MULTIPLE TESTING PROCEDURES AND SIMULTANEOUS INTERVAL ESTIMATES WITH THE INTERVAL PROPERTY

BY YINGQIU MA

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Statistics and Biostatistics

Written under the direction of

Arthur Cohen, Harold B. Sackrowitz

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

Oct, 2012

ABSTRACT OF THE DISSERTATION

Multiple Testing Procedures and Simultaneous Interval Estimates with the Interval Property

by YINGQIU MA

Dissertation Director: Arthur Cohen, Harold B. Sackrowitz

The use of step-wise multiple testing procedures arises because single step procedures are extremely conservative. Research into the construction of useful, computationally feasible interval estimates corresponding to step-wise procedures has been slow. We present an alternative method of constructing multiple testing procedures that easily admits corresponding interval estimates. The new approach has the desirable interval property not usually shared by step-wise procedures. Furthermore, the new method take dependency into account and easily carried out. In addition, these intervals are typically shorter, less likely to contain the null point falsely, and are more informative than those based on traditional methods. Applications include treatments vs control, change point and all pairwise comparisons. Examples and simulations are included.

Acknowledgements

I am grateful to my advisers, Dr. Arthur Cohen and Dr. Harold Sackrowitz, for their valuable guidance, insight and knowledge. I am also grateful to other committee members, Dr. Cun-Hui Zhang and Dr. Derek Gordon, for reading the manuscript and making a lot of helpful comments and suggestions.

Also I'd like to thank my parents and my husband for their continuous encouragement and support.

Table of Contents

List of Tables

Chapter 1

Introduction

Multiple testing models have become an important part of statistical applications. In the last two decades scores of multiple testing procedures (MTPs) have emerged to meet the call for applications in diverse fields such as microarrays, astronomy, genomics, bio-weapons use, mutual fund evaluations, proteomics, cytometry, blood analytes, imaging, school evaluations, and others.

Among the new MTPs, many are stepwise procedures. See for example Dudoit, Shaffer and Boldrick (2003), Hochberg and Tamhane (1987) and Lehmann and Romano (2005). Stepwise procedures are very valuable because they can be more powerful than single step procedures. Traditional single step procedures are conservative, because when the number of tests is large, critical values determined to control error rates in single step procedures are so large as to prevent detection of significant results.

In constructing stepwise testing procedures, it is common to begin with tests for the individual hypotheses that are known to have desirable properties. For example, the UMPU tests may have invariance properties and are likely to be admissible. Then a sequential rule is added that determines which hypotheses to accept or reject at each step and when to stop. Unfortunately, due to this sequential component, the stepwise procedures oftentimes do not retain all the desirable properties that the original test possessed.

We focus on an important practical and theoretical property called interval property. This property has been studied at length in Cohen and Sackrowitz (2012). Informally the interval property is simply that the resulting acceptance regions of each individual tests are all intervals. Suppose one is constructing a test for a two-sided hypothesis testing problem. Let $\mathbf{X} = (X_1, \ldots, X_M)$ be a sample point. There are often triples

of points, $\mathbf{X}, \mathbf{X}^*$ and $\mathbf{X}^{**}$ (on the same line) such that if both $\mathbf{X}$ and $\mathbf{X}^{**}$ are in the acceptance region then one would also want $\mathbf{X}^*$ to be in the acceptance region if in fact $\mathbf{X}^*$ was not the most indicative of the alternative of the three points. The interval property is necessary and sufficient for admissibility in many parametric models and intuitively desirable in both parametric and nonparametric models. This is a desirable property that the original tests might have but that the stepwise induced tests can easily lose.

In exponential family models, lack of the interval property not only means procedures exist with both better size and power for every individual hypothesis but it may also lead to very counterintuitive results. It is not desirable when testing an individual hypothesis, if relevant acceptance sections are not intervals. Because this procedure could yield a reject of a null hypothesis in one instance and then yield an accept of the same hypothesis in another instance when the evidence and intuition is more intuitively compelling in the latter case. It is a disturbing practical shortcoming of many of the usual stepwise procedures.

In a series of papers, Cohen and Sackrowitz (2005a), (2005b), (2007), (2008) and Cohen, Kolassa, and Sackrowitz (2007) have shown that many of the standard stepwise procedures under a wide variety of assumptions, do not have the interval property. It follows that often other such procedures are inadmissible for a variety of risk functions that involve both expected type I and expected type II errors. In particular, Cohen and Sackrowitz (2007) illustrated that for multivariate normal models with nonzero correlations, there exist procedures whose individual tests have smaller expected type I and type II errors.

In response to the inadmissibility property and the fact that many stepwise procedures are based on the marginal distributions of test statistics, even when they are statistically dependent, Cohen, Sackrowitz and Xu (CSX) (2009) recommend a new MTP method called maximum residual down (MRD) for models with dependent variables. MRD takes correlation into account and is admissible (in exponential family models) for a risk function that focuses on expected type I and type II errors.

Besides admissibility, consistency is also used to assess the quality of an MTP. The MRD method is not consistent for some models. Therefore, Chen, Cohen and Sackrowitz (2009) extended MRD to a new MTP called MRDSS, which adds a screening stage and a sign stage to MRD. When sample sizes are large, MRDSS is applicable to exponential family models and other models as well. MRDSS is admissible and consistent. Chen, Cohen and Sackrowitz (2011) show that the binary segmentation method (BSP) of Vostrikova (1981) and MRD of CSX(2009) are consistent for change point problems. The change point problem concerns all mean differences between any two consecutive means. That is, $H_{i,i+1} : \mu_i = \mu_{i+1}$ vs. $K_{i,i+1} : \mu_i \neq \mu_{i+1}$, for $i = 1, \ldots, k-1$. Chapter 2 of this thesis discusses the admissibility of MRDS, which is MRD with screening stage. We prove that it is also admissible for change point problems.

In addition to testing multiple hypotheses there is often a desire to obtain interval estimates for the parameters. When the number of parameters is large, typical simultaneous interval estimates such as Scheff*é*, Bonferroni, Tukey pairwise contrasts or Dunnett have excessively large lengths and are deemed too conservative because they are single step methods. See for example Miller (1966) for these latter methods. Unfortunately research into the construction of useful, computationally feasible interval estimates corresponding to step-wise procedures has been slow. A number of authors have commented on the difficulty of the problem of inverting stepwise MTPs, particularly in the case of two-sided hypotheses. These include Lehmann (1986, page 388), Stefansson, Kim, and Hsu (1988) and Benjamini and Stark (1996). Most attempts result in constructions that often lead to non-informative intervals as they contain the entire alternative space. For example, Stefansson et.al. (1988) give intervals for a one-sided treatment vs control model but (unless all the parameters are found to be significantly different from zero), the intervals are of the form $(0, \infty)$ for the significant parameters. As yet there is no general method of informative imultaneous interval estimates to accommodate a wide variety of models and a wide variety of problems represented in multiple testing situations.

MRD is a desirable MTP that takes correlation into account. However, we can't

get simultaneous interval estimates from MRD. In Chapter 3, we present an alternative method of constructing MTPs that easily admit corresponding interval estimates. The new methodology will be seen to be made up of a collection of individualized 2-stage processes each designed to test one of the hypotheses separately. The new method places greater focus on each hypothesis separately while still using all the data. We introduce our new approach in the setting of the classical two-sided multiple testing problem. An analogous development for 1-sided problems easily follows. We have M populations each with some parameter ν_i associated with it. From population i we would observe X_i whose marginal distribution depends only on ν_i. The X_i may or may not be independent. The hypotheses to be tested are $H_i : \nu_i = 0$ vs $K_i : \nu_i = 0$; $i = 1, \ldots, M$.

Not only do these new MTPs perform as well as commonly used step-wise procedures but they also have a practical interval property not usually shared by step-wise procedures. That is, acceptance regions have desirable interval properties. Furthermore, interval estimates associated with these tests are easily obtained. Like MRD, this method is also particularly effective in the dependent case. Its specific application in three common multiple testing models like treatment vs. control, change point problem and all pairwise comparisons are also described in detail. Simulations indicate that the new method does well in terms of false discovery rate (FDR) control and in terms of the total number of type I and type II mistakes. It performs better than MRD in sparse case. The new method has several advantages over the stepwise methods that are currently recommended in the literature. We list the following advantages of the interval estimates:

1. The estimated intervals are typically shorter than SCIs based on the Bonferroni, Scheffé or Tukey method when they are applicable.

2. They are less likely to contain zero falsely than other competitors.

3. They are informative, i.e., they are all finite, unlike some competitors often are $(0, \infty)$.

4. They are computationally easily obtained.

5. Both testing and interval procedures have an important interval property not shared by typical stepwise procedures.

Note that the interval property for simultaneous interval estimates, means that for some ordered sample points, say $\boldsymbol{X}, \boldsymbol{X}^*, \boldsymbol{X}^{**}$, suppose the interval for ν_i covers zero at $\boldsymbol{X}$. At $\boldsymbol{X}^*$ suppose the interval does not cover 0 because its center at $\boldsymbol{X}^*$ moves away from 0. Then at $\boldsymbol{X}^{**}$ where the center of the interval moves even further away from 0, it also does not cover 0.

In contrast to parametric models, the application of stepwise procedures to non-parametric statistics has been limited. The properties of these procedures have not been investigated in detail. Nonparametric multiple testing is discussed in Hochberg and Tamhane (1987). Campbell and Skillings (1985) explore various multiple rank tests for pairwise comparisons in a balanced one way layout without a normality assumption. The model is $Y_{ij} = \theta_i + \epsilon_{ij}$, $i = 1, \ldots, k$, $j = 1, \ldots, n_i$, where θ_i are unknowns and $\epsilon_{ij}{}'s$ are independent, identically distributed, continuous random variables. Hypotheses of interest are $H_{ij} : \theta_i = \theta_i$ vs $K_{ij} : \theta_i \neq \theta_i$ for every $i \neq j$ and $i, j = 1, \ldots, k$. That is, each population has the same distribution except for a translation parameter. All of the C_2^k pairwise differences in translation parameters are tested. They consider single step and stepwise procedures, finding that stepwise procedures have superior pairwise power compared to the commonly used single step procedures. Among the stepwise procedures are those that rerank at different steps as well as those that do not rerank. They recommend the ad hoc procedure labeled NAH.

Cohen and Sackrowitz (2012) propose a rank test method called RPADD+ for this model, the analogue of a method developed in Cohen, Sackrowitz and Chen (2010) for testing pairwise comparisons in a one way layout assuming normality called PADD+ (partitioned average difference down plus). RPADD+ has a desirable and intuitive interval property in terms of ranks that the ad hoc procedure does not have. For each individual pairwise hypothesis, relevant acceptance sections are intervals (not a

collection of disjoint sets) in terms of ranks. Simulations indicate that RPADD+ is comparable to the performance of the ad hoc procedure in terms of power.

Similarly as in the parametric models, we are still interested in the interval property of MTPs in terms of actual observations in nonparametric models. Ad hoc procedure NAH and RPADD+ do not have interval property in terms of the raw data. We are also interested in other pairwise differences of the parameters in testing other settings besides all pairwise comparisons, such as treatment vs. control and change point problem. We first explore several approaches such as step-down MTPs based on permutation test, bootstrap test and ranks, both on separate ranks and full ranks. We demonstrate that for $k = 2$, the two sample rank test, the two sample permutation test and the two sample bootstrap test have the interval property under certain conditions. For multiple testing, i.e. $k > 2$, the usual rank tests and permutation tests fail to have the interval property. Then we derive the new MTPs based on ranks which do have the interval property. The ranks are not the ranks of the original observations. The ranks are determined in a special way to ensure the interval property of MTPs. Simulations indicate that the new MTP procedures effectively control family wise error rate (FWER) and total number of type I and type II mistakes.

In addition to testing hypotheses regarding the difference in parameters, there is often a desire to obtain interval estimates for the difference in parameters in nonparametric models. Very little along these lines has been done in the past. This new nonparametric MTP provides an approach to simultaneous interval estimations. The methodology has its origins in the approach we propose for parametric models where simultaneous interval estimates are determined in a 2 stage process. The second stage of the process determines an interval estimate for each particular pairwise difference of interest. The length of each interval depends on the number of rejections made when testing all other pairwise parameters of interest in the first stage of the process. Test statistics for parameters in the first stage turn out to be two sample rank tests. However in comprising the two samples, data from more than 2 populations are used. The new method has several advantages over the stepwise methods that are currently

recommended in the literature. The interval has the following properties:

1. They are informative, i.e., they are all finite, unlike some competitors which often are $(0, \infty)$
2. They are computationally feasible
3. Both testing and interval procedures have an important interval property in terms of raw data not shared by typical stepwise procedures.

Chapter 2 discusses the admissibility of MRDS, which is MRD with screening stage. We proved that it is admissible for change point problems. Chapter 3 states a new method of constructing MTPs that easily admit corresponding interval estimates with the interval property. Chapter 4 shows the derivation of new nonparametric MTPs based on ranks that provide simultaneous interval estimates with the interval property in terms of observations. Simulations and analysis are given in Chapter 5. We present the comparisons of several MTPs including original MRD, Holm's step down and three new tests under the 2 stage framework proposed in Chapter 3. We also report the results a simulation study for the performance of nonparametric MTPs using the uniform, normal, exponential and double exponential distributions.

Chapter 2

Admissibility of MRDS in change point problems

This chapter considers the admissibility of MRD (maximum residual down) with a screening stage for the change point problems.

The change point problem considered here is as follows. Let $X_{ij}, i = 1, \ldots, M, j = 1, \ldots, n$ be independent normally distributed random variables with means μ_i and known variance σ^2. Consider the testing of consecutive mean change. That is, $H_i : \mu_i - \mu_{i+1} = 0$ vs. $K_i : \mu_i - \mu_{i+1} \neq 0, i = 1, \ldots, M-1$.

Cohen, Sackrowitz and Xu (2009) indicated that most standard stepwise MTPs (step-up, step-down and others) for the change point model are inadmissible for a vector risk function concerned with both Type I and Type II errors. On the other hand, the approach attributed to Vostrikova (1981) called binary segmentation procedure (BSP) and MRD procedure by Cohen, Sackrowitz and Xu (2009) are admissible.

The MRD procedures are described in detail in Chen, Cohen and Sackrowitz (2011).

Let $\bar{X}_i = \sum_{j=1}^{n} x_{ij}/n$ and $\bar{\mathbf{X}} = (\bar{X}_1, \ldots, \bar{X}_M)$. Let $I = \{1, 2, \ldots, M\}$. Let $B = B(t_1, t_2)$ be the subset of consecutive integers $t_1, t_1 + 1, \ldots, t_2$. Let $A(t_1, i) \subset B$ be the subset of B consisting of the consecutive integers $t_1, \ldots, i$. Next define

$$D_{\bar{\mathbf{x}}}(A(t_1, i); B) = n(\bar{\bar{X}}_A - \bar{\bar{X}}_{B\setminus A})^2/(\sigma^2(\frac{1}{i - t_1 + 1} + \frac{1}{t_2 - i})) \tag{2.0.1}$$

where

$$\bar{\bar{X}}_A = \sum_{j=t_1}^{i} \bar{X}_j/(i - t_1 + 1), \bar{\bar{X}}_{B\setminus A} = \sum_{j=i+1}^{t_2} \bar{X}_j/(t_2 - j).$$

Let

$$D^*_{\bar{\mathbf{x}}}(B) = \max_{\{t_1 \le i \le t_2\}} D_{\bar{\mathbf{x}}}(A(t_1, i); B).$$

Let $A(t_1, i^*)$ denote the set for which the max is attained. That is, $D^*_{\bar{\mathbf{x}}}(B) = D_{\bar{\mathbf{x}}}(A(t_1, i^*); B)$. Let $0 < C_1 < \ldots < C_{M-1}$ be constants.

At step 1 of MRD, some 2 set partitions of consecutive integers of I are considered. That is, $A(1, i(\bar{\mathbf{x}}))$, $I \backslash A(1, i(\bar{\mathbf{x}}))$, $i = 2, \ldots, M-1$. $D_{\bar{\mathbf{x}}}(A(1, i(\bar{\mathbf{x}})); I)$ is computed for $i = 2, \ldots, M-1$. If $D_1 = D^*_{\bar{\mathbf{x}}}(I) \le C_{M-1}$, stop and set $r = 0$. If $D_1 > C_{M-1}$ then partition I into $A(1, i^*(\bar{\mathbf{x}}))$ and $I \backslash A(1, i^*(\bar{\mathbf{x}}))$ and go to step 2 provided $i^* > 2$. If $i^* = 2$ just consider $I \backslash A(1, i^*(\bar{\mathbf{x}}))$ at step 2.

At step 2, each set $A(1, i^*(\bar{\mathbf{x}}))$ and $I \backslash A(1, i^*(\bar{\mathbf{x}}))$ is treated as I was at step 1 except now C_{M-1} is replaced by C_{M-2}. For $A(1, i^*(\bar{\mathbf{x}}))$ and $I \backslash A(1, i^*(\bar{\mathbf{x}}))$, either all hypotheses are accepted or one of the sets is split into 2 sets, leaving 3 sets to consider at step 3.

This process continues in succeeding steps with different constants. Once again at each step there can be no additional rejection or 1 rejection. The total number of rejections is r.

Chen, Cohen and Sackrowitz (2011) prove that MRD of CSX (2009) are consistent for change point problems.

Consistency is used to assess the quality of an MTP. There are different definitions of consistency. Some consider the accuracy of selection results when the number of variables goes to infinity and others focus on the accuracy when sample size is large. Our definition of consistency is that an MTP is consistent if the probability of not making mistakes, either type I or type II error, for each individual hypothesis goes to zero for sufficiently large sample size.

Consistency is a large sample property. For problems with fewer sample sizes in change point problems, sometimes the FWER of MRD can be large. To control FWER

at a reasonable level, MRD with a screening stage (MRDS) is considered as a remedy. Simulations show that MRDS is effective in controlling the FWER while maintaining desirable power.

2.1 MRD with a screening stage

Chen, Cohen and Sackrowitz (2009) extended MRD to a new MTP called MRDSS, which adds a screening stage and a sign stage to MRD. MRDSS is admissible and consistent. When sample sizes are large, MRDSS is applicable to exponential family models and other models as well.

Now we are interested in change point problems. In particular. We want to know if MRDS is admissible and consistent. If this is the case, we could simplify the procedures without adding a sign stage after screening stage for this type of problem.

We will describe the screening stage as in Chen, Cohen and Sackrowitz (2009). For change point problems, to add a screening stage to MRD, let $C_U > C_L > 0$ be two additional constants. Typically $C_L \leq C_M < C_U$. Note that $C_1 > C_2 > \ldots > C_M$ are the set of critical values used in the MRD stage. After MRD is done, each hypothesis is temporarily accepted or rejected.

Let $H_{j_1}, \ldots, H_{j_p}$ be those hypotheses that MRD rejected. Should any

$$|\overline{x}_{j_i+1} - \overline{x}_{j_i}|/(\sigma\sqrt{\frac{2}{n}}) < C_L$$

$i = 1, \ldots, p$, then reverse the reject decision to an accept decision.

For those hypotheses that MRD accepted, say, $H_{j_{p+1}}, \ldots, H_{j_M}$, Should any

$$|\overline{x}_{j_i+1} - \overline{x}_{j_i}|/(\sigma\sqrt{\frac{2}{n}}) > C_U$$

$i = p+1, \ldots, M$, then reverse the accept decision to a rejection decision.

