

FACULTÉ DES SCIENCES ET DE GÉNIE

REPRÉSENTATION BIPLLOT ROBUSTE

GAÉTAN DAIGLE

Mémoire
présenté
pour l'obtention
du grade de maître ès sciences (M. Sc.)

ÉCOLE DES GRADUÉS
UNIVERSITÉ LAVAL

MARS 1989

Je dédie ce mémoire à mes parents

REMERCIEMENTS

Je tiens à remercier mon directeur de mémoire, M. Louis-Paul Rivest, professeur au département de mathématiques et de statistique de l'Université Laval, pour sa direction, ses encouragements et ses conseils judicieux tout au long de cette recherche.

Je voudrais également remercier tous les étudiants gradués qui m'ont aidé de près ou de loin à la réalisation de ce mémoire et spécialement Richard Labrie pour ses précieux conseils dans l'utilisation du logiciel TEX.

Finalement, je voudrais remercier mes parents et ami(e)s pour leurs encouragements et leur soutien.

RÉSUMÉ

La représentation biplot est une méthode de statistique descriptive multivariée. Elle vise à extraire l'essentiel de l'information contenue dans un tableau de données (composé d'individus et de variables) et à en fournir une représentation imagée se prêtant aisément à l'interprétation.

Dans ce travail, on présente le biplot de façon à ce que les distances entre les individus ou variables soient représentées dans la métrique de notre choix. L'unicité de cette représentation est considérée. On étudie ensuite l'analyse biplot où chaque individu et variable reçoivent un poids déterminé au préalable lors d'une estimation robuste des paramètres de localisation ($\boldsymbol{\mu}$) et de dispersion ($\boldsymbol{\Sigma}$) de l'échantillon des individus. Ces estimateurs sont définis à partir d'une nouvelle paramétrisation d'une matrice définie positive. Un algorithme est présenté pour calculer explicitement les estimations robustes des paramètres multidimensionnels $\boldsymbol{\mu}$ et $\boldsymbol{\Sigma}$. En guise d'illustration, on effectue la représentation biplot d'un ensemble de données constitué de variables socio-économiques mesurées sur différents pays.

TABLE DES MATIÈRES

	page
Remerciements	i
Résumé	ii
Table des matières	iii
Liste des Tableaux	v
Liste des Figures	vi
Introduction	1
Chapitre I: Représentation biplot	3
I.1 Décomposition en valeurs singulières (DVS)	4
I.1.1 Forme usuelle de la DVS	4
I.1.2 Généralisation de la DVS au choix des métriques	5
I.1.3 Forme complète de la DVS	6
I.2 Représentation complète d'une matrice au moyen du biplot	6
I.2.1 Factorisation d'une matrice et construction du biplot	6
I.2.2 Unicité de la représentation biplot	10
I.2.3 Biplot des individus	12
I.2.4 Biplot des variables	15
I.3 Représentation approximative d'une matrice au moyen du biplot	16

I.4 Liens entre le biplot et l'analyse en composantes principales	19
I.4.1 Analyse en composantes principales	20
I.4.2 Relations avec l'analyse biplot	21
Chapitre II: Estimation robuste de μ et Σ	22
II.1 Estimation par le maximum de vraisemblance	23
II.2 Paramétrisation d'une matrice définie positive	26
II.2.1 Définition de la fonction matricielle logarithmique	27
II.2.2 Surjectivité de l'application exponentielle	28
II.2.3 Injectivité de l'application exponentielle	32
II.3 Estimation robuste de μ et Σ	33
II.3.1 Estimation robuste de Σ avec μ connu	33
II.3.2 Estimation robuste de μ avec Σ connu	38
II.4 Algorithme de calcul pour $\hat{\mu}$ et $\hat{\Sigma}$	38
Chapitre III: Représentation biplot robuste	41
III.1 Introduction	41
III.2 Biplot robuste des individus	43
III.3 Biplot robuste des variables réduites	43
III.4 Aides à l'interprétation	44
III.5 Exemple d'application	46
III.5.1 Interprétation du biplot des variables réduites	47
III.5.2 Interprétation du biplot des individus	51
Conclusion	56
Bibliographie	57
Annexe A: Ensemble de données analysé au chapitre III	59
Annexe B: Programme SAS effectuant une analyse biplot	63

LISTE DES TABLEAUX

Tableau		page
1.2.1	Caractéristiques démographiques de certaines tribus indiennes . . .	10
1.2.2	Résumé de la construction des deux types de biplot	16
3.5.1	Matrices des corrélations classiques et robustes	47
3.5.2	Aides à l'interprétation du biplot des variables réduites	48

LISTE DES FIGURES

Figure		page
1.2.1	Représentation biplot du tableau 1.2.1	9
3.5.1	Représentation biplot des variables réduites (analyse classique) . .	49
3.5.2	Représentation biplot des variables réduites (analyse robuste) . . .	50
3.5.3	Représentation biplot des individus (analyse classique)	54
3.5.4	Représentation biplot des individus (analyse robuste)	55

INTRODUCTION

L'analyse des données a vu au cours des dernières années son domaine d'application s'étendre considérablement. Elle a connu une effervescence bien légitime en statistique pour son aptitude à décrire et analyser de vastes ensembles de données. Les ouvrages fondamentaux sur l'analyse des données sont ceux de J.P. Benzécri.

Nous allons présenter dans ce texte une méthode d'analyse des données appelée *analyse biplot*. Cette méthode a vu le jour en 1971 grâce aux travaux de Gabriel. Elle vise à extraire l'essentiel de l'information apportée par un ensemble de données et à en fournir une représentation imagée se prêtant aisément à l'interprétation. Cette analyse a l'avantage de représenter simultanément les individus et les variables sur un même graphique. Elle est particulièrement révélatrice pour l'analyse en composantes principales car elle nous montre les distances entre les individus et du même coup les regroupements qu'il est possible de faire entre ces unités. De plus, elle exhibe les covariances et les corrélations entre les variables.

Toutefois, la présence d'observations extrêmes au sein d'une analyse biplot peut fortement influencer ses résultats. Pour résoudre ce problème, il est d'usage de refaire l'analyse sans les points aberrants et de mettre ces derniers en points supplémentaires. Ce travail étudie une autre voie: l'analyse biplot liée à des estimateurs robustes des paramètres de localisation (μ) et de dispersion (Σ) de l'échantillon des individus. Elle a l'avantage d'être résistante aux perturbations causées par la présence de données douteuses ou aberrantes.

Ce travail est divisé en trois chapitres. Le premier présente le biplot, une méthode de statistique descriptive multivariée. On y discute d'algèbre linéaire et d'approximation de matrices à l'aide de la méthode des moindres carrés pondérés. On y montre également la relation entre le biplot et l'analyse en composantes principales telle que développée par l'école française.

Le deuxième chapitre porte sur l'estimation robuste des paramètres de localisation et d'échelle multivariés. On se sert d'une nouvelle paramétrisation pour les matrices de variances-covariances pour donner une réponse partielle au problème de l'unicité des solutions du système d'équations non linéaires définissant les estimateurs.

Le troisième fait une synthèse des deux premiers chapitres. On y montre comment construire une représentation biplot à l'aide des estimateurs robustes du chapitre II. Une application de la méthode portant sur un ensemble de données constitué de plusieurs variables socio-économiques mesurées sur différents pays y est présentée.

CHAPITRE I

Représentation biplot

Soit $\mathbf{X} = (x_{i,j})$ une matrice $n \times p$ de rang r avec $2 \leq r \leq \min(n, p)$. Supposons que $x_{i,j}$ soit égal à la valeur de la $j^{\text{ième}}$ variable ($j = 1, \dots, p$) mesurée sur l'individu i ($i = 1, \dots, n$). Nous allons montrer qu'il est toujours possible d'associer un vecteur $\mathbf{a}_i \in \mathbb{R}^r$ à la $i^{\text{ième}}$ ligne de \mathbf{X} et un vecteur $\mathbf{b}_j \in \mathbb{R}^r$ à sa $j^{\text{ième}}$ colonne de telle sorte que $x_{i,j} = \mathbf{a}_i' \mathbf{b}_j$. La représentation graphique de ces vecteurs dans un espace de dimension r nous livre toute l'information sur \mathbf{X} . Cependant, cette représentation est souvent inaccessible, notre perception ne nous permettant pas de visualiser quoi que ce soit dans un espace dont la dimension est supérieure à trois. On a alors avantage à approcher cette représentation en projetant les vecteurs $\mathbf{a}_1, \dots, \mathbf{a}_n, \mathbf{b}_1, \dots, \mathbf{b}_p$ dans un sous-espace de dimension h , avec $h < r$. Gabriel (1971) réserve le terme biplot à la représentation de \mathbf{X} dans le meilleur sous-espace de dimension 2 et appelle "bimodel" les autres représentations avec $h \geq 3$.

Les applications du biplot sont nombreuses et diversifiées. Corsten et Gabriel (1976) s'en servent pour comparer les matrices de covariances de plusieurs populations et pour déceler les combinaisons linéaires de variables pour lesquelles ces populations diffèrent. Bradu et Gabriel (1978) l'utilisent à leur tour comme un outil analytique per-

mettant d'identifier le modèle qui s'ajuste le mieux aux données dans un plan d'analyse de variance à deux facteurs avec une seule répétition.

I.1 Décomposition en valeurs singulières (DVS)

La décomposition en valeurs singulières (DVS) est à la base de la représentation biplot. Sa compréhension est donc essentielle à la construction du biplot. Dans cette section, nous allons présenter la DVS sous ses différentes versions en insistant particulièrement sur ses propriétés géométriques. Ces résultats sont présentés sommairement par Greenacre (1984, annexe B) et Good (1969).

I.1.1 Forme usuelle de la DVS

Soit \mathbf{X} une matrice $n \times p$ de rang r , avec $r \leq \min(n, p)$. La décomposition en valeurs singulières de \mathbf{X} consiste à exprimer cette matrice comme le produit de trois autres de la façon suivante:

$$\mathbf{X} = \underset{n \times p}{\mathbf{P}} \underset{n \times r}{\Lambda} \underset{r \times r}{\mathbf{Q}'}, \quad (1.1.1)$$

où

- (i) \mathbf{P} et \mathbf{Q} admettent des colonnes orthonormales dans la norme euclidienne canonique, i.e. $\mathbf{P}'\mathbf{P} = \mathbf{I}_r$ et $\mathbf{Q}'\mathbf{Q} = \mathbf{I}_r$, où \mathbf{I}_r est la matrice identité $r \times r$.
- (ii) Λ est une matrice diagonale à termes positifs $\lambda_1 \geq \dots \geq \lambda_r$.

Cette matrice \mathbf{X} peut s'écrire de façon équivalente comme $\mathbf{X} = \sum_{\alpha=1}^r \lambda_{\alpha} \mathbf{p}_{\alpha} \mathbf{q}'_{\alpha}$ où $\{\mathbf{p}_1, \dots, \mathbf{p}_r\}$ et $\{\mathbf{q}_1, \dots, \mathbf{q}_r\}$ sont les colonnes respectives de \mathbf{P} et \mathbf{Q} . Ces deux ensembles de vecteurs forment respectivement une base orthonormale (au sens de la métrique euclidienne canonique) de l'espace vectoriel engendré par les colonnes de \mathbf{X} et de celui engendré par ses lignes. En fait, ces vecteurs correspondent respectivement aux vecteurs propres de norme unité associés aux valeurs propres non nulles de $\mathbf{X}\mathbf{X}'$ et de $\mathbf{X}'\mathbf{X}$. Les nombres $\lambda_1 \geq \dots \geq \lambda_r > 0$, quant à eux, sont appelés les valeurs singulières de \mathbf{X} et correspondent à la racine carrée de ces valeurs propres.

La décomposition en valeurs singulières admet comme cas particulier la décomposition en valeurs propres. En fait, ces deux types de décomposition coïncident pour

des matrices $\mathbf{X}_{n \times n}$ symétriques. Dans de tels cas, l'équation (1.1.1) devient alors

$$\mathbf{X} = \begin{matrix} \Gamma & \Lambda & \Gamma' \\ n \times n & n \times r & r \times r & r \times n \end{matrix}, \quad (1.1.2)$$

où les colonnes de Γ sont les vecteurs propres de norme unité associés aux valeurs propres non nulles de \mathbf{X}^2 , lesquelles sont données par les éléments diagonaux de Λ^2 . Il est à noter que la DVS s'applique sur n'importe quel tableau rectangulaire, alors que la décomposition en valeurs propres est restreinte aux matrices symétriques.

I.1.2 Généralisation de la DVS au choix des métriques

Soient \mathbf{M} et \mathbf{N} deux matrices définies positives établissant des métriques¹ dans R^p et R^n respectivement. Ces métriques s'intègrent à la décomposition en valeurs singulières de \mathbf{X} de la façon suivante:

$$\mathbf{X} = \begin{matrix} \mathbf{P} & \Lambda & \mathbf{Q}' \\ n \times p & n \times r & r \times r & r \times p \end{matrix}, \quad (1.1.3)$$

où

- (i) \mathbf{P} et \mathbf{Q} admettent des colonnes orthonormales dans la norme induite par les matrices \mathbf{N} et \mathbf{M} respectivement, i.e. $\mathbf{P}'\mathbf{N}\mathbf{P} = \mathbf{I}_r$ et $\mathbf{Q}'\mathbf{M}\mathbf{Q} = \mathbf{I}_r$
- (ii) $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r)$ avec $\lambda_1 \geq \dots \geq \lambda_r > 0$.

Cette décomposition diffère de la première par la notion de norme vectorielle. Nous la dénotons par $\text{DVS}(\mathbf{X}, \mathbf{M}, \mathbf{N})$. De cette décomposition et de la condition (i), on peut déduire les relations matricielles suivantes: $\mathbf{X}\mathbf{M}\mathbf{X}'\mathbf{N}\mathbf{P} = \mathbf{P}\Lambda^2$ et $\mathbf{X}'\mathbf{N}\mathbf{X}\mathbf{M}\mathbf{Q} = \mathbf{Q}\Lambda^2$. C'est donc dire que les colonnes des matrices \mathbf{P} et \mathbf{Q} correspondent respectivement aux vecteurs propres associés aux valeurs propres non nulles de $\mathbf{X}\mathbf{M}\mathbf{X}'\mathbf{N}$ et $\mathbf{X}'\mathbf{N}\mathbf{X}\mathbf{M}$, et que Λ^2 est la matrice diagonale des valeurs propres non nulles de l'une ou l'autre de ces deux dernières matrices.

Cette généralisation de la DVS se vérifie aisément en exprimant d'abord la matrice $\mathbf{N}^{1/2}\mathbf{X}\mathbf{M}^{1/2}$ sous la forme (1.1.1), i.e. en écrivant $\mathbf{N}^{1/2}\mathbf{X}\mathbf{M}^{1/2} = \mathbf{U}\Lambda\mathbf{V}'$ avec $\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}_r$. En isolant ensuite \mathbf{X} dans cette dernière expression, on obtient alors $\mathbf{X} = \mathbf{N}^{-1/2}\mathbf{U}\Lambda\mathbf{V}'\mathbf{M}^{-1/2}$. Il suffit donc de poser $\mathbf{P} = \mathbf{N}^{-1/2}\mathbf{U}$ et $\mathbf{Q} = \mathbf{M}^{-1/2}\mathbf{V}$ pour obtenir le résultat. Ces matrices vérifient la condition (i) puisque $\mathbf{P}'\mathbf{N}\mathbf{P} = \mathbf{U}'\mathbf{U} = \mathbf{I}_r$, $\mathbf{Q}'\mathbf{M}\mathbf{Q} = \mathbf{V}'\mathbf{V} = \mathbf{I}_r$.

¹Pour simplifier l'écriture, nous allons plus tard confondre la notation entre la métrique et la matrice la définissant.

I.1.3 Forme complète de la DVS

Les colonnes $\mathbf{p}_1, \dots, \mathbf{p}_r$ de la matrice \mathbf{P} définie en (1.1.3) forment une base du sous-espace vectoriel de R^n engendré par les colonnes de \mathbf{X} . On note ce sous-espace par $\mathcal{V}(\mathbf{X})$. En utilisant le théorème de la base incomplète (Lancaster et Tismenetsky 1985, p.87), il est toujours possible de trouver des vecteurs $\mathbf{p}_{r+1}, \dots, \mathbf{p}_n$ de telle sorte que les n vecteurs de l'ensemble $\{\mathbf{p}_1, \dots, \mathbf{p}_n\}$ soient orthonormaux dans la métrique \mathbf{N} et forment une base de R^n . De plus, nous savons que les colonnes $\mathbf{q}_1, \dots, \mathbf{q}_r$ de la matrice \mathbf{Q} définie en (1.1.3) forment une base du sous-espace vectoriel de \mathbb{R}^p engendré par les lignes de \mathbf{X} , à savoir $\mathcal{V}(\mathbf{X}')$. Par un raisonnement analogue au précédent, on peut alors trouver des vecteurs $\mathbf{q}_{r+1}, \dots, \mathbf{q}_p$ de telle sorte que les vecteurs de l'ensemble $\{\mathbf{q}_1, \dots, \mathbf{q}_p\}$ soient orthonormaux dans la métrique \mathbf{M} et forment une base de \mathbb{R}^p . Ainsi, en définissant $\tilde{\mathbf{P}} = (\mathbf{p}_1, \dots, \mathbf{p}_n)$ et $\tilde{\mathbf{Q}} = (\mathbf{q}_1, \dots, \mathbf{q}_p)$, on peut écrire la décomposition (1.1.3) sous la forme complète suivante:

$$X = \tilde{\mathbf{P}} \begin{pmatrix} \Lambda & O \\ O & O \end{pmatrix} \tilde{\mathbf{Q}}' = \tilde{\mathbf{P}} \tilde{\Lambda} \tilde{\mathbf{Q}}', \quad (1.1.4)$$

où

- (i) $\tilde{\mathbf{P}}' \mathbf{N} \tilde{\mathbf{P}} = \mathbf{N}^{1/2} \tilde{\mathbf{P}} \tilde{\mathbf{P}}' \mathbf{N}^{1/2} = \mathbf{I}_n$
- (ii) $\tilde{\mathbf{Q}}' \mathbf{M} \tilde{\mathbf{Q}} = \mathbf{M}^{1/2} \tilde{\mathbf{Q}} \tilde{\mathbf{Q}}' \mathbf{M}^{1/2} = \mathbf{I}_p$.

I.2 Représentation complète d'une matrice au moyen du biplot

Soit $\mathbf{X}_{n \times p}$ une matrice de rang r sur laquelle on a appliqué la transformation linéaire ramenant la moyenne de chacune de ses colonnes à 0. Nous allons commencer cette section en montrant comment on construit la représentation biplot de cette matrice. Nous allons en déduire que l'information sur \mathbf{X} apportée par ce type de représentation n'est complète que si $r = 2$. Nous allons ensuite discuter de l'unicité du biplot, ce qui nous amènera à considérer deux types de représentation: le biplot des individus et le biplot des variables.

I.2.1 Factorisation d'une matrice et construction du biplot

La relation (1.1.3) nous permet de factoriser la matrice \mathbf{X} de la façon suivante:

$$\mathbf{X} = \underset{n \times p}{\mathbf{A}} \underset{n \times r}{\mathbf{B}} \underset{r \times p}{\mathbf{B}'}, \quad (1.2.1)$$

où \mathbf{A} et \mathbf{B} sont toutes deux de rang r . En effet, on obtient cette égalité à partir de la relation (1.1.3) en prenant, par exemple, $\mathbf{A} = \mathbf{P}\Lambda^\beta$ et $\mathbf{B} = \mathbf{Q}\Lambda^{1-\beta}$ avec $\beta \in [0, 1]$. Cette remarque nous fait voir que la factorisation (1.2.1) n'est pas unique. Cependant, elle nous permet d'écrire chacune des colonnes de \mathbf{X} comme une combinaison linéaire des colonnes de \mathbf{A} . En fait, si on dénote par $(b_{i,j})$ les éléments de la matrice \mathbf{B} et par $\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(p)}$ et $\mathbf{a}_{(1)}, \dots, \mathbf{a}_{(r)}$ les colonnes de \mathbf{X} et de \mathbf{A} respectivement, nous avons

$$\mathbf{x}_{(j)} = \sum_{i=1}^r b_{j,i} \mathbf{a}_{(i)},$$

$\forall j \in \{1, \dots, p\}$. C'est donc dire que les vecteurs $\{\mathbf{a}_{(1)}, \dots, \mathbf{a}_{(r)}\}$ engendrent l'espace formé par les colonnes de la matrice \mathbf{X} , lequel est noté par $\mathcal{V}(\mathbf{X})$. De plus, il est clair que ces vecteurs sont linéairement indépendants puisque le rang de la matrice \mathbf{A} est r . On en déduit donc que ces vecteurs forment une base de $\mathcal{V}(\mathbf{X})$. De même, on peut montrer que les colonnes de \mathbf{B} forment une base de l'espace vectoriel engendré par les lignes de \mathbf{X} , à savoir $\mathcal{V}(\mathbf{X}')$.

La relation (1.2.1) permet d'exprimer chacun des éléments $x_{i,j}$ de \mathbf{X} au moyen des lignes des matrices \mathbf{A} et \mathbf{B} de la façon suivante:

$$x_{i,j} = \mathbf{a}'_i \mathbf{b}_j, \quad \forall i \in \{1, \dots, n\}, \quad \forall j \in \{1, \dots, p\}, \quad (1.2.2)$$

où \mathbf{a}'_i est la $i^{\text{ième}}$ ligne de \mathbf{A} et où \mathbf{b}'_j est la $j^{\text{ième}}$ ligne de \mathbf{B} . On peut donc reconstituer les np éléments de la matrice \mathbf{X} à partir des $n + p$ vecteurs $\mathbf{a}_1, \dots, \mathbf{a}_n, \mathbf{b}_1, \dots, \mathbf{b}_p$ de \mathbb{R}^r . En particulier, on peut obtenir la valeur prise par l'élément $x_{i,j}$ de \mathbf{X} en effectuant le produit scalaire du vecteur \mathbf{a}_i par le vecteur \mathbf{b}_j . Du point de vue géométrique, cela revient à calculer le produit de la longueur de la projection de l'un des vecteurs sur l'autre par la longueur de cet autre vecteur, ce produit étant négatif si et seulement si le vecteur résultant de la projection et celui sur lequel on projette sont de sens inverse.

Outre la reconstitution de la matrice \mathbf{X} , la factorisation (1.2.2) permet aussi d'établir des relations entre les lignes et entre les colonnes de cette matrice. En effet, les vecteurs \mathbf{a}_i et \mathbf{b}_j de cette factorisation sont associés respectivement au $i^{\text{ième}}$ individu et à la $j^{\text{ième}}$ variable du tableau de données \mathbf{X} . Il suffit alors d'étudier les relations entre les vecteurs $\{\mathbf{a}_i\}_{i=1}^n$ (resp. $\{\mathbf{b}_j\}_{j=1}^p$) pour savoir quels sont les liens entre les individus (resp. variables). Par exemple, on dira que les p mesures prises sur l'individu i sont

proportionnelles à celles prises sur l'individu e s'il existe un nombre k tel que $\mathbf{a}_i = k\mathbf{a}_e$. D'un point de vue géométrique, cela revient à dire que ces vecteurs sont colinéaires, i.e. qu'ils sont situés sur une même droite passant par l'origine. Ce sujet sera traité plus à fond dans les sections I.2.3 et I.2.4.

