SUR L'ANALYSE DES FAMILLES FINIES DE VARIABLES VECTORIELLES

Bases algèbriques et application à la description statistique

P.A. JAFFRENNOU

Pré-Publication n°4 Octobre 1978

UNIVERSITE DE SAINT-ETIENNE

département de mathématiques

23, rue du Docteur Paul Michelon

SUR L'ANALYSE DES FAMILLES FINIES DE VARIABLES VECTORIELLES

Bases algébriques

et application

à la description statistique

P.A. JAFFRENNOU

INTRODUCTION

K étant un corps commutatif muni d'un automorphisme involutif $\mathcal{J}$, et E et F étant deux K-espaces vectoriels respectivement munis de formes m et n, sesquilinéaires à droite pour $\mathcal{J}$, on montre dans un premier chapitre que l'application :

$$\Phi_{mn} : L_K(E,F) \to K$$

définie par :

$$\Phi_{mn}(f,g) = \mathrm{Trace}(g^* \circ f)$$

est une forme sesquilinéaire à droite pour $\mathcal{J}$ ($L_K(E,F)$ désigne l'espace des applications linéaires de E dans F, et g^* l'adjoint de g relativement à m et n). On étudie ensuite les propriétés de cette forme suivant celles attribuées à m et n et suivant la dimension, finie ou non, des espaces E et F. En particulier si E et F sont hermitiens (resp. euclidiens) on montre que $(L_K(E,F),\Phi_{mn})$ est hermitien (resp. euclidien).

Au second chapitre, on se place dans la situation où K est le corps des complexes et $\mathcal{J}$ est la conjugaison . On se donne à présent des formes hermitiennes p, n et m respectivement sur C^p, C^n et C^m; $E = (C^p,p)$, $F = (C^n,n)$ et $G = (C^m,m)$ sont les espaces hermitiens correspondants. Dans ces conditions une famille $\mathcal{G} = (g_j)_{j=1,\ldots,p}$ d'applications linéaires de F dans G étant donnée, on définit l'application linéaire :

$$f_G = E^* \to \mathcal{L}_C(F,G)$$

par :

$$f_G(x) = \sum_{j=1}^{p} \pi_j^*(x).g_j$$

où π_j^* désigne la j-ème fonction coordonnée de E^* relativement à la base duale de la base canonique de E. On note $d\mathrm{P}$ (resp. $d\Phi$) l'isomorphisme de E (resp. $\mathcal{L}_C(G,F)$) dans son dual associé à P (resp. à Φ_{mn}). Si $(u_\lambda)_{\lambda=1,\dots,p'}$ $(p' \leq p)$ est un système P-orthonormé de vecteurs propres de l'endomorphisme ${}^t f_G \circ d\Phi \circ f_G \circ d\mathrm{P}$ de E, associés aux valeurs propres $\alpha_\lambda > 0$ (les autres étant nulles), on montre que la famille :

$$a_\lambda = \frac{1}{\sqrt{\alpha_\lambda}}.f_G(d\mathrm{P}(u_\lambda)), \qquad \lambda = 1,\dots,p'$$

est Φ_{mn}-orthonormée dans $\mathcal{L}_C(G,F)$ et que tout élément de la famille $\mathcal{G}$ est combinaison linéaire des a_λ. Ces résultats constituent le théorème de décomposition.

Le vecteur ρ_j de $\mathbb{C}^{p'}$ ayant pour coordonnées dans la base canonique de cet espace les coefficients de la combinaison linéaire relative à g_j donne une représentation de g_j dans $\mathbb{C}^{p'}$, et si i est le produit hermitien canonique de cet espace, on a :

$$\mathrm{i}(\rho_j,\rho_\ell) = \Phi_{mn}(g_j,g_\ell).$$

On dit que $(\rho_j)_{j=1,\dots,p'}$ constitue dans $(\mathbb{C}^{p'},\mathrm{i})$ une représentation de $(g_j)_{j=1,\dots,p'}$; les i-distances mutuelles des ρ_j sont égales aux Φ_{mn}-distances des g_j. C'est le théorème de représentation.

Dans le troisième chapitre on considère une famille finie de p vecteurs aléatoires réels de même dimension m, et leurs réalisations sur un échantillon I de taille n. La matrice X_j des réalisations de V_j sur I est considérée comme matrice d'une application linéaire x_j de G^* dans F. On munit G^* de la forme m^{-1}, inverse de m, et le produit scalaire de x_j et x_ℓ est alors donné par :

$$\Phi_{m^{-1}n}(x_j, x_\ell) = \text{Trace}(M.{}^tX_\ell.N.X_j) \ .$$

En appliquant les théorèmes précédents à la famille (x_j), $j = 1,\ldots,p$, on montre que les vecteurs $(u_\lambda)_{\lambda=1,\ldots,p'}$ de E, p-orthonormés, sont les vecteurs propres de U.P, où U est la matrice des produits scalaires des x_j, et P la matrice associée à la forme p définie dans E. Dans $E' = (R^{p'}, i)$ on obtient une représentation (ρ_j) de (x_j) telle que si P est une matrice diagonale de poids affectés aux vecteurs V_j, alors les vecteurs e'_λ de la base canonique de E' sont les axes d'inertie du nuage (ρ_j). On montre qu'il est possible d'expliciter ces axes d'inertie par des représentations conjointes des éléments des deux autres ensembles $I = \{1,\ldots,n\}$ et $K = \{1,\ldots,m\}$.

Pour interpréter statistiquement cette analyse génrale, on suppose que la matrice N associée à la forme n de F, est une matrice diagonale de poids affectés aux éléments de l'échantillon. On définit alors la covariance vectorielle de deux vecteurs V_j et V_ℓ, notée $COV(V_j,V_\ell)$, comme le produit scalaire de leurs réalisations,centrées en colonne. Suivant la nature de la forme m on aura diverses interprétationsde la covariance vectorielle, fonctions des covariances usuelles des variables composantes des deux vecteurs. Le coefficient COR de corrélation linéaire entre variables vectorielles est défini comme à l'ordinaire, à partir du coefficient COV; si $COR(V_j,V_\ell) = +1$, on montre que les réalisations des variables composantes de même rang des vecteurs V_j et V_ℓ sont liées par une relation affine de type $y = ax + b$, où le coefficient a est identique pour tous les couples de variables correspondantes.

Si les données X_j sont a priori centrées ou normalisées (on propose deux types de normalisations) les propriétés de l'analyse générale correspondante sont identiques à celles de l'analyse en C.P classique, sur des données centrées ou normalisées.

Au chapitre IV, la famille des réalisations des p vecteurs aléatoires est considérée comme un "cube" de données de dimension $p\times n\times m$; on établit des propriétés algébriques concernant les trois analyses en C.P du cube, regardé successivement suivant l'une des trois directions. Notamment on montre que l'on perd les propriétés de dualité de l'analyse en C.P classique. Enfin, on établit une formule de reconstitution analytique des données, fondée sur les termes de l'analyse triple.

On examine au chapitre V, ce qu'il advient des résultats établis lorsque la dimension m des vecteurs aléatoires est égale à 1. Dans ce cas on montre que le produit $\Phi_{m^{-1}n}$ défini sur $L_R(G^*,F)$ est confondu avec le produit scalaire usuel n, défini sur F, que l'analyse de la famille (V_j) proposée au chapitre III est confondue avec l'analyse en C.P classique de cette même famille, et que les coefficients de covariance, variance et corrélation vectorielle sont égaux coefficients de covariance, variance et corrélation usuels. Des considérations sur l'analyse triple du cube réduit ici à une matrice viennent conforter les hypothèses avancées au chapitre III concernant les représentations conjointes.

Le chapitre VI est consacré à la comparaison de l'analyse en C.P proposée ici, avec les procédures issues des travaux de TUCKER, CARROLL et ESCOUFIER. Notamment on montre que la formule de reconstitution de TUCKER est identique à celle construite au chapitre IV, lorsque les trois métriques P, N et M sont égales à l'unité.

Enfin le dernier chapitre traite de l'analyse canonique de deux familles finies de vecteurs aléatoires dont on connait les réalisations sur un même échantillon I, sachant que les p vecteurs de la première famille et les q vecteurs de la seconde ont la même dimension m. Les formules de résolution strictement identiques à celles de l'analyse canonique classique, sont obtenues à partir du produit Φ défini sur $L_R(G^*,F)$ et de la définition de la covariance vectorielle proposée au chapitre III. Si m = 1, on retrouve bien entendu l'analyse canonique classique.

Table des matières

Quelques erreurs dactylographiques qui n'ont pu être corrigées avant le tirage :

page I de l'introduction, avant dernière ligne, lire : d'applications linéaires de G dans F étant ...

page II, première ligne, lire : $f_G : E^* \to \mathcal{L}_C(G,F)$

page II, ligne 9 du bas, lire : On dit que $(\rho_j)_{j=1,\ldots,p}$ au lieu de $(\rho_j)_{j=1,\ldots,p'}$ ...

page III, ligne 11, lire : générale au lieu de génrale ...

page 81, le diagramme du bas de page devrait comporter l'indication des applications définies au début de la page 82; l'opérateur a reculé devant la difficulté dactylographique !

page 86, le paragraphe 75 devait commencer par la phrase suivante : On suppose désormais que les matrices X_j^1 et X_ℓ^2 sont centrées pour D_{P_I}. Le lecteur voudra bien la replacer à cet endroit.

Le lecteur consciencieux trouvera certainement d'autres imperfections passées inaperçues; qu'il veuille bien en excuser le dactylographe !

Chapitre I

FORME SESQUILINEAIRE DEFINIE SUR L'ENSEMBLE DES APPLICATIONS LINEAIRES D'UN K-ESPACE VECTORIEL DANS UN AUTRE.

I. Rappels et compléments.

1. Avertissement.

Ce sous-chapitre I reprend les notations et quelques résultats de [2]. Des compléments sont apportés. Les définitions et propriétés présentées ici permettront de construire une forme sesquilinéaire sur l'ensemble des applications linéaires d'un K-espace vectoriel dans un autre. On verra au chapitre V que cette forme généralise le produit hermitien usuel; en outre elle permettra aux chapitres III et VII de généraliser l'analyse en composantes principales et l'analyse canonique à des cubes de données.

Dans tout ce qui suit K est un corps commutatif muni d'un automorphisme involutif J. Pour tout λ de K son image par J sera notée indifféremment $J(\lambda)$ ou $\overline{\lambda}$.

Si K = C, corps des complexes, $\overline{\lambda}$ sera le conjugué de λ.

Si K = R, corps des réels, $\overline{\lambda}$ sera égal à λ.

Sauf mention contraire, tous les espaces vectoriels rencontrés ici utilisent le même corps de base K.

E étant un tel espace vectoriel et E^* son dual, on notera pour tout x de E et tout y de E^* :

(1) $\qquad \langle x,y\rangle = y(x).$

E^{**} étant le bidual de E, pour tout y de E^* et tout z de E^{**}, on notera

(2) $\langle y,z\rangle_* = z(y)$.

Dans le cas de dimension finie, en identifiant E et son bidual, on pourra écrire :

(3) $\langle y,z\rangle_* = \langle z,y\rangle$.

2. Forme sesquilinéaire à droite et applications associées.

Soient $\mathfrak{m}$ une forme sesquilinéaire à droite pour $\mathcal{J}$, définie sur l'espace vectoriel E, et l'application $d\mathfrak{m} : E \to E^*$ définie par :

(4) $x \in E, y \in E, \forall x, \forall y \qquad d\mathfrak{m}(y)(x) = \mathfrak{m}(x,y)$.

Théorème 1.-

> i) $d\mathfrak{m}$ est une application semi-linéaire de E dans E^*;
>
> ii) $d\mathfrak{m}$ est injective ssi $\mathfrak{m}$ est non dégénérée à droite;
>
> iii) si E est de dimension finie et si $\mathfrak{m}$ est non dégénérée à droite, alors $d\mathfrak{m}$ est bijective.

Soit $E^{*\mathcal{J}}$ l'ensemble des formes semi-linéaires définies sur E. Pour tout y de E^* on note $y^{\mathcal{J}}$ l'application de E dans K définie par

(5) $x \in E, \forall x \qquad y^{\mathcal{J}}(x) = \mathcal{J}(y(x))$.

Il est clair que $E^{*\mathcal{J}} = \{y^{\mathcal{J}} \ / \ y \in E^*\}$ et que cet ensemble est un espace vectoriel sur K.

Soit l'application $s\mathfrak{m} : E \to E^{*\mathcal{J}}$ définie par :

(6) $x \in E, y \in E, \forall x, \forall y \qquad s\mathfrak{m}(x)(y) = \mathfrak{m}(x,y)$.

Théorème 2.-

> i) $s\mathfrak{m}$ est une application linéaire de E dans $E^{*\mathcal{J}}$;
>
> ii) $s\mathfrak{m}$ est injective ssi $\mathfrak{m}$ est non dégénérée à gauche;
>
> iii) si E est de dimension finie et si $\mathfrak{m}$ est non dégénérée à gauche alors $s\mathfrak{m}$ est bijective.

On définit également l'application $s\mathfrak{m}^{\mathcal{J}}$ par :

(7) $x \in E, y \in E, \forall x, \forall y \qquad s\mathfrak{m}^{\mathcal{J}}(x)(y) = \overline{\mathfrak{m}(x,y)}$.

Comme $d\mathfrak{m}$, $s\mathfrak{m}^{\mathcal{J}}$ est une application semi-linéaire de E dans son dual.

3. Formes sesquilinéaires à gauche et applications associées.

Si m est à présent une forme sesquilinéaire à gauche définie sur E, l'application dm définie par (4) possède les propriétés énoncées par le théorème 2 et l'application sm définie par (5) possède les propriétés énoncées par le théorème 1.

4. Cas particulier où l'espace vectoriel est de dimension finie.

On suppose ici que E est de dimension finie et que la forme m est sesquilinéaire à droite pour J. On a alors le théorème suivant :

Théorème 3.-

> Les propriétés suivantes sont équivalentes :
>
> i) m est non dégénérée à droite;
>
> ii) m est non dégénérée à gauche;
>
> iii) dm est injective;
>
> iv) sm est injective;
>
> v) dm est surjective;
>
> vi) sm est surjective.

m étant dans ce cas non dégénérée à droite et à gauche, on dit qu'elle est non dégénérée.

5. Forme inverse définie sur l'espace dual.

Hypothèses 1 : E est un espace vectoriel de dimension finie et m est une forme sesquilinéaire à droite pour J, non dégénérée. Dans ces conditions la forme m^{-1} définie sur E^* par (8) est appelée forme inverse de m :

(8) $$x \in E^*, y \in E^*, \forall x, \forall y \qquad m^{-1}(x,y) = m(dm^{-1}(x), dm^{-1}(y)).$$

Hypothèses 2 : E est un espace vectoriel de dimension finie et m est une forme sesquilinéaire à gauche pour J, non dégénérée. La forme inverse de m est alors définie par :

(9) $$x \in E^*, y \in E^*, \forall x, \forall y \qquad m^{-1}(x,y) = m(sm^{-1}(x), sm^{-1}(y)).$$

Théorème 4.-

> Sous les hypothèses 1 (resp. 2) on a les propriétés suivantes :
>
> i) m^{-1} est une forme sesquilinéaire à gauche (resp. à droite) pour J, non dégénérée, définie sur E^*;
>
> ii) m est hermitienne pour J ssi m^{-1} est hermitienne pour J;
>
> iii) dans le cas complexe ou réel m est positive ssi m^{-1} l'est;
>
> iv) en identifiant E avec son bidual, on a :
>
> $$s(m^{-1}) = (dm)^{-1} \quad (\text{resp. } d(m^{-1}) = (sm)^{-1});$$
>
> v) on a : $$(m^{-1})^{-1} = m.$$

Preuve (de iv)) : on se place sous les hypothèses 1; m^{-1} est sesquilinéaire à gauche, par suite $s(m^{-1})$ est une application semi-linéaire de E^* dans E^{**}. Pour tous x et y de E^* on a :

$$\begin{aligned}
<y,s(m^{-1})(x)>_* &= m^{-1}(x,y) && \text{d'après (6)}\\
&= m((dm)^{-1}(x),(dm)^{-1}(y)) && \text{d'après (8)}\\
&= dm((dm)^{-1}(y))((dm)^{-1}(x)) && \text{d'après (4)}\\
&= y((dm)^{-1}(x)) &&\\
&= <(dm)^{-1}(x),y> &&\\
&= <y,(dm)^{-1}(x)>_* && \text{d'après (3).}
\end{aligned}$$

Preuve (de v)) : toujours sous les hypothèses 1, on a :

$$\begin{aligned}
(m^{-1})^{-1}(x,y) &= m^{-1}((s(m^{-1}))^{-1}(x),(s(m^{-1})^{-1}(y)) && \text{d'après (9)}\\
&= m^{-1}(dm(x),dm(y)) && \text{d'après iv)}\\
&= m((dm)^{-1}(dm(x)),(dm)^{-1}(dm(y))) && \text{d'après (8)}\\
&= m(x,y)
\end{aligned}$$

pour tous x et y de E identifié à son bidual.

6. Propriétés de l'adjoint d'un homomorphisme d'un K-espace vectoriel dans un autre.

Hypothèses 3 : E est un K-espace vectoriel de dimension finie, F est un K-espace vectoriel quelconque, m est une forme sesquilinéaire à droite pour J, non dégénérée, définie sur E, et n est une forme sesquilinéaire à droite pour

$\mathcal{J}$, définie sur F.

Soit $L_K(E,F)$ l'ensemble des applications linéaires de E dans F, et soit f un élément de cet ensemble. On rappelle que sous les hypothèses 3, l'application f^* de F dans E définie par :

(10) $\quad x \in E, y \in F, \forall x, \forall y \qquad m(x,f^*(y)) = n(f(x),y)$

est une application linéaire appelée <u>homomorphisme adjoint de</u> f <u>relativement à</u> m <u>et</u> n.

A) sous les hypothèses 3, on a la propriété suivante :

(11) $\quad f \in L_K(E,F), \forall f \qquad f^* = (dm)^{-1} \circ {}^t f \circ dn.$

En effet, l'application transposée de f, notée ${}^t f$, est définie par :

(12) $\quad x \in E, y \in F^*, \forall x, \forall y \qquad \langle x, {}^t f(y)\rangle = \langle f(x),y\rangle;$

par suite, pour tout x de E, tout y de F et tout f de $L_K(E,F)$ on a :

$$\begin{aligned} \langle x, dm \circ f^*(y)\rangle &= m(x,f^*(y)) \\ &= \langle f(x), dn(y)\rangle && \text{d'après (10)} \\ &= \langle x, {}^t f \circ dn(y)\rangle && \text{d'après (12).} \end{aligned}$$

B) sous les hypothèses 3, on a également :

(13) $\quad y \in F, \forall y \qquad f^*(y) = (dm)^{-1}(dn(y) \circ f)$

propriété classique de la transposée, qui s'obtient immédiatement en écrivant la définition (10) sous la forme :

(14) $\quad y \in F, \forall y \qquad dm(f^*(y)) = dn(y) \circ f.$

C) si m et n sont hermitiennes pour $\mathcal{J}$:

(15) $\quad x \in E, y \in F, f \in L_K(E,F), \forall x, \forall y, \forall f \qquad m(f^*(y),x) = n(y,f(x)).$

D) si dans les hypothèses 3, les formes sont sesquilinéaires à <u>gauche</u>, la définition de l'adjoint de f relativement à m et n devient :

(16) $\quad x \in E, y \in F, f \in L_K(E,F), \forall x, \forall y, \forall f \qquad m(f^*(y),x) = n(y,f(x)).$

f^* est encore un élément de $L_K(F,E)$ et les propriétés ci-dessus s'écrivent :

(17) $\quad f \in L_K(E,F), \forall f \qquad f^* = (sm)^{-1} \circ {}^t f \circ sn$

(18) $\quad y \in F, \forall y \qquad f^*(y) = (sm)^{-1}(sn(y) \circ f).$

7. Propriétés de l'application d'adjonction.

On considère l'application de $L_K(E,F)$ dans $L_K(F,E)$ qui à tout élément du premier ensemble fait correspondre son adjoint relativement à m et n. Cette application est manifestement semi-linéaire.

A) Sous les hypothèses 3 et si n est non dégénérée à gauche, l'application d'adjonction est injective : f et g étant deux éléments de $L_K(E,F)$ tels que $f^* = g^*$, (14) permet d'écrire :

$$x \in E, y \in F, \forall x, \forall y \qquad dn(y)(f(x)) = dn(y)(g(x))$$

soit

$$n(f(x),y) = n(g(x),y)$$

B) Sous les hypothèses 3 et si dn est bijective, alors l'application d'adjonction est surjective : soit $g \in L_K(F,E)$ et $f = {}^t(dm \circ g \circ (dn)^{-1})$; on a immédiatement $g = f^*$ d'après (11).

C) Réciproquement, sous les hypothèses 3 et si l'application d'adjonction est surjective, alors dn est injective : sinon, il existe $y \in F$, non nul, tel que $dn(y) = 0$; soit $g \in L_K(F,E)$ tel que $g(y) \neq 0$, et $f \in L_K(E,F)$ tel que $g = f^*$; d'après (11) on a : $dm \circ g = {}^t f \circ dn$, or dm étant bijective, $dm \circ g(y) \neq 0$, ce qui est absurde (on suppose bien entendu que $E \neq \{0\}$).

II. Définition et propriétés fondamentales.

8. Définition.

Sous les hypothèses 3 du §.6, on pose :

(19) $$f \in L_K(E,F), g \in L_K(E,F), \forall f, \forall g \qquad \Phi_{mn}(f,g) = \mathrm{Trace}(g^* \circ f).$$

Dans cette définition les indices m et n attribués à l'application Φ rappellent que l'adjoint de g est relatif aux formes m et n.

On obtient immédiatement de (11) la forme équivalente suivante :

(20) $$f \in L_K(E,F), g \in L_K(E,F), \forall f, \forall g \qquad \Phi_{mn}(f,g) = \mathrm{Trace}((dm)^{-1} \circ {}^t g \circ dn \circ f).$$

Théorème 5.-

> Sous les hypothèses 3, l'application $\Phi_{mn} : \mathcal{L}_K(E,F)\times\mathcal{L}_K(E,F) \to K$ est une forme sesquilinéaire à droite pour $\mathcal{J}$.

En effet, $\mathcal{L}_K(E)$ désignant l'ensemble des endomorphismes de E, on sait que l'application "trace" de $\mathcal{L}_K(E)$ dans K est une forme linéaire. De ce rappel et de la définition (19) résulte la linéarité à gauche de Φ_{mn}. De la semi-linéarité de l'application d'adjonction résulte la semi-linéarité à droite de Φ_{mn}.

9. Conditions pour que Φ_{mn} soit non dégénérée.

Lemme 1.-

> Sous les hypothèses 3, h étant un élément de $\mathcal{L}_K(F,E)$ vérifiant Trace($h\circ f$) = 0, pour tout f de $\mathcal{L}_K(E,F)$, alors h = 0.

Si h n'est pas nul, il existe au moins un vecteur v de F, tel que $w = h(v) \neq 0$. Soit une base de E contenant w et l'application f appartenant à $\mathcal{L}_K(E,F)$ telle que f(w) = v et s'annulant pour tous les autres vecteurs de la base. Il est clair que $h\circ f$ est l'opération de projection sur la droite vectorielle définie par w; par suite Trace($h\circ f$) = 1, ce qui est contraire à l'hypothèse.

Lemme 2.-

> Sous les hypothèses 3, f étant un élément de $\mathcal{L}_K(E,F)$ vérifiant Trace($h\circ f$) = 0, pour tout h de $\mathcal{L}_K(F,E)$, alors f = 0.

Soit M le noyau de f et $(e_1,\dots,e_p)$ une base de M. Soit $(e'_1,\dots,e'_q)$ une base d'un supplémentaire N de M. Les e_i et les e'_j constituent une base de E et si f n'est pas nulle, $q \neq 0$. Les vecteurs $f(e'_j)$ forment une base de l'image de f dans F et en complétant cette famille libre par des vecteurs f_k, $k\in K$, on obtient une base de F. On définie alors $h \in \mathcal{L}_K(F,E)$ par :

$$h(f(e'_j)) = e'_j \qquad j=1,\dots,q$$
$$h(f_k) = 0 \qquad k \in K.$$

Il est clair que $h\circ f$ est l'opérateur de projection sur le sous-espace N et par suite Trace($h\circ f$) = q.1, ce qui est contraire à l'hypothèse.

Théorème 6.-

> Sous les hypothèses 3 et si n est non dégénérée à gauche, alors Φ_{mn} est non dégénérée à droite.

En effet, soit g un élément de $L_K(E,F)$ et supposons $\Phi_{mn}(f,g) = 0$, pour tout $f \in L_K(E,F)$. D'après le lemme 1 on a $g^* = 0$; mais d'après la propriété A) du §.7, l'application d'adjonction est injective, donc $g = 0$.

Si dans le théorème 6, on échange les termes gauche et droite, l'assertion obtenue est fausse : la condition $\Phi_{mn}(f,g) = \text{Trace}(g^* \circ f) = 0$, pour tout g, ne permet d'appliquer le lemme 2 que si l'application d'adjonction est surjective. Si F est de dimension finie, la propriété B) du §.7 et le théorème 3 prouvent que cette condition est alors vérifiée. Mais alors le théorème 6 est encore valable et :

Théorème 7.-

> Sous les hypothèses 3, si F est de dimension finie et si n est non dégénérée, alors Φ_{mn} est non dégénérée.

10. Conditions pour que Φ_{mn} soit hermitienne.

Lemme 3.-

> Soit E un K-espace vectoriel de dimension finie; alors pour tout endomorphisme a de E on a :
>
> $$\text{Trace}(a^*) = \overline{\text{Trace}(a)}.$$
>
> De plus a et a^* ont dans K des valeurs propres conjuguées pour J, avec le même ordre de multiplicité.

L'adjoint d'un endomorphisme a est défini par la formule simplifiée :

(21) $\qquad x \in E, y \in E, \forall x, \forall y \qquad m(x,a^*(y)) = m(a(x),y).$

M, A, B, X et Y désignant respectivement les matrices associées à m, a, a^*, x et y relativement à une base arbitraire de E, la relation (21) s'écrit :

$$^t\overline{Y}.{}^t\overline{B}.M.X = {}^t\overline{Y}.M.A.X$$

d'où $^t\overline{B}.M = M.A$, et M étant inversible, $B = {}^t\overline{M}^{-1}.{}^t\overline{A}.{}^t\overline{M}$. Soit T une indéterminée; le polynome caractéristique P_{a^*} de a^* s'écrit, en posant $P = {}^t\overline{M}$:

$$\begin{aligned}P_{a^*}(T) &= \det(P^{-1}.{}^t\overline{A}.P - T.I)\\ &= \det(P^{-1}.({}^t\overline{A}-T.I).P)\\ &= \det({}^t\overline{A}-T.I)\\ &= \overline{\det(A-T.I)}\\ &= \overline{P_a(T)}\end{aligned}$$

ce qui suffit pour montrer le lemme.

Lemme 4.-

> Sous les hypothèses 3 et si $\mathfrak{m}$ et $\mathfrak{n}$ sont hermitiennes pour $\mathcal{J}$, alors
>
> $f \in L_K(E,F)$, $g \in L_K(E,F)$, $\forall f$, $\forall g$ $\quad (g^* \circ f)^* = f^* \circ g$.

