# Un indicateur pour mesurer la dépendance aux paramètres des palmarès

Antoine Rolland

*Laboratoire ERIC, Université de Lyon*
*antoine.rolland@univ-lyon2.fr*

**Résumé.** La production des palmarès nécessite des données statistiques sur les alternatives à comparer d'une part, et une fonction d'agrégation de ces données pour obtenir un score global d'autre part. Le choix de la fonction d'agrégation (typiquement une moyenne pondérée) et plus encore le choix des paramètres (typiquement les poids) peut influer grandement le classement obtenu. Nous proposons ici un indice permettant de quantifier les parts respectives dans le classement dues au choix des paramètres d'une part, et à l'information contenue dans les données d'autre part.

**Mots-clés.** Palmarès, indicateur, robustesse

**Abstract.** A ranking depends on the data available on the alternatives for one hand, and for a second hand on a aggregation function in order to obtain a global score for each alternatives. The choice of the aggregation function (e.g. a weighted sum) and teh choice of the parameters of the function (e.g. the weights) should have a great influence on the obtained ranking. We introduce in this communication a ration index which can quantify the respective part of the ranking due to the parameters and due to the data.

**Keywords.** Ranking, index, robustness

## 1 Le problème des palmarès

La presse magazine, tant généraliste que spécialisée, est une grande productrice de classements et palmarès en tout genre : hôpitaux [1], départements [2], villes [3]... tout est prétexte à la production de classements. De même les organismes internationaux sont friands de ces comparaisons entre pays [4,5]. La méthodologie est souvent la même : les évaluateurs identifient un certain nombre de critères d'évaluation ; ils recherchent ensuite les informations permettant de donner une valeur à tous les objets comparés (que nous appellerons de manière générique "alternatives") sur chaque critère; enfin ils agrègent ces informations en un score unique de manière à pouvoir comparer les alternatives suivant un préordre total.

Nous nous intéressons ici à la dernière phase de ce processus : l'agrégation des scores sur chacun des critères. Nous laissons de côté toutes les problématiques portant sur la construction de critères pertinents pour répondre à la question [6], ainsi que celles portant sur la qualité, la précision et la disponibilité des données. L'agrégation des valeurs sur les critères en une valeur unique se fait par une fonction d'agrégation, typiquement une moyenne pondérée (mais il en existe beaucoup d'autres [7]). Le jeu de poids utilisé (ou tout jeu de paramètre de manière plus générale) influence bien entendu le classement proposé. Cependant, une alternative qui serait meilleure qu'une autre sur tous les critères serait nécessairement meilleure après agrégation, et cela indépendamment du jeu de poids. Le classement final obtenu est donc dépendant des données intrinsèques d'une part, et des paramètres d'agrégation d'autre part. On peut alors se demander quelle est l'influence du choix politique (choix des paramètres) sur le classement. Autrement dit, le choix du paramétrage conditionne-t-il fortement ou faiblement le classement final?

De manière générique, on se place dans le cadre de l'agrégation multicritère : soit $\mathcal{X}$ un ensemble d'alternative, et $\mathcal{N}$ un ensemble de $n$ critères. Chaque alternative $x \in \mathcal{X}$ est décrite par ses valeurs prises sur les $n$ critères $x_1, \dots, x_n \in \mathbb{R}$. Soit $f$ une fonction d'agrégation de $\mathcal{X}$ dans $\mathbb{R}$. On suppose que la fonction $f$ est dépendante d'un jeu de paramètre $w$. Par exemple, si $f$ est la moyenne pondérée, $w = (w_1, \dots w_n)$ est l'ensemble des poids des critères et $f(x) = \sum_{i=1}^{n} w_i x_i$. On note $W$ l'ensemble de tous les jeux de paramètres possibles. On suppose que les valeurs sur les critères sont ordonnées suivant un ordre de préférence, et qu'il est le même pour tous les critères. Par convention on peut par exemple supposer que l'on souhaite minimiser chacun des critères, et partant, minimiser la fonction $f$. Cela nous permet d'obtenir un classement des alternatives par rapport aux valeurs prises par cette fonction $f$.

La question que l'on se pose est : étant donné une fonction paramétrée $f_w$, quelle est l'influence de $w$ sur le classement final? A notre connaissance, cette question n'a pas encore été abordée dans la littérature. Certains travaux portent bien sur la robustesse des ranking [8,9,10] mais aucun n'analyse la robustesse du classement vis-à-vis des variations de poids des critères. Nous proposons dans la partie suivante un indicateur permettant de répondre à cette question.

