Revue de statistique appliquée

B. Monjardet

Concordance et consensus d'ordres totaux : les coefficients *K* et *W*

Revue de statistique appliquée, tome 33, nº 2 (1985), p. 55-87

<http://www.numdam.org/item?id=RSA_1985__33_2_55_0>

CONCORDANCE ET CONSENSUS D'ORDRES TOTAUX : LES COEFFICIENTS K ET W

B. MONJARDET

Université Paris V et Centre de Mathématique Sociale
54 Bd. Raspail, 75270 Paris Cedex 06

RESUME

Une des "données ordinales" les plus classiques est la relation d'ordre total, rencontrée par exemple, dans l'étude de préférences. Les problèmes soulevés par la comparaison de deux ou plusieurs ordres totaux, ou par leur "agrégation" en un ordre total "consensus", ont fait l'objet de maints travaux, mais sans qu'apparaisse généralement de lien entre l'aspect concordance et l'aspect consensus, ou d'idée commune sous-jacente aux diverses solutions proposées. Le but de ce texte est de présenter un traitement "géométrique" unifié de ces deux aspects, dans un cadre limité mais privilégié. Les ordres totaux sont codés dans des espaces euclidiens et nous nous limitons aux deux codages fondamentaux auxquels on peut associer les noms de KENDALL et de SPEARMAN (et les coefficients tau et rho). En définissant de manière générale l'homogénéité d'un nuage de points comme son inertie normée, les coefficients K et W de concordance entre plusieurs ordres totaux apparaissent alors comme des cas particuliers, correspondants aux deux codages de base. Ils admettent des variantes résultant de normalisations différentes. Au point de vue consensus, on montre comment la procédure d'agrégation associée canoniquement à un coefficient de concordance, redonne pour K et W deux procédures classiques. On étudie aussi la relation entre K et W qui découle d'une relation entre distances de KENDALL et de SPEARMAN (ou entre tau et rho). Au cours de l'exposé, on rectifie quelques erreurs rencontrées dans la littérature du sujet.

Mots-clefs : **Coefficient de concordance, Consensus, Inertie, Préférence, Ordre total.**

1. INTRODUCTION

Le cas le plus classique de "données ordinales" est celui où ces données sont modélisées par des préordres totaux, définis sur l'ensemble des objets considérés. Ce type de données se rencontre fréquemment, notamment en analyse des préférences ; le préordre total peut être obtenu "directement", par exemple, si l'on demande de ranger des candidats ; il peut être aussi induit à partir de notes ou de mesures d'utilité. Plus généralement, n'importe quelle variable numérique induit un préordre total, et cette réduction des données à leur seul aspect ordinal, qu'on peut faire remonter au moins à SPEARMAN (1904), est utile dans certaines situations (cf., par exemple, SIBSON, 1972 et HOLLANDER et WOLFE, 1973). Nous ne traiterons en fait ici, pour des raisons sur lesquelles nous reviendrons à la fin de cette introduction, que le cas "sans ex-aequo", où les données sont représentées par des ordres totaux.

Nous cherchons donc à évaluer la concordance entre deux ou plusieurs ordres totaux. Les deux coefficients τ de KENDALL et ρ de SPEARMAN, qui mesurent l'accord entre deux ordres totaux, sont bien connus. Dans le cas de plus de deux ordres totaux, quoique les coefficients W de KENDALL et U de KENDALL-EHRENBERG soient classiques, et qu'au moins le premier ait été souvent utilisé, ils semblent nettement moins bien connus. Ceci se traduit notamment par le fait que les manuels où ils sont présentés comportent souvent des imprécisions ou même des erreurs. Certains résultats, (par exemple, les bornes de U) peuvent aussi sembler étranges. Enfin, alors que le τ et le ρ sont souvent définis comme des coefficients de corrélation, objets familiers au statisticien, les définitions usuelles de W et de U (cf., par exemple, KENDALL 1962) ne font pas apparaître d'interprétations géométriques, qui cependant existent (DEGENNE, 1972).

Nous nous proposons dans ce texte d'essayer de remédier à cet état de fait, d'une part en donnant une présentation géométrique unifiée des coefficients de concordance évoqués ci-dessus, d'autre part en corrigeant les erreurs que nous avons relevées (ce qui ne veut pas dire que nous résoudrons tous les problèmes soulevés). Il apparaîtra alors qu'il y a deux coefficients fondamentaux, l'un que nous appelons K (pour KENDALL) et le W. Ces deux coefficients résultent de l'application de la même formule d'inertie normée d'un nuage de points (le coefficient d'homogénéité H_1 du paragraphe 2) à deux choix fondamentaux de codage euclidien de l'ensemble $\mathcal{O}$ des n ! ordres totaux définis sur l'ensemble à n éléments (paragraphe 3). Le premier de ces codages, qu'on peut appeler codage de KENDALL plonge $\mathcal{O}$ dans $\mathbf{R}^{n^2}$; la distance euclidienne entre les images de deux ordres totaux correspond dans ce cas à la "distance de KENDALL", i.e. au nombre de "désaccords" entre eux. Le second codage, qu'on peut appeler codage de SPEARMAN, plonge $\mathcal{O}$ dans $\mathbf{R}^n$, la distance euclidienne associée correspondant à celle dite des rangs. Dans les deux cas, ces codages sont "sphériques" en ce sens que dans l'espace affine euclidien de codage, les points représentatifs sont sur une (hyper) sphère. Il est de plus commode de prendre des codages "centrés", (les points sont sur l'hyperplan : somme des coordonnées nulle), ce qui dans le premier cas conduit à coder par un "vecteur caractéristique centré", dans le second par un "vecteur score" (de préférence au "vecteur rang"). Ces codages une fois définis, il suffit ensuite aux paragraphes 4 et 5, d'utiliser les formules générales de 2.4.1. pour obtenir différentes expressions des coefficients K et W, notamment comme moyenne des coefficients binaires τ et ρ. Les variantes étudiées dans ces mêmes paragraphes résultent soit d'un moyennage différent (c'est le cas du U qui est une variante de K), soit d'un choix différent de la normalisation de l'inertie d'un nuage de points qui conduit au coefficient d'homogénéité H_2 du paragraphe 2. Les coefficients de concordance résultant de ce dernier choix sont presque toujours égaux ou approximativement égaux aux coefficients K et W, ce qui explique qu'on ait pu abusivement les confondre avec eux.

Au paragraphe 6, on se pose le problème de la relation entre les coefficients K et W. Celle-ci résulte d'une relation fondamentale entre distance de KENDALL et distance de SPEARMAN, relation équivalente à l'inégalité de DANIELS (1948) ou à la formule de GUILBAUD (1980) pour les coefficients τ et ρ.

Une fois donné un coefficient de concordance entre ordres totaux, on peut lui associer une méthode d'agrégation de v ordres totaux ; celle consistant à chercher l'ordre total maximisant le coefficient de concordance des v + 1 ordres totaux. Nous montrerons au paragraphe 7, comment une telle utilisation des coefficients K et W permet de réobtenir deux procédures classiques d'agrégation : celle de "l'ordre médian" et celle dérivée de la "somme des rangs".

La manière dont nous introduisons dans ce texte les coefficients K et W. à partir de formules généralement classiques sur l'inertie d'un nuage de points, conduit à des démonstrations très simples de la plupart de leurs propriétés et nous ne les avons donc pas détaillées . Les démonstrations de quelques résultats moins évidents sont fournis dans le texte, ainsi que certains problèmes ouverts non résolus à notre connaissance.

Les références aux travaux antérieurs sont relativement réduites dans la version publiée ici. Toutefois, le lecteur pourra se reporter à une première version (MONJARDET, 1984), où il trouvera une note historique et des compléments variés. Nous nous contenterons ici d'attirer l'attention sur des articles de BENZECRI (1965), FISHBURN (1973), HAYS (1960), JACQUET-LAGREZE (1977), KRUSKAL (1958) et sur le Symposium on rankings methods (MORAN et autres, 1950). Insistons aussi sur le fait qu'on peut définir bien d'autres distances entre ordres totaux que celles de KENDALL ou de SPEARMAN, notamment celles associées à d'autres codages euclidiens, auxquelles correspondront d'autres coefficients de concordance et procédures d'agrégation associées ; on consultera à ce sujet l'ouvrage de MARCOTORCHINO et MICHAUD (1979). Les deux distances précédentes sont toutefois privilégiées à plusieurs égards – noter que la première a été caractérisée axiomatiquement (cf. notamment, BARTHELEMY 1979) – ce qui justifie l'étude approfondie – et non achevée – des mesures de concordances et des méthodes de consensus qui leur sont associées.

Comme nous l'avions écrit au début de cette introduction, le cas le plus classique de données ordinales est le cas des préordres totaux. Mais la définition de coefficients de concordance, ne serait-ce qu'entre deux préordres totaux. n'est pas un problème simple, comme en témoigne la littérature importante et parfois étrange sur ce sujet (cf., par exemple, le "τ_a" de KENDALL). Il nous a donc semblé utile avant l'étude de ce problème, de faire le point sur le cas assurément plus simple, mais non toujours bien compris des ordres totaux. Signalons aussi que dans le cadre de la recherche de modèles relationnels rendant compte de préférences exprimées, le préordre total est lui-même un modèle souvent trop contraignant, et doit être remplacé par des modèles moins restrictifs : tournois, matchs, quasi-ordres, ordres d'intervalles, relations valuées d'un certain type... (cf., par exemple, les n° 62 et 63 de Math. Sci. hum., 1978, ou BARTHELEMY, FLAMENT et MONJARDET, 1982). Le problème de définition de coefficients de concordance entre de telles données peut être alors considéré dans le cadre général du problème de la définition de coefficients de concordance entre structures discrètes d'un certain type, problème étudiée notamment par HUBERT (1979) et LERMAN (1981). Signalons toutefois que la plupart des notions et résultats des pages suivantes se généraliseraient aisément au cas des relations de tournoi (relations totales et antisymétriques) qu'on obtient par exemple dans des méthodes de comparaison par paires.

2. HOMOGENEITE D'UNE FAMILLE DE VECTEURS OU D'UN NUAGE DE POINTS

2.1. Notations et rappels

Dans ce texte, $\mathbf{R}^p$ désigne soit l'espace vectoriel euclidien de dimension p, soit l'espace affine euclidien canoniquement associé au vectoriel $\mathbf{R}^p$. Dans le premier cas, un élément de $\mathbf{R}^p$ est appelé vecteur et est noté x ; dans le second cas un élément de $\mathbf{R}^p$ est appelé point et est noté P. On peut écrire $x = \overrightarrow{OP}$ (O étant le point correspondant au vecteur nul). Dans la suite on utilisera ce double langage vectoriel ou affine en précisant éventuellement les correspondances. $E \subset \mathbf{R}^p$ est un sous-ensemble de $\mathbf{R}^p$ dont on suppose seulement qu'il ne contient que des vecteurs de norme $\|x\|$ bornée. $N = (x_1, \ldots, x_i, \ldots x_v)$ est une famille de vecteurs de E. On note $N = (P_1, \ldots, P_i, \ldots, P_v)$ le nuage de points correspondants. Si $N \in E^v$, on note g_N le vecteur moyen de N :

$$g_N = \sum_{i=1}^{v} \frac{x_i}{v} \tag{1}$$

et G_N le point moyen (ou barycentre) correspondant.

On a

$$\|g_N\|^2 = \sum_{(i,j)\in V^2} \frac{\langle x_i , x_j \rangle}{v^2} \tag{2}$$

où dans (2) $\langle x_i , x_j \rangle$ désigne le produit scalaire des vecteurs x_i et x_j, la sommation étant sur tous les couples de $V = \{1, \ldots, i, j, \ldots, v\}$.

