Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v16n2p323

Different views of the multivariate ranking problem

By Corain et al.

14 October 2023

Electronic Journal of Applied Statistical Analysis
Vol. 16, Issue 2, 2023, 323-338
DOI: 10.1285/i20705948v16n2p323

Different views of the multivariate ranking problem

Livio Corain[a], Rosa Arboretti[b], Stefano Bonnini[c], Riccardo Ceccato*[a], and Luigi Salmaso[a]

[a]*Department of Management and Engineering, University of Padova, Vicenza, Italy*
[b]*Deparment of Civil, Environmental and Architectural Engineering, University of Padova, Padova, Italy*
[c]*Department of Economics and Management, University of Ferrara, Ferrara, Italy*

14 October 2023

Multivariate ranking problems are characterized by the need of ordering C different items according to several different features. The multivariate nature of these problems makes them quite challenging and flexible multivariate statistical techniques are therefore required. In this study we focus on two different scenarios, where we need to rank C different populations. Under the first scenario, preliminary knowledge about the order of the populations is available, while under the second one no information is available. Two solutions, based on the Nonparametric combination (NPC) technique, are proposed to deal with these scenarios and two case studies are adopted to facilitate the comprehension of the methods and to highlights the main differences between the two considered multivariate ranking problems.

keywords: multivariate ranking, multivariate stochastic ordering, nonparametric combination, permutation test.

1 Introduction

The term ranking problem is used to refer to a wide variety of situations characterized by the need of ordering C different items according to one or more features and which are quite common in many industrial fields. For example, let us consider a new product development process and suppose that a company is interested in launching a new

*Corresponding authors: riccardo.ceccato.1@unipd.it

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

product with four different available configurations. For the product to be successful, they essentially need to rank the configurations according to their performances in order to understand which configurations should be prioritized. Appropriate statistical techniques need therefore to be applied in such situations, so that the first ranking position can be assigned to the item which significantly outperforms the others, the second one to the second best item and so on. However, the performance of a product is rarely defined in terms of a single feature or variable. The term multivariate ranking problem refers to such situations, in which multiple relevant features need to be considered to determine a global ranking of the items.

Given the multivariate nature of the problem, multivariate statistical techniques are required to solve it. Let us consider the basic scenario where we need to order N units within a sample according to V different features. When $V = 1$, a ranking can be simply derived by computing $\sum_{i=1}^{N} \mathbb{I}(X_j \geq X_i)$ for each unit j. When $V > 1$, the task becomes more challenging.

Several visual techniques have been proposed to deal with multivariate scenarios. Ordination methods (Syms, 2008), which are widely adopted in ecology, have been proposed to display units along a reduced number (commonly lower than 4) of dimensions (or axes), representing the existing differences between items. The basic idea is to summarize multivariate data by proposing an adequate low-dimensional ordination space onto which data are projected. The inspection of the resulting plot allows us to study the relative positions of units in a reduced space. Principal Component Analysis, Principal Coordinate Analysis, and Multidimensional Scaling are some popular ordination techniques (Gower, 1987).

According to Li and Liu (2004), the notion of data depth can also be used to introduce a natural center-outward ordering of the units in a multivariate sample. A depth value indeed measure the centrality or the outlyingness of an observed unit $\mathbf{x} \in \mathbb{R}^V$ with respect to its underlying multivariate distribution $\mathbf{F}$. It is therefore possible to associate a depth value to each sample point and order units accordingly. Data depth is also used to construct multivariate nonparametric tests (Li and Liu, 2004; Chenouri and Small, 2012; Chenouri et al., 2020).

A different and quite challenging ranking task consists in ordering multivariate populations. Corain et al. (2017) widely investigated the problem and identified several different families of statistical methods proposed to deal with it and among them:

- multiple comparison procedures,
- stochastic ordering techniques,
- selection and ranking methods,
- ranking models.

In this paper we decided to focus on multiple comparison procedures and stochastic ordering techniques and highlight similarities and dissimilarities between these approaches. In particular, we show how multiple comparison procedures rely on the idea of performing all the possible pair-wise comparisons between populations in order to identify their existing order, while stochastic ordering techniques are intended to test a specific order suggested by some a-priori knowledge. In both cases however comparisons between multivariate samples need to be performed and for this reason the adoption of the Non-parametric combination (NPC) methodology (Pesarin and Salmaso, 2010; Corain and Salmaso, 2013; Corain et al., 2014) represents a suitable solution as we are going to show in this paper.

Section 2 is dedicated to the introduction of the stochastic ordering problem and the related NPC-based solution. Then, Section 3 provides a solution to the generic multivariate ranking problem (when no preliminary knowledge is available), based on the conduction of multiple comparisons through the NPC methodology. Section 4 is dedicated to a couple of case studies, where the previously introduced techniques are adopted. Finally, Section 5 is devoted to conclusions and final remarks.

2 Stochastic ordering

Stochastic ordering refers to the specific scenario in which we are interested in evaluating the existing order among C different populations, but prior-knowledge is available. In other words, a specific stochastic order is considered and tested by means of appropriate statistical techniques.

Let us consider two multivariate variables $\mathbf{X}_1$ and $\mathbf{X}_2$ with distribution function $\mathbf{F}_1$ and $\mathbf{F}_2$ respectively. Let V be the number of components of each multivariate variable. $\mathbf{X}_1$ stochastically dominates $\mathbf{X}_2$, written $\mathbf{X}_1 \overset{d}{>} \mathbf{X}_2$, if and only if $E[g(\mathbf{X}_1)] > E[g(\mathbf{X}_2)$ for each increasing function $g : \mathbb{R}^V \to \mathbb{R}$ such that the expectation $E[\cdot]$ exists. It follows that $\mathbf{X}_1$ stochastically dominates $\mathbf{X}_2$ if and only if $F_{1v}(x) \leq F_{2v}(x), \forall x \in \mathbb{R}, v = 1, \ldots, V$ and $\exists I : F_{1v}(x) < F_{2v}(x), x \in I$ for at least one v (Scarsini and Shaked, 1990; Corain et al., 2017).

