Week 9

Spatial Point Processes

Overview

Welcome to Week 9 of Spatial Analysis.

This week we will cover two topics:

- Investigating Spatial Point Processes

- Modelling by using spatial Poisson processes

In the first topic of this week, we will consider several methods to work with spatial point processes. First, you will study several approaches in R to simulate spatial point processes with specified intensities. Then, you will learn about the Kolmogorov-Smirnov test of complete spatial randomness. This test is routinely used to check the randomness of spatial locations. We will demonstrate it using several datasets, in particular, locations of trees in New Zealand and tropical rainforests.

You will learn how to fit spatial Poisson process to data by using its intensity. We will examine the accuracy of such fitting by using spatial residuals and QQ plots. Then, we will consider spatial marked processes. Usually, marks take values from a finite set and indicate types of objects in the corresponding locations. We will learn how to split data depending on their marks and to analyse the corresponding intensities.

Finally, you will participate in a workshop, in which you can practise conducting an analysis of real spatial data.

By the end of this week, you will learn about:

Topic 1: Investigating Spatial Point Processes

- Simulation of homogeneous Poisson spatial processes.
- Simulation of inhomogeneous Poisson spatial processes.
- Kolmogorov-Smirnov test of complete spatial randomness.
- Applications to New Zealand and tropical rainforest trees data.

Topic 2: Modelling by using spatial Poisson processes

- Fitting a Poisson process to spatial data in R.
- Introduction to marked Poisson processes.
- Manipulating marks in R.
- Investigation of intensities of marked Poisson processes in R.
- Species interactions and cross-type pair correlation functions.

By completing this module, you will be working towards the following subject-intended learning outcomes:

1. Formulate purposeful questions to explore new statistical ideas and subsequently design valid statistical experiments.
2. Present clear, well-structured analysis of important statistical model results.
3. Creatively find solutions to real-world problems consistent with those commonly faced by practising statisticians.
4. Professionally defend or question the validity of existing statistical analyses and associated evidence-based conclusions that are derived via application of sound spatial statistical methodology.

Topic 1: Investigating Spatial Point Processes

In this topic we continue studying of spatial point processes and learn about their simulation and testing hypothesis of complete spatial randomness. We will demonstrate applications with R by using several simulated and real datasets, in particular, locations of trees in New Zealand and tropical rainforests.

Simulation of Poisson spatial processes

In this part, first, we formally define the homogeneous Poisson spatial processes. Then, you will learn how to simulate realisations of this process in R. We will consider the cases when the simulation is done for a given intensity and for a given number of points in a region. You will also learn how to define a region where the simulated points are placed. Then, you will study how to define the intensity function that is not constant and simulate the corresponding inhomogeneous spatial point process.

Read

You will read the first part of Studying Spatial Point Processes and will learn how to use R to simulate Poisson spatial processes.

From the folder, open the document titled **Week_9_Topic_1** and read **slides 1–6**.

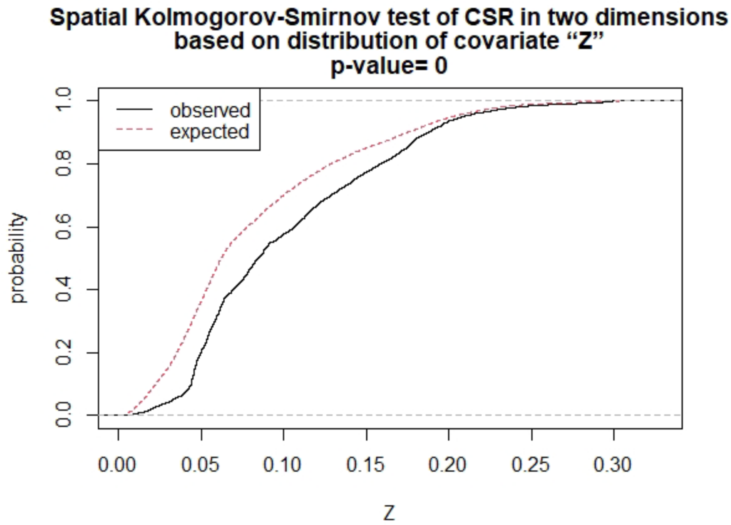
Kolmogorov-Smirnov test of CSR

In this section, you will learn testing hypothesis about distributions of points of a spatial process. The main model is the homogeneous Poisson point process. The intensity of this process is constant. In applied research, this model is often referred to as Complete Spatial Randomness (CSR). Testing hypotheses about this model means finding evidence against CSR. One of the standard approaches is to consider the statistic T, which depends on points' locations and has a known distribution under CSR. Then, some goodness-of-fit tests can be used to check whether there is sufficient evidence that the empirical distribution of data and the expected distribution under CSE are the same.

