Statistique et analyse des données

Jean-Louis Philoche
Jean-Jacques Daudin

Recherche d'un sous-espace vectoriel de dimension donnée traversant au mieux un nuage de points : une démonstration globale, simple et autonome

Statistique et analyse des données, tome 16, n° 3 (1991), p. 203-209.
<http://www.numdam.org/item?id=SAD_1991__16_3_203_0>

Statistique et Analyse des Données
1992 - Vol. 16 n° 3 - p . 203 - 209.

Brève communication présentée par B. Van Cutsem

RECHERCHE D'UN SOUS-ESPACE VECTORIEL DE DIMENSION DONNEE TRAVERSANT AU MIEUX UN NUAGE DE POINTS : UNE DEMONSTRATION GLOBALE, SIMPLE ET AUTONOME

Jean-Louis PHILOCHE* et Jean-Jacques DAUDIN**
INA-PG
Département de Mathématiques et Informatique
16, rue Claude Bernard
75231 Paris cedex 05

Résumé.

Dans cette note on donne une nouvelle démonstration du résultat bien connu selon lequel le sous-espace engendré par les d premiers vecteurs propres de l'opérateur d'inertie est optimal parmi les sous-espaces de même dimension.

Mots clés.

Analyse en composantes principales, inertie, optimisation.

Classification AMS :

Abstract.

In this paper a new proof is given of the well known property which states that the subspace generated by the d first eigenvectors of the inertia operator is optimal among the subspaces of the same dimension.

Key words.

Principal component analysis, inertia, optimization.

---

* INSEE/INA-PG
** INA-PG

Manuscrit reçu le 18 décembre 1991, révisé le 5 octobre 1992

En analyse factorielle, par exemple en analyse en composantes principales, on s'intéresse au problème fondamental suivant : dans l'espace dit des "individus", trouver un sous-espace de dimension d (le plus souvent petite) traversant au mieux un nuage de points donnés au sens du critère des moindres carrés. On souhaite, par là que l'inertie du nuage projeté soit la plus grande possible. Le résultat est bien connu : c'est le sous-espace engendré par des d premiers vecteurs propres de l'opérateur d'inertie. Ce problème est ancien, puisque, d'un point de vue statistique, on le voit apparaître pour la première fois, semble-t-il, dans [Pearson (1901)].

Malheureusement les démonstrations sont parfois incomplètes : on omet, par exemple, de montrer que le meilleur sous-espace de dimension p+1 contient le meilleur sous-espace de dimension p. Naturellement toutes les démonstrations ne présentent pas ce type de défaut ; la propriété d'"emboîtement" évoquée ci-dessus est en particulier clairement établie dans [Cailliez et Pagès (1976)], [Cazes (1987)], [Ladiray (1988)] et [Saporta (1990)]. Une démonstration directe d'optimalité d'un esprit différent du nôtre est donnée dans [Benzécri et Coll. (1973)], mais nous ne trouvons pas, pour notre part, qu'elle soit parfaitement claire.

Les auteurs anglo-saxons, eux, en général, ne posent pas le problème de la même façon : ils cherchent une combinaison linéaire des variables initiales de variance maximale, puis une seconde combinaison linéaire, non corrélée à la précédente, à nouveau de variance maximale et ainsi de suite. Le plus souvent les multiplicateurs de Lagrange sont utilisés et on ne s'intéresse guère aux propriétés d'optimalité de la suite des combinaisons ainsi extraites. Des présentations typiques sont celles qu'on trouve dans [Anderson (1958)] et [Morrison (1967)].

Une exception est particulièrement digne d'être mentionnée : dans [Rao, (1968)] C.R. Rao, sous l'angle des variables, traite (Résultat 8g.2(iv) (d)) le problème consistant à maximiser la somme des variances de d combinaisons linéaires (normées) des variables initiales. On peut montrer que ce problème est équivalent à celui que nous traitons, formulé, lui, en terme d'inertie dans l'espace des individus. Le résultat principal s'appuie sur le Lemme 1f-2(iv) dont la démonstration est plutôt esquissée qu'établie.

Il est bon de savoir aussi que certains auteurs se sont intéressés à un problème plus général consistant à montrer que la solution du problème de K. Pearson optimise d'autres critères que celui des moindres carrés [Okamoto (1969)], [Kobilinsky (1979)],

[Sabatier, Ian, Escoufier (1984)], etc. Notre objectif précis et limité, n'est naturellement pas de nous placer sur ce terrain.