Let us now consider C multivariate variables $\mathbf{X}_c, c = 1, \ldots, C$. In a stochastic ordering problem we are interested in investigating the following system of hypotheses:

$$\begin{cases} H_0 : \mathbf{F}_1 = \mathbf{F}_2 = \ldots = \mathbf{F}_{(C-1)} = \mathbf{F}_C \\ H_1 : \mathbf{F}_1 \leq \mathbf{F}_2 \leq \ldots \leq \mathbf{F}_{(C-1)} \leq \mathbf{F}_C \end{cases} \quad \text{and at least one strict inequality,} \tag{1}$$

where $\mathbf{F}_j \leq \mathbf{F}_k$ means that $\mathbf{X}_j$ stochastically dominates $\mathbf{X}_k$ for each $j < k$. It is worth noting that H_0 implies exchangeability of the data between groups. Under H_0 the C-level treatment yields no effects, so leaving unaffected the distributions (i.e. $\mathbf{F}_c = \mathbf{F}, c = 1, \ldots, C$), whereas under H1 it can yield non-decreasing non-negative effects.

For the univariate scenario (i.e. $V = 1$), several different solutions have been proposed in the literature, such as the Jonckheere-Terpstra test (Jonckheere, 1954; Terpstra, 1952) and Cuzick's test (Cuzick, 1985).

The first one is a non-parametric test which involves the conduction of multiple pair-wise comparisons using the Mann-Whitney test statistic:

$$MW_{jk} = \sum_{i=1}^{n_j} \sum_{l=1}^{n_k} [\mathbb{I}(X_{ji} < X_{kl}) + 0.5\mathbb{I}(X_{ji} = X_{kl})], j, k = 1, \ldots, C, j \neq k$$

where $\mathbb{I}(\cdot)$ is 1 if condition $\cdot$ is satisfied and 0 otherwise, and n_j and n_k are the sample sizes of group j and k respectively. $T^{JT} = \sum_{j=1}^{(C-1)} \sum_{k=j+1}^{C} MW_{jk}$ is finally used to evaluate the stochastic ordering problem, adopting critical values provided in the literature for small sample sizes and a normal approximation for large samples (Jonckheere, 1954).

Cuzick's test (Cuzick, 1985) firstly assign scores w_c to the C groups according to their ordering: 1 to the first group, 2 to the second, and so on. The test statistic $T^C = \sum_{c=1}^{C} w_c S_c$ is then retrieved, where S_c is the sum of the ranks calculated within group c. Finally, the test statistic is commonly standardized so that the distribution of $Z^C = (T^C - \mu^C)/\sigma^C$ can be approximated to a standard normal distribution.

Finally, permutation-based solutions involving the Nonparametric combination (NPC) methodology (Pesarin and Salmaso, 2010) have been proposed in the literature. The most appealing characteristic of these solutions is that they can be easily extended to multivariate scenarios, introducing an appropriate decomposition of the hypotheses and a further combination step.

Let us therefore focus directly on the comparison of multivariate populations. A two-step decomposition must be applied, with the first split of the hypotheses which is made in order to test each of the V different variables marginally (see system of hypotheses 2) and then the second split which is performed to recreate the conditions of a two-sample problem (see system of hypotheses 3).

$$\begin{cases} H_0 : \bigcap_{v=1}^{V} H_{0v} = \bigcap_{v=1}^{V} [F_{1v} = F_{2v} = \cdots = F_{(C-1)v} = F_{Cv}] \\ H_1 : \bigcup_{v=1}^{V} H_{1v} = \bigcup_{v=1}^{V} [F_{1v} \geq F_{2v} \geq \cdots \geq F_{(C-1)v} \geq F_{Cv}] \end{cases} \tag{2}$$

$$\begin{cases} H_{v0} : \bigcap_{c=2}^{C} H_{0cv} = \bigcap_{c=2}^{C} [(F_{1v} = \cdots = F_{(c-1)v}) = (F_{cv} = \cdots = F_{Cv})] \\ H_{v1} : \bigcup_{c=2}^{C} H_{1cv} = \bigcup_{c=2}^{C} [(F_{1v} = \cdots = F_{(c-1)v}) > (F_{cv} = \cdots = F_{Cv})] \end{cases} \tag{3}$$

For each sub-hypothesis H_{0cv}, the first $(c-1)$ and the last $(C-c+1)$ samples are pooled into two new samples Z_{1v}^c and Z_{2v}^c of size N and M respectively, so that a basic two-sample problem can be addressed:

$$\begin{cases} H_{0cv}^{new} : Z_{1v}^c \stackrel{d}{=} Z_{2v}^c \\ H_{1cv}^{new} : Z_{1v}^c \stackrel{d}{<} Z_{2v}^c. \end{cases}$$

Each sub-hypothesis H_{0cv}^{new} is tested marginally by means of appropriate permutation tests. It is worth noting that the NPC methodology, due to its conditioning on the whole sample data $\mathbf{X}$ considered as a set of sufficient statistics for the underlying common V-dimensional distribution $\mathbf{F}$, allows us to implicitly take into account dependency between the V components of the multivariate outcome, permuting rows of the matrix:

$$\begin{bmatrix} Z^c_{111} & Z^c_{112} & \dots & Z^c_{11(V-1)} & Z^c_{11V} \\ \dots & \dots & \dots & \dots & \dots \\ Z^c_{N11} & Z^c_{N12} & \dots & Z^c_{N1(V-1)} & Z^c_{N1V} \\ Z^c_{121} & Z^c_{122} & \dots & Z^c_{12(V-1)} & Z^c_{12V} \\ \dots & \dots & \dots & \dots & \dots \\ Z^c_{M21} & Z^c_{M22} & \dots & Z^c_{M2(V-1)} & Z^c_{M2V} \end{bmatrix}$$

