Functional summary statistics for Poisson processes on convex shapes in three dimensions

Scott Ward, Edward Cohen, Niall Adams

Imperial College London

2019

Outline

1. Motivation
2. Background
3. Statement of Problem
4. Methodology
5. Examples
6. Further Work
7. Astrophysics Applications

My Motivation

- Pseudomonas aeruginosa is a virulent and opportunistic bacteria
- A major cause of hospital borne infections and a serious danger to patients undergoing immunosuppressive therapy
- P. aeruginosa virulence is, in part, attributed to its type 6 secretion system (T6SS)
- The T6SS originates on the cell membrane of P. aeruginosa and we want to understand its spatial distribution with respect to its shape.
- Is the point pattern completely spatially random (CSR) or is there preferential placement where the T6SS builds and activates?

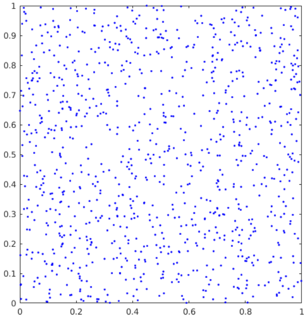
Spatial point pattern analysis in $\mathbb{R}^2$

- A spatial point process is the stochastic mechanism that gives rise to a point pattern
- Intensity measure and function of a process,

$$\mu(B) \equiv \mathbb{E}[N(B)] = \int_B \rho(\mathbf{x}) d\mathbf{x}.$$

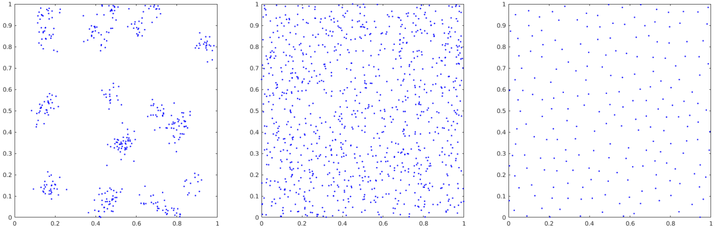
- A process is homogeneous if ρ is constant.
- A process is stationary if it is distributionally invariant to translations.

Complete spatial randomness on $\mathbb{R}^2$



- Poisson process with constant intensity function, ρ.
- Poisson number of points uniformly and independently distributed.

Typical point patterns



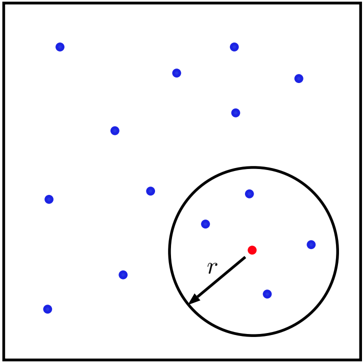
(a) Cluster (b) Complete Spatial Randomness (c) Regular

Scott Ward, Edward Cohen, Niall Adams

Summary statistics for planar and spatial data

- Determining whether a point pattern exhibits CSR is normally the first step.
- This is commonly achieved using functional summary statistics, such as nearest neighbour function.
- Based on these deviations we can suggest whether a pattern is more regular or clustered.

Ripley's K-Function



- A popular functional summary statistic is Ripley's K-function,

$$K(r) = \frac{1}{\rho}\mathbb{E}[N_{\mathbf{0}}(B(\mathbf{0}, r))],$$

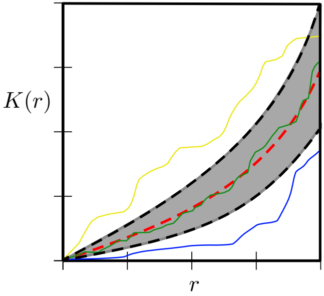
- For a homogeneous Poisson process $K(r) = \pi r^2$.

