TMA4250 Spatial Statistics
Assignment 2: Event Spatial Variables

IMF/NTNU/HO

February 2020

Introduction

This assignment contains problems related to event spatial variables and Poisson random fields (RF). The package R can be used for solving the problems and relevant functions may be found in the R library `spatial`.

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`

Problem 1: Analysis of Point Patterns

Consider the three real data point patterns defined above.

a) Display each point pattern and discuss their appearance. Try to relate the point patterns to real processes in nature.

b) Compute the empirical L or J-function for each of the point patterns, the function `Kfn( . )` may be used.
Display the functions for each point pattern and discuss their appearance.

Specify the expression for the theoretical L and J-functions for a stationary Poisson RF. Display the empirical L or J-function for each of the point patterns with the corresponding theoretical function for a stationary Poisson RF, and discuss whether a stationary Poisson RF appear as a suitable model for each of the point patterns.

c) An empirical Monte Carlo test for whether the stationary Poisson RF is a suitable model shall be performed.

Consider each point pattern - under the hypothesis that the point pattern originates from a stationary Poisson RF, conditional on the actual point count observed, and generate 100 realizations of the Poisson RF. For each realization compute the associated L or J-function and use the set of functions to empirically test whether the actual point pattern could originate from a Poisson RF.
Display the empirical 0.9-intervals for the functions jointly with the corresponding estimated functions for each point pattern. Discuss, for each point pattern, whether a Poisson RF is a suitable model.

Problem 2: Bayesian inversion in Poisson RF

Consider an area of size (300×300) m^2 containing a pine tree forest, and the actual locations of the pine trees are to be assessed. The pine tree locations are observed from a satellite by remote sensing, and due to partly cloudy weather the observation probability for individual trees vary across the area.

Consider a discretization of the area into a regular $([1, 30] \times [1, 30])$-grid $\mathtt{L}$ with grid unit size 100 m^2. The true, but unknown, number of pine trees located in each grid unit is $\{k(\boldsymbol{x}); \boldsymbol{x} \in \mathtt{L}\}$.

The probabilities for observing a pine tree occuring in a given grid unit is represented by $\{\alpha(\boldsymbol{x}); \boldsymbol{x} \in \mathtt{L}\}$ - and these probabilities varies across the area. The probabilities are listed in the file `obsprob.txt` together with the x and y coordinates of the centroids of the grid units. The number of pine trees observed in each grid unit is given by $\{d(\boldsymbol{x}); \boldsymbol{x} \in \mathtt{L}\}$ - and these numbers are listed in the file `obspines.txt` in a format similar to the probabilities above.

a) Display the observations and the observation probabilities. Assume that the observations in the grid units, given the true number of pine trees in each grid unit, are spatially uncorrelated. Specify the expression for the corresponding likelihood model for the observations.

b) Assume apriori that the distribution of pine trees is according to a stationary Poisson RF with model parameter λ_k. Specify the expression for the corresponding prior model for the discretized Poisson count model.

c) Estimate the intensity λ_k based on the observations in the grid units with associated observation probabilities. Generate six realizations from this prior stationary Poisson event-count model and the associated approximate Poisson event-location realizations. Display the approximate Poisson event-location realizations.

d) Develop the expression for the posterior discretized event-count model and justify that this posterior model is a discretized Poisson RF model. Generate six realizations of the associated approximate event-location model, and display these realizations. Discuss the similarities and differences between the prior and posterior approximate event-location realizations.

e) Simulate 100 realizations of the discretized event-count model, both for the prior and the posterior models. Compute the average of these 100 realizations, representing the expected event-count number, for each of the two models. Display these averages graphically, and compare the displays and explain the differences.

Problem 3: Clustered event spatial variables

Consider the redwood tree data listed above.

a) Consider the Neuman-Scott (Mother-Child) cluster model for spatial event variables, with Poisson mother model and Gaussian child intensity model. Describe shortly the model and specify the full set of model parameters. Discuss potential border problems, caused by the use of a finite domain `D`, when simulating realizations from the model. Suggest solutions to these problems.

Make an empirical fit of the model parameters to the redwood tree data. This fit need only be done by inspecting the tree pattern and guestimate the model parameter values from your intuitive understanding of their impact on the pattern. Evaluate your parameter values by a Monte Carlo test on the J or L-interaction function.

Iterate your guestimate procedure to improve the fit and try to make the Monte Carlo test appear as significant. List the final model parameters guestimates and justify them by displaying the Monte Carlo test results. Discuss the results.

Display the redwood tree data set next to three realizations from the guestimated Neuman-Scott model. Comment on the display.

Problem 4: Repulsive event spatial variables

Consider the biological cell data listed above.

a) Specify the expression for the Strauss repulsion model for spatial event variables with exponential interaction function and fixed event count. Describe the full set of model parameters. Discuss potential border problems, caused by the use of a finite domain `D`, when simulating realizations from the model. Suggest solutions to these problems.

Make an empirical fit of the model parameters to the biological cell data. This fit need only be done by inspecting the cell pattern and guestimate the model parameter values from your intuitive understanding of their impact on the pattern. Evaluate your parameter values by a Monte Carlo test on the J or L-interaction function.

Iterate your guestimate procedure to improve the fit and try to make the Monte Carlo test appear as significant. List the final model parameters guestimates and justify them by displaying the Monte Carlo test results. Discuss the results.

Display the biological cell data set next to three realizations from the guestimated Strauss repulsion model. Comment on the display.