



HAL
open science

Nouvelle démonstration et généralisation de la formule de Kanto et Puntanen sur les corrélations entre résidus des régressions linéaires

Guy Der Megreditchian, Pierre Cazes

► **To cite this version:**

Guy Der Megreditchian, Pierre Cazes. Nouvelle démonstration et généralisation de la formule de Kanto et Puntanen sur les corrélations entre résidus des régressions linéaires. *Annales de l'ISUP*, 1988, XXXIII (1), pp.25-39. hal-03672284

HAL Id: hal-03672284

<https://hal.science/hal-03672284>

Submitted on 19 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NOUVELLE DEMONSTRATION ET GENERALISATION DE LA FORMULE DE KANTO ET PUNTANEN SUR LES CORRELATIONS ENTRE RESIDUS DES REGRESSIONS LINEAIRES

Guy DER MEGREDITCHIAN
Pierre CAZES

I - INTRODUCTION

On considère le système (y_1, y_2, X) où y_1 et y_2 sont des variables aléatoires scalaires et X un vecteur aléatoire. Deux régressions linéaires multiples peuvent être réalisées pour expliquer y_i :

$\tilde{y}_i[X]$, si y_i est expliquée uniquement par X ;

$\tilde{y}_i[X, y_j]$, si y_i est expliquée par le couple (X, y_j) ,

où $j = 3 - i$, $j = 2$, si $i = 1$ et $j = 1$, si $i = 2$.

En notant les résidus respectivement e_i et \mathcal{E}_i , on a les relations :

$$y_i = \tilde{y}_i[X] + e_i = \tilde{y}_i[X, y_j] + \mathcal{E}_i, \quad (1)$$

d'où nous tirons :

$$\mathcal{E}_i = e_i - [\tilde{y}_i[X, y_j] - \tilde{y}_i[X]]. \quad (1')$$

Kanto et Puntanen ont attiré l'attention [4,5] sur une curieuse liaison statistique entre les résidus ξ_i et e_j :

$$r [\xi_i, \xi_j] = -r [e_i, e_j] \quad (2)$$

Ils ont présenté (2) trois démonstrations de cette relation, l'une basée sur la technique de Cramer, l'autre sur le recours aux projecteurs orthogonaux, et la troisième sur l'utilisation d'un résultat de Rao.

L'objet de cet article est de proposer deux nouvelles démonstrations, l'une dans le but de généraliser le résultat et l'autre basée sur des considérations purement géométriques.

2 - DEMONSTRATION ALGEBRIQUE

Elle repose sur une formule de l'accroissement de la valeur prévue \tilde{y} de y quand on passe d'un système de n prédicteurs X à un système de $n + 1$ prédicteurs $\{X, x\}$ établie par l'auteur dans (2).

$$\tilde{y}(X, x) - \tilde{y}(X) = \frac{r[y, x/X] \cdot \sigma(y/X)}{\sigma(x/X)} [x - \tilde{x}(X)] \quad (3)$$

On a utilisé ici les notations suivantes :

$r[y, x/X]$, la corrélation partielle entre y et x , quand X est fixé,
 $\sigma^2(y/X)$, la variance partielle de y , quand X est fixé,
 $\sigma^2(x/X)$, la variance partielle de x , quand X est fixé,
 $\tilde{x}(x)$, la valeur prévue par régression linéaire multiple du prédicteur x à l'aide du vecteur X des prédicteurs.

On peut maintenant établir la relation :

$$\tilde{y}_i(X, y_j) - \tilde{y}_i(X) = \frac{r[e_i, e_j] \cdot \sigma(e_i)}{\sigma(e_j)} e_j \quad (4)$$

dans le cas où le couple (X, y_j) est gaussien.

On peut écrire cette relation sous la forme :

$$\tilde{y}_i(X, y_j) - \tilde{y}_i(X) = a_{ij} \cdot e_j \quad (4')$$

où

$$a_{ij} = \frac{r[e_i, e_j] \cdot \sigma(e_i)}{\sigma(e_j)} \quad (4'')$$

En effet, l'application directe de la relation (3) conduit à la relation :

$$\hat{y}_i[X, y_j] - \hat{y}_i(X) = \frac{r(y_i, y_j/X) \cdot \sigma(y_i/X)}{\sigma(y_j/X)} [y_j - \hat{y}_j(X)]. \quad (5)$$

Or $r(y_i, y_j/X)$ est la corrélation partielle entre y_i et y_j quand X est fixé, c'est-à-dire la corrélation entre les résidus e_i et e_j :

$$r(y_i, y_j/X) = r(e_i, e_j).$$

De même $\sigma^2(y_i/X)$ désigne la variance partielle de y_i , quand X est fixé, c'est-à-dire la variance du résidu e_i :

$$\sigma^2(y_i/X) = \text{cov}(y_i, y_i/X) = \sigma^2(e_i).$$

Si l'on ajoute à ces considérations le fait que l'on a noté $y_j - \hat{y}_j(X) = e_j$, la relation (4) est établie.