A two-step combination is then performed. Firstly, for each $v = 1, \cdots, V$, the p-values related to the $C-1$ sub-problems $\{H_{0cv} \text{ vs } H_{1cv}\}$, $c = 2, \dots, C$ are combined using an appropriate combining function $\psi(\cdot)$. Secondly, the V combined p-value vectors just obtained are combined using a combining function $\theta(\cdot)$, in order to achieve a global p-value λ'' which can be compared with the desired significance level α to address system of hypotheses 1.

The main steps of this approach are summarized by the following algorithm:

1. For $c = 2, \dots, C$:
   a) Pool the first $(c-1)$ and the last $(C-c+1)$ samples. $\mathbf{Z}^c_1$ and $\mathbf{Z}^c_2$ are retrieved.
   b) Apply the test statistic to the pooled sample $\mathbf{Z}^c = \{\mathbf{Z}^c_1, \mathbf{Z}^c_2\}$. The vector $\mathbf{T}^o = \mathbf{T}(\mathbf{Z^c})$ is retrieved.
   c) For $b = 1, \dots, B$, retrieve $\mathbf{T}^{b*} = \mathbf{T}(\mathbf{Z^{c*}})$, where $\mathbf{Z}^{c*}$ is random permutation of $\mathbf{Z}^c$. $\mathbf{T}^{b*}, b = 1, \dots, B$ is a random sampling from the permutation V-variate distribution of $\mathbf{T}$.
   d) Estimate marginal p-values as $\hat{\lambda}_{cv} = \hat{L}_v(T^o_v|\mathcal{Z}^\rfloor{}_{/\mathbf{Z^c}}) = \frac{\left[\frac{1}{2}+\sum_b \mathbb{I}(T^{b*}_v \geq T^o_v)\right]}{(B+1)}$, $v = 1, \dots, V$ and simulate the permutation distributions $\hat{\lambda}^{r*}_{cv} = \hat{L}_v(T^{r*}_v|\mathcal{Z}^\rfloor{}_{/\mathbf{Z^c}}) = \frac{\left[\frac{1}{2}+\sum_b \mathbb{I}(T^{b*}_v \geq T^{r*}_v)\right]}{(B+1)}$, $v = 1, \dots, V,\ r = 1, \dots, B$.
2. Compute the combined second-order tests as $T^o_{\psi v} = \psi(\hat{\lambda}_{1v}, \dots, \hat{\lambda}_{Cv})$, $v = 1, \dots, V$ and simulate their permutation distribution as $T^{b*}_{\psi v} = \psi(\hat{\lambda}^{b*}_{1v}, \dots, \hat{\lambda}^{b*}_{Cv})$, $b = 1, \dots, B$, $v = 1, \dots, V$.
3. Calculate the combined p-values as $\hat{\lambda}^{\psi}_v = \sum_b \mathbb{I}(T^{b*}_{\psi} \geq T^o_{\psi})/B$, $v = 1, \dots, V$ and simulate their distribution as $\hat{\lambda}^{r\psi}_v = \sum_b \mathbb{I}(T^{b*}_{\psi} \geq T^{r*}_{\psi})/B$, $r = 1, \dots, B$, $v = 1, \dots, V$.
4. Achieve the third-order test as $T^o_\theta = \theta(\hat{\lambda}^{\psi}_1, \dots, \hat{\lambda}^{\psi}_V)$ and simulate its permutation distribution as $T^{b*}_\theta = \theta(\hat{\lambda}^{b\psi}_1, \dots, \hat{\lambda}^{b\psi}_V)$, $b = 1, \dots, B$.
5. Compute the global p-value as $\hat{\lambda}^{\theta} = \sum_b \mathbb{I}(T^{b*}_\theta \geq T^o_\theta)/B$.

When applying such a procedure, the user needs to make three fundamental choices of:

- the test statistics $T_v(\cdot),\ v = 1, \ldots, V$
- the first combining function $\psi(\cdot)$
- the second combining function $\theta(\cdot)$

With regard to the test statistic, a popular solution is the modified Anderson-Darling test statistic proposed in Arboretti et al. (2021) and Pesarin and Salmaso (2010). It is important to highlight that the combining functions must have the following properties:

- be non-increasing in each argument
- attain their supremum when at least one of their arguments attains 0
- provide a critical value which is finite and strictly smaller than the aforementioned supremum value
- have convex rejection regions (Birnbaum, 1954; Birnbaum et al., 1955)

Widely investigated combining functions satisfying all the previous conditions are in particular Fisher's $-2 \cdot \sum_v \log(\lambda_v)$ and Tippett's $\max_{1 \leq v \leq V}(1 - \lambda_v)$ where $\lambda_v,\ v = 1, \ldots, V$ is the vector of p-values to be combined.

