

# ANNALES DE L'I. H. P., SECTION B

SERGE DEGERINE

**Propriétés de moindres carrés de la matrice  
de covariance d'un vecteur aléatoire et  
applications statistiques**

*Annales de l'I. H. P., section B*, tome 18, n° 3 (1982), p. 237-247

[http://www.numdam.org/item?id=AIHPB\\_1982\\_\\_18\\_3\\_237\\_0](http://www.numdam.org/item?id=AIHPB_1982__18_3_237_0)

© Gauthier-Villars, 1982, tous droits réservés.

L'accès aux archives de la revue « Annales de l'I. H. P., section B » (<http://www.elsevier.com/locate/anihpb>) 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/>

---

## **Propriétés de moindres carrés de la matrice de covariance d'un vecteur aléatoire et applications statistiques**

par

**Serge DEGERINE**

Laboratoire IMAG-UII, B. P. 53 X, 38041 Grenoble Cedex

---

**SOMMAIRE.** — Soit  $X$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^p$ , de matrice de covariance  $\Sigma_X$  supposée non singulière. La minimisation de la trace du produit  $M\Sigma_X$ , où  $M$  désigne une matrice symétrique définie positive, nous permet, lorsque l'on se fixe une contrainte sur  $M$ , de caractériser  $\Sigma_X$ .

On examine le cas des contraintes  $\text{Tr}(M^{-1}) \leq 1$  et  $\det(M^{-1}) \leq 1$ . Dans chaque cas on obtient une représentation de  $X$  faisant intervenir une erreur de type multiplicatif qui conduit à interpréter la matrice de covariance empirique d'un échantillon issu de  $X$ , comme un estimateur des moindres carrés de  $\Sigma_X$ . D'autres applications statistiques sont abordées.

**SUMMARY.** — Let  $X$  be a  $p$ -dimensional random vector with a non-singular covariance matrix  $\Sigma_X$ . The minimisation of the trace of the product  $M\Sigma_X$ , with  $M$  a symmetric positive definite matrix, allows us, if  $M$  is subject to some constraint, to characterize  $\Sigma_X$ . We consider the constraints  $\text{Tr}(M^{-1}) \leq 1$  and  $\det(M^{-1}) \leq 1$ . In each case we obtain a representation of  $X$  with a multiplicative error which leads us to look on the sample covariance matrix as a least square estimator. Other statistical applications are approached.

---

## I. NOTATIONS ET POSITION DU PROBLÈME

Dans tout ce qui suit  $X$  désigne un vecteur aléatoire à valeurs dans  $\mathbb{R}^p$ , d'espérance  $m_X$  et de matrice de covariance  $\Sigma_X$ .  $X_1, \dots, X_n$  est un échantillon de taille  $n$ ,  $n \geq p + 1$ , issu de  $X$  auquel sont associées la moyenne et la matrice de covariance empiriques :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\Sigma}_X = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})'.$$

$\mathcal{S}^+$  est l'ensemble des matrices symétriques définies positives d'ordre  $p$ ; une matrice  $M$  de  $\mathcal{S}^+$  est le plus souvent assimilée à la norme euclidienne qu'elle définit sur  $\mathbb{R}^p$ .

On appelle décomposition spectrale d'une matrice symétrique positive  $\Sigma$  toute écriture de la forme :

$$\Sigma = \sigma^2 U' D^2 U,$$

où  $U$  est une matrice orthogonale, ( $U'U = I_p$ ), dont les vecteurs lignes, notés  $U^j$ , sont des vecteurs propres de  $\Sigma$ , où  $\sigma^2$  est un terme positif et où  $D^2$  est une matrice diagonale à éléments positifs ou nuls.

Il est bien connu que  $\bar{X}$  et  $\hat{\Sigma}_X$  sont des estimateurs de  $m_X$  et  $\Sigma_X$  qui possèdent de bonnes propriétés : ils sont sans biais  $\left(\frac{n}{n-1} \hat{\Sigma}_X\right)$ , convergents et de maximum de vraisemblance dans le cas gaussien (cf. par exemple [1]). On sait de plus que  $\bar{X}$  est un estimateur des moindres carrés ; par contre aucune propriété de ce genre ne semble jusqu'alors avoir été établie pour  $\hat{\Sigma}_X$ . On propose ici deux manières d'interpréter  $\hat{\Sigma}_X$  comme un estimateur des moindres carrés.

