ADMISSIBILITY AND CONSISTENCY FOR MULTIPLE COMPARISON PROBLEMS WITH DEPENDENT VARIABLES

BY KUO-MEI CHEN

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

Written under the direction of
Arthur Cohen, Harold B. Sackrowitz

and approved by

________________________________________

________________________________________

________________________________________

________________________________________

New Brunswick, New Jersey
January, 2011
ABSTRACT OF THE DISSERTATION

Admissibility and consistency for multiple comparison problems with dependent variables

by Kuo-mei Chen

Dissertation Director: Arthur Cohen, Harold B. Sackrowitz

In response to the need of dealing with high dimensional data, new multiple testing procedures (MTPs) based on p-values were developed to improve average power while controlling error rates at the same time. Although MTPs are well accepted and practiced in many disciplines, little attention has been put on a decision theory approach to evaluate MTPs. In a series of papers, Cohen and Sackrowitz (2005a; 2005b; 2007; 2008) laid out the foundation to assess MTPs by studying both type I and type II errors. Here we extend that work and focus on some properties, such as admissibility, p-value monotonicity and consistency, to assess frequently used and new MTPs. Applications include all-pairwise comparisons in anova models and change point problems. In addition, the development of admissible procedures for a matrix order problem is presented.
Acknowledgements

I am most grateful to my advisers, Dr. Arthur Cohen and Dr. Harold Sackrowitz, for their valuable guidance, insight and knowledge. Without their assistance and encouragement, this thesis might have been indefinitely delayed. I am also indebted to other committee members, Dr. Cun-Hui Zhang and Dr. Derek Gordon, for reading the manuscript and making a number of helpful comments and suggestions.

I gratefully acknowledge the faculty and staff of Department of statistics at Rutgers. Throughout the process of my study, I greatly benefited from their wonderful lectures and facilities.

Last but no means least, I wish to thank my husband, Ti-chiun Chang, for his patient understanding and loyal support.
Dedication

To my parents
# Table of Contents

Abstract ........................................................................................................... ii

Acknowledgements .......................................................................................... iii

Dedication .......................................................................................................... iv

List of Tables .................................................................................................... vii

List of Figures ................................................................................................... viii

1. Introduction .................................................................................................. 1

2. Admissibility and monotonicity of all-pairwise testing procedures .... 6
   2.1. Notations and preliminaries ................................................................. 6
   2.2. Single-step procedures ....................................................................... 7
   2.3. Stepwise procedures ......................................................................... 9
       2.3.1. Multiple range test ................................................................. 9
       2.3.2. Multiple $\chi^2$ test ............................................................ 12
       2.3.3. Closed testing method ......................................................... 14
       2.3.4. Fisher’s least significance method ....................................... 19
   2.4. Graphical interpretation of admissibility ......................................... 21
   2.5. P-value monotonicity ....................................................................... 21

3. Cluster separation measure and consistency of PADD and PADD+ ... 26
   3.1. PADD and PADD+ procedures .................................................... 27
   3.2. Cluster separation measure ............................................................. 28
       3.2.1. Definitions and notations .................................................... 28
       3.2.2. Simulation results ............................................................ 30
3.3. Consistency results .................................................. 32

4. Consistency of change point problem ............................. 46
   4.1. Notations and preliminaries ...................................... 47
   4.2. BSP and MRD procedures ......................................... 47
   4.3. Consistency results ................................................ 49

5. MTPs for the matrix order problem ............................... 55
   5.1. Step-down procedure ............................................. 57
   5.2. Simple order constraints ......................................... 59
   5.3. Closed testing method ........................................... 65
   5.4. MRD in matrix order ............................................. 70
      5.4.1. MRD using likelihood ratio test ......................... 71
      5.4.2. MRD using z test ........................................... 75
   5.5. Simulation results ................................................ 77

Vita ................................................................. 86
List of Tables

3.1. Seven p-clusters used in the simulation study ........................................ 31
5.1. Comparison of the modified MRD and step-down procedures using like-
    lihood ratio test statistics for a $2 \times 3$ matrix order problem .................... 79
5.2. Comparison of the modified MRD and step-down procedures using $z$
    statistics for a $2 \times 3$ matrix order problem ........................................ 80
List of Figures

2.1. Acceptance and rejection region for the single-step test . . . . . . . . 22
2.2. Acceptance and rejection region for the multiple $\chi^2$ test . . . . . 22
2.3. Acceptance and rejection region for the multiple range test . . . . . 23
3.1. Simulated FCRT and pure true outcome cluster rate of PADD+ and the
    Newman-Keuls for patterns 1–4 . . . . . . . . . . . . . . . . . . . . . . . 33
3.2. Simulated FCRT and pure true outcome cluster rate PADD+ and the
    Newman-Keuls for patterns 5–7 . . . . . . . . . . . . . . . . . . . . . . . 34
3.3. Simulated FCRT and pure true outcome cluster rate of PADD and the
    Newman-Keuls for patterns 1–4 . . . . . . . . . . . . . . . . . . . . . . . 35
3.4. Simulated FCRT and pure true outcome cluster rate PADD and the
    Newman-Keuls for patterns 5–7 . . . . . . . . . . . . . . . . . . . . . . . 36
5.1. Simulated average power using likelihood ratio test statistics for a $2 \times 3$
    matrix order problem . . . . . . . . . . . . . . . . . . . . . . . . . . . . 81
5.2. Simulated average power using z statistics for a $2 \times 3$ matrix order problem 82
Chapter 1
Introduction

The topic of multiple testing procedures (MTPs) had been around in the late 1940s and early 1950s. Most works were established by Duncan, Roy, Scheffé and Tukey. During the 1970s, the publication of Miller’s (1966) book helped make the use of MTPs more popular and also provoked an interest in new research in this field. Nowadays commonly used MTPs are discussed in many statistical textbooks, elementary or advanced, such as Stapleton (1995), Hochberg and Tamhane (1987) and Lehmann and Romano (2005).

A resurgent interest in MTPs has occurred in the last decade to satisfy the need to study high dimensional data in many fields, like microarrays, astronomy, finance, proteomics, cytometry and many others. However, traditional MTPs like Scheffé, Tukey and Bonferroni are deemed too conservative when the number of tests is large. Therefore, new methods were developed to improve average power while controlling error rates, either family wise error rate (FWER), weak or strong, or false discovery rate (FDR).

Most of those new approaches are classified as stepwise or multistep procedures as compared to the traditional single-step procedures. See Hochberg and Tamhane (1987) and Dudoit, Shaffer and Boldrick (2003) for discussion of these procedures. Typically, the stepwise procedures utilize p-values determined from marginal distributions of each testing problem. The implementation is usually easy and straightforward. Nevertheless, when the model of interest entails correlated variables, the p-values from marginal distributions, ignoring the information from the structure of the problem, will induce some drawbacks.

In connection with this issue, Cohen and Sackrowitz (2005a; 2005b; 2007; 2008) and Cohen, Kolassa and Sackrowitz (2007) showed that, typical step-up or step-down
procedures are inadmissible for various loss functions that involve both type I and type II errors. In particular, Cohen and Sackrowitz (2007) illustrated that for a multivariate normal models with nonzero correlations, there exist procedures whose individual tests have smaller expected type I and type II errors. Furthermore, Cohen, Sackrowitz and Xu (2009) offered an admissible MTP, maximal residual down (MRD), for models with dependent variables.

Many MTPs involve test statistics \( T_i \), where marginal distribution depends on a parameter \( \theta_i \). The problem concerns testing \( H_i : \theta_i = \theta_{0i} \). Let \( p_i \) represents the p-value for testing \( H_i \) based on \( T_i \). Then the MTP is said to be p-value monotone if given any \( i^* \) for which \( H_{i^*} \) is rejected, \( H_i \) would also be rejected for all \( i \) such that \( p_i \leq p_{i^*} \).

Consistency is often used to assess the quality of a variable selection procedure. Variable selection procedures choose important predictor variables step by step. In some sense, it is comparable to a stepwise MTP. See Cohen, Sackrowitz and Chen (2009) for more details. Thus, consistency can be used as another criteria to evaluate the quality of MTPs.

There are several definitions of consistency. Some consider the accuracy of selection results when the number of variables goes to infinity and other focus is on the accuracy when sample size is large. Throughout this thesis, we are concerned with the latter definition. That is, 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. A formal definition will be given in Chapter 3.

In this thesis, we aim to assess the admissibility, p-value monotonicity and consistency properties of MTPs for some models with dependent variables and also develop an extension of MRD to the matrix order problem. In particular, we examine the admissibility of commonly used MTPs for the all-pairwise testing problem, the interpretability and consistency of partitioned average difference down (PADD) and partitioned average difference down plus (PADD+), the procedures given in Cohen, Sackrowitz and Chen (2010), extensions of MRD for all-pairwise testing problem, consistency of MRD for the change point problem, and the admissibility of MRD for the matrix order problem.

To begin with, note that the testing results from MTPs would induce individual
tests, $\psi_i$, for individual hypotheses. Although tracking the individual test induced from MTPs can be complicated, the performance of the individual test is fundamental and essential to evaluate any MTP as a whole. Even more, recall that a testing procedure is admissible if there exists no single test that has uniformly greater or matching power while the size is not greater. Thus, if the individual test induced from a MTP is admissible, the MTP is admissible for a vector risk function consisting of risk functions of individual components. See Cohen and Sackrowitz (2005b).

Now assume a balanced fixed effects one-way analysis of variance model. Let $X$ be $1 \times k$, independent normally distributed random variable with mean $\mu$ and common known variance $\sigma^2$. The all-pairwise testing problem in this model concerns the mean difference among all $L = k(k - 1)/2$ pairs. That is, $H_{ij}: \mu_i = \mu_j$ vs. $K_{ij}: \mu_i \neq \mu_j$, for $i \neq j$.

Matthes and Truax (1967) offered a necessary and sufficient condition for an individual test to be admissible. It requires that the acceptance sections for an individual test to be convex in the testing variables while other variables being fixed. Cohen, Sackrowitz and Chen (2010) applied the MRD procedure to the all-pairwise testing problem and proved the extensions of MRD, namely PADD and PADD+, are admissible while typical step-down procedures are inadmissible. We extend the work to other commonly used all-pairwise testing MTPs and find that they possess the convexity property. In addition, the p-value monotonicity for those frequently used all-pairwise testing MTPs is also examined. It is concluded that, single-step procedures are p-value monotone while most stepwise procedures are not p-value monotone when the number of populations is greater than or equal to four.

Another practical issue concerning MTPs is logical restrictions. Oftentimes, one expects test results from MTPs to be interpretable. That is, following transitive order. For instance, If $\mu_1 = \mu_2$ and $\mu_2 = \mu_3$ are ascertained, one would suppose $\mu_1 = \mu_3$ is also accepted. However, that is not the case in practice.

Shaffer (2008) gave a framework to measure interpretability for all-pairwise testing MTPs. Her simulation study shows that the Newman-Keuls procedure performs better than the Benjamini-Hochberg procedure. Following Shaffer’s framework, we compare
the performances of the Newman-Keuls procedure, PADD and PADD+. The Newman-Keuls procedure dominates PADD+ for most of the cases and PADD has higher pure true outcome rate than the Newman-Keuls does in some cases. Here the pure true outcome rate is used to measure the interpretability of testing results, and a formal definition of cluster rate will be given in Chapter 3. Additionally, the proof of consistency for PADD and PADD+ are included.

Assume the same balanced fixed effects one-way analysis of variance model as stated before. The change point problem in this model 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 \( i = 1, \ldots, k-1 \). Cohen, Sackrowitz and Xu (2009) indicated that most stepwise procedures for the change point problem are inadmissible for a risk function concerning both type I and type II errors, while MRD and the binary segmentation procedure (BSP) are admissible. We investigate the consistency property of MRD and BSP and find them both consistent. That is, the probability of making any error for either of these two procedures is zero for sufficiently large sample size.

Consider a two-way design matrix. Let \( Y_{ijk} \) be the \( k \)th observation from the \( i \)th row and \( j \)th column, \( i = 1, \ldots, R \) and \( j = 1, \ldots, C \). Assume all observations \( Y_{ijk} \) to be independent normally distributed with mean \( \mu_{ij} \) and common known variance \( \sigma^2 \). The pair differences concerned in the matrix order problem can be formulated as \( H_{ij'} : \mu_{ij'} = \mu_{ij'+1} \) vs. \( K_{ij'} : \mu_{ij'} < \mu_{ij'+1} \) for \( i = 1, \ldots, R, j' = 1, \ldots, C-1 \), and \( H_{i'j} : \mu_{i'j} = \mu_{i'+1,j} \) vs. \( K_{i'j} : \mu_{i'j} < \mu_{i'+1,j} \) for \( i' = 1, \ldots, R-1 \) and \( j = 1, \ldots, C \). The conventional step-down procedure is inadmissible for the matrix order problem, while simple order constraint (SOC) procedure is admissible. Moreover, the closed testing method for a \( 2 \times 2 \) matrix order problem is admissible and a modification of MRD for a general matrix order problem is admissible. However, the modified MRD with orthogonal binary segmentation is not consistent.

In the next chapter we describe the admissibility result and p-value monotonicity of commonly used MTPs for the all-pairwise testing problem. Chapter 3 provides cluster interpretability and consistency of PADD and PADD+. The consistency of MRD and BSP for the change point problem is offered in Chapter 4. Chapter 5 gives
the admissibility results of various MTPs for the matrix order problem. The simulation comparison of the modified MRD and the step-down procedure is also mentioned.
Chapter 2

Admissibility and monotonicity of all-pairwise testing procedures

The all-pairwise testing problem is a multiple testing problem, which consists of \( L(= k(k - 1)/2) \) pairwise differences among \( k \) populations. In this chapter, we intend to examine the admissibility and p-value monotonicity of some commonly used all-pairwise testing MTPs, including single-step and stepwise procedures.

Cohen, Sackrowitz and Chen (2010) offered admissible procedures for the all-pairwise testing problem, namely PADD and PADD+. They also proved that conventional step-down procedures are inadmissible. Following the framework of Cohen, Sackrowitz and Chen (2010), we investigate the admissibility property of other all-pairwise testing procedures for a balanced fixed effects one-way analysis of variance model. The MTPs evaluated here include Newman-Keuls multiple range test, general step-down procedure and closed testing method. In addition to admissibility results, a graphical interpretation of admissibility and p-value monotonicity are also shown in this chapter.

The structure of this chapter is as follows. In section 1, the all-pairwise testing problem for a balanced fixed effects one-way analysis of variance model is briefly introduced. Sections 2 and 3 mention single-step and stepwise MTPs and their admissibility results. Section 4 gives the graphical interpretation of admissibility. Section 5 shows the p-value monotonicity property.

2.1 Notations and preliminaries

A balanced fixed effects one-way analysis of variance model is described below. Let \( X_{ij} \), \( i = 1, \cdots, k \) and \( j = 1, \cdots, n \) be independent, normally distributed random variables with mean \( \mu_i \) and common known variance \( \sigma^2 \). Without loss of generality, we assume...
\( n = 1 \) and suppress \( j \) in \( X_{ij} \). Later, this assumption will be relaxed for either unknown \( \sigma^2 \) or \( n > 1 \). Denote a sample point \( x = (x_1, \cdots, x_k) \). The all-pairwise testing problem in this model is meant to test \( L \) individual tests. Namely, \( H_{ij'} : \mu_i = \mu_{j'} \text{ vs. } K_{ij'} : \mu_i \neq \mu_{j'} \), for \( i' \neq j' \).

A vector risk function assigns each individual test a risk function. It possesses minimal restrictions on risk functions and is fundamental to many other risk functions. If a test procedure is inadmissible for a vector risk function, it is also inadmissible for any monotone linear combination of individual risk functions.

Suppose \( W \) has an exponential family distribution with natural parameter \( \nu \). Test \( H_1 : \nu_i = 0 \). Matthes and Truax (1967) provided a necessary and sufficient condition for an individual test of \( H_1 \), \( \phi(w) = \phi(w_1, \cdots, w_L) \) to be admissible: for almost every fixed \( w_2, \cdots, w_L \), the acceptance section of any specific test \( \phi(w) \) is convex in \( w_1 \). For the all-pairwise testing problem, due to the symmetry of tests and the assumption of vector risk function, we will focus on \( H_{12} : \mu_1 = \mu_2 \text{ vs. } K_{12} : \mu_1 \neq \mu_2 \), unless stated otherwise. The convexity property allows us to prove that a test is admissible, if and only if there exist no three sample points, \( \bar{x} \), \( \bar{x}^* \), and \( \bar{x}^{**} \), where \( \bar{x}^* = \bar{x} + r_1g \), \( \bar{x}^{**} = \bar{x} + r_2g \), \( 0 < r_1 < r_2 \), \( g = (-1, 1, 0, \cdots, 0) \), such that \( \bar{x}^* \) is a reject point, while \( \bar{x} \) and \( \bar{x}^{**} \) are accept points for testing \( H_{12} \). See Cohen, Sackrowitz and Chen (2010) for more details.

### 2.2 Single-step procedures

In this section, we consider three single-step procedures for the all-pairwise testing problem. They are Tukey, Fisher and Scheffé methods.

Tukey’s (1953) method provides simultaneous confidence intervals for all-pairwise differences. The procedure rejects \( \mu_i = \mu_j \) if \( |x_i - x_j| > q\sqrt{2} \) for all \( i < j \), and \( q \) satisfies

\[
\Pr \left( \frac{|x_i - x_j|}{\sqrt{2}} \leq q, \text{ for all } i < j \right) = 1 - \alpha.
\]

Fisher’s unprotected least significant difference method (Fisher (1935)) rejects \( \mu_i = \mu_j \) if \( |x_i - x_j| > \sqrt{2}Z_{\alpha/2} \), where \( Z_{\alpha/2} \) is the \( (1 - \alpha/2) \) quantile of a standard normal.

Scheffé (1953) uses a set of \( 100(1-\alpha)\% \) simultaneous confidence intervals for pairwise difference \( \mu_i - \mu_j \). Further \( \mu_i = \mu_j \) is rejected if \( |x_i - x_j| > \sqrt{(k-1)\chi^2_{k-1,\alpha}/2} \) for all
i < j, where \( \chi^2_{k-1, \alpha} \) is the \((1 - \alpha)\) quantile of an \( \chi^2 \) distribution with degrees of freedom \( k - 1 \).

It is observed that all three single step procedures use the same statistics, \(|x_i - x_j|\), but different critical values. Before we move on to the admissibility proof of single-step procedures, we start with the following lemma.

**Lemma 2.2.1.** Let \( \overline{x} = (x_1, \cdots, x_k), \overline{x}^* = \overline{x} + r_1 \overline{g}, \overline{x}^{**} = \overline{x} + r_2 \overline{g}, 0 < r_1 < r_2 \) and \( \overline{g} = (-1, 1, 0, \cdots, 0) \). Then,

\[
\max(|x_1 - x_2|, |x_1^* - x_2^*|, |x_1^{**} - x_2^{**}|) = \max(|x_1 - x_2|, |x_1^* - x_2^*|).
\]

**Proof.** Consider a function \( f(t) = |t| \). Since \( f(t) \) is a convex function, \((1 - a)f(t) + af(t - 2r_2) \geq f(t - 2ar_2)\) for \( 0 < a < 1 \). Now let \( a = r_2/r_1 \) and \( x_1 - x_2 = t \). We obtain

\[
(1-r_2/r_1)f(x_1-x_2) + r_2/r_1 f(x_1-x_2-2r_2) \geq f(x_1-x_2-2r_1).
\]

Thus, \( f(x_1-x_2-2r_1) \) can’t be larger than \( f(x_1-x_2) \) and \( f(x_1-x_2-2r_2) \) at the same time. That is,

\[
\max(|x_1 - x_2|, |x_1^* - x_2^*|, |x_1^{**} - x_2^{**}|) = \max(|x_1 - x_2|, |x_1^* - x_2^*|).
\]

\( \square \)

**Theorem 2.2.2.** Tukey, Fisher’s unprotected least significance and Scheffé’s methods for the all-pairwise testing problem for a vector risk function are admissible.

**Proof.** We prove that there exist no three sample points, \( \overline{x}, \overline{x}^*, \) and \( \overline{x}^{**} \), where \( \overline{x}^* = \overline{x} + r_1 \overline{g}, \overline{x}^{**} = \overline{x} + r_2 \overline{g} \), \( 0 < r_1 < r_2 \), \( \overline{g} = (-1, 1, 0, \cdots, 0) \), such that \( \overline{x}^* \) is a reject point, while \( \overline{x} \) and \( \overline{x}^{**} \) are accept points for testing \( H_{12} \).

For \( \overline{x} \) and \( \overline{x}^{**} \) to be accept points for \( \mu_1 = \mu_2 \), it implies that \(|x_1 - x_2| \leq \sqrt{2}C \) and \(|x_1^{**} - x_2^{**}| \leq \sqrt{2}C \). Here \( C \) is some critical value. Further more, for \( \overline{x}^* \) to be a reject point for \( \mu_1 = \mu_2 \), it requires \(|x_1^* - x_2^*| > \sqrt{2}C \). Thus,

\[
|x_1 - x_2| < |x_1^* - x_2^*| > |x_1^{**} - x_2^{**}|.
\]  \( (2.1) \)

(2.1) implies \(|x_1^* - x_2^*|\) is larger than \(|x_1 - x_2|\) and \(|x_1^{**} - x_2^{**}|\). However, Lemma 2.2.1 shows \(|x_1^* - x_2^*| \) can’t be greater than \(|x_1 - x_2|\) and \(|x_1^{**} - x_2^{**}|\) at the same time. Thus, there can’t exist three sample points \( \overline{x}, \overline{x}^*, \) and \( \overline{x}^{**} \), such that \( \overline{x}^* \) is a reject point, while
and $\bar{z}^{**}$ are accept points for $\mu_1 = \mu_2$. Therefore, all three single-step procedures are admissible.

2.3 Stepwise procedures

In this section, we assess the admissibility of some stepwise procedures for the all-pairwise testing problem, including multiple range test, multiple $\chi^2$ test, closed testing procedure and Fisher’s protected least significant distance (LSD) method.

2.3.1 Multiple range test

Several versions of multiple range tests have been attributed to Keuls (1952), Duncan (1955), Lehmann and Shaffer (1977) and many others. The main difference among various multiple range tests is the choices of critical values. Specifically, the significance level ($\alpha_m$) used at different stages, where

$$\alpha_m = P_{\mu_1=\ldots=\mu_m} \left( \max_{1 \leq i,j \leq m} \frac{x_i - x_j}{\sqrt{2}} > C_m \right), m = 2, \ldots, k.$$  

Here $C_m$ is the critical value for the range statistics encompassing $m$ samples.

We list different versions of the multiple range tests below for reference.

Newman-Keuls $\alpha_m = \alpha, \forall m = 2, \ldots, k$

Duncan $\alpha_m = 1 - (1 - \alpha)^{m-1}, \forall m = 2, \ldots, k$

Einot and Gabriel $\alpha_m = 1 - (1 - \alpha)^{m/k}, \forall m = 2, \ldots, k$

Ryan $\alpha_m = \begin{cases} (m/k)\alpha & m = 2, \ldots, k - 2 \\ \alpha & m = k - 1, k \end{cases}$

Welsch $\alpha_m = \begin{cases} 1 - (1 - \alpha)^{m/k} & m = 2, \ldots, k - 2 \\ \alpha & m = k - 1, k \end{cases}$

In general, the multiple range tests proceed as follows. First, order sample means $x_i$'s as

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(k)}.$$  

Let $\mu_{(i)}$ be the population mean ordered by the values of their associated sample means with corresponding means. At the first step, the range statistic $D_k = x_{(k)} - x_{(1)}$
is computed and compared with $\sqrt{2}C_k$. If $D_k \leq \sqrt{2}C_k$, then stop and accept all $L = k(k - 1)/2$ hypotheses. Otherwise, reject $\mu_{(1)} = \mu_{(k)}$ and go to the next step.

At the second step, compute two range statistics, $D_{k-1}^{(1)} = x_{(k)} - x_{(2)}$ and $D_{k-1}^{(2)} = x_{(k-1)} - x_{(1)}$, and compare both with $\sqrt{2}C_{k-1}$. If both ranges are less than or equal to $\sqrt{2}C_{k-1}$, then stop and accept $\mu_{(2)} = \mu_{(k)}$, $\mu_{(1)} = \mu_{(k-1)}$ and the remaining hypotheses. Otherwise, if $D_{k-1}^{(1)} > \sqrt{2}C_{k-1}$, then reject $\mu_{(2)} = \mu_{(k)}$. Likewise, reject $\mu_{(1)} = \mu_{(k-1)}$ if $D_{k-1}^{(2)} > \sqrt{2}C_{k-1}$.

For the following steps, once a range statistic is shown to be less than or equal to a specific critical value, its subranges are no longer tested.

In general, $s$ or less range statistics are computed and compared with $\sqrt{2}C_{k-s+1}$ at the $s$th step. Namely, $D_{k-s+1}^{(1)} = x_{(k)} - x_{(s)}$, $D_{k-s+1}^{(2)} = x_{(k-1)} - x_{(s-1)}$, \ldots, $D_{k-s+1}^{(s)} = x_{(k-s+1)} - x_{(1)}$. The total number of range statistics calculated at each step depends on the result of the previous steps.

The multiple range testing steps described above can be summarized as rejecting the hypothesis $H_{(ij)} : \mu_{(i)} = \mu_{(j)}$ if

$$x_{(j')} - x_{(i')} > \sqrt{2}C_{j' - i' + 1}, \forall j', i' \leq i.$$ 

In other words, $\mu_{(i)} = \mu_{(j)}$ is rejected if and only if $H_{(ik)}$ and all subranges covering $x_{(i)}$ and $x_{(j)}$ have been rejected.

**Theorem 2.3.1.** The multiple range test is admissible for the all-pairwise testing problem for a vector risk function.

**Proof.** Let $\bar{x}$, $\bar{x}^*$, and $\bar{x}^{**}$ be three sample points, where $\bar{x}^* = \bar{x} + r_1\bar{g}$, $\bar{x}^{**} = \bar{x} + r_2\bar{g}$, $0 < r_1 < r_2$, $\bar{g} = (-1, 1, 0, \ldots, 0)$, such that $\bar{x}^*$ is a reject point, while $\bar{x}$ and $\bar{x}^{**}$ are accept points for $\mu_1 = \mu_2$.

