MATHÉMATIQUES ET SCIENCES HUMAINES

N. RAHMANIA

Distances vectorielles entre mots

Mathématiques et sciences humaines, tome 97 (1987), p. 67-78
<http://www.numdam.org/item?id=MSH_1987__97__67_0>

Math. Sci. hum. (25^e année, n°97, *1987*, p.67-78)

DISTANCES VECTORIELLES ENTRE MOTS *

N. Rahmania **

I. Introduction.

Dans le traitement des données statistiques, le choix d'une distance entre individus ou entre variables est conditionné par la nature des observations ; ce choix peut se faire à partir d'un catalogue de distances bien connu de l'utilisateur (cf. par exemple [2]) ; cependant il peut arriver qu'aucune de ces distances ne s'applique directement aux données recueillies ; c'est par exemple le cas des enquêtes sur le devenir professionnel, la mobilité sociale, les trajectoires scolaires, les itinéraires biographiques, certaines préférences et d'une manière générale les données formalisables par des "mots", c'est-à-dire des suites de lettres d'un alphabet. On est donc amené à définir des distances entre mots. Une approche classique (cf. notamment [2]) utilise les propriétés structurelles de l'ensemble des mots considérés ; par exemple, si cet ensemble peut être structuré comme graphe on pourra utiliser une distance du "plus court chemin". Une autre approche consiste à définir une distance vectorielle classique. C'est cette dernière que nous utilisons ici.

Dans le 2ème paragraphe de ce texte, nous définissons les mots, leur codage vectoriel basé sur la notion de vecteur rang ainsi que la métrique L_γ associée ; nous présentons ensuite un autre codage vectoriel permettant une généralisation de cette métrique.

---

* Je tiens à remercier vivement Bernard Monjardet dont les suggestions ont permis la mise en forme de cet article.

** U.E.R. de Mathématiques Pures et Appliquées, Université de Lille I.
(Manuscrit reçu en janvier 1986).

Dans le 3ème paragraphe, nous considérons les mots particuliers que sont les arrangements ; nous mettrons principalement l'accent sur la relation entre trois distances possibles sur l'ensemble de ces mots. Un exemple d'application illustrera cette partie.

L'annexe sera consacrée à la représentation polyédrique des arrangements.

II. Distances entre mots.

1) *Notations et définitions*.

Soit A un alphabet de n lettres ; un mot est une suite finie $\mu = \mu_1\mu_2\ldots\mu_\ell$ de ℓ lettres ; l'entier $\ell \geq 1$ est la longueur du mot μ. On note A_ℓ l'ensemble des mots de longueur ℓ et $M = \bigcup_{\ell=1}^{q} A^\ell$ l'ensemble des mots de longueur $\leq q$; c'est sur cet ensemble M que l'on va définir des métriques.

Par définition, on dira que le rang de la lettre μ_i dans le mot μ est i et on écrira : $r(\mu_i) = i$.

Soit $\mu \in M$ et $x \in A$; on désigne par $k(x)$, ou k s'il n'y a pas d'ambiguïté, le nombre d'occurences de x dans le mot μ ; plus précisément si $\mu_{i_1} = \mu_{i_2} = \ldots = \mu_{i_k} = x$, on pose pour $j = 1,\ldots,k$:

$$r_j(x) = r(\mu_{i_j}) = i_j$$

autrement dit : $r_1(x),\ldots,r_j(x),\ldots,r_k(x)$ est la suite des rangs des k occurences de x dans le mot μ.

On définit alors le vecteur $\vec{r}_\mu(x) = (r_1,r_2,\ldots,r_q)$ de $\mathbb{R}^q$ en posant

$$\begin{array}{ll} r_j = r_j(x) & \text{pour} \quad 1 \leq j \leq k(x) \\ r_j = 0 & \text{pour} \quad k(x) < j \leq q \end{array}$$

Ce vecteur $\vec{r}_\mu(x)$, ou simplement $\vec{r}(x)$, est donc formé par la suite des rangs des k occurences de x dans le mot μ, suivie de $q-k$ termes nuls.

