

# REVUE DE STATISTIQUE APPLIQUÉE

AZIZ LAZRAQ

ROBERT CLÉROUX

## **Inférence robuste sur un indice de redondance**

*Revue de statistique appliquée*, tome 50, n° 4 (2002), p. 39-54

[http://www.numdam.org/item?id=RSA\\_2002\\_\\_50\\_4\\_39\\_0](http://www.numdam.org/item?id=RSA_2002__50_4_39_0)

© Société française de statistique, 2002, tous droits réservés.

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

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

## INFÉRENCE ROBUSTE SUR UN INDICE DE REDONDANCE <sup>1</sup>

Aziz LAZRAQ<sup>(1)</sup>, Robert CLÉROUX<sup>(2)</sup>

<sup>(1)</sup> *École Nationale de l'Industrie Minérale, Rabat,*  
*courriel : lazraq@enim.ac.ma*

<sup>(2)</sup> *Département de mathématiques et de statistique, Université de Montréal,*  
*courriel : cleroux@dms.umontreal.ca*

### RÉSUMÉ

Dans cet article on obtient l'estimation  $S$ -robuste de l'indice de redondance  $\rho I$  de Stewart et Love (1968) et on construit un test asymptotique robuste pour  $H_0 : \rho I = 0$  contre  $H_1 : \rho I > 0$  lorsque la loi sous-jacente est dans la classe des lois elliptiques. On étudie le point de rupture de l'estimateur. Sa fonction d'influence est aussi obtenue de même que la puissance asymptotique du test. Des exemples et des simulations illustrent bien les avantages du  $S$ -estimateur et du test robuste dans le cas de lois multivariées contaminées.

**Mots-clés :** *Indice de redondance, S-estimateur, point de rupture, fonction d'influence, test robuste.*

### ABSTRACT

In this paper the  $S$ -estimator of  $\rho I$ , the redundancy index introduced by Stewart and Love (1968), is obtained together with a robust asymptotic test for  $H_0 : \rho I = 0$  against  $H_1 : \rho I > 0$ . The breaking point of the estimator is discussed and its influence function is obtained when the underlying distribution is in the class of elliptical distributions. The asymptotic power of the test is also obtained. Some examples and simulation results show clearly the advantages of the  $S$ -estimator and the robust test in the case of contaminated multivariate normal data.

**Keywords :** *Redundancy index, S estimator, breakdown point, influence function, robust test.*

### 1. Introduction

Dans cet article on s'intéresse à l'inférence robuste sur un indice de redondance. On considère l'indice de redondance de Stewart et Love (1968) noté  $\rho I$  et on trouve la loi asymptotique de son estimateur  $RI$  obtenu lorsque la matrice de dispersion est estimée par une matrice  $V_n$  qui est  $S$ -robuste, dans le cas d'une population elliptique. On propose un test robuste pour  $H_0 : \rho I = 0$  contre  $H_1 : \rho I > 0$  et on obtient sa fonction de puissance asymptotique. Des exemples illustrent la supériorité du test robuste lorsque des données sont contaminées. Des simulations illustrent aussi la

---

<sup>1</sup> Cette recherche fut subventionnée en partie par le CRSNG du Canada.

supériorité de l'estimateur  $RI$  robuste sur l'estimateur habituel. On étudie également le point de rupture de l'estimateur robuste et on obtient la fonction d'influence du  $S$ -estimateur de  $\rho I$  dans le cas d'une loi elliptique.

L'indice de redondance de Stewart et Love (1968), généralisé par Gleason (1976), est une mesure de la qualité de la prédiction d'un vecteur aléatoire  $X^{(1)} : p \times 1$  par un vecteur aléatoire  $X^{(2)} : q \times 1$  dans trois contextes différents :

- (i) dans un modèle de régression linéaire multivarié où il généralise le coefficient de corrélation multiple,
- (ii) dans un modèle d'analyse de redondance, alternative à l'analyse canonique souvent utilisé en recherche psychométrique (van der Wollenberg (1977)),
- (iii) dans un modèle de régression PLS que l'on retrouve fréquemment en chimométrie ou dans les problèmes de calibration (Wold (1982)).

En fait, l'indice de redondance peut mesurer la qualité de prédiction à chaque étape dans un processus séquentiel de sélection de variables ou de construction de facteurs latents (voir par exemple Lazraq et Cléroux (1988, 2001)). Il fut utilisé, en outre, pour généraliser l'analyse canonique à plusieurs groupes (Lazraq, Cléroux et Kiers (1992)), pour construire des variables principales (McCabe (1984)), pour étudier l'optimalité des composantes principales (Okamoto (1969)) et, comme mentionné plus haut, pour développer une alternative à l'analyse canonique (van der Wollenberg (1977)).

L'article est organisé de la façon suivante. La Section 2 présente l'indice de redondance  $\rho I$ . Dans la Section 3 on rappelle brièvement les propriétés de la loi elliptique que nous utilisons par la suite. L'estimateur  $S$ -robuste de  $\rho I$  est introduit dans la Section 4 et dans la Section 5 on obtient sa fonction d'influence. La Section 6 présente la loi asymptotique du  $S$ -estimateur de  $\rho I$  sous l'hypothèse  $H_0 : \rho I = 0$  lorsque la loi sous-jacente est elliptique. Un test de  $H_0$  est ainsi obtenu. Des exemples et une étude empirique sont présentés dans la Section 7. Finalement, la puissance asymptotique du test est obtenue dans la Section 8 et une conclusion suit dans la Section 9.

## 2. Un indice de redondance

Considérons deux vecteurs aléatoires  $X^{(1)} : p \times 1$  et  $X^{(2)} : q \times 1$  avec moyennes  $E(X^{(i)}) = \mu^{(i)}$  et matrices de covariances  $\Sigma_{ij} = E(X^{(i)} - \mu^{(i)})(X^{(j)} - \mu^{(j)})'$  pour  $i = 1, 2$  et  $j = 1, 2$ . Soit  $m = p + q$  et écrivons

$$X = (X_1, X_2, \dots, X_m)' = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix} \quad \text{et} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

L'indice de redondance introduit par Stewart et Love (1968) et par la suite généralisé par Gleason (1976) est donné par

$$\rho I(X^{(1)}, X^{(2)}) = \frac{\text{tr}(\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})}{\text{tr}(\Sigma_{11})} \quad (1)$$

où  $tr(\cdot)$  désigne la trace d'une matrice. Il peut aussi s'exprimer comme

$$\rho I(X^{(1)}, X^{(2)}) = \frac{\sum_{i=1}^p \sigma_i^2 \bar{R}_{i:p+1, \dots, m}^2}{\sum_{i=1}^p \sigma_i^2} \quad (2)$$

où  $\bar{R}_{i:p+1, \dots, m}^2$  est le carré du coefficient de corrélation multiple entre  $X_i^{(1)}$ , la  $i^e$  composante de  $X^{(1)}$ , et le vecteur  $X^{(2)}$  et où  $\sigma_i^2$  est la variance de  $X_i^{(1)}$ . L'indice de redondance est donc une moyenne pondérée des carrés des coefficients de corrélation multiple entre les composantes de  $X^{(1)}$  et le vecteur  $X^{(2)}$ . Il sert à mesurer la qualité de la prédiction du vecteur  $X^{(1)}$  par le vecteur  $X^{(2)}$  ou par un sous-vecteur de  $X^{(2)}$ . Il fut utilisé, entre autres, pour la sélection de variables en régression linéaire multivariée (voir Lazraq et Cléroux (1988)) et pour la sélection de variables latentes en régression PLS multivariée (Lazraq et Cléroux (2001)).

