Lecture 5. Point Pattern Data

Spatial Big Data Analysis with GIS

Korean Statistical Society, Winter School, February 24, 2023

Spatial point patterns

- Geostatistical data: Random variables $Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)$ at fixed spatial locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$.
- Spatial point pattern: The spatial locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ and the number of points n are random, typically a realization from a point process.
- Marked spatial point pattern: $\mathbf{s}_1, \ldots, \mathbf{s}_n, n$, and $Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)$ are all random. $Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_n)$ are called 'marks'.
- For example, the patterns of trees in a forest, occurrence of disease, distribution of commercial properties ...

  Here, we will focus on point patterns (without marks).

Issues of interest

- Are the points completely random or do they have some meaningful patterns?

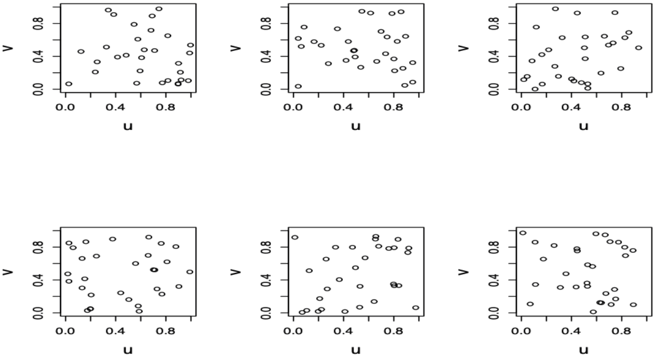


Figure 8.1 *The panels depict* spatial homogeneity *for six samples each of 30 points. The plots reveal that the eye cannot easily assess complete randomness and tends to look for structure.*

Issues of interest (conti-)

- Do the patterns show any clustering?

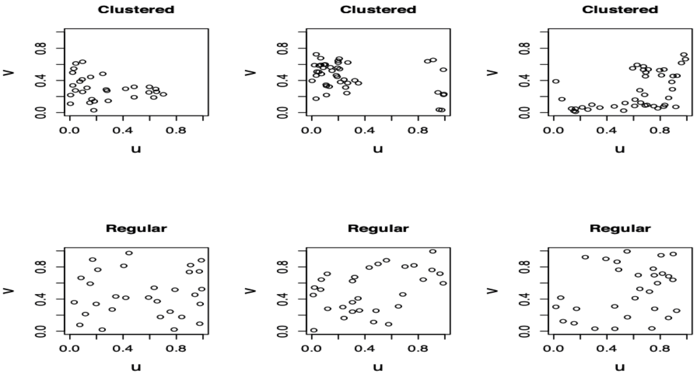


Figure 8.2 *Clustering and systematic (regular) pattern.*

Issues of interest (conti-)

- Does the *intensity* of occurrence change depending on the location?

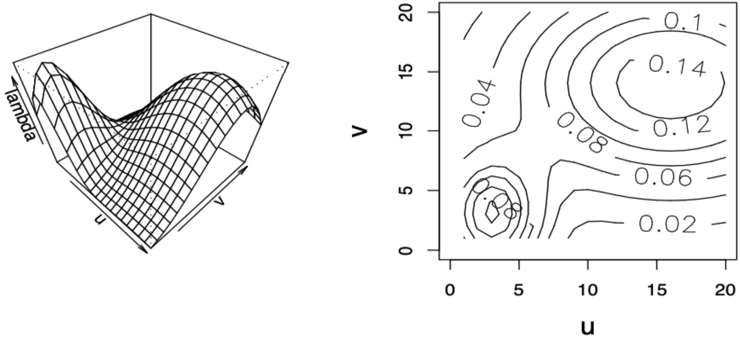


Figure 8.3 *Intensity surface used to generate point patterns.*

Issues of interest (conti-)

- How covariates affect the occurrence of events?

