Università degli Studi di Padova

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di scienze Statistiche
SCUOLA DI DOTTORATO DI RICERCA IN: SCIENZE STATISTICHE
CICLO XXII

METHODOLOGICAL ADVANCES IN PERMUTATION TESTS: MULTI-SIDED TESTS AND RELATED TOPICS

Direttore della Scuola: Ch.ma Prof.ssa Alessandra Salvan
Supervisore: Ch.mo Prof. Fortunato Pesarin
Co-supervisore: Ch.mo Prof. Luigi Salmaso

Dottorando: Francesco Bertoluzzo

Data: 01/02/2010

Sommario

L'interesse iniziale di questo lavoro era testare l'effetto di un generico trattamento applicato a superfici tridimensionali. L'analisi di superfici tridimensionali presenta diversi problemi di varia natura. Innanzi tutto, ai dati rilevati sulle superfici per mezzo di scansioni laser non sono direttamente applicabili test statistici per almeno due motivi: il numero di punti rilevati non è lo stesso per tutti i soggetti, i punti non sono sincronizzati nel senso che punti riferiti a soggetti diversi, ma aventi la stessa posizione nella sequenza digitale, possono essere riferiti ad aree diverse della superficie. Questo problema è stato risolto utilizzando Funzioni a Base Radiale che forniscono dei coefficienti a cui è possibile applicare direttamente le procedure statistiche. Il problema è complicato dal fatto che il trattamento può generare, per alcuni coefficienti, degli effetti positivi su alcuni soggetti e negativi su altri e inoltre il numero dei coefficienti fornito dalla rappresentazione è notevolmente superiore al numero delle osservazioni. Per risolvere il primo di questi problemi è nato il multi-sided test. Il suo sviluppo in ambiente non parametrico ha contribuito a risolvere il secondo problema. Questo test è applicabile nelle situazioni in cui l'effetto di un trattamento può essere positivo su alcuni individui e negativo sugli altri. Tale situazione è sostanzialmente diversa da quella considerata nei tradizionali test bilaterali nei quali si assume che solo una delle due alternative può essere attiva, non entrambe. Il test multi-sided considera attive congiuntamente le due direzioni anche se in soggetti diversi. Al fine di affrontare questa situazione atipica, si possono applicare prima due "goodness-of-fit" tests, uno per gli effetti positivi e l'altro per quelli negativi e procedere poi con la loro combinazione non parametrica per via permutazione. Nelle situazioni in cui il numero di variabili è maggiore del numero di osservazioni, come nell'analisi di superfici, è conveniente utilizzare i test di permutazione poiché la potenza del test globale che si ottiene dalla combinazione dei test parziali, fatte salve alcune condizioni, aumenta monotonicamente al crescere della noncentralità.
Infine, con opportune tecniche di correzione per la molteplicità è possibile identificare zone delle superfici maggiormente interessate dal trattamento.

Abstract

The initial objective of this work was to test the effect of a general treatment on three-dimensional surfaces. The analysis of three-dimensional surfaces has several problems of different nature. Firstly ordinary statistical tests are not directly applicable to collected data on the surface using laser scans for two reasons: the number of collected points is not the same for all subjects, moreover the points are not synchronized in the sense that points related to different subjects, but having the same position in the digital sequence, can be related to different areas of the surface. This problem has been solved by using Radial Basis Functions that provide coefficients to which statistical procedures can be applied directly applied. The problem is complicated by the fact that the treatment may generate on some coefficients positive effects on some subjects and negative effects on others and also the number of coefficients obtained from the representation is far greater than the number of subjects. The multi-sided test is born to solve the first of these problems. Its development in nonparametric environment has helped also to solve the second problem. The use of the multi-sided test has proved to be useful in many situations, where the effect of a treatment can be positive on some individuals and negative on the rest. Such a situation is essentially different from that of the traditional two-sided test, in which the alternative is assumed being active on only one of two directions, but not on both. The multisided-test allows the two sides alternative to be jointly active although in different subjects. In order to face such an atypical situation, one can first apply two goodness-of-fit tests, one for the positive effects and the other for the negative, and then to proceed with their nonparametric combination within a permutation framework. The use of permutation test is also useful in the context where the number of variables is much larger than the number of subjects, since it can be proven that, if some conditions are satisfied, the power function of permutation tests monotonically increases as the related noncentrality parameter increases. This property also holds for multivariate situations.

Finally, with appropriate multiplicity control techniques we can identify the areas of surfaces that are mostly affected by the treatment.

Contents

1 Introduction **1**
1.1 Main Contributions of the Thesis 2

2 Multi-sided permutation tests **5**
2.1 Introduction 5
2.2 The problem of negative estimates of variance components 8
2.3 Multi-sided permutation test 9
2.3.1 Introduction 9
2.3.2 Sufficient statistics, exchangeability and similarity . . . 11
2.3.3 The statistics 13
2.3.4 Exactness and Unbiasedness 14
2.3.5 Consistency 16
2.3.6 Combination of partial tests 16
2.4 Simulation study 17
2.5 Multivariate extension of the test 19
2.6 Conclusions 19

3 Finite-sample consistency of combination-based permutation tests **33**
3.1 Introduction 33
3.2 Finite sample consistency 34
3.2.1 Weak unconditional finite-sample consistency of T . . . 35
3.2.2 Unconditional finite sample consistency for $V \to \infty$ . . 37
3.2.3 Weak unconditional consistency of T for $n \to \infty$ 39
3.2.4 Weak unconditional finite-sample consistency for random effects 40
3.3 Consistency of multi-sided test 42
3.4 Simulation study 43
3.5 Conclusion 54

4 Nonparametric Weighted Step Down Holm Method with heteroscedastic variables 55
4.1 Introduction . . . 55
4.2 The multiple testing problem . . . 55
4.3 Weighted step-down method . . . 56
4.4 Permutation WSDH . . . 57
4.5 The choice of the weights . . . 59
4.6 Conclusions . . . 59

5 Nonparametric Functional Data Analysis of 3-D surfaces 63
5.1 Introduction . . . 63
5.2 A three-dimensional data in orthognathic surgery . . . 63
5.2.1 Oral-maxillofacial surgery . . . 63
5.2.2 The 3-Dimensional approach . . . 64
5.3 Functional Data . . . 65
5.3.1 Some properties of functional data . . . 65
5.3.2 The interplay between smooth and noisy variation . . . 66
5.3.3 Smoothing data using a basis system by least squares . . . 67
5.3.4 The penalized sum of squared errors fitting criterion . . . 69
5.4 Representation of 3D surfaces with Radial Basis Function . . . 70
5.4.1 Fitting an implicit function to a surface . . . 70
5.4.2 The Radial Basis Functions . . . 71
5.5 Fast Multipole Method . . . 73
5.5.1 RBF centers reduction . . . 74
5.6 Application of the tests . . . 74
5.7 Conclusion . . . 75

References 77

Chapter 1

Introduction

Why permutation multi-sided test? This type of approach is the solution, or better a part of the solution to a problem of shape analysis considered from the functional point of view. The initial goal of the work was the construction of a procedure to check the regions of discrepancy in three-dimensional surfaces before and after a general treatment (industrial, surgical, pharmacological, economic, etc.). The surfaces are represented in digital form using a laser scanning of the original surface. The problem has proved more complex than anticipated and equally unexpected was the usefulness of the solution, or rather, as stated previously, part of the solution represented precisely by multi-sided test. In the three-dimensional analysis there are two major statistical problems: the effect of the treatment on a component variable may be positive on some subjects and negative on others, and the number of variables (e.g. three times the points considered in the surface) is far greater than the number of observed units. The multi-sided test is born to solve the first of these problems. Its development in nonparametric environment has helped to solve the second problem. The use of the multi-sided test has proved to be useful in different fields such as clinical trials, the environment, epidemiology, genetics, pharmacology, etc., where there are situations in which the effect of a drug treatment can be positive on some individuals and negative on the rest. Formally this situation can be expressed with a model for responses where a random effect Δ in the alternative is such that $\Pr\{\Delta < 0\} > 0$, $\Pr\{\Delta > 0\} > 0$ and $\Pr\{\Delta < 0\} + \Pr\{\Delta > 0\} = 1$. Such a situation is essentially different from that of the traditional two-sided test, in which the alternative is assumed being active on only one of two directions, but not on both. We want to consider alternatives in which two sub-alternatives $(\Delta \overset{d}{<} 0, \Delta \overset{d}{>} 0)$ can be jointly true. Thus, starting for instance from an underlying unimodal distribution in H_0, the response distribution in the alternative may become bimodal. In order to face such an atypical situation,

one can first apply two goodness-of-fit tests, one for the positive effects and the other for the negative, and then to proceed with their nonparametric combination within a permutation framework. Of course the two partial tests are not independent, since are calculated on the same dataset and so some kind of dependence is generally present. This dependence is extremely difficult to model, to analyze and to take into account explicitly. Thus, it must be analyzed nonparametrically.
The use of permutation test is useful in the context of shape analysis where the number of variables is much larger than the number of observed subjects, since it can be proven that, the power function of permutation tests monotonically increases as the related noncentrality parameter increases. This property also holds for multivariate situations. In particular, for any added variable the power does not decreases if each variable makes larger noncentrality. For a given and fixed number of observations, when the number of variables and the associated noncentrality parameter both diverge, then the power of multivariate permutation tests based on nonparametric combining functions converges to one (finite-sample-consistency) provided that the test statistics in the null hypothesis converges to a random variable and in the alternative the global noncentrality diverges.
A strictly related topic to multivariate analysis is the multiplicity control. A major drawback of multiple testing is the greatly increased probability of declaring “false significances”, or statistically significant associations where none exists in reality. A related negative feature is that it is very easy to overstate the evidence for a particular association if the statistical test that best supports a given hypothesis is chosen. One solution for solving the multiplicity dilemma is to make the individual tests more conservative, or more difficult to arrive at rejecting partial null hypotheses H_{0i}. In this dissertation we propose a permutation-based test procedure controlling the family wise error rate (FWE) by Weighted Step-Down Holm methods (WSDH)
Finally, our approach to shape analysis that makes use of the three topics mentioned above with the representation of surfaces by a particular kind of three-dimensional splines is presented.

1.1 Main Contributions of the Thesis

An overview of the original results obtained during the Ph.D. thesis development and presented in the thesis is given below.

- The multisided-test is a method that checks the presence of an effect in a random effect model. Traditional two-sided tests require that the alternatives $\left\{\Delta \overset{d}{\neq} 0\right\}$ is either $(\Delta \overset{d}{>} 0)$ or $(\Delta \overset{d}{<} 0)$, but not both. So in the presence of random effects, this kind of alternatives and related tests are not appropriate because both alternatives can be active. Instead we need testing for alternatives such that $\Pr\{\Delta < 0\} > 0$, $\Pr\{\Delta > 0\} > 0$ and $\Pr\{\Delta < 0\} + \Pr\{\Delta > 0\} = 1$ so both sides can be jointly active although on different subjects. More generally we present a procedure that performs a goodness-of-fit test $H_0 : \{F_1 = F_2\}$ in which in the alternative hypothesis both stochastic dominance $(F_1 \leq F_2)$ and $(F_1 \geq F_2)$ can jointly hold in separate sets of units.

- Working with high dimensional data and low sample size a quite important problem usually occurs. In (Pesarin, 2001) it is shown that, under very mild conditions, the power function of permutation tests monotonically increases as the related noncentrality parameter increases. This is true also for multivariate situations. Specifically, we will see that, for a given and fixed number of subjects, when the number of variables and associated noncentrality parameter δ both diverge, then the power function of multivariate NPC test converges to one if some conditions are satisfied. Such a property looks very relevant to solve multivariate small sample problems since it ensures that it is possible to obtain powerful tests in a nonparametric framework by increasing the number of informative variables while the number of cases is held fixed. An exhaustive simulation study is also presented.

- We extend the Weighted Step-Down Holm method with data-driven weights to the permutation framework and in heteroscedastic situations provided that the chosen weights are permutation invariant. The simulation study shows that even with heteroscedastic variables, if the non-centrality parameters are in terms of signal to noise ratio, the sample variance is still an acceptable permutation invariant indicator for the construction of the weights.

- We propose a procedure for representing three-dimensional surfaces using Radial Basis Functions. We use this kind of representation since in this way we can minimize the penalized residual sum of square, an index useful for statistical representation of smoothed surfaces. With this type of representation the application of permutation tests with the above developments becomes particularly easy.

Chapter 2

Multi-sided permutation tests

2.1 Introduction

In fields such as chemical trials, the environment, epidemiology, genetics, pharmacological, etc., situations in which the effect of a drug treatment can be positive on some individuals and negative on the rest may often occur. Formally this situation can be expressed using a model for the responses where a random effect Δ in the alternative is such that $\Pr\{\Delta < 0\} > 0$, $\Pr\{\Delta > 0\} > 0$ and $\Pr\{\Delta < 0\} + \Pr\{\Delta > 0\} = 1$. Such a situation is essentially different from that of the traditional two-sided test, in which the alternative is assumed to be active on one of two directions, but not on both. We wish to consider alternatives in which two sub-alternatives $(\Delta \overset{d}{<} 0)$ and$(\Delta \overset{d}{>} 0)$, where $\overset{d}{<}$ and $\overset{d}{>}$ stand for dominance in distribution (i.e. stochastic dominance), can be jointly true. Thus, for instance from an underlying unimodal distribution in H_0, the response distribution in the alternative may become bimodal. In order to deal with such an atypical situation, we can firstly apply two goodness of fit tests, one for the positive effects and the other for negative effects, and then proceed with their nonparametric combination within a permutation framework. Firstly we introduce models with random effects and highlight the problems associated with estimates of parameters within the traditional framework and then we propose a methodology for the nonparametric testing of hypotheses on the random effects.
As every experimentalist knows, subject responses vary from trial to trial. Furthermore, responses vary from subject to subject. These two sources of variability, within-subject and between-subjects, must both be taken into account when making inferences on the population. If we consider the effects Δ as random, after being observed the permutation analysis treats them in the same way as fixed effects conditionally to subject, but random between-

subjects. This allows us to make inferences on population.
Underlying any analysis is a probability model defined as follows: let δ be the mean effect in the population (i.e. averaged across subjects) and σ_b^2 the variability of this effect between subjects. This process reflects the fact that we are drawing subjects at random from a large population. We take the within-subject variability into account by modelling the h-th observation in subject i as being drawn from a distribution F_w with mean Δ_i and variance σ_w^2. Given a data set of observations from n subjects with v replications of observations per subject, the population effect is modelled by a two-level process

$$\begin{aligned} y_{hi} &= \Delta_i + e_{hi} \qquad (2.1)\\ \Delta_i &= \delta + z_i \qquad (2.2) \end{aligned}$$

where e_{hi} is a random variable with distribution F_w, mean 0 and variance σ_w^2, z_i is a random variable with distribution F_b, mean 0 and variance σ_b^2, for $i = 1 \dots n$ and $h = 1 \dots v$. The first equation captures the within-subject variability and the second the between-subject variability. Note that the within-subject variability σ_w^2 is assumed to be the same for all subjects. This assumption is not always reasonable. Nevertheless, it is usually adopted because no results are available under more complicated models (Scheffé, 1959). Within the parametric normal approach it is also assumed that the errors $\{e_{hi}\}$ and $\{z_i\}$ are independent and both distributions F_w and F_b are Normal. This two-stage process is shown graphically in Figure 2.1. The dotted line is the Normal distribution with mean $\delta = 50$ and variance $\sigma_b^2 = 10$ from which the Δ_i are observed; the solid lines are the Normal distributions with mean Δ_i and variance $\sigma_w^2 = 3$ from which the y_{hi} are observed; and the crosses represent the observed data y_{hi}. Collapsing the two levels into one gives

$$y_{hi} = \delta + z_i + e_{hi} \qquad (2.3)$$

Considering equations (2.1), (2.2) and (2.3), and the above assumptions of independence and normality of errors we can write the conditional and unconditional distributions of the y_{hi} observations

$$\begin{aligned} Y_{hi}|Z_i &= N(\Delta_i, \sigma_w^2)\\ Y_{hi} &= N(\delta, \sigma_w^2 + \sigma_b^2). \end{aligned}$$

Two observations y_{hi} and $y_{h'i}$ $(h' \neq h)$ are not statistically independent (the so-called within-subject dependence). The statistical dependence in the above random-effects model is formulated in a concept, useful in applications,

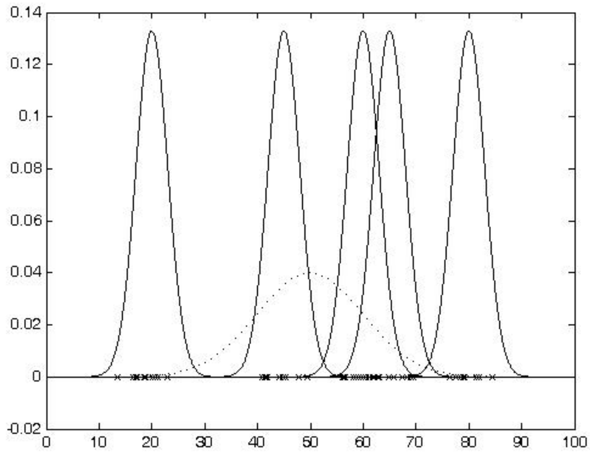

Figure 2.1: The two-stages generating process of the observed y_{hi}

called the interclass correlation coefficient defined as the ordinary correlation coefficient between observations in the same class (i.e. with the same i)

$$\begin{aligned}\tilde{\rho} &= E\left[(y_{hi}-\delta)(y_{h'i}-\delta)\right]/\sigma_y^2 \\ &= E\left[(z_i+e_{hi})(z_i+e_{h'i})\right]/\sigma_y^2 \\ &= E(z_i^2)/\sigma_y^2;\end{aligned}$$

hence

$$\tilde{\rho} = \frac{\sigma_b^2}{\sigma_b^2+\sigma_w^2}$$

If the number of $\{v_i\}$ are subject-invariant, i.e. are equal for $\forall i = 1 \ldots n$, the two-way layout is said to be balanced. When the $\{v_i\}$ are not equal, the model is said to be unbalanced. In the unbalanced framework with random effects the “best” tests and estimates are not known. The basic problem is that the distribution theory is much more complicated. The assumptions made thus far can be summarized as follows:

- (A.1) $y_{hi} = \delta + z_i + e_{hi}$;
- (A.2) the $n + nv$ random variables $\{e_{hi}\}$ and $\{z_i\}$ are independent;
- (A.3) the $\{e_{hi}\}$ are $N(0, \sigma_w^2)$
- (A.4) the $\{z_i\}$ are $N(0, \sigma_b^2)$.

2.2 The problem of negative estimates of variance components

Let $\bar{y}_{..}$ be the overall mean of the observations and $\bar{y}_{.i}$ be the mean of the observations made on the i-th subject. For model (2.3) we can define the following sums of squares

$$\begin{aligned} SS_b &= v\sum_{i=1}^{n}(\bar{y}_{.i}-\bar{y}_{..})^2 \\ SS_w &= \sum_{i=1}^{n}\sum_{h=1}^{v}(y_{hi}-\bar{y}_{.i})^2, \end{aligned}$$

Under model (2.3)

$$\begin{aligned} \bar{y}_{.i} &= \delta + z_i + \bar{e}_{.i} \\ \bar{y}_{..} &= \delta + \bar{z} + \bar{e}_{..} \end{aligned}$$

where $\bar{e}_{.i}$, $\bar{e}_{..}$ and $\bar{z}$ are respectively the mean error of the observations of subject i, the global mean error of all observations, and the mean error of the effect.
In order to obtain a distribution theory on which to base classical statistical analysis, we now add the normality assumption to the errors. Writing $g_i = z_i + \bar{e}_{.i}$ we have

$$SS_a = v\sum_{i=1}^{n}(g_i - \bar{g}_.)^2$$

and the random variables $\{g_i\}$ are independently $N(0, \sigma_g^2)$ distributed, where $\sigma_g^2 = \sigma_b^2 + v^{-1}\sigma_b^2$. Therefore in the null hypothesis $\sum_i(g_i - \bar{g}_.)^2/\sigma_g^2$ behaves as a central Chi-square variable with $n-1$ degrees of freedom, and hence

$$SS_b = v\sigma_g^2\chi^2_{n-1}.$$

In the same way, from assumption (A.3) we can derive the distribution of SS_w, which is $\sigma_w^2\chi^2_{n(v-1)}$. Therefore, we have the following expectations of the sample mean squares:

$$\begin{aligned} E\left[\frac{SS_b}{n-1}\right] &= v\sigma_b^2 + \sigma_w^2 \\ E\left[\frac{SS_w}{n(v-1)}\right] &= \sigma_w^2. \end{aligned}$$

We replace σ_b^2 and σ_w^2 in the above equations with observed values $\hat{\sigma}_b^2$ and $\hat{\sigma}_w^2$, equate the resulting expression to $SS_b/(n-1)$ and $SS_w/n(v-1)$, and solve for $\hat{\sigma}_b^2$ and $\hat{\sigma}_w^2$ to get

$$\begin{aligned}
\hat{\sigma}_b^2 &= v^{-1}\left(\frac{SS_b}{n-1} - \frac{SS_w}{n(v-1)}\right) \\
\hat{\sigma}_w^2 &= \frac{SS_w}{n(v-1)}.
\end{aligned}$$

Clearly the traditional estimate of σ_b^2 can sometimes be negative. Should this occur, we do not believe that any such statistical analysis would be useful until a decision is made as to what to do with the negative estimate. This is an example of what is known in the literature as "the problem of negative estimates of variance components" (Nelder, 1954; Thompson, 1962). Two possible explanations for a negative estimate present themselves: (1) the assumed model may be incorrect and (2) noise may have obscured the underlying physical situation. The literature generally treats test on variance, neglecting those on the model's random effect. The problem is clearly present, for example, in epidemiology (Davies et al., 1994; Khoury et al., 1988) but we have not found acceptable inferential solutions in the literature. This dissertation proposes a test designed for testing the presence or absence of this random effect in a nonparametric permutation way, so as to avoid the problem of estimating the variance components.

2.3 Multi-sided permutation test

2.3.1 Introduction

Our interest is focused on the analysis of the random effect Δ. By working in a nonparametric permutation framework the problem related to estimating the variance σ_b^2 becomes irrelevant, because no "standardization" is required to derive the reference null distribution when testing for main effects. The assumption of normality for errors z_i in models (2.1) and (2.2) implies they can assume both negative and positive values. This possible alternation of values, which occurs with distributions whose support is $\mathbb{R}$, implies that some individuals may have positive effects and others negative. If we wish to test the presence or absence of two kinds of effects it seems natural to use a test in which the null hypothesis and the alternative are

$$H_0 : \left\{\Delta \stackrel{d}{=} 0\right\} \qquad H_1 : \left\{\Delta \stackrel{d}{\neq} 0\right\}.$$

Traditional two-sided test requires that the alternative $(\Delta \overset{d}{\neq} 0)$ is either $(\Delta \overset{d}{>} 0)$ or $(\Delta \overset{d}{<} 0)$, but not both. So in the presence of random effects, this kind of alternative and related test are not appropriate. Instead we need testing for alternatives such that $\Pr\{\Delta < 0\} > 0$, $\Pr\{\Delta > 0\} > 0$ and $\Pr\{\Delta < 0\} + \Pr\{\Delta > 0\} = 1$ so both sides can be active although in different subjects. More generally, by expressing relationships in terms of cumulative distribution functions (c.d.f.s), we want to perform a goodness-of-fit test

$$H_0 : \{F_1 = F_2\}$$

in which, in the alternative hypothesis, both stochastic dominance $(F_1 \leq F_2)$ and $(F_1 \geq F_2)$ can jointly hold on separate sets of subjects. We call this kind of test *multi-sided test* and denote the alternative hypothesis with

$$H_{1M} : \{F_1 \neq F_2\} = \{[F_1 \leq F_2] \cup [F_1 \geq F_2]\},$$

where the union symbol $\cup$ means that one or both of two events can be satisfied. In Figure 2.2 the traditional two-sided alternative hypothesis is represented. The solid lines are the alternatives $(F_1 \leq F_2)$ and $(F_1 \geq F_2)$, of which only one is active. In Figure 2.3 the multi-sided H_{1M} hypothesis is represented, where in the alternative (solid line) both $(F_1 \leq F_2)$ and $(F_1 \geq F_2)$ are jointly active.

Within the traditional statistical methodology it is difficult to test this type

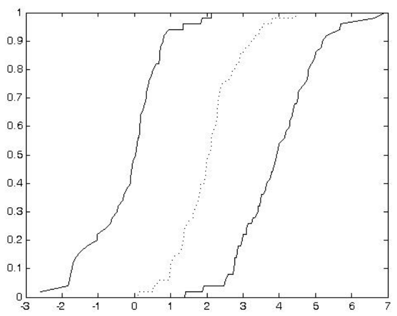

Figure 2.2: H_1 hypothesis

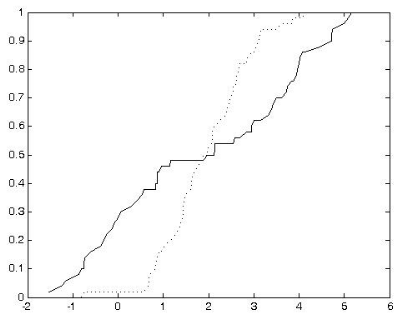

Figure 2.3: H_{1M} hypothesis

of hypothesis. However, it can easily be tested within the nonparametric combination of dependent permutation tests (NPC) framework.

