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

Christophe Biernacki, Julien Jacques

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

Christophe BIERNACKI [a], Julien JACQUES [a]

Abstract

Rank data arise from a sorting mechanism which is generally unobservable for the statistician. Assuming that this process relies on paired comparisons, the insertion sorting algorithm is known as being the best candidate for minimizing the number of potential paired misclassifications. Combining this optimality argument with a Bernoulli event at the paired comparison step, an original and very meaningful probabilistic generative model for rank data is obtained: This model is the first which takes into account the initial presentation order. Its theoretical properties are studied among which unimodality, symmetry and identifiability. In addition, maximum likelihood principle can be easily performed through an EM algorithm thanks to an unobserved latent variables interpretation. Finally, the practical relevance of the proposal is illustrated by both its adequacy with several real datasets and a comparison with a usual rank data model. In particular, its specific ability for revealing some fundamental hidden structure in the data mechanism generation is underlined.

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

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 respect this structure have been proposed, for instance the permutation polytope for plotting the rank vectors in Euclidean space ([Tho93a, Tho93b] and an example on Figure 1) or also suitable distances for defining the centre and

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

spread of a dataset [Ken38, Mal57, FV86].
From an inference point of view, distances are useful for testing the distribution of these data (uniformity, populations comparison [FA86, Mar95]) or for modeling the distribution itself (for instance the Mallows Φ model relies on the Kendall distance [Ken38, Cri85]). 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 basic 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 is the result of a sorting process, we make the natural assumption that it relies on successive object paired comparisons. In this case, potential errors in the final ranking (according to a reference rank) are only a consequence of some erroneous comparisons and, so, an optimal sorting strategy should minimize its total number of paired comparisons. Adopting the insertion sorting algorithm for this reason, we obtain a new kind of generative model which enjoys good theoretical properties and whose originality is to involve the initial presentation order of the objects.

The paper is organized as follows. 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 and builds 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. Numerical illustrations are presented in Section 6 to evaluate the relevance of the proposed model on real data sets both from a distributional adequacy point of view and from a comparison with the usual Mallows Φ model. Since this work sheds a new light on rank data modeling, numerous related perspectives are discussed in the last section (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 $\mathcal{O}_1, \ldots, \mathcal{O}_m$ by a judge (human or not). Two representations of these data are commonly used: Ranking or ordering. The *ranking* representation $x^{-1} = (x_1^{-1}, \ldots, x_m^{-1})$ contains the ranks given to the objects, and means that $\mathcal{O}_i$ is in the x_i^{-1}th position ($i = 1, \ldots, m$). A ranking is then an element of $\mathcal{P}$, the set of permutations of the m first integers. The *ordering* representation $x = (x_1, \ldots, x_m)$ is also an element of $\mathcal{P}$ and signifies that Object $\mathcal{O}_{x_i}$ is the ith ($i = 1, \ldots, m$). 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 [Mar95, Chap. 5 to 10]. The first class is derived from a paired comparison process [KS40]: 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:

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

with p_{ij} the probability that $\mathcal{O}_{x_i}$ is prefered to $\mathcal{O}_{x_j}$, 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. [BT52] associate to each object $\mathcal{O}_j$ a score u_j indicating an overall degree of preference of this object, and connect these scores to p_{ij} by $p_{ij} = u_i/(u_i + u_j)$, which defines the *Bradley–Terry–Mallows model*. [Mal57] goes forward into the simplification by imposing that $p_{ij} = \pi$ if and only if $\mu_{x_i}^{-1} < \mu_{x_j}^{-1}$ (p_{ij} only depends on the sign of $\mu_{x_i}^{-1} - \mu_{x_j}^{-1}$), where μ is a "reference" rank. It leads after reparameterization to the famous *Mallows Φ model*:

$$\mathrm{p}(x; \mu, \lambda) = \mathcal{C}(\lambda)^{-1} \exp^{-\lambda K(x,\mu)},$$

where K is the Kendall distance between two ranks [Ken38, Cri85], $\lambda = -\frac{1}{2} \ln \frac{\pi}{1-\pi}$ is a precision parameter ($\lambda \in \mathbb{R}$) and

$$\mathcal{C}(\lambda) = \prod_{j=1}^{m-1} \frac{1 - \exp(-(m-j+1)\lambda)}{1 - \exp(-\lambda)}$$

is a normalization constant [FV86]. For instance, a high λ value leads to strong unimodality around μ.

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 $\mathcal{O}_{x_i}$ is ranked first among the m objects, the corresponding *Plackett-Luce model* [Luc59, Pla75] defines the probability of a rank x as

$$\mathrm{p}(x) = \prod_{j=1}^{m-1} \frac{v_j}{v_j + v_{j+1} + \ldots + v_m}.$$

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 [Thu27, Böc93] with a Gumbel density. [FV86, FV88] introduce an alternative multistage model by considering another form of the probability at each step of the

ranking process. Let $V_j = \alpha$ if at the stage j the $(\alpha+1)$th best of the remaining objects is selected ($\alpha = 0, \ldots, m-j$), so $V_j = 0$ indicates a correct choice at stage j. The probability of a rank x according to the Fligner and Verducci's *strongly unimodal model* is:

$$\mathrm{p}(x) = \prod_{j=1}^{m-1} \mathrm{p}(V_j, j)$$

where $\mathrm{p}(\alpha, \beta)$ is a probability parameterized by α and β ($0 \leq \alpha \leq m-\beta$ and $1 \leq \beta \leq m-1$) satisfying $\sum_{\alpha=0}^{m-\beta} \mathrm{p}(\alpha, \beta) = 1$, the probability $\mathrm{p}(., \beta)$ being nonincreasing for all $1 \leq \beta \leq m-1$ and where $\mathrm{p}(0, \beta) > \mathrm{p}(1, \beta)$. Assuming specific forms for the probability $\mathrm{p}(\alpha, \beta)$ could lead to the Mallows Φ model or to a generalization of this latter named Φ component-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 used sorting algorithms 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 associated ranking process can be likened to a *selection* sorting algorithm. It is reasonnable to assume that it relies also on underlying paired comparisons even if it is not explicitly modeled in this way. Under this assumption, the selection sorting algorithm is one of the most simple but it is well known for its lack of optimality from the number of paired comparisons point of view [Knu73]. 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 $m \leq 10$ [Knu73, Chap. 5].

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

3.1 Motivation for an insertion sorting algorithm

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. Making also the natural assumption that a rank $x = (x_1, \ldots, x_m)$ is the result of a sorting process relying on successive object paired comparisons, any difference between the final rank x and μ is necessarily attributed to some incorrect paired comparisons. As a consequence, reducing the gap between x and μ is strongly correlated to minimizing the number of paired comparisons involved in the sorting process. Thus, an "optimal judge" should adopt the *insertion* sorting algorithm which is optimal for

a "reasonable" number of objects ($m \leq 10$) [Knu73, Chap. 5]. Since it is natural to model the reliability of the judge for the ranking by the risk of wrongly order a pair of objects, each paired comparison can be usefully interpreted as the result of a Bernoulli experiment whose outcome is a correct comparison (according to μ) with probability π and an incorrect comparison with probability $1 - \pi$. We assume also that each pair ranking operation is independent of the others and that the probability π is constant throughout the sorting process. Merging both deterministic insertion algorithm and random paired comparison leads to a meaningful generative model for rank data that is now presented at length.

