MATHÉMATIQUES ET SCIENCES HUMAINES

B. LE ROUX
H. ROUANET

Contrastes essentiels et directions principales d'un nuage

Mathématiques et sciences humaines, tome 95 (1986), p. 75-81
<http://www.numdam.org/item?id=MSH_1986__95__75_0>

Math. Sci. hum. (*24*e année, n°95, *1986*, p.75-81)

CONTRASTES ESSENTIELS ET DIRECTIONS PRINCIPALES D'UN NUAGE

B. LE ROUX (*) et H. ROUANET (*)

I. INTRODUCTION

Parmi les méthodes statistiques, certaines visent simplement à condenser les données (analyse en composantes principales, etc.), d'autres visent à répondre à des questions (analyse de variance et plus généralement ce que nous appelons analyse des comparaisons). Parmi ces dernières, on distinguera les méthodes de comparaisons a priori (les questions sont posées indépendamment des données) et celles de comparaisons a posteriori (les questions sont posées au vu des données). Dans cette note, nous nous proposons de montrer comment on peut rattacher la méthode fondamentale de condensation de données, la recherche des composantes principales d'un nuage, à l'analyse des comparaisons a posteriori. Dans ce but nous définirons et étudierons la notion de contraste essentiel d'un nuage.

Cette note s'inscrit dans l'entreprise de restructuration de la statistique linéaire poursuivie depuis 1968 [5] ; nous nous appuierons sur l'opposition dégagée antérieurement entre les variables, qui traduisent les données, et les mesures, qui formalisent les questions [6] ; également sur l'opposition entre l'espace des variables, qui relève de l'algèbre linéaire, et l'espace des observables, qui renvoie à une formalisation géométrique [2].

(*) Groupe Mathématiques et Psychologie - Sciences Humaines - Sorbonne - Université René Descartes - 12, rue Cujas, 75005 Paris.

II. ANALYSE DES COMPARAISONS

1 - Nous rappellerons d'abord les notions pertinentes d'analyse des comparaisons pour une variable numérique (cf [6], [3], [4]). On se donne un *support pondéré* (J, f_J), où J est un ensemble fini et $f_J = (f_j)j \in J$ une mesure de fréquence fondamentale :

$$f_j > 0 \quad \text{et} \quad \Sigma f_j = 1.$$

Rappelons [1] et [6] que nous désignons par $\mathbb{R}_J$ l'espace vectoriel euclidien des mesures sur J, et par $\mathbb{R}^J$ celui des variables (fonctions numériques) sur J.

. Un *contraste sur J* est une mesure $c_J = (c_j)j \in J$, dont la somme des coefficients (ou "masse totale") est nulle : $\Sigma c_j = 0$.

. La mesure f_J permet de définir la *norme de centre* f_J (ou f_J-norme) pour toute mesure (en particulier tout contraste) sur J :

$$\boxed{\text{Norme}\ (c_J) = (\Sigma c_j{}^2 / f_j)^{1/2}}$$

Soit (x^J, f_J), en bref x^J, une variable numérique sur (J, f_J), de moyenne $\hat{x} = \Sigma f_j\, x^j$ et de variance $\text{var}\ x^J = \Sigma f_j (x^j - \hat{x})^2$. Soit c_J un contraste sur J ; on définit :

. l'*effet* (numérique) *du contraste* c_J sur x^J par :

$$\boxed{\text{effet}\ (c_J\ ;\ x^J) = \Sigma\, c_j\, x^j}$$

(on écrira en bref : effet (c_J)) ; on remarque que :
effet $(c_J ; x^J)$ = effet $(c_J ; x^J - a.1^J)$, où $a.1^J$ est une fonction numérique de valeur constante égale à a.

. la *variance du contraste* (ou *part de variance* de la variable x^J prise en compte par le contraste c_J) par :

$$\boxed{\text{var}\ (c_J\ ; x^J) = \frac{(\Sigma\, c_j\, x^j)^2}{\Sigma c_j{}^2 / f_j}}$$

(on écrira en bref : var c_J) ; on a $\text{var}(c_J\ ;\ x^J) \leq \text{var}(x^J)$. En effet (inégalité de Schwartz) : $(\Sigma c_j x^j)^2 = (\Sigma c_j (x^j - \hat{x}))^2 \leqslant (\Sigma c_j^2/f_j).(\Sigma f_j (x^j - \hat{x})^2$.