Testing analysis by NPC methods requires that a problem can be broken down into a set of simpler sub-problems, for each of which a partial permutation test is available, and that these partial tests can be jointly processed.

Therefore, two different tests are to be applied to the same data. Of course, each partial test shall be appropriate for one kind of deviation from H_0. Moreover, two partial test statistics are not independent, since they are calculated on the same data set and so some kind of dependence is generally present. This dependence is extremely difficult to model, analyze and take into account explicitly. In the context of the NPC approach, it is particularly worth noting that researchers are not required to model an estimate dependence coefficients among variables and or partial tests because, due to conditioning on a set of sufficient statistics for F, NPC methods are also nonparametric with respect to these underling coefficients (Pesarin, 2001). Permutation methods are known to be conditional inferential procedures in which conditioning is made on a set of sufficient statistics in the null hypothesis for the underling and usually unknown population distribution F. In the next section we will discuss about the set of sufficient statistics and the concept of exchangeability.

2.3.2 Sufficient statistics, exchangeability and similarity

For all problems of practical interest (since not any sequence of numbers is a sample useful for statistical analyses!) the set of sufficient statistics in the null hypothesis is the observed data set for whatever underlying distribution. Let P be the underlying probability measure for the problem, $f_P(\mathbf{x})$ the corresponding density with respect to a suitable dominating measure μ of the sampling variable $\mathbf{X}$ which takes values on a sample space $\mathcal{X}$, and $\mathbf{x} \in \mathcal{X}$ a realization of $\mathbf{X}$, i.e. the observed data set. By sufficiency, given a sample point $\mathbf{x}$, if $\mathbf{x}^* \in \mathcal{X}$ and $\mathbf{x} \neq \mathbf{x}^*$ is such that the likelihood ratio $f_P(\mathbf{x})/f_P(\mathbf{x}^*) = \rho(\mathbf{x}, \mathbf{x}^*)$ is not dependent on f_P for whatever $P \in \mathcal{P}$, where $\mathcal{P}$ is a nonparametric family of non-degenerate distributions, then $\mathbf{x}$ and $\mathbf{x}^*$ are said to contain the same amount of information with respect to P. So that they are equivalent for inferential purposes. The set of point which are equivalent to $\mathbf{x}$, with respect to contained information, is called the orbit associated with $\mathbf{x}$ and is denoted by $\mathcal{X}_{/\mathbf{x}}$ so that $\mathcal{X}_{/\mathbf{x}} = \{\mathbf{x}^* : \rho(\mathbf{x}, \mathbf{x}^*) \; is f_P - independent\}$. The same conclusion is obtained if $f_P(\mathbf{x})$ is assumed to be invariant with respect to permutations of the arguments of $\mathbf{x}$, i.e. the elements $(x_1, \ldots, x_n)$. This happens when the assumption of independence for observable data is replaced by that of exchangeability:

$$f_P(x_1, \ldots, x_n) = f_P(x_{u_1^*}, \ldots, x_{u_n^*})$$

where $(u_1^*, \ldots, u_n^*)$ is any permutation of $(1, \ldots, n)$. In the context of permutation tests, this concept of exchangeability is often referred to as the

exchangeability of the observed data with respect to groups if H_0 is true. Orbits $\mathcal{X}_{/\mathbf{x}}$ are also called permutation sample spaces. It is important to note that orbits $\mathcal{X}_{/\mathbf{x}}$ associated with a data set $\mathbf{x} \in \mathcal{X}$ always contain a finite number of points, as sample size is finite.
Permutation tests are conditional statistical procedures, where conditioning is with respect to the orbit $\mathcal{X}_{/\mathbf{x}}$ associated with the observed data set $\mathbf{x}$. In this way, in the null hypothesis and assuming exchangeability, the conditional probability distribution of a generic point $\mathbf{x}' \in \mathcal{X}_{/\mathbf{x}}$, for whatever underlying population distribution $P \in \mathcal{P}$, is

$$\begin{aligned}\Pr\left\{\mathbf{x}^* = \mathbf{x}'|\mathcal{X}_{/\mathbf{x}}\right\} &= \frac{\sum_{\mathbf{x}^*=\mathbf{x}'} f_P(\mathbf{x}^*)\cdot d\mu}{\sum_{\mathbf{x}^*\in\mathcal{X}_{/\mathbf{x}}} f_P(\mathbf{x}^*)\cdot d\mu} \\ &= \frac{\#\left[\mathbf{x}^* = \mathbf{x}', \mathbf{x}^* \in \mathcal{X}_{/\mathbf{x}}\right]}{\#\left[\mathbf{x}^* \in \mathcal{X}_{/\mathbf{x}}\right]},\end{aligned}$$

which is P-independent. Of course, if there is only one point in $\mathcal{X}_{/\mathbf{x}}$ whose coordinates coincide with those of $\mathbf{x}'$, i.e. if there are no ties in the data set, this conditional probability becomes $1/n!$. Thus, $\Pr\left\{\mathbf{x}^* = \mathbf{x}'|\mathcal{X}_{/\mathbf{x}}\right\}$ is uniform on $\mathcal{X}_{/\mathbf{x}}$ for all $P \in \mathcal{P}$. These statements allow permutation inferences to be invariant with respect to P in H_0. Due to this invariance property, permutation tests are distribution-free and nonparametric. Another important property due to invariance is that the permutation tests enjoy the strong similarity property (Lehmann, 1986) in the sense that, for all distribution $P \in \mathcal{P}$, the conditional α-size of the tests are $\mathbf{X}$-invariant. This property means that if data come from continuous distributions, so that the probability of finding ties in the data set is zero, the rejection probability in H_0 is invariant with respect to observed data set $\mathbf{x}$, for almost all $\mathbf{x} \in \mathcal{X}$. Thus, conditional rejection regions are similar to the unconditional regions. When data come from non-continuous distributions, the similarity property is only asymptotically valid. Formally, let T_α be the permutation critical α-value associated to statistic T and data $\mathbf{X}$. Since T_α depends on $\mathcal{X}_{/\mathbf{x}}$, the probability of finding a sample point $\mathbf{X}^* \in \mathcal{X}_{/\mathbf{x}}$ such that $T(\mathbf{X}^*) \geq T_\alpha$ is precisely the attainable α-size

$$\begin{aligned}\Pr\left\{\mathbf{X}^* \in \mathcal{X}_{/\mathbf{X}} : T(\mathbf{X}^*) \geq T_\alpha\right\} &= \alpha \\ &= \mathbb{E}_{\mathcal{X}_{/\mathbf{X}}}\Big[\mathbf{I}\left\{\lambda(\mathbf{X}^*) \leq \alpha|\mathbf{X}\right\}\Big]\end{aligned}$$

if and only if H_0 is true whatever the data set $\mathbf{X} \in \mathcal{X}$, where $\mathbf{I}\{A\}$ is the indicator function, i.e. $\mathbf{I}\{A\} = 1$ if A is true, 0 otherwise, and $\lambda(\mathbf{X})$ is the attainable p-value. Moreover, due to invariance property and noting that the relationships $(T \geq T_\alpha) \Leftrightarrow (\lambda \leq \alpha)$ is true by definition, if and only if H_0 is true we have

$$\begin{aligned}
\Pr\{T(\mathbf{X}) \geq T_\alpha(\mathbf{X})|H_0\} &= \mathbb{E}_{\mathcal{X}\backslash\mathcal{X}_{/\mathbf{X}}}\Big\{\mathbb{E}_{\mathcal{X}_{/\mathbf{X}}}\Big[\mathbf{I}\{\lambda(\mathbf{X}^*) \leq \alpha|\mathbf{X}, H_0\}\Big]\Big\} \\
&= \mathbb{E}_{\mathcal{X}}\Big[\mathbf{I}\{\lambda(\mathbf{X}^*) \leq \alpha|H_0\}\Big] \\
&= \int_{\mathcal{X}} \mathbf{I}\{\lambda(\mathbf{X}^*) \leq \alpha\} f_P(\mathbf{X})d\nu(\mathbf{X}) = \alpha
\end{aligned}$$

where $\mathcal{X}\backslash\mathcal{X}_{/\mathbf{X}}$ represents the partition set induced on the sample space $\mathcal{X}$ by conditioning with respect to the sample point $\mathbf{X}$, so that if $\mathbf{X}$ and $\mathbf{X}'$ are two distinct points of $\mathcal{X}\backslash\mathcal{X}_{/\mathbf{X}}$, then $\mathcal{X}_{/\mathbf{X}}$ and $\mathcal{X}_{/\mathbf{X}'}$ are distinct, i.e. the intersection of the two orbits $\mathcal{X}_{/\mathbf{X}}$ and $\mathcal{X}_{/\mathbf{X}'}$ is empty. In the last equality the cardinality of $\mathcal{X}_{/\mathbf{X}}$ is considered to be $\mathbf{X}$-invariant. The unconditional statement suggests that, for a permutation test with data from continuous variables, the attainable α-size is similar for any underlying distribution $P \in \mathcal{P}$ provided that, in H_0, the exchangeability of error components is satisfied.

2.3.3 The statistics

To test the two sub-hypotheses we use the Anderson-Darling type test statistics. To this end let us use $\mathbf{X}^*$ to denote a random permutation of pooled data set $\mathbf{X}$. This is obtained as $\mathbf{X}^* = \{X(u_i^*), i = 1, \ldots, n; n_1, n_2\}$, where $(u_1^*, \ldots, u_n^*)$ is a permutation of $(1, \ldots, n)$. Two partial tests are

$$T_1^* = \sum_{i=1}^{n} S\Big\{F_1^*(X_i) - F_2^*(X_i)\Big\}\left(\hat{F}(X_i)\left[1 - \hat{F}(X_i)\right]\right)^{-1/2} \tag{2.4}$$

to test the sub-hypothesis $H_{11} : \{F_1 \geq F_2\}$ and

$$T_2^* = \sum_{i=1}^{n} S\Big\{F_2^*(X_i) - F_1^*(X_i)\Big\}\left(\hat{F}(X_i)\left[1 - \hat{F}(X_i)\right]\right)^{-1/2} \tag{2.5}$$

to test the sub-hypothesis $H_{12} : \{F_1 \leq F_2\}$ where

$$S\{\omega\} = \begin{cases} \omega & \text{if } \omega > 0 \\ 0 & \text{if } \omega \leq 0, \end{cases}$$

and $F_j^*(t)$, $j = 1, 2$, are the normalized empirical distribution functions on permuted samples (Brunner et al., 1995; Ruymgaart, 1980) given by

$$F_j^*(t) = \left[\#(X_{ji}^* < t) + \frac{1}{2}\#(X_{ji}^* = t)\right]/n_j$$

and $\hat{F}(t) = [n_1 F_1(t) + n_2 F_2(t)]/n$. We use the normalized empirical distribution functions because they are especially useful for discrete cases. In the set of units in which the sub-alternative $\Delta \overset{d}{<} 0$ is active, where $\hat{F}_1(t) \geq \hat{F}_2(t)$, T_1 is unbiased and consistent. Correspondingly, T_2 is unbiased and consistent for the sub-alternative $\Delta \overset{d}{>} 0$. In the following section we demonstrate the unbiasedness of tests T_1 and T_2 and in Section 2.3.6 we will see how to combine the two partial tests in order to obtain a global test.

2.3.4 Exactness and Unbiasedness

As regards the exactness property of two separate tests, let us argue on T_1^* and extend the same conclusions to T_2^*. To this end let us observe that: a) a permutation test is called exact if its null distribution depends only on exchangeable random errors; b) exactness is intended with respect to attainable α-values; c) the number of positive summands in one permutation of T_1^*, i.e. $\nu^* = \sum_i \mathbf{I}[S(F_1^* - F_2^*) > 0]$, is not invariant over the permutation sample space $\mathcal{X}_{/\mathbf{X}}$; d) then also the conditional p-value $\Pr\{T_1^* \geq T_1^o | \mathbf{X}, \nu^*\}$, where it is emphasized that the latter depends on the subset of permutations sharing the same number of summands ν^*, is not invariant over $\mathcal{X}_{/\mathbf{X}}$ as well; e) as a consequence the related attainable p-value becomes $\lambda(\mathbf{X}) = \sum_{\nu^*} \Pr\{T_1^* \geq T_1^o | \mathbf{X}, \nu^*\} \Pr(\nu^* | \mathbf{X})$, which then is a mixture of non-invariant permutation quantities.
This implies that attainable α-values $\Lambda(\mathbf{X})$ are not $\mathbf{X}$-invariant quantities even when the observed variable is continuous or there are no ties in $\mathbf{X}$. Thus, test T_1^*, which in the null hypothesis depends only on exchangeable random errors, is an exact test at its attainable α-values, which in turn depend on $\mathbf{X}$. And so $T_j^*, j = 1, 2$, satisfy the similarity property only asymptotically. Of course, due to this, in a simulation study we cannot expect to exactly obtain the "desired nominal" α-values in the null hypothesis. In this respect, reported simulations in section 2.4 show that T_j^*, $j = 1, 2$, behave as if they were approximate. Their apparent approximations are mostly due to the non-invariant property on $\mathbf{X}$ of attainable α-values $\Lambda(\mathbf{X})$. From simulation results reported below this approximation appears to be quite accurate even for small sample sizes and unbalanced situations. This may be due to the fact that, as is well known, the null distribution of Anderson–Darling statistic is practically invariant over sample sizes ν^*, so that $\Lambda(\mathbf{X})$ for fixed sample sizes is an "almost" $\mathbf{X}$-invariant set, i.e. T_j^*, $j = 1, 2$, are almost similar.

To show the unbiasedness of multi-sided test we assume the exchangeability of errors in H_0. We employ the pointwise representation of elements of sample space $\mathcal{X}$ in H_0 and H_1. To this end, for any given set of units we consider the associated sample points in $\mathcal{X}$ are denoted by $\mathbf{X}(0)$ in H_0 and by $\mathbf{X}(\Delta)$ in H_1, where $\mathbf{X(0)} = \mathbf{X_1(0)} \uplus \mathbf{X_2(0)}$ and $\mathbf{X(\Delta)} = \big(\mathbf{X_1(0)} + \mathbf{\Delta_1}\big) \uplus \big(\mathbf{X_2(0)} + \mathbf{\Delta_2}\big)$, in which

$$\mathbf{\Delta} = \mathbf{\Delta_1} \uplus \mathbf{\Delta_2} = \{\Delta_{1i} \sim F_\Delta, i = 1, \ldots, n_1\} \uplus \Big\{\Delta_{2i} \overset{d}{=} 0, i = 1, \ldots, n_2\Big\}$$

represent the pooled vector of stochastic effects Δ_{ji} with known or unknown c.d.f. F_Δ. First we show the unbiasedness of test T_1, so we consider only the negative part of effect Δ:

$$\Delta_{1i}^- = \begin{cases} \Delta_{1i} & \text{if } \Delta_{1i} \overset{d}{<} 0 \\ 0 & \text{if } \Delta_{1i} \overset{d}{>} 0, \end{cases}$$

the proof for T_2 is similar. In this context, the observed value of T_1 in H_1 is

$$T_1^0(\Delta) = \sum_{i=1}^{n} \frac{S\Big\{F_1\big(X_i(0) + \Delta_{1i}^-\big) - F_2\big(X_i(0) + \Delta_{2i}^-\big)\Big\}}{\Big(\hat{F}\big(X_i(\Delta_i^-)\big)\Big[1 - \hat{F}\big(X_i(\Delta_i^-)\big)\Big]^{1/2}}$$

To study the unbiasedness we analyze only the numerator of the statistic as the denominator is permutationally invariant. As Δ_{1i}^- decreases, some more summands in the sum become positive and the value of the positive summands do not decreases. So $T_1^0(\Delta) = T_1^0(0) + \tau$ where τ is a non-negative quantity. In order to compare the permutation structures of T_1 in H_0 and in H_1, we consider one generic permutation $(u_1^*, \ldots, u_n^*)$ of unit labels $(1, \ldots, n)$. Therefore, the associated values of T_1^* are $T_1^*(0)$ and $T_1^*(\Delta) = T_1^*(0) + \tau^*$, where τ^* is still non-negative since the element u_i^* in the sum of equation (2.4) give the same influence to T_1 as before. τ^* is much greater when more units with negative effect are assigned on the first sample from the random permutation. Clearly τ^* can not be larger than τ since all the units under H_1 are in the sample 1 of the observed data. For any generic permutation we have

$$\begin{aligned} \Pr\big\{T_1^*(\Delta) \geq T_1^0(\Delta)|\mathbf{X(\Delta)}\big\} &= \Pr\big\{T_1^*(0) + \tau^* \geq T_1^0(0) + \tau|\mathbf{X(0)}\big\} \\ &= \Pr\big\{T_1^*(0) + \tau^* - \tau \geq T_1^0(0)|\mathbf{X(0)}\big\} \\ &\leq \Pr\big\{T_1^*(0) \geq T_1^0(0)|\mathbf{X(0)}\big\} \end{aligned}$$

where the weak inequality holds since $\tau^* = \tau$ only for the observed sample and $\tau^* < \tau$ for all other permutations. This give rise to a pointwise dominance of $T_1^*(\Delta)$ with respect to $T_1^*(0)$ and proves the conditional unbiasedness of T_1 for all data sets $\mathbf{X}$. The unconditional unbiasedness for all sampling experiments and all underlying population distributions P is obtained by the similarity property. Similar results hold for T_2.

2.3.5 Consistency

The proof of consistency is easy considering the finite sample consistency properties of permutation tests, so we postpone the proof to the next chapter.

2.3.6 Combination of partial tests

In order to obtain an overall solution, one way is to properly combine the two partial tests T_1 and T_2. Of course, these partial tests and associated p-values are dependent in a way that in general is extremely difficult to take into account explicitly. Consequently, when considering their combination, we take account of such underlying dependence relations nonparametrically; hence we must work within the NPC approach. Therefore, to test H_0 against H_{1M} we need to combine the p-values λ_1 and λ_1 of the two partial tests by a non degenerate and measurable combining function $\psi : [0,1]^2 \to \mathbb{R}$. Of many functions, those which are appropriate for combination testing must at least satisfy the following mild properties:

1. every combining function ψ must be non increasing in each argument:

$$\psi(\lambda_1, \lambda_2) \geq \psi(\lambda_1, \lambda_2')$$
$$\psi(\lambda_1, \lambda_2) \geq \psi(\lambda_1', \lambda_2)$$

   if at least one $\lambda_j < \lambda_j'$, $j = 1, 2$;

2. every ψ must attain its supremum value $\tilde{\psi}$, possibly non finite, when even one argument attains zero: $\psi(\lambda_1, \lambda_2) \to \tilde{\psi}$ if $\lambda_j \to 0, j \in (1, 2)$;

3. $\forall \alpha > 0$, the critical value of every ψ is assumed to be finite and strictly smaller than the supremum value: $\psi_\alpha < \tilde{\psi}$.

Properties 1, 2 and 3, are generally easy to check and justify. In (Pesarin, 2001) it is proved that: (i) if the partial permutation tests are exact, then the combined test $T_\psi = \psi(\lambda_1, \lambda_2)$ is exact; (ii) if all partial permutation tests are marginally (i.e. separately) unbiased, then T_ψ is unbiased; (iii) if both partial tests are marginally unbiased and at least one is consistent, then T_ψ is consistent. Of the many combining functions ψ that satisfy properties 1, 2 and 3 those mostly often used are :

- Fisher: $T_F = -2\sum_j \log(\lambda_j)$;
- Liptak: $T_T = \sum_j \Phi^{-1}(1 - \lambda_j)$;

- Tippet: $T_L = \max_j(1 - \lambda_j)$;

where Φ is the standard normal cumulative distribution function. In the framework of permutation tests we do not require the assumptions (A.1)-(A.4) in section 2.1. All that is required is that data are exchangeable between groups in H_0.

2.4 Simulation study

In our simulation study we consider a simplified version of models (2.1) and (2.2). We consider a univariate model. In the first step we consider the model

$$
\begin{aligned}
y_i &= \mu + \Delta_i + e_i \\
\Delta_i &= \delta \eta_i,
\end{aligned}
$$

where η_i is a random variable that assumes the value -1 and 1, each with probability 1/2, and δ is a fixed effect. We generate a first sample for any $\delta \in \{0.1, 0.2, 0.5, 1, 2\}$ and a second sample with $\delta = 0$. Of course, as both test statistics are invariant on the nuisance population quantity μ, on all simulations we set $\mu = 0$. Different distributions are chosen for random errors e_i. The chosen distributions are: Normal (0,1), Chi square with 3 degrees of freedom, Student's t with one degree of freedom, Exponential(1), the Skew normal (Azzalini, 1999) with location parameter 0, scale parameter 1 and shape parameter equal to 5, and a mixture of Binomial distributions. We consider the sample sizes $n = 5, 10, 15$ and all possible combinations of n for each of the two samples. Hence we also considered unbalanced samples. We replicated the study with 1000 Monte Carlo simulations and considered 1000 samples from the permutation sample space. In some cases, when the permutation sample space was less than 1000 points, for example when the sample size was $n_1 = n_2 = 5$, we performed the exact test. We calculated the power of the test for $\alpha = 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.7, 0.8$. For each of the 1000 Monte Carlo Simulation tests (2.4) and (2.5) were performed and the two p-values were combined with the three functions, Fisher, Liptak and Tippet, to obtain three global tests. Figure 2.4 reports the power of the partial and the global tests obtained using Fisher's combining function, with $n_1 = n_2 = 10$ and $\delta = 1$ for different values of α, and normal errors. Figures 2.5 and 2.6, are as above but using the Liptak and Tippet combining function respectively. Figure 2.7 reports the three tests in terms of δ. Very similar results are obtained with the other considered distributions. In the simulation's second step, we considered a univariate version of models (2.1)

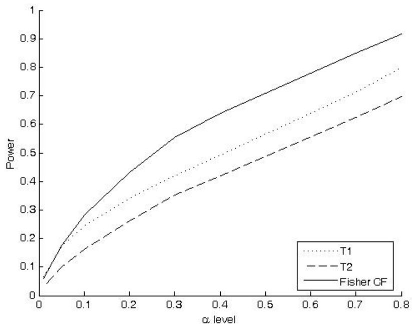

Figure 2.4: Fisher's combining function

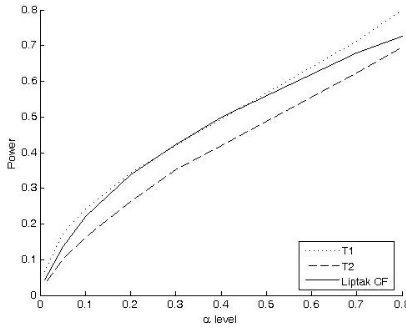

Figure 2.5: Liptak's combining function

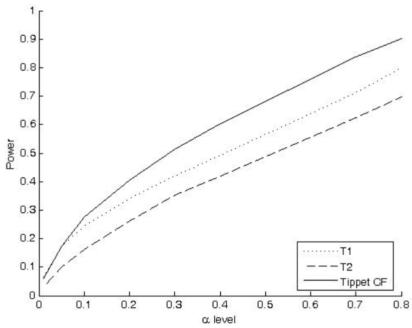

Figure 2.6: Tippet's combining function

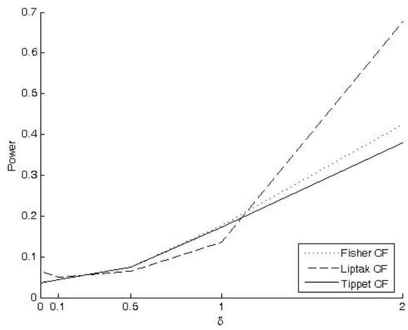

Figure 2.7: The three global tests

and (2.2)

$$y_i = \Delta_i + e_i \quad (2.6)$$
$$\Delta_i = z_i \eta_i, \quad (2.7)$$

where e_i has the same meaning as before and $z_i \sim U(0, \delta)$ so that $\Delta \sim U(-\delta, \delta)$. In this model, effect Δ_i is random in the absolute value and also in the sign. In this simulation $\delta = 10$. Figure 2.8 reports the plot of the power of T_1, T_2, and the global test obtained with Fisher's combining function, $e_i = \beta Y_1 + (1-\beta)Y_2$, where $\beta = 0.25$, $Y_1 \sim \text{Bin}(5, 0.5)$, and $Y_2 \sim \text{Bin}(3, 0.2)$. Tables 2.1-2.12 show the estimated power of the partial and global tests for the various values of α, with different distributions of errors. We can see that the power of the test increases with the size-effect even if the effect has a random sign. This happens for both partial tests and for the global.

