REVUE DE STATISTIQUE APPLIQUÉE

T. FOUCART

Transitivité du produit scalaire

Revue de statistique appliquée, tome 39, nº 3 (1991), p. 57-68
<http://www.numdam.org/item?id=RSA_1991__39_3_57_0>

Rev. Statistique Appliquée, 1991, **XXXIX** (3), 57-68

TRANSITIVITÉ DU PRODUIT SCALAIRE

T. FOUCART

Département de Mathématiques
Université d'Orléans

RÉSUMÉ

"Notons pour finir que la corrélation n'est pas transitive : **x** très corrélé avec **y**, **y** très corrélé avec **z**, n'implique nullement que **x** soit corrélé avec **z**." : cette phrase, extraite du livre de SAPORTA (1990, page 134), figure à quelques mots près dans d'autres ouvrages généraux de Statistique (CAILLIEZ et PAGES, 1976 page 177 par exemple).

L'objectif de cet article est de préciser cette notion de transitivité entre deux coefficients de corrélation.

Mots-clés : *produit scalaire, inégalité de Schwarz, matrice symétrique définie positive, corrélation, transitivité.*

ABSTRACT

"Let's note that correlation is not transitive : **x** very correlated to **y**, **y** very correlated to **z** does not imply that **x** is correlated to **z**." : this sentence, read from the book by SAPORTA (1990, page 134), can be found in about same words in other general books of Statistics (CAILLIEZ ET PAGES, 1976, page 177 for instance).

The aim of this paper is to give precisions about this notion of transitivity between two correlation coefficients.

Key-words : *dot product, Schwarz's inequality, positive semi- definite matrix, correlation, transitivity.*

1. Enoncé du problème

Soit R une matrice symétrique définie par les produits scalaires entre n vecteurs e_i d'un espace euclidien de dimension n que nous supposerons égal à $\mathbb{R}^n$.

Dans le cas général, la matrice symétrique R est semi-définie positive : elle possède n vecteurs propres orthogonaux, $n-k$ valeurs propres strictement positives et k valeurs propres nulles. Le système $(e_i)_{i=1,n}$ n'est pas une base de $\mathbb{R}^n$: il est de rang $n-k$.

Une condition nécessaire et suffisante pour que le système $(e_i)_{i=1,n}$ constitue une base de $\mathbb{R}^n$ est que toutes les valeurs propres de la matrice R soient strictement positives. La matrice est alors de rang n; par définition, elle est définie positive (on emploie parfois le terme non dégénérée pour définie). Elle possède la propriété supplémentaire d'être inversible.

On choisit un produit scalaire $e_i.e_j$ (égal aux termes $r_{i,j}$ et $r_{j,i}$), et l'on cherche les valeurs r qu'il peut prendre en laissant la matrice définie positive, les autres termes restant constants. On suppose $i \neq j$.

Il s'agit donc de généraliser l'inégalité de Schwarz qui est le cas particulier du problème précédent pour une matrice de taille 2, et qui s'exprime :

– par les produits scalaires (ou covariances, en probabilité) :

$$|e_i.e_j| \leq \| e_i \| \| e_j \|$$

– par les cosinus (ou coefficients de corrélation) :

$$|e_i.e_j| / \| e_i \| \| e_j \| \leq 1$$

On choisit ensuite un couple de produits scalaires $r_{i,j}$ et $r_{k,l}$ ($i \neq j$ et $k \neq l$) : à chaque valeur r' de $r_{i,j}$, on peut associer l'ensemble des valeurs r'' que peut prendre $r_{k,l}$ en laissant la matrice définie positive : on définit ainsi une relation de transitivité.

2. Propriétés utilisées (CIARLET, 1988)

2.1 Toute matrice déduite de R par une même permutation des lignes et des colonnes est définie positive.

Cette propriété découle du fait que le rang d'un système est indépendant de l'ordre de ses vecteurs. Elle est utilisée pour placer le terme étudié en dernière colonne et avant-dernière ligne : on n'étudiera donc par la suite que le terme $r_{n-1,n}$ les autres termes de la matrice (y compris $r_{n,n}$) étant fixés.

2.2 Toute matrice définie positive est le produit d'une matrice triangulaire inférieure B par sa transposée B^t (décomposition de Choleski) et les termes diagonaux de B sont strictement positifs.

