LES INDICES D'AUTO-CORRELATION SPATIALE

L'autocorrélation spatiale est l'absence d'indépendance entre observations géographiques. Ainsi, on constate très souvent que les variables spatialisées sont soumises à des dépendances spatiales (ou interactions spatiales), qui sont d'autant plus fortes que les localisations sont plus proches : "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970).

Les mesures d'autocorrélation spatiale permettent d'estimer la dépendance spatiale entre les valeurs d'une même variable en différents endroits de l'espace. Pour la mettre en évidence, les indices prennent en compte deux critères : la proximité spatiale et la ressemblance ou la dissemblance des valeurs de cette variable dans les unités spatiales de la zone d'étude.

On fait la distinction entre la mesure de l'autocorrélation spatiale globale d'une variable dans un territoire donné et celle de l'autocorrélation locale dans chaque unité spatiale. Cette dernière correspond à l'intensité et la significativité de la dépendance locale entre la valeur d'une variable dans une unité spatiale et les valeurs de cette même variable dans les unités spatiales environnantes (plus ou moins proches).

1. Interprétation des indices

Dans OASIS, nous proposons le calcul des indices d'autocorrélation spatiale pour les données en valeurs absolues, mais aussi en valeurs relatives. Les indices les plus utilisés sont l'indice de Moran et l'indice de Geary. Dans la littérature, le coefficient de Moran est souvent préféré à celui de Geary en raison d'une stabilité générale plus grande (Upton et Fingleton, 1985). Avant de calculer ces indices, il est nécessaire de définir les interactions spatiales prises en compte.

a) La matrice d'interactions spatiales

Comme son nom l'indique, la matrice d'interactions spatiales est une matrice carrée qui permet de mesurer les interactions entre les unités spatiales, indépendamment de la variable étudiée. Le nombre de lignes et de colonnes est égal au nombre d'unités spatiales dans la zone d'étude. Sur la diagonale la matrice a des valeurs nulles. Chaque élément de la matrice mesure l'interaction entre les unités spatiales correspondante en ligne et en colonne. Sa valeur est d'autant plus élevée que la « proximité » spatiale entre les unités est forte.

Pour utiliser une matrice d'interactions spatiales, il est nécessaire de donner a priori une forme fonctionnelle spécifique aux interactions spatiales. C'est pour cette raison que nous proposons dans OASIS les trois formes les plus utilisées dans la littérature :

- La matrice de contiguïté : généralement, les éléments de la matrice de contigüité d'ordre k sont définis comme le nombre minimal de frontières qu'il faut franchir pour aller d'une unité spatiale à l'autre. La matrice de contiguïté la plus utilisé est d'ordre 1 : les éléments de la matrice sont égaux à 1 si deux unités sont voisines et sont nuls dans le cas contraire. C'est la forme la plus simple d'interactions spatiales, souvent suffisante pour déterminer les niveaux autocorrélation spatiales d'une variable.

- La matrice des distances est une mesure d'intensité spatiale qui peut être utilisée quand les données sont spatialisées sous forme de points. Lorsqu'on étudie des unités spatiales, la distance prise en compte est généralement la distance entre leurs centroïdes. Cette distance peut être intégrée sous différentes formes : linéaire, quadratique ou logarithmique. Dans OASIS, la matrice d'interactions spatiales basée sur la distance utilise les distances linéaires.

- Une autre manière plus détaillée de décrire les interactions spatiales consiste à prendre en compte non seulement le voisinage (des valeurs 0 ou 1), mais de tenir compte aussi de l'intensité de cette proximité spatiale, qui dépend de l'importance de la frontière commune. Les éléments de la matrice sont ainsi des coefficients proportionnels à longueur de la frontière commune.

b) L'indice I de Moran

L'indice de Moran (Moran 1950) permet de mesurer le niveau d'autocorrélation spatiale d'une variable et de tester sa significativité. Il est égal au ratio de la covariance entre observations contigües (définies par la matrice d'interactions spatiales) à la variance totale de l'échantillon (Jayet, 2001).

L'indice a des valeurs comprises entre -1 (indiquant une dispersion parfaite) à 1 (corrélation parfaite). Une valeur nulle signifie que la distribution spatiale de la variable étudiée est parfaitement aléatoire dans le territoire. Les valeurs négatives (positives) de l'indice indiquent une autocorrélation spatiale négative (positive).

Pour le test d'hypothèse statistique, l'indice I de Moran peut être transformé en Z-scores, pour lesquels les valeurs plus grandes que le seuil de significativité positive ou plus petites que le seuil de significativité négative indiquent une autocorrélation spatiale significative, avec un taux d'erreur correspondant au seuil (hypothèse de loi normale).

