

REVUE DE STATISTIQUE APPLIQUÉE

R. CLÉROUX

A. LAZRAQ

Y. LEPAGE

Indice de redondance basé sur les rangs et inférence non paramétrique

Revue de statistique appliquée, tome 42, n° 2 (1994), p. 79-98

http://www.numdam.org/item?id=RSA_1994__42_2_79_0

© Société française de statistique, 1994, 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/>

INDICE DE REDONDANCE BASÉ SUR LES RANGS ET INFÉRENCE NON PARAMÉTRIQUE*

R. Cléroux (1), A. Lazraq (2), Y. Lepage (3)

(1) *Département d'informatique et de recherche opérationnelle*

Université de Montréal, Montréal, Québec, Canada

(2) *Département d'informatique*

École Nationale de l'Industrie Minérale, Rabat, Maroc

(3) *Département de mathématiques et de statistique*

Université de Montréal, Montréal, Québec, Canada

RÉSUMÉ

Une mesure de liaison basée sur les rangs est proposée pour tester l'absence de liaison entre deux vecteurs aléatoires. Les lois asymptotiques de la statistique du test sous l'hypothèse de l'absence de liaison et sous une série d'alternatives sont présentées. Une étude expérimentale montre la puissance de ce test non paramétrique par comparaison avec deux tests compétiteurs.

Mots-clés : *Corrélation vectorielle, mesures de redondance, test multivarié basé sur les rangs, test d'indépendance.*

SUMMARY

A measure of association based on ranks is introduced in order to test the independence between two vectors. Its asymptotic law is derived under the null hypothesis of independence and under a sequence of alternatives. An empirical study is performed to compare the new test with two other competitors.

Key-words : *Vector correlation, redundancy measures, multivariate rank test, test for independence.*

Introduction

Dans cet article nous nous intéressons aux mesures d'association entre deux ensembles de variables. Plusieurs auteurs ont considéré des mesures de corrélation, redondance ou association entre deux vecteurs aléatoires, deux ensembles de variables ou deux matrices de données : voir par exemple Hotelling (1936), Masuyama (1939, 1941), Roseboom (1965), Stewart et Love (1968), Kshirsagar (1969), Escoufier (1973), Coxhead (1974), Cramer (1974), Lingoes et Shönemann (1974), Shaffer et

* Cette recherche fut partiellement subventionnée par le CRSNG (Canada), le FCAR (Québec) ainsi que le FRSQ (Québec).

Gillo (1974), Gleason (1976), Robert et Escoufier (1976), Cramer et Nicewander (1979), Stephens (1979) et finalement Ramsay, ten Berge et Styan (1984). Ces mesures furent groupées en deux classes par Cramer et Nicewander (1979) : mesures de redondance qui visent à prédire un ensemble de variables par un autre et mesures d'association ou de corrélation vectorielle qui généralisent le concept de coefficient de corrélation à deux ensembles de variables.

On rappelle d'abord l'indice de redondance de Stewart et Love (1968) et Gleason (1976) dans la section 1 et on introduit l'indice de redondance basé sur les rangs en utilisant le principe de la transformée de rangs de Conover et Iman (1980, 1981) dans la section 2. Les sections 3 et 4 concernent une matrice d'association qui est nécessaire aux développements théoriques. Les résultats distributionnels sous l'hypothèse de l'indépendance de deux vecteurs et sous une suite d'alternatives donnant une structure de covariances entre deux vecteurs sont obtenus respectivement dans les sections 5 et 6. Un algorithme décrivant le test non paramétrique proposé est donné dans la section 7. La section 8 présente un exemple d'application du nouveau test. Finalement, la section 9 contient une étude expérimentale comparant le test construit avec l'indice de redondance basé sur les rangs et deux autres compétiteurs, un test non paramétrique et un test paramétrique.

Une telle étude a été effectuée dans Cléroux, Lazraq et Lepage (1993) pour la mesure d'association vectorielle RV d'Escoufier (1973).

1. Un indice de redondance

Soient $X^{(1)} : p \times 1$ et $X^{(2)} : q \times 1$ deux vecteurs aléatoires 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 $j = 1, 2$. Ecrivons

$$X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}, \mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

et supposons Σ définie positive. Soit $X_\alpha = \begin{pmatrix} X_\alpha^{(1)} \\ X_\alpha^{(2)} \end{pmatrix}$, $\alpha = 1, 2, \dots, n$, un échantillon aléatoire de X et soient $\bar{X}^{(i)} = \frac{1}{n} \sum_{\alpha=1}^n X_\alpha^{(i)}$ et $S_{ij} = \frac{1}{n-1} \sum_{\alpha=1}^n (X_\alpha^{(i)} - \bar{X}^{(i)})(X_\alpha^{(j)} - \bar{X}^{(j)})'$ pour $i = 1, 2$ et $j = 1, 2$. Finalement, soient $\bar{X} = \begin{pmatrix} \bar{X}^{(1)} \\ \bar{X}^{(2)} \end{pmatrix}$ et $S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$. Les paramètres μ et Σ sont estimés par $\bar{X}$ et S .

L'indice de redondance s'inscrit dans un contexte de prédiction de $X^{(1)}$ par $X^{(2)}$. Le modèle de régression linéaire multivariée s'écrit

$$X_{(1)} = X_{(2)}B + E \quad (1.1)$$

où $X_{(1)}$ est une matrice $n \times p$ de p variables mesurées sur n individus, $X_{(2)}$ est une matrice $n \times q$ de q variables mesurées sur les mêmes individus, B est la matrice

$q \times p$ de régression et E une matrice aléatoire $n \times p$ de résidus. On suppose que les moyennes des colonnes de $X_{(1)}$ et $X_{(2)}$ sont nulles, les deux matrices sont de plein rang et $p \leq q$. Alors, les matrices de covariances s'écrivent $S_{11} = \frac{1}{n-1} X_{(1)}' X_{(1)}$, $S_{22} = \frac{1}{n-1} X_{(2)}' X_{(2)}$, $S_{12} = \frac{1}{n-1} X_{(1)}' X_{(2)}$ et en minimisant $tr(E'E)$, où $tr(\cdot)$ désigne la trace, on obtient $\hat{B} = S_{22}^{-1} S_{21}$ et $\hat{X}_{(1)} = X_{(2)} \hat{B}$.

Si on écrit $S_{11}^* = \hat{X}_{(1)}' \hat{X}_{(1)} = (X_{(2)} \hat{B})' (X_{(2)} \hat{B})$ alors $S_{11}^* = S_{12} S_{22}^{-1} S_{21}$. Finalement, si $SS_{1(i)}$ est le i^e élément de la diagonale de S_{11} et $SS_{1(i)}^*$ celui de S_{11}^* , alors le carré du coefficient de corrélation multiple pour prédire la i^e composante $X_i^{(1)}$ de $X^{(1)}$ par $X^{(2)}$, pour $i = 1, 2, \dots, p$, est donné par

$$R_{1(i) \cdot x_1^{(2)}, \dots, x_q^{(2)}}^2 = \frac{SS_{1(i)}^*}{SS_{1(i)}}. \quad (1.2)$$

En notant, $\hat{\rho}_1^2, \hat{\rho}_2^2, \dots, \hat{\rho}_p^2$ les carrés des corrélations canoniques entre $X^{(1)}$ et $X^{(2)}$, $\wedge = \text{diag}(\hat{\rho}_1^2, \hat{\rho}_2^2, \dots, \hat{\rho}_p^2)$, U la matrice $p \times p$ telle que la i^e colonne est le vecteur canonique correspondant à $\hat{\rho}_i$ et si $D = U^{-1}$, alors on a $S_{11} = D'D$ et $S_{11}^* = D' \wedge D$.

