

Diplomado en Análisis de Información Geoespacial

Patrones puntuales

Autor:
M. en G. Camilo A. Caudillo Cos

Introducción

Uno de los objetivos principales del análisis de datos puntuales es detectar patrones. En particular, un conjunto de posiciones observadas que contienen aglomeraciones de eventos.

Asimismo, en el análisis espacial buscamos definiciones matemáticas para identificar patrones en la distribución espacial de un conjunto de lugares. Uno de los patrones de interés más común muestra la presencia o tendencia a la aglomeración.

Un punto de partida útil para abordar este tema implica distinguir entre un punto y un evento. Para ello, Diggle (1983) utiliza las siguientes definiciones:

Un punto es cualquier posición en un área de estudio donde un evento pudo suceder; y la posición de un evento resulta particular en donde éste ocurrió efectivamente.

Por otra parte, un conjunto de datos consiste en una colección de posiciones de eventos observados y en un dominio espacial de interés. El dominio espacial es un aspecto importante de los datos que puede definirse a partir de su disponibilidad (agregaciones de diferentes escalas, municipios, estados, ciudades, etc.), aunque en muchos casos el analista debe definir las fronteras más adecuadas.

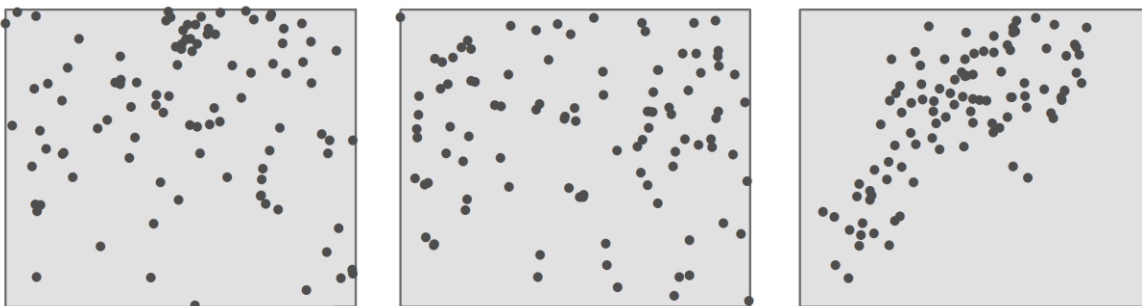


Figura 1. Camilo Caudillo (2010). Muestras de patrones puntuales. Elaboración propia.

La figura 1 muestra 100 eventos de diferentes fenómenos en los que a simple vista parecen existir aglomeraciones. Los métodos de análisis espacial tienen la finalidad de caracterizar los patrones extremos entre lo disperso y lo aglomerado.

Para definir el concepto aglomeración es más fácil comenzar por lo que no es, y para esto necesitamos considerar lo que significa un patrón "aleatorio".

En casi cualquier aplicación de análisis espacial, los conceptos patrón aleatorio, aleatoriamente o por azar se refieren a que la distribución de los eventos no está influenciada por los factores que se investiga.

El primer modelo de patrón relevante es la aleatoriedad espacial completa, definida como la misma probabilidad de que ocurra un evento en cualquier posición dentro del área de estudio, independientemente de las posiciones de otros eventos.

Esto es, los eventos se distribuyen uniformemente a través del área de estudio y son independientes unos de otros. La palabra uniforme en este caso implica una distribución de probabilidad uniforme, no que se distribuyen equitativamente en el área de estudio, ver figura 2.

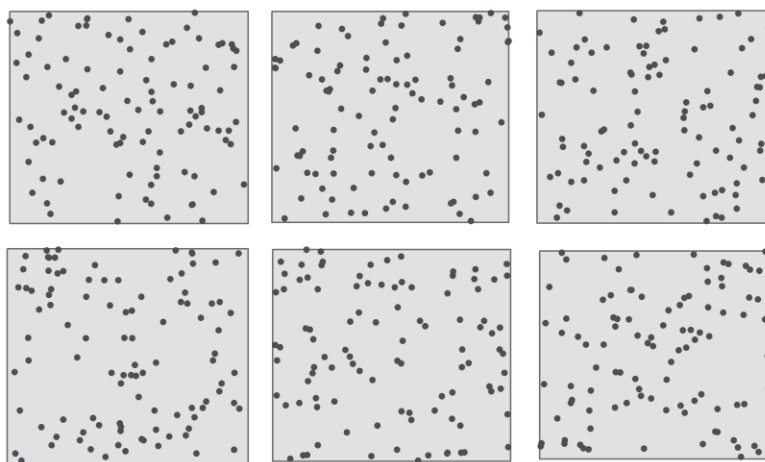


Figura 2. Camilo Caudillo (2010). Seis realizaciones de un proceso de aleatoriedad espacial completo. Elaboración propia.

La aleatoriedad no implica uniformidad espacial pues aunque hay puntos que se generaron cerca de otros, esto no viola el principio de aleatoriedad completa.

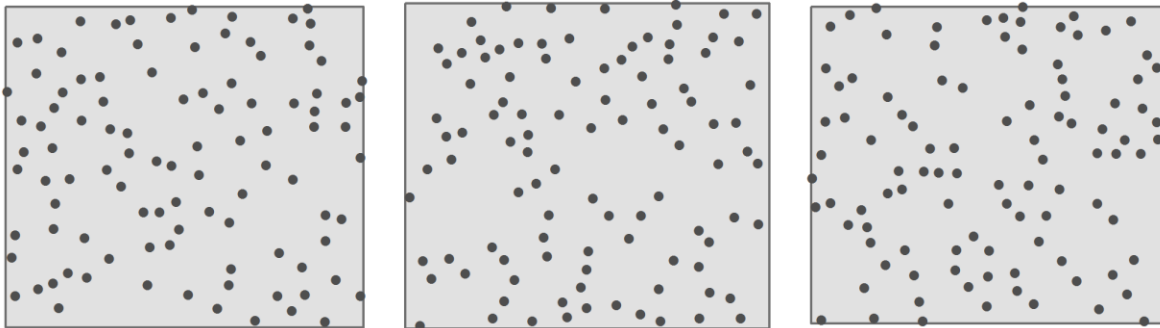


Figura 3. Camilo Caudillo (2010). Tres realizaciones de un proceso aleatorio uniforme.
Elaboración propia.

