Permutations based model for business performance

Arthur Fétiveau, Gilles Durrieu, Emmanuel Frénod, Claude-Henri Meledo, Benoît Prat

Permutations based model for business performance

Arthur Fétiveau, arthur.fetiveau@univ-ubs.fr[1,2], Gilles Durrieu, gilles.durrieu@univ-ubs.fr[1], Emmanuel Frénod, emmanuel.frenod@univ-ubs.fr[1], Claude-Henri Meledo, cmeledo@aldecis.com[2], and Benoît Prat, bprat@aldecis.com[2]

[1]Univ Bretagne Sud, CNRS UMR 6205, LMBA, F-56000 Vannes, France
[2]Aldecis, Paris, France

February 16, 2024

Abstract

This article is devoted to the development of a machine learning statistical framework to drive company's objectives. To this end, their sales data are used to target efficiently the issues or opportunities by a ranking. We implement a permutation based model using generalized Mallows models dealing with quantitative values, considering that a ranking is a permutation. The advantage of the generalized version is the possibility to differentiate the cost to move each element in the permutation. In our model, we differentiate the cost of an inversion in the permutation by using the gap value between the two elements. We propose model parameters estimators and we illustrate our estimation procedure on simulated data and on a real application.

Keywords : Computational statistics, Generalized Mallows models, Kemeny consensus, Kendall distance, Machine learning, Mallows models, Permutations.

1 Introduction

In this paper, we are interested in the development of a machine learning statistical framework to drive company's objectives. The defined approach helps the decision

makers choices using the company's sales data. We have to learn what is important to know from the decision makers point of view to efficiently help them to target the issues or opportunities.

To explain the objectives and for the sake of example, we study a company selling fishing products worldwide. This company sells a large number of different products like fish hooks, fishing nets, lures, fishing rods and more. We may have variations of each of those products in size, weight, material, color or even package, multiplying the total number of unique references to evaluate and analyze by the company. As this company makes profits worldwide, it sells each of those references in a bunch of stores, located in many countries with different climates. For example, this company can sell fishing products in France, Iceland, Mexico and Australia. In every country, we can also make a distinction between each selling region. For instance in France, we can sell in Brittany, New Aquitaine, Normandy, which are coastal regions but also in Great East as you can also fish in lakes or rivers. Depending on your responsibilities in the company, you are interested in a set of references in a geographical area. For example, the sales director of Brittany is interested in products sold in Brittany and is not interested if the company makes profit selling fishing rods in Australia.

For every product in stores, we can have information on a lot of indicators. For example, the revenue of the product, the selling price, the expected revenue leading to information on seasonality and many other information. Each criterion has to be taken into account to target the best interest of the user. For instance, the Normandy sales director is not interested to analyze the lure sales for salmons in the Channel during the winter as we can only fish them in spring and summer. Even with the deletion of the uninterested products, there is still a wide range of potentially interesting products. It would be really expensive, in money or time, to analyze every one of them at once. So, it is important to rank all of the references in a customized importance order. It is mandatory to prioritize the elements with the best potential for our Normandy sales director, but also every other sales director. For example, in September, it is more important for the Normandy sales director to analyze an important gap between the expected and the real revenue in the lures for cuttlefish than a huge gap between the expected and the real revenue in the lures for salmons. Indeed, as it is the beginning of the fishing season of cuttlefish and the end of the fishing season of salmons, a smaller gap may lead to a more important future profit. Once every sales director is provided with the indicator of his own importance order, he can study the elements in the decreasing order, and makes a plan for each one to increase their benefits. If our Normandy sales director has the time after analyzing the cuttlefish's lures sales, he may then study the salmon's lures sales.

Our objective is then to create a statistical and computational framework or-

dering every reference to match the user's preferences. This method uses the indicators given for all the products at every geographical level. However, this is not an easy supervised learning as we do not have any truth about the potential of the product. To help us with this, we use an information given by the user by reinforcement learning. First, our Normandy sales director, the user, have a ranking of his references; the ranking begins with the cuttlefish's lures, followed by the salmon's lures, and so on. He then analyzes the first references, the most as he can in the resources allowed to this task. Then, he returns an evaluation of his interest for every analyzed reference. Based on the user returns, we adjust the statistical model to fit the user preferences. This process iterates, month after month, to find the perfect ranking for the user.

As it is impossible for the user to give the perfect ranking of several hundreds of references or even to quantify the difference of two references, we will only ask for a value on a one to five ladder based on his interest. We then need to use a method able to deal with criteria composed of numerical continuous values, like the revenue where a difference of 10 euros is not the same as a difference of 100 euros, as well as ranking only criterion, like the partial ranking of the user preferences. To this end, we develop a permutation based procedure using generalized Mallows models. We consider here that a ranking is a permutation. The advantage of the generalized version it to be able to differentiate the cost to move each element in the permutation. However, we need to be able to differentiate the cost of an inversion in the permutation, for example by using the gap value between the two elements.

As we have seen in our example, the sales director of New Aquitaine where the fishermen fish the mackerel in the Bay of Biscay does not have the same kind of interest in the different criteria as the sales director of Great East. The seasonality of the mackerel is not as important as the salmon's seasonality in the large mountain rivers. Maybe, if you are the world sales director, the large amount of different climates and latitudes will smooth all the seasonality. It is then important to have a method able to be generic, to fit to every possible desire of the user. Each profile must be able to have its own set of parameters. We must also be able to compare each set of parameters between the different user to understand the differences between them.

This paper is organized as followed. In Section 2, we define the Mallows models, its generalization and limitations in our case. Section 3 deals with the parametric estimation of the model parameters. Section 4 introduces our modifications to integrate the cost of an inversion. Section 5 deals with the study of our procedure on simulated data while section 6 is devoted to real data application.

2 Mallows model and its generalization

2.1 Mallows model

The Mallows model [7] is based on the distances between the permutations of a set of n items to order. A permutations is a one to one application from $\{1, ..., n\}$ to itself. The set of all the permutations of n items is denoted by $\mathcal{S}_n$. We write $\pi(i)$ the position associated to the item i in the permutation named π for $i \in \{1, ..., n\}$ and $\pi^{-1}(i)$ the item associated to the i^{th} position of the permutation named π for $i \in \{1, ..., n\}$. The cardinal of $\mathcal{S}_n$ is equal to $n!$.

In this framework, we consider the permutation π as the perfect theoretical ranking for the user. The Mallows model is then defined as the probability to select randomly a permutation σ, in the set of all the permutations $\mathcal{S}_n$, based on the permutation π with a dispersion parameter θ. We have

$$P_\theta(\sigma) = \frac{\exp(-\theta\ d(\sigma, \pi))}{Z(\theta)} \tag{1}$$

with $\theta \geq 0$, $d(.,.)$ a right invariant distance between two permutations of $\mathcal{S}_n$ and

$$Z(\theta) = \sum_{\sigma \in \mathcal{S}_n} \exp(-\theta\, d(\sigma, \pi)), \tag{2}$$

the normalization term which is not influenced by π, thanks to the right invariant property. A distance $d(.,.)$ is right invariant [4] if $d(\pi, \sigma) = d(\pi\tau, \sigma\tau)$ for every π and σ in $\mathcal{S}_n$. The normalization term $Z(\theta)$ only varies with θ. This normalization term is then easy to compute in a time of order $O(n)$. Moreover, in the Mallows we tries as much as possible to avoid computing this value. As we can see in Section 3, we do not need it to estimate our parameters.

When $\theta = 0$, then each permutation σ of $\mathcal{S}_n$ has the same probability. The larger θ, the larger the density of σ around the modal permutation π. The distance is right invariant, meaning that a change of the labels in the permutations does not change the distance. To continue with our fishing example, if our permutation σ is first the mackerels fishing nets, then the salmon's lures and finally the fishing rods, and our permutation π is first the fishing rods, then the mackerels fishing nets and finally the salmon's lures, then renaming the mackerels fishing nets to 1, the salmon's lures to 2 and the fishing rods to 3 will gave the same distance as renaming the fishing rods to 1, the mackerels fishing nets to 2 and the salmon's lures to 3.

In the Figure 1, every node represents a permutation and every link represents a distance of 1 between 2 permutations. Here, we take the black node as π. Then, the probability of the black permutation, π which is at a distance of 0 is $\frac{\exp(0)}{Z(\theta)}$.

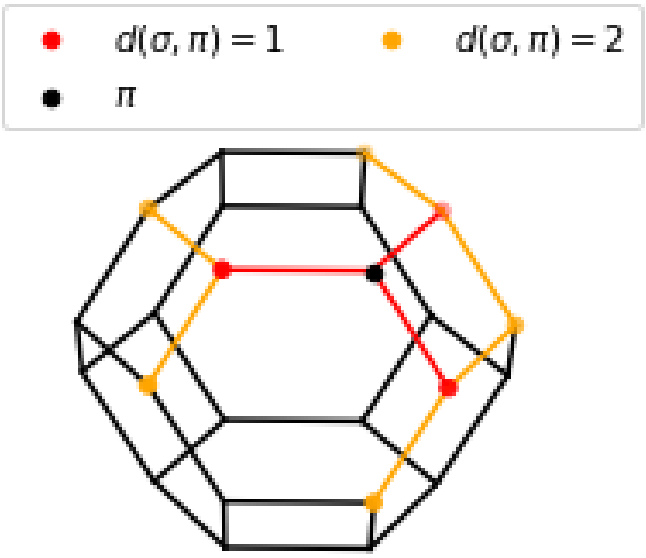

Figure 1: Permutohedron of order 4 for the Kendall Distance.

The probabilities of the red permutations are $\frac{\exp(-\theta)}{Z(\theta)}$ and the probabilities of the orange permutations are $\frac{\exp(-2\theta)}{Z(\theta)}$.

