Part 2: Point Processes

This week:
- Examples
- Definitions + 3.4.1
- More examples and 3.4.2 Spatial statistical dependency

Next week:
- Simulations + more models (3.4.3 + revisit ++).
- Estimation

Extra text book: Statistical Analysis and Modelling of Spatial Point Patterns by Janine Illian, Antti Penttinen, Helga Stoyan, Dietrich Stoyan (2008)

Where and when

- Tuesdays: room 734 Central Building 2 (February 14, **NOT March 7 and 14, April 6**)
- Thursdays: room 743, Central Building 2 (February 9, **NOT 16**, but March 9, 16, April 6)
- Friday: room 656, February 17, **March 10 and 17**

Spatial Point Processe (SPP)

From Cressie and Wikle:

4.3 SPATIAL POINT PROCESSES

A spatial point process is a stochastic process governing the location of events (equivalently, points) $\{\mathbf{s}_i\}$ in some set $D_s \subset \mathbb{R}^d$, where the number of such events in D_s is also random (e.g., Diggle, 2003). In the simplest case, the

and we only consider:

Only *simple* spatial point processes in $\mathbb{R}^d$ (i.e., almost surely, either no event or a single event occurs at any point) will be considered. We characterize the

Homogenous Poisson point process (HPPP)

An important, but simple, example is the *Poisson point process* Z, for which

$$Z(A)|\lambda^o \sim Poi(\lambda^o|A|), \qquad A \subset D_s, \qquad (4.159)$$

where $\lambda^o > 0$ is a parameter of the Poisson point process, and recall that $|A|$ is the d-dimensional volume of A. More details on this spatial point process are given in Section 4.3.1.

Sampling from Poisson point process

Trees in $1km \times 1km$ domain, with intensity $\lambda^0 = 15 trees/km^2$

- Sample $m \sim Pos(15)$
- for $i = 1 : m$
    - sample location randomly in domain $\{s_i\}$
- end

First order intensity

ds denote a small region located at $\mathbf{s}$ with volume $|ds|$. Then the *first-order intensity* function of the point process $Z(\cdot)$ is defined as

$$\lambda(\mathbf{s}) \equiv \lim_{|d\mathbf{s}| \to 0} E(Z(d\mathbf{s}))/|d\mathbf{s}|, \qquad \mathbf{s} \in D_s, \tag{4.160}$$

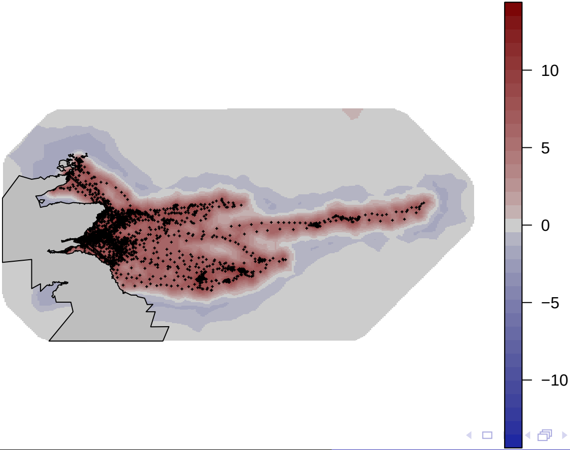
provided the limit exists. Hence,

$$E(Z(A)) = \int_A \lambda(\mathbf{s})\, d\mathbf{s}, \qquad A \subset D_s. \tag{4.161}$$

Estimation of intensity

We often want to estimate the $\lambda(s)$ from data (points).

Seals (from log-Gaussian Cox model):



Kernel estimation of intensity

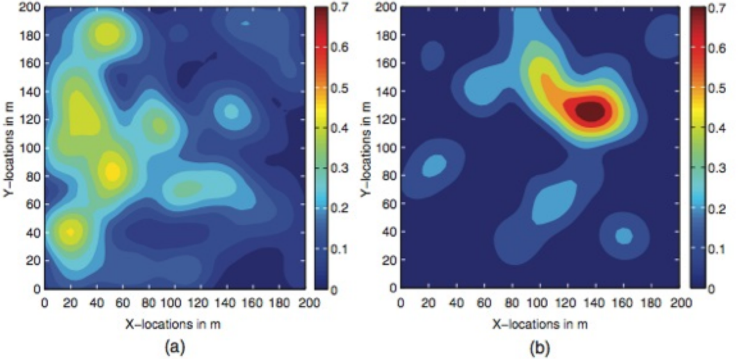


Figure 4.16 **(a)** Contour plot of estimated first-order intensity of $m_A = 271$ adult longleaf pine trees in the 4-ha study area in the Wade Tract; bandwidth $b = 30$ m. **(b)** Contour plot of estimated first-order intensity of $m_S = 159$ subadult longleaf pine trees in the same study area as **(a)**; bandwidth $b = 30$ m.

