# Spatial Point Patterns

## Interaction

Josep Lluís Carrasco

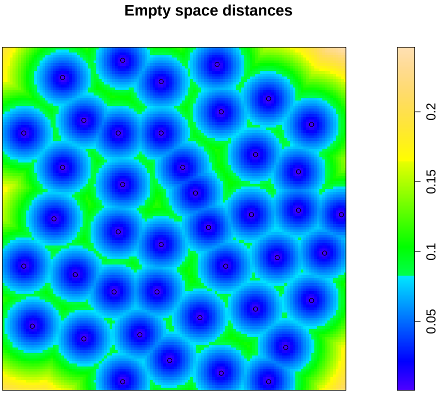
Departament de Fonaments Clínics
Universitat de Barcelona

# Outline

# Outline

# Empty space distances

- The distance from a fixed reference location $u$ in the window to the nearest data point.



# Empty space distances

$$d\left(u, \mathbf{x}\right) = min\left\{||u - x_i|| : x_i \in \mathbf{x}\right\}$$

- Cumulative distribution function

$$F\left(r\right) = P\left[d\left(u, \mathbf{X}\right) \leq r\right]$$

- Empirical distribution function

$$F^*\left(r\right) = \frac{1}{m}\sum_j \mathbf{1}\left\{d\left(u_j, x\right) \leq r\right\}$$

- Empirical *edge corrected* distribution function

$$\hat{F}\left(r\right) = \sum_j e\left(u_j, r\right)\mathbf{1}\left\{d\left(u_j, x\right) \leq r\right\}$$

where $e\left(u_j, r\right)$ is a weight to correct the edge effect.

# Edge correction

- Simplest approach: **border method**.
- Only use those points at a distance $r$ to the area boundary.
- For each $r$ the function is estimated as

$$\hat{F}^{b}(r) = \frac{\hat{F}(r) \cap A_{(-r)}}{|A_{(-r)}|}$$

- Other option: edge effect as a censoring issue (Baddeley and Gill, 1997)

# Empty space distances

- Homogeneous Poisson Point Process

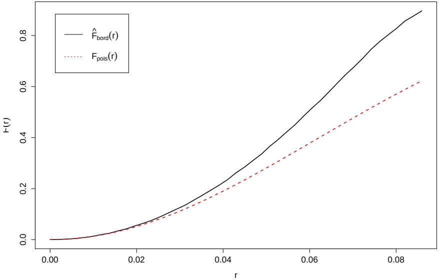
$$F_{pois}\left(r\right) = 1 - exp\left(-\lambda \pi r^2\right)$$

- Regular pattern

$$\hat{F}\left(r\right) > F_{pois}\left(r\right)$$

- Clustered pattern

$$\hat{F}\left(r\right) < F_{pois}\left(r\right)$$



# Nearest neighbour distances

$$t_i = \min_{j \neq i} ||x_i - x_j||$$

- Cumulative distribution function

$$G(r) = P\left[d\left(u, \mathbf{X}\{u\}\right) \leq r | u \in \mathbf{X}\right]$$

- Empirical distribution function

$$G^*(r) = \frac{1}{n} \sum_i \mathbf{1}\{t_i \leq r\}$$

- Empirical *edge corrected* distribution function

$$\hat{G}(r) = \sum_i e(x_i, r)\, \mathbf{1}\{t_i \leq r\}$$

# Nearest neighbour distances

- Homogeneous Poisson point process

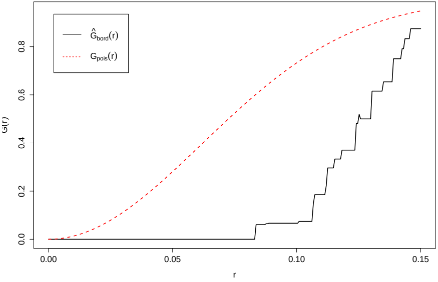
$$G_{Pois}\left(r\right) = 1 - exp\left(-\lambda\pi r^2\right)$$

- Regular pattern

$$\hat{G}\left(r\right) < G_{Pois}\left(r\right)$$

- Clustered pattern

$$\hat{G}\left(r\right) > G_{Pois}\left(r\right)$$



# Pairwise distances

$$s_{ij} = ||x_i - x_j||$$

- Expected number of other points of the process within a distance r

$$E\left[n\left(\mathbf{X}, r\right)\right] = \lambda \cdot K\left(r\right)$$

where $K\left(r\right)$ is the Ripley's K-function.

- Estimator of $K\left(r\right)$

$$\hat{E}\left[n\left(\mathbf{X}, r\right)\right] = \frac{1}{N}\sum_{i}\sum_{j \neq i}\mathbf{1}\left\{||x_i - x_j|| \leq r\right\}$$

$$\hat{\lambda} = \frac{N}{area\left(A\right)}$$

$$\hat{K}\left(r\right) = \frac{area\left(A\right)}{N^2}\sum_{i}\sum_{j \neq i}\mathbf{1}\left\{||x_i - x_j|| \leq r\right\} e\left(x_i, x_j, r\right)$$

- **Isotropic method**. $e\left(x_i, x_j, r\right)$ is the proportion of a disc centered at point $x_i$ and passing through point $x_j$ that lies within the study area.