Nous obtenons ainsi :

$$\begin{cases} \mathcal{E}_i = e_i - a_{ij} \cdot e_j ; \\ \mathcal{E}_j = e_j - a_{ji} \cdot e_i . \end{cases} \quad (6)$$

Calculons la covariances entre \mathcal{E}_i et \mathcal{E}_j :

$$\begin{aligned} \text{cov}(\mathcal{E}_i, \mathcal{E}_j) &= \text{cov}(e_i - a_{ij} e_j, e_j - a_{ji} e_i) = \\ &= \text{cov}(e_i, e_j) + a_{ij} \cdot a_{ji} \text{cov}(e_j, e_i) - a_{ji} \sigma^2(e_i) - a_{ij} \sigma^2(e_j) . \end{aligned} \quad (7)$$

On vérifie directement les relations :

$$\begin{aligned} a_{ij} \cdot a_{ji} &= r^2(e_i, e_j) . \\ a_{ij} \cdot \sigma^2(e_j) &= \text{cov}(e_i, e_j) = a_{ji} \sigma^2(e_i) . \end{aligned} \quad (8)$$

On en tire par conséquent :

$$\text{cov}(\mathcal{E}_i, \mathcal{E}_j) = [r^2(e_i, e_j) - 1] \cdot \text{cov}(e_i, e_j) . \quad (9)$$

On calcule de même la variance de $\hat{\varepsilon}_i$ en tenant compte de (8)

$$\begin{aligned}\sigma^2(\hat{\varepsilon}_i) &= \text{cov} [e_i - a_{ij} e_j, e_i - a_{ij} e_j] = \\ &= \sigma^2(e_i) + a_{ij} [a_{ij} \cdot \sigma^2(e_j)] - 2 a_{ij} (\text{cov}[e_i, e_j]) = \\ &= \sigma^2(e_i) + a_{ij} a_{ji} \sigma^2(e_i) - 2 a_{ij} a_{ji} \sigma^2(e_i) = \\ &= \sigma^2(e_i) [1 - a_{ij} a_{ji}] = \sigma^2(e_i) \cdot [1 - r^2(e_i, e_j)]\end{aligned}$$

Nous avons ainsi :

$$\begin{aligned}\sigma(\hat{\varepsilon}_i) &= \sigma(e_i) \sqrt{1 - r^2(e_i, e_j)} ; \\ \sigma(\hat{\varepsilon}_j) &= \sigma(e_j) \sqrt{1 - r^2(e_i, e_j)} .\end{aligned}\quad (10)$$

L'application de (9) et (10) conduit au résultat recherché :

$$r(\hat{\varepsilon}_i, \hat{\varepsilon}_j) = \frac{[r^2(e_i, e_j) - 1] \text{cov}[e_i, e_j]}{\sigma(e_i) [1 - r^2(e_i, e_j)] \cdot \sigma(e_j)} = -r(e_i, e_j)$$

La formule (2) est établie. On peut l'explicitier ainsi. La corrélation entre les résidus $\varepsilon_i, \varepsilon_j$ des prévisions $\hat{y}_i [X, y_j]$ et $\hat{y}_j [X, y_i]$ de y_i et y_j à l'aide des prédicteurs $\{X, y_j\}$ et $\{X, y_i\}$ respectivement est égale à la corrélation entre les résidus e_i et e_j des prévisions $\hat{y}_i [X]$ et $\hat{y}_j [X]$ de y_i et y_j à l'aide des prédicteurs X prise avec le signe moins.