L'indice de redondance, introduit par Stewart et Love (1968) et généralisé par Gleason (1976), est défini par

$$RI = \frac{tr S_{11}^*}{tr S_{11}} = \frac{tr(S_{12} S_{22}^{-1} S_{21})}{tr(S_{11})} = \frac{tr(D' \wedge D)}{tr(D'D)}. \quad (1.3)$$

Cet indice correspond à la proportion de la variance de $X^{(1)}$ expliquée par $X^{(2)}$ et il peut également s'écrire sous la forme

$$RI = \frac{\sum_{i=1}^p \sigma_{1(i)}^2 R_{1(i) \cdot x_1^{(2)}, \dots, x_q^{(2)}}^2}{\sum_{i=1}^p \sigma_{1(i)}^2} \quad (1.4)$$

où $\sigma_{1(i)}^2$ est la variance de la i^e composante de $X^{(1)}$. 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 fut utilisé dans une procédure de régression linéaire multivariée pas à pas par Lazraq et Cléroux (1988).

On peut voir que $0 \leq RI \leq 1$, que RI devient égal à r^2 , le carré du coefficient de corrélation simple quand $p = q = 1$ et qu'il se réduit au carré du coefficient de corrélation multiple si $p = 1$ et $q > 1$. En général il n'est pas invariant sous les transformations de variables. On peut trouver d'autres résultats concernant l'indice RI dans Lazraq et Cléroux (1992) de même que dans Lazraq, Cléroux et Kiers (1992).

Ces résultats concernent la loi asymptotique de RI et permettent de construire de nouveaux tests pour des hypothèses sur des structures de corrélations entre deux vecteurs aléatoires.

Enfin, au niveau de la population, l'indice de redondance est noté

$$\rho I = \frac{\text{tr}(\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})}{\text{tr}(\Sigma_{11})}. \quad (1.5)$$

2. Un indice de redondance basé sur les rangs

Dans le présent article, la mesure RI est calculée en se basant sur le rang des observations selon l'approche introduite par Conover et Iman (1980, 1981). Les avantages tirés en se basant sur les rangs sont multiples :

- (i) les données peuvent être de type ordinal,
- (ii) les échelles de mesures pour chaque variable peuvent être différentes,
- (iii) l'hypothèse classique de la multinormalité ou de l'ellipticité de la loi des observations peut être écartée,
- (iv) l'approche est plus robuste face au problème d'observations aberrantes.

Chacune des $m(= p + q)$ composantes de l'échantillon aléatoire $X_1, X_2, \dots, X_n$ est ordonné pour obtenir les vecteurs de rangs $R_1, R_2, \dots, R_n$ avec

$$R_\alpha = \begin{pmatrix} R_\alpha^{(1)} \\ R_\alpha^{(2)} \end{pmatrix} = \begin{pmatrix} R_{1\alpha} \\ R_{2\alpha} \\ \vdots \\ R_{m\alpha} \end{pmatrix}$$

où $R_\alpha^{(1)} : p \times 1$ et $R_\alpha^{(2)} : q \times 1$ pour $\alpha = 1, 2, \dots, n$. La matrice de covariances basée sur ces vecteurs de rangs est

$$S_r = (S_{r,k\ell}) = \begin{pmatrix} S_{r(1,1)} & S_{r(1,2)} \\ S_{r(2,1)} & S_{r(2,2)} \end{pmatrix}$$

où pour $i = 1, 2$ et $j = 1, 2$, $S_{r(i,j)} = \frac{1}{n-1} \sum_{\alpha=1}^n (R_\alpha^{(i)} - \bar{R}^{(i)})(R_\alpha^{(j)} - \bar{R}^{(j)})'$ avec

pour $i = 1, 2$, $\bar{R}^{(i)} = \frac{1}{n} \sum_{\alpha=1}^n R_\alpha^{(i)} = \frac{n+1}{2} e^{(i)}$ où $e^{(i)} = (1, 1, \dots, 1)'$ est le vecteur unité de $\mathbb{R}^p$ ($i = 1$) ou $\mathbb{R}^q$ ($i = 2$).

Ainsi, on obtient

$$S_{r,k\ell} = \frac{1}{n-1} \sum_{\alpha=1}^n \left(R_{k\alpha} - \frac{n+1}{2} \right) \left(R_{\ell\alpha} - \frac{n+1}{2} \right)$$

pour $k = 1, 2, \dots, m$ et $\ell = 1, 2, \dots, m$. Finalement, la mesure de redondance RI basée sur les rangs est définie par

$$RI_r = \frac{\text{tr}(S_{r(1,2)} S_{r(2,2)}^{-1} S_{r(2,1)})}{\text{tr}(S_{r(1,1)})}. \quad (2.1)$$

Les propriétés de RI_r sont les suivantes :

- (i) $0 \leq RI_r \leq 1$,
- (ii) $RI_r = 0$ si et seulement si $S_{r(1,2)} = 0$,
- (iii) RI_r est invariante sous toutes les transformations qui préservent les rangs; en particulier, RI_r est invariante sous des changements d'échelles ou d'origines.

Si $p = q = 1$, les propriétés de RI_r sont les suivantes :

- (i) $RI_r = r_r^2$ le carré du coefficient de corrélation de Spearman,
- (ii) si $R_{1j} = R_{2j}$, $j = 1, 2, \dots, n$, presque sûrement, alors $RI_r = 1$ presque sûrement.

3. Une matrice d'association

Chaque $S_{r(k,\ell)}$ converge en probabilité lorsque $n \rightarrow \infty$ vers une «covariance grade» (voir par exemple Hoeffding (1948) ou Puri et Sen (1971), p. 125). Donc RI_r converge en probabilité, lorsque $n \rightarrow \infty$, vers un indice de redondance ρI_r défini à partir de la matrice limite de la même manière que RI_r est défini à partir de S_r . Dans cette section, nous obtenons cette matrice d'association limite. L'approche suivie est similaire à celle de Puri, Sen et Gokhale (1970).

Supposons que les observations $X_1, X_2, \dots, X_n$ sont indépendantes et de même fonction de loi $F(x)$ et notons par $F^{(1)}(x^{(1)})$ et $F^{(2)}(x^{(2)})$ les fonctions de loi marginales de $X_\alpha^{(1)}$ et $X_\alpha^{(2)}$ pour $\alpha = 1, 2, \dots, n$. Dans cet article, nous considérons l'hypothèse nulle H_0 d'indépendance entre $X^{(1)}$ et $X^{(2)}$, c'est-à-dire

$$H_0 : F(x) = F^{(1)}(x^{(1)})F^{(2)}(x^{(2)}).$$

Soit $F_{[i]}(x_{i\alpha})$ la fonction de loi marginale de $X_{i\alpha}$ pour $i = 1, 2, \dots, m$ et $\alpha = 1, 2, \dots, n$ et soit $F_{[i,j]}(x_{i\alpha}, x_{j\alpha})$ la fonction de loi marginale de $(X_{i\alpha}, X_{j\alpha})$ pour $i = 1, 2, \dots, m$, $j = 1, 2, \dots, m$, $i \neq j$ et $\alpha = 1, 2, \dots, n$.

