

A. ABI-FARAH

**Sur la double représentation graphique en
analyse des correspondances de l'ensemble J
par le nuage $N(J)$ et le simplexe S_J**

Les cahiers de l'analyse des données, tome 12, n° 3 (1987),
p. 371-378

http://www.numdam.org/item?id=CAD_1987__12_3_371_0

© Les cahiers de l'analyse des données, Dunod, 1987, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

SUR LA DOUBLE REPRÉSENTATION GRAPHIQUE EN ANALYSE DES CORRESPONDANCES DE L'ENSEMBLE J PAR LE NUAGE N(J) ET LE SIMPLEXE S_J [REPR. SIMPLEXE]

A. ABI-FARAH*

1 Rappel : représentation simultanée et principe barycentrique

Partant d'un tableau de correspondance, on construit classiquement deux nuages :

$$N(I) = \{(f_J^i, f_i) / i \in I\} \text{ in } S_J \text{ in } H_J \text{ in } R_J ;$$

$$N(J) = \{(f_I^j, f_j) / j \in J\} \text{ in } S_I \text{ in } H_I \text{ in } R_I .$$

Ces notations peuvent être rappelées, e.g. pour $N(I)$: R_J est l'ensemble des systèmes de masses, de signes quelconques sur J ; H_J est dans R_J ; H_J est, dans R_J , l'hyperplan des systèmes dont la masse totale est 1; S_J est, dans H_J , le sous-ensemble, ou simplexe, des lois de probabilités, ou systèmes de masses, toutes positives ou nulles, dont le total est 1. En introduisant dans les espaces vectoriels ambiants (R_J et R_I) la métrique du chi 2, définie par les profils (f_j et f_i) marginaux du tableau considéré, profils qui sont aussi les centres de gravité des nuages (respectivement $N(I)$ et $N(J)$), on peut rapporter chaque nuage au système orthonormé de ses axes principaux d'inertie; et en identifiant les deux systèmes d'axes, obtenir une représentation simultanée de $N(I)$ et $N(J)$. L'intérêt de cette représentation étant lié à la formule de transition (ou barycentrique) liant les coordonnées (appelées facteurs des points des deux nuages):

$$F_\alpha(i) = \lambda_\alpha^{-1/2} \Sigma \{f_j^i G_\alpha(j) / j \in J\} ;$$

$$G_\alpha(j) = \lambda_\alpha^{-1/2} \Sigma \{f_i^j F_\alpha(i) / i \in I\} .$$

(*) Faculté des Sciences, Université Libanaise, Hadath - Beyrouth.

Si l'on fait abstraction du coefficient $\lambda_\alpha^{-1/2}$ l'interprétation de ces formules

est d'une simplicité parfaite. La coordonnée $F_\alpha(i)$ d'un élément i de I sur l'axe α , est la moyenne pondérée des coordonnées des éléments j de J sur le même axe, les coefficients de pondération étant fournis par la ligne i elle-même du tableau des données. En effet,

$$\Sigma\{f_j^i G_\alpha(j) \mid j \in J\},$$

n'est autre que :

$$\Sigma_j\{k(i,j) G_\alpha(j)\} / \Sigma_j\{k(i,j)\}.$$

Toujours en faisant abstraction du coefficient $\lambda_\alpha^{-1/2}$, on peut formuler un énoncé géométrique en mettant en relation des points et non des coordonnées: le point i (identifié au profil f_i^i) se place au barycentre des points j (identifiés au f_j^j), muni chacun de la masse f_j^j ; ou, ce qui revient au même, puisque le barycentre ne change pas quand on multiplie toutes les masses par une même constante: $k(i,j)$.

