C.R.E.D.O.C.
BIBLIOTHÈQUE

C. R. E. D. O. C.

INTRODUCTION
A
L'ANALYSE FACTORIELLE .

Sou1969-2080

1969

Introduction à l'analyse factorielle / C. Deniau et L. Lebart. (Mars 1969).

IERCHES ET DE DOCUMENTATION SUR LA CONSOMMATION
ULEVARD DE LA GARE (13e) — TÉL. 707-97-59

CREDOC•Bibliothèque

Groupe de Travail Statistique

N° 1

I N T R O D U C T I O N

A

L'A N A L Y S E F A C T O R I E L L E

CREDOC
N° 210

C. DENIAU

L. LEBART

3328 - MARS 1969

R 6 1
(1er ex.)

SOMMAIRE

Pages

Première partie

-- Rappel de quelques notions d'algèbre linéaire 1

Deuxième partie

I - ANALYSE EN COMPOSANTES PRINCIPALES 38

I.1. Le problème pour l'utilisateur 38
I.2. Le problème pour le statisticien 39
I.3. Présentation des résultats 47
I.4. Exemple d'application 52

II - ANALYSE EN FACTEURS COMMUNS ET SPECIFIQUES 55

III - ANALYSE DANS UNE METRIQUE DEFINIE PAR UNE FORME QUADRATIQUE Q DEFINIE POSITIVE 62

III.1. Méthodes 62
III.2. Exemple d'application à l'analyse discriminante 65

IV. ANALYSE FACTORIELLE DES CORRESPONDANCES 70

IV.1. Le problème pour l'utilisateur 70
IV.2. Le problème pour le statisticien 72
IV.3. Représentation simultanée des ensembles I et J. 83
IV.4. Exemple d'application 83

V - BIBLIOGRAPHIE 86

- I N T R O D U C T I O N -

Les pages qui suivent constituent une première rédaction d'un exposé portant sur "l'analyse des données", et devraient être accessibles au plus grand nombre.

S'il est nécessaire d'avoir une certaine expérience statistique pour apprécier pleinement l'intérêt des méthodes employées, le langage mathématique d'un "bon bachelier" devrait suffire. En effet, dans une première partie, figurent des rappels d'algèbre linéaire.

Les diverses méthodes d'analyse exposées ici ont toutes le même but : réduire de façon synthétique et optimale une masse de données statistiques trop encombrante pour être exploitée telle quelle.

Les idées générales à la base de ces "réductions" sont anciennes, mais l'essentiel de celles exposées ici ont reçue un "certain bain de jouvence" dans le Laboratoire de J.P. Benzécri à la Faculté des Sciences de Paris.

Nous avons commis le péché de vulgarisation (!) ; nous espérons cependant n'avoir pas trop déformé la pensée des auteurs. Enfin, nous pensons que ce travail pourra aider certains utilisateurs à approfondir les idées présentées ici.

Nous remercions Marie-Noëlle Savoye, Béatrice Carréo et Claude Drillet pour la frappe et la "relecture" des stencils ainsi que pour l'assemblage et la présentation.

RAPPELS D'ALGEBRE LINEAIRE

1 - ESPACES VECTORIELS

Définition (1,1)

Un ensemble E pour lequel on suppose données une application de $E \times E \longrightarrow E$ $((x, y) \longmapsto x + y)$ et une application de $\mathbb{R} \times E \longrightarrow E$ $((\lambda, x) \longmapsto \lambda x)$ s'appelle un espace vectoriel réel (1) si :

1. Pour tout couple (x, y) d'éléments de E, on a : $x + y = y + x$
2. Pour tout triplet (x, y, z) d'éléments de E on a :
$$x + (y + z) = (x + y) + z$$
3. Il existe $e \in E$ tel que pour tout élément $x \in E$:
$$x + e = e + x = x$$
4. Pour tout $x \in E$, il existe $x_2 \in E$ tel que $x_1 + x_2 = e$
5. Pour tout couple (α, β) de nombres réels et tout $x \in E$ on a : $(\alpha + \beta)x = \alpha x + \beta x$
6. Pour tout $\alpha \in \mathbb{R}$ et tout couple (x, y) d'éléments de E on a :
$$\alpha(x + y) = \alpha x + \alpha y$$
7. Pour tout couple (α, β) d'éléments de $\mathbb{R}$ et tout $x \in E$ on a : $\alpha(\beta x) = (\alpha\beta)x$
8. Pour tout $x \in E$, on a $1x = x$.

Propriétés (1,1) : (simples à démontrer)

- Pour tout $x, y, z, \in E$: $x + y = y + z$ entraîne $x = z$
- Pour tout scalaire α et tout élément $x \in E$, on a $\alpha x = \alpha 0 = 0$
- La relation $\lambda y = 0$ est équivalente à $y = 0$ ou $\lambda = 0$
- Pour tout $x \in E$, l'élément $x' = (-1)x$ est l'unique élément de E tel que $x + x' = 0$.

exemples (1,1) : $\mathbb{R}$ est un espace vectoriel pour les applications :

$$(\alpha, \beta) \longmapsto \alpha + \beta$$
$$(\lambda, \mu) \longmapsto \lambda\mu$$

où $\alpha, \beta, \lambda, \mu \in \mathbb{R}$

---

(1) Quand on écrira espace vectoriel cela sous-entendra "réel".

- soit $\mathbb{R}^n = \left\{ z \; ; \; z = (x_1, \ldots, x_n), \; x_i \in \mathbb{R} \qquad 1 \leqslant i \leqslant n \right\}$

Si on se donne sur $\mathbb{R}^n$ les deux opérations suivantes :

$$(x_1, \ldots, x_n) + (x'_1, \ldots, x'_n) = (x_1 + x'_1, x_2 + x'_2, \ldots, x_n + x'_n)$$

$$\alpha(x_1, \ldots, x_n) = (\alpha x_1, \ldots, \alpha x_n)$$

$\mathbb{R}^n$ est un espace vectoriel, <u>produit des n espaces $\mathbb{R}$</u>

<u>Proposition</u> (1,1) :

Si $E_1, \ldots, E_n$ sont n espaces vectoriels et si on se donne sur $E_1 \times \ldots \times E_n$ les opérations définies ci-dessus, $E_1 \times \ldots \times E_n$ est un espace vectoriel.

<u>Définition</u> (1,2)

On dit qu'une partie non vide F de E est un sous-espace vectoriel si pour tout couple (x, y) d'éléments de F et tout couple (λ, μ) d'éléments de $\mathbb{R}$, on a :

$$\lambda x + \mu y \in F$$

<u>Proposition</u> (1,2) : (Très simple à démontrer)

Si E_1 et E_2 sont deux sous espaces vectoriels de E il en est de même de $E_1 \cap E_2$ (la proposition s'étend à une famille quelconque de sous espaces vectoriels).

Si E_1 et E_2 sont deux sous espaces vectoriels de E

$S = \left\{ x \in E \; ; \; x = x_1 + x_2 \quad x_1 \in E_1, \; x_2 \in E_2 \right\}$ est un sous espace vectoriel de E.

On le note $E_1 + E_2$ et on l'appelle espace vectoriel <u>somme</u> de E_1 et E_2

Définition (1,3)

Soient deux sous espaces vectoriels E_1 et E_2 d'un espace E. Les propriétés suivantes sont équivalentes :

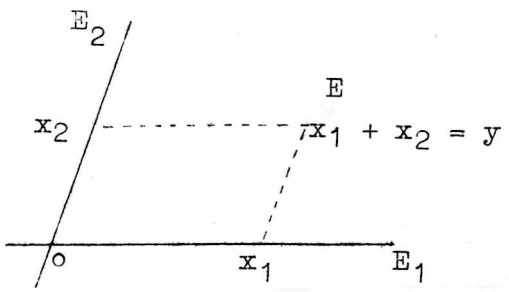
i/ $E_1 \cap E_2 = \{0\}$

ii/ Tout élément de $E_1 + E_2$ s'écrit d'une manière et d'une seule sous la forme $x_1 + x_2$ où $x_1 \in E_1$ et $x_2 \in E_2$.

On dit alors que la somme $E_1 + E_2$ est directe et on la note $E_1 \oplus E_2$ (on dit aussi que E_1 et E_2 sont supplémentaires dans E)

exemple (1,3)

$$E_1 \oplus E_2 = E \begin{cases} E_1 \cap E_2 = \{0\} \\ y \in E,\ y = x_1 + x_2 \end{cases}$$



Variété linéaire

Soit E un espace vectoriel réel et $a \in E$ et soit l'application :

$$t_a : E \longrightarrow E$$

$$t_a : x \longmapsto a + x$$

Definition (1,4)

Soient E_1 un sous espace de E et $a \in E$. L'image de E_1 par t_a (notée $a + E_1$) s'appelle une variété linéaire affine.

2 - APPLICATIONS LINEAIRES ET BILINEAIRES

Définition (2,1)

Soient E et F deux espaces vectoriels. On dit qu'une application f de E dans F est linéaire si :

$$f(\lambda x + \mu y) = \lambda f(x) + \mu f(y) \quad \text{quels que soient } x, y \in E, \lambda, \mu \in \mathbb{R}$$

(Montrer que $f(0) = 0$)

exemple (2,1)

Soient E un espace vectoriel et $a \in \mathbb{R}$. On définit une application h_a de E dans E par :

$$h_a : x \longmapsto ax$$

C'est une application linéaire.

Définition (2,2)

Soient E et F deux espaces vectoriels et f une application linéaire de E dans F

i/ On appelle image de f et on note $f(E)$ l'ensemble des $y \in F$ tels qu'il existe $x \in E$ vérifiant $y = f(x)$

ii/ On appelle noyau de f et on note $\mathrm{Ker}(f)$ l'ensemble des $x \in E$ tel que $f(x) = 0$

Propriétés (2,2)

a. $f(E)$ est un sous espace vectoriel de F et $\mathrm{Ker}(f)$ un sous espace vectoriel de E (facile à démontrer).

b. Une application linéaire bijective est appelée isomorphisme (si f est un isomorphisme, f^{-1} est aussi un isomorphisme).

Proposition (2,2)

Soient trois espaces vectoriels E, F, G, f une application linéaire de E dans F et g une application linéaire de F dans G. Alors $h = g \circ f$ est une application linéaire de E dans G.

3 - COMBINAISONS LINEAIRES, BASES, DIMENSION

Définition (3,1)

Soient $x_1, \ldots, x_n$ des éléments d'un espace vectoriel E ; on appelle combinaison linéaire de $x_1, \ldots, x_n$, tout élément $x \in E$ possédant la propriété suivante : il existe des scalaires $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ tels que

$$x = \lambda_1 x_1 + \ldots + \lambda_n x_n$$

On notera $x = \sum_{1 \leqslant i \leqslant n} \lambda_i x_i$

Exemple (3,1) soit $a \in \mathbb{R}^n$ $\quad a = (\lambda_1, \ldots, \lambda_n)$ $\lambda_i \in \mathbb{R}$ $\quad 1 \leqslant i \leqslant n$

considérons

$e_1 = (1, 0 \ldots, 0)$

$\ldots\ldots\ldots\ldots\ldots\ldots$

$e_i = (0, \ldots, 1, 0 \ldots 0)$ $\qquad a = \sum_{1 \leqslant i \leqslant n} \lambda_i e_i$

$\ldots\ldots\ldots\ldots\ldots\ldots$

$e_n = (0, \ldots, 0, 1)$

tout élément de $\mathbb{R}^n$ s'écrit comme combinaison linéaire de $e_1, \ldots e_n$. (On peut montrer que l'écriture est unique en remarquant que $(\xi_1, \ldots, \xi_n) = (\lambda_1, \ldots, \lambda_n)$ si et seulement si $\xi_i = \lambda_i$ $\quad 1 \leqslant i \leqslant n$).

Définition (3,2)

Soit E un espace vectoriel et $(x_i)_{1 \leqslant i \leqslant n}$ une famille d'éléments de E

i/ On dit que la famille $(x_i)_{1 \leqslant i \leqslant n}$ est libre si et seulement si :

$$\sum \lambda_i x_i = 0 \text{ entraîne } \lambda_i = 0 \qquad 1 \leqslant i \leqslant n$$

ii/ On dit que la famille $(x_i)_{1 \leqslant i \leqslant n}$ engendre E (ou est une famille génératrice de E) si tout élément $x \in E$ s'écrit comme combinaison linéaire d'éléments de la famille $(x_i)_{1 \leqslant i \leqslant n}$

iii/ On dit que la famille $(x_i)_{1 \leqslant i \leqslant n}$ est une base de E si tout élément $x \in E$ s'écrit de manière unique comme combinaison linéaire de la famille $(x_i)_{1 \leqslant i \leqslant n}$

Remarque (3,2)

On montrera à titre d'exercice que (i/, ii/) est équivalent à (iii/). On admettra que "tout espace vectoriel possède une base" et que si un espace vectoriel possède deux bases, elles ont même "nombre" d'éléments.

exemple (3,2)

Montrer que la famille $(\epsilon_i)_{1 \leqslant i \leqslant n}$ d'éléments de $\mathbb{R}^n$, définie dans l'exemple (3,1 page 5) est une base de $\mathbb{R}^n$.

Définition (3,3)

"Le nombre" d'éléments d'une base d'un espace vectoriel est appelé dimension de E. On notera $\dim(E)$. (Nous ne nous intéresserons qu'à des espaces vectoriels de dimension finie).

Définition (3,4)

Soient E et F deux espaces vectoriels et f une application linéaire de E dans F. On appelle rang de f la dimension de $f(E)$.

4 - ESPACE $\mathcal{L}(E, F)$

Proposition (4,1)

Soient E et F deux espaces vectoriels. L'ensemble des applications linéaires de E dans F pour lequel on suppose données les opérations :

$(f, g) \qquad f+g \quad$ [définie par $(f+g)(x) = f(x) + g(x)$ pour tout $x \in E$]

$(\lambda, f) \qquad \lambda f \quad$ [définie par $(\lambda f)(x) = \lambda f(x)$ pour tout $x \in E$]

est un espace vectoriel. On le notera $\mathcal{L}(E, F)$. (Démonstration triviale).

Définition (4,1)

L'ensemble des applications linéaires de E dans $\mathbb{R}$ est appelé l'espace dual de E et se note $E^* = \mathcal{L}(E, \mathbb{R})$.

Proposition (4,2)

Il existe une application linéaire unique f d'un espace vectoriel E dans un espace vectoriel F, telle que

$$\text{quel que soit } 1 \leqslant i \leqslant n \quad f(a_i) = b_i \qquad (1)$$

où $(b_i)_{1 \leqslant i \leqslant n}$ est une famille quelconque d'éléments de F et $(a_i)_{1 \leqslant i \leqslant n}$ une base de E.

Preuve :

. Unicité : supposons que f existe vérifiant (1)
$x \in E$ s'écrit de manière unique $x = \sum_{1 \leqslant i \leqslant n} \alpha_i a_i$

$(a_i)_{1\leqslant i\leqslant n}$ base de E, donc $f(x) = \sum \alpha_i f(a_i) = \sum \alpha_i b_i$

donc si f existe elle est unique.

. Existence : l'application $x \longmapsto \sum_{1\leqslant i\leqslant n} \alpha_i b_i$ de E dans F est linéaire. (Donc f est déterminée par ses valeurs sur les éléments de la base).

Proposition (4,3)

Soit E un espace vectoriel de dimension finie, alors $\dim(E) = \dim(E^*)$

Preuve :

Soit $(e_i)_{1\leqslant i\leqslant n}$ une base de E ; considérons la famille $(e_i^*)_{1\leqslant i\leqslant n}$ d'éléments de E^* tels que :

soient $x^* \in E^*$ et $x \in E$ $\quad x = \sum_{1\leqslant i\leqslant n} \lambda_i e_i$

$$x^*(x) = \sum \lambda_i x^*(e_i) \qquad (1)$$

Posons $x^*(e_i) = \lambda_i^* \quad 1 \leqslant i \leqslant n$

D'après la proposition (4,2), x^* est parfaitement déterminé.

Définissons une famille $(e^*)_{1\leqslant i\leqslant n}$ d'éléments de E^* définis par

$$e_j^*(x) = \lambda_j$$

ou ce qui revient au même, par $e_j^*(a_i) = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$

(1) s'écrit alors $x^*(x) = \sum_{1\leqslant i\leqslant n} \lambda_i^* e_i^*(x)$

d'où $x^* = \sum_{1\leqslant i\leqslant n} \lambda_i^* e_i^*$ quel que soit $x^* \in E^*$

donc $(e_i^*)_{1\leqslant i\leqslant n}$ engendre E^*.

Montrons que la famille $(e_i^*)_{1\leqslant i\leqslant n}$ est libre.

$$\sum_{1\leqslant i\leqslant n} \alpha_i\, e_i^* = 0 \Longrightarrow \text{quel que soit } x \in E \sum_{1\leqslant i\leqslant n} \alpha_i\, e_i^*(x) = 0$$

en particulier pour $x = a_j$

$$\sum \alpha_i\, e_i^*(a_j) = \alpha_i = 0 \quad \text{quel que soit} \quad 1 \leqslant i \leqslant n$$

ce qui achève la démonstration.

(par la suite on identifiera $\mathbb{R}^n$ à $(\mathbb{R}^n)^*$ la justification sera donnée par ailleurs).

