

CAD

**Mémoire reçu : analyse de correspondances
sur la sphère**

Les cahiers de l'analyse des données, tome 3, n° 4 (1978),
p. 493-496

http://www.numdam.org/item?id=CAD_1978__3_4_493_0

© Les cahiers de l'analyse des données, Dunod, 1978, 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/>

MÉMOIRE REÇU

analyse de correspondances sur la sphère

par M. Volle ⁽¹⁾

Le but de M. Volle est d'analyser un tableau rectangulaire de données concrètes en le comparant à un tableau modèle, ou tableau de référence. Les calculs de M. Volle sont faits en représentant une mesure $f_I = \{f_i \mid i \in I\}$; $\sum f_i = 1$, par le point de coordonnées $\{f_i^{1/2}\}$, point qui est sur la sphère de rayon 1 ; d'où le titre du mémoire. Mais quant aux idées, l'on peut reprendre les propositions de M. Volle en s'écartant le moins possible du cadre usuel de l'analyse des correspondances ; i.e. sans le changement $f_i \rightarrow f_i^{1/2}$ et d'autres analogues. C'est ce que nous ferons ici.

1 Rappel sur la décomposition d'un tenseur : Notons suivant l'usage reçu en analyse des correspondances :

R^I espace des fonctions sur un ensemble fini I .

R_I espace des mesures sur I ,

f_I mesure positive de référence sur I ayant masse totale 1 ;

et de même, pour un autre ensemble J ; espace R^J , R_J et mesure de référence f_J .

On a entre mesures et fonctions, les isomorphismes canoniques :

$$u^I \in R^I \rightarrow (uf)_I = \{u^i f_i \mid i \in I\} \in R_I ;$$

$$\mu_I \in R_I \rightarrow (\mu/f)^I = \{\mu_i/f_i \mid i \in I\} \in R^I ;$$

$(uf)_I$ est la mesure ayant pour densité u^I , relativement à f_I ; $(\mu/f)^I$ est la densité de la mesure μ_I relativement à f_I . Ces isomorphismes sont des isomorphismes de structure euclidienne si R^I est muni de la norme $\|u^I\|^2 = \sum \{f_i (u^i)^2 \mid i \in I\}$ et R_I de la norme du χ^2 : $\|\mu_I\|^2 = \sum \{(\mu_i)^2 / f_i \mid i \in I\}$; alors une mesure et sa densité ont même norme. En utilisant pour coordonnées d'une mesure μ_I les nombres $\mu_i f_i^{-1/2} = \mu \langle i \rangle$ et pour coordonnées d'une fonction u^I les nombres $u \langle i \rangle = u^i f_i^{1/2}$; on a pour une mesure et sa densité les mêmes coordonnées et la norme au carré s'écrit comme somme des carrés de celles-ci : $\|\mu_I\|^2 = \sum \{\mu \langle i \rangle^2 \mid i \in I\}$; $\|u^I\|^2 = \sum \{u \langle i \rangle^2 \mid i \in I\}$. (Il faut bien prendre garde que $u \langle i \rangle$ n'est pas la valeur en i de la fonction u : cette valeur est notée u^i ; et c'est pourquoi on a préféré l'écrire $u \langle i \rangle$ plutôt que $u(i)$). Ces coordonnées $u \langle i \rangle$ et $\mu \langle i \rangle$ sont couramment utilisées dans les calculs. En revanche dans les résultats définitifs on leur préfère les valeurs et masses u^i , μ_i , qui d'une part sont directement interprétables ;

(1) Administrateur de l'I.N.S.E.E..

d'autre part permettent de ne pas perdre de vue que le choix d'une formule de distance (avec des coefficients non tous égaux à 1) est une étape essentielle de l'analyse des données.