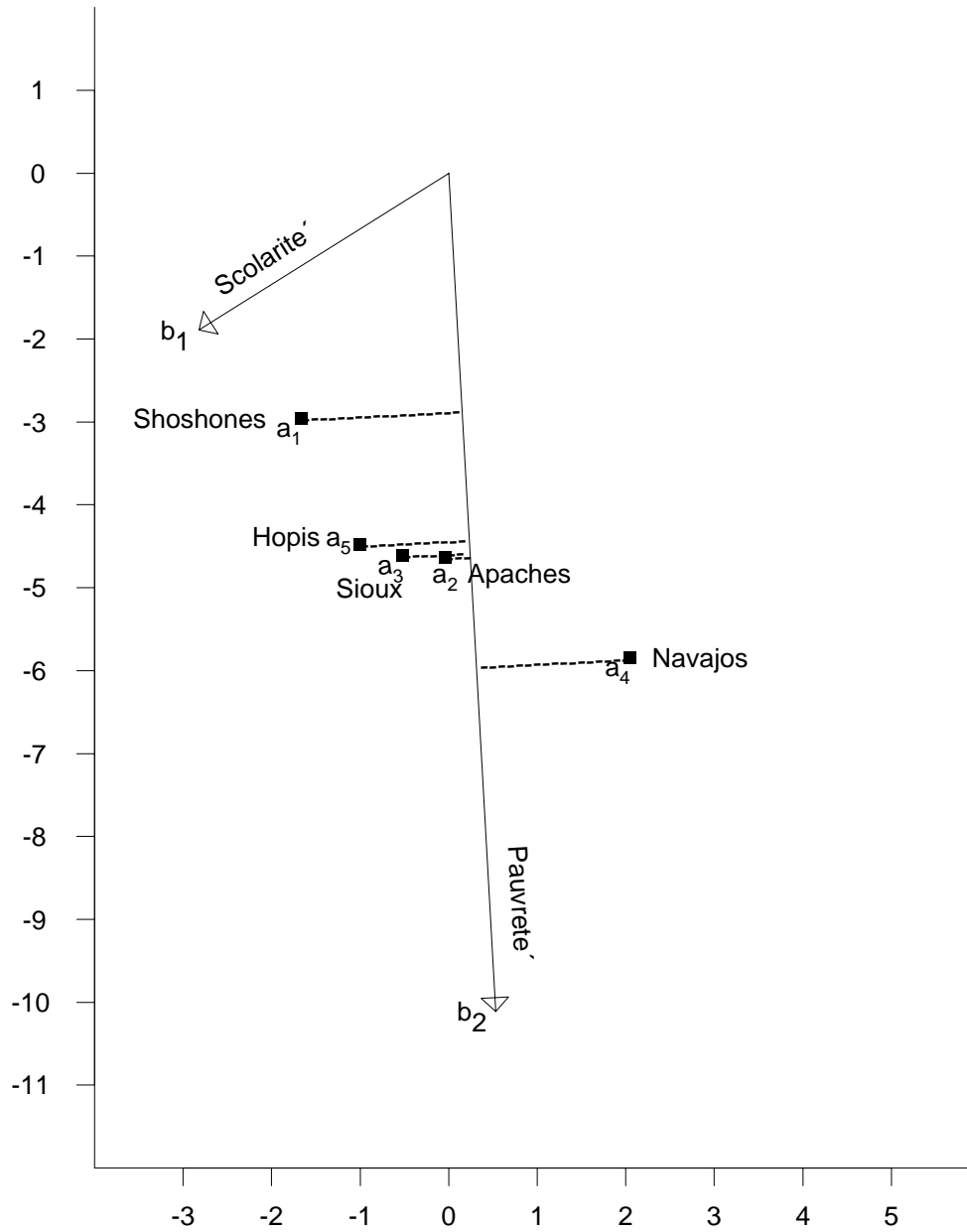
Pour utiliser à profit les propriétés géométriques de la factorisation (1.2.1), il est préférable de représenter les vecteurs $\mathbf{a}_1, \dots, \mathbf{a}_n, \mathbf{b}_1, \dots, \mathbf{b}_p$ sur un graphique. Cependant, cette représentation est souvent inaccessible à notre perception, celle-ci étant faite dans un espace de dimension r . La seule façon d'y voir quelque chose est de projeter les vecteurs $\mathbf{a}_1, \dots, \mathbf{a}_n, \mathbf{b}_1, \dots, \mathbf{b}_p$ dans un espace de dimension inférieure. On appelle biplot de \mathbf{X} le graphique qui représente la matrice \mathbf{X} dans le meilleur sous-espace de dimension 2. Bien que ce type de représentation soit facile à interpréter, il faut en général sacrifier une partie de l'information contenue dans \mathbf{X} pour l'obtenir. Le seul cas qui ne nécessite aucun sacrifice est lorsque les vecteurs $\{\mathbf{a}_i\}$ et $\{\mathbf{b}_j\}$ sont déjà éléments de \mathbb{R}^2 , i.e. lorsque le rang de la matrice \mathbf{X} est 2. C'est dans le cadre de cette hypothèse que nous allons travailler pour le reste de la section.

A titre d'illustration, nous avons effectué la représentation biplot de la matrice \mathbf{X} définie au tableau 1.2.1. Cette représentation est donnée à la figure 1.2.1 en termes de la factorisation \mathbf{AB}' suivante:

$$\mathbf{X} = \begin{pmatrix} -1.66 & -2.96 \\ -0.04 & -4.63 \\ -0.52 & -4.61 \\ 2.05 & -5.85 \\ -1.00 & -4.48 \end{pmatrix} \begin{pmatrix} -2.82 & 0.53 \\ -1.89 & -10.11 \end{pmatrix}. \quad (1.2.3)$$

Sur cette figure, on peut ordonner les éléments $x_{i,j}$ ($i = 1, \dots, 5$) de la $j^{\text{ième}}$ colonne de \mathbf{X} de la même façon qu'on ordonne les projections des vecteurs $\{\mathbf{a}_i\}$ sur le vecteur \mathbf{b}_j . Ceci nous permet de voir par exemple que le pourcentage d'Indiens vivant sous le seuil de la pauvreté est le plus élevé chez les Navajos, le plus faible chez les Shoshones, et à un niveau intermédiaire chez les autres. De plus, on peut voir que les Indiens de la tribu des Hopis sont ceux ayant la plus forte scolarité contrairement aux Indiens de la tribu des Navajos qui sont les moins instruits. On procède de la même façon pour ordonner les éléments $x_{i,j}$ ($j = 1, 2$) de la $i^{\text{ième}}$ ligne de \mathbf{X} . Le biplot de \mathbf{X} suggère également des regroupements possibles d'individus. Par exemple, la proximité des vecteurs \mathbf{a}_2 et

Figure 1.2.1 Représentation biplot du tableau 1.2.1 en termes de la factorisation (1.2.3)



\mathbf{a}_3 sur la figure 1.2.1 nous indique que les Indiens des tribus Apaches et Sioux ont un niveau de scolarité et un seuil de pauvreté comparables.

Tableau 1.2.1. *Caractéristiques démographiques de certaines tribus indiennes**

Tribus	Médiane des années de scolarité	Pourcentage vivant sous le seuil de la pauvreté
Shoshones	10.3	29.0
Apaches	8.9	46.8
Sioux	10.2	46.3
Navajos	5.3	60.2
Hopis	11.3	44.7

* Source: Kunitz (1976)

1.2.2 Unicité de la représentation biplot

La représentation biplot d'une matrice \mathbf{X} dépend essentiellement du choix des matrices \mathbf{A} et \mathbf{B} utilisées dans la factorisation (1.2.1). Or, ces matrices sont définies à une matrice inversible près comme le montre l'égalité suivante:

$$\mathbf{X} = \mathbf{A}\mathbf{B}' = (\mathbf{A}\mathbf{R}')(\mathbf{B}\mathbf{R}^{-1})', \quad \forall \mathbf{R} \text{ inversible.} \quad (1.2.4)$$

Ce que nous entendons faire ici, c'est déterminer en quoi les nombreuses représentations biplot diffèrent si on n'impose pas de conditions supplémentaires dans la factorisation de \mathbf{X} . Cela revient à étudier la nature des transformations $\mathbf{A} \rightarrow \mathbf{A}\mathbf{R}'$ et $\mathbf{B} \rightarrow \mathbf{B}\mathbf{R}^{-1}$ définies en (1.2.4). Ces transformations associent le vecteur $\mathbf{R}\mathbf{a}_i$ à chacun des vecteurs \mathbf{a}_i et le vecteur $\mathbf{R}^{-1}'\mathbf{b}_j$ à chacun des vecteurs \mathbf{b}_j . Pour savoir en quoi consistent ces transformations, écrivons la matrice \mathbf{R} sous la forme (1.1.1), i.e.

$$\mathbf{R} = \begin{matrix} \mathbf{U} & \Lambda & \mathbf{V}' \\ 2 \times 2 & 2 \times 2 & 2 \times 2 \end{matrix}, \quad (1.2.5)$$

où $\Lambda = \text{diag}(\lambda_1, \lambda_2)$ avec $\lambda_1 \geq \lambda_2 > 0$ et où $\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}_2$. Les transformations deviennent donc $\mathbf{a}_i \rightarrow \mathbf{U}\Lambda\mathbf{V}'\mathbf{a}_i$ et $\mathbf{b}_j \rightarrow \mathbf{U}\Lambda^{-1}\mathbf{V}'\mathbf{b}_j$, ce qui nous permet de voir que la différence entre les représentations biplot repose sur le rôle joué par les matrices \mathbf{U} , Λ et \mathbf{V} . Or, le fait que \mathbf{U} et \mathbf{V} aient des colonnes orthonormales dans la factorisation (1.2.5) oblige ces matrices à avoir l'une des deux formes suivantes:

$$\begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{pmatrix}.$$

Ces matrices définissent une application linéaire de \mathbb{R}^2 dans \mathbb{R}^2 ; la première consiste en une rotation des vecteurs d'un angle θ dans le sens positif, la seconde transforme tout vecteur en lui faisant d'abord subir une symétrie par rapport à l'axe des abscisses, puis ensuite une rotation d'un angle θ dans le sens positif. Les matrices Λ et Λ^{-1} , quant à elles, font correspondre à tout vecteur $(x, y)'$ les vecteurs $(\lambda_1 x, \lambda_2 y)'$ et $(\lambda_1^{-1} x, \lambda_2^{-1} y)'$ respectivement. Ainsi, une représentation biplot diffère d'une autre par une série d'applications linéaires consistant en des rotations, des symétries, et des changements d'échelle.

Exemple 1.2.1. La matrice \mathbf{X} définie au tableau 1.2.1 peut encore se factoriser sous la forme \mathbf{AB}' de la façon suivante:

$$\mathbf{X} = \begin{pmatrix} 1.29 & -5.91 \\ 4.59 & -9.27 \\ 4.09 & -9.22 \\ 7.90 & -11.70 \\ 3.48 & -8.95 \end{pmatrix} \begin{pmatrix} -2.82 & 0.53 \\ -2.36 & -4.79 \end{pmatrix}.$$

La matrice de passage d'une factorisation à l'autre est donnée par

$$\mathbf{R} = \begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix},$$

laquelle peut encore s'écrire sous la forme (1.2.4) avec

$$\begin{aligned} \mathbf{U} &= \begin{pmatrix} -.85 & .53 \\ -.53 & -.85 \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} & \theta = 212.0^\circ, \\ \mathbf{V}' &= \begin{pmatrix} -.97 & -.23 \\ .23 & -.97 \end{pmatrix} = \begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix} & \varphi = 166.7^\circ, \\ \Lambda &= \begin{pmatrix} 0.87 & 0 \\ 0 & 2.29 \end{pmatrix}. \end{aligned}$$

□

On peut reconstituer les éléments de la matrice \mathbf{X} à partir de n'importe quelle représentation biplot de \mathbf{X} . C'est seulement lorsque nous voulons dégager les relations (exception faite de la colinéarité) entre les individus et entre les variables que les choses se gâtent. En effet, ces relations sont établies à partir des distances et des angles entre

les vecteurs $\mathbf{a}_1, \dots, \mathbf{a}_n$ et entre les vecteurs $\mathbf{b}_1, \dots, \mathbf{b}_p$, mesures qui dépendent beaucoup de la factorisation (1.2.4) utilisée. On peut toutefois éliminer cette dépendance en contraignant \mathbf{R} à être orthogonale dans cette factorisation. En effet, avec cette propriété matricielle, les transformations que nous avons étudiées deviennent alors $\mathbf{a}_i \rightarrow \mathbf{R}\mathbf{a}_i$ et $\mathbf{b}_j \rightarrow \mathbf{R}\mathbf{b}_j$, lesquelles consistent en des opérations qui préservent les distances et les angles entre les vecteurs, comme le montre les égalités suivantes:

$$\begin{aligned} \|\mathbf{R}\mathbf{x}\|^2 &= (\mathbf{R}\mathbf{x})'(\mathbf{R}\mathbf{x}) = \mathbf{x}'\mathbf{R}'\mathbf{R}\mathbf{x} = \mathbf{x}'\mathbf{x} = \|\mathbf{x}\|^2 \\ \cos(\mathbf{R}\mathbf{x}, \mathbf{R}\mathbf{y}) &= \frac{(\mathbf{R}\mathbf{x})'(\mathbf{R}\mathbf{y})}{\|\mathbf{R}\mathbf{x}\| \|\mathbf{R}\mathbf{y}\|} = \frac{\mathbf{x}'\mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} = \cos(\mathbf{x}, \mathbf{y}), \end{aligned} \quad (1.2.6)$$

où (\mathbf{x}, \mathbf{y}) désigne le plus petit angle entre les vecteurs \mathbf{x} et \mathbf{y} . En fait, ces transformations consistent en des rotations et/ou symétries.

Gabriel (1971) affirme que le biplot peut être utilisé efficacement pour inspecter les relations entre les lignes et/ou colonnes de \mathbf{X} en imposant une certaine métrique dans l'étude de ces relations. Cependant, une fois qu'une métrique est spécifiée pour définir les distances entre les individus (resp. variables), une autre est alors induite pour les variables (resp. individus). Ainsi, il faudra construire le biplot en fonction de certaines priorités. Par exemple, si ce sont les individus que l'on choisit de bien représenter, il faudra spécifier la métrique dans laquelle on veut évaluer les distances inter-individus et construire le biplot en conséquence. Il en est de même pour les variables. Ceci nous amène à parler de deux types de représentations biplot: le biplot des individus et le biplot des variables.

1.2.3 *Biplot des individus*

Pour ce type de biplot, ce sont les individus que l'on choisit de mieux représenter. Si \mathbf{M} est la métrique dans laquelle nous voulons évaluer les distances inter-individus, il faudra alors imposer à la factorisation (1.2.1) la contrainte

$$\mathbf{B}'\mathbf{M}\mathbf{B} = \mathbf{I}_2, \quad (1.2.7)$$

de façon à ce que

$$\mathbf{X}\mathbf{M}\mathbf{X}' = \mathbf{A}\mathbf{A}'. \quad (1.2.8)$$

En fait, si \mathbf{x}'_i et \mathbf{x}'_e représentent deux lignes quelconques de \mathbf{X} , nous avons que

$$(\mathbf{x}_i - \mathbf{x}_e)' \mathbf{M} (\mathbf{x}_i - \mathbf{x}_e) = (\mathbf{a}_i - \mathbf{a}_e)' (\mathbf{a}_i - \mathbf{a}_e).$$

Ainsi, ce type de biplot permet de faire des regroupements homogènes d'individus en mettant dans une même classe tous les individus pour lesquels les vecteurs \mathbf{a}_i sont près les uns des autres.

La métrique \mathbf{M} la plus couramment utilisée est celle conduisant à la standardisation des variables, i.e. $\mathbf{M} = \text{diag}(s_1^{-2}, \dots, s_p^{-2})$ où s_j^2 est la variance de la $j^{\text{ième}}$ variable. Elle permet d'évaluer les distances entre les individus indépendamment des unités avec lesquelles les variables ont été mesurées. Elle est souvent utilisée lorsque les quantités mesurées apparaissant dans les diverses colonnes de \mathbf{X} ne sont pas de même nature.

Une fois que la métrique \mathbf{M} est spécifiée pour l'étude des relations entre les lignes de \mathbf{X} , une autre *métrique*² est alors induite pour comparer les colonnes de \mathbf{X} entre elles. Il s'agit de la *métrique* $(\mathbf{X}\mathbf{M}\mathbf{X}')^g$, comme le montre la relation matricielle suivante:

$$\mathbf{X}'(\mathbf{X}\mathbf{M}\mathbf{X}')^g \mathbf{X} = \mathbf{B}\mathbf{B}', \quad (1.2.9)$$

où \mathbf{Y}^g représente une inverse généralisée de \mathbf{Y} , i.e. une matrice \mathbf{Z} telle que $\mathbf{Y}\mathbf{Z}\mathbf{Y} = \mathbf{Y}$. Souvent, il est difficile d'interpréter correctement la proximité entre deux variables sur ce type de biplot puisque la *métrique* $(\mathbf{X}\mathbf{M}\mathbf{X}')^g$ peut être vide de sens. Et même dans de tels cas, les vecteurs $\{\mathbf{b}_j\}$ auront encore leur place sur le biplot puisqu'ils permettent de reconstituer chacun des éléments $x_{i,j}$ de la matrice \mathbf{X} (voir section I.2.1).

On démontre la relation (1.2.9) en se servant des équations (1.2.1) et (1.2.7) et en utilisant le fait que $\mathbf{A}'(\mathbf{A}\mathbf{A}')^g \mathbf{A} = \mathbf{I}_2$, ce qui est l'objet de la proposition suivante.

Proposition 1.2.1. *Si $(\mathbf{A}\mathbf{A}')^g$ est une inverse généralisée de $\mathbf{A}\mathbf{A}'$, alors*

$$\mathbf{A}'(\mathbf{A}\mathbf{A}')^g \mathbf{A} = \mathbf{I}_2.$$

Démonstration. On peut facilement vérifier que la matrice $\mathbf{Z}_0 = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-2}\mathbf{A}'$ est une inverse généralisée de $\mathbf{A}\mathbf{A}'$. A partir de ce cas particulier, Rao (1973, 1b.5.2, p.52)

²On donne à ce mot une extension de son sens usuel. Il peut s'agir en fait d'une pseudo-métrique.

montre que toutes les inverses généralisées de $\mathbf{A}\mathbf{A}'$ s'écrivent comme

$$\mathbf{Z} = \mathbf{Z}_0 + \mathbf{U} - \mathbf{Z}_0\mathbf{A}\mathbf{A}'\mathbf{U}\mathbf{A}\mathbf{A}'\mathbf{Z}_0,$$

avec une certaine matrice \mathbf{U} de dimension $n \times n$. On obtient le résultat compte tenu du fait que $\mathbf{A}'\mathbf{Z}_0\mathbf{A} = \mathbf{I}_2$. En effet,

$$\mathbf{A}'\mathbf{Z}\mathbf{A} = \mathbf{A}'\mathbf{Z}_0\mathbf{A} + \mathbf{A}'\mathbf{U}\mathbf{A} - \mathbf{A}'\mathbf{Z}_0\mathbf{A}\mathbf{A}'\mathbf{U}\mathbf{A}\mathbf{A}'\mathbf{Z}_0\mathbf{A} = \mathbf{I}_2.$$

□

Il reste maintenant à déterminer de façon explicite quelles sont les matrices \mathbf{A} et \mathbf{B} qui satisfont les équations (1.2.1) et (1.2.7). Pour ce faire, nous allons utiliser la décomposition $\text{DVS}(\mathbf{X}, \mathbf{M}, \mathbf{N})$ définie en (1.1.3) avec une métrique quelconque \mathbf{N} de dimension $n \times n$. Cette décomposition nous donne les matrices \mathbf{P} , Λ et \mathbf{Q} qui satisfont les conditions $\mathbf{P}'\mathbf{N}\mathbf{P} = \mathbf{Q}'\mathbf{M}\mathbf{Q} = \mathbf{I}_2$ et telles que $\mathbf{X} = \mathbf{P}\Lambda\mathbf{Q}'$. Les matrices que l'on cherche sont donc $\mathbf{A} = \mathbf{P}\Lambda$ et $\mathbf{B} = \mathbf{Q}$.

Mais qu'advient-il de ces matrices si on change la métrique \mathbf{N} pour une autre dans la décomposition (1.1.3)? La représentation biplot s'en trouve-t-elle affectée? Les réponses à ces questions sont données à la proposition suivante.

