DECOMPOSITION CANONIQUE D'UN OPERATEUR , ANALYSE EN COMPOSANTES PRINCIPALES ET INDICES DES VARIABLES ALEATOIRES ETAGEES

Jérôme MANUCEAU*

***Unité de Formation et de Recherche : Mathématiques , Informatique , Mécanique**

Université de Provence

Place Victor Hugo

13003 - MARSEILLE CEDEX 3

RESUME

Soit E et F deux espaces euclidiens et u un élément de $\mathcal{L}$ (F,E) . On montre par une méthode de moindres carrés que $u = \sum_i \sqrt{\lambda_i}\,\varepsilon_i \otimes \psi_i$, où $(\varepsilon_i)_i$ (resp.$(\psi_i)_i$) est une base orthonormale de val u (resp.$(\ker u)^{\perp}$) diagonalisant uu* (resp.u*u) , $(\lambda_i)_i$ les valeurs propres non nulles de uu* et u*u et $\forall x \in F$, $(\varepsilon_i \otimes \psi_i)(x) = \langle \psi_i \mid x \rangle . \varepsilon_i$. C'est la <u>décomposition canonique</u> de u.

En Analyse des Données , la représentation d'un système statistique comprenant n (= dim F) individus et ℓ (= dim E) variables aléatoires , est équivalente à la donnée d'un élément u de $\mathcal{L}$ (F,E). Nous établissons un critère (les $(\lambda_i)_i$ doivent être distincts deux à deux) pour qu'il y ait unicité de la représentation . L'Analyse en Composantes Principales revient à faire la décomposition canonique de u . En effet , les $(\varepsilon_i)_i$ sont les <u>vecteurs principaux</u> et les $(\psi_i)_i$ les <u>composantes principales</u> . On obtient ainsi un formalisme qui simplifie les calculs et généralise les résultats , en particulier en Analyse Factorielle Discriminante et en Analyse Factorielle des Correspondances.

L'Analyse Canonique de deux sous espaces vectoriels d'un espace euclidien , permet de définir les <u>indices de proximité</u> et d'<u>inclusion</u> entre deux variables aléatoires étagées . Ces indices jouent un rôle capital pour choisir, dans un système statistique, le plus petit nombre possible de variables aléatoires, contenant l'essentiel de l'information statistique .

1.DECOMPOSITION CANONIQUE

1.1. <u>Notations et position du problème</u>

Soient E et F deux espaces euclidiens de dimension ℓ et n respectivement , u une application linéaire de F dans E , u* sa transposée . Posons U = u*u et V = uu* .

Nous allons établir par une méthode de moindres carrés , la décomposition spectrale de U et de V. Nous chercherons ensuite une écriture de u qui " ressemble" à une décomposition spectrale , que nous appellerons <u>décomposition canonique</u> .

Soit $f_1, f_2, ..., f_n$ une base orthonormale de F et notons $u_i = u(f_i)$ .On appelle <u>indice d'inertie de</u> $(u_i)_i$ <u>par rapport à 0</u> le nombre $I_0 = \sum_{i=1}^{n} \|u_i\|^2$. Si W est un sous-espace vectoriel de E et si P_w est le projecteur orthogonal sur w , alors , en posant

$$I_{o,w} = \sum_{i=1}^{n} \| P_w u_i \|^2 \quad , \text{ on a}$$

$$I_o = I_{o,w} + I_{o,w^\perp} .$$

Il est évident que $I_o = \mathrm{Tr}\,(U) = \mathrm{Tr}\,(V)$ puisque $I_o = \sum_{i=1}^{n} < u(f_i) \mid u(f_i) > \; = \; \sum_{i=1}^{n} < f_i \mid U f_i >$.

Ce nombre ne dépend donc pas de la base choisie dans F .

1.2. Décomposition spectrale de V .

1.2.1. Proposition

> $\forall$ x $\in$ E, si [x] est l'espace vectoriel engendré par x , alors ,
>
> $$I_{o,[x]} = < \frac{x}{\|x\|} \mid V \; \frac{x}{\|x\|} > .$$

Démonstration : puisque $P_{[x]}(u_i) = < u_i \mid \frac{x}{\| x \|} > . \frac{x}{\| x \|}$, alors ,

$$I_{o,[x]} = \sum_{i=1}^{n} < u_i \mid \frac{x}{\| x \|} >^2 = \sum_{i=1}^{n} < f_i \mid \frac{u^*(x)}{\| x \|} >^2 = \frac{\| u^*(x) \|^2}{\| x \|^2} \; \square$$

1.2.2. Proposition

> L'application $x \in S(0,1) \rightarrow I_{o,[x]}$ (S(0,1) est la sphère unité),atteint son maximum . En outre si $I_{o,[x_o]}$ est maximum, x_o est un vecteur propre de V correspondant à sa plus grande valeur propre .

