A generative model for rank data based on sorting algorithm

Christophe Biernacki, Julien Jacques

A generative model for rank data based on sorting algorithm

Christophe BIERNACKI[a], Julien JACQUES[a]

Abstract

Rank data arise from a sorting mechanism which is generally unobservable for the statistician. Retaining the insertion sorting algorithm because of its well known optimality properties and combining it with a natural stochastic error in the pair comparison process allows to propose a parsimonious and meaningful parametric generative model for rank data. Its theoretical properties are studied like unimodality, symmetry and identifiability. In addition, maximum likelihood principle can be easily performed through an EM algorithm thanks to an unobserved latent variables interpretation of the model. Finally, an illustration of adequacy between the proposed model and rank data resulting from a general knowledge quiz suggests the relevance of our proposal.

Résumé

Les données de rang sont le résultat d'un processus de tri généralement non observable par le statisticien. En retenant un algorithme de tri par insertion pour ses propriétés d'optimalité et en introduisant une erreur stochastique dans le processus de comparaison par paire, nous proposons un modèle génératif parcimonieux pour les données de rang. Ses propriétés théoriques comme l'unimodalité, la symétrie et l'identifiabilité sont étudiées. L'estimation des paramètres du modèle par maximum de vraisemblance utilisant l'algorithme EM est présentée. Enfin, une illustration de l'adéquation du modèle proposé sur un jeu de données réelles résultant d'un quiz de culture générale met en évidence l'intérêt de notre proposition.

Key words and phrases. EM algorithm, insertion algorithm, quiz data, rank data, sorting process.

[a] Laboratoire P. Painlevé, UMR 8524 CNRS Université Lille I, Bât M2, Cité Scientifique, F-59655 Villeneuve d'Ascq Cedex, France.

1 Introduction

Ranking data are of great interest in human activities involving preferences, attitudes or choices like Web Page ranking, Sport, Politics, Economics, Educational Testing, Biology, Psychology, Sociology, Marketing, *etc.* Ranks are so meaningful that it is not unusual they result from a transformation of other kinds of data.
Rank data are multivariate but highly structured data. So, beyond standard but general data analysis methods (means, factor analysis, *etc.*), some specific descriptive methods which respects this structure have been proposed, for instance the permutation polytope for plotting the rank vectors in Euclidean space [16, 17] or also suitable distances for defining the centre and spread of a dataset [9, 13, 7].
From an inference point of view, distances are useful for testing the distribution of these data (uniformity, populations comparison [6, 14]) or for modeling the distribution itself (for instance the Mallows Φ model relies on the Kendall distance [9, 3]). More generally, parametric probabilistic models, if relevant and allowing easy parameter interpretation, are useful for summarizing and understanding such quite complex data and are a basis tool for density estimation, prediction or clustering. Major rank data models date from the mid 20th century and most of the current works on the topic uses these models. Pointing out that a rank data is the result of a sorting process, we suggest in this paper a generative model for rank data, based on a modeling of the sorting process which aims to be optimal in a sense explained in the next section.
So, Section 2 is devoted to the notation and the interpretation of usual rank data models as the modeling of particular sorting algorithms. Section 3 introduces the proposed model which is based on an insertion sorting algorithm, and its theoretical properties (unimodality, symmetry, identifiability) are detailed in Section 4. Maximum likelihood estimation is considered in Section 5 by the mean of an EM algorithm since a missing data interpretation of the proposed model can be pointed out, and numerical illustrations are presented in Section 6 to evaluate the relevance of the new model both from distributional visualization point of view and from adequacy to some real data sets. A discussion on the numerous perspectives concludes this work in Section 7.

2 Notation and usual rank data models

The rank datum, which is the statistical unit of interest in this paper, results from a ranking of m objects $\mathcal{O}_1, \mathcal{O}_2, \ldots, \mathcal{O}_m$ by a judge (human or not). Two representations of these data are commonly used : Ranking or ordering. The *ranking* representation $x^{-1} = (x_1^{-1}, x_2^{-1}, \ldots, x_m^{-1})$ contains the ranks given to the objects, and means that $\mathcal{O}_1$ is in the x_1^{-1}-th position, $\mathcal{O}_2$ is in the x_2^{-1}-th position, an so forth. A ranking is then an element of $\mathcal{P}_m$, the set of permutations of the m first integers. The *ordering* representation $x = (x_1, x_2, \ldots, x_m)$ is also an element of $\mathcal{P}_m$ and signifies that object $\mathcal{O}_{x_1}$ is the first, object $\mathcal{O}_{x_2}$ is the second, *etc.* Let consider the following example to illustrate these two notations : A judge, which has to rank by preference order three holidays destinations ($\mathcal{O}_1$ = Campaign, $\mathcal{O}_2$ = Mountain and $\mathcal{O}_3$ = Sea), ranks first Sea, second Campaign, and last Mountain. The ordering result of the judge is $x = (3, 1, 2)$ whereas the ranking result is $x^{-1} = (2, 3, 1)$. In the following both ordering and ranking notations will be used for rank data.

The two most popular classes of models for rank data consist in modeling directly the hypothetical ranking process followed by the judge. For a complete review, refer to [14, Chap. 5 to 10]. The first class is derived from a paired comparison process [10] : The judge constructs a rank by first comparing each pair of objects, and second ensuring the consistency of these paired comparisons (if $\mathcal{O}_1$ is prefered to $\mathcal{O}_2$ and $\mathcal{O}_2$ to $\mathcal{O}_3$, $\mathcal{O}_1$ must be prefered to $\mathcal{O}_3$). It follows the *Babington Smith model* for a rank x :

$$\mathrm{pr}(x) \propto \prod_{1 \leq i < j \leq m} p_{ij},$$

with p_{ij} the probability that $\mathcal{O}_i$ is prefered to $\mathcal{O}_j$, and where the proportionality is due to the need of consistency of the paired comparisons. The number of parameters of this model being very large, especially when m grows, some simplifications have been considered. [2] associate to each object $\mathcal{O}_j$ a score v_j indicating an overall degree of preference of this object, and connect these scores to

p_{ij} by $p_{ij} = v_i/(v_i + v_j)$, which defines the *Bradley–Terry–Mallows model*. [13] goes forward into the simplification by imposing that p_{ij} only depends on the sign of $i - j$, which leads to the famous *Mallows* Φ *model*. A property of this latter is that it can be linked to the Kendall distance between two rankings [9, 3].
The second popular class of rank data models is multistage models, which considers the following iterative ranking process : The judge selects firstly the best object among the m ones, then the best among the $m - 1$ remaining ones, and so forth. Noting v_i the probability that $\mathcal{O}_i$ is ranked first among the m objects, the corresponding *Plackett-Luce model* [12, 15] defines the probability of a rank x as

$$\mathrm{pr}(x) = \prod_{j=1}^{m} \frac{v_{x_j}}{v_{x_j} + v_{x_{j+1}} + \ldots + v_{x_m}}.$$

The term in the product means the probability that $\mathcal{O}_{x_j}$ is ranked first among objects $\mathcal{O}_{x_j}$ to $\mathcal{O}_{x_m}$. It could be noticed that this model corresponds to a Thurstonian model [18, 1] with a Gumbel density. [7, 8] introduce an alternative multistage model by considering another form of the probability at each step of the ranking process. The model deriving from this parametrization is particularly interesting because it leads, for a special value of its parameter, to the Mallows Φ model.

