BIOMETRICS 29, 751–760
December 1973

LE TRAITEMENT DES VARIABLES VECTORIELLES

YVES ESCOUFIER

Département d'Informatique, Institut Universitaire de Technologie, Montpellier, France

RESUME

Les méthodes d'analyse multivariable reposent sur la structure d'espace de Hilbert de l'ensemble des variables aléatoires de variance finie et sont donc généralisables à tout ensemble d'objets jouissant de cette structure. En particulier, parce que certaines applications conduisent à des familles de variables vectorielles, on est amené à chercher l'espace de Hilbert qui permettra la généralisation dans ce cas. Le produit scalaire utilisé possède des propriétés pratiques et théoriques qui sont discutées. Une application est présentée afin de faciliter la compréhension.

1. LA NATURE GEOMETRIQUE DES METHODES D'ANALYSE MULTIVARIABLE.

Soit $\mathbf{X}$ un vecteur aléatoire réel qui sera supposé centré. Plusieurs techniques de la statistique multivariable permettent d'étudier les liaisons existant entre composantes ou groupes de composantes de $\mathbf{X}$.

L'analyse en composante principale permet de substituer au vecteur $\mathbf{X} = (X_1, \cdots, X_p)$, le vecteur $\mathbf{Y} = (Y_1, \cdots, Y_p)$ choisi de telle sorte que pour tout i et j, $i \neq j$, $E(Y_iY_j) = 0$ et $E(Y_i^2) = \lambda_i \quad (i = 1, \cdots, p; j = 1, \cdots, p)$ òu λ_i est la $i^{\text{ième}}$ valeur propre de la matrice des variances et covariances de $\mathbf{X}$. Ces valeurs propres sont en général classées par ordre décroissant. Un des buts de l'étude est d'expliciter les éléments a_{ik} tels que:

$$\text{pout tout } i,\ X_i = \sum_{k=1}^{p} a_{ik} Y_k \qquad i = 1, \cdots, p.$$

On en déduit facilement que:

$$E(X_iX_j) = \sum_{k=1}^{p} a_{ik}a_{jk}\lambda_k$$

et

$$E(X_i^2) = \sum_{k=1}^{p} a_{ik}^2\lambda_k$$

si bien que l'on peut "expliquer" la variance des différentes composantes de $\mathbf{X}$ de même que les covariances.

Si $\mathbf{X}$ peut être partitionné en deux sous vecteurs $\mathbf{X}^1$ et $\mathbf{X}^2$, *l'analyse canonique* cherchera les combinaisons linéaires des éléments de $\mathbf{X}^1$ et de $\mathbf{X}^2$

qui ont les plus grandes covariances et fournira la valeur de ces covariances. Elle permet ainsi l'appréciation globale des liaisons existant entre $\mathbf{X}^1$ et $\mathbf{X}^2$. La présentation conjointe des composantes de $\mathbf{X}^1$ et des projections des composantes de $\mathbf{X}^2$ dans le sous espace engendré par $\mathbf{X}^1$ et rapporté aux variables canoniques permet de plus une appréciation des liaisons à l'intérieur de $\mathbf{X}^1$, à l'intérieur de $\mathbf{X}^2$ et entre éléments de $\mathbf{X}^1$ et $\mathbf{X}^2$.

Lorsque l'un des sous vecteurs, $\mathbf{X}^1$ par exemple, n'a qu'une seule composante, l'analyse canonique se réduit à une *analyse de la régression* de cette variable $\mathbf{X}^1$ sur le vecteur $\mathbf{X}^2$.

Toutes ces techniques reposent sur la structure d'espace de Hilbert de l'ensemble des combinaisons linéaires des X_i pour le produit scalaire défini par la covariance. Leur mise en oeuvre implique des projections orthogonales sur des espaces convenablement choisis ainsi que la construction de bases judicieuses. Les hypothèses gaussiennes qui sont faites en général pour l'emploi de ces techniques sont nécessaires pour l'inférence statisque mais ne sont d'aucune nécessité dans la construction des solutions. Il faut noter toutefois qu'elles ajoutent à l'intérêt de ces méthodes lorsqu'elles sont réalisées: normalité des composantes principales; équivalence des notions de non-corrélation et d'indépendance . . . Si ce superflu est agréable, il n'est pas fondamental. Les seuls éléments fondamentaux dans les démonstrations sont liés à la nature Hilbertienne de l'espace des combinaisons linéaires des éléments de $\mathbf{X}$. Il en découle alors inéluctablement que ces méthodes seront utilisables pour toute famille d'êtres mathématiques X_i pour laquelle il sera possible de parler d'espace de Hilbert pour l'ensemble des combinaisons linéaires des X_i .

