Working Paper Series, N. 3, February 2013

Rotation-based multiple testing in the multivariate linear model

Livio Finos, Aldo Solari, Jelle J. Goeman
Department of Statistical Sciences
University of Padua
Italy

Abstract: Permutation methods are very useful in several scientific fields. They have the advantage of making fewer assumptions about the data and of providing more reliable inferential results. They are also particularly useful in case of high-dimensional problems since they easily account for dependence between tests, thereby allowing for more powerful multiplicity control procedures. Indeed, Westfall and Young's min-p procedure often improves on the Holm procedure by providing more rejections. The advantage of being able to make fewer assumptions about the process generating the data unfortunately involves an inherent limitation in the way a process can be modeled (e.g. through multiple linear models).
In this work, we propose a permutation (and rotation) method which allows the inference in the multivariate linear model even in the presence of covariates (i.e. nuisance parameters, i.e. confounders). Also, the method allows for the immediate application of the min-p procedure.
We make clear how permutations are a particular case of rotations of the data. Permutation tests are exact, while rotation tests retain exactness under multiple-multivariate linear model with normal errors. When errors are not normal, the rotation tests are weakly exchangeable (i.e. approximated and asymptotically exact). A real application to genetic data is presented and discussed.

Keywords: Permutation Tests, Multivariate tests, confounders

Department of Statistical Sciences
University of Padua
Italy

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

First Draft (2013-02-01)

Contents

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
http://www.stat.unipd.it

Corresponding author:
Livio Finos
tel: +39 049 827 4113
livio@stat.unipd.it
http://www.stat.unipd.it/~livio

Rotation-based multiple testing in the multivariate linear model

Livio Finos, Aldo Solari, Jelle J. Goeman
Department of Statistical Sciences
University of Padua
Italy

Abstract: Permutation methods are very useful in several scientific fields. They have the advantage of making fewer assumptions about the data and of providing more reliable inferential results. They are also particularly useful in case of high-dimensional problems since they easily account for dependence between tests, thereby allowing for more powerful multiplicity control procedures. Indeed, Westfall and Young's min-p procedure often improves on the Holm procedure by providing more rejections. The advantage of being able to make fewer assumptions about the process generating the data unfortunately involves an inherent limitation in the way a process can be modeled (e.g. through multiple linear models).

In this work, we propose a permutation (and rotation) method which allows the inference in the multivariate linear model even in the presence of covariates (i.e. nuisance parameters, i.e. confounders). Also, the method allows for the immediate application of the min-p procedure.

We make clear how permutations are a particular case of rotations of the data. Permutation tests are exact, while rotation tests retain exactness under multiple-multivariate linear model with normal errors. When errors are not normal, the rotation tests are weakly exchangeable (i.e. approximated and asymptotically exact). A real application to genetic data is presented and discussed.

Keywords: Permutation Tests, Multivariate tests, confounders

1 Introduction

Multiple testing is a subject of great interest for applications in genomics, in which a large number of hypotheses are tested simultaneously. A typical example is a microarray study where the aim is to identify which of a large number of genes are differentially expressed between two groups of subjects. We will use this gene expression example for concreteness, although it is just one of the many instances of this problem.

Microarray studies rarely involve the analysis of independent genes, rather, many dependent genes are analyzed simultaneously. However, well-known corrections for multiplicity are unduly conservative when the genes are strongly dependent, resulting in a loss of power to detect truly differentially expressed genes. The reason behind this loss of power is that most multiple testing methods do not incorporate the depen-

dence structure of the p-values, rather are based on assumptions on this dependence structure. Popular methods for controlling the Familywise Error Rate (FWER) are the Holm's method, developed under no assumptions on the dependence structure, and the Hochberg's method, developed under the assumption that the Simes inequality holds for the p-values corresponding to the true null hypotheses. The same issue remains for methods controlling the false discovery rate (FDR). A more explicit use of the dependence structure should result in more powerful methods.

As a consequence, it is desirable to capture the unknown dependence structure from the data. In FWER control, one approach control is to use resampling-based methods, already applied widely in practice (Westfall and Young, 1993; Ge et al., 2003). Notably, the use of permutation-based methods provides an exact finite sample theory, although *exchangeability* is a key requirement for their validity (Calian et al., 2008; Westfall and Troendle, 2008).

In a randomized setting, the mechanism of randomization itself ensures that exchangeability holds. In microarray studies, however, random assignment of subjects to groups is not always feasible, and there is the possibility that differences in expression profiles might be due to confounding factors, such age and sex. Besides breaking exchangeability, confounding is a problem in that it biases the associations that are observed in the data (Ghosh, 2008). For example, if a gene is differentially expressed between the sexes but not between the two groups being compared, and there is an imbalance in sex in the two groups, then without adjusting for the imbalance in sex, it can happen to wrongly attribute the observed difference in expression to groups. One simple strategy for adjusting for measured confounders is to include them in a regression model (Heller et al., 2009).

We consider the multivariate linear model, where multivariate response is the gene expression profile, and the set of covariates is given by the binary group variable, which is of interest for the test, and the age and sex variables, which play the role of confounders. In this multivariate linear model, exchangeability of responses does not hold under a null hypothesis about a null group variable effect. However, it is possible to recover exchangeability of transformed residuals, but only at the cost of a normality assumption (Commenges, 2003). Accordingly, in the multivariate normal linear model it is possible to find exact permutation tests.

This result is linked to the theory of *rotation tests* by Langsrud (2005), since like permutations, rotations are *null-invariant transformations*, that is, transformations of the data that do not change its null distribution. Of the two approaches, the former is well known from the theory of permutation tests, but the latter is lesser known and just recently applied in the theory of rotation tests (Buja et al., 2009; Perry and Owen, 2010; Wu et al., 2010).

However, for low-dimensional normal linear models where the estimation of the covariance matrix is feasible, the practical relevance of having an exact permutation test is limited because a parametric exact solution is applicable. We argue that the importance of rotation-based methods can be enhanced for high-dimensional normal linear models, where in the ever-more-common situation of more responses than observations, we have the advantage that the dependence of the responses can be taken into account.