Proposition 1.2.2. *Sur le biplot des individus, les relations entre les vecteurs $\{\mathbf{a}_i\}$ et entre les vecteurs $\{\mathbf{b}_j\}$ sont indépendantes du choix de la métrique \mathbf{N} utilisée dans la décomposition (1.1.3).*

Démonstration. Utilisons les décompositions $\text{DVS}(\mathbf{X}, \mathbf{M}, \mathbf{N})$ et $\text{DVS}(\mathbf{X}, \mathbf{M}, \tilde{\mathbf{N}})$ pour écrire la matrice \mathbf{X} sous les deux formes équivalentes suivantes:

$$\mathbf{X} = \mathbf{P}_1\Lambda_1\mathbf{Q}'_1 \quad \text{avec } \mathbf{P}'_1\mathbf{N}\mathbf{P}_1 = \mathbf{Q}'_1\mathbf{M}\mathbf{Q}_1 = \mathbf{I}_2,$$

$$\mathbf{X} = \mathbf{P}_2\Lambda_2\mathbf{Q}'_2 \quad \text{avec } \mathbf{P}'_2\tilde{\mathbf{N}}\mathbf{P}_2 = \mathbf{Q}'_2\mathbf{M}\mathbf{Q}_2 = \mathbf{I}_2.$$

De plus, posons $\mathbf{A}_1 = \mathbf{P}_1\Lambda_1$, $\mathbf{B}_1 = \mathbf{Q}_1$, $\mathbf{A}_2 = \mathbf{P}_2\Lambda_2$ et $\mathbf{B}_2 = \mathbf{Q}_2$. Pour démontrer le résultat, il suffit de montrer qu'il existe une matrice \mathbf{R} orthogonale telle que $\mathbf{A}_2 = \mathbf{A}_1\mathbf{R}$ et $\mathbf{B}_2 = \mathbf{B}_1\mathbf{R}$ (voir section I.2.2). Nous allons commencer par vérifier qu'on peut toujours trouver une matrice \mathbf{R} telle que $\mathbf{B}_2 = \mathbf{B}_1\mathbf{R}$. En effet, en isolant \mathbf{R} dans cette égalité, i.e. en prémultipliant les deux membres de cette équation par $\mathbf{B}'_1\mathbf{M}$ et en

utilisant la relation (1.2.7), on trouve que $\mathbf{R} = \mathbf{B}'_1 \mathbf{M} \mathbf{B}_2$. Cette matrice est orthogonale puisque

$$\mathbf{R}'\mathbf{R} = \mathbf{B}'_2 \mathbf{M} \mathbf{B}_1 \mathbf{R} = \mathbf{B}'_2 \mathbf{M} \mathbf{B}_2 = \mathbf{I}_2.$$

On en déduit alors que

$$\mathbf{A}_2 = \mathbf{P}_2 \Lambda_2 = \mathbf{X} \mathbf{M} \mathbf{B}_2 = \mathbf{X} \mathbf{M} \mathbf{B}_1 \mathbf{R} = \mathbf{P}_1 \Lambda_1 \mathbf{R} = \mathbf{A}_1 \mathbf{R},$$

ce qui termine la démonstration. \square

I.2.4 *Biplot des variables*

Pour ce type de biplot, ce sont maintenant les variables que l'on choisit de mieux représenter. Si on désire que ce biplot représente les distances entre les variables dans la métrique \mathbf{N} , il faut exiger que la factorisation (1.2.1) se fasse sous la contrainte

$$\mathbf{A}'\mathbf{N}\mathbf{A} = \mathbf{I}_2 \tag{1.2.10}$$

au lieu de (1.2.7) de façon à obtenir

$$\mathbf{X}'\mathbf{N}\mathbf{X} = \mathbf{B}\mathbf{B}'. \tag{1.2.11}$$

Plaçons-nous dans le cas particulier où la matrice \mathbf{X} est centrée, i.e. où on a opéré sur \mathbf{X} la transformation linéaire ramenant la moyenne de chacune de ses colonnes à zéro. En choisissant $\mathbf{N} = [1/(n-1)] \mathbf{I}_n$, nous sommes alors en mesure de lire sur le biplot les covariances et les corrélations entre les variables. En effet, en notant respectivement la covariance et la corrélation entre la $j^{\text{ième}}$ et la $k^{\text{ième}}$ variable par $s_{j,k}$ et $r_{j,k}$ nous obtenons, $\forall j, k \in \{1, \dots, p\}$,

$$\begin{aligned} s_{j,k} &= \frac{1}{n-1} \sum_{i=1}^n x_{i,j} x_{i,k} = \mathbf{b}'_j \mathbf{b}_k \\ s_{j,j} &= \frac{1}{n-1} \sum_{i=1}^n x_{i,j}^2 = \mathbf{b}'_j \mathbf{b}_j = \|\mathbf{b}_j\|^2 \\ r_{j,k} &= \frac{s_{j,k}}{\sqrt{s_{j,j} s_{k,k}}} = \cos(\mathbf{b}_j, \mathbf{b}_k). \end{aligned} \tag{1.2.12}$$

Une fois la métrique \mathbf{N} spécifiée, c'est maintenant au tour des individus à être représentés dans une *métrique* $(\mathbf{X}'\mathbf{N}\mathbf{X})^g$. En effet, nous pouvons montrer de la même façon que nous l'avons fait pour (1.2.9) que

$$\mathbf{X}(\mathbf{X}'\mathbf{N}\mathbf{X})^g\mathbf{X}' = \mathbf{A}\mathbf{A}'. \quad (1.2.13)$$

Il est à noter que si \mathbf{X} est centrée et de plein rang et si $\mathbf{N} = [1/(n-1)]\mathbf{I}_p$, alors les distances euclidiennes entre les vecteurs $\{\mathbf{a}_i\}$ correspondent aux distances de Mahalanobis entre les individus. En effet, en notant par $\{\mathbf{x}'_i\}$ les lignes de la matrice \mathbf{X} , nous obtenons

$$(n-1)(\mathbf{x}_i - \mathbf{x}_e)'[\mathbf{X}'\mathbf{X}]^{-1}(\mathbf{x}_i - \mathbf{x}_e) = (\mathbf{a}_i - \mathbf{a}_e)'(\mathbf{a}_i - \mathbf{a}_e). \quad (1.2.14)$$

Le membre de gauche de (1.2.14) correspond à la distance de Mahalanobis entre le $i^{\text{ième}}$ et le $e^{\text{ième}}$ individu.

Pour déterminer les matrices \mathbf{A} et \mathbf{B} qui satisfont les équations (1.2.1) et (1.2.10), on se sert encore une fois de la décomposition DVS($\mathbf{X}, \mathbf{M}, \mathbf{N}$) définie en (1.1.3) mais cette fois avec une métrique quelconque \mathbf{M} de dimension $p \times p$. Cette décomposition nous donne les matrices \mathbf{P} , Λ et \mathbf{Q} qui satisfont les conditions $\mathbf{P}'\mathbf{N}\mathbf{P} = \mathbf{Q}'\mathbf{M}\mathbf{Q} = \mathbf{I}_2$ et telles que $\mathbf{X} = \mathbf{P}\Lambda\mathbf{Q}'$. Il suffit alors de prendre $\mathbf{A} = \mathbf{P}$ et $\mathbf{B} = \mathbf{Q}\Lambda$ pour construire le biplot des variables.

Tableau 1.2.2. *Résumé de la construction des deux types de biplot*

Biplot	Factorisation	Métrique pour les individus	Métrique pour les variables
individus variables	$\mathbf{A} = \mathbf{P}\Lambda, \mathbf{B} = \mathbf{Q}$ $\mathbf{A} = \mathbf{P}, \mathbf{B} = \mathbf{Q}\Lambda$	\mathbf{M} $(\mathbf{X}'\mathbf{N}\mathbf{X})^g$	$(\mathbf{X}\mathbf{M}\mathbf{X}')^g$ \mathbf{N}

I.3 Représentation approximative d'une matrice au moyen du biplot

Les matrices de rang supérieur à deux ne peuvent pas être représentées exactement au moyen du biplot. Cependant, si nous pouvons approcher de façon satisfaisante la matrice \mathbf{X} par une matrice $\mathbf{X}_{(2)}$ de rang 2, alors le biplot de $\mathbf{X}_{(2)}$ nous aidera à dégager la structure de la matrice \mathbf{X} elle-même.

La méthode des moindres carrés est souvent utilisée pour approcher une matrice $\mathbf{X}_{n \times p}$ de rang r par une autre de même dimension mais de rang inférieur. Cette méthode a été utilisée pour la première fois dans ce contexte par Householder et Young en 1938. Elle fait l'objet de la proposition suivante:

Proposition 1.3.1. *Soit \mathbf{X} une matrice $n \times p$ de rang r dont la décomposition est donnée en (1.1.1), i.e. $\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}' = \sum_{\alpha=1}^r \lambda_{\alpha} \mathbf{u}_{\alpha} \mathbf{v}'_{\alpha}$ avec $\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}_r$. Alors la matrice $\mathbf{X}_{(s)} = \sum_{\alpha=1}^s \lambda_{\alpha} \mathbf{u}_{\alpha} \mathbf{v}'_{\alpha}$ est la matrice $n \times p$ de rang s avec $s < r$ qui approche le mieux \mathbf{X} au sens des moindres carrés, i.e.*

$$\|\mathbf{X} - \mathbf{X}_{(s)}\|^2 = \min_{\mathbf{H}} \|\mathbf{X} - \mathbf{H}\|^2 = \min_{\mathbf{H}} \sum_{i=1}^n \sum_{j=1}^p (x_{ij} - h_{ij})^2. \quad \square \quad (1.3.1)$$

Les propriétés de cette approximation matricielle sont étudiées par Gabriel (1978). Il est à remarquer que cette façon d'approcher \mathbf{X} par une autre matrice ne fait pas intervenir les métriques dans lesquelles sont calculées les distances entre les individus et entre les variables sur le biplot. Il faut alors penser à une autre méthode d'approximation qui tient compte de ces métriques. Le critère des moindres carrés pondérés en est une.

Proposition 1.3.2. *Soient $\mathbf{M} = \text{diag}(\phi_1, \dots, \phi_p)$ et $\mathbf{N} = \text{diag}(\omega_1, \dots, \omega_n)$ les matrices définissant les métriques dans \mathbf{R}^p et \mathbf{R}^n respectivement. Si \mathbf{X} est une matrice $n \times p$ de rang r dont la décomposition est donnée en (1.1.3), alors la matrice $\mathbf{X}_{(s)} = \sum_{\alpha=1}^s \lambda_{\alpha} \mathbf{p}_{\alpha} \mathbf{q}'_{\alpha}$ est la matrice $n \times p$ de rang s avec $s < r$ qui satisfait les conditions suivantes:*

$$\begin{aligned} \text{trace}[\mathbf{N}(\mathbf{X} - \mathbf{X}_{(s)})\mathbf{M}(\mathbf{X} - \mathbf{X}_{(s)})'] &= \min_{\mathbf{H}} \text{trace}[\mathbf{N}(\mathbf{X} - \mathbf{H})\mathbf{M}(\mathbf{X} - \mathbf{H})'] \\ &= \min_{\mathbf{H}} \sum_{i=1}^n \sum_{j=1}^p \omega_i \phi_j (x_{i,j} - h_{i,j})^2 \end{aligned} \quad (1.3.2)$$

où ω_i est le poids associé à la $i^{\text{ième}}$ ligne de \mathbf{X} et ϕ_j , le poids associé à sa $j^{\text{ième}}$ colonne.

Démonstration. Posons $L(\mathbf{H}) = \text{trace}[\mathbf{N}(\mathbf{X} - \mathbf{H})\mathbf{M}(\mathbf{X} - \mathbf{H})']$. En utilisant la décom-

position DVS($\mathbf{X}, \mathbf{M}, \mathbf{N}$) définie en (1.1.4) avec les conditions (i) et (ii), nous obtenons:

$$\begin{aligned}
L(\mathbf{H}) &= \text{trace} [\mathbf{N}(\mathbf{X} - \mathbf{H})\mathbf{M}(\mathbf{X} - \mathbf{H})'] \\
&= \text{trace} \left[\mathbf{N}\tilde{\mathbf{P}}\tilde{\mathbf{P}}'\mathbf{N}(\mathbf{X} - \mathbf{H})\mathbf{M}\tilde{\mathbf{Q}}\tilde{\mathbf{Q}}'\mathbf{M}(\mathbf{X} - \mathbf{H})' \right], \quad \text{par (i) et (ii)} \\
&= \text{trace} \left[\tilde{\mathbf{P}}'\mathbf{N}(\mathbf{X} - \mathbf{H})\mathbf{M}\tilde{\mathbf{Q}}\tilde{\mathbf{Q}}'\mathbf{M}(\mathbf{X} - \mathbf{H})'\mathbf{N}\tilde{\mathbf{P}} \right] \\
&= \text{trace} \left[\tilde{\mathbf{P}}'\mathbf{N}(\tilde{\mathbf{P}}\tilde{\Lambda}\tilde{\mathbf{Q}}' - \mathbf{H})\mathbf{M}\tilde{\mathbf{Q}}\tilde{\mathbf{Q}}'\mathbf{M}(\tilde{\mathbf{P}}\tilde{\Lambda}\tilde{\mathbf{Q}}' - \mathbf{H})'\mathbf{N}\tilde{\mathbf{P}} \right] \quad \text{par (1.1.4)} \\
&= \text{trace} \left[(\tilde{\Lambda} - \mathbf{G})(\tilde{\Lambda} - \mathbf{G})' \right] = \sum_i (\lambda_i - g_{i,i})^2 + \sum_{i \neq j} g_{i,j}^2,
\end{aligned}$$

où $\mathbf{G} = \tilde{\mathbf{P}}'\mathbf{N}\mathbf{H}\mathbf{M}\tilde{\mathbf{Q}}$ et où $\tilde{\mathbf{P}}$, $\tilde{\Lambda}$ et $\tilde{\mathbf{Q}}$ sont des matrices définies en (1.1.4). Puisque \mathbf{G} a le même rang que \mathbf{H} (Graybill 1985, p.13), il est clair que (voir Greenacre 1984, p.343-344)

$$L(\mathbf{X}_{(s)}) = \min_{\mathbf{H}} L(\mathbf{H}) = \sum_{\alpha=s+1}^r \lambda_{\alpha}^2,$$

où $\mathbf{X}_{(s)} = \tilde{\mathbf{P}}\mathbf{G}\tilde{\mathbf{Q}}' = \sum_{\alpha=1}^s \lambda_{\alpha}\mathbf{p}_{\alpha}\mathbf{q}'_{\alpha}$. \square

Il est à remarquer que la matrice \mathbf{H} qui approche le mieux \mathbf{X} au sens des moindres carrés pondérés est la même que celle qui minimise la somme des distances pondérées entre ses lignes et celles de \mathbf{X} puisque

$$\begin{aligned}
L(\mathbf{H}) &= \sum_{i=1}^n \sum_{j=1}^p \omega_i \phi_j (x_{i,j} - h_{i,j})^2 \\
&= \sum_{i=1}^n \omega_i (\mathbf{x}_i - \mathbf{h}_i)' \mathbf{M} (\mathbf{x}_i - \mathbf{h}_i) \\
&= \sum_{i=1}^n \omega_i \|\mathbf{x}_i - \mathbf{h}_i\|_{\mathbf{M}}^2.
\end{aligned}$$

On peut faire un raisonnement analogue pour les colonnes.

La matrice de rang 2 qui approche le mieux \mathbf{X} au sens des moindres carrés pondérés est donc

$$\mathbf{X}_{(2)} = \sum_{\alpha=1}^2 \lambda_{\alpha}\mathbf{p}_{\alpha}\mathbf{q}'_{\alpha}. \quad (1.3.3)$$

On peut quantifier la qualité de cette approximation à partir de la mesure

$$\tau = 1 - \frac{\sum_{i=1}^n \sum_{j=1}^p \omega_i \phi_j (x_{i,j} - h_{i,j})^2}{\sum_{i=1}^n \sum_{j=1}^p \omega_i \phi_j x_{i,j}^2} = \frac{\lambda_1^2 + \lambda_2^2}{\sum_{\alpha=1}^r \lambda_{\alpha}^2}, \quad (1.3.4)$$

laquelle est établie à partir des égalités suivantes:

$$\sum_{i=1}^n \sum_{j=1}^p \omega_i \phi_j x_{i,j}^2 = \sum_1^r \lambda_\alpha^2$$

$$\sum_{i=1}^n \sum_{j=1}^p \omega_i \phi_j (x_{i,j} - h_{i,j})^2 = \sum_3^r \lambda_\alpha^2.$$

Plus la mesure τ est voisine de 1, meilleure est la qualité de l'approximation de \mathbf{X} par $\mathbf{X}_{(2)}$.

En posant $\mathbf{p}'_\alpha = (p_{\alpha,1}, \dots, p_{\alpha,n})$ et $\mathbf{q}'_\alpha = (q_{\alpha,1}, \dots, q_{\alpha,p})$ dans (1.3.3), on trouve que

$$\mathbf{X}_{(2)} = \begin{pmatrix} p_{1,1} & p_{2,1} \\ \vdots & \vdots \\ p_{1,n} & p_{2,n} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} q_{1,1} & \dots & q_{1,p} \\ q_{2,1} & \dots & q_{2,p} \end{pmatrix}. \quad (1.3.5)$$

Ainsi, pour construire le biplot des individus (voir section I.2.3), il suffit de factoriser la matrice $\mathbf{X}_{(2)}$ sous la forme (1.2.2) avec

$$\mathbf{a}'_i = (\lambda_1 p_{1,i}, \lambda_2 p_{2,i}) \quad (i = 1, \dots, n)$$

$$\mathbf{b}'_j = (q_{1,j}, q_{2,j}) \quad (j = 1, \dots, p) \quad (1.3.6)$$

de façon à ce que la contrainte (1.2.7) soit satisfaite. De plus, on obtient le biplot des variables (voir section I.2.4) en choisissant

$$\mathbf{a}'_i = (p_{1,i}, p_{2,i}) \quad (i = 1, \dots, n)$$

$$\mathbf{b}'_j = (\lambda_1 q_{1,j}, \lambda_2 q_{2,j}) \quad (j = 1, \dots, p) \quad (1.3.7)$$

laquelle factorisation vérifie la condition (1.2.10).

I.4 Liens entre le biplot et l'analyse en composantes principales

L'analyse en composantes principales (ACP) a connu son développement en 1933 avec Hotelling, mais les idées essentielles de cette analyse remontent à Pearson en 1901. Cette analyse est discutée dans tous ses détails par Volle (1980) et par Lebart, Morineau et Tabard (1977). L'ACP est, au même titre que le biplot, une technique de représentation graphique d'un ensemble de données. Nous nous proposons dans cette section d'établir les similitudes entre ces deux analyses. Mais d'abord, rappelons brièvement en quoi consiste l'ACP.

I.4.1 Analyse en composantes principales

Soit \mathbf{X} une matrice $n \times p$ sur laquelle on a opéré la transformation linéaire ramenant la moyenne de chacune de ses colonnes à 0. L'ACP de ce tableau de données comporte deux étapes: l'ajustement du nuage des individus formé de n points dans \mathbb{R}^p et l'ajustement du nuage des variables formé de p points dans \mathbb{R}^n .

Commençons d'abord par l'ajustement dans \mathbb{R}^p . Cet ajustement se fait à partir d'une métrique $\mathbf{M} = \text{diag}(\phi_1, \dots, \phi_p)$ qui définit les distances dans \mathbb{R}^p et d'une matrice de pondération $\mathbf{N} = \text{diag}(\omega_1, \dots, \omega_n)$ qui nous donne l'importance relative de chacun des individus. On ajuste le nuage des individus en cherchant dans un premier temps la direction dans l'espace \mathbb{R}^p qui explique le plus de variabilité du nuage, i.e. le plus d'inertie. Il s'agit en fait de trouver la direction \mathbf{q}_1 qui maximise la somme pondérée des carrés des n projections sur \mathbf{q}_1 , à savoir $\sum_{i=1}^n \omega_i (\mathbf{x}'_i \mathbf{M} \mathbf{q}_1)^2 = \mathbf{q}'_1 \mathbf{M} \mathbf{X}' \mathbf{N} \mathbf{X} \mathbf{M} \mathbf{q}_1$, sous la contrainte $\mathbf{q}'_1 \mathbf{M} \mathbf{q}_1 = 1$. En faisant appel au calcul classique d'extremums avec contrainte, on trouve que le maximum de cette forme quadratique doit satisfaire la relation matricielle suivante:

$$\mathbf{X}' \mathbf{N} \mathbf{X} \mathbf{M} \mathbf{u} = \lambda^2 \mathbf{u}. \quad (1.4.1)$$

En fait, \mathbf{q}_1 est le vecteur propre correspondant à λ_1^2 , la plus grande valeur propre de $\mathbf{X}' \mathbf{N} \mathbf{X} \mathbf{M}$. Plus généralement, on montre que le plan factoriel qui réalise le meilleur ajustement du nuage est formé par les vecteurs propres \mathbf{q}_1 et \mathbf{q}_2 correspondant aux deux plus grandes valeurs propres de $\mathbf{X}' \mathbf{N} \mathbf{X} \mathbf{M}$. Les coordonnées des individus sur ce plan factoriel sont données par les n lignes de la matrice

$$\mathbf{A} = (\mathbf{X} \mathbf{M} \mathbf{q}_1, \mathbf{X} \mathbf{M} \mathbf{q}_2). \quad (1.4.2)$$

A partir de cette représentation, on pourra effectuer des regroupements homogènes d'individus en mettant dans une même classe tous les individus qui sont près les uns des autres.

Réalisons maintenant l'ajustement dans \mathbb{R}^n . En général, cet ajustement est déterminé à partir du précédent en intervertissant les rôles des matrices \mathbf{M} et \mathbf{N} . La matrice \mathbf{N} définit maintenant la métrique dans \mathbb{R}^n alors que \mathbf{M} attribue les poids aux variables. Le sous-espace de dimension 2 qui réalise le meilleur ajustement du nuage des variables est formé par les vecteurs propres \mathbf{p}_1 et \mathbf{p}_2 correspondant aux deux plus

grandes valeurs propres de $\mathbf{XMX}'\mathbf{N}$. Ces vecteurs \mathbf{p}_α sont liés aux vecteurs \mathbf{q}_α de l'ajustement dans \mathbb{R}^p de la façon suivante:

$$\mathbf{p}_\alpha = \frac{\mathbf{XM}\mathbf{q}_\alpha}{\lambda_\alpha}, \quad \mathbf{q}_\alpha = \frac{\mathbf{X}'\mathbf{N}\mathbf{p}_\alpha}{\lambda_\alpha}. \quad (1.4.3)$$

Les coordonnées des p variables sur le plan factoriel formé par \mathbf{p}_1 et \mathbf{p}_2 sont données par les lignes de la matrice

$$\mathbf{B} = (\mathbf{X}'\mathbf{N}\mathbf{p}_1, \mathbf{X}'\mathbf{N}\mathbf{p}_2) = (\lambda_1\mathbf{q}_1, \lambda_2\mathbf{q}_2). \quad (1.4.4)$$

I.4.2 Relations avec l'analyse biplot

Nous avons établi en (1.3.3) que la représentation approximative de \mathbf{X} s'obtient en effectuant le biplot de

$$\mathbf{X}_{(2)} = \sum_{\alpha=1}^2 \lambda_\alpha \mathbf{p}_\alpha \mathbf{q}'_\alpha.$$

En factorisant cette matrice sous la forme \mathbf{AB}' présentée en (1.2.1), on obtient alors

$$\mathbf{X}_{(2)} = (\lambda_1^\beta \mathbf{p}_1, \lambda_2^\beta \mathbf{p}_2) (\lambda_1^{1-\beta} \mathbf{q}_1, \lambda_2^{1-\beta} \mathbf{q}_2)' = \mathbf{AB}', \quad (1.4.5)$$

$\forall \beta \in [0, 1]$. En particulier, si on fixe $\beta = 0$, cette décomposition s'établit en posant $\mathbf{A} = (\mathbf{p}_1, \mathbf{p}_2)$ et $\mathbf{B} = (\lambda_1 \mathbf{q}_1, \lambda_2 \mathbf{q}_2)$. Ces matrices sont celles utilisées dans la construction du biplot des variables (voir (1.3.7)). De par la relation (1.4.4), on déduit que la représentation des variables sur ce type de biplot est la même que celle effectuée en ACP. On peut se servir de la relation (1.2.12) pour interpréter la proximité des variables sur ces graphiques. Il est à remarquer que les axes utilisés en ACP pour représenter ces variables dans le meilleur sous-espace de dimension 2 correspondent aux colonnes de la matrice \mathbf{A} .

De plus, si on fixe $\beta = 1$ dans (1.4.5), on obtient $\mathbf{A} = (\lambda_1 \mathbf{p}_1, \lambda_2 \mathbf{p}_2)$ et $\mathbf{B} = (\mathbf{q}_1, \mathbf{q}_2)$. Ce sont les matrices utilisées dans la construction du biplot des individus (voir (1.3.6)). En insérant la première égalité de (1.4.3) dans (1.4.2), on constate que les coordonnées des individus sur ce type de biplot et en ACP sont les mêmes. De plus, les axes factoriels utilisés en ACP pour représenter ces individus dans le meilleur sous-espace de dimension 2 correspondent aux colonnes de la matrice \mathbf{B} .

CHAPITRE II

Estimation robuste de μ et Σ

L'influence des observations extrêmes sur l'estimation des paramètres préoccupe les statisticiens depuis fort longtemps (voir Stigler (1963)). Cela a permis d'élargir les cadres de la statistique en introduisant le concept de robustesse. De façon intuitive, un estimateur est dit robuste s'il possède les deux caractéristiques suivantes:

- 1) il a une bonne efficacité dans le modèle paramétrique choisi
- 2) il est peu affecté par un léger changement dans la plupart des observations ou par une grande modification de certaines d'entre elles.

Une définition rigoureuse de la robustesse d'une statistique est présentée par Hampel (1971). En général, on utilise une statistique robuste pour se prémunir de l'impact des observations extrêmes sur son comportement.

Les estimateurs robustes du paramètre de localisation μ et du paramètre de dispersion Σ présentés dans ce chapitre ont été suggérés par Maronna (1976) et Huber (1981, p.211-242). Ces estimateurs constituent une généralisation des estimateurs du maximum de vraisemblance (EMV). Ainsi, nous allons commencer ce chapitre par quelques rappels sur les EMV. Nous allons ensuite introduire une nouvelle paramétrisation de Σ qui facilitera l'étude des propriétés locales des fonctions à partir desquelles

les estimateurs robustes sont définis. Nous terminerons en présentant un algorithme nous permettant de calculer explicitement les estimations des paramètres $\boldsymbol{\mu}$ et $\boldsymbol{\Sigma}$.

