

# Parte VII

## Datos puntuales

### 18. Introducción

El análisis de procesos puntuales, esto es de la disposición de un conjunto de eventos sobre una región, se enmarca en una de las tres grandes ramas de la estadística espacial. Básicamente, pretende determinar si dichos eventos presentan un patrón de agregación (los eventos se producen cerca de otros eventos), de inhibición (los eventos aparecen diseminados) o de aleatoriedad espacial completa (los eventos se producen con igual probabilidad en cualquier punto del espacio, con independencia de dónde se hallen otros eventos). Además es posible la comparación de los patrones de dos conjuntos de eventos. En general, un proceso puntual es un proceso estocástico en el cual observamos las localizaciones  $x \in D$  de algún suceso de interés representado por  $Z$ , dentro de una región acotada  $D \subset \mathbb{R}^d$ ,  $d = 1, 2, 3$ . El conjunto de localizaciones observadas (habitualmente llamados eventos)  $\{x_1, x_2, \dots, x_n\}$  recibe el nombre de proceso puntual.

Referencias teóricas importantes para datos puntuales:

- Cressie, N. (1993). Statistics for spatial data. Wiley, New York.
- Diggle, P.J. (2003). Statistical analysis of spatial point patterns. Oxford University Press, New York.

El análisis de procesos puntuales aparece en áreas de investigación muy diferentes. En ecología, por ejemplo, el interés puede estar en determinar la distribución espacial, y sus causas, de una especie de árboles para la cual se han registrado las localizaciones dentro de una determinada zona de estudio  $D$ . Más aún, si se han registrado datos de dos o más especies diferentes puede ser interesante comprobar si estas especies se distribuyen aleatoriamente o si existe algún tipo de competición por el terreno entre ellas. En epidemiología se han aplicado con éxito en algunas situaciones tales como el análisis de posibles fuentes de contaminación alrededor de las cuales surgen casos

de alguna enfermedad. Su campo de aplicación es más amplio, y puede incluso servir también para cartografiar enfermedades.

## 19. Análisis preliminar

Cuando se estudia un proceso puntual, lo primero que suele comprobarse es la denominada “aleatoriedad espacial completa”. Intuitivamente se entiende que si los eventos presentan agregación podremos sospechar que existe atracción o contagio mientras que si el proceso presenta un comportamiento muy regular puede estar indicando inhibición o competición entre los elementos de estudio. Patrones intermedios indicarían un comportamiento aleatorio. Estos comportamientos suelen cuantificarse utilizando técnicas no paramétricas.

En la figura 14 se representan tres procesos puntuales correspondientes a tres simulaciones. El cuadro de la izquierda muestra la disposición aleatoria de los eventos, el cuadro del medio muestra las localizaciones muy regulares y el de la derecha localizaciones agregadas.

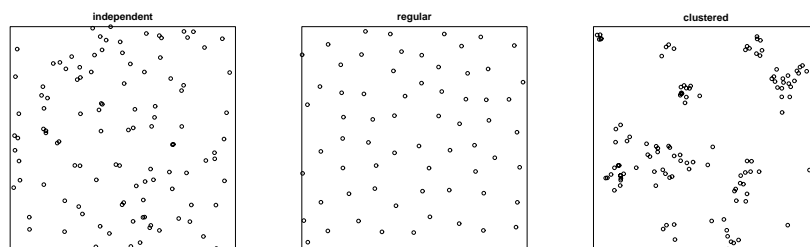


Figura 14: Ejemplo de tres procesos puntuales con diferentes grados de aleatoriedad espacial

### 19.1. Test para la aleatoriedad espacial completa

Bajo la hipótesis de aleatoriedad espacial completa los eventos se distribuyen según un proceso de Poisson de intensidad  $\lambda$ , donde  $\lambda$  es el número medio de eventos por unidad de superficie (área o volumen, según la dimensión  $d$ ). Bajo esta hipótesis la distribución

de la distancia entre un evento cualquiera y su evento más próximo se modeliza por

$$G(r) = 1 - \exp\{-\lambda\pi r^2\}, r \in \mathbb{R},$$

si  $D \subset \mathbb{R}^2$  o por

$$G(r) = 1 - \exp\left\{-\lambda\frac{4}{3}\pi r^3\right\}, r \in \mathbb{R},$$

si  $D \subset \mathbb{R}^3$ .

La distribución muestral  $\hat{G}$  de la distancia entre eventos más próximos se calcula utilizando los puntos observados mediante

$$\hat{G}(r) = \frac{\#\{\text{Pares de eventos que distan } \leq r\}}{n},$$

donde el numerador es el número de pares de puntos cuya distancia euclídea es menor o igual que  $r$  y  $n$  es el tamaño muestral. La compatibilidad de la aleatoriedad espacial con el proceso puntual se comprueba comparando gráficamente ambas distribuciones. Así, se representa gráficamente la función  $G(r)$  junto con su banda de confianza. Si la función empírica  $\hat{G}(r)$  se mantiene dentro de la banda de confianza de la función teórica  $G(r)$ , entonces se acepta la hipótesis nula de que el proceso puntual sigue un modelo de aleatoriedad espacial completa.