En el caso ilustrado en la figura 3, la distancia mínima entre dos eventos es una restricción. El proceso aleatorio uniforme, así como los procesos Poisson heterogéneos, pueden ser puestos a prueba de hipótesis con la ayuda de funciones de densidad auxiliares.

Estacionariedad e isotropía

Estos dos conceptos son importantes para el modelado de procesos puntuales. Matemáticamente, un proceso es estacionario cuando no varía al transformarlo dentro de un espacio d -dimensional, e isotrópico cuando no varía por la rotación del punto de origen.

En otras palabras, la relación entre dos eventos en un proceso estacionario depende únicamente de sus posiciones relativas, no de las posiciones de los eventos en sí mismas.

Si a partir de un proceso estacionario se incorpora el supuesto de isotropía, esto implica

que la restricción de las relaciones entre dos eventos depende únicamente de la distancia que los separa y no de la orientación entre ellos. Sin embargo, ninguna de estas propiedades implica a la otra.

Estas propiedades muestran el principio de replicación de un conjunto de datos. Por ejemplo, dos pares de eventos en la realización de un proceso estacionario, separados por la misma distancia y dirección relativa, deberán estar sujetos al mismo tipo de relación. De igual modo, dos pares de eventos de una realización de un proceso isotrópico y estacionario, separados por la misma distancia, deberán exhibir las mismas propiedades.

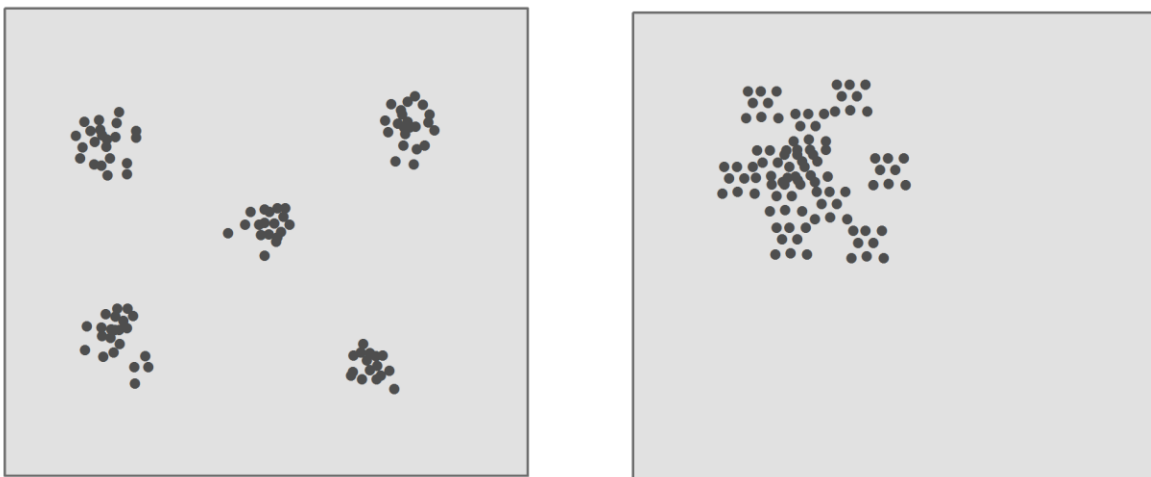


Figura 4. Camilo Caudillo (2010). Ejemplos de procesos puntuales hipotéticos que contienen aglomeración y regularidad en escalas diferentes. Elaboración propia.

En la figura 4 a la izquierda aparece un patrón regular de aglomeraciones y a la derecha, una aglomeración de patrones regulares.

Pruebas de hipótesis de aleatoriedad espacial completa a partir de métodos de inferencia Monte Carlo

Para detectar una aglomeración (o regularidad) estadísticamente, necesitamos comparaciones respecto a un proceso aleatorio. Es decir, se requiere un método que describa qué tanta variación esperaríamos, partiendo de un proceso de aleatoriedad espacial completa, para saber cuándo un patrón observado se diferencia significativamente del proceso aleatorio.

Las pruebas de hipótesis y la simulación Monte Carlo son las técnicas más útiles para esa valoración.

Las pruebas de hipótesis comparan típicamente el valor de un resumen cuantitativo de los datos observados (estadístico de prueba) con la distribución de probabilidad de dicho resumen bajo el supuesto de una hipótesis nula.

En el caso de los procesos espaciales, para evaluar la noción frecuentista de la comparación de un valor observado con su distribución bajo la hipótesis nula y combinada con la sencillez de la simulación de aleatoriedad espacial completa, se sugiere el uso de métodos de inferencia Monte Carlo (basados en simulación).

Para ello, primero se calcula el estadístico de prueba basado en los datos observados y después, el mismo estadístico para un número grande de datos simulados independientemente, bajo la hipótesis nula (aleatoriedad espacial completa).

Con ello se genera un histograma del estadístico obtenido por simulación, lo cual nos da la aproximación al estadístico de prueba bajo la hipótesis nula.