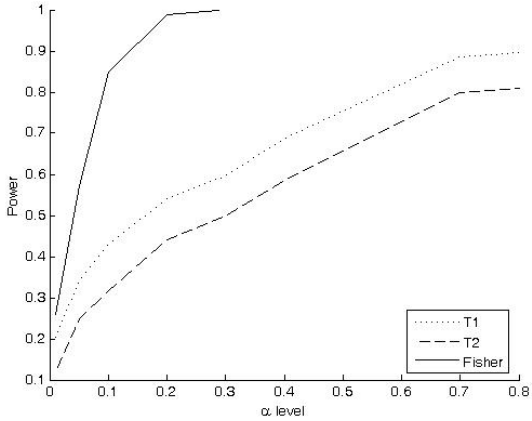

Figure 2.8: Testing H_0 against H_{1M} with $z_i \sim U(0, \delta)$

2.5 Multivariate extension of the test

In our simulations we considered only the univariate version of models (2.1) and (2.2). The extension to multi-sample and multivariate versions is quite easy in the framework of multivariate permutation tests and NPC of dependent partial tests. For instance, in the two-sample multivariate case, we can break down the global, multivariate hypothesis about the presence of random effects

$$H_0 : \{F_1(\mathbf{X}) = F_2(\mathbf{X}), \mathbf{X} \in \mathbb{R}^v\}$$

into v sub hypotheses

$$H_0 : \left[\{F_1(X_1) = F_2(X_1)\} \ldots \cap \ldots \{F_1(X_v) = F_2(X_v)\} \right] = \cap_{h=1}^{v} H_{0h}$$

where $X_h \in \mathbb{R}, h = 1, \ldots, v$, considering for each sub-hypothesis H_{0h} the alternative H_{1Mh}, thus computing v (global)partial tests as before, and proceeding, with their nonparametric combination in a third step.

2.6 Conclusions

The test proposed offers the opportunity to test the presence of random effects due, for example, to medical treatments, industrial process, particular economic policy etc. The usual two-sided tests in the presence of perfectly balanced random effects would give zero power so they would not provide acceptable results. Application of the multi-sided test in the nonparametric framework is easy as it is a particular form of multi-aspect testing (Salmaso, 2005). Once the global null hypothesis is rejected, it is straightforward to proceed with correction for multiplicity to check which of the two, if not both, tails is actually active.

α	Type	$\delta=0$	$\delta=0.1$	$\delta=0.5$	$\delta=1$	$\delta=2$	$z_i \sim U(0,\delta)$
0.01	T_1	0.004	0.011	0.022	0.060	0.170	0.178
	T_2	0.007	0.014	0.011	0.006	0.004	0.014
	T_F	0.005	0.009	0.013	0.038	0.138	0.161
	T_L	0.009	0.003	0.011	0.011	0.096	0.278
	T_T	0.005	0.009	0.013	0.038	0.138	0.162
0.05	T_1	0.042	0.053	0.080	0.147	0.300	0.340
	T_2	0.047	0.041	0.056	0.052	0.082	0.120
	T_F	0.038	0.041	0.066	0.114	0.237	0.262
	T_L	0.045	0.038	0.053	0.061	0.338	0.560
	T_T	0.040	0.042	0.064	0.121	0.227	0.216
0.10	T_1	0.091	0.123	0.142	0.249	0.366	0.391
	T_2	0.090	0.083	0.096	0.098	0.144	0.170
	T_F	0.089	0.093	0.131	0.200	0.409	0.504
	T_L	0.080	0.088	0.103	0.117	0.447	0.657
	T_T	0.089	0.094	0.136	0.199	0.382	0.460
0.20	T_1	0.176	0.219	0.261	0.361	0.488	0.544
	T_2	0.212	0.181	0.185	0.200	0.263	0.312
	T_F	0.177	0.208	0.237	0.335	0.605	0.782
	T_L	0.189	0.177	0.190	0.220	0.565	0.726
	T_T	0.181	0.206	0.238	0.347	0.510	0.561
0.30	T_1	0.263	0.320	0.362	0.458	0.565	0.587
	T_2	0.312	0.289	0.265	0.261	0.329	0.362
	T_F	0.277	0.305	0.337	0.460	0.774	0.896
	T_L	0.281	0.297	0.284	0.312	0.633	0.738
	T_T	0.275	0.307	0.340	0.454	0.586	0.614
0.40	T_1	0.359	0.420	0.474	0.535	0.625	0.649
	T_2	0.434	0.393	0.348	0.330	0.384	0.431
	T_F	0.371	0.409	0.444	0.557	0.851	0.962
	T_L	0.385	0.389	0.360	0.396	0.664	0.752
	T_T	0.388	0.400	0.446	0.561	0.751	0.856
0.70	T_1	0.671	0.713	0.724	0.753	0.798	0.817
	T_2	0.722	0.664	0.624	0.554	0.579	0.609
	T_F	0.686	0.703	0.725	0.793	0.969	0.997
	T_L	0.701	0.685	0.645	0.634	0.720	0.756
	T_T	0.675	0.709	0.723	0.792	0.933	0.978
0.80	T_1	0.766	0.809	0.813	0.846	0.919	0.958
	T_2	0.810	0.767	0.727	0.662	0.718	0.777
	T_F	0.787	0.805	0.810	0.880	0.983	0.999
	T_L	0.795	0.785	0.743	0.696	0.724	0.757
	T_T	0.795	0.805	0.818	0.873	0.974	0.995

Table 2.1: Power of the test, $n_1 = 5$, $n_2 = 10$, $N(0,1)$ distribution

α	Type	$\delta = 0$	$\delta = 0.1$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$	$z_i \sim U(0, \delta)$
0.01	T_1	0.005	0.005	0.022	0.066	0.180	0.203
	T_2	0.007	0.011	0.014	0.031	0.080	0.097
	T_F	0.004	0.009	0.017	0.060	0.188	0.233
	T_L	0.011	0.006	0.013	0.043	0.466	0.723
	T_T	0.004	0.009	0.017	0.060	0.185	0.225
0.05	T_1	0.047	0.038	0.089	0.174	0.288	0.329
	T_2	0.037	0.055	0.054	0.101	0.202	0.238
	T_F	0.037	0.045	0.075	0.177	0.424	0.516
	T_L	0.064	0.051	0.064	0.137	0.676	0.856
	T_T	0.036	0.043	0.074	0.173	0.380	0.409
0.10	T_1	0.090	0.092	0.145	0.243	0.366	0.396
	T_2	0.109	0.108	0.099	0.162	0.269	0.303
	T_F	0.089	0.101	0.135	0.281	0.611	0.766
	T_L	0.123	0.100	0.129	0.220	0.741	0.885
	T_T	0.084	0.093	0.143	0.275	0.490	0.567
0.20	T_1	0.196	0.190	0.263	0.342	0.467	0.506
	T_2	0.220	0.206	0.200	0.262	0.375	0.409
	T_F	0.198	0.201	0.247	0.430	0.847	0.944
	T_L	0.226	0.208	0.217	0.339	0.800	0.894
	T_T	0.199	0.200	0.244	0.405	0.635	0.699
0.30	T_1	0.302	0.292	0.358	0.419	0.544	0.594
	T_2	0.317	0.315	0.304	0.353	0.451	0.488
	T_F	0.320	0.288	0.354	0.553	0.920	0.986
	T_L	0.308	0.311	0.312	0.422	0.814	0.896
	T_T	0.300	0.301	0.346	0.513	0.760	0.801
0.40	T_1	0.386	0.384	0.441	0.494	0.633	0.662
	T_2	0.421	0.416	0.388	0.421	0.535	0.576
	T_F	0.403	0.402	0.448	0.640	0.961	0.996
	T_L	0.411	0.402	0.395	0.499	0.820	0.896
	T_T	0.416	0.396	0.463	0.604	0.841	0.905
0.70	T_1	0.692	0.682	0.721	0.712	0.832	0.886
	T_2	0.699	0.717	0.640	0.624	0.739	0.778
	T_F	0.699	0.703	0.731	0.850	0.997	0.999
	T_L	0.697	0.699	0.682	0.680	0.834	0.897
	T_T	0.708	0.715	0.749	0.838	0.983	0.999
0.80	T_1	0.779	0.781	0.801	0.799	0.871	0.898
	T_2	0.799	0.814	0.742	0.697	0.773	0.793
	T_F	0.789	0.793	0.820	0.917	0.999	0.999
	T_L	0.796	0.807	0.764	0.727	0.834	0.897
	T_T	0.803	0.803	0.828	0.902	0.996	0.999

Table 2.2: Power of the test, $n_1 = 10$, $n_2 = 10$, $N(0,1)$ distribution

α	Type	$\delta = 0$	$\delta = 0.1$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$	$z_i \sim U(0, \delta)$
0.01	T_1	0.006	0.011	0.064	0.123	0.187	0.182
	T_2	0.007	0.013	0.014	0.011	0.004	0.006
	T_F	0.008	0.013	0.054	0.108	0.184	0.173
	T_L	0.008	0.005	0.017	0.086	0.395	0.354
	T_T	0.008	0.013	0.055	0.108	0.184	0.173
0.05	T_1	0.051	0.057	0.141	0.262	0.367	0.360
	T_2	0.047	0.045	0.053	0.064	0.144	0.138
	T_F	0.048	0.048	0.130	0.203	0.282	0.279
	T_L	0.042	0.038	0.082	0.291	0.681	0.657
	T_T	0.048	0.052	0.129	0.191	0.200	0.212
0.10	T_1	0.099	0.107	0.219	0.317	0.383	0.379
	T_2	0.103	0.088	0.101	0.129	0.177	0.186
	T_F	0.099	0.102	0.203	0.352	0.529	0.525
	T_L	0.107	0.084	0.155	0.453	0.755	0.717
	T_T	0.098	0.102	0.194	0.326	0.511	0.498
0.20	T_1	0.188	0.211	0.352	0.478	0.562	0.552
	T_2	0.191	0.194	0.182	0.246	0.368	0.352
	T_F	0.199	0.189	0.322	0.535	0.878	0.855
	T_L	0.201	0.188	0.280	0.587	0.787	0.767
	T_T	0.202	0.195	0.320	0.446	0.560	0.565
0.30	T_1	0.307	0.323	0.445	0.530	0.569	0.566
	T_2	0.278	0.291	0.254	0.321	0.402	0.401
	T_F	0.302	0.293	0.445	0.707	0.958	0.943
	T_L	0.296	0.273	0.379	0.650	0.790	0.772
	T_T	0.300	0.298	0.432	0.534	0.598	0.608
0.40	T_1	0.408	0.419	0.546	0.625	0.639	0.622
	T_2	0.388	0.370	0.344	0.408	0.459	0.461
	T_F	0.383	0.379	0.564	0.838	0.988	0.973
	T_L	0.403	0.372	0.465	0.693	0.792	0.777
	T_T	0.379	0.405	0.534	0.724	0.930	0.904
0.70	T_1	0.714	0.705	0.808	0.832	0.846	0.817
	T_2	0.677	0.679	0.569	0.595	0.628	0.625
	T_F	0.691	0.708	0.829	0.973	0.998	0.998
	T_L	0.700	0.696	0.681	0.743	0.796	0.782
	T_T	0.684	0.709	0.791	0.918	0.992	0.981
0.80	T_1	0.804	0.804	0.893	0.959	0.996	0.988
	T_2	0.797	0.781	0.677	0.704	0.770	0.782
	T_F	0.804	0.797	0.899	0.985	0.999	0.999
	T_L	0.797	0.793	0.725	0.753	0.796	0.782
	T_T	0.802	0.794	0.883	0.984	0.999	0.997

Table 2.3: Power of the test, $n_1 = 5$, $n_2 = 10$, χ^2_3 distribution

α	Type	$\delta = 0$	$\delta = 0.1$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$	$z_i \sim U(0, \delta)$
0.01	T_1	0.010	0.013	0.066	0.160	0.225	0.205
	T_2	0.019	0.013	0.027	0.068	0.123	0.116
	T_F	0.016	0.012	0.070	0.171	0.282	0.269
	T_L	0.012	0.007	0.074	0.376	0.854	0.832
	T_T	0.016	0.012	0.070	0.171	0.271	0.257
0.05	T_1	0.052	0.065	0.172	0.290	0.362	0.352
	T_2	0.056	0.056	0.108	0.190	0.276	0.265
	T_F	0.055	0.069	0.199	0.405	0.618	0.588
	T_L	0.051	0.067	0.218	0.636	0.885	0.873
	T_T	0.055	0.069	0.193	0.346	0.456	0.434
0.10	T_1	0.106	0.118	0.247	0.366	0.428	0.417
	T_2	0.097	0.119	0.159	0.267	0.352	0.349
	T_F	0.101	0.121	0.289	0.569	0.910	0.877
	T_L	0.095	0.121	0.303	0.716	0.891	0.881
	T_T	0.108	0.121	0.280	0.480	0.638	0.617
0.20	T_1	0.208	0.239	0.350	0.477	0.527	0.524
	T_2	0.193	0.211	0.250	0.370	0.461	0.455
	T_F	0.210	0.235	0.463	0.808	0.990	0.988
	T_L	0.198	0.228	0.436	0.781	0.892	0.882
	T_T	0.203	0.237	0.406	0.633	0.780	0.766
0.30	T_1	0.299	0.334	0.444	0.546	0.576	0.568
	T_2	0.288	0.297	0.338	0.452	0.517	0.514
	T_F	0.301	0.339	0.577	0.910	0.998	0.997
	T_L	0.303	0.331	0.531	0.799	0.892	0.882
	T_T	0.311	0.341	0.513	0.737	0.883	0.868
0.40	T_1	0.417	0.439	0.540	0.620	0.673	0.664
	T_2	0.384	0.406	0.405	0.515	0.598	0.592
	T_F	0.397	0.451	0.683	0.959	0.999	1.000
	T_L	0.401	0.431	0.607	0.813	0.892	0.882
	T_T	0.401	0.450	0.600	0.846	0.964	0.965
0.70	T_1	0.714	0.713	0.806	0.850	0.900	0.887
	T_2	0.690	0.661	0.623	0.682	0.791	0.791
	T_F	0.700	0.742	0.894	0.994	1.000	1.000
	T_L	0.713	0.680	0.743	0.828	0.892	0.882
	T_T	0.692	0.735	0.863	0.989	1.000	0.999
0.80	T_1	0.800	0.807	0.850	0.870	0.902	0.894
	T_2	0.788	0.770	0.705	0.755	0.804	0.808
	T_F	0.809	0.826	0.945	0.999	1.000	1.000
	T_L	0.799	0.772	0.771	0.829	0.892	0.882
	T_T	0.804	0.840	0.913	0.996	1.000	1.000

Table 2.4: Power of the test, $n_1 = 10$, $n_2 = 10$, χ^2_3 distribution

α	Type	$\delta = 0$	$\delta = 0.1$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$	$z_i \sim U(0, \delta)$
0.01	T_1	0.011	0.009	0.017	0.024	0.064	0.109
	T_2	0.012	0.005	0.000	0.010	0.014	0.007
	T_F	0.012	0.009	0.008	0.018	0.044	0.094
	T_L	0.010	0.005	0.007	0.017	0.035	0.078
	T_T	0.012	0.009	0.008	0.018	0.044	0.093
0.05	T_1	0.039	0.053	0.075	0.098	0.140	0.238
	T_2	0.051	0.042	0.021	0.048	0.048	0.070
	T_F	0.046	0.046	0.047	0.083	0.127	0.191
	T_L	0.046	0.045	0.040	0.065	0.107	0.237
	T_T	0.046	0.048	0.047	0.083	0.125	0.176
0.10	T_1	0.088	0.105	0.130	0.175	0.211	0.301
	T_2	0.093	0.092	0.075	0.099	0.083	0.128
	T_F	0.089	0.089	0.095	0.150	0.197	0.324
	T_L	0.094	0.096	0.098	0.105	0.184	0.381
	T_T	0.090	0.095	0.096	0.146	0.188	0.308
0.20	T_1	0.187	0.211	0.246	0.295	0.340	0.431
	T_2	0.211	0.207	0.155	0.176	0.167	0.231
	T_F	0.185	0.188	0.202	0.269	0.311	0.492
	T_L	0.191	0.192	0.201	0.214	0.347	0.533
	T_T	0.181	0.197	0.205	0.274	0.294	0.429
0.30	T_1	0.297	0.304	0.360	0.404	0.423	0.507
	T_2	0.307	0.294	0.254	0.250	0.252	0.305
	T_F	0.298	0.288	0.292	0.363	0.429	0.653
	T_L	0.308	0.283	0.303	0.313	0.464	0.610
	T_T	0.290	0.292	0.302	0.388	0.395	0.525
0.40	T_1	0.389	0.395	0.449	0.476	0.526	0.598
	T_2	0.404	0.405	0.357	0.342	0.350	0.396
	T_F	0.388	0.401	0.395	0.470	0.542	0.763
	T_L	0.403	0.376	0.415	0.412	0.558	0.667
	T_T	0.398	0.418	0.401	0.471	0.507	0.662
0.70	T_1	0.679	0.676	0.742	0.740	0.779	0.798
	T_2	0.706	0.697	0.637	0.600	0.627	0.614
	T_F	0.718	0.703	0.714	0.737	0.813	0.915
	T_L	0.724	0.673	0.731	0.697	0.760	0.765
	T_T	0.704	0.686	0.706	0.723	0.782	0.889
0.80	T_1	0.792	0.787	0.846	0.815	0.872	0.881
	T_2	0.806	0.789	0.755	0.707	0.741	0.738
	T_F	0.804	0.791	0.824	0.839	0.891	0.944
	T_L	0.817	0.782	0.812	0.771	0.812	0.786
	T_T	0.797	0.797	0.811	0.827	0.870	0.941

Table 2.5: Power of the test, $n_1 = 5$, $n_2 = 10$, t_1 distribution

α	Type	$\delta = 0$	$\delta = 0.1$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$	$z_i \sim U(0, \delta)$
0.01	T_1	0.009	0.013	0.016	0.031	0.067	0.118
	T_2	0.012	0.006	0.012	0.013	0.033	0.044
	T_F	0.013	0.010	0.020	0.025	0.067	0.106
	T_L	0.011	0.011	0.011	0.022	0.091	0.271
	T_T	0.013	0.010	0.020	0.025	0.067	0.105
0.05	T_1	0.051	0.059	0.066	0.102	0.156	0.232
	T_2	0.054	0.043	0.051	0.058	0.101	0.135
	T_F	0.054	0.054	0.067	0.083	0.169	0.288
	T_L	0.057	0.042	0.063	0.084	0.231	0.528
	T_T	0.050	0.058	0.071	0.080	0.156	0.268
0.10	T_1	0.105	0.108	0.126	0.156	0.231	0.300
	T_2	0.105	0.087	0.095	0.097	0.171	0.186
	T_F	0.101	0.101	0.123	0.161	0.282	0.442
	T_L	0.096	0.092	0.112	0.153	0.354	0.634
	T_T	0.105	0.102	0.117	0.160	0.257	0.367
0.20	T_1	0.200	0.207	0.231	0.256	0.346	0.425
	T_2	0.213	0.200	0.177	0.181	0.266	0.309
	T_F	0.206	0.206	0.220	0.267	0.464	0.661
	T_L	0.180	0.178	0.203	0.272	0.483	0.729
	T_T	0.210	0.195	0.221	0.253	0.402	0.486
0.30	T_1	0.300	0.304	0.329	0.367	0.432	0.527
	T_2	0.299	0.318	0.270	0.270	0.352	0.393
	T_F	0.316	0.299	0.318	0.361	0.601	0.797
	T_L	0.275	0.302	0.292	0.376	0.568	0.776
	T_T	0.323	0.300	0.319	0.342	0.520	0.610
0.40	T_1	0.383	0.396	0.430	0.460	0.534	0.613
	T_2	0.396	0.411	0.371	0.351	0.435	0.487
	T_F	0.414	0.396	0.420	0.459	0.709	0.872
	T_L	0.380	0.393	0.388	0.490	0.631	0.804
	T_T	0.413	0.407	0.408	0.437	0.612	0.734
0.70	T_1	0.676	0.697	0.728	0.755	0.760	0.837
	T_2	0.700	0.696	0.666	0.677	0.680	0.726
	T_F	0.685	0.713	0.693	0.743	0.885	0.979
	T_L	0.695	0.691	0.687	0.740	0.770	0.843
	T_T	0.700	0.707	0.694	0.725	0.857	0.939
0.80	T_1	0.780	0.787	0.818	0.840	0.836	0.888
	T_2	0.789	0.788	0.770	0.784	0.774	0.782
	T_F	0.794	0.797	0.798	0.839	0.933	0.990
	T_L	0.795	0.792	0.795	0.823	0.797	0.848
	T_T	0.785	0.816	0.796	0.820	0.921	0.975

Table 2.6: Power of the test, $n_1 = 10$, $n_2 = 10$, t_1 distribution

α	Type	$\delta = 0$	$\delta = 0.1$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$	$z_i \sim U(0, \delta)$
0.01	T_1	0.006	0.011	0.038	0.113	0.189	0.173
	T_2	0.013	0.010	0.007	0.007	0.000	0.003
	T_F	0.008	0.013	0.027	0.090	0.189	0.168
	T_L	0.007	0.020	0.017	0.090	0.567	0.431
	T_T	0.008	0.013	0.027	0.090	0.189	0.168
0.05	T_1	0.045	0.062	0.117	0.242	0.407	0.373
	T_2	0.047	0.041	0.033	0.044	0.145	0.132
	T_F	0.049	0.049	0.097	0.190	0.347	0.305
	T_L	0.048	0.084	0.083	0.269	0.769	0.678
	T_T	0.050	0.046	0.094	0.186	0.207	0.199
0.10	T_1	0.089	0.113	0.192	0.325	0.423	0.405
	T_2	0.098	0.096	0.084	0.103	0.178	0.172
	T_F	0.098	0.108	0.161	0.319	0.584	0.536
	T_L	0.099	0.134	0.144	0.382	0.787	0.733
	T_T	0.097	0.104	0.153	0.289	0.553	0.505
0.20	T_1	0.181	0.232	0.329	0.475	0.621	0.601
	T_2	0.203	0.183	0.160	0.202	0.344	0.316
	T_F	0.196	0.219	0.285	0.523	0.952	0.865
	T_L	0.187	0.238	0.257	0.495	0.797	0.778
	T_T	0.196	0.211	0.280	0.429	0.603	0.578
0.30	T_1	0.274	0.342	0.429	0.547	0.637	0.616
	T_2	0.309	0.281	0.238	0.288	0.359	0.343
	T_F	0.288	0.333	0.400	0.681	0.990	0.941
	T_L	0.285	0.369	0.352	0.563	0.799	0.786
	T_T	0.292	0.314	0.385	0.532	0.626	0.608
0.40	T_1	0.379	0.455	0.537	0.651	0.723	0.699
	T_2	0.438	0.379	0.327	0.364	0.464	0.442
	T_F	0.391	0.433	0.525	0.775	0.996	0.974
	T_L	0.381	0.469	0.434	0.601	0.799	0.791
	T_T	0.389	0.416	0.492	0.678	0.965	0.918
0.70	T_1	0.656	0.739	0.804	0.861	0.921	0.869
	T_2	0.678	0.665	0.588	0.618	0.707	0.645
	T_F	0.714	0.748	0.787	0.936	1.000	0.998
	T_L	0.600	0.676	0.594	0.650	0.799	0.794
	T_T	0.718	0.729	0.765	0.896	0.999	0.981
0.80	T_1	0.732	0.802	0.865	0.929	0.996	0.989
	T_2	0.743	0.735	0.657	0.681	0.795	0.796
	T_F	0.803	0.819	0.845	0.959	1.000	1.000
	T_L	0.603	0.734	0.597	0.651	0.799	0.794
	T_T	0.824	0.822	0.853	0.957	1.000	0.997

Table 2.7: Power of the test, $n_1 = 5$, $n_2 = 10, e_i = \beta Y_1 + (1 - \beta) Y_2$, where $\beta = 0.25$, $Y_1 \sim \text{Bin}(5, 0.5)$, and $Y_2 \sim \text{Bin}(3, 0.2)$

α	Type	$\delta=0$	$\delta=0.1$	$\delta=0.5$	$\delta=1$	$\delta=2$	$z_i \sim U(0,\delta)$
0.01	T_1	0.008	0.021	0.048	0.140	0.222	0.203
	T_2	0.010	0.008	0.018	0.067	0.139	0.120
	T_F	0.007	0.020	0.038	0.156	0.298	0.258
	T_L	0.010	0.016	0.025	0.280	0.888	0.832
	T_T	0.008	0.020	0.036	0.147	0.276	0.250
0.05	T_1	0.055	0.065	0.146	0.258	0.370	0.343
	T_2	0.048	0.050	0.082	0.170	0.266	0.249
	T_F	0.060	0.065	0.140	0.338	0.638	0.572
	T_L	0.044	0.071	0.107	0.484	0.898	0.890
	T_T	0.063	0.063	0.133	0.306	0.466	0.442
0.10	T_1	0.101	0.119	0.228	0.340	0.450	0.429
	T_2	0.100	0.111	0.135	0.244	0.340	0.317
	T_F	0.108	0.124	0.227	0.512	0.953	0.848
	T_L	0.103	0.134	0.179	0.584	0.899	0.894
	T_T	0.103	0.115	0.229	0.428	0.636	0.592
0.20	T_1	0.196	0.229	0.337	0.448	0.559	0.541
	T_2	0.203	0.211	0.238	0.354	0.469	0.441
	T_F	0.204	0.233	0.381	0.751	0.999	0.988
	T_L	0.198	0.260	0.290	0.667	0.899	0.895
	T_T	0.202	0.230	0.363	0.584	0.790	0.746
0.30	T_1	0.302	0.326	0.414	0.519	0.625	0.596
	T_2	0.298	0.314	0.325	0.440	0.533	0.499
	T_F	0.301	0.340	0.504	0.860	1.000	1.000
	T_L	0.301	0.356	0.378	0.697	0.899	0.895
	T_T	0.294	0.326	0.476	0.710	0.885	0.851
0.40	T_1	0.386	0.423	0.486	0.591	0.708	0.689
	T_2	0.395	0.412	0.400	0.506	0.608	0.585
	T_F	0.399	0.449	0.611	0.914	1.000	1.000
	T_L	0.401	0.463	0.460	0.719	0.899	0.895
	T_T	0.400	0.440	0.575	0.802	0.985	0.965
0.70	T_1	0.676	0.697	0.731	0.814	0.887	0.886
	T_2	0.688	0.693	0.625	0.684	0.805	0.799
	T_F	0.693	0.742	0.825	0.979	1.000	1.000
	T_L	0.623	0.725	0.607	0.742	0.899	0.895
	T_T	0.690	0.730	0.804	0.960	1.000	1.000
0.80	T_1	0.769	0.803	0.804	0.866	0.914	0.895
	T_2	0.780	0.784	0.702	0.753	0.845	0.809
	T_F	0.806	0.834	0.892	0.994	1.000	1.000
	T_L	0.625	0.785	0.616	0.742	0.899	0.895
	T_T	0.791	0.838	0.867	0.982	1.000	1.000