Bien que la question soit donc, en un sens réglée, par exemple par la démonstration "pas à pas" évoquée plus haut, nous pensons qu'il est utile de donner une démonstration globale, simple et autonome. En d'autres termes on évite aussi bien l'argument "d'emboîtement" (assez peu naturel) de la démonstration "pas à pas" que la technique des multiplicateurs de Lagrange qui est manifestement issue d'un autre "horizon" mathématique. On prend le temps, auparavant, de mettre en place la situation euclidienne appropriée au problème.

On se donne n points pesants $\{x_i,p_i\}$ dans un espace euclidien E, dont le produit scalaire est noté (.,.). On introduit l'opérateur d'inertie $Q \in \mathfrak{L}(E)$, défini par :

$$Q = \sum_{i=1}^{n} p_i x_i \otimes x_i \ ; \tag{1}$$

pour $x \in E$, on a désigné par $x \otimes x$ l'opérateur tel que :

$$(x \otimes x)e = (x,e)x, \quad \text{pour } e \in E\,. \tag{2}$$

De la sorte il est naturel d'appeler Q l'opérateur d'inertie, puisque, pour e satisfaisant $\|e\| = 1$,

$$(e,Qe) = \sum_{i=1}^{n} p_i (x_i,e)^2 \tag{3}$$

est bien l'inertie[(i)] du nuage projeté sur l'axe e.
L'opérateur Q est symétrique et positif[(ii)] , il admet donc une diagonalisation orthonormale. Par la suite, les valeurs propres, non nulles, seront supposées distinctes et ordonnées :

$$\lambda_1 > \lambda_2 > \ldots > \lambda_K > 0 \ ; \quad (K = rgQ) \tag{4}$$

(i) Les inerties seront toujours prises par rapport à l'origine.
(ii) $(f,Qe) = (Qf,e)$ et $(e,Qe) \geq 0$, pour tous $e, f \in E$.

On désignera par φ_j un vecteur propre (normé) associé à λ_j ; de la sorte

$$Q = \sum_{j=1}^{K} \lambda_i \, \varphi_j \otimes \varphi_j \tag{5}$$

($\varphi_j \otimes \varphi_j$ est le projecteur sur l'axe engendré par φ_j).
On peut utiliser Q pour transformer l'inertie du nuage projeté sur un s.e.v. de E :

Lemme 1.

Soit F un s.e.v. de E, d = dim F et $e_1,...,e_d$ une base orthonormale de F

$$\sum_{i=1}^{n} p_i \|P_F x_i\|^2 = \sum_{k=1}^{d} (e_k, Qe_k) = \sum_{j=1}^{K} \lambda_j \|P_F \varphi_j\|^2 \quad (6) \tag{6}$$

Démonstration.

La première égalité s'obtient en remarquant que $\|P_F x_i\|^2 = \sum_{k=1}^{d} (x_i, e_k)^2$ et en utilisant (3) où e est remplacé par les e_k.
La seconde égalité s'obtient en commençant par utiliser (5) d'où

$$(e_k, Qe_k) = \sum_{j=1}^{K} \lambda_j (\varphi_j, e_k)^2 \quad .$$

Pour mener à bien la recherche du sous-espace optimal de dimension d nous aurons besoin d'un résultat, d'ailleurs intuitif : (dans le cas d =1, ce résultat est établi dans [Lebart et Fénelon (1971)]) :

Lemme 2.

Soit $(\lambda_k)_{1 \le j \le k}$ des nombres positifs satisfaisant (4). Pour toute famille $(m_j)_{1 \le j \le K}$ de nombres vérifiant

$$0 \le m_j \le 1, \quad \text{avec} \quad 1 \le j \le K, \tag{7}$$

(6) On a noté $P_F(x)$ la projection orthogonale de x sur F.

$$\sum_{j=1}^{K} m_j \leq d \text{ , avec } d \leq K \text{ ;} \tag{8}$$

alors on a

$$\sum_{j=1}^{K} m_j \lambda_j \leq \sum_{j=1}^{d} \lambda_j \text{ .} \tag{9}$$

Cette inégalité ne devient une égalité que si les m_j satisfont

$$m_j = 1, \text{ pour } 1 \leq j \leq d \text{ et } m_j = 0 \text{ sinon.} \tag{10}$$

Démonstration.