In the case of $x^*_1 < x^*_2$, assume the ranks of $x^*_1$ and $x^*_2$ among $\bar{x}^* = (x^*_1, \ldots, x^*_k)$ are $i$ and $j$, respectively. For $\bar{x}^*$ to be a reject point for $\mu_1 = \mu_2$, it requires that,

$$x^*_{(j')} - x^*_{(i')} > \sqrt{2}C_{j' - i' + 1}, \forall j', i' \leq i.$$

Now let the ranks of $x^*_1$ and $x^*_2$ be $u$ and $v$ among $\bar{x}^{**} = (x^{**}_1, \ldots, x^{**}_k)$. Since $r_2 > r_1$ and $x^*_1 < x^*_2$, it can be concluded that $x^{**}_1 < x^*_1$ and $x^{**}_2 > x^*_2$. Because $x^{**}_1 = x^*_1$ for
\[ i = 3, \ldots, k, \] it must be true that \( u \leq i \) and \( v \geq j \). Compare \( |x_{(v)}^* - x_{(u)}^*| \) and \( \sqrt{2C_{v-u+1}} \) for four cases in the following.

**case 1:** \( u = i, v = j \)

\[
|x_{(v)}^* - x_{(u)}^* - x_{(i)}^*| > \sqrt{2C_{v-u+1}} = \sqrt{2C_{v-u+1}}
\]

**case 2:** \( u = i, v > j \)

\[
|x_{(v)}^* - x_{(u)}^* - x_{(i)}^*| > \sqrt{2C_{v-u+1}}
\]

**case 3:** \( u < i, v = j \)

\[
|x_{(v)}^* - x_{(u)}^* - x_{(i)}^*| > \sqrt{2C_{v-u+1}}
\]

**case 4:** \( u < i, v > j \)

\[
|x_{(v)}^* - x_{(u)}^* - x_{(i)}^*| > \sqrt{2C_{v-u+1}}
\]

Thus, we see that, when \( x_1^* < x_2^* \), \( x_{(v)}^* - x_{(i)}^* \) > \( \sqrt{2C_{v-u+1}} \), \( v \geq j, v' \leq u \). It implies \( x_{(v)}^* \) can’t be an accept point for \( \mu_1 = \mu_2 \). By symmetry, it can be shown that in the case when \( x_1^* > x_2^* \), \( x \) can’t be an accept point for \( \mu_1 = \mu_2 \). In other words, \( x \) and \( x_{(v)}^* \) can’t be accept points while \( x^+ \) is a reject point for \( \mu_1 = \mu_2 \) at the same time. Thus, the multiple range tests are admissible for the all-pairwise testing problem for a vector risk function.
Remark 2.3.1. The proof of admissibility for the multiple range tests does not depend on the versions of the multiple range tests, thus it is a proof for general multiple range tests.

2.3.2 Multiple $\chi^2$ test

Multiple $\chi^2$ test is a special case of a general step-down testing scheme. It applies $\chi^2$ statistics to the step-down testing scheme and sets level of significance of $m$ samples, $\alpha_m$, as

$$\alpha_m = \begin{cases} 1 - (1 - \alpha)^{m/k} & m = 1, \ldots, k - 2 \\ \alpha & m = k - 1, k \end{cases}$$

A general step-down testing scheme proceeds as follows. At the first step, a global hypothesis $H_{12\ldots k} : \mu_1 = \cdots = \mu_k$ vs. $K_{12\ldots k}$ : not all $\mu_i$ are equal, is conducted at level $\alpha_k$. If $H_{12\ldots k}$ is accepted, then accept all $L = k(k - 1)/2$ individual hypotheses. Otherwise, move on to the next step.

At the second step, test all subsets of $\{1, 2, \cdots, k\}$ which contain $k - 1$ elements at level $\alpha_{k-1}$. For instance, $H_{23\ldots k} : \mu_2 = \mu_3 = \cdots = \mu_k$, $H_{13\ldots k} : \mu_1 = \mu_3 = \cdots = \mu_k$, and $H_{12\ldots k-1} : \mu_1 = \mu_2 = \cdots = \mu_{k-1}$, etc. If one hypothesis, say $H_I$ is accepted, then accept all hypotheses $H_J$, for all $J \subset I$ and there is no need to test those hypotheses. Continue the process until all current subgroups have been accepted.

Finally, the general step-down testing scheme rejects $\mu_i = \mu_j$ if all $H_U$’s, such that $\{i, j\} \in U$ are rejected. Take the case when $k = 4$ as an example. At the first step, $H_{1234}$ is tested at level $\alpha_4 = \alpha$. If $H_{1234}$ is accepted, then accept all 6 pair difference hypotheses. Otherwise, move on to test $H_{123}$, $H_{124}$, $H_{134}$ and $H_{234}$ at level $\alpha_3 = \alpha$. If $H_{123}$ is accepted, then $H_{12}$, $H_{13}$ and $H_{23}$ are also accepted. The process continues until all remaining tests are accepted. Thus, the general step-down procedure rejects $\mu_1 = \mu_2$ if $H_{1234}$, $H_{123}$, $H_{124}$ and $H_{12}$ are all rejected at levels $\alpha_4$, $\alpha_3$, $\alpha_3$ and $\alpha_2$, respectively.

Denote the $\chi^2$ statistics used for $H_U : \mu_i$ are equal, $\forall i \in U$ as $\chi_U(x)$. That is,

$$\chi_U(x) = \frac{1}{|U| - 1} \sum_{i \in U} \left( x_i - \frac{1}{|U|} \sum_{i \in U} x_i \right)^2.$$  \hspace{1cm} (2.2)

Here $|U|$ represents the number of elements in $U$. Thus, the multiple $\chi^2$ test rejects
\( \mu_i = \mu_j \) at \( \bar{x} \) if
\[
\chi_U(\bar{x}) > \chi_{|U|-1,\alpha|U|}^2, \forall U, \text{ such that } \{i, j\} \in U.
\]

Having the basic knowledge of the multiple \( \chi^2 \) test, we turn our attention to the admissibility of the multiple \( \chi^2 \) test.

**Theorem 2.3.2.** The multiple \( \chi^2 \) test is admissible for the all-pairwise testing problem for a vector risk function.

**Proof.** Let \( \bar{x}, \bar{x}^*, \) and \( \bar{x}^{**} \) be three sample points, where \( \bar{x}^* = \bar{x} + r_1g, \bar{x}^{**} = \bar{x} + r_2g, \)
\( 0 < r_1 < r_2, g = (-1, 1, 0, \cdots, 0), \) such that \( \bar{x}^* \) is a reject point, while \( \bar{x} \) and \( \bar{x}^{**} \) are accept points for \( \mu_1 = \mu_2. \)

Assume \( \bar{x}^* \) to be a reject point for \( \mu_1 = \mu_2. \) It implies,
\[
\chi_U(\bar{x}^*) > \chi_{|U|-1,\alpha|U|}^2, \forall U, \text{ such that } \{1, 2\} \in U. \tag{2.3}
\]

For \( \bar{x} \) and \( \bar{x}^{**} \) be accept points for \( \mu_1 = \mu_2, \) there exist \( U_1 \) and \( U_2, \) such that \( U_1 \) and \( U_2 \) contain \( \{1, 2\} \) and,
\[
\chi_{U_1}(\bar{x}) \leq \chi_{|U_1|-1,\alpha|U_1|}^2 \tag{2.4}
\]
\[
\chi_{U_2}(\bar{x}^{**}) \leq \chi_{|U_2|-1,\alpha|U_2|}^2 \tag{2.5}
\]

Note that \( \sum_{i \in U} x_i = \sum_{i \in U} x_i^* = \sum_{i \in U} x_i^{**}, \) for all \( U \ni \{1, 2\}. \) Combining the result of (2.2), (2.3) and (2.4), we obtain \( \chi_{U_1}(\bar{x}^*) > \chi_{U_1}(\bar{x}). \) Thus,
\[
\sum_{i \in U_1} \left( x_i^* - \sum_{i \in U_1} x_i^*/|U_1| \right)^2 > \sum_{i \in U_1} \left( x_i - \sum_{i \in U_1} x_i/|U_1| \right)^2
\]
\[
\Rightarrow (x_1 - r_1 - \sum_{i \in U_1} x_i/|U_1|)^2 + (x_2 + r_1 - \sum_{i \in U_1} x_i/|U_1|)^2
\]
\[
> (x_1 - \sum_{i \in U_1} x_i/|U_1|)^2 + (x_2 - \sum_{i \in U_1} x_i/|U_1|)^2
\]
\[
\Rightarrow 2r_1(x_2 - \sum_{i \in U_1} x_i/|U_1|) - 2r_1(x_1 - \sum_{i \in U_1} x_i/|U_1|) + 2r_1^2 > 0
\]
\[
\Rightarrow x_2 - x_1 + r_1 > 0 \tag{2.6}
\]
Meanwhile, it is derived from (2.2), (2.3) and (2.5) that,
\[
\sum_{i \in U_2} (x_i^* - \sum_{i \in U_2} x_i/|U_2|)^2 > \sum_{i \in U_2} (x_i^{**} - \sum_{i \in U_2} x_i/|U_2|)^2
\]
\[
\Rightarrow (x_1 - r_1 - \sum_{i \in U_2} x_i) + (x_2 + r_1) > (x_1 - r_2 - \sum_{i \in U_2} x_i^{**}) + (x_2 + r_2 - \sum_{i \in U_2} x_i)
\]
\[
\Rightarrow 2r_1(x_2 - \sum_{i \in U_2} x_i) - 2r_1(x_1 - \sum_{i \in U_2} x_i) + 2r_2^2
\]
\[
> 2r_2(x_2 - \sum_{i \in U_2} x_i - \sum_{i \in U_2} x_i) + 2r_2^2
\]
\[
\Rightarrow (r_2 - r_1)(x_2 - x_1) + r_2^2 - r_1^2 < 0
\]
\[
\Rightarrow (r_2 - r_1)(x_2 - x_1 + r_1 + r_2) < 0
\]
\[
\Rightarrow x_2 - x_1 + r_1 + r_2 < 0 \quad (2.7)
\]

Obviously, (2.6) and (2.7) are contradictory. Thus, \( \bar{x} \) and \( \bar{x}^{**} \) can’t be accept points while \( \bar{x}^* \) is a reject point for \( \mu_1 = \mu_2 \) at the same time. Consequently, the multiple \( \chi^2 \) test is admissible for the all-pairwise testing problem for a vector risk function. \( \square \)

### 2.3.3 Closed testing method

The closure method proposed by Marcus, Peritz and Gabriel (1976) provides a way to construct tests which control strong family wise error rate (FWER). The closure for a finite family of hypotheses \( \{H_i, i = 1, \cdots, L\} \) is a collection of all non-empty intersection \( H_c = \cap_{i \in c} H_i \) for all \( c \subseteq \{1, 2, \cdots, L\} \). Assume a level \( \alpha \) test exists for each \( H_c \). The closed testing method rejects individual hypothesis \( H_i \) if and only if every \( H_c \) with \( c \supseteq i \) is rejected by its corresponding level \( \alpha \) test.

For the all-pairwise testing problem, it can be shown that, when there are \( k = 3 \) populations, the closed testing method is almost equivalent to a general step-down procedure, except for the choices of significance levels.

When constructing a closed testing method, one often encounters a non-empty intersection of test, say \( H_c = \cap_{i \in c} H_i \), which is composed of several subhypotheses. In this situation, Peritz (1970) suggested testing each subhypothesis at level \( 1 - (1 - \alpha)^{1/m} \).
for \( m \) subhypotheses. Another method proposed Royen (1989) rejects the hypothesis based on the most significant individual test. For instance, consider the hypothesis \( H_{12,34} : \mu_1 = \mu_2, \mu_3 = \mu_4 \). It contains two subhypotheses: \( H_{12} : \mu_1 = \mu_2 \) and \( H_{34} : \mu_3 = \mu_4 \). Peritz’s suggestion would test each hypothesis at level \( 1 - (1 - \alpha)^{1/2} \). On the other hand, Royen’s suggestion is to use level \( \alpha \) for each hypothesis. And \( H_{12,34} \) is rejected if at least one of the hypotheses is rejected. Therefore, in general the hypothesis containing two or more subhypotheses is rejected if at least one subhypothesis is rejected at level \( 1 - (1 - \alpha)^{1/m} \) for \( m \) subhypotheses with Peritz’s suggestion. While the same hypothesis is rejected if at least one subhypothesis is rejected at level \( \alpha \) with Royen’s suggestion.

It is useful to keep in mind that the closure of all-pairwise tests is formed by a collection of possible partitions and subpartitions of \( \mu = \{\mu_1, \mu_2, \ldots, \mu_k\} \). For example, with \( k = 4 \) populations, the closure is formed by,

\[
\begin{align*}
H_{1234} : \mu_1 = \mu_2 = \mu_3 = \mu_4 & \quad H_{12,34} : \mu_1 = \mu_2, \mu_3 = \mu_4 \\
H_{123} : \mu_1 = \mu_2 = \mu_3 & \quad H_{13,24} : \mu_1 = \mu_3, \mu_2 = \mu_4 \\
H_{124} : \mu_1 = \mu_2 = \mu_4 & \quad H_{14,23} : \mu_1 = \mu_4, \mu_2 = \mu_3 \\
H_{134} : \mu_1 = \mu_3 = \mu_4 & \quad H_{23} : \mu_2 = \mu_3 \\
H_{234} : \mu_2 = \mu_3 = \mu_4 & \quad H_{24} : \mu_2 = \mu_4 \\
H_{34} : \mu_3 = \mu_4 &
\end{align*}
\]

In this example, the closed testing method rejects \( \mu_1 = \mu_2 \) if the tests that include \( \mu_1 = \mu_2 \), i.e., \( H_{1234}, H_{123}, H_{124}, H_{12,34} \) and \( H_{12} \) are all rejected.

**Remark 2.3.2.** In general, those hypotheses relevant to the test \( \mu_1 = \mu_2 \) are classified into two categories. One category, say category A, contains a single hypothesis. For example, \( H_{1234}, H_{124}, H_{12} \), etc. The other category, say category B, consists of hypotheses that have two or more subhypotheses, such as \( H_{12,34} \).

**Remark 2.3.3.** When the closed testing method is applied to the all-pairwise testing problem, the relevant samples for a specific individual test are always grouped in the same test statistics. And this is true for hypotheses in the category A and B. For example, consider the hypothesis \( \mu_1 = \mu_2 \) for \( k = 4 \) populations. The hypotheses relating
to test \( \mu_1 = \mu_2 \) are \( H_{1234}, H_{123}, H_{124}, H_{12,34} \) and \( H_{12} \). The corresponding test statistics, either using \( \chi^2 \) or range statistics, involve both \( \bar{x}_1 \) and \( \bar{x}_2 \).

Note that, for a sample to be an accept point for \( \mu_i = \mu_j \), at least one of the relevant hypotheses is not rejected. If that one is from the category B, that means the most significant test statistics among all subhypotheses has p-value larger than \( 1 - (1 - \alpha)^{1/m} \) for Peritz's and \( \alpha \) for Royen’s suggestions.

With the properties of the closed testing method applying to all-pairwise testing problem, we may direct our attention to its admissibility result. In the following, two statistics, range statistics and \( \chi^2 \) statistics are discussed to show that the closed testing method is admissible for the all-pairwise testing problem.

Closed test using range statistics

Let \( s_i \) be a subset of \( \{1, 2, \ldots, k\} \) and \( s_i(\bar{x}) = \{x_j : j \in s_i\} \). Letting \( S = \{s_1, s_2, \ldots, s_s\} \) be a collection of such \( s_i \). Denote \( R_{s_i}(\bar{x}) \) be a range statistics involving \( s_i(\bar{x}) \) and \( p(R_{s_i}(\bar{x})) \) be the p-value of \( R_{s_i}(\bar{x}) \). Let \( P_R(S(\bar{x})) \) be the smallest p-value calculated among \( \{R_{s_1}(\bar{x}), R_{s_2}(\bar{x}), \ldots, R_{s_s}(\bar{x})\} \). Take \( S = \{s = \{1, 2, 3\}\} \) as an example.

\[
R_{s_i}(\bar{x}) = R(x_1, x_2, x_3) = \max(x_1, x_2, x_3) - \min(x_1, x_2, x_3)
\]

\[
P_R(S(\bar{x})) = p(R_{s_i}(\bar{x}))
\]

Another example is when \( S = \{s_1 = \{1, 2\}, s_2 = \{3, 4, 5\}\} \). Then,

\[
R_{s_1}(\bar{x}) = R(x_1, x_2) = \max(x_1, x_2) - \min(x_1, x_2)
\]

\[
R_{s_2}(\bar{x}) = R(x_3, x_4, x_5) = \max(x_3, x_4, x_5) - \min(x_3, x_4, x_5)
\]

\[
P_R(S(\bar{x})) = \min\{p(R_{s_1}(\bar{x})), p(R_{s_2}(\bar{x}))\}
\]

**Theorem 2.3.3.** The closed testing method using range statistics in the all-pairwise testing problem is admissible for a vector risk function.

**Proof.** Let the collection of index sets from the hypotheses relevant to \( \mu_1 = \mu_2 \) be \( O = \{O_A \cup O_B\} \). \( O_A \) and \( O_B \) are index sets involved in the hypotheses being classified into the category A and B, respectively. Let \( \bar{x}, \bar{x}^*, \) and \( \bar{x}^{**} \) be three sample points, where \( \bar{x}^* = \bar{x} + r_1 \bar{g}, \bar{x}^{**} = \bar{x} + r_2 \bar{g}, 0 < r_1 < r_2, \bar{g} = (-1, 1, 0, \ldots, 0) \), such that \( \bar{x}^* \) is a reject point, while \( \bar{x} \) and \( \bar{x}^{**} \) are accept points for \( \mu_1 = \mu_2 \).
For $\bar{x}^*$ to be a reject point for $\mu_1 = \mu_2$, it satisfies the following inequalities.

$$P_R(S(\bar{x}^*)) < \alpha, \text{ if } S \in O_A$$

$$P_R(S(\bar{x}^*)) < \begin{cases} \alpha, & \text{for Royen’s method and } S \in O_B \\ 1 - (1 - \alpha)^{1/m}, & \text{for Peritz’s method and } S \in O_B \end{cases}$$

For $\bar{x}$ to be an accept point for $\mu_1 = \mu_2$, there exists $S' \in O$, such that

$$P_R(S'(\bar{x})) > P_R(S'(\bar{x}^*)). \tag{2.8}$$

Note that $R_s(\bar{x}) = R_s(\bar{x}^*)$ if $s$ exclude $\{1, 2\}$. Remark 2.3.3 asserts that the index sets in $O_A$ must contain both $\{1, 2\}$. Thus (2.8) implies $R_{S'}(\bar{x}) < R_{S'}(\bar{x}^*)$ if $S' \in O_A$. In addition, if $S' \in O_B$, some $s_1 \in S'$ containing both $\{1, 2\}$ also yields to $R_{s_1}(\bar{x}) < R_{s_1}(\bar{x}^*)$. Then, one may go on from this to the conclusion that, for some subset of $\{1, 2, \ldots, k\}$ containing both $\{1, 2\}$, say $\hat{s}$, $R_{\hat{s}}(\bar{x}) < R_{\hat{s}}(\bar{x}^*)$. For that to be true, either $x_2^*$ is the maximum of $\hat{s}(\bar{x}^*)$, or $x_1^*$ is the minimum of $\hat{s}(\bar{x}^*)$ since $x_1 > x_1^*$ and $x_2 < x_2^*$. Thus, we see that $x_2^* > x_1^*$.

Likewise, for $\bar{x}^{**}$ to be an accept point for $\mu_1 = \mu_2$, one may conclude $R_{\hat{s}}(\bar{x}^{**}) < R_{\hat{s}}(\bar{x}^*)$. However, $x_2^* > x_1^*$ implies $x_2^{**} > x_1^{**}$. And that means $R_{\hat{s}}(\bar{x}^{**}) \geq R_{\hat{s}}(\bar{x}^*)$, a contradiction to the above observation. Thus, the closed testing method using range statistics is admissible for the all-pairwise testing problem for a vector risk function. \hfill \Box

**Closed test using $\chi^2$ statistics**

Denote the $\chi^2$ statistics as $\chi_l(\bar{x}) = \sum_{i \in l} (x_i - \bar{x}_l)^2 / (|l| - 1)$ and $\bar{x}_l = \sum_{i \in l} x_i / |l|$. Let $s_i$ be a subset of $\{1, 2, \ldots, k\}$ and $s_i(\bar{x}) = \{x_j : j \in s_i\}$. Letting $S = \{s_1, s_2, \ldots, s_s\}$ be a collection of such $s_i$. Denote $\chi_{s_i}(\bar{x})$ be a $\chi^2$ statistic involving $s_i(\bar{x})$ and $p(\chi_{s_i}(\bar{x}))$ be the p-value of $\chi_{s_i}(\bar{x})$. Let $P_\chi(S(\bar{x}))$ be the smallest p-value calculated among $\{\chi_{s_1}(\bar{x}), \chi_{s_2}(\bar{x}), \ldots, \chi_{s_s}(\bar{x})\}$.

**Theorem 2.3.4.** The closed testing method using $\chi^2$ statistics for the all-pairwise testing problem is admissible for a vector risk function.

**Proof.** Let the collection of index sets from the hypotheses relevant to $\mu_1 = \mu_2$ be $O = \{O_A \cup O_B\}$. $O_A$ and $O_B$ are index sets involved in the hypotheses being classified.
into the category A and B, respectively. Letting \( \bar{x}, \bar{x}^*, \) and \( \bar{x}^{**} \) be three sample points, where \( \bar{x}^* = \bar{x} + r_1 g, \bar{x}^{**} = \bar{x} + r_2 g, 0 < r_1 < r_2, g = (-1, 1, 0, \cdots, 0), \) such that \( \bar{x}^* \) is a reject point, while \( \bar{x} \) and \( \bar{x}^{**} \) are accept points for \( \mu_1 = \mu_2. \)

For \( \bar{x}^* \) to be a reject point for \( \mu_1 = \mu_2, \) it implies,

\[
P_\chi(S(\bar{x}^*)) < \alpha, \text{ if } S \in O_A
\]

\[
P_\chi(S(\bar{x}^*)) < \begin{cases} \alpha, & \text{for Royen’s method and } S \in O_B \\ 1 - (1 - \alpha)^{1/m}, & \text{for Peritz’s method and } S \in O_B \end{cases}
\]

Note that \( \chi_l(\bar{x}) = \chi_l(\bar{x}^*) \) if \( l \) exclude \( \{1, 2\}. \) It follows that \( \chi_{l_1}(\bar{x}) < \chi_{l_1}(\bar{x}^*) \) for some \( l_1 \in O \) and \( \{1, 2\} \in l_1, \) i.e.,

\[
\sum_{i \in l_1} (x_i - \bar{x}_{l_1})^2 < \sum_{i \in l_1} (x_i^* - \bar{x}_{l_1}^*)^2
\]

\[
(x_1 - \bar{x}_{l_1})^2 + (x_2 - \bar{x}_{l_1})^2 < (x_1 - r_1 - \bar{x}_{l_1})^2 + (x_2 + r_1 - \bar{x}_{l_1})^2
\]

\[
\Rightarrow x_2 - x_1 + r_1 > 0 \tag{2.9}
\]

where \( \bar{x}_{l_1}^* \) is the average of sample means in \( l_1. \) Similarly, assume for some \( l_2 \) containing \( \{1, 2\}, \) such that, \( \chi_{l_2}(\bar{x}^{**}) < \chi_{l_2}(\bar{x}^*). \) Further calculation yields to the following result, where \( \bar{x}_{l_2}^* \) is the average of sample means in \( l_2. \)

\[
(x_1^* - \bar{x}_{l_2}^*)^2 + (x_2^* - \bar{x}_{l_2}^*)^2 < (x_1^* - \bar{x}_{l_2})^2 + (x_2^* - \bar{x}_{l_2})^2
\]

\[
(x_1 - r_2 - \bar{x}_{l_1})^2 + (x_2 + r_2 - \bar{x}_{l_1})^2 < (x_1 - r_1 - \bar{x}_{l_1})^2 + (x_2 + r_1 - \bar{x}_{l_1})^2
\]

\[
(r_2 - r_1)(x_2 - x_1 + r_2 + r_1) < 0
\]

\[
\Rightarrow x_2 - x_1 + r_2 + r_1 < 0 \tag{2.10}
\]

Thus, the contradiction of (2.9) and (2.10) is sufficient to show that \( \bar{x} \) and \( \bar{x}^{**} \) can’t be accept points while \( \bar{x}^* \) is a reject point for \( \mu_1 = \mu_2 \) at the same time. Therefore, the closed testing method using \( \chi^2 \) statistics is admissible for the all-pairwise testing problem for a vector risk function.

\[\square\]

**Remark 2.3.4.** The proof for the admissibility of the closed testing method has nothing to do with Peritz’s or Royen’s suggestion for hypotheses containing two or more subhypotheses. Thus, either version of the closed testing procedures is admissible.
2.3.4 Fisher’s least significance method

Fisher’s protected least significance method (1935) uses two-step tests for the all-pairwise testing problem. At the first step, a size $\alpha \chi^2$ test is performed for a global hypothesis $H_{12\cdots k} : \mu_1 = \mu_2 = \cdots = \mu_k$ vs. $K : \text{not all } \mu_i \text{ are equal}$. If $H_{12\cdots k}$ is accepted, all pair comparisons, $\mu_i = \mu_j, i \neq j$, are accepted. Otherwise, LSD proceeds to the second step. At the second step, individual hypothesis $\mu_i = \mu_j$ is rejected if

$$|x_i - x_j| > \sqrt{2Z_{1-\alpha/2}}.$$ 

The following shows that the Fisher’s protected LSD is admissible.

**Theorem 2.3.5.** Fisher’s protected LSD is admissible for the all-pairwise testing problem for a vector risk function.

**Proof.** Let $\bar{x}$, $\bar{x}^*$, and $\bar{x}^{**}$ be three sample points, where $\bar{x}^* = \bar{x} + r_1g$, $\bar{x}^{**} = \bar{x} + r_2g$, $0 < r_1 < r_2$, $g = (-1, 1, 0, \cdots, 0)$, such that $\bar{x}^*$ is a reject point, while $\bar{x}$ and $\bar{x}^{**}$ are accept points for $\mu_1 = \mu_2$.

