

# REVUE DE STATISTIQUE APPLIQUÉE

J. BENASSENI

## **Influence des poids des unités statistiques sur les valeurs propres en analyse en composantes principales**

*Revue de statistique appliquée*, tome 33, n° 4 (1985), p. 41-55

[http://www.numdam.org/item?id=RSA\\_1985\\_\\_33\\_4\\_41\\_0](http://www.numdam.org/item?id=RSA_1985__33_4_41_0)

© Société française de statistique, 1985, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

# INFLUENCE DES POIDS DES UNITÉS STATISTIQUES SUR LES VALEURS PROPRES EN ANALYSE EN COMPOSANTES PRINCIPALES

J. BENASSENI

Unité de Biométrie  
9, place Viala 34060 Montpellier Cedex

---

## RESUME

On étudie, en analyse en composantes principales, l'influence sur les valeurs propres des poids attribués aux unités statistiques (u.s.) du tableau de données. On considère tout d'abord la perturbation des valeurs propres résultant d'une modification quelconque de la matrice de poids. On développe ensuite plus particulièrement le cas où l'on modifie le poids d'une seule u.s., les autres poids restant inchangés à une normalisation près. Les aspects géométriques et graphiques de ce dernier point sont soulignés et une illustration pratique d'une partie des résultats est donnée à partir de l'exemple bien connu des poissons d'Amiard.

## ABSTRACT

This paper is concerned with the influence of the weights of the observations upon the eigenvalues in principal component analysis. The consequences of a general perturbation of the weights are studied in a first part. The following results are devoted to the perturbation of the weight of a particular observation, the weights of the other observations being only normalized. Geometrical and graphical aspects are pointed out. A practical illustration of a part of the results, based on the well known example of Amiard fishes, is given.

## 1. INTRODUCTION

On considère un tableau de données  $X$  regroupant les mesures de  $n$  u.s. (lignes du tableau) sur  $p$  variables. Chaque u.s.  $i$ ,  $i = 1, \dots, n$  est caractérisée par le vecteur  $x_i = (x_{i1}, \dots, x_{ip})$  de  $\mathbf{R}^p$ . L'Analyse en Composantes Principales (A.C.P.) du tableau est conditionnée par le choix d'une métrique  $Q$  sur l'espace  $\mathbf{R}^p$  représentatif des u.s. et d'une matrice de poids  $D = \text{diag } p_i$  où figurent les poids  $p_i$ ,  $i = 1, \dots, n$  ( $p_i \geq 0$ ,  $\sum_{i=1}^n p_i = 1$ ) attribués à ces différentes u.s. La matrice  $D$  définit le plan d'échantillonnage adopté et joue le rôle de métrique sur l'espace  $\mathbf{R}^n$  représentatif des variables.

Le statisticien, avant de réaliser l'A.C.P. des données, est en droit de se demander quelle importance aura sur les inerties des axes le choix des métriques  $Q$  et  $D$ .

En particulier une question importante est de savoir si un choix de métrique « légèrement différent » serait à même d'entraîner des variations « importantes » de ces inerties qui se définissent comme les valeurs propres des opérateurs  $WD = (I - ee'D) XQX'(I - Dee') D$  ou

$$VQ = X'(I - Dee') D (I - ee'D) XQ$$

avec  $e' = (1, \dots, 1) \in \mathbf{R}^n$ ,  $W$  et  $V$  étant les matrices de produits scalaires respectivement entre u.s. et entre variables. Il est immédiat de voir que  $WD$  et  $VQ$  ont mêmes valeurs propres et qu'en ce qui concerne ces dernières on peut raisonner sur les opérateurs symétriques  $D^{1/2}WD^{1/2}$  et  $LVL'$  avec  $D^{1/2} = \text{diag } \sqrt{p_i}$  et  $L$  défini par la décomposition de Choleski de  $Q$  selon  $Q = L'L$ .

Le problème d'une perturbation de la métrique  $Q$  ayant été traité par B. ESCOPIER en [5], nous nous proposons ici d'aborder celui d'une modification de la métrique  $D$ , ce qui nous semble d'autant plus important que dans la pratique  $Q$  est souvent déterminée à partir du tableau  $X$  et de  $D$ .

Après avoir rappelé quelques résultats algébriques dans le deuxième paragraphe, nous donnerons dans le troisième, des bornes aux variations des valeurs propres engendrées par une transformation de  $D$  en une matrice de poids  $\tilde{D} = \text{diag } \tilde{p}_i$  quelconque. On raisonnera alors sur l'opérateur  $D^{1/2}WD^{1/2}$ .

Nous nous intéresserons ensuite jusqu'à la fin à la modification du poids d'une seule u.s., les autres poids n'étant affectés que par un coefficient normalisateur. Ce cas nous semble en effet le plus important puisqu'il permet d'étudier l'influence d'une u.s. particulière sur les valeurs propres à partir de la variation continue que l'on exerce sur son poids. Les résultats se fonderont alors sur l'opérateur  $LVL'$ .

Dans le paragraphe 4 on supposera tout d'abord que la métrique  $Q$  est l'identité. Après avoir explicité les variations de la matrice de variance  $V$ , des bornes à la variation des valeurs propres seront proposées.

Le paragraphe 5 traitera du cas important dans la pratique où  $Q = \text{diag } 1/s_i^2$  est la matrice diagonale inverse des variances des variables. On notera qu'alors  $Q^{1/2}VQ^{1/2}$  (avec  $Q^{1/2} = \text{diag } 1/s_i$ ) n'est autre que la matrice des corrélations et l'on montrera comment évoluent les résultats du paragraphe 4.

Les trois derniers paragraphes développeront les résultats obtenus au paragraphe 5 pour la matrice de corrélation.

Dans le paragraphe 6, un réel positif  $\Theta$  étant arbitrairement choisi, on donnera des conditions géométriques, sur l'u.s. dont on modifie le poids, permettant d'assurer que les variations des valeurs propres seront d'amplitude inférieure à  $\Theta$ .

Dans le paragraphe suivant on proposera des abaques exprimant le pourcentage maximal de variation des valeurs propres, en fonction de la variation de poids et de certaines caractéristiques de l'u.s. concernée.

Un dernier paragraphe illustrera une partie des résultats dans le cadre de l'exemple des poissons d'Amiard dont on trouvera une présentation dans [6].

## 2. PRÉLIMINAIRES ALGÈBRIQUES

Tout au long de notre travail, les valeurs propres d'une matrice carrée d'ordre  $m$  seront notées  $\lambda_i(A)$   $i = 1, \dots, m$  avec  $\lambda_1(A) \geq \lambda_2(A) \geq \dots \geq \lambda_m(A)$  ou plus simplement  $\lambda_i$   $i = 1, \dots, m$  avec  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$  lorsqu'il n'y aura pas d'ambiguïté sur la matrice considérée.

### Proposition 2.1.

Soient  $A$  et  $B$  deux matrices symétriques d'ordre  $m$ . Alors pour  $i, j, k$  éléments de  $\{1, \dots, m\}$  vérifiant  $j + k \leq i + 1$  les inégalités suivantes se trouvent vérifiées.

