Spatial Analysis

Types of Spatial Data

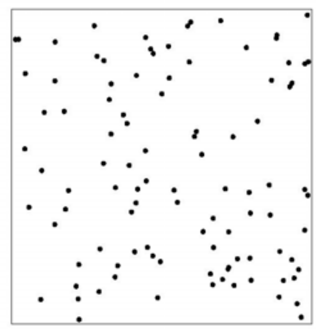
- Continuous Random Field
- Lattice Data
- Point Pattern Data

Note: Each type of data is analyzed differently

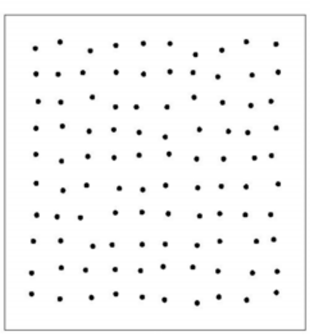
Point Pattern Data

- ✓ data are the locations of the objects of interest
- ✓ record the location and covariate information (often referred to as a marked point process)
- ✓ Ideally, all locations of the point pattern are recorded for the study region
- ✓ Some methods do not require all locations to be known

Random Spatial Pattern



Uniform Spatial Pattern



Clustered Spatial Pattern

Complete Spatial Randomness (CSR)

CSR is based on a homogeneous Poisson process, which has two conditions:

1) The number of events in any study region A with area |A| has a Poisson distribution with mean $\lambda|A|$
   - ✓ the intensity is constant throughout the study region
   - ✓ such processes are said to be stationary

2) The *N* events in a region A are an independent random sample from a uniform distribution on A
   - ✓ There is no interaction among events
   - ✓ The csr process is isotropic → no directional effects

Consequences of CSR

- If a spatial point pattern exhibiting csr is sampled with randomly placed quadrats, the expected distribution of the number of plants per quadrat is Poisson with parameter $\lambda|A_i|$, where λ is the mean number, or intensity, of plants in a unit of the study area and $|A_i|$ is the area of the ith quadrat.
- The distance between a randomly selected plant and its nearest neighbor is exponentially distribution with mean $1/\lambda$.

Poisson Distribution

One parameter: λ

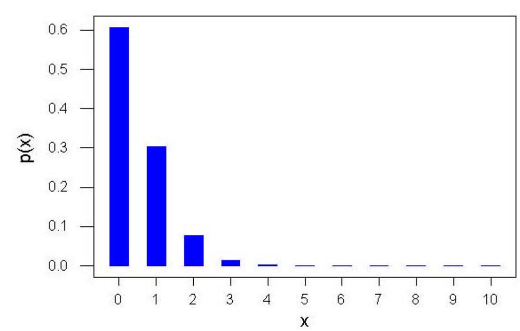
Mean: λ

Variance: λ

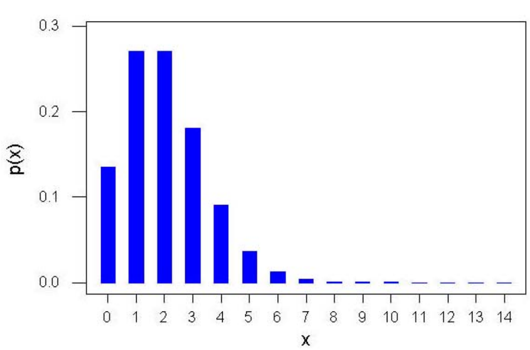
Probability Mass function:

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}, \qquad x = 0,\ 1,\ 2, \ldots$$

