REVUE DE STATISTIQUE APPLIQUÉE

R. BOUMAZA

Analyse en composantes principales de distributions gaussiennes multidimensionnelles

Revue de statistique appliquée, tome 46, n° 2 (1998), p. 5-20

<http://www.numdam.org/item?id=RSA_1998__46_2_5_0>

Rev. Statistique Appliquée, 1998, **XLVI** (2), 5-20

ANALYSE EN COMPOSANTES PRINCIPALES DE DISTRIBUTIONS GAUSSIENNES MULTIDIMENSIONNELLES

R. Boumaza

Département de Mathématiques, Université Mouloud Mammeri, Tizi-Ouzou (Algérie).

RÉSUMÉ

Plusieurs techniques de description de données à trois indices sont proposées dans la littérature, chacune adaptée à un objectif et à un type de données. Nous en proposons une adaptée à des données numériques gaussiennes avec l'objectif de décrire les variables considérées à partir de leur densité de probabilité.

Mots-clés : *données numériques à trois indices, distribution gaussienne, analyse en composantes principales.*

ABSTRACT

Several description techniques for three-way data have been proposed in the literature, each of them being adapted to particular goals and data. We propose a technique which is adapted to Gaussian data with the objective of describing the variables by means of their probability distribution function.

Keywords : *three-way data, Gaussian distribution, principal component analysis.*

1. Introduction

Les données auxquelles on s'intéresse ici sont du type données ternaires, $individus \times variables \times instants$ (three-way data), auxquelles sont consacrés de nombreux travaux dont on peut trouver des synthèses ([Kroonenberg, 1983], [Escoufier, 1985], [Coppi et Bolasco, 1989], [Kiers, 1991]), après les deux articles ([Tucker, 1966], [Escoufier, 1973]) fondateurs de deux approches différentes.

Ces données sont des tableaux $(n_t \times p)$ indexés par t, où à chaque instant t $(t = 1, \ldots, T)$ on dispose d'un échantillon de taille n_t d'un vecteur aléatoire gaussien à p dimensions : ainsi à chaque instant on observe les mêmes variables quantitatives mais pas nécessairement sur les mêmes individus.

L'objectif est de décrire de façon globale ces données pour en apprécier qualitativement l'évolution : le temps (dans le cas où t fait référence au temps)

n'intervenant que comme élément d'interprétation. Ce type de données peut être décrit au moyen de deux analyses distinctes, l'une portant sur les moyennes, l'autre portant sur les matrices de variance. Notre souci en proposant la méthode dite analyse en composantes principales (ACP) de distributions gaussiennes multidimensionnelles est de disposer d'une analyse globale qui prenne en compte aussi bien les moyennes que les variances / covariances. Cette méthode consiste à associer à chaque tableau t un objet qui est une densité de probabilité et d'en faire une ACP à la manière dont procède la méthode STATIS Dual dans sa première étape.

Le cas gaussien a été choisi pour son importance en statistique, cependant le problème de l'extension à d'autres types de distribution demeure.

Après un exposé de l'ACP de distributions gaussiennes multidimensionnelles, on proposera une estimation convergente dans le cas où les paramètres des distributions sont inconnus. Enfin on traitera deux exemples : le premier déjà traité par la méthode STATIS Dual dans [Lavit, 1988], le second étant celui d'une promenade aléatoire pour comparer les représentations de l'ACP classique de T variables et l'ACP de leurs distributions de probabilité.

2. ACP de T distributions gaussiennes multidimensionnelles

2.1. Hypothèses et position du problème

Soient $X_1, \ldots, X_T$ des variables aléatoires de distribution de Gauss non dégénérée à p dimensions, de moyennes $\mu_1, \ldots, \mu_T$ et de matrices de variance $\Sigma_1, \ldots, \Sigma_T$ respectivement; les densités de probabilité $f_1, \ldots, f_T$:

$$f_t(x) = \frac{1}{(2\pi)^{\frac{p}{2}}} \frac{1}{|\Sigma_t|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_t)' \Sigma_t^{-1}(x-\mu_t)}$$

sont de carré intégrable; elles constituent donc un nuage $\mathcal{F}$ dans l'espace de Hilbert $L^2(\mathbb{R}^p)$ noté H.

L'objectif de la méthode proposée étant d'obtenir une représentation approchée de ce nuage, dans une première étape on cherche g_1 dans H, de norme unité, tel que la quantité :

$$\sum_{t=1}^{T} \|P_{g_1}(f_t) - f_t\|_H^2 \tag{1}$$

soit minimum, P_{g_1} désignant le projecteur orthogonal sur le sous espace vectoriel de H engendré par g_1.

Ce critère des moindres carrés est encore équivalent à la maximisation de :

$$\sum_{t=1}^{T} \|P_{g_1}(f_t)\|_H^2 \,. \tag{2}$$

Puis on itère sous contraintes d'orthonormalité : à l'étape k, on cherche g_k dans $L^2(\mathbb{R}^p)$, de norme unité, orthogonal à $g_1, \ldots, g_{k-1}$, tel que la quantité :

$$\sum_{t=1}^{T} \|P_{g_k}(f_t) - f_t\|_H^2 \tag{3}$$

soit minimum.