Table 2.8: Power of the test, $n_1 = 10$, $n_2 = 10, e_i = \beta Y_1 + (1-\beta)Y_2$, where $\beta = 0.25$, $Y_1 \sim \text{Bin}(5, 0.5)$, and $Y_2 \sim \text{Bin}(3, 0.2)$

α	Type	$\delta = 0$	$\delta = 0.1$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$	$z_i \sim U(0, \delta)$
0.01	T_1	0.014	0.009	0.041	0.128	0.199	0.185
	T_2	0.011	0.007	0.014	0.007	0.000	0.010
	T_F	0.009	0.008	0.036	0.108	0.193	0.174
	T_L	0.007	0.009	0.012	0.056	0.416	0.363
	T_T	0.009	0.008	0.036	0.108	0.193	0.174
0.05	T_1	0.056	0.051	0.128	0.273	0.394	0.360
	T_2	0.047	0.044	0.049	0.060	0.174	0.150
	T_F	0.048	0.048	0.099	0.205	0.295	0.292
	T_L	0.040	0.052	0.057	0.205	0.711	0.643
	T_T	0.051	0.051	0.104	0.207	0.216	0.220
0.10	T_1	0.109	0.086	0.206	0.335	0.414	0.395
	T_2	0.094	0.095	0.098	0.141	0.207	0.194
	T_F	0.099	0.099	0.172	0.347	0.596	0.545
	T_L	0.094	0.096	0.107	0.335	0.758	0.709
	T_T	0.103	0.095	0.177	0.333	0.568	0.510
0.20	T_1	0.224	0.204	0.312	0.466	0.575	0.570
	T_2	0.185	0.183	0.179	0.231	0.356	0.352
	T_F	0.209	0.184	0.301	0.527	0.889	0.850
	T_L	0.200	0.195	0.208	0.467	0.784	0.763
	T_T	0.203	0.181	0.305	0.476	0.621	0.589
0.30	T_1	0.315	0.304	0.409	0.532	0.598	0.590
	T_2	0.272	0.299	0.257	0.311	0.388	0.387
	T_F	0.292	0.285	0.410	0.674	0.975	0.945
	T_L	0.298	0.295	0.311	0.537	0.787	0.768
	T_T	0.298	0.288	0.398	0.555	0.638	0.632
0.40	T_1	0.434	0.382	0.512	0.616	0.650	0.648
	T_2	0.363	0.407	0.350	0.385	0.434	0.437
	T_F	0.412	0.396	0.502	0.768	0.992	0.984
	T_L	0.392	0.385	0.413	0.591	0.787	0.777
	T_T	0.409	0.387	0.491	0.697	0.931	0.922
0.70	T_1	0.720	0.686	0.785	0.803	0.815	0.812
	T_2	0.684	0.688	0.573	0.554	0.606	0.606
	T_F	0.705	0.701	0.773	0.943	1.000	1.000
	T_L	0.695	0.724	0.655	0.679	0.787	0.780
	T_T	0.693	0.699	0.764	0.906	0.995	0.992
0.80	T_1	0.805	0.803	0.861	0.923	0.991	0.976
	T_2	0.777	0.789	0.680	0.663	0.777	0.780
	T_F	0.798	0.800	0.866	0.966	1.000	1.000
	T_L	0.786	0.810	0.718	0.693	0.787	0.780
	T_T	0.797	0.792	0.861	0.967	1.000	1.000

Table 2.9: Power of the test, $n_1 = 5$, $n_2 = 10$, $SN(1, 5)$ distribution

α	Type	$\delta = 0$	$\delta = 0.1$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$	$z_i \sim U(0, \delta)$
0.01	T_1	0.004	0.011	0.052	0.139	0.222	0.209
	T_2	0.013	0.007	0.014	0.068	0.116	0.102
	T_F	0.008	0.012	0.048	0.161	0.265	0.248
	T_L	0.004	0.013	0.032	0.247	0.882	0.848
	T_T	0.008	0.012	0.047	0.161	0.255	0.235
0.05	T_1	0.054	0.061	0.129	0.270	0.368	0.360
	T_2	0.063	0.044	0.073	0.164	0.276	0.254
	T_F	0.053	0.053	0.130	0.349	0.626	0.590
	T_L	0.050	0.050	0.131	0.475	0.904	0.903
	T_T	0.053	0.050	0.128	0.319	0.469	0.435
0.10	T_1	0.090	0.111	0.208	0.349	0.424	0.406
	T_2	0.122	0.093	0.123	0.242	0.338	0.325
	T_F	0.117	0.110	0.207	0.509	0.936	0.856
	T_L	0.096	0.093	0.202	0.577	0.904	0.906
	T_T	0.117	0.105	0.202	0.434	0.644	0.614
0.20	T_1	0.190	0.222	0.306	0.448	0.547	0.534
	T_2	0.199	0.207	0.241	0.345	0.458	0.441
	T_F	0.204	0.194	0.350	0.694	1.000	0.983
	T_L	0.187	0.208	0.331	0.679	0.905	0.909
	T_T	0.212	0.204	0.331	0.591	0.762	0.731
0.30	T_1	0.281	0.317	0.411	0.522	0.609	0.596
	T_2	0.301	0.307	0.319	0.423	0.524	0.511
	T_F	0.297	0.311	0.472	0.820	1.000	0.998
	T_L	0.286	0.304	0.414	0.727	0.905	0.909
	T_T	0.297	0.308	0.447	0.690	0.857	0.835
0.40	T_1	0.375	0.402	0.513	0.593	0.688	0.671
	T_2	0.382	0.396	0.388	0.489	0.602	0.597
	T_F	0.384	0.427	0.579	0.905	1.000	1.000
	T_L	0.396	0.418	0.500	0.751	0.905	0.909
	T_T	0.389	0.429	0.547	0.793	0.971	0.953
0.70	T_1	0.711	0.708	0.766	0.828	0.907	0.904
	T_2	0.705	0.679	0.626	0.657	0.794	0.789
	T_F	0.668	0.717	0.821	0.985	1.000	1.000
	T_L	0.709	0.721	0.717	0.783	0.905	0.909
	T_T	0.656	0.706	0.819	0.961	1.000	1.000
0.80	T_1	0.803	0.813	0.837	0.857	0.910	0.912
	T_2	0.809	0.776	0.723	0.716	0.806	0.804
	T_F	0.763	0.816	0.904	0.990	1.000	1.000
	T_L	0.814	0.816	0.767	0.784	0.905	0.909
	T_T	0.753	0.803	0.887	0.988	1.000	1.000

Table 2.10: Power of the test, $n_1 = 10$, $n_2 = 10$, $SN(0, 1, 5)$ distribution

α	Type	$\delta = 0$	$\delta = 0.1$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$	$z_i \sim U(0, \delta)$
0.01	T_1	0.009	0.016	0.039	0.096	0.160	0.164
	T_2	0.005	0.005	0.020	0.014	0.007	0.008
	T_F	0.006	0.008	0.038	0.082	0.150	0.159
	T_L	0.012	0.016	0.020	0.023	0.174	0.287
	T_T	0.006	0.008	0.038	0.082	0.150	0.159
0.05	T_1	0.056	0.069	0.122	0.195	0.319	0.357
	T_2	0.047	0.040	0.067	0.050	0.084	0.111
	T_F	0.045	0.056	0.115	0.175	0.245	0.271
	T_L	0.059	0.052	0.079	0.127	0.426	0.588
	T_T	0.045	0.058	0.113	0.175	0.211	0.216
0.10	T_1	0.100	0.112	0.185	0.265	0.346	0.382
	T_2	0.091	0.098	0.109	0.109	0.155	0.185
	T_F	0.102	0.108	0.191	0.261	0.431	0.509
	T_L	0.097	0.098	0.152	0.251	0.591	0.676
	T_T	0.103	0.109	0.189	0.245	0.403	0.468
0.20	T_1	0.214	0.220	0.313	0.404	0.526	0.545
	T_2	0.190	0.189	0.185	0.196	0.286	0.331
	T_F	0.200	0.211	0.305	0.401	0.658	0.792
	T_L	0.202	0.197	0.275	0.410	0.694	0.733
	T_T	0.191	0.210	0.294	0.374	0.501	0.567
0.30	T_1	0.302	0.319	0.406	0.485	0.548	0.559
	T_2	0.287	0.282	0.265	0.286	0.349	0.399
	T_F	0.292	0.310	0.427	0.549	0.826	0.920
	T_L	0.288	0.305	0.374	0.508	0.718	0.748
	T_T	0.299	0.308	0.398	0.468	0.595	0.625
0.40	T_1	0.394	0.422	0.527	0.604	0.654	0.620
	T_2	0.390	0.388	0.349	0.358	0.422	0.462
	T_F	0.393	0.394	0.527	0.669	0.919	0.968
	T_L	0.393	0.402	0.461	0.590	0.737	0.759
	T_T	0.404	0.409	0.498	0.600	0.812	0.876
0.70	T_1	0.706	0.727	0.792	0.829	0.851	0.809
	T_2	0.688	0.650	0.592	0.585	0.603	0.616
	T_F	0.679	0.708	0.801	0.921	0.990	0.996
	T_L	0.715	0.696	0.658	0.717	0.769	0.771
	T_T	0.685	0.707	0.762	0.844	0.951	0.981
0.80	T_1	0.798	0.825	0.883	0.930	0.980	0.978
	T_2	0.788	0.755	0.708	0.682	0.713	0.757
	T_F	0.793	0.802	0.873	0.962	0.998	0.999
	T_L	0.813	0.799	0.733	0.734	0.770	0.771
	T_T	0.790	0.815	0.873	0.942	0.994	0.997

Table 2.11: Power of the test, $n_1 = 5$, $n_2 = 10$, $Exp(1)$ distribution

α	Type	$\delta = 0$	$\delta = 0.1$	$\delta = 0.5$	$\delta = 1$	$\delta = 2$	$z_i \sim U(0, \delta)$
0.01	T_1	0.007	0.015	0.047	0.122	0.198	0.206
	T_2	0.012	0.006	0.022	0.049	0.107	0.128
	T_F	0.009	0.013	0.040	0.117	0.232	0.255
	T_L	0.016	0.009	0.048	0.167	0.590	0.764
	T_T	0.009	0.013	0.040	0.116	0.226	0.247
0.05	T_1	0.042	0.063	0.136	0.227	0.322	0.339
	T_2	0.050	0.048	0.086	0.135	0.239	0.252
	T_F	0.036	0.054	0.139	0.283	0.518	0.556
	T_L	0.062	0.058	0.141	0.331	0.769	0.869
	T_T	0.039	0.052	0.139	0.260	0.433	0.451
0.10	T_1	0.083	0.120	0.216	0.305	0.385	0.409
	T_2	0.099	0.091	0.140	0.217	0.304	0.332
	T_F	0.096	0.112	0.234	0.400	0.719	0.819
	T_L	0.111	0.111	0.227	0.445	0.812	0.881
	T_T	0.092	0.111	0.222	0.362	0.561	0.591
0.20	T_1	0.174	0.223	0.330	0.418	0.491	0.522
	T_2	0.203	0.174	0.230	0.320	0.403	0.435
	T_F	0.194	0.213	0.383	0.597	0.898	0.979
	T_L	0.213	0.196	0.355	0.581	0.832	0.881
	T_T	0.182	0.211	0.356	0.522	0.689	0.741
0.30	T_1	0.270	0.321	0.426	0.486	0.555	0.575
	T_2	0.315	0.276	0.323	0.384	0.471	0.498
	T_F	0.287	0.316	0.499	0.738	0.964	0.995
	T_L	0.306	0.300	0.454	0.653	0.841	0.882
	T_T	0.289	0.316	0.460	0.628	0.790	0.841
0.40	T_1	0.383	0.444	0.522	0.561	0.651	0.655
	T_2	0.415	0.376	0.400	0.458	0.559	0.590
	T_F	0.390	0.417	0.608	0.834	0.984	0.999
	T_L	0.410	0.396	0.536	0.707	0.844	0.882
	T_T	0.377	0.397	0.560	0.738	0.888	0.944
0.70	T_1	0.693	0.743	0.779	0.812	0.870	0.875
	T_2	0.724	0.669	0.647	0.636	0.736	0.769
	T_F	0.705	0.692	0.863	0.969	0.998	1.000
	T_L	0.710	0.697	0.731	0.774	0.847	0.882
	T_T	0.684	0.709	0.820	0.925	0.991	1.000
0.80	T_1	0.799	0.818	0.845	0.856	0.884	0.882
	T_2	0.822	0.774	0.719	0.716	0.762	0.790
	T_F	0.793	0.796	0.926	0.985	0.999	1.000
	T_L	0.814	0.792	0.763	0.780	0.847	0.882
	T_T	0.797	0.815	0.895	0.969	0.997	1.000

Table 2.12: Power of the test, $n_1 = 10$, $n_2 = 10$, $Exp(1)$ distribution

Chapter 3

Finite-sample consistency of combination-based permutation tests

3.1 Introduction

As we said in the introduction of this dissertation, the second problem to be addressed in the analysis of three-dimensional surfaces is that the number of variables (e.g. three times the points -landmarks- considered in the surface) is far greater than the number of observed units. A similar situation is not at all unusual, in many cases for example in analysis of microarrays and genomics (Salmaso and Solari, 2005, 2006), shape analysis (Bookstein, 1991), functional data analysis (Ramsay and Silverman, 1997, 2002; Ferraty and Vieu, 2006) it may happen that the number of observed variables is very much larger than that of subjects. In Pesarin (2001) it is shown that, under very mild conditions, the power function of permutation tests based on associative statistics monotonically increases as the related standardized noncentrality functional increases. This is true also for multivariate situations. In particular, for any added variable the power does not decreases if this variable makes larger standardized global noncentrality. This property allow us to define the notion of *finite-sample consistency* for those kinds of combination-based permutation tests. The concept of *finite-sample consistency* is different from the traditional property of consistency of a parametric test. Generally we are interested in studying the power W of a test when the sample size goes to infinity. A test is usually defined consistent if

$$\lim_{n\to\infty} W_n = 1 \tag{3.1}$$

when H_0 is not true. Within the *finite-sample consistency* sufficient conditions are established as to ensure that the power of the test goes to one when the number of "informative" variables V diverges, while the number of observations remains fixed, that is

$$\lim_{V\to\infty} W_{V,n} = 1 \tag{3.2}$$

when H_0 is not true.
In this chapter we will show some fundamental aspects about the *finite-sample consistency* giving sufficient conditions in order that the rejection rate converges to one, for fixed sample sizes at any attainable α-values, when the number of variables diverges. We will present a simulation study. At the end, using some results presented here we could easily prove the consistency of multi-sided test.

3.2 Finite sample consistency

As a guide, we refer to one-sided two-sample designs and we use the same notation of previous chapter. Here we discuss testing problems for stochastic dominance alternatives as are generated by symbolic treatments with non-negative random shift effects Δ. In particular, the alternative assumes that treatments produce effects Δ_1 and Δ_2, respectively, and that $\Delta_1 \overset{d}{>} \Delta_2$,. Thus, the hypotheses are $H_0 : \{X_1 \overset{d}{=} X_2\} \equiv \{P_1 = P_2\}$, and $H_1 : \{(X_1 + \Delta_1) \overset{d}{>} (X_2 + \Delta_2)\}$. Extensions to non-positive, two-sided alternatives are straightforward. Note that under H_0 data of two samples are exchangeable, in accordance with the notion that subjects are randomized to treatments. Without loss of generality, we assume that effects in H_1 are such that $\Delta_1 = \Delta \overset{d}{>} 0$ and $\Pr\{\Delta_2 = 0\} = 1$. Condition $\Delta_2 \overset{d}{=} 0$ agrees with the notion that an active treatment is only assigned to subjects of first sample and a placebo to those of the second. In this situation, since effects Δ may depend on null responses X_1, stochastic dominance $(X_1 + \Delta) \overset{d}{>} X_2 = X$ is compatible with non-homoscedasticities in the alternative. Thus, the null hypothesis may also be written as $H_0 : \{\Delta \overset{d}{=} 0\}$. In the context of this dissertation, it is also worth noting that observed variable X, random deviates Z, sample space $\mathcal{X}$, and random effect Δ are V-dimensional, with $V \geq 1$.
In what follows we consider associative test statistics defined as $T^*(\Delta) = \sum_i \varphi[X^*_{1i}(\Delta)]/n_1 - \sum_i \varphi[X^*_{2i}(\Delta)]/n_2$, where φ is any non-degenerate measurable non-decreasing function of the data and so $T^*(\Delta)$ corresponds to the

comparison of sampling φ-means: $T^*(\Delta) = \bar{\varphi}_1^* - \bar{\varphi}_2^*$ say. Of course, the observed value of $T(\Delta)$ is $T^o(\Delta) = \sum_i \varphi[X_{1i}(\Delta)]/n_1 - \sum_i \varphi[X_{2i}]/n_2$, and $T^o(0)$ and $T^*(0)$ are the related observed and permutation values when $\Delta \overset{d}{=} 0$.
We want investigate the rejection behaviour of permutation test T when the random effect Δ can diverge to the infinity. Any test statistic is a mapping from the sample space to the real line, $T : \mathcal{X}^n \to \mathbb{R}^1$. So that we investigate on a test T by comparing its behaviour in H_0 to that in H_1, that is $T(\mathbf{X}(0))$ to $T(\mathbf{X}(\Delta))$. Such a comparison, together with their respective asymptotic behaviour, will be perfectly clear in the permutation framework if we are able to write their related random variables in the form $T(\mathbf{X}(\Delta)) = T(\mathbf{X}(0)) + \phi_T(\Delta, \mathbf{X}(0))$, where the induced noncentrality $\phi_T(\Delta, \mathbf{X}(0))$ is a random function which may diverge in probability, i.e. such that $\lim_{\Delta\to\infty} \Pr\{\phi_T > t\} = 1$, for any real t.
Since main inferential conclusions associated with permutation tests are concerning the observed data set $\mathbf{X}$ related to the given set of $n = n_1 + n_2$ individuals, the notion of consistency that is truly useful is the *weak* form (or in probability) which essentially states that *for divergent values of noncentrality parameter induced by the test statistic, the limit rejection probability of test T is of one for any fixed $\alpha > 0$.* The sense of this is that, for fixed sample sizes and large values of induced noncentrality, the rejection probability of T approaches one. With reference for simplicity to fixed effects δ, in practice this means that the rejection rate is greater in H_1 than in H_0, that is when $\delta > 0$ than when $\delta = 0$. Similarly, it is easy to establish that the rejection rate of H_0 is greater for larger δ. That is, if $\delta < \delta'$, then for any attainable α-value

$$\Pr\{\lambda(\mathbf{X}(\delta)) \leq \alpha | \mathcal{X}^n_{/\mathbf{X}(\delta)}\} \leq \Pr\{\lambda(\mathbf{X}(\delta')) \leq \alpha | \mathcal{X}^n_{/\mathbf{X}(\delta')}\}$$

and

$$\mathbf{E}_P\left[\Pr\{\lambda(\mathbf{X}(\delta)) \leq \alpha | \mathcal{X}^n_{/\mathbf{X}(\delta)}\}\right] \leq \mathbf{E}_P\left[\Pr\{\lambda(\mathbf{X}(\delta')) \leq \alpha | \mathcal{X}^n_{/\mathbf{X}(\delta')}\}\right]$$

where $\mathbf{E}_P(\bullet)$ is the mean value of $(\bullet)$ with respect to P. Similar relations are true also for random effects Δ. Considering the finite-sample property of permutation test it will easy to show the consistency of multi-sided test.

3.2.1 Weak unconditional finite-sample consistency of T

Let us argue for fixed effects δ first. The extension to random effects Δ will be considered in the specific section.
Suppose that the following conditions are satisfied:

- T is any associative test statistic for one-sided hypotheses;
- sample sizes (n_1, n_2) are fixed and finite;
- the data set $\mathbf{X}(\delta) = (\mathbf{Z}_1 + \delta, \mathbf{Z}_2)$, where $(\mathbf{Z}_1, \mathbf{Z}_2) = \mathbf{Z} \in \mathcal{X}^n$ are i.i.d. measurable real random deviates whose parent distribution is $P_Z(z) = \Pr\{Z \leq z\}$ and $\delta = (\delta, \ldots, \delta)'$ is the vector of non-negative fixed effects;
- fixed effects δ diverge to the infinity according to whatever monotonic sequence $\{\delta_v, v \geq 1\}$, the elements of which are such that $\delta_v \leq \delta_{v'}$ for any pair $v < v'$.

then the permutation unconditional rejection rate of test T converges to 1 for all α-values not smaller than the minimum attainable α; so that T is weak unconditional finite-sample consistent.

To show the unconditional finite-sample consistency of T we consider the observed data set $\mathbf{X}(\delta) = (\mathbf{Z}_1 + \delta, \mathbf{Z}_2)$ for fixed deviates $\mathbf{Z}$; of course $\mathbf{X}(\delta)$ depends by δ. The permutation support induced by the test statistic T when applied to the data set $\mathbf{X}(\delta)$ is $\mathcal{T}_{\mathbf{X}(\delta)} = \{T^*(\delta) = T(\mathbf{X}^*(\delta)) : \mathbf{X}^*(\delta) \in \mathcal{X}^n_{/\mathbf{X}(\delta)}\}$. Depending on $\mathbf{Z}$, in the sequence $\{\delta_v, v \geq 1\}$ there is a value $\delta_{\mathbf{Z}}$ of δ such that the related observed value $T^o(\mathbf{X}(\delta_{\mathbf{Z}}))$ is right-extremal for the induced permutation support $\mathcal{T}_{\mathbf{X}(\delta_{\mathbf{Z}})}$, that is $T^o(\mathbf{X}(\delta_{\mathbf{Z}})) = \max_{\mathcal{T}_{\mathbf{X}(\delta_{\mathbf{Z}})}} \{T^*(\delta_{\mathbf{Z}}) : \mathbf{X}^*(\delta_{\mathbf{Z}}) \in \mathcal{X}^n_{/\mathbf{X}(\delta_{\mathbf{Z}})}\}$. This $\delta_{\mathbf{Z}}$ can be determined by observing that a sufficient condition for right-extremal property of T^o is that

$$\min_{n_1}(Z_{1i} + \delta_{\mathbf{Z}}) > \max_{n_2}(Z_{2i}), \tag{3.3}$$

indeed, since φ is monotonic non-decreasing, we necessarily have that

$$\sum_i \varphi(Z_{1i} + \delta_{\mathbf{Z}})/n_1 > \sum_i \varphi(Z_{2i})/n_2$$

and so $T^o(\mathbf{X}(\delta_{\mathbf{Z}}))$ is right-extremal because for all permutations $\mathbf{X}^*(\delta_{\mathbf{Z}}) \neq \mathbf{X}(\delta_{\mathbf{Z}})$ it is $T^o(\mathbf{X}^*(\delta_{\mathbf{Z}})) < T^o(\mathbf{X}(\delta_{\mathbf{Z}}))$.

Observing that the random deviates Z_{ji} are i.i.d., the probability of the event in equation (3.3) is

$$\Pr\left\{\min_{n_1}(Z_{1i} + \delta) > \max_{n_2}(Z_{2i})\right\} = \int_{\mathcal{X}} \{[1 - P_Z(t - \delta)]^{n_1}\} \, \mathrm{d}\,[P_Z(t)]^{n_2}, \tag{3.4}$$

the limit of which, as δ goes to the infinity according to the given sequence $\{\delta_v, v \geq 1\}$, is of 1 since the measurability of random deviates Z implies that $\lim_{z\to-\infty}\Pr(Z \leq z) = 0$, $\lim_{z\to+\infty}\Pr(Z \leq z) = 1$, and because, by the Lebesgue's monotone convergence theorem (see Lehmann, 1986, pg. 39) in force of which the limit of an integral is the integral of the limit, the associated sequence of probability measures $\{P_Z(t-\delta_v), v \geq 1\}$ converges to zero monotonically for any t.