- Estimator is,

$$\hat{K}(r) = \frac{\text{Area}^2(B)}{N(N-1)} \sum_{\mathbf{x},\mathbf{y} \in X \cap B}^{\neq} \mathbb{1}[d(\mathbf{x}, \mathbf{y}) \leq r].$$

Interpretation of Ripley's K-function

- CSR: $\hat{K}(r) \approx \pi r^2$
- Cluster: $\hat{K}(r) > \pi r^2$
- Regular: $\hat{K}(r) < \pi r^2$



Inhomogeneous K-Function

- Ripley's K-function can be extended to a class of inhomogeneous processes [1].
- For this class of inhomogeneous processes the K-function is,

$$K_{\text{inhom}}(r) = \mathbb{E} \sum_{\mathbf{x} \in X^{!}_{\mathbf{y}}} \frac{\mathbb{1}[\mathbf{x} \in B(\mathbf{y}, r)]}{\rho(\mathbf{x})},$$

- Estimators take the form,

$$\hat{K}_{\text{inhom}}(r) = \frac{1}{\text{Area}(B)} \sum_{\mathbf{x},\mathbf{y} \in X \cap B}^{\neq} \frac{\mathbb{1}[\mathbf{x} \in B(\mathbf{y}, r)]}{\hat{\rho}(\mathbf{x})\hat{\rho}(\mathbf{y})},$$

Statement of Problem

Let $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$*, with* $N = |X|$*, be a spatial point pattern on a convex space in* $\mathbb{R}^3$*,* $\mathbb{D}$*. Then we wish to determine if,*

$$H_0 : X \text{ is CSR on } \mathbb{D} \qquad H_1 : X \text{ is not CSR on } \mathbb{D}$$

Point processes on the unit Sphere

- The metric on the unit sphere is the great circle length,

$$d(\mathbf{x}, \mathbf{y}) = \cos^{-1}(\mathbf{x} \cdot \mathbf{y}).$$

- A point process is said to be isotropic on the sphere if it's distribution is invariant under rotations [2].
- Poisson processes with intensity function, $\rho : \mathbb{S}^2 \mapsto \mathbb{R}$, on a sphere is defined
  1. Poisson$(\mu(\mathbb{S}^2))$ number of points
  2. Given the number of points, these points are independently distributed with density proportional to $\rho(\mathbf{x})$ on $\mathbb{S}^2$.

K-Function on sphere for Poisson processes

- K-function for Poisson process is,

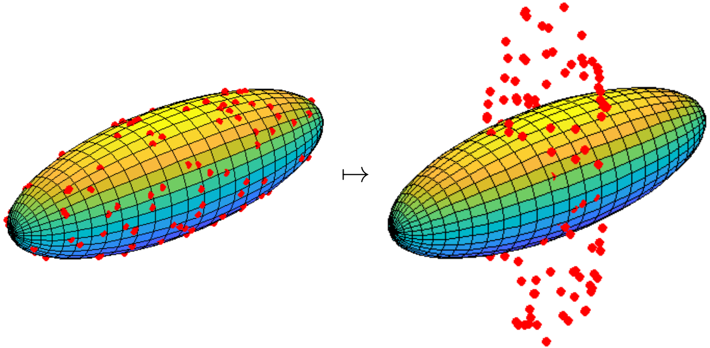
$$K(r) = K_{\text{inhom}}(r) = 2\pi(1 - \cos r).$$

- Non-parametric estimator of $K_{\text{inhom}}(r)$ is,

$$\hat{K}_{\text{inhom}}(r) = \frac{1}{4\pi} \sum_{\mathbf{x},\mathbf{y} \in X}^{\neq} \frac{\mathbb{1}[d(\mathbf{x},\mathbf{y}) \leq r]}{\rho(\mathbf{x})\rho(\mathbf{y})}, \tag{1}$$

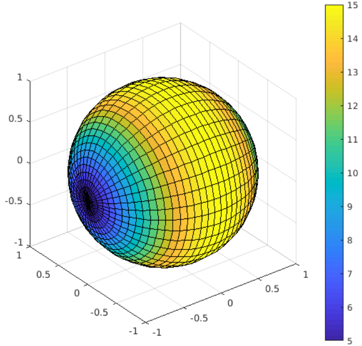
Main challenge for point pattern analysis on convex shapes

- K-functions rely on definitions of stationarity/isotropy.
- For a general convex shape these definitions do not extend.