# Pairwise distances

- Homogeneous Poisson point process

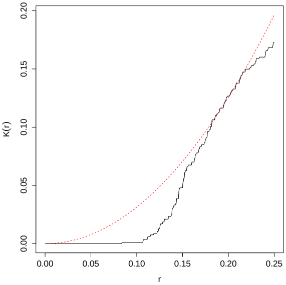
$$K_{Pois}(r) = \pi r^2$$

- Regular pattern

$$\hat{K}(r) < K_{Pois}(r)$$

- Clustered pattern

$$\hat{K}(r) > K_{Pois}(r)$$



# Additional related functions

- L-function

$$L(r) = \sqrt{\frac{K(r)}{\pi}} \quad L_{Pois}(r) = r$$

- *Pair correlation function*. Probability of observing a pair of points separated by a distance r, divided by the corresponding probability for a Poisson process.

$$g(r) = \frac{K'(r)}{2\pi r} \quad g_{Pois}(r) = 1$$

Clustering $g(r) > 1$; Regularity $g(r) < 1$

- J function

$$J(r) = \frac{1 - G(r)}{1 - F(r)} \quad J_{Pois} = 1$$

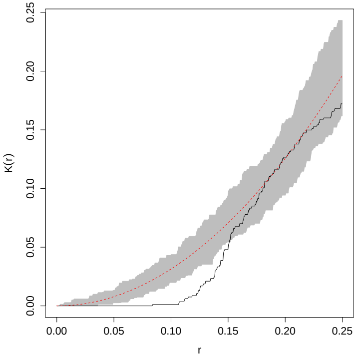
# Envelopes and Montecarlo tests

1. Suppose the reference curve is the theoretical K function for CSR.
2. Generate M independent simulations of CSR inside the study region A.
3. Compute the estimated K functions for each of these realizations.
4. Obtain the pointwise upper and lower envelopes of these simulated curves.

$$L(r) = \min_j \hat{K}^{(j)}(r)$$

$$U(r) = \max_j \hat{K}^{(j)}(r)$$

# Envelopes and Montecarlo tests



# Envelopes and Montecarlo tests

- For any fixed value of r the significance level is $\alpha = 2/\left(M+1\right)$.
- r must be fixed in advance
- If the test was used as *Does K(r) lie outside for any r?* the significance level would be higher.
- Simultaneous Monte Carlo test $\rightarrow$ simultaneous critical bands with a global significance level $1/\left(M+1\right)$.

1. For each estimate $\hat{K}_i\left(r\right)$, $i = 1,\ldots,M$ compute its maximum deviation from Poisson K function

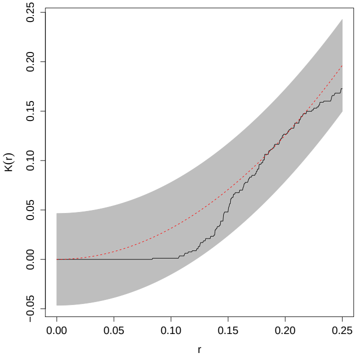
$$D_i = \max\left|\hat{K}_i\left(r\right) - K_{Pois}\left(r\right)\right|$$

2. $D_{max} = \max\left\{D_i\right\}$

$$L\left(r\right) = \pi r^2 - D_{max}$$

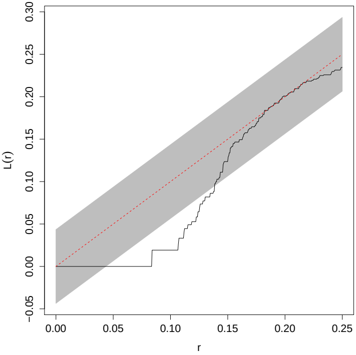
$$U\left(r\right) = \pi r^2 + D_{max}$$

# Envelopes and Montecarlo tests



# Envelopes and Montecarlo tests

A more powerful test is obtained if the variance is stabilized using $\sqrt{K(r)}$.

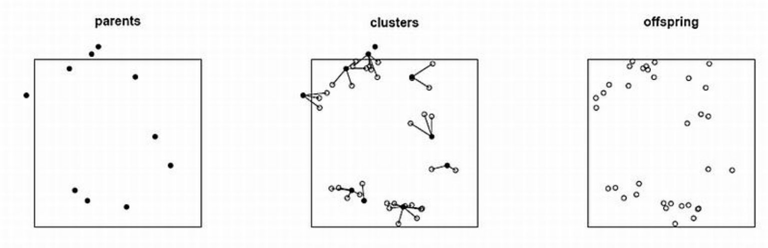


# Outline

# Outline

# Poisson cluster processes

- Each event belongs to a particular cluster.
- The process consist of a set of *parent locations* each of which generates a set of *child locations*.
- Only *child locations* are observed.