Soient $x \in E$, $x' \in E$, f et g; on a :

$$\begin{aligned}\mathfrak{m}(x,(g^*\circ f)^*(x')) &= \mathfrak{m}(g^*\circ f(x),x')\\ &= \overline{\mathfrak{m}(x',g^*\circ f(x))}\\ &= \overline{\mathfrak{n}(g(x'),f(x))}\\ &= \mathfrak{n}(f(x),g(x'))\\ &= \mathfrak{m}(x,f^*\circ g(x'))\end{aligned}$$

Théorème 8.-

> Sous les hypothèses 3 et si $\mathfrak{m}$ et $\mathfrak{n}$ sont hermitiennes, alors $\Phi_{\mathfrak{mn}}$ est hermitienne pour $\mathcal{J}$.

Ceci résulte immédiatement des lemmes précédents.

11. Espace hermitien (resp. euclidien) des applications linéaires d'un K-espace vectoriel dans un autre, lorsque K = C (resp. R).

Théorème 9.-

> Sous les hypothèses 3, si K = C (resp. R), et si $\mathfrak{m}$ et $\mathfrak{n}$ sont hermitiennes (resp. euclidiennes) positives, alors $\Phi_{\mathfrak{mn}}$ est hermitienne (resp. euclidienne) positive.

Du lemme 4 il vient : $(f^*\circ f)^* = f^*\circ f$, et $f^*\circ f$ est hermitienne, et toutes ses valeurs propres sont réelles. Soit λ une telle valeur propre et $x \neq 0$ un vecteur propre associé :

$$\lambda m(x,x) = m(f^{*}\circ f(x),x)$$
$$= n(f(x),f(x)) \geqq 0$$

ce qui implique que $\lambda \geqq 0$; par suite $\Phi_{mn}(f,f) = \mathrm{Trace}(f^{*}\circ f) = \Sigma\lambda \geqq 0$.

Théorème 10.-

> Si $K = \mathbb{C}$ (resp. $\mathbb{R}$) et si (E,m) et (F,n) sont des espaces hermitiens (resp. euclidiens), il en est de même de $(L_K(E,F),\Phi_{mn})$.

III. Propriétés annexes.

12. Basesorthonormées de $(L_K(E,F),\Phi_{mn})$.

Si $(e_i)_{i\in I}$ est une base m-orthonormée de E et $(f_j)_{j\in J}$ une base n-orthonormée de F, on définit la base $(g_{k\ell})_{k\in I,\ell\in J}$ de $L_K(E,F)$ par :

$$g_{k\ell}(e_\ell) = f_k$$
$$g_{k\ell}(e_h) = 0 \qquad \text{si } h \neq \ell.$$

Sous les conditions du théorème 10, c'est une base Φ_{mn}-orthonormée.

13. Propriétés de dualité.

Dans tout ce qui suit on se place sous les hypothèses :

Hypothèses 4 : E et F sont deux K-espaces vectoriels de dimension finie; m et n sont deux formes sesquilinéaires non dégénérées.

Dans ces conditions on sait que Φ_{mn} est une forme sesquilinéaire non dégénérée.

A) Transposé de l'adjoint :

Pour toute $f\in L_K(E,F)$, ${}^t(f^{*}) \in L_K(E^{*},F^{*})$ et $({}^tf)^{*} \in L_K(E^{*},F^{*})$. On a le schéma suivant, où les applications $dm, dn, s(m^{-1})$ et $s(n^{-1})$ sont bijectives :

$$\begin{tikzcd}
E \arrow[r, shift left, "f"] \arrow[d, shift left, "dm"] & F \arrow[l, shift left, "f^{*}"] \arrow[d, shift right, "dn"'] \\
E^{*} \arrow[u, shift left, "s(m^{-1})"] \arrow[r, shift right, "{{}^t(f^{*}),({}^tf)^{*}}"'] & F^{*} \arrow[u, shift right, "s(n^{-1})"'] \arrow[l, shift right, "{}^tf"']
\end{tikzcd}$$

($({}^t f)^*$ désigne l'adjoint de ${}^t f$ relativement à n^{-1} et m^{-1} (cf.(8)).

Théorème 11.-

> Sous les hypothèses 4, pour toute f de $L_K(E,F)$, on a :
> $${}^t(f^*) = ({}^t f)^*.$$

En effet, pour tout x de E et tout y de F on peut écrire :

$$n(y,(f^*)^*(x)) = m(f^*(y),x)$$
$$= n(y,f(x));$$

étant non dégénérée, on en déduit $(f^*)^* = f$. Par ailleurs :

$${}^t(f^*) = dn\circ f\circ (dm)^{-1}$$
$$= (s(n^{-1}))^{-1}\circ {}^{tt}f\circ s(m^{-1}) \qquad \text{(th. 4)}$$
$$= ({}^t f)^*.$$

B) Dualité :

On pose $\ell = g^*\circ f \in L_K(E)$ et $h = ({}^t f)^*\circ {}^t g \in L_K(F^*)$. De (11) on déduit :

$$\ell = (dm)^{-1}\circ {}^t g\circ dn\circ f$$
$$h = dn\circ f\circ (dm)^{-1}\circ {}^t g$$

et :

(22) $\qquad h\circ dn\circ f = dn\circ f\circ \ell$

(23) $\qquad \ell\circ (dm)^{-1}\circ {}^t g = (dm)^{-1}\circ {}^t g\circ h.$

Soit $\lambda \neq 0$ une valeur propre de ℓ et $x \neq 0$ un vecteur propre associé :

$$h(dn\circ f(x)) = dn\circ f(\ell(x))$$
$$= \bar{\lambda} dn\circ f(x)$$

et $\bar{\lambda}$ est valeur propre de h, car $dn\circ f(x) \neq 0$. De même, si μ est une valeur prorpre non nulle de h, $\bar{\mu}$ est valeur propre de ℓ.

Soit V_λ le sous-espace propre de E associé à la valeur propre λ de ℓ et W_μ le sous-espace propre de F^* associé à la valeur propre μ de h. On a immédiatement les inclusions :

$$dn\circ f(V_\lambda) \subset W_{\bar{\lambda}} \qquad \text{et} \qquad (dm)^{-1}\circ {}^t g(W_\mu) \subset V_{\bar{\mu}}\ .$$

Dans le cas où $\mu = \bar{\lambda}$, on en déduit :

$$h(W_{\bar\lambda}) = dn_\circ f_\circ (dm)^{-1}{}_\circ {}^t g(W_{\bar\lambda}) \subset dn_\circ f(V_\lambda) \subset W_{\bar\lambda}$$

et :

$$\ell(V_\lambda) = (dm)^{-1}{}_\circ {}^t g_\circ dn_\circ f(V_\lambda) \subset (dm)^{-1}{}_\circ {}^t g(W_{\bar\lambda}) \subset V_\lambda \ ;$$

la valeur propre λ étant non nulle, on a $h(W_{\bar\lambda}) = W_{\bar\lambda}$ et $\ell(V_\lambda) = V_\lambda$, et par suite : $dn_\circ f(V_\lambda) = W_{\bar\lambda}$ et $(dm)^{-1}{}_\circ {}^t g(W_{\bar\lambda}) = V_\lambda$. En définitive, $\dim(V_\lambda) \geqq \dim(W_{\bar\lambda})$ et $\dim(W_{\bar\lambda}) \geqq \dim(V_\lambda)$, ce qui assure l'égalité et les valeurs propres non nulles λ de ℓ et $\bar\lambda$ de h ont même ordre de multiplicité. Ainsi :

<u>Lemme</u> 5.-

> Sous les hypothèses 4, pour tout f et tout g de $L_K(E,F)$, les endomorphismes $\ell = g^*{}_\circ f$ et $h = ({}^t f)^*{}_\circ {}^t g$ admettent le même nombre de valeurs propres non nulles. En outre chaque valeur propre non nulle de ℓ est conjuguée, pour J, d'une valeur propre de h et réciproquement.

On obtient alors le résultat suivant :

<u>Théorème</u> 12.-

> Sous les hypothèses 4, et dans le cas particulier où $K = C$ (resp. R), on a, pour tout f et tout g de $L_K(E,F)$:
>
> $$\Phi_{n^{-1}m^{-1}}({}^t g, {}^t f) = \overline{\Phi_{mn}(f,g)}$$
>
> (resp. $\Phi_{n^{-1}m^{-1}}({}^t g, {}^t f) = \Phi_{mn}(f,g)$).

Dans le cas complexe cela résulte du caractère algébriquement clos de C, car $\mathrm{Trace}(\ell) = \Phi_{mn}(f,g) = \sum\lambda$ et $\mathrm{Trace}(h) = \Phi_{n^{-1}m^{-1}}({}^t g, {}^t f) = \sum\bar\lambda$. Dans le cas réel il suffit de prolonger m, n, f et g aux complexifiés de E et F, et de remarquer que les valeurs propres non réelles sont conjuguées.

<u>Remarque</u> : si l'on se place dans les conditions du théorème 10 et si l'on fait f = g, on a :

$$\ell = f^*{}_\circ f = (dm)^{-1}{}_\circ {}^t f_\circ dn_\circ f \quad \text{et} \quad h = ({}^t f)^*{}_\circ {}^t f = dn_\circ f_\circ (dm)^{-1}{}_\circ {}^t f.$$

D'après ce qui précède il est clair que ℓ et h sont des opérateurs hermitiens respectivement de E et de F^*. Sachant que ceux-ci admettent de valeurs propres positives, on retrouve alors facilement les propriétés classiques :

- les valeurs propres non nulles de ℓ et h sont réelles positives et égales;
- E (resp. F^*) admet des bases $\mathfrak{m}$-orthonormées (resp. $\mathfrak{n}^{-1}$-orthonormées)

formées de vecteurs propres de ℓ (resp. h).

Chapitre II

DECOMPOSITION ET REPRESENTATION D'UNE FAMILLE FINIE D'APPLICATIONS LINEAIRES D'UN K-ESPACE VECTORIEL DANS UN AUTRE

I. Théorèmes généraux.

14. Introduction.

Dans tout le chapitre II, K est le corps des complexes. On définira d'abord la notion d'opérateur hermitien associé à une application linéaire d'un espace hermitien dans un autre, et on établira les propriétés de cet opérateur. Dans le cas réel on retrouve les propriétés usuelles utilisées en analyse des données. Ces propriétés seront utilisées dans la suite du chapitre pour montrer deux théorèmes fondamentaux concernant une famille finie d'applications linéaires d'un espace hermitien dans un autre. Au chapitre III, ces théorèmes seront utilisés dans le cas réel pour construire l'analyse en composantes principales d'une famille finie de vecteurs aléatoires. On considérera un produit hermitien p (resp. n) sur C^p (resp. C^n) et les espaces hermitiens $E = (C^p,p)$ et $F = (C^n,n)$.

15. Opérateur hermitien de E associé à une forme sesquilinéaire de E^*.

Lemme 1.-

u étant une forme sesquilinéaire à droite, définie sur E^*, pour tout x de E on a :

(24) $$p(du \circ dp(x),x) = u(dp(x),dp(x)).$$

On a en effet :

$$\begin{aligned} p(du_\circ dp(x),x) &= dp(x)(du_\circ dp(x)) \\ &= du(dp(x))(dp(x)) \\ &= u(dp(x),dp(x)). \end{aligned}$$

Théorème 1.-

> u étant une forme hermitienne, définie sur E^*, l'opérateur $du_\circ dp$ est hermitien.

Il suffit de montrer que $p(x,du_\circ dp(x))$ est réel pour tout x de E; or $p(x,du_\circ dp(x)) = u(dp(x),dp(x))$ d'après le lemme, et u étant hermitienne le deuxième membre est réel.

Corollaire 1.-

> u étant une forme hermitienne définie sur E^*, l'opérateur hermitien $du_\circ dp$ vérifie les propriétés suivantes :
>
> i) les valeurs propres $(\alpha_\lambda)_{\lambda=1,\dots,p}$ de $du_\circ dp$ sont réelles;
>
> ii) il existe une famille p-orthonormée $(u_\lambda)_{\lambda=1,\dots,p}$ de vecteurs propres associés.

Ce rappel en forme de corollaire introduit les notations.

Corollaire 2.-

> Si la forme u définie sur E^* est hermitienne et positive, les valeurs propres $(\alpha_\lambda)_{\lambda=1,\dots,p}$ de $du_\circ dp$ sont positives ou nulles.

16. Propriétés de maximisation.

On suppose désormais u positive et $\alpha_1 \geqq \alpha_2 \geqq \dots \geqq \alpha_p \geqq 0$. Pour tout x de E on pose :

(25) $$\gamma_u(x) = u(dp(x),dp(x)).$$

On écrira souvent γ au lieu de γ_u pour simplifier. Il est clair que $\gamma(x)$ est réel. On a alors :

Théorème 2.-

> Le vecteur propre u_1 de $du_\circ dp$ réalise le maximum de $\gamma(x)$ sous la contrainte $p(x,x) = 1$. Pour tout $\lambda > 1$, le vecteur propre u_λ

réalise le maximum de $\gamma(x)$ sous les contraintes $p(x,x) = 1$ et $p(u_\nu,x) = 0$ pour tout $\nu < \lambda$. Enfin, pour tout λ, on a $\gamma(u_\lambda) = \alpha_\lambda$.

La démonstration de ce théorème est classique (cf. [4], par exemple) : en notant (x_λ) les coordonnées de x dans la base (u_λ) de E, le problème revient à maximiser $\sum_\lambda |x_\lambda|^2\alpha_\lambda$ sous les contraintes $\sum_\lambda |x_\lambda|^2 = 1$ et $x_\nu = 0$ pour $\nu \neq \lambda$.

Le système $(u_\lambda)_{\lambda=1,\dots,p}$ n'est pas la seule solution au problème; en effet, tout système $(e^{i\Theta_\lambda}.u_\lambda)_{\lambda=1,\dots,p}$ où $\Theta_\lambda \in R$, est aussi solution.

17. <u>Opérateur hermitien associé à une application linéaire</u>.

Soit f une application linéaire de E^* dans F. On a le schéma suivant :

$$\begin{tikzcd} E^* \arrow[r, "f"] & F \arrow[d, "dn"] \\ E \arrow[u, "dp"] & F^* \arrow[l, "{}^t f"'] \end{tikzcd}$$

Soit u la forme définie sur E^* par :

(26) $$du = {}^t f \circ dn \circ f.$$

<u>Lemme</u> 2.-

Pour tous x et y de E^* on a :

(27) $$u(x,y) = n(f(x),f(y)).$$

En effet :

$$\begin{aligned} u(x,y) = du(y)(x) &= {}^t f \circ dn \circ f(y)(x) \\ &= \langle x, {}^t f \circ dn \circ f(y)\rangle_* \\ &= \langle f(x), dn(f(y)\rangle \\ &= n(f(x),f(y)). \end{aligned}$$

<u>Théorème</u> 3.-

La forme u est hermitienne positive.

Ce résultat découle directement de (27). En conséquence de ce théorème, l'opérataur $du \circ dp$ de E vérifie les propriétés des corollaires 1 et 2 et du théorème 2.

On notera encore $(\alpha_\lambda)_{\lambda=1,\dots,p}$ les valeurs propres réelles positives ou nulles de $du\circ dp$, rangées dans l'ordre décroissant et $(u_\lambda)_{\lambda=1,\dots,p}$ un système p-orthonormé de vecteurs propres associés.

Définition.-

L'opérateur $du\circ dp$ s'appelle l'opérateur hermitien associé à f.

18. Dualité.

L'opérateur hermitien de F associé à l'application ${}^tf$ sera noté $dv\circ dn$; on a donc :

(28) $$dv\circ dn = f\circ dp\circ {}^tf\circ dn$$

et :

(29) $$dv = f\circ dp\circ {}^tf.$$

Théorème 4.-

Les opérateurs hermitiens associés à f et à ${}^tf$ ont les mêmes valeurs propres non nulles avec le même ordre de multiplicité. Pour tout $\lambda = 1,\dots,p$ tel que $\alpha_\lambda \neq 0$, on pose $v_\lambda = f\circ dp(u_\lambda)$; le système (v_λ) ainsi obtenu est n-orthogonal et ce sont des vecteurs propres de $dv\circ dn$. De plus $n(v_\lambda,v_\lambda) = \alpha_\lambda$.

On a en effet, pour $\alpha_\lambda \neq 0$:

$$\begin{aligned} dv\circ dn(v_\lambda) &= f\circ dp\circ {}^tf\circ dn\circ f\circ dp(u_\lambda)\\ &= f\circ dp\circ du\circ dp(u_\lambda)\\ &= f\circ dp(\alpha_\lambda.u_\lambda)\\ &= \overline{\alpha_\lambda}.f\circ dp(u_\lambda)\\ &= \overline{\alpha_\lambda}.v_\lambda. \end{aligned}$$

Les α_λ sont réels et comme $du\circ dp(u_\lambda) \neq 0$, a fortiori, $v_\lambda \neq 0$, ce qui prouve la première partie du théorème. Par ailleurs pour tous λ et λ' on a :

$$\begin{aligned} n(v_\lambda,v_{\lambda'}) &= n(f\circ dp(u_\lambda),f\circ dp(u_{\lambda'}))\\ &= u(dp(u_\lambda),dp(u_{\lambda'}))\\ &= p(du\circ dp(u_\lambda),u_{\lambda'})\\ &= \alpha_\lambda.\delta_{\lambda\lambda'} . \end{aligned}$$

Remarque : si $\alpha_\lambda = 0$, on aurait encore $n(v_\lambda,v_\lambda) = 0$, donc $v_\lambda = 0$.

II. Opérateur hermitien associé à une famille finie d'applications linéaires.

19. <u>Définition</u>.

Sauf mention contraire, on se place désormais sous les hypothèses suivantes :

<u>Hypothèses</u> 5 : n et m sont des produits hermitiens, respectivement sur C^n et C^m.

On notera $F = (C^n, \mathsf{n})$ et $G = (C^m, \mathsf{m})$ les espaces hermitiens correspondants. Soient $\mathcal{G} = (g_j)_{j=1,\dots,p}$ une famille finie d'applications linéaires de G dans F, et p un produit hermitien sur C^p. On considérera les espaces hermitiens $E = (C^p, \mathsf{p})$ et $L = (\mathcal{L}_C(G,F), \Phi_{\mathsf{mn}})$. Enfin, on notera π_j (resp. π_j^*) les fonctions coordonnées de E (resp. de E^*) relatives à la base canonique (resp. la base duale) de l'espace E (resp. du dual E^*).

A la famille $\mathcal{G}$ d'applications linéaires on associe l'application linéaire $f_{\mathcal{G}} : E^* \to L$ définie pour tout x de E^* par :

$$f_{\mathcal{G}}(x) = \sum_{j=1}^{p} \pi_j^*(x) . g_j . \tag{30}$$

Pour simplifier on notera f au lieu de $f_{\mathcal{G}}$. On a alors le schéma de dualité :

$$\begin{tikzcd} E^* \arrow[r, "f"] & L \arrow[d, "d\Phi_{\mathsf{mn}}"] \\ E \arrow[u, "d\mathsf{p}"] & L^* \arrow[l, "{}^t f"] \end{tikzcd} .$$

L'opérateur hermitien, noté $d\mathsf{u} \circ d\mathsf{p}$ associé à f est ${}^t f \circ d\Phi_{\mathsf{mn}} \circ f \circ d\mathsf{p}$. On a :

$$d\mathsf{u} = {}^t f \circ d\Phi_{\mathsf{mn}} \circ f . \tag{31}$$

On dira, de façon équivalente, que $d\mathsf{u} \circ d\mathsf{p}$ est l'opérateur hermitien associé à f ou à la famille $\mathcal{G}$.

D'après (27) la forme hermitienne positive u correspondant à $d\mathsf{u}$, et définie sur E^*, est donnée par :

$$x \in E^*, y \in E^*, \forall x, \forall y \qquad \mathsf{u}(x,y) = \Phi_{\mathsf{mn}}(f(x), f(y)) . \tag{32}$$

Comme précedemment on note $(\alpha_\lambda)_{\lambda=1,\dots,p}$ les valeurs propres (réelles, positives ou nulles) de $d\mathsf{u} \circ d\mathsf{p}$, ordonnées de façon décroissante, et $(u_\lambda)_{\lambda=1,\dots,p}$

un système p-orthonormé de vecteurs propres associés. On notera p' le plus grand entier inférieur ou égal à p pour lequel $\alpha_\lambda \neq 0$ quel que soit $\lambda \leq p'$.

L'opérateur hermitien $dV \circ d\Phi_{mn}$ dual associé à ${}^t f$ est donné par :

(33) $$dV \circ d\Phi_{mn} = f \circ dp \circ {}^t f \circ d\Phi_{mn}.$$

Pour tout $\lambda = 1,\ldots,p'$ on pose alors :

(34) $$\hat{a}_\lambda = f(dp(u_\lambda))$$

et d'après le théorème 4 on a, en tenant compte de la remarque :

<u>Lemme</u> 3.-

> L'opérateur hermitien $f \circ dp \circ {}^t f \circ d\Phi_{mn}$ de L admet les mêmes valeurs propres non nulles $(\alpha_\lambda)_{\lambda=1,\ldots,p'}$ que $du \circ dp$. De plus les $\hat{a}_\lambda$ sont les vecteurs propres associés aux α_λ, et ils forment une famille Φ_{mn}-orthogonale. Enfin pour tout λ on a $\Phi_{mn}(\hat{a}_\lambda,\hat{a}_\lambda) = \|\hat{a}_\lambda\|^2_{\Phi_{mn}} = \alpha_\lambda$.

20. <u>Propriétés des applications</u> $\hat{a}_\lambda$.

On a vu au théorème 2 que les u_λ maximisent successivement

$$\gamma(x) = u(dp(x),dp(x))$$

sous des contraintes de p-orthogonalité pour les solutions en x. Pour tout $x \in E^*$ on pose :

(35) $$a(x) = f \circ dp(x)$$

et l'égalité (32) montre alors que :

$$\gamma(x) = \Phi_{mn}(f(dp(x)),f(dp(x))) = \Phi_{mn}(a(x),a(x)).$$

Par suite les $\hat{a}_\lambda = a(u_\lambda)$ maximisent $\|a(x)\|^2_{\Phi_{mn}}$ sous des contraintes de p-orthogonalité pour les solutions en x :

<u>Théorème</u> 5.-

> Le vecteur propre $\hat{a}_1$ de $dV \circ d\Phi_{mn}$ réalise le maximum de $\|a(x)\|^2_{\Phi_{mn}}$ sous la contrainte $p(x,x) = 1$. Pour tout λ, $1 < \lambda \leq p'$, le vecteur propre $\hat{a}_\lambda$ réalise le maximum de $\|a(x)\|^2_{\Phi_{mn}}$ sous les contraintes $p(x,x) = 1$ et $p(u_\nu,x) = 0$ pour tout $\nu < \lambda$.

21. <u>Matrice associée à la forme</u> u <u>et à l'application</u> du.

Soit $(e_j^*)_{j=1,\ldots,p}$ la base duale de la base canonique de E; les égalités

(30) et (32) prouvent que, pour tout j et tout ℓ, on a :

(36) $$u(e_j^*,e_\ell^*) = \Phi_{mn}(g_j,g_\ell).$$

Théorème 6.-

La matrice hermitienne U, de terme général $\Phi_{mn}(g_j,g_\ell)$ est la matrice associée à la forme u de E^*.

P étant la matrice de p dans la base canonique de E, la matrice $U.P$ est la matrice de l'opérateur hermitien $du \circ dp$ associé à la famille G. C'est le théorème 6 qui permettra de mettre en oeuvre effectivement les calculs proposés par la suite.

III. Décomposition d'une famille finie d'applications linéaires.

22. Le théorème de décomposition.

Théorème 7.-

Sous les hypothèses 5, du §.19, pour toute famille $G = (g_j)_{j=1,\dots,p}$ d'applications linéaires de G dans F, il existe :

i) une famile $(\hat{a}_\lambda)_{\lambda=1,\dots,p}$ d'applications linéaires (éventuellement nulles) de G dans F, Φ_{mn}-orthogonale, telles que

$$\Phi_{mn}(\hat{a}_\lambda,\hat{a}_\lambda) = \alpha_\lambda;$$

ii) une famille $(a_{j\lambda})_{j=1,\dots,p;\lambda=1,\dots,p}$ de nombres complexes vérifiant :

(37) $$g_j = \sum_{\lambda=1}^{p} \overline{a_{j\lambda}}.\hat{a}_\lambda$$

(38) $$\hat{a}_\lambda = f(dp(u_\lambda)) = \sum_{j=1}^{p} \pi_j^*(dp(u_\lambda)).g_j$$

(39) $$a_{j\lambda} = \pi_j(u_\lambda).$$

Montrons simultanément les égalités (37) et (39). Pour tout $j = 1,\dots,p$ on pose :

$$b_j = \sum_{\lambda=1}^{p} \overline{\pi_j(u_\lambda)}.\hat{a}_\lambda$$
$$= \sum_{\lambda=1}^{p} (\overline{\pi_j(u_\lambda)}. \sum_{\ell=1}^{p} \pi_\ell^*(d\ (u_\lambda)).g_\ell)$$

(40) $$b_j = \sum_{\ell=1}^{p} \left(\sum_{\lambda=1}^{p} \overline{\pi_j(u_\lambda)}.\pi_\ell^*(dp(u_\lambda))\right).g_\ell$$

Par ailleurs on note U la matrice p×p ayant pour λ-ème colonne les ccordonnées de u_λ relativement à la base canonique de E.

(41) $$U = (u_1,\dots,u_p) = (\pi_j(u_\lambda))_{\substack{j=1,\dots,p\\ \lambda=1,\dots,p}}.$$

On a alors :

(42) $${}^t\overline{U}.P.U = I \qquad\qquad \text{(I matrice unité p×p).}$$

U est la matrice de passage de la base canonique à la base (u_λ), elle est donc inversible et $U.U^{-1} = I$ peut s'écrire $\overline{U}.\overline{U}^{-1} = I$; alors (42) devient :

(43) $$\overline{U}.{}^t(P.U) = I.$$

Analytiquement cette dernière égalité s'écrit :

$$\sum_{\lambda=1}^{p} \overline{\pi_j(u_\lambda)}.\pi_\ell^*(dp(u_\lambda)) = \delta_{j\ell}.$$

L'égalité (40) devient alors $b_j = g_j$ et le théorème est démontré.

Corollaire 3.-

Sous les hypothèses 5, pour toute famille $= (g_j)_{j=1,\dots,p}$ d'applications linéaires de G dans F on a :

(44) $$g_j = \sum_{\lambda=1}^{p'} \sqrt{\alpha_\lambda}.\overline{\pi_j(u_\lambda)}.a_\lambda$$

(45) $$a_\lambda = \frac{1}{\sqrt{\alpha_\lambda}}.\hat{a}_\lambda$$

et la famille $(a_\lambda)_{\lambda=1,\dots,p'}$ est Φ_{mn}-orthonormée.

On a vu (remarque du théorème 4) que si $\alpha_\lambda = 0$, on a nécessairement $\hat{a}_\lambda = 0$, ce qui assure que la sommation dans (37) se réduit aux λ inférieurs à p'. Le reste est conséquence immédiate du théorème 7.