Clairement (9) équivaut à

$$\sum_{j=1}^{d} (1-m_j)\lambda_j \geq \sum_{j=d+1}^{K} m_j \lambda_j \tag{11}$$

(si d =K, on convient que $\sum_{j=d+1}^{K} = 0$).

Or, avec (4), (8) et (7),

$$\sum_{j=1}^{d} (1-m_j)\lambda_j \geq \lambda_d \left(d - \sum_{j=1}^{d} m_j \right) \geq \lambda_d \sum_{j=d+1}^{K} m_j \geq \sum_{j=d+1}^{K} m_j \lambda_j \text{ ,} \tag{12}$$

d'où l'inégalité (9).

Naturellement, l'égalité est atteinte si les m_j satisfont (10). C'est d'ailleurs le seul cas, car l'égalité dans (9) impose des égalités partout dans (12), ce qui oblige les conditions (10), puisque les λ_i sont tous distincts.

On peut alors établir la proposition classique dont la démonstration fait l'objet de cette note.

Proposition.

Soit n points pesants $\{x_i,p_i\}$ dans un espace euclidien E, soit Q l'opérateur d'inertie associé. On désigne par $\{\lambda_j,\varphi_j\}$ les éléments propres de Q et on suppose que les λ_j , distinctes, satisfont $\lambda_1 > \lambda_2 > \ldots > \lambda_k > 0$.

Parmi les sous-espaces de dimension d ($d \leq K$), il en existe un seul maximisant l'inertie du nuage projeté : c'est le s.e.v. engendré par $\varphi_1,...,\varphi_d$. L'inertie associée vaut alors $\lambda_1 + ... + \lambda_d$.

Démonstration.

En vertu du lemme 1, l'inertie du nuage projeté sur F vaut $\sum_{j=1}^{K} m_j \lambda_j$, en notant $m_j = \|P_F\varphi_j\|^2$.

Or ces m_j satisfont les conditions (7) et (8).
La condition (7) découle de $\|\varphi_j\| = 1$; la condition (8), de ce que

$$\sum_{j=1}^{K} \|P_F\varphi_j\|^2 = \sum_{k=1}^{d} \sum_{j=1}^{K} (\varphi_j,e_k)^2 ,$$

où $\sum_{j=1}^{K} (\varphi_j,e_k)^2 \leq 1.$

Le lemme 2 montre alors que l'optimum n'est atteint que si

$\|P_F\varphi_j\|^2 = 1$, pour $1 \leq j \leq d$

$\|P_F\varphi_j\|^2 = 0$, sinon.

Ceci équivaut à

$\varphi_j \in F$, pour $1 \leq j \leq d$;

il y a un optimum unique : le sous-espace F engendré par les d premiers vecteurs propres de Q.

REFERENCES

Anderson, T.W. (1958) An introduction to multivariate Statistical Analysis, Wiley.

Benzécri, J.P. et Coll. (1973) L'Analyse des Données T.II - Correspondances, Dunod.

Cailliez, F. et Pagès, J.P. (1976) Introduction à l'Analyse des Données, SMASH.

Cazes, P. (1987) Cours Multigraphié, I.S.U.P.

Kobilinsky, A. (1979) Ordre entre formes quadratiques. Application à l'optimalité de sous-espaces en analyse des données, R.S.A., vol. 27 n° 1.

Ladiray, D. (1988) Cours Multigraphié, E.N.S.A.E.

Lebart, L. et Fénelon J.P. (1971) Statistique et Informatique Appliquées, Dunod.

Morrisson, D.F. (1967) Multivariate Statistical methods, Mc. Graw-Hill.

Okamoto, M. (1969) Optimality of principal components, in Multivariate Analysis, 2, P.R. Krishnaiah, ed., p. 673-685. Academic Press, New-York.

Pearson, K. (1901) On lines and planes of closest fit to points in space, Phil. Magazine, 2, 559-572.

Rao, C.R. (1968) Linear Statistical Inference and its Applications, (Second corrected printing) Wiley.

Sabatier, R., Ian, Y. et Escoufier, Y. (1984) Approximations d'applications linéaires et analyse en composantes principales. Data Analysis and Informatics, E. Diday et al. (editors) North-Holland.

Saporta, G. (1990) Probabilités, Analyse des Données et Statistique, Editions Technip.