PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 18, Fasc. 1 (1998), pp. 1–18

QUELQUES PROPRIÉTÉS EXTRÉMALES DES VALEURS SINGULIÈRES D'UN OPÉRATEUR COMPACT ET LEURS APPLICATIONS AUX ANALYSES FACTORIELLES D'UNE PROBABILITÉ OU D'UNE FONCTION ALÉATOIRE

II. CRITÈRES D'ANALYSES FACTORIELLES LINÉAIRES D'UNE PROBABILITÉ OU D'UNE FONCTION ALÉATOIRE

PAR

ALAIN POUSSE ET JEAN-JACQUES TÉCHENÉ (PAU)

Abstract. This paper extends the work by Rao [6] concerning factor analysis criteria equivalent to the principal component analysis of a finite set of random variables. We search for global (i.e. non-iterative) criteria for the factor analysis of a probability defined on a separable Hilbert space or of a real random function other than a finite or countable set of real random variables. We compare this analysis with principal component analysis defined in a general probabilistic setting by Dauxois and Pousse [2].

1. INTRODUCTION ET NOTATIONS

Les analyses factorielles sont inspirées, de manière plus ou moins directe, de techniques relativement anciennes élaborées dans un cadre probabiliste en dimension finie en vue d'application à des problèmes de statistique inférentielle (cf. [3]), puis adaptées plus récemment à des problèmes de statistique multivariée. Ces analyses ont été, à l'origine, définies comme des méthodes itératives (ou "pas à pas"), et alors soumises à des conditions relativement exigeantes d'existence. Mais un cadre probabiliste plus général et une approche globale s'imposent si l'on veut mieux comprendre les propriétés asymptotiques de ces analyses et leur stabilité par échantillonnage ou discrétisation (cf. [2]).

Cet article est une suite de [5]. Nous cherchons à obtenir, à partir des caractères extrémaux des valeurs singulières d'un opérateur compact, des critères globaux (i.e. non-itératifs) d'analyses factorielles linéaires d'une probabilité définie sur un espace de Hilbert réel séparable ou d'une fonction aléatoire réelle non-nécessairement réduite à une famille finie de variables aléatoires réelles. Outre l'avantage d'assurer dans tous les cas l'existence de l'analyse et de fournir un moyen global de l'obtenir, une telle approche permet de dégager plusieurs familles de critères, plus forts, équivalents, ou plus faibles que l'Analyse en Composantes Principales (A.C.P.) définie dans ce cadre probabiliste général par Dauxois et Pousse [2].

L'idée de chercher des critères globaux d'analyses factorielles linéaires équivalents à l'A.C.P. d'une famille finie de variables aléatoires réelles revient à Rao (cf. [6]) et à Darroch (cf. [1], voir aussi [4]). Notre étude constitue une extension de cette recherche dans un cadre probabiliste sensiblement plus général.

1.1. Intégration sur un espace de Hilbert. Dans cette section, H désigne un espace de Hilbert réel séparable de norme notée $\|\cdot\|$. Nous identifierons H à son dual topologique, identification qui conduit logiquement à noter le produit scalaire sur H par les crochets de dualité $\langle\cdot,\cdot\rangle$ (au lieu de $(\cdot\mid\cdot)$). Nous désignerons par $\mathscr{B}_H$ la tribu de Borel de H. On se reportera au § 1 de [5] pour tout ce qui concerne les espaces $\sigma_p(H)$ et leurs normes $\|\cdot\|_p$ pour tout réel $p \geqslant 1$ ou pour $p = \infty$.

Soit une mesure bornée μ sur un espace mesurable $(\Omega, \mathscr{A})$. Nous désignerons par $L^1(\mu)$ [resp. $L^2(\mu)$] l'ensemble des applications μ-intégrables [resp. de carré μ-intégrable] de Ω dans $\boldsymbol{R}$, par M_H l'ensemble des applications mesurables de $(\Omega, \mathscr{A})$ dans $(H, \mathscr{B}_H)$, et pour chaque application f de Ω dans H, et par $\|f\|$ l'application de Ω dans $\boldsymbol{R}$ qui à tout ω de Ω associe le réel $\|f(\omega)\|$.

Nous noterons $L^1_H(\mu)$ l'ensemble des applications μ-intégrables de Ω dans H, autrement dit l'ensemble

$$L^1_H(\mu) = \{f \in M_H:\ \|f\| \in L^1(\mu)\},$$

et l'intégrale de f par rapport à μ est l'unique élément $\boldsymbol{i}(f) = \int_\Omega f(\omega)\,d\mu(\omega)$ de H, encore noté $\boldsymbol{i}(f) = \int_\Omega f\,d\mu$, défini par la relation

$$(\forall y \in H)\ \ \langle \boldsymbol{i}(f), y\rangle = \int_\Omega \langle f(\omega), y\rangle\, d\mu(\omega).$$

Nous supposerons connues les propriétés usuelles de $\boldsymbol{i}(f)$. Si μ est une probabilité sur $(\Omega, \mathscr{A})$, $\boldsymbol{i}(f)$ est alors *l'espérance mathématique de* f que nous noterons $\boldsymbol{E}(f)$.

De la même manière, nous noterons $L^2_H(\mu)$ le sous-ensemble de $L^1_H(\mu)$ défini par

$$L^2_H(\mu) = \{f \in M_H:\ \|f\| \in L^2(\mu)\}.$$

Muni du produit scalaire

$$(f\mid g) = \int_\Omega \langle f(\omega), g(\omega)\rangle\, d\mu(\omega),$$

$L^2_H(\mu)$ est un espace de Hilbert séparable.

Lorsque $(\Omega, \mathscr{A}) = (H, \mathscr{B}_H)$, nous dirons que μ est *centrée* si l'intégrale $i(I_H)$ de l'application identique I_H de H par rapport à μ est nulle ($\int_H x d\mu(x) = 0$), et que μ *admet un moment d'ordre deux* si I_H est élément de $L_H^2(\mu)$.

1.2. Opérateur de covariance d'une mesure bornée sur un espace de Hilbert. Considérons maintenant une mesure bornée μ sur $(H, \mathscr{B}_H)$ que l'on suppose *centrée* et admettant un moment d'ordre deux. Pour chaque couple (x, y) d'éléments de H, nous noterons par $x \otimes y$ l'opérateur de rang 1 de H dans H défini par

$$x \otimes y(u) = \langle x, u \rangle y.$$

Il est immédiat que, pour tout x de H, l'opérateur $x \otimes x$ est élément de $\sigma_2(H)$, que l'application Φ de H dans l'espace de Hilbert $\sigma_2(H)$ (muni de la norme de Hilbert–Schmidt $\|\cdot\|_2$, cf. [5]) qui à tout x de H associe $x \otimes x$ est continue (donc mesurable), et que (par hypothèse):

$$\int_H \|\Phi(x)\|_2 \, d\mu(x) = \int_H \|x\|^2 \, d\mu(x) < +\infty.$$

Par suite, Φ est élément de $L^2_{\sigma_2(H)}(\mu)$. Nous appellerons *l'opérateur de covariance de* μ l'intégrale V de Φ par rapport à μ, i.e.

$$V = \int_H x \otimes x \, d\mu(x).$$

Nous supposerons connues les propriétés usuelles d'un tel opérateur intégral, élément de $\sigma_2(H)$ par définition, et notamment que:

(a) V est *l'unique opérateur* de H dans H tel que

$$(\forall u \in H) \quad Vu = \int_H x \otimes x(u) \, d\mu(x).$$

(b) $(\forall u \in H, \ \forall v \in H) \ \langle Vu, v \rangle = \int_H \langle u, x \rangle \langle v, x \rangle \, d\mu(x)$.

