SPATIAL POINT PROCESSES

The goal of spatial point processes is to interpret data that is assigned to a location. There are many different methods with different goals. Sometimes you want to characterize clusters, dispersion or randomness. Sometimes you want to test if a pattern is random or not. You can also model or interpolation to get predictions.

Definitions:

Stationary: the process is invariant under spatial translation, there is some constant mean & variance

Isotropic: invariant under rotation

Intensity: a scale of how many points you expect to find in a given volume.

Complete Spatial Randomness: a stationary Poisson point process where λ is constant accross the space

Uniformity Test

to see if a pattern is present you can use the e.d.f of the nearest neighbor distribution

for CSR $G_{CSR}(r) = 1 - e^{-\pi n r^2 / A}$ → A is area

you can compare the data to CSR to see if there are trends.

Moran I's

this is one test to see if there is spatial autocorrelation or just randomness.

1	1	0	0
1	1	0	0

Positive correlation

1	0	1	0
0	1	0	1

Negative correlation

$$w_{ij} = \begin{cases} 1 & \text{if } i,j \text{ Border} \\ 0 & \text{if not} \end{cases}$$

→ there are other options & you can consider further neighbors

Weight matrix

$$W = \sum w_{ij} = 20$$

1. Compute ave. value in the spatial locations $\bar{x}$
2. Find variation $\sum (x_i - \bar{x})$
3. Find variation weighted by w_{ij}
   $\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})$ ⇒ are the neighbors similar?

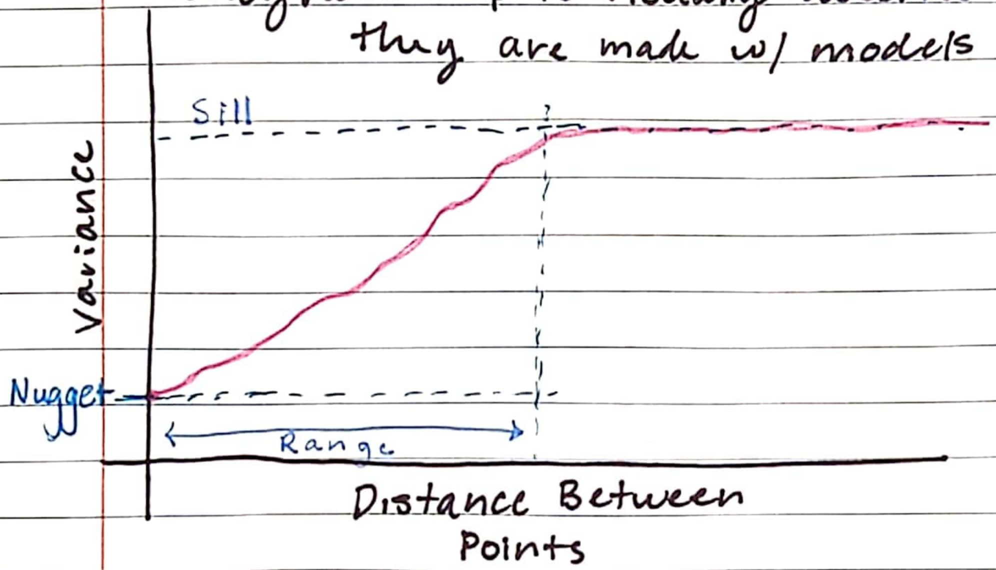
CamScanner

4. $I = \frac{N}{W} \frac{(\text{Step } 2)}{(\text{Step } 3)}$ → positive value = positive correlati
negative value = negative correla

For another test, try Geary's C
Or for local (not global) Local Moran's I can be used

Variogram

variograms help to visually describe the correlation
they are made w/ models that use least-squares or MLE



High Variance - the point have much different values

Low Variance - the points have similar values

Points that are further away should be less similar

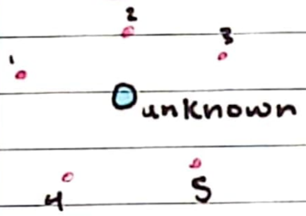
Range: where the semi-variance/variance stops increasing, after the range, points are not correlated

Sill: the level of variance when correlation is not present

Nugget: at a distance of 0 the variance, it occurs because of measurement errors, variations at scales smaller than the distances

Kriging

→ you have some sort of geographic area where you took some measurement, but now you want to predict what happens inbetween.



