MATHÉMATIQUES ET SCIENCES HUMAINES

G. TH. GUILBAUD

Relation entre les deux coefficients de corrélation de rangs

Mathématiques et sciences humaines, tome 72 (1980), p. 45-59

<http://www.numdam.org/item?id=MSH_1980__72__45_0>

Math. Sci. hum. (*18*^{e} année, n°72, *1980*, p.45-59)

RELATION ENTRE LES DEUX COEFFICIENTS DE CORRELATION DE RANGS

G.Th. GUILBAUD*

Première partie : Introduction au problème, au moyen d'un petit exemple numérique

1. Pour mesurer (si l'on ose dire) le degré de dérangement opéré par une permutation, on dispose de deux coefficients, dits de "corrélation de rangs" : le Spearman, le plus souvent nommé Rho (ρ), et le Kendall nommé Tau (τ). Leur calcul est enseigné par tous les manuels de statistique, même élémentaires.

Y a-t-il une relation, et de quelle nature, entre ces deux indicateurs ? On va montrer ici qu'il est avantageux d'introduire un troisième coefficient ; et, comme il fait le lien entre ρ et τ, je propose de le nommer Sigma (σ).

On a :

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

n étant le nombre d'éléments de la permutation. C'est dire que Tau est égal à une moyenne (pondérée) entre Rho et Sigma. Et lorsque *n* est grand, la pondération est voisine de : deux tiers de Rho, un tiers de Sigma.

2. Commençons par un petit exemple (n = 8).

Prenons un mot de huit lettres (toutes différentes), soit :

(m) C H A M B E R Y .

Si on les range dans l'ordre alphabétique :

(a) A B C E H M R Y .

Calculons les deux coefficients pour (a) et (m).

---

* Ecole des Hautes Etudes en Sciences Sociales, Paris.

Le Tau : on considère successivement les vingt-huit paires de lettres :

A et B : dans le même ordre en (a) et (m)

A et C : en ordres contraires

A et E : même ordre

etc.

Il y a sept désaccords (A et C, A et H, B et C, B et H, B et M, E et H, E et M) ou, si on préfère : le mot CHAMBERY présente *sept* renversements de l'ordre alphabétique.

Il y a donc vingt et un accords.

Et l'on pose :

$$\text{tau} = \frac{21 - 7}{21 + 7} = \frac{1}{2} .$$

(formule : Accords - Désaccords / Accords + Désaccords)

3. Le Rho : on numérote les places occupées par les lettres ; on obtient ainsi deux vecteurs, dont on calcule le produit scalaire.

Pour numéroter, on a le choix entre diverses procédures. Rien n'empêche, bien sûr, de prendre la plus banale :

C	H	A	M	B	E	R	Y
1	2	3	4	5	6	7	8

Mais il est avantageux de s'arranger pour que la somme des indicateurs de rangs soit nulle. On prendra, par exemple, comme échelle des rangs, la suite :

-7, -5, -3, -1, +1, +3, +5, +7

dont la signification intuitive est commode :

la lettre E est marquée +3 parce qu'elle a 5 lettres avant et 2 après, or 5-2=3.

L'indicateur de rang est donc ici : différences entre le nombre de lettres *avant* et le nombre de lettres *après*. Pour H : 1 avant et 6 après, cela donne bien 1-6 =-5, etc.

On constitue alors les deux vecteurs :

	Rangs en (m)	Rangs en (a)	Produits
A	-3	-7	+21
B	+1	-5	-5
C	-7	-3	+21
E	+3	-1	-3
H	-5	+1	-5
M	-1	+3	-3
R	+5	+5	+25
Y	+7	+7	+49
Total	0	0	+100

La somme des carrés (carré de la longueur) est la même pour les deux :

2 fois (1 + 9 + 25 + 49) = 168.

Le produit scalaire (somme des produits) = 100.

On prendra :

RhO = Somme des produits / Somme des carrés

soit :

Rho = 100/168 = 25/42.

4. Nous allons conserver l'ordre alphabétique (a) comme référence, mais modifier l'ordre (m), de façon à examiner les variations concomitantes de Rho et de Tau.

Pour commencer, je propose d'effectuer seulement des

permutations circulaires,

c'est-à-dire :

m0 = CHAMBERY
m1 = HAMBERYC
m2 = AMBERYCH
m3 = MBERYCHA
m4 = BERYCHAM
etc.