IV. Représentation d'une famille finie d'applications linéaires.

23. Théorème de représentation.

On se place sous les hypothèses 5. On rappelle que p' est le nombre de valeurs propres non nulles de du∘dp. Soit i le produit hermitien canonique de $C^{p'}$, ayant pour matrice $(\delta_{jj'})$ relativement à la base canonique de l'espace.

Soit ρ_j le vecteur de $E' = (C^{p'}, i)$ de composantes, par rapport à la base canonique :

(46) $$\sqrt{\alpha_\lambda}.\overline{\pi_j(u_\lambda)} \qquad \lambda = 1,\dots,p'$$

Dans ces conditions on obtient :

Théorème 8 (de représentation).-

(F,n) et (G,m) étant deux espaces hermitiens et i la forme canonique sur $C^{p'}$, pour toute famille finie $G = (g_j)_{j=1,\dots,p}$ d'applications linéaires de G dans F on a :

(47) $$i(\rho_j,\rho_\ell) = \Phi_{mn}(g_j,g_\ell).$$

En effet, à partir de (44) il vient :

$$\Phi_{mn}(g_j,g_\ell) = \sum_{\lambda=1}^{p'} \alpha_\lambda.\overline{\pi_\ell(u_\lambda)}.\pi_j(u_\lambda)$$

$$= i(\rho_j,\rho_\ell).$$

Ce théorème permet de représenter la famille (g_j) par la famille (ρ_j) de vecteurs de E', de façon telle que les i-distances de ces derniers pris deux à deux soient égales aux Φ_{mn}-distances des applications linéaires correspondantes. De plus, toutes les propriétés fondées sur des calculs de produit scalaire des ρ_j se transposent aux g_j et réciproquement.

24. Remarques.

Remarque 1.- La correspondance $g_j \in L \mapsto \rho_j \in E'$ peut-être étendue à toute application linéaire appartenant au sous-espace de dimension p' de L engendré par la famille (g_j), ou ce qui revient au même par la famille $(a_\lambda)_{\lambda=1,\dots,p'}$; pour cela il suffit de faire correspondre à une application linéaire g du sous-espace, de composantes $(g_\lambda)_{\lambda=1,\dots,p'}$, le vecteur de E' ayant les mêmes composantes dans la base canonique de cet espace. Dans cette correspondance, il est clair que les images des a_λ sont précisément les vecteurs de la base canonique de E'.

Remarque 2.- On considère les hypothèses suivantes, moins restrictives que les hypothèses 5 : G est un espace vectoriel sur C de dimension finie, F est un espace vectoriel quelconque (sur C), m est une forme sesquilinéaire à droite, hermitienne, non dégénérée, définie sur G et η est une forme sesquilinéaire à droite, hermitienne, définie sur F.

Dans ces conditions tous les théorèmes et corollaires de ce chapitre demeurent, sauf le théorème 5, le corollaire du théorème de décomposition et le théorème de représentation. Le théorème fondamental de décomposition reste valable.

25. Cas euclidien.

Si on remplace $\mathbb{C}$ par $\mathbb{R}$, tous les résultats de ce deuxième chapitre demeurent également, à condition de remplacer forme hermitienne par forme bilinéaire symétrique, et espace (resp. produit) hermitien par espace (resp. produit) euclidien.

Comme il est d'usage, en analyse des données, de travailler dans les réels, nous supposerons désormais $\mathbb{K} = \mathbb{R}$, bien que ce ne soit pas une nécessité théorique.

Chapitre III

ANALYSE EN COMPOSANTES PRINCIPALES D'UNE FAMILLE FINIE DE VECTEURS ALEATOIRES REELS DE DIMENSION m

I. Introduction.

26. Avertissement.

L'analyse générale présentée dans le sous-chapitre II introduira les bases algébriques de l'analyse en composantes principales d'une famille de vecteurs aléatoires réels de même longueur. Les termes de facteurs et de composantes principales seront définis dans cette section. L'analyse générale recherche une représentation euclidienne des vecteurs aléatoires (ou plus exactement de leurs réalisations sur un échantillon) possédant éventuellement suivant les produits euclidiens adoptés, des propriétés s'exprimant en terme d'inertie d'un nuage de points. Conjointement des propriétés de maximalisation et d'orthogonalité des composantes principales seront proposées. Les démonstrations de ces propriétés découlent directement des résultats établis dans le chapitre précédent. Ces propriétés n'ont pas véritablement d'interprétation statistique, cependant, comme en analyse classique, elles peuvent être utilisées pour mettre en valeur des effets de taille ou de moyenne.

La transformation des données initiales permet d'aborder l'analyse statistique. L'analyse des données centrées (chaque réalisation des variables aléatoires composantes des vecteurs aléatoires est centrée) permet d'interpréter l'analyse générale en terme de maximum de variance et d'indépendance linéaire. On proposera donc une définition de la covariance de deux vecteurs aléatoires; lorsque les vecteurs sont réduits à une seule composante, on retrouve la covariance usuelle des variables aléatoires. Cette définition ne coïncide pas avec celle donnée par ESCOUFFIER ([8] et [9]), comme on le verra dans un autre chapitre.

L'analyse des données normalisées permet de retrouver toutes les propriétés de l'analyse classique correspondante.

27. Notations.

On considère une famille $V_j = (V_j^k)_{k=1,\dots,m}$ de p vecteurs aléatoires de longueur m, et les réalisations de ces vecteurs sur un échantillon $I = \{1,\dots,n\}$. On pose $J = \{1,\dots,p\}$ et $K = \{1,\dots,m\}$ et on note X_j la matrice $n\times m$ exprimant la réalisation du vecteur V_j sur I.

Soient trois espaces euclidiens $E = (R^p,\mathrm{p})$, $F = (R^n,\mathrm{n})$ et $G = (R^m,\mathrm{m})$ où p, n, m désignent les produits euclidiens et P, N, M les matrices carrées associées par rapport aux bases canoniques. Les espaces duals sont rapportés aux bases duales. Ainsi, comme il est d'usage en analyse des données rectangulaires, les indices i et j se rapportent respectivement aux éléments de l'échantillon et aux unités aléatoires.; l'indice k introduit la troisième dimension.

On désigne par x_j l'application linéaire de G^* dans F définie par la matrice X_j et par m^{-1} la forme inverse de m définie sur G^* (cf. §.5). Pour toute forme bilinéaire symétrique les applications associées s et d étant identiques on a :

$$d(\mathrm{m}^{-1}) = (d\mathrm{m})^{-1} \tag{48}$$

qu'on notera sans ambiguïté $d\mathrm{m}^{-1}$.

Le produit euclidien $\Phi_{\mathrm{m}^{-1}\mathrm{n}}$ défini sur $L_R(G^*,F)$ est donné par :

$$\Phi_{\mathrm{m}^{-1}\mathrm{n}}(f,g) = \mathrm{Trace}(g^*\circ f) = \mathrm{Trace}(d\mathrm{m}\circ {}^tg\circ d\mathrm{n}\circ f). \tag{49}$$

En particulier pour tout j et tout ℓ de J on a :

$$\Phi_{\mathrm{m}^{-1}\mathrm{n}}(x_j,x_\ell) = \mathrm{Trace}(M.{}^tX_\ell.N.X_j). \tag{50}$$

Pour simplifier les écritures on notera Φ au lieu de $\Phi_{m^{-1}n}$ et on dira quelquefois "vecteurs aléatoires" au lieu de "réalisation des vecteurs aléatoires sur l'échantillon I".

II. Analyse générale.

28. Inertie d'un nuage de points.

Pour simplifier on confondra espace vectoriel et espace affine associé. Ainsi on dira nuage de points au lieu de nuage de vecteurs. Dans un même souci, si X est une matrice n×p, si N_I est le nuage de points de $E = (R^p, p)$ associé aux n lignes de X, et si f est l'application linéaire de matrice X, on dira indifféremment, opérateur hermitien associé à f, à X ou à N_I. Dans ce cadre on a le schéma de dualité suivant :

$$\begin{tikzcd} E^* \arrow[r, "f"] \arrow[d, shift left, "du"] & F \arrow[d, "dn"] \\ E \arrow[u, shift left, "dp"] & F^+ \arrow[l, "{}^t f"] \end{tikzcd}$$

On a vu que les vecteurs propres p-orthonormés $(u_\lambda)_{\lambda=1,\dots,p}$ de l'opérateur $du \circ dp = {}^t f \circ dn \circ f \circ dp$ maximisent :

$$\gamma(x) = u(dp(x), dp(x)) = n(f(dp(x)), f(dp(x))).$$

Par suite, si $N = (n_{i\ell})$ on a :

$$\gamma(x) = \sum_{i=1}^{n} \sum_{\ell=1}^{n} n_{i\ell} \cdot p(x_i, x) \cdot p(x_\ell, x),$$

où $(x_i)_{i=1,\dots,n}$ désignent les points du nuage N_I de E.

Si des poids positifs p_i ($\Sigma p_i = 1$) sont attribués aux n points du nuage N_I et si N est la matrice diagonale des poids D_{P_I} (resp. la matrice unité) on a :

$$\gamma(x) = \sum_{i=1}^{n} p_i \cdot p^2(x_i, x)$$

$$\text{(resp. } \gamma(x) = \sum_{i=1}^{n} p^2(x_i, x)).$$

Dans les deux cas cette quantité représente l'inertie du nuage N_I relativement à l'hyperplan p-orthogonal à x. On la note $I_{N_I}(\Delta_x^\perp)$.

Théorème 1.-

> Si N est la matrice diagonale des poids ou la matrice unité, pour tout x de E on a : $\gamma(x) = I_{N_I}(\Delta_x^{\perp})$.

D'après le théorème 2 du chapitre II, il vient immédiatement :

Corollaire 1.-

> Si N est la matrice diagonale des poids ou la matrice unité, les vecteurs propres de l'opérateur $du \circ dp$ associé à f sont les axes principaux d'inertie du nuage correspondant aux lignes de la matrice de f.

29. Image euclidienne de la famille de vecteurs aléatoires.

En revenant aux notations du §.27, on note à présent f_J l'application linéaire de E^* dans $L = (L_R(G^*,F),\Phi)$ associée à la famille $(x_j)_{j\in J}$ et définie par (30), et u la forme bilinéaire symétrique définie sur E^* par (32) :

(51) $$x \in E^*, \forall x \qquad f_J(x) = \sum_{j\in J} \pi_j^*(x).x_j$$

(52) $$x \in E^*, y \in E^*, \forall x, \forall y \qquad u(x,y) = \Phi(f_J(x),f_J(y)).$$

Comme dans le chapitre précédent $(\alpha_\lambda)_{\lambda=1,\dots,p}$ et $(u_\lambda)_{\lambda=1,\dots,p}$ sont respectivement valeurs propres (ordonnées) et vecteurs propres p-orthonormés de l'opérateur $du \circ dp$ associé à f_J; p' est le nombre de valeurs propres non nulles; E' est l'espace euclidien $(R^{p'},i)$ et N_J est le nuage $(\rho_j)_{j\in J}$ où ρ_j est le vecteur de E' de composantes $(\sqrt{\alpha_\lambda}.\pi_j(u_\lambda))_{\lambda=1,\dots,p'}$ par rapport à la base canonique de E'.

On rappelle alors (cf. [4]) les définitions d'image euclidienne et d'images euclidiennes équivalentes :

Image euclidienne : N étant un nuage de points de E muni du produit scalaire p, et J étant un ensemble muni d'une distance d_J, le triplet (N,E,p) est une image euclidienne de (J,d_J) s'il existe une application $h : J \to E$ vérifiant : i) $h(J) = N$ et ii) $d_J(j,j') = \|h(j)-h(j')\|_p$ pour tous j et j' de J.

Images euclidiennes équivalentes : (N,E,p) et (N',E',p') sont deux images euclidiennes équivalentes de (J,d_J) s'il existe une application bijective $f : N \to N'$ telle que :

$$\| f(x_i)-f(x_{i'})\|_{p'} = \|x_i - x_{i'}\|_p \text{ pour tous } x_i \text{ et } x_{i'} \text{ de } N.$$

L'application qui à x_j fait correspondre ρ_j est bijective; du théorème de représentation on déduit le résultat fondamental :

<u>Théorème</u> 2.-

Les images euclidiennes $(N_J, R^{p'}, \mathrm{i})$ et $((x_j)_{j\in J}, L_R(G^*,F), \Phi)$ sont équivalentes.

N_J sera appelé la représentation (ou le nuage) des vecteurs aléatoires V_j.

30. <u>Propriété de l'image euclidienne des vecteurs aléatoires.</u>

A) <u>Propriétés matricielles.</u>

Compte tenu de l'inutilité (ch.II.,th.4; Remarque) des vecteurs propres u_λ, pour $\lambda > p'$, on notera désormais U (resp. $\hat{U}$) la matrice $p\times p'$ dont les vecteurs colonnes sont les vecteurs propres u_λ, $\lambda = 1,\dots,p'$ (resp. les vecteurs $\hat{u}_\lambda$, $\lambda = 1,\dots,p'$) :

(53) $$\hat{u}_\lambda = \sqrt{\alpha_\lambda}.u_\lambda$$

(54) $$U = (u_1,\dots,u_{p'})$$

(55) $$\hat{U} = (\hat{u}_1,\dots,\hat{u}_{p'}).$$

D_α étant la matrice carrée $p'\times p'$ diagonale des valeurs propres non nulles (ordonnées), il est clair que :

(56) $$\hat{U} = U.D_\alpha^{1/2}.$$

Les vecteurs propres $(u_\lambda)_{\lambda=1,\dots,p'}$ étant p-orthonormés, la matrice de leurs produits scalaires dans E est la matrice unité. Par ailleurs, d'après le théorème de représentation, la matrice des produits scalaires des ρ_j dans E' est la matrice $(\Phi(x_j,x_{j'}))$, c'est à dire la matrice $\mathcal{U}$(cf. théorème 6 du §.21). On peut donc écrire :

Lemme 1.-

Les matrices p×p', $\mathcal{U}$ et $\hat{\mathcal{U}}$ vérifient les propriétés suivantes :

(57) $${}^t\mathcal{U}.P.\mathcal{U} = I$$

(58) $${}^t\hat{\mathcal{U}}.P.\hat{\mathcal{U}} = D_\alpha$$

(59) $$\hat{\mathcal{U}}.{}^t\hat{\mathcal{U}} = U.$$

$\hat{\mathcal{U}}$ est la matrice exprimant, en ligne, les composantes des ρ_j du nuage N_J dans E' :

$$\begin{tikzcd} E'^* \arrow[r, "\hat{\mathcal{U}}"] & E \arrow[d, "P"] \\ E' \arrow[u, "I"] & E^* \arrow[l, "{}^t\hat{\mathcal{U}}"] \end{tikzcd}$$

par suite l'opérateur associé à N_J a pour matrice :

$${}^t\hat{\mathcal{U}}.P.\hat{\mathcal{U}}.I = D_\alpha \;;$$

ses valeurs propres sont donc $(\alpha_\lambda)_{\lambda=1,\dots,p'}$ et les vecteurs $(e'_\lambda)_{\lambda=1,\dots,p'}$ de la base canonique de E' constituent un système i-orthonormé de vecteurs propres correspondants.

Lemme 2.-

Les vecteurs propres de l'opérateur hermitien de E' associé au nuage N_J correspondants aux valeurs propres $(\alpha_\lambda)_{\lambda=1,\dots,p'}$, sont les vecteurs de la base canonique de E, et sont i-orthonormés.

B) Axes d'inertie du nuage N_J.

On suppose ici que des poids réels positifs ($\Sigma p_j = 1$) sont attribués aux vecteurs aléatoires et que P est la matrice diagonale des poids D_{P_J} (ou éventuellement la matrice unité). D'après le lemme 2 et le corollaire 1 du §.28, les vecteurs propres (e'_λ) de l'opérateur associé à N_J sont les axes principaux du nuage :

Théorème 3.-

Si P est la matrice diagonale des poids attribués aux vecteurs aléatoires (ou la matrice unité) les vecteurs $(e'_\lambda)_{\lambda=1,\dots,p'}$ de la base canonique de E' = $(R^{p'},i)$ sont les axes principaux d'inertie du nuage pondéré $N_J = (\rho_j)_{j\in J}$. De plus, pour tout $\lambda = 1,\dots,p'$ on a $I_{N_J}(\Delta_{e'}^{\perp}) = \alpha_\lambda$.

Ce théorème permet donc d'interpréter la représentation N_J des vecteurs aléatoires en terme d'inertie.

C) Centrage du nuage N_J.

Définition.-

i) $x_J = \sum_{j\in J} p_j . x_j$ est appelé homomorphisme moyen associé à la famille $((x_j,p_j))_{j\in J}$ d'homomorphismes pondérés.

ii) La famille $((x_j,p_j))_{j\in J}$ est dite centrée ssi $x_J = 0$.

Il est clair que :

Proposition.-

Pour toute famille $((x_j,p_j))_{j\in J}$, la famille $((x_j-x_J,p_j))_{j\in J}$ est centrée.

On suppose à présent que la famille $(x_j,p_j)_{j\in J}$ associée aux vecteurs aléatoires pondérés V_j est centrée. Comme en (34) et (45) on pose alors :

(60) $\hat{a}_\lambda = f_J(dp(u_\lambda)) \qquad \lambda = 1,\dots,p'$

(61) $a_\lambda = \dfrac{1}{\sqrt{\alpha_\lambda}}.\hat{a}_\lambda \qquad \lambda = 1,\dots,p'$.

Pour tout j de J le corollaire du théorème de décomposition permet d'écrire :

$$x_j = \sum_{\lambda=1}^{p'} \sqrt{\alpha_\lambda}.\pi_j(u_\lambda).a_\lambda$$

$$x_J = \sum_{j\in J} p_j . \sum_{\lambda=1}^{p'} \sqrt{\alpha_\lambda}.\pi_j(u_\lambda).a_\lambda$$

$$= \sum_{\lambda=1}^{p'} \Big(\sum_{j\in J} p_j.\sqrt{\alpha_\lambda}.\pi_j(u_\lambda)\Big).a_\lambda = 0.$$

Le système $(a_\lambda)_{\lambda=1,\dots,p'}$ étant Φ-orthonormé, on en déduit pour tout $\lambda \leq p'$:

$$\sum_{j\in J} p_j(\sqrt{\alpha_\lambda}.\pi_j(u_\lambda)) = \sum_{j\in J} p_j.\pi'_\lambda(\rho_j) = 0$$

où π'_λ désigne la λ-ème fonction coordonnée de E'.

Cette dernière égalité exprime que la matrice $\hat{U}$ est centrée en colonnes pour D_{p_J} , c'est à dire que le nuage N_J est centré. La réciproque se démontre de façon identique.

Théorème 4.-

> La matrice P associée au produit scalaire p de l'espace E étant la matrice diagonale des poids, le nuage pondéré N_J représentation des vecteurs aléatoires V_j est centré ssi la famille $((x_j,p_j))_{j\in J}$ est centrée.

31. Facteurs et composantes principales.

Pour tout x de E on note $A(x)$ le vecteur aléatoire de longueur m défini par :

(62) $$A(x) = \sum_{j\in J} \pi_j^*(dp(x)).V_j$$

et $A(x)$ la matrice :

(63) $$A(x) = \sum_{j\in J} \pi_j^*(dp(x)).X_j,$$

matrice de sa réalisation sur l'échantillon I. Par abus d'écriture on notera aussi quelquefois :

(64) $$A(x) = f_J(dp(x))$$

confondant ainsi application linéaire et matrice associée.

Pour tout $\lambda = 1,\dots,p'$ on pose :

(65) $$\hat{A}_\lambda = A(u_\lambda)$$

(66) $$\hat{A}_\lambda = f_J(dp(u_\lambda)).$$

Définition.-

> Le vecteur aléatoire $\hat{A}_\lambda$ est appelé le λ-ème facteur, et la matrice $\hat{A}_\lambda$ de ses réalisations sur I est appelée la λ-ème composante principale.

Remarque importante.- En analyse des données rectangulaires il est traditionnel (cf. [1]) d'appeler facteur l'élément $dp(u_\lambda)$ du dual E^*. Ceci provient du fait que les éléments de l'échantillon I étant considérés comme élément de E, une variable aléatoire est prise comme une application linéaire de E dans R, c'est à dire comme un élément du dual. Dès lors il est légitime de confondre les p vecteurs de la base canonique du dual avec les p variables aléatoires.

Dans notre cas, ces considérations ne sont pas valables, et c'est pourquoi nous préférons la définition ci-dessus, où un facteur est une combinaison linéaire des vecteurs aléatoires initiaux, les coefficients de la combinaison restant les composantes d'un élément du dual E^*.

Reparquons de plus que pour tous j et j' de J on a :

(67) $$\Phi(x_j, x_{j'}) = u(e_j^*, e_{j'}^*) = i(\rho_j, \rho_{j'})$$

et pour tous λ et λ' inférieurs ou égaux à p' :

(68) $$\Phi(\hat{a}_\lambda, \hat{a}_{\lambda'}) = u(dp(u_\lambda), dp(u_{\lambda'})) = i(\sqrt{\alpha_\lambda}.e'_\lambda, \sqrt{\alpha_{\lambda'}}.e'_{\lambda'}).$$

Enfin on a :

$$\begin{aligned} p^{-1}(dp(u_\lambda), dp(u_{\lambda'})) &= dp^{-1} \circ dp(u_{\lambda'})(dp(u_\lambda)) \\ &= u_{\lambda'}(dp(u_\lambda)) \\ &= dp(u_\lambda)(u_{\lambda'}) \\ &= p(u_\lambda, u_{\lambda'}) \\ &= \delta_{\lambda\lambda'}. \end{aligned}$$

Par suite, pour des éléments correspondant en λ et λ', u_λ, $dp(u_\lambda)$, $\hat{a}_\lambda$ et $\sqrt{\alpha_\lambda}.e'_\lambda$, la p-orthogonalité dans E, la p^{-1}-orthogonalité dans E^*, la Φ-orthogonalité dans L et la i-orthogonalité dans E' sont équivalentes.

Le schéma ci-dessous, en forme de double schéma de dualité, résume les espaces et les applications utilisées :

$$\begin{tikzcd} E' = (R^{p'}, i) \arrow[d, "di"'] & E^* = (R^{p^*}, p^{-1}) \arrow[l, "{}^t\hat{u}"'] \arrow[r, "f_J"] \arrow[d, shift right, "du"'] & L = (L_R(G^*,F), \Phi) \arrow[d, "d\Phi"] \\ E'^* \arrow[r, "\hat{u}"'] & E = (R^p, p) \arrow[u, shift right, "dp"'] & L^* \arrow[l, "{}^t f_J"] \end{tikzcd}$$

32. Propriété des facteurs et des composantes principales.

Théorème 5.-

Les facteurs $\hat{A}_\lambda = A(u_\lambda)$, $\lambda = 1,\dots,p'$ sont de tous les vecteurs aléatoires $A(x) = \sum_{j\in J} \pi_j^*(dp(x)).V_j$, combinaisons linéaires des p vecteurs initiaux, ceux qui réalisent successivement le maximum de la Φ-norme de leur réalisation sur l'échantillon I, sous les contraintes de p-orthonormalité portant sur les solutions en x. En outre les composantes principales sont Φ-orthogonales et pour tout $\lambda = 1,\dots,p'$ on a : $\|\hat{A}_\lambda\|^2_\Phi = \alpha_\lambda$.

Ce théorème reprend simplement dans le cas qui nous intéresse ici les résultats généraux décrits par le lemme 3 du §.19 et le théorème 5 du §.20.

33. Interprétation.

En pratique, l'analyse en composantes principales d'une famille finie de vecteurs aléatoires s'appuie sur l'interprétation que l'on donne au théorème 5. Cette interprétation dépend elle-même du sens que l'on prête à la Φ-norme d'une matrice $n\times m$, et plus généralement au sens du Φ-produit de deux matrices $n\times m$, considérées comme réalisations de deux nouveaux vecteurs aléatoires, combinaisons linéaires des anciens. Or pour toutes matrices A et A' :

$$\Phi(A,A') = \mathrm{Trace}(M.{}^tA'.N.A).$$

L'interprétation statistique proprement dite dépendra donc en fin de compte à la fois du choix des matrices M et N, et des transformations éventuelles appliquées a priori aux matrices A et A'.

Dans les sous-chapitre III, IV et V suivants nous allons proposer un choix pour les matrices M et N, et des transformations particulières à appliquer aux réalisations des vecteurs V_j. Ces choix nous conduisent à proposer une définition de la covariance et du coefficient de corrélation de deux matrices. Ces définitions ont l'avantage de généraliser celles correspondant au cas usuel des variables aléatoires.et, en outre, de s'interpréter simplement en terme de covariance empirique usuelle.

III. Analyse en terme de variance.

34. Covariance de deux vecteurs aléatoires.

On dira que la matrice $n\times m$: X_j, réalisation du vecteur V_j sur l'échantillon I est centrée si les vecteurs de R^n, réalisations de chaque variable aléatoire composante de V_j, sont centrés pour le produit scalaire associé à la matrice diagonale des poids $N = D_{P_I}$.

Si X_j est la réalisation de V_j sur I, on note X_j' la matrice X_j centrée pour D_{P_I} au sens ci-dessus. Cela signifie que X_j' est la réalisation du vecteur $V_j' = (V'^k_j)_{k\in K}$, où chaque variable V'^k_j est centrée au sens usuel.

Les notations $\mathrm{cov}(V_j^k, V_j^\ell)$ et $\mathrm{var}(V_j^k)$ désignent respectivement la covariance et la variance empiriques des variables correspondantes. Enfin Φ désignera tout aussi bien le produit des applications linéaires que celui des matrices correspondantes. Dans ces conditions on pose :

Définition.-

(69) $$\mathrm{COV}(V_j, V_\ell) = \Phi(X_j', X_\ell')$$

(70) $$\mathrm{VAR}(V_j) = \mathrm{COV}(V_j, V_j) = \|X_j'\|^2_\Phi .$$

On suppose maintenant que des poids positifs p_k ($\Sigma p_k = 1$) sont attribués aux différentes composantes des vecteurs V_j. On pose $M = D_{P_K}$. Alors :

$$\Phi(X_j', X_\ell') = \mathrm{Trace}(D_{P_K} . {}^tX_\ell' . D_{P_I} . X_j').$$

Les vecteurs colonnes de X_j' et X_ℓ' étant centrés pour D_{P_I}, le terme général de $D_{P_K} . {}^tX_\ell' . D_{P_I} . X_j'$ est alors :

$$a_{kk'} = p_k . \mathrm{cov}(V_j^{k'}, V_j^k),$$

d'où le théorème :

Théorème 6.-

Si $N = D_{P_I}$ et $M = D_{P_K}$ on a :

(71) $$\mathrm{COV}(V_j, V_\ell) = \sum_{k\in K} p_k . \mathrm{cov}(V_j^k, V_\ell^k)$$

(72) $$\mathrm{VAR}(V_j) = \sum_{k\in K} p_k . \mathrm{var}(V_j^k).$$

On a de même :

Théorème 7.-

Si $N = D_{P_I}$ et $M = I$:

$$\left| \begin{array}{ll} (73) & COV(V_j,V_\ell) = \sum_{k\in K} cov(V_j^k,V_\ell^k) \\ (74) & VAR(V_j) = \sum_{k\in K} var(V_j^k). \end{array} \right.$$

Dans les §.36,37 et 38 on se placera dans le cas $M = D_{P_K}$. Sous forme de remarque à la fin du §.38 on indiquera les modifications à apporter aux résultats lorsque $M = I$.