MRD with screening stage for change point

Theorem 2.1.1. *MRD with a screening stage is admissible for change point problems.*

Proof. Suppose that MRD with a screening stage for change point problem is inadmissible. Let $\psi_{i,i+1}(\bar{\mathbf{x}})$ be the test of H_i determined by MRDS. Suppose $\psi_{i,i+1}(\bar{\mathbf{x}})$ is inadmissible. Let $\mathbf{g} = (0, \ldots, 1, -1, 0, \ldots, 0)'$ are in the i and $i+1$ positions. Then there exists three points $\bar{\mathbf{x}}, \bar{\mathbf{x}}^* = \bar{\mathbf{x}} + r_1\mathbf{g}, \bar{\mathbf{x}}^{**} = \bar{\mathbf{x}} + r_2\mathbf{g}$ with $0 < r_1 < r_2$, such that $\bar{\mathbf{x}}$ and $\bar{\mathbf{x}}^{**}$ are accept points and $\bar{\mathbf{x}}^*$ is a reject point. This can occur in 8 ways. They are listed in the Table 2.1.1. Without loss of generality, we assume $\sigma\sqrt{\frac{2}{n}} = 1$.

Sample points	$\bar{\mathbf{x}}$	$\bar{\mathbf{x}}^*$	$\bar{\mathbf{x}}^{**}$	$\bar{\mathbf{x}}$	$\bar{\mathbf{x}}^*$	$\bar{\mathbf{x}}^{**}$
	Case 1			**Case 2**		
Actions of $\psi_{i,i+1}$	A	A	A	A	A	R
$\mid \bar{x}_i - \bar{x}_{i+1} \mid, \mid \bar{x}_i^* - \bar{x}_{i+1}^* \mid, \mid \bar{x}_i^{**} - \bar{x}_{i+1}^{**} \mid$	$< C_U,$	$> C_U,$	$< C_U$	$< C_U,$	$> C_U,$	$< C_L$
	Case 3			**Case 4**		
Actions of $\psi_{i,i+1}$	A	R	A	A	R	R
$\mid \bar{x}_i - \bar{x}_{i+1} \mid, \mid \bar{x}_i^* - \bar{x}_{i+1}^* \mid, \mid \bar{x}_i^{**} - \bar{x}_{i+1}^{**} \mid$	$< C_U,$	$> C_L,$	$< C_U$	$< C_U,$	$> C_L,$	$< C_L$
	Case 5			**Case 6**		
Actions of $\psi_{i,i+1}$	R	A	A	R	A	R
$\mid \bar{x}_i - \bar{x}_{i+1} \mid, \mid \bar{x}_i^* - \bar{x}_{i+1}^* \mid, \mid \bar{x}_i^{**} - \bar{x}_{i+1}^{**} \mid$	$< C_L,$	$> C_U,$	$< C_U$	$< C_L,$	$> C_U,$	$< C_L$
	Case 7			**Case 8**		
Actions of $\psi_{i,i+1}$	R	R	A	R	R	R
$\mid \bar{x}_i - \bar{x}_{i+1} \mid, \mid \bar{x}_i^* - \bar{x}_{i+1}^* \mid, \mid \bar{x}_i^{**} - \bar{x}_{i+1}^{**} \mid$	$< C_L,$	$> C_L,$	$< C_U$	$< C_L,$	$> C_L,$	$< C_L$

Table 2.1.1: Possible behaviors leading to ARA for $\psi_{i,i+1}$

F1: As a function of r, $|\bar{x}_i - \bar{x}_{i+1} + 2r|$ is strictly decreasing for $r < (\bar{x}_{i+1} - \bar{x}_i)/2$ and strictly increasing for $r > (\bar{x}_{i+1} - \bar{x}_i)/2$.

Case 1,2,5,6 and 8 require $|\bar{x}_i - \bar{x}_{i+1}| \leq |\bar{x}_i^* - \bar{x}_{i+1}^*| \geq |\bar{x}_i^{**} - \bar{x}_{i+1}^{**}|$ at screen stage, but this would violate the F1.

Case 3 is impossible because MRD without screen stage is admissible.

In case 4, $|\bar{x}_i^* - \bar{x}_{i+1}^*| > C_L > |\bar{x}_i^{**} - \bar{x}_{i+1}^{**}|$. This indicates that $\bar{x}_i^* < \bar{x}_{i+1}^*$ because $\bar{x}_i^* = \bar{x}_i + r_1, \bar{x}_i^{**} = \bar{x}_i + r_2$ and $\bar{x}_{i+1}^* = \bar{x}_{i+1} - r_1, \bar{x}_{i+1}^{**} = \bar{x}_{i+1} - r_2$ where $0 < r_1 < r_2$, therefore $|\bar{x}_i^* - \bar{x}_{i+1}^*| = |\bar{x}_i - \bar{x}_{i+1} + 2r_1|$ and $|\bar{x}_i^{**} - \bar{x}_{i+1}^{**}| = |\bar{x}_i - \bar{x}_{i+1} + 2r_2|$, so it follows from F1 that $0 < r_1 < (\bar{x}_{i+1} - \bar{x}_i)/2$ and thus $\bar{x}_i^* = \bar{x}_i + r_1 < \bar{x}_i + (\bar{x}_{i+1} - \bar{x}_i)/2 = (\bar{x}_i + \bar{x}_{i+1})/2 = \bar{x}_{i+1} - (\bar{x}_{i+1} - \bar{x}_i)/2 < \bar{x}_{i+1} - r_1 = \bar{x}_{i+1}^*$.

Without loss of generality, we can denote $\sum_{j=t_1}^{i-1} \bar{x}_j^* = A, \bar{x}_i^* = -b, \bar{x}_{i+1}^* = b, \sum_{j=i+2}^{t_2} \bar{x}_j^* = D$ where $b > 0$ and set $m_1 = i - t_1 + 1$ and $m_2 = t_2 - i$. Since $\bar{\mathbf{x}}$ is an accept point of $\psi_{i,i+1}$ and $\bar{\mathbf{x}}^*$ is a reject point of $\psi_{i,i+1}$, $\psi_{i,i+1}$ is increasing from $\bar{\mathbf{x}}$ to $\bar{\mathbf{x}}^*$, i.e. $|\frac{\sum_{j=t_1}^{i} \bar{x}_j}{m_1} - \frac{\sum_{j=i+1}^{t_2} \bar{x}_j}{m_2}| < |\frac{\sum_{j=t_1}^{i} \bar{x}_j^*}{m_1} - \frac{\sum_{j=i+1}^{t_2} \bar{x}_j^*}{m_2}|$. $|\frac{A - b - r_1}{m_1} - \frac{D + b + r_1}{m_2}| < |\frac{A - b}{m_1} - \frac{D + b}{m_2}|$. Thus, $\frac{A - b}{m_1} - \frac{D + b}{m_2} > 0$.

Consider

(1)

$$\frac{(\frac{A}{m_1 - 1} - \frac{D}{m_2 + 1})^2}{\frac{1}{m_1 - 1} + \frac{1}{m_2 + 1}}$$

(2)

$$\frac{(\frac{A - b}{m_1} - \frac{D + b}{m_2})^2}{\frac{1}{m_1} + \frac{1}{m_2}}$$

(3)

$$\frac{(\frac{A}{m_1+1}-\frac{D}{m_2-1})^2}{\frac{1}{m_1+1}+\frac{1}{m_2-1}}$$

(1),(2) and (3) satisfy

$$(1)+(3)>2\times(2)$$

because

$$(1)+(3)>2\times\frac{(\frac{A}{m_1}-\frac{D}{m_2})^2}{\frac{1}{m_1}+\frac{1}{m_2}}$$

and

$$\frac{(\frac{A}{m_1}-\frac{D}{m_2})^2}{\frac{1}{m_1}+\frac{1}{m_2}}>(2)$$

because

$$\frac{A}{m_1}-\frac{D}{m_2}>\frac{A-b}{m_1}-\frac{D+b}{m_2}>0$$

Thus either (1) or (3) is larger than (2). This implies that H_i will not be rejected at $\bar{\mathbf{x}}^*$ since it is not the maximum test statistic at each stage until only $\bar{x}_i^*, \bar{x}_{i+1}^*$ is left to test. In that special case, the test statistic is the same as that at screen stage, so we can't inverse rejection to acceptance at the screening stage.

In case 7, $|\bar{x}_i - \bar{x}_{i+1}| < C_L < |\bar{x}_i^* - \bar{x}_{i+1}^*|$. This indicates that $\bar{x}_i^* > \bar{x}_{i+1}^*$. Because it follows from F1 that $r_1 > (\bar{x}_{i+1} - \bar{x}_i)/2$ and thus $\bar{x}_i^* = \bar{x}_i + r_1 > \bar{x}_i + (\bar{x}_{i+1} - \bar{x}_i)/2 = (\bar{x}_i + \bar{x}_{i+1})/2 = \bar{x}_{i+1} - (\bar{x}_{i+1} - \bar{x}_i)/2 > \bar{x}_{i+1} - r_1 = \bar{x}_{i+1}^*$.

Without loss of generality, let $\sum_{j=t_1}^{i-1} \bar{x}_j^* = A$, $\bar{x}_i^* = b$, $\bar{x}_{i+1}^* = -b$ and $\sum_{j=i+2}^{t_2} \bar{x}_j^* = D$

where $b > 0$. Since $\bar{\mathbf{x}}^*$ is an reject point of $\psi_{i,i+1}$ and $\bar{\mathbf{x}}^{**}$ is a accept point of $\psi_{i,i+1}$, $\psi_{i,i+1}$ is decreasing from $\bar{\mathbf{x}}^*$ to $\bar{\mathbf{x}}^{**}$, i.e. $|\frac{\sum_{j=t_1}^{i} \bar{x}_j^*}{m_1} - \frac{\sum_{j=i+1}^{t_2} \bar{x}_j^*}{m_2}| > |\frac{\sum_{j=t_1}^{i} \bar{x}_j^{**}}{m_1} - \frac{\sum_{j=i+1}^{t_2} \bar{x}_j^{**}}{m_2}|$. $|\frac{A+b}{m_1} - \frac{D-b}{m_2}| > |\frac{A+b+(r_2-r_1)}{m_1} - \frac{D-b-(r_2-r_1)}{m_2}|$. Thus, $\frac{A+b}{m_1} - \frac{D-b}{m_2} < 0$.

(1) and (3) are the same as in case 4, just (2) becomes

(4)

$$\frac{(\frac{A+b}{m_1} - \frac{D-b}{m_2})^2}{\frac{1}{m_1} + \frac{1}{m_2}}$$

(1),(4) and (3) satisfy $(1) + (3) > 2 \times (4)$ because

$$(1) + (3) > 2 \times \frac{(\frac{A}{m_1} - \frac{D}{m_2})^2}{\frac{1}{m_1} + \frac{1}{m_2}}$$

and

$$\frac{(\frac{A}{m_1} - \frac{D}{m_2})^2}{\frac{1}{m_1} + \frac{1}{m_2}} > (4)$$

because

$$\frac{A}{m_1} - \frac{D}{m_2} < \frac{A+b}{m_1} - \frac{D-b}{m_2} < 0$$

Similarly as in case 4, H_i will not be rejected at $\bar{\mathbf{x}}^*$ since it is not the maximum test statistic at each stage until only $\bar{x}_i^*, \bar{x}_{i+1}^*$ is left to test. In that special case, the test statistic is the same as that at screen stage, so we can't inverse rejection to acceptance at the screening stage. □

Chapter 3

Multiple testing procedures

In this chapter, we propose a new construction of MTPs having the interval property that also results in interval estimates. We will present and demonstrate the method by applying it to some common models of one dimensional normal means of variables that may or may not be correlated. The two basic models in this chapter are as follows:

I. Consider M independent normal populations with means μ_i, $i = 1, \ldots, M$ and common known variance σ^2. We consider various models involving collections of pairwise differences $\mu_i - \mu_j$.

II. Let $\mathbf{X}_\alpha$, $\alpha = 1, \ldots, M$ be multivariate normal with mean vector $\boldsymbol{\mu}$ and known covariance $\boldsymbol{\Sigma}$. We seek MTPs and interval estimates for the components of $\boldsymbol{\mu}$. An important special case is the general linear model $\mathbf{y} = A\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I)$. Then the vector of least squares estimators is normal with mean vector $\boldsymbol{\beta}$ and covariance matrix $\sigma^2(A'A)^{-1}$. We seek MTPs and interval estimates for the components of $\boldsymbol{\beta}$.

Section 3.1 will introduce the construction of the individualized 2-stage processes that leads to our new class of MTPs with corresponding interval estimates. We show that all typical step-down and step-up MTP procedures have equivalent representations belonging to this class. In Section 3.2, we will apply the method to the above mentioned models.

3.1 Description of new construction and first principles

Suppose we want to test the collection of hypotheses $H_i : \nu_i = 0$ vs $K_i : \nu_i \neq 0,\ i = 1, ..., M$. First we describe the generic step-down process and step-up process in a fashion similar to Lehmann and Romano (2005). That will be followed by a description of the new individualized 2-stage process.

For the stepwise procedures we suppose that the individual hypothesis H_i has a test based on the test statistic T_i with large values indicating evidence for K_i.

1. Step-down process. Fix constants $C_1 < ... < C_M$.

Step 1: Consider $U_{i_1} = \max_{\{1 \leq i \leq M\}} T_i$. If $U_{i_1} \leq C_M$, stop and accept all H_i. If $U_{i_1} > C_M$ reject H_{i_1} and go to step 2.

Step 2: Consider $U_{i_2} = \max_{\{1 \leq i \leq M, i \neq i_1\}} T_i$. If $U_{i_2} \leq C_{M-1}$ stop and accept all $H_i : i \in \{1, \ldots, M\} \backslash \{i_1\}$. If $U_{i_2} > C_{M-1}$ reject H_{i_2} and go to step 3.

Step m: Consider $U_{i_m} = \max_{i \in \{1,...,M\} \backslash \{i_1, i_2, ..., i_{m-1}\}} T_i$. If $U_{i_m} \leq C_{M-m+1}$ stop and accept all $H_i : i \in \{1, \ldots, M\} \backslash \{i_1, \ldots, i_{m-1}\}$. If $U_{i_m} > C_{M-m+1}$ reject H_{i_m} and go to step $(m+1)$.

2. Step-up process. Fix constants $C_1 > ... > C_M$.

Step 1: Consider $U_{i_1} = \min_{\{1 \leq i \leq M\}} T_i$. If $U_{i_1} \geq C_M$, stop and reject all H_i. If $U_{i_1} < C_M$ accept H_{i_1} and go to step 2.

Step 2: Consider $U_{i_2} = \min_{\{1 \leq i \leq M, i \neq i_1\}} T_i$. If $U_{i_2} \geq C_{M-1}$ stop and reject all $H_i : i \in \{1, \ldots, M\} \backslash \{i_1\}$. If $U_{i_2} < C_{M-1}$ accept H_{i_2} and go to step 3.

Step m: Consider $U_{i_m} = \min_{i \in \{1,\ldots,M\} \backslash \{i_1, i_2, \ldots, i_{m-1}\}} T_i$. If $U_{i_m} \geq C_{M-m+1}$ stop and reject all $H_i : i \in \{1, \ldots, M\} \backslash \{i_1, \ldots, i_{m-1}\}$. If $U_{i_m} < C_{M-m+1}$ accept H_{i_m} and go to step $(m+1)$.

Next we introduce a new method for constructing MTPs that immediately yields corresponding interval estimates.

3. Individualized 2-stage process.

All hypotheses are treated similarly but they are considered one at a time. We can begin with testing H_1 by subjecting it to the following two stage process. At stage 1 apply any (M-1 population) MTP to **ONLY** the other hypotheses, i.e., $H_i : \nu_i = 0$ vs $K_i : \nu_i \neq 0,\ i = 2, ..., M$, but using all the data. At the end of this stage, $r_1 =$ the number of rejections is recorded.

At stage 2 we construct intervals of the form

$$\widehat{\nu_1} \pm B(r_1)\sigma_{\widehat{\nu_1}}$$

where $B(r_1)$ is a decreasing function of r_1 and $\widehat{\nu_1}$ can depend on all the data. The corresponding test of H_1 is to reject if the interval does not contain 0. Typically this would mimic, except for the use of r_1, what would have been done if only the one population had been observed.

This process is repeated for each hypothesis. Due to the flexibility in choices of the modified MTP used at stage 1, the choice of $\widehat{\nu_1}$ and the choice of function $B(\cdot)$, this process generates a large family of MTP's with corresponding interval estimates. Every MTP in this family is associated with an easily obtained interval estimate. We point out that a natural choice of $\widehat{\nu_1}$ will often depend on only the data involving the parameter ν_1. This is the case in our examples.

Two MTPs are said to be equivalent if, with probability one, they make the same decisions for all the hypotheses in the collection being tested. The following theorem indicates the versatility of the new construction.

Theorem 3.1.1. *The class of MTPs made up of individualized 2-stage tests contains MTPs equivalent to the generic step-down and step-up procedures.*

Proof. Consider the generic step-down procedure using the statistics $T_j, j = 1, ..., M$ and constants $C_1 < C_2 < ... < C_M$. We focus on the decision for testing any individual hypothesis H_{j*}. Without loss of generality we can study testing H_1. To do this let $V_{(1)} < ... < V_{(M-1)}$ be the order statistics of $T_2, ..., T_M$.

Note that the family of individualized 2-stage procedures includes the following MTP. At stage 1 use the generic step-down based on $T_2, ..., T_M$ using the constants $C_2 < C_3 < ... < C_M$. At stage 2 reject H_1 if and only if $T_1 > C_{M-r}$ where $r = 0, 1, ..., M-1$ is the number of rejections at stage 1 among the $M-1$ tests performed at stage 1.

Clearly if $C_M < T_1$ both methods will reject H_1 and if $T_1 \leq C_1$ both methods will accept H_1. Thus we suppose $C_i < T_1 \leq C_{i+1}$ for some $i = 1, ..., M-1$. Based on the above definitions we have 1) The step-down procedure will reject H_1 if and only if

$$V_{(M-1)} > C_M, \ldots, V_{(i)} > C_{i+1}$$

2) When using the individualized 2-stage procedure and $C_i < T_1 \leq C_{i+1}$, H_1 will be rejected if and only if $r \geq M - i$. However, getting at least $M - i$ rejections at stage 1 is equivalent to (as in case (1)),

$$V_{(M-1)} > C_M, \ldots, V_{(i)} > C_{i+1}.$$

This completes the proof for the generic step-down MTP. The proof for the generic step-up MTP is similar. □

Our next concern is that our tests and interval estimates have the interval property.

Whether or not a particular procedure has the interval property depends on the model for which it is being used. For example, it can be seen in Cohen and Sackrowitz (CS) (2008) that in the case of dependence for the test statistics, the commonly used stepwise procedures often do not have the interval property. We will next establish sufficient conditions for an individualized 2-stage test to have the interval property.

Suppose we desire a test, ϕ_i, of H_i to have the interval property with respect to the direction $\mathbf{g}_i$. This will mean that, as a function of λ, $\phi_i(\mathbf{X} + \lambda\mathbf{g}_i) = 0$ only on an interval. The next two results, Theorem 2.2 and Lemma 2.3 that we will present are, essentially immediate consequences of definitions. However they are quite useful in finding MTPs having the interval property. We will state them without proof. The following theorem gives sufficient conditions for the new tests and interval estimates to have the interval property. In succeeding sections we will demonstrate the identification of appropriate vectors $\mathbf{g}$ in a number of different models.

Theorem 3.1.2. *Suppose for each fixed $\boldsymbol{x}, r_i(\boldsymbol{x} + \lambda\boldsymbol{g}_i)$ is constant as a function of λ and $\widehat{\nu}_i(\boldsymbol{x} + \lambda\boldsymbol{g}_i)$ is first non-increasing and then non-decreasing as a function of λ. Then the test and interval estimate will have the interval property with respect to the direction $\boldsymbol{g}_i$.*

Suppose an individualized 2-stage test is to have the interval property in the direction $\mathbf{g}$. The following simple lemma serves to demonstrate the ease in finding modified MTPs to be used at stage 1 that will guarantee the condition on $r(\cdot)$ in Theorem 3.1.2.