Pour $i = 1, 2, \dots, m$, soit $J_{(i)}(u)$, $0 < u < 1$, une fonction telle que

$$\begin{aligned} \text{(i)} \quad & J_{(i)}(u) \text{ est absolument continue,} \\ \text{(ii)} \quad & \int_0^1 J_{(i)}(u) du = 0 \text{ et } \int_0^1 J_{(i)}^2(u) du = 1. \end{aligned} \quad (3.1)$$

Posons $Y_i = J_{(i)}(F_{[i]}(\nu_i))$ la fonction grade de ν_i , la i^{e} composante du vecteur X , pour $i = 1, 2, \dots, m$ et $Y = (Y_1, Y_2, \dots, Y_m)$. Finalement, définissons pour $i = 1, 2, \dots, m$ et $j = 1, 2, \dots, m$,

$$\theta_{ij} = E_{F_{[i,j]}}(Y_i Y_j) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} J_{(i)}(F_{[i]}(x)) J_{(j)}(F_{[j]}(y)) dF_{[i,j]}(x, y) \quad (3.2)$$

et posons

$$\Theta = (\theta_{ij}) = \begin{pmatrix} \theta_{(1,1)} & \theta_{(1,2)} \\ \theta_{(2,1)} & \theta_{(2,2)} \end{pmatrix}$$

et $\theta_{(1,2)} = \theta'_{(2,1)}$.

La matrice d'association non paramétrique recherchée est Θ (voir Puri, Sen et Gokhale (1970)). On peut maintenant estimer Θ par un estimateur convergent qui est fonction de la matrice de covariances S_r .

Soit $F_{n[i]}(x)$ la fonction de répartition expérimentale de $X_{i\alpha}$, $i = 1, 2, \dots, m$, c'est-à-dire

$$F_{n[i]}(x) = \frac{(\#X_{i\alpha} \leq x, \alpha = 1, 2, \dots, n)}{n}$$

où $\#$ signifie «le nombre de». D'une façon similaire, soit

$$F_{n[i,j]}(x, y) = \frac{(\#(X_{i\alpha}, Y_{j\alpha}) \leq (x, y), \alpha = 1, 2, \dots, n)}{n}$$

pour $i = 1, 2, \dots, m$, $j = 1, 2, \dots, m$, $i \neq j$, $\alpha = 1, 2, \dots, n$, où $(X_{i\alpha}, Y_{j\alpha}) \leq (x, y)$ signifie que $X_{i\alpha} \leq x$ et $Y_{j\alpha} \leq y$.

Pour $i = 1, 2, \dots, m$, soit $\{J_{n(i)}(u)\}_{n=1}^{\infty}$ une suite de fonctions telles que $J_{n(i)}(u)$ converge uniformément, lorsque $n \rightarrow \infty$, vers $J_{(i)}(u)$ pour $0 < u < 1$. Afin d'être conforme avec les conditions (3.1), on considère des fonctions $J_{n(i)}(\cdot)$ telles que

$$\sum_{j=1}^n J_{n(i)}\left(\frac{j}{n}\right) = 0 \text{ et } \frac{1}{n} \sum_{j=1}^n J_{n(i)}^2\left(\frac{j}{n}\right) = 1 \quad (3.3)$$

pour $i = 1, 2, \dots, m$.

Finalement, les éléments θ_{ij} de la matrice Θ peuvent être estimés par

$$\begin{aligned} T_{n,ij} &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} J_{n(i)}(F_{n[i]}(x)) J_{n(j)}(F_{n[j]}(y)) dF_{n[i,j]}(x, y) \\ &= \frac{1}{n} \sum_{\alpha=1}^n J_{n(i)}(F_{n[i]}(X_{i\alpha})) J_{n(j)}(F_{n[j]}(X_{j\alpha})). \end{aligned}$$

Nous écrivons

$$T_n = (T_{n,ij}) = \begin{bmatrix} T_{n(1,1)} & T_{n(1,2)} \\ T_{n(2,1)} & T_{n(2,2)} \end{bmatrix}$$

où T_n est $m \times m$, $T_{n(1,1)}$ est $p \times p$, $T_{n(2,2)}$ est $q \times q$, $T_{n(1,2)}$ est $p \times q$ et $T_{n(1,2)} = T'_{n(2,1)}$. La matrice T_n est clairement un estimateur convergent de la matrice Θ .

Dans le cas particulier où on choisit $J_{n(i)}\left(\frac{\alpha}{n}\right) = \left(\frac{12}{n^2-1}\right)^{1/2} \left(\alpha - \frac{n+1}{2}\right)$ pour $i = 1, 2, \dots, m$ et $\alpha = 1, 2, \dots, n$, les conditions (3.3) sont vérifiées. En plus, $J_{n(i)}\left(\frac{\alpha}{n}\right)$ converge uniformément, lorsque $n \rightarrow \infty$, vers $J_{(i)}(u) = \sqrt{12} \left(u - \frac{1}{2}\right)$, $0 < u < 1$, et cette fonction vérifie les conditions (3.1). Notons que pour i et $j = 1, 2, \dots, m$ et $\alpha = 1, 2, \dots, n$,

$$F_{n[i]}(X_{i\alpha}) = \frac{R_{i\alpha}}{n} \text{ et } F_{n[j]}(X_{j\alpha}) = \frac{R_{j\alpha}}{n}.$$

On trouve alors

$$T_{n,ij} = \frac{12}{n(n^2-1)} \sum_{\alpha=1}^n \left(R_{i\alpha} - \frac{n+1}{2}\right) \left(R_{j\alpha} - \frac{n+1}{2}\right) = \frac{12}{n(n+1)} S_{r(i,j)},$$

qui correspond au coefficient de corrélation des rangs de Spearman entre les composantes i et j du vecteur X , i et $j = 1, 2, \dots, m$. Par ailleurs, nous avons pour i et $j = 1, 2, \dots, m$,

$$\theta_{ij} = 12 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(F_{[i]}(x) - \frac{1}{2}\right) \left(F_{[j]}(y) - \frac{1}{2}\right) dF_{[i,j]}(x, y),$$

qui est le coefficient de corrélation grade entre les composantes i et j de X (voir Hoeffding (1948) ou Puri et Sen (1971), p. 125) et $\theta_{ij} = 0$ si les composantes i et j du vecteur X sont indépendantes. Finalement, puisque T_n converge en probabilité vers Θ lorsque $n \rightarrow \infty$, on en déduit que

$$RI_r = \frac{\text{tr}(T_{n(1,2)} T_{n(2,2)}^{-1} T_{n(2,1)})}{\text{tr}(T_{n(1,1)})}$$

converge en probabilité lorsque $n \rightarrow \infty$ vers

$$\rho I_r = \frac{\text{tr}(\theta_{(1,2)} \theta_{(2,2)}^{-1} \theta_{(2,1)})}{\text{tr}(\theta_{(1,1)})}.$$

4. La loi asymptotique de T_n

Avant d'énoncer le résultat principal, nous allons rappeler un théorème de Puri, Sen et Gokhale (1970). Nous avons besoin des conditions suivantes.