2. LES FAMILLES DE VARIABLES VECTORIELLES

Certains ensembles de données, par leur nature même, ne sont pas aptes à subir les traitements dont nous avons parlé précédemment. C'est le cas, en particulier, des familles de variables vectorielles. Nous appelons de ce nom un ensemble $(\mathbf{X}^1, \mathbf{X}^2, \cdots, \mathbf{X}^n)$ dont les éléments $\mathbf{X}^i$ sont des vecteurs aléatoires réels. Envisageons des exemples de données de ce type.

1er exemple: Les différents examens que subit un patient au moment de son admission dans un service médical constituent une réalisation du vecteur aléatoire défini par le protocole des examens qui se font dans le dit service. Il n'est pas rare qu'au cours de son hospitalisation, le patient passe par plusieurs services. Chaque service aura alors son protocole particulier si bien que chaque patient sera caractérisé par une réalisation de la famille des vecteurs aléatoires liés au protocole des examens des services où il est passé. Il sera intéressant d'étudier en quoi ces protocoles se différencient, c'est-à-dire en quoi les vecteurs qui leur sont liés sont proches ou éloignés au sens d'une distance adéquate.

2ième exemple: Des objets sont successivement présentés à différents individus. Chaque individu utilise son propre ensemble de variables pour les décrire si bien que, à chaque objet, est attaché une réalisation de la famille des vecteurs aléatoires liés à chacun des individus. Etudier les relations entre les vecteurs aléatoires c'est étudier dans quelle mesure les individus fournissent des descriptions comparables ou non.

3ième exemple: Des êtres vivants sont placés successivement dans différents milieux. Pour chaque milieu des mesures sont faites qui décrivent les réactions des êtres vivants. Ainsi, à chaque être vivant est attaché une réalisation d'une famille de vecteurs aléatoires, chaque vecteur étant lié à un milieu. L'étude des liaisons entre les différents vecteurs renseigne sur les conséquences comparées de l'immersion dans tel ou tel milieu.

3. COVARIANCE ET CORRELATION DES VECTEURS ALEATOIRES

Le paragraphe 5 de cet article donnera les justifications des définitions qui suivent. Nous nous contenterons ici d'expliciter quelques conséquences immédiates susceptibles de montrer l'intérêt de la proposition qui est faite. Soient $\mathbf{X}^1$ et $\mathbf{X}^2$ deux vecteurs aléatoires; $\mathbf{\Sigma}$ sera la matrice des covariances de $\mathbf{X} = (\mathbf{X}^1; \mathbf{X}^2)$. On utilisera la factorisation habituelle:

$$\mathbf{\Sigma} = \left[\begin{array}{c|c} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \hline \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{array}\right].$$

Définition 1

Tr $(\mathbf{\Sigma}_{21}\,\mathbf{\Sigma}_{12})$ est appelée la "*covariance*" des vecteurs $\mathbf{X}^1$ et $\mathbf{X}^2$ et est notée COVV $(\mathbf{X}^1, \mathbf{X}^2)$.
Tr $(\mathbf{\Sigma}_{11}{}^2)$ est appelée la "*variance*" du vecteur $\mathbf{X}^1$ et est notée VAV $(\mathbf{X}^1)$.
COVV $(\mathbf{X}^1, \mathbf{X}^2)/[\text{VAV}(\mathbf{X}^1)\ \text{VAV}(\mathbf{X}^2)]^{1/2}$ est appelée la "*corrélation*" des vecteurs $\mathbf{X}^1$ et $\mathbf{X}^2$ et est notée RV $(\mathbf{X}^1, \mathbf{X}^2)$.

1ière Conséquence:

Si $\mathbf{X}^3$ est un vecteur aléatoire, on a:

$$\text{COVV}(\mathbf{X}, \mathbf{X}^3) = \text{COVV}(\mathbf{X}^1, \mathbf{X}^3) + \text{COVV}(\mathbf{X}^2, \mathbf{X}^3).$$

On en déduit: $\text{VAV}(\mathbf{X}) = \text{VAV}(\mathbf{X}^1) + \text{VAV}(\mathbf{X}^2) + 2\,\text{COVV}(\mathbf{X}^1, \mathbf{X}^2)$.