5 - MATRICES

Soit f une application de E dont une base est $(a_i)_{1\leqslant i\leqslant m}$ dans F dont une base est $(b_j)_{1\leqslant j\leqslant n}$

$$f : E \longrightarrow F$$

$x \in E$ s'écrit $x = \sum_{1\leqslant i\leqslant m} \lambda_i\, a_i$

$$f(x) = \sum_{1\leqslant i\leqslant m} \lambda_i\, f(a_i) \qquad (1)$$

mais $f(x) \in F$, donc $f(x) = \sum_{1\leqslant j\leqslant n} \mu_j\, b_j$ (1')

Or, quel que soit i $1\leqslant i\leqslant n$ les $f(a_i) \in F$ donc :

$$\begin{aligned} f(a_1) &= b_1 \alpha_{11} + \ldots\ldots + b_n \alpha_{1n} \\ &\vdots \\ f(a_m) &= b_1 \alpha_{m1} + \ldots\ldots + b_n \alpha_{mn} \end{aligned} \qquad (2)$$

Remplaçons dans ① les $f(a_i)$ $1 \leqslant i \leqslant m$ par leurs expressions ①

③' $f(x) = (b_1 \alpha_{11} + \ldots\ldots + b_n \alpha_{1n}) \lambda_1 + \ldots\ldots +$
$(b_1 \alpha_{m_1} + \ldots\ldots + b_n \alpha_{nm}) \lambda_m$

Ecrivons que les seconds membres de ①' et ③ sont égaux et identifiés. On en déduit :

$$\left.\begin{array}{l} \mu_1 = \alpha_{11} \lambda_1 + \ldots\ldots + \alpha_{m_1} \lambda_m \\ \vdots \\ \mu_n = \alpha_{n1} \lambda_1 + \ldots\ldots + \alpha_{mn} \lambda_m \end{array}\right\} \quad ④$$

L'expression ④ détermine d'une manière et d'une seule l'application linéaire f par rapport aux bases $(a_i)_{1 \leqslant i \leqslant m}$ et $(b_j)_{1 \leqslant j \leqslant n}$ de E et F. Pour connaître f il suffit donc de connaître le tableau ⑤

$$A = \left(\begin{array}{ccc} \alpha_{11} & \ldots\ldots & \alpha_{m_1} \\ \vdots & & \vdots \\ \alpha_{n_1} & & m_n \end{array}\right) \quad ⑤$$

Un tel tableau s'appelle une <u>matrice</u> à m colonnes et n lignes. La ième colonne de la matrice A est formée avec les composantes de $f(a_i)$ dans ②

Ecriture matricielle d'une application linéaire

La relation ④ du § précédent : $\begin{array}{l} \mu_1 = \alpha_{11} \lambda_1 + \ldots\ldots + \alpha_{m_1} \lambda_m \\ \vdots \\ \mu_n = \alpha_{n_1} \lambda_1 + \ldots\ldots + \alpha_{mn} \lambda_m \end{array}$

où $f(x)$ a pour composantes $(\mu_1, \ldots\ldots, \mu_n)$
où x a pour composantes $(\lambda_1, \ldots\ldots, \lambda_m)$

s'écrit $\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \ldots\ldots & \alpha_{m1} \\ \vdots & & \vdots \\ \alpha_{n1} & \ldots\ldots & \alpha_{mn} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix}$

$\boxed{Y = \triangle X}$

Remarque (5,1)

2 | A est la matrice de f associée aux bases $(a_i)_{1\leqslant i\leqslant m}$ de E et $(b_j)_{1\leqslant j\leqslant n}$ de F.

exemple (5,1) $E = F = \mathbb{R}^2$ $\quad a = (a_1, a_2) \in \mathbb{R}^2$

Choisissons pour $\mathbb{R}^2$ la base $e_1 = (1, 0)$ et $e_2 = (0, 1)$

soit $h_a : \mathbb{R} \longrightarrow \mathbb{R}$ $\quad (h_a : x \longrightarrow ax)$

Construisons la matrice de h_a dans la base (e_1, e_2)

$$\left.\begin{aligned} h_a(e_1) &= a\,e_1 + o\,e_2 \\ h_a(e_2) &= o\,e_1 + a\,e_2 \end{aligned}\right\} \Longrightarrow A = \begin{pmatrix} a & o \\ o & a \end{pmatrix}$$

Si on choisit pour $\mathbb{R}^2$ la base $e'_1 = (1/2, 0)$, $e'_2 = (0, 1/2)$ alors $A = \begin{pmatrix} a/2 & o \\ o & a/2 \end{pmatrix}$

Notations :

- Les coordonnées par rapport à la base $(a_i)_{1\leqslant i\leqslant m}$ d'un élément $x \in E$

$$x = \sum_{1\leqslant i\leqslant m} \lambda_i a_i$$

se disposent en une matrice du type "colonne" $M(x) = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix}$

- Une forme linéaire f est déterminée par les $\left(f(a_i)\right)_{1\leqslant i\leqslant m}$ où $(a_i)_{1\leqslant i\leqslant m}$ est une base de E. Les $f(a_i) = \alpha_i$ sont disposés en une matrice du type "ligne" $M(f) = (\alpha_1 \ldots\ldots \alpha_n)$.

Transposée d'une application linéaire

Soient E et F deux espaces vectoriels et $f : E \longrightarrow F$ une application linéaire. Soit u une forme linéaire sur F ($u \in F^*$) ; alors, l'application composée $u \circ f$ est une forme linéaire sur E.

On peut définir une application linéaire ${}^t f : F^* \longrightarrow E^*$ en posant : ${}^t f(u) = u \circ f$ quel que soit $u \in F^*$

Définition

L'application ${}^t f$ s'appelle la <u>transposée</u> de l'application linéaire f.

Proposition (à vérifier)

i/ Soient E, F deux espaces vectoriels et f, g deux applications linéaires de E dans F, alors ${}^t(f+g) = {}^t f + {}^t g$

ii/ Soit G un autre espace vectoriel et $f : E \longrightarrow F$, $h : F \longrightarrow G$ deux applications linéaires :

$$ {}^t(f \circ g) = {}^t g \circ {}^t f $$

Quelques éléments de calcul matriciel

Une matrice A à m colonnes et n lignes est appelée matrice de type (m, n) et notée :

$$A = (a_{i,j})_{\substack{1 \leqslant i \leqslant m \\ 1 \leqslant j \leqslant n}} \quad \text{ou} \quad A\ (m, n)$$

(Si $m = n$ la matrice est carrée)

Définition (5,1)

On appelle transposée de la matrice $A = (a_{ij})$ la matrice $B = (b_{ij})$ définie par $b_{ij} = a_{ji}$; on la notera $B = {}^t A$

exemple : ${}^t(x_1, \ldots\ldots, x_n) = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

(le vérifier à partir de la définition de la transposée d'une application linéaire).

Définition (5,2)

i/ On dit qu'une matrice A est symétrique si ${}^t A = A$

ii/ On dit qu'une matrice D est diagonale si $a_{ij} = 0$ pour $i \neq j$

iii/ On dit qu'une matrice A est antisymétrique si ${}^t A = -A$

(où $-A = (-a_{ij})_{\substack{1 \leqslant i \leqslant m \\ 1 \leqslant j \leqslant n}}$)

Remarque (5,2)

On vérifiera qu'une matrice symétrique (ou une matrice antisymétrique) est nécessairement carrée.

Addition des matrices (5,3)

Soient E et F deux espaces vectoriels réels, et f, g deux applications linéaires de E dans F. Choisissons $(a_i)_{1\leqslant i\leqslant m}$ et $(b_j)_{1\leqslant j\leqslant n}$ pour les bases respectives de E et F

$$A = (\alpha_{ij}) \quad \text{matrice de } f$$

$$B = (\beta_{ij}) \quad \text{matrice de } g$$

$$f(a_i) = b_1 \alpha_{i1} + \dots + b_n \alpha_{in}$$

$$g(a_i) = b_1 \beta_{i1} + \dots + b_n \beta_{in}$$

Posons $h = f + g$

Les éléments de la ième colonne de la matrice associée à h seront les composantes de $h(a_i)$, donc si on appelle $C = (\gamma_{ij})$ la matrice associée à a h :

$$\gamma_{ij} = \alpha_{ij} + \beta_{ij}$$

$1 \leqslant i, j \leqslant n$

et on écrit $C = A + B$

Définition (5,3)

Soient E, F et G trois espaces vectoriels ayant pour base respective $(a_i)_{1\leqslant i\leqslant m}$, $(b_j)_{1\leqslant j\leqslant n}$, $(c_k)_{1\leqslant k\leqslant p}$ et deux applications linéaires f et g

$$g : E \longrightarrow F \qquad f : F \longrightarrow G$$

soit $h = f \circ g$

Désignons par $A = (\alpha_{jk})_{\substack{1\leqslant j\leqslant n \\ 1\leqslant k\leqslant p}}$ $B = (\beta_{ij})_{\substack{1\leqslant i\leqslant m \\ 1\leqslant j\leqslant n}}$ et $C = (\gamma_{ik})_{\substack{1\leqslant i\leqslant m \\ 1\leqslant j\leqslant p}}$

les matrices associées à f, g et h
Vérifier à titre d'exercice que

$$\gamma_{ik} = \sum_{1 \leqslant j \leqslant n} \beta_{ij}\, \alpha_{jk}$$

On dit alors que C est la <u>matrice produit</u> de A et B et on note C = AB.

<u>Remarque</u> (5,4)

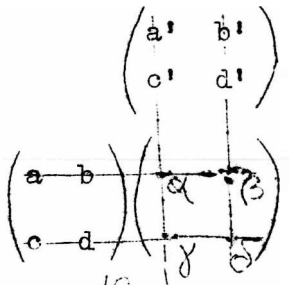
Le produit n'est défini que si le nombre de colonnes de A est égal au nombre de lignes de B (le vérifier)

<u>Recette</u> :

. $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$

$B = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}$

Calculer C = AB



$\alpha = aa' + bc'$

$\beta = ab' + bd'$

etc...

. Soit $A = (\alpha_1, \ldots\ldots, \alpha_m)$ $\quad B = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}$

$$A\,B = \alpha_1 \beta_1 + \ldots + \alpha_m \beta_m$$

. Soient $A = (\alpha_{ij})_{\substack{1 \leqslant i \leqslant m \\ 1 \leqslant j \leqslant n}}$ et $\lambda \in \mathbb{R}$

$$\lambda A = (\lambda \alpha_{ij})_{\substack{1 \leqslant i \leqslant m \\ 1 \leqslant j \leqslant n}}$$

Propriétés (5,5)

Soient A, B, C des matrices ; lorsque ces opérations sont définies :

$(AB)C = A(BC) = A\,B\,C$

$A\,(B + C) = AB + AC$

$A\,(\lambda B) = \lambda\, AB \qquad \lambda \in \mathbb{R}$

$AB \neq BA$

${}^t(A + B) = {}^tA + {}^tB$

${}^t(AB) = {}^tB\ \ {}^tA$

${}^t({}^tA) = A$

Définition (5,4)

Soit A une matrice carrée, on appelle trace de $A = (\alpha_{ij})_{1 \leqslant i, j \leqslant m}$ et on note tr (A) la somme des éléments de la forme α_{ii} (éléments diagonaux de A)

$$\text{tr } A = \sum_{1 \leqslant i \leqslant m} a_{ii}$$

Propriétés (5,6) (à vérifier)

$\text{tr } (A + B) = \text{tr } (A) + \text{tr } (B)$

$\text{tr } ({}^tA) = \text{tr } (A)$

$\text{tr } (\lambda A) = \lambda\, \text{Tr } (A)$

$$\text{tr } (AB) = \text{tr } (BA) = \sum_{\substack{1 \leqslant i \leqslant m \\ 1 \leqslant j \leqslant n}} a_{ij}\ b_{ji}$$

La matrice carrée (m, m) (matrice carrée d'ordre m) dont tous les éléments diagonaux sont égaux à 1 et dont tous les autres sont nuls

est notée

$$I_m = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

On a $IA_m = AI_m = A$ et $tr(I_m) = m$

Définition (5,5)

On dit qu'une matrice carrée A (m, m) est inversible si et seulement si il existe B (m, m) telle que

$$A\,B = B\,A = I_m$$

B est alors appelée matrice inverse de A et notée A^{-1}.

Définition (5,6)

On appelle rang d'une matrice le rang de l'application linéaire qu'elle représente.

6 - FORMES BILINEAIRES, FORMES QUADRATIQUES

Définition (6,1)

Soient E, F deux espaces vectoriels et f une application de E dans F ; f est bilinéaire si elle est linéaire par rapport à chacune des variables (c'est-à-dire), si $x \longrightarrow f(x, y)$ et $y \longrightarrow f(x, y)$ sont linéaires).

C.R.E.D.O.C.
BIBLIOTHÈQUE

INTRODUCTION A L'ANALYSE FACTORIELLE

C.DENIAU - L.LEBART - 3328 - mars 1969

E R R A T A

Page

1 Lignes 8, 9 et 10 lire : 3. Il existe $0 \in E$ tel que pour tout élément $x \in E$ $\quad x+0 = 0+x = x$

4. Pour tout $x_1 \in E$, il existe $x_2 \in E$ tel que $x_1 + x_2 = 0$

7 Définition (4,1) ajouter (un élément de E^* est appelé : forme linéaire sur E)

8 Ligne 17, lire : ou ce qui revient au même, par $e^*_j(e_i) = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$

9 Ligne 15, lire : or, quel que soit i, $1 \leqslant i \leqslant n$ les $f(a_i) \in F$ donc :

10 Lignes 1, 2 et 3 lire : remplaçons dans (1) les $f(a_i)$ $1 \leqslant i \leqslant m$ par leurs expressions (2) alors $f(x) = (b_1 \alpha_{11} + \ldots\ldots + b_n \alpha_{1n}) \lambda_1 + \ldots + (b_1 \alpha_{m1} + \ldots + b_n \alpha_{mn}) \lambda_m$ (3)

Ligne 7 lire : $\mu_n = \alpha_{1n} \lambda_1 + \ldots\ldots + \alpha_{mn} \lambda_m$

Ligne 12 lire :

$$A = \begin{pmatrix} \alpha_{11} & \ldots\ldots & \alpha_{m1} \\ \vdots & & \\ \alpha_{1n} & \ldots\ldots & \alpha_{mn} \end{pmatrix}$$

Ligne 17 lire : $\mu_n = \alpha_{1n} \lambda_1 + \ldots\ldots + \alpha_{mn} \lambda_m$

Ligne 20 lire :

$$\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \ldots\ldots & \alpha_{m1} \\ \vdots & & \vdots \\ \alpha_{1n} & \ldots\ldots & \alpha_{mn} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix} \qquad (6)$$

Ligne 21, supprimer $\boxed{Y = AX}$

CREDOC
N° 210

R6 1
(1er ex.)

Page

11 | Ligne 4, ajouter : au départ et à l'arrivée

Ligne 9, ajouter : au départ

Ligne 14, lire : se disposent en une matrice du type colonne notée

$$X \text{ ou } \mu(x) \quad \text{avec} \quad X = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix}$$

bas de la page, ajouter : enfin posons

$$Y = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

l'expression ⑥ s'écrit $\boxed{Y = AX}$

11bis Ligne 14, lire : $g \quad F \longrightarrow G$ deux applications linéaires

17 | Ligne 4, lire $f(x, y) = \sum_{\substack{1 \leqslant i \leqslant m \\ 1 \leqslant j \leqslant n}} f(a_i b_j) x_i y_j$

Ligne 7, lire :

$$f(x, y) = (y_1 \ldots\ldots y_n) \begin{pmatrix} \alpha_{11} & \ldots\ldots & \alpha_{m1} \\ \alpha_{1n} & \ldots\ldots & \alpha_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$$

21 | Ligne 7, lire : $\|\lambda x\| = |\lambda| \, \|x\| \qquad \lambda \in \mathbb{R}$

22 | Ligne 4, lire : i/ $d(x, y) = 0$ si et seulement si $x = y$

Ligne 11, lire : $d(\lambda x, \lambda y) = |\lambda| \, d(x, y)$

23 | Ligne 18, lire : donc

$f(x_1, \ldots, x_i, \ldots, x_j, \ldots, x_n) = - f(x_1, \ldots, x_j, \ldots, x_i, \ldots, x_n)$

bas de la page, ajouter : $f(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) = \varepsilon_\sigma f(x_1, \ldots, x_n)$

où $\varepsilon_\sigma = \pm 1$ suivant la parité de k

Page

24 Ligne 14, lire : $f(x_1, \ldots, x_n) = \sum_{\sigma \in \mathfrak{S}_n} \varepsilon\sigma \; \lambda_{\sigma(1)} \cdots$

$$\lambda_{n\,\sigma(n)} f(a_1, \ldots, a_n) \qquad (1)$$

supprimer la ligne suivante.

25 Ligne 14, lire : .... est celle des formes n - linéaires alternées.....

26 Ligne 16, lire $A_{i,j}$ est appelé le <u>mineur</u> associé à λ_{ij} et $(-1)^{i+j}A_{ij}$ est appelée le <u>cofacteur</u> associé à λ_{ij}

28 Ligne 14, ajouter : si $i = j$
Ligne 15, remplacer $(i \neq j)$ par si $i \neq j$

30 Propositio (9,4), lire :

Si λ est une valeur propre de A et si $r \in \mathbb{N}^*$, alors λ^r est une valeur propre de A^r ; si A est inversible, alors λ^{-r} est valeur propre de A^{-r}.

- Si $AX = \lambda X$, alors $A(AX) = A(\lambda X) = \lambda AX = \lambda^2 X$ d'où par récurrence sur r $A^r X = \lambda^r X$
- Si $AX = \lambda X$, alors $\lambda^{-1}A^{-1}AX = \lambda^{-1}A^{-1}\lambda X$ et $\lambda^{-1}X = A^{-1}X$ d'où par récurrence sur r $A^{-r}X = \lambda^{-r}X$.

32 Exemple, lire :

$$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

34 Ligne 14, lire : de $D = {}^tPAP$ où P est la matrice orthogonale

35 Ligne 4, lire : Soit $Y = {}^tPX$ telle que
Ligne 7, lire : $PY = X = 0$; donc A est définie positive
Ligne 9, lire : propre λ_1 soit non positive
Ligne 11, lire : mais ${}^tX'AX = {}^tY, {}^tPAPY' = {}^tY'DY = \lambda_1 \leqslant 0$

Page

41 Ligne 8, lire : soient x_k, x_e deux points observation du nuage.

Lignes 11 et 12, lire : la grandeur de la projection $B_k B_e$ est le produit scalaire du vecteur $x_k - x_e$ par le vecteur u.

43 Ligne 5, lire : son terme général s'écrit :

$$\vartheta_{ij} = \sum_{k,e} (x_{ki} - x_{ei})(x_{kj} - x_{ej}) \qquad (1)$$

44 Ligne 1, lire : $\vartheta_{ij} = 2n\left(\sum_k x_{ki}x_{kj} - 1/n \sum_e x_{ei} \sum_k x_{kj}\right)$

Ligne 18, lire $V = R \Delta^t R$

49 Ligne 7, lire : Ainsi, le coefficient $c(i,k)$ s'écrit : $c(i,k) = u_{ki} \cdot \sqrt{\lambda_k}$

62 Ligne 1, lire : Soient $x_1, \ldots, x_p$ des éléments de $\mathbb{R}^p$

Ligne 9, lire : Posons $y_i = Cx_i \qquad 1 \leqslant i \leqslant p$

Ligne 13, lire : et avec $\vartheta_i = \vartheta(a_i) \qquad 1 \leqslant i \leqslant p$ où les a_i sont les éléments

Ligne 16, lire : c'est une application linéaire "projection sur $\mathbb{R}$"

63 Ligne 5, lire : $\vartheta = (\vartheta_{ij})$ avec $\vartheta_{ij} = \sum_{1 \leqslant k,e \leqslant p} (y_{ik} - y_{ie})(y_{jk} - y_{je})$

Ligne 7, ajouter : où V est définie par l'expression (1) page 43.