Le point essentiel de la variante que nous présentons de la démonstration est évidemment l'utilisation de la formule (3) sous sa forme particulière (4) qu'elle prend dans le contexte de notre problème. Cela ouvre la voie à une généralisation dans le cadre du système (Y_1, Y_2, X) ,

de trois vecteurs aléatoires :

$$Y_1 = (m_1 \times 1) ; \quad Y_2 = (m_2 \times 1) ; \quad X = (n \times 1)$$

3 - EXTENSION DE LA FORMULE DE KANTO-PUNTANEN AU CAS VECTORIEL

On peut réaliser comme dans le cas de y_i scalaire deux régressions linéaires pour expliquer Y_i :

$$\begin{aligned}\hat{y}_i(X) &= A_i' X & , \text{ si } Y_i \text{ est expliqué uniquement par le vecteur } X ; \\ \hat{y}_i(X, Y_j) &= A_{ij}' X & , \text{ si } Y_i \text{ est expliqué par le couple de prédicteurs vectoriels } \{X, Y_j\} = X_{ij}\end{aligned}$$

Les coefficients de régression A_i et A_{ij} sont définis par les expressions bien connues :

$$A_i = V_{XX}^{-1} V_{Xy_i} ; A_{ij} = V_{X_j X_j}^{-1} V_{X_j y_i} \quad (11)$$

En notant respectivement $e_i = (m_i \times 1)$ et $\tilde{e}_i = (m_i \times 1)$ les résidus vectoriels des prévisions correspondantes, on a les relations :

$$y_i = \hat{y}_i(X) + e_i = \tilde{y}_i(X, y_j) + \tilde{e}_i, \quad (12)$$

ou encore :

$$\tilde{e}_i = e_i - \{ \tilde{y}_i(X, y_j) - \hat{y}_i(X) \}. \quad (12')$$

Nous allons utiliser maintenant une formule relative à l'accroissement des formes quadratiques et bilinéaires [3] (ch.25) et que nous formulerons sous forme de lemme.

Lemme : Soit $\tilde{a} = \{a_1, a_2\}$, $\tilde{b} = \{b_1, b_2\}$

$$\text{et } \tilde{V} = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} = \tilde{V}',$$

où $a_1 = (n_1 \times 1)$, $a_2 = (n_2 \times 1)$, $b_1 = (n_1 \times 1)$, $b_2 = (n_2 \times 1)$;

$$V_{11} = (n_1 \times n_1) ; V_{12} = (n_1 \times n_2) ; V_{21} = (n_2 \times n_1) ; V_{22} = (n_2 \times n_2).$$

On a alors la relation suivante pour l'accroissement des formes bilinéaires :

$$\tilde{a}' \tilde{V}^{-1} \tilde{b} - a_1' V_{11}^{-1} b_1 = (a_2' - \tilde{a}_2') \tilde{V}_{22}^{-1} (b_2 - \tilde{b}_2), \quad (13)$$

où l'on a introduit les notations complémentaires :

$$\tilde{a}_2 = V_{21} V_{11}^{-1} a_1 ; \tilde{b}_2 = V_{21} V_{11}^{-1} b_1 ; \quad (13')$$

$$\tilde{V}_{22} = V_{22} - V_{21} V_{11}^{-1} V_{12}.$$

Le lemme peut être démontré en appliquant la formule de Frobenius [1,3] sur l'inversion des matrices partitionnées en quatre blocs que l'on peut écrire avec les notations utilisées :

$$\widehat{V}^{-1} = \begin{pmatrix} V_{11}^{-1} + \widehat{V}^{-1} V_{12} \widehat{V}^{-1} V_{21} V_{11}^{-1} & -\widehat{V}^{-1} V_{12} \widehat{V}^{-1} \\ -\widehat{V}^{-1} V_{21} V_{11}^{-1} & \widehat{V}^{-1} \end{pmatrix}. \quad (14)$$

La formule (13) peut maintenant être vérifiée directement en appliquant (14) au prix de quelques transformations algébriques.