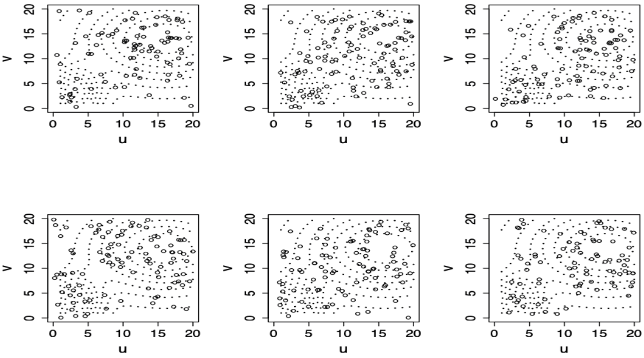


Figure 8.4 *Realizations from the intensity surface in Figure 8.3 with overlaid contours shown as dashed lines.*

Notation and basic definitions

- We focus on point patterns over $D \subset \mathbb{R}^2$. D is the domain of interest.
- A random realization of a point pattern $\mathbf{S} = (\mathbf{s}_1, \ldots, \mathbf{s}_n)$ where $\mathbf{s}_i \in D$ for $i = 1, \ldots, n$.
- We need distribution of (i) the total number of points $N(D)$ where $N(\cdot)$ is the number of points in an area, and (ii) the locations of points $\mathbf{s}_1, \ldots, \mathbf{s}_n$ given $N(D) = n$.
- Let $f(\mathbf{s}_1, \ldots, \mathbf{s}_n)$ be the *location density*. Since points are exchangeable, f must be symmetric in its arguments.
- Stationarity: $f(\mathbf{s}_1, \ldots, \mathbf{s}_n) = f(\mathbf{s}_1 + \mathbf{h}, \ldots, \mathbf{s}_n + \mathbf{h})$ for all $n, \mathbf{s}_i$, and $\mathbf{h} \in \mathbb{R}^2$.

Point process

- A point process $\{Z(\mathbf{s}) : \mathbf{s} \in D \subset \mathbb{R}^2\}$ consists of a pattern of points in the random set D.
- Bernoulli and Binomial process:
  - If a single event $\mathbf{s}$ is distributed in D such that $P(\mathbf{s} \in A) = \nu(A)/\nu(D)$ for all sets $A \subset D$, where $\nu(A)$ gives the "area" of the set A, then we call the process a Bernoulli process.
  - If n Bernoulli processes are supposed to form a process of n events in D, we call the resulting process a Binomial process.
- If $N(A)$ denotes the number of events in the set $A \subset D$, then for a Binomial process, $N(A)$ is a Binomial random variable with sample size $N(D)$ and success probability $\pi(A) = \nu(A)/\nu(D)$.

Point process (conti-)

- The *intensity* $\lambda(\mathbf{s})$ is the average number of events per unit area.
- We define

$$\lambda(\mathbf{s}) = \lim_{\nu(d\mathbf{s}) \to 0} \frac{E\{N(d\mathbf{s})\}}{\nu(d\mathbf{s})}$$

- If the intensity does not change with spatial location, we say the process is homogeneous. Binomial process is a homogeneous process.

Counting measure and Poisson process

- One easy way: Define a point process through $N(B) = \sum_{\mathbf{s}_i \in \mathbf{S}} 1(\mathbf{s} \in B)$ for any $B \subset D$.
- $N(B)$ is a counting measure for a sigma algebra $\mathcal{B}$ for D, with $\forall B \in \mathcal{B}$.
- **Poisson process**: For $B \subset D$, $N(B) \sim \text{Poisson}(\lambda(B))$ where $\lambda(B) = \int_B \lambda(\mathbf{s}) d\mathbf{s}$. $N(B_1)$ and $N(B_2)$ are independent if B_1 and B_2 are disjoint.
  - Note that $E(N(B)) = \text{Var}(N(B)) = \lambda(B)$.
  - The independence of disjoint sets implies

$$f(\mathbf{s}_1, \ldots, \mathbf{s}_n) = \prod_i f(\mathbf{s}_i) = \prod_i \lambda(\mathbf{s}_i)/\lambda(D)$$

where $\lambda(D) = \int_D \lambda(\mathbf{s}) d\mathbf{s}$.

Why $f(\mathbf{s}) = \lambda(\mathbf{s})/\lambda(D)$

- Note that, given $N(D) = n, N(B) \sim B(n, P(B))$ where $P(B) = \int_B f(\mathbf{s})d\mathbf{s}$ by the conditional independence of the locations.

