

Autocorrélation spatiale

Université Paris-Est Créteil
Serge Lhomme

Maître de conférences en géographie
<http://sergelhomme.fr/>
serge.lhomme@u-pec.fr

L'autocorrélation spatiale

Présentation

Compte tenu du caractère inégalitaire de nombreuses distributions, des objets géographiques se ressemblent plus que d'autres. Une question d'analyse spatiale que l'on est alors en droit de se poser est la suivante :

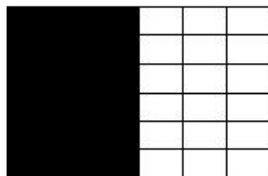
Est-ce que les objets géographiques qui sont proches se ressemblent plus que les objets géographiques qui sont éloignés ? C'est la question de l'autocorrélation spatiale.

Mesurer l'autocorrélation spatiale d'un phénomène (d'une distribution) revient à déterminer s'il semble exister une organisation spatiale sous-jacente à ce phénomène (à cette distribution) et donc qu'il (qu'elle) ne se répartit pas de façon aléatoire au sein du territoire étudié.

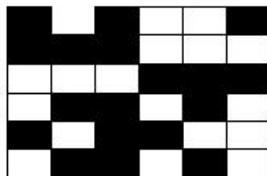
Par exemple : Les personnes riches se regroupent-elles ? Les communes très peuplées côtoient-elles des communes très peu peuplées ?

L'autocorrélation spatiale

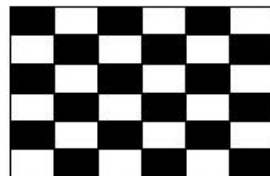
Présentation



*Autocorrélation spatiale
positive*



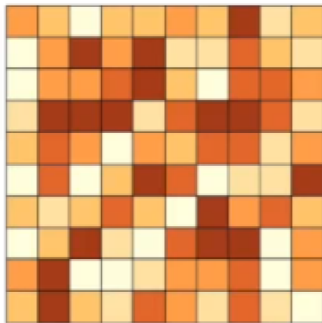
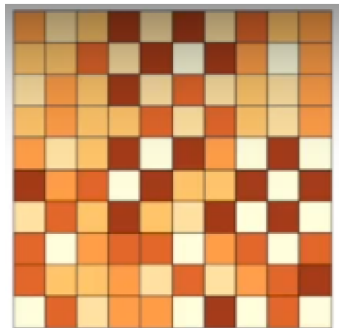
*Autocorrélation spatiale
nulle*



*Autocorrélation spatiale
négative*

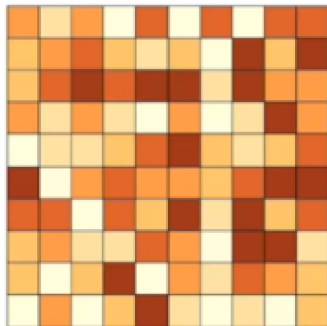
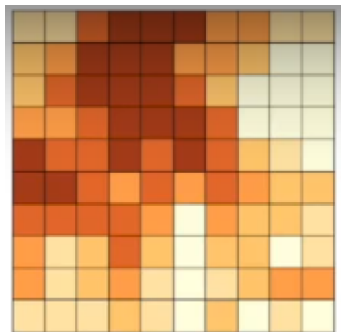
L'autocorrélation spatiale

Présentation : aléatoire ou pas



L'autocorrélation spatiale

Présentation : aléatoire ou pas



L'autocorrélation spatiale

Présentation statistique

Les coefficients d'autocorrélation spatiale sont alors construits statistiquement de telle manière qu'il soit possible de répondre à la question suivante :

La variation d'un caractère entre unités voisines (proches) est-elle plus ou moins grande que la variation de ce même caractère pour l'ensemble du territoire ou plus précisément entre unités non-voisines (éloignées) ?

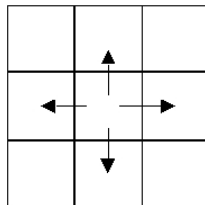
Il convient dès lors de définir ce qui est proche, de définir ce qui est voisin. Le plus simple est de le déterminer de manière binaire en s'appuyant par exemple sur la notion de contiguïté.

Il existe plusieurs indicateurs pour mesurer l'autocorrélation spatiale. Les deux principaux, c'est-à-dire les plus couramment utilisés, sont les indices de Moran et de Geary.