The ranking processes which have motivated these two classes of rank data models can be interpreted as two different sorting processes, in which stochastic errors are introduced to define a probability distribution on the whole rank data space. The natural question involved by this interpretation is whether the sorting algorithms used are the most appropriate. Effectively, in paired comparison models it seems not optimal to do so much comparisons since it leads to a sorting algorithm with excessively high computational complexity. In practice a human judge would probably not exhaustively proceed to all paired comparisons. For multistage models, the ranking process can be likened to a *selection* sorting algorithm. Even if this sorting algorithm is one of the most simple, it is well known for its lack of optimality [11]. Here, we propose a generative model for rank data based on the (straight) *insertion* sorting algorithm, which is one of the most powerful among the usual sorts when $m \leq 10$ [11, Chap. 5]. In addition, our proposal is potentially able to take into account the presentation order of the different objects to the judge, realistic situation which can have an impact on the resulting rank.

3 A generative model for rank data based on an insertion sorting

We assume there exists an ordering $\mu = (\mu_1, \ldots, \mu_m)$ on the m objects, so that a judge who perfectly sorts these objects returns this *reference* rank μ. Moreover, we assume that the judge adopts one of the best sorting strategy for a small number of objects ($m \leq 10$), which is the *insertion* sorting algorithm. We also introduce the possibility for the judge of making mistakes regarding to μ in his sorting, and such mistakes will be modeled by a random event in paired comparison. Merging both deterministic insertion algorithm and the random paired comparison leads to a meaningful generative model for rank data that is now presented at length.

Let the ordering $\sigma = (\sigma_1, \ldots, \sigma_m)$ be the presentation order of the objects to the judge, this latter using an insertion sorting algorithm to rank these objects. The current object to be sorted is placed on the left of the already sorted objects, and is compared to the first objet on its right. If the relative position of both objects in this pair is correct (according to μ), this pair order is unchanged and the next object in σ is inserted far left. Otherwise, the pair order is reversed and a new pair comparison is performed with the next object on the right (if it exists). And so forth. The result of this deterministic sorting algorithm would be μ if the judge was perfect. However, none judge is perfect and the mistakes he/she/it can do leads to a given rank $x = (x_1, x_2, \ldots, x_m)$, which could be different from μ. Since the sorting algorithm by insertion consists solely of a sequence of comparisons of pairs of objects, it is natural to model the reliability of the judge for the ranking by the risk of wrongly order a pair of objects. Each pair comparison can be interpreted as the result of a Bernoulli experiment whose outcome is a correct comparison (according to μ) with probability

p and an incorrect comparison with probability $1-p$. Moreover, it is reasonable to assume that each pair ranking operation is independent of the others. Based on this modeling of a stochastic insertion sorting, the first natural question is to calculate the probability $pr(x|\sigma;\mu,p)$ to obtain a rank x from an initial presentation order σ and a reference rank μ. To do so, let introduce the following notations, where $j=1,\ldots,m$ denotes the step in the sorting algorithm consisting of ranking the object $\mathcal{O}_{\sigma_j}$. The notations and their use in the proposed sorting algorithm are both illustrated in Table 1.

- $\delta_{ii'}(\mu)=\mathbf{1}\{\mu_i^{-1}<\mu_{i'}^{-1}\}$, which is equal to 1 if $\mathcal{O}_i$ is correctly ranked before $\mathcal{O}_{i'}$ (according to μ), 0 otherwise $(i,i'=1,\ldots,m, i\neq i')$.
- $j^-(x,\sigma)=\{i: x_{\sigma_i}^{-1}<x_{\sigma_j}^{-1}, 1\leq i<j\}$ is the set of the indices of the presentation order σ for which the already sorted objects $\mathcal{O}_{\sigma_1},\ldots,\mathcal{O}_{\sigma_{j-1}}$ are ranked in x before the current object $\mathcal{O}_{\sigma_j}$, and consequently *on its left*. Its cardinal $\#j^-(x,\sigma)$ is consequently the number of *all* comparisons of the current object with the objects already ranked (according to x) on its left, if they exist.
- $j^+(x,\sigma)=\{i: i=\arg\min_{1\leq i'<j}\{i': x_{\sigma_{i'}}^{-1}>x_{\sigma_j}^{-1}\}\}$ is the index of the rank σ designating the object sorted in x just after (so *on the right* of) $\mathcal{O}_{\sigma_j}$ among the already sorted objects $\mathcal{O}_{\sigma_1},\ldots,\mathcal{O}_{\sigma_{j-1}}$. This set has at most one element. Its cardinal $\#j^+(x,\sigma)$ indicates that the current object $\mathcal{O}_{\sigma_j}$ has been compared with the object ranked (according to x) just *on its right*, if it exists.
- $\eta_j^-(x,\sigma,\mu)=\sum_{i\in j^-(x,\sigma)}\delta_{\sigma_i\sigma_j}(\mu)$ is the number of *good* comparisons (according to μ) of the current object $\mathcal{O}_{\sigma_j}$ with the objects already ranked *on its left*, if they exist.
- $\eta_j^+(x,\sigma,\mu)=\sum_{i\in j^+(x,\sigma)}\delta_{\sigma_j\sigma_i}(\mu)$ is the indicator of *good* comparison (according to μ) of the current object $\mathcal{O}_{\sigma_j}$ with the object already ranked just *on its right*, if it exists.

