
Volume 137
Issue 8
1 August 2007
ISSN 0378-3758
ELSEVIER
journal of statistical planning and inference
Special Issue: 5th St. Petersburg Workshop on Simulation
St. Petersburg State University, St. Petersburg, Russia
26 June–2 July 2005
Edited by: Viatcheslav Melas
For Contents see page iii
Available online at
ScienceDirect
www.sciencedirect.com

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Journal of Statistical Planning and Inference 137 (2007) 2633–2641
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Conditional inference under simultaneous stochastic ordering constraints

Livio Finos[a], Luigi Salmaso[b], Aldo Solari[c,*]

[a]Department of Biomedical Sciences, University of Ferrara, via Fossato di Mortara 17/19, 44100 Ferrara, Italy
[b]Department of Management and Engineering, University of Padova, Str. S. Nicola 3, 36100 Vicenza, Italy
[c]Department of Statistical Sciences, University of Padova, via Cesare Battisti 241, 35121 Padova, Italy

Received 11 November 2005; accepted 3 April 2006
Available online 14 January 2007

Abstract

Testing for stochastic ordering is of considerable importance when increasing does of a treatment are being compared, but in applications involving multivariate responses has received much less attention. We propose a permutation test for testing against multivariate stochastic ordering. This test is distribution-free and no assumption is made about the dependence relations among variables. A comparative simulation study shows that the proposed solution exhibits a good overall performance when compared with existing tests that can be used for the same problem.

Keywords: Combining dependent permutation tests; Multiple testing; Order-restricted inference; Permutation test; Stochastic ordering

1. Introduction

In a dose–response experiment, c doses of a treatment are administered to independent groups of subjects.

Let $\boldsymbol{X}_{ji} = (X_{1ji}, \ldots, X_{pji})' \in \mathbb{R}^p$ be a vector of response on p variables for the ith subject randomly assigned to treatment dose j, $j = 1, \ldots, c$, $i = 1, \ldots, n_j$, and let the total number of observations be $n = \sum_{j=1}^{c} n_j$.

Assume that $\boldsymbol{X}_{j1}, \ldots, \boldsymbol{X}_{jn_j}$ are n_j independent and identically distributed random vectors with a continuous distribution function $\boldsymbol{F}_j$ defined on $\mathbb{R}^p$ and finite mean $E(\boldsymbol{X}_j) = \boldsymbol{\mu}_j$, $j = 1, \ldots, c$, and denote by F_{hj} the hth marginal distribution for $\boldsymbol{F}_j$, $h = 1, \ldots, p$.

An appealing framework, often used when increasing doses of a treatment are being compared, assumes that the c distributions are stochastically ordered.

Inference based on stochastically ordered univariate random variables has been studied extensively, whereas stochastically ordered random vectors (see Marshall and Olkin, 1979, Chapter 17) has received much less attention.

The c multivariate distributions are said stochastically ordered, written

$$\boldsymbol{X}_1 \overset{st}{\leqslant} \boldsymbol{X}_2 \overset{st}{\leqslant} \ldots \overset{st}{\leqslant} \boldsymbol{X}_c \tag{1}$$

* Corresponding author.
E-mail addresses: livio.finos@unife.it (L. Finos), salmaso@gest.unipd.it (L. Salmaso), solari@stat.unipd.it (A. Solari).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2006.04.014

if and only if $E[f(X_1)] \leqslant E[f(X_2)] \leqslant \ldots \leqslant E[f(X_c)]$ holds for all increasing functions $f : \mathbb{R}^p \to \mathbb{R}$ such that the expectation exists.

We wish to test the null hypothesis

$$\mathrm{H}_0 : \boldsymbol{X}_1 \overset{d}{=} \ldots \overset{d}{=} \boldsymbol{X}_c$$

where $\overset{d}{=}$ means 'equal in distribution', against the alternative hypothesis

$$\mathrm{H}_1 : \boldsymbol{X}_1 \overset{st}{\leqslant} \ldots \overset{st}{\leqslant} \boldsymbol{X}_c \text{ and } \boldsymbol{X}_1, \ldots, \boldsymbol{X}_c \text{ are not equal in distribution.}$$

Note that when we apply a test of hypotheses, it is assumed that either H_0 or H_1 is true (Silvapulle and Sen, 2005), i.e. we must make the prior assumption that (1) holds. From Corollary 3 in Bacelli and Makowski (1989), it follows that under (1), H_0 holds if and only if $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_c$ have the same marginal distributions, namely

$$F_{h1}(x) = \ldots = F_{hc}(x) \quad \forall x \in \mathbb{R}, \quad h = 1, \ldots, p.$$

As a consequence, H_1 holds if and only if $X_{1h} \overset{st}{\leqslant} \ldots \overset{st}{\leqslant} X_{hc}$, $h=1, \ldots, p$, and $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_c$ are not equal in distribution, expressed as

$$F_{h1}(x) \geqslant \ldots \geqslant F_{hc}(x) \quad \forall x \in \mathbb{R}, \quad h = 1, \ldots, p.$$

with at least one strict inequality holding at some $x \in \mathbb{R}$ for at least one h.