Les termes de B sont définis de la façon suivante :

$$\forall i = 1, n \quad b_{i,1} = r_{1,i} / \sqrt{r_{1,1}}$$

$$\forall i = 2, n \quad b_{i,i} = [r_{i,i} - \sum_{k=1}^{i-1} b_{i,k}^2]^{1/2}$$

$$\forall j = i+1, n \quad b_{j,i} = \frac{r_{i,j} - \sum_{k=1}^{i-1} b_{i,k} b_{j,k}}{b_{i,i}}$$

3. Méthode

3.1 Etude d'un produit scalaire

Pour un terme fixé $b_{j,i}(i \leq j)$, les formules précédentes ne font intervenir que les colonnes de B de rang k strictement inférieur à i (les $b_{i,k}$), les $i-1$ termes précédents de la ligne de rang j (les $b_{j,k}$), et les termes $b_{i,i}$ et $r_{i,j}$.

L'algorithme consiste à calculer les termes $b_{j,i}$ en faisant varier i de 1 à n et, pour chaque i fixé, j de i à n. Par exemple, dans le cas d'une matrice 4×4, le calcul s'effectue dans l'ordre suivant : $b_{1,1}$, $b_{2,1}$, $b_{3,1}$, $b_{4,1}$ (première colonne), $b_{2,2}$, $b_{3,2}$, $b_{4,2}$ (deuxième colonne), $b_{3,3}$, $b_{4,3}$ (troisième colonne), $b_{4,4}$ (dernier terme).

On en déduit que le terme $r_{n-1,n}$ n'intervient que dans le calcul de $b_{n,n-1}$ et de $b_{n,n}$.

Or, la matrice initiale étant définie positive, la matrice R_{n-1} constituée des $n-1$ premières lignes et premières colonnes est définie positive : tous les calculs antérieurs à celui de $b_{n,n-1}$ sont possibles et le terme $b_{n-1,n-1}$ est différent de 0.

Le terme $b_{n,n-1}$ étant alors défini quel que soit $r_{n-1,n}$, la matrice est donc définie positive si et seulement si $b_{n,n}$ existe et est strictement positif. On a :

$$b_{n,n} = [r_{n,n} - \sum_{k=1}^{n-1} b_{n,k}^2]^{1/2} \tag{0}$$

On en déduit la condition nécessaire et suffisante suivante :

$$r_{n,n} - \sum_{k=1}^{n-1} b_{n,k}^2 > 0$$

Dans l'expression ci-dessus, seul le terme $b_{n,n-1}$ dépend de $r_{n-1,n}$ et l'on a :

$$b_{n,n-1} = \frac{r_{n-1,n} - \sum_{k=1}^{n-2} b_{n-1,k} b_{n,k}}{b_{n-1,n-1}}$$

On remplace $b_{n,n-1}$ par son expression dans l'équation (0); la matrice R est donc définie positive si et seulement si $r_{n-1,n}$ appartient à l'intervalle $I =]a, b[$ où :

$$a = -b_{n-1,n-1}[r_{n,n} - \sum_{k=1}^{n-2} b_{n,k}^2]^{1/2} + \sum_{k=1}^{n-2} b_{n-1,k} b_{n,k} \tag{1}$$

$$b = b_{n-1,n-1}[r_{n,n} - \sum_{k=1}^{n-2} b_{n,k}^2]^{1/2} + \sum_{k=1}^{n-2} b_{n-1,k} b_{n,k} \tag{2}$$

On notera que les bornes a et b sont des fonctions continues des $b_{n,k}$, eux-mêmes fonctions continues des $r_{i,j}$ (dans la mesure où ces fonctions sont définies).

Lorsque le terme $b_{n,n}$ est nul, le déterminant de B est nul et donc celui de R : le système $(e_i)_{i=1,n}$ est lié alors que le système $(e_i)_{i=1,n-1}$ est libre. La matrice R possède une valeur propre nulle et une seule.

Supposons maintenant la matrice initiale semi-définie positive. Deux cas se présentent :

– le système $(e_i)_{i=1,n-1}$ est libre : on peut alors calculer les bornes de l'intervalle précédent.

– le système $(e_i)_{i=1,n-1}$ est lié : la matrice R_{n-1} n'est pas définie positive, ni, par suite, la matrice R quel que soit le terme $r_{n-1,n}$.

