Analyse Statistique de la Contiguïté *(Statistical Analysis of Contiguity)* [Publications de l'ISUP, XVIII, P 81 - 112]

Troisième partie (fin), pages 104 – 112 *(Third part (last part), p 104 - 112)*

La formule précédente demande, pour le calcul effectif de $\hat{\beta}$, l'inversion de matrices ayant autant de lignes qu'il y a d'observations, (matrices 88×88 pour le cas des statistiques départementales. Ces matrices faisant l'objet d'estimations assez approchées, le résultat des calculs risque de n'avoir pas grand sens. Il semble donc plus efficient de retenir pour $\hat{\beta}$ l'estimation *centrée* :

$$\hat{\beta} = (Z'Z)^{-1} Z'X$$

en tenant compte du fait que cette estimation *n'est plus à dispersion minimale*, et en estimant la nouvelle dispersion de $\hat{\beta}$.

$$V(\hat{\beta}) = (Z'Z)^{-1} Z' V(\varepsilon) Z (Z'Z)^{-1} \qquad (1)$$

c) Estimation de la matrice des covariances des résidus.

Si la variable endogène X et les variables exogènes Z_1, Z_2, Z_p sont susceptibles d'avoir été générées par un modèle stationnaire*, la variable $\varepsilon = X - Z\beta$ est également stationnaire, (la liaison entre les observations Z_{1i} et Z_{kj} des variables Z_1 et Z_k dépendant de l, k et de d (i, j), sans dépendre ni de i, ni de j).

On peut donc supposer que la matrice des covariances du vecteur résiduel est du type : $V(\varepsilon) = v_o I + v_1 M_1 + \ldots + v_e M_e$.

Pour les échantillons importants, on estimera $v_o, \ldots v_e$ à partir des résidus observés (comme précédemment, v_o est la variance résiduelle, v_i la covariance des deux résidus correspondant à des sommets distants de i sur le graphe, Mi la Matrice associé au graphe des distances « i »).

Le biais résultant de l'utilisation de $\hat{\beta}$ au lieu de la vraie valeur de β est d'autant plus réduit que la taille « n » de l'échantillon est importante : (si $\hat{\varepsilon}$ désigne le vecteur de résidus observés : $\hat{\varepsilon} = Z(\beta - \hat{\beta}) + \varepsilon$.

Par suite, quand $n \to \infty$, $\hat{\beta}$ tend en probabilité vers β, et $\hat{\varepsilon}$ tend en probabilité vers ε).

* c'est-à-dire tel que la covariance de deux observations ne dépend que de leur distance sur le graphe.

Sans faire intervenir les valeurs numériques des résidus on peut estimer $v_i(\varepsilon)$ à partir de $v_i(X)$ et des $v_i(Zk)$ à l'aide de la relation $E(\hat{\varepsilon}\hat{\varepsilon}') = E(X - Z\hat{\beta})(X - Z\hat{\beta})'$.

(Notons que les tests de significativité des coefficients de contigüité $C(\varepsilon)$ ne s'appliquent plus rigoureusement, car dans l'hypothèse où les résidus *théoriques* ont une matrice de covariance $\sigma^2 I$, les résidus *observés* ont une matrice de covariance : $\sigma^2 (I - \frac{U}{n} - Z(Z'Z)^{-1} Z')$, d'où une inextricable complication de la loi de $C(\hat{\varepsilon})$).

On aboutit donc à une estimation du type :

$$V(\varepsilon) = v_o I + v_1 M_1 + \ldots + v_e M_e$$

Par suite, en remplaçant $V(\varepsilon)$ par sa valeur dans la formule (1)

$$V(\hat{\beta}) = (Z'Z)^{-1} Z' (v_o I + v_1 M_1 + \ldots + v_e M_e) Z (Z'Z)^{-1}$$

soit :

$$(2)\quad V(\hat{\beta}) = v_o (Z'Z)^{-1} + v_1 (Z'Z)^{-1} (Z'M_1 Z) (Z'Z)^{-1} + \ldots + v_e (Z'Z)^{-1} (Z'M_e Z) (Z'Z)^{-1}$$

Notons que $v_o (Z'Z)^{-1}$ n'est autre que la matrice des covariances de $\hat{\beta}$ dans l'hypothèse de non contigüité des résidus.