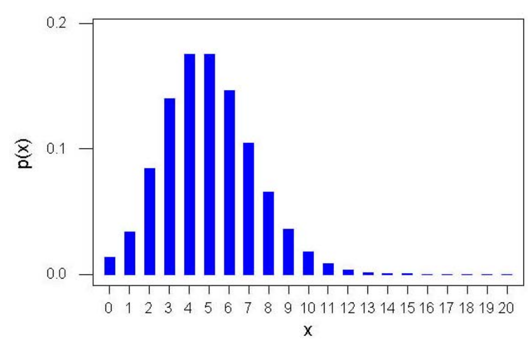
Poisson Distribution: λ = 0.5



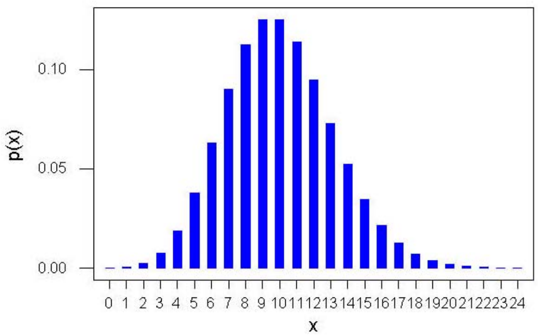
Poisson Distribution: λ = 2



Poisson Distribution: λ = 5



Poisson Distribution: $\lambda = 10$



Exponential Distribution

One parameter: λ

Mean: $1/\lambda$

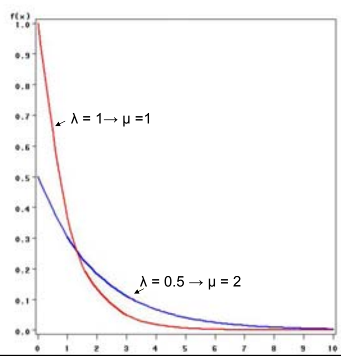
Variance: $1/\lambda^2$

Probability Density Function:

$$f(x) = \lambda e^{-\lambda x}, \qquad x > 0$$

The Exponential Distribution has the **memoryless property**

Exponential Distribution



Approaches

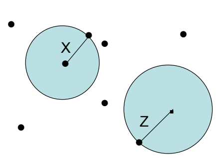
1) Map all events in the area
   - ✓ Most powerful
   - ✓ Sometimes (often?) infeasible
2) Use nearest neighbor approaches
   - ✓ Not as powerful
   - ✓ Easier to implement

We will begin with the nearest neighbor methods

Hopkins's Test

Measure the distance X from a randomly selected event to its nearest neighbor

Measure the distance Z from a randomly selected point to its nearest neighbor



Area searched around event: πX^2

Area searched around point: πZ^2

Twp areas are independent

Hopkins's Test

Measure the distance X from a randomly selected event to its nearest neighbor

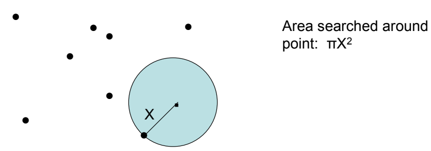
Measure the distance Z from a randomly selected event to its nearest neighbor

Test Statistic: $A = \dfrac{\sum_{i=1}^{n} X_i^2}{\sum_{i=1}^{n} Z_i^2}$

Under csr, *A* has an *F* distribution with 2*n* and 2*n* degrees of freedom. We reject the null hypothesis of csr if *A* becomes too small or too large.

Pielou's Index and Test

Measure the distance X_i, I = 1, 2, …, *n*, from *n* randomly selected points to their nearest neighbor



Pielou's Index

The average area of the n circles is $\dfrac{\pi\sum_{i=1}^{n} X_i^2}{n}$

Pielou proposed an **index** for measuring departures from csr:

$$C = \frac{\pi\lambda}{n}\sum_{i=1}^{n} X_i^2$$

Under the hypothesis of csr, the distance from a randomly selected plant to the nearest event is exponentially distributed with mean 1/λ. Thus C has an expected value of 1, and a variance of 1/n.

- ✓ C < 1 → events are more regular
- ✓ C > 1 → events are more clustered

Pielou's Test

The average area of the n circles is $\dfrac{\pi\sum_{i=1}^{n} X_i^2}{n}$

Pielou proposed a **test statistic** for testing the null hypothesis of csr:

$$D = 2\pi\lambda\sum_{i=1}^{n} X_i^2$$

Under the hypothesis of csr, *D* has a chi-squared distribution with 2*n* degrees of freedom.