Il existe plusieurs situations en analyse multivariée où l'estimation robuste des paramètres $\boldsymbol{\mu}$ et $\boldsymbol{\Sigma}$ a été employée avec succès. Parmi ces domaines d'application, il y a l'analyse en composantes principales (Rivest, Plante 1988), l'analyse discriminante (Lachenbruch et al. 1973) et bien sûr l'analyse biplot.

II.1 Estimation par le maximum de vraisemblance

Soit $f_{\mathbf{Y}}(\mathbf{y}) = f(\|\mathbf{y}\|)$ la fonction de densité du vecteur aléatoire $\mathbf{Y} \in \mathbb{R}^p$ où la norme euclidienne est représentée par $\|\cdot\|$. Cette densité est symétrique par rapport à l'origine et invariante pour des transformations orthogonales. Cela veut dire que \mathbf{Y} et $\mathbf{R}\mathbf{Y}$ ont même loi, $\forall \mathbf{R}_{p \times p}$ orthogonale. Le modèle paramétrique \mathcal{F} considéré dans ce chapitre est constitué par l'ensemble des lois obtenues en appliquant les transformations affines

$$\varphi_{\mathbf{A}, \boldsymbol{\mu}}(Y) = \mathbf{A}\mathbf{Y} + \boldsymbol{\mu}, \quad \boldsymbol{\mu} \in \mathbb{R}^p, \quad \mathbf{A} \text{ inversible} \quad (2.1.1)$$

au vecteur aléatoire \mathbf{Y} . La paire $(\boldsymbol{\mu}, \mathbf{A})$ ne constitue pas une paramétrisation acceptable pour \mathcal{F} . En effet, on peut obtenir la même loi à partir des deux paires $(\boldsymbol{\mu}, \mathbf{A})$ et $(\boldsymbol{\mu}, \mathbf{A}\mathbf{R})$ avec \mathbf{R} orthogonale. Ceci nous laisse voir cependant la possibilité de paramétriser le modèle \mathcal{F} à l'aide de la paire $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ où $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}'$. C'est la paramétrisation proposée par Hampel, Ronchetti, Rousseeuw et Stahel (1986, p.271,272). Avec cette paramétrisation, on pourra par exemple obtenir la loi de \mathcal{F} associée à $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ à partir de la transformation $\varphi_{\boldsymbol{\Sigma}_0^{1/2}, \boldsymbol{\mu}_0}$. La matrice $\boldsymbol{\Sigma}_0^{1/2}$ est celle qui résulte de la décomposition de Cholesky de $\boldsymbol{\Sigma}_0$; elle est triangulaire inférieure avec des éléments positifs sur la diagonale. Le modèle paramétrique \mathcal{F} est en fait un modèle de localisation-échelle généré par la loi du vecteur aléatoire \mathbf{Y} .

Les lois du modèle \mathcal{F} admettent des densités de la forme

$$g(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \det(\boldsymbol{\Sigma}^{-1/2}) f\left(\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})\|\right), \quad \forall \mathbf{x} \in \mathbb{R}^p. \quad (2.1.2)$$

Ces densités ont une forme *elliptique* puisqu'elles sont constantes en tous les points de \mathbb{R}^p situés sur une même ellipsoïde. Par exemple, sur l'ellipsoïde définie par

$\{\mathbf{x} \in \mathbb{R}^p; (\mathbf{x} - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}_0^{-1} (\mathbf{x} - \boldsymbol{\mu}_0) = c\}$ avec $c > 0$, la densité $g(\cdot; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ prend la valeur $\det(\boldsymbol{\Sigma}_0^{-1/2}) f(\sqrt{c})$.

Soit $\mathbf{X}_1, \dots, \mathbf{X}_n$ un échantillon d'une loi appartenant au modèle \mathcal{F} défini en (2.1.2). Les estimateurs du maximum de vraisemblance de $\boldsymbol{\mu}$ et $\boldsymbol{\Sigma}$ sont définis par les statistiques qui maximisent la fonction $\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{X}_1, \dots, \mathbf{X}_n) = \prod_{i=1}^n g(\mathbf{X}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, ou de façon équivalente, celles qui minimisent l'expression

$$\begin{aligned} \mathcal{S}(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{X}_1, \dots, \mathbf{X}_n) &= -\log \mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\Sigma}; \mathbf{X}_1, \dots, \mathbf{X}_n) \\ &= -n \log(\det \boldsymbol{\Sigma}^{-1/2}) - \sum_{i=1}^n \log f\left(\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{X}_i - \boldsymbol{\mu})\|\right). \end{aligned} \quad (2.1.3)$$

Cette fonction admet un extremum pour les valeurs de $\boldsymbol{\mu}$ et $\boldsymbol{\Sigma}$ qui annulent ses dérivées partielles. Or, en utilisant les résultats de Dwyer (1967), à savoir

$$\begin{aligned} \frac{d \left[-\log(\det \boldsymbol{\Sigma}^{-1/2}) \right]}{d\boldsymbol{\Sigma}} &= \frac{1}{2} \frac{[d \log(\det \boldsymbol{\Sigma})]}{d\boldsymbol{\Sigma}} = \frac{\boldsymbol{\Sigma}^{-1}}{2} \\ \frac{\partial \left[\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})\| \right]}{\partial \boldsymbol{\Sigma}} &= \frac{-\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}}{2\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})\|} \\ \frac{\partial \left[\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})\| \right]}{\partial \boldsymbol{\mu}} &= \frac{-\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})}{\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})\|}, \end{aligned}$$

on en déduit que les dérivées partielles de la fonction \mathcal{S} s'écrivent comme suit:

$$\frac{\partial \mathcal{S}(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu}} = \sum_{i=1}^n \frac{f'}{f}(\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{X}_i - \boldsymbol{\mu})\|) \frac{\boldsymbol{\Sigma}^{-1}(\mathbf{X}_i - \boldsymbol{\mu})}{\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{X}_i - \boldsymbol{\mu})\|} \quad (2.1.4)$$

$$\frac{\partial \mathcal{S}(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}} = \frac{n}{2} \boldsymbol{\Sigma}^{-1} + \frac{1}{2} \sum_{i=1}^n \frac{f'}{f}(\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{X}_i - \boldsymbol{\mu})\|) \frac{\boldsymbol{\Sigma}^{-1}(\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}}{\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{X}_i - \boldsymbol{\mu})\|}.$$

Ces résultats sont valides en autant que la fonction f est différentiable. De ces équations, on en déduit que les statistiques $\hat{\boldsymbol{\mu}}(\mathbf{X}_1, \dots, \mathbf{X}_n)$ et $\hat{\boldsymbol{\Sigma}}(\mathbf{X}_1, \dots, \mathbf{X}_n)$ qui annulent les dérivées partielles de la fonction \mathcal{S} sont les solutions du système implicite suivant:

$$\hat{\boldsymbol{\mu}} = \frac{\sum_{i=1}^n w \left[\|\hat{\boldsymbol{\Sigma}}^{-1/2}(\mathbf{X}_i - \hat{\boldsymbol{\mu}})\| \right] \mathbf{X}_i}{\sum_{i=1}^n w \left[\|\hat{\boldsymbol{\Sigma}}^{-1/2}(\mathbf{X}_i - \hat{\boldsymbol{\mu}})\| \right]} \quad (2.1.5)$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^n w \left[\|\hat{\boldsymbol{\Sigma}}^{-1/2}(\mathbf{X}_i - \hat{\boldsymbol{\mu}})\| \right] (\mathbf{X}_i - \hat{\boldsymbol{\mu}})(\mathbf{X}_i - \hat{\boldsymbol{\mu}})',$$

avec $w(y) = -f'(y)/[yf(y)]$. Pour vérifier si ces solutions correspondent aux estimateurs du maximum de vraisemblance, il faudrait aussi vérifier les conditions du second ordre de la fonction \mathcal{S} . En fait, il faudrait voir si la matrice des dérivées secondes est définie positive pour toutes les valeurs de $\boldsymbol{\mu}$ et $\boldsymbol{\Sigma}$ qui satisfont (2.1.5). Or, cette condition est difficile à vérifier avec la paramétrisation de \mathcal{F} utilisée dans cette section. Il faudra attendre la nouvelle paramétrisation de \mathcal{F} introduite dans la section suivante pour vérifier que les solutions de $\hat{\boldsymbol{\mu}}$ et $\hat{\boldsymbol{\Sigma}}$ du système (2.1.5) sont en effet les estimateurs du maximum de vraisemblance pour les paramètres $\boldsymbol{\mu}$ et $\boldsymbol{\Sigma}$.

Exemple 2.1.1. Soit \mathbf{Y} un vecteur aléatoire d'une loi $\mathcal{N}(0, I_p)$. La densité de ce vecteur aléatoire est donnée par

$$f(\|\mathbf{y}\|) = (2\pi)^{-p/2} \exp \frac{-\|\mathbf{y}\|^2}{2}, \quad \mathbf{y} \in \mathbb{R}^p.$$

Le modèle paramétrique formé en appliquant les transformations affines définies en (2.1.1) sur le vecteur aléatoire \mathbf{Y} est le modèle de localisation-échelle normal. Pour ce modèle, nous avons que $w(y) \equiv 1$. Si $\mathbf{X}_1, \dots, \mathbf{X}_n$ est un échantillon d'une loi appartenant à ce modèle, l'équation (2.1.5) devient alors

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \\ \hat{\boldsymbol{\Sigma}} &= \frac{1}{n} \sum_{i=1}^n (\mathbf{X}_i - \hat{\boldsymbol{\mu}})(\mathbf{X}_i - \hat{\boldsymbol{\mu}})'. \end{aligned}$$

Ces statistiques correspondent aux estimateurs du maximum de vraisemblance de $\boldsymbol{\mu}$ et $\boldsymbol{\Sigma}$ (voir Anderson (1984), p.63).

Exemple 2.1.2. Soit \mathbf{Y} un vecteur aléatoire d'une loi de Student à k degrés de liberté. La densité de ce vecteur aléatoire est donnée par

$$f(\|\mathbf{y}\|) = C(k + \|\mathbf{y}\|^2)^{-(p+k)/2}, \quad \mathbf{y} \in \mathbb{R}^p,$$

où C est une constante de normalisation. En appliquant les transformations affines définies en (2.1.1) sur le vecteur aléatoire \mathbf{Y} , on forme le modèle de localisation-échelle de la famille des lois Student à k degrés de liberté. Si $\mathbf{X}_1, \dots, \mathbf{X}_n$ est un échantillon

d'une loi de ce modèle, alors on peut montrer que les statistiques obtenues en posant

$$\begin{aligned} w(y) &= \frac{-1}{y} \frac{d \log f(y)}{dy} \\ &= \frac{-1}{y} \frac{d}{dy} \left[\log C - \frac{p+k}{2} \log(k+y^2) \right] \\ &= \frac{p+k}{k+y^2}. \end{aligned}$$

dans (2.1.5) correspondent aux estimateurs du maximum de vraisemblance de $\boldsymbol{\mu}$ et $\boldsymbol{\Sigma}$ (Voir Maronna (1976)).

II.2 Paramétrisation d'une matrice définie positive

Soit $\boldsymbol{\Sigma}$ une matrice $p \times p$ définie positive. Cette matrice est complètement déterminée par la matrice $\boldsymbol{\Sigma}^{1/2}$ résultant de la décomposition $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{1/2} \boldsymbol{\Sigma}^{1/2'}$. En fait, $\boldsymbol{\Sigma}^{1/2}$ est la seule matrice appartenant à l'ensemble G des matrices triangulaires inférieures $p \times p$ avec des éléments positifs sur la diagonale vérifiant cette condition. Cet ensemble G , lorsque muni de la multiplication matricielle comme loi de composition interne, forme un groupe multiplicatif. En effet, toutes les propriétés inhérentes à ce type de groupe sont satisfaites, i.e.

- 1) la multiplication matricielle est associative:

$$(\mathbf{A}\mathbf{B})\mathbf{C} = \mathbf{A}(\mathbf{B}\mathbf{C}) \quad \forall \mathbf{A}, \mathbf{B}, \mathbf{C} \in G$$

- 2) il existe dans G un élément neutre, à savoir \mathbf{I}_p , tel que

$$\mathbf{A}\mathbf{I}_p = \mathbf{I}_p\mathbf{A} = \mathbf{A}; \quad \forall \mathbf{A} \in G$$

- 3) $\forall \mathbf{A} \in G, \exists \mathbf{A}^{-1} \in G$ telle que

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_p.$$

Plus précisément, cet ensemble G forme un groupe de Lie. L'algèbre de Lie \mathcal{G} qui lui est associée est l'espace vectoriel engendré par les matrices triangulaires inférieures de dimension $p \times p$ (Warner 1983, p.82-84).

La nouvelle paramétrisation proposée dans cette section se base sur le fait que toute matrice $\boldsymbol{\Sigma}^{1/2} \in G$ peut s'exprimer comme l'exponentielle d'une certaine matrice

de l'algèbre \mathcal{G} . Par exponentielle d'une matrice, on entend la fonction qui à $\mathbf{B} \in \mathcal{G}$ associe la quantité

$$\exp(\mathbf{B}) = \mathbf{I}_p + \mathbf{B} + \frac{\mathbf{B}^2}{2!} + \dots + \frac{\mathbf{B}^n}{n!} + \dots \quad (2.2.1)$$

L'inverse de cette fonction matricielle est toujours défini puisque

$$\begin{aligned} \det [\exp(\mathbf{B})] &= \prod_{i=1}^p \exp(b_{i,i}) \\ &= \exp\left(\sum_{i=1}^p b_{i,i}\right) \\ &= \exp(\text{trace}(\mathbf{B})) > 0. \end{aligned} \quad (2.2.2)$$

En fait, on peut montrer que $[\exp(\mathbf{B})]^{-1} = \exp(-\mathbf{B})$ (Warner (1983), p.102-108). Nous allons consacrer cette section à montrer que l'on peut établir une bijection entre les deux ensembles \mathcal{G} et G à partir de la fonction exponentielle.

II.2.1 Définition de la fonction matricielle logarithmique

Le développement en série de Taylor de la fonction logarithmique autour du point $\lambda > 0$ s'écrit comme

$$\log(x) = \log(\lambda) - \sum_{i=1}^{\infty} \frac{1}{i} \left(\frac{\lambda - x}{\lambda}\right)^i. \quad (2.2.3)$$

Cette série converge pour toutes les valeurs de x situées à l'intérieur du rayon de convergence, i.e. pour toutes les valeurs de x telles que $|x - \lambda| < \lambda$. C'est dans la même veine d'idées que nous allons définir la fonction matricielle logarithmique. Cependant, cette fonction ne sera définie que pour les matrices dont les valeurs propres sont positives, condition qui est satisfaite pour les matrices de l'ensemble G .

Définition 2.2.1. Un bloc de Jordan $\mathbf{J}_n(\lambda)$ est une matrice $n \times n$ qui possède les caractéristiques suivantes:

- 1) Tous les éléments sur la diagonale principale sont égaux à λ .
- 2) Tous les éléments sur la diagonale supérieure sont égaux à 1.
- 3) Tous les autres éléments sont nuls.

Méthode pour déterminer le logarithme de $\Sigma^{1/2} \in G$ (Pullman 1976, p.160)

- 1) Trouvons une décomposition de Jordan de $\Sigma^{1/2}$, i.e. trouvons une matrice \mathbf{U} et des blocs de Jordan $\mathbf{J}_{m_1}(\lambda_1), \dots, \mathbf{J}_{m_s}(\lambda_s)$ tels que

$$\Sigma^{1/2} = \mathbf{U} \text{diag} [\mathbf{J}_{m_i}(\lambda_i)] \mathbf{U}^{-1}, \quad (2.2.4)$$

où

$$\text{diag} [\mathbf{J}_{m_i}(\lambda_i)] = \begin{pmatrix} \mathbf{J}_{m_1}(\lambda_1) & \mathbf{O} & \dots & \mathbf{O} \\ \mathbf{O} & \mathbf{J}_{m_2}(\lambda_2) & \dots & \mathbf{O} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O} & \dots & \mathbf{O} & \mathbf{J}_{m_s}(\lambda_s) \end{pmatrix}$$

et où $\lambda_1, \dots, \lambda_s$ sont les valeurs propres de $\Sigma^{1/2}$.

- 2) Ecrivons chacun des blocs $\mathbf{J}_{m_i}(\lambda_i)$ comme suit:

$$\mathbf{J}_{m_i}(\lambda_i) = \lambda_i \mathbf{I}_{m_i} + \mathbf{N}_i,$$

où les éléments $n_{i,j}$ de la matrice \mathbf{N}_i sont égaux à 1 si $i + 1 = j$ et sont nuls autrement.

- 3) En utilisant un argument analogue à celui de (2.2.3), on définit le logarithme de $\Sigma^{1/2}$ comme étant

$$\begin{aligned} \log(\Sigma^{1/2}) &= \mathbf{U} \text{diag} [\log \mathbf{J}_{m_i}(\lambda_i)] \mathbf{U}^{-1} \\ &= \mathbf{U} \text{diag} \left[\log(\lambda_i) \mathbf{I}_{m_i} - \sum_{j=1}^{\infty} \frac{1}{j} \left(\frac{-\mathbf{N}_i}{\lambda_i} \right)^j \right] \mathbf{U}^{-1} \\ &= \mathbf{U} \text{diag} \left[\log(\lambda_i) \mathbf{I}_{m_i} - \sum_{j=1}^{m_i-1} \frac{1}{j} \left(\frac{-\mathbf{N}_i}{\lambda_i} \right)^j \right] \mathbf{U}^{-1}. \end{aligned} \quad (2.2.5)$$

Pour établir la première égalité, on s'est servi des théorèmes 2 et 3 de Lancaster et Tismenetsky (1985, p.310-311). La dernière égalité vient du fait que les puissances de \mathbf{N}_i d'ordre supérieur à $m_i - 1$ sont nulles.

II.2.2 Surjectivité de l'application exponentielle

Pour montrer que l'application $\exp : \mathcal{G} \rightarrow G$ est surjective, il faut vérifier que pour toute matrice $\Sigma^{1/2} \in G$, il existe une matrice $\mathbf{B} \in \mathcal{G}$ telle que $\exp(\mathbf{B}) = \Sigma^{1/2}$.

Ceci va se faire en montrant que la matrice $\mathbf{B} = \log(\boldsymbol{\Sigma}^{1/2})$ définie en (2.2.5) vérifie cette condition. En fait, il suffit de vérifier que l'expression entre crochets dans la dernière égalité de (2.2.5), i.e.

$$\mathbf{L}_i = \log(\lambda_i)\mathbf{I}_{m_i} - \sum_{j=1}^{m_i-1} \frac{1}{j} \left(\frac{-\mathbf{N}_i}{\lambda_i} \right)^j,$$

satisfait l'égalité suivante $\exp(\mathbf{L}_i) = \mathbf{J}_{m_i}(\lambda_i)$, $\forall i \in \{1, \dots, s\}$. En effet, si cette égalité est satisfaite nous aurons alors que (voir Lancaster et Tismenetsky (1985), p.310-311)

$$\begin{aligned} \exp\left(\log \boldsymbol{\Sigma}^{1/2}\right) &= \mathbf{U} \text{diag} [\exp(\mathbf{L}_i)] \mathbf{U}^{-1} \\ &= \mathbf{U} \text{diag} [\mathbf{J}_{m_i}(\lambda_i)] \mathbf{U}^{-1} \\ &= \boldsymbol{\Sigma}^{1/2}, \end{aligned}$$

ce qui est à démontrer.

Posons $\mathbf{V}_i = (\mathbf{N}_i/\lambda_i)$ et $\mathbf{W}_i(t) = \sum_{j=1}^{m_i-1} (1/j)(t\mathbf{V}_i)^j$, $\forall t \in \mathbb{R}$. Pullman (1976, p.160,161) montre que les matrices $\mathbf{I}_{m_i} - t\mathbf{V}_i$ et $\exp[-\mathbf{W}_i(t)]$ sont toutes deux solutions de l'équation différentielle $(d\mathbf{X}_i/dt) = -\mathbf{V}_i(\mathbf{I}_{m_i} - t\mathbf{V}_i)^{-1}\mathbf{X}_i$. De plus, ces deux matrices sont égales à \mathbf{I}_{m_i} dans le cas particulier où $t = 0$. Par conséquent, en se servant du résultat de Pullman (1976, p.133 théorème 1) qui nous assure de l'unicité de la solution $\mathbf{X}_i(t)$ de l'équation

$$\frac{d\mathbf{X}_i}{dt} = -\mathbf{V}_i(\mathbf{I}_{m_i} - t\mathbf{V}_i)^{-1}\mathbf{X}_i(t)$$

sous la condition $\mathbf{X}_i(0) = \mathbf{I}_{m_i}$, nous avons alors que $[\mathbf{I}_{m_i} - t\mathbf{V}_i] = \exp(-\mathbf{W}_i(t))$, $\forall t \in \mathbb{R}$. En particulier, pour $t = -1$, on a que $[\mathbf{I}_{m_i} + \mathbf{V}_i] = \exp(-\mathbf{W}_i(-1))$. Cette égalité établie, nous pouvons alors en déduire les égalités matricielles suivantes:

$$\begin{aligned} \exp(\mathbf{L}_i) &= \exp[\log(\lambda_i)\mathbf{I}_{m_i}] \exp(-\mathbf{W}_i(-1)) \\ &= (\lambda_i\mathbf{I}_{m_i})(\mathbf{I}_{m_i} + \mathbf{V}_i) \\ &= \lambda_i\mathbf{I}_{m_i} + \mathbf{N}_i \\ &= \mathbf{J}_{m_i}(\lambda_i). \end{aligned}$$

Nous avons donc démontré la proposition ci-dessous.

Proposition 2.2.1. *L'application exponentielle $\exp : \mathcal{G} \rightarrow G$ est surjective, i.e. pour toute matrice $\boldsymbol{\Sigma}^{1/2} \in G$, il existe une matrice \mathbf{B} telle que $\exp(\mathbf{B}) = \boldsymbol{\Sigma}^{1/2}$. La matrice $\mathbf{B} = \log(\boldsymbol{\Sigma}^{1/2})$ définie en (2.2.5) satisfait cette condition.*

□

Pour calculer explicitement la matrice $\log(\Sigma^{1/2})$, il n'est pas toujours nécessaire d'exprimer $\Sigma^{1/2}$ sous la forme (2.2.4) comme le montre la proposition suivante.

Proposition 2.2.2. *Soit $\Sigma^{1/2} \in G$ une matrice qui admet p valeurs propres positives et distinctes $\lambda_1, \dots, \lambda_p$. Si $\Sigma^{1/2} = \mathbf{U} \text{diag}(\lambda_i) \mathbf{U}^{-1}$ est la décomposition de Jordan de cette matrice, alors il y a équivalence entre les énoncés suivants:*

- i) $\log(\Sigma^{1/2}) = \mathbf{U} \text{diag}(\log \lambda_i) \mathbf{U}^{-1}$ (voir 2.2.5)
- ii) $\log(\Sigma^{1/2}) = \sum_{k=1}^p \log(\lambda_k) \left[\prod_{j \neq k}^p (\Sigma^{1/2} - \lambda_j \mathbf{I}_p) / \prod_{j \neq k}^p (\lambda_k - \lambda_j) \right]$.

