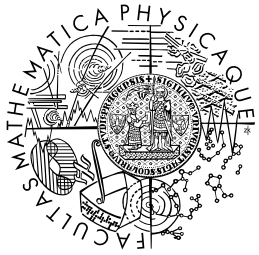

FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Kateřina Koňasová

Stochastic Reconstruction of Random Point Patterns

Department of Probability and Mathematical Statistics

Supervisor of the master thesis:	RNDr. Jiří Dvořák, Ph.D.
Study programme:	Mathematics
Study branch:	Probability, Mathematical Statistics and Econometrics

Prague 2018

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In ..... date ..... signature of the author

I would like to thank my supervisor, RNDr. Jiří Dvořák, Ph.D., for his enthusiasm, insightful advice, patient encouragement and unfailing support. His guidance helped me in all the time of research and writing of this thesis.

This work has been supported by the Charles University Grant Agency, project no. 472217.

The BCI forest dynamics research project was founded by S.P. Hubbell and R.B. Foster and is now managed by R. Condit, S. Lao, and R. Perez under the Center for Tropical Forest Science and the Smithsonian Tropical Research in Panama. Numerous organizations have provided funding, principally the U.S. National Science Foundation, and hundreds of field workers have contributed.

Title: Stochastic Reconstruction of Random Point Patterns

Author: Kateřina Koňasová

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Jiří Dvořák, Ph.D., Department of Probability and Mathematical Statistics

Abstract: Point procesess serve as stochastic models for locations of objects that are randomly placed in space, e.g. the locations of trees of a given species in a forest stand, earthquake epicenters or defect positions in industrial materials. Stochastic reconstruction is an algorithmic procedure providing independent replicates of point process data which may be used for various purposes, e.g. testing statistical hypothesis. The main advantage of this technique is that we do not need to specify any theoretical model for the observed data, only the estimates of selected summary characteristics are employed. Main aim of this work is to discuss the possibility of extension of the stochastic reconstruction algorithm for inhomogeneous point patterns.

Keywords: stochastic reconstruction, simulation, point process, inhomogeneity

Contents

Introduction

Point procesess serve as stochastic models for locations of objects that are randomly placed in space. They are widely used in many scientific disciplines as biology, ecology, particle physics, material science or astronomy. The points may represent locations of trees of a given species in a forest stand, earthquake epicenters, burrows or nests of animals, locations of stars or galaxies or defect positions in industrial materials. Authors of the book Møller and Waagepetersen [2004] claim that in last 30 years spatial point processes have been a major area of research in spatial statistics and they expect that research in spatial point processes will continue to be of particular importance. They also point out that thanks to new technologies spatial point process data become more available and a lot of new applications emerge.

In spatial statistics simulations can be used for a number of different aims, e.g. exploring the sampling variation of estimated summary characteristics or testing various hypotheses about the observed data using simulation-based tests. In this text we focus on situations when we do not want to look for an explicit parametric point process model (from which we could simulate) for the observed data or the null hypothesis which we want to test is not specific enough to enable simulations. In both of these cases we can use the stochastic reconstruction approach instead of simulations to provide independent replicates of point process data.

Stochastic reconstruction is an algorithmic procedure which has tradition in statistical physics, see Chapter 12 in Torquato [2002]. In the context of stationary point processes the algorithm was described in Tscheschel and Stoyan [2006]. Nowadays stochastic reconstruction is of particular interest in biology and ecology. It can be used when monitoring forest ecosystems, see Getzin et. al. [2014] or Lilleleht et al. [2014]. Of course, many other applications of this technique can be found in recent scientific papers, for example the quasi-plus sampling edge correction method in Tscheschel and Chiu [2008] or the model-free isotropy test in Wong and Chiu [2016].

The main idea of the stochastic reconstruction approach consists in measuring "similarity" of two point patterns. It is done by evaluating the deviation of their estimated summary characteristics. For this purpose, the energy functional E is established. While reconstructiong point process data, we start with an initial point configuration with the same number of points as the data has. In each iteration step of the stochastic reconstruction algorithm we try to move one point of the current configuration in order to obtain new configuration with lower energy (smaller value of the energy functional E). The smaller the value of the energy functional E is, the smaller the deviation of the estimated summary characteristics is. Since the summary characteristics contain information about the point pattern structure, by minimizing the energy functional we will obtain point pattern with similar form of the estimated summary characteristics as the observed data and thus similar structure. Note, however, that the summary

characteristics used in practice do not characterise distribution of a point process.

Example of the atochastic reconstruction method can be see in Figure 1. In the top left corner the point process data can be seen. The data are obviously clustered, i.e. we observe a number of different clusters randomly placed in the observation window. Next to the data the initial configuration can be seen. As the number of iteration steps growths, the values of the energy functional decrease and the intermediate states of the algorithm start to look similar to the observed data. In fact the points of the initial configuration are stepwise moved (each point may be moved several times) to form clusters with the size and shape similar to the observed data. Evolution of the values of the energy functional E for this particular run of the stochastic reconstruction algorithm can be see in Figure 2. In this case the estimated summary characteristics are the empirical distribution functions of the distances to the k-th nearest neighbour for $k = 1, 2, \ldots, 5$.

Let us now demonstrate the use of the stochastic reconstruction algorithm on an practical example. We will work with the **BCI** dataset (see Hubbell et al. [2005]) which contains information about locations of 311 species of trees in the tropical forest at Barro Colorado Island, Panama. For further information

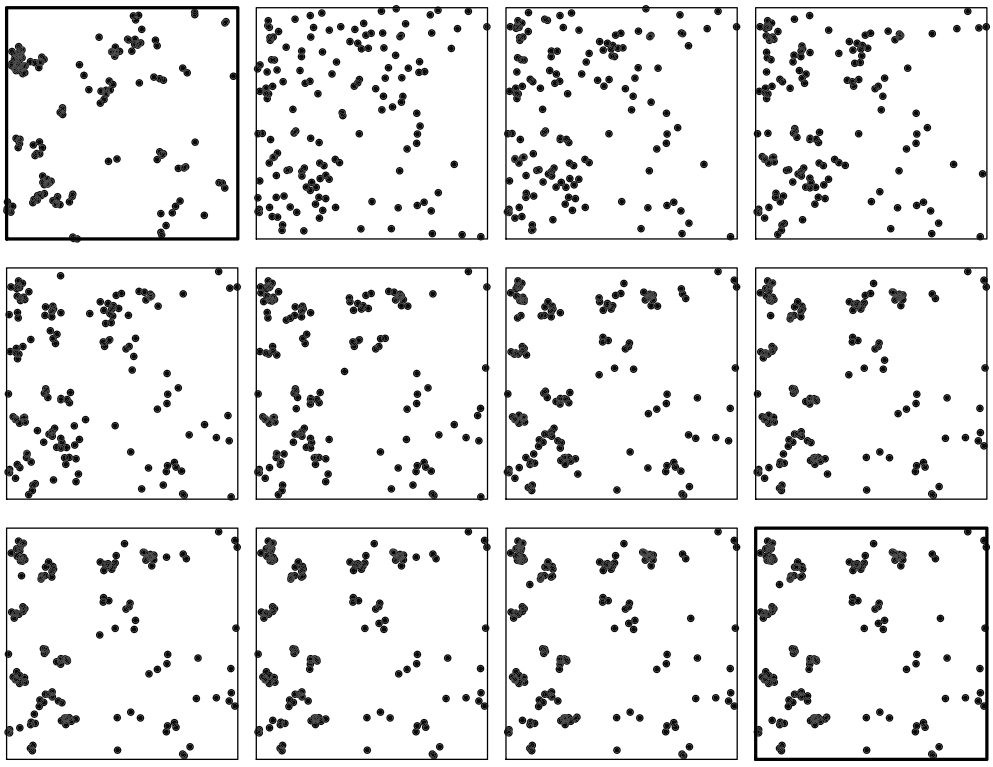

Figure 1: Illustration of the stochastic reconstruction method. Observed data to be reconstructed (top left corner), initial configuration, intermediate states of the stochastic reconstruction algorithm after 100, 200, 300, 400, 800, 1000, 1100, 1300 and 1400 iteration steps (from left to right and from top to bottom), output of the stochastic reconstruction algorithm (bottom right corner). Values of the energy functional can be seen in Figure 2.