2ième Conséquence:

Si $\mathbf{X}^2$ est un vecteur à $k2$ composantes; $\mathbf{X}^1$ un vecteur à $k1 \leq k2$ composantes et A une matrice, $k1 \times k2$, telle que $AA' = kI$, on a:

$$\text{COVV}(\mathbf{X}^2, \mathbf{X}^3) = k\,\text{COVV}(\mathbf{X}^1, \mathbf{X}^3)$$

On en déduit: $\text{VAV}(\mathbf{X}^2) = k^2\,\text{VAV}(\mathbf{X}^1)$

et $\text{RV}(\mathbf{X}^2, \mathbf{X}^3) = \text{RV}(\mathbf{X}^1, \mathbf{X}^3)$.

Ces deux premières conséquences sont intéressantes par leur analogie avec les propriétés de la covariance, de la variance et du coefficient de corrélation usuels. Il est important de noter toutefois que si $\mathbf{X}^1 = (X_1^{\ 1})$ et $\mathbf{X}^2 = (X_1^{\ 2})$, COVV $(\mathbf{X}^1, \mathbf{X}^2)$ est égal au carré de la covariance de $X_1^{\ 1}$ et $X_1^{\ 2}$.

3ième Conséquence:

Supposons $\mathbf{X}^1 = (X_1^{\ 1}, X_2^{\ 1}, \cdots, X_{k1}^{\ 1})$ et $\mathbf{X}^2 = (X_1^{\ 2}, \cdots, X_{k2}^{\ 2})$.
Soit D une matrice $k1 \times k2$ telle que:

$$D = \begin{bmatrix} a_1 & & & 0 & 0 & 0 \\ & a_2 & & & & \\ & & \ddots & & & \\ 0 & & & a_{k1} & 0 & 0 \end{bmatrix} \text{ où } |a_i| \leq 1 \text{ pour tout } i = 1, \cdots, k1.$$

$$\text{Si } \mathbf{\Sigma} = \left[\begin{array}{c|c} I_{k1k1} & D \\ \hline D' & I_{k2k2} \end{array}\right] \text{ alors } RV(\mathbf{X}^1, \mathbf{X}^2) = \sum_{i=1}^{k1} a_i^{\ 2}/(k1k2)^{1/2}.$$

Pour une telle matrice $\mathbf{\Sigma}$ le carré du coefficient de corrélation vectorielle (Hotelling [1936]) serait égal à $R^2(\mathbf{X}^1, \mathbf{X}^2) = \prod_{i=1}^{k1} a_i^{\ 2}$. Il suffit alors que l'un des a_i soit nul pour que $R^2(\mathbf{X}^1, \mathbf{X}^2)$ le soit aussi. RV $(\mathbf{X}^1, \mathbf{X}^2)$ apparait donc comme une mesure de similarité entre $\mathbf{X}^1$ et $\mathbf{X}^2$ plus raisonnable que $R^2(\mathbf{X}^1, \mathbf{X}^2)$. Cette considération est appuyée par le cas particulier $k1 = 1$ qui donne RV $(\mathbf{X}^1, \mathbf{X}^2) = a_1^{\ 2}/(k2)^{1/2}$ tandis que $R^2(\mathbf{X}^1, \mathbf{X}^2) = a_1^{\ 2}$. Or quand $k2$ croît, l'espace engendré par $\mathbf{X}^2$ est de plus en plus différent de l'espace engendré par $\mathbf{X}^1 = (X_1^{\ 1})$ et il est donc normal que la similarité entre les deux vecteurs diminue.

4ième Conséquence:

Soit H_{rn} une matrice $r \times n$; supposons que:

$$\mathbf{\Sigma} = \left[\begin{array}{c|c} H_{k1k1} & aH_{k1k2} \\ \hline aH_{k2k1} & H_{k2k2} \end{array}\right] \text{ où } \quad H_{rn} \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{bmatrix}.$$

Alors : RV $(\mathbf{X}^1, \mathbf{X}^2) = a^2$.