Il est connu que  $0 \leq \rho I \leq 1$ , que  $\rho I$  devient égal à  $\rho^2$ , le carré du coefficient de corrélation simple, lorsque  $p = q = 1$  et qu'il se réduit au carré du coefficient de corrélation multiple lorsque  $p = 1$  et  $q > 1$ . Le lecteur intéressé trouvera d'autres résultats concernant l'indice de redondance  $\rho I$  dans Lazraq et Cléroux (1992) de même que dans Lazraq, Cléroux et Kiers (1992).

### 3. Rappel des propriétés d'une loi elliptique

On rappelle brièvement la définition d'une loi elliptique ainsi que ses principales propriétés qui seront régulièrement utilisées dans la suite du texte. Les preuves se trouvent, par exemple, dans Giri (1993).

Le vecteur aléatoire  $X : m \times 1$  possède une loi appartenant à la famille des lois elliptiques  $E_m(\mu, V)$  avec paramètre de centrage  $\mu : m \times 1$  et paramètre d'échelle  $V : m \times m$ , matrice définie positive, si sa fonction de densité s'écrit

$$f(x) = |V|^{-1/2} g((x - \mu)'V^{-1}(x - \mu)) \quad (3)$$

pour une certaine fonction  $g(\cdot)$ , où  $|V|$  désigne le déterminant de la matrice  $V$ . On écrit  $X \sim E_m(\mu, V)$ . Si  $X \sim E_m(0, I)$  on dit que  $X$  possède la loi sphérique.

La famille des lois elliptiques contient en particulier les lois multinormales et multinormales contaminées ainsi que la loi de Student multivariée et on l'utilise souvent dans des études de robustesse à l'absence de multinormalité. Ses principales propriétés sont les suivantes :

- (i) si  $X \sim E_m(\mu, V)$ , sa fonction caractéristique est de la forme  $E(e^{it'x}) = e^{it'\mu} \varphi(t'Vt)$  où  $t : m \times 1$  et où  $\varphi(\cdot)$  est une fonction définie sur  $[0, \infty)$ . Aussi,  $var(X) = \Sigma = -2\varphi'(0)V$  et  $E(X) = \mu$ .
- (ii) si  $X \sim E_m(\mu, V)$ , si  $b : m \times 1$  et  $C : m \times m$  définie positive ne sont pas aléatoires, alors  $Y = CX + b \sim E_m(C\mu + b, CV C')$ .

(iii) si  $X \sim E_m(\mu, V)$  alors toute loi marginale est également elliptique. En particulier, si on écrit

$$X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix} \text{ et } V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$$

avec  $X^{(1)} : p \times 1$ ,  $X^{(2)} : q \times 1$ ,  $\mu^{(1)} : p \times 1$ ,  $\mu^{(2)} : q \times 1$ ,  $V_{11} : p \times p$ ,  $V_{22} : q \times q$ ,  $V_{21} = V_{12}' : p \times q$ ,  $p + q = m$ , alors  $X^{(1)} \sim E_p(\mu^{(1)}, V_{11})$ . Et si  $\text{var}(X) = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ , alors  $\Sigma_{11} = -2\varphi'(0)V_{11}$ .

#### 4. Estimateur $S$ -robuste de $\rho I$

Soit  $X_\alpha = \begin{pmatrix} X_\alpha^{(1)} \\ X_\alpha^{(2)} \end{pmatrix}$ ,  $\alpha = 1, 2, \dots, n$  un échantillon aléatoire iid provenant de la population  $X \sim E_m(\mu, V)$  et soit  $Z \sim E_m(0, I)$ . Posons  $t = |Z|$ . Dans le contexte de la régression multivariée, Rousseeuw et Yohai (1984) ont proposé des estimateurs  $(\mu_n, V_n)$  de  $(\mu, V)$  qui sont robustes et asymptotiquement multinormaux. Ce sont les  $S$ -estimateurs. Nous rappelons brièvement cette approche telle qu'on la retrouve dans Davies (1987), Lopuhaä (1989) et Bilodeau et Brenner (1999). Nous l'utiliserons par la suite pour obtenir un  $S$ -estimateur de  $\rho I$ .

##### 4.1. Rappel des $S$ -estimateurs

Soit  $\rho$  une fonction qui satisfait les conditions suivantes

$C_1$  :  $\rho$  est symétrique, possède une dérivée continue  $\psi$  et  $\rho(0) = 0$ .

$C_2$  : il existe une constante  $C_0 > 0$  telle que  $\rho$  est strictement croissante sur  $[0, C_0]$  et constante sur  $[C_0, \infty]$ . Soit  $a_0 = \sup \rho$ .

$C_3$  :  $\rho$  possède une dérivée seconde  $\psi'$  avec  $\psi'(t)$  et  $u(t) = \frac{\psi(t)}{t}$  continues et bornées.

Si  $t_i = [(X_i - \mu_n)' V_n^{-1} (X_i - \mu_n)]^{1/2}$ ,  $i = 1, 2, \dots, n$ , les  $S$ -estimateurs  $(\mu_n, V_n)$  sont définis comme les solutions du problème d'optimisation

$$\min |V_n| \text{ sujet à } \frac{1}{n} \sum_{i=1}^n \rho(t_i) = b_0 \quad (4)$$

où  $b_0$  est une constante qui sera spécifiée ultérieurement.

