MATHÉMATIQUES ET SCIENCES HUMAINES

MARC BARBUT

Médianes, Condorcet et Kendall

Mathématiques et sciences humaines, tome 69 (1980), p. 5-13
<http://www.numdam.org/item?id=MSH_1980__69__5_0>

Math. Sci. hum. (*18*^e^ année, n°69, *1980*, p.5-13)

MEDIANES, CONDORCET ET KENDALL

Marc BARBUT*

Au début des années 60, je rédigeai deux notes internes ; l'une, très brève, intitulée "Médianes, Condorcet et Kendall" ; l'autre, plus volumineuse, intitulée "Médiane, distributivité, éloignements".

Ces deux notes furent fréquemment utilisées et citées depuis, de sorte que leur publication sembla souhaitable ; la première paraît donc aujourd'hui ; la seconde, allégée par rapport à sa version primitive, sera incluse dans un prochain numéro de "Mathématiques et Sciences humaines".

La présente note a son origine dans les problèmes d'*agrégation de préférences* (exprimées dans des votes, ou résultant de choix à critères multiples). Il s'agissait de mettre en évidence quelques propriétés remarquables de la procédure dite de Condorcet pour cette agrégation. C'est pourquoi on y traite de préordres (partiels ou totaux) : ceux -ci sont sensés représenter les préférences.

En fait, il s'agit de propriétés du treillis des parties d'un ensemble fini, muni de la métrique de la différence symétrique (entre parties) ; le fait qu'on les applique à des préordres (i.e., des relations binaires transitives) n'a aucune conséquence sur les résultats ; de sorte que l'on peut partout, dans ce qui suit, remplacer le mot "préordre" par le mot "relation binaire", ou le mot "partie d'un ensemble donné". Et c'est d'ailleurs à la généralisation de ces propriétés (du treillis des parties d'un ensemble) à la classe des treillis distributifs qu'est consacrée la seconde note.

---

* Centre de Mathématique Sociale, Ecole des Hautes Etudes en Sciences Sociales, Paris.

Celle qui suit comprend cinq parties.

1. Distance-K entre préordres complets - Distance de deux graphes.
2. Relation (dite de Condorcet) à distance-K minimum de n préordres complets distincts.
3. Cas de préordres pondérés.
4. Préordres pondérés et relation pondérée la plus proche.
5. Unanimités équivalentes à n préordres pondérés.

1. DISTANCE-K ENTRE PREORDRES COMPLETS - DISTANCE DE DEUX GRAPHES.

Si X et Y sont deux parties d'un ensemble E fini, on pose :

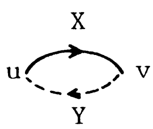
$d(X,Y) = |X \Delta Y|$ = Cardinal de la différence symétrique.

Si $E = A \times A$, X et Y sont deux relations (graphes) sur A, $d(X,Y)$ est le nombre des couples $(u,v) \in E$ tels que :

(u X v et non-u Y v) ou (non-u X v et u Y v).

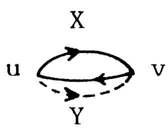
Si X et Y sont deux préordres complets sur A, $d(X,Y)$ est le nombre de désaccords (discordances) entre ces deux préordres : c'est *leur distance-K* (de M.G. Kendall).

Chaque paire {u,v} telle que :

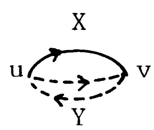


compte pour 2 dans le calcul de d.

Chaque paire {u,v} telle que :



ou



compte pour 1.

2. RELATION (GRAPHE A DISTANCE-K MINIMUM DE n PREORDRES COMPLETS DISTINCTS : C'EST LA RELATION DE CONDORCET DE CES PREORDRES.

a) Soient n préordres sur un ensemble A fini donnés par leurs graphes $X_1, X_2, \dots, X_n$ dans $E = A \times A$.

J'appelle relation de Condorcet de ces préordres la relation $C\,(X_1, X_2, \dots, X_n)$ sur A définie par :

$$(u,v) \in C \iff |\{i : u\, X_i v\}| > \left[\frac{n}{2}\right]$$

i.e. : u est avant v pour une majorité de ces n préordres.