An interpretation of this is that the probability of finding a set $\mathbf{Z} \in \mathcal{X}^n$ for which there does not exist a finite value of $\delta_{\mathbf{Z}} \in \{\delta_v, v \geq 1\}$ such that $\min_{n_1}(Z_{1i} + \delta_{\mathbf{Z}}) > \max_{n_2}(Z_{2i})$ converges to zero monotonically as δ diverges. This implies that the unconditional rejection rate

$$W_\alpha(\delta) = \int_{\mathcal{X}} \Pr\{\lambda(\mathbf{X}(\delta)) \leq \alpha | \mathcal{X}^n_{/\mathbf{X}(\delta)}\}\, \mathrm{d}P_{\mathbf{Z}}(\mathbf{z}),$$

where $P_{\mathbf{Z}}$ is the multivariate distribution of vector $\mathbf{Z}$, as δ tends to the infinity converges to 1 for all α-values not smaller to the minimum attainable α-value α_a, which for one-sided alternatives is of $1/\binom{n}{n_1}$ (it is of $2/\binom{n}{n_1}$ for two-sided alternatives).
It is to be emphasized that the notion of unconditional finite-sample consistency, defined for divergent fixed effects δ, is different from the traditional notion of (unconditional) consistency of a test, which in turn considers the behaviour of rejection rate for given δ when $\min(n_1, n_2)$ diverges. It is known that, in order to attain permutation unconditional consistency it is required that random deviates Z at least possess finite second moment (Lehmann, 1986; Pesarin, 2001). Here we only require they are measurable, so that in this respect it is to be emphasized that random deviates Z are not required to be provided with finite moments of any positive order. For instance, they can be distributed according to Cauchy $\mathcal{C}au(0,\sigma)$ or Pareto $\mathcal{P}a(\theta,\sigma)$, with shape parameter $0 < \theta \leq 1$, and both with finite scale coefficients $\sigma > 0$, etc.

3.2.2 Unconditional finite sample consistency for $V \to \infty$

To see the strict relation between this form of consistency and that described in equation (3.2), let us firstly consider a case where in a two-sample problem there are $V \geq 1$ homoschedastic variables $X = (X_1, ..., X_V)$, in which the observed data set is $\mathbf{X}(\delta) = \{\delta_h + Z_{h1i}, i = 1, \ldots, n_1; Z_{h2i}, i = 1, \ldots, n_2; h = 1, \ldots, V\}$, and the hypotheses are $H_0 : \{\mathbf{X}_1 \stackrel{d}{=} \mathbf{X}_2\} = \{\delta = \mathbf{0}\}$ against $H_1 : \{\mathbf{X}_1 \stackrel{d}{>} \mathbf{X}_2\} = \{\delta \geq \mathbf{0}\}$, where δ is the vector of fixed effects, i.e. $\delta = (\delta_1, ..., \delta_V)'$, in which δ_h is the effect for the h-th variable and $\mathbf{0}$ is the vector with V null components. Consider that the permutation test statistic has the form

$$T''^{*}(\delta) = \psi(V) \sum_{h=1}^{V} \left[\bar{X}^{*}_{h1}(\delta_h) - \bar{X}^{*}_{h2}(\delta_h) \right],$$

where $\psi(V)$ is such that the statistic $T(\mathbf{X}(\mathbf{0}))$ is measurable as V diverges, so that $\lim_{z\to\infty} \Pr\{T(\mathbf{X}(\mathbf{0})) \leq z; P_{\mathbf{Z}}\} = 1$, and

$$\bar{X}^{*}_{hj}(\delta_h) = \sum_{i=1}^{n_j} X^{*}_{hji}(\delta_h)/n_j = T^{*}_{h}(\delta_h),$$

$j = 1, 2$, are permutation sample means of the h-th variable. In other terms, the statistic T'' is a measurable sum of V partial tests T_h in accordance to the direct combination of several partial tests, that is a global test statistic T'' is given by the form $T'' = \sum_h T_h, T^o = \sum_h T^o_h$ and $T''^{*}_{r} = \sum_h T^{*}_{hr}, r = 1, \ldots, B$ for the combined test, observed, and permutation values respectively. Suppose now that the noncentrality parameter induced by the test statistic, that is the global effect $\bar{\delta}_V = \psi(V) \sum_{h\leq V} \delta_h$, diverges as V diverges. To see the unconditional finite-sample consistency of T, let us consider the permutationally equivalent form of the test statistics

$$\begin{aligned} T''^{*}(\delta) &= \psi(V) \sum_{h=1}^{V} \sum_{i=1}^{n_1} X^{*}_{h1i}(\delta_h) = \psi(V) \sum_{i=1}^{n_1} \sum_{h=1}^{V} X^{*}_{h1i}(\delta_h) \\ &= \sum_{i=1}^{n_1} Y^{*}_{1i}(\delta) = T''^{*}(0) + n_1 \bar{\delta}^{*}_{V}, \end{aligned}$$

where the $Y_{1i}(\delta) = \psi(V) \sum_{h\leq V} X_{h1i}(\delta_h)$, $i = 1, \ldots, n_1$, are univariate data transformations which summarize the whole set of information on effects δ collected by the V variables, $\bar{\delta}^{*}_{V} = \psi(V) \sum_{h\leq V} \delta^{*}_{h}$, and $T''^{*}(\mathbf{0})$ is the null permutation value of T'' which is a function only of random deviates $\mathbf{Z}^{*}_{1} \in \mathbf{Z}$. The right-hand side expression shows that a multivariate test statistic is reduced to one one-dimensional quantity. Thus conditions of section 3.2.1 are satisfied because, by assumption, $T''^{*}(\mathbf{0})$ is measurable and $\bar{\delta}_V$ is assumed to diverge. And so T is unconditionally finite-sample consistent.

A typical case occurs when all component variables $X_h(\delta_h)$, $h = 1, \ldots, V$ are provided with finite mean value, that is when $\mathbf{E}\left[|X_h(\delta_h)|\right] < \infty$, $h = 1, \ldots, V$. In such a case, we may put $\psi(V) = 1/V$. So that, under to conditions for the the law of large numbers for dependent variables (Feller, 1968), $T''^{*}(\mathbf{X}(\mathbf{0}))$ converges to zero in probability (at least). Thus, if $\bar{\delta}_V = \sum_{h\leq V} \delta_h/V$ is positive in the limit all assumptions at beginning of section 3.2.1 are met, T'' is finite-sample consistent.

3.2.3 Weak unconditional consistency of T for $n \to \infty$

In this section we will consider the relationship between the finite sample consistency and the traditional notion of consistency described in equation (3.1). Suppose that conditions of section 3.2.1 hold and so T is a finite sample consistent statistic. We consider the case of a two-sample problem for one-sided alternatives with the data set $\mathbf{X}(\delta) = \{\delta + Z_{1i}, i = 1, \ldots, n_1; Z_{2i}, i = 1, \ldots, n_2\}$, where $\mathbf{E}[Z_{ji}] = 0$ and the two sample sizes (n_1, n_2) satisfy the relation $(n_1 = vm_1, n_2 = vm_2)$ so that they can diverge according to the sequence $\{(vm_1, vm_2), v \geq 1\}$.
Let us observe that the effect δ is now a fixed and unknown constant and that sample sizes diverge, so that the traditional notion of consistency may be applied to T. For any integer $v \geq 1$, let us arrange the one-dimensional data set $\mathbf{X}_1(\delta) = (\delta+\mathbf{Z}_1) = \{\delta + Z_{1i}, i = 1, \ldots, n_1\}$ and $\mathbf{X}_2 = \mathbf{Z}_2 = \{Z_{2i}, i = 1, \ldots, n_2\}$ into respectively the V-dimensional sets $\mathbf{Y}_1(\delta) = \{Y_{11i} = X_{1i}, Y_{21i} = X_{1,v+i}, \ldots, Y_{v1i} = X_{1,(m_1-1)v+i}, i = 1, \ldots, m_1\}$ and $\mathbf{Y}_2 = \{Y_{12i} = X_{2i}, Y_{21i} = X_{2,v+i}, \ldots, Y_{v2i} = X_{2,(m_2-1)v+i}, i = 1, \ldots, m_2\}$, where $(n_1, n_2) = (vm_1, vm_2)$. That is

$$\begin{bmatrix} X_{1,1} \\ \vdots \\ X_{1,i} \\ \vdots \\ X_{1,vm_1} \end{bmatrix} = \begin{bmatrix} X_{1,1} & \ldots & X_{1,i} & \ldots & X_{1,v} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ X_{1,kv+1} & \ldots & X_{1,kv+i} & \ldots & X_{1,(k+1)v} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ X_{1,(m_1-1)v+1} & \ldots & X_{1,(m_1-1)v+i} & \ldots & X_{1,m_1v} \end{bmatrix} \tag{3.5}$$

$$= \begin{bmatrix} Y_{1,1,1} & \ldots & Y_{h,1,1} & \ldots & Y_{v,1,1} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ Y_{1,1,i} & \ldots & Y_{h,1,i} & \ldots & Y_{v,1,i} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ Y_{1,1,m_1} & \ldots & Y_{h,1,m_1} & \ldots & Y_{v,1,m_1} \end{bmatrix} \tag{3.6}$$

and

$$\begin{bmatrix} X_{2,1} \\ \vdots \\ X_{2,i} \\ \vdots \\ X_{2,vm_2} \end{bmatrix} = \begin{bmatrix} X_{2,1} & \ldots & X_{2,i} & \ldots & X_{2,v} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ X_{2,kv+1} & \ldots & X_{2,kv+i} & \ldots & X_{2,(k+1)v} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ X_{2,(m_2-1)v+1} & \ldots & X_{2,(m_2-1)v+i} & \ldots & X_{2,m_2v} \end{bmatrix} \tag{3.7}$$

$$= \begin{bmatrix} Y_{1,2,1} & \ldots & Y_{h,2,1} & \ldots & Y_{v,2,1} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ Y_{1,2,i} & \ldots & Y_{h,2,i} & \ldots & Y_{v,2,i} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ Y_{1,2,m_2} & \ldots & Y_{h,2,m_2} & \ldots & Y_{v,2,m_2} \end{bmatrix} \tag{3.8}$$

Thus the data vector $\mathbf{X}(\delta)$, with 1 column and $n = n_1 + n_2$ rows, is organized into a matrix $\mathbf{Y}(\delta)$ with ν columns and $m = m_1 + m_2$ rows. Of course, as v diverges also $\min(n_1, n_2)$ diverges. If we apply the same statistic as before we observe that, for any $v \geq 1$,:

$$\begin{aligned} T(\mathbf{X}(\delta)) &= \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1i}(\delta) \\ &= \frac{1}{m_1}\sum_{i=1}^{m_1}\frac{1}{v}\sum_{h=1}^{v} Y_{h1i}(\delta) = T(\mathbf{Y}(\delta)) \end{aligned}$$

that is the two statistics coincide. Test statistic T when applied to the data set $\mathbf{Y}(\delta)$, as in the previous example is unconditionally finite-sample consistent, because all the required conditions are satisfied by assumption.
Moreover, we may also write $T(\mathbf{X}(\delta)) = T(\mathbf{X}(0)) + \delta = T(\mathbf{Y}(\delta))$, stressing that two forms have the same null distribution and the same non-centrality parameter which does not vary as v diverges, whereas the null component $T(\mathbf{X}(0))$ as v diverges collapses almost surely towards zero by the strong law of large numbers because, by assumption, the random deviates Z have first moment equal to 0 and observations in $\mathbf{Z}$ are i.i.d.. Thus, the rejection probability for both ways converges to 1, $\forall\ \delta > 0$. And so weak unconditional finite-sample consistency implies weak unconditional consistency, in accordance with the traditional notion of consistency, for all $\alpha \geq \alpha$-attainable.

3.2.4 Weak unconditional finite-sample consistency for random effects

The previous results can be extended to divergent random effects Δ according to whatever sequence $\{\Delta_v, v \geq 1\}$, whose elements are stochastically non-decreasing, i.e. $\Delta_v \overset{d}{\leq} \Delta_{v+1}, \forall v \geq 1$,and provided that $\lim_{v\to\infty} \Pr\{\Delta_v > u\} \to 1$ for every finite u.
It is easy to verify that the finite sample consistency of e test T holds also for random effects if we consider that to apply the Lebesgue's monotone convergence theorem to (3.4) it suffices that $P_Z(t - \Delta'' \leq u)$ is stochastically dominated by $P_Z(t - \Delta' \leq u)$ for every u, whenever $\Delta' \overset{d}{\leq} \Delta''$. So that the associated sequence of probabilities $\{P_Z[t - \Delta_v], v \geq 1\}$ monotonically converges to zero.

This property is useful because it extends the validity of previous results to the case of heteroscedastic variables. Let us consider a heteroscedastic data set is $\mathbf{X}(\delta) = (\delta_h + \sigma_h Z_{h1i}, i = 1, \ldots, n_1, \sigma_h Z_{h2i}, i = 1, \ldots, n_2; h = 1, \ldots, V)$ for the hypotheses $H_0 : \{\mathbf{X}_1 \overset{d}{=} \mathbf{X}_2\} = \{\delta = \mathbf{0}\}$ against $H_1 : \{\mathbf{X}_1 \overset{d}{>} \mathbf{X}_2\} = \{\delta \geq \mathbf{0}\}$, where δ_h and σ_h are the fixed effect and the scale coefficient of the hth variable. Suppose also that the test statistic has the form

$$T''^*(\delta) = \psi(V) \sum_{h=1}^{V} [\bar{X}^*_{h1}(\delta_h) - \bar{X}^*_{h2}(\delta_h)]/S_h,$$

where, as in section 3.2.2,

$$\bar{X}^*_{hj}(\delta) = \sum_{i=1}^{n_j} X^*_{hji}(\delta_h)/n_j = T^*_h(\delta),$$

and S_h is a permutation invariant statistic for the hth scale coefficient σ_h, that is a function $S[X_{hji}(\delta_h), i = 1, \ldots, n_j, j = 1, 2]$ of pooled data, so that both conditional and unconditional distributions of $[\bar{X}_{h1}(\delta_h) - \bar{X}_{h2}(0)]/S_h$ are invariant with respect to scale σ_h, $h = 1, \ldots, V$, and $\psi(V)$ is such that the statistic $T''^*(0)$ is measurable as V diverges. Therefore, the statistic T''^* is a measurable sum of V scale-invariant partial tests T^*_h. Since S_h is a function of random data $\mathbf{Z} \in \mathcal{X}^n$, and thus is a random object, the scale-invariant non-centrality parameter $\psi(V) \sum_{h \leq V} \delta_h / S_h$ becomes a random quantity which we may denominate Δ_V. Also, we may denominate the tests statistic as $T(\Delta_V)$. Suppose now that the associated sequence of random effects $\{\Delta_V, V \geq 1\}$, being the sum of V stochastically non-negative quantities, diverges as V diverges. To see the finite-sample consistency of $T(\Delta_V)$, let us consider the permutationally equivalent form of the test statistics

$$\begin{aligned} T^*(\Delta_V) &= \psi(V) \sum_{h=1}^{V} \sum_{i=1}^{n_1} X^*_{h1i}(\delta_h)/S_h = \psi(V) \sum_{i=1}^{n_1} \sum_{h=1}^{V} X^*_{h1i}(\delta_h)/S_h \\ &= \sum_{i=1}^{n_1} Y^*_{1i}(\delta) = T''^*(0) + n_1 \Delta^*_V, \end{aligned}$$

where the $Y_{1i}(\delta)$, $i = 1, \ldots, n_1$, are univariate data transformations which summarize the whole set of information on effects δ collected by the V variables and $\Delta^*_V = \psi(V) \sum_{h \leq V} \delta^*_h / S_h$. The right-hand side expression shows that a multivariate test statistic is reduced to one univariate. It is worth noting that we do not ask that all δ_h are positive, what is important is that Δ_V diverges at least in probability as V diverges while $T''^*(0)$ is measurable. Therefore, T is unconditional finite-sample consistent at least in the weak form. It is also to be emphasized that it is not required that the V variables are independent, actually they can be dependent in any way, because their dependences are nonparametrically taken into consideration by the NPC procedure. What is important is that the distribution induced by $T(\mathbf{X}(\mathbf{0}))$ is measurable and that of $T(\mathbf{X}(\delta))$ diverges at least in probability. It is also important to observe that, since the statistics S_h are functions of the data, the resulting random effects Δ_V, being data dependent, are not independent on random deviates $\mathbf{Z}$.

3.3 Consistency of multi-sided test

In the previous chapter we introduced the multisided-test, a method useful to testing the presence of random effects. The test is given by the combination of two partial tests T_1 and T_2. Each partial test separately checks one side of deviation from H_0. In paragraph 2.3.4 we proved that the test is exact and unbiased. To prove its consistency in the usual way we should verify that the critical values $T_\alpha(\mathbf{X})$ are almost surely asymptotically finite for every $\alpha > 0$. This proof is not easy to obtain because asymptotically the test consists of an infinite sum of elements. We will see instead that, using the finite-sample consistency property, the proof is immediate.
Here we report the multi-sided test statistic for testing the sub-hypothesis $H_{01} : \left\{\Delta \overset{d}{=} 0\right\}$ against $H_{11} : \left\{\Delta \overset{d}{<} 0\right\}$.

$$T_1^* = \sum_{i=1}^{n} S\Big\{F_1^*(X_i) - F_2^*(X_i)\Big\} \left(\hat{F}(X_i)\left[1 - \hat{F}(X_i)\right]\right)^{-1/2} \tag{3.9}$$

where

$$S\{\omega\} = \begin{cases} \omega & \text{if } \omega > 0 \\ 0 & \text{if } \omega \leq 0, \end{cases}$$

We can rewrite the vector of observations

$$\mathbf{X}(\Delta) = \{X_{1i} = \mu + \Delta_i + Z_{1i},\ i = 1, \dots, n_1;\ X_{2i} = \mu + Z_{2i},\ i = 1, \dots, n_2\}$$

in the matrix form $\mathbf{Y}(\Delta)$ as in equation (3.6) and (3.8) whose rows are of the form $\mathbf{Y}_{ji}(\Delta) = \left\{Y_{1ji} = X_{j,v(i-1)+1}, \dots, Y_{hji} = X_{j,v(i-1)+h}, \dots, Y_{vji} = X_{j,vi}\right\}$, $j = 1, 2$, where $v \geq 1$, $n_1 = m_1 v$, $n_2 = m_2 v$ and $n = (m_1 + m_2)v$. As in examples above, we can rewrite the test (3.9) in the permutationally equivalent form:

$$\begin{aligned} T_1^* &= \frac{1}{n}\sum_{i=1}^{m_1+m_2}\sum_{h=1}^{v} S\Big\{F_1^*(Y_{hi}) - F_2^*(Y_{hi})\Big\}\left(\hat{F}(Y_{hi})\Big[1-\hat{F}(Y_{hi})\Big]\right)^{-1/2} \\ &= \frac{1}{v}\sum_{h=1}^{v} T_h^* \end{aligned}$$

where

$$T_h^* = \frac{1}{m_1+m_2}\sum_{i=1}^{m_1+m_2} S\Big\{F_1^*(Y_{hi}) - F_2^*(Y_{hi})\Big\}\left(\hat{F}(Y_{hi})\Big[1-\hat{F}(Y_{hi})\Big]\right)^{-1/2}$$

Where the random variables T_h, under H_0, are i.i.d. with 0 mean and finite variance. As n_1 or n_2 diverges, also v diverges so we can apply the Kolmogorov's strong law of large numbers (Lessi, 1993) which states that

$$\lim_{v\to\infty}\frac{1}{v}\sum_{h=1}^{v} T_h(0) \stackrel{a.s.}{=} 0$$

so the whole null distribution collapses towards 0 with probability one, hence for every α-value not smaller than the minimum attainable, the critical point of T_1 is zero. As shown in paragraph 2.3.4, in the alternative the statistic $T_1(\Delta)$ increases with the effect Δ and then falls in the critical region with probability one.

3.4 Simulation study

In this section we report some results of a simulation study performed with the goal to test the unconditional power behaviour of a two-sample multivariate test processed according to the direct combination of several partial tests. So the global test statistic is given by the form $T''_D = \sum_h T_h, T''^o_D = \sum_h T^o_h$ and $T''^*_{Dr} = \sum_h T^*_{hr}, r = 1,\ldots,B$ for the combined test, observed, and permutation values respectively. Hence the combined p-value is given by $\hat{\lambda}'' = \sum_r \mathbf{I}(T''^*_{Dr} \geq T''^o_D)/B$.

We consider a two-sample problem where there are $V \geq 1$ variables, $\mathbf{X} = (\mathbf{X}_h, h = 1,\ldots,V)$, $\mathbf{X}_h = \mathbf{X}_{h1} \uplus \mathbf{X}_{h2}$, where $\mathbf{X}_{h1} = (\delta_h + \sigma_h Z_{h1i}, i = 1,\ldots,n_1)$ are the observations of variable h on the first sample and $\mathbf{X}_{h2} = (\sigma_h Z_{h1i}, i = 1,\ldots,n_2)$ are the observations of variable h on the second sample, δ_h and σ_h are respectively the non-centrality parameter and scale coefficient of variable

h. The $Z_{hji}, j = 1, 2$ are the random errors generate with different, independent distributions.
We perform the multivariate one-sided test

$$H_0 : \left\{ \mathbf{X}_1 \stackrel{d}{=} \mathbf{X}_2 \right\} = \left\{ \bigcap_h X_{h1} \stackrel{d}{=} X_{h2} \right\}$$

against the dominance alternative $H_1 : \left\{ \mathbf{X}_1 \stackrel{d}{>} \mathbf{X}_2 \right\} = \left\{ \bigcup_h X_{h1} \stackrel{d}{>} X_{h2} \right\}$ and the two-sided test $H_0 : \left\{ \mathbf{X}_1 \stackrel{d}{=} \mathbf{X}_2 \right\} = \left\{ \bigcap_h X_{h1} \stackrel{d}{=} X_{h2} \right\}$ against the non-dominance alternative $H_1 : \left\{ \mathbf{X}_1 \stackrel{d}{\neq} \mathbf{X}_2 \right\} = \left\{ \bigcup_h X_{h1} \stackrel{d}{\neq} X_{h2} \right\}$.
For every simulations we used different combinations of the number of variables V, the sample size n_1 and n_2, the α-value and the non centrality parameter δ. In particular the following values are used:

- V set on the seven values 1, 2, 10, 20, 50, 100, 1000;
- $n_1 = n_2$ set on the four values 3, 5, 10, 20;
- α set on the six values 0.05, 0.1, 0.2, 0.3, 0.5, 0.8;
- δ set on the seven values 0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 1.

We replicated the study with 1000 Monte Carlo simulations and considered $B = 1000$ samples from the permutation sample space.
In the following presentation are reported only the result for $\delta = 0.2$ and for $\alpha = 0.05$. The full table are avaible on request to the author. We considered different distributions for the random variables Z_{hji} and for each distribution proper test statistic are used:

- Standard Normal Distribution: for $\sigma_h = 1$ (homoscedasticity) and for unilateral test:

$$T''^{*}_{N} = \sum_{h=1}^{V} \sum_{i=1}^{n_1} X^{*}_{h1i} = \sum_{i=1}^{n_1} \sum_{h=1}^{V} X^{*}_{h1i} = \sum_{i=1}^{n_i} T^{*}_{1i}$$

For non-directional test we used the statistic:

$$T''^{*}_{Nb} = \left(\frac{1}{n_1} \sum_{i=1}^{n_1} T^{*}_{1i} - \frac{1}{n_2} \sum_{i=1}^{n_2} T^{*}_{2i} \right)^2 .$$

V	$\delta = 0$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.3$	$\delta = 0.5$	$\delta = 0.7$	$\delta = 1.0$
1	0.049	0.054	0.063	0.081	0.103	0.18	0.269	0.433
2	0.053	0.066	0.082	0.131	0.166	0.277	0.443	0.662
10	0.059	0.098	0.139	0.264	0.415	0.726	0.922	0.998
20	0.051	0.083	0.142	0.347	0.577	0.928	0.993	1
50	0.059	0.128	0.271	0.639	0.912	1	1	1
100	0.048	0.169	0.398	0.882	0.995	1	1	1
1000	0.049	0.747	0.999	1	1	1	1	1

Table 3.1: Power of the T''_N test, $n_1 = n_2 = 5$, $Z_{hij} \sim N(0, 1)$, $\alpha = 0.05$

V	$\delta = 0$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.3$	$\delta = 0.5$	$\delta = 0.7$	$\delta = 1.0$
1	0.05	0.048	0.052	0.064	0.075	0.111	0.153	0.262
2	0.05	0.045	0.046	0.059	0.078	0.157	0.262	0.49
10	0.046	0.05	0.069	0.153	0.25	0.558	0.848	0.991
20	0.047	0.069	0.099	0.238	0.463	0.866	0.992	1
50	0.039	0.077	0.151	0.492	0.82	0.999	1	1
100	0.055	0.113	0.271	0.766	0.98	1	1	1
1000	0.059	0.578	0.983	1	1	1	1	1

Table 3.2: Power of the T''_{Nb} test, $n_1 = n_2 = 5$, $Z_{hij} \sim N(0, 1)$, $\alpha = 0.05$

where $T^*_{ji} = \sum_{h=1}^{V} X^*_{hji}$, $j = 1, 2$. The estimated power of the multivariate one-sided and two-sided test are reported respectively in Table 3.1 and 3.2, both for $\alpha = 0.05$ and with $n_1 = n_2 = 5$.