Dans ce cas où $R^2(\mathbf{X}^1, \mathbf{X}^2)$ n'est pas défini puisque $|H_{k1k1}| = |H_{k2k2}| = 0$, RV $(\mathbf{X}^1, \mathbf{X}^2)$ reste égal à a^2 quelles que soient les valeurs $k1$ et $k2$. Ceci permet de dire que RV supporte bien les cas dégénérés où une même variable est répétée plusieurs fois.

4. EXEMPLE

La Table 1 fournit la matrice de corrélations de neuf variables. Ces données extraites de Thurstone et Thurstone [1941] ont été souvent utilisées en particulier par Kettenring [1971]. Posons $\mathbf{X}^i = (S_i, N_i, R_i)$ pour $i = 1, 2, 3$. Les variables S_i mesurent les possibilités spatiales d'un individu; les variables N_i, les possibilités numériques; les R_i, les possibilités de raisonnement mental. On peut se poser deux questions sur ces données. D'abord

TABLE 1

MATRICES DES CORRELATIONS DES 9 VARIABLES

	N_1	R_1	S_2	N_2	R_2	S_3	N_3	R_3
S_1	0.249	0.271	0.636	0.183	0.185	0.626	0.369	0.279
N_1		0.399	0.138	0.654	0.262	0.190	0.527	0.356
R_1			0.180	0.407	0.613	0.225	0.471	0.610
S_2				0.091	0.147	0.709	0.254	0.191
N_2					0.296	0.103	0.541	0.394
R_2						0.179	0.437	0.496
S_3							0.291	0.245
N_3								0.429

peut-on considérer que $\mathbf{X}^1$, $\mathbf{X}^2$, $\mathbf{X}^3$ apportent les mêmes informations? Ensuite peut-on considérer que les vecteurs $\mathbf{S} = (S_1, S_2, S_3)$, $\mathbf{N} = (N_1, N_2, N_3)$ et $\mathbf{R} = (R_1, R_2, R_3)$ apportent les mêmes informations?

La Table 2a fournit la matrice des RV pour les vecteurs $\mathbf{X}^i$ tandis que la Table 2b fournit les valeurs propres et les vecteurs propres de cette matrice.

TABLE 2a

MATRICE RV POUR $\mathbf{X} = (\mathbf{X}^1, \mathbf{X}^2, \mathbf{X}^3)$

	$\mathbf{X}^2$	$\mathbf{X}^3$
$\mathbf{X}^1$	0.4582	0.4667
$\mathbf{X}^2$		0.4452

TABLE 2b

VALEURS PROPRES ET VECTEURS PROPRES DE RV POUR $\mathbf{X} = (\mathbf{X}^1, \mathbf{X}^2, \mathbf{X}^3)$

	Y_1	Y_2	Y_3
Valeurs propres	1.9135	0.5558	0.5307
$\mathbf{X}^1$	0.5822	0.1598	−0.7972
$\mathbf{X}^2$	0.5731	−0.7761	0.2630
$\mathbf{X}^3$	0.5767	0.6100	0.5434

Une interprétation tout à fait analogue à celle que l'on ferait dans une analyse classique en composantes principales montre que les trois vecteurs sont à peu près équidistants l'un de l'autre. Il est intéressant de noter que la distance entre les points représentants les vecteurs $\mathbf{X}^1$ et $\mathbf{X}^2$ par exemple aura un carré égal à $2(1 - \mathrm{RV}\ (\mathbf{X}^1, \mathbf{X}^2))$ si on a pris soin de donner aux vecteurs propres une norme égale à la valeur propre qui leur correspond (Gower [1966]). La difficulté est bien sûr d'apprécier la valeur de cette distance mutuelle. L'expérience que nous avons de tels traitements nous ferait dire que sans être très importante cette distance n'est pas négligeable et qu'il serait dangereux de prendre l'un des vecteurs pour l'autre.

Les Tables 3a et 3b fournissent les renseignements nécessaires à la comparaison des vecteurs **S**, **N**, **R**. On notera d'abord que ces vecteurs sont beaucoup moins "corrélés" entre eux que ne l'étaient les $\mathbf{X}^i$. C'est dire par exemple que **N** est moins semblable à **R** que $\mathbf{X}^1$ ne l'était de $\mathbf{X}^2$. L'analyse en composantes principales est intéressante qui met en évidence pour les deux premières composantes une opposition de **S** à **N** et **R**, ces deux derniers vecteurs s'opposant sur la troisième composante.