Ceci posé tout tableau $h_{IJ} = \{h_{ij} | i \in I; j \in J\}$, considéré comme une mesure sur $I \times J$ admet une décomposition canonique, (cf e.g. [Red Tens.] TII B n° 6) qu'on écrit :

$$h_{ij} = f_i f_j \sum \{\varphi_\alpha^i \varphi_\alpha^j \lambda_\alpha^{1/2} | \alpha = 1 \dots\}$$

ici les φ_α^I (et de même φ_α^J) sont des fonctions normalisées deux à deux orthogonales : i.e. $\sum \{\varphi_\alpha^i \varphi_\beta^i | i \in I\} = \delta_{\alpha\beta}$; il importe de noter qu'en général les φ_α^I (ou φ_α^J) n'ont pas moyenne nulle ; et que par conséquent "norme $\|\varphi\|^2 = 1$ ", n'implique pas "variance 1". Les φ_α^I sont caractérisées par une équation aux vecteurs propres :

$$\sum \{\varphi_\alpha^i h_{ij} (1/f_j) h_{ij} (1/f_i) | i \in I, j \in J\} = \lambda_\alpha \varphi_\alpha^i$$

pour comprendre ces formules il suffit de se souvenir que 1°) la contraction ne peut s'effectuer qu'entre un indice haut et un indice bas (ce qui correspond à la dualité entre R^I et R_I) ; 2°) que la division par f_j (ou f_i) élève l'indice j (ou i) ; i.e. fait passer d'une mesure (indice bas) à sa fonction densité (indice haut). On peut encore recourir aux coordonnées orthonormées $\varphi_\alpha^{<i>}$ et $h^{<i,j>} = h_{ij} (f_i f_j)^{-1/2}$; on a alors :

$$\sum \{\varphi_\alpha^{<i>} h^{<i,j>} h^{<i',j>} | i \in I, j \in J\} = \lambda_\alpha \varphi_\alpha^{<i'>}$$

Pour les facteurs φ^J on a de même, l'équation :

$$\sum \{\varphi_\alpha^j h_{ij} (1/f_i) h_{ij} (1/f_j) | i \in I, j \in J\} = \lambda_\alpha \varphi_\alpha^j$$

Avec entre les φ_α^I , φ_α^J de même rang les formules de transition :

$$\sum \{\varphi_\alpha^i h_{ij} / f_j | i \in I\} = \lambda_\alpha^{1/2} \varphi_\alpha^j$$

$$\sum \{\varphi_\alpha^j h_{ij} / f_i | i \in I\} = \lambda_\alpha^{1/2} \varphi_\alpha^i$$

qu'on peut récrire en $\varphi^{<>}$ et $h^{<>}$:

$$\sum \{\varphi_\alpha^{<i>} h^{<i,j>} | i \in I\} = \lambda_\alpha^{1/2} \varphi_\alpha^{<j>} ;$$

$$\sum \{\varphi_\alpha^{<j>} h^{<i,j>} | j \in J\} = \lambda_\alpha^{1/2} \varphi_\alpha^{<i>} .$$

De tous ces calculs M. Volle donne une interprétation en terme de nuages et d'axes principaux d'inertie. Faisons-le par exemple pour les φ_α^I . Notons $h_I^J = \{h_{ij} / f_j | i \in I\}$, $h_I^J \in R_I$: il importe de noter que dans le cas général où nous sommes, h_I^J n'a pas en général masse totale 1, et n'est pas une mesure positive. Dans R_I muni de la métrique du χ^2 de centre f_I , les h_I^J constituent un nuage : on peut attribuer à chaque h_I^J la masse f_j , et considérer la matrice d'inertie relativement à l'origine :

$$\begin{aligned} \sigma_{ii'} &= \sum \{h_I^j h_I^{j'} | j \in J\} \\ &= \sum \{h_{ij} (1/f_j) h_{i'j} | j \in J\} ; \end{aligned}$$

et l'on voit que l'équation des facteurs est l'équation usuelle :

$$\Sigma \{\varphi_{\alpha}^i \sigma_{ii} / f_i, | i \in I\} = \lambda_{\alpha} \varphi_{\alpha}^i$$

2 Analyse des correspondances et variances connues : Dans le cas de l'analyse des correspondances, le tableau h_{IJ} est un tableau f_{IJ} de nombres positifs dont le total est 1 : de plus ce tableau fournit par ses marges les métriques f_I et f_J de R_I et R_J . Le premier facteur souvent noté $\varphi_{\alpha}^I, \varphi_{\alpha}^J$ (rang α = zéro) et appelé facteur trivial est la fonction constante 1 ; et va avec la valeur propre $\lambda_0 = 1$. La formule de décomposition canonique de f_{IJ} , n'est autre que ce qu'on appelle communément : formule de reconstitution des données en fonction des facteurs :

$$f_{ij} = f_i f_j (1 + \Sigma \{\varphi_{\alpha}^i \varphi_{\alpha}^j \lambda_{\alpha}^{1/2} | \alpha = 1 \dots\})$$

ayant séparé le terme $\alpha = 0$ relatif au facteur trivial 1, on peut donc écrire :

$$(f_{ij} - f_i f_j) = f_i f_j (\Sigma \{\varphi_{\alpha}^i \varphi_{\alpha}^j \lambda_{\alpha}'^{1/2} | \alpha = 1, \dots\}).$$

Ainsi on voit que l'analyse des correspondances est à la fois une décomposition canonique de f_{IJ} ; et une décomposition canonique de $h_{IJ} = f_{IJ} - f_i f_j$.