For $\bar{x}^*$ to be a reject point for $\mu_1 = \mu_2$, it satisfies

$$\sum_{i=1}^{k}(x_i^* - \bar{x}^*)^2 > (k - 1)\chi^2_{k-1,\alpha} \tag{2.11}$$

and

$$|x^*_1 - x^*_2| > \sqrt{2Z_{1-\alpha/2}} \tag{2.12}$$

On the other hand, for $\bar{x}$ to be an accept point for $\mu_1 = \mu_2$, at least one of the inequalities, (2.11) or (2.12) is reversed. That is,

$$\sum_{i=1}^{k}(x_i - \bar{x})^2 \leq (k - 1)\chi^2_{k-1,\alpha} \tag{2.13}$$

and

$$|x_1 - x_2| \leq \sqrt{2Z_{1-\alpha/2}} \tag{2.14}$$

If (2.13) is correct, it implies $\sum_{i=1}^{k}(x_i^* - \bar{x}^*)^2 > \sum_{i=1}^{k}(x_i - \bar{x})^2$, and further calculation leads to

$$x_2 - x_1 + r_1 > 0. \tag{2.15}$$

If (2.14) is correct, we obtain

$$|x^*_1 - x^*_2| > |x_1 - x_2|. \tag{2.16}$$
Moreover, for $\bar{x}^{**}$ to be an accept point for $\mu_1 = \mu_2$, at least one of the inequalities below is satisfied.

\[
\sum_{i=1}^{k}(x_i^{**} - \bar{x}^{**})^2 \leq (k - 1)\chi^2_{k-1,\alpha} \tag{2.17}
\]

\[
|x_1^{**} - x_2^{**}| \leq \sqrt{2}Z_{1-\alpha/2} \tag{2.18}
\]

It can be shown from further calculation that at least one of the inequalities below should be true.

\[
x_2 - x_1 + r_1 + r_2 > 0, \tag{2.19}
\]

\[
|x_1^* - x_2^*| > |x_1^{**} - x_2^{**}|. \tag{2.20}
\]

Now, we show that there does not exist a situation where at least one of the inequalities from (2.15) and (2.16), and at least one of the inequalities from (2.19) and (2.20) is true. Note that (2.16) and (2.20) are contradictory. (2.15) and (2.19) are not valid at the same time because that would mean $|x_1^* - x_2^*| > \max(|x_1^* - x_2^*|, |x_1^{**} - x_2^{**}|)$. From Lemma 2.2.1, it is not possible.

For (2.15) to be true, it is either $x_1 < x_2$ or $x_1 > x_2$ and $r_1 > x_1 - x_2$. Both conditions contradict (2.20). Meanwhile, for (2.19) to be true, either $x_2 - x_1 > 0$ or $x_2 - x_1 < 0$ and $r_1 > x_1 - x_2$ is true. The former and (2.19) are contradictory. The latter also contradicts (2.19). Thus, we see $\bar{x}$ and $\bar{x}^{**}$ can’t be reject points and $\bar{x}^*$ be an accept point at the same time. Therefore, Fisher’s protected LSD is admissible for the all-pairwise testing problem for a vector risk function.

\[\square\]

Remark 2.3.5. Note that, in the situation when common variance $\sigma^2$ is unknown, a pooled sample variance $s^2 = \sum_{ij}(x_{ij} - \bar{x})^2/(k(n-1))$ can replace $\sigma^2$. The proof of the admissibility for various all-pairwise testing problem is still valid, except the distribution forms need to adjust. Moreover, when sample size for each population, $n > 1$, the proof remains valid with the modification of the degrees of freedom. Thus, in the general case with unknown variance or $n > 1$, those commonly used procedures for the all-pairwise testing problem are admissible.
2.4 Graphical interpretation of admissibility

Consider MTPs when the number of populations \( k = 3 \). Figure 2.1–2.3 are cross sections of the 3-dimensional sample space when sample sum \( x_1 + x_2 + x_3 \) is fixed and \( \sigma^2 \) is known. On this cross section, a unit movement from \( (x_1, x_2, x_3) \) in the positive direction of the horizontal axis would result in the new vector \( (x_1 - 1/2, x_2 + 1/2, x_3) \). On the other hand, an upward unit movement to the positive direction of the vertical axis results in \( (x_1 - 1/2, x_2 - 1/2, x_3 + 1) \). The three letters marking on the cross section represent the action of acceptance (A) and rejection (R) for hypotheses \( H_{12} : \mu_1 = \mu_2 \), \( H_{23} : \mu_2 = \mu_3 \) and \( H_{13} : \mu_1 = \mu_3 \) in a specific region. For example, ARA means \( \mu_1 = \mu_2 \) and \( \mu_1 = \mu_3 \) are accepted while \( \mu_2 = \mu_3 \) is rejected.

Geometrically, the necessary and sufficient condition for a procedure to be admissible is that, the acceptance region along every line parallel to the horizontal axis is an interval. The thick broken lines on Figure 2.1–2.3 represent an example of lines parallel to the horizontal axis. The black part is for rejection region and the red one is for acceptance region. The thick line is used to illustrate the separation of the acceptance and rejection region for the test \( \mu_1 = \mu_2 \).

Figure 2.1–2.3 shows the acceptance and rejection region for the single-step procedure, the multiple range test and the multiple \( \chi^2 \) test, respectively. It is observed that, if we draw any line parallel to the horizontal axis cross this section, we won’t see the action for \( H_{12} \) to be acceptance, rejection and acceptance. Instead, the acceptance region is always convex. Note that the acceptance and rejection region for the closed testing method when \( k = 3 \) is the same as the one for the multiple \( \chi^2 \) test. So is the Fisher’s protected LSD method.

2.5 P-value monotonicity

In this section, we investigate the p-value monotonicity of some all-pairwise testing MTPs. An all-pairwise testing MTP is said to be p-value monotone, if and only if, an individual test \( \mu_i = \mu_j \) is rejected, all other hypotheses \( \mu_{i'} = \mu_{j'}, i' \neq j' \) with \( |x_{i'} - x_{j'}| \geq |x_i - x_j| \) are rejected.
Figure 2.1: Acceptance and rejection region for the single-step test

Figure 2.2: Acceptance and rejection region for the multiple $\chi^2$ test
Figure 2.3: Acceptance and rejection region for the multiple range test

The p-value monotonicity results for some all-pairwise testing MTPs are listed in the following.

**Theorem 2.5.1.** The multiple $\chi^2$ test is p-value monotone when $k = 3$ but not when $k \geq 4$.

**Proof.** For $k = 3$, let $\bar{x} = (x_1, x_2, x_3)$ be a reject point for $\mu_1 = \mu_2$. Let $C_2, C_3$ be critical values for the $\chi^2$ test. It follows that

$$\chi_{123}(\bar{x}) = \frac{1}{2} \sum_{i=1}^{3} (x_i - \bar{x}_{123})^2 > C_3,$$

$$\chi_{12}(\bar{x}) = (x_1 - \bar{x}_{12})^2 + (x_2 - \bar{x}_{12})^2 > C_2.$$

If $|x_2 - x_3| \geq |x_1 - x_2|$, then

$$\chi_{23}(\bar{x}) = (x_2 - \bar{x}_{23})^2 + (x_3 - \bar{x}_{23})^2$$

$$= \frac{1}{2} (x_2 - x_3)^2$$

$$\geq \frac{1}{2} (x_1 - x_2)^2$$

$$= D_{12} > C_2$$
Therefore $\mu_2 = \mu_3$ is also rejected. Likewise, it can be shown that $\mu_1 = \mu_3$ is rejected if $|x_1 - x_3| \geq |x_1 - x_2|$. Thus, when $k = 3$, the multiple $\chi^2$ test for the all-pairwise testing problem is p-value monotone.

For $k = 4$, let $x = (x_1, x_2, x_3, x_4) = (0, \sqrt{2C_2 + 2\epsilon}, \sqrt{2C_2 + 2\epsilon}, \sqrt{2C_2 + 2\epsilon}, \epsilon > 0$, $s > 2, s \epsilon = 3/2C_3 - 2C_2$ and $\chi_{1234}(x) > C_4$. Note that $\chi_{12}(x) = C_2 + \epsilon > C_2$, $\chi_{134}(x) = C_3$ and $\chi_{123}(x) = \chi_{124}(x) > C_3$. Thus, $\mu_1 = \mu_2$ is rejected and $\mu_1 = \mu_3$ is accepted. But, it can be shown that $|x_1 - x_3| > |x_1 - x_2|$. So, the multiple $\chi^2$ test for all-pairwise testing problem when $k = 4$ is not p-value monotone.

For $k > 4$, consider $x_1, x_2, x_3, x_4$ defined as the previous case and $x_5 = x_6 = \cdots = x_k = M$, where $M$ is some large number, such that $\chi_I(x) > C_{|I|}$ for all possible subset $I$ of $\{1, 2, \cdots, k\}$. Then the p-value monotone proof for $k = 4$ remains valid in the case of $k > 4$.

\begin{theorem}
The multiple range test for the all-pairwise testing problem is p-value monotone when $k = 3$ and not when $k \geq 4$.
\end{theorem}

\textit{Proof.} For $k = 3$, let $x = (x_1, x_2, x_3)$ be a reject point for $\mu_1 = \mu_2$. That implies $x_2 - x_1 > C_2$ and $x_3 - x_1 > C_3$, and $C_2$ and $C_3$ are some critical values of range statistics encompassing three and two random variables, respectively. If $|x_2 - x_3| > |x_1 - x_2|$, then $x_1 - x_3 > C_3$ and $x_3 - x_2 > x_2 - x_1 > C_2$ would result in rejecting $\mu_2 = \mu_3$. Thus, the multiple range test is p-value monotone when $k = 3$.

For $k \geq 4$, consider $x = (x_1, x_2, x_3, x_4) = (C_4 - 3C_2 - 2\epsilon, C_4 - 2C_2 - 2\epsilon, C_2, C_4 + \epsilon)$, where $\epsilon = 1/2(C_3 - 2C_2) > 0$ and $C_3 + C_2 + \epsilon > C_4$. Since $x_4 - x_1 = 3C_2 + 3\epsilon > C_4$, $x_4 - x_2 = C_3 + \epsilon$ and $x_4 - x_3 = C_2 + \epsilon$, $x$ is a reject point for $\mu_3 = \mu_4$.

Meanwhile, because $x_3 - x_1 = 2C_2 + 2\epsilon = C_3$, the multiple range test accepts $\mu_2 = \mu_3$ at $x$. But the fact that $x_3 - x_2 = C_2 + 2\epsilon > x_4 - x_3 = C_2 + \epsilon$ implies the multiple range test is not p-value monotone when $k = 4$.

Note that in the general case when $k > 4$, let $x = (x_1, x_2, \cdots, x_k)$, where $x_1$, $x_2$, $x_3$, $x_4$ are defined as in the case of $k = 4$ and $x_5, \cdots$ and $x_k$ are some large positive numbers. Then the proof of p-value monotone for the case of $k = 4$ is still valid.
for \( k > 4. \)

The proof of p-value monotone for the multiple \( \chi^2 \) test can be used to show that, the closed testing method is not p-value monotone for \( k > 4. \) On the other hand, single-step tests are p-value monotone, because single-step procedures use only one statistics and the larger the test statistics, the less the p-value. In addition, since the rejection of any individual hypothesis means the global hypothesis is rejected at the first step and the p-value of the statistics used at the second step is monotone with respect \( |x_i - x_j| \), it can be shown that Fisher’s protected LSD method is also p-value monotone.
Chapter 3

Cluster separation measure and consistency of PADD and PADD+

Cohen, Sackrowitz and Chen (2010) provided two admissible procedures, partitioned average difference down (PADD) and partitioned average difference down plus (PADD+), for the all-pairwise testing problem. This chapter is intended as an investigation of two other interesting properties of PADD and PADD+. Namely, cluster interpretability and consistency.

One practical issue for the all-pairwise testing MTPs is logical restrictions. Often-times, one test procedure would accept $\mu_1 = \mu_2$ and $\mu_2 = \mu_3$ while rejecting $\mu_1 = \mu_3$ at the same time. Although the result has its theoretical reasoning, it provokes the issue of interpretation. To further explore this matter, Shaffer (2008) used cluster formation pattern as a measure of interpretability to compare two all-pairwise testing MTPs, the original Benjamini-Hochberg and the Newman-Keuls procedures. In that simulation study, Shaffer showed that the Newman-Keuls performs better than the Benjamini-Hochberg procedure does in several parameter configurations. Thus, it is of interest to compare the cluster interpretability of PADD and PADD+ with the Newman-Keuls procedure.

Another interesting property of PADD and PADD+ that we explore in this chapter is consistency. Although there are different definitions regarding consistency, we define a testing procedure to be consistent if the probability of making any mistakes, either false rejection or false acceptance, tends to zero as sample sizes of each population tend to infinity.

Section 1 contains the description of PADD and PADD+. We outline the framework for analyzing cluster interpretability and offer the simulation results in section 2. Section
3 states the consistency results for PADD and PADD+.

### 3.1 PADD and PADD+ procedures

For a balanced fixed effects analysis of variance model, denote $X_{ij}, i = 1, \ldots, n$ and $j = 1, \ldots, k$ be independent normally distributed random variables with mean $\mu_i$ and common known variance $\sigma^2$. Without loss of generality, assume $\sigma^2 = 1$. For each ordered sample point $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_k)$, where $\bar{x}_1 \leq \bar{x}_2 \leq \cdots \leq \bar{x}_k$, and $\bar{x}_i$ is the sample mean. Let $S = \{1, \ldots, k\}$ and define

$$D_\bar{x}(A; B) = \frac{\sqrt{n}(\bar{x}_A - \bar{x}_{B \setminus A})}{\sqrt{\frac{1}{|A|} + \frac{1}{|B \setminus A|}}}$$

and $D_\bar{x}^*(B) = \max_{A \subseteq B} D_\bar{x}(A; B)$, $A \subset B \subseteq S$, $A \neq \emptyset$. Here $\bar{x}_A = \sum_{i \in A} \bar{x}_i / |A|$ and $|A|$ is the number of elements in $A$. Note that $D_\bar{x}^*(B)$ is the largest standardized difference between two subset means, and these two subsets form the set $B$. Let $V_\bar{x}(B)$ denote the set for which the maximum is attained, i.e., $D_\bar{x}^*(B) = D_\bar{x}(V_\bar{x}; B)$.

The PADD proceeds like a binary segmentation procedure. At the first step, all possible two-set partitions of $S$ are computed for $D_\bar{x}(A; S)$. Let $D_1 = D_\bar{x}^*(S)$ and compare $D_1$ with $C(S)$, a constant at step 1. If $D_1 \leq C(S)$, stop and accept all null hypotheses. Otherwise, partition $S$ into $V_\bar{x}(S)$ and $S \setminus V_\bar{x}(S)$ and continue to the next step.

At the second step, compute $D_\bar{x}(A; V_\bar{x}(S))$ and $D_\bar{x}(A; S \setminus V_\bar{x}(S))$ for all possible two-set partitions. Let $D_2^{(1)} = D_\bar{x}^*(V_\bar{x}(S))$ and $D_2^{(2)} = D_\bar{x}^*(S \setminus V_\bar{x}(S))$. Compare $D_2 = \max(D_2^{(1)}, D_2^{(2)})$ with $C(V_\bar{x}(S), S \setminus V_\bar{x}(S))$, a constant determined by $\{V_\bar{x}(S), S \setminus V_\bar{x}(S)\}$. If $D_2 > C(V_\bar{x}(S), S \setminus V_\bar{x}(S))$, then partition one of the sets into two sets. Otherwise, stop and accept the remaining hypotheses.

At the end of the $(m - 1)$st step, if the procedure does not stop, $S$ would have been partitioned into $m$ subsets, say $\{B_1, \ldots, B_m\}$. Then at the $m$th step, compute $D_m = \max_{1 \leq k \leq m} D_\bar{x}^*(B_k)$ and compare with $C(B_1, \ldots, B_m)$. If $D_m \leq C(B_1, \ldots, B_m)$, stop the procedure. Otherwise, select $B_i$ so that $D_m = D_\bar{x}^*(B_i)$ and partition $B_i$ into $V_\bar{x}(B_i)$ and $B_i \setminus V_\bar{x}(B_i)$.
Let \( \{B_1, B_2, \cdots, B_l\} \) be the final partition of \( S \). Then \( H_{ij} : \mu_i = \mu_j \) is accepted if \( i \) and \( j \) are in the same set of the partition. Otherwise, \( H_{ij} \) is rejected.

PADD+ is a hybrid combination of PADD and a Tukey-like screening procedure. That is, two additional constants, \( C_l \leq C_u \), are used to modify the test result from PADD. Let \( Y_{ij} \) be the standardized pair difference between \( \bar{x}_i \) and \( \bar{x}_j \). If \( Y_{ij} > C_u \) and \( i, j \) lie in the same set of final partition, then override the action from PADD and reject \( H_{ij} \). On the other hand, if \( Y_{ij} < C_l \) and \( i, j \) are in different sets of the final partition, then accept \( H_{ij} \). Note that in the case when \( C_l = 0 \) and \( C_u = \infty \), two procedures are equivalent.

### 3.2 Cluster separation measure

Define a cluster as a set of populations with equal parameter values. For any set of \( k \) ordered parameters, one observes at least one and at most \( k \) clusters. A good testing procedure would not only detect the difference between two clusters but also recognize the similarity among those in a single cluster. Moreover, a good testing procedure is expected to provide the action that is easy to interpret, or follows logical restrictions. Shaffer (2008) constructed a framework to measure the quality of all-pairwise testing MTPs in terms of cluster formation. Loosely speaking, compared with population cluster pattern, the higher proportion of correct sample clusters, the better.

#### 3.2.1 Definitions and notations

Before describing the simulation study for comparing the Newman-Keuls, PADD and PADD+ procedures, we briefly outline the framework and definition following Shaffer (2008). See Shaffer (2008) for a detailed description.

Define a p-cluster as a set of populations with equal means. Note that any populations outside that cluster have means either less or greater than the mean of those populations within that cluster. For instance, if \( x_1, x_3, \) and \( x_5 \) form a p-cluster, then it means \( \mu_1 = \mu_3 = \mu_5 = \mu \) and \( \mu_i \neq \mu \), for all \( i \neq 1, 3, 5 \).

To express any combination of population means in terms of p-cluster, p-pattern is
defined as a set of numbers used to illustrate population pattern by ordered population means. The clusters within a p-pattern are separated by square brackets. For example, \([1\,[234\,[56]]\) represents a population pattern such that, \(\mu_1 < \mu_2 = \mu_3 = \mu_4 < \mu_5 = \mu_6\).

For any all-pairwise testing MTP, the testing result can be translated into several overlapped clusters. Let an outcome cluster be a maximal proper subset of sample means which are significantly different from all other means. That is, if \(\{x_i : i \in A, A \subset (1, 2, \ldots, k)\}\) form an outcome cluster, then \(H_{ij} : \mu_i = \mu_j, i \in A, j \notin A\) are all rejected and there exists no \(A' \subset A\) satisfying the same condition. Moreover, vertical lines are used to separate outcome clusters. It should be noted that the samples within an outcome cluster does not mean that they have the same population means. For example, if \(k = 4\) and \(x_2, x_3\) and \(x_4\) form an outcome cluster, then \(H_{12} : \mu_1 = \mu_2, H_{13} : \mu_1 = \mu_3\) and \(H_{14} : \mu_1 = \mu_4\) are all rejected.

In analog to p-pattern, outcome pattern is defined as a set of numbers used to illustrate all-pairwise testing results. When some sample means are declared to show significantly no differences among them, single parentheses would be used to encompass them. And the order of parentheses also correspond to the order of sample means. In addition, vertical lines in the outcome pattern are used to separate outcome clusters. An example of the outcome pattern \((124)(123)|(68)(57)\) shows two outcome clusters. One is formed by \(x_1, x_2, x_3\) and \(x_4\), and the other is formed by \(x_5, x_6, x_7\) and \(x_8\). Thus, \(H_{ij} : \mu_i = \mu_j, i = 1, 2, 3, 4, j = 5, 6, 7, 8\) are all rejected. Moreover, \(H_{34} : \mu_3 = \mu_4\), \(H_{56} : \mu_5 = \mu_6\), \(H_{58} : \mu_5 = \mu_8\), \(H_{67} : \mu_6 = \mu_7\), \(H_{78} : \mu_7 = \mu_8\) are also rejected since they are not included in the same single parentheses. And all other hypotheses are accepted. Note that \((123)\) alone is not an outcome cluster since \(H_{14} : \mu_1 = \mu_4\) or \(H_{24} : \mu_2 = \mu_4\) are accepted.

An outcome cluster is a true outcome cluster if the components’ corresponding population means are formed by one or more adjacent p-clusters. In this respect, the sample means within a true outcome cluster have their population means larger than all population means below it and smaller than all population means above it. Take one example to illustrate. Consider a p-pattern, \([1\,[234\,[56]]\). An outcome pattern \((1234)|(56)\) forms two true outcome clusters, while the outcome pattern \((123)|(456)\)
forms zero true outcome clusters.

In contrast to true outcome cluster, an outcome cluster that is not a true outcome cluster is a false outcome cluster. The example mentioned before, (123)|(456), forms two false outcome clusters.

Of all true outcome clusters, a true outcome cluster which contains no type I or type III errors is a pure true outcome cluster. Here, the type I error means that not all sample means within a population cluster are included in an outcome cluster. That is to say, some true hypotheses are falsely rejected. The type III error refers to the directionally incorrect rejection. For example, consider the same P-pattern, [1][234][56]. An outcome pattern (1234)|(56) forms two pure true outcome clusters. On the other hand, an outcome pattern (124)(123)|(56) contains two true outcome clusters. But only the second one is a pure true outcome cluster. The first one is not pure due to directional rejection error for $H_{34} : \mu_3 = \mu_4$.

To measure the quality of an outcome pattern, false cluster rate (FCLR) is defined as the expected ratio of false outcome clusters to the max of outcome clusters or 1. It is an analogous idea growing out of the false discovery rate (FDR). That is,

$$FCLR = E\left(\frac{\text{false outcome cluster}}{\max(\text{outcome cluster}, 1)}\right)$$

### 3.2.2 Simulation results

With the above definitions in mind, we now lay out the simulation plan. Following Shaffer (2008), we use the same seven p-patterns for comparing three testing procedures, the Newman-Keuls, PADD and PADD+. The seven p-patterns are listed in Table 3.1. The first column represents pattern ID, and the second column shows the number of populations. The last column lists the p-pattern. Recall that, the square bracket separates p-cluster in the p-pattern.

In this simulation study, data are simulated from a balanced fixed effects analysis of variance model with $\sigma^2 = 1$. The difference between any two adjacent p-clusters is set to $(0, 1, \cdots, 10)$ for 11 cases. For instance, in the p-pattern [12][34][56], we set $\mu_1 = \mu_2 = \mu, \mu_3 = \mu_4 = \mu + m$ and $\mu_5 = \mu_6 = \mu + 2m$, where $m$ ranges from 0 to 10.
Table 3.1: Seven p-clusters used in the simulation study

<table>
<thead>
<tr>
<th>Pattern ID</th>
<th>Number of populations</th>
<th>P-pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>[1] [234]</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>[12] [34]</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>[12] [34] [56]</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>[1] [2345678]</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>[1] [234567] [8]</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>[12] [345678]</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>[1234] [5678]</td>
</tr>
</tbody>
</table>

For each set of simulation, the number of replicates is 1000. Note that the critical values for PADD and PADD+ are obtained from simulations and they are chosen so as to control the false discovery rate being less than 0.05. The simulated pure true outcome cluster rate and false cluster rate for seven p-patterns are illustrated in Figure 3.1–3.4.

Figure 3.1 and 3.2 show the simulation results for PADD+ and the Newman-Keuls procedures for the seven p-patterns. The outcomes vary from almost equivalent (pattern 1 and pattern 2) to the Newman-Keuls greatly dominates PADD+ for other patterns. The same phenomenon also appears on the comparison simulation study of the Benjamini-Hochberg and the Newman-Keuls in Shaffer (2008). Shaffer suggests the large number of same means may account for the phenomenon since the FWER for the Newman-Keuls is limited while that for the Benjamini-Hochberg increases. The same reason may apply to PADD+. The simulation results from Cohen, Sackrowitz and Chen (2010) showed that, PADD+ has higher FWER in the cases when the number of same means is large.

Figure 3.3 and 3.4 list the simulation outcomes for PADD and the Newman-Keuls procedure. Unlike the PADD+, we see that the pure true outcome cluster rate for PADD and the Newman-Keuls are very similar in pattern 4 and PADD has higher pure true outcome rate than the Newman-Keuls for most patterns. However, it is observed that, the FCRTs for PADD are generally higher than that for the Newman-Keuls in all seven patterns, especially for moderate separation cases. Recall that without screening, higher
critical values are necessary for PADD to control false discovery rate. Our observation is that, decreasing FWER results in the increasing of the pure true outcome cluster rate.

The simulation study shows that, the cluster formation patterns for the Newman-Keuls, PADD and PADD+ are almost perfectly correct when cluster separation is greater or equal to 8. Meanwhile, when the cluster separation is less than or equal to 4, all three procedures have pure true outcome rate less than 0.2 in all seven p-patterns. When the number of populations is small (patterns 1 and 2), PADD has higher pure true outcome rate than PADD+ and the Newman-Keuls procedures do while possessing low FCLR. In the case of moderate number of populations (pattern 3), although PADD still has higher pure true outcome rate than the other two procedures do, the FCLR for PADD elevates quite a bit. For the case of large number of populations (patterns 4-7), the Newman-Keuls has the highest pure true outcome rate in pattern 5 and PADD dominates the other two procedures for the other patterns.