Calculons les deux coefficients, comme ci-dessus :

	Rho	Tau
(a) et (m0)	+25/42	+1/2
(m1)	+13/42	+2/7
(m2)	+17/42	+5/14
(m3)	-11/42	-1/7
(m4)	+ 1/42	+1/14
(m5)	-19/42	-2/7
(m6)	-23/42	-5/14
(m7)	- 3/42	0

Pour faciliter le comparaison, réduisons au même dénominateur, et rangeons selon l'ordre de grandeur :

Rho	+25	+17	+13	+1	-3	-11	-19	-23
Tau	+21	+15	+12	+3	0	- 6	-12	-15
	m0	m2	m1	m4	m7	m3	m5	m6

Est-il bien nécessaire de porter ces résultats sur un graphique pour apercevoir la liaison linéaire ? Quand Rho diminue de 4, de 8 ou de 12, Tau diminue de 3, de 6 ou de 9.

On constate que :

Tau = (3/4) Rho + (3/56).

En posant :

Sigma = 3/14, on obtiendra :

$$\tau = (3\rho + \sigma) / 4$$

(Le choix de Sigma est tel que Tau est une *moyenne* entre Rho et Sigma).

5. Anticipons (les démonstrations vont suivre) et donnons un mode de calcul pour l'indicateur Sigma.

Il est commode, parmi diverses présentations possibles, de définir ce coefficient à partir d'une comparaison, non plus des *paires* (comme pour Tau), mais des *trios*.

De plus, il faut entendre ici "accord" et "désaccord" au sens "circulaire".
Ce qui sera plus clair sur un exemple :

Pour le trio : (A,B,C), on a

l'ordre C..A..B... dans (m0)
l'ordre ..B..C..A dans (m3)
l'ordre ABC... dans (a)

nous dirons qu'il y a accord pour ce trio, parce que c'est partout le même ordre *circulaire* à savoir : ABCABCA...

Par contre pour le trio (A,B,M) on a :
dans (m0) : AMB, dans (m3) : MBA, mais dans (a) : ABM ;
il y a désaccord entre (a) et tous les (m).

Evidemment, les huit permutations (m) sont toujours d'accord (au point de vue circulaire).

On comptera les A accords et les D désaccords ; le nombre de trios est A+D.
Et l'on définira Sigma par

Sigma = A - D / A + D .

Pour notre exemple, on trouve entre (a) et n'importe quel (m) A = 34 et D = 22.
Ce qui donne $\sigma = 12/56$ ou 3/14 .

On peut trouver d'autres algorithmes, équivalents au précédent.
Signalons seulement celui qui à chaque paire attribue une "distance circulaire" pour une permutation donnée ; et, pour deux permutations, calcule Sigma comme un coefficient de corrélation (à la manière de Rho).

Reprenons notre exemple ; si nous voulons considérer (m) comme un ordre circulaire, il suffira d'écrire :

(m) ...CHAMBERYCHAMBERYCHA... ,

Pour décrire la situation respective des deux lettres A et B, on comptera le nombre de lettres intermédiaires :

...A (1) B (5) A (1) B ...

on attribuera au couple AB la distance

$$1-5 = -4$$

(pour BA ce sera +4)

Le même couple, dans l'ordre alphabétique

(a) ...ABCEHMRYABCEHMRYAB...

donne :

...A B (6) A B (6) A ...

d'où, pour distance :

$$0-6 = -6$$

On fera ce calcul pour tous les couples (ou pour la moitié d'entre eux).

couples	distances en (m)	distances en (a)	produits
AB	-4	-6	+24
CH	-6	-4	+24
AR	0	+4	0
HY	+4	-2	-8
...	...	...	...
		Total	+96

La somme des carrés est la même pour les deux colonnes, soit :

$$8 \text{ fois } (2^2+4^2+6^2) = 448.$$

On prendra la formule de la corrélation :

Somme des produits / Somme des carrés.

Soit, ici,

Sigma = 96/448 = 3/14 .

Deuxième partie : Kendall est à Spearman ce que Condorcet est à Borda

On pourrait établir la relation annoncée ci-dessus (une fois qu'elle a été énoncée et présumée vraie) par des calculs directs et pas trop compliqués.