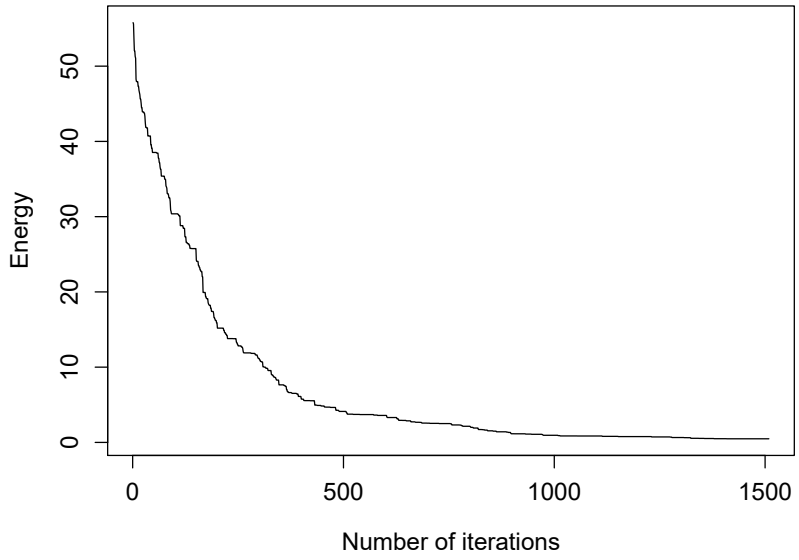

Figure 2: Values of the energy functional E during the run of the stochastic reconstruction algorithm from Figure 1. We have stopped the run of the algorithm after rejecting the proposed new configuration 100 times in a row.

see Condit [1998] and Hubbell et al. [1999].

Our attention will focus on two related species, namely the Zanthoxylum ekmanii which will be refered to as population A and Zanthoxylum panamense which will be refered to as population B. Our observation window is a rectangle, the shorter edge is 500 meters long and the longer edge is 1000 meters long. The area of the observation window is hence 50 hectares. The observed data which can be seen in Figure 3 consists of locations of 235 individuals from the population A and 188 individuals from the population B.

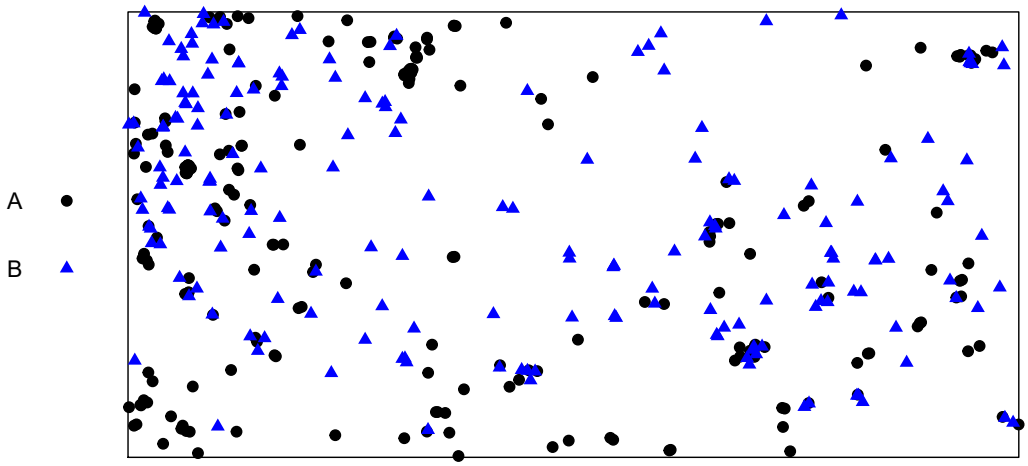

Figure 3: BCI dataset: locations of the Zanthoxylum ekmanii (black circles) and Zanthoxylum panamense (blue triangles) in the tropical forest at Barro Colorado Island. The data can be considered as clustered.

The data corresponds to a realisation of planar marked point process. It means that to each point of the process a mark (i.e. some supplemental information) is attached. In our case a point gets the mark A if it represents location of a tree from the population A and mark B if it represents location of a tree from the population B. We want to test the hypothesis that these two populations are independent, i.e. occurence of individuals from population A in the observation window is not affected by the locations of individuals from population B (and vice versa).

The hypothesis that populations A and B are independent corresponds to the so-called random superposition hypothesis. It means that the observed point pattern is formed by two independent point patterns (one cointains only the points with mark A and will be denoted as φ_A, the other contains only the points with mark B and will be denoted as φ_B) that have been joined into one point configuration. While dealing with real data, we can be never sure if it is a realisation of a stationary point process or not. Thus we will treat the data as a realisation of an inhomogeneous point pattern with non-constant intensity function. It means that we suppose that φ_A and φ_B are realisations of inhomogeneous point processes. Therefore we will use the inhomogeneous counterpart of the cross K-function $K_{AB}(r)$, $r > 0$, as the test statistic. More details can be seen in Section 4.4.1 in Møller and Waagepetersen [2004]. When estimating $K_{AB}(r)$, we are counting pairs of points with different marks that are closer together than the distance r. Such pairs of points do not contribute to the final sum with the same weight – for each of them the weight includes the edge correction factor and the normalisation by the non-constant intensity functions.

If we want to test a hypothesis about a point process data, the simulation-based tests are usually used. It means that we generate a large number of simulations from a null model, we compute the test statistics for these simulations and for the observed data and we compare how extreme is the value of the test statistic

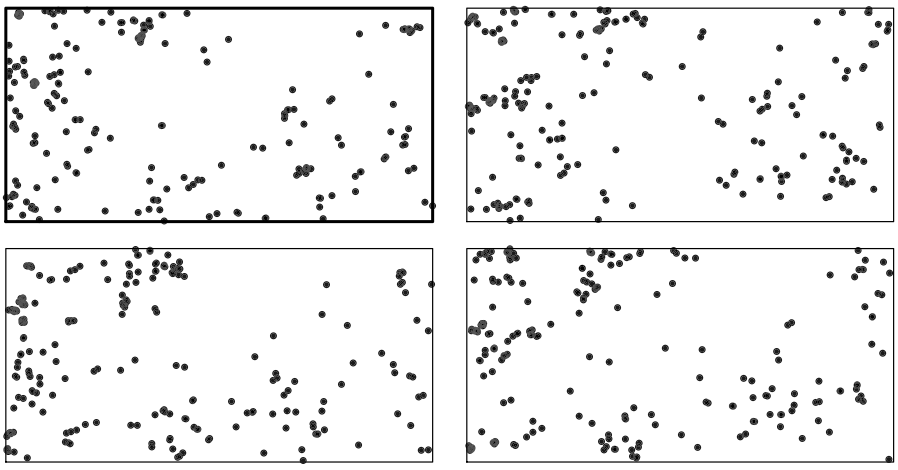

Figure 4: Reconstruction of the locations of points with mark A: the observed data (top left corner), three different reconstructions.

computed for the data compared to the values computed for the simulations. If the test statistics is a function (as the cross K-function) we can use the global rank envelope test. It gives us a recipe how to order the curves computed for the simulated patterns and how to decide whether the data curve is extreme (compared to the simulated ones) or not. Moreover, the global rank envelope test has a graphical representation. Based on the simulated curves we are able to draw the envelope (i.e. the bounded region in which the observed curve should lie, with a given probability, under the null hypothesis). If the data curve leaves the envelope for any value of the patameter r we reject the null hypothesis.

The key element to perform the simulation-based test are the simulations from the null model. But the null hypothesis that two different populations are independent is not specific enough to enable simulations. Fortunately, we can use the stochastic reconstruction instead. Locations of points with mark B stay fixed and we generate 2499 reconstructions of the locations of points with mark A (three of these 2499 reconstructions can be seen in Figure 4). Then we superimpose the patterns and we obtain 2499 marked point patterns. Once we have these patterns we can compute the test statistic $K_{AB}(r),\ r > 0$, and create the envelope.

The global rank envelope test is in this situation performed on the significance level $\alpha = 0.05$. For the stochastic reconstruction we use the energy functional E which is modified for inhomogeneous point patterns. In Section 3.3.1 this functional will be refered to as E_5'. The algorithm is stopped if the energy functional E does not decrease in the 500 iteration steps in a row. The global rank envelope can be seen in Figure 5. The p-interval (whose bouns correspond to the most liberal and the most conservative p-values) given by the global rank envelope test is $(0.029, 0.061)$ and thus it is not clear whether we should reject the null hypothesis or not.

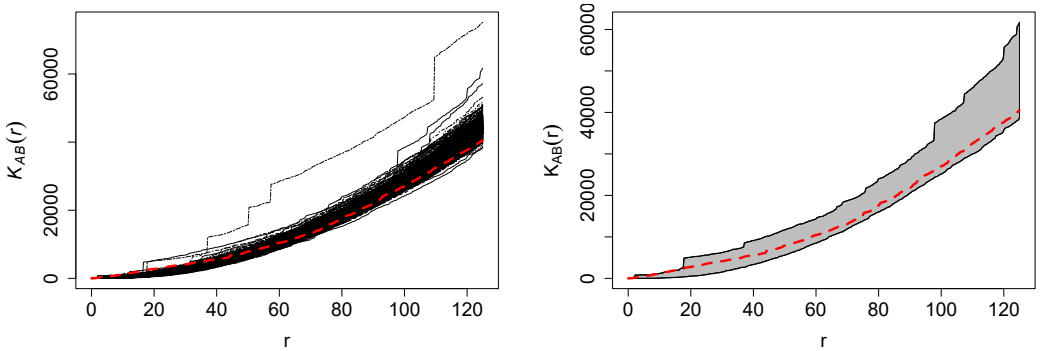

Figure 5: Left: test statistic $K_{AB}(r),\ r > 0$ computed for the data (red dashed curve) and for the 2499 superimposed patters. Right: global rank envelope (red dashed curve is the data curve) for the test statistic $K_{AB}(r),\ r > 0$, made from the 2499 superimposed point patterns.