For $\sigma_h \neq \sigma_k$, $h \neq k$ (heteroscedastic variables) we define the permutationally invariant square sum of deviation for variable h as:

$$SS(X_h) = \sqrt{\sum_{j=1}^{2} \sum_{i=1}^{n_j} X^2_{hji} - n\overline{X}^2_h}$$

where $\overline{X}_h = \frac{1}{n} \sum_{j=1}^{2} \sum_{i=1}^{n_j} X_{hji}$. The permutation test statistic becomes:

$$T''^*_N(\sigma) = \sum_{i=1}^{n_1} \sum_{h=1}^{V} \frac{X^*_{hji}(\sigma)}{SS(X_h)} = \sum_{i=1}^{n_1} T^*_{1i}(SS)$$

The two-sided test now becomes:

$$T''^*_{Nb}(\sigma) = \left(\frac{1}{n_1} \sum_{i=1}^{n_1} T^*_{1i}(SS) - \frac{1}{n_2} \sum_{i=1}^{n_2} T^*_{2i}(SS) \right)^2$$

where

$$T^*_{ji}(SS) = \sum_{h=1}^{V} \frac{X^*_{hji}(\sigma)}{SS(X_h)}$$

V	$\delta=0$	$\delta=0.05$	$\delta=0.1$	$\delta=0.2$	$\delta=0.3$	$\delta=0.5$	$\delta=0.7$	$\delta=1.0$
1	0.05	0.059	0.069	0.082	0.103	0.161	0.247	0.404
2	0.055	0.063	0.076	0.104	0.147	0.27	0.414	0.666
10	0.047	0.078	0.112	0.209	0.341	0.685	0.914	0.998
20	0.056	0.098	0.174	0.365	0.582	0.92	0.997	1
50	0.03	0.092	0.238	0.607	0.891	0.999	1	1
100	0.055	0.19	0.406	0.872	0.993	1	1	1
1000	0.053	0.711	0.999	1	1	1	1	1

Table 3.3: Power of the $T''_N(\sigma)$ test, $n_1 = n_2 = 5$, $Z_{hij} \sim N(0,1)$, $\alpha = 0.05$

V	$\delta=0$	$\delta=0.05$	$\delta=0.1$	$\delta=0.2$	$\delta=0.3$	$\delta=0.5$	$\delta=0.7$	$\delta=1.0$
1	0.044	0.045	0.046	0.051	0.064	0.108	0.16	0.258
2	0.045	0.046	0.047	0.053	0.072	0.156	0.244	0.444
10	0.043	0.047	0.067	0.142	0.255	0.57	0.844	0.988
20	0.045	0.054	0.079	0.188	0.401	0.833	0.979	1
50	0.059	0.078	0.149	0.442	0.801	0.998	1	1
100	0.047	0.1	0.269	0.73	0.975	1	1	1
1000	0.052	0.574	0.987	1	1	1	1	1

Table 3.4: Power of the $T''_{Nb}(\sigma)$ test, $n_1 = n_2 = 5$, $Z_{hij} \sim N(0,1)$, $\alpha = 0.05$

for $j = 1, 2$. The estimated power of the one-sided and two-sided test with heteroscedastic variables are reported respectively in Table 3.3 and 3.4, both for $\alpha = 0.05$ and with $n_1 = n_2 = 5$. In Figure 3.1 we report the power of the two-sided test with homoscedastic and heteroscedastic variables for different number of variables, normal error, $\delta = 0.2$. The power of the two tests are very similar.

- Student-t with two degree of freedom distribution. For one-sided and two-sided test with homoscedastic variables we can use the same test statistics as before and so $T''^{*}_{t} = T''^{*}_{N}$ and $T''^{*}_{tb} = T''^{*}_{Nb}$.. In Table 3.5 and 3.6 are reported the estimated power functions with Student-t errors. The remaining settings are as before.

  Since the Student's t_2 distribution has infinite second moment we can't use the $SS(X_h)$ statistic to standardize the variables. In place of $SS(X_h)$ we can use the sum of absolute deviates from mean:

$$S(X_h) = \sum_{j=1}^{2} \sum_{i=1}^{n_j} \left| X_{hji} - \overline{X}_h \right|$$

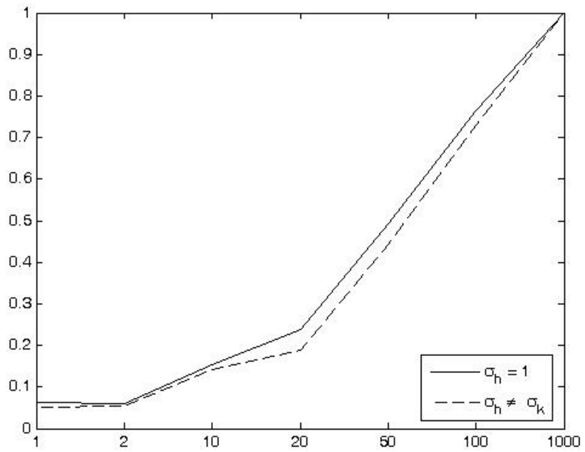

Figure 3.1: The two-sided tests with normal error

V	$\delta = 0$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.3$	$\delta = 0.5$	$\delta = 0.7$	$\delta = 1.0$
1	0.047	0.049	0.053	0.067	0.081	0.126	0.179	0.246
2	0.051	0.057	0.066	0.087	0.11	0.162	0.23	0.335
10	0.046	0.058	0.076	0.106	0.146	0.251	0.414	0.609
20	0.052	0.068	0.087	0.15	0.223	0.399	0.583	0.793
50	0.05	0.085	0.119	0.202	0.326	0.613	0.801	0.928
100	0.067	0.104	0.146	0.305	0.475	0.779	0.916	0.979
1000	0.049	0.163	0.417	0.855	0.966	0.993	0.997	0.999

Table 3.5: Power of the T_t'' test, $n_1 = n_2 = 5$, $Z_{hij} \sim t_2$, $\alpha = 0.05$

V	$\delta = 0$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.3$	$\delta = 0.5$	$\delta = 0.7$	$\delta = 1.0$
1	0.04	0.042	0.045	0.05	0.053	0.066	0.099	0.16
2	0.042	0.044	0.046	0.044	0.05	0.079	0.138	0.236
10	0.055	0.058	0.06	0.072	0.093	0.177	0.3	0.532
20	0.056	0.057	0.061	0.085	0.134	0.289	0.466	0.728
50	0.057	0.058	0.067	0.11	0.183	0.429	0.688	0.895
100	0.047	0.056	0.079	0.187	0.347	0.677	0.866	0.96
1000	0.046	0.1	0.281	0.765	0.944	0.988	0.996	0.999

Table 3.6: Power of the T_{tb}'' test, $n_1 = n_2 = 5$, $Z_{hij} \sim t_2$, $\alpha = 0.05$

V	$\delta = 0$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.3$	$\delta = 0.5$	$\delta = 0.7$	$\delta = 1.0$
1	0.054	0.057	0.064	0.081	0.092	0.127	0.181	0.262
2	0.056	0.065	0.073	0.088	0.104	0.154	0.226	0.345
10	0.032	0.044	0.054	0.109	0.178	0.37	0.571	0.815
20	0.051	0.081	0.113	0.197	0.303	0.579	0.818	0.966
50	0.047	0.092	0.148	0.334	0.545	0.883	0.978	0.999
100	0.044	0.113	0.202	0.484	0.777	0.988	1	1
1000	0.056	0.347	0.811	1	1	1	1	1

Table 3.7: Power of the $T''_t(\sigma)$ test, $n_1 = n_2 = 5$, $Z_{hij} \sim t_2$, $\alpha = 0.05$

V	$\delta = 0$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.3$	$\delta = 0.5$	$\delta = 0.7$	$\delta = 1.0$
1	0.034	0.038	0.041	0.052	0.062	0.084	0.112	0.181
2	0.056	0.054	0.055	0.055	0.061	0.087	0.133	0.218
10	0.052	0.054	0.061	0.081	0.106	0.216	0.388	0.642
20	0.046	0.048	0.059	0.094	0.168	0.396	0.659	0.918
50	0.054	0.06	0.08	0.194	0.355	0.778	0.948	0.998
100	0.044	0.064	0.122	0.342	0.639	0.966	0.998	1
1000	0.05	0.241	0.673	1	1	1	1	1

Table 3.8: Power of the $T''_{tb}(\sigma)$ test, $n_1 = n_2 = 5$, $Z_{hij} \sim t_2$, $\alpha = 0.05$

So for heteroscedastic variables, we used the statistics:

$$T''^{*}_{t}(\sigma) = \sum_{i=1}^{n_1} \sum_{h=1}^{V} \frac{X^*_{h1i}(\sigma)}{S(X_h)} = \sum_{i=1}^{n_1} T^*_{1i}(S)$$

for one-side test and:

$$T''^{*}_{tb}(\sigma) = \left(\frac{1}{n_1} \sum_{i=1}^{n_1} T^*_{1i}(S) - \frac{1}{n_2} \sum_{i=1}^{n_1} T^*_{2i}(S) \right)^2$$

for two-side test where

$$T^*_{ji}(S) = \sum_{h=1}^{V} \frac{X^*_{hji}(\sigma)}{S(X_h)}$$

for $j = 1, 2$. In Table 3.7 and 3.8 we report the estimated power for these last test.

- Standard Cauchy. This distribution has no moment so we must use the sample median as location index and the median absolute deviation (MAD) as scale indicator to standardize the variables. For homoscedastic variables and for one-sided and two-sided test we used

V	$\delta=0$	$\delta=0.05$	$\delta=0.1$	$\delta=0.2$	$\delta=0.3$	$\delta=0.5$	$\delta=0.7$	$\delta=1.0$
1	0.031	0.037	0.041	0.049	0.054	0.077	0.094	0.125
2	0.049	0.052	0.056	0.072	0.088	0.115	0.16	0.227
10	0.05	0.058	0.077	0.104	0.171	0.318	0.487	0.705
20	0.055	0.074	0.089	0.148	0.247	0.509	0.739	0.919
50	0.04	0.075	0.118	0.293	0.467	0.829	0.966	0.998
100	0.046	0.092	0.169	0.42	0.738	0.972	1	1
1000	0.053	0.311	0.757	0.996	1	1	1	1

Table 3.9: Power of the T_C'' test, $n_1 = n_2 = 5$, $Z_{hij} \sim Cau(0,1)$, $\alpha = 0.05$

V	$\delta=0$	$\delta=0.05$	$\delta=0.1$	$\delta=0.2$	$\delta=0.3$	$\delta=0.5$	$\delta=0.7$	$\delta=1.0$
1	0.036	0.041	0.042	0.042	0.043	0.051	0.064	0.085
2	0.045	0.045	0.047	0.051	0.06	0.065	0.083	0.126
10	0.037	0.042	0.047	0.062	0.083	0.188	0.312	0.511
20	0.041	0.052	0.061	0.092	0.149	0.329	0.55	0.808
50	0.052	0.055	0.071	0.156	0.322	0.677	0.892	0.985
100	0.04	0.057	0.093	0.26	0.543	0.91	0.99	1
1000	0.056	0.189	0.591	0.978	1	1	1	1

Table 3.10: Power of the T_{Cb}'' test, $n_1 = n_2 = 5$, $Z_{hij} \sim Cau(0,1)$, $\alpha = 0.05$

respectively:

$$T_C''^{*} = \sum_{i=1}^{n_1} \widetilde{T_{1i}^{*}} \qquad T_{Cb}''^{*} = \left(\frac{1}{n_1} \sum_{i=1}^{n_1} \widetilde{T_{1i}^{*}} - \frac{1}{n_2} \sum_{i=1}^{n_1} \widetilde{T_{2i}^{*}} \right)^2$$

where $\widetilde{T_{ji}^{*}} = Me(X_{hji}^{*})$, $j = 1,2$ and Me is the median operator. The estimated power function of these two tests are in Table 3.9 and 3.10. To standardize the variable we use the index

$$MAD\,(X_h) = Me \left| X_{hi} - \widetilde{X_h} \right| ,$$

where $\widetilde{X_h} = Me\,[X_{hi}]$ calculated on the pooled data set. If we indicate with $T_{ji}^{*}\widetilde{(MAD)} = Me\left[X_{hji}^{*}/MAD(X_h)\right]$, $j = 1,2$, for non-homoscedastic variables and for one side and two side test we can use respectively:

$$T_C''^{*}(\sigma) = \sum_{i=1}^{n_1} T_{1i}^{*}\widetilde{(MAD)}$$

$$T_{Cb}''^{*}(\sigma) = \left(\frac{1}{n_1} \sum_{i=1}^{n_1} T_{1i}^{*}\widetilde{(MAD)} - \frac{1}{n_1} \sum_{i=1}^{n_2} T_{2i}^{*}\widetilde{(MAD)} \right)^2$$

V	$\delta = 0$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.3$	$\delta = 0.5$	$\delta = 0.7$	$\delta = 1.0$
1	0.036	0.041	0.042	0.042	0.043	0.051	0.064	0.085
2	0.045	0.045	0.047	0.051	0.06	0.065	0.083	0.126
10	0.037	0.042	0.047	0.062	0.083	0.188	0.312	0.511
20	0.041	0.052	0.061	0.092	0.149	0.329	0.55	0.808
50	0.052	0.055	0.071	0.156	0.322	0.677	0.892	0.985
100	0.04	0.057	0.093	0.26	0.543	0.91	0.99	1
1000	0.056	0.189	0.591	0.978	1	1	1	1

Table 3.11: Power of the $T_C''(\sigma)$ test, $n_1 = n_2 = 5$, $\alpha = 0.05$

V	$\delta = 0$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.3$	$\delta = 0.5$	$\delta = 0.7$	$\delta = 1.0$
2	0.046	0.044	0.041	0.043	0.039	0.051	0.072	0.101
5	0.045	0.048	0.051	0.061	0.08	0.134	0.237	0.376
10	0.055	0.059	0.064	0.097	0.125	0.262	0.424	0.696
20	0.052	0.052	0.064	0.117	0.195	0.452	0.721	0.928
50	0.055	0.065	0.094	0.251	0.455	0.845	0.967	1
100	0.045	0.073	0.157	0.422	0.727	0.983	0.999	1
500	0.048	0.149	0.477	0.963	1	1	1	1
1000	0.041	0.298	0.779	1	1	1	1	1

Table 3.12: Power of the $T_{Cb}''(\sigma)$ test, $n_1 = n_2 = 5$, $\alpha = 0.05$

In Table 3.11 and 3.12 the estimated power of the $T_C''(\sigma)$ and $T_{Cb}''(\sigma)$ respectively. In Figure 3.2 we report the power of the two-sided test with homoscedastic and heteroscedastic variables for different number of variables with Cauchy distributed errors, $\delta = 0.2$. Again, the power of the two tests are very similar.
For the two-sided test with heteroscedastic variables we use another kind of statistic

$$T_{Cb}''^{*}(\sigma)_{Me} = \left(Me\left[T_{1i}^{*}(\widetilde{MAD})\right] - Me\left[T_{2i}^{*}(\widetilde{MAD})\right]\right)^2$$

the power obtained with this statistic is reported in Table 3.13, and in Figure 3.3 we report the power of the statistics $T_{Cb}''^{*}(\sigma)$ and $T_{Cb}''^{*}(\sigma)_{Me}$ both obtained with $\delta = 0.2$. In Figure 3.4 we report a similar comparison, obtained with t_2-Student distributed heteroscedastic random errors, between the statistic $T_{tb}(\sigma)$ and

$$T_{tb}''^{*}(\sigma)_{Me} = \left(Me\left[T_{1i}^{*}(S)\right] - Me\left[T_{2i}^{*}(S)\right]\right)^2$$

Clearly the statistics $T_{Cb}''(\sigma)_{Me}$ and $T_{tb}''(\sigma)_{Me}$ are not associative as using the median operator instead of the mean as in statistics $T_{Cb}''(\sigma)$ and $T_{tb}''(\sigma)$ anyway, the power of the test converges quickly to 1 as

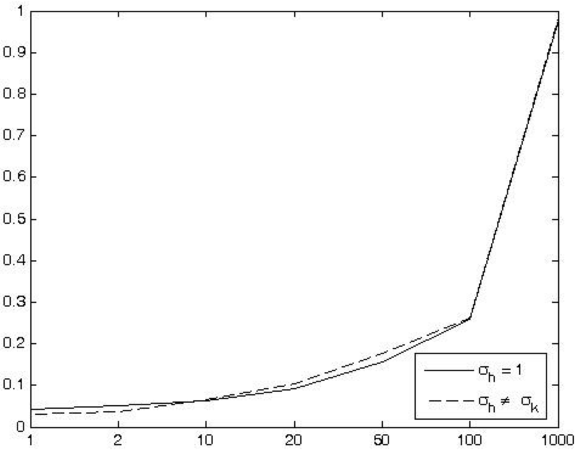

Figure 3.2: The two-sided tests with Cauchy distributed errors

V	$\delta = 0$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.3$	$\delta = 0.5$	$\delta = 0.7$	$\delta = 1.0$
2	0.042	0.043	0.047	0.048	0.058	0.058	0.079	0.11
5	0.052	0.056	0.054	0.067	0.089	0.148	0.22	0.344
10	0.045	0.048	0.051	0.067	0.108	0.233	0.398	0.628
20	0.058	0.063	0.076	0.12	0.198	0.429	0.66	0.902
50	0.042	0.049	0.085	0.231	0.427	0.785	0.955	0.992
100	0.046	0.07	0.136	0.372	0.672	0.963	0.997	1
500	0.045	0.166	0.46	0.949	0.999	1	1	1
1000	0.043	0.253	0.743	1	1	1	1	1

Table 3.13: Power of the $T''_{Cb}(\sigma)_{Me}$ test, $n_1 = n_2 = 5$, $\alpha = 0.05$

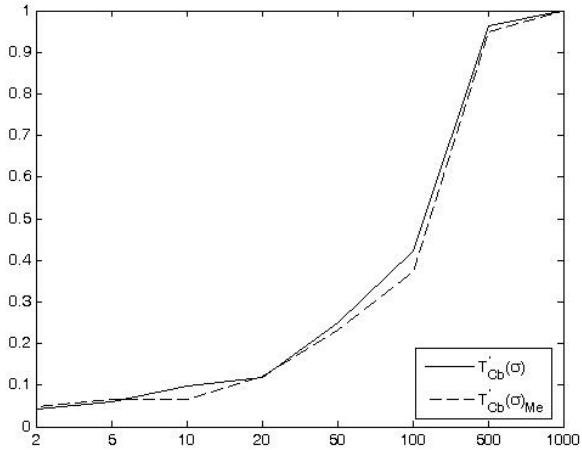

Figure 3.3: Comparison of statistics $T''_{Cb}(\sigma)$ and $T''_{Cb}(\sigma)_{Me}$

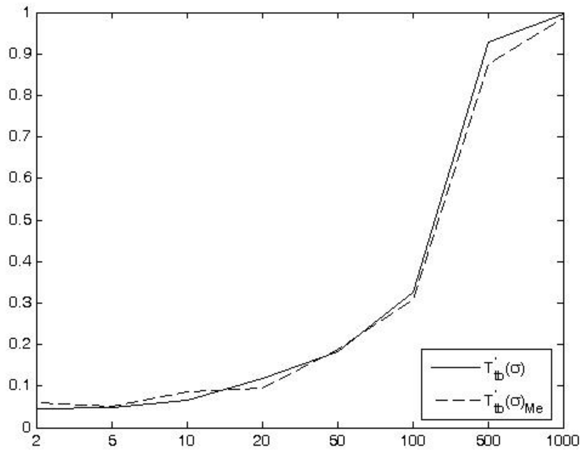

Figure 3.4: Comparison of statistics $T''_{tb}(\sigma)$ and $T''_{tb}(\sigma)_{Me}$

V	$\delta = 0$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.3$	$\delta = 0.5$	$\delta = 0.7$	$\delta = 1.0$
2	0.054	0.048	0.053	0.068	0.077	0.142	0.254	0.481
5	0.058	0.051	0.045	0.073	0.12	0.273	0.505	0.812
10	0.039	0.04	0.067	0.128	0.239	0.553	0.838	0.984
20	0.054	0.053	0.081	0.201	0.419	0.842	0.989	1
50	0.035	0.059	0.139	0.483	0.828	0.998	1	1
100	0.047	0.108	0.27	0.782	0.986	1	1	1
500	0.048	0.369	0.871	1	1	1	1	1
1000	0.047	0.568	0.983	1	1	1	1	1

Table 3.14: Power of the two-sided test, with mixture of a fixed and a random effect give by $\Delta_t = 0.5\Delta_{t-1} + e$, $e \sim N(0, 0.1)$, $n_1 = n_2 = 5$, $Z_{hij} \sim N(0, 1)$, $\alpha = 0.05$

V	$\delta = 0$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.3$	$\delta = 0.5$	$\delta = 0.7$	$\delta = 1.0$
2	0.123	0.11	0.098	0.071	0.056	0.041	0.048	0.128
5	0.033	0.035	0.032	0.058	0.104	0.243	0.453	0.764
10	0.063	0.044	0.026	0.016	0.027	0.124	0.37	0.822
20	0.012	0.022	0.037	0.101	0.225	0.623	0.919	1
50	0.03	0.067	0.134	0.425	0.783	0.997	1	1
100	0.025	0.074	0.224	0.701	0.958	1	1	1
500	0.359	0.032	0.02	0.815	1	1	1	1
1000	0.03	0.089	0.733	1	1	1	1	1

Table 3.15: Power of the two-sided test, with mixture of a fixed and a random effect give by $\Delta_t = 0.5\Delta_{t-1} + e$, $e \sim N(0, 1)$, $n_1 = n_2 = 5$, $Z_{hij} \sim N(0, 1)$, $\alpha = 0.05$

for the associative statistic. This convergence suggests the validity of the finite sample consistency also for non-associative statistics. In this work, however, we will not go further in this direction. This result also suggests that outside the exponential family, the sample mean is not necessarily the best choice because the statistic is not minimal sufficient.

- Mixture of a fixed and random correlated effects. In this simulations we add to the fixed effect an autocorrelated part given by the AR(1) process $\Delta_t = 0.5\Delta_{t-1} + e$, where e is a Normal innovation with mean 0 and with two different variances: $\sigma_{AR} = 0.1$ and $\sigma_{AR} = 1$. With this kind of processes we want to study the behaviour of the power of the test when the effects are in some way dependent. We performed the two-sided test with Normal(0,1) errors and $\sigma_h \neq \sigma_k$, $h \neq k$. The statistic used is the $T^*_{Nb}(\sigma)$. In Table 3.14 and 3.15 the estimated power of these two-sided tests with $\sigma_{AR} = 0.1$ and $\sigma_{AR} = 1$ respectively. When

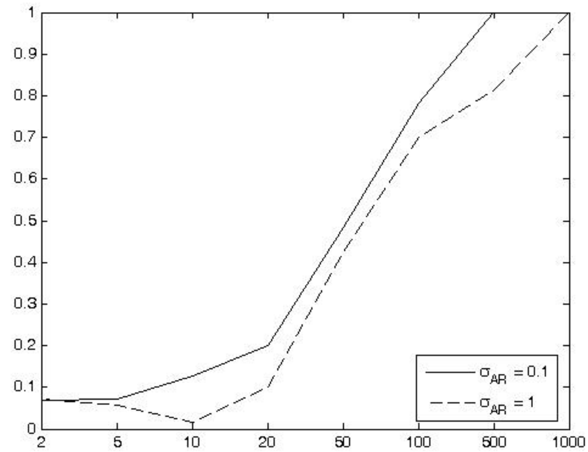

Figure 3.5: The two-sided tests with random effect

the variance of the AR process is greater is evident a non monotonically convergence of the power to one. This behaviour is due to the major noise introduced by the AR process. This situation is evident in Figure 3.5.

3.5 Conclusion

The finite sample consistency is a very important property which should be taken into account by experimenter when defining the design of the observational or experimental study. The simulation study confirmed what we have seen theoretically. Of course, it has to be underlined that only informative variables allow us to gain in power. With the NPC approach we can deal with situations where the number of variables is considerably larger than the number of observations. However, in these contexts the problem of multiplicity immediately arises. We will discuss about this topic in the next chapter.

Chapter 4

Nonparametric Weighted Step Down Holm Method with heteroscedastic variables

4.1 Introduction

In previous chapters we saw how the permutation methods deal with issues where the number of variables to be treated is far greater than the number of observations. In previous chapters the focus is placed on the global test obtained by the combination of partial tests. In this chapter we will instead consider the partial individual tests, we will see the problems of multiplicity and we will propose a permutation-based test procedure controlling the family wise error rate (FWE) by Weighted Step Down Holm methods (WSDH). It is shown that in this contest the choice of the weights must be permutation invariant. By a simulation study we "controlled" that the weights chosen as function of the variance of the pooled data set are good also for heteroscedastic variables.

4.2 The multiple testing problem

The issue of multiplicity control occurs in any situation where a problem is structured into more than one statistical test. This situation occurs very frequently in practice and there is an increasing tendency among researchers to analyze complex data sets from many viewpoints, formulating and testing myriads of hypotheses. In many cases a global multivariate test (e.g. when comparing two independent or dependent groups) is not sufficient for the experimenter who wishes to know which of the variables takes part in the observed effects. In this article, therefore, we consider statements about individual null hypothesis $H_{01}, \ldots, H_{0V}$, rather than just the global null hypothesis $H_0 = \bigcap_h H_{0h}$. A major drawback of multiple testing is the greatly increased probability of declaring "false significances", or statistically significant associations where none exists in reality. A related negative feature is that it is very easy to overstate the evidence for a particular association if the statistical test that best supports a given hypothesis is chosen. One solution for solving the multiplicity dilemma is to make the individual tests more conservative, i.e. to arrive at rejecting H_{0h} with more difficulty. Such a procedure is called a Multiple Testing Procedure (MTP). MTPs are commonly devised to control the Family-wise Error Rate (FWE). The strong form of FWE is the probability of rejecting any true null hypothesis H_h contained in a subset of true null hypotheses S; stated formally:

$$\text{FWE}(S) = \Pr(\text{Reject all least one } H_{0h},\, h \in S | H_{0h} \text{ is true for all } h \in S).$$

A simultaneous test procedure is said to control the FWE in the strong sense if $\text{FWE}(S) \leq \alpha$ for any subset S of hypotheses that happens to be true. Various MTPs have been proposed to control FWE. An overview can be found in Hochberg and Tamhane, 1987. Closed testing and step-wise methods are particularly popular because of their improved power (Marcus et al., 1976).

