UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche
Corso di Dottorato di Ricerca in Scienze Statistiche
Ciclo XXXII

Permutation tests for stochastic ordering with ordered categorical data

Coordinatore del Corso: Prof. Massimiliano Caporin

Supervisore: Prof. Luigi Salmaso

Co-supervisore: Prof. Fortunato Pesarin & Prof. Dabuxilatu Wang

Dottoranda: Huiting Huang

2nd December 2019

Abstract

Ordered categorical data are frequently encountered in many fields of research, such as, sociology, psychology, quality control, medical studies, and so forth. Especially in medical research, it is inevitable to meet a lot of problems containing ordered categorical data. Our specific interest is to find convincing solutions to some of the testing problems which include restrictions in the set of alternatives, such as testing for stochastic dominance and testing for monotonic stochastic ordering while using such a kind of data. When the number of nuisance parameters of underlying distributions or that of observed variables is small, there are some likelihood-based solutions. Our interest, however, is for cases where such numbers are not small. In these cases likelihood-based methods do not work, thus our interest is to proceed nonparametrically within permutation methods. Permutation methods are conditional on the pooled set of observed data which, in turn, are typically a set of sufficient statistics under the null hypothesis for the underlying distribution. Moreover, due to the evident complexity of such problems, according to Roy (1953), we also must use their Union-Intersection representation consisting on an equivalent break-down of the hypothesis under testing into a set of simpler sub-hypotheses for each of which a permutation test is available and such tests are jointly considered. So we must stay within the nonparametric combination of several dependent permutation tests. In the thesis, guided by two medical examples from the literature, we propose suitable solutions that are proved to be admissible combinations of optimal conditional partial tests and so enjoying good asymptotic properties.

Sommario

Dati di tipo categoriale ordinato si incontrano molto frequentemente in molti ambiti di ricerca, ad esempio: sociologia, psicologia, controllo della qualità, studi clinici, e così via. Specialmente nella ricerca medica è quasi invitabile incontrare un gran numero di problemi in cui i dati sono espressi con variabili di tipo categoriale ordinato. Il nostro specifico interesse è di trovare convincenti soluzioni ad alcuni problemi di verifica d'ipotesi che richiedono restrizioni sull'insieme delle alternative, come ad esempio test di dominanza stocastica e test di regressione monotona quando i dati disponibili sono appunto di tipo categoriale ordinato. Quando il numero di parametri di disturbo e/o quello delle variabili in gioco è piccolo vi sono disponibili, ancorché piuttosto problematiche, delle soluzioni desunte via likelihood. Il nostro interesse comunque è principalmente rivolto a situazioni in cui tali numeri non sono piccoli. In casi, cioè, in cui soluzioni via likelihood o non sono disponibili o non sono possibili. Perciò il nostro interesse è di procedere in modo non parametrico nell'ambito dei metodi di permutazione. I metodi di permutazione sono metodi inferenziali condizionati ai dati osservati che, tipicamente, sono un insieme di statistiche sufficienti sotto l'ipotesi nulla per la sottostante distribuzione, qualunque essa sia. Inoltre, per l'evidente complessità dei problemi, in accordo con Roy (1953) dobbiamo necessariamente adottare la loro rappresentazione di tipo Unione-Intersezione, consistente in una scomposizione delle ipotesi in un equivalente insieme di sotto-ipotesi per ognuna delle quali sia disponibile un test di permutazione e che tali test vengano processati congiuntamente. Per cui si dovrà operare nell'ambito della combinazione non parametrica di una pluralità di test di permutazione tra loro dipendenti. Nella tesi, guidati da due esempi di letteratura di ambito medico, vengono discusse alcune soluzioni a questo tipo di problemi che risultano combinazioni ammissibili di test parziali condizionatamente ottimi e perciò dotate di buone proprietà asintotiche.

Dedicated to my family

&

my supervisors

Acknowledgements

In presenting this thesis, I would first like to express my greatest gratitude to my supervisors Professor Luigi Salmaso and Professor Fortunato Pesarin, who have wisely guided me into the challenging field of permutation testing and have provided me with extraordinary support in this thesis process. I would like to thank Professor Luigi Salmaso for his support and expert advice, as well as Professor Fortunato Pesarin for his patience, encouragement, wisdom, and guidance throughout this journey of research. I would extend my thanks to Professor Dabuxilatu Wang, University of Guangzhou, who always helped me and encourages my studies during the Ph.D. period. Special thanks go to Professor Alessandra Brazzale who gently helped me during the first year of Ph.D., as well as Professor Nicola Sartori who, as a course coordinator during 2017-2018 of Ph.D., kindly helped me in solving administrative issues.

I would like to thank all professors who provided us with great and high quality of educational curriculums. Moreover, I wish to thank Patrizia Piacentini who tried her best to help us in the face of any administrative problems during the Ph.D. period.

I would like to thank my best colleagues from the 32nd Ph.D. cycle, for their support and wonderful collaboration from start to finish. Special thanks go to Ilaria Bussoli and Anastasiia Gorshechnikova, who helped me through the hard times at the beginning of Ph.D., and are always willing to help and encourage me.

I gratefully acknowledge the support and generosity of the University of Guangzhou, without which the present study could not have been completed.

I would like to thank my thesis reviewers, Professor Roberto Fontana and Professor Stefano Bonnini, for their precious comments and constructive recommendations which have led to improvements to my work.

I would also like to express my thanks to the friend Linjuan Chen, who guided and motivated me to apply for the Department of Statistical Sciences, University of Padova. I would extend my thanks to the friends Huijun Wen and Xiaomin Lai, who have been motivators during this long journey.

I would like to thank my friends I met during this journey, because there are you, this unfamiliar city has made me feel at home. We shared the most hilarious moments

during this journey, and I will always treasure our time spent together and our many precious memories. I would thank my friend Haitao Cao, who is always willing to help me. Special thanks go to my best housemates Qiaoling Wang, Yang Yang, Shiruo Wang, Lingzi Jiang, Qing Liu, I feel very lucky to spend our time together and have learned enormously from all of you. I would extend my thanks to my friends, they include but not limited to Xiuli Mao, Ju Cao, Guangyuan Liu, Qi Xiao, Fei Deng, Lishuai Jia.

I especially wish to express my thanks to my family, whose consistently love, support and encouragement provided me with strength, optimism, and confidence for the successful completion of the task. It has made me realize that family support is the driving force of my work, which helps me to stay active. I thank my brother for his concern and unconditional love. I wish to express my love and greatest gratitude to my parents for their support, understanding and unconditional love.

Contents

List of Figures

List of Tables

Preamble

Overview

Ordered categorical data are popular in many fields, such as psychology, medical studies, quality control, sociology, and so on. Such data are presented in the form of ordered categories, as for instance, {*Unqualified quality*, *Qualified quality*}; {*Unhappy*, *Neither happy nor unhappy*, *Happy*}; {*Death*, *Vegetative state*, *Major disability*, *Minor disability*, *Good recovery*}; and so forth. Various related experiments are designed to analyze the superior or inferior rank of comparable groups, and the corresponding sample sizes are always small or moderate. Such problems incorporating ordered categorical data are still challenging for researchers, especially with testing under constrained alternatives.

Most researchers tend to use likelihood approaches because, in general, the related solutions are provided with nice statistical inferential properties, but they must be based on too stringent assumptions including normality or other specific distributions assumptions of underlying populations, separable nuisance entities, etc. The stringent assumptions may be set up for one reason or another, such as researchers' prior information; statistical inference experiences; a reduction in the cost of computation; and so on. When problems involve the comparison of several groups or the data are multivariate distributed, in such a setting, the number of underlying parameters/nuisance parameters, and/or the number of observed variables can often be much larger than sample sizes, the corresponding solutions within likelihood frameworks, when available become an extremely difficult task. Therefore, unless there are clearly reasonable assumptions allowing for a considerable reduction of underlying complexity, the most interesting is that when one pseudo-parameter is represented as a function of many underlying nuisance parameters, there is no general correct testing solution within that approach. The likelihood approaches may provide questionable but interpretable results, and for more general and complex situations, there is no simple closed-form expression for the solution.

In the face of different or complicated situations, such as, the lack of underlying distribution's information, the existence of multidimensional variables, the existence of unknown dependence relations between variables, and so on, compared with parametric/semi-parametric methods, it seems to be more suitable for solving the problems by nonparametric approaches, which in turn are based on mild assumptions. Among nonparametric approaches, the nonparametric combination (NPC) of dependent permutation tests which work within the principles of sufficiency and conditionality, could provide admissible solutions for many complex situations that are also asymptotically coincident with the optimal ones when these exist. According to Roy's Union-Intersection approach, the NPC testing procedure can properly and equivalently breakdown the global problem into a set of sub-problems, each of which provided with proper partial permutation tests, the corresponding global result is obtained by jointly analyzed all of them. In this thesis, we tried to analyze some testing problems incorporating with ordered categorical data.

The rest of the thesis is organized as follows. Chapter 1 reviews the literature on the ordered constraint problems with the main focus on categorical ordinal variables. Chapter 2 introduces the notions of the permutation approach and some relevant definitions and properties. Chapter 3 discusses the two-sample problem for unidimensional and multidimensional cases, and the J-sample stochastic ordering case. Chapter 4 study approaches for stochastic ordering constraints with respect to time. Some concluding remarks are in Chapter 5. The overview of software tools and algorithms is in the Appendix Section.

Main contributions of the thesis

The main contributions of the thesis can be summarized as follows.

- We proposed related permutation solutions for $2 \times K$ unidimensional and multidimensional cases (Chapter 3).
- We produced the permutation solutions based on UI-NPC for the J-sample stochastic ordering case (Chapter 3).
- We provided an analysis of a unidimensional example of subarachnoid hemorrhage measured by the Glasgow outcome scale (Chapter 3).
- For the repeated measure designs, we proposed the permutation solutions to detect the stochastic ordering with respect to time (Chapter 4).

- We provided an analysis of the pain scores on the shoulder tip under two treatment schemes after laparoscopic cholecystectomy (Chapter 4).

Chapter 1

Introduction

1.1 Foreword

Throughout past decades, statisticians have continuously proposed and improved various approaches to analyzing contingency tables with categorical ordinal variables in order to make restricted inferences. There is plenty of occurrences of ordered categorical data in many fields of research and statistical consulting, such as sociology, psychology, medical studies, quality control, and so forth. Various investigations or experiments are designed to analyze superior or inferior rank of comparable groups, such as the ordering of treatment effects in clinical experiments; whether the treatment effect varies with taking different doses of the drug; respondents' inclinations toward some particular questions in social sciences and psychology; quality examination of products in marketing and technology; etc.

Such ordered categorical data are usually organized in contingency tables, and denote the rows by X be explanatory variable and the columns by Y be response variable, with J and K classes, respectively. Assume X take values at $\{\tau_1, \tau_2, \cdots, \tau_J\}$, and Y take values at $\{\upsilon_1, \upsilon_2, \cdots, \upsilon_K\}$. If rows or columns is/are ordered categorical, there are simple order $\prec$ in rows and/or columns (i.e. $\tau_1 \prec \tau_2 \prec \cdots \prec \tau_J$ or $\upsilon_1 \prec \upsilon_2 \prec \cdots \prec \upsilon_K$), and the table becomes doubly-ordered table if both are ordered categorical variables. The problem of comparing whether different levels in the row variable X satisfy stochastic dominance is our principal interest, therefore, we intend to provide tests of hypothesis with ordinal responses especially by testing for stochastic dominance between several levels of variable X.

It is well known that there are several difficulties to analyze such studies concerning some aspects: sample size is small or moderate; an occurrence of tied data; high dimensional contingency tables; the larger the sample size, the greater the number of

underlying nuisance parameters; and so forth. Such problems are still challenging and further research is going on.

1.2 Solutions within the likelihood framework

1.2.1 Prologue

A substantial of inferences, which study the restricted issues (especially stochastic ordering problems) defined on the identical ordered categorical scale, is made by the introduction of the likelihood framework schemes.

Under the likelihood framework, there are two main approaches to the problem of testing distributional equality in the null hypothesis against an ordered alternative. One is to consider the likelihood ratio principle, that is, to search for the maximum of the likelihood function under both the null parameter space and full parameter space, usually including the assumption of normality of the populations. The majority of the statistics by the introduction of the likelihood-based framework are established in a form similar to the chi-bar-squared $\bar{\chi}^2$ statistic or χ^2-mix type statistics, and few parts of approaches are to consider the F distribution and Student's t distribution.

It is well known that quantitative data have mathematical meaning and can be worked out with mathematical operations. Different from quantitative data, qualitative data or categorical data are usually descriptive. Categorical data could be represented by numerical data (i.e. "1" for *No pain*, "5" for *Severe pain*), but this does not have any clear mathematical meaning since numbers play the role of ordinal code symbols. Ordinal categorical data where the variables have order categories and the distance between categories is unknown if ordinal categories are represented by numbers (scores) having a ranking, analysis of ordinal data should incorporate natural ordering and numbers, to avoid loss of information when we only consider numerical data. Moreover, by representing ordered categorical variables by artificial variables, the restricted issues could be transformed into standard regression problems, and the isotonic regression tool is the well-known regression-based manner to make inference on such ordered restricted issues. It is an intuitive and natural idea for ordered categorical data analysis, however, that we need to maintain the sorting property of categories when we encounter issues in practice.

1.2.2 Regression-based approaches

Regression methods for making inferences of the fitting of the curves or mathematical functions to a series of data points or joint distributions have always been the focus of attention. It is also straightforward to give the definition of a regression function, which is chosen to fit a series of observed data points. Of course, there may be some problems to be considered, such as how to evaluate the quality of fit; and how to avoid the uncertainty as much as possible; and so on. The solutions to such questions involve selecting the appropriate measure of the quality of fit, the most common methods are based on restricted least squares or restricted least absolute deviations.

In different situations, in fact, regression methods could be chosen with different functions forms with respect to the shapes of the observed curves or the prior information of the researchers who may believe that the "true" distribution has a particular form with the appropriate functions. Unlike the liberal performance of unrestricted regression function, most approaches are working within several assumptions or subject to some particular constraints, and in fact, the ordered constraints have always been taken into account for the restricted classes of regression functions. Particularly, by the mathematical way, the ordered restricted regression functions can be summarized by the monotonic regression functions. A monotonic regression function is called *isotonic* regression when the function shows a nondecreasing trend, it indicates that the curve plotted has a nondecreasing tendency as the values of the independent variable increase. If the function has a non-increasing tendency, the regression function can be named as the *antitonic* regression. Here we will give the well-known definition of the *isotonic* regression (see more details in the book of Robertson *et al.* (1988)).

Definition 1.1. Let X be the finite set $\{\tau_1, \tau_2, \cdots, \tau_J\}$ with the simple order $\tau_1 \prec \tau_2 \prec \cdots \prec \tau_J$. A function f measuring on X is *isotonic* regarding the ordering if $f(\tau_1) \leq f(\tau_2) \leq \cdots \leq f(\tau_J)$, with $f(\tau_1) < f(\tau_J)$. Moreover, the simple order relation on X can be replaced with a more weak relation on X, such as *quasi-order* relation.

Simple order should satisfy four conditions: I) reflexive: $\tau_i \prec \tau_i$, for $\tau_i \in X$; II) transitive: $\tau_i \prec \tau_j$, $\tau_j \prec \tau_k$, imply $\tau_i \prec \tau_k$; III) antisymmetric: $\tau_i \prec \tau_j$, $\tau_j \prec \tau_i$, imply $\tau_i = \tau_j$; IV) comparable: if $\tau_i, \tau_j \in X$, implies either $\tau_i \prec \tau_j$ or $\tau_j \prec \tau_i$. *Quasi-order* need to satisfy reflexive and transitive but not necessarily satisfy antisymmetric.

It is worth noting that, the main objective of the statistical inference for order restricted regression theory is to find a solution to assess the quality of fit under constraints, thus, the common least squares method must be extended to the *restricted least squares*.

Definition 1.2. Let X be the finite set $\{\tau_1, \tau_2, \cdots, \tau_J\}$ with the simple order $\tau_1 \prec \tau_2 \prec \cdots \prec \tau_J$, w be a positive weight function defined on X and $\mathcal{F}$ is a restricted family of functions f measuring on X. Suppose g be a given function defined on X, a function g^* defined on X is an *isotonic regression* of g with weights $w(\tau)$ if and only if g^* is isotonic and g^* is the solution to the restricted least squares problem

$$\text{minimize} \quad \sum_{\tau \in X} [g(\tau) - f(\tau)]^2 \, w(\tau)$$

in the class of all isotonic functions f in $\mathcal{F}$. In the words, g^* is the least squares projection of g onto the collection $\mathcal{F}$. The emphasis of the solution is to find the function g^*.

The problems of comparing several populations or several mean values are of interest in practice. The majority of methods for dealing with such issues are considered under the assumption of normality. As a typical problem in the setting, it is desirable to test the null hypothesis of equality of means against the alternative where the means are isotonic under the normality assumptions. The following example was presented in the book by Barlow *et al.* (1972) and also in the book by Robertson *et al.* (1988).

Suppose $\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_J$ is a sequence of independent random samples from normal populations with unknown means $\{\mu_1, \mu_2, \cdots, \mu_J\}$ and known variances $\{\sigma_1^2, \sigma_2^2, \cdots, \sigma_J^2\}$. Let μ_j and σ_j^2 be means and variances functions defined on finite set X, respectively. Viz., $X = \{\tau_1, \tau_2, \cdots, \tau_J\}$ with the simple order $\tau_1 \prec \tau_2 \prec \cdots \prec \tau_J$, such that $\mu_j = \mu(\tau_j)$ and $\sigma_j^2 = \sigma^2(\tau_j)$. For the j-th samples, let n_j be the sample size corresponding to the j-th sample $\mathbf{Y}_j$, namely, $\mathbf{Y}_j = \{Y_{j1}, Y_{j2}, \cdots, Y_{jn_j}\}$, and sample mean be $\bar{Y}_j = \sum_{i=1}^{n_j} Y_{ji}/n_j$ for $j = 1, 2, \cdots, J$, and denote total sample size by $n = \sum_j n_j$.

We intend to find the maximum likelihood estimator (MLE) under the constraint $\mu(\tau_1) \leq \mu(\tau_2) \leq \cdots \leq \mu(\tau_J)$. We start from the likelihood function as follows:

$$L(\mathbf{Y}_1, \cdots, \mathbf{Y}_J | \boldsymbol{\mu}, \boldsymbol{\sigma}) = (2\pi)^{-n/2} \prod_{j=1}^{J} \sigma_j^{-n_j} \exp\left\{ -\frac{1}{2\sigma_j^2} \sum_{i=1}^{n_j} (Y_{ji} - \mu(\tau_j))^2 \right\},$$

The log-likelihood function is proportional to

$$l(\boldsymbol{\mu}, \sigma^2 | \mathbf{Y}) \propto -\sum_{j=1}^{J} \sigma_j^{-2} \sum_{i=1}^{n_j} (Y_{ji} - \mu(\tau_j))^2.$$

It is straightforward to minimize the $\sum_{j=1}^{J}\sigma_j^{-2}\sum_{i=1}^{n_j}(Y_{ji}-\mu(\tau_j))^2$ to obtain the restricted MLE subject to $\mu(\tau_1)\leq\mu(\tau_2)\leq\cdots\leq\mu(\tau_J)$. The sum can be rewritten as

$$\sum_{j=1}^{J}\sum_{i=1}^{n_j}\sigma_j^{-2}(Y_{ji}-\bar{Y}_j)^2+2\sum_{j=1}^{J}\sum_{i=1}^{n_j}\sigma_j^{-2}(Y_{ji}-\bar{Y}_j)[\bar{Y}_j-\mu(\tau_j)]+\sum_{j=1}^{J}[\bar{Y}_j-\mu(\tau_j)]^2n_j\sigma_j^{-2}.$$

The first term does depend on $\mu(\tau_j)$, the second term disappears, thus, the solution become to minimize $\sum_{j=1}^{J}[\bar{Y}_j-\mu(\tau_j)]^2n_j/\sigma_j^2$ for restricted MLE subject to $\mu(\tau_1)\leq\mu(\tau_2)\leq\cdots\leq\mu(\tau_J)$. Comparing with the definition above, let g and w be the functions defined on X, namely, $g(\tau_j)=\bar{Y}_j$ and weights $w(\tau_j)=n_j/\sigma_j^2$. Denote the restricted MLE of $\boldsymbol{\mu}$ by g^*, hence g^* is isotonic regression of g with weights w.

The problem of the case is to test the equality of means in the null hypothesis against the alternative that the means are isotonic.

$$H_0:\mu(\tau_1)=\mu(\tau_2)=\cdots=\mu(\tau_J)$$

against

$$H_1:\mu(\tau_1)\leq\mu(\tau_2)\leq\cdots\leq\mu(\tau_J)\quad\text{with}\quad\mu(\tau_1)<\mu(\tau_J)$$

It is clear that the MLEs of $\boldsymbol{\mu}=\{\mu(\tau_1),\mu(\tau_2),\cdots,\mu(\tau_J)\}$ under null hypothesis H_0 are given by

$$\widehat{\mu}(\tau_1)=\widehat{\mu}(\tau_2)=\cdots=\widehat{\mu}(\tau_J)=\sum_{j=1}^{J}w(\tau_j)\bar{Y}_j/\sum_{j=1}^{J}w(\tau_j)\quad\text{with}\quad w(\tau_j)=n_j/\sigma^2(\tau_j).$$

The likelihood ratio test can be taken into account to settle such problems. Furthermore, the finite set X could assume artificial variables by the introduction of "scores" that maintain the same order restrictions, it may be relabelled by numerical values based on actual problems.

1.2.3 Chi-squared-based statistics

The problem of comparing several populations or groups frequently occurs in many areas. Normality assumptions are often introduced to settle such issues, and it is expected to test for homogeneity of normal means against order restricted alternative. More details about the following example were presented in the book by Barlow *et al.* (1972) and also in the book by Robertson *et al.* (1988).

Suppose X be finite set $\{\tau_1, \tau_2, \cdots, \tau_J\}$ with simple order/quasi-order $\prec$ defined on X. To simplify, the components in X can be remarked by numerical values that maintain the same order relations, namely, $X = \{1, 2, \cdots, J\}$, and denote positive weights by w_j. At the same time, $\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_J$ is a sequence of independent random samples from normal populations with unknown means $\{\mu_1, \mu_2, \cdots, \mu_J\}$ and known variances $\{\sigma_1^2, \sigma_2^2, \cdots, \sigma_J^2\}$, denote sample size and sample mean by n_j and $\bar{Y}_j$ for $j = 1, 2, \cdots, J$, the simplified notations share the same meaning as presented in the previous subsection.

The interest of problem may test for homogeneity of normal means with order restricted alternative,

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_J \quad vs. \quad H_1 : \mu_1 \leq \mu_2 \leq \cdots \leq \mu_J, \text{ with } \quad \mu_1 < \mu_J.$$

Under H_0, the MLE $\widehat{\mu}$ of $\mu_1 = \mu_2 = \cdots = \mu_J$ are given by the previous subsection,

$$\widehat{\mu} = \sum_{j=1}^{J} w_j \bar{Y}_j / \sum_{j=1}^{J} w_j \quad \text{with} \quad w_j = n_j/\sigma_j^2$$

Assume the restricted MLE of $\boldsymbol{\mu}$ under H_1 be $\boldsymbol{\mu}^*$, it is clear that $\boldsymbol{\mu}^*$ is isotonic regression of $\bar{\mathbf{Y}}$ with weights $\mathbf{w} = \{w_1, w_2, \cdots, w_J\}$.

The likelihood ratio test (LRT) for testing H_0 against H_1 rejects H_0 for small values of the statistic

$$\lambda = \frac{\max_{\boldsymbol{\mu} \in H_0} L(\mathbf{Y}_1, \cdots, \mathbf{Y}_J | \boldsymbol{\mu}, \boldsymbol{\sigma})}{\max_{\boldsymbol{\mu} \in H_1} L(\mathbf{Y}_1, \cdots, \mathbf{Y}_J | \boldsymbol{\mu}, \boldsymbol{\sigma})}$$

The negative of the log-likelihood function is proportional to

$$-2 \log \lambda \propto \sum_{j=1}^{J} w_j (\mu_j^* - \widehat{\mu})^2$$

Therefore, the LRT rejects to the null hypothesis for large values of

$$\bar{\chi}_J^2 = \sum_{j=1}^{J} w_j (\mu_j^* - \widehat{\mu})^2, \quad \text{with} \quad w_j = n_j/\sigma_j^2$$

Here the subscript J implies the number of means being compared.

Furthermore, if the variances are unknown with form $\sigma_j^2 = a_j\sigma^2$, where a_j are known but σ^2 is unknown, the LRT of the statistic is given by

$$\bar{E}_J^2 = \frac{\sum_{j=1}^{J} w_j(\mu_j^* - \widehat{\mu})^2}{\sum_{j=1}^{J} a_j^{-1} \sum_{i=1}^{n_j}(Y_{ji} - \widehat{\mu})^2}$$

where the statistic $\bar{E}_J^2$ rejects H_0 for large values.

Although the statistic $\bar{\chi}_J^2$ is of form as $\sum_{j=1}^{J} w_j(\mu_j^* - \widehat{\mu})^2$, the null distribution of the $\bar{\chi}_J^2$ is not easy to determine especially when the value of J is large. Accordingly, the following definition is a powerful tool to give a solution to obtain the null hypothesis distribution (Robertson *et al.* (1988)).

Definition 1.3. Let $\boldsymbol{\mu} \in H_0$. The level probabilities are defined as

$$\Pr(l, J; \mathbf{w}) = \Pr(M = l), \quad l = 1, 2, \cdots, J,$$

where M is the number of level sets in $\mathbf{Y}^*$, which is in accordance with the isotonic regression of $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_J)$ with weights $\mathbf{w}$, and $Y_1, Y_2, \cdots, Y_J$ are independent normal variables with mean μ_j and variance w_j^{-1}. Obviously, the $\Pr(l, J; \mathbf{w})$ do not depend on the common value of μ_j and they are unchanged if $\mathbf{w}$ is multiplied by a positive constant. Particularly, when $w_1 = w_2 = \cdots = w_J$, level probabilities have the form of $\Pr(l, J)$.

The general form of the null distribution of the statistic $\bar{\chi}_J^2$ which need the values of level probabilities $\Pr(l, J; \mathbf{w})$ is given by

$$\Pr\left(\bar{\chi}_J^2 \geq c\right) = \sum_{l=1}^{J} \Pr(l, J; \mathbf{w}) \Pr\left(\chi_{l-1}^2 \geq c\right), \quad c > 0.$$

Obviously, the density of $\bar{\chi}_J^2$ can be expressed as a weighted sum of well-known densities. Therefore, it is necessary to find the level probabilities $\Pr(l, J; \mathbf{w})$ for getting the null hypothesis distribution. Unfortunately, simple closed-form expressions for $\Pr(l, J; \mathbf{w})$ are not possible to be determined except some particular orderings (Barlow *et al.* (1972) and Robertson *et al.* (1988)).

The general method can be implemented in two steps, and the main idea is by using recursive algorithms. Firstly, we need to compute the value of $\Pr(J, J; \mathbf{w})$. And then for $l < J$, the recursive formulas for $\Pr(J, J; \mathbf{w})$ in terms of $\Pr(j, i; \mathbf{w})$ for $j \leq i < J$, the values $\Pr(l, J; \mathbf{w})$ can be computed under the constraint $\sum_{l=1}^{J} \Pr(l, J; \mathbf{w}) = 1$.