The main aim of this work is to discuss the difficulties that may arrive when extending the stochastic reconstruction algorithm described in Tscheschel and Stoyan [2006] for inhomogeneous point processes. We will cover the direct method of minimizing the energy functional including a simulation study in which the quality of reconstructions will be investigated for three different point proces models. Also a brief discussion about the use of a special case of the Metropolis-Hastings algorithm described in Section 7.1.1 in Møller and Waagepetersen [2004] will be given.

1. Point processes in $\mathbb{R}^d$

In this chapter we will introduce the space of all locally finite point configurations in $\mathbb{R}^d$ and we will state the formal definition of a point process on a d-dimensional Euclidean space. We will also state the definition of some important summary characteristics and we will discuss the difference between a homogeneous and an inhomogeneous point process. Also four well-known models for spatial point paterrns will be mentioned. Note that most of notation and basic definitions are taken from Møller and Waagepetersen [2004].

1.1 Basic definitions

Take $d \in \mathbb{N}$. By $\mathbb{R}^d$ we mean the d-dimensional Euclidean space. Let us denote by $\mathfrak{B}^d$ the Borel σ-algebra on $\mathbb{R}^d$. Moreover denote by $\mathfrak{B}_0^d$ the set of all bounded Borel subsets of $\mathbb{R}^d$. Let ζ be a subset of $\mathbb{R}^d$. Denote by $N(\zeta)$ the cardinality of the set ζ, i.e. the number of elements of ζ. If ζ is not finite, set $N(\zeta) = \infty$.

Definition 1. *We say that $\zeta \subset \mathbb{R}^d$ is locally finite, if $N(\zeta \cap B) < \infty$ for all $B \in \mathfrak{B}_0^d$. Set*

$$\mathcal{N}_{lf} = \left\{\zeta \subset \mathbb{R}^d : \ N(\zeta \cap B) < \infty \ \forall B \in \mathfrak{B}_0^d\right\}.$$

$\mathcal{N}_{lf}$ is the set of all locally finite point configurations in $\mathbb{R}^d$. Elements of $\mathcal{N}_{lf}$ will be denoted by ζ, ξ, η ..., while x, y, z ... denote points in $\mathbb{R}^d$. We need to equip the space $\mathcal{N}_{lf}$ with some σ-algebra. Denote

$$\mathfrak{N}_{lf} = \sigma\left(\{\zeta \in \mathcal{N}_{lf} : \ N(\zeta \cap B) = m\}, \ B \in \mathfrak{B}_0^d, \ m \in \mathbb{N}_0\right),$$

where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Obvoiously, $\mathfrak{N}_{lf}$ is a σ-algebra. Hence $(\mathcal{N}_{lf}, \mathfrak{N}_{lf})$ is a measurable space. Since $\mathbb{R}^d$ is a Polish space, $\mathfrak{N}_{lf}$ is a countably generated σ-algebra, see Proposition B.1 in Møller and Waagepetersen [2004]. This fact will be used in Section 4.2.1. Now fix $(\Omega, \mathcal{A}, \mathbb{P})$ a probability space. We have already prepared all ingredients needed to state the definition of a point process.

Definition 2. *A measurable mapping $\Phi : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow (\mathcal{N}_{lf}, \mathfrak{N}_{lf})$ is called a point process on $\mathbb{R}^d$.*

Remark. The measurability of Φ implies that $N(\Phi \cap B)$ is a random variable for any $B \in \mathfrak{B}^d$, see Definition B.1 in Møller and Waagepetersen [2004].

Remark. Note that in a very same way we can define a point proces on $S \subseteq \mathbb{R}^d$. In Møller and Waagepetersen [2004] point processes defined on $S \subseteq \mathbb{R}^d$ are called spatial point processes.

This is not the only possibility how to define a point process in $\mathbb{R}^d$. It is also possible to consider all locally finite random counting measures μ as point processes, see Daley and Vere-Jones [2008]. Our definition 2 corresponds to so-called simple point process. It means that corresponding random counting measure μ satisfies $\mu(\{y\}) \leq 1 \ \forall y \in \mathbb{R}^d$. In this case we can identify μ with its support *supp* μ which is a locally finite point configuration in $\mathbb{R}^d$. We get

$\mu(B) = N\left(supp\ \mu \cap B\right)$ for all $B \in \mathfrak{B}^d$.

Let Φ be a point process, i.e. a random locally finite subset of $\mathbb{R}^d$. Let us denote $N_\Phi(B) = N(\Phi \cap B)$ for all $B \in \mathfrak{B}^d$. As we have mentioned above $N_\Phi(B)$ is a random variable for any $B \in \mathfrak{B}^d$. Naturally, its moments such as the expectation are of particular interest.

Definition 3. *The intensity measure Λ of a point process Φ is given by*

$$\Lambda(B) = \mathbb{E}\left[N_\Phi(B)\right],\ B \in \mathfrak{B}^d.$$

Once we have defined the intensity measure, we are able to speak about homogeneity and inhomogeneity of a point processes.

Definition 4. *Suppose that the intensity measure Λ of a point process Φ is translation invariant, i.e.*

$$\Lambda(x + B) = \Lambda(B), \quad for\ all\ x \in \mathbb{R}^d\ and\ B \in \mathfrak{B}^d,$$

where $x + B = \{x + b :\ b \in B\}$. Then Φ is said to be homogeneous point process. Ohterwise, Φ is said to be inhomogeneous.

Definition 5. *Suppose that the intensity measure Λ of a point process Φ is absolutely continuous with respect to d-dimensional Lebesgue measure, i.e. there exists a measurable function $\lambda : \mathbb{R}^d \longrightarrow [0, \infty)$ such that*

$$\Lambda(B) = \int_B \lambda(u)\ \mathrm{d}u \quad \forall\ B \in \mathfrak{B}.$$

Then λ is called the intensity function of Φ. If λ is a constant, then it is called intensity.

Remark. Roughly speaking, $\lambda(y)\ \mathrm{d}y$ is the probability that a point of the process Φ will occure in an infinitesimally small ball with centre y and volume $\mathrm{d}y$, see Section 4.1.1 in Møller and Waagepetersen [2004].

Instead of saying that Φ is homogeneous, we can call Φ first order stationary. It is not the same property as stationarity, which is defined as follows.

Definition 6. *A point process Φ on $\mathbb{R}^d$ is stationary, if its distribution (the probability measure $\mathbb{P}_\Phi(F) = \mathbb{P}\left[\Phi \in F\right],\ F \in \mathfrak{N}_{lf}$) is invariant under translations. It means that the distribution of $\Phi + s = \{X + s :\ X \in \Phi\}$ is the same as that of Φ for any $s \in \mathbb{R}^d$. Further Φ is called isotropic, if its distribution is invariant under rotations about the origin in $\mathbb{R}^d$. That is to say, the distribution of $\mathcal{O}\Phi = \{\mathcal{O}X :\ X \in \Phi\}$ is the same as that of Φ for any rotation $\mathcal{O}$ around the origin.*

Remark. Let Φ be a stationary point process. From the definiton of stationarity, it is clear that the intensity measure Λ of Φ must be translation invariant. It is know that the d-dimensional Lebesgue measure is (up to the multiplication by a constant) the unique translation invariant measure on $\mathbb{R}^d$, see Lemma 1.29 in Kallenberg [2002]. Thus, the intensity measure Λ of a stationary point process Φ is the d-dimensional Lebesgue measure multiplied by a positive constant λ. Therefore, stationary point process is also homogeneous with intensity λ.

1.2 Examples of point process models

As we have mentioned before point procesess serve to model the arrangement of objects that are randomly placed in space. In this section we will mention some well-known models for spatial point patterns. We will first state a definition of binomial and Poisson point processes which are used in case that there are no interactions among modeled objects. Then we will introduce Thomas point procees as an example of models for clustered patterns and Matérn hard-core process of type II which can be used when modelling regular data.

Definition 7. *Let ν be a diffuse measure on $\mathfrak{B}^d$, i.e $\nu(\{y\}) = 0$ for all $y \in \mathbb{R}^d$. Fix $n \in \mathbb{N}$ and take $B \in \mathcal{B}^d$ such that $0 < \nu(B) < \infty$. Let $X_1, X_2, \ldots, X_n$ be independent identically distributed random elements with values in B and suppose that*

$$\mathbb{P}[X_1 \in A] = \frac{\nu(A)}{\nu(B)}, \quad A \subseteq B,\ A \in \mathfrak{B}^d.$$

Then $\Phi = \{X_1, X_2, \ldots, X_n\}$ is called a binomial point process in B (with n points and a measure ν).