Here we consider a nonparametric permutation approach applied to step-down weighted methods. Weighted methods are useful when some H_h are more important than others. For example, main effect tests might be considered more important than interactions, primary endpoints in clinical trials might be considered more important than secondary endpoints, and so on.

4.3 Weighted step-down method

The simplest and the first weighted multiple testing procedure is the weighted single-step Bonferroni (WSSB) method (Westfall and Krishen, 2001): reject H_{0h} if $p_h \leq w_h \alpha$, where p_h is the unadjusted p-value of hypothesis H_h and w_h is the weight assigned to hypothesis H_h, $w_h \geq 0$, $\sum w_h = 1$.

Holm developed a weighted step-down testing method using the Bonferroni inequality and the min p_h/w_h statistic. Firstly we consider Holm's original step-down (SDH) method then we extend it to the weighted form. Given a set of p-values sorted in increasing order $p_{(1)} \leq \ldots \leq p_{(V)}$ corresponding to null hypotheses $H_{(1)}, \ldots, H_{(V)}$, hypothesis $H_{(k)}$ is rejected under the SDH method if $p_{(h)} \leq \alpha/(V - h + 1)$, for all $h = 1, \ldots, k$. The intuitive rationale is as follows: once $H_{(1)}$ has been rejected using Bonferroni critical value α/V,

we should believe that $H_{(1)}$ is false. Thus, there are only $V-1$ hypotheses which might still be true, implying the critical value $\alpha/(V-1)$ for $H_{(2)}$, and so on.
The method is popular because it is uniformly more powerful than the single-step Bonferroni method and yet retains control of the FWE in the strong sense. However, in many circumstances, the various hypotheses are not equally important. For example, in a two-way ANOVA model, main effect contrasts might be considered more important than interaction contrasts. If so, it is reasonable to allocate larger weights to the tests of primary importance. In this sense Holm extended his step-down testing method to incorporate weights as follows: once the weights have been assigned sort the weighted p-values $q_h = p_h/w_h$ into increasing order $q_{(1)} \leq q_{(2)} \dots \leq q_{(V)}$, where $q_{(k)} = q_{h_k}$ and h_k denotes the index of the kth ordered weighted p-value. Define the set $S_k = h_k, \dots, h_V$, $k = 1, \dots, V$. By letting $H_{(k)}$ denote the hypothesis corresponding to $q_{(k)}$, the weighted step-down Holm (WSDH) method rejects $H_{(k)}$ if $q_{(h)} \leq \alpha/\sum_{k\in S_h} w_k$ for all $h = 1, \dots, k$. When the weights are all equal to $1/V$, the method reduces to the ordinary SDH method.

4.4 Permutation WSDH

We consider a two-sample test assuming a model with fixed additive effects:

$$X_{hji} = \mu_h + \delta_{hj} + \sigma_{hj} Z_{hji} \tag{4.1}$$

where X_{hji} indicate the ith observation, $i = 1, \dots, n_j$, from the sample $j = 1, 2$ of the variable $h = 1, \dots, V$, μ_h represents a population constant for the hth variable, δ_{hj} represents effect on the hth variable in sample j, and Z_{hji} are V-dimensional random errors, which are assumed to be exchangeable with respect to groups or samples, independent with respect to units, with null mean vector $E(\mathbf{Z}) = 0)$ and with unspecified distribution. σ_{hj} is the scale coefficient of variable h and may depend on the treatment. Note that we do not assume homoscedasticity among variables, as in Kropf *et al.*, 2004 and Westfall and Krishen, 2001. We wish to choose the weights w_h on the basis of the experimental data so no a priori knowledge is required. As it is well known, the weights must be permutationally invariant quantities in the sense that for all points in the permutation sample space $\mathcal{X}_{/\mathbf{X}}$ the weight of variable h is the same. The WSDH method is implemented as follows:

1. Calculate the p-values for the usual permutation two-sample two-sided test for each of the V variables.

2. For each variable h, determine the permutation invariant weight $w_h = s_h^\eta$, where s_h is a chosen permutation invariant statistic and η is a positive fixed coefficient.

3. Calculate the weighted p-values $q_h = p_h/w_h$ and sort the variables for increasing values: $q_{h_1} \leq q_{h_2} \leq \ldots \leq q_{h_V}$ or $q_{(1)} \leq q_{(2)} \leq \ldots \leq q_{(V)}$ respectively. Define the index sets $S_u = h_u, h_{u+1}, \ldots, h_V$, for $u = 1, 2, \ldots, V$.

4. The ordered hypothesis $H_{(u)}$ for $u = 1, 2, \ldots, V$ is rejected as long as $q_{(u)} \leq \alpha / \sum_{k \in S_u} w_k$.

We can prove that this procedure maintains the FWE in the strong sense. We follow the arguments used in Kropf *et al.*, 2004 for the Wilcoxon test. Let S_0 be the subset of variables that satisfy H_0 and h_0 the first variable under H_0 after the ordering of step 3 above. If the procedure controls the FWE, the null hypothesis for variable $X_{(h_0)}$ is accepted with probability $1-\alpha$ at least. If the procedure stops before reaching this variable, a rejection of any true null hypothesis is avoided. Let S_{h_0} be the set constructed as at point 3. Obviously $S_0 \subseteq S_{h_0}$ since both sets contain all variables fulfilling the true null hypotheses but S_{h_0} possibly contains other variables. Note that for a fixed $\mathcal{X}_{/\mathbf{X}}$, the weights w_h for $h = 1, \ldots, V$ are fixed too because they depend on the pooled sample data, and so are permutation invariant quantities. Hence the variable with $\min q_h$ is also fixed in $\mathcal{X}_{/\mathbf{X}}$ as well as the ordering subscripts $h_1, \ldots, h_V$. So the permutation test for this variable is the usual one. Conditional on $\mathcal{X}_{/\mathbf{X}}$ we have:

$$\Pr\left(q_{(h_0)} \leq \frac{\alpha}{\sum_{k \in S_{h_0}} w_k} | \mathcal{X}_{/\mathbf{X}}\right) \leq \Pr\left(q_{(h_0)} \leq \frac{\alpha}{\sum_{k \in S_0} w_k} | \mathcal{X}_{/\mathbf{X}}\right)$$

since $\sum_{k \in S_0} w_k \leq \sum_{k \in S_{h_0}} w_k$. The probability of declaring a test $h \in S_0$ significant is equivalent to:

$$\begin{aligned}
\Pr\left(\min_{h \in S_0} q_h \leq \frac{\alpha}{\sum_{k \in S_0} w_k} | \mathcal{X}_{/\mathbf{X}}\right) &= \Pr\left(\bigcup_{h \in S_0} \left(q_h \leq \frac{\alpha}{\sum_{k \in S_0} w_k} | \mathcal{X}_{/\mathbf{X}}\right)\right) \\
&= \Pr\left(\bigcup_{h \in S_0} \left(p_h \leq \frac{\alpha w_h}{\sum_{k \in S_0} w_k} | \mathcal{X}_{/\mathbf{X}}\right)\right) \\
&\leq \sum_{h \in S_0} \Pr\left(p_h \leq \frac{\alpha w_h}{\sum_{k \in S_0} w_k} | \mathcal{X}_{/\mathbf{X}}\right)
\end{aligned}$$

Where the latter is the well-known Bonferroni inequality. As sample sizes tend to infinity, attainable p-values become dense in the unit interval, so when H_{0h} is true, p_h becomes uniformly distributed in the interval $[0, 1]$ (Pesarin, 2001). Therefore under H_{0h}, $\Pr(p_h \leq c) \leq c$ for any constant c and thus:

$$\begin{aligned}\Pr\left(q_{(h_0)} \leq \frac{\alpha}{\sum_{k\in S_{i_0}} w_k}|\mathcal{X}_{/\mathbf{X}}\right) &\leq \sum_{h\in S_0} \Pr\left(p_h \leq \frac{\alpha w_h}{\sum_{k\in S_0} w_k}|\mathcal{X}_{/\mathbf{X}}\right) \\ &\leq \sum_{h\in S_0} \frac{\alpha w_h}{\sum_{k\in S_0}} = \alpha\end{aligned}$$

The similarity property (see paragraph 2.3.2) of permutation tests in continuous non-degenerate situations is attained for almost all data set $\mathbf{X}$. This property allows us to extend the conditional inference to the unconditional inference so the inequality above is valid also unconditionally.

4.5 The choice of the weights

If we consider homoschedastic variables, the sample variance of the pooled data set is a good choice for the weights as shown in Kropf *et al.*, 2004. What happens if the variables are heteroscedastic? In our simulation study we consider a two-sample test with a data set composed of five observations per sample from $N(\mu_h, \sigma_h^2)$ where μ_h is a $U(0, 10)$ and σ_h^2 is a $U(1, 10000)$. Since the variables are heteroscedastic the non-centrality parameters δ_h are set $\delta_h = \delta\sqrt{\sigma_i^2}$, with $\delta = 2$. We wish to check the power behaviour. In the literature there have been several definitions of power given for multiple testing. We consider the total power i.e. the probability of detecting all true alternatives. In Figure 4.1 it is shown the behaviour of the sample variances for each of the generated 100 variables. The first 10 variables are generated under H_1, the other 90 under H_0. The sample variance appears to be a good indicator to identify the variables under H_1 since for these variables it assumes generally greater values than the variables under H_0.
Figure 4.2 shows the power of the test evaluated after 1000 runs of Monte Carlo Simulation. It also shows the type I error which is under control for each value of η.

4.6 Conclusions

The accurate interpretation of statistical data is a concern of physicians, politicians sociologists, engineers, and scientists everywhere. A problem that

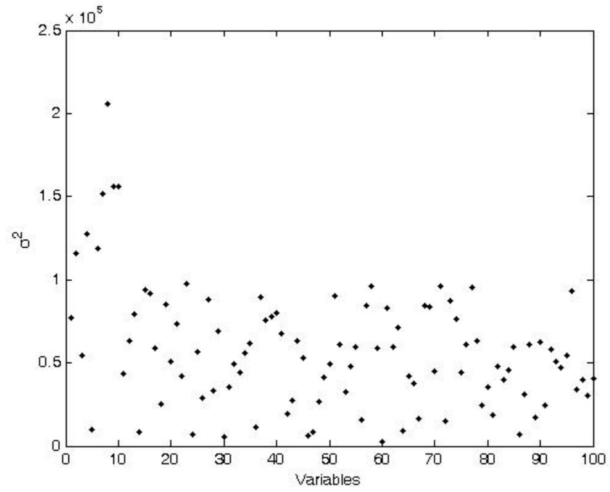

Figure 4.1: Sample variances of the variables

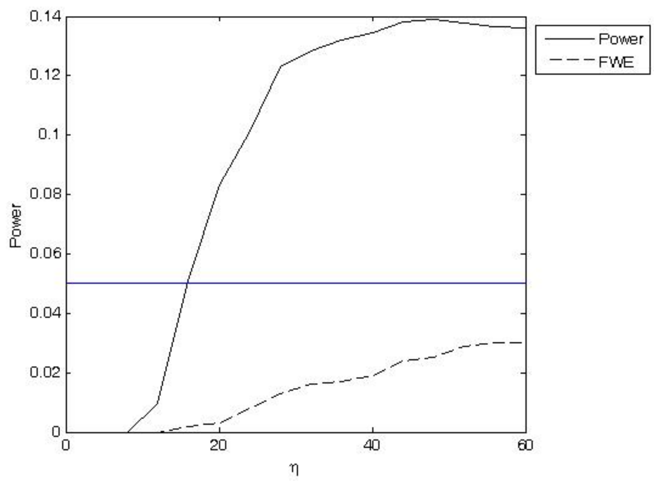

Figure 4.2: Power of the test

recurs in research studies, on which these professionals depend, is the extensive analysis of data. Modern computing equipment makes extensive analysis quite inexpensive, relative to the cost of obtaining the data. Once the data is available and on the computer, researchers question and analyze it from every possible angle, to miss no information. The result of such extensive data analysis, or "data mining", is the increased chance of inaccurately interpreted data. In particular, spurious results may be claimed to be real. For this reason some kind of corrections of the p-value obtained is necessary.
In the previous chapter we saw that in the non-parametric framework, the addition of informative variables increases the power of the combined test. Even the combination of partial tests can be seen as a kind of correction to solve the multiplicity dilemma, even if the combined test does not reveal what partial tests are actually significant. In this chapter we extended the WSDH method with data-driven weights to the permutation framework. The simulation study shows that even with heteroscedastic variables, if the non-centrality parameters are in terms of signal to noise ratio, the sample variance is still an acceptable indicator for the construction of the weights.
If $V = 1$, the application of methods for the multiplicity control to multi-sided test allows to verify which of the two tails, if not both, of the distributions of the random effect Δ are active. If $V > 1$ is possible to identify which variables are really effected by the treatment. In a three-dimensional surfaces analysis, this allows for the possibility of identifying the areas in which the treatment has produced an effect. This type of analysis will be discussed in detail in the next chapter.

Chapter 5

Nonparametric Functional Data Analysis of 3-D surfaces

5.1 Introduction

The theoretical aspects presented in previous chapters here are used to solve a testing problem connected to three-dimensional surface analysis. This chapter opens with a brief description of the surgical problem that motivated the research, then we discuss some concepts relating to functional data from which we borrow the theoretical rationale for our choice of representing three-dimensional surfaces by means of Radial Basis Functions. Within this type of representation the application of permutation tests becomes easily justified.

5.2 A three-dimensional data in orthognathic surgery

5.2.1 Oral-maxillofacial surgery

Dentofacial malformations are pathologies of the shape and size of the face. The oral-maxillofacial surgeon who attempts to correct these by deformations and segmentation of the facial skeleton into parts and recomposes them in order to modify the size, the form, and location of typical regions. A variation in the skeletal support induces a modification of the nearby soft tissues and thus of the facial aesthetics. Up until a few years ago the guiding principle behind the reconstruction of the maxilla was represented by the occlusion and by the mean statistical measure of the skeletal dimensions typical of the population to which the patient belonged. Experience has shown that this clinical principle does not necessarily mean that the soft tissues analogously achieve the standards and thus that facial aesthetics improves. At present time surgeons prefer to define - in a preliminary way- the aesthetic goal to achieve, that is the desired form of the soft issues. The movements of the teeth and maxilla that are necessary to obtain that goal are then decided. In the past, every clinical decision concerning the direction and the quantity of required skeletal movements was based upon the surgeon's intuition and experience. Recently, 2-dimensional software capable of modifying the features of the face in relation to dentoskeletal movements have joined the conventional support instruments used in clinics, such as models of the dental arches, radiographs and face photos. Naturally the basic problem behind these procedures is the accuracy of the prediction that they produce. Simulations have been shown to be moderately accurate when the surgical movements shift the skeletal hard tissue and thus the related soft tissues in a forward-backward direction. The inadequacy of the 2-dimensional approach for the prediction of variations induced by surgical treatment and the need of utilizing a representation and modification patterns containing 3-Dimensional data have become evident to the oral-maxillofacial surgeon.

5.2.2 The 3-Dimensional approach

A number of methods of facial reconstruction and 3-Dimensional analyses have been proposed in the literature. The technology based on the Structured Light Systems, such as those based on laser scanning (Moss et al., 1994), is capable of faithfully reproducing the features of the facial surface so that these can be evaluated. Its clinical applications, thus, permit an accurate evaluation of the modifications of the pre- and post-operative surfaces. In particular a complete laser scan of the face of a patient is composed by a collection of approximately 1,500,000 three-dimensional points which permit details up to order of 0.5 mm. Even with this new methodology the development, on statistical basis, of a prediction model of the modifications induced on soft tissues consequent to skeletal movements in maxillofacial surgery is far from being easy to achieve. It should be remembered that the soft tissues present non linear modifications with respect to the movements of the underlying skeletal structures connected to:

1. the type of surgical intervention to be carried out;
2. the diversity of the individual patient's response to the surgical trauma undertaking the same intervention;
3. the personal experience of the surgeon.

The complexity and the intrinsic non linearity of the induced modifications justify the decision to derive the possible correlations from statistical analysis. The problem is so complex that it is still unclear which areas of the soft tissues are really involved by the skeletal movements in the surgery. In particular, the subjectivity of patient response, as indicated in point 2 above, makes it difficult to assess the direction of changes in some areas. So, the effect of the same maxillofacial surgery, in one area, may be positive on some subjects and negative on others.
In the following paragraphs, we introduce the functional data in $\mathbb{R}$, then we extend the results from $\mathbb{R}$ to $\mathbb{R}^3$ to represent three-dimensional surfaces, since the scattered data supplied form the laser scan are considered observations of an underlying function $s : \mathbb{R}^3 \to \mathbb{R}$. Then, as the representation of a surface does involve a large (sometimes very large) set of data for each individual, we must apply to this representation the multi-sided tests, the NPC method for multivariate testing, and the multiple testing analysis on selected areas.

5.3 Functional Data

5.3.1 Some properties of functional data

The basic philosophy of functional data analysis is to think of observed data functions (typically curves) as singles, rather than merely as a sequence of individual observations. The term functional in reference to observed data refers to the intrinsic structure of the data rather than to their explicit form. In practice, functional data are usually observed and recorded discretely as v pairs (t_h, y_h), and y_h is a snapshot of the function at point t_h, possibly blurred by measurement error. Not always time is the continuum over which functional data are recorded; certainly other continua may be involved, such as spatial position, frequency, weight, and so forth.
What would it mean for functional observation to be known in functional form s, where s, in this contest, refers to a function? We do not mean that s is actually recorded for every value of t, because that would involve storing an uncountable number of values. Rather, it means that the existence of a function s giving rise to the observed data is assumed. In addition, for typical kind of analyses we have to carry out, we assume that the underlying function s is smooth, so that a pair of adjacent data values y_h and y_{h+1} are necessarily linked together to some extent and unlikely to be too different from each others. If this smoothness property do not apply, there would be nothing much to be gained by treating the data as functional rather than just multivariate observations.

Clearly, working with facial surfaces the assumption that the underlying function is smooth is justified. Smooth usually means that function s possesses one or more derivatives, indicate by Ds, D^2s and so on, so that $D^m s$ refers to the derivative of order m and $D^m s(t)$ is the value of that derivative at point t. We will usually want to use discrete data $y_h, h = 1, \dots, v$ to estimate the function s and at same time a certain number of its derivatives.
The actual observed data, however, may not be at all smooth due to the presence of noise or measurement error. Some of this externally induced variation may indeed have all the characteristics of noise, that is, be formless and unpredictable, or it may be high-frequency variation that we could in principle model, but for practical reasons choose to ignore. Sometimes this noise level is a tiny fraction of the size of the function that it reflects, and then we say that the signal-to-noise ratio (S/N ratio) is high. However, higher levels of variation of the y_h around the corresponding $s(t_h)$'s can make extracting a stable estimate of the function and some of its derivatives a real challenge.
Clearly we are concerned with a collection or sample of functional data, rather than just a single function s: one function for each sampled individual. Specifically, using the same indices used in previous chapters, the observation of the function of subject i, s_i might consist of v_i pairs (t_{hi}, y_{hi}), $h = 1, \dots, v_i$. If we consider the observations pre- and post-surgery, like in one-sample paired problems, we can indicate the underlying functions s_{1i} and s_{2i} whose observations are respectively the pairs (t_{h1i}, y_{h1i}), and (t_{h2i}, y_{h2i}), $h = 1, \dots, v_i$, $i = 1, \dots, n$. Until it is necessary, we will use the simplified notation s.

5.3.2 The interplay between smooth and noisy variation

Smoothness, in the sense of possessing a certain number of derivatives, is a property of the true underlying (latent) function s. Of course, it may not be at all obvious in the raw data vector $\mathbf{y} = (y_1, \dots, y_v)$, owing to the presence of observational error or noise that is superimposed on the underlying signal as a consequence of the measurement process, how to separate noise from signal. We express this in notation as:

$$y_h = s(t_h) + \epsilon_h \tag{5.1}$$

where the noise, disturbance, error, perturbation or otherwise exogenous term ϵ_j contributes a roughness to the raw data, for which, as usual we assume $\mathbb{E}\left[\epsilon_j\right] = 0$. Of course, in this model signal and noise are confounded.

Thus, one of the task in representing the raw data as functions may be to attempt to filter out this induced noise as efficiently as possible, instead to try separating them.
When comparing sample functions discretized according to (t_{hi}, y_{hi}), $h = 1, \ldots, v_i$ we meet some more problems:

1. the number of really observed points is not necessarily invariant with respect to individuals: v_i are not constant numbers;
2. points t_{hi} and t_{hj} are generally not synchronized in the sense that y_{hi} and y_{hj} do not correspond to observations on the same point in the surface and for all individuals, so that they are not directly comparable (e.g., points of two photos taken in different occasions on the same subject cannot be synchronized by just considering their ordering in the digital sequence).

A direct implication of 2. is that we cannot directly compare curves by means of standard permutation multivariate tools based on their multivariate discretized representation, because of lack of synchronization of observed points. In fact, it is well-known that in multivariate comparisons it is compulsory to compare variables having the same name: weights with weights, speeds with speeds, etc. To this end we must represent observed curves by means of suitable series expansions so that we can compare their ordered coefficients which are then synchronized due to their ordering. Of course, assumed suitable smoothness property of s allow for series expansions.

5.3.3 Smoothing data using a basis system by least squares

A basis function system is a set of known functions ϕ_k that are mathematically independent of each other and that have the property that we can approximate arbitrarily well any function by taking a weighted sum or linear combination of a sufficiently large number K of these functions. The most familiar basis system of functions is the collection of monomials that are used to construct power series

$$1, t, t^2, \ldots, t^K,$$

or the well known Fourier series system

$$1, \sin(\omega t), \cos(\omega t), \sin(2\omega t), \cos(2\omega t), \ldots, \sin(K\omega t), \cos(K\omega t).$$

Basis function procedures represent a function s by a linear expansion

$$s(t) = \sum_{k=1}^{K} c_k \phi_k(t), \tag{5.2}$$

in terms of K known basis functions ϕ_k.
If our goal is to fit the discrete observations y_j, $j = 1, \dots, n$ using the model (5.1), by a basis function expansion for $s(t)$ of the form (5.2), we can use a simple linear smoother choosing the expansions c_k that minimizes the least squares criterion

$$\mathrm{SMSSE}(\mathbf{y}|\mathbf{c}) = \sum_{h=1}^{v} \left[y_h - \sum_{k=1}^{K} c_k \phi_k(t_h) \right]^2 . \tag{5.3}$$

for a given basis functions ϕ_k.
How to choose the order of the expansion K? The larger K, the better the fit to the data, but of course we then risk also fitting noise or variation that we wish to ignore, but if K is too small, we may miss some important aspects of the smooth function s that we are trying to estimate. This trade-off can be expressed in another way. For large values of K, the bias in estimating $s(t)$, that is

$$\mathrm{Bias}\left[\hat{s}(t)\right] = s(t) - \mathbb{E}\left[\hat{s}(t)\right],$$

is small. In fact, if the notion of additive errors having null expectation holds, then we know that the bias will be zero for $K = v$. But, one of the main reasons that we do smoothing is to reduce the influence of noise on the estimate $\hat{s}$. Consequently we are also interested in the variance of the estimate

$$\mathrm{Var}\left[\hat{s}(t)\right] = \mathbb{E}\left[\{\hat{s}(t) - \mathbb{E}\left[\hat{s}(t)\right]\}^2\right].$$

If $K = v$, this is almost certainly to be unacceptably high. Reducing variance leads to look for smaller values of K, but of course not so small as to make the bias unacceptable. The worse the signal-to-noise ratio in the data, the more reducing sampling variance will outweigh controlling bias. One way of expressing what we really want to achieve is mean squared error

$$\mathrm{MSE}\left[\hat{s}(t)\right] = \mathbb{E}\left[\{\hat{s}(t) - s(t)\}^2\right].$$

In most applications we can't actually minimize this quantity since $s(t)$ assumed to be unknown. However, an important equation in statistics link mean squared error to bias and sampling variance by the simple additive decomposition

$$\mathrm{MSE}\left[\hat{s}(t)\right] = \mathrm{Bias}^2 + \mathrm{Var}\left[\hat{s}(t)\right].$$

What this relation tells us is that it would be worthwhile to tolerate a little bias if the result is a big reduction in sampling variance. In fact, on the one hand, we wish to ensure that the estimated curve gives a good fit to the data. On the other hand, we do not wish the fit to be good if this results in a curve

s that is excessively "wiggly" or locally variable. A completely unbiased estimate of the function value $s(t_h)$ can be produced by a curve fitting y_h exactly, since this observed value is itself an unbiased estimate of $s(t_h)$ according to our error model. But any such curve must have high variance, manifested in the rapid local variation of the curve. MSE can often be dramatically reduced by sacrificing some bias in order to reduce sampling variance, and this is a key reason for imposing smoothness on the estimated curve. By requiring that the estimate vary only gently from one value to another, we are effectively "borrowing information" from neighbouring data values, thereby expressing our faith in the regularity of the underlying function s that we are trying to estimate. This pooling of information is what makes our estimated curve more stable, at cost of some increase in bias. The roughness penalty makes explicit what we sacrifice in bias to achieve an improvement MSE.

