Two algorithms for finding optimal solutions of the Kemeny rank aggregation problem for full rankings
By D'Ambrosio, Amodio, Iorio

Published: 14 October 2015

Two algorithms for finding optimal solutions of the Kemeny rank aggregation problem for full rankings

Antonio D'Ambrosio[*a], Sonia Amodio[b], and Carmela Iorio[a]

[a]*Department of Economics and Statistics, University of Naples Federico II, Naples, Italy*
[b]*Department of Medical Statistics, Leiden University Medical Centre, Leiden, The Netherlands*

The analysis of ranking data has recently received increasing attention in many fields (i.e. political sciences, computer sciences, social sciences, medical sciences, etc.). Typically when dealing with preference rankings one of the main issue is to find a ranking that best represents the set of input rankings. Among several measures of agreement proposed in the literature, the Kendall distance is probably the most known. We propose a branch-and-bound algorithm to find the solution(s) even when we take into account a relatively large number of objects to be ranked. We also propose a heuristic variant of the branch-and-bound algorithm useful when the number of objects to rank is particularly high. We show how the solution(s) achieved by the algorithm can be employed in different analysis of rank data such as Mallow's ϕ model, mixtures of distance-based models, cluster analysis and so on.

keywords: Consensus ranking, Branch and bound, Mallows-ϕ model, exponential models.

1 Introduction

In the past few years preference rankings have gained an increased popularity in several fields of analysis, such as political sciences, computer sciences, marketing, etc. Preference rankings usually can be represented through either rank vectors or order vectors (Marden, 1996). Rank vectors are usually called rankings while order vectors are simply

---

*Corresponding author: antonio.dambrosio@unina.it

named orderings. Suppose there are n objects to be ranked by m judges: a ranking lists the ranks given to the objects, in which 1 means the best and n means the worst. An ordering, instead, lists the objects in order from the best to the worst. Suppose we have four objects to be ranked, and suppose that these objects are labeled A, B, C and D. Then, the ordering $< B\ A\ D\ C >$ corresponds to the ranking (2 1 4 3). Throughout the paper we use the terms rankings and orderings interchangeably since we give both the meaning of rank vectors.

Statistical methods and models for the analysis of rank data can be, broadly speaking, divided into methods based on badness-of-fit functions and methods based on probabilistic models (Marden, 1996; Heiser and D'Ambrosio, 2013). The former aim to describe the multidimensional structure of rank data. Statistical methods belonging to this category include Principal Component Analysis (Carroll, 1972), Multidimensional Scaling (Heiser and De Leeuw, 1981), Unfolding methods (Coombs, 1950, 1964; Busing et al., 2010), Preference mapping methods (Greenhoff and MacFie, 1994; Pagliuca and Scarpato, 2011).

Probabilistic methods that model the ranking process assume a homogeneous population of judges and include distance-based models (Mallows, 1957; Fligner and Verducci, 1986; Critchlow et al., 1991), multi-stage models (Fligner and Verducci, 1988) and Thurstonian models (Thurstone, 1927). Among the models that aim to identify homogeneous subpopulations, assuming for that reason heterogeneity among the judges, we recall mixtures of Bradley-Terry-Luce models, mixtures of distance-based models (Croon, 1989; Murphy and Martin, 2003; Gormley and Murphy, 2008) and clustering techniques (Heiser and D'Ambrosio, 2013). Regardless of whether homogeneity or heterogeneity among the judges is assumed or whether a probabilistic model is used, searching a ranking that is representative of a population (or of sub-populations) of judges is a central theme in the analysis of rank data. According to the Kemeny ranking problem (Kemeny, 1959), such ranking is called consensus or median ranking. The median ranking is defined as the ranking (or the rankings) that minimizes the sum of distances between itself and all input rankings. Unfortunately, finding the median ranking is a *Non-deterministic Polynomial-time* (NP) hard problem (Bartholdi III et al., 1989) because the searching space has the same dimension of the full permutations of the objects to be ranked. The problem is more complex if there are either incomplete or partial rankings. When a subject assigns integer values from 1 to n to all the n items we have a complete (or full) ranking. Whenever instead the judge fails to distinguish between two or more items and assigns to them the same integer number (expressing indifference about the relative order of this set of items), we deal with tied (or weak) rankings. Moreover we have a partial ranking when judges are asked to rank a subset of the entire set of objects (e.g. pick the q most favorite items out of a set of n). In these situations the dimension of the universe of rankings increases.

One of the main problem when working with rankings is the so-called *social choice* problem, also known as *rank aggregation* problem in computer science (Contucci et al., 2015), namely finding a ranking which most closely agrees with the set of the input rankings. Broadly speaking, there exist two classes of approaches to aggregate preference rankings to obtain a consensus (Cook, 2006): ad hoc methods and axiomatic based methods.

The ad hoc methods include elimination and non-elimination methods. The most famous non-elimination methods are the Bordas method of marks (de Borda, 1781) and the Condorcet's method (Condorcet, 1785): the first one is based on deriving the total of the ranks for each alternative assigned by the voters while the Condorcet's method is the simple majority rule on all pairwise comparisons between alternatives. The elimination methods evolved over time via parliamentary procedures; these methods include the Runoff from top method, the Runoff from bottom method, the American System and the Pairwise majority rule (Cook, 2006).