The example below is a special case of simple order, where $H_1 : \mu_1 \leq \mu_2 \leq \cdots \leq \mu_J$, with $\mu_1 < \mu_J$. Suppose $Y_1, Y_2, \cdots, Y_J$ are independent normally distributed variables, with a common zero mean and variances $w_1^{-1}, w_2^{-1}, \cdots, w_J^{-1}$ under H_0. The method firstly considers the probability $\Pr(Y_1 < Y_2 < \cdots < Y_J)$; by setting $U_{j-1} = Y_j - Y_{j-1}$ for $j = 2, 3, \cdots, J$, the orthant probabilities P_{j-1} are

$$P_{j-1} = \Pr(U_1 > 0, U_2 > 0, \cdots, U_{j-1} > 0), \text{ for } j = 2, 3, \cdots, J.$$

Clearly, $\mathbf{U} = (U_1, U_2, \cdots, U_{J-1})$ is multivariate normally distributed, viz. $\mathbf{U} \sim \mathcal{N}(\mathbf{0}, \Sigma)$. For $J > 5$, however, there does not exist a general closed expression for orthant probabilities P_{j-1}(see Abrahamson (1964)); for the case $J \leq 5$, the related solution of which can be obtained by the correlation coefficients ρ_{jk} and integrals. Let $B_1, B_2, \cdots, B_l$, for $1 \leq l \leq J$, be the (unordered) level sets of $\mathbf{Y}^*$, and $\mathcal{L}_{\{B_1, B_2, \cdots, B_l\}}$ be the collection of all decompositions of $\mathbf{Y}^*$ into l (unordered) level sets, and $\#(\cdot)$ is the number of elements of $\mathbf{w}^*_{B_j}$.

The level probabilities $\Pr(l, J; \mathbf{w})$ for the case $J = 2$, are straightforward to achieve, $\Pr(1, 2; \mathbf{w}) = \Pr(2, 2; \mathbf{w}) = P_1 = 1/2$.

For the case $J = 3$, $\Pr(3, 3; \mathbf{w}) = P_2 = 1/4 + 1/2\pi \cdot \sin^{-1} \rho_{12}$, and based on recursive algorithm,

$$\begin{aligned}\Pr(2, 3; \mathbf{w}) = &\Pr(2, 2; w_1 + w_2, w_3) \cdot \Pr(1, 2; w_1, w_2) \\ &+ \Pr(2, 2; w_1, w_2 + w_3) \cdot \Pr(1, 2; w_2, w_3)\end{aligned}$$

since $\Pr(1, 3; \mathbf{w}) = 1 - \Pr(3, 3; \mathbf{w}) - \Pr(2, 3; \mathbf{w})$.

Similarly, $\Pr(4, 4; \mathbf{w})$ takes the value of P_3. For $\Pr(3, 4; \mathbf{w})$, the number of level sets is 3 for simple order, therefore, there are three decomposition methods to weights $\mathbf{w}$; namely,

$$\Pr(3, 4; \mathbf{w}) = \sum_{(B_1, B_2, B_3) \in \mathcal{L}_{\{B_1, B_2, B_3\}}} \left[\Pr(3, 3; \mathbf{w}^*_{B_1}, \mathbf{w}^*_{B_2}, \mathbf{w}^*_{B_3}) \cdot \prod_{i=1}^{3} \Pr(1, \#(\mathbf{w}^*_{B_i}); \mathbf{w}^*_{B_i}) \right]$$

where the decompositions of $\mathbf{w}$ include $\{w_1, w_2\}$, $\{w_3\}$, $\{w_4\}$; $\{w_1\}$, $\{w_2, w_3\}$, $\{w_4\}$; $\{w_1\}$, $\{w_2\}$, $\{w_3, w_4\}$. Applying the decomposition methods to $\mathbf{w}$ of $\Pr(2, 4; \mathbf{w})$, the results obtained, consist of the following: $\{w_1\}$, $\{w_2, w_3, w_4\}$; $\{w_1, w_2, w_3\}$, $\{w_4\}$; $\{w_1, w_2\}$,

$\{w_3, w_4\}$; thus $\Pr(2, 4; \mathbf{w})$ is expressed as follows,

$$\begin{aligned}\Pr(2, 4; \mathbf{w}) = &\Pr(2, 2; w_1, w_2 + w_3 + w_4) \cdot \Pr(1, 3; w_2, w_3, w_4) \\ &+ \Pr(2, 2; w_1 + w_2 + w_3, w_4) \cdot \Pr(1, 3; w_1, w_2, w_3) \\ &+ \Pr(2, 2; w_1 + w_2, w_3 + w_4) \cdot \Pr(1, 2; w_1, w_2) \cdot \Pr(1, 2; w_3, w_4).\end{aligned}$$

It implies that $\Pr(2, 4; \mathbf{w}) = 1 - \Pr(2, 4; \mathbf{w}) - \Pr(3, 4; \mathbf{w}) - \Pr(4, 4; \mathbf{w})$. For more details, refer to the book (Robertson *et al.* (1988)).

Given that the condition is in some particular orderings, have the same weights or have some other stringent restrictions, the distributions of $\bar{\chi}^2_J$ can be determined by the level probabilities $\Pr(l, J; \mathbf{w})$ when the number of groups is not larger than 5, e.g. the above-mentioned example. There do not exist simple closed-form expressions of level probabilities for more general cases ($J > 5$), hence, the density of statistic $\bar{\chi}^2_J$ may differ depending on the choice of the weights $\Pr(l, J; \mathbf{w})$.

Obviously, in order to obtain the maximum likelihood estimator (MLE) through minimizing the negative of the log-likelihood function, both the regression-based approaches and chi-squared-based methods consider least squares to assess the quality of fit. Hence the restricted MLE is isotonic regression of the given function with weights. The method adopted is the likelihood ratio test (LRT), which has always been widely used in the parametric area and has many nice properties of making statistical inferences, the statistic named chi-bar-squared statistic $\bar{\chi}^2_J$ and its density is of form as a weighted sum of central chi-squared densities. Even though we have the form of the density of $\bar{\chi}^2_J$, the choice of the weights may either take expensive computations or cannot be possible to compute in weak restrictions, especially when the number of comparison groups is large.

Comparing the gain in the power of the chi-bar-squared statistic with that of the standard chi-squared statistic, it is an obvious advantage that the chi-bar-squared statistic is better than the chi-squared statistic when the alternative is of stochastic ordering (Barlow *et al.* (1972)). It is well known that, the classical Pearson's chi-squared test presents quite a poor performance when there exist the ordering columns in the contingency table (e.g., Graubard and Korn (1987); Nair (1987)).

Almost at the same time, Kudo (1963) also made a major contribution for the related similar problem, which tests equality constraint among means of a given multivariate normal population with known covariance matrix against one-sided ordered alternative hypothesis, namely, $H_0 : \mu_1 = \mu_2 = \cdots = \mu_J$ against $H_1 : \mu_1 < \mu_2 < \cdots < \mu_J$. The hypotheses could be equivalently represented by the differences between adjacent means, that is, the null hypothesis $H_0 : \mu_{l+1} - \mu_l = 0$ against the alternative hypothesis

$H_1 : \mu_{l+1} - \mu_l > 0$ for $l = 1, \cdots, J-1$. The author proposed a chi-bar-squared statistic $\bar{\chi}^2$ based on the likelihood ratio criterion. The key point of statistic $\bar{\chi}^2$ is to find the restricted MLE. However, the task is not so simple to achieve under the constraints. Colombi and Forcina (2016) introduced two schemes of testing procedures for detecting several relevant violations of the order relations. One of the schemes considers the decomposition of the log-likelihood ratio, and the other scheme is based on the assumption of asymptotically multivariate normal distributions. Zhu and Chen (2018) proposed a manner to decompose the dimensions of a given multivariate normal distributed sample into two disjoint subsets, and gave the likelihood ratio test statistic (mLR) based on these two complementary subsets, and the distribution of mLR can be described by a finite mixture of F distributions. When the dimension of the sample is less than 4, the authors gave the specific expressions of the LRT statistic. Obviously, the larger the problem dimension, the more difficult it is to express the mLRT statistic. Comparing the behavior of powers with the classical likelihood ratio test and the Perlman-Wu test (Perlman and Wu (2003)), the proposed test behaved generally better, but not uniformly better.

Further, by introducing the concept of a cumulative effective score, the statistic $\bar{\chi}^2$ can be extended to cumulative chi-squared tests for testing ordered alternatives (Hirotsu (1982); Hirotsu (1986)). The author proposed a test of the null hypothesis of homogeneity of means against one-sided ordered alternative for a multivariate normal problem for

$$H_0 : \mu_1 = \cdots = \mu_J \quad vs. \quad H_1 : \mu_1 \geq \cdots \geq \mu_J$$

and a test of the null hypothesis of linearity

$$H_0 : \lambda_1 - \lambda_2 = \lambda_2 - \lambda_3 = \cdots = \lambda_{J+K-2} - \lambda_{J+K-1}$$

against convexity alternative for adjacent cohort effect in the given cohort model

$$H_1 : \lambda_1 - \lambda_2 \geq \lambda_2 - \lambda_3 \geq \cdots \geq \lambda_{J+K-2} - \lambda_{J+K-1}$$

Given the introduction of cumulative effective scores, these tests are characterized as a weighted sum of independent chi-squared random variables each with one degree of freedom. The algorithm for deriving the weights is provided by both the Čebyšev's orthogonal polynomials for finite points and recurrence algorithms. The approximation of the cumulative chi-squared can be achieved by two-moment approximation. Nair

(1987) had also been done an in-depth study into the general case of cumulative chi-squared statistics for the ordered alternatives. As populations behave as multinomial or binomial, the testing problem of interest is to test homogeneity of parameters or of cumulative probabilities, against stochastic dominance or ordering in the alternatives. These two testing problems all yield test statistics which have an asymptotic $\bar{\chi}^2$ distribution. The limiting distribution of the proposed statistic has the form of a weighted sum of central chi-squared random variables with one degree of freedom under the null hypothesis distribution, namely, $T \xrightarrow{D} \sum_i c_i \chi^2_{1(i)}$, and is distributed as a weighted sum of non-central chi-squared random variables with one degree of freedom with non-central parameters δ_i under the ordered alternative; that is, $T \xrightarrow{D} \sum_i c_i \chi^2_{1(i)}(\delta_i)$.

When the samples involve the count random variables or discrete populations, for instance, we have J populations having independent distribution functions $F_1, \cdots, F_J$, the related likelihood functions have the form of the product of probability mass functions for finite points. A typical way for such a case under order constraints is based on the likelihood ratio criterion. However, the greater the number of compared distributions, the more difficult the computations. Dykstra (1982) considered the case of survival functions with right-censored data defined on finite time points $0 = T_0 < T_1 < \cdots < T_n < T_{n+1} = \infty$, and assume the related survival functions satisfy stochastic ordering relations

$$P_1 \overset{st}{\geq} P_2 \overset{st}{\geq} \cdots \overset{st}{\geq} P_J$$

The author gave restricted MLEs which are expressed in the form of Kaplan-Meier product limit estimators, provided by using iterative algorithms based on the pairwise scheme. The solution has good performance for statistical inference, however, the actual MLEs are difficult to realize. The restricted MLEs provided by the solution converge to the actual MLEs in probability. Such an iterative procedure for solutions that depend only on pairwise problems was also studied by Feltz and Dykstra (1985). For a similar problem under ordered restrictions, Wang (1996) indicated that there do not exist closed-form expressions for maximum likelihood estimators when the number of compared distributions is larger than 2. By the introduction of Monte Carlo simulation, Wang (1996) derived the limiting distribution of the log-likelihood ratio test under the null hypothesis; obviously, the limiting distribution can be expressed in terms of a chi-bar-squared statistic. Considering a typical problem, given two m-dimensional multivariate populations, in order to know if one stochastically dominates the other, Sampson and Whitaker (1989) proposed an algorithm that reduces the dimensionality of the problem to obtain a numerical approximation of MLEs. The algorithm transforms the two-sample

problem into two one-sample problems in which the MLEs of the one-sample problem can be achieved by a min-max formulation based on isotonic regression techniques, it is obvious that the LRT statistic has an asymptotic $\bar{\chi}^2$ distribution.

Davidov and Peddada (2011) proposed a manner to settle the problem of comparison stochastic ordering among groups for multivariate binary response data regardless of any dependence structure. The testing problem can be broken down into several sub-problems of pairwise comparison by using procedures of union-intersection tests and Bonferroni-based tests, and made statistical inferences of the combination of individual univariate tests.

1.2.4 Association-model-based approach

As given samples can be summarized in terms of two-way contingency tables, the simplest case is to consider a 2×2 table. Various statistical methods for examining homogeneity/association in 2×2 tables have been studied intensively over the years. One of the famous methods is Fisher's exact test which provides a powerful behavior for a testing association. An alternative exact test is Barnard's exact test. Both exact tests are based on the conditionality principle where the margins are held fixed. Other alternatives are to use the Pearson chi-squared statistic or chi-bar-squared statistic, both of them behave well for essentially testing independence/association between two categorical variables.

Further, given that the $J \times K$ two-way contingency table, the testing problem for homogeneity versus ordered restrictions becomes much more difficult and complicated, especially when columns or rows or both are defined on the ordering categorical scale. Methods of statistical inferences for such problems have been proposed and developed by many researchers under specific assumptions, such as variables from the regular exponential family, a nonnegative association of cell probabilities, and so forth. When available, likelihood-based solutions within stringent assumptions are provided with nice characterization statistical inferences. In general, however, it is quite difficult to obtain proper testing inference in practice, especially for the multivariate case. The multivariate case is much more difficult to be analyzed within likelihood frameworks than the univariate one. In such a setting, the number of underlying nuisance parameters to be removed and/or that of observed variables can often be much larger than sample sizes. So, unless clearly justified assumptions allowing for a considerable reduction of underlying complexity, the most intriguing of which is when one pseudo-parameter is expressed as a function of many underlying nuisance parameters, no general correct testing solution is possible within that approach. In what follows, the columns or rows refer

to ordered categorical variables, or both columns and rows refer to ordered categorical variables. In such a setting, the columns or rows can be assumed as multinomial or binomial distributions, and denote explanatory variable by rows and response variable by columns.

Method called *linear-by-linear association* model is a general scheme for studying whether an association between column variables and row variables satisfies a set of ordinal structures. Assigned scores $\alpha_1 < \cdots < \alpha_j < \cdots < \alpha_J$ for the rows and $\beta_1 < \cdots < \beta_k < \cdots < \beta_K$ for the columns, the *linear-by-linear association* model is in the form of

$$\log \Pr(Y = k \mid X = j) = \mu + \lambda_j^X + \lambda_k^Y + \psi \alpha_j \beta_k$$

where the parameter ψ indicates the direction of the association.

Agresti *et al.* (1990) proposed an algorithm of a statistic to test for independence against a one-sided alternative for small-sample inference in the doubly-ordered table. The significant value of the statistic for the testing problem is provided by both the linear-by-linear association model and the exact permutation distribution of ordinal odds ratios. Moreover, Agresti and Coull (2002) reviewed various approaches for the categorical responses with inequality constraints in the contingency tables. The emphasis of the proposed approach is to presume the inequality constraints on the parameters of these categorical responses, and the linear-by-linear association model should satisfy the assumption. For the problem of binomial responses with several ordered levels, they introduced a generalized linear model under the condition of monotonicity constraint on the underlying parameters. When the contingency table consists of more than two rows and columns respectively, the problems become the multinomial case with several ordered levels. In this case, the authors prefer to collapse rows and columns into the double dichotomy of the two variables to obtain a set of 2×2 tables, and proposed the association structure, that is described with several types of odds ratios and the log-linear models with monotonicity ordered parameters. The algorithm is processed in the cell probabilities. Kateri (2011) intended to find which one is better between two samples measured on the same ordinal scale, through the column effect association model with constrained column scores. That author also gave a unified scheme for comparing two treatment groups, when the testing problem of interest is to detect the null hypothesis of homogeneity against the ordered restricted alternative, and the ordered restrictions consist of stochastic ordering and umbrella ordering. By the introduction of the Bayesian approach, Kateri and Agresti (2013) assumed the conditional probabilities in each row as conjugate Dirichlet priors, and gave the stochastic ordering and other

types of ordinal odds ratios structure for the two-samples comparison.

1.2.5 Distance-based approach

In particular, some statisticians turned their attention to assigning specific norms to the given data, such as L^2-norm. Under the assumption of the norm, many statistical inferences related to distance can be derived immediately, such as location, dispersion or symmetry of random variables; confidence intervals; and so forth. All methods that decompose a $J \times K$ table into a number of 2×2 tables and consider the related partial tests on odd-ratios fall into the subsequent problem of combining that number of dependent tests. That combination process is traditionally carried out by using a Bonferroni's type of analysis, which in turn may become too conservative.

For testing whether the two mean vectors from two independent high-dimensional samples are equal, Zhang *et al.* (2019) proposed an L^2-norm-based test statistic $T_{n,p}$ with fixed p and any n,

$$T_{n,p} = (n_1 n_2/n)||\bar{y}_1 - \bar{y}_2||^2$$

which has a similar form to the Hotelling's T^2-test statistic $T_H = (n_1 n_2/n)(\bar{y}_1 - \bar{y}_2)^T \Sigma^{-1}(\bar{y}_1 - \bar{y}_2)$, where Σ is the estimated covariance matrix; and a non-exact test statistic $T_{BS} = (n_1 n_2/n)||\bar{y}_1 - \bar{y}_2||^2 - \text{tr}(\Sigma)$ proposed from Bai and Saranadasa (1996). The asymptotic distribution of the proposed test statistic can be achieved by using the Welch-Satterthwaite (W-S) χ^2-approximation under the null hypothesis. The proposed approach for the testing problem, is to avoid assuming any strong particular conditions of the unknown covariance matrix and the underlying distribution of the samples, it also works when the dimension of units is larger than the total sample size. The problem of testing is a goodness of fit problem, the key point of the literature is that the two samples are from unknown distributions but with the same covariance matrix. The authors derived the W-S χ^2-approximation approach from the fixed and low dimensional normal data to high-dimensional normal or non-normal data without any assumptions of the covariance matrix.

In the spirit of distance-based approach, Weiß (2019) considered a distance function defined on the ordinal random variables for the problem with ordinal categorical variables and ordinal times series, and expressed ordinal random variable with the rank-count variable. The author intended to consider the distance-based approach based on expected distances, which is structured by the framework with two necessary aspects: one is the definition of distance measures or metric measures and another one is possible

properties of ordinal distances, and then to obtain the measures of location, dispersion or symmetry of random variables and the measures of serial dependence within a given process. Three types of distance measures are introduced in the literature, such as Hamming distance $d_H(s_k, s_l) := 1 - \delta_{k,l}$, with $\delta_{k,l}$ denotes Kronecker delta; Minkowski distance $d_{o,1}(s_k, s_l) :=\mid k - l \mid$; and squared Euclidean distance $d_{o,2}(s_k, s_l) := (k - l)^2$; where both s_k and s_l belong to the ordered categorical range $\mathcal{S}$, and with $s_0 < \cdots < s_m$. Thus, a distance-based approach gives us the possibility to obtain the asymptotic distribution with the given measures. However, since it is not an easy way to achieve this aim, Weiß indicated "*Unfortunately, it is not possible to find simple closed-form formulae for this asymptotics being valid for any type of distance measure d. But focusing on a particular type of distance instead, such derivations are possible*". Furthermore, the author extended the case with I.I.D. ordinal random variables to the case with dependence ordinal random variables and the ordinal time series.

1.3 Solutions by nonparametric scheme

In the realistic problems, researchers may not desire to assume specific distributions to the given data in advance, but making assumptions based on prior information is inevitable. Indeed, researchers tend to assign specific parametric families to observations, and make statistical inferences within principles of parametric approaches. It is natural to obtain a nice and intuitive statistical characterization of inferences within parametric/semi-parametric assumptions. Perhaps the supposed parametric family is far from the potential population of a given data; however, it is intuitively appealing and useful for analyzing the problems.

In the clinical research, the biomedical studies are designed to answer some specific questions, including the feedback on several treatments (such as new drugs; dosage levels; toxin doses; and so forth), and to make a further comparison between the new treatment and other competitors in order to find the better one. This involves a comparison of the ranking levels of treatment effects, and the corresponding statistical studies lie within the order restriction inferences and hypothetical inference theorems. Methods of hypothetical inference for order restrictions take frequently account of parametric theorems, which always consider intrinsic assumptions including several aspects: normality; data homogeneity in the alternatives; random sampling from a pre-specific population; random effects independent of units; separable nuisance entities; and so forth. Under such assumptions, most parametric tests enjoy convergence theorem, and tend to some particular asymptotic distributions (such as mixed chi-squared distribution; mixed F

distribution; etc.), but generally with an unknown rate of convergence. It is worth noting that these methods are intuitively attractive for statistical inference, but they do not guarantee correctness or closeness to "true" latent populations. In some scenarios, the normality assumption is even quite questionable. What is more, expensive computation costs may be inevitable to deal with multidimensional problems under the likelihood framework scheme for making restricted inferences.

When the problems are so complex or involve multidimensional cases, algorithms of nonparametric approaches only need mild assumptions, therefore, we may suggest using nonparametric approaches instead of parametric/semi-parametric methods to avoid those disadvantages or stringent restrictions of parametric approaches. In particular, the permutation testing principle plays an important role in the nonparametric area to overcome most of the complex multidimensional problems within restricted assumptions. There exist two indispensable principles for permutation approaches, namely, the conditionality and sufficiency principles, where the conditioning is with respect to a set of sufficient statistics in the null hypothesis. Due to this, it shows a general good power behavior. The permutation method usually works in a so-called combination-based algorithm, which provides solutions for many complex situations under lack of knowledge on underlying distributions, with multidimensional variables, or with unknown dependence relations between variables. The approach is called nonparametric combination (NPC) of dependent permutation tests (Pesarin (2001), Pesarin (2015), Pesarin (2006); Basso *et al.* (2009); Pesarin and Salmaso (2006), Pesarin and Salmaso (2010); Pesarin *et al.* (2016)). The procedure of NPC testing performs Roy (1953)'s Union-Intersection (UI) approach in typical nonparametric settings. In other words, the procedure assumes that the original testing problem can be appropriately and equivalently broken-down into a set of sub-problems, that each sub-problem is provided with an appropriate permutation solution and that these sub-problems can be jointly analyzed. Especially, the idea of "dependence" is reflected in the underlying unknown dependence, which in turn is fully contained in the sufficient statistic, and due to conditioning, it remains invariant in the computational process. The subsequent combination process often adopts Fisher, Liptak or Direct combining functions. It is worth noting that, the power functions provided by NPC regarding permutation principles may not be the optimal one but are close to the best parametric counterparts when available, even when the sample sizes are moderate. The results provided by such an approach, being generally exact, do not require asymptotic approximations and so they would be more credible and closer to the latent real populations.

The testing problems for homogeneity against ordered restrictions defined on the

ordered categorical scale are our principle interest. Ordered categorical data have often appeared in many fields. Taking clinical experiments for an example, experiment is designed to analyze the treatment effects of drugs in order to provide convincing evidence that the treatment has an active effect; each subject is randomly assigned to receive either the treatment or a placebo, and choose one of several options {*No pain; Mild pain; Moderate pain; Severe pain*} to answer the study question. The options mentioned above obviously define on an ordered categorical variable. Ordered categorical data usually present many tied data in practice. Given such characteristics of ordered categorical data, the statistical inferences by NPC test become slightly more complicated, we should be more careful to select the permutation tests for sub-problems. Sometimes, ordered categorical data are usually organized with two-way contingency tables or even on longitudinal time tables.

Graubard and Korn (1987) discussed the choices of column scores for testing independence in ordered $2 \times K$ contingency tables, and listed 14 tests under different column scores preassigned methods based on the exact conditional permutation theorem. They preferred to consider equally spaced scores for column preassigned scores when the choice is not apparent in practice, and should examine the mid-rank as column scores carefully before using a rank test. They deemed, that one should be more cautious to choose rank statistics for testing independence when the distribution of column margin is not or far from uniform. It is worth noting that the selection of the schemes of the column scores should be careful when the column factors are interval measures. Lumley (1996) proposed a generalized estimating equation (GEE) model, which incorporates most association structures based on the cumulative odds ratio. Analyzing categorical ordinal data on longitudinal design, Brunner and Langer (2000) proposed an extension of the Wilcoxon-Mann-Whitney test to factorial designs, the pooling data were constructed on the nonparametric marginal model and on the relative treatment effects which were estimated by using the empirical method based on mid-ranks.

Considering the problem of testing distributional equality in the null hypothesis against stochastic dominance in the alternative between two levels of treatment for ordinal categorical variables, Arboretti *et al.* (2007) and Arboretti and Bonnini (2009) introduced a solution based on score transformation and on the finite moments of transformed variables, with $w_1 < w_2 < \cdots < w_k$ and $\mathbf{E}\{[w(X)]^r\} = \sum_k \Pr(X = A_k) w_k^r$, thus, the null hypothesis is in the form of

$$H_0 : \{X_1 \overset{d}{=} X_2\} \equiv \left\{ \bigcap_{r=1}^{K-1} \mathbf{E}\{[w(X_1)]^r\} = \mathbf{E}\{[w(X_2)]^r\} \right\}.$$

against the stochastic dominance as follows

$$H_1 : \{X_1 \overset{st}{>} X_2\} \equiv \left\{ \bigcup_{r=1}^{K-1} \mathbf{E}\{[w(X_1)]^r\} > \mathbf{E}\{[w(X_2)]^r\} \right\}.$$

which is based on the nonparametric combination of dependent permutation tests. The rationale for this is based on the notion that "joint equality of first $K-1$ moments of two discrete distributions over K classes entails equality of their probability generating functions and so of their distributions". Bazyari and Pesarin (2013) intended to test for the homogeneity of mean vectors against the alternatives of two-sided restricted in multivariate normal distributions. There are two cases under the multivariate normal distributions: one case which has known covariance matrices can be settled within classical likelihood ratio criterion, which has a known mixed chi-squared asymptotic distribution; the other, which has unknown and common covariance matrices. Firstly, the authors gave a reformulation of the test statistic and determined the upper bounds for its p-values; in the meantime, they also proposed a solution on the permutation approach. Gökpinar *et al.* (2017) preferred to utilize the so-called χ^2-P statistic, which indicates that the algorithm of separable χ^2 statistic is working within the permutation approach, for testing equality of linear, quadratic and cubic effects for categorical ordinal data with small sample size.

An nonparametric statistic called the modification of the Baumgartner-Weiß-Schindler (modified BWS) statistic was proposed by Neuhäuser (2006), for testing homogeneity of binomial proportions $H_0 : \pi_0 = \pi_1 = \cdots = \pi_J$ against one-sided ordered alternative $H_1 : \pi_0 \leq \pi_1 \leq \cdots \leq \pi_J$ with at least $\pi_0 < \pi_J$. Compared the exact test based on the modified BWS statistic with asymptotic Cochran-Armitage (asymptotic CA) test and exact Cochran-Armitage (exact CA) test, the modified BWS test shows better behavior than the exact CA test, but not uniformly better than the asymptotic CA test. Jelizarow *et al.* (2015) also provided an in-depth study on the global inference for two samples with multivariate high dimensional ordinal data on marginal distributions. The authors discussed and gave algorithms for several pairwise testing problems, such as identical joint distribution (IJD) against non-identical joint distribution (NJD); simultaneous marginal homogeneity (SMH) against marginal inhomogeneity (MI); simultaneous marginal homogeneity (SMH) against two-sided alternative as marginal order (MO). They gave the test statistics of multivariate quadratic forms, which variables are independent, for two testing problems (SMH against MI; SMH against MO), respectively. In the meantime, they discussed the permutation-based global inference about marginal distributions.

Chapter 2

Theory of Permutation Tests

2.1 Prologue

In practical investigations, the comparison of several groups for making inferences, researchers meet with a set of questions, this leads us to study in-depth and construct valid methods to either determine if such groups are well-matched or make a superior or inferior rank analysis. Especially in the field of clinical experiments, it is usually necessary to compare different levels of treatment effects. Researchers gave different levels of treatment to the subjects, and then compared the outcomes of the treated subjects to assess whether the subject's response differs as the treatment scheme changes, and so the related response distribution may depend on treatment effect. Researchers desire to provide convincing evidence to select the superior one or rating levels of treatments. In particular, such a challenging task can be converted into a statistical language description, the object is to determine whether there exists stochastic ordering among groups.

The methods we will discuss are to determine whether categorical explanatory variables satisfy stochastic ordering restriction when the response variables are expressed with ordered categories in contingency tables. It is well known that solutions to order restricted problems are mentioned by many kinds of literature. These solutions assume the underlying population response is normally distributed under some quite stringent assumptions; or the solution suggested by Nair (1987) considers the cumulative chi-squared method for multinomial or binomial data, but there is no closed-form of expressions for the related test statistics.

