B. MONJARDET

Quelques problèmes relatifs à la méthode des comparaisons par paires

Mathématiques et sciences humaines, tome 37 (1972), p. 69-71
<http://www.numdam.org/item?id=MSH_1972__37__69_0>

Math. Sci. hum. (10e année, n° 37, 1972 p. 69-71)

QUELQUES PROBLÈMES RELATIFS A LA MÉTHODE DES COMPARAISONS PAR PAIRES

par

B. MONJARDET [1]

La méthode des comparaisons par paires est utilisée au niveau individuel comme au niveau collectif. Dans le premier cas elle consiste à faire exprimer les préférences d'un sujet sur un ensemble d'objets en lui demandant de choisir pour chaque paire d'objets celui qu'il préfère. On obtient ainsi une relation totale, antisymétrique (si on refuse les ex-aequo), mais non transitive en général ; le graphe ainsi obtenu s'appelle un tournoi. Au niveau collectif, on a un ensemble de tournois exprimant les préférences individuelles de différents sujets ; on définit alors une préférence collective en choisissant pour chaque paire d'objets celui préféré par une majorité de sujets ; on obtient encore un tournoi.

Un cas particulier important est celui où tous les tournois individuels sont transitifs, ce sont alors des ordres totaux et on retrouve le problème bien connu d'agrégation des votes ; on sait que même dans ce cas, la préférence collective n'est pas en général transitive : c'est l'« Effet Condorcet » (*cf.* Guilbaud G. Th., et Rosenstiehl P., « L'analyse algébrique d'un scrutin», *Math. Sci. hum.*, n° 4, ou dans *Ordres totaux finis*, Paris, Gauthier-Villars, 1971).

Les problèmes liés à la méthode de comparaisons par paires sont nombreux et ont été plusieurs fois abordés dans cette revue ; nous les classerons en deux thèmes principaux, évidemment reliés : description et ajustement.

Commençons par la description d'un tournoi ou d'un ensemble de tournois. Un tournoi sera décrit par ses propriétés, par exemple, de connexité ou de transitivité ; ainsi, on définira des indices de transitivité basés sur le nombre de circuits de longueur trois ou sur la distance à un ordre total ; la définition de ces indices requiert la connaissance de l'amplitude maximum de ces caractéristiques, ce qui, dans le second cas, pose des problèmes combinatoires difficiles (*cf.* l'article de J. C. Bermond dans ce numéro). La description d'un ensemble de tournois suit la même démarche qu'en statistique descriptive classique ; on cherche à résumer un ensemble d'éléments en donnant un élément central (moyenne, médiane ... en statistique) et un indice de dispersion par rapport à cet élément (variance, étendue...). Habituellement, l'élément central est calculé comme l'élément à « distance minimum » des éléments donnés, l'indice de dispersion étant alors une fonction de cette distance minimum ; par exemple, si les éléments sont de nature vectorielle, on cherche l'élément minimisant le moment d'inertie (somme des carrés des distances) des points donnés par rapport à un point quelconque : on obtient le centre de gravité et la variance n'est autre, à un coefficient près, que le moment d'inertie par rapport au centre de gravité. Dans tous les cas, l'élément central et l'indice de dispersion dépendent des distances choisies et de la définition de la « distance minimum ».

1. CNRS, Centre de Mathématique Sociale.

Dans le cas des tournois, on peut prendre comme distance celle de la différence symétrique entre deux relations ; elle est égale au double du nombre d'arcs à inverser pour passer du premier tournoi au second, c'est-à-dire au double de la « distance de Kendall ». Si on définit alors la distance minimum d'un tournoi à un ensemble de tournois comme la somme des distances de ce tournoi à chacun des tournois donnés, le tournoi à distance minimum est précisément celui obtenu en utilisant la méthode des comparaisons par paires ; c'est aussi le tournoi médian au sens algébrique du terme (*cf.* Barbut M., et Monjardet B., *Ordre et classification*, Paris, Hachette, 1970, chap. IV). On peut aussi plonger l'ensemble des tournois dans un vectoriel et utiliser une distance euclidienne. En particulier, si on considère seulement les n ! tournois transitifs comparant n objets, on peut les plonger dans un espace euclidien à $n - 1$ dimensions et en donner une représentation géométrique polyédrique : le permutoèdre (*cf.* chap. III d'*Ordres totaux finis*). A un ensemble d'ordres totaux correspond alors un nuage de points sur le permutoèdre ; l'élément à distance euclidienne minimum des ordres totaux correspond au centre de gravité de ce nuage ; le moment d'inertie du nuage par rapport à son centre de gravité, convenablement normé, redonne le coefficient de concordance appelé W par Kendall (*cf.* Degenne A., *L'ajustement des modèles ordonnés en analyse de questionnaires*, thèse 3e cycle Lettres, Paris, 1969). On trouvera dans cette même référence, l'interprétation des coefficients de corrélation classiques entre deux ordres totaux (τ de Kendall, φ de Spearman) ou de leurs moyennes sur les couples possibles, au moyen des deux distances définies ci-dessus.