Mais, pour séduisante qu'elle soit, cette formule qui efface les coefficients $\lambda_\alpha^{-1/2}$ est fautive; mis à part le cas où le tableau analysé est décomposé en blocs diagonaux, et où (on se restreint à la proportion des nuages sur le sous-espace engendré par les axes non triviaux relatifs à la v.p. 1). Dans la pratique ordinaire, on s'accommode de l'effet des coefficients, parce qu'on ne considère pas la place exacte des points, mais seulement la direction dans laquelle ceux-ci s'éloignent de l'origine: tandis que classiquement, "le barycentre tombe à l'intérieur du polygone de sustentation" (c'est-à-dire que f_j^j devrait tomber à l'intérieur de l'enveloppe du nuage des f_j^j), les coefficients $\lambda_\alpha^{-1/2}$ (tous supérieurs à 1, parce que les valeurs propres λ_α sont inférieures à 1), font sortir chacun des nuages de l'enveloppe de l'autre; et c'est même par la nécessité que la double application du principe barycentrique ne fasse pas rentrer $N(I)$ strictement à l'intérieur de lui-même, qu'on démontre le plus simplement que les valeurs propres ne peuvent dépasser 1.

Si l'on demande aux graphiques plans issus de l'analyse factorielle non l'expression des tendances générales régissant les associations entre I et J, mais la traduction exacte des proportions de chacun des j dans la ligne i, on est contraint de sacrifier, dans une certaine mesure, à l'exactitude, la symétrie des rôles entre I et J. On le fera si l'ensemble J compte un très petit nombre d'éléments, ce qui permet d'embrasser véritablement du regard dans tous ses détails le nuage $N(I)$, placé dans un espace de dimension assez faible pour que notre intuition, (aidée par les constructions géométriques), y soit à l'aise.

L'objet du présent travail est de montrer comment se conjuguent, dans ces représentations, des constructions usuelles et les principes généraux de l'analyse des correspondances.

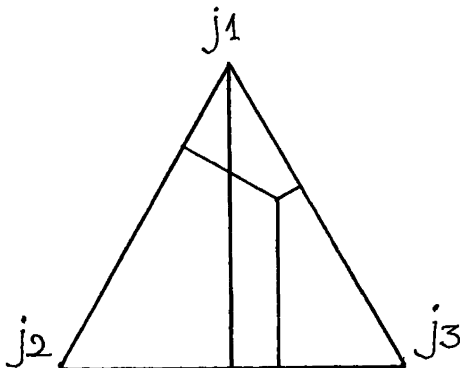
2 Simplexe régulier euclidien, simplexe des profils

Dans le cas d'un tableau à J colonnes, de nombreuses disciplines, (chimie, géographie, sociologie,...) utilisent ordinairement des diagrammes triangulaires: soit

$$J = \{ j_1, j_2, j_3 \} ;$$

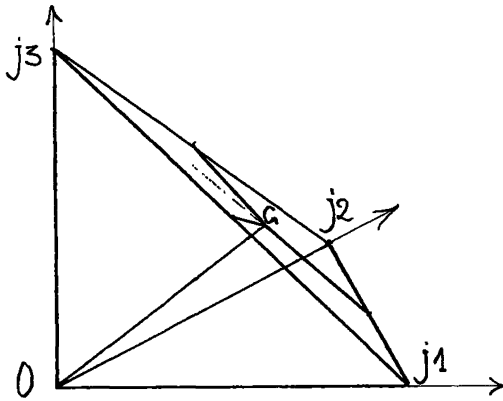
On place les points j_1, j_2, j_3 aux sommets d'un triangle équilatéral; et tout individu i est placé à l'intérieur de ce triangle, au barycentre des j affectés des masses $k(i,j)$. Pour les usagers non familiers avec le calcul de barycentre, on dit encore que i est placé à une distance de chacun des côtés d'autant plus faible, qu'est plus faible, la part dans la ligne i, du sommet opposé à ce côté: si $k(i,j_1) = 0$, alors i est sur le côté (j_2, j_3) ; au contraire si $k(i,j_2) = k(i,j_3) = 0$, i est, comme de raison placé au sommet. Plus précisément, si les trois hauteurs du triangle équilatéral sont égales à 1, la distance de i au côté (j_2, j_3) est exactement égale à