Problem: λ is unknown

Pielou's Test

Pielou proposed a **test statistic** for testing the null hypothesis of csr:

$$D = 2\pi\lambda\sum_{i=1}^{n} X_i^2$$

λ, which is unknown, is needed to compute the test statistic.

Solution: Select *m* quadrats randomly and obtain $\hat{\lambda}$. Then use $\hat{\lambda}$ in computing the test statistic *D*.

Mountford (1961) showed that the distribution of *D* is then asymptotically normal with mean 1 and variance

$$\frac{1+\frac{n+1}{m\lambda}}{n}$$

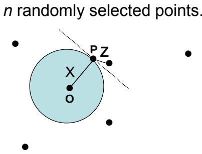
T-Square Sampling

Measure the distance *X* from a randomly selected point O to event to its nearest neighbor P

Construct a line perpendicular to the segment OP, dividing the sampling area (plane) into two regions (half-planes)

The distance *Y* from the event P to the nearest event Q **not** in the half-plane containing the sample point is measured.

Repeat for *n* randomly selected points.



T-Square Sampling

Test Statistic:

$$E = 2\frac{\sum_{i=1}^{n} X_i^2}{\sum_{i=1}^{n} Y_i^2}$$

Under the null hypothesis of csr, E has an F distribtuion with 2n and 2n degrees of freedom. Small values of E indicate a more regular spatial pattern, and large values provide evidence of an aggregated spatial pattern

T-square sampling has been found to be powerful against a variety of alternatives.

A Caution: Small Sampling Regions

A basic assumption for all nearest-neighbor methods is that the n nearest-neighbor measurements are independent.

This assumption may be violated if the sampling region is small

Cressie (1993) found that the hypothesis of csr is rejected too often when this assumption is violated. If the test statistic is close to a critical value, simulation methods should be considered.

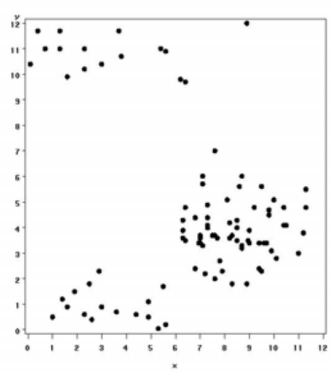
All Events in a Region are Mapped

The K and L functions are becoming the standard for analysis of such data

Tests based on the K and L functions are the most powerful.

If one can afford to do so, mapping all events in a region and using this information is the best thing to do.

12m x 12m plot



K Function

Consider a circle of radius h centered at an arbitrary event

Suppose the circle is wholly within study region

- Area is πh^2
- Expected number of other events in circle is $\lambda \pi h^2$

For a given h, we could put a circle of this radius about every event in the study region and compute the average number of other events in these circles of radius h.

Ignoring the problem of what happens when a circle lies partially outside of the region, we could consider this average to be an estimate of $\lambda \pi h^2$ if the process has csr.

K Function

We have the average number of other events in the circles of radius h resulting from putting a circle about every event in the study region as an estimate of $\lambda \pi h^2$.

So that the measure does not depend on λ, we divide the average by λ. Under csr, this quantity should be about πh^2. It will be larger than πh^2 when there is aggregation and less than πh^2 when the events are more regularly spaced.

Repeat the process for many different h's. This is the idea behind the *K* function:

$K(h) = \lambda^{-1}E$(number of other events within h units of an arbitrary event), $\lambda > 0$

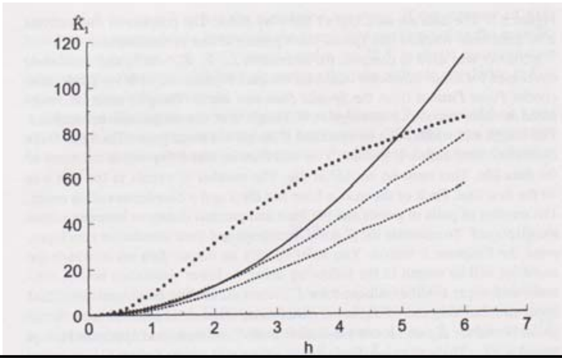
Monte Carlo Test for CSR

Monte Carlo tests are useful for testing hypotheses for which the distribution of a potential test statistic is intractable.

