REVUE DE STATISTIQUE APPLIQUÉE

J. FINE

Problèmes d'indétermination en analyse en facteurs et analyse en composantes principales optimale

Revue de statistique appliquée, tome 41, n° 4 (1993), p. 45-72
<http://www.numdam.org/item?id=RSA_1993__41_4_45_0>

Rev. Statistique Appliquée, 1993, **XLI** (4), 45-72

PROBLÈMES D'INDÉTERMINATION EN ANALYSE EN FACTEURS ET ANALYSE EN COMPOSANTES PRINCIPALES OPTIMALE

J. Fine

Laboratoire de Statistique et Probabilités, URA CNRS D0745
Université Paul Sabatier, 31062 Toulouse cedex, France

RÉSUMÉ

Pour un même type de données (observations de p variables quantitatives sur n individus) deux analyses sont couramment utilisées : l'Analyse en Composantes Principales (A.C.P.) et l'Analyse en Facteurs (A.F.).

L'A.F. présente trois problèmes d'indétermination que nous décrivons, le plus important étant l'indétermination des facteurs.

Par ailleurs, plusieurs études ont permis de montrer qu'une légère modélisation de l'A.C.P. (par le modèle à effets fixes) conduit à un choix optimal de la métrique M sur l'espace des individus.

Nous montrons alors que la seule approximation communément admise des facteurs communs de l'A.F. est obtenue à partir de l'A.C.P. définie ci-dessus.

Mots-clés : *Analyse en Composantes Principales, Analyse en Facteurs, Modèle à effets fixes, Indétermination des facteurs, Métrique optimale.*

SUMMARY

For the same data matrix (observations of p quantitative variables on n statistical units) two analyses are currently used : Principal Component Analysis (P.C.A.) and Factor Analysis (F.A.).

F.A. presents three indetermination problems that we expose, the most important of them beeing the factor score indetermination.

Besides, several studies show that a light modelization of P.C.A. (using the fixed effect model) leads to an optimal choice for the metric M on unit space.

Finally, we show that the only one commonly accepted approximation of F.A. common factors is obtained from the P.C.A. defined above.

Key-words : *Principal Component Analysis, Factor Analysis, Fixed effect model, Factor score Indetermination, Metric choice.*

1. Introduction

Soit Y une matrice $n \times p$ de données dont les éléments sont les valeurs prises par n individus sur p variables quantitatives. Pour traiter ces données et lorsqu'aucune information auxiliaire n'est disponible, deux méthodes sont couramment utilisées : l'Analyse en Composantes Principales (A.C.P.) et l'Analyse en Facteurs (A.F.). Nous présentons ces deux analyses et nous exposons les débats qui opposent les adeptes de chacune de ces deux méthodes.

Le paragraphe 2 concerne l'A.F. : le modèle de l'A.F. et les différents types d'indétermination. Le plus sévère est l'indétermination des facteurs, c'est-à-dire que, quelque soient les contraintes utilisées, le modèle d' l'A.F. admet une infinité de solutions. La pratique consiste à fournir un seul prédicteur pour l'ensemble des solutions, prédicteur obtenu à partir d'une A.C.P..

Dans le paragraphe 3 nous présentons l'A.C.P. classique (Hotteling, 1933) et l'A.C.P. descriptive avec choix arbitraire de la métrique M sur l'espace des individus. Plusieurs études ont permis de montrer qu'une légère modélisation de l'A.C.P. (par le modèle à effets fixes) conduit à un choix optimal pour M.

En conclusion nous montrons que la seule approximation communément admise des facteurs communs de l'A.F. est obtenue à partir de l'A.C.P. optimale définie ci-dessus.

2. Le modèle de l'analyse en facteurs

2.1 L'Analyse en Facteurs

L'Analyse en Facteurs (Communs et Spécifiques) est introduite par Spearman en 1904 pour un seul facteur commun et généralisée par Garnett en 1919 pour plusieurs facteurs communs. A partir de 1936, Thurstone, qui semble ignorer les travaux des anglais qui l'ont précédé, popularise l'analyse en facteurs communs multiples grâce à la revue Psychometrika. Cf. aussi Thurstone, 1935, 1947.

Soit L_2 l'espace des variables aléatoires réelles (v.a.r.) admettant des moments d'ordre 2. C'est un espace de Hilbert pour le produit scalaire qui à (f, g) associe $E(fg)$. Dans la suite, toutes les v.a.r. sont supposées appartenir à L_2.

Soit $y = (y^1 ... y^p)'$ un vecteur aléatoire de $\mathbb{R}^p$. On confond le vecteur et la matrice colonne de ses coordonnées dans la base canonique. Le modèle de l'A.F. est alors le suivant :

$$\begin{cases} \text{Il existe } \Lambda, \ p \times q, \mu, \ p \times 1, \\ f = (f^1 ... f^q)', E(f) = 0, \\ e = (e^1 ... e^p)', E(e) = 0, E(fe') = 0 \\ \text{et } E(ee') \text{ diagonale régulière, tels que :} \\ y = \Lambda f + \mu + e \end{cases} \tag{2.1}$$

On a alors $E(y) = \mu$.

On pose : $x = y - \mu$, $\Sigma = E(xx')$, $\Phi = E(ff')$ et $\Psi = E(ee')$.

On déduit du modèle de l'Analyse en Facteurs (2.1) la relation suivante entre les matrices de covariance, appelée dans la suite modèle de stucture de covariance :

$$\Sigma = \Lambda\Phi\Lambda' + \Psi \tag{2.2}$$

Avec des notations évidentes on a, pour tout j de $1, ..., p$:

$$x^j = \sum_{k=1}^{q} \lambda_j^k f^k + e^j.$$

Les éléments $f^1, ..., f^q$ de f sont appelés les *facteurs communs* aux p variables observables et les éléments $e^1, ..., e^p$ sont appelés les *facteurs spécifiques*. Ils sont parfois appelés facteurs uniques et chacun d'eux est supposé décomposable en la somme d'un facteur spécifique et d'une erreur. Nous ne ferons pas cette distinction et préférons utiliser le terme spécifique plutôt qu'unique.

La matrice Λ est appelée en anglais «*factor loadings matrix*» ou «pattern matrix». Lorsque l'on suppose $\Phi = I$, Λ est la matrice des covariances entre les variables observables et les facteurs communs : $\Lambda = E(xf')$.

On remarque que l'on a :

$$\operatorname{diag}(\Sigma) = \operatorname{diag}(\Lambda\Phi\Lambda') + \Psi$$

et

$$\Sigma - \operatorname{diag}(\Sigma) = \Lambda\Phi\Lambda' - \operatorname{diag}(\Lambda\Phi\Lambda').$$

Les covariances des variables observables sont complètement reconstituées à partir des facteurs communs. En revanche, les variances se décomposent en une part due aux facteurs communs appelée *communalité* et une part due aux facteurs spécifiques appelée *spécificité* (ou unicité).

Sans perte de généralité on suppose que Φ est de rang q, sinon on se ramène à des facteurs communs linéairement indépendants pour une valeur inférieure à q.

Si on pose $z = \Lambda f + \mu$ et $\Theta = E((z-\mu)(z-\mu)') = \Lambda\Phi\Lambda'$, on a un modèle à effets aléatoires dont la présentation est assez proche du modèle à effets fixes qui sera exposé au paragraphe 3.4. :

$y = z + e$, où z est un effet aléatoire appartenant presque sûrement à un sous-espace affine de dimension q, F_q, de $\mathbb{R}^p$ et e est une erreur aléatoire centrée, non corrélée avec z et de matrice de covariance diagonale régulière.

En fait e est ici le vecteur des facteurs spécifiques et les différences essentielles par rapport au modèle à effets fixes sont

1) le caractère aléatoire des effets,

2) l'hypothèse de non corrélation des erreurs (Ψ diagonale),

3) l'introduction des facteurs communs.

Si A est une matrice carrée d'ordre p diagonalisable, on note $[A]_q$ l'espace engendré par les vecteurs propres associés aux q plus grandes valeurs propres de $A(q < p)$, et on l'appelle q-espace principal de A.

Comme on a : $\Sigma = \Theta + \Psi$, donc aussi : $\Sigma\Psi^{-1} = \Theta\Psi^{-1} + I$, il est aisé de voir que l'on a : $F_q = \mu + E_q$ avec $E_q = Im(\Theta) = [\Sigma\Psi^{-1}]_q$.

L'existence d'une solution (Λ, Φ, Ψ) du modèle (2.2) est une condition nécessaire de l'existence d'une solution (Λ, f, e) du modèle (2.1) aussi la démarche consiste-t-elle à étudier d'abord le modèle (2.2) puis le modèle (2.1).

Il est important de différencier les deux aspects de l'A.F. :

1) l'estimation de Λ («factor loadings»), afin d'étudier les relations entre les variables observables et les facteurs communs.

2) l'«estimation» de f («factor scores»), afin de prédire les valeurs de f sur les individus à partir des valeurs observées sur les variables initiales.

L'A.F. a une propriété souvent mise en avant pour la préférer à l'A.C.P. : son invariance par changement d'échelle des variables. En effet, soit D une matrice diagonale, on déduit du modèle (2.2) que l'on a :

$$D\Sigma D = (D\Lambda)\Phi(\Lambda' D) + (D\Psi D), \text{ avec } D\Psi D \text{ diagonale.}$$

La solution (Λ, Φ, Ψ) pour Σ fournit donc la solution $(D\Lambda, \Phi, D\Psi D)$ pour $D\Sigma D$. En particulier il est indifférent de faire l'analyse sur la matrice de covariance ou sur la matrice de corrélation.

Il existe plusieurs niveaux d'indétermination :

1) Le problème d'identification de Λ («factors loadings») appelé encore problème de «rotation» des facteurs communs.

2) Le problème d'identification du modèle de structure de covariance (2.2).

3) Le problème d'indétermination des facteurs («factor score indetermination») dans le modèle (2.1).

2.2. Problème de «rotation» des facteurs communs

On peut remarquer que, si (Λ, f, e) est solution de (2.1) alors, pour tout A matrice $q \times q$ régulière, $(\Lambda A, A^{-1}f, e)$ est solution de (2.1). De même, si (Λ, Φ, Ψ) est solution de (2.2) alors $(\Lambda A, A^{-1}\Phi A'^{-1}, \Psi)$ est solution de (2.2). Cette indétermination correspond au choix d'une base $(\lambda^1, ..., \lambda^q)$ du sous-espace E_q de $\mathbb{R}^p$ dans lequel appartiennent $p.s.$ les effets aléatoires centrés, ou encore de façon duale, au choix d'une base $(f^1, ..., f^q)$ de l'espace des facteurs communs dans L_2.