3 Ranking of multiple populations

In this section we focus on a scenario in which we are interested in ranking C different multivariate populations with respect to V marginal variables when C samples $\mathbf{X}_1, .., \mathbf{X}_C$ are drawn from C populations and no preliminary knowledge on underlying distribution $\mathbf{F}$ is available. In other words, we want to estimate the relative ordering of each population when compared among all other populations which is defined as $r_j = r(\mathbf{X}) = 1 + \sum_{k \neq j} \mathbb{I}(\mathbf{X}_j \overset{d}{<} \mathbf{X}_k),\ j = 1, \ldots, C$ where $\mathbb{I}(\cdot)$ is 1 if condition $\cdot$ is satisfied and 0 otherwise. This definition (Gupta and Panchapakesan, 2002) of rank recalls the concept of stochastic dominance and is based on the idea of pairwise counting how many populations are stochastically larger than that a specific population. An alternative definition proposed in Corain et al. (2017) is $r_j = 1 + \#\{(C - \sum_{k \neq j} \mathbb{I}(\mathbf{X}_j \overset{d}{>} \mathbf{X}_k)) > (C - \sum_{k \neq j'} \mathbb{I}(\mathbf{X}_{j'} \overset{d}{>} \mathbf{X}_k)),\ j' = 1, \ldots, C,\ j \neq j'\},\ j = 1, \ldots, C$, where $\#\{\cdot\}$ counts the number of times $\cdot$ is true. In this case we simply pairwise count how many populations are stochastically smaller than a given population. The same ranking is achieved independently of the chosen definition of rank, but the first one involves moving from from the best to the worst population in a downward fashion while the second one from worst to best in an upward fashion. To distinguish the two different definitions, we are now going to use r_j^D for the first one and r_j^U for the second one.

Let us consider a dummy example. Let $\mathbf{\Gamma}$ be the matrix in which each cell Γ_{jk} contains the result of $\mathbb{I}(\mathbf{X}_j \overset{d}{>} \mathbf{X}_k)$ and be:

$$\begin{bmatrix} \bullet & 1 & 1 \\ 0 & \bullet & 0 \\ 0 & 0 & \bullet \end{bmatrix}.$$

It is immediate to see that r_j^D are essentially found as 1 plus the column-wise sums of $\mathbf{\Gamma}$, so that $r_1^D = 1+0+0$, $r_2^D = 1+1+0$, and $r_3^D = 1+1+0$ (bullets are ignored). The downward ranking is therefore $(1,2,2)$. On the other hand, according to Corain et al. (2017) r_j^U can be estimated by subtracting the row-wise sums of $\mathbf{\Gamma}$ to C, so that $r_1^D = 3-1-1$, $r_2^D = 3-0-0$, and $r_3^D = 3-0-0$, and the upward ranking is again $(1,2,2)$.

Multivariate directional pairwise comparisons need to be performed to estimate population ranks and in particular to assess $\mathbb{I}(\mathbf{X}_j \overset{d}{<} \mathbf{X}_k), j \neq k, j,k = 1\ldots,C$. It is worth noting that in this case we do not have any a priori knowledge about the real ordering among the C multivariate populations, differently from the typical stochastic ordering problem.

For each pair $(\mathbf{X}_j, \mathbf{X}_k)$, we need therefore to evaluate the following system of hypotheses:

$$\begin{cases} H_0^{(j,k)} : \mathbf{X}_j \overset{d}{=} \mathbf{X}_k \\ H_1^{(j,k)} : \mathbf{X}_j \overset{d}{>} \mathbf{X}_k \end{cases} \quad j,k = 1,\ldots,C,\ j \neq k \tag{4}$$

To do that, the NPC methodology could again represent a useful solution. We start by decomposing system of hypotheses 4 as follows:

$$\begin{cases} H_{0v}^{(j,k)} : \bigcap_{v=1}^{V}(X_{jv} \overset{d}{=} X_{kv}) \\ H_{1v}^{(j,k)} : \bigcup_{v=1}^{V}(X_{jv} \overset{d}{>} X_{kv}) \end{cases} \quad j,k = 1,\ldots,C,\ j \neq k,\ v = 1,\ldots,V \tag{5}$$

For each pair $(\mathbf{X}_j, \mathbf{X}_k)$, we then apply the following algorithm:

1. Apply the test statistic to the pooled sample $\mathbf{Z} = \{\mathbf{X}_j, \mathbf{X}_k\}$ and achieve the vector $\mathbf{T}^o = \mathbf{T}(\mathbf{Z})$.

2. For $b = 1,\ldots,B$, retrieve $\mathbf{T}^{b*} = \mathbf{T}(\mathbf{Z}^*)$, where $\mathbf{Z}^*$ is random permutation of $\mathbf{Z}$, so that a random sampling from the permutation V-variate distribution of $\mathbf{T}$ is obtained.

3. Compute $\hat{\lambda}_v = \frac{\left[\frac{1}{2}+\sum_b \mathbb{I}(T_v^{b*} \geq T_v^o)\right]}{(B+1)}$, $v = 1,\ldots,V$ and simulate their permutation distribution $\hat{\lambda}_v^{r*} = \frac{\left[\frac{1}{2}+\sum_b \mathbb{I}(T_v^{b*} \geq T_v^{r*})\right]}{(B+1)}$, $v = 1,\ldots,V,\ r = 1,\ldots,B$.

4. Combine p-values to achieve a second-order test $T_\psi^o = \psi(\hat{\lambda}_1,\ldots,\hat{\lambda}_V)$ and simulate its permutation distribution as $T_\psi^{b*} = \psi(\hat{\lambda}_1^{b*},\ldots,\hat{\lambda}_V^{b*}),\ b = 1,\ldots,B$.

5. Achieve the global p-value as $\hat{\lambda}^{(j,k)} = \sum_b \mathbb{I}(T_{\psi}^{b*} \geq T_{\psi}^{o})/B$.

Again, key aspects are the choices of appropriate test statistics and combining function.

