



Estimación non paramétrica da densidade no plano con R

María Bugallo Porto

Facultade de Matemáticas, Universidade de Santiago de
Compostela

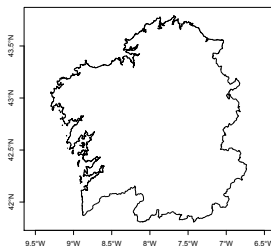
15 de outubro de 2020

Distribución das velutinas en Galicia

- Posición de niños de avespa velutina en Galicia en 2016, 2017 e 2018 [coordenadas cartesianas].



- Como se distribúe a concentración no mapa cada ano?
- Como é a evolución temporal da mesma?



Descripción dos datos

Temos 7387 observacións agrupadas en tres anos consecutivos

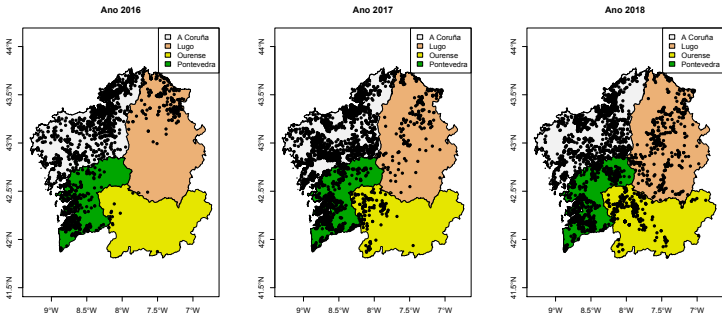
· 2784
do ano
2016

· 2968
do ano
2017

· 2635
do ano
2018

que conforman unha mostra aleatoria bidimensional (latitude e longitude) da posición dos niños no territorio galego.

Diagrama de dispersión



$\{sp\}$: lectura das coordenadas do mapa galego.

plot (mapa galego), points (niños das avespas da mostra),

legend (lenda).

Histograma bidimensional

Función estimada mediante o histograma en cada punto:

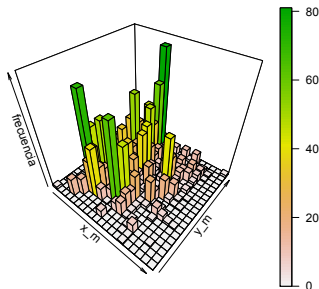
$$\hat{f}_{hist}((x, y), (b_x, b_y)) = \frac{\sum_{i=1}^n \mathbb{I}_{((X_i, Y_i) \in B_j)}}{nb_x b_y}, \forall (x, y) \in B_j, j \in \mathbb{Z}.$$

Os termos a determinar son:

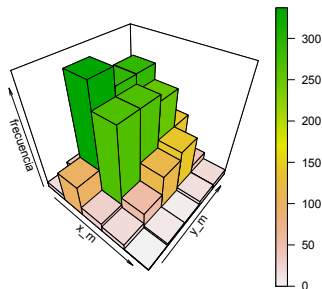
- punto de anclaxe (t_{0x}, t_{0y}) .
- lonxitude dos lados dos rectángulos b_x e b_y .
- orientación dos rectángulos que, por sinxeleza, serán paralelos aos eixos coordenados.

Histograma bidimensional

$bx=0.1194$ e $by=0.09604$



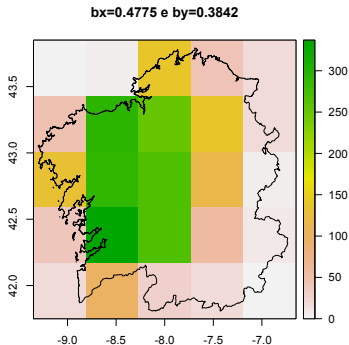
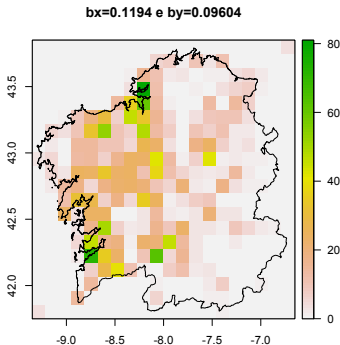
$bx=0.4775$ e $by=0.3842$