64 Ligne 14, lire : soit ${}^t\vartheta Q^{-1} \vartheta = 1$

67 Ligne 14, Lire : ${}^t x\, V^{-1} x = k$ (k = constante positive)

70 Ligne 3, lire (en fin de ligne) : $(i, j, k) \in I \times J \times K$

77 Ligne 2, lire : en faisant intervenir les poids $p(i)$, $i \in I$

81 Ligne 2, lire :

$$\sum_i \left\{ \sum_j \frac{p(i,j) \cdot r(e,j)}{p(e) \cdot p(j)} \right\} \sum_k \frac{p(i,k)}{p(i)} \vartheta_k = \lambda \sum_j \frac{p(e,j)}{r(e)} \vartheta_j$$

15 SEP. 1970

Matrice associée à une forme bilinéaire

(a_i) base de E $\qquad x = \sum_{1 \leqslant i \leqslant m} a_i x_i$

(b_j) base de F $\qquad y = \sum_{1 \leqslant j \leqslant n} b_j x_j$

$$f(x, y) = \sum_{1 \leqslant i, j \leqslant n} f(a_i\ b_j)\ x_i\ y_j$$

(On laisse le soin au lecteur de démontrer que les $f(a_i\ b_j) = \alpha_{ij}$ définissent complètement f)

On écrira $f(x, y) = (y_1, \ldots\ldots y_n) \begin{pmatrix} \alpha_{11} & \ldots\ldots & \alpha_{n1} \\ \vdots & & \\ \alpha_{1n} & \ldots\ldots & \alpha_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

soit en écriture matricielle $F = {}^t Y\ A\ X$

Mais ${}^t Y\ A\ X$ est un scalaire, donc ${}^t({}^t YAX) = {}^t YAX$

donc $\boxed{{}^t X\ {}^t A\ Y = {}^t Y\ A\ X}$

si le rang de A est égal à r, f est dite de rang r.

Définition (6,2)

On dit qu'une forme bilinéaire f, définie sur E x E est symétrique si quel que soit $x, y \in E$ $f(x, y) = f(y, x)$

Remarque (6,2)

f symétrique entraîne : quel que soit $1 \leqslant i, j \leqslant n$

$\alpha_{ij} = f(a_i, a_j) = f(a_j, a_i) = \alpha_{ji}$

donc la matrice associée à f est symétrique.

Definition (6,3)

f étant une forme bilinéaire symétrique on appelle forme quadratique q associée, l'application de E dans $\mathbb{R}$ définie par

$$q(x) = f(x, x) \quad \text{quel que soit} \quad x \in E$$

(f est appelée la forme polaire de q)

Ecriture matricielle : $Q(X) = {}^t X A X$ ou A est symétrique et Q la matrice associée à q.

exemple :

$$A = \begin{pmatrix} 2 & -1 \\ -1 & 3 \end{pmatrix} \qquad X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$Q(X) = (x_1, x_2) \begin{pmatrix} 2 & -1 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 2x_1^2 - 2x_1x_2 + 3x_2^2$$

Action d'un changement de base

1/ Sur un vecteur; soit E un espace vectoriel de dimension n, $B = (a_i)_{1 \leq i \leq n}$ et $B' = (a'_i)_{1 \leq i \leq n}$ deux bases de E.

Considérons la matrice $S = (s_{ij})$ ayant pour colonne i les coordonnées de a'_i sur B

$$a'_i = \sum_{1 \leq i \leq n} s_{ij} a_j \qquad (1)$$

S est inversible. (En effet, on sait qu'il existe $s \in \mathcal{L}(E,E)$ unique tel que $a'_i = s(a_i)$ (propriétés 4,2) et $s(a_i)$ est une famille libre, s est de rang n, donc s est inversible.

Soit $x \in F$

si on pose $x = \sum_{1 \leqslant i \leqslant n} a_i x_i = \sum_{1 \leqslant i \leqslant n} a'_i x'_i$

en tenant compte de ① $x_j = \sum_{1 \leqslant i \leqslant n} s_{ij} x'_i \qquad i \leqslant j \leqslant n$

que l'on peut écrire :

$$X = S X' \quad \text{et} \quad X' = S^{-1} X$$

La matrice S de passage donne les anciennes coordonnées $(x_1 \ldots x_n)$ en fonction des nouvelles $(x'_1, \ldots, x'_n)$.

2/ Sur les formes polaires et quadratiques

Soient $x \in E$ et X sa représentation dans B et Y sa représentation dans B'

Soient $x_1 \in E$ et X_1 sa représentation dans B et Y_1 sa représentation dans B'

On a $X = S Y \qquad Y = S^{-1} X$

$X_1 = S Y_1 \qquad Y_1 = S^{-1} X_1$

Soit q la forme quadratique et Q sa représentation dans B

$$Q(X) = {}^t X A X$$

Exprimons X et ${}^tX$ en fonction de Y

$$X = S Y \quad \text{et} \quad {}^t X = {}^t Y \, {}^t S$$

La représentation de q dans la base B'

et $Q(Y) = {}^t Y \, {}^t S A S Y$

$= {}^t Y B Y \qquad$ où $\boxed{{}^t S A S = B}$

La représentation de la forme polaire f dans B et B' sera :

$$F(X, X_1) = {}^tX\, A\, X_1$$

$$F(Y, Y_1) = {}^tY\, {}^tS\, A\, S\, Y_1$$

$$= {}^tY\, B\, Y_1$$

Proposition (6,1)

Etant donné E espace vectoriel, $(a_i)_{1\leqslant i\leqslant n}$ $(a'_i)_{1\leqslant i\leqslant n}$ deux bases de E et S la matrice de passage de $(a_i)_{1\leqslant i\leqslant n}$ à $(a'_i)_{1\leqslant i\leqslant n}$.
Pour toute application bilinéaire f de E dans E : $B = {}^tS\, A\, S$ où A est la matrice associée à f par rapport à $(a_i)_{1\leqslant i\leqslant n}$ et B la matrice associée à f par rapport à $(a'_i)_{1\leqslant i\leqslant n}$

Définition (6,4)

On dit qu'une forme quadratique q sur un espace vectoriel E est positive si $q(x, x) \geqslant 0$ pour tout $x \in E$

Définition (6,5)

Pour qu'une forme quadratique positive q sur un espace vectoriel E soit non dégénérée (ou définie) il faut et il suffit que $q(x, x) > 0$ pour tout $x \neq 0$ dans E.

Espace vectoriel norme, distance

Définition (6,6)

E étant un espace vectoriel, toute application $\| \ \|$ de E dans $\mathbb{R}^+$ vérifiant

$$\begin{cases} \| x \| = 0 \text{ si et seulement si } x = 0 \\ \| x + y \| \leqslant \| x \| + \| y \| \qquad x, y \in E \\ \| \lambda x \| = \lambda \ \|x\| \qquad \lambda \in R \end{cases}$$

est appelée une norme sur E (E est alors un espace vectoriel normé)

Norme associée à une forme quadratique positive non dégénérée

Soit q une forme quadratique positive non dégénérée alors :

quel que soit $x \in \mathbb{R}$ $\quad q(\lambda x) = \lambda^2 q(x)$

donc $\sqrt{q(\lambda x)} = |\lambda| \sqrt{q(x)}$

d'autre part $q(x) = 0$ si et seulement si $x = 0$ (q est définie positive)

enfin vérifier que $\sqrt{q\ (x + y)} \leqslant \sqrt{q(x)} + \sqrt{q\ (g)}$

Proposition (6,2)

L'application de E dans $\mathbb{R}^+$ définie par $x \longrightarrow \sqrt{q(x)}$ est une norme sur E.

Définition (6,7)

Soit E un espace vectoriel, muni d'une norme $x \longmapsto q(x)$

q étant une forme quadratique définie positive, E est alors appelé espace euclidien.

Définition (6,8)

On appelle distance sur un ensemble E, une application d de E dans $\mathbb{R}^+$ telle que :

i/ $d(x, y) = 0$ si et seulement si $x = 0$

ii/ $d(x, y) = d(y, x)$ quels que soient $x, y \in E$

iii/ $d(x, z) \leqslant d(x, y) + d(y, z)$ quels que soient $x, y, z, \in E$

(quelques notions sur les espaces métriques seront données ultérieurement).

Proposition (6,3)

Si $x \longmapsto \| x \|$ est une norme sur un espace vectoriel E, alors $d(x, y) = \| x - y \|$ est une distance sur E telle que $d(\lambda x, \lambda y) = \lambda d(x, y)$ et $d(x + z, y + z) = d(x, y)$

i/ $d(x, y) = 0$ si et seulement si $\|x - y\| = 0$

$\|x - y\| = 0$ si et seulement si $x = y$

ii/ $d(x, y) = \|x - y\|$

$= |-1| \, \|y - x\|$

$= \| y - x\|$

$= d(y, x)$

iii/ $d(x, z) = \|x - z\|$

$\leqslant \|x - y\| + \|y - z\|$

$\leqslant d(x, y) + d(y, z)$

Les deux autres propriétés se démontrent facilement.

7 - DETERMINANT

Définition (7,1)

Etant donné $(E_i)_{1 \leq i \leq n}$ des espaces vectoriels on dit qu'une application de $E_1 \times \ldots\ldots \times E_n$ dans $\mathbb{R}$ est une forme multilinéaire, si chaque application partielle $x_i \longrightarrow f(x_1, \ldots\ldots x_i \ldots\ldots x_n)$ est linéaire.

(On dit aussi "forme n-linéaire")

Définition (7,2)

Une forme n - linéaire définie sur E^n $(\overbrace{E \times \ldots\ldots \times E}^{n})$ où E est un espace vectoriel est alternée si :

$$f(x_1, \ldots\ldots, x_i, \ldots\ldots\ldots x_n) = 0 \quad (x_1, \ldots\ldots, x_i, \ldots\ldots x_n) \in E^n$$

pour tout élément ayant deux coordonnées égales.

Proposition(7,1)

i/ Une forme multilinéaire alternée f change de signe lorsque l'on transpose deux composantes.

En effet : $f(x_1, \ldots\ldots, x_i + x_j, \ldots, x_j + x_i, \ldots\ldots\ldots x_n) = 0$

$$+ f(x_1, \ldots\ldots, x_i, \ldots\ldots x_j, \ldots\ldots x_n)$$

$$+ f(x_1, \ldots\ldots, x_j, \ldots\ldots x_i, \ldots\ldots x_n)$$

donc $f(x_1, \ldots, x_i, \ldots, x_j, \ldots x_n) = f(x_1, \ldots, x_j, \ldots, x_i, \ldots, x_n)$

ii/ Si σ est une permutation et k le nombre de transposition de σ

$$\sigma = i \longmapsto \sigma(i)$$

alors $f(x_{\sigma(1)}, \ldots\ldots, x_{\sigma(n)}) = (-1)^k f(x_1, \ldots\ldots, x_n)$

Formes n - linéaires alternées sur un espace vectoriel de dimension n

Etudions les formes n - linéaires alternées sur E^n

- soient E un espace vectoriel, $(a_i)_{1 \leqslant i \leqslant n}$ une base et $(x_i)_{1 \leqslant i \leqslant n}$ n éléments de E

$$x_i = \sum_{1 \leqslant j \leqslant n} \lambda_{ij} a_j$$

$$f(x_1, \ldots\ldots, x_n) = f\left(\sum_j \lambda_{ij} a_j, \ldots\ldots, \sum_j \lambda_{nj} a_j\right) \qquad 1 \leqslant j \leqslant n$$

$$f(x_1, \ldots\ldots, x_n) = \sum_{\substack{1 \leqslant j_1 \leqslant n \\ \ldots\ldots \\ 1 \leqslant j_n \leqslant n}} \lambda_{1j1} \cdots \lambda_{njn} \; f(a_{j1}, \ldots\ldots, a_{jn})$$

Les termes de la forme $f(a_{j1}, \ldots\ldots, a_{jp}, \ldots\ldots, a_{jq}, \ldots\ldots, a_{jn})$ sont nuls si $j_q = j_p$.

Il ne reste que n ! termes (nombre de permutations d'un ensemble à n éléments ; ensemble noté $\mathfrak{S}_n$).

Soit la permutation σ : $(1, \ldots\ldots, n) \longrightarrow (j_1, \ldots\ldots, j_n)$

$$f(x_1, \ldots\ldots, x_n) = \sum_{\sigma \in \mathfrak{S}_n} \lambda_{1,\sigma(1)} \cdots \lambda_{n\sigma(n)} \; f(a_{\sigma(1)}, \ldots, a_{\sigma(n)})$$

$$f(x_1, \ldots\ldots, x_n) \sum_{\sigma \in \mathfrak{S}_n} \lambda_{1\,\sigma(1)} \cdots \lambda_{n\sigma(n)} \; f(a_1, \ldots\ldots a_n) \quad (1)$$

où $\varepsilon_\sigma = \pm 1$ (suivant la parité de la permutation).

Dans l'expression (1) les $f(a_1, \ldots\ldots, a_n)$ sont indépendants du n - uplet d'éléments de E.

Montrons que les formes construites répondent à la question.

- multilinéaire : évident
- alterné : supposons que $x_i = x_{i'}$ (c'est-à-dire $\lambda_{ij} = \lambda_{i'j}$ $1 \leqslant j \leqslant n$)

à tout terme

$$\varepsilon_\sigma \lambda_{1\sigma(1)} \cdots \lambda_{i\sigma(i)} \cdots \lambda_{i'\sigma(i')} \cdots \lambda_{n\sigma(n)} \quad \text{correspond}$$

$$\varepsilon_{\sigma'} \lambda_{1\sigma(1)} \cdots \lambda_{i'\sigma(i')} \cdots \lambda_{i\sigma(i)} \cdots \lambda_{n\sigma(n)}$$

ces deux termes se diffèrent d'une transposition donc : $\varepsilon_\sigma = -\varepsilon_{\sigma'}$ donc les termes de la somme s'éliminent deux à deux.

La condition ① est une condition nécessaire et suffisante.

Donc : sur un espace de dimension n, il existe une forme n - linéaire alternée déterminée à une constante multiplicative près.

Définition (7,3)

Le déterminant, par rapport à une base ordonnée, **d'un espace** vectoriel est celle des formes linéaires alternées qui prend la valeur 1 pour les n éléments de la base.

$$\det\left[x_1, \ldots, x_n\right] = \sum_{\sigma \in \mathfrak{S}_n} \varepsilon_\sigma \lambda_{1\sigma(1)} \cdots \lambda_{n\sigma(n)}$$

on le notera : $$\det\left[x_1, \ldots, x_n\right] = \begin{vmatrix} \lambda_{11} & \cdots & \lambda_{n1} \\ \vdots & & \vdots \\ \lambda_{1n} & \cdots & \lambda_{nn} \end{vmatrix} \qquad (11)$$

(déterminant d'ordre n)

Le déterminant est linéaire par rapport aux éléments d'une ligne ou d'une colonne (multilinéarité).

$$\det\left[x_1, \ldots\ldots, x_n\right] = \lambda_{i1}\ A_{i1} + \ldots + \lambda_{in}\ A_{in} \qquad (2)$$

Calcul de A_{11} (par exemple) coefficient de λ_{11} dans une expression du type (2) . Si nous mettons λ_{11} en facteur dans l'expression (1) , ce nombre est facteur d'une somme qui n'est autre que le déterminant d'ordre $(n - 1)$ obtenu en supprimant du tableau (1') la première ligne et la première colonne.

Pour déterminer A_{ij} on amène la $i^{ème}$ ligne à la première (ce qui revient à changer $(i - 1)$ fois le signe du déterminant).

Amenons ensuite la $j^{ème}$ ligne à la première (ce qui revient à changer $(j - 1)$ fois de signe).

<u>Conclusion</u> : $A_{i,j}$ est égal au produit par $(-1)^{i-1+j-1} = (-1)^{i+j}$ du déterminant obtenu à partir de (1') en supprimant la $i^{ème}$ ligne et la $j^{ème}$ colonne.

$A_{i,j}$ est appelé le mineur associé à λ_{ij} et $(-1)^{i+j}$ est appelé le cofacteur associé à λ_{ij}.

<u>exemple</u>

$$\begin{vmatrix} a & b & c \\ a' & b' & c' \\ a'' & b'' & c'' \end{vmatrix} = a \begin{vmatrix} b' & c' \\ b'' & c'' \end{vmatrix} - b \begin{vmatrix} a' & c' \\ a'' & c'' \end{vmatrix} + c \begin{vmatrix} a' & b' \\ a'' & c'' \end{vmatrix}$$

$$= a(b'c'' - b''c') - b(a'c'' - a''c') + c(a'c'' - a''b')$$

$$\begin{vmatrix} + & - & + \\ - & + & - \\ + & - & + \end{vmatrix}$$

Déterminant associé à une application linéaire

$(x_1, \ldots, x_n) \in E^n$ et $f, g \in \mathcal{L}(E, E)$

$$(x_1, \ldots, x_n) \longrightarrow \det \left[g(x_1), \ldots, g(x_n) \right]$$

C'est une forme multilinéaire : (g linéaire, déterminant multilinéaire)

C'est une forme alternée : $(x_i = x_j \Longrightarrow g(x_i) = g(x_j))$

Cette forme multilinéaire alternée est proportionnelle à $\det \left[x_1, \ldots x_n \right]$

$$\det \left[g(x_1), \ldots, g(x_n) \right] = h \det \left[x_1, \ldots, x_n \right] \qquad h \in \mathbb{R}$$

$$\det \left[g(a_1), \ldots, g(a_n) \right] = h$$

$$\det \left[g(x_1), \ldots, g(x_n) \right] = \det \left[g(a_1) \ldots g(a_n) \right] \det \left[x_{11}, \ldots, x_n \right]$$

Mais se donner n éléments $(x_1, \ldots, x_n)$ c'est se donner une application linéaire qui aux éléments $(a_1, \ldots, a_n)$ associe $(x_1, \ldots, x_n)$

donc : $\det \left[g \circ f(a_1), \ldots, g \circ f(a_n) \right] = \det \left[g(a_1), \ldots, g(a_n) \right]$

$$\det \left[f(a_1), \ldots, f(a_n) \right]$$

en introduisant les matrices

$f \longrightarrow A$ $\qquad\qquad$ $\det \left[BA \right] = \det B \ \det A$

$g \longrightarrow B$

Proposition (7,2)

i/ $\det A \cdot \det A^{-1} = 1$

ii/ $\det A = \det S^{-1} A S$ ou S matrice de changement (le déterminant d'une application linéaire de base de E dans E est indépendant de la base à laquelle E est rapportée).

iii/ "$\det f \neq 0$" est équivalent à "f isomorphisme de E dans E" (automorphisme).

Application

Proposition (7,3)

Soit $A = (a_{ij})_{1\leqslant i,j\leqslant n}$ une matrice carrée. A est inversible si et seulement si $\det A \neq 0$.