Lemma 3.1.1. *If every statistic used during the stage 1 modified MTP is based on functions of the form $\boldsymbol{a}'\boldsymbol{X}$, i.e., each is a linear combination of the $\boldsymbol{X}$'s, where $\boldsymbol{a}'\boldsymbol{g} = 0$ for each $\boldsymbol{a}$ then $r(\boldsymbol{X} + \lambda\boldsymbol{g}_i)$ is constant as a function of λ.*

3.2 Pairwise comparisons of normal means

3.2.1 Treatment vs Control

Let X_{ij}, $i = 1, \ldots, M$, $j = 1, \ldots, n$ be independent normal variables with means μ_i and known variance σ^2. For $i = 1$, the population represents a control, while $i = 2, \ldots, M$ represent treatment populations. We seek simultaneous interval estimates for $\nu_i = \mu_i - \mu_1$, $i = 2, \ldots, M$. Let $\bar{X}_i = \sum_{j=1}^{n} X_{ij}/n$. For ν_2 the confidence interval at stage 2 is

$$(\bar{X}_2 - \bar{X}_1) \pm B(r)\sigma\sqrt{\frac{2}{n}} \tag{3.2.1}$$

$r = 0, \ldots, M-2$, where $B(r)$ is sequence of decreasing constants depending on $r =$ number of rejections of hypotheses $H_i : \nu_i = 0$ vs $K_i : \nu_i \neq 0$, $i = 3, \ldots, M$. At step 1 hypotheses are tested by the RSD method of CS (2012).

This MTP method is as follows: Let $0 < C_1 < \ldots < C_{M-2}$ be constants.

Step 1: Let $t_M = (\bar{X}_1 + \ldots + \bar{X}_M)/M$. Let

$$U_{i_1} = \max_{3 \leq i \leq M} |\bar{X}_i - t_M| / (\frac{\sigma}{\sqrt{n}} \sqrt{(M+1)/M}) \tag{3.2.2}$$

If $U_{i_1} \leq C_{M-2}$, stop and set $r = 0$. If $U_{i_1} > C_{M-2}$ go to step 2.

At step 2: consider

$$U_{i_2} = \max_{\{3 \leq i \leq M\} \setminus \{i_1\}} |\bar{X}_i - t_{M-1}| / (\frac{\sigma}{\sqrt{n}} \sqrt{M/(M-1)})$$

where

$$t_{M-1} = \sum_{i \in \{1 \leq i \leq M\} \setminus \{i_1\}} \bar{X}_i/(M-1)$$

If $U_{i_2} \leq C_{M-3}$ stop and set $r = 1$. If $U_{i_2} > C_{M-3}$ go to step 3.

At step p, consider

$$U_{i_p} = \max_{\{3\leq i\leq M\}\setminus\{i_1,\ldots,i_{p-1}\}} |\bar{X}_i - t_{M-p+1}|/(\frac{\sigma}{\sqrt{n}}\sqrt{(M-p+2)/(M-p+1)}) \qquad (3.2.3)$$

If $U_{i_p} \leq C_{M-p-1}$ stop and set $r = p-1$. If $U_{i_p} > C_{M-p-1}$ go to step $(p+1)$.

Once r is finalized and the interval estimate for ν_2 is determined the process is repeated for the other ν_i.

Note that the above interval estimate will have the interval property. This will follow since r, the number of rejections determined by the MTP part of the process will not change as $\bar{X}_2$ increases and $\bar{X}_1$ decreases by the same amount. This follows because of the statistics defined in U_{i1},U_{i_2} and U_{i_p}.

Should σ^2 be unknown replace σ^2 with s^2, where s^2 is the usual unbiased estimator of σ^2.

An alternative to the method at stage one, called a shortcut method is as follows:
Step 1: Let

$$U_{i_1} = \max_{3\leq i\leq M} |\bar{X}_i - (\bar{X}_1 + \bar{X}_2)/2|/(\frac{\sigma}{\sqrt{n}}\sqrt{3/2}) = \max_{3\leq i\leq M} |W_i|.$$

If $U_{i_1} \leq C_{M-2}$, stop and set $r = 0$. If $U_{i_1} > C_{M-2}$ go to step 2.

At step 2, consider

$$U_{i_2} = \max_{\{3\leq i\leq M\}\setminus\{i_1\}} |W_i|.$$

If $U_{i_2} \leq C_{M-3}$, stop and set $r = 1$. If $U_{i_2} > C_{M-2}$ go to step 3.

At step p, consider

$$U_{i_p} = \max_{\{3 \le i \le M\} \setminus \{i_1, \ldots, i_{p-1}\}} |W_i|. \tag{3.2.4}$$

If $U_{i_p} \le C_{M-p-1}$ stop and set $r = p - 1$. If $U_{i_p} > C_{M-p-1}$ go to step $(p + 1)$.

Again, once r = number of rejections is determined then the confidence interval of ν_2 is determined. The process is repeated for the other ν_i's.

3.2.2 Change point

Let X_{ij}, $i = 1, \ldots, M$, $j = 1, \ldots, n$ be independent normal variables with means μ_i and known variance σ^2. We seek simultaneous interval estimates for $\nu_i = \mu_{i+1} - \mu_i$, $i = 1, \ldots, M - 1$. Let $B(r)$ be a decreasing set of constants and let $\bar{X}_i = \sum X_{ij}/n$. For ν_1 the confidence interval determined at stage 2 is

$$\bar{X}_2 - \bar{X}_1 \pm B(r)\sigma\sqrt{\frac{2}{n}} \tag{3.2.5}$$

$r = 0, 1, \ldots, M - 2$ where r = number of rejections of hypotheses $H_i : \nu_i = 0$ vs $K_i : \nu_i \neq 0$, $i = 2, \ldots, M - 1$.

At stage 1, the hypotheses are tested by the RSD method of CS (2012) which in this case is spelled out in more detail in Chen, Cohen, and Sackrowitz (2011). For the MTP part of the overall procedure let $I = \{1, 2, \ldots, M\}$. Let $B = B(t_1, t_2)$ be the subset of consecutive integers $t_1, t_1 + 1, \ldots, t_2$. Let $A(t_1, i) \subset B$ be the subset of B consisting of the consecutive integers $t_1, \ldots, i$. Next define

$$D_{\bar{\mathbf{x}}}(A(t_1, i); B) = n(\bar{\bar{X}}_A - \bar{\bar{X}}_{B \setminus A})^2 / (\sigma^2 (\frac{1}{i - t_1 + 1} + \frac{1}{t_2 - i})) \tag{3.2.6}$$

where

$$\bar{\bar{X}}_A = \sum_{j=t_1}^{i} \bar{X}_j/(i - t_1 + 1), \bar{\bar{X}}_{B \setminus A} = \sum_{j=i+1}^{t_2} \bar{X}_j/(t_2 - j).$$

Let

$$D^*_{\bar{\mathbf{x}}}(B) = \max_{\{t_1 \le i \le t_2\}} D_{\bar{\mathbf{x}}}(A(t_1, i); B).$$

Let $A(t_1, i^*(\bar{\mathbf{x}}))$ denote the set for which the max is attained. That is, $D^*_{\bar{\mathbf{x}}}(B) = D_{\bar{\mathbf{x}}}(A(t_1, i^*); B)$. Let $0 < C_1 < \ldots < C_{M-1}$ be constants.

At step 1 of the MTP part, some 2 set partitions of consecutive integers of I are considered. That is, $A(1, i(\bar{\mathbf{x}}))$, $I \backslash A(1, i(\bar{\mathbf{x}}))$, $i = 2, \ldots, M-1$. $D_{\bar{\mathbf{x}}}(A(1, i(\bar{\mathbf{x}})); I)$ is computed for $i = 2, \ldots, M-1$. If $D_1 = D^*_{\bar{\mathbf{x}}}(I) \le C_{M-1}$, stop and set $r = 0$. If $D_1 > C_{M-1}$ then partition I into $A(1, i^*(\bar{\mathbf{x}}))$ and $I \backslash A(1, i^*(\bar{\mathbf{x}}))$ and go to step 2 provided $i^* > 2$. If $i^* = 2$ just consider $I \backslash A(1, i^*(\bar{\mathbf{x}}))$ at step 2.

At step 2, each set $A(1, i^*(\bar{\mathbf{x}}))$ and $I \backslash A(1, i^*(\bar{\mathbf{x}}))$ is treated as I was at step 1 except now C_{M-1} is replaced by C_{M-2}. For $A(1, i^*(\bar{\mathbf{x}}))$ and $I \backslash A(1, i^*(\bar{\mathbf{x}}))$, either all hypotheses are accepted or one of the sets is split into 2 sets, leaving 3 sets to consider at step 3.

This process continues in succeeding steps with different constants. Once again at each step there can be no additional rejection or 1 rejection. The total number of rejections is r.

Note we described the MTP portion of the procedure in order to derive the interval estimate for ν_1. When we seek a interval estimate for ν_i, $i \neq 1$ then we need to consider two set partitions of consecutive integers of I but always keeping integers i and $(i+1)$ in the same set. For example if $M = 5$, $I = \{1, 2, 3, 4, 5\}$ and we seek a interval estimate for $\nu_4 - \nu_3$ then at step 1 the two set partitions are $(\{1\}, \{2, 3, 4, 5\})$, $(\{1, 2\}, \{3, 4, 5\})$ and $(\{1, 2, 3, 4\}, \{5\})$. Now we proceed as before always keeping i and $(i+1)$ together in the same set of any future partition of a subset in which i or $(i+1)$ appears. Again r is the total number of rejected hypotheses.

3.2.3 All pairwise

Let X_{ij}, $i = 1, \ldots, M$, $j = 1, \ldots, n$ be independent normal variables with means μ_i and known variance σ^2. We seek simultaneous interval estimates for $\nu_{ij} = \mu_i - \mu_j$, $i < j$, $i, j = 1, \ldots, M$.

At stage 2 the confidence interval for ν_{12} is

$$(\bar{X}_1 - \bar{X}_2) \pm B(r)\sigma\sqrt{\frac{2}{n}} \tag{3.2.7}$$

where r is the number of rejections of hypotheses $H_{ij} : \nu_{ij} = 0$ vs $K_{ij} : \nu_{ij} \neq 0$, ν_{12} is not included, $i < j$, $i, j = 1, \ldots, M$.

At stage 1 the hypotheses are to be tested by a modification of the PADD or PADD+ method of Cohen, Sackrowitz, and Chen (2010). To describe the method and modification, let $I = \{1, \ldots, \mathrm{M}\}$. For any subset of integers $A \subset I$, let $N(A) =$ the number of points in A. Let $\bar{X}_A = \sum_{i \epsilon A} \bar{X}_i / \mathrm{N(A)}$. Next define for all $A \subset H \subseteq I$ with $A \neq \phi$ and $H \backslash A \neq \phi$, for each sample point $(\bar{\mathbf{X}})$,

$$D_{\bar{\mathbf{x}}}(A; H) = (\bar{X}_A - \bar{X}_{H \backslash A}) / \sigma[(1/N(A) + 1/N(H \backslash A))^{1/2}] \tag{3.2.8}$$

and

$$D^*_{\bar{\mathbf{x}}}(H) = \max_{A \subset H} D_{\bar{\mathbf{x}}}(A; H) \tag{3.2.9}$$

Thus $D^*_{\bar{\mathbf{x}}}(H)$ is the largest possible standardized difference in subset means when the set of $\{\bar{X}_i : i \in H\}$ is broken into two non-empty subsets whose union is $\{\bar{X}_i : i \in H\}$. We further let $V_{\bar{\mathbf{x}}}(H)$ denote the set for which the maximum is attained. That is, $D^*_{\bar{\mathbf{x}}}(H) = D_{\bar{\mathbf{x}}}(V_{\bar{\mathbf{x}}}(H); H)$ when H is split into $V_{\bar{\mathbf{x}}}(H)$ and $H \backslash V_{\bar{\mathbf{x}}}(H)$.

At the first step of PADD all non-empty 2 set partitions of I are considered. For the modification here all 2 set partitions are considered except that indices 1 and 2 can

never be in separate sets. $D_{\bar{\mathbf{x}}}(A; I)$ is computed for all non-empty $A \subset I$. Let $\lambda_j =$ number of indices in the largest set of the partition at step j. Let C_{λ_1} be a constant at step 1 and let $D_1 = D^*_{\bar{\mathbf{x}}}(I)$. If $D_1 \leq C_{\lambda_1}$ then partition I into $V_{\bar{\mathbf{x}}}(I)$ and $I \backslash V_{\bar{\mathbf{x}}}(I)$ and continue to step 2.

At each successive step, until the procedure stops, one of the sets in the current partition will be split into 2 sets as follows: Suppose that after step m, I had been partitioned into $(H_1, H_2, \ldots, H_{m+1})$ and we continue. Recall indices 1 and 2 are never to be separated, i.e., they always lie in the same set of any partition. Compute

$$D_{m+1} = \max_{1 \leq k \leq m+1} D^*_{\bar{\mathbf{x}}(H_k)}$$

If $D_{m+1} \leq C_{\lambda_{m+1}}$ we stop. If $D_{m+1} > C_{\lambda_{m+1}}$ find k^* so that $D_{m+1} = D^*_{\bar{\mathbf{x}}(H_{k^*})}$. Next break H_{k^*} into $V_{\bar{\mathbf{x}}}(H_{k^*})$ and $H_{k^*} \backslash V_{\bar{\mathbf{x}}}(H_{k^*})$. Continue to step $(m+2)$.

Thus we see that as we enter step m, the partition consists of m sets. Denote these by $H_{m,1}(\bar{\mathbf{x}}), H_{m,2}(\bar{\mathbf{x}}), \ldots, H_{m,m}(\bar{\mathbf{x}})$. If $D_m \leq C_{\lambda_m}$, stop and then $(H_{m,1}(\bar{\mathbf{x}}), \ldots, H_{m,m}(\bar{\mathbf{x}}))$ is the final partition. If $D_m \geq C_{\lambda_m}$ we continue and the partition will become finer. If $H_{m,1}(\bar{\mathbf{x}}), H_{m,2}(\bar{\mathbf{x}}), \ldots, H_{m,m}(\bar{\mathbf{x}})$ is the final partition then $H_{ii'}$ is accepted provided i and i' are in the same set of the partition. Otherwise $H_{ii'}$ is rejected. The total number of rejections is r. Note that $\nu_{12} = 0$ is not one of the hypotheses to be tested, so that even though $(1, 2)$ will end up in the same set of the final partition, that hypothesis is not accepted.

Remark 3.2.1. *For the procedure based on PADD+ we refer to the reference of Cohen, Sackrowitz, and Chen (2010). The same modification of keeping indices 1 and 2 in the same set is made.*

Remark 3.2.2. *In case σ is unknown, the usual estimate s^2 of σ^2 is used and then s replaces σ above.*

Of course the method used for obtaining a interval estimates for ν_{12} is repeated for ν_{ij}.

3.2.4 Normal Models

Components of a mean vector

Let $\mathbf{X}$ be a $q \times 1$ normal random vector with mean $\boldsymbol{\mu}$ and known covariance matrix $\Sigma = (\sigma_{ij})$. We seek MTPs and interval estimates for the components of $\boldsymbol{\mu}$. We point out that for testing $H_i : \mu_i = 0$ vs $K_i : \mu_i \neq 0$ the appropriate $\mathbf{g}$ is the ith column of Σ. See CS(2012) for justification of the choice $\mathbf{g}$. First focus on μ_1.

The interval estimate for μ_1 will be

$$X_1 \pm B(r)\sigma_{11}^{\frac{1}{2}} \tag{3.2.10}$$

where $B(r)$, $r = 0, 1, \ldots, q-1$ is a set of decreasing numbers and r is the number of rejections determined by a modified MTP focused on testing $H_i : \mu_i = 0$ vs $K_i : \mu_i \neq 0$, $i = 2, \ldots, q$.

The modified MTP we use in this case is based on the MRD method of Cohen, Sackrowitz and Xu (2009) and goes as follows:

Let $\begin{pmatrix} X_1 \\ \mathbf{X_2} \end{pmatrix}$ and let $\begin{pmatrix} V_1 \\ V_2 \end{pmatrix} = \Sigma^{-1}$ where V_1 is the first row of Σ^{-1} and V_2 is the last $q-1$ rows of Σ.

Define the $(q-1) \times 1$ vector $\mathbf{U} = D^{-1/2}(\Gamma)V_2\mathbf{X} = (U_2, \ldots, U_q)'$
where $\Gamma = V_2 \Sigma V_2'$ and $D(\Gamma)$ is the diagonal matrix whose diagonal elements are those of Γ.

Next let $U_{i_1} = \max_{2 \leq j \leq q} |U_j|$ and let $0 < C_1 < C_2 < \ldots < C_q$ be the constants for the MTP part of the procedure. If $U_{i_1} \leq C_q$ stop and set $r = 0$. If $U_{i_1} > C_q$ go to the next step, namely step 2.

At step 2, consider $\mathbf{X}^{(i_1)}$ which is $\mathbf{X}$ with X_{i_1} left out. Let $\Sigma^{(i_1)}$ be the covariance matrix of $\mathbf{X}^{(i_1)}$. Let $V^{(i_1)} = {\Sigma^{(i_1)}}^{-1}$ and let $V_2^{(i_1)}$ be the last $q-2$ rows of $V^{(i_1)}$. Let $\Gamma^{(i_1)} = V_2^{(i_1)}\Sigma^{(i_1)}{V_2^{(i_1)}}'$ and define $D(\Gamma^{(i_1)})$ as a diagonal matrix as before. Next let $\mathbf{U}^{(i_1)} = D^{-1/2}(\Gamma^{(i_1)})V_2^{(i_1)}\mathbf{X}^{(i_1)}$ and let $U_{i_2} = \max_{\{2\le j\le q\}\setminus i_1} |U_j^{(i_1)}|$. If $U_{i_2} \le C_{q-1}$, stop and set $r = 1$. If $U_{i_2} > C_{q-1}$ go to step 3.

At step p, consider $\mathbf{X}^{(i_1,\ldots,i_{p-1})}$ which is $\mathbf{X}$ with $X_{i_1},\ldots,X_{i_{p-1}}$ left out. Let $\Sigma^{(i_1,\ldots,i_{p-1})}$ be the covariance matrix of $\mathbf{X}^{(i_1,\ldots,i_{p-1})}$. Let $V^{(i_1,\ldots,i_{p-1})} = {\Sigma^{(i_1,\ldots,i_{p-1})}}^{-1}$ and let $V_2^{(i_1,\ldots,i_{p-1})}$ be the last $q-i_p$ rows of $V^{(i_1,\ldots,i_{p-1})}$.

Let $\Gamma^{(i_1,\ldots,i_{p-1})} = V_2^{(i_1,\ldots,i_{p-1})}\Sigma^{(i_1,\ldots,i_{p-1})}{V_2^{(i_1,\ldots,i_{p-1})}}'$ and define $D(\Gamma^{(i_1,\ldots,i_{p-1})})$ as a diagonal matrix as before. Next let

$$\mathbf{U}^{(i_1,\ldots,i_{p-1})} = D^{-1/2}(\Gamma^{(i_1,\ldots,i_{p-1})})V_2{}^{(i_1,\ldots,i_{p-1})}\mathbf{X}^{(i_1,\ldots,i_{p-1})}$$

and let

$$U_{i_p} = \max_{2\le j\le q\setminus\{i_1,\ldots,i_{p-1}\}} |U_j^{(i_1,\ldots,i_{p-1})}|$$

If $U_{i_p} \le C_{q-p+1}$, stop and set $r = p-1$. If $U_{i_p} > C_{q-p+1}$ go to step $(p+1)$.

Once r is finalized and the interval estimate for μ_1 is determined the process is repeated for μ_2 and then $\mu_3,\ldots$ and μ_q.

That the interval estimate for μ_1 (and $\mu_3,\ldots$ and μ_q) will have the interval property in the direction of $\mathbf{g}_1$ can be seen to follow from Theorem 2.2.