The proposed approach provides the rotation joint null distribution of the p-

values, which can be used as a basis for carrying out exact multiple testing procedures. One such procedure is the permutation method of Meinshausen (2006), which controls the k-FWER simultaneously for all values of k. By this construction, it becomes possible to extend Meinshausen's method by allowing adjustments for confounders. Furthermore, we demonstrate how the general framework of Goeman and Solari (2011) might be used to uniformly improve Meinshausen's method.

The application of the proposed approach to microarrays data represents a considerable refinement of the permutation approach proposed by Wagner et al. (2008). This approach, based on permuting non exchangeable residuals, provides only an approximate solution. In addition, it does not work well with extremely small sample sizes, because enough subjects are needed in each group to be able to calculate sufficient permutations.

The proposed approach, in contrast, is based on rotating transformed residuals, retains finite sample exactness when the distribution of errors is multivariate normal, and the number of rotations can be chosen large enough to avoid any problem with granularity of p-values.

The outline of this paper is as follows. In the next section, we give the formulation of the multivariate general linear model and we discuss .... Section 3 ... Finally, Section 6 ... Software to perform the procedures described in this paper is available in the `flip` R package.

2 Multivariate linear model

Suppose we have a sample of n independent subjects, for which we have an $(n \times m)$ matrix of m responses $\mathbf{Y}$ and an $n \times (q+c)$ design matrix, with $c < n$. We partition the design matrix into an $n \times q$ design matrix $\mathbf{X}$ of covariates and an $n \times c$ matrix $\mathbf{Z}$ of potential confounders.

Consider the multivariate general linear model defined by

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{Z}\boldsymbol{\Gamma} + \mathbf{E} \tag{1}$$

where $\mathbf{B}$ and $\boldsymbol{\Gamma}$ are $(q \times m)$ and $(c \times m)$ matrices of parameters of interest and nuisance parameters, respectively. We usually suppose that the first column of $\mathbf{Z}$ equals $\mathbf{1}_n$, a n-vector of ones, which means that the first row of $\boldsymbol{\Gamma}$ contains the intercept terms.

The random matrix $\mathbf{E}$ is a matrix of errors. We do not assume a distributional form for $\mathbf{E}$, but we specify its first two moments:

$$\mathbf{E} \sim (\mathbf{0}_{n \times m}, \mathbf{I}_n \otimes \boldsymbol{\Sigma}),$$

where $\mathbf{I}_n$ denotes the $(n \times n)$ identity matrix and for a random matrix $\mathbf{M} \sim (\mathbf{A}, \mathbf{B} \otimes \mathbf{C})$, the notation means that its elements have $\mathrm{E}(m_{ij}) = a_{ij}$ and $\mathrm{Cov}(m_{ij}, m_{kl}) = b_{ik}c_{jl}$ for matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ of appropriate dimensions. Since we have assumed independent subjects, the rows of $\mathbf{E}$ are independent draws from some m-variate distribution with mean $\mathbf{0}$ and covariance $\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ in the microarray setting represents the $(m \times m)$ gene-gene covariance matrix.

The univariate linear model for the jth response is

$$\boldsymbol{y}_j = \mathbf{X}\boldsymbol{\beta}_j + \mathbf{Z}\boldsymbol{\gamma}_j + \boldsymbol{\varepsilon}_j$$

where $\boldsymbol{y}_j$, $\boldsymbol{\beta}_j$, $\boldsymbol{\gamma}_j$ and $\boldsymbol{\varepsilon}_j$ are the jth column of $\mathbf{Y}$, $\mathbf{B}$, $\boldsymbol{\Gamma}$ and $\mathbf{E}$, respectively. Here $\boldsymbol{\varepsilon}_j$ has mean $\mathbf{0}$ and covariance $\sigma_j^2\mathbf{I}_n$, where σ_j^2 is the jth diagonal element of $\boldsymbol{\Sigma}$. We are interested in testing the collection of null hypotheses

$$H_j : \boldsymbol{\beta}_j = \mathbf{0} \qquad j = 1, \ldots, m, \tag{2}$$

and the complete null hypothesis is denoted by

$$H = \bigcap_{j=1}^{m} H_j. \tag{3}$$

In microarray studies, this general framework looks as follows: m genes are measured in n independent subjects (represented by $\mathbf{Y}$), and the aim is to identify the genes that are associated with some variable of interest, such as a binary or continuous phenotype (represented by $\mathbf{X}$) while taking into account additional subject characteristics, such as sex and age (represented by $\mathbf{Z}$).

We can get rid of the part of the design matrix which is not of interest for the test by projecting the response into $\mathcal{Z}^{\perp}$, the subspace orthogonal to the subspace $\mathcal{Z}$ spanned by the columns of $\mathbf{Z}$. This can be done by premultiplying both sides of (1) by the projection matrix $\mathbf{I}_n - \mathbf{H}$, where $\mathbf{H} = \mathbf{Z}(\mathbf{Z}^{\mathsf{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathsf{T}}$, obtaining

$$\hat{\mathbf{E}} = (\mathbf{I}_n - \mathbf{H})\mathbf{X}\mathbf{B} + (\mathbf{I}_n - \mathbf{H})\mathbf{E}. \tag{4}$$

Here $\hat{\mathbf{E}} = (\mathbf{I}_n - \mathbf{H})\mathbf{Y}$ represents the estimated residuals from the reduced model $\mathbf{Y} = \mathbf{Z}\boldsymbol{\Gamma} + \mathbf{E}$, that is, model (1) under H. Wagner et al. (2008) and Zeng et al. (2011) used these type of residuals to obtain permutation tests that have exact significance levels only asymptotically. In fact, in order to obtain finite sample exactness, $\hat{\mathbf{E}}$ should have an exchangeable null distribution, i.e. a null distribution that does not change if we permute the rows of $\hat{\mathbf{E}}$.