5.3.4 The penalized sum of squared errors fitting criterion

The square of the second derivative $[D^2 s(t)]^2$ of a function at t is often called its curvature at t, since a straight line, which has no curvature, has a zero second derivative. Consequently, a natural measure of a function's roughness is the integrated squared second derivative

$$\mathrm{PEN}_2(s) = \int \left[D^2 s(t)\right]^2 dt.$$

Highly variable functions can be expected to yield high values of $\mathrm{PEN}_2(s)$ because their second derivatives are large over at least some of the range of interest.

Now we need to modify the last squares fitting criterion (5.3) so as to allow the roughness penalty $\mathrm{PEN}_2(s)$ to play a role in defining the estimate of s, We define a compromise that explicitly trades off smoothness against data fit by defining the penalized residual sum of squares as

$$\mathrm{PENSSE}_\lambda(s|\mathbf{y}) = \sum_{h=1}^{v} \left[y_h - \sum_{k=1}^{K} c_k \phi_k(t_h) \right]^2 + \lambda \mathrm{PEN}_2(s),$$

Our estimate of the function is obtained by finding the function s that minimize $\mathrm{PENSSE}_\lambda(s|\mathbf{y})$ over the space of functions s for which $\mathrm{PEN}_2(s)$ is defined.

The parameter λ is a smoothing parameter that measures the rate of exchange between fit to the data, as measured by the residual sum of squares

in the first term, and variability of the function s, as quantified by $\mathrm{PEN}_2(s)$ in the second term. As λ becomes larger and larger, functions which are not linear must incur a more substantial roughness penalty through the term $\mathrm{PEN}_2(s)$, and consequently the composite criterion $\mathrm{PENSSE}_\lambda(s|\mathbf{y})$ must place more and more emphasis on the smoothness of s and less and less on fitting the data. For this reason, as $\lambda \to \infty$ the fitted curve s must approach the standard linear regression to the observed data, where $\mathrm{PEN}_2(s) = 0$. On the other hand, for small λ the curve tends to become more and more variable since there is less and less penalty placed on its roughness, and as $\lambda \to \infty$ the curve s approaches a function interpolating the data and satisfying $s(t_h) = y_h$ for all h. However, even in this limiting case the interpolating curve is not arbitrarily variable; instead, it is the smoothest twice-differentiable curve that exactly fits the data. Generally the basis functions used in one-dimensional case are the Fourier basis, B-spline basis and Wavelets. How these basis functions work in two- or three-dimensional spaces? The tensor product is the easiest way since it uses rectangular partition of the domain and thus it is a very natural extension of the univariate case. But if the domain is not rectangular the tensor product does not work. Also the multivariate extension of Functional Principal Components Analysis (Bosq, 2000) is not applicable since data are not synchronized (see below) to the same argument.
We must use other basis function for the three-dimensional surfaces as we will see in the following paragraph.

5.4 Representation of 3D surfaces with Radial Basis Function

5.4.1 Fitting an implicit function to a surface

We wish to find a function f which implicitly defines a surface M and satisfies the equation $f(\mathbf{t}_h) = 0$ where $\mathbf{t}_h \in \mathbb{R}^3$ for $h = 1, \ldots, v$ are points lying on the surface. In order to avoid the trivial solution that f is zero everywhere, off-surface points are appended to the input data and are given non-zero values. This gives a more useful interpolation problem: Find f such that

$$\begin{array}{lll} f(\mathbf{t}_h) = 0 & h = 1, \ldots, v & \text{on-surface points,} \\ f(\mathbf{t}_h) = y_h \neq 0 & h = v+1, \ldots, V & \text{off-surface points.} \end{array}$$

This still leaves the problem of generating the off-surface points $\mathbf{t}_h$ for $h = v + 1, \ldots, V$ and the corresponding values y_h. An obvious choice for f is a signed-distance function, where the y_h are chosen to be the distance to the closest on-surface point. Points outside the object are assigned positive

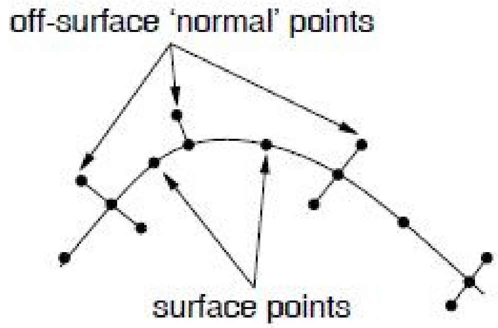

Figure 5.1: Off-surface points along surface normals.

values, while points inside are assigned negative values. These off-surface points are generated by projecting along surface normals as illustrated in Figure 5.1. Experience has shown that it is better to augment a data point with two off-surface points, one either side of the surface.

5.4.2 The Radial Basis Functions

Given a set of scattered data points pairs $(\mathbf{t}_h, y_h)$, $h = 1, \ldots, V$, where the points $h = 1, \ldots, v$ are zero-valued surface points and the points $h = v + 1, \ldots, V$ are non-zero off-surface points, we want to approximate the signed-distance function $f(\mathbf{t}_h) = y_h$, by an interpolating function $s(\mathbf{t})$.
If we consider the roughness penalty $\mathrm{PEN}_2(s)$ in $\mathbb{R}^3$ it becomes:

$$\begin{aligned}\mathrm{PEN}_2(s) &= \int_{\mathbb{R}^3} \left[\left(\frac{\partial^2 s(\mathbf{t})}{\partial t_1^2}\right)^2 + \left(\frac{\partial^2 s(\mathbf{t})}{\partial t_2^2}\right)^2 + \left(\frac{\partial^2 s(\mathbf{t})}{\partial t_3^2}\right)^2 \right. \\ &+ \left. 2\left(\frac{\partial^2 s(\mathbf{t})}{\partial t_1 t_2}\right)^2 + 2\left(\frac{\partial^2 s(\mathbf{t})}{\partial t_1 t_3}\right)^2 + 2\left(\frac{\partial^2 s(\mathbf{t})}{\partial t_2 t_3}\right)^2 \right] d\mathbf{t}\end{aligned}$$

In Duchon, 1977 is shown that the family of functions that minimize the $\mathrm{PEN}_2(s)$ among the functions with square integrable second derivatives has the form

$$s(\mathbf{t}) = p(\mathbf{t}) + \sum_{k=1}^{K} c_k \|\mathbf{t} - \mathbf{q}_k\|, \tag{5.4}$$

that is a particular form of a Radial Basis Function (RBF). In general, an RBF is a function of the form

$$s(\mathbf{t}) = p(\mathbf{t}) + \sum_{k=1}^{K} c_k \phi(\|\mathbf{t} - \mathbf{q}_k\|),$$

where:

- $s : \mathbb{R}^3 \to \mathbb{R}$ is the radial basis function,
- p is a low degree polynomial, typically linear or quadratic,
- c_k, $k = 1, \ldots, K$ are the coefficients,
- ϕ is a real valued function called the basis function, and $\| \bullet \|$ is the Euclidian norm in $\mathbb{R}^3$
- $\mathbf{q}_k$, $k = 1, \ldots, K$ are the RBF centers.

The RBF consists of a weighted sum of a radially symmetric basic function ϕ located at the centers $\mathbf{q}_k$ and a low polynomial p. RBF's are popular for approximate scattered data as the associated system of linear equations is guaranteed to be invertible under very mild conditions on the locations of the data points t_h (Carr et al., 1997). If we choose $\phi(r) = r$ we have the form (5.4) known as biharmonic spline.
If we impose the interpolation conditions $s(\mathbf{t_h}) = y_h$, for $h = 1, \ldots, V$ and we chose a polynomial $p(\mathbf{t_h}) = \beta_0 + \beta_1 t_1 + \beta_2 t_2 + \beta_3 + t_3$, where t_i, $i = 1, 2, 3$ are the elements of the vector $\mathbf{t}$, then the coefficients c_k of the (5.4) and of the polynomial $p(\mathbf{t_h})$ that minimize the $\mathrm{PEN}_2(s)$ can be found solving the linear system

$$\begin{bmatrix} \mathbf{A} & \mathbf{T} \\ \mathbf{T}' & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{c} \\ \beta \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix}, \tag{5.5}$$

where

$$\begin{aligned}
\mathbf{A} &= (a_{hk}) = (\|\mathbf{t_h} - \mathbf{q}_k\|), \\
\mathbf{T} &= \begin{bmatrix} 1 & t_{11} & t_{11} & t_{13} \\ \vdots & \vdots & \vdots & \\ 1 & t_{h1} & t_{h1} & t_{h3} \\ \vdots & \vdots & \vdots & \\ 1 & t_{V1} & t_{V1} & t_{V3} \\ \end{bmatrix}, \\
\mathbf{c} &= (c_1, \ldots, c_k, \ldots, c_K), \\
\beta &= (\beta_0, \beta_1, \beta_2, \beta_3), \\
\mathbf{y} &= (y_1, \ldots, y_h, \ldots, y_V).
\end{aligned}$$

However, if there is noise in the data, as we assumed with model (5.1), the interpolation conditions $s(\mathbf{t_h}) = y_h$, $h = 1, \ldots, V$ are too strict and we would prefer to place more emphasis on finding a smooth function, hence we prefer minimize the PENSSE$_\lambda(s|\mathbf{y})$ index. To have the coefficients c_k of the (5.4) and of the polynomial $p(\mathbf{t_h})$ that minimize the PENSSE$_\lambda(s|\mathbf{y})$ we must solve the linear system (Carr et al., 1997)

$$\begin{bmatrix} \mathbf{A} - 8V\pi\lambda\mathbf{I} & \mathbf{T} \\ \mathbf{T}' & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{c} \\ \beta \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix}, \tag{5.6}$$

where the parameter λ balances smoothness against fidelity to the data.

5.5 Fast Multipole Method

Solving the systems (5.5) or (5.6) by ordinary or direct methods is computationally expensive and rapidly becomes impossible as V becomes larger than a few thousand. We recall that in our problem each laser scan is composed by 1,500,000 points. Not only are the storage requirements for the systems (5.5) or (5.6) $O(V^2)$ and the work to solve the system $O(V^3)$, but the work associated with evaluating $s(\mathbf{t})$ is also $O(V)$. Greengard and Rokhlin (Greengard, Rokhlin, 1987) proposed the Fast Multipole Method (FMM) to reduce the processing time for the RBF. A full description of the FMM can be found in Beatson et al., 1992. We give a brief outline of the method.

The FMM makes use of the simple fact that when computations are performed, infinite precision is neither required nor expected. Once this is realized, the use of approximations is allowed. With the centers clustered in a hierarchical manner, far- and near-field expansions are used to generate an approximation to that part of the RBF due to the centers in a particular cluster. A judicious use of approximate evaluation for cluster "far" from evaluation point and direct evaluation for clusters "near" to an evaluation point allows the RBF to be computed to any predetermined accuracy and with a significant decrease in computation time compared with direct evaluation. These fast evaluation methods, when used together with fitting methods (Beatson et al., 1999), greatly reduce the storage and computational costs of using RBFs. They reduce the cost of solving the systems (5.5) or (5.6 from $O(V^3)$ to $O(V \log V)$ operations. The fast methods introduce two parameters: a fitting accuracy and evaluation accuracy. The fitting accuracy specifies the maximum allowed deviation of the fitted RBF value from specified value at the interpolation nodes. The evaluation accuracy specifies the precision with which the fitted RBF is then evaluated.

5.5.1 RBF centers reduction

Conventionally, an RBF approximation uses all the input data points as centers of the RBF, so $K = V$, $\mathbf{q}_k = \mathbf{t}_h$, $k, h = 1, \ldots, V$. However, the same input data may be able to be approximated to the desired accuracy using significantly fewer centers. A simple greedy algorithm consists of the following steps:

1. Choose a subset from the V points $\mathbf{t}_h$ and fit an RBF only to these.
2. Evaluate the residuals $\epsilon_h = y_h - s(\mathbf{t}_h)$.
3. If $\max[\epsilon_h] <$ fitting accuracy then stop.
4. Else append new centers where ϵ_h is large.
5. Re-fit RBF and go to 2.

It is important to note that the centers need not to correspond to points $\mathbf{t}_h$.

5.6 Application of the tests

Let us indicate with

$$\mathbf{X}_{1i} = \{\mathbf{t}_{h1i}, y_{h1i}, h = 1, \ldots, V\}$$

and

$$\mathbf{X}_{2i} = \{\mathbf{t}_{h2i}, y_{h2i}, h = 1, \ldots, V\},$$

$i = 1, \ldots, n$ the observations pre-and post-surgery, respectively, where the off-surface points are already included. Clearly V is far greater than the number of units n. Let $s(\mathbf{X}_{ji})$ the smoothing surfaces obtained by RBF methods above.
Considering that:

- given the centers $\mathbf{q}_k$ the choice of the coefficients c_k to approximate a surface is unique (Faul and Powell, 1999);
- the centers need not to correspond to points $\mathbf{t}_h$,

it is possible to use the same centers for all surfaces. Clearly, if the centers are the same the differences between surfaces are all detectable by the coefficients c_k. Hence is possible to apply the test to the new "derived variables" $\mathbf{Y}_{1i} = (c_{k1i}, k = 1, \ldots, K)$ and $\mathbf{Y}_{2i} = (c_{k2i}, k = 1, \ldots, K)$, $i = 1, \ldots, n$. Again K can be much larger than n, but as we seen in previous chapters, we can handle this situation easily with permutation test also if random effects are present using the multi-sided test extended to K-dimensional variables.

5.7 Conclusion

In this final chapter we have seen only a summary presentation of the approximation of surfaces by means of the RBF. In particular, we do not described how to generate points off-surface by projecting along surface normal and we have provided only some hints of a FMM algorithm. These algorithms are very complex and essential for the development of the surfaces but from a statistical point of view these arguments are not particularly interesting because they are of a mathematical and computational nature. Clearly, the proposed methodology is applicable in each field (automotive, aeronautical, geological, etc..) where a digitized surface is available. Unfortunately we have not found commercial software having these algorithms implemented. To write original software would have taken away a lot of resources not only in terms of time. For this reason is not possible to see a practical application of covered topics. We preferred to devote more attention to the study and development of the necessary (and new) statistical methods, such as multisided-tests, finite-sample consistency and weighted multiple testing procedures, as useful tools for analyzing 3-D surfaces. We consider such methods of great practical usefulness and of wide application.

References

Alves, P., Bolognese A. M., Zhao, L. (2007) Three-Dimensional Computerized Orthognathic Surgical Treatment Planning. *Clinics in Plastic Surgery*, 34, 3, 427-436.

Antonini, M., Barlaud, M., Daubechies, I., Mathieu, P. (1992) Image Coding Using Wavelet Transform. *IEEE Transactions on image processing*, 1, 2, 205-220.

Ashraf, A. K., Lifeng, Z. (2000) Rate-Scalable Object-Based Wavelet Codec with Implicit Shape Coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 10, 7, 1068-1079

Azzalini, A., Capitanio, A. (1999) Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society*, Series B, 61, 579-602.

Beatson, R. K., Newsman, G. N. (1992) Fast evaluation of radial basis functions. *Computers & Mathematics with Applications*, 24, 12, 7-19.

Beatson, R. K., Cherrie, J. B., Mouat, C. T. (1999) Fast fitting of radial basis functions: Methods based on preconditioned GMRES iteration. *Advances in Computational Mathematics*, 11, 253-270.

Beatson, R. K., Cherrie, J. B., Ragozin, D. L. (2001) Fast evaluation of radial basis functions: Methods for four-dimensional polyharmonic splines. *SIAM Journal on Mathematical Analysis*, 32, 6, 1272-1310.

Bell, C.B. (1964a) Some basic theorems of distribution-free statistics. *Annals of Mathematical Statistics*, 35, 150-156.

Bell, C.B. (1964b) A characterization of multisample distribution-free statistics. *Annals of Mathematical Statistics*, 35, 735-738.

Bookstein, F. L. (1991) *Morphometric Tools for Landmark Data: Geometry and Biology.* Cambridge: Cambridge University Press.

Brunner, E., Puri, M. L., Sun, S. (1995) Nonparametric methods for stratified two-sample designs with application to multi clinic trials. *Journal of the American Statistical Association*, 90, 1004-1014.

Bosq, D. (2000) *Linear processes in function spaces.* Springer, New York.

Box, G.E.P., Andersen, S.L. (1955) Permutation theory in the derivation of robust criteria and the study of departures from assumption. *Journal of the Royal Statistical Society B*, 17, 1-34.

Boyett, J.M., Shuster, J.J. (1977) Nonparametric one-sided tests in multivariate analysis with medical applications. *Journal of the American Statistical Association*, 72, 665-668.

Cakirer, B., Dean, D., Palomo, J. M., Hans, M. G. (2002) Orthognathic surgery outcome analysis: 3-dimensional landmark geometric morphometrics. *The International Journal of Adult Orthodontics & Orthognathic Surgery*, 17, 2, 116-132.

Carr, J. C., Fright, W. R., Beatson, R. K. (1997) Surface interpolation with radial basis functions for medical imaging. *IEEE Transactions on Medical Imaging*, 16, 1, 96-107.

Cox, D. D., Lee, J. S. (2008) Pointwise testing with functional data using the Westfall-Young randomization method. *Biometrika*, 95, 3, 621-634.

Davidson, R.R., Bradley, R.A. (1970) *Multivariate paired comparisons: some large-sample results on estimation and tests of equality of preference.* In M.L. Puri (ed.) Nonparametric techniques in statistical inference, Cambridge University Press, Cambridge.

Davies, J. L., Kawaguchi, Y, Bennet, S., et al. (1994) A genome-wide search for human type 1 diabetes susceptibility. *Nature*, 371, 130-136.

De Boor, C. (2001) *A practical Guide to Splines*, Revised Edition. Springer, New York.

De B. Pereria, B. (1977) A note on the consistency and on the finite sample comparisons of some tests of separate families of hypotheses. *Biometrika*, 64, 1, 109-113.

Duchon, J. (1977) Splines minimizing rotation-invariant seminorms in Sobolev spaces. In Schempp W. and Zeller K., editors. *Constructive Theory of Functions of Several Variables*, Lectures Notes in Mathematics, 571, 85-100, Springer-Verlag, Berlin.

Dyn, N., Levin, D., Rippa, S. (1986) Numerical procedures for surface fitting of scattered data by radial functions. *SIAM Journal on Scientific Computing*, 7, 2, 639-659.

Faul, A. C., Powell, M. J. D. (1999) Proof of convergence of an iterative technique for thin plate spline interpolation in two dimensions. *Advances in Computational Mathematics*, 11, 183-192.

Feller, W. (1968) *An Introduction to Probability Theory and Its Applications*, Vol. 1. Wiley, New York

Ferraty, F., Vieu, P. (2006) *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, New York.

Finos, L, Salmaso L. (2007) FDR and FWE controlling methods using data-driven weights. *Journal of Statistical Planning an Inference*, 137, 3859-3870.

Greengard, L., Rokhlin, A. (1987) A fast algorithm for particle simulations, *Journal of Computational Physics*, 73, 325-348.

Hall, P., Van Keilegom, I. (2008) Two-sample tests in functional data analysis starting from discrete data. *Statistica Sinica*, 17, 1511-1531.

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

Hollander, M., Wolfe, D.A. (1999) *Nonparametric Statistical Methods*. 2nd Edition, Wiley, New York.

Jihong, C., Xiaoming H. (2005) JBEAM: Multiscale Curve Coding via Beamlets. *IEEE Transactions on image processing*, Vol. 14, No. 11.

Kazuyoshi Y., Makoto A. (2009) PCA Consistency for Non-Gaussian Data in High Dimension, Low Sample Size Context. *Communications in Statistics - Theory and Methods*, Volume 38, Issue 16-17, 2634 - 2652.

Khoury, M. J., Beaty, H., Liang, K. Y. (1988) Can familial aggregation of disease be explained by familial aggregation of environmental risk factors? *American Journal of Epidemiology*, 127, 674-683.

Klassen, E., Srivastava, A. (2004) Analysis of Planar Shapes Using Geodesic Paths on Shape Spaces. *IEEE Transactions on pattern analysis and machine intelligence*, Vol. 26, No. 3.

Kropf, S., Lauter, J., Eszlinger, M., Krohn, K., Paschke, R. (2004) Nonparametric multiple test procedures with data-driven order of hypotheses and with weighted hypotheses. *Journal of Statistical Planning and Inference*,125, 31-47.

Lehmann, E.L (1975) *Nonparametrics: Statistical Methods Based on Ranks.* Holden Day, San Francisco.

Lehmann, E.L. (1986) *Testing Statistical Hypotheses* (2nd edn). Wiley, New York.

Lessi, O. (1993) *Corso di probabilità.* Metria, Padova

Makarov, A., Moniri M. (2006) Binary shape coding using finite automata. *IEE Proceedings - Vision, Image, and Signal Processing*, Vol. 153, No. 5.

Mallat, S. G. (1989) A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine intelligence*, Vol. 11, No. 7.

Marcus, R., Peritz, E., Gabriel, K.R. (1976) On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 66, 655-660.

Moss, J. P., McCance, A. M., Fright, W. R., Linney, A. D., James, D. R. (1994) A three-dimensional soft tissue analysis of fifteen patients with Class II, division 1 malocclusions after bimaxillary surgery. *American Journal of Orthodontics and Dentofacial Orthopaedics*, 105,5, 430-437.

Nelder, J. A. (1954) The interpretation of negative components of variance. *Biometrika*, 41, 544-548.

Peretta, R., Concheri, G., Comelli, D., Meneghello, R., Galzignato, P. F., Ferronato, G. (2008) A 3-Dimensional facial Morpho-Dynamic Database in the development of a prediction model in orthognathic surgery. *Progress in Orthodontics*, 9(2), 8-19.

Pesarin, F. (2001) *Multivariate Permutation Tests.* Willey, Chichester.

Puri, M.L., Sen, P.K. (1971) *Nonparametric Methods in Multivariate Analysis.* Wiley, New York.

Ramsey, J.O., Silverman, B.W. (2005) *Functional Data Analysis, Second Edition.* Springer.

Ruymgaart, F. H. (1980) A unified approach to the asymptotic distribution theory of certain midrank statistics. *Statistique non Parametrique Asymptotique*, 118, J. P. Rault (Ed.), Lectures Notes on Mathematics, No. 821, Springer, Berlin.

Salmaso, L., Solari A. (2005) Multiple aspect testing for case-control designs. *Metrika*, 12, 1-10.

Salmaso, L., Solari, A. (2006) Nonparametric iterated combined tests for genetic differentiation. *Computational Statistics & Data Analysis*, 50, 1105-1112.

Scheffé, H. (1956) *The Analysis Of Variance*. Wiley, New York.

Thompson, W. A., Jr (1962) The problem of negative estimates of variance components. *Annals of Mathematical Statistics*, 33, 273-289.

Westfall, P.H. (1985) Simultaneous Small-Sample Multivariate Bernoulli Confidence Intervals. *Biometrica*, 41, 1001-1013.

Westfall, P.H., Krishen, A. (2001). Optimally weighted, fixed sequence and gatekeeper multiple testing procedures. *Journal of Statistical Planning and Inference*, 99, 25-40.

Bertoluzzo Francesco - Curriculum Vitae

Personal information

Adress: Department of Statistics, University of Padova, Via Cesare Battisti, 241 - 35121 Padova (Italy)

Phone: +39 339 156 0978

E-Mail: fbertoluzzo@stat.unipd.it

Education

University

January 2007:
Enrolled at the Doctoral School in Statistical Sciences at the Department of Statistical Sciences, University of Padova. I'm working on nonparametric methods for statistical analysis of three-dimensional surfaces, under the supervision of Prof. Fortunato Pesarin. Expected completation by March 2010.

March 2004:
Laurea Degree (four years) in Economics at the Faculty of Economics, University Ca' Foscari Venice, Italy. Thesis on "Evolutionary Algorithms of GMDH type: theoretical advances and applications on italian financial market", advisor: Prof. Marco Corazza and Prof. Elio Canestrelli.

Secondary School

July 1990:
Secondary School Diploma in Accountant and Commercial Expert, at I.T.C. G. Girardi, Cittadella (Padova), Italy.

Conferences and Presentations

June 28 - July 4, 2009
Saint Petersburg
6th St. Petersburg Workshop on Simulation, talk titled "Nonparametric Weighted Step Down Holm Method with heteroscedastic variables".

September 3-6 2007
Lecce (Italy)
XXXI Convegno AMASES, talk titled "Making Financial Trading by Recurrent Reinforcement Learning"

Collaborations

MD Peretta R., Dept. of Medical and Surgical Specialities, University of Padova, Italy

Teaching experiences

February 2008
Descriptive Statistics at Department of Physics, University of Padova (Prof. Bruno Scarpa)

Publications

Bertoluzzo, F., Pesarin, F., Salmaso, L., (2009) Multi-sided permutation tests: an approach to random effects. Submitted to Journal of Statistical Planning and Inference.

Bertoluzzo, F., Pesarin, F., Salmaso, L., (2009) Non-parametric Weighted Step Down Holm Method with heteroscedastic variables. Submitted to Quaderni di STATISTICA.

Bertoluzzo, F., Corazza, M., (2007) Macking Financial Trading by Recurrent Reinforcement Learning. Atti XXXI Convegno A.M.A.S.E.S. Lecce 2007.