E. Yagolnitzer, dans son analyse comparative de deux correspondances sur $I \times J$ $f_{IJ}^{(1)}$ et $f_{IJ}^{(2)}$, construit un tableau en quatre blocs et aboutit à analyser outre le tableau moyenne (ce qui est une analyse de correspondance usuelle) le tableau différence $f_{IJ}^{(1)} - f_{IJ}^{(2)}$; dont l'analyse est une décomposition canonique (au sens du § 1) avec des poids f_i, f_j qui sont fournis par les marges du tableau moyenne (cf Cahiers Vol. II n° 3, pp 251-264).

Dans l'analyse d'un tableau symétrique (cf [Cor. Sym.] TII B n° 9 ; et Cahiers Vol. III n° 2, pp 219-234) ; on utilise essentiellement la remarque suivante : l'analyse canonique d'un tableau symétrique f_{II} ou de toute combinaison linéaire entre $f_{II}, f_i f_i$ et $f_i \delta_{II}^I$ (tableau diagonal ayant pour éléments diagonaux les f_i) conduit aux mêmes facteurs normalisés φ_{α}^I ; les valeurs propres et avec elles le rang relatif des facteurs changeant seuls.

3 La proposition de M. Volle : On trouve dans le mémoire de M. Volle, les résultats que nous avons repris au § 1, (à l'usage près des coordonnées sphériques) : définition des facteurs ; interprétation de ceux-ci par construction des nuages $N(I)$ et $N(J)$; formule de transition et de décomposition (ou reconstitution canonique). Quant aux applications, M. Volle propose d'effectuer la décomposition canonique de la différence h_{IJ} entre les données concrètes f_{IJ}^C et un modèle de référence f_{IJ}^R : $h_{IJ} = f_{IJ}^C - f_{IJ}^R$; les f_i et les f_j pouvant être tirés soit du modèle soit des données ; (éventuellement, f_{IJ}^C et f_{IJ}^R ont mêmes marges). Pour M. Volle l'analyse de correspondance usuelle utilise systématiquement comme modèle de référence $f_i f_j = f_{IJ}^R$, la loi produit des marges du tableau concret : or parfois l'écart est si grand entre f_{IJ} et $f_i f_j$ que les comparer semble injustifiable. (C'est le cas en particulier selon M. Volle pour les tableaux

d'échange inter-industriels (TEI) dont la diagonale est indéfinie : pour de telles données, M. V. envisage de prendre un f_{II}^x (les T.E.I. sont en effet des tableaux carrés) diagonal avec des f_{ii}^x proportionnels aux f_i et de poids tendant vers 1^∞ .

Selon-nous, les formules de M. Volle sont intéressantes ; et ses suggestions méritent d'être suivies d'expériences. Toutefois les remarques faites au § 2 limitent la portée de ses critiques sur la comparaison entre f_{IJ} et $f_I f_J$: en réalité, on peut aussi bien dire que l'on décompose $f_{IJ} - f_I f_J$ ou f_{IJ} lui-même (sans référence à aucun modèle d'indépendance ou autre) ; dans le cas des tableaux symétriques on peut même soustraire de f_{IJ} n'importe quelle combinaison linéaire du modèle d'indépendance et du modèle diagonal, sans changer rien aux facteurs. A une réserve près, certes importante : l'ordre des facteurs est modifié, ce qui dans l'interprétation est essentiel. Quant aux T.E.I. il nous paraît que ce sont des données artificielles dont une méthode d'analyse si heureuse soit-elle ne tirera que des structures qui se ressentent des hypothèses économiques ayant présidé à la construction des tableaux. Mais c'est là une question qui demanderait de grands développements et sur laquelle nous nous réservons de revenir après qu'on aura de multiples manières, tenté d'analyser les T.E.I. .