

Parte I

Introducción a la estadística espacial

- En términos muy generales podemos decir que datos espaciales son todos aquellos que vienen asociados a una localización del espacio. Cuando la distribución espacial de los datos es importante para su estudio e interpretación, la aplicación de técnicas específicas para datos espaciales cobra importancia puesto que puede proporcionar mayor información que las técnicas tradicionales.

Un primer ejemplo de datos espaciales podría ser el estudio de la temperatura en superficie, donde cada dato debe venir asociado a una coordenada tridimensional. Es fácil darse cuenta de que realizar un estudio de la temperatura sin tener en cuenta el lugar en donde se ha medido carece de sentido.

Las bases de datos espaciales deben contener observaciones de una (o varias) variable estadística de interés y una referencia cartográfica. En muchas ocasiones están asociadas a Sistemas de Información Geográfica (SIG).

Las variables estadísticas de interés pueden ser continuas o discretas. La única restricción que se considera es que exista algún tipo de dependencia entre dos variables con distinta referencia cartográfica. Dos características de este tipo de datos son que las observaciones son más similares cuanto más cercanas son sus localizaciones muestrales (independientemente del valor observado) y que a medida que la distancia de las localizaciones muestrales aumenta la correlación entre las variables tenderán a anularse.

Tampoco hay ninguna restricción en las referencias cartográficas, que pueden ser una referencia territorial explícita como latitud y longitud (datos geográficos) o una referencia implícita como domicilio o código postal (datos socio-económicos).

- Un dato espacio-temporal es una observación correspondiente a una variable asociada a una localización del espacio en un instante determinado.

1. Conceptos generales

- Habitualmente utilizaremos $x \in D$ para denotar una localización en un espacio euclídeo d -dimensional, $d = 1, 2, 3$
- $D \subset \mathbb{R}^d$ es un dominio sobre el cual consideramos n localizaciones muestrales

$$\{x_1, x_2, \dots, x_i, \dots, x_n\} \subset D$$

- Para un x_i específico denotaremos por $Z(x_i)$ a la variable aleatoria, unidimensional o multidimensional, de interés. Generalmente se verifica que, para dos localizaciones próximas x_i y x_j , las variables $Z(x_i)$ y $Z(x_j)$ son dependientes.
- En el contexto de datos espaciales se considera que $Z(x)$, $x \in D$ es un proceso aleatorio. Por esto se interpreta el conjunto de observaciones muestrales

$$\{z(x_1), \dots, z(x_n)\} \subset \mathbb{R}^n$$

como una única realización discreta del proceso $Z(x)$

También podemos referirnos a $Z(x)$ como función aleatoria, variable regionalizada o variable aleatoria georeferenciada.

2. Tipos de datos espaciales

Según como sea el conjunto de localizaciones estaremos en una situación diferente. Esto implicará métodos diferentes de trabajo. Así:

1. Localizaciones fijas (no aleatorias)

- Cuando D es un intervalo o unión de intervalos: Geoestadística (geostatistical data)

Ejemplo: $Z(x)$ es un proceso que representa la temperatura en superficie y D es la provincia de Pontevedra. Las localizaciones x_i son puntos en $D \subset \mathbb{R}^3$

- Cuando D es un conjunto numerable: Datos en áreas (areal data)

Ejemplo: $Z(x)$ representa la renta per-cápita de una provincia y D es el conjunto de las provincias españolas. Las localizaciones x_i pueden ser las coordenadas de las capitales de provincia, pero representan el área de la provincia.

2. Localizaciones aleatorias

- Procesos puntuales (point patterns): Un proceso puntual determina las localizaciones en las que ocurre un suceso de interés.

Ejemplo: $Z(x)$ representa la intensidad de un terremoto y D es el estado de California

Independientemente del tipo de datos, los objetivos principales del estudio de la Estadística Espacial son, como en casi todos los campos de la Estadística, dos:

1. Descripción de los datos.

Esto puede incluir no sólo el estudio descriptivo del proceso $Z(x)$, tal como se entiende dentro de la estadística clásica, sino también la modelización del tipo de dependencia espacial.

2. Predicción.

Sin duda en muchas ocasiones la predicción es el objetivo que motiva el estudio de datos espaciales

Ejercicio 1. El propósito de este primer ejercicio es comprobar vuestros conocimientos básicos de R. Desde R deberéis cargar la librería `geoR` y escoger alguno de los ficheros de datos de ejemplo que incluye esta librería:

```
>library(geoR)
>data(package='geoR')
```

Tenéis que realizar un estudio descriptivo del fichero elegido. Quiero que calculéis coeficientes de las variables (ningún estudio de la dependencia espacial, que se estudiará posteriormente). Se valorará positivamente la inclusión de alguna representación gráfica. Podéis ver algunos ejemplos en el fichero: `Examples.Diggle.pdf`

Si necesitáis recordar algunas funciones básicas de **R**, puede seros de utilidad el siguiente libro: `Introduction.to.R.pdf`.

Ambos ficheros se encuentran incluidos en el presente módulo de esta secuencia de aprendizaje, bajo la etiqueta de “Bibliografía adicional”

Antes del día 26 de febrero debéis hacerme llegar un script de **R** que yo pueda ejecutar. No es necesario entregar un informe en documento de texto.