Considérons le modèle linéaire

$$X = m_X + \varepsilon$$

dans lequel  $\varepsilon$  est un vecteur aléatoire à valeurs dans  $\mathbb{R}^p$ , centré, de matrice de covariance  $\Sigma_X$ . Dans ce modèle  $\varepsilon$  s'interprète comme une erreur additive sur la mesure de  $m_X$  que représente  $X$ . Ainsi, pour estimer  $m_X$ , il est naturel d'utiliser la méthode des moindres carrés, c'est-à-dire de retenir le vecteur  $m$  de  $\mathbb{R}^p$  qui minimise la somme des carrés des normes des écarts entre les différents  $X_i$  et  $m$ . Il reste cependant à choisir une norme sur  $\mathbb{R}^p$  ;

en se restreignant aux normes euclidiennes et en notant  $M$  la matrice de  $\mathcal{S}^+$  correspondante on a :

$$\sum_{i=1}^n (X_i - m)' M (X_i - m) = \sum_{i=1}^n (X_i - \bar{X})' M (X_i - \bar{X}) + n(\bar{X} - m)' M (\bar{X} - m).$$

Ceci prouve que  $\bar{X}$  est l'estimateur des moindres carrés de  $m_X$  indépendamment du choix de  $M$  et de la matrice de covariance  $\Sigma_X$ . La relation  $E \{ (X - m)' M (X - m) \} = E \{ X - m_X \}' M (X - m_X) \} + (m_X - m)' M (m_X - m)$ , montre que  $m_X$  est caractérisé, dans le modèle, comme le vecteur  $m$  de  $\mathbb{R}^p$  qui minimise l'espérance du carré de la norme de l'écart entre  $X$  et  $m$ . Ainsi l'estimateur des moindres carrés  $\bar{X}$  possède relativement à la loi empirique de l'échantillon la même propriété caractéristique que  $m_X$  relativement à la loi de  $X$ .

Dans la relation ci-dessus le minimum de  $E \{ (X - m)' M (X - m) \}$  n'est atteint que pour  $m = m_X$  et vaut  $\text{Tr} (M \Sigma_X)$ . On peut envisager alors de minimiser  $\text{Tr} (M \Sigma_X)$  par le choix de  $M$  mais on constate que, sans contrainte sur  $M$ , le problème n'a pas d'intérêt car :

$$\text{Inf} \{ \text{Tr} (M \Sigma_X), M \in \mathcal{S}^+ \} = 0.$$

On a considéré dans [2] la contrainte  $\text{Tr} (M^{-1}) \leq 1$  que l'on reprend ici en la comparant avec la contrainte  $\det (M^{-1}) \leq 1$ . On obtient ainsi au § II deux caractérisations de  $\Sigma_X$ . Ces caractérisations conduisent à proposer au § III deux représentations de  $X$  dans lesquelles intervient une erreur de type multiplicatif dont la minimisation, dans le cadre empirique, permet d'interpréter  $\hat{\Sigma}_X$  comme un estimateur des moindres carrés. On termine ce paragraphe sur deux applications possibles en analyse des données.

## II. CARACTÉRISATIONS DE LA MATRICE DE COVARIANCE

II.1 LEMME. — Soit  $\Sigma$  une matrice symétrique définie positive. On a :

$$\text{Min} \{ \text{Tr} (M \Sigma) ; M \in \mathcal{S}^+, \text{Tr} (M^{-1}) \leq 1 \} = [\text{Tr} (\Sigma^{1/2})]^2,$$

le minimum n'étant réalisé que pour  $M = \text{Tr} (\Sigma^{1/2}) \Sigma^{-1/2}$ .