Remark. We want ν to be diffuse to ensure that the binomial point process will be simple.

It is easy to see that in this case random variable $N_\Phi(A)$ has binomial distribution with parameters n and $\frac{\nu(A \cap B)}{\nu(B)}$ for all $A \in \mathfrak{B}^2$. Three different realisations of a planar binomial point process can be seen in Figure 1.1.

Definition 8. *A measure μ on $\left(\mathbb{R}^d, \mathfrak{B}^d\right)$ is said to be locally finite if $\mu(K) < \infty$ for each $K \subset \mathbb{R}^d$ a compact set.*

Definition 9. *Let Λ be a locally finite diffuse measure on $\mathfrak{B}^d$. A planar point process Φ is said to be (simple) Poisson point process with intensity measure Λ if the following properties are satisfied:*

(a) for any $B \in \mathfrak{B}_0^d$ the random variable $N_\Phi(B)$ has the Poisson distribution with parameter $\Lambda(B)$ (if $\Lambda(B) = 0$ then $N_\Phi(B) = 0$),

(b) $N_\Phi(B_1), N_\Phi(B_2), \ldots, N_\Phi(B_n)$ are independent random variables for all $n \in \mathbb{N}$ and $B_1, B_2, \ldots, B_n \in \mathfrak{B}_0^d$ pairwise disjoint.

Remark. Our assumpltions that ν is locally finite and diffuse ensures existence and uniqueness of Poisson point process, see Proposition 9.2.III and Corollary 9.2.VIII in Daley and Vere-Jones [2008].

Remark. A homogeneous Poisson process is stationary and isotropic, this observation is discussed in Section 2.3.2 of Illian et al. [2004].

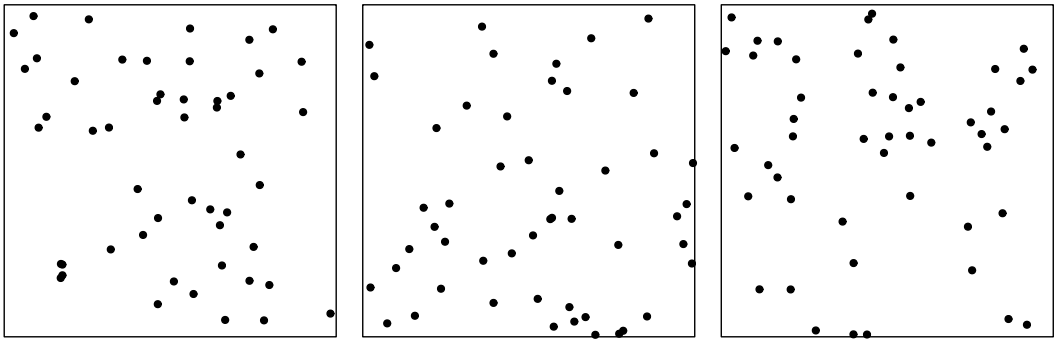

Figure 1.1: Three different realisations of planar binomial point process with 50 uniformly distributed points on unit square.

Poisson point process is a very fundamental point process model. It plays the role of a referential object in point pattern analysis and there exist more complex models that are constructed from the Poisson process. Homogeneous Poisson process represents the hypothesis of complete spatial randomness, i.e. the situation when there are no interactions between points. Three realisations of a planar stationary Poison point process with differen intensities can be seen in Figure 1.2.

Another natural situation is when the observed objects form divers clusters which are randomly placed in space. Imagine for example that we want to model locations of young plants dispersed around the adult ones. It is clear that Poisson point process is not a good model for this situation. We will need a finer model that captures the interactions between the adult "parent" plant and its young "daughter" plants. Definitions and notation concerning cluster point processes is taken from Chiu et al. [2013].

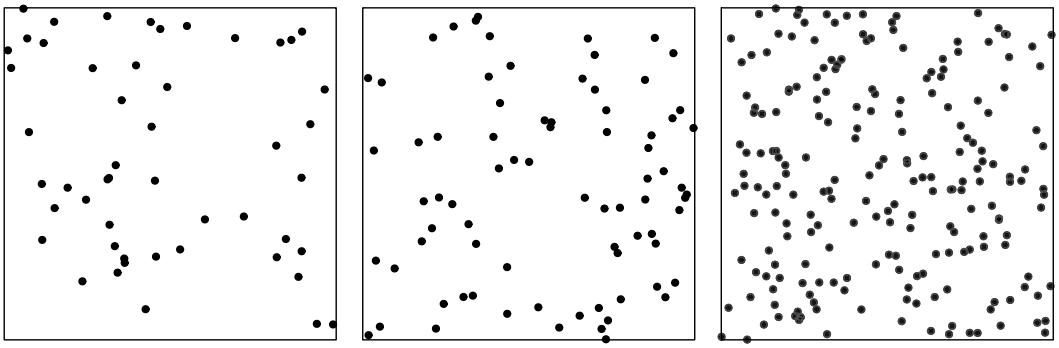

Figure 1.2: Three realisations of a planar stationary Poisson point process with intensity λ equal to 50, 100 and 200. Numbers of observed points are 50, 77 and 211, respectively.

Definition 10. *Let us have a point proces Φ_p (the so-called parent point process) and a collection of finite point processes $\left\{\Theta_x : \ x \in \mathbb{R}^d\right\}$. Finite means that for all $x \in \mathbb{R}^d$ $N_{\Theta_x}(\mathbb{R}^d)$ is almost surely finite. Take*

$$\Phi = \bigcup_{X \in \Phi_p} \Theta_X.$$

If $N_\Phi(B) < \infty$ almost surely for all $B \in \mathfrak{B}_0^d$ then Φ is called a cluster point process. For every $X \in \Phi_p$ the process Θ_X is said to be the daughter process corresponding to the parent point X.

Definition 11. *Cluster point process Φ such that $\left\{\Theta_x : \ x \in \mathbb{R}^d\right\}$ are mutually independent and independent on the parent point process Φ_p is called cluster point process with independent clusters. If Φ_p is moreover a Poinsson process, the Φ is said to be Poisson cluster point process.*

Definition 12. *Let us have a probability density p on $\mathbb{R}^d$. Assume that Φ is a Poisson cluster process such that:*

- *$\Theta_x(\mathbb{R}^d), \ x \in \mathbb{R}^d$, are independent identically distributed random variables,*
- *for all $x \in \mathbb{R}^d$ Θ_x is formed by a random number of independent identically distributed random vectors with probability density $p(\bullet - x)$.*

Then Φ is called Neyman-Scott process. If moreover $\Theta_x(\mathbb{R}^d)$ have Poisson distribution with parameter $\gamma > 0$, then Φ is called Neyman-Scott Poisson point process.

Remark. Suppose that Φ is a Neyman-Scott Poisson point proces. Denote by κ the intensity of the underlying Poisson point process Φ_p. Then Φ is a stationary point process with intensity $\lambda = \gamma\kappa$, see Section 5.3 in Møller and Waagepetersen [2004]. Constant γ corresponds to the expected number of points per cluster.

Definition 12 says that parent points form a stationary Poisson point process with intensity κ and a random number of daugter points in every cluster is scattered independently and with identical distribution around the parent point. Note that in the resulting process, we do not observe the parent points. Neyman-Scott processes were introduced in Neyman and Scott [1958] in order to model patterns formed by the locations of galaxies in space. Of course they can be used to model many others natural phenomena such as trees in a forest (for more details see Section 6.3.2 in Illian et al. [2004]). Let us now define a special type of Neyman-Scott Poisson processes, so-called Thomas process.

Definition 13. *If p in Definition 12 is given by*

$$p(y) = \frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}} \exp\left\{-\frac{\|y\|^2}{2\sigma^2}\right\}, \quad y \in \mathbb{R}^d,$$

then Φ is called Thomas point process.

Remark. Note that p is a density of centered multivariate normal distribution with the covariance matrix $\sigma^2 I_d$, where I_d denotes the d-dimensional identity matrix.

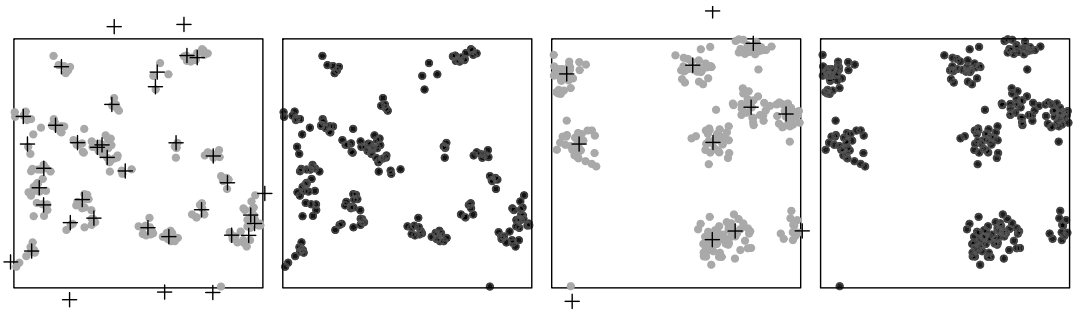

Figure 1.3: Two realisations of planar Thomas point process with different parametrs. Two left pictures: daughter points (grey circles) and the underlying parent points (black crosses), daughter points form a realisation of a Thomas process (black circles) with parameters $\kappa = 40$, $\sigma = 0.02$ and the expected number of points per cluster equal to 6. The same situation is in the two right pictures, only the daughter points form a realisation of a Thomas process with parameters $\kappa = 8$, $\sigma = 0.04$ and the expected number of points per cluster equal to 30.

