Testing For and Against an Order Restriction Among Bivariate Populations

by

Yu Yang

A thesis submitted to the Faculty of Graduate Studies and Research

in partial fulfillment of the requirements for the degree of

Master of Science

Ottawa-Carleton Institute for Mathematics and Statistics

Carleton University

1125 Colonel By Drive, Ottawa, Ontario

Canada K1S 5B6

©Yu Yang, 2009

Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-64427-0
Our file Notre référence
ISBN: 978-0-494-64427-0

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

■ ■ ■
Canada

Abstract

In this thesis, we study order restricted statistical inference among bivariate populations. We first use the concept of a latent variable to derive a joint distribution model for a bivariate response with one binary and the other continuous. A likelihood ratio test is proposed to detect a possible ordering among several populations. We also develop likelihood ratio tests with one for testing the goodness-of-fit of the ordering assumption on both responses and the other for homogeneity against order restriction. We then apply this model to a data set on a toxicity study involving laboratory animals and investigate the effect of a food color additive on the development of reticuloendothelial (RE) tumors as well as their life times.

Acknowledgements

I wish to express my sincere gratitude to my supervisor, Dr. Chul Gyu Park, for his guidance, encouragement, patience, and support throughout the preparation of this thesis. His many useful suggestions and comments have been received with great appreciation.

I wish to dedicate this thesis to my beloved grandma and mother, Wei Zhao and Xia Zhao. It is their love, support and encouragement accompanying me in walking through each step of my life. None of this would be possible without their love and sacrifice.

Contents

Abstract	ii
Acknowledgements	iii
1 Introduction	1
1.1 Motivation	1
1.2 Overview of the thesis	3
2 Background	5
2.1 Isotonic Regression and Linear Inequality Constraints	5
2.1.1 Convex cones and orthogonal projections	5
2.1.2 Isotonic regression	7
2.1.3 Linear inequality constraints	9
2.2 Tests of Constrained Hypotheses	10
2.2.1 The one-way layout with ordered parameters	11
2.2.2 Testing one-sided hypothesis	14
2.2.3 Testing hypothesis of linear inequality	17
3 Inference for Ordered Groups	20

3.1	Model and Related Hypotheses	20
3.1.1	The model	20
3.1.2	Hypotheses for joint analysis	22
3.2	Maximum Likelihood Estimation	23
3.2.1	Unrestricted estimates	24
3.2.2	Constrained estimation	28
3.2.3	Least squares estimation under linear inequality constraints	31
3.3	Asymptotic Properties of the Estimators	35
3.4	Likelihood Ratio Tests	37
4	Application to RE Tumor Data	41
4.1	The RE Tumor Data	41
4.2	Model and Maximum Likelihood Estimates	43
4.3	Test Statistics	44
5	Conclusion and Generalizations	47
6	Appendix	49

List of Figures

3.1	Schematic Drawing of Dykstra's Algorithm	33
-----	--	----

List of Tables

4.1	Summary Statistics for Red Dye No.40 Data	42
4.2	Parameter Estimates under Each Hypothesis.	44
4.3	Estimates of p_i under Each Hypothesis.	45
4.4	Simulation Results for Weight Vectors	45
4.5	Test Statistics and Associated p-value	46

Chapter 1

Introduction

1.1 Motivation

Many types of problems in real world situations involve the identification of ordered structures. Inequality constraints on parameters are encountered frequently in areas such as clinical trials, biostatistics and econometrics. For example, the probability of a particular response may increase with the treatment level. Such constraints contain statistical information, and so, if properly incorporated, they would be more efficient than their counterparts where such constraints are ignored. Statistical methods for incorporating such constraints have advanced significantly since the early 1950s. Bartholomew [4] proposed a likelihood ratio test (LRT) for normal distribution means. He tested equality of normal means against a simple linear ordering among them. Later in 1963, Kudô [13] considered a testing problem for one-sided hypothesis on multivariate normal mean vector. Shapiro (1988)[24] proposed a more general testing problem with linear inequality on multivariate

normal means. General theories on order restricted inference are well discussed in the book by Robertson et al.(1988) [22]. Sen and Silvapulle [25] published a book in 2004, which included more up-to-date accomplishments in this area.

On the other hand, correlated binary observations arise in many investigations in biostatistics and psychology experiments. In particular, jointly modeling with one discrete and one continuous variables has gained a lot of attention. However, the lack of a natural multivariate distribution has created difficulty in joint modeling bivariate data comprising a continuous and a binary response [1]. There is a substantial literature focused on developing appropriate joint models for such bivariate responses. Tate (1954)[26] studied the simplest such model, when x has a binomial distribution, and the conditional distribution of y for fixed x is normal, assuming there is a correlation between these two variables. A multivariate extension model was considered by Olkin and Tate (1961)[18]. They proposed the general “location model”. This model is based on a multinomial model for the discrete outcomes and a multivariate Gaussian model for the continuous outcomes conditional on the discrete outcomes. Fitzmaurice and Laird (1995)[12] discussed marginal regression models, in which the marginal expectation of the bivariate response vector is related to a set of covariates by some known link functions. Cox and Wermuth (1992)[7] compared a number of models for the joint distribution of quantitative and binary response variables. Catalano and Ryan (1992)[5] introduced the concept of a continuous latent variable underlying the discrete response. They assumed the latent variable and the continuous response have a joint Gaussian distribution. They later extended the model to clustered data. Their model allows the joint distribution to be written as a product of a standard

random effect model for the continuous variable and a correlated probit model for the discrete variable. Later, Regan and Catalano (1999)[20] proposed a likelihood-based model that is an extension of a correlated probit model of Ochi and Prentice (1984)[17] to incorporate continuous outcomes. More recently, Agresti and Gueorguieva (2001)[1] proposed a Monte Carlo EM algorithm for finding maximum likelihood estimates of the correlated probit model.

The studies mentioned above focus on the relationship between the mean response and covariates. However, in a data set consisting of several group, these groups might be ordered in some way (for example, by treatment level). To date, very little attention has been paid to an ordering problem involves groups of bivariate dichotomous and continuous outcomes. Various examples of this type can be found in clinical trials, econometrics, psychology and many other areas. Farrell and Park (2007)[11] proposed a likelihood ratio test for the detection of an ordered group effect on bivariate responses based on a conditional logistic model. However, this thesis aims to develop a testing procedure to detect such an ordering effect in a joint model constructed using the concept of latent variables.

1.2 Overview of the thesis

This thesis is organized as follows: Chapter 2 gives some background information and some preliminary results about constrained statistical inference. In Chapter 3, we first describe a joint model for bivariate responses with one being binary and the other continuous. Then we discuss constrained estimation and testing problems under the joint model. In Chapter 4, we apply the proposed testing

methods to the RE tumor data, which are originally discussed in Lagakos and Mosteller (1981) [14]. Finally, we discuss possible generalization of the proposed model and future directions of order restricted statistical inference. Chapter 6 contains R programs of the algorithms presented in this thesis.

Chapter 2

Background

This chapter provides the background information that is relevant to the research on order restricted statistical inference, and it is organized as follows.

Section 2.1 briefly describes convex cones and orthogonal projections and defines isotonic regression which utilize prior ordering information when finding maximum likelihood estimator. Section 2.2 is an overview of some well-known results in inequality constrained testing problems.

2.1 Isotonic Regression and Linear Inequality Constraints

2.1.1 Convex cones and orthogonal projections

A subset K of $\mathbf{R}^n$ is called a convex cone if $\mathbf{x}, \mathbf{y} \in K$ implies that $a\mathbf{x} + b\mathbf{y} \in K$ for every positive scalars a and b . Let $\mathbf{w}(w_i > 0)$ be a fixed point in $\mathbf{R}^n$, we denote

the inner product of $\mathbf{x}$ and $\mathbf{y}$, with weight vector $\mathbf{w}$, as

$$(\mathbf{x}, \mathbf{y})_w = \sum_{i=1}^n x_i y_i w_i$$

So the norm of $\mathbf{x}$ associated with this inner product is

$$\|\mathbf{x}\|_w = (\mathbf{x}, \mathbf{x})_w^{1/2} = \left(\sum_{i=1}^n x_i^2 w_i \right)^{1/2}$$

The dual cone associated with K is given by

$$K^* = \{\mathbf{y}; (\mathbf{y}, \mathbf{x})_w = \sum_{i=1}^n y_i x_i w_i \leq 0, \text{ for all } \mathbf{x} \in K\}$$

Similarly to the case of linear spaces, if K is closed and convex, then K^* is also a closed convex cone with the property $K^{**} = K$.

Let $P_w(\mathbf{x}|K)$ denote the orthogonal projection of $\mathbf{x}$ onto K , which assigns to each $\mathbf{x}$ the closest point in K . The corresponding distance from $\mathbf{x}$ to K is given by $\|\mathbf{x} - P_w(\mathbf{x}|K)\|_w$. For all $\mathbf{x}$, $\mathbf{x} - P_w(\mathbf{x}|K) = P_w(\mathbf{x}|K^*)$. Similarly to Pythagorean theorem, we have

$$\|\mathbf{x}\|_w^2 = \|P_w(\mathbf{x}|K)\|_w^2 + \|\mathbf{x} - P_w(\mathbf{x}|K)\|_w^2. \quad (2.1)$$

In applications, K is often a polyhedral cone, that is

$$K = \{\mathbf{x} \in \mathbf{R}^k : A\mathbf{x} \leq 0\} \quad (2.2)$$

for some $r \times k$ matrix A .

2.1.2 Isotonic regression

The word *isotonic* means having the same tone. Isotonic regression is concerned with finding a regression function satisfying various order or shape restrictions and is obtained in a different way depending on the underlying order restrictions.

Let X be the finite set $\{x_1, x_2, \dots, x_k\}$. Some partial orders of interest defined on X are the simple order $(x_1 \preceq x_2 \preceq \dots \preceq x_k)$, the simple tree order $(x_1 \preceq x_i, i = 2, \dots, k)$, the simple loop order $(x_1 \preceq x_i \preceq x_k, i = 2, \dots, k-1)$, and the unimodal order $(x_1 \preceq x_2 \preceq \dots \preceq x_h \succeq x_{h+1} \succeq x_{h+2} \succeq \dots \succeq x_k)$.

A real valued function, f , on X is said to be isotonic with respect to the partial ordering $\preceq$ on X if $f(x) \leq f(y)$ for any $x, y \in X$ with $x \preceq y$. Let w be a positive function defined on X . Then the isotonic regression of a given function on X can be defined as follows.

Definition 2.1 Suppose g is a given function on X . A function g^* on X is an isotonic regression of g , with weights w if and only if g^* is isotonic and g^* minimizes

$$\sum_{x \in X} [g(x) - f(x)]^2 w(x) \quad (2.3)$$

in the class of all isotonic functions f on X .

Robertson et al.(1988)[22] point out the conditions that under which g^* achieves the minimal values.

Theorem 2.1 If K is a convex cone of functions on X and g and w are given

functions on X with $w > 0$, then a function g^* on X solves

$$\min \sum_{x \in X} [g(x) - f(x)]^2 w(x) \quad \text{subject to } f \in K \quad (2.4)$$

if and only if $g^* \in K$ satisfies

$$\sum_{x \in X} [g(x) - g^*(x)] g^*(x) w(x) = 0 \quad (2.5)$$

and

$$\sum_{x \in X} [g(x) - g^*(x)] f(x) w(x) \leq 0, \quad \text{for all } f \in K. \quad (2.6)$$

The orthogonal projection $g^* = P_w(g|K)$ is the solution of (2.4). It is usually the case that the isotonic regression function g^* can not be found in an explicit form. Since each f in K is a mapping from $X = \{x_1, x_2, \dots, x_k\}$ to a point in $\mathbf{R}^k$, the problem in (2.4) can be viewed as a more familiar minimization problem. Let $C = \{\boldsymbol{\mu} \in \mathbf{R}^k : \mu_i = f(x_i), i = 1, 2, \dots, k, f \in K\}$. Then C is a convex cone in $\mathbf{R}^k$ and depends on the partial ordering on X . If the partial order is simple linear $(x_1 \preceq x_2 \preceq \dots \preceq x_k)$, $C = \{\boldsymbol{\mu} \in \mathbf{R}^k : \mu_1 \leq \dots \leq \mu_k\}$. In case of simple-tree ordering $(x_1 \preceq x_i, i = 2, \dots, k)$, $C = \{\boldsymbol{\mu} \in \mathbf{R}^k : \mu_1 \leq \mu_j, j = 2, \dots, k\}$. Letting $\mathbf{y} = (y_1, y_2, \dots, y_k)'$ and $\mathbf{w} = (w_1, w_2, \dots, w_k)'$ where $y_i = g(x_i)$, and $w_i = w(x_i)$, $i = 1, \dots, k$, the problem in (2.4) becomes a usual least squares problem:

$$\min \|\mathbf{y} - \boldsymbol{\mu}\|_w^2 \quad \text{subject to } \boldsymbol{\mu} \in C. \quad (2.7)$$

The solution, $\boldsymbol{\mu}^* = P_w(\mathbf{y}|C)$, of this problem can be characterized by the conditions

$$(\mathbf{y} - \boldsymbol{\mu}^*, \boldsymbol{\mu}^*)_w = 0, \quad \text{and}$$

$$(\mathbf{y} - \boldsymbol{\mu}^*, \mathbf{z})_w \leq 0, \quad \text{for any } \mathbf{z} \in C.$$

There are a number of algorithms to compute the least squares projection and they depend on the type of convex cone C . In the case of simple linear order where $C = \{\boldsymbol{\mu} \in \mathbf{R}^k : \mu_1 \leq \mu_2 \leq \dots \leq \mu_k\}$, the most common algorithm is PAVA (the Pool-Adjacent-Violators Algorithm) that keeps pooling neighbor components until they satisfy the required ordering. The solution obtained by this algorithm can be represented as

$$\mu_j^* = \max_{i \leq j} \min_{l \geq j} Av(i, l), \quad j = 1, 2, \dots, k,$$

where $Av(i, l) = \frac{\sum_{m=i}^l w_m y_m}{\sum_{m=i}^l w_m}$, $1 \leq i \leq l \leq k$.