*Démonstration.* — Soit  $\mathcal{S}$  l'espace vectoriel des matrices symétriques

d'ordre  $p$  muni du produit scalaire défini par la trace. Soit  $\mathcal{C}_K$  le compact défini par :

$$\mathcal{C}_K = \{ M \in \mathcal{S}^+ ; \text{Tr}(M^{-1}) \leq 1 ; \text{Tr}(M) \leq K \}.$$

Pour  $K$  suffisamment grand,  $\mathcal{C}_K$  est non vide et la fonction continue  $M \mapsto \text{Tr}(M\Sigma)$  y atteint ses *extremum*. Soit  $M_0$  une valeur en laquelle est réalisé le minimum. Les conditions  $M_0 \in \mathcal{S}^+$  et  $\text{Tr}(M_0^{-1}) \leq 1$  impliquent que toutes les valeurs propres de  $M_0$  sont minorées par 1 ; il existe donc un voisinage de  $M_0$  dans  $\mathcal{S}$  entièrement contenu dans  $\mathcal{S}^+$  et ainsi  $M_0$  réalise un minimum local de la fonction  $\text{Tr}(M\Sigma)$  considérée sur  $\mathcal{S}$  avec les contraintes  $M \in \mathcal{S}^+$ ,  $\text{Tr}(M^{-1}) \leq 1$  et  $\text{Tr}(M) \leq K$ .  $M_0$  étant situé à l'intérieur de  $\mathcal{S}^+$ , la contrainte  $M \in \mathcal{S}^+$  n'est pas active. De même la contrainte  $\text{Tr}(M) \leq K$  n'est plus active dès que  $K$  est suffisamment grand. En effet la valeur de  $\text{Tr}(M_0\Sigma)$  ne peut pas croître lorsque  $K$  augmente puisqu'il s'agit du minimum de  $\text{Tr}(M\Sigma)$  sur  $\mathcal{C}_K$  et que  $\mathcal{C}_K$  croît avec  $K$ . Or on a :

$$\text{Tr}(M\Sigma) \geq \lambda_p(\Sigma) \text{Tr}(M),$$

où  $\lambda_p(\Sigma)$  désigne la plus petite valeur propre de  $\Sigma$  ; ainsi si la contrainte  $\text{Tr}(M) \leq K$  restait active la valeur de  $\text{Tr}(M_0\Sigma)$  deviendrait infinie en même temps que  $K$ .

$M_0$  doit alors satisfaire les conditions suivantes :

$$\begin{aligned} d_M \{ \text{Tr}(M\Sigma) - \lambda(1 - \text{Tr}(M^{-1})) \} &= 0, & \lambda &\geq 0, \\ \lambda(1 - \text{Tr}(M^{-1})) &= 0, \end{aligned}$$

où  $d_M$  signifie que l'on différencie selon  $M$ . L'égalité

$$d_M \{ \text{Tr}(M\Sigma) - \lambda(1 - \text{Tr}(M^{-1})) \} = \text{Tr} \{ dM[\Sigma - \lambda M^{-2}] \} = 0$$

devant être réalisée pour tout choix de la direction  $dM$ , on obtient les conditions nécessaires suivantes :

$$\Sigma - \lambda M^{-2} = 0, \quad \lambda(1 - \text{Tr}(M^{-1})) = 0, \quad \lambda \geq 0.$$

La seule solution est obtenue pour :

$$M_0 = \text{Tr}(\Sigma^{1/2})\Sigma^{-1/2}, \quad \lambda = \lambda \text{Tr}(M_0^{-1}) = \text{Tr}(M_0\Sigma) = [\text{Tr}(\Sigma^{1/2})]^2. \quad \blacksquare$$

Ce résultat est à rapprocher du suivant :

II.2. LEMME. — Soit  $\Sigma$  une matrice symétrique définie positive. On a :

$$\text{Min} \{ \text{Tr}(M\Sigma) ; M \in \mathcal{S}^+, \det(M^{-1}) \leq 1 \} = p [\det(\Sigma)]^{1/p},$$

le minimum n'étant réalisé que pour  $M = [\det(\Sigma)]^{1/p}\Sigma^{-1}$ .