I — Inégalités de Weyl

$$(1) \lambda_i(A + B) \leq \lambda_j(A) + \lambda_k(B)$$

$$(2) \lambda_{m-i+1}(A + B) \geq \lambda_{m-j+1}(A) + \lambda_{m-k+1}(B)$$

$$(3) \lambda_m(B) + \lambda_i(A) \leq \lambda_i(A + B) \leq \lambda_i(A) + \lambda_1(B)$$

II — Si  $A$  et  $B$  sont semi-définies positives

$$(4) \lambda_i(AB) \geq \lambda_j(A) \lambda_k(B)$$

$$(5) \lambda_{m-i+1}(AB) \geq \lambda_{m-j+1}(A) \lambda_{m-k+1}(B)$$

$$(6) \lambda_m(B) \lambda_i(A) \leq \lambda_i(AB) \leq \lambda_i(A) \lambda_1(B)$$

### Remarque :

En ce qui concerne les inégalités de Weyl on pourra consulter [7]. Les inégalités II ont été obtenues en [1] dans le cas où  $B$  est supposée définie positive. Il est immédiat d'étendre le résultat au cas où  $B$  est semi-définie positive seulement. Pour cela on pourra se reporter à [2]. On notera enfin que (3) et (6) s'obtiennent comme cas particuliers des inégalités (1), (2) et (4), (5) respectivement.

## 3. MODIFICATION QUELCONQUE DE LA MÉTRIQUE D

Dans ce paragraphe on étudie la modification des valeurs propres engendrée par une transformation quelconque de la matrice de poids  $D = \text{diag } p_i$  en  $\tilde{D} = \text{diag } \tilde{p}_i$ . Les résultats se fondent sur le lemme technique suivant :

### Lemme 3.1.

Soit  $\Delta = \text{diag } \pi_i$  une matrice de poids dans  $\mathbf{R}^n$

$$\left( \pi_i \geq 0, i = 1, \dots, n, \sum_{i=1}^n \pi_i = 1 \right).$$

On pose  $\Delta^{1/2} = \text{diag } \sqrt{\pi_i}$ . Alors l'opérateur  $I - \Delta^{1/2} e e' \Delta^{1/2}$  est le projecteur orthogonal de  $\mathbf{R}^n$  (pour la métrique usuelle) parallèlement à la direction de  $u = (\sqrt{\pi_1}, \dots, \sqrt{\pi_n})$ .

La démonstration ne présente aucune difficulté si l'on remarque que  $e'\Delta e = 1$ .

Supposons tout d'abord que la métrique  $Q$  n'est pas effectuée par la transformation de  $D$ . Il s'agit alors de comparer les inerties des axes factoriels associés aux triplets statistiques  $(X, Q, D)$  et  $(X, Q, \tilde{D})$ .

Les matrices de produits scalaires entre u.s. pour les matrices de poids  $D$  et  $\tilde{D}$  s'écrivent respectivement :

$$W = (I-ee'D) XQX'(I-Dee') \quad \text{et} \quad \tilde{W} = (I-ee'\tilde{D}) XQX'(I-\tilde{D}ee').$$

Considérons les opérateurs  $D^{1/2} W D^{1/2}$  et  $\tilde{D}^{1/2} \tilde{W} \tilde{D}^{1/2}$  qui sont au plus de rang  $p$  par construction et désignons par  $\lambda_i$  et  $\tilde{\lambda}_i$ ,  $i = 1, \dots, p$  leurs  $p$  plus grandes valeurs propres respectives. On a alors le résultat suivant :

**Proposition 3.2.**

Pour  $i = 1, \dots, p$  les inégalités suivantes sont vérifiées

(a) si  $D$  est inversible ( $p_k \neq 0, \forall k = 1, \dots, n$ ) :

$$\tilde{\lambda}_i \leq \lambda_i \left[ \max_k \frac{\tilde{p}_k}{p_k} \right]$$

(b) si  $D$  et  $\tilde{D}$  sont inversibles ( $p_k \neq 0$  et  $\tilde{p}_k \neq 0, \forall k = 1, \dots, n$ ) :

$$\lambda_i \left[ \min_k \frac{\tilde{p}_k}{p_k} \right] \leq \tilde{\lambda}_i$$

**Preuve**

Montrons tout d'abord (a). Il est immédiat de constater que

$$\tilde{W} = (I-ee'\tilde{D}) W (I-\tilde{D}ee')$$

car  $(I-ee'\tilde{D})(I-ee'D) = (I-ee'\tilde{D})$ .

Aussi avec la notation naturelle  $D^{-1/2} = [D^{1/2}]^{-1} = \text{diag}(1/\sqrt{p_i})$  on peut écrire :

$$\tilde{D}^{1/2} \tilde{W} \tilde{D}^{1/2} = [\tilde{D}^{1/2} (I-ee'\tilde{D}) D^{-1/2}] (D^{1/2} W D^{1/2}) [D^{-1/2} (I-\tilde{D}ee') \tilde{D}^{1/2}]$$

et cet opérateur  $\tilde{D}^{1/2} \tilde{W} \tilde{D}^{1/2}$  a alors les mêmes valeurs propres que :

$$(D^{1/2} W D^{1/2}) [D^{-1/2} (I-\tilde{D}ee') \tilde{D}^{1/2}] [\tilde{D}^{1/2} (I-ee'\tilde{D}) D^{-1/2}]$$

i.e.

$$(D^{1/2} W D^{1/2}) [D^{-1/2} \tilde{D}^{1/2} (I-\tilde{D}^{1/2} ee'\tilde{D}^{1/2}) \tilde{D}^{1/2} D^{-1/2}] \quad (1)$$

Posons  $A = D^{-1/2} \tilde{D}^{1/2} (I-\tilde{D}^{1/2} ee'\tilde{D}^{1/2}) \tilde{D}^{1/2} D^{-1/2}$ . Cet opérateur a les mêmes valeurs propres que  $(I-\tilde{D}^{1/2} ee'\tilde{D}^{1/2}) \tilde{D} D^{-1}$ . Or d'après le lemme 3.1,  $I-\tilde{D}^{1/2} ee'\tilde{D}^{1/2}$  est un projecteur et sa plus grande valeur propre est donc égale à l'unité. Aussi en appliquant la relation (6) de la proposition 2.1 à  $A$  on obtient :

$$\lambda_1(A) \leq \lambda_1(\tilde{D} D^{-1}) \Leftrightarrow \lambda_1(A) \leq \max_k \frac{\tilde{p}_k}{p_k} \quad (2)$$

Comme  $\tilde{D}^{1/2} \tilde{W} \tilde{D}^{1/2}$  a les mêmes valeurs propres que l'expression matricielle (1), la proposition 2.1 permet encore d'obtenir pour  $i = 1, \dots, p$  :

$$\tilde{\lambda}_i \leq \lambda_i \lambda_1(A) \Rightarrow \tilde{\lambda}_i \leq \lambda_i \max_k \frac{\tilde{p}_k}{p_k}$$

Pour démontrer la relation (b) on utilise la relation (a). La matrice de poids  $\tilde{D}$  étant supposée inversible on obtient à partir de (a) en permutant les rôles de  $D$  et  $\tilde{D}$  :

$$\lambda_i \leq \tilde{\lambda}_i \max_k \frac{p_k}{\tilde{p}_k} \Leftrightarrow \lambda_i \leq \tilde{\lambda}_i \left[ \min_k \frac{\tilde{p}_k}{p_k} \right]^{-1} \Leftrightarrow \lambda_i \min_k \frac{\tilde{p}_k}{p_k} \leq \tilde{\lambda}_i$$

c'est-à-dire la relation désirée.