Condition 4.1 :

$\lim_{n \rightarrow \infty} J_{n(i)}(u) = J_{(i)}(u)$ existe pour $0 < u < 1$ et est non constante pour $i = 1, 2, \dots, m$. En plus, cette limite vérifie les conditions (3.1) imposées aux fonctions $J_{(i)}(u)$.

Condition 4.2 :

$$J_{n(i)}(1) = o(n^{1/4}), \quad i = 1, 2, \dots, m.$$

Condition 4.3 :

Pour $i, j = 1, 2, \dots, m$ et pour $A = I_{n(i)} \times I_{n(j)}$,

$$\int \int_A \{ J_{n(i)}(F_{n[i]}(x_i)) J_{n(j)}(F_{n[j]}(x_j)) - J_{(i)}(F_{[i]}(x_i)) J_{(j)}(F_{[j]}(x_j)) \} dF_{n[i,j]}(x_i, x_j) = O_p(n^{-1/2})$$

où $I_{n(k)} = \{x_k : 0 < F_{n[k]}(x_k) < 1, \quad k = 1, 2, \dots, m\}$.

Condition 4.4 :

Il existe une constante H telle que pour $i = 1, 2, \dots, m$,

$$\text{a) } |J_{(i)}(u)| \leq H(u(1-u))^{-\alpha} \text{ pour } 0 < \alpha < \frac{1}{8},$$

$$\text{b) } |J'_{(i)}(u)| \leq H(u(1-u))^{-1},$$

$$\text{c) } |J''_{(i)}(u)| \leq H(u(1-u))^{-2}$$

où $J'_{(i)}(\cdot)$ et $J''_{(i)}(\cdot)$ désignent respectivement les dérivées premières et secondes de $J_{(i)}(\cdot)$.

Le lemme suivant peut être trouvé dans Puri, Sen et Gokhale (1970). Dans ce lemme $uvec(T_n)$ désigne le vecteur $\frac{m(m-1)}{2} \times 1$ formé en empilant les colonnes de T_n et en omettant les éléments sur et sous la diagonale principale; $uvec(\Theta)$ est défini de la même façon.

Lemme 4.1. *Sous les conditions (4.1) à (4.4), nous avons*

$$\sqrt{n}(uvec(T_n) - uvec(\Theta)) \xrightarrow{L} N_k(0, \Phi),$$

lorsque $n \rightarrow \infty$, avec $k = \frac{m(m-1)}{2}$ et $\Phi = n(\phi_{n[(i,j),(r,s)]})$ où

$$n \phi_{[n(i,j),(r,s)]} = \begin{cases} \text{var} \sum_{\ell=1}^3 U_{(i,j);\ell}^{(\alpha)} & \text{si } (i,j) = (r,s), \\ \sum_{\ell=1}^3 \sum_{\ell'=1}^3 \text{cov} \left(U_{(i,j);\ell}^{(\alpha)}, U_{(r,s);\ell'}^{(\alpha)} \right) & \text{si } (i,j) \neq (r,s) \end{cases}$$

où

$$\begin{aligned} U_{(r,s);1}^{(\alpha)} &= J_{(r)} (F_{[r]}(X_{r\alpha})) J_{(s)} (F_{[s]}(X_{s\alpha})), \\ U_{(r,s);2}^{(\alpha)} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (F_{1X_{r\alpha}}(x_r) \\ &\quad - F_{[r]}(x_r)) J_{(s)} (F_{[s]}(x_s)) J'_{(r)} (F_{[r]}(x_r)) dF_{[r,s]}(x_r, x_s), \\ U_{(r,s);3}^{(\alpha)} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (F_{1X_{s\alpha}}(x_s) \\ &\quad - F_{[s]}(x_s)) J_{(r)} (F_{[r]}(x_r)) J'_{(s)} (F_{[s]}(x_s)) dF_{[r,s]}(x_r, x_s), \end{aligned}$$

avec

$$F_{1X_{s\alpha}}(u) = \begin{cases} 1 & \text{si } X_{s\alpha} \leq u, \\ 0 & \text{sinon.} \end{cases}$$

Lorsque $J_{n(i)} \left(\frac{\alpha}{n} \right)$ est choisi comme dans le cas particulier de la section 3, on peut montrer après quelques manipulations algébriques que

a) Les conditions (4.1) à (4.4) sont vérifiées.

b) Si $F_{[r,s]}(x_r, x_s) = F_{[r]}(x_r)F_{[s]}(x_s)$, alors

b.1) $E \left(U_{(r,s);\ell}^{(\alpha)} \right) = 0$ pour $\ell = 1, 2, 3$.

b.2) $U_{(r,s);2}^{(\alpha)}$ et $U_{(r,s);3}^{(\alpha)}$ sont des variables aléatoires dégénérées en 0 c'est-à-dire presque sûrement nulles.

b.3) Une sous-matrice particulière de Φ , notée Φ_{33} , est obtenue de la façon suivante :

b.3.1) Si $(i, j) = (r, s)$, il suit du lemme 4.1,

$$\begin{aligned} n \phi_{n[(i,j),(r,s)]} &= \text{var} \sum_{\ell=1}^3 U_{(i,j);\ell}^{(\alpha)} = \text{var} U_{(i,j);1}^{(\alpha)} \\ &= 144E \left(\left(F_{[i]}(X_{i\alpha}) - \frac{1}{2} \right) \left(F_{[j]}(X_{j\alpha}) - \frac{1}{2} \right) \right)^2 \\ &= 1. \end{aligned}$$

b.3.2) Si $(i, j) \neq (r, s)$ pour $1 \leq i, r \leq p$ et $p+1 \leq j, s \leq m$, et $F_{[i,j,r,s]}(x_i, x_j, x_r, x_s) = F_{[i,j]}(x_i, x_j)F_{[r,s]}(x_r, x_s)$, nous avons

$$\begin{aligned} n \phi_{n[(i,j),(r,s)]} &= \sum_{\ell=1}^3 \sum_{\ell'=1}^3 \text{cov} \left(U_{(i,j);\ell}^{(\alpha)}, U_{(r,s);\ell'}^{(\alpha)} \right) \\ &= \text{cov} \left(U_{(i,j);1}^{(\alpha)}, U_{(r,s);1}^{(\alpha)} \right) \\ &= 144E \left(\left(F_{[i]}(X_{i\alpha}) - \frac{1}{2} \right) \left(F_{[r]}(X_{r\alpha}) - \frac{1}{2} \right) \right) \\ &\quad E \left(\left(F_{[j]}(X_{j\alpha}) - \frac{1}{2} \right) \left(F_{[s]}(X_{s\alpha}) - \frac{1}{2} \right) \right) \\ &= \theta_{ir} \theta_{js}. \end{aligned}$$

La matrice Φ_{33} est $\theta_{(1,1)} \otimes \theta_{(2,2)}$ où $\otimes$ désigne le produit de Kronecker et les éléments de Φ_{33} sont définis pour $1 \leq i, r \leq p$ et $p+1 \leq j, s \leq m$; les éléments diagonaux de Φ_{33} sont égaux à 1 puisque $\theta_{ii} = 1$ pour tout i .

