Spatial point processes: introduction

Aila Särkkä

Mathematical sciences
Chalmers University of Technology and University of Gothenburg
Gothenburg, Sweden
What is spatial statistics?

Locations of observations (in $\geq 2$ dimensions) play an important role. Observations are given in the form

$$[x, m(x)],$$

where

$x$ is a location

$m(x)$ is an observation at $x$, value of some variable
Some examples

- $x$ is the location of a cell and $m(x)$ the size or state of health
- $x$ is the location of a picture element and $m(x)$ is the colour
- $x$ is the location of a tree in a forest and $m(x)$ is the height/diameter/species of the tree
- $x$ is the location of a weather station and $m(x)$ is the mean daily temperature
1) **Where** do we observe something? \( \rightarrow x \) interesting

2) **What** do we observe somewhere? \( \rightarrow m(x) \) interesting

3) **Where** and **what** do we observe? \( \rightarrow \) both \( x \) and \( m(x) \) interesting
Some remarks

- There is no natural ordering in higher (> 1) dimensions.
- Spatial dependencies between the observations
  - make statistical analysis difficult
  - are the basis for spatial statistics
Some examples of data structures

- tessellation
- fibre data
- lattice data
- point pattern
Tessellation: materials science

Spatial point processes: introduction
Fibre pattern: epidermal nerve fibers

3151 (normal), skeletonized image
Lattice data: poverty level in the different counties in Illinois, Indiana, Michigan, Ohio, and Wisconsin
Point pattern: crime incidents in London
A point process $N$ is a stochastic mechanism or rule to produce point patterns or realisations according to the distribution of the process.

A marked point process is a point process where each point $x_i$ of the process is assigned a quantity $m(x_i)$, called a mark. Often, marks are integers or real numbers but more general marks can also be considered.

Diggle (2013)
Illian et al. (2008)

R library spatstat (Baddeley and Turner, 2005)
Remark 1: We assume that all point processes are simple, i.e. that there are no multiple points ($x_i \neq x_j$ if $i \neq j$).

Remark 2: There is a large literature on point processes in time. There is an overlap of methods for point processes in space and in time but the temporal case is not only a special case of the spatial process with $d = 1$. Time is 1-directional.

Remark 3: To avoid confusion between points of the process and points of $\mathbb{R}^d$, the points of the process or point pattern (realization) are called events.
Spatial point patterns

- clustered
- completely random
- regular
Examples

- Locations of betacells within a rectangular region in a cat’s eye (regular)
- Locations of Finnish pine saplings (clustered)
- Locations of Spanish towns (regular)
- Locations of galaxes (clustered)

Remark: Very different scales, from microscopic to cosmic
Finnish pine saplings: locations and diameters

Beta-type retina cells in the retina of a cat: locations and type (red triangles "on", blue circles "off")
(Compare to the mean of a real-valued random variable.)

The mean number of points of $N$ in $B$ is $\mathbb{E}(N(B))$ (depends on the set $B$). We use the notation

$$\Lambda(B) = \mathbb{E}(N(B))$$

and call $\Lambda$ the intensity measure.

Under some continuity conditions, a density function $\lambda$, called the intensity function, exists, and

$$\Lambda(B) = \int_B \lambda(x) \, dx.$$
Some properties of point processes: stationarity and isotropy

- A point process $N$ is stationary (translation invariant) if $N$ and the translated point process $N_x$ have the same distribution for all translations $x$, i.e.

$$N = \{x_1, x_2, \ldots\} \text{ and } N_x = \{x_1 + x, x_2 + x, \ldots\}$$

have the same distribution for all $x \in \mathbb{R}^d$.

- A point process is isotropic (rotation invariant) if its characteristics are invariant under rotations, i.e.

$$N = \{x_1, x_2, \ldots\} \text{ and } rN_x = \{rx_1, rx_2, \ldots\}$$

have the same distribution for any rotation $r$ around the origin.

- If a point process is both stationary and isotropic, it is called motion-invariant.
If $N$ is stationary, then

$$\Lambda(B) = \lambda |B|,$$

where $0 < \lambda < \infty$ is called the intensity of $N$ and $|B|$ is the area/volume of $B$.

$\lambda$ is the mean number of points of $N$ per unit area/volume, i.e.

$$\lambda = \frac{\Lambda(B)}{|B|} = \frac{\mathbb{E}(N(B))}{|B|}.$$
Let $D_1$ be the distance from an arbitrary event to the nearest other event.

The nearest neighbour distance function is defined as

$$G(r) = P(D_1 \leq r)$$

If the pattern is completely spatially random (CSR),

$$G(r) = 1 - \exp(-\lambda \pi r^2)$$

For regular patterns $G(r)$ tends to lie below and for clustered patterns above the CSR curve.
Let $D_2$ denote the distance from an arbitrary point to the nearest event.