The algorithm for other type of cones are also available in Robertson et. al. (1988) [22].

2.1.3 Linear inequality constraints

A natural extension of isotonic regression problem is to use a Mahalanobis distance in the objective function and also to consider a set of linear inequality constraints. For this, we need to extend the concept of inner product of $\mathbf{x}$ and $\mathbf{y}$ in $\mathbf{R}^k$ as $(\mathbf{x}, \mathbf{y})_W = \sum_{i=1}^k \sum_{j=1}^k x_i y_j w_{ij}$, where W is a given positive definite matrix with

its (i, j) th element being w_{ij} . Then, the new problem is to

$$\min \|\mathbf{y} - \boldsymbol{\mu}\|_W^2 = (\mathbf{y} - \boldsymbol{\mu})' W (\mathbf{y} - \boldsymbol{\mu}) \quad \text{subject to } A\boldsymbol{\mu} \leq \mathbf{0}$$

where A is a $r \times k$ matrix of full row rank.

The polyhedral cone $C = \{\boldsymbol{\mu} \in \mathbf{R}^k : A\boldsymbol{\mu} \leq \mathbf{0}\}$ is more general and represents many cases depending on how we set up matrix A . For example, if A is a $(k-1) \times k$ matrix with

$$a_{ij} = \begin{cases} 1 & \text{if } i = j \\ -1 & \text{if } j = i + 1 \\ 0 & \text{otherwise} \end{cases} \quad (2.8)$$

for $i = 1, \dots, k-1$, then C is a simple linear order cone.

The computation algorithm for this problem was suggested by Dykstra(1983)[9] and will be described in Section 3.2.3 because we will use it as a part of our estimation procedure.

2.2 Tests of Constrained Hypotheses

The first monograph on order restricted statistical inference is the classical book *Statistical Inference Under Order Restrictions* by Barlow, Bartholomew, Bremner, and Brunk (1972)[3], which synthesized the development of this field during the 1950s and 1960s. A follow-up book by Robertson, Wright and Dykstra (1988) [22] provides a comprehensive review of researches until 1988. Recently, Silvapulle and Sen (2004)[25] published a book on the same subject but included more up-to-date

researches.

2.2.1 The one-way layout with ordered parameters

Many problems of statistical inference involves comparing several normal population means. The analysis of variance (ANOVA) is a commonly used statistical procedure for comparing the means of several groups in the one-way layout design.

Let X_{ij} be the j th observation in the i th group, $i = 1, 2, \dots, k$ and $j = 1, 2, \dots, n_i$, $X_{ij} \sim N(\mu_i, \sigma_i^2)$. Assume that the σ_i 's are known. Suppose that $\bar{X}_i$'s are sample means and independent. Then $\bar{X}_i \sim N(\mu_i, \frac{\sigma_i^2}{n_i})$.

The following hypotheses are of interest:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k,$$

$$H_1 : \mu_1 \leq \mu_2 \leq \dots \leq \mu_k,$$

$$H_2 : \text{No restriction on } \boldsymbol{\mu}.$$

The usual ANOVA tests H_0 against $H_2 - H_0$. The tests of interests with ordered parameters would be either H_0 versus $H_1 - H_0$, or H_1 versus $H_2 - H_1$. The latter test is a goodness-of-fit test which is used to justify the ordered parameter space in H_1 . Because $\bar{X}_i$ is sufficient for μ_i , the LRTs of H_0 against $H_2 - H_0$, H_0 versus $H_1 - H_0$ and of H_1 versus $H_2 - H_1$ can be determined from the likelihood function $L(\bar{X}_1, \bar{X}_1, \dots, \bar{X}_k; \boldsymbol{\mu}, \boldsymbol{\sigma})$.

Under H_0 , the MLE of $\mu_1 = \mu_2 = \cdots = \mu_k$ is given by

$$\bar{\mu} = \frac{\sum_{i=1}^k w_i \bar{X}_i}{\sum_{i=1}^k w_i}, \quad \text{with } w_i = \frac{n_i}{\sigma_i^2}. \quad (2.9)$$

Under H_1 , the MLE of $\boldsymbol{\mu}$ is the isotonic regression, $\boldsymbol{\mu}^*$, of $\bar{\mathbf{x}}$ with weight vector $\mathbf{w} = (w_1, w_2, \dots, w_k)'$. The unrestricted MLE of $\boldsymbol{\mu}$ is $\bar{\mathbf{X}} = (\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k)'$.

The likelihood ratio for testing H_0 versus $H_1 - H_0$ is

$$\Lambda = \frac{\max_{\boldsymbol{\mu} \in H_0} L(\bar{\mathbf{X}}; \boldsymbol{\mu}, \boldsymbol{\sigma})}{\max_{\boldsymbol{\mu} \in H_1} L(\bar{\mathbf{X}}; \boldsymbol{\mu}, \boldsymbol{\sigma})}$$

Hence, the LRT rejects H_0 in favor of $H_1 - H_0$ for large values of

$$T_{01} = \sum_{i=1}^k (\mu_i^* - \bar{\mu})^2 w_i. \quad (2.10)$$

Similarly, for testing H_1 versus $H_2 - H_1$, the LRT rejects H_1 for large values of

$$T_{12} = \sum_{i=1}^k (\bar{x}_i - \mu_i^*)^2 w_i \quad (2.11)$$

Geometrically, the LRT statistics are just the squared weighted Euclidean distance between the least squares projections onto the regions defined in the null and alternative hypotheses. It is instructive to note that, applying Pythagorean theorem in (2.1), the omnibus test statistics for H_0 against $H_2 - H_0$ is

$$T_{02} = \sum_{i=1}^k (\bar{x}_i - \bar{\mu}_i)^2 w_i = T_{01} + T_{12}. \quad (2.12)$$

As is well known, the null distribution of T_{02} is $\chi^2_{(k-1)}$. However, the LRT

statistics, T_{01} and T_{02} , have Chi-bar-square($\bar{\chi}^2$) distributions under H_0 . These distributions are mixtures of Chi-square distributions as shown in Bartholomew (1959)[4].

Theorem 2.2 *Let T_{01} and T_{12} be the LRT statistics for H_0 against $H_1 - H_0$ and H_1 against $H_2 - H_1$, respectively. Then, for any $t \geq 0$, we have*

$$i) P[T_{01} > t] = \sum_{l=1}^k p(l, k; \mathbf{w}) P[\chi_{l-1}^2 > t]$$

$$ii) P[T_{12} > t] = \sum_{l=1}^k p(l, k; \mathbf{w}) P[\chi_{k-l}^2 > t]$$

under H_0 , where χ_l^2 is a standard chi-square random variable with l degrees of freedom ($\chi_0^2 \equiv 0$) and $p(l, k; \mathbf{w})$ is the level probability as defined in (2.13).

The weights, $p(l, k; w)$, in the $\bar{\chi}^2$ distributions are called “level probabilities” that form a probability distribution for the number of distinct component values in the isotonic regression of $\mathbf{Z} = (Z_1, \dots, Z_k)'$ where Z_i are independent and follow $N(0, w_i^{-1})$.

Definition 2.2 *Let $\mu \in H_0$, the level probabilities are defined by*

$$p(l, k; \mathbf{w}) = P[L = l], \quad l = 1, 2, \dots, k \quad (2.13)$$

with L denoting the number of level sets in the isotonic regression, $\mathbf{X}^$, of $\mathbf{X} = (X_1, X_2, \dots, X_k)'$ with weight vector $\mathbf{w}$, where X_i are independent normal variables with mean μ_i and variance w_i^{-1}*

Since level probabilities do not depend on μ_i 's if $\mu_1 = \dots = \mu_k$, we may assume that $\mu_i = 0$ in the definition of the level probabilities. The level probabilities are

generally unknown unless $k \leq 4$. In the simple order case with equal weights ($w_1 = w_2 = \dots = w_k$), we have an explicit recursive relation given by

$$p(l, k) = \frac{1}{k}p(l-1, k-1) + \frac{k-1}{k}p(l, k-1), \quad \text{for } l = 2, 3, \dots, k-1$$

where $p(1, k) = \frac{1}{k}$ and $p(k, k) = \frac{1}{k!}$. However, if weights are not all equal, the level probabilities are approximately evaluated by a simulation with independent normal variables following $N(0, w_i^{-1})$, $i = 1, 2, \dots, k$. Robertson et al. (1988) [22] discuss the approximation of level probabilities in Section 3.4 for the simple order.

2.2.2 Testing one-sided hypothesis

Kudô (1963) [13] considered testing problem for one-sided hypothesis on multivariate normal mean vector. Suppose $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ are observations in a random sample from a k -dimensional multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a known positive definite covariance matrix.

The problem is now to test H_0 versus $H_1 - H_0$ and H_1 versus $H_2 - H_1$ where hypotheses are:

$$H_0 : \boldsymbol{\mu} = \mathbf{0},$$

$$H_1 : \boldsymbol{\mu} \geq \mathbf{0},$$

$$H_2 : \text{No restriction on } \boldsymbol{\mu}.$$

Obviously, the unconstrained MLE of $\boldsymbol{\mu}$ is $\hat{\boldsymbol{\mu}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i$. The MLE of $\boldsymbol{\mu}$ under

the cone $K = \{\boldsymbol{\mu} \in \mathbf{R}^k : \boldsymbol{\mu} \geq 0\}$ is obtained by solving

$$\min_{\boldsymbol{\mu} \in K} (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}). \quad (2.14)$$

The least square projection, $\boldsymbol{\mu}^*$, of $\hat{\boldsymbol{\mu}}$ onto K with inner product $(\mathbf{x}, \mathbf{y})_{\boldsymbol{\Sigma}^{-1}} = \mathbf{x}' \boldsymbol{\Sigma}^{-1} \mathbf{y}$ is the solution to (2.14) and will be denoted by $\boldsymbol{\mu}^* = P_{\boldsymbol{\Sigma}^{-1}}(\hat{\boldsymbol{\mu}}|K)$. Robertson et. al. (1988)[22] point out that $\boldsymbol{\mu}^*$ is the unique element of K which satisfies

$$(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}^*, \boldsymbol{\mu}^*)_{\boldsymbol{\Sigma}^{-1}} = 0$$

and

$$(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}^*, \mathbf{y})_{\boldsymbol{\Sigma}^{-1}} \leq 0, \quad \text{for all } \mathbf{y} \in K.$$

Kudô (1963) [13] describes a heuristic method to find $\boldsymbol{\mu}^*$. However, his method involves intuitive judgment in the selection of subspaces onto which to project $\bar{\mathbf{X}}$. In application, $\boldsymbol{\mu}^*$ can be easily obtained by Dykstra's [9] algorithm using software such as MATLAB, or R as will be discussed later.

Once estimators, $\hat{\boldsymbol{\mu}}$ and $\boldsymbol{\mu}^*$, are available, the LRTs can be constructed based on test statistics:

$$T_{01} = (\boldsymbol{\mu}^*)' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}^*$$

$$T_{12} = (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}^*)' \boldsymbol{\Sigma}^{-1} (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}^*)$$

The LRT rejects H_0 in favor of $H_1 - H_0$ for large values of T_{01} , and rejects H_1

in favor of $H_2 - H_1$ for large values of T_{12} . The null distributions of these test statistics are given in the following theorem.

Theorem 2.3 *Let T_{01} and T_{12} be the LRT statistics for H_0 against $H_1 - H_0$ and H_1 against $H_2 - H_1$, respectively. If H_0 is true, if Σ is known and positive definite, then, for any $t \geq 0$, we have*

$$i) P[T_{01} > t] = \sum_{l=0}^k W(l, k; \Sigma^{-1}) P[\chi_l^2 > t]$$

$$ii) P[T_{12} > t] = \sum_{l=0}^k W(l, k; \Sigma^{-1}) P[\chi_{k-l}^2 > t]$$

where χ_l^2 is a standard chi-square random variable with l degrees of freedom ($\chi_0^2 \equiv 0$) and $W(l, k; \Sigma^{-1})$ is the probability that $P_{\Sigma^{-1}}(\mathbf{Z}|K)$ with $\mathbf{Z} \sim N(\mathbf{0}, \Sigma)$ has exactly l positive component values. The weights $W(l, k; \Sigma^{-1}), l = 0, 1, 2, \dots, k$, sum to one.