*Démonstration.* — Avec la contrainte  $\det(\mathbf{M}) = 1$  au lieu de  $\det(\mathbf{M}^{-1}) \leq 1$  on peut trouver le résultat dans [3] (lemme 4, chap. II). La relation suivante permet alors de conclure :

$$\begin{aligned} \text{Tr}(\mathbf{M}\Sigma) &= [\det(\mathbf{M})]^{1/p} \text{Tr}(\tilde{\mathbf{M}}\Sigma) \geq \text{Tr}(\tilde{\mathbf{M}}\Sigma), \\ \tilde{\mathbf{M}} &= [\det(\mathbf{M})]^{-1/p}\mathbf{M}, \quad \det(\tilde{\mathbf{M}}) = 1. \quad \blacksquare \end{aligned}$$

Une interprétation géométrique des contraintes permet de mieux les comparer. Pour cela considérons un système d'axes orthonormés représentant le vectoriel  $\mathbb{R}^p$  muni de la métrique identité. Soit  $\mathbf{M} = \mathbf{U}'\mathbf{D}^2\mathbf{U}$  une décomposition spectrale d'une nouvelle métrique  $\mathbf{M}$  sur  $\mathbb{R}^p$ ; les vecteurs lignes de la matrice  $\mathbf{D}^{-1}\mathbf{U}$  constituent une base  $\mathbf{M}$ -orthogonale de  $\mathbb{R}^p$ ; ils définissent un nouveau système d'axes orthogonaux, les longueurs des unités sur les axes étant données par les termes diagonaux de  $\mathbf{D}^{-1}$ . Un élément de  $\mathbb{R}^p$  de coordonnées  $x$  dans le système d'axes initial a pour coordonnées  $y = \mathbf{D}\mathbf{U}x$  dans le nouveau et on a :

$$\|y\|_p^2 = y'y = x'\mathbf{U}'\mathbf{D}^2\mathbf{U}x = x'\mathbf{M}x = \|x\|_{\mathbf{M}}^2.$$

Le parallélépipède formé sur les vecteurs unitaires du nouveau système d'axes a pour volume  $[\det(\mathbf{M}^{-1})]^{1/2}$  et la longueur de sa diagonale principale est  $(\text{Tr}(\mathbf{M}^{-1}))^{1/2}$ . Ainsi la conservation du volume se traduit par  $[\det(\mathbf{M}^{-1})] = 1$  et la conservation de la longueur de la diagonale par  $\text{Tr}(\mathbf{M}^{-1}) = p$ .

Les propositions suivantes sont des conséquences directes des lemmes précédents.

**II.3. PROPOSITION.** — Soit  $\mathbf{X}$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^p$ , d'espérance  $m_{\mathbf{X}}$  et de matrice de covariance  $\Sigma_{\mathbf{X}}$  non singulière. On a :

$$E\{(X-m)'M(X-m)\} \geq [\text{Tr}(\Sigma_{\mathbf{X}}^{1/2})]^2, \quad \forall m \in \mathbb{R}^p, \quad \forall M \in \mathcal{S}^+, \quad \text{Tr}(\mathbf{M}^{-1}) \leq 1,$$

l'égalité n'étant réalisée que pour  $m = m_{\mathbf{X}}$  et  $\mathbf{M} = \text{Tr}(\Sigma_{\mathbf{X}}^{1/2})\Sigma_{\mathbf{X}}^{-1/2}$ .

**II.4. PROPOSITION.** — Soit  $\mathbf{X}$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^p$ , d'espérance  $m_{\mathbf{X}}$  et de matrice de covariance  $\Sigma_{\mathbf{X}}$  non singulière. On a :

$$E\{(X-m)'M(X-m)\} \geq p [\det(\Sigma_{\mathbf{X}})]^{1/p}, \\ \forall m \in \mathbb{R}^p, \quad \forall M \in \mathcal{S}^+, \quad \det(\mathbf{M}^{-1}) \leq 1,$$

l'égalité n'étant réalisée que pour  $m = m_{\mathbf{X}}$  et  $\mathbf{M} = [\det(\Sigma_{\mathbf{X}})]^{1/p}\Sigma_{\mathbf{X}}^{-1}$ .