2 - Toujours pour une variable numérique (x^J, f_J), nous définirons maintenant le *contraste essentiel* associé à cette variable comme le contraste e_J dont la f_J-densité est la variable e^J des écarts à la moyenne, avec donc $e^j = x^j - \hat{x}$, c'est-à-dire :

$$\boxed{e_j = f_j (x^j - \hat{x})}$$

La variance du contraste e_J est égale à la variance de la variable x^J, d'où l'appellation d'essentiel (elle est aussi égale à l'effet du contraste e_J, et à son carré de norme). Nous écrirons donc :

$$\boxed{\mathrm{Var}(x^J) = \mathrm{var}(e_J)}$$

Effet, variance, contraste essentiel sont des notions qui ne font intervenir la variable x^J qu'à une constante additive près ; ce qui suggère les généralisations multidimensionnelles affines présentées ci-après.

3 - On se donne désormais, dans un espace euclidien $\mathcal{U}$, un nuage sur le support pondéré (J, f_J), noté (M^J, f_J) ou en bref M^J. Le *point moyen* G du nuage est défini par $\Sigma f_j \vec{GM^j} = \vec{0}$ (vecteur nul), et la *variance du nuage* par $\mathrm{var}(M^J) = \Sigma f_j (GM^j)^2$.

> <u>Notations</u> : nous désignons les points de $\mathcal{U}$ par des lettres majuscules : M,M', etc ; nous fléchons les vecteurs de l'espace vectoriel $\mathcal{V}$ sous-jacent à $\mathcal{U}$: $\vec{MM'}$, etc. ; nous noterons $\langle .|. \rangle$ le produit scalaire dans $\mathcal{V}$, et $\|.\|$ la norme euclidienne ; enfin nous noterons simplement MM' la distance euclidienne $\|\vec{MM'}\|$.

Soit c_J un contraste sur J ; on définit :

. l'*effet* (vectoriel), ou *vecteur-effet*, du contraste c_J sur le nuage M^J comme le vecteur $\Sigma \overrightarrow{c_j PM^j}$, où P désigne un point quelconque de $\mathcal{U}$; (dans la suite on prendra souvent P=G) ; le vecteur-effet ne dépend pas de P et appartient au support affin du nuage ; il sera également noté $\Sigma \overrightarrow{c_j M^j}$, donc :

$$\boxed{\mathrm{effet}(c_J \ ; \ M^J) = \Sigma \overrightarrow{c_j M^j}}$$

(en bref effet (c_J)).

. la *variance du contraste* c_J (ou *part de variance* du nuage (M^J, f_J) prise en compte par le contraste c_J) comme le rapport du carré (scalaire) du vecteur-effet au carré de la norme du contraste, soit :

$$\boxed{\mathrm{var}(c_J, M^J) = \frac{\|\Sigma\, c_j \overrightarrow{M^j}\|^2}{\Sigma\, c_j^2/f_j}}$$

(en bref : var c_J).

Etant donné un nuage M^J, nous appellerons *homomorphisme-effet*, noté eff, l'homomorphisme qui à tout contraste $c_J \in \mathbb{R}_J$ associe le vecteur-effet $\Sigma\, c_j \overrightarrow{M^j}$: eff : $\mathbb{R}_J \longrightarrow \mathcal{V}$.

4 - Etant donné un vecteur $\vec{\alpha}$ de $\mathcal{V}$, posons :

$$\alpha^j = \langle \overrightarrow{GM^j} \,|\, \vec{\alpha} \rangle = \langle \overrightarrow{GM^j_{\vec{\alpha}}} \,|\, \vec{\alpha} \rangle ,$$

en désignant par $M^j_{\vec{\alpha}}$ la projection orthogonale du point M^j sur la droite $(G, \vec{\alpha})$; le nombre α^j est la coordonnée covariante du vecteur $\overrightarrow{GM^j}$ par rapport à $\vec{\alpha}$, et aussi la coordonnée covariante du point M^j par rapport à $(G, \vec{\alpha})$.

La variable α^J, qu'on peut appeler <u>variable covariante</u> du nuage associée à $\vec{\alpha}$, est f_J-centrée, et l'on a :

$$(GM^j_{\vec{\alpha}})^2 = \frac{(\alpha^j)^2}{\|\vec{\alpha}\|^2} \quad \text{d'où} \quad \mathrm{var}\, M^J_{\vec{\alpha}} = \frac{1}{\|\vec{\alpha}\|^2}\, \mathrm{var}\,(\alpha^J).$$

On posera $\alpha_j = f_j \alpha^j = f_j \langle \overrightarrow{GM^j} \,|\, \vec{\alpha} \rangle$; α_J est le contraste essentiel associé à la variable numérique α^J au sens défini au 2) précédent.