Let the ordering $y = (y_1, \ldots, y_m)$ be the presentation order of the objects to the judge, this latter using the following insertion sorting algorithm to rank these objects. First, the current object to be sorted is placed on the left of the already sorted objects, and is compared to the first object 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 y 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.

3.2 Modeling of the resulting distribution

Based on this modeling of a stochastic insertion sorting, the question is now to calculate the probability $\mathrm{p}(x|y;\mu,\pi)$ to obtain a rank x from an initial presentation order y 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 in ranking the object $\mathcal{O}_{y_j}$. The notations and their use in the proposed sorting algorithm are both illustrated by an example in Table 1.

- $\delta_{ii'}(\mu) = \mathbf{1}\{\mu_i^{-1} < \mu_{i'}^{-1}\}$ is equal to 1 if $\mathcal{O}_i$ is correctly ranked before $\mathcal{O}_{i'}$ (according to μ), 0 otherwise ($i, i' = 1, \ldots, m, i \neq i'$).

- $\mathcal{A}_j^-(x,y) = \{i : x_{y_i}^{-1} < x_{y_j}^{-1}, 1 \leq i < j\}$ is the set of the indices of the presentation order y for which the already sorted objects $\mathcal{O}_{y_1}, \ldots, \mathcal{O}_{y_{j-1}}$ are ranked in x before the current object $\mathcal{O}_{y_j}$, and consequently *on its left*. Its cardinal $A_j^-(x,y)$ 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.

- $\mathcal{A}_j^+(x,y) = \{i : i = \arg\min_{1 \leq i' < j}\{i' : x_{y_{i'}}^{-1} > x_{y_j}^{-1}\}\}$ is the index of the rank y designating the object sorted in x just after (so *on the right* of) $\mathcal{O}_{y_j}$ among the already sorted objects $\mathcal{O}_{y_1}, \ldots, \mathcal{O}_{y_{j-1}}$, if it exists. This set has at most one element. Its cardinal $A_j^+(x,y)$ indicates if the current object $\mathcal{O}_{y_j}$ is compared, at the j step of the sorting, with the object ranked in x just *on its right*.

- $G_j^-(x,y,\mu) = \sum_{i \in \mathcal{A}_j^-(x,y)} \delta_{y_i y_j}(\mu)$ is the number of *good* comparisons (according to μ) of the current object $\mathcal{O}_{y_j}$ with the objects already ranked *on its left*, if they exist.

- $G_j^+(x, y, \mu) = \sum_{i \in \mathcal{A}_j^+(x,y)} \delta_{y_j y_i}(\mu)$ is the indicator of *good* comparison (according to μ) of the current object $\mathcal{O}_{y_j}$ with the object already ranked just *on its right*, if it exists.

We will use also intensively the following shorter and meaningful notations:

- $A_j(x, y) = A_j^-(x, y) + A_j^+(x, y)$ and $A(x, y) = \sum_{j=1}^m A_j(x, y)$ are the total number of *all* paired comparisons respectively for the step j and for the whole process.
- $G_j(x, y, \mu) = G_j^-(x, y, \mu) + G_j^+(x, y, \mu)$ and $G(x, y, \mu) = \sum_{j=1}^m G_j(x, y, \mu)$ are the total number of *good* paired comparisons respectively for the step j and for the whole process.

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

step j	$\mathcal{A}_j^-$	$\mathcal{A}_j^+$	A_j^-	A_j^+	A_j	G_j^-	G_j^+	G_j	$x^{(j)}$
1	$\{\}$	$\{\}$	0	0	0	0	0	0	(1)
2	$\{\}$	$\{1\}$	0	1	1	0	0	0	$(3, 1)$
3	$\{3, 1\}$	$\{\}$	2	0	2	1	0	1	$(3, 1, 2)$
					$A = 3$			$G = 1$	

With these notations, the probability to obtain a rank x from an initial presentation order y is given by:

$$\mathrm{p}(x|y; \mu, \pi) = \pi^{G(x,y,\mu)}(1 - \pi)^{A(x,y)-G(x,y,\mu)}. \tag{3.1}$$

The proof of this formula is given in Appendix A. The first term corresponds to the probability of performing $G(x, y, \mu)$ *good* paired comparisons and the second term is the probability of performing $A(x, y) - G(x, y, \mu)$ *wrong* paired comparisons. Finally, if the presentation order is unknown but of probability $\mathrm{p}(y)$, the marginal distribution of x is given by:

$$\mathrm{p}(x; \mu, \pi) = \sum_{y \in \mathcal{P}} \mathrm{p}(x|y; \mu, \pi)\mathrm{p}(y). \tag{3.2}$$

In this paper, we assume the presentation orders are uniformly distributed, and then $\mathrm{p}(y) = m!^{-1}$ for all $y \in \mathcal{P}$. In the following the rank data model defined by Distribution (3.2) will be named ISR for Insertion Sorting Rank data model. We will note shortly $\mathrm{ISR}(\mu, p)$ this model and its associated parameters.

Remark The conditional probability (3.1) is invariant to an inversion of the first two elements of the presentation order (Lemma B.1 of Appendix B). Consequently, the number $m!$ of presentation orders y to be considered in the calculus of the probability (3.2) may be reduced by half, what will be computationally helpful for the model parameters estimation.

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 π, the existence of modal and anti-modal ranks, the symmetry of the ISR distribution, and its identifiability. The proofs rely on applying permutation properties on both ranking and ordering notations on $\mathcal{P}$. Composition $\tau \circ x$ is noted shortly τx for any τ and x in $\mathcal{P}$.

4.1 Uniformity for $\pi = \frac{1}{2}$

Proposition 4.1 proves the uniformity for $\pi = \frac{1}{2}$, and requires Lemma B.3 of Appendix B.

Proposition 4.1. *For all $x, \mu \in \mathcal{P}$, $p(x; \mu, \frac{1}{2}) = m!^{-1}$.*

Proof. Let e be the identity permutation of $\mathcal{P}$. Using firstly Lemma B.3 of Appendix B and then using the fact that $\mathrm{p}(.|e; \mu, \frac{1}{2})$ is a probability distribution on $\mathcal{P}$, we have

$$\mathrm{p}(x; \mu, \tfrac{1}{2}) \propto \sum_{y \in \mathcal{P}} \mathrm{p}(x|y; \mu, \tfrac{1}{2}) = \sum_{y \in \mathcal{P}} \mathrm{p}(y^{-1}x|y^{-1}y; \mu, \tfrac{1}{2}) = \sum_{y \in \mathcal{P}} \mathrm{p}(y^{-1}x|e; \mu, \tfrac{1}{2}) = 1.$$

□

4.2 Mode and anti-mode

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 $\pi > \frac{1}{2}$ (Proposition 4.2). Let $\bar{\mu}$ be defined by $\bar{\mu} = \mu\bar{e}$ where $\bar{e} = (m, \ldots, 1)$ is the permutation of total inversion. This rank $\bar{\mu}$ is the furthest from μ for the Kendall distance. We symmetrically prove in this section that the unique anti-mode (the rank of smallest probability) is $\bar{\mu}$ if $\pi > \frac{1}{2}$ (Corollary 4.1). Finally, Proposition 4.3 establishes that the mode is uniformly more pronounced when π grows. Proofs require Lemmas B.2 and B.6 of Appendix B.

