Spatial point process

Odd Kolbjørnsen

13.March 2017

Today

- Point process
- Poisson process
  - Intensity
    - Homogeneous
    - Inhomogeneous
- Other processes
  - Random intensity (Cox process)
    - Log Gaussian cox process (LGCP)
  - Clustering
    - Parent – Child
    - Neyman-Scott
  - Repulsion (regularity)
    - Markov point process
    - Strauss

- Inference
  - Estimation of intensity

  - Detection of clustering and repulsion
    - Pair correlation function
    - K-function/L-function

Point process

- Stochastic process $Z(.)$ describing the location of events (points) in a region $D_s \subset \mathbb{R}^d$

- Both the number of points and the location of points are random

- e.g The number of earth quakes and their locations

http://www.norsardata.no/NDC/recenteq/lastweek.html
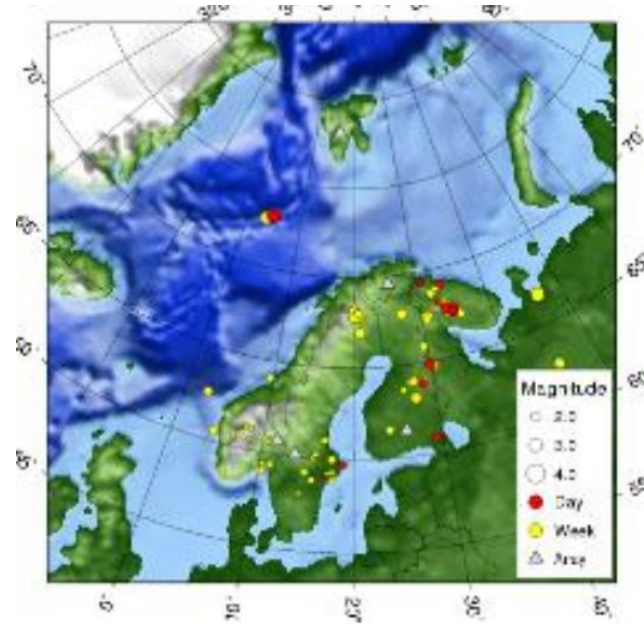
Marked point process:

- Additional parameters related to the event are denoted marks

- e.g magnitude

Poisson point process
(homogeneous)

Poisson distribution:

- The number of points in a region A has a Poisson distribution with intensity: $\lambda|A|$

$$p(N(A) = n)$$
$$= \frac{(\lambda|A|)^n \exp(-\lambda|A|)}{n!}$$

Complete independence:

- The number of points in two non-overlapping regions are independent

- $\lambda|A| = \int_A \lambda \, d\boldsymbol{s}$
- $\lambda > 0$

CSR= Complete spatial randomness

Poisson point process
(inhomogeneous)

Poisson distribution:

- The number of points in a region A has a Poisson distribution with intensity: $\Lambda(A)$

$$p(N(A) = n)$$
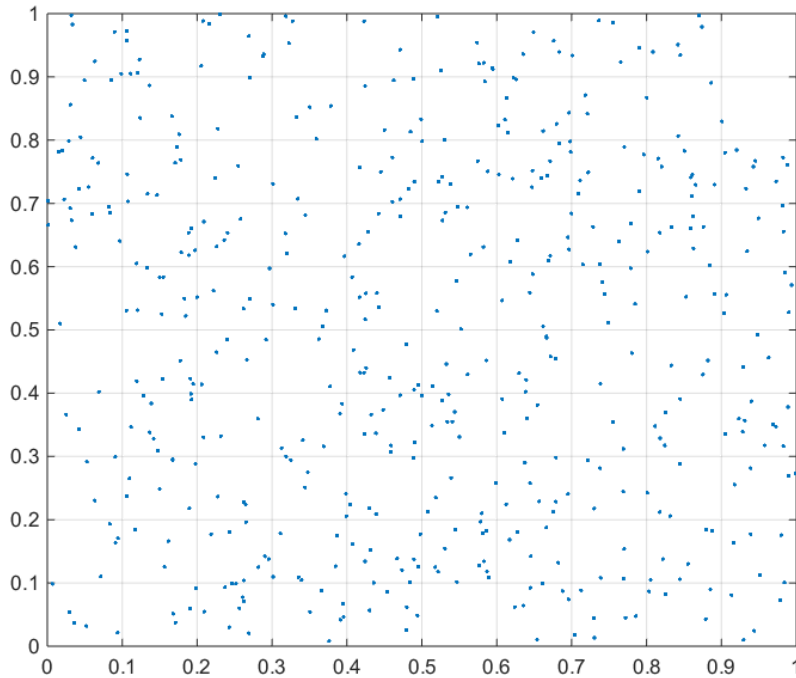$$= \frac{(\Lambda(A))^n \exp(-\Lambda(A))}{n!}$$