In complex problems, as discussed in Roy et al. (1971), it is preferable to look upon H_0 as the intersection of a number of more primitive hypotheses and upon H_1 as the union of the same number of corresponding primitive alternatives, in symbols

$$H_0 \quad \Leftrightarrow \quad \bigcap_{h=1}^{p} \{H_{0h}\} : \bigcap_{h=1}^{p} \{F_{h1}(x) = \ldots = F_{hc}(x), \forall x \in \mathbb{R}\} \tag{2}$$

and

$$H_1 \quad \Leftrightarrow \quad \bigcup_{h=1}^{p} \{H_{1h}\} : \bigcup_{h=1}^{p} \{F_{h1}(x) \geqslant \ldots \geqslant F_{hc}(x), \forall x \in \mathbb{R} \text{ and not } H_{0h}\}. \tag{3}$$

In a nonparametric setup, let $\mathscr{F}$ be the class of all continuous distribution functions on $\mathbb{R}^p$, thus the simultaneous stochastically ordered random variable model is

$$\mathscr{F}^{\otimes c} = \{\boldsymbol{F}_1, \ldots, \boldsymbol{F}_c \in \mathscr{F} : F_{h1}(x) \geqslant \cdots \geqslant F_{hc}(x), \ \forall x \in \mathbb{R}, \quad h = 1, \ldots, p\}.$$

If we assume that the distributions $\boldsymbol{F}_1, \ldots, \boldsymbol{F}_c$ may differ only with respect to locations, i.e. $\boldsymbol{F}_j(\boldsymbol{x}) = \boldsymbol{F}(\boldsymbol{x} + \boldsymbol{\delta}_j)$, $\forall \boldsymbol{F} \in \mathscr{F}$, $\forall \boldsymbol{x}, \boldsymbol{\delta}_j \in \mathbb{R}^p$, $j = 1, \ldots, c$, then the previous model reduces to the homoscedastic location model

$$\mathscr{F}_{\boldsymbol{\delta}} = \{\boldsymbol{F} \in \mathscr{F} : \boldsymbol{F}_1(\boldsymbol{x} - \boldsymbol{\delta}_1) = \cdots = \boldsymbol{F}_c(\boldsymbol{x} - \boldsymbol{\delta}_c), \ \forall \boldsymbol{x}, \boldsymbol{\delta}_1 \leqslant \cdots \leqslant \boldsymbol{\delta}_c \in \mathbb{R}^p\},$$

where $\boldsymbol{a} \leqslant \boldsymbol{b}$ means $a_h \leqslant b_h$ for $h = 1, \ldots, p$.

Let $\mathscr{F}_{\mathscr{N}_p}$ be the class of all p-variate normal distributions; the normal homoscedastic location model corresponds to

$$\mathscr{F}_{\mathscr{N}_p} = \{\boldsymbol{F} \in \mathscr{F}_{\mathscr{N}_p} : \boldsymbol{F}_1(\boldsymbol{x} - \boldsymbol{\delta}_1) = \cdots = \boldsymbol{F}_c(\boldsymbol{x} - \boldsymbol{\delta}_c), \ \forall \boldsymbol{x}, \boldsymbol{\delta}_1 \leqslant \cdots \leqslant \boldsymbol{\delta}_c \in \mathbb{R}^p\}.$$

Obviously, $\mathscr{F}_{\mathscr{N}_p} \subseteq \mathscr{F}_{\boldsymbol{\delta}} \subseteq \mathscr{F}^{\otimes c}$. Under the homoscedastic location model, the testing problem becomes

$$\mathrm{H}_0^{\star} \quad \Leftrightarrow \quad \bigcap_{h=1}^{p} \{\mathrm{H}_{0h}^{\star}\} : \bigcap_{h=1}^{p} \{\mu_{h1} = \cdots = \mu_{hc}\} \tag{4}$$

against

$$\mathrm{H}_1^\star \;\Leftrightarrow\; \bigcup_{h=1}^{p}\{\mathrm{H}_{1h}^\star\} : \bigcup_{h=1}^{p}\{\mu_{h1}\leqslant\cdots\leqslant\mu_{hc},\ \text{and not } \mathrm{H}_{0h}^\star\}. \tag{5}$$

1.1. Further decomposition of the problem

Let us consider the hth sub-problem of testing H_{0h} against $\mathrm{H}_{1h}-\mathrm{H}_{0h}$. A nonparametric rank solution of this problem is given by the Jonckheere–Tepstra test. Alternatively, El Barmi and Mukerjee (2005) proposed an asymptotic test by using the sequential testing procedure developed in Hogg (1965). This procedure consists in testing sequentially $\mathrm{H}_{0h}^k : F_{h1}=\ldots=F_{h(k-1)}=F_{hk}$ against $\mathrm{H}_{1h}^k : F_{h1}=\cdots=F_{h(k-1)}\not\leqslant F_{hk}$, $k=2,\ldots,c$, as two-sample tests by pooling the first $(k-1)$ samples for testing H_{0h}^k, where $F_{h(k-1)}\not\leqslant F_{hk}$ means $F_{h(k-1)}(x)\leqslant F_{hk}(x)\forall x\in\mathbb{R}$ with strict inequality holding at some $x\in\mathbb{R}$. Thus, H_{0h} and H_{1h} can be expressed as $\mathrm{H}_{0h} \Leftrightarrow \bigcap_{k=2}^{c}\{\mathrm{H}_{h0}^k\} : \bigcap_{k=2}^{c}\{F_{h1}=\cdots=F_{h(k-1)}=F_{hk}\}$ and $\mathrm{H}_{1h} \Leftrightarrow \bigcup_{k=2}^{c}\{\mathrm{H}_{h1}^k\} : \bigcup_{k=2}^{c}\{F_{h1}=\cdots=F_{h(k-1)}\not\leqslant F_{hk}\}$, respectively. However, H_{0h} and H_{1h} can also be expressed as