Permutation tests' construction can be seen as a special case of the more general randomization tests' construction discussed in Lehmann and Romano (2005). This construction requires a finite algebraic group of transformations $\mathcal{O}$ such that

$$\hat{\mathbf{E}} \stackrel{\mathrm{d}}{=} \mathbb{O}\hat{\mathbf{E}} \quad \text{under } H, \tag{5}$$

for every $n \times n$ matrix $\mathbb{O} \in \mathcal{O}$, where '$\stackrel{\mathrm{d}}{=}$' denotes equality in distribution. We refer to this condition as the *randomization hypothesis* and to such transformations as *null-invariants*. The name null-invariants comes from the fact that the transformation of $\hat{\mathbf{E}}$ by $\mathbb{O}$ does not change its null distribution. In this construction, $\mathbb{O}\hat{\mathbf{E}}$ is, intuitively speaking, 'look-alike' of $\hat{\mathbf{E}}$ if H is true, and repeat observations are made again and again, the p-values calculated for each, and their joint null distribution is built up.

For permutation tests, $\mathbb{O}$ is a permutation matrix, that is, a square matrix that have exactly one entry 1 in each row and each column and 0s elsewhere, and all the $n!$ possible permutation matrices form the group $\mathcal{O}$. The randomization hypothesis (5) corresponds to require row-exchangeability of $\hat{\mathbf{E}}$, i.e. that $\mathbb{O}\hat{\mathbf{E}}$ has the same null distribution as $\hat{\mathbf{E}}$ for every permutation matrix $\mathbb{O}$.

To see the lack of exchangeability of $\hat{\mathbf{E}}$, note that a necessary condition for (5) to hold is

$$(\mathbf{I}_n - \mathbf{H}) = \mathbb{O}(\mathbf{I}_n - \mathbf{H})\mathbb{O}^\mathsf{T} \qquad \text{under } H,$$

for every permutation matrix $\mathbb{O}$, which, in general, is not true.

This can be seen by looking at the rank of $\mathbf{I}_n - \mathbf{H}$, which is $n-c$. We known from Commenges (2003) that $\mathbf{I}_n - \mathbf{H}$ in order to satisfy $(\mathbf{I}_n - \mathbf{H}) = \mathbb{O}(\mathbf{I}_n - \mathbf{H})\mathbb{O}^\mathsf{T}$ must have rank n, $n-1$, 1 or 0. The only interesting case is of rank equal to $n-1$, which occurs when $\mathbf{Z} = \mathbf{1}_n$, that is, when under H the model contains only an intercept. In this case, $\hat{\mathbf{E}} = (\mathbf{I}_n - \frac{1}{n}\mathbf{1}_{n\times n})\mathbf{Y}$ and (5) does hold whatever the distribution of $\mathbf{Y}$.

The randomization hypothesis may be recovered, but only at the cost of an additional assumption. Write $(\mathbf{I}_n - \mathbf{H}) = \mathbf{Q}\mathbf{Q}^\mathsf{T}$ according to its eigenvalue decomposition, where $\mathbf{Q}$ is a semi-orthogonal $(n \times n-c)$ matrix whose columns are the eigenvectors corresponding to non-zero eigenvalues and $\mathbf{Q}^\mathsf{T}\mathbf{Q} = \mathbf{I}_{n-c}$. By premultiplying both sides of (4) by $\mathbf{Q}^\mathsf{T}$, we obtain

$$\widetilde{\mathbf{Y}} = \widetilde{\mathbf{X}}\mathbf{B} + \widetilde{\mathbf{E}}, \tag{6}$$

where $\widetilde{\mathbf{Y}} = \mathbf{Q}^\mathsf{T}\mathbf{Y}$, $\widetilde{\mathbf{X}} = \mathbf{Q}^\mathsf{T}\mathbf{X}$ and $\widetilde{\mathbf{E}} = \mathbf{Q}^\mathsf{T}\mathbf{E}$. Here $\widetilde{\mathbf{Y}} \sim (\mathbf{0}_{(n-c)\times m}, \mathbf{I}_{n-c} \otimes \mathbf{\Sigma})$ under H, and in order to achieve the randomization hypothesis for (6), i.e.

$$\widetilde{\mathbf{Y}} \stackrel{\mathrm{d}}{=} \widetilde{\mathbb{O}}\widetilde{\mathbf{Y}} \quad \text{under } H, \tag{7}$$

for every $(n-c) \times (n-c)$ matrix $\widetilde{\mathbb{O}} \in \widetilde{\mathcal{O}}$, it is necessary for the null-invariants $\widetilde{\mathbb{O}}$ to satisfy $\mathbf{I}_{n-c} = \widetilde{\mathbb{O}}\mathbf{I}_{n-c}\widetilde{\mathbb{O}}^\mathsf{T}$. It follows that $\widetilde{\mathcal{O}}$ is the orthogonal group of degree $n-c$ formed by the orthogonal matrices $\widetilde{\mathbb{O}}$ with $\widetilde{\mathbb{O}}\widetilde{\mathbb{O}}^\mathsf{T} = \mathbf{I}_{n-c}$, which we refer to as *rotations*.

In addition, it is known that the class of distributions satisfying (7) when $\widetilde{\mathbb{O}}$ is an orthogonal matrix is the class of *left-spherical distributions* (Dawid, 1981). This means that to fulfill the randomization hypothesis (7), we need to assume that $\widetilde{\mathbf{E}}$ is left-spherically distributed.

Because this implies also that $\mathbf{E}$ is left-spherically distributed, the gain in generality that flows from allowing any left-spherical distribution for $\mathbf{E}$ is somewhat illusory. Indeed, the requirement of independence of observations, coupled with left-sphericity, implies normality, i.e.

$$\mathbf{E} \sim \mathcal{N}(\mathbf{0}_{n\times m}, \mathbf{I}_n \otimes \mathbf{\Sigma}). \tag{8}$$

As a result, the multivariate normal linear model provides the basis for rotation-based multiple testing procedures with finite sample validity that explicitly account for the correlation structure of the data.