Définition : nous appelons intervalle de variation I_v du coefficient $r_{n-1,n}$ d'une matrice semi-définie positive l'intervalle $]a, b[$ s'il existe, où a et b sont définis par les formules (1) et (2).

Par commodité de langage, nous dirons lorsque la matrice n'est pas définie positive que l'intervalle de variation est égal à l'intervalle $]a, b[$, avec $a = b$.

L'intervalle de variation I_v est inclus dans l'intervalle I_s obtenu par l'inégalité de Schwarz appliquée à la matrice $R_{n-1,n}$ définie par les deux dernières lignes et les deux dernières colonnes :

$$I_s = \left]-[r_{n,n}r_{n-1,n-1}]^{1/2},\ [r_{n,n}r_{n-1,n-1}]^{1/2}\right[$$

En effet, la matrice R est définie positive pour $r_{n-1,n}$ appartenant à l'intervalle I_v. Cela implique que la matrice $R_{n-1,n}$ l'est aussi, et donc que :

$$I_v \subset I_s$$

Définition : nous appelons coefficient de contrainte l'indice c_k défini par :

$$c_k = 1 - (b - a)/(2[r_{n,n}r_{n-1,n-1}]^{1/2})$$

Ce coefficient prend la valeur 1 pour $a = b$: l'intervalle de variation est vide et la contrainte est maximale, et la valeur 0 lorsque $I_v = I_s$: les termes des lignes et des colonnes de rang inférieur à $n - 1$ n'imposent aucune contrainte supplémentaire à l'inégalité de Schwarz.

Plus généralement, l'intervalle de variation est inclus dans les intervalles de variation calculés sur les sous-matrices de R contenant la ligne et la colonne du coefficient considéré, qui sont nécessairement définies positives. Ainsi, en supprimant la ligne i et la colonne i de la matrice pour $i = 1, ..., n - 2$, on obtient $n - 2$ intervalles de variation $]a(i), b(i)[$ contenant tous l'intervalle $]a, b[$ et inclus dans I_s. Les coefficients de contrainte correspondants $c_k(i)$ sont compris entre 0 et c_k. On peut facilement généraliser le processus.

3.2 Etude d'un couple de produits scalaires

La méthode précédente est limitée à l'étude des variations d'un seul produit scalaire, les autres restant constants. On considère maintenant le cas de deux produits scalaires.

L'inégalité de Schwarz nous donne un intervalle $I_{s'}$ contenant des valeurs r' du premier terme $r_{i,j}$ à l'extérieur duquel la matrice n'est pas définie positive.

En chaque point de $I_{s'}$, on peut alors calculer par la méthode précédente un intervalle de variation $I_{v''}(r')$ (éventuellement vide) contenant les valeurs r'' du second terme $r_{k,l}$ telles que la matrice soit définie positive.

On obtient ainsi l'ensemble E_{ijkl} des couples (r', r'') tels que la matrice soit définie positive pour $r_{i,j} = r'$ et $r_{k,l=r''}$, que l'on appelle ensemble de variation du couple $(r_{i,j},\ r_{k,l})$.

L'ensemble E_{ijkl} est évidemment inclus dans l'ensemble produit $I_{s'} \times I_{s''}$ déduit de l'inégalité de Schwarz. Par construction, il est mesurable.

On peut définir les ensembles $E_{r'}$ et $E_{r''}$ de la façon suivante :

$E_{r'} = \{r' ;$ il existe r'' tel que R soit définie positive$\}$

$E_{r''} = \{r'' ;$ il existe r' tel que R soit définie positive$\}$

Les ensembles $E_{r'}$ et $E_{r''}$ contiennent les intervalles de variation $I_{r'}$ et $I_{r''}$ définis précédemment.

On peut en déduire deux paramètres :

– un coefficient de contrainte :

$$c_k = 1 - (m_{r'} m_{r''})/(m_{s'} m_{s''})$$

où $m_{s'}$ et $m_{s''}$ sont les longueurs des intervalles $I_{s'}$ et $I_{s''}$ et $m_{r'}$ et $m_{r''}$ les mesures des ensembles $E_{r'}$, $E_{r''}$ (les longueurs si ce sont des intervalles).

Le coefficient c_k est d'autant plus proche de 1 que les contraintes supplémentaires réduisent les ensembles $E_{r'}$ et $E_{r''}$ par rapport aux intervalles obtenus par l'inégalité de Schwarz, et d'autant plus proche de 0 que l'ensemble $E_{r'} \times E_{r''}$ est voisin de $I_{s'} \times I_{s''}$.