35. Définition de l'analyse en C.P en terme de variance.

On appellera analyse en composantes principales en terme de variance de la famille $(V_j)_{j\in J}$ (ou de la famille $(X_j)_{j\in J}$) ou plus simplement analyse des données centrées l'analyse générale correspondante lorsque $N = D_{P_I}$ et $M = D_{P_K}$, et lorsque les réalisations des vecteurs V_j sont centrées pour D_{P_I}.

Il ne faut pas confondre le centrage de la famille $(X_j)_{j\in J}$ pour D_{P_J} (correspondant au centrage du nuage N_J) étudié au §.30, avec le centrage de chaque matrice X_j pour D_{P_I} proposé ici. Ces deux opérations ne sont pas exclusives l'une de l'autre. Cependant le centrage de la famille $(X_j)_{j\in J}$ pour D_{P_J} modifie évidemment les produits scalaires des matrices et par suite l'analyse elle-même. On ne doit donc procéder à cette opération que si celle-ci a un sens statistique.

36. Représentation des vecteurs aléatoires.

Soit N_J la représentation des vecteurs aléatoires issue de l'analyse en terme de variance. Les X_j étant par définition centrées pour D_{P_I}, (69) fournit :

(75) $$\Phi(X_j,X_\ell) = COV(V_j,V_\ell)$$

donc du théorème 6 du §.21 on obtient :

Propriété 1.-

> La matrice U associée à la forme u définie sur E^* est la matrice des variances-covariances des vecteurs aléatoires.

et le théorème de représentation donne :

Propriété 2.-

> $N_j = (\rho_j)_{j\in J}$ étant la représentation des vecteurs aléatoires, pour tout j et tout ℓ de J on a :
>
> (76) $$i(\rho_j,\rho_\ell) = COV(V_j,V_\ell)$$
>
> (77) $$\|\rho_j\|_i^2 = VAR(V_j).$$

Cette propriété qui permet de passer de considérations géométriques portant sur les représentations obtenues à des considérations statistiques sur les données constitue une aide à l'interprétation.

Enfin comme précédemment, si $P = D_{P_J}$ ou $P = I$ les axes de la représentation sont les axes d'inertie du nuage N_J.

37. Interprétation statistique.

Ce sera simplement la traduction du théorème 5 du §.32 :

Théorème 8.-

> Les facteurs $\hat{A}_\lambda = A(u_\lambda)$, $\lambda = 1,\dots,p'$ sont de tous les vecteurs aléatoires $A(x) = \sum_{j\in J} \pi_j^*(dP(x)).V_j$ combinaisons linéaires des p vecteurs initiaux ceux qui réalisent successivement le maximum de la variance de leur réalisation sur l'échantillon I, sous les contraintes de P-orthonormalité portant sur les solutions en x. En outre, les covariances inter-composantes principales sont nulles et pour tout $\lambda = 1,\dots,p'$ on a $VAR(\hat{A}_\lambda) = \alpha_\lambda$.

D'un point de vue statistique c'est ce théorème qui donne son sens à l'analyse. On verra ci-après en IV que la nullité des covariances des composantes principales s'interprète également comme une non-corrélation linéaire. Remarquons enfin que chaque composante principale est centrée pour D_{P_I}.

On notera maintenant $ca_\lambda(V_j)$ la contribution absolue de j-ème vecteur V_j au λ-ème facteur, et $cr_\lambda(V_j)$ la contribution relative du λ-ème facteur au j-ème vecteur. Comme en analyse en C.P. classique on posera lorsque $P = D_{P_J}$:

(78) $$ca_\lambda(V_j) = p_j . \pi_j^2(u_\lambda)$$

(79) $$cr_\lambda(V_j) = \frac{\alpha_\lambda . \pi_j^2(u_\lambda)}{\| X_j \|_\Phi^2} = \frac{\alpha_\lambda . \pi^2(u_\lambda)}{\| \rho_j \|_1^2} .$$

Les propriétés de ces coefficients sont les propriétés usuelles.

38. Espaces factoriels de représentations conjointes.

En analyse en C.P. classique on cherche, entre autre, à expliciter les proximités et les oppositions des éléments relatifs à un indice en fonction des éléments relatifs au deuxième indice. Il s'agit notamment des représentations conjointes. On espérait ici produire un espace de représentations conjointes contenant les éléments relatifs aux trois indices. Comme on le verra plus loin il ne semble pas que cela soit possible car dans le cas des données cubiques on perd la propriété fondamentale de dualité qui exprime le fait que les deux opérateurs hermitiens duaux possèdent les mêmes valeurs propres. Cependant d'après le corollaire du théorème de décomposition pour tout j on peut écrire :

(80) $$X_j = \sum_{\lambda=1}^{p'} X_j^\lambda$$

où

(81) $$X_j^\lambda = \pi_j(u_\lambda) . \hat{A}_\lambda .$$

Toutes les matrices X_j s'exprimant comme combinaison linéaire des $\hat{A}_\lambda$, avec la part d'inertie que l'on sait, on en déduit que la matrice $\hat{A}_\lambda$ doit permettre d'expliquer comment les éléments relatifs aux deux autres indices I et K ont contribués à la séparation des V_j le long du λ-ème axe factoriel.

Ces considérations nous amènent à proposer les définitions suivantes :

i) On dira que la matrice $\hat{A}_\lambda$ est la λ-ème matrice de référence des vecteurs aléatoires;

ii) on appelle λ-ème espace factoriel de représentation conjointe relative aux vecteurs aléatoires, l'espace de la représentation conjointe issu de l'analyse en composantes principales classique de la matrice de référence $\hat{A}_\lambda$.

A la matrice de référence $\hat{A}_\lambda$, de Φ-norme $\sqrt{\alpha_\lambda}$ est associé le schéma de dualité suivant :

$$\begin{array}{ccc} G^* & \xrightarrow{\hat{A}_\lambda} & F \\ {\scriptstyle D_{P_K}=M}\uparrow & & \downarrow{\scriptstyle N=D_{P_I}} \\ G & \xleftarrow[{}^t\hat{A}_\lambda]{} & F^* \end{array}$$

N_I^λ est le nuage de G associé aux n lignes de $\hat{A}_\lambda$;

N_K^λ est le nuage de F associé aux m colonnes de $\hat{A}_\lambda$.

Puisque chaque composante principale $\hat{A}_\lambda$ est centrée (en colonne) pour D_{P_I}, on obtient :

<u>Propriété</u> 1.-

> Pour tout λ, le nuage N_I^λ du λ-ème espace factoriel de représentation conjointe est centré.

N étant la matrice diagonale des poids on a d'après le corollaire 1 du §.28 :

<u>Propriété</u> 2.-

> Les axes factoriels du λ-ème espace factoriel de représentation conjointe sont également les axes principaux d'inertie du nuage N_I^λ.

Rappelons que ces axes factoriels sont les vecteurs propres M-orthonormés de l'opérateur hermitien associé à $\hat{A}_\lambda$, de matrice : ${}^t\hat{A}_\lambda . D_{P_I} . \hat{A}_\lambda . D_{P_K}$.
La matrice M étant également une matrice de poids la propriété 2 est aussi valable pour le nuage N_K^λ :

<u>Propriété</u> 3.-

> Les axes factoriels du λ-ème espace de représentation conjointe sont les axes principaux d'inertie du nuage N_K^λ.

Soit $\hat{A}_\lambda$ le λ-ème facteur; on pose pour tout $\lambda = 1,\ldots,p'$:

(82) $\qquad \hat{A}_\lambda = (V_\lambda^1,\ldots,V_\lambda^m) = (V_\lambda^k)_{k\in K}$

et d'après (65) chaque variable aléatoire V_λ^k est donnée par :

(83) $\qquad V_\lambda^k = \sum_{j\in J} \pi_j^*(dp(u_\lambda)).V_j^k$.

Par suite :

Propriété 4.-

(84) La matrice de l'opérateur hermitien associé à la λ-ème matrice de référence $\hat{A}_\lambda$ est la matrice des variances-covariances pondérées des m variables aléatoires V_λ^k :

$$^t\hat{A}_\lambda . D_{P_I} . \hat{A}_\lambda . D_{P_K} = (p_{k'} . \mathrm{cov}(V_\lambda^k, V_\lambda^{k'}))_{k\in K, k'\in K} .$$

Soient $(\alpha_{\lambda\nu})_{\nu=1,\dots,m}$ les valeurs propres de l'opérateur hermitien associé à $\hat{A}_\lambda$. On connait le résultat classique de l'analyse en C.P. d'une matrice :

$$\sum_{\nu=1}^{m} \alpha_{\lambda\nu} = \mathrm{Trace}(^t\hat{A}_\lambda . D_{P_I} . \hat{A}_\lambda . D_{P_K}).$$

Or :

$$\begin{aligned} \mathrm{Trace}(^t\hat{A}_\lambda . D_{P_I} . \hat{A}_\lambda . D_{P_K}) &= \mathrm{Trace}(D_{P_K} . {}^t\hat{A}_\lambda . D_{P_I} . \hat{A}_\lambda) \\ &= \Phi(\hat{A}_\lambda, \hat{A}_\lambda) \\ &= \alpha_\lambda . \end{aligned}$$

Théorème 9.-

Pour tout $\lambda = 1,\dots,p'$ on a :

(85) $\sum_{\nu\in K} \alpha_{\lambda\nu} = \alpha_\lambda$ i.e.

l'inertie expliquée par la totalité des facteurs dans le λ°espace factoriel de représentation conjointe est égale à l'inertie expliquée par le λ-ème facteur relatif à la famille $(V_j)_{j\in J}$ de vecteurs aléatoires.

Remarque 1.- On aurait pu prendre la matrice Φ-normée A_λ comme matrice de référence au lieu de $\hat{A}_\lambda$. Les nuages N_I^λ et N_K^λ correspondants se déduiraient alors simplement des précédents par une homothétie centrale de rapport $\sqrt{\alpha_\lambda}^{-1}$, ce qui ne change rien à l'interprétation. Par contre la relation (85) devient : $\sum_{\nu\in K} \alpha_{\lambda\nu} = 1$. Dans cette optique on perd l'importance relative des différents facteurs des espaces factoriels de représentation conjointe, de l'un à l'autre de ces espaces. Il n'y a pas en effet, sur l'ensemble $\{\alpha_{\lambda\nu}\}_{\substack{\lambda=1,\dots,p' \\ \nu\in K}}$ de relation d'ordre compatible avec les relations d'ordre naturelles des ensembles de variation des indices λ et ν.

Remarque 2.- Pour un λ donné l'annalyse en C.P. des différentes matrices X_j^λ, $j \in J$ permet de représenter dans le λ-ème espace factoriel les ensembles I et K relatifs à chaque j. Pour tout j les vecteurs propres de l'opérateur hermitien associé à X_j^λ sont en effet identiques à ceux de l'opérateur hermitien associé à $\hat{A}_\lambda$, comme le prouvent les relations (81). Ces mêmes relations montrent que pour λ donné, et pout tout j les nuages N_I^λ et N_K^λ relatifs à j sont homothétiques des nuages N_I^λ et N_K^λ relatifs à $\hat{A}_\lambda$, dans le rapport $\pi_j(u_\lambda)$. De telles représentations ne sont donc pas souhaitables puisque les informations portant sur les différences entre les vecteurs V_j, relativement au λ-ème facteur, sont apportées par le λ-ème axe factoriel du nuage N_J.

Remarque 3.- Si on pose M = I au lieu de D_{P_K}, les résultats de l'analyse présentée aux paragraphes précédents restent valablesà ceci près :

i) la conduite de l'analyse, on l'a vu, repose sur la recherche des éléments propres de la matrice U.P. Si M = I la matrice U des variances-covariances des vecteurs a pour terme général $\sum_{k \in K} cov(V_j^k, V_{j'}^k)$ au lieu de $\sum_{k \in K} p_k . cov(V_j^k, V_{j'}^k)$.

ii) Dans la propriété 4 du §.38 la matrice de l'opérateur hermitien associé à la λ-ème matrice de référence $\hat{A}_\lambda$ est la matrice ${}^t\hat{A}_\lambda . D_{P_I} . \hat{A}_\lambda$ de terme général $cov(V_\lambda^k, V_\lambda^{k'})$.

IV. Corrélations vectorielles et normalisations.

39. Définition.

On suppose N = D_{P_I}. Comme précédemment X_j' et X_ℓ' désignent les matrices X_j et X_ℓ centrées pour D_{P_I}.

On appelle coefficient de corrélation vectoriel le réel noté $COR(V_j, V_\ell)$ et défini par :

$$(86) \qquad COR(V_j, V_\ell) = \frac{COV(V_j, V_\ell)}{\sqrt{VAR(V_j) . VAR(V_\ell)}}$$

De (69) et (70) on déduit :

(87) $$\mathrm{COR}(V_J, V_\ell) = \frac{\Phi(X'_j, X'_\ell)}{\|X'_j\|_\Phi \|X'_\ell\|_\Phi} .$$

40. Propriétés du coefficient COR.

On a clairement :

(88) $$-1 \leqq \mathrm{COR}(V_j, V_\ell) \leqq +1.$$

On suppose toujours $M = D_{P_K}$:

Théorème 10.-

> $\mathrm{COR}(V_j, V_\ell) = +1$ (resp. -1) si et seulement si $X_\ell = aX_j + B$, où a est positif (resp. négatif) et B est une matrice $n\times m$ ayant toutes ses lignes égales, ou ce qui revient au même ayant pour colonnes des vecteurs constants.

L'interprétation statistique de ce théorème est que, dans ces conditions, tout couple (V_j^k, V_ℓ^k) de deux composantes de même rang des vecteurs V_j et V_ℓ est formé de deux variables aléatoires liées fonctionnelement par une relation affine. En outre toutes ces fonctions ont même coefficient angulaire.

Preuve : supposons $\mathrm{COR}(V_j, V_\ell) = +1$, ce qui d'après (86), (71) et (72) est équivalent à :

(89) $$\sum_{k\in K} p_k . \mathrm{cov}(V_j^k, V_\ell^k) = \sqrt{\sum_{k\in K} p_k . \mathrm{var}(V_j^k) . \sum_{k\in K} p_k . \mathrm{var}(V_\ell^k)} .$$

Pour tout k on sait que :

(90) $$p_k . \mathrm{cov}(V_j^k, V_\ell^k) = p_k . \sigma(v_j^k) . \sigma(V_\ell^k) . \mathrm{cor}(V_j^k, V_\ell^k)$$

où σ désigne l'écart type empirique sur I. Par suite :

$$p_k . \mathrm{cov}(V_j^k, V_\ell^k) \leqq p_k . \sigma(V_j^k) . \sigma(V_\ell^k) \qquad \text{et :}$$

(91) $$\sum_{k\in K} p_k . \mathrm{cov}(V_j^k, V_\ell^k) \leqq \sum_{k\in K} p_k . \sigma(V_j^k) . \sigma(V_\ell^k) .$$

D'autre part il est clair que :

$$\left(\sum_{k\in K} p_k . \sigma(V_j^k) . \sigma(V_\ell^k)\right)^2 \leqq \sum_{k\in K} p_k . \sigma^2(V_j^k) . \sum_{k\in K} p_k . \sigma^2(V_\ell^k)$$

(92) $$\sum_{k\in K} p_k . \sigma(V_j^k) . \sigma(V_\ell^k) \leqq \sqrt{\sum_{k\in K} p_k . \mathrm{var}(V_j^k) . \sum_{k\in K} p_k . \mathrm{var}(V_\ell^k)} .$$

De (89), (91) et (92) on déduit :

$$\sum_{k\in K} p_k.\mathrm{cov}(V_j^k,V_\ell^k) \leqq \sum_{k\in K} p_k.\sigma(V_j^k).\sigma(V_\ell^k) \leqq \sum_{k\in K} p_k.\mathrm{cov}(V_j^k,V_\ell^k)$$

soit :

(93) $$\sum_{k\in K} p_k.\mathrm{cov}(V_j^k,V_\ell^k) = \sum_{k\in K} p_k.\sigma(V_j^k).\sigma(V_\ell^k).$$

De ce dernier résultat et de (90) on tire :

$$\sum_{k\in K} p_k.\sigma(V_j^k).\sigma(V_\ell^k)(\mathrm{cor}(V_j^k,V_\ell^k) - 1) = 0.$$

On supposera ici, qu'aucun écart type des variariables V_j^k et V_ℓ^k, $k = 1,\dots,m$ n'est nul, c'est à dire qu'aucune colonne des matrices X_j et X_ℓ n'est constante. Dans ces conditions l'égalité ci-dessus implique :

(94) $$\forall k \qquad \mathrm{cor}(V_j^k,V_\ell^k) = 1.$$

Les relations (94) signifient que pour tout k les variables aléatoires V_j^k et V_ℓ^k sont liées par une relation fonctionnelle affine que l'on notera :

$$V_\ell^k = a_k.V_j^k + B_k$$

où B_k est une variable vérifiant $B_k(i) = \beta_k$ pour tout $i \in I$. Dans cette situation on sait que :

$$\mathrm{var}(V_\ell^k) = a_k^2.\mathrm{var}(V_j^k)$$
$$\mathrm{cov}(V_j^k,V_\ell^k) = a_k.\mathrm{var}(V_j^k).$$

Par suite (89) s'écrit :

$$\Big(\sum_{k\in K} p_k.a_k.\mathrm{var}(V_j^k)\Big)^2 = \sum_{k\in K} p_k.\mathrm{var}(V_j^k).\sum_{k\in K} p_k.a_k^2.\mathrm{var}(V_j^k)$$

soit :

$$\sum_{k\in K}\sum_{h>k} p_k p_h.\mathrm{var}(V_j^k)\mathrm{var}(V_j^h)(a_k - a_h)^2 = 0.$$

Sous la réserve exprimée ci-dessus on en déduit :

$$\forall\, k, \forall\, h \qquad a_k = a_h = a.$$

En outre (94) implique $a > 0$. On aurait un résultat semblable en supposant $\mathrm{COR}(V_j,V_\ell) = -1$. On note alors b_k le vecteur de F ayant toutes ses coordonnées égales à β_k et on pose $B = (b_1,\dots,b_m)$. La proposition directe est établie.

Réciproquement, j et ℓ étant donnés dans J, on suppose que les variables V_j^k et V_ℓ^k sont liées, pour tout k, par :

$$V_\ell^k = a.V_j^k + B_k \qquad \text{où } B_k(i) = \beta_k.$$

Pour tout k on en déduit :

$$\mathrm{cov}(V_j^k,V_\ell^k) = \sigma(V_j^k).\sigma(V_\ell^k)$$

soit :

$$\sum_{k\in K} p_k.\mathrm{cov}(V_j^k,V_\ell^k) = \sum_{k\in K} p_k.\sigma(V_j^k).\sigma(V_\ell^k)$$

et sachant que $\sigma(V_\ell^k) = a.\sigma(V_j^k)$, on a :

(95) $$\sum_{k\in K} p_k.\mathrm{cov}(V_j^k,V_\ell^k) = a.\sum_{k\in K} p_k \mathrm{var}(V_j^k).$$

Par ailleurs on a également :

(96) $$\sum_{k\in K} p_k.\mathrm{var}(V_j^k).\sum_{k\in K} p_k.\mathrm{var}(V_\ell^k) = a.(\sum_{k\in K} p_k.\mathrm{var}(V_j^k))^2.$$

De (95) et (96) il vient :

$$\mathrm{COR}(V_j,V_\ell) = \frac{a}{|a|}$$

et tout est dit.

Remarque : de ce théorème il résulte que l'expression :"les vecteurs V_j et V_ℓ sont correlés linéairement", signifie que les matrices X_j et X_ℓ de leurs réalisations sont sensiblement liées par une relation affine de type

$$X_\ell = a.X_j + B$$

la matrice B étant formée de vecteurs colonnes constants. Cela signifie également que les variables aléatoires de même rang k, composantes des vecteurs V_j et V_ℓ sont sensiblement correlées linéairement, les pentes des m droites de régression étant en outre sensiblement égales.

La recherche d'une telle corrélation entre vecteurs aléatoires peut présenter un intérêt certain et une interprétation naturelle dans de nombreux problèmes concrets. Par exemple, dans les séries chronologiques, supposons que j soit l'indice des dates, i celui des pays européens et k celui de biens de consommation ou de matières premières. La matrice X_j indiquant le prix de ces biens dans tous les pays de l'ensemble I et à la date j. Une relation $\mathrm{COR}(V_j,V_\ell) \simeq +1$, s'exprimant matriciellement par

$$X_\ell \simeq a.X_j + B$$

permettra de dégager la tendance inflationniste générale exprimée par a (coefficient multiplicatif) et pour chaque bien k une variation conjecturelle (spéculative?)additive exprimée par le terme constant de la k-ème colonne de B. Un autre exemple classique est celui où la matrice X_j indique les notes attribuées par un juge j, à un ensemble I de candidats, suivant un ensemble K de critères. Une relation de type $X_\ell \simeq a.X_j + B$ exprimera de la part des juges j et ℓ des jugements très voisins pour tout i de I et tout k de K, mais révélera des échelles différentes caractérisées par a (étendue des jugements) et par B (hauteur moyenne des jugements pour chaque critère).

Par ailleurs $COR(V_j,V_\ell) = 0$ implique $\sum_{k\in K} p_k.cov(V_j^k,V_\ell^k) = 0$. Cette dernière relation est vérifiée en particulier lorsque les coefficients de corrélation usuels de chacune des paires de composantes correspondantes des vecteurs V_j et V_ℓ est nul. Cependant elle peut être vérifiée, même lorsque ces coefficients valent ± 1, les m droites de régression ayant alors des pentes différentes.

Il reste une situation intéressante, que ne décèle pas le coefficient COR proposé ici : c'est celle où chacun des coefficients de corrélation usuel, de chacune des paires de composantes correspondantes, vaut sensiblement +1 (resp. -1), mais où les pentes des m droites de régression sont différentes. On verra au §.42 comment résoudre ce problème.

41. Normalisation globale des vecteurs aléatoires.

Définition.-

> On dira que le vecteur aléatoire V_j est globalement normalisé si la matrice X_j de sa réalisation sur I est à la fois centrée pour D_{P_I} et Φ-normée.

Propriété 1.-

> Si V_j est un vecteur aléatoire quelconque de matrice X_j sur I alors le vecteur ayant sur I la matrice :
>
> (97) $$\frac{X'_j}{\|X'_j\|_\Phi}$$
>
> est globalement normalisé.

Propriété 2.-

(98) Si V_j et V_ℓ sont globalement normalisés alors :

$$COR(V_j,V_\ell) = \Phi(X_j,X_\ell).$$

En effet, $X'_j = X_j$ et $X'_\ell = X_\ell$ et de plus :

$\|X_j\|_\Phi = \|X'_j\|_\Phi = \|X_\ell\|_\Phi = \|X'_\ell\|_\Phi = 1$. le résultat découle alors de (87).

42. Normalisation forte des vecteurs aléatoires.

Définition.-

On dira que le vecteur aléatoire est fortement normalisé si chaque variable aléatoire V^k_j, $k \in K$ composante de V_j est normalisée.

E désignant l'espérance mathématique d'une variable aléatoire, la définition est équivalente à :

$$k \in K, \forall k \qquad E(V^k_j) = 0 \text{ et } var(V^k_j) = 1.$$

Propriété 1.-

Si $N = D_{P_I}$ et $M = D_{P_K}$ et si V_j est fortement normalisée, alors il est globalement normalisé.

En effet :

$$\|X_j\|^2_\Phi = \Phi(X_j,X_j) = \mathrm{Trace}(D_{P_K}.{}^tX_j.D_{P_I}.X_j).$$

Le terme général de cette matrice produit est :

$$a_{kk'} = p_k.\mathrm{cov}(V^k_j,V^{k'}_j).$$

Les variables V^k_j étant normalisées, on a :

(99) $$a_{kk'} = p_k.\mathrm{cor}(V^k_j,V^{k'}_j)$$

et par suite :

$$\|X_j\|^2_\Phi = \sum_{k\in K} a_{kk} = \sum_{k\in K} p_k.\mathrm{cor}(V^k_j,V^k_j) = \sum_{k\in K} p_k = 1.$$

D'après le paragraphe précédent on en déduit immédiatement :

Propriété 2.-

(100) Si V_j et V_ℓ sont fortement normalisés alors :

$$COR(V_j,V_\ell) = \Phi(X_j,X_\ell).$$

En étudiant de la même façon la matrice $D_{P_K}.{}^tX_\ell.D_{P_I}.X_j$ on obtient :

Théorème 11.-

Si $N = D_{P_I}$ et $M = D_{P_K}$ et si les vecteurs aléatoires V_j et V_ℓ sont fortement normalisés on a :

(101) $$COR(V_j,V_\ell) = \sum_{k\in K} p_k.cor(V_j^k,V_\ell^k).$$

On note V_j'' le vecteur V_j fortement normalisé; on a alors :

Théorème 12.-

Si $N = D_{P_I}$ et $M = D_{P_K}$, alors $COR(V_j'',V_\ell'') = +1$ (resp. -1) si et seulement si, pour tout k de K, les variables aléatoires V_j^k et V_ℓ^k sont liées par une relation fonctionnelle affine du type

$$V_\ell^k = a_k.V_j^k + B_k$$

où B_k est une variable vérifiant $B_k(i) = \beta_k$ et où $a_k > 0$ (resp. $a_k < 0$).

Supposons $COR(V_j'',V_\ell'') = +1$; du théorème 10 il vient :

$$X_\ell'' = a.X_j'' + B \qquad \text{où } a > 0.$$

Les matrices X_ℓ'' et X_j'' étant centrées pour D_{P_I}, B l'est aussi et chacun de ses vecteurs colonnes constant est le vecteur nul. On a donc $B = 0$ et $X_\ell'' = a.X_j''$, ce qui implique,en prenant les Φ-normes,$a = 1$. Enfin l'égalité $X_\ell'' = X_j''$ implique

$$X_\ell = X_j.\begin{pmatrix} a_1 & & 0 \\ & \ddots & \\ 0 & & a_m \end{pmatrix} + B$$

où B est une matrice à vecteurs colonnes constants.

Inversement les relations fonctionnelles indiquées dans le th. 12 impliquent $X_\ell'' = X_j''$. Par suite, d'après la propriété 2 :

$$COR(V_j'',V_\ell'') = \Phi(X_j'',X_\ell'') = \|X_j''\|_\Phi^2 = 1.$$

L'analyse des corrélations COR portant sur des vecteurs aléatoires fortement normalisés permettra donc de déceler les couples de vecteurs tels que, pour tout k, les variables V_j^k et V_ℓ^k correspondantes soient linéairement corrélées. Il reste la restriction que les pentes des droites de régression correspondantes doivent toutes être de même signe. Sachant $COR(V_j'',V_\ell'') = \pm 1$, on obtiendra la valeur a_k de ces pentes, en calculant directement pour tout k la régression de la variable V_ℓ^k à la variable V_j^k.

V. Analyse en terme de corrélation.

43. Définition de l'analyse en C.P en terme de corrélation.

On appellera analyse en composantes principales en terme de corrélation de la famille $(V_j)_{j\in J}$ (ou $(X_j)_{j\in J}$), ou plus simplement analyse des données normalisées, l'analyse générale correspondante, lorsque $N = D_{P_I}$, $M = D_{P_K}$ et lorsque les vecteurs $(V_j)_{j\in J}$ sont globalement ou fortement normalisés.