Une interprétation facile de l'indice de Moran peut être faite avec l'aide du diagramme de Moran, qui représente, sous la forme d'un nuage de points, les couples de valeurs correspondant à la valeur de la variable dans chaque unité spatiale (en abscisse) et la moyenne des valeurs des zones contiguës (en ordonnée) définies par la matrice d'interactions spatiale (cette moyenne est appellée spatial lag ou décalage spatial). Dans OASIS, les unités spatiales où l'autocorrélation spatiale n'est pas significative sont représentées sous la forme de cercles, tandis que les unités avec une autocorrélation locale statistiquement significative sont en forme de losange, avec en libellé le nom de l'unité spatiale. Le graphique est complété avec la droite de pente égale à l'indice de Moran. Une pente négative signifie une autocorrélation spatiale négative tandis qu'une pente positive signifie une autocorrélation spatiale positive.

c) L'indice de Geary

Une autre alternative pour mesurer l'autocorrélation spatiale est l'indice de Geary qui est, à un facteur ½ près, égal au ratio de la variance des écarts entre observations contigües à la variance totale (Jayet, 2001). Si l'indice de Moran est une mesure de l'autocorrélation spatiale globale, l'indice de Geary est plus sensible à l'autocorrélation spatiale locale.

L'indice de Geary varie de 0 à l'infini et vaut 1 s'il y a indépendance spatiale. Pour toute valeur inférieur à l'unité, il y a une autocorrélation spatiale positive, et inversement pour des valeurs supérieures à l'unité. L'indice de Geary varie en sens inverse de l'indice de Moran. Comme pour l'indice de Moran, l'indice de Geary peut être testé statistiquement, par une transformation en Z-scores et en déterminant son seuil de significativité.

d) L'autocorrélation spatiale locale (LISA)

L'analyse de l'autocorrélation spatiale locale a été introduite par Anselin (1995). L'indice Moran local permet de mesurer le degré de corrélation spatiale au niveau local pour chaque unité spatiale. Comme dans le cas de l'indice de Moran, on peut calculer les Z-scores et tester la significativité du degré d'autocorrélation spatiale locale. Les enregistrements significatifs peuvent être représentés sous formes de cartes.

2. Calcul des indices d'autocorrélation spatiale

L'indice de Moran :

$$I_{Moran} = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(x_i - \overline{x}\right)\left(x_j - \overline{x}\right)}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}$$

où *i,j* = unité spatiale ; *n* = nombre d'unités spatiales ; x_i est la valeur de la variable dans l'unité *i* ; $\overline{x}$ est la moyenne de *x* ; et w_{ij} sont les éléments de la matrice d'interactions spatiales, définie sous la forme de la contigüité, les distances ou les frontières communes.

Pour le test d'hypothèse statistique nulle (pas d'autocorrélation spatial), l'indice I de Moran peut être transformé en Z-scores (les valeurs critiques) et p-value (la significativité).L'espérance mathématique de l'indice de Moran (hypothèse de non autocorrélation spatiale) est:

$$E\left(I_{Moran}\right) = -\frac{1}{n-1}$$

Et la variance est égale à :

$$V\left(I_{Moran}\right) = E\left(I_{Moran}{}^2\right) - \left(E\left(I_{Moran}\right)\right)^2$$

Ainsi, la valeur critique z-score est égale à :

$$z_I = \frac{I_{Moran} - E\left(I_{Moran}\right)}{\sqrt{V\left(I_{Moran}\right)}}$$

La valeur de p-value est l'approximation numérique de la superficie au-dessous de la courbe de la distribution de moyenne et variances connues.

L'indice de Geary

$$C_{Geary} = \frac{(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(x_i - x_j\right)^2}{2\left(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\right)\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$$

où *i,j* = unité spatiale ; *n* = nombre d'unités spatiales ; x_i , x_j sont les valeurs de la variable dans l'unité *i* respectivement *j* ; $\bar{x}$ est la moyenne de x ; et w_{ij} sont les éléments de la matrice d'interactions spatiales, définie sous la forme de la contigüité, les distances ou les frontières communes. Comme pour l'indice de Moran, on peut tester l'indice de Geary par transformation en z-scores.

L'indice de Moran local

$$I_i = \frac{(n-1)\left(x_i - \bar{x}\right)\sum_{j=1}^{n} w_{ij}\left(x_j - \bar{x}\right)}{\sum_{j=1}^{n}\left(x_j - \bar{x}\right)^2}$$

Pour tester statistiquement l'indice local, l'espérance et la variance de l'indice a été proposé par Anselin (1995).

Bibliographie

ANSELIN, L. (1995) : Local indicators of spatial association, Geographical Analysis, 27, 93–115

ANSELIN, L. (1996) The Moran scatterplot as an ESDA tool to assess local instability in spatial association. pp. 111–125 in M. M. Fischer, H. J. Scholten and D. Unwin (eds) Spatial analytical perspectives on GIS, London, Taylor and Francis;

GEARY R. C. (1954): The contiguity ratio and statistical mapping, The Incorporated Statistician, 5, pp. 115-145.

JAYET H. (2001): Econométrie et données spatiales, Cahiers d'économie et sociologie rurales, 58-59

MORAN P. A. P. (1950): A test for serial dependence of residuals, Biometrika, 37, pp. 178-181.

TOBLER W.R. (1970): A computer movie simulating urban growth in the Detroit region, Economic geography, Supplement 46, , 234-40.

UPTON G. and FINGLETON B. (1985): Spatial data analysis by example, New York, Wiley