C.N Supposons A inversible et soit A^{-1} son inverse, alors $\det(AA^{-1}) = \det A \det A^{-1} = I_n = 1$

donc $\det A \neq 0$

C.S Supposons $\det A \neq 0$. Si $A = (a)$, $B = (a^{-1})$ est telle que $AB = BA = I$

Soit $A = (a_{ij})_{1\leqslant i,j\leqslant n}$ $n \geqslant 2$; considérons $B = (b_{ij})_{1\leqslant ij\leqslant n}$ où $b_{ij} = c_{ji}/\det A$ et c_{ji} le cofacteur de a_{ij}.

L'élément (i, j) de AB est $\sum a_{ik}\, b_{kj} = \left(\sum_{1\leqslant k\leqslant n} a_{ik}\, c_{jk}\right) \Big/ \det A$

pour $(1\leqslant i, j\leqslant n)$.

Or, $\sum_{1\leqslant k\leqslant n} a_{ik}\, c_{jk} = \det A$

et $\sum_{1\leqslant k\leqslant n} a_{ik}\, c_{jk} = 0 \quad (j \neq i)$

Définition (7,4)

Soit $A = (a_{ij})_{1\leqslant ij\leqslant n}$ une matrice carrée. On appelle adjoint de A et on note $\underline{Adj(A)} = (d_{ij})_{1\leqslant ij\leqslant n}$ la matrice transposée des cofacteurs de $(a_{ij})_{1\leqslant ij\leqslant n}$

$$\text{donc } A^{-1} = Adj\,(A) \,/\, \det A$$

exercice

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 5 \\ 1 & 5 & 12 \end{pmatrix} \qquad \det A = 3$$

$$\text{Adj } A = \begin{pmatrix} 11 & -9 & 1 \\ -7 & 9 & -2 \\ 2 & -3 & 1 \end{pmatrix}$$

$$A^{-1} = \begin{pmatrix} 11/3 & -9/3 & 1/3 \\ -7/3 & 9/3 & -2/3 \\ 2/3 & -1 & 1/3 \end{pmatrix}$$

8 - VECTEURS PROPRES D'UNE MATRICE

Définition (8,1)

Soit $A = (a_{ij})_{1 \leqslant i,j \leqslant n}$ une matrice carrée ; on appelle valeur propre de A (respectivement vecteur propre de A) tout élément $\lambda_i \in \mathbb{R}$ $(1 \leqslant i \leqslant n)$ (respectivement tout élément $x_i \neq 0$) tel que

$$Ax_i = \lambda_i x_i$$

$$\text{ou} \quad (A - \lambda_i) x_i = 0$$

Or un tel système homogène a une solution non triviale si et seulement si $\det (A - \lambda_i) = 0$.

Proposition (8,1)

Soit $A = (a_{ij})_{1 \leqslant ij \leqslant n}$ une matrice. Les valeurs propres de A sont les n racines (distinctes ou non) de $\det (A - \lambda) = 0$

Le polynôme $\det (A - \lambda)$ est appelé "polynôme caractéristique" de A, (il est de degré n par rapport à λ).

Proposition (8,2)

Soient $A = (a_{ij})_{1 \leqslant ij \leqslant n}$ une matrice , P une matrice régulière et $B = PAP^{-1}$; les valeurs propres de A sont identiques à celles de B ; d'autre part trA = trB et rang A = rang B.

En effet,

. $\det (B - \lambda I) = \det (PAP^{-1} - \lambda I) = \det (PAP^{-1} - \lambda PP^{-1}) =$

$\det \left[P(A - \lambda) P^{-1} \right] = \det (A - \lambda)$

donc les valeurs propres de A sont égales à celles de B.

. $trB = trPAP^{-1} = tr(P^{-1}PA) = tr(A)$

. P et P^{-1} régulières donc rang (A) = rang (B)

Proposition (8,3)

Si D est une matrice diagonale ses valeurs propres sont ses éléments diagonaux (trivial).

Proposition (8,4)

Si λ est valeur propre de A et $p \in \mathbb{N}^*$, alors λ^2 est valeur propre de A^2 et λ^{-2} valeur propre de A^{-2} (si $\det A \neq 0$).

. Si $AX = \lambda X$ alors $A(AX) = A(\lambda X) = \lambda AX = \lambda^2 X$ d'où par récurrence sur $A^2X = \lambda^2 X$.

. Si $AX = \lambda X$ alors $\lambda^{-1} A^{-1} AX = \lambda^{-1} A^{-1} \lambda X$ et $\lambda^{-1} X = A^{-1}X$ et par récurrence $\lambda^{-R} X = A^{-r} X$.

Proposition (8,5)

Si $(A - \lambda_1 I) X_1 = 0$ et $(A - \lambda_2 I) X_2 = 0$ où A est une matrice symétrique et $\lambda_1 \neq 0$, $\lambda_2 \neq 0$, $\lambda_1 \neq \lambda_2$, alors ${}^t X_1 X_2 = 0$.

Par extension si A et B sont symétriques et si $(B - \lambda_1 A) X_1 = 0$ et $(B - \lambda_2 A) X_2 = 0$ alors ${}^t X_1 A X_2 = 0$,

· ${}^t X_2 (A - \lambda I) X_1 = {}^t X_2 A X_1 - {}^t X_2 \lambda X_1 = 0$ donc :

$$\lambda {}^t X_2 X_1 = {}^t X_2 A X_1$$

de même $\lambda {}^t X_1 X_2 = {}^t X_1 A X_2$

mais ${}^t X_2 X_1 = {}^t X_1 X_2$ et ${}^t X_2 A X_1^t = ({}^t X_2 A X_1) = {}^t X_1 {}^t A X_2$

mais A symétrique donc ${}^t X_1 {}^t A X_2 = {}^t X_1 A X_2$

$$\lambda_1 {}^t X_2 X_1 = \lambda_2 {}^t X_1 X_2 \quad \text{et} \quad {}^t X_1 X_2 (\lambda_1 - \lambda_2) = 0$$

$$\text{or, } \lambda_1 \neq \lambda_2 \quad \text{donc} \quad {}^t X_1 X_2 = 0$$

· ${}^t X_2 B X_1 = \lambda_1 {}^t X_2 A X_1$ et ${}^t X_1 A X_2 = \lambda_2 {}^t X_1 A X_2$

or, $\lambda_1 {}^t X_1 A X_2 = \lambda_2 {}^t X_2 A X_1$ d'où ${}^t X_1 A X_2 = \lambda_2 / \lambda_1 {}^t X_1 A X_2$

$$\lambda_2 / \lambda_1 \neq 1 \quad \text{donc} \quad {}^t X_1 A X_2 = 0$$

Matrices orthogonales

Définition (8,2)

On appelle matrice orthogonale une matrice $P = (m_{ij})_{1 \leq i, j \leq n}$ telle que

$${}^t P P = 1$$

On remarque que si P_i et P_j sont deux colonnes de P

$${}^tP_jP_i = 0 \quad \text{si} \quad i \neq j \quad \text{et} \quad {}^tP_iP_i = I$$

Proposition (8,6)

Si P est orthogonale, $|\det P| = 1$

En effet, $\det I = \det {}^tP \det P$ et $\det {}^tP = \det P$

exemple :

a. La matrice unité est orthogonale

b. $\begin{pmatrix} \cos\theta & \sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$ est orthogonale

Diagonalisation d'une matrice symétrique

Proposition (8,7)

Si $A = (a)_{1\leqslant i,j\leqslant n}$ est une matrice symétrique, il existe une matrice orthogonale $P = (p_{ij})_{1\leqslant ij\leqslant n}$ telle que ${}^tPAP$ soit diagonale (la démonstration est laissée provisoirement de côté).

exemple

$$A = \begin{bmatrix} 4 & 3 \\ 3 & -4 \end{bmatrix} \qquad \det (A - \lambda) = \lambda^2 - 25$$

les valeurs propres sont racines de $\lambda^2 - 25 = 0$, soit $\lambda = \pm 5$

vecteurs propres : $-x + 3y = 0$ $\qquad V_1 = (3, 1)$

$9x + 3y = 0$ $\qquad V_2 = (1, -3)$

vecteurs propres normés : $U_1 = (3/\sqrt{10},\ 1/\sqrt{10})$

$U_2 = (1/\sqrt{10},\ -3/\sqrt{10})$

On vérifie que :

$$P = \begin{pmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{pmatrix}$$

et $P^{-1} AP = \begin{pmatrix} 5 & 0 \\ 0 & -5 \end{pmatrix}$

Proposition (8,7)

Soient A une matrice symétrique et P la matrice orthogonale qui diagonalise A. Alors les valeurs propres de A sont les éléments diagonaux de $D = {}^tPAP$. Le rang de A est le nombre d'éléments diagonaux non nuls de D.

D matrice diagonale, donc ses valeurs propres sont ses éléments diagonaux et son rang le nombre des éléments diagonaux non nuls.

Puisque P est orthogonale $({}^tP)^{-1} = P$ donc $D = {}^tPA\,({}^tP)^{-1}$ à même rang et les mêmes valeurs propres que A.

Matrices idempotentes

Définition (8,3)

A est idempotente si et seulement si ${}^tA = A$ et $A^2 = A$

exemple $\begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix}$ est idempotente

Proposition (8,8)

Si A est idempotente, les valeurs propres de A sont soit 0, soit 1.

Soit λ valeur propre de A, $AX = \lambda X$ ou $X \neq 0$ vecteur propre

$$A(AX) = \lambda AX = \lambda^2 X$$

or $AA = A$ donc $AX = \lambda^2 X$

$\lambda^2 = \lambda$ donc $\lambda = 0$ ou 1

Remarque : Le rang de A est le nombre valeurs propres non nulles de $D = {}^tP\,AP$ ou P est la matrice orthogonale qui diagonalise A.

Propriétés des matrices définies positives

Proposition (8,9)

Une matrice symétrique A est définie positive si et seulement si les valeurs propres de A sont positives.

. Soit P la matrice orthogonale qui diagonalise A

$$D = {}^tPAP = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

où les λ_i sont les valeurs propres de A

Soit $y = {}^tPX$ telle que $X = ({}^tP)^{-1}Y = PY$

Alors ${}^tXAX = {}^t(PY)\,A\,(PY) = {}^tYDY = \sum_{1 \leqslant i \leqslant n} \lambda_i y_i^2$

Or, $\sum \lambda_i y_i^2 = 0$ entraîne $y_i = 0$ $\quad 1 \leqslant i \leqslant n$

$x_i = Py_i = 0$; donc A est définie positive.

. Réciproquement si A est définie positive ; supposons qu'une valeur propre λ_1 soit nulle

soit $Y' = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ et $X' = PY'$ donc $X' \neq 0$

mais ${}^tX'\,A\,X = {}^tY'\,{}^tP\,A\,P\,Y' = {}^tY'\,A\,Y = \lambda_1 \leqslant 0$

(contraire à l'hypothèse)

Proposition (8,10)

Si A est une matrice symétrique définie positive. Alors $\det A > 0$, le rang de A est égal à n (évident).

Proposition (8,11)

Si $A = (a_{ij})_{1\leqslant i,j\leqslant n}$ est une matrice définie positive et $C = (c_{ij})_{\substack{1\leqslant i\leqslant n \\ 1\leqslant j\leqslant n}}$ une matrice telle que rang $C = n$, alors ${}^tCAC$

est définie positive.

Proposition (8,12)

Si A est définie positive et C régulière, alors ${}^tCAC$ est définie positive.

- ${}^tCAC$ est symétrique
- $Y \neq 0$, ${}^tY({}^tCAC)Y = {}^tXAX$ où $X = CY$. Puisque A est définie positive et $X \neq 0$, ${}^tXAX > 0$

  Mais ${}^tY({}^tCAC)Y$ est donc positif pour $Y \neq 0$ et on a gagné !

Proposition (8,13)

Si A est définie positive, il en est de même de A^{-1}.

(On pose $C = A^{-1}$ dans la démonstration de : Proposition 8,12).

Proposition (8,14)

Soit A une matrice symétrique définie positive, il existe une matrice régulière C telle que $CA{}^tC = I$ et ${}^tCC = A^{-1}$.

Soit P une matrice orthogonale telle que

$$D = {}^tPAP = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \qquad \lambda_i > 0 \qquad 1\leqslant i\leqslant n$$

et soit $M = \begin{pmatrix} 1/\sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & 1/\sqrt{\lambda_n} \end{pmatrix}$

Alors $C = {}^tM\,{}^tP$ produit de matrice régulière est régulière
et $CA{}^tC = {}^tM{}^tPAPM = {}^tMDM = I$

Enfin, si $CA{}^tC = I$ alors ${}^tC(CA{}^tC)C = {}^tCIC = {}^tCC$
d'où ${}^tCCA = I$ et ${}^tCC = A^{-1}$. On a gagné !

BIBLIOGRAPHIE

.GODEMENT Algèbre (Hermann) (exceptionnel)

.QUEYSANNE Algèbre (A.Colin)

.BARBUT Mathématiques des Sciences Humaines
Tomes I et II - (PUF collection Sup) -
(à recommander aux psychologues et sociologues)

I - ANALYSE EN COMPOSANTES PRINCIPALES.

I.1. Le problème pour l'utilisateur.

Le problème posé par l'analyse factorielle peut schématiquement se résumer ainsi :

A : On dispose de données statistiques qui ne peuvent se représenter que dans un espace de dimension élevé ; ce qui signifie le plus couramment que l'on dispose d'un nombre assez important de variables et d'observations, par exemple 60 caractères (variables) mesurés sur 1.000 individus (observations), ou même 8 caractéristiques économiques (variables) mesurées sur 21 régions (observations). Quelquefois, la distinction entre variables et observations est artificielle ou conventionnelle ; en principe, on réservera le nom d'observation à tout ce qui a un caractère répétitif, à tout ce qui peut être considéré comme une réalisation d'un vecteur aléatoire qui aurait autant de composantes qu'il existe de variables.

Pratiquement donc, on dispose d'un tableau rectangulaire de valeurs numériques dont les dimensions sont telles qu'elles sont un obstacle à l'assimilation rapide par le statisticien de l'information contenue dans cet ensemble de nombres.

B : On désire représenter, dans la mesure du possible, ces données statistiques dans un espace de faible dimension, avec le minimum de perte d'information ; ainsi, par exemple, on résumerait les 60 variables mesurées sur les 1.000 individus par cinq variables. (les données de ces cinq variables permettant de reconstituer approximativement les 60).

On passe ainsi de 1.000 points dans un espace à 60 dimensions, à 1.000 points dans un espace à cinq dimensions.

En fait, cinq dimensions sont encore beaucoup pour notre rétine.

Ainsi, nous chercherons d'abord des espaces à une ou deux ou trois dimensions, en les complétant éventuellement par quelques dimensions supplémentaires...

C : On ne fait aucune hypothèse particulière préalable; ceci pour insister sur le côté purement descriptif de l'opération, à l'opposé des démarches des pionniers de l'analyse factorielle, cherchant un petit nombres de facteurs interprétables, à partir de modèles assouplis et rendus plus dociles par l'introduction de paramètres supplémentaires. Nous dirons quelques mots de l'analyse en facteur commun et spécifique en fin de chapitre.

I.2. Le problème pour le statisticien.

Le statisticien a le choix entre plusieurs espaces de départ, pour raisonner : l'espace dit "des observations" (où les données de base seront par exemple représentées par 60 points variables dans un espace à 1.000 dimensions) ou "l'espace des variables" (à 60 dimensions, avec 1.000 points observations dans le cas de notre exemple, chaque point observation ayant 60 coordonnées qui sont les valeurs des 60 variables lui correspondant).

Dans ce dernier espace, par exemple, la distance entre deux points constitue un nombre caractéristique de la ressemblance des deux observations correspondantes. Soit x_1 et x_2 ces observations (x_1 et x_2 sont donc des vecteurs à 60 composantes que nous désignerons par x_{1i} et x_{2i} l'indice i variant de 1 à 60).

La distance $d(x_1, x_2)$ s'écrit :

$$d^2(x_1, x_2) = \sum_{i=1}^{n} (x_{1i} - x_{2i})^2$$

Elle est d'autant plus faible que les composantes x_{1i} et x_{2i} sont proches pour toutes les valeurs de i.

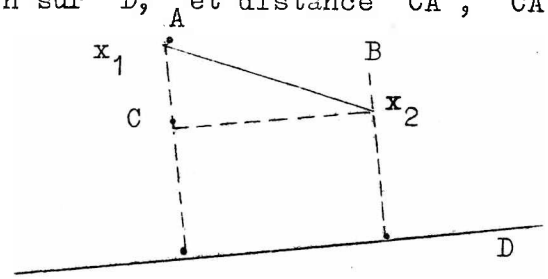
Dire que l'observation x_3 est à la même distance de x_1 que x_2 pose néanmoins, en général, beaucoup de problèmes d'interprétation car l'égalité $\sum_{i=1}^{n} (x_{1i} - x_{2i})^2 = \sum_{i=1}^{n} (x_{1i} - x_{3i})^2$ ne renseigne pas beaucoup sur les composantes responsables de cette égalité... (cette difficulté pourra être levée par la "représentation simultanée").

Notons qu'intervient ici un important problème d'unité et de pondération entre les différentes variables : en effet, si la j^{eme} composante a des valeurs relativement élevées, le terme $(x_{1j} - x_{2j})^2$ jouera un rôle prépondérant dans la somme : $\sum_{i=1}^{n} (x_{1i} - x_{2i})^2$, par suite, la j^{eme} variable aura un poids supérieur aux autres dans la définition des proximités entre observations.

Plaçons-nous donc dans l'espace à p dimensions des variables, $\mathbb{R}^p$, où se trouvent les n points observations.

A : Nous allons chercher le sous-espace à une dimension D (la droite) tel que les distances entre les projections des points-observations mesurées sur cette droite soient le plus proche possible des distances définies plus haut, dans l'espace à p dimensions.

D'après le théorème de Pythagore, la distance entre deux points dans l'espace $\mathbb{R}^p$: AB^2 se décompose entre distance BC^2 de leur projection sur D, et distance CA^2, CA étant



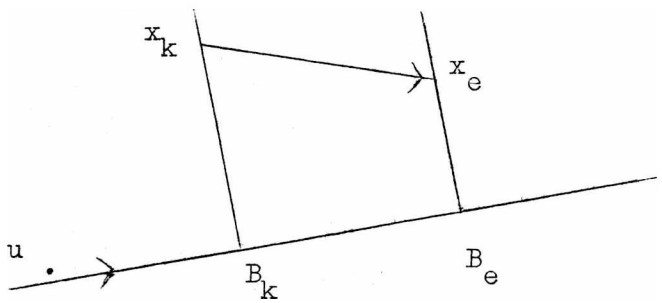
un vecteur de l'espace à $p - 1$ dimension orthogonal à D.

Il est donc naturel de chercher un sous-espace D qui maximise les distances projetées du type BC^2 ou, ce qui revient au même, qui minimise les distances "résiduelles" telles que la distance AC^2.

Soient x_k, x_2 deux points-observations du nuage.