La conduite algébrique de l'analyse normalisée est indépendante de l'un ou l'autre mode de normalisation proposé, mais bien entendu les conclusions pratiques ne seront pas les mêmes en général.

Dans tout ce sous-chapitre V on supposera une fois pour toute que les vecteurs V_j sont normalisés (sans autre précision).

44. Représentation des vecteurs aléatoires.

D'après les propriétés 2 des §.41 et 42, on a :

Propriété 1.-

> La matrice U associée à la forme u définie sur E^* est la matrice des corrélations des vecteurs aléatoires.

Du théorème de représentation il vient alors :

Propriété 2.-

> $N_J = (\rho_j)_{j\in J}$ étant la représentation des vecteurs aléatoires, pour tout j et pour tout ℓ de J on a :
>
> (102) $$i(\rho_j,\rho_\ell) = COR(V_j,V_\ell)$$
>
> (103) $$\|\rho_j\|_i^2 = 1.$$

Comme en analyse classique on en déduit :

Propriété 3.-

> Dans l'espace E' de représentation des vecteurs, le nuage N_J est sur la sphère unité de centre l'origine.

Et enfin :

<u>Propriété</u> 4.-

(104) Pour tous j et ℓ de J on a :

$$\cos(\rho_j, \rho_\ell) = \mathrm{COR}(V_j, V_\ell).$$

Dans l'espace de représentation des vecteurs, deux vecteurs ayant un coefficient COR égal à 1 (resp. -1) sont donc représentés par deux points confondus (resp. opposés) sur la sphère unité. Si ce coefficient est nul, les points représentatifs sont à angle droit sur la sphère.

45. Interprétation statistique.

Le théorème 5 du §.32 s'énonce dans ce cas :

<u>Théorème</u> 13.-

Les facteurs $\hat{A}_\lambda = A(u_\lambda)$, $\lambda = 1,\ldots,p'$ sont de tous les vecteurs aléatoires $A(x) = \sum_{j\in J} \pi_j^*(d\rho(x)).V_j$, combinaisons linéaires des p vecteurs initiaux, ceux qui réalisent successivement le maximum de la variance de leurs réalisations sur l'échantillon I, sous les contraintes de ρ-orthonormalité portant sur les solutions en x. En outre les composantes principales $\hat{A}_\lambda$ sont de corrélation nulle entre elles, et pour tout λ on a : $\mathrm{VAR}(\hat{A}_\lambda) = \alpha_\lambda$.

Les facteurs sont bien entendu centrés pour D_{P_I}. Pour chaque λ et chaque k, la variance de V_λ^k est donnée par :

$$\mathrm{var}(V_\lambda^k) = \sum_{j\in J} \sum_{\ell\in J} \pi_j^*(d\rho(u_\lambda)).\pi_\ell^*(d\rho(u_\lambda)).\mathrm{cor}(V_j^k, V_\ell^k).$$

46. Composantes principales et espaces factoriels de représentation.

Les propriétés 1, 2 et 3 et le théorème 9 de l'analyse des données centrées (§.38) restent bien entendu valables ici. L'égalité (84) qui donne le terme général de la matrice associée à la λ-ème matrice de référence devient :

$$(105) \qquad {}^t\hat{A}_\lambda D_{P_I}.\hat{A}_\lambda.D_{P_K} = (p_k.\mathrm{cor}(V_\lambda^k, V_\lambda^{k'}))_{k\in K, k'\in K}.$$

La matrice A_λ, associée à $\hat{A}_\lambda$, étant par définition Φ-normée, on a pour tout λ et tout j :

(106) $$\Phi(A_\lambda, X_j) = COR(A_\lambda, V_j) = \sqrt{\alpha_\lambda}.\pi_j(u_\lambda)$$

où :

(107) $$A_\lambda = \frac{1}{\sqrt{\alpha_\lambda}}.\hat{A}_\lambda.$$

De (106) on déduit pour tout j de J :

(108) $$X_j = \sum_{\lambda=1}^{p'} COR(A_\lambda, V_j).A_\lambda.$$

Enfin, dans l'espace E' de représentation des vecteurs, la relation (106) se traduit par :

(109) $$\cos(e'_\lambda, \rho_j) = COR(A_\lambda, V_j)$$

soit, pour tout j :

(110) $$\rho_j = \sum_{\lambda=1}^{p'} COR(A_\lambda, V_j).e'_\lambda .$$

VI. Reconstitution des données.

46. Inertie totale de la représentation des vecteurs V_j.

Ce sous-chapitre VI introduit brièvement le problème de la reconstitution des données, et donne l'expression qui, connaissant les composantes principales, permet de les reconstruire exactement ou approximativement. Le problème sera repris au chapitre suivant.

D'après le théorème 3 du paragraphe 30, les axes factoriels du nuage N_J sont aussi les axes d'inertie lorsque $P = D_{P_J}$. Dans ces conditions, on déduit que l'inertie totale du nuage N_J par rapport à l'origine est donne par :

$$I_{N_J}(0) = \mathrm{Trace}(D_\alpha) = \mathrm{Trace}(U.P) = \sum_{\lambda=1}^{p} \alpha_\lambda.$$

Cette quantité représente également l'inertie de $(X_j)_{j\in J}$ par rapport à l'origine de L et pour la métrique Φ.

47. <u>Reconstitution globale des données</u>.

Pour tout j de J la matrice X_j est reconstruite à partir des composantes principales à l'aide de l'expression suivante, issue du corollaire du théorème de décomposition :

(111) $$X_j = \sum_{\lambda=1}^{p'} \sqrt{\alpha_\lambda}.\pi_j(u_\lambda).A_\lambda = \sum_{\lambda=1}^{p'} \pi_j(u_\lambda).\hat{A}_\lambda \ .$$

Si pour tout j, la sommation sur λ est effectuée de 1 à r, avec $r < p'$, il y a alors reconstitution approchée des données. L'inertie du nuage de E' correspondant est $\sum_{\lambda=1}^{r} \alpha_\lambda$, et la part d'inertie reconstruite est :

(112) $$\sum_{\lambda=1}^{r} \alpha_\lambda \Big/ \sum_{\lambda=1}^{p'} \alpha_\lambda \ .$$

On dira que les expressions (111) constituent une <u>reconstitution globale</u> des données, par opposition à la reconstitution analytique donnée plus loin.

Chapitre IV

ANALYSE TRIADIQUE

I. Propriétés triadiques.

48. Introduction et notations.

Jusqu'à présent nous avons fait jouer des rôles dissymétriques aux ensembles I, J et K. Nous abandonnerons cette attitude, dans ce chapitre, pour mettre en valeur les liens existant entre les trois analyses en C.P que nous proposons ci-dessous. Ces liens exprimés en langage algébrique seront traduits en terme de "vecteurs aléatoires" et de "facteurs" chaque fois que cela sera possible. Soient :

(113) $$A = (a_{ijk})_{i\in I, j\in J, k\in K}$$

un "cube" de données réelles, $E = (R^p, p)$, $F = (R^n, n)$, $G = (R^m, m)$ trois espaces euclidiens, respectivement rapportés à leur base canonique $\mathcal{E}$, $\mathcal{F}$, $\mathcal{G}$; les espaces duals seront rapportés aux bases duales $\mathcal{E}^*$, $\mathcal{F}^*$, $\mathcal{G}^*$.

On pose à présent, pour des raisons de symétrie dans les écritures :

$$I = \{1,\dots,p\}, \quad J = \{1,\dots,n\}, \quad K = \{1,\dots,m\}$$

ce qui par rapport aux notations antérieures revient uniquement à permuter les indices i et j.

On définit alors les trois familles de matrices :

(114) $\quad i \in I, \forall i \quad X_i = (x^i_{jk})_{j\in J, k\in K} \quad$ où $\quad x^i_{jk} = a_{ijk}$;

(115) $\quad j \in J, \forall j \quad Y_j = (y^j_{ki})_{k\in K, i\in I} \quad$ où $\quad y^j_{ki} = a_{ijk}$;

(116) $\quad k \in K, \forall k \quad Z_k = (z^k_{ij})_{i\in I, j\in J} \quad$ où $\quad z^k_{ij} = a_{ijk}$.

D'un point de vue algébrique, on peut considérer ces différentes matrices comme associées à des applications linéaires x_i, y_j, z_k définies par :

$$\begin{array}{lll} \forall i & X_i = M(x_i, G^*, F) & x_i \in \mathcal{L}_R(G^*, F) \\ \forall j & Y_j = M(y_j, E^*, G) & y_j \in \mathcal{L}_R(E^*, G) \\ \forall k & Z_k = M(z_k, F^*, E) & z_k \in \mathcal{L}_R(F^*, E). \end{array}$$

D'un point de vue statistique, on définira trois familles de vecteurs aléatoires :

$$\begin{array}{ll} \forall i & V^I_i = (V^{Ik}_i)_{k\in K} \\ \forall j & V^J_j = (V^{Ji}_j)_{i\in I} \\ \forall k & V^K_k = (V^{Kj}_k)_{j\in J} \end{array}$$

telles que X_i (resp. Y_j, Z_k) soit la réalisation de V^I_i sur J (resp. de V^J_j sur K, de V^K_k sur I).

49. Formes sur les espaces vectoriels d'applications linéaires.

On note respectivement :

$\Phi_I = \Phi_{m^{-1}n}$ le produit scalaire défini sur $\mathcal{L}_R(G^*, F)$

$\Phi_J = \Phi_{p^{-1}m}$ le produit scalaire défini sur $\mathcal{L}_R(E^*, G)$

$\Phi_K = \Phi_{n^{-1}p}$ le produit scalaire défini sur $\mathcal{L}_R(F^*, E)$.

On a donc :

$$\begin{aligned} \Phi_I(X_i, X_{i'}) &= \mathrm{Trace}(x^*_{i'} \circ x_i) = \mathrm{Trace}(M.{}^tX_{i'}.N.X_i) \\ \Phi_J(Y_j, Y_{j'}) &= \mathrm{Trace}(y^*_{j'} \circ y_j) = \mathrm{Trace}(P.{}^tY_{j'}.M.Y_j) \\ \Phi_K(Z_k, Z_{k'}) &= \mathrm{Trace}(z^*_{k'} \circ z_k) = \mathrm{Trace}(N.{}^tZ_{k'}.P.Z_k). \end{aligned}$$

Enfin, on note U, V, W, les matrices de dimensions respectives p×p,

$n\times n$, $m\times m$, définies par :

$$U = (\Phi_I(X_i,X_{i'}))_{i\in I,i'\in I}$$

$$V = (\Phi_J(Y_j,Y_{j'}))_{j\in J,j'\in J}$$

$$W = (\Phi_K(Z_k,Z_{k'}))_{k\in K,k'\in K'}$$

et u, v, w, les formes correspondantes définies sur E^*, F^*, G^*.

50. Propriétés des représentations relatives aux ensembles I, J et K.

$p_{ii'}$, $n_{jj'}$, $m_{kk'}$ désignent respectivement les termes généraux des matrices P, N, et M associées aux produits euclidiens p, n et m. En explicitant les produits $U.P$, $V.N$ et $W.M$ on montre alors :

Propriété 1.-

(117)
$$\text{Trace}(U.P) = \text{Trace}(V.N) = \text{Trace}(W.M)$$
$$= \sum_i \sum_{i'} \sum_j \sum_{j'} \sum_k \sum_{k'} p_{ii'} n_{jj'} m_{kk'} a_{ijk} a_{i'j'k'}.$$

Propriété 2.-

$(\alpha_\lambda)_{\lambda\in I}$, $(\beta_\mu)_{\mu\in J}$, $(\gamma_\nu)_{\nu\in K}$ étant respectivement les valeurs propres des endomorphismes $du\circ dp$, $dv\circ dn$ et $dw\circ dm$ on a :

(118)
$$\sum_\lambda \alpha_\lambda = \sum_\mu \beta_\mu = \sum_\nu \gamma_\nu .$$

On note alors respectivement p', n' et m' le nombre de valeurs propres non nulles de $du\circ dp$, $dv\circ dn$ et $dw\circ dm$ et $E' = (R^{p'},i)$, $F' = (R^{n'},j)$ et $G' = (R^{m'},k)$. Enfin, $N_I = (\rho_i)_{i\in I}$, $N_J = (\sigma_j)_{j\in J}$ et $N_K = (\tau_k)_{k\in K}$ désignent les représentations associées aux éléments de I, J et K dans les espaces E', F' et G'. De la propriété 2 on déduit alors :

Théorème 1.-

Les matrices P, N et M étant respectivement égales à D_{P_I}, D_{P_J} et D_{P_K}, les inerties totales, par rapport à l'origine : $I_{N_I}(0)$, $I_{N_J}(0)$ et $I_{N_K}(0)$ des nuages N_I, N_J et N_K sont égales.

Remarque : on perd ici la propriété essentielle exprimant la dualité dans l'analyse en C.P classique, qui assure suivant le théorème 4 du §.18, que les opérateurs hermitiens duals : $du\circ dp$ et $dv\circ dn$ admettent les mêmes valeurs propres. C'est pourquoi, dans notre cas, on ne propose pas de représentations conjointes des trois ensembles associés à I, J et K. On abordera à nouveau ce point, lorsqu'on examinera, au chapitre suivant le cas particulier $m = 1$.

Enfin, pour certaines valeurs particulières des matrices P, N et M, on donne ci-dessous des expressions qui constituent des moyens rapides de calcul des matrices U, V et W. En outre, au niveau de l'interprétation, ces expressions permettent éventuellement de lier les analyses des trois ensembles associés à I, J et K. Elles se démontrent sans difficulté :

<u>Propriété</u> 3.-

i) Si P, N et M sont des matrices unités on a :

$$U = \sum_{j\in J} {}^tY_j.Y_j = \sum_{k\in K} Z_k.{}^tZ_k$$

$$V = \sum_{k\in K} {}^tZ_k.Z_k = \sum_{i\in I} X_i.{}^tX_i$$

$$W = \sum_{i\in I} {}^tX_i.X_i = \sum_{j\in J} Y_j.{}^tY_j \;.$$

ii) Si P et N sont des matrices unités et si $M = D_{P_K}$ alors :

$$U = \sum_{j\in J} {}^tY_j.D_{P_K}.Y_j = \sum_{k\in K} p_k.Z_k.{}^tZ_k \;.$$

iii) Si P et M sont des matrices unités et si $N = D_{P_J}$ alors :

$$U = \sum_{j\in J} p_j.{}^tY_j.Y_j = \sum_{k\in K} Z_k.D_{P_J}.{}^tZ_k \;.$$

Dans les cas ii) et iii) on a des expressions semblables pour V et W.

Comme on l'a dit ces expressions sont importantes au niveau de l'interprétation. Par exemple, si i, j et k se réfèrent respectivement à des variables, aux éléments d'un échantillon et à des époques d'observations, alors chaque variable vectorielle est constituée par les m observations de la variable initiale. Si P et N sont des matrices unités et si $M = D_{P_K}$ on a alors :

$$U = \sum_{k\in K} p_k.Z_k.{}^tZ_k.$$

Chaque matrice Z_k est une matrice variables×échantillon, relative à l'époque k d'observation. L'égalité ci-dessus, dans ces conditions, signifie que la matrice des produits scalaires des variables vectorielles est égale à la moyenne, pour toutes les époques, des matrices des produits scalaires des variables !.

51. Propriétés triadiques des composantes principales.

On note $(u_\lambda)_{\lambda=1,\dots,p'}$, $(v_\mu)_{\mu=1,\dots,n'}$ et $(w_\nu)_{\nu=1,\dots,m'}$ des systèmes de vecteurs propres de $dU \circ dp$, $dV \circ dn$ et $dW \circ dm$, respectivement p, n et m-orthonormés, associés aux valeurs propres non nulles, d'indice correspondant.

Pour tout λ, μ et ν satisfaisant à ces conditions, on pose :

(119) $$\hat{u}_\lambda = \sqrt{\alpha_\lambda}.u_\lambda,\ \mathcal{U} = (u_1,\dots,u_{p'}) \text{ et } \hat{\mathcal{U}} = (\hat{u}_1,\dots,\hat{u}_{p'})$$

(120) $$\hat{v}_\mu = \sqrt{\beta_\mu}.v_\mu,\ \mathcal{V} = (v_1,\dots,v_{n'}) \text{ et } \hat{\mathcal{V}} = (\hat{v}_1,\dots,\hat{v}_{n'})$$

(121) $$\hat{w}_\nu = \sqrt{\gamma_\nu}.w\ ,\ \mathcal{W} = (w_1,\dots,w_{m'}) \text{ et } \hat{\mathcal{W}} = (\hat{w}_1,\dots,\hat{w}_{m'}).$$

f_I, f_J et f_K représentant les applications linéaires respectivement associées aux familles $(x_i)_{i\in I}$, $(y_j)_{j\in J}$ et $(z_k)_{k\in K}$, définies par des expressions semblables à (51) du §.29, le schéma général qui suit, composé des trois schémas semblables à celui du §.31, explicite les espaces concernés et les applications linéaires de l'un à l'autre. Il permet notamment de lire les applications composées utilisées ci-dessous. Pour plus de clarté, ne figurent pas les applications ${}^t x_i$, ${}^t y_j$ et ${}^t z_k$.

Voir schéma sur feuille hors-texte.

Pour tout $\lambda = 1,\dots,p'$ et tout $j \in J$ on pose :

(122) $$\hat{a}_j^\lambda = y_j(dp(u_\lambda)) \in G$$

De même :

(123) $$\hat{b}_k^\mu = z_k(dn(v_\mu)) \in E$$

(124) $$\hat{c}_i^\nu = x_i(dm(w_\nu)) \in F.$$

On obtient alors le résultat suivant :

Théorème 2.-

Les composantes principales $\hat{A}_\lambda, \hat{B}_\mu$ et $\hat{C}_\nu$ vérifient les relations :

$$(125)\quad \hat{A}_\lambda = {}^t(\hat{a}_1^\lambda,\ldots,\hat{a}_n^\lambda)$$

$$(126)\quad \hat{B}_\mu = {}^t(\hat{b}_1^\mu,\ldots,\hat{b}_m^\mu)$$

$$(127)\quad \hat{C}_\nu = {}^t(\hat{c}_1^\nu,\ldots,\hat{c}_p^\nu).$$

Montrons (125) : on sait que $\hat{A}_\lambda = \sum_{i\in I} \pi_i^*(dp(u_\lambda)).X_i$; par suite le terme (j,k) de cette matrice s'écrit :

$$\sum_{i\in I} \pi_i^*(dp(u_\lambda)).x_{jk}^i = \sum_{i\in I} \pi_i^*(dp(u_\lambda)).a_{ijk}$$

et la k-ème coordonnée de $\hat{a}_j^\lambda$ s'écrit d'après (122) et (115) :

$$\sum_{i\in I} \pi_i^*(dp(u_\lambda)).y_{ki}^j = \sum_{i\in I} \pi_i^*(dp(u_\lambda)).a_{ijk}.$$

De la même façon, si on pose :

$$(128)\quad \hat{a}'^\lambda_k = {}^t z_k(dp(u_\lambda)) \in F$$

$$(129)\quad \hat{b}'^\mu_i = {}^t x_i(dn(v_\mu)) \in G$$

$$(130)\quad \hat{c}'^\nu_j = {}^t y_j(dm(w_\nu)) \in E$$

on obtient :

Théorème 3.-

Les composantes principales $\hat{A}_\lambda, \hat{B}_\mu$ et $\hat{C}_\nu$ vérifient les relations :

$$(131)\quad \hat{A}_\lambda = (\hat{a}'^\lambda_1,\ldots,\hat{a}'^\lambda_m)$$

$$(132)\quad \hat{B}_\mu = (\hat{b}'^\mu_1,\ldots,\hat{b}'^\mu_p)$$

$$(133)\quad \hat{C}_\nu = (\hat{c}'^\nu_1,\ldots,\hat{c}'^\nu_n).$$

Enfin, pour chaque ensemble I, J et K on a les relations fondamentales suivantes, provenant du théorème de décomposition et constituant trois façons de reconstituer globalement les données.

$$(134)\quad \forall\, i \qquad X_i = \sum_{\lambda=1}^{p'} \pi_i(u_\lambda).\hat{A}$$

$$(135)\quad \forall\, j \qquad Y_j = \sum_{\mu=1}^{n'} \pi_j(v_\mu).\hat{B}$$

$$(136)\quad \forall\, k \qquad Z_k = \sum_{\nu=1}^{m'} \pi_k(w_\nu).\hat{C}\ .$$

II. Reconstitution analytique des données.

52. Transformation des composantes principales.

En reprenant l'écriture (65) du pargraphe 31, on dira que :

$$\hat{A}_\lambda = \sum_{i\in I} \pi_i^*(dp(u_\lambda)).V_i^I \text{ est le } \lambda\text{-ème facteur I,}$$

$$\hat{B}_\mu = \sum_{j\in J} \pi_j^*(dn(v_\mu)).V_j^J \text{ est le } \mu\text{-ème facteur J,}$$

$$\hat{C}_\nu = \sum_{k\in K} \pi_k^*(dm(w_\nu)).V_k^K \text{ est le } \nu\text{-ème facteur K,}$$

où π_i^*, π_j^* et π_k^* désignent respectivement la i-ème, j-ème et k-ème fonction coordonnée de E^*, F^* et G^*; les réalisations de $\hat{A}_\lambda$, $\hat{B}_\mu$ et $\hat{C}_\nu$ respectivement sur I, J et K étant données par les matrices $\hat{A}_\lambda$, $\hat{B}_\mu$ et $\hat{C}_\nu$.

On propose l'écriture symbolique suivante :

$$\lambda = \sum_{i\in I} \pi_i^*(dp(u_\lambda)).i$$

$$\mu = \sum_{j\in J} \pi_j^*(dn(v_\mu)).j$$

$$\nu = \sum_{k\in K} \pi_k^*(dm(w_\nu)).k$$

où la première égalité (par exemple) signifie pour tout vecteur ou variable aléatoire V :

$$V(\lambda) = \sum_{i\in I} \pi_i^*(dp(u_\lambda)).V(i).$$

On dira alors que λ (resp. μ et ν) est l'<u>élément associé au</u> λ-<u>ème facteur</u> I (resp. μ-<u>ème facteur</u> J et ν-<u>ème facteur</u> K). Cela revient à construire trois nouveaux ensembles : $\mathcal{I} = \{\lambda\}_{\lambda=1,\dots,p'}$, $\mathcal{J} = \{\mu\}_{\mu=1,\dots,n'}$ et $\mathcal{K} = \{\nu\}_{\nu=1,\dots,m'}$ à partir des trois ensembles initiaux I, J et K.

Dans ces conditions il apparait que le réel :

$${}^t(dn(v_\mu)).\hat{A}_\lambda.(dm(w_\nu))$$

est la réalisation du λ-ème facteur I, pour les éléments respectivement associés au μ-ème facteur J et au ν-ème facteur K. Il vient alors :

<u>Proposition</u>.-

> La matrice $n'\times m'$:
>
> $${}^t(N.\mathcal{V}).\hat{A}_\lambda.(M.\mathcal{W}) = {}^t\mathcal{V}.N.\hat{A}_\lambda.M.\mathcal{W}$$
>
> exprime les réalisations du λ-ème facteur I sur les ensembles $\mathcal{J}$ et $\mathcal{K}$.

Algébriquement, d'après le schéma hors-texte, cette matrice apparait comme associée à une application linéaire de G'^* dans F', tandis que $\hat{A}_\lambda$ correspond à une application linéaire de G^* dans F. Dans le cas où $m' = m$ et $n' = n$, on interprète encore cette matrice comme matrice de l'application linéaire associée à $\hat{A}_\lambda$ par rapport aux nouvelles bases $(dm(w_\nu))_{\nu\in K}$ de G^* et $(v_\mu)_{\mu\in J}$ de F.

En conséquence on adoptera les notations suivantes :

(138) $$\forall \lambda \qquad \hat{H}^I_\lambda = {}^t\mathcal{V}.N.\hat{A}_\lambda.M.\mathcal{W}$$

(139) $$\forall \mu \qquad \hat{H}^J_\mu = {}^t\mathcal{W}.M.\hat{B}_\mu.P.\mathcal{U}$$

(140) $$\forall \nu \qquad \hat{H}^K_\nu = {}^t\mathcal{U}.P.\hat{C}_\nu.N.\mathcal{V}$$

et les matrices $\hat{H}^J_\mu$ et $\hat{H}^K_\nu$ s'interprètent de façon semblable à $\hat{H}^I_\lambda$.

53. <u>Transformation du cube initial de données.</u>

Soit $h^{I\lambda}_{\mu\nu}$ (resp. $h^{J\mu}_{\nu\lambda}$, $h^{K\nu}_{\lambda\mu}$) le terme général de $\hat{H}^I_\lambda$ (resp. $\hat{H}^J_\mu$, $\hat{H}^K_\nu$). On a :

<u>Théorème</u> 4.-

(141) Pour tous λ, μ et ν on a :
$$h^{I\lambda}_{\mu\nu} = h^{J\mu}_{\nu\lambda} = h^{K\nu}_{\lambda\mu}.$$

Montrons la première égalité :
$$h^{I\lambda}_{\mu\nu} = {}^t(dn(v_\mu)).\hat{A}_\lambda.(dm(w_\nu))$$
soit, d'après (128) et (131) :
$$h^{I\lambda}_{\mu\nu} = {}^t(dn(v_\mu)).({}^tZ_1.dp(u_\lambda),\ldots,{}^tZ_m.dp(u_\lambda)).dm(w_\nu)$$
$$= ({}^t(dn(v_\mu))\,{}^tZ_1.dp(u_\lambda),\ldots,{}^t(dn(v_\mu))\,{}^tZ_m.dp(u_\lambda)).dm(w_\nu)$$
or, ${}^t(dp(u_\lambda))Z_k.dn(v_\mu) = {}^t(dn(v_\mu)).{}^tZ_k.dp(u_\lambda)$; par suite :
$$h^{I\lambda}_{\mu\nu} = ({}^t(dp(u_\lambda)).Z_1.dn(v_\mu),\ldots,{}^t(dp(u_\lambda)).Z_m.dn(v_\mu)).dm(w_\nu)$$
$$= {}^t(dp(u_\lambda)).(Z_1.dn(v_\mu),\ldots,Z_m.dn(v_\mu)).dm(w_\nu).$$

La matrice centrale, d'après (122) et (126) n'est autre que ${}^t\hat{B}_\mu$, donc :
$$h^{I\lambda}_{\mu\nu} = {}^t(dp(u_\lambda)).{}^t\hat{B}_\mu.dm(w_\nu)$$
$$= {}^t(dm(w_\nu)).\hat{B}_\mu.dp(u_\lambda)$$
$$= h^{J\mu}_{\nu\lambda}.$$

Ce théorème apporte une confirmation, s'il en était besoin, de l'interprétation des matrices $\hat{A}$, à savoir que l'un quelconque des termes de (141) représente la réalisation de l'un quelconque des trois facteurs λ, μ et ν sur les éléments associés aux deux autres.