Proposition 4.2. *For all $x \neq \mu \in \mathcal{P}$ and $\pi > \frac{1}{2}$, $p(\mu; \mu, \pi) > p(x; \mu, \pi)$.*

Proof. Using successively the fact that $\{\pi > \frac{1}{2} \Leftrightarrow \pi > 1 - \pi\}$, $x \neq \mu$ and then Lemma B.2, we obtain:

$$m!\,\mathrm{p}(x; \mu, \pi) < \sum_{y \in \mathcal{P}} \pi^{A(x,y)} = \sum_{y \in \mathcal{P}} \pi^{A((\mu x^{-1})x, (\mu x^{-1})y)} = \sum_{y' \in \mathcal{P}} \pi^{A(\mu, y')} = m!\,\mathrm{p}(\mu; \mu, \pi).$$

The last equality comes from the fact that $A(\mu, y') = G(\mu, y', \mu)$. □

Corollary 4.1. *For all $x \neq \bar{\mu} \in \mathcal{P}$ and $\pi > \frac{1}{2}$, $p(\bar{\mu}; \mu, \pi) < p(x; \mu, \pi)$.*

The proof, symmetrical to that of Proposition 4.2, is left to the reader.

Proposition 4.3. *For all $x, \mu \in \mathcal{P}$, $p(\mu; \mu, \pi) - p(x; \mu, \pi)$ is an increasing function of $\pi \geq \frac{1}{2}$.*

Proof. Noting $\Delta(\pi) = \mathrm{p}(\mu; \mu, \pi) - \mathrm{p}(x; \mu, \pi)$, $\partial\Delta(\pi)/\partial\pi$ can be written

$$\frac{\partial\Delta(\pi)}{\partial\pi} = \frac{1}{m!} \sum_{y \in \mathcal{P}} \left\{ A(\mu, y)\pi^{A(\mu,y)-1} - G(x, y, \mu)\pi^{G(x,y,\mu)-1}(1-\pi)^{A(x,y)-G(x,y,\mu)} \right\} + c$$

where c is a non-negative term. Since $\pi \geq \frac{1}{2}$, we deduce that

$$G(x, y, \mu)\pi^{G(x,y,\mu)-1}(1-\pi)^{A(x,y)-G(x,y,\mu)} \leq G(x, y, \mu)\pi^{A(x,y)-1}.$$

Using the fact that $A(\mu, y) \geq G(x, y, \mu)$, we deduce that $\partial\Delta(\pi)/\partial\pi \geq 0$. □

4.3 Symmetry

In this section a symmetry of the ISR distribution is proved with the following sense: Distributions $\mathrm{ISR}(\mu, \pi)$ and $\mathrm{ISR}(\bar{\mu}, 1-\pi)$ are equivalent (Proposition 4.4 below). This property will be especially useful to exhibit the identifiability conditions of the ISR distribution in the next section. Proposition 4.4 requires Lemma B.4 in Appendix B.

Proposition 4.4. *For all $x, \mu \in \mathcal{P}$ and all $\pi \in [0, 1]$, $p(x; \bar{\mu}, 1-\pi) = p(x; \mu, \pi)$.*

Proof. Using Lemma B.4, we can write:

$$\mathrm{p}(x; \bar{\mu}, 1-\pi) \propto \sum_{y \in \mathcal{P}} \pi^{A(x,y)-(A(x,y)-G(x,y,\mu))}(1-\pi)^{A(x,y)-G(x,y,\mu)} \propto \mathrm{p}(x; \mu, \pi).$$

□

4.4 Identifiability

A necessary identifiability condition is immediately suggested by Propositions 4.1 and 4.4: The uniformity for $\pi = \frac{1}{2}$ of the ISR distribution and its symmetry lead to impose $\pi > \frac{1}{2}$. The sufficiency of this condition is proved in the next proposition. Its proof needs Lemma B.5 of Appendix B.

Proposition 4.5. *The ISR distribution is identifiable since $\pi > \frac{1}{2}$.*

Proof. The identifiability problem can concern parameters π and/or μ.

- First, there exists none couple $(\mu, \mu') \in \mathcal{P}^2$ with $\mu \neq \mu'$ such that $\mathrm{p}(x; \mu, \pi) = \mathrm{p}(x; \mu', \pi)$ for any $x \in \mathcal{P}$ and any $\pi > \frac{1}{2}$. Indeed, choosing $x = \mu$, from Lemma B.5 we have $\mathrm{p}(\mu; \mu, \pi) \neq \mathrm{p}(\mu; \mu', \pi)$.

- Second, for a given $\mu \in \mathcal{P}$, assume there exists $\pi \neq \pi'$ such that $\mathrm{p}(x;\mu,\pi) = \mathrm{p}(x;\mu,\pi')$ for any $x \in \mathcal{P}$. In particular, for $x = \mu$, in the proof of Lemma B.5 it is obtained that $G(x,y,x) = A(x,y)$, thus $\sum_{y\in\mathcal{P}} \pi^{A(\mu,y)} = \sum_{y\in\mathcal{P}} {\pi'}^{A(\mu,y)}$. The strict increasing of the function $p \mapsto \pi^n$ on the interval $[\frac{1}{2},1]$ for all $n \in \mathbb{N}^*$ ensures that $\pi = \pi'$.

- Assume finally there exists $(\mu,\mu') \in \mathcal{P}^2$ with $\mu \neq \mu'$ and $\pi < \pi'$ such that $\mathrm{p}(x;\mu,\pi) = \mathrm{p}(x;\mu',\pi')$ for any $x \in \mathcal{P}$. In the proof of Lemma B.5, it is obtained also that $G(x,y,\mu) < A(x,y)$ when $x \neq \mu$, thus

$$\mathrm{p}(x|y;\mu,\pi) < \pi^{A(x,y)} < {\pi'}^{A(x,y)} = \mathrm{p}(x|y;x,\pi')$$

  and then by averaging over all y in $\mathcal{P}$ gives $\mathrm{p}(x;\mu,\pi) < \mathrm{p}(x;x,\pi')$. 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 π, which is a real in $[\frac{1}{2},1]$ and the reference rank, or modal rank, μ, which can take its values in $\mathcal{P}$. Note that the case $\pi = \frac{1}{2}$ 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. Considering $(x^1,\ldots,x^n)$ as an independent sample of n ranks from ISR(μ,π), we present in this section estimation of (μ,π) by maximizing the log-likelihood of the ISR model which is given by

$$l(\mu,\pi) = \sum_{i=1}^{n} \ln\left(\frac{1}{m!}\sum_{y\in\mathcal{P}} \mathrm{p}(x^i|y;\mu,\pi)\right).$$

5.1 Using an EM algorithm

As the presentation orders $(y^1,\ldots,y^n)$ are unknown (latent variables), we use an EM algorithm [DLR77] to maximize this *observed* data log-likelihood. Denoting by $(\mu,\pi)^{\{0\}}$ the starting parameter of EM and by $(\mu,\pi)^{\{q\}}$ the current value of the parameters at the step q ($q \in \mathbb{N}$), the two steps (E and M) of this algorithm are described as follows. We have assumed that pairs (x^i,y^i) arise independently $(i = 1,\ldots,n)$.