Il est possible, à partir d'une solution f, en utilisant une transformation régulière A dans l'espace des facteurs communs, d'obtenir des facteurs plus facilement interprétables. Les éléments de Λ étant les coordonnées des variables observables sur les facteurs communs, si Λ contient de nombreux éléments nuls (structure simple pour Λ), l'interprétation des facteurs communs devient plus facile. La transformation A a pour objectif d'obtenir une structure approximativement simple pour Λ, c'est-à-dire, avec le plus grand nombre possible d'éléments proches de zéro.

Si la solution initiale f est composée de facteurs orthogonaux et que l'on désire conserver cette propriété, on utilisera des rotations (transformations orthogonales).

La recherche d'une structure approximativement simple pour Λ repose sur des méthodes procrustes (Mosier 1939, Green 1952, Schönemann 1966, Browne 1967). Des algorithmes sont implémentés sur les logiciels usuels de statistique tels que BMDP, SPSS, SAS. Les plus connus sont VARIMAX (Kaiser, 1958) ou QUARTIMAX pour les transformations orthogonales, QUARTIMIN, OBLIMIN (Carroll, 1953,1957), OBLIMAX (Saunders, 1961), PROMAX (Hendrickson et White, 1964), direct OBLIMIN (Jennrich et Sampson, 1966) pour les transformations obliques. Cf. Clarkson et Jennrich, 1988 pour une revue.

Ce problème de «rotation» des facteurs est de faible importance. Il n'affecte en rien l'intérêt et les fondements de l'A.F.. Pourtant certains défenseurs de l'A.F. ne semblent connaître que ce problème et trouvent injustifiées les méfiances de certains statisticiens à l'égard de l'A.F. qu'ils pensent être dues à ce problème (cf. Everitt et Dunn, 1991). Cf. aussi Jackson, 1991, p. 396 : «Furthermore, factor analysis methodologie usually includes a rotation to simple structure...Therefore, even if a solution is determinate, it is only up to a rotation...there has been considerable furor about this practically since the beginning of F.A. and even at this point in time there is some controversy about the properties of this condition. Rather than dwell on this any further here, the reader may consult Steiger (1979) for more details of this problem. A pair of papers by Heermann (1964, 1966) are also of interest because they discuss both the geometric and algebraic concepts of indeterminacy.»

Il faut noter que les trois références citées concernent le problème d'indétermination des facteurs qui sera étudié au § 2.4..

Dans la mesure où, pour Ψ fixé, à partir d'une solution (Λ, f) de (2.1), on peut, grâce à des transformations A, obtenir toutes les solutions $(\Lambda A, A^{-1}f)$, on choisit dans la suite, sans perte de généralité, de chercher une solution (Λ, f) de (2.1) vérifiant

$$E(ff') = I \tag{2.3}$$

$$\Lambda'\Psi^{-1}\Lambda \text{ diagonale } (= \Gamma) \tag{2.4}$$

Ceci peut encore être traduit de la façon suivante. Si l'on pose :

$\mathcal{S} = \{(\Lambda, f, e) \mid (\Lambda, f, e) \text{ solution de (2.1)}\}$, alors on a :

$\mathcal{S} = \{(\Lambda A, A^{-1}f, e) \mid (\Lambda, f, e) \text{ sol. de (2.1),(2.3) et (2.4), } A\, q \times q \text{ régul.}\}$.

L'hypothèse (2.4) est souvent mal acceptée, considérée comme une facilité d'ordre mathématique. Elle est pourtant du même type que l'hypothèse (2.3). L'hypothèse (2.3) revient à choisir dans l'espace des facteurs communs une base orthonormée pour le produit scalaire L_2, l'hypothèse (2.4) revient à choisir dans E_q une base orthogonale pour la métrique Ψ^{-1} de $\mathbb{R}^p$.

Il aurait été équivalent de choisir l'hypothèse «$\Lambda'\Lambda$ diagonale» car il existe une matrice $A\, q \times q$ régulière faisant passer d'une solution Λ vérifiant $\Lambda'\Lambda$ diagonale à une solution Λ vérifiant (2.4).

On considère donc le modèle :

$$\Sigma = \Lambda\Lambda' + \Psi, \Psi \text{ diagonale}, \Lambda'\Psi^{-1}\Lambda (= \Gamma) \text{ diagonale}.$$

Remarque :

On a : $\Sigma\Psi^{-1} = \Lambda\Lambda'\Psi^{-1} + I$ et les q valeurs propres de $\Lambda\Lambda'\Psi^{-1}$, qui sont égales aux valeurs propres de $\Psi^{-1/2}\Lambda\Lambda'\Psi^{-1/2}$, sont strictement positives, aussi la matrice $\Sigma\Psi^{-1}$ a q valeurs propres strictement supérieures à 1 et la valeur propre 1 d'ordre de multiplicité $p - q$.

On a de plus : $\Sigma\Psi^{-1}\Lambda = \Lambda(\Gamma + I)$. La matrice $\Gamma + I$ est donc la matrice diagonale des valeurs propres strictement supérieures à 1 de $\Sigma\Psi^{-1}$ et Λ la matrice des vecteurs propres associés vérifiant la condition de normalisation $\Lambda'\Psi^{-1}\Lambda = \Gamma$.

Pour Ψ fixé, la matrice Λ peut donc être obtenue de façon unique à partir de la diagonalisation de $\Sigma\ \Psi^{-1}$, à l'orientation des vecteurs propres près et sous réserve que les valeurs propres différentes de 1 soient distinctes.

Il faut noter que l'hypothèse «Ψ diagonale» n'intervient pas dans cette remarque.

2.3 Problème d'identification du modèle de structure de covariance

On considère le modèle :

$$\Sigma = \Lambda\Lambda' + \Psi, \Psi \text{ diagonale}, \Lambda'\Psi^{-1}\Lambda (= \Gamma) \text{ diagonale.}$$

Soit r la différence entre le nombre de paramètres de Σ sans contraintes, c'est-à-dire le nombre d'équations $(= p(p+1)/2)$ et le nombre de paramètres libres du modèle de l'A.F. $(= pq + p - q(q-1)/2)$. On a : $r = [(p-q)^2 - (p+q)]/2$. On suppose qu'il n'existe pas de relation entre les paramètres ni entre les équations.

Si $r < 0$, il y a plus de paramètres que d'équations. Il existe donc une infinité de solutions. Le modèle est surparamétré.

Si $r = 0$, c'est-à-dire si $q = [2p + 1 - \sqrt{8p+1}]/2$, il existe une solution unique.

Si $r > 0$, il y a plus d'équations que de paramètres. Il n'existe pas de solution.

(Cf. Anderson et Rubin, 1956, pour une discussion sur ce sujet).

Il n'y a actuellement aucune méthode permettant de savoir s'il y a des relations entre les paramètres ou entre les équations et donc de connaître la valeur de r. On renvoie à Hakim *et al.* (1976) et Bekker et de Leeuw (1987) pour la recherche de la dimension q minimum telle que Σ-Ψ soit semi-définie positive de rang q.

Le premier cas étant évidemment à éviter, le deuxième étant impossible à réaliser de façon pratique, on se place en général dans le troisième cas. De plus, pour des raisons de parcimonie, il est conseillé d'utiliser peu de paramètres et donc de se placer dans cette situation. Il est alors nécessaire de chercher des solutions approchées.

Soit F une fonction de divergence («discrepancy function», Browne, 1974, 1982), c'est-à-dire, une fonction qui pour $\Sigma_i, i = 1, 2$, matrices $p \times p$ symétriques positives, vérifie :

$F(\Sigma_1, \Sigma_2) \geqslant 0$

$F(\Sigma_1, \Sigma_2) = 0$ si et seulement si $\Sigma_1 = \Sigma_2$

$F(\Sigma_1, \Sigma_2)$ deux fois continûment différentiable par rapport à Σ_1 et Σ_2.

Les exemples les plus utiles sont des fonctions construites à partir de critères d'estimation. Par exemple, si S désigne la matrice de covariance empirique obtenue à partir d'un échantillon $(y_i)_{i=1,\ldots,n}$ composé de variables indépendantes et de même distribution que y :

$F(S, \Sigma) = \text{Log} \mid \Sigma \mid - \text{Log} \mid S \mid + tr(S\Sigma^{-1}) - p$, critère du maximum de vraisemblance gaussien,

$F(S, \Sigma) = tr[(S - \Sigma)\Sigma^{-1}]^2$, critère des moindres carrés généralisés.

Soit F une fonction de divergence et (Λ_F, Ψ_F) la solution minimisant $F(\Sigma, \Lambda\Lambda' + \Psi)$, solution supposée unique.

Le modèle dépend à présent de la fonction de divergence utilisée, fonction qu'il est souhaitable ensuite d'utiliser dans la phase d'estimation.

En effet, soit $(\hat{\Lambda}_F, \hat{\Psi}_F)$ la solution minimisant $F(S, \Lambda\Lambda' + \Psi)$. Browne (1984) montre, sous certaines conditions de régularité, que c'est un estimateur consistant. L'efficacité asymptotique est étudiée dans des articles ultérieurs (cf. par exemple, Browne 1987).

Dans cet article, nous ne nous considérons que le *modèle* de l'A.F. Aussi, les problèmes d'estimation, qui ne font que rendre plus obscures les difficultés inhérentes au modèle lui-même ne sont pas exposés. De plus, la distinction entre modèle et estimation n'est pas toujours clairement faite. Cf. par exemple : Velicer et Jackson, 1990, p. 4 :

«Hence, when working with sample covariance matrix C, we fit the model : $C = AA' + U^2 + E$, where E is to be kept as small as possible by choosing A and U^2 to minimize a loss-function.»

2.4 Problème d'indétermination des facteurs

Afin d'aborder le problème d'indétermination des facteurs, il est d'usage de supposer le modèle de structure de covariance identifié, c'est-à-dire, (Λ, Ψ) défini de façon unique. (En fait on peut envisager le cas «$r > 0$» du paragraphe précédent, choisir une fonction de divergence F et supposer (Λ_F, Ψ_F) défini de façon unique. Les développements qui suivent peuvent être alors appliqués à condition d'admettre que la différence entre Σ et $\Lambda_F\Lambda_F' + \Psi_F$ est négligeable.)

Bien que (Λ, Ψ) soit défini de façon unique, le modèle de l'A.F. (2.1) admet une infinité de solutions.

En effet, soit P une matrice $q \times q$ telle que $PP' = I - \Lambda'\Sigma^{-1}\Lambda$.