$$\mathrm{H}_{h0} \;\Leftrightarrow\; \bigcap_{k=2}^{c}\{\mathrm{H}_{0hk}\} : \bigcap_{k=2}^{c}\{F_{h1}=\cdots=F_{h(k-1)}=F_{hk}=\cdots=F_{hc}\},$$

$$\mathrm{H}_{h1} \;\Leftrightarrow\; \bigcup_{k=2}^{c}\{\mathrm{H}_{1hk}\} : \bigcup_{k=2}^{c}\{F_{h1}=\ldots=F_{h(k-1)}\not\leqslant F_{hk}=\ldots=F_{hc}\},$$

where for testing H_{0hk} against H_{1hk} we refer to a two-sample problem by pooling the first $(k-1)$ and the last $(c-k+1)$ samples, $k=2,\ldots,c$.

Analogously, under the homoscedastic location model, we may decompose $\mathrm{H}_{0h}^\star$ and $\mathrm{H}_{1h}^\star$ as

$$\mathrm{H}_{h0}^\star \;\Leftrightarrow\; \bigcap_{k=2}^{c}\{\mathrm{H}_{0hk}^\star\} : \bigcap_{k=2}^{c}\{\mu_{h1}=\cdots=\mu_{h(k-1)}=\mu_{hk}=\cdots=\mu_{hc}\},$$

$$\mathrm{H}_{h1}^\star \;\Leftrightarrow\; \bigcup_{k=2}^{c}\{\mathrm{H}_{1hk}^\star\} : \bigcup_{k=2}^{c}\{\mu_{h1}=\cdots=\mu_{h(k-1)}<\mu_{hk}=\cdots=\mu_{hc}\}.$$

Thus, we can express the null hypotheses H_0 in (2) and $\mathrm{H}_0^\star$ in (4) as

$$\mathrm{H}_0 \;\Leftrightarrow\; \bigcap_{h=1}^{p}\left\{\bigcap_{k=2}^{c}[\mathrm{H}_{0hk}]\right\},\quad \mathrm{H}_0^\star \;\Leftrightarrow\; \bigcap_{h=1}^{p}\left\{\bigcap_{k=2}^{c}[\mathrm{H}_{0hk}^\star]\right\}, \tag{6}$$

and the alternative hypotheses H_1 in (3) and $\mathrm{H}_1^\star$ in (5) as

$$\mathrm{H}_1 \;\Leftrightarrow\; \bigcup_{h=1}^{p}\left\{\bigcup_{k=2}^{c}[\mathrm{H}_{1hk}]\right\},\quad \mathrm{H}_1^\star \;\Leftrightarrow\; \bigcup_{h=1}^{p}\left\{\bigcup_{k=2}^{c}[\mathrm{H}_{1hk}^\star]\right\}. \tag{7}$$

2. Combining dependent permutation tests

Let us now consider the general problem of simultaneous testing for a finite number of hypotheses. We shall assume that tests for the individual hypotheses are available and the problem is how to combine them into a simultaneous test procedure.

Since in general the considered tests and associated p-values are positively dependent, it is not allowed to use methods for combining independent p-values (Loughin, 2004).

A simple way to conduct a global test is with the Bonferroni procedure, which is potentially conservative, both because it is based on the Bonferroni inequality and because it ignores potential dependence between separate individual inferences.

Hence, we shall use a nonparametric combination method within the permutation paradigm to combine the 'partial' tests into a global test. For details on the nonparametric combination method we refer to Pesarin (2001).

In the permutation context let us denote the data set by the unit-by-unit representation $\mathbf{X}=\{X_l, l=1,\ldots,n;\ n_1,\ldots,n_c\}$ where it is intended that the first n_1 vectors in the list belong to the first sample, the following n_2 vectors to the second sample, and so on. Let us define the permutation sample space $\mathscr{X}^n_{/\mathbf{X}}$ as the set that contains all the $n!$ permutations of the observed data set $\mathbf{X}$. Consider the bth permutation $(\pi_1^{*b},\ldots,\pi_l^{*b}\ldots,\pi_n^{*b})$ of $(1,\ldots,l,\ldots,n)$, $b=1,\ldots,n!$, then $\mathbf{X}^{*b}=\{X_l^{*b}=X_{\pi_l^{*b}},\ l=1,\ldots,n;n_1,\ldots,n_c\}$ denotes the bth permutation of $\mathbf{X}$, and $\mathbf{X}_j^{*b}=\{X_{ji}^{*b}=X_{\pi^{*b}_{m_{j-1}+i}},\ i=1,\ldots,n_j\}$, $j=1,\ldots,c$, denotes the corresponding c permuted samples, where $m_0\equiv 0$ and $m_j=\sum_{l=1}^{j}n_l$.

In order to see how to combine dependent permutation tests for testing H_0 in (6) against H_1 in (7), suppose that a 'partial' permutation test of H_{0hk} against H_{1hk} is based on a test statistic $T_{hk}(\cdot)$ for which large values are significant, $h=1,\ldots,p$, $k=2,\ldots,c$. Let us denote the value of the test statistic computed on $\mathbf{X}$ and $\mathbf{X}^{*b}$ by $T_{hk}(\mathbf{X})=T^0_{hk}$ and $T_{hk}(\mathbf{X}^{*b})=T^{*b}_{hk}$, respectively.