`{plot3D}`

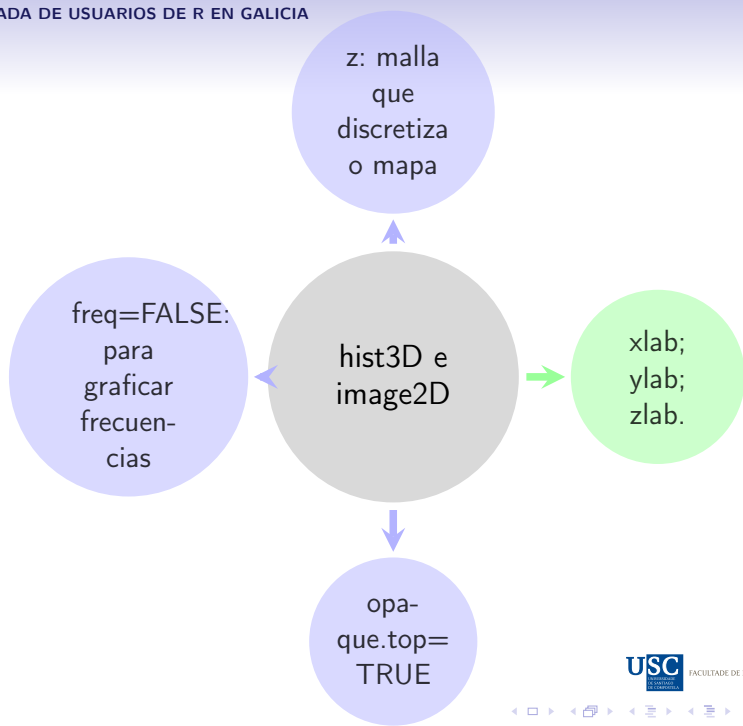
hist3D: discretizamos o espazo mostral, na figura da dereita catro veces menos que na da esquerda.

Mapas de calor

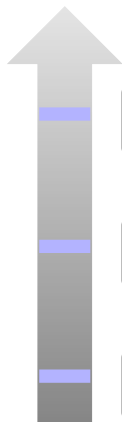


`{plot3D}`

image2D: proporciona a vista aérea dos histogramas anteriores.



Histograma bidimensional



Ventaxa: estimador moi intuitivo
cunha formulación moi sinxela.

Desventaxa: falta de suavidade na estimación.

Desventaxa: dependencia do punto de anclaxe.

Estimador tipo núcleo bidimensional

Función de densidade estimada en cada punto:

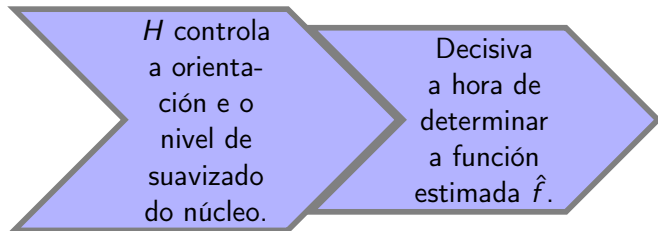
$$\hat{f}((x, y), H) = \frac{1}{n} \sum_{i=1}^n K_H(x - X_i, y - Y_i), (x, y) \in \mathbb{R}^2, \text{ onde:}$$

- K é a función núcleo.
- $H \in M_{2 \times 2}$ é a matriz de parámetros de suavizado.
- $K_H(x, y) = |H|^{-1/2} K(H^{-1/2}(x, y)^T)$.

Se $K = \phi_{\vec{0}, l_2}$ e $H = \text{diag}(h_1, h_2)$,

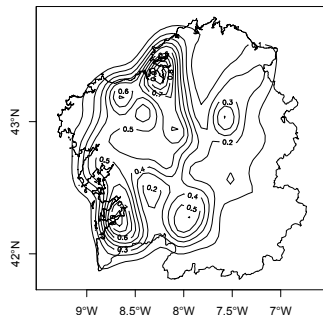
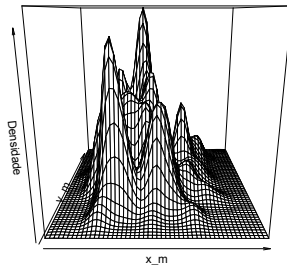
$$\hat{f}((x, y), H) = \frac{1}{n\sqrt{h_1}\sqrt{h_2}} \sum_{i=1}^n \phi\left(\frac{x - X_i}{\sqrt{h_1}}\right) \phi\left(\frac{y - Y_i}{\sqrt{h_2}}\right).$$

Selección da matriz de parámetros ventá



- Empregaremos a **Regra do Polgar de Scott** para estimar H .
- Básase en distribucións paramétricas.
- Logo de probar con outros criterios como os de **validación cruzada nesgada e inesgada**, é o que mellor se axusta ao nosos datos.

