Sensitivity index to measure dependence on parameters for rankings and top- k rankings

Antoine Rolland, Jairo Cugliari

ORIGINAL RESEARCH PAPER

Sensitivity index to measure dependence on parameters for rankings and top-k rankings

Antoine Rolland[a] and Jairo Cugliari[a]

[a]ERIC EA 3083, Université de Lyon, Université Lumière Lyon 2, France

ARTICLE HISTORY
Compiled June 16, 2019

ABSTRACT
In a multivariate framework, ranking a data set can be done by using an aggregation function in order to obtain a global score for each individual, and then by using these scores to rank the individuals. The choice of the aggregation function (e.g. a weighted sum) and the choice of the parameters of the function (e.g. the weights) may have a great influence on the obtained ranking. We introduce in this communication a ratio index that can quantify the sensitivity of the data set ranking up to a change of weights. This index is investigated in the general case and in the restricted case of top k rankings. We also illustrate the interest to use such an index to analyze ranked data sets.

KEYWORDS
Ranking, index, sensitivity, top-k list

1. Ranking issues

Ranking items using composite indices is a very common issue that can be faced in several application fields. For example, ranking universities using the well-known Shanghai index [2] or others [16]; ranking countries regarding their development level [7]; ranking various topics in newspapers (like "best place to live, to study,..."); but also in information retrieval as when a web search engine ranks web pages following a specific query, to mention but a few.

The methodology to obtain such rankings often follows the same procedure. First, an appropriate set of evaluation criteria is identified. Then, data are collected in order to obtain a complete multivariate dataset. In Multi Criteria Data Analysis (MCDA) jargon the individuals of this dataset are called alternatives and the variables are called criteria. Last, this information is aggregated into a single score in order to obtain a general ranking which is, mathematically speaking, a total pre-order (i.e. a complete order on the set of individuals possibility including *ex-aequo*).

The choice and construction of the set of variables for a composite index concerns delicate issues about quality, precision and availability of the data are crucial. They can have a great influence on the final ranking. However, in an optimistic point of view, we assume that all the difficulties pointed out above have been overcome in the first steps. In this paper we focus on the last step of the rank construction, i.e. the

CONTACT Email: Antoine.Rolland@univ-lyon2.fr

aggregation process.

A unique synthesized score from several variables can be obtained through the use of an aggregation function. A relatively large number of aggregation functions have been identified in the literature [1, 10, 11]. In this work we choose to study the weighted mean as aggregation function, since it has the advantages of making intuition simpler and computations faster. In a weighted mean the final score, and therefore the ranking, is of course highly dependent on the chosen weights. On the other hand, if an individual a Pareto-dominates an individual b (i.e. a has a better score than b on each variable) then a should necessarily be better ranked than b in the final ranking, no matter what the weights are. The final ranking then appears to depend both on the data and on the aggregation function parameters. If we consider the choice of the variables weights as a political act by the ranking-maker, i.e. a real choice from a human being and not a given parameter, then we can wonder if this choice has a weak or strong influence on the final ranking. This question is equivalent to studying the rank sensitivity with respect to the parameters (i.e. the weights) of the aggregation function.

Several previous works studied the ranking sensitivity up to the data, but not much have explored the ranking sensitivity up to the weights. Among others, in [15] Saisana, Saltelli and Tarantola present a methodology to test the sensitivity of a composite indicator, using sensitivity and uncertainty analysis, and pointed out several sources of uncertainty to obtain a composite ranking. In [12] Permanyer presents descriptive tools that can be seen as a measure of sensitivity of a given ranking with respect to changes in the chosen weighting scheme. In [6] Forster, McGillivray and Seth focus on the relationship between sensitivity and the statistical association between component variables. In [3], D'Agostino and Dardanoni have investigated the problem of measuring social mobility when the social status of individuals is given by their rank. However, in these different works the sensitivity has merely ever been precisely quantified.

We propose in this paper a global index which measures the sensitivity of a ranking with respect to the weights. The index proposed in this paper can be extended to aggregation functions more generalized than the weighted mean, modulo perhaps for computational reasons that are out of the scope of this paper.

The rest of the article is structured as follows. On the next section we give some intuitive examples that ground the construction of our index. Section 3 presents the new Rank Sensitivity Index (RSI) which aims at answering the question raised. In section 4 we study the RSI in the case of top-k lists, i.e. only the first k items of the ranking are a matter of interest. Numerical experiments are detailed in section 5 and we conclude in section 6.

2. Introducing examples

We consider three different examples, each one containing 3 individuals $\mathcal{X}_3 = \{x, y, z\}$ and 3 variables v_1, v_2, v_3. The aggregation function used to determine the ranking by increasing order is a weighted sum, i.e. the individual with the lowest weighted sum of the 3 variables will be ranked first. As shown in table 1 it exists 6 different possible rankings on $\mathcal{X}_3$.

Let us now present three different situations. In situation 1, shown in table 2, x is preferred to y and z for all the variables, and y is preferred to z for all the variables. Whatever the vector of weights is, the global score of x, y and z would always be the same (10 for x, 20 for y and 30 for z), and so the ranking in this situation is x first, y

rank	R_1	R_2	R_3	R_4	R_5	R_6
1	x	x	y	y	z	z
2	y	z	x	z	x	y
3	z	y	z	x	y	x

Table 1. The 6 different possible rankings (in columns) for $\mathcal{X}_3 = \{x, y, z\}$

second and z third. The weights have no influence on the final ranking.

	v_1	v_2	v_3
x	10	10	10
y	20	20	20
z	30	30	30

Table 2. Situation 1: In presence of an absolute winner the ranking is fully determined by data

In situation 2, as shown in table 3, the 3 individuals have totally symmetric evaluations on the 3 variables. So each of the 6 rankings presented in table 1 is equally possible and the choice of the weights totally determine the final ranking. The ranking is very sensitive to a change of weights.

	v_1	v_2	v_3
x	10	30	20
y	20	10	30
z	30	20	10

Table 3. Situation 2: The ranking is fully determined by the weight vector.