Fixons-nous maintenant un ordre $x_1,\ldots,x_t,\ldots x_n$ sur les éléments de A ; au mot μ, on associera la suite des n vecteurs $\vec{r}_\mu(x_1),\ldots,\vec{r}_\mu(x_t),\ldots,\vec{r}_\mu(x_n)$.

On définit ainsi une application notée S, de M dans $\mathbb{R}^{qn}$:

$$\begin{array}{l} M \xrightarrow{\ S\ } \mathbb{R}^{qn} \\ \mu \rightsquigarrow S(\mu) = (\vec{r}_\mu(x_1),\ldots,\vec{r}_\mu(x_t),\ldots,\vec{r}_\mu(x_n)). \end{array}$$

Soient μ et μ' deux mots et, pour x_t dans A, $r(x_t)$ et $r'(x_t)$ les deux vecteurs des rangs de x_t dans μ et μ'.

Une manière de définir des distances entre μ et μ' consiste à calculer d'abord une distance, notée δ_{x_t}, entre $r(x_t)$ et $r'(x_t)$, pour chaque x_t ; ensuite une fonction de δ_{x_t}.

Prenons pour δ_{x_t} la distance associée à la norme L_γ, $\gamma \geq 1$:

$$\delta_{x_t} = d_\gamma[\vec{r}(x_t),\vec{r}\,'(x_t)] = [\sum_{j=1}^{q} |r_j^t - r_j'^t|^\gamma]^{1/\gamma}$$

r_j^t et $r_j'^t$ étant les jièmes coordonnées de x_t dans $r(x_t)$ et $r'(x_t)$.

Pour discriminer entre μ et μ', on peut prendre la distance associée à la norme L_β, $\beta \geq 1$:

$$d_{\gamma,\beta}(\mu,\mu') = [\sum_{t=1}^{n} \delta_{x_t}^\beta]^{1/\beta}$$

2) *Remarques*.

a) une autre méthode consisterait à considérer $S(\mu)$ et $S(\mu')$ comme des vecteurs de $\mathbb{R}^{qn}$, ou matrices $n\times q$, et à utiliser une distance dans $\mathbb{R}^{qn}$.

b) si $\gamma = \beta = 2$, on obtient

$$d_2(\mu,\mu') = [\sum_{t=1}^{n} [[\sum_{j=1}^{q} (r_j^t - r_j'^t)^2]^{1/2}]^2]^{1/2}$$

$$= [\sum_{t=1}^{n} [\sum_{j=1}^{q} (r_j^t - r_j'^t)^2]]^{1/2}$$

qui est identique à la distance euclidienne sur les r_j^t.

c) D'une manière générale, si $\gamma = \beta$, on obtient la distance associée à la norme L_γ dans $\mathbb{R}^{qn}$; on retrouve ainsi la méthode évoquée dans la 1ère remarque.

d) Par définition, ces métriques ne tiennent pas compte des propriétés structurelles des mots et de leur signification concrète. Ce sont de telles propriétés ou significations qui permettent de préciser en quel sens on considère que deux mots diffèrent plus ou moins.Il peut alors se faire que deux mots différant "peu", en un certain sens, aient une valeur élevée de la distance définie ci-dessus. Ceci motive la généralisation suivante.

3) *Généralisation*.

Au lieu d'attribuer au couple (μ,x), $x \in A$, le vecteur $\vec{r}_\mu(x)$, on peut considérer un vecteur $\vec{\alpha}_\mu(x)$ défini abstraitement : à chaque mot μ, on associe l'application :

$$\begin{array}{ccc} A & \xrightarrow{\alpha_\mu} & \mathbb{R}^q \\ x & \rightsquigarrow & \vec{\alpha}_\mu(x) \end{array}$$

Notons A_μ l'ensemble des lettres distinctes du mot μ et $\vec{\mu}$ l'élément de $\mathbb{R}^{qn}$ défini par :

$$\vec{\mu} = \begin{pmatrix} \vec{\mu}_1 \\ \vec{\mu}_2 \\ \vdots \\ \vec{\mu}_n \end{pmatrix}$$

avec $\vec{\mu}_t = \vec{\alpha}_\mu(x_t)$ si $x_t \in A_\mu$

$= (0,0,\ldots,0)$ sinon.