The E step The *complete-data* log-likelihood is given by

$$l_c(\mu,\pi) = \sum_{i=1}^{n}\sum_{y\in\mathcal{P}} \mathbf{1}\{y = y^i\} \ln\left(\frac{1}{m!}\mathrm{p}(x^i|y;\mu,\pi)\right).$$

The E step consists in computing the conditional expectation $\mathcal{Q}$ of l_c expressed by:

$$\mathcal{Q}((\mu,\pi),(\mu,\pi)^{\{q\}}) = \sum_{i=1}^{n}\sum_{y\in\mathcal{P}} t_{iy}^{\{q\}} \ln\left(\frac{1}{m!}\mathrm{p}(x^i|y;\mu,\pi)\right)$$

where the conditional probability that $y^i = y$ is noted

$$t_{iy}^{\{q\}} = \frac{\mathrm{p}(x^i|y;(\mu,\pi)^{\{q\}})}{\sum_{\tau\in\mathcal{P}}\mathrm{p}(x^i|\tau;(\mu,\pi)^{\{q\}})}.$$

The M step The M step consists in choosing the value $(\mu,p)^{\{q+1\}}$ which maximizes the conditional expectation $\mathcal{Q}$ computed at the E step:

$$(\mu,\pi)^{\{q+1\}} = \underset{(\mu,\pi)\in\mathcal{P}\times[\frac{1}{2},1]}{\operatorname{argmax}} \mathcal{Q}((\mu,p),(\mu,\pi)^{\{q\}}).$$

As the parameter space $\mathcal{P}$ for μ is discrete, the maximization simply consists, but potentially computationally expensively, of browsing the entire $\mathcal{P}$ (we give a more cute strategy in Section 5.2). For the probability π, maximizing $\mathcal{Q}$ leads to the following maximum:

$$\pi^{\{q+1\}} = \frac{\sum_{i=1}^{n}\sum_{y\in\mathcal{P}} t_{iy}^{\{q\}} G(x^i,y,\mu^{\{q\}})}{\sum_{i=1}^{n}\sum_{y\in\mathcal{P}} t_{iy}^{\{q\}} A(x^i,y)}.$$

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

5.2 Initializing EM and reducing its computational cost

We propose first an immediate asymptotic bound on π and then a strategy to reduce, often drastically, the number of possible values for μ. Both results are useful for initializing EM and also for reducing highly the computational cost of the M step. They rely on the following two propositions.

Proposition 5.1. *Denoting by f_0 the empirical modal relative frequency, the interval $[\hat{\pi}^-,\hat{\pi}^+]$ asymptotically contains π where*

$$\hat{\pi}^- = f_0^{\frac{1}{m-1}} \quad \textit{and} \quad \hat{\pi}^+ = f_0^{\frac{2}{m(m-1)}}. \tag{5.1}$$

Proof. Using Lemma B.6 and also the fact that, for any μ and y, $\mathrm{p}(\mu|y;\mu,\pi) = \pi^{A(\mu,y)}$ (see the proof in Lemma B.5), it leads to the following bounds for the probability of μ:

$$\pi^{m(m-1)/2} \le \mathrm{p}(\mu;\mu,\pi) \le \pi^{m-1}.$$

Since f_0 is a consistent estimator of $\mathrm{p}(\mu;\mu,\pi)$, it ends the proof. □

As soon as $\hat{\pi}^-$ and $\hat{\pi}^+$ are greater than $\frac{1}{2}$, this result is useful for initializing π in EM by choosing uniformly at random $\pi^{\{0\}}$ in the interval given by (5.1). If only $\hat{\pi}^+ \geq \frac{1}{2}$, the interval becomes $[\frac{1}{2}, \hat{\pi}^+]$. If both bounds are lower than $\frac{1}{2}$, then the interval $[\frac{1}{2}, 1]$ must be used. In Table 2 of Section 6, bounds associated to all data sets are greater than $\frac{1}{2}$ and the retained intervals are quite narrow in comparison to $[\frac{1}{2}, 1]$, so the strategy makes the job. The next proposition is now focused on μ but requires the result of Proposition 5.1.

Proposition 5.2. *Let N_x be the number of individuals equal to $x \in \mathcal{P}$ among a n random sample from* $\text{ISR}(\mu, \pi)$. *Denoting by*

$$h_\alpha(\pi) = \#\{x : p(N_x \geq N_\mu; \mu, \pi) \geq \alpha\}$$

the number of ranks for which the empirical frequency can be greater or equal (with probability at least $\alpha \in [0, 1]$) than the empirical frequency associated to the theoretical modal rank μ, then the following inequality asymptotically holds for any $\mu \in \mathcal{P}$ and $\pi \in [\frac{1}{2}, 1]$:

$$h_\alpha(\pi) \leq h_\alpha\left(\hat{\pi}^-\right).$$

Proof. We know from Proposition 5.1 that asymptotically $\hat{\pi}^- \leq \pi$. For concluding the proof, it is sufficient to use Proposition 4.3. □

The following strategy can be only used if $\hat{\pi}^- \geq \frac{1}{2}$. Firstly, $h_\alpha\left(\hat{\pi}^-\right)$ is estimated with a parametric bootstrap [ET93] of M replications from $\text{ISR}(\mu, \hat{\pi}^-)$. The key point is that it is independent on μ, so any $\mu \in \mathcal{P}$ can be used. Then the $h_\alpha\left(\hat{\pi}^-\right)$ distinct most frequent distinct ranks in the sample $(x^1, \ldots, x^n)$ are retained as possible μ values among the potential $m!/2$ possibilities and are used both as potential initial values $\mu^{\{0\}}$ and also as values to browse at the M step. In other words, the idea is to browse the emprirical modal rank in association with some other ranks having quite high empirical relative frequency.

The proposed strategy is aimed to significantly reduce the number of candidates for μ. It decreases when the size of the observed sample n grows since $h_\alpha\left(\hat{\pi}^-\right) \xrightarrow{\text{p}} 1$ when $n \to \infty$. So, the browsed ranks are asymptotically reduced to the empirical modal rank which is known to be a consistent estimate of μ. Note that the selection of the possible ranks should be carried out only once before to start the EM algorithm.

Table 2 (Column "$\#\mu$") of Section 6 illustrates through numerical examples that this procedure effectively reduces the number of possible ranks for μ in comparison to the $m!/2$ possible values.

6 Numerical illustration

6.1 Presentation of the five real data sets

The ISR distribution is now compared to the Mallows Φ model on five real data sets: Two general knowledge quizzes (the answers of the 40 questioned students are

in Appendix C), four nations rugby league rankings, Fligner and Verducci's words associations rankings [FV86] and Louis Roussos's sports rankings [Mar95].

- *Football* quiz. This first quiz consists in ranking four national football teams according to increasing number of victories in the football World Cup: $\mathcal{O}_1 =$ France, $\mathcal{O}_2 =$ Germany, $\mathcal{O}_3 =$ Brasil, $\mathcal{O}_4 =$ Italy. The correct answer is $\mu^* = (1, 2, 4, 3)$.

- *Cinema* quiz. This quiz consists in ranking chronologically the following 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 answer is $\mu^* = (3, 2, 4, 1)$.

- *Rugby*. This data set is the result of the four nations rugby league, from 1883 to 1909 (except years 1888 and 1889 because only three nations were in the tournament, and except years 1886, 1890, 1897, 1898 and 1906 due to tie), which opposed $\mathcal{O}_1 =$ England, $\mathcal{O}_2 =$ Scotland, $\mathcal{O}_3 =$ Ireland and $\mathcal{O}_4 =$ Walles.