The axiomatic (or distance) based methods are based on a measure of distance of the desired consensus from the individual voter responses. In this paper we propose a branch-and-bound algorithm to find the median ranking under the experimental condition of both full rankings and *pick q out of n* experiments. The branch-and-bound programming strategy is one of the main tools in construction of solutions to NP-hard discrete optimization problems (Clausen, 1999). A branch-and-bound algorithm searches the complete space of solutions for a given problem by creating nested exhaustive branches to be evaluated by setting an upper (lower) bound to the optimal objective criterion value. The bounding process of the algorithm is used to evaluate partial solutions during the implementation of the procedure and cutting those branches that do not lead to optimal solutions (Brusco and Stahl, 2006). Our proposal consists of a variation of the algorithms proposed by Emond and Mason (2002) and Amodio et al. (2015) for partial rankings.

The paper is organized as follows: Section 2 introduces the geometry of rankings and the Kendall distance as the natural distance on such geometrical space, in Section 3 we describe our algorithm, Section 4 contains some evaluations of our proposal. The paper concludes with Section 5 in which we comment the obtained results.

2 Geometry of rankings and Kendall distance

The sample space of rankings consists of the vertices of a *permutation polytope* (Thompson, 1993; Heiser, 2004). This space is a convex hull of all vectors that are obtained by permuting the coordinates of a vector containing the first j integers forming an $(j-1)$-dimensional object. For such reason it can only be represented for $j = 3$ of with $j = 4$. In the case of $j = 4$, the polytope is a truncated octahedron with eight hexagonal and six squared faces (Figure 1).

As you can notice from Figure 1 there are four hexagons whose vertices show always a given object ranked in the first position, for example the red hexagon indicates the rankings for which object A is the most preferred. Moreover there are four hexagons whose vertices contain always the same object ranked in the last position, for example the hexagon placed in the opposite side with respect to the red hexagon contains the rankings for which A is always the less preferred object. The vertices of each square have always the same two objects ranked first and the remain objects ranked last (for example, the square connecting the red and the blue hexagons contains rankings which always present object A and B ranked before objects C and B). The vertices of the
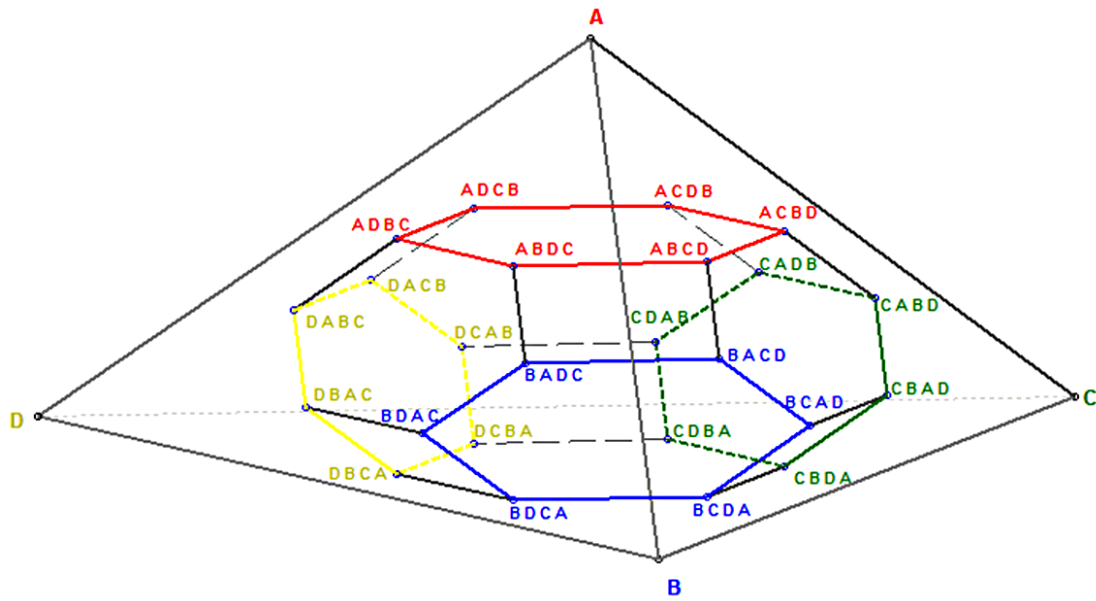
Figure 1: Permutation polytope for 4 objects embedded in a pyramid

polytope can also be interpreted as centers of gravity of the four items A, B, C, and D. If we allow the truncated octahedron to be embedded in a pyramid, the vertices of the pyramid indicate the poles of attraction of the objects (Heiser, 2004). Each of the four hexagon that present all the rankings in which a particular object is the most preferred are directed towards one vertex of the pyramid that represent the specific object. By following the same idea, the center of each face of the pyramid can be interpreted as a pole of repulsion for a given object. The edges of the polytope have all the same length. Two vertices are connected if and only if the corresponding permutations differ by an adjacent transposition. The distance between two vertices corresponds to the minimum number of transpositions of adjacent objects needed to transform one ranking into the other (Heiser, 2004, pag. 526). Note that, even if the polytope can be embedded in an Euclidean space, the shortest path distances are quite different from Euclidean distance. A geodesic distance, namely a distance based on the shortest path between two points in a curved space (Deza and Deza, 2009; Borg et al., 2012), is the most appropriate distance in this space. The natural distance measure defined on the permutation polytope is the Kendall distance. Given two different rankings on the permutation polytope of n objects, the Kendall distance is equal to the total number of steps to migrate from the first to the second ranking by reversing adjacent pairs of objects (Heiser, 2004). Note that if ties are allowed, the natural distance on the polytope is the Kemeny distance, which counts the number of interchanges of couples of elements that are required to transform one (partial) ranking into another (Emond and Mason, 2002; Heiser and D'Ambrosio, 2013; Amodio et al., 2015). More formally, let α and β be two rankings of n objects. A discordant pair is defined as a pair of distinct objects in which α and β have opposite