There is a similarity between the level probabilities discussed in Section 2.2.1 and the $W(l, k; \Sigma^{-1})$ in Theorem 2.3. The weight $W(l, k; \Sigma^{-1})$ depends upon both the matrix Σ^{-1} and the cone K .

There is a connection between ordered and one-sided testing problems. Consider linear transformations

$$Y_i = \bar{X}_{i+1} - \bar{X}_i \quad \text{for } i = 1, 2, \dots, k-1.$$

Then $\mathbf{Y} = (Y_1, \dots, Y_{k-1})'$ follows $N(\theta_i, \Sigma)$ where $\theta_i = \mu_{i+1} - \mu_i, i = 1, \dots, k-1$,

and the (i, j) th element of Σ is given by

$$\delta_{ij} = \begin{cases} w_{i+1}^{-1} + w_i^{-1} = \frac{\sigma_{i+1}^2}{n_{i+1}} + \frac{\sigma_i^2}{n_i} & \text{if } |i - j| = 0 \\ -w_{i+1}^{-1} = -\frac{\sigma_{i+1}^2}{n_{i+1}} & \text{if } |i - j| = 1 \\ 0 & \text{for } |i - j| \geq 2. \end{cases} \quad (2.15)$$

Hence, those hypotheses in the testing problems in Section 2.2.1 are expressed as

$$H_0 : \theta_i = 0, \quad i = 1, 2, \dots, k-1$$

$$H_1 : \theta_i \geq 0, \quad i = 1, 2, \dots, k-1$$

$$H_2 : \text{No restriction on } \boldsymbol{\theta}.$$

Since the random vector $\mathbf{Y} = (Y_1, Y_2, \dots, Y_{k-1})'$ is sufficient for $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_{k-1})'$, we can use $\mathbf{Y}$ for constructing the test statistics when testing H_0 versus $H_1 - H_0$, and H_1 versus $H_2 - H_1$. Therefore, the testing problems discussed in Section 2.2.1 can also be formulated as one-sided problems developed by Kudô (1963) [13].

2.2.3 Testing hypothesis of linear inequality

In Section 2.2.2, the constraints in H_1 is assumed to be in a orthant cone K . Shapiro (1988)[24] assumes that the matrix Σ is known and positive definite but considers arbitrary convex cones, $K = \{\boldsymbol{\mu} \in \mathbf{R}^k : A\boldsymbol{\mu} \geq 0\}$.

Let $\mathbf{X}$ be a k -dimensional random vector following $N(\boldsymbol{\mu}, \Sigma)$, where Σ is known positive definite matrix. Shapiro (1988)[24] considered problems of testing

H_0 versus $H_1 - H_0$ and H_1 versus $H_2 - H_1$ with hypotheses given by

$$H_0 : A\boldsymbol{\mu} = \mathbf{0},$$

$$H_1 : A_1\boldsymbol{\mu} \geq \mathbf{0},$$

$$H_2 : \text{No restriction on } \boldsymbol{\mu}.$$

where A is a $r \times k$ matrix of full row rank and A_1 is a $q \times k$ submatrix of A . The estimator of $\boldsymbol{\mu}$ under H_0 is expressed as

$$\bar{\boldsymbol{\mu}} = [I - \Sigma A' (A \Sigma A')^{-1} A] \mathbf{X}.$$

However, the constrained estimator, $\boldsymbol{\mu}^*$, of $\boldsymbol{\mu}$ under H_1 does not have a closed form of expression. Instead, $\boldsymbol{\mu}^*$ can be obtained iteratively as discussed in Dykstra (1983) [9]. The following theorem summarizes the results on the likelihood ratio tests based on $\bar{\boldsymbol{\mu}}$, $\boldsymbol{\mu}^*$, and the unrestricted estimator $\hat{\boldsymbol{\mu}}$.

Theorem 2.4 *Let T_{01} and T_{12} be the likelihood ratio statistics for testing H_0 versus $H_1 - H_0$ and H_1 versus $H_2 - H_1$, respectively. Then, for any value $\boldsymbol{\mu} \in H_0$, we have for any $t \geq 0$,*

$$i) P[T_{01} > t] = \sum_{i=0}^q W(i, q; (A_1 \Sigma A_1')^{-1}) P[\chi_{r-q+i}^2 > t]$$

$$ii) P[T_{12} > t] = \sum_{i=0}^r W(i, r; (A_1 \Sigma A_1')^{-1}) P[\chi_{r-i}^2 > t]$$

where χ_l^2 is a standard chi-square random variable with l degrees of freedom ($\chi_0^2 \equiv 0$) and $W(\cdot, \cdot; \cdot)$ are the weights as those in Theorem 2.3.

Typically, we have interests in the case with $A_1 = A$ (i.e. $q = r$). If we let $\mathbf{Y} =$

AX , and $\boldsymbol{\theta} = A\boldsymbol{\mu}$, then $\mathbf{Y} \sim \mathbf{N}(\boldsymbol{\theta}, A\Sigma A')$. Based on this reparameterization, the hypotheses in the previous problem can be rewritten as

$$H_0: \boldsymbol{\theta} = \mathbf{0},$$

$$H_1: \boldsymbol{\theta} \geq \mathbf{0},$$

$$H_2: \text{No restriction on } \boldsymbol{\theta}.$$

So, if $q = r$, the problem can be reformulated as Kudô's problem (1963) [13]. We can now perform hypothesis tests based upon random variable $\mathbf{Y}$, following the same procedure as discussed in Section 2.2.2.

Chapter 3

Inference for Ordered Groups

In this chapter we first use the concept of a latent variable to derive a joint distribution for a bivariate response that is composed of a continuous and a binary outcome. Based on this model we will formulate ordered hypothesis to compare several groups. Then we will discuss how the maximum likelihood estimates can be obtained under each hypothesis. Finally, we will derive the asymptotic null distributions of the LRT statistics.

3.1 Model and Related Hypotheses

3.1.1 The model

For $i = 1, 2, \dots, k$, let $\mathbf{X}_{ij} = (X_{ij1}, X_{ij2})'$ represent j th bivariate observation in a random sample of size n_i taken from the i th group. We assume X_{ij1} is a binary outcome that is observed jointly with a normal variable X_{ij2} . Since it is not possible to provide an omnibus joint model for $\mathbf{X}_{ij}$, we consider a bivariate

latent variable, $\mathbf{Y}_{ij} = (Y_{ij1}, Y_{ij2})'$, that follows $N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2})'$ and

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{bmatrix}$$

Then, we define

$$X_{ij1} = \begin{cases} 1 & \text{if } Y_{ij1} \geq 0 \\ 0 & \text{if } Y_{ij1} < 0 \end{cases} \quad (3.1)$$

and $X_{ij2} = Y_{ij2}$.

In this setup $X_{ij1} \sim b(1, p_i)$ where $p_i = 1 - \Phi(-\mu_{i1})$ and $X_{ij2} \sim N(\mu_{i2}, \sigma^2)$ and they are dependent unless $\rho \equiv \text{Corr}(Y_{ij1}, Y_{ij2}) = 0$. We also note that order relation $p_1 \leq \dots \leq p_k$ is equivalent to the ordering $\mu_{11} \leq \mu_{21} \leq \dots \leq \mu_{k1}$. The assumption of $\text{Var}(Y_{ij1}) = 1$ is to ensure the identifiability of the joint model for $\mathbf{X}_{ij}$.

Let $f(\mathbf{Y}_{ij}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ be the joint probability density function of $\mathbf{Y}_{ij}$ and define the following functions:

$$h_1(y_{ij2}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}) = \int_0^\infty f(\mathbf{y}_{ij}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}) dy_{ij1},$$

and

$$h_2(y_{ij2}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}) = \int_{-\infty}^0 f(\mathbf{y}_{ij}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}) dy_{ij1}.$$

Since $\mathbf{Y}_{ij} \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$, it follows that

$$h_1(y_{ij2}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}) = \phi\left(\frac{y_{ij2} - \mu_{i2}}{\sigma}\right) \Phi\left(-\frac{\mu_{i1} + \frac{\rho}{\sigma}(y_{ij2} - \mu_{i2})}{\sqrt{1 - \rho^2}}\right)$$

$$h_2(y_{ij2}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}) = \phi\left(\frac{y_{ij2} - \mu_{i2}}{\sigma}\right) \Phi\left(\frac{\mu_{i1} + \frac{\rho}{\sigma}(y_{ij2} - \mu_{i2})}{\sqrt{1 - \rho^2}}\right)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are standard normal p.d.f. and c.d.f., respectively.

Therefore, the density of $\mathbf{X}_{ij}$ can be expressed as

$$f_{X_{ij}}(\mathbf{x}_{ij}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}) = [h_1(x_{ij2}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma})]^{x_{ij1}} [h_2(x_{ij2}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma})]^{1-x_{ij1}} \quad (3.2)$$

$$x_{ij1} = 0 \text{ or } 1, \quad -\infty < x_{ij2} < \infty.$$

Throughout the thesis, we will assume the density in (3.2) is the common density for i.i.d. observations $\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i}$, for group i .

3.1.2 Hypotheses for joint analysis

The group effect on the outcome variables is determined by location parameters, μ_{i1} and μ_{i2} , respectively. Both of the group effects are different from treatment to treatment. The binary and continuous response variables depend on each other through the coefficient ρ . If the groups of experimental units are treated with treatments that follow some natural ordering, for example, different dosage levels from low to high, one may be curious to know whether the group effects on the responses are also ordered. This question can be formulated as a testing problem which involves an ordered hypothesis. That is, in order to investigate the possible

ordering of group effects on the responses, we consider the following hypotheses:

$$H_0 : \mu_{11} = \mu_{21} = \cdots = \mu_{k1}, \mu_{12} = \mu_{22} = \cdots = \mu_{k2}$$

$$H_1 : \mu_{11} \geq \mu_{21} \geq \cdots \geq \mu_{k1}, \mu_{12} \geq \mu_{22} \geq \cdots \geq \mu_{k2}$$

$$H_2 : \text{No restriction on } \boldsymbol{\mu}$$

The goodness-of-fit for the assumption of order restriction can be evaluated by testing H_1 versus $H_2 - H_1$. However, we are primarily interested in testing H_0 versus $H_1 - H_0$ for the ordered group effects, given that H_1 is concluded in the previous test. If we reject H_0 in favor of $H_1 - H_0$, then there is evidence that the group effects exist as an ordering instead of homogeneity.

3.2 Maximum Likelihood Estimation

In this section, we discuss how to obtain maximum likelihood estimates (MLEs) of the model parameters under each of the three hypotheses, H_0, H_1 and H_2 . Based on the model introduced in Section 3.1.1 the likelihood function can be expressed as

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{x}) = \prod_{i=1}^k \prod_{j=1}^{n_i} [h_1(x_{ij2}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma})]^{x_{ij1}} [h_2(x_{ij2}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma})]^{1-x_{ij1}}$$

where $\boldsymbol{\mu} = (\mu_{11}, \mu_{21}, \cdots, \mu_{k1}, \mu_{12}, \mu_{22}, \cdots, \mu_{k2})'$ and $\mathbf{x} = (\mathbf{x}_1', \mathbf{x}_2')'$ with $\mathbf{x}_1 = (x_{111}, x_{121}, \cdots, x_{1n_11}, x_{211}, x_{221}, \cdots, x_{2n_21}, \cdots, x_{k11}, x_{k21}, \cdots, x_{kn_k1})'$, and $\mathbf{x}_2 = (x_{112}, x_{122}, \cdots, x_{1n_12}, x_{212}, x_{222}, \cdots, x_{2n_22}, \cdots, x_{k12}, x_{k22}, \cdots, x_{kn_k2})'$.

The log-likelihood function becomes

$$l(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{x}) = \sum_{i=1}^k \sum_{j=1}^{n_i} \{x_{ij1} \ln[h_1(x_{ij2}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma})] + (1 - x_{ij1}) \ln[h_2(x_{ij2}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma})]\}.$$

Since no closed form of expressions of the MLEs are available, we rely on numerical approaches to obtain the estimates.

3.2.1 Unrestricted estimates

Let $\theta = (\boldsymbol{\mu}', \sigma^2, \rho)'$. Under H_2 , the unrestricted maximum likelihood estimate of θ can be obtained by solving the equations $\frac{\partial \ln L}{\partial \boldsymbol{\mu}} = \mathbf{0}$, $\frac{\partial \ln L}{\partial \sigma^2} = 0$, and $\frac{\partial \ln L}{\partial \rho} = 0$.