On remarquera que si l'on suppose simultanément  $D$  et  $\tilde{D}$  inversibles on a donc

$$\lambda_i \min_k \frac{\tilde{p}_k}{p_k} \leq \tilde{\lambda}_i \leq \lambda_i \max_k \frac{\tilde{p}_k}{p_k} \quad i = 1, \dots, p \quad (3)$$

Il est d'autre part trivial de constater que la transformation de  $D$  en  $\tilde{D}$  entraîne une modification des variances  $s_j^2$  des variables en  $\tilde{s}_j^2$  ( $j = 1, \dots, p$ ). En pratique lorsqu'on travaille avec la métrique

$$Q = \text{diag } 1/s_j^2$$

celle-ci se trouve modifiée en

$$\tilde{Q} = \text{diag } 1/\tilde{s}_j^2$$

et à partir du triplet statistique initial  $(X, Q, D)$  on obtient donc le nouveau triplet  $(X, \tilde{Q}, \tilde{D})$  auquel se trouve alors associées les valeurs propres  $\tilde{\lambda}_i$ . Il est possible de comparer les  $\tilde{\lambda}_i$  aux valeurs propres  $\lambda_i$  de  $(X, Q, \tilde{D})$  en considérant la matrice de variance  $\tilde{V} = X'(I - \tilde{D}e'e') \tilde{D} (I - ee'\tilde{D}) X = X'(\tilde{D} - \tilde{D}e'e'\tilde{D}) X$ . On constate simplement que l'opérateur  $\tilde{Q}^{1/2} \tilde{V} \tilde{Q}^{1/2}$  a les mêmes valeurs propres que  $(Q^{1/2} \tilde{V} Q^{1/2}) (\tilde{Q} \tilde{Q}^{-1})$ . La proposition 2.1 permet alors d'obtenir pour  $i = 1, \dots, p$

$$\tilde{\lambda}_i \min_j \frac{s_j^2}{\tilde{s}_j^2} \leq \tilde{\lambda}_i \leq \lambda_i \max_j \frac{s_j^2}{\tilde{s}_j^2} \quad (4)$$

$D$  et  $\tilde{D}$  étant supposées inversibles la combinaison de (3) et (4) permet d'avoir en définitive

$$\tilde{\lambda}_i \min_j \frac{s_j^2}{\tilde{s}_j^2} \min_k \frac{\tilde{p}_k}{p_k} \leq \tilde{\lambda}_i \leq \lambda_i \max_j \frac{s_j^2}{\tilde{s}_j^2} \max_k \frac{\tilde{p}_k}{p_k}$$

D'une manière générale on remarquera que l'amplitude de la variation des valeurs propres semble tributaire des rapports extrêmes des poids et des variances. Il semble donc qu'une perturbation importante du poids d'une seule u.s. (les autres poids étant seulement normalisés) peut avoir des conséquences plus sensibles sur les valeurs propres qu'une perturbation moyenne de l'ensemble des poids.

#### 4. MODIFICATION DU POIDS D'UNE UNITÉ STATISTIQUE EN A.C.P. SUR MATRICE DE VARIANCE

En pratique envisager une modification quelconque de  $D$  en  $\tilde{D}$  n'amène souvent que peu de renseignement car l'A.C.P. effectuée avec  $\tilde{D}$  est en général trop différente de l'A.C.P. initiale faite avec  $D$ , pour que la comparaison présente un réel intérêt. Il est souvent plus instructif de s'intéresser aux variations de poids d'une u.s.  $j$  particulière. En modifiant de manière progres-

sive le poids de cette u.s. on peut alors réellement apprécier son influence sur l'A.C.P. et en particulier, comme nous allons le faire ici, son influence sur les valeurs propres qui correspondent aux inerties des axes factoriels. Nous étudions tout d'abord dans ce paragraphe le cas où la métrique Q est supposée être l'identité. Tout revient alors à travailler sur la matrice de variance.

Lorsqu'on transforme le poids  $p_j$  de l'u.s. j en  $\tilde{p}_j$  quelconque ( $0 \leq \tilde{p}_j < 1$ ) les poids  $p_i$  ( $i \neq j$ ) des autres u.s. doivent être normalisés en  $\tilde{p}_i = \alpha_j p_i$  avec  $\alpha_j = (1 - \tilde{p}_j)/(1 - p_j)$  afin d'avoir  $\sum_{i=1}^n \tilde{p}_i = 1$ . Soient alors  $g$  et  $\tilde{g}$  les centres de gravités des u.s. pour les poids de  $D = \text{diag } p_i$  et  $\tilde{D} = \text{diag } \tilde{p}_i$  respectivement :

$$g = \sum_{k=1}^n p_k x_k, \quad \tilde{g} = \sum_{k=1}^n \tilde{p}_k x_k.$$

Les matrices de variance associées aux études statistiques  $(X, I, D)$  et  $(X, I, \tilde{D})$  peuvent respectivement s'exprimer sous la forme

$$V = \sum_{k=1}^n p_k (x_k - g)' (x_k - g) \quad \text{et} \quad \tilde{V} = \sum_{k=1}^n \tilde{p}_k (x_k - \tilde{g})' (x_k - \tilde{g})$$

#### 4.1. Relation entre V et $\tilde{V}$

Le théorème d'Huygens permet encore d'exprimer  $\tilde{V}$  sous la forme :

$$\begin{aligned} \tilde{V} &= \sum_{k=1}^n \tilde{p}_k (x_k - g)' (x_k - g) - (g - \tilde{g})' (g - \tilde{g}) \\ &= \alpha_j V + (\tilde{p}_j - \alpha_j p_j) (x_j - g)' (x_j - g) - (g - \tilde{g})' (g - \tilde{g}) \end{aligned}$$

En posant  $\beta_j = 1 - \alpha_j = (\tilde{p}_j - p_j)/(1 - p_j)$ , il est facile de voir que :

$$\tilde{p}_j - \alpha_j p_j = \beta_j, \quad \tilde{g} = \alpha_j g + \beta_j x_j \quad \text{et} \quad g - \tilde{g} = -\beta_j (x_j - g)$$

et l'on obtient alors l'expression suivante pour  $\tilde{V}$  :

$$\tilde{V} = \alpha_j \{V + \beta_j (x_j - g)' (x_j - g)\} \quad (4.1)$$

qui permet d'exprimer très simplement les nouvelles variances et covariances à partir des anciennes.

