20. Modelos para procesos puntuales

Una primera descripción de los procesos puntuales puede ser la estimación de la función de densidad espacial a partir de los datos observados. La densidad espacial tiene las mismas propiedades que la densidad unidimensional, pero su dominio es la región de interés donde el proceso puntual tienen lugar. Una alternativa es a la densidad es trabajar con la estimación de la intesidad $\lambda(x)$ del proceso puntual, que es proporcional a su densidad espacial. La intensidad y la densidad espacial son propiedades de primer orden del proceso puntual porque miden la distribución de los eventos en la región de estudio. Ninguna de ellas dan información sobre una posible interacción entre dos puntos. La interacción se cuantifica mediante las propiedades de segundo orden, que reflejan la tendencia de los eventos a aparecer agrupados, independiéntemente o regularmente espaciados.
Nos limitaremos a introducir los procesos de Poisson puesto que estos modelos son los que ofrecen las aproximaciones más sencillas a un mayor número de problemas. En ambos casos se asume que los eventos ocurren de manera independiente. En caso de detectarse patrones de agregación o de regularidad deben ajustarse otro tipo de modelos más complejos.

20.1. Procesos puntuales homogéneos de Poisson

Es la definición formal de un proceso puntual con aleatoriedad espacial completa. La función de intensidad λ es constante y se verifica que:

1. Si $A \subset D$, entonces el número de eventos en A es una variable aleatoria Poisson con parámetro $\lambda |A|$, con $|A|$ área o volumen de A.

2. Si $\{A_i, i \in I\}$ es una familia de subconjuntos de D disjuntos dos a dos, entonces el número de eventos en A_i constituyen variables aleatorias mutuamente independientes.

Para este tipo de procesos, un estimador insesgado de λ es $\hat{\lambda} = n/|D|$

20.2. Procesos puntuales no homogéneos de Poisson

En este caso la función de intensidad $\lambda(x), x \in D$ no es constante, sino que varía espacialmente. Se verifica que:

1. Si $A \subset D$, entonces el número de eventos en A es una variable aleatoria de Poisson con parámetro $\lambda(A) = \int_A \lambda(x)\, dx$.

2. Si $\{A_i, i \in I\}$ es una familia de subconjuntos de D disjuntos dos a dos, entonces el número de eventos en A_i constituyen variables aleatorias mutuamente independientes.

Es una generalización de los procesos puntuales homogéneos de Poisson.
Para este tipo de procesos $\lambda(x)$ puede estimarse

1. No paramétricamente, utilizando un estimador tipo núcleo. Si $D \subset \mathbb{R}^2$ la expresión del estimador no paramétrico sería
$$\hat{\lambda}(x) = \frac{1}{q(\|x\|)\, h^2} \sum_{i=1}^{n} K\left(\frac{\|x - x_i\|}{h}\right),$$
donde $\|u\| = \sqrt{u_1^2 + u_2^2}$ con $u = (u_1, u_2) \in D$, $K(u)$ es una función simétrica bidimensional, $q(\|x\|)$ es una función que corrige el efecto frontera que se produce cuando x está cerca de los límites de la región D y h es el parámetro de suavización.

   El problema de este tipo de estimación radica no en la elección de la función $K(u)$, sino en la selección adecuada del parámetro de suavización h.

2. Paramétricamente, maximizando la función de verosimilitud del proceso puntual.
$$L(\lambda) = \sum_{i=1}^{n} \log \lambda(x_i) - \lambda(D),$$
donde $\lambda(D) = \int_D \lambda(x)\, dx$ es el número esperado de casos del proceso puntual no homogéneo con intensidad $\lambda(x)$ en la región D.

Ejemplo: Puesto que los eventos correspondientes al estudio del pino japonés ocurren de manera independiente, este es el único fichero de datos que se ajusta a las hipótesis de procesos puntuales de Poisson. Lo utilizaremos para la estimación de $\lambda(x)$.

La función *ppm* de la librería *spatstat* ajusta paramétricamente la intensidad de un proceso puntual no homogéneo de Poisson obteniéndose los siguientes valores:

$$\lambda(x) = 4.069384 + 1.127699x - 1.561184y - 0.737021x^2 - 1.197023xy + 2.500548y^2$$

En la figura 17 se representan conjuntamente los eventos utilizados en la estimación así como, mediante gráfico de isolíneas, los valores estimados de $\lambda(x)$.

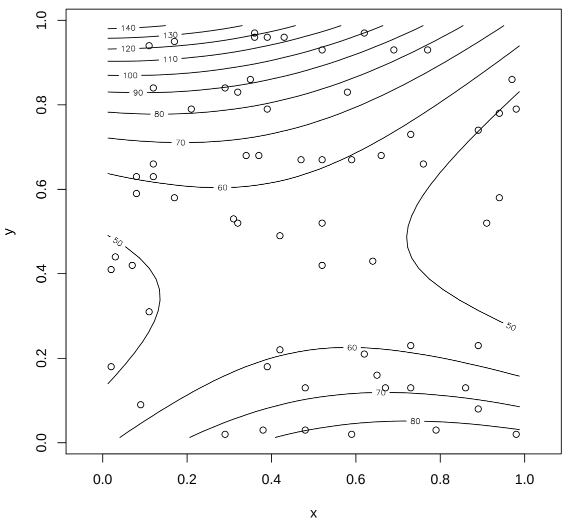

Figura 17: Localizaciones de pino japonés y estimación paramétrica de la intesidad del modelo.

Ejercicio 7. El último ejercicio de esta materia será nuevamente un cuestionario tipo test. Como en la anterior ocasión tendréis que contestar a 7 cuestiones que tratarán sobre datos en áreas y datos puntuales.

Observaciones para el ejercicio 7.

1. Para el cuestionario dispondréis de un tiempo limitado de 20 minutos.

2. Las preguntas serán seleccionadas aleatoriamente de una base de datos, por lo que las cuestiones no coincidirán en todos los cuestionarios.

3. Sólo podréis contestar al cuestionario una vez, sin posibilidad de repetir.

4. Obviamente podréis utilizar todo el material que consideréis necesario, pero conviene que lo tengáis todo revisado para que podáis contestar dentro de los 20 minutos previstos.

5. El cuestionario estará disponible desde el lunes 17 al viernes 21 de mayo.