The (conditional) 'partial' p-value related to the permutation test T^0_{hk} is

$$\lambda^0_{hk}=\lambda_{hk}(\mathbf{X})=\frac{\sum_{b=1}^{n!}\mathbb{I}\{T^{*b}_{hk}\geqslant T^0_{hk}\}}{n!},$$

where $\mathbb{I}\{\cdot\}$ denotes the indicator function. Let $S_{T_{hk}}(x)=\Pr\{T^{*b}_{hk}\geqslant x|\mathscr{X}^n_{/\mathbf{X}}\}=\sum_{b=1}^{n!}\mathbb{I}\{T^{*b}_{hk}\geqslant x\}/n!$, $\forall x\in\mathbb{R}$, be the permutation null 'survival'distribution of $T_{hk}(\cdot)$.

The nonparametric combination is developed in two stages:

- In the first step, for each component variable h, the combination is performed with respect to the $(c-1)$ partial tests T^0_{hk}, $k=2,\ldots,c$, thus we obtain p first-order combined permutation tests

  $$T'^0_h=\psi_f(\lambda^0_{h2},\ldots,\lambda^0_{hc}),\quad h=1,\ldots,p,$$

  where ψ_f is an admissible combining function. The first-order p-value related to the permutation test T'^0_h is then $\lambda'^0_h=\sum_{b=1}^{n!}\mathbb{I}\{T'^{*b}_h\geqslant T'^0_h\}/n!$, where $T'^{*b}_h=\psi_f(S_{T_{h2}}(T^{*b}_{h2}),\ldots,S_{T_{hc}}(T^{*b}_{hc}))$. Let $S_{T'_h}(x)=\sum_{b=1}^{n!}\mathbb{I}\{T'^{*b}_h\geqslant x\}/n!$, $\forall x\in\mathbb{R}$, be the permutation null 'survival' distribution of $T'_h(\cdot)$.
- In the second step, the combination is done with respect to the p variables, thus we obtain the (second-order) global test

  $$T''^0=\psi_s(\lambda'^0_1,\ldots,\lambda'^0_p),\tag{8}$$

  where ψ_s is an admissible combining function, which may not necessarily coincide with ψ_f.

Therefore, the null hypothesis H_0 is rejected at significance level α if

$$\frac{\sum_{b=1}^{n!}\mathbb{I}[T''^{*b}\geqslant T''^0]}{n!}=\lambda''^0\leqslant\alpha$$

because under the null hypothesis H_0 all permutations $\mathbf{X}^{*b}$ of $\mathbf{X}$ are equally likely, thus λ''^0 satisfies $\Pr_{\mathrm{H}_0}\{\lambda''^0\leqslant u|\mathscr{X}^n_{\mathbf{X}}\}\leqslant u$ for all $0\leqslant u\leqslant 1$, where $T''^{*b}=\psi_s(S_{T'_1}(T'^{*b}_1),\ldots,S_{T'_p}(T'^{*b}_p))$.

When the cardinality of $\mathscr{X}^n_{/\mathbf{X}}$ is large, we can inspect it by random sampling from $\mathscr{X}^n_{/\mathbf{X}}$ (with or without replacement), thus λ''^0 can be estimated to the desired degree of accuracy.

One more feature of this method is that, when the global analysis rejects H_0, one can apply multiple testing procedures in order to find which partial test (or group of tests) is (are) mostly responsible of global rejection. Thus, after p-value adjustment for multiplicity (for instance, by using the closed-testing approach, Westfall, Kropf and Finos, 2004), we can identify which sub-set of variables presents statistically significant ordering, that is, which adj-$\lambda'^0_h\leqslant\alpha$, $h=1,\ldots,p$.

2.1. Choice of combining function

Some well-known admissible combining functions which we consider in this paper include

$$\psi_T = \max_r (1-\lambda_r), \quad \psi_F = -2 \cdot \sum_r \log(\lambda_r), \quad \psi_L = \sum_r \Phi^{-1}(1-\lambda_r),$$

proposed, respectively, by Tippett, Fisher, and Liptak, and where Φ denotes the standard normal distribution function. In what follows, in order to emphasize the chosen combining functions, we use the notation $T''^{0}_{\psi_s\psi_f}$ and $\lambda''^{0}_{\psi_s\psi_f}$, where $f, s \in \{T, F, L\}$.

Of course, p-value $\lambda''^{0}_{\psi_s\psi_f}$ of combined test $T''^{0}_{\psi_s\psi_f}$ depends in general on which combining functions ψ_f and ψ_s are used, because different admissible combining functions have different convex acceptance regions.

In order to use a sort of neutral combining function, we may iterate the combining procedure (Salmaso and Solari, 2006), by applying to the same partial tests more than one combining function (we consider ψ_T, ψ_F, ψ_L), and then combine the resulting p-values by means of one combining function (we consider ψ_F).

Fisher's iterated combining function is denoted by

$$\psi_I = \psi_F(\lambda_{\psi_T}, \lambda_{\psi_F}, \lambda_{\psi_L}).$$

For instance, $T''^{0}_{\psi_I\psi_f} = \psi_F(\lambda''^{0}_{\psi_T\psi_f}, \lambda''^{0}_{\psi_F\psi_f}, \lambda''^{0}_{\psi_L\psi_f})$, $f \in \{T, F, L, I\}$, when we combine the p first-order permutation partial tests.