5 - Envisageons d'abord le cas particulier où le nuage est <u>unidimensionnel</u>, c'est-à-dire porté par une droite $\mathcal{D}$; dans ce cas :

> la variance du contraste essentiel α_J associé à tout vecteur $\vec{\alpha}$ de $\mathcal{V}$ est égale à la variance du nuage.

En effet : soit $\vec{\alpha}_1$ la projection orthogonale de $\vec{\alpha}$ sur $\mathcal{D}$, support affin du nuage. On a :

$$\langle \vec{GM}^j \mid \vec{\alpha} \rangle = \langle \vec{GM}^j \mid \vec{\alpha}_1 \rangle \quad \text{et} \quad \text{var } M^J = \sum_j f_j \left[\frac{\langle \vec{GM}^j \mid \vec{\alpha}_1 \rangle}{\|\vec{\alpha}_1\|} \right]^2$$

Par définition : $\text{var } \alpha_J = \dfrac{\|\Sigma \alpha_j \vec{GM}^j\|^2}{\Sigma \alpha_j^2 / f_j}$

Or $\Sigma \alpha_j \vec{GM}^j = \text{var } M^J . \vec{\alpha}_1$

puisque $\Sigma f_j \langle \vec{GM}^j \mid \vec{\alpha} \rangle . \vec{GM}^j = \Sigma f_j \langle \vec{GM}^j \mid \vec{\alpha}_1 \rangle . \dfrac{\langle \vec{GM}^j \mid \vec{\alpha}_1 \rangle}{\|\vec{\alpha}_1\|} . \vec{\alpha}_1$.

De plus, $\Sigma \alpha_j^2 \mid f_j = \text{var } M^J . \|\vec{\alpha}_1\|^2$

puisque $\Sigma (f_j \langle \vec{GM}^j \mid \vec{\alpha} \rangle)^2 / f_j = \sum \dfrac{f_j \langle \vec{GM}^j \mid \vec{\alpha}_j \rangle^2}{\|\vec{\alpha}_1\|^2} . \|\vec{\alpha}_1\|^2$

D'où $\quad \text{var } \alpha_J = \dfrac{(\text{var } M^J)^2 . \|\vec{\alpha}_1\|^2}{\text{var } M^J . \|\vec{\alpha}_1\|^2} = \text{var } M^J$.

6 - La propriété précédente motive, dans le cas d'un nuage de dimension quelconque, la définition générale suivante :

> On appelle *contraste essentiel* d'un nuage tout contraste admettant pour densité une variable covariante du nuage.

Notons ess($\vec{\alpha}$) le contraste essentiel associé à $\vec{\alpha}$. En vertu de la propriété précédente, l'effet de ce contraste essentiel appliqué au nuage projeté $M^J_{\vec{\alpha}}$ est proportionnel à $\vec{\alpha}$ effet (ess($\vec{\alpha}$) ; $M^J_{\vec{\alpha}}$) = var $M^J_{\vec{\alpha}} . \vec{\alpha}$, et sa variance vaut var $M^J_{\vec{\alpha}}$. Bien entendu, on n'a pas une telle propriété de proportionnalité pour l'effet de ess($\vec{\alpha}$) sur le nuage M^J.

7 - Nous appellerons *homomorphisme essentiel*, noté ess, l'homomorphisme qui à tout vecteur $\vec{\alpha}$ de $\mathcal{V}$ associe le contraste essentiel sur J : ess($\vec{\alpha}$) = $\alpha_J \in R_J$;
On aura donc :

$$\begin{aligned} \text{ess} : \mathcal{V} &\longrightarrow \mathbb{R}_J \\ \vec{\alpha} &\longmapsto \alpha_J = (f_j \langle \vec{GM}^j \mid \vec{\alpha} \rangle)j \in J. \end{aligned}$$