Figure number 1.3 shows two different realisations of Thomas point process. We should realize that some of the parent points can lie outside the observation window even though we can see their daughter points in the resulting pattern. Also sometimes clusters that corresponds to different parent points may overlap. Therefore it is sometimes imposible to decide which daughter points match with one particular parent point.

Third situation that we would like to cover includes point patterns in which there are no pairs of points that are closer to each other than a specific minimum distance r_0. Such situations are modeled by so-called hard-core point processes. Hard-core point patterns arise when representing centres of non-overlapping objects, typically circles or spheres with radius $R \leq \frac{r_0}{2}$.

Hard-core processes provide a typical examples of processes with a tendency towards regularity, which is caused by repulsive interactions among the points of the process. As it is said in Section 6.5 in Illian et al. [2004], there are two main types of hard-core processes: processes resulting from interaction of hard objects (objects represented by the pattern are hard and non-penetrable, hence they cannot be closer together than permitted by their sizes) and processes resulting from thinning operations (these oparations will with given probability remove or retain every point of a process).

Definition 14. *Let Φ_p be a stationary Poisson point process with intensity $\kappa > 0$. We equip each point $X \in \Phi_p$ independently with a mark $m(X)$ which is a random number uniformly distributed in $(0, 1)$. Fix $r > 0$. For each $X \in \Phi_p$ denote by $b(X, r)$ the ball centred at X with radius r and put*

$$\beta_X = \mathbf{1}\left\{N\left(b(X,r) \cap \{Y \in \Phi_p : Y \neq X,\ m(Y) \leq m(X)\}\right) = 0\right\},$$

the indicator that $b(X, r)$ contains no points of Φ_p with marks smaller than $m(X)$. Then $\Phi = \{X \in \Phi_p : \beta_X = 1\}$ is called Matérn hard-core process of type II with hard-core distance r.

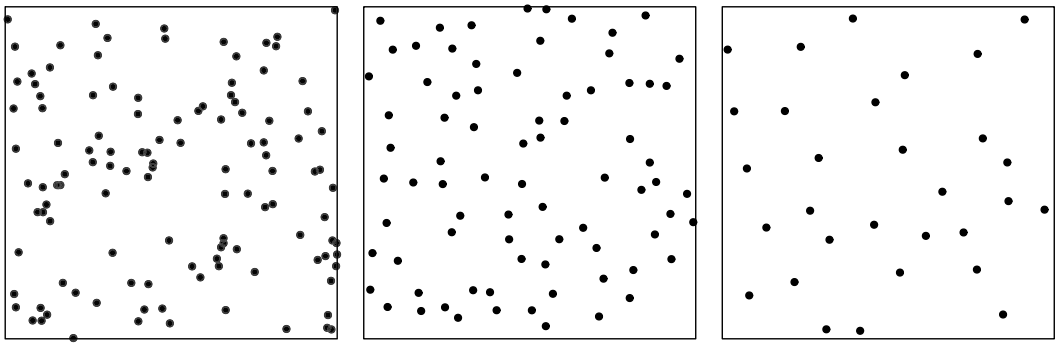

Figure 1.4: From left to right: realisation of stationary planar Poisson point process with intensity $\lambda = 100$, two realizations of Matérn hard-core process of type II with the intensity κ of the underlying Poisson point process equal to 100 and hard-core distance r equal to 0.05 and 0.1 respectively.

Remark. Matérn hard-core process of type II is a stationary point process with intensity

$$\lambda = \frac{1 - e^{-\kappa\omega_d r^d}}{\omega_d r^d},$$

where ω_d is the d-dimensional Lebesgue measure of the unit ball. Sketch of the proof can be found in Section 6.5.2 in Illian et al. [2004].

Definition 14 says that for every pair of distinct points $X, Y \in \Phi_p$ satisfying $0 < \|X - Y\| \leq r$ we delete the point with bigger mark. Two realisations of Matérn hard-core process of type II can be seen in Figure 1.4.

1.3 Summary characteristics and point pattern analysis

According to the authors of the book Illian et al. [2004], the aim of summary characteristics is to provide a brief and concise description of point patterns using numbers, functions or diagrams. They are used when analyzing properties of point process data (as the regular or clustered nature of the data) and they serve as a basis for parametric statistical approaches.

Let Φ be a spatial point process. We have already mentioned the random counts $N_\Phi(B)$, $B \in \mathfrak{B}^d$. Note that joint distribution of

$$N_\Phi(B_1), N_\Phi(B_2), \ldots, N_\Phi(B_m), \quad B_1, B_2, \ldots, B_m \in \mathfrak{B}_0^d,\ m \in \mathbb{N}$$

determines the distribution of Φ. Proof can be found in Appendix B of Møller and Waagepetersen [2004], Lemma B.2. The first and second order properties of these random variables are described by the intensity measure and the second order factorial moment measure defined as follows.

Definition 15. *The second-order factorial moment measure* $\alpha^{(2)}$ *on* $\mathfrak{B}^d \otimes \mathfrak{B}^d$ *is given by*

$$\alpha^{(2)}(C) \;= \sum_{X,Y\in\Phi}^{\neq} \mathbf{1}\left\{(X,Y)\in C\right\}, \quad C \in \mathfrak{B}^d \otimes \mathfrak{B}^d.$$

The symbol $\sum_{X,Y\in\Phi}^{\neq}$ *means that we sum trough all pairs of distinct points of* Φ.

Remark. Take $B \in \mathfrak{B}^d$. The second-order factorial moment measure $\alpha^{(2)}$ can be expressed as $\alpha^{(2)}(B\times B) = \mathbb{E}\left[N_\Phi(B)\left(N_\Phi(B)-1\right)\right]$. This formula follows from Equation (4.1) in Section 4.1.1 of Møller and Waagepetersen [2004] The second-order factorial moment measure is thus related with second-order factorial moment of N_Φ.

At the same time we can define the second-order product density in a very similar way as we did it with the intensity function.

Definition 16. *Suppose that the second-order factorial moment measure* $\alpha^{(2)}$ *can be written as*

$$\alpha^{(2)}(C) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{1}\left\{(x,y)\in C\right\}\; \lambda^{(2)}(x,y)\;\mathrm{d}x\;\mathrm{d}y, \quad C\in\mathfrak{B}^d\otimes\mathfrak{B}^d.$$

Then the function $\lambda^{(2)} : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow [0,\infty)$ *is called the second-order product density.*

Remark. Roughly speaking, $\lambda^{(2)}(x,y)\,\mathrm{d}x\,\mathrm{d}y$ is the probability that a pair of points of a point process Φ will occure jointly in each of two infinitesimally small balls with centres x, y and volumes $\mathrm{d}x$, $\mathrm{d}y$. Therefore, $\lambda^{(2)}$ contains information about interaction between pairs of points.

One of the first things to do in point pattern analysis is to study whether (and how) a point pattern deviates from a realization of Poisson point process. For this purpose it is useful to normalise the second-order product density $\lambda^{(2)}$.

Definition 17. *If both* λ *and* $\lambda^{(2)}$ *exist, then the pair correlation function* g *is given by the ratio*

$$g(x,y) = \frac{\lambda^{(2)}(x,y)}{\lambda(x)\lambda(y)}, \quad \textit{for all } x,y\in\mathbb{R}^d \textit{ such that } \lambda(x),\lambda(y)>0.$$

If $\lambda(x)$ *or* $\lambda(y)$ *equals zero, we set* $g(x,y)=0$.

Remark. The second-order product density $\lambda^{(2)}$ gives us some information about interactions between pairs of points of the process. At the same time the information can be distorted by the intensity function λ. We can always ask if the fact that we see lot of points in one particular area is caused by interactions or high intensity function in this area. Hence we define the pair correlation function g as the normalised second-order product density in order to obtain better information about interactions (free of the effect of the intensity function). For this reason we can use the pair correlation function to compare character of interactions between pairs of points for two point processes with diferent intensity functions.

It is know that for Poisson point process $g(x,y) \equiv 1$ (see Section 4.1.1. in Møller and Waagepetersen [2004]). Suppose that we have $g(x,y) > 1$ for some point process Φ. It means that pairs of points of Φ are more likely to appear jointly at the locations x, y than for a Poisson point process with the same intensity function as Φ, see the discussion after Definition 4.4 in Møller and Waagepetersen [2004].