(c) V est *auto-adjoint positif, nucléaire*, et

$$\operatorname{tr}(V) = \|V\|_1 = \int_H \|x\|^2 \, d\mu(x).$$

(d) Pour tout opérateur continu A de H dans un espace de Hilbert réel séparable H' (pouvant coïncider avec H), l'opérateur de covariance V' de la mesure image μ_A de μ par A (centrée comme μ) et V sont liés par la relation

$$V' = A \circ V \circ A^*,$$

où A^* désigne l'adjoint de A.

(e) Si H est l'espace euclidien $\boldsymbol{R}^n$ rapporté à sa base canonique et muni du produit scalaire canonique, on déduit de (b) que la matrice de représentation $\mathscr{V}$ de V dans cette base est la matrice de terme générique

$$v_{ij} = \int_{\boldsymbol{R}^n} x_i x_j \, d\mu(x),$$

où x_i et x_j sont respectivement les i-ème et j-ème composantes canoniques de x. Par suite, si μ est la loi de probabilité P_X d'un vecteur aléatoire centré $X = (X_1, \ldots, X_n)$ à valeurs dans $\boldsymbol{R}^n$, on obtient

$$v_{ij} = E(X_i X_j),$$

et $\mathscr{V}$ est bien la matrice de covariance des X_i au sens usuel.

1.3. Essai de classification des propriétés extrémales des valeurs singulières d'un opérateur compact. Dans [5], nous avons dégagé deux types de critères dont l'analyse conduit, pour un opérateur compact T donné défini sur H, à une base orthonormale $\{u_1, \ldots, u_q\}$ d'un sous-espace F de H.

Les uns conduisent à une *solution unique à une équivalence près*, les vecteurs u_i cherchés constituant une famille orthonormale de vecteurs propres de $T^* \circ T$, et F est engendré par cette famille. Nous rangerons parmi ces critères ceux faisant l'objet:

– de l'assertion (ii) du théorème 3.1 et de son corollaire 3.2,
– de la proposition 3.3 pour $\alpha > 1$,
– de la partie (b) de la proposition 3.4,
– de la proposition 3.6 et de ses corollaires 3.7 et 3.8.

Le critère de la première partie du théorème 2.7 fait également partie de ces critères. En effet, on peut interpréter la minimisation de $\|T - U\|_p$ pour U de rang fini q donné comme la recherche de deux familles orthonormales $\{u_1, \ldots, u_q\}$ et $\{v_1, \ldots, v_q\}$ respectivement de H et H' et de q nombres complexes $\alpha_1, \ldots, \alpha_q$ tels que

$$\Big\|T - \sum_{i=1}^{q} \alpha_i u_i \otimes v_i\Big\|_p$$

soit minimum. Comme $p \neq \infty$, la solution coïncide avec le développement de Schmidt

$$T_q = \sum_{i=1}^{q} s_i e_i \otimes f_i$$

d'ordre q de T, et on obtient bien (pour chaque $i \in \{1, \ldots, q\}$):

$$\alpha_i = s_i, \quad u_i = e_i \quad \text{et} \quad v_i = f_i.$$

Les autres critères conduisent à une *solution non-unique*, le sous-espace F étant unique à une équivalence près, toute base orthonormale de F répondant à la question. Ce sont les critères faisant l'objet:

– du théorème 2.3, et plus généralement de la proposition 2.5,
– de la proposition 2.8,
– de l'assertion (ii) du théorème 3.1,
– de la partie (a) de la proposition 3.4,
– de la proposition 3.5.

Les premiers apparaissent comme des *critères forts*, les seconds comme des *critères moyens*. Il faut noter le peu de différence pouvant séparer la formulation de certains critères moyens de certains critères forts.

Il est possible de dégager une troisième catégorie de critères qui ne conduisent pas à l'unicité de F, et que nous nommerons *critères faibles*. Considérons l'exemple suivant. Soient T un opérateur compact auto-adjoint et positif sur H, $\{u_1, \ldots, u_q\}$ une famille orthonormale de H engendrant un sous-espace F, P le projecteur orthogonal de H sur F, $\{\lambda_i\}_{i\in I}$ la suite pleine décroissante des valeurs propres de T, $\{\mu_j\}_{j\in J}$ celle des valeurs propres de $U = P \circ T_F$, et $\{e_i\}_{i\in I}$ une base orthonormale de vecteurs propres de T associée à la suite $\{\lambda_i\}_{i\in I}$. Alors:

(a) Nous avons vu (cf. théorème 3.1 de [5]) que

$$\sum_{i=1}^{q} \langle Tu_i, u_i\rangle^2 \leqslant \sum_{i=1}^{q} \lambda_i^2,$$

et que cette relation est une égalité si et seulement si $u_i = e_i$ pour chaque i de $\{1, \ldots, q\}$. C'est un critère *fort*.

(b) L'identité de Parseval donne

$$\sum_{i,j=1}^{q} \langle Tu_i, u_j\rangle^2 = \sum_{i,j=1}^{q} \langle Uu_i, u_j\rangle^2 = \sum_{i=1}^{q} \|Uu_i\|^2 = \|U\|_2^2 \leqslant \sum_{i=1}^{q} \lambda_i^2.$$

D'après le théorème 2.2, pour qu'il y ait égalité, il faut et il suffit que F soit engendré par $\{e_1, \ldots, e_q\}$. On obtient un critère *moyen*.

(c) La différence entre les deux critères (a) et (b) s'exprime par

$$\sum_{i\neq j} \langle Tu_i, u_j\rangle^2 \geqslant 0.$$

Le minimum, zéro, est atteint si et seulement si, pour chaque couple (i, j), $i \neq j$, on a

$$\langle Tu_i, u_j\rangle = \langle Uu_i, u_j\rangle = 0.$$

Chaque vecteur u_i doit donc être un vecteur propre de U ou encore de $P \circ T \circ P$. On peut donc choisir F arbitraire de dimension q; l'opérateur compact $P \circ T \circ P$ est alors déterminé, et toute base orthonormale de F formée de vecteurs propres de $P \circ T \circ P$ répond clairement à la question.

On obtient ainsi un critère *faible*. Toutefois, si H est de dimension finie q, F coïncide nécessairement avec H et ce critère devient alors un critère fort. Un critère faible pour la recherche simultanée d'un certain nombre de vecteurs peut donc se transformer en critère fort lorsqu'on les cherche globalement tous.