On notera alors :

(142) $$h_{\lambda\mu\nu} = h^{I\lambda}_{\mu\nu} = h^{J\mu}_{\nu\lambda} = h^{K\nu}_{\lambda\mu}$$

et :

$$H = (h_{\lambda\mu\nu})_{\lambda=1,\dots,p',\mu=1,\dots,n',\nu=1,\dots,m'}.$$

Proposition.-

> Le cube H est l'homologue du cube $A = (a_{ijk})$ lorsque l'on substitue aux ensembles I, J et K initiaux les ensembles I, J et K associés respectivement aux facteurs relatifs à I, J et K.

Le cube H constitue en quelque sorte la réecriture du cube A relativement à de "nouvelles dimensions idéalisées" au sens de la littérature américaine (cf. |5|).

54. Reconstitution analytique des données.

L'expression (111) du §.47 permet, on l'a vu, de reconstruire les données, connaissant les composantes principales et les vecteurs propres p-orthonormés (u_λ) de $du_\circ dp$. Pour simplifier, on pose $\pi_i(u_\lambda) = u^i_\lambda$, $\pi_j(v_\mu) = v^j_\mu$ et $\pi_k(w_\nu) = w^k_\nu$, et le théorème suivant donne une façon analytique de reconstruire les données connaissant le cube H et les trois familles de vecteurs propres :

Théorème 5.-

> Pour tous i, j et k on a :
>
> (143) $$a_{ijk} = \sum_{\lambda=1}^{p'} \sum_{\mu=1}^{n'} \sum_{\nu=1}^{m'} u^i_\lambda . v^j_\mu . w^k_\nu . h_{\lambda\mu\nu}.$$

En effet, on a d'après (111) :

$$X_i = \sum_{\lambda=1}^{p'} u^i_\lambda . \hat{A}_\lambda$$

d'où, en notant $\hat{a}^\lambda_{jk}$ le terme général de $\hat{A}_\lambda$, il vient

$$x^i_{jk} = a_{ijk} = \sum_{\lambda=1}^{p'} u^i_\lambda . \hat{a}^\lambda_{jk} .$$

D'après (43), on peut écrire :

$$\mathcal{V}.{}^t\mathcal{V}.P = I \qquad \text{et} \qquad \mathcal{W}.{}^t\mathcal{W}.M = I$$

ou ce qui est équivalent :

$$M.\mathcal{W}.{}^t\mathcal{W} = I .$$

D'après (138) :

$$\mathcal{V}.\hat{H}^I_\lambda.{}^t\mathcal{W} = \mathcal{V}.{}^t\mathcal{V}.P.\hat{A}_\lambda.M.\mathcal{W}.{}^t\mathcal{W} = \hat{A}_\lambda$$

$$\hat{a}^\lambda_{jk} = \sum_{\mu=1}^{n'} \sum_{\nu=1}^{m'} v^j_\mu . h^{I\lambda}_{\mu\nu} . w^k_\nu = \sum_{\mu=1}^{n'} \sum_{\nu=1}^{m'} v^j_\mu . w^k_\nu . h_{\lambda\mu\nu}$$

d'où le résultat.

On aura l'occasion de revenir sur cette expression au chapitre 6.

Chapitre V

CAS PARTICULIER OU LES VECTEURS ALEATOIRES SONT DE DIMENSION $m = 1$

I. Analyse simple.

55. Notations.

Comme il est d'usage en analyse des données on considère une famille $(V_j)_{j\in J}$ de variables aléatoires interprétée ici comme une famille de vecteurs aléatoires de dimension 1. On pose $J = \{1,\dots,p\}$ et soit $I = \{1,\dots,n\}$ un échantillon. Pour tout j, Y_j est la matrice $n\times 1$ des réalisations de V_j sur I; on la considère comme la matrice des composantes d'un vecteur y_j de $F = (R^n,\eta)$. Comme $m = 1$, l'espace euclidien $G = (R^m,m)$ est réduit à la droite des réels, pour laquelle il est naturel de choisir la métrique unité.

56. Forme Φ_J.

D'après la définition générale de Φ on a :

$$\Phi_J(Y_j,Y_\ell) = \text{Trace}({}^tY_\ell.N.Y_j)$$
$$= \eta(y_j,y_\ell).$$

Propriété.-

> Si $m = 1$, le produit Φ défini sur $L_R(G^*,F)$ est confondu avec le produit η défini sur F.

Remarque : soit X_i la matrice ligne de dimension $1 \times p$ et de terme général $V_j(i)$. Elle représente les valeurs prises pour i par toutes les variables V_j. On considère X_i comme exprimant les coordonnées d'un vecteur x_i de $E = (R^p, \rho)$. On a alors :

$$\Phi_I(X_i, X_\ell) = \text{Trace}(P.{}^tX_\ell.X_i).$$

Le terme général $a_{jj'}$ de cette matrice est donné par :

$$a_{jj'} = \sum_{j''=1}^{p} p_{jj''}.x_\ell^{j''}.x_i^{j'}$$

et :

$$\text{Trace}(P.{}^tX_\ell.X_i) = \sum_{j\in J} a_{jj} = \sum_{j\in J} \sum_{j''\in J} x_i^j.p_{jj''}.x_\ell^{j''}$$

$$= \rho(x_i, x_\ell).$$

La propriété ci-dessus est encore valable pour Φ_I.

57. Analyse en C.P de la famille $(V_j)_{j\in J}$, considérée comme famille devecteurs.

On applique les résultats établis au chapitre III. Toute la conduite de l'analyse revient à déterminer les vecteurs propres ρ-orthonormés (u_λ) de la matrice U.P, où U est la matrice de terme général $\Phi_J(Y_j, Y_{j'})$ D'après la propriété ci-dessus, cette matrice est précisément la matrice des produits scàlaires des vecteurs y_j de F. Cela signifie que si Z est la matrice $n \times p$ exprimant les réalisations des variables V_j sur I; alors :

$$U.P = {}^tZ.N.Z.P$$

On reconnait là l'expression usuelle utilisée en analyse en C.P classique;

La λ-ème composante principale $\hat{A}_\lambda$ est donnée par :

$$\hat{A}_\lambda = \sum_{j\in J} \pi_j^*(d\rho(u_\lambda)).Y_j$$

et cette dernière quantité s'écrit également :

$$\hat{A}_\lambda = Z.P.u_\lambda \ .$$

On retrouve là encore les composantes principales au sens classique dans cette expression;

Enfin, en ce qui concerne les représentations, on a vu que l'image ρ_j de V_j a pour coordonnées $(\sqrt{\alpha_\lambda}.\pi_j(u_\lambda))$. Cela signifie que le vecteur $\sqrt{\alpha_\lambda}.u_\lambda$ donne sur l'axe factoriel λ, les coordonnées de tous les ρ_j'.

En analyse classique ces valeurs sont données par la λ-ème composante principale de ${}^tZ$, définie par :

$$ {}^tZ.N.v\lambda \ , $$

où $v_\lambda = \frac{1}{\sqrt{\alpha_\lambda}}.Z.P.u_\lambda$. On peut donc écrire :

$$ {}^tZ.N.v_\lambda = \frac{1}{\sqrt{\alpha_\lambda}}.{}^tZ.N.Z.P.u_\lambda = \frac{1}{\sqrt{\alpha_\lambda}}.U.P.u_\lambda = \sqrt{\alpha_\lambda}.u_\lambda \ . $$

On constate alors que, dans les deux cas, les coordonnées des représentations des variables V_j, sur le λ-ème axe factoriel coïncident. On a donc :

Propriété.-

> L'analyse en C.P classique d'une famille de variables aléatoires sur un échantillon I, et l'analyse en C.P de cette famille, considérée comme famille de vecteurs aléatoires de dimension 1, sont identiques.

58. Covariance, variance et corrélation.

On suppose $N = D_{P_I}$. Il est clair que le centrage pour D_{P_I} de la matrice Y_j proposé au paragraphe 34 coïncide avec le centrage de la famille Y_j, lorsque $m = 1$.

Par ailleurs, d'après (71) ou (73) il vient immédiatement :

(144) $$ COV(V_j,V_\ell) = cov(V_j^1,V_\ell^1) $$

et :

(145) $$ VAR(V_j) = var(V_j^1) $$

où V_j^1 désigne l'unique composante de V_j.

Enfin, de (86) et des deux relations ci-dessus on tire :

(146) $$ COR(V_j,V_\ell) = cor(V_j^1,V_\ell^1). $$

Propriété.-

> Si les vecteurs sont de dimension 1, alors les coefficients COV, VAR et COR sont identiques aux coefficients cov, var et cor.

II. Analyse triple.

59. Analyse de la matrice Z considérée comme un cube.

La matrice Z de dimension $n \times p$ exprime les réalisations de toutes les variables V_j sur l'échantillon I. Puisque K = {1}, on pose :

$$Z = (a_{ij})_{\substack{i\in I\\ j\in J}} = (a_{ij1})_{\substack{i\in I\\ j\in J}}$$

et par suite :

$$\forall j \qquad Y_j = (a_{ij1})_{i\in I} \qquad \text{(matrice colonne)}$$

$$\forall i \qquad X_i = (a_{ij1})_{j\in J} \qquad \text{(matrice ligne).}$$

Y_j représente donc la j-ème colonne de Z et X_i la i-ème ligne.

On a vu que l'analyse vectorielle de la famille $(Y_j)_{j\in J}$ coïncide avec l'analyse en C.P classique des colonnes de Z, et que l'analyse vectorielle de la famille $(X_i)_{i\in I}$ coïncide avec celle des lignes de Z. Il reste à **analyser** (!) la famille (Z) réduite à un seul élément. La matrice W de terme général $\Phi_K(Z_k,Z_{k'})$ se réduit à un scalaire; il existe donc un seul facteur K dont la réalisation sur I est précisément la matrice Z elle-même. Pour la même raison on n'aura qu'un seul espace factoriel de représentations conjointes, issu de l'analyse en C.P classique de $\hat{C}_1 = Z$.

60. Problème des représentations conjointes.

L'analyse triple de la matrice Z considérée comme un cube se réduit donc à l'analyse en C.P classique de la matrice Z. La raison profonde de ce fait est que les opérateurs hermitiens $dU \circ dP$ et $dV \circ dN$ **associés aux** familles $(y_j)_{j\in J}$ et $(x_i)_{i\in I}$ admettent le même système de valeurs propres (α_λ) et que leurs vecteurs propres orthonormés sont liés par les relations classiques de dualité :

$$v_\lambda = \frac{1}{\sqrt{\alpha_\lambda}}.Z.P.\ u_\lambda$$

$$u_\lambda = \frac{1}{\sqrt{\alpha_\lambda}}.{}^tZ.N.v_\lambda \ .$$

Comme on l'a vu cette propriété est caractéristique du cas m = 1, et c'est pourquoi nous avons proposé la notion d'espaces factoriels de représentations conjointes, dans le cas général, au lieu d'une représentation conjointe des trois ensembles d'indices I, J et K.

Remarquons enfin, que les applications linéaires f_I et f_J du schéma général (hors-texte) sont ici données par :

$$f_I(x) = \sum_{i \in I} \pi_i^*(dn(x)).X_i = {}^t({}^tZ.N.x) \qquad \text{où } x \in R^n$$

$$f_J(x) = \sum_{j \in J} \pi_j^*(dp(x)).Y_j = Z.P.x \qquad \text{où } x \in R^p.$$

Quant à l'application f_K elle se réduit à l'identité.

Chapitre VI

COMPARAISONS AVEC D'AUTRES PROCEDURES

I. Introduction.

61. Présentation.

Nous abordons dans ce sixième chapitre la comparaison des éléments théoriques que nous avons développés ci-avant, avec des procédures dont la problématique est proche de la notre. Pour simplifier disons qu'il s'agit, à partir d'un tableau donné à trois indices, prenant leurs valeurs dans trois ensembles I, J et K, d'obtenir une description la plus simple possible des éléments de l'un ou plusieurs de ces ensembles; cette description permettant en outre une reconstitution approximative des données, possédant des propriétés sympathiques, et offrant dans certains cas la possibilité d'une interprétation statistique.

A partir des travaux des précurseurs de l'analyse en composantes principales, notamment d'ECKART et YOUNG ([7]) et de HOTELLING ([11]), il existe actuellement, de façon schématique, deux courants de pensée en la matière, dont les finalités sont distinctes. L'un, représenté notamment par TUCKER et CARROLL, issu de la recherche en psychologie, développe des modèles appliqués au domaine du positionnement multidimensionnel (multidimensional scaling), principalement dans le cadre de l'explicitation des différences individuelles de jugement ou de perception. L'autre, plus soucieux du développement de la description statistique proprement dite, est issu des travaux d'ESCOUFIER sur le traitement des variables vectorielles.

On examinera d'abord le modèle de l'analyse factorielle de mode-trois de TUCKER ([17] et [18]), faisant suite aux travaux de LEVIN ([13]), puis le modèle de décomposition canonique de CARROLL ([5] et [6]). Les programmes d'analyse des différences individuelles INDSCAL et IDIOSCAL de CARROLL, qui découlent directement de son modèle de décomposition canonique, sont probablement les plus connus et les plus utilisés à l'heure actuelle, dans le champ concerné. Par ailleurs ce modèle peut être regardé comme un cas particulier de celui de TUCKER. Ce sont là les raisons qui nous conduisent à les analyser, en regard de notre étude.

Enfin, dans la même optique, on examinera à la lumière des chapitres antérieurs, comment interpréter le produit défini sur les opérateurs d'ESCOUFIER ([8] et [9]), qui donne lieu à divers compléments et prolongements, notamment dans l'introduction à l'analyse des données de CAILLEZ et PAGES ([4]), de la part de BRAUN ([3]), PAGES ([16]) et L'HERMIER DES PLANTES ([14]).

Nous n'aborderons pas ici le problème du positionnement multidimensionnel proprement dit.

62. Signification des modèles de TUCKER et de CARROLL.

Soit $(x_{ijk})_{i\in I, j\in J, k\in K}$ un tableau à trois indices. TUCKER et CARROLL recherchent tous deux, pour tous i, j et k une décomposition de x_{ijk}. Le premier postule une décomposition du type :

$$x_{ijk} = \sum_m \sum_p \sum_q a_{im} . b_{jp} . c_{kq} . g_{mpq} \tag{147}$$

où m, p, q représentent trois "modes dérivés" (ou dimensions idéalisées), a_{im}, b_{jp}, c_{kq} l'expression des modes observations en termes de modes dérivés, et g_{mpq} les inter-relations entre ces modes dérivés.

CARROLL quant à lui recherche une décomposition du type :

(148) $$x_{ijk} = \sum_t a_{it}.b_{jt}.c_{kt} .$$

La relation (148) se déduit évidemment de (147) avec :

$$\begin{cases} g_{mpq} = 1 & \text{si } m = p = q = t \\ g_{mpq} = 0 & \text{sinon.} \end{cases}$$

Une différence essentielle, entre ces deux modèles, réside en la résolution de ceux-ci. La solution du modèle de TUCKER s'obtient rigoureusement par des voies algébriques, à partir de décompositions factorielles classiques ([7]), tandis que celle du modèle de CARROLL s'obtient par la procédure itérative NILES de WOLD ([19]) dont la convergence n'est pas prouvée.

Pour rechercher l'origine et la signification de ces modèles plaçons nous dans le cas plus connu d'un tableau (x_{ij}) à deux indices. On notera X la matrice n×p de terme général x_{ij}, $E = (R^p, P)$ et $F = (R^n, N)$.

Si (u_λ) et (v_λ) sont respectivement les vecteurs propres P et N-orthonormés des opérateurs hermitiens $U.P$ et $V.N$ associés à X et à ${}^tX$, on pose comme précédemment (si p < n) :

$$\mathcal{U} = (u_1, \ldots, u_p) \qquad \text{et} \qquad \mathcal{V} = (v_1, \ldots, v_p)$$

et :

(149) $$G = {}^t\mathcal{V}.N.X.P.\mathcal{U} .$$

On a vu précédemment que :

$$P.\mathcal{U}.{}^t\mathcal{U} = I \qquad \text{et} \qquad \mathcal{V}.{}^t(N.\mathcal{V}) = I$$

donc de (149) il vient :

(150) $$X = \mathcal{V}.G.{}^t\mathcal{U}$$

soit, pour tous i et j :

(151) $$x_{ij} = \sum_\lambda \sum_{\lambda'} v^i_\lambda.u^j_{\lambda'}.g_{\lambda\lambda'}$$

où $g_{\lambda\lambda'}$ est le terme général de G.

On constate donc que formellement le modèle (147) de TUCKER généralise (151). On verra au sous-chapitre suivant que les propriétés des termes figurant sous les signes de sommation de (151) sont conservés dans (147).

Si ζ_λ est la composante principale (vecteur de F) associée à u_λ, on connait l'expression classique :

$$X = \sum_\lambda \zeta_\lambda . {}^t u_\lambda$$

qui pour tous i et j s'écrit :

(152) $$x_{ij} = \sum_\lambda \zeta_\lambda^i . u_\lambda^j \quad .$$

Formellement ce modèle (152) est identique au modèle de CARROLL. Cependant il semble plutôt que ce soit celui de l'analyse en facteurs communs et spécifiques qui inspire ce dernier. Ce modèle s'écrit :

(153) $$x_{ij} = \sum_\alpha a_{i\alpha} . b_{j\alpha} + e_{ij}$$

où $a_{i\alpha}$ est "l'influence" du α-ème facteur commun sur le sujet i et $b_{j\alpha}$ son influence sur la variable j. Le terme e_{ij} exprime un "résidus" spécifique au sujet i et à la variable j. Si en première approximation on suppose $e_{ij} = 0$, pour tous i et j, on obtient :

(154) $$x_{ij} = \sum_\alpha a_{i\alpha} . b_{j\alpha}$$

dont peut dériver le modèle (148) de CARROLL.

Cette approche en terme de modèles, posés a priori, est particulièrement floue, puisqu'elle ne fixe ni les présupposés théoriques du modèle ni les propriétés vérifiées par ses termes.

flexible

II. Le modèle de TUCKER.

63. Résolution générale du modèle.

On résume ici brièvement les étapes algébriques de la solution proposée par TUCKER. Les notations sont celles utilisées dans [17]. Plus précisément :

i) $_iA_m$ désigne une matrice ayant i (resp. m) pour indice ligne (resp. colonne);

ii) $X = (x_{ijk})$, $i \in I$, $j \in J$, $k \in K$;

iii) $_iX_{(jk)}$ désigne la matrice de dimension Card(I)×(Card(J).Card(K)), dont la i-ème ligne est formée des termes x_{ijk}, où $(j,k) \in J\times K$, ordonné par l'ordre lexicographique;

iv) iAmBj désigne le produit des matrices A et B;

v) (ij)H(mp) = aAm×jBq désigne la matrice, produit de Kronecker des matrices A et B donnée par :

$$(ij)H(mp) = \begin{pmatrix} (a_{1i1m}.jBq) & (a_{1i2m}.jBq) & \cdots \\ (a_{2i1m}.jBq) & \cdots & \\ \cdots & & \\ \cdots & & \cdots \end{pmatrix} .$$

Dans ces conditions le modèle (147) s'écrit aussi, et indifféremment, sous l'une des trois formes suivantes :

(155) iX(jk) = iAmG(pq)(pBj×qCk)

(156) jX(ik) = jBpG(mq)(mAi×qCk)

(157) kX(ij) = kCqG(mp)(mAi×pBj)

où $G = (g_{mpq})$, $A = (a_{im})$, $B = (b_{jp})$, $C = (c_{kq})$. On pose alors :

(158) iMi = iX(jk)Xi

(159) jPj = jX(ik)Xj

(160) kQk = kX(ij)Xk.

En remplaçant dans ces trois dernières égalités les différentes matrices X par leurs valeurs données en (155), (156) et (157), on obtient finalement:

(161) iMi = iAmMmAi

(162) jPj = jBpPpBj

(163) kQk = kCqQqCk

où l'on connait les expressions des matrices M, P et Q en fonction des matrices A, B et C et du cube G. La conduite de la résolution est alors la suivante :

i) calcul des matrices M, P et Q par (158), (159) et (160);

ii) factorisation de M, P et Q sous la forme donnée par (161), (162) et (163), par une méthode quelconque, pourvu que le rang des matrices A, B et C obtenues soit respectivement égal au nombre de leurs colonnes.

iii) le cube G est alors donné indifféremment par l'un ou l'autre des trois expressions suivantes :

(164) $$mG(pq) = m\overset{+}{A}iX(jk)(j\overset{+}{B}p \times k\overset{+}{C}q)$$

(165) $$pG(mq) = p\overset{+}{B}jX(ik)(i\overset{+}{A}m \times k\overset{+}{C}q)$$

(166) $$qG(mp) = q\overset{+}{C}kX(ij)(i\overset{+}{A}m \times j\overset{+}{B}p)$$

où

(167) $$m\overset{+}{A}i = (mAiAm)^{-1}mAi$$

(168) $$p\overset{+}{B}j = (pBjBp)^{-1}pBj$$

(169) $$q\overset{+}{C}k = (qCkCq)^{-1}qCk \ .$$

64. Résolution dans un cas particulier.

TUCKER explicite la procédure de factorisation des matrices iMi, jPj et kQk, lorsque l'on impose aux matrices A, B et C d'être des sections verticales de matrices orthogonales. On a alors les relations (161), (162) et (163) où les matrices mMm, pPp et qQq sont les matrices diagonales des valeurs propres non nulles de iMi, jPj et kQk, et où les matrices A, B et C expriment en colonnes les vecteurs propres orthonormés correspondants. On a donc dans ce cas :

$$mAiAm = I$$

$$pBjBp = I$$

$$qCkCq = I$$

et les relations (164), (165) et (166) deviennent :

(170) $$mG(pq) = mAiX(jk)(jBp \times kCq)$$

(171) $$pG(mq) = pBjX(ik)(iAm \times kCq)$$

(172) $$qG(mp) = qCkX(ij)(iAm \times jBp).$$

65. Comparaison.

On compare à présent, la solution du modèle de TUCKER présentée au paragraphe précédent à l'analyse en composantes principales d'une famille de variables vectorielles. Comme au chapitre IV, le cube $A = (a_{ijk})$ peut s'écrire sous l'une ou l'autre des trois formes :

$$X_i = (x^i_{jk}),\ Y_j = (y^j_{ki}),\ Z_k = (z^k_{ij}),\ \text{avec } x^i_{jk} = y^j_{ki} = z^k_{ij} = a_{ijk}.$$

Les métriques choisies dans E, F et G sont celles associées aux matrices unités. D'après la définition de iX(jk) donnée au paragraphe 63, on a :

$$iX(jk) = (a_{11},\dots,a_{1m},a_{21},\dots,a_{2m},\dots,a_{n1},\dots,a_{nm})$$

où a_{jk} est la matrice colonne $(a_{ijk})_{i\in I}$. Par suite :

$$(a_{j1},\dots,a_{jm}) = (a_{ijk})_{\substack{i\in I\\ k\in K}} = (y^j_{ki})_{\substack{i\in I\\ k\in K}} = {}^tY_j.$$

On a alors :

$$iX(jk) = ({}^tY_1,\dots,{}^tY_n)$$

et :

$$\begin{aligned} iMi = iX(jk)Xi &= ({}^tY_1,\dots,{}^tY_n).{}^t({}^tY_1,\dots,{}^tY_n) \\ &= \sum_{j\in J} {}^tY_j.Y_j \\ &= U \end{aligned}$$

d'après la propriété 3 du §.50, où $U = (\Phi_I(X_i,X_{i'}))$.

De la même façon :

$$jPj = V \qquad \text{et } kQk = W.$$

Lemme 1.-

> Les matrices iMi, jPj et kQk du modèle de TUCKER sont égales aux matrices U, V et W de l'analyse triadique, lorsque les formes p, n et m respectivement définies sur E, F et G sont les formes canoniques.

On en déduit :

Lemme 2.-

> Dans les conditions du lemme 1, les matrices aAm, jBp et kCq et mMm, pPp et qQq du modèle de TUCKER sont respectivement égales aux matrices $\mathcal{U}$, $\mathcal{V}$ et $\mathcal{W}$ et D_α, D_β et D_γ de l'analyse triadique.

Il s'ensuit que les indices courants m, p et q utilisés par TUCKER sont respectivement les indices λ, μ et ν des trois familles de valeurs propres. On écrira donc λ, μ et ν à la place de m, p et q. On pose à présent :

(173) $$\lambda G(\mu\nu) = (G_1^J,\ldots,G_\mu^J,\ldots,G_{n'}^J)$$

où G_μ^J est une matrice p'×m'. On utilisera également la notation $\pi_j(v_\mu) = v_\mu^j$.

Dans ces conditions l'égalité (170) devient :

$$\lambda G(\mu\nu) = {}^t\mathcal{U}.({}^tY_1,\ldots,{}^tY_n).(\mathcal{V}\times\mathcal{W})$$

soit :

$$(G_1^J,\ldots,G_n^J) = ({}^t\mathcal{U}.{}^tY_1,\ldots,{}^t\mathcal{U}.{}^tY_n)\begin{pmatrix} v_1^1.\mathcal{W} & v_2^1.\mathcal{W} & \ldots & v_{n'}^1.\mathcal{W} \\ \ldots & \ldots & & \ldots \\ \ldots & \ldots & & \ldots \\ v_1^n.\mathcal{W} & v_2^n.\mathcal{W} & \ldots & v_{n'}^n.\mathcal{W} \end{pmatrix}$$

$$= ((\sum_j v_1^j.{}^t\mathcal{U}.{}^tY_j).\mathcal{W},\ldots,(\sum_j v_{n'}^j.{}^t\mathcal{U}.{}^tY_j).\mathcal{W})$$

soit pour tout $\mu = 1,\ldots,n'$:

$$G_\mu^J = \sum_{j\in J} v_\mu^j.{}^t\mathcal{U}.{}^tY_j.\mathcal{W}$$

ou encore :

(174) $$G_\mu^J = {}^t\mathcal{U}.(\sum_{j\in J} v_\mu^j.{}^tY_j).\mathcal{W} \quad .$$

L'application $d\eta$ étant ici l'isomorphisme canonique de F dans F^* associé à la forme canonique j, définie sur F, on a : $\pi_j^*(d\eta(v_\mu)) = \pi_j(v_\mu) = v_\mu^j$. Par suite la μ-ème composante principale $\hat{B}_\mu$ associée au μ-ème facteur J de l'analyse triadique est donnée par :

$$\hat{B}_\mu = \sum_{j\in J} \pi_j^*(d\eta(v_\mu)).Y_j = \sum_{j\in J} v_\mu^j.Y_j$$

et (174) s'écrit :

$$G_\mu^J = {}^t\mathcal{U}.{}^t\hat{B}_\mu.\mathcal{W}$$

soit, d'après (138) :

(175) $$G_\mu^J = {}^t\hat{H}_\mu^J \ .$$

En posant $\mu G(\lambda\nu) = (G_1^I,\ldots,G_\lambda^I,\ldots,G_{p'}^I)$ où G_λ^I est une matrice n'×m' de (171) on tire de même :

$$G_\lambda^I = \hat{H}_\lambda^I \ .$$

On a alors :

Lemme 3.-

Dans les conditions du lemme 1, les matrices mG(pq), pG(mq) et qG(mp) du modèle de TUCKER sont respectivement égales aux matrices

$({}^t\hat{H}_1^J, \ldots, {}^t\hat{H}_\mu^J, \ldots, {}^t\hat{H}_{n'}^J)$

$(\hat{H}_1^I, \ldots, \hat{H}_\lambda^I, \ldots, \hat{H}_{p'}^I)$

$({}^t\hat{H}_1^K, \ldots, {}^t\hat{H}_\nu^K, \ldots, {}^t\hat{H}_{m'}^K)$

de l'analyse triadique.

