REVUE DE STATISTIQUE APPLIQUÉE

H. CAUSSINUS
S. HAKAM
A. RUIZ-GAZEN

Projections révélatrices contrôlées: groupements et structures diverses

Revue de statistique appliquée, tome 51, n° 1 (2003), p. 37-58
<http://www.numdam.org/item?id=RSA_2003__51_1_37_0>

Rev. Statistique Appliquée, 2003, **LI** (1), 37-58

PROJECTIONS RÉVÉLATRICES CONTRÔLÉES : GROUPEMENTS ET STRUCTURES DIVERSES

H. CAUSSINUS*, S. HAKAM*(1), A. RUIZ-GAZEN*(2)

* *Laboratoire de Statistique et Probabilités, U.M.R. - C.N.R.S. C5583, Université Paul Sabatier, 118, route de Narbonne, 31062 Toulouse cedex 4.*

(1) *Département de Mathématique et Informatique, Faculté des Sciences, B.P. 1014, Ave. Ibn Battouta, Rabat, Maroc.*

(2) *Gremaq, U.M.R. C.N.R.S. C5604, Université Toulouse I, 21, allée de Brienne, 31000 Toulouse.*

RÉSUMÉ

Les méthodes de projections révélatrices recherchent des projections exhibant au mieux d'éventuelles structures particulières de la distribution d'individus caractérisés par p variables numériques. Parmi les diverses techniques qui ont été proposées figurent des analyses en composantes principales généralisées grâce à des métriques convenables sur l'espace des individus. Après avoir étudié les individus atypiques au moyen d'une d'entre elles dans un précédent article, nous examinons ici des analyses fondées sur une autre métrique, susceptibles de mettre en évidence des structures plus complexes. Les développements donnés vont essentiellement dans deux directions. On montre d'abord comment contrôler la signification statistique des projections obtenues en utilisant des techniques inférentielles fondées sur des résultats théoriques récents (Hakam, 2002) donnant la loi asymptotique des matrices aléatoires considérées; on voit ensuite comment, dans le cas particulier d'une structuration en groupes homogènes, nos propositions peuvent fournir un préalable utile à la classification. Plusieurs exemples sont examinés pour illustrer ces divers points tout en montrant comment se présente la mise en œuvre concrète des analyses proposées.

Mots-clés : *Analyse en composantes principales, Classification, Projections révélatrices, Techniques graphiques.*

ABSTRACT

Projection pursuit aims to find low-dimensional projections of units characterised by p real variables. Various techniques have been proposed. One of them consists in generalised principal components analyses with suitable choices of the metric on the units space. We have considered the problem of outliers in a previous paper; we are now interested in the display of more complex structures by means of another choice of the metric. The developments therein are mainly in two directions. On the one hand, we show how to assess the statistical significance of the obtained projections by using recent theoretical results (Hakam, 2002) concerning the asymptotic distribution of the involved random matrices. On the other hand, in the special case where the data are divided into homogeneous groups, we show how our proposals may provide

a useful preliminary step to clustering techniques. Several examples are given to illustrate these various points as well as to show the practical implementation of the proposed analyses.

Keywords : *Clustering, Graphical displays, Principal component analysis, Projection pursuit.*

1. Introduction

Dans un précédent article de cette Revue (Caussinus, Hakam et Ruiz-Gazen, 2002), nous avons rappelé le principe des méthodes de projections révélatrices à partir d'Analyses en Composantes Principales (ACP) généralisées et nous avons illustré notre propos avec la recherche d'individus atypiques. C'est une première question d'intérêt évident dans l'analyse d'un tableau individus $\times$ variables, mais la recherche d'autres types de structures est aussi envisageable par une méthode analogue. Un exemple important est la recherche de groupes, un peu comme le ferait une technique de classification, mais avec le souci de visualiser les groupes en les séparant au mieux et en les positionnant les uns par rapport aux autres. On peut noter que l'analyse factorielle discriminante est justement conçue pour séparer des groupes de façon optimale et les visualiser, mais il s'agit de groupes connus à l'avance alors que nous nous proposons au contraire de les rechercher : d'une certaine façon, on peut dire que nous souhaitons faire de l'analyse factorielle discriminante sans connaissance *a priori* de l'affectation des individus aux groupes et sans connaître le nombre ni la position des groupes (à supposer qu'il y en ait...). Plus généralement, les techniques que nous allons envisager cherchent à détecter une « structure » affectée par un « bruit » et visualiser celle-ci en projetant les individus sur le sous-espace le plus approprié. Comme dans Caussinus, Hakam et Ruiz-Gazen (2002), nous reprenons des techniques précédemment présentées par deux des auteurs (Caussinus et Ruiz-Gazen, 1993, 1995) et nous complétons l'analyse par une méthode objective de choix de la dimension utile, ce qui permet de contrôler la signification statistique des graphiques obtenus et d'éviter ainsi sans doute bon nombre d'interprétations hasardeuses; cette démarche repose sur des résultats théoriques récents (Hakam, 2002).

Les principes généraux de la méthode sont exposés dans le paragraphe 2 et les problèmes de dimension dans le paragraphe 3. Le paragraphe 4 approfondit les connexions avec les problèmes de classification automatique : après avoir rappelé quelques démarches précédentes de nature voisine, on montre comment nos méthodes peuvent s'articuler avec des techniques de classification et quelle utilité elles peuvent prendre dans cette optique. Plusieurs exemples sont présentés dans le paragraphe 5 afin de préciser l'utilisation pratique des méthodes proposées et d'illustrer leurs diverses potentialités.

2. Une ACP généralisée pour la recherche de structure : modèle et principe

2.1. Méthode de base

A chaque individu i ($i = 1, \ldots, n$) est associé un vecteur X_i de $\mathbb{R}^p$. On considère les deux matrices $p \times p$ suivantes :

$$V_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)(X_i - \bar{X}_n)'$$

$$T_n(\beta) = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \exp\left(-\frac{\beta}{2} \|X_i - X_j\|^2_{V_n^{-1}}\right) (X_i - X_j)(X_i - X_j)'}{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \exp\left(-\frac{\beta}{2} \|X_i - X_j\|^2_{V_n^{-1}}\right)} \quad (1)$$

avec $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ et $\|X\|^2_M = X' M X$.

V_n est la matrice des variances et covariances empiriques usuelles, tandis que $T_n(\beta)$ peut s'interpréter comme une matrice de variances et covariances « locale » (à un facteur près). En effet, sans le terme exponentiel de pondération (ou, si l'on veut, avec $\beta = 0$), $T_n(\beta)$ est égale à $\frac{2n}{n-1} V_n$, mais la pondération introduite vise à diminuer l'importance des couples d'individus éloignés au profit des individus proches. Dans le cas de groupes, par exemple, $T_n(\beta)$ jouera le rôle d'une matrice de variances et covariances intragroupe. On comprend alors pourquoi, par analogie avec l'analyse factorielle discriminante, il a été proposé de réaliser une ACP de V_n avec, pour métrique sur l'espace $\mathbb{R}^p$ des individus, la matrice inverse de $T_n(\beta)$. On diagonalise donc $V_n T_n^{-1}(\beta)$ et les individus X_i sont projetés $T_n^{-1}(\beta)$-orthogonalement sur les sous-espaces principaux.

Quelques propriétés théoriques de cette méthode, d'abord proposée de façon heuristique dans Caussinus et Ruiz (1990), ont été étudiées par Caussinus et Ruiz-Gazen (1993, 1995). Ces auteurs considèrent que les X_i sont des vecteurs aléatoires indépendants, de même loi, cette loi étant un mélange de lois normales. De façon générale, on peut écrire cette loi sous la forme intégrale

$$\int \mathcal{N}_p(x, W)\, dP(x) \quad (2)$$

où P est une probabilité concentrée sur un sous espace (ou une variété linéaire) de $\mathbb{R}^p$ de dimension q. Dans le cas où la probabilité P est discrète, chargeant $q + 1$ points μ_j de $\mathbb{R}^p$ avec les probabilités p_j, on est en présence de $q + 1$ groupes de lois respectives $\mathcal{N}(\mu_j, W)$ en proportions respectives p_j ($j = 1, \ldots, q + 1$); les centres des groupes sont évidemment contenus dans une variété linéaire de dimension q. Mais P peut très bien être une loi plus générale, par exemple répartie sur une courbe d'un sous-espace à q dimensions. La restriction la plus importante de la formulation ci-dessus est clairement le fait que les diverses lois du mélange ont la même matrice de variances et covariances.