Mais ce n'est pas ainsi que je l'ai trouvée, et je préfère donner ici tous les détours qui m'y ont conduit : il s'agit de l'analyse spectrale de quelques matrices que je crois dignes d'intérêt. Ces détours auront aussi le mérite de montrer l'indépendance (au sens probabiliste) de Rho et de Sigma ; en effet, la somme des produits ($\rho \times \sigma$), pour les (n!) permutations, est nulle.

1. Depuis Condorcet, il est admis que, pour décrire l'état des préférences individuelles dans une collectivité, il est commode de dresser un tableau tel que celui-ci :

		A>B	A>C	A>D	B>C	B>D	C>D
Pierre pense que	A>B>C>D	oui	oui	oui	oui	oui	oui
Jacques :	B>C>A>D	non	non	oui	oui	oui	oui
André :	A>C>B>D	oui	oui	oui	non	oui	oui
Bernard :	C>D>B>A	non	non	non	non	non	oui
	&c						

Pour *dépouiller* un vote, à la manière de Condorcet, on comptera, dans chaque colonne les ouis et les nons, et on dira où va la *majorité*. Pour mécaniser l'affaire, il convient de *coder* : le codage que je considère comme le plus idoine remplace "OUI" par "+1", "NON" par "-1" ; il suffit alors de faire l'addition : si la somme est positive on dira que la majorité a voté "oui" ; "non" si la somme est négative. (Si vous préférez le codage par 0 et 1, libre à vous - tout ce qui va suivre reste vrai - mais il faut quelques modifications de pure forme).

2. Disons la même chose en langage matriciel.

J'appellerai matrice de CONDORCET (ou, pour aller vite : "CONDOR") ceci :

$$\begin{array}{c|ccc|}
\cline{2-4}
ABC & +1 & +1 & +1 \\
ACB & +1 & +1 & -1 \\
BAC & -1 & +1 & +1 \\
BCA & -1 & -1 & +1 \\
CAB & +1 & -1 & -1 \\
CBA & -1 & -1 & -1 \\
\cline{2-4}
\multicolumn{1}{c}{} & \multicolumn{1}{c}{AB} & \multicolumn{1}{c}{AC} & \multicolumn{1}{c}{BC}
\end{array} = \text{(CONDOR, pour } n=3\text{)}$$

et les analogues avec n! lignes et n(n-1)/2 colonnes.

(variante : rien n'empêcherait d'écrire les n(n-1) couples - mais comme ils sont deux à deux opposés, je trouve que ça ne vaut pas la peine - si on le fait cependant, quelques-unes des formules qui suivent devront être corrigées, par un facteur 2, çà et là).

L'état descriptif de l'opinion sera une liste de nombres, un vecteur, indiquant combien de citoyens ont choisi chacune des n! opinions possibles. Alors le dépouillement consiste à faire le produit du vecteur "opinion" par la matrice "Condor".

Je reprendrai ici les deux exemples choisis en 1785 par Condorcet lui-même (Cf. G.Th. Guilbaud, *Eléments de la Théorie des jeux*, Paris, Dunod, 1968, pages 48 et 49) :

3. Un exemple numérique :

Opinions I	Opinions II	Condor		
0	23	+1	+1	+1
23	0	+1	+1	-1
0	2	-1	+1	+1
19	17	-1	-1	+1
2	10	+1	-1	-1
16	8	-1	-1	-1

-10	-14	+22	Dépouillement I (à la Condorcet)
+6	-10	+24	Dépouillement II (à la Condorcet)

soit, en notation classique :

Opinion × Condor = Dépouillement.

vecteur ligne × matrice = vecteur ligne.

4. Il existe une autre façon de dépouiller : celle de BORDA. On analyse non plus selon les couples, mais selon les individus et leur rang. Quel codage ?

On pourrait évidemment choisir le numéro d'ordre

B C A D ⟶ B = 1, C = 2, A = 3, D = 4

mais ici encore je préfère un codage *centré*, les numéros étant :

soit ..., 4, 2, 0, -2, -4, ...

soit ..., 5, 3, 1, -1, -3, ...

selon que n est pair ou impair.

En d'autres termes je prends pour indicatif du rang d'un objet dans une suite, la différence entre le nombre des objets *après* et le nombre des objets *avant*.