2. ANALYSES FACTORIELLES LINÉAIRES D'UNE MESURE BORNÉE SUR UN ESPACE DE HILBERT

Dans toute cette section, μ désigne une mesure bornée sur un espace de Hilbert réel séparable H. On suppose que μ est *centrée* et *admet un moment d'ordre deux*. On note K le *support de* μ, autrement dit, le plus petit sous-ensemble fermé de H en dehors duquel μ est nulle. Par hypothèse, K est *implicitement* muni de sa tribu de Borel $\mathscr{B}_K$, sous-tribu de $\mathscr{B}_H$, et nous noterons simplement (K, μ) l'espace mesuré $(K, \mathscr{B}_K, \mu)$. Les autres notations sont celles de la section 1.

On considère un sous-espace de Hilbert F de H, de supplémentaire orthogonal noté $F^\perp$, et on désigne par A et B respectivement les projecteurs orthogonaux de H sur F et $F^\perp$, par μ_A et μ_B les mesures images de μ par A et B sur F et $F^\perp$.

2.1. Sous-espace de dimension finie le plus proche d'un espace mesuré. Comme μ est centrée et admet un moment d'ordre deux, les mesures μ_A et μ_B sont centrées et admettent chacune un moment d'ordre deux. Nous conviendrons de la définition suivante:

DÉFINITION 2.1. Nous appellerons *moment d'inertie de l'espace mesuré* (K, μ) *relativement à un sous-espace de Hilbert* F *de* H le réel

$$I_F = \int_{F^\perp} \|u\|^2 \, d\mu_B(u) = \int_H \|Bx\|^2 \, d\mu(x).$$

Soit $\mathscr{F}_q$ la famille des sous-espaces vectoriels (fermés) de H de dimension finie au plus égale à un entier q donné ($q \leqslant \dim H$). S'il existe un élément F_0 de $\mathscr{F}_q$ tel que

$$I_{F_0} = \operatorname{Inf}\{I_F \colon F \in \mathscr{F}_q\},$$

nous dirons que F_0 est un *sous-espace de dimension finie au plus égale à q le plus proche de* (K, μ).

On définit de manière analogue le *moment d'inertie* $I_{F^\perp}$ *de* (K, μ) *relativement au sous-espace* $F^\perp$. Il est immédiat que

$$I_F + I_{F^\perp} = \|V\|_1 = \operatorname{tr}(V). \tag{1}$$

En effet, comme $A+B = I_H$, on a

$$I_F + I_{F^\perp} = \int_H \|Bx\|^2 \, d\mu(x) + \int_H \|Ax\|^2 \, d\mu(x) = \int_H \|x\|^2 \, d\mu(x) = \operatorname{tr}(V).$$

Remarque. Prenant $F = \{0\}$, le moment d'inertie de (K, μ) par rapport à l'origine est donc égal à la trace $\operatorname{tr}(V)$. On retrouve ainsi un résultat connu lorsque K est un sous-ensemble fini de points pondérés d'un espace euclidien $\boldsymbol{R}^d$ (cf. [6] par exemple).

THÉORÈME 2.2. *Pour qu'un sous-espace F de H de dimension finie au plus égale à un entier q donné soit le plus proche de (K, μ), il faut et il suffit que F soit engendré par q premiers vecteurs propres de l'opérateur de covariance V de μ.*

En effet, de la relation (1) on déduit que

$$\operatorname{Inf}\{I_F\colon F \in \mathscr{F}_q\} = \|V\|_1 - \operatorname{Sup}\{I_{F^\perp}\colon F \in \mathscr{F}_q\},$$

où, par définition:

$$I_{F^\perp} = \int_F \|u\|^2 \, d\mu_A(u) = \|V'\|_1 = \|A \circ V \circ A\|_1 = \|A \circ V_F\|_1,$$

$V' = A \circ V \circ A$ désignant l'opérateur de covariance de μ_A (A étant auto-adjoint) et V_F la restriction de V au sous-espace F. Le résultat apparaît alors comme une conséquence immédiate du théorème 2.3 de [5].

Remarque. Le sous-espace F obtenu n'est unique qu'à une équivalence près. Néanmoins, si $\{\lambda_j\}_{j \in J}$ désigne la suite pleine des valeurs propres de V et si $\lambda_q \neq \lambda_{q+1}$ ($q < \operatorname{card} J$), alors le sous-espace F cherché est unique et

$$I_{F^\perp} = \sum_{i=1}^{q} \lambda_i.$$

2.2. Sous-espace de codimension finie donnée le plus éloigné d'un espace mesuré. De manière analogue, on peut chercher s'il existe un sous-espace G de H, de codimension finie donnée q (ou de codimension finie au plus égale à q), tel que I_G soit maximum. Cela revient à maximiser $\|A \circ V \circ A\|_1$ lorsque A est un projecteur orthogonal de H d'image de codimension finie donnée q (ou de codimension au plus égale à q). Le résultat découle directement de la proposition 2.8 de [5]:

PROPOSITION 2.3. *Pour qu'un sous-espace G de H de codimension finie au plus égale à un entier q donné soit le plus éloigné de (K, μ), il faut et il suffit que le supplémentaire orthogonal $G^\perp$ de G soit engendré par q premiers vecteurs propres de l'opérateur de covariance V de μ.*

Remarque. On peut aussi déduire ce résultat du théorème 2.2 précédent en observant qu'imposer $\operatorname{codim} G = q$ revient à imposer $\dim G^\perp = q$, et par suite que l'on a

$$\operatorname{Sup}\{I_G\colon \operatorname{codim} G = q\} = \|V\|_1 - \operatorname{Inf}\{I_{G^\perp}\colon \dim G^\perp = q\}.$$

On est ainsi ramené à chercher un sous-espace de dimension q (ou au plus égale à q) le plus proche de (K, μ). Ceci provient du fait que l'opérateur de covariance V de μ vérifie la propriété

$$\|A \circ V \circ A\|_1 + \|(I_H - A) \circ V \circ (I_H - A)\|_1 = \|V\|_1,$$

alors qu'en général, pour tout T de $\sigma_p(H)$, avec $p \in [1, +\infty[$, on a seulement (cf. [5]):

$$\|A \circ T \circ A\|_p + \|(I_H - A) \circ T \circ (I_H - A)\|_p \leqslant \|V\|_p.$$

Cette remarque renforce l'intérêt de la proposition 2.8 de [5].

2.3. Maximisation de la moyenne des carrés des distances entre les points d'un espace mesuré. Nous conviendrons de la définition suivante:

DÉFINITION 2.4. Nous nommerons *moyenne des carrés des distances entre les points de l'espace mesuré* (K, μ) le réel

$$M = \frac{1}{2} \int_{K^2} \|x-y\|^2 \, d(\mu \otimes \mu)(x, y).$$

PROPOSITION 2.5. *On a*: $M = \mu(K) \operatorname{tr}(V)$.