3.3 Consistency results

Although there are some other definitions of consistency, we focus on the consistency in the case of sufficiently large sample sizes. Particularly, for a vector risk function containing $L = k(k - 1)/2$ elements, $\Psi = (\psi_{12}(x), \cdots, \psi_{k-1,k}(x))$, where $\psi_{ij}(x) = 1$ if $H_{ij}: \mu_i = \mu_j$ is rejected and $\psi_{ij}(x) = 0$ if $H_{ij}$ is accepted. A procedure is consistent if for each pair of $i, j$,

\[
\lim_{n \to \infty} P_{\mu_i \neq \mu_j} \left\{ \psi_{ij}(x) = 1 \right\} = 1 \\
\text{and} \\
\lim_{n \to \infty} P_{\mu_i = \mu_j} \left\{ \psi_{ij}(x) = 0 \right\} = 1
\]

Cohen, Sackrowitz and Chen (2009) showed that, MRD without screening for treatment vs. control problem is not a consistent procedure. However, MRD with screening and sign adjustment (MRDSS) is both admissible and consistent. Thus, it is of interest to examine the consistency of PADD and PADD+, the extension of MRD to the all-pairwise testing problem.
Figure 3.1: Simulated FCRT and pure true outcome cluster rate of PADD+ and the Newman-Keuls for patterns 1–4
Figure 3.2: Simulated FCRT and pure true outcome cluster rate PADD+ and the Newman-Keuls for patterns 5–7
Figure 3.3: Simulated FCRT and pure true outcome cluster rate of PADD and the Newman-Keuls for patterns 1–4
Figure 3.4: Simulated FCRT and pure true outcome cluster rate PADD and the Newman-Keuls for patterns 5–7
Consistency is often used to assess the quality of variable selection procedures. In some sense, MRD is comparable to a variable selection procedure. See Cohen, Sackrowitz and Chen (2009) for full description.

We prove that PADD and PADD+ are consistent and the consistency results can be easily extended to the case of unknown \( \sigma^2 \). We start with some lemmas that lead to consistency.

**Lemma 3.3.1.** Define three functions, \( f_1(q) \), \( f_2(q) \) and \( f_3(q) \) as follows.

\[
\begin{align*}
    f_1(q) &= \frac{n_2\mu_2 + n_1(1-q)\mu_1}{n_2 + n_1(1-q)} - \mu_1, \\
    f_2(q) &= \frac{\mu_2 - \frac{n_1\mu_1 + n_2q\mu_2}{n_1 + n_2q}}{\sqrt{\frac{1}{n_1 + n_2q} + \frac{1}{n_2(1-q)}}}, \\
    f_3(q) &= \frac{n_2\mu_2 + n_3(1-q)\mu_3 - n_1\mu_1 + n_3q\mu_3}{n_2 + n_3(1-q) + \frac{1}{n_1 + n_3q}}.
\end{align*}
\]

For \( 0 \leq q \leq 1, n_1, n_2, \) and \( n_3 \) positive numbers, and \( \mu_1 < \mu_3 < \mu_2 \),

1. \( f_1(q) \) and \( f_2(q) \) attain their maximum at \( q = 1 \) and \( q = 0 \), respectively.
2. \( f_3(q) \) has its maximum at either \( q = 1 \) or \( q = 0 \).

**Proof.** First, we simplify \( f_1(q) \) as the following.

\[
\begin{align*}
    f_1(q) &= \frac{n_2\mu_2 + n_1(1-q)\mu_1}{n_2 + n_1(1-q)} - \mu_1 \\
    &= \frac{n_1qm_2 + n_1n_3(1-q)\mu_1}{\sqrt{n_1q(n_2 + n_1(1-q))(n_1 + n_2)}} \\
    &= \frac{n_1n_2q(\mu_2 - \mu_1)}{\sqrt{n_1q(n_2 + n_1(1-q))(n_1 + n_2)}}.
\end{align*}
\]

Next, consider the logarithm transformation of \( f_1(q) \). Since \( f_1(q) \) is positive for \( \mu_2 > \mu_1 \), the point at which \( f_1 \) attains its maximum is the same as that of \( \log f_1 \). Write down \( \log f_1(q) \) as

\[
\begin{align*}
    \log f_1(q) &= \log(n_1n_2q(\mu_2 - \mu_1)) - \frac{1}{2}(\log(n_1q(n_2 + n_1(1-q))(n_1 + n_2))) \\
    &= \frac{1}{2}\log q - \frac{1}{2}\log(n_2 + n_1(1-q)) + \log(n_1n_2(\mu_2 - \mu_1)) - \frac{1}{2}\log n_1(n_1 + n_2).
\end{align*}
\]
The derivative of \( \log f_1(q) \) with respect to \( q \),

\[
\partial \log f_1/\partial q = 1/2(1/q + n_1/(n_2 + n_1(1-q))).
\]

Note that (3.1) is positive for \( 0 \leq q \leq 1 \), therefore \( \log f_1(q) \) has its maximum at \( q = 1 \), so does \( f_1(q) \).

Likewise, we also simplify \( f_2(q) \) and take the logarithm transformation,

\[
\log f_2(q) = \log(n_1n_2(1-q)(\mu_2 - \mu_1)) - \frac{1}{2} \left( \log(n_2(1-q)(n_1 + n_2(1-q))(n_1 + n_2)) \right)
\]

\[
= \frac{1}{2} \log(1-q) - \frac{1}{2} \log(n_1 + n_2q) + \log(n_1n_2(\mu_2 - \mu_1)) - \frac{1}{2} \log n_2(n_1 + n_2).
\]

Since \( \partial \log f_2(q)/\partial q = -1/2(1/(1-q) + n_2/(n_1 + n_2q)) < 0 \) for \( 0 \leq q \leq 1 \), therefore \( \log f_2(q) \) attains its maximum at \( q = 0 \), so does \( f_2(q) \).

Last, without loss of generality, consider a special case where \( \mu_3 = 0 \) and \( \mu_1 < 0 < \mu_2 \). In a general case with \( \mu_1 < \mu_3 < \mu_2 \), the following proof can be extended by redefining \( \mu_1' = \mu_1 - \mu_3, \mu_3' = 0 \) and \( \mu_2' = \mu_2 - \mu_3 \).

\( f_3(q) \) is simplified as

\[
f_3(q) = \frac{n_2}{n_2 + n_3(1-q)} + \frac{n_1}{n_1 + n_3q}.
\]

Since \( f_3(q) \geq 0 \), the point at which \( \log f_3(q) \) has its extrema is the same as that of \( f_3(q) \). And \( \log f_3(q) \) can be written as

\[
\log f_3(q) = \log(n_1n_2(\mu_1 + \mu_2) + n_3(qn_2\mu_2 + (1-q)n_1\mu_1))
\]

\[
- \frac{1}{2} \log(n_1 + n_3q)(n_2 + n_3(1-q)) - \frac{1}{2} \log(n_1 + n_2 + n_3)
\]

The function \( \log f_3 \) has its extremum at \( q = 0, 1 \) or \( \tilde{q} \), which satisfies the following equation.

\[
\frac{n_3(n_2\mu_2 - n_1\mu_1)}{n_1n_2(\mu_1 + \mu_2) + n_3(\tilde{q}n_2\mu_2 + (1-\tilde{q})n_1\mu_1)} = \frac{1}{2} \left( \frac{n_3}{n_1 + n_3\tilde{q}} - \frac{n_3}{n_2 + n_3(1-\tilde{q})} \right)
\]

The second derivative of \( \log f_3 \) evaluated at \( \tilde{q} \) is,

\[
\frac{n_3^2(n_2\mu_2 - n_1\mu_1)^2}{(n_1n_2(\mu_1 + \mu_2) + n_3(\tilde{q}n_2\mu_2 + (1-\tilde{q})n_1\mu_1))^2} + \frac{n_3^2}{2(n_1 + n_3\tilde{q})^2} + \frac{n_3^2}{2(n_2 + n_3(1-\tilde{q}))^2}.
\]

(3.2)
Note that (3.2) is always positive. Thus, the maximum of \( \log f_3(q) \) would be at \( q = 0 \) or \( q = 1 \).

**Lemma 3.3.2.** Let \( C(n) \) be a constant so that \( \lim_{n \to \infty} C(n) = 0 \) and \( C(n) = o(\sqrt{n}) \). For a balanced fixed effects one-way analysis of variance model, let \( S = \{1, 2, \ldots, k\} \) and \( X_{ij} \sim N(\mu_i, 1) \). If \( \mu_1 = \mu_2 = \cdots = \mu_k \), then \( \lim_{n \to \infty} P_{\mu}(D_x^*(S) > C(n)) = 0 \).

**Proof.**

\[
\lim_{n \to \infty} P_{\mu}(D_x^*(S) > C(n)) = 1 - \lim_{n \to \infty} P_{\mu}(D_x(A; S) \leq C(n), \forall A \subset S, A \neq \emptyset)
\leq (k - 1) - \lim_{n \to \infty} \sum_{A \subset S} P_{\mu}(D_x(A; S) \leq C(n)) \tag{3.3}
\]

Note that \( D_x(A; S) \) is a normal random variable with mean 0 and variance 1 when \( \mu_1 = \mu_2 = \cdots = \mu_k \). Therefore,

\[
\lim_{n \to \infty} P_{\mu}(D_x(A; S) \leq C(n)) = \lim_{n \to \infty} \Phi(C(n)) - \Phi(-C(n)) = 1
\]

where \( \Phi(\cdot) \) is the cdf of a standard normal distribution. Thus, (3.3)=0

**Lemma 3.3.3.** Let \( X_{ij} \) be independent normally distributed random variables with mean \( \mu_i \) and variance 1, where \( i = 1, \ldots, k \) and \( j = 1, \ldots, n \). Let \( \bar{x}_i \) be the sample mean of the \( i \)th population. If \( \mu_1 \leq \mu_2 \leq \cdots \leq \mu_k \), then \( \lim_{n \to \infty} P_{\mu}(\bar{x}_1 \leq \bar{x}_2 \leq \cdots \leq \bar{x}_k) = 1 \).

**Proof.**

\[
\lim_{n \to \infty} P_{\mu}(\bar{x}_1 \leq \bar{x}_2 \leq \cdots \leq \bar{x}_k) = 1 - \lim_{n \to \infty} P_{\mu}(\text{at least a pair of } i < j \text{ such that } \bar{x}_i \geq \bar{x}_j)
\geq 1 - \lim_{n \to \infty} P_{\mu}(\bar{x}_1 \geq \bar{x}_2)
= 1 - \lim_{n \to \infty} (1 - \Phi(\sqrt{n/2}(\mu_2 - \mu_1)))
= 1
\]
Lemma 3.3.4. For a balanced fixed effects one-way analysis of variance model, let $S = \{1, 2, \cdots, k\}$ and $X_{ij} \sim N(\mu_i, 1)$, $i = 1, \cdots, k$, $j = 1, \cdots, n$. Denote $D_\mathcal{E}(i; S) = D_\mathcal{E}(A = \{1, \cdots, i\}; S)$. If $\mu = (\mu_1, \cdots, \mu_k) = u_1(1_{n_1}, 0_{n_2}) + u_2(0_{n_1}, 1_{n_2})$, then for $i = 1, 2, \cdots, n_1 - 1, n_1 + 1, \cdots, k - 1$, $A \subset S$, $A \neq \emptyset$, 

$$
\lim_{n \to \infty} P_\mu(D_\mathcal{E}(n_1; S) > D_\mathcal{E}(i; S)) = 1,
$$

$$
\lim_{n \to \infty} P_\mu(D_\mathcal{E}(n_1; S) > D_\mathcal{E}(A; S)) = 1.
$$

Proof. For the first part, define $f_1(q)$ and $f_2(q)$ as in Lemma 3.3.1. Note that $D_\mathcal{E}(i; S)$ is a normal random variable with mean $\sqrt{n}f_1(\frac{i}{n_1})$ and variance 1 if $i \leq n_1$. And if $n_1 < i < k$, $D_\mathcal{E}(i; S)$ is a normal random variable with mean $\sqrt{n}f_2(\frac{i-n_1}{n_2})$ and variance 1. Since $D_\mathcal{E}(n_1; S)$ is a normal random variable with mean $\sqrt{n}f_1(1)$ or $\sqrt{n}f_2(0)$ and variance 1, 

$$
\lim_{n \to \infty} P_\mu(D_\mathcal{E}(n_1; S) > D_\mathcal{E}(i; S))
$$

$$
= \lim_{n \to \infty} P_\mu(D_\mathcal{E}(n_1; S) - D_\mathcal{E}(i; S) > 0)
$$

$$
= \begin{cases} 
\lim_{n \to \infty} 1 - \Phi(-\sqrt{n}(f_1(1) - f_1(\frac{i}{n_1}))) & \text{if } i < n_1 \\
\lim_{n \to \infty} 1 - \Phi(-\sqrt{n}(f_2(0) - f_2(\frac{i-n_1}{n_2}))) & \text{if } i > n_1
\end{cases}
$$

From Lemma 3.3.1, $f_1(1) - f_1(\frac{i}{n_1}) > 0$ and $f_2(0) - f_2(\frac{i-n_1}{n_2}) > 0$. Thus,

$$
\lim_{n \to \infty} P_\mu(D_\mathcal{E}(n_1; S) > D_\mathcal{E}(i; S)) = 1.
$$

Next, define the event $M$ be $\{\bar{x}_1 \leq \bar{x}_2 \leq \cdots \leq \bar{x}_k\}$. Note that, 

$$
\lim_{n \to \infty} P_\mu(D_\mathcal{E}(n_1; S) > D_\mathcal{E}(A; S))
$$

$$
= \lim_{n \to \infty} P_\mu(D_\mathcal{E}(n_1; S) > D_\mathcal{E}(A; S)|M)P_\mu(M)
$$

$$
+ \lim_{n \to \infty} P_\mu(D_\mathcal{E}(n_1; S) > D_\mathcal{E}(A; S)|\text{not } M)P_\mu(\text{not } M)
$$

From Lemma 3.3.3, $\lim_{n \to \infty} P_\mu(M) = 1$. The above equation can be simplified as

$$
\lim_{n \to \infty} P_\mu(D_\mathcal{E}(n_1; S) > D_\mathcal{E}(A; S)|M)
$$

$$
= \lim_{n \to \infty} P_\mu(D_\mathcal{E}(n_1; S) > D_\mathcal{E}(i; S))
$$

$$
= 1
$$

$$
\square
$$
Lemma 3.3.5. Let $C(n)$ be a constant so that $\lim_{n \to \infty} C(n) = 0$ and $C(n) = o(\sqrt{n})$. For a balanced fixed effects one-way analysis of variance model, let $S = \{1, 2, \cdots, k\}$ and $X_{ij} \sim N(\mu_i, 1)$, $i = 1, \cdots, k$, $j = 1, \cdots, n$. Denote $D_x(i; S) = D_x(A = \{1, \cdots, i\}; S)$. If $\mu = (\mu_1, \cdots, \mu_k) = u_1(1_{n_1}, 0_{n_2}) + u_2(0_{n_1}, 1_{n_2})$, then

\[
\lim_{n \to \infty} P_\mu(\psi_{ij} = 1) = 1 \quad \text{if} \quad \mu_i \neq \mu_j,
\]
\[
\lim_{n \to \infty} P_\mu(\psi_{ij} = 0) = 1 \quad \text{if} \quad \mu_i = \mu_j.
\]

Proof. For the first part, in the case of $\mu_i \neq \mu_j$, let the event $M$ be \{\(\bar{x}_1 \leq \bar{x}_2 \leq \cdots \leq \bar{x}_k\}\).

\[
\lim_{n \to \infty} P_\mu(\psi_{ij} = 1) = \lim_{n \to \infty} P_\mu(D_x(A'; S) > C(n), i \in A', j \in S\setminus A') = \lim_{n \to \infty} P_\mu(D_x(A'; S) > C(n)|M)P_\mu(M)
\]
\[
+ \lim_{n \to \infty} P_\mu(D_x(A'; S) > C(n)|\text{not } M)P_\mu(\text{not } M)
\]
\[
= \lim_{n \to \infty} P_\mu(D_x(n_1; S) > C(n))
\]
\[
= \lim_{n \to \infty} 1 - \Phi(C(n) - \sqrt{n}f_1(1))
\]
\[
= 1
\]

For the second part, if $\mu_i = \mu_j$,

\[
\lim_{n \to \infty} P_\mu(\psi_{ij} = 0)
\]
\[
= \lim_{n \to \infty} 1 - P_\mu(\psi_{ij} = 1)
\]
\[
= \lim_{n \to \infty} 1 - P_\mu(D_x^*(S) = D_x(A'; S) > C(n), i \in A', j \in S\setminus A')
\]
\[
= \lim_{n \to \infty} 1 - P_\mu(D_x^*(S) = D_x(A'; S) > C(n), i \in A', j \in S\setminus A'|M)
\]
\[
= \lim_{n \to \infty} 1 - P_\mu(D_x^*(S) = D_x(i); S, i \neq n_1 \text{ and } D_x^*(S) > C(n))
\]
\[
\geq \lim_{n \to \infty} 1 - P_\mu(D_x^*(S) = D_x(i); S, i \neq n_1)
\]

From Lemma 3.3.4, $\lim_{n \to \infty} P_\mu(D_x^*(S) = D_x(i); S, i \neq n_1) = 0$. Thus, $\lim_{n \to \infty} P_\mu(\psi_{ij} = 0) = 1$ if $\mu_i = \mu_j$. \qed

Lemma 3.3.6. For a balanced fixed effects one-way analysis of variance model, let $S = \{1, 2, \cdots, k\}$ and $X_{ij} \sim N(\mu_i, 1)$, $i = 1, \cdots, k$, $j = 1, \cdots, n$. Denote $D_x(i; S) = D_x(A = \{1, \cdots, i\}; S)$. If $\mu = (\mu_1, \cdots, \mu_k) = u_1(1_{n_1}, 0_{n_2}) + u_3(0_{n_1}, 1_{n_2}) + u_2(0_{n_1}, 0_{n_3}, 1_{n_2})$, then for $i = n_1 + 1, n_1 + 2, \cdots, n_1 - 1, n_1 + 1, \cdots, n_1 + n_3 - 1, A \subset S$, \(\mu_i \neq \mu_j\).
For a balanced fixed effects one-way analysis of variance model, let
Lemma 3.3.7.

Proof. Let $f_3(q)$ be defined as in Lemma 3.3.1. Note that $D_x(i; S)$ for $i = n_1 + 1, n_1 + 2, \cdots, n_1 - 1, n_1 + 1, \cdots, n_1 + n_3 - 1$ is a normal random variable with mean $\sqrt{n} f_3(\frac{j - n_1}{n_3})$ and variance 1. And note that $D_x(n_1; S) \sim N(\sqrt{n} f_3(0), 1)$, $D_x(n_1 + n_3; S) \sim N(\sqrt{n} f_3(1), 1)$. Thus,

$$
\lim_{n \to \infty} P_\mu(\max(D_x(n_1; S), D_x(n_1 + n_3; S)) > D_x(i; S)) = \begin{cases} 
\lim_{n \to \infty} P_\mu(D_x(n_1; S) > D_x(i; S)) & \text{if } D_x(n_1; S) > D_x(n_1 + n_3; S) \\
\lim_{n \to \infty} P_\mu(D_x(n_1 + n_3; S) > D_x(i; S)) & \text{if } D_x(n_1 + n_3; S) > D_x(n_1; S) \\
\lim_{n \to \infty} P_\mu(D_x(n_1; S) - D_x(i; S) > 0) & \\
\lim_{n \to \infty} P_\mu(D_x(n_1 + n_3; S) - D_x(i; S) > 0) & \\
\lim_{n \to \infty} 1 - \Phi(-\sqrt{n}(f_3(0) - f_3(\frac{j - n_1}{n_3}))) & \\
\lim_{n \to \infty} 1 - \Phi(-\sqrt{n}(f_3(1) - f_3(\frac{j - n_1}{n_3}))) & \\
1 & 
\end{cases}
$$

Next, let the event $M$ be $\{\bar{x}_1 \leq \bar{x}_2 \leq \cdots \leq \bar{x}_k\}$,

$$
\lim_{n \to \infty} P_\mu(\max(D_x(n_1; S), D_x(n_1 + n_3; S)) > D_x(A; S)) = \lim_{n \to \infty} P_\mu(\max(D_x(n_1; S), D_x(n_1 + n_3; S)) > D_x(A; S)|M) P_\mu(M) \\
+ \lim_{n \to \infty} P_\mu(\max(D_x(n_1; S), D_x(n_1 + n_3; S)) > D_x(A; S)|\text{not } M) P(\text{not } M) \\
= \lim_{n \to \infty} P_\mu(\max(D_x(n_1; S), D_x(n_1 + n_3; S)) > D_x(A; S)|M) \\
= \lim_{n \to \infty} P_\mu(\max(D_x(n_1; S), D_x(n_1 + n_3; S)) > D_x(i; S)|M) \\
= 1
$$

Lemma 3.3.7. Let $C(n)$ be a constant so that $\lim_{n \to \infty} C(n) = 0$ and $C(n) = o(\sqrt{n})$. For a balanced fixed effects one-way analysis of variance model, let $S = \{1, 2, \cdots, k\}$ and $X_{ij} \sim N(\mu_i, 1)$, $i = 1, \cdots, k$, $j = 1, \cdots, n$. Denote $D_x(i; S) = D_x(A = \{1, \cdots, i\}; S)$. 

\[\square\]
If $\underline{\mu} = (\mu_1, \cdots, \mu_k) = u_1(1_{n_1}, 0_{n_3}, 0_{n_2}) + u_3(0_{n_1}, 1_{n_3}, 0_{n_2}) + u_2(0_{n_1}, 0_{n_3}, 1_{n_2})$, then

$$
\lim_{n \to \infty} P_{\underline{\mu}}(\psi_{ij} = 0) = 1 \quad \text{if } \mu_i = \mu_j,
$$

$$
\lim_{n \to \infty} P_{\underline{\mu}}(\psi_{ij} = 1) = 1 \quad \text{if } \mu_i \neq \mu_j.
$$

**Proof.** We begin with the first part. Let the event $M$ be $\{\bar{x}_1 \leq \bar{x}_2 \leq \cdots \leq \bar{x}_k\}$. In the case of $n_1 < i, j < n_1 + n_3$, when $\mu_i = \mu_j$,

$$
\lim_{n \to \infty} P_{\underline{\mu}}(\psi_{ij} = 0)
= \lim_{n \to \infty} 1 - P_{\underline{\mu}}(\psi_{ij} = 1)
= \lim_{n \to \infty} 1 - P_{\underline{\mu}}(D^*_\underline{x}(S) = D_{\underline{x}}(A'; S) > C(n), i \in A', j \in S \backslash A'|M) P_{\underline{\mu}}(M)
+ \lim_{n \to \infty} 1 - P_{\underline{\mu}}(D^*_\underline{x}(S) = D_{\underline{x}}(A'; S) > C(n), i \in A', j \in S \backslash A'|\not M) P_{\underline{\mu}}(\not M)
= \lim_{n \to \infty} 1 - P_{\underline{\mu}}(D^*_\underline{x}(S) = D_{\underline{x}}(i; S) > C(n)|M)
= \lim_{n \to \infty} 1 - P_{\underline{\mu}}(D^*_\underline{x}(S) = D_{\underline{x}}(i; S) > C(n))
\geq \lim_{n \to \infty} 1 - P_{\underline{\mu}}(D^*_\underline{x}(S) = D_{\underline{x}}(i; S), i = n_1 + 1, \cdots, n_1 + n_3 - 1)
$$

Note that Lemma 3.3.6 implies $\lim_{n \to \infty} P_{\underline{\mu}}(D^*_\underline{x}(S) = D_{\underline{x}}(i; S), i = n_1 + 1, \cdots, n_1 + n_3 - 1) = 0$. Thus, $\lim_{n \to \infty} P_{\underline{\mu}}(\psi_{ij} = 0) = 1$ if $\mu_i = \mu_j$ for $n_1 < i, j < n_1 + n_3$.

In the case of $i, j < n_1$ or $i, j > n_1 + n_3$, let $\mu'_{\underline{\mu}}$ be the weighted mean of $\mu_2$ and $\mu_3$. From Lemma 3.3.5, it can be shown that $\lim_{n \to \infty} P_{\underline{\mu}}(\psi_{ij} = 0) = 1$ if $\mu_i = \mu_j$ for $i, j < n_1$. Likewise, letting $\mu'_1$ be the weighted mean of $\mu_1$ and $\mu_3$, Lemma 3.3.5 implies $\lim_{n \to \infty} P_{\underline{\mu}}(\psi_{ij} = 0) = 1$ if $\mu_i = \mu_j$ for $i, j > n_1 + n_3$.

For the second part, if $\mu_i \neq \mu_j$, we obtain

$$
\lim_{n \to \infty} P_{\underline{\mu}}(\psi_{ij}(\underline{x}) = 1)
= \lim_{n \to \infty} P_{\underline{\mu}}(D^*_\underline{x}(S) = D_{\underline{x}}(A'; S) > C(n), i \in A', j \in S \backslash A')
= \lim_{n \to \infty} P_{\underline{\mu}}(D^*_\underline{x}(S) = D_{\underline{x}}(A'; S) > C(n), i \in A', j \in S \backslash A'|M)
= \lim_{n \to \infty} P_{\underline{\mu}}(\max(D_{\underline{x}}(n_1; S), D_{\underline{x}}(n_1 + n_3; S)) = D^*_\underline{x}(S) > C(n))
= \lim_{n \to \infty} P_{\underline{\mu}}(D^*_\underline{x}(S) > C(n)|\max(D_{\underline{x}}(n_1; S), D_{\underline{x}}(n_1 + n_3; S)) = D^*_\underline{x}(S))
\times P_{\underline{\mu}}(\max(D_{\underline{x}}(n_1; S), D_{\underline{x}}(n_1 + n_3; S)) = D^*_\underline{x}(S))
$$

(3.4)
Lemma 3.3.6 implies \( \lim_{n \to \infty} P_{\mu}(\max(D_{\mu}(n_1; S), D_{\mu}(n_1 + n_3; S)) = D_{\mu}^{*}(S)) = 1 \). Consequently, (3.4) becomes

\[
\lim_{n \to \infty} P_{\mu}(D_{\mu}^{*}(S) > C(n) | \max(D_{\mu}(n_1; S), D_{\mu}(n_1 + n_3; S)) = D_{\mu}^{*}(S)) = \begin{cases} 
\lim_{n \to \infty} (1 - \Phi(C(n) - \sqrt{n}f_3(0))) & \text{if } D_{\mu}(n_1; S) > D_{\mu}(n_1 + n_3; S) \\
\lim_{n \to \infty} (1 - \Phi(C(n) - \sqrt{n}f_3(1))) & \text{if } D_{\mu}(n_1; S) < D_{\mu}(n_1 + n_3; S)
\end{cases}
= 1
\]

\[\square\]

**Theorem 3.3.8.** For \( x \sim N(\mu, I_k) \), \( \mu = (\mu_1, \cdots, \mu_k) \), PADD is a consistent procedure provided that all critical values, \( C(n) \)'s, grows with the order of magnitude \( o(\sqrt{n}) \).