Appliquons la formule (13) au calcul de l'accroissement de la valeur prévue \widehat{y}_i de y_i , quand on passe du système X des prédicteurs au système :

$$\widehat{X}_j = \{X, y_j\}.$$

$$\widehat{y}_i(X, y_j) - \widehat{y}_i(X) = V_{y_i X}^+ V_{\widehat{X}_j}^{-1} \{X, y_j\}' - V_{y_i X} V_{\widehat{X}_j}^{-1} X. \quad (14')$$

En posant :

$$\begin{aligned} a_1 &= V_{X y_i}; \quad b_1 = X; \quad a_2 = V_{y_j y_i}; \quad b_2 = y_j; \\ V_{11} &= V_{XX}; \quad V_{12} = V_{X y_j}; \quad V_{21} = V_{y_j X}; \quad V_{22} = V_{y_j y_j}. \end{aligned} \quad (15)$$

$$\widehat{a}_2 = V_{y_j X} V_{XX}^{-1} V_{X y_i}; \quad \widehat{b}_2 = V_{y_j X} V_{XX}^{-1} X.$$

Nous obtenons :

$$\begin{aligned} &\widehat{y}_i(X, y_j) - \widehat{y}_i(X) = \\ &= (V_{y_i y_j} - V_{y_i X} V_{XX}^{-1} V_{X y_j}) (V_{y_j y_j} - V_{y_j X} V_{XX}^{-1} V_{X y_j})^{-1} (y_j - V_{y_j X} V_{XX}^{-1} X). \end{aligned} \quad (16)$$

Or nous avons par définition :

$$\begin{aligned} y_j - V_{y_j X} V_{XX}^{-1} X &= y_j - \widehat{y}_j(X) = e_j. \\ V_{y_i y_j} - V_{y_i X} V_{XX}^{-1} V_{X y_j} &= V_{e_i} e_j'. \\ V_{y_j y_j} - V_{y_j X} V_{XX}^{-1} V_{X y_j} &= V_{e_j} e_j'. \end{aligned} \quad (17)$$

Par conséquent la relation (16) devient :

$$\hat{y}_i(x, y_j) - \hat{y}_i(x) = V_{e_i e_j} V_{e_j e_j}^{-1} e_j = G_{ij}' e_j, \quad (18)$$

où $G_{ij}' = V_{e_i e_j}^{-1} V_{e_j e_i}$ est le vecteur des coefficients de régression de e_j sur e_i .

En portant (18) dans (12') nous obtenons :

$$\begin{cases} \varepsilon_i = e_i - G_{ij}' e_j, \\ \varepsilon_j = e_j - G_{ji}' e_i. \end{cases} \quad (19)$$

On peut maintenant calculer les matrices de covariances entre les résidus vectoriels $V_{\varepsilon_i \varepsilon_j}$ et $V_{\varepsilon_i \varepsilon_i}$:

$$\begin{aligned} V_{\varepsilon_i \varepsilon_j} &= \text{cov}[\varepsilon_i, \varepsilon_j] = \text{cov}[e_i - G_{ij}' e_j, e_j - G_{ji}' e_i] = \\ &= \text{cov}[e_i, e_j] + G_{ij}' \text{cov}[e_j, e_i] G_{ji}' - G_{ij}' \text{cov}[e_j, e_j] - \text{cov}[e_i, e_i] G_{ji}'. \end{aligned} \quad (20)$$

Pour évaluer $V_{\varepsilon_i \varepsilon_j}$, calculons au préalable les expressions

$$\begin{aligned} G_{ij}' V_{e_j e_i} G_{ji}' &= V_{e_i e_j} V_{e_j e_j}^{-1} V_{e_j e_i} V_{e_i e_i}^{-1} V_{e_i e_j} \\ G_{ij}' V_{e_j e_j} &= V_{e_i e_j} V_{e_j e_j}^{-1} V_{e_j e_j} = V_{e_i e_j} \end{aligned} \quad (21)$$

$$V_{e_i e_i} G_{ji}' = V_{e_i e_i} V_{e_i e_i}^{-1} V_{e_i e_j} = V_{e_i e_j}.$$

Nous avons en définitive en portant (21) dans (20) :

$$V_{\varepsilon_i \varepsilon_j} = V_{e_i e_j} V_{e_j e_j}^{-1} V_{e_j e_i} V_{e_i e_i}^{-1} V_{e_i e_j} - V_{e_i e_j}. \quad (22)$$

Calculons de même la matrice de covariances $V_{\varepsilon_i \varepsilon_i}$:

$$\begin{aligned} V_{\varepsilon_i \varepsilon_i} &= \text{cov}[\varepsilon_i, \varepsilon_i] = \text{cov}[e_i - G_{ij}' e_j, e_i - G_{ij}' e_j] = \\ &= \text{cov}(e_i, e_i) - G_{ij}' \text{cov}(e_j, e_i) - \text{cov}(e_i, e_j) G_{ij}' + G_{ij}' \text{cov}(e_j, e_j) G_{ij}' = \\ &= V_{e_i e_i} - V_{e_i e_j} V_{e_j e_j}^{-1} V_{e_j e_i}. \end{aligned}$$