L'ensemble des solutions du modèle (2.1) est :

$$f = \Lambda'\Sigma^{-1}x + Ps \text{ et } e = \Psi\Sigma^{-1}x - \Lambda Ps$$

pour tout s q-vecteur aléatoire vérifiant $E(s) = 0, E(xs') = 0$ et $E(ss') = I$. (Kestelman, 1952, Guttman, 1955).

Le vecteur f des facteurs communs est donc défini comme la somme d'un vecteur de composantes (*i.e.* de combinaisons de variables observables) $\Lambda'\Sigma^{-1}x$, noté C_f dans la suite, et d'une part arbitraire non nulle, Ps.

L'ensemble des solutions est sur un cône multidimensionnel dont l'axe (multidimensionnel aussi!) est le support du vecteur de composantes C_f. Pour $q = 1$, cf. figure 1.

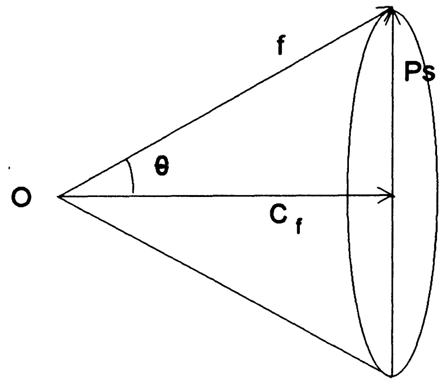

FIGURE 1

Ce problème d'indétermination des facteurs vient du fait que, vu les hypothèses du modèle, l'espace des facteurs communs et spécifiques est de dimension $p + q$ alors que l'espace des variables observables est de dimension p. On peut dire aussi que l'effet z étant supposé aléatoire, il n'est pas possible, vu les hypothèses du modèle, de le distinguer totalement du vecteur aléatoire e des facteurs spécifiques. En effet, le problème d'indétermination n'existe pas dans le cadre du modèle à effets fixes.

On ne connaît de façon unique que la part composante des facteurs, c'est-à-dire, C_f pour f et $\Psi\Sigma^{-1}x$ pour e. C_f est aussi appelée part de régression de f, mais on préfère réserver ce terme pour les composantes de régression proposées par Schönemann et Steiger, 1976, (cf. § 2.5.).

Historique.

Le problème de l'indétermination des facteurs a une longue histoire.

En 1927, Spearman présente un livre qui reprend 20 ans de travaux. A partir de nombreux tests psychologiques soumis à des individus, il espère trouver, grâce à l'Analyse en Facteurs qu'il propose (1 seul facteur commun), le facteur commun à tous ces tests qui serait interprété comme le facteur général d'intelligence (facteur g). Hart et Spearman (1912), présentant le modèle, envisagent même une société dans laquelle chaque citoyen aurait son facteur g mesuré et enregistré par le gouvernement. Les décisions concernant le droit de vote pourraient être prises à partir de cet indice!

En 1928, Wilson, ayant à reviewer le livre, montre par des exemples que le modèle n'admet pas une solution unique pour le facteur commun g.

Il s'ensuit, de 1928 à 1938 de nombreux articles et débats sur le sujet. Spearman (1933,1934) caractérise le problème d'indétermination des facteurs comme étant un problème d'échantillonnage de variables. L'erreur, selon Spearman, peut être réduite autant que l'on veut en ajoutant de plus en plus de variables de la même façon que l'erreur d'échantillonnage des individus est réduite en augmentant la taille de l'échantillon.

En plus des articles cités, on peut indiquer aussi Wilson 1929, Piaggio 1931,1933, Camp 1932,1934, Thompson 1934, 1935, Lederman 1938.

A partir de 1936 et grâce à la revue Psychometrika, l'Analyse en Facteurs, généralisée à plusieurs facteurs communs, est popularisée par Thurstone et son équipe qui occultent complètement le problème d'indétermination. A tel point que Wolfe en 1940, dans le cadre d'une revue historique, présente de nombreux développements de l'A.F. incluant les limitations de l'analyse sans mentionner l'existence de l'indétermination des facteurs.

Malgré la parution de quelques articles, Kestelman 1952, Guttman 1955, Heerman 1964, 1966, le débat sur l'indétermination des facteurs n'est relancé que depuis 1971 par Schönemann. Pourtant, après Kestelman (1952) et par une autre approche, Guttman en 1955 donne explicitement l'ensemble des solutions du modèle de l'A.F. ainsi que de nombreux développements encore d'actualité. Sa conclusion, «the Spearman-Thurstone approach may have to be discarded for lack of determinacy of its factor scores» pouvait laisser penser que le débat allait être relancé. Il n'en a rien été.

Depuis 1971, des articles paraissent régulièrement sur le sujet. Schönemann et Wang 1972, Mulaik 1972, 1976, 1981, Meyer 1973, Green, 1976, McDonald 1974, 1977, 1978, Mulaik et McDonald, 1978, Schönemann et Steiger, 1978, Steiger, 1979a, Williams 1978, Rozeboom 1982, 1988, Kano 1984, Haagen 1986, 1991, Schönemann et Haagen 1987, Bentler et Kano 1990. Il s'agit essentiellement de mesurer l'indétermination (cf.§ 2.5.) et de construire des variables supplémentaires pour réduire l'indétermination (on sort alors du cadre de l'A.F. définie pour p fixé).

Pour des éléments historiques on pourra consulter Steiger et Schönemann 1978, Steiger 1979b et Mulaik 1986.

Le sujet est aussi débattu dans les articles récents opposant les avantages et les inconvénients de l'A.C.P. et de l'A.F. : Borgotta, Kercher et Stull 1986, Hubbard et Allenn 1987, Snook et Gorsuch, 1989, Wilkinson 1989, Velicer et Jackson 1990, Schönemann 1990.

Bien plus nombreux sont les articles et les livres même récents ne mentionnant pas le problème. Parmi les livres on peut citer évidemment Thurstone 1935,1947, mais aussi Mardia et al (1979), Srivastava et Carter (1983), Anderson (1984b), Krzanowski (1988), Everitt et Dunn (1991).

Afin de contourner le problème d'indétermination des facteurs certains auteurs proposent des solutions alternatives au modèle de l'A.F. En 1962, Jöreskog propose de remplacer l'A.F. par l'«Image Factor Analysis», modèle s'inspirant de la théorie des images de Guttman (1953). En 1976 Schönemann et Steiger proposent de remplacer l'A.F. par une analyse en composantes de régression (cf.§ 2.5.). Bartholomew (1984, 1985a, 1987) développe une étude intitulée «les fondements de l'A.F.». Cf aussi les commentaires d'Aitkin 1985, Bentler 1985, Browne 1985, McDonald 1985b et la réponse de l'auteur Bartholomew 1985b.

La solution la plus communément adoptée, pour contourner le problème d'indétermination des facteurs est de ne s'intéresser qu'au modèle de stucture de covariance et d'abandonner l'idée d'«estimer» les facteurs. A partir de là se sont développés les modèles de structure de covariance, Bock et Bargmann 1966, Jöreskog 1970, 1978, 1981, McDonald 1978, 1980, 1985a, McArdle 1979, McArdle et McDonald 1984, Jöreskog et Sorbom 1984.

Le modèle en facteurs présenté ici est le modèle en facteurs de l'analyse dite exploratoire. Lorsque certains éléments de la matrice Ψ sont supposés nuls, le modèle est dit de régression avec erreurs sur les variables (Anderson, 1984a). C'est un modèle statistique de nature différente, les p variables ne jouent pas des rôles symétriques, et les problèmes mathématiques qu'il soulève sont plus simples. Lorsque certains éléments de Λ sont supposés nuls, hypothèses reposant sur des connaissances a priori sur les variables, on parle d'analyse en facteurs confirmatoire. Seule, la phase d'estimation des paramètres (Λ, Ψ) du modèle de structure de covariance (2.2) diffère de l'analyse en facteurs exploratoire. Le problème d'indétermination des facteurs demeure, contrairement à ce qui est souvent écrit en référence à McDonald et Mulaik, 1979.

Quelque soit la structure de covariance (cf. le logiciel Lisrel, Joreskog et Sorbom, 1984), dès lors que les hypothèses du modèle définissent les variables aléatoires latentes en dehors de l'espace engendré par les variables manifestes, le problème d'indétermination des facteurs apparaît lorsque que l'on veut prédire ces variables latentes. (Cf. Vittadini, 1989).

2.5 Equivalence des approximations usuelles de f

Ce paragraphe pourrait s'intituler «estimation des facteurs» (en anglais «factor score estimates») mais cette terminologie est source d'erreur. Tout d'abord, les facteurs étant aléatoires, il serait plus correct de parler de prédicteurs que d'estimateurs. Mais là n'est pas le principal reproche fait à l'usage de cette expression. Il réside dans le fait que cette expression sous-entend qu'il est possible d'estimer les facteurs. Or on a vu qu'il y a une infinité de solutions. Il faudrait donc proposer une infinité d'estimateurs (un pour chaque solution). De fait, il est proposé un seul estimateur pour l'ensemble des solutions.

De nombreux auteurs écrivent qu'il n'est pas possible d'écrire la solution f, qu'elle ne peut qu'être estimée. En fait, nous avons vu dans le paragraphe précédent qu'il est tout à fait possible d'écrire l'ensemble des solutions.

Evidemment, on peut arguer que le modèle postule l'existence d'une seule solution. Le problème d'indétermination réside alors dans le fait que la position de cette unique solution sur le cône est inconnue.

En tout état de cause, la critique essentielle faite à l'A.F. repose sur l'infinité de solutions du modèle (ou sur l'infinité de positions de la solution du modèle) et donc sur un arbitraire de la part de l'utilisateur quant au choix de la solution (ou de la position sur le cône de la solution).

Ce problème d'indétermination est de nature très différente du problème de rotation des facteurs communs. Il faut noter que, dans le cas d'un seul facteur commun, le problème de rotation n'existe plus alors que le problème d'indétermination des facteurs demeure.

La plupart des ouvrages déjà cités, réservant au moins un chapitre à l'A.F. et ne parlant pas du problème d'indétermination des facteurs traite cependant de l'estimation du vecteur des facteurs communs («factor score estimates»).

Nous verrons plus loin que l'arbitraire est en pratique considérablement réduit car tous les auteurs s'accordent à penser qu'on ne peut «estimer» les facteurs qu'à partir d'un vecteur de composantes.