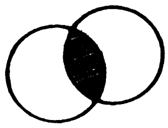
On a évidemment si $n = 2p+1$ ou $n = 2p$

$$C = \bigcup_{|P| = p+1} \left(\bigcap_{i \in P} X_i \right) \qquad P \subset \{1,2,\dots,n\}.$$

C'est, lorsque n est impair, l'*ensemble médian* de $X_1, X_2, \ldots, X_n$; lorsque n est pair, c'est la borne inférieure de l'*intervalle médian* de $X_1, X_2, \ldots, X_n$.

Exemples :

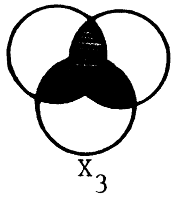
n = 2



$C = X_1 \cap X_2$

(intervalle médian : $(x_1 \cap X_2,\ X_1 \cup X_2)$)

n = 3



$$C = (X_1 \cap X_2) \cup (X_2 \cap X_3) \cup (X_3 \cap X_1)$$
$$= (X_1 \cup X_2) \cap (X_2 \cup X_3) \cap (X_3 \cup X_1)$$

n = 4

$$C = (X_1 \cap X_2 \cap X_3) \cup (X_1 \cap X_2 \cap X_4) \cup$$
$$\cup (X_1 \cap X_3 \cap X_4) \cup (X_2 \cap X_3 \cap X_4)$$

b) Posons d'une façon générale :

$$Y_1 = \bigcup_{i=1}^{n} X_i \qquad Y_2 = \bigcup_{i \neq j} (X_i \cap X_j)$$

$$Y_k = \bigcup_{|K|=k} \left(\bigcap_{i \in K} X_i \right) \qquad 1 \leqslant k \leqslant n$$

$$Y_n = \bigcap_{i=1}^{n} X_i$$

On a : $Y_n \subset Y_{n-1} \subset Y_{n-2} \subset \ldots \subset Y_2 \subset Y_1$.

Y_k est la relation : au moins k des préordres X_i mettent u avant v.
Selon le contexte, ce peut être la"majorité des $k/_n$", ou le premier quantile $k/_n$, par exemple.

c) Soit T une partie quelconque de E. On a :

$$\sum_{i=1}^{n} d\ (T, X_i) = \sum_{i=1}^{n} d\ (T, Y_i)$$

Montrons le par récurrence.

La propriété est évidente pour n = 2 :

$$d(T,X_1) + d(T,X_2) = d(T,\ X_1 \cap X_2) + d(T, X_1 \cup X_2)$$

Pour tout k $(1 < k < n)$, on a :

$$Y_k = (Z_{k-1} \cap X_n) \cup Z_k$$

où Z_k est défini par rapport à $X_1,\ldots,X_{n-1}$ comme Y_k l'est par rapport à $X_1,\ldots,X_{n-1},X_n$.

En particulier :

$$Y_1 = X_n \cup Z_1 \qquad\qquad Y_n = X_n \cap Z_{n-1}$$

D'où, puisque $Z_k \subset Z_{k-1}$:

$$d(T,Y_k) = d(T,Z_{k-1} \cap X_n) + d(T,Z_k) - d(T,X_n \cap Z_k).$$

Soit par sommation :

$$\sum_{k=1}^{n} d(T,Y_k) = \sum_{k=1}^{n-1} d(T,Z_k) + d(T,X_n)$$

$$= \sum_{k=1}^{n-1} d(T,X_i) + d(T,X_n)$$

d) D'où résulte immédiatement :

La somme $\sum_{i=1}^{n} d(T,X_i)$ est rendue minimum lorsque T est égale à la relation de Condorcet C associée aux X . (Parmi les T rendant cette somme minimum, il y a C.)