La méthode proposée est invariante par transformations affines (voir chapitre 8 de l'ouvrage édité par Droesbecke, Fichet et Tassi, 1992), une propriété importante quand il s'agit de faire ressortir la structure d'un nuage de points, laquelle n'a justement de sens qu'à une affinité près. Un avantage technique est qu'on peut, sans perte de généralité, supposer P centrée (en pratique les X_i sont centrés) et admettant une variance telle que la matrice des variances et covariances V de X_i soit égale à l'identité I_p.

Pour faire ressortir au mieux la structure (c'est-à-dire la loi P par opposition aux lois normales qui représentent le bruit), il faut que la méthode conduise à projeter les individus sur le sous-espace contenant P (contenant les $q+1$ vecteurs μ_j dans le cas discret). Caussinus et Ruiz-Gazen (1993, 1995) montrent que c'est le cas dans un certain nombre de situations dès que n est assez grand; la valeur optimale de β dépend de la situation réelle (inconnue), mais une valeur voisine de 2 est en général la plus convenable en fonction de notre expérience et de quelques résultats théoriques (encore insuffisants). Nous utiliserons dans les exemples la valeur $\beta = 2$, mais il peut être utile en pratique d'essayer quelques valeurs entre, disons, 1.5 et 2.5 et comparer les résultats obtenus.

Comme des valeurs atypiques (outliers) constituent un cas particulier de structure des données (un groupe important et un ou plusieurs petits groupes), la méthode s'avère sensible à la présence de telles valeurs qui peuvent avoir tendance à déterminer les premiers plans de projection, masquant ainsi d'autres faits plus importants dans la mesure où les valeurs atypiques sont faciles à exhiber au préalable par une première méthode plus simple (Caussinus, Hakam et Ruiz-Gazen, 2002). Pour éliminer ce type d'effet, deux procédures sont envisageables :

- retrancher des données les outliers précédemment détectés,
- robustifier en remplaçant la matrice V_n par un estimateur de variance robuste.

La première de ces procédures ne demande aucun commentaire supplémentaire, si ce n'est qu'elle exige des choix de type «0 ou 1» pas toujours évidents. Nous présentons ci-dessous la seconde qui nous semble en général préférable et sera largement utilisée dans les exemples du paragraphe 5.

2.2. Méthode robustifiée

Pour rendre la technique moins sensible aux valeurs atypiques, la procédure de base exposée plus haut est modifiée en remplaçant V_n par un estimateur de variance robuste. Nous choisissons l'estimateur U_n introduit par Ruiz-Gazen (1996), estimateur qui a le double mérite d'être d'un calcul simple et d'être dans l'esprit des divers opérateurs utilisés ici. Nous poserons donc :

$$U_n(\beta) = \left(S_n(\beta)^{-1} - \beta V_n^{-1}\right)^{-1}$$

avec

$$S_n(\beta) = \frac{\sum_{i=1}^n \exp(-\frac{\beta}{2}\,\|X_i - \mu_n\|^2_{V_n^{-1}})(X_i-\mu_n)(X_i-\mu_n)'}{\sum_{i=1}^n \exp(-\frac{\beta}{2}\,\|X_i - \mu_n\|^2_{V_n^{-1}})},$$

où μ_n est un estimateur de la moyenne générale; en pratique, il est possible d'utiliser la moyenne empirique ce qui donne la matrice $S_n(\beta)$ introduite par Caussinus, Hakam et Ruiz-Gazen (2002); c'est ce qui est fait plus loin. (L'utilisation d'un centrage robuste pourrait cependant améliorer certains résultats.)

On diagonalise alors $U_n(\beta_1)T_n^{-1}(\beta_2)$. Soit u_j $(j = 1, \ldots, q)$ des vecteurs propres associés aux q plus grandes valeurs propres; ils peuvent être choisis $T_n^{-1}(\beta_2)$ – orthonormés et constituent alors une base $T_n^{-1}(\beta_2)$ – orthonormée d'un sous-espace à q dimensions sur lequel les X_i sont projetés $T_n^{-1}(\beta_2)$ – orthogonalement. Les coordonnées de la projection de X_i dans cette base sont $X_i' T_n^{-1}(\beta_2) u_j$ $(j = 1, \ldots, q)$.

Remarque. – Les paramètres β_1 et β_2 sont indépendants. Le second doit être choisi comme dans la méthode de base ($\beta_2 = 2$ ou voisin de 2). Le premier doit être petit. Cependant, il faut souligner que le problème n'est pas ici tout à fait le même que dans Caussinus, Hakam et Ruiz-Gazen (2002); il s'agit maintenant d'avoir une bonne estimation robuste de la variance et les préceptes donnés par Ruiz-Gazen (1996) s'appliquent conduisant à une valeur de β_1 un peu plus élevée (entre 0,1 et 0,5) : cette question sera reprise dans les exemples.

3. Choix de la dimension : valeurs critiques

Dans le modèle (2) exprimé ci-dessus, la dimension utile d'une projection est q. Mais, en pratique, q est inconnu. Nous montrons dans ce paragraphe comment choisir q de façon objective. Nous détaillons le cas de la méthode de base. Celui de la méthode robuste est tout à fait similaire.

Dans Caussinus et Ruiz-Gazen (1993, 1995) on remarque que, sous des conditions assez générales, les q plus grandes valeurs propres de $V_n T_n^{-1}(\beta)$ convergent vers un nombre strictement supérieur à $\beta + 1/2$, quand n tend vers l'infini, alors que les autres convergent vers $\beta + 1/2$, ce qui suggère de ne conserver que les dimensions correspondant à des valeurs propres significativement supérieures à $\beta + 1/2$. Afin de préciser ce «significativement supérieur» , il faut établir la loi de probabilité (au moins asymptotique) de $V_n T_n^{-1}(\beta)$ sous l'hypothèse nulle $q = 0$. C'est ce qui est fait dans Hakam (2002) (en même temps qu'une étude plus générale pour des hypothèses alternatives). Donnons le résultat qui nous est utile ici. Pour une matrice M carrée d'ordre p et symétrique, nous notons M^* le vecteur colonne à $p(p+1)/2$ lignes constitué, dans l'ordre, des éléments diagonaux de M et des éléments extra-diagonaux du triangle supérieur. On a alors :

PROPOSITION (Hakam, 2002). – *Définissons les matrices V_n et $T_n(\beta)$ comme en (1) ci-dessus et posons :*

$$M_n = \sqrt{n}\left(\frac{2}{2\beta+1} V_n^{1/2} T_n^{-1}(\beta) V_n^{1/2} - I_p\right).$$

Si les vecteurs X_i suivent indépendamment la même loi normale $\mathcal{N}_p(0, I_p)$, la loi asymptotique, quand n tend vers l'infini, de la matrice M_n^* est normale (à $p(p+1)/2$ dimensions) de moyenne nulle et de matrice des variances et covariances

de la forme

$$\left(\begin{array}{cc|cc} b & d & & \\ & \ddots & & 0 \\ d & b & & \\ \hline & & c & 0 \\ & 0 & & \ddots \\ & & 0 & c \end{array}\right)$$

où le bloc sud-est est diagonal de dimension $p(p-1)/2 \times p(p-1)/2$ *avec sur la diagonale*

$$c = \frac{(2\beta+1)^{p+2}}{[(\beta+1)(3\beta+1)]^{p/2+2}} - \frac{1}{(2\beta+1)^2},$$

et le bloc nord-ouest est de dimension $p \times p$ *avec tous les éléments extra-diagonaux égaux à*

$$d = \frac{(2\beta+1)^p \beta^2}{[(\beta+1)(3\beta+1)]^{p/2+2}},$$

et les éléments diagonaux égaux à

$$b = 2c + d.$$

Le terme c est la variance asymptotique des termes extra-diagonaux de M_n, les covariances correspondantes étant nulles, tandis que b correspond à la variance asymptotique et d à la covariance asymptotique des termes diagonaux de M_n. Les covariances asymptotiques entre termes diagonaux et termes extra-diagonaux de M_n sont toutes nulles.

