Università degli Studi di Ferrara

DIPARTIMENTO DI ECONOMIA, ISTITUZIONI, TERRITORIO

Corso Ercole I D'Este n.44, 44100 Ferrara

Quaderni del Dipartimento

n.7/2003

May 2003

NPC of Dependent Rankings: applications to multivariate problems in management and industry

by Rosa Arboretti Giancristofaro, Simona Boari, Livio Corain

May 2003

NPC of Dependent Rankings: applications to multivariate problems in management and industry

Rosa Arboretti Giancristofaro[*]

Simona Boari[§]

Livio Corain[^]

Abstract: in this paper we propose some extensions and applications of the nonparametric combination of dependent rankings (Pesarin and Lago, 2000). This methodology is applied to Conjoint Analysis in order to aggregate preferences from a group of individuals. Furthermore, the PPT (Pesarin and Salmaso, 2002) for the inferential analysis is applied in a multifactorial ANOVA.

Keywords: ranking, nonparametric combination, multivariate.

[*] Data Medica Spa - Padova; via Zanchi, 89 - 35100 Padova; rosa.arboretti@datamedica.it
[§] Facoltà di Economia - Università di Ferrara; via del Gregorio, 13 - 44100 Ferrara; s.boari@economia.unife.it
[^] Dipartimento di Tecnica e Gestione dei sistemi industriali - Università di Padova; stradella S. Nicola, 3 - 36100 Vicenza; livio.corain@gest.unipd.it

1 Introduction

In many real situations we encounter the need to compare entities of a different nature (products, services, companies, behaviour and so on) in order to obtain a ranking among the considered statistical units. If the comparison is based on only one feature the result is obtained in a trivial way but difficulties may arise when we are dealing with two or more informative variables jointly. We can build up as many rankings as the number of features we are dealing with. Apart from the case where units occupy the same position in every ranking, the need to summarise a set of rankings into one single global ranking arises. The main purpose of the NPC ranking method (Pesarin and Lago, 2000) is to obtain a single ranking criterion for the statistical units being studied, which summarises the many initial partial (univariate) criteria. This method is defined as nonparametric since it needs neither the knowledge of the underlying statistical distribution for the variables being studied, nor the dependence structure among variables, apart from the assumption that all dependences are monotonic regressions.

2 NonParametric Combination (NPC) of Dependent Rankings

Given a multivariate phenomenon $\mathbf{X}=[X_1, X_2,..., X_k]$, observed on N statistical units, and once we have calculated the *k partial rankings* $R_1, R_2,..., R_k$, starting from the variables X_i, $i=1,\ldots,k$, each one being informative about a partial aspect of phenomenon $\mathbf{X}$, we want to build up a *global combined ranking* Y:

$$Y = \psi\,(X_1, X_2,..., X_k; \omega_1, \omega_2,..., \omega_k), \psi : \Re^{2k} \rightarrow \Re^1$$

where ψ is a real function allowing us to combine the partial dependent rankings and where $\omega_1, \omega_2,\ldots, \omega_k$ is a set of weights, defined on the basis of technological, functional or economic considerations, which measure the relative degree of importance among the k aspects of $\mathbf{X}$.

In order to build up Y we introduce a set of minimal reasonable conditions related to the variables X_i $i=1,\ldots,k$:

1. for each of the k informative variables a partial ordering criterion is well established, in the sense that "large is better"; if it is not so, it is possible to recode the variables by means of any appropriate transformation φ :
   a. if "large is worse" $\Rightarrow \varphi\,(X)=1/X$ or $\varphi\,(X)= -X$;
   b. if "δ is better" (central target value) $\Rightarrow \varphi\,(X)=|X-\delta|$;
2. regression relationships within the k informative variables are monotonic (increasing or decreasing)
3. the marginal distribution of each informative variable is non-degenerate.

Moreover, we need not make any further assumptions either on the statistical distribution of the informative variables, or on their dependence structure. Finally, note that we do not need to assume the continuity of X_i $i=1,\ldots,k$, so that the probability of ex-equo can be different from zero.