Cette méthode est bien une ACP particulière dont la solution est immédiate à obtenir. Pour l'écrire, nous adoptons le cadre formel de l'ACP d'opérateur compact proposé par [Dauxois et Pousse, 1976] en adaptant les calculs et résultats aux données précédentes.

2.2. Solution

Soit U l'opérateur de $\mathbb{R}^T$, muni du produit scalaire classique, dans H défini par :

$$Uu = \sum_{t=1}^{T} u_t f_t \ ;$$

en identifiant les espaces de Hilbert séparables H et $\mathbb{R}^T$ à leur dual respectif, l'opérateur adjoint U^* de U est :

$$\begin{array}{rccl} U^* : & H & \longrightarrow & \mathbb{R}^T \\ & g & \longmapsto & U^*g = (< f_1, g >_H, \ldots, < f_T, g >_H) \end{array} \tag{4}$$

car :

$$< Uu, g >_H = \sum_{t=1}^{T} u_t < f_t, g >_H = < u, U^*g >_{\mathbb{R}^T} ,$$

$< , >_{\mathbb{R}^T}$ désignant le produit scalaire usuel dans $\mathbb{R}^T$.

Avec ces notations, on remarque que la quantité (2) à maximiser s'écrit $\|U^*g_1\|_{\mathbb{R}^T}^2$; en effet g_1 étant de norme unité :

$$P_{g_1}(f_t) = < f_t, g_1 >_H \ g_1$$

et donc :

$$\sum_{t=1}^{T} \|P_{g_1}(f_t)\|_H^2 = \sum_{t=1}^{T} < f_t, g_1 >_H^2 = \|U^*g_1\|_{\mathbb{R}^T}^2 .$$

Le problème posé revient donc à chercher g de norme unité qui maximise $\|U^*g\|_{\mathbb{R}^T}^2$ puis itérations sous contraintes d'orthonormalité. Comme :

$$\|U^*g\|_{\mathbb{R}^T}^2 = < U^*g, U^*g >_{\mathbb{R}^T} = < g, U \circ U^*g >_H$$

la solution est obtenue en faisant l'analyse spectrale de l'opérateur V égal à $U \circ U^*$ qui est autoadjoint, positif et de rang fini, ou encore de l'opérateur $U^* \circ U$ noté W qui a les mêmes valeurs propres non nulles que V; de plus si u de $\mathbb{R}^T$ est vecteur propre normé de W associé à la valeur propre non nulle λ alors :

$$g = \frac{Uu}{\sqrt{\lambda}} \tag{5}$$

est un vecteur propre normé de V associé à la même valeur propre λ.

2.3. Ecriture matricielle de W

Si $e_1, \ldots, e_T$ désigne la base canonique de $\mathbb{R}^T$, chaque $U^* f_t$ s'écrira comme combinaison linéaire de ces vecteurs, et dans cette base l'endomorphisme W de $\mathbb{R}^T$ a pour matrice, notée aussi W :

$$W = (< f_s, f_t >_H)_{(s,t)\in T\times T} \tag{6}$$

car :

$$We_t = U^* \circ Ue_t = U^* f_t = \sum_{s=1}^{T} < f_s, f_t >_H e_s .$$

Le terme général $< f_s, f_t >_H$ de la matrice W est égal à :

$$\frac{1}{(2\pi)^p} \frac{1}{|\Sigma_s|^{\frac{1}{2}} |\Sigma_t|^{\frac{1}{2}}} \int_{\mathbb{R}^p} e^{-\frac{1}{2}[(x-\mu_s)'\Sigma_s^{-1}(x-\mu_s)+(x-\mu_t)'\Sigma_t^{-1}(x-\mu_t)]}\, dx ;$$

le calcul de cette intégrale (*Cf.* Annexe) aboutit à :

$$< f_s, f_t >_H = \frac{1}{(2\pi)^{\frac{p}{2}}} \frac{1}{|\Sigma_s + \Sigma_t|^{\frac{1}{2}}} e^{-\frac{1}{4}\|\mu_s-\mu_t\|^2_{st}} \tag{7}$$

où $\|\mu_s - \mu_t\|^2_{st}$ désigne la norme carrée de $(\mu_s - \mu_t)$ pour la métrique de matrice $(\frac{\Sigma_s + \Sigma_t}{2})^{-1}$.

Les éléments propres normalisés $(\lambda_1, u_1), \ldots, (\lambda_K, u_K)$ de W correspondant aux valeurs propres non nulles et rangées dans l'ordre décroissant permettent d'obtenir en utilisant l'expression (5), les vecteurs propres $g_1, \ldots, g_K$ de V.

Ces fonctions de $L^2(\mathbb{R}^p)$ sont des combinaisons linéaires des densités de probabilité gaussiennes f_t mais ne sont pas des densités de probabilité sauf éventuellement g_1. En effet deux densités de probabilité donc positives ne peuvent être orthogonales pour le produit scalaire de $L^2(\mathbb{R}^p)$.