⇒ the prediction of the unknown point is a linear combo. of the closest neighbors

$$y_{unknown} = w^T y + \varepsilon = w_1 y_1 + w_2 y_2 + \dots w_i y_i + \varepsilon$$

↓ w's are the weights, some points are more important

$\gamma(x_i, x_j) = \frac{1}{2}(y_i - y_j)^2$ → variance between 2 points

$Aw = b$ → solve for the weights matrix

A ↓ $\gamma(x_i, x_j)$ variogram function

b → $\gamma(x_{unk}, x_i)$ variogram function

$w = A^{-1} b$

→ Use the y_{unk} formula w/ the predicted weights

To use Kriging:

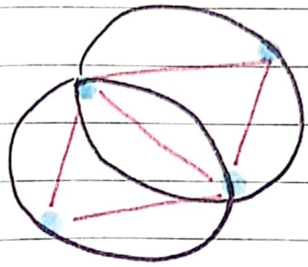
1. The data must be stationary
2. The variogram must be constant, different areas have same variogram

But, you can also transform data too.

Delaunay Triangulation

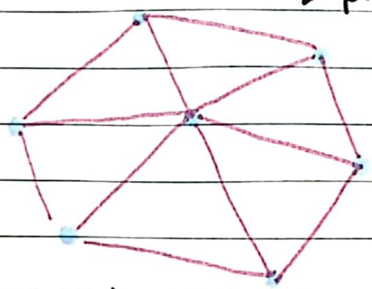
is a set of triangles that connects the data points so that no vertex is in the circumcircle of any triangle in the set

This is used for making triangle meshes

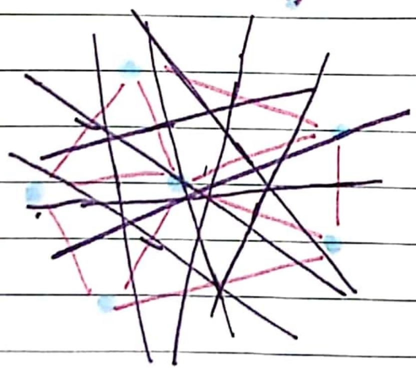


Voroni Tesselations

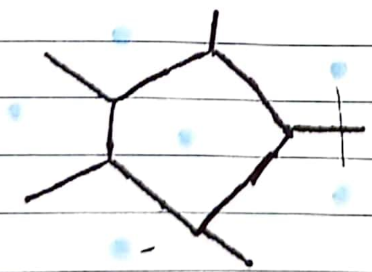
↳ provide a visual tool for analyzing spatial distrib.



line segment connects all the points



Perpendicular bisectors go through each line



Then you remove the unneccesary stuff and are left w/ something like

The lines are equidistant to 2 points

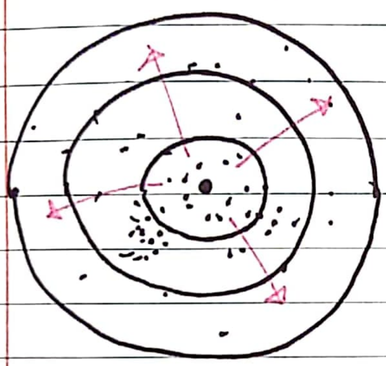
Ripley's K Function

is used as a global measure of clustering

$$K(d) = \frac{1}{\lambda n} \sum_{i=1}^{n} \#[S \text{ in } C(s_i, d)]$$

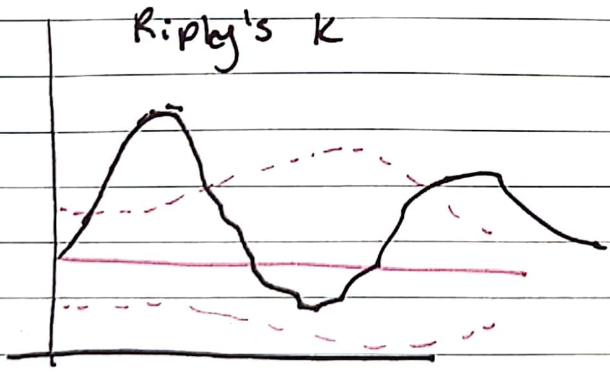
↳ circle of radius d

↳ average # of points w/in distance d of observed points divided by the process intensity λ.



you move outward & count how many points are at each dist.

positive = more points than expected
negative = less than expected



if the function goes outside of the bounds, then you know its not random

Besag's L function:

Is similar to the K function, but stabilizes it more

$$L^*(d) = \sqrt{\frac{K(d)}{\pi}} - d$$

Geographically Weight Regression

This is a type of model that fits spatial data by regression. Used if an effect is not constant accross space. There is a map of coef. values.

$$\underbrace{Y(x)}_{\text{response variable}} = \underbrace{\mu(x)}_{\text{the mean value at a location}} + \overbrace{\beta(x)}^{\text{coefficient}} \underbrace{X(x)}_{\text{covariates}} + \varepsilon$$

this is solved by a weights matrix & a least square algorithm.