Recall the expressions of h_1 and h_2 are

$$h_1(x_{ij2}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}) = \phi\left(\frac{x_{ij2} - \mu_{i2}}{\sigma}\right) \Phi\left(-\frac{\mu_{i1} + \frac{\rho}{\sigma}(x_{ij2} - \mu_{i2})}{\sqrt{1 - \rho^2}}\right)$$

and

$$h_2(x_{ij2}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}) = \phi\left(\frac{x_{ij2} - \mu_{i2}}{\sigma}\right) \Phi\left(\frac{\mu_{i1} + \frac{\rho}{\sigma}(x_{ij2} - \mu_{i2})}{\sqrt{1 - \rho^2}}\right).$$

After some calculations, we have the first-order derivatives of the log-likelihood function:

$$\frac{\partial \ln L}{\partial \mu_{i1}} = \frac{1}{\sqrt{1 - \rho^2}} \sum_{j=1}^{n_i} \phi(t_{ij}) \left(x_{ij1} \cdot \frac{1}{\Phi(t_{ij})} - (1 - x_{ij1}) \cdot \frac{1}{\Phi(-t_{ij})} \right), \quad i = 1, 2, \dots, k$$

$$\begin{aligned} \frac{\partial \ln L}{\partial \mu_{i2}} &= (\sigma^2)^{-1/2} \sum_{j=1}^{n_i} \left[x_{ij1} \left(z_{ij} - \frac{\phi(t_{ij})}{\Phi(t_{ij})} \frac{\rho}{\sqrt{1-\rho^2}} \right) + \right. \\ &\quad \left. (1 - x_{ij1}) \left(z_{ij} + \frac{\phi(t_{ij})}{\Phi(-t_{ij})} \frac{\rho}{\sqrt{1-\rho^2}} \right) \right], \quad i = 1, 2, \dots, k \end{aligned}$$

$$\begin{aligned} \frac{\partial \ln L}{\partial \sigma^2} &= \left(\frac{1}{2} \right) (\sigma^2)^{-1} \sum_{i=1}^k \sum_{j=1}^{n_i} \left\{ [x_{ij1}(-1 + (z_{ij})^2) - (z_{ij}) \frac{\phi(t_{ij})}{\Phi(t_{ij})} \frac{\rho}{\sqrt{1-\rho^2}}] + \right. \\ &\quad \left. (1 - x_{ij1}) [(-1 + (z_{ij})^2) - (z_{ij}) \frac{\phi(t_{ij})}{\Phi(-t_{ij})} \frac{\rho}{\sqrt{1-\rho^2}}] \right\} \end{aligned}$$

$$\begin{aligned} \frac{\partial \ln L}{\partial \rho} &= \frac{1}{\sqrt{1-\rho^2}} \sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij1} \phi(t_{ij}) \left[\frac{1}{\Phi(t_{ij})} (z_{ij} + \rho t_{ij}) - \right. \\ &\quad \left. (1 - x_{ij1}) \frac{1}{\Phi(-t_{ij})} (z_{ij} + \rho t_{ij}) \right] \end{aligned}$$

where $z_{ij} = \frac{x_{ij2} - \mu_{i2}}{\sigma}$ and $t_{ij} = \frac{\mu_{i1} + \rho z_{ij}}{\sqrt{1-\rho^2}}$.

After setting the above equations equal to zero, one can find the unconstrained MLE of $\boldsymbol{\theta}$. However it is not easy to solve those highly nonlinear equations directly. One may use Newton's method to get the approximation for $\hat{\boldsymbol{\theta}}$. Let $l'(\boldsymbol{\theta})$ be a $(2k+2) \times 1$ vector, which stores all the first-order derivatives of the log-likelihood function $l(\boldsymbol{\theta})$. Let $l''(\boldsymbol{\theta})$ be the $(2k+2) \times (2k+2)$ Hessian matrix of the log-

likelihood function as follows:

$$l''(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial^2 l}{\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}'} & \frac{\partial^2 l}{\partial \boldsymbol{\mu} \partial \sigma^2} & \frac{\partial^2 l}{\partial \boldsymbol{\mu} \partial \rho} \\ \frac{\partial^2 l}{\partial \sigma^2 \partial \boldsymbol{\mu}'} & \frac{\partial^2 l}{\partial \sigma^2 \partial \sigma^2} & \frac{\partial^2 l}{\partial \sigma^2 \partial \rho} \\ \frac{\partial^2 l}{\partial \rho \partial \boldsymbol{\mu}'} & \frac{\partial^2 l}{\partial \rho \partial \sigma^2} & \frac{\partial^2 l}{\partial \rho \partial \rho} \end{bmatrix}.$$

Then, after some tedious algebra, we get

$$\frac{\partial^2 l}{(\partial \mu_{i1})^2} = \frac{1}{1 - \rho^2} \sum_{j=1}^{n_i} [-x_{ij1} \pi(t_{ij}) - (1 - x_{ij1}) \pi(-t_{ij})]$$

$$\frac{\partial^2 l}{\partial \mu_{i1} \partial \mu_{i2}} = \frac{\rho}{\sigma(1 - \rho^2)} \sum_{j=1}^{n_i} [x_{ij1} \pi(t_{ij}) + (1 - x_{ij1}) \pi(-t_{ij})]$$

$$\frac{\partial^2 l}{\partial \mu_{i1} \partial \sigma^2} = \frac{1}{\sqrt{1 - \rho^2}} \sum_{j=1}^{n_i} [x_{ij1} z_{ij} \pi(t_{ij}) + (1 - x_{ij1}) z_{ij} \pi(-t_{ij})]$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \mu_{i1} \partial \rho} = & -\frac{1}{1 - \rho^2} \sum_{j=1}^{n_i} \left\{ x_{ij1} \left[\left(z_{ij} + \frac{\rho t_{ij}}{\sqrt{1 - \rho^2}} \right) \pi(t_{ij}) + \frac{\phi(t_{ij})}{\Phi(t_{ij})} \frac{\rho}{\sqrt{1 - \rho^2}} \right] \right. \\ & \left. + (1 - x_{ij1}) \left[\left(z_{ij} + \frac{\rho t_{ij}}{\sqrt{1 - \rho^2}} \right) \pi(-t_{ij}) + \frac{\phi(t_{ij})}{\Phi(-t_{ij})} \frac{\rho}{\sqrt{1 - \rho^2}} \right] \right\} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \mu_{i2} \partial \sigma^2} = & (-\frac{1}{2})(\sigma^{-3}) \sum_{j=1}^{n_i} \left[x_{ij1} \left(2z_{ij} - \frac{\rho}{\sqrt{1 - \rho^2}} \frac{\phi(t_{ij})}{\Phi(t_{ij})} + \frac{\rho z_{ij}}{1 - \rho^2} \pi(t_{ij}) \right) \right. \\ & \left. + (1 - x_{ij1}) \left(2z_{ij} + \frac{\rho}{\sqrt{1 - \rho^2}} \frac{\phi(t_{ij})}{\Phi(-t_{ij})} + \frac{\rho z_{ij}}{1 - \rho^2} \pi(-t_{ij}) \right) \right] \end{aligned}$$

$$\frac{\partial^2 l}{(\partial \mu_{i2})^2} = -(\sigma^{-2}) \sum_{j=1}^{n_i} \left[x_{ij1} \left(1 + \frac{\rho^2}{1 - \rho^2} \pi(t_{ij}) \right) + (1 - x_{ij1}) \left(1 + \frac{\rho^2}{1 - \rho^2} \pi(-t_{ij}) \right) \right]$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \mu_{i2} \partial \rho} &= \frac{1}{\sigma(1-\rho^2)} \sum_{j=1}^{n_i} \left\{ x_{ij1} \left[\left(z_{ij} + \frac{\rho t_{ij}}{\sqrt{1-\rho^2}} \right) \pi(t_{ij}) + \frac{\phi(t_{ij})}{\Phi(t_{ij})} \frac{\rho}{\sqrt{1-\rho^2}} \right] \right. \\ &\quad \left. + (1-x_{ij1}) \left[\left(z_{ij} + \frac{\rho t_{ij}}{\sqrt{1-\rho^2}} \right) \pi(-t_{ij}) + \frac{\phi(t_{ij})}{\Phi(-t_{ij})} \frac{\rho}{\sqrt{1-\rho^2}} \right] \right\} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l}{(\partial \sigma^2)^2} &= \left(\frac{1}{2} \right) (\sigma^2)^{-2} \sum_{i=1}^k \sum_{j=1}^{n_i} \left\{ x_{ij1} \left[-z_{ij}^2 + \frac{z_{ij}\rho}{2\sqrt{1-\rho^2}} \frac{\phi(t_{ij})}{\Phi(t_{ij})} - \frac{z_{ij}^2 \rho^2}{2(1-\rho^2)} \pi(t_{ij}) \right] \right. \\ &\quad \left. + (1-x_{ij1}) \left[-z_{ij}^2 + \frac{z_{ij}\rho}{2\sqrt{1-\rho^2}} \frac{\phi(t_{ij})}{\Phi(-t_{ij})} + \frac{z_{ij}^2 \rho^2}{2(1-\rho^2)} \pi(-t_{ij}) \right] \right\} \\ &\quad - (\sigma^2)^{-1} \frac{\partial l}{\partial \sigma^2} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \sigma^2 \partial \rho} &= \left(\frac{1}{2} \right) \frac{1}{\sigma^2 \sqrt{1-\rho^2}} \sum_{i=1}^k \sum_{j=1}^{n_i} \left\{ x_{ij1} z_{ij} \left[-(1-\rho^2)^{-1} \frac{\phi(t_{ij})}{\Phi(t_{ij})} + \rho \left(\frac{z_{ij}}{\sqrt{1-\rho^2}} + \frac{t_{ij}\rho}{1-\rho^2} \right) \pi(t_{ij}) \right] \right. \\ &\quad \left. + (1-x_{ij1}) z_{ij} \left[-(1-\rho^2)^{-1} \frac{\phi(t_{ij})}{\Phi(-t_{ij})} - \rho \left(\frac{z_{ij}}{\sqrt{1-\rho^2}} + \frac{t_{ij}\rho}{1-\rho^2} \right) \pi(-t_{ij}) \right] \right\} \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l}{(\partial \rho)^2} &= \frac{1}{\sqrt{1-\rho^2}} \sum_{i=1}^k \sum_{j=1}^{n_i} \left\{ -x_{ij1} \left[\left(\frac{z_{ij}}{\sqrt{1-\rho^2}} + \frac{t_{ij}\rho}{1-\rho^2} \right) \pi(t_{ij}) (z_{ij} + t_{ij}\rho) + \frac{t_{ij}}{1-\rho^2} \frac{\phi(t_{ij})}{\Phi(t_{ij})} \right] \right. \\ &\quad \left. - (1-x_{ij1}) \left[\left(\frac{z_{ij}}{\sqrt{1-\rho^2}} + \frac{t_{ij}\rho}{1-\rho^2} \right) \pi(-t_{ij}) (z_{ij} + t_{ij}\rho) + \frac{t_{ij}}{1-\rho^2} \frac{\phi(t_{ij})}{\Phi(-t_{ij})} \right] \right\} \\ &\quad + \frac{\rho}{1-\rho^2} \frac{\partial l}{\partial \rho} \end{aligned}$$

where $\pi(t_{ij}) = \frac{\phi(t_{ij})}{\Phi(t_{ij})^2} [t_{ij}\Phi(t_{ij}) + \phi(t_{ij})]$.

The unconstrained estimate of $\boldsymbol{\theta}$ can be obtained by Newton's method.

Starting from $\boldsymbol{\theta}^0$, this method keeps updating the estimate by

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - [l''(\boldsymbol{\theta}^{(t)})]^{-1} l'(\boldsymbol{\theta}^{(t)})$$

until satisfactory convergence.

Since the second-order derivatives of the log-likelihood function are complicated, we may approximate $l''(\boldsymbol{\theta}^{(t)})$ by

$$l''(\boldsymbol{\theta}^{(t)})_{ij} \doteq \frac{1}{\Delta} \left[\frac{\partial l}{\partial \theta_i} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(t)}+\frac{1}{2}\Delta\mathbf{e}_j} - \frac{\partial l}{\partial \theta_i} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(t)}-\frac{1}{2}\Delta\mathbf{e}_j} \right],$$

$$j = 1, \dots, 2k+2, \quad i = 1, \dots, 2k+2$$

where Δ is a sufficiently small value and $\mathbf{e}_j = (0, \dots, 0, 1, 0, \dots, 0)'$ is a $(2k+2) \times 1$ vector with '1' located in the j th entry.

Hereafter, the unconstrained estimator of $\boldsymbol{\theta}$ obtained by Newton's method will be denoted by $\hat{\boldsymbol{\theta}}$.

3.2.2 Constrained estimation

To determine the order restricted estimates under the conditions stated in H_1 , we need to minimize $-\ln L(\boldsymbol{\theta})$, under $\mu_{11} \geq \mu_{21} \geq \dots \geq \mu_{k1}$ and $\mu_{21} \geq \mu_{22} \geq \dots \geq \mu_{k2}$. Similarly to the Newton's method, we can find the constrained estimate by solving repeatedly quadratic minimization under the order restrictions.