La formule précédente montre que si les variables *exogènes* sont « non contigües », la zone de confiance des coefficients de régression est inchangée, quelle que soit la contigüité des résidus.

Si les variables exogènes dépendent significativement du graphe, les éléments de $Z'M_k Z$ ont des valeurs élevées pour les premières valeurs de k.

Le terme rectificatif est alors donné par la formule (2).

d) Exemple d'application numérique.

Nous nous proposons d'étudier la dépendance qui peut exister entre un indicateur de dépenses départementales D, et le taux de population urbaine U d'un même département.

$$D_i = \alpha U_i + \beta + e_i$$

La M. de C. du vecteur $\binom{D}{U}$ est $V = \begin{pmatrix} 653{,}273 & 332{,}715 \\ 332{,}715 & 271{,}839 \end{pmatrix}$

On a donc : $D_i = 1{,}224\, U_i + \beta + \varepsilon_i$ avec un coefficient de corrélation $\rho = 0{,}7897$.

On trouve : $\sigma_\alpha^2 = 0{,}01053$, sans faire d'hypothèse sur la liaison des résidus.

Ce qui donne approximativement la zone de confiance au seuil 0,05 :

$$1{,}02 \leqslant \alpha \leqslant 1{,}43$$

En fait, le coefficient de contigüité des résidus est $C(\hat{\varepsilon}) = 0{,}396$; bien que le test de contigüité ne s'applique plus rigoureusement puisqu'il s'agit de résidus observés, la taille de l'échantillon (88) et la valeur très faible trouvée pour $C(\hat{\varepsilon})$ (à 7 écart-types de la moyenne) nous conduisent à considérer que les résidus sont significativement contigus. En ne retenant que le terme rectificatif qui fait intervenir des liaisons au niveau 1 :

On trouve $\sigma_\alpha^2 = 0{,}01053 + 0{,}0083 = 0{,}01883$ $(\sigma_\alpha = 0{,}137)$.

La nouvelle zone de confiance est alors : $0{,}950 \leqslant \alpha \leqslant 1{,}498$, pour le même seuil de 0,05 .

D'une manière générale, on notera que, comme dans le cas des séries temporelles, la précision des coefficients est souvent *surestimée* lorsqu'il existe des liaisons directes.

B - Perte d'information et « pseudo-taille » de l'échantillon.

Le modèle théorique le plus simple, susceptible de générer des réalisations dépendantes du graphe est le modèle « stationnaire », où il existe

entre deux observations un coefficient d'autocorrélation qui ne dépend que de leur distance sur le graphe.

Ainsi, le vecteur des observations d'une variable X a pour matrice des covariances :

$$V = v_0 I + v_1 M_1 + \ldots + v_k M_k$$

(les différentes observations ont même moyenne μ , et même variance v_0).

I est la matrice unité.

$M_1, \ldots, M_k$ les matrices associées aux graphes $G_1, \ldots, G_k$ $(m_{uij} \in)$

$M_u = 1$ si $d(i,j) = u$, $m_{uij} = 0$ si $d(i,j) \neq u$) .

$\bar{\mu}$ est le vecteur dont toutes les composantes sont égales à μ .

Calculons les quantités d'informations relatives aux paramètres μ et v_u $(u = 1, \ldots, k)$.

La densité de probabilité du vecteur X s'écrit :

$$P(X) = \frac{1}{(2\pi)^{\frac{n}{2}} |V|} \exp \left\{ \frac{1}{2} (X - \bar{\mu})' V^{-1} (X - \bar{\mu}) \right\}$$

$$\text{Log } P(X) = -\frac{n}{2} \text{ Log } 2\pi + \frac{1}{2} \text{ Log } |V^{-1}| - \frac{1}{2} \text{tr } (V^{-1} (X - \bar{\mu}) (X - \bar{\mu})')$$

Posons $V^{-1} = C = (c_{ij})$,

$(X - \bar{\mu}) (X - \bar{\mu})' = D(\mu)$,

et $\frac{1}{2} (\text{Log } |C| - \text{tr } (CD)) = L$.

$$\text{On a : } \frac{\partial L}{\partial v_u} = \sum_{i,j} \frac{\partial L}{\partial c_{ij}} \cdot \frac{\partial c_{ij}}{\partial v_u} = \text{tr} \left(\frac{\partial L}{\partial C} \cdot \frac{\partial C}{\partial v_u} \right)$$

où la matrice $\frac{\partial L}{\partial C}$ a pour terme générique $\frac{1}{2}\left(\frac{\partial L}{\partial c_{ij}}\right)$ et $\left(\frac{\partial L}{\partial c_{ii}}\right)$ sur la diagonale :