Démonstration : La première partie est évidente d'aprés 1.2.1. . D'aprés l'inégalité de Schwartz on voit que :

$$I_{o,[x]} = \frac{< x \mid Vx >}{< x \mid x >} \quad \leq \quad \frac{< x \mid V^2 x >}{< x \mid Vx >} .$$

V étant positif , il en est de même de la forme bilinéaire symétrique < x | Vy > et une nouvelle application de l'inégalité de Schwartz donne

$$\frac{< x \mid V^2 x >}{< x \mid Vx >} \leq \frac{< x \mid V^3 x >}{< x \mid V^2 x >} = I_{o,[Vx]} \, .$$

D'où $\forall\, x \in E$, $I_{o,[x]} \leq I_{o,[Vx]}$.

Si x_o rend maximum $I_{o,[x_o]}$, alors, $I_{o,[x_o]} = I_{o,[Vx_o]}$, ce qui implique

$$< x_o \mid Vx_o >^2 = < x_o \mid x_o > < V x_o \mid Vx_o > \, . \qquad \square$$

1.2.3. Corollaire

V admet une décomposition spectrale , i.e. , $\exists\ \lambda_1,\dots,\lambda_r$ tels que $\lambda_1 > \lambda_2 > \dots > \lambda_r > 0$, $\exists\ E_1,\dots, E_r$ sous espaces vectoriels de E deux à deux orthogonaux tels que

$V = \sum_{i=1}^{r} \lambda_i P_{E_i}$, où P_{E_i} est le projecteur orthogonal sur E_i .

1.2.4. Théorème

$\forall\, k \in \mathbb{N}$, $k < \dim(E)$, $\exists$ W sous espace vectoriel de E , tel que dim W=k et $I_{o,w}$ soit maximum. En outre W admet une base formée de k vecteurs propres de V correspondants aux valeurs propres les plus grandes .

Démonstration : Soit $x_1, x_2, \dots, x_\ell$ une base orthonormale qui diagonalise V telle que $V x_i = \mu_i x_i$ et $\mu_1 \geq \dots \geq \mu_\ell$. Pour $k = 1$ voir 1.2.2. Supposons $k = 2$ et W' un sous espace vectoriel de dimension deux . Evidemment $\exists\ y \in [x_1]^\perp \cap W'$ et $\exists$ x orthogonal à y tels que $W' = [\{x,y\}]$.

Comme $[x_1]^\perp$ est stable par V , x_2 est le vecteur de $[x_1]^\perp$ qui rend maximum $I_{o,[x]}$, $x \in [x_1]^\perp$.

D'où :

$$I_{o,w'} = I_{o,[x]} + I_{o,[y]} \leq I_{o,[x_1]} + I_{o,[x_2]} = I_{o,w} \quad \text{où } W = [\{x_1, x_2\}] \, .$$

On conclut par itération . $\square$

L'espace w de 1.2.4 s'appelle , espace principal de dimension k.

1.3. Décomposition canonique de u .

Il est évident que ker V = ker u* puisque $\| u^*x \| = < V x \mid x >$ et ker U = ker u .

1.3.1. Théorème

En adoptant les notations de 1.2.3. , il existe $F_1, F_2, \ldots, F_r$, sous espaces vectoriels de F deux à deux orthogonaux et il existe $P_1, P_2, \ldots, P_r$ des isométries partielles (d'espaces euclidiens) de F dans E , telles que

$$- P_i \mid F_i = \frac{1}{\sqrt{\lambda_i}} \quad u \mid F_i \ ,$$

$$- P_i \mid_{F_i^{\perp}} = 0 \ ;$$

$$- \text{ et } \quad u = \sum_{i=1}^{r} \sqrt{\lambda_i} \ P_i \ ,$$

écriture unique que nous appellerons décomposition canonique de u.

Démonstration : Notons F_i le sous-espace propre de U à la valeur propre λ_i . On voit que $u^*(E_i) \subset F_i$ et $u(F_i) \subset E_i$. Comme u et u* sont des isomorphismes entre $(\ker u)^{\perp}$ et $(\ker u^*)^{\perp}$, $u^*(E_i) = F_i$ et $u(F_i) = E_i$. On déduit le théorème, des inégalités suivantes :

$$\forall (x,x') \in E_i^2 \ , \quad \forall (y,y') \in F_i^2 \ ,$$

$$< u^*(x) \mid u^*(x') > \ = \ < x \mid V x' > \ = \ \lambda_i < x \mid x' >$$

et

$$< u(y) \mid u(y') > = < y \mid U y' > \ = \ \lambda_i < y \mid y' > \ .$$

□

1.3.2. Corollaire

$\exists \, (\varepsilon_j)_j$ (resp. $(\psi_j)_j$) une base orthogonale de val u (resp . $(\ker u)^{\perp}$)

diagonalisant uu* (resp. u*u) et $\exists \, (\beta_j)_j$, valeurs propres de uu* et u*u - telles que

$$u = \sum_j \sqrt{\beta_j} \ \varepsilon_j \otimes \psi_j \ \text{ où } \ (\varepsilon_j \otimes \psi_j)(x) = < \psi_j \mid x > \varepsilon_j \ , \ \forall x \in F .$$

1.4 . Espaces Principaux

Avec les notations 1.1. , on suppose donné $p = (p_1, \ldots, p_n)$ un n-uplet de nombres strictement positifs vérifiant $\sum_{i=1}^{n} p_i = 1$ et l'opérateur $D_p = \sum_{i=1}^{n} \sqrt{p_i}\, P_{[f_i]}$.