Recall the second order Taylor's approximation to $l(\boldsymbol{\theta})$ at $\boldsymbol{\theta}^{(m)}$ gives

$$l(\boldsymbol{\theta}) \doteq l(\boldsymbol{\theta}^{(m)}) + [l'(\boldsymbol{\theta}^{(m)})]'(\boldsymbol{\theta} - \boldsymbol{\theta}^{(m)}) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(m)})' l''(\boldsymbol{\theta}^{(m)}) (\boldsymbol{\theta} - \boldsymbol{\theta}^{(m)})$$

Based on this second-order approximation, we can update intermediate estimate by solving

$$\min Q(\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})' [-l''(\boldsymbol{\theta}^{(m)})] (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}), \text{ subject to } A\boldsymbol{\theta} \leq \mathbf{0} \quad (3.3)$$

where $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(m)} - l''(\boldsymbol{\theta}^{(m)})^{-1}l'(\boldsymbol{\theta}^{(m)})$ and A is a $(2k-2) \times (2k+2)$ matrix of the form

$$A = \begin{bmatrix} B & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & B & \mathbf{0} \end{bmatrix}$$

with submatrix B being $(k-1) \times k$ and given by

$$B = \begin{bmatrix} -1 & 1 & & \cdots & 0 \\ 0 & -1 & 1 & & \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & & \cdots & -1 & 1 \end{bmatrix}$$

Hence, the constrained estimation procedure can be summarized in the following steps:

Step 1: Find the unrestricted estimate, $\hat{\boldsymbol{\theta}}$, of $\boldsymbol{\theta}$ based on usual Newton's method.

Step 2: If $\hat{\boldsymbol{\theta}}$ satisfy $A\boldsymbol{\theta} \leq \mathbf{0}$, let $\boldsymbol{\theta}^* = \hat{\boldsymbol{\theta}}$ and stop. Otherwise, pick up a point $\boldsymbol{\theta}^{(0)}$ that satisfies $A\boldsymbol{\theta}^{(0)} \leq \mathbf{0}$ and is close to $\hat{\boldsymbol{\theta}}$ as much as possible.

Step 3: After m iterations, let $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(m)} - l''(\boldsymbol{\theta}^{(m)})^{-1}l'(\boldsymbol{\theta}^{(m)})$.

Step 4: Solve the quadratic minimization problem in (3.3) and denote the solution by $\boldsymbol{\theta}^{(m+1)}$.

Step 5: Let $m = m + 1$ and go to Step 3.

In this algorithm, Steps 3 - 5 are iterated until the required convergence. This constrained estimate will be denoted by $\boldsymbol{\theta}^*$. In order to solve the problem in Step 4, one can use standard software such as MATLAB, R or SAS to find the constrained estimates. The iterative algorithm for restricted least squares estimation by Dykstra (1983) [9] can be implemented in a simple way, and will be discussed in the next section.

To estimate $\boldsymbol{\theta}$ under the null hypothesis H_0 , we consider $\bar{\boldsymbol{\mu}} = (\mu_1 \mathbf{1}', \mu_2 \mathbf{1}')$ in the likelihood function, and then apply the usual Newton's method to find the MLE. We shall denote this estimate by $\bar{\boldsymbol{\theta}} = (\bar{\boldsymbol{\mu}}', \bar{\sigma}^2, \bar{\rho})'$.

One may also use the same approximation to the likelihood function as in constrained estimation. In this approach, estimates are updated by solving

$$\min Q(\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})'[-l''(\boldsymbol{\theta}^{(m)})](\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}), \quad \text{subject to } A\boldsymbol{\theta} = \mathbf{0} \quad (3.4)$$

where matrix A is the same as what we described earlier in this section. Let

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = Q(\boldsymbol{\theta}) + \boldsymbol{\lambda}' A\boldsymbol{\theta}$$

where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_{k-1})$ and λ_i is Lagrange multiplier for the equality constraints $\mu_{i1} - \mu_{(i+1)1} = 0$ and $\mu_{i2} - \mu_{(i+1)2} = 0$, $i = 1, 2, \dots, k-1$.

The Kuhn-Tucker conditions (details can be found in Luenberger(1984) [16]) for this minimization are obtained by first setting

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}} = (-l''(\boldsymbol{\theta}^{(m)}))(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) + A' \boldsymbol{\lambda} = \mathbf{0}$$

So we have

$$A' \lambda = l''(\theta^{(m)})(\theta - \tilde{\theta})$$

Then

$$\theta = \tilde{\theta} + [l''(\hat{\theta})]^{-1} A' \lambda \quad (3.5)$$

Putting this solution into the constraint $A\theta = \mathbf{0}$ leads to

$$\lambda = -[A(l''(\theta^{(m)}))^{-1} A']^{-1} A\theta^{(m)}.$$

Then we can plug λ into Equation (3.5) and get the new updated value

$$\theta^{(m+1)} = \theta^{(m)} - (l''(\theta^{(m)}))^{-1} A' \left(A[l''(\theta^{(m)})]^{-1} A' \right)^{-1} A\theta^{(m)}.$$

Using this update, we can suggest an algorithm similar to the constrained estimation procedure under H_1 . Starting with an arbitrary initial value $\theta^{(0)}$ satisfying $A\theta^{(0)} = \mathbf{0}$, we are only to iterate Steps 3 - 5 with the modified updating procedure in Step 4. We will denote this estimate by $\bar{\theta}$.

3.2.3 Least squares estimation under linear inequality constraints

As discussed in the previous section, constrained estimate, θ^* , of θ under H_1 can be obtained by solving the least squares problem 3.3 repeatedly. Since this prob-

lem also requires iterative procedures, we introduce briefly Dykstra's algorithm [9].

Let $\mathbf{g}$ and $\mathbf{w}$ be fixed points in $\mathbf{R}^k$. Components of the weight vector $\mathbf{w} = (w_1, \dots, w_k)'$ are assumed to be all positive. Suppose we wish to solve

$$\min \sum_{i=1}^k (g_i - x_i)^2 w_i$$

subject to

$$\sum_{j=1}^k a_{ij} x_j \leq 0, \text{ for } i = 1, 2, \dots, r$$

If we let $K_i = \{\mathbf{x}; \mathbf{a}_i' \mathbf{x} \leq 0\}$, where $\mathbf{a}_i = (a_{i1}, \dots, a_{ik})'$ is a $k \times 1$ vector. We assume $\mathbf{a}_i$, $i = 1, \dots, r$ are linearly independent. This problem involves finding the projection of $\mathbf{g}$ onto K_i 's. A closed form for the projections can be easily found:

$$\begin{aligned} P_w(\mathbf{g}|K_i) &= \mathbf{g}, & \text{if } \sum_{j=1}^k a_{ij} g_j \leq 0, \\ &= (g_1^*, \dots, g_k^*)', & \text{if } \sum_{j=1}^k a_{ij} g_j > 0, \end{aligned}$$

where

$$g_j^* = g_j - \frac{(\sum_{l=1}^k g_l a_{il}) a_{ij} w_j^{-1}}{\sum_{l=1}^k a_{il}^2 w_l^{-1}}.$$

However, we need to find $P_w(\mathbf{g}|K)$ where $K = \cap_{i=1}^r K_i$. Dykstra's algorithm (1983) [9] finds sequentially the least squares projection of $\mathbf{g}$ onto K_i over $i = 1, \dots, r$, and then adjust the last least squares projection before starting

the next iteration. Dykstra's algorithm is summarized in a step-by step fashion emphasizing the role played by the increments. A schematic diagram of the algorithm is given in Figure 3.1 (this diagram is taken from Dykstra [9] with slight modification).

Figure 3.1: Schematic Drawing of Dykstra's Algorithm

Step 1: Let $g_{1,1}$ denote the projection of g onto the cone K_1 . We let $I_{1,1} = g_{1,1} - g$ denote the incremental change incurred by the projection, so that $g_{1,1} = g + I_{1,1}$.

Step 2: Let $g_{1,2}$ denote the projection of $g_{1,1}$ onto K_2 . The incremental change is $I_{1,2} = g_{1,2} - g_{1,1} = g_{1,2} - (g + I_{1,1})$, so that $g_{1,2} = g + I_{1,1} + I_{1,2}$.

Step 3: Let $g_{1,3}$ be the projection of $g_{1,2}$ onto K_3 . The incremental change is $I_{1,3} = g_{1,3} - g_{1,2} = g_{1,3} - (g + I_{1,1} + I_{1,2})$, so that $g_{1,3} = g + I_{1,1} + I_{1,2} + I_{1,3}$.

Step 4: Continue. After $g_{1,r}$ and $I_{1,r} = g_{1,r} - g_{1,r-1} = g_{1,r} - (g + I_{1,1} + I_{1,2} + \dots + I_{1,r-1})$ are found, we let $g_{2,1}$ denote the projection of $g + I_{1,2} + \dots + I_{1,r}$ onto K_1 . Note that we have removed the increment $I_{1,1}$ before the projection. Our new increment is $I_{2,1} = g_{2,1} - (g + I_{1,2} + \dots + I_{1,r})$, so that $g_{2,1} = g + I_{2,1} + I_{1,2} + \dots + I_{1,r}$.

Step 5: Continue until $g_{m,r}$ converges where $g_{m,r}$ is the projection onto the r th cone during the m th cycle.

This algorithm can be easily implemented in an R program as presented in the Appendix. This problem can be extended to a more general case. Let us consider the following problem:

$$\min (\mathbf{g} - \mathbf{x})' F (\mathbf{g} - \mathbf{x})$$

subject to the constraints of the form $\sum_{i=1}^n a_{ij} x_j = \mathbf{a}_i' \mathbf{x} \leq 0, i = 1, 2, \dots, r$, where F is a given, symmetric, positive definitive matrix. By Spectral Decomposition theorem we can always find an orthogonal matrix P such that $F = PDP'$ where D is the diagonal matrix consisting of the eigenvalues of F . In this case,

the problem can be transformed as

$$\min (P' \mathbf{g} - \mathbf{y})' D (P' \mathbf{g} - \mathbf{y})$$

subject to

$$(P \mathbf{a}_i)' \mathbf{y} \leq 0, \quad i = 1, 2, \dots, r,$$

where $\mathbf{y} = P' \mathbf{x}$. Since D is a diagonal matrix with positive elements, we can use Dykstra's algorithm to solve for $\mathbf{y}$. Then, the desired solution is $\mathbf{x} = P \mathbf{y}$.

3.3 Asymptotic Properties of the Estimators

Based on the density function in (3.2), the log-density function of $\mathbf{X}_{ij}$ is given by

$$\ln f_{X_{ij}}(\mathbf{x}_{ij}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}) = x_{ij1} \ln[h_1(x_{ij2}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma})] + (1 - x_{ij1}) \ln[h_2(x_{ij2}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma})] \quad (3.6)$$

where $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2})'$ and $\boldsymbol{\Sigma}$ is the covariance matrix in Section 3.1.1.

Recall the notation for the combined parameter vector is

$$\boldsymbol{\theta} = (\mu_{11}, \mu_{21}, \dots, \mu_{k1}, \mu_{12}, \dots, \mu_{k2}, \sigma^2, \rho)'$$

Since $\mathbf{X}_{i1}, \mathbf{X}_{i2}, \dots, \mathbf{X}_{in_i}$ are independent and identically distributed, the Fisher information matrices contained in those individual observations are commonly

given by

$$I_i^*(\boldsymbol{\theta}) = -E \left[\frac{\partial^2 \ln f_{X_{ij}}(\mathbf{x}_{ij}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] \quad (3.7)$$

Note that elements in $I_i^*(\boldsymbol{\theta})$ are zero unless both of their row and column are associated with μ_{i1} , μ_{i2} , σ^2 and ρ .

Since there are k samples of sizes $\mathbf{n} = (n_1, n_2, \dots, n_k)'$, we let $N = \sum_{i=1}^k n_i$ and assume $\mathbf{n} \rightarrow \infty$ indicates $n_i \rightarrow \infty$ such that $\frac{n_i}{N} \rightarrow \gamma_i (> 0)$ for each i .

If we define the information matrix as

$$-E \left[\frac{1}{N} \cdot \frac{\partial^2 \ln L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] = I_n(\boldsymbol{\theta}), \quad (3.8)$$

then, $I_n(\boldsymbol{\theta}) \rightarrow I(\boldsymbol{\theta}) = \sum_{i=1}^k \gamma_i I_i^*(\boldsymbol{\theta})$ as $\mathbf{n} \rightarrow \infty$.

From the large sampling theory (see Section 4.4.2 of Serfling (1980) [23]), the unconstrained estimator, $\hat{\boldsymbol{\theta}}$, of $\boldsymbol{\theta}$ is asymptotically normal as summarized in the following theorem.

Theorem 3.1 *Let $\hat{\boldsymbol{\theta}}$ be the unconstrained estimator obtained by the Newton's method as in Section 3.2.1. Then, we have*

- i) $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, I^{-1}(\boldsymbol{\theta}))$, as $\mathbf{n} \rightarrow \infty$
 - ii) $[l(\hat{\boldsymbol{\theta}}) - l(\boldsymbol{\theta})] - \frac{1}{2}N(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' I(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{p} 0$, as $\mathbf{n} \rightarrow \infty$
 - iii) $2[l(\hat{\boldsymbol{\theta}}) - l(\boldsymbol{\theta})] \xrightarrow{d} \chi_{2(k+1)}^2$, as $\mathbf{n} \rightarrow \infty$,
- where $N = \sum_{i=1}^k n_i$ and $I(\boldsymbol{\theta}) = \lim_{n \rightarrow \infty} I_n(\boldsymbol{\theta})$.

This theorem implies $\hat{\boldsymbol{\theta}}$ is strongly consistent. From this fact, we can also show the strong consistency of the constrained estimator $\boldsymbol{\theta}^*$.