Corollaire 1.-

Le cube "central" G du modèle de TUCKER est égal au cube H de l'analyse triadique.

Des lemmes 1 et 2, du corollaire 1 et du fait que les expressions (155), (156) et (157) sont équivalentes à l'écriture (147) du modèle de TUCKER, on déduit immédiatement le résultat fondamental :

Théorème 1.-

Dans le cas particulier où les formes $\mathfrak{p}$, $\mathfrak{n}$ et $\mathfrak{m}$ de l'analyse triadique sont les formes canoniques sur E, F et G, les deux décompositions des données initiales :

$$a_{ijk} = \sum_m \sum_p \sum_q a_{im} . b_{jp} . c_{kq} . g_{mpq}$$

issue de la procédure de TUCKER, et :

$$a_{ijk} = \sum_\lambda \sum_\mu \sum_\nu u_\lambda^i . v_\mu^j . w_\nu^k . h_{\lambda\mu\nu}$$

issue de l'analyse triadique, sont identiques point par point.

65. Propriété caractéristique.

TUCKER indique la propriété suivante vérifiée par les termes du cube central :

(176) $$\sum_i \sum_j \sum_k x_{ijk}^2 = \sum_m \sum_p \sum_q g_{mpq}^2 .$$

D'après le théorème précédent cette propriété est également vraie en analyse triadique dans le cas singulier du théorème. Nous allons montrer que cette propriété est caractéristique de ce cas singulier.

Les matrices $\hat{H}^I_\lambda$ correspondent à une application de $(R^{m'},k)^* = G'^*$ dans $(R^{n'},j) = F'$, donc :

$$\begin{aligned}\|\hat{H}^I_\lambda\|^2_{\Phi_I} &= \mathrm{Trace}({}^t\hat{H}^I_\lambda.\hat{H}^I_\lambda)\\ &= \mathrm{Trace}({}^t\mathcal{W}.M.{}^t\hat{A}_\lambda.N.\mathcal{V}.{}^t\mathcal{V}.N.\hat{A}_\lambda.M.\mathcal{W})\\ &= \mathrm{Trace}({}^t\mathcal{W}.M.{}^t\hat{A}_\lambda.N.\hat{A}_\lambda.M.\mathcal{W})\\ &= \mathrm{Trace}(M.{}^t\hat{A}_\lambda.N.\hat{A}_\lambda.M.\mathcal{W}.{}^t\mathcal{W})\\ &= \mathrm{Trace}(M.{}^t\hat{A}_\lambda.N.\hat{A}_\lambda)\\ &= \|\hat{A}_\lambda\|^2_{\Phi_I}\\ &= \alpha_\lambda .\end{aligned}$$

Par ailleurs il est clair que $h^{I\lambda}_{\mu\nu}$ étant le terme général de $\hat{H}^I_\lambda$,

$$\|\hat{H}^I_\lambda\|^2_{\Phi_I} = \sum_\nu \sum_\mu (h^{I\lambda}_{\mu\nu})^2 = \sum_\mu \sum_\nu h^2_{\lambda\mu\nu} \qquad \text{(cf. (142))}$$

donc :

$$\sum_\lambda \|\hat{H}^I_\lambda\|^2_{\Phi_I} = \sum_\lambda \sum_\mu \sum_\nu h^2_{\lambda\mu\nu} = \sum_\lambda \alpha_\lambda \quad .$$

Enfin, on **a** vu en (117) et (118) :

$$\sum_i \sum_{i'} \sum_j \sum_{j'} \sum_k \sum_{k'} p_{ii'}.n_{jj'}.m_{kk'}.a_{ijk}.a_{i'j'k'} = \sum_\lambda \alpha_\lambda = \sum_\mu \beta_\mu = \sum_\nu \gamma_\nu$$

on a donc :

<u>Théorème</u> 2.-

Le cube initial $A = (a_{ijk})$ et le cube transformé $H = (h_{\lambda\mu\nu})$ vérifient :

(177)
$$\begin{aligned}\sum_i \sum_{i'} \sum_j \sum_{j'} \sum_k \sum_{k'} p_{ii'}.n_{jj'}.m_{kk'}.a_{ijk}.a_{i'j'k'} &= \sum_\lambda \sum_\mu \sum_\nu h^2_{\lambda\mu\nu}\\ &= \sum_\lambda \alpha_\lambda = \sum_\mu \beta_\mu = \sum_\nu \gamma_\nu\end{aligned}$$

On retrouve bien la propriété avancée par TUCKER lorsque les matrices P, N et M associées aux formes p, n et m sont les matrices unités.

67. <u>Reconstitution approximative des données.</u>

A plusieurs reprises TUCKER dans ([17]) insiste sur le fait que l'équation (147) de son modèle ne permet pas la reconstitution approchée des données, au sens des moindres carrés entre données initiales et données approchées, lorsque un ou plusieurs indices m, p et q des modes dérivés décrivent une partie commençante de son ensemble de variation. La raison en est que la troncature correspondante d'une face du cube central G, affecte du même coup les autres faces. CARROLL dans [5], du fait de l'inexistence d'un cube central dans son modèle, prétend que celui-ci réalise cette approximation. C'est pour des raisons semblables à celles avancées par TUCKER, que nous préférons l'expression (111) comme formule de reconstitution approchée des données :

$$X_i = \sum_{\lambda=1}^{p'} \pi_i(u_\lambda).\hat{A}_\lambda \ .$$

En effet, si on pose pour tout $i \in I$:

(178) $$X_i^r = \sum_{\lambda=1}^{r} \pi_i(u_\lambda).\hat{A}_\lambda \quad \text{où } r < p'$$

et :

(179) $$X_i^r = (a_{ijk}^r)_{j\in J, k\in K}$$

on est alors assuré de minimiser la quantité

(180) $$\sum_i \sum_j \sum_k p_i.p_j.p_k(a_{ijk} - a_{ijk}^r)^2$$

qui vaut alors $\sum_{\lambda=r+1}^{p'} \alpha_\lambda$, pour toute projection Φ_I-orthogonale de la famille $(x_i)_{i\in I}$ sur un sous-espace de dimension r de $L_R(G^*,F)$ et lorsque les matrices P , N et M sont des matrices diagonales de poids (ou pour toute projection i-orthogonale du nuage N_I sur un sous-espace de dimension r de E').

Remarquons cependant que si on opérait de même avec les matrices $(Y_j)_{j\in J}$ ou $(Z_k)_{k\in K}$ on n'obtiendrait pas la même approximation. Cela signifie évidemment que la minimisation de (180) dépend à la fois de la famille que l'on projette et du produit Φ correspondant.

III. Le modèle de CARROLL.

68. Résolution du modèle.

Pour trouver la décomposition du cube $A = (a_{ijk})$ sous la forme :

(181) $$a_{ijk} = \sum_{t=1}^{r} a_{it}.b_{jt}.c_{kt} \ ,$$

CARROLL procède de façon itérative par une succession de régressions semblables à la procédure NILES ([19]). Les b_{jt} et c_{kt} étant fixés arbitrairement, on détermine les a_{it} qui ajustent au mieux (181) au sens des moindres carrés. De la même façon, pour déterminer les b_{jt}, on donne aux a_{it} les valeurs trouvées précédemment et on conserve les valeurs arbitraires fixées pour les termes c_{kt}. Puis on procède de même avec les c_{kt}, et on recommence jusqu'à convergence.

Plus exactement avec les notations définies au paragraphe 48, on pose :

$X = (X_1,\dots,X_p)$ matrice $n\times(pm)$

$Y = (Y_1,\dots,Y_n)$ matrice $m\times(pn)$

$Z = (Z_1,\dots,Z_m)$ matrice $p\times(nm)$

et :

$A = (a_{it})$ matrice $p\times r$

$B = (b_{jt})$ matrice $n\times r$

$C = (c_{kt})$ matrice $m\times r$.

1.- On fixe B et C de façon arbitraire et on pose

$$G_1^I = (b_{jt}.c_{kt})_{\substack{j,k\in J\times K\\ r\in R=\{1,\dots,r\}}} \qquad \text{matrice } (nm)\times r.$$

Alors la solution des moindres carrés pour A est donnée par :

$$A_1 = Z.G_1^I.({}^tG_1^I.G_1^I)^{-1} .$$

2.- On pose $A = A_1$ et :

$$G_1^J = (a_{it}.c_{kt})' \qquad \text{matrice } (pm)\times n;$$

la solution pour B est donnée par :

$$B_1 = X.G_1^J.({}^tG_1^J.G_1^J)^{-1} .$$

3. On pose $A = A_1$ et $B = B_1$ et :

$$G_1^K = (a_{it}.b_{jt}) \qquad \text{matrice } (pn)\times m;$$

et C est approchée par :

$$C_1 = Y.G_1^K.({}^tG_1^K.G_1^K)^{-1}.$$

4.- On commence alors la deuxième étape avec $B = B_1$ et $C = C_1$, et :

$$G_2^I = (b_{jt} . c_{kt})$$

d'où une nouvelle approximation de A :

$$A_2 = Z.G_2^I.({}^tG_2^I.G_2^I)^{-1}$$

et le processus se poursuit ...

69. <u>Commentaires.</u>

La preuve de la convergence des différentes matrices A, B et C n'est pas apportée, mais en pratique il y a convergence. CARROLL prétend que pour un r donné la partie droite de (181) ajuste les a_{ijk} au sens des moindres carrés. Formellement, comme on l'a déjà dit, le modèle de CARROLL est un cas particulier de celui de TUCKER; cependant il ne semble pas que ce soit réellement le cas, au sens où les différentes matrices A, B et C ainsi que pseudo-cube central $G = (g_{\lambda\mu\nu})$ défini par $g_{\lambda\mu\nu} = 1$ si $\lambda = \mu = \nu = t$ et $g_{\lambda\mu\nu} = 0$ sinon, devraient vérifier les propriétés établies dans les pages précédentes. N'ayant pas de définitions algébriques de A, B et C et ne connais-sant pas les propriétés de ces matrices, la comparaison est ardue. Ni TUCKER, ni CARROLL, se citant mutuellement, ne l'ont effectuée.

Cependant si les deux modèles coïncidaient, pour tout λ, par exemple, on aurait :

$$\|\hat{H}_\lambda^I\|_\Phi^2 = \text{Trace}({}^t\hat{H}_\lambda^I.\hat{H}_\lambda^I) = \sum_{\mu,\nu} g_{\lambda\mu\nu} = 1 \text{ ou } 0,$$

suivant la valeur de λ. Or on a vu, au §.66, que pour tout :

$$\|\hat{H}_\lambda^I\|_\Phi^2 = \alpha_\lambda \quad .$$

ce qui signifierait $\alpha_\lambda = 1$ ou 0, pour tout λ, ce qui est absurde.

Notre opinion est donc qu'il n'y a pas de rapport entre les décomposi-tions des a_{ijk} proposées par TUCKER et CARROLL. En outre les indices m, n et p (ou λ, μ et ν) de TUCKER, se référant à trois familles distinctes de va-leurs propres, la réduction à un seul indice t, conforte notre point de vue.

On songe alors à une décomposition généralisant (152) ou (154). Dans les termes de notre analyse, la généralisation de (152) s'écrirait :

$$X_i = \sum_\lambda \pi_i(u_\lambda).\hat{A}_\lambda = \sum_\lambda u_\lambda^i.\hat{A}_\lambda$$

soit :

(182) $$a_{ijk} = \sum_{\lambda} u^i_{\lambda}.\hat{a}^{\lambda}_{jk}$$

où $\hat{a}^{\lambda}_{jk}$ est le terme général de la composante principale $\hat{A}_{\lambda}$.

Le problème reviendrait donc à trouver une décomposition de la matrice $\hat{A}_{\lambda}$ telle que :

$$\hat{a}^{\lambda}_{jk} = b^j_{\lambda}.c^k_{\lambda}$$

pour tout j et tout k.

D'après CARROLL les procédures NILES et ECKART-YOUNG sont équivalentes pour effectuer la décomposition d'une matrice. Cela signifie que la décomposition factorielle et la procédure d'ajustement itératif de CARROLL conduisent à la même décomposition de type (152) dans le cas d'une matrice. Dans le cas d'un cube, l'identité des deux procédures n'est pas prouvée, et cette hypothèse apparait même douteuse après ce que l'on vient de dire.

IV. Les opérateurs d'ESCOUFIER.

70. Rappel.

On se place dans le cadre habituel de l'analyse des données, où une matrice rectangulaire est considérée comme matrice associée à une application linéaire d'unespace dual dans un espace vectoriel. Cela signifie que les vecteurs colonnes des matrices sont les réalisations de variables sur un échantillon donné. Soit X_1 (resp. X_2) une matrice de dimensions $n{\times}m_1$ (resp. $n{\times}m_2$) et soient $F = (R^n,n)$, $G_1 = (R^{m_1},m_1)$ et $G_2 = (R^{m_2},m_2)$.

On a le double schéma de dualité suivant :

$$\begin{tikzcd}
G_1^* \arrow[r, "X_1"] & F \arrow[d, "N"] & G_2^* \arrow[l, "X_2"'] \\
G_1 \arrow[u, "M_1"] & F^* \arrow[l, "{}^tX_1"] \arrow[r, "{}^tX_2"'] & G_2 \arrow[u, "M_2"']
\end{tikzcd}$$

Soient u_1 et u_2 les endomorphismes de F de matrices associées :

$$U_1 = X_1.M_1.{}^tX_1.N$$
$$U_2 = X_2.M_2.{}^tX_2.N \quad .$$

On dit que u_1 et u_2 sont les <u>opérateurs d'ESCOUFIER</u> respectivement associés à X_1 et X_2.

Les opérateurs d'ESCOUFIER font partie de la classe des opérateurs u de F, de matrice associée $U = V.N$, où V est la matrice associée à une forme de F^*, symétrique et positive. Les opérateurs u engendrent un sous-espace vectoriel de $\mathcal{L}_R(F)$. Remarquons que u_1 (resp. u_2) est ce que nous avons appelé opérateur hermitien associé à ${}^tX_1$ (resp. ${}^tX_2$).

Sur le sous-ensemble des opérateurs u on définitle produit scalaire Π par :

$$\Pi(u_1,u_2) = \mathrm{Trace}(u_1 \circ u_2) = \mathrm{Trace}(U_1.U_2).$$

71. <u>Opérateurs à Π-distance nulle</u>.

Notons $N_1 = (x_1^i)_{i=1,\dots,n}$ (resp. $N_2 = (x_2^i)_{i=1,\dots,n}$) le nuage de G_1 (resp. G_2) associé aux n lignes de X_1 (resp. X_2). On a alors :

$$X_1 = {}^t(x_1^1,\dots,x_1^n)$$
$$X_2 = {}^t(x_2^1,\dots,x_2^n)$$

et supposons qu'il existe une application linéaire $h : G_2 \to G_1$ faisant correspondre à chaque vecteur de N_2 le vecteur de même indice de N_1 :

$$i = 1,\dots,n, \forall i \qquad h(x_2^i) = x_1^i.$$

Si H est la matrice $m_1 \times m_2$ associée à h, on a :

$$H.{}^tX_2 = {}^tX_1$$

et par suite :

$$U_1 = X_1.M_1.{}^tX_1.N = X_2.{}^tH.M_1.H.{}^tX_2.N \ .$$

Ainsi, lorsque :

(183) $\qquad {}^tH.M_1.H = M_2$

il vient :

$$U_1 = U_2.$$

Pour tous x et y de G_2, la relation (183) s'écrit aussi :

$$\mathfrak{m}_1(h(x),h(y)) = \mathfrak{m}_2(x,y),$$

et en particulier, si $(e_1,\ldots,e_{m_2})$ est une base m_2-orthonormée de G_2, on a :

$$m_1(h(e_k),h(e_{k'})) = \delta_{kk'}.$$

Cette dernière relationassure que h est injective.

En résumé, on peut dire :

i) Si il existe une application injective h de G_2 dans G_1 telle que, au nuage N_2 corresponde termes à termes le nuage N_1, alors les opérateurs d'ESCOUFIER correspondants u_1 et u_2 sont égaux.

Dans le cas particulier où $G_1 = G_2 = G = (R^m, m)$ et $M = I$, la relation (183) devient :

$$^tH.H = I\ ,$$

et par suite le nuage N_1 se déduit de N_2 par des rotations et des symétries.

ii) Si deux nuages N_1 et N_2 de G se déduisent l'un de l'autre par des rotations et des symétries, leurs opérateurs associés sont à Π-distance nulle.

Dans une optique d'analyse en C.P les représentations des matrices X_1 et X_2 seront donc confondues. Ce n'est pas le cas dans l'analyse que nous avons présentée, puisque :

$$d\Phi(X_1,X_2) = 0 \iff X_1 = X_2.$$

72. <u>Comparaison</u>.

On se place ici sur le terrain de l'analyse en C.P que nous avons proposée. Les notations sont celles du chapitre III. On cherche une description des vecteurs $(V_j)_{j\in J}$. Si j et ℓ sont deux éléments de J on a le double schéma de dualité :

$$\begin{tikzcd}
G^* \arrow[r, "X_j"] & F \arrow[d, "N"] & G^* \arrow[l, "X_\ell"'] \\
G \arrow[u, "M"] & F^* \arrow[l, "{}^tX_j"] \arrow[r, "{}^tX_\ell"'] & G \arrow[u, "M"']
\end{tikzcd}$$

et l'on pose :

$$V_{\ell j} = {}^tX_\ell.N.X_j$$
$$V_{j\ell} = {}^tX_j.N.X_\ell = {}^tV_{\ell j}$$
$$W_{jj} = X_j.M.{}^tX_j$$
$$W_{\ell\ell} = X_\ell.M.{}^tX_\ell \quad .$$

U_j et U_ℓ étant les opérateurs d'ESCOUFIER associés à X_j et X_ℓ on a :

$$\Pi(U_j,U_\ell) = \mathrm{Trace}(X_j.M.{}^tX_j.N.X_\ell.M.{}^tX_\ell.N)$$

soit :

(184) $$\Pi(U_j,U_\ell) = \mathrm{Trace}(M.{}^tX_j.N.X_\ell.M.{}^tX_\ell.N.X_j)$$

(185) $$\Pi(U_j,U_\ell) = \mathrm{Trace}(N.X_j.M.{}^tX_j.N.X_\ell.M.{}^tX_\ell).$$

De (184) on tire :

(186) $$\Pi(U_j,U_\ell) = \Phi_{m^{-1}m}(V_{\ell j},V_{\ell j}) = \|V_{\ell j}\|^2_{\Phi_{m^{-1}m}} = \|V_{j\ell}\|^2_{\Phi_{m^{-1}m}}$$

et de (185) :

(187) $$\Pi(U_j,U_\ell) = \Phi_{n^{-1}n}(W_{\ell\ell},W_{jj}) = \Phi_{n^{-1}n}(W_{jj},W_{\ell\ell}).$$

Si les éléments de l'échantillon I sont munis de poids p_i, si $N = D_{P_I}$ et $M = I$, et si les matrices X_j sont centrées pour D_{P_I}, alors :

$$V_{\ell j} = (\mathrm{cov}(V_j^k,V_\ell^{k'}))_{\substack{k\in K\\ k'\in K}} \quad .$$

Par ailleurs, W_{jj} est la matrice des produits scalaires des n vecteurs lignes de la matrice X_j, c'est-à-dire des éléments du nuage N_j correspondant;

Définitions.-

On appelle <u>Π-analyse en composantes principales d'une famille d'opérateurs d'ESCOUFIER</u> $(U_j)_{j\in J}$, associés à la famille $(V_j)_{j\in J}$, l'analyse en composantes principales issue de la décomposition factorielle de U.P , où $U = (\Pi(U_j,U_\ell))$.

On appelle <u>Φ-analyse en composantes principales d'une famille</u> $(X_j)_{j\in J}$ l'analyse en composantes principales issue de la décomposition factorielle de U.P , où $U = (\Phi(X_j,X_\ell))$.

De (187) on obtient :

Théorème 3.-

> La Π-analyse en composantes principales de la famille d'opérateurs d'ESCOUFIER $(U_j)_{j\in J}$ associée à la famille $(X_j)_{j\in J}$ est identique à la Φ-analyse en composantes principales de la famille $(W_{jj})_{j\in J}$, où W_{jj} est la matrice des produits scalaires des lignes de X_j.

De (186) il vient également :

Théorème 4.-

> Si $N = D_{P_I}$ et $M = I$, et si les matrices X_j sont centrées pour D_{P_I}, la matrice U de la Π-analyse en composantes principales de la famille d'opérateurs d'ESCOUFIER $(U_j)_{j\in J}$ associée à la famille $(X_j)_{j\in J}$ a pour terme général :

(188) $$u_{j\ell} = \|V_{j\ell}\|^2_{\Phi_{m^{-1}m}} = \sum_{k\in K} \sum_{k'\in K} \mathrm{cov}^2(V_j^k, V_\ell^{k'}).$$

Ces deux théorèmes indiquent donc la différence entre les deux analyses. Notamment les relations (75) et (73) montrent que sous les hypothèses du théorème 4, le terme général $u_{j\ell}$, de la matrice U, dans le cas de la Φ-analyse de la famille $(V_j)_{j\in J}$ est donné par :

$$u_{j\ell} = \sum_{k\in K} \mathrm{cov}(V_j^k, V_\ell^k) .$$

Ajoutons enfin, que dans le cas particulier où $m = 1$, la Π-analyse des opérateurs associés à la famille $(V_j)_{j\in J}$ des variables ne coïncide pas avec l'analyse en C.P classique de la matrice $n\times p$ correspondante. Les vecteurs x_j et x_ℓ associés aux deux matrices colonnes X_j et X_ℓ vérifient en effet :

$$\Pi(U_j, U_\ell) = n^2(x_j, x_\ell).$$

La matrice U de l'opérateur $U.P$ est donc la matrice des carrés des produits scalaires au lieu d'être la matrice des produits scalaires.

Chapitre VII

ANALYSE CANONIQUE DE DEUX FAMILLES FINIES DE VECTEURS ALEATOIRES

I. Problématique.

73. Introduction.

On sait qu'il ya deux façons distinctes d'envisager la généralisation de l'analyse canonique. L'une consiste en l'analyse de m familles de variables dont on connait les réalisations sur un échantillon. Si $m = 2$ on retrouve l'analyse canonique classique. On pose $K = \{1,\dots,m\}$ et $F_k = (V_k^j)_{j\in J_k}$ où V_k^j est une variable et $J_k = \{1,\dots,p_k\}$. La matrice X_k, de dimension $n\times p_k$, exprime la réalisation de la famille F_k sur l'échantillon I de cardinal n. Le problème consiste alors à rechercher, pour chaque famille F_k, des combinaisons linéaires des variables de la famille, telles que ces combinaisons, pour la totalité des familles, satisfassent à des conditions dérivées de l'analyse classique, et interprétables dans les termes de la statistique descriptive. Une vision plus algébrique de cette problématique, rejoignant le problème général de PROCRUSTE, est de rechercher m matrices de transformation T_k, telles que la totalité des matrices transformées $X_k.T_k$, soient aussi proches que possible les unes des autres, en un certain sens.

Dans cette voie, où la littérature est abondante, on peut citer par exemple [10], [12] et [15].

L'autre généralisation possible consiste en l'étude de deux familles de vecteurs, dont on connait les réalisations sur un échantillon I. Plus exactement on recherchera des combinaisons linéaires des vecteurs d'une famille, puis de l'autre famille, telles que ces combinaisons linéaires satisfassent également des conditions dérivées de l'analyse canonique classique. Si pour tout vecteur, la dimension est réduite à 1, on retrouve alors l'analyse habituelle.

74. <u>Notations.</u>

Les notations sont encore celles du chapitre III, avec les précisions suivantes : $I = \{1,\dots,n\}$ est un échantillon, $J_1 = \{1,\dots,p\}$ est un ensemble d'indices relatifs à une famille de p variables vectorielles V_j^1, $j \in J_1$, et $J_2 = \{1,\dots,q\}$ est un ensemble d'indices relatifs à une seconde famille de q variables vectorielles : V_ℓ^2, $\ell \in J_2$. On pose $K = \{1,\dots,m\}$ et

$$V_j^1 = (V_j^{1k})_{k\in K} \qquad V_\ell^2 = (V_\ell^{2k})_{k\in K}.$$

Pour fixer les idées, si k par exemple, fait référence à une "circonstance d'observation", les notations adoptées impliquent que V_j^{1k} et V_ℓ^{2k}, pour tout j et tout ℓ se rapportent à la même circonstance k. De plus elles sont observées sur le même échantillon I. On note alors X_j^1 (resp. X_ℓ^2) la matrice n×m des réalisations de V_j^1 (resp. V_ℓ^2) sur I.

Les matrices $(X_j^1)_{j\in J_1}$ et $(X_\ell^2)_{\ell\in J_2}$ fournissent deux tableaux réels à trois indices :

$$A^1 = (a_{ijk}^1)_{\substack{i\in I\\ j\in J_1\\ k\in K}} \qquad A^2 = (a_{ijk}^2)_{\substack{i\in I\\ j\in J_2\\ k\in K}} .$$

On supposera enfin que les espaces euclidiens $E_1 = (R^p,p)$, $E_2 = (R^q,q)$, $F = (R^n,n)$ et $G = (R^m,m)$ sont tels que : $P = I$, $Q = I$, $N = D_{P_I}$, matrice diagonale de poids, et $M = I$.

En fait, on pourrait, sans grandes modifications de ce qui va suivre, choisir des matrices diagonales de poids, pour toutes les matrices associées aux formes définies sur ces espaces. Le choix que nous imposons ici a seulement pour but de simplifier les écritures de ce chapitre.

Dans ces conditions, il résulte que, pour tous j et j', et pour tous ℓ et ℓ', on a :

(189) $\Phi_{J_1}(X^1_j, X^1_{j'}) = \text{Trace}({}^tX^1_{j'} . D_{P_I} . X^1_j)$

(190) $\Phi_{J_2}(X^2_\ell, X^2_{\ell'}) = \text{trace}({}^tX^2_{\ell'} . D_{P_I} . X^2_\ell)$. Les formes Φ_{J_1} et Φ_{J_2} étant égales, on notera :

(191) $\Phi = \Phi_{J_1} = \Phi_{J_2}$.

Les applications linéaires x^1_j et x^2_ℓ respectivement associées à X^1_j et X^2_ℓ appartiennent à $L_R(G^*, F)$. Donc la notation $\Phi(X^1_j, X^2_\ell)$ est autorisée quels que soient j dans J_1 et ℓ dans J_2.

75. <u>Formes sur E^*_1 et E^*_2.</u>

D'après (73) on peut écrire :

(192) $\Phi(X^1_j, X^1_{j'}) = \text{COV}(V^1_j, V^1_{j'}) = v^1_{jj'}$

(193) $\Phi(X^2_\ell, X^2_{\ell'}) = \text{COV}(V^2_\ell, V^2_{\ell'}) = v^2_{\ell\ell'}$.