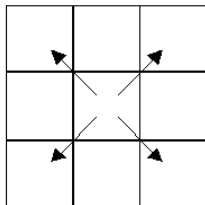
L'autocorrélation spatiale

La contiguïté

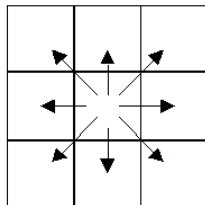
Rooks Case



Bishops Case



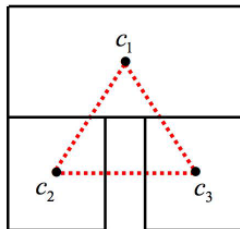
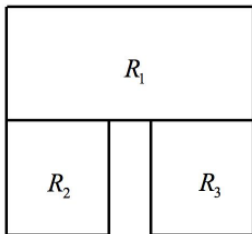
Queen's (Kings) Case



L'autocorrélation spatiale

La proximité

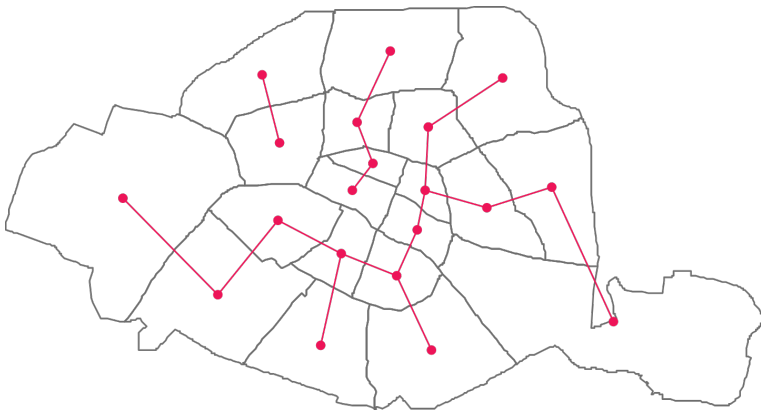
Comme on travaille souvent avec des entités géographiques zonales, on passera souvent par les centroïdes pour évaluer la proximité, ici les trois entités sont alors équidistantes.



L'autocorrélation spatiale

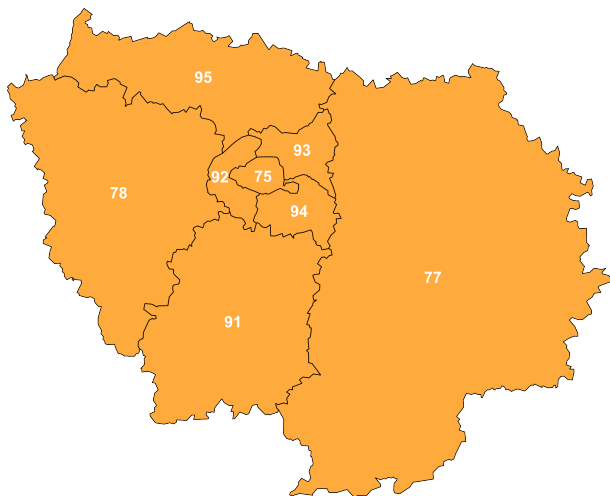
La proximité

A partir de ces centroïdes, on peut alors créer des graphes de voisinage fondés sur la proximité, par exemple au plus proche voisin.



L'autocorrélation spatiale

Contiguïté Vs Proximité



L'autocorrélation spatiale

Les indices de Moran et Geary

$$G = \frac{N-1}{2L} \times \frac{\sum_{i,j} l_{ij} \times (X_i - X_j)^2}{\sum_i (X_i - \bar{X})^2}$$

Les valeurs de l'indice de Geary s'étendent de 0 à 2. La valeur 1 signifie qu'aucune autocorrélation spatiale n'est présente dans les mesures effectuées. Une valeur plus petite que 1 signifie une autocorrélation spatiale positive.

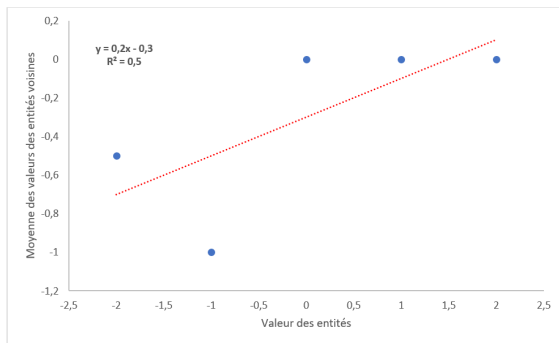
$$I = \frac{N}{L} \times \frac{\sum_{i,j} l_{ij} \times (X_i - \bar{X}) \times (X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}$$

Les valeurs de l'indice de Moran s'étendent de -1 (corrélation négative) à +1 (corrélation positive). Une valeur nulle correspond à un modèle spatial parfaitement aléatoire.