Des tables de la loi asymptotique des valeurs propres de M_n sous l'hypothèse nulle peuvent alors être établies par simulation comme dans Caussinus, Hakam et Ruiz-Gazen (2002) [1] (c'est le même problème avec les nouvelles valeurs ci-dessus de c et d). A partir de là, et puisqu'en pratique on considère plutôt les valeurs propres de $L_n = V_n T_n^{-1}(\beta)$ que celles de M_n, on pourra par exemple procéder à un test d'hypothèses multiples de la façon suivante :

- choisir un niveau de signification α,
- transformer les valeurs propres λ de L_n en les valeurs propres correspondantes μ de M_n (ou réciproquement) au moyen de la relation :

$$\lambda = (\beta + 1/2)\left(1 + \frac{\mu}{\sqrt{n}}\right), \tag{3}$$

[1] A ce sujet, notons que la légende des tables de ce précédent article est incorrecte : il faut lire « Valeurs critiques des 10 plus grandes » au lieu de « Valeurs critiques des 10 plus petites » valeurs propres de M.

– comparer chaque valeur propre empirique ainsi obtenue à la valeur critique correspondante en commençant par la plus grande valeur propre et en validant les dimensions comme significatives tant que la valeur empirique de la valeur propre correspondante est supérieure à la valeur critique.

Remarques.

1. La proposition ci-dessus est énoncée pour une matrice de covariances $V = I_p$. A partir de là, on obtient le résultat général pour V quelconque en multipliant les X_i par $V^{1/2}$ ce qui multiplie L_n à gauche par cette matrice et à droite par son inverse. La proposition ci-dessus est donc suffisante pour étudier la loi des valeurs propres puisque les matrices L et ALA^{-1} ont mêmes valeurs propres.

2. Nous donnons des tables de valeurs critiques asymptotiques pour $\beta = 2$ et les niveaux 1% et 5% (tables 1 et 2). Nous avons fait quelques simulations afin de vérifier la qualité de l'approche asymptotique pour des valeurs modérées de n. Il faut admettre que celle-ci n'est pas très bonne pour, disons, n inférieur à 500, les niveaux réels étant supérieurs aux niveaux nominaux. Les conclusions doivent donc être tirées très prudemment dans ces cas. Des indications sur l'ordre de grandeur des approximations peuvent être trouvées à partir du troisième exemple traité dans le paragraphe 5 où l'on approche la loi exacte (non asymptotique) par simulation.

3. La distribution asymptotique à la base des tables fournies est établie sous l'hypothèse nulle $q = 0$, ce qui est tout à fait justifié pour contrôler le niveau du test relatif à la plus grande valeur propre (hypothèse $q = 0$ contre hypothèse $q > 0$). Mais, si celui-ci s'avère significatif, l'hypothèse nulle testée ensuite au moyen de la seconde plus grande valeur propre est $q = 1$ (contre $q > 1$); sous cette nouvelle hypothèse nulle, la loi asymptotique exacte de cette seconde valeur propre dépend de paramètres inconnus et ne peut être calculée : la loi pour $q = 0$ n'en donne qu'une approximation; on connaît cependant le sens de l'erreur commise : cette seconde valeur propre tend à être plus grande que ne l'indique la distribution pour $q = 0$, le niveau réel est donc supérieur au niveau nominal donné par les tables. La même remarque vaut pour les valeurs propres d'ordre supérieur.

4. Les mêmes raisonnements s'appliquent lorsque l'on fait l'analyse robuste en changeant $V_n T_n^{-1}(\beta)$ en $U_n(\beta_1)T_n^{-1}(\beta_2)$. Il suffit (Hakam, 2002) de prendre les nouvelles valeurs de c et d suivantes :

$$c = \frac{(\beta_1+1)^{p+4}}{(2\beta_1+1)^{p/2+2}} + \frac{(2\beta_2+1)^{p+2}}{[(\beta_2+1)(3\beta_2+1)]^{p/2+2}} - 2\,\frac{(\beta_1+1)^{p/2+2}(2\beta_2+1)^{p/2+1}}{(2\beta_2+\beta_1\beta_2+\beta_1+1)^{p/2+2}} - \frac{4\beta_2^2}{(2\beta_2+1)^2},$$

$$d = \frac{(\beta_1+1)^{p+2}\beta_1^2}{(2\beta_1+1)^{p/2+2}} + \frac{(2\beta_2+1)^p\beta_2^2}{[(\beta_2+1)(3\beta_2+1)]^{p/2+2}} - 2\,\frac{\beta_1\beta_2(\beta_1+1)^{p/2+1}(2\beta_2+1)^{p/2}}{(2\beta_2+\beta_1\beta_2+\beta_1+1)^{p/2+2}}.$$

4. Projections révélatrices et classification

Parmi les structures les plus couramment recherchées figure la répartition des données en groupes homogènes. Et c'est l'une des situations que les analyses étudiées ici sont susceptibles de faire ressortir. En ce sens, elles se rapprochent de la classification automatique. Même si les méthodes de cette dernière sont plus orientées vers l'affectation d'individus à des groupes alors que les nôtres sont d'abord orientées vers la visualisation au moyen d'une réduction de dimension, il est bien clair d'une part que les utilisateurs de la classification automatique ont souvent le souci de visualiser les groupes obtenus (et plusieurs ont souligné la complémentarité de l'approche classification et de l'approche visualisation pour une analyse pertinente de leurs données, par exemple Jambu, 1977, dans un article sur lequel nous reviendrons plus loin; on notera toutefois qu'il s'agit là de juxtaposer des techniques sous leur forme usuelle, même si on les fait dialoguer de façon judicieuse), d'autre part que bien des travaux en classification automatique cherchent parallèlement des réductions de dimension en combinant techniques de classification et techniques factorielles. La première idée est de faire une ACP préalable et de classer les individus à partir d'un certain nombre de composantes principales. Il n'est cependant pas certain que les premières composantes principales contiennent l'essentiel de l'information pertinente pour la classification comme l'ont souligné plusieurs auteurs, certains proposant des techniques alternatives. Une mise en garde déjà ancienne est Chang (1983); pour une revue récente de ces questions et une nouvelle proposition performante, voir Vichi and Kiers (2001). Lebart (2001) applique le principe de l'analyse de contiguïté avec un graphe de contiguïté construit à partir des données elles-mêmes, ce qui conduit à une approche descriptive tout à fait dans l'esprit des techniques de projections révélatrices permettant, selon les termes de l'auteur «*de mettre en évidence des zones d'inégales densités et des structures intermédiaires entre celles que détectent les méthodes factorielles et celles mises en évidence par les méthodes de classification*». On peut aussi mentionner que Montanari et Lizzani (2001) proposent d'utiliser un indice de projections révélatrices pour la sélection de variables en vue d'une classification; mais, contrairement à l'esprit des méthodes factorielles qui nous anime ici, il s'agit pour ces auteurs de retenir ou pas chacune des variables initiales, et non de rechercher les combinaisons les plus utiles de celles-ci. Enfin, pour l'utilisation combinée de techniques factorielles et d'algorithmes de classification (dans un esprit cependant un peu différent) une importante référence est Diday *et al.* (1979).

Puisque l'analyse en composantes principales généralisée que nous présentons ici vise à faire apparaître les effets structuraux, les composantes principales qu'elle fournit devraient être bien adaptées pour une classification. Avant de préciser ce point, notons qu'il s'agit de la motivation principale de Art, Gnanadesikan et Kettenring (1982) qui, les premiers, ont introduit une variance locale dans le même esprit que T_n quoique sensiblement différente.