Pour tout couple (μ,μ') de M^2, on pose pour $\gamma \geqslant 1$ et $\beta \geqslant 1$:

$$d_{\gamma,\beta}(\mu,\mu') = d_{\gamma,\beta}(\vec{\mu},\vec{\mu}') = [\sum_{i=1}^{n} \|\vec{\mu}_i - \vec{\mu}'_i\|_\gamma^\beta]^{1/\beta}$$

où $\|\vec{\mu}_i\|_\gamma$ désigne la norme L_γ de l'élément $\vec{\mu}_i$.

PROPOSITION 1. Si $\forall \mu \in M$, l'application α_μ est telle que :

- $\forall x \in A$, $\vec{\alpha}_\mu(x) \neq (0,0,\ldots,0)$
- (si $A_\mu = A_{\mu'}$ et $\vec{\alpha}_\mu(x) = \vec{\alpha}_{\mu'}(x)$, $\forall x \in A_\mu$) $\Rightarrow \mu = \mu'$

alors $d_{\gamma,\beta}$ est une distance.

(La preuve est évidente).

Remarques.

- Le choix du vecteur $\vec{\alpha}_\mu(x)$ ne peut se justifier que dans des problèmes spécifiques ; ainsi dans les problèmes d'ordonnancement, chaque lettre, appelée tâche, peut être soumise à de diverses contraintes : contraintes de localisation temporelle, contraintes de cumul liées aux impératifs technologiques, contraintes disjonctives en rapport avec les disponibilités des moyens... chaque lettre peut alors être repérée par sa durée, la quantité de moyens requis pour sa réalisation, son coût...

- Parmi les mots que le praticien peut rencontrer, on peut distinguer essentiellement :

* les mots où les lettres en nombre n n'apparaissent qu'une et une seule fois ; c'est le cas, par exemple, des questionnaires où on demande de classer n "choix".

* les mots qui comportent p lettres distinctes, $p < n$, choisies dans une liste de n lettres ; c'est par exemple le cas des préférences partielles ; l'exemple d'application ci-dessous en fournit une bonne illustration.

* les mots où les lettres peuvent apparaître zéro, une ou plusieurs fois ; c'est le cas des trajectoires professionnelles : un même individu peut successivement occuper des positions professionnelles différentes et récurrentes (stage de formation, emploi, arrêt-maladie, nouvel emploi, chômage, etc...)

- On peut aussi signaler que le praticien peut collecter dans certaines enquêtes (cf [4]) des données dont une partie seulement est modélisable par des mots ; c'est le cas des questionnaires qui combinent à la fois des variables nominales telles que : lieu de naissance, origine sociale, nationalité... et des trajectoires professionnelles par exemple ; la mesure de la proximité entre deux individus ω et ω' peut s'écrire :

$$D(\omega,\omega') = a\ \delta(\omega,\omega') + b\ d_{\alpha,\beta}(\mu,\mu')$$

où

* μ et μ' sont respectivement les mots associés à ω et ω'

* $d_{\alpha,\beta}$ est la distance définie ci-dessus

* δ est un indice de dissimilarité portant sur les variables nominales

* a et b sont des coefficients de pondération

III. Cas des arrangements

La distance définie ci-dessus permet d'évaluer la proximité entre les arrangements de n objets pris p à p.

Désignons par M_n^p l'ensemble des mots composés de p lettres distinctes choisies dans l'alphabet $A = \{x_1,x_2,\dots,x_n\}$.

Un élément μ de M_n^p est alors un arrangement p à p des n lettres de A.