Quelque soit $u : F \to E$, on note $u_p = u.\, D_p$.

Dans tout ce qui va suivre nous remplacerons u par u_p et retrouverons entre autre les résultats classiques des statistiques et de l'Analyse des Données . En particulier , l'indice d'inertie des $(u_p(f_i))_i$ par rapport à 0 ,est $I_0^p = \sum_{i=1}^{n} p_i \| u_i \|^2$

$$\forall\ W \text{ s.e.v.} \subset E,\ I_{0,W}^p = \sum_i p_i \| P_W(u_i) \|^2 \text{ et}$$

$$I_0^p = I_{0,W}^p + I_{0,W^\perp}^p .$$

$I_{0,W^\perp}^p$ s'appellera la distance des $(u_i)_i$ à W et $I_{0,W}^p / I_0^p$ la part d'inertie expliquée par W .

De même , les espaces principaux seront ceux obtenus par la décomposition spectrale de $V_p = u_p u_p^* = u\, D_p^2\, u^*$ (ou la décomposition canonique de u_p) .

Le centre de gravité g_p des $(u_i)_i$ vérifie

$$g_p = \sum_i p_i u_i = u_p(\varphi_p) \text{ où } \varphi_p = \sum_i \sqrt{p_i}\ f_i .$$

Ainsi $g_p = 0$ si et seulement si $u_p = u_p . P_{[\varphi_p]^\perp}$ et la transformation $u_p \to v_p = u_p . P_{[\varphi_p]^\perp}$ consiste à ramener le centre de gravité des $(u_i)_i$ à l'origine .

$\forall\ a \in E$, $\forall\ W$ s.e.v. $\subset E$, posons $I_a^p = \sum_i p_i \| u_i - a \|^2$ l'inertie des $(u_i)_i$ par rapport à a et $I_{a,W}^p = \sum_i p_i \| P_W(u_i - a) \|^2$ l'inertie des $(P_W(u_i))_i$ par rapport à Pw (a) .

$I_{a,W^\perp}^p$ représente la distance des $(u_i)_i$ à W + a , qui est aussi celle des $(u_i - a)_i$ à W .

On voit que cette distance est minimum lorsque $P_{W^\perp}(a) = P_{W^\perp}(g_p)$, autrement dit lorsque W + a contient g_p , puisque $I_{a,W^\perp}^p = I_{g_p,W^\perp}^p + \| P_{W^\perp}(a - g_p) \|^2$.

On déduit aisément de ce qui précède le résultat général suivant :

Théorème

> $\forall\ k \in \{1,2,\ldots,\ell-1\}$, le sous-espace affine de dimension k , le plus proche des $(u_i)_i$ est $W + g_p$ où g_p est le centre de gravité des $(u_i)_i$ et W un espace principal de dimension k, donné par la décomposition canonique de l'opérateur $v_p = u_p \cdot P_{[\varphi_p]^\perp}$.

2 . REPRESENTATIONS D'UN SYSTEME STATISTIQUE (S.S) . ANALYSE EN COMPOSANTES PRINCIPALES (A.C.P)

2.1 Définitions

2.1.1. Système statistique

> Un système statistique est la donnée d'un ensemble probabilisé (Ω , P) de n individus et d'un ensemble de ℓ variables aléatoires réelles (v.a.r).Le couple (n,ℓ) , s'appelle la dimension du S.S.

On se donne (Ω,P) pour toute la suite de cet exposé avec $\Omega = \{\omega_1, \ldots, \omega_n\}$ et $P = \sum_{i=1}^{n} p_i\, \delta_{\omega_i}$. Deux S.S. de même dimension dont les variables aléatoires de l'un sont multiples des variables aléatoires de l'autre , avec le même coefficient de multiplicité , seront considérés comme identiques.

2.1.2. Représentations

> Une représentation d'un S.S. de dimension (n,ℓ) est un triplet $(u,(f_i)_{1\le i\le n}, (e_j)_{1\le j\le \ell})$, où $u \in \mathcal{L}(F,E)$, F et E étant deux espaces euclidiens de dimensions n et ℓ respectivement , $(f_i)_{1\le i\le n}, (e_j)_{1\le j\le \ell}$) des bases orthonormales respectives , tel que
>
> $$u_i^j = \langle e_j \mid u(f_i) \rangle$$
>
> soit la valeur que prend la $j^{e.}$ variable aléatoire sur le $i^{e.}$ individu.

E est l'espace des individus : $\forall\, i \in \{1,\ldots,n\}$, $u_i = u(f_i)$ représente le i^e individu puisque ses composantes sur la base $(e_j)_{1 \leq j \leq \ell}$ sont les valeurs que prennent les variables aléatoires sur le i^e individu . En fait comme nous l'avons vu plus haut nous préférons prendre comme i^e individu

$u_p(f_i) = \sqrt{p_i}\,.u_i$.