Soient u_1, u_2, ... u_p les composantes d'un vecteur unitaire u porté par la droite cherchée D.

La grandeur de la projection BC est le produit scalaire du vecteur $x_e - x_k$, par le vecteur u.



$$\overline{B_k B_e} = {}^t u\,(x_k - x_e) = \sum_{i=1}^{p} u_i\,(x_{ki} - x_{ei})$$

Si l'on veut que, en moyenne, les longueurs des projections soient maxima, de façon à ce que la déformation du nuage soit minimum, il faut donc que en sommant pour tous les couples (k,e) de points-observations :

$$S = \sum_{k,e} \overline{B_k B_e}^2 = \sum_{k,e} \left[{}^t u \, (x_k - x_e) \right]^2 \quad \text{soit maximum,}$$

avec la condition :

$$\sum u_i^2 = {}^t u \; u = 1 \quad \text{puisque } u \text{ est supposé unitaire.}$$

$$\text{Or} \quad \left[{}^t u \, (x_k - x_e) \right]^2 = \left[\sum_{i=1}^{R} u_i \, (x_{ki} - x_{ei}) \right]^2$$

$$= \sum_{1 \leq i,j \leq p} u_i \, u_j \, (x_{ki} - x_{ei}) \, (x_{kj} - x_{ej})$$

Par suite, en intervertissant les indices (i,j) et (k,e)

$$S = \sum_{1 \leq i,j \leq p} u_i \, u_j \left(\sum_{k,e} (x_{ki} - x_{ei}) \, (x_{kj} - x_{ej}) \right)$$

La quantité entre parenthèse est le terme général v_{ij} d'une matrice V ;

On peut donc écrire $S = \sum_{1 \leq i,j \leq p} u_i \, u_j \, v_{ij}$

En notation matricielle, il faut donc maximiser la forme quadratique :

$$S = {}^t u \, V u \quad \text{avec la condition} \quad {}^t u \, u = 1$$

Précisons quelle est cette matrice V :

Son terme général s'écrit : $v_{ij} = \sum_{k,e} (x_{ki} - x_{ei})(x_{kj} - x_{ej})$

C'est, à un coefficient près, la covariance des variables x_i et x_j

Vérifions le :

$$v_{ij} = \sum_{k,e} (x_{ki} \, x_{kj} - x_{ei} \, x_{kj} - x_{ki} \, x_{ej} + x_{ei} \, x_{ej})$$

$$v_{ij} = n \sum_{k} x_{ki} \, x_{kj} + n \sum_{e} x_{ei} \, x_{ej} - 2 \sum x_{ei} \sum x_{kj}$$

$$v_{ij} = 2n \sum_k x_{ki}\, x_{kj} - 1 \sum_n \sum_e x_{ei} \cdot \sum_k x_{kj}$$

On reconnaît là le terme général de la covariance des variables x_i et x_j, au facteur $2n^2$ près.

Ainsi, les composantes d'un vecteur u unitaire porté par la droite D sur laquelle se projette le nuage en se déformant le moins possible maximisant ${}^t u\, Vu$, (où V est la matrice des covariances des variables 2 à 2) avec la contrainte ${}^t u\, u = 1$

Il s'agit là d'un problème classique en algèbre et en géométrie analytique. (Recherche des axes principaux d'une quadrique).

La matrice des covariances V est symétrique et définie positive. Nous savons (cf. rappel) qu'une telle matrice admet des valeurs propres positives correspondant à des vecteurs propres r_i orthogonaux. Soit R la matrice ayant les vecteurs propres r_i en colonne : on a donc ici :

$VR = R\Delta$ Δ étant une matrice diagonale ayant les valeurs propres i sur sa diagonale.

Multipliant par la transposée ${}^t R$ (qui est aussi, ici, l'inverse de R)

$$V = {}^t R\, \Delta\, R$$

Notre forme quadratique s'écrit : ${}^t u\, Vu = {}^t u\, R\, \Delta\, {}^t Ru$

Posons : $^{t}y = {}^{t}uR$

Et notons que $y = {}^{t}({}^{t}u\,R) = {}^{t}R\,u$ et que $Ry = u$

On a alors : $^{t}u\,V\,u = {}^{t}y\,\Delta\,y = \sum \sigma_i y_i^2$

La condition $^{t}u\,u = 1$ devient, puisque $u = Ry$

$$^{t}y\ {}^{t}RRy = 1 = {}^{t}yy \qquad ({}^{t}RR = I)$$

On est donc ramené au problème suivant :

chercher la maximum de $\sum \sigma_i y_i^2$ sachant que $\sum y_i^2 = 1$

C'est, si l'on veut, un petit problème de programmation linéaire : la solution se trouve nécessairement en un sommet du domaine défini par la relation : $\sum y_i^2 = 1$, donc pour une valeur $y_i = 1$,

$y_j = 0, \quad j \neq i$

La somme $\sum \sigma_i y_i^2$ devant être rendue maximum, l'indice i choisi sera celui correspondant à la plus grande valeur propre σ_i, qui sera la valeur affective du maximum.

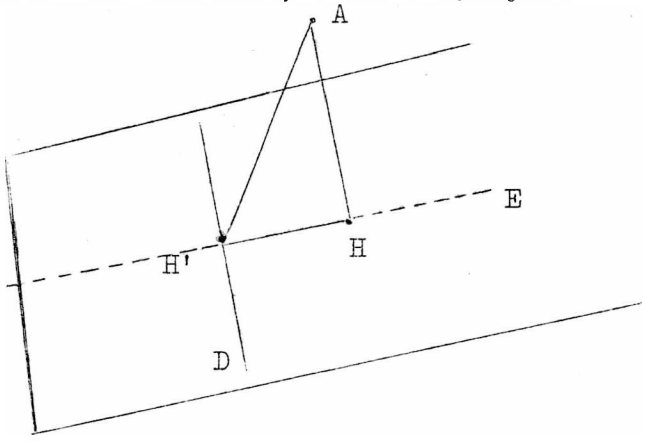
Comme $u = Ry$, le vecteur u cherché sera le vecteur propre correspondant à cette plus grande valeur propre.

Ce vecteur u s'appelle la première composante principale du nuage de points.

On peut donc énoncer le résultat suivant :

"les cosinus directeurs de l'espace à une dimension sur laquelle se projette le nuage des points-observations en se déformant le moins possible (avec notre définition de la non-déformation) sont les composantes du vecteur propre correspondant à la plus grande valeur propre de la matrice des covariances associée au nuage".

B : Si l'on cherche maintenant le sous-espace à deux dimensions (par conséquent représentable sur un graphique) ayant toujours ces propriétés de déformation minimale, l'on s'aperçoit



que ce sous-espace contient D , et qu'il contient le sous-espace à une dimension E ayant également la propriété de "déformation minimale" tout en étant assujetti à être orthogonal à D.

On voit facilement qu'une base de ce sous-espace est constituée par les deux vecteurs propres de la matrice V correspondant aux deux plus grandes valeurs propres. Il en serait de même pour l'espace à p dimensions de déformation minimale, dont une base serait constituée par les p premiers vecteurs propres de V.

Nous avons ainsi trouvé des axes factoriels qui ont la propriété d'extraire progressivement le plus d'information possible concernant les proximités entre les points.

La qualité de cette information peut être mesurée par l'importance de la, ou des plus grandes valeurs propres, selon que l'on utilise un ou plusieurs facteurs pour la représentation.

I.3. Présentation des résultats. (pour le statisticien, pour l'utilisateur)

Le plus souvent, et particulièrement lorsqu'on dispose d'un ensemble hétérogène de variables, on utilise des transformations simples sur ces variables afin de les rendre maniables et comparables.

Ainsi, la transformation la plus couramment utilisée, héritée de la statistique classique, consiste à centrer et réduire les variables, c'est-à-dire à leur retrancher leurs moyennes et à les diviser par leurs écarts-types.

L'analyse précédente revient alors à chercher les vecteurs propres de la matrice des corrélations des variables.

Toutefois, lorsque l'ensemble des variables est homogène, et lorsque les différences entre les échelles des différentes variables ont une signification, il est plus raisonnable de s'en tenir aux données initiales brutes.

La représentation la plus courante consistera à faire figurer les proximités entre points-observations sur un graphique-plan dont les axes sont les deux premiers axes factoriels.

Si ceux-ci s'avèrent insuffisants pour résumer ces proximités, on adjoindra, sur un deuxième graphique les positions de ces points dans le système des troisième et quatrième axes factoriels, etc...

Soit x_{ij} la j^{eme} observation de la variable i.

Les coordonnées du j^{eme} point-observation sur les deux premiers axes factoriels (de cosinus directeurs u et s) sont les produits scalaires ${}^t u\, x_j$ avec :

$$ {}^t u\, x_j = \sum_i u_i\, x_{ij} $$

et ${}^t s\, x_j$ avec : $ {}^t s\, x_j = \sum_i s_i\, x_{ij} $

Ainsi, ces nouvelles coordonnées apparaissent comme des combinaisons linéaires, des variables initiales.

Corrélations variables-composantes :

Supposons les variables centrées et réduites ; calculons le coefficient de corrélation entre la i^{eme} variable x_i et le k^{eme} facteur u_k :

$$ c(i,k) = \frac{\frac{1}{n} \sum_j x_{ij} \left(\sum_e u_{ke}\, x_{ej} \right)}{\sqrt{\text{var}\,(x_i) \cdot \text{var}\,(u_k)}} $$

$\text{var}(x_i) = 1$ puisque la variable x_i est réduite.

$\text{var}(u_k) = \lambda_k$ ($k^{\text{ème}}$ valeur propre = $\max \left({}^t u_k V u_k \right)$ (maximum lié)

D'autre part la somme $\frac{1}{n} \sum_j x_{ij} \left(\sum_e u_{ke} x_{ej} \right) = S$

s'écrit :

$$S = \frac{1}{n} \sum_e \left(\sum_j x_{ij} x_{ej} \right) u_{ke} = \sum_e v_{ie} u_{ke}$$

Comme u_k est vecteur propre de V :

$$\sum_e v_{ik} u_{ke} = \lambda_k u_i$$

Ainsi, le coefficient $c(i,k)$ s'écrit : $c(i,k) = u_i \cdot \sqrt{\lambda_k}$

On a donc le résultat suivant :

- <u>Dans le cas où l'analyse est effectuée sur des variables centrées réduites, les coefficients de corrélation entre une variable "i" et un facteur "k" sont proportionnels à la $i^{\text{ème}}$ composante de l'axe factoriel "k".</u>

Ainsi, si l'on porte sur deux axes les p points ayant pour coordonnées (u_i, s_i), $i^{\text{èmes}}$ composantes des premier et second axes factoriels, on obtiendra dans le plan p points-variables,

dont les proximités peuvent s'interpréter en terme de corrélation.

Notons que si l'on porte ces p points-variables sur le même graphique que les n points-observations, on enrichit considérablement sa lecture.

En effet, au vu des coordonnées des points-variables, on a aussitôt une idée de la signification de facteurs, et l'on sait quelles variables sont responsables de la proximité entre telle ou telle observation.

Il se dégage ainsi une notion de proximité entre variables et observations, confirmée par la remarque suivante :

- les points variables sont des "barycentres" des points-observations, chaque point-observation étant affecté du poids "valeur de la variable pour cette observation". De la même façon, les points-observations sont des barycentres des points-variables.

En effet, nous avons vu que la coordonnée de la j^{eme} observation sur le k^{eme} axe factoriel s'écrivait :

$$c_{kj} = \sum_{i} u_{ki}\, x_{ij} \qquad (\text{pour } k = 1, 2, \ldots.)$$

(ce qui prouve que le point c_j est barycentre des i points-variables u_i avec les "poids" x_{ij}, dont la somme est nulle).

Comme u_k est vecteur propre de $V = (v_{ij}) = (\frac{1}{n} \sum_{e} x_{ie}\, x_{je})$

On a de même, puisque u_k est vecteur propre de V :

$$V\, u_k = \lambda_k\, u_k$$

Soit $\sum_i \left(\frac{1}{n} \sum_e x_{ie}\, x_{je}\right) u_{ki} = \lambda_k\, u_{kj}$

Que l'on peut écrire : $\frac{1}{n} \sum_e \left(\sum_i u_{ki}\, x_{ie}\right) x_{je} = \lambda_k\, u_{kj}$

On reconnaît c_{ke} défini plus haut dans la parenthèse :

$$\frac{1}{n} \sum_e c_{ke}\, x_{je} = \lambda_k\, u_{kj}$$

Ainsi, en remplaçant u_k par $u_k / \sqrt{n \lambda_k}$ on a les deux relations :

$$\begin{cases} \sqrt{n \lambda_k}\; c_{kj} = \sum_e u_{ke}\, x_{ej} \\ \sqrt{n \lambda_k}\; u_{kj} = \sum_e c_{ke}\, x_{je} \end{cases}$$

I.4. Exemple d'application.

1) Variables :

A. Catégories socio-professionnelles (8 variables)

B. Consommations médicales. 11 variables : consultation, visites, actes en B,K,R pour ville et hôpital, pharmacie, soins dentaires et fréquentation hospitalière.
(B=analyses, K=actes techniques, R=radiologie)

2) Observations : les départements français (repérés sur la figure 1 par leurs numeros minéralogiques).

3) Remarques générales :

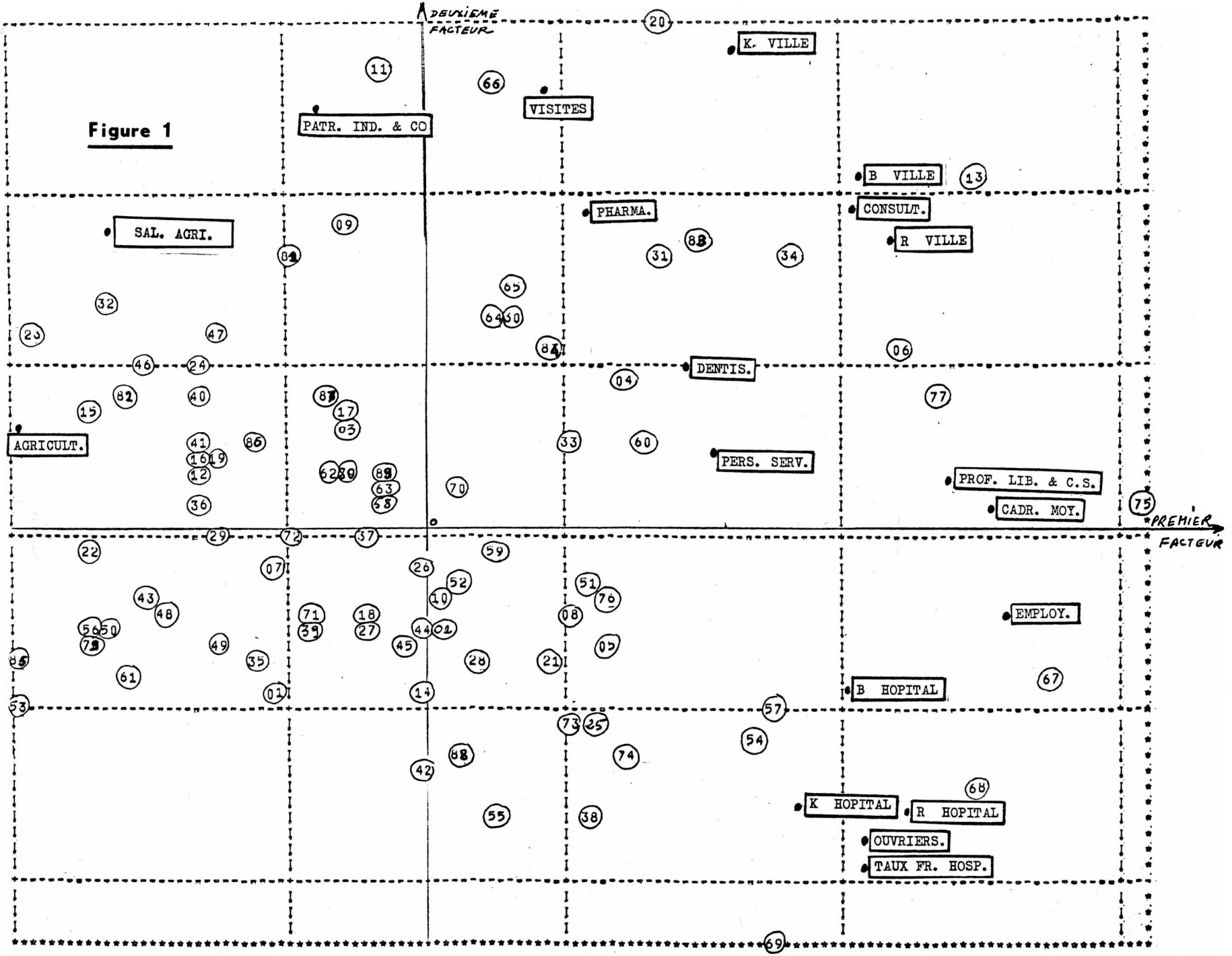
Le premier facteur correspond à une valeur propre représentant 40 % de la somme des valeurs propres ; le second est relatif à une valeur propre représentant 20 % de cette somme (trace de la matrice des corrélations).

La figure 1, qui donne les positions des points-variables et des points-observations résume donc 60 % de la dispersion totale.

Mais cela ne signifie pas que nous n'avons que 60 % de l'information disponible, car les 40 % restant ne sont pas "intéressants" du point de vue de l'interprétation. (En effet, si nous ajoutons d'autres variables sans aucune corrélation avec les précédentes ni entre elles, celles-ci augmenteront la trace (somme des éléments diagonaux de la matrice des corrélations) sans augmenter les premières valeurs propres. Tout en ayant la même information, le "pourcentage d'explication" diminuera.

4) Lecture de la figure 1 : (voir page 53)

Figure 1


DEUXIEME FACTEUR
PREMIER FACTEUR
K. VILLE
PATR. IND. & CO
VISITES
B VILLE
PHARMA.
CONSULT.
R VILLE
SAL. AGRI.
DENTIS.
AGRICULT.
PERS. SERV.
PROF. LIB. & C.S.
CADR. MOY.
EMPLOY.
B HOPITAL
K HOPITAL
R HOPITAL
OUVRIERS.
TAUX FR. HOSP.

Le premier axe factoriel, porté en abscisse, peut être qualifié "d'axe d'urbanisation" au vu des variables qui y participent. (abscisses, sur cet axe,des divers points-variables).

Il oppose en effet le pourcentage d'agriculteurs,à gauche, aux pourcentages de professions libérales, de cadres et d'employés, à droite.

Les abscisses des points-départements sur cet axe sont également significatives : à gauche la Creuse, la Mayenne et la Vendée, à droite la Seine. Notons que la qualification de ce premier axe factoriel "d'axe d'urbanisation" n'est qu'une appréciation qui a un caractère approché.