The above results are in agreement with the results of Perry and Owen (2010) and Commenges (2003). Perry and Owen (2010) showed that the rotations

$$\mathbb{O} = \mathbf{P}\mathbf{P}^\mathsf{T} + \mathbf{Q}\widetilde{\mathbb{O}}\mathbf{Q}^\mathsf{T} \tag{9}$$

satisfy (5) if we assume (8), where the $(n \times c)$ matrix $\mathbf{P}$ comes from the eigenvalue decomposition $\mathbf{H} = \mathbf{P}\mathbf{P}^\mathsf{T}$. They proved that in order to satisfy $\mathbb{O}(\mathbf{I}_n - \mathbf{H})\mathbb{O} = \mathbf{I}_n - \mathbf{H}$,

$\mathbb{O}$ must take the form given in (9). This construction transforms $\hat{\mathbf{E}}$ even if its covariance matrix is not of full rank, as opposed to that of $\widetilde{\mathbf{Y}}$, but it is essentially equivalent to transform $\widetilde{\mathbf{Y}}$, since for transformations $\mathbb{O}$ of the form (9), we have $\mathbb{O}\hat{\mathbf{E}} = \mathbb{O}(\mathbf{I}_n - \mathbf{H})\mathbf{Y} = \mathbf{Q}\widetilde{\mathbb{O}}\widetilde{\mathbf{Y}}$.

Commenges (2003) restricted attention to the finite subgroup of $\widetilde{\mathcal{O}}$ formed by all the permutation matrices, and showed that by assuming (8), $\widetilde{\mathbf{Y}}$ has an exchangeable null distribution and thus it is possible to find exact permutation tests.

Without assuming (8), what we obtain is a weaker form of invariance, that we refer to as *second-moment null-invariance*, which replaces (7) by

$$\mathrm{E}(\widetilde{\mathbf{Y}}) = \mathrm{E}(\widetilde{\mathbb{O}}\widetilde{\mathbf{Y}}) \text{ and } \mathrm{Cov}(\widetilde{\mathbf{Y}}) = \mathrm{Cov}(\widetilde{\mathbb{O}}\widetilde{\mathbf{Y}}) \text{ under } H. \tag{10}$$

This corresponds to second-moment exchangeability (Commenges, 2003) when $\widetilde{\mathbb{O}}$ is a permutation matrix.

3 Multiple testing

A key feature of the rotation-based construction is that it provides the joint null distribution of the p-values which can be used as a basis for carrying out multiple testing procedures. We focus on controlling the k-FWER, that is, the probability of making at least $k \geq 1$ false rejections, but simultaneously for all values of k. This consents to the user to choose the number of rejections adaptively while still retaining control over the number of false rejections.

We develop our procedure in the framework of Goeman and Solari (2011), which requires the closed testing procedure (Marcus et al., 1976). To use the closed testing procedure, we should consider not only the hypotheses $H_1, \ldots, H_m$, but also the intersection hypotheses $H_J = \bigcap_{j \in J} H_j$, where $J \subseteq \{1, \ldots, m\}$ is a non-empty collection of indices.

We demand that two conditions should be fulfilled. On one hand, we focus on Simes type local tests, by requiring α-level tests for each intersection hypothesis to guarantee a valid closed testing procedure. On the other hand, we require conditions on the critical values, which admit shortcuts for avoiding the calculation of all the $2^m - 1$ tests required by the closed testing procedure.

For notational convenience, we will denote $\widetilde{\mathbf{Y}}$, $\widetilde{\mathbf{X}}$, $\widetilde{\mathbf{E}}$ and $\widetilde{\mathbb{O}}$ simply as $\mathbf{Y}$, $\mathbf{X}$, $\mathbf{E}$ and $\mathbb{O}$, respectively, keeping in mind that now the number of observations is $n - c$. Suppose we have a p-value p_j for each null hypothesis H_j, which depends on $\mathbf{Y}$ only through its jth component $\boldsymbol{y}_j$. For example, when $q < n - c$, we may obtain p_j from the F statistic

$$\frac{\boldsymbol{y}_j^\top \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top \boldsymbol{y}_j / q}{\boldsymbol{y}_j^\top (\mathbf{I}_{n-c} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top)\boldsymbol{y}_j / (n - c - q)}$$

which follows the F distribution with $q, (n-c)-q$ degrees of freedom when the errors are normally distributed and H_j is true.

3.1 Local tests

We focus on Simes type local tests, that is, we reject H_J whenever

$$p^J_{(k)} < c^J_k$$

for at least one $1 \le k \le \#J$, where $p^J_{(k)}$ is the kth smallest p-value among $\{p_j, j \in J\}$ and $\{c^J_1, \ldots, c^J_{\#J}\}$ are critical values such that $c^J_k \le c^J_{k'}$ if $k \le k'$. Note that when $c^J_k = c^J$ for $1 \le k \le \#J$, it corresponds to reject H_J when

$$p^J_{(1)} < c^J$$

i.e., when the smallest p-value among $\{p_j, j \in J\}$ is less than some critical value, as in Westfall and Young's procedure.

We require α-level tests for every intersection hypothesis, that is,

$$\mathrm{P}\Big(\bigcup_{k=1}^{\#J} \{p^J_{(k)} < c^J_k\}\Big) \le \alpha \qquad \text{under } H_J, \tag{11}$$

for every $J \subseteq \{1, \ldots, m\}$. When the Simes inequality holds for $\{p_j, j \in J\}$, $c^J_k = k\alpha/\#J$ is a valid choice.

However, instead of making assumptions on the correlation structure of the p-values, we can find some α-quantile $(\hat{c}^J_1, \ldots, \hat{c}^J_{\#J})$ of the rotation distribution of the ordered p-values $(p^J_{(1)}, \ldots, p^J_{(\#J)})$. This generalizes the Westfall and Young's permutation-based procedure, which locates the α-quantile $\hat{c}^J$ of the permutation distribution of the minimum p-value $p^J_{(1)}$.