F est l'espace des variables aléatoires : $\forall\, j \in \{1,\ldots,\ell\}$, $u_j^* = u^*(e_j)$ représente la j^e variable aléatoire puisque ses composantes sur la base $(f_i)_{1 \leq i \leq n}$ sont les valeurs que prennent la j^e variable aléatoire. Comme ci-dessus nous préférons prendre comme j^e variable aléatoire

$$u_p^*(e_j) = D_p \,.\, u_j^* \quad (\text{i.e. } (u_p^*)_j^i = \sqrt{p_i}\,(u^*)_j^i) .$$

Remarques :

$$\text{1- } P_{[\varphi_p]} u_p^*(e_j) = \left(\sum_{i=1}^{n} p_i (u^*)_j^i\right) \varphi_p = m_j\, \varphi_p \ ,$$

donc la projection de la j^e variable aléatoire sur φ_p donne sa moyenne m_j .

$$\text{2- } \| P_{[\varphi_p]^\perp}\, u_p^*(e_j) \|^2 = \sum_{i=1}^{n} p_i \left((u^*)_j^i - m_j \right)^2 ,$$

donc la norme au carré de la projection de la j^e variable aléatoire sur $[\varphi_p]^\perp$ est égale à sa variance.

3 - $g_p = 0$ si et seulement si toutes les variables aléatoires sont centrées et la transformation $u_p \rightarrow v_p = u_p \,.\, P_{[\varphi_p]^\perp}$ consiste à centrer les variables aléatoires et à ramener le centre de gravité des $(u_i)_i$ à l'origine.

2.2 Représentation d'un S.S.

2.2.1 Définition

Deux ensembles de vecteurs,d'un espace euclidien ayant même nombre d'éléments $(x_i)_{1 \leq i \leq n}$ et $(y_i)_{1 \leq i \leq n}$, seront dit isomorphes si et seulement si, $\exists$ une permutation σ de $\{1, 2, \ldots, n\}$ et un opérateur $T = \lambda\,.\,O$ où λ est un scalaire et O une isométrie tels que $y_{\sigma(i)} = T(x_i)$, $\forall\, i \in \{1, \ldots, n\}$.

T vérifie $TT^* = T^*T = \lambda^2 I$; il conserve les angles et les rapports des longueurs des vecteurs. Pour simplifier les écritures,nous supposons dans la suite $\sigma = I$.

2.2.2 Définition

> Deux représentations de même dimension $(u, (f_i)_{1\leq i\leq n}, (e_j)_{1\leq j\leq \ell})$ et $(v , (g_i)_{1\leq i\leq n} , (h_j)_{1\leq j\leq \ell}$ seront dites isomorphes, si et seulement si, $(u(f_i))_{1\leq i\leq n}$ (resp. $(u^*(e_j))_{1\leq j\leq \ell}$) est isomorphe à $(v(g_i))_{1\leq i\leq n}$ (resp.$(v^*(e_j))_{1\leq j\leq \ell}$) .

Il est clair que deux représentations isomorphes peuvent être considérées, sur le plan algébrique, comme identiques. Par conséquent, l'association S.S.-représentation, ne peut avoir un intérêt, que si elle est bijective à un isomorphisme près. C'est le problème que nous allons examiner dans la suite.

2.2.3 Proposition

> Si $(u,(f_i)_{1 \leq i \leq n}, (e_j)_{1 \leq j \leq \ell})$est une représentation d'un S.S., alors, quel que soit les bases orthonormales $(f'_i)_{1 \leq i \leq n}$ de F et $(e'_j)_{1 \leq j \leq \ell}$ de E , il existe une représentation $(v, (f'_i)_{1 \leq i \leq n} , (e'_j)_{1 \leq j \leq \ell})$du même S.S., isomorphe à la première .

Démonstration : Considérons les isométries définies par $O e_j = e'_j$ et $N f'_i = f_i$.

Alors , si $v = O u N$, $v(f'_i) = O u(f_i)$, $v^*(e'_j) = N^* u^*(e_j)$ et $< v^*(e'_j) \mid f'_i > = <u^*(e_j) |f_i >$

□

Cette proposition nous permet de prendre la même base pour toutes les representations. Nous prendrons donc dans la suite $F = \mathbb{R}^n$, $E = \mathbb{R}^\ell$, avec leur base canoniqure respective. Ainsi, une representation d'un S.S. de dimension (n, ℓ) s'identifie à un élément de $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^\ell)$.

Par définition (voir 2.1.1) nous savons que deux représentations u et v d'un même S.Ssont isomorphes car de la forme $u = \lambda v$ avec $\lambda \in \mathbb{R}$. Nous allons montrer que la réciproque est fausse en général et chercher une condition, appelée de non-dégénérescence du S.S.,pour qu'elle soit vraie.