# Neyman-Scott processes

1. Parent locations follow a Poisson process
2. Numbers of children per parent are i.i.d..
3. Locations of children around their respective parent are i.i.d.

- Matérn process.
  - Parent points come from a Poisson process with intensity $\lambda\left(x\right)$.
  - Each parent has a $Poisson\left(\mu\right)$ number of offspring.
  - Children has an independently and uniformly distributed displacement from its parent in a disc of radius r centred around the parent.
- Thomas process.
  - Parent points come from a Poisson process with intensity $\lambda\left(x\right)$.
  - Each parent has a $Poisson\left(\mu\right)$ number of offspring.
  - Children has an independently and isotropic Gaussian $N\left(0,\sigma^2\mathbf{I}\right)$ distributed displacement from its parent in a disc of radius r centred around the parent.

# Cox processes

- Poisson process with a random intensity function.
- Most used: log-Gaussian Cox Process.
- Intensities $\lambda(\mathbf{u}) \sim N(exp(\beta \mathbf{X}), C(r))$
- Covariance between locations takes the form:

$$C(r) = \sigma^2 f(r, \alpha)$$

- $\sigma^2$ and $\alpha$ controls the scale and the strength of the autocorrelation.
- $f(r, \alpha)$ is a decay function. Two cases are:
  - Exponential. $f(r, \alpha) = exp(-r/\alpha)$
  - Gaussian $f(r, \alpha) = exp\left(-(r/\alpha)^2\right)$

# Model fitting: estimation method

- **Minimum contrast** (Diggle, 1983)

$$D(\theta) = \int_a^b |\hat{K}(r)^q - K_\theta(r)^q|^p dr$$

- $a$, $b$ and $q$ are used to control the sampling fluctuations in the estimates of K.
    - $a$, $b$: limits of the range of distances.
    - $q > 0$ is used to transform the $K$ function to more powerful options.
    - If $q = 0.5$ the contrast uses the $L$ function.
    - In spatial processes commonly $p = 2$.
- It gives consistent estimates.
- It can be very computationally intensive.

Diggle, PJ (1983). Statistical Analysis of Spatial Point Patterns. Chapter 5. Academic Press, London 1983

# Model fitting: estimation method

- **Composite likelihood** (Guan, 2006).
- It maximises a composite likelihood based on knowledge of the second moment of intensity (covariance between locations).
- **Palm likelihood** (Tanaka et al, 2008).
- It assumes a Palm distribution for the point process.
- Asymptotic normality of estimates (if the likelihood is correct).

Guan, Y. (2006). Journal of the American Statistical Association 101, 1502-1512

Tanaka, U. and Ogata, Y. and Stoyan, D. (2008). Biometrical Journal 50, 43-57

# Gibbs point processes

- Spatial point process models that are constructed by writing down their probability densities.

$$f(\mathbf{X}) = \alpha \left[ \prod_{i=1}^{n} b(x_i) \right] \left[ \prod_{i<j} c(x_i, x_j) \right]$$

- $\alpha$ is a normalising constant.
- $b(x_i)$ are the *first order* terms.
- $c(x_i, x_j)$ are the *second order* terms.
- Intensity. $\lambda(u) = b(u) \left[ \prod_{i=1}^{n} c(u, x_i) \right]$
- Gibbs point processes can model regular patterns.
- Estimation approach: Maximum pseudolikelihood

# Gibbs point processes for regular patterns

**Hard core process**

- Intensity. $b\left(u\right) = \lambda$
- Pairwise interaction.

$$c(u,v) = \begin{cases} 1 & \text{if} ||u-v|| > r \\ 0 & \text{if} ||u-v|| \leq r \end{cases}$$

**Strauss process**

- Pairwise interaction.

$$c(u,v) = \begin{cases} 1 & \text{if} ||u-v|| > r \\ \gamma & \text{if} ||u-v|| \leq r \end{cases}$$

$$0 \leq \gamma \leq 1$$

- Intensity. $b\left(u\right) = \lambda \cdot \gamma^{t(u)}$

where $t\left(u\right)$ is the number of points at a distance lower than r.

# Gibbs point processes for cluster patterns

**Geyer process**

- Pairwise interaction.

$$c(u,v) = \begin{cases} 1 & \text{if} ||u-v|| > r \\ \gamma & \text{if} ||u-v|| \leq r \end{cases}$$

- Intensity. $b\left(u\right) = \lambda \cdot \gamma^{min(s,t(u))}$

where

- $t\left(u\right)$ is the number of points at a distance lower than r.
- $s$ is the saturation parameter.

# Gibbs point processes for cluster patterns

**Area Interaction process**

- Two points interact if their discs of radius *r* overlaps.
- Pairwise interaction.

$$c(u,v) = \begin{cases} 1 & \text{if} ||u-v|| > 2r \\ \gamma & \text{if} ||u-v|| \leq 2r \end{cases}$$

- Intensity. $b\left(u\right) = \lambda \cdot \gamma^{-A}$

where A is the area of the region formed by the union of discs of radius r centred at the observed points.