5. JUSTIFICATIONS THEORIQUES

Tous les vecteurs aléatoires $\mathbf{X} = (X_1, X_2, \cdots, X_k)$ considérés sont supposés centrés et toutes leurs composantes sont des variables aléatoires de variance finie sur le même espace de probabilité (Ω, α, P). On notera $L_2\ (\Omega, \alpha, P)$ ou plus simplement L_2, l'ensemble des variables aléatoires centrées de variance finie sur (Ω, α, P).

La méthode proposée consiste à substituer à $\mathbf{X}$ un opérateur $U_{\mathbf{X}}$ de L_2 en lui-même qui le caractérise. Il est alors facile de montrer que l'ensemble des opérateurs $U_{\mathbf{X}}$ est un espace de Hilbert pour le produit scalaire défini

TABLE 3a

MATRICE RV POUR $\mathbf{X} = (\mathbf{S}, \mathbf{N}, \mathbf{R})$

	N	R
S	0.0860	0.0792
N		0.2724

TABLE 3b

VALEURS PROPRES ET VECTEURS PROPRES DE RV POUR $\mathbf{X} = (\mathbf{S}, \mathbf{N}, \mathbf{R})$

	Y_1	Y_2	Y_3
Valeurs propres	1.3157	0.9568	0.7275
S	0.3472	0.9376	0.0194
N	0.6651	−0.2316	−0.7099
R	0.6611	−0.2593	0.7040

par COVV ce qui justifie l'utilisation de cette quantité dans les paragraphes précédents et nous place dans le contexte mathématique adéquat pour la généralisation des méthodes d'analyse multivariable aux familles de variables vectorielles.

Définition 2: On appelle opérateur associé à $\mathbf{X}$, l'opérateur $U_{\mathbf{X}}$ défini sur L_2 par:

$$\text{pour tout } Y \varepsilon L_2 , \qquad U_{\mathbf{X}}(Y) = \sum_{r=1}^{k} [E(X_r Y)] X_r . \tag{1}$$

La définition fait de $U_{\mathbf{X}}$ un opérateur de L_2 dans lui-même. Appelons $\mathbf{\Sigma}$ la matrice des covariances de $\mathbf{X} = (X_1 , X_2 , \cdots , X_k)$.

Proposition 1: Pour tout vecteur de R^k, $\mathbf{Z} = (z_1 , z_2 , \cdots , z_k)$ tel que $\mathbf{\Sigma}\ \mathbf{Z}' = \lambda \mathbf{Z}'$, la variable aléatoire $Y = \mathbf{XZ}'$ est telle que $U_{\mathbf{X}}(Y) = \lambda Y$. Inversement, toute variable aléatoire Y telle que $U_{\mathbf{X}}(Y) = \lambda Y$ est de la forme $Y = \mathbf{XZ}'$ avec $\mathbf{\Sigma}\ \mathbf{Z}' = \lambda \mathbf{Z}'$.

$$\text{De } Y = \mathbf{XZ}' = \sum_{r=1}^{k} z_r X_r , \quad \text{on tire:}$$

$$U_{\mathbf{X}}(Y) = \sum_{r=1}^{k} E(X_r Y) X_r = \sum_{r=1}^{k} \sum_{h=1}^{k} (E(X_r X_h) z_h) X_r$$

$$= \sum_{r=1}^{k} \lambda z_r X_r = \lambda Y$$

et la première partie de la proposition est ainsi démontrée. Pour démontrer la seconde partie remarquons que si $U_{\mathbf{X}}(Y) = \lambda Y$ alors par définition de $U_{\mathbf{X}}$, $Y = \mathbf{XZ}'$ avec $\mathbf{Z} = (E(X_1 Y),\ E(X_2 Y),\ \cdots ,\ E(X_k Y))/\lambda$. Or $\sum_{r=1}^{k} E(X_r Y) X_r = \lambda Y$ implique $\sum_{r=1}^{k} E(X_r Y) E(X_r X_h) = \lambda E(Y X_h)$ ce qui établit que $(E(X_1 Y),\ E(X_2 Y),\ \cdots ,\ E(X_k Y))$, ou tout multiple de ce vecteur, est vecteur propre de $\mathbf{\Sigma}$.