Complete independence:

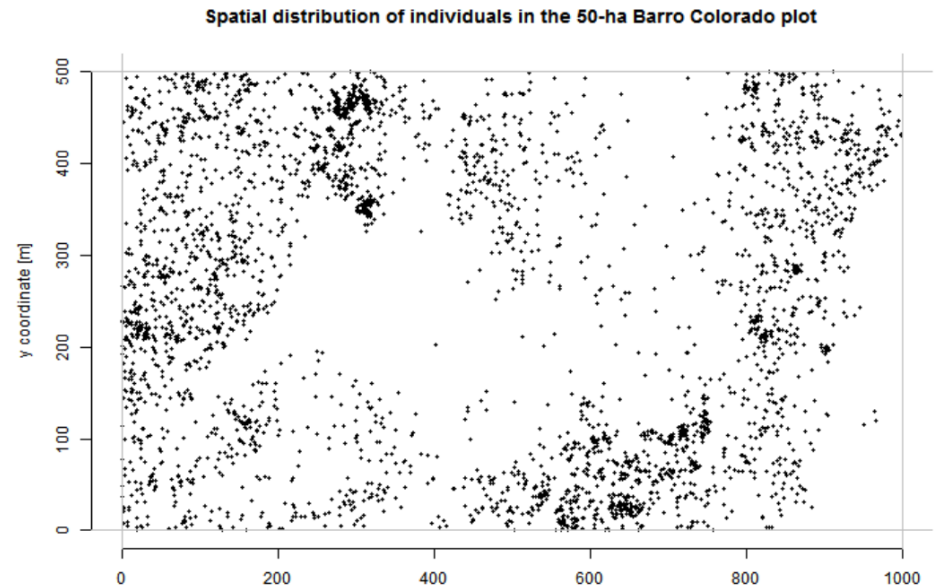- The number of points in two non-overlapping regions are independent

- $\Lambda(A) = \int_A \lambda(\boldsymbol{s})d\boldsymbol{s}$
- $\lambda(\boldsymbol{s}) > 0 \quad \forall \boldsymbol{s}$

Homogeneous versus inhomogeneous

Homogeneous

Inhomogeneous

Spatial distribution of individuals in the 50-ha Barro Colorado plot

Homogeneous does not mean regular, common to see areas
which appears as inhomogeneity but can equally well be caused by
random fluctuations

Density of Poisson process

- In the Poisson process what is the probability for the point pattern $\{\boldsymbol{s}_i\}_{i=1}^m$?

- $\Pr(Z(d\boldsymbol{s}_1) = 1, \ldots, Z(d\boldsymbol{s}_n) = 1, Z(A\backslash\{\boldsymbol{s}_i\}_{i=1}^m) = 0)$

$$= \prod_{i=1}^m \lambda(\boldsymbol{s}_i) d\boldsymbol{s}_i \exp\{-\lambda(\boldsymbol{s}_i)d\boldsymbol{s_i}\} \exp\{-\Lambda(A\backslash\{\boldsymbol{s}_i\}_{i=1}^m)\}$$

$$= \prod_{i=1}^m \lambda(\boldsymbol{s}_i) \exp\{-\Lambda(A)\} \prod_{i=1}^m d\boldsymbol{s}_i$$

$$p(N = n|\Lambda) = \frac{\Lambda^n \exp(-\Lambda)}{n!}$$

Density of point process

- In the general case what is the probability for the point pattern $\{s_i\}_{i=1}^m$?