#### 4.2. Valeurs propres de V et $\tilde{V}$

Il est facile de donner l'expression des valeurs propres de

$$B = \beta_j (x_j - g)' (x_j - g)$$

en utilisant la norme habituelle  $\| \cdot \|$  sur  $\mathbf{R}^p$  :

$$\begin{aligned} \text{si} \quad & \beta_j > 0 \quad (\tilde{p}_j > p_j) & \lambda_1(B) &= \beta_j \|x_j - g\|^2 \\ \text{et} & & \lambda_s(B) &= 0 \quad s = 2, \dots, p \\ \text{si} \quad & \beta_j < 0 \quad (\tilde{p}_j < p_j) & \lambda_s(B) &= 0 \quad s = 1, \dots, p-1 \\ \text{et} & & \lambda_p(B) &= \beta_j \|x_j - g\|^2 \end{aligned}$$

La proposition 2.1.I appliquée à la relation (4.1) précédente (avec  $A = V$ ) permet d'obtenir toute une famille d'encadrements des valeurs propres de  $\tilde{V}$  par celles de V. Les plus intéressants, qui correspondent à la relation (3) de la proposition et aux relations (1) et (2) avec  $j = i - 1$  et  $k = 2$  sont les suivants :

dans le cas  $\beta_j > 0$   
 pour  $i = 1, \dots, p$   $\alpha_j \lambda_i(V) \leq \lambda_i(\tilde{V}) \leq \alpha_j \{\lambda_i(V) + \beta_j \|x_j - g\|^2\}$  (4.2)

pour  $i = 2, \dots, p$   $\lambda_i(\tilde{V}) \leq \alpha_j \lambda_{i-1}(V)$  (4.3)

dans le cas  $\beta_j < 0$   
 pour  $i = 1, \dots, p$   $\alpha_j \{\lambda_i(V) + \beta_j \|x_j - g\|^2\} \leq \lambda_i(\tilde{V}) \leq \alpha_j \lambda_i$  (4.4)

pour  $i = 1, \dots, p - 1$   $\alpha_j \lambda_{i+1}(V) \leq \lambda_i(\tilde{V})$  (4.5)

Les inégalités (4.2) et (4.4) valables pour n'importe quelle valeur propre permettent de donner des bornes à la  $i$ -ième valeur propre de  $\tilde{V}$  en utilisant la  $i$ -ième valeur propre de  $V$ . Les formules (4.3) et (4.5) sont intéressantes par leur simplicité. Leur précision peut éventuellement être meilleure que celle de (4.2) ou (4.4) si  $\lambda_i(V)$  est très proche de  $\lambda_{i-1}(V)$  ou  $\lambda_{i+1}(V)$ .

Dans le cas où  $V$  est inversible, il est possible d'obtenir d'autres encadrements faisant intervenir la norme de Mahalanobis. En décomposant  $V$  définie positive sous la forme  $V = SS'$ , on constate à partir de la relation (4.1) que  $\tilde{V}$ , qui peut s'écrire  $\tilde{V} = \alpha_j S \{I + \beta_j S^{-1} (x_j - g)' (x_j - g) (S')^{-1}\} S'$ , a les mêmes valeurs propres que :

$$\alpha_j (S'S) \{I + \beta_j S^{-1} (x_j - g)' (x_j - g) (S')^{-1}\} \quad (4.6)$$

Cette expression fait apparaître le produit des deux matrices symétriques  $A = S'S$  (qui a les mêmes valeurs propres que  $V$ ) et

$$B = I + \beta_j S^{-1} (x_j - g)' (x_j - g) (S')^{-1}.$$

Les valeurs propres de  $B$  s'expriment simplement à partir de :

$$\|x_j - g\|_{V^{-1}}^2 = (x_j - g)' V^{-1} (x_j - g) :$$

si  $\beta_j > 0$  :  $\lambda_1(B) = 1 + \beta_j \|x_j - g\|_{V^{-1}}^2$ ;  $\lambda_s(B) = 1$ ,  $s = 2, \dots, p$

si  $\beta_j < 0$  :  $\lambda_s(B) = 0$ ,  $s = 1, \dots, p - 1$ ;  $\lambda_p(B) = 1 + \beta_j \|x_j - g\|_{V^{-1}}^2$

L'application de la proposition 2.1 II à l'expression (4.6) précédente permet alors d'obtenir les inégalités suivantes :

si  $\beta_j > 0$   $\lambda_i(\tilde{V}) \leq \alpha_j \{\lambda_i(V)\} \{1 + \beta_j \|x_j - g\|_{V^{-1}}^2\}$  (4.7)

si  $\beta_j < 0$   $\lambda_i(\tilde{V}) \geq \alpha_j \{\lambda_i(V)\} \{1 + \beta_j \|x_j - g\|_{V^{-1}}^2\}$  (4.8)

On remarquera que d'un point de vue théorique la précision de ces deux dernières relations est difficile à comparer avec celle de la majoration de (4.2) ou de la minoration de (4.4).

## 5. MODIFICATION DU POIDS D'UNE UNITÉ STATISTIQUE EN A.C.P. SUR MATRICE DE CORRÉLATION

Nous prolongeons ici les résultats du paragraphe précédent au cas où  $Q = \text{diag}(1/s_k^2)$  avec  $s_k^2$  variance de la variable  $k$ , pour  $k = 1, \dots, p$ . La modification du poids de l'u.s.  $j$  entraîne alors une transformation de  $Q$  en  $\tilde{Q} = \text{diag}(1/\tilde{s}_k^2)$ . Nous allons tout d'abord chercher une relation entre les matrices de corrélation  $R = (r_{k\ell}) = Q^{1/2} V Q^{1/2}$  et  $\tilde{R} = (\tilde{r}_{k\ell}) = \tilde{Q}^{1/2} \tilde{V} \tilde{Q}^{1/2}$  avant et après modification du poids de l'u.s.  $j$  ( $Q^{1/2} = \text{diag } 1/s_k$ ,  $\tilde{Q} = \text{diag } 1/\tilde{s}_k$ ).



### 5.1. Relation entre R et $\tilde{R}$

En utilisant la relation (4.1) liant V et  $\tilde{V}$  on peut écrire :

$$\tilde{R} = \tilde{Q}^{1/2} \tilde{V} \tilde{Q}^{1/2} = \alpha_j (\tilde{Q}^{1/2} Q^{-1/2}) \{R + \beta_j Q^{1/2} (x_j - g)' (x_j - g) Q^{1/2}\} (Q^{-1/2} \tilde{Q}^{1/2}) \quad (5.1)$$

Notons  $y_j$  l'u.s. j centrée réduite pour les poids initiaux :

$$y_j' = Q^{1/2} (x_j - g)' = (y_{j1}, \dots, y_{jp})'$$

avec  $y_{jk} = (x_{jk} - g_k)/s_k$  et considérons la matrice  $T = \sqrt{\alpha_j} (\tilde{Q}^{1/2} Q^{-1/2}) = \text{diag} t_k$ .