Quant à l'éventualité que g_1 soit une densité, elle pourrait se produire si par exemple la somme des composantes de u_1 qui sont nécessairement de même signe,

vaut $\sqrt{\lambda_1}$; en effet la matrice W a tous ses éléments positifs, son premier vecteur propre u_1 aura donc des composantes toutes de même signe (Théorème de Frobenius) que l'on choisit positif, et g_1 calculé par (5) apparaît comme une combinaison convexe de densités de probabilité.

2.4. Reconstitution des densités de probabilité

Pour tout t la densité de probabilité f_t peut se décomposer suivant le système $(g_k)_{k=1,\ldots,K}$:

$$f_t = \sum_{k=1}^{K} < f_t, g_k >_H \; g_k \tag{8}$$

La coordonnée de chaque f_t suivant g_k étant $< f_t, g_k >_H$, ces coordonnées sont donc les composantes du vecteur $U^* g_k$ (4) égal à $\sqrt{\lambda_k} u_k$ (5) et donc (8) devient :

$$f_t = \sum_{k=1}^{K} \sqrt{\lambda_k}\, u_{kt}\, g_k \tag{9}$$

où u_{kt} désigne la t^{e} composante du vecteur u_k.

On peut obtenir une représentation approchée du nuage $\mathcal{F}$ dans un sous-espace de dimension réduite en tronquant la décomposition précédente et calculer les aides à l'interprétation ([Volle, 1981]) :

– la qualité globale de l'ACP se mesure par la somme des proportions d'inertie expliquée par les premiers axes retenus, chaque axe k expliquant une proportion égale à $\dfrac{\lambda_k}{tr(W)}$ où $tr(W)$ désigne la trace de la matrice W et vaut $\displaystyle\sum_{k=1}^{K} \lambda_k$ qui est l'inertie totale du nuage par rapport à l'origine;

– la qualité de représentation de f_t suivant g_k se mesure par le rapport $\dfrac{\|P_{g_k}(f_t)\|_H^2}{\|f_t\|_H^2}$, soit $\lambda_k\,(u_{kt})^2\,(2\sqrt{\pi})^p\,|\Sigma_t|^{\frac{1}{2}}$, puisque d'après (7), $\|f_t\|_H^2$ est égale à $(2\sqrt{\pi})^{-p}\,|\Sigma_t|^{-\frac{1}{2}}$;

– la valeur propre λ_k mesurant l'inertie (par rapport à l'origine) du nuage obtenu par projection de $\mathcal{F}$ sur l'axe k, on mesure l'importance de chaque f_t dans la détermination du facteur g_k par sa contribution relative à cette inertie qui vaut $(u_{kt})^2$.

2.5. ACP des fonctions caractéristiques

Plutôt que de représenter la variable X_t par sa densité de probabilité f_t dans $L^2(\mathbb{R}^p)$, il aurait été possible de la représenter par sa fonction caractéristique φ_t dans l'espace de Hilbert $L^2_{\mathbb{C}}(\mathbb{R}^p)$. Cette ACP conduit aux mêmes représentations. En effet le théorème de Plancherel [Vinograd, 1987] montre que :

$$< \varphi_s, \varphi_t > = (2\pi)^p < f_s, f_t >_H ; \tag{10}$$

ainsi la matrice W correspondante à cette ACP est égale à celle définie en (6) au coefficient $(2\pi)^p$ près.

3. Distance entre densités de probabilité

La distance carrée entre les densités de probabilité f_s et f_t est obtenue à partir de $\|f_s - f_t\|_H^2$ et vaut :

$$\frac{1}{(2\sqrt{\pi})^p}\left[\frac{1}{|\Sigma_s|^{\frac{1}{2}}}+\frac{1}{|\Sigma_t|^{\frac{1}{2}}}-\frac{2}{\left|\frac{\Sigma_s+\Sigma_t}{2}\right|^{\frac{1}{2}}}e^{-\frac{1}{4}\|\mu_s-\mu_t\|_{st}^2}\right].$$

Cette fonction qu'on ne peut représenter dans le cas général, car dépendant des trois paramètres Σ_s, Σ_t et $(\mu_s - \mu_t)$, sera précisée dans le cas particulier où $\Sigma_t = c\Sigma_s$ $(c > 0)$. Elle dépend dans ce cas du paramètre c et de d la distance entre μ_s et μ_t pour la métrique de matrice Σ_s^{-1} :

$$\|f_s - f_t\|_H^2 = \frac{1}{(2\sqrt{\pi})^p}\frac{1}{|\Sigma_s|^{\frac{1}{2}}}\underbrace{\left[1+\frac{1}{c^{\frac{p}{2}}}-\frac{2}{\left(\frac{1+c}{2}\right)^{\frac{p}{2}}}e^{-\frac{1}{2}\frac{d^2}{1+c}}\right]}_{h(c,d)}$$

Pour $p = 4$, la fonction h a sur le pavé $[0.5\,,\,2] \times [0\,,\,4.5]$ l'allure de la figure ([Fig. 1]) qui donne aussi les coupes à $(d = 0)$ et $(c = 1)$ respectivement.