- $\Pr(Z(ds_1) = 1, \ldots, Z(ds_n) = 1 | Z(A) = m)$

$$\propto \lambda_m(s_1, s_2, \ldots s_m) \prod_{i=1}^m ds_i$$

- *Full distribution:*

$\Pr(Z(ds_1) = 1, \ldots, Z(ds_n) = 1 | Z(A) = m) \Pr(Z(A) = m)$

$$= \lambda_m(s_1, s_2, \ldots s_m) \prod_{i=1}^m ds_i \Pr(Z(A) = m)$$

Higher order intensity functions

- $\lambda_m(s_1, s_2, \ldots s_m)$
  - Sequence m=1,2,3,…
  - Nonnegative
  - Permutation invariant

  For fixed number of points this is a
  Joint multivariate distribution,
  but the number of samples m vary.

Other processes

- Cox model:
  - Random intensity
    $$p(N(A) = n|\Lambda) = \frac{(\Lambda(A)\,)^n \exp(-\Lambda(A)\,)}{n!}$$
    $p(\Lambda) = $ gives different models
  - Log Gaussian Cox Model
- Clustering
  - Parent child process
  - Neyman-Scott
- Repulsion
  - Markov point process
  - Strauss process

Log Gaussian Cox process

$$p(N(A) = n|\Lambda) = \frac{(\Lambda(A))^n \exp(-\Lambda(A))}{n!}$$

$$\lambda(\boldsymbol{s}) = \exp(Y(\boldsymbol{s})),$$

$$\Lambda(A) = \int_A \lambda(\boldsymbol{s})d\boldsymbol{s}$$

$$Y(\boldsymbol{s}) = \beta^T \boldsymbol{x}(\boldsymbol{s}) + R(\boldsymbol{s})$$

$$R(s) \sim N(0, C_R(\boldsymbol{h}))$$

Recall:

Spatial disease mapping

$$Z_i|Y_i \overset{ind}{\sim} \text{Poisson}(E_i \exp(Y_i))$$

$Z_i =$ Observed disease count
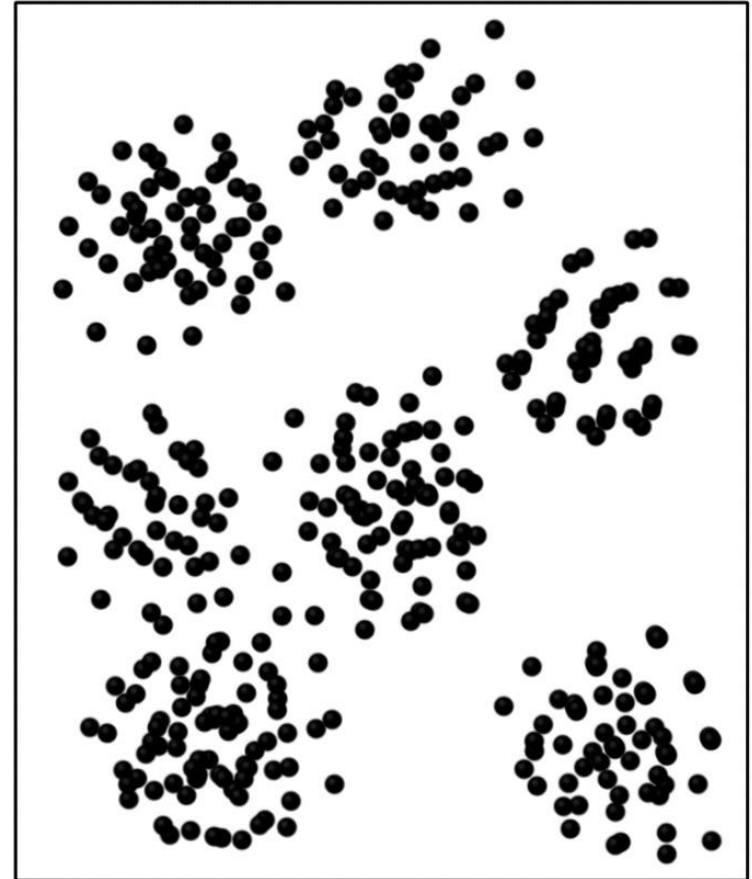
$E_i =$ Expected count (known), and