On note d la distance euclidienne entre les vecteurs, ou les points, de $\mathbf{R}^p$ et on appelle *inertie* de la famille de vecteurs ou du nuage de points N, la quantité :

$$\text{in}(N) = \sum_{i=1}^{v} d^2(x_i , g_N) = \sum_{i=1}^{v} d^2(P_i , G_N) \tag{3}$$

Noter que cette quantité est parfois appelée le "moment centré d'ordre 2". On a les formules classiques :

$$\text{in}(N) = \sum_{i=1}^{v} \|x_i\|^2 - v\,\|g_N\|^2 \tag{4}$$

$$\text{in}(N) = \sum_{(i,j)\in V^2} \frac{d^2(x_i , x_j)}{2v} \tag{5}$$

Enfin, on pose pour $N \in E^v$:

$$R^2(N) = \sum_{i=1}^{v} \|x_i\|^2 \tag{6}$$

$$r^2(N) = \frac{R^2(N)}{v} \tag{7}$$

On a

$$\|g_N\|^2 \leqslant r^2(N)$$

$$in(N) = v(r^2(N) - \|g_N\|^2) \tag{8}$$

Soient $x_i , x_j \in N$; on pose

$$\alpha_{ij}(N) = 1 - \frac{d^2(x_i , x_j)}{2r^2(N)} \tag{9}$$

On a donc $\alpha_{ij} = \alpha_{ji}$ et $\alpha_{ii} = 1$. Un calcul simple donne :

$$\alpha_{ij}(N) = \frac{2\langle x_i , x_j\rangle + 2r^2(N) - (\|x_i\|^2 + \|x_j\|^2)}{2r^2(N)} \tag{10}$$

En particulier, si pour tout i, $\|x_i\| = r$, $\alpha_{ij} = \cos(x_i , x_j)$, on peut exprimer l'inertie de N en utilisant les coefficients α_{ij} :

$$in(N) = R^2(N)\left(1 - \sum_{(i,j)\in v^2} \frac{\alpha_{ij}}{v^2}\right) \tag{11}$$

N.B. : pour alléger les formules, on pourra supprimer parfois le symbole N dans les notations.

2.2. Coefficient d'homogénéité H_1

On obtient l'homogénéité H_1 de la famille N en normant de façon "intrinsèque" l'inertie de cette famille :

$$H_1(N) = \left[1 - \frac{in(N)}{R^2(N)}\right] \tag{12}$$

On a immédiatement, à partir des formules (7) et (8) :

$$H_1(N) = \frac{\|g_N\|^2}{r^2(N)} \tag{13}$$

qui montre que H_1 est le rapport du carré de la norme du vecteur moyen à la moyenne des carrés des normes.

D'autre part les formules (12) et (5) donnent :

$$H_1(N) = 1 - \sum_{(i,j)\in v^2} \frac{d^2(x_i , x_j)}{2v^2 r^2(N)} \tag{14}$$

d'où l'on déduit, d'après (9), que $H_1(N)$ est aussi la moyenne des v^2 coefficients α_{ij} :

$$H_1(N) = \sum_{(i,j)\in V^2} \frac{\alpha_{ij}}{v^2} \tag{15}$$

La formule (12) montre qu'on a toujours

$$0 \leqslant H_1(N) \leqslant 1$$

avec $H_1(N) = 0$ ssi $g_N = 0$
et $H_1(N) = 1$ ssi $N = (x, \ldots, x)$ (i.e. $\mathrm{in}(N) = 0$)

Dans l'expression (15), on moyenne sur les v^2 couples α_{ij}. Mais on pourrait aussi moyenner sur les $\binom{v}{2}$ paires $\{i, j\}$. Du fait des égalités $\alpha_{ij} = \alpha_{ji}$, et $\alpha_{ii} = 1$ on obtient l'expression suivante :

$$H'_1(N) = \sum_{\{i,j\}} \frac{\alpha_{ij}}{\binom{v}{2}} = \frac{v H_1 - 1}{v - 1} \tag{16}$$

2.3. Coefficient d'homogénéité H_2

On se place maintenant dans le cas où les familles N considérées appartiennent toutes à un ensemble connu $\mathcal{N} : \mathcal{N} \subseteq E^v$. Pour définir le coefficient H_2, on utilise l'inertie de N normée par l'inertie maximum qu'on peut obtenir dans $\mathcal{N}$:

$$H_2(N) = 1 - \frac{\mathrm{in}(N)}{\max\limits_{N\in \mathcal{N}} \mathrm{in}(N)} \tag{17}$$

ou encore

$$H_2(N) = 1 - \frac{r^2(N) - \|g_N\|^2}{\max\limits_{N\in \mathcal{N}} (r^2(N) - \|g_N\|^2)} \tag{18}$$

On a clairement

$$0 \leqslant H_2(N) \leqslant 1$$

avec $\quad H_2(N) = 0$ ssi $\mathrm{in}(N) = \max\limits_{N\in \mathcal{N}} \mathrm{in}(N)$

$H_1(N) = 1$ ssi $N = (x, \ldots, x)$

2.4. Homogénéité de nuages sphériques

On suppose maintenant que le sous-ensemble E de $\mathbf{R}^p$ auquel on s'intéresse est formé des vecteurs de même norme r. Autrement dit E est une sphère centrée à l'origine et de rayon r ; on la note B_r. Un nuage de points sur cette sphère est appelée sphérique. D'autre part, on considérera toujours qu'on s'intéresse à un certain ensemble $\mathcal{N}$ de nuages, ensemble qui est donc inclus dans $(B_r)^v$. De plus, on supposera toujours dans la suite que $\mathcal{N}$ contient des nuages "constants", c'est-à-dire formés de v points identiques.

Soit $N = (x_1, \ldots, x_i, \ldots, x_v) \in \mathcal{N}$; puisque pour tout i, $\|x_i\| = r$, on a :

$$R^2(N) = v\, r^2 \qquad r^2(N) = r^2 \qquad \langle x_i , x_j \rangle = r^2 \cos(x_i , x_j)$$

d'où
$$\|g_N\|^2 = r^2 \sum_{(i,j)\in V^2} \frac{\cos(x_i , x_j)}{v^2} \tag{19}$$

$$\alpha_{ij} = \cos(x_i , x_j) = 1 - \frac{d^2(x_i , x_j)}{2r^2} \tag{20}$$

Les formules (8) et (11) donnent alors :

$$\text{in}(N) = v(r^2 - \|g_N\|^2) \tag{21}$$

$$\text{in}(N) = \frac{r^2}{v}\left[v^2 - \sum_{(i,j)\in V^2} \cos(x_i , x_j)\right] \tag{22}$$

Nous considérons maintenant ce que deviennent les coefficients d'homogénéité H_1 et H_2 définis ci-dessus.

2.4.1. Coefficient d'homogénéité H_1

La formule (13) donne

$$H_1(N) = \frac{\|g_N\|^2}{r^2} \tag{23}$$

d'où,d'après (19)

$$H_1(N) = \sum_{(i,j)\in V^2} \frac{\cos(x_i , x_j)}{v^2} \tag{24}$$

Ainsi $H_1(N)$ a une double interprétation : rapport entre le carré de la norme du vecteur moyen et le carré du rayon de B_r ; moyenne par rapport aux v^2 couples (i , j) des cosinus des vecteurs de la famille.

L'homogénéité H_1 est maximum et égale à 1 pour les familles constantes ; elle est minimum (et $\geqslant 0$) pour les familles de N minimisant $\|g_N\|$.

On pose

$$\epsilon_{\mathcal{N}} = \min_{N\in\mathcal{N}} \|g_N\| \tag{25}$$

On a donc

$$0 \leqslant \frac{\epsilon^2_{\mathcal{N}}}{r^2} \leqslant H_1(N) \leqslant 1$$

Nous allons donner d'autres interprétations de $H_1(N)$ dans le cas où les vecteurs x_i considérés sont centrés (la somme de leurs coordonnées est nulle).

Soit donc $N \in (B_r \cap H_0)^v$, où H_0 désigne l'hyperplan $\sum_{k=1}^{p} x^k = 0$ et x^k la $k^{\text{ième}}$ coordonnée du vecteur x. $X_N = \sum_{i=1}^{v} x_i$ est centré, de même que $g_N = \frac{X_N}{v}$ et on a, en notant var x la variance du vecteur x :

$$\operatorname{var} g_N = \frac{\|g_N\|^2}{p} = \frac{\operatorname{var} X_N}{v^2}$$

De plus pour une famille constante $N = (x, \ldots, x) \in (B_r \cap H_0)^v$, on a

$$\|g_N\|^2 = r^2 = \max_{\mathfrak{N}} \|g_N\|^2 = p \max_{\mathfrak{N}} \operatorname{var} g_N = \frac{p}{v^2} \max_{\mathfrak{N}} \operatorname{var} X_N$$

De (23) et des expressions ci-dessus, on déduit :

$$H_1(N) = \frac{\operatorname{var} g_N}{\max\limits_{\mathfrak{N}} \operatorname{var} g_N} = \frac{\operatorname{var} X_N}{\max\limits_{\mathfrak{N}} \operatorname{var} X_N} \tag{26}$$

ce qui donne l'homogénéité H_1 comme rapport de variances (c'est une variance normée).

Supposons maintenant que N étant toujours formé de vecteurs centrés, on opère une transformation affine sur ces vecteurs : $x \mapsto y = \alpha(x + b)$ où α est un scalaire et b le vecteur constant de coordonnées toutes égales à β. Appelons N' la famille des transformés, $N' = (y_1, \ldots, y_i, \ldots, y_n)$ et y_i^k la $k^{\text{ième}}$ coordonnée du vecteur y_i.

On a

$$\|y_i\|^2 = \alpha^2 \|x_i + b\|^2 = \alpha^2(r^2 + p\beta^2)$$

$$\sum_{k=1}^{p} y_i^k = p\,\alpha\,\beta \tag{26 bis}$$

Donc N' est un nuage sphérique pour la sphère de rayon $r' = \alpha\sqrt{r^2 + p\beta^2}$, contenue dans l'hyperplan d'équation donnée par (26 bis).

D'autre part

$$Y_{N'} = \sum_{i=1}^{v} y_i = \alpha \sum_{i=1}^{v} (x_i + b) = \alpha(X_N + vb)$$

$$\operatorname{Var} Y_{N'} = \alpha^2 \operatorname{var} X_N$$

$$\max_{\mathfrak{N}} \operatorname{var} Y_{N'} = \alpha^2 \max_{\mathfrak{N}} \operatorname{var} X_N = \frac{\alpha^2 v^2 r^2}{p} = (r'^2 - p\beta^2 \alpha^2)\frac{v^2}{p}$$

D'où, en utilisant (26) et en posant $\mathfrak{N}' = \{N', N \in \mathfrak{N}\}$:

$$H_1(N) = \frac{\operatorname{var} Y_{N'}}{\max\limits_{\mathfrak{N}'} \operatorname{var} Y_{N'}} = \frac{p \operatorname{var} Y_{N'}}{v^2 \alpha^2 r^2} \tag{27}$$

2.4.2. Coefficient d'homogénéité H_2

On déduit des formules (17), (18) et de la définition (25) l'expression :

$$H_2(N) = 1 - \frac{\operatorname{in}(N)}{v(r^2 - \epsilon_{\mathfrak{N}}^2)} \tag{28}$$

On a

$$0 \leqslant H_2(N) \leqslant 1$$

avec $H_2(N)$ nul pour les familles d'inertie maximum (i.e. minimisant $\|g_N\|$ et $H_2(N) = 1$ pour les familles constantes.