Nous obtenons ainsi :

$$V_{\mathcal{E}_i, \mathcal{E}_i} = V_{e_i, e_i} - V_{e_i, e_j} V_{e_j, e_j}^{-1} V_{e_j, e_i} \quad (22')$$

Introduisons maintenant les coefficients de régression de la prévision de \mathcal{E}_i à l'aide de \mathcal{E}_j :

$$\zeta_{\mathcal{E}_i} = \Gamma_{ij}' \mathcal{E}_j, \quad \text{où } \Gamma_{ij}' = V_{\mathcal{E}_j, \mathcal{E}_j}^{-1} V_{\mathcal{E}_j, \mathcal{E}_i}, \quad (23)$$

Nous obtenons alors de (22) et (22') la relation :

$$\begin{aligned} -V_{\mathcal{E}_j, \mathcal{E}_i} &= [V_{e_j, e_i} V_{e_i, e_i}^{-1} V_{e_i, e_j}] V_{e_j, e_j}^{-1} V_{e_j, e_i} - V_{e_j, e_i} = \\ &= [V_{e_j, e_j} - V_{e_j, \mathcal{E}_j}] V_{e_j, e_j}^{-1} V_{e_j, e_i} - V_{e_j, e_i} = \\ &= -V_{\mathcal{E}_j, \mathcal{E}_j} V_{e_j, e_j}^{-1} V_{e_j, e_i}, \end{aligned}$$

que nous pouvons mettre sous la forme :

$$-V_{\mathcal{E}_j, \mathcal{E}_j}^{-1} V_{\mathcal{E}_j, \mathcal{E}_i} = -V_{e_j, e_j}^{-1} V_{e_j, e_i}, \quad (24)$$

ou encore :

$$\Gamma_{ij}' = -G_{ij}. \quad (24')$$

Nous avons ainsi obtenu dans le cas des résidus vectoriels une relation qui généralise le résultat (2) de KANTO et PUNTANEN pour les résidus scalaires. Nous la formulerons sous forme de théorème :

Théorème 1 : La matrice Γ_{ij}' des coefficients de régression de la prévision de \mathcal{E}_i à l'aide de \mathcal{E}_j et la matrice G_{ij}' de la prévision de e_i à l'aide de e_j sont liées par une relation simple :

$$\Gamma_{ij}' = -G_{ij}'. \quad (24'')$$

La liaison statistique entre les vecteurs \mathcal{E}_i et \mathcal{E}_j (respectivement entre e_i et e_j) peut être également appréciée à l'aide des corrélations canoniques.

$$\begin{aligned} \text{Soit } z_k[\mathcal{E}_i] &= F_k'[\mathcal{E}_i] \cdot \mathcal{E}_i, \\ \text{et } z_k[\mathcal{E}_j] &= F_k'[\mathcal{E}_j] \cdot \mathcal{E}_j, \end{aligned} \quad (25)$$

Le couple des K-ièmes variables canoniques associées aux vecteurs aléatoires \mathcal{E}_i et \mathcal{E}_j . Les coefficients $F_k[\mathcal{E}_i]$ et $F_k[\mathcal{E}_j]$ de ces transformations de variables (les K-ièmes facteurs canoniques) sont respectivement les K-ièmes vecteurs propres des matrices :

$$H[\mathcal{E}_i] = V_{\mathcal{E}_i}^{-1} \cdot V_{\mathcal{E}_i \mathcal{E}_j} \cdot V_{\mathcal{E}_j}^{-1} \cdot V_{\mathcal{E}_j \mathcal{E}_i} \quad (26)$$

$$\text{et } H[\mathcal{E}_j] = V_{\mathcal{E}_j}^{-1} \cdot V_{\mathcal{E}_j \mathcal{E}_i} \cdot V_{\mathcal{E}_i}^{-1} \cdot V_{\mathcal{E}_i \mathcal{E}_j},$$

$$F_k[\mathcal{E}_i] = C_k[H[\mathcal{E}_i]], \quad (26')$$

$$F_k[\mathcal{E}_j] = C_k[H[\mathcal{E}_j]],$$

où l'on a noté $C_k[H[\mathcal{E}_i]], C_k[H[\mathcal{E}_j]]$ le K-ième vecteur propre de la matrice correspondante.