# 2 Un indice de mesure de la diversité potentielle d'un classement

## 2.1 Définition de l'indice

Nous proposons ici un indice de mesure de la dépendance du classement d'un jeu de données vis-à-vis des paramètres d'agrégation. L'indice de dépendance aux paramètres (IDP) utilise la notion de distance moyenne entre deux classements. Il existe plusieurs manières de mesurer la distance entre deux classements, les plus connues étant les distances de Kendall, de Spearman ou de Kemeny [11]. On note $d(R_1, R_2)$ la distance entre deux classements $R_1$ et $R_2$. On note $IDP_f(\mathcal{X})$ l'IDP de l'ensemble $\mathcal{X}$ par rapport à la fonction $f$. L'idée de l'$IDP$ est de calculer un ratio entre la distance moyenne entre

| rang | R1 | R2 | R3 | R4 | R5 | R6 |
|---|---|---|---|---|---|---|
| 1 | x | x | y | y | z | z |
| 2 | y | z | x | z | x | y |
| 3 | z | y | z | x | y | x |

Table 1: Exemple: les six classements possibles

deux classements différents obtenus pour deux jeux de paramètres différents de $W$, et la distance moyenne entre tous les classements possibles.

Dans le cas de paramètres discrets, l'IDP s'écrit:

$$IDP_f(\mathcal{X}) = \frac{\sum_{w \in W} \sum_{w' \in W} d(R, R')^2}{n^2 \bar{d}^2}$$

avec $d(R, R')$ la distance entre le classement obtenu avec le jeu de paramètres $w$ et le classement obtenu avec le jeu de paramètres $w'$, $n$ le cardinal de $W$, et $\bar{d}$ la distance moyenne entre deux classements de $\mathcal{X}$.

Dans le cas de paramètres continus, l'IDP s'écrit:

$$DPI_f(\mathcal{X}) = \int_{W \times W} \frac{d(R, R')^2}{\bar{d}^2} dw dw'$$

avec $d(R, R')$ la distance entre le classement obtenu avec le jeu de paramètres $w$ et le classement obtenu avec le jeu de paramètres $w'$, et $\bar{d}$ la distance moyenne entre deux classements de $\mathcal{X}$.

## 2.2 Exemples

Prenons un exemple à 3 alternatives $\mathcal{X} = \{x, y, z\}$ décrites sur 3 critères $c_1, c_2, c_3$ devant être minimisés. La fonction d'agrégation $f$ est une somme pondérée : $f(x) = \sum_i w_i x_i$. Il existe 6 classements différents présentés dans le tableau 1. Prenons comme distance $d$ la distance de Kendall; ici, $d(R1, R2) = 1$ et $d(R1, R6) = 3$, distance maximale possible. On vérifie aisément que la distance moyenne $\bar{d}$ entre deux classements est égale à 1,5.

Considérons maintenant 3 situations différentes. Dans la situation 1, présentée en tableau 2, $x$ domine $y$ et $z$ sur tous les critères, et $y$ domine $z$ également sur tous les critères. Quel que soit le jeu de poids, le résultat de la fonction d'agrégation sera toujours $f(x) = 1$, $f(y) = 2$ et $f(z) = 3$ ; le classement obtenu sera toujours $x$ en première position, puis $y$, puis $z$. Donc la distance entre deux classements obtenus avec deux jeux de poids différents sera toujours 0. Donc $IDP_f(\mathcal{X}) = 0/1,5^2 = 0$. Le jeu de poids n'a aucune influence sur le classement final.

Dans la situation 2, présentée en tableau 3, les trois alternatives différentes sont totalement symétriques par rapport aux critères. Alors chacun des six classements présentés

| | $c_1$ | $c_2$ | $c_3$ |
|---|---|---|---|
| x | 10 | 10 | 10 |
| y | 20 | 20 | 20 |
| z | 30 | 30 | 30 |

Table 2: Exemple: situation 1

| | $c_1$ | $c_2$ | $c_3$ |
|---|---|---|---|
| x | 10 | 30 | 20 |
| y | 20 | 10 | 30 |
| z | 30 | 20 | 10 |

Table 3: Exemple: situation 2

dans le tableau 1 sont également possibles, et donc $DPI_f(\mathcal{X}) = 1,5^2/1,5^2 = 1$ : le classement est entièrement dépendant du choix des paramètres.

La situation 3 est présentée au tableau 4. A titre de simplification, supposons que les poids sont des entiers entre 0 et 5, avec $\sum_{i=1}^{3} w_i = 5$. Après un calcul de la somme, nous trouvons que $IDP_f(\mathcal{X}) = 0.49/1,5^2 = 0.22$. Nous pouvons interpréter ce chiffre en disant que le classement obtenu dépend à 22% du choix politique des paramètres, et à 78% des valeurs des alternatives sur les critères.

Enfin, nous pouvons citer l'exemple du classement des villes étudiantes du magazine "L'Etudiant" [3]. Ce classement prend en compte 9 critères. A partir d'une estimation par méthode de Monte-Carlo, nous obtenons un indicateur $IDP = 0.095$. Cela montre que le classement donné par "L'Etudiant" dépend très peu du choix éditorial des paramètres de la fonction d'agrégation, et dépend beaucoup des données; autrement dit, il est raisonnable de penser que l'Etudiant est assez objectif dans le classement proposé.