Dependency

from Illian et al 2008:

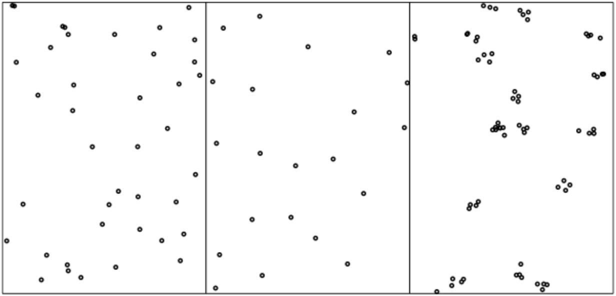


Figure 1.1 Three simulated point patterns: (left) random, (centre) regular, (right) clustered.

2nd order intensity and Pair-correlation function

2nd order intensity

$$\lambda_2(s,x) = \lim_{|ds|\to 0, |dx|\to 0} \frac{E(Z(ds)Z(dx))}{|dx||dy|}$$

Pair correlation function

$$g(s,x) = \frac{\lambda_2(x,s)}{\lambda(s)\lambda(x)}$$

For HPPP: $\lambda_2(s,x) = \lambda(s)\lambda(x)$ and $g(s,x) = 1$

Stationarity (from f3.pdf)

Second order (weakly) stationary

The random field $Y(s)$ for $s \in D$ is second order stationary if

- $E(Y(s)) = \mu$ for all $s \in D$
- $Cov(Y(s+h), Y(s)) = C_Y(h)$ for all $s, s+h \in D$
- I.e. the covariance only depends on the vector difference between the locations.

Strong stationarity

Let $F(Y(s_1), Y(s_2), \ldots, Y(s_n))$ be the cdf of $Y(s_1), Y(s_2), \ldots, Y(s_n)$. The random field $Y(s)$ is strongly stationary if $F(Y(s_1), Y(s_2), \ldots, Y(s_n)) = F(Y(s_1+h), Y(s_2+h), \ldots, Y(s_n+h))$ for all s_i and $s_i + h \in D$

- General: Strong stationarity $\Rightarrow$ second order stationarity.
- For GRF: Strong stationarity $\Leftrightarrow$ second order stationarity

Isotropy (from f3.pdf)

A subclass of weakly stationary covariance functions are:

Isotropic covaraince function

A covariance function is isotropic if it only depends on the distance between the locations:

$$C_Y(s, s+h) = C_Y(\|h\|)$$

- Not isotropic = anisotropic
- The exponential covariance function is isotropic.

Example: Gold particles

From Illian et al (2008)

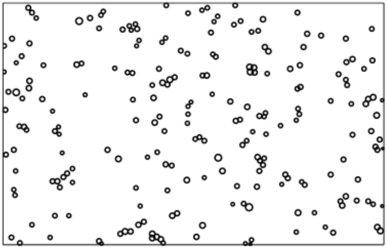


Figure 1.3 Ultrathin section of a pellet of purified tobacco rattle virus after immunogold labelling with a goat antirabbit gold (size 15 nm) probe in a rectangular window of size 1064.7×676 nm. The 218 gold particles are identifiable as dark spots in the electron-microscopic image. The diameters of the small circles are proportional to the gold particle diameters. Data courtesy of C. Glasbey.

Pair-correlation function for gold particles

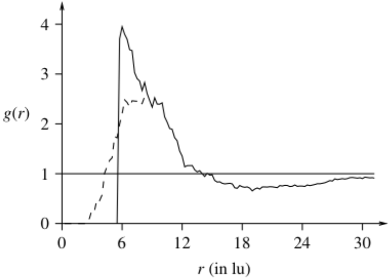


Figure 4.20 The empirical pair correlation function of the pattern of gold particles, obtained with the estimator (4.3.38) and bandwidths $h = 3$ lu for $r \leq 20$ lu and 6 lu for $r > 20$ lu and improved with the reflection method. The dashed line shows the result without this correction. A comparison with Figure 4.18 reveals the advantages of using $g(r)$ as opposed to $L(r)$ as an instructive summary characteristic.