Finally, we analyse situation 3, shown in table 4. The rank of x is always 1, whereas rank of y is 2 in half of the cases and 3 in the other half (same for z). The final ranking is then dependent of the chosen weights, but in a less sensitive way than situation 2, as only 2 out of 6 possible rankings can be obtained. This last example shows that a quantitative index of the respective contributions of parameters choice and variables values can give an insight on the ranking sensitivity with respect to the parameters.

This last example shows that a quantitative index of the respective contributions of parameters choice and variables values can give an insight on the ranking sensitivity with respect to the parameters. We introduce in the next section such an index based on an analysis of variance approach.

3. Ranking Sensitivity index (RSI)

Following [1], we formally introduce the aggregation functions as follows:

	v_1	v_2	v_3
x	10	10	20
y	20	30	30
z	30	20	30

Table 4. Situation 3: the ranking will be determined by both data and parameters.

Definition 3.1. [1] Let $\mathbb{I} = [a, b]$ be an interval of $\mathbb{R}$. An aggregation function f is a function of $p > 1$ arguments that maps the p-dimensional cube $\mathbb{I}^p$ onto $\mathbb{I}$, $f : \mathbb{I}^p \mapsto \mathbb{I}$, with the properties:

(1) $f(\underbrace{a, a, \ldots, a}_{p\ times}) = a$ and $f(\underbrace{b, b, \ldots, b}_{p\ times}) = b$
(2) $\forall j = 1, \ldots, p$, $x_j \leq y_j \Longrightarrow f(x) \leq f(y)$ $\forall x = (x_1, \ldots, x_p), y = (y_1, \ldots, y_p) \in \mathbb{I}^p$.

For the sake of simplicity, and as already mentioned, we focus in this paper only on weighted mean as aggregation function. The study of the Ranking Sensitivity Index for other aggregation function is out of the scope of this paper and will be developed in future works.

Definition 3.2. We denote $\mathcal{W}$ the simplex in $\mathbb{R}^p$, i.e. the set of probability vectors of p positive real values $\mathcal{W} = \{(w_1, \ldots, w_p) \in \mathbb{R}^p | w_j \geq 0, j = 1, \ldots p; \sum_{j=1}^{p} w_j = 1\}$.

Definition 3.3. A weighted mean f_w, $w \in \mathcal{W}$ is an aggregation function $f_w : \mathbb{I}^p \mapsto \mathbb{I}$ such that, $\forall x = (x_1, \ldots, x_p)$,

$$f_w(x) = \sum_{j=1}^{p} w_j x_j = < w, x > .$$

A specific weighted mean depends on a vector of parameters $w = (w_1, \ldots, w_p)$.

We also consider a multicriteria data set containing m individuals (alternatives) each described by a set of p variables (criteria). Let x_{ij} be the value of the j-th criterion on the i-th individual, with $i = 1, \ldots, m$ and $j = 1, \ldots, p$. We use the weighted mean f_w to determine a preference ranking for the m considered individuals, with respect to the values taken on the variables and the weight vector w. The question we try to ask is the following: given a set of individuals and the weighted mean family $\{f_w, w \in \mathcal{W}\}$, what is the influence of the parameters w on the final ranking? In other words, is the final ranking sensitive to a modification of weights in the aggregation function?

Our goal is to answer this question through the use of a global index measuring the sensitivity of a ranking on a specific data set with respect to a change of the aggregation function parameters. We suppose in the following that a specific multicriteria data set containing m individuals is given and denoted $\mathcal{X}_m$. The Rank Sensitivity Index (RSI) of $\mathcal{X}_m$ is based on an analysis of variance approach. Two sources of variation are to be considered:

(1) the values taken by each individual x on each variable $j \in \{1, \ldots, p\}$,
(2) the parameter set w for the aggregation function.

In order to measure how different choices of the vector $w \in \mathcal{W}$ induces different rankings on the m individuals, we propose in section 3.1 to introduce RSI in the finite case and in section 3.2 in the general case.

3.1. Rank Sensibility Index : the finite case

	R_1	R_2	R_3	...	R_n	...
x_1	1	1	5	...	2	...
x_2	2	4	2	...	5	...
x_3	3	2	1	...	4	...
...	...	...	...	...	...	...
x_m	5	3	3	...	1	...

Table 5. Example of different rankings on a set $\mathcal{X}_m$ of m individuals

We consider in this section that the weight vectors w are from an arbitrary finite subset of $\mathcal{W}$ of size n that we call $\mathcal{W}_n = \{w^1, \dots, w^n\} \subset \mathcal{W}$. Each vector $w^l \in \mathcal{W}_n$ leads to a possibly different ranking R_l on the m individuals of $\mathcal{X}_m$. Table 5 presents an example of such a situation where rankings are in columns. Such a table, for a subset $\mathcal{W}_n \subset \mathcal{W}$ contains rich information about the influence that a weight vector has on the specific rank of an individual. Intuitively, the lower the variability across the rows, the less influence the parameters have on the ranking.