- *Words*. [FV86] examined the data collected under the auspices of the Graduate Record Examination Board. A sample of 98 college students were asked to rank five words according to strength of association (least to most associated) with the target word "Idea": $\mathcal{O}_1 =$ Thought, $\mathcal{O}_2 =$ Play, $\mathcal{O}_3 =$ Theory, $\mathcal{O}_4 =$ Dream and $\mathcal{O}_5 =$ Attention.

- *Sports*. This last data set is due to Louis Roussos [Mar95] who asked 130 students at the University of Illinois to rank seven sports according to their preference in participating: $\mathcal{O}_1 =$ Baseball, $\mathcal{O}_2 =$ Football, $\mathcal{O}_3 =$ Basketball, $\mathcal{O}_4 =$ Tennis, $\mathcal{O}_5 =$ Cycling, $\mathcal{O}_6 =$ Swimming, $\mathcal{O}_7 =$ Jogging.

Empirical distribution of the first three data sets (for which the number of objects to rank is 4) is graphically displayed on the *left* column of Figure 1 in the ranking space (orderings are displayed on each node).

6.2 Estimation results

For each dataset, the ISR distribution and the Mallows Φ model are estimated. For ISR the convergence threshold for the growth of the log-likelihood in the EM algorithm was fixed to $1e{-}6$ and only one initialization of π in $[\hat{\pi}^-, \hat{\pi}^+]$ has been used (no change on the results have been observed with several initializations). For Mallows Φ model, the numerical optimization has been carried out by the *optim* function of R (programmed in C) with a quasi-Newton method and the same convergence threshold than for ISR ($1e-6$).
The ISR distribution of the first three data sets is graphically displayed on the *right* column of Figure 1 for a visual comparison with the empirical distribution. In addition, a χ^2 adequacy test, where the distribution under the null assumption is estimated by bootstrap [ET93] based on 1000 replications, is performed for

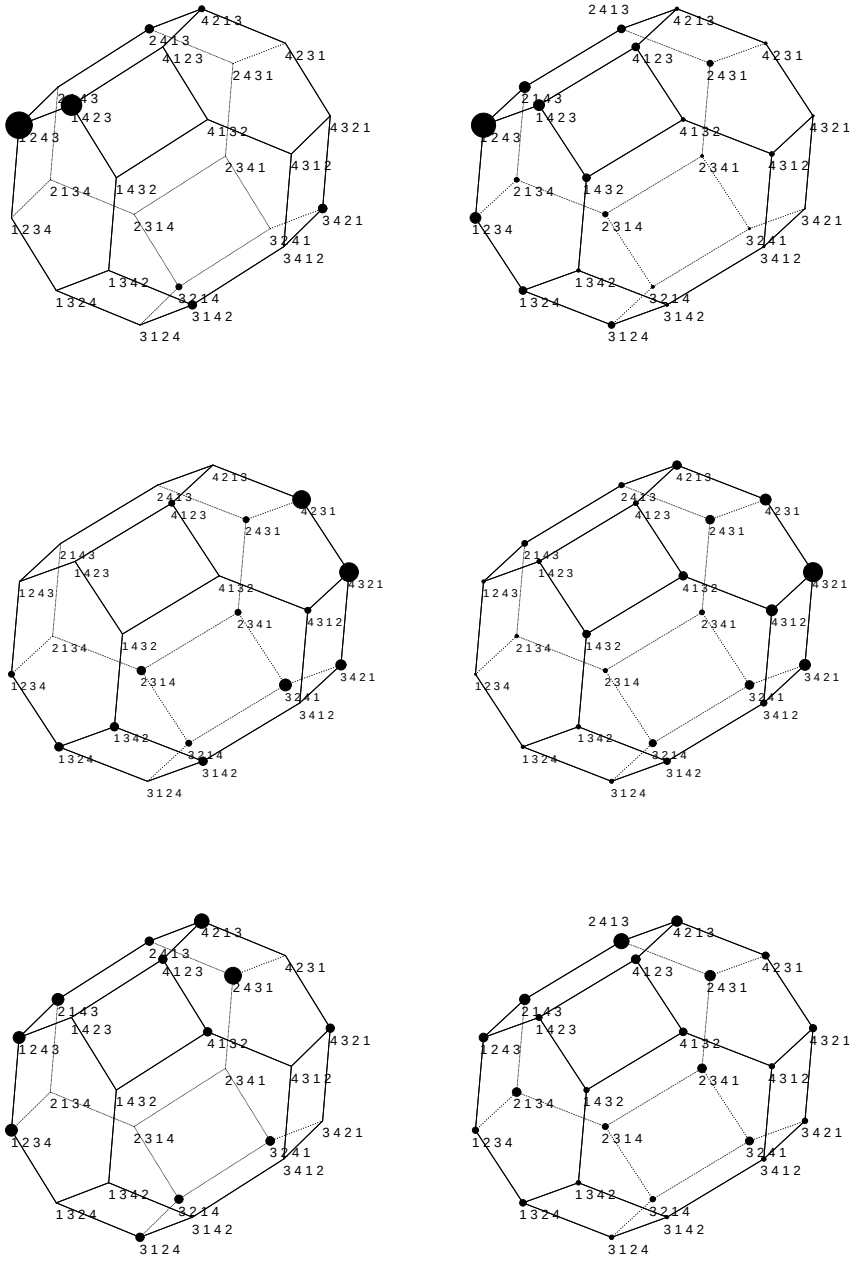

Figure 1: Empirical (left) and estimate ISR (right) distributions for Football and Cinema quizzes and four nations rugby league (from top to bottom).

both models and for all data sets and the results are displayed in Table 2 (Column "$\widehat{\text{p-value}}$"). We notice that both models can be suitable for some data sets but not all of them and not necessarily the same ones. This fact is corroborated when comparing maximum log-likelihood values (Column "l"; The highest likelihoods are in bold). Consequently, ISR could be a natural competitor to be considered beside other classical models in any rank data analysis. Additional arguments for using ISR appear also when analysing further Table 2.

Table 2: ISR and Mallows Φ models estimation results: Estimate parameters $\hat{\mu}$, $\hat{\pi}$ (ISR) and $\hat{\lambda}$ (Mallows), maximum log-likelihood l, estimated p-value of the χ^2 adequacy test, number of possible μ explored ($\#\mu$; For ISR it corresponds to $\hat{h}_\alpha(\hat{\pi}^-)$ with $\alpha = 0.05$ and $M = 100$ replications), lower and upper bounds $\hat{\pi}^-$ and $\hat{\pi}^+$ for π (ISR only) and times of execution (in seconds).