En effet, il vient immédiatement en utilisant le théorème de Fubini:

$$2M = \Big[\int_K \|x\|^2 d\mu(x) \int_K d\mu(y) + \int_K \|y\|^2 d\mu(y) \int_K d\mu(x) - 2 \int_{K^2} \langle x, y \rangle \, d(\mu \otimes \mu)(x, y)\Big].$$

Par une nouvelle application du théorème de Fubini, il suit, μ étant centrée:

$$\begin{aligned}\int_{K^2} \langle x, y \rangle \, d(\mu \otimes \mu)(x, y) &= \int_{K^2} [\langle x, y \rangle \, d\mu(x)] \, d\mu(y) \\ &= \int_K \Big[\Big\langle \int_K x d\mu(x), y \Big\rangle\Big] \, d\mu(y) = 0,\end{aligned}$$

avec, d'autre part:

$$\int_K \|x\|^2 d\mu(x) \int_K d\mu(y) = \int_K \|y\|^2 d\mu(y) \int_K d\mu(x) = \|V\|_1 \, \mu(K) = \operatorname{tr}(V) \, \mu(K),$$

ce qui achève la preuve. ■

Remarque. On rapprochera ce résultat de celui obtenu dans un cadre matriciel par Rao dans [6] (p. 335) lorsque K est un nuage de n points (de poids 1) d'un espace $\boldsymbol{R}^d$ muni de sa structure euclidienne canonique; il a établi que $M = n \operatorname{tr}(V)$, où V est la matrice d'inertie du nuage.

Soit H' un espace de Hilbert réel séparable. On peut chercher une "représentation" de (K, μ) dans H', de dimension donnée q, telle que la moyenne des carrés des distances entre les points de cette représentation soit la plus proche possible de ce qu'elle était dans K. Une telle représentation peut se faire au moyen d'un opérateur $\varDelta$ de $L(H, H')$. Nous nous limitons ici au cas où cet opérateur est une *isométrie partielle*, ce qui revient éventuellement à conserver ou à représenter isométriquement une partie de K. Sous une telle hypothèse, la moyenne $M_\varDelta$ des carrés des distances entre les points de $\varDelta(K)$ est (cf. proposition 2.5):

$$M_\Delta = \mu_\Delta[\Delta(K)]\,\|\Delta\circ V\circ\Delta^*\|_1 = \mu(K)\,\|\Delta\circ V\circ\Delta^*\|_1,$$

et, par application de la propriété 5 du § 1.3 de [5], est inférieure ou égale à la moyenne des carrés des distances entre les points de K. Par suite, on cherche une isométrie partielle Δ de H dans H' maximisant $\|\Delta\circ V\circ\Delta^*\|_1$. Or, $\Delta^*\circ\Delta$ est un projecteur orthogonal P de $L(H)$ tel que $\Delta = \Delta\circ P$ (cf. § 1.1 de [5]), et, V étant auto-adjoint positif, $\Delta\circ V\circ\Delta^*$ l'est aussi, et par suite:

$$\|\Delta\circ V\circ\Delta^*\|_1 = \operatorname{tr}(\Delta\circ V\circ\Delta^*) = \operatorname{tr}(\Delta^*\circ\Delta\circ V) = \operatorname{tr}(P\circ V) = \|P\circ V\circ P\|_1,$$

et le problème revient donc à chercher un sous-espace $\operatorname{Im}P$ de dimension finie donnée q tel que $\|P\circ V\circ P\|_1$ soit maximum, autrement dit, un sous-espace de dimension q le plus proche de (K, μ), et toute isométrie partielle Δ dont l'ensemble initial est un tel sous-espace convient.

Nous énoncerons donc la

PROPOSITION 2.6. *Soit H' un espace de Hilbert réel de dimension finie q. Pour que la moyenne des carrés des distances entre les points d'une représentation de (K, μ) au moyen d'une isométrie partielle Δ de H dans H' soit maximum, il faut et il suffit que l'ensemble initial de Δ soit un sous-espace de H engendré par q premiers vecteurs propres de l'opérateur de covariance V de μ.*

2.4. Représentation d'un espace mesuré par un système orthonormal de cardinal fini donné. On cherche maintenant une représentation optimale de (K, μ) par un système *orthonormal* $\{u_i\colon i = 1, \ldots, q\}$ de cardinal fini donné q ($q \leqslant \dim H$), chaque vecteur u_i étant muni d'un poids ν_i ($\nu_i > 0$). Le critère d'optimalité retenu ici est la minimisation de la norme de Hilbert–Schmidt de $V - V'$, où

$$V' = \sum_{i=1}^{q} \nu_i u_i \otimes u_i$$

est l'opérateur de covariance de la mesure ν sur $\{u_i\colon i = 1, \ldots, q\}$ affectant chaque point u_i de son poids ν_i ($\nu(\{u_i\}) = \nu_i$). Lorsque H est un espace $\boldsymbol{R}^d$ muni de la structure euclidienne canonique, la matrice de représentation canonique de V' n'est autre que la matrice d'inertie du nuage pondéré $\{(u_i, \nu_i)\colon i = 1, \ldots, q\}$ (cf. [6]).

En vertu du théorème 2.7 de [5], $\|V - V'\|_2$ est minimum si et seulement si V' coïncide avec le développement de Schmidt

$$V_q = \sum_{i=1}^{q} \lambda_i e_i \otimes e_i$$

d'ordre q de V (cf. [5], § 1.3). Il est alors immédiat que, pour chaque i de $\{1, \ldots, q\}$, on a nécessairement $\nu_i = \lambda_i$ et $u_i = e_i$, d'où la

PROPOSITION 2.7. *Pour qu'un système orthonormal pondéré de H de cardinal fini donné q,* $\{(u_i, v_i)\colon i = 1, \ldots, q\}$ $(v_i > 0)$, *constitue une représentation optimale de* (K, μ), *il faut et il suffit que*

$$(\forall i \in \{1, \ldots, q\})\ u_i = e_i \ \text{et}\ v_i = \lambda_i,$$

où $\{e_i\colon i = 1, \ldots, q\}$ *est un système orthonormal de q premiers vecteurs propres associés à la suite pleine décroissante* $\{\lambda_i\colon i = 1, \ldots, q\}$ *de q premières valeurs propres (strictement positives) de l'opérateur de covariance V de* μ.

Remarque 1. La mesure ν sur $\{u_i\colon i = 1, \ldots, q\}$ n'est pas centrée, et donc une telle représentation ne respecte pas la condition d'égalité des moments d'ordre 1.

Remarque 2. Au lieu de la norme de Hilbert–Schmidt, qui est celle utilisée historiquement par Rao [6] ou Darroch [1] puisqu'elle correspond à la norme euclidienne des matrices, le théorème 2.7 de [5] permet d'envisager, lorsque K est un *sous-ensemble fini* d'un espace de Hilbert séparable, la minimisation de $\|V - V'\|_p$ pour tout p de $[1, +\infty[$ et indique que le résultat est conservé.

2.5. Analyse en Composantes Principales d'une mesure bornée. On peut comparer les analyses factorielles précédentes de la mesure μ avec son Analyse en Composantes Principales (A.C.P.), introduite par Dauxois et Pousse dans [2] (pp. 303–306), à laquelle on associe le *schéma de dualité*:

$$\begin{array}{ccc} H & \xleftarrow{U} & L^2(\mu) \\ {\scriptstyle I}\downarrow\uparrow{\scriptstyle V} & & {\scriptstyle W}\downarrow\uparrow{\scriptstyle I} \\ H & \xrightarrow{U^*} & L^2(\mu) \end{array} \qquad (V = U \circ U^*,\ W = U^* \circ U),$$

où U et son adjoint U^* sont définis par

$$(\forall f \in L^2(\mu))\ Uf = \int_H x f(x)\, d\mu(x), \qquad (\forall y \in H)\ U^* y = \langle \cdot, y \rangle.$$

L'A.C.P. de μ est fournie par l'analyse spectrale de V, opérateur de covariance de μ. Les facteurs principaux de μ constituent une base orthonormale de H formée de vecteurs propres de V associés à la suite pleine décroissante de ses valeurs propres.

