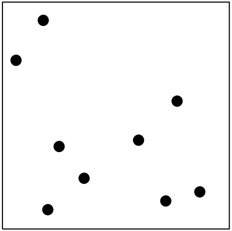
Spatial analysis

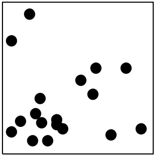
Josemari Sarasola

Gizapedia

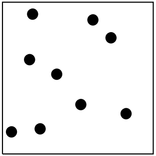
Let's take a space distribution of points (cancer cases, flats on sale, ...):



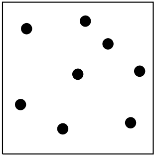
Spatial pattern types



Clustered

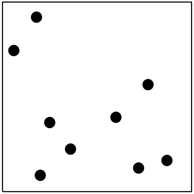


Random



Regular (overdispersion)

How to decide *statistically* if a given point distribution is clustered, random or regular?



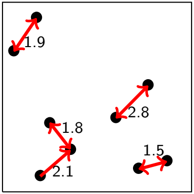
?

Spatial patterns and Poisson processes

Random point disributions are a special kind of Poisson processes. Usually Possion processes apply to time (one dimension) but we can also extrapolate them to more dimensions (2 dim., for spatial patterns). In those cases, lambda parameter expresses the mean number of points per space unit.

Random? Nearest neighbor method

Calculate distance for each point to the nearest neighbor (double arrow shows bidirectional nearest neighbor) :



We calculate the mean:

$$\bar{d} = \frac{2.1 + 1.8 + 1.8 + 1.9 + 1.9 + 2.8 + 2.8 + 1.5 + 1.5}{9} = 2.01$$

Random? Nearest neighbor method

In clustered patterns, average distance is small, intermediate for random patterns and big for overdispersed patterns. But when is the average small, intermediate or big?
In random patterns, the average distance is a function of λ point density. We calculate λ density in this way:

$$\lambda = \frac{n}{A}$$

From this value, the expected value for the average distance in random patterns is this one:

$$E[d] = \frac{1}{2\sqrt{\lambda}}$$

Random? Nearest neighbor method

In the example, the area is $10 \times 10 = 100$ unit. We have 9 points. So density is: 9/100=0.09.
And the expected value for distance is:

$$E[d] = \frac{1}{2\sqrt{0.09}} = 1.66$$

So, we could decide in this way:

- $\overline{d} \approx E(d) \rightarrow$ random
- $\overline{d} < E(d) \rightarrow$ clustered)
- $\overline{d} > E(d) \rightarrow$ regular (overdispersed)

Random? Nearest neighbor method

For example, if the average distance is 2.01. We should conclude that the pattern is somehow clustered. But, *statistically* we should decide if 2.01 is significantly far from 1.66, because in fact a 2.01 value may be compatible with randomness. (It's like when after throwing a dice 60 times, you cannot conclude that the dice is unfair because number 1 didn't appear exactly 10 times. So, we should'nt take 1.66 as a fixed value, but as an approximate one. So we must decide with a statistical test or a confidence interval.

Random? Nearest neighbor method

For a 90% confidence level, this is the interval of acceptance for randomess, around the average distance given by lambda, having n points:

$$\frac{1}{2\sqrt{\lambda}} \pm 1.64\sqrt{\frac{4-\pi}{4\pi n\lambda}}$$

As the numer of points grows, the interval shortens (it's logical: with more points, the certainty for the average distance is bigger).
This in the interval for our data:

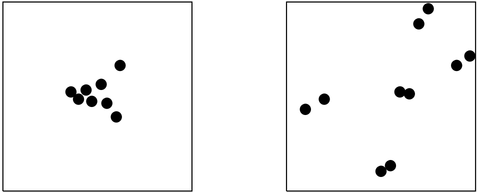
$$\frac{1}{2\sqrt{0.09}} \pm 1.64\sqrt{\frac{4-\pi}{4\pi \times 20 \times 0.09}} : (1.66 \pm 0.32) : (1.34, 1.98)$$

2.01 is on the upper side of the interval of randonmess, so the spatial pattern in overdispersed or regular with 90% of confidence.

Random? Nearest neighbor method

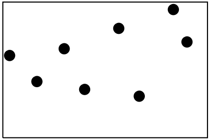
Disadvantages

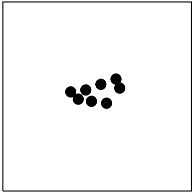
(1) Both patterns are different but the give a similar mean distance:



Random? Nearest neighbor method

(2) Perspective is relevant. In this pattern, from a close POV, is regular, but looking into it from far away is clustered:





Random? K function method

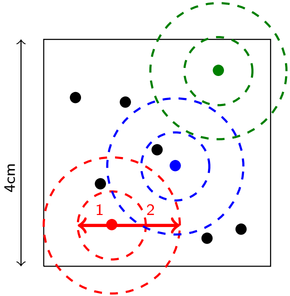
Better than nearest neighbor: overcomes the first disadvantage

1. From each point we draw a circle of radius h and count the no. of points inside: p_i . Sum of p_i is P.
2. Caculate K function, for different h values:

$$K(h) = \frac{P}{n\lambda}$$

3. We compare $K(h)$ values with a values corresponding to randomness:

Random? K function method



$$n = 9$$

$$azalera = 4cm \times 4cm = 16$$

$$\lambda = 9/16 = 0.56$$

Point	K(1cm)	K(2cm)
p_1	0	1
p_2	1	1
p_3	0	0
P	1	2
$K(h)$	$\frac{1}{3 \times 0.56} = 0.59$	$\frac{2}{3 \times 0.56} = 1.18$

We must count for all points. We have done for only 3 points as an example.