La relation (4.1) permet de voir que  $\tilde{s}_k^2 = \alpha_j [s_k^2 + \beta_j (x_{jk} - g_k)^2]$  on en déduit  $t_k = \sqrt{\alpha_j} (s_k/\tilde{s}_k) = (1 + \beta_j y_{jk}^2)^{-1/2}$ . La relation (5.1) devient alors :

$$\tilde{R} = T(R + \beta_j y_j' y_j) T \quad (5.2)$$

Elle se traduit en terme de coefficient de corrélation par :

$$\tilde{r}_{k\ell} = \frac{r_{k\ell} + \beta_j y_{jk} y_{j\ell}}{\{(1 + \beta_j y_{jk}^2)(1 + \beta_j y_{j\ell}^2)\}^{1/2}} \quad (5.3)$$

Si  $y_{jk}$  et  $y_{j\ell}$  sont proches de zéro, dans le cas fréquent où  $p_i = 1/n$   $i = 1, \dots, n$ , on peut retrouver par un développement limité de (5.3) l'approximation suivante utilisée de manière classique en statistique (voir par exemple [3] ou [4]) :

$$(n - 1) (r_{k\ell} - \tilde{r}_{k\ell}) \approx - \frac{1}{2} (y_{jk}^2 + y_{j\ell}^2) r_{k\ell} + y_{jk} y_{j\ell}$$

### 5.2. Valeurs propres de R et $\tilde{R}$

Comme pour l'étude de V, il est facile de constater que :

si  $\beta_j > 0$   $\lambda_1(\beta_j y_j' y_j) = \beta_j \|y_j\|^2$  et  $\lambda_s(\beta_j y_j' y_j) = 0$   $s = 2, \dots, p$

si  $\beta_j < 0$   $\lambda_s(\beta_j y_j' y_j) = 0$   $s = 1, \dots, p - 1$   $\lambda_p(\beta_j y_j' y_j) = \beta_j \|y_j\|^2$

Alors en remarquant d'après (5.2) que  $\tilde{R}$  a les mêmes valeurs propres que  $(R + \beta_j y_j' y_j) T^2$ , on obtient, en utilisant conjointement les parties I et II de la proposition 2.1, une famille d'encadrements des valeurs propres de  $\tilde{R}$  dont les plus intéressants sont les suivants pour  $i \leq p$  :

si  $\beta_j > 0$

$$\{\lambda_i(R)\} \left\{ \min_k (1 + \beta_j y_{jk}^2)^{-1} \right\} \leq \lambda_i(\tilde{R}) \leq \left\{ \lambda_i(R) + \beta_j \|y_j\|^2 \right\} \left\{ \max_k (1 + \beta_j y_{jk}^2)^{-1} \right\} \quad (5.4)$$

$$\lambda_i(\tilde{R}) \leq \{\lambda_{i-1}(R)\} \left\{ \max_k (1 + \beta_j y_{jk}^2)^{-1} \right\} \text{ pour } i \leq p - 2$$

si  $\beta_j < 0$

$$\{\lambda_i(R) + \beta_j \|y_j\|^2\} \left\{ \min_k (1 + \beta_j y_{jk}^2)^{-1} \right\} \leq \lambda_i(\tilde{R}) \leq \left\{ \lambda_i(R) \right\} \left\{ \max_k (1 + \beta_j y_{jk}^2)^{-1} \right\} \quad (5.5)$$

$$\{\lambda_{i+1}(R)\} \left\{ \min_k (1 + \beta_j y_{jk}^2)^{-1} \right\} \leq \lambda_i(\tilde{R}) \text{ pour } i \leq p - 1$$

De plus comme dans le cas de la matrice de variance, il est possible d'obtenir d'autres encadrements lorsque R est de plein rang. En décomposant R sous la forme  $R = SS'$  on voit à partir de la relation (5.2) que  $\tilde{R}$  a mêmes valeurs propres que  $BT^2$  avec  $B = S(I + \beta_j S^{-1} y_j' y_j (S')^{-1}) S'$ .

La proposition 2.1 permet alors d'écrire pour  $i \leq p$  :

$$\{\min_k (1 + \beta_j y_{jk}^2)^{-1}\} \lambda_i(\mathbf{B}) \leq \lambda_i(\tilde{\mathbf{R}}) \leq \lambda_i(\mathbf{B}) \{\max_k (1 + \beta_j y_{jk}^2)^{-1}\}$$

et

$$\lambda_p(\mathbf{I} + \beta_j \mathbf{S}^{-1} y_j' y_j (\mathbf{S}')^{-1}) \lambda_i(\mathbf{R}) \leq \lambda_i(\mathbf{B}) \leq \lambda_i(\mathbf{R}) \lambda_i(\mathbf{I} + \beta_j \mathbf{S}^{-1} y_j' y_j (\mathbf{S}')^{-1})$$

En explicitant comme au paragraphe 4 les valeurs propres de  $\mathbf{I} + \beta_j \mathbf{S}^{-1} y_j' y_j (\mathbf{S}')^{-1}$  à partir de  $\|y_j\|_{\mathbf{R}^{-1}}^2 = y_j \mathbf{R}^{-1} y_j'$ , on obtient :

si  $\beta_j > 0$

$$\lambda_i(\tilde{\mathbf{R}}) \leq \{\lambda_i(\mathbf{R})\} \{1 + \beta_j \|y_j\|_{\mathbf{R}^{-1}}^2\} \{\max_k (1 + \beta_j y_{jk}^2)^{-1}\} \quad (5.6)$$

si  $\beta_j < 0$

$$\lambda_i(\tilde{\mathbf{R}}) \geq \{\lambda_i(\mathbf{R})\} \{1 + \beta_j \|y_j\|_{\mathbf{R}^{-1}}^2\} \{\min_k (1 + \beta_j y_{jk}^2)^{-1}\} \quad (5.7)$$

## 6. ASPECT GÉOMÉTRIQUE DE LA MODIFICATION DU POIDS D'UNE UNITÉ STATISTIQUE

L'objet de ce paragraphe est de souligner l'aspect géométrique sous-jacent aux résultats du paragraphe 5. Etant donné un réel positif  $\theta$ , on propose à partir des encadrements précédents, des conditions géométriques, sur l'u.s.  $j$  dont on perturbe le poids, qui assurent que cette modification n'entraînera pas de perturbation des valeurs propres d'amplitude supérieure à  $\theta$ . Outre les résultats précédents, ces développements géométriques utilisent le lemme suivant dont la démonstration ne présente aucune difficulté :

### Lemme 6.1

- (i) si  $\beta_j > 0$  ( $\tilde{p}_j > p_j$ )  $\lambda_1(\mathbf{T}^2) = \max_{\ell} \{(1 + \beta_j y_{j\ell}^2)^{-1}\} \leq 1$
- (ii) si  $\beta_j < 0$  ( $\tilde{p}_j < p_j$ )  $\lambda_p(\mathbf{T}^2) = \min_{\ell} \{(1 + \beta_j y_{j\ell}^2)^{-1}\} \geq 1$

Dans le cas d'une augmentation de poids ( $\beta_j > 0$ ) on peut alors formuler les deux propositions qui suivent :

### Proposition 6.2.

Une augmentation du poids  $p_j$  en  $\tilde{p}_j$  ( $\beta_j > 0$ ) n'entraînera pas une augmentation de plus de  $\theta \in \mathbf{R}^{+*}$  fixé, pour chaque valeur propre, si l'u.s. centrée réduite  $y_j$  est dans la sphère  $S_{\theta j}$  de  $\mathbf{R}^p$  de rayon  $r_{\theta j} = \sqrt{\theta/\beta_j}$  centrée sur l'origine.