III. ENDOMORPHISME SOM ET DIRECTIONS PRINCIPALES DU NUAGE

Si on compose l'homomorphisme-essentiel ess avec l'homomorphisme-effet (qui à chaque contraste sur J associe son vecteur-effet), on retrouve l'endomorphisme classique de $\mathcal{V}$ qui engendre les directions principales du nuage, endomorphisme qu'à la suite de Benzécri [1] nous noterons Som :

$$\begin{tikzcd} \mathcal{V} \arrow[r, "\text{ess}"] \arrow[rr, bend right, "\text{Som}"'] & \mathbb{R}_J \arrow[r, "\text{eff}"] & \mathcal{V} \end{tikzcd}$$

$$\vec{\alpha} \xmapsto{\text{ess}} \alpha_J = (f_j < \overrightarrow{GM^j} \mid \vec{\alpha} >) j \in J \xmapsto{\text{eff}} \Sigma \alpha_j \overrightarrow{M^j}$$

A la direction principale engendrée par $\vec{\alpha}_\ell$, l'homomorphisme essentiel associera le *contraste essentiel principal* $\text{ess}(\vec{\alpha}_\ell)$, dont la densité est la variable principale y^J_ℓ associée à $\vec{\alpha}_\ell$. L'effet, <u>sur le nuage M^J</u>, d'un contraste essentiel principal, associé à la direction principale engendrée par $\vec{\alpha}_\ell$, vaut $\lambda_\ell \vec{\alpha}_\ell$, et sa variance vaut $\lambda_\ell (= \text{var}(y_{J\ell}, M^J))$. Il est clair que la valeur propre λ_1 est la valeur maximum que peut atteindre la variance d'un contraste c_J appliqué au nuage (M^J, f_J).

<u>Cas particulier de l'analyse des correspondances</u>

L'analyse des correspondances d'un tableau $J \times K$ consiste en la recherche des directions principales d'un nuage (M^J, f_J) représentant géométriquement la famille des profils $(f^j_K, f_j) j \in J$, où $\mathcal{V}$ espace vectoriel sous-jacent à $\mathcal{U}$, est identifié à l'espace $\mathbb{R}_K$ des mesures sur K ; d'où les deux propriétés suivantes :

- l'homomorphisme essentiel : $\text{ess} : \mathbb{R}_K \longrightarrow \mathbb{R}_J$ est la transition F^K_J de K vers J. En effet :

$$\begin{aligned} \text{ess}(c_K) &= (f_j < f^j_K \mid c_K >) j \in J \\ &= (\Sigma f^j_K \, c_K) j \in J \\ &= f^K_J \, (c_K) \end{aligned}$$

- l'homorphisme effet : $\text{eff} : \mathbb{R}_J \longrightarrow \mathbb{R}_K$ est la transition f^J_K de J vers K : $\text{eff}\,(c_J) = \Sigma c_j f^j_K = f^J_K\,(c_J)$.

Ainsi, en analyse des correspondances, les deux homomorphismes "essentiel" et "effet" s'identifient respectivement aux deux transitions de J vers K et de K vers J ; ils sont adjoints l'un de l'autre et jouent des rôles symétriques vis-à-vis des deux nuages puisque :

$$\mathrm{eff}(c_K\,;\,f_K^J) = \mathrm{eff}(c_K\,;\,f_J^K)$$

et $$\mathrm{ess}(c_J\,;\,f_J^K) = \mathrm{eff}(c_J\,;\,f_K^J).$$

B I B L I O G R A P H I E

[1] BENZECRI J.-P. et F. (1980) *Pratique de l'analyse des données*, Tome 1, Dunod, Paris.

[2] LE ROUX B., ROUANET H. (1979). "L'analyse statistique des protocoles multidimensionnels : analyse en composantes principales", *Publications de l'Institut de Statistique de l'Université de Paris*, XXIV, fasc. 1-2, pp. 47-74.

[3] LE ROUX B., ROUANET H. (1983) "L'analyse statistique des protocoles multidimensionnels : analyse des comparaisons", *Publications de l'Institut de Statistique de l'Université de Paris*, XXVIII, fasc. 12, pp. 47-70.

[4] LE ROUX B., ROUANET H. (1984) "L'analyse multidimensionnelle des données structurées", *Mathématiques et Sciences Humaines*, n° 85, pp. 5-18.

[5] ROUANET H., ROGALSKI J., LEPINE D. (1968) "Algèbre linéaire et la formalisation de la notion de comparaison", *Mathématiques et Sciences Humaines*, 1968, n° 24, pp. 5-16.

[6] ROUANET H., LEPINE D. (1976) "Structures linéaires et analyse des comparaisons" *Mathématiques et Sciences Humaines*, n° 56, pp. 5-46.