At this point we apply the new procedure to a linear regression model. The general linear model assumes $\mathbf{y} = A\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{y}$ is an $n\times 1$ vector, A is an $n\times p$ fixed design matrix of rank p and $\boldsymbol{\epsilon} \sim N(0,\sigma^2 I)$. We assume σ^2 is known. It is well known that $\hat{\boldsymbol{\beta}} = S^{-1}A'\mathbf{y}$ where $S = A'A$ and $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2 S^{-1})$.

Now let $S = \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}$ where S_1 is $1 \times p$ and S_2 is $(p-1) \times p$ and let $S^{-1} = \mathrm{V}$. We seek a interval estimate for β_1.

Step 1:

Let $\mathbf{U} = \begin{pmatrix} U_2 \\ \vdots \\ U_p \end{pmatrix} = S_2\hat{\boldsymbol{\beta}} = A_2'\mathbf{y}$ where $A = \begin{pmatrix} a_1 \\ A_2 \end{pmatrix}$,a_1 is the ith row of A and A_2 is $n \times (p-1)$. Let $U_{i_1} = \max_{2 \le j \le p} |U_j|/\sigma_{v_j} = \max_{2 \le j \le p} |a_j'\mathbf{y}|/(\sigma||\mathbf{a_j}||)$. Let $0 < C_1 < \ldots < C_p$ be constants of the first stage of the procedure. If $U_{i_1} \le C_p$, stop and use

$$\hat{\beta}_1 \pm B(0)\sigma v_{11}^{1/2}$$

If $U_{i_1} > C_p$, go to step 2.

At step 2 replace S^{-1} by $(S^{-1})^{(i_1)}$ with row i_1 and column i_1 deleted from S^{-1}. Let $V^{(i_1)} = [(S^{-1})^{(i_1)}]^{-1}$. Then let $V_2^{(i_1)}$ be the last $(p-2)$ rows of $V^{(i_1)}$. Let $\Gamma^{(i_1)} = V_2^{(i_1)}(S^{-1})^{(i_1)}V_2^{(i_1)'}$ and define $D(\Gamma^{(i_1)})$ as a diagonal matrix as before. Then let

$$\mathbf{U}^{(i_1)} = D^{-1/2}(\Gamma^{(i_1)})V_2^{(i_1)}\hat{\boldsymbol{\beta}}^{(i_1)}$$

and find

$$U_{i_2} = \max_{\{2 \le j \le p\} \setminus \{i_1\}} |U_j^{(i_1)}|.$$

If $U_{i_2} \le C_{p-1}$, stop and set $r = 1$ and use

$$\hat{\beta}_1 \pm B(1)\sigma v_{11}^{1/2}$$

If $U_{i_2} > C_{p-1}$, go to step 3.

Following the steps as in the previous description with $\hat{\boldsymbol{\beta}}$ playing the role of $\mathbf{X}$ and

S^{-1} playing the role of Σ we continue the procedure and get the interval estimate for β_1 as

$$\hat{\beta}_1 \pm B(r)\sigma v_{11}^{1/2}$$

The process is repeated to obtain interval estimates for all components of $\boldsymbol{\beta}$ as well as tests of $H_i : \beta_i = 0$.

Remark 3.2.3. *In all the normal models above we assumed σ^2 to be known. When σ^2 is unknown we recommend replacing σ^2 by s^2 in interval estimate and test statistic formulas. We use s^2 to denote the natural independent estimate of σ^2 that is typically available in linear models. The consequence is that the desirable interval property is retained for fixed s^2.*

3.3 Determination of constants and simulation results

We now turn to the related issues of implementing and evaluating the procedures. Basically the constants C_i used in the stage 1 MTPs as well as the $B(r)$ constants used in stage 2 are chosen by trial and error using simulation. We begin by searching for a procedure that can be seen, by simulation, to perform well as an overall MTP in the practical problem at hand. For the C_i a modification of the Benjamini and Gavrilov (2009) critical values tend to work well while, for the $B(r)$, a modification of the Holm (1979) critical values work well. That is, begin by fixing an α_1 and α_2 and taking

$$C(i) = i\alpha_1/(M - i(1 - \alpha_1)) \tag{3.3.1}$$

and

$$B(r) = \Phi^{-1}(1 - \alpha_2/(M - r)). \tag{3.3.2}$$

Modifying α_1 and α_2 typically leads to an effective procedure. In the treatments versus control model of Section 3.2.1 we did extensive simulations for 5, 10 and 25 treatments and one control. When viewed as an MTP we compared the performance of the MRD and standard step-down procedure with those of the new procedures. In

(3.3.1) and (3.3.2) the new procedure used $\alpha_1 = 0.1$ and $\alpha_2 = 0.03$ in order to control the FDR of the overall procedure at 0.05.

Chapter 4

Nonparametric MTPs and simultaneous interval estimates

Consider a one way layout without a normal assumption. That is, assume each of k populations has the same continuous distribution with an unknown translation parameter μ_i, $i = 1, \ldots, k$. We are interested in various pairwise comparisons emanating from a treatment vs control model, a change point model or all pairwise differences model. We propose new MTPs and also derive simultaneous interval estimates for the various pairwise differences of interest. Both modes of inference are based on ranks. The ranks in question however are not the ranks of the original observations. The ranks are determined in a special way to ensure that the resulting simultaneous interval estimates and MTP procedures have the interval property. The MTPs derived have interval property for testing $H_{ij} : \mu_i - \mu_j = 0$ vs $K_{ij} : \mu_i - \mu_j \neq 0$. The simultaneous interval estimates for $\mu_i - \mu_j$ in the various settings are typically shorter than those based on Bonferroni methods, are informative (not infinite), computationally feasible and they will also have a desirable interval property to be described later.

Simultaneous interval estimates are determined in a 2 stage process. The second stage of the process determines an interval estimate for each particular pairwise difference of interest. The length of each interval depends on the number of rejections made when testing all other pairwise parameters of interest in the first stage of the process. Test statistics for parameters in the first stage turn out to be two sample rank tests. However in comprising the two samples, data from more than 2 populations are used.

Section 4.1 contains the models considered along with the definitions of the interval property for testing and for intervals. In Section 4.2 and 4.3, to provide motivation for what follows, we demonstrate that for $k = 2$, the two sample rank test and the two sample permutation test have the interval property, and we also give examples

of instances where typical stepwise rank tests and permutation test do not have the interval property in multiple testing situations, i.e., $k > 2$. The method and details will be given in Section 4.4. This special rank tests used in determining simultaneous interval estimates and MTPs do have the interval property.

4.1 Model and Definitions

Let $x_{ij} = \mu_i + \epsilon_{ij}$, $j = 1, \ldots, n_i$, $i = 1, 2, \ldots, k$ be independent random variables all with the same continuous distribution except for the translation parameter μ_i. Let $\nu_{ii'} = \mu_i - \mu_{i'}$ be parameters of interest. For treatment vs. control models the parameters of interest are ν_{i1}, $i = 2, \ldots, k$ as $i' = 1$ corresponds to the control population. For change point models $\nu_{i+1,i}$, $i = 1, \ldots, k-1$ are parameters of interest. For all pairwise models, $i' < i$, $i = 2, \ldots, k$, $i' = 1, 2, \ldots, k-1$. For each of the three models we will be seeking simultaneous interval estimates for $\nu_{ii'}$. We will also be concerned with multiple testing of hypotheses $H_{ii'} : \nu_{ii'} = 0$ vs $K_{ii'} : \nu_{ii'} \neq 0$. Our approach for all models and all problems will be to focus on one of the $\nu_{ii'}$ and get an interval estimate for it or a test for it and then in turn to focus individually on each of the other $\nu_{ii'}$ of interest. It will therefore be helpful to think first about an interval estimate or a test for ν_{21}, let

$$\mathbf{x} = (x_{11}, x_{12}, \ldots, x_{1n_1}, x_{21}, x_{22}, \ldots, x_{2n_2}, \ldots, x_{k1}, x_{k2}, \ldots, x_{kn_k})' \tag{4.1.1}$$

be an $n \times 1$ vector of observations with $n = \sum_{i=1}^{k} n_i$.

Now let $\Delta > 0$, $\gamma > \Delta$, also let

$$\begin{aligned} \mathbf{x}^* = \quad & (x_{11} - \frac{\Delta}{n_1}, x_{12} - \frac{\Delta}{n_1}, \ldots, x_{1n_1} - \frac{\Delta}{n_1}, x_{21} + \frac{\Delta}{n_2}, x_{22} + \frac{\Delta}{n_2}, \ldots, x_{2n_2} + \frac{\Delta}{n_2}, \\ & x_{31}, x_{32}, \ldots, x_{3n_3}, \ldots, x_{k_1}, \ldots, x_{kn_k})' \end{aligned} \tag{4.1.2}$$

and let

$$\mathbf{x}^{**} = (x_{11} - \frac{\gamma}{n_1}, x_{12} - \frac{\gamma}{n_1}, \ldots, x_{1n_1} - \frac{\gamma}{n_1}, x_{21} + \frac{\gamma}{n_2}, x_{22} + \frac{\gamma}{n_2}, \ldots, x_{2n_2} + \frac{\gamma}{n_2}, x_{31}, x_{32}, \ldots, x_{3n_3}, \ldots, x_{k_1}, \ldots, x_{kn_k})' \quad (4.1.3)$$

If φ_{21} is a one sided test function for H_{21}, then φ_{21} has the interval property if whenever $\varphi_{21}(\mathbf{x}) = 1$ then $\varphi_{21}(\mathbf{x}^*) = 1$. If φ_{21} is a two sided test function for H_{21}, then φ_{21} has the interval property provided whenever $\varphi_{21}(\mathbf{x}) = 0$ and $\varphi_{21}(\mathbf{x}^*) = 1$, then $\varphi_{21}(\mathbf{x}^{**}) = 1$. An interval estimate $(L(\mathbf{x}), U(\mathbf{x}))$ for ν_{21} has the interval property provided $(L(\mathbf{x}^{**}), U(\mathbf{x}^{**}))$ does not cover zero whenever $(L(\mathbf{x}), U(\mathbf{x}))$ covers zero but $(L(\mathbf{x}^*), U(\mathbf{x}^*))$ does not.

An overall MTP or overall simultaneous interval estimates have the interval property provided each test or each interval has the interval property for each parameter of interest.

4.2 Two sample problem

In this subsection we assume k=2 and demonstrate that for testing $H_{21} : \nu_{21} = \mu_2 - \mu_1 = 0$ vs $K_{21} : \nu_{21} \neq 0$, the rank sum test and permutation test have the interval property.

4.2.1 two sample rank sum test

Without loss of generality let $n_2 \leq n_1$. For the rank test let W be the sum of the ranks for the second population. Assume that at sample point $\mathbf{x}$ given in (4.1.1) with $k = 2$ after ranking the $n_1 + n_2$ observations $W = w$ lies in the acceptance region. Also assume that at $\mathbf{x}^*$ given in (4.1.2) $W = w^*$ lies in the rejection region. Now for this to happen, W at $\mathbf{x}$ had to be between the lower critical value and upper critical value for the rank test (See Devore (2012)) while W at $\mathbf{x}^*$ would have to exceed the upper critical value. This is true since W could only increase as $\mathbf{x} \to \mathbf{x}^*$. That is the ranks for the second population could not decrease. Since this fact is also true as $\mathbf{x}^* \to \mathbf{x}^{**}$

it follows that W at $\mathbf{x}^{**}$ would also exceed the upper critical value.

4.2.2 Two sample Permutation test

Now we study the two sample permutation test. Within a permutation framework it is known (See Basso, Paserin, Salmaso and Solari (2009)) that the usual t-statistic,

$$t = |\overline{x}_2 - \overline{x}_1| / \{[1/n_1 + 1/n_2]s^2\}^{1/2}$$

is equivalent to the statistic

$$T = |\overline{x}_2 - \overline{x}_1| \tag{4.2.1}$$

The two-sample permutation test is as follows:
The vector $\mathbf{x}$ consists of $\mathbf{x}_1$ and $\mathbf{x}_2$. There are $C_{n_1}^{n_1+n_2}$ possible permutation vectors $\mathbf{x}^p$, corresponding to all possible ways of partitioning $n_1 + n_2$ elements of $\mathbf{x}$ into two subsets of size n_1 and n_2. For each $\mathbf{x}^p$, the permutation replication of T is defined as $T^p = T(\mathbf{x}^p)$. Evaluate all T^p permutation replications and then we can calculate

$$P = Prob(T^p \geq T) = \#(T^p \geq T)/C_{n_1}^{n_1+n_2}.$$

For the two sample permutation test, we can extend the definition of interval property to more general situations. Given

$$\mathbf{x} = (x_{11}, x_{12}, \ldots, x_{1n_1}, x_{21}, x_{22}, \ldots, x_{2n_2})'$$

Let $\Delta > 0$, $\gamma > 0$ and let

$$\mathbf{x}^* = (x_{11} - q\Delta, \ldots, x_{1k_1} - q\Delta, \ldots, x_{1n_1}, x_{21} + \Delta, \ldots, x_{2k_2} + \Delta, \ldots, x_{2n_2})' \tag{4.2.2}$$

$$\mathbf{x}^{**} = (x_{11} - q\gamma, \ldots, x_{1k_1} - q\gamma, \ldots, x_{1n_1}, x_{21} + \gamma, \ldots, x_{2k_2} + \gamma, \ldots, x_{2n_2})' \tag{4.2.3}$$

Note when $k_1 < n_1$ and $k_2 < n_2$, only some of the observations are changed.

Again, if φ_{21} is a one sided test function for H_{21}, then φ_{21} has the interval property if whenever $\varphi_{21}(\mathbf{x}) = 1$ then $\varphi_{21}(\mathbf{x}^*) = 1$. If φ_{21} is a two sided test function for H_{21}, then φ_{21} has the interval property provided whenever $\varphi_{21}(\mathbf{x}) = 0$ and $\varphi_{21}(\mathbf{x}^*) = 1$, then $\varphi_{21}(\mathbf{x}^{**}) = 1$.

We now state

Theorem 4.2.1. *The one-sided two sample permutation test has the interval property. The two-sided two sample permutation test has the interval property if* $n_1 = n_2$ *or if* $q \geq \frac{k_2}{k_1}$ *when* $n_1 < n_2$ *or* $q \leq \frac{k_2}{k_1}$ *when* $n_1 > n_2$.

Proof. Fact 1: The test statistic (4.2.1) evaluated at $\mathbf{x}^*$ is either an increasing function of Δ or first decreasing and later increasing as a function of Δ. This follows from (4.2.2) since

$$|\overline{\mathbf{x}_2^*} - \overline{\mathbf{x}_1^*}| = |\overline{\mathbf{x}_2} - \overline{\mathbf{x}_1} + (\frac{k_1 q}{n_1} + \frac{k_2}{n_2})\Delta| \tag{4.2.4}$$

Fact 2: The test statistic evaluated at $\mathbf{x}^*$ changes at the fastest rate among all permutations under the condition if $n_1 = n_2$ or $q \geq \frac{k_2}{k_1}$ when $n_1 < n_2$ or $q \leq \frac{k_2}{k_1}$ when $n_1 > n_2$.

To see this, suppose we swap m ($m \leq \min\{n_1, n_2\}$) observations from one population to the other. Suppose after the swap, m_1 observations which are originally from x_{11} to x_{1k_1} are now in $\mathbf{x}_2$, thus $m_1 \leq k_1$. m_2 observations which are originally from x_{21} to x_{2k_2} are now in $\mathbf{x}_1$, thus $m_2 \leq k_2$. $\max\{m_1, m_2\} \leq m$. Then the change in $\overline{\mathbf{x}}_2 - \overline{\mathbf{x}}_1$ is

$$\begin{aligned} & \frac{(k_2 - m_2)\Delta - m_1\Delta q}{n_2} + \frac{(k_1 - m_1)\Delta q - m_2\Delta}{n_1} \\ \leq \quad & (\frac{k_1 q}{n_1} + \frac{k_2}{n_2})\Delta \end{aligned}$$

This ensures that the one-sided sample permutation test has the interval property.

On the other hand,

$$\begin{aligned} & \frac{(k_2-m_2)\Delta - m_1\Delta q}{n_2} + \frac{(k_1-m_1)\Delta q - m_2\Delta}{n_1} \\ \geq \quad & -(\frac{k_1 q}{n_2} + \frac{k_2}{n_1})\Delta \end{aligned}$$

Because $\frac{(k_2-m_2)\Delta - m_1\Delta q}{n_2} + \frac{(k_1-m_1)\Delta q - m_2\Delta}{n_1}$ is a decreasing function of m_1 and m_2, its minimum is achieved when $m_1 = k_1$ and $m_2 = k_2$. Note this minimum may not always be reached because for example when $k_1 > n_2$, $m_1 < k_1$. In this case, a strict inequality holds.

The condition $n_1 = n_2$ or $q \geq \frac{k_2}{k_1}$ when $n_1 < n_2$ or $q \leq \frac{k_2}{k_1}$ when $n_1 > n_2$ implies that

$$-(\frac{k_1 q}{n_2} + \frac{k_2}{n_1})\Delta \quad \geq \quad -(\frac{k_1 q}{n_1} + \frac{k_2}{n_2})\Delta$$

Therefore,

$$\frac{(k_2-m_2)\Delta - m_1\Delta q}{n_2} + \frac{(k_1-m_1)\Delta q - m_2\Delta}{n_1} \quad \geq \quad -(\frac{k_1 q}{n_1} + \frac{k_2}{n_2})\Delta$$

Now suppose the two sample test will reject when the P-value of the test statistic (4.2.1) $< \alpha$. Suppose that at $\Delta = 0$ we are at an accept point and at $\Delta = a^*$ we are at a reject point. This means that the P-value decreases from above α at $\Delta = 0$ to below level α at $\Delta = a^*$. This implies that the test statistic for at least one of the other permutations fell below the test statistic for the original point with the increase $\Delta = a^*$. Once the statistic in (4.2.4) as a function of Δ is increasing it does so at the fastest rate so the corresponding P-value can only be nonincreasing. Thus once there is a rejection at $\mathbf{x}^*$ following an acceptance at $\mathbf{x}$, a rejection at $\mathbf{x}^{**}$ must ensue. □

Corollary 4.2.1. *Under the definition of (4.1.2) and (4.1.3), one-sided and two-sided two sample permutation tests have the interval property.*

Because in that case, $k_1 = n_1$, $k_2 = n_2$, $q\Delta = \frac{\Delta}{n_1}$, $\Delta = \frac{\Delta}{n_2}$, we have $q = \frac{n_2}{n_1} = \frac{k_2}{k_1}$. It satisfies the conditions of the theorem and thus is a special case.

4.2.3 Two sample bootstrap test

The bootstrap algorithm is quite similar to the permutation algorithm. The main difference is the sampling is carried out with replacement rather than without replacement. We need to evaluate all $(n_1+n_2)^{n_1+n_2}$ bootstrap replications of T denoted as T^b. Then we can calculate P-value=$Prob(T^b \geq T) = \#(T^b \geq T)/(n_1+n_2)^{n_1+n_2}$. When $(n_1+n_2)^{n_1+n_2}$ is large, in practice, the bootstrap P-value is approximated by Monte Carlo methods, that is, instead of evaluating T^b for all $(n_1+n_2)^{n_1+n_2}$ $\mathbf{x}^b$, choose B vectors randomly from $(n_1+n_2)^{n_1+n_2}$ $\mathbf{x}^b$. B will usually be at least 1000.

Efron and Tibshirani pointed out in their book *An introduction to the Bootstrap* that the permutation test is exact, while the bootstrap test is approximate. In practice, the two methods often give quite similar results.

The interval property of a bootstrap test in practice is difficult to analyze because of the randomness in generating the bootstrap sample. For theoretical purpose, we still could study the property of the limiting form of bootstrap test, that is, when all $(n_1+n_2)^{n_1+n_2}$ T^b are evaluated.