relative preferences, $(\alpha_i - \alpha_j)(\beta_i - \beta_j) < 0$ for $i \neq j$. The Kendall distance is then defined as (Marden, 1995, pag.25):

$$d(\alpha, \beta) = \sum \sum_{1 \leq i < j \leq n} I\left[(\alpha_i - \alpha_j)(\beta_i - \beta_j) < 0\right], \qquad (1)$$

where I is the indicator function. Kendall distance is popular because it is also largely used in the framework of distance-based models. The famous Mallows-ϕ model (Mallows, 1957) is a Mallows model with the Kendall distance.

3 The branch-and-bound algorithm and its variants

Given m rankings of n objects, we are interested in finding the ranking that best represents the consensus opinion. Let $\mathcal{Z}^n$ be the universe of all rankings with n objects with cardinality equal to $n!$. Let $\alpha_1, \ldots, \alpha_m$ be a set of rankings not necessarily distinct. According to the Kemeny ranking problem (Kemeny, 1959) the median ranking is the point σ for which

$$\sum_{k=1}^{m} d(\alpha_k, \sigma) = \min, \qquad (2)$$

among all $\sigma \in \mathcal{Z}^n$.

In the definition of his rank correlation coefficient, Kendall (1948) used the following matrix representation of a ranking. Let a_{ij} (b_{ij}) be the generic element of the $n \times n$ squared preference matrix $\boldsymbol{\alpha}$ $(\boldsymbol{\beta})$ called *score matrix*, with $i, j \in 1, \cdots, n$, in which $a_{ij} = 1$ if the ith object is preferred to the jth object, $a_{ij} = -1$ if the jth object is preferred to the ith object. Then the Kendall correlation coefficient is expressed as:

$$\tau_{(\alpha, \beta)} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{ij}}{n(n-1)}, \qquad (3)$$

in which $a_{ij} b_{ij}$ indicates the product term by term of the elements of the score matrices. Since there is always a one-to-one correspondence between a correlation coefficient and a distance measure, it is well known that the Kendall τ correlation coefficient can also be expressed in terms of the distance defined in Equation 1 as:

$$\tau_{(\alpha, \beta)} = 1 - 2 \frac{d(\alpha, \beta)}{max(d)}, \qquad (4)$$

with $max(d) = n(n-1)/2$.

This consideration allows us to define the median ranking as that ranking σ for which

$$\sum_{k=1}^{m} \tau_{(\alpha_k, \sigma)} = \max, \qquad (5)$$

among all $\sigma \in \mathcal{Z}^n$.

Emond and Mason (2002) defined the combined input matrix as a $n \times n$ matrix resulting from the aggregation of all the m score matrices calculated on the individual rankings in the input data. In other words, the combined input matrix is a matrix containing all the information about the sample of judges. Let $\alpha_1, \ldots, \alpha_m$ be a set of rankings with score matrices respectively equal to $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m$ with elements $a_{ij}^{(1)}, \ldots, a_{ij}^{(m)}$. The combined input matrix $\mathbf{C}$ is then defined as:

$$\mathbf{C} = \sum_{k=1}^{m} \boldsymbol{\alpha}_k, \tag{6}$$

whose elements are denoted by c_{ij}, obtaining

$$c_{ij} = \sum_{k=1}^{m} a_{ij}^{(k)}. \tag{7}$$

If s_{ij} represents the score matrix associated to the candidate consensus ranking, the Kemeny ranking problem can be written as:

$$\frac{\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} s_{ij}}{mn(n-1)} = max. \tag{8}$$

Moreover, if the m input rankings are not all distinct, Equation 8 can be rewritten by adding a positive weight w_k that indicates the frequency associated to the kth ordering, giving an equivalent formulation

$$\frac{\sum_{k=1}^{m} w_k \left[\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^{(k)} s_{ij} \right]}{[n(n-1)] \sum_{k=1}^{m} w_k} = max. \tag{9}$$

By looking at both Equations 8 and 9, we can notice that the maximization problem is then reduced to the search of the ranking σ whose score matrix s_{ij}, when multiplied term by term with the combined input matrix, returns the maximum. Recall that a score matrix contains only 1 and -1 in the cells that do not belong to the main diagonal. For this reason, if there is a ranking that can be exactly represented in terms of the same sign of the combined input matrix, then that particular ranking is the solution. In the same way as in Emond and Mason (2002), we compute the maximum possible value of Equation 9 as:

$$P^* = \sum_{i=1}^{n} \sum_{j=1}^{n} |c_{ij}|, \tag{10}$$

and then we consider a candidate σ with score matrix $\{s_{ij}\}$ and we evaluate the value of Equation 9 as:

$$p = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} s_{ij}}{mn(n-1)} = \frac{\sum_{k=1}^{m} w_k \left[\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^{(k)} s_{ij} \right]}{[n(n-1)] \sum_{k=1}^{m} w_k}. \tag{11}$$