5. La loi asymptotique de RI_r sous H_0

Pour une matrice M de dimension $r \times s$, nous désignons par $\text{vec}(M)$ le vecteur $rs \times 1$ formé en empilant les colonnes de M . On utilisera le lemme suivant dont la démonstration est une conséquence immédiate des résultats b du lemme 4.1.

Lemme 5.1. *Sous les conditions (4.1) à (4.4) et sous l'hypothèse H_0 d'indépendance de $X^{(1)}$ et $X^{(2)}$, on a*

$$\sqrt{n} \text{vec}(T_{n(2,1)}) \xrightarrow{L} N_{pq}(0, \Phi_{33})$$

lorsque $n \rightarrow \infty$.

Rappelons que l'indépendance de $X^{(1)}$ et $X^{(2)}$ implique qu'au niveau de la population $\rho I_r = 0$. Maintenant, nous pouvons énoncer le théorème donnant la loi

asymptotique de RI_r sous H_0 . La démonstration suit la même démarche que dans Lazraq et Clérout (1992).

Théorème 5.2. *Sous l'hypothèse H_0 d'indépendance de $X^{(1)}$ et $X^{(2)}$, la loi asymptotique, lorsque $n \rightarrow \infty$, de $n RI_r$ est celle de*

$$\frac{1}{tr(\theta_{(1,1)})} \sum_{i=1}^p \lambda_i \sum_{j=1}^q U_{ij}^2$$

où $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ sont les valeurs propres de $\theta_{(1,1)}$ et les U_{ij} sont des variables aléatoires indépendantes et de même loi $N(0, 1)$.

Démonstration. On a $RI_r = \frac{tr(T_{n(1,2)} T_{n(2,2)}^{-1} T_{n(2,1)})}{tr(T_{n(1,1)})}$. Comme, lorsque $n \rightarrow \infty$, on a les convergences en probabilité $T_{n(1,1)} \xrightarrow{P} \theta_{(1,1)}$, et $T_{n(2,2)} \xrightarrow{P} \theta_{(2,2)}$ de même que la convergence en loi $T_{n(2,1)} \xrightarrow{L} W$ alors, par le théorème 4.4 de Billingsley (1968), on a $(T_{n(2,1)}, T_{n(2,2)}) \xrightarrow{L} (W, \theta_{(2,2)})$. Soit $Q(W, \theta_{(2,2)}) = tr(W' \theta_{(2,2)} W)$. Comme Q est continue en chaque élément de W et de $\theta_{(2,2)}$, le théorème 5.1 de Billingsley (1968) nous permet d'écrire $Q(T_{n(2,1)}, T_{n(2,2)}) \xrightarrow{L} Q(W, \theta_{(2,2)})$.

Par conséquent, $tr(T_{n(1,2)} T_{n(2,2)}^{-1} T_{n(2,1)})$ a la même loi asymptotique que $tr(T_{n(1,2)} \theta_{(2,2)}^{-1} T_{n(2,1)})$ et RI_r a la même loi asymptotique que

$RI_r^* = \frac{tr(T_{n(1,2)} \theta_{(2,2)}^{-1} T_{n(2,1)})}{tr(\theta_{(1,1)})}$. Considérons maintenant le numérateur de RI_r^* ou plus précisément, considérons

$$Q^* = n tr(T_{n(1,2)} \theta_{(2,2)}^{-1} T_{n(2,1)}) = n \text{vec}(T_{n(2,1)})' (I_p \otimes \theta_{(2,2)}^{-1}) \text{vec}(T_{n(2,1)}).$$

Du lemme 5.1 et sous l'hypothèse d'indépendance, nous avons, lorsque $n \rightarrow \infty$,

$$\sqrt{n} \text{vec}(T_{n(2,1)}) \xrightarrow{L} N_{pq}(0, \Phi_{33})$$

et par conséquent, la forme quadratique Q^* est asymptotiquement distribuée (voir Johnson et Kotz (1970), p. 150) comme $\sum_{i=1}^{pq} w_i U_i^2$ où les U_i sont des variables aléatoires indépendantes et de même loi $N(0, 1)$ et les w_i sont les valeurs propres de

$$\Phi_{33}(I_p \otimes \theta_{(2,2)}^{-1}) = (\theta_{(1,1)} \otimes \theta_{(2,2)})(I_p \otimes \theta_{(2,2)}^{-1}) = (\theta_{(1,1)} \otimes I_q).$$

Ainsi, les valeurs propres w_i , $i = 1, 2, \dots, pq$, correspondent avec une multiplicité q aux valeurs propres $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ de $\theta_{(1,1)}$. La démonstration du théorème est complète. $\square$

Corollaire 5.3. *Sous H_0 , on a que $E(n RI_r)$ converge vers q et $\text{var}(n RI_r)$ converge vers $2q \text{tr}(\theta_{(1,1)}^2) / \text{tr}^2(\theta_{(1,1)})$.*

Démonstration. Le résultat découle immédiatement du théorème 5.2. $\square$

6. La loi asymptotique de RI_r sous une série d'alternatives

Considérons la suite d'hypothèses alternatives $H_{1n} : \theta_{(2,1)} = \frac{A}{\sqrt{n}}$ où A est une matrice $q \times p$ fixe quelconque. Dans cette section, on obtient la loi asymptotique de $n RI_r$ sous cette suite d'hypothèses alternatives.

Théorème 6.1. *Sous H_{1n} , la loi asymptotique lorsque $n \rightarrow \infty$ de $n RI_r$ est celle de*

$$\frac{\sum_{i=1}^p \lambda_i \sum_{j=1}^q V_{ij}}{\text{tr}(\theta_{(1,1)})}$$

où $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ sont les valeurs propres de $\theta_{(1,1)}$ et V_{ij} sont des variables aléatoires indépendantes de loi khi-deux décentrée avec un degré de liberté et paramètre de décentralité δ_{ij} correspondant à $P'_{ij}(\theta_{(1,1)} \otimes \theta_{(2,2)})^{-1/2} \text{vec}(A)$, P_{ij} est le vecteur propre orthonormé correspondant pour $j = 1, 2, \dots, q$, à la valeur propre λ_i de multiplicité q de $(\theta_{(1,1)} \otimes I_q)$.