Rousseeuw et Yohai (1984) démontrent la proposition suivante :

**PROPOSITION 4.1.** – *Supposons les conditions  $C_1$ ,  $C_2$  et  $C_3$  satisfaites et supposons en outre  $E[\psi'(t)] > 0$  et  $E[\psi'(t)t^2 + (m+1)\psi(t)t] > 0$  avec  $t = |Z|$ .*

Posons

$$\beta_1 = \frac{1}{m} E[\psi^2(t)], \beta_2 = E\left[\left(1 - \frac{1}{m}\right)u(t) + \frac{1}{m}\psi'(t)\right],$$

$$\sigma_1 = \frac{m(m+2)E[\psi^2(t)t^2]}{E^2[\psi'(t)t^2 + (m+1)\psi(t)t]}$$

et

$$\sigma_2 = -\frac{2\sigma_1}{m} + 4\frac{E[\rho(t) - b_0]^2}{E^2[\psi(t)t]}.$$

Alors  $\sqrt{n}(V_n - V, \mu_n - \mu) \xrightarrow{\mathcal{L}} (N_1, n_1)$  où  $n_1$  et  $N_1$  sont indépendantes,  $n_1 \sim N_m(0, \frac{\beta_1 V}{\beta_2^2})$  et  $\text{vec}(N_1) \sim N_{m^2}[0, \sigma_1(I + K_m)(V \otimes V) + \sigma_2 \text{vec}(V)\text{vec}'(V)] \equiv N_{m^2}(0, \wedge)$ .

**Note :**  $\xrightarrow{\mathcal{L}}$  signifie “convergence en loi”,  $\text{vec}(N_1)$  est le vecteur  $m^2 \times 1$  formé en empilant les colonnes de  $N_1$  :  $m \times m$  et  $K_m$  est la matrice de commutation  $m^2 \times m^2$  définie par  $K_m = \sum_{i=1}^m \sum_{j=1}^m (e_i e_j') \otimes (e_j e_i')$  où  $e_i$  est un vecteur  $m \times 1$  dont les composantes sont 0 sauf la  $i^e$  qui prend la valeur 1.

Une fonction  $\rho(t)$  typique est la fonction double poids de Tuckey

$$\rho(t) = \begin{cases} t^2/2 - t^4/2c_0^2 + t^6/6c_0^4 & \text{si } |t| \leq c_0 \\ c_0^2/6 & \text{si } |t| \geq c_0. \end{cases} \quad (5)$$

La constante  $b_0$ ,  $0 < b_0 < a_0$  doit être choisie de façon que  $0 < b_0/a_0 = r \leq (n-m)/2n$ . Ce choix permet d’obtenir un point de rupture empirique égal à  $\varepsilon_n^* = \frac{[nr]}{n}$  (voir Lopuhaä et Rousseeuw (1991)). Le choix  $r = \frac{n-m}{2n}$  permet d’obtenir un point de rupture égal à  $\left\lfloor \frac{(n-m+1)}{2} / n \right\rfloor$  asymptotiquement égal à 50%.

En pratique, le point de rupture est le pourcentage minimal d’observations aberrantes nécessaires pour faire exploser un estimateur au delà de toute limite. La moyenne empirique ne nécessite pour cela qu’une seule observation et possède donc un point de rupture égal à  $\frac{1}{n}$  ou asymptotiquement 0. Pour obtenir simultanément un point de rupture empirique égal à  $\varepsilon_n^* = \frac{[nr]}{n}$  et un estimateur  $V_n$  convergent en probabilité vers  $V$  dans le cas d’une loi  $E_m(\mu, V)$ , la constante  $c_0$  est choisie de telle façon que  $\frac{E[\rho(t)]}{a_0} = r$  et donc  $b_0 = a_0 r = E[\rho(t)]$ .

Remarquons aussi que le volume d’un ellipsoïde de dimension  $n$  défini par  $z'V^{-1}z \leq 1$  est  $\frac{|V|^{1/2} 2\pi^{n/2}}{\Gamma(\frac{n}{2})m}$  de sorte que minimiser  $|V|$  revient à trouver un ellipsoïde de volume minimal (Rousseeuw (1985)).

#### 4.2. Estimateur $S$ -robuste de $\rho I$

Puisque  $\Sigma = -2\varphi'(0)V$ , si le  $S$ -estimateur de  $V$  est  $V_n = \begin{pmatrix} V_{11}^{(n)} & V_{12}^{(n)} \\ V_{21}^{(n)} & V_{22}^{(n)} \end{pmatrix}$  alors le  $S$ -estimateur de  $\rho I$  est donné par

$$RI = \frac{\text{tr}(V_{12}^{(n)} V_{22}^{(n)-1} V_{21}^{(n)})}{\text{tr}(V_{11}^{(n)})} \quad (6)$$

et comme  $V_n \xrightarrow{p} V$  on aura également  $RI \xrightarrow{p} \rho I$ . On s'intéresse maintenant au point de rupture de  $RI$ . Désignons par  $Z_n$  l'échantillon  $\begin{pmatrix} X_\alpha^{(1)} \\ X_\alpha^{(2)} \end{pmatrix}$ ,  $\alpha = 1, 2, \dots, n$ , de  $n$  observations iid provenant de  $X \sim E_m(\mu, V)$  et désignons par  $T_{\cdot Z_n}$  un estimateur d'un paramètre  $T$  calculé à partir de  $Z_n$ .

Le point de rupture de  $T_{\cdot Z_n}$  peut aussi être défini par

$$\varepsilon_n^*(T_{\cdot Z_n}) = \min_{1 \leq m \leq n} \left\{ \frac{m}{n} \setminus \sup_{Z'_n} \| T_{\cdot Z'_n} - T_{\cdot Z_n} \| = \infty \right\} \quad (7)$$

le sup étant pris sur tous les sous-ensembles  $Z'_n$  qui diffèrent de  $Z_n$  par au plus  $m$  points. C'est donc dire que  $\varepsilon_n^*(T_{\cdot Z_n})$  est la fraction minimale d'observations de  $Z_n$  que l'on doit remplacer pour faire exploser  $T_{\cdot Z_n}$  au delà de toute limite.

Le théorème suivant montre que le numérateur et le dénominateur de  $RI = \rho I_{\cdot Z_n}$  ont le même point de rupture. La preuve s'inspire de Rousseeuw et al (2000).

**THÉORÈME 4.2.** – Soit  $Z_n$  un ensemble de  $n \geq p + q + 1$  observations et soit  $V_n = V_{\cdot Z_n}$  un estimateur de la matrice  $V$  avec  $\varepsilon_n^* = \frac{\lfloor nr \rfloor}{n}$  où  $r \leq \frac{n - (p + q)}{2n}$ . Alors le numérateur et le dénominateur de  $\rho I_{\cdot Z_n}$  ont le même point de rupture  $\frac{\lfloor nr \rfloor}{n}$ .