La matrice $\frac{\partial L}{\partial C}$ est telle que $\begin{cases} \frac{\partial L}{\partial c_{ij}} = \left(\frac{\text{cof}\,(c_{ij})}{|C|} - d_{ij}\right) \\ \frac{\partial L}{\partial c_{ii}} = \frac{1}{2}\left(\frac{\text{cof}\,(c_{ii})}{|C|} - d_{ii}\right) \end{cases}$

Donc $\frac{\partial L}{\partial C} = \frac{1}{2}(C^{-1} - D) = \frac{1}{2}(V - D)$

On a aussi : $\frac{\partial C}{\partial v_u} = \frac{\partial V^{-1}}{\partial v_u} = -V^{-1}\frac{\partial V}{\partial v_u}V^{-1} = -V^{-1}M_u V^{-1}$

donc : $\frac{\partial L}{\partial v_u} = -\frac{1}{2}\,\text{tr}\,(V - D)V^{-1}M_u V^{-1}$

D'autre part : $\frac{\partial L}{\partial \mu} = \text{tr}\,\left(C.\frac{\partial D}{\partial \mu}\right)$

Calculons : $\frac{\partial^2 L}{\partial v^2}$ et $\frac{\partial^2 L}{\partial \mu^2}$:

$$\frac{\partial V^{-1}M_u V^{-1}}{\partial v_u} = \frac{\partial V^{-1}}{\partial v_u}.M_u V^{-1} + V^{-1}\frac{\partial M_u V^{-1}}{\partial v_u}$$

$$= -V^{-1}M_u V^{-1}M_u V^{-1} - V^{-1}(M_u . V^{-1}M_u V^{-1})$$

$$= 2V^{-1}(M_u V^{-1})^2$$

Par suite :

$$\frac{\partial^2 L}{\partial v_u^2} = -\frac{1}{2}\,\text{tr}\,(2DV^{-1} - I)(M_u V^{-1})^2$$

Si les observations sont indépendantes $v = \sigma^2 I$ et $M_u = M_o = I$, l'équation $\frac{\partial L}{\partial \sigma^2} = 0$ redonne $\sigma^2 = \frac{1}{2}\Sigma (x_i - \mu)^2$ et l'on a bien

$$\frac{\partial^2 L}{\partial \sigma^2} = \frac{-1}{2\sigma^4}\quad (2n - 1) < 0$$

On a de même $\frac{\partial^2 L}{\partial \mu^2} = -\sum_{ij} c_{ij} = -\,\text{tr}\,(V^{-1}U)$

(avec $u_{ij} = 1$ pour tout i et j).

Par suite :

$$I_{v_u} = E\left(\frac{-\partial^2 L}{\partial v_u^2}\right) = \frac{1}{2}\,\text{tr}\,(M_u V^{-1}) \quad \text{et} \quad I\mu = E\left(-\frac{\partial^2 L}{\partial \mu_2}\right) = \ldots$$

$$\ldots = +\,\text{tr}\,(V^{-1}U)$$

I_{v_u} et $I\mu$ étant les quantités d'information de Fischer, relatives aux paramètres v_u et μ.

Dans le cas où $V = \sigma^2(I + \rho M)$, cas où la contigüité ne se manifeste qu'au niveau (1), on a :

$$I_m = \frac{1}{\sigma^2}\,\text{tr}\,(I + \rho M)^{-1}U = \frac{1}{\sigma^2}\,\text{tr}\,(I - \rho M + \rho^2 M^2 - \rho^3 M^3 + \ldots)U$$

Or, $\text{tr}\,M^\alpha U$ est le nombre « c_α » de chemins de longueur α du graphe.

Donc, $I_m = \frac{1}{\sigma^2}(n - c_1\rho + c_2\rho^2 - c_3\rho^3 + \ldots)$

Si $\rho = 0$, on retrouve la quantité connue $\frac{n}{\sigma^2}$. On constate ce qui est intuitif, que la quantité d'information sur la moyenne, les autres paramètres étant connus, est plus élevée si les liaisons (caractérisées par ρ) sont « négatives ». Elle est, en général, plus faible si les liaisons sont « positives ».