Dans ce cas sphérique, on peut aussi exprimer H_2 en fonction de H_1. D'abord, puisque $H_1 = 1 - \frac{\text{in N}}{vr^2}$, on a toujours ;

$$O \leqslant H_2(N) \leqslant H_1(N) \leqslant 1 \tag{29}$$

De plus, en reportant $\text{inN} = vr^2(1 - H_1(N))$ et $\epsilon^2_{\mathfrak{N}} = r^2 \min_{\mathfrak{N}} H_1(N)$ dans (28), on obtient :

$$H_2(N) = \frac{H_1(N) - \min_{\mathfrak{N}} H_1(N)}{1 - \min_{\mathfrak{N}} H_1(H)} \tag{30}$$

En particulier,

. si $\min_{\mathfrak{N}} H_1(N) = 0$, on a $H_2 \equiv H_1$ (31)

. si $\min_{\mathfrak{N}} H_1(N) = \frac{1}{v^2} \left(\text{i.e. } \epsilon^2_{\mathfrak{N}} = \frac{r^2}{v^2}\right)$, on a $H_2 \equiv \frac{v^2 H_1 - 1}{v^2 - 1}$ (32)

3. CODAGES VECTORIELS DES ORDRES TOTAUX

3.1. Notations et rappels

On considère un ensemble $X = \{a, b, c, \ldots, x, y, z, \ldots, n\}$ à n éléments. O désigne un ordre total sur X, i.e. une relation binaire, réflexive, transitive, antisymétrique et totale sur X. Noter que par O relation binaire sur X, on entend que O est un ensemble de couples de X, et qu'on écrit indifféremment $(x, y) \in O$ ou $x\,O\,y$. O^r désigne l'ordre réciproque (parfois appelé ordre dual ou opposé) de l'ordre O : $x\,O^r$ ssi $y\,O\,x$.

On peut écrire un ordre total sur X sous la forme d'une "permutation" des n éléments de X, rangés du "plus petit" au "plus grand" élément. Par exemple, si $X = \{a, b, c\}$, bca désigne l'ordre total $\{(b, c), (b, a), (c, a), (b, b), (c, c), (a, a)\}$. Nous utiliserons cette écriture dans les exemples.

Nous notons $\mathcal{O}_X$, $\mathcal{O}_n$ ou simplement $\mathcal{O}$ l'ensemble des n ! ordres totaux définis sur l'ensemble X. Soient O et $O' \in \mathcal{O}$; la différence symétrique entre O et O' est définie par :

$$O \,\Delta\, O' = (O - O') \cup (O' - O) = (O \cup O') - (O \cap O')$$

La classique *distance de la différence symétrique* entre O et O' est définie par :

$$d_\Delta(O, O') = |O \,\Delta\, O'| = |O \cup O'| - |O \cap O'| \tag{33}$$

elle dénombre les couples en "désaccord" entre O et O'(i.e. les couples (x, y) avec, par exemple, $x\,O\,y$ et $y\,O'x$). La distance de KENDALL entre O et O' est définie par :

$$d_K(O, O') = |O - O'| = |O' - O| = \frac{d_\Delta(O, O')}{2} \tag{34}$$

elle dénombre les paires d'éléments de X sur lesquelles O et O' sont en "désaccord". Cette distance varie donc de O à $\frac{n(n-1)}{2}$.

Le coefficient appelé τ de KENDALL est obtenu par une transformation affine de la distance de KENDALL conduisant à attribuer la valeur + 1 à l'accord complet (O' = O) et − 1 au désaccord total ($O' = O^r$).

Donc $$\tau(O , O') = 1 - \frac{4\, d_K(O , O')}{n(n-1)} \tag{35}$$

Exemple 1

$O = abcd, O' = cbda \,;\, d_\Delta(O , O') = 8 \,;\, d_K(O , O') = 4 \,;\, \tau(O , O') = -\frac{1}{3}$

Soient $O \in \mathcal{O}$ et $x \in X$; on appelle *rang* de x dans O le nombre :

$$r_O(x) = |\{y \in X : y\, O\, x\}|$$

On appelle *score* de x dans O le nombre :

$$s_O(x) = |\{y \in X : x\, O\, y\}| - |\{y \in X : y\, O\, x\}|$$

On a donc

$$s_O(x) = n + 1 - 2r_O(x)$$

Soit un ordre total arbitraire défini sur X ; on peut l'écrire sous la forme $x_1\ x_2 \,..\, x_k \,..\, x_n$. Soient O et O' deux ordres totaux ; posons $r_O(x_k) = r_k$, $r_{O'}(x_k) = r'_k$.

La distance de SPEARMAN d_S entre O et O' est définie par :

$$d_S(O , O') = \sqrt{\sum_{k=1}^{n} (r_k - r'_k)^2} \tag{36}$$

On sait que cette distance varie de O à $\sqrt{\frac{n^3 - n}{3}}$ (valeur obtenue pour $O' = O^r$).

Le coefficient appelé ρ de SPEARMAN est obtenu par une transformation affine de d_S^2 conduisant à attribuer la valeur + 1 à l'accord complet et la valeur − 1 au désaccord complet.

Donc $$\rho(O , O') = 1 - \frac{6\, d_S^2(O , O')}{n^3 - n} \tag{37}$$

Exemple 2

$O = abcd, O' = cbda \,;\, d_S(O , O') = \sqrt{14} \,;\, \rho(O , O') = \frac{2}{5}$

On considère un ensemble $V = \{1, 2, .. , i, .. , v\}$ à v éléments. On appelle V-profil ou simplement profil d'ordres totaux, la donnée d'une application π de V dans $\mathcal{O}$. On écrit un profil sous la forme $\pi = (O_1 , .. O_i , .. O_v)$. On désigne par $\mathcal{O}^V$ l'ensemble de tous ces profils. Pour un profil $\pi = (O_1 , .. , O_i , .. O_v)$, on pose.

$$V(x , y) = \{i \in V : (x , y) \in O_i\} \,;\, v(x , y) = |V(x , y)| \,;\, w(x , y) = v(x , y) - v(y , x)$$

Si $x \neq y$, $v(x, y) + v(y, x) = v$; $w(x, y) = 2v(x, y) - v$
Si $x = y$, $v(x, y) + v(y, x) = 2v$; $w(x, x) = 0$.

Exemple 3

π = (abcd, bcad, cbda, bacd, dcba)

On a écrit ci-dessous dans des tableaux de "comparaison par paires", les valeurs des $v(x, y)$ et $w(x, y)$:

v(x , y)	a	b	c	c
a	5	1	2	3
b	4	5	3	4
c	3	2	5	4
d	2	1	1	5

w(x , y)	a	b	c	d
a	0	−3	−1	1
b	3	0	1	3
c	1	−1	0	3
d	−1	−3	−3	0

3.2. Codages de KENDALL.

Nous appellerons "codage de KENDALL" d'un ordre total, un codage qui associe à un ordre total un vecteur de $\mathbf{R}^{n^2}$. Nous considérons deux tels codages : δ et γ

3.2.1. Codage caractéristique δ

Les n^2 couples de X^2 étant rangés dans un ordre total arbitraire, le codage δ de l'ordre O associe à chacun de ces couples le nombre 0 ou 1 :

$$\delta(x, y) = 1 \quad \text{ssi} \quad (x, y) \in 0$$
$$\delta(x, y) = 0 \quad \text{ssi} \quad (x, y) \notin 0$$

On note δ_0, ou simplement δ, le vecteur ainsi associé à 0.

On a :
$$\|\delta_0\|^2 = \frac{n(n+1)}{2} \tag{38}$$

Si l'on note $\delta_{0'}$, ou simplement δ', le vecteur associé à l'ordre O', on remarque que le carré de la distance euclidienne entre ces deux vecteurs de $\mathbf{R}^{n^2}$ vérifie :

$$d^2(\delta, \delta') = d_\Delta(O, O') = 2d_K(O, O') \tag{39}$$

3.2.2. Codage caractéristique centré γ

On peut définir ce codage à partir de δ en posant pour tout couple (x, y) de X^2 :

$$\gamma(x, y) = \delta(x, y) - \delta(y, x)$$

Donc
$$\gamma(x, y) = +1 \quad \text{ssi} \quad (x, y) \in 0 \text{ et } x \neq y$$
$$\gamma(x, y) = -1 \quad \text{ssi} \quad (x, y) \notin 0$$
$$\gamma(x, y) = 0 \quad \text{ssi} \quad x = y$$

En notant γ_0, ou γ, le vecteur de $\mathbf{R}^{n^2}$ associé à 0, γ' le vecteur associé à O' on a :

$$\|\gamma_0\|^2 = n(n-1) \tag{40}$$

$$d^2(\gamma, \gamma') = 4\, d_\Delta(O, O') = 8\, d_K(O, O') \tag{41}$$

On a donc un codage dans la sphère de rayon $[n(n-1)]^{1/2}$. A partir des formules (20), (34), (40) et (41), on obtient immédiatement :

$$\underline{\tau(O, O') = \cos(\gamma, \gamma')} \tag{42}$$

Puisque les vecteurs γ sont centrés, le coefficient de corrélation τ de KENDALL est donc aussi égal au coefficient de corrélation (linéaire) entre les vecteurs γ et γ' de $\mathbf{R}^{n^2}$, et il est parfois défini de cette manière.

Exemple 4

π = (abc, bac, cba). Les vecteurs γ correspondant aux 3 ordres de ce profil sont représentés ci-dessous :

	(a, a)	(a, b)	(a, c)	(b, b)	(b, a)	(b, c)	(c, c)	(c, b)	(c, a)
abc	0	+1	+1	0	−1	+1	0	−1	−1
bac	0	−1	+1	0	+1	+1	0	−1	−1
cba	0	−1	−1	0	+1	−1	0	+1	+1

3.3. Codages de SPEARMAN

Un "codage de SPEARMAN" va associer à un ordre total un vecteur de $\mathbf{R}^n$. On va considérer les codages r et s.

3.3.1. Codage du rang

Soit un ordre total aribitraire défini sur X écrit sous la forme $x_1\, x_2 \ldots x_k \ldots x_n$. Le codage du rang associe à l'ordre total 0 le vecteur de $\mathbf{R}^n$:

$$r_0 = [r_0(x_1), \ldots r_0(x_k), \ldots r_0(x_n)]$$

où $r_0(x_k)$ est le rang de x_k dans l'ordre 0.

Puisque les rangs varient de 1 à n, le carré de la norme d'un tel vecteur est :

$$\|r_0\|^2 = \frac{n(n+1)(2n+1)}{6} \tag{43}$$

Soient 0 et 0' deux ordres totaux, r et r' les vecteurs rangs associés. La distance euclidienne entre r et r' dans $\mathbf{R}^n$ n'est autre que la distance de SPEARMAN précédemment définie entre 0 et 0' :

$$d(r, r') = d_S(0, 0') \tag{44}$$

N.B. le codage du rang est relié au codage caractéristique δ, par la formule :

$$r_0(x) = \sum_{y \in X} \delta_0(y, x) \tag{45}$$

3.3.2. Codage du score

Etant donné l'ordre de référence $x_1\ x_2 \ldots x_k \ldots x_n$, le codage du score associe à l'ordre total 0 le vecteur de $\mathbf{R}^n$:

$$s_0 = [s_0(x_1), \ldots s_0(x_k), \ldots s_0(x_n)]$$

où $s_0(x_k)$ est le score de x_k dans l'ordre 0. On a :

$$s_0 = (n+1\ , \ldots n+1) - 2r_0$$

Le carré de la norme d'un tel vecteur score est :

$$\|s_0\|^2 = \frac{n^3 - n}{3} \tag{46}$$

On a donc un codage dans la sphère de rayon $\left(\frac{n^3 - n}{3}\right)^{1/2}$.

En notant s et s' les vecteurs scores associés à 0 et 0' on a :

$$d(s\ , s') = 2d_S(0\ , 0') \tag{47}$$

D'autre part, à partir des formules (20), (37), (46) et (47), on obtient :

$$\underline{\rho(0\ , 0') = \cos(s\ , s')} \tag{48}$$

De plus les vecteurs scores étant centrés, leurs cosinus sont donc égaux à leurs coefficients de corrélation (linéaire), et puisque le vecteur score est transformé affine du vecteur rang, on obtient

$$\underline{\rho(0\ , 0') = \text{coefficient de corrélation entre r et r}'} \tag{49}$$

formule qui, depuis SPEARMAN, est souvent prise comme définition de ρ.

N.B. 1 : Il existe d'autres variantes de ces codages. Par exemple, BENZECRI (1965) utilise le codage du score "normé", i.e. les vecteurs $\dfrac{s}{\sqrt{\dfrac{n^3 - n}{3}}}$.