Theorem 4.2.2. *Under definition (4.2.2) and (4.2.3), if $k_1 = n_1$ and $k_2 = n_2$, both the one-sided bootstrap test and the two-sided bootstrap test have the interval property.*

Proof.

$$\mathbf{x}^* = (x_{11} - q\Delta, x_{12} - q\Delta, \ldots, x_{1n_1} - q\Delta, x_{21} + \Delta, x_{22} + \Delta, \ldots, x_{2n_2} + \Delta)'$$

Fact 1: The test statistic evaluated at $\mathbf{x}^*$ is either an increasing function of Δ or first decreasing and later increasing as a function of Δ. This follows from (4.2.2) since

$$|\overline{\mathbf{x}_2^*} - \overline{\mathbf{x}_1^*}| = |\overline{\mathbf{x}_2} - \overline{\mathbf{x}_1} + (1+q)\Delta| \tag{4.2.5}$$

Fact 2: The test statistic evaluated at $\mathbf{x}^*$ changes at the fastest rate among all

bootstrap samples if $k_1 = n_1$ and $k_2 = n_2$.

Because each observation now can be sampled with replacement from $(n_1+n_2)^{n_1+n_2}$ observations. Suppose now $\mathbf{x}_2$ contains m_1 observations from $\mathbf{x}_1$, note duplication is possible, thus $m_1 \leq n_2$. $\mathbf{x}_1$ contains m_2 observations from $\mathbf{x}_2$, thus $m_2 \leq n_1$. Then the change in $\overline{\mathbf{x}}_2 - \overline{\mathbf{x}}_1$ is

$$\begin{aligned} & \frac{(n_2 - m_1)\Delta - m_1\Delta q}{n_2} + \frac{(n_1 - m_2)\Delta q - m_2\Delta}{n_1} \qquad (4.2.6) \\ \leq \; & (1+q)\Delta \end{aligned}$$

On the other hand,

$$\begin{aligned} & \frac{(n_2 - m_1)\Delta - m_1\Delta q}{n_2} + \frac{(n_1 - m_2)\Delta q - m_2\Delta}{n_1} \qquad (4.2.7) \\ \geq \; & \frac{(n_2 - n_2)\Delta - n_2\Delta q}{n_2} + \frac{(n_1 - n_1)\Delta q - n_1\Delta}{n_1} \\ = \; & -(1+q)\Delta \qquad (4.2.8) \end{aligned}$$

Suppose the two sample test will reject when the P-value of the test statistic (4.2.1) $< \alpha$. Note, suppose that at $\Delta = 0$ we are at an accept point and at $\Delta = a^*$ we are at a reject point. This means that the P-value decreases from above α at $\Delta = 0$ to below level α at $\Delta = a^*$. This implies that the test statistic for at least one of the other permutations fell below the test statistic for the original point with the increase $\Delta = a^*$. Once the statistic in (4.2.5) as a function of Δ is increasing it does so at the fastest rate so the corresponding P-value can only be nonincreasing. Thus once there is a rejection at $\mathbf{x}^*$ following an acceptance at $\mathbf{x}$, a rejection at $\mathbf{x}^{**}$ must ensue. □

Corollary 4.2.2. *: Under the definition of (4.1.2) and (4.1.3), one sample and two sample bootstrap test have the interval property.*

Because in that case, $k_1 = n_1$, $k_2 = n_2$, it satisfies the conditions of the theorem. Therefore, it is a special case of this theorem. In the theorem there is no restriction on the value of q, while in the special case $q = \frac{n_2}{n_1} = \frac{k_2}{k_1}$.

4.3 Multiple testing problem

4.3.1 Step-down MTP based on ranks

We will consider examples where there are three populations, one control and two for treatments. We will perform a step-down MTP and focus on the test for the difference in translation parameters for treatment labeled 2, and control which is labeled 1. That is, let μ_i, $i = 1, 2, 3$, represent the translation parameters for the 3 populations. Multiple test $H_{21} : \nu_{21} = \mu_2 - \mu_1 = 0$ vs $K_{21} : \nu_{21} \neq 0$ and $H_{31} : \nu_{31} = \mu_3 - \mu_1 = 0$ vs. $K_{31} : \nu_{31} \neq 0$ using the step-down method. The tests performed will be based rank tests. We offer two examples. One where the ranks are based on full ranking of all observations from the 3 populations and another where ranking is based on each pair of populations separately.

For the example we will take 3 independent observations from each population. That is, let x_{ij}, $i = 1, 2, 3; j = 1, 2, 3$ and let $\mathbf{x} = (x_{11}, x_{12}, x_{13}, x_{21}, x_{22}, x_{23}, x_{31}, x_{32}, x_{33})'$. Let T_i, $i = 2, 3$ stand for treatments with T_1 being the control.

Let $U_i(\mathbf{x}) = |R_i - R_1|$, $i = 2, 3$, where R_i is the sum of the ranks for population i , be test statistics for H_{i1}. For the step-down method at step 1, let C_1 be a constant and let $U^{(1)}_{i^*}(\mathbf{x}) = \max_{i\in\{2,3\}} U_i(\mathbf{x})$. If $U^{(1)}_{i^*}(\mathbf{x}) \leq C_1$, accept H_{21} and H_{31}.
If $U^{(1)}_{i^*}(\mathbf{x}) > C_1$, reject H_{i^*1} and go to step 2.

At step 2, let $C_2 < C_1$ be a constant and consider $U^{(2)}(\mathbf{x}) = U_i(\mathbf{x})$ where $i \neq i^*$. If $U^{(2)}(\mathbf{x}) \leq C_2$, reject H_{i^*1}, but accept H_{i1}. If $U^{(2)}(\mathbf{x}) > C_2$, reject both hypotheses.

Now we offer numerical examples to demonstrate that the step-down MTP for testing H_{21} does not have the interval property. Note that the sample points in the examples follow the behavior given in (4.2.1), (4.2.2) and (4.2.3) with $\Delta = 0.2$ and $\gamma = 0.4$. Thus the scores for the treatment T_2 and the control are getting closer together as

$\mathbf{x} \to \mathbf{x}^* \to \mathbf{x}^{**}$ while the (independent) scores for treatment T_3 remain fixed.

Step-down MTP based on joint ranks

The step-down MTP based on joint ranks doesn't have the interval property.

A counter example is as follows:

Let $C_1 = 8, C_2 = 5$.

for sample points at x

Raw Scores			**Ranks**		
T_2	T_1	T_3	T_2	T_1	T_3
1.6	2.2	6	1	2	6
2.8	5.2	8	3	5	7
3.8	8.1	9	4	8	9

$|R_2 - R_1| = 7$, $|R_3 - R_1| = 7$, accept H_{21} and H_{31}.

for sample points at $\mathbf{x}^*$

Raw Scores			**Ranks**		
T_2	T_1	T_3	T_2	T_1	T_3
1.8	2.0	6	1	2	6
3.0	5.0	8	3	5	8
4.0	7.9	9	4	7	9

$|R_2 - R_1| = 6$, $|R_3 - R_1| = 9$, reject H_{21} and H_{31}.

for sample points at $\mathbf{x}^{**}$

Raw Scores			**Ranks**		
T_2	T_1	T_3	T_2	T_1	T_3
2.0	1.8	6	2	1	6
3.2	4.8	8	3	5	8
4.2	7.7	9	4	7	9

$|R_2 - R_1| = 4$, $|R_3 - R_1| = 10$, reject H_{31} and accept H_{21}.

Thus we see that for testing $H_{21} : \nu_{21} = 0$, the step down procedure goes from accept to reject to accept as $\mathbf{x} \to \mathbf{x}^* \to \mathbf{x}^{**}$.

MTP based on separate ranks

The step-down MTP based on separate ranks doesn't have the interval property.
A counter example is as follows:

Let $C_1 = 6, C_2 = 4$.

for sample points at x

Raw Scores			**Ranks**			
T_2	T_1	T_3	T_2	T_1	T_1	T_3
1.6	2.2	6	1	2	1	3
2.8	5.2	8	3	5	2	4
3.8	8.1	9	4	6	5	6

$|R_2 - R_1| = 5$, $|R_3 - R_1| = 5$, accept H_{21} and H_{31}.

for sample points at $\mathbf{x}^*$

Raw Scores			**Ranks**			
T_2	T_1	T_3	T_2	T_1	T_1	T_3
1.8	2.0	6	1	2	1	3
3.0	5.0	8	3	5	2	5
4.0	7.9	9	4	6	4	6

$|R_2 - R_1| = 5$, $|R_3 - R_1| = 7$, reject H_{21} and H_{31}.

for sample points at $\mathbf{x}^{**}$

Raw Scores			**Ranks**			
T_2	T_1	T_3	T_2	T_1	T_1	T_3
2.0	1.8	6	2	1	1	3
3.2	4.8	8	3	5	2	5
4.2	7.7	9	4	6	4	6

$|R_2 - R_1| = 3$, $|R_3 - R_1| = 7$, reject H_{31} and accept H_{21}.

Thus we see that for testing $H_{21} : \nu_{21} = 0$, the step down procedure goes from accept to reject to accept as $\mathbf{x} \to \mathbf{x}^* \to \mathbf{x}^{**}$.

4.3.2 Step-down MTP based on permutation test

When $k > 2$, the multiple permutation test doesn't have the interval property. Two counter-examples are offered here. They demonstrate that when $k = 3$ the step-down procedure based on pairwise permutation tests does not have the interval property.

Consider the change point model when there are 3 populations at times t_1, t_2, t_3. We demonstrate that the step-down MTP based on pairwise permutation tests for $\nu_{21} = \mu_2 - \mu_1$ and $\nu_{32} = \mu_3 - \mu_2$ does not have the interval property for testing

$H_{21} : \nu_{21} = 0$. For critical values we let $C_1 = 0.05$ and $C_2 = 0.10$. The data for 3 populations are as follows:

for sample points at x

$\mathbf{t_1}$	$\mathbf{t_2}$	$\mathbf{t_3}$
-0.55	0.21	1.02
-1.31	0.12	1.45
-0.06	1.32	2.12
0.54	0.58	1.88
-0.89	-0.77	0.66

The means $(\overline{x}_1, \overline{x}_2, \overline{x}_3) = (-0.454, 0.292, 1.426)$. The statistics are $|\overline{x}_2 - \overline{x}_1| = 0.746$, $|\overline{x}_3 - \overline{x}_2| = 1.134$ with permutation P-value of 0.151 and 0.032 respectively. This means that at $\mathbf{x}$, H_{21} is accepted and H_{32} is rejected. Note the permutation P-values are determined by listing all C_5^{10} possible outcomes for the permutations and finding the percent of these outcomes where statistic is less than or equal to the observed statistic at $\mathbf{x}$.

for sample points at $\mathbf{x}^*$

$\mathbf{t_1}$	$\mathbf{t_2}$	$\mathbf{t_3}$
-0.65	0.31	1.02
-1.41	0.22	1.45
-0.16	1.42	2.12
0.44	0.68	1.88
-0.99	-0.67	0.66

The means $(\overline{x}_1^*, \overline{x}_2^*, \overline{x}_3^*) = (-0.554, 0.392, 1.426)$. The statistics are $|\overline{x}_2^* - \overline{x}_1^*| = 0.946$, $|\overline{x}_3^* - \overline{x}_2^*| = 1.034$ with p-value of 0.087 and 0.048 respectively. This means that at $\mathbf{x}^*$, H_{21} and H_{32} are both rejected. At $\mathbf{x}^{**}$

for sample points at $\mathbf{x}^{**}$

$\mathbf{t_1}$	$\mathbf{t_2}$	$\mathbf{t_3}$
-0.70	0.36	1.02
-1.46	0.27	1.45
-0.21	1.47	2.12
0.39	0.73	1.88
-1.04	-0.62	0.66

The means are $(-0.604, 0.442, 1.426)$. The statistics are $|\overline{x}_2^{**} - \overline{x}_1^{**}| = 1.046$, $|\overline{x}_3^{**} - \overline{x}_2^{**}| = 0.984$ with P-values 0.071 and 0.071 respectively. This means that at $\mathbf{x}^{**}$, H_{21} and H_{32} are both accepted. Thus for H_{12}, as $\mathbf{x} \to \mathbf{x}^{*} \to \mathbf{x}^{**}$ we have an accept, followed by a reject, followed by an accept. This violates the interval property.

Consider the treatment vs. control model when there are 3 populations at times T_1, T_2, T_3. We demonstrate that the step-down MTP based on pairwise permutation tests for $\nu_{21} = \mu_2 - \mu_1$ and $\nu_{31} = \mu_3 - \mu_1$ does not have the interval property for testing $H_{21} : \nu_{21} = 0$. For critical values we let $C_1 = 0.05$ and $C_2 = 0.10$. The data for 3 populations are as follows:

for sample points at x

$\mathbf{T_2}$	$\mathbf{T_1}$	$\mathbf{T_3}$
0.41	1.55	2.22
0.32	0.79	2.65
1.52	2.04	3.32
0.78	2.64	3.08
-0.57	1.21	1.86

The means $(\overline{x}_1, \overline{x}_2, \overline{x}_3) = (1.65, 0.49, 2.63)$. The statistics are $|\overline{x}_2 - \overline{x}_1| = 1.16$, $|\overline{x}_3 - \overline{x}_1| = 0.98$ with permutation P-value of 0.071 and 0.056 respectively. This means

that at $\mathbf{x}$, both H_{21} and H_{31} are accepted. Note the permutation P-values are determined by listing all C_5^{10} possible outcomes for the permutations and finding the percent of these outcomes where statistic is less than or equal to the observed statistic at $\mathbf{x}$.

for sample points at $\mathbf{x}^*$

$\mathbf{T_2}$	$\mathbf{T_1}$	$\mathbf{T_3}$
0.51	1.45	2.22
0.42	0.69	2.65
1.62	1.94	3.32
0.88	2.54	3.08
-0.47	1.11	1.86

The means $(\overline{x}_1^*, \overline{x}_2^*, \overline{x}_3^*) = (1.55, 0.59, 2.63)$. The statistics are $|\overline{x}_2^* - \overline{x}_1^*| = 0.96$, $|\overline{x}_3^* - \overline{x}_1^*| = 1.08$ with P-value of 0.071 and 0.040 respectively. This means that at $\mathbf{x}^*$, H_{21} and H_{31} are both rejected.

for sample points at $\mathbf{x}^{**}$

$\mathbf{T_2}$	$\mathbf{T_1}$	$\mathbf{T_3}$
0.56	1.40	2.22
0.47	0.64	2.65
1.67	1.89	3.32
0.93	2.49	3.08
-0.42	1.06	1.86

The means are $(1.50, 0.64, 2.63)$. The statistics are $|\overline{x}_2^{**} - \overline{x}_1^{**}| = 0.86$, $|\overline{x}_3^{**} - \overline{x}_1^{**}| = 1.13$ with P-values 0.111 and 0.040 respectively. This means that at $\mathbf{x}^{**}$, H_{21} is accepted and H_{31} is rejected. Thus for H_{21}, as $\mathbf{x} \to \mathbf{x}^* \to \mathbf{x}^{**}$ we have an accept, followed by a reject, followed by an accept. This violates the interval property.

4.4 New Approach

In this section, we will describe the new nonparametric MTPs which have associated interval estimates and the practical interval property.

Let M be the total number of parameters, $\nu_{ii'}$, of interest for a particular pairwise difference model. Our approach to obtaining MTPs and simultaneous interval estimates is the same for all pairwise comparisons models. It is to focus on each individual parameter separately and treat it with a 2-stage process that depends on all the data. All parameters are treated in a similar fashion. For example, we can focus on the parameter ν_{21}.

At stage 1 we apply an (M-1 population) MTP to **ONLY** all the other hypotheses involving parameters of interest, (i.e., omitting only ν_{21}). However, this MTP is based on all the data. At the end of this stage the number of rejections, $r = r(\mathbf{x})$, with possible values $r = 0, 1, ..., M-1$ is recorded.

At stage 2 the process we present produces an interval estimate based on the r produced at stage 1 and on ranking the pairwise differences of observations from populations 1 with those from population 2. The corresponding MTP decision is defined by accepting the hypothesis $H_{21} : \nu_{21} = 0$ if and only if the interval covers 0. With this relationship between the MTP and interval estimates it immediately follows that

Theorem 4.4.1. *The MTP has the interval property if and only if the corresponding interval estimates have the interval property.*

The particular interval estimate we use at stage 2 is based on the Wilcoxon rank-sum interval as described in Devore (2012), page 643. The interval is determined as follows:

Let $B(r)$ be a decreasing set of constants. Also let, for $e = 1, \ldots, n_2$, $f = 1, \ldots, n_1$, $d_{ef} = x_{2e} - x_{1f}$ and let the ordered differences be $d_{ef(1)}, d_{ef(2)}, \ldots, d_{ef(n_1 n_2)}$. Then the

interval estimate for ν_{21} is

$$(d_{ef(n_1n_2-B(r)+1)}, d_{ef(B(r))}) \tag{4.4.1}$$

For the constants $B(r)$ we take the approximation also given in Devore (2012).

$$B(r) = \text{largest integer in} \quad n_1n_2/2 + z_{\gamma(r)/2}\sqrt{n_1n_2(n_1+n_2+1)/12}$$

where z denotes the standard normal critical values.

The values of $\gamma(r)$ need to be determined by simulation. Also the constants used with the particular choice of the MTP used at stage 1 need to be determined by simulation. These issues will be discussed later.

There is considerable flexibility in choosing the M-1 dimensional MTP to be used at stage 1. Those that we suggest here require some of what we refer to as amalgamation of samples. The amalgamation of two samples consists of first converting the observed samples of (possibly) unequal sizes to samples with equal sizes and then combining them. We use the following algorithm to accomplish the amalgamation.

a. Sample size conversion: Suppose $x^{(j)}_{(1)}, \ldots, x^{(j)}_{(n_j)}$ are n_j ordered values for some population j. We wish to convert them into n_m ordered values $z^{(j)}_{(1)}, \ldots, z^{(j)}_{(n_m)}$. To do so define n_jn_m scores y_β, $\beta = 1, \ldots, n_jn_m$ as follows. Fix $i = 1, \ldots, n_j$. For all $\beta = (i-1)n_m + 1, \ldots, in_m$. Let $y_\beta = x^{(j)}_{(i)}$. That is, y_β is the sequence of $x^{(j)}_{(i)}$ scores and each repeated n_m times. Next go along the y_β sequence taking them n_j at a time. For $i = 1, \ldots, n_m$, define the converted score to be

$$z^{(j)}_{(i)} = \sum_{\beta=(i-1)n_j+1}^{in_j} y_\beta/n_j \tag{4.4.2}$$

b. Combining scores: Suppose we want to amalgamate the samples from populations i and i' that have had their sample sizes converted to n_m. We form a sample of size n_m of ordered values by taking a weighted average of all corresponding n_m order statistics. For example, the first weighted ordered value of the combined sample is

$$U_{(1)}(\mathbf{x}) = (n_i z_{(1)}^{(i)} + n_{i'} z_{(1)}^{(i')})/(n_i + n_{i'}) \tag{4.4.3}$$

Continue and determine $U_{(2)}(\mathbf{x}), \ldots, U_{(n_m)}(\mathbf{x})$. These are the amalgamated scores. It is easy to check that

Lemma 4.4.1. *Suppose we are testing $H_{21} : \nu_{21} = 0$ and all the statistics used in the stage 1 MTP depend on the samples from populations 1 and 2 only through their amalgamation. Then r will be constant for $\boldsymbol{x}, \boldsymbol{x}^*, \boldsymbol{x}^{**}$ as defined in (4.1.1), (4.1.2) and (4.1.3). That is, $r(\boldsymbol{x}) = r(\boldsymbol{x}^*) = r(\boldsymbol{x}^{**})$.*

Clearly the analogous result holds for testing $H_{ii'} : \nu_{ii'} = 0$ for any i, i'. We can now state

Theorem 4.4.2. *If all the statistics used in the stage 1 MTP for testing $H_{ii'} : \nu_{ii'} = 0$ depend on the samples from populations i and i' only through their amalgamation then the interval estimate for $\nu_{ii'}$ has the interval property.*

Proof. It suffices to show

(1) $B(r)$ does not change as the sample point $\mathbf{x}$ changes to $\mathbf{x}^*$ and $\mathbf{x}^*$ changes to $\mathbf{x}^{**}$.