A toute lettre de x_i de A, on peut associer le réel :

$$\alpha_\mu(x_i) = \begin{cases} 0 & \text{si } x_i \notin A_\mu \\ p-r_{x_i}+1 & \text{si } x_i \in A_\mu \quad \text{au rang } r_{x_i} \end{cases}$$

On peut ainsi identifier M_n^p à un sous ensemble $\vec{M}_n^p$ de $\mathbb{R}^n$ comme suit : à μ on fait correspondre $\vec{\mu}$ de $\mathbb{R}^n$ défini par :

$$\vec{\mu} = (y_1, y_2, \ldots, y_n)$$

où

$$y_i = \alpha_\mu(x_i), \qquad i = 1,2,\ldots,n$$

Par ailleurs, on peut aussi considérer le codage qui consiste à associer à x_i le réel :

$$\alpha_\mu^{(K)}(x_i) = \begin{cases} n-r_{x_i}+1 & \text{si } x_i \in A_\mu \quad \text{au rang } r_{x_i} \\ \frac{1}{2}(n-p+1) & \text{si } x_i \notin A_\mu \end{cases}$$

μ peut alors être identifié à $\vec{\mu}^{(K)}$ de $\mathbb{R}^n$ défini par :

$$\vec{\mu}^{(K)} = (t_1, t_2, \ldots, t_n)$$

où

$$t_i = \alpha_\mu^{(K)}(x_i), \qquad i = 1,2,\ldots,n.$$

L'apparition de la lettre K dans ce dernier codage est une référence à Kendall. On peut en effet considérer qu'un arrangement de p lettres est équivalent à un préordre total : les p lettres retenues étant totalement ordonnées, les autres étant exaequo et inférieures aux précédentes. Le dernier codage est alors une transformation affine du codage des préordres totaux utilisés notamment par Kendall (cf [7]) ; si k_i est le codage de Kendall, on a $t_i = n+1-k_i$.

Par contre les codages $\vec{\mu}$ et $\vec{\mu}^{(K)}$ ne se déduisent pas en général l'un de l'autre par transformation affine. On a effet :

$$\alpha_\mu^{(K)}(x) = \begin{cases} \alpha_\mu(x)+n-p & \text{si} \quad x \in A_\mu \\ \alpha_\mu(x)+\frac{n-p+1}{2} & \text{si} \quad x \notin A_\mu. \end{cases}$$

L'utilisation des distances associées aux normes L_β (paragraphe 2) peut donc conduire à des résultats différents, sauf si $p = n$ ou $p = n-1$, i.e. dans le cas des permutations. Dans ce dernier cas, si on prend $\beta = 2$ ou 1 on retrouve des distances classiques entre ordres totaux, la lère étant souvent appelée distance de Spearman (cf. par exemple [2] et [9]). Dans le cas général $(p < n-1)$, nous prenons $\beta = 1$ et nous étudions la relation des deux distances entre arrangements obtenues par l'utilisation des deux codages précédents.

PROPOSITION 2. Si A_μ et $A_{\mu'}$ désignent respectivement l'ensemble des lettres de μ et de μ', on a

$$d_1(\vec{\mu}^{(K)},\vec{\mu}'^{(K)}) = d_1(\vec{\mu},\vec{\mu}') + \tfrac{1}{2}(n-p+1)d_\Delta(\vec{\mu},\vec{\mu}')$$

où $d_\Delta(\mu,\mu') = |A_\mu \,\Delta\, A_{\mu'}|$ est la distance de la différence symétrique entre A_μ et $A_{\mu'}$.

Preuve : Il suffit de remarquer que :

$$\begin{aligned} d_1(\mu^{(K)},\mu'^{(K)}) &= \sum_{x\in A} |\alpha_\mu^{(K)}(x)-\alpha_{\mu'}^{(K)}(x)| \\ &= \sum_{x\in A_\mu\cap A_{\mu'}} |\alpha_\mu^{(K)}(x)-\alpha_{\mu'}^{(K)}(x)| + \\ &+ \sum_{x\in A_\mu - A_{\mu'}} |\alpha_\mu^{(K)}(x)-\alpha_{\mu'}^{(K)}(x)| + \\ &+ \sum_{x\in A_{\mu'} - A_\mu} |\alpha_\mu^{(K)}(x)-\alpha_{\mu'}^{(K)}(x)| + \\ &+ \sum_{x\in A_\mu^c\cap A_{\mu'}^c} |\alpha_\mu^{(K)}(x)-\alpha_{\mu'}^{(K)}(x)| = \\ &= \sum_{x\in A_\mu\cap A_{\mu'}} |\alpha_\mu(x)-\alpha_{\mu'}(x)| + \sum_{x\in A_\mu - A_{\mu'}} \alpha_\mu(x) \;+ \end{aligned}$$