The classical analysis of variance theory (ANOVA) is then a natural framework to study the influence of the weights on the global ranking. We note r_{il} the rank that f_{w^l} attributes to the i^{th} individual, $i = 1, \dots, m$, with the use of weights set $w^l \in \mathcal{W}_n$. The global variability for a vectors set $\mathcal{W}_n$ is expressed as the sum of square deviations (SSD) of the rankings r_{il}, $i \in \{1, \dots, m\}$, $l \in \{1, \dots, n\}$,

$$SSD_{\text{total}}^{\mathcal{X}_m, \mathcal{W}_n} = \sum_{i=1}^{m} \sum_{l=1}^{n} (r_{il} - \bar{r}_{..})^2,$$

where $\bar{r}_{..} = (nm)^{-1} \sum_{i=1}^{m} \sum_{l=1}^{n} r_{il}$ is the global mean. Applied to our framework, the $SSD_{\text{total}}^{\mathcal{X}_m, \mathcal{W}_n}$ can be split into two factors: the variability due to the parameters ($SSD_{\text{parameters}}^{\mathcal{X}_m, \mathcal{W}_n}$) and the intrinsic variability of the individual ranks that does not depends on the parameters ($SSD_{\text{intrinsic}}^{\mathcal{X}_m, \mathcal{W}_n}$). As in usual ANOVA, we can compute the part of variance which is due to parameters variation, and the part of the intrinsic variance, i.e. the rank variability that is not sensitive to a change of weights.

The total Sum of Squared Deviations (SSD) can easily be computed. It only depends of n an m and is equal to

$$SSD_{\text{total}}^{\mathcal{X}_m, \mathcal{W}_n} = SSD_{\text{total}}^{n,m} = \frac{n(m-1)m(m+1)}{12}.$$

We define the SSD due to the parameters by:

$$SSD_{\text{parameters}}^{\mathcal{X}_m, \mathcal{W}_n} = \sum_{i=1}^{m} \sum_{l=1}^{n} (r_{il} - \bar{r}_{i.})^2$$

where $\bar{r}_{i.}$ is the mean rank of individual i. The term $\sum_{l=1}^{n} (r_{il} - \bar{r}_{i.})^2$ is therefore the exact sum of square deviations of n different rankings of x_i. For all $i = 1, \dots, m$, let denote var_i^n the exact variance of the ranks r_{ik}, $w_k \in W_n$:

$$var_i^n = n^{-1} \sum_{l=1}^{n} (r_{il} - \bar{r}_{i.})^2$$

Therefore,

$$SSD_{\text{parameters}}^{\mathcal{X}_m, \mathcal{W}_n} = n \sum_{i=1}^{m} var_i^n$$

We define the intrinsic SSD by:

$$SSD_{\text{intrinsic}}^{\mathcal{X}_m, \mathcal{W}_n} = n \sum_{i=1}^{m} (\bar{r}_{i.} - \bar{r}_{..})^2$$

where as above $\bar{r}_{..} = (nm)^{-1} \sum_{i=1}^{m} \sum_{l=1}^{n} r_{il}$ is the global mean and $\bar{r}_{i.}$ is the mean rank of individual i. As in usual ANOVA we have:

$$SSD_{\text{total}}^{\mathcal{X}_m, \mathcal{W}_n} \quad = \quad SSD_{\text{parameters}}^{\mathcal{X}_m, \mathcal{W}_n} + SSD_{\text{intrinsic}}^{\mathcal{X}_m, \mathcal{W}_n}. \tag{1}$$

As in usual ANOVA, we can focus on the ratio $\frac{SSD_{\text{intrinsic}}^{\mathcal{X}_m, \mathcal{W}_n}}{SSD_{\text{total}}^{\mathcal{X}_m, \mathcal{W}_n}}$, which determine the ratio between the intrinsic SSD and the global SSD, and which can be interpreted as the proportion of the global variance which is explained by the intrinsic variance of the data. We have:

$$\frac{SSD_{\text{intrinsic}}^{\mathcal{X}_m, \mathcal{W}_n}}{SSD_{\text{total}}^{\mathcal{X}_m, \mathcal{W}_n}} = 1 - \frac{SSD_{\text{parameters}}^{\mathcal{X}_m, \mathcal{W}_n}}{SSD_{\text{total}}^{\mathcal{X}_m, \mathcal{W}_n}} = 1 - \frac{12n \sum_{i=1}^{m} var_i^n}{nm(m^2-1)}. = 1 - \frac{12 \sum_{i=1}^{m} var_i^n}{m(m^2-1)}.$$

For any data set $\mathcal{X}_m$ and any weight vectors set $\mathcal{W}_n$, we can therefore define a rank sensibility index RSI_n as

$$RSI(\mathcal{X}_m, \mathcal{W}_n) = 1 - \frac{12 \sum_{i=1}^{m} var_i^n}{m(m^2-1)}$$

$RSI(\mathcal{X}_m, \mathcal{W}_n)$ expresses the sensibility of a ranking on $\mathcal{X}_m$ obtained through the use of a weighted mean, with respect to the weight vectors set $\mathcal{W}_n$. We propose in the following section to extend the definition of RSI with respect to $\mathcal{W}$, the set of all possible weight vectors.

3.2. Rank Sensibility Index : the general case

We formally state the RSI. Let us consider a given set of m alternatives $\mathcal{X}_m = \{x_1, \dots, x_m\}$, each one containing an evaluation over p criteria, i.e. $x_i \in \mathbb{R}^p, i = 1, \dots, m$. Then, we define a random variable W taking value on the simplex $\mathcal{W}$. The law of $W$is denoted $\mathcal{L}(W)$ For each alternative, we may define the random weighted averages $Y_i =< W, x_i >$, $i = 1, \dots, m$. We are interested in the ranking of these random weighted averages $\{\rho_i := \text{rank}(Y_i), i = 1, \dots, m\}$.