- Therefore,

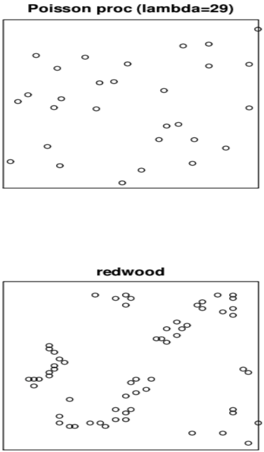
$$\begin{aligned} E(N(B)) &= E(E(N(B)|N(D))) = E(N(D)P(B)) \\ &= E\left(N(D)\int_B f(\mathbf{s})d\mathbf{s}\right) \\ &= \int_B E(N(D))f(\mathbf{s})d\mathbf{s} \\ &= \int_B \lambda(D)f(\mathbf{s})d\mathbf{s} \end{aligned}$$

This implies that $f(\mathbf{s}) = \lambda(\mathbf{s})/\lambda(D)$.

Homogeneous Poisson process (HPP)

- Arises when $\lambda(\mathbf{s}) = \lambda$ (a constant over D), defining the notions of complete spatial randomness (CSR).
- $N(B) \sim \text{Poisson}(\lambda(B))$ where $\lambda(B) = \lambda|B|$ and $|B| =$ (the area of B).
- The location density is given by $f(\mathbf{s}_1, \ldots, \mathbf{s}_n) = 1/|D|^n$.
- Note that stationarity implies $\lambda(\mathbf{s}) = \lambda$, because $\lambda(B) = \lambda(B + \mathbf{h}) = \lambda|B|$, which in turn means that $\lambda(\mathbf{s}) = \lambda$. Therefore, a stationary Poisson process has to be homogeneous.
- Note also that there are other types of stationary processes. HPP is the one with both stationarity and conditional independence.

Examples of Poisson process



Second-order properties of point patterns

- Second-order intensity function is defined as

$$\lambda_2(\mathbf{s}_i, \mathbf{s}_j) = \lim_{|d\mathbf{s}_i| \to 0, |d\mathbf{s}_j| \to 0} \frac{E\{N(d\mathbf{s}_i)N(d\mathbf{s}_j)\}}{|d\mathbf{s}_i||d\mathbf{s}_j|}.$$

- A point process is stationary if $\lambda_2(\mathbf{s}_i, \mathbf{s}_j) = \tilde{\lambda}_2(\mathbf{s}_i - \mathbf{s}_j)$.
- Isotropy should be defined in the obvious way.

Estimation of the intensity function

- Suppose k is a kernel function.
- Kernel function is of a simpler shape to covariance functions. It is usually nonnegative and has largest mass in the center (origin). Examples of kernel functions are as follows:
  - Gaussian function
  - $k(x) = \mathbf{1}_{(|x| \leq h)}$
  - $k(x) = 0.75(1 - x^2)\mathbf{1}_{(|x| \leq 1)}$

Estimation of the intensity function (conti-)

- We use a kernel to estimate the intensity function in $\mathbb{R}$:

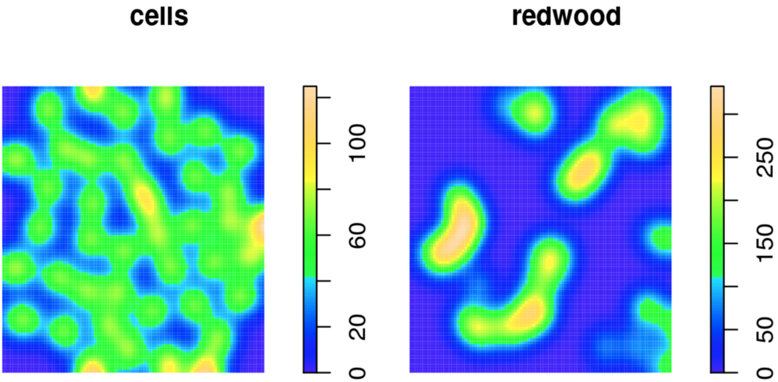
$$\hat{\lambda}(s_0) = \frac{1}{\nu(A)h} \sum_{i=1}^{n} k\Big(\Big|\frac{s_i - s_0}{h}\Big|\Big).$$