**Proof.** Since PADD is a stepwise procedure, it is consistent if it does not make any mistake at any step. Without loss of generality, we consider the action at the first step. If all null hypotheses are true, Lemma 3.3.2 shows that all nulls will accepted for sufficiently large \( n \). If there are two means among \( \mu \), Lemma 3.3.5 implies that, PADD will reject false nulls and stop when the remaining hypotheses are all true, for sufficiently large \( n \). Lemma 3.3.7 describes the consistent action of PADD for three means among \( \mu \). For a general \( \mu \) that contains \( p \) means, \( \mu_1 < \mu_2 < \cdots < \mu_p \), and

\[\mu = (u_1(1_{n_1}, 0_{n_2}, \cdots, 0_{n_p}) + u_2(0_{n_1}, 1_{n_2}, \cdots, 0_{n_p}) + \cdots + u_p(0_{n_1}, 0_{n_2}, \cdots, 1_{n_p}))\]

Consider the actions relevant to the samples with mean \( u_j \). Denote \( u'_{j-1} \) as the weighted mean calculated from \( u_1, \cdots, u_{j-1} \), and \( u'_{j+1} \) be the one calculated from \( u_{j+1}, u_{j+2}, \cdots, u_p \). Then one can apply the proof for the case of three-mean case to \( u'_{j-1} \), \( u_j \) and \( u'_{j+1} \), and show that, for sufficiently large \( n \), PADD will not reject true nulls \( H_{uv} : \mu_u = \mu_v = u_j \) and will always split \( S = \{1, 2, \cdots, k\} \) at \( n_1 + n_2 + \cdots + n_{j-1} \) and \( n_1 + n_2 + \cdots + n_j \). \[\square\]

**Theorem 3.3.9.** For a balanced fixed effects one-way analysis of variance model, let \( S = \{1, 2, \cdots, k\} \) and \( X_{ij} \sim N(\mu_i, 1) \), \( i = 1, \cdots, k; j = 1, \cdots, n \). PADD+ is a consistent procedure provided that all critical values, \( C(n), C_u \) and \( C_l \), grow with the order of magnitude \( o(\sqrt{n}) \).
Proof. To prove PADD+ is consistent, it suffices to show that, for sufficiently large sample size, the actions of PADD and PADD+ are the same.

Let $\psi_{ij}(\bar{x})$ and $\psi'_{ij}(\bar{x})$ be the test function of PADD and PADD+, respectively.

$$
\Pr(\psi_{ij}(\bar{x}) \neq \psi'_{ij}(\bar{x}))
= \Pr(\psi_{ij}(\bar{x}) = 1, \psi'_{ij}(\bar{x}) = 0) + \Pr(\psi_{ij}(\bar{x}) = 0, \psi'_{ij}(\bar{x}) = 1)
$$

$$
= \Pr(\psi_{ij}(\bar{x}) = 1, \psi'_{ij}(\bar{x}) = 0|\mu_i = \mu_j) + \Pr(\psi_{ij}(\bar{x}) = 0, \psi'_{ij}(\bar{x}) = 1|\mu_i = \mu_j)
$$

$$
+ \Pr(\psi_{ij}(\bar{x}) = 1, \psi'_{ij}(\bar{x}) = 0|\mu_i \neq \mu_j) + \Pr(\psi_{ij}(\bar{x}) = 0, \psi'_{ij}(\bar{x}) = 1|\mu_i \neq \mu_j)
\leq \Pr(\psi_{ij}(\bar{x}) = 1|\mu_i = \mu_j) + \Pr(\psi_{ij}(\bar{x}) = 0, \psi'_{ij}(\bar{x}) = 1|\mu_i = \mu_j)
$$

$$
+ \Pr(\psi_{ij}(\bar{x}) = 1, \psi'_{ij}(\bar{x}) = 0|\mu_i \neq \mu_j) + \Pr(\psi_{ij}(\bar{x}) = 0|\mu_i \neq \mu_j)
$$

From Theorem 3.3.8, $\lim_{n \to \infty} P(\psi_{ij}(\bar{x}) = 1|\mu_i = \mu_j) = 0$ and $\lim_{n \to \infty} P(\psi_{ij}(\bar{x}) = 0|\mu_i \neq \mu_j) = 0$. Besides,

$$
\lim_{n \to \infty} P(\psi_{ij}(\bar{x}) = 1, \psi'_{ij}(\bar{x}) = 0|\mu_i \neq \mu_j)
$$

$$
\leq \lim_{n \to \infty} P(|\bar{x}_i - \bar{x}_j| < C_l|\mu_i \neq \mu_j)
$$

$$
= \lim_{n \to \infty} \Phi(C_l - \sqrt{n}(|\mu_i - \mu_j|)) - \Phi(-C_l - \sqrt{n}(|\mu_i - \mu_j|))
$$

$$
= 0,
$$

$$
\lim_{n \to \infty} P(\psi_{ij}(\bar{x}) = 0, \psi'_{ij}(\bar{x}) = 1|\mu_i = \mu_j)
$$

$$
\leq \lim_{n \to \infty} P(|\bar{x}_i - \bar{x}_j| > C_u|\mu_i = \mu_j)
$$

$$
= \lim_{n \to \infty} 1 - \Phi(C_u) + \Phi(-C_u)
$$

$$
= 0
$$

Therefore, $\lim_{n \to \infty} \Pr(\psi_{ij}(\bar{x}) \neq \psi'_{ij}(\bar{x})) = 0$. \qed

Remark 3.3.1. In the case of unknown constant variance, one can replace $\sigma^2$ by $s^2 = \frac{\sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2}{k(n-1)}$. The consistency results remain valid, since $s^2$ converges in probability to $\sigma^2$ for sufficiently large $n$. And the change of distribution form will not alter consistency results.
Chapter 4
Consistency of change point problem

This chapter considers the consistency of MRD (maximum residual down) and BSP (binary segmentation procedure) for the change point problem.

The change point problem considered here is as follows. Let $X_{ij}, i = 1, \cdots, k, j = 1, \cdots, n$ be independent normally distributed random variables with means $\mu_i$ and known variance $\sigma^2$. Consider the testing of consecutive mean change. That is, $H_i: \mu_i - \mu_{i+1} = 0$ vs. $K_i: \mu_i - \mu_{i+1} \neq 0, i = 1, \cdots, k - 1$. Various approaches to this change point problem have been considered. See Chen and Gupta (2000) for details. One such approach attributed to Vostrikova (1981) is a stepwise procedure called binary segmentation procedure (BSP). Note that the statistics used in BSP are the same as the ones used by Cohen, Sackrowitz and Xu (2009).

Consistent here means that the probability of making any mistakes, regarding any of the $k - 1$ hypotheses testing problems, goes to zero as $n \to \infty$. Consistency in this sense for other multiple testing problems has previously been considered by An and Gu (1985) and Bunea, Wegkamp and Auguste (2006).

Cohen, Sackrowitz and Xu (2009) indicated that most standard stepwise multiple testing procedures (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 BSP and MRD procedures are admissible.

The consistency results are easily extended to the case of unknown common variance $\sigma^2$.

Section 1 provides some notations and preliminaries. Section 2 gives the description of BSP and MRD procedures. Section 3 contains the lemmas and the theorem establishing consistency. Extensions are also mentioned in Section 3.
4.1 Notations and preliminaries

Denote sample means for each population as $\bar{x}_i = \sum_{j=1}^{n} x_{ij}/n$, $i = 1, \cdots, k$. Let $\bar{x} = (\bar{x}_1, \cdots, \bar{x}_k)$ be a sample point. Without loss of generality, assume the known variance $\sigma^2$ to be 1, unless stated otherwise. Define $\nu_i = \mu_i - \mu_{i+1}$, $i = 1, \cdots, k - 1$. Here $\nu_i \neq 0$ represents a change point exists between $i$ and $i+1$. Consider all procedures as finite action procedures and each procedure (nonrandomized) either accepts or rejects each of the $k - 1$ hypotheses. A vector test function consists of $k - 1$ elements, $\Psi = (\psi_1(\bar{x}), \cdots, \psi_{k-1}(\bar{x}))$, where $\psi_i(\bar{x}) = 1$ if $H_i$ is rejected and $\psi_i(\bar{x}) = 0$ if $H_i$ is accepted. A procedure is consistent if for every $i$,

$$\lim_{n \to \infty} P_{\mu}\{\psi_i(\bar{x}) = 1\} = 1 \quad \text{if } \nu_i \neq 0$$

and

$$\lim_{n \to \infty} P_{\mu}\{\psi_i(\bar{x}) = 0\} = 1 \quad \text{if } \nu_i = 0.$$

In other words, as $n \to \infty$, the probability of any mistake tends to 0.

4.2 BSP and MRD procedures

Let $I = \{1, 2, \cdots, k\}$. Let $B = B(t_1, t_2) = \{t_1, t_1 + 1, \cdots, t_2\}$, be the subset of consecutive integers from $t_1$ to $t_2$. Let $A(t_1, i) = \{t_1, \cdots, i\}$, a subset of $B$ consisting of the consecutive integers from $t_1$ to $i$. Define

$$D_{\bar{x}}(A(t_1, i); B) = n(\bar{x}_A - \bar{x}_{B\setminus A})^2/\sigma^2\left(\frac{1}{i-t_1+1} + \frac{1}{t_2-i}\right)$$ (3.1)

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

$$D^*_{\bar{x}}(B) = \max_{t_1 \leq i < t_2} D_{\bar{x}}(A(t_1, i); B)$$ (3.2)

Note that, $D_{\bar{x}}(A(t_1, i); B)$ represents the square of the standardized difference of two means calculated from $x_{t_1}$ to $x_i$ and $x_{i+1}$ to $x_{t_2}$ and $D^*_{\bar{x}}(B)$ is the normalized maximum squared difference of two means among possible segmentation. Let $A(t_1, i^*(\bar{x}))$ denote the set for which the maximum is attained. That is,

$$D^*_{\bar{x}}(B) = D_{\bar{x}}(A(t_1, i^*(\bar{x}); B).$$
The MRD and BSP procedures are as follows. At the first step of BSP or MRD, consider all set partitions of consecutive integers of $I$. That is, $A(1, i)$, $I\backslash A(1, i)$, $i = 1, \cdots, k - 1$. Compute corresponding $D_\mathcal{E}(A(1, i); I)$ for $i = 1, \cdots, k - 1$. Let $C_n(I)$ be a constant at step 1. Letting $D_1 = D_\mathcal{E}(I)$ and compare with $C_n(I)$. If $D_1 \leq C_n(I)$, stop and accept all null hypotheses. If $D_1 > C_n(I)$, then partition $I$ into $A(1, i^*(x))$ and $I\backslash A(1, i^*(x))$ and continue to step 2.

At step 2, for BSP, each of $A(1, i^*(x))$ and $I\backslash A(1, i^*(x))$ is treated as $I$ was at step 1. That is, calculate $D_\mathcal{E}(A(1, i); A(1, i^*(x)))$ for $i = 1, \cdots, i^*(x) - 1$, and $D_\mathcal{E}(A(i^*(x) + 1, j); I\backslash A(1, i^*(x)))$ for $j = i^*(x) + 1, \cdots, k - 1$. For $A(1, i^*(x))$, either all hypotheses determined by the indices in $A(1, i^*(x))$ are accepted, or $A(1, i^*(x))$ is split into two sets. Similarly for $I\backslash A(1, i^*(x))$. At each successive step, until the procedures stops, each set of the current partition could possibly be split into two sets. Note that, once an action of not splitting has been taken on a set, the hypotheses determined by the indices of the set are accepted.

Although MRD also treats $A(1, i^*(x))$ and $I\backslash A(1, i^*(x))$ as $I$ was at step 1, MRD only splits one of the two subsets. More specifically, at step 2, BSP uses two, possibly different constants, $C_n(A(1, i^*(x)))$ and $C_n(I\backslash A(1, i^*(x)))$, depending on the size of the sets. On the other hand, MRD uses the same constant for both sets at step 2. In other words, only the most significant test statistics from $A(1, i^*(x))$ and $I\backslash A(1, i^*(x))$ is considered.

At successive steps, BSP takes each subset into consideration and either keeps or splits the set based on the comparison with an appropriate constant. On the other hand, MRD always considers to split the most significant subset and only a single constant is used for all the current subsets. That is to say, subsets whose maximum statistic is less than the given constant do not give rise to acceptance unless all subsets yield maximum statistics less than the critical value for that stage. That is, the MRD procedure stops.

Critical values for MRD can be decided on a priori with deference to a variety of criteria. At step 1, typically the critical value is determined by a Bonferroni bound using a significance level $\alpha$. That is, if there are $k$ elements in $I$, one determines the critical value as the $1 - \alpha / (k - 1)$ percentile of a chi-square distribution with 1
degree of freedom. At successive stages, one can also choose the critical value to be the \(1 - \alpha/(r - 1)\) percentile of the chi-square distribution if there are \(r\) elements in the set under study. Such selections would certainly control weak FWER. Other methods of choosing critical values can be used, depending on how conservative one wishes to be.

4.3 Consistency results

In this section we give a series of lemmas which will be used to prove the consistency of the BSP and MRD procedures. Note that the idea of proof is an extension from Theorem 3.3.8. We start with

**Lemma 4.3.1.** Define three functions, \(f_1(i), f_2(i)\) and \(f_3(i)\) as follows:

\[
 f_1(i) = \frac{(k - n_1)\theta_i^2}{\frac{1}{i} + \frac{1}{k - i}}, i = 1, \ldots, n_1
\]

\[
 f_2(i) = \frac{(\theta - \frac{(i - n_1)\theta_i}{i})^2}{\frac{1}{i} + \frac{1}{k - i}}, i = n_1, \ldots, k
\]

\[
 f_3(i) = \frac{\left(n_1\theta_3 + \frac{(i - n_1)\theta_3}{i} - \frac{n_2\theta_2 + (k - n_2 - i)\theta_3}{k - i}\right)^2}{\frac{1}{i} + \frac{1}{k - i}}, i = n_1, \ldots, k - n_2
\]

for \(0 < n_1, n_2 < k\),

(1) \(f_1(i)\) and \(f_2(i)\) attain their maximum at \(i = n_1\).

(2) \(f_3(i)\) has its maximum at either \(i = n_1\) or \(i = k - n_2\).

**Proof.** With some calculation, \(f_1(i)\) can be simplified as

\[
 f_1(i) = \frac{i((k - n_1)\theta_i)^2}{k(k - i)}.
\]

Since \(f_1(i)\) is nonnegative for \(i \leq n_1\), the point at which \(f_1\) attains its maximum is the same as that of \(\log f_1(i)\). And

\[
 \log f_1(i) = 2\log(k - n_1) + 2\log \theta + \log i - \log(k - i) - \log k.
\]

Notice that \(\partial \log f_1(i)/\partial i = \frac{1}{i} + \frac{1}{k - i}\) is positive for \(i \leq n_1\), therefore \(\log f_1(i)\) attains its maximum at \(i = n_1\), so does \(f_1(i)\).
Likewise, simplify $f_2(i)$ and its logarithm transformation is

$$
\log f_2(i) = 2 \log n_1 \theta + \log(k - i) - \log i - \log k.
$$

Note that $\partial \log f_2(i)/\partial i = \frac{-1}{k - i} - \frac{1}{i}$, is decreasing for $i \geq n_1$. Consequently, $\log f_2(i)$ attains its max at $i = n_1$, so does $f_2(i)$.

For the last part, let $n_1 \theta_1 = s_1$ and $n_2 \theta_2 = s_2$. $f_3(i)$ can be rewritten as

$$
f_3(i) = \frac{(s_1 + (i - n_1) \theta_3 \cdot s_2 + (k - n_2 - i) \theta_3)^2}{(k - i) + i(k - i)}
= \frac{(s_1 + (k - i)(i - n_1) \theta_3 - is_2 - i(k - n_2 - i) \theta_3)^2}{i(k - i)}
= \frac{(ks_1 - (s_1 + s_2) i - n_1 k \theta_3 + (n_1 + n_2) \theta_3 i)^2}{ki(k - i)}
$$

and the logarithm transformation of $f_3(i)$,

$$
\log f_3(i) = 2 \log (ks_1 - n_1 k \theta_3 + (n_1 + n_2) \theta_3 i - (s_1 + s_2) i) - \log k - \log i - \log (k - i).
$$

The function $\log f_3(i)$ has its extremum at $i = n_1, k - n_2$, or $\tilde{i}$, which satisfies the following equation,

$$
2 \frac{(n_1 + n_2) \theta_3 - (s_1 + s_2)}{ks_1 - n_1 k \theta_3 + (n_1 + n_2) \theta_3 i - (s_1 + s_2) i} - \frac{1}{i} + \frac{1}{k - i} = 0.
$$

The second derivative of $\log f_3$ evaluated at $\tilde{i}$,

$$
\frac{\partial^2 \log f_3}{\partial i^2} = 2 \frac{((n_1 + n_2) \theta_3 - (s_1 + s_2))^2}{(ks_1 - n_1 k \theta_3 - (s_1 + s_2) i + (n_1 + n_2) \theta_3 i)^2} + \frac{1}{i^2} + \frac{1}{(k - i)^2}
$$

is always positive. Thus, the maximum of $\log f_3$ would be at $i = n_1$ or $i = k - n_2$.

\textbf{Lemma 4.3.2.} Let $C(n)$ be a constant so that $C(n) \to \infty$ as $n \to \infty$ and $C(n) = o(n)$. Let $I = \{1, 2, \cdots, k\}$. For $X \sim N(\mu, I_k)$, $\mu = (\mu_1, \mu_2, \cdots, \mu_k)$, and $\mu_1 = \mu_2 = \cdots = \mu_k$, then

$$
\lim_{n \to \infty} P_\mu \{D^*_x(I) > C(n)\} = 0.
$$

\textbf{Proof.} We start with

$$
\lim_{n \to \infty} P_\mu \{D^*_x(I) \leq C(n)\}
\geq \lim_{n \to \infty} \sum_{i=1}^{k-1} P_\mu \{D^*_x(A(1, i); I) \leq C(n), \forall i\} - (k - 2)
$$

(4.1)
Note that $D_\mu(A(1,i);I)$ is the square of a normal random variable with mean 0 and variance 1 when $\mu_1 = \mu_2 = \cdots = \mu_k$. Then, (4.1) becomes $\lim_{n \to \infty}(k-1)(\Phi(\sqrt{C(n)}) - \Phi(-\sqrt{C(n)})) - (k-2) = 1$. Thus, $\lim_{n \to \infty} P_\mu\{D_\mu^*(I) > C(n)\} = 0$. \hfill \Box

**Lemma 4.3.3.** Let $I = \{1, 2, \cdots, k\}$. For $X \sim N(\mu, I_k)$ and $\mu = u_1(1_{n_1}, 0_{k-n_1}) + u_2(0_{n_1}, 1_{k-n_1})$, then
\[
\lim_{n \to \infty} P\{D_\mu^*(I) = D_\mu^*(A(1,n_1); I)\} = 1.
\]

**Proof.**
\[
\lim_{n \to \infty} P_\mu\{D_\mu^*(I) = D_\mu^*(A(1,n_1); I)\}
= \lim_{n \to \infty} P_\mu\{D_\mu^*(A(1,n_1); I) \geq D_\mu^*(A(1,i); I), \forall i \neq n_1\}
\geq \lim_{n \to \infty} \sum_{i \neq n_1} P_\mu\{D_\mu^*(A(1,n_1); I) \geq D_\mu^*(A(1,i); I)\} - (k-2) \quad (4.2)
\]

Note that $D_\mu^*(A(1,n_1); I)$ is the square of a normal random variable with mean $\sqrt{n_1f_1(n_1)}$ and variance 1, and $D_\mu^*(A(1,i); I)$ is the square of a normal random variable with mean $\sqrt{n_1f_1(i)}$ and variance 1 for $i \leq n_1$, and mean $\sqrt{n_2f_2(i)}$ and variance 1 for $i > n_1$, where $f_1(i)$ and $f_2(i)$ are defined as in Lemma 4.3.1. Thus, (4.2) can be further expressed as
\[
\lim_{n \to \infty} \left( \sum_{i=1}^{n_1-1} 1 - \Phi(\sqrt{n_1f_1(i)} - \sqrt{n_1f_1(n_1)}) \right.
+ \left. \sum_{i=n_1+1}^{k-1} 1 - \Phi(\sqrt{n_2f_2(i)} - \sqrt{n_2f_2(n_1)}) \right) - (k-2)
= 1
\]

\hfill \Box

**Lemma 4.3.4.** Let $C(n)$ be a constant so that $C(n) \to \infty$ as $n \to \infty$ and $C(n) = o(n)$. Let $I = \{1, 2, \cdots, k\}$. For $X \sim N(\mu, I_k)$ and $\mu = u_1(1_{n_1}, 0_{k-n_1}) + u_2(0_{n_1}, 1_{k-n_1})$, then
\[
\lim_{n \to \infty} P_\mu\{\psi_\ell(x) = 1\} = 1 \quad \text{if } \mu_\ell \neq \mu_{\ell+1}
\]
and
\[
\lim_{n \to \infty} P_\mu\{\psi_\ell(x) = 0\} = 1 \quad \text{if } \mu_\ell = \mu_{\ell+1}.
\]
Proof. For the first part, if \( \mu_i \neq \mu_{i+1} \),

\[
\lim_{n \to \infty} P_\mu \{ \psi_1(x) = 1 \} = \lim_{n \to \infty} P_\mu (D^*_{\mu}(I) = D_{\mu}(A(1, n_1); I) > C(n))
\]

\[
\lim_{n \to \infty} P_\mu \{ D^*_{\mu}(I) > C(n) | D^*_{\mu}(I) = D_{\mu}(A(1, n_1); I) \}
\]

\[
\times P_\mu \{ D^*_{\mu}(I) = D_{\mu}(A(1, n_1); I) \}
\]

(4.3)

Lemma 4.3.3 implies \( \lim_{n \to \infty} P_\mu \{ D^*_{\mu}(I) = D_{\mu}(A(1, n_1); I) \} = 1 \). Since \( D_{\mu}(A(1, n_1); I) \) is the square of a normal distribution with mean \( \sqrt{nf_1(n_1)} \) or \( \sqrt{nf_2(n_1)} \) and variance 1, (4.3) can be rewritten as

\[
\lim_{n \to \infty} P_\mu (D^*_{\mu}(I) > C(n) | D^*_{\mu}(I) = D_{\mu}(A(1, n_1); I))
\]

\[
= \lim_{n \to \infty} P_\mu (D_{\mu}(A(1, n_1); I)) > C(n)
\]

\[
= \lim_{n \to \infty} 1 - \Phi(\sqrt{C(n)} - \sqrt{nf_1(n_1)}) + \Phi(-\sqrt{C(n)} - \sqrt{nf_1(n_1)})
\]

\[
= 1
\]

For the second part, note that when \( \mu_i = \mu_{i+1} \), \( P_\mu(\psi_1(x) = 0) = P_\mu(D_{\mu}(A(1, i); I) \leq C(n), \text{ for } i \neq n_1) \). Since \( D_{\mu}(A(1, i); I) \) is the square of a normal random variable with mean \( \sqrt{nf_1(i)} \) and variance 1 if \( i < n_1 \), and for \( i > n_1 \), \( D_{\mu}(A(1, i); I) \) is the square of a normal random variable with mean \( \sqrt{nf_2(i)} \) and variance 1. Therefore,

\[
\lim_{n \to \infty} P_\mu (D_{\mu}(A(1, i); I) \leq C(n))
\]

\[
= \begin{cases} 
  \lim_{n \to \infty} 1 - \Phi(\sqrt{C(n)} - \sqrt{nf_1(i)}) + \Phi(-\sqrt{C(n)} - \sqrt{nf_1(i)}) & \text{if } i < n_1 \\
  \lim_{n \to \infty} 1 - \Phi(\sqrt{C(n)} - \sqrt{nf_2(i)}) + \Phi(-\sqrt{C(n)} - \sqrt{nf_2(i)}) & \text{if } i > n_1
\end{cases}
\]

\[
= 1
\]

\[
\square
\]

**Lemma 4.3.5.** Let \( I = \{1, 2, \cdots, k\} \). For \( X \sim N(\mu, I_k) \), and \( \mu = u_1(1_{n_1}, 0_{n_2}, 0_{n_2}) + u_3(0_{n_1}, 1_{n_3}, 0_{n_2}) + u_2(0_{n_1}, 0_{n_3}, 1_{n_2}) \), then

\[
\lim_{n \to \infty} P\{ \max(D_{\mu}(A(1, n_1); I), D_{\mu}(A(1, n_1 + n_3); I)) \geq D_{\mu}(A(1, i); I) \} = 1.
\]
Proof.
\[
\lim_{n \to \infty} P\{\max(D_x(A(1, n_1); I), D_x(A(1, n_1 + n_3); I)) \geq D_x(A(1, i); I)\} \\
\geq \lim_{n \to \infty} \sum_{i \neq n_1, n_1 + n_3} P\{\max(D_x(A(1, n_1); I), D_x(A(1, n_1 + n_3); I)) \geq D_x(A(1, i); I)\} \\
-(k - 3) \\
= (k - 2) - (k - 3) = 1
\]

\[\square\]

**Lemma 4.3.6.** Let \(C(n)\) be a constant so that \(C(n) \to \infty\) as \(n \to \infty\) and \(C(n) = o(n)\). Let \(I = \{1, 2, \cdots, k\}\). For \(X \sim N(\mu, I_k)\) and \(\mu = u_1(1_{n_1}, 0_{n_2}) + u_2(0_{n_1}, 1_{n_3})\), then

\[
\lim_{n \to \infty} P_{\mu}\{\psi_1(\bar{x}) = 1\} = 1 \quad \text{if } \mu_i \neq \mu_{i+1}
\]

and

\[
\lim_{n \to \infty} P_{\mu}\{\psi_1(\bar{x}) = 0\} = 1 \quad \text{if } \mu_i = \mu_{i+1}.
\]

**Proof.** Without loss of generality, let \(\theta_1 = u_1 - u_3\), \(\theta_2 = u_2 - u_3\) and \(\theta_3 = u_3 - u_3 = 0\).