Démonstration. Sous H_{1n} , on déduit directement du lemme 5.2 de Puri, Sen et Gokhale (1970) que $\sqrt{n} \text{vec}(T_{n(2,1)}) \xrightarrow{L} N_{pq}(\text{vec}(A), \theta_{(1,1)} \otimes \theta_{(2,2)})$ lorsque $n \rightarrow \infty$. Maintenant, par le même argument qu'au théorème 5.2, $n RI_r$ a la même loi asymptotique que

$$\frac{n(\text{vec}'(T_{n(2,1)})(I_p \otimes \theta_{(2,2)}^{-1})(\text{vec}(T_{n(2,1)}))}{\text{tr}(\theta_{(1,1)})}.$$

Or, en utilisant Johnson et Kotz (1970) p. 150, on déduit que la loi asymptotique du numérateur est celle de

$$\sum_{i=1}^p \lambda_i \sum_{j=1}^q \chi_1^2(\delta_{ij})$$

où $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ sont les valeurs propres de $\theta_{(1,1)}$ et $\chi_1^2(\delta_{ij})$ sont des variables aléatoires indépendantes de loi khi-deux décentrée avec un degré de liberté et paramètre de décentralité donné dans l'énoncé du théorème. $\square$

7. Test non paramétrique de l'absence de liaison entre deux vecteurs

Les résultats des sections précédentes peuvent être utilisés pour construire un test non paramétrique de l'absence de liaison entre deux vecteurs aléatoires. Nous testons

$$H_0 : \rho I_r = 0 \quad \text{contre} \quad H_1 : \rho I_r > 0$$

au niveau asymptotique α en rejetant H_0 si

$$n RI_r > c_\alpha$$

où c_α est le $100(1 - \alpha)^e$ centile de la loi de la statistique

$$Q = \frac{1}{tr(\theta_{(1,1)})} \sum_{i=1}^p \lambda_i \sum_{j=1}^q U_{ij}^2$$

du théorème 5.2. Sous H_1 , RI_r converge en probabilité vers $\rho I_r > 0$ c'est-à-dire que la puissance asymptotique de ce test converge vers 1 lorsque $n \rightarrow \infty$. Le test est donc qualifié de convergent.

Plusieurs algorithmes sont disponibles pour calculer c_α . L'algorithme d'Imhof (1961), dont un code informatique se trouve dans Koerts et Abrahamse (1969), peut être considéré comme exact et c'est celui que nous recommandons. En pratique, les paramètres de la loi sont inconnus et doivent être remplacés par des estimateurs convergents. Dans une étude détaillée faite dans Roy et Clérout (1993), on montre la précision des valeurs critiques obtenues en remplaçant les paramètres inconnus par des estimateurs convergents.

Le test asymptotique est proposé par l'algorithme suivant :

Entrée :

$$\begin{pmatrix} X_1^{(1)} \\ X_1^{(2)} \end{pmatrix}, \begin{pmatrix} X_2^{(1)} \\ X_2^{(2)} \end{pmatrix}, \dots, \begin{pmatrix} X_n^{(1)} \\ X_n^{(2)} \end{pmatrix}$$

où $X_i^{(1)} : p \times 1$ et $X_i^{(2)} : q \times 1$ pour $i = 1, \dots, n$.

Etape 1 :

Ordonner les vecteurs par composantes et remplacer chaque observation par son rang pour obtenir

$$R_1 = \begin{pmatrix} R_1^{(1)} \\ R_1^{(2)} \end{pmatrix}, R_2 = \begin{pmatrix} R_2^{(1)} \\ R_2^{(2)} \end{pmatrix}, \dots, R_n = \begin{pmatrix} R_n^{(1)} \\ R_n^{(2)} \end{pmatrix}.$$

Etape 2 :

Calculer $S_r = \frac{1}{n-1} \sum_{\alpha=1}^n \left(R_\alpha - \frac{n+1}{2} e \right) \left(R_\alpha - \frac{n+1}{2} e \right)'$ et partitionner S_r en

$$\begin{pmatrix} S_{r(1,1)} & S_{r(1,2)} \\ S_{r(2,1)} & S_{r(2,2)} \end{pmatrix}$$

où $S_{r(1,1)} : p \times p$ et $S_{r(2,2)} : q \times q$, e étant le vecteur $(p+q) \times 1$ dont toutes les composantes valent 1.

Etape 3 :

Calculer

$$RI_r = \frac{\text{tr}(S_{r(1,2)} S_{r(2,2)}^{-1} S_{r(2,1)})}{\text{tr}(S_{r(1,1)})}$$

et $V = n RI_r$.

Etape 4 :

Calculer

$$T_n = \begin{pmatrix} T_{n(1,1)} & T_{n(1,2)} \\ T_{n(2,1)} & T_{n(2,2)} \end{pmatrix} = \frac{12}{n(n+1)} \begin{pmatrix} S_{r(1,1)} & S_{r(1,2)} \\ S_{r(2,1)} & S_{r(2,2)} \end{pmatrix}$$

et obtenir les valeurs propres $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_p$ de $T_{n(1,1)}$.

Etape 5 :

Utiliser l'algorithme d'Imhof pour obtenir le 95^{ième} centile, $Q_{0,95}$, de la loi de

$$Q = \frac{1}{\text{tr}(T_{n(1,1)})} \sum_{i=1}^p \hat{\lambda}_i \sum_{j=1}^q U_{ij}^2.$$

Etape 6 :

Rejeter H_0 au niveau 0,05 si $V > Q_{0,95}$.

Sortie :

RI_r , V et $Q_{0,95}$.

8. Exemple

Considérons les données de Werner, Tolls et Mellecker (1970) que l'on peut aussi trouver dans le logiciel statistique BMDP (1981), p. 38. Les données représentent 8 variables mesurées sur 188 patients : âge, taille, poids, prise de contraceptif, cholestérol, albumine, calcium et acide urique. La variable prise de contraceptif étant

binaire ainsi que 7 vecteurs avec des données manquantes ne seront pas considérés. On se retrouve ainsi avec 181 vecteurs de taille 7. Soit $X^{(1)}$ le vecteur obtenu à partir des variables physiques (les trois premières variables) et $X^{(2)}$ le vecteur formé des variables biochimiques (les quatre dernières variables). On teste l'absence de liaison entre $X^{(1)}$ et $X^{(2)}$ par la procédure non paramétrique de la section 7. Nous avons $n = 181$, $p = 3$ et $q = 4$. On obtient $RI_r = 0,136$, $V = n RI_r = 24,624$ et l'algorithme d'Imhof donne $Q_{0,95} = 7,378$. Ainsi H_0 est rejetée au niveau $\alpha = 0,05$.

9. Étude expérimentale

Afin de mesurer le comportement et la performance du test basé sur RI_r , une étude de Monte Carlo est conduite pour comparer son niveau expérimental et sa puissance expérimentale avec deux autres tests compétiteurs.