– un coefficient de transitivité :

$$c_t = 1 - A_E/(m_{r'} m_{r''})$$

où A_E est l'aire de l'ensemble E_{ijkl}. Ce coefficient est d'autant plus proche de 1 que cette aire est petite par rapport à l'aire de $I_{r'} \times I_{r''}$ et inversement.

Nous n'avons pu montrer la connexité de l'ensemble E_{ijkl} ni mettre en évidence un contrexemple. On peut simplement montrer que, si r_0 et r_1 sont des valeurs de $r_{i,j}$ telles que les intervalles de variation de $r_{k,l}$ associés ne sont pas disjoints, alors toute valeur r comprise entre r_0 et r_1 appartient à l'ensemble $E_{r'}$.

En effet, par permutation, on peut placer $r_{i,j}$ en $r_{n-1,n}$: en posant $r_{k,l}$ égal à une valeur appartenant simultanément aux deux intervalles de variation de $r_{k,l}$, on

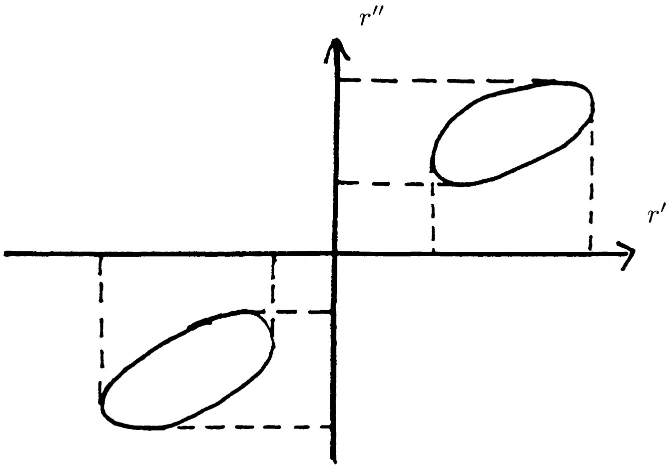

FIGURE 1
Exemple de non-connexité de E_{ijkl}

montre que les deux valeurs r_0 et r_1 appartiennent à l'intervalle de variation de $r_{i,j}$, et donc que l'intervalle $[r_0,\ r_1]$ y est inclus. Les cas possibles de non-connexité sont illustrés par la figure 1 ci-dessous.

4. Exemples numériques

Tous les calculs ci-dessous ont été effectués par un programme fonctionnant sur PC muni d'une carte graphique CGA, EGA ou VGA, que l'on peut demander à l'auteur de l'article (le programme est limité à l'étude des matrices de corrélation).

4.1 Analyse d'une matrice 4 × 4

On considère la matrice définie positive ci-dessous :

1.000			
0.800	1.000		
0.000	0.000	1.000	
−0.600	-0.500	0.500	1.000

Matrice de corrélation

4.1.1 Etude de chaque coefficient de corrélation

Nous donnons tout d'abord les résultats concernant chaque coefficient :

coefficients de contrainte	valeurs minimales	valeurs maximales
1.000	1.000	1.000
0.411 1.000	−0.189 1.000	0.989 1.000
0.548 0.532 1.000	−0.585-0.484 1.000	0.319 0.452 1.000
0.576 0.625 0.201 1.000	−0.824-0.855-0.799 1.000	0.024-0.105 0.799 1.000

Pour examiner l'influence de la 4^{e} variable sur les résultats précédents, on supprime la ligne et la colonne 4 :

coefficients de contrainte	valeurs minimales	valeurs maximales
1.000	1.000	1.000
0.000 1.000	−1.000 1.000	1.000 1.000
0.400 0.400 1.000	−0.600-0.600 1.000	0.600 0.600 1.000

Ainsi, lorsque le terme $r_{1,2}$ est égal à 0.8 et le terme $r_{1,3}$ à 0, le terme $r_{2,3}$ est compris entre -0.6 et 0.6. Nous donnons ci- dessous l'intervalle de variation et le coefficient de contrainte de $r_{2,3}$ pour différentes valeurs de $r_{1,3}$:

$$\begin{array}{lll}
r_{1,3} = 0.2 & c_k = 0.412 & I_v =]-0.428, 0.748[\\
r_{1,3} = 0.4 & c_k = 0.450 & I_v =]-0.230, 0.870[\\
r_{1,3} = 0.6 & c_k = 0.520 & I_v =]0.000, 0.960[\\
r_{1,3} = 0.8 & c_k = 0.640 & I_v =]0.280, 1.000[
\end{array}$$