The starting point of the searching problem is the definition of an initial penalty, which is defined as:

$$P = P^* - p. \tag{12}$$

At the first stage of the algorithm we consider a candidate σ to be the consensus ranking. We fix the favorite object and consider the second ranked object. This object can be ranked either ahead or behind with respect to the first one. So, the set of all ordering can be divided into two branches based on the relative position of these first two objects. For each branch, an incremental penalty is computed by looking at the values of the combined input matrix in this way:

- object i is preferred to object j:
  if $c_{ij} > 0$ *and* $c_{ji} < 0$, then $\delta P = 0$
  if $c_{ij} < 0$ *and* $c_{ji} > 0$, then $\delta P = c_{ji} - c_{ij}$

- object j is preferred to object i:
  if $c_{ij} > 0$ *and* $c_{ji} < 0$, then $\delta P = c_{ij} - c_{ji}$
  if $c_{ij} < 0$ *and* $c_{ji} > 0$, then $\delta P = 0$,

where δP represents the incremental penalty.
If the incremental penalty for a branch is greater than the initial penalty, then we do not consider it any longer because all orderings in that branch will have a penalty larger than the initial one. If the incremental penalty of a branch is smaller or equal to the initial penalty, then we consider the next object in the initial solution and we create new branches by placing this object in all possible positions relative to the objects already considered. For each new branch, the incremental penalties are computed just as before and they are summed to the existing cumulated penalty for the branch. The algorithm continues in this way by cutting useless branches and evaluating useful branches until a solution is achieved. The algorithm is able to find the solution(s) in a reasonable time (in a blink of an eye for up to 15 objects, in a few seconds up to 25 objects, in a few minutes up to 35 objects). The execution time depends on both the quality of the initial solution and the degree of consensus around the median ranking. When there is a large number of objects to be ranked both the Quick and Fast heuristic algorithms for partial rankings as defined by Amodio et al. (2015) can be used.
The Quick algorithm processes only one branch at time by adding and interchanging at turn one object in all possible position, and evaluating Equation 11 taking in account also all the remaining objects. The Fast algorithm repeats a pre-specified number of times the Quick algorithm by selecting as initial candidate a random permutation of the objects. For more details on these algorithms we refer to Amodio et al. (2015).
The proposed algorithm also works with experimental situations of the type *pick q out of n* (or with some missing ranked object) with the convention that unranked objects do not add any information in the formation of the combined input matrix. By the way, note that if there are missing ranked objects, the distance measure is not defined.

4 Experimental evaluation

In this section we evaluate the proposed algorithm with both simulated and real data sets. Moreover we show an application of our algorithm to a clustering problem via mixture of distance-based models as defined by Murphy and Martin (2003). Data analyses were performed with our own program written in MatLab language in a Computer Intel Core i5-3317U 1.70 GHz and 4GB of RAM. MatLab codes are available upon request to authors. The R-package `ConsRank` (D'Ambrosio and Amodio, 2015) implementing the same routines is freely available (`https://cran.r-project.org/web/packages/ConsRank/index.html/`).

4.1 Simulation setting: Mallows-ϕ models

We simulated data under a Mallows-ϕ model. In the Mallows model, for a given consensus σ, a distance function d and some real parameter λ, the density with respect to the *Uniform* is (Marden, 1996)

$$f_\lambda (\alpha; \sigma) = \exp \left(\lambda d (\sigma, \alpha) - \psi (\lambda) \right), \tag{13}$$

where α is a ranking and ψ is a normalizing factor.

Let $\mathcal{Z}^n$ be the universe of all orderings with n objects and let h be a number in the range ς of $d(\sigma, \alpha)$, then

$$P_\lambda \left[d (\sigma, A) = h \right] = c (h) \exp \left(\lambda h - \psi (\lambda) \right), \tag{14}$$

where $\exp \left(\psi (\lambda) \right) = \sum_{h \in \varsigma} c (h) \exp (\lambda h)$ and $c (h) = \# \{ \alpha \in \mathcal{Z}^n | d (\sigma, A) = h \}$.

Rankings nearer to the median ranking σ show a higher probability of occurrence and this is controlled by λ. In general we have: $\sum_{k=1}^m d(\sigma, \alpha^{(k)}) = 0, \lambda \Rightarrow -\infty$; $\sum_{k=1}^m d(\sigma, \alpha^{(k)}) = m \left[n(n-1) \right], \lambda \Rightarrow \infty$; if $\lambda = 0$, then each ranking in $\mathcal{Z}^n$ is equally likely (cfr. Marden, 1996, pag. 150). In particular, when the Kendall distance is used, the model is called Mallows-ϕ model.