If the response variable is measured on a binary scale (*Yes* or *No*; *A* or *B*; *Agree* or *Disagree*; etc.), there are many kinds of literature that provide solutions for detecting stochastic ordering among explanatory variables (e.g. Agresti *et al.* (1990); Agresti

and Coull (2002); Kateri and Agresti (2013); etc.). When the responses are binary data ($K \geq 2, J = 2$), the testing problem pertains to comparing several independent binomial samples. If the response variable defined on a multinomial scale (i.e. *extreme disagree, disagree, neither disagree nor agree, agree, extreme agree*; *No pain, Mild pain, Moderate pain, Severe pain*; etc.), the problem we will discuss pertains to comparing multinomial samples. Further, the solutions for comparing two multinomial samples ($K = 2, J \geq 2$) were already mentioned by existing literature (i.e. Arboretti and Bonnini (2009) ; Kateri and Agresti (2013); etc.).

The solutions to the ordered restricted problems incorporating ordered categorical response variables become extremely difficult when both K and J are large, especially if the sample size is small or moderate. Solutions to these problems may be handled with parametric/semi-parametric methods, so it seems to be more suitable for solving them by nonparametric approaches.

In the field of nonparametric permutation tests, Pesarin (see Pesarin (1990), Pesarin (1992), Pesarin (2001)) performed an in-depth investigation and built a rigorous theoretical framework for nonparametric combination (NPC) methodology which motivates many researchers to be able to cope with more general and complicated problems (i.e. Salmaso, Arboretti Giancristofaro, etc.). In this chapter, we will introduce the basic and important ideas behind the permutation approach and give some necessary relevant definitions and properties. The majority of the notions that will be explored in this chapter are treated in the books Pesarin (2001) and Pesarin and Salmaso (2010) unless otherwise stated. For more details and proofs, please refer to the books (Pesarin (2001); Pesarin and Salmaso (2010)).

2.2 Data Layout

In the beginning, we will start with a two-sample Q-dimensional design. The extensions to more general settings are straightforward to obtain within the NPC.

Let $\mathcal{X}$ be the sample space, on where elementary event $\mathcal{A}$ is defined. Let $\mathcal{F}$ be the collection of events of interest for statistical problems, namely, $\mathcal{A} \in \mathcal{F}$; and P be an assigned probability measure defined on the set of such subsets $\mathcal{F}$ on sample space $\mathcal{X}$, where each P pertains to a nonparametric family $\mathcal{P}$, namely, $P \in \mathcal{P}$. Therefore, the related probability space is in the form of the triple $(\mathcal{X}, \mathcal{F}, \mathcal{P})$.

Let X be a non-degenerate variable defined on the sample space $\mathcal{X}$; also denote an unknown parent distribution by P, and f_P be the underlying likelihood related to P. In what follows, the statistical model is defined by $(X, \mathcal{X}, \mathcal{A}, P)$. Furthermore, in the

following we do not distinguish between the variable X and its observed sample points, the context is sufficient to avoid misunderstanding.

Let $\mathbf{X}_{\cdot j} = \{X_{j,i}, i = 1, \ldots, n_j\} = \{X_{j,1}, X_{j,2}, \ldots, x_{j,n_j}\} \in \mathcal{X}^{n_j}$ be a realization of independent and identically distributed (IID) Q-dimensional random variables with latent distribution P_j with sample size n_j, $j = 1, 2$, where $\min(n_1, n_2) \geq 1$. Each $X_{j,i}$ is a Q-dimensional vector, with the form of $X_{j,i} = (X_{1,j,i}, X_{2,j,i}, \cdots, X_{Q,j,i})^T \in \mathcal{X}_{\mathcal{Q}}$, for $i = 1, 2, \cdots, n_j$, $Q \geq 1$. Denote pooled data by $\mathbf{X} = (\mathbf{X}_{\cdot 1} \biguplus \mathbf{X}_{\cdot 2})$, where $\biguplus$ is the symbol for pooling data. Obviously, the related model of pooled samples $\mathbf{X}$ is defined by $(\mathbf{X}, \mathcal{X}^n, \mathcal{A}^{(n)}, P^{(n)} \in \mathcal{P}^{(n)})$ with $n = n_1 + n_2$ and $P^{(n)} = P_1^{(n_1)} \cdot P_2^{(n_2)}$.

With the so-called *unit - by - unit* representation, the aforementioned samples can be summarized as

$$\mathbf{X} = \{X_i = X(i), i = 1, \cdots, n; n_1, n_2\} = \{X_1, \ldots, X_{n_1}, X_{n_1+1}, \ldots, X_n\},$$

where the first n_1 data in the list belong to first sample and the rest belong to the second sample.

Define unit labels as $\mathbf{u} = (1, \ldots, n)$, assume $\mathbf{u}^* = (u_1^*, \ldots, u_n^*)$ is one permutation of $\mathbf{u}$. Denote a set of all permutations of $\mathbf{u}$ by $\Pi(\mathbf{u})$, namely, $\mathbf{u}, \mathbf{u}^* \in \Pi(\mathbf{u})$. Similarly, one permutation $\mathbf{X}^*$ of $\mathbf{X}$ can be obtained by two steps: one is to permute $\mathbf{u}$ which is the subscript of X; second is to reorganize elements by sorting the indices of subscripts $\mathbf{u}^*$. And the related complete enumerations are $n!$. Denote the set of all permutations $\mathbf{X}^*$ of $\mathbf{X}$ by $\Pi(\mathbf{X})$, thus any component of $\Pi(\mathbf{X})$ is in the form of

$$\mathbf{X}^* = \{X_i^* = X(u_i^*), i = 1, \ldots, n; n_1, n_2\} = \{X_1^*, \ldots, X_{n_1}^*, X_{n_1+1}^*, \ldots, X_n^*\},$$

and so $\mathbf{X}^* = (\mathbf{X}_{\cdot 1}^* \biguplus \mathbf{X}_{\cdot 2}^*)$, where $\mathbf{X}_{\cdot 1}^* = \{X_i^* = X(u_i^*), i = 1, \cdots, n_1\}$ and $\mathbf{X}_{\cdot 2}^* = \{X_i^* = X(u_i^*), i = n_1 + 1, \cdots, n\}$ denote the permuted sample and the two permuted samples, respectively.

2.3 Basic testing problem

The testing problem for detecting the distributional equality in the null hypothesis against stochastic dominance in the alternative typically leads to a one-sided testing problem.

Take the same clinical trial example, studied subjects are randomly assigned two levels of treatment, X_1 received *Standard treatment*, the other X_2 received *Placebo*. We presuppose that both treatment schemes have yielded nonnegative effects in H_1, denote

the *Standard treatment* effect by δ_1 and *Placebo* effect by δ_2. And further, we expect that if the standard treatment has an effect, thus δ_1 is larger than or at least equal to δ_2 with reference from prior information, such as $\delta_1 \geq \delta_2 \geq 0$.

The related hypotheses under testing are expressed as

$$H_0 : X_1 \overset{d}{=} X_2 \overset{d}{=} X \equiv P_1 = P_2 \quad vs. \quad H_1 : X_1 + \delta_1 \overset{d}{>} X_2 + \delta_2.$$

where $\overset{d}{=}$ and $\overset{d}{>}$ take the meaning of *equal in distribution* and *stochastic dominance in distribution*, respectively.

Note that under H_0, $\overset{d}{=}$ infers that variables from both sides of the equation have the same underlying distribution P, namely, $P_1 = P_2 = P$. Thus, the joint null likelihood of $\mathbf{X}$ takes the form of

$$f_P(\mathbf{X}) = \prod_{i=1}^{n_1} f_{P_1}(X_{1,i}) \prod_{i=1}^{n_2} f_{P_2}(X_{2,i}) = \prod_{j=1}^{2} \prod_{i=1}^{n_j} f_P(X_{j,i})$$

which is invariable with respect to any permutation $\mathbf{X}^*$,

$$f_P(\mathbf{X}^*) = \prod_{i=1}^{n_1} f_{P_1}(X^*_{1,i}) \prod_{i=1}^{n_2} f_{P_2}(X^*_{2,i}) = \prod_{j=1}^{2} \prod_{i=1}^{n_j} f_P(X^*_{j,i}) = f_P(\mathbf{X})$$

This shows that data under H_0 are exchangeable, which is permutable. The aforementioned exchangeability can be directly presented as

$$(X_{1,1}, \cdots, X_{1,n_1}, X_{2,1}, \cdots, X_{2,n_2}) \overset{d}{=} (X^*_1, \cdots, X^*_{n_1}, X^*_{n_1+1}, \cdots, X^*_n)$$

where the marginal distributions of $X_{j,i}$ are identical.

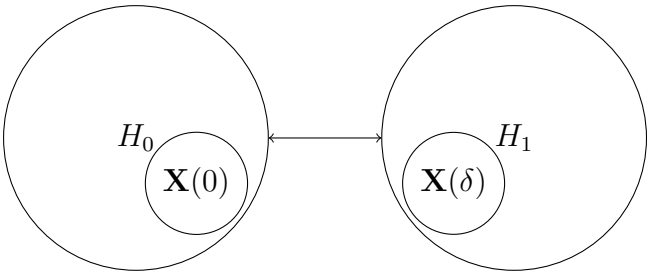
For simplicity purposes, let $\delta_1 > 0$, and $\delta_2 = 0$, thus we only need to retain one effect $\delta = \delta_1 > 0$, the one-sided alternative can be rewritten as $H_1 : X_1 + \delta \overset{d}{>} X_2 = X$. Without loss of generality, random effects may differ depend on units, that is $(X_{1,i}, \delta_i)$ for $i = 1, 2, \cdots, n_1$, and provided that $X_{1,i} + \delta_i \geq X_{1,i}$ with at least one strict inequality.

Thus the hypotheses can be rewritten as follows

$$H_0 : \delta \overset{d}{=} 0 \quad vs. \quad H_1 : \delta \overset{d}{>} 0$$

We may denote the data set by $\mathbf{X}(\delta) = \{X_{1,1} + \delta_1, \ldots, X_{1,n_1} + \delta_{n_1}, X_{2,1}, \ldots, X_{2,n_2}\}$ under H_1, and $\mathbf{X}(0) = \{X_{1,1}, \ldots, X_{1,n_1}, X_{2,1}, \ldots, X_{2,n_2}\}$ under H_0. In the Figure (2.1), the relationship of $\mathbf{X}(0)$ and $\mathbf{X}(\delta)$ are clearly visualized,

FIGURE 2.1: Representation of $\mathbf{X}(0)$ and $\mathbf{X}(\delta)$ in H_0 and H_1, respectively



2.4 Permutation principles

By the definition of sufficient statistic, the pooled data set $\mathbf{X}$ is a sufficient statistic for P in H_0, and we will give a short proof of the description as follows:

Let $T(\cdot)$ be a statistic, and assume it is an identity function, namely, $T(x) = x$. The related conditional distribution of X given T is

$$P(\mathbf{X} = x \mid T(\mathbf{X}) = T(x)) = \frac{P(\mathbf{X} = x \text{ and } T(\mathbf{X}) = T(x))}{P(T(\mathbf{X}) = T(x))} = \frac{P(\mathbf{X} = x)}{P(\mathbf{X} = x)} = 1.$$

It is worth noting that no information on P contained in $\mathbf{X}$ is left on the residual (conditional) distribution $(\mathbf{X} \mid \mathbf{X})$. The related latent likelihood is f_P, thus, $\mathbf{X}$ is the sufficient statistic for the underlying distribution P. Similarly, let us consider $S(\mathbf{X}) = (\mathbf{X} \mid \mathbf{X})$, whose distribution does depend on a constant, in other words, not related to the underlying distribution P. Therefore $S(\mathbf{X})$ plays the role of an *auxiliary/ancillary* statistic.

When P pertains to a nonparametric family $\mathcal{P}$; or P is parametric but the number of its parameters is larger than sample size; or in most cases in which it lies outside the regular exponential family, the pooled data set $\mathbf{X}(0)$ in H_0 is *minimal sufficient*.

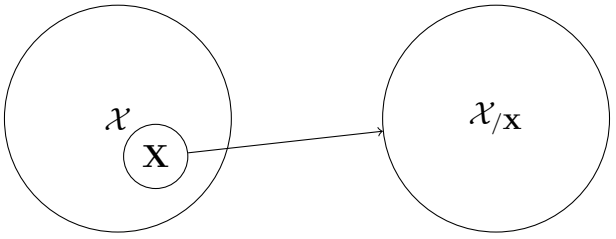
For every permutation $\mathbf{u}^* = (u_1^*, \ldots, u_n^*)$ of $\mathbf{u} = (1, 2, \cdots, n)$, the related permutation of $\mathbf{X}$ is $\mathbf{X}^*$, the likelihood ratio $f_P(\mathbf{X})/f_P(\mathbf{X}^*) = 1$, $\forall \mathbf{u}^* \in \Pi(\mathbf{u})$, thus each permutation $\mathbf{X}^*$ is of equally likely nature. Additionally, pooled data $\mathbf{X}(0)$ are minimal sufficient, and are exchangeable in H_0. Those bases are often termed the *Permutation principle*, and the conditional tests which use that principle, are called *Permutation tests*.

Permutation tests lie within conditional methods of inference. Since it is a conditional procedure, it is worth noting that a permutation test cannot be performed in general until after the data have been observed, and it is also termed the *conditional principle*. Within the principle of conditionality, the permutation procedure for making statistical inferences can be established based on the pooled data $\mathbf{X}$ in H_0.

Further, considering the conditionality principle, we define the related conditional

reference space, like the set that contains points that are equivalent to data $\mathbf{X}$ in the sense that they preserve all its information about P as that contained in $\mathbf{X}$. Denote such related conditional reference space by $\mathcal{X}^n_{/\mathbf{X}}$. To simplify, in what follows, we intend to unify the symbol of the related conditional reference space with $\mathcal{X}_{/\mathbf{X}}$. All distinct

FIGURE 2.2: Conditional reference space $\mathcal{X}_{/\mathbf{X}}$



points $\mathbf{X}^*$ in $\mathcal{X}_{/\mathbf{X}}$ provide the same information as $\mathbf{X}$ (i.e. $\mathbf{X}^* \in \mathcal{X}_{/\mathbf{X}}$ in the Figure (2.2)), and a new related conditional reference space $\mathcal{X}_{/\mathbf{X}^*}$ is generated by data $\mathbf{X}^*$; obviously, $\mathcal{X}_{/\mathbf{X}^*}$ also contains all points which provide the same information as $\mathbf{X}^*$, that is, $\mathbf{X} \in \mathcal{X}_{/\mathbf{X}^*}$, then we can immediately conclude that $\mathcal{X}_{/\mathbf{X}} = \mathcal{X}_{/\mathbf{X}^*}$. In addition, provided that $f_P(\mathcal{X}) > 0$, it is easy to know that the likelihood ratio $f_P(\mathbf{X})/f_P(\mathbf{X}^*)$ is $(\mathbf{X}, \mathbf{X}^*)$-invariant, and P-independent, and so it is distributional free over the class of all possible distributions. In other words, $\mathcal{X}_{/\mathbf{X}}$ is a set of sufficient statistics, corresponding to the *orbit* of equivalent points associated with $\mathbf{X}$. The conditional space $\mathcal{X}_{/\mathbf{X}}$ can be presented in the form of

$$\mathcal{X}_{/\mathbf{X}} = \left\{ \bigcup_{\mathbf{u}^* \in \Pi(\mathbf{u})} [X(u_i^*), i = 1, 2, \cdots, n] \right\} = \Pi(\mathbf{X})$$

In H_1, the set of sufficient statistics is the pair $(\mathbf{X}_{\cdot 1}; \mathbf{X}_{\cdot 2})$. It is obvious that the pair data are not exchangeable between two groups, but are exchangeable within groups.

2.4.1 Some relevant properties

Based on the principles of the permutation procedure, some relevant properties can be inferred for making statistical inferences. In addition, suppose the symbols in what follows share the same meaning as aforementioned notations.

(1) The act of conditioning on a set of sufficient statistics for P in H_0 entails that any conditional inference is independent of the underlying population distribution P, where the distribution P may be univariate, multivariate, normal, categorical, and so forth. This conditionality principle gives rise to the following fundamental property:

Let $(\mathcal{X}, \mathcal{F}, \mathcal{P})$ be the probability space related to variable X, then sufficiency of $\mathbf{X}$ for underlying P, under H_0, implies that the null conditional probability of any event $\mathcal{A} \in \mathcal{F}$, given $\mathbf{X}$, is independent of P, i.e.

$$\Pr\{\mathbf{X}^* \in \mathcal{A}; P \mid \mathbf{X}\} = \Pr\{\mathbf{X}^* \in \mathcal{A} \mid \mathbf{X}\} = \Pr[\mathcal{A} \mid \mathbf{X}].$$

Since every $\mathbf{X}^* \in \mathcal{X}_{/\mathbf{X}}$ is sufficient for underlying distribution P in H_0, $\mathcal{X}_{/\mathbf{X}}$ can be also considered as playing the role of a sufficient space, i.e.

$$\Pr[\mathcal{A} \mid \mathbf{X}] = \Pr[\mathcal{A} \mid \Pi(\mathbf{X})] = \Pr[\mathcal{A} \mid \mathcal{X}_{/\mathbf{X}}]$$

Three relevant consequences of this property are:

(i) Under H_0, all distinct permutations $\mathbf{X}^*$ of $\mathbf{X}$ are of equally likely nature.

As for finite n, the number $M = M^{(n)} = \sum_{\Pi(\mathbf{X})} I[\mathbf{X}^* \in \Pi(\mathbf{X})]$ of points in $\Pi(\mathbf{X})$ is finite and $\forall \mathbf{X}^* \in \Pi(\mathbf{X})$ it is $f_P^{(n)}(\mathbf{X}^*)d\mathbf{X}^* = f_P^{(n)}(\mathbf{X})d\mathbf{X}$, the conditional null probability of any $\mathcal{A} \in \mathcal{F}$ given $\Pi(\mathbf{X})$ is the count ratio:

$$\begin{aligned}\Pr\{\mathbf{X}^* \in \mathcal{A} \mid \Pi(\mathbf{X})\} &= \frac{\sum_{\mathbf{X}^* \in \mathcal{A}} f_P^{(n)}(\mathbf{X}^*)d\mathbf{X}^*}{\sum_{\mathbf{X}^* \in \Pi(\mathbf{X})} f_P^{(n)}(\mathbf{X}^*)d\mathbf{X}^*} \\ &= \sum_{\Pi(\mathbf{X})} \frac{I(\mathbf{X}^* \in \mathcal{A})}{M} = \#(\mathbf{X}^* \in \mathcal{A})/M\end{aligned}$$

where $\#(\cdot)$ is the number of elements of $\Pi(\mathbf{X})$ that satisfy condition $(\cdot)$.

(ii) In H_0, the data set $\mathbf{X}$ is uniformly distributed over $\Pi(\mathbf{X})$ conditionally, i.e. all points of $\Pi(\mathbf{X})$ are equally likely:

$$\Pr(\mathbf{X}^* = \mathbf{x} \mid \Pi(\mathbf{X})) = \begin{cases} M^{-1} & \forall \mathbf{x} \in \Pi(\mathbf{X}) \\ 0 & \forall \mathbf{x} \notin \Pi(\mathbf{X}) \end{cases}$$

(iii) Let $\mathbf{T} = (T_1, \cdots, T_S)^T$ be a vector of $S \geq 1$ permutation statistics (e.g. tests), and $\varphi : \mathcal{R}^S \to \mathcal{R}^1$ is any measurable function, then the conditional null distribution of φ is independent of P; indeed,

$$\begin{aligned}\Pr(\varphi(T_1^*, \cdots, T_S^*) \leq z; P \mid \mathbf{X}) &= \Pr(\varphi(T_1^*, \cdots, T_S^*) \leq z \mid \mathbf{X}) \\ &= \Pr(\varphi_{\mathbf{T}}^{-1}(z) \mid \mathbf{X}) = \frac{\#(\mathbf{X}^* \in \varphi_{\mathbf{T}}^{-1}(z))}{M}\end{aligned}$$

since, due to measurability of φ, $\forall z \in \mathcal{R}^1$, it is $\varphi_{\mathbf{T}}^{-1}(z) \in \mathcal{F}$.

It is worth noting that: i) the conditional probability $\Pr[\mathcal{A} \mid \mathbf{X}]$ has always an *objective existence*; ii) the conditional null distribution of φ is independent of all dependence parameters underlying $\mathbf{T}$; iii) to characterize sufficiency of $\mathbf{X}$ in H_0, permutation tests require the existence of a likelihood $f_P(\mathbf{X}) > 0$, not its calculability; iv) when $\mathbf{X}$ is minimal sufficient for P, it makes no sense to work outside the permutation testing principle (Pesarin (2015)); v) permutation tests are nonparametric, distribution-free and intrinsically robust.

(2) Based on property (1), if X is a continuous variable and T is a continuous non-degenerate function, then in H_0 and $\forall \mathbf{X} \in \mathcal{X}^n$, the ***p*-value-like statistic** $\lambda_T^0 = \lambda_T(\mathbf{X}) = \Pr\{T(\mathbf{X}^*) \geq T^0 = T(\mathbf{X}) \mid \Pi(\mathbf{X})\}$ is uniformly distributed over its attainable support:

$$\Pr\{\lambda_T(\mathbf{X}) \leq \alpha \mid \Pi(\mathbf{X})\} = \alpha, \quad 0 \leq \alpha \leq 1$$

so, α (attainable) plays the role of critical value for λ, whatever the test statistic T is chosen. It is worth noting that λ is a proper p-value for test T only if H_0 is true; thus, it can be viewed as a unifying way to define test statistic.

(3) A permutation test T is exact if its null distribution essentially depends only on exchangeable null error deviates $\mathbf{X}$.

(4) For unidimensional one-sided testing problem, suppose a permutation test is based on difference between two non-degenerate non-decreasing sample statistics, i.e. $T^*(\delta) = S_1(\mathbf{X}_1^*(\delta)) - S_2(\mathbf{X}_2^*(0))$, for $j = 1, 2$. The above-mentioned permutation test being conditionally on data $\mathbf{X}$, is **conditionally unbiased** for every attainable α and every $\delta \overset{d}{>} 0$.

Let us assume that S_j, $j = 1, 2$, are symmetric functions which are invariant with respect to data entry: the value $S_j(\cdot)$ remains the same for all permutations of $\mathbf{X}_j$, namely, $S_j(\mathbf{X}_j) = S_j(\mathbf{X}_j^\star)$, with $\mathbf{X}_j^\star$ being any rearrangement (i.e. within-sample permutation) of $\mathbf{X}_j$ holds for $j = 1, 2$.

In particular:

$$\Pr(\lambda_T(\mathbf{X}(\delta)) \leq \alpha \mid \Pi(\mathbf{X})) \geq \Pr(\lambda(\mathbf{X}(0))_T \leq \alpha \mid \Pi(\mathbf{X})) = \alpha$$

be equivalent to

$$\lambda_T(\mathbf{X}(\delta)) = \Pr(T^*(\delta) \geq T^o(\delta) \mid \mathcal{X}_{/\mathbf{X}(\delta)}) \leq \Pr(T^*(0) \geq T^o(0) \mid \mathcal{X}_{/\mathbf{X}(0)}) = \lambda_T(\mathbf{X}(0))$$

here, $T^o(\cdot)$ means that the observed value of test statistic T.

Thus, $\forall \delta \overset{d}{>} 0$, p-value statistics are such that $\lambda(\mathbf{X}(\delta)) \overset{d}{\leq} \lambda(\mathbf{X}(0))$ (i.e., the uniform dominance property); this also confers α (attainable) the role of critical value for λ. Moreover, if there exist δ_1, δ_2 and δ_3, satisfying $\delta_1 \overset{d}{>} \delta_2 \overset{d}{>} 0 \overset{d}{>} \delta_3$, then

$$\lambda(\mathbf{X}(\delta_1)) \overset{d}{\leq} \lambda(\mathbf{X}(\delta_2)) \overset{d}{\leq} \lambda(\mathbf{X}(0)) \overset{d}{\leq} \lambda(\mathbf{X}(\delta_3))$$

(5) The latter property can be used to define the unconditional/population power of a permutation test T, as a function of $(\delta, \alpha, T, P, n)$. Of course, such population power function is defined as

$$W(\delta, \alpha, T, P, n) = E_{P^n}\left[\Pr(\lambda_T(\mathbf{X}(\delta)) \leq \alpha \mid \Pi(\mathbf{X}))\right].$$

Of course, $W(\delta, \alpha, T, P, n) \geq W(0, \alpha, T, P, n) = \alpha \geq W(\delta', \alpha, T, P, n)$, $\forall \alpha > 0$, since, by (4), the integrand is $\geq = \leq \alpha$ as $\delta \overset{d}{>} 0 \overset{d}{>} \delta'$, for all $P \in \mathcal{P}$, all sample sizes (n_1, n_2) and all designs (i.e., the so-called uniform monotonicity of unconditional power).

(6) The null permutation distribution of any S-dimensional statistic $\mathbf{T}$ given $\mathbf{X}$ could be determined by the property (1), say $F(\mathbf{t} \mid \Pi(\mathbf{X}))$. The procedure based on conditional sufficient requires the complete enumerations of $\Pi(\mathbf{X})$. When the sample size is large or the structure of data is complex, to examine the complete enumerations of elements in $\Pi(\mathbf{X})$ becomes extremely difficult. Thus, based on current computational knowledge, the Conditional Monte Carlo procedure can provide reasonable estimates of it at any required degree of accuracy. The simplest correct way to do this requires $R > 1$ independent random permutations from $\Pi(\mathbf{X})$. Thus, $\forall \mathbf{t} \in \mathcal{R}^S$,

$$\widehat{F}(\mathbf{t} \mid \Pi(\mathbf{X})) = \#(\mathbf{T}^* \leq \mathbf{t})/R$$

gives an unbiased and strongly consistent estimate of $F(\mathbf{t} \mid \Pi(\mathbf{X}))$.

(7) Under H_0, the empirical probability measure (EPM) is a permutation invariant function over $\Pi(\mathbf{X})$. Define EPM by $\widehat{P}_{\mathbf{X}}(\mathcal{A}) = \#(X_i \in \mathcal{A})/n$, for every $\mathcal{A} \in \mathcal{F}$. If $\mathbf{X}^*$ is a realization of one permutation of $\mathbf{X}$, based on the exchangeability in H_0, then $\#(X_i^* \in \mathcal{A}) = \#(X_i \in \mathcal{A})$ holds for $\forall \mathbf{X}^* \in \mathcal{X}_{/\mathbf{X}}$, i.e. $\widehat{P}_{\mathbf{X}^*}(\mathcal{A}) = \widehat{P}_{\mathbf{X}}(\mathcal{A})$, $\widehat{P}_{\mathbf{X}}(\cdot)$ is also a symmetric function.

Indeed, we can infer the data information provided by $\widehat{P}_{\mathbf{X}}$, such inferred information is the same with the information provided by $\Pi(\mathbf{X})$. In other words, conditioning on $\Pi(\mathbf{X})$ is equivalent to conditioning on $\widehat{P}_{\mathbf{X}}$. Thus, EPM $\widehat{P}_{\mathbf{X}}$ is a sufficient function in H_0 for any underlying P. Define the empirical distribution function (EDF) by $\widehat{F}_{\mathbf{X}^*} = \#(\mathbf{X}_i^* \leq x)/n = \widehat{F}_{\mathbf{X}}(x)$, which is also a permutation invariant function over $\Pi(\mathbf{X})$ and is also a sufficient function.

2.5 The Nonparametric Combination (NPC) Methodology

When the testing problems involve some complex dilemmas, such as: complex-structural data; unknown dependent structure among variables; multivariate problems; homogeneity of several groups; and so forth, the nonparametric combination (NPC) approach provides a general methodology for such dilemmas. In particular, the NPC testing solution performs Roy (1953)'s Union-Intersection approach in a general nonparametric setting when an equivalent set of sub-problems is properly carried out.

Suppose the original testing problem can be appropriately and equivalently broken-down into a set of sub-problems, each sub-problem describes partially the original one, but the joint sub-problems equivalently describe the overall testing problem. Therefore, the overall hypotheses H_0 and H_1 are equivalently broken-down into $S \geq 2$ pairwise sub-hypotheses: H_{0s} against H_{1s} for $s = 1, 2, \cdots, S$.

Thus the hypotheses can be presented as follows:

$$H_0 \equiv \bigcap_{s=1}^{S} H_{0s} \quad vs. \quad H_1 \equiv \bigcup_{s=1}^{S} H_{1s}$$

It is worth noting that S sub-hypotheses can be one- and/or two-sided, simple and/or composite.