### III. APPLICATIONS STATISTIQUES

Avec les hypothèses de la proposition II.3. soit  $\Sigma_X = U'_X \Lambda^2 U_X$  une décomposition spectrale de  $\Sigma_X$ . Le choix optimal de  $M$  conduit à définir un terme d'erreur  $\varepsilon$  en posant :

$$\text{Tr}(\Sigma_X^{1/2})(X - m_X)' \Sigma_X^{-1/2}(X - m_X) = \varepsilon' \varepsilon,$$

avec

$$\varepsilon = [\text{Tr}(\Lambda_X)]^{1/2} \Lambda_X^{-1/2} U_X (X - m_X).$$

Cela suggère de poser pour  $X$  le modèle suivant :

$$X = m_X + U'_X D_X^{1/2} \varepsilon, \quad D_X = [\text{Tr}(\Lambda_X)]^{-1} \Lambda_X.$$

Dans ce modèle l'erreur  $\varepsilon$  est centrée, de matrice de covariance diagonale  $\sigma_\varepsilon^2 D_X$  avec  $\text{Tr}(D_X) = 1$ . En notant  $\alpha_X^j$  les éléments diagonaux de  $D_X$ ,  $U_X^j$  les vecteurs lignes de  $U_X$  et  $\varepsilon^j$  les composantes de  $\varepsilon$  on a :

$$X = m_X + \sum_{j=1}^p \varepsilon^j \sqrt{\alpha_X^j} U_X^j.$$

Lorsque les valeurs propres de  $\Sigma_X$  sont distinctes, ce modèle exprime que  $X$  est la résultante des composantes principales  $\langle X, U_X^j \rangle U_X^j$ , chacune traduisant l'effet sur  $X$  de la variable correspondante  $\langle X, U_X^j \rangle$ ; cette variable mesure un effet moyen  $\langle m_X, U_X^j \rangle$  avec une erreur  $\varepsilon^j$  de type multiplicatif dont la variance est proportionnelle à un paramètre d'échelle  $\alpha_X^j$ . On peut aussi considérer que les  $\sqrt{\alpha_X^j}$  constituent un système d'unités de mesure pour les variables  $\langle X, U_X^j \rangle$  et qu'alors les erreurs  $\varepsilon^j$  exprimées dans les unités ainsi déterminées ont même variance.

Pour une valeur propre multiple les variables  $\langle X, U_X^j \rangle$  qui lui sont associées mesurent un effet vectoriel dont les composantes ne peuvent être dissociées au vu de la variabilité qu'elles induisent sur  $X$ .

Cette représentation de  $X$  possède un caractère optimal, y compris dans le cas où  $\Sigma_X$  est singulière.

III.1. PROPOSITION. — Parmi les représentations de  $X$  sous la forme :

$$X = m + U'D^{1/2}\varepsilon, \quad m \in \mathbb{R}^p, \quad UU' = I_p, \quad \text{Tr}(D) \leq 1,$$

où  $D$  est une matrice diagonale à éléments positifs ou nuls, celles obtenues à partir de  $m_X$  et d'une décomposition spectrale de  $\Sigma_X$  :

$$\Sigma_X = \sigma_\varepsilon^2 U'_X D_X^2 U_X, \quad \text{Tr}(D_X) = 1,$$

sont les seules à réaliser le minimum de  $E\{\varepsilon'\varepsilon\}$ . La valeur de ce minimum est donnée par :

$$\sigma_\varepsilon^2 = \left[ \sum_{j=1}^p \lambda_j \right]^2,$$

où  $\lambda_1, \dots, \lambda_p$  sont les racines carrées des valeurs propres de  $\Sigma_X$ .

*Démonstration.* —  $E(\varepsilon'\varepsilon) = E\{[\varepsilon - E(\varepsilon)]'[\varepsilon - E(\varepsilon)]\} + [E(\varepsilon)]'[E(\varepsilon)]$ ,  $E(\varepsilon'\varepsilon)$  est donc minimum lorsque  $\varepsilon$  est centré, c'est-à-dire  $m = m_X$ , puisque, dans le cas contraire, en posant  $\tilde{\varepsilon} = \varepsilon - E(\varepsilon)$  on a :

$$X = m + U'D^{1/2}E(\varepsilon) + U'D^{1/2}\tilde{\varepsilon}$$

$$m + U'D^{1/2}E(\varepsilon) = m_X \in \mathbb{R}^p, \quad E(\tilde{\varepsilon}'\tilde{\varepsilon}) = E(\varepsilon'\varepsilon).$$