2 : Le codage du score est relié au codage caractéristique centré γ, par la formule :

$$s_0(x) = \sum_{y \in X} \gamma_0(x\ , y) \tag{50}$$

Exemple 5

Pour π = (abcd, bcad, cbda, bacd, dcba) et l'ordre de référence abcd, on a représenté ci-dessous les vecteurs rangs et scores correspondants :

	a	b	c	d
r_1	1	2	3	4
r_2	3	1	2	4
r_3	4	2	1	3
r_4	2	1	3	4
r_5	4	3	2	1

	a	b	c	d
s_1	3	1	−1	−3
s_2	−1	3	1	−3
s_3	−3	1	3	−1
s_4	1	3	−1	−3
s_5	−3	−1	1	−3

Remarque :

Chacun des quatre codages δ, γ, r, s, définis ci-dessus plonge l'ensemble des n ! ordres totaux dans un espace affine (euclidien). Il est facile de voir que les n ! points obtenus sont les sommets d'un polyèdre convexe. Nous appellerons permutoèdre un tel polyèdre et pour préciser, nous appellerons, par exemple, r-permutoèdre le polyèdre obtenu avec le codage r. On remarque que pour le r-permutoèdre la distance de SPEARMAN entre deux ordres totaux est la distance euclidienne entre les deux sommets correspondants, tandis que la distance de KENDALL est le nombre minimum d'arêtes du polyèdre qu'il faut parcourir pour aller d'un de ces sommets à l'autre.

4. LE COEFFICIENT DE CONCORDANCE K ET SES VARIANTES

Ce coefficient est obtenu à partir du coefficient d'homogénéité H_1 et du codage γ dans $\mathbf{R}^{n^2}$. Nous avons besoin de notations supplémentaires. Soit $\pi = (O_1, \dots, O_i, \dots O_v)$ un profil d'ordres totaux ; nous notons $\pi_\gamma = (\gamma_1, \dots \gamma_i, \dots \gamma_v)$ la famille de vecteurs de $\mathbf{R}^{n^2}$, définie par $\gamma_i = \gamma(O_i)$; nous notons g_γ le vecteur moyen de cette famille : $g_\gamma = \sum_{i=1}^{v} \frac{\gamma_i}{v}$. On définit de même π_δ et g_δ. D'après les définitions des nombres v(x , y) et w(x , y) (3.1), il est immédiat que :

$$\text{coordonnées de } g_\delta = \text{les } n^2 \text{ nombres } \frac{v(x\,,y)}{v} \tag{51}$$

$$\text{coordonnées de } g_\gamma = \text{les } n^2 \text{ nombres } \frac{w(x\,,y)}{v} \tag{52}$$

Définition

$$K(\pi) = H_1(\pi_\gamma) \tag{53}$$

Puisque π_γ correspond à un nuage de points dans la sphère de rayon $\sqrt{n(n-1)}$ (formule 40), on peut utiliser les formules du § 2.4.1. Les formules (23), (52) d'une part et (24), (42) d'autre part donnent les deux expressions suivantes pour K :

$$K(\pi) = \frac{\|g_\gamma\|^2}{n(n-1)} = \sum_{(x,y)\in X^2} \frac{w^2(x\,,y)}{v^2\, n(n-1)} \tag{54}$$

$$K(\pi) = \sum_{(i,j)\in V^2} \frac{\tau_{ij}}{v^2} \tag{55}$$

La formule (54) donne l'interprétation géométrique de K. Le γ-permutoèdre (remarque 3.3.2) est inscrit dans la sphère de rayon $\sqrt{n(n-1)}$. Au profil π est associé un nuage de points, sommets de ce permutoèdre, et K est le rapport entre le carré de la distance du point moyen au centre de la sphère, et le carré du rayon de cette sphère. La formule (54) permet aussi de calculer facilement K à partir des nombres $w(x, y)$, déduits du tableau des "comparaisons par paires" entre x et y. On peut aussi exprimer K directement en fonction des nombres $v(x, y)$, il suffit de remplacer dans (54), $w(x, y)$ par $2v(x, y) - v$. On obtient alors :

$$K(\pi) = 4 \sum_{\substack{(x,y) \\ x \neq y}} \frac{v^2(x, y)}{v^2 n(n-1)} - 1 \tag{56}$$

La formule (55) montre que K n'est autre que la moyenne pour tous les couples d'ordres des coefficients τ de KENDALL. Si on moyenne sur toutes les paires d'ordres, on obtient le coefficient K' de la formule (16), coefficient traditionnellement appelé U, on a donc :

$$K'(\pi) = U(\pi) = \frac{vK(\pi) - 1}{v - 1} = 4 \sum_{\substack{(x,y) \\ x \neq y}} \frac{v^2(x, y)}{v(v-1)n(n-1)} - \frac{(v+1)}{v-1} \tag{57}$$

Evidemment, on peut aussi écrire K (ou U) en fonction des distances de KENDALL entre les ordres du profil. On obtient, par exemple, à partir des formules (14) et (41)

$$K(\pi) = 1 - \frac{4 \sum\limits_{(i,j)} d_K(O_i, O_j)}{v^2 n(n-1)} \tag{58}$$

La valeur maximum 1 du coefficient K pour tous les profils de O^v est obtenue ssi tous les ordres totaux sont identiques. La valeur minimum dépend de la parité de v. Si v est pair, elle est nulle (considérer un profil composé de v/2 ordres identiques et de v/2 ordres réciproques du premier). Si v est impair, elle est égale à $(v^2)^{-1}$ (considérer le même profil que ci-dessus auquel on ajoute un ordre arbitraire).

En conclusion, K varie entre les bornes suivantes :

$$\begin{array}{ll} v \text{ pair} & 0 \leqslant K(\pi) \leqslant 1 \\ v \text{ impair} & \dfrac{1}{v^2} \leqslant K(\pi) \leqslant 1 \end{array} \tag{59}$$

On en déduit les bornes pour $U(\pi)$:

$$\begin{array}{ll} v \text{ pair} & -\dfrac{1}{v-1} \leqslant U(\pi) \leqslant 1 \\ v \text{ impair} & -\dfrac{1}{v} \leqslant U(\pi) \leqslant 1 \end{array} \tag{60}$$

N.B. : on utilise parfois une quantité $\Sigma(\pi)$ ne prenant que des valeurs entières. On a :

$$\Sigma(\pi) = \sum_{\substack{(x,y)\in X^2 \\ x \neq y}} \binom{v(x,y)}{2} = \frac{v\,n(n-1)}{8}[v\,K(\pi) + v - 2] =$$

$$\binom{v}{2}\binom{n}{2}\left(\frac{1+U}{2}\right) = \frac{1}{8}\left[\sum_{(x,y)\in X^2} w^2(x,y) + n(n-1)\,v(v-2)\right] =$$

$$\binom{n}{2}\binom{v}{2} - \frac{1}{2}\sum_{(i,j)\in V^2} d_K(O_i, O_j)$$

$$\text{si v pair : } \binom{n}{2}\frac{v(v-2)}{4} \leqslant \Sigma(\pi) \leqslant \binom{n}{2}\binom{v}{2}$$

$$\text{si v impair : } \binom{n}{2}\left(\frac{v-1}{4}\right)^2 \leqslant \Sigma(\pi) \leqslant \binom{n}{2}\binom{v}{2}$$

On a défini ci-dessus K au moyen de H_1. L'utilisation de H_2 conduit pratiquement au même résultat. En effet, compte-tenu des bornes précédentes pour K, on peut utiliser les formules (31) et (32) § 2.4.2, et on obtient :

$$\text{si v est pair } H_2(\pi_\gamma) = K(\pi) = \frac{(v-1)\,U(\pi) + 1}{v} \tag{61}$$

$$\text{si v est impair } (> 1)\ H_2(\pi_\gamma) = \frac{v^2 K(\pi) - 1}{v^2 - 1} = \frac{v\,U(\pi) + 1}{v+1} \tag{62}$$

Le coefficient ainsi obtenu varie donc dans tous les cas de 0 à 1 ; il est toujours égal à K pour v pair ; pour v impair, il ne pourra être sensiblement différent de K que si v et K sont tous les deux très petits.

Exemple 6

Pour le profil de l'exemple 3, on obtient $K(\pi) = \dfrac{\Sigma w^2(x,y)}{25 \times 4 \times 3} = \dfrac{60}{300} = \dfrac{1}{5}$; $U(\pi) = \dfrac{5\,K(\pi) - 1}{5-1} = 0$; $H_2(\pi_\gamma) = \dfrac{25\,K(\pi) - 1}{25 - 1} = \dfrac{1}{6}$.

$K(\pi)$ s'obtient aussi à partir des deux tableaux suivants donnant les distances et les τ de KENDALL entre tous les couples d'ordre du profil :

d_K	1	2	3	4	5
1	0	2	4	1	6
2	2	0	2	1	4
3	4	2	0	3	2
4	1	1	3	0	5
5	6	4	2	5	0

τ	1	2	3	4	5
1	1	1/3	−1/3	2/3	−1
2	1/3	1	1/3	2/3	−1/3
3	−1/3	1/3	1	0	1/3
4	2/3	2/3	0	1	−2/3
5	−1	−1/3	1/3	−2/3	1

$$K(\pi) = 1 - \frac{4\,\Sigma\, d_K(O_i, O_j)}{25 \times 4 \times 3} = \frac{1}{5} \qquad K(\pi) = \frac{\Sigma\,\tau_{ij}}{v^2} = \frac{5}{25} = \frac{1}{5}$$

Remarques

1) On peut aussi définir le coefficient $K(\pi)$ comme une variance normée. En effet, le codage caractéristique γ est sphérique et centré. Si donc on considère $\pi = (\gamma_1, \dots \gamma_1, \dots \gamma_u)$ et le vecteur $\Gamma = \sum_{i=1}^{v} \gamma_i$, on a par la formule (26)

$$K(\pi) = H_1(\pi_\gamma) = \frac{\text{var}\,\Gamma}{\max \text{var}\,\Gamma} \qquad (63)$$

Par contre, si on considère le codage caractéristique δ, et qu'on pose $\Delta = \sum_{i=1}^{v} \delta_i$, on obtient après calcul

$$\frac{\text{var}\,\Delta}{\max \text{var}\,\Delta} = \frac{nK(\pi) + 1}{n + 1}$$

Pour obtenir une formule donnant aussi $K(\pi)$ comme rapport de variances, il faut modifier le codage Σ, par exemple, en posant $\Sigma(x, x) = 1/2$ pour tout x, ou en considérant la restriction de δ à $\mathbf{R}^{n(n-1)}$ obtenue en ne codant pas les couples (x, x). Ainsi, si on note δ^r ce dernier codage et $\Delta^r = \sum_{i=1}^{v} \delta_i^r$, on obtient

$$K(\pi) = \frac{\text{var}\,\Delta^r}{\max \text{var}\,\Delta^r} \qquad (64)$$

2) La définition (53) de $K(\pi)$, et les formules (61) et (62) utilisent les coefficients d'homogénéité H_1 ou H_2 avec le codage caractéristique centré γ. On pourrait également utiliser ces mêmes coefficients avec les codages caractéristiques δ ou δ^r. Mais les formules (5), (39) et (41) donnent :

$$\text{in}(\pi_\delta) = \text{in}(\pi_\delta^r) = \frac{\text{in}(\pi_\gamma)}{4}$$

On en déduit $H_2(\pi_\delta) = H_2(\pi_\delta^r) = H_2(\pi_\gamma)$

et après calcul :

$$H_1(\pi_\delta) = \frac{(n-1)\,K(\pi) + (n+3)}{2(n+1)}$$

Ainsi, par exemple, pour v pair, ce dernier coefficient varie de $\frac{n+3}{2(n+1)}$ à 1.

5. LE COEFFICIENT DE CONCORDANCE W ET SES VARIANTES

Ce coefficient est obtenu à partir du coefficient d'homogénéité H_1 et du codage score dans $\mathbf{R}^n$. Nous notons $\pi_s = (s_1, \dots s_i, \dots s_v)$ la famille des vecteurs scores codant le profil $\pi = (O_1, \dots O_i, \dots O_v)$: $s_i = s_{O_i}$.