Ainsi la distance entre les densités de probabilité apparaît comme une synthèse de deux différences celle entre les moyennes et celle entre les matrices de variance.

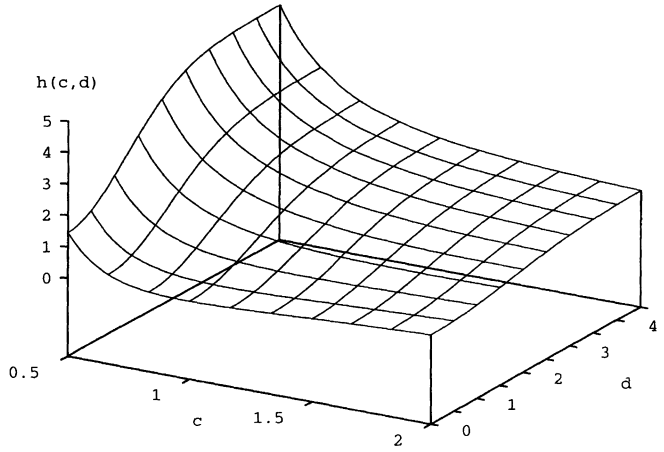
4. ACP normée et ACP centrée

La présentation de l'ACP précédente a considéré le nuage des densités de probabilité sans aucune transformation.

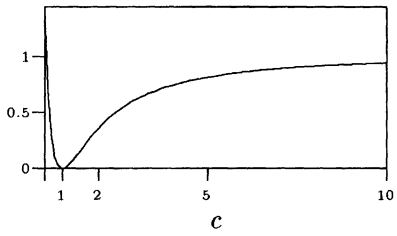
4.1. ACP normée

En effet il est possible de réduire ce nuage en normant à 1 au sens de la norme L^2 les fonctions f_t , c'est-à-dire en considérant les fonctions $\overline{f_t}$ (ces fonctions ne sont plus en général des densités de probabilité) :

$$\overline{f_t} = \frac{f_t}{\|f_t\|_H}$$

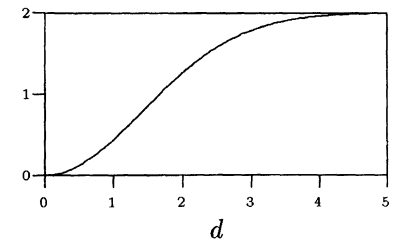


$h(c,0)$



$d = 0$
Les moyennes sont égales.

$h(1,d)$



$c = 1$
Les variances sont égales.

FIGURE 1
Graphe de la fonction h

soit :

$$\overline{f_t}(x) = \frac{1}{\pi^{\frac{p}{4}}} \frac{1}{|\Sigma_t|^{\frac{1}{4}}} e^{-\frac{1}{2}(x-\mu_t)' \Sigma_t^{-1}(x-\mu_t)} .$$

L'ACP dite normée conduit à diagonaliser la matrice $\overline{W}$ de terme général $\overline{W}_{st}$ égal à :

$$\overline{W}_{st} = < \overline{f_s}, \overline{f_t} >_H = 2^{\frac{p}{2}} \frac{|\Sigma_s|^{\frac{1}{4}} |\Sigma_t|^{\frac{1}{4}}}{|\Sigma_s + \Sigma_t|^{\frac{1}{2}}} e^{-\frac{1}{4}\|\mu_s - \mu_t\|^2_{st}} ;$$

Cette normalisation conserve dans H les angles entre les densités mais déforme les distances entre elles.

4.2. ACP centrée

Une autre transformation possible qui elle respecte les distances est le centrage du nuage $\mathcal{F}$ des densités f_t en opérant une translation du nuage amenant son centre de gravité $\frac{1}{T}\sum_{t=1}^{T} f_t$ noté f_G, sur l'origine de l'espace H. Ces nouvelles fonctions f_t° sont égales à :

$$f_t^\circ = f_t - f_G .$$

La matrice W° à diagonaliser pour obtenir l'ACP centrée est obtenue à partir de W; en effet le terme général W_{st}° de W° étant égal à $< f_s^\circ, f_t^\circ >_H$ son calcul est immédiat.

5. Estimation et convergence

En pratique on ne connaît pas les paramètres μ_t et Σ_t de la distribution (gaussienne) de la variable aléatoire X_t $(t = 1, \ldots, T)$; si pour tout t on dispose d'un n_t-échantillon $X_{t1}, \ldots, X_{tn_t}$ de X_t, on estime ces paramètres par $\overline{X_t}$ et S_t les estimateurs du maximum de vraisemblance respectivement de μ_t et Σ_t :

$$\overline{X_t} = \frac{1}{n_t} \sum_{i=1}^{n_t} X_{ti} ,$$

$$S_t = \frac{1}{n_t} \sum_{i=1}^{n_t} (X_{ti} - \overline{X_t})(X_{ti} - \overline{X_t})' .$$

On note $W^{(n)}$ où n désigne la plus petite taille d'échantillon $inf\, n_t$, l'estimateur de W obtenu en remplaçant, dans le terme général W_{ts} (7) de W, les paramètres

μ_t , μ_s , Σ_s et Σ_t par leur estimateur respectif. La convergence presque sûre de ces estimateurs lorsque n croît indéfiniment assure la convergence presque sûre de chaque terme $W_{ts}^{(n)}$ de la matrice $W^{(n)}$ vers le terme respectif W_{ts} de W et donc la convergence uniforme presque sûre de $W^{(n)}$ vers W.