By Taylor expansion of $l(\boldsymbol{\theta})$ about $\hat{\boldsymbol{\theta}}$ and then by the fact that $l'(\boldsymbol{\theta}) = \mathbf{0}$,

$$\begin{aligned} l(\hat{\boldsymbol{\theta}}) - l(\boldsymbol{\theta}) &= \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' [-l''(\tilde{\boldsymbol{\theta}})](\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \\ &= \frac{N}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \left[-\frac{l''(\tilde{\boldsymbol{\theta}})}{N} \right] (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \end{aligned} \quad (3.9)$$

for some $\tilde{\boldsymbol{\theta}}$ between $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$. Let $W_n(\tilde{\boldsymbol{\theta}}) = -\frac{l''(\tilde{\boldsymbol{\theta}})}{N}$ for convenience. Since $\boldsymbol{\theta}^*$ minimizes the quadratic function in (3.3), it follows that $\boldsymbol{\theta}^* = P_{W_n(\tilde{\boldsymbol{\theta}})}(\hat{\boldsymbol{\theta}}|K)$ for some $\tilde{\boldsymbol{\theta}}$ between $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$ where $K = \{\boldsymbol{\theta} \in \mathbf{R}^{2(k+1)} : A\boldsymbol{\theta} \leq \mathbf{0}\}$. Since $\hat{\boldsymbol{\theta}}$ is strongly consistent and $\tilde{\boldsymbol{\theta}}$ is between $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$, $\tilde{\boldsymbol{\theta}} \rightarrow \boldsymbol{\theta}$ almost surely as $\mathbf{n} \rightarrow \infty$. This implies $W_n(\tilde{\boldsymbol{\theta}}) \rightarrow I(\boldsymbol{\theta})$ almost surely. Thus, for any $\boldsymbol{\theta} \in K$ and any $\varepsilon > 0$, there exists $\mathbf{n}^0 = (n_1^0, \dots, n_k^0)'$ such that $\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|_{W_n(\tilde{\boldsymbol{\theta}})}^2 \leq \frac{\varepsilon}{2}$ almost surely for all $\mathbf{n} \geq \mathbf{n}^0$. Since $\boldsymbol{\theta}^* = P_{W_n(\tilde{\boldsymbol{\theta}})}(\hat{\boldsymbol{\theta}}|K)$, it follows that

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_{W_n(\tilde{\boldsymbol{\theta}})}^2 \leq \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|_{W_n(\tilde{\boldsymbol{\theta}})}^2 \leq \frac{\varepsilon}{2} \quad \text{almost surely for all } \mathbf{n} \geq \mathbf{n}^0.$$

Now, by the triangular law, we have

$$\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|_{W_n(\tilde{\boldsymbol{\theta}})}^2 \leq \|\boldsymbol{\theta}^* - \hat{\boldsymbol{\theta}}\|_{W_n(\tilde{\boldsymbol{\theta}})}^2 + \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_{W_n(\tilde{\boldsymbol{\theta}})}^2 \leq \varepsilon \quad \text{almost surely for all } \mathbf{n} \geq \mathbf{n}^0.$$

Therefore, by letting $\varepsilon \rightarrow 0$, $\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\|_{W_n(\tilde{\boldsymbol{\theta}})}^2 \rightarrow 0$ almost surely as $\mathbf{n} \rightarrow \infty$.

3.4 Likelihood Ratio Tests

The usual ANOVA test for H_0 against $H_2 - H_0$, which is based on the statistics $T_{02} = 2[l(\hat{\boldsymbol{\theta}}) - l(\bar{\boldsymbol{\theta}})]$, is a well known Chi-square test because $T_{02} \xrightarrow{d} \chi_{2(k-1)}^2$ as

$\mathbf{n} \rightarrow \infty$ under H_0 . The LRT statistic for testing H_0 versus $H_1 - H_0$ is $T_{01} = 2[l(\hat{\boldsymbol{\theta}}^*) - l(\hat{\boldsymbol{\theta}})]$ while the LRT statistic for testing H_1 versus $H_2 - H_1$ is $T_{12} = 2[l(\hat{\boldsymbol{\theta}}) - l(\hat{\boldsymbol{\theta}}^*)]$. We shall derive their limiting null distributions because the finite-sample distributions for these LRT statistics are completely unknown.

By the large sampling theory to the log-likelihood function (refer to Section 4.4 of Serfling (1980) [23]), it follows that $2[l(\hat{\boldsymbol{\theta}}) - l(\boldsymbol{\theta})] \xrightarrow{d} \chi_{2k+2}^2$ as $\mathbf{n} \rightarrow \infty$. In the similar way, we can show that the asymptotic null distributions of T_{01} and T_{02} are a mixture of several chi-square distributions as stated in the following theorem.

Theorem 3.2 *Let T_{01} and T_{12} be the LRT statistics for H_0 against $H_1 - H_0$ and H_1 against $H_2 - H_1$, respectively. Then, for any $\boldsymbol{\theta}_0 \in H_0$ and $t \geq 0$, we have*

- i) $\lim_{\mathbf{n} \rightarrow \infty} P[T_{01} > t] = \sum_{j=0}^{2(k-1)} W(j, 2(k-1); [AI^{-1}(\boldsymbol{\theta}_0)A']^{-1}) P[\chi_j^2 > t]$
 - ii) $\lim_{\mathbf{n} \rightarrow \infty} P[T_{12} > t] = \sum_{j=0}^{2(k-1)} W(j, 2(k-1); [AI^{-1}(\boldsymbol{\theta}_0)A']^{-1}) P[\chi_{2(k-1)-j}^2 > t]$
- where χ_j^2 is a standard chi-square random variable with j degrees of freedom ($\chi_0^2 \equiv 0$) and $W(j, m; \Sigma^{-1})$ is the probability that $P_{\Sigma^{-1}}(\mathbf{Z}|O)$ with $\mathbf{Z} \sim N(\mathbf{0}, \Sigma)$ and $O = \{\mathbf{x} \in \mathbf{R}^m : x_i \geq 0, i = 1, 2, \dots, m\}$ has exactly j positive component values (and hence, $\sum_{j=0}^m W(j, m; \Sigma^{-1}) = 1$ for any m and Σ^{-1}).

Proof: Let $D(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}) = N(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' I(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$. Then, by strong consistency of $\hat{\boldsymbol{\theta}}$ and the property in ii) of Theorem (3.1), we have $2[l(\hat{\boldsymbol{\theta}}) - l(\boldsymbol{\theta})] - D(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}) \xrightarrow{p} 0$, as $\mathbf{n} \rightarrow \infty$. Let $\boldsymbol{\theta}^0$ and $\boldsymbol{\theta}^1$ denote the least squares estimators obtained by minimizing $D(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}})$ under H_0 and H_1 , respectively. That is, $\boldsymbol{\theta}^0 = P_{I(\hat{\boldsymbol{\theta}})}(\hat{\boldsymbol{\theta}}|K_0)$ and $\boldsymbol{\theta}^1 = P_{I(\hat{\boldsymbol{\theta}})}(\hat{\boldsymbol{\theta}}|K)$ where $K_0 = \{\boldsymbol{\theta} \in \mathbf{R}^{2(k+1)} : A\boldsymbol{\theta} = \mathbf{0}\}$ and $K = \{\boldsymbol{\theta} \in \mathbf{R}^{2(k+1)} : A\boldsymbol{\theta} \leq \mathbf{0}\}$. Since $\boldsymbol{\theta}_0 \in K_0$, from the property of projection operator, it follows that

$$P_{I(\hat{\boldsymbol{\theta}})}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0|K_0) = P_{I(\hat{\boldsymbol{\theta}})}(\hat{\boldsymbol{\theta}}|K_0) - \boldsymbol{\theta}_0, \text{ and}$$

$$P_{I(\hat{\theta})}(\hat{\theta} - \theta_0|K) = P_{I(\hat{\theta})}(\hat{\theta}|K) - \theta_0.$$

Now,

$$\begin{aligned} D(\theta^0; \hat{\theta}) - D(\theta^1; \hat{\theta}) &= N(\hat{\theta} - \theta^0)' I(\hat{\theta})(\hat{\theta} - \theta^0) - N(\hat{\theta} - \theta^1)' I(\hat{\theta})(\hat{\theta} - \theta^1) \\ &= N(\theta^1 - \theta^0)' I(\hat{\theta})(\theta^1 - \theta^0) \quad (\because (\hat{\theta} - \theta^1, \theta^0)_{I(\hat{\theta})} = 0) \\ &= N[P_{I(\hat{\theta})}(\hat{\theta}|K) - P_{I(\hat{\theta})}(\hat{\theta}|K_0)]' I(\hat{\theta})[P_{I(\hat{\theta})}(\hat{\theta}|K) - P_{I(\hat{\theta})}(\hat{\theta}|K_0)] \\ &= N[P_{I(\hat{\theta})}(\hat{\theta} - \theta_0|K) - P_{I(\hat{\theta})}(\hat{\theta} - \theta_0|K_0)]' I(\hat{\theta}) \\ &\quad [P_{I(\hat{\theta})}(\hat{\theta} - \theta_0|K) - P_{I(\hat{\theta})}(\hat{\theta} - \theta_0|K_0)]. \end{aligned}$$

Since $\alpha P_W(\mathbf{x}|C) = P_W(\alpha \mathbf{x}|C)$ for any $\alpha \geq 0$ and cone C , it follows that

$$\begin{aligned} D(\theta^0; \hat{\theta}) - D(\theta^1; \hat{\theta}) &= \left[P_{I(\hat{\theta})}(\sqrt{N}(\hat{\theta} - \theta_0)|K) - P_{I(\hat{\theta})}(\sqrt{N}(\hat{\theta} - \theta_0)|K_0) \right]' I(\hat{\theta}) \\ &\quad \left[P_{I(\hat{\theta})}(\sqrt{N}(\hat{\theta} - \theta_0)|K) - P_{I(\hat{\theta})}(\sqrt{N}(\hat{\theta} - \theta_0)|K_0) \right]. \end{aligned}$$

By the asymptotic normality of $\hat{\theta}$ (see Theorem 3.1) and the continuity property of the projection operator,

$$\begin{aligned} D(\theta^0; \hat{\theta}) - D(\theta^1; \hat{\theta}) &\xrightarrow{d} [P_{I(\hat{\theta})}(\mathbf{U}|K) - P_{I(\hat{\theta})}(\mathbf{U}|K_0)]' I(\theta_0) \\ &\quad [P_{I(\hat{\theta})}(\mathbf{U}|K) - P_{I(\hat{\theta})}(\mathbf{U}|K_0)] \\ &\text{as } \mathbf{n} \rightarrow \infty \end{aligned}$$

where $\mathbf{U} \sim N(\mathbf{0}, I^{-1}(\theta_0))$.

Now, let $\mathbf{Z} = A\mathbf{U}$ with $m = 2(k-1)$, then $\mathbf{Z} \sim N(\mathbf{0}, AI^{-1}(\theta_0)A')$. Since $T_{01} - [D(\theta^0; \hat{\theta}) - D(\theta^1; \hat{\theta})] \xrightarrow{p} 0$, it follows by Theorem 2.4 in Section 2.2.3 that T_{01} has the asymptotic null distribution as given in the theorem. The derivation of

the asymptotic distribution of T_{12} under H_0 can be proved similarly. $\square$

In Theorem 3.2, the null distributions depend on the unknown parameter θ_0 through $I(\theta_0)$. However, we may use its estimates obtained during estimation procedures. In general, it is recommended to use $\widehat{I(\theta_0)} = -\frac{1}{N}l''(\theta^*)$ for T_{01} and $\widehat{I(\theta_0)} = -\frac{1}{N}l''(\hat{\theta})$ for T_{12} similarly to Wald's statistic.

The weights, $W(j, m; \Sigma^{-1})$, $j = 0, 1, \dots, m$, are generally unknown unless $k \leq 4$. But, they can be easily estimated from a simple simulation procedure as follows.

Step 1 : Generate $\mathbf{Z}$ from $N(\mathbf{0}, \Sigma)$.

Step 2 : Find the least squares projection $P_{\Sigma^{-1}}(\mathbf{Z}|O)$.

Step 3 : Count the number of strictly positive component values in $P_{\Sigma^{-1}}(\mathbf{Z}|O)$ and then let the count be denoted by M_i .

Step 4 : Repeat Steps 1 - 3 for $i = 1, 2, \dots, M$.

Then $W(j, m; \Sigma^{-1}) \doteq \frac{\#\{i: M_i=j\}}{M}$, $j = 0, 1, \dots, m$.

Since the Chi-bar-square distributions are not sensitive to their weights, this simulation approach is widely accepted in practical situation.