4.1.2 Etude de la transitivité de deux coefficients de corrélation

Les coefficients de contrainte et de transitivité calculés sur tous les couples de produits scalaires de la matrice 4×4 sont les suivants :

couples	contrainte	transitivité	couples	contrainte	transitivité
$(3,1)$ $(2,1)$	0.5185	0.2616	$(4,2)$ $(3,1)$	0.7841	0.2714
$(3,2)$ $(2,1)$	0.4769	0.2971	$(4,2)$ $(3,2)$	0.7120	0.3876
$(3,2)$ $(3,1)$	0.4804	0.4566	$(4,2)$ $(4,1)$	0.2509	0.5282
$(4,1)$ $(2,1)$	0.1350	0.3579	$(4,3)$ $(2,1)$	0.4457	0.1735
$(4,1)$ $(3,1)$	0.6910	0.3535	$(4,3)$ $(3,1)$	0.4810	0.2742
$(4,1)$ $(3,2)$	0.7305	0.2978	$(4,3)$ $(3,2)$	0.5200	0.2157
$(4,2)$ $(2,1)$	0.1350	0.4329	$(4,3)$ $(4,1)$	0.5505	0.2136
			$(4,3)$ $(4,2)$	0.6160	0.2150

Sur les quinze valeurs, les résultats statistiques sont les suivants :

	Minimum	Maximum	Moyenne	Ecart-type
contrainte	0.13500	0.78410	0.50183	0.19403
transitivité	0.17350	0.52820	0.31577	0.09838

Pour la matrice étudiée, les coefficients de contrainte et de transitivité varient plutôt en sens inverse l'un de l'autre (leur coefficient de corrélation est égal à −0.4157).

4.2 Intervalles de variation numériques des coefficients de corrélation

Nous étudions ci-dessous la matrice de corrélation entre les notes d'étudiants en MIAGE aux trois premiers trimestres de la première année, aux deux premiers trimestres de la seconde, de la note de stage et de l'âge. Le nombre d'étudiants est de 67.

On notera que les coefficients de corrélation entre les 6 premières variables sont positifs et que la 7^{e} ligne est constituée de termes peu élevés.

1.000						
0.456	1.000					
0.465	0.623	1.000				
0.401	0.571	0.442	1.000			
0.200	0.420	0.366	0.668	1.000		
0.186	0.230	0.188	0.355	0.161	1.000	
−0.095	-0.017	0.001	0.040	0.082	0.071	1.000

Matrice de corrélation donnée

4.2.1 Résultats numériques

1.000						
0.403	1.000					
0.336	0.354	1.000				
0.454	0.524	0.531	1.000			
0.384	0.492	0.451	0.311	1.000		
0.229	0.353	0.309	0.370	0.320	1.000	
0.169	0.309	0.264	0.380	0.281	0.082	1.000

coefficients de contrainte

1.000						
−0.232	1.000					
−0.369	−0.305	1.000				
−0.256	−0.038	−0.024	1.000			
−0.330	−0.107	−0.250	−0.421	1.000		
−0.607	−0.430	−0.519	−0.450	−0.441	1.000	
−0.836	−0.687	−0.754	−0.584	−0.686	−0.916	1.000

valeurs minimales

1.000						
0.961	1.000					
0.960	0.986	1.000				
0.835	0.915	0.914	1.000			
0.901	0.909	0.849	0.956	1.000		
0.935	0.865	0.862	0.809	0.919	1.000	
0.827	0.695	0.719	0.656	0.751	0.920	1.000

valeurs maximales

Le coefficient de contrainte très faible du terme $r_{6,7}(c_k = 0.082)$ s'explique par la faiblesse des coefficients de corrélation de l'âge (variable 7) avec toutes les autres variables.

Les deux coefficients de contrainte les plus élevés concernent les coefficients de corrélation $r_{2,4}$ et $r_{3,4}$, dont les valeurs minimales sont -0.038 et -0.024. Toutes choses égales par ailleurs, chacun de ces coefficients ne peut donc être significativement négatif : on observe ici une relation de transitivité entre la matrice et chacun de ces deux coefficients.