Broadly speaking, researchers using distance-based models are interested in testing the discrepancy of the collected data from the null hypothesis of uniform distribution. The parameters of this model are usually estimated by Maximum Likelihood. The Maximum Likelihood Estimation (MLE) of λ can be found, for example, through Central Limit Theorem or Newton-Raphson algorithms when the distribution of the distance is known, as in the case of the Kendall distance (Marden, 1996). An issue is that a good estimate of λ depends on the correct identification of the median ranking. There is a one-to-one correspondence between searching the median ranking and formally testing for uniform distribution of a (sub)population of judges. Our experimental setting aims to show the one-to-one correspondence between the estimate of parameters of Mallows-ϕ models and the search of the median ranking. The results of the first experiment are shown in Table 1. To obtain such results, we randomly extracted 80 rankings from the population of four

objects by setting $\lambda = -0.40$ and $\sigma = < A\ B\ C\ D >$. We then computed the consensus ranking ($\hat{\sigma} = < ABCD >$, $\tau = 0.308$, elapsed time 0.008 seconds) and we got the ML estimate of λ ($\hat{\lambda} = -0.446$). We estimated also the λ parameter and the value of τ as in Equation 9 for all rankings $\in \mathcal{Z}^n$.

Table 1: Population of rankings with 4 objects. ML estimate of λ and average value of τ for each ranking in $\mathcal{Z}^n$

Orderings	$\hat{\lambda}$	SE λ	τ
$< ACBD >$	-0.306	0.078	0.217
$< ADBC >$	-0.246	0.077	0.175
$< ABCD >$	**-0.446**	0.081	**0.308**
$< ABDC >$	-0.381	0.080	0.267
$< ACDB >$	-0.174	0.077	0.125
$< ADCB >$	-0.116	0.076	0.083
$< DABC >$	-0.035	0.076	0.025
$< CABD >$	-0.116	0.076	0.083
$< DACB >$	0.092	0.076	-0.067
$< CADB >$	0.012	0.076	-0.008
$< BADC >$	-0.222	0.077	0.158
$< BACD >$	-0.282	0.078	0.200
$< DCAB >$	0.282	0.078	-0.200
$< CDAB >$	0.222	0.077	-0.158
$< DBAC >$	0.116	0.076	-0.083
$< CBAD >$	0.035	0.076	-0.025
$< BCAD >$	-0.092	0.076	0.067
$< BDAC >$	-0.012	0.076	0.008
$< DCBA >$	0.446	0.081	-0.308
$< CDBA >$	0.381	0.080	-0.267
$< DBCA >$	0.306	0.078	-0.217
$< CBDA >$	0.246	0.077	-0.175
$< BCDA >$	0.116	0.076	-0.083
$< BDCA >$	0.174	0.077	-0.125

The Pearson correlation coefficient between the estimates of λ and the values of τ is equal to -0.999. This negative correlation is due to the fact that always larger values of Kendall τ correspond to lower values of $\hat{\lambda}$ and this represents an important result. The maximum likelihood estimate of λ is conditioned to the maximum likelihood estimate of the consensus ranking. In general for large n, a global search algorithm for MLE of σ is not practical because the number of possible choices is too large. In these cases local

search algorithms are used (Alvo and Yu, 2014). An accurate method to compute the median ranking is then a necessary step to get a better maximum likelihood estimate of the dispersion parameter, whatever distance measure is used.

Moreover, we repeated the experiment with three values of $\lambda = (-0.02, -0.1, -0.5)$ for the universes with $n = 5, 6, 7, 8, 9$ and 10 and a sample size always equal to 150. The consensus ranking was always set as $\sigma = (1\ 2\ \ldots\ n)$.

Table 2: Experimental evaluation: main results

Experiment		$\hat{\lambda}$	SE $\hat{\lambda}$	τ	corr($\hat{\lambda}.\tau$)
	$\hat{\lambda} = -0.02$	-0.067	0.040	0.056	-1.000
5 objects	$\hat{\lambda} = -0.10$	-0.132	0.041	0.109	-1.000
	$\hat{\lambda} = -0.50$	-0.463	0.044	0.361	-0.999
	$\hat{\lambda} = -0.02$	-0.075	0.031	0.07	-1.000
6 objects	$\hat{\lambda} = -0.10$	-0.141	0.031	0.132	-1.000
	$\hat{\lambda} = -0.50$	-0.539	0.036	0.454	-0.999
	$\hat{\lambda} = -0.02$	-0.049	0.025	0.053	-1.000
7 objects	$\hat{\lambda} = -0.10$	-0.154	0.025	0.16	-1.000
	$\hat{\lambda} = -0.50$	-0.488	0.029	0.456	-0.999
	$\hat{\lambda} = -0.02$	-0.071	0.020	0.082	-1.000
8 objects	$\hat{\lambda} = -0.10$	-0.126	0.021	0.145	-1.000
	$\hat{\lambda} = -0.50$	-0.463	0.025	0.471	-0.999
	$\hat{\lambda} = -0.02$	-0.062	0.017	0.079	-1.000
9 objects	$\hat{\lambda} = -0.10$	-0.115	0.017	0.145	-1.000
	$\hat{\lambda} = -0.50$	-0.497	0.022	0.529	-0.999
	$\hat{\lambda} = -0.02$	-0.018	0.014	0.051	-1.000
10 objects	$\hat{\lambda} = -0.10$	-0.101	0.014	0.146	-1.000
	$\hat{\lambda} = -0.50$	-0.491	0.019	0.552	-0.999

As expected, there is perfect correlation between the values of $\hat{\lambda}$ and the averaged τ: each $\sigma \in \mathcal{Z}^n$ corresponds to a value of $\hat{\lambda}$ unequivocally related to τ. The branch-and-bound algorithm always obtained the solution(s) in less than half second. In the case with 5 objects and $\lambda = -0.02$ two solutions were found, while in the case with 7 objects and $\lambda = -0.02$ seven solutions were found.