2.2.4. Théorème

Soient u et v deux représentations isomorphes et $u = \sum_{i=1}^{r} \sqrt{\lambda_i}\ P_i$ la décomposition canonique de u (voir 1.3.). Alors, vv^* (resp. v^*v) admet les mêmes sous-espaces propres que uu^* (resp. u^*u) et il existe $\lambda > O$, tel que

$$v = \lambda \sum_{i=1}^{r} \sqrt{\lambda_i}\ Q_i .$$

Démonstration : Par hypothèse, il existe $T \in \mathcal{L}(F)$ et $S \in \mathcal{L}(E)$ tels que $T^*T = TT^* = \alpha^2 I$, $S^*S = SS^* = \beta^2 I$, $u(f_i) = Sv(f_i)$ et $u^*(e_j) = T^*V^*(e_j)$. D'où $u = Sv = vT$.

On voit aisement que ker u=ker v, val u=val v , T(ker u)=ker u, S(val v)=val v. On peut donc supposer u et v bijectifs et alors , $S = u \circ v^{-1}$ et $T = v^{-1} \circ u$.

Puisque $u = \sum_{i=1}^{r} \sqrt{\lambda_i}\ P_i$, si $x = \sum_{i=1}^{r} x_i \in F = \bigoplus_{i=1}^{r} F_i$ où $x_i \in F_i$, on a ,

$$\alpha^2 \|x\|^2 = \| v^{-1}(\sum_{i=1}^{r} \sqrt{\lambda_i}\ P_i\ x_i) \|^2$$

et (*) $\quad \alpha^2 \|x_i\|^2 = \lambda_i \| v^{-1} (P_i\ x_i) \|^2 = \alpha^2 . \|P_i x_i \|^2$.

Ainsi $\|x\|^2 = \sum_{i=1}^{r} \|x_i\|^2$ implique $\| v^{-1} \sum_{i=1}^{r} \sqrt{\lambda_i} P_i\ x_i \|^2 = \sum_{i=1}^{r} \|v^{-1} (\sqrt{\lambda_i} P_i\ x_i)\|^2$,

ou encore , $\forall\ y = \sum_{i=1}^{r} y_i \in E = \bigoplus_{i=1}^{r} E_i$, où $y_i \in E_i$, $\| v^{-1} (\sum_{i=1}^{r} y_i) \|^2 = \sum_{i=1}^{r} \| v^{-1} (y_i)\|^2$.

Cela prouve que $v^{-1}(E_i) \perp v^{-1}(E_j)$ si $i \neq j$ et d'après (*) $\|v^{-1}(y_i)\| = \frac{\alpha}{\sqrt{\lambda_i}} \|y_i\|$.

En posant $Q_i^* \,|_{E_i^\perp} = 0$ et $Q_i^* \,|_{E_i} = \frac{\alpha}{\sqrt{\lambda_i}} \, v^{-1}\,|_{E_i}$, nous avons la décomposition

spectrale de $v = \frac{1}{\alpha} \sum \sqrt{\lambda_i} Q_i$. Pour démontrer que $v^{-1}(E_i) = F_i$, il faut recommencer un raisonnement analogue avec S, au lieu de T.

□

2.2.5. Corollaire

Si $\forall i \in \{1,......, r\}$ dim $F_i = 1$, alors, u et v isomorphes, implique, u et v représentent le même S.S. (i,e. $u = \lambda v$ avec $\lambda > 0$) .

On voit donc que la réciproque que nous cherchions admet " $\forall i \in \{1,......, r\}$ dim $F_i = 1$" appelée propriété de non-dégénérescence, comme condition suffisante.Montrons qu'elle est necessaire.

2.2.6.Proposition

Si u est la représentation d'un S.S. et si (avec les notations 1.2.3. et 1.3.) $\exists\, i_o \in \{1,....,r\}$ tel que dim $F_{i_o} > 1$, il existe une représentation v d'un S.S. différent de celui de u qui soit isomorphe à u.

Démonstration : Considérons une base orthogonale $\left(\psi_k^i\right)_k$ de F_i $i \in \{1, ..., r\}$ et posons $\varepsilon_k^i = P_i\left(\psi_k^i\right)$.

Soit l'opérateur $v = \sum_{i \neq i_o} \sqrt{\lambda_i}\, P_i + \sqrt{i_o}\, Q_{i_o}$ où $Q_{i_o}(\psi_1^{i_o}) = \varepsilon_2^{i_o}$, $Q_{i_o}(\psi_2^{i_o}) = \varepsilon_1^{i_o}$

et $Q_{i_o}(\psi_k^{i_o}) = \varepsilon_k^{i_o}$ pour $k \notin \{1,2\}$.

On voit que u et v sont isomorphes (2.2.2.). Par contre, si u et v représentaient le même S.S., on aurait $u = \lambda v$ avec $\lambda \in \mathbb{R}$, ce qui est faux. □

Conclusion

Un S.S. non-dégénéré de dimension (n,ℓ), admet une représentation unique $u_p : \mathbb{R}^n \to \mathbb{R}^\ell$, à un scalaire multiplicatif près, admettant pour décomposition canonique

$$u_p = \sum_{i=1}^{r} \sqrt{\lambda_i}\ \varepsilon_i \otimes \psi_i$$

où $(\varepsilon_i)_{1 \le i \le r}$ (resp. $(\psi_i)_{1 \le i \le r}$) est une base orthogonale de val u (resp.val u*),

$\lambda_1 > \lambda_2 > \dots\dots > \lambda_r > 0$. Dans ce cas, les espaces principaux des $(u_p(f_i))_{1 \le i \le n}$

et des $(u_p^*(e_j))_{1 \le j \le \ell}$ sont uniques.