*Démonstration.* – Soit  $Z'_n$  un ensemble obtenu en remplaçant  $m < \frac{\lfloor nr \rfloor}{n}$  observations de l'ensemble original  $Z_n$  par des valeurs arbitraires et soit la norme matricielle définie sur une matrice  $A$  par  $\| A \|_2 = \sup_{\|x\| \neq 0} \frac{\|Ax\|}{\|x\|}$  où  $\| \cdot \|$  est la norme euclidienne usuelle. Alors

$$\| V_{12 \cdot Z'_n} V_{22 \cdot Z'_n}^{-1} V_{21 \cdot Z'_n} \|_2 \leq \| V_{12 \cdot Z'_n} \|_2 \| V_{22 \cdot Z'_n}^{-1} \|_2 \| V_{21 \cdot Z'_n} \|_2.$$

Si  $\lambda_1(V_{22 \cdot Z'_n})$  est la plus petite valeur propre de  $V_{22 \cdot Z'_n}$  et si  $\lambda_{p+q}(V_{\cdot Z'_n})$  et  $\lambda_p(V_{11 \cdot Z'_n})$  sont respectivement les plus grandes valeurs propres de  $V_{\cdot Z'_n}$  et de  $V_{11 \cdot Z'_n}$ , on peut écrire

a)

$$\begin{aligned} \|V_{22 \cdot Z'_n}^{-1}\|_2 &= \sup_{\|x\| \neq 0} \frac{\|V_{22 \cdot Z'_n}^{-1} x\|}{\|x\|} \\ &= \frac{1}{\lambda_1(V_{22 \cdot Z'_n})} \end{aligned}$$

qui est borné puisque la matrice  $V_{\cdot Z'_n}$  ne peut se rompre (au sens du point de rupture) si  $m < \frac{[nr]}{n}$ , d'après les résultats de la fin du paragraphe 4.1.

b)

$$\|V_{11 \cdot Z'_n}\|_2 = \lambda_p(V_{11 \cdot Z'_n}) \text{ et } \|V_{12 \cdot Z'_n}\|_2 \leq \|V_{\cdot Z'_n}\|_2 = \lambda_{p+q}(V_{\cdot Z'_n})$$

sont bornés pour la même raison (l'inégalité vient de Golub et Van Loan (1989), p. 60).

Il suit que le numérateur et le dénominateur de  $\rho I_{\cdot Z'_n}$  ont le même point de rupture  $\frac{[nr]}{r}$ . □

Notons que puisque  $0 \leq \rho I_{\cdot Z'_n} \leq 1$ ,  $\rho I_{\cdot Z'_n}$  ne se rompra jamais même si l'on modifie toutes les observations de  $Z'_n$ .

### 5. Fonction d'influence du $S$ -estimateur de $\rho I$

Dans cette section on obtient la fonction d'influence du  $S$ -estimateur de  $\rho I$  lorsque la population suit une loi  $E_m(\mu, V)$  et on réalise qu'elle est bornée. Mais rappelons d'abord la définition de la fonction d'influence telle qu'introduite par Hampel (1974). Soit  $F$  une fonction de répartition et  $F_\varepsilon = (1 - \varepsilon)F + \varepsilon\delta_x$  une perturbation de  $F$  par  $\delta_x$ , la fonction de répartition qui affecte la probabilité 1 en  $x$ . Soit  $\theta(F) = T(F)$  un paramètre exprimé comme fonction de  $F$  et  $\theta(F_\varepsilon) = T(F_\varepsilon)$ . La fonction d'influence en  $x$  est définie par

$$I(x; \theta, F) = \lim_{\varepsilon \rightarrow 0} \left[ \frac{\theta(F_\varepsilon) - \theta(F)}{\varepsilon} \right] = \left. \frac{\partial}{\partial \varepsilon} \theta(F_\varepsilon) \right|_{\varepsilon=0}. \tag{8}$$

Soit  $C(F)$  le  $S$ -estimateur de  $V$  calculé à partir d'une loi théorique  $F$ . Rappelons que  $V_n$  est le  $S$ -estimateur de  $V$  calculé à partir de la loi empirique  $F_n$ . On écrit  $C(F_n) = V_n$ .

La fonction d'influence du  $S$ -estimateur de  $\rho I$  s'obtiendra à partir du lemme suivant démontré dans Lopuhaä (1989). Ici  $x \sim E_m(\mu, V)$ ,  $Z \sim E_m(0, I)$  et  $F_{\mu, V} = E_m(\mu, V)$ .

LEMME 5.1. – *Sous les conditions de la Proposition 4.1*



a) la fonction d'influence  $IF(Z; C, F_{0,I})$  du  $S$ -estimateur  $C$  au point  $Z$  calculé sur la loi  $F_{0,I}$  est donné par

$$IF(Z; C, F_{0,I}) - \frac{1}{m} \text{tr}[IF(Z; C, F_{0,I})]I = \frac{1}{\gamma_1} m \psi(\|Z\|) \|Z\| \left( \frac{ZZ'}{\|Z\|^2} - \frac{1}{m} I \right) \quad (9)$$

et

$$\frac{1}{m} \text{tr}[IF(Z; C, F_{0,I})] = \frac{2}{\gamma_2} (\rho(\|Z\|) - b_0)$$

où

$$\gamma_1 = \frac{1}{m+2} E[\psi'(\|x\|) \|x\|^2 + (m+1)\psi(\|x\|) \|x\|]$$

et

$$\gamma_2 = E[\psi(\|Z\|) \|Z\|]$$

l'espérance étant prise sur la loi de  $Z \sim E_m(0, I_m)$ .

b) la fonction d'influence  $IF(x; C, F_{\mu,V})$  est obtenue à partir de  $IF(Z; C, F_{0,I})$  comme suit :

$$IF(x; C, F_{\mu,V}) = B IF(B^{-1}(x - \mu); C, F_{0,I}) B' \quad (10)$$

où  $B$  vérifie  $BB' = V$ .

Désignons par  $\rho IC(X, Y)$  la valeur de  $\rho I$  calculée sur la matrice  $C$ , c'est-à-dire sur le  $S$ -estimateur de  $V$  avec la loi théorique  $F \sim E_m(\mu, V)$ . Posons

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \text{ avec } C_{11} : p \times p, C_{22} : q \times q, C_{12} = C_{21}' : p \times q.$$

Alors

$$\rho IC(X^{(1)}, X^{(2)}) = \frac{\text{tr}(C_{12} C_{22}^{-1} C_{21})}{\text{tr}(C_{11})} = \rho IC(F).$$

THÉORÈME 5.2. – Sous les conditions du Lemme 5.1, la fonction d'influence  $IF(x; \rho IC, F)$  où  $F \sim E_m(\mu, V)$  est donnée par

$$IF(x; \rho IC, F) = \rho IC(F) \left[ \frac{2 \text{tr}(IF(x; C_{12}, F) B')}{\text{tr}(C_{11}^*)} - \frac{\text{tr}(IF(x; C_{11}, F))}{\text{tr}(C_{11})} - \frac{\text{tr}(BIF(x; C_{22}, F) B')}{\text{tr}(C_{11}^*)} \right] \quad (11)$$

où  $C_{11}^* = C_{12}C_{22}^{-1}C_{21}$  et  $B = C_{12}C_{22}^{-1}$ , la matrice des coefficients de régression de  $X^{(1)}$  sur  $X^{(2)}$ .