data set	model	$\hat{\mu}$	$\hat{\pi}$ or $\hat{\lambda}$	l	$\widehat{\text{p-value}}$	$\#\mu$	$\hat{\pi}^-$	$\hat{\pi}^+$	times
Football	ISR	$(1,2,4,3)$	0.834	**-88.53**	0.001	1	0.794	0.891	2
	Mallows	$(1,2,4,3)$	1.106	-89.17	0.001	1	-	-	1
Cinema	ISR	$(4,3,2,1)$	0.723	**-111.97**	0.042	14	0.630	0.794	4
	Mallows	$(4,3,2,1)$	0.628	-112.12	0.029	2	-	-	1
Rugby	ISR	$(2,4,1,3)$	0.681	-58.68	0.538	12	0.585	0.765	3
	Mallows	$(2,4,1,3)$	0.528	**-58.33**	0.395	2	-	-	2
Words	ISR	$(2,5,4,3,1)$	0.879	-275.43	0.001	1	0.762	0.897	6
	Mallows	$(2,5,4,3,1)$	1.431	**-251.27**	0.019	1	-	-	2
Sports	ISR	$(1,3,2,4,5,7,6)$	0.564	**-1102.12**	0.999	1	0.534	0.836	1069
	Mallows	$(1,3,4,2,5,6,7)$	0.083	-1102.84	0.045	11	-	-	187

Firstly, we note that the ISR model estimation can be achieved in a reasonable time (column "times", obtained with a Bi Xeon processor running at 3.0GHz and 32GB of RAM) with a program[1] in R, although greater than Mallows Φ model estimation (in C): Only few seconds until five objects to rank, less than 20 minutes for 7 objects. This estimation time is allocated as follows: About 1% for the strategy leading to reduce the number of possible reference ranks (Section 5.2), 98% for a pre-processing step consisting of the computation of the terms $G(x, y, \mu)$ and $A(x, y)$ for all observed ranks x, all retained reference ranks μ and all presentation orders y, and finally only 1% for the EM algorithm. Since the pre-processing step is done by three nested loops (on x, y and μ), we can expect that a C program can drastically reduce the corresponding computing time such that we can deal with 9, perhaps 10, objects in few minutes (recall that $m = 10$ is the "optimality limit" for ISR). Contributing to the reduction of the computing time, the strategy selecting the number of possible reference ranks to explore (Section 5.2) is very effective. Indeed, only one candidate for μ has been selected by this strategy for the three data sets Football, Words and Sports (Column "$\#\mu$"). Concerning the Mallows Φ model, the estimation of μ is

[1]Software available on the authors website: http://math.univ-lille1.fr/∼jacques/soft.html

carried out by a quite empirical iterative local research (in the sense of the Kendall distance) around the modal rank [FV88] which appears to be effective yet.

We discuss now the meaningful interpretation of ISR parameters. For each of the Football and Cinema quizzes, the estimation of the reference rank μ coincides with the real rank. This underlines that the right answers are, on the whole, known by this population of students, and the accuracy level of students knowledge in these areas is reflected by the probability π of well paired comparison: 0.834 for the Football quiz and 0.723 for the Cinema one. Thus, these students have better knowledge in Football than in Cinema. The ISR model estimation on the Rugby data set enhances a *natural* ranking between these four nations: During this time Scotland were the best, then Walles, England and finally Ireland. But the low value of the probability π (0.681) means that this ranking was not very flagrant. On the opposite, the high value of π (0.879) for the Fligner and Verducci's Word data set shows that the questionned students overall had the same thinking for the association with the word Idea: Play is the least associated then Attention, Dream, Theory and finally Thought is the most associated. The last dataset is also very interesting. The reference rank $(1, 3, 2, 4, 5, 7, 6)$ estimated for the ISR model reflects a preference of the students at the University of Illinois for collective sports: Baseball, Basketball and Football are at the top three places while individual sports are at the end of the ranking: Cycling, Jogging and Swimming. Tennis, which is intermediate between a collective sport and an individual sport, is rationally ranked between these two groups.

From the Mallows Φ parameters point of view, most results are highly consistent with ISR: Main modal ranks are identical and the dispersion parameter λ is also well correlated with π, though λ is more abstract and could be less easy to understand by a practitioner. Only the modal rank of the last data set (Sports data set) differs: The Mallows Φ model classifies Tennis inside the collective sports collection instead of being put at the borderline of collective and individual sports, as ISR does.

6.3 Specificity and coherence of ISR

Here, we propose to exploit a specificity of the ISR model: It is possible to retrieve information on the order with which the objects to rank have been presented. For this purpose, we propose to compute for each possible $y \in \mathcal{P}$ its probability conditionally to the observed sample $(x^1, \ldots, x^n)$ and to the fact that all ranks of this sample have been generated with the same presentation order: $\mathrm{p}(y|x^1, \ldots, x^n, y^1 = \ldots = y^n)$. Thus, if there exists a common presentation order, we can expect to retrieve one rank with a high probability compared to the others, and in the contrary case, the probabilities must be more equidistributed. Figure 2 displays these probabilities for each possible presentation order y. In fact only the half of $\mathcal{P}$ is taken into consideration since the relative presentation order of the first two objects has no importance (last remark in Section 3). On this figure the probabilities are ranked by decreasing order of importance, and only the twenty largest are presented for the dataset with $m > 4$.

As expected, for the two Football and Cinema quizzes, where we know that

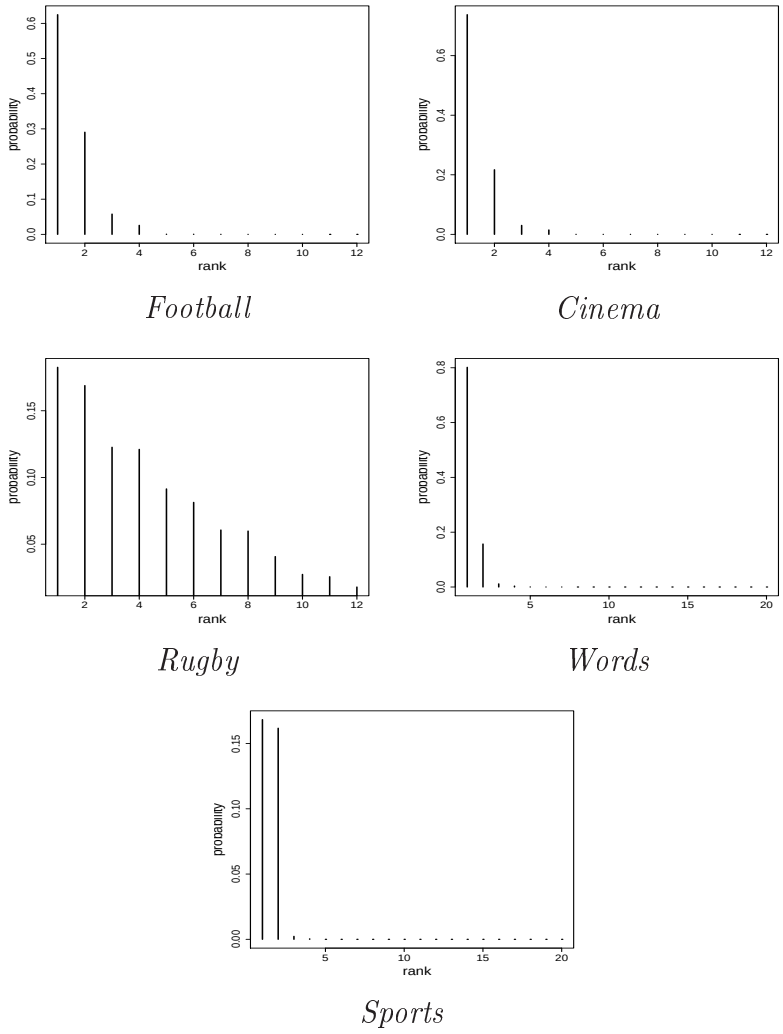

Figure 2: Presentation order probability for *Football* and *Cinema* quizzes, 4 Nations *Rugby* League, Fligner and Verducci's *Words* association data set, and Louis Roussos's *Sport* ranking.

the objects have been presented in the same order for all the students (recall it is a quizz the authors have built), we retrieve the fact that one presentation order is more probable than the other. Clearly, the same result is also obtained for the Fligner and Verducci's Words and the Louis Roussos's Sports data sets, which suggests that exactly the same quizz has been also presented to all students. Finally, for the Rugby dataset, where no presentation order exists with as much evidence (recall it is not a quizz), the probabilities of the presentation orders are more balanced. These experiments highlight the high coherence of ISR since it is able to retrieve some important but hidden information in the data.