En Analyse en Composantes Principales (A.C.P) les $(\varepsilon_i)_i$ s'appellent les vecteurs principaux et les $(\psi_i)_i$ les composantes principales. Ainsi la recherche de ces derniers est équivalente à la recherche de la décomposition canonique de u_p.

Remarques

1- Les représentations u_p et $v_p = u_p \circ P_{[\varphi_p]^\perp}$ (voir 2.1 remarque 3) ne sont pas isomorphes. Elles représentent donc des S.S. différents. Cette transformation qui est faite systématiquement en ACP n'est donc pas justifiée.

2- Faire un changement d'unité pour chaque variable aléatoire $u_p^*(e_j)$, revient à la multiplier par un scalaire $\alpha_j > 0$. Si $\alpha = (\alpha_1, \dots\dots, \alpha_\ell)$ et $\delta_\alpha = \sum_{j=1}^{\ell} \alpha_j P_{[e_j]}$, la représentation du nouveau S.S. est $w = \delta_\alpha \circ u$. u et w ne sont isomorphes que si $\alpha_1 = \dots = \alpha_\ell$.

3. ANALYSE CANONIQUE ET INDICES DES VARIABLES ALEATOIRES ETAGEES.

3.1. Analyse canonique de deux sous espaces vectoriels

Soient F_1 et F_2 deux sous espaces vectoriels de F, P_{F_1} et P_{F_2} les projecteurs orthogonaux correspondants. Posons $A_1 = P_{F_1} P_{F_2} P_{F_1}$ et $A_2 = P_{F_2} P_{F_1} P_{F_2}$.

Evidemment spect $(A_i) \subset [0,1]$, $i = 1, 2$.

3.1.1. Théorème (définition de l'analyse canonique)

Avec les notations ci-dessus, il existe r sous espaces vectoriels $(H_1^i)_{1 \leq i \leq r}$ (resp. $(H_2^i)_{1 \leq i \leq r}$) de F_1 (resp. F_2) et r nombres réels $(\beta_i)_{1 \leq i \leq r}$ tels que :

* $1 > \beta_1 > \beta_2 > \ldots > \beta_r > 0$
* $F_1 = (F_1 \cap F_2) \oplus \left(\bigoplus_{i=1}^{r} H_1^i \right) \oplus (F_1 \cap F_2^{\perp})$,
* $F_2 = (F_1 \cap F_2) \oplus \left(\bigoplus_{i=1}^{r} H_2^i \right) \oplus (F_2 \cap F_1^{\perp})$,
* $A_1 = P_{F_1 \cap F_2} + \sum_{i=1}^{r} \beta_i \, P_{H_1^i}$,
* $A_2 = P_{F_1 \cap F_2} + \sum_{i=1}^{r} \beta_i \, P_{H_2^i}$,
* $\forall \; i = 1, \ldots, r$, $H_2^i = P_{F_2}(H_1^i)$ et $\dim H_2^i = \dim H_1^i$,
* $F_1 + F_2 = (F_1 \cap F_2) \oplus \left(\bigoplus_{i=1}^{r} (H_1^i + H_2^i) \right) \oplus (F_1 \cap F_2^{\perp}) \oplus (F_2 \cap F_2^{\perp})$

Les sommes directes signifient que les s.e.v. sont orthogonaux.

Démonstration : Nous allons construire par étapes les décompositions spectrales de A_1 et A_2 dont nous connaissons l'existence (1.2.).

(i) $F_1 \cap F_2$ est évidemment le sous espace propre à la valeur propre 1 de A_j, j = 1 ou 2.

(ii) $\ker A_j \supset F_j^{\perp}$, j = 1 ou 2 .

(iii) $F_1 \cap F_2^{\perp} = F_1 \cap \ker A_1$: l'inclusion $F_1 \cap F_2^{\perp} \subset F_1 \cap \ker A_1$ est évidente.

Inversement soit $x \in F_1 \cap \ker A_1$; $x = y+z$ où $y \in F_2$, $z \in F_2^{\perp}$ et $\|x\|^2 = \|y\|^2 + \|z\|^2$.

$A_1(x) = P_{F_1} P_{F_2}(x) = P_{F_1}(y) = 0$,donc $y \in F_1^{\perp}$ et $z = x-y$ où $x \in F_1$ et $y \in F_1^{\perp}$.

Ainsi $\|z\|^2 = \|x\|^2 + \|y\|^2$ ce qui implique $y = 0$ et $x \in F_2^{\perp}$.