Le test basé sur RI_r décrit dans la section 7 est noté par TR . Le test basé sur la mesure RI de Stewart et Love (1968) et présenté pour les lois elliptiques par Lazraq et Cléroux (1992) consiste à rejeter H_0 au niveau α si $n RI > d_\alpha$ où d_α est le $100(1 - \alpha)^e$ centile de la loi de

$$G = \frac{1 + K}{tr(\Sigma_{11})} \sum_{i=1}^p \lambda_i \sum_{j=1}^q Z_{ij}^2,$$

la quantité K est le paramètre d'aplatissement de la loi elliptique de base, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ sont les valeurs propres des Σ_{11} et les Z_{ij} sont des variables aléatoires indépendantes et de même loi $N(0, 1)$. L'algorithme d'Imhof est aussi utilisé pour obtenir la valeur critique d_α après avoir remplacé Σ_{11} et K par des estimateurs convergents S_{11} et $\hat{K}$. Ce test paramétrique est noté par TP . L'autre test compétiteur utilisé est le test non paramétrique basé sur $-n \ell n(H)$ où

$$H = \frac{|T_n|}{|T_{n(1,1)}| |T_{n(2,2)}|}.$$

Ce test est décrit dans Puri, Sen et Gokhale (1970) et il consiste à rejeter H_0 si $-n \ell n(H) > \chi_{pq;\alpha}^2$ où $\chi_{pq;\alpha}^2$ est le $100(1 - \alpha)^e$ centile d'une loi khi-deux avec pq degrés de liberté. Ce test est noté par TN .

Cette étude comparative dépend d'un certain nombre de paramètres qui sont choisis de la façon suivante :

- i) $p = 2, 3$ et $q = 2, 3$;
- ii) $\alpha = 1\%, 2, 5\%, 5\%$ et 10% ;
- iii) $n = 30, 50, 75, 100, 200, 300$;
- iv) le nombre d'échantillons générés à chaque fois est de 1000;
- v) deux types de lois sont considérées :

- a) la loi multinormale $N_{p+q}(0, \Sigma)$ où $\Sigma_{11} = I_p, \Sigma_{22} = I_q$ et $\Sigma_{12} = \Sigma'_{21} = C_{00}, C_{10}, C_{15}, C_{20}$ et C_{25} où la matrice C_{xy} représente la matrice $p \times q$ avec tous ses éléments égaux à la valeur réelle $0, xy$.
- b) une loi multivariée générale obtenue à partir d'un vecteur Y centré et de matrice de covariances unité. Les composantes du vecteur Y sont indépendantes et respectivement une loi normale $N(0, 1)$, une loi exponentielle (avec paramètre 1) moins 1, une loi gamma (avec paramètres 1 et 1) moins 1, une loi bêta (avec paramètres 1 et 1) moins 0, 5 et multipliée par $\sqrt{12}$, une loi logistique multipliée par $\sqrt{3}/\pi$ et une loi khi-deux avec un degré de liberté moins 1 puis standardisée par $\sqrt{2}$. Le vecteur X est obtenu par la transformation $X = CY$ où C est une matrice telle que $\Sigma = CC'$. Ainsi, nous obtenons une loi générale multivariée avec une structure de covariances Σ . Le choix de Σ est fait de la même façon qu'en a). Selon les valeurs de p et q , on prend pour $X^{(1)}$ les p premières composantes de X et pour $X^{(2)}$ les q suivantes, $p + q \leq 6$.

Les résultats de la simulation selon a) et b) sont présentés respectivement dans les tableaux 1 et 2.

En général, la puissance expérimentale de chaque test considéré augmente lorsque l'alternative s'éloigne de l'hypothèse nulle c'est-à-dire lorsque la valeur xy de C_{xy} augmente. Les niveaux expérimentaux sont obtenus en considérant l'alternative avec C_{00} et en général, le niveau expérimental des tests TP et TR sont conservateurs c'est-à-dire inférieur au niveau α fixé tandis que le niveau expérimental du test TN est libéral c'est-à-dire supérieur au niveau α fixé.

Pour la loi multinormale, la puissance expérimentale du test TP est supérieure à celle du test TR qui à son tour est supérieure à celle du test TN si $xy \geq 15$ ou $n \geq 75$. Cette supériorité du test TP n'est pas surprenante puisque justement, toute la théorie distributionnelle pour ce test repose sur le prérequis de la multinormalité ou de l'ellipticité qui est vérifié dans cette étude (Tableau 1).

Pour la loi multivariée générale, la puissance expérimentale du test TR est supérieure à la puissance expérimentale du test TN qui est supérieure à celle du test TP . Le résultat repose sur le caractère non paramétrique des deux tests.

Références

- BILLINGSLEY P. (1968). Convergence of probability measure. John Wiley and Sons, New York.
- BMDP Statistical Software, W.J. Dixon, Chief Editor (1981). University of California Press, Berkeley.
- CLÉROUX R., LAZRAQ A., LEPAGE Y. (1993). Vector correlation based on ranks and a nonparametric test of independence between vectors. Rapport de recherche, département de mathématiques et de statistique, Université de Montréal.
- CONOVER W.J., IMAN R.L. (1980). The rank transformation as a method of discrimination with some examples. Comm. in Statist., A9(5), 465-487.

TABLEAU 1
Puissances expérimentales ($\times 1000$) pour une loi multinormale

n	Test	$p = q = 2; \alpha = 5\%$					$p = 2, q = 3; \alpha = 1\%$					$p = 3, q = 2; \alpha = 2, 5\%$					$p = q = 3; \alpha = 10\%$				
		C_{00}	C_{10}	C_{15}	C_{20}	C_{25}	C_{00}	C_{10}	C_{15}	C_{20}	C_{25}	C_{00}	C_{10}	C_{15}	C_{20}	C_{25}	C_{00}	C_{10}	C_{15}	C_{20}	C_{25}
30	TP	055	106	180	321	424	005	028	056	120	200	028	075	129	246	404	127	227	365	529	684
	TN	073	138	199	323	412	028	055	106	177	256	059	096	149	246	363	201	292	429	532	632
	TR	051	099	158	288	371	008	021	047	091	156	023	056	114	210	362	103	210	338	478	621
50	TP	040	154	286	474	706	007	047	121	300	469	028	103	273	506	708	092	312	551	781	912
	TN	052	167	278	437	640	017	057	127	280	455	042	112	220	414	587	130	327	514	725	856
	TR	039	144	272	428	652	009	033	097	258	419	023	089	245	449	649	098	281	511	733	874
75	TP	043	224	408	694	880	013	083	229	505	790	024	175	440	712	897	093	427	723	921	986
	TN	052	232	369	628	815	017	092	198	463	723	027	169	353	575	821	115	405	644	866	961
	TR	046	209	375	655	838	009	080	188	451	731	016	164	386	637	863	092	393	667	896	978
100	TP	054	308	573	814	951	008	140	396	710	911	030	267	595	856	974	114	531	865	983	999
	TN	057	274	496	760	927	011	117	347	640	859	035	230	475	744	922	128	485	804	952	993
	TR	047	272	519	777	939	006	113	351	654	867	026	239	540	815	951	116	494	832	973	997
200	TP	041	535	899	990	998	014	407	827	979	999	033	561	934	999	1000	111	849	989	999	1000
	TN	050	464	844	975	996	019	343	755	967	998	029	455	867	984	999	106	807	975	999	1000
	TR	048	490	865	983	998	018	360	787	971	998	029	510	911	995	1000	095	828	985	999	1000
300	TP	044	747	983	1000	1000	007	639	970	1000	1000	025	770	992	1000	1000	088	964	1000	1000	1000
	TN	052	695	958	998	1000	007	562	944	999	1000	032	684	975	998	1000	096	934	998	1000	1000
	TR	042	719	965	1000	1000	007	584	958	1000	1000	031	734	985	1000	1000	093	947	999	1000	1000