### Proposition 6.3.

On considère une valeur propre  $\lambda_i(\mathbf{R})$  particulière vérifiant

$$\lambda_i(\mathbf{R}) > \theta \quad (\theta \in \mathbf{R}^{+*}).$$

Alors, si l'u.s. centrée réduite  $y_j$  est dans le cube  $C_{i\theta j}$  de  $\mathbf{R}^p$  défini par

$$C_{i\theta j} = \{z = (z_1, \dots, z_p) \in \mathbf{R}^p / \max_k |z_k| \leq a_{i\theta j}\}$$

avec  $a_{i\theta_j} = [\theta/\{\beta_j(\lambda_i(\mathbf{R}) - \theta)\}]^{1/2}$ , une augmentation de  $p_j$  en  $\tilde{p}_j$  ne pourra pas entraîner une diminution de la  $i$ -ième valeur propre de plus de  $\theta$ .

La démonstration de ces deux propositions ne présentant aucune difficulté, nous nous contenterons à titre illustratif de donner la preuve de la proposition 6.2.

Supposons que  $y_j$  appartienne à la sphère  $S_{\theta_j}$ . Alors

$$\|y_j\|^2 \leq r_{\theta_j}^2 = \theta/\beta_j \Leftrightarrow \beta_j \|y_j\|^2 \leq \theta$$

On obtient par conséquent pour tout  $i \leq p$  :

$$\lambda_i(\mathbf{R}) + \beta_j \|y_j\|^2 \leq \lambda_i(\mathbf{R}) + \theta$$

On déduit de la relation (i) du lemme précédent :

$$\{\lambda_i(\mathbf{R}) + \beta_j \|y_j\|^2\} \max_{\ell} \{(1 + \beta_j y_{j\ell}^2)^{-1}\} \leq \lambda_i(\mathbf{R}) + \theta$$

i.e.  $\lambda_i(\tilde{\mathbf{R}}) \leq \lambda_i(\mathbf{R}) + \theta$  (en utilisant la relation (5.4)) ce qui achève la preuve.

Les propositions 6.2 et 6.3 permettent d'affirmer que lorsqu'on augmente le poids d'une u.s., la  $i$ -ième valeur propre ne variera pas de plus de  $\theta$  si  $y_j \in C_{i\theta_j} \cap S_{\theta_j}$ .

D'une manière analogue dans le cas d'une diminution de poids on peut énoncer les deux propositions suivantes :

#### Proposition 6.4.

Si l'u.s. centrée réduite  $y_j$  appartient à la sphère  $S_{\theta_j}$  de  $\mathbf{R}^p$  de rayon  $r_{\theta_j} = \sqrt{(-\theta/\beta_j)}$ , centrée sur l'origine, une diminution de son poids  $p_j$  en  $\tilde{p}_j$  ne pourra entraîner une diminution de plus de  $\theta$  de chacune des valeurs propres.

#### Proposition 6.5.

Si l'u.s. centrée réduite  $y_j$  appartient au cube de  $\mathbf{R}^p$  défini par

$$C_{i\theta_j} = \{z = (z_1, \dots, z_p) \in \mathbf{R}^p / \max_k |z_k| \leq a_{i\theta_j}\}$$

avec  $a_{i\theta_j} = [(-\theta)/\{\beta_j(\lambda_i(\mathbf{R}) + \theta)\}]^{1/2}$  une diminution de son poids de  $p_j$  en  $\tilde{p}_j$  ne pourra entraîner une augmentation de plus de  $\theta$  de la  $i$ -ième valeur propre.

## 7. ASPECTS PRATIQUES ET GRAPHIQUES

On remarquera que les résultats précédents répondent en particulier au problème pratique de la suppression ou de l'ajout d'une u.s. Il est intéressant d'examiner ces deux situations dans le cas le plus courant où les poids  $p_i$  sont supposés tous égaux. La suppression d'une u.s.  $j$  est alors caractérisée par la considération des poids suivants :  $p_i = 1/n$ ,  $i = 1, \dots, n$ ,  $\tilde{p}_j = 0$ ,  $\tilde{p}_i = 1/(n-1)$ ,  $i = 1, \dots, n$ ,  $i \neq j$  et l'on a  $\alpha_j = n/(n-1)$ ,  $\beta_j = -1/(n-1)$ . Les résultats obtenus dans ce cas peuvent alors être rattachés au problème général du jackknife. D'une manière symétrique lorsqu'on veut ajouter une  $(n+1)^{\text{ième}}$  u.s. aux  $n$  u.s. initiales, on peut considérer que cette u.s. est initialement affectée d'un poids nul et la situation se définit alors par :

$$p_i = 1/n, \quad i = 1, \dots, n, \quad p_{n+1} = 0, \quad \tilde{p}_i = \frac{1}{n+1}, \quad i = 1, \dots, n+1,$$

$$\alpha_{n+1} = \frac{n}{n+1}, \quad \beta_{n+1} = \frac{1}{n+1}$$

Si l'on veut étudier d'une manière plus nuancée l'influence d'une u.s. sur les valeurs propres il est possible de s'aider de graphiques donnant le pourcentage maximal de variation des valeurs propres en fonction de l'u.s. dont on désire modifier le poids et de la variation de poids envisagée. Cette approche graphique n'est valable que lorsqu'on travaille sur une matrice de corrélation de plein rang.

Lorsqu'on s'intéresse à une augmentation de l'influence d'une u.s. sur les valeurs propres, l'approche se fonde sur l'utilisation conjointe de la relation (5.6) et du résultat (i) du lemme 6.1 qui permettent d'écrire puisque  $\beta_j > 0$  :

$$\lambda_i(\tilde{\mathbf{R}}) \leq \{\lambda_i(\mathbf{R})\} \{1 + \beta_j \|y_j\|_{R-1}^2\} \quad (7.1)$$

Alors pour que l'augmentation du poids  $p_j = 1/n$  en  $\tilde{p}_j$  avec  $p_j < \tilde{p}_j < 1$  n'entraîne pas une augmentation de la  $i$ -ième valeur propre de plus de 100  $\rho$  % avec  $0 < \rho \leq 1$  il suffit d'avoir d'après la relation (7.1) précédente :

$$\{\lambda_i(\mathbf{R})\} \{1 + \beta_j \|y_j\|_{R-1}^2\} \leq (1 + \rho) \{\lambda_i(\mathbf{R})\}$$

i.e. 
$$\beta_j \|y_j\|_{R-1}^2 \leq \rho$$

Si l'on trace pour différentes valeurs de  $\rho$ , les hyperboles  $\|y_j\|_{R-1}^2 = \rho/\beta_j$ , ces dernières permettent de mettre en relation graphique la variation de poids et l'u.s. sur laquelle elle porte avec le pourcentage maximum d'augmentation qui peut en découler pour n'importe laquelle des valeurs propres. Les tracés correspondant aux valeurs  $\rho = 0.01, 0.05, 0.1$  sont donnés dans la figure 1.