A matrix of p-values $\mathbf{\Lambda}$ can therefore be achieved:

$$\mathbf{\Lambda} = \begin{bmatrix} \bullet & \hat{\lambda}^{(1,2)} & \dots & \hat{\lambda}^{(1,C-1)} & \hat{\lambda}^{(1,C)} \\ \hat{\lambda}^{(2,1)} & \bullet & \dots & \hat{\lambda}^{(2,C-1)} & \hat{\lambda}^{(2,C)} \\ \dots & \dots & \dots & \dots & \dots \\ \hat{\lambda}^{(C-1,1)} & \hat{\lambda}^{(C-1,2)} & \dots & \bullet & \hat{\lambda}^{(C-1,C)} \\ \hat{\lambda}^{(C,1)} & \hat{\lambda}^{(C,2)} & \dots & \hat{\lambda}^{(C,C-1)} & \bullet \end{bmatrix}$$

where $\hat{\lambda}^{(j,k)}$ refers to the alternative hypothesis $\mathbf{X}_j \overset{d}{>} \mathbf{X}_k$

Given that $C \times (C-1)$ one-sided pairwise comparisons are performed via p-value statistics, we need to properly control the global type I error. The multiplicity issue indeed occurs in case of simultaneous testing when multiple statistical tests are jointly considered. In particular, we need to apply appropriate techniques to control the family-wise error rate, i.e. the probability of committing at least one type I error (Shaffer, 1995; Pesarin and Salmaso, 2010). To this aim several methods have been proposed in the literature, which are commonly grouped into single step (such as Bonferroni's) and step-down procedures (such as Holm's) (Ge et al., 2003).

After applying one of these procedures, a matrix of adjusted p-values $\mathbf{\Lambda}_{adj}$ can be achieved. Let α be the chosen significance level, then a matrix of significances $\mathbf{\Omega}$ can be computed:

$$\mathbf{\Omega} = \begin{bmatrix} \bullet & \omega^{(1,2)} & \dots & \omega^{(1,C-1)} & \omega^{(1,C)} \\ \omega^{(2,1)} & \bullet & \dots & \omega^{(2,C-1)} & \omega^{(2,C)} \\ \dots & \dots & \dots & \dots & \dots \\ \omega^{(C-1,1)} & \omega^{(C-1,2)} & \dots & \bullet & \omega^{(C-1,C)} \\ \omega^{(C,1)} & \omega^{(C,2)} & \dots & \omega^{(C,C-1)} & \bullet \end{bmatrix}$$

where $\omega^{(j,k)} = \mathbb{I}(\hat{\lambda}^{(j,k)} \leq \frac{\alpha}{2})$. This matrix synthesize results from all multivariate directional pairwise comparisons and can be used to estimate rankings.

The downward ranks r_j^D can indeed be estimated by $\hat{r}_j^D = 1 + \sum_{k=1}^{C} \omega^{(k,j)},\ j = 1, \dots, C$ and the upward ranks r_j^U by $\hat{r}_j^U = 1 + \#\{(C - \sum_{k=1}^{C} \omega^{(j,k)}) > (C - \sum_{k=1}^{C} \omega^{(j',k)}),\ j' = 1, \dots, C,\ j \neq j'\},\ j = 1, \dots, C$. These estimates are achieved by counting how many times the jth population is significantly stochastically smaller (for downward ranks) or larger (for upward ranks) than the other populations at the chosen significance level.

Although $\hat{r}_j^D$ and $\hat{r}_j^U$ lead to the same ranking under many scenarios, sometimes the so called intransitivity issue occurs (Corain et al., 2017; Arboretti et al., 2014), with possible inconsistencies arising from pairwise results which may pose serious problems in the ranking estimation. This results in $\hat{r}_j^D$ and $\hat{r}_j^U$ not matching and to overcome this issue Corain et al. (2017) suggests a revised ranking estimator $\bar{r}_j = 1 + \#\{(\hat{r}_j^D + \hat{r}_j^U) > (\hat{r}_k^D + \hat{r}_k^U),\ k = 1, \dots, C,\ j \neq k\},\ j = 1, \dots, C$, which essentially is an average rank from the ranks derived from the countings of the significant observed stochastic inferiorities and superiorities.

$\bar{r}_j$ can therefore be used to solve the problem of ranking C different multivariate populations addressed in this section.

4 Explanatory case studies

A couple of case studies are considered in order to show potential applications of the aforementioned techniques through real-world ranking problems.

4.1 Case 1

The first case refers to the analysis of customer satisfaction data. It is well known that packaging plays a fundamental role in protecting and preserving the quality of the food or beverage stored within it. Packages used to store wine can be made of different materials such as glass, plastic, aluminum, and plastic covered paper. In this study we wanted to evaluate customer satisfaction towards four types of packaging (i.e. Prod1, Prod2, Prod3, Prod4), using a specific type of wine.

Several individuals were asked to answer an online questionnaire, with 517 of them accepting this task. A selection-biased sample was therefore collected and the related inferences are to be taken with some cautions. They were thereafter asked to rate a product (randomly chosen among the four possible ones) from 1 (= bad) to 5 (= extremely good) in terms of several key aspects. After receiving the product, the participant tested and evaluated it and nine different key performance indicators (KPIs) were considered, including purchase intention, overall quality, and specific evaluations of the color, the flavor, and the taste of the wine. The company was then interested in ordering the four types of packaging according to their global performances and this was non other than a multivariate ranking problem.