La proposition montre donc que $U_{\mathbf{X}}$ est caractéristique de $\mathbf{X}$ au sens où les valeurs propres et les vecteurs propres de $\mathbf{\Sigma}$ le sont. C'est une caractérisation très forte puisqu'elle définit non seulement les directions des composantes principales de $\mathbf{X}$ mais aussi les variances de ces composantes.

Proposition 2: L'opérateur $U_{\mathbf{X}}$ appartient à la classe des opérateurs de Hilbert–Schmidt (i.e. la somme des carrés des valeurs propres est finie) et sur cette classe, on peut définir un produit scalaire qui la munit d'une structure d'espace de Hilbert.

La première partie de la propriété est évidente puisque $U_{\mathbf{X}}$ n'a qu'un nombre fini de valeurs propres finies. La seconde partie est un résultat connu d'analyse fonctionnelle (Kato [1966] p. 262). Le produit scalaire de deux opérateurs B_1 et B_2 de Hilbert–Schmidt sur L_2 est défini à partir d'une famille orthonormale $(\phi_k)_{k \varepsilon K}$ complète dans L_2 mais est indépendant du choix de la famille orthornormale. Il est donné par:

$$\langle B_1\,,B_2\rangle = \sum_{k\varepsilon K} E(B_1(\phi_k)B_2(\phi_k)).$$

En particulier, si l'on prend pour famille orthonormale les vecteurs propres de B_1 (disons ξ_k, $k \varepsilon K$) associés aux valeurs propres non nulles μ_k comptées avec leur multiplicité et complétées par une famille orthonormale complète du noyau de B_1, on a:

$$\langle B_1\,,B_2\rangle = \sum_{k\varepsilon K} \mu_k E(\xi_k B_2(\xi_k)).$$

En écrivant alors pour les vecteurs propres de B_1 le développement convergent dans L_2

$$\xi_k = \sum_{r\varepsilon L} E(\xi_k\psi_r)\psi_r + h_k$$

où h_k appartient au noyau de B_2 et où les $\psi_r(r \varepsilon L)$ sont les vecteurs propres de B_2 associés aux valeurs propres λ_r comptées avec leur multiplicité, on a:

$$\langle B_1\,,B_2\rangle = \sum_{k\varepsilon K}\sum_{r\varepsilon L} \mu_k\lambda_r[E(\xi_k\psi_r)]^2. \tag{2}$$

Remarque 1: On déduit de ce qui précède la distance entre les opérateurs B_1 et B_2 :

$$\begin{aligned} d^2(B_1\,,B_2) &= \|B_1 - B_2\|^2 = \|B_1\|^2 + \|B_2\|^2 - 2\langle B_1\,,B_2\rangle \\ &= \sum_{k\varepsilon K}\mu_k{}^2 + \sum_{r\varepsilon L}\lambda_r{}^2 - 2\sum_{k\varepsilon K}\sum_{r\varepsilon L}\mu_k\lambda_r E(\xi_k\psi_r)^2. \end{aligned}$$

On peut également définir un "coefficient de corrélation" entre les opérateurs:

$$Q(B_1\,,B_2) = \frac{\langle B_1\,,B_2\rangle}{\|B_1\|\,\|B_2\|}$$

Ce coefficient pourra être interprété comme le produit scalaire des opérateurs

$$\tilde{B}_1 = \frac{B_1}{\|B_1\|} \text{ et } \qquad \tilde{B}_2 = \frac{B_2}{\|B_2\|}.$$

On voit sans difficulté que $Q(B_1\,,B_2) = 0$ est équivalent à $E(\xi_k\psi_r) = 0$ pour tout k et tout r. C'est dire que les espaces propres de B_1 et B_2 sont orthogonaux. De la même manière $Q(B_1\,,B_2) = 1$ est équivalent à $d^2\,(\tilde{B}_1\,,\tilde{B}_2) = 0$ qui signifie que les opérateurs $\tilde{B}_1$ et $\tilde{B}_2$ ont les mêmes espaces propres, les mêmes noyaux et des valeurs propres égales. Il en découle que B_1 et B_2 ont les mêmes espaces propres, les mêmes noyaux et des valeurs propres proportionnelles. $\left(\text{pour tout } k : \frac{\mu_k}{\lambda_k} = \frac{\|B_1\|}{\|B_2\|}\right)$.