TAB. 1 – An example to illustrate both the notations and the insertion sorting process with $\mu=(1,2,3)$, $\sigma=(1,3,2)$, and $x=(3,1,2)$. The notation $x^{(j)}$, defined in Appendix 1, means the ranking of the j first objects in σ in the order imposed by x.

step j	$j^-(x,\sigma)$	$\#j^-(x,\sigma)$	$j^+(x,\sigma)$	$\#j^+(x,\sigma)$	$\eta_j^-(x,\sigma,\mu)$	$\eta_j^+(x,\sigma,\mu)$	$x^{(j)}$
1	{}	0	{}	0	0	0	(1)
2	{}	0	{1}	1	0	0	(3,1)
3	{3,1}	2	{}	0	1	0	(3,1,2)

With these notations, the probability to obtain a rank x from a initial presentation order σ is :

$$pr(x|\sigma;\mu,p)=\prod_{j=1}^{m} p^{\eta_j^-(x,\sigma,\mu)}(1-p)^{\#j^-(x,\sigma)-\eta_j^-(x,\sigma,\mu)} p^{\eta_j^+(x,\sigma,\mu)}(1-p)^{\#j^+(x,\sigma)-\eta_j^+(x,\sigma,\mu)}. \quad (1)$$

The proof of this formula is given in Appendix 1. The first term $p^{\eta_j^-(x,\sigma,\mu)}(1-p)^{\#j^-(x,\sigma)-\eta_j^-(x,\sigma,\mu)}$ corresponds to the probability of shifting $\#j^-(x,\sigma)$ times to the right the object $\mathcal{O}_{\sigma_j}$ coming at the step j, and the second term $p^{\eta_j^+(x,\sigma,\mu)}(1-p)^{\#j^+(x,\sigma)-\eta_j^+(x,\sigma,\mu)}$ is the probability for this object of being no longer shifted to the right. Finally, if the presentation order is unknown but of probability $pr(\sigma)$, the marginal distribution of x is given by :

$$pr(x;\mu,p)=\sum_{\sigma\in\mathcal{P}_m} pr(x|\sigma;\mu,p)pr(\sigma). \quad (2)$$

In this paper, we assume the presentation orders are uniformly distributed, and then $pr(\sigma)=1/m!$ for all $\sigma\in\mathcal{P}_m$. In the following the rank data model defined by Distribution (2) will be named ISR for Insertion Sorting Rank data model. We will note shortly ISR(μ,p) this model and its associated parameters.

4 Properties of the ISR model

In this section the main properties of the ISR model are stated : The possibility for the ISR distribution to be uniform for a special value of p, the existence of modal and anti-modal ranks, the

symmetry of the ISR distribution, and finally its identifiability. The proofs relie on applying permutation properties on both ranking and ordering notations on $\mathcal{P}_m$. In the following, the composition $\tau \circ \sigma$ is noted shortly $\tau\sigma$ and the set $\{\sigma^1, \ldots, \sigma^{m!}\}$ describes all possible ranks in $\mathcal{P}_m$.

4.1 Uniformity of the ISR distribution for $p = 1/2$

Proposition 1 proves the uniformity for $p = 1/2$, and requires Lemma 8 of Appendix 2.

Proposition 1. *If $p = 1/2$, for all $x, \mu \in \mathcal{P}_m$ then $pr(x; \mu, 1/2) = 1/m!$.*

Démonstration. Let σ^1 be a given permutation of $\mathcal{P}_m$, and for any permutation σ^s of $\mathcal{P}_m$, let τ^s be the only permutation such that : $\sigma^1 = \tau^s\sigma^s$ $(s = 1, \ldots, m!)$. The probability of x according to ISR$(\mu, 1/2)$ is :

$$\begin{aligned} pr(x; \mu, 1/2) &= \frac{1}{m!}\sum_{s=1}^{m!} pr(x|\sigma^s; \mu, 1/2) \\ &= \frac{1}{m!}\sum_{s=1}^{m!} pr(\tau^s x|\tau^s\sigma^s; \mu, 1/2) \quad \text{(Lemma 8)} \\ &= \frac{1}{m!}\sum_{s=1}^{m!} pr(\tau^s x|\sigma^1; \mu, 1/2). \end{aligned}$$

The proof is concluded by noting that $\sum_{s=1}^{m!} pr(\tau^s x|\sigma^1; \mu, 1/2) = 1$ because $p(.|\sigma^1; \mu, 1/2)$ is a probability distribution on $\mathcal{P}_m$ and $\{\tau^s x : \tau^s \in \mathcal{P}_m\} = \mathcal{P}_m$. □

4.2 Mode and anti-mode of the ISR distribution

We prove in this section one of the most important properties which can be expected from the ISR distribution : The reference rank μ is the unique mode of the distribution if $p > 1/2$ (Proposition 2). Let $\bar{\mu}$ be defined by $\bar{\mu} = \mu\rho$ where $\rho = (m, \ldots, 1)$ is the permutation of total inversion. This rank $\bar{\mu}$ is the furthest from μ for the Kendall distance. We symetrically prove in this section that the unique anti-mode (the rank of smallest probability) is $\bar{\mu}$ if $p > 1/2$ (Corollary 3). Proofs require Lemmas 6 and 7 of Appendix 1.

Proposition 2. *If $p > 1/2$, for all $x, \mu \in \mathcal{P}_m$, $x \neq \mu$, we have $pr(\mu; \mu, p) > pr(x; \mu, p)$.*

Démonstration. Let τ be the only permutation such that $\mu = \tau x$. The probability of x according to ISR(μ, p) is :

$$\begin{aligned} pr(x; \mu, p) &= \frac{1}{m!}\sum_{s=1}^{m!} p^{\sum_{j=1}^m \eta_j^-(x,\sigma^s,\mu)+\eta_j^+(x,\sigma^s,\mu)} \\ &\qquad (1-p)^{\sum_{j=1}^m \#j^-(x,\sigma^s)+\#j^+(x,\sigma^s)-\eta_j^-(x,\sigma^s,\mu)-\eta_j^+(x,\sigma^s,\mu)} \\ &< \frac{1}{m!}\sum_{s=1}^{m!} p^{\sum_{j=1}^m \#j^-(x,\sigma^s)+\#j^+(x,\sigma^s)} \\ &\qquad \text{(because } p > 1/2 \Leftrightarrow p > 1-p \text{ and } x \neq \mu) \\ &= \frac{1}{m!}\sum_{s=1}^{m!} p^{\sum_{j=1}^m \#j^-(\tau x,\tau\sigma^s)+\#j^+(\tau x,\tau\sigma^s)} \quad \text{(from Lemmas 6 et 7)} \\ &= \frac{1}{m!}\sum_{s=1}^{m!} p^{\sum_{j=1}^m \#j^-(\mu,\sigma'^s)+\#j^+(\mu,\sigma'^s)} \quad (\text{with } \sigma'^s = \tau\sigma^s) \\ &= pr(\mu; \mu, p). \end{aligned}$$

The last line comes from the fact that σ'^s and σ^s are in bijection, $\#j^-(\mu, \sigma^s) = \eta_j^-(\mu, \sigma^s, \mu)$ and also $\#j^+(\mu, \sigma^s) = \eta_j^+(\mu, \sigma^s, \mu)$. □

Corollary 3. *If $p > 1/2$, for all $x, \mu \in \mathcal{P}_m$, $x \neq \bar{\mu}$, we have $pr(\bar{\mu}; \mu, p) < pr(x; \mu, p)$.*

The proof, symmetrical to that of Proposition 2, is left to the reader.