On note g_s le vecteur moyen de cette famille : $g_s = \sum_{i=1}^{v} \frac{s_i}{v}$.

En posant $S(x) = \sum_{i=1}^{v} s_i(x)$ (somme des scores de x), on a donc :

$$\text{coordonnées de } g_s = \text{les n scores moyens } \frac{S(x)}{v}$$

On définit de même pour le codage rang, π_r, g_r et $R(x) = \sum_{i=1}^{v} r_i(x)$ (somme des rangs de x). On a :

$$\text{coordonnées de } g_r = \text{les n rangs moyens } \frac{R(x)}{v}$$

Définition

$$W(\pi) = H_1(\pi_s) \tag{65}$$

Puisque π_s correspond à un nuage de points de la sphère de rayon $\left(\frac{n^3 - n}{3}\right)^{1/2}$ (formule 46) on peut utiliser les formules du § 2.4.1. . On obtient :

$$W(\pi) = \frac{\|g_s\|^2}{\frac{1}{3}(n^3 - n)} = 3 \frac{\sum_{x \in X} S^2(x)}{v^2(n^3 - n)} \tag{66}$$

$$W(\pi) = \sum_{(i,j) \in V^2} \frac{\rho_{ij}}{v^2} \tag{67}$$

La formule (66) donne l'interprétation géométrique de W, lorsque π_γ est considéré comme un ensemble de sommets du s-permutoèdre, inscrit dans la sphère de rayon $\left(\frac{n^3 - n}{3}\right)^{1/2}$. Elle conduit au calcul simple de W à partir des sommes des scores S(x). On peut aussi calculer W à partir des sommes des rangs R(x). Il suffit de remplacer S(x) par $v(n + 1) - 2R(x)$ pour obtenir :

$$W(\pi) = \frac{12 \sum_x R^2(x)}{v^2(n^3 - n)} - \frac{3(n + 1)}{n - 1} \tag{68}$$

Noter aussi la formule suivante, utilisée par KENDALL :

$$W(\pi) = \frac{12\,S}{v^2(n^3 - n)}, \text{avec } S = \sum_{x \in X} \left[\left(\frac{S(x)}{2}\right)\right]^2 = \sum_{x \in X} \left[\left(R(x) - \frac{v(n + 1)}{2}\right)\right]^2 \tag{69}$$

On peut aussi obtenir des formules de calcul de W, ne faisant intervenir que les tableaux de comparison par paires v(x , y) ou w(x , y). En effet, des formules (45) et (50) reliant les codages du rang ou du score aux codages caractéristiques, on déduit immédiatement :

$$R(x) = \sum_{i=1}^{v} \sum_{y \in X} \delta_i(y, x) = \sum_{y \in X} v(y, x) \qquad (70)$$

$$S(x) = \sum_{i=1}^{v} \sum_{y \in X} \gamma_i(x, y) = \sum_{y \in X} w(x, y) \qquad (71)$$

d'où,par exemple

$$W(\pi) = \frac{3 \sum_{x \in X} \left[\sum_{y \in X} w(x, y) \right]^2}{v^2(n^3 - n)} \qquad (72)$$

L'expression (67) montre que $W(\pi)$ est la moyenne sur tous les couples d'ordres des coefficients ρ de Spearman. Si on moyenne sur toutes les paires d'ordres, on obtient un coefficient que nous appellons W' et on a d'après (16) et (68)

$$W'(\pi) = \frac{v W(\pi) - 1}{v - 1} = \frac{12 \sum_{x} R^2(x)}{v(v-1)(n^3 - n)} + 1 - \frac{v(4n+2)}{(v-1)(n-1)} \qquad (73)$$

Evidemment, on peut aussi écrire W (ou W′) en fonction des distances de SPEARMAN entre les ordres du profil. On obtient, par exemple, à partir de (14) et (47) :

$$W(\pi) = 1 - \frac{6 \sum_{(i, j)} d_S^2(O_i, O_j)}{v^2(n^3 - n)} \qquad (74)$$

Le coefficient de concordance W prend sa valeur maximum, égale à 1, ssi tous les ordres du profil sont identiques. La détermination de la valeur minimum est moins évidente (et donne souvent lieu à des imprécisions ou des erreurs). On trouve dans KENDALL (1970) l'assertion que ce minimum est nul pour v pair ou n impair. Dans le cas $v = 2p$ c'est évident, puisqu'il suffit de considérer un profil composé de p ordres identiques et de p ordres réciproques du précédent. Pour $v = 2p + 1$, et n impair, on obtient le résultat de KENDALL en montrant qu'il existe toujours un profil π avec $W(\pi) = 0$ i.e. $S(x) = 0$ pour tout x de X. Enfin,dans le cas $v = 2p + 1$ et n pair, on remarque qu'on a toujours $|S(x)| \geqslant 1$ pour tout x de X, donc $W(\pi) \geqslant \dfrac{3}{v^2(n^2 - 1)}$; de plus, il existe toujours un profil réalisant ce minimum, i.e. pour lequel $|S(x)| = 1$, pour tout x. Dans ce cas où $v = 2p + 1$, on obtient des profils réalisant le minimum de W en prenant $(p - 1)$ fois un ordre O quelconque, $(p - 1)$ fois l'ordre O^r, et les trois ordres O_1, O_2, O_3 définis ci-dessous.

Si $n = 2q + 1$, on prend

$O_1 = x_1 \ldots\ldots\ldots\ldots x_q \; x_{q+1} \ldots\ldots\ldots\ldots x_n$

$O_2 : x_{n-1} \; x_{n-3} \ldots\ldots x_2 \; x_n \; x_{n-2} \ldots\ldots\ldots x_1$

$O_3 : x_n \; x_{n-2} \ldots\ldots x_3 \; x_1 \; x_{n-1} \ldots\ldots\ldots x_2$

En effet, en notant $s_i(x)$ le score de x dans l'ordre O_i :

$$\sum_{i=1}^{3} s_i(x_{2k+1}) = s_1(x_{2k+1}) + \left[-q - \frac{s_1(x_{2k+1})}{2} \right] + \left[q - \frac{s_1(x_{2k+1})}{2} \right] = 0$$

$$\sum_{i=1}^{3} s_i(x_{2k}) = s_1(x_{2k}) + \left[q + 1 - \frac{s_1(x_{2k})}{2} \right] + \left[-q - 1 - \frac{s_1(x_{2k})}{2} \right] = 0$$

Si $n = 2q$, on prend

$O_1 : x_1 \ldots\ldots\ldots\ldots\ldots x_q \; x_{q+1} \ldots\ldots\ldots\ldots\ldots x_n$

$O_2 : x_n \; x_{n-2} \ldots\ldots\ldots x_2 \; x_{n-1} \; x_{n-3} \ldots\ldots\ldots x_1$

$O_3 : x_{n-1} \; x_{n-3} \ldots\ldots\ldots x_1 \; x_n \; x_{n-2} \ldots\ldots\ldots x_2$

$$\sum_{i=1}^{3} s_i(x_{2k+1}) = s_1(x_{2k+1}) + \left[-q - \frac{(s_1(x_{2k+1}) - 1)}{2} \right] + \left[q - \frac{(s_1(x_{2k+1}) - 1)}{2} \right] = +1$$

$$\sum_{i=1}^{3} s_i(x_{2k}) = s_1(x_{2k}) + \left[q - \frac{(s_1(x_{2k}) + 1)}{2} \right] + \left[-q - \frac{(s_1(x_{2k}) + 1)}{2} \right] = -1$$

On a donc finalement

$$\left.\begin{array}{l} \text{si v pair ou n impair, } \min W(\pi) = 0 \\ \text{si v impair et n pair, } \min W(\pi) = \dfrac{3}{v^2(n^2 - 1)} \end{array}\right\} \qquad (75)$$

Evidemment dans le second cas, ce minimum est pratiquement nul dès que v ou n sont supérieurs à 4.

Si maintenant on utilise le coefficient d'homogénéité H_2, on a des résultats pratiquement semblables. En effet, la formule (30) et les résultats ci-dessus donnent

$$\text{si v pair ou n impair, } H_2(\pi_s) = W(\pi) \qquad (76)$$

$$\text{si v impair et n pair, } H_2(\pi_s) = \frac{v^2(n^2 - 1)\, W - 3}{v^2(n^2 - 1) - 3} \qquad (77)$$

Donc, dans tous les cas H_2 est égal ou presque égal à W.

Exemple 7

Pour le profil de l'exemple 3, on obtient $W(\pi) = \dfrac{3 \, \Sigma \, S^2(x)}{25 \times 60} = \dfrac{116}{500} = .232$, $W'(\pi) = \dfrac{5\, W(\pi) - 1}{4} = .04$; $H_2(\pi_s) = .226$. On obtient aussi $W(\pi)$ à partir des tableaux suivants donnant les carrés des distances de SPEARMAN et les ρ de SPEARMAN entre tous les couples d'ordres du profil.

d_S^2	1	2	3	4	5
1	0	6	14	2	20
2	6	0	4	2	14
3	14	4	0	10	6
4	2	2	10	0	18
5	20	14	6	18	0

ρ	1	2	3	4	5
1	1	2/5	−2/5	4/5	−1
2	2/5	1	3/5	4/5	−2/5
3	−2/5	3/5	1	0	2/5
4	4/5	4/5	0	1	−4/5
5	−1	−2/5	2/5	−4/5	1

$$W(\pi) = 1 - \frac{\Sigma\, d_S^2(O_i\,,O_j)}{250} = \frac{58}{250} = .\,232 \qquad W(\pi) = \frac{\Sigma\,\rho_{ij}}{25} = \frac{29}{5.25} = .232$$

Remarques

1) On peut donner une interprétation de W comme rapport de variances, en utilisant les formules (25) et (27). Les vecteurs rangs r_i sont en effet obtenus comme transformés affines des vecteurs scores s_i :

$$r_i = -\frac{1}{2}\,[s_i - (n+1)] = \alpha(s_i + b) \quad \text{avec} \quad \alpha = -\frac{1}{2}\,, b = -(n+1)$$

Si on pose $R = \sum_{i=1}^{v} r_i$ et $S = \sum_{i=1}^{v} s_i$,

On a donc

$$W(\pi) = H_1(\pi_s) = \frac{\text{var } S}{\text{max var } S} = \frac{\text{var } R}{\text{max var } R} = \frac{12\sum_x \left[R(x) - \frac{v(n+1)}{2}\right]^2}{v^2(n^3 - n)} \tag{78}$$

La formulation $\dfrac{\text{var } R}{\text{max var } R}$ correspond à la manière dont KENDALL introduit le coefficient W (1962, Chap. 6).

2) Si on utilise le codage du score normé (cf. 3.3.2. N.B.1.) $W(\pi)$ est le carré de la norme du vecteur moyen de la famille correspondante de vecteurs (BENZECRI, 1971).

3) La définition (65) de $W(\pi)$ ou les formules (76) et (77) utilisent les coefficients d'homogénéité H_1 ou H_2 avec le codage du score. On pourrait aussi utiliser le codage du rang. Mais d'après (5) et (47) on a :

$$\text{in}\,(\pi_r) = \frac{\text{in}(\pi_s)}{4}$$

On en déduit :

$$H_2(\pi_r) = H_2\,(\pi_s)$$

et après calcul :

$$H_1\,(\pi_r) = \frac{(n-1)\,W + 3(n+1)}{4n+2}$$

6. RELATION ENTRE LES COEFFICIENTS K ET W

Cette relation va dériver d'une relation fondamentale entre la distance de KENDALL et celle de SPEARMAN entre ordres totaux, dont on va montrer qu'elle est équivalente à l'inégalité de DANIELS (1944) et à la formule de GUILBAUD (1980) reliant le τ et le ρ. Pour cela, nous allons commencer par définir une notion d'écart entre ordres totaux, dérivée du "coefficient σ" de GUILBAUD, et que nous noterons pour cette raison e_G.

Définition

Soient O et O′ deux ordres totaux sur X, {x, y, z} une partie de X à 3 éléments ; on note $O_{\{x,y,z\}}$ la restriction de O à cette partie.