4.2.2 Ensemble de variation numérique d'un couple de coefficients de corrélation

Le nombre de couples de coefficients de corrélation augmente très vite en fonction de la taille de la matrice (le temps de calcul également). En étudiant les coefficients de contrainte et de transitivité par des méthodes statistiques élémentaires, on peut déterminer les couples les plus intéressants à étudier.

Examinons tout d'abord le cas du couple $(r_{2,4}, r_{3,4})$, égal à $(0.571, 0.442)$ dans la matrice de corrélation donnée. Sur la figure 2, le terme $r_{2,4}$ est représenté en abscisse et le terme $r_{3,4}$ en ordonnée.

Comme dans les figures suivantes, le cadre est un carré dont les côtés représentent l'intervalle $[-1, 1]$: les coefficients étudiés ici étant des coefficients de corrélation, ce cadre représente l'ensemble $I_{s'} \times I_{s''}$ déduit de l'inégalité de Schwarz.

A l'intérieur de ce cadre, nous avons représenté un rectangle défini par les ensembles de variation $E_{r'}$ et $E_{r''}$ qui sont ici les intervalles $[-0.132, 0.946]$ et

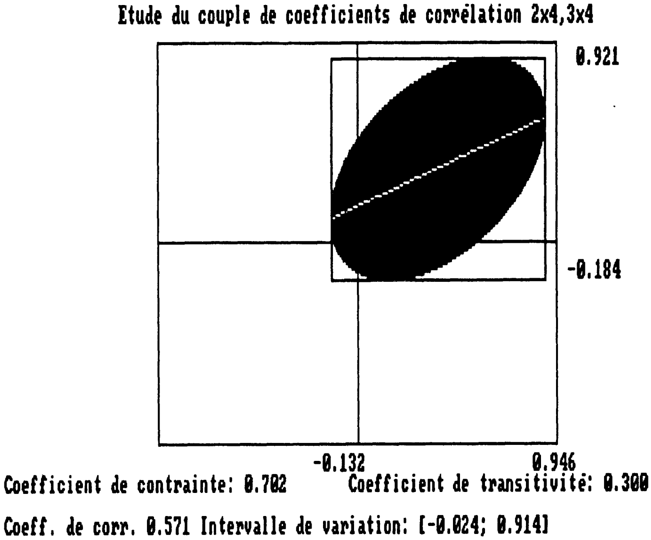

FIGURE 2
Ensemble de variation du couple $(r_{2,4}, r_{3,4})$

$[-0.184, 0.921]$: pour toute valeur r' de $r_{2,4}$ appartenant à $E_{r'}$, il existe une valeur r'' de $r_{3,4}$ appartenant à $E_{r''}$ telle que la matrice R soit définie positive et inversement.

Le coefficient de contrainte du couple $(r_{2,4}, r_{3,4})$ est calculé à partir de l'aire de ce rectangle (cf paragraphe 3.2). Il est ici parmi les plus forts observés $(c_k = 0.702)$.

L'ensemble dessiné à l'intérieur du rectangle précédent est l'ensemble de variation du couple; son aire détermine le coefficient de transitivité. Dans le cas du couple $(r_{2,4}, r_{3,4})$, ce coefficient est égal à $c_t = 0.300$.

La ligne blanche qui sépare cet ensemble en deux représente le milieu des intervalles de variation de $r_{3,4}$ (en ordonnée) suivant les valeurs de $r_{2,4}$ (en abscisse). La transitivité, mise en évidence par le graphique, n'est pas négligeable et semble linéaire.

Examinons maintenant le couple $(r_{1,2}, r_{5,7})$ (fig. 3) : les valeurs observées sont $r_{1,2} = 0.456$ et $r_{5,7} = 0.082$.

La contrainte reste assez forte $(c_k = 0.569)$ mais la transitivité est très faible $(c_t = 0.031)$: une variation de l'un des coefficients n'a presque pas de répercussion sur l'intervalle de variation de l'autre, sauf au voisinage des bornes des ensembles de variation de chacun d'entre eux.