Exemple : B F C A E D rang de C = 3 - 2 = 1, rang de E = 1 - 4 = -3,
rang de F = 4 - 1 = 3, etc.

Mais si quelqu'un préfère les numéros habituels, libre à lui : les calculs qui vont suivre devront être un peu modifiés en conséquence, ce n'est pas très difficile.

5. Voici donc les matrices que j'appellerai "BORDA" :

	A	B	C
ABC	+2	0	-2
ACB	+2	-2	0
BAC	0	+2	-2
BCA	-2	+2	0
CAB	0	-2	+2
CBA	-2	0	+2

pour n = 3

	A	B	C	D
ABCD	+3	+1	-1	-3
ABDC	+3	+1	-3	-1
ACBD	+3	-1	+1	-3
ACDB	+3	-3	+1	-1
ADBC	+3	-1	-3	+1
ADCB	+3	-3	-1	+1
BACD	+1	+3	-1	-3
....	..	..	..	..
DCAB	-3	-3	+1	+3
DCBA	-3	-1	+1	+3

pour n = 4

(sommes nulles en lignes comme en colonnes)

Et pour l'exemple numérique ci-dessus :

Opinion I	Opinion II
0	23
23	0
0	2
19	17
2	10
16	8

-24	-12	-36	Dépouillement de I (à la Borda)
- 4	+17	-14	Dépouillement de II (à la Borda)

6. La matrice "CONDOR" décrit, au moyen d'un simple codage, la relation entre l'ensemble des (n!) rangements et l'ensemble des n(n-1) couples (la moitié des couples suffit). La matrice "BORDA" décrit d'une façon analogue la relation entre l'ensemble des (n!) rangements et l'ensemble des n individus.

Cela invite (même le plus lent des cerveaux y penserait) à écrire une troisième matrice qui décrira la relation entre les couples et les individus. Il n'est pas utile d'hésiter longtemps sur le codage à choisir. Je propose :

$$\begin{array}{c|ccc|}
\mathrm{AB} & +1 & -1 & 0 \\
\mathrm{AC} & +1 & 0 & -1 \\
\mathrm{BC} & 0 & +1 & -1 \\
\hline
 & \mathrm{A} & \mathrm{B} & \mathrm{C}
\end{array}$$

pour n = 3

$$\begin{array}{c|cccc|}
\mathrm{AB} & +1 & -1 & 0 & 0 \\
\mathrm{AC} & +1 & 0 & -1 & 0 \\
\mathrm{AD} & +1 & 0 & 0 & -1 \\
\mathrm{BC} & 0 & +1 & -1 & 0 \\
\mathrm{BD} & 0 & +1 & 0 & -1 \\
\mathrm{CD} & 0 & 0 & +1 & -1 \\
\hline
 & \mathrm{A} & \mathrm{B} & \mathrm{C} & \mathrm{D}
\end{array}$$

pour n = 4

Comme d'habitude et par économie je n'écris que la moitié des couples.
De telles matrices sont célèbres, sous le nom de matrices d'incidence.
Si j'en crois Oswald Veblen (2e ed. de son *Analysis Situs*, pp. 26-27), avant leur exploitation systématique par Henri Poincaré, elles auraient été introduites par G. Kirchhoff dans son étude sur les réseaux électriques (1847).
Alors pourquoi ne pas les appeler des "KIRCHHOFF" ?

7. C'est un jeu de prouver le premier théorème :

$$\boxed{\text{"CONDORCET"} \times \text{"KIRCHHOFF"} = \text{"BORDA"}}$$

Il en résulte que si l'on a fait le dépouillement selon Condorcet un simple calcul linéaire donne le dépouillement selon Borda.
Exemple numérique ci-dessus :

$$\text{dépouillement Condorcet (I)} = \left[\begin{array}{c|c|c} -10 & -14 & -22 \\ \mathrm{AB} & \mathrm{AC} & \mathrm{BC} \end{array}\right]$$

formule de Kirchhoff pour passer à Borda

$$\left\{\begin{array}{l} \mathrm{A} = \mathrm{AB} + \mathrm{AC} \\ \mathrm{B} = \mathrm{BC} - \mathrm{AB} \\ \mathrm{C} = -\ \mathrm{AC} - \mathrm{BC} \end{array}\right.$$

d'où :

$$\text{dépouillement Borda (I)} = \left[\begin{array}{r|r|r} -10 - 14 = -24 & -22 + 10 = -12 & +14 + 22 = +36 \\ \mathrm{A} & \mathrm{B} & \mathrm{C} \end{array}\right]$$

On vérifiera de même pour l'opinion (II).
Ainsi, pour le dire en bref : Borda est une fonction linéaire de Condorcet.
Mais l'inverse n'est pas vrai car cette fonction n'est pas inversible (question de rang !). Ça explique bien des choses.

