



HAL
open science

Explications axiomatisées pour décisions équitables

Hénoïk Willot, Sébastien Destercke, Khaled Belahcene

► **To cite this version:**

Hénoïk Willot, Sébastien Destercke, Khaled Belahcene. Explications axiomatisées pour décisions équitables. ROADEF - 25ème congrès annuel de la Société française de recherche opérationnelle et d'aide à la décision, Société française de recherche opérationnelle et d'aide à la décision, Mar 2024, Amiens, France. hal-04590407

HAL Id: hal-04590407

<https://hal.science/hal-04590407>

Submitted on 28 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Explications axiomatisées pour décisions équitables

Hénoïk Willot¹, Khaled Belahcene², Sébastien Destercke¹

¹ Heudiasyc, Université de Technologie de Compiègne, France

{henoik.willot, sebastien.destercke}@hds.utc.fr

² MICS, CentraleSupélec, Université Paris-Saclay, France

khaled.belahcene@centralesupelec.fr

Mots-clés : *Explications, Préférences, Ordered Weighted Average*

Introduction Afin de prendre des décisions informées et de qualité, la partie prenante des décisions, le décideur doit être aidé par un système d'aide à la décision qui est i) robuste et ii) explicable. La robustesse est représentée par un ensemble de modèles, dont les décisions sont prises de manière sceptiques : seules les décisions unanimes sur l'ensemble de modèle sont présentées au décideur. L'explication d'une décision est toujours un élément important pour le décideur afin d'améliorer sa compréhension et sa confiance envers le modèle et de lui permettre de répondre à un audit extérieur ou aux questions des utilisateurs impactés.

Nous proposons de définir un cadre de décision robuste permettant de prendre des décisions équitables tout en proposant des explications se reposant sur l'axiomatique que nous définissons.

Contexte Depuis plus d'un siècle, les progrès en économie ont mené à la définition d'outils de mesure et de décision sur le problème de répartition des richesses dans le monde. Un scénario de répartition de richesse correspond à une situation, notée a , représentant les revenus d'un groupe de n agents, c'est-à-dire chaque agent i est associé à son revenu, noté $a_i \in \mathbb{R}_+$. Divers outils ont été introduits tels que l'Indice de Gini, la dominance de Lorenz restreinte, les Redistributive Ordered Weighted Average (OWA redistributif) et la dominance de Lorenz généralisée. Ces outils permettent de définir des relations de préférences entre différents scénarios de répartition de richesses.

Propriétés des opérateurs L'OWA redistributif est l'outil le plus général et a été à plusieurs reprises caractérisé par des axiomatiques (ensemble de propriétés permettant de le définir uniquement) [6, 1, 5]. Le principe de *symétrie* (parfois appelée anonymat) stipule que toute situation a' obtenue en échangeant les indices des agents doit être identiquement préférée à la situation a - il devient donc équivalent de raisonner sur la distribution a ou sur la distribution \vec{a} , obtenue en ordonnant les agents par ordre croissant de leurs revenus ($\vec{a}_1 \leq \vec{a}_2 \leq \dots \leq \vec{a}_n$). Le principe de *transfert redistributif* stipule qu'une situation \vec{a} doit être au plus aussi préférée que la situation \vec{a}' si \vec{a}' est obtenue à partir de \vec{a} en prenant une quantité $\epsilon > 0$ à un agent j , que l'on considère comme "riche", pour la donner à un agent $i < j$, qui est donc moins "riche" que j , tout en conservant l'ordre social, ni l'un ni l'autre ne monte ou ne descende dans l'échelle sociale. L'axiome de *décidabilité* (complétude de la relation) permet de définir que toutes situations a et b doivent être comparables, c'est-à-dire que a est au plus aussi préférée que b , ou que b est au plus aussi préférée que a , ou alors que les deux sont autant préférées l'une que l'autre.