Considérons la méthode de base (la discussion est analogue pour la méthode robuste) et supposons retenues q dimensions (où q n'est pas nécessairement la «vraie» valeur de la dimension). Notons u_j $(j = 1, \ldots, q)$ des vecteurs propres $T_n^{-1}(\beta)$ - orthonormés de $V_n T_n^{-1}(\beta)$ associés aux q plus grandes valeurs propres, et U_q la matrice $p \times q$ dont les colonnes sont les u_j. Afin d'obtenir une variance «locale» des individus représentés par leurs composantes principales (c'est-à-dire une variance locale pour les projections obtenues), il est naturel d'utiliser une formule du type (1),

avec les mêmes poids, mais en remplaçant chaque $X_i - X_j$ par sa projection, c'est-à-dire par $U_q' T_n^{-1}(X_i - X_j)$, en notant simplement T_n pour $T_n(\beta)$. La variance locale est donc maintenant $U_q' T_n^{-1} T_n T_n^{-1} U_q = I_q$ puisque les u_j sont $T_n^{-1}(\beta)$ - orthonormés. En ce sens, on voit que les groupes, si groupes il y a, sont « sphériques » et seront donc faciles à détecter par tout algorithme de classification « naïf » , nous voulons dire par là ne cherchant pas à « adapter » les distances euclidiennes canoniques, comme un algorithme k-means élémentaire, par exemple.

Par ailleurs, les techniques du paragraphe 3 peuvent permettre de réduire la dimension en s'en tenant aux composantes significatives, puisque les autres contiennent essentiellement des bruits non structurels.

Ces divers aspects sont illustrés et approfondis avec les exemples du paragraphe suivant.

5. Exemples

Précisons avant tout que, selon nous (et bien d'autres !), une analyse exploratoire doit commencer par les choses les plus simples, sans doute d'abord des analyses élémentaires variable par variable, mais nous n'alourdirons pas l'exposé avec celles-ci, ensuite des analyses multidimensionnelles qui, pour le type de données considérées ici, commenceront en général par l'analyse en composantes principales (ACP). Ce n'est qu'après (mais, pensons-nous, quels que soient les résultats de l'ACP) que l'on peut passer à des analyses en composantes principales généralisées, en commençant par la recherche de valeurs atypiques selon la méthode donnée dans Caussinus, Hakam et Ruiz-Gazen (2002) pour continuer avec les techniques développées dans le présent article. Mais, pour les besoins de notre illustration, c'est évidemment ce dernier point qui recevra ici le maximum d'attention.

5.1. Un exemple simulé

Nous considérons un exemple semblable à l'exemple 4.3 de Caussinus et Ruiz (1995). La loi de probabilité P est la loi uniforme sur le cercle de centre 0 et rayon $\sqrt{2}$ situé dans le plan des deux premières coordonnées, tandis que le bruit est une loi normale centrée, à p coordonnées indépendantes, les deux premières de variance σ^2 et les autres de variance $1 + \sigma^2$. La loi d'un X_i est alors centrée de matrice de variances et covariances $(1 + \sigma^2) I_p$. Nous avons pris ici $p = 4$, $\sigma = 0,2$ et nous avons généré $n = 500$ individus.

La figure 1 donne la projection des points simulés sur le plan des deux premières coordonnées. Vu la matrice de variance des X_i, il est clair que l'ACP fournit des résultats (projections) totalement arbitraires, ce que l'on peut vérifier avec la figure 2 qui montre les projections sur le premier plan principal. On a enfin utilisé notre méthode, méthode de base avec $\beta = 2$. La figure 3 donne les projections sur le premier plan principal. Les valeurs propres sont :

$$3,29 \quad 3,12 \quad 2,50 \quad 2,47$$

Les valeurs critiques asymptotiques déduites des tables 1 et 2 au moyen de la formule (3) sont :

au niveau 5% : $2,61 \quad 2,56 \quad 2,52 \quad 2,48$

au niveau 1% : $2,63 \quad 2,58 \quad 2,54 \quad 2,50$

Les deux premières valeurs propres observées sont significatives à 1% alors que les suivantes sont non significatives à 5%, ce qui est évidemment cohérent avec la nature de l'exemple (en fait la troisième valeur propre empirique est proche de la valeur critique à 5% mais la conclusion précédente est renforcée par la remarque 3 du paragraphe 3). La loi P légèrement bruitée se retrouve bien visible dans le premier plan principal.

TABLE 1

Valeurs critiques des plus grandes valeurs propres de M_n pour $p = 1, 2, ..., 14$ au niveau de 1% (obtenues à partir de 100000 simulations)

p	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	μ_9	μ_{10}
2	0.742	0.301								
3	0.929	0.494	0.139							
4	1.127	0.672	0.329	-0.02						
5	1.317	0.86	0.507	0.178	-0.163					
6	1.511	1.032	0.675	0.356	0.038	-0.31				
7	1.72	1.212	0.85	0.525	0.214	-0.108	-0.461			
8	1.932	1.413	1.032	0.698	0.383	0.075	-0.263	-0.625		
9	2.161	1.615	1.221	0.882	0.56	0.246	-0.074	-0.408	-0.791	
10	2.354	1.829	1.42	1.061	0.743	0.418	0.098	-0.23	-0.584	-0.9
11	2.648	2.047	1.649	1.249	0.917	0.59	0.268	-0.048	-0.378	-0.735
12	2.909	2.276	1.834	1.452	1.103	0.773	0.447	0.133	-0.203	-0.546
13	3.181	2.522	2.058	1.662	1.305	0.96	0.629	0.297	-0.039	-0.367
14	3.452	2.78	2.295	1.883	1.507	1.154	0.814	0.482	0.149	-0.187

TABLE 2

Valeurs critiques des plus grandes valeurs propres de M_n pour $p = 1, 2, ..., 14$ au niveau de 5% (obtenues à partir de 100000 simulations)

p	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	μ_9	μ_{10}
2	0.583	0.156								
3	0.761	0.35	-0.011							
4	0.948	0.524	0.185	-0.162						
5	1.13	0.698	0.356	0.033	-0.314					
6	1.319	0.869	0.527	0.206	-0.113	-0.471				
7	1.519	1.048	0.693	0.374	0.061	-0.262	-0.631			
8	1.725	1.236	0.867	0.54	0.226	-0.86	-0.418	-0.794		
9	1.94	1.43	1.049	0.713	0.395	0.081	-0.236	-0.574	-0.968	
10	2.168	1.634	1.245	0.892	0.565	0.249	0.076	-0.388	-0.764	-1.125
11	2.293	1.84	1.431	1.072	0.738	0.417	0.098	-0.221	-0.557	-0.917
12	2.648	2.066	1.64	1.268	0.923	0.596	0.271	-0.051	-0.382	-0.728
13	2.902	2.303	1.856	1.47	1.112	0.775	0.444	0.118	-0.215	-0.522
14	3.174	2.546	2.08	1.676	1.316	0.965	0.625	0.294	-0.042	-0.38

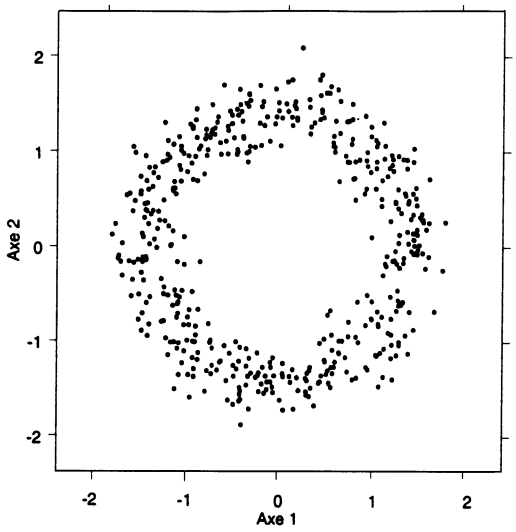

FIGURE 1
Projection des points sur le plan des deux premières coordonnées pour l'exemple simulé