Chapter 4

Application to RE Tumor Data

4.1 The RE Tumor Data

We consider data from a lifetime feeding experiment involving mice received Red Dye No. 40, a colour additive widely used in foods in the U.S. We want to assess whether Red 40 had any effect on the development of reticuloendothelial (RE) tumors and lifetimes. It was generally believed that RE tumors kill their mouse hosts shortly after onset and hence that time to RE death approximates time to RE tumor onset. This data set was originally discussed in Lagakos and Mosteller [14] in 1981. However, we use the data set provided in Table 42.1 of Andrews and Herzberg (1985)[2] for this experiment. In this study, there were four treatments groups, a control group and three dose level groups, which was set at low, medium or high. For each mouse, one of the four treatments were randomly assigned to it. After each mouse's death, the presence or absence of RE tumors and time to death (in weeks) were recorded. Thus, we have bivariate responses, with one

binary and the other continuous. Summary statistics for these data by dosage level groups are provided in Table 4.1 below. Please note that we have switched the notation originally used in Table 42.1 of Andrews and Herzberg (1985) [2]. That is, we used 0 for presence of tumor and 1 for absence of tumor. This is just to illustrate our model in Chapter 3, with positive correlation between the two outcomes.

Table 4.1: Summary Statistics for Red Dye No.40 Data

Dosage Level	Sample Size	Deaths without Tumor (%)	Mean Time to Death (Weeks)
Control	33	72.73	79.67
Low	30	73.33	75.83
Medium	31	74.19	80.71
High	28	64.29	68.50

Since the treatments in this study are ordered according to dosage level, one may naturally expect that both the proportion of normal deaths without tumors and the mean life time will be non-increasing with the amount of Red 40 applied. This pattern can be seen from the summary statistics by treatment. Thus, any statistical inference applied to assess the effect of Red 40 on the development of RE tumors should acknowledge this natural ordering in the treatment levels.

4.2 Model and Maximum Likelihood Estimates

In the model introduced in Chapter 3, the binary variable, X_1 , represents the presence or absence of an RE tumor at the time of death, and the continuous variable, X_2 , represents the time to death. We shall assume that the time to death is normally distributed for illustrating the proposed methodologies. Since the Red Dye No.40 data set is comprised of mice that died over the course of the experiment(104 weeks), mice that survived to week 104 were killed as planned. A few cases of this censoring were assumed to be complete observations. However, the data set we used in Andrews and Herzberg (1985) [2] has already excluded those censored cases.

As we mentioned earlier, we expect certain ordering among the μ_{i1} 's and μ_{i2} 's. Both the RE tumor absence proportion and the mean life time should be nonincreasing with the dosage level of Red 40. Thus, it is reasonable to test H_0 against $H_1 - H_0$ for simultaneous presence of ordered group effects on the two responses, and test H_1 against $H_2 - H_1$ for the goodness-of-fit of the ordering assumption, where H_0 , H_1 and H_2 are specified as follows:

$$H_0 : \mu_{11} = \mu_{21} = \cdots = \mu_{k1}, \mu_{12} = \mu_{22} = \cdots = \mu_{k2}$$

$$H_1 : \mu_{11} \geq \mu_{21} \geq \cdots \geq \mu_{k1}, \mu_{12} \geq \mu_{22} \geq \cdots \geq \mu_{k2}$$

$$H_2 : \text{No restriction on } \boldsymbol{\mu}.$$

Using the model described in Chapter 3 and Dykstra's least squares projection algorithm [9], we obtained the parameter estimates under each hypothesis. Table

4.1 lists all the parameter estimates.

Table 4.2: Parameter Estimates under Each Hypothesis.

Parameters	Estimates		
	under H_0	under H_1	under H_2
μ_{11}	0.5628	0.6252	0.5951
μ_{21}	0.5628	0.6252	0.6251
μ_{31}	0.5628	0.6201	0.6532
μ_{41}	0.5628	0.3703	0.3703
μ_{12}	76.4262	79.7143	79.6667
μ_{22}	76.4262	78.2857	75.8333
μ_{32}	76.4262	78.2857	80.7097
μ_{42}	76.4262	68.4999	68.5000
σ^2	614.6216	595.5862	592.6138
ρ	0.1333	0.1154	0.1157

Estimates of p_i can be obtained by equation $p_i = 1 - \Phi(-\mu_{i1})$ under each of the three hypotheses, which are shown in Table 4.3.

4.3 Test Statistics

Recall the LRT statistics for testing H_0 versus $H_2 - H_0$, H_0 versus $H_1 - H_0$, and H_1 versus $H_2 - H_1$ are $T_{02} = 2[l(\hat{\boldsymbol{\theta}}) - l(\bar{\boldsymbol{\theta}})]$, $T_{01} = 2[l(\boldsymbol{\theta}^*) - l(\bar{\boldsymbol{\theta}})]$, and $T_{12} = 2[l(\hat{\boldsymbol{\theta}}) - l(\boldsymbol{\theta}^*)]$, respectively. By the parameter estimates, we obtained these test statistics: $T_{02} = 4.993576$, $T_{01} = 4.356328$ and $T_{12} = 0.637248$. Let W_{01} and W_{12} denote the weight vectors for the test statistics T_{01} and T_{12} , respectively.

Table 4.3: Estimates of p_i under Each Hypothesis.

Parameters	Estimates		
	under H_0	under H_1	under H_2
p_1	0.7132	0.7341	0.7241
p_2	0.7132	0.7341	0.7340
p_3	0.7132	0.7324	0.7432
p_4	0.7132	0.6444	0.6444

The simulation for the weights using appropriate estimate for θ gave us the weight vectors in the calculation for p-value. The weight vectors are presented in Table 4.4. Then by Theorem 3.2 we got the associated p-values for T_{01} and T_{12} . Recall

Table 4.4: Simulation Results for Weight Vectors

Number of positive component values	Weight Vectors	
	W_{01}	W_{12}
0	0.0021	0.0021
1	0.0529	0.0430
2	0.2350	0.2280
3	0.5238	0.5471
4	0.1453	0.1405
5	0.0389	0.0372
6	0.0020	0.0021

that $T_{02} \sim \chi^2_{2(k-1)}$, we also calculated the associated p-value for T_{02} . As we can see from Table 4.5, the formal test of goodness-of-fit of the ordering assumption has associated p-value of 0.8433, which strongly supports the assumption of the

Table 4.5: Test Statistics and Associated p-value

Test Statistics	Associated p-value
$T_{01} = 4.356328$	0.2196
$T_{12} = 0.637248$	0.8433
$T_{02} = 4.993576$	0.5446

restricted parameter space defined by H_1 . Contrast to the p-value ($= 0.5446$) of the usual unrestricted test by testing H_0 versus $H_2 - H_0$, the inferential assessment of the simultaneous presence of ordered group effects on the two responses conducted by testing H_0 versus $H_1 - H_0$ leads to the p-value of 0.2196.

Neither the usual unrestricted test nor the order restricted test reject H_0 at usual significance levels, however, they differ in the strength of supporting H_0 and the test based on T_{01} is more powerful in the presence of ordering effect. Remember that the main purpose of the Red Dye No. 40 experiment is to assess whether Red 40 had any effect on the development of reticuloendothelial (RE) tumors. The null hypothesis suggests that there is ‘no harmful effect’, while the alternative hypothesis H_1 suggests that there is ‘non-decreasing risk of harmful effect’. So, comparing to Type I error, we have greater sensitivity towards Type II error here. With the p-value of 0.2196 associated with the ordered restricted test, an investigator might suspect there is non-negligible evidence that as the dosage level of Red Dye No. 40 increase, the mean time to live or the absence rate of RE tumor, or both, would be decreasing. But, such a conclusion could not be expected from the p-value of 0.5446 based on T_{02} of the usual unrestricted test.

Chapter 5

Conclusion and Generalizations

In this thesis, we present likelihood ratio tests for and against ordered group effects in binary and continuous outcomes. Specifically, hypotheses tests for ordered group effects and the goodness-of-fit of the ordering assumption are implemented that are based on a joint distribution model using the concept of the latent variable. There are several ways in which the testing hypothesis of linear inequality problem can be generalized by relaxing the assumptions. In the model introduced in Chapter 4, the marginal model for the continuous response was assumed to be normally distributed. This model could be extended to a larger class of continuous distributions using probability integral transformation. However, the usual regularity conditions for the density function are required in order to apply the large sample theory regarding the information matrix as discussed in Section 3.3. Details of these conditions can be found in Serfling (1980) [23]. We might also consider introducing another latent variable for the continuous outcome. Order restricted tests based on univariate responses are rigorously discussed in Robert-

son et al. [22]. In these tests, one of H_0 and H_1 is expressed as a closed convex cone of the parameters, while the other is a subspace. The null distribution are a mixtures of chi-square or beta distributions depending on the distributional assumptions (see Shapiro [24] for more details).

The discussion given in this thesis has primarily been concerned with likelihood ratio tests. Other tests are possible. Wald type tests and Lagrange multiplier tests are two alternative tests. However, they are identical to the LRT if the information matrix is known and asymptotically equivalent otherwise. Robertson et. al (1988) [22] discussed the topic of Lagrange multiplier tests.

In the context of the methodology presented in this thesis for bivariate data, one might want to do a marginal analysis for testing H_0 versus $H_1 - H_0$ on the binary outcome where

$$H_0 : \mu_{11} = \mu_{21} = \cdots = \mu_{k1}, \mu_{12} \geq \mu_{22} \geq \cdots \geq \mu_{k2}$$

and

$$H_1 : \mu_{11} \geq \mu_{21} \geq \cdots \geq \mu_{k1}, \mu_{12} \geq \mu_{22} \geq \cdots \geq \mu_{k2}.$$

Unfortunately, the development of an asymptotic distribution theory for such a test is challenging even under the normality assumption with known covariance matrix. In this case, the asymptotic p-value can only be determined via a simulation study with sufficient replications.

Chapter 6

Appendix

This chapter contains an implementation in R of the algorithms presented in this thesis. The first two programs are based on Dykstra's iterative least squares projection algorithm [9]. The third program is for the application of RE tumor data set, it includes Newton's method for maximum likelihood estimators under each hypothesis assumption and simulation for the weight vectors.

```
# min (g-x)'w(g-x), s.t. Ax<=0,
# simple least square projection with constrain matrix A,
# r number of linear inequality constraints, n size of vector x
# w is a n by n diagonal weight matrix
# g is a given point in R^{n}, eps constrols the stopping condition

ls.proj=function(A,r,n,w,g,eps){
  incmt=rep(0,n)          #increment for each projection
  first=rep(0,n)          #first increment of each cycle
  gprev=rep(0,n)          #previous projectoin onto the rth cone
  gnow=rep(0,n)           #current projection onto the rth cone
  gnew=rep(0,n)           #projection on each cone
  d=1                     #distance between gnow and gprev

  while(d > eps){
    gprev=gnow
    counter=0
    # finds sequentially the least square projection of g onto K_i
```

```

# K_i = {x; A[i,]' x <= 0}
for(i in 1:r){
  for(j in 1:n){
    if((t(A[i,])%*%g) <= 0){
      gnew=g
    }
    else{
      gnew[j]=g[j]-(((t(g)%*%A[i,])*A[i,j]*((solve(w))[j,j]))/(t(A[i,])%*%
        (solve(w))%*%A[i,]))
    }
  }
  incmt=gnew-g
  #record the first increment of each cycle
  if(counter==0) {
    first=incmt
    counter=1
  }
  g=gnew
}
gnow=g
d=sqrt((t(gnow-gprev))%*%w%*(gnow-gprev))
# remove the first increment before a new iteration cycle
g=gnew-first
}
gnow
}

# least square projection with Mahalanobis distance which calls the function
# ls.proj and return the point that solves
# min(g-x)'F(g-x) s.t Ax<=0,
# F is a given, symmetric,postive definitive weight matrix
# F=P'DP, where D is the diagonal matrix consisting of the eigenvalues of F
# A is the r x n constrain matrix
# f is the n x n weight matrix, g the given point,
# eps constrols the stopping condition
m.proj=function(A,r,n,f,g,eps){
  Dmatrix=diag(eigen(f)$values) # find the diagonal matrix D
  P=eigen(f)$vectors           # find the corresponding eigen vector P
  A2=A%*%P                     # transform the constrain matrix into A2
  pg=t(P)%*%g                  # transform the given point to P'g
  # use ls.proj to solve min (pg - y)'Dmatrix (pg - y), s.t A2'y <= 0
  y.star=ls.proj(A2,r,n,Dmatrix,pg,eps)
  x.star=P%*%y.star
  x.star
}

#read the RED 40 data set
red=read.table("E:/carleton/thesis/RED.DAT")

