Spatial point processes: repetitions, inhomogeneity, anisotropy, 2D→3D

Replicated patterns

Example (Diggle *et al.*, 1991): Pyramidal neuron positions in three groups, normal, schizoaffective and schizophrenic, are of interest.

Data: Pyramidal neuron positions measured from 12 normal, 9 schizoaffective and 10 schizophrenic subjects.

Questions of interest

- What is the spatial structure of pyramidal neuron positions in each group?
- Is the pattern different in the three groups?

Replicated patterns

Let us assume that the spatial point patterns are generated by stationary and isotropic spatial point processes.

The spatial pattern studied by using Ripley's K function.

Remark: Other summary statistics (e.g. nearest neighbour distance function) can be used as well.

Pooled K function

- First, we estimate the K function for each replicate, i.e. $\widehat{K}_{ij}$, $i = 1, 2, 3$ and $j = 1, \ldots, m_i$, where m_i is the number of replicates (subjects) in the group i.
- Group specific mean functions can then be estimated by

$$\bar{K}_i(r) = \sum_{j=1}^{m_i} w_{ij} \widehat{K}_{ij}(r), \quad i = 1, 2,$$

where $w_{ij} = n_{ij} / \sum_{k=1}^{m_i} n_{ik} = n_{ij}/n_i$

- Finally, the mean over all groups can be estimated by

$$\bar{K}(r) = \frac{1}{n} \sum_{i=1}^{3} n_i \bar{K}_i(r),$$

where $n = \sum_{i=1}^{3} n_i$

Pooled K function: comparing the groups

The test statistic used by Diggle *et al.* (1991) is

$$D = \sum_{i=1}^{3} \int_0^{r_0} \left(\sqrt{\bar{K}_i(r)} - \sqrt{\bar{K}(r)} \right)^2 dr.$$

The following bootstrap procedure can be used

1. define residual K functions $R_{ij}(r) = n_{ij}^{\frac{1}{2}}(\widehat{K}_{ij}(r) - \bar{K}_i(r))$
2. obtain an empirical approximation to the distribution of D by recomputing its value from a large number of bootstrap samples

$$\widehat{K}^*_{ij}(r) = \bar{K}(r) + n_{ij}^{-\frac{1}{2}} R^*_{ij}(r),$$

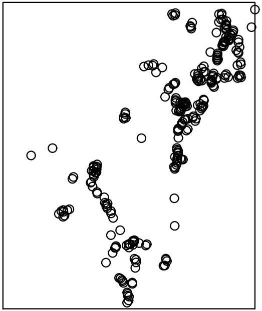
   where $R^*_{ij}(r)$'s are obtained by drawing at random and with replacement from the empirical distribution of the $R_{ij}(r)$, keeping the group sizes fixed

Remark

- Similar "pooling" ideas can be used when e.g. parameters of pairwise interaction processes are estimated based on replicated point patterns using the so-called pseudo-likelihood method (Diggle *et al.*, 2000)
- Spatial covariate information can be included in such models (see e.g. Bell and Grunwald, 2004; Illian and Hendrichsen, 2010).

Inhomogeneity

Example: Murchison data consist of 255 gold deposit locations in $330 \times 400km$ study region (Baddeley *et al.*, 2012)



Intensity can be estimated parametrically or non-parametrically

Kernel estimation of intensity

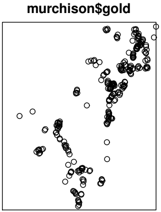
A fixed-bandwidth kernel estimate for point patterns was suggested by Diggle (1985)

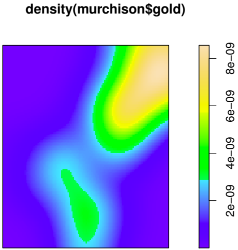
Intensity can be estimated by

$$\lambda(u) = \sum_{i=1}^{n} k(x_i - u) e(x_i),$$

where k is the isotropic Gaussian kernel with point masses at each of the data points in the point pattern, $e(x_i)^{-1} = \int_{y \in W} k(y - x_i)\, dy$ is an edge correction factor, and W is the observation window.

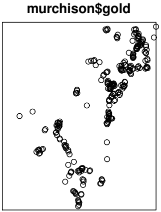
Kernel estimated intensity surface





Spatial covariate

255 gold deposit locations (left) and geological faults (right) in the same region



How to estimate the intensity as a function of spatial covariates?