Ainsi l'ACP des densités estimées obtenue par l'analyse spectrale de $W^{(n)}$ est convergente ([Dauxois et Pousse, 1976], [Romain, 1979]) et la représentation de ces densités est une bonne approximation de la représentation des densités parentes.

6. Comparaisons avec STATIS Dual

6.1. ACP des distributions gaussiennes et STATIS Dual

Si les données dont on dispose sont telles qu'elles sont présentées en introduction, la méthode STATIS Dual ([L'hermier des Plantes, 1976], [Glaçon, 1981], [Lavit, 1988]) peut les décrire en utilisant soit les matrices de variance (Σ_t), soit les matrices de corrélation (R_t); cependant cette description ne tient pas compte des moyennes (μ_t) des variables observées, dans le calcul du compromis à l'étape de l'interstructure. La description des moyennes nécessite une étude séparée. La méthode proposée reste dans l'esprit de la première étape de STATIS Dual, la différence réside dans les objets associés à chaque tableau : ici l'objet associé est f_t la densité de probabilité de la distribution $\mathcal{N}(\mu_t, \Sigma_t)$ ce qui conduit à diagonaliser la matrice W de terme général (7), tandis que dans STATIS Dual l'objet associé est Σ_t (ou R_t) ce qui conduit à diagonaliser la matrice D de terme général :

$$D_{st} = tr(\Sigma_s \Sigma_t) . \tag{11}$$

Si on dispose d'un programme informatique de mise en oeuvre de STATIS Dual, la réalisation de l'ACP de distributions gaussiennes multidimensionnelles ne nécessite que l'adjonction d'un module permettant de calculer l'inverse et le déterminant d'une matrice symétrique définie positive afin d'évaluer la matrice à diagonaliser de terme général W_{ts} (7), le coefficient $\dfrac{1}{(2\pi)^{\frac{p}{2}}}$ étant bien évidemment inutile; le reste du programme reste sans changement.

Si on utilise le premier facteur principal u_1 de l'ACP des densités de probabilité pour obtenir la densité de probabilité compromis :

$$f_c = \sum_{t=1}^{T} \alpha_t f_t ,$$

la moyenne μ_c et la matrice de variance Σ_c du compromis sont respectivement :

$$\mu_c = \sum_{t=1}^{T} \alpha_t \mu_t$$
$$\Sigma_c = \sum_{t=1}^{T} \alpha_t (\Sigma_t + (\mu_t - \mu_c)(\mu_t - \mu_c)');$$

on rappelle que $\sum_{t=1}^{T} \alpha_t$ vaut 1 car les T composantes $u_{11}, \ldots, u_{1T}$ du vecteur u_1 sont toutes de même signe et on prend α égal à $\frac{u_1}{\sum_{t=1}^{T} u_{1t}}$.

6.2. ACP de développements en série de fonctions caractéristiques et STATIS Dual

La fonction caractéristique φ_t du vecteur aléatoire X_t, dont les composantes seront notées $X_{t1}, \ldots, X_{tp}$ se décompose au voisinage de zéro (Formule de Taylor à l'ordre q) comme suit :

$$\varphi_t(z) = 1 + \sum_{r=1}^{q} \frac{i^r}{r!} \sum_{j_1=1}^{p} \cdots \sum_{j_r=1}^{p} z_{j_1} \ldots z_{j_r} E[X_{tj_1} \ldots X_{tj_r}] + o(\|z\|^q)$$

Pour $q = 2$ et si le vecteur X_t est centré ($\mu_t = 0$), cela donne :

$$\varphi_t(z) = 1 - \underbrace{\frac{1}{2} \sum_{j=1}^{p} \sum_{k=1}^{p} z_j z_k E[X_{tj}\, X_{tk}]} + o(\|z\|^2)$$

La quantité soulignée par l'accolade se développe en :

$$\begin{array}{ccccccc} & z_1^2\, E[X_{t1}^2] & + & z_1 z_2\, E[X_{t1} X_{t2}] & + \ldots + & z_1 z_p\, E[X_{t1} X_{tp}] \\ + & z_2 z_1\, E[X_{t2} X_{t1}] & + & z_2^2\, E[X_{t2}^2] & + \ldots + & z_2 z_p\, E[X_{t2} X_{tp}] \\ & \ldots & & & & \\ + & z_p z_1\, E[X_{tp} X_{t1}] & + & z_p z_2\, E[X_{tp} X_{t2}] & + \ldots + & z_p^2\, E[X_{tp}^2] \end{array}$$

A tout développement en série d'ordre 2 au voisinage de zéro d'une fonction caractéristique centrée on peut associer un vecteur Φ_t de $\mathbb{R}^{p^2}$ dont les composantes sont les coefficients des polynômes du second degré

$$z_j z_k\ ,\ \ 1 \leqslant j, k\ \leqslant p\ .$$

(les polynômes sont répétés uniquement pour faciliter la numérotation et les écritures).