4.3 Symmetry of the ISR distribution

In this section the symmetry of the ISR distribution is proved. The sense of this symmetry is the following : If the judge sorts the objects according to the ISR distribution with parameters (μ, p), the same sorting will be obtain (in distribution) with parameters $(\bar{\mu}, 1-p)$ (Proposition 4 below). This property will be especially useful in order to exhibit the identifiability conditions of the ISR distribution in the next section. Proposition 4 requires Lemmas 9 and 10 in Appendix 2.

Proposition 4. *For all $x, \mu \in \mathcal{P}_m$, we have $pr(x; \bar{\mu}, 1-p) = pr(x; \mu, p)$.*

Démonstration. The probability of x according to ISR$(\bar{\mu}, 1-p)$ is :

$$\begin{aligned}
pr(x;\bar{\mu},1-p) &= \frac{1}{m!}\sum_{s=1}^{m!}(1-p)^{\sum_{j=1}^{m}\eta_j^-(x,\sigma^s,\bar{\mu})+\eta_j^+(x,\sigma^s,\bar{\mu})} \\
&\qquad \times p^{\sum_{j=1}^{m}\#j^-(x,\sigma^s)+\#j^+(x,\sigma^s)-\eta_j^-(x,\sigma^s,\bar{\mu})-\eta_j^+(x,\sigma^s,\bar{\mu})} \\
&= \frac{1}{m!}\sum_{s=1}^{m!}(1-p)^{\sum_{j=1}^{m}\#j^-(x,\sigma^s)-\eta_j^-(x,\sigma^s,\mu)+\#j^+(x,\sigma^s)-\eta_j^+(x,\sigma^s,\bar{\mu})} \\
&\qquad \times p^{\sum_{j=1}^{m}\#j^-(x,\sigma^s)+\#j^+(x,\sigma^s)} \\
&\qquad \times p^{\sum_{j=1}^{m}-(\#j^-(x,\sigma^s)-\eta_j^-(x,\sigma^s,\mu))-(\#j^+(x,\sigma^s)-\eta_j^+(x,\sigma^s,\mu))} \\
&= \frac{1}{m!}\sum_{s=1}^{m!}(1-p)^{\sum_{j=1}^{m}\#j^-(x,\sigma^s)-\eta_j^-(x,\sigma^s,\mu)+\#j^+(x,\sigma^s)-\eta_j^+(x,\sigma^s,\bar{\mu})} \\
&\qquad \times p^{\sum_{j=1}^{m}\eta_j^-(x,\sigma^s,\mu)+\eta_j^+(x,\sigma^s,\mu)} \\
&= pr(x;\mu,p).
\end{aligned}$$

□

4.4 Identifiability of the ISR distribution

A necessary identifiability condition is immediatly suggested by Propositions 1 and 4 : The uniformity for $p = 1/2$ of the ISR distribution and its symmetry lead to impose $p > 1/2$. The sufficiency of this condition is proved in the next proposition. Its proof needs Lemma 11 of Appendix 2.

Proposition 5. *The ISR distribution is identifiable since $p > 1/2$.*

Démonstration. As the ISR model has two parameters, the probability p and the reference rank μ, the identifiability problem can concern only one of these two parameters or both.

- Firstly there exists none couple $(\mu, \mu') \in \mathcal{P}_m^2$ with $\mu \neq \mu'$ such that $pr(x; \mu, p) = pr(x; \mu', p)$ for any $x \in \mathcal{P}_m$ and any $p > 1/2$. Indeed, choosing $x = \mu$, from Lemma 11 we have $pr(\mu; \mu, p) \neq pr(\mu; \mu', p)$.
- Secondly, for a given $\mu \in \mathcal{P}_m$, assume there exists $p \neq p'$ such that $pr(x; \mu, p) = pr(x; \mu, p')$ for any $x \in \mathcal{P}_m$. In particular, for $x = \mu$, Equation (9) in the proof of Lemma 11 leads to

$$\frac{1}{m!}\sum_{s=1}^{m!}p^{\sum_{j=1}^{m}\#j^-(\mu,\sigma^s)+\#j^+(\mu,\sigma^s)} = \frac{1}{m!}\sum_{s=1}^{m!}p'^{\sum_{j=1}^{m}\#j^-(\mu,\sigma^s)+\#j^+(\mu,\sigma^s)}. \tag{3}$$

  The strict increasing of the function $p \mapsto p^n$ on the interval $[\frac{1}{2}, 1]$ for all $n \in \mathbb{N}^*$ ensures that $p = p'$.
- Assume finally there exists $(\mu, \mu') \in \mathcal{P}_m^2$ with $\mu \neq \mu'$ and $p < p'$ such that $p(x; \mu, p) = p(x; \mu', p')$ for any $x \in \mathcal{P}_m$. From Equations (9) and (10) in the proof of Lemma 11 we have with $x \neq \mu$

$$\begin{aligned}
pr(x|\sigma;\mu,p) &< p^{\sum_{j=1}^{m}\#j^-(x,\sigma)+\#j^+(x,\sigma)} \\
&< p'^{\sum_{j=1}^{m}\#j^-(x,\sigma)+\#j^+(x,\sigma)} \\
&= pr(x|\sigma;x,p')
\end{aligned}$$

and then by averaging over all σ in $\mathcal{P}_m$

$$pr(x;\mu,p) < pr(x;x,p').$$

Choosing $x = \mu'$ ensures the identifiability of the ISR model.