Mapping homogeneous Poisson processes from an ellipsoid to the unit sphere

- Mapping Theorem [3]: A Poisson process is invariant under transformations between metric spaces.
- Use $(x, y, z) \mapsto (x/a, y/b, z/c)$
- The intensity function becomes



$$\rho^*(\mathbf{x}) = \rho ab \left[1 - \left(1 - \frac{c^2}{a^2}\right)\mathbf{x}_1^2 - \left(1 - \frac{c^2}{b^2}\right)\mathbf{x}_2^2\right]^{\frac{1}{2}}.$$

Constructing K-functions for Poisson processes on ellipsoids

- We can construct $\hat{K}_{\text{inhom}}(r)$ as,

$$\hat{K}_{\text{inhom}}(r) = \frac{1}{4\pi(\rho ab)^2} \sum_{\mathbf{x}\in X} \sum_{\mathbf{y}\in X\setminus\{x\}} \frac{\mathbb{1}[d(\mathbf{x},\mathbf{y})]}{\tilde{\rho}(\mathbf{x})\tilde{\rho}(\mathbf{y})},$$

where $\tilde{\rho}(\mathbf{x}) = \left[1 - \left(1 - \frac{c^2}{a^2}\right)\mathbf{x}_1^2 - \left(1 - \frac{c^2}{b^2}\right)\mathbf{x}_2^2\right]^{\frac{1}{2}}$.

- Use the following unbiased estimator for ρ^2,

$$\hat{\rho}^2 = \frac{N(N-1)}{(\text{Area of ellipsoid})^2}$$

Test statistic for complete spatial randomness

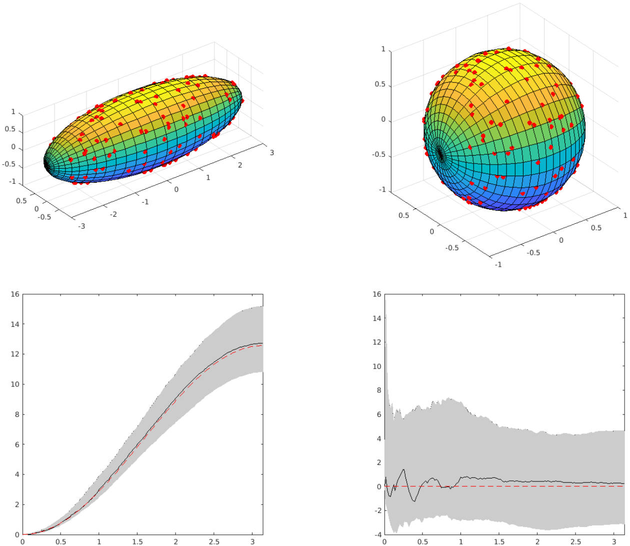
- In $\mathbb{R}^2$, the L-function is used to determine CSR,

$$L(r) = \left(\frac{K(r)}{\pi}\right)^{\frac{1}{2}} = r,$$

  as it is variance stabilised.
- We derive $\mathrm{Var}(\hat{K}_{\mathrm{inhom}}(r))$ and standardise $\hat{K}_{\mathrm{inhom}}(r)$ for each r to stabilise variance [4].
- We suggest using the following test statistic,

$$T = \sup_{r \in [0,\pi]} \left| \frac{\hat{K}_{\mathrm{inhom}}(r) - 2\pi(1 - \cos r)}{\sqrt{\mathrm{Var}(\hat{K}_{\mathrm{inhom}}(r))}} \right|$$

Example: Homogeneous Poisson process

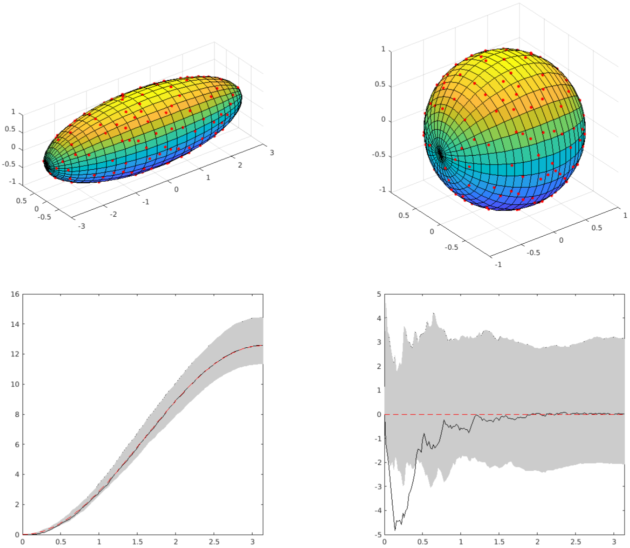


Example: Regular process

We simulate the Matèrn 2 process on ellipsoids.

1. Simulate a Poisson process with constant intensity function ρ.
2. For each point, simulate a *mark*, from a mark distribution independently of the locations of all the points and each other
3. For any two points within a distance R of each other remove the one with the smaller mark.