2.5.1 Partial permutation tests

Considering each pairwise sub-hypothesis: H_{0s} against H_{1s}, for $s = 1, 2, \cdots, S$. Suppose also that each sub-hypothesis is provided with a partial permutation test T_s. In particular, all partial permutation test T_s should satisfy some specific assumptions which are required for the nonparametric combination approach:

(i) All partial permutation tests T_s hold marginally unbiased and are significant for large values. This implies that the values of partial tests are stochastically larger in H_{1s} than in H_{0s} in both conditionally and unconditionally cases, for $s = 1, 2, \cdots, S$, so that its p-value-like statistics should satisfy:

$$\Pr(\lambda_s(\mathbf{X}(\delta_s) \leq \alpha \mid \Pi(\mathbf{X}), H_{1s}) \geq \Pr(\lambda_s(\mathbf{X}(0) \leq \alpha \mid \Pi(\mathbf{X}), H_{0s}) = \alpha, \quad \forall \alpha > 0$$

It is worth noting that in the large majority of situations the set of p-values $\lambda_1, \lambda_2, \cdots, \lambda_S$ are positively dependent (in the sense of Lehmann (1986)).

(ii) At least one partial permutation test T_s is consistent, so implying that the power of T_s in H_{1s} converges to 1 as the sample size n tends to infinity, that shows as follows,

$$\Pr(T_s \geq T_{s\alpha} \mid H_{1s}) \overset{n\to\infty}{\longmapsto} 1, \quad \forall \alpha > 0, \text{ for at least one } s = 1, 2, \cdots, S.$$

where $T_{s\alpha}$ is assumed to be finite, and playing the role of the permutation α-critical value of T_s.

2.5.2 Some important properties of Combining functions

The nonparametric combination approaches can be achieved by both Ψ and ψ combining functions assumed to be measurable. Combining functions Ψ are applied to the values of a set of partial tests $T_1, \cdots, T_S$, and ψ are applied to a set of p-values statistics $\lambda_1, \cdots, \lambda_S$ associated to partial tests $T_1, \cdots, T_S$.

The global hypotheses are tested by combining S dependent partial permutation tests:

$$T_\psi = \Psi(T_1, \cdots, T_S) \equiv \psi(\lambda_1, \cdots, \lambda_S)$$

Take into account combining functions ψ, multiple testing problems and the significant evidence of large values motivate ψ to satisfy some specific properties.

a1) If one p-value is dominant than the counterpart one, the corresponding combining functions ψ have a non-increasing tendency, namely, $\psi(\cdots, \lambda_s, \cdots) \leq \psi(\cdots, \lambda'_s, \cdots)$ holds, if $\lambda_s > \lambda'_s$.

a2) Each ψ must attain its supremum value $\bar{\psi}$, even when only one argument attains zero: $\psi(\cdots, \lambda_s, \cdots) \to \bar{\psi}$ if λ_s tends to zero.

a3) For all $\alpha > 0$, critical value $T_{\psi\alpha}$ of ψ is finite and satisfies $T_{\psi\alpha} < \bar{\psi}$.

To facilitate understanding of the procedure of NPC approach, the following table shows the steps of the procedure. It corresponds to the NPC procedure for a general problem with S partial tests, R random permutations and combining function ψ.

TABLE 2.1: Representation of the conditional Monte Carlo method

$\mathbf{X}$	$\mathbf{X}_1^*$	$\cdots$	$\cdots$	$\mathbf{X}_r^*$	$\cdots$	$\mathbf{X}_R^*$
T_1^o	T_{11}^*	$\cdots$	$\cdots$	T_{1r}^*	$\cdots$	T_{1R}^*
$\vdots$	$\vdots$			$\vdots$		$\vdots$
T_S^o	T_{S1}^*	$\cdots$	$\cdots$	T_{Sr}^*	$\cdots$	T_{SR}^*
			$\downarrow$			
$\widehat{\lambda}_1^o$	$\widehat{L}_{11}^*$	$\cdots$	$\cdots$	$\widehat{L}_{1r}^*$	$\cdots$	$\widehat{L}_{1R}^*$
$\vdots$	$\vdots$			$\vdots$		$\vdots$
$\widehat{\lambda}_S^o$	$\widehat{L}_{S1}^*$	$\cdots$	$\cdots$	$\widehat{L}_{Sr}^*$	$\cdots$	$\widehat{L}_{SR}^*$
			$\downarrow T_\psi$			
T_ψ^o	$T_{\psi 1}^*$	$\cdots$	$\cdots$	$T_{\psi r}^*$	$\cdots$	$T_{\psi R}^*$
			$\downarrow$			
			$\widehat{\lambda}_\psi$			

The Table (2.1) displays the procedure of the representation, the corresponding steps are as follows:

b1) Compute the observed values of permutation test $\mathbf{T} = (T_1, \cdots, T_S)^T$, namely, $\mathbf{T}^o = \mathbf{T}(\mathbf{X}) = (T_1^o, \cdots, T_S^o)^T$.

b2) We intend to compute the values of permutation test $\mathbf{T} = (T_1, \cdots, T_S)^T$ in the permutation space. Indeed, it is obviously difficult to compute the complete enumerations of $\Pi(\mathbf{X})$, thus, we randomly select R permutations from the related permutation space and calculate the corresponding values of permutation tests, that is, $\mathbf{T}_r^* = \mathbf{T}(\mathbf{X}_r^*) = (T_{1r}^*, \cdots, T_{Sr}^*)^T$, for $r = 1, \cdots, R$. R defaults to a sufficiently large value (i.e., 5 000; 10 000; etc.).

b3) The consistent estimate of null permutation distribution $F(\mathbf{t} \mid \Pi(\mathbf{X}))$ of statistic $\mathbf{T}$ be

$$\widehat{F}(\mathbf{t} \mid \Pi(\mathbf{X})) = \frac{\frac{1}{2} + \sum_r I(\mathbf{T}_r^* \leq \mathbf{t})}{R+1}, \quad \forall \mathbf{t} \in \mathcal{R}^S$$

Therefore, the corresponding estimate of the permutation normalized empirical significance level functions be

$$\widehat{L}^*_{sr} = \widehat{L}^*_s(T^*_{sr}) = \frac{\frac{1}{2} + \sum_{j=1}^{R} I(T^*_{sj} \geq T^*_{sr})}{R+1}$$

Compute $\lambda^o_s = \widehat{L}^*_s(T^o_s)$, and let $\lambda^*_{sr} = \widehat{L}^*_s(T^*_{sr})$, it is worth noting that all these p-values are measured in $(0, 1)$.

b4) Since we obtain a set of p-values, $\widehat{\boldsymbol{\lambda}}^o = (\widehat{\lambda}^o_1, \cdots, \widehat{\lambda}^o_S)$ and $\widehat{\mathbf{L}}^*_r = (\widehat{L}^*_{1r}, \cdots, \widehat{L}^*_{Sr})$. The combined observed $T^o_\psi = \psi(\widehat{\lambda}^o_1, \cdots, \widehat{\lambda}^o_S)$ and permutation values $T^*_{\psi r} = \psi(\widehat{L}^*_{1r}, \cdots, \widehat{L}^*_{Sr})$, for $r = 1, \cdots, R$.

b5) Compute the estimate of the p-value of the test as $\widehat{\lambda}_\psi = \sum_r I(T^*_{\psi r} \geq T^o_\psi)/R$.

Under H_0, the sub-matrix $\{T^*_{sr}\}_{S\times R}$ simulates the S-dimensional null distribution of S partial permutation tests. The sub-matrix $\{\widehat{\lambda}_{sr}\}_{S\times R}$ presents the permutation normalized empirical significance level functions corresponding to the sub-matrix $\{T^*_{sr}\}_{S\times R}$. And the sub-vector $\{T^*_{\psi r}\}_R$ simulates the null permutation distribution of combined test T_ψ.

Thus, the statistic $\widehat{\lambda}_\psi$ gives an unbiased and, as R diverges, a strongly consistent estimate of the p-value statistic λ_ψ of T_ψ.

Under H_1, at least one T^o_s presents larger observed values than in H_0; so, if the combining function ψ is non-decreasing in each argument, the p-value statistic satisfies the relation: $\widehat{\lambda}_{\psi;H_1} \overset{d}{\leq} \widehat{\lambda}_{\psi;H_0}$ uniformly for every data set $\mathbf{X}$ and every underlying distribution F. Hence, the latter justifies that H_0 is rejected when $\widehat{\lambda}_\psi \leq \alpha$; moreover, it can be proved that T_ψ is provided with the unbiasedness and consistency properties.

More specifically, let's zoom in on the sub-matrix $\{T^*_{sr}\}_{S\times R}$ of Table (2.1), to see how the procedure works within two-sample Q-dimensional design,

TABLE 2.2: Representation of a two-sample Q-dimensional permutation

	$\mathbf{X}$			$\mathbf{X}_1^*$			$\cdots$	$\mathbf{X}_R^*$		
	$X_1(1)$	$\cdots$	$X_Q(1)$	$X_1(u_1^*)$	$\cdots$	$X_Q(u_1^*)$	$\cdots$	$X_1(v_1^*)$	$\cdots$	$X_Q(v_1^*)$
$\mathbf{X}_{\cdot 1}^*$	$\vdots$		$\vdots$	$\vdots$		$\vdots$	$\cdots$	$\vdots$		$\vdots$
	$X_1(n_1)$	$\cdots$	$X_Q(n_1)$	$X_1(u_{n_1}^*)$	$\cdots$	$X_Q(u_{n_1}^*)$	$\cdots$	$X_1(v_{n_1}^*)$	$\cdots$	$X_Q(v_{n_1}^*)$
	$X_1(1+n_1)$	$\cdots$	$X_Q(1+n_1)$	$X_1(u_{1+n_1}^*)$	$\cdots$	$X_Q(u_{1+n_1}^*)$	$\cdots$	$X_1(v_{1+n_1}^*)$	$\cdots$	$X_Q(v_{1+n_1}^*)$
$\mathbf{X}_{\cdot 2}^*$	$\vdots$		$\vdots$	$\vdots$		$\vdots$	$\cdots$	$\vdots$		$\vdots$
	$X_1(n)$	$\cdots$	$X_Q(n)$	$X_1(u_n^*)$	$\cdots$	$X_Q(u_n^*)$	$\cdots$	$X_1(v_n^*)$	$\cdots$	$X_Q(v_n^*)$
	$\downarrow$			$\downarrow$				$\downarrow$		
$\mathbf{T}$	$(T_1^o, \cdots, T_S^o)$			$(T_{11}^*, \cdots, T_{S1}^*)$			$\cdots$	$(T_{1R}^*, \cdots, T_{SR}^*)$		
	$\downarrow$			$\downarrow$				$\downarrow$		
$\boldsymbol{\lambda}$	$(\widehat{\lambda}_1^o, \cdots, \widehat{\lambda}_S^o)$			$(\widehat{L}_{11}^*, \cdots, \widehat{L}_{S1}^*)$			$\cdots$	$(\widehat{L}_{1R}^*, \cdots, \widehat{L}_{SR}^*)$		
	$\downarrow$			$\downarrow$				$\downarrow$		
$\mathbf{T}_\psi$	T_ψ^o			$T_{\psi 1}^*$			$\cdots$	$T_{\psi R}^*$		

In the Table (2.2), $\mathbf{X}$ presents the observed data, and $(\mathbf{X}_1^*, \cdots, \mathbf{X}_R^*) \in \mathcal{X}_{/\mathbf{X}}$ be a set of permutations of $\mathbf{X}$. $\mathbf{X}_{\cdot 1}^*$ presents a sub-matrix including the first n_1 data corresponding to $\{\mathbf{X}, \mathbf{X}_1^*, \cdots, \mathbf{X}_R^*\}$, and the rest belong to the sub-matrix $\mathbf{X}_{\cdot 2}^*$. $\mathbf{T}$ be the S partial permutation tests, namely, $(T_1, \cdots, T_S)^T$.

Take the sub-matrix of $\mathbf{X}$ (Yellow part) for example, each row represents Q-dimensional data related to each element of $\mathbf{X}$. According to the data layout, the aforementioned representation of $\mathbf{X}$ is $\mathbf{X} = (X_1, \cdots, X_{n_1}, X_{1+n_1}, \cdots, X_n)$, the element is, $X_i = (X_1(i), \cdots, X_Q(i))$, for $i = 1, 2, \cdots, n$.

In the case of two-sample Q-dimensional design, the testing problem is to determine the stochastic dominance between two groups, the permutation procedure lies within sufficient space $\mathcal{X}_{/\mathbf{X}}$.

2.5.3 Some relevant Combining Functions

Combining functions ψ define a class $\mathcal{C}$ of possibilities. A sub-class $\mathcal{C}_A \subset \mathcal{C}$ contains admissible functions. A combining function ψ is admissible if its acceptance region is convex in the $(\lambda_1, \cdots, \lambda_S)$ representation (Birnbaum (1954), Birnbaum (1955)).

The admissible combining functions mostly used in practice are:

c1) The Fisher combining function is based on the statistic

$$T_F = -2 \cdot \sum_s \log(L_s^*)$$

T_F is the most popular combining function and corresponds to the *product rule*.

c2) The Liptak combining function is based on the statistic

$$T_L = \sum_s \Phi^{-1}(1 - L_s^*)$$

where Φ is the standard normal cumulative distribution function.

c3) The Tippett combining function is based on the statistic

$$T_T = \max_{1 \le s \le S}(1 - L_s^*)$$

Tippett test is significant for large values.

c4) The Direct combining function Ψ applied to the set of partial tests is:

$$T_D = \sum_s T_s^*$$

It is worth mentioning that, the direct combining function is suitable if all partial permutation tests T_k^* at least should share the same asymptotic null distribution so satisfying Anderson-Darling's spirit.

NPC gives us a significant evidence to reject the null hypothesis when the value of T_ψ is sufficient large. It is worth emphasizing that, if the partial tests are exact, unbiased and at least one is consistent, then the NPC methodology yields exact, unbiased and consistent solutions. Further, if we can determine the best power among the S partial tests, then the power of combined tests is close to that of the best one in theory.

2.6 Conclusion

In this chapter, we introduced some of the most important ideas and properties of permutation tests. We take the simple two-sample Q-dimensional design as a typical example, to establish a preliminary impression of the data layout. The approaches are suitable for balanced samples, and are also equally applicable to the unbalanced case. We assume these two samples share the same underlying distribution P in H_0,

in which case the pooled data $\mathbf{X}$ are exchangeable and sufficient. When underlying distribution P lies within nonparametric family $\mathcal{P}$, or incorporates infinitely many parameters, or cannot be converted to the regular exponential family, $\mathbf{X}$ plays the role of minimum sufficient in H_0. The aforementioned bases give rise to the Permutation principle within the conditionality principle of inference. The conditional tests based on the permutation principle are termed permutation tests. When we are trying to make use of a nonparametric permutation approach, it is essential to keep in mind that nonparametric permutation approaches must work within the principles of sufficiency and conditionality.

The nonparametric combination (NPC) is a well-known methodology based on the permutation theory. According to Roy's Union-Intersection approach, the NPC procedure subdivides the global hypothesis into some suitable pairwise sub-hypotheses, and each pairwise sub-hypotheses has appropriate partial permutation tests, the resulting is achieved by jointly analyzed all of them. Common combining functions contain Fisher's, Liptak's, Tippett's and the Direct; all of them are significant for large values to reject H_0, are unbiased, consistent and admissible. Of course, in the face of different or complicated situations, the NPC method is a flexible general tool for facing with most complex testing problems.

In the next chapter, we will further explore the testing problem which involves monotonic stochastic ordering case, especially, the case including multivariate data in which comparison groups are more than two.

Chapter 3

Testing for restricted alternatives with ordered data

3.1 Prologue

Without loss of generality, we will start with a clinical example. Suppose a study where subjects are randomly assigned J-levels of treatment, and then at a fixed observation point, subjects are asked to rate the response score for assigned treatment scheme or researchers determine the subsequent treatment effects according to a categorical scale. Some important aspects should be highlighted here:

d1) Each subject can be seen to behave equally likely. Assume each subject has a similar physical and health condition before receiving the treatment scheme, regardless of either the place where they are from, or their educational background, or some other aspects which should influence the outcomes in theory.

d2) Each subject is randomly assigned to the treatment scheme, moreover, they can be assumed to behave independently.

d3) The outcomes lie within categorical measurement, and they are deemed to be ordered because of psychological anticipated based on common sense. Thus, outcomes are defined on K ordinal categorical measurement with simple ordering $v_1 \prec v_2 \prec \cdots \prec v_K$.

d4) The sample size is usually not very large, so that asymptotic considerations may become improper.

d5) Usually, those treatment schemes would yield nonnegative effects, say δ, and *Standard Treatment* behavior better than *Placebo* based on prior information of researchers.

This typical realistic example can be described in statistical language, and the resulting data can be organized in a contingency table, as shown in Table (3.1).

TABLE 3.1: General Clinical Example

$\mathbf{X}$	Cats.					
	v_1	v_2	v_3	$\cdots$	v_K	Total
$\mathbf{X}_{\cdot 1}$	n_{11}	n_{12}	n_{13}	$\cdots$	n_{1K}	n_1
$\mathbf{X}_{\cdot 2}$	n_{21}	n_{22}	n_{23}	$\cdots$	n_{2K}	n_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\mathbf{X}_{\cdot J}$	n_{J1}	n_{J2}	n_{J3}	$\cdots$	n_{JK}	n_J
Total	$N_{\cdot 1}$	$N_{\cdot 2}$	$N_{\cdot 3}$	$\cdots$	$N_{\cdot K}$	n

A general typical example Table (3.1), the green region presents J groups (explanatory) of random sample receiving J-levels of treatment scheme, the larger the J, the better the effect is expected by the treatment scheme (Psychological anticipated); and the yellow area (responses) includes I categories which satisfy simple ordering $v_1 \prec v_2 \prec \cdots \prec v_K$. n_j, $j = 1, 2, \cdots, J$, are the sample sizes of treated samples, and $N_{\cdot k}$, $k = 1, 2, \cdots, K$ are frequencies of the related categories. The total sample size is $n = \sum_j n_j$.

The problem of interest is to determine whether there exist stochastic dominance or stochastic ordering among J treatments; in other words, whether the explanatory variable satisfies the stochastic ordering restriction, namely, $X_1 \overset{d}{\leq} X_2 \overset{d}{\leq} \cdots \overset{d}{\leq} X_J$, with $X_1 \overset{d}{<} X_J$.

It is termed the typical one-sided testing problem with stochastic dominance or stochastic ordering. Therefore, the related null hypothesis is as follows

$$H_0 : X_1 \overset{d}{=} X_2 \overset{d}{=} \cdots \overset{d}{=} X_J$$

against the one-sided alternative hypothesis

$$H_1 : X_1 \overset{d}{\leq} X_2 \overset{d}{\leq} \cdots \overset{d}{\leq} X_J, \quad \text{with} \quad X_1 \overset{d}{<} X_J.$$

Obviously, the responses are measured on the same ordered categorical scale $\{v_1, \cdots, v_K\}$, they can be considered multinomially distributed. Define the cumulative distribution of responses by F at ordered categories $v_1 \prec v_2 \prec \cdots \prec v_K$, as $F_X(v_k) = \Pr\{X \preceq v_k\}$ for $k = 1, \cdots, K$. Thus, the aforementioned hypotheses are equivalently expressed as

$$H_0 : \{F_{X_1} = F_{X_2} = \cdots = F_{X_J}\} \quad vs. \quad H_1 : \{F_{X_1} \geq F_{X_2} \geq \cdots \geq F_{X_J}\} \quad \text{with } F_{X_1} > F_{X_J}$$

With clear meaning of the symbols, the rationale for this formulation resides in that if, according to increasing j, non-decreasing treatment effects δ yield at latent null variables X, i.e. $\delta_h \leq \delta_j$, $1 \leq h < j \leq J$, then latent variables should behave as

$$X_h = (X + \delta_h) \overset{d}{\leq} (X + \delta_j) = X_j$$

The related testing problem has a rather difficult solution within the likelihood ratio theory, which with categorical data, in addition, presents quite a serious difficulty: even for moderate number of cells it is recognized to be not unique (Cohen *et al.* (2000), Cohen *et al.* (2003); Wang (1996); Silvapulle and Sen (2005); Colombi and Forcina (2016); etc.). Moreover, to get a solution, important supplementary options, difficult to justify in terms of the real problem under study, are required. This difficulty mostly consists in that the set of alternatives is restricted to lie in the $(J-1) \times (K-1)$-*Dimensional positive orthant* where the likelihood cannot be maximized under H_0 by ordinary methods of maximization.

Our solution does firstly consider the setting of two treatment schemes, and then, according to Roy (1953)'s UI and Jonckheere–Terpstra's approaches, by a break-down of the hypothesis into $S-1$ pairs of sub-hypotheses. Later, all resulting dependent partial tests are combined by an NPC method.

3.2 The two-sample unidimensional case

Let us firstly consider the two-sample one-dimensional case, where data are in a $2 \times K$ table, and denote two samples by $\mathbf{X}_{\cdot j}$, $j = 1, 2$, respectively. For the sake of clarity, the related table are shown as

TABLE 3.2: Two-sample one-dimensional case

X	Cats.					
	v_1	v_2	v_3	$\cdots$	v_K	Total
$\mathbf{X}_{\cdot 1}$	n_{11}	n_{12}	n_{13}	$\cdots$	n_{1K}	n_1
$\mathbf{X}_{\cdot 2}$	n_{21}	n_{22}	n_{23}	$\cdots$	n_{2K}	n_2
Total	$N_{\cdot 1}$	$N_{\cdot 2}$	$N_{\cdot 3}$	$\cdots$	$N_{\cdot K}$	n

Thus, the specific hypotheses are expressed as

$$H_0 : X_1 \stackrel{d}{=} X_2 \equiv \{F_1(v_k) = F_2(v_k), k = 1, \cdots K\}$$

against

$$H_1 : X_1 \stackrel{d}{<} X_2 \equiv \bigcup_{k=1}^{K-1} [F_1(v_k) > F_2(v_k)]$$

The related testing problem can be equivalently set as

$$H_0 : F_1 = F_2 \equiv \bigcap_{k=1}^{K-1} [F_1(v_k) = F_2(v_k)]$$

against the set of *restricted alternatives*

$$H_1 : F_1 > F_2 \equiv \bigcup_{k=1}^{K-1} [F_1(v_k) \geq F_2(v_k)], \quad \text{with at least one strict inequality}$$

It is worth noting that:

e1) According to Roy (1953), in H_0, where the two distribution are considered to be identical, such a null hypothesis is equivalent to the intersection of a set of sub-null-hypotheses, where the two cumulative distribution functions for each category in each sub-null-hypothesis are the same. The alternative in H_1 is equivalent to the union of a set of inequalities. Thus, the testing problem has been equivalently broken-down into $K - 1$ *one-sided sub-problems*;

e2) H_1 defines a *multi-one-sided* set of alternatives;

e3) Under both H_0 and H_1, cumulative empirical distribution function satisfies $F_1(v_K) = F_2(v_K) = 1$, and so category v_K is not considered since it does not provide information on any possible diversity;

e4) Thus, the global solution requires the joint comparison of $K-1$ differences random relative frequencies: $\widehat{F}_1(v_k) - \widehat{F}_2(v_k), k = 1, \cdots, K-1$.

For such testing problem, however, the number of unknown nuisance parameters to take care in any $2 \times K$ testing process is $(2 \times K - 1)$, and the likelihood is to be maximized in the $(K-1)$-dimensional positive orthant, indeed a very difficult task especially when $K > 4$. Due to these well-recognized difficulties of likelihood approaches, our approach is to stay within the conditionality principle of inference (please refer to Chapter 2), i.e. by incorporating both the sufficiency principle and the permutation theory. Since, the crucial point for that joint analysis is the proper handling of all underlying dependencies, to attain general solutions we must work within the UI-NPC of related dependent permutation tests because, due to (iii) (see Chapter 2), the estimation of dependence coefficients is not required since NPC works independently of such dependencies, regardless of how complex these are.

Accordingly, the $K-1$ partial test statistics are:

$$T_k^* = C(n_1, n_2) \cdot [\hat{F}_{1k}^* - \hat{F}_{2k}^*] \left[\bar{F}_{\cdot k}(1 - \bar{F}_{\cdot k})\right]^{-\frac{1}{2}}, \;\; k = 1, \cdots, K-1,$$

where this partial permutation tests are based on the cumulative empirical distribution functions which symmetric, non-decreasing and non-degenerate measurable functions; $\hat{F}_{jk}^* = \hat{F}_j^*(v_k) = \sum_{i=1}^{n_j} I(X_{j,i}^* \preceq v_k)/n_j$, $j = 1, 2$, $\bar{F}_{\cdot k} = \sum_{k=1}^{k} N_{\cdot k}/n$ are permutation and marginal empirical distribution functions (EDFs), respectively. Further, $C(n_1, n_2) \left[\bar{F}_{\cdot k}(1 - \bar{F}_{\cdot k})\right]^{-\frac{1}{2}}$ is the permutation variance of $\hat{F}_{1k}^* - \hat{F}_{2k}^*$. It is easy to see that, $\bar{F}_{\cdot k}$, $k = 1, \cdots, K$, since $(n_{1k}^* + n_{2k}^*) = (n_{1k} + n_{2k})$, are permutationally invariant cumulative frequencies obtained from the permuted table $\{f_{jk}^*, k = 1, \cdots, K, j = 1, 2\}$ where margins are fixed due to conditioning (as shown in the Table 3.3).

TABLE 3.3: Two-sample one-dimensional permuted table

$\mathbf{X}$	Cats.					
	v_1	v_2	v_3	$\cdots$	v_K	Total
$\mathbf{X}^*_{\cdot 1}$	n^*_{11}	n^*_{12}	n^*_{13}	$\cdots$	n^*_{1K}	n_1
$\mathbf{X}^*_{\cdot 2}$	n^*_{21}	n^*_{22}	n^*_{23}	$\cdots$	n^*_{2K}	n_2
Total	$N_{\cdot 1}$	$N_{\cdot 2}$	$N_{\cdot 3}$	$\cdots$	$N_{\cdot K}$	n

The aforementioned partial test statistics are suitable for the comparison of two-sample one-dimensional cases, and include some noteworthy properties:

f1) Marginal EDFs $\widehat{F}_{jk}$ are maximum likelihood unbiased estimates of population CDFs $F_j(v_k)$, $k = 1, \cdots, K-1$, $j = 1, 2$;

f2) Each partial permutation test T^*_k is a reformulation of Fisher's exact probability test, and so it is a *best conditional test*;

f3) Large values of each partial test T^*_k are significant against its related null sub-hypothesis H_{1k};

f4) The $K-1$ partial tests are positively dependent;

f5) For the computation of T^*_k, 0 is assigned to expressions with the form 0/0;

f6) $C(n_1, n_2) = [n_1 n_2 (n-1)/n^2]^{1/2}$ is a permutation constant not dependent on k;

f7) For increasing sample sizes, each T^*_k under H_0 converges to the standardized normal distribution: $T^*_k \xrightarrow{d} \mathcal{N}(0, 1)$.

According to the approach discussed in Pesarin (2001), Pesarin (2006) and Pesarin and Salmaso (2010), the global testing solution can be obtained by their UI-NPC while using any admissible combining function, such as Fisher's, Liptak's, Tippett's, the Direct, and so forth.

The simplest admissible combination is by the direct sum of partial tests T^*_k:

$$T^*_{AD} = \sum_{k=1}^{K-1} T^*_k = C(n_1, n_2) \cdot \sum_{k=1}^{K-1} [\hat{F}^*_{1k} - \hat{F}^*_{2k}] \left[\bar{F}_{\cdot k}(1 - \bar{F}_{\cdot k})\right]^{-\frac{1}{2}} .$$

Such a solution looks like the discrete version of the Anderson-Darling goodness-of-fit type test for multi-one-sided alternatives, which is the sum of standardized partial tests.

It is worth emphasizing that: g1) each partial test T_k^* is unbiased and so combined test T_{AD}^* is unbiased; g2) combined test T_{AD}^* is consistent if at least one partial test T_k^* is consistent; g3) T_{AD}^* is an admissible combination of partial best tests and so provided with *good power behavior*.