TABLEAU 2
Puissances expérimentales ($\times 1000$) pour une loi générale multivariée avec une structure de covariance donnée

n	Test	$p = q = 2; \alpha = 5\%$					$p = 2, q = 3; \alpha = 1\%$					$p = 3, q = 2; \alpha = 2, 5\%$					$p = q = 3; \alpha = 10\%$				
		C_{00}	C_{10}	C_{15}	C_{20}	C_{25}	C_{00}	C_{10}	C_{15}	C_{20}	C_{25}	C_{00}	C_{10}	C_{15}	C_{20}	C_{25}	C_{00}	C_{10}	C_{15}	C_{20}	C_{25}
30	TP	053	108	204	337	430	017	031	060	117	242	027	081	173	257	424	118	245	396	553	735
	TN	086	169	292	387	510	028	071	124	208	360	074	103	170	208	360	187	398	513	662	776
	TR	052	122	255	373	502	011	034	073	133	294	028	068	129	222	374	102	299	481	638	783
50	TP	070	162	313	523	693	013	059	157	319	498	038	118	282	503	686	094	326	553	752	913
	TN	083	217	404	597	750	021	087	215	374	557	045	114	209	391	518	153	473	670	801	918
	TR	069	195	401	613	772	013	065	200	349	560	030	090	230	454	610	102	476	703	836	943
75	TP	062	237	452	712	880	013	085	269	548	772	021	186	450	726	904	105	446	717	909	977
	TN	065	320	548	781	908	010	127	343	615	817	022	143	331	563	758	129	603	812	937	981
	TR	053	327	567	804	934	007	122	364	630	847	018	154	393	677	855	100	632	862	963	992
100	TP	050	291	611	823	955	017	157	408	707	913	026	269	577	867	976	108	564	843	977	1000
	TN	057	393	732	900	969	015	208	508	776	946	032	202	447	735	903	124	716	911	991	1000
	TR	048	411	761	925	977	009	217	543	813	965	024	230	522	829	954	104	750	940	999	1000
200	TP	059	559	868	986	999	010	402	831	984	1000	023	559	928	997	1000	093	835	999	1000	1000
	TN	054	685	962	996	999	007	526	901	992	1000	032	401	800	974	1000	101	951	1000	1000	1000
	TR	052	712	969	999	1000	006	561	933	997	1000	026	493	894	990	1000	085	969	1000	1000	1000
300	TP	052	716	984	999	1000	005	636	974	1000	1000	034	793	991	1000	1000	108	950	1000	1000	1000
	TN	046	856	994	999	1000	012	759	988	1000	1000	027	632	964	998	1000	091	995	1000	1000	1000
	TR	045	886	997	1000	1000	009	796	995	1000	1000	022	719	987	1000	1000	081	999	1000	1000	1000

- CONOVER W.J., IMAN R.L. (1981). Rank transformation as a bridge between parametric and non parametric statistics. *The American Statistician*, 35, 124-133.
- COXHEAD P. (1974). Mesuring the relationships between two sets of variables. *Brit. Jour. Math. Stat. Psycho.*, 27, 205-212.
- CRAMER E.M. (1974). A generalization of vector correlation and its relation to canonical correlation. *Mult. Behav. Res.*, 9, 347-352.
- CRAMER E.M., NICEWANDER W.A. (1979). Some symmetric invariant measures of multivariate association. *Psychometrika*, 44, 43-54.
- ESCOUFIER Y. (1973). Le traitement des variables vectorielles. *Biometrics*, 29, 751-760.
- GLEASON T.C. (1976). On Redundancy in Canonical Analysis. *Psychological Bulletin*, 83, 1004-1006.
- HOEFFDING W. (1948). A class of statistics with asymptotically normal variables. *Ann. Math. Statist.*, 19, 293-325.
- HOTELLING H. (1936). Relations between two sets of variables. *Biometrika*, 28, 321-377.
- IMHOF P. (1961). Computing the loi of quadratic forms in normal variates. *Biometrika*, 48, 419-426.
- JOHNSON N.L., KOTZ S. (1970). Continuous univariate loi -2. Houghton Mifflin, Boston.
- KOERTS J., ABRAHAMSE A.P.J. (1969). On the theory and application of the general linear model. Rotterdam University Press, Rotterdam.
- KSHIRSAGAR A.M. (1969). Correlation between two vector variables. *JRSS, B*, 31, 477-485.
- LAZRAQ A., CLÉROUX R. (1988). Un algorithme pas à pas de sélection de variables en régression linéaire multivariée. *Stat. Anal. Don.*, 13, 15-18.
- LAZRAQ A., CLÉROUX R., KIERS H.A.L. (1992). Mesures de liaison vectorielle et généralisation de l'analyse canonique. *Rev. Statistique Appliquée*, XXXX(1), 23-35.
- LAZRAQ A., CLÉROUX R. (1992). Tests d'homogénéité entre indices de redondance pour des lois elliptiques. *Rev. Statistique Appliquée*, XXXX(3), 19-33.
- LINGOES J.C., SCHÖNEMANN P.H. (1974). Alternative Measures of Fit for the Schönemann-Carroll Matrix Fitting Algorithm. *Psychometrika*, 39, 423-427.
- MASUYAMA M. (1939). Correlation between tensor quantities. *Proc. Physico-Math. Soc. Japan, Series 3*, 31, 638-647.
- MASUYAMA M. (1941). Correlation between two sets of complex vectors. *Proc. Physico-Math. Soc. Japan, Series 3*, 33, 918-924.
- PURI M.L., SEN P.K. (1971). Nonparametric methods in multivariate analysis. John Wiley and Sons, New York.

- PURI M.L., SEN P.K., GOKHALE D.V. (1970). On a class of rank order tests of independence. *Sankhya*, A, 271-298.
- RAMSAY J.O., ten Berge, J., STYAN G. (1984). Matrix correlation. *Psychometrika*, 49, 403-423.
- ROBERT P., ESCOUFIER Y. (1976). A unifying tool for linear multivariate statistical methods : the RV-coefficient. *Appl. Statist.*, 25, 257-265.
- ROSEBOOM N.W. (1965). Linear correlation between sets of variables. *Psychometrika*, 30, 57-71.
- ROY R., CLÉROUX R. (1993). On vector cross-correlation in time series and application. To appear in the *International Statistical Review*, Vol. 61, No. 3.
- SHAFFER J.P., GILLO M.W. (1974). A multivariate extension of the correlation ratio. *Educ. Psycho. Meas.*, 34, 521-524.
- STEPHENS M. (1979). Vector correlation. *Biometrika*, 66, 41-98.
- STEWART D., LOVE W. (1968). A general canonical correlation index. *Psycho. Bull.*, 70, 160-163.
- WERNER M., TOLLS M.E., HULTIN J.V., MELLECKER J. (1970). Sex and Age Dependence of Serum Calcium, Inorganic Phosphorus, Total Protein and Albumin in a Large Ambulatory Population. *Fifth International Congress on Automation, Advances in Automated Analysis*, 2, 59-65.