La K-ième corrélation canonique $\rho_k[\mathcal{E}_i, \mathcal{E}_j]$ entre les vecteurs \mathcal{E}_i et \mathcal{E}_j est définie par la K-ième valeur propre de $H[\mathcal{E}_i]$ ou $H[\mathcal{E}_j]$:

$$\rho_k^2[\mathcal{E}_i, \mathcal{E}_j] = \lambda_k[H[\mathcal{E}_i)] = \lambda_k[H[\mathcal{E}_j)].$$

On définira de même les variables canoniques, les facteurs canoniques, les corrélations canoniques entre les résidus vectoriels e_i et e_j :

$$z_k[e_i] = F_k'[e_i] \cdot e_i; \quad (27)$$

$$z_k[e_j] = F_k'[e_j] \cdot e_j;$$

$$F_k[e_i] = C_k[H[e_i]],$$

$$F_k[e_j] = C_k[H[e_j]]. \quad (27')$$

$$H_k [e_i] = V_{e_i, e_i}^{-1} V_{e_i, e_j} V_{e_j, e_j}^{-1} V_{e_j, e_i}$$

$$H_k [e_j] = V_{e_j, e_j}^{-1} V_{e_j, e_i} V_{e_i, e_i}^{-1} V_{e_i, e_j}$$

On peut maintenant établir quelques relations de corrélation canoniques pour préciser les liaisons statistiques entre les résidus vectoriels.

Théorème 2 : Les matrices $H(\mathcal{Z}_i)$ et $H(e_i)$ définissant les facteurs et les corrélations canoniques entre résidus vectoriels sont égales :

$$\begin{aligned} H(\mathcal{Z}_i) &= H(e_i) , \\ H(\mathcal{Z}_j) &= H(e_j) . \end{aligned} \quad (28)$$

La démonstration est immédiate en appliquant la relation (24') :

$$H(\mathcal{Z}_i) = \Gamma_{j' i} \Gamma_{j' j} = G_{j' i} \cdot G_{j' j} = H(e_i) ,$$

$$H(\mathcal{Z}_j) = \Gamma_{j' j} \Gamma_{j' i} = G_{j' j} \cdot G_{j' i} = H(e_j) .$$

Théorème 3 : Les facteurs canoniques des couples $(\mathcal{Z}_i, \mathcal{Z}_j)$ et (e_i, e_j) sont égaux :

$$\begin{aligned} F_k(\mathcal{Z}_i) &= F_k(e_i) , \\ F_k(\mathcal{Z}_j) &= F_k(e_j) . \end{aligned} \quad (29)$$

Théorème 4 : Les corrélations canoniques entre couples de résidus vectoriels $(\mathcal{Z}_i, \mathcal{Z}_j)$ et (e_i, e_j) sont égales :

$$\rho_k^2(\mathcal{Z}_i, \mathcal{Z}_j) = \rho_k^2(e_i, e_j) . \quad (30)$$

Les théorèmes 3 et 4 découlent immédiatement du théorème 2 puisque par définition :

$$F_k(\mathcal{Z}_i) = C_k[H(\mathcal{Z}_i)] ; \rho_k^2(\mathcal{Z}_i, \mathcal{Z}_j) = \lambda_k[H(\mathcal{Z}_i)] = \lambda_k[H(\mathcal{Z}_j)]$$

$$F_k(e_i) = C_k[H(e_i)] ; \rho_k^2(e_i, e_j) = \lambda_k[H(e_i)] = \lambda_k[H(e_j)]$$

Les résultats obtenus apportent une information intéressante sur les structures des erreurs d'interpolation des valeurs des champs aléatoires, que l'on pratique couramment dans les sciences de la Terre (Météorologie, Géologie, Géographie, Géophysique). Elles permettent d'évaluer avant même la prise en compte d'un nouveau groupe de prédicteurs potentiels son apport complémentaire dans la structure des résidus.

4 - DEMONSTRATION GEOMETRIQUE

On se place dans l'espace $L^2_{\mathbb{Z}}$ des fonctions aléatoires du second ordre de :

$$\mathbb{Z} = \{y_1, y_2, X\}, \quad X \in \mathbb{R}^n,$$

muni du produit scalaire associé à la covariance. On supposera, sans nuire à la généralité, que \mathbb{Z} est centré. Cela entraîne que le produit scalaire entre deux éléments centrés est égal à la covariance, et que leur cosinus est égal à leur coefficient de corrélation.