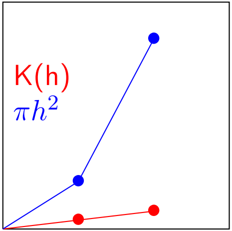
Interpretation

Under randomess the expected value for K(h) is πh^2. So:

- $K(h) \sim \pi h^2 \rightarrow$ random
- $K(h) > \pi h^2 \rightarrow$ cluster
- $K(h) < \pi h^2 \rightarrow$ regular (overdispersion)

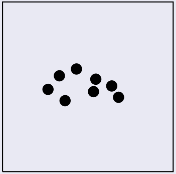
h	$K(h)$	πh^2
1	0.59	3.14
2	1.18	12.56

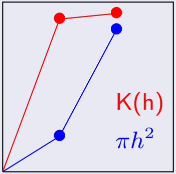
In our example:
$K(h) < \pi h^2 \rightarrow regular$.



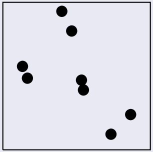
Random? K function method

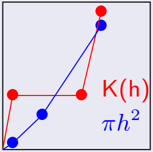
Usual patterns: one-clustered





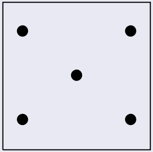
Usual patterns: several clusters

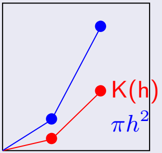




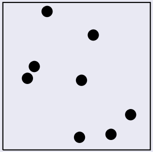
Random? K function method

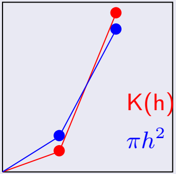
Usual patterns: overdispersion





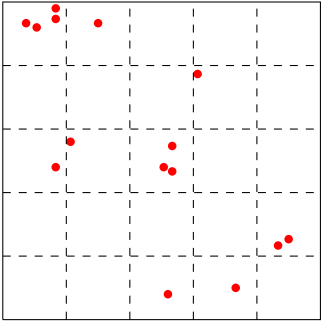
Usual patterns: random





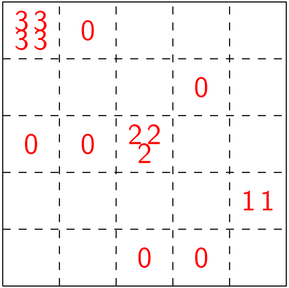
Random? Method of squares

Divide the area into equal-sized squares and count the no. of points into each one:



4	1	0	0	0
0	0	0	1	0
1	1	3	0	0
0	0	0	0	2
0	0	1	1	0

For each point in each square, we count how many "friends" has each point:



Crowding mean is the mean of the no. of "friends"

$$m^* = \frac{3+3+3+3+0+0+0+0+2+2+2+1+1+0+0}{15 \text{ points}} = 1.33$$

Now, we calculate the mean number of points for each square:

$$m = \frac{4+1+0+\cdots+1+0}{25 \text{ squares}} = 0.66$$

Random? Method fo squares: Lloyd's patchiness index

Patchiness index

$$\frac{m^*}{m}$$

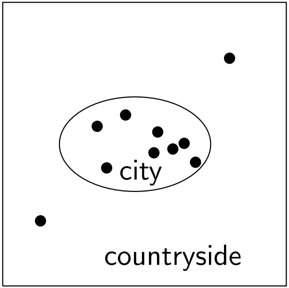
The bigger the index is, the clustering becomes more and more evident.

- $\dfrac{m*}{m} \approx 1 \rightarrow$ random (Poisson)
- $\dfrac{m*}{m} > 1 \rightarrow$ clustering
- $\dfrac{m*}{m} < 1 \rightarrow$ overdispersion

In our example, $\dfrac{m*}{m} = \dfrac{1.33}{0.66} = 1.33 \rightarrow cluster$

Randomness in a previously clustered pattern

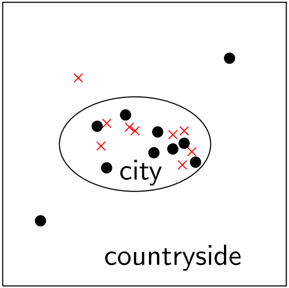
This distribution of points shows the distribution of flats on sale in a geographical area.



The question is: Are the flats clustered? In a strict sense yes (and the methods we have learnt would reflect that), but we have to take into account that all flats are previously clustered into the city. Actually, into the city flats seem to distribute randomly.

Randomness in a previously clustered pattern

A solution is taking a sample from the population of all flats, on sale or not (red cross), analyze its clustering and comparing it to the clustering of flats on sale:



If clustering of flats on sales bigger than that of the sample of all flats, we really have clustering for flats on sale; if not, flats on sale would be (more or less) overdispersed related to all the population of flats.