For the first part, if \(\mu_i \neq \mu_{i+1}\),

\[
P_{\mu}\{\psi_1(\bar{x}) = 1\} \\
= P_{\mu}\{\max(D_x(A(1, n_1); I), D_x(A(1, n_1 + n_3); I)) = D_x^*(I) > C(n)\} \\
= P_{\mu}(D_x^*(I) > C(n)|\max(D_x(A(1, n_1); I), D_x(A(1, n_1 + n_3); I)) = D_x^*(I)) \\
\times P_{\mu}(\max(D_x(A(1, n_1); I), D_x(A(1, n_1 + n_3); I)) = D_x^*(I))
\]

From Lemma 4.3.5, \(\lim_{n \to \infty} P_{\mu}\{\max(D_x(A(1, n_1); I), D_x(A(1, n_1 + n_3); I)) = D_x^*(I)\} = 1\). Thus,

\[
\lim_{n \to \infty} P_{\mu}\{D_x^*(I) > C(n)|\max(D_x(A(1, n_1); I), D_x(A(1, n_1 + n_3); I)) = D_x^*(I)\} \\
\times P_{\mu}(\max(D_x(A(1, n_1); I), D_x(A(1, n_1 + n_3); I)) = D_x^*(I)) \\
= \begin{cases} 
\lim_{n \to \infty} 1 - \Phi\left(\sqrt{C(n)} - \sqrt{n f_3(n_1)}\right) + \Phi\left(-\sqrt{C(n)} - \sqrt{n f_3(n_1)}\right) \\
\quad \text{if } D_x(A(1, n_1); I) \geq D_x(A(1, n_1 + n_3); I) \\
\lim_{n \to \infty} 1 - \Phi\left(\sqrt{C(n)} - \sqrt{n f_3(n_1 + n_3)}\right) + \Phi\left(-\sqrt{C(n)} - \sqrt{n f_3(n_1 + n_3)}\right) \\
\quad \text{if } D_x(A(1, n_1); I) < D_x(A(1, n_1 + n_3); I)
\end{cases}
\]

\[= 1\]
Here \( f_3(i) \) is defined as in Lemma 4.3.1.

For the second part, note that \( \lim_{n \to \infty} P_\mu(\psi_i(\bar{x}) = 1) = 1 \) if \( \mu_i \neq \mu_{i+1} \) implies for sufficiently large sample size, MRD and BSP would reject either \( H_{n_1} : \mu_{n_1} = \mu_{n_1+1} \) or \( H_{n_1+n_3} : \mu_{n_1+n_3} = \mu_{n_1+n_3+1} \) at the current step. Assume \( H_{n_1} \) is rejected before \( H_{n_1+n_3} \), then \( I \) would be partitioned into \( A(1, n_1) \) and \( I \setminus A(1, n_1) \). Since all samples with indices in \( A(1, n_1) \) have equal mean \( u_1 \), lemma 4.3.2 implies \( \lim_{n \to \infty} P_\mu(\{ \psi_i(\bar{x}) = 0 \}) = 1 \) for \( i < n_1 \). On the other hand, the samples with indices in \( I \setminus A(1, n_1) \) have a mean vector \( \theta_3(1_{n_3}, 0_{n_2}) + \theta_2(0_{n_3}, 1_{n_2}) \), lemma 4.3.4 implies \( \lim_{n \to \infty} P_\mu(\{ \psi_{n_1+n_3}(\bar{x}) = 1 \}) = 1 \) and \( \lim_{n \to \infty} P_\mu(\{ \psi_i(\bar{x}) = 0 \}) = 1 \) for \( n_1 < i < k, i \neq n_1 + n_3 \).

\[ \square \]

**Theorem 4.3.7.** Let \( C_w(n) \) be constants for the BSP and MRD procedures at various stages. Suppose \( \lim_{n \to \infty} C_w(n) = \infty \) and suppose \( C_w(n) = o(n) \). Then the BSP and MRD procedures are consistent.

**Proof.** Lemma 4.3.2 implies that for sufficiently large \( n \), all nulls will be accepted. If there is one \( v_j \neq 0 \), Lemma 4.3.4 implies that for sufficiently large \( n \), BSP and MRD will reject false hypotheses and accept true nulls. If there are two \( v_j \neq 0 \), Lemma 4.3.6 implies that, for sufficiently large \( n \), BSP and MRD will not make any mistakes. For more than two change points, suppose there are \( h \) change points. That is,

\[ \mu = \mu_1 = \mu_2 = \mu_{\tau_1} \neq \mu_{\tau_1+1} = \cdots = \mu_{\tau_2} \neq \mu_{\tau_2+1} = \cdots = \mu_{\tau_h} \neq \mu_{\tau_h+1} = \cdots = \mu_k. \]

Consider the subgroup of \( \mu \) indexed from \( \tau_s + 1 \) to \( \tau_{s+1} \). Let \( \mu'_1 \) be the weighted mean of \( \mu_{1}, \cdots, \mu_{\tau_s} \) and \( \mu'_2 \) be the weighted mean of \( \mu_{\tau_{s+1}}, \cdots, \mu_k \). Then Lemma 4.3.6 can be extended to the case of general \( h \) change points.

\[ \square \]

**Remark 4.3.1.** If the assumption of constant known variance is replaced by constant unknown variance, then \( \sigma^2 \) would be replaced by \( s^2 \), where \( s^2 = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2}{k(n-1)} \). This will not alter the consistency results since \( s^2 \) converges in probability to \( \sigma^2 \) as \( n \to \infty \) and all probability assertions will remain the same. The chi-square distribution would be replaced by the F-distribution with degree of freedom 1 and \( k(n-1) \).
Chapter 5

MTPs for the matrix order problem

The focus of this chapter is on the admissibility of some MTPs for the partial matrix order problem. Conventional step-down and closed testing procedures and two other MTPs emanating from MRD are evaluated for the admissibility property for the partial matrix order problem.

Consider a two-way $R \times C$ design matrix. Let $X_{ijk}$ be the $k$th observation from the $i$th row and $j$th column, $i = 1, \ldots, R$ and $j = 1, \ldots, C$. Assume all observations $X_{ijk}$ to be independent normally distributed random variables with mean $\mu_{ij}$ and common known variance $\sigma^2$. For the sake of convenience, let $\sigma^2 = 1$ and $k = 1$, and suppress $X_{ijk}$ to $X_{ij}$.

The binary order $\preceq$ on $\{\mu_{ij}\}$ considers all binary relations for $\mu_{ij} \preceq \mu_{kl}$ if and only if $i \leq k$ and $j \leq l$. Note that the binary relation defined in the above sense is a partial order, because it is reflective, transitive and antisymmetric, but not comparable. That is, the binary relation between $\mu_{ij}$ and $\mu_{kl}$ is not restricted if $i \leq k$ and $j \geq l$.

In terms of multiple comparisons, a partial matrix order problem considers a binary relation for some pairs within a matrix. Take a $2 \times 2$ matrix order problem as an example. The matrix order problem for $\{\mu_{ij}\}$, $i = 1, 2; j = 1, 2$ can be formulated as the following hypotheses:

$H$: $\mu_{11} = \mu_{12} = \mu_{21} = \mu_{22}$

$K$: $\mu_{11} \leq \mu_{12}, \mu_{21} \leq \mu_{22}, \mu_{11} \leq \mu_{21}, \mu_{12} \leq \mu_{22}$, with at least one strict inequality.

There are other similar multiple testing problems concerning matrix order. For example, the minimum effective dose (MED) problem is expressed as

$H_{rs}: \mu_{ij} = \mu_{rs}, \forall(i, j) \preceq (r, s)$

$K_{rs}: \mu_{ij} \preceq \mu_{rs}, \forall(i, j) \preceq (r, s)$, and $\mu_{11} < \mu_{rs}$
For details, see Hellmich and Lehmacher (2005).

Note that the symmetry of the matrix order problem in terms of hypotheses and the vector risk function allow us to treat any particular individual hypothesis equally. To examine the admissibility property of MTPs, we will focus on $\mu_{11} = \mu_{12}$ vs. $\mu_{11} < \mu_{12}$. To start, we rearrange the normal random variables in a $R \times C$ matrix into a $S \times 1$ random vector, where $S = RC$. Furthermore, denote the $i$th element as $X_i$ with mean $\mu_i$ and variance 1. That is,

$$X = (X_{11}, X_{12}, \ldots, X_{1C}, X_{21}, X_{22}, \ldots, X_{2C}, \ldots, X_{R1}, \ldots, X_{RC})'$$

$$= (X_1, X_2, \ldots, X_S)'$$

Let $U = \Gamma X$ be a $S \times 1$ vector, where

$$\Gamma = \begin{pmatrix} -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & -1 & 1 & \cdots & 0 \\ \vdots \\ -1 & \cdots & \cdots & \cdots & 1 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ \vdots \\ 1 & 1 & 1 & 1 \\ \end{pmatrix}.$$ 

Then $U$ is multivariate normally distributed with mean $\nu = \Gamma \mu$ and variance-covariance matrix $\Gamma \Gamma'$. The density of $U$ can be written as

$$f_U(u|\nu) = (2\pi)^{-S/2} |\Gamma \Gamma'|^{-1/2} \exp\{-\frac{1}{2}(u - \nu)'(\Gamma \Gamma')^{-1}(u - \nu)\}.$$ 

Let $H = (\Gamma \Gamma')^{-1} U$, the density of $H$ is

$$f_H(h|\nu) = \alpha(h)\beta(\nu) \exp \sum_{i=1}^{S} h_i \nu_i.$$ 

The hypothesis of interest $\mu_1 = \mu_2$ vs. $\mu_1 < \mu_2$ is equivalent to test $H_1 : \nu_1 = 0$ vs. $K_1 : \nu_1 > 0$. Matthes and Truax (1967) provided that, for the exponential family, a necessary and sufficient condition for a test $\phi(h)$ of $H_1$ vs. $K_1$ to be admissible, is that for almost every fixed $h_2, \ldots, h_S$, the acceptance regions of the test are convex in $h_1$. 
Notice that $\mathbf{H} = (\Gamma')^{-1} \mathbf{U} = (\Gamma')^{-1} \Gamma \mathbf{X} = (\Gamma')^{-1} \mathbf{X}$. To study a test function $\phi(h) = \phi_U(x)$ as $h_1$ varies while $(h_2, h_3, \cdots, h_S)$ remain fixed, we can consider sample points $x + rg$, where $g$ is the first row of $\Gamma$ and $r$ varies. This is true since when $h$ is evaluated from $x$ to $x + rg$, we obtain $(\Gamma')^{-1}(x + rg) = h + (r, 0, \cdots, 0)'$. Thus, to evaluate the admissibility property of a MTP for a vector risk function in the matrix order problem, we will examine the acceptance regions of an individual test along the direction of $g$.

This chapter is organized as follows. Section 1 gives the conventional step-down procedure. We introduce the simple order constraint method in section 2. Section 3 provides a closed testing method for the matrix order problem. MRD is illustrated in section 4. Section 5 shows simulation results.

5.1 Step-down procedure

Ignoring the structure of the problem, a conventional step-down procedure is a convenient method for the matrix order problem. For an $R \times C$ matrix order problem, the step-down procedure considers individual $M = (R(C - 1) + C(R - 1))$ hypotheses. One orders the $M$ individual test statistics by their significances and continues to reject hypotheses as long as the test statistics are larger than some specific critical values. Although the application is easy and straightforward, we illustrate in the following theorem that the step-down testing procedure is inadmissible for a vector risk function. First, we show the case of a $2 \times 2$ matrix order problem and the idea is extended to a general $R \times C$ matrix order problem.

**Theorem 5.1.1.** The step-down procedure for a $2 \times 2$ matrix order problem is inadmissible for a vector risk function.

**Proof.** For the sake of convenience, denote a sample point $x = (x_1, x_2, x_3, x_4)$, and $x_1 = x_{11}$, $x_2 = x_{12}$, $x_3 = x_{21}$, $x_4 = x_{22}$ and $D_{ij} = (x_j - x_i)/\sqrt{2}$. Let $\mu_i$ be the mean for $X_i$ and letting $C_4 < C_3 < C_2 < C_1$ be the critical values for a step-down procedure. Consider 3 sample points, $x, x^*, x^{**}$, where $x^* = x + r_1g$, $x^{**} = x + r_2g$, $0 < r_1 < r_2$, and $g = (-1, 1, 0, 0)$. 
For a vector risk function and the symmetry of the test, we may focus on one individual test $\mu_1 = \mu_2$. Consider 
\[ x = (0, \sqrt{2}(C_3 - 3\epsilon), \sqrt{2}(C_3 - \epsilon), \sqrt{2}(C_1 + C_3 - \epsilon)), \]
where $\epsilon$ is some positive number and satisfies $C_3 + 3\epsilon < C_2$. We get $D_{24} = C_1 + 2\epsilon$, $D_{34} = C_1$, $D_{13} = C_3 - \epsilon$ and $D_{12} = C_3 - 3\epsilon$. The order of the above statistics is $D_{12} < D_{13} < D_{34} < D_{24}$. When applying the step-down procedure on $x$, one would reject $\mu_2 = \mu_4$, $\mu_3 = \mu_4$ and accept $\mu_1 = \mu_3$ and $\mu_1 = \mu_2$. Thus, $x$ is an accept point for $\mu_1 = \mu_2$.

Let $r_1 = \frac{7\sqrt{2}\epsilon}{4}$. Denote $D_{ij}^* = (x_j^* - x_i^*)/\sqrt{2}$. We obtain $D_{24}^* = D_{24} - \frac{r_1}{\sqrt{2}} = C_1 + \frac{1}{4}\epsilon$, $D_{34}^* = D_{34} = C_1$, $D_{13}^* = D_{13} + \frac{r_1}{\sqrt{2}} = C_3 + \frac{3}{4}\epsilon$, and $D_{12}^* = D_{12} + \sqrt{2}r_1 = C_3 + \frac{1}{2}\epsilon$. Then, the order of the four statistics is $D_{12}^* < D_{13}^* < D_{34}^* < D_{24}^*$. But one would reject all four hypotheses. That is, $x^*$ is a reject point for $\mu_1 = \mu_2$.

Now, let $r_2 = 3\sqrt{2}\epsilon$ and $D_{ij}^{**} = (x_j^{**} - x_i^{**})/\sqrt{2}$. In this case, $D_{24}^{**} = D_{24} - \frac{r_2}{\sqrt{2}} = C_1 - \epsilon$, $D_{34}^{**} = D_{34} = C_1$, $D_{13}^{**} = D_{13} + \frac{r_2}{\sqrt{2}} = C_3 + 2\epsilon$, and $D_{12}^{**} = D_{12} + \sqrt{2}r_2 = C_3 + 3\epsilon$. Recall that $C_3 + 3\epsilon < C_2$, and the order of four statistics is $D_{13}^{**} < D_{12}^{**} < D_{24}^{**} < D_{34}^{**}$. Since $D_{34}^{**} = C_1$, all four hypotheses will be accepted. Thus, $x^{**}$ is an accept point. The above example shows that the acceptance region for individual test $\mu_1 = \mu_2$ is not an interval. Thus, the step-down procedure in a $2 \times 2$ matrix order problem is inadmissible. \qed

**Theorem 5.1.2.** The step-down procedure for an $R \times C$ matrix order problem is inadmissible for a vector risk function.

*Proof.* Let $M = R(C - 1) + C(R - 1)$ be the number of individual tests in a $R \times C$ matrix order problem. Denote $C_1 > C_2 > C_3 > C_4 > \cdots > C_M$ be some critical values for the step-down procedure. Consider a sample point, $\{x_{ij}\}$, where $x_{11} = 0$, $x_{12} = \sqrt{2}(C_3 - 3\epsilon)$, $x_{21} = \sqrt{2}(C_3 - \epsilon)$, $x_{22} = \sqrt{2}(C_1 + C_3 - \epsilon)$, and the rest of elements equal to 0.

Essentially, it is the example used in the proof of Theorem 5.1.1 and now it is augmented into an $R \times C$ matrix. The additional comparison tests for this $R \times C$ matrix order would have test statistics being less than or equal to 0. Thus the proof for a $2 \times 2$ matrix remains valid in the $R \times C$ matrix order problem. \qed
5.2 Simple order constraints

An $R \times C$ partial matrix order problem can be formulated as for $i = 1, \cdots, R$,

$$H_i : \mu_{ij} = \mu_i, \forall j$$

$$K_i : \mu_{i,1} \leq \mu_{i,2} \leq \cdots \leq \mu_{i,c}, \text{with at least one strict inequality}$$

and for $j = 1, \cdots, C$,

$$H_j : \mu_{ij} = \mu_j, \forall i$$

$$K_j : \mu_{1,j} \leq \mu_{2,j} \leq \cdots \leq \mu_{R,j}, \text{with at least one strict inequality}$$

That is, the $R \times C$ matrix order problem is formulated as $R + C$ individual tests. And the parameter space under the alternative hypothesis for each of the individual test is formed by simple order. It is important to know that the above formulation is not unique. There are many other ways to reconstruct the general matrix order problem with the combination of simple order problems.

With the observation of connection between matrix order and simple order constraints, we develop a testing procedure for a matrix order problem and call it simple order constraint (SOC). SOC uses likelihood ratio tests for each simple order constraint problem. It proceeds as follows.

At the first step, decompose the matrix order restriction on $\{\mu_{ij}\}$ as $h$ simple order constraints.

$$H : \mu_{ij} = \mu, \forall (i,j)$$

$$K : \begin{cases} \mu_{1,1} \leq \mu_{1,2} \leq \cdots \\ \mu_{1,1} \leq \mu_{2,1} \leq \cdots \\ \vdots \end{cases}$$

with at least one strict inequality.

Compute the likelihood ratio test for each of the $h$ simple order constraints and select the most significant constraint. Let $\alpha_1$ be a constant and $p(LRT^{(1)})$ be the p-value of the likelihood ratio test statistic for the most significant constraint, $LRT^{(1)}$. If $p(LRT^{(1)}) \geq \alpha_1$, accept all nulls and stop. Otherwise, split the most significant constraint into two subconstraints at the largest deviation.
At step \( p \), if the procedure has not stopped, compute at most \( h + p - 1 \) likelihood ratio statistics and select the most significant one. Compare \( p(LRT^{(p)}) \) with \( \alpha_p \), a constant for the \( p \)th step. If \( p(LRT^{(p)}) \geq \alpha_p \), stop the procedure. Otherwise, split the most significant constraint into two subconstraints at the largest deviation.

Now take a \( 2 \times 2 \) matrix order problem as an example. The global test for a \( 2 \times 2 \) matrix order problem is

\[
H : \mu_{11} = \mu_{12} = \mu_{21} = \mu_{22}
\]

\[
K : \mu_{11} \leq \mu_{12}, \mu_{21} \leq \mu_{22}, \mu_{11} \leq \mu_{21}, \mu_{12} \leq \mu_{22},
\]

with at least one strict inequality.

For the sake of convenience, denote a sample point \( \mathbf{x} = (x_1, x_2, x_3, x_4) \), where \( x_1 = x_{11}, x_2 = x_{12}, x_3 = x_{21}, x_4 = x_{22} \) and \( \mu_i \) be the mean for \( X_i \). Now formulate the \( 2 \times 2 \) matrix order problem as the following multiple testing problem with 2 individual hypotheses.

\[
H_{124,134} : \mu_1 = \mu_2 = \mu_3 = \mu_4
\]

\[
K_{124,134} : \mu_1 \leq \mu_2 \leq \mu_4, \mu_1 \leq \mu_3 \leq \mu_4, \text{with at least one strict inequality.}
\]

Define the following hypotheses:

\[
H_{124} : \mu_1 = \mu_2 = \mu_4
\]

\[
K_{124} : \mu_1 \leq \mu_2 \leq \mu_4, \text{with at least one strict inequality.}
\]

\[
H_{134} : \mu_1 = \mu_3 = \mu_4
\]

\[
K_{134} : \mu_1 \leq \mu_3 \leq \mu_4, \text{with at least one strict inequality.}
\]

Let the global test be formulated as \( H_{124,134} \) vs. \( K_{124,134} \). Let \( LRT_{124} \) and \( LRT_{134} \) be the corresponding likelihood ratio statistics for \( H_{124} \) vs. \( K_{124} \) and \( H_{134} \) vs. \( K_{134} \), respectively. Letting \( p(LRT_{124}) \) and \( p(LRT_{134}) \) be the corresponding p-values for \( LRT_{124} \) and \( LRT_{134} \).

SOC in this example proceeds as follows:

Step 1: If \( \min(p(LRT_{134}), p(LRT_{124})) \geq \alpha_1 \), accept all nulls. Otherwise, reject one of the following and go to the next step.
1. \( \mu_1 = \mu_2 \) if \( p(\text{LRT}_{124}) < p(\text{LRT}_{134}) \) and \( x_2 - x_1 > x_4 - x_2 \)

2. \( \mu_2 = \mu_4 \) if \( p(\text{LRT}_{124}) < p(\text{LRT}_{134}) \) and \( x_4 - x_2 > x_2 - x_1 \)

3. \( \mu_1 = \mu_3 \) if \( p(\text{LRT}_{134}) < p(\text{LRT}_{124}) \) and \( x_4 - x_1 > x_4 - x_3 \)

4. \( \mu_3 = \mu_4 \) if \( p(\text{LRT}_{134}) < p(\text{LRT}_{124}) \) and \( x_4 - x_3 > x_3 - x_1 \)

Step 2: Consider the remaining simple order tests. If the most significant constraint has p-value less than the specific significance level, split the constraint into two subconstraints where the deviation is largest.

Step 3: Stop the procedure when no single constraints under consideration has p-value less than some significance level.

Let’s, for a moment, consider the likelihood ratio test in order restricted problem.

Textbook discussion of order restricted problems can be found in Robertson, Wright and Dykstra (1988), Barlow, Bartholomew, Bremner and Brunk (1972) and Silvapulle and Sen (2004). Let \( X \) be independent normally distributed random vector with mean \( \mu \) and variance-covariance matrix \( I \). The likelihood function for \( \mu \) given \( \mathbf{x} \) is

\[
f(\mu; \mathbf{x}) = c \exp\left\{-\frac{1}{2} (\mathbf{x} - \mu)'(\mathbf{x} - \mu)\right\},
\]

where \( c \) is a constant. Note that the kernel of the likelihood ratio test in order restricted problem can expressed as

\[
(\mathbf{x} - \bar{x})'(\mathbf{x} - \bar{x}) - (\mathbf{x} - \mu(\mathbf{x}))'(\mathbf{x} - \mu(\mathbf{x}))
\]

where \( \mu(\mathbf{x}) \) is a point in restricted order space that has the minimal distance to \( \mathbf{x} \). Since the inner product of \( \mathbf{x} - \bar{x} \) and \( \bar{x} - \mu(\mathbf{x}) \) is zero, we further simplify (5.1) as

\[
(\bar{x} - \mu(\mathbf{x}))'(\bar{x} - \mu(\mathbf{x})) = \|\bar{x} - \mu(\mathbf{x})\|^2
\]

Notice that (5.2) depends on \( \mu(\mathbf{x}) \), the projection of \( \mathbf{x} \) onto the restricted space.

Note that the parameter space under these order restrictions is a polyhedral cone, i.e., the intersection of some half-spaces. Based on the projection formula onto this polyhedral cone (the faces of the restricted space), for the aforementioned \( 2 \times 2 \) matrix order example, the whole sample space can be partitioned into four subsample spaces
by the maximum likelihood estimates under $K_{124} : \mu_1 \leq \mu_2 \leq \mu_4$. The results are listed below.

1. $F_1 = \{ \bar{x} : \mu(\bar{x}) = (x_1, x_2, x_4) \}$
2. $F_2 = \{ \bar{x} : \mu(\bar{x}) = (\bar{x}_{12}, \bar{x}_{12}, x_4) \}$
3. $F_3 = \{ \bar{x} : \mu(\bar{x}) = (x_1, \bar{x}_{24}, \bar{x}_{24}) \}$
4. $F_4 = \{ \bar{x} : \mu(\bar{x}) = (\bar{x}_{124}, \bar{x}_{124}, \bar{x}_{124}) \}$

The following lemma shows that the kernel of the likelihood ratio test statistic for $H_{124}$ vs. $K_{124}$ is monotonic in the direction $g = (-1, 1, 0, 0)$.

**Lemma 5.2.1.** Let $L_{124}$ be the kernel of the likelihood ratio test statistic for $H_{124}$ vs. $K_{124}$. For a given sample point $\bar{x}$, and any point $\bar{x}^* = \bar{x} - rg$, $\bar{x}^{**} = \bar{x} + rg$, where $g = (-1, 1, 0, 0)$, $L_{124}(\bar{x}^*) \leq L_{124}(\bar{x}) \leq L_{124}(\bar{x}^{**})$ if $r > 0$ and $L_{124}(\bar{x}^*) \geq L_{124}(\bar{x}) \geq L_{124}(\bar{x}^{**})$ if $r < 0$.

**Proof.**

**Case 1:** $\bar{x}, \bar{x}^*, \bar{x}^{**} \in F_1$

$$L_{124}(\bar{x}^*) = L_{124}(\bar{x}) + \| \mu(\bar{x}) - \mu(\bar{x}^*) \|^2 + 2 \langle \bar{x}_{124} - \mu(\bar{x}), \mu(\bar{x}) - \mu(\bar{x}^*) \rangle$$

$$= L_{124}(\bar{x}) + 2r^2 + 2 \langle \bar{x}_{124} - \mu(\bar{x}), (-r, r, 0) \rangle$$

$$= L_{124}(\bar{x}) + 2r^2 - 2r(x_2 - x_1)$$

$$= L_{124}(\bar{x}) + 2r(r - (x_2 - x_1)) < L_{124}(\bar{x})$$

The last inequality is true since $\bar{x}^* \in F_1$ implies $x_1 + r \leq x_2 - r$.

$$L_{124}(\bar{x}^{**}) = L_{124}(\bar{x}) + 2r^2 + 2 \langle \bar{x}_{124} - \mu(\bar{x}), (r, -r, 0) \rangle$$

$$= L_{124}(\bar{x}) + 2r^2 + 2r(x_2 - x_1) \geq L_{124}(\bar{x})$$

**Case 2:** $\bar{x}, \bar{x}^*, \bar{x}^{**} \in F_2, F_4$

Because $\mu(\bar{x}) = \mu(\bar{x}^*) = \mu(\bar{x}^{**})$, we get $L_{124}(\bar{x}) = L_{124}(\bar{x}^*) = L_{124}(\bar{x}^{**})$. 