Introduisons les notations :

P_X , le projecteur sur l'espace H_X engendré par les fonctions linéaires des variables aléatoires composant X ,

$P_{j,X}$ le projecteur sur l'espace $H_{j,X}$ engendré par les fonctions linéaires du couple (y_j, X) ($1 \leq j \leq 2$).

On a alors :

$$y_i = \hat{y}_i(X) + e_i = P_X(y_i) + e_i \quad (31)$$

$$= \hat{y}_i(X, y_j) + \varepsilon_i = P_{j,X}(y_i) + \varepsilon_i, \quad (32)$$

où $P_X[y_i]$ désigne la projection de y_i sur H_X et $P_{j,X}[y_i]$ la projection de y_i sur $H_{j,X}$.

$$P_X[y_i] = X'a, \quad \text{ou} \quad a = V_{XX}^{-1} V_{Xy_i}.$$

De même on a :

$$y_j = P_X(y_j) + e_j.$$

Comme e_j est orthogonal à H_X , et que e_j et H_X engendrent $H_{j,X}$ on en déduit que :

$$P_{j,X} = P_X + P_{\Delta} e_j. \quad (33)$$

Ici $P_{\Delta e_j}$ désigne le projecteur sur la droite Δe_j dont e_j est le vecteur directeur. On déduit alors des relations (31) et (33) que :

$$\begin{aligned}\widehat{y}_i[X, y_j] &= P_{d, X} [y_i] = P_{d, X} [\tilde{y}_i(X) + e_i] = \\ &= P_{d, X} [\tilde{y}_i(X)] + P_{d, X} [e_i].\end{aligned}$$

Or comme $\tilde{y}_i(X)$ appartient à l'espace $H_{d, X}$ il est identique à sa projection sur cet espace :

$$P_{d, X} [\tilde{y}_i(X)] = \tilde{y}_i(X),$$

d'où il suit, puisque e_i est orthogonal à H_X :

$$P_X [e_i] = 0, \quad \text{d'où}$$

$$\widehat{y}_i[X, y_j] = \tilde{y}_i(X) + P_{\Delta e_j} [e_i];$$

on a alors :

$$e_i = y_i - \widehat{y}_i[X, y_j] = y_i - \tilde{y}_i(X) - P_{\Delta e_j} [e_i],$$

ou encore :

$$e_i = e_i - P_{\Delta e_j} [e_i]. \quad (34)$$

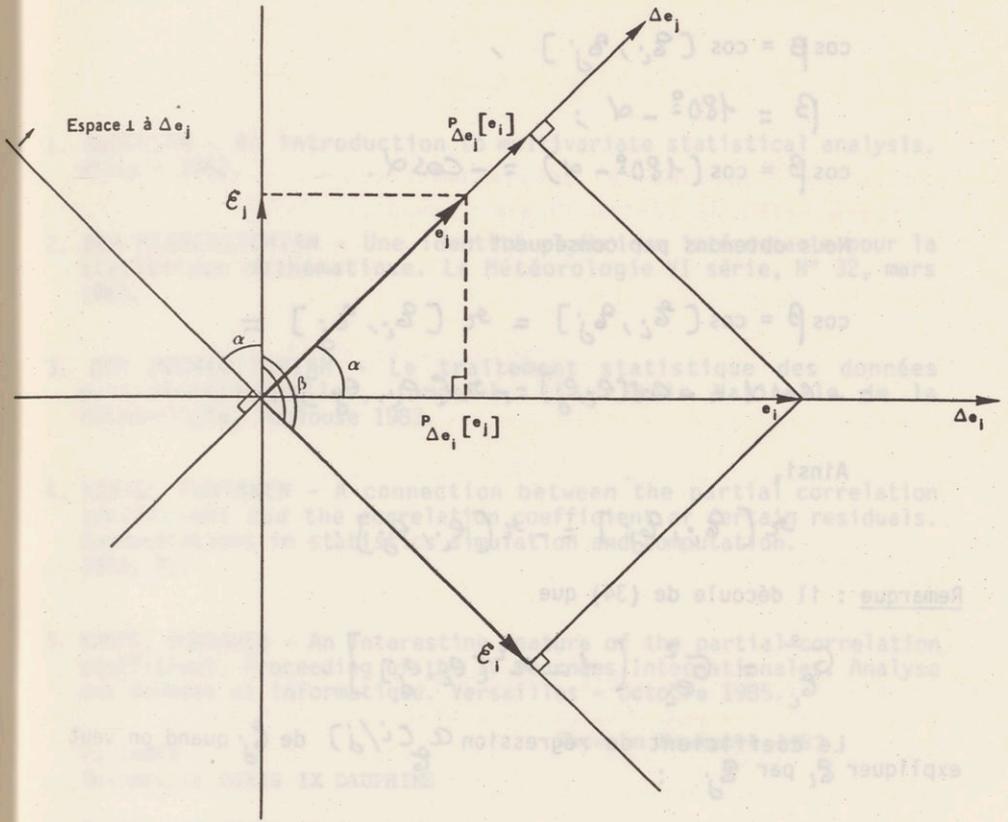
e_i est donc égal à la projection de e_i sur l'espace orthogonal à Δe_j .