2.2 Generalized Mallows model

The generalized Mallows model [5] is an extension of the Mallows model based on a decomposition of a right invariant distance. If we have

$$d(\sigma, \pi) = \sum_{k=1}^{p} S_k(\sigma\pi^{-1}, e), \tag{3}$$

with e the identity permutation obtained by $\pi\pi^{-1}$, then, the parameter θ can be replaced by θ_k for $k = 1, \ldots, p$. We obtain

$$P_{\boldsymbol{\theta}}(\sigma) = \frac{1}{Z(\boldsymbol{\theta})} \exp\left(\sum_{k=1}^{p} -\theta_k \, S_k(\sigma\pi^{-1}, e)\right) \tag{4}$$

where

$$Z(\boldsymbol{\theta}) = \sum_{\pi \in \mathcal{S}_n} \exp\left(\sum_{k=1}^{p} -\theta_k \, S_k(\sigma\pi^{-1}, e)\right) \tag{5}$$

corresponds to the normalization term not influenced by π. The Mallows model is a particular case of the generalized Mallows model with $\theta = \theta_k$, $k = 1, \ldots, p$ when all θ_k are the same.

We decide here to use the Kendall distance [6]. The Kendall distance represents the minimal number of adjacent transpositions to transform a permutation σ into a permutation π, both from the set of permutations $\mathcal{S}_n$:

$$d_k(\pi, \sigma) = \sum_{i=1}^{n-1} \sum_{j>i} \mathbb{1}\left((\pi(i) > \pi(j) \cap \sigma(i) < \sigma(j)) \cup (\pi(i) < \pi(j) \cap \sigma(i) > \sigma(j))\right) \tag{6}$$

where $\mathbb{1}(\mathcal{P})$ takes the value 1 or 0 depending on whether the condition $\mathcal{P}$ is satisfied or not. This distance is right invariant.

Using our fishing example, the Kendall distance between our permutation σ, the mackerels fishing nets (A), the salmon's lures (B) and the fishing rods (C), and our permutation π, the fishing rods (C), the mackerels fishing nets (A) and the salmon's lures (B), is 2 as the mackerels fishing nets (A) is inverted with the fishing rods (C) and the salmon's lures (B) is inverted with the fishing rods (C).

With the Kendall discrepancy function, we can determine for $k = 1, \ldots, n-1$, $S_k(\sigma\pi^{-1}, e)$. The minimum number of adjacent inversion to assign the first item of e in the first position of $\sigma\pi^{-1}$ is S_1. We then remove this item from both permutations to compute S_2 which will be the minimum number of adjacent inversion to assign the remaining first item in the remaining first position. We repeat this process until there is not any remaining item. We have

$$S_k(\sigma\pi^{-1}, e) = \sum_{l>k} \mathbb{1}((\sigma\pi^{-1})^{-1}(k) > (\sigma\pi^{-1})^{-1}(l)). \tag{7}$$

With this distance, we assume that each item of π has its own parameter θ_k, depending on its position in π, to move and does not depend on the elements inverted with this item. However, in our problem, for some criteria, it may be important to make a difference between each inversion, and we want to determine a new parameter associated to each element inversion.

3 Estimation of the model parameters

We assume that the permutations σ_j are known and represent the ranking of the J criteria, each of them picked according to a Mallows model with the same modal permutation π but with different dispersion parameters θ_j for $j \in \{1, \ldots, J\}$. We want to estimate π and every θ_j. Each θ_j is invariant through time, meaning that each month the criteria keep the same importance for the user. The permutation π

changes each month however, due to the variation of important products according to the criteria. So, each month, we have J permutations picked from the Mallows models, with still the same θ_j but a different π. In this Section, the Kendall distance is considered.

3.1 Estimations of π

Knowing, for $j \in \{1, \dots, J\}$, all the σ_j and their associated dispersion parameter θ_j for a month, we can estimate the permutation π associated to this month. We use the maximum likelihood estimator method. The likelihood function is given using (1) by

$$L(\pi|\boldsymbol{\sigma}, \boldsymbol{\theta}) = \prod_{j=1}^{J} \frac{\exp(-\theta_j \, d(\sigma_j, \pi))}{Z(\theta_j)}. \tag{8}$$

As $Z(\theta_j)$ does not depend on π and the exponential function is monotonous, the maximum likelihood estimator $\hat{\pi}$ is given by

$$\hat{\pi} = \underset{\pi \in \mathcal{S}_n}{\operatorname{argmax}} \sum_{j=1}^{J} -\theta_j \, d(\sigma_j, \pi) \tag{9}$$

where $\mathcal{S}_n$ is the set of every possible n items permutations.

We use a method based on Blin idea [2], to find the median of J permutations. First, we need to initialize $\hat{\pi}$. We use for this initialization the permutation σ_j associated with the biggest θ_j. Then, we compute the matrix of discrepancy between $\hat{\pi}$ and σ_j, that we name $G_{\hat{\pi},\sigma_j}$. The matrix $G_{\hat{\pi},\sigma_j}$ is a $n \times n$ symmetric matrix with 0 on the diagonal since an item cannot be inverted with itself. Each item of this discrepancy matrix between $\hat{\pi}$ and σ_j, $G_{\hat{\pi},\sigma_j,a,b}$ equal 1 if items a and b are in a different ordered between $\hat{\pi}$ and σ_j and 0 otherwise. We create each discrepancy matrix in the same order as the initialization of $\hat{\pi}$ in order to sum them. Finally, this matrix is multiplied by θ_j.

Each item of our discrepancy matrix $G = \sum_{j=1}^{J} G_{\hat{\pi},\sigma_j}$ takes its value x between 0 and $\sum_{j=1}^{J} \theta_j$. Furthermore, if we invert two elements of our permutation π, then the new value of the element in our discrepancy function x_{new}, is

$$x_{new} = \sum_{j=1}^{J} \theta_j - x. \tag{10}$$

While an inversion of two items can diminish the sum of elements of G, we move items in $\hat{\pi}$. When there is no more possible moves, we stop and return $\hat{\pi}$, the estimation of π.

With this method, we face a few issues. First, we can have a lot of permutations maximizing the likelihood function. But the main issue appears if our parameters θ_j are too different from each other. If there is j such as $\sum_{i=1,i\neq j}^{J} \theta_i - \theta_j < 0$ then, the maximum likelihood is given for the permutation $\pi = \sigma_j$. So, we cannot have any compromise. To solve this issue, we can introduce, to estimate π the following minimization problem

$$\hat{\pi} = \underset{\pi \in \mathcal{S}_n}{\operatorname{argmin}} \sum_{j=1}^{J} (\theta_j \; d(\sigma_j, \pi))^2. \tag{11}$$

Using (11), we need to modify our algorithm. To this end, we use the matrix $G_{\hat{\pi},\sigma_j}$ and modify $\hat{\pi}$ while we can reduce $\sum_{j=1}^{J} (\theta_j \; d(\sigma_j, \hat{\pi}))^2$ where $d(\sigma_j, \hat{\pi})$ is the sum of each element of $G_{\hat{\pi},\sigma_j}$, corresponding to the distance between $\hat{\pi}$ and σ_j.

We can also use the Borda [3] method to find the permutation π. In this method, we compute a weighted sum of each element position and sort the results in the ascending order. For our example with the same weight to each permutation, the best compromise between, the mackerels fishing nets, the salmon's lures and the fishing rods permutation and the fishing rods, the mackerels fishing nets and the salmon's lures permutation, is the mackerels fishing nets (1+2), the fishing rods (3+1) and finally the salmon's lures (2+3) permutation. This method is significantly faster than the previous ones. A lot of different algorithms were presented and tested in [1].

Algorithm 1 Estimate π

Require: $\sigma_1, \ldots, \sigma_J \in \mathcal{S}_n^J$ and $\theta_1, \ldots, \theta_J \in \mathcal{R}^J$
  $\hat{\pi} \leftarrow \sigma_j$ where $j = \operatorname{argmax}_{j \in 1,\ldots,J} \theta_j$
  for $\{j \in 1, \ldots, J\}$ **do**
    for $\{i \in 1, \ldots, n\}$ **do**
      for $\{k \in i, \ldots, n\}$ **do**
        if $(\hat{\pi}(i) > \hat{\pi}(k) \cap \sigma_j(i) < \sigma_j(k)) \cup (\hat{\pi}(i) < \hat{\pi}(k) \cap \sigma_j(i) > \sigma_j(k))$ **then**
          $G_{\hat{\pi},\sigma_j,i,k} \leftarrow \theta_j$
        else
          $G_{\hat{\pi},\sigma_j,i,k} \leftarrow 0$
        end if
      end for
    end for
  end for
  $G \leftarrow \sum_{j \in J} G_{\hat{\pi},\sigma_j}$
  G_{new} is an actualization of G with one element moved in $\hat{\pi}$, $\sum G_{new} \leq \sum G$
  We stop the algorithm when there is no more $\sum G_{new} \leq \sum G$

3.2 Estimation of θ

The estimator of θ is discussed in Fligner and Verducci [5], among others. To estimate the θ parameter in the Mallows model, we use the moment generative function of the random variable $D(\pi, \sigma)$, the distance between π and σ, in the uniform case, when $\theta = 0$.