To answer to this request, we decided to rely on the aforementioned NPC-based approach. We decided to adopt the modified Anderson-Darling test statistic to perform each pair-wise comparison $X_{jv} \overset{d}{>} X_{kv},\ j,k = 1,\dots,C,\ j \neq k,\ v = \dots, V$:

$$T_v^{(j,k)} = \sum\nolimits_{i=1}^{N} [\hat{F}_{kv}(X_{ikv}) - \hat{F}_{jv}(X_{ijv})]/\{\bar{F}_v(Z_{iv})[1-\bar{F}_v(Z_{iv})]\}^{\frac{1}{2}}$$

where $Z_v = \{X_{jv}, X_{kv}\}$ is the pooled sample of size $N = n_j + n_k$, n_j and n_k are sample sizes, $\hat{F}_{jv}(t) = \sum_{i=1}^{n_j} \mathbb{I}(X_{ijv} \leq t)/n_j$, $\hat{F}_{kv}(t) = \sum_{i=1}^{n_k} \mathbb{I}(X_{ikv} \leq t)/n_k$, $\bar{F}_v(t) = \sum_{i=1}^{N} \mathbb{I}(Z_{iv} \leq t)/N$, and $t \in \mathcal{R}^1$. Fisher's combining function was used, the Bonferroni-Holm-Shaffer method (Shaffer, 1986) was adopted for multiplicity adjustment and the number of permutations B was set to 2000.

The descriptive analysis showed that Prod1 appears to be substantially outperformed by the others (see Figure 1). Additionally, Prod2 appears to have slightly better performances than the remaining ones.

The adoption of the NPC-based ranking procedure provided us with the p-values reported in Table 1 and the adoption of the Bonferroni-Holm-Shaffer adjustment produced the adjusted p-values in Table 2. We can see that Prod2 significantly outperforms all the other products, while Prod3 and Prod4 perform better than Prod1. Estimating rankings according to the aforementioned three different approaches (see Table 3), we can see that they all agree and assign the first position to Prod2, the second position to Prod3 and Prod4, and the last position to Prod1. In other words, the adoption of the multivariate ranking procedure introduced in Section 3 allowed us to order the four types of packaging according to their overall performances from best to worst. The company was therefore able to choose the best innovative packaging to be launched into the marked.

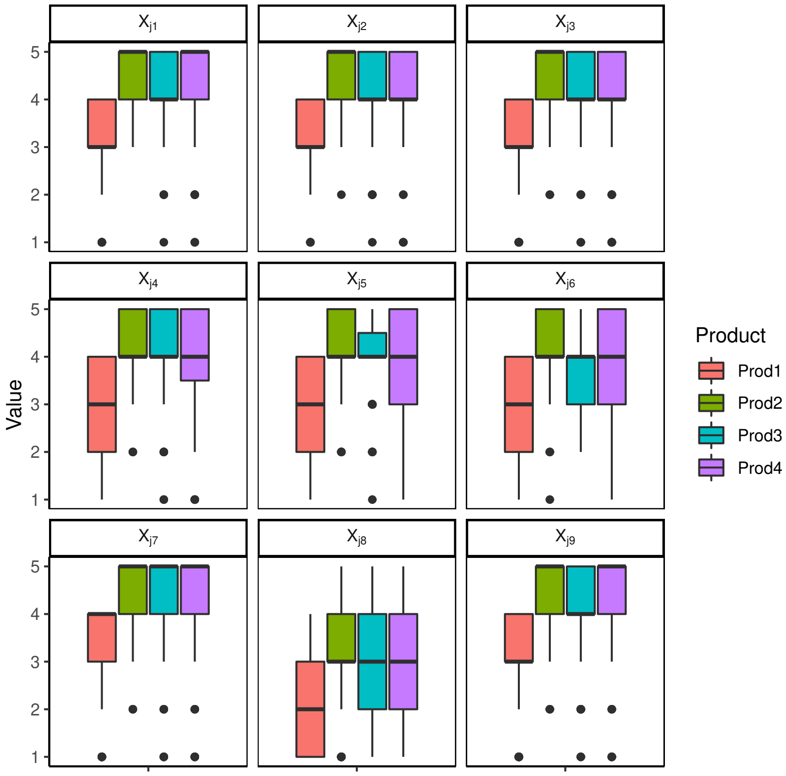

Figure 1: Box-plots: case 1.

Supposing then that the order $\mathbf{X}_{Prod1} \overset{d}{\leq} \mathbf{X}_{Prod3} \overset{d}{\leq} \mathbf{X}_{Prod4} \overset{d}{\leq} \mathbf{X}_{Prod2}$ (with at least one strict inequality) was known a-priori, we applied the NPC-based procedure for stochastic ordering problems.

To this purpose, we used again the aforementioned Anderson-Darling test statistic and a number of permutation equal to 2000. The achieved p-value was smaller than 0.001 and therefore much smaller than the chosen significance level 0.05. Evidence was found in favor of the proposed stochastic order.

It is worth noting that in this case we know that at least a strict inequality (i.e. $\mathbf{X}_i \overset{d}{<} \mathbf{X}_j$) holds, but we cannot exactly state which of the inequalities are strict. The

Table 1: P-values table: case 1.

	Prod1	Prod2	Prod3	Prod4
Prod1	$\bullet$	1.000	1.000	1.000
Prod2	< 0.001	$\bullet$	0.001	0.005
Prod3	< 0.001	0.996	$\bullet$	0.457
Prod4	< 0.001	0.996	0.588	$\bullet$

Table 2: Adjusted p-values table: case 1.

	Prod1	Prod2	Prod3	Prod4
Prod1	$\bullet$	1.000	1.000	1.000
Prod2	0.003	$\bullet$	0.006	0.015
Prod3	0.001	0.996	$\bullet$	1.000
Prod4	0.001	1.000	1.000	$\bullet$

Table 3: Estimated ranking: case 1.