*Démonstration.* – Soit  $F_\varepsilon = (1 - \varepsilon)F + \varepsilon\delta_x$  et soit  $\rho IC(F_\varepsilon)$  la valeur de  $\rho IC$  calculée sur la loi  $F_\varepsilon$ . Alors

$$\rho IC(F_\varepsilon) = \frac{\text{tr}(C_{12}(F_\varepsilon)C_{22}^{-1}(F_\varepsilon)C_{21}(F_\varepsilon))}{\text{tr}(C_{11}(F_\varepsilon))} = \frac{N(F_\varepsilon)}{D(F_\varepsilon)}. \quad (12)$$

Donc

$$\begin{aligned} IF(x; \rho IC, F) &= \left. \frac{\partial}{\partial \varepsilon} \rho IC(F_\varepsilon) \right|_{\varepsilon=0} \\ &= \frac{\left. \frac{\partial N(F_\varepsilon)}{\partial \varepsilon} \right|_{\varepsilon=0} D(F) - N(F) \left. \frac{\partial D(F_\varepsilon)}{\partial \varepsilon} \right|_{\varepsilon=0}}{(\text{tr}(C_{11}))^2}. \end{aligned} \quad (13)$$

Calculons d'abord  $\left. \frac{\partial D(F_\varepsilon)}{\partial \varepsilon} \right|_{\varepsilon=0}$ .

Si les éléments de la diagonale de  $C_{11}$  sont  $a_{11}, a_{22}, \dots, a_{pp}$  il suit  $\text{tr}(C_{11}(F_\varepsilon)) = \sum_{j=1}^p a_{jj}(F_\varepsilon)$  d'où on tire

$$\left. \frac{\partial D(F_\varepsilon)}{\partial \varepsilon} \right|_{\varepsilon=0} = \sum_{j=1}^p \left. \frac{\partial a_{jj}(F_\varepsilon)}{\partial \varepsilon} \right|_{\varepsilon=0} = IF(x; \text{tr}(C_{11}), F) = \text{tr}(IF(x; C_{11}, F)). \quad (14)$$

Pour calculer  $\left. \frac{\partial N(F_\varepsilon)}{\partial \varepsilon} \right|_{\varepsilon=0}$ , on note, en utilisant la formule  $\text{tr}(ABA') = \text{vec}'(A')(I \otimes B)\text{vec}(A')$  (Muirhead (1982), p. 76), que

$$N(F_\varepsilon) = \text{vec}'(C_{21}(F_\varepsilon))(I \otimes C_{22}^{-1}(F_\varepsilon))\text{vec}(C_{21}(F_\varepsilon)) \quad (15)$$

et en dérivant on obtient

$$\begin{aligned} \left. \frac{\partial N(F_\varepsilon)}{\partial \varepsilon} \right|_{\varepsilon=0} &= IF'(x; \text{vec}(C_{21}), F)(I \otimes C_{22}^{-1})\text{vec}(C_{21}) \\ &\quad + \text{vec}'(C_{21})(I \otimes IF(x; C_{22}^{-1}, F))\text{vec}(C_{21}) \\ &\quad + \text{vec}'(C_{21})(I \otimes C_{22}^{-1})IF(x; \text{vec}(C_{21}), F). \end{aligned} \quad (16)$$

En réutilisant la même formule de la trace d'un triple produit matriciel ainsi que  $\text{vec}(AXB) = (B' \otimes A)\text{vec}(X)$  (Muirhead (1982), p. 74) il suit

$$\left. \frac{\partial N(F_\varepsilon)}{\partial \varepsilon} \right|_{\varepsilon=0} = 2IF'(x; \text{vec}(C_{21}), F)\text{vec}(C_{22}^{-1}C_{21}) + \text{tr}(C_{12}IF(x; C_{22}^{-1}, F)C_{21}). \quad (17)$$

Finalement, en utilisant  $\text{tr}(AB) = \text{vec}'(A)\text{vec}(B)$  puis  $IF(x; C_{22}^{-1}, F) = -C_{22}^{-1}IF(x; C_{22}, F)C_{22}^{-1}$ , cette dernière formule étant obtenue en dérivant  $C_{22}^{-1}(F_\varepsilon)C_{22}(F_\varepsilon) = I_q$ , (16) devient

$$\begin{aligned} \left. \frac{\partial N(F_\varepsilon)}{\partial \varepsilon} \right|_{\varepsilon=0} &= 2\text{vec}'(IF(x; C_{21}, F))\text{vec}(C_{22}^{-1}C_{21}) \\ &\quad - \text{tr}(C_{12}C_{22}^{-1}IF(x; C_{22}, F)C_{22}^{-1}C_{21}) \\ &= 2\text{tr}(IF(x; C_{12}, F))C_{22}^{-1}C_{21} \\ &\quad - \text{tr}(C_{12}C_{22}^{-1}IF(x; C_{22}, F)C_{22}^{-1}C_{21}). \end{aligned} \quad (18)$$

En reportant (14) et (18) dans (13) et après quelques manipulations algébriques, on obtient le résultat désiré (11).  $\square$

Le Théorème 5.2 nous donne la fonction d'influence  $IF(x; \rho IC, F)$  en termes de  $IF(x; C_{ij}, F)$  pour  $i, j = 1, 2$ . Ces dernières fonctions sont obtenues du Lemme 5.1. On réalise que la forme explicite de  $IF(x; \rho IC, F)$  est relativement complexe. D'autre part, puisque  $IF(x; C, F)$  est bornée (Lopuhaä (1989)), il suit que  $IF(x; \rho IC, F)$  l'est aussi contrairement à la fonction d'influence de  $\rho I$  dans le cas multinormal (Lazraq et Cléroux (1989)).

## 6. Loi asymptotique de $RI$ robuste sous $H_0 : \rho I = 0$

La loi cherchée nous est fournie par le théorème suivant.

**THÉORÈME 6.1.** – *Sous les hypothèses de la Proposition 4.1 et sous  $H_0 : \rho I = 0$ , la loi asymptotique de  $nRI$ , lorsque  $n \rightarrow \infty$ , est la loi de  $\sigma_1(\text{tr}(V_{11}))^{-1} \sum_{i=1}^p \delta_i \sum_{j=1}^q W_{ij}$  où  $\delta_1 \geq \delta_2 \geq \dots \geq \delta_p$  sont les valeurs propres de  $V_{11}$  chacune étant prise  $q$  fois et où les  $W_{ij}$  sont iid  $N(0, 1)$ .*