De tous les critères étudiés dans cette section, seul celui du § 2.4 est équivalent à l'A.C.P. de μ: ce sont des *critères forts*. Par contre, les deux autres analyses correspondent à des *critères moyens* dont l'A.C.P. pas à pas de μ fournit seulement *une* solution.

2.6. Repère mobile d'inertie. Soient F un sous-espace de H de dimension finie égale à q, et P le projecteur orthogonal de H sur F. Pour analyser la mesure μ, on se propose d'utiliser le critère faible étudié au § 1.3: on cherche une base orthonormale $\{u_i\colon i = 1, \ldots, q\}$ de F telle que la somme $\sum_{i \neq j} \langle Vu_i, u_j \rangle^2$ soit minimum, autrement dit soit *nulle*.

Avec les notations du § 2.5, relevons que si l'on considère les variables aléatoires $Y_i = U^* u_i$ $(i = 1, \dots, q)$, on a

$$\langle V u_i, u_j \rangle = \mathrm{Cov}(Y_i, Y_j),$$

car μ est centrée.

On obtient (cf. § 1.3) une base orthonormale $\{u_i\colon i = 1, \dots, q\}$ de F formée de vecteurs propres de $P \circ V \circ P = V'$, opérateur de covariance de la mesure image μ_P de μ par le projecteur orthogonal P. Or, si les valeurs propres associées sont rangées dans l'ordre décroissant, la droite de direction u_1 est le *premier axe d'inertie de* (F, μ_P), autrement dit, la droite vectorielle de F la plus proche de (F, μ_P) (cf. théorème 2.2); de même la droite de direction u_2 est le *deuxième axe d'inertie de* (F, μ_P), et ainsi de suite...

A chaque sous-espace F, on peut donc ainsi associer un *repère d'inertie* de la projection orthogonale μ_P de μ sur F dont les vecteurs de base sont les facteurs principaux de l'A.C.P. de μ_P puisque ce sont des vecteurs propres de V' associés à la suite pleine décroissante de ses valeurs propres. Ce repère est évidemment variable avec le sous-espace F.

3. ANALYSES FACTORIELLES LINÉAIRES D'UNE FONCTION ALÉATOIRE RÉELLE

Dans toute cette section, on désigne par $X = (X_t)_{t \in T}$ une fonction aléatoire réelle définie sur un espace probabilisé $(\Omega, \mathscr{A}, P)$, l'espace paramétrique T étant supposé muni d'une tribu $\mathscr{T}$ et d'une mesure *bornée* μ si T n'est pas dénombrable. Si T est l'ensemble $\boldsymbol{N}$ des entiers naturels, ou un sous-ensemble de $\boldsymbol{N}$, la tribu T sera la tribu des parties de T et μ la mesure de dénombrement.

Les produits scalaires sur les espaces de Hilbert $L^2(P)$ et $L^2(\mu)$ seront respectivement notés $\langle \cdot, \cdot \rangle$ et $\langle \cdot, \cdot \rangle_T$, et leurs normes associées $\|\cdot\|$ et $\|\cdot\|_T$. Nous désignerons par $P \otimes \mu$ la mesure produit des mesures P et μ sur l'ensemble $\Omega \times T$ muni de la tribu produit $\mathscr{A} \otimes \mathscr{T}$.

Nous désignerons par X l'application de $\Omega \times T$ dans $\boldsymbol{R}$ qui à tout (ω, t) de $\Omega \times T$ associe le réel $X(\omega, t) = X_t(\omega)$, et supposerons que:

(a) X est élément de l'espace $L^2(P \otimes \mu)$, ce qui implique notamment que

(i) $(\forall t \in T)$ $X(\cdot, t) = X_t \in L^2(P)$;

(ii) $(\forall \omega \in \Omega)$ $X(\omega, \cdot) \in L^2(\mu)$.

(b) X est *centrée*, i.e. $(\forall t \in T)$ $\boldsymbol{E}(X_t) = 0$.

Enfin, on considère l'application X' de Ω dans $L^2(\mu)$ qui à tout ω de Ω associe sa *trajectoire* $X(\omega, \cdot)$ dans $L^2(\mu)$. On munit l'espace $L^2(\mu)$ de sa tribu borélienne. Cette application X' apparaît alors comme une variable aléatoire dont on note $P_{X'}$ la loi de probabilité. Il est immédiat que $P_{X'}$ est centrée: en effet, d'après le théorème de Fubini, l'espérance mathématique $\boldsymbol{E}(X')$ de X' est l'unique élément de $L^2(\mu)$ tel que

$$(\forall t \in T) \; [\boldsymbol{E}(X')](t) = \boldsymbol{E}(X_t) = 0.$$

De plus, $P_{X'}$ admet un opérateur de covariance car

$$\int_{L^2(\mu)} \|u\|^2 \, dP_{X'}(u)$$
$$= \int_\Omega \|X(\omega, \cdot)\|^2 \, dP(\omega) = \int_\Omega \Big[\int_T X^2(\omega, t)\, d\mu(t)\Big] \, dP(\omega) = |||X|||^2 < +\infty,$$

où $|||\cdot|||^2$ désigne la norme de l'espace de Hilbert $L^2(P \otimes \mu)$.

Nous désignerons par V cet opérateur que nous nommerons *opérateur de covariance de la fonction aléatoire* X; V est nucléaire, auto-adjoint positif et, par définition, donné par la relation (cf. § 1.2):

$$V = \int_{L^2(\mu)} u \otimes u \, dP_{X'}(u) = \int_\Omega X'(\omega) \otimes X'(\omega) \, dP(\omega).$$

Ainsi V apparaît comme l'espérance mathématique de l'opérateur aléatoire $X' \otimes X'$ de l'espace de Banach $\sigma_1[L^2(\mu)]$ — et donc aussi de l'espace de Hilbert $\sigma_2[L^2(\mu)]$ — que l'on devrait noter $\boldsymbol{E}(X' \otimes X')$ mais que l'on notera (par abus d'écriture):

$$V = \boldsymbol{E}(X \otimes X).$$

3.1. Le concept d'analyse factorielle linéaire d'une fonction aléatoire réelle. Considérons le schéma de dualité associé à la fonction aléatoire X (cf. [2]):

$$\begin{array}{ccc} L^2(\mu) & \xleftarrow{U} & L^2(P) \\ {\scriptstyle I}\downarrow\uparrow{\scriptstyle V} & & {\scriptstyle W}\downarrow\uparrow{\scriptstyle I} \\ L^2(\mu) & \xrightarrow{U^*} & L^2(P) \end{array} \qquad (V = U \circ U^*, \ W = U^* \circ U),$$

où U et son adjoint U^* sont définis par:

$$(\forall Y \in L^2(P)) \ [U(Y)](t) = \int_\Omega X(\omega, t) \, Y(\omega) \, dP(\omega) \ \ \mu\text{-presque-partout}$$
$$= \langle Y, X_t \rangle \ \ \mu\text{-presque-partout},$$

$$(\forall u \in L^2(\mu)) \ [U^*(u)](\omega) = \int_T X(\omega, t) \, u(t) \, d\mu(t) \ \ P\text{-presque-partout}$$
$$= \langle u, X'(\omega) \rangle_T \ \ P\text{-presque-partout}.$$

On a $U \circ U^* = V$, opérateur de covariance de X, et $U^* \circ U = W$, opérateur de covariance de la mesure image μ_Ξ de μ par l'application Ξ de T dans $L^2(P)$ qui à tout t de T associe X_t. Les opérateurs nucléaires V et W, auto-adjoints positifs, ont les mêmes valeurs propres. De plus, si v est vecteur propre de V de valeur propre associée λ, alors $Z = U^* v$ est vecteur propre de W associé à la même valeur propre; la réciproque est évidemment vraie si U^* est injectif.