D'après McDonald et Burr (1967), un «bon» prédicteur $\hat{f}$ de f est une fonction des variables d'échantillonnage qui devrait vérifier les propriétés suivantes :

a) être assez proche de f, par exemple «minimiser» $E((\hat{f}-f)(\hat{f}-f)')$;

b) avoir des composantes orthogonales comme f : $E(\hat{f}\hat{f}')$ diagonale;

c) avoir la propriété d'univocalité définie par Guilford et Michael (1948) : $E(\hat{f}f')$ diagonale;

d) être conditionnellement à f sans biais : $E(\hat{f} \mid f) = f$.

Afin de rester dans le cadre du modèle, on parlera d'approximation plutôt que de prédicteurs.

Une «bonne» approximation $\tilde{f}$ de f est un vecteur aléatoire, ne dépendant que du modèle, qui devrait vérifier les propriétés a) b) c) d) décrites ci-dessus.

En fait tous les prédicteurs proposés dans la littérature sont de «bons» prédicteurs de vecteurs aléatoires, qui sont en fait des composantes, (*i.e.* de la forme $C = Bx$). Ces composantes sont donc des approximations de f dont on se propose d'étudier les propriétés.

Les approximations usuelles de f

1) La première approximation proposée (Thurstone, 1935) est évidemment la part composante de f, c'est-à-dire :

$$C_f = \Lambda'\Sigma^{-1}x$$

Il est possible de montrer que C_f est la composante C qui minimise l'erreur quadratique moyenne : $E((C-f)(C-f)')$.

2) Si l'on considère f fixe et si l'on suppose x gaussienne alors x est gaussienne d'espérance Λf et de matrice de covariance Ψ. L'estimateur par maximum de vraisemblance de f est alors (Bartlett, 1937, 1938) :

$$C_B = Bx = (\Lambda'\Psi^{-1}\Lambda)^{-1}\Lambda'\Psi^{-1}x$$

Ici, B minimise $trE(\Psi^{-1/2}(x-\Lambda Bx)(x-\Lambda Bx)'\Psi^{-1/2})$ et donc $\Lambda\hat{f}$ est la projection Ψ^{-1}-orthogonale de x sur vec(Λ).

3) Dans une approche bayésienne, en supposant que la densité *a priori* de f est gaussienne d'espérance nulle et de matrice de covariance l'identité, on peut montrer que la densité *a posteriori* est proportionnelle à $\exp[-\frac{1}{2}(x-\Lambda f)'\Psi^{-1}(x-\Lambda f)-\frac{1}{2}f'f]$ et donc f est gaussienne d'espérance :

$$C_T = (I + \Lambda'\Psi^{-1}\Lambda)^{-1}\Lambda'\Psi^{-1}x$$

(Thompson 1939, Dolby 1976, Bartholomew 1981).

4) En 1956, Anderson et Rubin proposent $C_A = Bx$ avec B minimisant le critère de Bartlett sous la contrainte $E(C_A C'_A) = I$. On a alors :

$$C_A = (\Lambda'\Psi^{-1}\Sigma\Psi^{-1}\Lambda)^{-1/2}\Lambda'\psi^{-1}x$$

5) Schönemann et Steiger (1976) propose ce qu'ils appellent des composantes de régression :

$$C_S = (\Lambda'\Sigma^{-1}\Lambda)^{-1}\Lambda'\Sigma^{-1}x$$

Ces composantes qui auraient de meilleures propriétés que C_f sont définies à partir de ΛC_S, projection Σ^{-1}-orthogonale de x sur vec(Λ).

6) Dans son étude sur les fondements de l'A.F., Bartholomew (1984, 1985a, 1987) utilise également une approche bayésienne et aboutit pour approcher f à :

$$C_W = \Lambda'\Psi^{-1}x$$

7) Enfin nous proposons le vecteur des q premières composantes principales de l'A.C.P. centrée du vecteur aléatoire y dans $(\mathbb{R}^p, \Psi^{-1})$ (cf. 2.3. avec $M = \Psi^{-1}$), que nous notons C_P :

$$C_P = (\Lambda'\Psi^{-1}\Lambda)^{-1/2}\Lambda'\Psi^{-1}x$$

Propriétés

On considère le modèle :

$\Sigma = \Lambda\Lambda' + \Psi$, Λ $p \times q$ de rang q, Ψ régulière.

On pose $\Gamma = \Lambda'\Psi^{-1}\Lambda$.

Dans tout ce paragraphe Ψ et Γ ne sont pas supposées diagonales.

On vérifie aisément les propriétés suivantes (Guttman, 1955) :

(i) $\Sigma\Psi^{-1} = \Lambda\Lambda'\Psi^{-1} + I$

(ii) $\Sigma\Psi^{-1}\Lambda = \Lambda(\Gamma + I)$

(iii) $\Lambda'\Psi^{-1} = (\Gamma + I)\Lambda'\Sigma^{-1}$ et donc aussi $\Lambda'\Sigma^{-1} = (\Gamma + I)^{-1}\Lambda'\Psi^{-1}$

(iv) $\Lambda'\Psi^{-1}\Sigma\Psi^{-1}\Lambda = \Gamma(\Gamma + I)$

(v) $(\Gamma + I)(I - \Lambda'\Sigma^{-1}\Lambda) = I$

(vi) $\Lambda'\Sigma^{-1}\Lambda = \Gamma(\Gamma + I)^{-1}$

(vii) $\Sigma^{-1} = \Psi^{-1} - \Psi^{-1}\Lambda(\Gamma + I)^{-1}\Lambda'\Psi^{-1}$

Les approximations 1) à 7) définies précédemment vérifient les propriétés suivantes. A part celles concernant C_P, ces propriétés sont bien connues. Elles se déduisent aisément des propriétés (i) à (vii) ci-dessus.

$C_T = C_f$, $\quad C_W = (\Gamma + I)C_f$, $\quad C_B = C_S = \Gamma^{-1}(\Gamma + I)C_f$,
$C_A = \Gamma^{-1/2}(\Gamma + I)^{1/2}C_f$, $\quad C_P = \Gamma^{-1/2}(\Gamma + I)C_f$.

Propriétés de C_f et donc de C_T : $E(C_f f') = \Gamma(\Gamma + I)^{-1}$,
$E(C_f \mid f) = \Gamma(\Gamma + I)^{-1} f$,
$E(C_f C_f') = \Gamma(\Gamma + I)^{-1}$, $\quad E((C_f - f)(C_f - f)') = (\Gamma + I)^{-1}$.

Propriétés de C_W : $E(C_W f') = \Gamma$,
$E(C_W \mid f) = \Gamma f$,
$E(C_W C_W') = \Gamma(\Gamma + I)$, $\quad E((C_W - f)(C_W - f)') = \Gamma^2 - \Gamma + I$.

Propriétés de C_B et donc de C_S : $E(C_B f') = I$,
$E(C_B \mid f) = f$,
$E(C_B C_B') = \Gamma^{-1}(\Gamma + I)$, $\quad E((C_B - f)(C_B - f)') = \Gamma^{-1}$.

Propriétés de C_A : $E(C_A f') = \Gamma^{1/2}(\Gamma + I)^{-1/2}$,
$E(C_A \mid f) = \Gamma^{1/2}(\Gamma + I)^{-1/2} f$,
$E(C_A C_A') = I$, $\quad E((C_A - f)(C_A - f)') = 2I - 2\Gamma^{1/2}(\Gamma + I)^{-1/2}$

Propriétés de C_P : $E(C_P f') = \Gamma^{1/2}$,
$E(C_P \mid f) = \Gamma^{1/2} f$,
$E(C_P C_P') = \Gamma + I$, $\quad E((C_P - f)(C_P - f)') = \Gamma - 2\Gamma^{1/2} + 2I$.

Equivalence des approximations

On a vu au paragraphe 2.2. (Problème de «rotation» des facteurs communs), qu'à partir d'une solution f telle que $E(ff') = I$, on obtenait les autres solutions en transformant f par A matrice $q \times q$ régulière. A partir d'un prédicteur C de f, on obtient donc les prédicteurs des autres solutions en transformant C par A. A ce titre toutes les approximations définies précédemment sont équivalentes puisque l'on passe d'une approximation C à une autre par un tel type de transformation.

Toutes les approximations usuelles vérifient les propriétés b) et c) de Mc Donald et Burr si et seulement si Γ est diagonale. De plus, si Γ est diagonale, ces approximations sont toutes «proportionnelles» et ont donc même vecteur de composantes réduites, c'est-à-dire C_A, qui est aussi le vecteur des composantes principales réduites. Si l'on considère que la variance des facteurs communs (ou de leurs prédicteurs) est arbitraire, elle peut donc être posée égale à un et toutes les approximations sont équivalentes dans ce sens.

Bien que toutes les approximations soient équivalentes, elles sont toutes mauvaises. En effet, considérons une cinquième propriété proposée par Guttman (1955), que devrait vérifier une bonne approximation $\tilde{f}$ de f (ou un bon prédicteur $\hat{f}$ de f) :

si $\tilde{e} = x - \Lambda\tilde{f}$, la matrice $E(\tilde{e}\tilde{e}')$ devrait être diagonale régulière.

Or, pour toutes les approximations 1) à 7), cette matrice n'est ni diagonale, ni régulière puisqu'elle est de rang $p - q$.

Mesure de l'indétermination des facteurs

Si l'on revient à l'objectif de l'«estimation» des facteurs, il s'agit de prédire les valeurs de f sur les individus à partir des valeurs observées sur les variables initiales, ceci, afin de classer les individus.

Si l'on considère la figure 1, C_f donnera à peu près le même classement qu'un facteur f pris arbitrairement sur le cône seulement si l'angle θ est assez petit.

Il est donc important, avant d'utiliser une composante pour approcher f, de mesurer l'indétermination, c'est-à-dire le cosinus de l'angle que fait chaque facteur avec son approximation.

La matrice de corrélation de n'importe quelle composante 1) à 7) avec f est :

$$E(C_A f') = \Gamma^{1/2}(\Gamma + I)^{-1/2}.$$

Si Γ est diagonale, on a donc, pour tout $k = 1, ..., q$ et pour toute composante C :

$$\text{cor}(C^k, f^k) = (\gamma_k/(\gamma_k + 1))^{1/2} = \cos(\theta_k).$$

La contrainte $\Lambda'\Psi^{-1}\Lambda(= \Gamma)$ diagonale (en supposant les valeurs rangées dans l'ordre décroissant) permet de classer les facteurs communs dans l'ordre décroissant de leur mesure d'indétermination. En effet, $cos(\theta_k)$ diminue lorsque k augmente. De plus, cette contrainte Γ diagonale permet, à q fixé, de trouver les approximations les moins indéterminées.

Si l'on veut réduire l'indétermination des facteurs, il faut choisir q assez petit. Ce choix est malheureusement contradictoire avec la nécessité d'obtenir un bon ajustement du modèle; l'ajustement est d'autant meilleur que q augmente.