□

5 Estimation of the model parameters

The ISR model for rank data has two parameters : The probability p, which is a real in $[1/2, 1]$ and the reference rank, or modal rank, μ, which can take its values in $\mathcal{P}_m$. Note that the case $p = 1/2$ is kept although this is a non-identifiability situation because it leads to the uniformity of the ISR distribution, what can be of interest for practical applications. We present in this section maximum likelihood estimation.
Let $(x^1, \ldots, x^n)$ be a sample of n ranks. The log-likelihood of the ISR model is :

$$l(\theta) = \sum_{i=1}^{n} \ln \left(\frac{1}{m!} \sum_{s=1}^{m!} pr(x^i|\sigma^s;\mu,p) \right), \tag{4}$$

with $\theta = (\mu, p)$. As the presentation orders σ are unknown, we use the EM algorithm [4] to maximize this oberved data log-likelihood. The completed log-likelihood is :

$$l_c(\theta) = \frac{1}{m!} \sum_{i=1}^{n} \sum_{s=1}^{m!} \zeta_{is} \ln \left(pr(x^i|\sigma^s;\mu,p) \right), \tag{5}$$

where ζ_{is} is a random variable equal to 1 is the rank x^i is the result of a sorting with σ^s as presentation order, 0 otherwise. The EM algorithm is an iterative procedure composed of two steps, which intends to maximize this completed log-likelihood. Let $\theta^{(q)} = (\mu^{(q)}, p^{(q)})$ be a current value of the parameters ($q \in \mathbb{N}$), $\theta^{(0)}$ being the starting parameter of EM.
The E step consists in computing the conditional expectation of the completed log-likelihood :

$$\mathcal{Q}(\theta, \theta^{(q)}) = E_{\theta^{(q)}}[l_c(\theta)|x^1, \ldots, x^n] = \frac{1}{m!} \sum_{i=1}^{n} \sum_{s=1}^{m!} \tau_{is}^{(q)} \ln \left(pr(x^i|\sigma^s;\mu,p) \right)$$

where

$$\tau_{is}^{(q)} = E_{\theta^{(q)}}[\zeta_{is}|x^1, \ldots, x^n] = \frac{pr(x^i|\sigma^s;\mu^{(q)},p^{(q)})}{\sum_{r=1}^{m!} pr(x^i|\sigma^r;\mu^{(q)},p^{(q)})}$$

is the conditional probability for the rank i to be the result of a sorting with σ^s as presentation order.
The M step of the EM algorithm consists in choosing the value $\theta^{(q+1)}$ which maximizes the conditional expectation $\mathcal{Q}$ computed at the E step :

$$\theta^{(q+1)} = \underset{\theta \in \Theta}{\operatorname{argmax}}\, \mathcal{Q}(\theta; \theta^{(q)})$$

where Θ is the parameter space $\mathcal{P}_m \times [1/2, 1]$. As the parameter space $\mathcal{P}_m$ for μ is discrete, the maximization consists simply, but potentially computationally expensively, of browsing the entire $\mathcal{P}_m$. However in practice, thanks to the symmetry of the ISR distribution, it is enough to browse the half of $\mathcal{P}_m$. For the probability p, maximizing $\mathcal{Q}$ leads to the following maximum :

$$p^{(q+1)} = \frac{\sum_{i=1}^{n} \sum_{s=1}^{m!} \tau_{is}^{(q)} \sum_{j=1}^{m} \eta_j^-(x^i, \sigma^s, \mu^{(q)}) + \eta_j^+(x^i, \sigma^s, \mu^{(q)})}{\sum_{i=1}^{n} \sum_{s=1}^{m!} \tau_{is}^{(q)} \sum_{j=1}^{m} \#j^-(x^i, \sigma^s) + \#j^+(x^i, \sigma^s)}.$$

Note that this value of $p^{(q+1)}$ can be interpreted as the proportion of good manipulations (switching to the right or stop) in the insertion sorting algorithm.

6 Numerical illustration

6.1 Visualising the ISR distribution

Likening a ranking $x^{-1} = (x_1^{-1}, \ldots, x_m^{-1})$ to an element of $\mathbb{R}^m$, the representation of all ranks on a polytope lives only in a subspace of dimension $m-1$, since the knowledge of the $m-1$ first components $x_1^{-1}, \ldots, x_{m-1}^{-1}$ imposes the last one x_m^{-1}. A polytope representation of the distribution [16, 17] is then interesting when the number m of objects to rank is lower or equal to 4. Conventional representation of these polytopes links ranks of Kendall distance equal to 1 [14]. Figure 1 shows two generic permutation polytopes for $m = 4$ and $p = 0.6$ or $p = 0.9$, where orderings are displayed on each node. On these polytopes, the points surfaces are proportional to the ISR rank probabilities. They illustrate in particular that the decrease of the rank probability is even stronger as p increases.

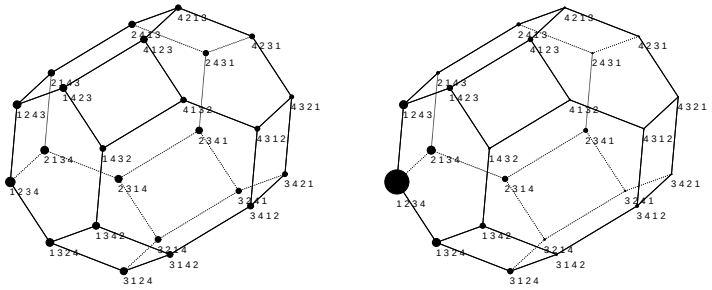

FIG. 1 – Polytope representation of the ranks ($m = 4$), with size of the points proportional to the ISR probability : $\mu = (1, 2, 3, 4)$, $p = 0.6$ (left) and $p = 0.9$ (right).

6.2 Estimation of the ISR distribution on real data sets

In order to assess the adequacy of the ISR distribution to a real data set, we submitted the following quiz to our students, consisting of three items Q_1, Q_2 and Q_3 of ascending difficulty :

- Q_1. Rank the following numbers in ascending order :
$$\mathcal{O}_1 = \pi/3,\ \mathcal{O}_2 = \log 1,\ \mathcal{O}_3 = \exp 2,\ \mathcal{O}_4 = \tfrac{1+\sqrt{5}}{2}.$$
- Q_2. Rank the following French writers in chronological order of birth :
$\mathcal{O}_1$ = Victor Hugo, $\mathcal{O}_2$ = Molière, $\mathcal{O}_3$ = Albert Camus, $\mathcal{O}_4$ = Jean-Jacques Rousseau.
- Q_3. Rank chronologically these Quentin Tarantino movies :
$\mathcal{O}_1$ = Inglourious Basterds, $\mathcal{O}_2$ = Pulp Fiction, $\mathcal{O}_3$ = Reservoir Dogs, $\mathcal{O}_4$ = Jackie Brown.