Pour simplifier les notations on pose  $m_X = m = E(\varepsilon) = 0$ . On se place dans le cas où  $\Sigma_X$  est singulière.  $X$  est presque sûrement à valeurs dans le sous-espace  $V$  de  $\mathbb{R}^p$  de dimension  $q$ ,  $q < p$ , orthogonal dans  $\mathbb{R}^p$  du noyau de  $\Sigma_X$ . En notant  $\alpha^j$ , les éléments diagonaux de  $D$ , la condition  $U'U = I_p$  dans la représentation

$$X = U'D^{1/2}\varepsilon = \sum_1^p \varepsilon^j \sqrt{\alpha^j} U^j,$$

montre que seuls  $q$  termes, parmi les  $\varepsilon^j \sqrt{\alpha^j}$ , sont non nuls et que les  $U^j$  correspondant aux termes nuls constituent une base de l'orthogonal de  $V$ . Moyennant une réindication la représentation s'écrit :

$$X = \sum_1^q \varepsilon^j \sqrt{\alpha^j} U^j, \quad \sum_1^q \alpha^j \leq 1.$$

En posant  $\alpha = \sum_1^q \alpha^j$  on a :

$$E(\varepsilon'\varepsilon) = \sum_1^q \frac{1}{\alpha^j} E(\langle X, U^j \rangle^2) \geq \sum_1^q \frac{\alpha}{\alpha_j} E(\langle X, U^j \rangle^2).$$



$E(\varepsilon'\varepsilon)$  est donc minimum lorsque  $\sum_1^q \alpha^j = 1$ . Ainsi  $\alpha^j = 0, j = q + 1, \dots, p$ ,

et les vecteurs  $U^j$  correspondants forment une base du sous-espace propre associé à la valeur propre nulle de  $\Sigma_X$ .

En se plaçant dans  $V$  on retrouve le problème initial en dimension  $q$  avec une matrice  $\Sigma_X$  non singulière et la contrainte  $\text{Tr}(D) = 1$ . Pour ne pas changer les notations on suppose  $\Sigma_X$  non singulière; on a :

$$E(\varepsilon'\varepsilon) = \text{Tr}(M\Sigma_X), \quad M = U'D^{-1}U, \quad \text{Tr}(M^{-1}) = 1.$$

La solution est alors donnée par le lemme II.1. ■

On peut considérer une représentation de  $X$  basée sur le résultat de la proposition II.4. On est alors conduit au modèle suivant :

$$X = m_X + U'_X D_X^{1/2} \varepsilon = m_X + \sum_{j=1}^p \varepsilon^j \sqrt{\alpha_X^j} U_X^j, \quad \det(D) = \prod_1^p \alpha_X^j = 1.$$

Dans ce cas l'erreur  $\varepsilon$  est centrée, de matrice de covariance  $\sigma_\varepsilon^2 I_p$ . On a une interprétation analogue mais avec des erreurs  $\varepsilon^j$  de type multiplicatif qui ont même variance ou encore dont la variance exprimée dans les unités définies par les  $\sqrt{\alpha_X^j}$  est d'autant plus faible que l'unité est grande. De plus on doit se restreindre au cas où  $\Sigma_X$  est non singulière pour énoncer :

III.2. PROPOSITION. — Lorsque  $\Sigma_X$  est non singulière, parmi les représentations de  $X$  sous la forme :

$$X = m + U'D^{1/2}\varepsilon, \quad m \in \mathbb{R}^p, \quad UU' = I_p, \quad \det(D) \leq 1,$$

où  $D$  est une matrice diagonale à éléments positifs, celles obtenues à partir de  $m_X$  et d'une décomposition spectrale de  $\Sigma_X$  :

$$\Sigma_X = \sigma_\varepsilon^2 U'_X D_X^2 U_X, \quad \det(D_X) = 1$$

sont les seules à réaliser le minimum de  $E\{\varepsilon'\varepsilon\}$ . La valeur de ce minimum est donnée par :

$$p\sigma_\varepsilon^2 = p \left[ \prod_1^p \lambda_j^2 \right]^{1/p},$$

où  $\lambda_1^2, \dots, \lambda_p^2$  sont les valeurs propres de  $\Sigma_X$ .