*Démonstration.* – Puisque  $RI$  défini par (6) est invariant sous les transformations orthogonales de  $X^{(1)}$  et sous les transformations de plein rang de  $X^{(2)}$ , on peut supposer, sans perte de généralité, que  $V_{11} = \Delta = \text{diag}(\delta_1, \delta_2, \dots, \delta_p)$  et  $V_{22} = I_q$ . Donc, sous  $H_0 : \rho I = 0$ ,  $V = \begin{pmatrix} \Delta & 0 \\ 0 & I_q \end{pmatrix}$ . Lorsque  $n \rightarrow \infty$ ,  $V_{11}^{(n)} \xrightarrow{p} V_{11} = \Delta$  et  $V_{22}^{(n)} \xrightarrow{p} V_{22} = I_q$ . Donc, sous  $H_0$ ,  $RI$  possède la même loi asymptotique que

$$RI^* = \frac{\text{tr}(V_{12}^{(n)} V_{22}^{-1} V_{21}^{(n)})}{\text{tr}(V_{11})} = \frac{\text{tr}(V_{12}^{(n)} V_{21}^{(n)})}{\text{tr}(\Delta)}. \quad (19)$$

Considérant le numérateur de (19), on peut écrire

$$n \text{tr}(V_{12}^{(n)} V_{21}^{(n)}) = n \text{vec}'(V_{21}^{(n)}) \text{vec}(V_{21}^{(n)}). \quad (20)$$

Par la proposition 4.1 et sous  $H_0$ , il suit

$$\sqrt{n} \text{vec}(V_{21}^{(n)}) = \sqrt{n} B \text{vec}(V^{(n)}) \xrightarrow{\mathcal{L}} N(0, B \wedge B') \quad (21)$$

où  $B = B_1 \otimes B_2$ ,  $B_1 = (I_p \ 0) : p \times m$  et  $B_2 = (0 \ I_q) : q \times m$  avec  $\Lambda = \sigma_1(I + K_m)(V \otimes V) + \sigma_2 \text{vec}(V) \text{vec}'(V)$ .

Donc (20) est asymptotiquement distribué comme  $\sum_{i=1}^{pq} w_i U_i^2$  où les  $U_i$  sont iid  $N(0, \Lambda)$  et où les  $w_i$  sont les valeurs propres de  $B \wedge B'$ . Mais  $B(V \otimes V)B' = \Delta \otimes I_q$ ,  $BK_m(V \otimes V)B' = 0$  et  $B \text{vec}(V) \text{vec}'(V)B' = 0$  d'où  $B \wedge B' = \sigma_1(\Delta \otimes I_q)$ .

Les valeurs propres de cette dernière matrice sont  $\sigma_1 \delta_i$ ,  $i = 1, 2, \dots, p$ , chaque  $\delta_i$  étant de multiplicité  $q$  et où  $\delta_1, \delta_2, \dots, \delta_p$  sont les valeurs propres de  $V_{11}$ . Cela complète la preuve du théorème.  $\square$

**COROLLAIRE 6.2.** – *Sous les mêmes hypothèses,  $E(n \ RI)$  converge vers  $q\sigma_1$  et  $\text{var}(n \ RI)$  converge vers  $2q\sigma_1^2 \frac{\text{tr}(V_{11}^2)}{(\text{tr}(V_{11}))^2}$  lorsque  $n \rightarrow \infty$ .*

Les résultats de cette section permettent d'effectuer un test asymptotique robuste de  $H_0 : \rho I = 0$  contre  $H_1 : \rho I > 0$ . Ce test est le suivant : rejeter  $H_0$  au niveau  $\alpha$  si  $n \ RI > C_\alpha$  où  $C_\alpha$  est le  $100(1 - \alpha)^e$  centile de la loi de  $\sigma_1(\text{tr}(V_{(11)}))^{-1} \sum_{i=1}^p \delta_i \sum_{j=1}^q U_{ij}^2$ . La valeur de  $C_\alpha$  est obtenue en utilisant l'algorithme exact de Imhof (1961) dont un programme FORTRAN est donné dans Koerts et Abrahamse (1989).

La loi précédente comporte des paramètres inconnus qui devront, en pratique, être remplacés par des estimateurs convergents. Bien que l'estimation des  $\delta_i$  ne pose pas de problème (on utilise  $V_{11}^{(n)}$  à la place de  $V_{11}$ ), on ne peut estimer  $\sigma_1$  en général. Pour les calculs de la prochaine section on se situe dans le contexte particulier de la loi multinormale contaminée. Si la proportion de données contaminées est faible, même si leurs valeurs sont très grandes, la matrice de covariance  $S$ -robuste affectera un poids presque nul à ces données et la valeur correspondante pour  $\sigma_1$  sera approximativement celle calculée à partir de la loi multinormale non contaminée.

## 7. Exemples et simulations

Tous les programmes utilisés pour les calculs de cette section ont été écrits en  $S$ -Plus. Le calcul de la matrice  $S$ -robuste  $V_n$  a été fait en utilisant les fonctions disponibles sur le site [www.dms.umontreal.ca/~bilodeau](http://www.dms.umontreal.ca/~bilodeau).

### 7.1. Premier exemple

Considérons les données sur le tabac que l'on retrouve dans Woltz et al (1948) et qui sont analysées également dans Lazraq et Cléroux (2001). Ici  $n = 25$ ,  $p = 3$  et  $q = 6$ . On veut tester  $H_0 : \rho I = 0$  contre  $H_1 : \rho I > 0$ . Le test développé par Lazraq et Cléroux (1992) quand les données proviennent d'une loi elliptique conduit à  $RI = 0.735$  et à une  $p$ -valeur égale à 0.0038.

Contaminons 2 observations comme suit : la première composante de  $X_1$  ainsi que la dernière de  $X_9$  sont multipliées par 100. Alors le même test appliqué aux nouvelles données conduit à  $RI = 0.058$  et à une  $p$ -valeur égale à 0.997. Par contre si l'on applique le test de la Section 6 à ces nouvelles données en choisissant un

point de rupture  $r = 0.1$  on obtient  $RI = 0.734$  et une  $p$ -valeur égale à 0.0042. Cet exemple illustre bien la robustesse de ce nouveau test.

### 7.2. Deuxième exemple

On recommence le même exercice sur les données de BMDP (1981), p. 38. Ici  $n = 181$  (après avoir éliminé les 7 individus possédant des données manquantes),  $p = 3$  et  $q = 4$ . Le test de Lazraq et Cléroux qui suppose une loi elliptique conduit à  $RI = 0.1684$  et à une  $p$ -valeur  $\simeq 0$ .

Si on contamine 4 observations en multipliant par 10 la première composante de  $X_1$ , la septième composante de  $X_{59}$  et les deux premières composantes de  $X_{181}$  et si l'on applique le même test à ces nouvelles données on obtient  $RI = 0.0458$  et une valeur- $p$  égale à 1. Par contre le test robuste proposé à la Section 6 conduit à  $RI = 0.1640$  et à une  $p$ -valeur  $\simeq 0$ .