Definition 18. *We say that g is translation invariant, if $g(x,y) = \tilde{g}(x-y)$ for all $x, y \in \mathbb{R}^d$ and some $\tilde{g} : \mathbb{R}^d \longrightarrow [0,\infty)$.*

Remark. Definition 18 says that a function g of two variables x and y can be expressed as a function $\tilde{g}$ of the difference $x - y$. We will abuse the notation and denote by g also the function $\tilde{g}$ which describes how the pair correlation depends on differences.

Remark. If Φ is stationary, then g is translation invariant, see Section 4.1.1 in Møller and Waagepetersen [2004]. At the same time there exist cases where g is translation invariant but Φ is an inhomogeneous process, we will se an example later. If Φ is isotropic, then $g(x,y)$ is a function of the distance from x to y, see discussion in Section 4.3.1 in Illian et al. [2004]. Again we abuse the notation and write $g(x,y) = g(\|x-y\|)$.

Suppose that Φ is a point process with translation invariant pair correlation function g. Then for $r > 0$ $g(r) > 1$ reveals clustering on a scale r and $g(r) < 1$ indicates regularity. If Φ is a hard-core process with the hard-core distance r_0, then $g(r) = 0$ for all $r \leq r_0$. For more details see Section 4.3.1 in Illian et al. [2004]. We will now introduce the class of second-order intensity reweighted stationary point processes. It includes for example every Poisson point process with intensity function λ (see [Baddeley et al. , 2000]).

Definition 19. *Suppose that Φ has the intensity function λ. Fix $A \in \mathfrak{B}^d$ such that $0 < |A| < \infty$, where $|A|$ denotes the d-dimensional Lebesgue measure of A. We define a measure $\mathcal{K}_A$ by*

$$\mathcal{K}_A(B) = \frac{1}{|A|}\mathbb{E}\sum_{X,Y\in\Phi}^{\neq} \frac{\mathbf{1}\{X \in A,\ Y - X \in B\}}{\lambda(X)\lambda(Y)}, \quad B \in \mathfrak{B}^d.$$

We set $\dfrac{\mathbf{1}\{X \in A,\ Y - X \in B\}}{\lambda(X)\lambda(Y)} = 0$ *if* $\lambda(X)$ *or* $\lambda(Y)$ *is equal to 0.*

Definition 20. *Let Φ be a point process with intensity function λ. If the measure $\mathcal{K}_A$ from definition 19 does not depend on the choice of A, then Φ is called second-order intensity reweighted stationary (SOIRS). $\mathcal{K}_A$ is in this case denoted by $\mathcal{K}$ and it is called the second-order reduced moment measure.*

Remark. Stationarity of Φ implies that Φ is also second-order intensity reweighted stationary, see Section 4.1.2 in Møller and Waagepetersen [2004].

If the pair correlation function exists and is translation invariant, then Φ is second-order intensity reweighted and $\mathcal{K}$ can be represented as follows:

$$\mathcal{K}(B) = \int_B g(z)\ \mathrm{d}z, \quad B \in \mathfrak{B}^d.$$

For the sketch of the proof see Section 4.1.2 in Møller and Waagepetersen [2004]. Once we have defined the second-order reduced moment measure, we are able to introduce two of the most popular point process summary characteristics: K and L-function.

Definition 21. *Let Φ be a second-order intensity reweighted stationary point process. Then the K-function is given by*

$$K(r) = \mathcal{K}\left(b(o,r)\right) = \int_{b(o,r)} g(z)\ \mathrm{d}z, \quad r > 0,$$

where $b(o,r)$ denotes the ball centered at the origin with radius r.

Remark. Definition 21 is an extension of the definition of Ripley's K-function (see Ripley [1976]) and originally comes from the paper Baddeley et al. [2000].

Definition 22. *Let us have a second-order intensity reweighted stationary point process Φ. L-function is then defined by*

$$L(r) = \left(\frac{K(r)}{\omega_d}\right)^{\frac{1}{d}}, \quad r > 0.$$

If we have a stationary point process Φ with intensity function λ, the quantity $\lambda K(r)$ can be interpreted as the expected number of further points of the process Φ within distance r from the origin given that Φ has a point at the origin. It is known that for the Poisson point process $K(r) = \omega_d r^d$. There is one to one correspondence between K and L-function and in the applications, L-function is used more often. It is due to the fact that for homogeneous Poisson process, the transformation from K to L is variance stabilising when K is estimated by nonparametric methods, see Section 4.2.1 in Møller and Waagepetersen [2004]. Also, for Poisson point process L-function becomes the identity, i.e. $L(r) = r$ for all $r > 0$. We can again compare the L-function of an arbitrary point process with the one of the Poisson point process. Bigger values (for some $r > 0$) suggests aggregation or clustering (at distances less than r), smaller values indicates regularity. Again you can see the discussion in Section 4.2.1 in Møller and Waagepetersen [2004].

Last but not least summary characteristics that we want to mention are based on interpoint distances. Namely it is the spherical contact distribution function and the nearest-neighbour distance distribution function.

Definition 23. *Assume that Φ is stationary. Then*

$$F(r) = \mathbb{P}\left[N_\Phi\left(b(o,r)\right) > 0\right], \quad r > 0,$$

is called the spherical contact distribution function.

In fact the spherical contact distribution function in nothing else than the distribution function of the distance from the origin (or, due to stationarity, any other fixed point in $\mathbb{R}^d$) to the nearest point of the process Φ. When analyzing point patterns we can go a bit further. Sometimes it can be useful to ask how does the (empirical) distribution function of the distances from the origin to the second or third or even k-th nearest point looks like. Now we will focuse on the distances between points of the process Φ.

Definition 24. *Let us have a stationary point process Φ with intensity λ. Then the nearest-neighbour distance distribution function D is given by*

$$D(r) = \frac{1}{\lambda|A|}\mathbb{E}\sum_{X\in\Phi\cap A}\mathbf{1}_{\{N(\Phi\setminus\{X\}\cap b(X,r))>0\}}, \quad r>0,$$

where A is an arbitrary Borel set satisfying $0<|A|<\infty$.

Under the assuption of stationarity, D can be interpreted as the distribution function of the distance from the "typical" point of the process to its nearest neighbour. The word typical refers to the Palm distribution and reduced Palm distribution of a point process. We will not give here the exact definition but it can be found in Appendix C of the book Møller and Waagepetersen [2004]. Again we can be interested in the (empirical) distribution function of the distances to the second or k-th neighbour. For the stationary Poisson point proces with finite intensity λ the explicit formula for F and D is known. So let Φ be a stationary Poisson point proces with finite intensity λ. Then

$$F(r) = D(r) = 1 - \exp\left\{-\lambda\omega_d r^d\right\}, \quad r>0.$$

It is possible to define another characteristic, so-called J-function, as the ratio

$$J(r) = \frac{1-D(r)}{1-F(r)}, \quad r>0.$$

Then obviously $J\equiv 1$ for the Poisson point process and at least for small values of r it holds that $J(r)<1$ implies agregation or clustering whereas $J(r)>1$ implies regularity. All of the stated facts can be found in Section 4.2.3 of the book Møller and Waagepetersen [2004].

1.4 Nonparametric estimation

Take $A\in\mathfrak{B}^d$. The d-dimensional Lebesgue measure of A wil be denoted as $|A|$. In the next two chapters we will have one single point pattern ζ on a bounded observation window $W\in\mathfrak{B}_0^d,\ 0<|W|<\infty$. Thus we will need some estimators for summary characteristics mentioned above. The observation window W is bounded and hence we will have to deal with edge effects. Imagine that for each point $x\in\zeta$ we want to measure the distance to its nearest neighbour. What if x lies very close to the boundary of W? It it possible that the nearest neighbour of x lies outside the observation widnow, hence we have no information about it. In fact we will take the distance from x to the nearest point from ζ that lies in W. But it is possibly not the right distance. This situation is illustrated

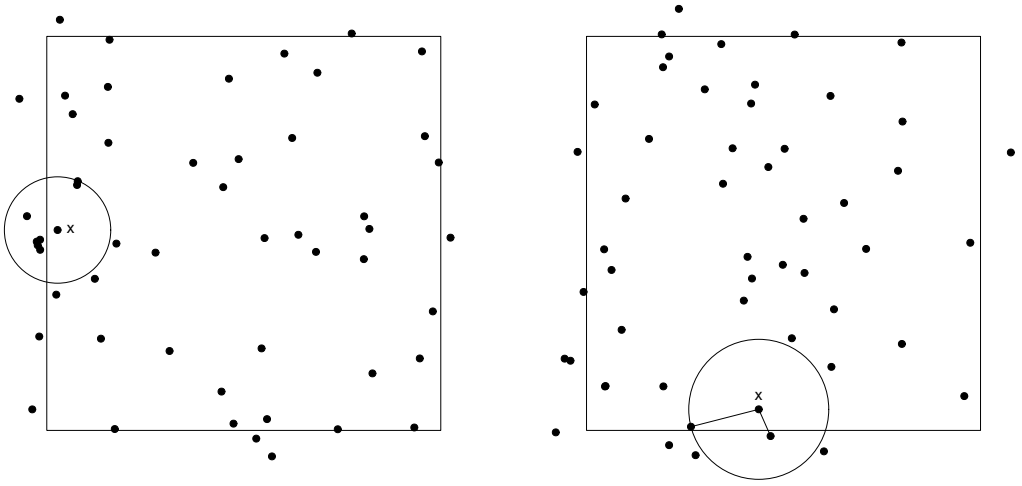

Figure 1.5: Illustration of the phenomenon called edge effects. Left: only 2 out of 7 points that hit the disc centered at point x with some radius r lie in the observation window. Right: the nearest neighbour of the point x lies outside the observation window W, if we take the nearest neighbour lying in W, this distance will be much bigger than the true one.

in Figure number 1.5.