Under mild conditions, for instance if $\mathcal{L}(W)$ admits an absolute continuous probability density function, we may define the expectation $\bar{\rho}_i = \mathbb{E}_W[\rho_i]$ and the variance $var_i = \mathbb{E}_W[\rho_i - \bar{\rho}_i]^2$. Then, we set two non random quantities that are of interest in what follows. On one side the mean value of the ranks $\bar{\rho} = m^{-1}\sum_{i=1}^m \rho_i = \frac{m+1}{2}$, and on the other side the sum-of-squares of the total variation

$$SSD_{\text{total}}^{\mathcal{X}_m,\mathcal{W}} = \sum_{i=1}^m (\rho_i - \bar{\rho})^2.$$

Notice that each one of the terms on the last expression can be analysed using the famous Huygens decomposition as follows

$$\mathbb{E}_W[\rho_i - \bar{\rho}]^2 = var_i + \mathbb{E}_W[\bar{\rho}_i - \bar{\rho}]^2.$$

We define the RSI index as the ratio of mean square error explained by the variation on the mean ranks, or analogously as the remaining part of variation, i.e.

$$RSI(\mathcal{X}_m, \mathcal{L}(W)) = 1 - \frac{12\sum_{i=1}^m var_i}{m(m^2-1)} \tag{2}$$

A case of particular interest is the case where $\mathcal{L}(W)$ is the uniform distribution on $\mathcal{W}$.

We develop further an estimation procedure of the RSI using n parameters sets. Beforehand, we discuss the examples given in section 2 and the link our index has with the Friedman statistic.

3.3. Examples

Let us analyse the situations introduced above by means of the RSI index. In situation 1, shown in table 2, the ranking variances are $var_x = var_y = var_z = 0$, thus $RSI = 1$. This means that whatever the parameters, the rank of a specific individual is always the same: the final ranking is not sensitive to any change in the parameters set. In situation 2, as shown in table 3, the 3 individuals have totally symmetric evaluations on the 3 variables. So the three variances are identical, $var_x = var_y = var_z$, and coincides with the variance of the ranking $\{1, 2, 3\}$ which is $2/3$. Hence, $RSI = 1 - \frac{12\times(3\times 2/3)}{3\times(9-1)} = 0$. This means that the whole variation comes from the parameters, and therefore that the final ranking is extremely sensitive to a modification of weights. In situation 3, as shown in table 4, the rank of x is always 1, whereas the rank of y is 2 for half of the cases and 3 for the other half. The situation for z is similar. Therefore $var_x = 0$, $var_y = var_z = 0.25$. Hence, $RSI = 1 - \frac{12\times(0+0.25+0.25)}{3\times(9-1)} = 0.25$. The interpretation is that the variance of the final ranking is dependent for $1/4$ on the intrinsic variance of the data and $3/4$ on the choice of the parameters: the ranking is quite sensitive to a change of weights.

3.4. Links with Friedman statistic

Friedman non parametric test (see e.g. [9]) is an independence test on ranks for repeated measures. The test is based on the Friedman statistic

$$F_r = \frac{12}{n \times m \times (m+1)} \sum_{i=1}^{m} S_i^2 - 3n(m+1)$$

with n treatments/blocks and m samples. The quantity S_i is the sum of ranks for each sample. Our RSI_n is linked to F_r as the individuals can be seen as samples and the different parameters set as different treatments on the samples. It is straightforward to see that $n(m-1)RSI_n = F_r$; therefore RSI_n can be interpreted as the mean of the Friedman statistic on the samples. It is well known that under the assumption of independence F_r converges when $n \to \infty$ to a χ^2 law with $m-1$ degrees of freedom. In our framework this means that under the assumption of a total independence of the ranking from the parameters, $n(m-1)RSI_n$ converge to a χ^2 law with $m-1$ degrees of freedom. The exact distribution of F_r under the assumption of partial dependence is not known. Therefore the distribution of observed RSI_n on a sample under the assumption of $RSI = \alpha$, $\alpha \neq 0$ is not known either, and we are unable to estimate a confidence interval of RSI using statistical inference theory. To circumvent this obstacle we turn to an approximated estimate of the distribution.

3.5. Estimation and computational issues

We now tackle the problem of estimating the RSI. For this, our approach is to approximately compute the ranking variances $var_i, i = 1, \ldots, m$. The nature of the possible space of weights generally makes impossible to obtain an exact computation of the values of var_i, except for very simple cases like the ones presented above. Therefore, we have to determine an estimate of RSI. We propose to use a Monte Carlo method to determine RSI as follows. Weight vectors are randomly generated using a Dirichlet distribution of parameters $(1/p, \ldots, 1/p)$ as it is well known that this distribution corresponds to the uniform distribution on the simplex $\mathcal{W}$.

(1) Generate n different sets of weights drawing randomly from the space of weights vectors. Then use these n parameters sets to obtain n different score vectors and so n different rankings on the set of m individuals.
(2) Compute the exact rank variance of each individual var_i^* of the n sampled rankings obtained in the previous step.
(3) Compute the estimation $\widehat{RSI}$ as one minus the ratio of the mean of individual variances divided by the exact total variance:

$$\widehat{RSI} = 1 - \frac{12 \sum_{i=1}^{m} var_i^*}{m(m^2-1)}$$

(4) It is possible to repeat the point estimation to estimate also the variance of the estimator, and then to obtain a confidence interval of $\widehat{RSI}$. We call $\overline{RSI}$ the average value of n estimates of the RSI, and s^2 is the estimated variance associated. Then, a confidence interval of RSI estimator is $]\overline{RSI} - zs/\sqrt{n}; \overline{RSI} + zs/\sqrt{n}[$, with z the normal corresponding value of the chosen quantile (typically $z = 1.96$ for a 95% confidence interval).

Experiments show that the estimator converges very quickly to the RSI when n grows. Typically, simulation of only 500 parameters set leads to very precise estimates of RSI. As an example, figure 1 and table 6 show the variation of RSI through the (estimated) standard deviation obtained by our procedure. For this illustration, the set of individuals is described by 3, 5 or 7 variables, the aggregation function is the weighted mean and the average RSI is around 0.75. The standard deviation is around 10^{-3}.