$$k(i,j_1)/(k(i,j_1) + k(i,j_2) + k(i,j_3)) = f_{j_1}^i$$



Cette représentation appelle notre attention sur le fait qu'est constante la somme des distances d'un point i , (intérieur à un triangle équilatéral), aux trois côtés de ce triangle. Cette propriété peut être démontrée de bien des manières; il vaut la peine d'en donner une démonstration générale, valant en toute dimension, pour un simplexe régulier, (généralisation du triangle équilatéral).

Plaçons-nous dans l'espace R_J , rapporté à Card J axes de coordonnées; et supposons, pour l'instant que la métrique est la somme des carrés des coordonnées sans coefficients de pondération, (afin de soutenir l'intuition nous faisons une figure pour le cas de Card $J = 3$).



Les extrémités des vecteurs unitaires des axes sont disposés aux sommets d'un simplexe régulier: En effet dans $d(j,j')^2$, tous les carrés de différence de coordonnées sont nuls, exceptés les termes en j et j' qui valent chacun $(1-0)$; donc si $(j \neq j')$ $d(j,j') = \sqrt{2}$. On a ainsi, croyons-nous, la manière la plus simple, parce que parfaitement symétrique, d'introduire un simplexe régulier en géométrie analytique.

Le centre du simplexe n'est autre que le point G dont toutes les coordonnées valent $(1/\text{Card } J)$; il est clair en effet que ce point est le centre de gravité du simplexe (au sens usuel, avec masses des sommets toutes égales); d'ailleurs ce point satisfait à l'équation du support du simplexe; hyperplan, lieu des points dont la somme des coordonnées vaut 1.

On peut démontrer par de multiples voies qu'est constante la somme des distances d'un point du simplexe aux faces de celui-ci. D'abord, la distance de i à la face $J - \{j\}$ engendrée par les sommets autres que j n'est autre que:

$$d_j(i) = f_j^{-1} (\text{Card } J / \text{Card } J - 1)^{1/2};$$

En effet, la coordonnée f_j^i et la distance $d_j(i)$ sont (sur l'hyperplan H support du simplexe) des formes linéaires s'annulant sur la face opposée au sommet j, et donc proportionnelles entre elles; et le coefficient de proportionnalité peut être calculé commodément en se plaçant au sommet j, qui correspond au profil δ_j^j dont toutes les composantes sont nulles sauf la jème qui vaut 1). Ce sommet se projette sur la face qui est opposée, au centre de gravité de celle-ci, c'est-à-dire un point dont toutes les coordonnées sont égales à $1/(\text{Card } J - 1)$, sauf la jème qui vaut 0. D'où pour le carré de la distance:

$$\begin{aligned} d_j^2(\delta_j^j) &= (1-0)^2 + (\text{Card } J - 1)(1/\text{Card } J - 1)^2 \\ &= 1 + (1/(\text{Card } J - 1)) = \text{Card } J/(\text{Card } J - 1); \end{aligned}$$

$$d_j(\delta_j^j) = (\text{Card } J/\text{Card } J - 1)^{1/2}$$

On a aussi montré que la somme des distances d'un point intérieur au simplexe aux faces de celui-ci n'est autre que $(\text{Card } J/(\text{Card } J - 1))^{1/2}$, car les sommes des coordonnées f_j^i de i, quand à elle vaut 1).

Une autre démonstration se fait dans l'hyperplan H_G rapporté à l'origine G dont les coordonnées valent $1/\text{Card } J$. Dans cet hyperplan, les distances aux faces sont des formes linéaires; et la somme des distances (à condition de se placer à l'intérieur du simplexe pour éviter les changements de signe de ces formes linéaires) est également une forme linéaire; celle-ci par raison de symétrie ne peut être que constante; et on s'assure qu'il en est bien ainsi en remarquant que les vecteurs associés à ces formes linéaires étant proportionnels aux vecteurs G_j ont, comme ceux-ci, une somme vectorielle nulle.

