

Exercice 4.3 – Sur l'indice de Moran

Géographie quantitative – Autocorrélation spatiale

AGNANGMA SANAM David Landry

ISE1 CYCLE LONG

Énoncé : Montrer que si la matrice \mathbf{W} est normalisée en ligne et si le vecteur \mathbf{X} est centré, le vecteur spatialement décalé \mathbf{WX} est également centré. En déduire que la pente de la droite de régression simple de \mathbf{WX} sur \mathbf{X} est égale à l'indice de Moran si la matrice \mathbf{W} est normalisée en ligne.

Partie 1 : Centrage du vecteur spatialement décalé

Définition 1 (Matrice normalisée en ligne). Une matrice $\mathbf{W} = (w_{ij})$ de taille $n \times n$ est dite **normalisée en ligne** si pour tout $i = 1, \dots, n$:

$$\sum_{j=1}^n w_{ij} = 1$$

Définition 2 (Vecteur centré). Un vecteur $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$ est dit **centré** si :

$$\sum_{i=1}^n X_i = 0$$

Preuve (Que \mathbf{WX} est centré). Soit \mathbf{X} un vecteur centré et \mathbf{W} une matrice normalisée en ligne.

La i -ème composante du vecteur spatialement décalé est :

$$(\mathbf{WX})_i = \sum_{j=1}^n w_{ij} X_j$$

Calculons la somme des composantes de \mathbf{WX} :

$$\begin{aligned} \sum_{i=1}^n (\mathbf{WX})_i &= \sum_{i=1}^n \sum_{j=1}^n w_{ij} X_j \\ &= \sum_{j=1}^n X_j \left(\sum_{i=1}^n w_{ij} \right) \\ &= \sum_{j=1}^n X_j \cdot S_j \end{aligned}$$

où $S_j = \sum_{i=1}^n w_{ij}$ est la somme de la colonne j de \mathbf{W} .

Si \mathbf{W} est **symétrique** (cas fréquent en analyse spatiale), alors $S_j = 1$ pour tout j (car chaque colonne est aussi normalisée), d'où :

$$\sum_{i=1}^n (\mathbf{WX})_i = \sum_{j=1}^n X_j = 0$$

Ce qui prouve que \mathbf{WX} est centré.

Remarque 1. Dans le cas général d'une matrice non symétrique normalisée en ligne, la somme $\sum_{i=1}^n (\mathbf{WX})_i$ n'est pas nécessairement nulle. Cependant, pour la plupart des matrices de contiguïté utilisées en pratique, cette propriété reste vérifiée.

Partie 2 : Lien avec l'indice de Moran

Rappel : Pente de la régression linéaire simple

Pour deux vecteurs centrés \mathbf{Y} et \mathbf{X} , la pente de la régression de \mathbf{Y} sur \mathbf{X} est :

$$\hat{\beta} = \frac{\text{Cov}(\mathbf{X}, \mathbf{Y})}{\text{Var}(\mathbf{X})} = \frac{\frac{1}{n} \sum_{i=1}^n X_i Y_i}{\frac{1}{n} \sum_{i=1}^n X_i^2} = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}$$

Application à notre cas

Posons $\mathbf{Y} = \mathbf{WX}$. D'après la Partie 1, \mathbf{Y} est centré lorsque \mathbf{W} est symétrique et normalisée en ligne. Alors :

$$\begin{aligned} \hat{\beta} &= \frac{\sum_{i=1}^n X_i (\mathbf{WX})_i}{\sum_{i=1}^n X_i^2} \\ &= \frac{\sum_{i=1}^n X_i \left(\sum_{j=1}^n w_{ij} X_j \right)}{\sum_{i=1}^n X_i^2} \\ &= \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} X_i X_j}{\sum_{i=1}^n X_i^2} \end{aligned}$$

Indice de Moran

Définition 3 (Indice de Moran). Pour un vecteur \mathbf{X} et une matrice de poids \mathbf{W} , l'indice de Moran est défini par :

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \cdot \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

où $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ est la moyenne de \mathbf{X} .

Égalité des deux expressions

Si \mathbf{X} est centré ($\bar{X} = 0$) et si \mathbf{W} est normalisée en ligne :

$$\begin{aligned} I &= \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \cdot \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} X_i X_j}{\sum_{i=1}^n X_i^2} \\ &= \frac{n}{n} \cdot \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} X_i X_j}{\sum_{i=1}^n X_i^2} \quad (\text{car } \sum_{i=1}^n \sum_{j=1}^n w_{ij} = \sum_{i=1}^n 1 = n) \\ &= \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} X_i X_j}{\sum_{i=1}^n X_i^2} \\ &= \hat{\beta} \end{aligned}$$

Conclusion

Résultat principal :

Si la matrice de voisinage \mathbf{W} est normalisée en ligne et symétrique, et si le vecteur \mathbf{X} est centré, alors :

1. Le vecteur spatialement décalé \mathbf{WX} est également centré
2. La pente $\hat{\beta}$ de la droite de régression simple de \mathbf{WX} sur \mathbf{X} est égale à l'indice de Moran I

Remarque 2 (Interprétation géographique). Ce résultat montre que l'indice de Moran peut être interprété comme le coefficient de régression du retard spatial sur la variable elle-même. Une valeur positive de I indique une autocorrélation spatiale positive (similarité entre voisins), tandis qu'une valeur négative indique une dissimilarité.