Choix du nombre de facteurs communs

Nous ne présenterons pas les nombreuses propositions qui sont faites. Notons seulement, qu'une condition naturelle découle du modèle : $\gamma_q > 1$. En effet, si $\cos(\theta_k) \leqslant \sqrt{2}/2$ (*i.e.* $\gamma_k \leqslant 1$) deux solutions distinctes correspondant au même facteur commun f^k peuvent être corrélées négativement ou non corrélées ! On peut se référer à la figure 1 pour $q = 1$ et à cette même figure pour chaque composante f^k dans le cas général.

Représentation des individus

Se poser la question de la représentation des individus, c'est revenir à la question de la variance des approximations. En effet considérons le cas $q = 2$ et Γ diagonale. On peut représenter les individus par rapport aux composantes réduites ou par rapport à d'autres composantes définies précédemment, qui ne diffèrent des composantes réduites que par le choix des variances. Implicitement, la pratique de la rotation des facteurs amène à choisir de représenter les individus par rapport aux composantes réduites.

3. Analyse en Composantes Principales

3.1 L'A.C.P. classique

Il y a différentes présentations possibles de l'A.C.P. (cf., par exemple, les références de Meredith et Millsap, 1985, Rao, 1964, ou Jolliffe, 1986, pour une revue récente sur l'A.C.P.). L'A.C.P. dite classique, proposée par Hotelling en 1933, est définie à partir des variables.

A.C.P. sur population

On considère un vecteur aléatoire $y = (y^1, ..., y^p)$ de $\mathbb{R}^p$ de variables dites observables, de moyenne $\mu = (\mu^1, ..., \mu^p)$ et de matrice de covariance Σ. Les variables peuvent être supposées de variance 1, Σ est alors la matrice des corrélations des variables. L'A.C.P. d'ordre q est la recherche de q variables aléatoires combinaisons linéaires des p variables observables (*i.e.* q composantes), non corrélées et de variance maximale. En fait, la première est de variance maximale, la deuxième est orthogonale à la première et de variance maximale, etc... Ces variables sont appelées composantes principales. Si L est la matrice diagonale des valeurs propres de Σ rangées dans l'ordre décroissant $l_1 \geqslant ... \geqslant l_p$ et V une matrice de vecteurs propres orthonormés associés alors le k-ème vecteur propre v^k définit la k-ème composante principale $c^k = \sum_{j=1}^p (y^j - \mu^j) v_j^k$. On a $\mathrm{cov}(c^k, c^{k'}) = 0, k \neq k'$ et $\mathrm{var}(c^k) = l_k$.

A.C.P. sur échantillon

On considère un échantillon $(y_i)_{i=1,\ldots,n}$ de vecteurs aléatoires de $\mathbb{R}^p$, indépendants et identiquement distribués (i.i.d.) comme y.

Soit $\bar{y} = n^{-1} \sum_{i=1}^{n} y_i$ et $S = n^{-1} \sum_{i=1}^{n} (y_i - \bar{y})(y_i - \bar{y})'$ la moyenne et la matrice de covariance empirique. L'A.C.P. sur l'échantillon est définie comme précédemment et conduit à la diagonalisation de S.

Convergence et statistique inférentielle

Lorsque la taille de l'échantillon augmente indéfiniment les différents éléments de l'A.C.P. sur l'échantillon convergent vers les éléments correspondants de l'A.C.P. sur la population. L'étude est réalisée par Anderson (1963) pour des variables gaussiennes, des résultats sont obtenus par Waternaux (1976) et Davis (1977) pour des variables dont on ne précise pas la distribution. Dauxois *et al* (1982) proposent une étude plus complète (valeurs propres multiples, projecteurs propres) dans un cadre plus large (A.C.P. d'une fonction aléatoire hilbertienne). On se trouve dans le cadre de la statistique classique. Il est donc possible, par exemple, de tester des hypothèses sur les valeurs propres et vecteurs propres de la matrice de covariance de la population.

Il est important de noter que l'on privilégie l'étude des variables. Les individus constituent un échantillon tiré de façon aléatoire dans une population dont on cherche à connaître les caractéristiques.

Cas de l'A.C.P. réduite

Dans le cas de l'A.C.P. réduite, les résultats sont obtenus à partir de la diagonalisation de la matrice de corrélation de l'échantillon. L'étude asymptotique pour variables gaussiennes et valeurs propres simples est réalisée de façon assez complète par Konishi en 1979. Dans un cadre plus général, variables aléatoires hilbertiennes de distribution quelconque, valeurs propres multiples, projecteurs propres, l'étude asymptotique de l'A.C.P. réduite est réalisée par Fine et Romain en 1984.

3.2 A.C.P. descriptive ou non stochastique

Il est possible de présenter l'A.C.P. sans référence à un modèle probabiliste (A.C.P. descriptive ou non stochastique). On insiste alors sur les propriétés algébriques et géométriques et sur le traitement simultané des individus et des variables. Cf. Cailliez et Pagès (1976), Lebart et al (1984), Saporta (1990).

Soit Y la matrice $n \times p$ des valeurs prises par n individus sur p variables quantitatives. Les individus peuvent être un échantillon d'une population plus vaste ou la population tout entière. Les lignes $y_i, i = 1, ..., n$, sont n vecteurs de $\mathbb{R}^p$, les colonnes $y^j, j = 1, ..., p$, sont p vecteurs de $\mathbb{R}^n$. On note de la même façon un vecteur et sa matrice dans la base canonique. Afin de représenter le nuage des individus (resp. des variables) on utilise une structure euclidienne. Un produit scalaire naturel sur $\mathbb{R}^n$ (et sa métrique associée) est donné par la matrice D des poids des individus. Dans le cas où les individus sont équipondérés on a $D = n^{-1}I_n$. Produit scalaire naturel car il fournit une interprétation géométrique des indices statistiques. Si $\bar{y}^j$ et $\bar{y}^k$ sont les moyennes des variables y^j et y^k et si x^j et x^k sont les variables centrées, alors le produit scalaire de x^j et x^k est la covariance des variables, la norme de x^j son écart-type. En revanche, il n'y a pas de choix naturel pour la métrique M sur $\mathbb{R}^p$ permettant de mesurer les proximités entre les individus.

Si l'on privilégie l'étude des individus (on les suppose ici équipondérés), l'A.C.P. d'ordre q peut être définie comme la recherche du sous-espace affine F_q de dimension q de $\mathbb{R}^p$ le plus proche possible du nuage des individus au sens des moindres carrés dans $(\mathbb{R}^p, M)$, c'est-à-dire, minimisant le critère :

$$\mathcal{F} = \frac{1}{n} \sum_{i=1}^{n} \| y_i - P_{F_q}^M y_i \|_M^2$$

où $P_{F_q}^M$ est le projecteur M-orthogonal sur F_q.

C'est l'approche proposée par Pearson en 1901 pour $q = 1$ et $M = I$. Cf aussi Gabriel, 1971, Kruskal, 1978. Pour l'introduction de la métrique M, cf. Dempster, 1969.

On note X la matrice des données centrées et $S = X'DX$ la matrice des covariances. Le sous-espace solution passe par $\bar{y}$ et est parallèle à l'espace engendré par les vecteurs propres associés aux q plus grandes valeurs propres de SM, c'est-à-dire à $[SM]_q$. Les individus sont alors représentés par les projections des y_i sur ce sous-espace, c'est-à-dire aussi par les projections des x_i sur $[SM]_q$, $i = 1, ..., n$. De façon duale, les variables sont représentées par des projections sur un sous-espace de dimension q de $\mathbb{R}^n$.

Si $M = TT'$ est une décomposition de Choleski de M, alors l'A.C.P. de (Y, M, D) et l'A.C.P. des données transformées par T avec métrique identité (YT, I, D) fournissent des résultats équivalents (mêmes valeurs propres et composantes principales, même représentation des individus).

Lorsque M est l'identité (resp. lorsque M est la matrice diagonale des inverses des variances), l'A.C.P. conduit à la diagonalisation de la matrice de covariance (resp. de corrélation). Les résultats de l'A.C.P. dépendent de l'échelle de mesure.

Lorsque les variables sont mesurées dans des unités différentes, il est d'usage de réduire les variables. Si elles sont mesurées dans la même unité, choisir de ne pas réduire les variables revient à pondérer les variables réduites par l'écart-type des variables avant réduction.

Une métrique intéressante et peu utilisée est la métrique $M = \mathrm{diag}(\Sigma^{-1})$, proposée par Jöreskog en 1962. Ce choix revient à pondérer les variables réduites par l'inverse de la racine carrée de $1 - r_j^2$ où r_j est le coefficient de corrélation multiple de la variable j avec l'ensemble des autres variables. On pondère ainsi d'autant plus les variables qu'elles sont mieux prédites par l'ensemble des autres variables. L'A.C.P. ainsi obtenue est invariante par changement de l'échelle de mesure.

3.3 Extension : A.C.P. d'un vecteur aléatoire de $(\mathbb{R}^p, M)$

Il est évidemment possible de définir l'A.C.P. d'ordre q d'un vecteur aléatoire y à valeurs dans $\mathbb{R}^p$ muni de la métrique M. Si $M = TT'$, c'est la recherche de q combinaisons linéaires des p variables observables transformées par T, non corrélées et de variance maximale. Les résultats sont obtenus à partir de la diagonalisation de ΣM (Σ étant la matrice de covariance de y).

3.4 Un modèle pour l'A.C.P. : le modèle à effets fixes.

Modèle à effets fixes

Bien souvent dans la pratique, les individus ne peuvent pas être considérés comme interchangeables et l'utilisateur de l'A.C.P., dans une approche exploratoire, s'intéresse à la représentation des individus et tente d'en dégager des structures.

Si l'on veut introduire une modélisation probabiliste, le modèle d'échantillonnage *i.i.d.* n'est plus acceptable puisqu'une structure particulière qui pourrait être observée ne serait due qu'à une fluctuation d'échantillonnage.

Caussinus (1986) propose alors de supposer que les lignes de la matrice de données sont des observations de vecteurs aléatoires vérifiant les hypothèses du modèle à effets fixes :

y_i, $i = 1, ..., n$, sont des p-vecteurs aléatoires indépendants, de même matrice de covariance Ψ régulière et d'espérances différentes z_i, $i = 1, ..., n$, supposées appartenir à un sous-espace affine F_q de dimension q de $\mathbb{R}^p$.

Ce modèle peut encore s' écrire :

$$y_i = z_i + e_i \tag{3.1}$$

z_i l'espérance de y_i, appelé effet fixe, appartient à un sous-espace de dimension réduite, e_i est une erreur aléatoire centrée, de matrice de covariance ne dépendant pas de i.