III.3. REMARQUE. — Dans les représentations  $X = m + U'D^{1/2}\varepsilon$ , le critère  $E(\varepsilon'\varepsilon)$  fait abstraction des éventuelles corrélations entre les composantes de  $\varepsilon$  mais il est minimisé précisément dans le cas où ces composantes sont non corrélées.

Pour l'estimation des paramètres  $m_X$  et  $\Sigma_X$  à partir de l'échantillon  $\{X_1, \dots, X_n\}$ , la proposition III.1. montre qu'il est naturel de considérer le couple  $(m, \Sigma)$  qui minimise la somme des carrés des normes des écarts  $\varepsilon_i$  définis par :

$$X_i = m + U'D^{1/2}\varepsilon_i, \quad i = 1, \dots, n; \quad m \in \mathbb{R}^p, \quad UU' = I_p, \quad \text{Tr}(D) \leq 1,$$

où  $D$  est une matrice diagonale à éléments positifs ou nuls,  $\Sigma$  étant alors donnée par :

$$\Sigma = \sigma_\varepsilon^2 U'D^2U, \quad \sigma_\varepsilon^2 = \frac{1}{n} \sum_{i=1}^n \|\varepsilon_i\|^2.$$

Par analogie avec le modèle linéaire on peut considérer qu'il s'agit d'une méthode de moindres carrés pour un modèle dans lequel les erreurs ont un effet multiplicatif.

III.4. PROPOSITION. — Soit  $X$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^p$ , d'espérance  $m_X$  et de matrice de covariance  $\Sigma_X$ . Soit  $\{X_1, \dots, X_n\}$  un échantillon de taille  $n$ ,  $n \geq p + 1$ , issu de  $X$ . Les éléments empiriques :

$$\hat{m}_X = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\Sigma}_X = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})',$$

sont les estimateurs des moindres carrés de  $m_X$  et  $\Sigma_X$  en ce sens que  $(\hat{m}_X, \hat{\Sigma}_X)$  est l'unique solution du problème de minimisation de la somme des carrés des normes des erreurs  $\varepsilon_i$  dans la représentation :

$$X_i = m + U'D^{1/2}\varepsilon_i, \quad i = 1, \dots, n; \quad m \in \mathbb{R}^p, \quad UU' = I_p, \quad \text{Tr}(D) \leq 1,$$

où  $D$  est une matrice diagonale à éléments positifs ou nuls associée à  $\Sigma$  par :

$$\Sigma = \sigma_\varepsilon^2 U'D^2U, \quad \sigma_\varepsilon^2 = \frac{1}{n} \sum_{i=1}^n \|\varepsilon_i\|^2.$$

Ce résultat est une simple application de la proposition III.1 au vecteur aléatoire  $Z$ , à valeurs dans  $\mathbb{R}^p$ , muni de la loi empirique associée à l'observation de  $\{X_1, \dots, X_n\}$ .

Il va de soi que l'on peut énoncer un résultat analogue basé sur la pro-

position III.2. ; toutefois dans ce cas on doit se restreindre à une matrice  $\Sigma_x$  non singulière alors que dans le cas gaussien le couple  $(\hat{m}_x, \hat{\Sigma}_x)$  est de maximum de vraisemblance y compris lorsque  $\Sigma_x$  est singulière. (La vraisemblance étant nulle pour tout  $\Sigma$  dont le sous-espace propre associé à la valeur propre nulle ne coïncide pas avec celui de  $\Sigma_x$ ).