Si cet axe n'avait aucune interprétation évidente, l'analyse n'en aurait pas été moins intéressante.

Le deuxième axe sépare principalement les consommations de ville des consommations d'hôpital.

Dans la partie supérieure du graphique, se trouvent essentiellement des départements du Midi.

Une proximité entre deux points-variables (respectivement deux points-départements) s'interprète en terme de corrélation (respectivement de comportement similaire vis à vis des variables), et ceci d'autant plus que ces points occupent des positions excentriques (ce qui signifie alors qu'ils sont bien corrélés, dans un sens ou dans un autre, avec les facteurs).

S'il n'est pas important de connaître les échelles choisies sur les axes, pour interpréter les diverses proximités, la position de l'origine, elle, doit donc être connue.

II - ANALYSE FACTORIELLE "EN FACTEURS COMMUNS ET SPECIFIQUES"

Cette méthode, qui est à l'origine de l'analyse factorielle, n'est pas destinée à réduire des données sans hypothèses préalables, dans un but descriptif.

Elle pose au contraire un modèle "a priori" qui est le suivant :

II.1. Le modèle

Les variables que l'on observe ne dépendent linéairement que d'un petit nombre de facteurs sous-jacents, à un rédisu aléatoire près.

Appelons comme précédemment x_{ij} la $j^{ème}$ observation de la $i^{ème}$ variable. x_j est donc un vecteur-observation.

Le modèle stipule que : (p étant le nombre de variables)

$$x_j = \Lambda f_j + a_j \qquad (1)$$

$$(p,1)\ (p,r)\ (r,1)\ \ (p,1)$$

- Les nombres entre parenthèses indiquent la taille des matrices.
- r est le nombre de facteurs.

Pour $r = 2$, par exemple, comme dans le cas des facteurs "intelligence" et "mémoire" que cherchaient les psychologues, le système matriciel (1) s'écrit :

$$\begin{cases} x_{1j} = \lambda_1 f_{1j} + \lambda'_1 f_{2j} + a_{1j} \\ x_{2j} = \lambda_2 f_{1j} + \lambda'_2 f_{2j} + a_{2j} \\ \cdots\cdots \\ x_{ij} = \lambda_i f_{1j} + \lambda'_i f_{2j} + a_{ij} \\ x_{pj} = \lambda_p f_{1j} + \lambda'_p f_{2j} + a_{pj} \end{cases}$$

Chaque observation est donc la réalisation d'un vecteur aléatoire à p composantes, qui ne dépendrait principalement en réalité que de deux variables aléatoires : les deux facteurs f_1 et f_2, à quelques variations résiduelles près.

- Supposons les variables centrées, et calculons la matrice des covariances expérimentale des p variables,
- Supposons également que les résidus aient des corrélations nulles, avec les facteurs, et entre eux :

$$v_{ie} = \sum_j x_{ij}\, x_{ej} = \sum_j \left(\sum_k \lambda_{ik} f_{kj} + a_{ij}\right)\left(\sum_{k'} \lambda_{ek'} f_{k'j} + a_{ej}\right)$$

$$v_{ie} = \sum_{kk'} \lambda_{ik}\, \lambda_{ek'} \sum_j f_{kj} f_{k'j} + \sum_j a_{ij}\, a_{ej} \qquad (2)$$

Notons F la matrice de covariance des facteurs, de terme général :

$$F_{kk'} = \sum_j f_{kj}\, f_{k'j}$$

Notons Δ la matrice des covariances des résidus, de terme général

$$\Delta_{ie} = \sum_{j} a_{ij}\, a_{ej}$$

(d'après nos hypothèses précédentes, $\delta_{ie} = 0$ si $i \neq e$, donc Δ est diagonale).

L'équation (2) s'écrit matriciellement :

$$V = \Lambda\, F\, {}^{t}\Lambda + \Delta$$

- Supposons (encore !) les facteurs indépendants, et de variance 1. Alors $F = I$, (matrice unité)

$$\text{et} \quad V = \Lambda\, {}^{t}\Lambda + \Delta \qquad (3)$$

Sur cette équation matricielle, le lien avec l'analyse en composante principale est clair : en effet, l'analyse en composante principale revient à décomposer la matrice de covariances V de la façon suivante :

$$V = R\, \Delta^{t}\, R \quad (R : \text{matrice orthogonale}$$

dont les colonnes sont les vecteurs propres).

Δ étant diagonale, à élément positif, on peut poser :

$$\delta_{ii}^{1/2} = \sqrt{\delta_{ii}}$$

Par suite : $V = (R\, \Delta^{1/2})\ (\Delta^{1/2}\, {}^{t}R)$

Posant : $\Gamma = R\, \Delta^{1/2}$ on peut noter $V = \Gamma\, {}^{t}\Gamma$ (4)

Γ est une matrice dont les colonnes sont les vecteurs propres de V, multipliés par les racines carrées des valeurs propres (rangées par ordre décroissant).

L'analyse factorielle en facteurs communs et spécifiques suppose donc qu'en retranchant une matrice diagonale à éléments positifs de la matrice des covariances, on peut obtenir une décomposition analogue à celle donnée par la relation ④ , mais dont la matrice Γ aurait un certain nombre de colonnes nulles ; le nombre r de facteurs étant précisément le nombre de colonnes non nulles.

Autrement dit, la matrice des covariances peut se "factoriser" sous la forme $V = \Gamma\,{}^t\Gamma$, Γ étant une matrice carrée (p, p). Mais la matrice $V - \triangle$ n'est plus que de rang r.

Ayant (p - r) valeurspropresnulles, elle peut donc s'écrire :

$$V - \triangle = \Lambda\,{}^t\Lambda \quad (\Lambda \text{ étant une matrice (p, r) seulement})$$

On voit par comparaison des formules ③ et ④ que lorsque les résidus ont une très petite variance, donc que les éléments diagonaux δ_{ii} de $\triangle$ sont petits, l'analyse en facteur commun et spécifique donne des résultats voisins de l'analyse en composantesprincipales.

II.2. Problèmes d'estimation

Nous n'insisterons pas ici sur les problèmes posés par un tel modèle, d'un volume d'ailleurs peu compatible avec l'étroitesse du champ d'application de la méthode.

Nous allons donner ici, quelques moyens pratiques de calcul, en renvoyant aux ouvrages cités en référence les lecteurs "curieux" !

Le problème essentiel est d'estimer Δ, matrice diagonale des variances des résidus spécifiques.
Une fois Δ connue, il suffit de chercher les composantes principales de $V - \Delta$; on ne doit normalement trouver qu'un petit nombre de composantes différentes (statistiquement) de zéro. Nous allons donner ici deux spécifications particulières du modèle :

a. Modèle à variances spécifiques égales

Pour ce modèle, les éléments de Δ sont tous identiques à un même scalaire σ^2. Chaque observation est générée par un schéma du type :

$$x_j = \Lambda f_j + \sigma a_j$$

Avec pour matrice des covariances : $V = E(x^t x) = \Lambda^t \Lambda + \sigma^2 I$

On obtiendra donc une estimation de Λ en prenant les composantes principales de $\Psi_o = V - \sigma^2 I$.

Si $V = R D^t R$ (R orthogonale, D diagonale)

On a alors :

$$\begin{aligned} V - \sigma^2 I &= R D {}^t R - \sigma^2 I \\ &= R D {}^t R - \sigma^2 R {}^t R \\ &= R (D - \sigma^2 I) {}^t R \end{aligned}$$

Ainsi, les valeurs propres de Ψ_o sont donc celles de V, diminuées de $\sigma^2 I$.
(comme Ψ_o doit être de rang $r < p$, V doit avoir σ^2 comme racine multiple d'ordre $p - r$).

Donc, si dans une analyse en composantes principales, les petites valeurs propres sont sensiblement égales, on peut considérer les données comme ayant été générées par un modèle à variances spécifiques égales.

b. Modèle de Joreskog

Les variances spécifiques ne sont pas égales, mais proportionnelles pour la $i^{ème}$ variable à $(1 - R_i^2)$, complément à 1 des carrés des coefficients de corrélation multiples de la $i^{ème}$ variable avec toutes les autres.

Si diag A désigne la matrice diagonale des éléments diagonaux de A.

On a :

$$d\,(1 - R_i^2) = (\text{diag } V^{-1})^{-1}$$

$d(1 - R_i^2)$ est la matrice diagonale ayant pour $i^{ème}$ élément $1 - R_i^2$.

On peut vérifier cette relation en développant par bloc le déterminant $|V|$. Le modèle s'écrit donc :

$$V = \Lambda\Lambda' + \alpha\,(\text{diag } V^{-1})^{-1}$$

On revient au modèle à variance spécifique égale en faisant le changement de variable :

$$z_j = (\text{diag } v^{-1})^{1/2} \; x_j$$

On est amené à diagonaliser la matrice $(\text{diag } V^{-1})^{1/2} \; V(\text{diag } V^{-1})^{1/2}$

Ce modèle donne alors les résultats les plus raisonnables compatibles avec les hypothèses retenues (et voisins de la méthode de Lawley).

.../...

III - ANALYSE DANS UNE METRIQUE ASSOCIEE A UNE FORME QUADRATIQUE DEFINIE POSITIVE

III.1. Méthode

Soient $x_1, \ldots\ldots, x_n$ des éléments de $\mathbb{R}^p$. On définit la distance entre deux éléments x_i et x_j de $\mathbb{R}^p$, par :

$$d^2(i, j) = {}^t(x_i - x_j)\ Q\ (x_i - x_j)$$

où Q est une matrice associée à une forme quadratique définie positive. Or, on sait qu'il existe une matrice régulière C telle que :

$$Q = {}^tC\ C$$

donc $d^2(i, j) = {}^t(x_i - x_j)\ {}^tC\ C\ (x_i - x_j)$

Posons : $y_i = Cx_i \qquad 1 \leqslant i \leqslant n$ (représentant x_i dans la nouvelle base)

$$d^2(i, j) = {}^t(y_i - y_j)(y_i - y_j)$$

Nous cherchons une forme linéaire v, telle $v(x_j) = \sum_{1 \leqslant i \leqslant p} v_i\ x_{ij}$ et avec $v_i = v(a_i) \quad 1 \leqslant i \leqslant n$ où les a_i sont les éléments de base initiale. Cette forme linéaire v étant astreinte à "résumer" l'information contenue dans l'observation j. C'est l'application linéaire "projection sur $\mathbb{R}^n$". Pour avoir cette bonne approximation de l'information contenue il faut que les distances entre les projections soient maxima (c'est-à-dire telles que les ${}^tv(x_i - x_j)$ soient maxima) dans la nouvelle base :

$${}^tv\ x = ({}^tv_o C^{-1})y$$

Comme au (I.2, page 39) dire que la déformation est minimum c'est dire que :

$$ {}^t u \, \mathcal{V} u \quad \text{est maximum avec} \quad {}^t u u = 1 $$

où $\mathcal{V}$ est à un coefficient près la matrice de covariance des y

$$ \mathcal{V} = (\mathcal{V}_{ij}) \quad \text{avec} \quad v_{ij} = \sum_{1 \leqslant k, \ell \leqslant p} (y_{ik} - y_{i\ell})(y_{jk} - y_{j\ell}) $$

comme $y = Cx$, on voit facilement (en français dans le texte) que :

$$ \mathcal{V} = C V {}^t C $$

donc que la forme linéaire u cherchée est solution de (cf. I.2).

$$ \begin{cases} C V {}^t C u = \lambda u \\ {}^t u u = 1 \end{cases} \qquad (1) $$

Nous savons résoudre ce problème.

En effet, la forme linéaire cherchée est telle que :

$$ {}^t u = {}^t v C^{-1} $$

donc $$ u = {}^t (C^{-1}) v $$

et $$ v = {}^t C u $$

Multiplions (1) à gauche par ${}^t C$

$$ {}^t C C V ({}^t C U) = \lambda ({}^t C u) $$

c'est-à-dire $$ (Q V)(v) = \lambda v $$

On sait que la forme linéaire cherchée est le vecteur propre de $Q \cdot V$ relatif à la plus grande valeur propre.
(Nous notons au passage que $C\ V\ {}^tC$ est symétrique et a les mêmes valeurs propres que $Q\ V$).

On constate alors que l'on a les normalisations suivantes :

i/ Norme de v :

$(Q\ V)\ (v) = \lambda v$ en multipliant à gauche par V

$(V\ Q)\ (Vv) = \lambda\ Vv$

alors que v est appelé facteur du nuage de points, $Vv = x$ sera appelé axe factoriel du nuage.

On avait la relation :

$$v = {}^tC\ u$$

$${}^tC^{-1}\ v = u \quad \text{or,} \quad {}^tuu = 1 \quad \text{donc} \quad {}^tv\ C^{-1}\ {}^tC^{-1}\ v = 1$$

$$\text{soit} \quad \underline{v\ Q^{-1}\ v = 1}$$

ii/ Norme de $x = Vv$

$$Q\ V\ (v) = \lambda v \qquad \text{donc} \quad V(v) = \lambda\ Q^{-1}\ v$$

$${}^tv\ V\ v = \lambda\ {}^t v\ Q^{-1}\ v$$

d'où : ${}^tv\ V\ v = \lambda$

$$x = Vv \qquad {}^tv = {}^tx\ V^{-1}$$

alors $VQ\ x = \lambda x$

Multiplions à gauche par ${}^t v$

$$ {}^t x\, V^{-1}\, V\, Q\, x = \lambda\, {}^t x\, V^{-1}\, x $$

$$ {}^t x\, Q\, x = \lambda\, {}^t x\, V^{-1}\, x $$

$$ = \lambda\, {}^t v\, V\, (V^{-1}\, V)\, v $$

$$ = \lambda\, {}^t v\, V\, v $$

$$ \underline{{}^t x\ Q\ x = \lambda^2} $$

III.2. Exemple d'application ; analyse discriminante

Le problème général de l'analyse discriminante est le suivant : il existe plusieurs populations multidimensionnelles $P_1, P_2, \ldots\ldots P_q$.

Les observations statistiques forment donc q "paquets" d'effectifs divers, le "paquet" i contenant par exemple N_i observations du vecteur x.

[exemple : on mesure 15 variables sur 1.000 ménages répartis en 8 catégories socio-professionnelles (qui constituent une partition des 1.000 ménages)].

Le problème est le suivant : disposant d'une observation multidimensionnelle supplémentaire (c'est-à-dire : un vecteur observation), nous nous posons la question :

De quelle population cette observation provient-elle le plus vraisemblablement ?

[Pour notre exemple : peut-on prédire la catégorie socio-professionnelle d'un nouveau ménage caractérisé par ses 15 variables ?]

Le problème prend tout son sens lorsque les valeurs des variables sont les seuls indices permettant de classer les individus. Si les populations P_1, P_2, ... P_q correspondent à des maladies, et les variables mesurées sur les individus à des résultats d'analyses et à des symptômes, l'analyse discriminante permettra de réaliser un diagnostic. (D'où le nom de "fonction diagnostic" donné parfois aux facteurs extraits).

La figure A représente le cas où l'on mesure seulement deux variables par individu, avec quatre populations.

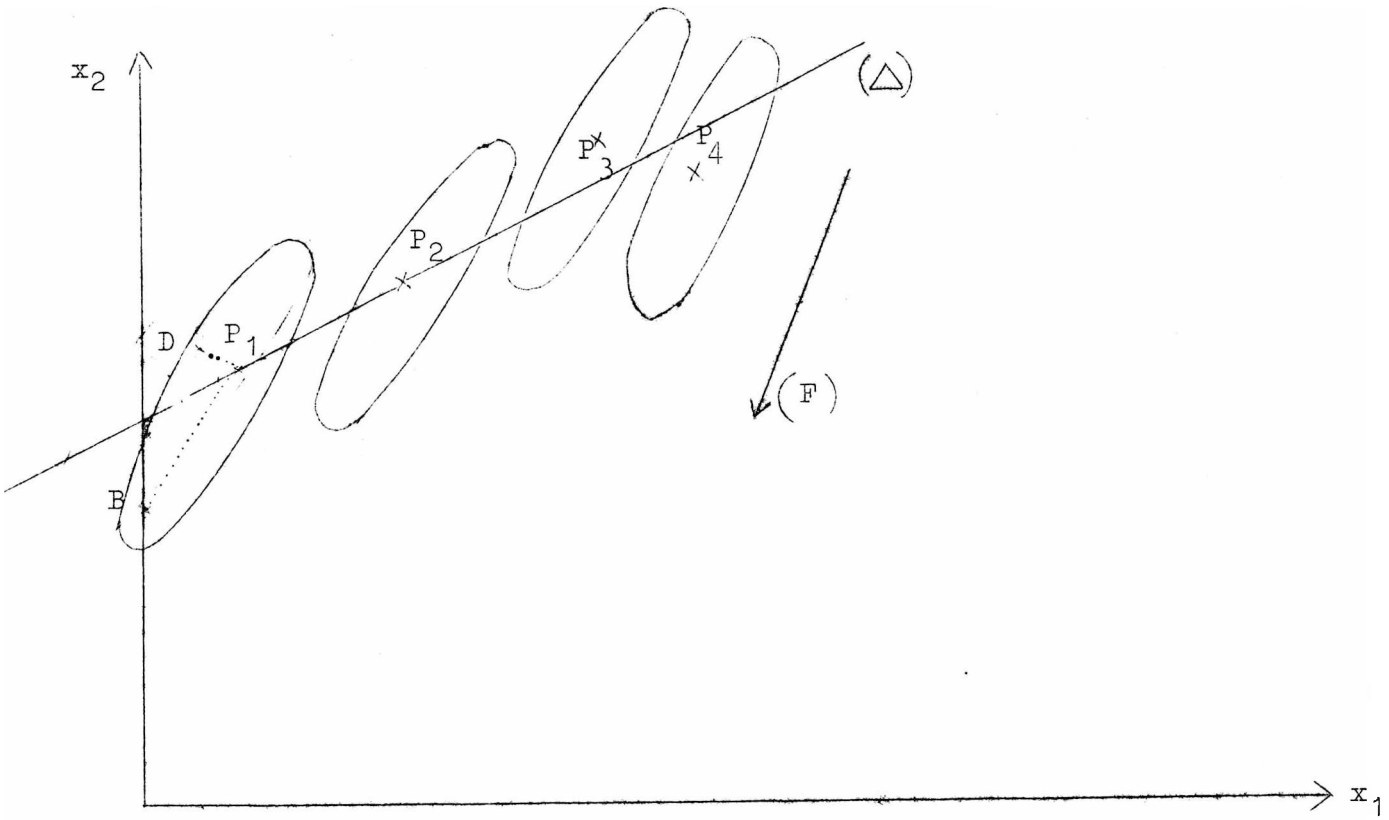


Figure A

Nous supposerons (ce qui est une hypothèse très forte) que les matrices des covariances sont les mêmes dans les diverses populations. Leur valeur commune est notée "V".