Estimador tipo núcleo bidimensional



`{KernSmooth}`

bkde2D: estimación tipo núcleo da densidade bidimensional.

Empregouse a **Regra do Polgar de Scoot** para elixir H .

bkde2D:

Bandwidth escolle $h_{12} = 0$ e o núcleo é a densidade normal estándar bivalente por defecto.

persp

Para obter a gráfica de perspectiva da estimación densidade.

contour

Para obter os contornos de nivel asociados á estimación da densidade.

Comparativa das dúas estimacións

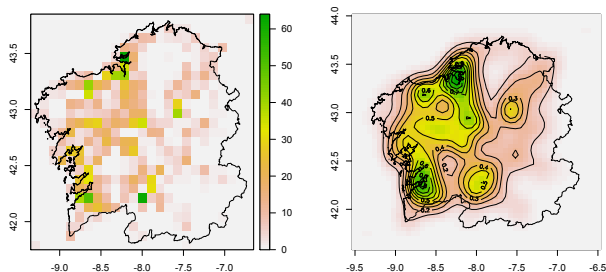


Figura: Dúas estimacións distintas da función de densidade dos niños de avespa velutina en Galicia no ano 2018. A esquerda, gráfica de calor asociada a un histograma bidimensional; a dereita, curvas de nivel da estimación tipo núcleo, xunto co mapa de calor e o contorno de Galicia en negro, igual que as curvas de nivel.

Evolución temporal

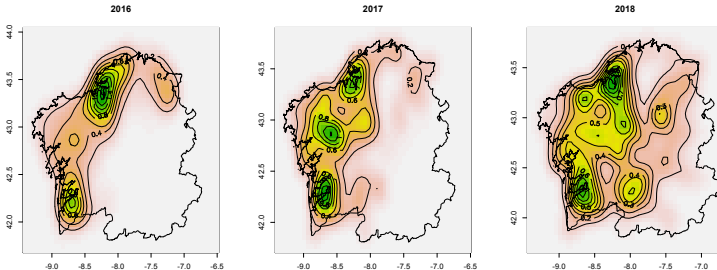


Figura: Evolución temporal da distribución da densidade dos niños de velutina empregando curvas de nivel e mapas de calor.

Evolución temporal dos lugares altamente densos

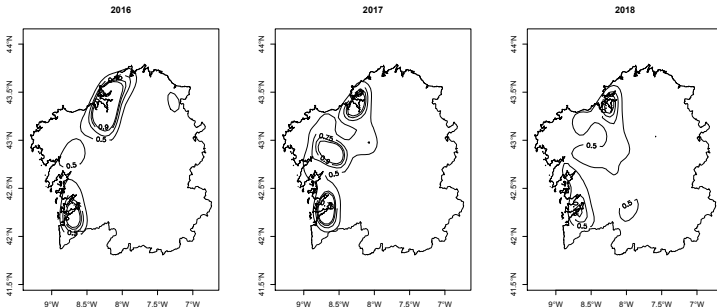


Figura: Curvas de nivel dos lugares con concentración de niños de velutina superior ao 50 % nos anos 2016, 2017 e 2018.

Engadimos a opción `levels = c(0,5, 0,75, 0,90, 0,99)` do comando `contour`.

Paquetes de R

→ *sp*: proporciona métodos para posicionar datos nun mapa, para empregar coordenadas en dúas e tres dimensións...

→ *plot3D*: contén funcións para debuxar datos bidimensionais e tridimensionais.

→ *KernSmooth*: contén funcións para realizar a estimación da densidade tipo núcleo unidimensional e bidimensional.



Estimación non paramétrica da densidade no plano con R

María Bugallo Porto

Facultade de Matemáticas, Universidade de Santiago de
Compostela

15 de outubro de 2020