4.2 Real data

In this section, we present the results obtained analyzing some real ranking data. Note that in some of these analyses, it is not always possible to get the ML estimates of λ because of the specific characteristics of data.

APA data set

APA data set is a collection of rankings about voters for the 1980 election of American Psychological Association president (Diaconis, 1988). This data set contains the rankings expressed by 15,449 psychologists on five candidates: A = Bevan, B = Iscoe, C = Kiesler, D = Siegle and E = Wriths. Of these rankings only 5,738 are complete while the remaining are partials. For the 5,738 full rankings we have this results: $\sigma = <ACEDB>$, $\tau = 0.060$, elapsed time = 0.914 seconds, $\hat{\lambda} = -0.072$, SE $\hat{\lambda} = 0.006$.

For the complete data set with 15,449 judges we have these results: $\sigma = <CAEDB>$, $\tau = 0.023$, elapsed time = 4.113 seconds.

Irish election data set

Irish election data set was described by Gormley and Murphy (2008). The data set consists of an opinion pool conducted by Irish Marketing Surveys in 1997. There are 1083 valid (full and/or incomplete) rankings (and also 6 covariates). The Candidate McAleese won the electoral competition, candidates Nally, Roche and Scallon were eliminated after the first votes counting. The candidate Banotti arrived at the second stage of counting. The number of full rankings is 807 and the number of incomplete rankings is 276. Ties were not permitted, but voters were asked to rank some or all candidates. The results for this data set are equal to: $\sigma = <$McAleese Banotti Roche Nally Scallon$>$, $\tau = 0.285$, elapsed time 0.331 seconds.

European Values Study data

European Values Study (EVS) data are the ranked version of the data set coming from the European Values Studies, which is a large-scale, cross-national, and longitudinal survey research program on basic human values. The data correspond to the random sub-sample of 3584 cases of the survey conducted in 1999 in 32 countries analyzed by Vermunt (2003). Rankings were obtained by applying the post-materialism scale developed by Inglehart (1977). This scale in based upon an experiment of the type *pick 2 out of 4* most important political goals for your Governments. The alternatives were (A) Maintain order in Nation, (B) Give people more say in Government decisions, (C) Fight rising prices, (D) Protect freedom of speech. The results of the analysis of this data were: $\sigma = <ABCD>$, $\tau = 0.051$, elapsed time 0.845 seconds.

Sports data set

The Sports data set comes from Louis Roussos and is described in Marden (1996). Roussos asked 130 students at the University of Illinois to rank seven sports according to their preference of participating in. The sports to choose from were: (A) baseball, (B) football, (C) basketball, (D) tennis, (E) cycling, (F) swimming and (G) jogging. On this data we report the following results: $\sigma = <EFCADBG>$, $\tau = 0.144$, elapsed time 0.155 seconds, $\hat{\lambda} = -0.138$, SE $\hat{\lambda} = 0.026$.

The shown results confirm that values of $\hat{\lambda}$ close to zero (which means low degree of consensus around the median ranking) are linked to extremely low values of the averaged τ. On the other hand, large values of $\hat{\lambda}$ (in absolute value) are linked to large values of

the averaged τ, which means a strong degree of consensus around the median ranking. This is an important remark for situations in which distance-based models for rankings cannot be used (as in example for either APA or Irish election data sets) because there are several incomplete rankings which are not governed by specific patterns of missingness or ties (Marden, 1996, chapter 11). Note that in real situations both tied and incomplete rankings are not always governed by fixed patterns but arise empirically in the data.

4.3 Mixture of distance-based models

Mallows-ϕ model has been extended to mixture models by Murphy and Martin (2003). A mixture model assumes that the observed rankings are coming from H probability distributions where each distribution characterizes a cluster. Every cluster has mixing probability p_H, with $0 \leq p_H \leq 1$ of appearing in the population. Each density function f_H is characterized by a central ranking σ_H and a spread parameter λ_H. The log-likelihood function of the mixture of the weighted distance-based model on a data set $\boldsymbol{\alpha}$ of m rankings can be written as:

$$l(\sigma, \lambda, p | \boldsymbol{\alpha}_m) = \sum_{k=1}^{m} \log \left\{ \sum_{h=1}^{H} p_h \exp \left(\lambda_h d \left(\sigma_h, \alpha_k \right) - \psi \left(\lambda_h \right) \right) \right\} \tag{15}$$

The model, fitted by maximum likelihood via EM algorithm, obtains maximum likelihood estimates. Note that with respect to Equation 13, Equation 15 presents a small difference, i.e. $\exp(\lambda)$ instead of $\exp(-\lambda)$. This consideration allows to consider an inversion of the sign of the values of $\hat{\lambda}$ (see also Diaconis, 1988; Critchlow et al., 1991). For more details about mixture of distance-based models, we refer to Murphy and Martin (2003). We ran the algorithm on the Sports data set by setting $H = 2, 3, 4, 5, 6, 7, 8, 9, 10$. BIC criterion suggested us the solution with $H = 2$ as well as the ICL criterion suggested us the solution with $H = 5$. Log-likelihood was maximum for the solution with $H = 2$. As suggested by Murphy and Martin (2003) we repeated each analysis with 30 different random starting points for each H. Results are reported in Table 3.

