

B. ESCOFIER

B. LE ROUX

Affinement d'une partition et stabilité en analyse factorielle

Les cahiers de l'analyse des données, tome 4, n° 3 (1979),
p. 299-300

http://www.numdam.org/item?id=CAD_1979__4_3_299_0

© Les cahiers de l'analyse des données, Dunod, 1979, 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/>

AFFINEMENT D'UNE PARTITION ET STABILITÉ

EN ANALYSE FACTORIELLE

[PARTITION ET STABILITÉ]

par B. Escofier ⁽¹⁾ et B. Le Roux ⁽²⁾

La présente note qui complète les résultats de Deniau et Oppenheim (cf [Partition et V.P.] ce cahier pp 289-297) est un résumé des pages 19 à 35 de l'article intitulé "Etude de 3 problèmes de stabilité en analyse factorielle" publié in Pub. Inst. Stat. Vol XXI - Fascicule 3-4-1972 (paru 1976) - Paris.

Etant donné une correspondance sur $I \times J$ notée k_{IJ} , à partir d'une partition de J en $\text{Card } C$ classes : $J = \{J_c | c \in C\}$, on déduit une correspondance k_{IC} entre I et l'ensemble C des classes de J en posant $k_{ic} = \sum \{k_{ij} | j \in J_c\}$. On se propose de comparer les valeurs propres et les facteurs des correspondances k_{IJ} et k_{IC} .

On définit d'ordinaire la masse totale $k = \sum \{k_{ij} | i \in I, j \in J\} = \sum \{k_{ic} | i \in I, c \in C\}$, et les mesures de fréquence :

$$f_{IJ} = \{(k_{ij}/k | i \in I, j \in J); f_{IC} = \{(k_{ic}/k | i \in I, c \in C) ;$$

et les mesures marginales :

$$f_I = \{f_i | i \in I\} ; f_i = \sum \{f_{ij} | j \in J\} = \sum \{f_{ic} | c \in C\} ;$$

$$f_J = \{f_j | j \in J\} ; f_j = \sum \{f_{ij} | i \in I\} ; f_C = \{f_c | c \in C\} ; f_c = \sum \{f_{ic} | i \in I\}.$$

Comparaison des valeurs propres des correspondances $I \times J$ et $I \times C$

Supposons les valeurs propres des deux correspondances rangées par valeurs décroissantes et notons $\lambda_s(J)$ (resp. $\lambda_s(C)$) la s -ème valeur propre de la correspondance $I \times J$ (resp. $I \times C$). On montre dans cet article que :

$$\forall s \in [\text{Card } I] : \lambda_s(C) \leq \lambda_s(J) \leq \lambda_s(C) + \sum \{I_n(J_c) | c \in C\} \quad (1)$$

où $I_n(J_c)$ désigne l'inertie du sous-nuage $N(J_c) \subset N(J)$:

(1) Maître-Assistant I.N.S.A. - Rennes.

(2) Maître-Assistant UER MLFI - Université René Descartes - Paris.

$$I_n(J_C) = \Sigma \{f_j d^2(j, c) \mid j \in J_C\}$$

$$\text{avec } d^2(j, c) = \Sigma \{(1/f_i)((f_{ij}/f_j) - (f_{ic}/f_c))^2 \mid i \in I\}$$

et d'autre part que :

$$\forall s \in \{1, \dots, \text{Card } I\} :$$

$$\lambda_s(C) \leq \lambda_s(J) \leq \lambda_s(C) + \Sigma \{\rho_1(c) \sup\{f_{ic}/f_i \mid i \in I\} \mid c \in C\} \quad (2)$$

où $\rho_1(c)$ désigne la plus grande valeur propre (non triviale) de la sous-correspondance $I \times J_C$.

Les deux majorations de $\lambda_s(J)$ ainsi obtenues ne peuvent se comparer, cependant si on ne connaît pas la valeur propre $\rho_1(c)$; on peut la majorer par l'inertie de la sous-correspondance $I \times J_C$ ce qui donne une majoration moins fine que (1), ou même par 1 (à condition de ne sommer que sur $C^* = \{c \in C \mid \text{Card } J_C > s-1\}$)

Entre les taux d'inertie $\tau_s(J)$ de la correspondance $I \times J$, et $\tau_s(C)$ de la correspondance $I \times C$, on a par exemple :

$$\tau_s(C) (1 - Q_{in}) \leq \tau_s(J) \leq \tau_s(C) (1 - Q_{in}) + Q_{in} ;$$

Q_{in} étant défini par la formule ci-dessous (où $I_n(J)$ désigne la trace de la correspondance $I \times J$) :

$$Q_{in} = \Sigma \{I_n(J_C) \mid c \in C\} / I_n(J) ,$$

Comparaison des facteurs sur I des correspondances $I \times J$ et $I \times C$

Pour comparer les sous-espaces factoriels associés aux valeurs propres de rang $s, s+1, \dots, s+l$ des deux correspondances, on majore l'angle maximum θ entre un vecteur de l'un des sous-espaces et sa projection orthogonale sur l'autre (θ est le plus grand angle canonique) :

Posant $\varepsilon = \inf\{\lambda_{s-1}(J) - \lambda_s(J), \lambda_{s+1}(J) - \lambda_{s+l-1}(J)\}$, on a :

si $\Sigma \{I_n(J_C) \mid c \in C\} < \varepsilon$ alors

$$(\theta < \pi/4) \text{ et } \sin 2\theta \leq \Sigma \{I_n(J_C)/\varepsilon \mid c \in C\} ;$$

si $\Sigma \{\rho_1(c) \sup\{f_{ic}/f_i \mid i \in I\} \mid c \in C\} < \varepsilon$ alors

$$(\theta < \pi/4) \text{ et } \sin 2\theta \leq \Sigma \{(\rho_1(c)/\varepsilon) \sup\{f_{ic}/f_i \mid i \in I\} \mid c \in C\} ;$$

Pour comparer deux facteurs de même rang on pose $l = 0$.

Pour comparer les sous-espaces factoriels associés aux s plus grandes valeurs propres, on pose $\varepsilon = \lambda_s(J) - \lambda_{s+1}(J)$.

Commentaire : Ces résultats montrent que le regroupement d'éléments distributionnellement proches change peu les résultats de l'analyse des correspondances. Cependant le regroupement diminue les valeurs propres sans que l'on puisse assurer que les taux d'inertie décroissent. Un facteur est d'autant plus stable que son inertie est plus éloignée de celle des facteurs précédents suivants. On rapprochera ce résultat de ceux relatifs à la stabilité des v. p. et des facteurs d'une analyse de correspondance, dans le cas où l'on met des éléments en supplémentaire (cf *Cahiers* Vol I n° 3 pp 297- sqq).