	Prod1	Prod2	Prod3	Prod4
$\hat{r}_j^U$	4	1	2	2
$\hat{r}_j^D$	4	1	2	2
$\bar{r}_j$	4	1	2	2

ranking procedure by Arboretti et al. (2014) on the other hand allowed us to describe the existing relationships between samples in a more detailed way, showing which types of packages were perceived in a similar way (i.e. Prod3 and Prod4) and which not.

4.2 Case 2

The second case study refers to the analysis of laboratory data. Four different formulations of a pharmaceutical product were evaluated through multiple laboratory experiments in terms of six different KPIs: three measures of its effectiveness on a scale from 1 (= totally ineffective) to 10 (= completely effective) and three measures of the severity of three possible side effects on a scale from 10 (= low) to 1 (= critical) were considered. The final data set contained twenty observations per formulation. The four formulations were characterized by an increasing amount of a specific active ingredient which was expected to have a positive impact on the global performances of the product, i.e. $\mathbf{X}_1 \overset{d}{\leq} \mathbf{X}_2 \overset{d}{\leq} \mathbf{X}_3 \overset{d}{\leq} \mathbf{X}_4$ (with at least a strict inequality) was expected. In other words, we had a typical stochastic ordering problem and we needed to confirm the expectations of the company (i.e. we had a-priori knowledge about the ordering).

To test the hypothesis of interest, we decided to rely on the aforementioned NPC-based procedure for multivariate stochastic ordering problems. Again, a modified Anderson-Darling test statistic was adopted, so that for each of the $C-1$ iterations in the first step of the algorithm presented in Section 2 we used:

$$T_v^c = \sum_{i=1}^{N+M} [\hat{F}_{1v}^c(Z_{i1v}^c) - \hat{F}_{2v}^c(Z_{i2v}^c)]/\{\bar{F}_v^c(Z_{iv}^c)[1 - \bar{F}_v^c(Z_{iv}^c)]\}^{\frac{1}{2}},$$

where $c = 2, \ldots, C$, $Z_v^c = \{Z_{1v}^c, Z_{2v}^c\}$ is the pooled sample of size $N+M$, N and M are sample sizes, $\hat{F}_{1v}^c(t) = \sum_{i=1}^{N} \mathbb{I}(Z_{i1v}^c \leq t)/N$, $\hat{F}_{2v}^c(t) = \sum_{i=1}^{M} \mathbb{I}(Z_{i2v}^c \leq t)/M$, $\bar{F}_v^c(t) = \sum_{i=1}^{N+M} \mathbb{I}(Z_{iv}^c \leq t)/N+M$, and $t \in \mathcal{R}^1$. Fisher's combining function was used as $\theta(\cdot)$ and Tippett's as $\psi(\cdot)$. The number of permutations B was set to 2000. In this case we decided to report also the partial p-values (i.e. indications about the stochastic ordering are provided for each single KPI) and for this reason we had to address the multiplicity issue by applying the Min$P$ Bonferroni-Holms adjustment described in Pesarin and Salmaso (2010).

Figure 2 appears to confirm the expected order, with Form4 (i.e. the one with the highest amount of the active ingredient) outperforming the others. The only exception is X_{j5} for which an odd behavior can be seen.

The application of the NPC-based technique provided the partial (adjusted) and global p-values reported in Table 4. It is possible to see that the only non-significant p-value is $\hat{\lambda}_{5,adj}^{\psi}$, confirming the different order observed for X_{j5}, $j = 1, \ldots, 5$. Overall, it appears that data support the expected $\mathbf{X}_1 \overset{d}{\leq} \mathbf{X}_2 \overset{d}{\leq} \mathbf{X}_3 \overset{d}{\leq} \mathbf{X}_4$ (with at least a strict inequality) and the positive impact of the active ingredient of interest.

Supposing then that we had no information about the expected order, we applied the NPC-based multivariate ranking procedure. To this purpose, we used again the aforementioned Anderson-Darling test statistic and a number of permutation equal to

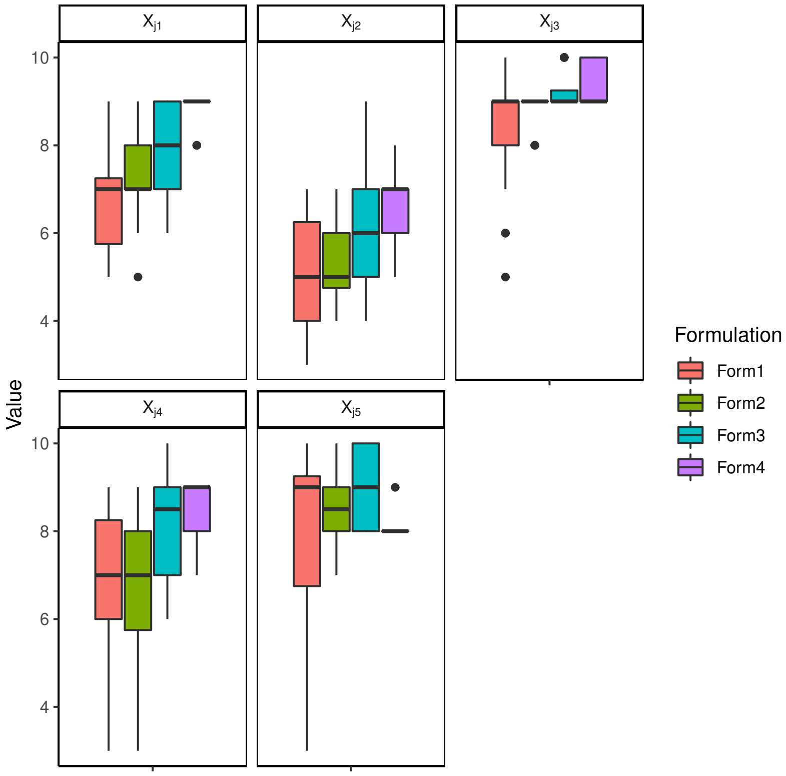

Figure 2: Box-plots: case 2.