$$Y_i = \mathbf{x}_i^T \beta + \delta_i + \varepsilon_i$$
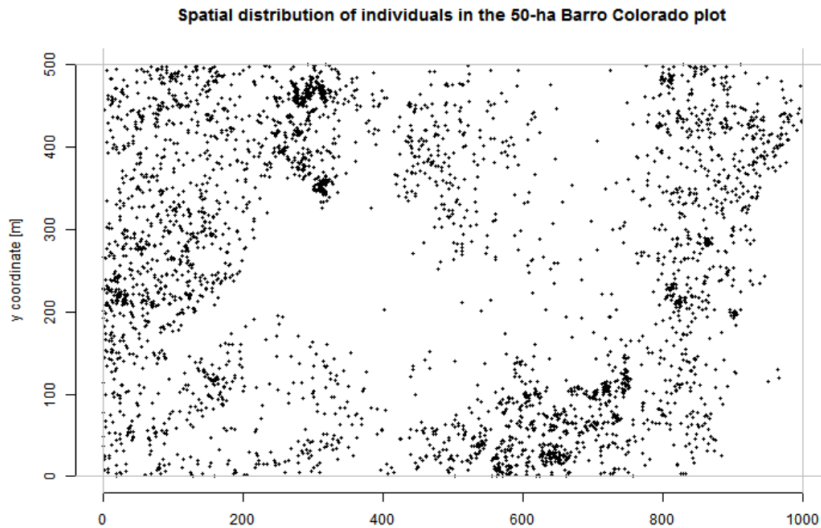
Clustering point process

- In some occasions Data points tend to group together

- Stars in the universe cluster in galaxies

- The clusters are "self organizing" i.e. no external force
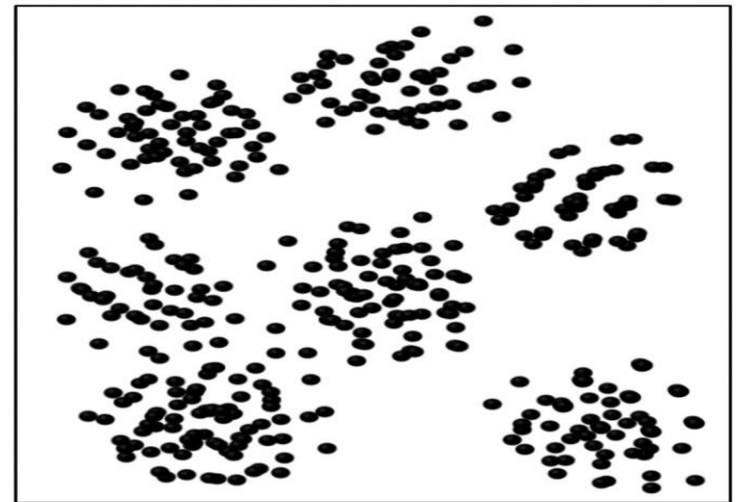
Informal definition of cluster:

- A cluster is a group of points whose intra-point distance is below the average distance in the pattern as a whole.

Clustering versus inhomogeneous

Spatial distribution of individuals in the 50-ha Barro Colorado plot

Inhomogeneous : model there is an external feature which causes the inhomogeneity in the model

Clustering model:
Then the point process "organize" , the cause is internal

Not always clear distinction: The term clusters are sometimes used for describing high intensity areas in inhomogeneous distributions
"Earth quakes cluster around tectonic boundaries"