Pour conclure on considère deux applications possibles de la contrainte  $\text{Tr}(M^{-1}) \leq 1$  en analyse des données. Soit  $\underline{x}_1, \dots, \underline{x}_n$  un nuage de  $n$  points dans  $\mathbb{R}^p$  avec  $\underline{x}'_i = (x_i^1, \dots, x_i^p)$ . On note  $\Sigma_x$  la matrice de covariance associée au nuage :

$$\Sigma_x = \frac{1}{n} \sum_{i=1}^n (\underline{x}_i - \bar{\underline{x}})(\underline{x}_i - \bar{\underline{x}})', \quad \bar{\underline{x}} = \frac{1}{n} \sum_{i=1}^n \underline{x}_i.$$

Dans [3] la contrainte  $\det(M) = 1$  est utilisée pour définir une mesure de la dispersion du nuage. En notant  $\mathcal{D}_M$  la somme des carrés des distances entre les points pris deux à deux on a :

$$\mathcal{D}_M = \sum_{i \neq k} \|\underline{x}_i - \underline{x}_k\|_M^2 = \sum_{i \neq k} (\underline{x}_i - \underline{x}_k)' M (\underline{x}_i - \underline{x}_k) = 2n^2 \text{Tr}(M \Sigma_x).$$

Ainsi, sous la contrainte  $\det(M) = 1$ ,  $\mathcal{D}_M$  est minimum pour

$$M^* = [\det(\Sigma_x)]^{1/p} \Sigma_x^{-1}$$

et vaut  $\mathcal{D}_{M^*} = 2n^2 p [\det(\Sigma_x)]^{1/p}$ . Dans le cas où  $\Sigma_x$  est singulière il est naturel de poser  $\mathcal{D}_{M^*} = 0$ .

On peut envisager de minimiser  $\mathcal{D}_M$  sous la contrainte  $\text{Tr}(M^{-1}) = 1$  ; la solution est donnée par  $M^* = \text{Tr}(\Sigma_x^{1/2}) \Sigma_x^{-1/2}$  et vaut  $\mathcal{D}_{M^*} = 2n^2 [\text{Tr}(\Sigma_x^{1/2})]^2$ . Dans ce cas il est naturel de poser  $\mathcal{D}_{M^*} = 2n^2 [\text{Tr}(\Sigma_x^{1/2})]^2$  lorsque  $\Sigma_x$  est singulière et cette quantité est nulle si et seulement si tous les points du nuage sont identiques.

Cependant, à des fins d'analyse discriminante, la contrainte  $\det(M) = 1$  présente l'avantage, par rapport à la contrainte  $\text{Tr}(M^{-1}) = 1$ , d'accorder une importance plus grande aux variables de faible dispersion.

Remarquons enfin que, dans le cadre d'une analyse en composantes principales, on peut envisager d'utiliser la métrique  $M = \text{Tr}(\Sigma_x^{1/2}) \Sigma_x^{-1/2}$  sur l'espace des individus. Si  $\lambda_1^2, \dots, \lambda_p^2$  sont les valeurs propres de  $\Sigma_x$  et  $U_x^1, \dots, U_x^p$  les vecteurs propres normalisés correspondants, c'est-à-dire les variances des composantes principales et les facteurs principaux pour la métrique identité, les valeurs propres de  $\Sigma_x M$  sont égales à  $\lambda_j(\Sigma \lambda_j)$ ,  $j = 1, \dots, p$ , les vecteurs propres  $M$ -orthonormés correspondants étant  $\sqrt{\lambda_j(\Sigma \lambda_j)}^{-1/2} U_x^j$ ,  $j = 1, \dots, p$ . Ainsi les facteurs principaux et les compo-

santes principales restent proportionnels à ceux obtenus avec la métrique identité puisque le  $J^{\text{ème}}$  facteur est égal à  $(\sum \lambda_j)^{1/2} \lambda_j^{-1/2} U_x^j$ . La forme du nuage représenté dans les axes principaux est modifiée ; les axes sont ceux obtenus avec la métrique identité mais les pourcentages d'inertie, bien que restant dans le même ordre, sont modifiés. En particulier au premier axe principal correspond un pourcentage plus faible ce qui a pour conséquence de diminuer un éventuel effet de taille.

### RÉFÉRENCES

- [1] T. W. ANDERSON, *An Introduction to Multivariate Statistical Analysis*, John Wiley, 1958.
- [2] S. DEGERINE, *Une caractérisation de la matrice de covariance d'un vecteur aléatoire et interprétation statistique*. Séminaire de Statistique, Grenoble, 1980-1981.
- [3] J. M. ROMEDER, *Méthodes et programmes d'analyse discriminante*, Dunod, éditeur, Paris, 1973.

(Manuscrit reçu le 20 octobre 1981)