

LES TESTS DE MONTE CARLO

Les simulations de Monte Carlo permettent de tester l'hypothèse que la valeur d'un indicateur est le résultat d'un processus aléatoire. Si cette hypothèse est rejetée, alors la valeur de cet indicateur est jugée significative.

Le principe de base est relativement simple : les valeurs d'une variable sont distribuées de manière « aléatoire » (suivant un critère de répartition aléatoire choisi) entre les unités spatiales, afin de pouvoir calculer une valeur simulée de l'indicateur. Les simulations sont répétées un nombre suffisamment élevé de fois pour que les valeurs simulées puissent être approximées par une loi de distribution. Cette distribution des valeurs simulées sera utilisée comme base de comparaison pour la valeur « observée » de l'indicateur. Si la valeur « observée » se trouve à l'intérieur des valeurs critiques données par les seuils de significativité, alors on peut conclure qu'elle est non significative. Au contraire, si la valeur « observée » de l'indicateur se trouve en dehors des valeurs critiques, elle n'apparaît pas comme étant le résultat d'un processus aléatoire et peut donc être considérée comme significative.

Deux types de simulations sont faites, en fonction du critère de répartition aléatoire choisi. Le premier permet de tester plus spécifiquement si la composante spatiale de l'indicateur est significative ou non : il est basé sur le test des permutations des données, proposé par Anselin (1995) pour les indices d'autocorrélation spatiale et utilisé par Feitosa et al. (2007) pour les indices de ségrégation. Le second permet de tester la significativité globale de l'indicateur : il est basé sur des localisations aléatoires avec ou sans contrainte.

1. Test des permutations

Pour le test des permutations, on garde la même structure des données, mais pour chaque simulation les données sont permutées de manière aléatoire entre les unités spatiales. Cette méthode permet ainsi de tester si la composante spatiale de l'indicateur est significative ou non.

Dans le cas des indicateurs d'autocorrélation spatiale, la composante spatiale est donnée par la matrice de décalage spatial. Dans OASIS, l'utilisateur peut choisir la définition des interactions spatiales qui lui paraît la plus appropriée parmi les modalités proposées : matrice de contiguïté, matrice de distances et matrice des frontières communes.

Dans le cas des indices de ségrégation, la composante spatiale des indicateurs est incluse dans leur définition. Certains prennent en compte la forme des unités spatiales (périmètres et superficies) ou



OASIS

Outil d'Analyse de la Ségrégation et des Inégalités Spatiales

les interactions spatiales entre unités (distances, frontières communes, contiguïté). Les indices qui ne tiennent pas compte de manière explicite des caractéristiques de l'espace sont insensibles aux tests des permutations et ont toujours la même valeur simulée, égale à l'indice observé.

Une fois les simulations avec données permutées réalisées, l'indicateur observé peut être comparé avec la distribution des valeurs simulées : son rang dans cette distribution permet de calculer une pseudo p-value et de conclure quant à la significativité de la composante spatiale de l'indicateur.

2. Test des localisations aléatoires

Les tests des localisations aléatoires permettent de déterminer la significativité globale de l'indice. Les localisations sont simulées sans contrainte (probabilité identique pour chaque unité spatiale) ou avec contraintes spatiales (probabilités déterminées par différents critères : la superficie totale des unités spatiales, ou leur superficie artificialisée). En comparant la distribution des valeurs simulées avec la valeur calculée de l'indice, nous pouvons conclure qu'il est le résultat d'un processus aléatoire ou d'un schéma de localisation déterminé. Si l'indice calculé se trouve à l'extérieur de la distribution (au-dessus du seuil de significativité), il n'apparaît pas comme étant le résultat d'un processus aléatoire et peut donc être considéré comme significatif.