Clustering processes

- ## Parent-child process

  - Parent events are distributed according to one process

  - Children events are distributed conditioned to parent distribution

  - The final process is just the children
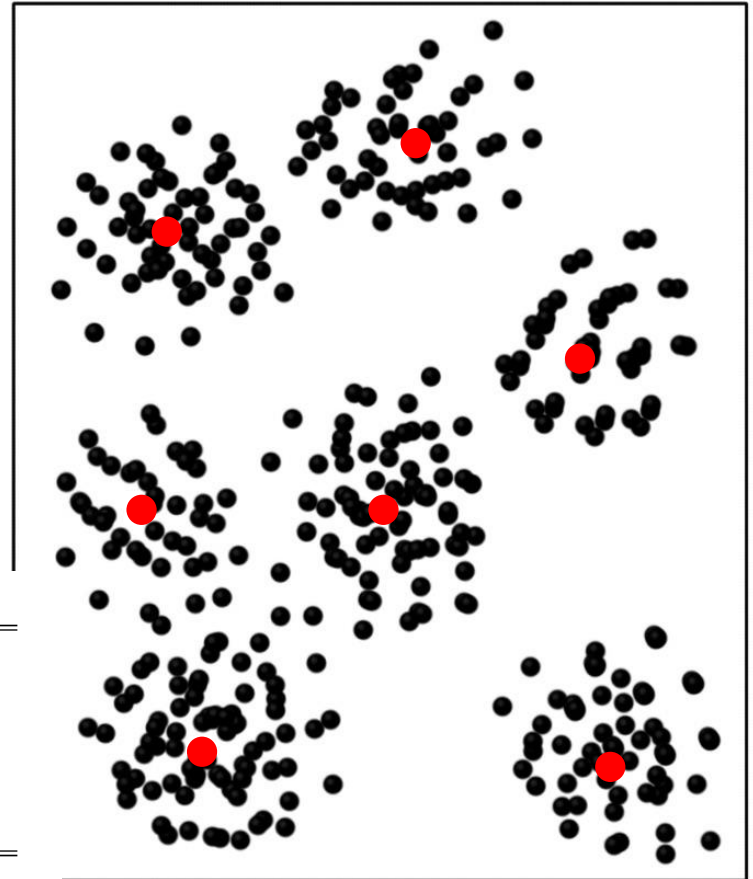
| Name of Process	Clusters	Parents
Independent cluster process	general	general
Poisson cluster process	general	Poisson
Cox cluster process	Poisson	general
Neyman-Scott process	Poisson	Poisson
Matérn cluster process	Poisson (uniform in ball)	Poisson (homog.)
Modified Thomas Process	Poisson (Gaussian)	Poisson (homog.)

Table 2: Nomenclature for independent cluster processes.

● parent
● child

Neyman-Scott process

- Parent $Z_p(s)$ is Poisson

- Children given parents $\{s_{p,i}\}_{i=1}^{N_p}$

  is inhomogeneous Poisson:

  $$\lambda(s|Z_p) = \sum_{i=1}^{N_p} \alpha \cdot \frac{1}{\sigma} f((s - s_p)/\sigma)$$

  - α : cluster intensify
  - σ : cluster spread
  - $f(s)$: probability density

- Neyman-Scott is also a Cox process (random intensity)