$$+ |A_\mu - A_{\mu'}|(\frac{n-p-1}{2}) + \sum_{x\in A_{\mu'}-A_\mu} \alpha_{\mu'}(x) +$$

$$+ |A_{\mu'} - A_\mu|(\frac{n-p-1}{2}) =$$

$$= \sum_{x\in A} |\alpha_\mu(x) - \alpha_{\mu'}(x)| + |A_\mu \Delta A_{\mu'}| (\frac{n-p-1}{2})$$

Ainsi, si $A_\mu = A_{\mu'}$, $d_1(\vec{\mu}^{(K)}, \vec{\mu}'^{(K)}) = d_1(\vec{\mu}, \vec{\mu}')$
et si $A_\mu \cap A_{\mu'} = \emptyset$, $d_1(\vec{\mu}^{(K)}, \vec{\mu}'^{(K)}) = d_1(\vec{\mu}, \vec{\mu}') + p(n-p-1)$. D'où le

COROLLAIRE.

$$d_1(\vec{\mu}^{(K)}, \vec{\mu}'^{(K)}) - p(n-p-1) \leq d_1(\vec{\mu}, \vec{\mu}') \leq d_1(\vec{\mu}^{(K)}, \vec{\mu}'^{(K)}).$$

Exemple d'application :

Dans l'enquête portant sur les chômeurs de Roubaix-Tourcoing (cf. [1]) il s'agissait pour chaque individu de caractériser son existence quotidienne, celle de son groupe d'âge et du groupe d'âge opposé ; pour cela, on lui proposait de choisir, dans l'ordre et pour chaque groupe 4 sommets ou états parmi 26 possibles (bal, cinéma, stade, café, formation, école, A.N.P.E., jardin public, super-marché, commissariat, maison des parents, hôpital,...) ; chaque individu est alors identifié à un mot μ à 4 "lettres".

A chaque lettre x, on attribue le réel :

$$\alpha_\mu(x) = \begin{cases} 0 & \text{si } x \notin A_\mu \\ 5 - r_x & \text{si } x \in A_\mu \text{ au rang } r_x. \end{cases}$$

La distance définie ci-dessus (β=2) a permis d'évaluer la proximité entre les individus et de produire une typologie cohérente (en appliquant une méthode de classification automatique). Par exemple, pour les mots des "jeunes", on obtient des groupes assez homogènes. Les quatre premiers groupes partagent tous une même préoccupation, celle de l'emploi ; le sommet ANPE est le plus souvent en tête et les différences entre les groupes peuvent s'interpréter comme des adaptations secondaires à ce point de départ commun ; celles-ci s'organisent soit autour de la famille, soit autour de la formation, soit enfin autour du café. A l'inverse des groupes précédents, les groupes 5 et 6 choisissent d'abord la maison familiale et se distinguent ensuite essentiellement par la place accordée à la recherche de l'emploi ou à la formation ; les groupes 7 et 8 rassemblent les mots qui partent de l'école ou du centre de formation; si le groupe 7 passe par l'ANPE, il évite complètement la maison des parents, le groupe 8 exclut toujours l'ANPE pour les lieux de détente et de rencontre. Le groupe 9, enfin, est caractérisé par l'évitement complet des points de départ de tous les groupes précédents : le plus grand nombre de mots tourne autour du quatuor : bal, café, ciné, stade : c'est le groupe qui privilégie ce qui peut apparaître comme les côtés agréables de la jeunesse.