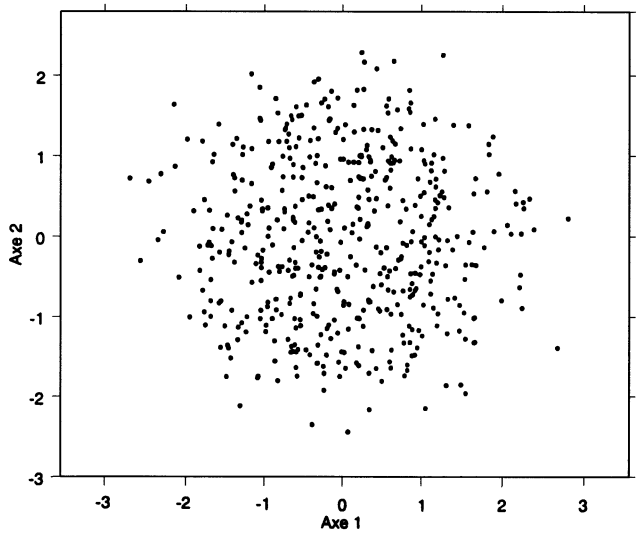

FIGURE 2
Plan principal (1,2) de l'ACP pour l'exemple simulé

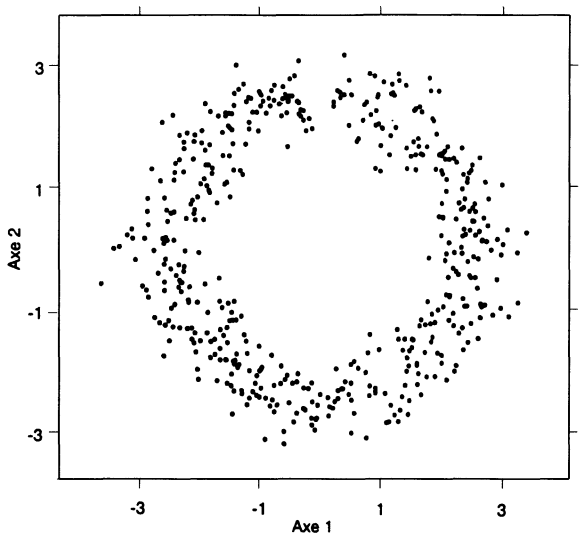

FIGURE 3
Plan principal (1,2) de l'ACP avec la métrique $T_n^{-1}(2)$ pour l'exemple simulé

5.2. Deux exemples réels

On considère maintenant deux exemples réels dans lesquels une classification est plus ou moins connue *a priori*; bien entendu cette connaissance n'est pas utilisée dans l'analyse, mais elle permet de contrôler l'efficacité des méthodes proposées qui cherchent, rappelons-le, à découvrir une structure « vraie » en contrôlant le risque de « découvrir » une structure illusoire. Nous allons voir que la méthodologie proposée est efficace pour retrouver la structure connue et aussi peut-être pour enrichir l'analyse sur quelques autres points.

5.2.1. Quelques insectes incontournables

Nous reprenons les données de Lubischew (1962) qui sont devenues un passage obligé en matière de projections révélatrices. Il s'agit de n=74 insectes sur chacun desquels sont mesurées p=6 grandeurs morphologiques. L'entomologiste pense *a priori* qu'il y a trois classes différentes constituées respectivement des individus numérotés 1 à 21, 22 à 43, 44 à 74.

La visualisation de ces données a été examinée par l'un de nous dans le chapitre 8 de l'ouvrage édité par Droesbecke, Fichet et Tassi (1992). On renvoie à cet ouvrage pour voir d'abord que l'ACP usuelle ne donne qu'une pâle information sur la structure du nuage des individus. Selon le précepte énoncé plus haut, nous avons ensuite procédé à la détection de possibles valeurs aberrantes par la méthode exposée dans Caussinus, Hakam et Ruiz-Gazen (2002). La projection sur le premier plan principal (figure 4) montre que les individus 67 et 68 sont un peu « périphériques » , mais cela n'est pas significatif : en effet la plus grande valeur propre de l'analyse vaut seulement 1,08

alors que la valeur critique selon les tables de notre article précèdent est ici de 1,13 au niveau 5%, pour β=0.05. On peut néanmoins penser qu'une analyse faisant entrer dans le rang ces valeurs extrêmes sera sans doute plus efficace. C'est bien vérifié si l'on compare les figures 5 et 6 qui donnent respectivement les projections des individus sur le premier plan principal de la méthode de base et de la méthode robuste exposées au paragraphe 2. Pour le choix de β_1 dans cette dernière, nous avons peu d'information sur la proportion «d'outliers» attendue, si ce n'est qu'il devrait y en avoir peu ou pas selon l'analyse précédente; dans ces conditions (Ruiz-Gazen, 1996), il convenait d'essayer quelques valeurs assez petites : nous l'avons fait et obtenu des vues très similaires, celle que nous donnons est pour $\beta_1 = 0,1$.

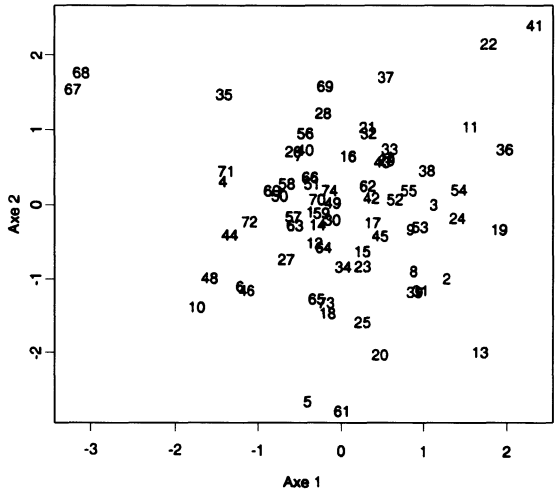

FIGURE 4
Plan principal (1,2) de l'ACP avec la métrique $S_n^{-1}(0{,}05)$
pour l'exemple des insectes

A cette étude, déjà présentée pour l'essentiel dans Droesbecke, Fichet et Tassi (1992), ajoutons maintenant plusieurs points importants.

– La suite des valeurs propres pour la méthode de base ($\beta = 2$) est :

$$5,27 \quad 3,12 \quad 2,64 \quad 2,36 \quad 2,14 \quad 2,12$$

A partir des tables 1 et 2 et de la relation (3), on calcule les valeurs critiques :

au niveau 5% :	2,88	2,75	2,65	2,56	2,47	2,36
au niveau 1% :	2,94	2,80	2,70	2,60	2,51	2,41

Seules les deux premières valeurs observées sont significatives. Nous devons donc conclure que les dimensions suivantes ne représentent que du bruit :

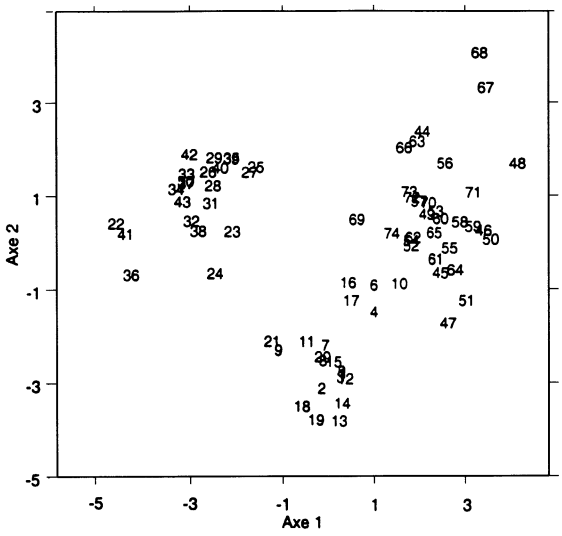

FIGURE 5
Plan principal (1,2) de l'ACP avec la métrique $T_n^{-1}(2)$ pour l'exemple des insectes