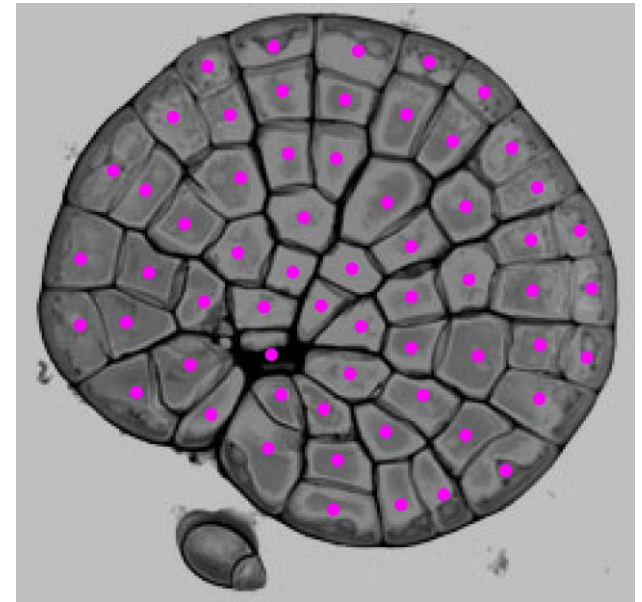- Many special cases and generalizations
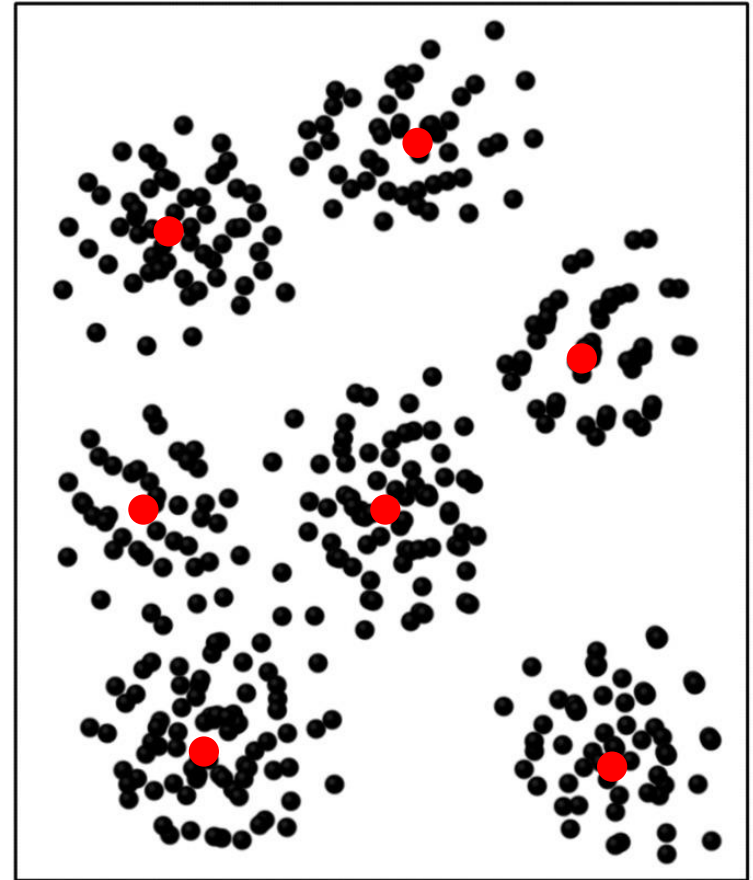
● parent
● child

Repulsion

- In some occasions data points tend to be more regular than the Poisson process

- Cell centers in tissue repels each other

- The repulsion is "self organizing"(not due to external forces)

Coleochaete

Organism with multicellular organization

Repulsion is often modelled using Markov point process

$$\lambda_m(\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots \boldsymbol{s}_m)$$

$$\propto \exp\left\{\sum_{i=1}^{m} g_i(\boldsymbol{s}_i) + \sum_{i<j} g_{ij}(\boldsymbol{s}_i, \boldsymbol{s}_j) + \cdots + g_{12\ldots m}(\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_m)\right\}$$

Strauss process:

$$g_i(\boldsymbol{s}_i) = \log \lambda$$

$$g_{ij}(\boldsymbol{s}_i, \boldsymbol{s}_j) = \log \gamma \, I(\|\boldsymbol{s}_i - \boldsymbol{s}_j\| < R)$$

$$g_{\text{Higher order}} \equiv 0$$

Strauss process

$$\lambda_m(\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots \boldsymbol{s}_m) = \frac{\lambda^m \gamma^{Y_m(R)}}{c(\lambda, \gamma, R)}$$

With $Y_m(R) = \Sigma_{i<j} I(\|\boldsymbol{s}_i - \boldsymbol{s}_j\| < R)$
number of pairs separated by less than R.

$c(\lambda, \gamma, R)$ is independent of m, thus gives the global scaling
(i.e sum over all m)

Hard core model:
No points within a distance R

$\lambda > 0$

$R > 0$

$0 \leq \gamma \leq 1$

$\quad \gamma > 1$ is impossible to normalize

$\quad \gamma = 0$ Hard core model

$\quad \gamma = 1$ Poisson process
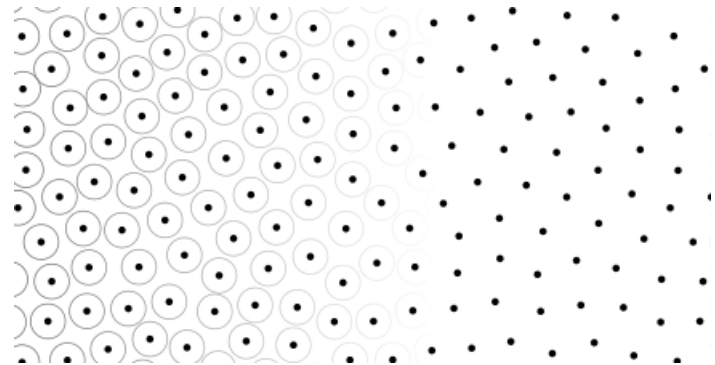
Estimation