- In $\mathbb{R}^2$, we may do:

$$\hat{\lambda}(s_0) = \frac{1}{\nu(A)h_x h_y} \sum_{i=1}^{n} k\Big(\frac{x_i - x_0}{h_x}\Big) k\Big(\frac{y_i - y_0}{h_y}\Big).$$

Examples of estimated intensity functions

Use R package "spatstat"



Ripley's K-function

- The Ripley's K-function (dectects deviations from spatial homogeneity) of a stationary or isotropic process is defined as

$$K(h) = \frac{2\pi}{\lambda^2} \int_0^h x\lambda_2(x)dx.$$

  Here λ is the global intensity estimator $(\lambda(\mathbf{s}) = \lambda)$[1].

  - if the process is simple, $\lambda K(h)$ represents the expected number of extra events within the distance h from an arbitrary event.
  - If $K(h)$ is known, then we can derive λ_2 from it.

- You can determine whether points have a random, dispersed or clustered distribution pattern at a certain scale.

[1] The second-order methods considered here assume that marginal distributions of points have a fixed intensity, but that the joint distribution of all points is such that individual distributions of points are not independent.

Estimation of K- and L-functions

- The L-function is defined as $L(h) = \sqrt{K(h)/\pi}$.
- Note $\lambda K(h) = E(h)$ is the expected number of extra events within distance h.
- If h_{ij} is the distance between $\mathbf{s}_i$ and $\mathbf{s}_j$, then a naive moment estimator for $E(h)$ is $\tilde{E}(h) = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} \sum_{j \neq i}^{n} I(h_{ij} \leq h)$.

  Then we can estimate K-function by $\tilde{K}(h) = \hat{\lambda}^{-1} \tilde{E}(h)$. Usually this estimator is negatively biased.

Estimation of K- and L-functions (conti-)

- Ripley's suggests $\hat{E}(h) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n} w(\mathbf{s}_i, \mathbf{s}_j)^{-1} I(h_{ij} \leq h)$ where w is proportional to the circumference of a circle that is within the study region.
- You would compare your K-estimate to that of the complete spatial random process (πh^2).
- Estimator for L-function has better statistical properties.

(Optional) Exploratory data analysis: G and F functions

- Objective of EDA: Examine departure from HPP to see if more elaborate modeling is needed.
- For a random point pattern $\mathbf{S}$, we define the following two cdf's:
  1. G function: $G(d) = P(N(\mathbf{s}, d; \mathbf{S}) > 0)$ for $\mathbf{s} \in \mathbf{S}$, "nearest neighbor distribution"
  2. F function: $F(d) = P(N(\mathbf{s}, d; \mathbf{S}) > 0)$ for $\mathbf{s} \notin \mathbf{S}$, "empty space distribution"

  where $N(\mathbf{s}, d; \mathbf{S})$ is the number of points in $\mathbf{S}$ within a circle centered at $\mathbf{s}$ with radius d.
- Under HPP, $G(d) = F(d) = 1 - \exp(-\lambda\pi d^2)$, because the number of events in this circle follows a Poisson$(\lambda\pi d^2)$.

Estimating G and F

- For nearest neighbor distances $d_1, \ldots, d_n$ (i.e., distance to the nearest neighbors for $\mathbf{s}_1, \ldots, \mathbf{s}_n$):

$$\hat{G}(d) = \frac{\sum_i I(d_i \le d < b_i)}{\sum_i I(d < b_i)},$$

  where b_i is the distance from $\mathbf{s}_i$ to edge of D.

- Edge correction by accounting for the fact that the event $\{d_i < d\}$ is not observed if $d > b_i$.

- We can also compute $\hat{F}(d)$ with the same formula except that now we use m distances from randomly selected m points within D, which are not in $\mathbf{S}$. Often, $\hat{J}(d) = \frac{1-\hat{G}(d)}{1-\hat{F}(d)}$ (not sensitive to edge effects) is plotted. $J(d) > 1$ indicates dispersion and $J(d) < 1$ indicates clustering.