8. Pour approfondir la relation entre Borda et Condorcet, le calcul matriciel nous invite à faire le produit des deux matrices (après transposition de l'une d'elles).

$$\underset{\text{"Condor" transposée}}{\begin{array}{c|cccccc|} \text{AB} & +1 & +1 & -1 & -1 & +1 & -1 \\ \text{AC} & +1 & +1 & +1 & -1 & -1 & -1 \\ \text{BC} & +1 & -1 & +1 & +1 & -1 & -1 \\ \hline & \text{ABC} & \text{ACB} & \text{BAC} & \text{BCA} & \text{CAB} & \text{CBA} \end{array}} \quad \underset{\text{"Borda"}}{\begin{array}{|ccc|} +2 & 0 & -2 \\ +2 & -2 & 0 \\ 0 & +2 & -2 \\ -2 & +2 & 0 \\ 0 & -2 & +2 \\ -2 & 0 & +2 \\ \hline \text{A} & \text{B} & \text{C} \end{array}} = \begin{array}{|ccc|} +8 & -8 & 0 \\ +8 & 0 & -8 \\ 0 & +8 & -8 \end{array}$$

c'est-à-dire (pour n = 3) :

$$\text{"CONDOR transp."} \times \text{"BORDA"} = 8 \times \text{"KIRCH"}$$

Ici, délassement proposé au lecteur : démontrez le

Théorème 2 : $\boxed{\text{"CONDOR transp."} \times \text{"BORDA"} = \frac{(n+1)!}{3} \times \text{"KIRCH"}}$

PARENTHESE (pour qui connaît un peu l'Analyse Spectrale, dite aussi Analyse Factorielle)

On a trouvé :

$$n! \updownarrow \boxed{\text{CONDOR}} \quad \underset{\longleftarrow n \longrightarrow}{\boxed{\text{KIRCH}}} = (2) \quad \underset{\longleftarrow n \longrightarrow}{\boxed{\text{BORDA}}}$$

et

$$\boxed{\text{BORDA}^{\text{tr}}} \quad \boxed{\text{CONDOR}} = (N) \quad \boxed{\text{KIRCH}^{\text{tr}}}$$

Portons l'attention sur une colonne *k* de KIRCHHOFF et la colonne *b* de même nom dans BORDA :

$$\boxed{\text{CONDOR}} \quad \Big[k = 2. \Big[b \qquad \underline{\quad b' \quad} \ \boxed{\text{CONDOR}} = N. \ \underline{\quad k' \quad}$$

(b' et b sont transposées, comme aussi k et k').

On voit bien que b et k fonctionnent comme Vecteurs Propres vis-à-vis de CONDOR. Et toutes ces paires de vecteurs propres correspondent à une seule et même valeur propre (2N). On a donc un espace propre, et on n'aura pas envie de choisir l'un plutôt que l'autre de ses vecteurs générateurs (et encore moins de choisir une base, ce qui détruirait la belle symétrie à laquelle tout algébriste consciencieux est farouchement attaché).

Mais quand on tient une paire de vecteurs propres, comme k et b ci-dessus, on se dépêche de former le produit b × k' qui donnera une matrice de rang un, bonne approximation de CONDOR(et même la meilleure en un sens facile à préciser).

On est donc conduit à tenter d'élargir l'Analyse Factorielle traditionaliste, en traitant *ensemble* tous les vecteurs propres qu'on a eu la chance de trouver. On peut toujours essayer.

9. Pour qui ne saurait rien de l'analyse spectrale, ce serait une occasion de l'apprendre un peu, sans grande peine, en partant de soucis purement algébriques :

$$\begin{tikzcd} & \boxed{n'} \arrow[dl, "\text{CONDOR}"'] \arrow[dr, "\text{BORDA}"] & \\ \boxed{n(n-1)} \arrow[rr, "\text{KIRCH}"'] & & \boxed{n} \end{tikzcd}$$

On vient d'examiner deux relations simples entre trois matrices.