Si $\mathbb{R}^{p^2}$ est muni du produit scalaire $< , >_{\mathbb{R}^{p^2}}$ de matrice identité alors :

$$< \Phi_t , \Phi_s >_{\mathbb{R}^{p^2}} = \frac{1}{4} \sum_{j=1}^{p} \sum_{k=1}^{p} E[X_{sj} X_{sk}] E[X_{tj} X_{tk}]$$

ce qui est le terme général de la matrice à diagonaliser si on procède à l'ACP des T «individus» $\Phi_1, \ldots, \Phi_T$ de $\mathbb{R}^{p^2}$. Ce terme est encore égal au quart de la trace (11) du produit des matrices de covariance Σ_s et Σ_t :

$$< \Phi_s , \Phi_t >_{\mathbb{R}^{p^2}} = \frac{1}{4} tr(\Sigma_s \Sigma_t) = \frac{1}{4} D_{st} .$$

Ainsi si les vecteurs $X_1, \ldots, X_T$ sont centrés STATIS Dual est équivalente à l'ACP des développements en série d'ordre 2 de leur fonction caractéristique dans $\mathbb{R}^{p^2}$ muni de la métrique identité; STATIS Dual serait donc dans le cas centré une forme d'approximation d'une ACP de fonctions caractéristiques qui peut être rapprochée de celle définie au paragraphe (2.5), équivalente à l'ACP des densités lorsqu'elles sont de carré intégrable, mais qui n'est cependant pas la même.

7. Exemples

7.1. Exemple [Lavit, 1988]

Nous reprenons les données publiées et analysées de façon détaillée dans [Lavit, 1988] au moyen, entre autres, de la méthode STATIS Dual appliquée aux matrices de corrélation. Les données décrivent le suivi de la morphologie (poids, taille, buste, périmètres crânien, thoracique, du bras gauche, du mollet gauche, largeur du bassin) de 30 filles de 4 à 15 ans.

Les premiers plans principaux obtenus par l'application de trois analyses différentes sont donnés ci-après. Le premier plan principal de l'ACP normée des densités de probabilité ([Fig. 3]) donne une représentation des densités qui dénote une régularité dans l'évolution morphologique; cette évolution apparaît comme une synthèse des évolutions qui peuvent être observées aussi bien au niveau des moyennes ([Fig. 2]) que des matrices de variance ([Fig. 4]).

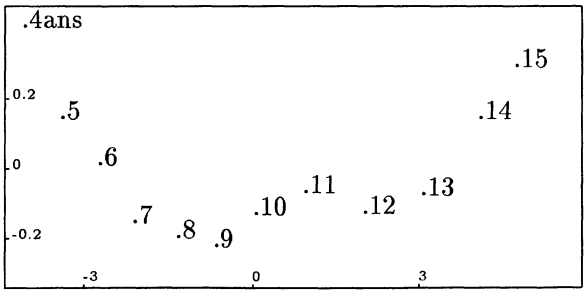


FIGURE 2
ACP centrée et normée des moyennes par âge.

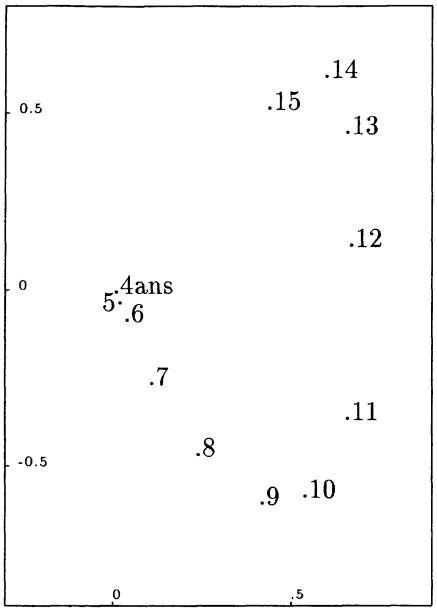


FIGURE 3
ACP normée des densités.

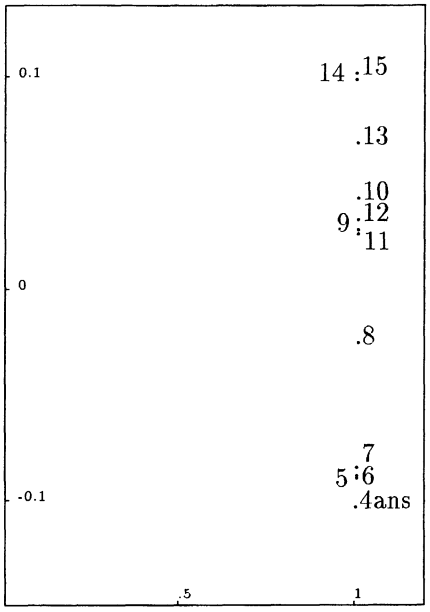


FIGURE 4
STATIS Dual sur matrices de variance.