Of course, by using other admissible combining functions one can obtain other so-called *good solutions*, none of which, however, iss uniformly better than any other. The corresponding p-value-like statistics can be written as $\lambda_{AD} = \Pr\{T_{AD}^* \geq T_{AD}^o \mid \mathbf{X}\}$, where $T_{AD}^o = T_{AD}(\mathbf{X})$ is the observed value of T_{AD} on the pooled data $\mathbf{X}$. So, remembering that p-value-like statistics play the role of tests whose common critical value is α, if $\lambda_{AD} \leq \alpha$, the null hypothesis is rejected at significance level $\alpha > 0$.

Similarly, there are some other useful partial tests for the same unidimensional problem, which can be tackled by considering the *comparison of two probability generating functions* (Arboretti *et al.* (2007); Arboretti and Bonnini (2009); Pesarin and Salmaso (2010)). One is based on Mid-ranks, which converts the original categories into the mid-rank of the categories among the pooled data set $\mathbf{X}$, namely,

$$\begin{aligned} MR(X_{j,i}) =& \#(X_{j,s} \leq X_{j,i}) - \frac{1}{2}\#(X_{j,s} = X_{j,i}) \\ =& \sum_{j=1}^{2}\sum_{i=1}^{n_j} I(X_{j,s} \leq X_{j,i}) - \frac{1}{2}\sum_{j=1}^{2}\sum_{i=1}^{n_j} I(X_{j,s} = X_{j,i}) \end{aligned}$$

Mid-ranks are also symmetric, nondecreasing and non-degenerate measurable functions. To be simplified, let $\varphi(\cdot)$ indicates the Mid-ranks transformation $MR(\cdot)$, namely, $\varphi(X_{j,i}) = MR(X_{j,i})$, and then to proceed by comparing means of mid-ranks. The corresponding test is

$$T_{MR}^* = \overline{\varphi(X_1^*)} - \overline{\varphi(X_2^*)} = \sum_i \varphi(X_{1,i}^*)/n_1 - \sum_i \varphi(X_{2,i}^*)/n_2$$

Since $\sum_j \sum_i \varphi(X_{j,i}^*)$ is a permutation invariant quantity, T_{MR}^* is equivalent to the standardized version $(\overline{\varphi(X_1^*)} - \overline{\varphi(X_2^*)})/Var_{|\mathbf{X}}(\overline{\varphi(X_1^*)} - \overline{\varphi(X_2^*)})^{1/2}$, to $\sum_i \varphi(X_{1,i}^*)/n_1$, and to $-\sum_i \varphi(X_{2,i}^*)/n_2$, where $Var_{|\mathbf{X}}(\overline{\varphi(X_1^*)} - \overline{\varphi(X_2^*)})^{1/2}$ is the permutation variance of $\overline{\varphi(X_1^*)} - \overline{\varphi(X_2^*)}$. Obviously, due to conditioning on pooled data $\mathbf{X}$, the permutation variance $Var_{|\mathbf{X}}(\overline{\varphi(X_1^*)} - \overline{\varphi(X_2^*)})^{1/2}$ is always finite.

It is worth cautioning that, in H_0, pooled data $\mathbf{X}$ are considered to be from the same underlying distribution F. With transformation $\varphi(\cdot)$, we assume that the φ-expectation of any transformed data be $E_F[\varphi(X)] = E_F[\varphi(X(0))] = \int_{\mathcal{X}} \varphi(x)dF(x) < \infty$; i.e. we assume that population mean-φ is finite. The finite φ-mean in H_1 will be such that

$E_F[\varphi(X(\delta))] = \int_{\mathcal{X}} \varphi(z) dP_{H_1}(z) > E_F[\varphi(X(0))]$ for every $\delta > 0$. It is clearly easy to see that, under H_0, the standardized version of T^*_{MR} also converges to standardized normal distribution: $T^*_{MR} \xrightarrow{d} \mathcal{N}(0,1)$. The corresponding p-value-like statistic is $\lambda_{MR} = \Pr\{T^*_{MR} \geq T^o_{MR} \mid \mathbf{X}\}$, where, T^o_{MR} means the observed value of T_{MR} on $\mathbf{X}$. If $\lambda_{MR} \leq \alpha$, the alternative is accepted at significant level $\alpha > 0$.

Another common approach is based on score functions, which were tackled by assigning non-decreasing appropriate W scores to ordered categories, e.g. as $v_k \to w_k$, for $k = 1, \cdots, K$, where $w_1 \leq \cdots \leq w_K$, with at least one strict inequality. In such a case, the data are transformed into $w_{ki} = w_k \cdot \mathbf{1}(X(i) = v_k)$, for $i = 1, \cdots, n$, where $\mathbf{1}(\cdot)$ is the counting function. It is clearly easy to see that, $w^*_1 = \sum_{i=1}^{n_1} \sum_{k=1}^{K} w^*_{ki}$ presents the sum of the first permuted samples, and $w^*_2 = \sum_{i=n_1+1}^{n} \sum_{k=1}^{K} w^*_{ki}$ presents the sum of the second one. Therefore, the permutation solution is nothing else than a comparison of sample means of scores, the related partial permutation test is

$$T^*_W = \bar{w}^*_1 - \bar{w}^*_2$$

Similarly, $\sum_j w^*_j$ is a permutation invariant quantity, T^*_W is equivalent to the standardized version $(\bar{w}^*_1 - \bar{w}^*_2)/Var_{|\mathbf{X}}(\bar{w}^*_1 - \bar{w}^*_2)^{1/2}$, to $\bar{w}^*_1$, and also to $-\bar{w}^*_2$. The aforementioned $\bar{w}^*_j$ is also a symmetric nondecreasing and non-degenerate measurable function. Standardized partial test T^*_W, under H_0, also converges to the standard normal distribution: $T^*_W \xrightarrow{d} \mathcal{N}(0,1)$. Further, a common score transformation is to assign the categories to equidistant numerical data (the distance between categories is normalized to one), integers numbers are common simple equidistant numerical data.

It is clearly easy to see that both Mid-ranks and score solutions share the spirit of Cramér-von Mises goodness-of-fit type statistic for multi-one-sided alternatives, which is the standardized sum of partial tests, and reject the alternatives for the large values of the test statistic.

3.3 The two-sample multidimensional case

In the general multidimensional case, let us start from two-sample Q-dimensional problem, $Q \geq 2$. The formulation of testing for multidimensional hypotheses are $H_0 : X_1 \overset{d}{=} X_2$ against $H_1 : X_1 \overset{d}{<} X_2$. The multidimensional hypotheses H_0 and H_1, according to Roy (1953) are assumed to be equivalently broken-down into $K \geq 2$ unidimensional sub-hypotheses, $H_0 \equiv \bigcap_{k=1}^{K} H_{0k}$ and $H_1 \equiv \bigcup_{k=1}^{K} H_{1k}$. Thus, with Q dimensional ordinal data and K ordered categories for each variable, the hypotheses are equivalently written

as

$$H_0 \equiv \bigcap_{q=1}^{Q} \left\{ \bigcap_{k=1}^{K-1} [F_{1q}(v_k) = F_{2q}(v_k)] \right\}$$

against

$$H_1 \equiv \bigcup_{q=1}^{Q} \left\{ \bigcup_{k=1}^{K-1} [F_{1q}(v_k) > F_{2q}(v_k)] \right\}$$

It is clearly easy to see that the global testing problem is broken-down into $V = Q \times (K-1)$ partial sub-problems.

Thus, for each component variable $q = 1, \cdots, Q$, partial test T^*_{ADq} can be obtained according to Chapter 2. Since all these partial tests are standardized and so, sharing the same asymptotic null distribution, for their combination we can proceed with their direct sum.

This provides for the Q-dimensional extension of the Anderson-Darling test for multi-one-sided alternatives:

$$T^*_{AD} = \sum_{q=1}^{Q} T^*_{ADq} = C(n_1, n_2) \cdot \sum_{q=1}^{Q} \sum_{k=1}^{K-1} [\hat{F}^*_{1qk} - \hat{F}^*_{2qk}] \left[\bar{F}_{\cdot qk}(1 - \bar{F}_{\cdot qk}) \right]^{-\frac{1}{2}}.$$

It is worth highlighting that, now, with symbol $\mathbf{X}$ it is represented the Q-dimensional variable and the pooled sample data matrix, the context generally suffices avoiding misunderstandings. Of course, the Q-dimensional T^*_{AD} enjoys the same *good* properties as the unidimensional.

In place of the direct combination of Q partial tests T^*_{ADq}, i.e. one Anderson-Darling test for each variable, it is possible to think of a more general combination like for instance $T^*_{\psi} = \psi(T^*_{AD1}, \cdots, T^*_{ADQ})$. The most commonly used combining functions ψ are Fisher's $T_F = -2 \sum_q \log(\lambda^*_{ADq})$, or Liptak's $T^*_L = \sum_q \Phi^{-1}(1 - \lambda^*_{ADq})$, where λ^*_{ADq} is the p-value statistic of T^*_{ADq} and $\Phi(\cdot)^{-1}$ is the inverse standard normal CDF. Since in T^*_{AD} all summands are well defined, it is also of some interest to observe that the double summation can equivalently be computed as $\sum_k \sum_q$.

Similarly to the unidimensional setting, the multidimensional problem can, however, be tackled by Mid-ranks T^*_{MR} and Scores T^*_W. For each component variable q, the permutation solution based on Mid-ranks is to proceed by comparing sample means of mid-ranks, $T^*_{MR} = \psi\left((\overline{\varphi_1(X^*_1)} - \overline{\varphi_1(X^*_2)}), \cdots, (\overline{\varphi_Q(X^*_1)} - \overline{\varphi_Q(X^*_2)}) \right)$. It is well-known that all standardized versions of partial tests are sharing the same limiting distribution function, and it is reasonable to combine the resulting partial tests by using their direct

sum. The corresponding test statistic by direct combining function is as follows,

$$T^*_{MR} = \sum_{q=1}^{Q} T^*_{MRq} = \sum_{q=1}^{Q} \left(\overline{\varphi_q(X^*_1)} - \overline{\varphi_q(X^*_2)}\right) Var_{|\mathbf{X}} \left(\overline{\varphi_q(X^*_1)} - \overline{\varphi_q(X^*_2)}\right)^{-\frac{1}{2}}$$

For $Q > 1$, we can assign non-decreasing W_q scores to ordered categories, e.g. $v_{qk} \to w_{qk}$, for $k = 1, \cdots, K$ and $q = 1, \cdots, Q$, where $w_{q1} \leq \cdots \leq w_{qK}$, with at least one strict inequality $\forall q$. In such a case, the data are transformed into $w_{qki} = w_{qk} \cdot \mathbf{1}(X_{qji} = v_{qk})$, for $j = 1, 2$ and $i = 1, \cdots, n$. Thus, the permutation solution is to compare of sample means of scores, $T^*_W = \psi\left((\bar{w}^*_{11} - \bar{w}^*_{12}), \cdots, (\bar{w}^*_{Q1} - \bar{w}^*_{Q2})\right)$. The test statistic with Q-dimensional data based on score function by direct combination is immediately obtained as follows,

$$T^*_W = \sum_{q=1}^{Q} T^*_{Wq} = \sum_{q=1}^{Q} \left(\bar{w}^*_{q1} - \bar{w}^*_{q2}\right) Var_{|\mathbf{X}} \left(\bar{w}^*_{q1} - \bar{w}^*_{q2}\right)^{-\frac{1}{2}}$$

It is suitable to use the Direct combination for T^*_{Wq}, due to the same limiting distribution function of all standardized partial tests. The related p-values for its counterpart-partial tests T^*_{MR} and T^*_W can be summarized as $\{\lambda^*_{MR1}, \cdots, \lambda^*_{MRQ}\}$ and $\{\lambda^*_{W1}, \cdots, \lambda^*_{WQ}\}$, respectively. Other combining functions are also suitable for those p-values, such as Fisher's, Liptak's and Tippett's.

3.4 The J-sample stochastic ordering problem

Let's turn our attention to the Table (3.1), the testing problem of interest is to detect whether there exists stochastic ordering among J samples, namely, $H_1 : X_1 \overset{d}{\leq} X_2 \overset{d}{\leq} \cdots \overset{d}{\leq} X_J$, with $X_1 \overset{d}{<} X_J$. According to the Jonckheere-Terpstra idea, the $J \times K$ table can be broken down into $(J-1)$ sub-tables. Thus, the global testing problem is broken down into $(J-1)$ sub-problems each based on a $2 \times K$ sub-table.

To be specific, for any $j \in \{1, \cdots, J-1\}$, we divide the data set $\mathbf{X}$ into two pooled pseudo-groups, where the first pseudo-group is obtained by pooling data of the first j ordered groups and the second by pooling the rest. With the symbol $\biguplus$ for pooling data sets, the procedure considers the first pooled pseudo-group as

$$\mathbf{Y}_{1(j)} = \mathbf{X}_{\cdot 1} \biguplus \mathbf{X}_{\cdot 2} \biguplus \cdots \biguplus \mathbf{X}_{\cdot j}$$

and the second as

$$\mathbf{Y}_{2(j)} = \mathbf{X}_{\cdot j+1} \biguplus \mathbf{X}_{\cdot j+2} \biguplus \cdots \biguplus \mathbf{X}_{\cdot J}$$

for $j = 1, \cdots, J-1$, where $\mathbf{X}_{\cdot j} = \{X_{j,i}, i = 1, \cdots, n_j\}$ is the data set in the j-th group.

In the null hypothesis H_0, related pooled variables satisfy the relationships $Y_{1(j)} \overset{d}{=} Y_{2(j)}$, $j = 1, \cdots, J-1$, hence, data from every pair of pseudo-groups are exchangeable. In the alternative H_1, as for at least one j the relation inequality $X_j \overset{d}{\leq} X_{j+1}, 1 \leq j \leq J-1$ is strict, which leads to the corresponding stochastic dominance between each pair of pseudo-groups $Y_{1(j)} \overset{d}{<} Y_{2(j)}$ for $j \leq J-1$, where the strict equality is satisfied for all $J-1$ pseudo-groups. Therefore, the hypotheses for monotonic stochastic ordering problem can be equivalently written as

$$H_0 : \left\{ \bigcap_{j=1}^{J-1} (Y_{1(j)} \overset{d}{=} Y_{2(j)}) \right\} \quad vs. \quad H_1 : \left\{ \bigcup_{j=1}^{J-1} (Y_{1(j)} \overset{d}{<} Y_{2(j)}) \right\}$$

which are emphasizing a break-down into a set of $J-1$ sub-hypotheses.

For each sub-problem, the analysis of Subsection 3.2 provides for proper tests statistic, and then the global problem can be jointly analyzed by their UI-NPC. Here, to be clearly easy to understand, first take $j = 1$ for example, and then extend to the general case. When $j = 1$, the first pooled pseudo-groups is $\mathbf{Y}_{1(1)} = \mathbf{X}_{\cdot 1}$, and the second one is $\mathbf{Y}_{2(1)} = \mathbf{X}_{\cdot 2} \biguplus \cdots \biguplus \mathbf{X}_{\cdot J}$, which leads to a two-sample one/multi-dimensional design, when j is fixed, the problem becomes to the $2 \times K$ one/multi-dimensional case; and sample sizes of two comparison pseudo-groups will differ as j changing. The specific sub-hypotheses are $H_0 : Y_{1(1)} \overset{d}{=} Y_{2(1)}$ against $H_1 : Y_{1(1)} \overset{d}{<} Y_{2(1)}$. The related combined test statistics can use Anderson-Darling, Mid-ranks and scores tests for one-sided alternatives, we already described. It is worth noting that, sample sizes for j-th pair of pseudo-groups, denoted by $n_{1(j)}$ and $n_{2(j)}$, respectively.

Therefore, for any j, we can consider the Anderson Darling test as follows:

$$T^*_{AD(j)} = C(n_{1(j)}, n_{2(j)}) \cdot \sum_{k=1}^{K-1} \left[\hat{F}^*_{1(j)k} - \hat{F}^*_{2(j)k} \right] [\bar{F}_{\cdot(j)k}(1 - \bar{F}_{\cdot(j)k})]^{-\frac{1}{2}}, \; j = 1, \cdots, J-1,$$

where: $n_{1(j)} = n_1 + \cdots + n_j$, $n_{2(j)} = n - n_{1(j)}$; the permutation relative frequencies are $\hat{F}^*_{l(j)k} = \#(X^*_{l(j)} \preceq v_k)/n_{l(j)} = \sum_{i=1}^{n_l} I(X^*_{l(j),i} \preceq v_k)/n_{l(j)}$, $l = 1, 2$; the marginal relative frequencies are $\bar{F}_{\cdot(j)k} = [\#(X^*_{1(j)} \preceq v_k) + \#(X^*_{2(j)} \preceq v_k)]/n$; partial tests $T^*_{AD(j)}$ are positively dependent; and $C(n_{1(j)}, n_{2(j)})$ are the permutation k-invariable constants. So the global test is solved by combining the $J-1$ partial tests within the UI-NPC as, for

instance by

$$T^*_{AD} = \sum_{j=1}^{J-1} T^*_{AD(j)}.$$

Of course, if $Q > 1$ variables were involved, the multivariate stochastic ordering solution would require one stochastic ordering partial test for each component variable, $q = 1, \cdots, Q$. So, with clear meanings of the symbols the global test, by Direct combination, is:

$$T^*_{AD,Q} = \sum_{j=1}^{J-1} C(n_{1(j)}, n_{2(j)}) \cdot \sum_{q=1}^{Q} \sum_{k=1}^{K-1} \left[\hat{F}^*_{1q(j)k} - \hat{F}^*_{2q(j)k}\right] [\bar{F}_{\cdot q(j)k}(1 - \bar{F}_{\cdot q(j)k})]^{-\frac{1}{2}}.$$

When $J > 2$, such a case can be also settled by Mid-ranks and Scores data transformations. For each j, under the unidimensional setting, the permutation solution based on Mid-ranks for unidimensional data is to proceed by comparing sample means of mid-ranks $T^*_{MR} = \psi\left((\overline{\varphi_{(1)}(X^*_1)} - \overline{\varphi_{(1)}(X^*_2)}), \cdots, (\overline{\varphi_{(J-1)}(X^*_1)} - \overline{\varphi_{(J-1)}(X^*_2)})\right)$. And all these standardized versions of partial tests share the same asymptotic distribution function, the corresponding partial tests based on Mid-ranks are:

$$T^*_{MR(j)} = \left(\overline{\varphi_{(j)}(X^*_1)} - \overline{\varphi_{(j)}(X^*_2)}\right) \cdot Var_{|\mathbf{X}}\left(\overline{\varphi_{(j)}(X^*_1)} - \overline{\varphi_{(j)}(X^*_2)}\right)^{-\frac{1}{2}}$$

The global test is given by combining $J - 1$ partial tests by Direct combination as,

$$T^*_{MR} = \sum_{j=1}^{J-1} T^*_{MR(j)}$$

When $Q > 1$, the problem becomes a multidimensional case, the permutation solution is to combine all partial tests $T^*_{MR(j)Q}$ or all counterpart p-values $\lambda^*_{MR(j)Q}$ with any suitable combining function ϕ, namely, $T^*_{MR,Q} = \phi\left(T^*_{MR(1)Q}, \cdots, T^*_{MR(J-1)Q}\right)$ or $T^*_{MR,Q} = \phi\left(\lambda^*_{MR(1)Q}, \cdots, \lambda^*_{MR(J-1)Q}\right)$. Due to the same limiting distribution of all partial tests $T^*_{MR(j)Q}$, the global test by Direct combination is as follows,

$$T^*_{MR,Q} = \sum_{j=1}^{J-1} \sum_{q=1}^{Q} \left(\overline{\varphi_{q(j)}(X^*_1)} - \overline{\varphi_{q(j)}(X^*_2)}\right) \cdot Var_{|\mathbf{X}}\left(\overline{\varphi_{q(j)}(X^*_1)} - \overline{\varphi_{q(j)}(X^*_2)}\right)^{-\frac{1}{2}}$$

For $J > 2$, with unidimensional setting, the approach based on score function is to assign non-decreasing W scores to ordered categories, e.g. as $v_{k(j)} \to w_{k(j)}$ for $k = 1, \cdots, K$, where $w_{1(j)} \leq \cdots \leq w_{K(j)}$, with at least one strict inequality. The data are transformed

into $w_{ki(j)} = w_{k(i)} \cdot \mathbf{1}(X(i) = v_k)$, for $i = 1, \cdots, n$, where $\mathbf{1}(\cdot)$ is the counting function. It is clearly easy to see that, $w^*_{1(j)} = \sum_{i=1}^{n_{1(j)}} \sum_{k=1}^{K} w^*_{ki}$ presents the sum of the first permuted samples, and $w^*_{2(j)} = \sum_{i=n_{1(j)}+1}^{n} \sum_{k=1}^{K} w^*_{ki}$ presents the sum of the second one. The partial tests based on score transformation, for any j, can be given by

$$T^*_{W(j)} = \left(\bar{w}^*_{1(j)} - \bar{w}^*_{2(j)}\right) Var_{|\mathbf{X}} \left(\bar{w}^*_{1(j)} - \bar{w}^*_{2(j)}\right)^{-\frac{1}{2}}$$

Thus, the global test is given by combining $J-1$ partial tests as follows,

$$T^*_{W} = \sum_{j=1}^{J-1} \left(\bar{w}^*_{1(j)} - \bar{w}^*_{2(j)}\right) Var_{|\mathbf{X}} \left(\bar{w}^*_{1(j)} - \bar{w}^*_{2(j)}\right)^{-\frac{1}{2}}$$

and for $Q > 1$, the unidimensional case extent to multidimensional problem, the permutation solution is to proceed by combining all partial tests $T^*_{W(j)Q}$ or all counterpart p-values $\lambda^*_{W(j)Q}$ with any suitable function ϕ, that is, $T^*_{W,Q} = \phi\left(T^*_{W(1)Q}, \cdots, T^*_{W(J-1)Q}\right)$ or $T^*_{W,Q} = \phi\left(\lambda^*_{W(1)Q}, \cdots, \lambda^*_{W(J-1)Q}\right)$. Due to the same asymptotic distribution function of all partial tests $T^*_{W(j)Q}$, the aforementioned global test by Direct combination is as follows

$$T^*_{W,Q} = \sum_{j=1}^{J-1} \sum_{q=1}^{Q} \left(\bar{w}^*_{1q(j)} - \bar{w}^*_{2q(j)}\right) Var_{|\mathbf{X}} \left(\bar{w}^*_{1q(j)} - \bar{w}^*_{2q(j)}\right)^{-\frac{1}{2}}$$

According to our experience, except for the Direct, the most suitable combining functions for this problem are Fisher's and Liptak's. Since in the stochastic ordering problem under the alternative all $J-1$ partial tests contain a positive non-centrality quantity, i.e. all lie in their respective sub-alternatives, Tippett's combination is less sensitive than others.

3.5 A typical medical example

Let us consider the example in Table (3.4) from Chuang-Stein and Agresti (1997), also reported by Agresti and Coull (1998) and Agresti and Coull (2002), Silvapulle and Sen (2005), Kateri and Agresti (2013) and Colombi and Forcina (2016). It regards a unidimensional survey on subarachnoid hemorrhage measured by the Glasgow outcome scale, where 210 patients received a Placebo, 190 received a Low dose, 207 a Medium dose and 195 a High dose. Response data, related to the extent of trauma, measured on the same ordinal categorical scale, are classified according to $J = 4$ levels of treatment: {*Placebo*, *Low*, *Medium*, *High*}, with clinical outcome classified in $K = 5$ ordered

categories {*Death*, *Vegetative state*, *Major disability*, *Minor disability*, *Good recovery*}.

TABLE 3.4: Dose and Extent of trauma due to subarachnoid hemorrhage

Treatment	Cats.					
	Death	Veget	Major	Minor	Recov	Total
Placebo	59	25	46	48	32	210
Low	48	21	44	47	30	190
Medium	44	14	54	64	31	207
High	43	4	49	58	41	195
Total	194	64	193	217	134	802

Based on our intuition, but also in accordance with quoted authors, patients taking Placebo are expected to achieve lower treatment effect than those taking Low dose, patients taking Low dose are expected to have lower effects than those with Medium dose, and so forth. Therefore, it is expected that patients exhibit monotonically non-decreasing responses as the dose increases. Thus, it is required to test whether there is a monotonic stochastic ordering on distributions related to the 4 treatment levels. Formally, the hypotheses to consider are:

$$H_0 : X_P \overset{d}{=} X_L \overset{d}{=} X_M \overset{d}{=} X_H$$

against

$$H_1 : X_P \overset{d}{\leq} X_L \overset{d}{\leq} X_M \overset{d}{\leq} X_H, \quad \text{with} \quad X_P \overset{d}{<} X_H$$

If responses were quantitative, this problem is also termed of the isotonic regression. Defining the cumulative distribution function for X at ordered categories $v_1 \prec \cdots \prec v_K$ as $F_X(v_k) = \Pr\{X \preceq v_k\}$, namely, the hypotheses are equivalently expressed as

$$H_0 : \{F_{X_P} = F_{X_M} = F_{X_L} = F_{X_H}\}$$

against the alternative

$$H_1 : \{F_{X_P} \geq F_{X_M} \geq F_{X_L} \geq F_{X_H}\} \quad \text{with at least one strict inequality}$$

From the aforementioned subsection 3.4, the analysis of the data from the medical example is termed the J-sample one-dimensional stochastic ordering problem. In the face of such a realistic problem, based on Jonckheere-Terpstra's idea, the problem can equivalently be broken-down into $J-1=3$ sub-problems, each sub-problem is based on a 2×5 contingency table. We intend to analyze this problem within the UI-NPC approach, based on Conditional Monte Carlo with $R=100\,000$ random permutations.

Firstly, we take into account one 2×5 contingency table resulting from collapsing the three levels of treatment, in the other words, it is equivalent to compare whether the one receiving *Standard treatment* stochastically dominates the other receiving *Placebo*. The resulting p-value based on Anderson-Darling equals 0.0141. For the test based on Mid-ranks transformation, giving p-value 0.0144. For the test based on score function which assigns the categories into equidistant scores, the related p-value is 0.0131. It is interesting to note that p-values from three partial tests are comparable, as they only differ in the third digit.

When we collapse the first two rows and the rest two rows, respectively. The resulting reorganized table, is equivalent to the sub-problem when $j=2$. For the Anderson-Darling, Mid-ranks, Score partial tests, the p-values are all at most 0.002, only slightly differ at the fourth digit.

Similarly, collapse the first three rows, the sub-problem implies that compare the highest level of treatment with the rest, to prove if one taking the highest treatment stochastically dominates than the other. The resulting p-values are comparable from the Anderson-Darling and Score partial tests, and the Mid-ranks test behaves slightly better, giving p-value 0.0062.

For the combination functions, by using Fisher's T_F'', Liptak's T_L'', Tippett's T_T'' and Direct T_D'', the corresponding results are shown in the Table (3.5). It is worth mentioned that: h1) W scores are assigned to equidistant ordering integer numbers as ($w_1=1, w_2=2, w_3=3, w_4=4, w_5=5$); h2) since small p-value statistics are evidence for H_1, Fisher's, Liptak's and Tippett's are non-increasing functions of partial p-values. The p-values based on UI-NPC method are:

TABLE 3.5: p-values based on UI-NPC approach

	$T^*_{(1)}$	$T^*_{(2)}$	$T^*_{(3)}$	T_D''	T_F''	T_L''	T_T''
$\hat{\lambda}_{AD(j)}$	0.0141	0.0025	0.0074	0.0017	0.0015	0.0012	0.0068
$\hat{\lambda}_{W(j)}$	0.0131	0.0021	0.0076	0.0010	0.0012	0.0010	0.0053
$\hat{\lambda}_{MR(j)}$	0.0144	0.0024	0.0062	0.0011	0.0014	0.0011	0.0068

Results in Table (3.5) clearly show that the p-values based on four different combination functions T_D'', T_F'', T_L'' and T_T'', all reject the null hypothesis at significance level $\alpha = 0.01$ of monotonic stochastic ordering among the $J = 4$ levels of treatment. So the inferential conclusion is that patients present non-decreasing responses as the dose increases.