Theorem 1. *Let $P_\theta(\sigma)$ defined by (1) be the distribution of σ, with π the modal permutation parameter and θ the dispersion parameter, and let $D(\pi, \sigma)$ the random variable of the right invariant distance $d(\pi, \sigma)$ between π and σ. Then the expected mean of the random variable $D(\pi, \sigma)$ for $t = -\theta$ is given by*

$$E_\theta(D) = \frac{\partial}{\partial t} \log(M_{D,0}(t)), \tag{12}$$

and the theoretical variance of the random variable $D(\pi, \sigma)$ for $t = -\theta$ is

$$Var_\theta(D) = \frac{\partial^2}{\partial t^2} \log(M_{D,0}(t)). \tag{13}$$

Proof of Theorem 1. In the sequel, we use D instead of $D(\pi, \sigma)$. The moment generative function of D for $\theta = 0$ is given by

$$M_{D,0}(t) = \sum_{\sigma \in \mathcal{S}_n} \exp(t\ d(\sigma, \pi)) P_0(\sigma), \tag{14}$$

Where $P_\theta(\sigma)$ is defined by (1) and corresponds to the probability to select σ. As given in Section 2.1, every permutation has the same probability when $\theta = 0$. We obtain

$$M_{D,0}(t) = \sum_{\sigma \in \mathcal{S}_n} \exp(t\ d(\sigma, \pi)) \frac{1}{n!}. \tag{15}$$

The normalization term $Z(\theta)$ given in (2) can be written using the moment generative function of D for $\theta = 0$ and $t = -\theta$:

$$Z(\theta) = n!\ M_{D,0}(-\theta). \tag{16}$$

Finally, we express the moment generative function of D in θ by

$$M_{D,\theta}(t) = \sum_{\sigma \in \mathcal{S}_n} \exp(t\ d(\sigma, \pi)) P_\theta(\sigma). \tag{17}$$

The probability of σ knowing θ is given by (1) and using (16) on

$$M_{D,\theta}(t) = \sum_{\sigma \in \mathcal{S}_n} \exp(t\ d(\sigma, \pi)) \frac{\exp(-\theta\ d(\sigma, \pi))}{Z(\theta)} = \sum_{\sigma \in \mathcal{S}_n} \frac{\exp((t - \theta)\ d(\sigma, \pi))}{Z(\theta)}, \tag{18}$$

we obtain

$$M_{D,\theta}(t) = \sum_{\sigma \in \mathcal{S}_n} \frac{\exp((t-\theta)\ d(\sigma,\pi))}{n!\ M_{D,0}(-\theta)} = \frac{1}{M_{D,0}(-\theta)} \sum_{\sigma \in \mathcal{S}_n} \frac{\exp((t-\theta)\ d(\sigma,\pi))}{n!} \quad (19)$$

where we recognize the expression (15) with $t = t - \theta$. Finally, we have

$$M_{D,\theta}(t) = \frac{M_{D,0}(t-\theta)}{M_{D,0}(-\theta)}. \quad (20)$$

The moment generative function properties allow us to calculate the mathematical expectation and the theoretical variance of D by a simple derivative. The mathematical expectation knowing θ of $D(\pi, \sigma)$ is given for $t = 0$ by

$$E_\theta(D) = \frac{\partial}{\partial t} M_{D,\theta}(t). \quad (21)$$

We know with (20) the link between the moment generative function with $\theta \neq 0$ and the moment generative function where $\theta = 0$, the case where every permutation σ has the same probability. We can then replace $M_{D,\theta}(t)$ by its equivalent with $M_{D,0}(t)$ in $t = -\theta$

$$E_\theta(D) = \frac{1}{M_{D,0}(-\theta)} \frac{\partial}{\partial t} M_{D,0}(t). \quad (22)$$

We recognize in the mathematical expectation $E_\theta(D)$ that the numerator is the derivative of the denominator and thus for $t = -\theta$ we obtain (12)

$$E_\theta(D) = \frac{\partial}{\partial t} \log(M_{D,0}(t)).$$

The variance is given in $t = 0$ by

$$Var_\theta(D) = \frac{\partial^2}{\partial t^2} M_{D,\theta}(t) - \left(\frac{\partial}{\partial t} M_{D,\theta}(t) \right)^2. \quad (23)$$

Using (20), we have in $t = -\theta$

$$Var_\theta(D) = \frac{\frac{\partial^2}{\partial t^2} M_{D,0}(t)}{M_{D,0}(-\theta)} - \left(\frac{\frac{\partial}{\partial t} M_{D,0}(t)}{M_{D,0}(-\theta)} \right)^2 = \frac{M_{D,0}(-\theta) \frac{\partial^2}{\partial t^2} M_{D,0}(t) - \left(\frac{\partial}{\partial t} M_{D,0}(t) \right)^2}{M_{D,0}(-\theta)^2}. \quad (24)$$

We recognize here the second derivative of the logarithm. Then for $t = -\theta$, we obtain (13)

$$Var_\theta(D) = \frac{\partial^2}{\partial t^2} \log(M_{D,0}(t)).$$

□

However, even with those simplifications with $\theta = 0$, the calculations of the mathematical expectation and the theoretical variance are still time consuming due to the sum of $n!$ terms. As previously, we use the Kendall distance. This distance helps us to simplify the calculations of the mathematical expectation and the theoretical variance of $D(\pi, \sigma)$.

Theorem 2. *Under the same assumptions of Theorem 1 with $d(\pi, \sigma)$ the Kendall distance. Then the expected mean of the random variable $D(\pi, \sigma)$ is given by*

$$E_\theta(D) = \frac{n \exp(-\theta)}{1 - \exp(-\theta)} - \sum_{j=1}^{n} \frac{j \exp(-j\theta)}{1 - \exp(-j\theta)}, \tag{25}$$

and the theoretical variance of the random variable $D(\pi, \sigma)$ is

$$Var_\theta(D) = \frac{n \exp(-\theta)}{(1 - \exp(-\theta))^2} - \sum_{j=1}^{n} \frac{j^2 \exp(-j\theta)}{(1 - \exp(-j\theta))^2}. \tag{26}$$

Proof of Theorem 2. If we can rewrite the random variable $D(\sigma, \pi)$ as

$$\sum_{j=1}^{p} S_j(\sigma\pi^{-1}, e), \tag{27}$$

where each $S_j(\sigma\pi^{-1}, e)$ is independent for $\theta = 0$, then we have the following property between $M_{S_j,0}(t)$, the moment generative functions of $S_j(\sigma\pi^{-1}, e)$, and $M_{D,0}(t)$, the moment generative function of $D(\sigma, \pi)$:

$$M_{D,0}(t) = \prod_j M_{S_j,0}(t). \tag{28}$$

The Kendall distance (6), as shown in Section 2.2, can be rewritten by $d(\sigma, \pi) = \sum_{j=1}^{n-1} S_j(\sigma\pi^{-1}, e)$ where each $S_j(\sigma\pi^{-1}, e)$ is independent from the others. As the Kendall distance is right invariant, we can use $\pi = e$ where e is the identity permutation. The product of the moment generative functions $S_j(\sigma, e)$ is given by

$$\prod_{j=1}^{n-1} M_{S_j,0}(t) = \prod_{j=1}^{n-1} \frac{1}{n!} \sum_{\sigma \in \mathcal{S}_n} \exp(t\, S_j(\sigma, e)). \tag{29}$$

No matter the permutation $\sigma \in \mathcal{S}_n$, $S_j(\sigma, e)$, the minimum number of inversion to assign the j^{th} element in the first remaining position of σ, takes its values in $\{0, \ldots, n - j\}$. As we are in the case where each permutation has the same probability, we need to count all the possibilities for each value of $S_j(\sigma, e)$. Each

value of $S_j(\sigma, e)$ is obtained by the same number of permutations $\frac{n!}{n-j+1}$. Then, we can write

$$\prod_{j=1}^{n-1} M_{S_j,0}(t) = \prod_{j=1}^{n-1} \frac{1}{n-j+1} \sum_{i=0}^{n-j} \exp(t\, i). \tag{30}$$

As we can calculate easily the sum of the geometric sequence $\sum_{i=0}^{n-j} \exp(t\, i)$ and using (30), the derivative of $E_\theta(D)$ given in (12) in $t = -\theta$ is

$$\frac{\partial}{\partial t} \log(M_{D,0}(t)) = \frac{\partial}{\partial t} \log \left(\prod_{j=1}^{n-1} \frac{1}{n-j+1} \frac{1-\exp((n-j+1)\, t)}{1-\exp(t)} \right). \tag{31}$$

Calculating our derivative, we have with $t = -\theta$

$$\frac{\partial}{\partial t} \log(M_{D,0}(t)) = \frac{(n-1)\exp(t)}{1-\exp(t)} - \sum_{j=1}^{n-1} \frac{(n-j+1)\exp((n-j+1)\, t)}{1-\exp((n-j+1)\, t)}. \tag{32}$$

We can rewrite (32) as

$$\frac{\partial}{\partial t} \log(M_{D,0}(t)) = \frac{(n-1)\exp(t)}{1-\exp(t)} - \sum_{j=2}^{n} \frac{j\exp(jt)}{1-\exp(jt)}. \tag{33}$$

Finally, for $j = 1$, both parts of the equation are the same and we obtain (25)

$$E_\theta(D) = \frac{n\exp(-\theta)}{1-\exp(-\theta)} - \sum_{j=1}^{n} \frac{j\exp(-j\theta)}{1-\exp(-j\theta)}$$

However, this last equation is not defined for $\theta = 0$, which can induce some difficulties during the optimization process. To stop those issues, we can keep the sum of exponential. In this case, we have

$$E_\theta(D) = \sum_{j=1}^{n-1} \frac{\sum_{i=0}^{n-j} i\exp(-i\theta)}{\sum_{i=0}^{n-j} \exp(-i\theta)}. \tag{34}$$

To calculate the variance of $D(\sigma, \pi)$ in θ, we determine the derivative of (25) for $t = -\theta$

$$\frac{\partial}{\partial t} E_{-t}(D) = \frac{\partial^2}{\partial t^2} \left(\frac{n\exp(t)}{1-\exp(t)} - \sum_{j=1}^{n} \frac{j\exp(jt)}{1-\exp(jt)} \right). \tag{35}$$