(2) $d_{ef(n_1n_2-B(r)+1)}$ given in (4.4.1) increases as $\mathbf{x} \to \mathbf{x}^* \to \mathbf{x}^{**}$.

To see (2) assuming (1), suppose the INTERVAL ESTIMATES of (4.4.1) covers zero at $\mathbf{x}$ but not at $\mathbf{x}^*$. Since the lower end point of the interval increases as $\mathbf{x} \to \mathbf{x}^*$ this implies that at $\mathbf{x}^*$, $d_{ef(n_1n_2-B(r)+1)} > 0$. Now with $d_{ef(n_1n_2-B(r)+1)} > 0$ increasing further as $\mathbf{x}^* \to \mathbf{x}^{**}$, the lower end point > 0. The reason $d_{ef(n_1n_2-B(r)+1)}$ increases follows from (1) and the definition of d_{ef} which implies that all ordered values of the combined n_1n_2 differences increase.

(1) follows from Lemma 4.4.1. which completes the proof. □

Corollary 4.4.1. *If all the statistics used in the stage 1 MTP for testing $H_{ii'} : \nu_{ii'} = 0$ depend on the samples from populations i and i' only through their amalgamation then the MTP resulting from the interval procedure (rejecting a hypothesis if and only if the interval does not cover zero) has the interval property.*

Proof. The proof follows from Theorem 4.4.1. □

In the specific models that follow we will demonstrate the use of a variety of possible stage 1 MTPs. They will all be based on 2-sample Wilcoxon rank sum statistics. However the particular situation will determine how to form the two samples on which the test is applied.

We will use the following notation and terminology with regard to use of the Wilcoxon rank sum test. Suppose our samples to be compared have sizes m_1 and m_2 with $m_1 \leq m_2$. Then the Wilcoxon rank sum statistic, W, is equal to the sum of the ranks of the scores in the smaller sample when the ranks come from the ranking of the combined sample. By the corresponding two sided P-value we mean

$$P = 2[1 - \Phi(W - m_1(m_1 + m_2 + 1)/2)/\sqrt{m_1 m_2(m_1 + m_2 + 1)/12})]$$

where Φ denotes the standard normal cdf.

4.4.1 Treatments vs. Control

In keeping with Section 3.2.1, the treatment vs control setup we offer 2 different MTPs that can be used at stage 1 of the individual tests. When focusing on ν_{21} we need a stage 1 MTP for testing the $k-2$ hypotheses $H_{i1} : \nu_{i1} = 0$ vs $K_{i1} : \nu_{i1} \neq 0$, $i = 3, \ldots, k$.

The first procedure we present is a very easily computable stepwise procedure. The first step of this stage 1 MTP is to form $k-2$ two sample Wilcoxon rank sum statistics. One sample of the two is just the sample of size n_i from population i. The other sample

comes from amalgamating the samples from populations 1 and 2 resulting in a single sample of size is $n_m = min(n_1, n_2)$.

Compute W_i the two sample Wilcoxon rank sum statistic for $i = 3, \ldots, k$ and its corresponding two sided P-value denoted by P_i. Then let $P_{i^*} = \min_{3 \le i \le k} P_i$. If $P_{i^*} > C_1$, then accept all H_{i1}, $i = 3, \ldots, k$. If $P_{i^*} \le C_1$, reject H_{i^*1} and go to step 2.

At step 2, repeat step 1 except now leave population i^* out of all calculations. That is, get the same set of P-values but leave out P_{i^*} (this time there are $k-3$ $P-$values) and get the min of these to compare with $C_2 > C_1$. If this min P-value corresponds to population i^{**} and is $< C_2$, reject $H_{i^{**}1}$ and go to step 3. Otherwise stop and accept all hypotheses other than H_{i^*1}. In this latter case r =number of rejections is 1. Should the process continue, eventually r will be determined as a number $0, 1, \ldots, (k-2)$ which in turn will determine $B(r)$ to be used in stage 1.

It follows from Theorem 4.4.2 and Corollary 4.4.1 that

Theorem 4.4.3. *The interval estimate for ν_{21} has the interval property.*

and

Corollary 4.4.2. *Corollary: The MTP resulting from the interval procedure (rejecting a hypothesis if and only if the interval does not cover zero) has the interval property.*

An alternative stepwise multiple testing procedure to be used at stage 1 for the simultaneous interval estimation procedure parallels the MRD based method in Section 3.2.1. It is as follows: Again the first step of the stepwise procedure is to form $(k-2)$ two sample Wilcoxon rank sum statistics. Assuming we are seeking the confidence interval for ν_{21} the one sample of the two is the sample of size n_i from population i, $i \neq 1$, $i \neq 2$. The other sample comes from pooling the amalgamation of samples 1 and 2 with the other $k-3$ samples (i.e., not including samples from populations $1, 2, i$).

Obtain, W_i, the two sample Wilcoxon rank sum statistic for $i = 3, \ldots, k$ and its corresponding two sided P-value denoted by P_i. Then let $P_{i^*} = \min_{3 \le i \le k} P_i$. If $P_{i^*} > C_1$,

then accept all H_{i1}, $i = 3, \ldots, k$. If $P_{i^*} \leq C_1$, reject H_{i^*1} and go to step 2.

At step 2, repeat step 1 except now leave population i^* out of all calculations. Once again get a set of $P-$values (this time there are $k-3$ P-values) and get the min of these to compare with $C_2 > C_1$. If this min P-value corresponds to population i^{**} and is $< C_2$, reject $H_{i^{**}1}$ and go to step 3. Otherwise stop and accept all hypotheses other than H_{i^*1}. In this latter case r =number of rejections is 1. Should the process continue, eventually r will be determined as a number $0, 1, \ldots, (k-2)$ which in turn will determine $B(r)$ to be used in stage 1.

It follows from Theorem 4.4.2 and Corollary 4.4.1 that both the interval estimates and the MTP using this alternative procedures at stage 1 has the interval property.

We illustrate the simultaneous interval estimation procedure with an example. We consider data from exercise 26, P.649 of Devore (2012). The data consists of observations from a control population and 3 treatment populations. We deliberately left out one observation to illustrate the method. Letting control be treatment 1 the raw data for the 4 populations are as follows:

$$T_1 : 33.2, 25.3, 20.2, 20.3, 18.3, 19.3, 17.3, 17.0, 16.7, 18.3$$

$$T_2 : 37.1, 31.8, 28.0, 25.9, 25.5, 25.3, 23.7, 24.4, 21.7$$

$$T_3 : 58.9, 54.2, 49.2, 47.9, 38.2, 48.8, 47.8, 40.2, 44.0, 46.4$$

$$T_4 : 56.7, 49.6, 46.4, 40.9, 39.4, 37.1, 37.5, 39.6, 35.1, 36.5$$

It will be convenient to order the values for each population. The ordered values are as

follows:

$$T_1 : 16.7, 17.0, 17.3, 18.3, 18.3, 19.3, 20.2, 20.3, 25.3, 33.2 \tag{4.4.4}$$

$$T_2 : 21.7, 23.7, 24.4, 25.3, 25.5, 25.9, 28.0, 31.8, 37.1 \tag{4.4.5}$$

$$T_3 : 38.2, 40.2, 44.0, 46.4, 47.8, 47.9, 48.8, 49.2, 54.2, 58.9$$

$$T_4 : 35.1, 36.5, 37.1, 37.5, 39.4, 39.6, 40.9, 46.4, 49.6, 56.7$$

We first seek an interval estimate for $\nu_{21} = \mu_2 - \mu_1$. We will derive such at stage 2 of the process. At stage 1 we wish to do an MTP for $H_{31} : \nu_{31} = 0$ and $H_{41} : \nu_{41} = 0$. The MTP will be a step-down procedure where at step 1 we will form 2 two sample rank statistics. The first statistic will arise from the ranks of population 3 and the combined ranks of population 1 and 2. The second statistic will arise from the ranks of population 4 and the combined ranks of populations 1 and 2. To find the combined ranks of populations 1 and 2, first note that the sample sizes are $n_1 = 10, n_2 = 9$.

We convert the 10 ordered observations from population 1 into 9 using (4.4.2) as follows:

$$\begin{aligned} & \{\frac{9}{10}16.7 + \frac{1}{10}17.0, \frac{8}{10}17.0 + \frac{2}{10}17.3, \frac{7}{10}17.3 + \frac{3}{10}18.3, \frac{6}{10}18.3 + \frac{4}{10}18.3, \frac{5}{10}18.3, \\ & + \frac{5}{10}19.3\frac{4}{10}19.3 + \frac{6}{10}20.2, \frac{3}{10}20.2 + \frac{7}{10}20.3, \frac{2}{10}20.3 + \frac{8}{10}25.3, \frac{1}{10}25.3 + \frac{9}{10}33.2\} \\ = \; & \{16.73, 17.06, 17.60, 18.3, 18.8, 19.84, 20.27, 24.30, 32.41\} \end{aligned} \tag{4.4.6}$$

At this point we use (4.4.3) to get a weighted average of the corresponding order statistics for population 1 and 2 combined. For this example these are as follows by

virtue of (4.4.5) and (4.4.6):

$$U_{(1)} = (10(16.73) + 9(21.7))/19 = 19.08$$

$$U_{(2)} = (10(17.06) + 9(23.7))/19 = 20.21$$

$$U_{(3)} = (10(17.6) + 9(24.4))/19 = 20.82$$

$$U_{(4)} = (10(18.3) + 9(25.3))/19 = 21.62$$

$$U_{(5)} = (10(18.8) + 9(25.5))/19 = 21.97$$

$$U_{(6)} = (10(19.84) + 9(25.9))/19 = 22.71$$

$$U_{(7)} = (10(20.27) + 9(28.0))/19 = 23.93$$

$$U_{(8)} = (10(24.30) + 9(31.8))/19 = 27.85$$

$$U_{(9)} = (10(32.41) + 9(37.1))/19 = 34.63$$

Next we find the 2 sample Wilcoxon rank sum statistics for testing population 3 with 10 observations against population 1 represented by the 9 order statistics just computed. That is, compare the two samples
(i) 38.2, 40.2, 44.0, 46.4, 47.8, 47.9, 48.8, 49.2, 54.2, 58.9
(ii) 19.08, 20.21, 20.82, 21.62, 21.97, 22.71, 23.93, 27.85, 34.63.

The rank sum statistic is $W = -45$ with a 2-sided P-value of 0.0004.
Similarly we find the rank sum statistic for testing population 4 with 10 observations against population 1 represented by the 9 order statistics computed. This time $W = -45$ again with a 2-sided $P-$value of 0.0004.

We now complete stage 1, which is an MTP for testing H_{31} and H_{41}. Toward this end consider constants as prescribed in Section 4. We pick $\alpha_1 = 0.05$ and $\alpha_2 = 0.01$ so that $C_1 = 0.024$ and $C_2 = 0.091$. Note both P-values < 0.01 which means the step-down procedure prescribed at stage 1 rejects both H_{31} and H_{41}. Thus $r = 2$.

For stage 2 we use (4.4.1) with $r = 2$, $\gamma(2) = 0.01$ as derived from (4.4.7) and thus

$B(r) = 77$ which means we need the 14th and 77th ordered values from the 90 ordered pairwise differences of T_2 and T_1 gives in (4.4.5) and (4.4.4) respectively. Hence the interval estimate using (4.4.1) is $[0.6, 12.7]$. This completes the two stage procedure for ν_{21}. We would repeat these same procedures to get interval estimates for ν_{31} and ν_{41}.

4.4.2 Change Point

We will focus on $\nu_{(\widehat{i}+1)\widehat{i}}$, so at stage 1 we do a stepwise multiple testing procedure for $H_{(i+1)i} : \nu_{(i+1)i} = 0$ vs. $K_{(i+1)i} : \nu_{(i+1)i} \neq 0$, $i = 1, \ldots, k-1$ but $i \neq \widehat{i}$. The first step of the stepwise procedure is to form $k-2$ two sample Wilcoxon rank sum statistics. If $i < \widehat{i}$ one sample of the two comes from pooling the samples from populations 1 through i into one large sample. The other sample comes from pooling the samples from populations $i+1$ through k but with the samples from populations $\widehat{i}$ and $\widehat{i}+1$ being amalgamated. If $i > \widehat{i}$ one sample of the two comes from pooling the samples from populations $i+1$ through k. The other sample comes from pooling the samples from populations 1 through i but with the samples from populations $\widehat{i}$ and $\widehat{i}+1$ being amalgamated.

Let those $k-2$ statistics be denoted by W_i, $i = 1, \ldots, k-1$ but $i \neq \widehat{i}$ and let P_i denote the corresponding P-value. Bear in mind as we continue the description that $H_{(\widehat{i}+1)\widehat{i}}$ is never tested and that the samples from $\widehat{i}+1$ and $\widehat{i}$ are always amalgamated. Thus d, i^* and i^{**} below cannot be equal to $\widehat{i}$.

Then let $P_{i^*}^{(1)} = \min_{1 \leq i \leq k-1} P_i$. If $P_{i^*}^{(1)} > C_1$, then accept all $H_{(i+1)i}$ and set $r =$ number of rejections to 0. If $P_{i^*}^{(1)} < C_1$, then reject $H_{(i^*+1)i^*}$ and go to step 2.

At step 2 the populations are divided into 2 subsets. Namely $\{1, \ldots, i^*\}$ and $\{i^*+1, \ldots, k\}$. Each subset now is treated as the original set of k populations to determine if there is a change point within a subset. That is, for the subset $\{1, \ldots, i^*\}$ form i^*-2 two sample Wilcoxon rank sum statistics. One sample of the two comes

from combining the order statistics from populations 1 through d, $d = 2, \ldots, i^* - 1$ and the second sample comes from combining the order statistics from populations $(d+1)$ to i^*. Also form $k - (i^* + 1)$ two sample Wilcoxon rank sum statistics from the populations in $\{i^* + 1, \ldots, k\}$. Once the $(i^* - 2) + k - (i^* + 1) = k - 3$ statistics and their corresponding P-values are determined then proceed as in step 1 except now use the constant $C_2 > C_1$, instead of C_1. That is, if $P^{(2)}_{i^{**}}$ is the minimum P-value of $(k-3)$ P-values and $P^{(2)}_{i^{**}} < C_2$, reject $H_{(i^{**}+1)i^{**}}$ and go to step 3. If $P^{(2)}_{i^{**}} > C_2$ accept all hypotheses except $H_{(i^*+1)i^*}$ and set $r = 1$.

At step 3, the original set $\{1, \ldots, k\}$ could now be partitioned into 3 subsets. If $2 \leq i^{**} \leq i^*$, the subsets are $\{1, \ldots, i^{**}\}$, $\{i^{**} + 1, \ldots, i^*\}$,$\{i^* + 1, \ldots, k\}$. If $i^* + 1 < i^{**} < k - 1$ then the 3 subsets are $\{1, \ldots, i^*\}$, $\{i^* + 1, \ldots, i^{**}\}$ and $\{i^{**} + 1, \ldots, k\}$. Now each of the 3 subsets is treated as $\{1, \ldots, k\}$ was treated in step 1. At each step any single subset could potentially be split into 2 smaller subsets. If the process is concluded at step 3, r is set equal to 2. If the process is not stopped at step 3, then it continues until r is determined. We conclude this section with

Theorem 4.4.4. *The simultaneous interval estimation procedure and MTP resulting from the interval estimation procedure for the change point problem have the interval property.*

Proof. The proof follows from Theorem 4.4.2 and Corollary 4.4.1. □

Remark 4.4.1. *An MTP for the change point problem for normal populations is given in Chen, Cohen and Sackrowitz (2011).*

Remark 4.4.2. *The statistics in the stage 1 MTPs are actually a non parametric version of cusum type statistics where data from i and $i+1$ are never separated.*

4.4.3 All pairwise

Once again we are focused on ν_{21}, so at stage 1 we do a stepwise MTP for $H_{ii'} : \nu_{ii'} = 0$ vs. $K_{ii'} : \nu_{ii'} \neq 0$, $i' < i$, $i' = 1, \ldots, k-1$ and $i = 2, \ldots, k$ but not $i' = 1, i = 2$. That

is, there are $C_2^k - 1$ hypotheses to be tested.

The first step of the stepwise procedure entails forming two sample Wilcoxon rank sum statistics. The pairs of statistics are determined as follows: All 2 set partitions of $I = \{1, 2, \ldots, k\}$ are considered except those for which indices 1 and 2 are separated. For each set of the 2 set partition the samples corresponding to the indices are pooled into one large sample. The samples coming from populations 1 and 2 are amalgamated within the set that they (both must) appear. Then for each 2 set partition the rank sum statistic is computed along with its corresponding P-value.

Let $P_{A_1}^{(1)}$ be the minimum of all P-values at step 1 where $A_1, I\backslash A_1$ is the partition that gave rise to $P_{A_1}^{(1)}$. If $P_{A_1}^{(1)} > C_1$, then accept all $H_{ii'}$. If $P_{A_1}^{(1)} < C_1$, then reject all $H_{ii'}$ where $i \in A_1$ and $i' \in I\backslash A_1$ and go to step 2.

At step 2 treat A_1 and $I\backslash A_1$ as I was treated at step 1. Again find P-values for all 2 sample Wilcoxon rank sum tests and get $P_{A_2}^{(2)}$ the minimum of all such P-values. This time let the set corresponding to this minimum P-value be denoted by A_2 where $A_2 \subset A_1$ or $A_2 \subset I\backslash A_1$. If $P_{A_2}^{(2)} > C_2$ accept the remaining hypotheses and stop, setting $r =$ number of hypotheses rejected at step 1. If $P_{A_2}^{(2)} < C_2$, reject all hypotheses $H_{ii'}$ where $i \in A_2$, $i' \in A_1\backslash A_2$ if $A_2 \subset A_1$. If $A_2 \subset I\backslash A_1$, reject $H_{ii'}$ with $i \in A_2$, $i' \in I\backslash A_1\backslash A_2$. Now go to step 3 where the partition of I consists of 3 sets. The 3 sets are either A_2, $A_1\backslash A_2$, $I\backslash A_1$ or A_1, A_2, $I\backslash A_1\backslash A_2$. Now treat each of the 3 sets as I was at step 1. Continue this process until r is determined.

We conclude this section with

Theorem 4.4.5. *The simultaneous interval estimation procedure and MTP resulting from the estimation procedure for all pairwise contrasts has the interval property.*

Proof. The proof follows from Theorem 4.4.2 and Corollary 4.4.1. □

Remark 4.4.3. *An MTP for all pairwise contrasts for normal distributions is given in Cohen, Sackrowitz and Chen (2010). Also an MTP using rank tests with a different type of interval property is offered in Cohen and Sackrowitz (2012b).*

4.5 Determination of constants and simulation results

We now turn to the related issues of implementing and evaluating the procedures. Basically the constants C_i used in the stage 1 MTPs as well as the $\gamma(r)$ and hence the $B(r)$ constants used in stage 2 are chosen by trial and error using simulation. We begin by searching for a procedure that can be seen, by simulation, to perform well as an overall MTP in the practical problem at hand. This chore is not as daunting as it may seem since there are logical paths to follow. For the C_i a modification of the Benjamini and Gavrilov (2009) critical values tend to work well while, for the $\gamma(r)$, a modification of the Holm (1979) critical values work well. That is, begin by fixing an α_1 and α_2 and taking

$$C(i) = i\alpha_1/(M - i(1 - \alpha_1)) \tag{4.5.1}$$

and

$$\gamma(r) = \alpha_2/(M - r). \tag{4.5.2}$$

Modifying α_1 and α_2 typically leads to an effective procedure. In the treatments versus control model of Section 4.4.1 we did extensive simulations for ten treatments and one control. When viewed as an MTP we compared the performance of the standard step-down procedure based on the Wilcoxon rank sum test with that of the new procedure. In (4.5.1) and (4.5.2) the new procedure used $\alpha_1 = 0.025$ and $\alpha_2 = 0.03$ in order to control the FWER of the overall procedure at 0.05. We studied their performance for a wide variety of possible parameter points. We used the Normal, Uniform, Exponential, Double Exponential and Cauchy distributions. In all cases the two procedures essentially matched one another in terms of FWER control and expected total number of mistakes (Type I plus Type II errors).