2.2. Choice of test statistic

The most obvious feature that has not been dealt with is the choice of test statistics.

The proposed approach consists of breaking down the problem so that each of the component problems $(\mathrm{H}_{0hk}, \mathrm{H}_{1hk})$, $h = 1, \ldots, p$, $k = 2, \ldots, c$, can be tested using simple two-sample test statistics for stochastic ordering alternatives; for instance, using the Kolmogorov–Smirnov or the Wilcoxon–Mann–Whitney test statistics.

Under the normal homoscedastic location model, the one-sided Student t test

$$t_{hk}(\mathbf{X}) = \frac{(\bar{X}_{h,k:c} - \bar{X}_{h,1:k-1})/\sqrt{1/\sum_{j=1}^{k-1} n_j + 1/\sum_{j=k}^{c} n_j}}{\sqrt{\left[\sum_{j=1}^{k-1}\sum_{i=1}^{n_j}(X_{hji} - \bar{X}_{h,1:k-1})^2 + \sum_{j=k}^{c}\sum_{i=1}^{n_j}(X_{hji} - \bar{X}_{h,k:c})^2\right]/(n-2)}},$$

where $\bar{X}_{h,r:s} = \sum_{j=r}^{s}\sum_{i=1}^{n_j} X_{hji}/\sum_{j=r}^{s} n_j$, $1 \leqslant r \leqslant s \leqslant c$, is the uniformly most powerful similar test for testing $\mathrm{H}^{\star}_{0hk}$ against $\mathrm{H}^{\star}_{1hk}$ and it is asymptotically equivalent to (Hoeffding, 1952) the corresponding permutation test based on

$$T_{hk}(\mathbf{X}) = \bar{X}_{h,k:c}. \tag{9}$$

This is the most powerful permutation test against the normal alternatives with common variance among all permutation tests which are unbiased and of level α. More generally, it is also unbiased against stochastic ordering alternatives for all pairs of continuous distributions (see Lehmann and Romano, 2005, Section, 5.9).

3. A brief outline of the literature

The literature contains a variety of test statistics for testing the equality of vector means against a multivariate ordered alternative under the homoscedastic location model, that is, for testing $\mathrm{H}^{\star}_0$ in (4) against $\mathrm{H}^{\star}_1$ in (5).

Dietz (1989) propose an asymptotically distribution-free test that generalizes the Jonckheere–Tepstra test to the multivariate case, which is based on

$$J(\mathbf{X}) = \frac{n^{-3/2}\sum_{h=1}^{p} J_h}{\sqrt{\mathbf{1}'\boldsymbol{\Upsilon}\mathbf{1}}}, \tag{10}$$

where $\mathbf{1}$ is a $p \times 1$ vector of ones, J_h, $h = 1, \ldots, p$, is the Jonckheere–Tepstra statistic computed on the h variable, and the covariance matrix $\boldsymbol{\Upsilon}$ has diagonal and off-diagonal elements $n^{-3}\,\mathrm{Var}(J_h)$ and $n^{-3}\,\mathrm{Cov}(J_h, J_g)$, respectively, for $h \neq g = 1, \ldots, p$. The asymptotic null distribution of J is standard normal.

By using the maxmin criterion of Abelson and Tukey (1989), Sim and Johnson (2004) generalize the contrast test developed by Schaafsma and Smid (1966) to the multivariate case, under the assumption that the covariance matrix $\boldsymbol{\Sigma} = \mathrm{Var}(\boldsymbol{X}_1)$ is known. The test rejects $\mathrm{H}_0^\star$ for large values of

$$L(\mathbf{X}) = \sum_{h=1}^{p}\sum_{g=1}^{p} \sigma^{hg}\left(\sum_{j=1}^{c} b_j \sum_{i=1}^{n_j} X_{hji}\right), \tag{11}$$

where the $\{\sigma^{hg}\}$, $h, g = 1, \ldots, p$, are the elements of $\boldsymbol{\Sigma}^{-1}$ and the b_j's, which satisfy $\sum_{j=1}^{c} n_j b_j = 0$, are the contrast coefficients obtained by Schaafsma and Smid (1966):

$$b_j = \frac{1}{\sqrt{n n_j}}\{\sqrt{m_{j-1}(n - m_{j-1})} - \sqrt{m_j(n - m_j)}\}, \quad j = 1, \ldots, c, \tag{12}$$

where $m_0 \equiv 0$ and $m_l = \sum_{j=1}^{l} n_j$, $l = 1, \ldots, c$. When the underlying distribution is multivariate normal, the null distribution of $L/\sqrt{(\boldsymbol{\Sigma}^{-1}\mathbf{1})'\mathbf{1}\sum_{j=1}^{c} n_j b_j^2}$ is standard normal.