So with

$$\frac{\partial^2}{\partial t^2}\frac{n\exp(t)}{1-\exp(t)} = n\frac{\exp(t)(1-\exp(t)) - \exp(t)(-\exp(t))}{(1-\exp(t))^2} = \frac{n\exp(t)}{(1-\exp(t))^2} \quad (36)$$

and

$$\frac{\partial^2}{\partial t^2}\frac{j\exp(jt)}{1-\exp(jt)} = \frac{j^2\exp(jt)(1-\exp(jt)) - j\exp(jt)(-j\exp(jt))}{(1-\exp(jt))^2} = \frac{j^2\exp(jt)}{(1-\exp(jt))^2}, \quad (37)$$

we obtain (26)

$$Var_\theta(D) = \frac{n\exp(-\theta)}{(1-\exp(-\theta))^2} - \sum_{j=1}^{n}\frac{j^2\exp(-j\theta)}{(1-\exp(-j\theta))^2}.$$

□

For a number of months M, we have all the σ_m, for $m \in \{1,\ldots,M\}$, each one of them selected according to a Mallows model with a known modal permutation π_m, not necessarily the same each month, and an unknown dispersion parameter θ, always the same, that we want to find. We then estimate θ by the maximum likelihood method. We have the following log-likelihood function for θ

$$\log(L(\boldsymbol{\sigma}|\theta,\pi)) = \sum_{m=1}^{M}\log(P_\theta(\sigma_m)) = \sum_{m=1}^{M}\log\left(\frac{\exp(-\theta\ d(\sigma_m,\pi))}{Z(\theta)}\right) \quad (38)$$

and so

$$\log(L(\boldsymbol{\sigma}|\theta,\pi)) = \sum_{m=1}^{M}\left(-\theta\ d(\sigma_m,\pi) - \log(Z(\theta))\right). \quad (39)$$

Using (16), we obtain

$$\log(L(\boldsymbol{\sigma}|\theta,\pi)) = \sum_{m=1}^{M}\left(-\theta\ d(\sigma_m,\pi) - \sum_{i=1}^{n}\log(i) - \log(M_{D,0}(-\theta))\right). \quad (40)$$

To maximize the likelihood over the parameter θ, we differentiate the log-likelihood with respect to θ

$$\frac{\partial}{\partial\theta}\log(L(\boldsymbol{\sigma}|\theta,\pi)) = \sum_{m=1}^{M}\left(d(\sigma_m,\pi) - \frac{\partial}{\partial\theta}\log(M_{D,0}(-\theta))\right). \quad (41)$$

Then, the maximum likelihood estimator $\hat{\theta}$ of θ is solution of

$$\frac{\partial}{\partial\theta}\log(M_{D,0}(-\theta)) = \frac{1}{M}\sum_{m=1}^{M} d(\sigma_m, \pi). \tag{42}$$

Using the Kendall distance, with (12) and (25), we have the maximum likelihood estimator $\hat{\theta}$ of θ the solution of the following equation :

$$\frac{n\exp(-\hat{\theta})}{1-\exp(-\hat{\theta})} - \sum_{j=1}^{n}\frac{j\exp(-j\hat{\theta})}{1-\exp(-j\hat{\theta})} = \frac{1}{M}\sum_{m=1}^{M} d(\sigma_m, \pi) \tag{43}$$

The Fisher information of the θ parameter is obtained differentiating (42) with respect to θ

$$I(\theta) = E\left(\frac{\partial^2}{\partial\theta^2}\log(L(\boldsymbol{\sigma}|\theta,\pi))\right) = E\left(\frac{\partial^2}{\partial\theta^2}M_{D,0}(-\theta)\right). \tag{44}$$

With the Kendall distance, we use (13) and (26), and obtain

$$I(\theta) = \frac{n\exp(-\theta)}{(1-\exp(-\theta))^2} - \sum_{j=1}^{n}\frac{j^2\exp(-j\theta)}{(1-\exp(-j\theta))^2}. \tag{45}$$

This second derivative does not depend on the observations σ_m for $m \in \{1,\dots,M\}$. The asymptotic behavior of the maximum likelihood estimator $\hat{\theta}$ of θ is given by the central limit theorem.

Theorem 3. *Under regularity conditions, we have when M tend to infinity :*

$$\sqrt{M}\left(\hat{\theta}-\theta\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, I(\theta)^{-1}\right) \tag{46}$$

where $I(\theta)$ is the Fisher information of θ given in (45).

As said in the introduction however, the real permutation π is really difficult to obtain. Our best option can give us a partial permutation on a smaller set of element. The formulas (25) and (26) only stands in the case π is a permutation, a ranking without any tie. In a case of partial rankings, one needs to adapt its distance to take the ties into account.

4 New method

When we use a criterion based on a numerical continuous value, we need to use the difference between two values to determine the cost of an inversion so a transformation of the difference is applied. For example, if our Normandy sales director

has a reference of lures for the trout with a difference between the expected and the real revenue of 1 euro and a reference for the salmon's lures with a difference between the expected and the real revenue of 101 euros, then our Normandy sales director will clearly be more interested to analyze the salmon's lures reference than the trout's lures reference. However, if the difference between the expected revenue and the real revenue for the trout's lures is 1000 euros and the difference for the salmon's lures reference is 1100 euros, the Normandy sales director will only be slightly more interested to analyze the salmon's lures reference. So in the first case, the cost, for this criterion, of the inversion between the two references must be higher than the cost of the inversion of the two references in the second case.

As it is really difficult to have labeled data, non parametric statistical method is difficult to use. Moreover, all the information we have are on the final result : "Is this reference interesting or not?" based on the user's opinion. This function's parameters and the Mallows model dispersion parameters θ need to be computed at the same moment. As a consequence, a logit transformation is used and we assign the weight of an inversion at the value of the difference between two references after this transformation. With this transformation, the Mallows model corresponds to the special case where each weight is the same.

We still use the same formalism as before but we do not assume that σ_j for $j \in 1, \ldots, J$, with J the number of criteria used, is selected randomly based on the the modal permutation π, the theoretical best ranking for the user. We determine the probability that π is the best ranking for the user. We have the values of each σ_j, the ranking obtained for the criterion j and its associated vector $x_j \in \mathbb{R}^n$ where n is the number of items to order. Then, we give the probability of a permutation π based on a modal permutation σ_j and the matrix of inversions costs $\boldsymbol{\theta}_j$ associated to the criterion j with

$$P_{\sigma_j,\boldsymbol{\theta}_j}(\pi) = \frac{\exp(-d_p(\sigma,\pi))}{Z(\sigma_j,\boldsymbol{\theta}_j)} \tag{47}$$

where $d_p(.,.)$ is a weighted distance between two permutations of $\mathcal{S}_n$ and

$$Z(\sigma_j,\boldsymbol{\theta}_j) = \sum_{\tau\in\mathcal{S}_n} \exp(-d_p(\sigma_j,\tau)), \tag{48}$$

the normalization term influenced by σ_j due to a change of values for each inversion and $\boldsymbol{\theta}_j$ but does not depend on π.

We use as $d_p(.,.)$ a variation of the Kendall distance given by

$$d_{pk}(\pi,\sigma) = \sum_{i=1}^{n-1}\sum_{k>i} \theta_{jik}\, \mathbb{1}\left((\pi(i) > \pi(k) \cap \sigma(i) < \sigma(k)) \cup (\pi(i) < \pi(k) \cap \sigma(i) > \sigma(k))\right) \tag{49}$$

where θ_{jik} is the difference of the transformations

$$\theta_{jik} = f(x_j(i)) - f(x_j(k)). \tag{50}$$

The value of the item i for the criterion j is given by $x_j(i)$. Here, we use the logit function for the transformation

$$f(x_j(i)) = \frac{\alpha}{1 + \exp(r(a - x_j(i)))} + \beta, \tag{51}$$

with β the minimum value after the transformation, $\alpha + \beta$ the maximum value after the transformation, r the velocity parameter to go from the minimum to the maximum and a the symmetrical point. These 4 parameters α, β, r and a fully determine all the θ_{jik}. Learning them give the matrix $\boldsymbol{\theta}_j$.

4.1 Estimation of π

To choose π, assuming that all the criteria are independent. We maximize the function

$$\Psi(\boldsymbol{\sigma}, \boldsymbol{\theta}_j, \pi) = \prod_{j=1}^{J} P_{\sigma_j, \boldsymbol{\theta}_j}(\pi), \tag{52}$$

where $P_{\sigma_j, \boldsymbol{\theta}_j}(\pi)$ is defined in (47). Then, taking the logarithm of $\Psi(\boldsymbol{\sigma}, \boldsymbol{\theta}_j, \pi)$, we have

$$\log(\Psi(\boldsymbol{\sigma}, \boldsymbol{\theta}_j, \pi)) = \sum_{j=1}^{J} -d_p(\pi, \sigma_j) - \sum_{j=1}^{J} \log\left(\sum_{\tau \in \mathcal{S}_n} \exp(-d_p(\tau, \sigma_j)\right). \tag{53}$$

As the second part of the equation (53) does not depend on π, we can ignore it and maximize

$$\log(\Psi(\boldsymbol{\sigma}, \boldsymbol{\theta}_j, \pi)) = \sum_{j=1}^{J} -d_p(\pi, \sigma_j). \tag{54}$$

To maximize this value, we use the method described in Section 3.1 to maximize (9). In this variation, we can reinitialize $G_{\hat{\pi}, \sigma_j, ik}$ with

$$G_{\hat{\pi}, \sigma_j, ik} = \theta_{jik} \mathbb{1}\left((\pi(i) > \pi(k) \cap \sigma(i) < \sigma(k)) \cup (\pi(i) < \pi(k) \cap \sigma(i) > \sigma(k))\right). \tag{55}$$