(iv) $\ker(A_1) = (F_1 \cap F_2^{\perp}) \oplus F_1^{\perp}$: l'inclusion $\ker A_1 \supset (F_1 \cap F_2^{\perp}) \oplus F_1^{\perp}$ est évidente. Inversement si $x \in \ker A_1$, $x = y + z$, où $y \in F_1$ et $z \in F_1^{\perp}$;

d'après (ii) $y = x - z \in \ker A_1$ et d'après (iii) , $y \in F_1 \cap F_2^{\perp}$. De même, on montre que

$$\ker A_2 = (F_2 \cap F_1^{\perp}) \oplus F_2^{\perp} .$$

(v) De ce qui précède, on déduit que

$$A_1 = P_{F_1 \cap F_2} + \sum_{i=1}^{r} \beta_i P_{H_1^i} \quad \text{où } 1 > \beta_1 > \beta_2 > \ldots\ldots > \beta_r > 0 .$$

Montrons que $A_2 = P_{F_1 \cap F_2} + \sum_{i=1}^{r} \beta_i P_{H_2^i}$ où $H_2^i = P_{F_2}(H_1^i)$et $\dim H_2^i = \dim H_1^i$.

$\forall x \in H_1^i$, $A_2(P_{F_2}(x)) = P_{F_2}(A_1(x)) = \beta_i P_{F_2}(x)$. D'où $P_{F_2}(H_1^i) \subset H_2^i$

et $\dim (H_1^i) \leq \dim (H_2^i)$ puisque $P_{F_2}|_{H_1^i}$ est injectif. Les indices 1 et 2 jouant des rôles symétriques, on déduit $\dim (H_1^i) = \dim(H_2^i)$ et $P_{F_2}(H_1^i) = H_2^i$.

(vi) de (iv) et (v) on tire les décompositions de F_1 et F_2 en sommes directes.

(vii) Pour obtenir la décomposition de F_1+F_2 en somme directe, il suffit de montrer que $i \neq j$ implique $H_1^i \perp H_2^j$. En effet $\forall x \in H_1^i$, $\forall y \in H_2^j$,

$$< x \mid y > = < x \mid P_{F_2} y > = < P_{F_2} x \mid y > = 0 \quad \text{puisque } P_{F_2} x \in H_2^i.$$

□

3.1.2. Proposition

$$\forall\, i \in \{1, \ldots, r\},\ \forall\, x \in H_1^i,\ d\left(\frac{x}{\|x\|}, F_2\right) = \sqrt{1 - \beta_i}.$$

Démonstration: c'est immédiat car $d\left(\frac{x}{\|x\|}, F_2\right) = \| x - P_{F_2}(x) \| \,/\, \| x \|$ □

Ainsi l'Analyse Canonique de deux espaces F_1 et F_2, permet de les décomposer en somme directe de sous-espaces qui sont les sous-espaces propres de A_1 et A_2 respectivement .
Ces deux opérateurs ont les mêmes valeurs propres. Leurs sous -espaces propres, correspondant à une même valeur propre, forment un angle compris entre 0 et $\frac{\pi}{2}$,dont le cosinus est la valeur propre.

Par exemple :
$\forall\, i \in \{1,\ldots, r\}$, $\cos(\widehat{H_1^i, H_2^i}) = \beta_i^{1/2}$, i.e. $\forall\, x \in H_1^i$, $\langle x \mid P_{F_2}(x) \rangle = \beta_i^{1/2} \cdot \| x \| \cdot \| P_{F_2}(x) \| = \beta_i \cdot \| x \|^2$.

3.1.3. Définitions

Avec les notations **3.1.1.**, on appellera :

(i) indice de proximité de F_1 et F_2 le nombre

$$\mathcal{P}(F_1, F_2) = \frac{\mathrm{Tr}\,(P_{F_1} P_{F_2})}{\max(\dim F_1, \dim F_2)};$$

(ii) si $\dim F_1 < \dim F_2$, indice d'inclusion de F_1 dans F_2, le nombre,

$$\mathcal{I}(F_1, F_2) = \frac{\mathrm{Tr}\,(P_{F_1} P_{F_2})}{\dim F_1}.$$

Evidemment $\mathcal{P}(F_1,F_2)$ et $\mathcal{I}(F_1,F_2) \in [0,1]$,

$\mathcal{P}(F_1,F_2)$ ou $\mathcal{I}(F_1,F_2) = 0 \Leftrightarrow F_1 \perp F_2$, $\mathcal{P}(F_1,F_2) = 1 \Leftrightarrow F_1 = F_2$ et

$\mathcal{I}(F_1,F_2) = 1 \Leftrightarrow F_1 \subset F_2$.

Remarque : si $(\varphi_i)_{1 \leq i \leq n_1}$ est une base orthogonale de F_1 ($n_1 = \dim F_1$), alors l'indice d'inertie de $(\varphi_i)_i$ par rapport à 0 est, $I_o((\varphi_i)_i) = \frac{1}{n_1} \sum_{i=1}^{n_1} \| \varphi_i \|^2 = 1$;

$I_{o,F_2}((\varphi_i)_i) = \mathcal{Y}(F_1, F_2)$ et $I_{o,F_2}{}^{\perp} = 1 - \mathcal{Y}(F_1, F_2)$ représente la "distance" de F_1 à F_2.