Remarque importante. Lorsque T n'est pas dénombrable, *l'hypothèse que la mesure μ est bornée est essentiellement faite pour assurer la mesurabilité des images $U(Y)$ des éléments Y de $L^2(P)$ par U.* Lorsque T est un sous-ensemble de $\boldsymbol{N}$ (pouvant coïncider avec $\boldsymbol{N}$), cette mesurabilité est acquise puisque T est alors muni de la tribu de ses parties, et nous prenons naturellement pour μ la mesure de dénombrement (qui n'est évidemment pas bornée si T n'est pas fini).

Nous conviendrons de la définition suivante:

DÉFINITION 3.1. Nous appellerons *analyse factorielle linéaire de la fonction aléatoire X* toute recherche d'une famille libre $\{u_i\colon i \in I\}$ de $L^2(\mu)$ éventuellement orthogonale (où I est une section commençante de N^*, ou N^*) telle que la suite $\{Y_i = U^* u_i\colon i \in I\}$ de variables aléatoires de $L^2(P)$ soit une représentation optimale de X (en divers sens que nous allons définir).

Nous distinguerons trois types de critères d'optimalité: des *critères forts*, des *critères moyens* et des *critères faibles*. Sauf éventuellement dans le cas de l'Analyse en Composantes Principales (A.C.P.) de X, *l'ensemble I sera supposé fini de cardinal donné q ($q \leqslant \dim [L^2(\mu)]$).*

Les variables aléatoires Y_i ($i \in I$) ainsi définies sont centrées car

$$(\forall i \in I)\ \ E(Y_i) = E(\langle u_i, X_t \rangle) = \langle u_i, E(X') \rangle_T = 0,$$

et par suite, pour tout couple (i, j):

$$\operatorname{Cov}(Y_i, Y_j) = \langle Y_i, Y_j \rangle = \langle U^* u_i, U^* u_j \rangle = \langle U \circ U^* u_i, u_j \rangle_T = \langle V u_i, u_j \rangle_T.$$

A.C.P. linéaire de X. Nous rappelons (cf. [2]) que l'A.C.P. linéaire de X consiste en la recherche d'une suite *orthonormale* $\{u_i\colon i \in I\}$ de $L^2(\mu)$ par le procédé itératif (ou "pas à pas") suivant, nécessairement fini lorsque $L^2(\mu)$ est de dimension finie:

$Y_1 = U^* u_1$ est de variance maximum,
$Y_2 = U^* u_2$ est de variance maximum (avec la contrainte u_2 orthogonal à u_1),
. . . .
$Y_k = U^* u_k$ est de variance maximum (avec la contrainte u_k orthogonal à $u_1, \ldots, u_{k-1}$), etc.

Comme, pour tout i de I, $\operatorname{Var}(Y_i) = \langle V u_i, u_i \rangle_T$, l'analyse spectrale de V donne immédiatement la réponse à ce critère d'optimisation: la suite $\{u_i\colon i \in I\}$ est une suite orthonormale de vecteurs propres de V associée à la suite pleine décroissante des valeurs propres strictement positives de V; ces vecteurs s'appellent les *facteurs principaux* et les $Y_i = U^* u_i$ les *composantes principales* de l'A.C.P. linéaire de X.

3.2. Critères forts d'analyses factorielles linéaires. Nous distinguerons trois types de critères forts:

(a) *Maximisation de $\sum_{i=1}^{q} f[\operatorname{Var}(Y_i)]$, f strictement convexe sur R_+.* Du corollaire 3.2 de [5], on déduit clairement le

THÉORÈME 3.2. *Lorsque $\{u_i\colon i = 1, \ldots, q\}$ est un système orthonormal de $L^2(\mu)$ et f une application strictement convexe de R_+ dans R, le maximum de*

$\sum_{i=1}^{q} f\,[\mathrm{Var}(Y_i)]$ *est atteint si et seulement si* $\{u_i\colon i = 1, \ldots, q\}$ *est une famille orthonormale de* q *premiers vecteurs propres de* V.

Il en découle aussitôt le résultat suivant:

Corollaire 3.3. *Pour tout réel* $\alpha > 1$, *le maximum de*

$$\sum_{i=1}^{q} [\mathrm{Var}(Y_i)]^{\alpha}$$

est atteint si et seulement si $u_1, \ldots, u_q$ *sont* q *premiers vecteurs propres orthonormaux de* V.

(b) *Représentation optimale de* X *par une famille finie* $\{Y_i = U^* u_i\colon i = 1, \ldots, q\}$ *pondérée de* $L^2(P)$. On peut considérer que la famille $\{Y_i = U^* u_i\colon i = 1, \ldots, q\}$ est une représentation optimale de la fonction aléatoire X si $\{(u_i, v_i)\colon i = 1, \ldots, q\}$ $(v_i > 0)$, système orthonormal de $L^2(\mu)$ pondéré, est une représentation optimale de l'espace probabilisé $(\overline{X'(\Omega)}, P_{X'})$ au sens du § 2.4, la tribu des événements étant (implicitement) la tribu de Borel de $\overline{X'(\Omega)}$: le critère retenu est donc la minimisation de $\|V - V'\|_2$, où

$$V' = \sum_{i=1}^{q} v_i u_i \otimes u_i$$

est l'opérateur de covariance de la mesure v sur $\{u_i\colon i = 1, \ldots, q\}$ affectant chaque point u_i de son poids v_i. De la proposition 2.7, il découle donc immédiatement la

Proposition 3.4. *Pour que* $\{Y_i = U^* u_i\colon i = 1, \ldots, q\}$, *où* $\{(u_i, v_i)\colon i = 1, \ldots, q\}$ *est un système orthonormal pondéré de* $L^2(\mu)$ $(v_i > 0,\ q \leqslant \dim L^2(\mu)$ *donné*), *soit une représentation optimale de la fonction aléatoire* X *au sens où*

$$\Big\|V - \sum_{i=1}^{q} v_i u_i \otimes u_i\Big\|_2$$

est minimum, il faut et il suffit que

$$(\forall i \in \{1, \ldots, q\})\ u_i = e_i \ \text{et}\ v_i = \lambda_i,$$

où $\{e_i\colon i = 1, \ldots, q\}$ *est un système orthonormal de* q *premiers vecteurs propres associés à la suite pleine décroissante* $\{\lambda_i\colon i = 1, \ldots, q\}$ *de* q *premières valeurs propres (strictement positives) de* V.