L'étude du pourcentage maximum de diminution des valeurs propres (toujours dans le cadre d'une augmentation du poids d'une u.s.), se fonde sur l'utilisation du membre de gauche de la relation (5.4).

Pour  $\beta_j > 0$  on a :

$$\{\lambda_i(\mathbf{R})\} \{\min_k (1 + \beta_j y_{jk}^2)^{-1}\} \leq \lambda_i(\tilde{\mathbf{R}}) \Leftrightarrow \{\lambda_i(\mathbf{R})\} \{(1 + \beta_j \max_k y_{jk}^2)^{-1}\} \leq \lambda_i(\tilde{\mathbf{R}})$$

Alors pour avoir  $\lambda_i(\tilde{\mathbf{R}}) \geq (1 - \rho) \lambda_i(\mathbf{R})$  il suffit que

$$\{\lambda_i(\mathbf{R})\} \{(1 + \beta_j \max_k y_{jk}^2)^{-1}\} \geq (1 - \rho) \lambda_i(\mathbf{R})$$

i.e. 
$$\beta_j \max_k y_{jk}^2 \leq \frac{\rho}{1 - \rho}$$

Si l'on trace à nouveau pour différentes valeurs de  $\rho$  les hyperboles  $\max_k y_{jk}^2 = \rho/\beta_j(1 - \rho)$ , ces dernières permettent comme précédemment de mettre en relation graphique la variation de poids et l'u.s. sur laquelle elle porte avec le pourcentage maximum de diminution qui peut en résulter pour n'importe quelle valeur propre. Les tracés correspondant à  $\rho = 0.01, 0.05, 0.1$  sont donnés en figure 2. On notera la ressemblance des tracés des figures 1 et 2 liée au comportement voisin de  $\rho$  et  $\rho/(1 - \rho)$  pour les petites valeurs de  $\rho$ .

Dans le cas d'une diminution de  $p_j$  en  $\tilde{p}_j$  ( $\beta_j < 0$ ) on obtient les contours hyperboliques  $\|y_j\|_{R-1}^2 = -\rho/\beta_j$  et  $\max_k y_{jk}^2 = -\rho/\beta_j(1 + \rho)$  qui correspondent respectivement au pourcentage maximal de diminution et d'augmentation de n'importe quelle valeur propre.

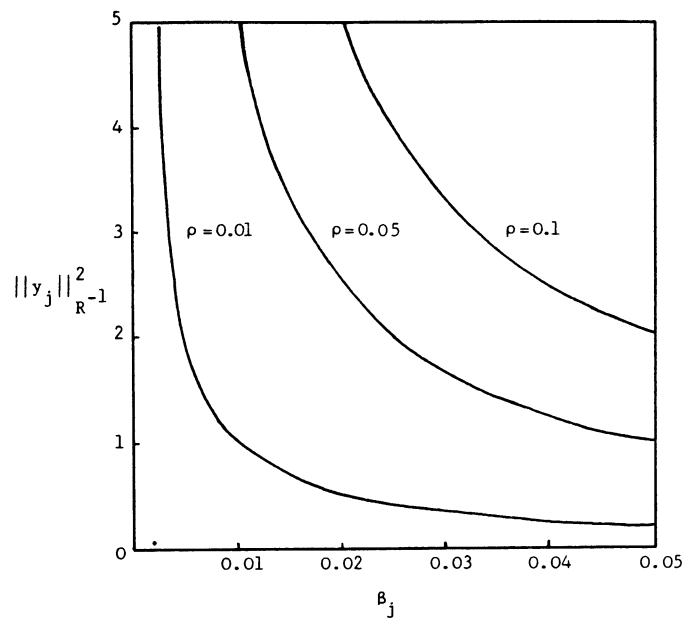


Figure 1. — Pourcentage maximum d'augmentation d'une valeur propre lorsque le poids  $p_j$  augmente jusqu'à  $\bar{p}_j$ .

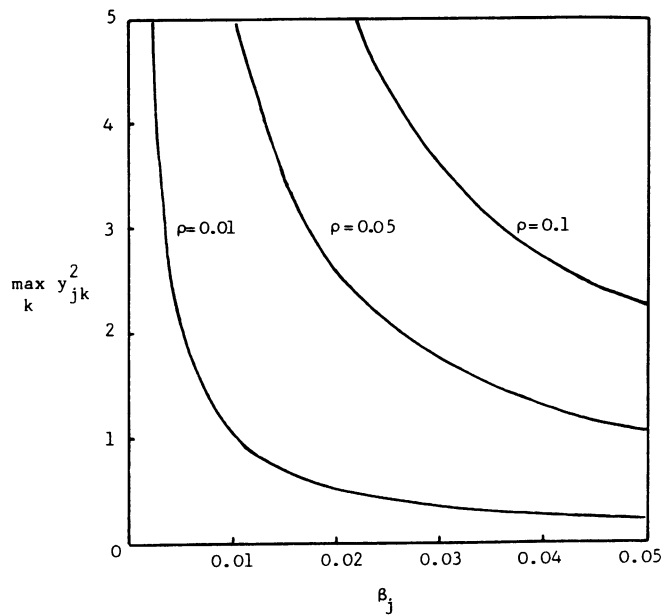


Figure 2. — Pourcentage maximum de diminution d'une valeur propre lorsque le poids  $p_j$  augmente jusqu'à  $\bar{p}_j$ .

## 8. ILLUSTRATION

En conclusion nous proposons une brève illustration de certains des encadrements des valeurs propres que nous avons obtenus. Pour cela nous avons choisi l'exemple des « poissons d'Amiard » que l'ouvrage de PAGES et CAILLIEZ [6] a rendu classique et auquel nous renvoyons pour une présentation détaillée. Dans un but de concision nous nous bornerons à rappeler que les données sont constituées par un ensemble de 24 poissons placés dans un milieu radioactif et dont on mesure à la fois la taille et la teneur en radioactivité de différents organes (16 variables au total). Dans cette même optique de concision nous limitons notre illustration aux deux plus grandes valeurs propres et aux encadrements les plus simples fournis par les relations (5.4) et (5.5). Cette approche sera toutefois suffisante pour dégager les aspects les plus importants.

Le poisson 17 mort en cours d'expérience n'est pas pris en compte dans le traitement statistique si bien qu'un poids de  $1/23$  est initialement attribué aux 23 poissons restants. Nous étudions les variations des valeurs propres engendrées par la modification du poids du poisson 20 qui dans l'A.C.P. initiale contribue de la manière la plus importante au premier axe factoriel.