L'autocorrélation spatiale

Exemple Moran : Le diagramme de Moran

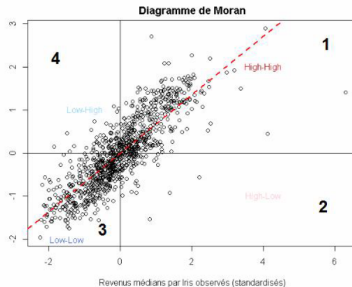
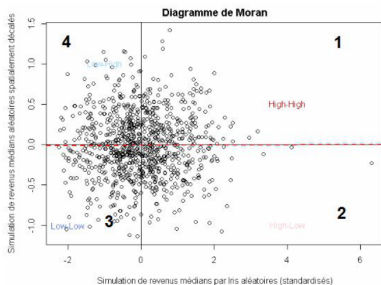
Derrière sa formulation mathématique qui peut paraître peu intuitive, l'indice de Moran revient simplement à mesurer la pente de la relation linéaire entre les valeurs prises par les entités géographiques et leurs entités voisines ou proches (en prenant pour celles-ci une valeur moyenne). Le diagramme de Moran est alors fondamental.



L'autocorrélation spatiale

Exemple Moran : Le diagramme de Moran

On aura tendance à présenter le diagramme de Moran avec des valeurs standardisées et les moyennes correspondantes.



L'autocorrélation spatiale

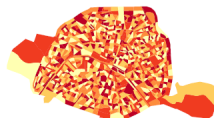
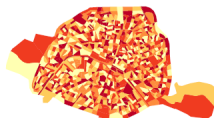
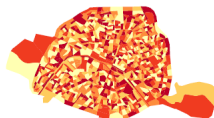
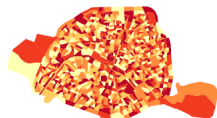
Exemple Moran : Le diagramme de Moran

$\rho = -0.1$

$\rho = -0.3$

$\rho = -0.6$

$\rho = -0.9$

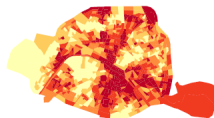
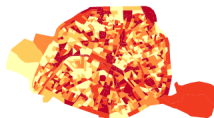
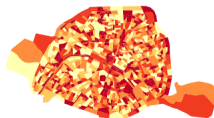
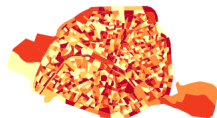


$\rho = 0.1$

$\rho = 0.3$

$\rho = 0.6$

$\rho = 0.9$

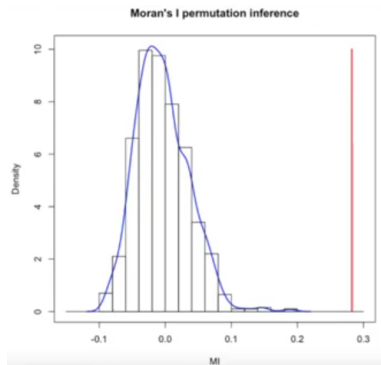


L'autocorrélation spatiale

Aléatoire ou pas : la p-value

La p-value calcule la probabilité que la valeur d'autocorrélation soit obtenue par une distribution aléatoire de la variable étudiée.

Il existe deux manières de calculer la p-value : de manière analytique ou par simulation (des permutations).



999 permutations et une valeur de $i = 0.289$

L'autocorrélation spatiale

Aléatoire ou pas : la p-value

Attention, il ne faut pas confondre p-value et z-value.

La z-value est une valeur qui peut être comparée à d'autres z-value, contrairement au i de Moran qui ne sont pas comparables.

Elle est dépendante de la manière dont on calcule la p-value...

L'autocorrélation spatiale

La pondération spatiale

La mesure de l'auto-corrélation spatiale requiert une précaution essentielle, ne pas accorder trop d'importance à des entités géographiques qui auraient plus de voisins que les autres.