7.2. Exemple d'une promenade aléatoire

Soit (Y_t) une promenade aléatoire réelle construite à partir du bruit blanc (ε_t) de variance σ^2 :

$$\forall t \geqslant 0, \quad Y_t = \sum_{j=0}^{t} \varepsilon_j \; ;$$

on suppose de plus que le bruit blanc est gaussien; ainsi les variables aléatoires $Y_t\,(t = 0, \ldots, T)$ sont centrées et pour tout t la variance de Y_t est $(t+1)\sigma^2$. Dans cet exemple, on a pris le cas (arbitraire) T égal à 25; cependant la configuration reste la même pour T allant jusqu'à la cinquantaine. L'ACP classique de ces variables qui conduit à diagonaliser la matrice de corrélation dont le terme général est :

$$R_{st} = \frac{Min(s,t) + 1}{\sqrt{s+1}\,\sqrt{t+1}},$$

donne la représentation [Fig. 5] sur le premier cercle des corrélations.

L'ACP normée des densités de probabilité est obtenue en diagonalisant la matrice de terme général :

$$\overline{W}_{st} = \sqrt{\frac{2\sqrt{(t+1)(s+1)}}{s+t+2}} \; ;$$

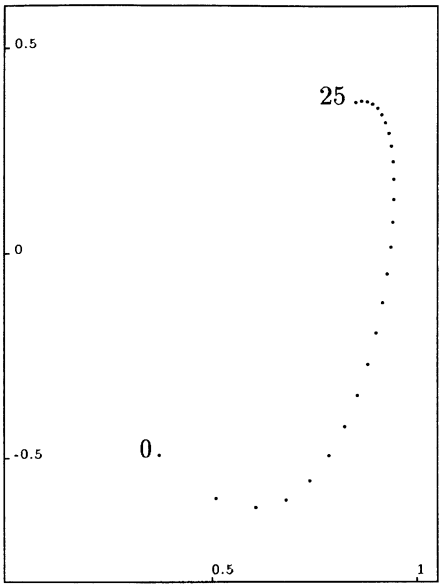


FIGURE 5
ACP classique des Y_t.

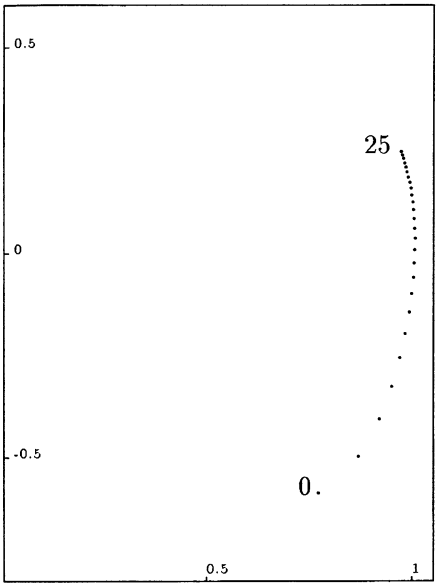


FIGURE 6
ACP des densités des Y_t.

la représentation des densités des variables Y_t sur le premier plan principal est donnée par [Fig. 6].

Dans les deux cas l'évolution temporelle des variables Y_t est bien marquée; on peut même avancer que la régularité est meilleure pour l'ACP normée des densités.

7.3. Remarques

Deux autres exemples ont été traités. Le premier a porté sur des données archéologiques qui ont été à l'origine de cette méthode mais qui n'a pas été exposé car non concluant; en effet ces données d'une part n'obéissent pas toutes à la loi de Gauss — ce qui pousse à étendre la méthode à des distributions non gaussiennes —, d'autre part l'évolution temporelle qu'on aurait souhaité déceler n'est partiellement évidente que sur les moyennes des variables mais pas du tout sur les variances et covariances, ni les corrélations.

Le deuxième exemple traité est celui de cas — construits à la main — de processus linéaires multivariés ([Gourieroux et Montfort, 1990]) dont la promenade aléatoire est un cas particulier; les représentations obtenues dénotent une évolution temporelle régulière semblable à celle de la promenade aléatoire.

Annexe

Calcul du produit scalaire de densités gaussiennes

On a remarqué (10) qu'on peut calculer indifféremment soit $< f_s, f_t >_H$ soit $< \varphi_s, \varphi_t >$ où φ_s et φ_t sont les fonctions caractéristiques associées à f_s et f_t

respectivement et :

$$< \varphi_s, \varphi_t > = \int_{\mathbb{R}^p} \varphi_s(u)\, \bar{\varphi}_t(u)\, du \;=\; \int_{\mathbb{R}^p} e^{iu'\mu_s - \frac{1}{2}u'\Sigma_s u}\, e^{-iu'\mu_t - \frac{1}{2}u'\Sigma_t u}\, du\,;$$

soit encore :

$$\int_{\mathbb{R}^p} e^{iu'(\mu_s-\mu_t) - u'\left(\frac{\Sigma_s+\Sigma_t}{2}\right)u}\, du \;=\; \int_{\mathbb{R}^p} e^{M(u)}\, du.$$

La matrice Σ égale à $\left(\dfrac{\Sigma_s+\Sigma_t}{2}\right)$ étant définie positive, on fait le changement de variables : $v = \Sigma^{\frac{1}{2}}u$. En notant a le vecteur $\Sigma^{-\frac{1}{2}}(\mu_s - \mu_t)$ de composantes $\alpha_1, \ldots, \alpha_p$, la quantité $M(u)$ s'écrit $(iv'a - v'v)$ c'est-à-dire $\sum_{j=1}^{p}(i\alpha_j v_j - v_j^2)$.