Remarque. Désignant par $\{u_i\colon i = 1, \ldots, q\}$ une famille de $L^2(\mu)$, on peut également considérer que $\{Y_i = U^* u_i\colon i = 1, \ldots, q\}$ est une représentation optimale de la fonction aléatoire X si $\{(Y_i, \bar{\omega}_i)\colon i = 1, \ldots, q\}$ est un système orthonormal pondéré de $L^2(P)$ (avec $\bar{\omega}_i > 0$) constituant une représentation optimale de l'espace mesuré $(\overline{X(T)}, \mu_{\Xi})$, où μ_{Ξ} est l'image de μ par l'application Ξ de T dans $L^2(P)$ qui à tout t de T associe X_t (cf. § 3.1). L'opérateur

$W = U^* \circ U$ étant l'opérateur de covariance de μ_{Ξ}, la réponse est encore fournie par la proposition 2.7: il faut et il suffit que $Y_1, \ldots, Y_q$ soient q premiers vecteurs propres orthonormaux de W associés à q premières valeurs propres (strictement positives) $\lambda_1, \ldots, \lambda_q$ de W (ou de V) et que $\bar{\omega}_i = \lambda_i$ pour chaque i de $\{1, \ldots, q\}$. Lorsque U^* est injectif (hypothèse assez restrictive, mais "réaliste" dans les applications courantes), $\{u_i\colon\ i = 1, \ldots, q\}$ est alors un système orthogonal de q premiers vecteurs propres de V.

(c) *Critères spécifiques de la dimension finie.* Lorsque la fonction aléatoire X se réduit au vecteur aléatoire $X = (X_1, \ldots, X_p)$ de $\boldsymbol{R}^p$, l'espace $L^2(\mu)$ étant alors identifié à l'espace $\boldsymbol{R}^p$ muni de son produit scalaire canonique, on déduit de la proposition 3.6 de [5] et de ses corollaires 3.7 et 3.8 la

PROPOSITION 3.5. *Soient $\{u_i\colon\ i = 1, \ldots, p\}$ une base orthonormale de $\boldsymbol{R}^p$, n un entier de $\{2, \ldots, p\}$ et S_n la fonction symétrique élémentaire sur $\boldsymbol{R}^p$ définie par*

$$S_n\colon\ (x_1, \ldots, x_p) \to \sum_{1 \leqslant i_1 < \ldots < i_n \leqslant p} x_{i_1} \ldots x_{i_n}.$$

Alors:

(i) *le minimum de* $S_n[\operatorname{Var}(Y_1), \ldots, \operatorname{Var}(Y_p)]$,

(ii) *le maximum de* $\sum_{i,j=1}^{p} |\operatorname{Var}(Y_i) - \operatorname{Var}(Y_j)|^{\alpha}$, *où α est un réel de* $[1, +\infty[$ *sont atteints si et seulement si $u_1, \ldots, u_q$ sont q premiers vecteurs propres orthonormaux de V.*

CONCLUSION. Chacun de ces critères forts est équivalent à l'A.C.P. linéaire de X. La représentation optimale de X obtenue est celle donnée par q premières composantes principales de X.

3.3. Critères moyens d'analyses factorielles linéaires. Dans cette partie les espaces $L^2(\mu)$ et $L^2(P)$ sont munis de leurs tribus boréliennes et respectivement des mesures bornées $P_{X'}$ et μ_{Ξ}. Rappelons que

$$X'(\Omega) = \{X(\omega, \cdot)\colon\ \omega \in \Omega\}$$

est l'ensemble des trajectoires de X et que

$$\Xi(T) = \{X(\cdot, t)\colon\ t \in T\} = \{X_t\colon\ t \in T\}.$$

(a) *Recherche de $F_u = \operatorname{vect}\{u_1, \ldots, u_q\}$ sous-espace le plus proche de $(\overline{X'(\Omega)}, P_{X'})$ dans $L^2(\mu)$ ou de $G_Y = \operatorname{vect}\{Y_1, \ldots, Y_q\}$ sous-espace le plus proche de $(\overline{X(T)}, \mu_{\Xi})$ dans $L^2(P)$.* Du théorème 2.2, il découle donc clairement la

PROPOSITION 3.6. *Soit $\{u_i\colon\ i = 1, \ldots, q\}$ une famille libre de $L^2(\mu)$.*

(1) *Pour que $F_u = \operatorname{vect}\{u_1, \ldots, u_q\}$ soit un sous-espace de dimension q le plus proche de l'espace probabilisé $(\overline{X'(\Omega)}, P_{X'})$, il faut et il suffit que $\{u_1, \ldots, u_q\}$ soit un système libre d'un sous-espace de dimension q de $L^2(\mu)$ engendré par q premiers*

vecteurs propres de V; alors $Y_1, \ldots, Y_q$ appartiennent à un sous-espace de $L^2(P)$ engendré par q premiers vecteurs propres de W.

(2) *Pour que $G_Y = \operatorname{vect}\{Y_1, \ldots, Y_q\}$ soit un sous-espace de dimension finie au plus égale à q le plus proche de l'espace mesuré $(\overline{X(T)}, \mu_\Xi)$, il faut et il suffit que $\{Y_1, \ldots, Y_q\}$ soit une famille libre d'un sous-espace de dimension q de $L^2(P)$ engendré par q premiers vecteurs propres de W; si l'opérateur U^* est injectif, alors $u_1, \ldots, u_q$ appartiennent à un sous-espace de $L^2(\mu)$ engendré par q premiers vecteurs propres de V.*

(b) *Maximisation de la somme des variances, ou de la somme des carrés des covariances, ou de la variance généralisée.* La conjonction du théorème 3.1 de [5] (assertion (iii)), de la propriété (b) du § 1.3 et de la proposition 3.5 de [5], entraîne clairement le

Théorème 3.7. *Pour toute famille orthonormale $\{u_i\colon i = 1, \ldots, q\}$ de $L^2(\mu)$, le maximum de:*

(i) $\sum_{i=1}^{q} \operatorname{Var}(Y_i)$, *ou*

(ii) $\sum_{i,j=1}^{p} \operatorname{Cov}^2(Y_i, Y_j)$, *ou*

(iii) $\det\left[(\operatorname{Cov}(Y_i, Y_j))_{i,j}\right]$,

sont atteints si et seulement si $u_1, \ldots, u_q$ appartiennent à un sous-espace de dimension q de $L^2(\mu)$ engendré par q premiers vecteurs propres de $V = U \circ U^*$; $Y_1, \ldots, Y_q$ appartiennent alors à un sous-espace de $L^2(P)$ engendré par q premiers vecteurs propres de W.