On pose $\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i$, $\Theta = \frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})(z_i - \bar{z})'$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $S = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})(y_i - \bar{y})'$ et $\Sigma = E(S)$.

On a alors la relation suivante sur les matrices de covariance :

$$\Sigma = \Theta + (1 - n^{-1})\Psi \tag{3.2}$$

avec rang $(\Theta) \leqslant q$. On suppose sans perte de généralité que rang $(\Theta) = q$.

On suppose tout d'abord que l'on a $\Psi = \sigma^2\Psi_0$ où Ψ_0 est une matrice régulière connue et σ un réel strictement positif inconnu. On ne fait aucune hypothèse sur la distribution des variables.

Estimation des paramètres

L'estimation des paramètres du modèle, F_q et z_i, $i = 1, ..., n$, par les moindres carrés dans $(\mathbb{R}^p, M)$ conduit à minimiser :

$$\mathcal{G} = \frac{1}{n}\sum_{i=1}^{n} \| y_i - z_i \|_M^2$$

La solution est alors $\hat{F}_q = \bar{y} + \hat{E}_q$ avec $\hat{E}_q = [SM]_q$ et $\hat{z}_i = P_{\hat{F}_q}^M y_i = \bar{y} + P_{\hat{E}_q}^M (y_i - \bar{y})$.

L'A.C.P. d'ordre q de (Y, M, D) fournit l'estimation des paramètres du modèle. C'est en ce sens que l'on considère que le modèle à effets fixes est un modèle pour l'A.C.P.. On peut en effet dans ce cadre répondre à des questions que se pose l'utilisateur. Par exemple : quelle métrique M utiliser? quelle dimension q choisir? Pour la réponse à cette dernière question nous renvoyons à Besse *et al* (1988), Ferré (1990) et Besse (1992).

Choix optimal de métrique : $M = \Psi^{-1}$

Dans le cadre ainsi proposé plusieurs études montrent que des propriétés d'optimalité amènent à considérer $M = \Psi^{-1}$.

- *Cas gaussien.* Notons tout d'abord que dans le cas où les variables sont supposées gaussiennes les estimateurs par maximum de vraisemblance des paramètres du modèle sont les estimateurs des moindres carrés avec $M = \Psi^{-1}$.

Ce résultat est encore valable lorsqu'on ne fait pas d'hypothèse sur la distribution des variables.

- *Cas σ^2 petit.* La qualité de l'ajustement peut être mesurée par

$$f(M, A) = \sum_{i=1}^{n} \| \hat{z}_i - z_i \|_A^2$$

où A est une métrique quelconque sur $\mathbb{R}^p$ non nécessairement égale à M. Pour σ suffisamment petit, en utilisant un développement limité par rapport à σ et certains résultats de théorie des perturbations, Besse *et al* (1988) montrent que, quelque soit A, le premier terme du développement de $f(M, A)$ est minimum pour $M = \Psi_0^{-1}(k > 0)$.

Le même résultat peut être montré à partir de propriétés asymptotiques et ceci, quelque soit σ.

– *Etude asymptotique du modèle à effets fixes.* Il est aisé de voir que le sous-espace affine du modèle à effets fixes est défini par

$$F_q = \bar{z} + E_q \text{ avec } E_q = Im(\Theta)$$

Comme on a : $\Sigma = \Theta + (1 - n^{-1})\sigma^2\Psi_0$, donc aussi :

$$\Sigma\Psi_0^{-1} = \Theta\Psi_0^{-1} + (1 - n^{-1})\sigma^2 I, \text{ on en déduit que } E_q = [\Sigma\Psi_0^{-1}]_q.$$

Or, on a vu que l'estimateur des moindres carrés de F_q est $\hat{F}_q = \bar{y} + \hat{E}_q$ avec $\hat{E}_q = [SM]_q$.

Sous des hypothèses assez faibles sur la convergence des suites $\bar{z}$ et Θ lorsque n tend vers l'infini, Fine et Pousse (1992) montrent que, quelque soit σ, le projecteur M-orthogonal sur $\hat{E}_q$ est consistant si et seulement si $M = k\Psi_0^{-1}$, $k > 0$.

– *Etude asymptotique du modèle à effets fixes dans le cas d'un nombre fini de sous-populations.* Dans l'étude asymptotique précédente on a autant d'observations que de sous-populations. On peut envisager le cas où on a un nombre fini K de sous-populations. L'étude asymptotique du modèle à effets fixes avec cette hypothèse et sous un modèle d'échantillonnage aléatoire simple conduit Fine (1993) à montrer que $P_{\hat{E}_q}^M$ est consistant pour tout M mais que, si L désigne l'opérateur de covariance asymptotique de $\sqrt{n}(P_{\hat{E}_q}^M - P_{E_q}^M)$, alors trace (L) est minimum si et seulement si $M = \Psi^{-1}$. C'est par un argument d'efficacité asymptotique que l'on obtient encore ici le même choix.

Il est possible, dans ce modèle d'échantillonnage, d'estimer Ψ par la matrice de covariance intra-classes de l'échantillon, ou encore, pour tenir compte du fait que la matrice de covariance inter-classes de la population est supposée de rang q, par une moyenne pondérée entre les matrices de covariance inter-classes et intra-classes de l'échantillon (cf. Anderson, 1984a). Quelque soit l'estimation de Ψ choisie, l'A.C.P. optimale, telle qu'elle est définie ici, est une Analyse Factorielle Discriminante (Cailliez et Pagès 1976, Lebart *et al* 1984, Saporta 1990).

Introduction des facteurs

L'hypothèse $z_i \epsilon F_q$, $i = 1, ..., n$, peut s'écrire :

il existe $\Lambda, p \times q, \mu, p \times 1$ et $f_i, q \times 1, i = 1, .., n, \frac{1}{n}\sum_{i=1}^{n} f_i = 0, \frac{1}{n}\sum_{i=1}^{n} f_i f_i' = I$ tels que : $z_i = \Lambda f_i + \mu$.

On a alors : $\mu = \bar{z}$ et $\Theta = \frac{1}{n}\sum_{i=1}^{n}(z_i - \mu)(z_i - \mu)' = \Lambda\Lambda'$.

Les estimations des $f_\imath, i = 1, ..., n$, par les moindres carrés dans $(\mathbb{R}^p, \Psi^{-1})$ sous la contrainte $\frac{1}{n}\sum_{i=1}^n f_\imath f_i' = I$ sont les observations des q premières composantes principales réduites de l'A.C.P. de (Y, Ψ^{-1}, D).

Rotation des composantes

En ce qui concerne les variables représentées de façon duale dans un sous-espace de $\mathbb{R}^{\ n}$ il est possible, comme en A.F., par des transformations régulières dans ce sous-espace de remplacer les composantes principales par des composantes plus facilement interprétables. Les méthodes sont celles utilisées pour la rotation des facteurs (cf. paragraphe 2.2.).

4. Conclusion

4.1 Pratique de l'Analyse en Facteurs

Finalement, la pratique de l'Analyse en Facteurs est exactement celle de l'estimation des paramètres du modèle à effets fixes suivant :

les lignes $y_\imath$, $i = 1, ..., n$, de la matrice $n \times p$ des données sont supposées être les observations de n vecteurs aléatoires de $\mathbb{R}^p$ de même matrice de covariance ψ diagonale régulière et d'espérances distinctes, $z_\imath, i = 1, ..., n$, (les effets fixes) pouvant s'écrire :

$z_\imath = \Lambda f_\imath + \mu, i = 1, ..., n$, avec $\Lambda, p \times q, \mu, p \times 1$ et $f_\imath, q \times 1, i = 1, ..., n$, vérifiant : $\displaystyle\frac{1}{n}\sum_{\imath=1}^n f_\imath = 0$ et $\displaystyle\frac{1}{n}\sum_{i=1}^n f_i f_i' = I$.

On a alors $\mu = \bar{z}$ et si on pose $\Sigma = E(S)$ on a : $\Sigma = \Lambda\Lambda' + (1 - n^{-1})\Psi$ avec Ψ diagonale régulière. Les corrélations observées dans S ne sont dues qu'à la corrélation entre les effets.

Pour $(\hat{\Lambda}, \hat{\Psi})$ fixé, les estimations des $f_\imath$, $i = 1, ..., n$, par les moindres carrés dans $(\mathbb{R}^p, \hat{\Psi}^{-1})$ sous la contrainte $\displaystyle\frac{1}{n}\sum_{i=1}^n f_\imath f_i' = I$ sont alors les observations des q premières composantes principales réduites de $(Y, \hat{\Psi}^{-1}, D)$.

En résumé, que ce soit dans le modèle à effets aléatoires (Analyse en Facteurs) ou dans le modèle à effets fixes, il faut distinguer le modèle lui-même (2.1) défini sur les variables aléatoires et le modèle de structure de covariance (2.2) :

Σ=$\Lambda\Lambda' + \Psi$ dans le modèle à effets aléatoires, $\Sigma = E(S) = \Lambda\Lambda' + (1 - n^{-1})\Psi$ dans le modèle à effets fixes.

La première étape, estimation de Λ et Ψ dans le modèle (2.2), est identique dans les deux modèles; c'est là qu'intervient l'hypothèse sur Ψ : Ψ régulière connue, $\Psi = \sigma^2\Psi_0$, Ψ_0 régulière connue, Ψ diagonale, régulière inconnue, ... avant d'étudier l'identification de ce modèle.

La deuxième étape est l'«estimation des facteurs» dans le modèle (2.1). Là, seule l'hypothèse «Ψ régulière» a son importance. Dans le modèle à effets aléatoires les facteurs sont indéterminés (infinité de solutions). Cependant, la pratique consiste à «estimer les facteurs communs», comme dans le cadre du modèle à effets fixes, par les q premières composantes principales réduites de l'A.C.P. de $(Y, \hat{\Psi}^{-1}, D)$.

4.2 Comparaison de l'A.C.P. et de l'A.F

Les deux méthodes, A.F. et A.C.P. ont les mêmes objectifs : réduction de dimension et étude des relations entre les individus et les variables. Ces objectifs sont clairement exprimés pour l'A.C.P. lorsqu'on fait ressortir la dualité entre les individus et les variables. Pour l'A.F. certains considèrent que son objectif est uniquement de reconstituer les corrélations entre les variables (Kruskal, 1978 par exemple). C'est en fait le premier objectif de l'A.F.; ceux qui considèrent qu'un objectif important est aussi d'estimer les facteurs s'intéressent de fait aux individus.