ANNEXE

Représentation polyédrique des arrangements

Rappelons que nous notons $\vec{M}_n^p$ l'ensemble des vecteurs de $\mathbb{R}^n$ codant les arrangements de p lettres parmi n au moyen du codage $\vec{\mu}$. Il est bien connu que $\vec{M}_n^n$ est l'ensemble des sommets d'un polyèdre convexe, le permutoèdre, contenu dans une sphère de $\mathbb{R}^{n-1}$, de centre $G = (\frac{n+1}{2},\ldots,\frac{n+1}{2})$ et de rayon $\frac{1}{2}\sqrt{\frac{n^3-n}{3}}$ (cf. [5],[6],[8] ou [10]). Ce résultat se généralise sans difficulté pour $p < n$: $\vec{M}_n^p$ est l'ensemble des sommets d'un polyèdre convexe contenu dans une sphère de $\mathbb{R}^{n-1}$ de centre $G = (\frac{p(p+1)}{2n},\ldots,\frac{p(p+1)}{2n})$ et de rayon $(\frac{p(p+1)}{12n}[2n(2p+1)-3p(p+1)])^{1/2}$. En outre si μ et μ' sont deux éléments distincts de M_n^p, on pose : $d_2(\mu,\mu') = d_2(\vec{\mu},\vec{\mu}')$ leur distance ; on a alors la :

PROPOSITION 3.

a) $d_2(\mu,\mu') \geq \sqrt{2}$ et $d_2(\mu,\mu') = \sqrt{2}$ si et seulement si on se trouve dans l'un ou l'autre des cas :

- μ et μ' ne différent que par la transposition de 2 éléments successifs ; on dira alors qu'ils sont adjacents du 1er genre.
- μ et μ' ne différent que par le dernier élément ; on dira qu'ils sont adjacents du 2ème genre

b) Tout arrangement μ admet exactement $(n-1)$ arrangements adjacents dont $(p-1)$ du 1er genre et $(n-p)$ du 2ème genre ; le nombre total de paires d'arrangements adjacents est $\frac{1}{2}(n-1)\frac{n!}{(n-p)!}$, dont $\frac{1}{2}(p-1)\frac{n!}{(n-p)!}$ du 1er genre et $\frac{1}{2}(n-p)\frac{n!}{(n-p)!}$ du 2ème genre.

Exemples.

- <u>$n = 3, p = 1$</u> : $\vec{M}_3^1$ est formé des points (1,0,0),(0,1,0) et (0,0,1) correspondant respectivement à 1, 2, 3 ; figure 1

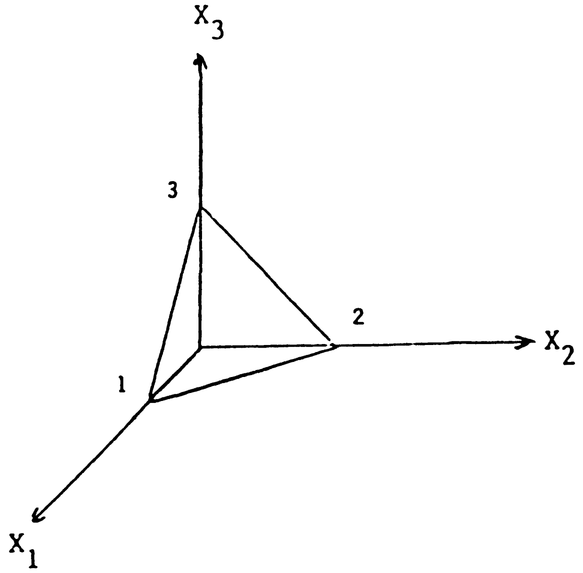

Fig. 1

- <u>$n = 3, p = 2$</u>. $\vec{M}_3^2$ est formé des points (1,2,0), (2,1,0) (1,0,2), (2,0,1), (0,1,2) et (0,2,1) correspondant respectivement à 21, 12, 31, 13, 32 et 23 ; c'est un hexagone régulier ; figure 2