Comme le poids initial du poisson 20 est  $p_{20} = 1/23 = 92/2116$  on envisage de manière arbitraire des modifications successives d'amplitude  $22/2116$  qui ont pour effet de faire varier à chaque pas les poids des autres poissons de  $1/2116$ . Les résultats sont résumés dans le tableau qui suit.

|  |                   |                   |                   |                   |                                  |                    |                    |                    |                    |
|--|-------------------|-------------------|-------------------|-------------------|----------------------------------|--------------------|--------------------|--------------------|--------------------|
| $\tilde{p}_{20}$                         | $\frac{4}{2116}$  | $\frac{26}{2116}$ | $\frac{48}{2116}$ | $\frac{70}{2116}$ | $\frac{1}{23} = \frac{92}{2116}$ | $\frac{114}{2116}$ | $\frac{136}{2116}$ | $\frac{158}{2116}$ | $\frac{180}{2116}$ |
| $\tilde{p}_i \quad i \neq 20$            | $\frac{96}{2116}$ | $\frac{95}{2116}$ | $\frac{94}{2116}$ | $\frac{93}{2116}$ | $\frac{92}{2116}$                | $\frac{91}{2116}$  | $\frac{90}{2116}$  | $\frac{89}{2116}$  | $\frac{88}{2116}$  |
| $\min_k (1 + \beta_{20} y_{20k}^2)^{-1}$ | 1.018             | 1.013             | 1.009             | 1.004             | 1.                               | 0.914              | 0.841              | 0.779              | 0.726              |
| $\max_k (1 + \beta_{20} y_{20k}^2)^{-1}$ | 1.608             | 1.396             | 1.233             | 1.104             | 1.                               | 0.996              | 0.991              | 0.987              | 0.983              |
| $\lambda_1(\tilde{R})$                   | 6.938             | 7.107             | 7.278             | 7.445             | 7.607                            | 7.764              | 7.915              | 8.060              | 8.119              |
| $m_1$                                    | 5.481             | 6.032             | 6.559             | 7.092             | 7.607                            | 6.947              | 6.399              | 5.928              | 5.520              |
| $M_1$                                    | 12.234            | 10.611            | 9.382             | 8.401             | 7.607                            | 8.112              | 8.587              | 9.078              | 9.554              |
| $\lambda_2(\tilde{R})$                   | 4.121             | 4.023             | 3.931             | 3.845             | 3.763                            | 3.686              | 3.613              | 3.543              | 3.476              |
| $m_2$                                    | 1.562             | 2.134             | 2.691             | 3.230             | 3.763                            | 3.438              | 3.161              | 2.932              | 2.731              |
| $M_2$                                    | 6.052             | 5.251             | 4.636             | 4.161             | 3.763                            | 4.281              | 4.793              | 5.288              | 5.767              |

Pour chaque valeur de  $\tilde{p}_{20}$  choisie dans l'intervalle  $[4/2116, 180/2116]$  on donne les valeurs des  $\tilde{p}_i \quad i \neq 20$ , de  $\min_{k=1 \dots 16} (1 + \beta_{20} y_{20k}^2)^{-1}$  et  $\max_{k=1 \dots 16} (1 + \beta_{20} y_{20k}^2)^{-1}$  les deux valeurs propres de l'A.C.P. correspondant aux

nouveaux poids ainsi que les bornes inférieures  $m_i$  et supérieures  $M_i$  ( $i = 1, 2$ )<sup>1</sup> qu'en donnent les relations (5.4) et (5.5).

Les commentaires principaux qu'amènent ces résultats nous semblent les suivants :

On note que la première valeur propre varie dans le même sens que le poids du poisson 20, ce qui semble être explicable par la forte contribution initiale de ce poisson au premier axe factoriel. La variation est assez sensible (de 6.9 à 8.1) lorsqu'on passe d'un poids pratiquement nul (4/2116) à un poids (180/2116) de l'ordre du double du poids initial 92/2116. Ainsi, dans le cas d'un tableau de taille modeste, la variation du poids d'une seule u.s. bien choisie (et en particulier sa suppression) peut avoir des conséquences importantes sur l'inertie des axes.

On remarque par ailleurs que les variations de la deuxième valeur propre vont en sens inverse de celles de la première ce qu'on peut en première analyse imputer à un phénomène de compensation, la trace de la matrice restant constante.

Les bornes  $m_i$  et  $M_i$  sont principalement liées au changement d'échelle qui intervient par l'intermédiaire des rapports extrémaux de variance  $\min_k (1 + \beta_{20} y_{20k}^2)^{-1}$  et  $\max_k (1 + \beta_{20} y_{20k}^2)^{-1}$ . Elles sont donc tributaires du manque de robustesse de la variance d'un petit échantillon lorsqu'on perturbe le poids d'une u.s. éloignée du centre de gravité.

Ainsi leur comportement satisfaisant pour de petites variations de poids se détériore parfois de manière trop importante (pour des variations plus grandes) pour qu'elles présentent alors un réel intérêt (pour  $\bar{p}_{20} = 4/2116$ ,  $M_1 = 12,2$  car  $\max_k (1 + \beta_{20} y_{20k}^2)^{-1} = 1,6$ ).

On notera cependant que pour une augmentation de poids le coefficient  $\max_k (1 + \beta_{20} y_{20k}^2)^{-1}$  reste très près de l'unité alors que  $\min_k (1 + \beta_{20} y_{20k}^2)^{-1}$  s'en écarte sensiblement le phénomène s'inversant pour une diminution de poids. Ainsi l'une des bornes reste toujours acceptable. De plus la proximité avec l'unité de  $\max_k (1 + \beta_{20} y_{20k}^2)^{-1}$  lorsque  $\beta_{20} > 0$ , et de  $\min_k (1 + \beta_{20} y_{20k}^2)^{-1}$  lorsque  $\beta_{20} < 0$ , laisse penser que la perte de précision, due à l'utilisation du lemme 6.1 pour passer à une interprétation géométrique ou graphique des résultats (paragraphes 6 et 7), est généralement faible.

Pour conclure notons que le choix d'un petit tableau de données nous a permis de faire ressortir de manière nette certains caractères essentiels des encadrements proposés. Il n'est pas douteux que les traits soulignés apparaîtraient de manière beaucoup plus atténuée dans l'étude d'un tableau plus important.

D'une manière plus globale on fera le rapprochement entre l'ensemble des résultats proposés et les recherches effectuées dans le cadre de la méthode du bootstrap.

## BIBLIOGRAPHIE

- [1] T.W. ANDERSON, S. DASGUPTA. — « Some inequalities on characteristic roots of matrices », *Biometrika*, 1963, Vol. 50, p. 522-524.
- [2] J. BENASSENI. — Une contribution à l'étude de la stabilité en analyse factorielle, *Thèse de 3<sup>e</sup> cycle*, U.S.T.L., Montpellier, 1984.
- [3] J. BRENOT, M. PARMENTIER, J.P. PAGES. — Sur l'univers des variables et la stabilité en analyse factorielle, *Data Analysis and Informatics, III*, North Holland, 1984.
- [4] S.J. DEVLIN, R. GNANADESIKAN, J.R. KETTENRING. — « Robust estimation and outlier detection with correlation coefficients », *Biometrika*, 1975, Vol. 63, p. 531-545.
- [5] B. ESCOFIER. — Stabilité et approximation en analyse factorielle, *Thèse de Doctorat d'Etat*, Université P. et M. Curie, Paris VI, 1979.
- [6] J.P. PAGES, F. CAILLIEZ. — *Introduction à l'Analyse des Données*, Smash, 1976.
- [7] J.H. WILKINSON. — *The algebraic eigenvalue problem*, Clarendon Press, Oxford, 1965.