It is worth noting that the three combined p-value statistics T_D'', T_F'' and T_L'' differ only slightly in the fourth digit. This means that related tests are all suitable for testing unidimensional dominance and stochastic ordering alternatives, and so the resulting combined p-values are likely close to the best one. In our case, if the stochastic ordering alternative is true it is jointly true by construction for all $J - 1$ partial tests $T^*_{(j)}$. So, Tippett's T_T'' differs from other combination functions because its power behavior is mostly sensitive when only one partial test lies in the alternative. Due to too many ties in the data set, the test with rank transformations was not considered.

Since all p-values statistics related to $T_{AD(3)}$ are $< 0.05/3$, by simple Bonferroni's rule it results that subjects taking High dose exhibit significantly lower responses than those taking lower doses.

Chapter 4

Testing for restricted alternatives with repeated data

4.1 Prologue

Given that two groups with unidimensional data, the related permutation solutions for stochastic dominance testing problems have been proposed in the former Chapter. When these two groups are observed repeatedly at several recording times, we are wondering if there still exists one subject that stochastically dominates the other. Such repeated measure design is a common occurrence in experimental situations, in which subject is observed at a sequence of recording time points. The responses of the unit may be expected to comply with an unknown tendency, or be viewed as the discretized stochastic process. To be clearly understood, we also take the clinical experiment, for example, the corresponding table is shown as follows,

TABLE 4.1: General two subjects with repeated data

Treatment	A						B						
	Time						Time						
ID	1	2	$\cdots$	$\cdots$	$\cdots$	T	**ID**	1	2	$\cdots$	$\cdots$	$\cdots$	T
1	$X_{1,1,1}$	$X_{1,1,2}$	$\cdots$	$\cdots$	$\cdots$	$X_{1,1,T}$	1	$X_{2,1,1}$	$X_{2,1,2}$	$\cdots$	$\cdots$	$\cdots$	$X_{2,1,T}$
2	$X_{1,2,1}$	$X_{1,2,2}$	$\cdots$	$\cdots$	$\cdots$	$X_{1,2,T}$	2	$X_{2,2,1}$	$X_{2,2,2}$	$\cdots$	$\cdots$	$\cdots$	$X_{2,2,T}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
i	$X_{1,i,1}$	$X_{1,i,2}$	$\cdots$	$\cdots$	$\cdots$	$X_{1,i,T}$	i	$X_{2,i,1}$	$X_{2,i,2}$	$\cdots$	$\cdots$	$\cdots$	$X_{2,i,T}$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
n_1	$X_{1,n_1,1}$	$X_{1,n_1,2}$	$\cdots$	$\cdots$	$\cdots$	$X_{1,n_1,T}$	n_2	$X_{2,n_2,1}$	$X_{2,n_2,2}$	$\cdots$	$\cdots$	$\cdots$	$X_{2,n_2,T}$

Suppose study subjects are randomly assigned to two levels of treatment schemes, the subjects will be asked to rate their post-treatment responses at several recording time points, or to score the symptoms before being given medication at several recording time points, or the researchers proceed with measurements of subjects' post-treatment responses at fixed time points. Some noteworthy aspects should be emphasized here:

i1) Physical fitness of subjects we consider may be of equally likely nature, or in general, the physical fitness may be seen as distinct due to different factors. In what follows, the contents will be sufficient to distinguish whether study subjects are of equally likely nature or not.

i2) Study subjects are independently observed of each other regardless of which treatment regimens they receive.

i3) Outcomes of distinct subjects either at the same time point or different time points are independent.

i4) The outcomes of the subject's pre-treatment or post-treatment are defined on the same ordered categorical scale. In other words, subjects need to rate categories of their pre-treatment or post-treatment effects according to an ordered categorical scale, such as: {*No pain; Mild Pain; Moderate Pain; Severe Pain*}; {*Unhappy; Neither happy nor unhappy; Happy*}; and so forth.

i5) The measurement contents are the same, since subjects are asked to answer the same question at different time occasions.

i6) The sample sizes are usually moderate, and the existence of unbalanced comparison groups is allowed.

i7) Both treatment regimens would yield non-negative treatment effects δ, and researchers expect one receiving *Standard treatment* would yield better effect than those receiving *Placebo*. Assume the effects of Placebo are not active nor non-negative, and so are set at zero.

According to these basic settings, let's turn our attention to the Table (4.1). There are two levels of treatment regimens in the Table (4.1), one is *Standard treatment* denoted by A, the other one is *Placebo* denoted by B. Both of them repeatedly measure T recording time points. Sample sizes of A and B are n_1 and n_2, respectively; in addition, sample sizes could be unbalanced: $n_1 = (\neq) n_2$. The layout of unit is $X_{j,i,t}$, for $j = 1, 2$, $i = 1, \cdots, n_j$, $t = 1, \cdots, T$, it implies that unit $X_{j,i,t}$ belongs to the j-th treatment

scheme, i-th unit of j-th treatment scheme, t-th measurement. To be simplified, assumed $j = 1$ stands for *Standard treatment*, and $j = 2$ for *Placebo*.

In such a case, we have some problems of interest to study:

j1) In the group being given standard treatment, whether the treatment effects will tend to decrease as the number of measurements increases, in other words, whether the treatment effects will tend to decrease over time.

j2) In the group being given placebo, whether the post-treatment effects also tend to decrease over time.

j3) Whether one taking placebo stochastically dominates the other receiving standard treatment, i.e. whether he/she presents stochastically lower outcome.

4.2 Repeated measurements with two occasions

To be more clearly understood, we will start from one sample in which subjects' post-treatment responses are measured at two recording time points. There are n study subjects taking the clinical experiments, and post-treatment responses of subjects are measured at time points $t = 1, 2$. Subjects are independent of each other, the measurements at two time points are not independent. Suppose samples are set to be $\mathbf{X}_{\cdot t} = \{X_{1,t}, X_{2,t}, \cdots, X_{n,t}\} \in \mathcal{X}^n$ for $t = 1, 2$.

The associated table is shown as follows,

TABLE 4.2: Two recording time points

ID \ Time	1	2
1	$X_{1,1}$	$X_{1,2}$
2	$X_{2,1}$	$X_{2,2}$
3	$X_{3,1}$	$X_{3,2}$
...	...	...
i	$X_{i,1}$	$X_{i,2}$
...	...	...
j	$X_{j,1}$	$X_{j,2}$
...	...	...
n	$X_{n,1}$	$X_{n,2}$

The researchers intend to determine whether the first-time treatment effect stochastically dominates the second-time treatment effect. For this case, we can consider the question in two ways: one is that, specific to the responses of each unit, whether there exists a dominating relation between two measurements; second is that, assuming each column as a whole to be considered, whether one stochastically dominates the other. These two ways can be considered equivalent, thus and so, the two related hypotheses can be shown as follows,

I) Specific to each unit: testing the responses pertain to the i-th unit,

$$H_{0i} : \{X_{i,1} \overset{d}{=} X_{i,2} \overset{d}{=} X_{i\cdot}\} \quad vs. \quad H_{1i} : \{X_{i,1} \overset{d}{>} X_{i,2} \overset{d}{=} X_{i\cdot}\}.$$

Since the related global hypotheses are shown as

$$H_0 : \left\{ \bigcap_i \{X_{i,1} \overset{d}{=} X_{i,2}\} \right\} \equiv \{X_{\cdot 1} \overset{d}{=} X_{\cdot 2}\} \; vs. \; H_1 : \left\{ \bigcup_i \{X_{i,1} \overset{d}{>} X_{i,2}\} \right\} \equiv \{X_{\cdot 1} \overset{d}{>} X_{\cdot 2}\}.$$

II) Specific to each column: testing the responses at two recording time points,

$$H_0 : \{X_{\cdot 1} \overset{d}{=} X_{\cdot 2}\}, \quad vs. \quad H_1 : \{X_{\cdot 1} \overset{d}{>} X_{\cdot 2}\}.$$

It is worth noting the strong similarity of I) to II). In I) data paired are assumed to be dependent within each pair and pairs are independent; in II) it is apparently assumed the traditional two-sample setting.

4.2.1 Some typical regression functions

To analyze such problem, suppose outcome $X_{i,t}$ can be expressed as a regression function,

$$X_{i,1} = \mu + \eta_i + \sigma_1 \cdot Z_{i,1}, \quad X_{i,2} = \mu + \eta_i - \delta + \sigma_2 \cdot Z_{i,2}, \; i = 1, \cdots, n.$$

There are some noteworthy aspects to point out: k1) μ be the population constant; k2) η_i be unknown components specific to the i-th unit, which are not dependent on treatment schemes; k3) δ be a non-negative and be defined as a treatment effect for subjects under treatment; k4) σ_1 and σ_2 are scale coefficients pertaining to the counterpart groups; k5) Assume $Z_{i,1}$ and $Z_{i,2}$ be identically distributed error components with zero median: $\Pr\{Z_{i,t} < 0\} = \Pr\{Z_{i,t} > 0\}$, for every i and t. Actually, the error components are assumed to be at least exchangeable within units but are independent between units.

It is worth highlighting that solutions to the regression function will differ depending on the addition or removal of some components.

l1) Assume $\eta_i = 0$, $\sigma_1 = \sigma_2 = 1$, all other components assume the same meaning as aforementioned content. The regression functions are presented as

$$X_{i,1} = \mu + Z_{i,1}, \quad X_{i,2} = \mu - \delta + Z_{i,2},\ i = 1, \cdots, n.$$

Considering the differences between measurement of i-th unit, that is, $\{\Delta_i = X_{i,1} - X_{i,2} = \delta + Z_{i,1} - Z_{i,2}, i = 1, \cdots, n\}$ which are equivalently represented as $\{\Delta_i - \delta = Z_{i,1} - Z_{i,2}, i = 1, \cdots, n\}$.

l2) Assume $\sigma_1 = \sigma_2 = 1$, all other components have the same meaning as the foregoing. The regression functions are expressed as

$$X_{i,1} = \mu + \eta_i + Z_{i,1}, \quad X_{i,2} = \mu + \eta_i - \delta + Z_{i,2},\ i = 1, \cdots, n.$$

The differences between measurement of i-th unit are $\{\Delta_i = X_{i,1} - X_{i,2} = \delta + Z_{i,1} - Z_{i,2}, i = 1, \cdots, n\} \equiv \{\Delta_i - \delta = Z_{i,1} - Z_{i,2}, i = 1, \cdots, n\}$.

l3) Assume $\sigma_1 = \sigma_2$, all other components share the same meaning as described above. The regression functions are in the form of

$$X_{i,1} = \mu + \eta_i + \sigma \cdot Z_{i,1}, \quad X_{i,2} = \mu + \eta_i - \delta + \sigma \cdot Z_{i,2},\ i = 1, \cdots, n,$$

where σ is the unknown standard deviation assumed to be independent of unit and treatment levels, and satisfying the condition $0 < \sigma < \infty$. Further, the differences between measurements of i-th unit are $\{\Delta_i = X_{i,1} - X_{i,2} = \delta + \sigma \cdot (Z_{i,1} - Z_{i,2}), i = 1, \cdots, n\} \equiv \{\Delta_i - \delta = \sigma \cdot (Z_{i,1} - Z_{i,2}), i = 1, \cdots, n\}$.

l4)

$$X_{i,1} = \mu + \eta_i + \sigma_1 \cdot Z_{i,1}, \quad X_{i,2} = \mu + \eta_i - \delta + \sigma_2 \cdot Z_{i,2},\ i = 1, \cdots, n,$$

where two scale coefficients σ_1 and σ_2 may not be guaranteed to be identical. When $\sigma_1 = \sigma_2$, the model becomes the third case. Moreover, in the case with $\sigma_1 \neq \sigma_2$, if σ_1 and σ_2 are only scale coefficients but not dependent on the treatment effects, the testing problem falls into logical fallacy because $X_{i,1}$ and $X_{i,2}$ are not equal in distribution under H_0. Therefore, when σ is the function of treatment effects, with $\sigma(0) = \sigma_1$ and $\sigma(\delta) = \sigma_2$ $(0 < \sigma < \infty)$, $X_{i,1}$ and $X_{i,2}$ are equal in distribution

under H_0, we only consider this situation. The associated differences between measurements of i-th unit are $\{\Delta_i = X_{i,1} - X_{i,2} = \delta + (\sigma_1 Z_{i,1} - \sigma_2 Z_{i,2}), i = 1, \cdots, n\} \equiv \{\Delta_i - \delta = (\sigma_1 Z_{i,1} - \sigma_2 Z_{i,2}), i = 1, \cdots, n\}$.

It is clearly easy to see that, there are three types of the differences of responses pertaining to i-th unit, namely, $\{Z_{i,1} - Z_{i,2}\}$; $\{\sigma \cdot (Z_{i,1} - Z_{i,2})\}$; and $\{\sigma_1 Z_{i,1} - \sigma_2 Z_{i,2}\}$. Thus, instead of studying the left side of equations $\Delta_i - \delta$, we turn our attention to work on the right side of equations. It is already known that, $Z_{i,1}$ and $Z_{i,2}$ are at least exchangeable and identically distributed, and independent with respect to units. Suppose the underlying likelihood related to $Z_{i,j}$ is f_Z, then it is straightforward to obtain $f_Z(Z_{i,1})f_Z(Z_{i,2}) = f_Z(Z_{i,2})f_Z(Z_{i,1}) = \prod_j f_Z(Z_{i,j})$, which characterizes exchangeability of $Z_{i,1}$ and $Z_{i,2}$ within units. Now, let us analyze those three cases individually:

m1) For the case of $\{Z_{i,1} - Z_{i,2}\}$, within transformation function $\{\zeta(x, y) = 1, y > x; 0, y \leq x\}$, we can obtain

$$\begin{aligned}(Z_{i1}, Z_{i2}) \overset{d}{=} (Z_{i2}, Z_{i1}) \overset{\zeta}{\to} \zeta(Z_{i1}, Z_{i2}) &\overset{d}{=} \zeta(Z_{i2}, Z_{i1}) \\ &\to E[\zeta(Z_{i1}, Z_{i2})] = E[\zeta(Z_{i2}, Z_{i1})] \\ &\to \Pr(Z_{i2} > Z_{i1}) = \Pr(Z_{i1} > Z_{i2}) \\ &\to \Pr(Z_{i2} - Z_{i1} > 0) = \Pr(Z_{i1} - Z_{i2} > 0) \\ &\to Z_{i2} - Z_{i1} \overset{d}{=} Z_{i1} - Z_{i2}.\end{aligned}$$

Therefore, $Z_{i2} - Z_{i1}$ and $Z_{i1} - Z_{i2}$ are equal in distribution, and it also implies that $Z_{i2} - Z_{i1}$ is symmetric around zero.

m2) For the case of $\{\sigma \cdot (Z_{i,1} - Z_{i,2})\}$, according to the symmetry of $Z_{i2} - Z_{i1}$, it is straightforward to obtain that

$$\begin{aligned}Z_{i2} - Z_{i1} \overset{d}{=} Z_{i1} - Z_{i2} \to E[\sigma \cdot (Z_{i1}, Z_{i2})] &= E[\sigma \cdot (Z_{i2}, Z_{i1})] \\ \to \sigma \cdot (Z_{i2} - Z_{i1}) &\overset{d}{=} \sigma \cdot (Z_{i1} - Z_{i2}).\end{aligned}$$

Since, $\sigma \cdot (Z_{i2} - Z_{i1})$ and $\sigma \cdot (Z_{i1} - Z_{i2})$ are equal in distribution, they are also symmetric around zero.

m3) For the case of $\{\sigma_1 Z_{i,1} - \sigma_2 Z_{i,2}\}$, due to the exchangeability between $Z_{i,1}$ and $Z_{i,2}$, $\sigma_1 \cdot Z_{i,1}$ and $\sigma_2 \cdot Z_{i,2}$ are not exchangeable unless the error components $Z_{i,1}$ and $Z_{i,2}$ are symmetrically distributed around zero. Thus, only within the transformation

function $\{\zeta(x,y) = 1, y > x; 0, y \leq x\}$, we have

$$\begin{aligned}(\sigma_1 Z_{i,1}, \sigma_2 Z_{i,2}) \stackrel{d}{=} (\sigma_2 Z_{i,2}, \sigma_1 Z_{i,1}) &\xrightarrow{\zeta} \zeta(\sigma_1 Z_{i,1}, \sigma_2 Z_{i,2}) \stackrel{d}{=} \zeta(\sigma_2 Z_{i,2}, \sigma_1 Z_{i,1}) \\ &\to E\{\zeta(\sigma_1 Z_{i,1}, \sigma_2 Z_{i,2})\} \stackrel{d}{=} E\{\zeta(\sigma_2 Z_{i,2}, \sigma_1 Z_{i,1})\} \\ &\to \Pr\{\sigma_1 Z_{i,1} > \sigma_2 Z_{i,2}\} = \Pr\{\sigma_2 Z_{i,2} > \sigma_1 Z_{i,1}\}\end{aligned}$$

Even though the underlying distributions of Z_{i1} and Z_{i2}, $i = 1, \cdots, n$, are unknown, the conclusion about symmetry of $\sigma_1 Z_{i,1} - \sigma_2 Z_{i,2}$ can be obtained.

All three types of the differences of responses are symmetric, equivalent to say that $\Delta_i - \delta$ is also symmetric around zero, and further, Δ_i is symmetric around δ. It is well known that, in H_0, it is expected that there are no differences in treatment effects between two levels of treatment schemes, thus, δ takes the value of zero, which implies that Δ_i is symmetric around zero. In H_1, the expected effect under the standard treatment is not worse than that under placebo, or we say that treatment effect receiving standard treatment is better than that being given placebo. Thus, δ is active, namely, $\delta > 0$, and it is straightforward to say Δ_i is symmetric around a location or a so-called treatment effect δ.

4.2.2 A general regression function

In general, related assumptions on components of regression function could be appropriately relaxed, regression function can be expressed as

$$X_{i,1} = \mu + \eta_i + \sigma \cdot Z_{i,1}, \quad X_{i,2} = \mu + \eta_i - \delta_i + \sigma(\delta_i) \cdot Z_{i,2},\ i = 1, \cdots, n,$$

where: n1) σ is a function of treatment effect δ_i. Suppose $\sigma(0) = \sigma$, for $\delta_i > 0$, we have $\sigma(\delta_i) \neq \sigma$, which implies that the scale coefficients are not constant σ in the alternative under conditioning $0 < \sigma < \infty$. n2) Error components $Z_{i,1}$ and $Z_{i,1}$ which are generally non-Gaussian, are assumed to behave as a stationary stochastic process with null mean and median and unknown distribution P_Z, where these error terms are exchangeable within units, but independent with respect to units. n3) δ_i, $i = 1, \cdots, n$, are treatment effects pertaining to the i-th unit. Let $\boldsymbol{\delta}$ be the vector of δ_i, if $\boldsymbol{\delta} = 0$, it implies that there are no differences between two levels of treatment schemes; if $\boldsymbol{\delta} > 0$, it means that there exist at least one δ_i active. The corresponding differences between measurements of i-th unit are $\{\Delta_i - \delta_i = \sigma Z_{i,1} - \sigma(\delta_i) Z_{i,2}, i = 1, \cdots, n\}$. For any $i = 1, \cdots, n$,

o1) If $\delta_i = 0$, the right side of equations become $\sigma(Z_{i,1} - Z_{i,2})$, so it pertains to the case of m2).

o2) If $\delta_i > 0$, the right side of equations become $\sigma Z_{i,1} - \sigma(\delta_i) Z_{i,2}$, it belongs to the case of m3).

Whether $\delta_i = 0$ or $\delta_i > 0$ holds, Δ_i is symmetric around δ_i. In H_0, $\boldsymbol{\delta}$ should take the value of the zero vector, there are no differences between two treatment schemes. Furthermore, in H_1, $\boldsymbol{\delta}$ is not a negative vector nor a zero vector, in which at least one δ_i non-negative, since, $\boldsymbol{\delta}$ can be expressed as a union of $\delta_i \geq 0$ with at least one strict inequality. It means that, in H_1, that may exist some $\{\delta_i = 0\}$ and the rest $\{\delta_i > 0\}$. Thus and so, the hypotheses can be expressed as

$$H_0 : \boldsymbol{\delta} = \mathbf{0} \equiv \bigcap_i \{\delta_i = 0\} = \bigcap_i \{H_{0i}\},$$

against

$$\begin{aligned} H_1 : \boldsymbol{\delta} > \mathbf{0} &\equiv \bigcup_i \{\delta_i \geq 0\}, \text{ with at least one strict inequality} \\ &\equiv \bigcup_i \{\delta_i > 0\} = \bigcup_i \{H_{1i}\}. \end{aligned}$$

The global testing problem is broken down into a set of sub-problems, that allows us to handle it by studying each paired hypotheses $H_{0i} : \{\delta_i = 0\}$ and $H_{1i} : \{\delta_i > 0\}$ for $i = 1, \cdots, n$.

4.2.3 Related solutions

To study the properties of $\Delta_i - \delta_i$, it would be inevitable to analyze the types of data $X_{i,t}$. If the existence of continuity assumption on $X_{i,t}$ holds, there are some suggested solutions for such problems:

p1) Assume error terms follow a specific underlying distribution and the identical treatment effect δ. For the simplest case like l1), $Z_{i,t}$ are independent with respect to units, and $Z_{i,t} \sim \mathcal{N}(0,1)$, $t = 1, 2$, $i = 1, \cdots, n$. In such a case, the Student's t-test is suitable for the differences Δ_i, that is,

$$T = \bar{\Delta} \cdot \frac{\sqrt{n}}{\hat{\Sigma}},$$

where $\bar{\Delta} = \sum_i \Delta_i / n$, $\hat{\Sigma}^2 = \sum_i (\Delta_i - \bar{\Delta})^2 / (n-1)$. Due to the normality assumption on error terms, differences Δ_i are also normally distributed: $\Delta_i \sim \mathcal{N}(\delta, \Sigma_\Delta^2)$. The sampling distribution of test T is central Student's t with $n-1$ degrees of freedom

in H_0; under H_1 the test T is distributed as a non-central Student's t with a positive non-central parameter, and so it is significant for large values.

p2) Assume the underlying distribution of paired $(X_{i,1}, X_{i,2})$ are IID from a bivariate normal distribution, and so the differences of paired data are also normally distributed. Of course, the inverse is not true, because the differences of paired data are normally distributed does not guarantee for normality of $(X_{i,1}, X_{i,2})$.

p3) If the underlying distribution of Δ_i is unknown, all study subjects are of equally likely nature, i.e. pairs $(X_{i,1}, X_{i,2})$ are IID, and so all Δ_i can be assumed to be independently identically distributed. In such a case, it is suitable to use a nonparametric approach based on the rank function. Suppose ties are assumed to occur with probability zero, Wilcoxon signed-rank test could be reasonable for this case,

$$W = \sum_i w_i \cdot R_i,$$

where $R_i = \sum_{1 \le h \le n} I(|\Delta_h| \le |\Delta_i|)$, and weights $w_i = 1$ if $\Delta_i > 0$ and $w_i = 0$ if $\Delta_i < 0$. It is worth mentioning that Wilcoxon signed-rank test works within the conditionality principle of inference.

p4) Δ_i is symmetric around δ_i, and with the unknown underlying distribution. Due to the continuity of $X_{i,t}$, Δ_i are also continuous.

It is well known that, Δ_i is symmetric around 0 in H_0, and then assuming one variable Y_i be an indicator function of Δ_i as follows,

$$Y_i = \begin{cases} 1 & \text{if } \Delta_i > 0 \\ 0 & \text{if } \Delta_i \le 0 \end{cases}$$

According to continuity of Δ_i, we have $\Pr(Y_i = 1) = \Pr(Y_i = 0) = 1/2$, since Y_i are independent and identically distributed Bernoulli random variables with successful probability $1/2$, $Y_i \sim Ber(1, 1/2)$, $i = 1, \cdots, n$. Denote the sum of Y_i by U, namely $U = \sum_i^n Y_i$, it is immediately clear to see that U is Binomially distributed, that is, $U \sim Bi(n, 1/2)$ under H_0. So U corresponds to a version of the well-known McNemar test.

If all δ_i are identical with value δ, and $\delta > 0$ in H_1, then U is also binomially distributed with parameters n and $p = \Pr(Y_i = 1) > 1/2$, respectively, and

namely, $U \sim Bi(n, p)$. Thus and so, the large values of U are significant to reject the null hypothesis.

If we cannot determine the values of δ_i in H_1, and Δ_i is independent of each other, U can be considered as a sum of independent Bernoulli random variables that are not necessarily identically distributed, therefore, U is termed Poisson binomial distribution. Then the probability mass function having u successful trials out of a total of n trails can be rewritten as

$$\Pr(U = u) = \sum_{A \in A_u} \prod_{i \in A} p_i \prod_{j \in A^c} (1 - p_j),$$

where $p_i = \Pr(\Delta_i > 0) \geq 1/2$, $i = 1, \cdots, n$; and A_u is a set of all subsets containing u integers which means the existence of u successful trials, A^c is the complement of A. The complete enumerations of A_u is $n!/\{u!(n-u)!\}$, however, it is impractical to compute the sum when the sample size is large. The mean value and variance of U are $E(U) = \sum_{i=1}^{n} p_i$ and $Var(U) = \sum_{i=1}^{n} p_i(1-p_i)$, respectively. It is immediately easy to know that large value of U gives us a significant evidence to reject H_0.

If no assumption of continuity of $X_{i,t}$ exist, Viz. $X_{i,t}$ are taken from a discrete distribution, then the differences Δ_i are also discrete. Under the assumptions of symmetry, both in H_0 and H_1, Δ_i is symmetric around δ_i, for $i = 1, \cdots, n$.

q1) Let p_{+i}, p_{-i} and p_{0i} be the probabilities $\Pr(\Delta_i > 0)$, $\Pr(\Delta_i < 0)$ and $\Pr(\Delta_i = 0)$, respectively. Assume one variable Y_i presents the indicator function of Δ_i as follows,

$$Y_i = \begin{cases} 1 & \text{if } \Delta_i > 0 \\ 0 & \text{if } \Delta_i \leq 0 \end{cases}$$

Clearly, $\Pr(Y_i = 1) = p_{+i}$ and $\Pr(Y_i = 0) = p_{-i} + p_{0i}$, the sum U of successful trails is also termed Poisson binomial distribution. Then the probability mass function having u successful trails out of a total of n trails can be also expressed as follows,

$$\Pr(U = u) = \sum_{A \in A_u} \prod_{i \in A} p_{+i} \prod_{j \in A^c} (1 - p_{+j}),$$

where $p_{+i} = \Pr(\Delta_i > 0)$, $i = 1, \cdots, n$; and A_u is a set of all subsets containing u integers which means the existence of u successful trials, A^c is the complement of

A. And the mean value and variance of U are $E(U) = \sum_{i=1}^{n} p_{+i}$ and $Var(U) = \sum_{i=1}^{n} p_{+i}(1 - p_{+i})$, respectively.

Accordingly, due to the existence of p_{0i}, we have $p_{+i} = p_{-i} \leq 1/2$ in H_0; the corresponding mean value is $E_{H_0}(U) = \sum_{i=1}^{n} p_{+i} \leq n/2$. In H_1, $p_{+i} > \{p_{-i} + p_{0i}\}$ holds because of $\delta_i \geq 0$, which is slipped to the right compared with zero; the mean value of U in H_1, is $E_{H_1}(U) = \sum_{i=1}^{n} p_{+i} \geq n/2$. Comparing with the underlying population of U in H_0, the population of U in H_1 is slipping to the right, thus, large values are significant for rejecting H_0.