It is important to realize that ignoring edge effects can lead to distorted conclusion about the investigated point process characteristic. There exist several methods how to deal with this phenomenon and how to reduce the effect of wrongly evaluated quantities. We will suggest one estimator for each of the characteristic λ, g, K, F and D. These particular estimators will be used in Chapter 3.

Suppose that Φ is a stationary point process with intensity λ. Suppose that we observe Φ only on a bounded observation window W, $|W| > 0$. It follows from the definition 5 that $\widehat{\lambda} = \dfrac{N_\Phi(W)}{|W|}$ is an unbiased estimator of λ.

For an inhomogeneous process Φ we use so-called non-parametric kernel estimators. Take $b > 0$ and u_b a kernel function with bandwidth b. It means that $u_b(x) = \frac{u(\frac{x}{b})}{b^d}$, where u is some probability density. Id $d = 1$ we often take

$$u(x) = \frac{3}{4}\left(x-1\right)^2, \quad x \in [-1,1]\,,$$

the Epanechnikov kernel. Other common choices are the density functions of uniform distribution on a ball or the Gaussian distribution. Put

$$C_{W,b}(x) = \int_W u_b(x-y)\ \mathrm{d}y$$

the correction on edge effects. Then the estimator of λ is given by

$$\widehat{\lambda}(x) = \frac{1}{C_{W,b}(x)} \sum_{Y \in \Phi \cap W} u_b(x-Y), \quad x \in W. \tag{1.1}$$

If u is symmetric then the following holds:

$$\mathbb{E}\int_W \widehat{\lambda}(x)\,\mathrm{d}x = \int_W \lambda(x)\,\mathrm{d}x.$$

It can be prooved in a very similar way as Lemma 4.1 in Møller and Waagepetersen [2004] (authors use a slightely different version of the estimator (1.1)).

Suppose that Φ is stationary and isotropic. Then the kernel estimator of the pair correlation function g is given by

$$\widehat{g}(r) = \frac{1}{\widehat{\lambda^2}} \sum_{X,Y\in\Phi\cap W}^{\neq} \frac{u_b\left(r - \|X-Y\|\right)}{\sigma_d r^{d-1}|W|} e_{W,r}(X,Y), \tag{1.2}$$

where u_b is a suitable kernel function, σ_d is the surface of a unit ball in $\mathbb{R}^d$ and $e_{W,r}(X,Y)$ is the so-calle edge correction factor. We use the translation edge correction factor in a form

$$e_{W,r}(X,Y) = \frac{|W|}{|W\cap(W+X-Y)\,|}.$$

Translation edge correction factor is not the only possibility, we can use for example the Ripley's isotropic edge correction or the minus sampling. For more details see Section 4.2.2 in Illian et al. [2004]. Note that the common estimator of $\widehat{\lambda^2}$ is $\frac{\Phi(W)(N_\Phi(W)-1)}{|W|}$ instead of $\left(\frac{N_\Phi(W)}{|W|}\right)^2$. In the case of inhomogeneous SOIRS process we replace $\widehat{\lambda^2}$ by $\widehat{\lambda(X)}\widehat{\lambda(Y)}$ in the denominator of each summand.

Suppose that Φ is an SOIRS process. Then the estimator of K-function with translation edge correction factor has the form

$$\widehat{K}(r) = \sum_{X,Y\in\Phi\cap W}^{\neq} \frac{\mathbf{1}\left\{\|X-Y\|\le r\right\}}{\widehat{\lambda(X)}\widehat{\lambda(Y)}|W\cap(W+X-Y)\,|}. \tag{1.3}$$

The estimate (1.3) is biased as well as the estimate of $L(r)$ which is obtained from transforming (1.3). We say that an estimate is ratio-unbiased, if it is in the form $\widehat{\theta} = Y/Z$ where $\theta = \mathbb{E}Y/\mathbb{E}Z$. If Φ is stationary then $\widehat{\lambda(X)}\widehat{\lambda(Y)} = \widehat{\lambda^2}$ for all $X,Y\in\Phi\cap W$. If $\widehat{\lambda^2}$ is an unbiased estimator of λ^2, then (1.3) is ratio-unbiased, see Section 4.3.2. in Møller and Waagepetersen [2004].

For the spherical contact distribution function and the nearest-neighbour distance distribution function we will present so-called raw estimate. There are other possibilities as reduced-sample or Kaplan-Meier method, see Section 4.3.6 in Møller and Waagepetersen [2004]. Our choice of none edge correction will be commented in Section 3.1. Suppose that Φ is stationary. Take $X\in\Phi$. Set $e(X) = d\left(X, \Phi\setminus\{X\}\cap W\right)$ the distance from X to its nearest neighbour in $\Phi\cap W$. The estimate is then given by

$$\widehat{D}(r) = \frac{1}{N_\Phi(W)} \sum_{X\in\Phi\cap W} \mathbf{1}\left\{e(X)\le r\right\}.$$

Now chose a regular grid I_a in $\mathbb{R}^d$, i.e.

$$I_a = y + a\mathbb{Z}^d = \left\{(y_1 + a_1 z_1, \ldots, y_d + a_d z_d) \in \mathbb{R}^d : \ z_i \in \mathbb{Z}\right\},$$

where $y = (y_1, y_2, \ldots, y_d) \in \mathbb{R}^d$ and $a = (a_1, a_2, \ldots, a_d) \in \mathbb{R}^d$ such that $a_i > 0$ for all $i \in \{1, 2, \ldots, d\}$. The estimator of F is then given by

$$\widehat{F}(r) = \frac{1}{N(I_a \cap W)} \sum_{x \in I_a \cap W} \mathbf{1}\left\{d(x, \Phi \cap W) \leq r\right\},$$

where $d(x, \Phi \cap W)$ denotes the distance from x to the nearest point of the process Φ lying in W.

In Chapter 3 we will use $\widehat{D}_k$ the empirical distribution function of the distances to the k-the nearest neighbour and $\widehat{F}_k$ the empirical distrubution function of the distances from a fixed set of points to the nearest point of the process. Let us now denote by $e_k(X)$ the distance of the point $X \in \Phi$ to its k-th nearest neighbour in $\Phi \cap W$. Then we will compute $\widehat{D}_k$ as follows:

$$\widehat{D}_k(r) = \frac{1}{N_\Phi(W)} \sum_{X \in \Phi \cap W} \mathbf{1}\left\{e_k(X) \leq r\right\}.$$

Similarly, let use denote by $d_k(x, \Phi \cap W)$ the distance from a fixed point x to the k-th nearest point of the process Φ lying in W. Then $\widehat{F}_k$ is given by

$$\widehat{F}_k(r) = \frac{1}{N(I_a \cap W)} \sum_{x \in I_a \cap W} \mathbf{1}\left\{d_k(x, \Phi \cap W) \leq r\right\}.$$

2. Stochastic reconstruction

In this chapter we will focuse on an algorithmic procedure described in Tscheschel and Stoyan [2006]. Even though the authors call it statistical reconstruction, concerning this work we will call the procedure stochastic reconstruction or stochastic reconstruction algorithm. It is due to the fact that the term stochastic reconstruction is commonly used in the literature when talking about this procedure. Before we start with description and explanations, let us talk for a while about point pattern analysis.

When analysing point pattern data, various hypotheses can be tested using summary characteristics introduced in Secction 1.3. Often the distribution of the test statistics (i.e. the estimator of some summary characteristic) is very complicated or untracable. For this reason simulation-based tests are needed. In this text we are primarily interested in situations when we are not able to simulate from the null model (the null hypothesis is not specific enough). Example of such situation was given in Introduction (testing independence of two diverse populations). In this context, the stochastic reconstruction algorithm can be used instead of simulations to generate independent point configurations with the same form of selected summary characteristics as the data.

2.1 The algorithm

Authors of the paper Tscheschel and Stoyan [2006] mention that the stochastic reconstruction has some tradition in statistical physics, where it is used when simulating random closed sets. More details can be found in Chapter 12 of Torquato [2002]. The procedure was for the first time adapted for spatial point processes in the diploma thesis of André Tscheschel (see Tscheschel [2001]). Another variant of the stochastic reconstruction approach can be found in Pommerening [2006], where it is discussed in a forestry context. In the next paragraphs we are going to describe the algorithm as it is presented in Tscheschel and Stoyan [2006].