For CSR, we would do the following:

1. Generate 93 observations under CSR in a 12m x 12m plot.
2. Compute K for that plot
3. Repeat (1) and (2) 100 times
4. Plot the observed K function, the mean of the 100 K functions and the min and max of the K functions.

K-Function, No Adjustment for Edge



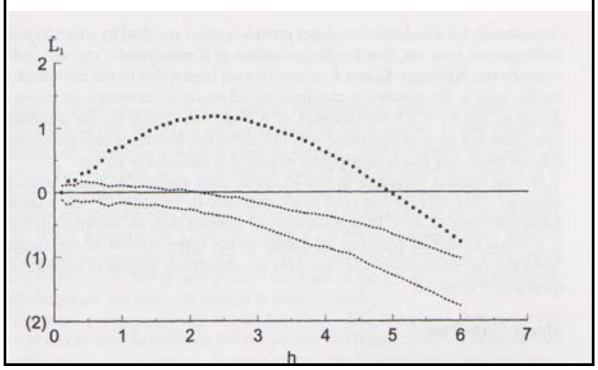
L Function

The *K* function increases with *h*. In measures of aggregation, a 0 or 1 implied no aggregation. To obtain a function that would be constant under CSR, we define

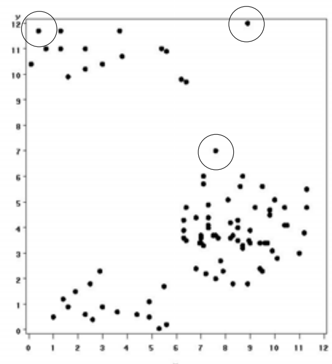
$$\hat{L}_1 = \sqrt{\frac{\hat{K}_1(h)}{\pi}} - h$$

Under the assumption of csr, *L*(*h*) is 0. Under regularity, *L*(*h*) tends to be less than 0. Under clustering, *L*(*h*) tends to be greater than 0.

L-Function, No Adjustment for Edge



12m x 12m plot



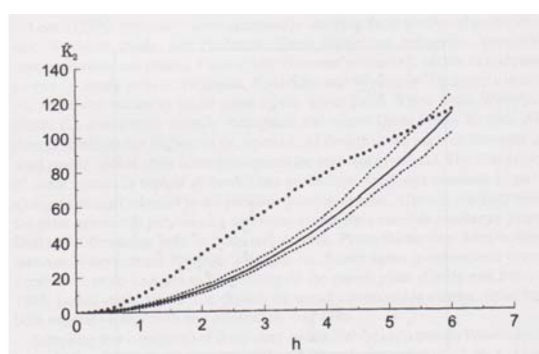
K Function: Accounting for Edge Effects

Suppose that a circle lies partially outside of the region. If *w* is the proportion of the area of the circle that lies within *A*, then we would expect $w\lambda\pi h^2$ events of *A* to be within the circle.

Then, we could use

$$\hat{K}_2(h) = \frac{1}{\hat{\lambda}} \sum_{\substack{i=1 \\ i \neq j}}^{N} \sum_{j=1}^{N} \frac{\frac{1}{w(x_i, x_j)} I(\| x_i - x_j \| \le h)}{N}$$

K-Function, Adjustment for Edge



L Function

The *K* function increases with *h*. In measures of aggregation, a 0 or 1 implied no aggregation. To obtain a function that would be constant under CSR, we define

$$\hat{L}_2 = \sqrt{\frac{\hat{K}_2(h)}{\pi}} - h$$

Under the assumption of csr, $L(h)$ is 0. Under regularity, $L(h)$ tends to be less than 0. Under clustering, $L(h)$ tends to be greater than 0.

L-Function, Adjustment for Edge