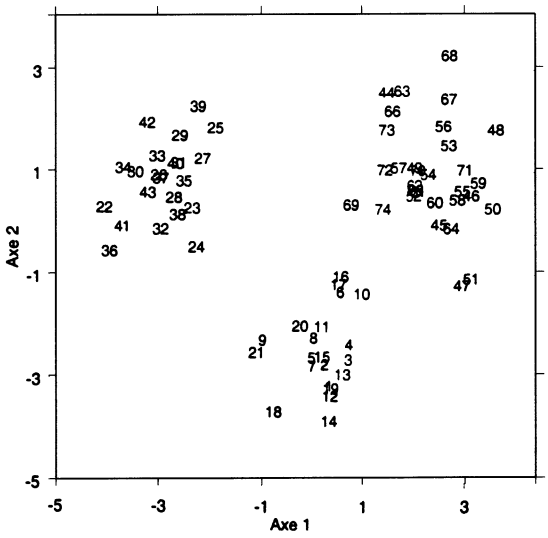

FIGURE 6
Plan principal (1,2) de l'ACP robuste avec $U_n(0,1)$ et la métrique $T_n^{-1}(2)$ pour l'exemple des insectes

l'étude nous dit ainsi que toute interprétation au delà du premier plan principal risque d'être illusoire. Ici encore notons que, si la troisième valeur propre est tout juste non significative au niveau nominal 5%, le niveau réel ne peut être que supérieur conformément aux remarques 2 et 3 du paragraphe 3.

- Nous avons utilisé l'algorithme k-means tel qu'il est donné en routine dans S-plus pour la recherche de 3 groupes. Sur les données d'origine (réduites), quatre individus sont mal affectés : les numéros 6 et 10 d'une part, 47 et 68 de l'autre. L'utilisation de l'algorithme sur les premières composantes de l'ACP (réduite) donne les quatre mêmes erreurs au moins quel que soit le nombre de composantes considéré. Nous avons ensuite utilisé ce même algorithme à partir des composantes principales des techniques discutées ici. Avec les deux premières composantes principales de la méthode de base, seul l'individu 10 reste mal classé, et il n'y a plus aucune erreur si l'on utilise les trois premières composantes ou plus. En fait, bien que non significative, la troisième composante peut encore contenir de l'information; d'autre part, il ne semble pas préjudiciable ici d'introduire des variables «inutiles» (sans doute grâce à la sphéricité des classes montrée au paragraphe 4). Enfin, avec les deux premières composantes de l'analyse robuste, l'algorithme k-means fournit la bonne classification, l'individu 10 étant maintenant rentré dans le rang, et il en va de même avec un nombre plus élevé de composantes. La supériorité de l'analyse robuste s'avère confirmée.

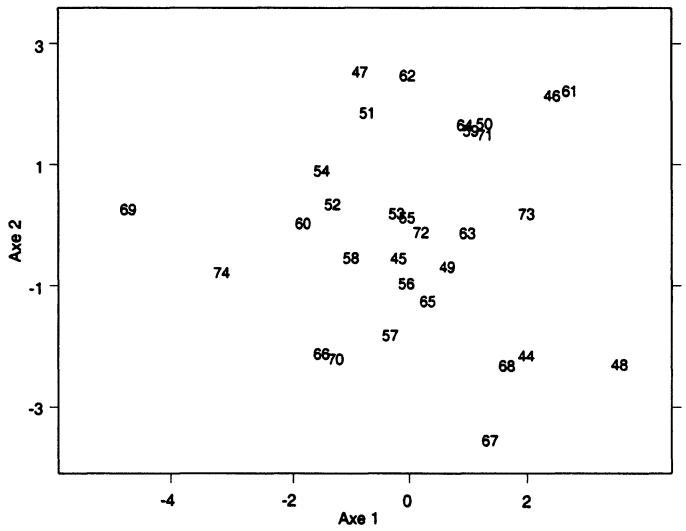

FIGURE 7
Plan principal (1,2) de l'ACP avec la métrique $T_n^{-1}(2)$ pour les individus 44 à 74 de l'exemple des insectes

- Nous n'avons pas discuté ci-dessus du choix du nombre de classes, car il semble s'imposer d'après les graphiques obtenus et d'après la dimension de

l'espace «significatif» . Nous avons cependant essayé une classification en quatre classes; l'algorithme retenu coupe alors en deux la troisième classe, celle des individus 44 à 74. Cela nous a amenés à examiner la structure de ce nuage de 31 points au moyen des analyses du paragraphe 2. Pour s'en tenir à la méthode de base, la figure 7 montre les projections des individus dans le premier plan principal. Il est difficile d'y voir une structuration sans faire preuve de beaucoup d'imagination et, pour le confirmer, la plus grande valeur propre est égale à 2,85 et elle est non significative (les tables et la relation (3) donnent une valeur critique de 3,09 au niveau nominal de 5%, lequel pour ce petit échantillon est très sous-évalué selon la remarque 2 du paragraphe 3 et les simulations que nous avons effectuées).

5.2.2. Entre chiens et loups

Nous revisitons maintenant l'exemple avec lequel Jambu (1977) illustre l'utilisation conjointe de l'ACP et de techniques de classification hiérarchique. C'est un petit fichier de 43 individus (des canidés) et 6 variables, mais, en adaptant nos ambitions à sa taille, il est suffisant pour montrer le parti qu'on peut tirer des méthodes que nous avons exposées. Il y a 12 loups numérotés 31 à 42 (nous avons conservé la numérotation de l'article d'origine) et 31 chiens de races diverses (individus 1 à 30 et 43, ce dernier ayant un statut spécial sur lequel nous ne reviendrons pas). Jambu (1977) fait d'abord une ACP (réduite) et note que le premier facteur tend à séparer chiens et loups quoiqu'il reste un recouvrement certain des deux groupes. Il en va de même pour les classifications qu'il établit ensuite. Pour notre part, conformément au précepte énoncé en préambule de ce paragraphe, après l'ACP nous avons fait une analyse de structure visant d'abord à détecter d'éventuelles valeurs atypiques. La méthode de notre précédent article conduit ici à la représentation de la figure 8 dans le premier plan principal. Les deux premières valeurs propres sont significatives et nous sommes conduits à accepter la présence de trois individus atypiques qui sont deux loups et un bull-dog. L'analyse de structure est alors poursuivie par les méthodes du présent papier. La méthode de base apprend peu de nouveau : elle redonne surtout (en moins clair) les points atypiques, illustrant nos remarques de la fin du paragraphe 2.1. sur le masquage par ceux-ci d'autres types de structure (voir figure 9). La méthode robuste s'impose donc. Pour l'appliquer, il faut choisir une valeur judicieuse de β_1, sans doute pas trop petite étant donné la proportion apparente d'«outliers» . On peut faire quelques essais, mais les recommandations de Ruiz-Gazen (1996) peuvent aider à les diriger : le nombre de variables étant $p = 6$ et la proportion d'outliers étant estimée à $\varepsilon = 3/43$, cet article suggère qu'une bonne valeur de départ est donnée par $26 \times \frac{3}{43} \times \frac{1}{6} = 0,3$; nous avons finalement retenu $\beta_1 = 0,2$ qui donne les projections des figures 10 et 11 respectivement sur les plans (1,2) et (1,3). La suite des valeurs propres observées est :

$$4,32 \quad 3,70 \quad 3,06 \quad 2,84 \quad 1,71 \quad 1,44$$

Comme nous l'avons dit, les résultats asymptotiques sont très sujets à caution pour $n = 43$. Nous avons donc préféré calculer ici des valeurs critiques par simulation de 1000 jeux de données sous l'hypothèse nulle. A tous les niveaux les plus usuels, ces valeurs propres ne sont pas significatives; par exemple la valeur critique pour la première valeur propre au niveau 5% est 5,07. Cependant, à cause de la petite taille de l'échantillon, la puissance du test considéré est très faible. Pour respecter