Another metric: K function (equivalent to pages 18-20)

- Another way to examine clustering/repulsion: The expected number of points within d of an arbitrary point.
- Under HPP, the K function is defined as

$$K(d) = \frac{1}{\lambda} E_{\mathbf{s}} \left(\sum_{\mathbf{s}_i \in \mathbf{S}, \mathbf{S} \subset D} N(\mathbf{s}_i, d; \mathbf{S}) \right).$$

- Note that the scaling $1/\lambda$ makes $K(d)$ free of λ. (Under HPP, $K(d) = E(N(\mathbf{s}, d; \mathbf{S})) = \lambda \pi d^2 / \lambda$).

Estimating K

- A customary estimate of $K(d)$ is

$$\hat{K}(d) = (\hat{\lambda})^{-1} \sum_i \sum_{j \neq i} \frac{1}{w_{ij}} 1(d_{ij} \leq d)/n,$$

  where $\hat{\lambda} = n/|D|$ and w_{ij} is the probability that an event is in D given its distance from $\mathbf{s}_i$ is exactly d_{ij}.

- Ripley's correction: $w_{ij} = \dfrac{\text{length}(c(\mathbf{s}_i, \|\mathbf{s}_i - \mathbf{s}_j\|) \cap D)}{2\pi \|\mathbf{s}_i - \mathbf{s}_j\|}$, where $c(u, r)$ is a circle centered at u with radius r.

- Often $L(d) = \sqrt{\dfrac{\hat{K}(d)}{\pi}} - d$ is plotted. $L(d) = 0$ for HPP, a peak at distance d suggests clustering at that distance.

Empirical estimates of intensity

- Note that G, F, and K rely on the notion of a 'typical point', i.e., all points are treated equally and hence stationarity is assumed.
- If the process is inhomogeneous, the following kernel estimate can be used to nonparametrically estiamte the spatially-varying intensity $\lambda(\mathbf{s})$:

  $$\hat{\lambda}(\mathbf{s}) = \sum_i h(\|\mathbf{s}_i - \mathbf{s}_j\|/\tau)/\tau^2, \ \mathbf{s} \in D,$$

  where h is a kernel function.
- An edge correction is often required to have a consistent estimate (dividing by $\int_D h(\|\mathbf{s} - \mathbf{s}_i\|/\tau) d\mathbf{s}$).

(Optional) Non-homogeneous Poisson process (NHPP)

- A NHPP is a Poisson process with a spatially varying intensity $\lambda(\mathbf{s})$. Often called as 'inhomogeneous Poisson process'
- The joint density of the total number points $N(D)$ and the locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$ is given by

$$\begin{aligned} f(\mathbf{s}_1, \ldots, \mathbf{s}_n, N(D) = n) &= f(\mathbf{s}_1, \ldots, \mathbf{s}_n | N(D) = n) P(N(D) = n) \\ &= \prod_i \frac{\lambda(\mathbf{s}_i)}{\lambda(D)^n} \times \lambda(D)^n \frac{\exp(-\lambda(D))}{n!} \end{aligned}$$

- Therefore, the likelihood function is given by

$$L(\lambda(\mathbf{s}), \mathbf{s} \in D; \mathbf{s}_1, \ldots, \mathbf{s}_n) = \prod_i \lambda(\mathbf{s}_i) \exp(-\lambda(D)).$$

Linear model for intensity function

- Note that the likelihood function depends on the function $\lambda(\mathbf{s})$ itself. We need a parametric model for $\lambda(\mathbf{s})$ to avoid having an uncountable dimensional model.
- One solution: Set $\log \lambda(\mathbf{s}) = \mathbf{X}^\top(\mathbf{s})\boldsymbol{\beta} + w(\mathbf{s})$, where $\mathbf{X}(\mathbf{s})$ contains covariates and $w(\mathbf{s})$ is a spatial process.
- Still need to evaluate $\lambda(D) = \int_D \exp(\mathbf{X}^\top(\mathbf{s})\boldsymbol{\beta} + w(\mathbf{s}))d\mathbf{s}$.
- Assume $\mathbf{X}^\top(\mathbf{s})$ and $w(\mathbf{w})$ are constant in each 'tile' B_m for $m = 1, \ldots, M$ where $\cup_{m=1} B_m = D$, to have

$$\int_D \lambda(\mathbf{s})d\mathbf{s} = \sum_{m=1}^{M} \exp(\mathbf{X}^\top(B_m)\boldsymbol{\beta} + \phi_m).$$

- The spatial effect ϕ_m can be modeled by GMRF or GP.