Remark However, we have noticed that ISR fails in retrieving the true initial ordering y for the first two quizzes (we know y for both). This observation will lead to interesting comments for future research directions in the discussion of the last section below.

7 Discussion

In this paper we suggest to consider rank data as the result of a paired comparisons sorting algorithm, where the possibility of wrong comparisons exists and can be randomly modelized. It opens a new way for proposing many distributions on rankings, all of them benefiting from very meaningful parameters (the reference rank μ and the probability π of good paired comparison) and also allowing to retrieve the latent initial rank y at the beginning of the process. Considering the case of $m \leq 10$ objects to rank and aiming to minimize the number of paired comparisons for avoiding as much as possible the potential wrong comparisons, the insertion sorting algorithm has been retained in this paper for its optimality in this context. The resulting distribution, the so-called ISR, has been established and many desirable properties have been pointed out. In addition, the latent variable interpretation of the model allows to derive a specific EM algorithm which can be easily accelerated by drastically reducing the number of potential reference ranks μ to consider.

Thus, the insertion sorting algorithm can be view as a first step in this new class of models, choice guided by optimality arguments. However, a selection sorting procedure can be for instance closer to the process followed by human judges and a first perspective of this work should be to establish the distribution and the corresponding properties in this case. This intuition is corroborated by the fact that the Mallows Φ model, which is somewhat based on a selection sorting algorithm, appears to be a hard competitor for the ISR model in our previous experiments. Moreover, this could be the reason for which we fail in retrieving exactly which presentation order has been used for the data sets studied in Section 6. For many objects to sort ($m \geq 10$), we can alternatively consider other procedures as the quick sort, and so on. Note that, for "high" m values, a particular attention should be paid to the computation cost involved in our models.

Another interesting prospect initialized by the present work is the possibility to include some information about the initial ranking y in the model and its corre-

sponding estimation. Indeed, in questionnaires this initial order is often known and it is a useful information which can be naturally used by our class of models. It is also possible to consider some more diffuse information about y, for instance to ignore the exact y value but to know that all y are the same for all questionnaires (realistic situation for many ranks coming from quiz studies), or other realistic variants [BJ10].

Although the ISR is unimodal (as many other distributions for ranks), multimodality can be easily taking into account through mixture of ISR distributions and a specific identifiability study. For instance, we can think that in our football quiz, girls and boys responses will probably not follow the same distribution, as it is suggested by the quite low estimated p-value [JB10].

At last, there is also a need to adapt our models to other situations than full rank data. 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.

A Building the ISR distribution

The goal of this appendix is to prove that Formula (3.1) corresponds to the stochastic left insertion sorting algorithm with probability π of good paired comparison and independence between the paired comparisons. The notations are those defined in Section 3.

Proof. Let $x^{(j)}$ be the ordering of the j $(1 \leq j \leq m)$ first objects in y in the order imposed by x (so $x^{(m)} = x$). An example of this notation is in Table 1. Thus, there exists the following relationship between $x^{(j)}$ and $x^{(j-1)}$:

$$x^{(j)} = (x_1^{(j-1)}, \ldots, x_{A_j^-(x,y)}^{(j-1)}, y_j, x_{A_j^-(x,y)+1}^{(j-1)}, \ldots, x_{j-1}^{(j-1)}).$$

Formula (3.1) is now proved by induction on j. It is true for $j = 1$ while there is only one object y_1 to sort: $\mathrm{p}(x^{(1)}|y;\mu,\pi) = 1$. Since the result of the ranking $x^{(j)}$ from $x^{(j-1)}$ is the result of $A_j(x,y)$ independent Bernoulli experiments of parameter π, then, conditionally to $x^{(j-1)}$, the probability of $x^{(j)}$ is

$$\mathrm{p}(x^{(j)}|x^{(j-1)}, y;\mu,\pi) = \pi^{G_j(x,y,\mu)}(1-\pi)^{A_j(x,y)-G_j(x,y,\mu)}.$$

We conclude the proof by noticing that

$$\mathrm{p}(x^{(j)}|y;\mu,\pi) = \mathrm{p}(x^{(j)}|x^{(j-1)}, y;\mu,\pi)\mathrm{p}(x^{(j-1)}|y;\mu,\pi),$$

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

B Lemmas

Lemma B.1. *Let $\tilde{e} = (2,1,3,\ldots,m)$ be the permutation inverting the first two elements. For all $x, y, \mu \in \mathcal{P}$ and $\pi \in [0,1]$, $p(x|y;\mu,\pi) = p(x|y\tilde{e};\mu,\pi)$.*

Proof. We use notations $x^{(j)}$ already introduced in Appendix A. The key point of the proof is to notice that the first two objects in y lead to the same paired comparison at the second step of the sorting process whatever is their order in y, so $\mathrm{p}(x^{(2)}|y\tilde{e},\pi) = \mathrm{p}(x^{(2)}|y,\pi)$. Combining this result with the fact that $\mathrm{p}(x|x^{(2)},y\tilde{e},\pi) = \mathrm{p}(x|x^{(2)},y,\pi)$, since $\tilde{e}$ only affects the first two objects, concludes the proof. □

Lemma B.2. *For all* $x,y,\tau \in \mathcal{P}$, $A(x,y) = A(\tau x,\tau y)$.

Proof. First we prove that $A_j^-(x,y) = A_j^-(\tau x,\tau y)$. For any $j = 1,\ldots,m$, we have (notice that i is always such that $1 \leq i < j$)

$$\begin{aligned} &A_j^-(\tau x,\tau y) \\ &\quad = \#\{i : (\tau x)^{-1}_{(\tau y)_i} < (\tau x)^{-1}_{(\tau y)_j}\} = \#\{i : (x^{-1}\tau^{-1}\tau y)_i < (x^{-1}\tau^{-1}\tau y)_j\} \\ &\quad = \#\{i : (x^{-1}y)_i < (x^{-1}y)_j\} = \#\{i : x^{-1}_{y_i} < x^{-1}_{y_j}\} = A_j^-(x,y). \end{aligned}$$

By noticing that $A_j^+(x,y) = \mathbf{1}\{A_j^-(x,y)+1 \leq j-1\}$ we deduce also that $A_j^+(x,y) = A_j^+(\tau x,\tau y)$. Consequently, $A_j(x,y) = A_j(\tau x,\tau y)$ and, so, $A(x,y) = A(\tau x,\tau y)$. □

Lemma B.3. *For all* $x,y,\mu,\tau \in \mathcal{P}$, $p(x|y;\mu,\frac{1}{2}) = p(\tau x|\tau y;\mu,\frac{1}{2})$.