Note that, as opposed to the univariate case, a whole range of α-quantiles is possible. Any choice of $(\hat{c}^J_1, \ldots, \hat{c}^J_{\#J}) \in [0,1]^{\#J}$ ensuring that the corresponding critical region has size α is valid.

Let $\mathcal{O}_r = \{\mathbb{O}_1, \ldots, \mathbb{O}_r\}$ be a finite subset of $\mathcal{O}$ consisting of r rotations.

Let $p_j \circ \mathbb{O}_b$ be the p-value obtained from the transformed data $\mathbb{O}_b \mathbf{y}_j$, and $p^J_{(k)} \circ \mathbb{O}_b$ be the kth smallest p-value among $\{p_j \circ \mathbb{O}_b, j \in J\}$. Define $\big(p^J_{(k)} \circ \mathcal{O}_r\big)_{(s)}$ as the sth smallest value among $p^J_{(k)} \circ \mathcal{O}_r = \{p^J_{(k)} \circ \mathbb{O}_b, b = 1, \ldots, r\}$.

One possible α-quantile of the rotation distribution of $(p^J_{(1)}, \ldots, p^J_{(\#J)})$ can be defined as

$$\hat{c}^J_k = \big(p^J_{(k)} \circ \mathcal{O}_r\big)_{(s^J_\alpha)}, \quad 1 \le k \le \#J \tag{12}$$

where

$$s^J_\alpha = \max\Big\{s \in \{1, \ldots, \lfloor \alpha r \rfloor\} : \#\Big\{b : \bigcup_{k=1}^{\#J} \Big\{p^J_{(k)} \circ \mathbb{O}_b < \big(p^J_{(k)} \circ \mathcal{O}_r\big)_{(s)}\Big\}\Big\} \le \lfloor \alpha r \rfloor\Big\}$$

and $\lfloor \alpha r \rfloor$ is the largest integer less than or equal to αr.

Observe that another possible α-quantile of the rotation distribution of $(p^J_{(1)}, \ldots, p^J_{(\#J)})$ is $(\hat{c}^J, \ldots, \hat{c}^J)$, where

$$\hat{c}^J = \big(p^J_{(1)} \circ \mathcal{O}_r\big)_{(\lfloor \alpha r \rfloor)} \tag{13}$$

is the α-quantile of the rotation distribution of $p^J_{(1)}$. This corresponds to the Westfall and Young's procedure.

In order to check condition (11) with c^J_k equal to $\hat{c}^J_k$ in (12) or to $\hat{c}^J$ in (13), we need to generalize the randomization hypothesis to all subsets of hypotheses because $\mathbf{Y}$ and $\mathbb{O}_b\mathbf{Y}$ need not have the same distribution if only a subset $\{H_j, j \in J\}$ of the hypotheses is true. Let $\mathbf{Y}_J$ be the $(n-c) \times (\#J)$ matrix consisting of the columns $\boldsymbol{y}_j, j \in J$.

The important point here is that for the subset of the data $\mathbf{Y}_J$ that is used for the calculation of $\{p_j, j \in J\}$, the randomization hypothesis holds:

$$\mathbf{Y}_J \stackrel{\mathrm{d}}{=} \mathbb{O}_b\mathbf{Y}_J \quad \text{under } H_J,$$

for every $J \subseteq \{1, \ldots, m\}$ and every $b \in \{1, \ldots, r\}$. This implies the *universal null-invariance condition*, which says that the joint distribution of the p-values of the true null hypotheses and their transformations by $\mathcal{O}_r$ is not altered by another transformation in $\mathcal{O}_r$, i.e. under H_J

$$\{p_j \circ \mathcal{O}_r, j \in J\} \stackrel{\mathrm{d}}{=} \{p_j \circ \mathcal{O}_r \circ \mathbb{O}_b, j \in J\}$$

for every $b \in \{1, \ldots, r\}$ and every $b \in \{1, \ldots, r\}$. It follows from Theorem 2 in Goeman and Solari (2010) that the local tests with c^J_k equal to $\hat{c}^J_k$ in (12) or to $\hat{c}^J$ in (13) are α-level tests.

3.2 Shortcuts

A valid closed testing procedure can be made on the basis of these local tests, but it becomes computationally intractable because it requires $2^m - 1$ rotation tests to be performed.

We focus on rejecting hypotheses that have p-values smaller than a threshold value λ, that is, hypotheses with indices in $R_\lambda = \{j : p_j < \lambda\}$. For a chosen λ, the multiple testing procedure returns a lower bound $f_\alpha(\lambda)$ for the number $\phi(\lambda)$ of false hypotheses among R_λ, which satisfies

$$\mathrm{P}\Big\{\phi(\lambda) \geq f_\alpha(\lambda)\Big\} \geq 1 - \alpha \quad \text{for all } \lambda \in [0, 1].$$

The important thing to note is that the confidence statement is simultaneous over all choices of λ, and consequently such threshold value can be chosen post hoc rather than a priori. By choosing λ, the user determines the number of rejections, and $\#R_\lambda$ with the largest λ such that $f_\alpha(\lambda) > \#R_\lambda - k$ holds gives the maximum number of rejections preserving k-FWER control. In this way the user can compute the maximum number of rejections allowed with $k = 1, \ldots, m$ and choose the value of k he or she likes best while maintaining k-FWER control.

We require the critical values to depend only on the cardinality of J and not on J itself, i.e.

$$c^J_k = c^{\#J}_k \tag{14}$$

and to satisfy

$$c_k^s \leq c_k^t \quad \text{for } s \geq t. \tag{15}$$

When the critical values satisfy both (14) and (15), we have the shortcut

$$f_\alpha(\lambda) > \max\{S_r : 1 \leq r \leq \#R_\lambda\},$$

where $S_r = \max\{s \geq 0 : p_{(r)} \leq c_{r-s}^m\}$ if such s exists, -1 otherwise.

Note that the choice

$$c_k^J = c_k$$

for any given $c_1 \leq \ldots \leq c_m$, fulfils both (14) and (15) for every k and J.