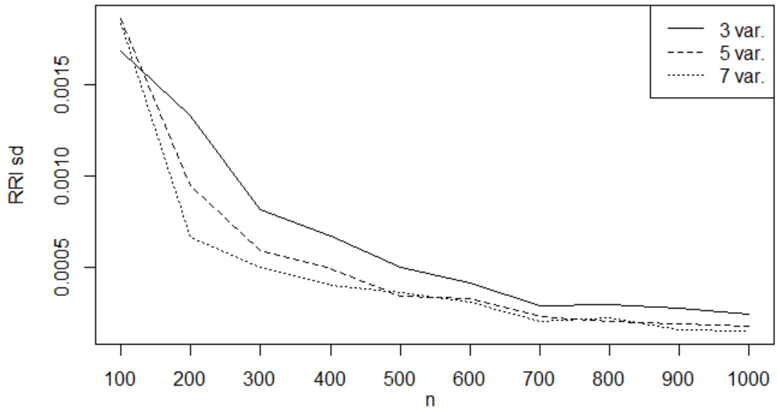


Figure 1. Estimates of the RSI standard deviation as a function of n

n	3 var.	5 var	7 var
100	$1.7\ 10^{-3}$	$1.9\ 10^{-3}$	$1.8\ 10^{-3}$
500	$4.1\ 10^{-4}$	$3.4\ 10^{-4}$	$3.6\ 10^{-4}$
1000	$2.0\ 10^{-4}$	$1.8\ 10^{-4}$	$1.5\ 10^{-4}$

Table 6. Estimates of theRSIstandard deviation as a function of n

3.6. Links with ranking distances

We discuss now alternative approaches to our ANOVA based definition of a rank sensitivity index. An interesting variant could be to use a metric between two rankings (see working papers [13] and [14]). The metric can be for example the Kendall distance or the Spearman's foot-rule distance (see [8] or [4] for an introduction to these metrics). Let us denote by $d(R, R')$ the similarity measure between two ranking R and R'. Now let $RSI_d(f, \mathcal{X})$ be the distance-based Rank Sensitivity Index of set X with respect to the parameters of the function f. The main idea in RSI_d is to compute the ratio of the average distance between two rankings obtained with two different weight vectors, and the average distance between all the possible rankings. Therefore $RSI_d(f, \mathcal{X})$ is defined as follow:

$$RSI_d(f, X) = \frac{\overline{d_{f,W}(R, R')}}{\overline{d}}, \tag{3}$$

where

- $\overline{d} = \mathbb{E}[d(R, R'), R, R' \in X]$ is the average distance between two rankings of X

- $\overline{d_{f,W}(R,R')} = \mathbb{E}[d(R,R'), R, R' \in X_W]$ is the average distance between two rankings obtained by using the aggregation function f and the set of possible parameters W

While appealing from a robust point of view, this approach has poorer properties than the ANOVA-based index proposed above.

- It is not proved that $RSI_d(f,X)$ is always less than 1. First, at the best of our knowledge, there is no theoretical result that proves that $\overline{d} \geq \mathbb{E}[d(R,R'), R, R' \in \mathcal{R}^P]$ where $\mathcal{R}^P$ represents the set of all possible rankings on X, each ranking associated with a probability $p \in P$. Second, the experiments we made show that sometimes we obtain RSI estimations that are greater than 1.
- The $RSI_d(f,X)$ index is very dependent on the distance used. The value obtained for $RSI_d(f,X)$ is not the same using Kendall or Spearmann distance. So it is almost impossible to use $RSI_d(f,X)$ as an absolute index of sensitivity. We can only use it in a comparison framework.
- Therefore it is difficult to have a semantic interpretation of $RSI_d(f,X)$ and to use it to determine the real sensitivity to the parameters of a ranking situation.

4. Top *k*-list and ranking sensitivity index (RSI)

4.1. Top k list definition

Until now we studied the case of an RSI index computed on the whole ranking. But often only the first elements of a ranking are a matter of interest. For example, a page-rank user will be interested in the first 10 results. A newspaper will typically focus on top 3 rankings. Therefore it is interesting to focus on the sensitivity of the top-k ranking to a variation of parameters. In this case, the RSI index will be more in accordance with the ranking user's impression. As a matter of fact, variations in the first (or last) elements of a ranking appear to be more important than variations in the middle of the ranking. Please note that we focus on top-k rankings. Similar results can be obtained for bottom-k rankings.

Let us take for example the situation A (on the left) described in table 7, where the objective is to minimize the value on each variable (the less is better). It is obvious to see that as the individual a get the lowest values on each variable, a will be ranked first independently from the choice of the weights for the weighted average. With a similar reasoning the individual b will be ranked in the second place. As c, d and e have totally symmetric evaluations on the three variables, c, d and e will be equi-probably ranked 3rd, 4th and 5th. Hence, this case gives $var_a = 0$, $var_b = 0$, $var_c = var_d = var_e = 2/3$, which results on a global $RSI = 1 - \frac{12\times(0+0+2/3+2/3+2/3)}{5\times(25-1)} = 0.8$. However, if we focus only on the first two elements of the ranking we obtain $RSI_2 = 1$ (as they are always the same in the same order). This indicates that the top-2 ranking does not depend on the parameters of the aggregation function.

In situation B, described in the same table 7, individuals d and e present the highest two value and so will be at the bottom of the ranking, while a, b and c will be equi-probably ranked as 1st, 2nd and 3rd. In this case, we have obviously the same global value $RSI = 0.8$, as in situation A. However, if we focus only on the first two elements of the ranking (i.e. by giving the same rank 3 to all the individuals not ranked in the first two), we obtain the value $RSI_2 = 0.375$ (see computation below), which shows that the top-2 depends largely on the choice of the parameters.

	c_1	c_2	c_3
a	10	10	10
b	20	20	20
c	30	40	50
d	40	50	30
e	50	30	40

	c_1	c_2	c_3
a	10	20	30
b	20	30	10
c	30	10	20
d	40	40	40
e	50	50	50

Table 7. Top-k analysis, an example through two situations: A (left) and B (right).