Proof. When $\pi = \frac{1}{2}$, we obtain by using Lemma B.2

$$\mathrm{p}(\tau x|\tau y;\mu,\tfrac{1}{2}) = \left(\tfrac{1}{2}\right)^{A(\tau x,\tau y)} = \left(\tfrac{1}{2}\right)^{A(x,y)} = \mathrm{p}(x|y;\mu,\tfrac{1}{2}).$$

□

Lemma B.4. *For all* $x,y,\mu \in \mathcal{P}$ $G(x,y,\bar{\mu}) = A(x,y) - G(x,y,\mu)$.

Proof. Let $\bar{e}$ be the permutation of total inversion previously introduced in Section 4.2 and $i,i' = 1,\ldots,m, i \neq i'$. We first prove that $G_j^-(x,y,\bar{\mu}) = A_j^-(x,y) - G_j^-(x,y,\mu)$. Using successively the fact that $\bar{\mu} = \mu\bar{e}$, $\bar{e} = \bar{e}^{-1}$, $\{i < i' \Leftrightarrow \bar{e}_i > \bar{e}_{i'}\}$ and $i \neq i'$, we have

$$\begin{aligned} &\delta_{ii'}(\bar{\mu}) \\ &\quad = \mathbf{1}\{(\mu\bar{e})_i^{-1} < (\mu\bar{e})_{i'}^{-1}\} = \mathbf{1}\{\bar{e}^{-1}_{\mu_i^{-1}} < \bar{e}^{-1}_{\mu_{i'}^{-1}}\} = \mathbf{1}\{\bar{e}_{\mu_i^{-1}} < \bar{e}_{\mu_{i'}^{-1}}\} \\ &\quad = \mathbf{1}\{\mu_i^{-1} > \mu_{i'}^{-1}\} = 1 - \mathbf{1}\{\mu_i^{-1} < \mu_{i'}^{-1}\} = 1 - \delta_{ii'}(\mu), \end{aligned}$$

we deduce then that:

$$G_j^-(x,y,\bar{\mu}) = \sum_{i \in \mathcal{A}_j^-(x,y)} (1 - \delta_{y_i y_j}(\mu)) = A_j^-(x,y) - G^-(x,y,\mu).$$

In a similar way, we can prove that $G_j^+(x,y,\bar{\mu}) = A_j^+(x,y) - G_j^+(x,y,\mu)$. The proof follows immediately from these two results. □

Lemma B.5. *For all* $x,\mu \in \mathcal{P}$, $x \neq \mu$ *and* $\pi > \frac{1}{2}$, $p(x;\mu,\pi) < p(x;x,\pi)$.

Proof. Remark first that $G(x, y, \mu) < A(x, y)$ for $\mu \neq x$. Since $\{\pi > \frac{1}{2} \Leftrightarrow 1-\pi < \pi\}$, we deduce for $\mu \neq x$ that $\mathrm{p}(x|y; \mu, \pi) < \pi^{A(x,y)}$. Notice also that $G(x, y, x) = A(x, y)$, thus $\mathrm{p}(x|y; x, \pi) = \pi^{A(x,y)}$. Consequently, we have $\mathrm{p}(x|y; \mu, \pi) < \mathrm{p}(x|y; x, \pi)$ and the proof is concluded by averaging over all possible presentation orders y in $\mathcal{P}$. □

Lemma B.6. *For all $\mu, y \in \mathcal{P}$, $m - 1 \leq A(\mu, y) \leq m(m-1)/2$.*

Proof. Left bound: There is no comparison when the first element arises and at least one comparison for the $m - 1$ others. Right bound: There is still no comparison when the first element arises and at most $j - 1$ comparisons at the jth step for each new object to rank, so $A(\mu, y) \leq \sum_{j=1}^{m}(j - 1) = m(m-1)/2$. □

C Quiz data sets

Table 3: Quiz answers of the 40 students.

Cinema		Football	
ordering	frequency	ordering	frequency
$(4,3,2,1)$	10	$(1,2,4,3)$	20
$(4,2,3,1)$	9	$(1,4,2,3)$	12
$(3,2,4,1)$	4	$(2,4,1,3)$	2
$(3,4,2,1)$	3	$(3,1,4,2)$	2
$(1,3,2,4)$	2	$(3,4,2,1)$	2
$(1,3,4,2)$	2	$(3,2,1,4)$	1
$(2,3,1,4)$	2	$(4,2,1,3)$	1
$(3,1,4,2)$	2	other	0
$(1,2,3,4)$	1		
$(2,3,4,1)$	1		
$(2,4,3,1)$	1		
$(3,2,1,4)$	1		
$(4,1,2,3)$	1		
$(4,3,1,2)$	1		
other	0		

References

[BJ10] C. Biernacki and J. Jacques. Modèles génératifs de rangs relatifs à un algorithme de tri par insertion. In *42th Journées de Statistique organisée par la Société Française de Statistique*, Marseille, France, 2010.

[Böc93] U. 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.

[BT52] R.A. Bradley and M.E. Terry. Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika*, 39:324–345, 1952.

[Cri85] D. E. Critchlow. *Metric methods for analyzing partially ranked data*, volume 34 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin, 1985.

[DLR77] 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.

[ET93] B. Efron and R.J. Tibshirani. *An introduction to the bootstrap*, volume 57 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York, 1993.

[FA86] P.D. Feigin and M. Alvo. Intergroup diversity and concordance for ranking data: an approach via metrics for permutations. *Ann. Statist.*, 14(2):691–707, 1986.

[FV86] M.A. Fligner and J.S. Verducci. Distance based ranking models. *J. Roy. Statist. Soc. Ser. B*, 48(3):359–369, 1986.

[FV88] M.A. Fligner and J.S. Verducci. Multistage ranking models. *J. Amer. Statist. Assoc.*, 83(403):892–901, 1988.

[JB10] J. Jacques and C. Biernacki. Model-based clustering for rank data based on an insertion sorting algorithm. In *17th Rencontres de la Société Francophone de Classification*, La Réunion, 2010.

[Ken38] M.G. Kendall. A new measure of rank correlation. *Biometrika*, 30:81–93, 1938.

[Knu73] D.E. Knuth. *Sorting and Searching: Volume3. The art of Computer Programming*. Addison-Wesley, Massachusetts, 1973.

[KS40] M.G. Kendall and B.B. Smith. On the method of paired comparisons. *Biometrika*, 31:324–345, 1940.

[Luc59] R.D. Luce. *Individual choice behavior: A theoretical analysis*. John Wiley & Sons Inc., New York, 1959.

[Mal57] C.L. Mallows. Non-null ranking models. I. *Biometrika*, 44:114–130, 1957.

[Mar95] J.I. Marden. *Analyzing and modeling rank data*, volume 64 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1995.

[Pla75] R.L. Plackett. The analysis of permutations. *J. Roy. Statist. Soc. Ser. C Appl. Statist.*, 24(2):193–202, 1975.

[Tho93a] G.L. Thompson. Generalized permutation polytopes and exploratory graphical methods for ranked data. *Ann. Statist.*, 21(3):1401–1430, 1993.

[Tho93b] 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.

[Thu27] L.L. Thurstone. A law of comparative judgment. *Psychological Review*, 79:281–299, 1927.