Meinshausen (2006) considered

$$c_k^J = \hat{c}_k^0$$

where $\hat{c}_k^0 = \hat{c}_k^{\{1,\ldots,m\}}$ in (12), which, by construction, ensures $\hat{c}_k^0 \leq \hat{c}_k^J$ for every k and J and guarantees a valid, though conservative, local test. The corresponding lower bound for the number of false hypotheses is

$$f_\alpha(\lambda) > \max\{S_r^0 : 1 \leq r \leq \#(R_\lambda)\} \quad \text{for all } \lambda \in [0,1]. \tag{16}$$

where $S_r^0 = \max\{s \geq 0 : p_{(r)} \leq c_{r-s}^0\}$.

Westfall and Young (1993) considered in their single-step procedure

$$c_k^J = \hat{c}^0$$

where $\hat{c}^0 = \hat{c}^{\{1,\ldots,m\}}$ in (13), which, by construction, ensures $\hat{c}^0 \leq \hat{c}^J$ for every J. In this case we have

$$f_\alpha(\lambda) \geq \#R_{\hat{c}^0} \quad \text{for all } \lambda \in (\hat{c}^0, 1]$$

that is, we can be confident that by rejecting the first $\#R_\lambda$ hypotheses that have the smallest p-values, we have at least $\#R_{\min\{\lambda,\hat{c}^0\}}$ correct rejections. This is because the single-step procedure rejects the hypotheses H_j, $j \in R_{\hat{c}^0}$ with familywise error control, thus all these rejections are correct rejections with probability at least $1-\alpha$.

However, it is well-known that the Westfall and Young's step-down procedure rejects at least as much as the single-step procedure, thus a sharper bound for the number of correct rejections may be obtained. This procedure discards the rejected hypotheses in the previous step and recalculates the α-quantile until any step fails to result in additional rejections. Let $\hat{c}^{i+1} = \hat{c}^{M_{i+1}}$, $i = 0, 1, \ldots$ where $M_{i+1} = \{1, \ldots, m\} \setminus R_{\hat{c}^i}$ is the collection of indices of hypotheses not rejected after step i. Note that by construction $\hat{c}^{i+1} \geq \hat{c}^i$. At the final step we have

$$f_\alpha(\lambda) \geq \#R_{\hat{c}^\infty} \quad \text{for all } \lambda \in (\hat{c}^\infty, 1]$$

where $\hat{c}^\infty = \lim_{i\to\infty} \hat{c}^i$ and $\#R_{\hat{c}^\infty}$ is the number of hypotheses rejected by the step-down procedure with familywise error control.

It is interesting to consider the Westfall and Young's step-down procedure from a closed testing perspective. In the first step, for every $j \in R_{\hat{c}^0}$, $p_j < \hat{c}^0$ implies

$p^J_{(1)} < \hat{c}^J$ for every $J \ni j$, thus the rejection of all the intersection hypotheses H_J with $J \ni j$. This partitions the set of all non-empty subsets of $\{1, \ldots, m\}$ into $\{J \neq \emptyset : J \subseteq M_1\}$ and $\bigcup_{j \in R_{\hat{c}^0}} \{J \subseteq \{1, \ldots, m\} : J \ni j\}$, where all the intersection hypotheses with indices in the latter set are rejected. Thus we can reject the first $\#R_{\hat{c}^0}$ hypotheses with smallest p-values and obtain $f_\alpha(\hat{c}^0) = \#R_{\hat{c}^0}$. In the next step, the closed testing procedure may be restarted with the rest of the hypotheses, i.e. H_J with $J \subseteq M_1$, calculate $\hat{c}^1$ and obtain $f_\alpha(\hat{c}^1) = \#R_{\hat{c}^1}$, where possibly $\#R_{\hat{c}^1} > \#R_{\hat{c}^0}$. This process is repeated until any step fails to result in additional rejections.

The same reasoning may be used to construct a sequential procedure which rejects at least as much as Meinshausen's procedure. Observe that for every $j \in R_{\hat{c}^0_1}$, $p_j < \hat{c}^0_1$ implies $p^J_{(k)} < \hat{c}^J_k$ for at least one k and every $J \ni j$, thus the rejection of all intersection hypotheses H_J, $J \ni j$. Thus we can reject the first $\#R_{\hat{c}^0_1}$ hypotheses with smallest p-values and obtain $f_\alpha(\hat{c}^0_1) = \#R_{\hat{c}^0_1}$. In the next step, we discard these hypotheses, we restart the closed testing procedure with H_J with $J \subseteq M_1$, and we recalculate the α-quantile $(\hat{c}^{M_1}_1, \ldots, \hat{c}^{M_1}_{\#M_1})$. This process is repeated until for some $i \geq 0$ occurs that $p_j \geq \hat{c}^{M_{i+1}}_1$ for every $j \in M_{i+1}$. At the final step we obtain

$$f_\alpha(\lambda) > \#R_{\hat{c}^\infty_1} + \max\{S^\infty_r : 1 \leq r \leq \#(R_\lambda \setminus R_{\hat{c}^\infty_1})\} \quad \text{for all } \lambda \in (\hat{c}^\infty_1, 1]. \tag{17}$$

where $S^{i+1}_r = \max\{s \geq 0 : p_{(r)} \leq \hat{c}^{M_{i+1}}_{r-s}\}$ for $i = 0, 1, \ldots$ and $S^\infty_r = \lim_{i \to \infty} S^i_r$.

The lower bound (17) lies between the Meinshausen's lower bound (16) and the one obtained by the full closed testing procedure. Further refinements towards the full closed testing lower bound can be obtained along the lines of remampling-based k-FWER controlling procedures (Romano and Wolf, 2007), but the algorithm becomes computationally more complex.