Spatial covariate

- We assume that the point process intensity is a function of the covariate(s)
- Data set is a finite set y of points representing the locations of events observed in a sampling window W
- The values of the covariate function $X : W \to \mathbb{R}$ are known at every spatial location $u \in W$.
- y is modeled as a realisation of a spatial point process Y with intensity function $\lambda(u)$ depending on $X(u)$,

$$\lambda(u) = \rho(X(u)),$$

where ρ is the function to be determined.

Parametric estimation of ρ

- The simplest and most popular parametric model for dependence of Y and X is the loglinear model

$$\lambda(u) = \exp(\beta^T X(u)),$$

where β is a parameter vector.

- For the gold deposit data, distance from the nearest geological fault is a natural covariate
- The model above would say that the abundance of gold deposits depends on distance from the nearest fault.
- If such a model can be extrapolated to other spatial regions, then the observed pattern of faults in another area can be used to identify areas where gold deposits are more likely to be

Gold deposit data

- Baddeley *et al.* (2012) suggest first kernel estimates ρ by
  - pointing out that the values $x_i = X(y_i)$ constitute a point process in $\mathbb{R}$
  - relating the intensity function of the spatial point process Y and the intensity function of the "covariate" point process
  - suggesting kernel estimates for ρ (related to estimating a probability density from a biased sample)
- The estimated ρ is approximately an exponential function
- Then, the loglinear model

$$\lambda(u) = \exp(\alpha + \beta d(u)),$$

where $d(u)$ is the distance to the nearest geological fault, is fitted to the data.

Inhomogeneous K function

For certain non-stationary process (e.g. inhomogeneous Poisson process), the inhomogeneous K function can be defined as

$$K_{\text{inhom}}(r) = \frac{1}{|B|}\mathbb{E}\left(\sum_{y_i \in Y \cap B} \sum_{y_j \in Y \setminus \{y_i\}} \frac{\mathbf{1}(||y_i - y_j|| \leq r)}{\lambda(y_i)\lambda(y_j)}\right), \ r \geq 0,$$

for any bounded set B with $|B| > 0$. (We define $a/0 = 0$ for $a \geq 0$.) (Baddeley *et al.*, 2000)

Remarks:

- Stationary K is K_{inhom}, where $\lambda(y_i) = \lambda(y_j) = \lambda$
- There is no natural reference process
- Estimation of this function is sensitive to the choice of intensity estimate

Anisotropy: directional summary statistics

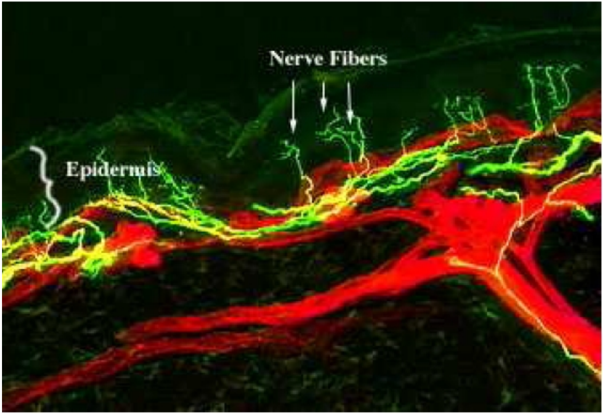
- Fry method (Fry, 1979)
- Point-pair-rose-density (Stoyan and Beneš, 1991)
  - Choose a pair of points with distance in a certain interval (r_1, r_2) at random and determine the angle β between the line going through the points and the 0-direction.
  - This angle is a random variable taking values between 0 and π, whose distribution gives information on the arrangement of the points.
  - The point-pair rose density is the corresponding probability density function
- Directional version of Ripley's K function (Stoyan, Kendall and Mecke, 1995): The K function is estimated separately in the directions of interest. Different behaviour in different directions can then reveal anisotropies in a point pattern.

Difficulties in 3D

- Definitions of 2D functions carry over to the 3D case
- The practical evaluation (e.g. edge corrections) as well as the visualisation of the results becomes more challenging.
- Which partition of the ball is suitable?