Remarque 2: Dans la formule (2) introduisons

$$\xi_k' = \sqrt{\mu_k}\,\xi_k \text{ et } \psi_r' = \sqrt{\lambda_r}\,\psi_r\,.$$

On a

$$\langle B_1\,,B_2\rangle = \sum_{k\varepsilon K}\sum_{r\varepsilon L} [E(\xi_k'\psi_1')]^2.$$

Appliquons ce résultat aux opérateurs $U_{\mathbf{X}1}$ et $U_{\mathbf{X}2}$ respectivement associés aux vecteurs $\mathbf{X}_1 = (X_1^1, X_2^1, \cdots, X_{k1}^1)$ et $\mathbf{X}_2 = (X_1^2, X_2^2, \cdots, X_{k2}^2)$. Notons $\mathbf{Y}_1 = (Y_1^1, Y_2^1, \cdots, Y_{k1}^1)$ les vecteurs propres de $U_{\mathbf{X}1}$ normés de façon à ce que $E((Y_i^1)^2) = \mu_i$ et de même $\mathbf{Y}_2 = (Y_1^2, Y_2^2, \cdots, Y_{k2}^2)$ où $E((Y_j^2)^2) = \lambda_j$. Alors:

$$\langle U_{X1}, U_{X2} \rangle = \sum_{i=1}^{k1} \sum_{j=1}^{k2} [E(Y_i^1 Y_j^2)]^2. \tag{3}$$

Lemme: Soient $\mathbf{X}_1 = (X_1^1, X_2^1, \cdots, X_{k1}^1)$ et $\mathbf{X}_2 = (X_1^2, X_2^2, \cdots, X_{k2}^2)$. Soient également A_1 (resp. A_2) une matrice orthogonale $k1 \times k1$ (resp. $k2 \times k2$). Posons $\mathbf{X}_1\mathbf{A}_1 = \mathbf{T}_1 = (T_1^1, T_2^1, \cdots, T_{k1}^1)$ et $\mathbf{X}_2 A_2 = \mathbf{T}_2 = (T_1^2, T_2^2, \cdots, T_{k2}^2)$. Soient $\mathbf{\Sigma}$ la matrice des covariances de $\mathbf{X} = (\mathbf{X}_1 ; \mathbf{X}_2)$ et $\mathbf{\Gamma}$ celle de $\mathbf{T} = (\mathbf{T}_1 ; \mathbf{T}_2)$

$$\text{Alors } \sum_{i=1}^{k1} \sum_{j=1}^{k2} [E(T_i^1 T_j^2)]^2 = \sum_{i=1}^{k1} \sum_{j=1}^{k2} [E(X_i^1 X_j^2)]^2.$$

$$\text{Posons } \mathbf{\Sigma} = \left[\begin{array}{c|c} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \hline \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{array}\right] \text{ et } \quad \mathbf{\Gamma} = \left[\begin{array}{c|c} \mathbf{\Gamma}_{11} & \mathbf{\Gamma}_{12} \\ \hline \mathbf{\Gamma}_{21} & \mathbf{\Gamma}_{22} \end{array}\right].$$

De $\mathbf{T}_1 = \mathbf{X}_1 A_1$ et $\mathbf{T}_2 = \mathbf{X}_2 A_2$ on tire $\mathbf{\Gamma}_{21} = A_2' \mathbf{\Sigma}_{21} A_1$. Il en découle alors:

$$\begin{aligned} \sum_{i=1}^{k1} \sum_{j=1}^{k2} [E(T_i^1 T_j^2)]^2 &= \mathrm{Tr}\,(\mathbf{\Gamma}_{21} \mathbf{\Gamma}_{12}) \\ &= \mathrm{Tr}\,(A_2' \mathbf{\Sigma}_{21} A_1 A_1' \mathbf{\Sigma}_{12} A_2) \\ &= \mathrm{Tr}\,(A_2' \mathbf{\Sigma}_{21} \mathbf{\Sigma}_{12} A_2) \\ &= \mathrm{Tr}\,(\mathbf{\Sigma}_{12} A_2 A_2' \mathbf{\Sigma}_{21}) \\ &= \mathrm{Tr}\,(\mathbf{\Sigma}_{12} \mathbf{\Sigma}_{21}) = \mathrm{Tr}\,(\mathbf{\Sigma}_{21} \mathbf{\Sigma}_{12}) \\ &= \sum_{i=1}^{k1} \sum_{j=1}^{k2} [E(X_i^1 X_j^2)]^2. \end{aligned}$$