However, the new procedure has the interval property while step-down does not. Also the new procedures admit corresponding interval estimates while step-down does not.

Chapter 5

Simulations

5.1 Simulations for comparisons of 5 MTPs

We did some simulations to compare 5 MTPs. We studied MRD, three kinds of two-stage methods and Holm's step down method in the treatments vs control problem. The three new methods are denoted as FSM (first stage MRD), SCM (shortcut method) and FSR (first stage residual) in the tables.

FSM is using original MRD method in the first stage, described in Section 3.2.1.

SCM is also described in Section 3.2.1, is an alternative to FSM. We call it shortcut because the test statistic is simplified.

FSR is using step down in the first stage, with the first step residual from MRD as the statistics and do not recalculate residuals.

Note that the second stage of these three methods are the same.

We consider two different treatments vs control problems by letting the number of treatments M equal to 25 Table 5.1.2-5.1.4). We studies both the sparse case and the non-sparse case of alternative hypotheses. FDR is controled at level $\alpha = 0.05$ for all five MTPs. To control FDR, usually we need to choose a set of critical values by simulation. In our simulation, for the first stage, using $\alpha_1 = 0.1$ in the Benjamini and Gavrilov (2009) critical values in (3.3.1) tends to work well. For the $B(r)$ in the second stage, using $\alpha_2 = 0.031$ for SCM and $\alpha_2 = 0.0375$ for FSM and FSR in the Holm (1979) critical values in (3.3.2) works well.

We studied the performance of these methods for a wide variety of possible parameter points. For each configuration, the number of iterations is 10000. We report the expected number of Type I errors, the expected number of Type II errors, FDR and total number of mistakes which is the sum of the expected number of Type I and Type II errors.

From the simulation results we find that the three new methods FSM, SCM and FSR are all uniformly better than step-down procedures in terms of total errors for most the parameter settings we considered. For a small proportion of true alternatives ($< 20\%$) MRD has fewer numbers of mistakes compared to other procedures. For the proportion of alternatives $> 20\%$, other methods performs quite similarly and much better than MRD. Also, the three new methods control FDR at a similar level as step-down procedures while MRD fails to control FDR at a lower level for non-sparse case. The performances of new MTPs are satisfying.

Table 5.1.1: Comparisons of 5 MTPs for 25 treatments vs 1 control problem

# of means equal to					type I error					type II error				
0	2	-2	4	-4	MRD	FSM	SD	SCM	FSR	MRD	FSM	SD	SCM	FSR
25	0	0	0	0	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0
23	2	0	0	0	0.1	0.1	0.1	0.1	0.1	1.2	1.7	1.7	1.5	1.7
23	0	0	2	0	0.2	0.1	0.1	0.1	0.1	0	0.3	0.3	0.1	0.3
21	2	2	0	0	0.2	0.1	0.1	0.1	0.1	2	3.3	3.3	2.9	3.3
21	4	0	0	0	0.2	0.1	0.1	0.1	0.1	2.5	3.3	3.3	3.1	3.3
21	2	0	0	2	0.2	0.1	0.1	0.1	0.1	1	1.9	2	1.5	1.9
21	2	0	2	0	0.2	0.1	0.1	0.1	0.1	1	1.9	1.9	1.5	1.9
21	0	2	0	2	0.3	0.1	0.1	0.1	0.1	0	0.5	0.6	0.1	0.6
21	0	4	0	0	0.3	0.1	0.1	0.1	0.1	0	0.5	0.6	0.1	0.5
19	4	2	0	0	0.2	0.1	0.1	0.1	0.1	2.8	4.9	5	4.4	4.9
19	6	0	0	0	0.2	0.1	0.1	0.1	0.1	3.8	4.9	5	4.8	4.9
19	4	0	0	2	0.3	0.1	0.1	0.1	0.1	1.9	3.5	3.6	3	3.5
19	4	0	2	0	0.3	0.1	0.1	0.1	0.1	1.9	3.5	3.6	3.1	3.4
19	2	0	2	2	0.3	0.1	0.1	0.1	0.1	0.8	2.1	2.2	1.5	2.1
19	2	0	4	0	0.3	0.1	0.1	0.1	0.2	0.8	2.1	2.2	1.6	1.9
19	0	0	4	2	0.4	0.1	0.1	0.1	0.1	0	0.7	0.8	0.1	0.8
19	0	0	6	0	0.4	0.1	0.1	0.1	0.5	0	0.7	0.8	0.2	0.4
21	4	0	0	0	0.2	0.1	0.1	0.1	0.1	2.5	3.3	3.3	3.1	3.3
21	0	0	4	0	0.3	0.1	0.1	0.1	0.1	0	0.5	0.6	0.1	0.5
17	4	4	0	0	0.3	0.1	0.1	0.1	0.1	3.2	6.4	6.7	5.8	6.4
17	8	0	0	0	0.3	0.1	0.1	0.1	0.1	5.5	6.5	6.6	6.5	6.4
17	4	0	0	4	0.4	0.1	0.1	0.1	0.1	1.6	3.6	3.8	3	3.6
17	4	0	4	0	0.4	0.1	0.1	0.1	0.4	1.7	3.6	3.8	3.1	3
17	0	0	4	4	0.5	0.1	0.1	0.1	0.1	0	0.8	1	0.2	1
17	0	0	8	0	0.5	0.1	0.1	0.1	1	0	0.8	1.1	0.2	0.2
13	8	4	0	0	0.6	0.1	0.1	0.1	0.1	4.9	9.1	9.9	8.7	9
13	12	0	0	0	0.9	0.1	0.1	0.1	0.2	10.2	9.5	9.9	9.9	8.7
13	8	0	0	4	0.8	0.1	0.1	0.1	0.1	4.2	6.3	7.1	6.1	6.4
13	8	0	4	0	0.9	0.1	0.1	0.1	0.4	4.4	6.4	7.1	6.4	5
13	4	0	4	4	0.7	0.1	0.1	0.1	0.2	1.3	3.4	4.2	2.9	3.7
13	4	0	8	0	0.7	0.1	0.1	0.1	0.7	1.3	3.4	4.2	3	2.1
13	0	0	8	4	0.7	0.1	0.1	0.2	0.4	0	0.6	1.4	0.2	0.8
13	0	0	0	12	2.5	0.2	0.1	0.2	1	1.5	0.7	1.5	0.3	0.2

Table 5.1.2: Comparisons of 5 MTPs for 25 treatments vs 1 control problem

# of means equal to					FDR					total errors				
0	2	-2	4	-4	MRD	FSM	SD	SCM	FSR	MRD	FSM	SD	SCM	FSR
25	0	0	0	0	0.05	0.05	0.05	0.05	0.05	0.1	0.1	0.1	0.1	0.1
23	2	0	0	0	0.05	0.039	0.038	0.037	0.039	1.3	1.7	1.8	1.6	1.7
23	0	0	2	0	0.051	0.03	0.029	0.022	0.031	0.2	0.4	0.4	0.2	0.4
21	2	2	0	0	0.045	0.026	0.024	0.023	0.026	2.2	3.4	3.4	3	3.4
21	4	0	0	0	0.057	0.034	0.031	0.033	0.034	2.6	3.4	3.4	3.2	3.4
21	2	0	0	2	0.051	0.021	0.02	0.018	0.022	1.2	2	2	1.6	2
21	2	0	2	0	0.051	0.026	0.024	0.019	0.028	1.2	2	2	1.6	2
21	0	2	0	2	0.05	0.018	0.017	0.015	0.019	0.3	0.6	0.6	0.2	0.6
21	0	4	0	0	0.05	0.022	0.021	0.015	0.026	0.3	0.6	0.6	0.2	0.6
19	4	2	0	0	0.049	0.022	0.019	0.02	0.022	3	5	5.1	4.5	5
19	6	0	0	0	0.074	0.031	0.028	0.032	0.033	4.1	5	5	4.8	5
19	4	0	0	2	0.056	0.018	0.016	0.017	0.018	2.2	3.6	3.7	3.1	3.6
19	4	0	2	0	0.056	0.024	0.022	0.019	0.029	2.2	3.6	3.7	3.2	3.6
19	2	0	2	2	0.051	0.015	0.014	0.014	0.016	1.1	2.2	2.3	1.6	2.2
19	2	0	4	0	0.051	0.019	0.019	0.015	0.035	1.1	2.2	2.3	1.6	2.1
19	0	0	4	2	0.051	0.014	0.013	0.013	0.016	0.4	0.8	0.9	0.3	0.9
19	0	0	6	0	0.051	0.016	0.017	0.013	0.059	0.4	0.8	0.9	0.3	0.9
21	4	0	0	0	0.057	0.034	0.031	0.033	0.034	2.6	3.4	3.4	3.2	3.4
21	0	0	4	0	0.05	0.022	0.021	0.015	0.026	0.3	0.6	0.6	0.2	0.6
17	4	4	0	0	0.047	0.018	0.014	0.016	0.019	3.5	6.5	6.7	5.9	6.5
17	8	0	0	0	0.108	0.03	0.024	0.033	0.037	5.8	6.6	6.7	6.6	6.5
17	4	0	0	4	0.055	0.013	0.011	0.013	0.016	2.1	3.7	3.9	3.1	3.7
17	4	0	4	0	0.055	0.017	0.017	0.014	0.052	2.1	3.7	3.9	3.2	3.4
17	0	0	4	4	0.051	0.011	0.01	0.011	0.012	0.5	0.9	1.1	0.3	1
17	0	0	8	0	0.052	0.013	0.014	0.012	0.093	0.5	0.9	1.1	0.3	1.2
13	8	4	0	0	0.076	0.016	0.009	0.014	0.021	5.6	9.2	10	8.8	9.2
13	12	0	0	0	0.275	0.031	0.018	0.04	0.044	11.1	9.6	9.9	10	8.9
13	8	0	0	4	0.093	0.013	0.008	0.014	0.017	5	6.4	7.1	6.2	6.6
13	8	0	4	0	0.101	0.017	0.012	0.017	0.047	5.2	6.5	7.1	6.5	5.4
13	4	0	4	4	0.053	0.011	0.007	0.012	0.016	1.9	3.5	4.3	3	3.9
13	4	0	8	0	0.058	0.012	0.01	0.013	0.06	2.1	3.5	4.3	3.1	2.8
13	0	0	8	4	0.05	0.01	0.007	0.011	0.027	0.7	0.7	1.5	0.4	1.1
13	0	0	0	12	0.179	0.017	0.009	0.012	0.066	4	0.9	1.5	0.4	1.1

Table 5.1.3: Comparisons of 5 MTPs for 25 treatments vs 1 control problem

# of means equal to					type I error					type II error				
0	2	-2	4	-4	MRD	FSM	SD	SCM	FSR	MRD	FSM	SD	SCM	FSR
18	7	0	0	0	0.3	0.1	0.1	0.1	0.1	4.6	5.7	5.8	5.6	5.6
18	0	0	7	0	0.4	0.1	0.1	0.1	0.8	0.0	0.8	0.9	0.2	0.3
11	7	7	0	0	0.8	0.1	0.0	0.1	0.2	4.8	9.9	11.6	9.8	10.1
11	14	0	0	0	1.4	0.1	0.1	0.1	0.2	12.9	11.1	11.5	11.5	10.2
11	7	0	0	7	1.0	0.2	0.0	0.2	0.4	3.1	5.1	6.5	5.0	4.7
11	7	0	7	0	1.3	0.2	0.1	0.2	0.4	3.5	5.4	6.5	5.4	4.1
11	0	0	7	7	0.8	0.2	0.1	0.2	0.2	0.0	0.4	1.5	0.3	1.1
11	0	0	14	0	8.8	0.6	0.1	0.2	0.8	10.1	0.8	1.6	0.3	0.2
11	14	7	0	0	2.2	0.1	0.0	0.2	0.2	13.1	13.8	17.2	13.4	10.4
4	21	0	0	0	1.3	0.0	0.0	0.2	0.0	20.8	17.2	16.9	16.5	17.2
4	14	0	0	7	2.3	0.1	0.0	0.2	0.3	13.2	10.4	11.8	9.7	6.8
4	14	0	7	0	2.5	0.1	0.0	0.2	0.0	15.9	11.2	11.8	10.5	11.3
4	7	0	7	7	2.8	0.3	0.0	0.2	0.3	6.5	4.0	6.6	4.3	3.6
4	7	0	14	0	3.8	0.3	0.0	0.2	0.1	16.8	4.9	6.7	4.8	4.7
4	0	0	14	7	4.0	0.3	0.0	0.2	0.3	12.7	0.8	1.6	0.4	0.3
4	0	0	0	21	4.0	0.1	0.1	0.2	0.0	20.7	2.7	1.8	0.6	2.2

Table 5.1.4: Comparisons of 5 MTPs for 25 treatments vs 1 control problem

# of means equal to					FDR					total errors				
0	2	-2	4	-4	MRD	FSM	SD	SCM	FSR	MRD	FSM	SD	SCM	FSR
18	7	0	0	0	0.086	0.031	0.026	0.033	0.034	4.9	5.8	5.9	5.7	5.7
18	0	0	7	0	0.051	0.014	0.015	0.012	0.086	0.4	0.9	1.0	0.3	1.1
11	7	7	0	0	0.072	0.016	0.007	0.014	0.022	5.6	10.0	11.6	9.9	10.3
11	14	0	0	0	0.407	0.034	0.016	0.047	0.036	14.3	11.2	11.5	11.7	10.3
11	7	0	0	7	0.081	0.013	0.005	0.013	0.032	4.1	5.3	6.6	5.2	5.1
11	7	0	7	0	0.11	0.017	0.01	0.016	0.038	4.9	5.5	6.6	5.5	4.5
11	0	0	7	7	0.051	0.012	0.005	0.013	0.015	0.8	0.6	1.6	0.5	1.3
11	0	0	14	0	0.736	0.046	0.008	0.014	0.051	18.9	1.4	1.7	0.5	1.0
11	14	7	0	0	0.212	0.012	0.002	0.017	0.019	15.4	14.0	17.3	13.6	10.6
4	21	0	0	0	0.622	0.006	0.007	0.068	0.006	22.2	17.3	16.9	16.6	17.2
4	14	0	0	7	0.215	0.01	0.002	0.015	0.018	15.4	10.5	11.9	9.9	7.0
4	14	0	7	0	0.357	0.014	0.004	0.026	0.005	18.4	11.3	11.9	10.7	11.3
4	7	0	7	7	0.165	0.016	0.002	0.012	0.014	9.3	4.3	6.7	4.5	3.8
4	7	0	14	0	0.541	0.022	0.004	0.018	0.009	20.6	5.1	6.8	5.0	4.8
4	0	0	14	7	0.33	0.015	0.002	0.011	0.014	16.7	1.1	1.6	0.6	0.6
4	0	0	0	21	0.936	0.005	0.004	0.014	0.003	24.6	2.8	1.9	0.8	2.2

5.2 Simulations for the MTP in interval estimation

We did some simulations for the shortcut method (SCM) as an interval estimates method in 25 treatments vs 1 control problem. We compare the performance of SCM

with Dunnett's (DNT in the table) method in terms of false coverage rate (FCR), average length of intervals and probabilities of false coverage of zero. FCR is the expected proportion of parameters not covered by their interval estimates among the selected parameters, where the proportion is 0 if no parameter is selected. See Benjamini, Y. and Yekutieli, Y. (2005) for further discussion of FCR.

The Dunnett's method is a single step method giving equal length intervals for all parameters given the number of treatments and sample size. For SCM, the length of the interval is determined in the second stage, which is related to the number of rejections in the first stage. The set of critical values for SCM in both stages are chosen to make the FDR controlled at level $\alpha = 0.05$ when SCM is performed as a MTP. In our simulation, for the first stage, using $\alpha_1 = 0.1$ in the Benjamini and Gavrilov (2009) critical values in (3.3.1) tends to work well. For the $B(r)$ in the second stage, using $\alpha_2 = 0.031$ in the Holm (1979) critical values in (3.3.2) works well.

From the results we can see that SCM controls the FCR at 0.05 for most parameter settings we considered, while Dunnett's method controls the FCR too strictly. SCM always gives narrower intervals. The probabilities of false coverage of zero using SCM are always less than those using Dunnett's method.

Table 5.2.1: Comparison of interval estimates for 25 treatments vs control problem

# of parameters equals					FCR		average length of intervals					
							SCM					DNT
0	2	-2	4	-4	SCM	DNT	0	2	-2	4	-4	all
25	0	0	0	0	0.047	0.045	6.02					6
23	2	0	0	0	0.050	0.037	5.99	5.53				6
23	0	0	2	0	0.063	0.028	5.95			4		6
21	4	0	0	0	0.042	0.030	5.95	5.61				6
21	2	2	0	0	0.045	0.023	5.95	5.35	5.35			6
21	2	0	2	0	0.049	0.021	5.90	5.34		4.16		6
21	2	0	0	2	0.053	0.016	5.90	5.34			3.94	6
21	0	0	4	0	0.054	0.018	5.86			3.96		6
21	0	0	2	2	0.057	0.016	5.86			3.95	3.95	6
19	6	0	0	0	0.039	0.026	5.89	5.71				6
19	4	2	0	0	0.043	0.019	5.89	5.44	5.2			6
19	4	0	2	0	0.040	0.019	5.84	5.44		4.4		6
19	4	0	0	2	0.051	0.014	5.85	5.43			3.9	6
19	2	0	4	0	0.046	0.017	5.80	5.19		4.12		6
19	2	0	2	2	0.049	0.012	5.79	5.18		4.11	3.91	6
19	0	0	6	0	0.053	0.015	5.76			3.93		6
19	0	0	4	2	0.052	0.011	5.76			3.91	3.9	6
21	4	0	0	0	0.050	0.037	5.95	5.6				6
21	0	0	4	0	0.063	0.028	5.86			3.96		6
17	8	0	0	0	0.042	0.030	5.81	5.8				6
17	4	4	0	0	0.045	0.023	5.82	5.28	5.28			6
17	4	0	4	0	0.049	0.021	5.72	5.26		4.37		6
17	4	0	0	4	0.053	0.016	5.72	5.25			3.87	6
17	0	0	8	0	0.054	0.018	5.64			3.9		6
17	0	0	4	4	0.057	0.016	5.63			3.88	3.88	6
13	12	0	0	0	0.039	0.026	5.58	5.88				6
13	8	4	0	0	0.043	0.019	5.58	5.45	4.92			6
13	8	0	4	0	0.040	0.019	5.48	5.44		5.1		6
13	8	0	0	4	0.051	0.014	5.48	5.43			3.82	6
13	4	0	8	0	0.046	0.017	5.38	4.9		4.32		6
13	4	0	4	4	0.049	0.012	5.38	4.89		4.25	3.83	6
13	0	0	12	0	0.053	0.015	5.29			3.88		6
13	0	0	8	4	0.052	0.011	5.28			3.85	3.83	6

Table 5.2.2: Comparison of interval estimates for 25 treatments vs control problem