En effet, il faut et il suffit de rendre minimum $\sum_{i=1}^{n} d(T,Y_i)$ et les Y_i forment une chaîne d'inclusions. $\sum d(T,Y_i)$ est nécessairement diminué en rendant vides les deux parties hachurées, sur la figure (1) ci-contre, c'est-à-dire en choisissant T telle que :

$$Y_n \subset T \subset Y_1.$$

Puis on recommence sur Y_2 et Y_{n-1}, etc.

On obtient bien la médiance (ou l'intervalle médian) des Y_i, i.e., des X_i.

N.B. On notera l'analogie de la dernière partie de ce raisonnement avec celui qui montre que la médiane (ou l'intervalle médian) d'une suite de points $x_1,x_2,\ldots,x_n$ sur une droite rend minimum la somme

$$\sum_{i=1}^{n} |x_i - t| .$$

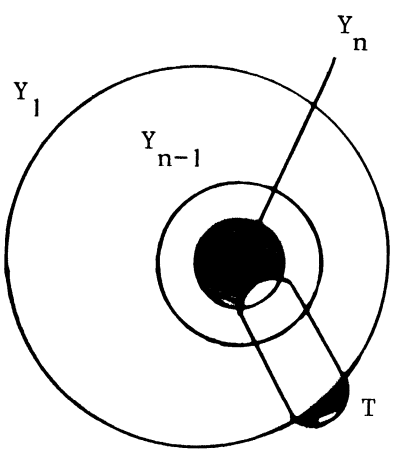


Figure 1

3. CAS DE PREORDRES ADMETTANT DES ORDRES DE MULTIPLICITE

Soient encore $X_1, X_2, \ldots, X_n$, n préordres complets ayant pour ordres de multiplicité ou "poids" $p_1, p_2, \ldots, p_n$ respectivement.

Selon les contextes, il pourra s'agir du nombre de voix (parmi $p = p_1+p_2+\ldots+p_n$ votants) qui se seront portées sur chaque préordre ; ou d'une valuation positive attribuée à chacun des préordres ; ou d'une distribution de fréquences sur ces préordres ; etc.

Ce qui suit ne suppose que la positivité des p_i.

J'appelle relation de Condorcet C de ces préordres "pondérés" la relation définie par :

$$(u,v) \in C \iff \left(\sum_{i:uX_iv} p_i \right) > \frac{1}{2} \left(\sum_{i=1}^{n} p_i \right) = \frac{1}{2} p$$

i.e., u est avant v pour une majorité simple de préordres, chacun étant compté avec son ordre de multiplicité.

On a alors le même résultat que supra en 2, à savoir :

$\sum_{i=1}^{n} p_i \ d(T,X_i)$ atteint son minimum lorsque $T = C$.

Nous allons le montrer par un raisonnement différent de celui du paragraphe précédent, qui utilisait les chaines d'inclusions. Ce raisonnement est plus général car il est applicable au cas particulier où :

$$\forall i \ , \quad p_i = 1$$

qui est celui du paragraphe 2 supra.

$X_1, X_2, \ldots, X_n$ déterminent dans les parties de E un simplexe -n dont les éléments sont les 2^n classes X_I :

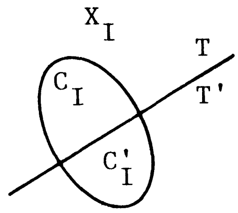
$$X_I = \bigcap_{i \in I} X_i \quad \bigcap_{j \in I'} X'_j \qquad I \subset \{1,2,\ldots,n\} = N$$

où l'on a posé : $\forall j \ , \ X'_j = E - X_j$ et $\forall I, \ I' = N - I$.

Chaque classe X_I est divisée par T en deux parties complémentaires dans X_I : $C_I = X_I \cap T$, $C'_I = X_I \cap T'$. $\qquad (T' = E - T)$

Dans le calcul de : $f(T) = \sum p_i \ d(T,X_i)$, chaque couple (u,v) appartenant à C_I intervient avec le poids :

$$p_{I'} = \sum_{j \in I'} p_j \quad \text{(où } I' = N-I)$$



et chaque élément de C'_I avec le poids $p_I = p - p_{I'}$.