On note alors $V_{11} = (v^1_{jj'})_{j\in J_1, j'\in J_1}$ la matrice des variances-covariances empiriques des vecteurs de la première famille, et $V_{22} = (v^2_{\ell\ell'})_{\ell\in J_2, \ell'\in J_2}$, celle relative à la seconde famille. On supposera que la famille $(x^1_j)_{j\in J_1}$ (resp. $(x^2_\ell)_{\ell\in J_2}$) est linéairement indépendante, c'est à dire qu'elle engendre un sous-espace L_1 (resp. L_2) de $L_R(G^*, F)$ de dimension p (resp. q).

Cela signifie également que V_{11} (resp. V_{22}) est de rang p (resp. q), ou encore qu'elle admet p (resp. q) valeurs propres strictement positives. Compte tenu de cette restriction, on dira que V_{11} (resp. V_{22}) est la matrice associée à la forme bilinéaire symétrique non dégénérée positive, notée $\mathbf{v}_{11}$ (resp. $\mathbf{v}_{22}$), définie sur R^{p*} (resp. R^{q*}). On écrira alors : $E^*_1 = (R^{p*}, \mathbf{v}_{11})$ et $E^*_2 = (R^{q*}, \mathbf{v}_{22})$.

Remarque : d'après les propriétés 3 du §.50, et avec les notations utilisées ici, on a :

(194) $$V_{11} = \sum_{i\in I} p_i . {}^tY_i^1 . Y_i^1 = \sum_{k\in K} Z_k^1 . D_{P_I} . {}^tZ_k^1 ,$$

(195) $$V_{22} = \sum_{i\in I} p_i . {}^tY_i^2 . Y_i^2 = \sum_{k\in K} Z_k^2 . D_{P_I} . {}^tZ_k^2 ,$$

où Y_i^1 et Z_k^1 (resp. Y_i^2 et Z_k^2) sont les matrices m×p et p×n issues des deux faces du cube A^1 (resp. A^2) ainsi que cela est indiqué au §.48.

Toutefois ces écritures qui permettent le calcul rapide de V_{11} et V_{22} ne sont plus valables si M ou P et Q sont également des matrices diagonales de poids.

76. Application linéaire de E_2^* dans E_1 (resp. E_1^* dans E_2).

Toujours d'après (73) on a :

(196) $$\Phi(X_j^1, X_\ell^2) = \mathrm{COV}(V_j^1, V_\ell^2) = v_{j\ell}^{12} .$$

On note alors $V_{12} = (v_{j\ell}^{12})_{j\in J_1, \ell\in J_2}$ la matrice p×q des covariances empiriques des vecteurs de la première famille avec ceux de la seconde, et $V_{21} = {}^tV_{12}$. Dans ces conditions, on obtient une application linéaire v_{12} (resp. $v_{21} = {}^tv_{12}$) de E_2^* dans E_1 (resp. de E_1^* dans E_2), de matrice V_{12} (resp. V_{21}); l'isomorphisme dv_{11} (resp. dv_{22}) de E_1^* sur E_1 (resp. E_2^* sur E_2) associé à V_{11} (resp. V_{22}) permet d'écrire le schéma suivant :

$$\begin{tikzcd}
E_1^* \arrow[r, "f_{J_1}"] \arrow[drr, "v_{21}"] \arrow[d, shift left, "dv_{11}"] & L_R(G^*, F) & E_2^* \arrow[l, "f_{J_2}"'] \arrow[dll, "v_{12}"'] \arrow[d, shift right, "dv_{22}"'] \\
E_1 \arrow[u, shift left, "dv_{11}^{-1}"] & & E_2 \arrow[u, shift right, "dv_{22}^{-1}"']
\end{tikzcd}$$

où f_{J_1} et f_{J_2} sont définis par (51) :

(197) $$\alpha \in E_1^* \qquad f_{J_1}(\alpha) = \sum_{j\in J_1} \pi_j^{1*}(\alpha) . x_j^1$$

(198) $$\beta \in E_2^* \qquad f_{J_2}(\beta) = \sum_{\ell\in J_2} \pi_\ell^{2*}(\beta) . x_\ell^2 .$$

Remarque : comme précédemment on a l'expression :

(199) $$V_{12} = \sum_{i\in I} p_i . {}^tY_i^1 . Y_i^2 = \sum_{k\in K} Z_k^1 . D_{P_I} . {}^tZ_k^2 .$$

77. Problématique.

Comme en (62) et (63), si $\alpha \in E_1^*$ et $\beta \in E_2^*$, on note $F^1(\alpha)$ et $F^2(\beta)$ les vecteurs aléatoires définis par :

(200) $$F^1(\alpha) = \sum_{j\in J_1} \pi_j^{1*}(\alpha).V_j^1$$

(201) $$F^2(\beta) = \sum_{\ell\in J_2} \pi_\ell^{2*}(\beta).V_\ell^2 ,$$

Les matrices $F^1(\alpha)$ et $F^2(\beta)$ de leurs réalisations surI vérifient :

(202) $$F^1(\alpha) = \sum_{j\in J_1} \pi_j^{1*}(\alpha).X_j^1$$

(203) $$F^2(\beta) = \sum_{\ell\in J_2} \pi_\ell^{2*}(\beta).X_\ell^2$$

On notera $f_{J_1}(\alpha)$ et $f_{J_2}(\beta)$ les applications linéaires de G^* dans F de matrices $F^1(\alpha)$ et $F^2(\beta)$.

Comme en analyse canonique classique, la problématique de l'analyse canonique généralisée présentée ici est alors la suivante :

i) On cherche le couple (α_1,β_1) de $E_1^*\times E_2^*$ réalisant le maximum de

$$COR(F^1(\alpha),F^2(\beta))$$

sous les contraintes : $VAR(F^1(\alpha)) = VAR(F^2(\beta)) = 1$.

ii) On cherche le couple (α_2,β_2) de $E_1^*\times E_2^*$ réalisant le maximum de

$$COR(F^1(\alpha),F^2(\beta))$$

sous les contraintes $VAR(F^1(\alpha)) = VAR(F^2(\beta)) = 1$ et $COR(F^1(\alpha_1),F^1(\alpha)) = COR(F^2(\beta_2),F^2(\beta)) = 0$.

iii) etc ...

Pour tout r on pose : $F_r^1 = F^1(\alpha_r)$, $F_r^2 = F^2(\beta_r)$, $F_r^1 = F^1(\alpha_r)$, $F_r^2 = F^2(\beta_r)$, $f_r^1 = f_{J_1}(\alpha_r)$ et $f_r^2 = f_{J_2}(\beta_r)$.

Définition.-

Les couples (F_r^1,F_r^2) sont les couples de variables (vectorielles) canoniques; les couples (f_r^1,f_r^2) sont les couples de vecteurs canoniques.

Remarque : suivant le type de transformation que l'on fera subir aux données initiales, on interprétera la maximisation de $COR(F^1(\alpha),F^2(\beta))$ ainsi qu'il est indiqué dans la section IV du chapitre III. En particulier, si les données initiales sont fortement normalisées (cf. §.42), maximiser ce coefficient reviendra à rechercher la combinaison linéaire $\sum_{j\in J_1} \pi_j^*(\alpha).V_{jk}^1$ de la famille $(V_{jk}^1)_{j\in J_1}$ des k-èmes composantes des vecteurs $(V_j^1)_{j\in J_1}$, et la combinaison linéaire $\sum_{\ell\in J_2} \pi_\ell^*(\beta).V_{\ell k}^2$ de la famille $(V_{\ell k}^2)_{\ell\in J_2}$ des k-èmes composantes des vecteurs $(V_\ell^2)_{\ell\, J_2}$, telles que :

$$\sum_{k\in K} cor(\sum_{j\in J_1} \pi_j^*(\alpha).V_{jk}^1, \sum_{\ell\ J_2} \pi_\ell^*(\beta).V_{\ell k}^2)$$

soit maximum.

78. Equivalences.

Les différentes matrices X_j^1 et X_ℓ^2 étant centrées pour D_{P_I}, les matrices $F^1(\alpha)$ et $F^2(\beta)$ le sont aussi. D'après (70) il vient :

$$VAR(F^1(\alpha)) = \|F^1(\alpha)\|_\Phi^2 = \|f_{J_1}(\alpha)\|_\Phi^2$$

$$VAR(F^2(\beta)) = \|F^2(\beta)\|_\Phi^2 = \|f_{J_2}(\beta)\|_\Phi^2$$

Théorème 1.-

Pour tout α de E_1^* (resp. tout β de E_2^*) $VAR(F^1(\alpha)) = 1$ ssi $\|f_{J_1}(\alpha)\|_\Phi^2 = 1$ (resp. $VAR(F^2(\beta)) = 1$ ssi $\|f_{J_2}(\beta)\|_\Phi^2 = 1$).

D'après (52) et le théorème 6 du §.21 :

(204) $\qquad (\alpha,\alpha')\in E_1^*\times E_1^* \qquad V_{11}(\alpha,\alpha') = \Phi(f_{J_1}(\alpha),f_{J_1}(\alpha'))$

(205) $\qquad (\beta,\beta')\in E_2^*\times E_2^* \qquad V_{22}(\beta,\beta') = \Phi(f_{J_2}(\beta),f_{J_2}(\beta'))$

d'où l'on déduit :

Théorème 2.-

Pour tous α et α' de E_1^*, $f_{J_1}(\alpha)$ et $f_{J_1}(\alpha')$ sont Φ-orthonormés ssi α et α' sont V_{11}-orthonormés. Pour tous β et β' de E_2^*, $f_{J_2}(\beta)$ et

$f_{J_2}(\beta')$ sont Φ-orthonormés ssi β et β' sont V_{22}-orthonormés.

La relation (98) nous conduit au

<u>Théorème</u> 3.-

Pour tous α et α' de E_1^* la relation $\|f_{J_1}(\alpha)\|_\Phi^2 = \|f_{J_1}(\alpha')\|_\Phi^2 = 1$ implique $COR(F^1(\alpha),F^1(\alpha')) = \Phi(f_{J_1}(\alpha),f_{J_1}(\alpha'))$. On a un résultat semblable pour tous β et β' de E_2^*.

Du théorème 3 et de (204) et (205), on déduit :

<u>Théorème</u> 4.-

Pour tous α et α' de E_1^*, si $f_{J_1}(\alpha)$ et $f_{J_1}(\alpha')$ sont Φ-normés, les trois propriétés suivantes sont équivalentes :

i) $COR(F^1(\alpha),F^1(\alpha')) = 0$;

ii) $f_{J_1}(\alpha)$ et $f_{J_1}(\alpha')$ sont Φ-orthogonaux;

iii) α et α' sont V_{11}-orthogonaux.

On a un résultat semblable pour tous β et β' de E_2^*.

Le théorème 3 est encore vrai si on remplace $f_{J_1}(\alpha')$ par $f_{J_2}(\beta)$:

$$\|f_{J_1}(\alpha)\|_\Phi^2 = \|f_{J_2}(\beta)\|_\Phi^2 = 1 \Rightarrow COR(F^1(\alpha),F^2(\beta)) = \Phi(f_{J_1}(\alpha),f_{J_2}(\beta)).$$

Par ailleurs le cosinus de l'angle des vecteurs $f_{J_1}(\alpha)$ et $f_{J_2}(\beta)$ dans $L_R(G^*,F)$ est donné par :

$$\cos(f_{J_1}(\alpha),f_{J_2}(\beta)) = \frac{\Phi(f_{J_1}(\alpha),f_{J_2}(\beta))}{\|f_{J_1}(\alpha)\|_\Phi \|f_{J_2}(\beta)\|_\Phi}$$

d'où le théorème suivant :

<u>Théorème</u> 5.-

Pour tout α de E_1^* et tout β de E_2^*, la relation $\|f_{J_1}(\alpha)\|_\Phi^2 = \|f_{J_2}(\beta)\|_\Phi^2 = 1$ implique :

$$\cos(f_{J_1}(\alpha),f_{J_2}(\beta)) = COR(F^1(\alpha),F^2(\beta)).$$

II. Méthode de résolution.

79. <u>Interprétation géométrique</u>.

D'après les théorèmes précédents on peut interpréter géométriquement la problématique exposée au §.77.

i) On cherche le couple (α_1,β_1) de $E_1^*\times E_2^*$ réalisant le maximum de $\cos(f_{J_1}(\alpha),f_{J_2}(\beta))$ sous les contraintes $\|f_{J_1}(\alpha)\|_\Phi^2 = \|f_{J_2}(\beta)\|_\Phi^2 = 1$.

ii) On cherche le couple (α_2,β_2) de $E_1^*\times E_2^*$ réalisant le maximum de $\cos(f_{J_1}(\alpha),f_{J_2}(\beta))$ sous les contraintes $\|f_{J_1}(\alpha)\|_\Phi^2 = \|f_{J_2}(\beta)\|_\Phi^2 = 1$ et $\cos(f_{J_1}(\alpha_1),f_{J_1}(\alpha)) = 0$ et $\cos(f_{J_2}(\beta_1),f_{J_2}(\beta)) = 0$.

iii) etc ...

A ce stade le problème posé ici est donc strictement identique à celui de l'analyse canonique classique. La solution suivra les mêmes voies.

Les deux sous-espaces $L_1 = \mathrm{Im}(f_{J_1})$ et $L_2 = \mathrm{Im}(f_{J_2})$ de $\mathcal{L}_R(G^*,F)$ sont respectivement engendrés par les $(x_j^1)_{j\in J_1}$ et les $(x_\ell^2)_{\ell\in J_2}$; on notera p_{L_1} et p_{L_2} les projecteurs Φ-orthogonaux de $\mathcal{L}_R(G^*,F)$ sur L_1 et L_2. On sait alors, (cf. [4]), que les vecteurs canoniques f_r^1 et f_r^2 sont respectivement les vecteurs propres de $p_{L_1}\circ p_{L_2}$ et $p_{L_2}\circ p_{L_1}$, associés à la même valeur propre λ_r.

80. <u>Projection d'un vecteur de</u> L_1 <u>sur</u> L_2.

Soient $\alpha\in E_1^*$ et $\hat\alpha\in E_2^*$ satisfaisants à :

(206) $$p_{L_2}(f_{J_1}(\alpha)) = f_{J_2}(\hat\alpha).$$

Cette relation est équivalente à :

(207) $$\ell = 1,\dots,q \qquad \Phi(f_{J_1}(\alpha)-f_{J_2}(\hat\alpha),x_\ell^2) = 0$$

qui s'écrit aussi :

$$\ell = 1,\dots,q \qquad \langle x_\ell^2, d\Phi(f_{J_1}(\alpha)-f_{J_2}(\hat\alpha))\rangle = 0$$

soit :

$$\ell = 1,\dots,q \qquad \langle f_{J_2}(f_\ell^{2*}), d\Phi(f_{J_1}(\alpha)-f_{J_2}(\hat\alpha))\rangle = 0$$

où $(f_\ell^*)_{\ell\, J_2}$, est la base de E_2^* duale de la base canonique de E_2. Cette dernière relation s'écrit également :

$$\ell = 1,\dots,q \qquad \langle f_\ell^{2*}, {}^t f_{J_2}\circ d\Phi(f_{J_1}(\alpha)-f_{J_2}(\hat\alpha))\rangle_* = 0$$

soit :

(208) $${}^t f_{J_2}\circ d\Phi\circ f_{J_1}(\alpha) = {}^t f_{J_2}\circ d\Phi\circ f_{J_2}(\hat\alpha).$$

En premier lieu, V_{22} étant la matrice des produits scalaires des éléments de la famille $(x^2_\ell)_{\ell\in J_2}$, de (31) il vient :

(209) $$dV_{22} = {}^t f_{J_2} \circ d\Phi \circ f_{J_2} \ .$$

Cette relation peut aussi s'établir à partir de (205).

En second lieu, sachant que v_{21} est l'application linéaire de E_1^* dans E_2, ayant pour matrice associée (relativement à la base duale dans E_1^* et à la base canonique dans E_2) la matrice V_{21} de terme général :

$$\Phi(x^2_\ell, x^1_j) = \Phi(f_{J_2}(f^{2*}_\ell), f_{J_1}(f^{1*}_j))$$

il vient pour tout $j \in J_1$ et tout $\ell \in J_2$:

$$\begin{aligned} \langle v_{21}(f^{1*}_j), f^{2*}_\ell \rangle &= \Phi(f_{J_2}(f^{2*}_\ell), f_{J_1}(f^{1*}_j)) \\ &= \langle f_{J_2}(f^{2*}_\ell), d\Phi(f_{J_1}(f^{1*}_j)) \rangle \\ &= \langle f^{2*}_\ell, {}^t f_{J_2} \circ d\Phi(f_{J_1}(f^{1*}_j)) \rangle_* \\ &= \langle {}^t f_{J_2} \circ d\Phi \circ f_{J_1}(f^{1*}_j), f^{2*}_\ell \rangle \ . \end{aligned}$$

On en déduit donc :

(210) $$v_{21} = {}^t f_{J_2} \circ d\Phi \circ f_{J_1} \ .$$

En reportant les expressions (209) et (210) dans (208) on a :

(211) $$v_{21}(\alpha) = dV_{22}(\hat{\alpha}) .$$

La matrice V_{22} étant de rang q, l'application dV_{22} est bijective et il vient alors :

(212) $$\hat{\alpha} = dV_{22}^{-1} \circ v_{21}(\alpha) \ .$$

Si $\hat{\beta}$ est l'élément de E_1^* tel que pour $\beta \in E_2^*$ on ait :

$$P_{L_1}(f_{J_2}(\beta)) = f_{J_1}(\hat{\beta})$$

on montre de même que $\hat{\beta}$ est donné par :

(213) $$\hat{\beta} = dV_{11}^{-1} \circ v_{12}(\beta) .$$

Le théorème suivant reprend ces résultats :

Théorème 6.-

On a les égalités suivantes entre applications linéaires :

$$p_{L_2} \circ f_{J_1} = f_{J_2} \circ dV_{22}^{-1} \circ v_{21}$$

$$p_{L_1} \circ f_{J_2} = f_{J_1} \circ dV_{11}^{-1} \circ v_{12} \; .$$

C'est ce théorème qui va permettre le calcul effectif des variables et des vecteurs canoniques.

81. Calcul des couples de vecteurs canoniques.

L'application $f_{J_1}(\alpha)$ est vecteur canonique ssi c'est un vecteur propre de $p_{L_1} \circ p_{L_2}$, c'est à dire :

$$p_{L_1} \circ p_{L_2}(f_{J_1}(\alpha)) = \lambda . f_{J_1}(\alpha).$$

Or d'après le théorème 6 :

$$p_{L_1} \circ p_{L_2}(f_{J_1}(\alpha)) = p_{L_1}(f_{J_2}(\hat{\alpha})) = f_{J_1}(\hat{\hat{\alpha}}).$$

Pour un vecteur canonique $f_{J_1}(\alpha)$ on a donc :

(214) $$f_{J_1}(\hat{\hat{\alpha}} - \lambda . \alpha) = 0$$

et comme f_{J_1} est injective, puisque la famille $(x_j^1)_{j \in J_1}$ est linéairement indépendante, on a $\hat{\hat{\alpha}} = \lambda . \alpha$ soit :

$$dV_{11}^{-1} \circ v_{12} \circ dV_{22}^{-1} \circ v_{21}(\alpha) = \lambda \alpha$$

On obtiendrait un résultat identique pour les vecteurs propres de $p_{L_2} \circ p_{L_1}$, d'où :

Théorème 7.-

> L'application $f_{J_1}(\alpha)$ est vecteur propre de $p_{L_1} \circ p_{L_2}$ ssi α est vecteur proprede $dV_{11}^{-1} \circ v_{12} \circ dV_{22}^{-1} \circ v_{21}$ pour la même valeur propre. De même $f_{J_2}(\beta)$ est vecteur propre de $p_{L_2} \circ p_{L_1}$ ssi β est vecteur propre de $dV_{22}^{-1} \circ v_{21} \circ dV_{11}^{-1} \circ v_{12}$ pour la même valeur propre.

82. Dualité.

Une partie du schéma présenté au §.76 peut s'écrire :

$$\begin{tikzcd}[row sep=large, column sep=huge]
E_1^* \arrow[r, "v_{21}"] \arrow[d, shift right, "dV_{11}"'] & E_2 \arrow[d, shift right, "dV_{22}^{-1}"'] \\
E_1 \arrow[u, shift right, "dV_{11}^{-1}"'] & E_2^* \arrow[l, "v_{12}={}^t v_{21}"] \arrow[u, shift right, "dV_{22}"']
\end{tikzcd}$$

On considère alors que R^p et R^q sont respectivement munis des produits scalaires V_{11}^{-1} et V_{22}^{-1}, c'est à dire que $E_1 = (R^p, V_{11}^{-1})$ et $E_2 = (R^q, V_{22}^{-1})$.

Dans ces conditions, si v_{12}^* désigne l'adjoint de v_{12} relativement aux formes V_{22} et V_{11}^{-1} de E_2^* et E_1, on a :

$$v_{12}^* = dV_{22}^{-1} \circ {}^t v_{12} \circ dV_{11}^{-1} \qquad \text{cf. (11).}$$

Par suite :

(215) $$dV_{22}^{-1} \circ v_{21} \circ dV_{11}^{-1} \circ v_{12} = v_{12}^* \circ v_{12}$$

et de même :

(216) $$dV_{11}^{-1} \circ v_{12} \circ dV_{22}^{-1} \circ v_{21} = v_{21}^* \circ v_{21} \ .$$

On peut alors appliquer les résultats de la remarque finale du B) du §.13.

<u>Théorème</u> 8.-

i) Les opérateurs hermitiens $dV_{11}^{-1} \circ v_{12} \circ dV_{22}^{-1} \circ v_{21}$ de E_1^* et $dV_{22}^{-1} \circ v_{21} \circ dV_{11}^{-1} \circ v_{12}$ de E_2^* admettent le même système de valeurs propres non nulles (λ_r). Celles-ci sont en outre positives.

ii) Il existe un système (α_r) (resp. (β_r)), V_{11}-orthonormé (resp. V_{22}-orthonormé) de vecteurs propres de $dV_{11}^{-1} \circ v_{12} \circ dV_{22}^{-1} \circ v_{21}$ (resp. $dV_{22}^{-1} \circ v_{21} \circ dV_{11}^{-1} \circ v_{12}$).

iii) Pour tout r on a :

$$\alpha_r = \frac{1}{\sqrt{\lambda_r}} . dV_{11}^{-1} \circ v_{12}(\beta_r)$$

$$\beta_r = \frac{1}{\sqrt{\lambda_r}} . dV_{22}^{-1} \circ v_{21}(\alpha_r) .$$

Ce dernier point se démontre facilement de façon directe.

83. Résolution.

Les familles $(f_{J_1}(\alpha_r))$ et $f_{J_2}(\beta_r))$ satisfont aux conditions du §.79, en tant que système Φ-orthonormé de vecteurs propres de $p_{L_1} \circ p_{L_2}$ et $p_{L_2} \circ p_{L_1}$. Par suite les variables vectorielles $F^1(\alpha_r)$ et $F^2(\beta_r)$ satisfont aux conditions posées au §.77. Pour tout r les couples $(f_{J_1}(\alpha_r), f_{J_2}(\beta_r))$ sont donc les couples (f_r^1, f_r^2) de vecteurs canoniques et les couples $(F^1(\alpha_r), F^2(\beta_r))$ sont les couples (F_r^1, F_r^2) de variables canoniques. En résumé :

Théorème 9.-

i) Etant donné un système (α_r) (resp. (β_r)) V_{11}-orthonormé (resp. V_{22}-orthonormé) de vecteurs propres de l'opérateur hermitien $dV_{11}^{-1} \circ V_{12} \circ dV_{22}^{-1} \circ V_{21}$ de E_1^* (resp. $dV_{22}^{-1} \circ V_{21} \circ dV_{11}^{-1} \circ V_{12}$ de E_2^*) associé à la famille (λ_r) de valeurs propres non nulles, les couples $(f_r^1, f_r^2) = (f_{J_1}(\alpha_r), f_{J_2}(\beta_r))$ sont les couples de vecteurs canoniques.

ii) Si :

$$F_r^1 = \sum_{j\, J_1} \pi_j^{1*}(\alpha_r).V_j^1$$

$$F_r^2 = \sum_{\ell\, J_2} \pi_\ell^{2*}(\beta_r).V_\ell^2$$

alors les couples (F_r^1, F_r^2) sont les couples de variables canoniques.

Nous terminons là ce dernier chapitre. Avec les notations adoptées ici, tous les autres résultats de l'analyse canonique classique se transposent dans les mêmes termes à l'analyse canonique de deux familles de variables vectorielles. Si m = 1, on retrouve strictement l'analyse classique.

Bibliographie

[1] BERTIER P. et BOUROCHE J.M. , *Analyse des données multidimensionnelles*, P.U.F., 1975.

[2] BOURBAKI N., *Algèbre, Chapitre 9*, Hermann.

[3] BRAUN J.M., *Séries chronologiques multiples : recherche d'indicateurs*, R.S.A. 1973, vol. 21, n°1, p. 81-106.

[4] CAILLEZ F. et PAGES J.P., *Introduction à l'analyse des données*, SMASH 1976.

[5] CARROLL J.D. and CHANG J.J., *Analysis of individual différences in multidimensional scaling via an N-way généralisation of "Eckart-Young" decomposition*, Psychometrika, 1970, vol.35, n°3, p.283-319.

[6] CARROLL J.D., *Individual differences and multidimensional scaling*, in R.N. SHEPPARD, A.K. ROMNEY and NERLOVE (Ed.), *Multidimensional scaling : theory and applications in the behavioral sciences*, vol. 1, New-York, Seminar Press, 1972, p. 105-155.

[7] ECKART C. and YOUNG G., *The approximation of one matrix by another of lower rank*, Psychometrika, 1936, vol. 1, p. 211-218.

[8] ESCOUFIER Y., *Le traitement des variables vectorielles*, Biometrics, vol. 29, n°4, 1973, p. 751-760.

[9] ESCOUFIER Y., *Le traitement des familles de variables vectorielles*, Bulletin de l'Institut International de Statistiques, Actes de la 39-ème session, Vienne, 1973.

[10] HORST P., *Relations among m sets of measures*, Psychometrika, vol. 16, 1961, p. 129-149.

[11] HOTELLING H.,*Analysis of a complexe of statistical variables into principal components*, J. Educ. Psychol.,1933, vol. 24 p. 417-441, p. 498-520.

[12] KETTENRING J.R., *Canonical analysis of several sets of variables*, Biometrika, vol. 58, n°3, 1971.

[13] LEVIN J., *Three-mode factor analysis*, Psychol. Bull. 1965, 64, p. 442-452.

[14] L'HERMIER DES PLANTES H., *Structuration des tableaux à trois indices de la statistique*, Thèse 3-ème cycle, Montpellier, 1976.

[15] MAC DONALD R.P., *Three common factor models for groups of variables*, Psychometrika, vol. 35, n°1, 1970, p. 111-128.

[16] PAGES J.P., *A propos des opérateurs d'Escoufier*, Séminaire IRIA sur la classification automatique et la perception par ordinateur, 1974.

[17] TUCKER L.R., *Some mathematical notes on three-mode factor analysis*, Psychometrika, vol. 31, n°3, 1966, p. 279-311.

[18] TUCKER L.R., *Relation between multidimensional scaling and three-mode factor analysis*, Psychometrika, vol. 37, n°1, 1972, p. 3-27.

[19] WOLD H., *Estimation of principal components and related models by iterative least squares*, in KRISHNAIAH P.R. (Ed.), *Multivariate analysis*, New-York, Academic Press, 1966, p. 391-420.