One may rewrite the hypotheses by the reparametrization $\vartheta_{hk} = \mu_{hk} - \mu_{h(k-1)}$ for $k = 2, \ldots, c$, thus the problem is rearranged in one-sample setting as $\mathrm{H}_0^\star : \bigcap_{h=1}^{p}\bigcap_{k=2}^{c}\{\vartheta_{hk} = 0\}$ against $\mathrm{H}_1^\star : \bigcup_{h=1}^{p}\bigcup_{k=2}^{c}\{\vartheta_{hk} > 0\}$ with at least one strict inequality at some pair (h, k). Note that the dimensionality of the problem is now $q = p(c-1)$. The likelihood ratio test based on a single observation $((\bar{\boldsymbol{X}}_2 - \bar{\boldsymbol{X}}_1)', \ldots, (\bar{\boldsymbol{X}}_c - \bar{\boldsymbol{X}}_{c-1})')'$ from a multivariate normal distribution $N_q(\boldsymbol{\vartheta}, \boldsymbol{\Lambda})$, when $\boldsymbol{\Lambda}$ is assumed to be known, has null distribution Chi-Bar-Square, i.e. $\Pr\{\bar{\chi}^2 \leqslant c|\mathrm{H}_0\} = \sum_{h=0}^{q} w_h(q, \boldsymbol{\Lambda}, \mathbb{R}^{+q})\Pr\{\chi_h^2 \leqslant c\}$,where χ_h^2 are central Chi-square variables with h degrees of freedom ($h = 0, \ldots, q$ and $\chi_0^2 \equiv 0$), and weights $w_h(q, \boldsymbol{\Lambda}, \mathbb{R}^{+q})$ that can be explicitly determined up to $q = 4$.

In order to keep the dimension more manageable in the c-sample multivariate setting, Sim and Johnson (2004) propose a likelihood ratio test by using the contrast coefficients obtained by Schaafsma and Smid (1966), which is based on

$$LR(\mathbf{X}) = \frac{1}{\sum_{j=1}^{c} n_j b_j^2}\left\{\mathbf{Y}'\boldsymbol{\Sigma}^{-1}\mathbf{Y} - \min_{\boldsymbol{\theta} \geqslant \mathbf{0}}[(\mathbf{Y} - \boldsymbol{\theta})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\theta})]\right\}, \tag{13}$$

where the b_j's are given in (12), $\mathbf{Y} = \sum_{j=1}^{c} b_j \sum_{i=1}^{n_j} \boldsymbol{X}_{ji}$ and $\boldsymbol{\theta} = E(\mathbf{Y}) = \sum_{j=1}^{c} b_j n_j \boldsymbol{\mu}_j$. When $\boldsymbol{\Sigma}$ is assumed to be known, LR in (13) has null distribution $\Pr\{LR \leqslant c\} = \sum_{h=0}^{p} w_h(p, \boldsymbol{\Sigma}, \mathbb{R}^{+p})\Pr\{\chi_h^2 \leqslant c\}$.

4. Comparative simulation study

Sim and Johnson (2004) consider an empirical power study under the normal homoscedastic location model, concerning three samples ($c = 3$) of bivariate ($p = 2$) normal variables with correlation ρ_{12}, where the mean for the first sample was $\mathbf{0}$, and for the second and third sample, they added $\boldsymbol{\mu}_2$ and $\boldsymbol{\mu}_3$, respectively.

We replicate this study at $\alpha = 5\%$ with 1000 Monte Carlo simulations and 1000 samples from the permutation sample space, with equal sample size $n_1 = n_2 = n_3 = 5$ and $\rho_{12} = -0.7, 0, 0.7$, but we multiply $\boldsymbol{\mu}_j$ by $\sqrt{1 - \rho_{12}^2}$, in order to reduce the shifts when there is correlation between the two variables.

As in Sim and Johnson (2004), we consider the test LR in (13) and L in (11), but when $\boldsymbol{\Sigma}$ is estimated from the data by $\mathbf{S}_n = \{s_{hg}\}$, with

$$s_{hg} = \frac{\sum_{j=1}^{c}\sum_{i=1}^{n_j}(X_{hji} - \bar{X}_{hj})(X_{gji} - \bar{X}_{gj})}{n - c}, \quad h, g = 1, \ldots, p,$$

because knowledge of $\boldsymbol{\Sigma}$ can be rarely assumed.

Table 1
Empirical power of the considered tests when $\mu_{11}=\mu_{12}=0$

ρ	$\boldsymbol{\mu}_2'$	$\boldsymbol{\mu}_3'$											
		(0.5, 0.5)				(1.0, 1.0)				(1.5, 1.5)			
		LR	*L*	*J*	$T''^{0}_{\psi_L\psi_L}$	*LR*	*L*	*J*	$T''^{0}_{\psi_L\psi_L}$	*LR*	*L*	*J*	$T''^{0}_{\psi_L\psi_L}$
−0.7	(0.0, 0.0)	264	351	314	355	676	786	709	798	924	963	923	974
	(0.0, 0.5)	275	366	331	368	721	830	759	828	955	974	954	983
	(0.0, 1.0)					745	840	783	850	975	985	970	987
	(0.0, 1.5)									976	986	966	990
					$T''^{0}_{\psi_L\psi_I}$				$T''^{0}_{\psi_L\psi_I}$				$T''^{0}_{\psi_L\psi_I}$
0	(0.0, 0.0)	202	255	259	262	481	609	592	658	792	868	856	929
	(0.0, 0.5)	200	265	272	271	511	643	629	667	817	887	890	929
	(0.0, 1.0)					523	646	645	689	835	904	900	940
	(0.0, 1.5)									836	914	899	949
0.7	(0.0, 0.0)	118	144	148	146	235	289	288	312	414	488	484	532
	(0.0, 0.5)	118	147	146	153	236	295	308	313	425	501	503	523
	(0.0, 1.0)					221	311	313	325	417	513	518	532
	(0.0, 1.5)									394	533	518	556