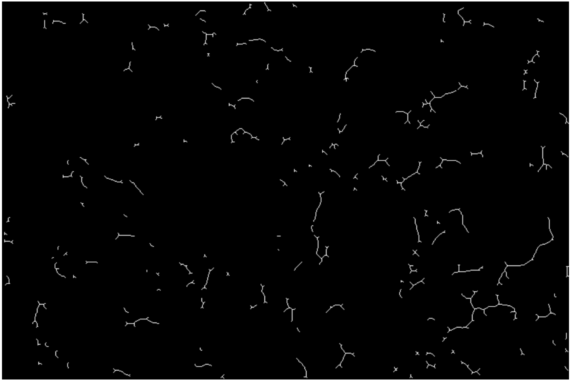
Example 1: epidermal nerve fiber patterns

Epidermal nerve fibers are thin nerve fibers in the epidermis (the outmost living layer of the skin)

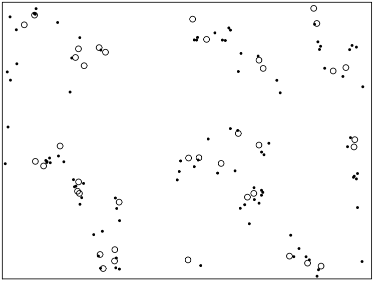


Example 1: fiber pattern

3151 (normal), skeletonized image



Example 1: reduced to a point pattern of entry (base) and end points

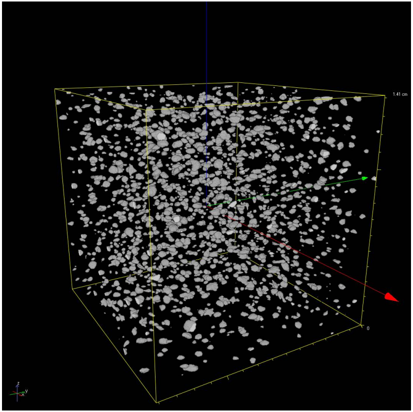


Base points (open circles) and end points (small black dots)

Example 1: problem

- Main questions:
  - Is the spatial pattern of nerve fibers of a subject suffering from some small fiber neuropathy (e.g. diabetic neuropathy) different from the pattern of healthy subjects?
  - Do some other covariates (age, gender, BMI) affect the nerve fiber pattern?
- Main methodological question: How to include non-spatial covariates in the spatial analysis?
- Repetitions and non-spatial covariates

Example 2: air inclusion pattern in polar ice



Example 2: problem

- Anisotropic patterns in 3D and missing information
- Main question: How can we estimate the deformation history in polar ice?
- Method: Study the anisotropy (deformation) of air inclusions in the ice
- Final goal is to determine how old the ice is

References

- Baddeley, A., Chang, Y.-M., Song, Y., Turner, R. Nonparametric estimation of the dependence of a spatial point process on spatial covariates. *Statistics and its interface* 5 (2012) 221-236.
- Baddeley, A.J., Møller, J., Waagepetersen, R. Non- and semiparametric estimation of interaction in inhomogeneous point patterns. *Stat. Neerl.* 54 (2000) 329-350.
- Baddeley, A.J., Moyeed, R.A., Howard, C.V., Boyde, A. Analysis of a three-dimensional point pattern with replication. *Applied Statistics* 42 (1993) 641-66.
- Bell, M.L., Grunwald, G.K. Mixed models for the analysis of replicated spatial point patterns. *Biostatistics* 5 (2004) 633648.

References (continues)

- Diggle, P. A Kernel method for smoothing point process data. *Applied Statistics* 34 (1985) 138-147.
- Diggle, P.J., Lange, N., Beneš, F.M. Analysis of variance for replicated spatial point patterns in clinical neuroanatomy *Journal of the American Statistical Association* 86 (1991) 618-625.
- Diggle, P.J., Mateu, J., Clough, H.E. A comparison between parametric and non-parametric approaches to the analysis of replicated spatial point patterns. *Advances in Applied Probability* 32 (2000) 331-343.
- Fry, N. Random point distributions and strain measurements in rocks. *Tectonophysics* 60 (1977) 89-105
- Illian, J., Hendrichsen, D.K. Gibbs point process models with mixed effects. *Environmetrics* 21 (2010) 341353.

References (continues)

- Stoyan, D. Describing the anisotropy of marked planar point processes. *Statistics* 22 (1991) 449462.
- Stoyan, D., Beneš, V. Anisotropy analysis for particle systems. *Journal of microscopy* 164 (1991) 159168.
- Stoyan, D., Kendall, W.S., Mecke, J. *Stochastic Geometry and its Applications* (Second ed.). Chichester: Wiley (1995).