- Inhomogeneous intensity
- Non parametric estimator : (Kernel estimator)

$$\hat{\lambda}(\boldsymbol{s}) = \frac{1}{p_{\boldsymbol{b}}(s)} \sum_{i=1}^{m} K_b(\boldsymbol{s} - \boldsymbol{s}_i)$$

- $p_{\boldsymbol{b}}(s)$ account for edge effect
- K_b probability density
- $b > 0$ bandwidth

Pair correlation function

- $g(s, x) = \lambda_2(\boldsymbol{s}, \boldsymbol{x}) / (\lambda(\boldsymbol{s})\lambda(\boldsymbol{x}))$

- Intensity

$$\lambda(\boldsymbol{s}) = E(Z(d\boldsymbol{s}))/d\boldsymbol{s}$$

- Second order intensity

$$\lambda_2(\boldsymbol{s}, \boldsymbol{x}) = E(Z(d\boldsymbol{s})Z(d\boldsymbol{x}))/(|d\boldsymbol{s}||d\boldsymbol{x}|)$$

If $Z(.)$ is stationary : $\lambda_2(\boldsymbol{s}, \boldsymbol{x}) = \lambda_2(\boldsymbol{s} - \boldsymbol{x})$

and also isotropic : $\lambda_2(\boldsymbol{s}, \boldsymbol{x}) = \lambda_2(\|\boldsymbol{s} - \boldsymbol{x}\|)$

Pair correlation function

- Poisson distribution:
$$\lambda_2(\boldsymbol{s}, \boldsymbol{x}) = 1$$

- Co occurrence more frequent than Poisson
$$\lambda_2(\boldsymbol{s}, \boldsymbol{x}) > 1$$

- Co occurrence less often than Poisson
$$\lambda_2(\boldsymbol{s}, \boldsymbol{x}) < 1$$

- Estimation using kernel based methods

- Clustering we have $\lambda_2(\boldsymbol{s}, \boldsymbol{x}) > 1$ when $\|\boldsymbol{x} - \boldsymbol{s}\|$ is small
- Repulsion we have $\lambda_2(\boldsymbol{s}, \boldsymbol{x}) < 1$ when $\|\boldsymbol{x} - \boldsymbol{s}\|$ is small

K-function

- Assume stationary and isotropic point process

$$K(h) = E\left\{\begin{array}{c}\text{number of events within}\\\text{distance h of an abitrary event}\end{array}\right\}/\lambda$$

- Related to pair correlation function
- Estimator:

$$\widehat{K}(h) = \frac{1}{m\hat{\lambda}}\sum_{i=1}^{m}\sum_{\substack{j=1\\j\neq i}}^{m}I\left(\|s_i - s_j\| \leq h\right)$$

$$\hat{\lambda} = m/|A|$$

Also edge effects se book 4.180

K-function for an Poisson process

- In $\mathbb{R}^d$

$$K(h) = \pi^{\frac{d}{2}} / \Gamma\left(1 + \frac{d}{2}\right) h^d$$

- In $\mathbb{R}^2$

$$K(h) = \pi h^2$$

To compare with Poisson distribution use L function which center and normalize K for Poisson distribution

L-function

- In $\mathbb{R}^d$

$$L(h) = \left(\frac{K(h)\Gamma\left(1 + \frac{d}{2}\right)}{\pi^{d/2}}\right)^{1/d} - h$$

- In $\mathbb{R}^2$

$$L(h) = \left(\frac{K(h)}{\pi}\right)^{1/2} - h$$

Some definitions do not subtract h

Clustering / Repulsion

- Clustering
  - More events at short distance than Poisson
  - If Estimated K function is larger than the response from Poisson process
  - If L-function is positive (significantly positive)
- Repulsion
  - Less events at short distance than Poisson
  - If Estimated K function is less than the response from Poisson process
  - If L-function is negative (significantly negative)
- Example Page 213 in book