Démonstration. Notons par $\{e_1, \dots, e_p\}$ la base euclidienne usuelle de l'espace à p dimensions. Les deux expressions (i) et (ii) sont équivalentes puisque

$$\begin{aligned}
 \log(\Sigma^{1/2}) &= \sum_{k=1}^p \log(\lambda_k) \left[\prod_{j \neq k}^p (\Sigma^{1/2} - \lambda_j \mathbf{I}_p) / \prod_{j \neq k}^p (\lambda_k - \lambda_j) \right] \\
 &= \sum_{k=1}^p \log(\lambda_k) \left\{ \frac{\prod_{j \neq k}^p \mathbf{U} [\text{diag}(\lambda_i) - \lambda_j \mathbf{I}_p] \mathbf{U}^{-1}}{\prod_{j \neq k}^p (\lambda_k - \lambda_j)} \right\} \\
 &= \sum_{k=1}^p \log(\lambda_k) \left\{ \frac{\mathbf{U} \left\{ \prod_{j \neq k}^p [\text{diag}(\lambda_i) - \lambda_j \mathbf{I}_p] \right\} \mathbf{U}^{-1}}{\prod_{j \neq k}^p (\lambda_k - \lambda_j)} \right\} \\
 &= \sum_{k=1}^p \log(\lambda_k) \mathbf{U} e_k e'_k \mathbf{U}^{-1} \\
 &= \mathbf{U} \left[\sum_{k=1}^p e_k \log(\lambda_k) e'_k \right] \mathbf{U}^{-1} \\
 &= \mathbf{U} \text{diag}(\log \lambda_i) \mathbf{U}^{-1}.
 \end{aligned}$$

□

Une généralisation de ce résultat est présentée par Lancaster et Tismenetsky (1985, p.308). Ces auteurs discutent du cas où certaines valeurs propres de $\Sigma^{1/2}$ se répètent.

Exemple 2.2.1. Soient a et c deux réels positifs distincts. Si on définit

$$\Sigma^{1/2} = \begin{pmatrix} a & 0 \\ b & c \end{pmatrix},$$

on trouve alors par la proposition 2.2.2 que

$$\begin{aligned}\log(\boldsymbol{\Sigma}^{1/2}) &= \frac{\log(a)}{a-c} \left[\boldsymbol{\Sigma}^{1/2} - c\mathbf{I}_2 \right] + \frac{\log(c)}{c-a} \left[\boldsymbol{\Sigma}^{1/2} - a\mathbf{I}_2 \right] \\ &= \begin{pmatrix} \log(a) & 0 \\ \beta & \log(c) \end{pmatrix}\end{aligned}$$

où $\beta = b[\log(a) - \log(c)]/(a - c)$. Notons par \mathbf{B} la matrice $\log(\boldsymbol{\Sigma}^{1/2})$. Cette matrice appartient à l'algèbre de Lie \mathcal{G} . Elle est celle qui, d'après la proposition 2.2.1, satisfait la condition $\exp(\mathbf{B}) = \boldsymbol{\Sigma}^{1/2}$. En effet, on peut vérifier que les puissances de \mathbf{B} sont de la forme

$$\mathbf{B}^n = \begin{pmatrix} \log^n(a) & 0 \\ \gamma_n & \log^n(c) \end{pmatrix},$$

avec $\gamma_n = \beta \log^{n-1}(c) + \log(a)\gamma_{n-1}, \forall n \in \{1, 2, \dots\}$ et $\gamma_0 \equiv 0$. On obtient alors par récurrence $\gamma_n = \beta \sum_{j=0}^{n-1} \log^{n-1-j}(c) \log^j(a)$. On montre que $\exp(\mathbf{B}) = \boldsymbol{\Sigma}^{1/2}$ en se servant de la définition de la fonction matricielle exponentielle et du fait que

$$\begin{aligned}\sum_{n=0}^{\infty} \frac{\gamma_n}{n!} &= \beta \sum_{n=1}^{\infty} \sum_{j=0}^{n-1} \frac{\log^{n-1-j}(c) \log^j(a)}{n!} \\ &= \beta \sum_{k=0}^{\infty} \sum_{j=0}^k \frac{\log^{k-j}(c) \log^j(a)}{(k+1)!} \\ &= \beta \sum_{k=0}^{\infty} \frac{\log^k(c)}{(k+1)!} \sum_{j=0}^k (\log(a)/\log(c))^j \\ &= \beta \sum_{k=0}^{\infty} \frac{\log^k(c)}{(k+1)!} \left\{ \frac{1 - [\log(a)/\log(c)]^{k+1}}{1 - [\log(a)/\log(c)]} \right\} \\ &= \frac{\beta}{\log(c) - \log(a)} \left[\sum_{k=0}^{\infty} \frac{\log^{k+1}(c)}{(k+1)!} - \sum_{k=0}^{\infty} \frac{\log^{k+1}(a)}{(k+1)!} \right] \\ &= \frac{\beta(c-a)}{\log(c) - \log(a)} = b.\end{aligned}$$

□

II.2.3 Injectivité de l'application exponentielle

Supposons que $\exp(\mathbf{A}) = \exp(\mathbf{B})$ où \mathbf{A} et \mathbf{B} appartiennent à \mathcal{G} , on veut montrer que $\mathbf{A} = \mathbf{B}$.

Lemme 2.2.3. *Soient \mathbf{A} et \mathbf{B} deux matrices appartenant à G et définissons $\mathbf{A}^{1/2}$ et $\mathbf{B}^{1/2}$ des matrices triangulaires inférieures avec des éléments positifs sur la diagonale telles que*

$$\left[\mathbf{A}^{1/2}\right]^2 = \mathbf{A} \quad \text{et} \quad \left[\mathbf{B}^{1/2}\right]^2 = \mathbf{B}.$$

Alors $\mathbf{A} = \mathbf{B} \Leftrightarrow \mathbf{A}^{1/2} = \mathbf{B}^{1/2}$.

Démonstration. Considérons le cas où \mathbf{A} est $p \times p$. Pour démontrer le résultat, il suffit de vérifier que la relation qui associe $\mathbf{A}^{1/2}$ à \mathbf{A} est bien une fonction. La preuve se fait par récurrence. Le résultat est évidemment vrai pour $p = 1$. Pour déterminer $\mathbf{A}^{1/2}$ lorsque $p > 1$, on doit résoudre

$$\mathbf{X}^2 = \mathbf{A}. \tag{2.2.6}$$

Pour ce faire, posons

$$\mathbf{X} = \begin{pmatrix} X_{1,1} & \mathbf{O} \\ \mathbf{X}_{2,1} & \mathbf{X}_{2,2} \end{pmatrix}$$

où $X_{1,1}$ est 1×1 , $\mathbf{X}_{2,1}$ est $(p-1) \times 1$ et $\mathbf{X}_{2,2}$ est $(p-1) \times (p-1)$. On peut alors écrire (2.2.6) comme

$$\begin{pmatrix} X_{1,1}^2 & \mathbf{O} \\ [X_{1,1}\mathbf{I}_{p-1} + \mathbf{X}_{2,2}] \mathbf{X}_{2,1} & \mathbf{X}_{2,2}^2 \end{pmatrix} = \begin{pmatrix} A_{1,1} & \mathbf{O} \\ \mathbf{A}_{2,1} & \mathbf{A}_{2,2} \end{pmatrix}.$$

En se servant de l'hypothèse de récurrence qui suppose le résultat vrai pour $p-1$, nous avons que les matrices $X_{1,1}$ et $\mathbf{X}_{2,2}$ définies par

$$X_{1,1}^2 = A_{1,1} \quad \text{et} \quad \mathbf{X}_{2,2}^2 = \mathbf{A}_{2,2}$$

sont uniques. En fait, ce sont les seules matrices triangulaires inférieures avec des éléments positifs sur la diagonale satisfaisant les relations ci-haut. Quant à la matrice $\mathbf{X}_{2,1}$, elle est déterminée par

$$[X_{1,1}\mathbf{I}_{p-1} + \mathbf{X}_{2,2}] \mathbf{X}_{2,1} = \mathbf{A}_{2,1}, \tag{2.2.7}$$

qui est un système d'équations linéaires à $p-1$ inconnus. Ce système définit $\mathbf{X}_{2,1}$ de façon unique puisque $X_{1,1}\mathbf{I}_{p-1} + \mathbf{X}_{2,2}$ est de plein rang, i.e. inversible. \square

Revenons au problème initial. Puisque $\exp(\mathbf{A}/2)$ est une racine de $\exp(\mathbf{A})$, elle est la seule racine et $\exp(\mathbf{A}) = \exp(\mathbf{B})$ implique que $\exp(\mathbf{A}/2) = \exp(\mathbf{B}/2)$. De façon générale, nous avons que

$$\exp(\mathbf{A}/2^k) = \exp(\mathbf{B}/2^k), \quad \forall k > 0.$$

De plus, lorsque k est assez grand, nous pouvons écrire

$$\exp(\mathbf{A}/2^k) = \mathbf{I}_p + \frac{\mathbf{A}}{2^k} + o\left(\frac{1}{2^k}\right).$$

Ainsi, $\exp(\mathbf{A}/2^k) = \exp(\mathbf{B}/2^k)$ implique que $\mathbf{A} = \mathbf{B}$. Ceci nous assure que la fonction logarithmique matricielle est uniquement définie sur G .

II.3 Estimation robuste de $\boldsymbol{\mu}$ et $\boldsymbol{\Sigma}$

Soit $\mathbf{X}_1, \dots, \mathbf{X}_n$ un échantillon d'une loi $F(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ définie sur \mathbb{R}^p . Le vecteur $\boldsymbol{\mu} \in \mathbb{R}^p$ et la matrice $\boldsymbol{\Sigma}$ sont respectivement les paramètres de localisation et d'échelle de l'échantillon multivarié. On définit les estimateurs robustes de $\boldsymbol{\mu}$ et $\boldsymbol{\Sigma}$ par les statistiques $\hat{\boldsymbol{\mu}}(\mathbf{X}_1, \dots, \mathbf{X}_n)$ et $\hat{\boldsymbol{\Sigma}}(\mathbf{X}_1, \dots, \mathbf{X}_n)$ qui minimisent la fonction

$$Q(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = -n \log \left(\det \boldsymbol{\Sigma}^{-1/2} \right) + \sum_{i=1}^n \rho(\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{X}_i - \boldsymbol{\mu})\|), \quad (2.3.1)$$

où ρ est une certaine fonction non constante. La classe d'estimateurs ainsi définis constitue une généralisation multivariée de la classe des estimateurs du maximum de vraisemblance. On obtient ces derniers en posant $\rho = -\log f$ dans (2.3.1).

Dans cette section, nous allons étudier les propriétés locales de la fonction Q . En fait, nous allons examiner le comportement de Q dans un certain voisinage de l'un des deux paramètres en considérant l'autre fixé.

II.3.1 Estimation robuste de $\boldsymbol{\Sigma}$ avec $\boldsymbol{\mu}$ connu

Supposons sans perte de généralité que le paramètre de localisation $\boldsymbol{\mu}$ est nul. Pour simplifier la notation, convenons d'utiliser $Q(\boldsymbol{\Sigma})$ pour désigner $Q(0, \boldsymbol{\Sigma})$. Dans ce cas particulier, la fonction Q s'écrit comme suit:

$$Q(\boldsymbol{\Sigma}) = -n \log(\det \boldsymbol{\Sigma}^{-1/2}) + \sum_{i=1}^n \rho(\|\boldsymbol{\Sigma}^{-1/2} \mathbf{X}_i\|). \quad (2.3.2)$$

Nous allons étudier le comportement local de cette fonction dans un voisinage d'une certaine matrice Σ_0 . Ce voisinage est déterminé par l'ensemble des matrices Σ telles que

$$\Sigma^{1/2} = \Sigma_0^{1/2} + \epsilon,$$

avec $\|\epsilon\| < \delta$ pour δ fixé. Cette dernière équation peut encore s'écrire comme

$$\begin{aligned} \Sigma^{1/2} &= \Sigma_0^{1/2} \left(\mathbf{I}_p + \Sigma_0^{-1/2} \epsilon \right) \\ &= \Sigma_0^{1/2} \exp(\mathbf{B}) \end{aligned} \tag{2.3.3}$$

puisque $\mathbf{I}_p + \Sigma_0^{-1/2} \epsilon$ est une matrice triangulaire inférieure avec des éléments positifs sur la diagonale. On peut alors écrire l'inverse de chacune de ces matrices comme

$$\Sigma^{-1/2} = \exp(-\mathbf{B}) \Sigma_0^{-1/2}. \tag{2.3.4}$$

Servons-nous maintenant de cette nouvelle paramétrisation pour écrire le premier terme de la fonction Q définie en (2.3.2).

Lemme 2.3.1. *Si $\Sigma^{1/2} = \Sigma_0^{1/2} \exp(\mathbf{B})$, alors*

$$-n \log \left[\det \Sigma^{-1/2} \right] = n \text{trace}(\mathbf{B}) - n \log \left[\det \Sigma_0^{-1/2} \right].$$

Démonstration. Le résultat découle des propriétés du déterminant et des égalités (2.2.2) et (2.3.4). En effet,

$$\begin{aligned} -n \log \left[\det \Sigma^{-1/2} \right] &= -n \log \left[\det \left(\exp(-\mathbf{B}) \Sigma_0^{-1/2} \right) \right], \quad \text{par (2.3.4)} \\ &= -n \log \left[\det \exp(-\mathbf{B}) \right] - n \log \left[\det \Sigma_0^{-1/2} \right] \\ &= n \text{trace}(\mathbf{B}) - n \log \left[\det \Sigma_0^{-1/2} \right], \quad \text{par (2.2.2)}. \end{aligned}$$

□

Pour ce qui est du deuxième terme de la fonction Q , nous avons l'approximation donnée par le lemme suivant.

Lemme 2.3.2. *Soit ρ une fonction deux fois différentiable et posons $\psi = \rho'$. Pour $\Sigma^{1/2}$ dans un voisinage de $\Sigma_0^{1/2}$, on peut écrire*

$$\begin{aligned} \rho(\|\Sigma^{-1/2} \mathbf{X}_i\|) &= \rho(\|\mathbf{Z}_i\|) + \left[\frac{-\mathbf{Z}_i \mathbf{B} \mathbf{Z}_i}{\|\mathbf{Z}_i\|} + \frac{\mathbf{Z}_i' (\mathbf{B}^2 + \mathbf{B}' \mathbf{B}) \mathbf{Z}_i}{2\|\mathbf{Z}_i\|} - \frac{(\mathbf{Z}_i' \mathbf{B} \mathbf{Z}_i)^2}{2\|\mathbf{Z}_i\|^3} \right] \psi(\|\mathbf{Z}_i\|) \\ &\quad + \frac{(\mathbf{Z}_i' \mathbf{B} \mathbf{Z}_i)^2}{2\|\mathbf{Z}_i\|^2} \psi'(\|\mathbf{Z}_i\|) + o(\|\mathbf{B}^2\|_s^2), \end{aligned}$$

où $\mathbf{Z}_i = \boldsymbol{\Sigma}_0^{-1/2} \mathbf{X}_i$ et où $\|\mathbf{B}^2\|_s = \sup_{\|\mathbf{v}\|=1} \|\mathbf{B}^2 \mathbf{v}\|$ que l'on appelle la norme subordonnée de \mathbf{B}^2 .

Démonstration. Pour cerner le voisinage de $\boldsymbol{\Sigma}_0^{1/2}$, considérons comme négligeable les puissances de \mathbf{B} d'ordre supérieur à 2. Ceci nous permet d'écrire

$$\begin{aligned} \|\boldsymbol{\Sigma}^{-1/2} \mathbf{X}_i\| &= \|\exp(-\mathbf{B}) \boldsymbol{\Sigma}_0^{-1/2} \mathbf{X}_i\| = \|\exp(-\mathbf{B}) \mathbf{Z}_i\| \\ &\simeq \|(\mathbf{I}_p - \mathbf{B} + \frac{\mathbf{B}^2}{2}) \mathbf{Z}_i\| \\ &= \left[\mathbf{Z}_i' (\mathbf{I}_p - \mathbf{B} + \frac{\mathbf{B}^2}{2})' (\mathbf{I}_p - \mathbf{B} + \frac{\mathbf{B}^2}{2}) \mathbf{Z}_i \right]^{1/2} \\ &\simeq \left[\mathbf{Z}_i' \mathbf{Z}_i - \mathbf{Z}_i' (\mathbf{B}' + \mathbf{B}) \mathbf{Z}_i + \mathbf{Z}_i' \left(\frac{\mathbf{B}^{2'}}{2} + \mathbf{B}' \mathbf{B} + \frac{\mathbf{B}^2}{2} \right) \mathbf{Z}_i \right]^{1/2} \\ &= \|\mathbf{Z}_i\| \left[1 - \left(\frac{2\mathbf{Z}_i' \mathbf{B} \mathbf{Z}_i}{\|\mathbf{Z}_i\|^2} - \frac{\mathbf{Z}_i' (\mathbf{B}^2 + \mathbf{B}' \mathbf{B}) \mathbf{Z}_i}{\|\mathbf{Z}_i\|^2} \right) \right]^{1/2}. \end{aligned}$$

En utilisant maintenant le fait que $(1 - x)^{1/2} = 1 - x/2 - x^2/8 + o(x^2)$, nous avons l'approximation suivante:

$$\|\boldsymbol{\Sigma}^{-1/2} \mathbf{X}_i\| \simeq \|\mathbf{Z}_i\| \left[1 - \frac{\mathbf{Z}_i' \mathbf{B} \mathbf{Z}_i}{\|\mathbf{Z}_i\|^2} + \frac{\mathbf{Z}_i' (\mathbf{B}^2 + \mathbf{B}' \mathbf{B}) \mathbf{Z}_i}{2\|\mathbf{Z}_i\|^2} - \frac{(\mathbf{Z}_i' \mathbf{B} \mathbf{Z}_i)^2}{2\|\mathbf{Z}_i\|^4} \right],$$

où les puissances de \mathbf{B} d'ordre supérieur à 2 ont été négligées. Nous arrivons finalement au résultat en développant la fonction ρ autour du point $\|\mathbf{Z}_i\|$, i.e.

$$\begin{aligned} \rho(\|\boldsymbol{\Sigma}^{-1/2} \mathbf{X}_i\|) &\simeq \rho(\|\mathbf{Z}_i\|) + \left[\frac{-\mathbf{Z}_i' \mathbf{B} \mathbf{Z}_i}{\|\mathbf{Z}_i\|} + \frac{\mathbf{Z}_i' (\mathbf{B}^2 + \mathbf{B}' \mathbf{B}) \mathbf{Z}_i}{2\|\mathbf{Z}_i\|} - \frac{(\mathbf{Z}_i' \mathbf{B} \mathbf{Z}_i)^2}{2\|\mathbf{Z}_i\|^3} \right] \psi(\|\mathbf{Z}_i\|) \\ &\quad + \frac{(\mathbf{Z}_i' \mathbf{B} \mathbf{Z}_i)^2}{2\|\mathbf{Z}_i\|^2} \psi'(\|\mathbf{Z}_i\|). \end{aligned}$$

Tous les termes que nous avons négligés dans ces approximations sont $o(\|\mathbf{B}^2\|_s^2)$, i.e. que ces termes tendent plus vite vers 0 que $\|\mathbf{B}^2\|_s^2$. \square

En combinant les résultats de ces deux derniers lemmes et en ordonnant les termes selon les puissances de \mathbf{B} , on obtient la proposition suivante.

Proposition 2.3.3. *Sous les mêmes conditions que le lemme 2.3.2, la fonction Q définie en (2.3.2) peut être approchée par*

$$\begin{aligned} & Q(\boldsymbol{\Sigma}_0^{1/2}) - \sum_{i=1}^n \left\{ \frac{\psi(\|\mathbf{Z}_i\|)}{\|\mathbf{Z}_i\|} \mathbf{Z}_i' \mathbf{B} \mathbf{Z}_i - \text{trace}(\mathbf{B}) \right\} \\ & + \sum_{i=1}^n \left\{ \left(\frac{\mathbf{Z}_i' (\mathbf{B}^2 + \mathbf{B}' \mathbf{B}) \mathbf{Z}_i}{2\|\mathbf{Z}_i\|} - \frac{(\mathbf{Z}_i' \mathbf{B} \mathbf{Z}_i)^2}{2\|\mathbf{Z}_i\|^3} \right) \psi(\|\mathbf{Z}_i\|) + \frac{(\mathbf{Z}_i' \mathbf{B} \mathbf{Z}_i)^2}{2\|\mathbf{Z}_i\|^2} \psi'(\|\mathbf{Z}_i\|) \right\} \\ & + o(\|\mathbf{B}^2\|_s^2) \end{aligned} \quad (2.3.6)$$

avec $\mathbf{Z}_i = \boldsymbol{\Sigma}_0^{-1/2} \mathbf{X}_i$ et $\|\mathbf{B}^2\|_s = \sup_{\|\mathbf{v}\|=1} \|\mathbf{B}^2 \mathbf{v}\|$.

□

Cette proposition nous permet d'approcher la valeur prise par la fonction Q pour les matrices $\boldsymbol{\Sigma}$ au voisinage de $\boldsymbol{\Sigma}_0$. De plus, elle nous permet de reconnaître les matrices pour lesquelles la fonction Q admet un extremum. Ce sont les matrices $\boldsymbol{\Sigma}_0$ qui annulent le deuxième terme de (2.3.6), i.e. qui satisfont l'équation suivante:

$$\sum_{i=1}^n \left\{ \frac{\psi(\|\mathbf{Z}_i\|)}{\|\mathbf{Z}_i\|} \mathbf{Z}_i' \mathbf{B} \mathbf{Z}_i - \text{trace}(\mathbf{B}) \right\} = \text{trace} \left[\mathbf{B} \sum_{i=1}^n \left\{ \frac{\psi(\|\mathbf{Z}_i\|)}{\|\mathbf{Z}_i\|} \mathbf{Z}_i \mathbf{Z}_i' - \mathbf{I}_p \right\} \right] = 0. \quad (2.3.7)$$

Les statistiques $\boldsymbol{\Sigma}_0 = \hat{\boldsymbol{\Sigma}}(\mathbf{X}_1, \dots, \mathbf{X}_n)$ qui annulent cette dernière expression coïncident en fait avec les solutions de l'une ou l'autre des équations suivantes:

$$\frac{1}{n} \sum_{i=1}^n w(\|\hat{\boldsymbol{\Sigma}}^{-1/2} \mathbf{X}_i\|^2) \hat{\boldsymbol{\Sigma}}^{-1/2} \mathbf{X}_i \mathbf{X}_i' \hat{\boldsymbol{\Sigma}}^{-1/2} = \mathbf{I}_p, \quad (2.3.8)$$

ou encore

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^n w(\|\hat{\boldsymbol{\Sigma}}^{-1/2} \mathbf{X}_i\|^2) \mathbf{X}_i \mathbf{X}_i', \quad (2.3.9)$$

avec $w(x) = \psi(x)/x$. Dans ce travail, nous allons seulement considérer les cas où la fonction w est décroissante sur $\mathbb{R}^+ \cup \{0\}$. Ceci aura pour effet de diminuer l'importance des observations extrêmes sur le calcul de $\hat{\boldsymbol{\Sigma}}$. En fait, plus $\mathbf{X}_i' \boldsymbol{\Sigma}^{-1} \mathbf{X}_i$ sera grand, plus faible sera la contribution de l'observation \mathbf{X}_i dans le calcul de $\hat{\boldsymbol{\Sigma}}$.