4.4 Analysis on a simulated data set with 100 objects

We tested both the Quick and Fast algorithms on a really difficult case. We generated a data set with 15 different rankings of 100 objects. Note that rank data are an excellent example of 'big data with few observations' in the sense that the complexity of the problem is exclusively governed by the number of objects involved in the choice process and not by the number of judges. These rankings were generated by randomly permuting the first 100 numbers. Then we randomly extracted 15 values from a normal distribution with mean equal to 100 and variance equal to 10. Finally we got the frequencies to associate to rankings by dividing each value by their sum. The sample size was set equal to 200. The branch-and-bound algorithm is not practical in such a case due to the huge number of objects. Even if a large number of branches can be eliminated, the universe counts 100! different rankings and the algorithm could run for several weeks

Table 3: Mixture of distance-based models results, Sports data set

H	σ_h	$\hat{\lambda}_h$	p_h
2	$< EDCAB >$	-0.294	0.317
	$< ACBDE >$	-0.157	0.683
5	$< CDBAE >$	-0.325	0.259
	$< EBACD >$	-0.492	0.113
	$< EADCB >$	-0.373	0.153
	$< DEACB >$	-0.403	0.151
	$< ABCDE >$	-0.226	0.324

and the storage capacity of the computer could be not enough. The data set is available at *http* : *//wpage.unina.it/antdambr/hundredrankdata/*. The Quick algorithm found one solution in 128.457 seconds with an average $\tau = 0.177$. The Fast algorithm was ran with 100 replications. It found a better solution (average $\tau = 0.178$) in $6,570.023$ seconds.

5 Concluding remarks

In this paper we proposed a branch-and-bound algorithm to find the median ranking under the experimental condition of both full rankings and *pick q out of n* experiment. The algorithm consists of a variation of the algorithms proposed by Emond and Mason (2002) and Amodio et al. (2015) for partial rankings.

We showed that maximum likelihood estimates of the parameters when considering the Mallows-ϕ model largely depends on the estimate of the median ranking. To estimate such a ranking people can either use exhaustive searches by considering the universe of rankings or use locally searching or heuristics algorithms. By adapting the ideas developed by Emond and Mason (2002) and Amodio et al. (2015) for partial rankings we propose an algorithm able to find all the solutions in a reasonable time up to a number of objects equal to 35. For very difficult problems, an accurate solution is also provided. The algorithm by Amodio et al. (2015) was used in finding the median rankings for recursive partitioning algorithms for preference rankings (D'Ambrosio and Heiser, 2009, 2015) and for non-parametric clustering for rankings (Heiser, 2004; Heiser and D'Ambrosio, 2014). If the experimental setting does not allow for ties, the proposed method for full rankings can be used for similar approaches.

We showed useful applications of the proposed algorithms on problems such as ML estimation of Mallows-ϕ model, mixture of distance-based models, computation of the consensus ranking when distance-based models cannot be estimated, estimation of the consensus ranking with extremely complicated data sets.

Recently, the analysis of preference rankings has been integrated with the use of weighted measures of agreement between two rankings, proposed in the literature for giving more importance to the concordance on most important ranks (the top ranks) rather than the concordance on lower ranks. There exist several weighted measures dealing with preference rankings such as weighted Spearman ρ and Spearman footrule (Shieh et al., 2000; Tarsitano, 2009), weighted Kendall τ Shieh (1998), weighted Kemeny distance (Garcıa-Lapresta and Pérez-Román, 2010). There are possibilities of further improvements by considering the use of weighted measures, especially when dealing with a large number of objects. This is the case of the web searching, where the discounted cumulative gain (DCG) (Järvelin and Kekäläinen, 2002) is usually used as a measure of ranking quality. Another future issue could be the merging of a weighted measure with the DCG to improve both searching process and interpretability.

Acknowledgement

References

Alvo, M. and Yu, P. L. (2014). *Statistical Methods for Ranking Data*. Springer.

Amodio, S., D'Ambrosio, A., and Siciliano, R. (2015). Accurate algorithms for identifying the median ranking when dealing with weak and partial rankings under the kemeny axiomatic approach. *European Journal of Operational Research*. doi: http://dx.doi.org/10.1016/j.ejor.2015.08.048.

Bartholdi III, J., Tovey, C. A., and Trick, M. A. (1989). Voting schemes for which it can be difficult to tell who won the election. *Social Choice and welfare*, 6(2):157–165.

Borg, I., Groenen, P. J., and Mair, P. (2012). *Applied multidimensional scaling*. Springer Science & Business Media.

Brusco, M. J. and Stahl, S. (2006). *Branch-and-bound applications in combinatorial data analysis*. Springer Science & Business Media.

Busing, F. M., Heiser, W. J., and Cleaver, G. (2010). Restricted unfolding: Preference analysis with optimal transformations of preferences and attributes. *Food quality and preference*, 21(1):82–92.

Carroll, J. D. (1972). Individual differences and multidimensional scaling. *Multidimensional scaling: Theory and applications in the behavioral sciences*, 1:105–155.