The correct answers are $\mu^* = (2, 1, 4, 3)$ for Q_1, $\mu^* = (2, 4, 1, 3)$ for Q_2 and $\mu^* = (3, 2, 4, 1)$ for Q_3. The answers of the 40 questioned students are in turn in Table 2.

For each item of the quiz, the ISR distribution is estimated and a χ^2 adequacy test, where the distribution under the null assumption is estimated by bootstrap [5] based on 1000 replications, is performed :

- Q_1. $\hat{\mu} = (2, 1, 4, 3)$, $\hat{p} = 0.962$ and p-value $= 0.593$,
- Q_2. $\hat{\mu} = (2, 4, 1, 3)$, $\hat{p} = 0.815$ and p-value $= 0.342$,
- Q_3. $\hat{\mu} = (4, 3, 2, 1)$, $\hat{p} = 0.754$ and p-value $= 0.264$.

We can first note that the adequacy of the ISR distribution is accepted for these three questions. This adequacy can be also found graphically on Figures 2, 3 and 4 displaying polytopes (orderings are displayed on each node) of both the empirical distribution and the ISR estimated one. We remark also the decrease of the number of good answers when one move away from the modal rank.

The growth of the questions difficulty is reflected by a decrease in the probability p : For the easy first question, 80% of the students gives the right answer and 15% makes one mistake by reversing $\mathcal{O}_1$ and $\mathcal{O}_4$. This leads to a high value of the probability p : $\hat{p} = 0.962$. For the second

TAB. 2 – Quiz answers of the 40 students.

Quiz Q_1		Quiz Q_2		Quiz Q_3	
ordering	frequency	ordering	frequency	ordering	frequency
$(2,1,4,3)$	32	$(2,4,1,3)$	15	$(4,3,2,1)$	10
$(2,4,1,3)$	6	$(2,4,3,1)$	8	$(4,2,3,1)$	9
$(2,1,3,4)$	2	$(2,1,4,3)$	4	$(3,2,4,1)$	4
other	0	$(4,2,1,3)$	4	$(3,4,2,1)$	3
		$(2,3,1,4)$	2	$(1,3,2,4)$	2
		$(1,2,3,4)$	1	$(1,3,4,2)$	2
		$(1,3,4,2)$	1	$(2,3,1,4)$	2
		$(1,4,2,3)$	1	$(3,1,4,2)$	2
		$(2,1,3,4)$	1	$(1,2,3,4)$	1
		$(2,3,4,1)$	1	$(2,3,4,1)$	1
		$(3,1,4,2)$	1	$(2,4,3,1)$	1
		$(3,2,1,4)$	1	$(3,2,1,4)$	1
		other	0	$(4,1,2,3)$	1
				$(4,3,1,2)$	1
				other	0

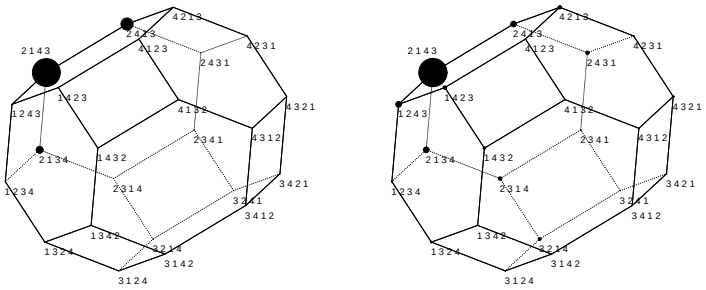

FIG. 2 – Empirical (left) and estimated (right) distributions for quiz Q_1, related to numbers.

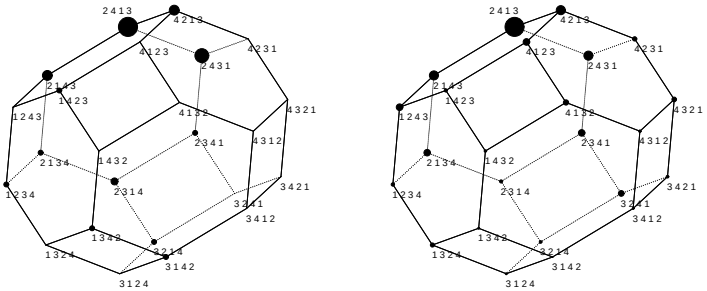

FIG. 3 – Empirical (left) and estimated (right) distributions for quiz Q_2, related to French writers.

question, only 37.5% of the students gives the right answer, and the number of wrong answers decreases gradually with the number of bad comparisons made in the ranking process. Finally, the last question leads to more mixed answers with a smaller decrease in the number of responses gradually as the distance from the modal rank $\hat{\mu} = (4,3,2,1)$, which moreover is different from the right rank $\mu^* = (3,2,4,1)$.

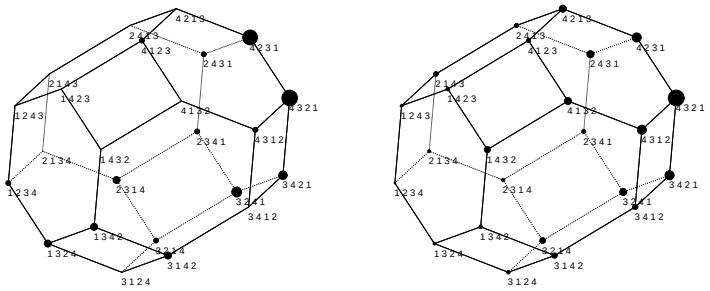

FIG. 4 – Empirical (left) and estimated (right) distributions for quiz Q_3, related to Quentin Tarantino movies.