Les deux méthodes ont donc des objectifs assez semblables. Cependant, l'A.C.P. consiste à chercher q variables latentes (les composantes principales) reconstituant au mieux la somme des éléments diagonaux de la matrice de covariance (ou de corrélation). L'A.F. consiste à chercher q variables latentes (les facteurs communs) reconstituant au mieux les éléments non diagonaux de la matrice de covariance (ou de corrélation).

L'A.F. est invariante par changement d'échelle de mesure des variables. Ce n'est pas le cas de l'A.C.P., du moins sous sa forme usuelle (avec métrique identité sur l'espace des individus). En revanche, si l'on utilise l'A.C.P. sur des variables transformées linéairement, ou, ce qui est équivalent, si on définit l'A.C.P. avec une métrique de matrice M sur l'espace des individus différente de l'identité, certains choix de M conduisent à une A.C.P. invariante par changement de l'échelle de mesure, par exemple $M = \text{diag}(\Sigma^{-1})$.

Les «solutions» de l'A.C.P. sont emboîtées. C'est-à-dire que, pour passer de k composantes principales à $k+1$, il suffit de rajouter aux k premières composantes principales la composante de variance maximale orthogonale aux k premières composantes principales. Les «solutions» de l'A.F. ne sont pas emboîtées.

L'A.F. révèle plusieurs problèmes d'indétermination, le plus sévère étant l'indétermination des facteurs, c'est-à-dire que, quelque soient les contraintes utilisées, le modèle de l'A.F. liant les variables aléatoires manifestes aux variables aléatoires latentes a une infinité de solutions. Quelle est la bonne solution?

Nous n'avons pas abordé ici les problèmes d'estimation des paramètres du modèle de l'A.F., de nombreuses méthodes, de nombreux algorithmes, différents choix pour la solution initiale, problèmes de solutions impropres (Heywood cases). En particulier, les estimations vérifiant la contrainte $\Lambda'\Psi^{-1}\Lambda$ diagonale sont différentes des estimations vérifiant la contrainte $\Lambda'\Lambda$ diagonale. En revanche, la solution de l'A.C.P. est obtenue par une procédure directe.

L'A.F. repose sur un modèle ce qui n'est pas le cas de l'A.C.P.. Nous avons vu que le modèle à effets fixes permet d'apporter des réponses à certaines questions qui se posent dans la pratique de l'A.C.P.. Dans ce modèle n'apparaît pas de problèmes d'indétermination; c'est un modèle plus «léger» que le modèle à effets aléatoires de l'A.F..De plus, la pratique des modèles à effets fixes ou aléatoires et la pratique de l'A.C.P. avec métrique M sur $\mathbb{R}^p$ égale à Ψ^{-1} se rejoignent. A ce titre nous pouvons parler d'un rapprochement entre les deux méthodes.

Références

(1) AITKIN M., 1985, Comments on D.J. Bartholomew, Foundations of Factor Analysis : some practical implications, Brit. J. Math. Statist. Psych., 38, 127-128.

(2) ANDERSON T.W., 1963, Asymptotic theory for Principal Component Analysis. Ann. Math. Statist., 34, 122-148.

(3) ANDERSON T.W., 1984a, Estimating linear statistical relationships, Ann. Statist., 12, 1-45.

(4) ANDERSON T.W., 1984b, 2ème éd., An introduction to Multivariate statistical analysis, Wiley.

(5) ANDERSON T.W. et RUBIN M., 1956, Statistical inference in Factor Analysis. Proceed. 3rd Berkeley Symp. Math. Statist. Probab., 5, 11-150.

(6) BARTHOLOMEW D.J., 1981, Posterior analysis of Factor Model, Brit. J. Math. Statist. Psych., 34, 92-99.

(7) BARTHOLOMEW D.J., 1984, The Foundations of Factor Analysis, Biometrika, 71, 221-232.

(8) BARTHOLOMEW D.J., 1985a, Foundations of Factor Analysis : some practical implications, Brit. J. Math. Statist. Psych., 38, 1-10.

(9) BARTHOLOMEW D.J., 1985b, Reply to discussion on Foundations of Factor Analysis, Brit. J. Math. Statist. Psych., 38, 138-140.

(10) BARTHOLOMEW D.J., 1987, Latent variable models and Factor Analysis, Griffin.

(11) BARTLETT M.S., 1937, The statistical conception of mental factors, Brit. J. Psych., 28, 97-104.

(12) BARTLETT M.S., 1938, Methods of estimating mental factors, Nature, 141, 609-610.

(13) BEKKER P.A. et de LEEUW, J., 1987, The rank of reduced dispersion matrices, Psychometrika, 52, 125-135.

(14) BENTLER P.M., 1985, On the implications of Bartholomew's approach to Factor Analysis, Brit. J. Math. Statist. Psych., 38, 129-131.

(15) BENTLER P.M. et KANO, Y., 1990, On the equivalence of Factors and Components, Mult. Behav. research, 25, 67-74.

(16) BESSE P., 1992, PCA stability and choice of dimensionality, Statist. Probab. Letters, 13, 405-410.

(17) BESSE P., CAUSSINUS, H., FERRÉ, L., FINE, J., 1988, Principal Component Analysis and optimization of graphical displays, Statistics, 19, 301-312.

(18) BOCK R.D. et BARGMANN, R.E., 1966, Analysis of covariance structures, Psychometrika, 31,507-534.

(19) BORGOTTA E.F., KERCHER, K. et STULL, D.E., 1986, A cautionary note on the use of PCA, Sociol. Meth. Research, 15, 160-168.

(20) BROWNE M.W., 1967, On oblique procrustes rotation, Psychometrika, 32, 375-394.

(21) BROWNE M.W., 1974, Generalized least squares estimators in the analysis of covariance structures, South Afric. J., 8, 1-24.

(22) BROWNE M.W., 1982, Covariances structures, In Topics in Applied Multivariate Analysis, D.M. Hawkins (ed), Cambridge Univ. Press, 72-141.

(23) BROWNE M.W., 1984, Asymptotically distribution-free methods for the analysis of covariance structures, Brit. J. Math. Statist. Psych., 37, 62-83.

(24) BROWNE M.W., 1985, Comments on D.J. Bartholomew, Foundations of Factor Analysis : some practical implications, Brit. J. Math. Statist. Psych., 38, 132-133.

(25) BROWNE M.W., 1987, Robustness of statistical inference in Factor Analysis and related models, Biometrika, 74, 2, 375-384.

(26) CAILLIEZ F. et PAGÈS J.P., 1976, Introduction à l'Analyse des Données, SMASH, Paris.

(27) CAMP B.H., 1932, The converse of Pearson's two factor theorem, Biometrika, 24, 418-427.

(28) CAMP B.H., 1934, Spearman's factor again, Biometrika, 26, 260-261.

(29) CARROLL J.B., 1953, An analytic solution for approximating simple structure in factor analysis, Psychometrika, 18, 23-38.

(30) CARROLL J.B., 1957, Biquartimin criterion for rotation to oblique simple structure in factor analysis, Science, 126, 1114-1115.

(31) CAUSSINUS H., 1986, Models and uses of Principal Component Analysis, in Mult. Data Analysis, J. de Leeuw *et al* (eds), DSWO Press, Leiden, 149-170.

(32) CLARKSON D.B. et Jennrich, R.I., 1988, Quartic rotation criteria and algorithms, Psychometrika, 53, 251-259.

(33) DAUXOIS J., Pousse, A., et Romain, Y., 1982, Asymptotic theory for the Principal Component Analysis of a random vector function. J. Mult. Analysis, 12, 136-154.

(34) DAVIS A.W., 1977, Asymptotic theory for Principal Component Analysis : non normal case, Austr. J. Statist., 19, 206-212.

(35) DEMPSTER A.P., 1969, Elements of continuous multivariate analysis, Addison-Wisley Publishing Co.

(36) DOLBY G.R., 1976, Stuctural relations and Factor Analysis, The Math. Scientist, Supp.1, 25-29.

(37) EVERITT B.S. et DUNN G., 1991, Applied Multivariate Data Analysis, Arnold, E., (ed), London.

(38) FERRÉ L., 1990, A mean square error criterion to determine the number of components in generalized Principal Component Analysis, Preprint, Lab. Statist. Prob., Toulouse.

(39) FINE J., 1993, Asymptotic study of the functional models in the case of a random number of observations for each mean, Proposé pour publication.

(40) FINE J. et POUSSE A., 1992, Asymptotic study of the multivariate functional model. Application to the metric choice in Principal Component Analysis, Statistics, 23, 63-83.

(41) FINE J. et ROMAIN Y, 1984, Reduced Principal Component Analysis, Statistics, 15, 493-512.

(42) GABRIEL K.R., 1971, The biplot graphic display of matrices with application to PCA, Biometrika, 58, 453-467.

(43) GARNETT J.C., 1919, General ability, cleverness and purpose, Brit. J. Psych., 9, 345-366.

(44) GREEN B.F., 1952, The orthogonal approximation of an oblique structure in factor analysis, Psychometrika, 27, 429-440.

(45) GREEN B.F., 1976, On the factor score controversy, Psychometrika, 41, 263-266.

(46) GUILFORD J.P. et MICHAEL W.B., 1948, Approaches to univocal factor scores, Psychometrika, 13, 1-22.

(47) GUTTMAN L., 1953, Image theory for the structure of quantitative variates, Psychometrika, 18, 277-296.

(48) GUTTMAN L., 1955, The determinacy of factor score matrices with implications for five other basic problems of common factor theory, Brit. J. Statist. Psych., 8, 65-81.

(49) HAAGEN K., 1986, The indeterminacy problem in factor analysis and its consequences for prediction, Osterr. Zeit. Statist. Inform., 15, 197-205.

(50) HAAGEN K., 1991, Necessary and sufficient conditions for vanishing the arbitrariness in a common-factor model, The Statist., 40, 401-408.

(51) HAKIM M., LOCHARD E.Q., OLIVIER J.P., TEROUANNE E., 1976, Sur les traces de Spearman, I. Le problème des communautés, II. Le problème des rangs, Cahiers du BURO, 25.

(52) HART B. et SPEARMAN C., 1912, General ability, its existence and nature, Brit. J. Psych., 51-84.

(53) HEERMAN E.F., 1964, The geometry of factorial indeterminacy, Psychometrika, 29, 371-381.

(54) HEERMAN E.F., 1966, The algebra of factorial indeterminacy, Psychometrika, 31, 531-543.

(55) HENDRICKSON, A.E. et WHITE P.O., 1964, Promax : a quick method for rotation to oblique simple structure, Brit. J. Statist. Psych., 17, 65-70.