Les ordres O et O′ sont dits en *désaccord (cyclique)* sur {x , y , z} si et seulement si :

$$d_K(O_{\{x,y,z\}}, O'_{\{x,y,z\}}) \equiv 1 \pmod 2).$$

Donc, si par exemple, x O y O z, O′ est en désaccord avec O sur {x , y , z}, ssi on a y O′x O′z, ou x O′z O′y, ou z O′y O′x.

On dit que O et O′ sont *en accord (cyclique)* sur {x , y , z} ssi ils ne sont pas en désaccord, i.e. ssi $d_K(O_{\{x,y,z\}}, O'_{\{x,y,z\}}) \equiv 0 \bmod (2)$.

On pose :

$$D(O\,,O') = \{\text{parties à 3 éléments de X où O et O}' \text{ sont en désaccord}\}$$

$$A(O\,,O') = \{\text{parties à 3 éléments de X, où O et O}' \text{ sont en accord}\}$$

$$e_G(O\,,O') = |D(O\,,O')|\,,\ a_G(O\,,O') = |A(O\,,O')|$$

Par exemple, si O = abcd et O′ = bcad, D(O , O′) = {abd, acd} ;

$$A(O\,,O') = \{abc, bcd\} \text{ et } e_G(O\,,O') = a_G(O\,,O') = 2.$$

Proposition

$e_G(O\,,O')$ est un écart sur l'ensemble $\mathcal{O}_X$ des ordres totaux.

Il est clair que $e_G(O\,,O') = e_G(O'\,,O)$, et que cette quantité est nulle si O égale O′. Par contre, $e_G(O\,,O')$ peut être nul, avec O différent de O′ ; ceci se produit chaque fois que O′ est "transformé circulaire" de O (et seulement dans ce cas) ; par exemple, si X = {a, b, c, d}, O = abcd, et O′ = bcda, ou cdab, ou dabc (ou abcd).

Pour montrer que e_G est un écart, il reste à montrer l'inégalité triangulaire : $e_G(O_1\,,O_3) \leqslant e_G(O_1\,,O_2) + e_G(O_2\,,O_3)$.

Notons $D_{ij} (i \neq j)$ l'ensemble des parties à 3 éléments de X où O_i et O_j sont en désaccord, et montrons

$$D_{13} \subseteq D_{12} \cup D_{23}$$

ce qui implique immédiatement l'inégalité triangulaire.

Soit $\{x , y , z\} \in D_{13}$; on doit montrer que si, par exemple, $\{x , y , z\} \in D_{12}$ $\{x , y , z\} \in D_{23}$. On peut par exemple, supposer $x\, O_1\, y\, O_1\, z$. Puisque $\{x , y , z\} \in D_{13}$, on a alors soit $y\, O_3\, x\, O_3\, z$, soit $x\, O_3\, z\, O_3\, y$, soit $z\, O_3\, y\, O_3\, x$. Puisque $\{x , y , z\} \in D_{12}$, on a alors soit $x\, O_2\, y\, O_2\, z$, soit $y\, O_2\, z\, O_2\, x$, soit $z O_2\, x\, O_2\, y$. On vérifie que dans les neuf cas possibles, $O_{2,\{x,y,z\}}$ et $O_{3,\{x,y,z\}}$ sont en désaccord, i.e. que $\{x , y , z\} \in D_{23}$, c.q.f.d.

Théorème

pour tout $O , O' \in \mathcal{O}_X$, on a

$$d_S^2(O , O') = n\, d_K(O , O') - e_G(O , O') \tag{79}$$

Ce résultat peut s'obtenir à partir de l'inégalié de DANIELS (Corollaire 2, ci-dessous). On peut donc en donner une démonstration analogue, qui utilise les expressions des quantités d_S^2, d_K et e_G en fonction des vecteurs caractéristiques centrés γ (cf. MONJARDET, 1984). Nous donnons ici les étapes d'une démonstration purement combinatoire qu'on trouvera développée dans MONJARDET (1985). Nous avons besoin de définitions supplémentaires. On pose :

$$D_1(O , O') = \left\{ \{x , y , x\} \subseteq X : O_{\{x,y,z\}} = O'^{r}_{\{x,y,z\}} \right\}$$
$$D_2(O , O') = D(O , O') - D_1(O , O') \quad \text{(différence ensembliste)}$$
$$A_1(O , O') = \left\{ \{x , y , z\} \subseteq X : O_{\{x,y,z\}} = O'_{\{x,y,z\}} \right\}$$
$$A_2(O , O') = A(O , O') - A_1(O , O')$$

pour $i \in \{1, 2\}$, $d_i(O , O') = |D_i(O , O')|$ et $a_i(O , O') = |A_i(O , O')|$.

On a donc :

$$d_1(O , O') + d_2(O , O') = e_G(O , O') \,;\, a_1(O , O') + a_2(O , O') = a_G(O , O')$$
$$d_1(O , O') + d_2(O , O') + a_1(O , O') + a_2(O , O') = \binom{n}{3} \tag{80}$$

On pose aussi :

$$S(O , O') = |O \cap O'| - n$$

et on a donc :

$$S(O , O') + d_K(O , O') = \binom{n}{2} \tag{81}$$

Lemme 1

$$3a_1(O , O') + 2d_2(O , O') + a_2(O , O') = (n - 2)\, S(O , O') \tag{82}$$

Cette égalité s'obtient en dénombrant de deux manières les paires $\}\{x , y\}, \{x , y, z\}\{$ où x et y sont deux éléments "comparables" pour l'ordre partiel $O \cap O'$ (i.e. tels que $x\, O\, y$ et $x\, O'\, y$, ou $y\, O\, x$ et $y\, O'\, x$).

Lemme 2

Soient $r(x)$ et $r'(x)$ les rangs de x dans les ordres totaux O et O'. On a :

$$\sum_{x \in X} r(x)\, r'(x) = n^2 + S(O , O') + d_1(O , O') + 2d_2(O , O') + a_2(O , O') + 2a_1(O , O') \tag{83}$$

Cette égalité fondamentale est démontrée dans la référence citée ci-dessus. Considérons alors le carré de la distance de SPEARMAN entre les deux ordres totaux O et O'. On a, en utilisant les lemmes 1 et 2 et la formule (80) :

$$\begin{aligned}
d_S^2(O\,,O') &= \sum_{x\in X} [r(x) - r'(x)]^2 = 2 \sum_{x\in X} r^2(x) - 2 \sum_{x\in X} r(x)\,r'(x) \\
&= \frac{n(n+1)(2n+1)}{3} - [2n^2 + 2S(O\,,O') + (n-2)\,S(O\,,O') + \binom{n}{3} \\
&\qquad + d_1(O\,,O') + d_2(O\,,O')] \\
&= \frac{n^2(n-1)}{2} - n\,S(O\,,O') - e_G(O\,,O') \\
&= n\left[\binom{n}{2} - S(O\,,O')\right] - e_G(O\,,O') \\
&= n\,d_K(O\,,O') - e_G(O\,,O') \quad \text{c.q.f.d.}
\end{aligned}$$

Le coefficient σ de GUILBAUD est défini en normant l'écart e_G de façon à obtenir l'intervalle de variation $[-1 + 1]$:

$$\sigma = 1 - \frac{2e_G}{\binom{n}{3}} = \frac{a_G - e_G}{\binom{n}{3}} \tag{84}$$

Les formules (35), (37) et (84) donnent

$$d_K = \binom{n}{2}\frac{1-\tau}{2} \quad d_S^2 = \frac{(n^3 - n)(1-\rho)}{6} \quad e_G = \binom{n}{3}\left(\frac{1-\sigma}{2}\right) \tag{85}$$

En reportant ces valeurs dans l'égalité du théorème, on obtient :

Corollaire 1 (GUILBAUD, 1980).

Pour tout $O\,,O' \in \mathcal{O}_X$

$$\tau(O\,,O') = \frac{2}{3}\left(1 + \frac{1}{n}\right)\rho(O\,,O') + \frac{1}{3}\left(1 - \frac{2}{n}\right)\sigma(O\,,O') \tag{86}$$

Si maintenant, on calcule σ en fonction de τ et ρ, on obtient

$$\sigma = \frac{3n}{n-2}\tau - \frac{2(n+1)}{n-2}\rho$$

D'où :

Corollaire 2 (DANIELS, 1944)

Pour tout $O\,,O' \in \mathcal{O}_X$

$$-1 \leqslant \frac{3n}{n-2}\tau(O\,,O') - \frac{2(n+1)}{n-2}\rho(O\,,O') \leqslant +1 \tag{87}$$

Noter que cette inégalité de DANIELS, apparaît sous une forme inexacte dans KENDALL (1970).

Définissons maintenant un coefficient G pour un profil π d'ordres totaux, en posant

$$G(\pi) = \sum_{i,j=1}^{v} \frac{\sigma(O_i , O_j)}{v^2} = 1 - \frac{2 \sum_{i,j=1}^{v} e_G(O_i , O_j)}{v^2 \binom{n}{3}} \tag{88}$$

Il résulte immédiatement des formules (55), (67) et du corollaire 1 :

Corollaire 3

Pour tout $\pi \in \mathcal{O}_X^v$

$$K(\pi) = \frac{2}{3}\left(1 + \frac{1}{n}\right) W(\pi) + \frac{1}{3}\left(1 - \frac{2}{n}\right) G(\pi) \tag{89}$$

Le coefficient G, qui s'obtient à partir de K et W, peut donc se calculer à partir des seules quantités w(x , y). On obtient

$$G(\pi) = \frac{n \sum_{\{x,y\} \subset X} w^2(x , y) - \sum_{x \in X} \left[\sum_{y \in X} w(x , y) \right]^2}{v^2 \binom{n}{3}} = \frac{n \sum_{\{x,y\} \subset X} w^2(x , y) - \sum_{x \in X} S^2(x)}{v^2 \binom{n}{3}} \tag{90}$$

Exemple 8

En reprenant le profil de cinq ordres traité aux exemples 3, 5, 6 et 7, on obtient les résultats suivants :

e_G	1	2	3	4	5
1	0	2	2	2	4
2	2	0	4	2	2
3	2	4	0	2	2
4	2	2	2	0	2
5	4	2	2	2	0

σ_G	1	2	3	4	5
1	1	0	0	0	−1
2	0	1	−1	0	0
3	0	−1	1	0	0
4	0	0	0	1	0
5	−1	0	0	0	1

$$G(\pi) = \Sigma \frac{\sigma_{ij}}{v^2} = \frac{1}{25}$$

$$G(\pi) = \frac{n \, \Sigma \, w^2(x , y) - \Sigma \, S^2(x)}{v^2 \binom{n}{3}} = \frac{4.30 - 116}{25.4} = \frac{1}{25}$$

On vérifie

$$K = \frac{2}{3}\,\frac{(n+1)}{n}\,W + \frac{1}{3}\,\frac{(n-2)}{n}\,G = \frac{10}{12}\cdot\frac{29}{125} + \frac{2}{12}\cdot\frac{1}{25} = \frac{1}{5}$$

Remarque

Soient O_1, O_2, O_3, O_4 quatre ordres totaux $\in \mathcal{O}_X$. Il résulte immédiatement de la définition de τ, la propriété suivante :

$$O_1 \cap O_2 \subseteq O_3 \cap O_4 \Rightarrow \tau(O_1, O_2) \leqslant \tau(O_3, O_4)$$

Le théorème de ce paragraphe permet de montrer que cette propriété souhaitable pour un coefficient d'accord entre deux ordres totaux, est également vérifiée pour le ρ de SPEARMAN :

$$O_1 \cap O_2 \subseteq O_3 \cap O_4 \Rightarrow \rho(O_1, O_2) \leqslant \rho(O_3, O_4)$$

Ce résultat n'est toutefois pas évident. On en trouvera la démonstration dans MONJARDET (1985), en même temps qu'une étude plus approfondie de la comparaison entre ρ et τ.