Ces poids se retrouvent d'ailleurs dans les parties $C_{I'}$, pour ce qui est de p_I et C'_I pour ce qui est de $p_{I'}$, avec

$$X_{I'} = \bigcap_{i \in I'} X_i \bigcap_{j \in I} X'_j .$$

On minimise donc f(T) en choisissant T de façon que chaque classe X se réduise à la partie C_I, si $p_I > p_{I'}$, et à la partie C'_I dans le cas contraire. En cas d'égalité, la coupure est arbitraire.

Remarque :

La propriété énoncée pour des parties d'ensembles pondérées :leur relation de Condorcet minimise la somme des distances pondérées, ne fait que généraliser la propriété connue de la médiane d'une distribution de nombres $x_1, x_2, \ldots, x_n$ pondérés par des poids $p_1, p_2, \ldots, p_n$.

Dans ce calcul de la médiane, chaque nombre x_i doit être compté comme p_i ex-aequo avec la note x_i ; la médiane ainsi calculée minimise $\sum p_i \; |x_i - t|$.

4. PREORDRES PONDERES ET RELATION PONDEREE LA PLUS PROCHE

La relation de Condorcet définie supra est, pour un ensemble de n préordres pondérés par $p_1, p_2, \ldots, p_n$, la relation qui minimise la somme des distances aux préordres (pondérés) correspondant à ces préordres.

On peut vouloir rechercher une relation T, *elle-même pondérée par un coefficient t*, qui soit à distance minimum des n préordres. On prendra alors naturellement comme définition de la distance entre T et un préordre X_i (avec les mêmes notations que précédemment) :

$$d(T,X_i) = |X_i \cap T|\,|p_i - t| + |X_i \cap T'|\; p_i + |T \cap X'_i|\; t .$$

Et on minimise $\quad F(T) = \sum_i d(T,X_i)$.

Intuitivement, t peut être interprété comme un poids moyen (un nombre de voix moyen) attribué à la relation T qui est supposée résumer l'ensemble des n préordres pondérés ; t devra donc être pris du même ordre de grandeur que ces poids.

Traitons deux cas particuliers intéressants :

1) $t = 1$

On a : $d(T,X_i) = p_i|X_i - T| + |p_i - 1|\ |X_i \cap T| + |T - X_i|$.

Dans chaque classe X_I, définie comme supra au paragraphe 3, les couples (u,v) appartenant à C_I interviennent dans le calcul de F(T) avec le poids :

$p_I - |I| + |I'|$. (où $|I|$ et $|I'|$ désignent respectivement le nombre d'éléments de I et de I'.)

Ceux de C'_I avec le poids p_I.

F(T) est donc minimisée en réduisant C_I à $\emptyset$ si $|I'| > |I|$ et à I si $|I| > |I'|$; en cas d'égalité, la coupure est arbitraire.

La relation T ainsi obtenue est donc la relation de Condorcet de $X_1, X_2, \ldots, X_n$ non pondérés : la médiane de $X_1, X_2, \ldots, X_n$.

2) $t > \max\limits_i p_i$

Les éléments de $X_I \cap T$ comptent alors pour le poids $nt - p_I$, ceux de $X_I \cap T'$ pour le poids p_I.

Si $p_I > \frac{nt}{2}$, on prend $T \supset X_I$.

Si $p_I < \frac{nt}{2}$, on prend $T' \supset X_I$.

Autrement dit, T est la relation obtenue en gardant tous les couples (u,v) tels que u est préféré à v par une majorité $\frac{nt}{2}$ au moins des votants ; comme $\max\limits_i p_i > \frac{p}{n}$, cette majorité est supérieure à la majorité simple $\frac{p}{2}$.

5. UNANIMITES EQUIVALENTES A n PREORDRES PONDERES

On considère maintenant la relation X valuée suivante déduite des n préordres $X_1, X_2, \ldots, X_n$ pondérés par $p_1, p_2, \ldots, p_n$: X est valuée, chaque arc (u,v) étant valué par la somme des p_i tels que $(u,v) \in X_i$.