Maronna (1976) et Huber (1981, p.212-213) ont défini de manière analogue l'estimateur robuste de $\boldsymbol{\Sigma}$. En fait, pour obtenir l'estimateur de Huber, il suffit de prendre

$$w_H(x) = \min(1, c/|x|), \quad (2.3.10)$$

pour une constante positive c quelconque. Certaines propriétés reliées à cet estimateur sont étudiées par Carroll (1978) et Tyler(1983, 1986).

Nous allons maintenant montrer que la matrice $\hat{\Sigma}$ définie en (2.3.9) correspond à un minimum local de la fonction Q .

Proposition 2.3.4. *Soit ψ une fonction différentiable telle que $x\psi(x)$ soit croissante, i.e. telle que $\forall x \in \mathbb{R}^+$, $\psi(x) + x\psi'(x) > 0$. Si $\hat{\Sigma}$ est solution de (2.3.9), alors*

$$Q(\Sigma) \geq Q(\hat{\Sigma}),$$

$\forall \Sigma$ dans un voisinage de $\hat{\Sigma}$.

Démonstration. Pour obtenir le résultat, il suffit de montrer que le troisième terme de (2.3.6) est positif. Or,

$$\begin{aligned} & \sum_{i=1}^n \left\{ \left(\frac{\mathbf{Z}'_i(\mathbf{B}^2 + \mathbf{B}'\mathbf{B})\mathbf{Z}_i}{2\|\mathbf{Z}_i\|} - \frac{(\mathbf{Z}'_i\mathbf{B}\mathbf{Z}_i)^2}{2\|\mathbf{Z}_i\|^3} \right) \psi(\|\mathbf{Z}_i\|) + \frac{(\mathbf{Z}'_i\mathbf{B}\mathbf{Z}_i)^2}{2\|\mathbf{Z}_i\|^2} \psi'(\|\mathbf{Z}_i\|) \right\} \\ &= \sum_{i=1}^n \left\{ \frac{\mathbf{Z}'_i(\mathbf{B}^2 + \mathbf{B}'\mathbf{B})\mathbf{Z}_i}{2\|\mathbf{Z}_i\|} - \frac{(\mathbf{Z}'_i\mathbf{B}\mathbf{Z}_i)^2}{\|\mathbf{Z}_i\|^3} \right\} \psi(\|\mathbf{Z}_i\|) \\ & \quad + \sum_{i=1}^n \frac{(\mathbf{Z}'_i\mathbf{B}\mathbf{Z}_i)^2}{2\|\mathbf{Z}_i\|^3} \{ \psi(\|\mathbf{Z}_i\|) + \|\mathbf{Z}_i\| \psi'(\|\mathbf{Z}_i\|) \} \\ & \geq \sum_{i=1}^n \left\{ \frac{\mathbf{Z}'_i(\mathbf{B}^2 + \mathbf{B}'\mathbf{B})\mathbf{Z}_i}{2\|\mathbf{Z}_i\|} - \frac{(\mathbf{Z}'_i\mathbf{B}\mathbf{Z}_i)^2}{\|\mathbf{Z}_i\|^3} \right\} \psi(\|\mathbf{Z}_i\|), \quad \text{car } \psi(x) + x\psi'(x) > 0 \\ &= \sum_{i=1}^n \left\{ \frac{\mathbf{Z}'_i(\frac{\mathbf{B}+\mathbf{B}'}{2})^2\mathbf{Z}_i}{\|\mathbf{Z}_i\|} - \frac{[\mathbf{Z}'_i(\frac{\mathbf{B}+\mathbf{B}'}{2})\mathbf{Z}_i]^2}{\|\mathbf{Z}_i\|^3} \right\} \psi(\|\mathbf{Z}_i\|) \\ & \geq \sum_{i=1}^n \left\{ \frac{\mathbf{Z}'_i(\frac{\mathbf{B}+\mathbf{B}'}{2})^2\mathbf{Z}_i}{\|\mathbf{Z}_i\|} - \frac{\mathbf{Z}'_i(\frac{\mathbf{B}+\mathbf{B}'}{2})^2\mathbf{Z}_i}{\|\mathbf{Z}_i\|} \right\} \psi(\|\mathbf{Z}_i\|) = 0 \quad \text{par Cauchy-Schwarz.} \end{aligned}$$

□

Pour vérifier que la matrice $\hat{\Sigma}$ est en fait un minimum global de la fonction Q , il faudrait montrer que cette fonction est constante sur la frontière du domaine de variation du paramètre multivarié Σ (Mäkeläinen, Schmidt et Styan 1981).

Il faudra cependant modifier légèrement la statistique $\hat{\Sigma}$ définie en (2.3.9) pour lui conférer de bonnes propriétés. Par exemple, si on divise le membre de droite de (2.3.9) par $CTE = E[Yw(Y)]/p$, où Y est une variable aléatoire χ_p^2 , et si on suppose que les observations \mathbf{X}_i sont gaussiennes, on peut montrer que la nouvelle statistique $\hat{\Sigma}$ converge en probabilité vers le paramètre Σ . Pour les fonctions w à l'étude, i.e. pour les fonctions w décroissantes sur $\mathbb{R}^+ \cup \{0\}$ avec $w(0) < \infty$, la constante CTE est toujours définie puisque

$$\frac{1}{p}E[Yw(Y)] \leq \frac{1}{p}E[Yw(0)] = w(0).$$

De plus, si on substitue le facteur $1/(n-1)$ à $1/n$ dans (2.3.9), cela permet de rendre l'estimateur $\hat{\Sigma}$ sans biais pour Σ lorsque $w(x) \equiv 1$.

En conclusion, lorsque le paramètre de localisation μ est fixé, l'estimateur robuste de Σ est donné par la statistique $\hat{\Sigma}(\mathbf{X}_1, \dots, \mathbf{X}_n)$ correspondant à la solution implicite de l'équation suivante:

$$\hat{\Sigma} = \frac{1}{(n-1)CTE} \sum_{i=1}^n w \left(\|\hat{\Sigma}^{-1/2}(\mathbf{X}_i - \mu)\|^2 \right) (\mathbf{X}_i - \mu)(\mathbf{X}_i - \mu)'. \quad (2.3.11)$$

II.3.2 Estimation robuste de μ avec Σ connu

Par un raisonnement analogue à celui de la section II.3.1, on peut montrer que la statistique $\hat{\mu}(\mathbf{X}_1, \dots, \mathbf{X}_n)$ qui est solution de l'une des deux équations suivantes

$$\sum_{i=1}^n w \left(\|\Sigma^{-1/2}(\mathbf{X}_i - \hat{\mu})\| \right) (\mathbf{X}_i - \hat{\mu}) = \mathbf{O}, \quad (2.3.12)$$

$$\hat{\mu} = \frac{\sum_{i=1}^n w \left(\|\Sigma^{-1/2}(\mathbf{X}_i - \hat{\mu})\| \right) \mathbf{X}_i}{\sum_{i=1}^n w \left(\|\Sigma^{-1/2}(\mathbf{X}_i - \hat{\mu})\| \right)}, \quad (2.3.13)$$

pour Σ fixé, correspond à un minimum local de la fonction Q définie en (2.3.1). Cette statistique est l'estimateur robuste de μ .

II.4 Algorithmes pour le calcul de $\hat{\mu}$ et $\hat{\Sigma}$

Dans cette section, nous allons présenter un algorithme mis au point par Huber (1981, p.238) pour résoudre simultanément les équations à partir desquelles $\hat{\mu}$ et $\hat{\Sigma}$ sont

définis. Ces équations sont respectivement

$$\hat{\boldsymbol{\mu}} = \frac{\sum_{i=1}^n w \left[\|\hat{\boldsymbol{\Sigma}}^{-1/2}(\mathbf{X}_i - \hat{\boldsymbol{\mu}})\| \right] \mathbf{X}_i}{\sum_{i=1}^n w \left[\|\hat{\boldsymbol{\Sigma}}^{-1/2}(\mathbf{X}_i - \hat{\boldsymbol{\mu}})\| \right]} \quad (2.4.1)$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{(n-1)CTE} \sum_{i=1}^n w \left[\|\hat{\boldsymbol{\Sigma}}^{-1/2}(\mathbf{X}_i - \hat{\boldsymbol{\mu}})\| \right] (\mathbf{X}_i - \hat{\boldsymbol{\mu}})(\mathbf{X}_i - \hat{\boldsymbol{\mu}})'$$

L'algorithme sera construit avec la fonction de poids de Huber définie en (2.3.10).

Définissons $\hat{\mathbf{V}} = (\hat{\boldsymbol{\Sigma}}^{1/2})^{-1} = \hat{\boldsymbol{\Sigma}}^{-1/2}$. Nous allons débiter l'algorithme avec certaines valeurs initiales \mathbf{t}_0 et \mathbf{V}_0 . Chaque itération sera composée de deux étapes: l'étape échelle et l'étape localisation. On dénotera par \mathbf{t}_j et \mathbf{V}_j les valeurs des paramètres obtenues à la $j^{\text{ième}}$ itération. Si le critère de convergence est obtenu après la $k^{\text{ième}}$ itération, nous prendrons alors $\hat{\boldsymbol{\mu}} = \mathbf{t}_k$ et $\hat{\boldsymbol{\Sigma}} = (\mathbf{V}_k^{-1})(\mathbf{V}_k^{-1})'$.

Valeurs de départ

Les valeurs de départ de l'algorithme sont les suivantes:

$$\mathbf{t}_0 = \bar{\mathbf{X}}, \quad \mathbf{V}_0 = \left[\frac{\sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'}{n-1} \right]^{-1/2}.$$

Elles correspondent aux estimateurs classiques pour $\boldsymbol{\mu}$ et $\boldsymbol{\Sigma}^{-1/2}$ dans la famille des lois normales.

Etape échelle

En partant de \mathbf{t}_j et \mathbf{V}_j , on calcule \mathbf{V}_{j+1} de la façon suivante: posons $\mathbf{Y}_i = \mathbf{V}_j(\mathbf{X}_i - \mathbf{t}_j)$ et calculons \mathbf{C}_{j+1} à partir de la relation

$$\mathbf{C}_{j+1} = \frac{1}{(n-1)CTE} \sum_{i=1}^n w_H(\mathbf{Y}_i' \mathbf{Y}_i) \mathbf{Y}_i \mathbf{Y}_i'$$

où w_H est définie en (2.3.10) et où $CTE = P[\chi_{p+2}^2 \leq c] + \frac{c}{p} P[\chi_p^2 > c]$. Nous obtenons finalement la valeur de \mathbf{V}_{j+1} en calculant $\mathbf{V}_{j+1} = \mathbf{C}_{j+1}^{-1/2} \mathbf{V}_j$.

Etape localisation

En partant de \mathbf{t}_j et \mathbf{V}_{j+1} , on calcule \mathbf{t}_{j+1} de la façon suivante: posons $\mathbf{Y}_i = \mathbf{V}_{j+1}(\mathbf{X}_i - \mathbf{t}_j)$ et calculons \mathbf{h}_{j+1} par la relation

$$\mathbf{h}_{j+1} = \frac{\sum_{i=1}^n w_H(\|\mathbf{Y}_i\|)(\mathbf{X}_i - \mathbf{t}_j)}{\sum_{i=1}^n w_H(\|\mathbf{Y}_i\|)}.$$

On obtient \mathbf{t}_{j+1} en calculant $\mathbf{t}_{j+1} = \mathbf{t}_j + \mathbf{h}_{j+1}$.

Critère de convergence

Le critère de convergence est donné par: "arrêter si la plus grande composante, en valeur absolue, de \mathbf{h}_j et de $\mathbf{C}_j - \mathbf{I}_p$ est inférieure à un ϵ prédéterminé" (en général $\epsilon = 10^{-3}$). C'est le critère proposé par Rivest et Plante (1988).

CHAPITRE III

Représentation biplot robuste

La présence d'observations extrêmes au sein d'une analyse statistique peut fortement influencer ses résultats. Ce problème se pose en particulier dans la représentation biplot d'un ensemble de données. Pour pallier ce problème, nous allons présenter une alternative robuste à la représentation biplot suggérée par Gabriel.

III.1 Introduction

Soit $\mathbf{Y} = (y_{i,j})$ une matrice $n \times p$ où l'élément $y_{i,j}$ est égal à la valeur de la $j^{\text{ième}}$ variable mesurée sur l'individu i . Nous allons associer à chaque individu et à chaque variable des poids déterminés au préalable lors d'une estimation robuste des paramètres de localisation ($\boldsymbol{\mu}$) et de dispersion ($\boldsymbol{\Sigma}$) de l'échantillon des individus. Si on note par $\{\mathbf{y}'_i\}_{i=1}^n$ les lignes du fichier de données, ces estimateurs sont définis implicitement comme étant la solution du système composé des équations (2.3.11) et (2.3.13), à savoir

$$\begin{aligned}\hat{\boldsymbol{\mu}} &= \frac{\sum_{i=1}^n m_i^{1/2} \mathbf{y}_i}{\sum_{i=1}^n m_i^{1/2}}, \\ \hat{\boldsymbol{\Sigma}} &= \frac{\sum_{i=1}^n m_i (\mathbf{y}_i - \hat{\boldsymbol{\mu}})(\mathbf{y}_i - \hat{\boldsymbol{\mu}})'}{(n-1)CTE},\end{aligned}\tag{3.1.1}$$

avec $m_i = w[(\mathbf{y}_i - \hat{\boldsymbol{\mu}})' \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{y}_i - \hat{\boldsymbol{\mu}})]$. L'argument de la fonction w est la distance de Mahalanobis entre l'observation \mathbf{y}_i et sa moyenne $\boldsymbol{\mu}$. En général, on choisira une fonction w décroissante afin de diminuer l'importance des observations pour lesquelles la distance de Mahalanobis est grande. On utilisera dans ce chapitre la fonction de poids de Huber définie par $w_H(x) = \min(1, c/|x|)$ avec $c = p - .5\sqrt{p}$. Le fait de choisir cette constante fait en sorte que lorsque $p = 1$, on obtient le *Proposal 2* de Huber avec constante de robustesse égale à environ 0.71. (Huber, 1964).

La constante CTE intervenant dans la définition de $\hat{\boldsymbol{\Sigma}}$ assure la convergence de cet estimateur dans le modèle paramétrique normal. En fait, cette constante, que nous avons déterminée à la section II.3, est égale à $CTE = E[Zw(Z)]/p$ où Z est une variable aléatoire suivant une loi χ^2 à p degrés de liberté. Pour l'estimateur de Huber, cette constante peut s'écrire de façon équivalente comme

$$CTE = \frac{E[Zw_H(Z)]}{p} = P[\chi_{p+2}^2 \leq c] + \frac{c}{p}P[\chi_p^2 \geq c].$$

L'algorithme présenté à la section II.4 est utilisé pour calculer l'estimation des paramètres multivariés $\boldsymbol{\mu}$ et $\boldsymbol{\Sigma}$. Une fois ce calcul effectué, nous attribuons à chacun des individus un poids

$$\omega_i = \frac{m_i}{(n-1)CTE}, \quad i = 1, \dots, n \quad (3.1.2)$$

et à chacune des variables, un poids ϕ_j correspondant à l'inverse de sa variance robuste, i.e.

$$\phi_j = 1/\hat{\boldsymbol{\Sigma}}_{j,j}, \quad j = 1, \dots, p, \quad (3.1.3)$$

où $\hat{\boldsymbol{\Sigma}}_{j,j}$ est l'élément (j, j) de la matrice $\hat{\boldsymbol{\Sigma}}$ définie en (3.1.1). Cette pondération a pour effet de diminuer l'importance relative des individus extrêmes et des variables pour lesquelles les observations ont une grande fluctuation.

La représentation biplot que nous allons étudier porte sur le tableau des données centrées, i.e. sur le tableau $\mathbf{X}_{n \times p}$ défini par

$$\mathbf{X} = \mathbf{Y} - \mathbf{1}\hat{\boldsymbol{\mu}}'. \quad (3.1.4)$$

Quant aux matrices,

$$\mathbf{M} = \text{diag}(\phi_1, \dots, \phi_p), \quad \text{et} \quad \mathbf{N} = \text{diag}(\omega_1, \dots, \omega_n), \quad (3.1.5)$$

elles serviront de métriques pour calculer les distances dans \mathbb{R}^p et \mathbb{R}^n respectivement.

III.2 Biplot robuste des individus

Ce type de biplot est construit de façon à mieux représenter les distances entre les individus. Une version non robuste de ce biplot est présentée à la section I.2.3. La version robuste diffère de la version classique par le choix de la métrique \mathbf{M} intervenant dans le calcul des distances et par une pondération différente associée à chaque individu. En fait, le poids associé à l'individu i dans l'analyse robuste est égal à $\omega_i = m_i / ((n - 1)CTE)$ alors qu'il est constant et égale à $1/(n - 1)$ dans l'analyse classique. L'avantage du biplot robuste est sa *résistance* aux observations extrêmes.

Les étapes dans la construction du biplot robuste sont les suivantes: il faut d'abord utiliser la décomposition en valeurs singulières du triplet $(\mathbf{X}, \mathbf{M}, \mathbf{N})$ telle que présentée en (1.1.3) pour écrire $\mathbf{X} = \mathbf{P}\mathbf{\Lambda}\mathbf{Q}' = \sum_{\alpha=1}^r \lambda_{\alpha} \mathbf{p}_{\alpha} \mathbf{q}'_{\alpha}$ avec $\mathbf{P}'\mathbf{N}\mathbf{P} = \mathbf{Q}'\mathbf{M}\mathbf{Q} = \mathbf{I}_r$. On utilise en second lieu la proposition 1.3.2 pour déterminer la matrice $\mathbf{X}_{(2)}$ de rang 2 qui approche le mieux le tableau de données \mathbf{X} . On obtient alors

$$\mathbf{X}_{(2)} = \sum_{\alpha=1}^2 \lambda_{\alpha} \mathbf{p}_{\alpha} \mathbf{q}'_{\alpha}. \quad (3.2.1)$$

On factorise ensuite cette matrice sous la forme $\mathbf{A}\mathbf{B}'$ définie en (1.2.1) de façon à ce que la contrainte (1.2.7) soit satisfaite, ce qui se fait en posant

$$\mathbf{A} = (\lambda_1 \mathbf{p}_1, \lambda_2 \mathbf{p}_2) \quad \text{et} \quad \mathbf{B} = (\mathbf{q}_1, \mathbf{q}_2). \quad (3.2.2)$$

Il reste finalement à représenter les lignes de ces deux matrices sur un graphique que l'on appelle *biplot robuste des individus*.

III.3 Biplot robuste des variables réduites

Cette représentation, contrairement à la précédente, est construite de façon à mieux représenter les distances entre les variables. La version classique de ce biplot est présentée dans la section I.2.4. Ces deux versions diffèrent fortement l'une de l'autre, les poids et les distances étant définis autrement. Dans l'analyse robuste, la matrice \mathbf{N} définie en (3.1.5) sert de métrique pour calculer les distances dans \mathbb{R}^n alors que les éléments ϕ_j de la matrice \mathbf{M} sont les poids associés aux variables.

Le biplot des variables réduites est réalisé en incorporant le poids des variables à leurs coordonnées sur le graphique. On s'en sert en général lorsque les variables ne sont pas mesurées dans les mêmes unités. Pour le construire, on se sert de la même décomposition en valeurs singulières que celle utilisée dans le biplot robuste des individus. Il en résulte que la matrice $\mathbf{X}_{(2)}$ définie en (3.2.1) est encore celle qui réalise le meilleur ajustement de \mathbf{X} . Toutefois, nous devons maintenant imposer à la factorisation de $\mathbf{X}_{(2)}$ la contrainte (1.2.10) au lieu de (1.2.7) de façon à ce que les distances entre les variables soient mieux représentées sur le biplot que celles entre les individus. Cette factorisation de $\mathbf{X}_{(2)}$ sous la forme \mathbf{AB}' s'effectue en posant

$$\mathbf{A} = (\mathbf{p}_1, \mathbf{p}_2) \quad \text{et} \quad \mathbf{B} = (\lambda_1 \mathbf{q}_1, \lambda_2 \mathbf{q}_2). \quad (3.3.1)$$

La $j^{\text{ième}}$ variable est identifiée sur le biplot par un point dont la coordonnée est $\phi_j^{1/2}(\lambda_1 q_{1,j}, \lambda_2 q_{2,j})$. Ces variables sont toutes situées à l'intérieur du cercle unité sur le biplot. En effet, une variable réduite est un vecteur de \mathbb{R}^n dont la norme dans la métrique \mathbf{N} est 1. Ainsi, toute projection de ce vecteur aura forcément une norme inférieure à 1.

III.4 Aides à l'interprétation

Plusieurs personnes adoptent un mouvement de recul devant les aides à l'interprétation, celles-ci étant beaucoup moins suggestives que les graphiques de l'analyse biplot. Pourtant l'usage des aides à l'interprétation est indispensable, une analyse biplot pouvant être erronée si elle est fondée seulement sur l'examen des graphiques. De plus, cet usage est facile pour peu que l'on en ait pris l'habitude. Une discussion des aides à l'interprétation pour les A.C.P est présentée par Volle (1981, p.89-92).

Pour interpréter un axe, il importe de savoir quels sont les points qui ont le plus contribué à sa position dans l'espace. C'est dans cet optique que nous allons introduire une mesure de la contribution individuelle à l'inertie expliquée par un axe. On définit la contribution relative du $i^{\text{ième}}$ individu à l'inertie expliquée par l'axe α par la quantité

$$CN_i = \frac{\omega_i (\mathbf{x}'_i \mathbf{M} \mathbf{q}_\alpha)^2}{\lambda_\alpha^2}, \quad i = 1, \dots, n. \quad (3.4.1)$$

Cette quantité joue un rôle essentiel dans l'interprétation du biplot car elle permet de classer les individus selon les rôles plus ou moins grands qu'ils ont joué dans la détermination des axes factoriels.

La quantité CN_i appartient toujours à l'intervalle $[0, 1]$ puisque

$$\lambda_\alpha^2 = \mathbf{q}'_\alpha \mathbf{M} \mathbf{X}' \mathbf{N} \mathbf{X} \mathbf{M} \mathbf{q}_\alpha = \sum_{i=1}^n \omega_i (\mathbf{x}'_i \mathbf{M} \mathbf{q}_\alpha)^2. \quad (3.4.2)$$

En fait, Rivest et Plante (1988) ont montré que la valeur maximale prise par CN_i est $c / ((n-1)CTE)$.

Il est à noter que $\mathbf{x}'_i \mathbf{M} \mathbf{q}_\alpha$ est en fait la coordonnée du $i^{\text{ième}}$ individu sur l'axe α . On peut donc dire que la contribution relative des individus à l'inertie expliquée par l'axe α est fonction à la fois de leurs coordonnées sur cet axe et de ω_i . Or, les poids associés aux individus ne sont pas représentés sur le biplot. Ainsi, un examen du graphique seul risque d'attirer notre attention sur des individus à fortes coordonnées sur l'axe α mais qui, en raison de leurs faibles poids, n'ont pas vraiment contribué à la direction de cet axe dans l'espace. Inversement, des individus influents dans l'analyse peuvent avoir contribué fortement à la position de l'axe factoriel sans pour autant se distinguer des autres sur le biplot.