3.3 Rotation null distribution of P-values

In this section we consider simulated dataset to understand how the dependences are dealt by permutation tests. The problem is bivariate with both test under the null hypothesis : $H : H_1 \cap H_2$

We compare the rejection regions of Bonferroni method (which does not take in account dependence among p-values) with the min-p method. The former has a region defined by $c = \alpha/2$, while the latter is based on the joint distribution of sampled (i.e. rotated) p-values and the frontier c is such that $\Pr_H(\min(p_1, p_2) < c) \leq \alpha$. Results given in Figure 1 show that the area defined by min-p method is broader than the Bonferroni one.

4 Application

The dataset by Chiaretti et al. (2004), available in the Bioconductor ALL package at www.bioconductor.org. The dataset was collected to identify genes that distinguish subgroups of leukemia patients. One hundred and twenty eight patients are split into 95 with B-cell and 33 with T-cell type acute lymphoblastic leukemia (ALL). The data consist of 12 625 expression profiles from the HGU95aV2 Affymetrix chip

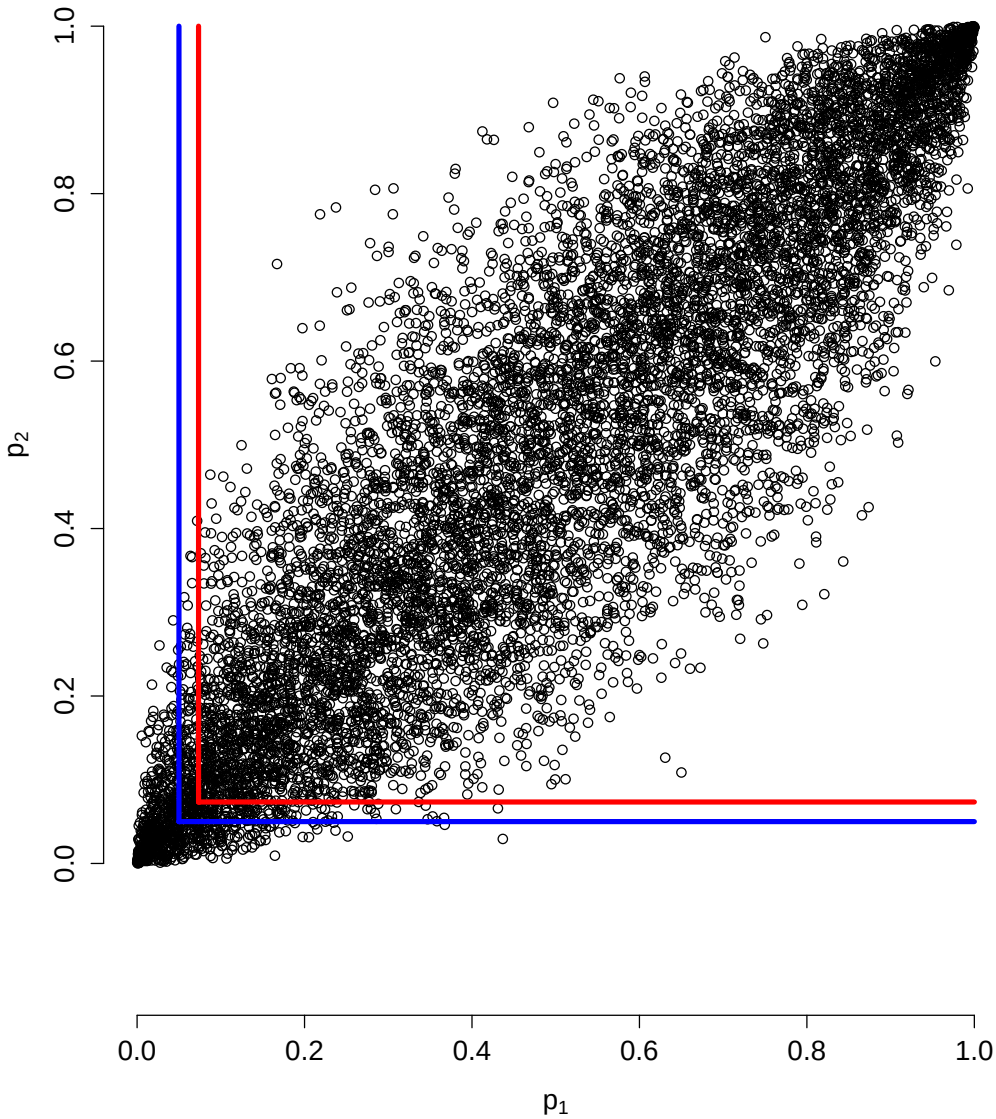

Figure 1: Joint rotation null distribution of p-values and rejection regions of rotation test with min-p (in red) and parametric with Bonferroni (in blue)