Nous voulons donc trouver des facteurs, c'est-à-dire des opérateurs-projection sur des espaces à une dimension, qui permettent de représenter les points d'observations en accentuant l'appartenance à telle ou telle population.

La nécessité d'une métrique particulière s'impose comme on le voit sur la figure A : les points B et D sont à des distances différentes du point moyen P_1, et pourtant ils ont visiblement la même probabilité d'appartenir à cette population.
L'équation des divers ellipsoïdes d'égale densité étant de la forme

$$ {}^t x \, V^{-1} \, x = k \qquad (k = \text{constante positive}), $$

nous choisirons une distance de la forme :

$$ d^2(x_i, x_j) = {}^t(x_i - x_j) \quad V^{-1} \quad (x_i - x_j) $$

C'est-à-dire une distance qui fait en sorte que B et D soient équidistants de P_1.

Avec cette métrique, nous analyserons le nuage des points moyens, dont la matrice des covariances sera notée M.

On doit donc chercher des fonctions u telles que :

$$ V^{-1} \, M \, u = \lambda \, u $$

Les fonctions u sont appellées les <u>fonctions discriminantes</u>.

Remarquons que s'il y a q populations, le rang de M est inférieur ou égal à $q - 1$. Il y a donc moins de $q - 1$ fonctions discriminantes.

Dans le cas de la figure A, la première fonction discriminante serait l'opérateur : "projection sur $\triangle$ parallèlement à F"

S'il n'y a que <u>deux populations</u>, la matrice M (de rang 1) n'est que le produit du vecteur "différence des vecteurs moyens" par son transposé (à un coefficient près)

$$M = (\bar{x}_1 - \bar{x}_2) \quad {}^t(\bar{x}_1 - \bar{x}_2)$$

On a donc :

$$V^{-1} (\bar{x}_1 - \bar{x}_2) \quad {}^t(\bar{x}_1 - \bar{x}_2) \quad u = \lambda u \qquad (1)$$

Multipliant à gauche par ${}^t(\bar{x}_1 - \bar{x}_2)$

$$\left[{}^t(\bar{x}_1 - \bar{x}_2) \quad V^{-1} \quad (\bar{x}_1 - \bar{x}_2) \right] \quad \left\{ {}^t(\bar{x}_1 - \bar{x}_2) \quad u \right\} = \lambda \left({}^t(\bar{x}_1 - \bar{x}_2) \, u \right)$$

La quantité entre crochet est un nombre réel (on a une forme bilinéaire); elle est égale à λ, valeur propre cherchée.

On l'appelle <u>distance généralisée</u> des deux populations multidimensionnelles.

On voit facilement que $u = V^{-1} (\bar{x}_1 - \bar{x}_2)$ est vecteur propre de $V^{-1} M$, pour la valeur propre λ précédente.

Cette fonction qui s'exprime simplement en fonction de V, de $\bar{x}_1$, et $\bar{x}_2$ (respectivement points moyens des premières et secondes populations) est la fonction discriminante des deux populations ;(la matrice des covariances interne V commune aux diverses populations se calcule,en pratique, en faisant une moyenne pondérée des diverses matrices de covariances internes expérimentales).

IV - ANALYSE FACTORIELLE DES CORRESPONDANCES

IV.1. Le problème pour l'utilisateur

L'utilisateur se trouve souvent en présence de "gros" tableaux de correspondances I x J (ou I x J x K) ("Contingency tables"), où, à un couple $(i, j) \in I \times J$ (ou un triplet $(i, j, k) \in I \times J \times K$), correspond un nombre.

Donnons comme exemple le tableau (13 x 90) qui croise les causes de décès et les départements, pour une classe d'âge. A l'intersection de la ligne du département i et de la colonne de la cause j se trouve un nombre k(i, j) de décédés.

De tels tableaux pourraient, d'ailleurs, être étudiés par l'analyse en composantes principales.

Il est plus naturel de chercher une méthode qui tienne compte du caractère probabiliste de ce type de données.

Si $k = \sum_{(i,j) \in I \times J} k(i,j)$; $p(i,j) = \frac{k(i,j)}{k}$ est une

estimation de probabilité.

$$p(i) = \sum_{j \in J} k(i,j)/k$$

$$p(j) = \sum_{i \in I} k(i,j)/k$$

peuvent être interprétés en termes de lois marginales.

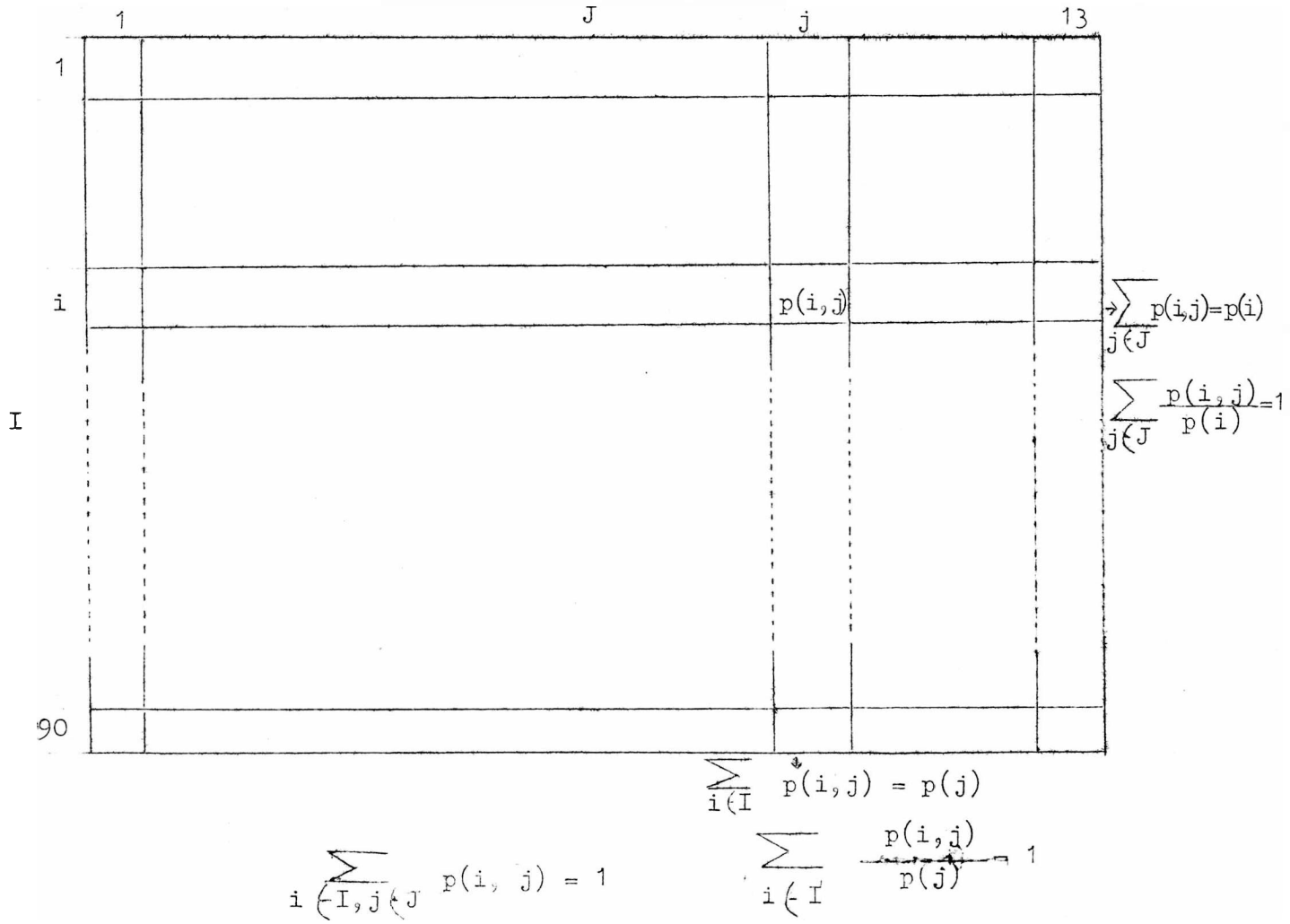
Dans notre exemple p(i) caractérise l'importance du département "i", p(j) l'importance de la cause de décès "j".

Ce qui est surtout intéressant, lorsque l'on compare les causes de décès de deux départements, c'est de confronter la part de chaque cause dans le total des décès départementaux, et non pas les nombres absolus de décès.

En un mot, il est intéressant de comparer les probabilités conditionnelles de décès des départements.

Il s'agit donc de trouver une méthode qui permette de décrire les éventuelles proximités existant entre les lignes et entre les colonnes d'un tableau de correspondances (proximités entre formes indépendamment des niveaux ou tailles), compte tenu des poids différents de ces lignes et de ces colonnes.

Tableau de correspondances



IV.2. Le problème pour le statisticien

1) Position du problème

Comme dans le cas des composantes principales, on va se placer dans un espace ayant autant de dimensions qu'il existe d'éléments dans une ligne ou une colonne du tableau de correspondances.

Choisissons les colonnes pour fixer les idées (pour notre exemple, on se place donc dans un espace à 13 dimensions, dans lequel on aura par conséquent 90 points). Nous verrons plus loin qu'il y a intérêt à prendre comme dimensions celles correspondant au plus petit côté du tableau rectangulaire.

Nous n'allons pas, comme nous l'avions fait précédemment, placer directement les valeurs $p(i,j)$ ou $k(i,j)$ sur les axes de cet espace.

Lorsqu'aucune confusion n'est à craindre I, J désigneront aussi bien un ensemble que le nombre des éléments de cet ensemble.

Dans l'espace $\mathbb{R}^J$ nous construirons un nuage de I points, chaque point ayant pour coordonnées les quantités :

$$\left(\frac{p(i,\ j)}{p(i)} \right)_{j \in J}$$

et étant affecté de la masse $p(i)$.

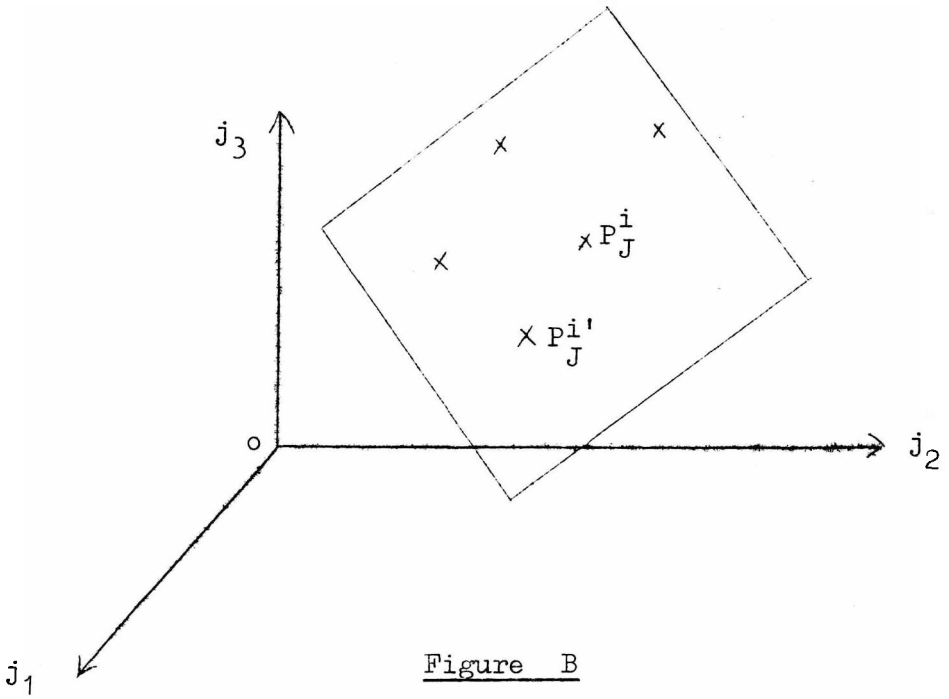


Figure B

Nous noterons P_J^i le vecteur dont les J composantes sont

$$\left(\frac{p(i,j)}{p(i)} \right) \qquad j \in J$$

Remarquons que les I points P_J^i sont tous situés dans un hyperplan (sous-espace à J-1 dimensions), puisque leurs coordonnées vérifient la relation :

$$\sum_j \frac{p(i,\ j)}{p(i)} = 1 \qquad \text{pour} \quad i = 1,\ 2,\ \ldots\ldots\ I$$

Remarquons également que si deux points sont proches dans cet espace, cela signifie que les "profils" des lignes représentées par ces points sont voisins.

Il n'est pas raisonnable de mesurer la distance de deux points par la formule classique :

$$d^2(i, i') = \sum_{j \in J} \left[\frac{p(i,j)}{p(i)} - \frac{p(i', j)}{p(i')} \right]^2 \qquad (1)$$

En effet, dans cette formule intervient la comparaison terme à terme des J éléments des profils de i et de i', en donnant à ces j éléments le même poids.

Supposons que les effectifs, mesurés par $p(j_0)$, de la colonne j_0 soient considérables.

Dans $d^2(i, i')$ le terme :

$$\left[\frac{p(i, j_0)}{p(i)} - \frac{p(i', j_0)}{p(i')} \right]^2$$

sera très grand par rapport aux autres, et jouera un rôle excessif dans la détermination des proximités.

L'expression pondérée :

$$(2) \qquad d^2(i, i') = \sum_{j \in J} \frac{1}{p(j)} \left[\frac{p(i, j)}{p(i)} - \frac{p(i', j_0)}{p(i')} \right]^2$$

a le mérite d'atténuer ces disparités.

Elle a également l'avantage de vérifier le principe "d'équivalence distributionnelle", c'est-à-dire que si deux points P_J^i et $P_J^{i'}$ sont confondus et si on les considère comme un seul point affecté de la somme des masses de i et de i', alors les distances entre les éléments de J ne sont pas modifiées. Cette propriété se vérifie en faisant jouer un rôle symétrique aux indices i et j dans la formule (2) .

.../...

2) Résolution du problème

Le problème se situe donc dans le cadre du Chapitre III, puisque la distance entre P_J^i et $P_J^{i'}$ n'est pas une somme de carrés.

Cependant, la métrique Q utilisée ici est particulièrement simple, puisque la matrice associée à la forme quadratique est diagonale, et vérifie la relation (cf. formule ②)

$$q_{jj} = \frac{1}{p(j)} \qquad (q_{ij} = 0 \quad \text{si} \quad i \neq j)$$

Le changement de base effectué précédemment ($y = Cx$ avec C telle que $Q = {}^tCC$) est ici immédiat puisque Q est diagonale. C est également diagonale (simple changement de l'échelle des axes), et telle que

$$c_{jj} = \sqrt{q_{jj}} = \frac{1}{\sqrt{p(j)}} \qquad (c_{ij} = 0 \quad \text{si} \quad i \neq j)$$

On est donc ramené à une analyse simple en prenant comme coordonnées des points du nuage les quantités :

$$\frac{p(i,\ j)}{p(i)\ \sqrt{p(j)}}$$

Le nuage est maintenant dans l'hyperplan H d'équation :

$$\sum_j \sqrt{p(j)}\ x_j = 1 \qquad (2')$$

Le terme $v_{jj'}$ de la matrice des covariances V s'écrit (en faisant intervenir les poids $p(i)d$, $i \in I$).

$$③ \quad v_{jj'} = \sum_i p(i) \cdot \frac{p(i, j)}{p(i)\sqrt{p(j)}} \cdot \frac{p(i, j')}{p(i)\sqrt{p(j')}} - \left(\sqrt{p(j)}\sqrt{p(j')}\right)$$

En effet, la $j^{ème}$ composante du point moyen (centre de gravité) s'écrit :

$$m_j = \sum p(i) \cdot \frac{p(i, j)}{p(i)\sqrt{p(j)}} = \frac{p(j)}{\sqrt{p(j)}} = \sqrt{p(j)}$$

La relation ③ utilise la décomposition classique de la covariance

$$\left(\sum f_i(x_i - \bar{x})\,(y_i - \bar{y}) = \sum f_i\, x_i y_i - \bar{x}\,\bar{y} \right)$$

$v_{jj'}$ s'écrit finalement :

$$v_{jj'} = \sum_i \frac{p(i, j)\;\; p(i, j')}{p(i)\sqrt{p(j)}\;\sqrt{p(j')}} - \sqrt{p(j)}\;\sqrt{p(j')}$$

Les facteurs cherchés vérifient l'équation

$$Vu = \lambda_u$$

Soit :

$$④ \qquad \sum_k \sum_i \frac{p(i,j)\ p(i,k)}{p(i)\sqrt{p(j)}\sqrt{p(k)}}\ u_k - \sum_k \sqrt{p(j)}\sqrt{p(k)}\ u_k = -\lambda u_j$$

Il y a J équations de ce type $(j = 1, \ldots\ldots, J)$.

Notons que le vecteur propre u défini par $u_k = \sqrt{p(k)}$ est racine évidente de ce système.

En effet, le premier membre de ④ se réduit à

$$\sqrt{p(j)} - \sqrt{p(j)} = 0$$

et le second membre à $-\lambda \sqrt{p(j)}$.

u est donc vecteur propre relatif à la valeur propre 0.

Remarquons que tout vecteur propre v solution de ④, et différent de u, est solution de l'équation simplifiée ④' :

$$④' \qquad \sum_k \sum_i \frac{p(i,\ j)\ p(i,\ k)}{p(i)\sqrt{p(j)}\sqrt{p(k)}}\ u_k = -\lambda u_j$$

En effet, $v \in H$, support du nuage, défini par l'équation ②' ; par suite :

$$\sqrt{p(j)} \sum_k \sqrt{p(k)}\ v_k = 0$$

Finalement les calculs se résument à ceci :

on cherche les vecteurs propres de la matrice <u>symétrique</u> <u>S</u> de terme général :

$$s_{jk} = \sum_{i} \frac{p(i,j)\ \ p(i,k)}{p(i)\sqrt{p(j)\ \ p(k)}}$$

La projection du point "i" de l'ensemble I sur le $r^{ème}$ axe factoriel est égale à $f_r(i)$ tel que :

$$f_r(i) = \sum_{k} u_{rk} \quad \frac{p(i,\ k)}{p(i)\sqrt{p(k)}}$$

Comme nous cherchons des facteurs qui s'appliquent sur les données initiales

$$\frac{p\ (i,\ k)}{p(i)} \qquad \text{(sans modification d'échelle)}$$

il nous faudra prendre comme facteur v_r :

$$v_{rk} = u_{rk} \Big/ \sqrt{p(k)}$$

N.B. Le vecteur ayant pour coordonnées $\sqrt{p(k)}$, vecteur propre de V correspondant à la valeur propre 0, est vecteur propre de S, correspondant à la valeur propre 1.

[Ce vecteur, q-orthogonal à H, n'est pas à prendre en compte dans l'analyse].

3) Liaisons entre les représentations des ensembles I et J.