Etablissons maintenant une mesure de la qualité de la représentation des individus sur les axes factoriels. Soit θ l'angle entre les vecteurs \mathbf{x}_i et \mathbf{q}_α ; on définit la qualité de la représentation du $i^{\text{ième}}$ individu sur le $\alpha^{\text{ième}}$ axe factoriel par le carré du cosinus de cet angle, que l'on note par $CO2_\alpha(i)$. En termes algébriques, cela revient tout simplement à

$$CO2_\alpha(i) = \frac{(\mathbf{x}'_i \mathbf{M} \mathbf{q}_\alpha)^2}{\|\mathbf{x}_i\|_{\mathbf{M}^2}}, \quad i = 1, \dots, n, \quad (3.4.3)$$

où $\|\mathbf{x}_i\|_{\mathbf{M}^2}^2 = \mathbf{x}'_i \mathbf{M} \mathbf{x}_i$. Plus cette mesure est grande, meilleure est la représentation de l'individu i . Cette mesure étant additive, on peut apprécier la qualité de la représentation du $i^{\text{ième}}$ individu sur le biplot par

$$QLT_2(i) = \sum_{\alpha=1}^2 CO2_\alpha(i), \quad i = 1, \dots, n. \quad (3.4.4.)$$

Il y a une symétrie dans le rôle joué par les matrices \mathbf{M} et \mathbf{N} dans les deux types de représentation biplot discutés dans ce chapitre. Nous pouvons donc utiliser

ces mêmes mesures pour interpréter le biplot des variables en interchangeant les deux matrices \mathbf{M} et \mathbf{N} dans les équations (3.4.1)-(3.4.4).

III.5 Exemple d'application et interprétation

Les représentations biplot robuste et classique donnent souvent des résultats semblables. C'est seulement en présence de données hétérogènes que ces deux représentations diffèrent. L'exemple que nous allons étudier utilise de telles données. En effet, comme le disent Rivest et Plante, il est plus intéressant de présenter un cas où apparaissent des différences entre les deux analyses car il illustre mieux les caractéristiques propres de chacune.

L'ensemble de données que nous allons analyser provient de World Statistics in Brief (1986). On y retrouve plusieurs variables socio-économiques mesurées sur 70 pays. Parmi celles-ci, il y a

- 1) la population totale mesurée en millions d'habitants (POP),
- 2) le nombre de mortalités infantiles par 1000 naissances (MORT),
- 3) le nombre de véhicules motorisés par 100 habitants (MOTOR),
- 4) le nombre de médecins par 1000 habitants (DOCTEUR),
- 5) le produit national brut par habitant mesuré en milliers de dollars U.S. (PNB),
- 6) la densité de la population par km^2 (DENS).
- 7) la fraction du total des transactions consacrée à l'exportation (EXPORT).

Le programme que nous avons utilisé pour analyser les données ainsi que les données elles-mêmes sont présentés en annexe.

L'itinéraire que nous allons suivre pour analyser cet exemple est le suivant: nous allons commencer par dégager les liens entre les diverses variables sur le biplot des variables réduites. Ceci nous amènera à donner une signification aux axes factoriels de cette représentation. Nous allons ensuite étudier le biplot des individus en se référant à l'interprétation des axes effectuée précédemment.

III.5.1 *Interprétation du biplot des variables réduites*

Avant de commencer l'interprétation proprement dite de ce biplot, rappelons d'abord que le cosinus de l'angle sous lequel on voit deux variables dans \mathbb{R}^n n'est autre que le coefficient de corrélation entre ces deux variables. Cette propriété est plus ou moins bien conservée sur le biplot. En fait, elle est d'autant mieux conservée que la qualité de l'ajustement est bonne. Dans de tels cas, le biplot des variables réduites constitue un moyen efficace pour représenter visuellement les corrélations entre les variables.

Dans cette section, nous allons étudier deux types de corrélations: les corrélations de l'analyse classique (au sens de Pearson) et les corrélations robustes (au sens de Huber). Ces mesures peuvent être lues directement sur le tableau 3.5.1 ou indirectement sur les biplot classique et robuste de la figure 3.5.1 et 3.5.2. Il est à noter que les corrélations robustes sont généralement plus grandes en valeur absolue que celles de l'analyse classique. Ainsi, en diminuant l'importance des individus éloignés du centre de gravité, l'analyse robuste raffermi la dépendance entre les variables.

Tableau 3.5.1 *Matrices de corrélations classiques et robustes**

	pop	mort	motor	docteur	pnb	dens	export
pop	1	.06 (.13)	-.10 (-.00)	.01 (-.05)	-.09 (-.04)	.03 (.15)	.03 (.13)
mort		1	-.68 (-.73)	-.75 (-.80)	-.54 (-.60)	-.14 (-.19)	-.21 (-.30)
motor			1	.63 (.76)	.86 (.89)	.02 (.08)	.21 (.28)
docteur				1	.48 (.59)	.02 (.16)	.18 (.22)
pnb					1	-.02 (.06)	.19 (.22)
dens						1	-.19 (-.09)
export							1

* *Les chiffres entre parenthèses sont les corrélations robustes de Huber*

Si on commence par regarder l'information présente le long du premier axe factoriel sur les graphiques 3.5.1 et 3.5.2, on constate que la position des variables MOTOR, PNB, DOCTEUR et MORT est sensiblement la même. Ceci laisse entendre que ces variables sont liées dans le même sens et qu'elles ont le même degré d'association dans les deux analyses. En fait, nous pouvons dire que les variables MOTOR, PNB et DOCTEUR sont très fortement corrélées entre elles puisque celles-ci sont situées très près l'une de l'autre. De plus, chacune de ces variables est liée à MORT dans le sens opposé puisque cette dernière variable est séparée des trois autres par un angle de près de 180° . Ceci nous permet de dire, entre autres, que le nombre de médecins dans un pays est inversement proportionnel au taux de mortalité infantile.

Si on consulte maintenant les aides à l'interprétation présentées au tableau 3.5.2, on remarque que les quatre variables MOTOR, PNB, DOCTEUR et MORT ont contribué presque à elles seules à la direction du premier axe factoriel dans les deux analyses. En fait, la part de chacune de ces variables est sensiblement la même, soit à peu près 25% chacune. Ceci dit, nous ne serons guère surpris de constater que ces variables sont les seules à être bien représentées sur cet axe. De plus, puisque ces variables mesurent en quelque sorte la richesse d'un pays, on peut alors conclure que cet axe factoriel oppose les pays riches aux pays pauvres.

Tableau 3.5.2 Aides à l'interprétation du biplot des variables réduites

	Biplot classique				Biplot robuste			
	$CTR_\alpha(j)$		$CO2_\alpha(j)$		$CTR_\alpha(j)$		$CO2_\alpha(j)$	
	$\alpha = 1$	$\alpha = 2$	$\alpha = 1$	$\alpha = 2$	$\alpha = 1$	$\alpha = 2$	$\alpha = 1$	$\alpha = 2$
pop	.00	.00	.01	.00	.00	.57	.00	.66
mort	.24	.02	.74	.02	.23	.00	.78	.00
motor	.28	.00	.84	.00	.27	.00	.88	.00
docteur	.22	.00	.67	.00	.23	.00	.76	.00
pnb	.23	.00	.69	.00	.22	.00	.72	.00
dens	.00	.61	.00	.71	.01	.42	.03	.48
export	.04	.37	.11	.43	.04	.00	.15	.01

A regarder la position relative des variables EXPORT et DENS sur la figure 3.5.1, on serait porté à croire que ces variables sont liées par une forte relation linéaire.

Figure 3.5.1 *Représentation biplot des variables réduites (analyse classique)*

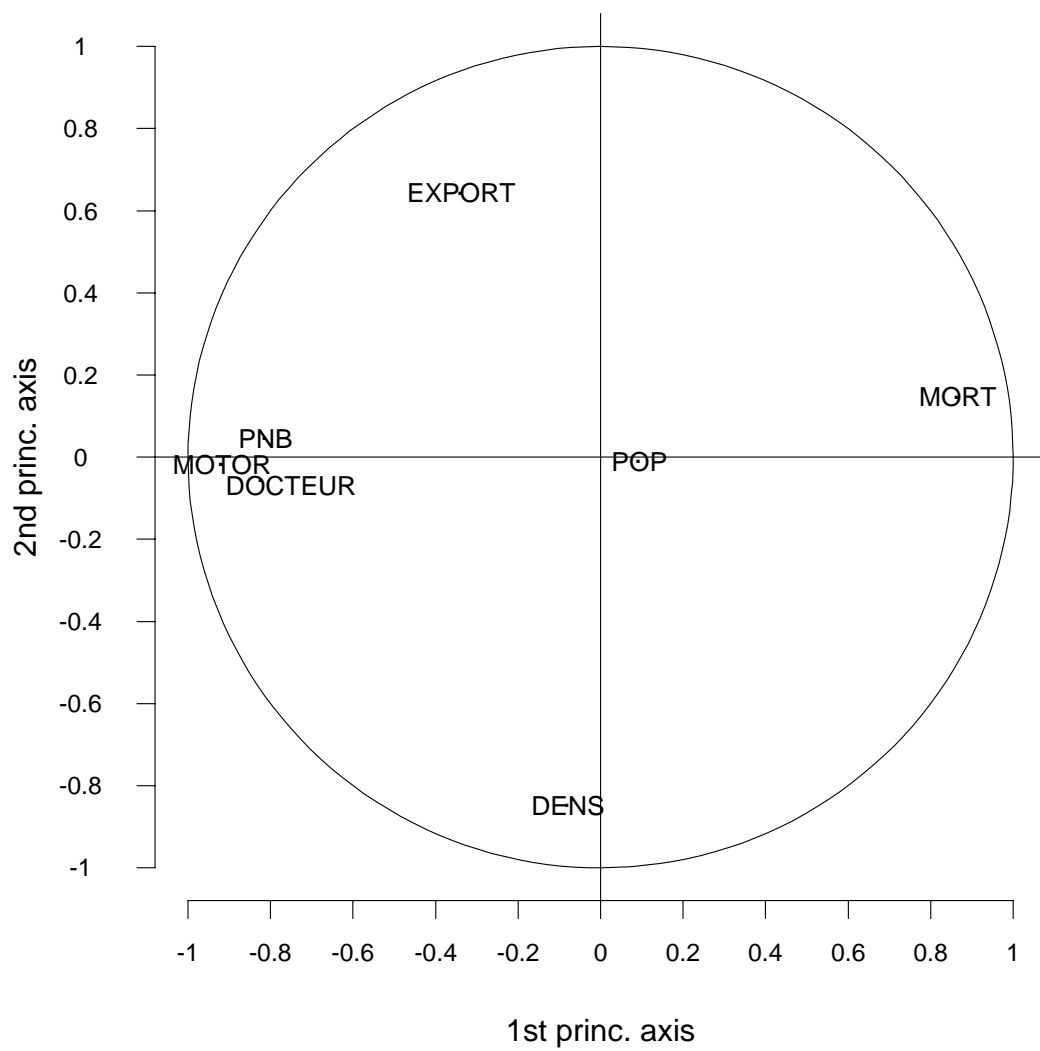
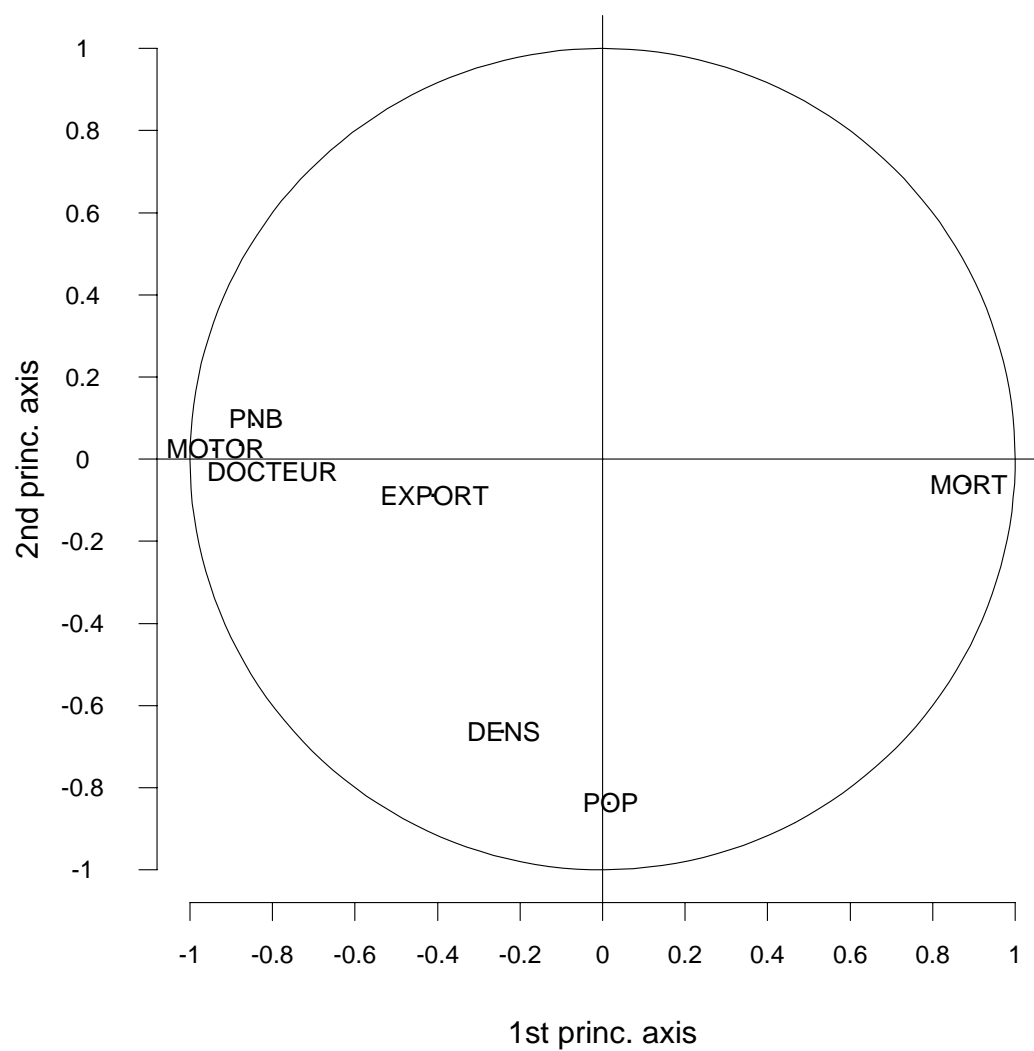


Figure 3.5.2 Représentation biplot des variables réduites (analyse robuste)



Or, cette relation n'est pas aussi forte qu'on le pense si on examine le tableau 3.5.1. En fait, la corrélation de Pearson entre ces deux variables est seulement de -0.19. Ceci est dû au fait que la variable EXPORT est mal représentée sur le biplot classique. En fait, elle y est représentée à seulement 54%, comme le montre le tableau 3.5.2. Mais il y a pire encore, la variable POP étant représentée seulement à 1%. Une situation semblable se produit dans l'analyse robuste. En effet, si on regarde maintenant la proximité entre les variables DENS et POP sur la figure 3.5.2, on est porté à croire que ces variables sont fortement liées entre elles. Or, la corrélation de Huber est seulement de .15. La raison de cette mauvaise interprétation est encore la même: seulement 51% de la variabilité de DENS est présente sur cette figure.

C'est dans la direction du deuxième axe factoriel que l'analyse classique diffère de l'analyse robuste. Les deux analyses diffèrent en fait au niveau de la contribution relative des variables à l'inertie expliquée par cet axe. Ces contributions sont présentées au tableau 3.5.2. On peut voir sur ce tableau que seulement deux des sept variables ont participé à la position du 2^{ième} axe factoriel dans les deux analyses. Il s'agit des variables DENS et EXPORT pour l'analyse classique et des variables DENS et POP pour l'analyse robuste. De plus, ces variables sont les seules à être bien représentées dans leurs analyses respectives. Nous en concluons donc que l'analyse robuste a favorisé la représentation de la variable POP au détriment de la variable EXPORT. Ceci s'explique par le fait que POP permet de mieux discriminer les pays, compte tenu de sa grande variabilité.

Il est à noter toutefois qu'aucune conclusion ne peut être faite à l'égard de la variable POP dans l'analyse biplot classique, celle-ci étant mal représentée sur la figure 3.5.1. Il en est de même pour la variable EXPORT sur la figure 3.5.2.

III.5.2 *Interprétation du biplot des individus*

Les représentations biplot classique et robuste sont illustrées respectivement à la figure 3.5.3 et 3.5.4. Nous y avons représenté les variables sur une échelle différente, celles-ci étant à la base trop loin de l'origine. Nous les avons représentées en fait sur un cercle de rayon 3 mais en agissant de la sorte, on perd toute l'information à l'égard de la proximité des variables. Toutefois, cette information n'est pas pertinente puisque

la *métrique* induite dans le calcul des distances (voir (1.2.9)) est vide de sens. Même si cette information est perdue, elle ne nous empêche pas pour autant de reconstituer le tableau de données en lui-même puisque la direction des vecteurs-variables est sauvegardée.

Les représentations biplot des deux analyses apportent sensiblement la même quantité d'information. Nous avons en fait 60.4% de l'inertie totale qui est représentée sur le biplot classique comparativement à 64% sur l'analyse robuste. Cependant, cette information est différente comme nous l'avons remarqué à la section III.5.1.

Commençons par interpréter la représentation biplot de la figure 3.5.3. Les pays qui sont bien représentés sur l'axe factoriel 1 sont soit très riches, soit très pauvres selon qu'ils sont du côté négatif ou positif de l'axe. Par exemple, les pays de l'Amérique du Nord, de l'Australie, de l'Europe du Nord et de l'ouest (à l'exception de Malte et du Portugal) sont assez riches. En effet, ce sont des pays qui sont caractérisés par un produit national brut assez élevé, par un grand nombre de véhicules motorisés et de médecins par habitant, et par un faible taux de mortalité infantile. Ces constatations viennent du fait que ces pays sont relativement près les uns des autres et qu'ils sont tous du côté négatif de l'axe 1. Les pays pauvres sont représentés sur le côté positif de cet axe. Ce sont des pays comme Haïti, l'Afghanistan, la Chine, l'Inde et la majorité des pays d'Afrique. Ces pays possèdent un taux de mortalité infantile assez élevé puisque ses habitants n'ont pas la même facilité à se déplacer d'un endroit à un autre et à se faire soigner. Le produit national brut par habitant de chacun de ces pays est également assez faible. Ces conclusions sont les mêmes que celles déduites de la représentation biplot robuste illustrée à la figure 3.5.4.

Etudions maintenant l'information présente le long du 2^{ième} axe factoriel sur la figure 3.5.3. Les pays qui sont bien représentés sur cet axe sont soit des pays à très forte densité, soit des pays qui exportent beaucoup. On les reconnaît à leur position sur le côté négatif ou positif de l'axe. Par exemple, l'Arabie Séoudite et la majorité des pays de l'Amérique du Sud et de l'Amérique central sont des pays dont le taux d'exportation est relativement élevé. A l'opposé, les pays comme le Bangladesh, la Barbade et Malte sont des pays dont la densité de la population par km^2 est très forte, ces pays étant situés du côté négatif de l'axe.

Cependant, cette représentation ne nous permet pas de distinguer les pays très peuplés de ceux qui le sont moins. Il faut regarder la représentation biplot robuste (figure 3.5.4) pour le savoir. Par exemple, le fait que le Canada et les Etats-Unis soient près l'un de l'autre sur le biplot classique et qu'ils soient relativement éloignés sur le biplot robuste indique que ces pays ont des caractéristiques similaires compte tenu des variables à l'étude, exception faite pour la population. Les Etats-Unis sont en fait plus peuplés que le Canada. On reconnaît les pays à forte population sur cette figure par ceux qui ont une grande composante négative sur l'axe 2. On retrouve cette caractéristique pour les pays comme le Japon, les Etats-Unis, l'U.R.S.S, la Chine, l'Inde et le Bangladesh.

Figure 3.5.3: Representation biplot des individus (analyse classique)