3.2. Variables aléatoires étagées

Nous appelons variable aléatoire étagée, toute variable aléatoire X ayant un nombre fini s de valeurs que ces dernières soient numériques ou non. Ce vocable inclut donc les variables aléatoires qualitatives. Evidemment s est " beaucoup plus petit " que n = card Ω.

3.2.1. Définition

Soit (Ω, P) un espace de probabilité de n individus et F l'espace des paramètres de base $(f_i)_{1 \leq i \leq n}$. $\forall\, M \subset \{1, \ldots, n\}$, on appelle modalité associé à M, le vecteur

$$\varphi_M = \sum_{i \in M} \sqrt{p_i} \cdot f_i .$$

Toute variable aléatoire étagée induit une partition de Ω et donc de $\{1, \ldots, n\}$.
Soit X une telle variable aléatoire et $(M^k)_{i \leq k \leq s}$ la partition associée de $\{1, \ldots, n\}$.
Les $(\varphi_M^k)_k$ sont les modalités de X et le sous-espace vectoriel F_x de F engendré par les modalités s'appelle le potentiel de prévision de X. En fait F est isomorphe à $\mathcal{L}^2(\mathcal{P}(\Omega))$, F_x à $\mathcal{L}^2(\sigma(x))$ où $\sigma(x)$ est l'algèbre engendrée par les $(M^k)_k$ et $[\varphi_p]$ au sous espace des fonctions constantes.

Evidemment $k \neq k' \Rightarrow \varphi_M{}^k \perp \varphi_M{}^{k'}$ et $\| \varphi_M{}^k \|^2 = P(M^k)$.

$\varphi_M{}^k$ peut être considéré comme le representant dans F de la variable aléatoire réelle $\mathbb{1}_M{}^k$ (fonction indicatrice de M^k) ; $\dim F_x = s$ et $\varphi_p = \sum_k \varphi_M{}^k \in F_x$. Enfin $\langle \varphi_M | \varphi_N \rangle = P(M \cap N)$.

Si un S.S. contient une ou plusieurs variables aléatoires étagées, on peut remplacer ces dernières par leurs modalités. On obtient ainsi un nouveau S.S. et en faisant une A.C.P.dans F, on sait que l'on a la meilleure description possible de l'ensemble de ces vecteurs. Cette méthode permet d'étudier les S.S. contenant en plus des variables aléatoires numériques, une ou plusieurs variables aléatoires qualitatives. Elles est donc plusgénérale et plus fine que l'Analyse Factorielle Discriminante et l'Analyse Factorielle des correspondances .

3.3. Indices des variables aléatoires étagées.

Soient X et Y deux variables aléatoires étagées, F_X et F_Y leur potentiel de prévision. Evidemment $[\varphi_p] \subset F_X \cap F_Y$ et X et Y sont indépendants si et seulement si $F_X - [\varphi_p] \perp F_Y - [\varphi_p]$, i.e. $T_r(P_{F_X} P_{F_Y}) = 1$. D'où les définitions inspirées de **3.1.3.**

Définitions

On appellera

(i) Indice de proximité de X et Y , le nombre,

$$\mathcal{P}(X,Y) = \frac{\mathrm{Tr}\,(P_{F_X} P_{F_Y}) - 1}{\max\,(\dim F_X , \dim F_Y) - 1} \quad ;$$

(ii) Si $\dim F_X < \dim F_Y$, indice d'inclusion de X dans Y, le nombre

$$\mathcal{Y}(X,Y) = \frac{\mathrm{Tr}\,(P_{F_X} P_{F_Y}) - 1}{\dim F_X - 1} \quad .$$

Evidemment $\mathcal{P}(X,Y)$ et $\mathcal{Y}(X,Y) \in [0,1]$,

$\mathcal{P}(X,Y)$ ou $\mathcal{Y}(X,Y) = 0 \Leftrightarrow$ X et Y sont indépendantes , $\mathcal{P}(X,Y) = 1 \Leftrightarrow F_X = F_Y$

et $\mathcal{Y}(F_X,F_Y) = 1 \Leftrightarrow F_X \subset F_Y$.

Si $(M^k)_{1\leq k\leq s}$ (resp. $(N^\ell)_{1\leq \ell\leq t}$) est la partition induite par X (resp. Y) sur Ω ,

$$\text{alors, } \operatorname{Tr}(P_{F_X} \cdot P_{F_Y}) = \sum_{k,\ell} \frac{P\,(M^k \cap N^\ell)^2}{P\,(M^k)\ P\,(N^\ell)}$$

Dans certains cas (*), ces indices permettent de sélectionner parmi un ensemble de variables aléatoires étagées, un sous ensemble, contenant l'essentiel de l'information.

(*) J.MANUCEAU et A.KHATTABI : Analyse en Composantes Principales et Inférence Statistique. Valeur pronostiq~~s~~ue des variables aléatoires étagées. A paraître.