Under the H_0, $X_{i,1}$ and $X_{i,2}$ are equal in distribution, and are exchangeable within unit. Thus, the two observed data of i-th unit are considered as if they were randomly assigned to two time points, it seemingly randomly assign a sign "+" or "−" to the difference Δ_i, and each with probability 1/2. The related permutation test statistic is

$$U^* = \sum_{i}^{n} Y_i^* = \sum_{i}^{n} I(\Delta_i^* > 0) = \sum_{i}^{n} I(\Delta_i \cdot S_i^* > 0),$$

where S_i^* are IID random variables and each takes the value of $+1$ or -1 with probability 1/2, and are independent of Δ_i. In H_0, the expectation and variance of U^* are shown as follows, respectively

$$E(U^*) = E(\sum_{i}^{n} I(\Delta_i \cdot S_i^* > 0)) = \sum_{i}^{n} \Pr(\Delta_i \cdot S_i^* > 0) = \sum_{i}^{n} [\frac{1}{2} p_{+i} + \frac{1}{2} p_{-i}] = \sum_{i}^{n} p_{+i},$$

$$Var(U^*) = Var(\sum_{i}^{n} I(\Delta_i \cdot S_i^* > 0)) = \sum_{i=1}^{n} p_{+i}(1 - p_{+i}).$$

Therefore, the standardized version of related permutation test can be obtained immediately, $T_U^* = (U^* - E(U^*))/\sqrt{Var(U^*)}$, which for large sample size is normally distributed, namely, $T_U^* \xrightarrow{d} \mathcal{N}(0, 1)$.

q2) In H_0, according to the exchangeability between $X_{i,1}$ and $X_{i,2}$ within unit, it seems as if the two observations pertaining to i-th unit are randomly assigned to two time points, which is equivalent to assign the sign "+" or "−" to Δ_i and each with probability 1/2. One solution for such a case is to take into consideration a test of the form $T = \sum_i^n \Delta_i$. The permutation solution is based on the form $T^* = \sum_i^n \Delta_i^* = \sum_i^n \Delta_i \cdot S_i^*$, where S_i^* are IID random variables and each takes the

value of -1 or $+1$ with probability $1/2$ each, independent of Δ_i. As we known,

$$\sum_i^n \Delta_i(0) \cdot S_i^* \quad \textit{(in the null hypotheses)} \overset{d}{<} \sum_i^n \Delta_i(\delta_i) \cdot S_i^* \quad \textit{(in the alternatives)},$$

where $\delta_i > 0$, therefore, large values are significant to reject the null hypothesis. The mean value and the variance of T^* are shown as follows,

$$\begin{aligned} E(\sum_i \Delta_i^*/n) &= \sum_i E(\Delta_i \cdot S_i^*)/n = \sum_i (\Delta_i \cdot 1/2 - \Delta_i \cdot 1/2)/n = 0, \\ Var(\sum_i \Delta_i^*/n) &= Var(\sum_i \Delta_i \cdot S_i^*/n) = \sum_i \Delta_i^2/n^2. \end{aligned}$$

Therefore, the permutation standardized version:

$$T_S^* = \left(\sum_i \Delta_i \cdot S_i^*\right) / \left(\sum_i \Delta_i^2\right)^{1/2}.$$

4.3 Repeated measurements with T occasions

When the recording time points T is larger than 2, such a case, termed the repeated measure designs, is shown in the Table (4.1). As described in the Prologue, each subject is independently and randomly assigned to one of the two levels of treatment (A or B), and the measurements at different recording time points are assumed to follow a suitable stochastic process. Denote $j = 1$ for the standard treatment and $j = 2$ for placebo.

A general regression function can be expressed as follows,

$$X_{j,i,t} = \mu + \eta_{j,i} + \delta_{j,i,t} + \sigma(\delta_{j,i,t}) \cdot Z_{j,i,t},\ j = 1, 2;\ i = 1, \cdots, n_j;\ t = 1, \cdots, T,$$

where: r1) μ be the population constant; r2) $\eta_{j,i}$ be unknown components specific to the i-th unit of j-th scheme, which are not dependent on treatment schemes; r3) $\delta_{j,i,t}$ be the treatment effect specific to the i-th unit of j-th scheme at t time point; r4) σ be a function of $\delta_{j,i,t}$; suppose $\sigma(0) = \sigma$, for $\delta_{j,i,t} > 0$ we have $\sigma(\delta_{j,i,t}) \neq \sigma$; r5) $Z_{j,i,t}$ are assumed to generally non-Gaussian error terms distributed as a stationary stochastic process with null mean and unknown distribution P_Z, these error terms are assumed to be exchangeable within units but independent with respect to units.

The testing problem of interest is to determine whether the treatment effect will tend to decrease over time. For such a case, we start from the analysis of preliminary stratification testing, and then analyze the whole entity.

In short, the related hypotheses for j-th scheme can be presented as follows

$$H_{0[j]}:\left\{\bigcap_i\{\delta_{j,i,1}=\delta_{j,i,2}=\cdots=\delta_{j,i,T}=0\}\right\}=\left\{\bigcap_i H_{0i[j]}\right\},$$

against

$$H_{1[j]}:\left\{\bigcup_i\{\delta_{j,i,1}>\delta_{j,i,2}>\cdots>\delta_{j,i,T}\}\right\}=\left\{\bigcup_i H_{1i[j]}\right\}.$$

Thus, in $H_{0[j]}$, the treatment effects are identical over time, which leads to that $X_{j,i,t}$ are stationary distributed over time. Specific to the j-th treatment scheme, for any i, suppose the underlying distribution of $X_{j,i}$ is $f_{j,i}$, the joint null likelihood $\mathbf{X}_{\cdot j}$ takes the form of

$$f(\mathbf{X}_{\cdot j})=\prod_{i=1}^{n_j}\left\{\prod_{t=1}^{T}f_{j,i}(X_{j,i,t})\right\},$$

which, due to assumed exchangeability, is invariable with respect to any permutation $\mathbf{X}^*_{\cdot j}$,

$$f(\mathbf{X}^*_{\cdot j})=\prod_{i=1}^{n_j}\left\{\prod_{t=1}^{T}f_{j,i}(X^*_{j,i,t})\right\}=f(\mathbf{X}_{\cdot j}).$$

Thus, the related conditional reference space is $\mathcal{X}_{/\mathbf{X}_{\cdot j}}$. Further, considering the whole entity $\mathbf{X}$, which also plays the role of related conditional reference space, because of $f(\mathbf{X})=\prod_{j=1}^{2}f(\mathbf{X}_{\cdot j})=\prod_{j=1}^{2}f(\mathbf{X}^*_{\cdot j})=f(\mathbf{X}^*)$, denote it by $\mathcal{X}_{/\mathbf{X}}$. It is worth emphasizing that the exchangeability only works within units, which means that each unit also plays the role of sufficient space, the joint likelihood of $\mathbf{X}_{j,i}=(X_{j,i,1},X_{j,i,2},\cdots,X_{j,i,T})$ is in the form of

$$f(\mathbf{X}_{j,i})=\prod_{t=1}^{T}f_{j,i}(X_{j,i,t})=\prod_{t=1}^{T}f_{j,i}(X^*_{j,i,t})=f(\mathbf{X}^*_{j,i})$$

Which proves that the sufficient space of $\mathbf{X}_{j,i}$ is $\mathcal{X}_{/\mathbf{X}_{j,i}}$, in which the data are exchangeable. Similarly, it can also be explained by the meaning of the orbit, in which the likelihoods of all points are assumed to be the same. In our case, under H_0, to each study subject, it is defined a sub-orbit, where these components are exchangeable within orbit. In short, each unit $\mathbf{X}_{j,i}$ can be considered as a sub-orbit, in which the data are exchangeable. Further, the exchangeability of data does not hold between subjects,

specifically, data $X_{1,1,2}$ of subject $X_{1,1}$ and data $X_{1,4,2}$ of subject $X_{1,4}$ cannot be exchanged. It also implies that sub-orbits are independent of each other, the intersection of any two sub-orbits yields an empty set.

Accordingly, the global conditional reference space $\mathcal{X}_{/\mathbf{X}}$ can be expressed with the product of sub-sufficient space of $\mathbf{X}_{j,i}$, namely, $\mathcal{X}_{/\mathbf{X}_{\cdot j}} = \mathcal{X}_{/\mathbf{X}_{j,1}} \times \cdots \times \mathcal{X}_{/\mathbf{X}_{j,n_j}}$. The data are only permuted and exchangeable within each conditional reference space, which yield the permutation sufficient space $\mathcal{X}_{/\mathbf{X}^*_{\cdot j}} = \mathcal{X}_{/\mathbf{X}^*_{j,1}} \times \cdots \times \mathcal{X}_{/\mathbf{X}^*_{j,n_j}}$.

The aforementioned hypotheses are equivalent to

$$H_{0[j]} : \left\{ \bigcap_i \{X_{j,i,1} \overset{d}{=} X_{j,i,2} \overset{d}{=} \cdots \overset{d}{=} X_{j,i,T} \overset{d}{=} X_{j,i\cdot}\} \right\},$$

against

$$H_{1[j]} : \left\{ \bigcup_i \{X_{j,i,1} \overset{d}{>} X_{j,i,2} \overset{d}{>} \cdots \overset{d}{>} X_{j,i,T}\} \right\}.$$

Since, considering the testing for the whole entity $\mathbf{X}$, the global hypotheses can be expressed as the union or intersection of a sequence of sub-hypotheses,

$$H_0 : \left\{ \bigcap_j H_{0[j]} \right\}, \quad vs. \quad H_1 : \left\{ \bigcup_j H_{1[j]} \right\},$$

Where it is to be emphasized that treatments ($j = 1, 2$) play the role of strata. Accordingly, the related conditional reference space can be expressed by the product of $\mathcal{X}_{/\mathbf{X}_{\cdot j}}$ in H_0, that is, $\mathcal{X}_{/\mathbf{X}} = \mathcal{X}_{/\mathbf{X}_{\cdot 1}} \times \mathcal{X}_{/\mathbf{X}_{\cdot 2}}$. In what follows, we will study the testing problem with respect to the strata, and then jointly analyze the whole entity.

4.3.1 Two relevant solutions

For determining the tendency of treatment effects over time, relevant solutions can be motivated by the case with two recording time points. First of all, we intend to find a proper way to decompose the whole entity into either a sequence of pairwise comparable columns in two parts. And then the problem can be handled by the UI-NPC approaches accordingly, in which each sub-problems can be assigned proper test statistics. It is worth cautioning that, all solutions should be based on the related conditional reference space, such as $\mathcal{X}_{/\mathbf{X}_{\cdot j}}$, $\mathcal{X}_{/\mathbf{X}}$. On the basis of permutation sufficient space, there are at least two approaches to tackle such a problem. In what follows, we will discuss two

decomposing methods under the conditional reference space $\mathcal{X}_{/\mathbf{X}_{\cdot j}}$, and the solutions to the whole entity are straightforward to obtain based on $\mathcal{X}_{/\mathbf{X}}$.

(I) The first method is to decompose the global problem into a set of pairwise comparable columns, namely, sub-problem has termed the case with two recording time points. On the basis of the permutation space $\mathcal{X}_{/\mathbf{X}_{\cdot j}}$ or $\mathcal{X}_{/\mathbf{X}}$, the permutation algorithm is working within all sub-orbits concurrently, and then proceed solutions to sub-problems after permuting. For such a case, some relevant solutions have been already proposed. Firstly, let's see how the decomposing method works.

For each j scheme, the hypotheses are equivalent to

$$H_{0[j]} \equiv \{\delta_{j,\cdot,1} = \cdots = \delta_{j,\cdot,T} = 0\} = \left\{\bigcap_{hk}\{\delta_{j,\cdot,h} = \delta_{j,\cdot,k}\}\right\} = \left\{\bigcap_{hk} H_{0hk[j]}\right\},$$

against

$$H_{1[j]} \equiv \{\delta_{j,\cdot,1} \geq \cdots \geq \delta_{j,\cdot,T}\} = \left\{\bigcup_{hk}\{\delta_{j,\cdot,h} > \delta_{j,\cdot,k}\}\right\} = \left\{\bigcup_{hk} H_{1hk[j]}\right\},\ 1 \leq h < k \leq T,$$

where denote the treatment effects belong to the j-th treatment scheme, t-th measurement by $\delta_{j,\cdot,t}$. For each pair h and k of time points with $h < k$, the paired sub-hypotheses H_{0hk} and H_{1hk} can be expressed with the intersection and union of a set of sub-sub-hypotheses, respectively. Such as

$$H_{0hk[j]} : \left\{\bigcap_{i}\{\delta_{j,i,h} = \delta_{j,i,k}\}\right\} \equiv \left\{\bigcap_{i}\left\{X_{j,i,h} \overset{d}{=} X_{j,i,k}\right\}\right\},$$

against

$$H_{1hk[j]} : \left\{\bigcup_{i}\{\delta_{j,i,h} > \delta_{j,i,k}\}\right\} \equiv \left\{\bigcup_{i}\left\{X_{j,i,h} \overset{d}{>} X_{j,i,k}\right\}\right\}.$$

Thus, the global problem is converted into studying a set of pairwise variables, and then by jointly analyze them within the UI-NPC approach. The related algorithm can be seen in Tables (4.3) .

In our case, the data are defined on the ordered categorical scale, in short, no assumption of continuity of data exists. Obviously, after permuting all data in sub-orbits concurrently, for each paired h and k, the related partial permutation tests are given by the previous discussion, one is based on the Poisson Binomial distribution, another one is based on the direct sum of differences.

▷ Based on Poisson Binomial distribution: for paired h and k, the partial permutation test is $U^*_{hk[j]} = \sum_i^{n_j} I(\Delta^*_{j,i,hk} > 0)$. Obviously, for any i, $\Delta_{j,i,hk}$ is symmetric around zero in H_0, $\Delta_{j,i,hk}$is symmetric around $(\delta_{j,i,h} - \delta_{j,i,k}) > 0$ in H_1. The underlying distribution of U in H_1 is slipped to the right side comparing with that in H_0; thus, large values of U are significant.

The standardized version of the permutation test statistic is

$$T^*_{U_{hk[j]}} = (U^*_{hk[j]} - E(U^*_{hk[j]}))/\sqrt{Var_{|\mathbf{X}_{\cdot j}}(U^*_{hk[j]})},$$

which is standard normally distributed as the sample size tends to infinity. The corresponding p-value-like statistics can be written as $\lambda_{T_{U_{hk[j]}}} = \Pr(T^*_{U_{hk[j]}} \geq T^o_{U_{hk[j]}} \mid \mathcal{X}_{/\mathbf{X}_{\cdot j}})$, where $T^o_{U_{hk[j]}} = T_{U_{hk[j]}}(\mathbf{X}_{\cdot j})$ is the observed value of T conditioning on the $\mathbf{X}_{\cdot j}$ for paired h and k.

As discussed previously, the p-value-like statistics play the role of tests whose common critical value is α; thus, if $\lambda_{T_{U_{hk[j]}}} \leq \alpha$, the null hypothesis is rejected at significance level $\alpha > 0$.

TABLE 4.3: From original to permutation

Original					
ID	Time				
	1	2	$\cdots$	$T-1$	T
1	$X_{j,1,1}$	$X_{j,1,2}$	$\cdots$	$X_{j,1,T-1}$	$X_{j,1,T}$
2	$X_{j,2,1}$	$X_{j,2,2}$	$\cdots$	$X_{j,2,T-1}$	$X_{j,2,T}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
n_j	$X_{j,n_j,1}$	$X_{j,n_j,2}$	$\cdots$	$X_{j,n,T-1}$	$X_{j,n_j,T}$

$\Downarrow$

Permutation					
ID	Time				
	1	2	$\cdots$	$T-1$	T
1	$X^*_{j,1,1}$	$X^*_{j,1,2}$	$\cdots$	$X^*_{j,1,T-1}$	$X^*_{j,1,T}$
2	$X^*_{j,2,1}$	$X^*_{j,2,2}$	$\cdots$	$X^*_{j,2,T-1}$	$X^*_{j,2,T}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
n_j	$X^*_{j,n_j,1}$	$X^*_{j,n,2}$	$\cdots$	$X^*_{j,n_j,T-1}$	$X^*_{j,n_j,T}$

Accordingly, we can obtain a set of partial test values $(T^*_{U_{12[j]}}, T^*_{U_{13[j]}}, \cdots, T^*_{U_{(T-1)T[j]}})$ or a set of p-values $(\lambda_{T_{U_{12[j]}}}, \lambda_{T_{U_{13[j]}}}, \cdots, \lambda_{T_{U_{(T-1)T[j]}}})$ which are necessarily dependent. The hypotheses can be tested by combining such dependent partial tests values or p-values, as

$$T^*_\psi = \Psi(T^*_{U_{12[j]}}, T^*_{U_{13[j]}}, \cdots, T^*_{U_{(T-1)T[j]}}) \equiv \psi(\lambda_{T_{U_{12[j]}}}, \lambda_{T_{U_{13[j]}}}, \cdots, \lambda_{T_{U_{(T-1)T[j]}}}).$$

It is well known that, since all partial tests share the same asymptotic distribution, the Direct combination function Ψ can be applied; Fisher's, Liptak's and Tippett's combination functions ψ are suitable more than for partial test statistics, for combining a set of p-values.

▷ Based on Direct sum of differences: for any paired h and k, with $h < k$, the partial permutation test is

$$\begin{aligned} T^*_{S_{hk[j]}} &= \sum_i^{n_j} \Delta^*_{j,i,hk} = \sum_i^{n_j} \left(X^*_{j,i,h} - X^*_{j,i,k}\right) \\ &= \sum_i^{n_j} \left\{(\delta^*_{j,i,h} - \delta^*_{j,i,k}) + \sigma(\delta^*_{j,i,h}) Z^*_{j,i,h} - \sigma(\delta^*_{j,i,k}) Z^*_{j,i,k}\right\}. \end{aligned}$$

The standardized version of permutation test statistic is

$$T^*_{S_{hk[j]}} = (\sum_i \Delta^*_{j,i,hk} - E(\sum_i \Delta^*_{j,i,hk}))/(Var_{|\mathbf{X}_{\cdot j}}(\sum_i \Delta^*_{j,i,hk}))^{1/2},$$

which is asymptotically standard normally distributed. The corresponding p-value-like statistics can be written as $\lambda_{T_{S_{hk[j]}}} = \Pr(T^*_{S_{hk[j]}} \geq T^o_{S_{hk[j]}} \mid \mathcal{X}_{/\mathbf{X}_{\cdot j}})$, where $T^o_{S_{hk[j]}} = T_{S_{hk[j]}}(\mathbf{X}_{\cdot j})$ is the observed value of T conditioning on the $\mathbf{X}_{\cdot j}$ for paired h and k. Obviously, for any i, $\Delta^*_{j,i,hk}$ is symmetric around zero in H_0, $\Delta^*_{j,i,hk}$ is symmetric around $(\delta^*_{j,i,h} - \delta^*_{j,i,k})$ in H_1. And the p-value-like statistics play the role of tests whose common critical value is α; thus, if $\lambda_{T_{S_{hk[j]}}} \leq \alpha$, the null hypothesis is rejected at significance level $\alpha > 0$.
Accordingly, we can obtain a set of partial test values $(T^*_{S_{12[j]}}, T^*_{S_{13[j]}}, \cdots, T^*_{S_{(T-1)T[j]}})$ or a set of p-values $(\lambda_{T_{S_{12[j]}}}, \lambda_{T_{S_{13[j]}}}, \cdots, \lambda_{T_{S_{(T-1)T[j]}}})$. The global hypotheses can be tested by combining such partial tests values or p-values, as

$$T^*_\psi = \Psi(T^*_{S_{12[j]}}, T^*_{S_{13[j]}}, \cdots, T^*_{S_{(T-1)T[j]}}) \equiv \psi(\lambda_{T_{S_{12[j]}}}, \lambda_{T_{S_{13[j]}}}, \cdots, \lambda_{T_{S_{(T-1)T[j]}}}).$$

It is well known that the Direct combination function Ψ can be applied to the

set of partial tests; Fisher's, Liptak's and Tippett's combination functions ψ are suitable for combining a set of p-values.

(II) The second approach is to decompose the global problem into $T-1$ sub problems, where the original data set can be divided into two pooled pseudo-groups in each sub-problem, the first pseudo-group is obtained by pooling data within units of the first t time points, and the second one by pooling the rest. To be specific, the first part is $\mathbf{X}_{1,j,(t)i} = X_{j,i,1} \biguplus \cdots \biguplus X_{j,i,t}$ and the second one is $\mathbf{X}_{2,j,(t)i} = X_{j,i,t+1} \biguplus \cdots \biguplus X_{j,i,T}$, for $t = 1, 2, \cdots, T-1$. The related decomposition scheme for $t = 1$ is shown in Table (4.4). When $t = 1$, the first pseudo-group is the first column (green part), and the second pseudo-group is by pooling the rest (yellow part). The global hypotheses can be rewritten as follows

$$H_{0[j]} : \left\{ \bigcap_{t=1}^{T-1} \left\{ \bigcap_{i}^{n_j} \left\{ \mathbf{X}_{1,j,(t)i} \overset{d}{=} \mathbf{X}_{2,j,(t)i} \right\} \right\} \right\}, \ vs. \ H_{1[j]} : \left\{ \bigcup_{t=1}^{T-1} \left\{ \bigcup_{i}^{n_j} \left\{ \mathbf{X}_{1,j,(t)i} \overset{d}{>} \mathbf{X}_{2,j,(t)i} \right\} \right\} \right\}.$$

TABLE 4.4: Two pooled pseudo-groups for $t = 1$

Original					
ID	Time				
	1	2	$\cdots$	$T-1$	T
1	$X_{j,1,1}$	$X_{j,1,2}$	$\cdots$	$X_{j,1,T-1}$	$X_{j,1,T}$
2	$X_{j,2,1}$	$X_{j,2,2}$	$\cdots$	$X_{j,2,T-1}$	$X_{j,2,T}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
n_j	$X_{j,n_j,1}$	$X_{j,n_j,2}$	$\cdots$	$X_{j,n_j,T-1}$	$X_{j,n_j,T}$

$\Downarrow$

Permutation					
ID	Time				
	1	2	$\cdots$	$T-1$	T
1	$X^*_{j,1,1}$	$X^*_{j,1,2}$	$\cdots$	$X^*_{j,1,T-1}$	$X^*_{j,1,T}$
2	$X^*_{j,2,1}$	$X^*_{j,2,2}$	$\cdots$	$X^*_{j,2,T-1}$	$X^*_{j,2,T}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
n_j	$X^*_{j,n_j,1}$	$X^*_{j,n,2}$	$\cdots$	$X^*_{j,n_j,T-1}$	$X^*_{j,n_j,T}$

For each t, the partial permutation test statistic can be

$$T^*_{(t)[j]} = \sum_i^n \left(\sum_{t=1}^{t} X^*_{j,i,t}/t - \sum_{t=t+1}^{T} X^*_{j,i,t}/(T-t) \right).$$

We are wondering if the partial permutation test is significant to reject H_0 for large value. When $t = 1$, the test statistic is in the form of

$$\begin{aligned} T^*_{(1)[j]} &= \sum_i^{n_j} \left(X^*_{j,i,1} - \sum_{t=2}^{T} X^*_{j,i,t}/(T-1) \right) \\ &= \sum_i^{n_j} \left(\sum_{t=2}^{T} \left(X^*_{j,i,1} - X^*_{j,i,t} \right) \right) /(T-1) = \sum_i^{n_j} \sum_{t=2}^{T} \left(\Delta^*_{j,i,1t} \right) /(T-1). \end{aligned}$$

For $t = T-2$, the test is in the form of

$$\begin{aligned} T^*_{(T-2)[j]} &= \sum_i^{n_j} \left(\sum_{t=1}^{T-2} X^*_{j,i,t}/(T-2) - \sum_{t=T-1}^{T} X^*_{j,i,t}/2 \right) \\ &= \sum_i^{n_j} \left(\sum_{t=1}^{T-2} \Delta^*_{j,i,t(T-1)} + \sum_{t=1}^{T-2} \Delta^*_{j,i,tT} \right) /2(T-2). \end{aligned}$$

When $t = T-1$, the test is

$$T^*_{(T-1)[j]} = \sum_i^{n_j} \left(\sum_{t=1}^{T-1} X^*_{j,i,t}/(T-1) - X^*_{j,i,T} \right) = \sum_i^{n_j} \sum_{t=1}^{T-1} \left(\Delta^*_{j,i,tT} \right) /(T-1).$$

In the spirit of mathematical induction, when $t = 1$, and $t = T-2$, both test statistics are the functions of differences, and we find the test is also a function of differences when $t = T-1$, we can infer that the test statistics $T^*_{(t)[j]}$ are the functions of differences. In H_0, $\boldsymbol{\delta}$ should take the value of a zero vector, and $\boldsymbol{\delta}$ is not a negative vector nor a zero vector in H_1. Clearly, we have

$$T^*_{(t)[j]|H_0}(\mathbf{0}) \overset{d}{<} T^*_{(t)[j]|H_1}(\boldsymbol{\delta}).$$

The partial permutation test is significant to reject H_0 for large values. The standardized version of that partial test is $(T^*_{(t)[j]} - E(T^*_{(t[j])}))/(Var_{|\mathbf{X}_{\cdot j}}(T^*_{(t)[j]}))^{1/2}$ which is standard normally distributed. Accordingly, the related p-value-like is $\lambda_{T_{(t)[j]}} = \Pr(T^*_{(t)[j]} \geq T^o_{(t)[j]} \mid \mathbf{X}_{\cdot j})$, where $T^o_{(t)[j]} = T_{(t)[j]}(\mathbf{X}_{\cdot j})$ is the observed value of $T_{(t)[j]}$ based on $\mathbf{X}_{\cdot j}$. Assume the critical value is α, the null hypothesis is rejected at significant value $\alpha > 0$ when $\lambda_{T_{(t)[j]}} \leq \alpha$. We can obtain a

set of p-values as $(\lambda_{T_{(1)[j]}}, \lambda_{T_{(2)[j]}}, \cdots, \lambda_{T_{(T-1)[j]}})$ and a set of partial test values as $(T^*_{(1)[j]}, T^*_{(2)[j]}, \cdots, T^*_{(T-1)[j]})$. There are some suitable combination functions for combining partial test values or p-values, such as Fisher's, Liptak's, Tippett's and Direct combining functions.

4.4 A medical example with repeated data

Now, let us consider a practical medical example in Table (4.4) from Lumley (1996), and also mentioned by Brunner and Langer (2000). The related table is shown in Table (4.5). The study subjects after receiving laparoscopic cholecystectomy are randomly assigned to two levels of treatment schemes, one is standard treatment (active drug), the other

TABLE 4.5: Pain scores on shoulder tip under two treatment schemes after laparoscopic cholecystectomy

Treatment	*Y*							*N*						
	Time							Time						
ID	1	2	3	4	5	6	**ID**	1	2	3	4	5	6
1	1	1	1	1	1	1	23	5	2	3	5	5	4
2	3	2	1	1	1	1	24	1	5	3	4	5	3
3	3	2	2	2	1	1	25	4	4	4	4	1	1
4	1	1	1	1	1	1	26	4	4	4	4	4	3
5	1	1	1	1	1	1	27	2	3	4	3	3	2
6	1	2	1	1	1	1	28	3	4	3	3	3	2
7	1	3	2	1	1	1	29	3	3	4	4	4	3
8	2	2	1	1	1	1	30	1	1	1	1	1	1
9	1	1	1	1	1	1	31	1	1	1	1	1	1
10	3	1	1	1	1	1	32	1	5	5	5	4	3
11	1	1	1	1	1	1	33	1	3	2	2	1	1
12	2	1	1	1	1	2	34	2	2	3	4	2	2
13	1	2	2	2	2	2	35	2	2	1	3	3	2
14	3	1	1	1	3	3	36	1	1	1	1	1	1
15	2	1	1	1	1	1	37	1	1	1	1	1	1
16	1	1	1	1	1	1	38	5	5	5	4	3	3
17	1	1	1	1	1	1	39	3	3	3	3	1	1
18	2	1	1	1	1	1	40	5	4	4	4	2	2
19	4	4	2	4	2	2	41	1	3	3	3	3	1
20	4	4	4	2	1	1							
21	1	1	1	2	1	1							
22	1	1	1	2	1	2							

is placebo. The patients are asked to rate the scores for shoulder tip pain before being given medication, the pain scores are defined on the ordered categorical scale with five categories, in which the pain scores are from 1 (low) to 5 (high), so denoting the five categories by $(v_1 = 1 \prec v_2 = 2 \prec \cdots \prec v_5 = 5)$. The total sample is made up of 41 study subjects, in which 22 of these belong to the treatment group Y and 19 pertain to the placebo or control group N. Further, the clinical data are observed at six different recording times.

Comparing with the time points and the number of ordered categories, the total sample size is quite small. The number of parameters for each unit is $5^6 - 1 = 15\,624$, and the total number of parameters is $41 \times (5^6 - 1) = 640\,584$. In such a framework, it looks impossible to handle the problem by using the likelihood approach.