(c) *Minimisation de la norme $\|\cdot\|_p$ de l'opérateur de covariance résiduel.* Pour chaque t de T, nous désignerons par $\hat{X}_t$ la projection orthogonale de X_t sur le sous-espace $G_Y = \operatorname{vect}\{Y_1, \ldots, Y_q\}$ dans $L^2(P)$. Les applications $\hat{X}$ et $X - \hat{X}$ de $\Omega \times T$ dans $\boldsymbol{R}$ telles que

$$\hat{X}\colon (\omega, t) \to \hat{X}_t(\omega) \quad \text{et} \quad X - \hat{X}\colon (\omega, t) \to X_t(\omega) - \hat{X}_t(\omega)$$

sont éléments de l'espace $L^2(P \otimes \mu)$ et définissent deux fonctions aléatoires réelles que nous noterons $\hat{X}$ et $X - \hat{X}$. Nous conviendrons de la

Définition 3.8. Nous appellerons *opérateur de covariance résiduel de la fonction aléatoire X* l'opérateur de covariance de la fonction aléatoire $X - \hat{X}$; nous le noterons V'.

Avec les conventions d'écriture définies en début de ce § 3, on a donc

$$V' = \boldsymbol{E}[(X - \hat{X}) \otimes (X - \hat{X})].$$

Le critère d'optimisation retenu ici est la minimisation de la norme $\|V'\|_p$, le réel $p \geqslant 1$ étant arbitraire et $\{u_i\colon i = 1, \ldots, q\}$ une famille *libre* de $L^2(\mu)$. Commençons par établir le

Lemme 3.9. *On a $V' = V - \hat{V}$, où $\hat{V}$ est l'opérateur de covariance de $\hat{X}$.*

En effet, la linéarité de l'espérance mathématique entraîne:

$$V' = E[(X-\hat{X})\otimes(X-\hat{X})] = E[X\otimes(X-\hat{X})] - E[\hat{X}\otimes(X-\hat{X})],$$

et par suite, les fonctions aléatoires $\hat{X}$ et $X-\hat{X}$ étant orthogonales:

$$V' = E[X\otimes(X-\hat{X})] = E(X\otimes X) - E(X\otimes\hat{X}).$$

De la relation $X = (X-\hat{X})+\hat{X}$, on déduit que

$$E(X\otimes\hat{X}) = E\{[(X-\hat{X})+\hat{X}]\otimes\hat{X}\} = E[(X-\hat{X})\otimes\hat{X}] + E(\hat{X}\otimes\hat{X}) = E(\hat{X}\otimes\hat{X}),$$

ce qui achève la preuve. ∎

L'opérateur $\hat{V} = E(\hat{X}\otimes\hat{X})$ est l'opérateur de covariance de $\hat{X}$, et donc

$$\hat{V} = \hat{U}\circ\hat{U}^*,$$

où $\hat{U}$ est l'opérateur de Hilbert–Schmidt de $L^2(P)$ dans $L^2(\mu)$ défini μ-presque-partout par

$$(\forall Y\in L^2(P))\ (\hat{U}Y)(t) = \langle Y, \hat{X}_t\rangle = \langle Y, AX_t\rangle,$$

où A désigne le projecteur orthogonal de $L^2(P)$ sur $G_Y = \mathrm{vect}\{Y_1, \ldots, Y_q\}$. Par suite, pour tout Y de $L^2(P)$, on a, μ-presque-partout:

$$(\hat{U}Y)(t) = \langle AY, X_t\rangle = [(U\circ A)\,Y](t),$$

d'où l'on tire que $\hat{U} = U\circ A$, et donc que

$$\hat{V} = U\circ A\circ U^*.$$

Il en résulte clairement que, pour chaque réel $p \geqslant 1$, on a

$$(\|V'\|_p)^p = (\|U\circ U^* - U\circ A\circ U^*\|_p)^p = (\|U - U\circ A\|_p)^{2p}.$$

Comme G_Y est un sous-espace de dimension q, $U\circ A$ est un opérateur de rang au plus égal à q. Ainsi, en vertu du théorème 2.7 de [5], $\|U-U\circ A\|_p$ est minimum si et seulement si $U\circ A$ est égal à U_q, développement de Schmidt d'ordre q de U. Or, d'après le corollaire 2.4 de [5], pour que $U\circ A$ soit égal à U_q, il faut et il suffit que A soit le projecteur orthogonal sur un sous-espace de dimension q engendré par q premiers vecteurs propres de $U^*\circ U = W$.

Nous énoncerons donc le

THÉORÈME 3.10. *Soient un réel $p \geqslant 1$ et V' l'opérateur de covariance résiduel de la fonction aléatoire X. Pour que $\|V'\|_p$ soit minimum, il faut et il suffit que $\{Y_1, \ldots, Y_q\}$ soit une famille libre d'un sous-espace de dimension q de $L^2(P)$ engendré par q premiers vecteurs propres de W; si l'opérateur U^* est injectif, alors $u_1, \ldots, u_q$ appartiennent à un sous-espace de $L^2(\mu)$ engendré par q premiers vecteurs propres de V.*

Conclusion. Tout système libre $\{Y_i = U^* u_i\colon i = 1, \ldots, q\}$ d'un sous--espace de dimension q de $L^2(P)$ engendré par q premiers vecteurs propres de W constitue, pour chacun de ces critères, une représentation optimale de la fonction aléatoire X. Les q premières composantes principales de l'A.C.P. de X n'en sont qu'une solution particulière, et cette solution, fournie par chacun des critères forts étudiés précédemment, est la seule qui soit formée de q variables aléatoires non corrélées deux à deux.

3.4. Critère faible d'analyse factorielle linéaire. Le critère que nous pouvons envisager ici est la minimisation de

$$\sum_{i,j=1}^{p} \mathrm{Cov}^2(Y_i, Y_j)$$

lorsque $\{u_i\colon i = 1, \ldots, q\}$ est un système orthonormal de $L^2(\mu)$. De l'étude menée dans le § 3 de [5], on déduit immédiatement que le minimum de cette expression est égal à 0 et atteint si et seulement si $\{u_i\colon i = 1, \ldots, q\}$ est un système orthonormal de vecteurs propres de $A \circ V \circ A$, où A est un projecteur orthogonal *arbitraire* de $L^2(\mu)$ *de rang* q.

TRAVAUX CITÉS

[1] J. N. Darroch, *An extremal property of principal components*, Ann. Math. Statist. 36 (1965), pp. 1579–1582.

[2] J. Dauxois et A. Pousse, *Les analyses factorielles en calcul des probabilités et en statistique: Essai d'étude synthétique*, Thèse d'Etat, Université Paul-Sabatier, Toulouse 1976.

[3] H. Hotteling, *Analysis of a complex of statistical variables into principal components*, J. Educ. Psychol. 24 (1933).

[4] M. Okamoto, *Optimality of principal components*, in: *Multivariate Analysis. II*, P. R. Krishnaiah (Ed.), Academic Press, New York 1969, pp. 673–685.

[5] A. Pousse et J.-J. Téchené, *Quelques propriétés extrémales des valeurs singulières d'un opérateur compact et leurs applications aux analyses factorielles d'une probabilité ou d'une fonction aléatoire: I. Quelques propriétés extrémales des valeurs singulières d'un opérateur compact*, Probab. Math. Statist. 17 (1997), pp. 197–221.

[6] C. R. Rao, *The use and interpretation of principal component analysis in applied research*, Sankyā Ser. A 26 (1964), pp 329–358.

Département de Mathématiques-Recherche
I.P.R.A.–Université de Pau et Pays de l'Adour
Avenue de l'Université — 64000 Pau, France

Received on 7.1.1997