Following Fagin *et al.* [5], we use top-k list, i.e. the top-k members of the ordering given by a ranking. Formally speaking a top-k list $\mathcal{R}$ is a bijection from a domain $\mathcal{D}$ (intuitively, the members of the top-k list) to the set $\{1, \dots, k\}$. In our paper, we extend this definition assuming that $\mathcal{D}$ is a subset of a discrete and possibly infinite set N of size $n \in \mathbb{N} \cup +\infty$. In order to formalize this presentation of a top-k list into a set of n elements, we then choose what is called the *optimistic approach* in [5] and assume that all the $n-k$ elements that are not in the top k list are ranked *ex aequo* at the $k+1-$th position. Therefore, a top-k list $\mathcal{R}$ is a bijection from N to the set $\{1, 2, 3..., k, k+1, ..., k+1\}$.

4.2. Computing RSI on top-k lists

RSI for top k lists has the same definition as the one proposed in equation 2, that is the ratio between the sum of squares due to the intrinsic variance and the total one. For top-k list, the only difference is on the specific form taken for the quantities $SSD_{\text{intrinsic}}$ and SSD_{total}.

Proposition 4.1. *RSI for top k lists (RSI_k) can be expressed as:*

$$RSI_k = \frac{SSD^{(k)}_{intrinsic}}{SSD^{(k)}_{total}} = 1 - \frac{6\sum_{i=1}^{m} var_i}{k(k+1)\left(k + (k+1)(1 - \frac{3k}{2m})\right)}. \tag{4}$$

Proof. The mean of the set $\{1, 2, \dots, k, k+1, \dots k+1\}$ is $\frac{(k+1)(m-k/2)}{m}$ and its variance can be computed as the mean of squares minus the squared mean. The mean of squares is

$$\frac{1}{m}\left(\frac{k(k+1)(2k+1)}{6} + (k+1)^2(m-k)\right),$$

and the square of the mean is

$$\frac{(k+1)^2(m-k/2)^2}{m^2}.$$

Thus, the wanted variance is

$$\begin{aligned} var_{tot} &= \frac{1}{m}\left(\frac{k(k+1)(2k+1)}{6} + (k+1)^2(m-k)\right) - \frac{(k+1)^2(m-k/2)^2}{m^2} \\ &= \frac{k(k+1)}{6m}\left(k + (k+1)(1 - \frac{3k}{2m})\right). \end{aligned}$$

Therefore as $SSD^{(k)}_{\text{total}} = SSD^{(k)}_{parameters} + SSD^{(k)}_{\text{intrinsic}}$ we have that the RSI for top-k lists can be written as:

$$\begin{aligned} RSI_k &= \frac{SSD^{(k)}_{\text{intrinsic}}}{SSD^{(k)}_{\text{total}}} \\ &= 1 - \frac{SSD^{(k)}_{\text{parameters}}}{SSD^{(k)}_{\text{total}}} \\ &= 1 - \frac{\sum_{i=1}^{m} var_i}{m \times var_{tot}} \\ &= 1 - \frac{6\sum_{i=1}^{m} var_i}{k(k+1)\left(k + (k+1)(1-\frac{3k}{2m})\right)} \end{aligned}$$

□

As before, the computation of RSI_k needs to determine the variance of the rank of each individual. We use an estimation via a Monte Carlo procedure as in section 3.5.

5. Experiments

We explore in this section three different rankings as study cases where RSI can help to better understand the importance of the weights concerning a ranking. Most of these rankings use normalized scores on each criterion where the best score is automatically set to 100. Although this should be avoided since it can lead to rank reversals, we are not interested here in the criticism of the procedure to obtain each ranking. We focus on the fact that these rankings are obtained by a weighted average over several criteria, which is exactly the situation where RSI is useful. In what follows, all the estimates are obtained using simulated samples of 1000 parameters vectors, which allows one to obtain an estimated value of RSI_k as proposed in section 3.5. Repeating this estimation 100 times gives also an estimate of the variance of RSI_k. The estimates of RSI_k are the mean of 100 estimations of RSI_k. We present results for the the top-k positions with $k \in \{1, 2, 3, 5, 10, m\}$.

5.1. Three different situations

We present here the computations concerning three real-life cases of rankings. Results and discussion are postponed until section 5.2. Graphs representing RSI_k distributions for each situation can be found in appendix.

5.1.1. OECD's countries ranking

In May 2011, the Organization for Economic Co-operation and Development (OECD) proposed a new well-being index named "Better Life Index" (BLI)[1]. The BLI evaluates the 34 states members of the OECD. It uses 11 criteria labelled as housing, income, education, governance, etc. Each criterion is evaluated on a scale ranging between 0 and 10. A global country score is obtained by a weighted mean of the criteria.

[1] http://www.oecdbetterlifeindex.org/

As emphasized by OECD, the innovative aspect of the BLI is the possibility offered to anyone to choose her/his own weights (weights are integers between 0 and 5) in order to represent her/his own preferences on well-being indicators: “The OECD is NOT deciding what makes for better lives. YOU decide for yourself.” A study of this index was proposed in [7], and enlightened the fact that the final ranking was poorly dependant of the selected parameters, i.e. whatever the parameters are, it is always one of the same three countries which is ranked at the first place.

We present in table 8 the estimation of the Rank sensitivity Index (and its standard deviation σ_{RSI}) for OECD’s Better Life index (data obtained in may 2012).

k	1	2	3	5	10	34
RSI_k	0.245	0.345	0.422	0.571	0.770	0.927
σ_{RSI}	0.010	0.010	0.009	0.008	0.004	0.001

Table 8. OECD’s Better Life Index *RSI* estimation.