As for spatial data the statistics T can be chosen in many different ways, one usually checks several possible choices of T. For example, the distributions in coordinate axes directions, directions for some angles or distributions with respect to covariates. If at least one of such tests rejects the specific CSR hypothesis, then the data do not support the CSR hypothesis in whole.

Often, the Kolmogorov-Smirnov test is used as a goodness-of-fit test and p=0.05 as the p-value for rejections (see, for example, Fig 5.3). We will illustrate these methods by using locations of trees in New Zealand and tropical rainforests. We also demonstrate that for simulated data with known properties, the results of the Kolmogorov-Smirnov test are consistent with the expected results.

Figure 5.3



Read

In the second part of the reading Studying Spatial Point Processes, you will learn about testing hypothesis about complete spatial randomness by using the Kolmogorov-Smirnov test.

From the folder, open the document titled **Week_9_Topic_1** and read **slides 7–14**.

Revise key R commands used in this topic’s materials.

Read

In the reading [Key R commands](), you will revise some of the key R commands that were used in this topic’s materials.

From the folder, open the document titled **Week_9_Topic_1** and read **slide 15**.

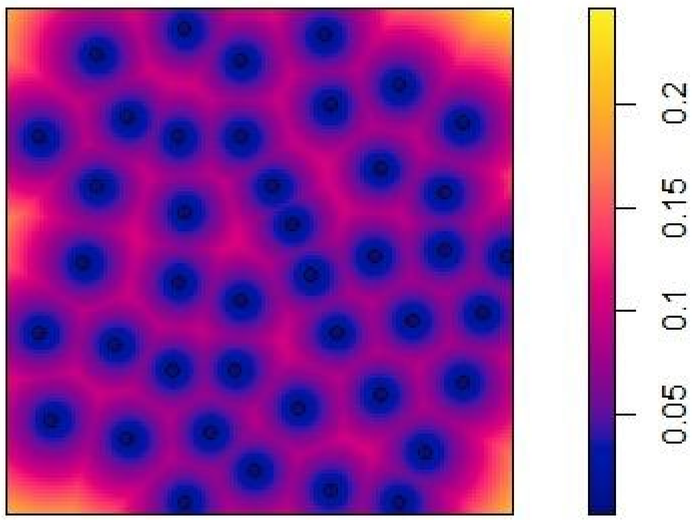
Topic 2: Modelling by using spatial Poisson processes

In this topic, we will mainly concentrate on spatial data modelled by Poisson processes. We will consider how to fit the spatial intensities of such processes and how to use different distance characteristics of spatial point patterns to compare them with the point distributions expected for spatial Poisson processes. We will also study spatial points with attached attributes which are called marks. For such marked point processes, we will learn how to split points due to their mark values and analyse interactions between data with different marks.

We will use several real datasets to illustrate applications of these approaches in R. These datasets are mainly from the forestry industry, but similar approaches can be applied to other point patterns as well.

Figure 6.1

Visualisation of empty space distance distribution around points



You will read the content, watch videos and complete activities to test your understanding of this topic.

Upon completion of this topic, you will be able to analyse spatial distributions, estimate their parametric intensity, and be familiar with distance characteristics of point patterns and their applications to distinguishing between 3 types of distributions. You will further develop your spatial computational skills using the spatstat package from R.

Fitting intensities of Poisson processes

In many applications, it is important to investigate the intensity of spatial points. If points in a spatial point pattern are from a realization of the Poisson process, then their intensity can be estimated by using parametric models. In this topic, you will fit Poisson processes under the assumption that their intensity function is log-linear. As you will see, it covers a rather wide class of spatial models. You will investigate the quality of such fits using spatial residuals. These residuals are used to determine areas, where the model should be improved.

Read

In the first part of the reading Studying Poisson process, you will explore how to fit the intensities of spatial Poisson processes. You will apply these methods to the data about trees from the bei dataset.