3 Propriétés orthocentriques du simplexe des profils

Cependant une question se pose: est-il légitime de représenter le simplexe euclidien régulier, c'est-à-dire un simplexe dont toutes les arêtes sont égales ? Il faut d'abord remarquer que le principe barycentrique lui-même ne requiert pas de structure métrique euclidienne: car il s'agit de géométrie affine. Les dimensions euclidiennes ne s'introduisent que par la nécessité physique de faire une figure... et d'imaginer dans l'espace. Reste donc à choisir au mieux les dimensions pour compléter les propriétés affines.

Le simplexe régulier a pour lui la symétrie parfaite des rôles dévolus à tous les éléments j de J. Mais cette symétrie est-elle dans la nature du problème ? Nous répondons: oui *a priori* ; mais non, si l'on considère le nuage $N(I)$. On fait l'analyse des correspondances introduit dans l'espace R_J où nous nous sommes d'emblée intentionnellement placés (au § 2) une métrique dont les coefficients sont fournis par le nuage $N(I)$; la métrique du chi 2. Pour cette métrique, le

simplexe S_J s'il n'est pas régulier, a cependant une propriété géométrique remarquable: il est orthocentrique.

Rappelons en quoi consiste cette propriété. En géométrie plane, on appelle hauteur d'un triangle issue du sommet j la droite passant par ce point et perpendiculaire à l'arête définie par les deux autres sommets. En toute dimension, on appelle hauteur d'un simplexe S_J issue du sommet j ($\in J$) la droite (sous espace affiné de dimension 1) passant par j et perpendiculaire à l'hyperplan engendré par les $\text{Card } J - 1$ autres sommets.

Il est bien connu que les hauteurs de tout triangle sont concourantes, passent par un même point appelé orthocentre. En revanche dès le cas $\text{Card } J = 4$ les hauteurs ne concourent pas nécessairement: les tétraèdres orthocentriques (ceux dont les hauteurs concourent) forment une classe particulière dont l'étude est classique; on sait en particulier que, dans un tel tétraèdre, les arêtes opposées sont orthogonales entre elles; et que la perpendiculaire commune à deux arêtes opposées passe par l'orthocentre (point de concours des hauteurs). En dimension quelconque, l'existence d'un point G tel que tout j , la droite Gj soit orthogonale à l'hyperplan engendré par les $(\text{Card } J - 1)$ sommets autres que j entraîne également que si A est une partie quelconque de J (non vide et distincte de J), les deux sous-espaces engendrés respectivement par G et A et H et $J - A$ se coupent perpendiculairement suivant une droite qui passe donc par H ; cette droite joint les projections orthogonales de G sur les sous-espaces affines engendrés respectivement par A et $J - A$. En effet soit α la projection orthogonale de G sur le sous-espace engendré par A ; H_α est combinaison linéaire des vecteurs ha ($a \in A$); donc ha , comme les ha est orthogonal à tout vecteur bb' ($b, b' \in B = J - A$); car puisque ha est une hauteur, h_α est orthogonal à tout vecteur jj' ($j \in a, j' \in a$), de la face opposée à a ; et H_α est orthogonal au sous-espace engendré par $B = J - A$.

Toutes ces propriétés des simplexes orthocentriques se retrouvent sous une forme particulièrement concrète, (parqu'accessible aux calculs de profils), mais aussi tout à fait générale, (sans imposer des relations entre les dimensions des arêtes, à une réserve près qui sera précisée plus bas), dans le cas de la métrique du χ^2 .