7 Discussion

In this paper we suggest a probability distribution for rank data, modeling a stochastic version of the insertion sorting algorithm. The main force of the ISR distribution consists in the naturalness and the optimality of this sorting algorithm for a moderate number m of objects to rank. In this sense, we can expect in many cases a higher modeling power of the ISR model than usual rank data models which can be interpreted as the modeling of poorly performing sorting algorithms. Obviously, this claim relies on the assumption that the judge is somewhat *optimal*. Moreover, the ISR model allows to take into account, if it is known, the presentation order of the objects to rank, what could be particulary interesting since this last could influence the ranking in such situations. Two other benefits to consider such a generative model is that it allows an interpretation of the ranking results *via* its parameters μ (the modal ranking) and p (the probability of good paired comparison during the sorting), and leads to easy maximum likelihood estimation.
Perspectives of this work are numerous. First ones concern computational aspects : When m is reasonable (smaller than 10), the interest of the ISR distribution has been previously underlined. For higher number of objects to rank, the insertion sorting algorithm is no more the most powerful, and other algorithms should be considered. However, as for usual rank data models, ISR estimation needs to browse all the possible reference ranks μ, what becomes computationally infeasable when m grows beyond 10. In addition, specific rank data models can be also defined by modeling specific sorting algorithm according to the knowledge on the sorting process employed by the judges.
Second perspectives concern the application aspects : In this paper the ISR distribution is successfully used to model students answers to a quiz, but mixture of ISR distributions could also be of great interest for modeling more complex situations, typically multimodality. For instance, assume we have put in our quiz a football question, then girls and boys responses will probably not follow the same distribution.
A third axis of perspective looks on the nature of the rank data : Here only full rank data for single judgment criterion have been considered. This approach needs to be extended to other types of ranks, frequently encountered in practice, as partially ranked data, tied data or even ranks resulting from multiple preference responses.

Appendix 1

Proof of ISR distribution for a known presentation order

The goal of this appendix is to prove that Formula (1) corresponds to the stochastic left insertion sorting algorithm with probability p of good paired comparison and independence between the paired comparisons. The notation are those defined in Section 3.

Démonstration. Let $x^{(j)} = (x_i : x_i \in \{\sigma_1, \ldots, \sigma_j\})$ be the ranking obtained at the step j ($1 \leq j \leq m$) : It is the ranking of the j first objects in σ in the order imposed by x (thus $x^{(m)} = x$). Note

that there exists the following relationship between $x^{(j)}$ and $x^{(j-1)}$:

$$x^{(j)} = (x_1^{(j-1)}, \ldots, x_{\#j^-(x,\sigma)}^{(j-1)}, \sigma_j, x_{\#j^-(x,\sigma)+1}^{(j-1)}, \ldots, x_{j-1}^{(j-1)}).$$

Formula (1) will be proved by induction on j. It is true for $j = 1$ while there is only one object σ_1 to sort : $pr(x^{(1)}|\sigma) = 1$. The result of the ranking $x^{(j)}$ from $x^{(j-1)}$ is the result of $\#j^-(x,\sigma)+\#j^+(x,\sigma)$ independent Bernoulli experiments of parameter p. Conditionally to $x^{(j-1)}$, the probability of $x^{(j)}$ is then

$$\begin{aligned} pr(x^{(j)}|x^{(j-1)}, \sigma; \mu, p) = \quad & p^{\eta_j^-(x,\sigma,\mu)+\eta_j^+(x,\sigma,\mu)} \\ & (1-p)^{\#j^-(x,\sigma)+\#j^+(x,\sigma)-\eta_j^-(x,\sigma,\mu)-\eta_j^+(x,\sigma,\mu)}. \end{aligned}$$

We conclude the proof by remarking that

$$pr(x^{(j)}|\sigma; \mu, p) = pr(x^{(j)}|x^{(j-1)}, \sigma; \mu, p) pr(x^{(j-1)}|\sigma; \mu, p),$$

because we have the following implication relationship between events : $x^{(j)} \Rightarrow x^{(j-1)}$. □

Appendix 2

Lemma 6. *For all permutations $x, \sigma, \tau \in \mathcal{P}_m$, we have $\#j^-(x,\sigma) = \#j^-(\tau x, \tau\sigma)$.*

Démonstration. For any $j = 1, \ldots, m$

$$\begin{aligned} \#j^-(\tau x, \tau\sigma) &= \#\{i : (\tau x)^{-1}_{(\tau\sigma)_i} < (\tau x)^{-1}_{(\tau\sigma)_j}, 1 \le i < j\} \\ &= \#\{i : (\tau x)^{-1}(\tau\sigma)_i < (\tau x)^{-1}(\tau\sigma)_j, 1 \le i < j\} \\ &= \#\{i : (x^{-1}\tau^{-1}\tau\sigma)_i < (x^{-1}\tau^{-1}\tau\sigma)_j, 1 \le i < j\} \\ &= \#\{i : (x^{-1}\sigma)_i < (x^{-1}\sigma)_j, 1 \le i < j\} \\ &= \#\{i : x^{-1}_{\sigma_i} < x^{-1}_{\sigma_j}, 1 \le i < j\} \\ &= \#j^-(x,\sigma). \end{aligned}$$

□

Lemma 7. *For all permutations $x, \sigma, \tau \in \mathcal{P}_m$, we have $\#j^+(x,\sigma) = \#j^+(\tau x, \tau\sigma)$.*

Démonstration. It suffices to note that $\#j^+(x,\sigma) = \mathbf{1}\{\#j^-(x,\sigma) + 1 \le j - 1\}$ and to use Lemma 6 to conclude. □

Lemma 8. *If $p = 1/2$ then for all permutations $x, \tau \in \mathcal{P}_m$, we have $pr(x|\sigma; \mu, 1/2) = pr(\tau x|\tau\sigma; \mu, 1/2)$.*

Démonstration. When $p = 1/2$, we obtain by using Lemmas 6 and 7

$$\begin{aligned} pr(\tau x|\tau\sigma; \mu, 1/2) &= \left(\frac{1}{2}\right)^{\sum_{j=1}^m \#j^-(\tau x,\tau\sigma)+\#j^+(\tau x,\tau\sigma)} \\ &= \left(\frac{1}{2}\right)^{\sum_{j=1}^m \#j^-(x,\sigma)+\#j^+(x,\sigma)} \\ &= pr(x|\sigma; \mu, 1/2). \end{aligned}$$

□

Lemma 9. *For all $x, \sigma, \mu \in \mathcal{P}_m$ we have $\eta_j^-(x, \sigma, \bar{\mu}) = \#j^-(x,\sigma) - \eta_j^-(x, \sigma, \mu)$.*