5.1.2. “Where-to-study” cities ranking

The French magazine “L’Etudiant” is specialized in issues concerning student life, student jobs, student housing and particularly student guidance. It publishes each year a ranking of 41 French cities to determine “where is the best city to study in France”. It does not rank universities, but the cities where the universities are installed. For this, 9 criteria are considered including for example life quality, housing possibilities, employment possibilities or international environment. Each city is evaluated on these 9 criteria, and the scores are averaged into a global score which is used to rank the cities. We present in table 9 the estimation of the Rank sensitivity Index for ”Where to study”.

k	1	2	3	5	10	20	41
RSI_k	0.237	0.361	0.438	0.527	0.623	0.692	0.720
σ_{RSI}	0.010	0.011	0.012	0.011	0.009	0.010	0.007

Table 9. “Where to study” cities ranking *RSI* estimation

5.1.3. Universities ranking

Since the famous Shanghai ranking appeared in 2003, many others organisms proposed a world universities ranking, all based more or less on the same scheme, i.e. evaluating the universities by giving a score on several criteria, and then averaging these scores on a global value. The Academic Ranking of World Universities (ARWU - best known as ”Shanghai ranking”[2]) considers *every university that has any Nobel Laureates, Fields Medalists, Highly Cited Researchers, or papers published in Nature or Science. In addition, universities with significant amount of papers indexed by Science Citation Index-Expanded (SCIE) and Social Science Citation Index (SSCI) are also included.* The ranking is obtained by averaging normalized score on 6 criteria with specific weights. The QS World University Ranking[3] has been also chosen for our study as the data are easily available, and the ranking is obtained by averaging the score on 6 criteria which weights are given in the methodology section of the website. We restrict our study to the 20 first universities in the final ranking.

[2] `http://www.shanghairanking.com/ARWU2017.html`

[3] `http://www.topuniversities.com/qs-world-university-rankings`

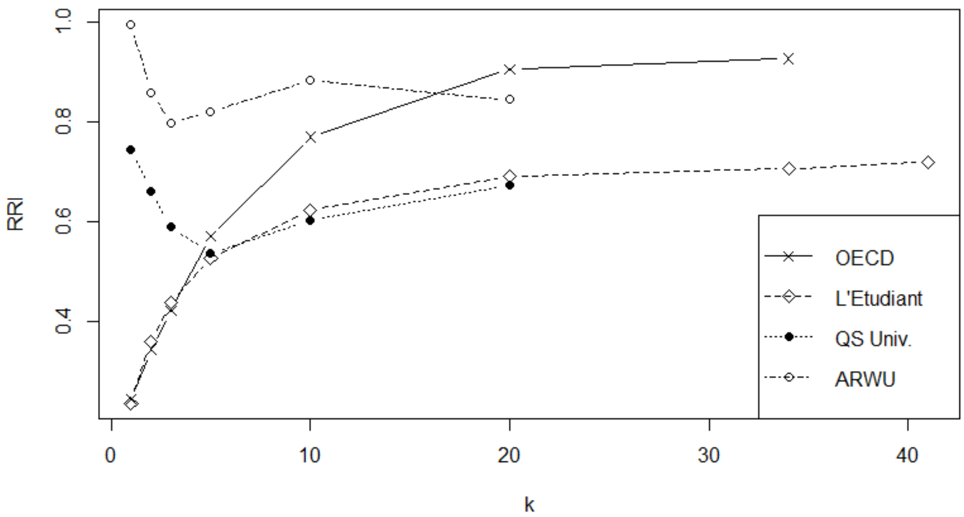


Figure 2. RSI_k for three different example rankings.

We present in table 10 some estimation of the Rank Sensitivity Index for ARWU (dated 2017) and QS World University Ranking (dated 2018).

	k	1	2	3	5	10	20
ARWU	RSI_k	0.995	0.858	0.798	0.820	0.884	0.845
	σ_{RSI}	0.003	0.002	0.003	0.004	0.003	0.004
QS	RSI_k	0.744	0.662	0.590	0.536	0.603	0.673
	σ_{RSI}	0.018	0.012	0.009	0.005	0.004	0.005

Table 10. Universities ARWU and QS ranking RSI estimation.

5.2. Discussion

Results of the precedent section are plotted in figure 2. As we can see, the RSI changes according to k. In OECD's index, the RSI increases with k, which means that the first elements of the ranking are more sensitive to a change on the weights than the global ranking, i.e. the top of the ranking is relatively more dependent on the weights. The RSI score of 0.92 for the whole ranking shows the whole ranking is very robust with respect to the parameters, i.e. that countries ranked on the top of the ranking are always ranked in the top, whatever the weights are, and similarly for countries ranked in the bottom of the ranking. The RSI score of 0.57 for the top 5 shows that when focusing on the only first five countries, their ranks (inside the top 5) are more dependent of the weights. The RSI score is even 0.25 for the first place of the ranking, which corresponds to the situation where two different countries can be ranked first with about the same probability of 50% each (see [7] for a detailed study). In this situation 25% of the variance for the first ranked country is given by the data (mainly only two countries can be ranked first) and 75% of the variance is due to the weights of the weighted sum (these two countries can be equally ranked first depending of the weights).

In the "where to study" ranking, the RSI for the whole ranking is 0.72, which means that this ranking can be significantly modified by changing weights. Note that the magazine “L'Etudiant” generally focus on the best 3 cities. The RSI for $k = 3$ is about 0.44, which means that the best 3 ranking reflects at least the political choice from the magazine, through the choice of parameters, than just an objective situation.

In world universities ranking, the RSI is about 0.8 for ARWU and 0.6 for QS. It means that the rankings (and specially the first position) depend marginally of the chosen weights. It means also that the QS ranking is more dependent to the choice of weights than the ARWU ranking, which is less sensitive to the set of weights. Please note that ARWU data are also product of political choices from the ranking builder, and then RSI=1 does not mean that the ranking is the ground truth, but that the ranking isn't sensitive to changes on the weights.