Let us define the set of variables X_i as $\{Z_{ji}\ ,\ i=1,\ldots,k,\ j=1,\ldots,N\}$, possibly after proper transformations. Without loss of generality, they are assumed to behave in accordance with the rule "large is better". In this setting, we consider the rank transformations R_{ji} (*partial rankings*):

$$\{\mathrm{R}_{ji}= \mathrm{R}(Z_{ji}) = \#\,(Z_{ji} \geq Z_{hi}),\ i=1,\ldots,k,\ j,h=1,\ldots,N\}.$$

Associated with these ranks are the scores:

$$\left\{ \lambda_{ji} = \frac{R_{ji} + 0.5}{N+1}, i = 1,...,k \;\; j = 1,...,N \right\}.$$

Once a combining function ψ (for details of combining functions see below) has been chosen, we compute the transformation

$$\psi : \{ Y_j = \psi(\lambda_{j1},\ldots, \lambda_{jk}; \omega_1,\ldots, \omega_k), j=1,\ldots,N\},$$

finally, applying the rank transformation, we obtain the *global combined ranking Y*:

$$\{ Y_j = \mathrm{R}(Y_j) = \# (Y_j \geq Y_h),\ j,h=1,\ldots,N\}.$$

In the global ranking Y, each statistical units is ranked in a unique way, by taking into consideration the whole set of the k informative variables.
It seems reasonable that each combining real function ψ should satisfy at least the following minimal properties:

a) it must be continuous in all $2k$ arguments, in that small variations in any subset of arguments imply small variations in the ψ-index;
b) it must be non-increasing in each argument: $\psi(...,\lambda_i,...;\omega_1,..., \omega_k) \geq \psi(...,\lambda'_i,...; \omega_1,..., \omega_k)$ if $1 > \lambda_i > \lambda'_i > 0$, for whatever $i \in \{1,\ldots,k\}$;
c) it must be symmetric with respect to permutations of the arguments, in that if for instance $u_1,...,u_k$ is any permutation of $1,...,k$, then $\psi(\lambda_{u1},..., \lambda_{uk};\omega_{u1},..., \omega_{uk}) = \psi(\lambda_1,..., \lambda_k;\omega_1,..., \omega_k)$.

The above properties define the class Ψ of combining functions. Some of the functions most often used to combine independent tests (Fisher, Lancaster, Liptak, Tippett, Mahalanobis, etc.) are included in this class. The Fisher combining function has the form $Y = \sum_{i=1}^{k} \omega_i \cdot \log(1-\lambda_i)$.
It is worth noting that combining functions in class Ψ are nonparametric with respect to the underlying dependence structure among informative variables, in that all kinds of monotonic dependences are implicitly captured by these functions. In fact, no explicit model for this dependence structure is needed and no dependent coefficient need be estimated directly from the data.

3 Applications to multivariate problems in management and industry

The methodology of the Nonparametric Combination of Dependent Rankings can be naturally applied to various company and industrial contexts where the need to establish an order of a set of elements (e.g. products or services) proves to be strategic for the success of the company.
In this work we deal with a problem linked to the development of a new consumer product, more specifically of a moveable, modular, assembled wall for office use. In general, during the initial stages of development of a new product, managing to establish the influence of the product's characteristics in relation to the preferences of potential consumers is crucial. To deal with this problem, a methodology called Conjoint Analysis (CA) has been developed in business-statistical literature. The Conjoint Analysis identifies a set of predominantly statistical methodologies aimed at studying consumer choice models using opinions of preference expressed by consumers about various profiles of a product/service (Gustafsson et al., 2001).

In the case study in question three basic characteristics of the new product have been identified - flexibility, versatility and type of software. Two levels were associated to each of these three factors, thus generating a 2^3 experimental design with a total of eight product profiles (full profile design). They were subjected to the opinions of a group of nine individuals who gave each profile a score on a scale of 0-100. During the data gathering stage, to facilitate the task of respondents, they were given the possibility of assigning the same score to various profiles.

Traditionally, at this stage of the CA, the next step is to estimate the parameters associated to the experimental design, using OLS estimators, starting with each vector of preferences given by each respondent. To then obtain an overall result an individual parameters' mean is calculated, thus determining the presumed optimal product profile for the entire group of respondents. It is clear, however, that this solution is rather unsatisfactory and even misleading in some cases since it fully suffers the application limitations of the sample mean.

As a proposed solution to this problem, in this work we present a multi-stage procedure which substantially uses the Nonparametric Combination Dependent Rankings. This procedure is made up of the following stages:

1. considering the product profiles as statistical units needing ordered and the preferences of each respondent as partial informative variables, the NPC Ranking methodology is applied, obtaining a global combined ranking of the product profiles;
2. on the basis of the global ranking, the multifactor ANOVA for unreplicated levels is applied, considering both the main effects and the interactions of any order;
3. a permutation *p*-value, obtained using the PPT (Paired Permutation Test, Pesarin and Salmaso, 2002), is associated to each factor.