From the folder, open the document titled **Week_9_Topic_2** and read **slides 1–10**.

Marked point processes

In this part, you will learn about point patterns which have attached information for each point. They are called marked point processes. First, you will study several methods to work with marks. Then, we split spatial data in subsets with respect to the values of a mark and analyse their intensities. Finally, we will consider how to characterise between-mark interactions by using cross-type pair correlation functions. A cross-type pair correlation function is a function of the distance between marks. If this function takes values substantially below 1 it indicates inhibition between marks at a specific distance. Otherwise, for the values close to 1 and above, the marks can coexist at the corresponding distances.

You will apply these methods to several real datasets with R.

Read

In this part of the reading Studying Poisson process, you will explore how to work with marked spatial processes. You will apply the methods to datasets from the R package spatstat.

From the folder, open the document titled **Week_9_Topic_2** and read **slides 11–21**.

Revise key R commands used in this topic's materials.

Read

In the reading Key R commands, you will revise some of the key R commands that were used in this topic's materials.

From the folder, open the document titled **Week_9_Topic_2** and read **slide 22**.

Workshop

Activity

Workshop

This activity will be completed in R. Repeat the R programming content covered in Week 9. Modify the code and understand the impact of different R parameters on changes in results.

Your task

- Repeat R commands learnt in Week 9 prior to the two-hour workshop session.
- Try to modify the code and understand the impact and meaning of different R function parameters. Interpret the observed changes in plots and analysis results.
- Feel free to discuss questions with other students as you go in the forum, and please also take the time to help others. It is amazing how much we all can learn from each other's questions, and how in helping others we strengthen our own understanding.
- Come along to the two-hour scheduled workshop session and discuss any challenges, seek advice and work through some problems with your peers and facilitator.
- Revisit these problems in later weeks and challenge yourself to get a deeper understanding to build on what you learn later.

Guidelines

- This activity is not graded but is an essential part of your learning. It will be held synchronously and facilitated by your instructor.
- You don't need to submit your R code; however, to be successful in this subject it is necessary to work through all R coding materials from this week and understand how to apply the corresponding R commands.
- You should repeat all R commands in this week's materials before the workshop. This will give you an opportunity to efficiently work with the facilitator during the workshop and get your questions answered.
- You should spend around two hours on this activity.

Workshop solutions

The R code for Workshop is provided for your reference. Click this link to access the R code. From the folder, please open the document titled – **Week_9_Workshop_9_RCode**.

Summary

This week, we continued the studies of statistical and data science models based on spatial point processes. We learned about different applications of Poisson processes. Then, we considered marked point processes and their applications. We practised these point processes methods with the spatstat package.

Next week, we will continue learning about other applications of the obtained results for several problems in spatial data analysis.

Here's a list of tasks that you should be working on or have completed:

- **Required readings**
- **Workshop**

The following resources provide you with this week's references and additional suggested readings.

Additional suggested readings and resources

While these readings and resources are not essential, they provide greater insight into the concepts covered in the week's lectures and give you the choice to enhance your learning or pursue an area of interest in greater detail.

Software and data:

- CRAN documentation for the spatstat package: https://cran.r-project.org/web/packages/spatstat/index.html
- amacrine data set: https://rdrr.io/cran/splancs/man/amacrines.html
- Tropical rain forest trees data: https://rdrr.io/cran/spatstat.data/man/bei.html
- longleaf data: https://rdrr.io/cran/spatstat.data/man/longleaf.html
- lansing data: https://rdrr.io/cran/spatstat.data/man/lansing.html
- northcumbria data: https://rdrr.io/cran/stpp/man/northcumbria.html

Theoretical concepts:

- Spatial point processes: https://www.apps.stat.vt.edu/leman/VTCourses/BaddeleyPointProcesses.pdf
- Kolmogorov-Smirnov test: https://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm

Books:

- Bivand, R. S., Pebesma, E., & Gomez-Rubio, V. (2013). Applied spatial data analysis with R (2nd ed.). Springer. https://doi.org/10.1007/s12061-014-9118-y Avalilable on-line in La Trobe EBL ebook Library
- Baddeley, A., Rubak, E., & Turner, R. (2015). Spatial Point Patterns: Methodology and Applications with R. Chapman and Hall/CRC. https://doi.org/10.1201/b19708 Avalilable on-line in La Trobe EBL ebook Library