Ainsi la relation de préférence vérifiant symétrie, réflexivité, transitivité, transfert redistributif, monotonie (et deux autres axiomes assez spécifiques que nous ne résumerons pas ici) peut s'écrire comme une fonction de valeurs (l'OWA redistributive) exprimée comme : $Score(a) = \sum_{i=1}^n w_i * \vec{a}_i$ avec $w_1 \geq w_2 \geq \dots \geq w_n$ et $\sum_{i=1}^n w_i = 1$. Cependant, cet axiome de décidabilité est trop restrictif car il attribue un trop grand pouvoir de décision au jeu de poids

w choisi. Certaines de ces décisions reposent uniquement sur la valeur choisie, ce qui rends les rends peu explicable, d'autant plus que trouver cette valeur est une tâche complexe.

Nous souhaitons donc trois choses : i) remplacer l'axiome de décidabilité par un axiome moins restrictif, mais permettant tout de même d'incorporer des préférences du décideur, afin d'obtenir ce que l'on nommera OWA redistributif robuste ; ii) définir une inférence sceptique sur cet OWA redistributif robuste ; iii) trouver des explications de ces décisions.

Robustification Nous commençons par remplacer l'axiome de décidabilité par de l'information préférentielle récoltée auprès du décideur selon les principes de l'apprentissage de préférences robustes [3]. Nous définissons donc un OWA redistributif robuste : l'ensemble de modèles contenant ceux qui vérifient l'information préférentielle. L'inférence sceptique correspond donc aux décisions unanimes sur cet ensemble.

Ces décisions sceptiques sont donc celles qui sont prises par les propriétés de l'OWA redistributif (dont l'information préférentielle), et non par la valeur précise des paramètres. Nous cherchons donc à utiliser l'axiomatique pour proposer des explications. Nous commençons par les décisions sceptiques prises en ne considérant que les axiomes de réflexivité, transitivité, symétrie et transfert redistributifs. Nous montrons que cela correspond à la dominance de Lorenz restreinte. Il est connu depuis [4] que la dominance de Lorenz restreinte est explicable par une explication transitive reposant uniquement sur l'utilisation des transferts redistributifs. Une explication revient donc à transformer la situation b , jugée au mieux aussi bonne que la situation a , par une succession de transferts, que nous avons utilisés comme propriété désirable du modèle. Nous prouvons que le problème de décision associé est NP-complet, et proposons une nouvelle heuristique de calcul et la comparons à l'algorithme de [4] sur des données générées.

Nous montrons ensuite que lorsque l'on ajoute l'axiome de monotonie, l'inférence sceptique correspond alors à la dominance de Lorenz généralisée. Cette dernière a eu un regain d'intérêt récent en machine learning [2], et peut elle aussi être expliquée par des explications transitives, en utilisant les transferts redistributifs et la monotonie. Ce problème étant une généralisation du précédent, sa complexité est donc au moins aussi grande, il est donc NP-difficile.

OWA robuste L'OWA redistributif robuste, étant obtenu avec l'ajout de nouveaux axiomes, nous cherchons donc à généraliser les explications transitives que nous avons pu obtenir pour la dominance de Lorenz généralisée. Nous montrons que, sous certaines conditions, ajouter aux actions disponibles des transferts proportionnels à l'information préférentielle est suffisant pour produire une explication. Nous devons cependant changer la structure de nos explications impliquant une hausse de la complexité (ressentie par l'utilisateur) des explications pour être complet. Nous proposons la formulation théorique permettant de calculer ces explications ainsi que plusieurs heuristiques pour obtenir plus efficacement une partie des explications, avec des comparaisons sur des données synthétiques.

Références

- [1] E. Ben-Porath and I. Gilboa. Linear Measures, the Gini Index and the Income-Equality Tradeoff. *Journal of Economic Theory*, 64 :443–467, 1994.
- [2] V. Do et al. Two-sided fairness in rankings via Lorenz dominance. In *Advances in Neural Information Processing Systems*, volume 34, pages 8596–8608, 2021.
- [3] S. Greco et al. Rough sets theory for multicriteria decision analysis. *European Journal of Operational Research*, 129(1) :1–47, 2001. Number : 1.
- [4] G.H. Hardy et al. Inequalities. *Press, Cambridge*, 1934.
- [5] P. Perny et al. A decision-theoretic approach to robust optimization in multivalued graphs. *Ann. Oper. Res.*, 147(1) :317–341, 2006.
- [6] J. A. Weymark. Generalized gini inequality indices. *Mathematical Social Sciences*, 1(4) :409–430, 1981.