L'intégrale précédente devient :

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{j=1}^{p} e^{-(v_j^2 - i\alpha_j v_j)}\; |\Sigma|^{-\frac{1}{2}}\, dv_1 \ldots dv_p.$$

On a :

$$\forall \alpha \in \mathbb{R}, \quad \int_{-\infty}^{\infty} e^{-(x^2 - i\alpha x)} dx \;= \sqrt{\pi}\, e^{-\frac{\alpha^2}{4}};$$

ce résultat découlant de :

$$\int_{-\infty}^{\infty} e^{-(x-i\alpha)^2} dx \;= \sqrt{\pi}$$

qui s'obtient en intégrant la fonction holomorphe $e^{-(z-i\alpha)^2}$ sur le contour rectangulaire du plan complexe délimité par les points A, B, C et D de coordonnées respectives $(-R, 0)$, $(-R, i\alpha)$, $(R, i\alpha)$, $(R, 0)$ et en faisant tendre R vers l'infini.

Ainsi :

$$< \varphi_s, \varphi_t > = \; |\Sigma|^{-\frac{1}{2}} \prod_{j=1}^{p} (\sqrt{\pi}\, e^{-\frac{\alpha_j^2}{4}}) \;=\; |\Sigma|^{-\frac{1}{2}}\, \pi^{\frac{p}{2}}\, e^{-\frac{\sum_{j=1}^{p}\alpha_j^2}{4}},$$

et en remplaçant $\sum_{j=1}^{p} \alpha_j^2$ par sa valeur $(\mu_s - \mu_t)' \left(\frac{\Sigma_s + \Sigma_t}{2} \right)^{-1} (\mu_s - \mu_t)$, noté $\|\mu_s - \mu_t\|_{ts}^2$, il vient :

$$< \varphi_s, \varphi_t > = (2\pi)^{\frac{p}{2}} \frac{1}{|\Sigma_s + \Sigma_t|^{\frac{1}{2}}} e^{-\frac{1}{4}\|\mu_s - \mu_t\|_{st}^2}$$

d'où, compte tenu de (10) :

$$< f_t, f_s >_H = \frac{1}{(2\pi)^{\frac{p}{2}}} \frac{1}{|\Sigma_s + \Sigma_t|^{\frac{1}{2}}} e^{-\frac{1}{4}\|\mu_s - \mu_t\|_{st}^2}$$

Remerciements

Nous remercions vivement Y. Escoufier, B. Ycart et un lecteur anonyme pour les corrections et remarques qu'ils ont apportées.

Références bibliographiques

COPPI R., BOLASCO S., (1989), *Multiway data analysis*, North-Holland, Amsterdam. Proceedings of the International Meeting on the Analysis of Multiway Data Matrices, Rome, March 28-30, 1988.

DAUXOIS J., POUSSE A., (1976), *Les analyses factorielles en calcul des probabilités et en statistique : essai d'étude synthétique*, Thèse d'état, Université Paul Sabatier, Toulouse, France.

ESCOUFIER Y., (1973), Le traitement des variables vectorielles, *Biometrics*, 29, 751-760.

ESCOUFIER Y., (1985), Objectifs et procédures de l'analyse conjointe de plusieurs tableaux de données, *Statistique et analyse de données*, 10, 1-10.

GLAÇON F., (1981), *Analyse conjointe de plusieurs matrices de données*, Thèse de 3e cycle, Université Joseph Fourier, Grenoble, France.

GOURIEROUX C., MONFORT A., (1990), *Séries temporelles et modèles dynamiques*, Economica, Paris.

KIERS H. A.L., (1991), Hierarchical relations among three-way methods, *Psychometrika*, 56, 3, 449-470.

KROONENBERG P.M., (1983), *Three-mode principal component analysis. Theory and applications*, DSWO Press, Leiden, Reprint 1989.

LAVIT C., (1988), *Analyse conjointe de tableaux quantitatifs*, Masson, Paris.

L'HERMIER DES PLANTES H., (1976), *Structuration des tableaux à trois indices de la statistique*, Thèse de 3e cycle, Université Montpellier II, Montpellier, France,

ROMAIN Y., (1979), *Etude asymptotique des approximations par échantillonnage de l'analyse en composantes principales d'une fonction aléatoire. Quelques applications*, Thèse de 3e cycle, Université Paul Sabatier, Toulouse, France.

TUCKER L.R., (1966), Some mathematical notes on three-mode factor analysis, *Psychometrika*, 31, 279-311.

VINOGRAD I.M., (1987), *Encyclopaedia of mathematics*, Reidel, 163-164.

VOLLE M., (1981), *Analyse des données*, Economica, Paris,