## 2.3 Prolongements

D'un point de vue journalistique, il est fréquent que seuls les $k$ premiers éléments du classement soient médiatisés (typiquement 3 ou 10). Il est alors intéressant de mettre

| | $c_1$ | $c_2$ | $c_3$ |
|---|---|---|---|
| x | 10 | 10 | 20 |
| y | 20 | 30 | 30 |
| z | 30 | 20 | 30 |

Table 4: Exemple: situation 3

| | $c_1$ | $c_2$ | $c_3$ |
|---|---|---|---|
| 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 |
| a | 20 | 30 | 10 |
| a | 30 | 10 | 20 |
| a | 40 | 40 | 40 |
| a | 50 | 50 | 50 |

Table 5: Exemple: situations 4 et 5

en avant la robustesse de ces $k$ premiers éléments aux variations de paramètres. Il suffit alors pour chaque classement de $n$ alternatives de déclarer *ex-æquo* les $n - k$ derniers éléments pour que l'indicateur $IDP$ donne une valeur de robustesse plus en accord avec l'impression du lecteur : en effet, les variations du premier (ou du trio de tête) sont plus importantes, d'un point de vue médiatique, que des variations au cœur du classement.

Prenons par exemple la situation 4 décrite dans le tableau 5. Il est évident, si l'on cherche à minimiser les scores, que $a$ sera toujours classé premier, $b$ deuxième, et que $c$, $d$ et $e$ seront classés de la troisième à la cinquième place de manière équiprobable. Dans ce cas là, nous obtenons la valeur $IDP = 0.09$. Si nous réduisons la situation à l'analyse des deux premiers du classement, comme ce sont toujours les deux mêmes, nous obtenons le score $IDP = 0$, ce qui montre que les deux premiers ne dépendent pas des paramètres choisis.

Dans la situation 5 décrite dans le même tableau 5, et toujours en maximisant les scores, nous avons maintenant $d$ et $e$ toujours classés $4^{eme}$ et $5^{eme}$, et $a$, $b$ et $c$ classés aux trois premières places de manière équiprobable. Dans ce cas là, nous obtenons la valeur $IDP = 0.09$, comme dans la situation 4. Par contre si nous réduisons la situation à l'analyse des deux premiers du classement, nous obtenons le score $IDP = 0.37$, ce qui montre que les deux premiers dépendent de manières non négligeable des paramètres choisis.

# 3 Conclusion et perspectives

Nous avons posé ici la définition succincte d'un indicateur de dépendance vis-à-vis des paramètres pour une fonction d'agrégation pour un ensemble d'alternatives données. Nous pensons que ce travail mérite d'être poursuivi dans plusieurs directions et proposons ici des pistes de perspectives :

1. savoir calculer la distance moyenne entre deux classements équiprobable à l'aide d'une formule, y compris pour les cas de classement partiel tel que proposé en partie 2.3

2. définir un protocole de mise en œuvre pratique dans le cadre de jeux de poids continus, utilisant une simulation par méthode de Monte-Carlo.

3. étudier les propriétés mathématiques d'un tel indicateur, en essayant de faire le lien avec l'analyse de la variance.

4. tester sur de nombreux exemples le pouvoir descriptif d'un tel indicateur.

# Bibliographie

[1] F. Malye et J. Vincent (2013), *le Palmarès des hôpitaux*, le Point.
[2] O. Nouaillas et A. Culat (2013), *La France Verte, Palmarès écologique* (2013), La Vie, 3559, 14 novembre 2013.
[3] V. Bertereau, T. Brisson, M. Brochard et P. Falga, (2012) *Dossier : le palmarès 2012-2013 des villes où il fait bon étudier*, L'Étudiant, septembre 2012.
[4] J. Kasparian et A. Rolland (2012), *OECD's'Better Life Index: can any country be well ranked?*, Journal of Applied Statistics, 39 (10), p. 2223-2230.
[5] A. Rolland and J. Kasparian (2012), *Etude critique du palmarès de la qualité de vie dans les pays de l'OCDE*, 44 ème journées de la statistiques de la SFdS, Bruxelles.
[6] B. Roy (1985), Méthodologie multicritère d'aide à la décision, Economica.
[7] J-L. Marichal (2009), *Aggregation functions for decision making*, Decision-Making Process - Concepts and Methods, ISTE/John Wiley.
[8] M. Saisana, A. Saltelli, S. Tarantola (2005), *Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators*, Journal of the Royal Statistical Society, 168 (2), p. 307-323.
[9] I. Permanyer (2012), *Uncertainty and robustness in composite indices rankings* Oxf. Econ. Pap. 64: 57-79.
[10] J. E. Foster, M. McGillivray, and S. Seth (2013), *Composite Indices: Rank Robustness, Statistical Association, and Redundancy*, Econometric Reviews, 32 (1), p. 35-56,
[11] M. Kendall and J. D. Gibbons (1990) Rank Correlation Methods. Edward Arnold, London.