4.2 Estimation of $\boldsymbol{\theta}$

We assume independence between all the criteria. With this independence, maximizing the probability $P_{\sigma_j,\boldsymbol{\theta}_j}(\pi)$, for any $j \in 1,\ldots,J$, maximize the function $\Psi(\boldsymbol{\sigma},\boldsymbol{\theta}_j,\pi)$. Figure 2 gives a graphical representation of what we want to do. In this representation, we want to minimize the cost of the red links and maximize the cost of the black links. Thus, we can estimate the parameter $\boldsymbol{\theta}_j$ for any criterion j independently of the others. We have, for a number of months M, π_m, the wanted permutation considering the criteria. This permutation is the best for the user for the month m. We also know the real permutations σ_{jm}, associated to their real values x_{jm} for each month $m \in \{1,\ldots,M\}$. Also, $\boldsymbol{\theta}_j$ is the same for each month. As $\boldsymbol{\theta}_j$ is defined with f, we need to obtain the parameters α, β, r and a. To determine those values, we use the maximum likelihood estimators. The likelihood function is given by :

$$L(x_j,\boldsymbol{\sigma}_j,\boldsymbol{\pi};a,\alpha,\beta,r) = \prod_{m=1}^{M} \frac{\exp(-d_{pk}(\sigma_{jm},\pi_m))}{Z(a,\alpha,\beta,r,x_j,\sigma_{jm})}, \tag{56}$$

where $d_{pk}(.,.)$ is defined in (49). Thus, the log likelihood function is given by

$$\log(L(x_j,\boldsymbol{\sigma}_j,\boldsymbol{\pi};a,\alpha,\beta,r)) = \sum_{m=1}^{M} -d_{pk}(\sigma_{jm},\pi_m) - \sum_{m=1}^{M} \log(Z(a,\alpha,\beta,r,x_j,\sigma_{jm})). \tag{57}$$

In this case, not as before, the time complexity of $Z(a,\alpha,\beta,r,x_j,\sigma_{jm})$ cannot be reduced in $O(n)$. It is really expensive to compute as the time complexity is in $O(n!)$. For this reason, it is necessary to approximate $Z(a,\alpha,\beta,r,x_j,\sigma_{jm})$. We know how to determine this function with the generalized Mallows model, when all the inversions of the first element with its followings are the same cost, the second with its followings are the same cost as well and so on. To approximate, we can bound the value of $Z(a,\alpha,\beta,r,x_j,\sigma_{jm})$ by using the minimum and maximum value of the transposition cost of an element with its followings. A central value of this bounding, as the mean, is often a satisfying approximation of $Z(a,\alpha,\beta,r,x_j,\sigma_{jm})$.

In the case of the Kendall distance, we have the possibility to separate the distance in a sum of independent distances $d_k(\pi,\sigma) = \sum_{i=1}^{n-1} S_k(\pi,\sigma)$ as described in (7).

Using this decomposition, we have for an approximation of $Z(\boldsymbol{\theta}_j,\sigma_{jm})$ using a unique value of the inversion cost of an element with its followings

$$Z_{approx}(\boldsymbol{\theta}_j,\sigma_{jm}) = \sum_{\tau\in\mathcal{S}_n} \exp\left(\sum_{k=1}^{n-1} -\theta_k\ S_k(\tau,\sigma_{jm})\right), \tag{58}$$

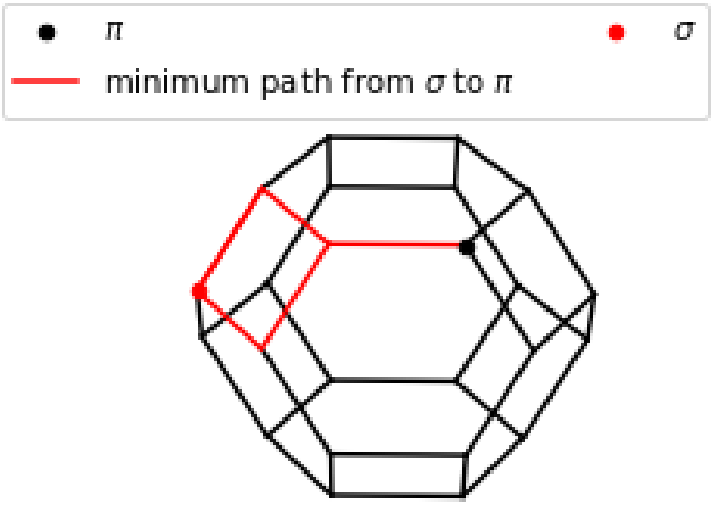

Figure 2: Permutohedron of order 4 for the Kendall Distance.

where θ_k is the cost to invert the k^{th} element with its followings. As the $S_k(\tau, \sigma_{jm})$ are independent, we can, for every permutation of $\mathcal{S}_n$, combine every possible number of inversion of the first element, with every possible number of inversion of the second element, with every possible number of inversion of every other element. We obtain

$$Z_{approx}(\boldsymbol{\theta}_j, \sigma_{jm}) = \prod_{k=1}^{n-1} \sum_{i=0}^{n-k} \exp(-i\,\theta_k). \tag{59}$$

Finally, the logarithm of the approximation of the normalization term is given by

$$\log(Z_{approx}(\boldsymbol{\theta}_j, \sigma_{jm})) = \sum_{k=1}^{n-1} \log(1 - \exp(-(n-k)\,\theta_k)) - \log(1 - \exp(-\theta_k)). \tag{60}$$

5 Simulation studies

The estimators of π in the Mallows model are studied through simulation. The estimators considered are described in Section 3.1. The real parameter θ_j for each permutation σ_j, for $j \in 1, \ldots, J$ is fixed. Each σ_j is selected following the Mallows

model with the parameters π and θ_j. We evaluate $\hat{\pi}$, the estimator of π, using two different criteria: the distance between $\hat{\pi}$ and π must be the lowest as possible and the estimator $\hat{\pi}$ must be a compromise of every σ_j, for $j \in 1, \ldots, J$. We consider different combinations of θ_j to describe the behavior of the estimators. The real value of π is not important to know as it doesn't affect the results due to the right invariant property of the distance used. In this simulation, the Kendall distance (6) is used.

For the results, we denote the estimator of π by MLE the Maximum Likelihood Estimator method given in (9), by QUAD the QUADratic sum of the distances method given in (11) and by BORDA the BORDA method [3].

The simulation results given in Table 1 compare the three methods for 50 σ_j, permutations with $n = 30$ items, each one associated to the same θ over 3000 realizations. The Kendall distance distance between $\hat{\pi}$ and π is considered and the means and standard deviations results on the distance are given. This distance is between 0, the minimum value, and $\frac{n(n-1)}{2}$, the maximum value. The best results are for a distance of 0 meaning that the two permutations on n items are the same, and a distance of $\frac{n(n-1)}{2}$ as $\hat{\pi}$ is the unique exact inverse permutation of π. The worst distance value is $\frac{n(n-1)}{4}$ corresponding to an uninformative random noise. For $n = 30$, the maximum value for the Kendall distance is 435 and the worst value is 217.5.

In Table 1, we observe that the BORDA method gives the best estimation results for π considering the distance to π criterion for the lowest θ parameters. The lowest θ parameters values, the less information contains the criteria associated to the permutations σ_j. The BORDA method gives in our simulations, the best results in the sense of the minimum variance comparing to the MLE and QUAD methods for the lowest parameters θ. The maximum likelihood estimator method is better than the BORDA method when $\theta \geq 0.1$ and the difference with the BORDA method for the distance to π is more important when θ increases. For the standard deviation, the maximum likelihood estimator has a smaller value than the BORDA method when $\theta \geq 0.25$. Finally, the QUAD estimator is only slightly worst, less than 1 adjacent inversion, than the MLE one but the gap between the two never really change too much.

In Table 2, we consider the same simulation data but we study a compromise of every σ_j, for $j \in 1, \ldots, J$. To evaluate this information, the Kendall distance between each σ_j and the estimator $\hat{\pi}$ is considered. The standard deviations of those distances reflect the harmony between the use of all σ_j in the decision of a compromise $\hat{\pi}$. The lower this value, the best our estimators for compromises.

In Table 2, we observe that the QUAD method is clearly the best for this simulation. The QUAD method is better than the maximum likelihood estimator no matter the value of θ and better than the BORDA method until $\theta = 0.5$ when

Choice of θ	BORDA	MLE	QUAD
0.01	166.19 ± 26.269	168.03 ± 26.491	168.447 ± 26.287
0.05	63.08 ± 10.843	64.61 ± 11.530	65.09 ± 11.590
0.1	31.13 ± 6.050	30.65 ± 6.456	31.05 ± 6.525
0.15	19.43 ± 4.298	18.10 ± 4.505	18.41 ± 4.610
0.2	13.36 ± 3.410	11.55 ± 3.439	11.80 ± 3.538
0.25	9.53 ± 2.769	7.77 ± 2.722	8.02 ± 2.739
0.3	7.16 ± 2.351	5.30 ± 2.234	5.52 ± 2.281
0.35	5.35 ± 2.054	3.70 ± 1.851	3.80 ± 1.879
0.4	4.025 ± 1.785	2.54 ± 1.557	2.70 ± 1.608
0.45	3.08 ± 1.569	1.79 ± 1.269	1.89 ± 1.295
0.5	2.22 ± 1.367	1.18 ± 1.045	1.30 ± 1.089
0.55	1.65 ± 1.189	0.79 ± 0.876	0.87 ± 0.898
0.6	1.23 ± 1.060	0.54 ± 0.712	0.62 ± 0.779

Table 1: Means $\pm$ standard deviations of Kendall distances between π and $\hat{\pi}$ for 50 σ_j, permutations with $n = 30$ items, with the same θ and the same π over 3000 realizations.

the average values are close in values. However, as seen previously in Table 1 for $\theta \geq 0.5$, the QUAD method give better results than the BORDA method with a lower distance between $\hat{\pi}$ and π.

For this criterion however, the results of the maximum likelihood estimator, even though they are the worst, are the most similar to the results obtained with the real value of π instead of any of its estimators (results not shown).

We show in Figure 3 the proximity to π criterion, when the number of criteria used is reduce to 5, still keeping the same value for all θ. We do not observe a significant difference in the distances with the 50 criteria simulation. The BORDA method is better than the maximum likelihood estimator method when $\theta \leq 0.35$ and the differences between every estimator are lower than 3 adjacent inversions. It is really interesting however to study the differences on the compromise criterion.