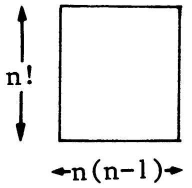
Théor. 1 CONDOR × KIRCH = 2 × BORDA

Théor. 2 $(\text{CONDOR})^{\text{transp.}} \times \text{BORDA} = N \times \text{KIRCH}$

On a évidemment envie de compléter l'étude en calculant le troisième produit :

$$\text{BORDA} \times (\text{KIRCH})^{\text{transp.}}$$

On peut prévoir que le résultat aura la même "forme" que CONDOR à savoir



Calculons d'abord le cas n = 3. On trouve sans grande peine :

$$B \times K_n^{tr} \qquad\qquad \text{CONDOR}$$

$$\begin{array}{c|ccc|} \text{ABC} & +2 & +4 & +2 \\ \text{ACB} & +4 & +2 & -2 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \hline & \text{AB} & \text{AC} & \text{BC} \end{array} \qquad \begin{array}{|ccc|} +1 & +1 & +1 \\ +1 & +1 & -1 \\ & & \\ & & \\ \end{array}$$

On a bien la même forme ; mais il y a plus : les signes des termes sont aussi les mêmes pour les deux matrices. Par contre les valeurs absolues (qui sont toujours égales à l'unité chez Condorcet) varient dans l'autre : 2 si les deux lettres sont contigues, 4 sinon. Il s'agit en quelque sorte d'une *mesure* (mais orientée : c'est-à-dire positive ou négative) de l'intervalle qui sépare les deux lettres d'un couple pour chacune des n! permutations.

10. On est ainsi amené à définir une nouvelle matrice, de même espèce que la matrice de CONDORCET (c'est-à-dire relation valuée entre les n! permutations et n(n-1)/2 couples), au moyen de la règle évidente :
dans la ligne notée : "B F A C G E D" et dans la colonne notée "C F" on écrit -2, parce que F est le 2ème à gauche à partir de C. Dans la colonne A B ce sera encore -2 et dans la colonne F G ce sera +3, et ainsi de suite.
Il faut lui donner un nom. Pourquoi pas "INTERVALLES"?
On établira alors :

Théorème 3. "BORDA" × "KIRCH. transp." = 2 × "INTERVALLES"

Il convient maintenant d'ajuster la matrice INTER pour la rapprocher la plus possible de son modèle CONDOR. Une première méthode simple : multiplier INTER par un scalaire de façon à rendre minimum la somme des carrés des termes de la différence.
Une autre, plus simple (et équivalente) :
On a vu que :

$$\begin{aligned} \text{CONDOR} \times \text{KIRCH} &= 2 \times \text{BORDA} \\ \text{BORDA}^{tr} \times \text{CONDOR} &= N \times \text{KIRCH}^{tr} \end{aligned}$$

Calculons encore :

$$\begin{aligned} \text{INTER} \times \text{KIRCH} &= n \times \text{BORDA} \\ \text{BORDA}^{tr} \times \text{INTER} &= N' \times \text{KIRCH}^{tr} \end{aligned}$$

(on peut ainsi comparer le comportement de CONDOR et celui de INTER quand on les multiplie par KIRCH ou BORDA).
On voit par conséquent que si l'on pose

$$\text{DIFF} = \text{CONDOR} - \frac{2}{n}\ \text{INTER},$$

on a $\text{DIFF} \times \text{KIRCH} = 0$ et $\text{BORDA}^{tr} \times \text{DIFF} = 0$ (orthogonalité)
ou bien $\text{DIFF} \times \text{INTER} = \text{INTER}^{tr} \times \text{DIFF} = 0.$

D'où l'approximation optimale : " $\frac{2}{n}$ INTER " ..

Théorème 4. Parmi toutes les matrices : K × "INTERV." celle qui est la plus proche (au sens des moindres carrés) de "CONDOR" est celle qui correspond à :

$$\hat{K} = \frac{2}{n}$$

11. Ainsi on peut compléter la proposition qui a été faite ci-dessus (au n°7) :
BORDA est fonction de CONDORCET,
par : connaissant BORDA on ne peut avoir qu'une idée approximative de CONDORCET.

$$\text{approx (CONDOR)} = \frac{1}{n}\,\text{(BORDA)} \times \text{(KIRCH . transp.)}.$$

Mais il convient d'étudier le biais, c'est-à-dire la matrice d'écart :

$$\text{(CONDOR)} - \frac{1}{n}\,\text{(BORDA) (KIRCHH. transp.)}.$$

Le cas n = 3 comme d'habitude nous ouvrira les yeux.