```

```

dat=cbind(red$V4,red$V2)
x1=dat[1:33,]
x2=dat[34:63,]
x3=dat[64:94,]
x4=dat[95:122,]
n1=33
n2=30
n3=31
n4=28
x=list(x1,x2,x3,x4)
k=4
N=122
n=c(n1,n2,n3,n4)
theta=c(0,0,0,0,79.667,75.833,80.7097,68.5,619,0.2)
A=matrix(0,6,10)
#construct constrain matrix A
for(i in 1:3){
  A[i,i]=-1
  A[i,i+1]=1
}
for(i in 4:6){
  A[i,i+1]=-1
  A[i,i+2]=1
}
#find the first order derivatives of the log-likelihood function
lder=function(theta){
  z=list(rep(0,n[1]),rep(0,n[2]),rep(0,n[3]),rep(0,n[4]))
  a=list(rep(0,n[1]),rep(0,n[2]),rep(0,n[3]),rep(0,n[4]))
  l1=list(rep(0,n[1]),rep(0,n[2]),rep(0,n[3]),rep(0,n[4])) #lmui1
  l2=list(rep(0,n[1]),rep(0,n[2]),rep(0,n[3]),rep(0,n[4])) #lmui2
  l3=list(rep(0,n[1]),rep(0,n[2]),rep(0,n[3]),rep(0,n[4])) #lsigma^2
  l4=list(rep(0,n[1]),rep(0,n[2]),rep(0,n[3]),rep(0,n[4])) #lrho

  for(i in 1:k){
    for(j in 1:n[i]){
      z[[i]][j]=(x[[i]][j,2]-theta[i+k])/(theta[9]^(1/2))
      a[[i]][j]=(theta[i]+theta[10]*z[[i]][j])/((1-theta[10]^2)^(1/2))
      if(x[[i]][j,1]==1){
        l1[[i]][j]=(dnorm(a[[i]][j])/pnorm(a[[i]][j]))/((1-theta[10]^2)^(1/2))
        l2[[i]][j]=(theta[9]^(-1/2))*
          (z[[i]][j]-((dnorm(a[[i]][j])/pnorm(a[[i]][j]))*theta[10]*
            (1-theta[10]^2)^(-1/2)))
        l3[[i]][j]=((1/2)/theta[9])*(-1+(z[[i]][j]^2)-z[[i]][j]*
          ((dnorm(a[[i]][j])/pnorm(a[[i]][j]))*theta[10]/((1-theta[10]^2)^(1/2))))
        l4[[i]][j]=(dnorm(a[[i]][j])/pnorm(a[[i]][j]))*(z[[i]][j]/
          ((1-theta[10]^2)^(1/2))+ (theta[10]*a[[i]][j]/(1-theta[10]^2)))
      }
      if(x[[i]][j]==0){

```

```

11[[i]][j]=-(dnorm(a[[i]][j])/pnorm(-a[[i]][j]))/((1-theta[10]^2)^(1/2))
12[[i]][j]=(theta[9]^(-1/2))*
  (z[[i]][j]+((dnorm(a[[i]][j])/pnorm(-a[[i]][j]))*theta[10]*
  (1-theta[10]^2)^(1/2)))
13[[i]][j]=((1/2)/theta[9])*(-1+(z[[i]][j]^2)+z[[i]][j]*
  ((dnorm(a[[i]][j])/pnorm(-a[[i]][j]))*theta[10]/((1-theta[10]^2)^(1/2))))
14[[i]][j]=-(dnorm(a[[i]][j])/pnorm(-a[[i]][j]))*(z[[i]][j]/
  ((1-theta[10]^2)^(1/2))+ (theta[10]*a[[i]][j]/(1-theta[10]^2)))
}
}
}
lprime=c(sapply(11,sum),sapply(12,sum),sum(sapply(13,sum)),sum(sapply(14,sum)))
lprime
}
delta=10^(-8)
#Approximate the hessian matrix of the log-likelihood function by
# (f(x+delta)-f(x))/delta, with delta sufficiently small
hess=function(theta,delta){
  hessian=matrix(0,10,10)
  theta.m=matrix(0,10,10)
  for(i in 1:10){
    theta.m[,i]=theta
  }
  diag(theta.m)=diag(theta.m)+delta
  for(i in 1:10){
    for(j in 1:10){
      hessian[i,j]=(lder(theta.m[,j])[i]-lder(theta)[i])/delta
    }
  }
  hessian
}
d=1
theta.new=theta
while(d>10^(-6)){
  theta=theta.new
  theta.new=theta-(solve(hess(theta,delta)))*%lder(theta)
  d=sqrt((t(theta-theta.new))%*(theta-theta.new))
}
theta.hat=theta.new          #unconstrained mle theta.hat
theta.hat

##### iterated Newton's method for theta.star #####
d2=1
theta.n=c(0.625,0.625,0.625,0.625,79.7,78,78,68.5,592.6,0.1149)
iter=1
for(iter in 1:50) {
  theta.star=theta.n
  f= -hess(theta.star,delta)

```

```

# theta.t is \tilde{\theta}, theta.n is the new updated theta.star
theta.t = theta.star + solve(f)%%lder(theta.star)
theta.n = m.proj(A,6,10,f,theta.t,10^(-6))
d2=sqrt((t(theta.n-theta.star))%*f%(theta.n-theta.star))
iter=iter+1
}
theta.star=theta.n

##### estimator under Ho (theta.bar) #####
#least square projection of theta.hat      n-steps est.
d3=1
delta=10^(-6)
theta.m=c(0.562,0.562,0.562,0.562,76.426,76.426,76.426,76.426,592.659,0.117)
while(d3>10^(-8)) {
  theta.bar=theta.m
  f= -hess(theta.bar,delta)  #weight matrix
  theta.tilde = theta.bar + solve(f)%%lder(theta.bar)
  theta.m=theta.tilde-solve(f)%%t(A)%%solve(A)%%solve(f)%%t(A))%*%
    A)%%theta.tilde
  d3=sqrt((t(theta.m-theta.bar))%*f%(theta.m-theta.bar))
}
theta.bar=theta.m

##### simulation for the weights #####
library(MASS) #load package MASS
weight = function(A,f,r,n,N){
  w=rep(0,r+1)
  mu = rep(0,r)
  inverse = solve(f/n)
  variance = A)%%inverse)%%t(A)
  v.inverse = solve(variance)
  B=diag(1,r,r)
  fre=rep(0,r+1)
  for ( i in 1:N){
    y = mvrnorm(1,mu,variance)
    y.star=m.proj(B,r,r,v.inverse,y,10^(-10))
    count=0
    for(j in 1:r){
      if (y.star[j] > -10^(-10)){
        count=count + 1
      }
    }
    for(j in 0:r){
      if (count == j){
        fre[j+1]=fre[j+1]+1
      }
    }
  }
}

```

```

w=fre/N
w
}
f.hat=-hess(theta.hat,10^(-7))
w12=weight(A,f.hat,6,122,10000)
f.star=-hess(theta.star, 10^(-7))
w01=weight(A,f.star,6,122,10000)

##### calculate the test statistics #####
# parameter estimates under each hypothesis
theta.bar=c(0.5627606,0.5627606,0.5627606,0.5627606,
            76.4262295,76.4262295,76.4262295,76.4262295,614.6216071,0.1332521)
theta.star=c(0.6251647,0.6251647,0.6200710,0.3702848,
            79.7143454,78.2856918,78.2856918,68.4999320,595.5861952,0.1153976)
theta.hat=c(0.5951377,0.6250694,0.6532347,0.3703223,
            79.6666667,75.8333333,80.7096774,68.5000000,592.6138287,0.1157380)

# log-likelihood function, return the value of the log-likelihood
# x the data set, theta the parameter
# k number of groups, n a vector stores sample size for each group
lnL=function(x,theta,k,n){
  z=list(rep(0,n[1]),rep(0,n[2]),rep(0,n[3]),rep(0,n[4]))
  a=list(rep(0,n[1]),rep(0,n[2]),rep(0,n[3]),rep(0,n[4]))
  logg=list(rep(0,n[1]),rep(0,n[2]),rep(0,n[3]),rep(0,n[4]))
  for(i in 1:k){
    for(j in 1:n[i]){
      z[[i]][j]=(x[[i]][j,2]-theta[i+k])/(theta[9]^(1/2))
      a[[i]][j]=(theta[i]+theta[10]*z[[i]][j])/((1-theta[10]^2)^(1/2))
      logg[[i]][j]=(x[[i]][j,1]*log(dnorm(z[[i]][j])*pnorm(a[[i]][j])*
        (theta[9]^(-1/2))))+
        (1-x[[i]][j,1])*log(dnorm(z[[i]][j])*pnorm(-a[[i]][j])*(theta[9]^(-1/2))))
    }
  }
  lnLikelihood=sum(sapply(logg,sum))
  lnLikelihood
}

# calculate test statistics T01, T12 and T02
t01=2*(lnL(x,theta.star,k,n) - lnL(x,theta.bar,k,n))
cat("t01 = ",t01,"\n")

t12=2*(lnL(x,theta.hat,k,n) - lnL(x,theta.star,k,n))
cat("t12 = ",t12,"\n")

t02=2*(lnL(x,theta.hat,k,n) - lnL(x,theta.bar,k,n))
cat("t02 = ",t02,"\n")

#t01 = 4.356328

```

```
#t12 = 0.637248
# weight vector from simulation result
w01=c(0.0021, 0.0529, 0.2350, 0.5238, 0.1453, 0.0389, 0.0020)
w12=c(0.0021, 0.0430, 0.2280, 0.5471, 0.1405, 0.0372, 0.0021)

pvalue=function(k,w,tstat){
  pchi=rep(0,2*k-1)
  for(i in 1:(2*k-1)){
    pchi[i]=pchisq(tstat,i-1,lower.tail=FALSE)
  }
  pv=t(w)%*%pchi
  pv
}

p01=pvalue(4,w01,t01)
cat("p-value for T01=", p01,"\n")

p12=pvalue(4,w12,t12)
cat("p-value for T12=", p12,"\n")

p02=pchisq(t02,6,lower.tail=FALSE)
cat("p-value for T02=", p02,"\n")
```

Bibliography

- [1] Agresti, A., and Gueorguieva, R. V. (2001). A Correlated Probit Model for Joint Modeling of Clustered Binary and Continuous Responses, *Journal of the American Statistical Association*, **96**, 1102–1112.
- [2] Andrews, D. F. and Herzberg, A. M. (1985). *Data: A Collection of Problems from Many Fields for the Student and Research Worker*, Springer-Verlag, New York.
- [3] Barlow, R. E., Bartholomew D. J., Bremner, J. M., and Brunk, H. D. (1972). *Statistical Inference Under Order Restrictions: Theory and Application of Isotonic Regression*, John Wiley&Sons, London.
- [4] Bartholomew, D. J. (1959). A test of homogeneity for ordered alternatives, *Biometrika* **46**, 36–48.
- [5] Catalano, P. J., and Ryan, L. M. (1992), Bivariate Latent Variable Models for Clustered Discrete and Continuous Outcomes, *Journal of the American Statistical Association*, **87**, 651–658.
- [6] Cox, D. R. (1972). The analysis of multivariate binary data. *Applied Statistics*, **21**, 113–120.

- [7] Cox, D. R., and Wermuth, N. (1992), Response Models for Mixed Binary and Quantitative Variables, *Biometrika*, **79**, 441–461.
- [8] Dobler, Carolyn Pillers (2002), The one-way layout with ordered parameters: a survey of advances since 1988, *Journal of Statistical Planning and Inference*, **107**, 75–88.
- [9] Dykstra, Richard L. (1983), An Algorithm for Restricted Least Squares Regression, *Journal of the American Statistical Association*, **78**, 837–842.
- [10] Dykstra, R. L., Kochar, S., and Robertson, T. (1991), Statistical Inference for Uniform Stochastic Ordering in Several Populations. *The Annals of Statistics*, **19**, 870–888.
- [11] Farrell, P. J., and Park, C. G. (2007), Testing for Ordered Group Effects in Binary and Continuous Outcomes, *Biometrical Journal*, **49**, 585–598.
- [12] Fitzmaurice, G. M., and Laird, N. M. (1995), Regression Models for Discrete and Continuous Outcome with Clustering, *Journal of the American Statistical Association*, **90**, 845–852.
- [13] Kudô, A. (1963). A Multivariate analogue of the one-sided test, *Biometrika*, **50**, 403–418.
- [14] Lagakos, S. W. and Mosteller, F. (1981). A case study of statistics in the regulatory process: The FD&C Red No. 40 experiments, *Journal of the National Cancer Institute*, **66**, 197–212.

- [15] Lee, C. I. C. (1987). Chi-square tests for and against an order restriction on multinomial parameters. *Journal of the American Statistical Association*, **82**, 611–618.
- [16] Luenberger, D. G. (1984). *Linear and Nonlinear Programming*. Addison Wesley, Reading.
- [17] Ochi, Y. and Prentice, Ross L. (1984). Likelihood inference in a correlated probit regression model, *Biometrics*, **71**, 531–543.
- [18] Olkin, I. and Tate, R. F. (1961). Multivariate correlation models with mixed discrete and continuous variables, *Annals of Mathematical Statistics*, **32**, 448–465.
- [19] Perlman, M. D. (1969). One-sided testing problems in multivariate analysis. *The Annals of Mathematical Statistics*, **40**, 549–567.
- [20] Regan, M. M. and Catalano, P. J. (1999). Likelihood models for clustered binary and continuous outcomes: Application to developmental toxicology, *Biometrics*, **55**, 760–768.
- [21] Robertson, T. and Wright, F. T. (1981). Likelihood ratio tests for and against stochastic ordering between multinomial populations. *The Annals of Statistics*, **9**, 1248–1257.
- [22] Robertson, T., Wright, F. T., and Dykstra, R. L. (1988). *Order Restricted Inference*, Wiley, New York.

- [23] Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.
- [24] Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis, *International Statistical Review*, **56**, 49–62.
- [25] Silvapulle, M. J. and Sen, P. K. (2004). *Constrained Statistical Inference*, Wiley-Interscience, Hoboken, N.J.
- [26] Tate, R. F. (1954). Correlation between a Discrete and a Continuous Variable. *Annals of Mathematical Statistics*, **25**, 603–607.