In Table 3, we observe the value of the compromise criterion for the different estimators. We observe that the maximum likelihood estimator is clearly outplayed by the others and that the QUAD estimator is clearly better than the others. When we reduce the number of permutations σ_j, for $j \in \{1, \ldots, J\}$, taken into account, it is obvious that the maximum likelihood estimator cannot match the QUAD estimator performance on this criterion as we approach the tyranny described in the end of Section 3.1.

In Table 4 and Figure 4, the mean and the standard deviations of the Kendall distance between π and $\hat{\pi}$ are given for different values of J, the number of cri-

Choice of θ	BORDA	MLE	QUAD
0.01	27.40 ± 2.785	27.86 ± 2.801	25.21 ± 2.571
0.05	26.17 ± 2.616	26.64 ± 2.646	25.13 ± 2.492
0.1	23.95 ± 2.449	24.38 ± 2.491	23.46 ± 2.388
0.15	21.15 ± 2.180	21.55 ± 2.210	20.90 ± 2.158
0.2	18.46 ± 1.929	18.79 ± 1.972	18.31 ± 1.918
0.25	16.17 ± 1.686	16.42 ± 1.707	16.07 ± 1.668
0.3	14.21 ± 1.509	14.41 ± 1.532	14.15 ± 1.505
0.35	12.67 ± 1.331	12.83 ± 1.350	12.63 ± 1.329
0.4	11.38 ± 1.196	11.51 ± 1.210	11.35 ± 1.195
0.45	10.31 ± 1.116	10.42 ± 1.118	10.30 ± 1.108
0.5	9.43 ± 0.999	9.52 ± 1.005	9.42 ± 0.994
0.55	8.65 ± 0.920	8.72 ± 0.927	8.65 ± 0.920
0.6	8.02 ± 0.847	8.07 ± 0.848	8.02 ± 0.844

Table 2: Mean $\pm$ standard deviation of the standard deviation of the Kendall distances between σ and $\hat{\pi}$ for 50 σ_j, permutations of $n = 30$ items, with the same θ and the same π over 3000 realizations.

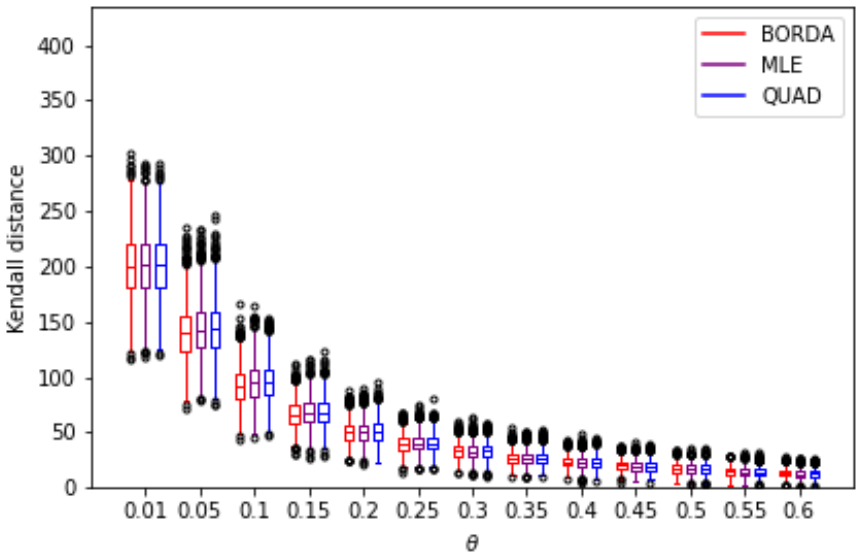

Figure 3: Boxplot of the Kendall distances between $\hat{\pi}$ and π for 5 σ_j, permutations of $n = 30$ items, with the same θ and the same π over 3000 realizations.

Choice of θ	BORDA	MLE	QUAD
0.01	21.43 ± 7.759	27.08 ± 9.761	16.76 ± 6.101
0.05	20.07 ± 7.318	25.20 ± 8.963	16.32 ± 6.013
0.1	17.42 ± 6.425	22.04 ± 8.096	15.05 ± 5.618
0.15	15.23 ± 5.623	19.43 ± 7.126	13.79 ± 5.096
0.2	13.00 ± 4.823	16.75 ± 6.070	12.30 ± 4.554
0.25	11.38 ± 4.322	14.47 ± 5.468	10.92 ± 4.170
0.3	9.89 ± 3.721	12.57 ± 4.702	9.57 ± 3.674
0.35	8.86 ± 3.315	11.12 ± 4.144	8.63 ± 3.226
0.4	8.00 ± 2.924	9.98 ± 3.616	7.91 ± 2.909
0.45	7.29 ± 2.745	9.04 ± 3.338	7.25 ± 2.647
0.5	6.63 ± 2.417	8.13 ± 3.009	6.61 ± 2.403
0.55	6.21 ± 2.281	7.58 ± 2.762	6.23 ± 2.175
0.6	5.77 ± 2.117	7.00 ± 2.573	5.79 ± 2.011

Table 3: Mean $\pm$ standard deviation of the standard deviation of the Kendall distances between σ and $\hat{\pi}$ for 5 σ_j, permutations of $n = 30$ items, with the same θ and the same π over 3000 realizations.

teria used. All the criteria have the same importance. The value of θ and n are respectively 0.3 and 30. The BORDA method is better than the two others in the simulation, for less than 7 permutations to use. We observe that the maximum likelihood estimator is the worst when we only have two permutations. In this particular case, we are in the tyranny due to the limitations of the method. However, the maximum likelihood estimator become the best as the number of permutations used increases. Concerning the QUAD estimator, the results are only slightly worst than the maximum likelihood estimator on the proximity to π criterion.

The Table 5 represents the compromise criterion for the same set of parameters than the Table 4. The tyranny of the maximum likelihood estimator when we only have two permutations to take a decision, is also shown by this criterion. Here, we observe also the perfect compromise with the QUAD estimator, which does not necessarily lead to the best result as it is 12 adjacent inversions away in average from the BORDA method for the proximity to π criterion. We also notice that all the methods tend to the same level of compromise when the number of permutations used to find $\hat{\pi}$ is increased. For less than 10 permutations however, the QUAD estimator is better, especially when J is even. When J is odd, the compromise is not as good.

We are now interested to evaluate the performances of the three methods with different associations of parameters to find witch one is the more robust to every

Choice of J	BORDA	MLE	QUAD
2	51.52 ± 10.300	69.13 ± 14.736	63.59 ± 12.554
3	42.51 ± 8.530	43.76 ± 9.118	43.83 ± 9.074
4	37.00 ± 7.409	39.49 ± 8.238	38.81 ± 7.908
5	32.73 ± 6.766	32.91 ± 7.222	33.10 ± 7.117
6	29.97 ± 6.166	30.65 ± 6.768	30.40 ± 6.661
7	27.42 ± 5.642	26.66 ± 5.921	26.94 ± 5.929
8	25.37 ± 5.320	25.20 ± 5.752	25.27 ± 5.797
9	23.77 ± 5.037	22.79 ± 5.371	23.12 ± 5.307
10	22.07 ± 4.728	21.45 ± 5.157	21.53 ± 5.137
15	17.33 ± 3.992	15.79 ± 4.178	16.25 ± 4.244
20	14.29 ± 3.588	12.78 ± 3.635	13.02 ± 3.648
25	12.23 ± 3.255	10.46 ± 3.398	10.82 ± 3.314
30	10.66 ± 3.055	8.95 ± 2.976	9.15 ± 3.044
40	8.57 ± 2.655	6.82 ± 2.504	7.05 ± 2.570
50	7.07 ± 2.431	5.29 ± 2.204	5.49 ± 2.242

Table 4: Means $\pm$ standard deviations of Kendall distances between π and $\hat{\pi}$ for 2 to 50 σ_j, permutations of $n = 30$ items, with the same $\theta = 0.3$ and the same π over 3000 realizations.

possibilities. We determine the effect of adding low informative criteria into the decision. We are still comparing on two different objectives. The distance between $\hat{\pi}$ and π does not need any change to adapt for this kind of tests but the compromise criterion cannot be taken the same way as before. Obviously, the distance between $\hat{\pi}$ and σ strongly depends on θ. As θ increases, we have lower distances between $\hat{\pi}$ and σ. Then, we adapt the criterion using the expected distance between π and σ knowing θ given for the Kendall distance with (25). We then have, for the compromise criterion, to study

$$C_2 = \sqrt{Var(d(\hat{\pi}, \sigma) - E_\theta(\sigma))}. \tag{61}$$

We want to study what happen when we use a different mix of very informative criteria, with $\theta = 0.6$, and low informative criteria with $\theta = 0.05$. In Table 6, we study the Kendall distance between $\hat{\pi}$ and π for those cases, with $n = 30$. We notice that for two parameters $\theta = 0.6$, the set of parameters P_3, the BORDA method is clearly the best and the maximum likelihood estimator is the worst due to the tyranny problem. We observe particularly for the set of parameters P_5 that adding a lot of low informative criteria strengthen the results for the maximum likelihood estimator. However, the BORDA and QUAD method decrease the quality of the estimation while adding low informative criteria. The set of parameters P_4 shows