Nous venons de représenter les proximités entre les éléments de l'ensemble I, vis à vis de leurs associations avec ceux de l'ensemble J.

On pourrait, en renversant les rôles de I et de J dans les calculs ci-dessus, obtenir une représentation analogue des éléments de J. Mais I peut être très grand (I = 90 pour notre exemple) ; les calculs faisant alors intervenir des extractions de valeurs propres, relatives à des matrices (I, I), deviennent alors coûteuses, ou parfois impossibles. Fort heureusement, il existe des relations simples entre les facteurs représentant I et ceux représentant J :

Les facteurs $u_k / \sqrt{p(k)} = v_k$, issus de l'équation 4', vérifient :

$$(5) \quad \sum_k \sum_i \frac{p(i,j)}{p(i)} \frac{p(i,k)}{p(j)} v_k = \lambda v_j$$

((5) est obtenu à partir de 4' en remplaçant u_k par sa valeur en fonction de v_k)

Sommons les deux membres de l'équation (5) par rapport à j, en pondérant les termes par la quantité

$$\frac{p(e,j)}{p(e)} :$$

On obtient, en intervertissant les signes $\sum$:

$$(6)\quad \sum_i \left\{ \sum_j \frac{p(i,j)\cdot p(l,j)}{p(e)\quad p(j)} \right\} \sum_k \frac{p(i,k)}{p(i)}\, v_k = \lambda \sum_j \frac{p(e,j)}{p(e)}\, v_j$$

On reconnaît, entre les accolades, le terme général d'une matrice analogue à celle de l'équation (5), où i et j auraient été permutés.

Posons $\displaystyle\sum_k \frac{p(i,k)}{p(i)}\, v_k = w_i$

Alors l'équation (6) nous montre que le vecteur w, dont les composantes sont les w_i, est un vecteur propre qui joue pour J le même rôle que le vecteur v pour I.

On passe donc très simplement de la représentation de l'ensemble I à celle de J.

Un point reste cependant à éclaircir, celui de la norme du vecteur v.

Calculons-la :

$$\sum_i p(i)\, w_i^{\,2} = \sum_i \left(\sum_k \frac{p(i,k)}{p(i)}\, v_k \right)^2 \cdot p(i)$$

$$= \sum_i \sum_k \sum_{k'} \frac{p(i,k)\, p(i,k')}{p(i)}\, v_k\, v_{k'}$$

D'après l'équation (5) :

$$= \sum_{k'} \left(\lambda\, v_{k'}^{\,2} \cdot p(k') \right) = \lambda \sum_{k'} p(k')\, v_{k'}^{\,2}$$

$= \lambda$ (puisque V est q. normé)

Posons donc maintenant :

$$(7) \qquad w_i = \frac{1}{\sqrt{\lambda}} \sum_k \frac{p(i,k)}{p(i)}\, v_k$$

Le vecteur ainsi obtenu est donc de norme 1.

Ainsi, les facteurs du nuage J sont proportionnels aux coordonnées des points représentatifs de I sur les axes factoriels du nuage I, (et réciproquement).

IV.3. Représentation simultanée des ensembles I et J.

On peut représenter les proximités entre les éléments de I dans le plan des deux premiers axes factoriels. (les coordonnées du point "i" sont alors les nombres w_{1i} et w_{2i})

On peut représenter sur le même graphique les proximités entre les éléments de J (les coordonnées du point "j" étant les nombres v_{1j} et v_{2j}). Cette représentation simultanée est justifiée par le fait que les w_i apparaissent comme des barycentres des v_j, chaque v_j étant affecté du poids $\frac{p(i,j)}{p(i)}$, probabilité

conditionnelle d'occurence de j sachant que i est réalisé.

[ainsi, dans notre exemple, le département i est d'autant plus près de la variable j que celle-ci intervient fortement dans le profil de ce département]

IV.4. Exemple d'application.

L'analyse des correspondances possède un bon pouvoir descriptif des tableaux de nombres positifs. (Sans que ceux-ci soient forcément des tableaux de probabilités ou de fréquences).

Les résultats obtenus lorsqu'interviennent des variables qualitatives (variables prenant les valeurs 0 ou 1) sont sensiblement meilleurs que ceux que donne l'analyse en composante principale.

Ainsi, par exemple, l'analyse de la matrice associée à un graphe à 25 sommets (représentant un damier 5 x 5) par cette méthode (figure 3) donne une représentation plus satisfaisante dans le plan des deux premiers facteurs que l'analyse en composante principale (figure 4).

Application au tableau de 13 causes de décès départementales :

Ce tableau a pour terme générique k(i,j), nombre de décédés masculins de la classe d'âge 65-74 ans en 1962 pour le département i, par la cause j.

Les 13 causes de décès sont les suivantes, avec entre parenthèse les effectifs totaux correspondants :

TU : Tuberculose de l'appareil respiratoire (1.279)
CA : Cancer de toutes formes (non compris les leucémies) (14.425)
CB : Cancer bronco-pulmonaire (2.186)
LE : Leucémie (376)
DI : Diabète sucré (836)
LV : Lésions vasculaires intra-craniennes (8.020)
CO : Maladies du coeur (14.249)
GR : Grippe (477)
PN : Pneumonie, Bronco-pneumonie (1.186)
AL : Alcoolisme (917)
CI : Cirrhose du foie (2.736)
SU : Suicide (733)
AC : Accidents et morts violentes (non compris les suicides) (2.051)

Le graphique II représente les proximités entre les départements, entre les causes de décès, et entre départements et causes de décès.

Rappelons que chaque département intervient avec un poids qui est le nombre total de décédés qu'il contient ; chaque cause de décès avec un poids qui est le nombre total de décédés par cette cause.

On a une approximation de la répartition des décès par cause dans un département "i" en regardant les distances du point "i" aux 13 points-causes.

Notons que le premier facteur est caractérisié par une forte participation de la cause "alcoolisme" s'opposant à "cancer bronco-pulmonaire" et "maladie du coeur".

Nous laissons au lecteur le soin d'analyser plus précisément cette représentation.

V. - Bibliographie sommaire -

1) Publication du Laboratoire de J.-P. Benzécri (Fac. Sciences Paris).

J.-P. BENZECRI (ronéoté) : Leçon sur l'analyse statistique des données dimensionnelles.

J.-P. BENZECRI (ronéoté) : Distance distributionnelle et métrique du χ^2 en analyse factorielle des correspondances.

B. CORDIER (thèse) : Analyse factorielle des correspondances.

J.-P. BENZECRI (ronéoté) : Métrique du χ^2 entre loi de probabilité sur un ensemble fini.

J.-P. BENZECRI : Réduction d'un élément du produit tensoriel de deux espaces euclidiens.

2) Divers.

MC. KENDALL and A. STUART (Griffin) : The advanced theory of statistic.

Donald F. MORRISSON (Mc Graw Hill) : Multivariate statistical methods.

LAWLEY and MAXWELL (Butterworths, London, 1963) : Factor analysis as a statistical méthod.

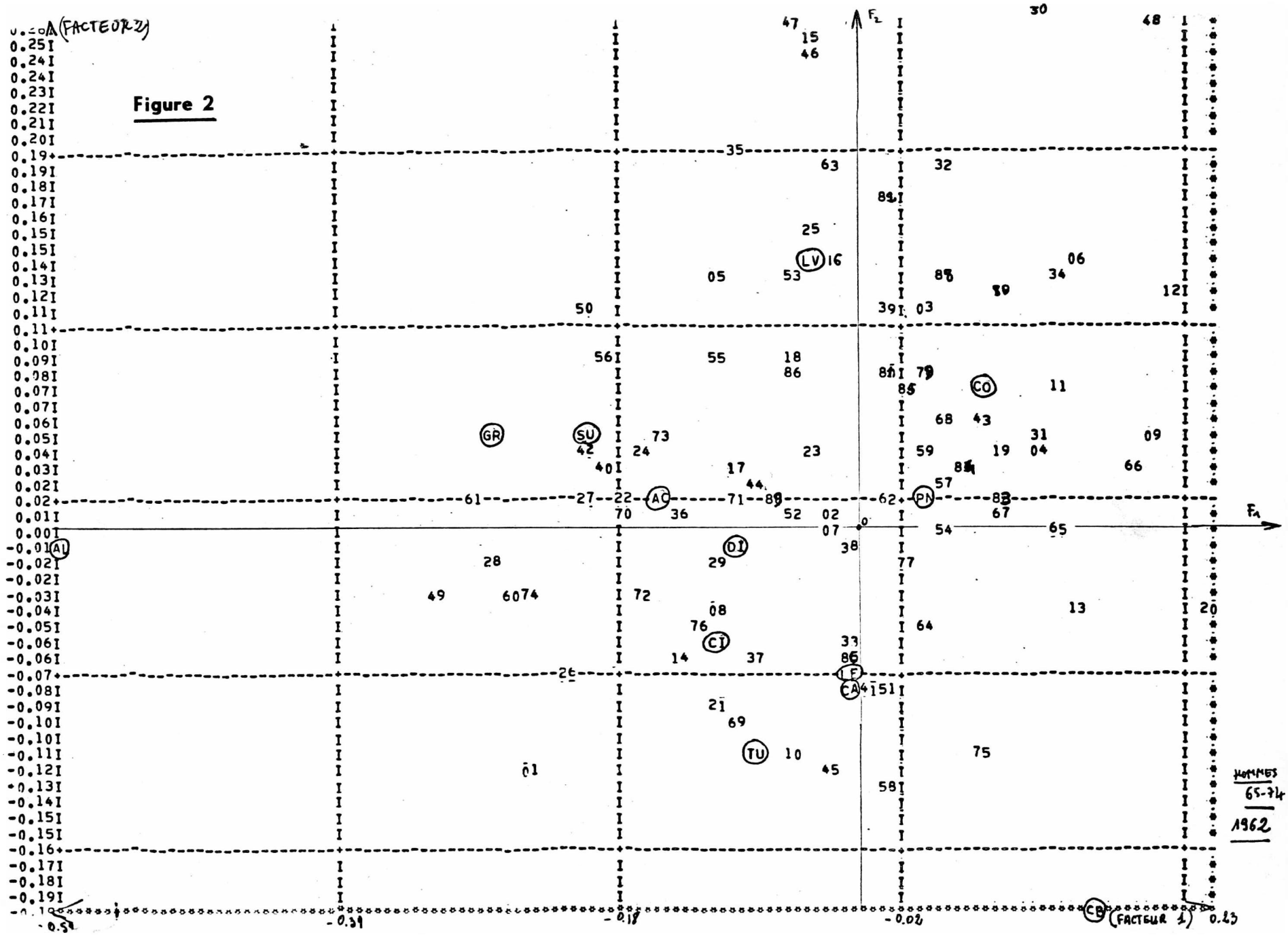
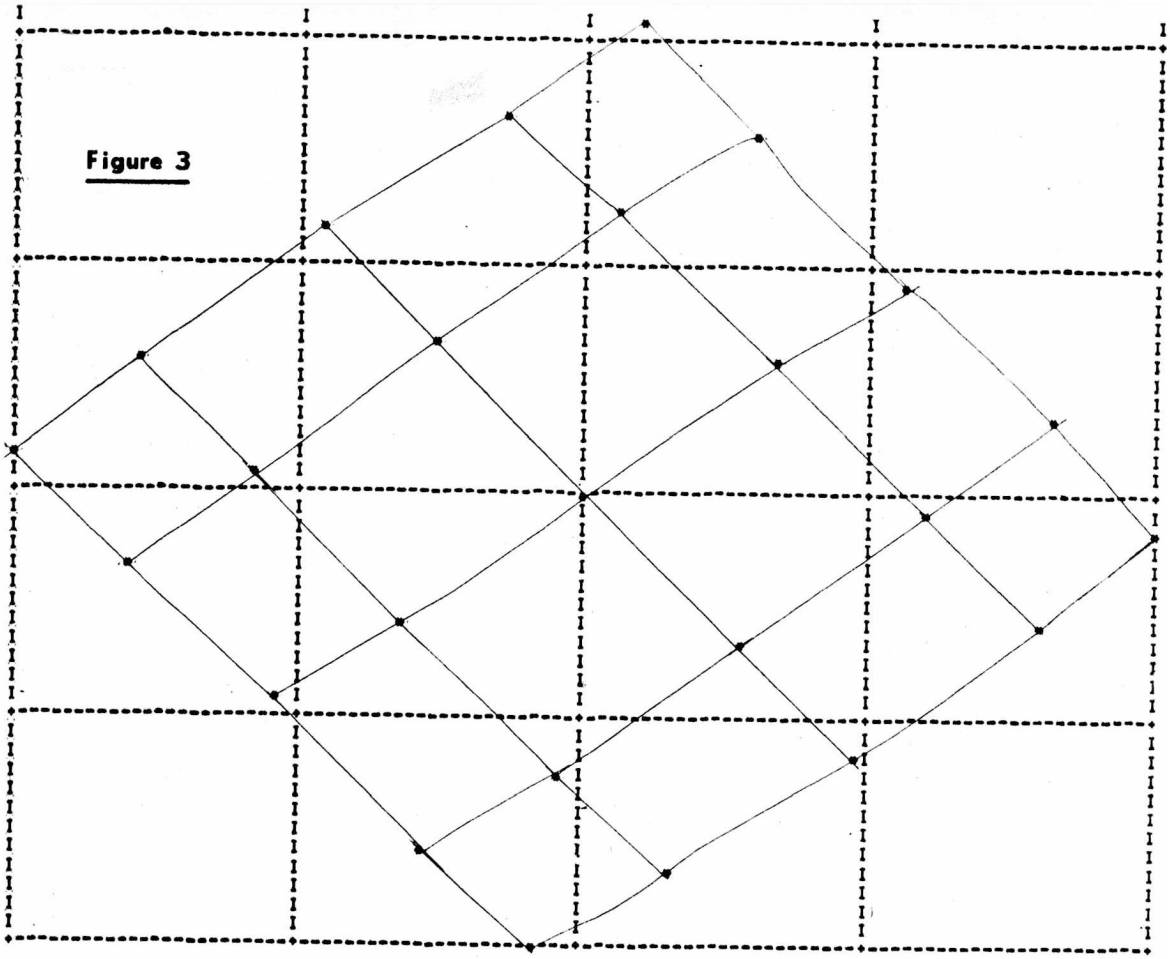


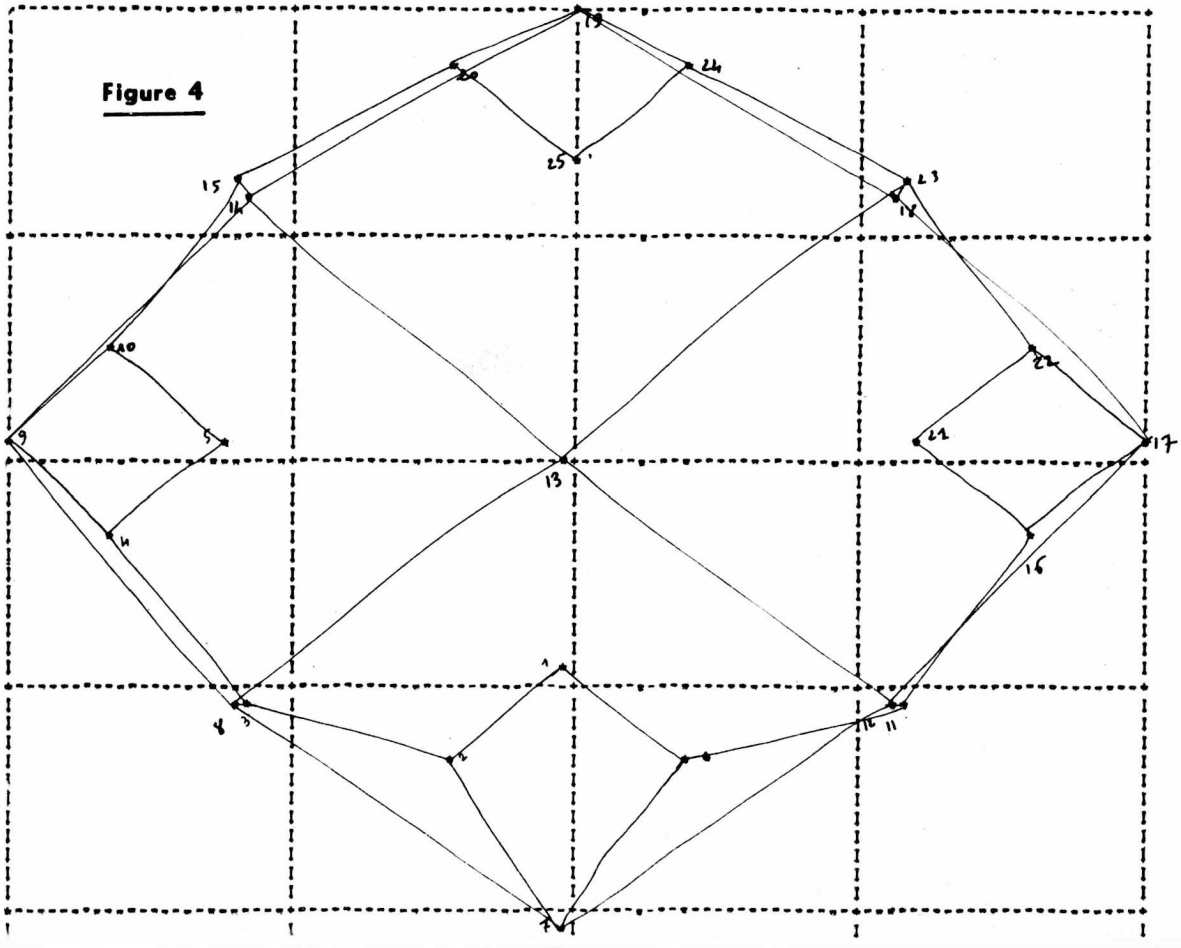
Figure 2

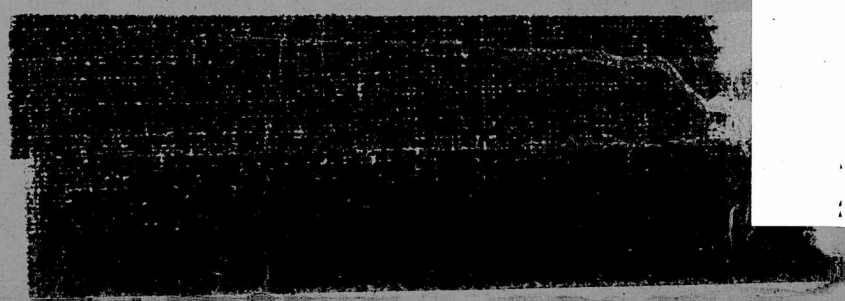
Figure 3



(COMMENTAIRES DANS LE TEXTE)

Figure 4



15 SEP. 1970

2 ex - nº 1