Démonstration. Let ρ be the permutation of total inversion previously introduced in Section 4.2 and $i, i' = 1, \ldots, m, i \neq i'$. We have :

$$\begin{aligned}
\delta_{ii'}(\bar{\mu}) &= \mathbf{1}\{\bar{\mu}_i^{-1} < \bar{\mu}_{i'}^{-1}\} \\
&= \mathbf{1}\{(\mu\rho)_i^{-1} < (\mu\rho)_{i'}^{-1}\} \quad \text{(because } \bar{\mu} = \mu\rho\text{)} \\
&= \mathbf{1}\{\rho^{-1}\mu_i^{-1} < \rho^{-1}\mu_{i'}^{-1}\} \\
&= \mathbf{1}\{\rho\mu_i^{-1} < \rho\mu_{i'}^{-1}\} \quad \text{(since } \rho = \rho^{-1}\text{)} \\
&= \mathbf{1}\{\mu_i^{-1} > \mu_{i'}^{-1}\} \quad \text{(since } i < i' \Leftrightarrow \rho_i > \rho_{i'}\text{)} \\
&= 1 - \mathbf{1}\{\mu_i^{-1} < \mu_{i'}^{-1}\} \quad \text{(since } i \neq i'\text{)} \\
&= 1 - \delta_{ii'}(\mu).
\end{aligned}$$

This proof is concluded by :

$$\begin{aligned}
\eta_j^-(x, \sigma, \bar{\mu}) &= \sum_{i \in j^-(x,\sigma)} \delta_{\sigma_i \sigma_j}(\bar{\mu}) \\
&= \sum_{i \in j^-(x,\sigma)} (1 - \delta_{\sigma_i \sigma_j}(\mu)) \\
&= \#j^-(x, \sigma) - \sum_{i \in j^-(x,\sigma)} \delta_{\sigma_i \sigma_j}(\mu) \\
&= \#j^-(x, \sigma) - \eta^-(x, \sigma, \mu).
\end{aligned}$$

□

Lemma 10. *For all $x, \sigma, \mu \in \mathcal{P}_m$ we have $\eta_j^+(x, \sigma, \bar{\mu}) = \#j^+(x, \sigma) - \eta_j^+(x, \sigma, \mu)$.*

The proof is similar to that of Lemma 9.

Lemma 11. *For all $x, \mu \in \mathcal{P}_m$, $x \neq \mu$ and $p > 1/2$, we have $pr(x; \mu, p) < pr(x; x, p)$.*

Démonstration. Remark first that

$$\eta_j^-(x, \sigma, x) = \sum_{i \in j^-(x,\sigma)} \delta_{\sigma_i \sigma_j}(x) = \#j^-(x, \sigma) \tag{6}$$

$$\eta_j^+(x, \sigma, x) = \#j^+(x, \sigma) \tag{7}$$

because all the already sorted objects at the step j are necessarily well sorted. In addition, for any $\mu \neq x$ we have necessarily

$$\sum_{j=1}^{m} \eta_j^-(x, \sigma, \mu) + \eta_j^+(x, \sigma, \mu) < \sum_{j=1}^{m} \#j^-(x, \sigma) + \#j^+(x, \sigma) \tag{8}$$

Effectively, the equality occurs only if, at each step j of the sorting process, all the already sorted objects are well sorted, what is possible only for $\mu = x$. We have then

$$\begin{aligned}
pr(x|\sigma; \mu, p) &= p^{\sum_{j=1}^m \eta_j^-(x,\sigma,\mu) + \eta_j^+(x,\sigma,\mu)} \\
&\quad (1-p)^{\sum_{j=1}^m \#j^-(x,\sigma) + \#j^+(x,\sigma) - \eta_j^-(x,\sigma,\mu) - \eta_j^+(x,\sigma,\mu)} \\
&< p^{\sum_{j=1}^m \#j^-(x,\sigma) + \#j^+(x,\sigma)}
\end{aligned}$$

since $p > 1/2 \Leftrightarrow 1 - p < p$ and the exponent of $1 - p$ is positive as just seen before in (8). From Equations (6) and (7) we deduce

$$pr(x|\sigma; x, p) = p^{\sum_{j=1}^m \#j^-(x,\sigma) + \#j^+(x,\sigma)} \tag{9}$$

and finally

$$pr(x|\sigma; \mu, p) < pr(x|\sigma; x, p). \tag{10}$$

The proof is concluded by averaging over all possible presentation orders σ in $\mathcal{P}_m$. □

Références

[1] Ulf Böckenholt. Applications of Thurstonian models to ranking data. In *Probability models and statistical analyses for ranking data (Amherst, MA, 1990)*, volume 80 of *Lecture Notes in Statist.*, pages 157–172. Springer, New York, 1993.

[2] Ralph Allan Bradley and Milton E. Terry. Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika*, 39 :324–345, 1952.

[3] Douglas E. Critchlow. *Metric methods for analyzing partially ranked data*, volume 34 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin, 1985.

[4] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, 39(1) :1–38, 1977. With discussion.

[5] Bradley Efron and Robert J. Tibshirani. *An introduction to the bootstrap*, volume 57 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York, 1993.

[6] Paul D. Feigin and Mayer Alvo. Intergroup diversity and concordance for ranking data : an approach via metrics for permutations. *Ann. Statist.*, 14(2) :691–707, 1986.

[7] M. A. Fligner and J. S. Verducci. Distance based ranking models. *J. Roy. Statist. Soc. Ser. B*, 48(3) :359–369, 1986.

[8] Michael A. Fligner and Joseph S. Verducci. Multistage ranking models. *J. Amer. Statist. Assoc.*, 83(403) :892–901, 1988.

[9] M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30 :81–93, 1938.

[10] M. G. Kendall and B. Babington Smith. On the method of paired comparisons. *Biometrika*, 31 :324–345, 1940.

[11] D.E. Knuth. *Sorting and Searching : Volume3. The art of Computer Programming*. Addison-Wesley, Massachusetts, 1973.

[12] R. Duncan Luce. *Individual choice behavior : A theoretical analysis*. John Wiley & Sons Inc., New York, 1959.

[13] C. L. Mallows. Non-null ranking models. I. *Biometrika*, 44 :114–130, 1957.

[14] John I. Marden. *Analyzing and modeling rank data*, volume 64 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1995.

[15] R. L. Plackett. The analysis of permutations. *J. Roy. Statist. Soc. Ser. C Appl. Statist.*, 24(2) :193–202, 1975.

[16] G. L. Thompson. Generalized permutation polytopes and exploratory graphical methods for ranked data. *Ann. Statist.*, 21(3) :1401–1430, 1993.

[17] G. L. Thompson. *Probability models and statistical analyses for ranking data*, chapter Graphical techniques for ranked data, pages 294–298. Springer-Verlag, New-York, 1993.

[18] L.L. Thurstone. A law of comparative judgment. *Psychological Review*, 79 :281–299, 1927.