Table 2
Empirical power of the considered tests when $\mu_{11}=\mu_{12}=\mu_{13}=0$

ρ	$\boldsymbol{\mu}_2'$	$\boldsymbol{\mu}_3'$											
		(0.0, 0.5)				(0.0, 1.0)				(0.0, 1.5)			
		LR	*L*	*J*	$T''^{0}_{\psi_L\psi_L}$	*LR*	*L*	*J*	$T''^{0}_{\psi_L\psi_L}$	*LR*	*L*	*J*	$T''^{0}_{\psi_L\psi_L}$
−0.7	(0.0, 0.0)	117	176	180	179	287	374	335	388	491	554	506	596
	(0.0, 0.5)	112	182	164	186	293	392	348	414	538	608	560	632
	(0.0, 1.0)					271	380	339	388	536	601	545	643
	(0.0, 1.5)									495	563	513	597
					$T''^{0}_{\psi_T\psi_I}$				$T''^{0}_{\psi_T\psi_I}$				$T''^{0}_{\psi_T\psi_I}$
0	(0.0, 0.0)	128	140	141	163	268	233	257	342	474	304	383	629
	(0.0, 0.5)	132	147	144	156	293	242	280	330	512	344	431	598
	(0.0, 1.0)					266	236	263	356	515	339	438	601
	(0.0, 1.5)									470	316	387	619
0.7	(0.0, 0.0)	114	89	89	86	257	111	131	184	472	118	198	373
	(0.0, 0.5)	113	86	88	89	280	113	137	182	516	118	206	357
	(0.0, 1.0)					256	107	132	191	505	110	195	361
	(0.0, 1.5)									463	108	193	370

In Tables 1–3, we display the number of rejections out of 1000 tests for *LR* and *L* based on $\mathbf{S}_n$, for *J* in (10), and for $T''^{0}_{\psi_s\psi_f}$ in (8), by choosing $T_{hk}(\cdot)$ in (9) as partial test statistics and

- when $\rho=-0.7$, $f, s=L$;
- when $\rho\neq-0.7$, $f=I$, and $s=L$ when both alternatives H_{11} and H_{12} are true (Tables 1 and 3) or $s=T$ when only one of these is true (Table 2).

Sim and Johnson (2004) find that the power of *L* is higher than *LR* when $\boldsymbol{\mu}_2$ and/or $\boldsymbol{\mu}_3$ are in the direction of $\mathbf{1}=(1,1)$, whereas is lower than *LR* when both $\boldsymbol{\mu}_2$ and $\boldsymbol{\mu}_3$ are along the *y*-axis, particularly when ρ_{12} is close to 1. This pattern

Table 3
Empirical power of the considered tests when $\boldsymbol{\mu}_2$ and $\boldsymbol{\mu}_3$ are in equal angular direction

ρ	$\boldsymbol{\mu}_2'$	$\boldsymbol{\mu}_3'$											
		(0.5, 0.5)				(1.0, 1.0)				(1.5, 1.5)			
		LR	L	J	$T''^{0}_{\psi_L\psi_L}$	LR	L	J	$T''^{0}_{\psi_L\psi_L}$	LR	L	J	$T''^{0}_{\psi_L\psi_L}$
−0.7	(0.5, 0.5)	236	370	302	368	771	851	780	861	974	984	972	989
	(1.0, 1.0)					699	809	721	824	969	984	968	987
	(1.5, 1.5)									925	965	948	972
					$T''^{0}_{\psi_L\psi_I}$				$T''^{0}_{\psi_L\psi_I}$				$T''^{0}_{\psi_L\psi_I}$
0	(0.5, 0.5)	181	269	261	274	542	649	651	657	835	901	913	921
	(1.0, 1.0)					498	607	603	669	838	900	909	920
	(1.5, 1.5)									772	864	867	930
0.7	(0.5, 0.5)	114	143	138	140	235	283	297	297	408	493	506	512
	(1.0, 1.0)					231	270	282	302	406	497	510	516
	(1.5, 1.5)									387	472	469	521

Table 4
Serum enzyme levels for 40 rats; serum enzyme levels in international units/liter; dosage of vinylidene fluoride in parts/million

Dosage	Enzyme	Rat within dosage									
		1	2	3	4	5	6	7	8	9	10
0	SDH	18	27	16	21	26	22	17	27	26	27
	SGOT	101	103	90	98	101	92	123	105	92	88
1500	SDH	25	21	24	19	21	22	20	25	24	27
	SGOT	113	99	102	144	109	135	100	95	89	98
5000	SDH	22	21	22	30	25	21	29	22	24	21
	SGOT	88	95	104	92	103	96	100	122	102	107
15,000	SDH	31	26	28	24	33	23	27	24	28	29
	SGOT	104	123	105	98	167	111	130	93	99	99

remains the same when $\boldsymbol{\Sigma}$ is estimated by $\mathbf{S}_n$, except when $\rho_{12} = -0.7$; in this case L has higher power than LR. The conclusions that emerge from the simulation study are that powers of considered tests can vary greatly depending on the case considered. Apart for the case when $\mu_{11} = \mu_{12} = \mu_{13} = 0$ and $\rho = 0.7$, the power of $T''^{0}_{\psi_s\psi_f}$ is very often the highest.