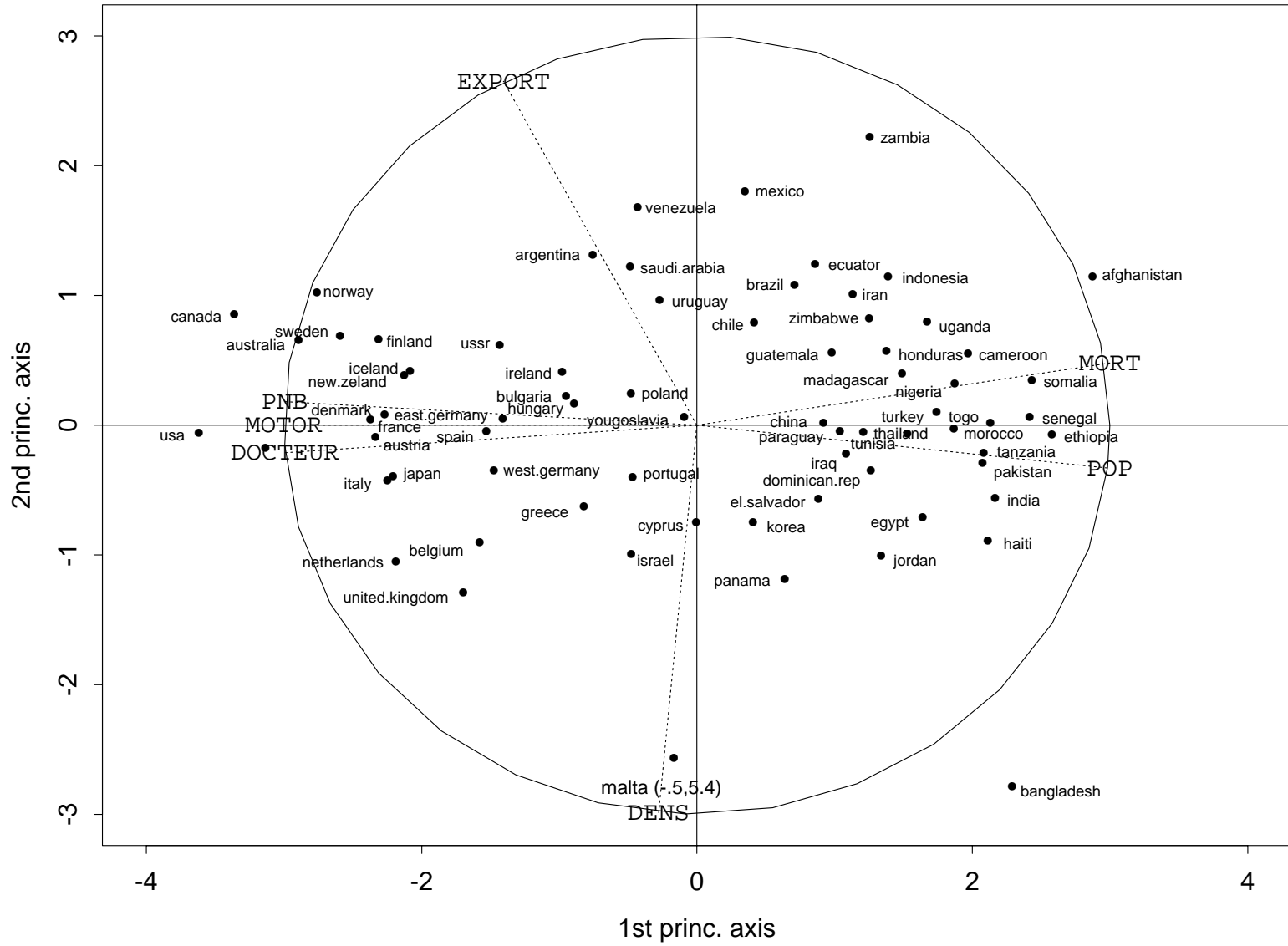
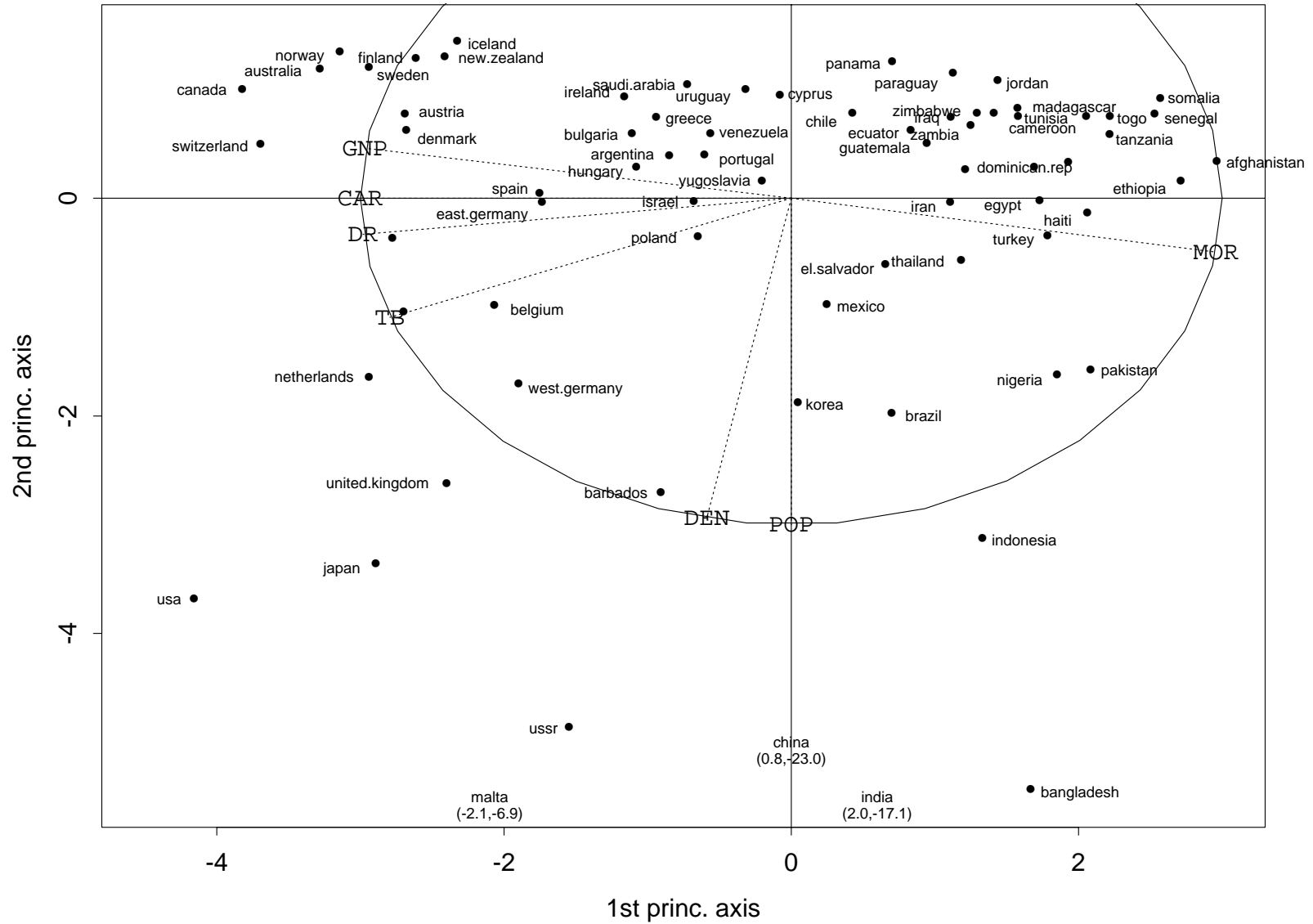


Figure 3.5.4: Representation biplot des individus (analyse robuste)



CONCLUSION

L'analyse biplot a permis de dégager l'essentiel de l'information contenue dans un tableau de données et d'en fournir une représentation imagée se prêtant aisément à l'interprétation. En ayant synthétisé cette information, le biplot a permis de repérer assez rapidement certaines particularités dans les relations inter-individus et inter-variables. Le biplot a été construit selon qu'on veuille mieux représenter les distances entre les individus ou celles entre les variables. Nous avons interprété différemment chacune des représentations biplot qui en découlaient.

Nous avons présenté l'analyse biplot robuste comme un complément de l'analyse classique. Son utilisation en présence de données hétérogènes a permis de mettre en lumière certaines caractéristiques que l'analyse classique passe sous silence. Nous l'avons constaté à partir d'un exemple.

On suggère de faire l'analyse robuste même en présence de données homogènes car elle permet de valider les résultats d'une analyse biplot ordinaire. En effet, si les résultats de ces deux analyses sont identiques, la crédibilité des conclusions s'en trouve alors augmentée.

BIBLIOGRAPHIE

- [1] Anderson, T.W., *An Introduction to Multivariate Statistical Analysis (2^{ième} édition)*, Wiley, New York, 1984,, pp. 63.
- [2] Bradu, D., Gabriel, K.R., *The Biplot as a Diagnostic Tools for Models of Two-way Tables*, Technometrics, **20** (1978), no. no 1, 47–67.
- [3] Carroll, R.J., *On the Asymptotic Distribution of Multivariate M-estimates*, J. Multivar. Anal. **8** (1978), 361–371.
- [4] Corsten, L.C.A., Gabriel, K.R., *Graphical Exploration in Comparing Variance Matrices*, Biometrics **32** (1976), 851–863.
- [5] Dwyer, P.S., *Some Applications of Matrix Derivatives in Multivariate Analysis*, J. Am. Statis. Assoc. **62**, (1967), 607–625.
- [6] Gabriel, K.R., *Least Squares Approximations of Matrices by Additive and Multiplicative Models*, J. R. Statis. Soc. B **40** (1978), no. no 2, 182–196.
- [7] Gabriel, K.R., *The Biplot Graphic Display of Matrices with Application to Principal Component Analysis*, Biometrika **58** (1971), no. no 3, 453–467.
- [8] Good, I.J., *Some Applications of the Singular Decomposition of a Matrix*, Technometrics **11** (1969), no. no 4, 823–831.
- [9] Greenacre, M.J., *Theory and Applications of Correspondence Analysis*, Academic Press, London, 1984.
- [10] Hampel, F.R., Rousseeuw, P.J., Ronchetti, E.M., Stahel, W.A., *Robust statistics: The Approach Based on Influence Functions*, Wiley, New York, 1986.
- [11] Huber, P.J., *Robust Estimation of a Location Parameter*, Annals Math. Statist. **35** (1964), 73–101.

- [12] Huber, P.J., *Robust Statistics*, Wiley, New York, 1981.
- [13] Householder, A.S., Young, G., *Matrix Approximations and Latent Roots*, Am. Math. Monthly **45** (1938), 165–171.
- [14] Kunitz, S.J., Human biol. **48** (1976), 361–377.
- [15] Lachenbruch, P.A., Sneeringer, C., Revo, L.T., *Robustness of the Linear and Quadratic Discriminant Function to Certain Types of Nonnormality*, Comm. Stat. **1** (1973), 39–56.
- [16] Lancaster, P., Tismenetsky, M., *The theory of Matrices with Applications (2^{ième} édition)*, Academic Press, New York, 1985.
- [17] Lebart, L., Morineau, A., Tabard, N., *Techniques de la Description Statistique*, Dunod, Paris, 1977.
- [18] Mäkeläinen, T., Schmidt, K., Styan, G.P.H., *On the Existence and Uniqueness of the Maximum Likelihood Estimate of a Vector-valued Parameter in Fixed-size Samples*, The Annals of Statistics **9** (1981), no. no 4, 758—767.
- [19] Maronna, R.A., *Robust M-estimators of Multivariate Location and Scatter*, the Annals of Statistics **4** (1976), no. no 1, 51–67.
- [20] Pullman, N.J., *Matrix Theory and its Applications Selected Topics*, Marcel Dekker Inc., New York, 1976.
- [21] Rao, C.R., *Linear Statistical Inference and its Applications (2^{ième} édition)*, Wiley, New York, 1973.
- [22] Rivest, L.P., Plante N., *L'analyse en Composantes Principales Robuste*, Rev, Statistique appliquée **36** (1988), no. no 1, 55-66.
- [23] Stigler, S.M., *Simon Newcomb, Percy Daniell, and the History of Robust Estimation*, J. Am. Statist. Assoc. **68** (1973), 872–879.
- [24] Tyler, D.E., *Robustness and Efficiency Properties of Scatter Matrices*, Biometrika **70** (1983), 411–420.
- [25] Volle, M., *Analyse des Données (2^{ième} édition)*, Economica, Paris, 1981.
- [26] Warner, F.W., *Foundations of Differentiable Manifolds and Lie Groups*, Springer-Verlag, New York, 1971.
- [27], *World Statistics in Brief, number 10*, United Nations, New York, 1986.

ANNEXE A

Données analysées au chapitre III

PAYS	POP	MORT	EXPORT	MOTOR	DOCTEUR	PNB	DENS
Canada	25.13	9	0.54	57.50	1.82	12.03	2.52
U.S.A	236.63	11	0.40	70.00	1.82	13.84	25.21
Barbados	0.25	23	0.37	9.80	0.86	4.22	625.00
Dominic.Rep	6.10	64	0.38	1.80	0.25	1.41	127.08
El Salvador	5.39	44	0.45	3.20	0.37	0.72	256.67
Guatemala	7.60	66	0.51	2.70	0.46	1.19	69.72
Haiti	5.19	108	0.30	0.30	0.12	0.19	185.36
Honduras	4.23	82	0.47	1.50	0.32	0.70	37.77
Mexico	76.79	53	0.72	9.10	0.37	1.16	38.98
Panama	2.13	26	0.18	7.80	0.96	2.06	28.40
Argentina	30.10	36	0.64	16.50	2.38	0.13	10.83
Brazil	132.58	71	0.57	8.80	0.88	0.29	15.58
Chile	11.85	24	0.53	7.50	0.62	0.06	15.97
Ecuador	9.12	77	0.60	4.10	0.62	0.91	33.65
Paraguay	3.28	45	0.36	1.20	0.57	1.04	8.06
Uruguay	2.99	38	0.57	9.90	1.87	0.85	15.99
Venezuela	16.85	39	0.69	15.30	1.06	2.30	18.48
Austria	7.55	12	0.44	34.50	2.78	7.24	89.88
Belgium	9.88	11	0.48	36.60	2.70	0.01	318.71
France	54.94	9	0.47	43.30	2.08	7.47	99.53
Italia	56.98	14	0.47	39.00	2.89	4.89	189.30
Malta	0.38	15	0.35	20.80	1.15	2.45	1266.67
Netherlands	14.42	8	0.51	35.0	1.85	7.39	424.12
Portugal	10.16	26	0.40	14.30	2.19	1.33	110.43
Spain	38.72	10	0.45	26.90	2.56	3.38	76.37
Sweden	6.44	8	0.47	42.00	2.44	12.25	153.33
Danemark	5.11	8	0.49	32.00	2.07	8.96	118.84
Finland	4.88	6	0.52	32.90	2.08	8.63	14.48
Iceland	0.24	7	0.47	39.50	2.14	5.76	2.33
Ireland	3.54	10	0.50	22.70	1.29	4.10	50.57

PAYS	POP	MORT	EXPORT	MOTOR	DOCTEUR	PNB	DENS
Norway	4.14	8	0.58	32.50	2.03	11.87	12.78
Sweden	8.34	7	0.53	38.70	2.09	9.40	18.53
Bulgaria	8.96	17	0.50	10.70	2.54	2.66	80.72
Cyprus	0.65	17	0.30	16.30	0.94	2.64	72.22
Greece	9.90	16	0.34	16.40	2.54	2.41	74.44
Hungary	10.67	19	0.51	13.20	2.56	1.35	114.73
Poland	36.91	19	0.52	10.70	1.85	1.32	118.30
U.R.S.S.	274.00	26	0.53	7.00	4.15	2.44	12.23
Yougoslavia	23.02	30	0.46	13.30	1.49	0.83	89.92
Cameroon	9.47	117	0.44	0.50	0.07	0.70	19.98
Egypt	45.66	113	0.24	1.80	1.32	0.77	45.66
Ethiopia	35.42	143	0.32	0.13	0.01	0.13	28.63
Madagascar	9.73	67	0.43	0.70	0.10	0.19	16.58
Morocco	22.85	99	0.36	3.00	0.06	0.44	51.12
Nigeria	92.04	114	0.46	1.40	0.10	0.65	99.61
Senegal	6.35	141	0.34	0.90	0.08	0.32	32.23
Somalia	5.42	143	0.38	0.40	0.06	0.11	8.50
Togo	2.84	113	0.36	0.08	0.05	0.21	49.82
Tunisia	7.04	85	0.36	3.90	0.27	0.90	45.42
Ouganda	15.15	94	0.53	0.08	0.04	0.00	62.35
Tanzania	21.06	98	0.31	0.30	0.04	0.17	22.43
Zambia	6.45	101	0.74	1.20	0.14	0.29	8.65
Zimbabwe	7.98	70	0.51	3.10	0.16	0.42	20.46
Australia	15.54	10	0.51	53.50	1.79	9.94	2.02
New Zealand	3.23	12	0.47	52.90	1.65	5.17	11.96
Iran	43.41	101	0.53	4.10	0.39	2.64	26.34
Iraq	15.16	72	0.33	2.60	0.56	3.06	34.93
Israel	4.19	15	0.37	7.70	2.66	0.57	199.52
Jordan	3.38	63	0.20	3.70	0.76	1.09	35.21
Saudi.Arabia	10.82	103	0.54	20.20	0.60	10.71	6.18

PAYS	POP	MORT	EXPORT	MOTOR	DOCTEUR	PNB	DENS
Turkey	48.27	110	0.40	1.80	0.65	0.00	61.88
Afghanistan	17.67	205	0.50	0.20	0.07	0.08	27.18
Bangladesh	96.73	133	0.31	0.05	0.14	0.13	676.43
China	1051.55	38	0.47	0.09	1.27	0.16	107.52
India	746.74	118	0.40	0.30	0.39	0.21	228.50
Indonesia	159.90	87	0.61	1.10	0.09	0.41	84.16
Japon	120.02	7	0.56	36.00	1.36	9.12	325.26
Pakistan	89.73	120	0.35	0.50	0.34	0.30	111.60
Korea	40.58	29	0.49	1.90	0.69	1.78	335.37
Thailand	50.40	51	0.42	2.20	0.15	0.68	98.05

ANNEXE B

Programme SAS utilisé pour faire une analyse biplot

```

PROC MATRIX;
fetch INIT DATA=lecture;
n=nrow(INIT);
k=ncol(INIT);

/*ANALYSE BIPLLOT CLASSIQUE*/

/*Quelques statistiques descriptives*/
moy=INIT(.,);
note moyennes arithmétiques des variables;
print moy;

V=inv((half(((INIT**INIT)-(n#moy**moy))#/(n-1)))));
COVAR=inv(V)*(inv(V));
note matrice de variances-covariances;
print COVAR;

var=vecdiag(COVAR);
CORREL=diag(var###(-.5))*COVAR*diag(var###(-.5));
note skip=5 matrice de corrélations;
print CORREL;

/*Décomposition en valeurs singulières DVS(X, M_I, M_J)*/
CENTR=INIT-J.(n,1,1)*moy;
MI=J.(n,1,1#/(n-1));
MJ=var###(-1);
Y=diag(MI###(0.5))*CENTR*diag(MJ###(0.5));
svd P lambda Q Y;
r=k+1-rank(lambda);
TEMP=P; P(,r)=TEMP;
TEMP=lambda; lambda(r,)=TEMP;
TEMP=Q; Q(,r)=TEMP;
P=diag(MI###(-0.5))*P;
Q=diag(MJ###(-0.5))*Q;

/*résultats concernant les valeurs propres*/
lambda2=lambda##2;
prop2=lambda2#/lambda2(+,);
cumul2=J.(k,1,prop2(1,1));
do j=1 to k-1;
    cumul2(j+1,1)=cumul2(j,1)+prop2(j+1,1);
end;
MATVAL=lambda?/?/prop2?/?/cumul2;
note page qualité de l'ajustement du biplot;
ajust="lambda" "prop2" "cumul2";
print MATVAL colname=ajust;

/*Biplot des variables réduites*/
VARIABLE=diag(MJ###(0.5))*Q*diag(lambda);
note page coordonnées des variables sur le biplot;
print VARIABLE;

```

```

/*aides à l'interprétation*/
CO2J=VARIABLE##2;
CTRJ=CO2J#/(J.(k,1,1)*lambda2');
note page contribution des variables à l'inertie expliquée par les axes;
print CTRJ;
note qualité de la représentation des variables sur les axes factoriels;
print CO2J;
QLTJ=CO2J(,1)+CO2J(,2);
note qualité de la représentation des variables sur le biplot;
print QLTJ;

/*biplot des individus*/
INDI=P*diag(lambda);
note coordonnées des individus sur le biplot;
print INDI;
note coordonnées des variables sur le biplot;
print Q;

/*aides à l'interprétation*/
INDI2=INDI##2;
CTRI=(MI*J.(1,K,1))#INDI2#/(J.(N,1,1)*lambda2');
note page contribution des individus à l'inertie expliquée par les axes;
print CTRI;
CO2I=INDI2#/(INDI2(+)*J.(1,k,1));
note qualité de la représentation des individus sur les axes factoriels;
print CO2I;
QLTI=CO2I(,1)+CO2I(,2);
note qualité de la représentation des individus sur le biplot;
print QLTI;

/*sortie des résultats dans un data set*/
output VARIABLE out=biplot1(rename=(col1=comp1 col2=comp2));
output INDI out=biplot3(rename=(col1=comp1 col2=comp2));

/*ESTIMATION ROBUSTE */

c=k-.5#sqrt(k);
ctel=probchi(c,k+2)+(1-probchi(c,k))#c#/k;
note page constante utilisée dans l'estimation robuste;
print ctel;
cte=(n-1)#ctel;
maxiter=300 /* nombre max. d'itérations toléré */;
wi2=J.(n,1,1);

do iter=1 to maxiter;

/* étape localisation*/

/* données centrées réduites */
CENTR=INIT-J.(n,1,1)*moy;
CENTRED=CENTR*V';

```

```

/* détermination des poids des individus */
do i=1 to n;
    dist=CENTRED(i)*CENTRED(i)';
    if dist<=c then wi2(i,1)=1;
    else wi2(i,1)=c#/dist;
end;

/* estimation du vecteur des moyennes */
h=(sqrt(wi2)*CENTR)#/sum(sqrt(wi2));
moy=moy+h;

/* étape échelle */

/* données centrées réduites */
CENTR=INIT-J.(n,1,1)*moy;
CENTRED=CENTR*V';

/* détermination des poids des individus */
do i=1 to n;
    dist=CENTRED(i)*CENTRED(i)';
    if dist<=c then wi2(i,1)=1;
    else wi2(i,1)=c#/dist;
end;

/* estimation de la matrice de covariances */
wi2centr=(wi2*J.(1,k,1))#CENTRED;
W=inv((half(CENTRED'*wi2centr#/cte)'));
V=W*V';

/* critères de convergence */

critère1=max(abs(W-I.(k)));
critère2=max(abs(V*h'));
if critère1<0.001 and critère2<0.001 then goto conv;

end;

conv:note nombre d"itérations effectuées;
print iter;
if iter>maxiter then do;
    note le nombre d"itérations permis n"as pas assuré la convergence;
    stop;
end;

/* résultats de l'estimation robuste */
note page moyenne robuste des variables;
print moy;

COVAR=inv(V)*(inv(V)');
note skip=5 matrice de variances-covariances robuste;
print COVAR;

```

```

var=vecdiag(COVAR);
CORREL=diag(var##(-.5))*COVAR*diag(var##(-.5));
note skip=5 matrice de corrélations robuste;
print CORREL;

/* ANALYSE BIPLLOT ROBUSTE */

/* décomposition en valeurs singulières DVS(X, MI, MJ) */
MI=wi2#/cte;
note page poids attribués aux INDIVIDUS;
print MI;
MJ=var##(-1);
Y=diag(MI##(0.5))*CENTR*diag(MJ##(0.5));
svd P lambda Q Y;
r=k+1-rank(lambda);
TEMP=P; P(,r)=TEMP;
TEMP=lambda; lambda(r,)=TEMP;
TEMP=Q; Q(,r)=TEMP;
P=DIAG(MI##(-.5))*P;
Q=DIAG(MJ##(-.5))*Q;

/* résultats concernant les valeurs propres */
lambda2=lambda##2;
prop2=lambda2#/lambda2(+,);
cumul2=J.(k,1,prop2(1,1));
do j=1 to k-1;
  cumul2(j+1,1)=cumul2(j,1)+prop2(j+1,1);
end;
MATVAL=lambda?/?/prop2?/?/cumul2;
note page qualité de l'ajustement du biplot robuste;
ajust="lambda" "prop2" "CUMUL2";
print MATVAL colname=ajust;

/*biplot des variables réduites*/
VARIABLE=diag(MJ##(0.5))*Q*diag(lambda);
note page coordonnées des variables sur le biplot robuste;
print VARIABLE;

/*aides à l'interprétation */
CO2J=VARIABLE##2;
CTRJ=CO2J#/(J.(k,1,1)*lambda2');
note page contribution des variables à l'inertie expliquée par les axes;
print CTRJ;
note qualité de la représentation des variables sur les axes factoriels;
print CO2J;
QLTJ=CO2J(,1)+CO2J(,2);
note qualité de la représentation des variables sur le biplot;
print QLTJ;

```

```

/* biplot des individus */
INDI=P*diag(lambda);
note coordonnées des individus sur le biplot;
print INDI;
note coordonnées des variables sur le biplot;
print Q;

  /* aides à l'interprétation */
  INDI2=INDI##2;
  CTRI=(MI*J.(1,k,1))#INDI2#/(J.(n,1,1)*lambda2');
  note page contribution des individus à l'inertie expliquée par les axes;
  print CTRI;
  CO2I=INDI2#/(INDI2(+)*J.(1,k,1));
  note qualité de la représentation des individus sur les axes factoriels;
  print CO2I;
  QLTI=CO2I(,1)+CO2I(,2);
  note qualité de la représentation des individus sur le biplot;
  print QLTI;

/* sortie des résultats dans un data set */
output VARIABLE out=biplot2(rename=(col1=comp1 col2=comp2));
output INDI out=biplot4(rename=(col1=comp1 col2=comp2));

```

/* REPRÉSENTATION GRAPHIQUE */

Biplot classique des variables réduites

```

PROC PLOT DATA=biplot1;
plot comp2*comp1='*/href=0 vref=0 haxis=-1 to 1 by .1
                                vaxis=-1 to 1 by .1 hpos=45 vpos=45;

```

Biplot robuste des variables réduites

```

PROC PLOT=DATA=biplot2
plot comp2*comp1='*/href=0 vref=0 haxis=-1 to 1 by .1
                                vaxis=-1 to 1 by .1 hpos=45 vpos=45;

```

Biplot classique des individus

```

PROC PLOT=DATA biplot3;
plot comp2*comp1='*/href=0 vref=0;

```

Biplot robuste des individus

```

PROC PLOT DATA=biplot4;
plot comp2*comp1='*/href=0 vref=0;

```