The empty space function can be defined as

$$F(r) = P(D_2 \leq r)$$

If the pattern is completely spatially random,

$$F(r) = 1 - \exp(-\lambda \pi r^2)$$

For regular patterns $F(r)$ tends to lie above and for clustered patterns below the CSR curve.
Summary: combination of $G$ and $F$

- Using $G$ and $F$ we can define the so-called $J$ function as

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

(whenever $1 - F(r) > 0$)

- If the pattern is completely spatially random,

$$J(r) \equiv 1$$

- For regular patterns $J(r) > 1$ and for clustered patterns $J(r) < 1$. 
Ripley's $K$ function

- The 2nd order properties (compare to variance) of a stationary and isotropic point process can be characterized by Ripley's $K$ function (Ripley, 1977)

$$\lambda K(r) = \mathbb{E}[\# \text{ further events within distance } r \text{ of an arbitrary event}].$$

- If the pattern is completely spatially random,

$$K(r) = \pi r^2.$$  

- For regular patterns $K(r) < \pi r^2$ and for clustered patterns $K(r) > \pi r^2$. 

Aila Särkkä

Spatial point processes: introduction
Often, a variance stabilizing and centered version of the $K$ function (Besag, 1977) is used, namely (in 2D)

$$L(r) - r = \sqrt{K(r)/\pi} - r,$$

which equals 0 under CSR. Values less than zero indicate regularity and values larger than zero clustering.
Note on estimation of $G$, $F$, $J$ and $K$

- Typically, a point pattern is observed in a (bounded) observation window and points outside the window are not observed.
- Estimators (except for $J(r)$) need to be edge-corrected.
- Edge correction methods include plus sampling, minus sampling, Ripley’s isotropic correction and translation (stationary) correction.
A point process is a homogeneous Poisson process (CSR) if
(P1) for some $\lambda > 0$ and any finite region $B$, $N(B)$ has a Poisson distribution with mean $\lambda |B|$
(P2) given $N(B) = n$, the events in $B$ form an independent random sample from the uniform distribution on $B$

Inhomogeneous Poisson process: intensity $\lambda$ (in homogeneous Poisson process) replaced by an intensity function $\lambda(x)$
Realization of a homogeneous Poisson process

Poisson process with intensity 100
Summary statistics

\textbf{allstats(simPoisson)}

\begin{align*}
F \text{ function} & : \hat{F}_{km}(r), \hat{F}_{bord}(r), \hat{F}_{cs}(r), \hat{F}_{pois}(r) \\
G \text{ function} & : \hat{G}_{km}(r), \hat{G}_{bord}(r), \hat{G}_{han}(r), \hat{G}_{pois}(r) \\
J \text{ function} & : \hat{J}_{km}(r), \hat{J}_{han}(r), \hat{J}_{cs}(r), \hat{J}_{pois}(r) \\
K \text{ function} & : \hat{K}_{iso}(r), \hat{K}_{trans}(r), \hat{K}_{bord}(r), \hat{K}_{pois}(r)
\end{align*}
Cluster processes are models for aggregated spatial point patterns

For Neyman-Scott cluster process

1. **NS1** parent events form a Poisson process with intensity $\lambda$
2. **NS2** each parent produces a random number $S$ of daughters (offsprings), realized independently and identically for each parent according to some probability distribution
3. **NS3** the locations of the daughters relative to their parent are independently and identically distributed according to some bivariate distribution

The cluster process consists only of the daughter points.
Examples of Neyman-Scott processes

Matérn cluster process: The points in a cluster are independently and uniformly scattered in a disc (ball) of radius $R$ centered at the parent point.

Thomas process: The distribution of the daughter points around the parent point is a radially symmetric normal distribution with variance $\sigma^2$. 
Realization of a Matérn cluster process

**Left:** Parent point intensity 20, cluster radius 0.05, average number of daughter points per cluster 5

**Right:** Poisson process with intensity 100
Summary statistics

allstats(simClust)

F function

G function

J function

K function

Spatial point processes: introduction
Matérn hard-core processes

- Hard-core processes are models for regular spatial point patterns
- There is a minimum allowed distance, called hard-core distance $\delta$, between any two points
- **Matérn I hard-core process**: A Poisson process with intensity $\lambda$ is thinned by delating all pairs of points that are at distance less than the hard-core radius apart.
- **Matérn II hard-core process**: Events of a Poisson process are given "marks" from (e.g.) the uniform $U(0, 1)$ distribution. An event $x$ with mark $m$ is retained if $b(x, \delta)$ contains no points of the process that has a mark smaller than $m$. 
Realization of a Matérn I hard-core process

**Left:** Hard-core process with the initial Poisson intensity 300, hard-core radius 0.04

**Right:** Poisson process with intensity 100
Summary statistics

allstats(simHC)

F function

$\hat{F}_{\text{km}}(r)$
$\hat{F}_{\text{bord}}(r)$
$\hat{F}_{\text{cs}}(r)$
$\hat{F}_{\text{pois}}(r)$

G function

$\hat{G}_{\text{km}}(r)$
$\hat{G}_{\text{bord}}(r)$
$\hat{G}_{\text{han}}(r)$
$\hat{G}_{\text{pois}}(r)$

J function

$\hat{J}_{\text{km}}(r)$
$\hat{J}_{\text{han}}(r)$
$\hat{J}_{\text{cs}}(r)$
$\hat{J}_{\text{pois}}(r)$

K function

$\hat{K}_{\text{iso}}(r)$
$\hat{K}_{\text{trans}}(r)$
$\hat{K}_{\text{bord}}(r)$
$\hat{K}_{\text{pois}}(r)$
You have gotten a short introduction to the analysis of homogeneous (stationary) and isotropic spatial point patterns.

- Hard to check stationarity/isotropy.

- If stationarity/isotropy cannot be assumed, the summary statistics and models have to be modified to the new conditions.