Example: Regular process

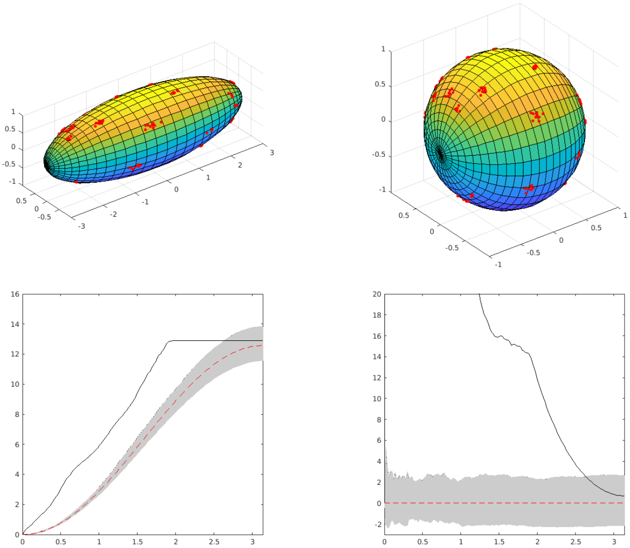


Example: Cluster process

We simulate a *Thomas* type ellipsoid process.

1. Simulate a parent Poisson process with constant intensity ρ.
2. For each parent draw N Poisson random variable offspring.
3. Independently distribute each offspring with a von Mises Fisher distribution centred at the parent point.

Example: Cluster process



Scott Ward, Edward Cohen, Niall Adams

Interpretation of K-function ECF

- From the previous plots we see that,
  - Regular processes: $K(r)$ is below the lower simulation envelope for small values of r
  - Cluster processes: $K(r)$ is larger than the upper simulation envelope for many values of r
- This coincides with the interpretation of $K(r)$ for planar and spatial data.

H-, F-, and J-function

- The K-function only captures part of the information within an observed spatial point pattern.
- Other functional summary statistics include,

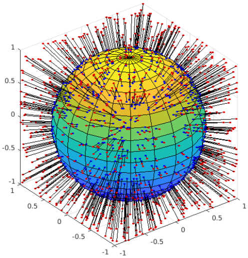
$$F(r) = 1 - P_{X \cap B(\mathbf{0},r)}(\emptyset),$$
$$H(r) = 1 - P_{X_{\mathbf{x}}^{!} \cap B(\mathbf{0},r)}(\emptyset),$$
$$J(r) = \frac{1 - H(r)}{1 - F(r)},$$

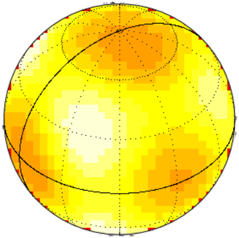
General convex shapes

- We can use,

$$f(\mathbf{x}) = \frac{\mathbf{x}}{\sqrt{\mathbf{x}_1^2 + \mathbf{x}_2^2 + \mathbf{x}_3^2}}, \qquad (2)$$

which maps each point of the space uniquely to the sphere.





Potential astrophysics applications

- This methodology is not restricted to convex shapes within $\mathbb{R}^3$ but can be applied to convex shapes in any dimension.
- In particular it could be extended for point patterns on ellipses.
- Consider an event on a planet. Can we determine if these events occur more frequently when closer to one of the foci of it's orbit?
- For example do natural disasters occur more when the Earth is closer to the Sun?
- Are the other potential applications?

Acknowledgements

- Supervisors: Edward Cohen & Niall Adams
- Wellcome Trust

References I

A J Baddeley, J. Møller, and R Waagepetersen.
Non- and semi-parametric estimation of interaction in inhomogeneous point patterns.
Statistica Neerlandica, 54(3):329–350, 2000.

Jesper Møller and Ege Rubak.
Functional summary statistics for point processes on the sphere with an application to determinantal point processes.
Spatial Statistics, 18(Section 2):4–23, 2016.

J F C Kingman.
Poisson Processes.
Oxford University Press, 1993.

References II

Thibault Lagache, Gabriel Lang, Nathalie Sauvonnet, and Jean Christophe Olivo-Marin.
Analysis of the spatial organization of molecules with robust statistics.
PLoS ONE, 8(12):1–7, 2013.