(56) HOTELLING H., 1933, Analysis of a complex of statistical variables into Principal Components, J. Educ. Psych., 24, 417-441, 498-520.

(57) HUBBARD R. et ALLENN S.J., 1987, A Cautionary note on the use of PCA; supportive empirical evidence, Sociol. Meth. Research, 16, 301-308.

(58) JACKSON J.E., 1991, A user's guide to Principal Components, Wiley.

(59) JENNRICH R.I. et Sampson, P.F., 1966, Rotation for simple loadings, Psychometrika, 31, 313-323.

(60) JOLLIFFE I., 1986, Principal Component Analysis, Springer-Verlag, New-York

(61) JÖRESKOG K.G., 1962, On the statistical treatment of residuals in Factor Analysis, Psychometrika, 27, 335-345.

(62) JÖRESKOG K.G., 1970, A general method for the analysis of covariance stuctures, Biometrika, 57, 239-251.

(63) JÖRESKOG K.G., 1978, Structural analysis of covariance and correlation matrices, Psychometrika, 43, 443-477.

(64) JÖRESKOG K.G., 1981, Analysis of covariance structures, Scand. J. Statist., 8, 65-92.

(65) JÖRESKOG K.G. et SORBOM, D., 1984, LISREL VI. A user's guide., 3°ed, Mooresville, In Scientific Software, Inc.

(66) KAISER H.F., 1958, The varimax criterion for analytic rotation in Factor Analysis, Psychometrika, 23, 187-200.

(67) KANO Y., 1984, Construction of additional variables conforming to a common factor model, Statist. Prob. Letters, 2, 241-244.

(68) KESTELMAN H., 1952, The fundamental equation of factor analysis, Brit. J. Psych. Statist., 5, 1-6.

(69) KONISHI S., 1979, Asymptotic expansions for the distributions of statistics based on the sample correlation matrix in Principal Component Analysis, Hiroshima Math. J., 9, 647-700.

(70) KRUSKAL J.B., 1978, Factor Analysis and Principal Components : Bilinear Methods, Intern. Ency. Statist. Kruskal, W.H. and Tanur, J.M. (eds), The Free Press, MacMillan, London.

(71) KRZANOWSKI W.J., 1988, Principles of Multivariate Analysis, Clarendon Press, Oxford.

(72) LEBART L., MORINEAU A. et WARWICK K.M., 1984, Multivariate Descriptive Statistical Analysis, Wiley, New-York.

(73) LEDERMANN W., 1938, The orthogonal transformation of a factorial matrix into itself, Psychometrika, 3, 181-187.

(74) MARDIA K.V., KENT J.T. et BIBBY J.M. 1979, Multivariate Analysis, Academic Press, London.

(75) McARDLE J.J., 1979, The development of general multivariate software, Proceed. Ass. Dev. Computer Based Instruct. Syst., Hirschbuhl (Ed), Univ. Akron, OH.

(76) McARDLE J.J. et McDONALD R.P., 1984, Some algebraic properties of the Reticular Action Model for moment structures, Brit. J. Math. Statist. Psych., 37, 234-251.

(77) McDONALD R.P., 1974, The measurement of factor indeterminacy, Psychometrika, 39, 203-222.

(78) McDONALD R.P., 1977, The indeterminacy of components and the definition of common factors, Brit. J. Math. Statist. Psych., 30, 165-176.

(79) McDONALD, R.P., 1978, A simple comprehensive model for the analysis of covariance structures, Brit. J. Math. Statist. Psych., 31, 59-72.

(80) McDONALD R.P., 1980, A simple comprehensive model for the analysis of covariance structures; some remarks and applications, Brit. J. Math. Statist. Psych., 33, 161-183.

(81) McDONALD R.P., 1985a, Factor Analysis and related methods, Lawrence Erlbaum Associates Publishers, London.

(82) McDONALD R.P., 1985b, Comments on D.J. Bartholomew, Foundations of Factor Analysis : some practical implications, Brit. J. Math. Statist. Psych., 38, 134-37.

(83) McDONALD R.P. et BURR E.J., 1967, A comparaison of four methods of constructing factor scores, Psychometrika, 32, 381-401.

(84) McDONALD R.P. et MULAIK, S.A., 1979, Determinacy of common factors : a non technical review, Psych. Bull., 86, 297-306.

(85) MEREDITH W. et MILLSAP R.E., 1985, On component analyses, Psychometrika, 50, 495-507.

(86) MEYER E.P., 1973, On the relationship between ratio of number of variables to number of factors and factorial indeterminacy, Psychometrika, 38, 375-378.

(87) MOSIER C.I., 1939, Determining a simple structure when loadings for certain tests are known, Psychometrika, 4, 149-162.

(88) MULAIK S.A., 1972, The foundations of factor analysis, McGraw-Hill, New-York.

(89) MULAIK S.A., 1976, Comments on "The measurement of factorial indeterminacy", Psychometrika, 41, 249-262.

(90) MULAIK S.A., 1981, A note on sufficient conditions that common factor will be determinate in an infinite domain of variables, Psychometrika, 46, 105-107.

(91) MULAIK S.A., 1986, Factor Analysis and Psychometrika : major developments, Psychometrika, 51, 23-33.

(92) MULAIK S.A. et McDonald, R.P., 1978, The effect of additional variables on factor indeterminacy in models with a single common factor, Psychometrika, 43, 177-192.

(93) PEARSON K, 1901, On lines and planes of closest fit to systems of points in space, Philos. Mag., 2, 559-572.

(94) PIAGGIO H.T.H., 1931, The general factor in Spearman's theory of intelligence, Nature, 127, 56-57.

(95) PIAGGIO H.T.H., 1933, Approximate general and specific factors without indeterminate parts, Brit. J. Psych., 25, 485-489.

(96) RAO C.R., 1964, The use and interpretation of Principal Component Analysis in applied research, Sankhya, A, 26, 329-358.

(97) ROZEBOOM W.W., 1982, The determinacy of common factors in large item domains, Psychometrika, 47, 281-295.

(98) ROZEBOOM W.W., 1988, Factor indeterminacy : the saga continues, Brit. J. Math. Statist. Psych., 41, 209-226.

(99) SAPORTA G., 1990, Probabilités, Analyse des Données et Statistique, Ed. Technip, Paris.

(100) SAUNDERS D.R., 1961, The rationale for an oblimax method of transformation in factor analysis, Psychometrika, 26, 317-324.

(101) SCHÖNEMANN P.H., 1966, The generalized solution for the orthogonal procrustes problem, Psychometrika, 31, 1-16.

(102) SCHÖNEMANN P.H., 1971, The minimum average correlation between equivalent sets of uncorrelated factors, Psychometrika, 36, 21-30.

(103) SCHÖNEMANN P.H., 1990, Facts, fictions, and common sense about Factors and Components, Mult. Behav. Research, 25, 47-51.

(104) SCHÖNEMANN P.H. et HAAGEN K., 1987, On the use of factor scores for prediction, Biometrical J., 29, 835-847.

(105) SCHÖNEMANN P.H. et Steiger, J.H., 1976, Regression Component Analysis, Brit. J. Math. Statist. Psych., 29, 175-189.

(106) SCHÖNEMANN P.H. et STEIGER J.H., 1978, On the validity of indeterminate factor scores, Bull. Psych. Society, 12, 287-290.

(107) SCHÖNEMANN P.H. et WANG M.M., 1972, Some new results on factor indeterminacy, Psychometrika, 37, 61-91.

(108) SNOOK S.C. et GORSUCH R.L., 1989, Principal Component Analysis versus Factor Analysis, a Monte Carlo study, Psych. Bull., 106, 148-154.

(109) SPEARMAN C., 1904, General intelligence, objectively determined and measures, Amer. J. Psych., 15, 201-293.

(110) SPEARMAN C., 1927, The abilities of man; their nature and measurement, MacMillan, New-York.

(111) SPEARMAN C., 1933, The uniqueness of «g», J. Educ. Psych., 24, 106-108.

(112) SPEARMAN C., 1934, The factor theory and its trouble, IV Uniqueness of g, J. Educ. Psych., 25, 142-153.

(113) SRIVASTAVA M.S. et Carter, E.M., 1983, An introduction to Applied Multivariate Statistics, North-Holland.

(114) STEIGER J.H., 1979a, The relationships between external variables and indeterminate factors, Psychometrika, 44, 93-97.

(115) STEIGER J.H., 1979b, Factor indeterminacy in the 1930's and the 1970's. Some interesting parallels. Psychometrika, 44, 157-167.

(116) STEIGER J.H. et SCHÖNEMANN P.H., 1978, A history of factor score indeterminacy, In Theory Construction and Data Analysis in the Behavioural Sciences, Skye (ed), San Francisco, Jossey-Bass.

(117) THOMPSON G.H., 1934, On measuring g and s by tests which break the g-hierarchy, J. Educ. Psych., 25, 204-210.

(118) THOMPSON G.H., 1935, The meaning of «i» in the measurement of «g» (general intelligence), J. Educ. Psych., 26, 241-262.

(119) THOMPSON G.H., 1939, 5° éd 1951, The Factorial Analysis of human ability, Houghton Mifflin, Boston.

(120) THURSTONE L.L., 1935, The vectors of mind : multiple factor analysis for the isolation of primary traits, Univ. Chicago Press, Chicago.

(121) THURSTONE L.L., 1947, Multiple Factor Analysis, Univ. Chicago Press, Chicago.

(122) VELICER W.F., et JACKSON D.N., 1990, Component Analysis versus Common Factor Analysis : some issues in selecting an appropriate procedure, Mult. Behav. Research, 25, 1-28.

(123) VITTADINI G., 1989, Indetermination problems in the LISREL model, Mult. Behav. Research, 24, 4, 397-414.

(124) WATERNAUX C.M., 1976, Asymptotic distribution of the sample roots for a non normal population, Biometrika, 63, 639-646.

(125) WILKINSON L., 1989, A cautionary note on the use of FA; a response to Borgotta, Kercher and Stull and Hubbard and Allen, Sociol. Meth. Research, 17, 449-459.

(126) WILLIAMS J.S., 1978, A definition for the common factor model and the elimination of problems of factor score indeterminacy, Psychometrika, 43, 293-306.

(127) WILSON E.B., 1928, Review of «The abilities of man, their nature and measurement» by C. Spearman, Science, 67, 244-248.

(128) WILSON E.B., 1929, Comments on Professor Spearman's note, J. Educ. Psych., 20, 217-223.

(129) WOLFE D., 1940, Factor Analysis to 1940, Psychometric Monographs, 3, Univ. Chicago Press, Chicago.