Proposition 3:

$$\langle U_{\mathbf{X}1}, U_{\mathbf{X}2} \rangle = \mathrm{Tr}\,(\mathbf{\Sigma}_{21} \mathbf{\Sigma}_{12}) = \sum_{i=1}^{k1} \sum_{j=1}^{k2} [E(X_i^1 X_j^2)]^2$$

La matrice qui permet de passer de $\mathbf{X}_1$ aux vecteurs propres $\mathbf{Y}_1 = (Y_1^1, Y_2^1, \cdots, Y_{k1}^1)$ de $U_{\mathbf{X}1}$ est, d'après la proposition 1, la matrice des vecteurs propres de $\mathbf{\Sigma}_{11}$; on sait que cette matrice est orthogonale. Il en est de même de la matrice qui permet de passer de $\mathbf{X}_1$ aux vecteurs propres $\mathbf{Y}_2 = (Y_1^1, Y_2^1, \cdots, Y_{k2}^2)$ de $U_{\mathbf{X}2}$. Alors, d'après le lemme précédent

$$\sum_{i=1}^{k1} \sum_{j=1}^{k2} [E(X_i^1 X_j^2)]^2 = \sum_{i=1}^{k1} \sum_{j=1}^{k2} [E(Y_i^1 Y_j^2)]^2$$

et le résultat annoncé découle de la formule (3).

Remarque 3: On voit que les quantités COVV et RV utilisées dans les paragraphes précédents sont définies par

$$\text{COVV}\ (\mathbf{X}_1\ , \mathbf{X}_2) = \langle U_{\mathbf{X}1}\ , U_{\mathbf{X}2} \rangle$$

et

$$\text{RV}\ (\mathbf{X}_1\ , \mathbf{X}_2) = Q(U_{\mathbf{X}1}\ , U_{\mathbf{X}2}).$$

Le lemme permet d'affirmer que COVV $(X_1\ , X_2)$ est invariant dans des transformations orthogonales affectant $\mathbf{X}_1$ et $\mathbf{X}_2$. On aimerait pouvoir énoncer une propriété semblable pour les matrices non singulières. On doit renoncer à cet espoir. Il ne faut pas s'étonner de cette impossibilité: elle est liée au fait bien connu (Anderson [1958] p. 279) de la non invariance des composantes principales d'un vecteur $\mathbf{X}$ dans une transformation $\mathbf{T} = \mathbf{X}\ A$ où $AA' \neq kI$.

REMERCIEMENTS

Je tiens à remercier les professeurs P. Robillard et R. Cléroux pour l'aide constante qu'ils m'ont apportée dans la rédaction de ce texte pendant que je séjournais au département Informatique de l'Université de Montréal. J'adresse également mes remerciements les plus chaleureux à J. Mosiman du National Institute of Health: mon travail doit beaucoup à ses commentaires précis et à sa bienveillante compréhension.

SUMMARY

Multivariate statistical techniques are based upon the Hilbert space structure of the set of random variables with finite variance, and are therefore generalizable to any set having such a structure. In particular, since certain applications lead to families of vector random variables, we are led to the problem of finding the Hilbert space which permits the above generalization to families of vector variables. The inner product which is used possesses theoretical and practical properties which are discussed. An illustrative example is given.

REFERENCES

Anderson, T. W. [1958]. *An Introduction to Multivariate Statistical Analysis.* Wiley, New York.

Gower, J. C. [1966]. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika 53*, 325–38.

Hotelling, H. [1936]. Relations between two sets of variates. *Biometrika 28*, 321–77.

Kato, T. [1966]. *Perturbation Theory for Linear Operators.* Springer–Verlag.

Kettenring, J. R. [1971]. Canonical analysis of several sets of variates. *Biometrika 58*, 433–51.

Thurstone, L. L. and Thurstone, T. G. [1941]. Factorial studies of intelligence. Psychometric Monograph 2.

Received March 1972, *Revised January* 1973

Key Words: Famille de vecteurs aléatoires; Coefficient de corrélation entre vecteurs; Opérateur associé à un vecteur.