Clausen, J. (1999). Branch and bound algorithms-principles and examples. *Department of Computer Science, University of Copenhagen*, pages 1–30.

Condorcet, M. (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. L'imprimerie royale.

Contucci, P., Panizzi, E., Ricci-Tersenghi, F., and Sîrbu, A. (2015). Egalitarianism in

the rank aggregation problem: a new dimension for democracy. *Quality & Quantity*, pages 1–16.

Cook, W. D. (2006). Distance-based and ad hoc consensus models in ordinal preference ranking. *European Journal of Operational Research*, 172(2):369–385.

Coombs, C. H. (1950). Psychological scaling without a unit of measurement. *Psychological review*, 57(3):145.

Coombs, C. H. (1964). A theory of data.

Critchlow, D. E., Fligner, M. A., and Verducci, J. S. (1991). Probability models on rankings. *Journal of mathematical psychology*, 35(3):294–318.

Croon, M. A. (1989). Latent class models for the analysis of rankings. *Advances in psychology*, 60:99–121.

D'Ambrosio, A. and Amodio, S. (2015). *ConsRank, Compute the Median Ranking(s) According to the Kemeny's Axiomatic Approach.* R package version 1.0.2. `https://cran.r-project.org/web/packages/ConsRank/index.html`.

D'Ambrosio, A. and Heiser, W. J. (2009). Decision trees for preference rankings. In *Procceedings of the 7th Meeting of the Classification and Data Analysis Group, Catania - Italy*, pages 133–136. CLEUP–Padova. Invited talk.

D'Ambrosio, A. and Heiser, W. J. (2015). A recursive partitioning method for the prediction of preference rankings based upon kemeny distances. *Technical report, University of Naples Federico II*. Submitted manuscript.

de Borda, J. C. (1781). Mémoire sur les élections au scrutin.

Deza, M. M. and Deza, E. (2009). *Encyclopedia of distances.* Springer.

Diaconis, P. (1988). Group representations in probability and statistics. *Lecture Notes-Monograph Series*, pages i–192.

Emond, E. J. and Mason, D. W. (2002). A new rank correlation coefficient with application to the consensus ranking problem. *Journal of Multi-Criteria Decision Analysis*, 11(1):17–28.

Fligner, M. A. and Verducci, J. S. (1986). Distance based ranking models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 359–369.

Fligner, M. A. and Verducci, J. S. (1988). Multistage ranking models. *Journal of the American Statistical association*, 83(403):892–901.

Garcıa-Lapresta, J. L. and Pérez-Román, D. (2010). Consensus measures generated by weighted kemeny distances on weak orders. In *Procceedings of the 10th International Conference on Intelligent Systems Design and Applications, Cairo*.

Gormley, I. C. and Murphy, T. B. (2008). Exploring voting blocs within the irish electorate: A mixture modeling approach. *Journal of the American Statistical Association*, 103(483):1014–1027.

Greenhoff, K. and MacFie, H. (1994). Preference mapping in practice. In *Measurement of food preferences*, pages 137–166. Springer.

Heiser, W. J. (2004). Geometric representation of association between categories. *Psychometrika*, 69(4):513–545.

Heiser, W. J. and D'Ambrosio, A. (2013). Clustering and prediction of rankings within a kemeny distance framework. In *Algorithms from and for Nature and Life*, pages 19–31. Springer.

Heiser, W. J. and D'Ambrosio, A. (2014). K-median cluster component analysis. *Technical report, University of Naples Federico II*. Submitted manuscript.

Heiser, W. J. and De Leeuw, J. (1981). Multidimensional mapping of preference data. *Mathématiques et Sciences humaines*, 73:39–96.

Järvelin, K. and Kekäläinen, J. (2002). Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4):422–446.

Kemeny, J. G. (1959). Mathematics without numbers. *Daedalus*, 88(4):577–591.

Kendall, M. G. (1948). *Rank correlation methods*. Griffin.

Mallows, C. L. (1957). Non-null ranking models. i. *Biometrika*, pages 114–130.

Marden, J. I. (1996). *Analyzing and modeling rank data*. CRC Press.

Murphy, T. B. and Martin, D. (2003). Mixtures of distance-based models for ranking data. *Computational statistics & data analysis*, 41(3):645–655.

Pagliuca, M. M. and Scarpato, D. (2011). Food quality, consumer perception and preferences: an analysis on olive oil. *Electronic Journal of Applied Statistical Analysis*, 4(2):215–226.

Shieh, G. S. (1998). A weighted kendall's tau statistic. *Statistics & Probability Letters*, 39(1):17–24.

Shieh, G. S., Bai, Z., and Tsai, W.-Y. (2000). Rank tests for independence-with a weighted contamination alternative. *Statistica Sinica*, 10(2):577–594.

Tarsitano, A. (2009). Comparing the effectiveness of rank correlation statistics. *Working Papers 200906, Universitá della Calabria, Dipartimento di Economia e Statistica*.

Thompson, G. (1993). Generalized permutation polytopes and exploratory graphical methods for ranked data. *The Annals of Statistics*, pages 1401–1430.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological review*, 34(4):273.

Vermunt, J. K. (2003). Multilevel latent class models. *Sociological methodology*, 33(1):213–239.