Plus précisément, il découle de (34) que :

a) e_i appartient au plan engendré par (e_i, e_j) .

b) dans ce plan, e_i est la projection de e_i sur la droite orthogonale à e_j :

De même e_j se trouve dans le plan engendré par (e_i, e_j) et dans ce plan e_j est la projection de e_j sur la droite orthogonale à e_i comme le montre la figure.



Interprétation géométrique de la propriété des corrélations partielles

On a alors avec les notations de la figure :

$$\cos \beta = \cos [\vartheta_i, \vartheta_j] ,$$

$$\beta = 180^\circ - \alpha ;$$

$$\cos \beta = \cos (180^\circ - \alpha) = -\cos \alpha .$$

Nous obtenons par conséquent :

$$\begin{aligned} \cos \beta &= \cos [\vartheta_i, \vartheta_j] = r [\vartheta_i, \vartheta_j] = \\ &= -\cos \alpha = -\cos [e_i, e_j] = -r [e_i, e_j] . \end{aligned}$$

Ainsi,

$$r [\vartheta_i, \vartheta_j] = -r [e_i, e_j] .$$

Remarque : il découle de (34) que

$$\sigma_{\vartheta_i}^2 = \sigma_{e_i}^2 [1 - r^2 [e_i, e_j]] .$$

Le coefficient de régression $a_{\vartheta} [i/j]$ de ϑ_j quand on veut expliquer ϑ_i par ϑ_j :

$$\begin{aligned} a_{\vartheta} [i/j] &= r [\vartheta_i, \vartheta_j] \frac{\sigma_{\vartheta_i}}{\sigma_{\vartheta_j}} = r [\vartheta_i, \vartheta_j] \frac{\sigma_{e_i}}{\sigma_{e_j}} = \\ &= -r [e_i, e_j] \frac{\sigma_{e_i}}{\sigma_{e_j}} = -a_e [i/j] . \end{aligned}$$

Il est donc l'opposé du coefficient de régression $a_e [i/j]$ de e_j quand on veut expliquer e_i par e_j :

$$a_{\vartheta} [i/j] = -a_e [i/j] .$$

REFERENCES BIBLIOGRAPHIQUES

1. ANDERSON - An introduction to multivariate statistical analysis. Wiley - 1962.
2. DER MEGREDITCHIAN - Une identité algébrique intéressante pour la statistique mathématique. La Météorologie VI série, N° 32, mars 1983.
3. DER MEGREDITCHIAN - Le traitement statistique des données multidimensionnelles, tomes I, II - Ecole Nationale de la Météorologie, Toulouse 1983.
4. KANTO, PUNTANEN - A connection between the partial correlation coefficient and the correlation coefficient of certain residuals. Communications in statistics simulation and computation. 1983, 72.
5. KANTO, PUNTANEN - An interesting feature of the partial correlation coefficient. Proceeding of the IV Journées Internationales. Analyse des données et informatique. Versailles - Octobre 1985.

Reçu en Novembre 1987

P. CAZES
Université PARIS IX DAUPHINE

G. DER MEGREDITCHIAN
Météorologie Nationale

J. H. G. is Dozent at the Department of Statistics, University of Bern, Sidlerstrasse 5, 3012 Bern, Switzerland. J. Tiago de Oliveira is Professor at the Department of Statistics, University of Lisbon, 58 Rua da Escola Politecnica, 1294 Lisboa Codex, Portugal. The work was done while the second author was Visiting Professor at the Department of Statistics, University of Bern.