En effet, à partir des graphes de voisinage il est facile de produire une matrice de poids. Néanmoins, la somme de chaque ligne traduit alors l'importance accordée à chaque entité. Une matrice de poids devra toujours être normalisée.

	a	b	c	d	e	Somme des poids des voisins
a	0	1	1	0	0	2
b	1	0	0	1	0	2
c	1	0	0	1	0	2
d	0	1	1	0	1	3
e	0	0	0	1	0	1

L'autocorrélation spatiale

Normalisation de la matrice de poids

Il existe alors plusieurs méthodes de normalisation :

- Normalisation en ligne (schéma de codage "W") : pour une zone, le poids accordé à chaque voisin est divisé par la somme des poids de ses voisins. Cette standardisation facilite l'interprétation de la matrice de poids, puisqu'en fin de compte on obtient la moyenne de la variable x calculée sur tous les voisins de l'observation i .
- Normalisation globale (schéma de codage "C") : les poids sont standardisés de sorte que la somme de tous les poids soit égale au nombre total d'entités.
- Normalisation par stabilisation de la variance (schéma de codage "S") : elle permet de réduire l'hétérogénéité dans les poids liée aux différences de taille et de nombre de voisins entre les zones. En effet, la normalisation en ligne donne plus de poids aux observations situées en bordure de la zone d'étude, avec un faible nombre de voisins.

L'autocorrélation spatiale

L'autocorrélation spatiale locale : les LISA

Les statistiques globales font l'hypothèse de **stationnarité** du processus spatial : l'autocorrélation spatiale serait la même dans tout l'espace. Or, cette hypothèse est d'autant moins réaliste que le nombre d'observations est élevé. Les données spatiales sont souvent caractérisées par de **l'hétérogénéité spatiale**.

Le diagramme de Moran montre même qu'il existe des endroits où l'autocorrélation aurait tendance à être négative alors même que la tendance globale est positive et inversement.

Deux questions se posent alors : Comment mesurer une valeur d'autocorrélation spatiale locale ? Ces valeurs sont-elles significatives ?

Pour chaque observation, ces indicateurs indiquent l'intensité du regroupement de valeurs similaires (ou de tendance opposée) autour de cette observation et la somme des indices locaux sur l'ensemble des observations doit être proportionnelle à l'indice global correspondant.

L'autocorrélation spatiale

L'autocorrélation spatiale locale : les LISA

Concrètement, il est possible pour chaque lieu de calculer uniquement le numérateur de l'indicateur de Moran pour le décliner localement.

$I_i > 0$ indique un regroupement de valeurs similaires (plus élevées ou plus faibles que la moyenne). $I_i < 0$ indique un regroupement de valeurs dissimilaires (des valeurs élevées entourées de valeurs faibles).

$$I = \frac{N}{L} \times \frac{\sum_i \sum_j l_{ij} \times (X_i - \bar{X}) \times (X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}$$

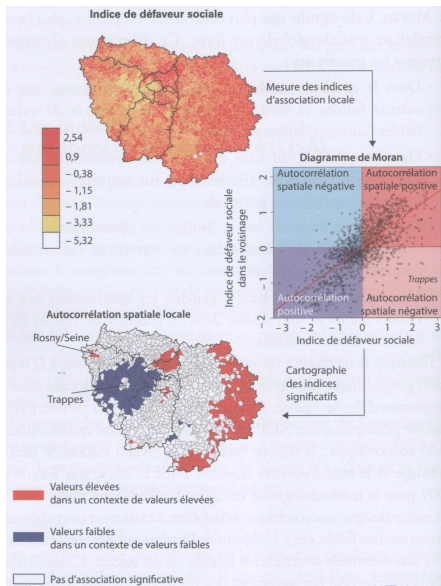
$$I_i = (X_i - \bar{X}) \sum_j l_{ij} \times (X_j - \bar{X})$$

Attention au calcul de la p-value

S'il y a 100 indices d'autocorrélation spatiale locaux, on multiplie par 100 le risque d'en détecter au moins un significatif à tort. Différentes méthodes ont été développées pour éviter cette inflation du risque α .

L'autocorrélation spatiale

L'autocorrélation spatiale locale : les LISA



L'autocorrélation spatiale

L'autocorrélation spatiale locale : Getis and Ord

L'indice de Getis-Ord est un autre moyen simple de détecter des clusters de valeurs fortes (hot spot) ou faibles (cold spot).

Pour cela, cet indice va simplement rapporter pour chaque observation, la valeur moyenne dans le voisinage à la somme des valeurs totales.

Une fois standardisée en un Z-score les valeurs négatives pourront être considérées comme des cold spot (notamment celles inférieures à 1,96) et les valeurs positives comme des hot spot (notamment celles supérieures à 1,96).

Cet indice présente l'inconvénient de ne pas détecter l'autocorrélation spatiale négative.

Cet indice est globalement moins robuste que les i locaux (LISA).