K-function (left) subadult longleaf pines

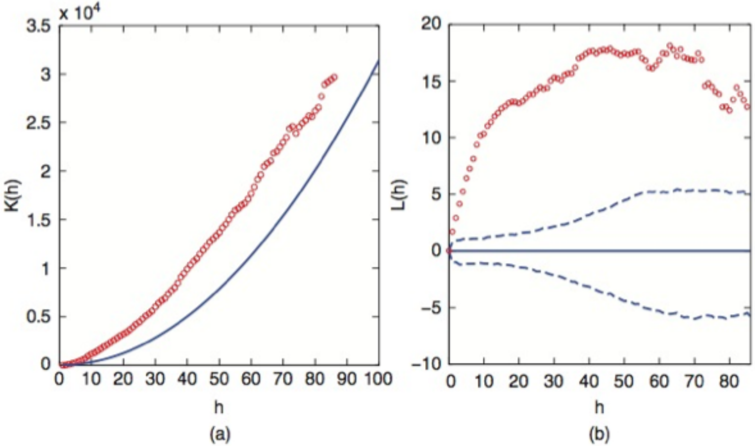


Figure 4.17 (**a**) Estimated K function for the $m_S = 159$ subadult longleaf pine trees (red circles) in the 4-ha study area in the Wade Tract; the theoretical K function for CSR is superimposed (solid blue line). (**b**) Estimated L function for the $m_S = 159$ subadult longleaf pine trees (red circles) obtained from (**a**); the 95% pointwise confidence limits for L values based on 1000 CSR realizations (dashed blue lines) are superimposed.

L-function (left) adult longleaf pine

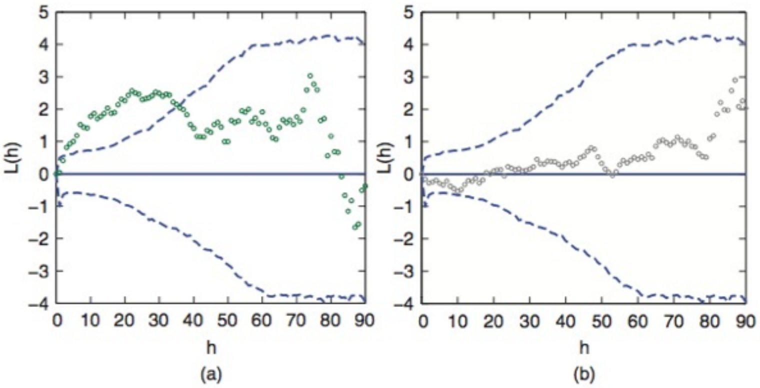


Figure 4.18 (**a**) Estimated L function for the $m_A = 271$ adult longleaf pine trees (green circles) in the 4-ha study area in the Wade Tract; the 95% confidence limits for L values based on 1000 CSR realizations are superimposed (dashed blue lines). (**b**) Estimated L function for $m_A = 271$ locations generated from a single realization of a CSR process (black circles); the same 95% confidence limits for L values, as shown in (**a**), are superimposed (dashed blue lines).

L-function gold particles

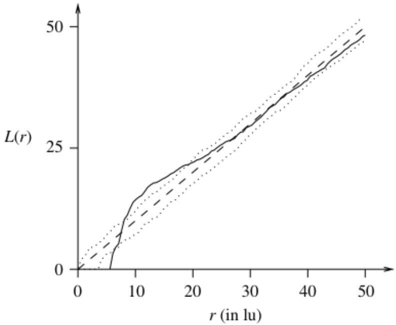


Figure 4.18 The empirical L-function (solid line) for the pattern of gold particles, minimum and maximum envelopes from 99 simulations of a Poisson process of intensity 0.000 865 in a 630×400 rectangle (dotted lines) and the theoretical L-function of a Poisson process (dashed line). The curves indicate micro-scale repulsion and meso-scale clustering.

Data model: $[Z|Y,\theta]$

Process model: $[Y|\theta]$

Parameter model: $[\theta]$

Data model: Conditional on $\lambda(\cdot)$, the point process $Z(\cdot)$ is an inhomogeneous Poisson point process with intensity $\lambda(\cdot)$.

Process model: Conditional on $\boldsymbol{\beta}$ and $C_Y(\cdot,\cdot)$, $Y(\cdot) \equiv \log\lambda(\cdot)$ is a Gaussian process with the following properties:

$$E(Y(\mathbf{s})) = \mathbf{x}(\mathbf{s})'\boldsymbol{\beta} \quad \text{and} \quad C_Y(\mathbf{s},\mathbf{x}) \equiv \operatorname{cov}(Y(\mathbf{s}), Y(\mathbf{x})).$$