TMA4250 Spatial Statistics
Assignment 2: Point Processes

IMF/NTNU

February 2017

Introduction

This assignment contains problems related to spatial Point Processes. R can be used for solving the problems and relevant functions can be found in the R library `spatial`.

Problem 1: Poisson Point Process (PPP)

Consider the point process $\{\boldsymbol{s_i} \in \mathcal{D}_s \subset \mathcal{R}^2 \; ; i = 1, \ldots, Z(\mathcal{D}_s)\}$ with $\mathcal{D}_s = (0,1) \times (0,1)$. Assume that the point process is an inhomogenuous Poisson point process (PPP) with intensity function

$$\lambda(s_v, s_h) = \mu \exp\{-(\alpha s_v + \beta s_h)\} \; ; \; \boldsymbol{s} = (s_v, s_h) \in \mathcal{D}_s$$

where $\boldsymbol{\theta} = (\mu, \alpha, \beta)$ are model parameters.

a) The number of points in $\mathcal{D}_s$, $Z(\mathcal{D}_s)$, is a random variable. Write down the distribution $[Z(\mathcal{D}_s)|\lambda(\cdot)]$. Develop an expression for the expected number of points in $\mathcal{D}_s$, i.e find $\mathrm{E}\{Z(\mathcal{D}_s)|\lambda(\cdot)\}$. What is the variance $\mathrm{Var}\{Z(\mathcal{D}_s)|\lambda(\cdot)\}$?

b) Set the model parameters to $\theta = (200, 2, 1)$ in the intensity function.
Generate realisations of $[\{\boldsymbol{s_1}, ..., \boldsymbol{s}_{Z(\mathcal{D}_s)}\}|Z(\mathcal{D}_s), \lambda(\cdot)]$ for $Z(\mathcal{D}_s) = 20$, 50 and 200 respectively by using a rejection sampling algorithm.
Visualize the results and comment.
Use the option `par(pty="s")` to plot the points on a square.

c) Generate 10 realisations of $[Z(\mathcal{D}_s), \{\boldsymbol{s_1}, ..., \boldsymbol{s}_{Z(\mathcal{D}_s)}\}|\lambda(\cdot)]$ for $\theta = (200, 2, 1)$.
Visualize the results and comment.

d) Assume that you have generated one realisation of the PPP. Specify the likelihood for the PPP. Use this to develop expressions for the maximum likelihood estimates for the parameters α, β and μ.

Choose one of the realisations in **c)**. Compute the numerical values of the maximum likelihood estimates for α, β and μ given the points in the chosen realisation. The function `optim()` may be useful for maximizing the likelihood numerically.
Compare the estimates of α, β and μ with the true values, and visualize the estimated and true intensity function by using the function image.plot(). Comment the results.

Problem 2: Neymann-Scott Point Process (N-SPP)

In this exercise you are going to explore the properties of a Neymann-Scott Point Process. A Neymann-Scott PP can be used to model clustered point processes. It is defined by a mother-model and a child-model: The mother-model is a statistical distribution that specifies the number of clusters and their locations in our spatial domain. The child-model specifies the point pattern around each of the clusters.

Consider a Neymann-Scott PP (N-SPP) $\{\boldsymbol{s_i} \in \mathcal{D}_s \subset \mathcal{R}^2 \ ; i = 1, \ldots, Z(\mathcal{D}_s)\}$ with $\mathcal{D}_s = (0,1) \times (0,1)$. Let the mother-model be given by a homogeneous Poisson PP with parameter λ_m. Further, let the child-model be specified as follows: The number of points is given by a Poisson distribution with parameter λ_c, while the point locations are given by an uncorrelated bi-Gaussian distribution with covariance matrix $\boldsymbol{I}\sigma^2$. (These model assumptions defines the so called Thomas PP.)

a) Vary the set of model parameters and generate realisations of the N-SPP. Explain how you perform the simulations by writing a simple pseudocode. Pay particular attention to boundary-problems in your simulation: You need to find a way to deal with points falling outside the spatial domain $\mathcal{D}_s$. Include some of the realisations in your report. Describe the resulting point patterns, and explain the relation between the realisations and the parameter values of the mother- and child-model.

b) Use the function Kfn() in the R library spatial to compute the empirical L-function for five extreme realisations from **a)**. For this purpose, you need the function ppregion() to specify the spatial domain.