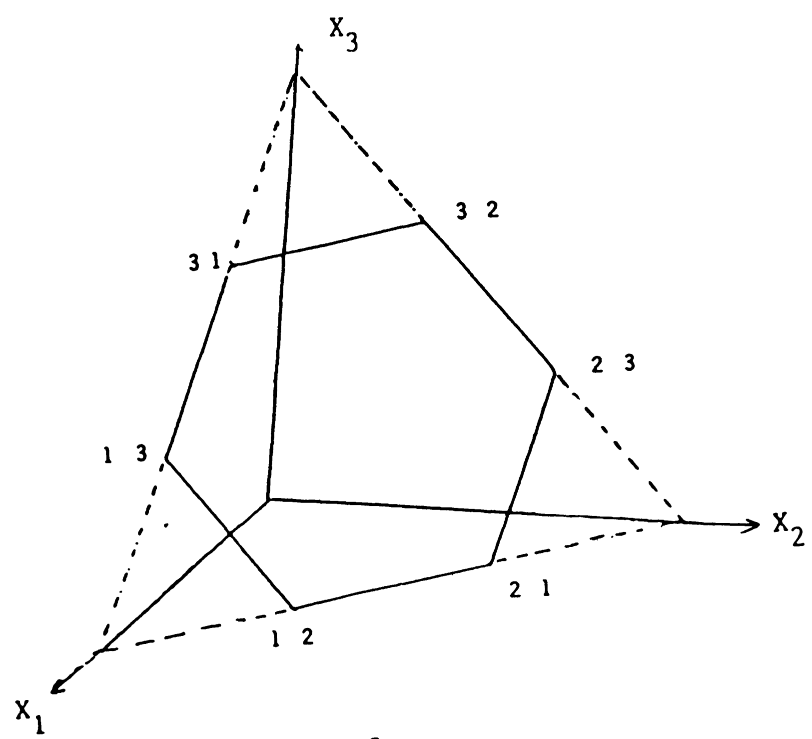

Fig. 2

- <u>n = 4, p = 2</u>. $\vec{M}_4^2$ est formé de 12 sommets et peut être représenté dans l'hyperplan d'équation : $\sum_{i=1}^{4} x_i = 3$; on obtient un polyèdre semi-régulier à 12 sommets, 18 arêtes et 8 faces ; quatre des 8 faces sont des hexagones réguliers et les 4 autres des triangles équilatéraux ; figure 3

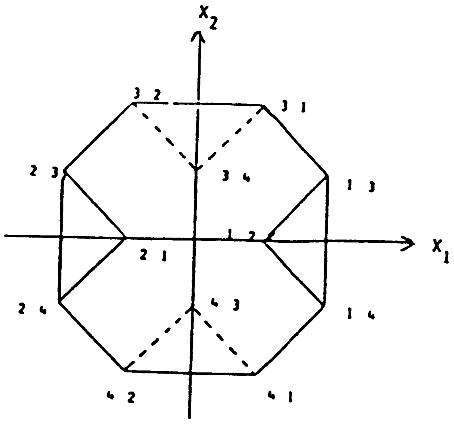

Fig. 3

- <u>n = 4, p = 3</u>. $\vec{M}_4^3$ a 24 sommets dans R^4 et peut être représenté dans l'hyperplan d'équation $\sum_{i=1}^{4} x_i = 6$; l'enveloppe convexe est un polyèdre à 36 arêtes et 14 faces ; 8 des 14 faces sont des hexagones réguliers et les 6 autres des carrés ; figure 4.