Voici le résultat du calcul :

$$\begin{array}{l|ccc|}
\text{ABC} & +1/3 & -1/3 & +1/3 \\
\text{ACB} & -1/3 & +1/3 & -1/3 \\
\text{BAC} & -1/3 & +1/3 & -1/3 \\
\text{BCA} & +1/3 & -1/3 & +1/3 \\
\text{CAB} & +1/3 & -1/3 & +1/3 \\
\text{CBA} & -1/3 & +1/3 & -1/3 \\
\hline
 & \text{AB} & \text{AC} & \text{BC}
\end{array}$$

C'est une matrice de rang un qui classe les six permutations en deux classes :

ABC, BCA et CAB : + − +
ACB, CBA et BAC : − + −

Il s'agit des permutations circulaires.

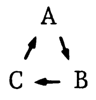
12. On trouve aisément le schéma général pour *n* quelconque.

Il faut introduire deux nouvelles matrices qui établissent le lien entre les permutations (arrangements linéaires) et les permutations circulaires (cycles de trois suffisent) d'une part − et d'autre part les cycles et les couples.

Il suffira ici d'esquisser le tableau pour n = 4.

$$\begin{array}{l|cccc|}
\text{ABCD} & +1 & +1 & +1 & +1 \\
\text{ABDC} & +1 & +1 & -1 & -1 \\
\text{ACBD} & -1 & +1 & +1 & -1 \\
\text{ACDB} & -1 & -1 & +1 & +1 \\
\text{ADBC} & +1 & -1 & -1 & +1 \\
\dots & \dots & \dots & \dots & \dots \\
\text{DCBA} & & & & \\
\hline
 & \text{(ABC)} & \text{(ABD)} & \text{(ACD)} & \text{(BCD)}
\end{array}
\qquad
\begin{array}{l|ccccc|}
\text{(ABC)} & +1 & -1 & 0 & \dots & 0 \\
\text{(ABD)} & +1 & 0 & -1 & \dots & 0 \\
\text{(ACD)} & 0 & +1 & -1 & \dots & +1 \\
\text{(BCD)} & 0 & 0 & 0 & \dots & +1 \\
\hline
 & \text{AB} & \text{AC} & \text{AD} & \dots & \text{CD}
\end{array}$$

(ABC) signifie (ABC *ou* BCA *ou* CAB) c'est-à-dire l'ordre circulaire :



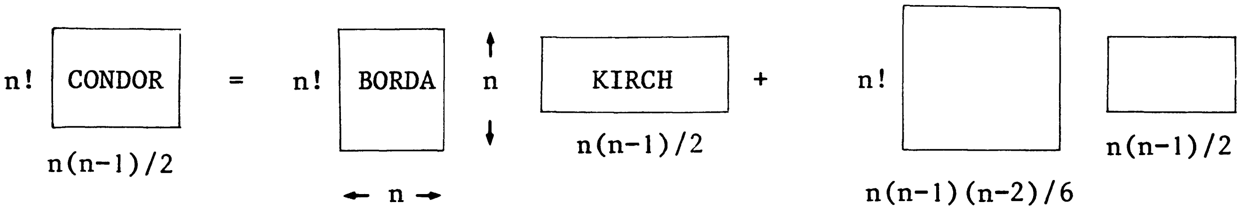
+1 : accord −1 : désaccord 0 = indifférence.

En effectuant le produit de ces deux matrices, on obtient une matrice de même espèce que "CONDOR" et que "INTER" et que nous appellerons "CYCLES". J'aurais souhaité donner des noms à ces deux qui sont ci-dessus : mais pour le moment j'hésite encore. Qu'on se le dise.