5. An example from the literature

This example is taken from Dietz (1989). Group of 10 ‘Fisher 344’ male rats received by inhalation exposure one of four dosages of vinylidene fluoride, a chemical suspected of causing liver damage. Among the response variables measured on the rats were two serum enzymes: SDH and SGOT. Increasing levels of these enzymes are often associated with liver damage. It is of interest to test whether each enzyme level stochastically increases with increasing doses of vinylidene fluoride. The data are shown in Table 4.

For testing H_0 against $\mathrm{H}_1 : \bigcup_{h=1}^{2}\{X_{h1} \overset{st}{\leqslant} \cdots \overset{st}{\leqslant} X_{h4}\}$, $T''^{0}_{\psi_L\psi_I}$ based on 10,000 samples from the permutation sample space rejects H_0 at $\lambda''^{0}_{\psi_L\psi_I} = 0.0007$, with adj-$\lambda'^{0}_1 = 0.0035$ and adj-$\lambda'^{0}_2 = 0.0368$. If we perform two Bonferroni-corrected univariate Jonckheere–Tepstra tests, we obtain $J_1/\mathrm{Var}(J_1) = 2.45$ and $J_2/\mathrm{Var}(J_2) = 1.44$, thus only SDH is significant at $2\ \Pr\{\mathcal{N}(0,1) > 2.45\} = 0.0143$.

If we assume the normal homoscedastic location model, for testing $\mathrm{H}_0^\star : \bigcap_{h=1}^2 \{\mu_{h1} = \cdots = \mu_{h4}\}$ against $\mathrm{H}_1^\star : \bigcup_{h=1}^2 \{\mu_{h1} \leqslant \cdots \leqslant \mu_{h4}\}$, the J statistic equals 2.72, significant at $\Pr\{\mathcal{N}(0,1) > 2.72\} = 0.0033$, the L statistic equals 2.886, significant at $\Pr\{\mathcal{N}(0,1) > 2.886\} = 0.0020$, and the LR statistic equals 9.4225, significant at $\Pr\{\bar{\chi}^2 > 9.4225\} = 0.0041$.

6. Conclusions

When testing for multivariate ordered alternatives and multivariate normality and knowledge of Σ cannot be assumed, it might not be possible to obtain a test that performs optimally. The proposed solution, without assuming knowledge

(1) of underlying distribution,
(2) of fixed effect model for responses,
(3) of any form of dependence among variables,

is conveniently usable in this complex setup, because it decomposes the global problem into a finite number of component problems that are more manageable from a statistical point of view. The conclusions that emerge from the simulation study are that the powers of considered tests can vary greatly depending on the case considered, but the proposed permutation test exhibits good performances under the most favorable conditions for the competitor tests.

References

Abelson, R.P., Tukey, J.W., 1989. Efficient utilization of non-numerical information in quantitative analysis: general theory and the case of the simple order.. Ann. Math. Statist. 34, 1347–1369.

Bacelli, F., Makowski, A.M., 1989. Multidimensional stochastic ordering and associated random variables. Oper. Res. 37, 478–487.

Dietz, E.J., 1989. Multivariate generalizations of Jonckheere's test for ordered alternatives. Comm. Statist—Theory Methods 18, 3763–3783.

El Barmi, H., Mukerjee, H., 2005. Inferences under a stochastic ordering constraint: the *k*-sample case. J. Amer. Statist. Assoc. 100, 252–261.

Hoeffding, W., 1952. The large-sample power of tests based on permutations of observations. Ann. Math. Statist. 23, 169–192.

Hogg, R.V., 1965. On models and hypotheses with restricted alternatives. J. Amer. Statist. Assoc. 60, 1153–1162.

Lehmann E.L., Romano J., 2005. Testing Statistical Hypotheses. third ed. Springer, Berlin.

Loughin, T.M., 2004. A systematic comparison of methods for combining *p*-values from independent tests. Comput. Statist. Data Anal. 47, 467–485.

Marshall A., Olkin I., 1979. Inequalities: Theory of Majorization and Its Applications. Academic Press, New York.

Pesarin F., 2001. Multivariate Permutation Tests with Applications in Biostatistics, Wiley, Chichester.

Roy S.N., Gnanadesikan R., Srivastava J.N., 1971. Analysis & Design of Certain Quantitative Multiresponse Experiments. Pergamon Press Ltd., Oxford.

Salmaso, L., Solari, A., 2006. Nonparametric iterated combined tests for genetic differentiation. Comput. Statist. Data Anal. 50, 1105–1112.

Schaafsma, W., Smid, L.J., 1966. Most stringent somewhere most powerful tests against alternatives restricted by a number of linear inequalities. Ann. Math. Statist. 37, 1161–1172.

Silvapulle J.S., Sen P.K., 2005. Constrained Statistical Inference. Inequality, Order, and Shape Restictions. Wiley, New Jersey.

Sim, S., Johnson, R.A., 2004. New tests for multivariate ordered alternatives. Comm. Statist. Theory Methods 33, 2027–2039.

Westfall P.H., Kropf S., Finos L., 2004. Weighted FWE-controlling methods in high-dimensional situations. In: Y. Benjamini, F. Bretz, S. Sarkar (Eds.), Recent developments in Multiple Comparison Procedures, Institute of Mathematical Statistics Lecture Notes — Monograph Series 47, pp. 143-154.