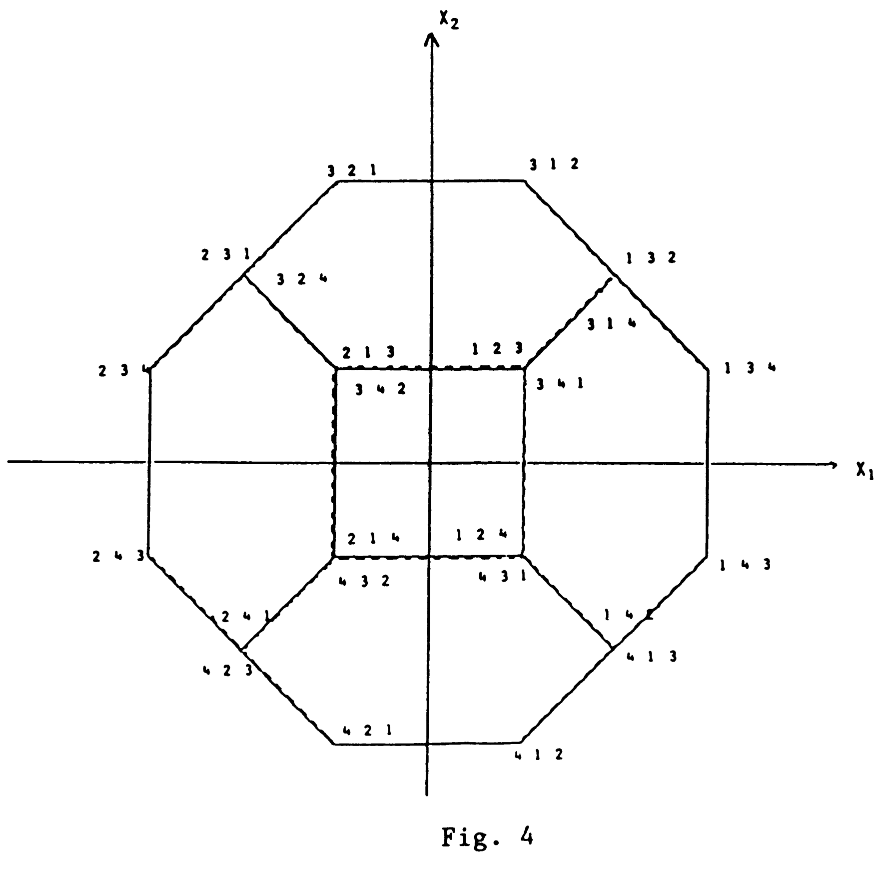

Fig. 4

Remarque.

L'ensemble des (n-p)! permutations correspondant à un arrangement p à p définissent les sommets d'une face du permutoèdre usuel ; cet arrangement peut donc être représenté par cette face en vertu des résultats de Kreweras (cf. [8])sur la bijection entre les faces du permutoèdre et les préordres totaux ; M_n^p peut donc être représenté par les faces du permutoèdre usuel.

BIBLIOGRAPHIE

[1] ABECEDAIRE, (BECHAIRE C., CUKROWICZ H., DUBUS A., DUPREZ J.M., RAHMANIA N.)
Calculs sioux pour réponses rusées : l'analyse de réponses non verbales ; choix d'itinéraires, calcul de distances et typologie ; application au cas des chômeurs de Roubaix-Tourcoing, Cahiers lillois d'économie et de sociologie, Vol.6, 3-17, 1985.

[2] CAILLEZ F., PAGES J.P.,
Introduction à l'analyse des données , SMASH, (1976).

[3] COHEN G., DEZA M.,
Distances invariantes et L-cliques sur certains demi-groupes finis. Math. Sci. Hum., 67 (1979), 49-69.

[4] GADREY N., RAHMANIA N.,
Insertion professionnelle et trajectoires de jeunes, Cahiers lillois d'économie et de sociologie, Vol. 8, 77-88, 1986.

[5] GAIHA P., GUPTA S.K.,
Adjacent vertices on a permutohedron, SIAM Jal. Appl. Math. Vol 32, n°2, 1977.

[6] GUILBAUD G.Th., ROSENSTIEHL P.,
"Analyse algébrique d'un scrutin", Math. Sci. Hum., 4 (1963), 9-33

[7] KENDALL M.G.,
Rank correlation methods, Griffin, Londres, 1948.

[8] KREWERAS G.
Représentation polyédrique des préordres complets finis, in Ordres totaux finis, 71-100, Gauthier-Villars, Paris, 1971.

[9] MONJARDET B.,
Concordance et consensus d'ordres totaux, les coefficients K et W, Revue de Statistique appliquée , 23, 2, (1985), 55-85.

[10] MONJARDET B.,
Polyèdres, permutoèdres, U.E.R de Mathématiques, logique formelle et informatique, Université René Descartes, 1986.