Autrement dit, la valuation est constante dans chacune des 2^n classes X_I définies en 3, et vaut p_I dans cette classe X_I.

Soit T une relation valuée par t (tous les arcs sont valués par t ou 0) : on peut l'interpréter comme une relation d'*unanimité* pour t votants (ou t critères), chacun compté à son ordre de multiplicité .

Posons : $d(X,T) = \sum_{u,v} |p_{uv} - t_{uv}|$

où : $p_{uv} = p_I$ si $(u,v) \in X_I$

et : $t_{uv} = 0$ ou t selon que $(u,v) \in T'$ ou $(u,v) \in T$,

et minimisons $d(X,T)$.

On a, pour chaque classe X_I, des termes qui contribuent à $d(X,T)$:

$$|p_I - t|\ |X_I \cap T| + p_I\ |X_I \cap T'|.$$

Si $|p_I - t| > p_I$, on minimise la somme en prenant $T' \supset X_I$, si $|p_I - t| < p_I$, on prend $T \supset X_I$.

Ce qui revient à prendre pour T l'union des classes X_I telles que $p_I > \frac{t}{2}$.

Si $t = 1$: $t = \bigcup_{i=1}^{n} X_i$.

Si $t = p$: $T = C\,(X_1, X_2, \ldots, X_n\ ;\ p_1, \ldots, p_n)$ = relation de Condorcet de $X_1, \ldots, X_n$ pondérés.

Si $2p>t>p$: T comprend tous les arcs qui ont une majorité supérieure à $\frac{t}{2}$ des suffrages.

Par exemple, si $t \doteq 1{,}8 \times p$, c'est la majorité 90%.

Pour $t = 2p$, $T = \bigcap_{i=1}^{n} X_i$ (unanimité).

Si $t > 2p$, $T = \emptyset$.

Une interprétation de T est la suivante : la relation X analyse le résultat des comparaisons par paires selon les n préordres (fournis par exemple par n votants, ou n critères de choix) pondérés. T est une relation réalisant, sur tous les couples lui appartenant, l'unanimité de t votants (critères), et qui soit la plus proche de X. Il est donc naturel de prendre ici t de l'ordre de grandeur de p.

La cas particulier $t = p$ s'interprète alors ainsi : j'ai p votants répartis sur n préordres $X_1, X_2, \ldots, X_n$; si ces votants avaient été unanimes, leur vote aurait donné la relation de Condorcet.

Les cas $t > p$ permettent d'avoir des seuils de majorité supérieurs à 1/2 ; un tel seuil correspond à une unanimité de plus de votants qu'il n'y en a effectivement.

b) Si l'on pose :

$$d(X,T) = \max_{(u,v)} |p_{uv}-t|$$

(métrique de la convergence uniforme), les résultats sont analogues ; c'est encore à t/2 que se fait la coupure qui définit la relation T.

BIBLIOGRAPHIE

[1] CONDORCET A. de, *Essai d'application de l'analyse à la probabilité des décisions prises à la pluralité des voix*, 1786.

[2] GUILBAUD G.Th., "Les théories de l'intérêt général", *Economie Appliquée*, n°4, 1952 (épuisé). Réimpression *in Eléments de la Théorie Mathématique des Jeux*, Paris, Dunod, 1968.

[3] ARROW K.J., *Social Choice and Individual Values*, New York, Londres , Sidney, Wiley, 1963, *Choix collectif et préférences individuelles*, trad. de l'anglais, Paris, Calmann-Lévy, 1974.

[4] KENDALL M.G., *Rank Correlation Methods*, Londres, Griffin, 1955.

[5] KREWERAS G., "Les décisions collectives", *Math. Sci. hum.*, 2, 1962.

[6] GUILBAUD G.Th. et ROSENSTIEHL P., "Analyse algébrique d'un scrutin", *Math. Sci. hum.*, 4, 1962.

[7] LUCE R.D. and RAIFFA H., *Games and Decisions*, New York, Wiley, Londres, Chapman, 1958.