# of treatments equal to					False coverage probability							
					SCM				Dunnett			
0	2	-2	4	-4	2	-2	4	-4	2	-2	4	-4
25	0	0	0	0								
23	2	0	0	0	0.759				0.841			
23	0	0	2	0			0.031				0.159	
21	4	0	0	0	0.774				0.841			
21	2	2	0	0	0.732	0.726			0.842	0.837		
21	2	0	2	0	0.732		0.036		0.848		0.155	
21	2	0	0	2	0.731			0.026	0.840			0.161
21	0	0	4	0			0.028				0.156	
21	0	0	2	2			0.025	0.026			0.162	0.156
19	6	0	0	0	0.791				0.837			
19	4	2	0	0	0.756	0.709			0.845	0.843		
19	4	0	2	0	0.753		0.047		0.846		0.158	
19	4	0	0	2	0.753			0.023	0.846			0.162
19	2	0	4	0	0.707		0.035		0.838		0.163	
19	2	0	2	2	0.703		0.034	0.025	0.843		0.154	0.157
19	0	0	6	0			0.024				0.157	
19	0	0	4	2			0.024	0.024			0.155	0.162
21	4	0	0	0	0.774				0.841			
21	0	0	4	0			0.027				0.159	
17	8	0	0	0	0.810				0.841			
17	4	4	0	0	0.720	0.022			0.842	0.837		
17	4	0	4	0	0.720		0.024		0.848		0.158	
17	4	0	0	4	0.719			0.023	0.840			0.16
17	0	0	8	0			0.024				0.156	
17	0	0	4	4			0.022	0.022			0.162	0.16
13	12	0	0	0	0.820				0.837			
13	8	4	0	0	0.762	0.666			0.845	0.838		
13	8	0	4	0	0.758		0.084		0.846		0.16	
13	8	0	0	4	0.760			0.021	0.846			0.157
13	4	0	8	0	0.664		0.041		0.838		0.165	
13	4	0	4	4	0.654		0.036	0.022	0.843		0.151	0.156
13	0	0	12	0			0.022				0.157	
13	0	0	8	4			0.020	0.019			0.155	0.164

Table 5.2.3: Comparison of interval estimates for 25 treatments vs control problem

# of parameters equals					FCR		average length of intervals					
							SCM					DNT
0	2	-2	4	-4	SCM	DNT	0	2	-2	4	-4	all
18	7	0	0	0	0.041	0.025	5.85	5.75				6
18	0	0	7	0	0.054	0.014	5.7			3.9		6
11	14	0	0	0	0.051	0.018	5.39	5.88				6
11	7	7	0	0	0.036	0.008	5.37	5.1	5.09			6
11	7	0	7	0	0.030	0.009	5.2	5.09		4.89		6
11	7	0	0	7	0.052	0.005	5.19	5.07			3.82	6
11	0	0	14	0	0.053	0.006	5.03			3.88		6
11	0	0	7	7	0.056	0.005	5.02			3.83	3.83	6
11	21	0	0	0	0.075	0.009	4.3	5.66				6
4	14	7	0	0	0.050	0.005	4.17	5.04	5.03			6
4	14	0	7	0	0.033	0.004	4.1	5.03		4.25		6
4	14	0	0	7	0.054	0.004	4.09	5.01			3.78	6
4	7	0	14	0	0.036	0.003	3.96	4.25		4.91		6
4	7	0	7	7	0.047	0.004	3.97	4.24		4.7	3.78	6
4	0	0	21	0	0.055	0.004	3.91			3.94		6
4	0	0	14	7	0.055	0.003	3.91			3.83	3.77	6

Table 5.2.4: Comparison of interval estimates for 25 treatments vs control problem

# of treatments equal to					False coverage probability							
					SCM				Dunnett			
0	2	-2	4	-4	2	-2	4	-4	2	-2	4	-4
18	7	0	0	0	0.802				0.841			
18	0	0	7	0			0.025				0.025	
11	14	0	0	0	0.826				0.841			
11	7	7	0	0	0.696	0.696			0.842	0.842		
11	7	0	7	0	0.695		0.695		0.848		0.848	
11	7	0	0	7	0.692			0.692	0.840			0.84
11	0	0	14	0			0.023				0.156	
11	0	0	7	7			0.019	0.019			0.162	0.162
11	21	0	0	0	0.780				0.837			
4	14	7	0	0	0.692	0.550			0.845	0.849		
4	14	0	7	0	0.689		0.130		0.846		0.157	
4	14	0	0	7	0.685			0.017	0.846			0.156
4	7	0	14	0	0.546		0.073		0.838		0.157	
4	7	0	7	7	0.540		0.060	0.019	0.843		0.151	0.158
4	0	0	21	0			0.025				0.157	
4	0	0	14	7			0.020	0.018			0.155	0.163

5.3 Simulations for nonparametric MTPs

We did some simulations to compare the proposed nonparametric shortcut method (SCM) and the step-down method based on the Wilcoxon rank sum test in the treatments vs control problem for unknown distribution. The nonparametric SCM is described in detail in Section 4.4.1. We considered five different distributions including double exponential, normal, Cauchy, exponential, and uniform (Table 5.3.1-5.3.5). We present the results of the 25 treatments vs 1 control problem with sample size 10. Independent observations have been generated for each shift and distribution. 5,000 replications of each translation or shift configuration were performed.

FWER is controled at level $\alpha = 0.05$ for both methods. To control FWER, usually we need to choose a set of critical values by simulation. In our simulation, we found that in the first stage, using $\alpha_1 = 0.045$ in the Benjamini and Gavrilov (2009) critical values in (4.5.1) tend to work well. For the $B(r)$ in the second stage, using $\alpha_2 = 0.045$ in the Holm (1979) critical values in (4.5.2) work well.

We studied their performance for a wide variety of possible treatment shifts. The shifts shown in the table have been adjusted for the sample size we chose. We report the expected number of Type I errors, the expected number of Type II errors, FWER and total errors.

From the results we can see that for all 5 different distributions, using the same sets of critical values we selected, FWER can be controlled very well at all the parameter points we explored using the new method. The new method SCM is uniformly better than than step-down procedures in terms of total errors for double exponential, exponential and uniform distribution. For normal and Cauchy distribution, SCM matches the performance of step-down procedure. We have pointed out that the new procedures have the interval property while the step-down procedure does not. Also the new procedures admit corresponding interval estimates while the step-down procedure does not.

Table 5.3.1: MTPs in 25 treatments vs 1 control with double exponential distribution

# of means equal to					type I		type II		FWER		total errors	
0	0.63	0.63	1.27	-1.27	SCM	SD	SCM	SD	SCM	SD	SCM	SD
25	0	0	0	0	0.1	0.1	0	0	0.047	0.043	0.1	0.1
23	2	0	0	0	0.1	0.1	1.8	1.9	0.04	0.036	1.9	1.9
23	0	0	2	0	0.1	0.1	0.9	1.3	0.029	0.029	1	1.3
21	2	2	0	0	0.1	0.1	3.6	3.8	0.03	0.027	3.7	3.8
21	4	0	0	0	0.1	0.1	3.7	3.8	0.035	0.03	3.8	3.8
21	2	0	0	2	0.1	0.1	2.6	3.2	0.023	0.022	2.7	3.2
21	2	0	2	0	0.1	0.1	2.8	3.1	0.026	0.026	2.9	3.2
21	0	0	2	2	0.1	0.1	1.5	2.5	0.018	0.019	1.6	2.6
21	0	0	4	0	0.1	0.1	1.8	2.5	0.021	0.023	1.9	2.5
19	4	2	0	0	0.1	0.1	5.5	5.6	0.024	0.021	5.5	5.7
19	6	0	0	0	0.1	0.1	5.6	5.6	0.031	0.025	5.7	5.7
19	4	0	0	2	0.1	0.1	4.4	5	0.019	0.017	4.4	5.1
19	4	0	2	0	0.1	0.1	4.7	5	0.024	0.021	4.8	5
19	2	0	2	2	0.1	0.1	3.3	4.4	0.016	0.016	3.4	4.4
19	2	0	4	0	0.1	0.1	3.8	4.3	0.02	0.019	3.9	4.4
19	0	0	4	2	0.1	0.1	2.2	3.7	0.015	0.014	2.3	3.7
19	0	0	6	0	0.1	0.1	2.8	3.7	0.019	0.017	2.9	3.7

Table 5.3.2: MTPs in 25 treatments vs 1 control with normal distribution

# of means equal to					type I		type II		FWER		total errors	
0	0.63	0.63	1.27	-1.27	SCM	SD	SCM	SD	SCM	SD	SCM	SD
25	0	0	0	0	0.1	0.1	0	0	0.04	0.038	0.1	0.1
23	2	0	0	0	0.1	0.1	1.9	1.9	0.035	0.032	2	2
23	0	0	2	0	0.1	0.1	1.2	1.5	0.026	0.028	1.3	1.6
21	2	2	0	0	0.1	0.1	3.8	3.9	0.028	0.026	3.9	3.9
21	4	0	0	0	0.1	0	3.9	3.9	0.031	0.027	3.9	3.9
21	2	0	0	2	0.1	0.1	3.1	3.4	0.022	0.022	3.1	3.5
21	2	0	2	0	0.1	0.1	3.2	3.4	0.024	0.024	3.3	3.5
21	0	0	2	2	0.1	0.1	2.2	3	0.017	0.018	2.3	3
21	0	0	4	0	0.1	0.1	2.6	3	0.02	0.023	2.6	3
19	4	2	0	0	0.1	0	5.7	5.8	0.024	0.021	5.8	5.8
19	6	0	0	0	0.1	0	5.8	5.8	0.028	0.024	5.8	5.8
19	4	0	0	2	0.1	0.1	4.9	5.3	0.019	0.017	4.9	5.4
19	4	0	2	0	0.1	0.1	5.2	5.4	0.023	0.022	5.3	5.4
19	2	0	2	2	0.1	0.1	4.1	4.9	0.015	0.016	4.2	4.9
19	2	0	4	0	0.1	0.1	4.6	4.9	0.02	0.02	4.7	4.9
19	0	0	4	2	0.1	0.1	3.3	4.4	0.014	0.014	3.4	4.5
19	0	0	6	0	0.1	0.1	4	4.4	0.021	0.019	4.1	4.5

Table 5.3.3: MTPs in 25 treatments vs 1 control with Cauchy distribution

# of means equal to					type I		type II		FWER		total errors	
0	0.63	0.63	1.27	-1.27	SCM	SD	SCM	SD	SCM	SD	SCM	SD
25	0	0	0	0	0.1	0.1	0	0	0.047	0.036	0.1	0.1
23	2	0	0	0	0.1	0	2	2	0.041	0.032	2	2
23	0	0	2	0	0.1	0.1	1.9	1.9	0.04	0.03	1.9	2
21	2	2	0	0	0.1	0	3.9	4	0.035	0.027	4	4
21	4	0	0	0	0.1	0	4	4	0.038	0.028	4	4
21	2	0	0	2	0.1	0	3.9	3.9	0.031	0.024	3.9	3.9
21	2	0	2	0	0.1	0	3.9	3.9	0.034	0.026	3.9	3.9
21	0	0	2	2	0.1	0	3.7	3.8	0.03	0.022	3.8	3.9
21	0	0	4	0	0.1	0	3.8	3.8	0.034	0.025	3.8	3.9
19	4	2	0	0	0.1	0	5.9	5.9	0.032	0.025	6	6
19	6	0	0	0	0.1	0	5.9	5.9	0.035	0.025	6	6
19	4	0	0	2	0.1	0	5.8	5.9	0.029	0.023	5.9	5.9
19	4	0	2	0	0.1	0	5.8	5.9	0.034	0.024	5.9	5.9
19	2	0	2	2	0.1	0	5.7	5.8	0.026	0.021	5.8	5.8
19	2	0	4	0	0.1	0	5.8	5.8	0.033	0.023	5.8	5.8
19	0	0	4	2	0.1	0	5.6	5.7	0.025	0.02	5.7	5.8
19	0	0	6	0	0.1	0	5.7	5.7	0.033	0.022	5.7	5.8

Table 5.3.4: MTPs in 25 treatments vs 1 control with exponential distribution

# of means equal to					type I		type II		FWER		total errors	
0	0.63	0.63	1.27	-1.27	SCM	SD	SCM	SD	SCM	SD	SCM	SD
25	0	0	0	0	0.1	0	0	0	0.044	0.033	0.1	0
23	2	0	0	0	0.1	0	1.7	1.8	0.029	0.025	1.8	1.9
23	0	0	2	0	0.1	0.1	0.4	1.1	0.022	0.023	0.5	1.2
21	2	2	0	0	0.1	0	3.4	3.6	0.023	0.019	3.5	3.7
21	4	0	0	0	0.1	0	3.5	3.6	0.024	0.021	3.6	3.7
21	2	0	0	2	0.1	0.1	2.5	2.9	0.018	0.016	2.5	3
21	2	0	2	0	0.1	0	2.3	2.9	0.019	0.019	2.4	2.9
21	0	0	2	2	0.1	0.1	1.1	2.2	0.016	0.015	1.2	2.2
21	0	0	4	0	0.1	0	1	2.2	0.017	0.018	1	2.2
19	4	2	0	0	0.1	0	5.2	5.4	0.019	0.015	5.3	5.5
19	6	0	0	0	0	0	5.4	5.4	0.02	0.017	5.4	5.5
19	4	0	0	2	0.1	0.1	2.8	3.7	0.019	0.017	2.9	3.7
19	4	0	2	0	0.1	0	4.2	4.7	0.017	0.016	4.3	4.7
19	2	0	2	2	0.1	0	2.8	3.9	0.013	0.012	2.9	4
19	2	0	4	0	0.1	0	3	3.9	0.016	0.015	3	4
19	0	0	4	2	0.1	0	1.4	3.2	0.014	0.011	1.5	3.2
19	0	0	6	0	0.1	0	1.6	3.2	0.018	0.014	1.7	3.2

Table 5.3.5: MTPs in 25 treatments vs 1 control with uniform distribution

# of means equal to					type I		type II		FWER		total errors	
0	0.63	0.63	1.27	-1.27	SCM	SD	SCM	SD	SCM	SD	SCM	SD
25	0	0	0	0	0.1	0.1	0	0	0.044	0.043	0.1	0.1
23	2	0	0	0	0.1	0.1	1.9	2	0.039	0.038	2	2
23	0	0	2	0	0.1	0.1	1.4	1.6	0.029	0.031	1.5	1.7
21	2	2	0	0	0.1	0.1	3.9	3.9	0.031	0.03	3.9	4
21	4	0	0	0	0.1	0.1	3.9	3.9	0.033	0.031	3.9	4
21	2	0	0	2	0.1	0.1	3.3	3.6	0.023	0.024	3.4	3.6
21	2	0	2	0	0.1	0.1	3.4	3.5	0.029	0.028	3.5	3.6
21	0	0	2	2	0.1	0.1	2.6	3.2	0.018	0.019	2.7	3.3
21	0	0	4	0	0.1	0.1	2.9	3.2	0.026	0.026	3	3.2
19	4	2	0	0	0	0.1	5.8	5.9	0.026	0.024	5.9	5.9
19	6	0	0	0	0.1	0.1	5.8	5.8	0.028	0.026	5.9	5.9
19	4	0	0	2	0.1	0.1	5.2	5.5	0.019	0.019	5.2	5.6
19	4	0	2	0	0.1	0.1	5.4	5.5	0.026	0.024	5.4	5.5
19	2	0	2	2	0.1	0.1	4.6	5.1	0.017	0.017	4.6	5.2
19	2	0	4	0	0.1	0.1	5	5.1	0.024	0.023	5	5.2
19	0	0	4	2	0.1	0.1	4	4.7	0.016	0.015	4	4.8
19	0	0	6	0	0.1	0.1	4.5	4.7	0.025	0.021	4.6	4.8

References

[1] Basso, D., Pesarin, F., Salmaso, L. and Solari, A. (2009). *Permutation Tests for Stochastic Ordering and ANOVA*. Springer.

[2] Benjamini, Y., Hochberg, Y and Stark, P.B. (1998). Confidence intervals with more power to determine the sign. *Journal of the Statistical Association* **93**, 309-317.

[3] Benjamini, Y. and Gavrilov, Y. (2009). A simple forward selection procedure based on false discovery rate control. *Annals of Applied Statistics* **3**, 179-198.

[4] Benjamini, Y. and Stark, P.B. (1996). Nonequivariant Simultaneous Confidence intervals less likely to contain zero. *Journal of the American Statistical Association* **91**, 329-337.

[5] Benjamini, Y. and Yekutieli, Y. (2005). False discovery rate controlling confidence intervals for selected parameters. *Journal of the American Statistical Association* **100**, 71-80.

[6] Campbell, G. and Skillings,J.H. (1985). Nonparametric Stepwise Multiple Comparison Procedures. *Journal of the American Statistical Association* **80**, 998–1003.

[7] Chen, C., Cohen, A. and Sackrowitz, H.B. (2009). Admissible, consistent multiple testing with applications including variable selection. *Electronic Journal of Statistics* **3**, 633-650.

[8] Chen, K., Cohen, A. and Sackrowitz, H.B. (2011). Consistent multiple testing for change points. *Journal of Multivariate Analysis* **102**, 1339-1343.

[9] Cohen, A., Ma, Y. and Sackrowitz, H.B. (2012). Multiple testing procedures and simultaneous interval estimates. Submitted.

[10] Cohen, A. and Sackrowitz, H.B. (2012a). The interval property in multiple testing of pairwise differences. *Statistical Science*. to appear.

[11] Cohen, A. and Sackrowitz, H.B. (2012b). Multiple rank test for pairwise comparisons. *IMS Collections*. to appear.

[12] Cohen, A. and Sackrowitz, H.B. (2008). Multiple testing of two-sided alternatives with dependent data. *Statistica Sinica* **18**, 1593 - 1602.

[13] Cohen, A., Sackrowitz, H.B. and Chen, C. (2010). Multiple testing of pairwise comparisons. *IMS Collections. Borrowing Strength: Theory Powering Applications - A Festschrift for Lawrence D. Brown* **6**, 144-157.

[14] Cohen, A., Sackrowitz, H.B. and Xu, M. (2009). A new multiple testing method in the dependent case, *Annals of Statistics* **37(3)**, 1518-544.

[15] Devore, J.L. (2012). *Probability and Statistics for Engineering and the Sciences*, Eighth Edition, Brooks/Cole.

[16] Dudoit, S., Shaffer, J.P., and Boldrick, J.C. (2003). Multiple hypothesis testing in microarray experiments. *Statistical Science* **18**, 71-103.

[17] Efron, B. and Tibshirani, R.J. (1994). *An Introduction to the Bootstrap* . Chapman and Hall/CRC.

[18] Guilbaud, O. (2008). Simultaneous confidence regions corresponding to Holm's step-down procedure and other closed-testing procedures. *Biometrical Journal* **50**, 678-692.

[19] Hayter, A.J. and Hsu, J.C. (1994). On the relationship between stepwise decision procedures and confidence sets. *Journal of the American Statistical Association* **89,** 128-136.

[20] Hochberg, Y. and Tamhane, A.C. (1987). *Multiple Comparison Procedures*. Wiley, New York.

[21] Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**, 65-70.

[22] Hsu, J.C. (1996). *Multiple Comparisons: Theory and Methods*. Chapman and Hall, London.

[23] Lehmann, E.L. (1986). *Testing Statistical Hypotheses*, 2^{nd} Ed. Springer

[24] Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses*, 2^{nd} Ed. Springer

[25] Miller, R.G. (1996). *Simultaneous Statistical Inference*. McGraw-Hill, New York.

[26] Stefansson, G., Kim, W., and Hsu, J.C. (1988). On confidence sets in multiple comparisons. *Statistical Decision Theory and Related Topics IV (Eds.: Gupta, S.S. and Berger, J.O.)* Volume 2, pages 89-104. Springer-Verlag, New York.

[27] Vostrikova, L. J. (1981). Detecting disorder in multidimensional random processes. *Soviet Math Doklady* **24(1)**, 55-59.

Vita

Yingqiu Ma

- **EDUCATION**

2006 B. S. in Mathematics, Fudan University, Shanghai, China.

2011 M. S. in Mathematics with Option in Mathematical Finance, Rutgers University, Piscataway, NJ.

2012 Ph.D. candidate in Statistics, Rutgers University, Piscataway, NJ.

- **PROFESSIONAL EXPERIENCE**

2006 Fellowship, Department of Statistics and Biostatistics, Rutgers University.

2007-2010 Teaching assistant, Department of Statistics and Biostatistics, Rutgers University.