6. Conclusion

We propose in this paper a Ranking Sensitivity Index based on an ANOVA approach. This index is able to determine whether a specific ranking obtained by multicriteria aggregation is robust to a modification of the aggregation operator parameters. The main interest of such an index is to specify if a given ranking is highly dependent on the choice of the parameters or not. As illustrated by the examples, application fields of RSI are various and covers most of the public rankings with available criteria data. Readers and users of these rankings (such as universities rankings) can then measure the influence of choices done by the ranking maker. Of course, RSI is just an index which gives an useful insight when comparing two rankings. It gives no absolute information about a single ranking since the same value can recover several different situations. We recommend using RSI more as a relative index for the ranking comparison, like for instance as we did in the examples to compare ARWU and QS rankings. Experiments have been done only on weighted average. Future work should focus on other aggregation operators like Ordered Weighted Average or Choquet integral.

References

[1] G. Beliakov, A. Pradera, and T. Calvo. *Aggregation Functions: A Guide for Practitioners*. Springer Publishing Company, Incorporated, 1st edition, 2008.

[2] J.-C. Billaut, D. Bouyssou, and Ph. Vincke. Should you believe in the Shanghai ranking? An mcdm view. *Scientometrics*, 84:237–263, 2010.

[3] M. D'Agostino and V. Dardanoni. The measurement of rank mobility. *Journal of Economic Theory*, 144(4):1783–1803, July 2009.

[4] P. Diaconis. Number 11 in ims lecture series. In *Group Representation in Probability and Statistics*. Institute of Mathematical Statistics, 1988.

[5] R. Fagin, R. Kumar, and D. Sivakumar. Comparing top k lists. *J. Discrete Mathematics*, 17(1):134–160, 2003.

[6] J. E. Foster, M. McGillivray, and S. Seth. Composite indices: Rank robustness, statistical association, and redundancy. *Econometric Reviews*, 32(1):35–56, 2013.

[7] J Kasparian and A Rolland. Oecd's'better life index: can any country be well ranked? *Journal of Applied Statistics*, 39(10):2223–2230, 2012.

[8] M. Kendall and J. D. Gibbons. *Rank Correlation Methods*. Edward Arnold, London, 1990.

[9] E.L. Lehmann. *Nonparametrics: Statistical Methods Based on Ranks*. Holden-Day, San Francisco, 1975.

[10] S. Lin. Rank aggregation methods. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(5), 2010.

[11] J-L. Marichal. *Aggregation functions for decision making, Decision-Making Process - Concepts and Methods*. ISTE/John Wiley, 2009.

[12] I. Permanyer. Uncertainty and robustness in composite indices rankings. *Oxford Economic Papers*, 64:57–79, 2012.

[13] A. Rolland. Un indicateur pour mesurer la dépendance aux paramètres des palmarès. In *46 ème journées de la statistiques de la SFdS, Rennes*, 2014.

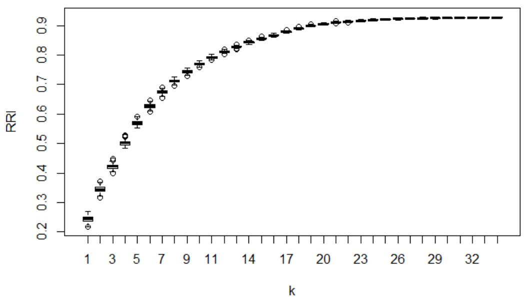
[14] A. Rolland. A note on top-k lists: average distance between two top-k lists. In *48 ème journées de la statistiques de la SFdS, Montpellier*, 2016.

[15] M. Saisana, A. Saltelli, and S. Tarantola. Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society A*, 168(2):307–323, 2005.

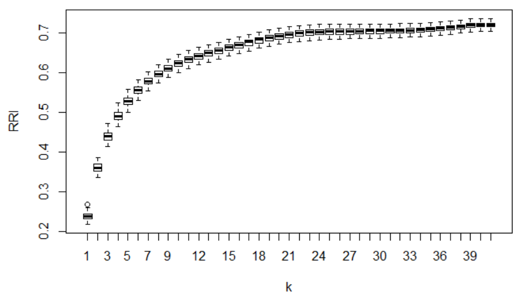
[16] A. Telcs, Z. T. Kosztyan, and A. Török. Unbiased one-dimensional university ranking - application-based preference ordering. *Journal of Applied Statistics*, to appear:1–17, 2015.

Appendix

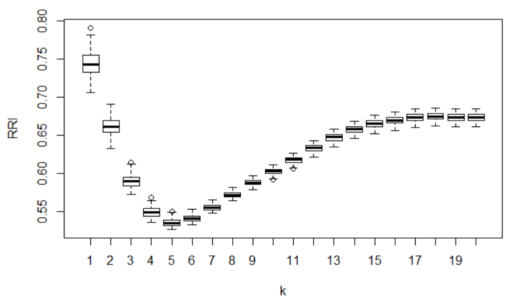
Figures below represent more in detail the information summarized in figure 2. For each of them, a boxplot of RSI_k is represented as a function of k. Boxplots show that the variability of the estimates is reasonably low.



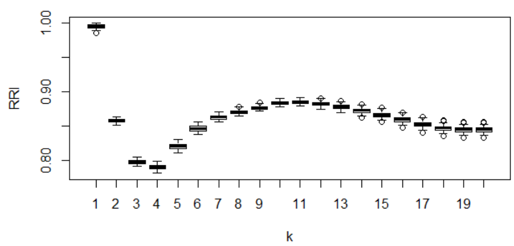
(a) OECD ranking



(b) "L'Etudiant" ranking



(c) QS ranking



(d) ARWU ranking

Figure 3. RSI_k for the different examples studied.