### 7.3. Simulations

Les deux exemples précédents illustrent bien la robustesse du test de la Section 6 en présence de données contaminées. Dans cette sous-section on étudie par simulation la qualité de l'estimateur  $RI$  calculé à partir d'une matrice de dispersion  $V_n$  qui est  $S$ -robuste en comparaison avec l'estimateur de  $RI$  calculé à partir d'une matrice de covariance empirique habituelle  $S_n$ . On simule des données vectorielles à partir des paramètres suivants :

- (i)  $p = 2$  et  $q = 3$ ,
- (ii)  $n = 50, 75$
- (iii)  $L = 100$  échantillons générés pour chaque cas,
- (iv) deux types de lois sont considérés
  - a) données non contaminées : on génère les données à partir de la loi  $N_{p+q}(0, \Sigma)$  avec  $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$  où  $\Sigma_{11} = I_p$ ,  $\Sigma_{22} = I_q$ ,  $\Sigma_{12} = \Sigma'_{21} = C_{00}, C_{10}, C_{20}, C_{30}$  où la matrice  $C_{xy}$  a tous ses éléments égaux à  $0.xy$
  - b) données contaminées : pour  $n_C = [0.1n]$  (partie entière de 10% de  $n$ ) on génère  $n - n_C$  données de la loi précédente et  $n_C$  de la loi  $N_{p+q}(0, 49I)$ .

Dans chaque cas, pour chacune des  $L = 100$  simulations  $\ell = 1, 2, \dots, L$ , on calcule  $\rho I$ ,  $RI(S_n)$  et  $RI(V_n)$ . Par analogie à Dehon et al (2000) on stabilise la variance par la transformation  $w(RI) = \tanh^{-1}(\sqrt{RI})$  appliquée successivement à  $\rho I$ ,  $RI(S_n)$  et  $RI(V_n)$  puis on calcule la moyenne des carrés des erreurs d'estimation  $e^2 = MCE(RI) = \frac{1}{L} \sum_{\ell=1}^L [w(RI) - w(\rho I)]^2$  avec  $RI(S_n)$  et  $RI(V_n)$  pour obtenir respectivement  $e^2(S_n)$  et  $e^2(V_n)$ . Les résultats sont présentés aux Tableaux 1 et 2. Comme prévu, on remarque que (i) pour les données non contaminées  $e^2(S_n)$  et  $e^2(V_n)$  sont à toutes fins pratiques égales et (ii) pour les données contaminées,

TABLEAU 1

Sommes des carrés des erreurs d'estimation pour les données non contaminées

$n$	mesure	$C_{00}$	$C_{10}$	$C_{20}$	$C_{30}$
50	$e^2(S_n)$	0.059	0.023	0.019	0.012
	$e^2(V_n)$	0.060	0.023	0.019	0.012
75	$e^2(S_n)$	0.046	0.009	0.008	0.008
	$e^2(V_n)$	0.047	0.009	0.008	0.008

TABLEAU 2

Sommes des carrés des erreurs d'estimation pour les données contaminées

$n$	mesure	$C_{00}$	$C_{10}$	$C_{20}$	$C_{30}$
50	$e^2(S_n)$	0.558	0.326	0.185	0.086
	$e^2(V_n)$	0.190	0.100	0.047	0.032
75	$e^2(S_n)$	0.338	0.186	0.084	0.027
	$e^2(V_n)$	0.089	0.033	0.018	0.012

$e^2(S_n) > e^2(V_n)$  dans tous les cas, illustrant bien la supériorité de l'estimateur robuste.

**8. Loi asymptotique de RI sous  $H_1 : \rho I > 0$**

Nous aurons d'abord besoin de définir la dérivée matricielle  $\frac{da}{db} = \left( \frac{\partial a_i}{\partial b_j} \right)$  où  $i$  parcourt les lignes et  $j$  les colonnes.

Soit  $V$  une matrice  $m \times m$  et soit  $h$  une fonction de  $\mathbb{R}^{m^2}$  dans  $\mathbb{R}^q$ . La dérivée de  $h$  est définie par (voir Tyler (1983) ou Bilodeau et Brenner (1999), p. 228)

$$h'(V) = \frac{1}{2} [dh(V)/dvec(V)] [I + J_m] \in \mathbb{R}_{m^2}^q$$

où  $J_m = \sum_{i=1}^m e_i e_i' \otimes e_i e_i'$  et  $e_i \in \mathbb{R}^m$  contient 1 en position  $i$  et des 0 ailleurs.

La proposition suivante se trouve dans Bilodeau et Brenner (1999), pp. 228–229.

PROPOSITION 8.1. – Si  $h(\gamma V) = h(V)$  pour  $V$  définie positive et  $\forall \gamma > 0$  et si  $V_n$  est un estimateur de  $V$  qui vérifie  $\sqrt{n}(\text{vec}(V_n) - \text{vec}(V)) \xrightarrow{\mathcal{L}} Z$  avec  $Z \sim N_{m^2}(0, \sigma_1(I + K_m)(V \otimes V) + \sigma_2 \text{vec}(V) \text{vec}'(V))$  alors

$$\sqrt{n}[h(V_n) - h(V)] \xrightarrow{\mathcal{L}} Z_h \text{ où } Z_h \sim N_q(0, 2\sigma_1(h'(V))(V \otimes V)(h'(V))').$$

En prenant  $h(V) = RI \in R$ , on a  $q = 1$  et  $h(\gamma V) = h(V)$ ,  $\forall \gamma > 0$  et d'après Lazraq et Cléroux (1992) nous avons le théorème suivant :

THÉORÈME 8.2. – Sous les hypothèses de la proposition 4.1, si  $n \rightarrow \infty$  et si  $0 < \rho I < 1$ , la loi asymptotique de  $\sqrt{n}(RI - \rho I)$  est  $N(0, 2\sigma_1 \rho I^2 \sigma^2)$  où

$$\sigma^2 = \frac{\text{tr}(V_{11}^2)}{(\text{tr}(V_{11}))^2} - \frac{4\text{tr}(V_{11}V_{11}^*) - 2\text{tr}(V_{11}^{*2})}{4\text{tr}(V_{11})\text{tr}(V_{11}^*)} + \frac{2\text{tr}(V_{11}V_{11}^*) - \text{tr}(V_{11}^{*2})}{(\text{tr}(V_{11}^*))^2}$$

où  $V_{11}^* = V_{12}V_{22}^{-1}V_{21}$ .

Ce résultat permet de calculer la puissance asymptotique du test de  $H_0 : \rho I = 0$  contre  $H_1 : \rho I > 0$ . Elle est égale à  $p[nRI > C_\alpha \mid H_1] = p\left[N(0, 1) > \frac{C_\alpha - n\rho I}{\rho I \sigma \sqrt{2n\sigma_1}}\right]$ . En pratique, cependant, on devra connaître un estimateur convergent de  $\sigma_1$ .