Le cadre spatial est R_J . L'orthocentre n'est autre que f_I , centre de gravité de $N(I)$ et aussi barycentre du système des δ_j^j (notés j) muni chacun de la masse f_j . En effet, on a vectoriellement:

$$G_j = GO + O_j$$

le vecteur GO est orthogonal à l'hyperplan H support du simplexe, car quel que soit j on a

$$\langle O_j, O \rangle = 1 = \langle OG, OG \rangle$$

(si on calcule les produits scalaires par la distance du χ^2 :

$$\langle u_j, v_j \rangle = \Sigma \{ u_j v_j / f_j \mid i \in J \});$$

d'où

$$\langle OG, jj' \rangle = \langle OG, O_j \rangle - \langle OG, O_{j'} \rangle = 0 .$$

Et de même :

$$\langle G_j'', jj' \rangle = 0 .$$

On peut encore écrire :

$$j, j' : (j \neq j') \implies \langle G_j, G_{j'} \rangle = - \|GO\|^2$$

Montrons d'abord que (comme c'est le cas pour tout simplexe orthocentrique, dont l'orthocentre est noté G) le produit scalaire $\langle G_j, G_{j'} \rangle$ ne dépend pas de la paire j, j' considérée. En effet on a:

$$\langle G_j, G_{j'} \rangle - \langle G_j, G_j'' \rangle = \langle G_j, j''j' \rangle = 0$$

(où j, j', j'' sont tous trois différents)

on a de plus (si $j \neq j'$) :

$$\begin{aligned} \langle O_j, O_{j'} \rangle &= \langle (OG + G_j) - (OG + G_{j'}) \rangle \\ &= \|OG\|^2 + \langle OG, G_{j'} \rangle + \langle G_j, OG \rangle + \langle G_j, G_{j'} \rangle; \end{aligned}$$

or le produit scalaire $\langle OG, G_j \rangle$ (ainsi que $\langle OG, G_{j'} \rangle$) est nul, parce que OG est orthogonal à l'hyperplan H support du simplexe; d'où

$$0 = \|OG\|^2 + \langle G_j, G_{j'} \rangle ; \quad \|OG\|^2 = -\langle G_j, G_{j'} \rangle$$

Réciproquement, tout simplexe orthocentrique J (d'orthocentre G) peut être placé sur un système d'axes orthonormés dont chacun passe par l'un des sommets. Il suffit de prendre l'origine O des axes sur la perpendiculaire en G au support du simplexe à une distance $\|GO\|$ telle que $\|GO\|^2$ soit égale à l'opposé de la valeur commune des produits scalaires $\langle G_j, G_{j'} \rangle$. En reprenant le calcul précédent on voit en effet que $(j \neq j') \implies \langle O_j, O_{j'} \rangle = 0$; car les produits scalaires $\langle OG, G_j \rangle$ sont nuls (puisque OG est perpendiculaire au support du simplexe). Cette construction de O tombe toutefois en défazut si la valeur

commune aux produits scalaires $\langle G_j, G_j \rangle$ n'est pas négative: ce qui est le cas si l'orthocentre G n'est pas intérieur au simplexe: cas qui peut se produire dès la dimension 2 (Card $J = 3$) avec un triangle ayant un angle obtus. On peut alors, éventuellement, introduire des coordonnées imaginaires pour achever la construction d'un système d'axes orthonormés passant par les sommets du simplexe.

Dans Pratique de l'A des D en Economie, Vol 5, (pp. 209-214) et CAD Vol IX (1984) n°2, pp. 212-226, on trouve une analyse factorielle où le nuage $N(I)$ s'inscrit dans un rectangle, qui est la section par un plan parallèle à deux arêtes opposées, du tétraèdre orthocentrique T . En général, il peut être utile de figurer sur un graphique d'analyse des correspondances non seulement les f_j^i et les f_j^j , mais aussi les δ_j^j , sommets du simplexe; ou certains autres profils correspondants à des cas modèles, qui servent de termes de comparaison avec les f_j^i .