13. L'analyse spectrale de la matrice de CONDORCET est quasiment achevée :

$$\text{"CONDOR"} = \frac{2}{n}\ \text{"INTERV."} + \frac{1}{n}\ \text{"CYCLES"}$$

(les deux composantes sont orthogonales, vérifiez !)
selon les formats suivants :



Pour se faire une idée de la valeur de l'approximation qu'on fait quand on néglige les "CYCLES", il faut jauger l'*importance* des matrices.
On peut choisir la moyenne des *carrés* des éléments.
Pour CONDOR c'est évidemment l'unité.
La décomposition est alors :

$$\underset{\substack{\uparrow \\ \text{CONDOR}}}{1} = \underset{\substack{\uparrow \\ (\frac{2}{n}\ \text{INTERV})}}{\frac{2}{3}\left(1 + \frac{1}{n}\right)} + \underset{\substack{\uparrow \\ (\frac{1}{n}\ \text{CYCLES})}}{\frac{1}{3}\left(1 - \frac{2}{n}\right)}$$

Grosso modo : partage de l'importance (ou Norme) en 2/3 + 1/3. A retenir.
Si j'osais parler à la mode : BORDA retient environ les 2/3 de l'information de CONDORCET, quand n est grand.
Resterait enfin à prouver qu'on ne peut poursuivre l'analyse : pour chacun des deux morceaux les valeurs propres sont égales : on ne peut décomposer sans altérer la symétrie.

14. J'ai parlé ci-dessus de "paire propre" pour une matrice rectangulaire (à savoir "CONDOR").
Pour ceux qui ont des préjugés, je veux dire : pour ceux qui croient que les seules matrices raisonnables sont les matrices carrées, j'énoncerai ici deux corollaires de nos deux théorèmes :

Corollaire 1 : ["CONDOR" × "CONDOR transp."] × "BORDA" = S × "BORDA"
Corollaire 2 : ["CONDOR transp." × "CONDOR"] × "KIRCH" = S × "KIRCH"
(S étant valeur propre)

La première matrice carrée (de format : n! × n!), je l'appellerai matrice de KENDALL : elle fournit en effet des "distances" entre permutations et pour obtenir le coefficient de corrélation TAU de M. Kendall il suffit de diviser par : n(n-1)/2. Nous poserons donc :

"KENDALL" = "CONDOR" × "CONDOR transp." et puisqu'on a "KENDALL" × "BORDA" = S × "BORDA", nous voyons que Borda fournit des vecteurs propres à Kendall, qui l'en remercie.

15. Si BORDA peut être considéré comme une approximation de CONDORCET, alors, pour les mêmes raisons le coefficient RHO de Spearman (corrélation entre les numéros d'ordre) peut être considéré comme une approximation du coefficient TAU de Kendall.
Pour exprimer l'écart entre ces deux coefficients on pourra utiliser les relations matricielles qui précèdent (KENDALL est à CONDORCET ce que SPEARMAN est à BORDA).
Voici le résultat :
pour deux permutations de n lettres, on compte pour chacune des $\frac{n(n-1)}{2}$ paires le nombre A d'accords et celui D des désaccords.

$$\text{TAU} = A - D \ / \ A + D$$

D'une façon analogue on comptera pour toutes les permutations circulaires de trois éléments les accords et les désaccords.
(exemple : pour (ABC) il y a accord entre BDCAE et CEDAB);
Appelons SIGMA le coefficient : A-D/A+D pour les cycles de trois. Si on nomme RHO comme d'habitude le coefficient de Spearman :

$$\text{TAU} = \frac{2}{3}\ (\text{RHO}) \times \left(1 + \frac{1}{n}\right) + \frac{1}{3}\ (\text{SIGMA}) \times \left(1 - \frac{2}{n}\right).$$

16. Revenons aux problèmes de votations. C'est toujours d'actualité.
On sait que le dépouillement à la manière de CONDORCET peut conduire à une impasse : parfois on ne peut décider (ex. ci-dessus n°3 : l'état d'opinions II).
Est-ce un grand risque ?
J'ai donné autrefois une estimation (*Eléments de la Théorie des Jeux*, Paris, Dunod, 1968, page 54).
On pose : cos u = -1/3, cos v = -1/2 et on calcule : $\frac{u}{v}$ = 0,91226 ..., qui est la probabilité pour n = 3 de pouvoir décider quand, a priori, toutes les opinions sont équiprobables. On m'a demandé comment se fait ce calcul : il suffit d'étudier la matrice "CONDOR" et comment elle transforme les "aires, volumes, etc" ; c'est classique (déterminants).