(Optional) Modeling interactions

- Poisson processes (both homogeneous and non-homogenous) assume conditional independence and do not have any 'interaction' between points.
- The Papangelou conditional intensity for a point process: $\lambda(\mathbf{s}, \mathbf{S}) = \frac{f(\mathbf{S})}{f(\mathbf{S} \backslash \{\mathbf{s}\})}$
- Homogeneous Poisson process: $\lambda(\mathbf{s}, \mathbf{S}) = \lambda$.
- Non-homogeneous Poisson process: $\lambda(\mathbf{s}, \mathbf{S}) = \lambda(\mathbf{s})$.
- Strauss process: $\lambda(\mathbf{s}, \mathbf{S}) = \lambda\gamma^{N(\mathbf{s}, d; \mathbf{S})}$, where $N(\mathbf{s}, d; \mathbf{S})$ is the number of points in $\mathbf{S} \backslash \{\mathbf{s}\}$ within a circle centered at $\mathbf{s}$ with raidus d. $0 < \gamma < 1$ means inhibition and $\gamma = 1$ means no interaction (note that γ cannot be greater than 1).

Fitting Strauss process using Pseudo-likelihood

- The original likelihood function for Strauss process contains an intractable norming constant.
- We use the following pseudo likelihood (Besag, 1977, Baddeley and Turner, 2000):

$$PL(\lambda, \gamma; \mathbf{S}) = \lambda^{N(\mathbf{S})} \gamma^{2a(\mathbf{S})} \exp\Big(- \lambda \int_D \gamma^{N(\mathbf{s},d;\mathbf{S})} d\mathbf{s} \Big)$$

  where $a(\mathbf{S}) = \#\{(i,j) | i < j, \|\mathbf{s}_i - \mathbf{s}_j\| \leq d\}$.

- Note that the integral $\int_D \gamma^{N(\mathbf{s},d,\mathbf{S})} d\mathbf{s} = \alpha_0 + \alpha_1 \gamma + \cdots + \alpha_K \gamma^K$, where $\alpha_k = |A_k|$ with $A_k = \{\mathbf{s} \in D | N(\mathbf{s}, d; \mathbf{S}) = k)\}$.

Generating point patterns

- For HPP: Determine $N(D) = n$ by sampling $N(D) \sim \text{Poisson}(\lambda(D))$, and then generate n points from the uniform distribution on D.
- For NHHP:
  1. Find $\lambda_{\max} = \max_{\mathbf{s} \in D} \lambda(\mathbf{s})$.
  2. Generate $n = N(D) \sim \lambda_{\max}|D|$.
  3. Sample $\mathbf{s}_1, \ldots, \mathbf{s}_n$ uniformly from D.
  4. For each $\mathbf{s}_i$, keep $\mathbf{s}_i$ with probability $\lambda(\mathbf{s}_i)/\lambda_{\max}$.
- This process is often called 'thinning'.
- If $\lambda(\mathbf{s})$ is random (e.g., $\log \lambda(\mathbf{s}) = \mathbf{X}^\top(\mathbf{s})\boldsymbol{\beta} + w(\mathbf{s})$), $\lambda(\mathbf{s})$ has to be generated first.

Reference

- Cressie, N. Statistics for Spatial Data. Wiley. Chapter 1.
- Banerjee, S., Carlin, B., and Gelfand, A. Hierarchical Modeling and Analysis for Spatial Data (2nd). CRC Press.
- Jun, M., Genton, M. G., Chang, W., and Jeong, J. Lecture Notes for Spatial Statistics. UH, KAUST, UC, and HYU.