Table 4: P-values table: case 2 - stochastic ordering.

$\hat{\lambda}^{\psi}_{1,adj}$	$\hat{\lambda}^{\psi}_{2,adj}$	$\hat{\lambda}^{\psi}_{3,adj}$	$\hat{\lambda}^{\psi}_{4,adj}$	$\hat{\lambda}^{\psi}_{5,adj}$	$\hat{\lambda}^{\theta}$
< 0.001	< 0.001	< 0.001	< 0.001	0.241	< 0.001

Table 5: Adjusted p-values table: case 2 - multiple comparisons.

	Form1	Form2	Form3	Form4
Form1	•	1.000	0.998	1.000
Form2	0.154	•	1.000	1.000
Form3	0.003	0.015	•	0.831
Form4	0.001	0.001	0.099	•

Table 6: Estimated ranking: case 2.

	Form1	Form2	Form3	Form4
$\hat{r}_j^U$	3	3	1	1
$\hat{r}_j^D$	3	3	1	1
$\bar{r}_j$	3	3	1	1

2000. Bonferroni-Holm-Shaffer adjustment was adopted to deal with the multiplicity issue.

Adjusted p-values related to the performed multiple comparisons are reported in Table 5. The results of the estimation of rankings according to the presented three different approaches are displayed in Table 6. It appears that the pair of formulations with the highest concentration of the active ingredient (i.e. Form3 and Form4) significantly outperformed the remaining formulations, while no significant differences were detected between Form1 and Form2 and between Form3 and Form4.

Again, the NPC-based solution to the multivariate stochastic ordering problem allowed us to confirm the supposed order, but the ranking procedure allowed us to better appreciate pair-wise differences between formulations.

5 Conclusions

Multivariate ranking refers to problems where the main interest is to assess the existing order among C different items in terms of multiple features. This is a quite common task which is relevant in a wide variety of fields. Multivariate ranking however is a generic term which includes many different scenarios and in this paper we decided to consider two of them in which populations need to be ranked, highlighting differences and similarities.

Under the first scenario we suppose to have preliminary knowledge about the ranking of the populations and to be interested in testing this specific stochastic order. This is called a stochastic ordering problem and a solution based on the NPC methodology (Pesarin and Salmaso, 2010) was proposed.

The second considered problem was characterized by the absence of a-priori information, so that the conduction of multiple pair-wise comparisons was required. To solve this challenge, we suggested the adoption of the NPC-based solution proposed by Arboretti et al. (2014) and widely described in Corain et al. (2017).

The differences between these two multivariate ranking problems were then highlighted by means of two different case studies. The analysis of real-world data allowed us to better illustrate the proposed procedures and show how the use of the NPC methodology should be considered to deal with such multivariate problems.

Acknowledgment

In memory of Prof. Livio Corain for his great contribution to the study of multivariate ranking problems.

References

Arboretti, R., Bonnini, S., Corain, L., and Salmaso, L. (2014). A permutation approach for ranking of multivariate populations. *Journal of Multivariate Analysis*, 132:39–57.

Arboretti, R., Ceccato, R., and Salmaso, L. (2021). Permutation testing for goodness-of-fit and stochastic ordering with multivariate mixed variables. *Journal of Statistical Computation and Simulation*, 91(5):876–896.

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

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

Chenouri, S., Mozaffari, A., and Rice, G. (2020). Robust multivariate change point analysis based on data depth. *Canadian Journal of Statistics*, 48(3):417–446. cited By 1.

Chenouri, S. and Small, C. (2012). A nonparametric multivariate multisample test based on data depth. *Electronic Journal of Statistics*, 6:760–782. cited By 20.

Corain, L., Arboretti, R., and Bonnini, S. (2017). *Ranking of multivariate populations: A permutation approach with applications*. CRC press.

Corain, L., Melas, V. B., Pepelyshev, A., and Salmaso, L. (2014). New insights on permutation approach for hypothesis testing on functional data. *Advances in Data Analysis and Classification*, 8(3):339–356.

Corain, L. and Salmaso, L. (2013). Nonparametric permutation and combination-based multivariate control charts with applications in microelectronics. *Applied Stochastic Models in Business and Industry*, 29(4):334–349.

Cuzick, J. (1985). A wilcoxon-type test for trend. *Statistics in medicine*, 4(1):87–90.

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

Gower, J. (1987). *Introduction to ordination techniques.* cited By 70.

Gupta, S. S. and Panchapakesan, S. (2002). *Multiple decision procedures: theory and methodology of selecting and ranking populations*. SIAM.

Jonckheere, A. R. (1954). A distribution-free k-sample test against ordered alternatives. *Biometrika*, 41(1/2):133–145.

Li, J. and Liu, R. (2004). New nonparametric tests of multivariate locations and scales using data depth. *Statistical Science*, 19(4):686–696. cited By 45.

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

Scarsini, M. and Shaked, M. (1990). Stochastic ordering for permutation symmetric distributions. *Statistics & probability letters*, 9(3):217–222.

Shaffer, J. P. (1986). Modified sequentially rejective multiple test procedures. *Journal of the American Statistical Association*, 81(395):826–831.

Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual review of psychology*, 46(1):561–584.

Syms, C. (2008). Ordination. In Jørgensen, S. E. and Fath, B. D., editors, *Encyclopedia of Ecology*, pages 2572–2581. Academic Press, Oxford.

Terpstra, T. J. (1952). The asymptotic normality and consistency of kendall's test against trend, when ties are present in one ranking. *Indagationes Mathematicae*, 14(3):327–333.