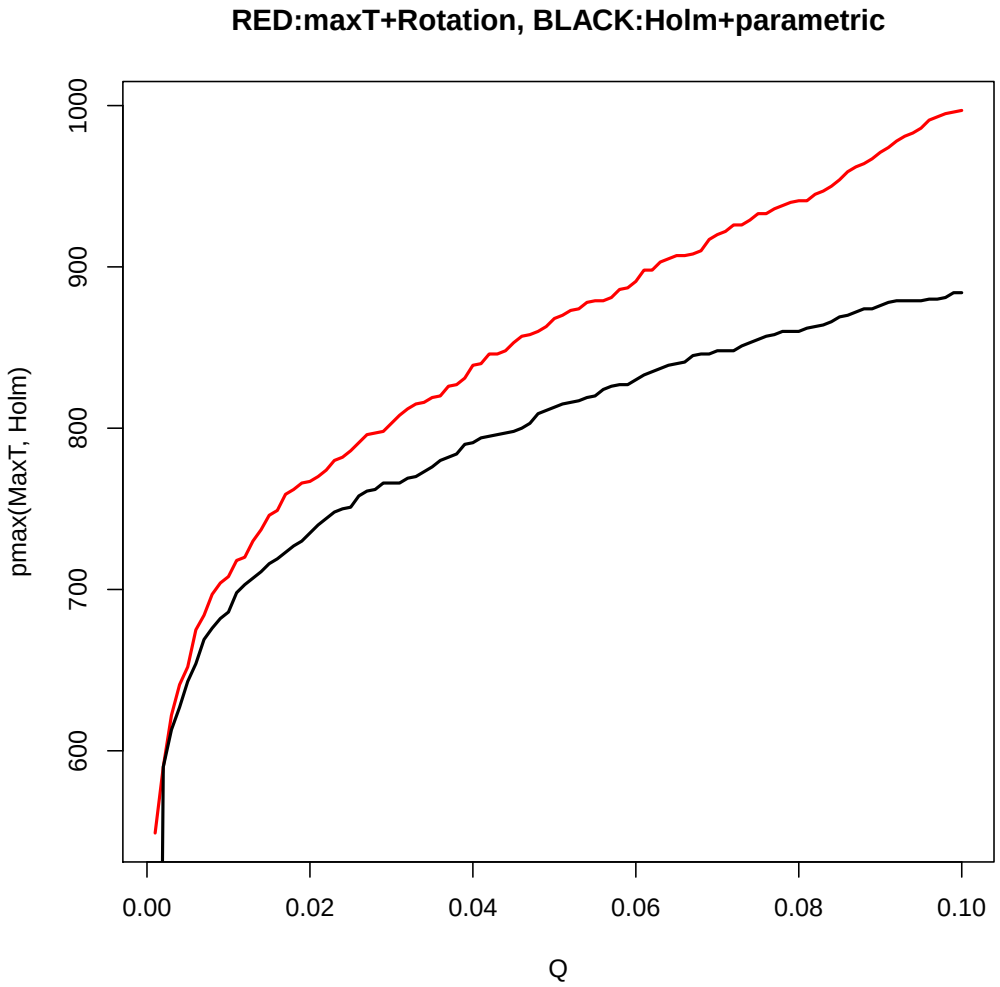

Figure 2: Number of rejections for max-t rotation method and Hom method with t-test as a function of significance level.

for each patient. Information was available for the following additional covariates: age, sex, multi drug resistance (mdr), the stage of cell differentiation (stage) and an indicator of whether the chromosome number was larger than 46 (kinet).

The results of Figure 2 compare the rejections of two methods of multiplicity control: the rotation + max-T method discussed here is compared with a parametric linear model followed by a Holm correction. Let note that max-T method is actually the same as min-p, with the only exception that test statistic is based on max-T instead of min-p. This becomes convenient for large dataset, where computation of p-values is time-consuming.

References

A. Buja, D. Cook, H. Hofmann, M. Lawrence, E.-K. Lee, D.F. Swayne, and Wickham H. Statistical inference for exploratory data analysis and model diagnostics. *Phil. Trans. R. Soc. A*, 367:4361–4383, 2009.

V. Calian, D. M. Li, and J. C. Hsu. Partitioning to uncover conditions for permutation tests to control multiple testing error rates. *Biometrical Journal*, 50(5): 756–766, October 2008.

D. Commenges. Transformations which preserve exchangeability and application to permutation tests. *Journal of Nonparametric Statistics*, 15:171–185, 2003.

A.P. Dawid. Some matrix-variate distribution theory: Notational considerations and a bayesian application. *Biometrika*, 68:265–274, 1981.

Y. Ge, S. Dudoit, and T. P. Speed. Resampling-based multiple testing for microarray data analysis. *Test*, 12(1):1–77, 2003.

D. Ghosh. Multiple testing procedures under confounding. *N. Balakrishnan, Edsel A. PeÛa and Mervyn J. Silvapulle, eds., Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, 1:243–256, 2008.

J.J. Goeman and A. Solari. The sequential rejection principle of familywise error control. *The Annals of Statistics*, 38:3782–3810, 2010.

J.J. Goeman and A. Solari. Multiple testing for exploratory research. *Statistical Science*, 26:584–597, 2011.

R. Heller, E. Manduchi, and Small D.S. Matching methods for observational microarray studies. *Bioinformatics*, 25:904–909, 2009.

O. Langsrud. Rotation tests. *Statistics and Computing*, 15:53–60, 2005. ISSN 0960-3174.

E. L. Lehmann and J. P. Romano. *Testing statistical hypotheses.* Springer, New York, 2005.

R Marcus, E. Peritz, and K.R. Gabriel. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63:655–660, 1976.

N. Meinshausen. False discovery control for multiple tests of association under general dependence. *Scandinavian Journal of Statistics*, 33:227–237, 2006.

P.O. Perry and A.B. Owen. A rotation test to verify latent structure. *Journal of Machine Learning Research*, 11:603–624, 2010.

J.P. Romano and M. Wolf. Control of generalized error rates in multiple testing. *Annals of Statistics*, 35:1378–1408, 2007.

Brandie D. Wagner, Gary O. Zerbe, Sharon Mexal, and Sherry S. Leonard. Permutation-based adjustments for the significance of partial regression coefficients in microarray data analysis. *Genetic Epidemiology*, 32(1):1–8, 2008. ISSN 1098-2272. doi: 10.1002/gepi.20255. URL `http://dx.doi.org/10.1002/gepi.20255`.

P. H. Westfall and J. F. Troendle. Multiple testing with minimal assumptions. *Biometrical Journal*, 50(5):745–755, 2008.

P. H. Westfall and S. S. Young. *Resampling-based multiple testing: examples and methods for p-value adjustment*. Wiley, New York, 1993.

Di Wu, Elgene Lim, FranÃğois Vaillant, Marie-Liesse Asselin-Labat, Jane E. Visvader, and Gordon K. Smyth. Roast: rotation gene set tests for complex microarray experiments. *Bioinformatics*, 26(17):2176–2182, 2010.

C. Zeng, Z. Pan, S. MaWhinney, A.E. BarŮn, and G.O. Zerbe. Permutation and f distribution of tests in the multivariate general linear model. *The American Statistician*, 65:31–36, 2011.

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it

Most of the working papers can also be found at the following url: http://wp.stat.unipd.it

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

Department of Statistical Sciences
University of Padua
Italy