## 9. Conclusion

Dans cet article on a considéré un  $S$ -estimateur de  $\rho I$  dans le cas des lois elliptiques et on a construit un test robuste pour  $H_0 : \rho I = 0$  contre  $H_1 : \rho I > 0$ . Des exemples ainsi que des simulations en ont montré les avantages par rapport à l'estimateur et au test classique. On a aussi étudié le point de rupture et la fonction d'influence de l'estimateur. Et finalement on a obtenu la puissance asymptotique du test robuste. Toutefois, un problème demeure : celui d'obtenir un estimateur convergent de  $\sigma_1$ . Dans les calculs, nous avons contourné ce problème en utilisant des lois multinormales contaminées dont le pourcentage de données contaminées est faible.

Mentionnons en terminant que le travail précédent aurait pu être réalisé en utilisant, plutôt que le  $S$ -estimateur de  $\rho I$ , l'estimateur MCD (Minimum Covariance Determinant Estimator) (voir par exemple Rousseeuw et Van Driessen (1999)). Mais les comparaisons numériques de Croux et Haesbroeck (1999) nous ont incités à préférer le  $S$ -estimateur.

## 10. Remerciements

Les auteurs désirent remercier l'arbitre anonyme pour ses commentaires. Rarement avait-on vu un arbitrage aussi efficace.

**Bibliographie**

- BILODEAU M. and BRENNER D. (1999), *Theory of Multivariate Statistics*, Springer, New York.
- BMDP Statistical Software (1981), W.J. Dixon, chief Editor, University of California Press, Berkeley.
- CROUX C. and HAESBROECK G. (1999), Influence Function and Efficiency of the Minimum Covariance Determinant Scatter Matrix Estimator, *Journal of Multivariate Analysis*, **71**, 161–190.
- DAVIES P.L. (1987), Asymptotic Behavior of  $S$ -Estimates of Multivariate Location Parameters and Dispersion Matrices, *Annals of Statistics*, **15**, 1269–1292.
- DEHON C., FITZMOSER P. and CROUX C. (2000), *Robust Methods for Canonical Correlation Analysis*, in *Data Analysis, Classification and Related Methods*, Eds. H.A.L. Kiers, J.-P. Rosson, P.J.E. Groenen and M. Schader, Springer, Berlin, pp. 321–326.
- GIRI N.C. (1993), *Introduction to Probability and Statistics*, Marcel Dekker Inc., New York.
- GLEASON T.C. (1976), On Redundancy in Canonical Analysis, *Psychological Bulletin*, **83**, 1004–1006.
- GOLUB G.H. and VAN LOAN C.F. (1989), *Matrix Computations*, 2nd ed., The Johns Hopkins University Press.
- HAMPEL F.R. (1974), The Influence Curve and its Role in Robust Estimation, *Journal of the American Statistical Association*, **69**, 383–393.
- IMHOF P. (1961), Computing the Distribution of Quadratic Forms in Normal Variables, *Biometrika*, **48**, 419–426.
- KOERTS J. and ABRAHAMSE A.P.J. (1969), *On the Theory and Application of the General Linear Model*, Rotterdam University Press, Rotterdam.
- LAZRAQ A. et CLÉROUX R. (1988), Un algorithme pas à pas de sélection de variables en régression linéaire multivariée, *Statistique et Analyse des Données*, **13**, 39–58.
- LAZRAQ A. and CLÉROUX R. (1989), *On the Detection of Multivariate Data Outliers and Regression Outliers*, in *Data Analysis, Learning Symbolic and Numeric Knowledge*, Ed. E. Diday, Nova Science Publishers, New York, pp. 133–140.
- LAZRAQ A. et CLÉROUX R. (1992), Tests d'homogénéité entre indices de redondance pour des lois elliptiques, *Revue de Statistique Appliquée*, XXXX, No. 3, 19–33.
- LAZRAQ A. and CLÉROUX R. (2001), The PLS Multivariate Regression Model : Testing the Significance of Successive PLS Components, *Journal of Chemometrics*, **15**, 523–536.



- LAZRAQ A., CLÉROUX R. et KIERS H.A.L. (1992), Mesures de liaison vectorielle et généralisation de l'analyse canonique, *Revue de Statistique Appliquée*, XXXX, No. 1, 23–35.
- LOPUHAÄ H.P. (1989), On the Relation Between  $S$ -Estimators and  $M$ -estimators of Multivariate Location and Covariance, *Annals of Statistics*, **17**, 1662–1683.
- LOPUHAÄ H.P. and ROUSSEUW P.J. (1991), Breakdown Points of Affine Equivariant Estimators of Multivariate Location and Covariance Matrices, *Annals of Statistics*, **19**, 229–248.
- MCCABE G.P. (1984), Principal Variables, *Technometrics*, **26**, 137–144.
- MUIRHEAD R.J. (1982), *Aspects of Multivariate Statistical Theory*, John Wiley and Sons, New York.
- OKAMOTO M. (1969), *Optimality of Principal Components*, in *Multivariate Analysis II*, Ed. P.R. Krishnaiah et al., Academic Press, New York, pp. 673–685.
- ROUSSEUW P.J. (1985), *Multivariate Estimation with High Breakdown Point*, in *Mathematical Statistics and Applications*, Eds. W. Grossmann, G. Pflug, I. Veneze and W. Wertz, vol. 8, Reidel, Dordrecht, pp. 283–297.
- ROUSSEUW P.J., VAN AELST S. and DRIESSEN K.V. (2000), *Robust Multivariate Regression*, Technical Report, Dept. of Mathematics and Computer Science, University of Antwerp.
- ROUSSEUW P.J. and VAN DRIESSEN K. (1999), A Fast Algorithm for the Minimum Covariance Determinant Estimator, *Technometrics*, **41**, 212–223.
- ROUSSEUW P.J. and YOHAI V.J. (1984), Robust Regression by Means of  $S$ -Estimator. Robust and Nonlinear Time Series Analysis, *Lecture Notes in Statistics*, Vol. 26, Springer, New York, pp. 256–272.
- STEWART D. and LOVE W. (1968), A General Canonical Correlation Index, *Psychological Bulletin*, **70**, 160–163.
- TYLER D.E. (1983), Robustness and Efficiency Properties of Scatter Matrices, *Biometrika*, **70**, 411–420.
- VAN DEN WOLLENBERG A.L. (1977), Redundancy Analysis, and Alternative for Canonical Correlation Analysis, *Psychometrika*, **42**, 207–219.
- WOLD H. (1982), *Soft Modeling : the Basic Design and Some Extensions*. In *Systems Under Indirect Observation, Part II*, Jöreskog K.G., Wold H. (eds). North Holland, Amsterdam, pp. 1–54.
- WOLTZ W.G., REID W.A. and COLWELL W.E. (1948), Sugar and Nicotine in Cured Bright Tobacco as Related to Mineral Elements Composition, *Proc. Soil Sci. Soc. Am.*, **13**, 385–387.