In this way it is possible to unequivocally establish the preference order of the product profiles for a group of individuals, and which factors significantly influence the preferences of the those interviewed. A key point in the procedure is without doubt the accounting of the preference opinions carried out a priori on the basis of the NPC of Dependent Rankings.

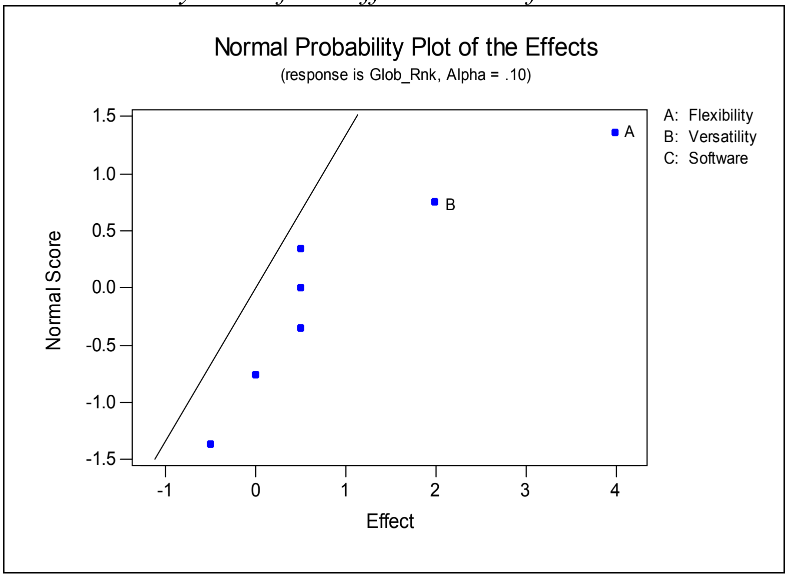
The following table illustrates the gathered data and the obtained results.

Table 1: *Expressed preferences, partial rankings and global ranking per product profile.*

Profile	Global Ranking	Expressed preferences (S) and partial rankings (pR) per individual																	
		S_1		S_2		S_3		S_4		S_5		S_6		S_7		S_8		S_9	
		S	pR	S	pR	S	pR	S	pR	S	pR	S	pR	S	pR	S	pR	S	pR
P_1	8	20	8	10	8	50	6	40	5	20	8	10	7	30	7	30	8	30	5
P_2	6	50	5	40	5	20	8	30	6	30	7	10	7	40	6	50	5	20	7
P_3	7	30	7	20	7	30	7	10	8	40	6	30	6	20	8	40	6	30	5
P_4	5	40	6	30	6	60	5	20	7	50	5	40	4	50	5	40	6	20	7
P_5	3	70	3	50	4	65	4	60	3	60	4	50	2	70	3	75	2	80	1
P_6	2	80	2	90	1	95	1	80	1	90	2	50	2	80	2	70	3	70	3
P_7	4	60	4	70	2	85	2	50	4	80	3	40	4	60	4	70	3	75	2
P_8	1	90	1	70	2	75	3	70	2	100	1	70	1	90	1	80	1	70	3

The favourite product is therefore shown to be profile number 8 and the least favourite number 1. With regard to the multifactor ANOVA carried out on the global ranking, the following normal plot is obtained for the effects.

Graph 1: *Normal Probability Plot of the Effects – multifactor ANOVA.*



It can be deduced that the most important factor is flexibility, followed by versatility. The software factor, as with all interactions, does not appear to be important.
In conclusion we present the *p*-values of PPT which confirm the significance of the flexibility factor (p=0.048) and of the versatility factor (p=0.054) at a significance level of 0.063, while for the other effects, the PPT does not reject the null hypothesis.

4 Conclusions

This work highlights how the Nonparametric Combination of Dependent Rankings method can be efficiently applied to, for instance, the problem of studying relative opinions of preference about various profiles of a new product, plus a new CA procedure based on aggregated preferences has been proposed.

References

Folks J.L. (1984) Combinations of independent tests. In P.R. Krishnaiah and P.K. Sen (eds.), *Handbook of Statistics*, 4, 113-121, North-Holland, Amsterdam.

Gustafsson A., Herrmann, A., Huber, F. (2001), *Conjoint measurement: methods and applications*, Berlin: Springer.

Joe H. (1997) *Multivariate Models and Dependence Concepts*, Chapman & Hall, London.

Pesarin F., Lago, A. (2000) Nonparametric combination of dependent rankings with application to the quality assessment of industrial products, *Metron*, LVIII, 1-2.

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

Pesarin F., Salmaso, L. (2002) Exact permutation tests for unreplicated factorials, *Appl. Stochastic Models Bus. Ind.*, 18: 287 – 299.