7. PROCEDURES D'AGREGATION ASSOCIEES AUX COEFFICIENTS K ET W

Le problème d'agrégation d'ordres totaux consiste à définir de "bonnes" procédures qui à tout profil π d'ordres totaux associe un – ou plusieurs – ordres totaux qu'on considérera comme les "bons" résumés du profil donné. Un tel problème se rencontre dans différents contextes : théorie du choix social, méthodes multicritères, analyse de préférences, etc. . . et il a fait l'objet d'une abondante littérature. Les conditions imposées sur les procédures peuvent être de type varié : axiomatique (c'est l'approche du théorème bien connu d'ARROW), constructive (par exemple, la règle majoritaire), par optimisation d'un certain critère. Une procédure du troisième type est associée canoniquement à chaque coefficient de concordance entre ordres totaux. Il suffit en effet de prendre comme ordre total résumé d'un profil π. un ordre total maximisant le coefficient de concordance du profil $\pi + 0$ (0 variant parmi tous les ordres totaux possibles). On pourra aussi déduire de la comparaison des coefficients de concordance, une mesure de la valeur du résumé. On va montrer que les procédures ainsi associées au choix des coefficients K (ou U) et W (ou W′) sont classiques. Pour cela, nous commençons par énoncer une proposition valable pour un nuage de points sphérique, dans un espace affine euclidien.

Les notations sont celles du paragraphe 2.4 ; B_r est une sphère de $\mathbf{R}^p$ (de centre l'origine et de rayon r) : $N = (x_1, \ldots x_i, \ldots, x_v)$ est une famille de v vecteurs de E ; g_N est le vecteur moyen de N ; x_0 est un vecteur quelconque de E ; $N + x_0$ est la famille $(x_1, \ldots x_i, \ldots x_v, x_0)$; H_1 et H'_1 sont les coefficients définis respectivement par les formules (23) et (16).

Proposition

Soient B_r sphère de $\mathbf{R}^p$, $N \in B_r^v$ et $x_0 \in B_r$. Les conditions (1) à (7) suivantes sont équivalentes.

(1) x_0 maximise $H_1(N + x_0)$, dans B_r.

(2) x_0 maximise $H'_1(N + x_0)$, dans B_r.

(3) x_0 maximise $\sum_{i=1}^{v} \cos(x_0 , x_i)$, dans B_r.

(4) x_0 maximise $\cos(x_0 , g_N)$, dans B_r.

(5) x_0 maximise $\langle x_0 , g_N \rangle$, dans B_r.

(6) x_0 minimise $\sum_{i=1}^{v} d^2(x_0 , x_i)$, dans B_r.

(7) x_0 minimise $d^2(x_0 , g_N)$, dans B_r.

De plus, pour x_0 vérifiant ces conditions on a,

$$H_1(N + x_0) \geqslant H_1(N) + \frac{in(N)}{v(v + 1)^2 r^2}$$

Démonstration

(1) est équivalent à (2), puisque H'_1 est une transformation affine de H_1.
(1) est équivalent à (3), puisque en utilisant la formule (24), on obtient

$$H_1(N + x_0) = \frac{\sum_{(i,j)\in V^2} \cos(x_i , x_j) + 2 \sum_{i\in V} \cos(x_0 , x_i) + 1}{(v + 1)^2} .$$

L'équivalence de (3), (4) et (5) résulte des égalités

$$\Sigma \cos(x_0 , x_i) = \Sigma \frac{\langle x_0 , x_i \rangle}{r^2} = \frac{v}{r^2} \langle x_0 , g_N \rangle = \frac{v}{r \| g_N \|} \cos(x_0 , g_N) .$$

L'équivalence de (3) et (6) provient de l'égalité $\cos(x_0 , x_i) = 1 - \dfrac{d^2(x_0 , x_i)}{2r^2}$, et l'équivalence entre (6) et (7) de la formule de Huyghens.

Un calcul simple montre que pour $x \in B_r$, on a

$$H_1(N + x) = H_1(N) + \frac{(2v + 1)\, in(N) - v \sum_{i=1}^{v} d^2(x , x_i)}{v(v + 1)^2 r^2}$$

Pour x_0 vérifiant (6), on a $\sum_{i=1}^{v} d^2(x_0 , x_i) \leqslant \sum_{i=1}^{v} d^2(x_j , x_i)$ pour tout $j = 1 , . . v$.

D'où $v \sum_{i=1}^{v} d^2(x_0 , x_i) \leqslant 2v\, in(N)$, et le résultat.

Soit maintenant $\mathcal{O}$ l'ensemble des ordres totaux définis sur X. On va appliquer la proposition précédente d'abord dans le cas ou $B_r = \gamma(\mathcal{O}) \subset \mathbf{R}^{n^2}$ (γ codage de KENDALL des ordres totaux), puis dans celui où $B_r = s(\mathcal{O}) \subset \mathbf{R}^n$ (s codage de SPEARMAN des ordres totaux).

Corollaire 1

Soient $\pi = (O_1, \dots O_i, \dots O_v)$ un profil d'ordres totaux définis sur X, O un ordre total défini sur X. Les propriétés suivantes sont équivalentes :

(1) O maximise dans $\mathcal{O}$, $K(\pi + O)$.

(2) O maximise dans $\mathcal{O}$, $U(\pi + O)$.

(3) O maximise dans $\mathcal{O}$, $\sum_{i=1}^{v} \tau(O, O_i)$.

(4) O maximise dans $\mathcal{O}$, $\sum_{(x,y)\in O} w(x, y)$ (ou $\sum_{(x,y)\in O} v(x, y)$)

(5) O minimise dans $\mathcal{O}$, $\sum_{(x,y)\in O} v(y, x)$.

(6) O minimise dans $\mathcal{O}$, $\sum_{i=1}^{v} d_K(O, O_i)$.

(7) O minimise dans $\mathcal{O}$, $\sum_{(x,y)\in X^2} [w(x, y) - v\,\gamma_0(x, y)]^2$.

De plus pour O, vérifiant ces propriétés, on a $K(\pi + O) \geqslant K(\pi) + \dfrac{1 - K(\pi)}{(v+1)^2}$.

La formule (6), définit la procédure d'agrégation de l'ordre médian, proposée initialement par KEMENY (1959), redécouverte maintes fois, et étudiée par de nombreux auteurs (cf. à ce propos, BARTHELEMY et MONJARDET, 1981). On se contentera de noter ici que l'ordre médian, qui n'est pas nécessairement unique, coïncide d'après (4) avec la relation "majoritaire" (ou relation de Condorcet, définie en prenant pour chaque paire, le couple majoritaire) si celle-ci est un ordre total. Signalons aussi que cette procédure de l'ordre médian a été axiomatisée par YOUNG et LEVENGLICK (1978), et que des progrès récents dûs à MARCOTORCHINO et MICHAUD (1979) d'une part, et à ARDITTI (1983) d'autre part, permettent de calculer exactement de tels ordres (pour $|X|$ voisin de 100) ou d'avoir de bonnes heuristiques.

Corollaire 2

Soient $\pi = (O_1, \dots O_i, \dots O_v)$ un profil d'ordres totaux définis sur X et O un ordre total défini sur X (s_i, s, r_i, r, les vecteurs scores ou rangs correspondants). Les propriétés suivantes sont équivalents :

(1) O maximise dans $\mathcal{O}$, $W(\pi + O)$.

(2) O maximise dans $\mathcal{O}$, $W'(\pi + O)$.

(3) O maximise dans $\mathcal{O}$, $\sum_{i=1}^{v} \rho(O, O_i)$.

(4) O maximise dans $\mathcal{O}$, $\sum_{x\in X} S(x)\, s(x)$.

(4′) O maximise dans $\mathcal{O}$, $\sum_{x \in X} R(x)\, r(x)$.

(5) O minimise dans $\mathcal{O}$, $\sum_{x \in X} [S(x) - vs(x)]^2$.

(5′) O minimise dans $\mathcal{O}$, $\sum_{x \in X} [R(x) - vr(x)]^2$

(6) O minimise dans $\mathcal{O}$, $\sum_{i=1}^{v} d^2(s, s_i)$.

(6′) O minimise dans $\mathcal{O}$, $\sum_{i=1}^{v} d^2(r, r_i)$.

(7) O minimise dans $\mathcal{O}$, $\sum_{i=1}^{v} d_S^2(O, O_i)$.

(8) O est un ordre total vérifiant la condition : pour tout x , y

$$x\, O\, y \Rightarrow S(x) \geqslant S(y)\ (R(x) \leqslant R(y)).$$

De plus pour un tel ordre total, on a $W(\pi + O) \geqslant W(\pi) + \dfrac{1 - W(\pi)}{(v+1)^2}$.

La formule (8) montre que la procédure d'agrégation ainsi obtenue est dérivée de celle dite de la somme des rangs, qui remonte (au moins) à BORDA. De façon précise, la méthode de BORDA consiste à classer les objets au moyen de la somme des rangs (ou des scores) qu'ils ont obtenu dans les différents ordres de préférences. On définit ainsi une relation de préordre total ; les ordres totaux inclus dans ce préordre (i.e. ceux obtenus en ordonnant arbitrairement les ex-aequo) sont les solutions du problème d'optimisation (1). On a donc là une procédure particulièrement simple à mettre en œuvre ; elle a été proposée par KENDALL (1942) qui a montré l'équivalence des conditions (3), (5′) et (8) ; pour cette raison elle est parfois appelée méthode de BORDA-KENDALL.

Ajoutons deux remarques. La méthode de BORDA donne en général un préordre total ; mais il faut se garder de croire que cette procédure conduit nécessairement au préordre total "optimal" au sens de celui qui minimiserait parmi tous les préodres totaux possibles, l'équivalent de l'expression de (6) ; la solution de ce dernier problème d'optimisation se trouve dans LEMAIRE (1977) et COOK et SEIFORD (1982).

D'autre part, il est bien connu (depuis CONDORCET. . . !) que la méthode d'agrégation d'ordres totaux en un ordre total définie au corollaire 2 ne conduit pas en général au même résultat que la méthode de l'ordre médian, même si l'ordre médian est la relation majoritaire, ou si n est de "taille modérée" (contrairement à ce qu'écrit KENDALL, dans Rank correlation methods). Le problème des conditions de concordance entre ces deux méthodes n'a actuellement que des résultats très partiels (cf. notamment, FISHBURN, 1974), et est un des aspects des relations entre distance de KENDALL et distance de SPEARMAN.

Exemple 9

. Soit le profil des cinq ordres totaux de l'exemple 5 : π = (abcd, bcad, cbda, bacd, dcba) ; d'après le tableau des scores, on a S(a) = − 3, S(b) = 7, S(c) = 3, S(d) = − 7 ; la méthode de BORDA-KENDALL donne donc comme ordre consensus, l'ordre bcad ; cet ordre étant aussi obtenu par la procédure majoritaire, c'est l'unique ordre médian du profil π.

. Par contre, pour le profil (acdb, cdba, bcad, bcda, adbc) l'ordre médian est bcad alors que l'ordre de BORDA-KENDALL est cbad.

8. REFERENCES

O. ARDITTI. – "Un nouvel algorithme de recherche d'un ordre induit par des comparaisons par paires", in *Data Analysis and Informatics III*, E. Diday et autres edit., North-Holland, 1984 (323-343).

M. BARBUT. – Médianes, Condorcet et Kendall, Note SEMA, Paris, 1967, et *Math. Sci. hum.* 69 (1980) 5-13.

M. BARBUT et B. MONJARDET. – *Ordre et classification, Algèbre et Combinatoire,* 2 tomes, Hachette, Paris, 1970.

J.P. BARTHELEMY. – "Caractérisations axiomatiques de la distance de la différence symétrique entre des relations binaires", *Math. Sci. Hum.* 67 (1979) 85-114.

J.P. BARTHELEMY, Cl. FLAMENT et B. MONJARDET. – "Ordered sets and social sciences", in *Ordered sets,* I. Rival edit., D. Reidel, 1982 (721-758).

J.P. BARTHELEMY et B. MONJARDET. – "The median procedure in cluster analysis and social choice theory", *Math. Soc. Sci.,* 1, 3 (1981) 235-267.