Le couple dont la transitivité est la plus forte $(c_t = 0.382)$ est $(r_{4,2}, r_{5,2})$ (fig. 4). Les valeurs observées sont $r_{4,2} = 0.571$ et $r_{5,2} = 0.420$, significativement positives pour la taille de l'effectif de l'échantillon (67). On notera que ces coefficients concernent tous les deux la variable 2.

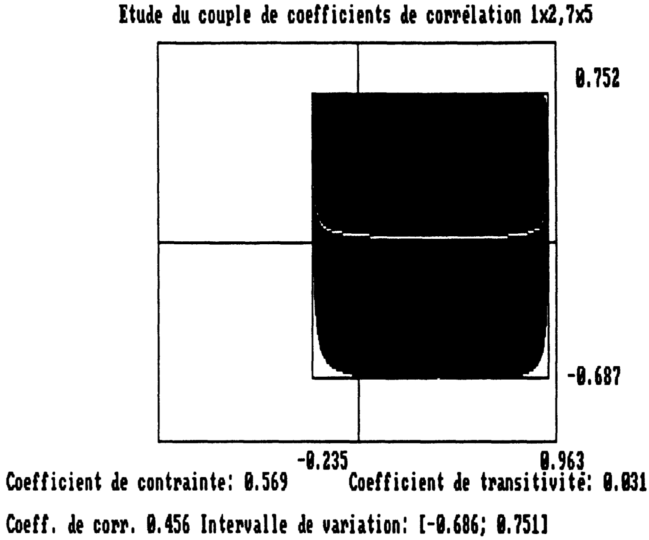

FIGURE 3
Ensemble de variation du couple $(r_{1,2}, r_{5,7})$

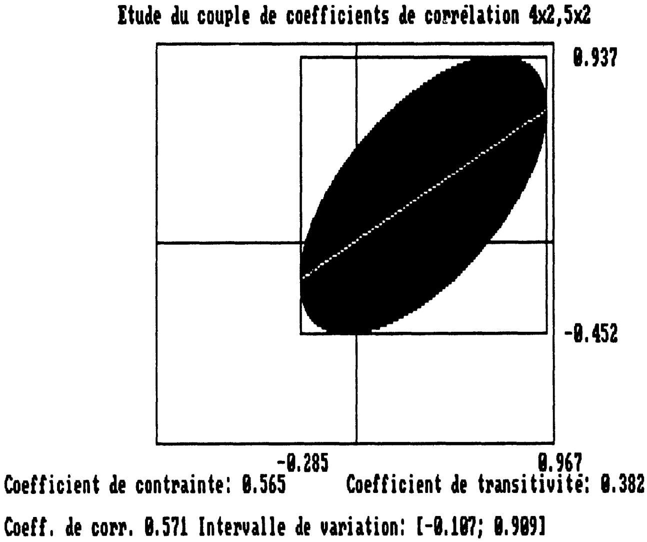

FIGURE 4
Ensemble de variation du couple $(r_{4,2}, r_{5,2})$

La transitivité de la corrélation est nettement plus forte que dans le premier exemple : un coefficient $r_{4,2}$ égal à 0.7 détermine comme intervalle de variation sur $r_{5,2}[0.042, 0.936]$: nécessairement, le coefficient $r_{5,2}$ sera positif.

5. Conclusion

Chaque terme non diagonal d'une matrice de produit scalaire est soumis à une contrainte numérique qui se traduit par l'appartenance à un intervalle de R inclus dans l'intervalle déduit de l'inégalité de Schwarz. Lorsque deux termes sont considérés simultanément, l'ensemble des couples est soumis à une contrainte analogue complétée par une liaison plus ou moins forte entre les deux termes. La relation (0) donnée dans le paragraphe 3.1 montre de façon évidente que les termes diagonaux de la matrice sont minorés.

Les contraintes sont d'autant plus grandes que la taille de la matrice est élevée. Par la multiplication des variables, on peut donc aboutir à un artefact néfaste à l'interprétation des résultats , la liaison statistique étant doublée d'une liaison numérique dont les effets méritent d'être examinés .

Bibliographie

CAILLIEZ F. PAGES J.P. (1976) Introduction à l'analyse des données (Paris, SMASH, 9 rue Duban, 616 pages)

CIARLET P.-G. (1988) Introduction à l'analyse matricielle et à l'optimisation (Paris, Masson, 292 pages)

SAPORTA G. (1990) Probabilités, analyse des données et statistique (Paris, Editions TECHNIP, 493 pages).