Case 3: $\bar{x}, \bar{x}^*, \bar{x}^{**} \in F_3$

\[ L_{124}(\bar{x}^*) = L_{124}(\bar{x}) + \|\mu(\bar{x}) - \mu(\bar{x}^*)\|^2 + 2 \langle \bar{x}_{124} - \mu(\bar{x}), \mu(\bar{x}) - \mu(\bar{x}^*) \rangle \]

\[ = L_{124}(\bar{x}) + \frac{3}{2}r^2 + 2 \langle \bar{x}_{124} - \mu(\bar{x}), (r, r, r) \rangle \]

\[ = L_{124}(\bar{x}) + 2r \left( \frac{3}{4}r + x_1 - \bar{x}_{24} \right) < L_{124}(\bar{x}) \]

The last inequality is true since $\bar{x}^* \in F_3$ implies $x_1 + r \leq \bar{x}_{24} - \frac{r}{2}$. 

\[ L_{124}(\bar{x}^{**}) = L_{124}(\bar{x}) + \frac{3}{2}r^2 + 2 \langle \bar{x}_{124} - \mu(\bar{x}), (r, -r, -r) \rangle \]

\[ = L_{124}(\bar{x}) + 2r \left( \frac{3}{4}r + x_1 - \bar{x}_{24} \right) \geq L_{124}(\bar{x}) \]

Cases 1-3 enable us to conclude that, for a small value of $r$, $L_{124}(\bar{x}^*) \leq L_{124}(\bar{x}) \leq L_{124}(\bar{x}^{**})$ if $r > 0$. By symmetry, it can be shown that $L_{124}(\bar{x}^*) \geq L_{124}(\bar{x}) \geq L_{124}(\bar{x}^{**})$ if $r < 0$. Since $L_{124}$ is continuous, the conclusion remains valid for a large value of $r$, i.e., the whole sample space.

\[ \square \]

**Lemma 5.2.2.** Let $L_{12}$ be the kernel of the likelihood ratio test statistic for $H_{12}: \mu_1 = \mu_2$ vs. $K_{12}: \mu_1 < \mu_2$. For a given sample point $\bar{x}$, and any point on $\bar{x}^* = \bar{x} - rg$, $\bar{x}^{**} = \bar{x} + rg$, where $g = (-1, 1, 0, 0)$, then $L_{12}(\bar{x}^*) \leq L_{12}(\bar{x}) \leq L_{12}(\bar{x}^{**})$ if $r > 0$ and $L_{12}(\bar{x}^*) \geq L_{12}(\bar{x}) \geq L_{12}(\bar{x}^{**})$ if $r < 0$.

**Proof.** Same as the proof of Lemma 5.2.1. \[ \square \]

**Theorem 5.2.3.** The SOC procedure is admissible for a $2 \times 2$ matrix order problem with a vector risk function.

**Proof.** For the symmetry of the test, let’s focus on the test $H_{12}: \mu_1 = \mu_2$ vs. $K_{12}: \mu_1 < \mu_2$. To prove that the acceptance region of the hypothesis $\mu_1 = \mu_2$ is convex in the direction of $g = (-1, 1, 0, 0)$, we show that there does not exist three points $\bar{x}$, $\bar{x}^* = \bar{x} - rg$, $\bar{x}^{**} = \bar{x} + rg$, such that $\bar{x}$ is a reject point and $\bar{x}^*$ and $\bar{x}^{**}$ are accept points.

Assume $\bar{x}$ is a reject point and $\mu_1 = \mu_2$ is rejected at the $m$th step. That implies $p(LRT_{124}(\bar{x})) < \alpha_i, i \leq m$. And if $i = m, x_2 - x_1 > x_4 - x_2$. Otherwise, $x_2 - x_1 < x_4 - x_2$ and $p(LRT_{12}(\bar{x})) < \alpha_m$. 

\[ \square \]
Lemma 5.2.1 and 5.2.2 imply \( p(\text{LRT}_{124}(x^{**})) \leq p(\text{LRT}_{124}(\bar{x})) \), and \( p(\text{LRT}_{12}(x^{**})) \leq p(\text{LRT}_{12}(\bar{x})) \) for \( r > 0 \). In addition, \( r > 0 \) implies \( x^{**}_2 - x^{**}_4 > x_2 - x_1 \) and \( x^{**}_4 - x^{**}_2 \leq x_4 - x_2 \). Therefore, we see that, \( p(\text{LRT}_{124}(x^{**})) < \alpha_i \) and \( p(\text{LRT}_{12}(x^{**})) < \alpha_m \). In the case of \( i = m \), \( x^{**}_2 - x^{**}_4 > x^*_4 - x^*_2 \). Thus, \( \mu_1 = \mu_2 \) is also rejected before or at the \( m \)th step at \( x^{**} \). Consequently, \( x^{**} \) can’t be an accept point for \( \mu_1 = \mu_2 \). Likewise, it can be shown that \( x^* \) can’t be an accept point for \( \mu_1 = \mu_2 \) if \( r < 0 \). Thus, SOC is admissible for a \( 2 \times 2 \) matrix order problem for a vector risk function.

Now we consider the admissibility result of SOC in a \( R \times C \) matrix order problem.

Before that we use the following lemma as the generalization of Lemma 5.2.1.

**Lemma 5.2.4.** Let \( A \) be the set of \( \mu_i \)'s that forms a simple order constraint and contains both \( \mu_1 \) and \( \mu_2 \). Letting \( L_A(x) \) be the kernel of the likelihood ratio test statistic for \( H_A : \mu_i = \mu_j \forall i, j \in A \) vs. \( K_A : \mu_1 \leq \mu_2 \cdots \). Denote sample points \( x \), \( x^* = \bar{x} - rg \) and \( x^{**} = \bar{x} + rg \), where \( g = (-1, 1, 0, \cdots, 0) \). Then \( L_A(x^*) \leq L_A(x) \leq L_A(x^{**}) \) if \( r > 0 \) and \( L_A(x^*) \geq L_A(x) \geq L_A(x^{**}) \) if \( r < 0 \).

**Proof.** Consider a small \( r > 0 \) such that \( x \), \( x^* \) and \( x^{**} \) are all in the same face from the simple order constraint formed by the elements in \( A \). Note that the projection of \( x^* \) onto the face can be expressed as

\[
\mu(x^*) = (\mu_1(x^*), \cdots, \mu_S(x^*)) = (\mu_1(x) + \frac{r}{m}, \mu_2(x) - \frac{r}{n}, \cdots)
\]

Thus,

\[
L_A(x^*) = L_A(x) + \|\mu(x) - \mu(x^*)\|^2 + 2 \langle \bar{x}_A - \mu(x), \mu(x) - \mu(x^*) \rangle
\]

\[
= L_A(x) + \left( \frac{1}{m} + \frac{1}{n} \right) r^2 + 2 r \left( \bar{x}_A - \mu(x), \left( \frac{-1}{m}, \frac{1}{n}, \cdots \right) \right)
\]

\[
= L_A(x) + \left( \frac{1}{m} + \frac{1}{n} \right) r^2 + 2 r (\bar{x}_A - \mu_2(x) - (\bar{x}_A - \mu_1(x)))
\]

\[
= L_A(x) + \left( \frac{1}{m} + \frac{1}{n} \right) r^2 + 2 r (\mu_1(x) - \mu_2(x))
\]

\[
= L_A(x) + 2 r (\mu_1(x) - \mu_2(x) + \frac{r}{2m} + \frac{r}{2n}) \leq L_A(x)
\]
The last inequality is true since \( \mu_1(x) + \frac{\xi}{m} \leq \mu_2(x) - \frac{\xi}{n} \). Furthermore,

\[
L_A(x^{**}) = L_A(x) + \left( \frac{1}{m} + \frac{1}{n} \right)r^2 + 2r(\mu_2(x) - \mu_1(x)) \geq L_A(x).
\]

Thus, \( L_A(x) \leq L_A(x) \leq L_A(x^{**}) \) for small and nonnegative \( r \). Similarly, it can be shown that \( L_A(x) \geq L_A(x) \geq L_A(x^{**}) \) for small and nonpositive \( r \). Since \( L_A(x) \) is continuous, the monotone property remains valid for a large \( r \).

**Theorem 5.2.5.** SOC is admissible for a \( R \times C \) matrix order problem for a vector risk function.

**Proof.** Again we focus on test \( H_{12} : \mu_1 = \mu_2 \) vs. \( K_{12} : \mu_1 < \mu_2 \). We will show in the following that, there does not exist three points \( x, x^* = x - rg, x^{**} = x + rg \), where \( g = (-1, 1, 0, \cdots, 0) \), such that \( x \) is a reject point and \( x^* \) and \( x^{**} \) are accept points.

Assume \( x \) is a reject point for \( \mu_1 = \mu_2 \) and the procedure stops at the \( m \)th step. Let \( A_1 \subset A_2 \subset \cdots \subset A_\tau \) be the sets of \( \mu_i \)'s that form simple order constraints and contain both \( \mu_1 \) and \( \mu_2 \). This implies \( p(LRT_{A_i}(x)) < \alpha_j \), for \( i = 1, \cdots, \tau \) and \( j \leq m \). Lemma 5.2.4 states that, if \( r \geq 0 \), \( p(L_{A_i}(x^{**})) \leq p(L_{A_i}(x)) \) for all \( A_i \), \( i = 1, \cdots, \tau \). In addition, \( x^{**}_2 - x^*_1 \geq x_2 - x_1 \) for \( r \geq 0 \). Thus, \( x^{**} \) can’t be an accept point when \( r \geq 0 \). Likewise, \( x^* \) can’t be an accept point when \( r < 0 \). Consequently, SOC is admissible for a \( R \times C \) matrix order problem for a vector risk function.

**5.3 Closed testing method**

As previously mentioned in chapter 2, the closed testing method can be used to construct a method for a multiple testing problem. It starts with the formulation of a closure which contains all non-empty intersections of individual tests. And the method rejects a specific hypothesis if and only if those tests concerning the hypothesis within the closure are all rejected.

For example, consider a \( 2 \times 2 \) matrix order problem as defined in the previous section. The closed testing method would reject \( \mu_1 = \mu_2 \) if and only if the following tests are all rejected at level \( \alpha \).
$H_{12} : \mu_1 = \mu_2$

$K_{12} : \mu_1 < \mu_2$

$H_{124} : \mu_1 = \mu_2 = \mu_4$

$K_{124} : \mu_1 \leq \mu_2, \mu_2 \leq \mu_4$

$H_{12,134} : \mu_1 = \mu_2, \mu_1 = \mu_3 = \mu_4$

$K_{12,134} : \mu_1 \leq \mu_2, \mu_1 \leq \mu_3 \leq \mu_4$

$H_{12,134} : \mu_1 = \mu_2 = \mu_3 = \mu_4$

$K_{12,134} : \mu_1 \leq \mu_2 \leq \mu_4, \mu_1 \leq \mu_3 \leq \mu_4$

$H_{12,13} : \mu_1 = \mu_2 = \mu_3$

$K_{12,13} : \mu_1 \leq \mu_2, \mu_1 \leq \mu_3$

$H_{12,34} : \mu_1 = \mu_2, \mu_3 = \mu_4$

$K_{12,34} : \mu_1 \leq \mu_2, \mu_3 \leq \mu_4$

$H_{12,34} : \mu_1 = \mu_2 = \mu_4, \mu_3 = \mu_4$

$K_{12,34} : \mu_1 \leq \mu_2 \leq \mu_4, \mu_3 \leq \mu_4$

Now consider the kernel of the likelihood ratio test for $H_{12,134}$ vs. $K_{12,134}$. It is observed that the whole sample space can be partitioned into 10 sub sample spaces by the maximum likelihood estimates under $K_{12,134}$. Let $\bar{x}_l = \sum_{i \in l} x_i / |l|$. We list the results below.

1. $F_1 = \{ \underline{x} : \mu(\underline{x}) = (x_1, x_2, x_3, x_4) \}$

2. $F_2 = \{ \underline{x} : \mu(\underline{x}) = (\bar{x}_{12}, x_2, x_3, x_4) \}$

3. $F_3 = \{ \underline{x} : \mu(\underline{x}) = (\bar{x}_{13}, x_2, \bar{x}_{13}, x_4) \}$

4. $F_4 = \{ \underline{x} : \mu(\underline{x}) = (x_1, \bar{x}_{24}, x_3, \bar{x}_{24}) \}$

5. $F_5 = \{ \underline{x} : \mu(\underline{x}) = (x_1, x_2, \bar{x}_{34}, \bar{x}_{34}) \}$

6. $F_6 = \{ \underline{x} : \mu(\underline{x}) = (\bar{x}_{12}, \bar{x}_{12}, \bar{x}_{34}, \bar{x}_{34}) \}$

7. $F_7 = \{ \underline{x} : \mu(\underline{x}) = (\bar{x}_{13}, \bar{x}_{24}, \bar{x}_{13}, \bar{x}_{24}) \}$

8. $F_8 = \{ \underline{x} : \mu(\underline{x}) = (x_1, \bar{x}_{234}, \bar{x}_{234}, \bar{x}_{234}) \}$

9. $F_9 = \{ \underline{x} : \mu(\underline{x}) = (\bar{x}_{123}, \bar{x}_{123}, \bar{x}_{123}, x_4) \}$

10. $F_{10} = \{ \underline{x} : \mu(\underline{x}) = (\bar{x}, \bar{x}, \bar{x}, \bar{x}) \}$

The following lemma examines how the kernel of the likelihood ratio statistics for $LRT_{1234}$ changes along the direction of $g = (-1, 1, 0, 0)$, from a given sample point $\underline{x}$, in each sub sample space listed above.
Lemma 5.3.1. Let $L_{1234}(\bar{x})$ be the kernel of the likelihood ratio statistics for $H_{124,134}$ vs. $K_{124,134}$. For a given sample point $\bar{x}$, and any point $\bar{x}^* = \bar{x} - rg$, $\bar{x}^{**} = \bar{x} + rg$, where $g = (-1,1,0,0)$, $L_{1234}(\bar{x}^*) \leq L_{1234}(\bar{x}) \leq L_{1234}(\bar{x}^{**})$ if $r > 0$ and $L_{1234}(\bar{x}^*) \geq L_{1234}(\bar{x}) \geq L_{1234}(\bar{x}^{**})$ if $r < 0$.

Proof. For a small $r > 0$ such that $\bar{x}$, $\bar{x}^*$ and $\bar{x}^{**}$ are in the same subspace. Consider the following cases.

**Case 1:** $\bar{x}$, $\bar{x}^*$, $\bar{x}^{**} \in F_1$, $F_5$

$$L_{1234}(\bar{x}^*) = L_{1234}(\bar{x}) + \|\mu(\bar{x}) - \mu(\bar{x}^*)\|^2 + 2 \langle \bar{x} - \mu(\bar{x}), \mu(\bar{x}) - \mu(\bar{x}^*) \rangle$$

$$= L_{1234}(\bar{x}) + \|((x_1, x_2, x_3, x_4) - (x_1 + r, x_2 - r, x_3, x_4)\|^2$$

$$+ 2 \langle \bar{x} - \mu(\bar{x}), (-r, r, 0, 0) \rangle$$

$$= L_{1234}(\bar{x}) + 2r^2 - 2r(x_2 - x_1)$$

$$= L_{1234}(\bar{x}) + 2r(r - (x_2 - x_1)) < L_{1234}(\bar{x})$$

The last inequality is true since $\bar{x}^* \in F_1$ implies $x_1 + r < x_2 - r$.

$$L_{1234}(\bar{x}^{**}) = L_{1234}(\bar{x}) + \|\mu(\bar{x}) - \mu(\bar{x}^{**})\|^2 + 2 \langle \bar{x} - \mu(\bar{x}), \mu(\bar{x}) - \mu(\bar{x}^{**}) \rangle$$

$$= L_{1234}(\bar{x}) + \|((x_1, x_2, x_3, x_4) - (x_1 - r, x_2 + r, x_3, x_4)\|^2$$

$$+ 2 \langle \bar{x} - \mu(\bar{x}), (r, -r, 0, 0) \rangle$$

$$= L_{1234}(\bar{x}) + 2r^2 + 2r(x_2 - x_1)$$

$$= L_{1234}(\bar{x}) + 2r(r + x_2 - x_1) > L_{1234}(\bar{x})$$

When $\bar{x}$, $\bar{x}^*$ and $\bar{x}^{**} \in F_5$, the proof is similar.

**Case 2:** $\bar{x}$, $\bar{x}^*$, $\bar{x}^{**} \in F_2$, $F_6$, $F_9$, $F_{10}$

Because $\mu(\bar{x}) = \mu(\bar{x}^*) = \mu(\bar{x}^{**})$, we get $L_{1234}(\bar{x}) = L_{1234}(\bar{x}^*) = L_{1234}(\bar{x}^{**})$.

**Case 3:** $\bar{x}$, $\bar{x}^*$, $\bar{x}^{**} \in F_3$

$$L_{1234}(\bar{x}^*) = L_{1234}(\bar{x}) + \|((\bar{x}_{13}, x_2, \bar{x}_{13}, x_4) - (\bar{x}_{13} + r/2, x_2 - r, \bar{x}_{13} + r/2, x_4)\|^2$$

$$+ 2 \langle \bar{x} - \mu(\bar{x}), (-r/2, r, -r/2, 0) \rangle$$

$$= L_{1234}(\bar{x}) + \frac{3}{2}r^2 - 2r(x_2 - \bar{x}_{13}) < L_{1234}(\bar{x})$$
The last inequality is true since $x^* \in F_3$ implies $\bar{x}_{13} + r/2 < x_2 - r$.

$$L_{1234}(\mathcal{x}^{**}) = L_{1234}(\mathcal{x}) + \|(\bar{x}_{13}, x_2, \bar{x}_{13}, x_4) - (\bar{x}_{13} - r/2, x_2 + r, \bar{x}_{13} - r/2, x_4)\|^2$$

$$+ 2 \langle \bar{x} - \mu(\mathcal{x}), (r/2, -r, r/2, 0) \rangle$$

$$= L_{1234}(\mathcal{x}) + \frac{3}{2} r^2 + 2r(x_2 - \bar{x}_{13}) > L_{1234}(\mathcal{x})$$

**Case 4: $x, x^*, x^{**} \in F_4$**

$$L_{1234}(\mathcal{x}^*) = L_{1234}(\mathcal{x}) + \|(x_1, \bar{x}_{24}, x_3, \bar{x}_{24}) - (x_1 + r, \bar{x}_{24} - r/2, x_3, \bar{x}_{24} + r/2)\|^2$$

$$+ 2 \langle \bar{x} - \mu(\mathcal{x}), (-r, r/2, 0, r/2) \rangle$$

$$= L_{1234}(\mathcal{x}) + \frac{3}{2} r^2 - 2r(\bar{x}_{24} - x_1) < L_{1234}(\mathcal{x})$$

The last inequality is true since $x^* \in F_4$ implies $x_1 + r < \bar{x}_{24} - r/2$.

$$L_{1234}(\mathcal{x}^{**}) = L_{1234}(\mathcal{x}) + \|(x_1, \bar{x}_{24}, x_3, \bar{x}_{24}) - (x_1 - r, \bar{x}_{24} + r/2, x_3, \bar{x}_{24} + r/2)\|^2$$

$$+ 2 \langle \bar{x} - \mu(\mathcal{x}), (r, -r/2, 0, -r/2) \rangle$$

$$= L_{1234}(\mathcal{x}) + \frac{3}{2} r^2 + 2r(\bar{x}_{24} - x_1) > L_{1234}(\mathcal{x})$$

**Case 5: $x, x^*, x^{**} \in F_7$**

$$L_{1234}(\mathcal{x}^*) = L_{1234}(\mathcal{x}) + 2 \langle \bar{x} - \mu(\mathcal{x}), (-r/2, r/2, -r/2, r/2) \rangle$$

$$+ \|(\bar{x}_{13}, \bar{x}_{24}, \bar{x}_{13}, \bar{x}_{24}) - (\bar{x}_{13} + r/2, \bar{x}_{24} - r/2, \bar{x}_{13} + r/2, \bar{x}_{24} - r/2)\|^2$$

$$= L_{1234}(\mathcal{x}) + r^2 - 2r(\bar{x}_{24} - \bar{x}_{13}) < L_{1234}(\mathcal{x})$$

The last inequality is true since $x^* \in F_7$ implies $\bar{x}_{13} + r/2 < \bar{x}_{24} - r/2$.

$$L_{1234}(\mathcal{x}^{**}) = L_{1234}(\mathcal{x}) + 2 \langle \bar{x} - \mu(\mathcal{x}), (r/2, -r/2, +r/2, -r/2) \rangle$$

$$+ \|(\bar{x}_{13}, \bar{x}_{24}, \bar{x}_{13}, \bar{x}_{24}) - (\bar{x}_{13} - r/2, \bar{x}_{24} + r/2, \bar{x}_{13} - r/2, \bar{x}_{24} + r/2)\|^2$$

$$= L_{1234}(\mathcal{x}) + r^2 + 2r(\bar{x}_{24} - \bar{x}_{13}) > L_{1234}(\mathcal{x})$$

**Case 6: $x, x^*, x^{**} \in F_8$**

$$L_{1234}(\mathcal{x}^*) = L_{1234}(\mathcal{x}) + 2 \langle \bar{x} - \mu(\mathcal{x}), (-r, r/3, r/3, r/3) \rangle$$

$$+ \|(x_1, \bar{x}_{234}, \bar{x}_{234}, \bar{x}_{234}) - (x_1 + r, \bar{x}_{234} - r/3, \bar{x}_{234} - r/3, \bar{x}_{234} - r/3)\|^2$$

$$= L_{1234}(\mathcal{x}) + \frac{4}{3} r^2 - 2r(\bar{x}_{234} - x_1) < L_{1234}(\mathcal{x})$$
The last inequality is true since $\bar{x}^* \in F_\delta$ implies $x_1 + r < \bar{x}_{234} - r/3$.

\[ L_{1234}(\bar{x}^{**}) = L_{1234}(\bar{x}) + 2 \langle \bar{x} - \mu(\bar{x}), (r, -r/3, -r/3, -r/3) \rangle + \| (x_1, \bar{x}_{234}, \bar{x}_{234}, \bar{x}_{234}) - (x_1 - r, \bar{x}_{234} + r/3, \bar{x}_{234} + r/3, \bar{x}_{234} + r/3) \|^2 = L_{1234}(\bar{x}) + \frac{4}{3} r^2 + 2r(\bar{x}_{234} - x_1) > L_{1234}(\bar{x}) \]

Based on the results from cases 1-6, we conclude that, for a small value of $r$, $L_{1234}(\bar{x}^*) \leq L_{1234}(\bar{x}) \leq L_{1234}(\bar{x}^{**})$ if $r > 0$. By symmetry, it can be shown that $L_{1234}(\bar{x}^*) \geq L_{1234}(\bar{x}) \geq L_{1234}(\bar{x}^{**})$ if $r < 0$. Since $L_{1234}$ is continuous, the conclusion can be extended to a large value of $r$, i.e., the whole sample space.

Now consider using likelihood ratio test for each of the hypotheses within the closure. The following theorem shows that the closed testing method using likelihood ratio test statistics is admissible for the $2 \times 2$ matrix order problem.

**Theorem 5.3.2.** The closed testing method for a $2 \times 2$ matrix order problem is admissible for a vector risk function.

**Proof.** Let $\bar{x}$ and $\mu$ be defined as in Section 5.1. To prove that the acceptance region for the test $\mu_1 = \mu_2$ is convex in the direction of $g = (-1, 1, 0, \cdots, 0)$, we show that there does not exist three points $\bar{x}$, $\bar{x}^* = \bar{x} - rg$, $\bar{x}^{**} = \bar{x} + rg$ such that $\bar{x}$ is a reject point and $\bar{x}^*$ and $\bar{x}^{**}$ are accept points for $\mu_1 = \mu_2$.

For $\bar{x}$ to be a reject point for $\mu_1 = \mu_2$, it implies the p-value of the following test statistics, $LRT_{12}(\bar{x})$, $LRT_{12,13}(\bar{x})$, $LRT_{12,34}(\bar{x})$, $LRT_{12,134}(\bar{x})$, $LRT_{12,34}(\bar{x})$ and $LRT_{1234}(\bar{x})$ are all less than $\alpha$. Lemma 5.3.1 and Lemma 5.2.2 imply that the p-value for $LRT_{12,134}(\bar{x})$, $LRT_{124}(\bar{x})$ and $LRT_{12}(\bar{x})$ are monotone with respect to $g$. Thus the p-value of $LRT_{12}(\bar{x}^{**})$, $LRT_{12,13}(\bar{x}^{**})$, $LRT_{124}(\bar{x}^{**})$, $LRT_{12,34}(\bar{x}^{**})$, $LRT_{12,134}(\bar{x}^{**})$ and $LRT_{1234}(\bar{x}^{**})$ are all less than $\alpha$. Therefore, $\bar{x}^{**}$ can’t be an accept point for $\mu_1 = \mu_2$ when $r > 0$. On the other hand, $\bar{x}^*$ can’t be an accept point for $\mu_1 = \mu_2$ when $r < 0$. Consequently, $\bar{x}^*$ and $\bar{x}^{**}$ can’t be accept points for $\mu_1 = \mu_2$ at the same time and the closed testing method for a $2 \times 2$ matrix order problem is admissible for a vector risk function. □
Remark 5.3.1. To extend the proof of the admissibility to the general matrix order problem would require the likelihood ratio statistics to be monotone along the direction $\mathbf{g} = (-1,1,0,\cdots,0)$. The generalization of Lemma 5.3.1 would involve many cases for large dimensions.

5.4 MRD in matrix order

Cohen, Sackrowitz and Xu (2009) provided a multiple testing method, maximum residual down (MRD), for dependent cases. Originally, the method applies to a full rank case, where residuals exist. In a general matrix order problem, MRD can not be applied directly since the contrast matrix is singular.

Recall that MRD for the all-pairwise testing problem is essentially a binary segmentation procedure for one dimensional data. To extend the idea of binary segmentation and solve the singularity problem, we propose a revised version of MRD. The revised MRD procedure combines the idea of binary segmentation for the singular case (usually in early stages) and uses the original MRD when the remaining undetermined hypotheses form a full rank case.

More specifically, in the early steps, we consider two kinds of binary segmentations: orthogonal binary segmentation and flexible binary segmentation. They both divide samples under consideration into two sets, the lower and the upper sets. The orthogonal binary segmentation groups samples from the same row or column together. The flexible binary segmentation allows the flexibility of separating samples from the same row or column into two groups. That is, it accounts for the partial order restrictions on population means. Nevertheless, the allocation of lower set(L) and the upper set(U) does not exist the following case: $x_{ij} \in L$ and $x_{i'j} \in U$ for $i' < i$ or $x_{ij} \in L$ and $x_{ij'} \in U$ for $j' < j$.