Les résumés étudiés ci-dessus se limitent à un élément central et un indice de dispersion ; on peut aller plus loin en adaptant la démarche de l'analyse factorielle. L'analyse des préférences a été introduite par Benzécri dans le cas où ces préférences sont des ordres totaux (*cf.* chap. V d'*Ordres totaux finis*). On considère le nuage de points formé sur le permutoèdre par un ensemble d'ordres totaux ; on calcule l'inertie du nuage par rapport à un sous-espace, et le sous-espace minimisant cette inertie ; cette inertie minimum est décomposée suivant les axes principaux de ce sous-espace ; on obtient ainsi une expression pondérée des principales tendances de l'opinion (représentées par les axes principaux). D'autre part, on peut obtenir une représentation globale visuelle des proximités entre les préférences et éventuellement les préférences et les objets comparés ; ceci, en ne retenant qu'un sous-espace de faible dimension sur lequel on projette le nuage. A cette approche on peut relier l'article de Jacquet-Lagrèze (« Analyse d'opinions valuées et graphes de préférence », *Math. Sci. hum.*, n° 33) conduisant à une représentation plane simultanée des préférences et des objets comparés ; les axes sont ici déterminés à l'aide d'algorithmes d'agrégation ; à signaler aussi que ce même article définit des indices de cohésion pour des préférences valuées.

Un des résultats de l'analyse descriptive d'un ensemble de préférences est de définir des indices de cohérence ou d'homogénéité de ces préférences. Si ces indices sont suffisamment élevés on peut chercher à ajuster à ces préférences un modèle appartenant à une classe déterminée ; la méthode usuelle est de définir une distance et de chercher le modèle de la classe à distance minimum des préférences données. Plusieurs classes de modèles sont envisageables : préordres, ordres partiels, fuseaux d'ordres, ordres totaux... ; nous ne considérons ici que la classe la plus couramment utilisée, celle des ordres totaux. Ainsi, dans le cas d'une préférence traduite par un tournoi, on cherchera les ordres totaux à distance minimum de ce tournoi ; il en est de même pour un ensemble de tournois lorsque ceux-ci ont été résumés par leur tournoi médian.

Là encore, on peut prendre la distance de la différence symétrique ou une distance euclidienne. Dans le premier cas la recherche des ordres optimaux conduit à un problème combinatoire ; M. Barbut avait déjà remarqué (« Note sur les ordres totaux à distance minimum d'une relation binaire donnée », *Math. Sci. hum.*, n° 20) que ce problème est équivalent à un problème de recouvrement minimum donc de « transversale» minimum. D'autre part, des algorithmes pour obtenir les ordres optimaux ont été présentés par B. Durand (« A propos du nombre minimum d'arcs pour supprimer les circuits d'un graphe », *Math. Sci. hum.*, n° 20) et C. Heuchenne pour un cas plus général (« Un algorithme général pour trouver un sous-ensemble d'un certain type à distance minimum d'une partie donnée », *Math. Sci. hum.*, nº 30). Ces problèmes sont étudiés dans l'article de J. C. Bermond, dans ce numéro ; il présente la synthèse des résultats obtenus en utilisant souvent des démonstrations originales et unifiantes et il apporte des résultats nouveaux.

Pour obtenir les ordres totaux à distance minimum d'un ensemble de tournois on peut également utiliser la distance euclidienne ; dans le cas où ces tournois sont transitifs cela revient à chercher les ordres les plus proches du centre de gravité du nuage qu'ils forment sur le permutoèdre ; on peut montrer que ceci est équivalent à choisir les ordres obtenus par la « méthode de Borda », c'est-à-dire en attribuant à chaque objet la somme de ses rangs (*cf.* Degenne, *op. cit.*).

Il existe d'autres méthodes pour ajuster un ordre total à un tournoi ou un ensemble de tournois ; par exemple, dans le premier cas, celles basées sur les scores ou des procédés d'itération des scores. Il reste à situer ce genre de méthodes par rapport à celles fondées sur la minimisation d'une distance et aussi à comparer l'ensemble de ces méthodes. Par exemple, à quelles conditions, distance de la différence symétrique et distance euclidienne donnent-elles les mêmes résultats ?

Signalons aussi quelques problèmes connexes à ces questions : la détermination des ensembles de préférences dont le résumé à distance minimum pour la distance de la différence symétrique est un ordre ; en particulier, la caractérisation des ensembles d'ordres totaux stables pour l'opération médiane, tels que par exemple, les ordres Blackiens (*cf.* Kreweras G., « Les décisions collectives », *Math. Sci. hum.*, nº 2 et Frey L., « Parties distributives du treillis des permutations », dans *Ordres totaux finis*) ; l'inversion de la relation de Condorcet, c'est-à-dire la détermination du nombre minimum d'ordres totaux dont la médiane est un tournoi donné.

Aucun modèle probabiliste n'était introduit dans les méthodes d'ajustement précédentes ; si on introduit un tel modèle on rencontre des problèmes d'estimation et de décision statistique, étudiés dans l'article de A. Astié (« Comparaisons par paires et problèmes de classement, estimation et tests statistiques », *Math. Sci. hum.*, nº 32 et aussi dans celui de G. Th. Guilbaud (« Préférences et ochastiques », *ibid.*).

Cette rapide revue des problèmes soulevés par la méthode des comparaisons par paires aura montré au lecteur qu'ils sont nombreux et que malgré les contributions apportées, bien des questions restent ouvertes. Nous souhaitons donc pouvoir accueillir prochainement d'autres articles sur ce thème dans *Mathématiques et Sciences humaines*.

N. B. — Nous avons surtout cité dans cette revue, les articles parus dans *Math. Sci. hum.* ; le lecteur trouvera des références plus complètes dans la bibliographie de ces articles ou des autres ouvrages cités.