The theoretical L-function for a Thomas PP is

$$L(h) = \sqrt{h^2 + (\pi\lambda_m)^{-1}(1 - \exp\{-h^2/4\sigma^2\})},$$

where λ_m is the intensity of the mother-model, σ^2 is the variance of the bi-Gaussian distribution, and h is the distance from an arbitrary event. Display the estimated and the theoretical L-function for five extreme realisations.

Comment the results, i.e relate the shape of the L-functions to the point patterns.

Problem 3: Strauss Point Process (SPP)

A Strauss point process can be used to model repulsion. Consider a Strauss point process (SPP) $\{\boldsymbol{s_i} \in \mathcal{D}_s \subset \mathcal{R}^2 \; ; \, i = 1, \ldots, Z(\mathcal{D}_s)\}$ with $\mathcal{D}_s = (0,1) \times (0,1)$. Assume that the pdf, given the number of points $Z(\mathcal{D}_s)$, related to this PP is:

$$[\{\boldsymbol{s_1}, ..., \boldsymbol{s}_{Z(\mathcal{D}_s)}\}|Z(\mathcal{D}_s), \theta] = \text{const} \times \exp\{-\sum_{i=1}^{Z(\mathcal{D}_s)} \sum_{j=1}^{Z(\mathcal{D}_s)} \varphi(\tau_{ij}; \theta)\}$$

with $\tau_{ij} = |\boldsymbol{s_i} - \boldsymbol{s_j}|$ being the euclidean distance between $\boldsymbol{s_i}$ and $\boldsymbol{s_j}$, and

$$\varphi(\tau_{ij}; c, d) = \begin{cases} c & \text{if } \tau_{ij} < d \\ 0 & \text{else} \end{cases}$$

a) How can you interpret the parameters c and d?

b) Set $Z(\mathcal{D}) = 50$ and $d = 0.1$ and generate realisations of the SPP with $c = 0.01$, 1 and 100 respectively. You can for example use a one-point updating Metropolis-Hastings MCMC algorithm for making the simulations. Show that the algorithm has converged for each value of c.
Visualize the results and comment.

Problem 4: Analysis of Point Patterns

Consider three real data point patterns in the R library MASS:

- biological cell data, available at cells.dat
- redwood tree data, available at redwood.dat
- pine tree data, available at pines.dat

Use the command `data<-ppinit("file.dat")` to load the data files.

a) Display the three point patterns and describe what you see. Try to relate the point patterns to real processes in nature. Use google to find information about the datasets.

b) Compute the empirical L-function for each of the point patterns by using the function Kfn().
Visualize the results and comment.

Specify the theoretical L-function for a homogeneous Poisson PP. Compare the empirical L-function for each of the point patterns with the theoretical L-function for a homogeneous Poisson PP.
Visualize the results and comment.

c) Take the number of points in each point pattern into account. Generate 100 realisations of a corresponding homogeneous Poisson PP, and compute the empirical L-function for each realisation. Use the results to estimate the expected L-function with associated 0.95 confidence intervals. Use the results to test informally whether each of the point patterns in cells.dat, redwood.dat and pines.dat could come from an underlying Poisson PP.
Visualize the results and comment.

d) Consider a homogeneous Poisson PP in $\mathcal{R}^2$ with intensity λ. Derive the distribution of the distance r from one arbitrary location in our spatial domain to the closest data point.
Hint: Consider a disc with radius r and with its origin at this arbitrary location.

For a homogeneous PPP the pdf of the distance r from an arbitrary location to the closest datapoint has the same distribution as the distance from a given datapoint to its closest neighbour. This pdf is called the r-function of the point process.

Use the information above to perform the same procedure as in **c)**, but now by using the r-function of the homogeneous PPP: The pdf of the distance from a given point to its closest neighbour. The functions `rdist()` and `density()` may be useful for computing an empirical r-function. As before, use the results to test informally whether each of the point patterns could come from an underlying homogeneous Poisson PP.