Take a $2 \times 2$ matrix order problem as used in the previous section as an example. Let $l$ and $u$ be the set of indices of the samples in the lower and upper set, respectively. At the first step, the orthogonal binary segmentation considers two possible orthogonal partitions, \{l = (1, 2), u = (3, 4)\} and \{l = (1, 3), u = (2, 4)\}. The flexible binary
segmentation would consider four possible partitions, \( \{ l = (1), u = (2, 3, 4) \} \), \( \{ l = (1, 2), u = (3, 4) \} \), \( \{ l = (1, 3), u = (2, 4) \} \) and \( \{ l = (1, 2, 3), u = (4) \} \).

Note that we use singularity as a condition to determine if MRD should replace the orthogonal or flexible binary segmentation to provide all possible segmentations at a given step. When the relevant hypotheses for the set of samples which is nonsingular, consider the segmentation calculated from MRD. Here the singularity only refers to the subset that has two or more elements.

The modified MRD proceeds as follows. Denote the set of indices of samples from the original matrix by \( G \). At the first step, consider all possible segmentations for \( G \). Compute test statistics for each segmentation and its p-value. Let the p-value from the most significant segmentation be \( p_1 \). Compare \( p_1 \) with a constant, \( \alpha_1 \) for the first step. If \( p_1 \geq \alpha_1 \), then stop and accept all nulls. Otherwise, split the sample into two subgroups based on the most significant segmentation.

At the second step, let the collection of indices of samples from each subgroup be \( G_1 \) and \( G_2 \). For each of the subgroups, if the contrast matrix formed by the relevant hypotheses is singular, consider the binary segmentation rather than the segmentation based on MRD. Otherwise, use the segmentation based on MRD. The test statistics and p-value for each possible segmentation are calculated. Select the most significant segmentation across all subgroups and record the p-value from the most significant segmentation as \( p_2 \). Compare \( p_2 \) with a constant \( \alpha_2 \) for step 2. If \( p_2 \geq \alpha_2 \), then stop. If \( p_2 < \alpha_2 \), further split the subgroup with the most significant segmentation into two.

Continue the process until no subgroups need to be split. At the end, reject \( \mu_i = \mu_j \) if \( x_i \) and \( x_j \) lie in different subgroups.

Although MRD use z/t statistics to measure the significance of each segmentation, we consider using two statistics, likelihood ratio and z statistics, for the modified MRD.

5.4.1 MRD using likelihood ratio test

For the purpose of convenience, we denote \( K (= R \times C) \) as the number of populations from a \( R \times C \) matrix. Let \( \mu_1 \) and \( \mu_2 \) be the population means for \( X_{1,1} \) and \( X_{1,2} \). Letting \( \{ g_1, g_2 \} \) represent a partition or segmentation for \( G = \{ 1, 2, \cdots, K \} = g_1 \cup g_2 \), where
Let \( g_1 \) be a set containing the indices of samples in the lower set and \( g_2 \) be the one for the upper set.

Let \( f(\cdot) \) be the likelihood function. The likelihood ratio test statistics for \( H : \mu_i = \mu_j, \forall i,j \in G \) vs. \( K : \mu_i < \mu_j, i \in g_1, j \in g_2 \), \( \text{LRT}_{g_1,g_2}(x;G) \), is as follows.

\[
\text{LRT}_{g_1,g_2}(x;G) = \sup_{K: \mu_i \leq \mu_j} f(x, \mu) \sup_{H: \mu_i = \mu_j} f(x, \mu)
= \begin{cases}
\exp \frac{1}{2} \|x - \bar{x}_{g_1} - \bar{x}_{g_2}\|^2 & \text{if } \bar{x}_{g_1} \leq \bar{x}_{g_2} \\
\exp \frac{1}{2} \|x - \bar{x}_g\|^2 & \text{otherwise}
\end{cases}
\]

(5.3)

Here,

\[
\bar{x}_{g_1,g_2} = \bar{x}_{g_1} I_{g_1} + \bar{x}_{g_2} I_{g_2}
\]
\[
\bar{x}_{g_1} = \frac{1}{|g_1|} \sum_{i \in g_1} x_i
\]
\[
\bar{x}_{g_2} = \frac{1}{|g_2|} \sum_{i \in g_2} x_i
\]
\[
\bar{x}_g = \frac{1}{K} \sum_i x_i 1_K
\]

\( I_l \) is a \( 1 \times K \) vector with its \( i \)th element equals 1 if \( i \in l \) and 0 if not. \( 1_K \) is a \( 1 \times K \) vector with every element equals 1 and \( |g| \) is the number of elements in \( g \). (5.3) leads to the observation that the kernel of the likelihood ratio statistics is formulated as

\[
\|x - \bar{x}_g\|^2 - \|x - \bar{x}_{g_1,g_2}\|^2.
\]

For simplicity, denote the likelihood ratio statistics for \( H : \mu_i = \mu_j, \forall i,j \in G \) vs. \( K : \mu_i < \mu_j, i \in g_1, j \in g_2 \) as

\[
L_{g_1,g_2}(x;G) = \begin{cases}
\|x - \bar{x}_g\|^2 - \|x - \bar{x}_{g_1,g_2}\|^2 & \text{if } \bar{x}_{g_1} \leq \bar{x}_{g_2} \\
0 & \text{otherwise}
\end{cases}
\]

(5.4)

**Lemma 5.4.1.** Let \( x^* = x + r g \) and \( g = (-1, 1, 0, \cdots, 0) \). Letting \( \{g_1, g_2\} \) represent a partition or segmentation for \( G = \{1, 2, \cdots, K\} = g_1 \cup g_2 \), where \( g_1 \) be a set containing the indices of samples in the lower set and \( g_2 \) be the one for the upper set. Assume \( X_i \) are independent normally distributed random variables with mean \( \mu_i \) and variance 1.

For \( L_{g_1,g_2}(x;G) \) as defined in (5.4),

(1). If \( \{1, 2\} \in g_1 \), then \( L_{g_1,g_2}(x^*;G) = L_{g_1,g_2}(x;G) \).
(2). If \( \{1\} \in g_1, \{2\} \in g_2 \), then

\[
L_{g_1,g_2}(\bar{x}^*; G) = \begin{cases} 
L_{g_1,g_2}(\bar{x}; G) + r^2(1_{|g_1|} + 1_{|g_2|}) + 2r(\bar{x}_{g_1} - \bar{x}_g) & \text{if } \bar{x}_{g_1} \leq \bar{x}_g \\
\|x - \bar{x}_g\|^2 - \|x - \bar{x}_{g_1,g_2}\|^2 + r^2(1_{|g_1|} + 1_{|g_2|}) + 2r(\bar{x}_{g_2} - \bar{x}_g) & \text{if } \bar{x}_{g_1} > \bar{x}_g, \text{ and } r > \tilde{r} \\
0 & \text{if } r < \tilde{r}
\end{cases}
\]

where \( \tilde{r} \) is the minimal \( r \) such that \( \bar{x}_{g_1}^* \leq \bar{x}_{g_2}^* \) when \( \bar{x}_{g_1} > \bar{x}_{g_2} \).

Proof. For the first part, if \( \{1, 2\} \in g_1 \), then \( \bar{x}_g^* = \bar{x}_g \) and \( \bar{x}_{g_1,g_2}^* = \bar{x}_{g_1,g_2} \). Thus, the likelihood ratio statistics evaluated at \( \bar{x} \) or \( \bar{x}^* \) are the same.

For the second part, when \( \{1\} \in g_1, \{2\} \in g_2 \), we obtain

\[
\bar{x}_{g_1}^* = \bar{x}_{g_1} - \frac{r}{|g_1|} \\
\bar{x}_{g_2}^* = \bar{x}_{g_2} + \frac{r}{|g_2|}
\]

\[
L_{g_1,g_2}(\bar{x}^*; G) = \begin{cases} 
\|\bar{x}^* - \bar{x}_g\|^2 - \|\bar{x}^* - \bar{x}_{g_1,g_2}\|^2 & \text{if } \bar{x}_{g_1} \leq \bar{x}_{g_2} \\
0 & \text{otherwise}
\end{cases}
\]

And \( \|\bar{x}^* - \bar{x}_g\|^2 - \|\bar{x}^* - \bar{x}_{g_1,g_2}\|^2 \) can be simplified as follows.

\[
\|\bar{x}^* - \bar{x}_g\|^2 - \|\bar{x}^* - \bar{x}_{g_1,g_2}\|^2 = \|\bar{x} + rg - \bar{x}_g\|^2 - \|\bar{x} + rg - (\bar{x}_{g_1,g_2} + r(\frac{1}{|g_1|}, \frac{1}{|g_2|}, \cdots))\|^2 \\
= \|\bar{x} - \bar{x}_g\|^2 + 2r^2 + 2r(x_2 - x_1) \\
- \{\|\bar{x} - \bar{x}_{g_1,g_2}\|^2 + |r g - r(\frac{1}{|g_1|}, \frac{1}{|g_2|}, \cdots)|\}^2 \\
+ 2\left(\bar{x} - \bar{x}_{g_1,g_2}, rg - r(\frac{1}{|g_1|}, \frac{1}{|g_2|}, \cdots)\right) \\
= \|\bar{x} - \bar{x}_g\|^2 - \|\bar{x} - \bar{x}_{g_1,g_2}\|^2 + r^2(1_{|g_1|} + 1_{|g_2|}) + 2r(\bar{x}_{g_2} - \bar{x}_g)
\]

Therefore, \( L_{g_1,g_2}(\bar{x}^*; G) \) can be formulated as

\[
L_{g_1,g_2}(\bar{x}^*; G) = \begin{cases} 
L_{g_1,g_2}(\bar{x}; G) + r^2(1_{|g_1|} + 1_{|g_2|}) + 2r(\bar{x}_{g_2} - \bar{x}_g) & \text{if } \bar{x}_{g_1} \leq \bar{x}_g \\
\|x - \bar{x}_g\|^2 - \|x - \bar{x}_{g_1,g_2}\|^2 + r^2(1_{|g_1|} + 1_{|g_2|}) + 2r(\bar{x}_{g_2} - \bar{x}_g) & \text{if } \bar{x}_{g_1} > \bar{x}_g, r > \tilde{r} \\
0 & \text{if } \bar{x}_{g_1} > \bar{x}_g^*, r < \tilde{r}
\end{cases}
\]

where \( \tilde{r} \) is the minimal \( r \) such that \( \bar{x}_{g_1}^* \leq \bar{x}_{g_2}^* \) when \( \bar{x}_{g_1} > \bar{x}_{g_2} \).
Theorem 5.4.2. The modified MRD using likelihood ratio test for the matrix order problem is admissible for a vector risk function.

Proof. Assume \( \bar{x} \) be an accept point for \( \mu_1 = \mu_2 \) and the modified MRD stops at the \( j \)th step. Let \( p_i \) be the corresponding p-values of the likelihood ratio test and \( \alpha_i \) is a constant for the \( i \)th step.

At the first step, all possible segmentations can be classified into two categories. One group \( x_1 \) and \( x_2 \) into the same subgroup and the other doesn’t. Let the first one be category A and the second one be category B. Letting \( L_{g_1^A,g_2^A}(\bar{x};G) \) and \( L_{g_1^B,g_2^B}(\bar{x};G) \) be the most significant likelihood ratio statistics from the category A and B, respectively. Since \( \bar{x} \) is an accept point, it implies the p-value of \( L_{g_1^B,g_2^B}(\bar{x};G) \), \( p(L_{g_1^B,g_2^B}(\bar{x};G)) \), is greater or equal to \( \alpha_1 \). Assume the modified MRD splits samples into two subsets, \( G_1^{(2)} \) and \( G_2^{(2)} \). Let \( G_1^{(2)} \) be the one containing both \( x_1 \) and \( x_2 \).

At the second step, consider possible segmentations for \( G_1^{(2)} \) and \( G_2^{(2)} \) separately. For \( G_1^{(2)} \), there are two kinds of splitting, either splitting \( x_1 \) and \( x_2 \) into different groups or not. Let \( L_{g_1^A,g_2^A}(\bar{x};G_1^{(2)}) \) and \( L_{g_1^B,g_2^B}(\bar{x};G_1^{(2)}) \) be the most significant likelihood ratio statistics from the category A and B for \( G_1^{(2)} \). Letting \( L_{g_1^A,g_2^A}(\bar{x};G_2^{(2)}) \) be the most significant likelihood ratio statistics for \( G_2^{(2)} \). Note that there is no need to distinguish segmentation types for \( G_2^{(2)} \) since the segmentation no longer relates to \( x_1 \) and \( x_2 \). Again, the fact that \( \bar{x} \) being an accept point for \( \mu_1 = \mu_2 \) implies \( p(L_{g_1^B,g_2^B}(\bar{x};G_1^{(2)})) \geq \alpha_2 \).

Generally speaking, let \( G_1^{(i)} \) be the subgroups that contains both \( x_1 \) and \( x_2 \) at the \( i \)th step. For \( \bar{x} \) to be an accept point and the procedure stops at the \( j \)th step, it requires \( p(L_{g_1^B,g_2^B}(\bar{x});G_1^{(i)}) \geq \alpha_i \) for \( i \leq j \).

Now we consider the rejection region along the direction \( g = (-1, 1, 0, \cdots, 0) \) from \( \bar{x} \). Let \( \bar{x}_* = \bar{x} + r_\bar{g} \). At the first step, the choices of segmentation for \( \bar{x}_* \) are the same as \( \bar{x} \).

If \( j = 1 \), it implies \( p(L_{g_1^A,g_2^A}(\bar{x}_*);G) \geq \alpha_1 \). If \( j > 1 \), then \( p(L_{g_1^A,g_2^A}(\bar{x}_*);G) < \alpha_1 \). If \( \bar{x}_* \) is a reject point and reject \( \mu_1 = \mu_2 \) at step 1, the only possible way is \( p(L_{g_1^B,g_2^B}(\bar{x}_*);G) < \alpha_1 \). From Lemma 5.4.1, it implies \( |r| \geq r_1 \) for some \( r_1 \). Otherwise, the modified MRD will not separate \( x_1^* \) and \( x_2^* \) at the first step and the split of \( \bar{x}_* \) at the first step is the
same as that of $\bar{x}$.

At the second step, consider possible segmentations for $G_1^{(2)}$ and $G_2^{(2)}$. Since $G_2^{(2)}$ does not contain $x_1$ and $x_2$, all likelihood ratio statistics have the same values evaluated at $\bar{x}$ and $\bar{x}^*$. That is, $L_{g_1,g_2}(\bar{x}; G_2^{(2)}) = L_{g_1,g_2}(\bar{x}^*; G_2^{(2)})$ for all $g_1 \cup g_2 = G_2^{(2)}$. Now treat $G_2^{(2)}$ as $G$ was and use the same argument. It leads to the conclusion that, the modified MRD would split $x_1$ and $x_2$ at the second step, if and only if $|r| \geq r_2$ for some $r_2$. Otherwise, the segmentation for $\bar{x}^*$ is the same for $\bar{x}$ at the second step. The argument can continue to the $j$th step. If $\bar{x}^*$ is a reject point and split $\mu_1 = \mu_2$ at the $j$th step, it means $|r| \geq r_j$. Otherwise, the modified MRD would stop at the $j$th step.

To sum up, for $\bar{x}^*$ to be a reject point, it implies $|r| \geq \min_{s \leq j} r_s$. Thus, the accept region for $\mu_1 = \mu_2$ along with the direction $g = (-1, 1, 0, \cdots, 0)$ from $\bar{x}$ is $|r| < \min_{s \leq j} r_s$, an interval. Therefore, the modified MRD is admissible for the matrix order problem for a vector risk function.

**5.4.2 MRD using $z$ test**

Again, consider $K(= R \times C)$ means from a $R \times C$ matrix. Let $\{g_1, g_2\}$ be a possible binary segmentation for $G = \{1, 2, \cdots, K\}$. And $g_1$ is the index set for the samples in the lower set and $g_2$ is the one for the upper set. Denote the z-statistics for $H : \mu_i = \mu_j, \forall i, j \in G$ vs. $K : \mu_i < \mu_j, i \in g_1, j \in g_2$ as

\[
D_{g_1,g_2}(\bar{x}; G) = \frac{\bar{x}_{g_2} - \bar{x}_{g_1}}{\sqrt{\frac{1}{|g_1|} + \frac{1}{|g_2|}}} \tag{5.5}
\]

**Lemma 5.4.3.** Define $D_{g_1,g_2}(\bar{x}; G)$ as in (5.5). Let $\bar{x}^* = \bar{x} + rg$ and $g = (-1, 1, 0, \cdots, 0)$.

1. If $\{1, 2\} \in g_1$, then $D_{g_1,g_2}(\bar{x}^*; G) = D_{g_1,g_2}(\bar{x}; G)$.

2. If $\{1\} \in g_1$ and $\{2\} \in g_2$, then $D_{g_1,g_2}(\bar{x}^*; G) = D_{g_1,g_2}(\bar{x}; G) + r \sqrt{\frac{1}{|g_1|} + \frac{1}{|g_2|}}$.

**Proof.** For the first part, note that $\bar{x}_{g_1}^* = \bar{x}_g$ and $\bar{x}_{g_2}^* = \bar{x}_{g_2}$ when $\{1, 2\} \in g_1$. Thus,

\[
D_{g_1,g_2}(\bar{x}^*; G) = \frac{\bar{x}_{g_2}^* - \bar{x}_{g_1}}{\sqrt{\frac{1}{|g_1|} + \frac{1}{|g_2|}}} = D_{g_1,g_2}(\bar{x}; G).
\]
For the second part, when \( \{1\} \in g_1 \) and \( \{2\} \in g_2 \),

\[
D_{g_1,g_2}(\bar{x}^*; G) = \frac{\bar{x}_{g_2} + \frac{r}{|g_2|} - (\bar{x}_{g_1} - \frac{r}{|g_1|})}{\sqrt{\frac{1}{|g_1|} + \frac{1}{|g_2|}}} = D_{g_1,g_2}(\bar{x}; G) + r \sqrt{\frac{1}{|g_1|} + \frac{1}{|g_2|}}.
\]

**Theorem 5.4.4.** The modified MRD using z test for the matrix order problem is admissible for a vector risk function.

**Proof.** Let \( \bar{x} \) be an accept point and stop at the \( j \)th step. Now consider \( \bar{x}^* = \bar{x} + rg \) and \( g = (-1, 1, 0, \cdots, 0) \).

At the first step, there are the same segmentation choices for \( \bar{x}^* \) as for \( \bar{x} \). Lemma 5.4.3 implies, the modified MRD using z statistics either has the same segmentation for \( \bar{x}^* \) as for \( \bar{x} \), or reject \( \mu_1 = \mu_2 \) if \( r \geq r_1 \) for some \( r_1 \). If the modified MRD using z statistics does not split \( x^*_1 \) and \( x^*_2 \) at the first step and continue to the second step, it would either follow the same segmentation as \( \bar{x} \) has, or split \( x^*_1 \) and \( x^*_2 \) if \( r \geq r_2 \) for some \( r_2 \). Following the same argument, it is concluded that, \( \bar{x}^* \) is a reject point for \( \mu_1 = \mu_2 \) if and only if \( r \geq \min_{i \leq j} r_j \). Thus, the acceptance region is an interval. Consequently, it is an admissible procedure for a vector risk function.

**Remark 5.4.1.** The proof of the admissibility for the modified MRD is independent of early segmentation methods. Therefore, both modified MRD with the orthogonal and flexible binary segmentations are admissible.

**Remark 5.4.2.** In the general case where \( n > 1 \) or unknown variance, the admissibility results remain valid. Since sample variance \( s^2 = \sum_{i=1}^{k} \sum_{j=1}^{n}(x_{ij} - \bar{x}_i)^2 / k(n - 1) \) can replace \( \sigma^2 \). The z distribution would be replaced by the t-distribution.

**Remark 5.4.3.** Although the modified MRD with the orthogonal binary segmentation for the matrix order problem is admissible, it can be shown it is not consistent. We illustrate the result in the following theorem.

**Theorem 5.4.5.** The modified MRD with the orthogonal binary segmentation for the matrix order problem is not consistent.
Proof. Consider a $2 \times 2$ matrix order problem as used in the previous section. If $\mu_1 = \mu_2 = \mu_3 < \mu_4$, then $\Pr(LRT_{12,34} \geq LRT_{13,24}) = 1/2$ whatever the sample size is. Thus, the probability of making mistakes is positive even as sample size goes to infinity.

5.5 Simulation results

The simulation study illustrates a $2 \times 3$ matrix order problem. 232 sets of parameter settings are considered (not listed here). For each parameter configuration, FDR and power is estimated from 250 replicates. The screening value $P_u = 0.06$ is set to have at least 95% of the set of parameter configurations with FDR less than or equal to 0.05. As space is limited, we show partial results in Table 5.1 and Table 5.2 and use Figure 5.1 and Figure 5.2 to illustrate full results.

Table 5.1 lists partial results for the comparison of average power using likelihood ratio test statistics among three methods. And Table 5.2 lists those using $z$ statistics. It is observed that the average power of the step-down procedure would be mildly greater than the modified MRD for some cases where the number of false nulls is small. On the other hand, when the number of false nulls is large, the modified MRD would dominate the step-down procedure for most of the case.

Figure 5.1–5.2 shows comparison of the average power among the modified MRD with the orthogonal binary segmentation, the modified MRD with the flexible binary segmentation and the step-down procedure using likelihood ratio test and $z$ statistics, respectively. Each point on the diagram represents a combination of average power for one set of parameter configuration. The 45 degree line is used as a reference line. A point under the 45 degree line illustrates its x-value is bigger than the y-value.

The results from the figure 5.1 show that, two modified MRD procedures have higher power than the step-down procedure most of the time. As a matter of fact, only around 10% of the parameter settings that the step-down procedure has higher power than the modified MRD with the orthogonal binary segmentation, and less than 1% the parameter settings that the step-down procedure has higher power than the modified
MRD with the flexible binary segmentation. In addition, the modified MRD with the flexible binary segmentation performs better than the one with the orthogonal binary segmentation. The result comes no surprise since the former accounts for more possible segmentation than the latter does.

Figure 5.2 shows the simulated power of three methods using z statistics. Like the results using likelihood ratio test statistics, the modified MRD with the flexible binary segmentation dominates the other two methods. Around 9% of the parameter settings that the step-down procedure has higher power than the modified MRD with the orthogonal binary segmentation, and less than 2% the parameter settings that the step-down procedure has higher power than the modified MRD with the flexible binary segmentation.

As mentioned previously, the original MRD use z/t statistics only. Here we depict the segmentation idea from MRD and consider using likelihood ratio statistics. One interesting finding is that, all three test procedures have higher average power when using z statistics in most of the parameter settings.
### Table 5.1: Comparison of the modified MRD and step-down procedures using likelihood ratio test statistics for a $2 \times 3$ matrix order problem

<table>
<thead>
<tr>
<th>Population means</th>
<th>#of false nulls</th>
<th>FDR</th>
<th>Power</th>
<th>Power increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{11}$</td>
<td>$\mu_{12}$</td>
<td>$\mu_{13}$</td>
<td>$\mu_{21}$</td>
<td>$\mu_{22}$</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>1.2</td>
</tr>
<tr>
<td>0.0</td>
<td>2.6</td>
<td>2.6</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>1.2</td>
</tr>
<tr>
<td>0.0</td>
<td>1.2</td>
<td>1.2</td>
<td>0.8</td>
<td>1.2</td>
</tr>
<tr>
<td>0.8</td>
<td>1.2</td>
<td>2.6</td>
<td>0.8</td>
<td>1.2</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>1.2</td>
<td>0.0</td>
<td>0.8</td>
</tr>
<tr>
<td>0.8</td>
<td>1.2</td>
<td>2.6</td>
<td>0.8</td>
<td>2.6</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.8</td>
</tr>
</tbody>
</table>
| *SD, Ort and Flex in the table stand for the step-down, the modified MRD with orthogonal binary segmentation and the modified MRD with flexible binary segmentation.*
Table 5.2: Comparison of the modified MRD and step-down procedures using z statistics for a $2 \times 3$ matrix order problem

<table>
<thead>
<tr>
<th>Population means</th>
<th>#of false nulls</th>
<th>FDR SD</th>
<th>Ort</th>
<th>Flex</th>
<th>Power SD</th>
<th>Ort</th>
<th>Flex</th>
<th>Power increase SD</th>
<th>Ort</th>
<th>Flex</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{11}$</td>
<td>$\mu_{12}$</td>
<td>$\mu_{13}$</td>
<td>$\mu_{21}$</td>
<td>$\mu_{22}$</td>
<td>$\mu_{23}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.04</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>0.0</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>2</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.03</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>2</td>
<td>0.03</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>1.2</td>
<td>1.2</td>
<td>1.2</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>0.0</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>2</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.8</td>
<td>2.6</td>
<td>4</td>
<td>0.01</td>
<td>0.05</td>
<td>0.04</td>
<td>0.09</td>
</tr>
<tr>
<td>0.0</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>2.6</td>
<td>4</td>
<td>0.01</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>2</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.04</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>0.0</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>2</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>2</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.8</td>
<td>2.6</td>
<td>4</td>
<td>0.01</td>
<td>0.05</td>
<td>0.04</td>
<td>0.09</td>
</tr>
<tr>
<td>0.0</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>2.6</td>
<td>4</td>
<td>0.01</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.04</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

*SD, Ort and Flex in the table stand for the step-down, the modified MRD with orthogonal binary segmentation and the modified MRD with flexible binary segmentation.*
Figure 5.1: Simulated average power using likelihood ratio test statistics for a $2 \times 3$ matrix order problem.
Figure 5.2: Simulated average power using z statistics for a $2 \times 3$ matrix order problem.
Bibliography


Vita

Kuo-mei Chen

2005-2011  Ph. D. in Statistics, Rutgers University
2000-2002  M. S. in Statistics from Purdue University
1995-1997  M. A. in Economics from National Taiwan University
1991-1995  B. A. in Economics from National Taiwan University

2008-2010  Graduate assistant, Department of Genetics, Rutgers University
2006-2008  Teaching assistant, Department of Statistics, Rutgers University