BST 6200 Spatial Statistics and Disease Mapping

Steven E. Rigdon

Spring 2020

Section 6.7: Interpolation of Point Patterns with Continuous Attributes

A **point process** is a set of point locations $(x_1, y_1),\ (x_2, y_2),\ \ldots,\ (x_n, y_n)$ in some domain.

A **marked point process** is a set of point locations $(x_1, y_1),\ (x_2, y_2),\ \ldots,\ (x_n, y_n)$ in some domain together with some "marked" variable z_i.

In this section we consider marked point processes where the "mark" is a continuous measurement such as a measure of air pollution. Data is of the form $[(\boldsymbol{x}_1, z_1),\ (\boldsymbol{x}_2, z_2),\ \ldots,\ (\boldsymbol{x}_n, z_n)]$.

Longitude	Latitude	Measurement
$\boldsymbol{x}_{11}$	$\boldsymbol{x}_{12}$	z_1
$\boldsymbol{x}_{21}$	$\boldsymbol{x}_{22}$	z_2
$\boldsymbol{x}_{31}$	$\boldsymbol{x}_{32}$	z_3

The Problem and Three Solutions

Objective: estimate the value of z at a point $\boldsymbol{x}_{\text{new}}$ that is not one of the original points

Three solutions:

1. Nearest neighbor interpolation
2. Inverse distance weighting
3. Kriging

Example to be used for illustration

We have eight monitoring sites for PM 2.5.

x	y	z
0.10	0.35	4.5
0.22	0.23	5.5
0.84	0.63	10.8
0.61	0.90	7.4
0.88	0.21	7.5
0.55	0.81	9.3
0.04	0.78	4.1
0.47	0.59	6.2

Solution 1: Nearest Neighbor Interpolation

Let $\boldsymbol{x}_{\text{new}}$ denote a new point that is not one of the observation points.

Objective: Estimate the outcome at $\boldsymbol{x}_{\text{new}}$.

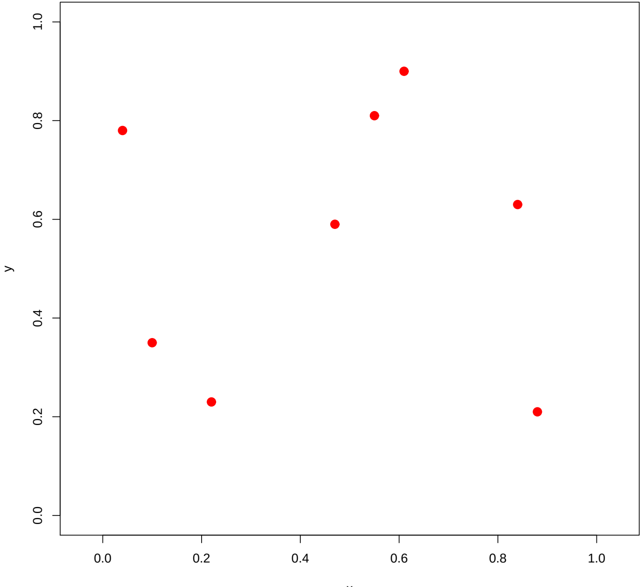
Solution: Find the nearest observation point to $\boldsymbol{x}_{\text{new}}$ and use the PM2.5 at that point as an estimate of the PM2.5 at $\boldsymbol{x}_{\text{new}}$.

More precisely:

Step 1: Find i that minimizes $\|\boldsymbol{x}_i - \boldsymbol{x}_{\text{new}}\|$. Call it i_{min}.

Step 2: Estimate $\hat{z}_{\text{new}} = z_{i_{\text{min}}}$.

Voronoi Diagram and Voronoi Polygons



Work through the R code to do plot the nearest neighbor estimator.

The Fulmar Data Set

Fulmaris glacialis is a sea bird that live in the North Sea between England and the Netherlands.

Counts of the bird were taken in 1998 and 1999 using arial photos. The recorded value for z is the density of birds observed: number/area.

The path taken by the plane is called the a transect and is visible in plots of the observation points.

Students should work through the R code on the fulmar data described in the textbook, section 6.7. Note however, that the book's website gives the code (incorrectly) under Section 6.8.

Solution 2: Inverse Distance Weighting

Idea: Use a weighted average of outcomes at all other points $\boldsymbol{x}_{\text{new}}$ and make the weights be inversely related to distance. Thus, close points get high weight and far away points get less weight.

Step 1: For a given $\boldsymbol{x}_{\text{new}}$ point, assign the weights

$$w_i = \left\| \boldsymbol{x} - \boldsymbol{x}_{\text{new}} \right\|^{-\alpha}$$

for all other observation points.

Step 2: Estimate the outcome at $\boldsymbol{x}_{\text{new}}$ to be

$$\hat{z}(\boldsymbol{x}_{\text{new}}) = \frac{\sum_{i=1}^{n} w_i z_i}{\sum_{i=1}^{n} w_i}$$

The Effect of α

If $\alpha = 0$ then all weights are equal to 1, and the estimate of the outcome is the average of all z values.

If $\alpha = 1$ then the weights are inversely proportional to the distance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_{\text{new}}$.

If $\alpha = 2$ then the weights are inversely proportional to the squared distance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_{\text{new}}$. This gives nearer points higher weight and farther away points lesser weight than using $\alpha = 1$.

If α is very large, then the weight of the nearest neighbor dominates all of the other weights and IDW behaves like the nearest neighbor method.

If $\boldsymbol{x}_{\text{new}}$ is actually one of the observation points, say the ith one, then $\hat{z} = z_i$

Apply IDW to the Fulmar Data Set in R