On peut définir la pseudo-taille n_1 de l'échantillon comme l'effectif qui suffirait à donner la même information sur la moyenne, si les observations étaient indépendantes.

n_1 est donc tel que :

$$\frac{n_1}{\sigma^2} = I_n = \frac{1}{\sigma^2}(n - c_1 \rho + c_2 \rho^2 \ldots)$$

Soit :

$$n_1 = n - \Sigma c_i \rho^i = \mathrm{tr}\,(I + \rho M)^{-1} U$$

Supposons ainsi que le graphe soit formé de doublets disjoints : l'échantillon est donc formé de couples d'observations corrélées. Le nombre de chemin de longueur k est constant et égal à n quelque soit k .

On a donc ici : $n_1 = n(1 - \rho + \rho^2 - \rho^3 \ldots) = \dfrac{n}{(1+\rho)}$

Pour $\rho = 1$, c'est-à-dire lorsque les n observations se décomposent en $\dfrac{n}{2}$ couples contenant la même observation, on a bien : $n_1 = \dfrac{n}{2}$.

Supposons au contraire que le graphe soit complet (tout couple de sommet est relié par une arête). Toutes les observations sont corrélées entre elles de la même façon. Le calcul direct de l'inverse de $V = I + \rho M$ est immédiat, et l'on trouve :

$$n_1 = \frac{n}{1 + (n-1)\rho}$$

Si $\rho = 1$, tout se passe comme si l'on n'avait qu'une seule observation.

La pseudo-taille de l'échantillon est $n_1 = 1$.

Pour des graphes plus complexes, le calcul de $n_1 = \mathrm{tr}\left\{(I+\rho M)^{-1} U\right\}$ doit être fait automatiquement ; dans ce cas, on peut estimer ρ par le complément à 1 du coefficient de contigüité ($\rho = 1 - c.$) .

Les valeurs trouvées dans le cas des statistiques départementales (n_1 varie de 40 à 80 pour les variables des exemples précédents, où n=88) doivent inciter à une grande prudence en ce qui concerne l'interprétation des tests.

BIBLIOGRAPHIE

[1] BERGE (C.) – *Théorie des graphes et ses applications.* (Dunod 1963).

[2] CHARTIER (F.) – *L'estimation statistique dans le cas d'observations non indépendantes.* Etude d'un cas particulier I.S.U.P. Vol. 1 Fasc. 2 - 1952.

[3] FERREZ (J.) – *Inégalité régionale des possibilités d'accès à l'Education.* (Publ. O.C.D.E. 1961).

[4] GEARY (R.C.) – A general expression for the moments of certain symmetrical fonctions of normal sample. (*Biometrika* XXV - p. 184 - 1933).

The contiguity ratio and statistical mapping « the incorporated statistician » Vol. 5, n° 3, p. 115-45, 1954).

[5] GURLAND (J.) – Quadratic forms in normally distributed random variables - *Sankhya* - Vol. 17 (1956).

[6] ISARD (W.) – *Methods of regional analysis. An introduction to regional science* (The MIT Press - Cambridge - Massachussett).

[7] JORESKOG (K.G.) – *Statistical estimation in factor analysis.* A new techrique and its foundations (Almqvist & Wiksell-Stockolm, Uppsala - 1963).

[8] LEDERMANN (S.) – *Cours d'analyse multivariate,* 1966, I.S.U.P. (Ronéo).

[9] MALINVAUD (E.) –*Méthodes statistiques de l'économétrie.* Dunod 1964.

[10] TRANQUILLI (G.B.) – *Les lois jointes des statistiques et de l'écart carré moyen des échantillons normaux* (thèse 1963).

[11] Von NEUMANN (J.) – Distribution of the ratio of the mean square successive difference to the variance (*Annals of math. stat* Vol. 12 - 1941).

[12] *Cahiers de l'I.S.E.A.* – Série L – N° 1 à 13, en particulier n° 10 – Bibliographie de science économique régionale.

[13] *Bulletin de l'Institut International de Statistique* - 1958 - Tome 36 (Divers articles sur l'économie régionale).

[14] «L'espace économique français» (1955, 1962) et diverses remises à jour. *(n° spécial de la revue « Etudes et conjoncture »).*