un équilibre entre les deux types d'erreurs, on peut donc être tenté d'augmenter le niveau. La première valeur propre s'avère significative à partir du niveau 0,19 (P-value). Si ce risque est accepté, nous sommes conduits à valider le premier axe c'est-à-dire la séparation en deux groupes visible sur la figure 10 et surtout sur la figure 12 qui montre la distribution des individus sur l'axe 1 (histogramme et lissage par un estimateur de densité à noyau) : les deux groupes, respectivement en-deçà et au-delà de l'abscisse 1 (environ), correspondent aux chiens et aux loups, à la seule exception du chien 29 (un saint-bernard égaré chez les loups !). Jetant un coup d'œil aux deux dimensions suivantes pour lesquelles les probabilités de dépassement (P-values) sous l'hypothèse nulle sont de l'ordre de 0.10, la deuxième ne semble pas apporter beaucoup d'information, mais sur le plan (1,3) (figure 11) on peut remarquer trois individus (1, 2 et 23) qui ont l'air un peu à part du côté des chiens. Cela peut très bien être un artefact car les niveaux de signification sont assez élevés; mais d'une part nous rappelons que la puissance des tests est nécessairement faible pour cet échantillon de petite taille si bien que des aspects structuraux réels ont de bonnes chances de ne pas être jugés significatifs, d'autre part on sait ici que ces trois individus sont des chiens aux ressemblances certaines (au moins pour les non cynologues que nous sommes) puisqu'il s'agit de deux bull-dogs et d'un boxer (les seuls animaux de ces races dans l'échantillon) : si aucune information nouvelle n'est vraiment découverte, la configuration observée correspond au moins à une certaine cohérence de notre méthode de construction des axes principaux. Cette cohérence est encore renforcée si l'on examine les variables qui contribuent le plus à la détermination des axes; ce n'est pas notre objectif d'insister ici sur ce point (voir pour cela Gabriel, 2002), mais on peut noter que le premier axe dépend avant tout de la variable « longueur de la carnassière supérieure » (plus longue pour tous les loups que pour tous les chiens, sauf le saint-bernard), tandis que le troisième est largement déterminé par la variable « longueur de la mâchoire supérieure » , nettement plus faible chez les trois chiens mentionnés ci-dessus.

Nous avons enfin vérifié l'aide à la classification que la méthode pouvait apporter, comme dans l'exemple qui précède. Avec l'algorithme k-means pour la recherche de deux classes sur les données d'origine réduites, les loups se retrouvent dans le même groupe mais avec 18 chiens, les autres chiens forment le second groupe, les chiens de même espèce n'étant pas nécessairement dans le même groupe. Par contre, le même algorithme sur la première composante de notre analyse robuste fournit une classe de chiens et une de loups à seulement deux « erreurs » près, le bull-dog 2 et le saint-bernard 29 se retrouvant avec les loups. Il en va de même lorsqu'on utilise les deux premières composantes principales, mais classer sur trois composantes fournit un résultat moins bon. En rapprochant ces résultats de ceux de l'exemple précédent, on suggère en pratique de commencer en utilisant seulement les composantes significatives (à un niveau adapté à la taille de l'échantillon), d'essayer éventuellement dans un second temps d'en introduire une ou deux autres avec précaution (de toutes façons, le caractère significatif dépend d'un niveau arbitrairement choisi, d'où la nécessaire souplesse d'application). Il ne faut pas perdre de vue que l'examen des projections obtenues doit aider aussi bien au choix du nombre de classes qu'au choix du nombre de composantes utiles (et il y a un lien entre les deux) et à la classification finale. Ainsi, dans l'exemple présent, l'examen de la figure 10 permet de supprimer une des « erreurs » puisque le chien 2 doit alors assez clairement rejoindre son groupe.

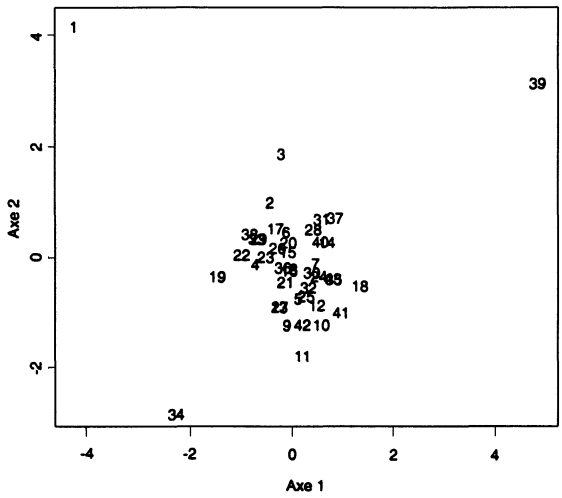

FIGURE 8
Plan principal (1,2) de l'ACP avec la métrique $S_n^{-1}(0,05)$
pour l'exemple des chiens et des loups

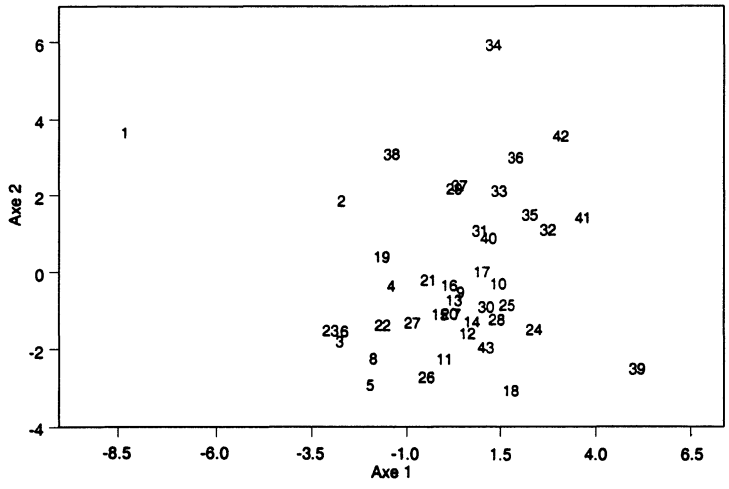

FIGURE 9
Plan principal (1,2) de l'ACP avec la métrique
$T_n^{-1}(2)$ *pour l'exemple des chiens et des loups*

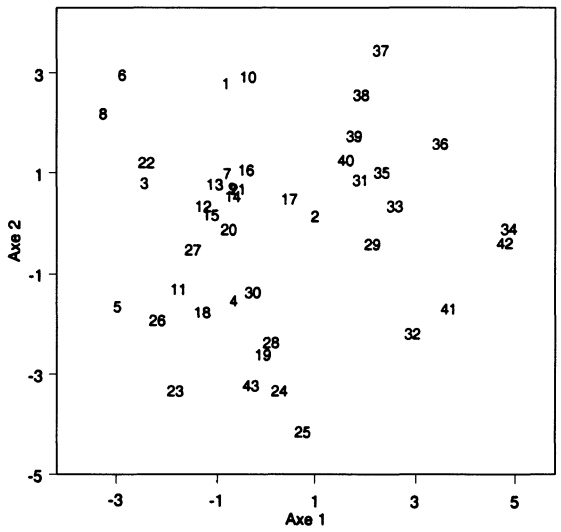

FIGURE 10
Plan principal (1,2) de l'ACP robuste avec $U_n(0,2)$ et la métrique $T_n^{-1}(2)$ pour l'exemple des chiens et des loups

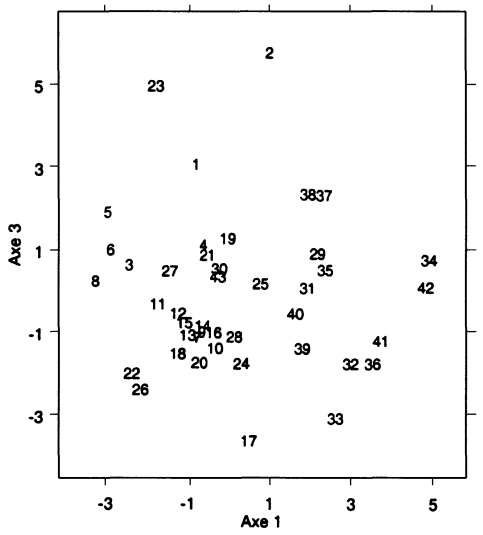

FIGURE 11
Plan principal (1,3) de l'ACP robuste avec $U_n(0,2)$ et la métrique $T_n^{-1}(2)$ pour l'exemple des chiens et des loups

Nous avons procédé aussi à une classification hiérarchique avec distance euclidienne pour comparer aux analyses de Jambu (1977). Pour simplifier en ne donnant que les résultats de dernier niveau, disons que cet auteur, à partir des données initiales réduites, tend à classer quatre dogues allemands et le saint-bernard avec les loups, alors qu'avec les premières composantes de notre analyse, nous avons seulement trois « erreurs » : outre le saint-bernard, les chiens 2 et 17 sont classés avec les loups; ces deux derniers chiens sont regroupés à un niveau très bas pour des raisons qui sont claires à la vue de la figure 10, mais clairement injustifiées à la vue de la figure 11.