4.4.1 Analysis for Treatment group

Firstly, we intend to answer the question of interest mentioned at the beginning of the chapter, whether the treatment effects will tend to decrease over time for the standard treatment scheme. For such a multidimensional case, the related regression function can be expressed as follows,

$$X_{1,i,t} = \mu + \eta_{1,i} + \delta_{1,i,t} + \sigma(\delta_{1,i,t}) \cdot Z_{1,i,t}, \quad i = 1, \cdots, 22;\ t = 1, \cdots, 6.$$

The related hypotheses are

$$H_{0[1]} : \left\{ \bigcap_i \{\delta_{1,i,1} = \delta_{1,i,2} = \cdots = \delta_{1,i,T} = 0\} \right\} = \left\{ \bigcap_i H_{0i[1]} \right\},$$

against

$$H_{1[1]} : \left\{ \bigcup_i \{\delta_{1,i,1} > \delta_{1,i,2} > \cdots > \delta_{1,i,T}\} \right\} = \left\{ \bigcup_i H_{1i[1]} \right\}.$$

As the solutions discussed previously, we can tackle such a testing problem by two decomposition schemes.

- ▷ For the first decomposition scheme, we intend to handle this problem within UI-NPC approach, based on Conditional Monte Carlo with $R = 100\,000$ random permutations. The corresponding sub-p-values are shown in Table (4.6),

TABLE 4.6: Sub-p-values for Treatment group under the first decomposition scheme

	p_{12}	p_{13}	p_{14}	p_{15}	p_{16}	p_{23}	p_{24}
Bi	0.05243	0.00349	0.00348	0.00341	0.01520	0.30785	0.30821
Sum	0.14683	0.00501	0.01147	0.00036	0.00211	0.09851	0.14508

	p_{25}	p_{26}	p_{34}	p_{35}	p_{36}	p_{45}	p_{46}	p_{56}
Bi	0.05424	0.05318	0.92187	0.76591	0.76439	0.30962	0.53850	1.0000
Sum	0.02188	0.06307	0.63614	0.28075	0.45179	0.20738	0.36377	0.72054

TABLE 4.7: p-values under combination functions of Treatment group

	Fisher	Liptak	Tippett	Direct
Binomial	0.00033	0.00015	0.01823	0.00027
Sum	0.00028	0.00005	0.00318	0.00007

▷ For the second decomposition scheme, based on Conditional Monte Carlo with $R = 100\,000$ random permutations. The related results are shown in Table (4.8),

TABLE 4.8: p-values for Treatment group under the second decomposition scheme

	H_{01}	H_{02}	H_{03}	H_{04}	H_{05}	Fisher	Liptak	Tippett	Direct
p	0.00119	0.00021	0.00426	0.00757	0.15323	0.00008	0.00008	0.00018	0.00007

Results in the Table (4.6), clearly show that the sub-p-values under the first decomposition algorithm. Obviously, we can obtain $15 = T \times (T-1)/2$ paired sub-hypotheses according to 6 recording time points, and each is assigned two test statistics, which are based on Poisson binomial test and direct sum of differences. Based on the related 15 sub-p-values, the global p value of preliminary stratification testing for the Treatment group is obtained by combination functions Fisher, Liptak, Tippett, Direct, are shown in the Table (4.7). All reject the null hypothesis at significant level $\alpha = 0.05$ of monotonic stochastic ordering among 6 recording time points. The combined result which is

given by Tippett's combination function, differs from that given by other combination functions because its power behavior is very sensitive to the small sub-p-values. Similar to Tippett, combining function Fisher is slightly sensitive to small p-values. The combined p-values of Liptak and the Direct are very similar.

Results in the Table (4.8), so obtained following the second decomposition scheme, also show that the 5 sub-p-values. The global p-values are given by Fisher's, Liptak's, Tippett's, the Direct combination functions. All p-values reject the null hypothesis at significant level $\alpha = 0.01$, which is acceptance of the alternative of monotonic stochastic ordering over time. In the second decomposition scheme, combining function Tippett's also shows slightly different from other combination functions, because it is very sensitive to the small p-values. Both decomposition schemes give the inferential conclusion, which is that the treatment effects in the Treatment group present a statistically significant decreasing tendency over time.

4.4.2 Analysis for Control group

Similarly, the testing problem is to detect whether the treatment effects will tend to decrease over time in the control group. For such a multidimensional case, the related regression function can be expressed as follows,

$$X_{2,i,t} = \mu + \eta_{2,i} + \delta_{2,i,t} + \sigma(\delta_{2,i,t}) \cdot Z_{2,i,t}, \quad i = 1, \cdots, 19; \; t = 1, \cdots, 6.$$

The related hypotheses are

$$H_{0[2]} : \left\{ \bigcap_i \{\delta_{2,i,1} = \delta_{2,i,2} = \cdots = \delta_{2,i,T} = 0\} \right\} = \left\{ \bigcap_i H_{0i[2]} \right\},$$

against

$$H_{1[2]} : \left\{ \bigcup_i \{\delta_{2,i,1} > \delta_{2,i,2} > \cdots > \delta_{2,i,T}\} \right\} = \left\{ \bigcup_i H_{1i[2]} \right\}.$$

As the solutions discussed previously, we can tackle such a testing problem by two decomposition schemes.

- ▷ For the first decomposition scheme, we intend to handle this problem within UI-NPC approach, based on Conditional Monte Carlo with $R = 100\,000$ random permutations. The corresponding sub-p-values are shown in Table (4.9),

TABLE 4.9: Sub-p-values for Control group under the first decomposition scheme

	p_{12}	p_{13}	p_{14}	p_{15}	p_{16}	p_{23}	p_{24}
Bi	0.98256	0.92525	0.98301	0.79938	0.20501	0.79918	0.79821
Sum	0.96640	0.94997	0.99154	0.60069	0.07158	0.46501	0.72653

	p_{25}	p_{26}	p_{34}	p_{35}	p_{36}	p_{45}	p_{46}	p_{56}
Bi	0.08812	0.00179	0.98248	0.03028	0.00020	0.08706	0.0000	0.03020
Sum	0.07062	0.00054	0.78154	0.09860	0.00085	0.02256	0.00005	0.04993

Based on the sub-p-values, the global p values of preliminary stratification testing for the control group are obtained by combination test functions Fisher's, Liptak's, Tippett's, Direct,

TABLE 4.10: p-values under combination functions of Control group

	Fisher	Liptak	Tippett	Direct
Binomial	0	0.00846	0	0.00301
Sum	0.00072	0.0212	0.00029	0.02162

▷ For the second decomposition scheme, based on Conditional Monte Carlo with $R = 100\,000$ random permutations. The related results are shown in Table (4.11),

TABLE 4.11: p-values for Control group under the second decomposition scheme

	H_{01}	H_{02}	H_{03}	H_{04}	H_{05}	Fisher	Liptak	Tippett	Direct
p	0.88294	0.37906	0.09764	0.00045	0.00058	0.00465	0.02441	0.00125	0.0232

Results in the Table (4.9), show 15 sub-p-values under the first decomposition algorithm, all paired sub-hypotheses are assigned two test statistics, which are based on Poisson binomial test and direct sum of differences. The global p values of preliminary stratification testing for the Control group are obtained by combination functions Fisher, Liptak, Tippett, the Direct, are shown in the Table (4.10). All results in the Table (4.10) give us

a significant evidence to reject the null hypothesis at significant level $\alpha = 0.05$, in which results from Fisher's and Tippett's are comparable, and those from Liptak's and the Direct are similar. Because both Fisher and Tippett are sensitive to the small values of p-value-like, which are likely close to the small value of sub-p-values. The combination functions Liptak and the Direct may be susceptible to the large p-values.

Results in the Table (4.11), given that the second decomposition scheme, show 5 sub-p-values of corresponding 5 partial tests $T_{(t)2}$. The global p-values are also given by Fisher's, Liptak's, Tippett's, the Direct combination functions. In the second decomposition scheme, p-values give us a significant evidence to reject the null hypothesis at significant level $\alpha = 0.05$, to reject the homogeneity effects in the null hypothesis over time. Due to the influence of small p-values in the partial tests, Fisher and Tippett yield similar results. The inference is that the treatment effects in the Control group present a statistically significant decreasing tendency over time.

4.4.3 Analysis for Whole entity data

For such a multidimensional case, the related regression function can be expressed as follows,

$$X_{j,i,t} = \mu + \eta_{j,i} + \delta_{j,i,t} + \sigma(\delta_{j,i,t}) \cdot Z_{j,i,t},\ j = 1, 2;\ i = 1, \cdots, n_j;\ t = 1, \cdots, 6.$$

The related hypotheses are

$$H_0 : \left\{ \bigcap_j H_{0[j]} \right\}, \quad vs. \quad H_1 : \left\{ \bigcup_j H_{1[j]} \right\}.$$

Similarly to the solutions discussed previously, we can tackle such a testing problem by two decomposition schemes.

- ▷ For the first decomposition scheme, we intend to handle this problem within UI-NPC approach, based on Conditional Monte Carlo with $R = 100\,000$ random permutations. The corresponding sub-p-values are shown in Table (4.12),

TABLE 4.12: Sub-p-values for Whole entity data under the first decomposition scheme

	p_{12}	p_{13}	p_{14}	p_{15}	p_{16}	p_{23}	p_{24}
Bi	0.53318	0.13262	0.23523	0.06668	0.01177	0.53339	0.53238
Sum	0.77848	0.41838	0.68570	0.04006	0.00188	0.18367	0.41565

	p_{25}	p_{26}	p_{34}	p_{35}	p_{36}	p_{45}	p_{46}	p_{56}
Bi	0.01184	0.00034	0.98958	0.13319	0.01236	0.06742	0.00011	0.53251
Sum	0.00718	0.00010	0.77859	0.06913	0.00454	0.01393	0.00060	0.14928

TABLE 4.13: p-values under combination functions of Whole entity data

	Fisher	Liptak	Tippett	Direct
Binomial	0	0	0.00073	0
Sum	0.00004	0.00014	0.00091	0.00016

▷ For the second decomposition scheme, based on Conditional Monte Carlo with $R = 100\,000$ random permutations. The related results are shown in Table (4.14),

TABLE 4.14: p-values for Whole entity data under the second decomposition scheme

	H_{01}	H_{02}	H_{03}	H_{04}	H_{05}	Fisher	Liptak	Tippett	Direct
p	0.1857	0.00839	0.00382	0.00002	0.00036	0.00008	0.00016	0.00004	0.00014

From the results given from the treatment group and control group, we can conjecture that the treatment effects show a decreasing tendency over time when we take the whole entity to be considered, i.e. irrespectively of treatment levels. Based on the sub-p-values, the global p values are obtained by combination functions Fisher's, Liptak's, Tippett's, Direct. All results shown in the Table (4.13) and in the Table (4.14) reject the null hypothesis at significant level $\alpha = 0.01$, also support our conjecture.

4.4.4 Third question

To answer the dominance question whether one taken placebo stochastically dominates the other received active drug, which is termed the $2 \times K$ multidimensional problem in the previous chapter. Formally, the multi-one-sided hypotheses are:

$$H_0 : X_T \overset{d}{=} X_N, \quad vs. \quad H_1 : X_T \overset{d}{<} X_N.$$

Equivalent to

$$H_0 \equiv \bigcap_{q=1}^{6} \left\{ \bigcap_{k=1}^{4} [F_{1q}(v_k) = F_{2q}(v_k)] \right\}, \; vs. \; H_1 \equiv \bigcup_{q=1}^{6} \left\{ \bigcup_{k=1}^{4} [F_{1q}(v_k) > F_{2q}(v_k)] \right\}.$$

We intend to deal with the problem within UI-NPC approach, based on Conditional Monte Carlo with $R = 100\,000$ random permutations. For tests: Anderson-Darling T^*_{AD}, on Scores T^*_W and on Mid-ranks T^*_{MR}. Given that the suitable combination functions: Fisher's T''_F, Liptak's T''_L, Tippett's T''_T and the Direct T''_D, the p-values based on UI-NPC method are shown in Table (4.15),

TABLE 4.15: p-values based on UI-NPC approach

	$T^*_{(1)}$	$T^*_{(2)}$	$T^*_{(3)}$	$T^*_{(4)}$	$T^*_{(5)}$	$T^*_{(6)}$	T''_D	T''_F	T''_L	T''_T
$\hat{\lambda}_{AD(j)}$	0.05974	0.00066	0.00002	0.00001	0.00029	0.00595	0.00002	0.00002	0.00002	0
$\hat{\lambda}_{W(j)}$	0.09162	0.00076	0.00002	0.00001	0.00031	0.00733	0.00001	0.00001	0.00002	0
$\hat{\lambda}_{MR(j)}$	0.11871	0.00053	0.00005	0.00001	0.00036	0.00752	0.0001	0.00002	0.00007	0

In the Table (4.15), results show sub-p-values from 5 sub-hypotheses and the combined p-values based on four combination functions T''_F, T''_L, T''_T, and the Direct T''_D. All combined p-values reject the null hypothesis at significant level $\alpha = 0.01$. Thus, we can infer that patients taking placebo stochastically dominates those taking standard treatment.

It is worth noting that, for permutation test statistic T^*_{AD} and T^*_W, the combined p-values are almost comparable, as they only differ slightly in the fifth digit under four combination functions. All the combined p-values by T^*_{AD} and T^*_W are very close to the best among that of sub-hypotheses.

Given that permutation test T^*_{MR}, the result 0.0001 by the combining function T''_D differs from other combination functions due to the result of $T^*_{(1)}$. Sub-p-value of $T^*_{(1)}$ shows great different from that of other partial tests based on Mid-ranks.

Chapter 5

Conclusions

Conclusions and future directions

In this thesis, we discuss the testing problems of interest, where one study is concerning whether there exists stochastic ordering with respect to time, and another study is concerning whether there are stochastic ordering or stochastic dominance among the comparison groups. All those measurements are defined on an ordered categorical scale.

First of all, we discussed $2 \times K$ unidimensional and multidimensional cases, the related permutation solutions based on Anderson-Darling, Mid-ranks, and Scores functions have been provided. It is worth noting that handling such a problem by traditional parametric approaches may become extremely difficult if not impossible. In contrast, the permutation method within the UI-NPC methodology may provide reasonable solutions. For the J-sample stochastic ordering case, the basic idea is to properly break down the global problem into a set of unidimensional sub-problems, where each sub-problem is processed with a proper permutation method, and then they are jointly analyzed. We named this the UI-NPC approach based on the permutation theory. The UI-NPC approach works within the conditionality principle of inference, where the conditioning is with respect to a set of sufficient statistics in the null hypothesis like the pooled observed data. So, it is based on the permutation testing approach and the NPC of dependent permutation tests. It is worth noting that when the set of observed data is minimal sufficient in the null hypothesis there is no reason to stay outside the permutation strategy.

In the face of repeated measure designs, to answer the question about detecting the stochastic ordering concerning time, the traditional approaches may exist dilemmas. Due to repeated measurements with outcome expressed on ordered categories, since a great number of nuisance parameters are to be removed when we are working within likelihood approaches, so that the total sample size is generally too small for obtaining

any useful joint estimation and related inference may become problematic. Secondly, the data look like a discrete stochastic process with unknown time-dependent transition matrices. While maintaining the integrity of information of each unit, and constructing an effective model without sacrificing calculation speed, it is indeed a dilemma to balance between both sides. Such problems have quite difficult solutions within the likelihood ratio theory which, in turn, and when available, have nice characterizations under their usually too stringent assumptions.

As it seems that the UI-NPC solutions are effective for the considered testing problems with ordered categorical data, there are at least some further future lines of research to be done. One line is concerning the power behavior of our solutions in some scenarios of practical interest. One more line may concern the analysis to establish under which conditions one specific test statistic is better than the others. A further line of research is to study the rate of convergence of our UI-NPC solutions to the best parametric competitor under the stringent conditions useful for the existence and availability of the latter.

Appendix A

Computational tools

The thesis deals with ordered constraint problems with a central focus on ordered categorical data. Main NPC routines have been achieved in statistical software tools `R`, `Python`, `SAS`®, StatXact®, MATLAB®, etc. All computing routines in the thesis were implemented in `R` statistical software (R Core Team (2017)).

In Chapter 3, we discuss the testing problem focussing on ordered categorical data and propose approaches to address whether there are stochastic ordering or stochastic dominance among comparison groups, the required computational time takes around 6 minutes with a laptop.

In Chapter 4, we deal with testing problems still focussing on ordered categorical data, in which one considers whether the treatment effects will tend to decrease over time, another one is detecting whether the placebo stochastically dominates standard treatment observations. For the first question, we equivalently divided the original problem into a set of sub-problems with two decomposition methods, for details refer to chapter 4. The required computational time is lower than 2 minutes. For the second question, to detect stochastic dominance between two repeated-measure-design samples, the related analysis would be addressed by the method of chapter 3, and the computational time costs around 3 minutes.

Bibliography

Abrahamson, I. G. (1964) Orthant probabilities for the quadrivariate normal distribution. *Annals of Mathematical Statistics* **35**, 1685–1703.

Agresti, A. and Coull, B. A. (1998) Order-restricted inference for monotone trend alternatives in contingency tables. *Computational Statistics & Data Analysis* **28**, 139–155.

Agresti, A. and Coull, B. A. (2002) The analysis of contingency tables under inequality constraint. *Journal of Statistical Planning and Inference* **107**, 45–73.

Agresti, A., Mehta, C. R. and Patel, N. R. (1990) Exact inference for contingency tables with ordered categories. *Journal of the American Statistical Association* **85**, 453–458.

Arboretti, G. R. and Bonnini, S. (2009) Some new results on univariate and multivariate permutation tests for ordinal categorical variables under restricted alternatives. *Statistical Methods and Applications* **18**, 221–236.

Arboretti, G. R., Pesarin, F. and Salmaso, L. (2007) Permutation Anderson-Darling type and moment-based test statistics for univariate ordered categorical data. *Communications in Statistics: Simulation and Computation* **36**, 139–150.

Bai, Z. and Saranadasa, H. (1996) Effect of high dimension: by an example of a two sample problem. *Statistica Sinica* **6**, 311–329.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972) *Statistical inference under order restrictions: The theory and application of isotonic regression.* New York: John Wiley & Sons, Inc.

Basso, D., Pesarin, F., Salmaso, L. and Solari, A. (2009) *Permutation tests for stochastic ordering and ANOVA: theory and applications with R.* New York: Springer.

Bazyari, A. and Pesarin, F. (2013) Parametric and permutation testing for multivariate monotonic alternatives. *Statistics and Computing* **23**, 639–652.

Birnbaum, A. (1954) Combining independent tests of significance. *Journal of the American Statistical Association* **49**, 559–574.

Birnbaum, A. (1955) Characterizations of complete classes of tests of some multiparametric hypotheses, with applications to likelihood ratio tests. *The Annals of Mathematical Statistics* **26**, 21–36.

Brunner, E. and Langer, F. (2000) Nonparametric analysis of ordered categorical data in designs with longitudinal observations and small sample sizes. *Biometrical Journal* **42**, 663–675.

Chuang-Stein, C. and Agresti, A. (1997) A review of tests for detecting a monotone dose–response relationship with ordinal response data. *Statistics in Medicine* **16**, 2599–2618.

Cohen, A., Kemperman, J. H. B. and Sackrowitz, H. B. (2000) Properties of likelihood inference for order restricted models. *Journal of Multivariate Analysis* **72**, 50–77.

Cohen, A., Madigan, D. and Sackrowitz, H. B. (2003) Effective directed tests for models with ordered categorical data. *Australian & New Zealand Journal of Statistics* **45**, 285–300.

Colombi, R. and Forcina, A. (2016) Testing order restrictions in contingency tables. *Metrika* **79**, 73–90.

Davidov, O. and Peddada, S. (2011) Order-restricted inference for multivariate binary data with application to toxicology. *Journal of the American Statistical Association* **106**, 1394–1404.

Dykstra, R. L. (1982) Maximum likelihood estimation of the survival functions of stochastically ordered random variables. *Journal of the American Statistical Association* **77**, 621–628.

Feltz, C. J. and Dykstra, R. L. (1985) Maximum likelihood estimation of the survival functions of N stochastically ordered random variables. *Journal of the American Statistical Association* **80**, 1012–1019.

Gökpinar, F., Gökpinar, E. and Bayrak, H. (2017) Permutation approach for ordinal preference data. *Communications in Statistics - Simulation and Computation* **46**, 2321–2332.

Graubard, B. I. and Korn, E. L. (1987) Choice of column scores for testing independence in ordered $2 \times K$ contingency tables. *Biometrics* **43**, 471–476.

Hirotsu, C. (1982) Use of cumulative efficient scores for testing ordered alternatives in discrete models. *Biometrika* **69**, 567–577.

Hirotsu, C. (1986) Cumulative chi-squared statistic as a tool for testing goodness of fit. *Biometrika* **73**, 165–173.

Jelizarow, M., Cieza, A. and Mansmann, U. (2015) Global permutation tests for multivariate ordinal data: Alternatives, test statistics and the null dilemma. *Journal of the Royal Statistical Society. Series C: Applied Statistics* **64**, 191–213.

Kateri, M. (2011) On the comparison of two ordinal responses. *Communications in Statistics - Theory and Methods* **40**, 3748–3763.

Kateri, M. and Agresti, A. (2013) Bayesian inference about odds ratio structure in ordinal contingency tables. *Environmetrics* **24**, 281–288.

Kudo, A. (1963) A multivariate analogue of the one-sided test. *Biometrika* **50**, 403–418.

Lehmann, E. L. (1986) *Testing Statistical Hypotheses, 2nd edn.* New York: John Wiley & Sons, Inc.

Lumley, T. (1996) Generalized estimating equations for ordinal data: A note on working correlation structures. *Biometrics* **52**, 354–361.

Nair, V. N. (1987) Chi-squared–type tests for ordered alternatives in contingency tables. *Journal of the American Statistical Association* **82**, 283–291.

Neuhäuser, M. (2006) An exact test for trend among binomial proportions based on a modified Baumgartner-Weiß-Schindler statistic. *Journal of Applied Statistics* **33**, 79–88.

Perlman, M. D. and Wu, L. (2003) On the validity of the likelihood ratio and maximum likelihood methods. *Journal of Statistical Planning and Inference* **117**, 59–81.

Pesarin, F. (1990) On a nonparametric combination method for dependent permutation tests with applications. *Psychotherapy and Psychosomatics* **54**(2-3), 172–179.

Pesarin, F. (1992) A resampling procedure for nonparametric combination of several dependent tests. *Statistical Methods & Applications* **1**(1), 87–101.

Pesarin, F. (2001) *Multivariate Permutation Tests: With Applications to Biostatistics*. Chichester: John Wiley & Sons, Ltd.

Pesarin, F. (2006) Permutation test: Multivariate. In *Encyclopedia of Statistical Sciences*. New York: John Wiley & Sons, Inc.

Pesarin, F. (2015) Some elementary theory of permutation tests. *Communications in Statistics - Theory and Methods* **44**, 4880–4892.

Pesarin, F. and Salmaso, L. (2006) Permutation tests for univariate and multivariate ordered categorical data. *Austrian Journal of Statistics* **35**, 315–324.

Pesarin, F. and Salmaso, L. (2010) *Permutation tests for complex data: theory, applications and software*. Chichester: John Wiley & Sons, Ltd.

Pesarin, F., Salmaso, L., Carrozzo, E. and Arboretti, R. (2016) Union–intersection permutation solution for two-sample equivalence testing. *Statistics and Computing* **26**, 693–701.

R Core Team (2017) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988) *Order Restricted Statistical Inference*. Chichester: John Wiley & Sons, Ltd.

Roy, S. N. (1953) On a heuristic method of test construction and its use in multivariate analysis. *The Annals of Mathematical Statistics* **24**, 220–238.

Sampson, A. R. and Whitaker, L. R. (1989) Estimation of multivariate distributions under stochastic ordering. *Journal of the American Statistical Association* **84**, 541–548.

Silvapulle, M. J. and Sen, P. K. (2005) *Constrained Statistical Inference: Inequality, Order and Shape Restrictions*. Hoboken, NJ: John Wiley & Sons, Inc.

Wang, Y. (1996) A likelihood ratio test against stochastic ordering in several populations. *Journal of the American Statistical Association* **91**, 1676–1683.

Weiß, C. H. (2019) Distance-based analysis of ordinal data and ordinal time series. *Journal of the American Statistical Association* pp. 1–37.

Zhang, J. T., Guo, J., Zhou, B. and Cheng, M. Y. (2019) A simple two-sample test in high dimensions based on L^2 -norm. *Journal of the American Statistical Association* pp. 1–42.

Zhu, G. and Chen, J. (2018) Multi-parameter one-sided monitoring tests. *Technometrics* **60**, 398–407.

Huiting Huang

CURRICULUM VITAE

Contact Information

University of Padova
Department of Statistical Sciences
via Cesare Battisti, 241
35121 Padova. Italy.

Tel. +39 377 849 8239
e-mail: huang@stat.unipd.it

Current Position

Since October 2016 (expected completion: April 2020);
Ph.D. Student in Statistical Sciences, University of Padova.
Thesis title: Permutation tests for stochastic ordering with ordered categorical data.
Supervisor: Prof. Luigi Salmaso
Co-supervisor: Prof. Fortunato Pesarin and Prof. Dabuxilatu Wang.

Research interests

- *Ordered constraints problems with ordered categorical data;*
- *Ordered constraints problems with complex designed data;*
- *Analysis to establish under which conditions one specific test statistic is better than the others;*
- *Study the rate of convergence of our UI-NPC solutions to the best parametric competitor under the stringent conditions useful for the existence and availability of the latter.*

Education

September 2013 – June 2016
Master (*laurea specialistica/magistrale***) degree in Science**.
University of Guangzhou, Faculty of Mathematics and Information Science
Title of dissertation: "Some evaluation methods for the outstanding claims reserve based on fuzzy linear regression models"
Supervisor: Prof. Dabuxilatu Wang

September 2009 – June 2013
Bachelor degree (*laurea triennale***) in Science**.
University of Huizhou, Faculty of Mathematics
Title of dissertation: "Discussion on the branches of a polyhedral chain-link model with chiral"
Supervisor: Prof. Xiaosheng Cheng

Visiting periods

December 2018 – March 2019
University of Guangzhou,
Guangzhou, China.
Supervisor: Prof. Dabuxilatu Wang

Work experience

September 2015 – May 2016
Prof. Dabuxilatu Wang.
The editorial member of the book ***Insurance Actuarial Science****. This book is for undergraduate students in the major of Financial Mathematics which is only used for internal circulation, not for publication.*

September 2014 – July 2015
Prof. Dabuxilatu Wang.
A tutor for undergraduates.

September 2010 – December 2011
Parents of a high-school student.
A mathematics private teacher.

March 2012 – May 2012
Zijin middle school.
A trainee teacher.

Awards and Scholarship

2016-2019
PhD scholarship, Guangzhou University.

11/2015
Graduate Students Scholarship of Guangzhou University (Grade 2), Guangzhou University.

Computer skills

- *Proficient at LaTeX;*
- *R;*
- *Proficient at Office software system;*
- *Further statistical software that I use: SPSS, Matlab.*

Language skills

- *Chinese: mother tougue.*
- *English: moderate (written).*

Publications

Articles in journals
Wang D., Huang H., Jiang C. and Zhang Q. (2015) A Fuzzy CUSUM Control Chart based on Bootstrap Distribution. Journal of Guangzhou University **14**, *4–9.*

Chapters in books

Pesarin F., Salmaso L., Huang H., Arboretti R. and Ceccato R. (2018) Permutation tests for stochastic ordering with ordinal data. In Advances in Statistical Modelling of Ordinal Data, eds. Capecchi S., Di Iorio F. and Simone R. pp. 163–170, Publisher, Italy.

Huang, H., Pesarin, F., Arboretti, R., and Ceccato, R. (2019) Multivariate permutation tests for ordered categorical data. Manuscript submitted for publication.

References

Prof. Luigi Salmaso
University of Padova
Stradella S. Nicola, 336100 Vicenza - Italy
Phone: +39 0444 99 87 20
e-mail: luigi.salmaso@unipd.it

Prof. Fortunato Pesarin
University of Padova
Via C. Battisti, 241 - 35121 Padova - Italy
e-mail: pesarin@stat.unipd.it

Prof. Dabuxilatu Wang
University of Guangzhou
230 Wai Huan Xi Road,Guangzhou Higher Education Mega Center, Guangzhou 510006 - P.R.China
Phone: +86 1852 0770 116
e-mail: wangdabu@gzhu.edu.cn