P. BATTEAU, E. JACQUET-LAGREZE et B. MONJARDET. – (edit.), *Analyse et Agrégation des Préférences,* Economica, Paris, 1981.

J.P. BENZECRI. – "Sur l'analyse des préférences", polycopié (1965) et in *Ordres totaux finis,* Gauthier-Villars, Paris, 1971 (195-205).

F. CAILLIEZ et J.P. PAGES. – *Introduction à l'analyse des données,* SMASH, Paris, 1976.

W.D. COOK et L.M. SEIFORD. – "On the Borda-Kendall consensus method for priority ranking problems", *Manag. Sci., 28,* 6 (1982) 621-627.

H.E. DANIELS. – "The relation between measures of correlation in the universe of sample permutations", *Biometrika,* 33 (1944) 129-135.

A. DEGENNE. – *Techniques ordinales en analyse des données : Statistiques,* Hachette, Paris, 1972.

A.S.C. EHRENBERG. – "On sampling from a population of rankers", *Biometrika,* 39, (1952) 82-87.

P.C. FISHBURN. – "Voter concordance, simple majorities and group decision methods", *Behav. Sci.,* 18 (197) 364-376.

G. Th. GUILBAUD. – "Relation entre les deux coefficients de correlation de rang", *Math. Sci. hum.,* 72 (1980) 45-59.

W.L. HAYS. – "A note on average τ as a measure of concordance", *J. Amer, Statist. Ass.,* 55, 290 (1960) 331-341.

M. HOLLANDER et D.A. WOLFE. – *Non parametric statistical methods,* Wiley, New-York, 1973.

L.J. HUBERT. – "Generalized concordance" *Psychometrika,* 44,2 (1979) 135-142.

E. JACQUET-LAGREZE. – "Analyse des préférences", in *Analyse des données en marketing,* J.M. Bouroche edt., Masson, Paris (1977) 63-97.

J.G. KEMENY. – "Mathematics without numbers", *Daedalus,* 88 (1959) 577-591.

M.G. KENDALL. – "Note on the estimation of a ranking", *J. Roy. Statist. Soc.,* 105 (1942) 119.

M.G. KENDALL. – *Rank correlation methods,* 4e ed., Hafner, New-York, 1970. (1e ed. 1948, 3e éd. 1962).

W.H. KRUSKAL. – "Ordinal measures of association", *J. Am. Statist. Ass.,* 53, 284 (1958) 814-861.

J. LEMAIRE. – "Agregation typologique de données de préférences", *Math. Sci. hum.* 58 (1977) 31-50.

I.C. LERMAN. – *Classification et analyse ordinales des données,* Dunod, Paris, 1981.

J.F. MARCOTORCHINO et P. MICHAUD. – *Optimisation en analyse des données,* Masson, Paris, 1979.

B. MONJARDET. – "Tournois et ordres médians pour une opinion", *Math. Sci. hum.* 43 (1973) 55-70.

B. MONJARDET. – Concordance et consensus d'ordres totaux : K et W (Kendall-Condorcet et Spearman-Borda), *rapport de recherche* P. 008 du *Centre de Mathématique Sociale,* Paris, 1984, (45 pages).

B. MONJARDET. – On an ordinal property of rho and its comparison with tau, *rapport de recherche du Centre de Mathématique Sociale,* Paris, 1985.

P.A.P. MORAN, J.W. WHITFIELD et H.E. DANIELS. – "Symposium on ranking methods, *J. Roy. Statist. Soc.,* B, 12, (1950), 153-191.

ORDRES TOTAUX FINIS. – Travaux du séminaire d'Aix-en-Provence sur les ordres totaux finis, Gauthiers-Villars, Paris, 1971.

R. SIBSON. – "Order invariant methods for data analysis", *J. Roy. Statist. Soc.* B, 34, (1972), 311-349.

C. SPEARMAN. – "The proof and measurement of association between two things", *Am. J. Psychol.,* 15, (1904), 72-101.

H.P. YOUNG et A. LEVENGLICK. – "A consistent extension of Condorcet's election principle", *SIAM J. Appl. Math.,* 25, 2, (1978), 285-300.

FORMULAIRE I

COEFFICIENTS	FORMULES	VARIATION
$d_K(O,O')$	$\lvert O-O'\rvert$ $\quad \dfrac{\lvert O\cup O'\rvert-\lvert O\cap O'\rvert}{2}$ $\quad \dfrac{d_\Delta(O,O')}{2}$	0 à $\dfrac{n(n-1)}{2}$
$\tau(O,O')$	$1-\dfrac{4d_K(O,O')}{n(n-1)}$ $\quad \dfrac{4\,S(O,O')}{n(n-1)}-1$ $\quad \dfrac{2[S(O,O')-d_K(O,O')]}{n(n-1)}$	-1 à $+1$
$K(\Pi)$	$\dfrac{\sum\limits_{(i,j)}\tau_{ij}}{v^2}$ $\quad 1-\dfrac{4\sum\limits_{(i,j)}d_{K,ij}}{v^2n(n-1)}$ $\dfrac{\sum\limits_{(x,y)}w^2_{xy}}{v^2n(n-1)}$ $\quad \dfrac{4\sum\limits_{x\neq y}v^2_{xy}}{v^2n(n-1)}-1$	0 à 1 si v pair $\dfrac{1}{v^2}$ à 1 si v impair
$U(\Pi)$	$\dfrac{\sum\limits_{i\neq j}\tau_{ij}}{v(v-1)}$ $\quad \dfrac{4\sum\limits_{x\neq y}v_{xy}(v_{xy}-1)}{v(v-1)n(n-1)}$ $\quad 1-\dfrac{4\sum\limits_{x\neq y}v_{xy}v_{yx}}{v(v-1)n(n-1)}$	$-\dfrac{1}{v-1}$ à 0 si v pair $-\dfrac{1}{v}$ à 0 si v impair
$H_2(\Pi_\gamma)$	K si v pair $\quad 1-\dfrac{4\sum\limits_{x\neq y}v_{xy}v_{yx}}{(v^2-1)n(n-1)}$ si v impair	0 à 1
$d^2_S(O,O')$	$\sum\limits_x[r(x)-r'(x)]^2$ $\quad \dfrac{1}{4}\sum\limits_x[s(x)-s'(x)]^2$	0 à $\dfrac{n^3-n}{3}$
$\rho(O,O')$	$1-\dfrac{6\,d^2_S(O,O')}{n^3-n}$	-1 à $+1$
$W(\Pi)$	$\dfrac{\sum\limits_{(i,j)}\rho_{ij}}{v^2}$ $\quad 1-\dfrac{6\sum\limits_{(i,j)}d^2_{S,ij}}{v^2(n^3-n)}$ $\dfrac{3\sum\limits_x S^2(x)}{v^2(n^3-n)}$ $\quad \dfrac{12\sum\limits_x R^2(x)}{v^2(n^3-n)}-\dfrac{3(n+1)}{n-1}$	0 à 1 si v pair ou n impair $\dfrac{3}{v^2(n^2-1)}$ à 1 si v impair et n pair
$W'(\Pi)$	$\dfrac{\sum\limits_{i\neq j}\rho_{ij}}{v(v-1)}$ $\quad \dfrac{3\sum\limits_x S^2(x)}{v(v-1)(n^3-n)}-\dfrac{1}{v-1}$ $\quad \dfrac{12\sum\limits_x R^2(x)}{v(v-1)(n^3-n)}-\dfrac{v(4n+2)}{(v-1)(n-1)}+1$	$-\dfrac{1}{v-1}$ à 1 si v pair ou n impair
$H_2(\Pi_S)$	W si v pair ou n impair $\quad \dfrac{v^2(n^2-1)W-3}{v^2(n^2-1)-3}$ sinon	0 à 1

$$S(O,O') = \lvert O\cap O'\rvert - n$$

$$\tau_{ij} = \tau(O_i,O_j) \qquad d_{K,ij} = d_K(O_i,O_j) \qquad d^2_{S,ij} = d^2_S(O_i,O_j)$$

$$v_{xy} = \lvert\{i \in V : x\ O_i\ y\}\rvert \qquad w_{xy} = v_{xy} - v_{yx}$$

$$r(x) = \lvert\{y \in X : y\ O\ x\}\rvert \qquad s(x) = n+1-2r(x)$$

$$\sum_{(i,j)} = \sum_{(i,j)\in V^2} \qquad \sum_{i\neq j} = \sum_{(i\neq j)\in V^2} \qquad \sum_{(x,y)} = \sum_{(x,y)\in X^2} \qquad \sum_x = \sum_{x\in X}$$

II DISTANCES ET COEFFICENTS ASSOCIES AUX CODAGES SPHERIQUES D'ORDRES TOTAUX

	$\mathbb{R}^{n^2}$			$\mathbb{R}^n$	
codage α	δ caractéristique	γ caractéristique centré	ε caractéristique pondéré	r rang	s score
$\|\alpha\|^2$	$\frac{n(n+1)}{2}$	$n(n-1)$	$\frac{n^2(n^2-1)}{6}$	$\frac{n(n+1)(2n+1)}{6}$	$\frac{n^3-n}{3}$
$d^2(\alpha,\alpha')$	$2d_K$	$8d_K$	$2n\ d_S^2$	d_S^2	$4d_S^2$
$\cos(\alpha,\alpha')$	$\frac{(n-1)\tau+(n+3)}{2(n+1)}$	τ	ρ	$\frac{(n-1)\rho+3(n+1)}{2(2n+1)}$	ρ
$\mathrm{corr}(\alpha,\alpha')$	$\frac{n\tau+1}{n+1}$	τ	ρ	ρ	ρ
$H_1(\Pi_\alpha)$	$\frac{(n-1)K+(n+3)}{2(n+1)}$	K	W	$\frac{(n-1)W+3(n+1)}{2(2n+1)}$	W
$H_1'(\Pi_\alpha)$	$\frac{(n-1)U+(n+3)}{2(n+1)}$	$U=\frac{vK-1}{v-1}$	$W'=\frac{vW-1}{v-1}$	$\frac{(n-1)W'+3(n+1)}{2(2n+1)}$	$W'=\frac{vW-1}{v-1}$
$H_2(\Pi_\alpha)$	K si v pair $\frac{v^2K-1}{v^2-1}$ si v impair	K si v pair $\frac{v^2K-1}{v^2-1}$ si v impair		W si v pair ou n impair $\frac{v^2(n^2-1)W-3}{v^2(n^2-1)-3}$ sinon	W si v pair ou n impair $\frac{v^2(n^2-1)W-3}{v^2(n^2-1)-3}$ sinon

CODAGE α, DANS LA SPHERE $B(0,r)$

$\|\alpha\|$ = norme du vecteur $\alpha = r$; $d(\alpha,\alpha')$ = distance euclidienne entre α et α'

$\mathrm{corr}(\alpha,\alpha')$ = coefficient de corrélation linéaire entre α et α'

Π = profil de V ordres totaux, sur un ensemble à n éléments

Π_α = famille des V vecteurs α, codant les V ordres totaux de Π

$$g_{\Pi_\alpha} = \frac{1}{v}\sum_{i=1}^{v} \alpha_i \qquad \mathrm{in}\ (\Pi_\alpha) = \sum_{(i,j)} d^2(\alpha_i, g_{\Pi_\alpha})$$

$$H_1(\Pi_\alpha) = 1 - \frac{\mathrm{in}\ (\Pi_\alpha)}{v\ r^2} = \frac{\|g_{\Pi_\alpha}\|^2}{r^2}$$

$$= \sum_{(i,j)} \frac{\cos(\alpha_i,\alpha_j)}{v^2} = 1 - \sum_{(i,j)} \frac{d^2(\alpha_i,\alpha_j)}{2v^2\ r^2}$$

$$H_1'(\Pi_\alpha) = \sum_{\{i,j\}} \frac{\cos(\alpha_i,\alpha_j)}{\binom{v}{2}} = \frac{vH_1(\Pi_\alpha) - 1}{v-1}$$

$$H_2(\Pi_\alpha) = 1 - \frac{\mathrm{in}(\Pi_\alpha)}{\max\ \mathrm{in}\ (\Pi_\alpha)}$$