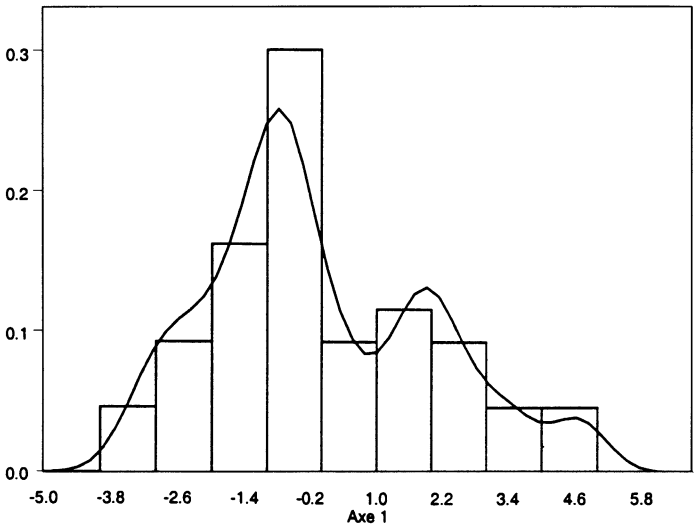

FIGURE 12
Distribution des individus sur le premier axe principal de l'ACP robuste avec $U_n(0,2)$ et la métrique $T_n^{-1}(2)$ pour l'exemple des chiens et des loups

6. Conclusion

Le premier exemple présenté ci-dessus n'a pas été choisi tout à fait au hasard, mais en écho à celui qui illustre le troisième principe de Benzécri (1973, page 9) : « *... le problème de la validité (du test) – troublant parfois avouons-le – passe au second plan... tandis que de trouver dans un espace de dimension deux cinquante points approximativement rangés sur un cercle est sûrement une découverte (à moins que la méthode de calcul ne soit une duperie)...* ». Nous sommes bien d'accord sur le fond ce qui nous a justement conduits à nous intéresser aux techniques de projections révélatrices : puisqu'il est important de découvrir un cercle (ou telle autre structure), il convient de se donner tous les moyens pour cela; nous avons cherché à faire partager l'idée que la méthode proposée ici est l'un d'entre eux. Par ailleurs, pour que la méthode de calcul ne soit pas une duperie, il faut se poser la question de la validation des résultats fournis (même si c'est au second plan ou au moins dans un second

temps) surtout si le nombre de techniques exploratoires disponibles augmente, ce qui n'était certainement pas dans l'esprit de Benzécri (1973) mais devient aujourd'hui une réalité; et le problème est d'autant plus aigu en matière de projections révélatrices : une technique construite pour chercher les aspects structurels d'un nuage de points risque fort d'en trouver à tout prix, réels ou non... Que l'on raisonne en terme de tests ou d'autres principes (choix de modèle,. . .) il nous paraissait donc important de pouvoir valider non pas un modèle (ce serait bien ambitieux) mais au moins l'outil de représentation, ici la projection, qui peut conduire l'utilisateur à telle ou telle conclusion; *a contrario*, cela revient à ne pas valider une représentation susceptible de ne contenir que des bruits informels, à partir de laquelle ne pourraient être tirées que des conclusions trompeuses ou, tout au moins, fort hasardeuses.

Les méthodes discutées ici présentent clairement des insuffisances, certaines sans doute incontournables, d'autres pouvant être aménagées. Bien que les problèmes de temps de calcul soient de moins en moins aigus, il convient de noter que ce temps est ici en n^2, c'est-à-dire devient vite grand avec n. Par ailleurs, nos résultats asymptotiques ne sont valables que pour n assez grand. Notons cependant que ces deux inconvénients ne se cumulent pas : en effet, pour n petit le recours à l'asymptotique peut être évité par des simulations exigeant dans ce cas des moyens raisonnables, tandis que pour n grand les résultats asymptotiques évitent le recours à des simulations qui seraient très coûteuses en temps. Enfin, et surtout, les exemples proposés illustrent des possibilités intéressantes, mais nous ne voudrions pas laisser croire que les méthodes présentées sont aussi efficaces dans tous les cas : en fait, il s'agit de méthodes exploratoires qui doivent s'intégrer dans un ensemble d'outils d'analyse. Elles ne sauraient être universelles. Cela dit, manipulées avec doigté et bien articulées avec d'autres techniques (pour commencer avec celles discutées dans Caussinus, Hakam et Ruiz-Gazen, 2002) elles peuvent apporter des éléments utiles à la compréhension d'un ensemble de données, soit directement (représentations graphiques), soit indirectement (comme l'aide à la classification).

7. Références

Art, D., Gnanadesikan, R., Kettenring, J. R. (1982), Data-based metrics for cluster analysis, Utilitas Mathematica, 21A, 75-99.

Benzécri, J.-P. (1973), L'analyse des données, t.II : l'analyse des correspondances. Bordas, Paris.

Caussinus, H., Hakam, S., Ruiz-Gazen, A. (2002), Projections révélatrices contrôlées - recherche d'individus atypiques, Revue de Statistique Appliquée, L(4), 5-37.

Caussinus, H., Ruiz A. (1990), Interesting projections of multidimensional data by means of generalized principal component analyses, COMPSTAT 90, Physica-Verlag, Heidelberg, 121-126.

Caussinus H., Ruiz-Gazen A. (1993), Projection Pursuit and Generalized Principal Component Analyses, In New Directions in Statistical Data Analysis and Robustness, Eds. Morgenthaler S. et al., Birkhäuse Verlag, Basel Boston Berlin, 35-46.

Caussinus H., Ruiz-Gazen A. (1995), Metrics for finding typical structures by means of principal component analysis, In Data Science and its Applications, Eds. Escoufier Y. et al., Academic Press, 177-192.

Chang, W. C. (1983), On using principal components before separating a mixture of two multivariate normal distributions, Applied Statistics, 32, 3, 267-275.

Diday, E. et al. (1979), Optimisation en classification automatique, vol. 1 et 2, INRIA, Le Chesnais, France.

Droesbecke J.-J., Fichet B. et Tassi Ph. (1992), Modèles pour l'analyse des données multidimensionnelles, Economica, Paris.

Gabriel K.R. (2002), Le Biplot, outil d'exploration de données multidimensionnelles, Journal de la Société Française de Statistique, à paraître.

Hakam S. (2002), Tests de signification pour quelques méthodes de projections révélatrices et applications, Thèse, Université Mohammed V-Agdal, Rabat.

Jambu, M. (1977), Sur l'utilisation conjointe d'une classification hiérarchique et de l'analyse factorielle en composantes principales, Revue de Statistique Appliquée, XXV, 4, 5-35.

Lebart, L. (2001), Classification et analyse de contiguïté, La Revue de Modulad, 27, 1-22.

Lubischew, A.A. (1962), On the use of discriminant functions in taxonomy, Biometrics, 18, 455-477.

Montanari, A. and Lizzani, L. (2001), A projection pursuit approach to variable selection, Computational Statistics and Data Analysis, 35, 4, 463-473.

Ruiz-Gazen A. (1996), A very simple robust estimator for a dispersion matrix, Computational Statistics and Data Analysis, 21, 149-162.

Vichi, M. and Kiers, H. A. L. (2001), Factorial k-means analysis for two-way data, Computational Statistics and Data Analysis,37,1, 49-64.