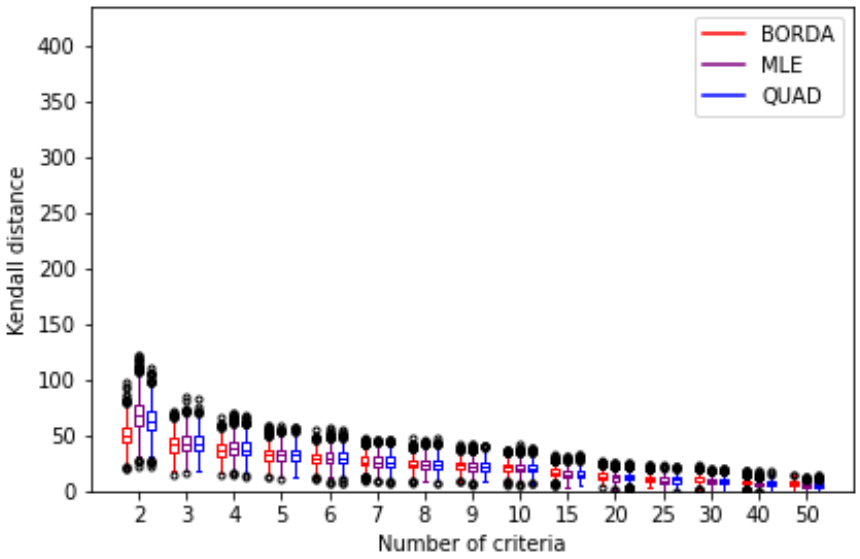

Figure 4: Kendall Distance between $\hat{\pi}$ and π for different numbers of σ_j, permutations of $n = 30$ items, with the same $\theta = 0.3$ and the same π over 3000 realizations.

that adding only a small amount of low informative criteria has the same effect as adding a lot of low informative criteria, like in the set of parameters P_5, for the BORDA method and the maximum likelihood estimator. This set of parameters P_4 also shows better results for the QUAD estimator than only very informative criteria in the set of parameters P_3. A mix with a lot of low informative criteria can have better results for the maximum likelihood estimator than the others methods but using only very informative criteria with the BORDA method can show a better precision in the estimation of π.

We need to be careful when we have a lot of very informative criteria. With the set of parameters P_6 and P_7, we witness that with 25 criteria having a parameter $\theta = 0.6$, adding 25 criteria with a parameter $\theta = 0.05$ does not increase the results, even with the maximum likelihood estimator. However, for the maximum likelihood estimator, the loss in precision is not really a problem as it is only 0.02 more adjacent inversions in average to reach π and this estimator is better than the BORDA and QUAD methods on this criterion anyway.

We use for the Table 7 the same sets of parameters than in the Table 6 but for the compromise criterion using (61). With this criterion, we cannot compare the different lines as different values of θ will also change the variance around the expected value of the distance, as it is well shown in Figure 5. This Figure, for the set of parameter P_4 shows also the main difference between the QUAD method

Choice of J	BORDA	MLE	QUAD
2	2.40 ± 1.845	50.24 ± 8.921	0.25 ± 0.250
3	7.01 ± 3.675	11.74 ± 6.181	7.55 ± 4.217
4	8.81 ± 3.824	13.38 ± 5.848	5.09 ± 3.924
5	9.92 ± 3.670	12.63 ± 4.659	9.66 ± 3.640
6	10.73 ± 3.554	13.64 ± 4.446	8.89 ± 3.733
7	11.25 ± 3.440	13.10 ± 3.899	10.93 ± 3.208
8	11.74 ± 3.224	13.47 ± 3.670	10.56 ± 3.262
9	12.16 ± 3.221	13.57 ± 3.581	11.77 ± 2.949
10	12.22 ± 3.083	13.57 ± 3.408	11.38 ± 3.116
15	13.16 ± 2.575	13.95 ± 2.724	12.83 ± 2.448
20	13.58 ± 2.233	14.16 ± 2.331	13.28 ± 2.204
25	13.77 ± 2.054	14.23 ± 2.127	13.57 ± 2.016
30	13.84 ± 1.867	14.21 ± 1.930	13.69 ± 1.855
40	14.07 ± 1.618	14.33 ± 1.658	13.97 ± 1.609
50	14.22 ± 1.482	14.41 ± 1.498	14.14 ± 1.480

Table 5: Mean $\pm$ standard deviations of the standard deviation of the Kendall distances between σ and $\hat{\pi}$ for 2 to 50 σ_j, permutations of $n = 30$ items, with the same θ and the same π over 3000 realizations.

Choice of θ	BORDA	MLE	QUAD
P_1	63.08 ± 10.843	64.61 ± 11.530	65.09 ± 11.590
P_2	1.23 ± 1.060	0.54 ± 0.712	0.62 ± 0.779
P_3	22.83 ± 5.811	32.67 ± 8.107	30.90 ± 7.436
P_4	25.20 ± 5.661	27.38 ± 6.716	29.76 ± 6.395
P_5	26.94 ± 5.456	22.74 ± 5.631	33.13 ± 6.380
P_6	3.32 ± 1.625	2.04 ± 1.374	2.25 ± 1.437
P_7	3.51 ± 1.664	2.06 ± 1.372	2.70 ± 1.564

Table 6: Mean $\pm$ standard deviations of Kendall distances between π and $\hat{\pi}$ for a set of parameters θ : P_1 : 50 parameters at 0.05 ; P_2 : 50 parameters at 0.6 ; P_3 : 2 parameters at 0.6 ; P_4 : 2 parameters at 0.6 and 5 parameters at 0.05 ; P_5 : 2 parameters at 0.6 and 48 parameters at 0.05 ; P_6 : 25 parameters at 0.6 ; P_7 : 25 parameters at 0.6 and 25 parameters at 0.05. The permutations are of $n = 30$ items.

and the two others. The QUAD method tends to favor low informative criteria more than high informative criteria in comparison with the BORDA method and obviously the maximum likelihood estimator. When coming to the compromise

criterion, the QUAD method gives the best results. However, in the sets of parameters tested here and excepting the special case P_3, the difference between the three methods is not statistically significant.

Parameters θ	BORDA	MLE	QUAD
P_1	26.17 ± 2.616	26.64 ± 2.646	25.13 ± 2.492
P_2	8.02 ± 0.847	8.07 ± 0.848	8.02 ± 0.844
P_3	1.73 ± 1.344	25.12 ± 5.053	0.25 ± 0.25
P_4	20.36 ± 6.735	20.97 ± 6.689	19.53 ± 6.405
P_5	25.99 ± 2.639	26.20 ± 2.657	25.45 ± 2.570
P_6	7.76 ± 1.187	7.91 ± 1.209	7.70 ± 1.177
P_7	19.62 ± 2.511	19.66 ± 2.507	19.58 ± 2.505

Table 7: Mean $\pm$ standard deviations of the standard deviation of the Kendall distances between σ and $\hat{\pi}$ for a set of parameters θ : P_1 : 50 parameters at 0.05 ; P_2 : 50 parameters at 0.6 ; P_3 : 2 parameters at 0.6 ; P_4 : 2 parameters at 0.6 and 5 parameters at 0.05 ; P_5 : 2 parameters at 0.6 and 48 parameters at 0.05 ; P_6 : 25 parameters at 0.6 ; P_7 : 25 parameters at 0.6 and 25 parameters at 0.05. The permutations are of $n = 30$ items.

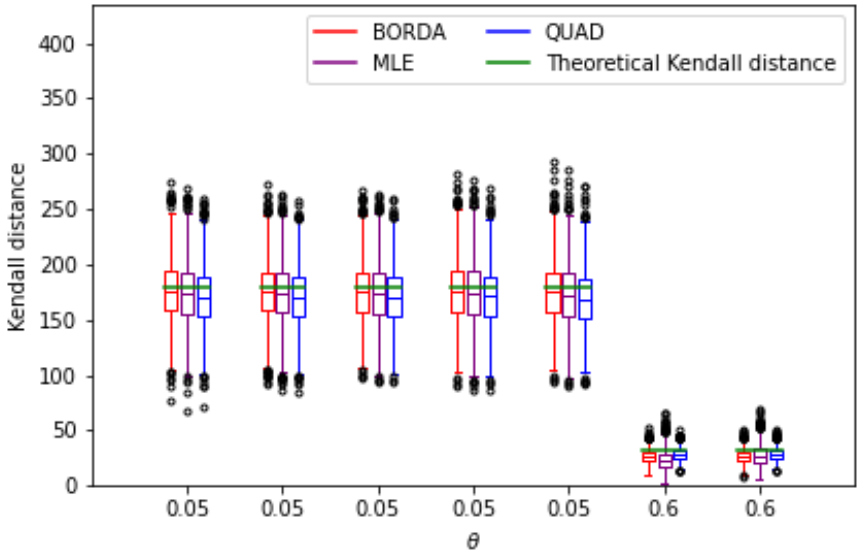

Figure 5: Boxplot of the Kendall distance between $\hat{\pi}$ and σ for $n = 30$, in the set of parameters P_4.

The behavior of our different estimators is studied in two cases: all different values of θ and with a big value of θ equalized by a sum of small values of θ. Those two cases are shown in Table 8, for the criterion of the distance between $\hat{\pi}$ and π, and are denoted respectively by P_8 and P_9. We can see that for this criterion and the set of parameters P_8, the maximum likelihood estimator, as expected, is the best. The surprising part of this Table is with the set of parameters P_9. We would have expected the maximum likelihood estimator to be the worst as with the set of parameters P_3 due to the tyranny. The results for the maximum likelihood estimator are the same as P_3. However, it is by far the best, meaning that the confusion induced by the low informative criteria with $\theta = 0.1$ confuses the two other methods and more significantly the QUAD method.

Choice of θ	BORDA	MLE	QUAD
P_8	23.95 ± 5.075	21.78 ± 5.368	24.74 ± 5.531
P_9	38.01 ± 7.309	32.66 ± 7.921	45.98 ± 8.356

Table 8: Mean $\pm$ standard deviations of the standard deviation of the Kendall distances between σ and $\hat{\pi}$ for a set of parameters θ : P_8 : 8 different parameters, $\theta_1 = 0.1$, $\theta_2 = 0.15$, $\theta_3 = 0.2$, $\theta_4 = 0.25$, $\theta_5 = 0.3$, $\theta_6 = 0.35$, $\theta_7 = 0.4$ and $\theta_8 = 0.6$; P_9 : 1 parameters at 0.6 and 6 parameters at 0.1. The permutations are of $n = 30$ items.