Ejemplo: En la figura 15 se representan tres procesos puntuales observados en dominios de dimensión  $d = 2$  diferentes. Los dominios han sido reescalados al cuadrado unidad para poder realizar una comparación gráfica. El proceso de la izquierda representa la distribución espacial de células estudiado por Ripley (1977), el proceso del centro muestra la distribución de un tipo de pino japonés estudiado por Numata (1961). Por último, el gráfico de la derecha muestra la distribución de árboles del tipo secuoyas rojas en California estudiado por Strauss (1975). Estos tres ficheros de datos se encuentran en la librería *spatstat* de *R* con los nombres de *cells*, *japanesepines* y *redwood* respectivamente. Utilizando la función *envelope* de la librería *spatstat* de *R* se pueden calcular las bandas de confianza del 96 % de la función teórica  $G(r)$ , así como los valores de la función empírica  $\hat{G}(r)$ .

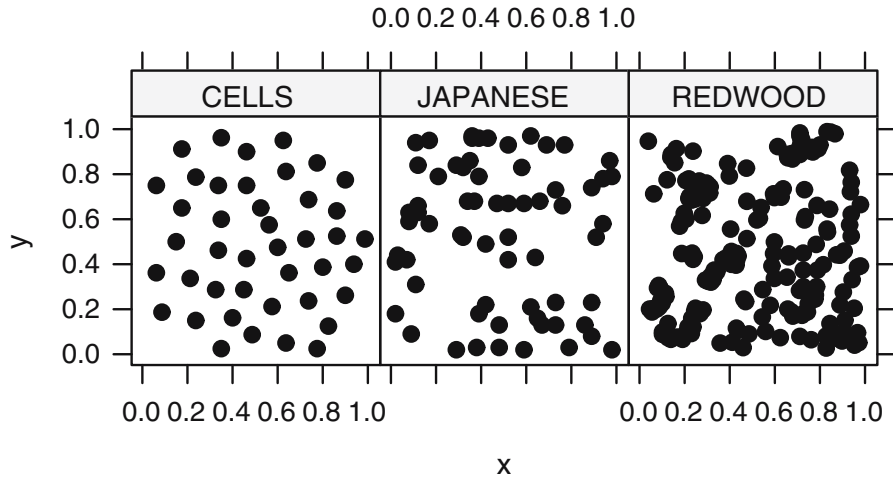


Figura 15: Ejemplo de tres procesos puntuales reescalados en el cuadrado de área uno.

En la figura 16 se han representado los valores obtenidos mediante la función *envelope* para los procesos puntuales descritos anteriormente. En el gráfico de la izquierda se aprecia que los valores de  $\hat{G}(r)$  están por debajo de la banda de confianza, lo cual se interpreta como un comportamiento regular de los eventos. El gráfico de la derecha, con valores de  $\hat{G}(r)$  por encima de la banda de confianza indica un patrón de agregación. Sólo el ejemplo relativo a los datos de pino japonés parece estar aleatoriamente distribuido.

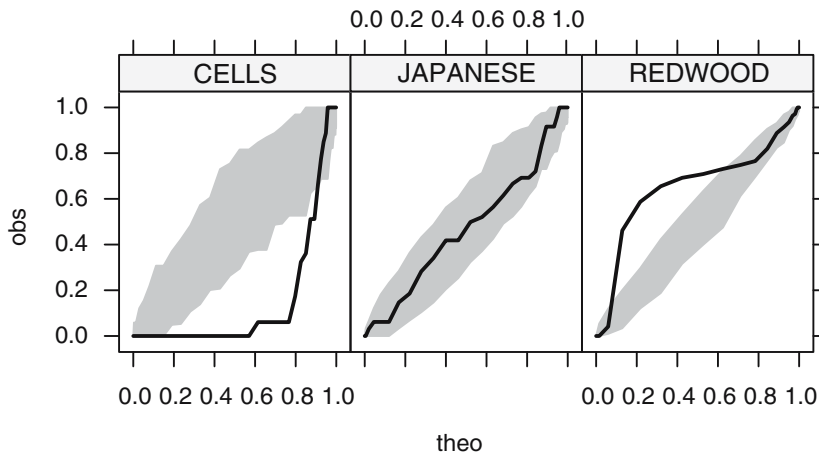


Figura 16: Bandas de confianza de  $G$  y valores de la función empírica  $\hat{G}$ .