The Figures 6 and 7 represent, for respectively the sets of parameters P_8 and P_9, the distances between $\hat{\pi}$ and σ. We can see in Figure 6 that for the set of parameters P_8, the maximum likelihood estimator over-valuate the criterion with the highest value and under-valuate the others. This is the exact opposite as the QUAD method which tends to under-valuate the criterion with the highest value and over-valuate the others. Here, only the BORDA method seems to respect the importance of every criterion. For the Figure 7, it is essentially the same problems than Figure 6 at a higher level. Here, the maximum likelihood estimator uses the ranking of the criterion associated to the θ with a value of 0.6 as $\hat{\pi}$ without any compromise. This is a tyranny.

6 Application on real data

Now, we want to compare the different models on real data. The data set is composed of items to order for 4 consecutive months. Each month has a different number of items. The first month has 325 items, the second 328 items, the third 331 items and the last 340 items. Those items have to be ordered with respect to 7 different criteria. The 7 criteria are not developed here as this is not important

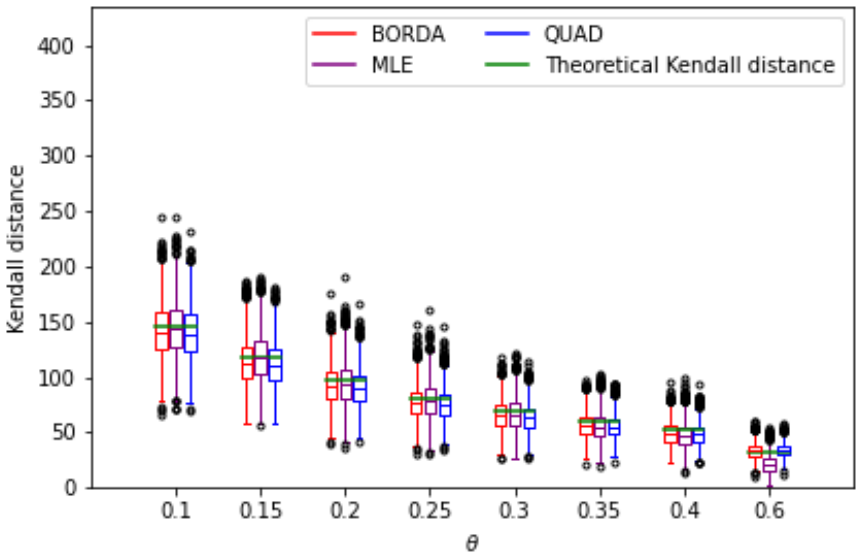

Figure 6: Boxplot of the Kendall distance between $\hat{\pi}$ and σ for $n = 30$, in the set of parameters P_8.

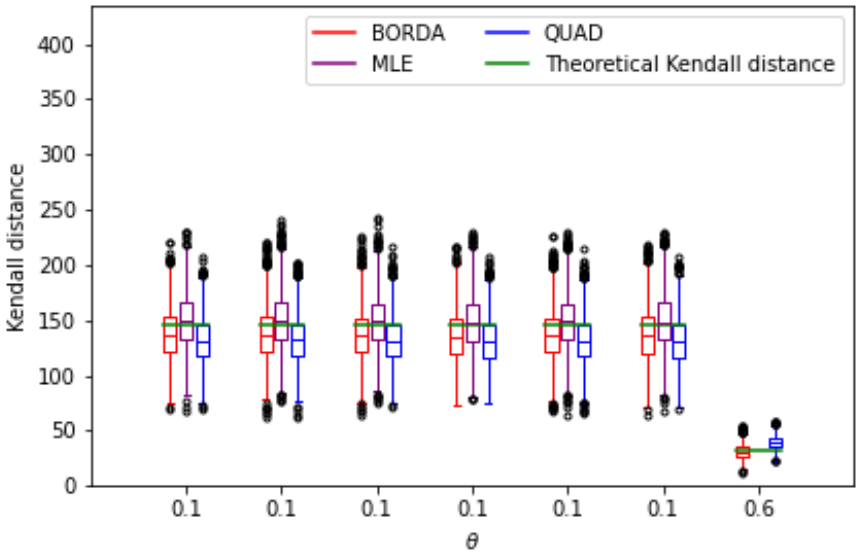

Figure 7: Boxplot of the Kendall distance between $\hat{\pi}$ and σ for $n = 30$, in the set of parameters P_9.

to understand the results. Due to the complexity to get the real value of π, in this experiment, we first use a Mallows model with values of θ given by the user as a first intuition. Considering the results, the user evaluates the 40 best items obtained by this Mallows model considering the maximum likelihood estimator (9) used to find $\hat{\pi}$. The user gives to the 40 items a notation between 0 and 9, where 0 correspond to the best notation, a very interesting product to study and 9 correspond to the worst notation, a useless product to study. Based on these notations, we make a partial ranking of π for those 40 items. So, we learn the parameters θ_j for the Mallows Model (1), denoted as MM, and a, α, β, r for the model given in (47) which uses the real values of each criteria, considering different approximations of $\log(Z(a, \alpha, \beta, r, x_j, \sigma_{jm}))$ in the likelihood (57).

As an approximation of $\log(Z(a, \alpha, \beta, r, x_j, \sigma_{jm}))$, we use 3 possibilities. The MEDian, denoted as MED, the MEAN denoted as MEAN, and MINMAX, the mean of the MINimum and MAXimum values of the logarithm. For the 4 models, we also use the MLE and QUAD estimators of $\hat{\pi}$ to compare the results.

All these models are compared over the 40 evaluated items of each month with the $\hat{\pi}$ obtained and the one given by the user. We compare also the order of the 40 items when we rank only those 40 items and when we rank every item of the month. For 40 items, the maximum Kendall distance is 780.

Month	MM + MLE	MM + QUAD	MINMAX + MLE	MINMAX + QUAD
1	160	181	189	183
2	91	147	121	137
3	128	135	168	168
4	137	173	182	198
Month	MED + MLE	MED + QUAD	MEAN + MLE	MEAN + QUAD
1	135	136	128	133
2	106	106	97	97
3	125	132	123	125
4	132	132	135	134

Table 9
Kendall distance between π and $\hat{\pi}$ when we rank only the 40 evaluated items.

In the Table 9, we compare for every model with the 2 different estimators of π the Kendall distance between π and $\hat{\pi}$. Here, we observe that using the real values, the results are better. The MEAN estimator of the logarithm of the normalization term is the best. Concerning the estimator of π for the MEAN model, both QUAD estimator and Maximum Likelihood estimator give similar results.

An interesting property for a prioritization is the indifference of irrelevant alternative, meaning that if we add an element to rank, then the order found without

this element has to be the same than the order of the same items found with this element. This property does not stand in the case of Mallows models. However, it can be interesting to have a proximity in the orders given for the full ranking and the orders given with the 40 items.

Month	MM + MLE	MM + QUAD	MEAN + MLE	MEAN + QUAD
1	0	40	0	7
2	0	111	0	0
3	0	82	0	0
4	0	108	0	13

Table 10
Kendall distance between $\hat{\pi}$ ranked over every items and $\hat{\pi}$ ranked over the 40 evaluated items, on the 40 evaluated items.

In Table 10, we observe for the 4 most interesting models, the Kendall distance between $\hat{\pi}$ ranked over every items and $\hat{\pi}$ ranked over the 40 evaluated items, on the 40 evaluated items. A distance of 0 may seems interesting but with the Maximum Likelihood Estimator, we have the possibility of a tyranny, which in the case of the Mallows Model case is the tyranny of the third criterion. For the MEAN model with the Maximum Likelihood Estimator, it is not exactly the case as we have a close proximity to the first and third criteria but we do not have the exact same permutation. Having a so close proximity between the 2 different $\hat{\pi}$ with the MEAN model and the QUAD estimator is good.

7 Conclusions

In this article, we focus on a permutation based method to find the best ranking of items to study. We choose the permutation space to use information given in rankings and explain our method to include other types of information given in the value of the variable.

An estimation of the dispersion parameters of the model is given and we propose different algorithms to find the best compromise for the ranking to study, according to the estimation of those parameters. For each estimation of the compromise ranking, the expected advantages and limitations of our choices is explained.

The limitations and the benefits of the three algorithms for the Mallows model is studied on simulated data. The behavior of the estimation of the compromise is studied with respect to the number of ranking to take into account and their dispersion parameter.

Finally, we apply our new model on a real data set and describe our procedure idea to find the dispersion parameters and the best compromise for the user with

a minimal time investment of the user. The results are promising, even though those models use the independence assumption.

Bibliography

[1] Alnur Ali and Marina Meilă. Experiments with kemeny ranking: What works when? *Mathematical Social Sciences*, 64(1):28–40, 2012.

[2] Guillaume Blin, Maxime Crochemore, Sylvie Hamel, and Stéphane Vialette. Median of an odd number of permutations. *Pure Mathematics and Applications*, 21(2):161–175, 2011.

[3] J-C de Borda. Mémoire sur les élections au scrutin: Histoire de l'académie royale des sciences. *Paris, France*, 12, 1781.

[4] Persi Diaconis. Group representations in probability and statistics. *Lecture Notes-Monograph Series*, 11:i–192, 1988. ISSN 07492170. URL `http://www.jstor.org/stable/4355560`.

[5] Michael A Fligner and Joseph S Verducci. Distance based ranking models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 48(3):359–369, 1986.

[6] M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938. ISSN 00063444. URL `http://www.jstor.org/stable/2332226`.

[7] C. L. Mallows. Non-null ranking models. i. *Biometrika*, 44(1/2):114–130, 1957. ISSN 00063444. URL `http://www.jstor.org/stable/2333244`.