Suppose we have a bounded observation window W such that $0 < |W| < \infty$. Let us have φ a point pattern observed in W. We suppose that φ is a realisation of a stationary point process Φ. Take I and J natural numbers. Let $n_i,\ i = 1, 2, \ldots, I$ be some numerical summary characteristics, for example the intensity, Pielou's index of randomness or Clark-Evans index. The last two can be found in Section 4.2.4 of the book Illian et al. [2004]. Similarly, let $f_j,\ j = 1, 2, \ldots, J$ be some functional summary characteristics. We have seen examples of functional summary characteristics in Section 1.3. To be more concrete we can mention the L-function or the nearest-neighbour distance distribution function D. Note that for every f_j we have to choose a constant R_j which depends on the observation window W and which limits the domain of f_j. It means that we are interested in the values $f_j(r),\ 0 < r \leq R_j,\ j = 1, 2, \ldots, J$. By $\widehat{n}_i(\varphi)$ and $\widehat{f}_j(\varphi, r)$ we denote empirical estimators of n_i and $f_j(r)$ for the observed point pattern φ. Examples of estimators for intensity and various functional summary characteristics (with particular edge correction) were given

in Section 1.4.

The main idea of the algorithm is to generate a point configuration ζ in the observation window W which has the estimated characteristics $\widehat{n}_i(\zeta),\ i = 1, 2, \ldots, I$ and $f_j(\zeta, r),\ j = 1, 2, \ldots, J$ similar to the observed point pattern φ. We emphasize that the algorithm is designed in such a way that all generated point configurations have the same number of points as the observed pattern φ which is reconstructed. In detail, we want to create a point configuration ζ such that $N(\zeta \cap W) = N(\varphi \cap W)$ and

$$\begin{aligned} & & \widehat{n}_i(\zeta) &\approx \widehat{n}_i(\varphi),\ i = 1, 2, \ldots, I, \\ \text{for all } 0 < r \le R_j & & \widehat{f}_j(\zeta, r) &\approx \widehat{f}_j(\varphi, r),\ j = 1, 2, \ldots, J. \end{aligned}$$

The "similarity" of two point configurations φ and ζ, i.e. the deviation of estimates of a collection of numerical and functional summary characteristics, is measured by a functional E which we will call the energy functional. Let us define

$$\begin{aligned} E_{n_i}(\varphi, \zeta) &= \left[\widehat{n}_i(\varphi) - \widehat{n}_i(\zeta)\right]^2,\ i = 1, 2, \ldots, I, \\ E_{f_j}(\varphi, \zeta) &= \int_0^{R_j} \left[\widehat{f}_j(\varphi, r) - \widehat{f}_j(\zeta, r)\right]^2 \mathrm{d}r,\ j = 1, 2, \ldots, J. \end{aligned}$$

Then the energy functional E is given by the weighted sum

$$E(\varphi, \zeta) = \sum_{i=1}^{I} w_{n_i} E_{n_i}(\varphi, \zeta) + \sum_{j=1}^{J} w_{f_j} E_{f_j}(\varphi, \zeta), \tag{2.1}$$

where $w_{n_i},\ i = 1, 2, \ldots, I$ and $w_{f_j},\ j = 1, 2, \ldots, J$ are positive real numbers.

Example. Suppose that we have observed a planar point pattern φ in the observation window $W = [0, 1]^2$. Based on our knowledge about the data, we have decided that the pair correlation function g reveals the property of φ that is of particular interest. We hawe also decided that we do not want to use any other characteristics in the energy functional E. Thus $I = 0,\ J = 1,\ f_1 = g$, $R_1 = 0.25$, $w_g = 1$ and for any ζ point configuration in W we have

$$E(\varphi, \zeta) = E_g(\varphi, \zeta) = \int_0^{0.25} \left[\widehat{g}(\varphi, r) - \widehat{g}(\zeta, r)\right]^2 \mathrm{d}r.$$

The weight w_g can be choosen arbitrarily and for the choice of R_1 we follow the common recommendation of the one fourth of the shortest side of the observation window.

Let us now make few remarks about the energy functional. The weights w_{n_i} and w_{f_j} could be used to give different importance to the individual terms in (2.1). Note that different terms may take values on different scales and thus the weights may serve to ensure that each of the terms E_{n_i}, E_{f_j} contributes to the final sum by values on the same scale. Usually the weights are chosen experimentally. Imagine we want to base the neregy functional on two different terms. First we will run the algorithm based only on the first term to see the scale of the values. Then we will do the same for the second term. Once

we have did these experimental runs we are able to compare the scale of values of the first and the second term and thus we are able to chose the weights appropriately.

Also, it is possible to modify the term E_{f_j} by adding weights $u_j(r)$ for all of the admissible values of the parameter r. The term hence becomes

$$E_{f_j} = \int_0^{R_j} u_j(r) \left[\widehat{f}_j(\varphi, r) - \widehat{f}_j(\zeta, r)\right]^2 \, \mathrm{d}r.$$

This modification can be useful when we have a reason to think that for some values of r the estimate $\widehat{f}_j$ is unreliable. Described situation may arise for example when we are estimating the pair correlation function g with r close to zero or when we want to compensate the increasing variability of estimator of K-function. Summary characteristics in the energy functional E are chosen based on our information about the data and the purpose for which the reconstructed patterns would be used. For example, if we want to use a simulation-based test to test the null hypothesis that φ is a realisation of a stationary Poisson point process, the test statistic could be L-function (because for the Poisson point process we know the analytic formula for L-function). Then the energy functional should not be based on L-function.

Once we have described the energy functional E, we can start describing the algorithm itself. The observed point pattern φ will be sometimes called the input of the stochastic reconstruction algorithm. Suppose now that $N(\varphi \cap W) = n \in \mathbb{N}$, then the initial step of the algorithm is a realisation of a binomial point process on W with n points and a 2-dimensional Lebesgue measure. It will be denoted by $\zeta^{(0)}$. Suppose that after l iteration steps we have obtained point configuration $\zeta^{(l)}$ with n points. In the iteration step number $l+1$ we:

- randomly (with probability $\frac{1}{n}$) choose one point z from the configuration $\zeta^{(l)}$ to be deleted,
- generate a new point y uniformly (with respect to the Lebesgue measure) in W,
- denote by ζ^{new} the configuration $\zeta^{new} = \left(\zeta^{(l)} \setminus \{z\}\right) \cup \{y\}$,
- accept the configuration ζ^{new} to be $\zeta^{(l+1)}$ if and only if

$$E\left(\varphi, \zeta^{new}\right) \leq E\left(\varphi, \zeta^{(l)}\right),$$

- otherwise we set $\zeta^{(l+1)} = \zeta^{(l)}$.

Individual parts of each iteration step are executed independently and each iteration step is independet on the previous steps. The run of the algorithm is stopped when an exact number (apriori chosen) of iteration steps have been executed or when the energy functional $E\left(\varphi, \zeta^{(l)}\right)$ is small enough, i.e. smaller than some constant $\epsilon > 0$. It is also possible to stop the algorithm if the new configuration

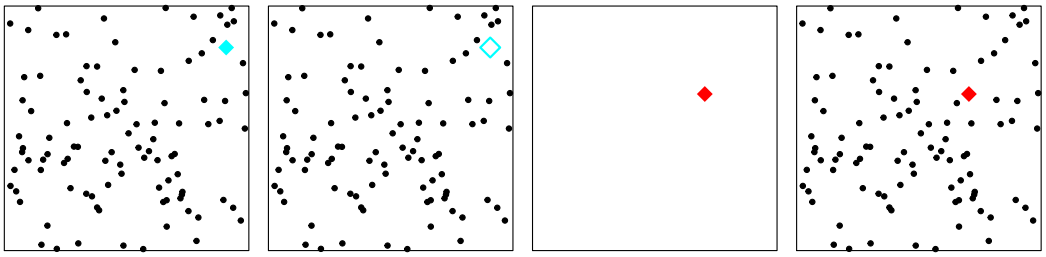

Figure 2.1: Illustration of an iteration step of the stochastic reconstruction algorithm: we randomly choose one point (blue diamond), choosen point is deleted, we generate a new point (red diamond) uniformly in the observation window, we compute the enrgy functional for the possible new configuration (blue dimond deletad and replaced by red one), if it is smaller than it was before (for the configuration with blue diamond) the deleted point is replaced by the new one.

ζ^{new} is not accepted for m times in a row, where $m \in \mathbb{N}$. Illustration of an iteration step ot the stochastic reconstruction algorithm can be found in Figure 2.1.

In other words, in every iteration step of the stochastic reconstruction algorithm we move one (randomly chosen) point of the current point configuration $\zeta^{(l)}$ to another place in the observation window W in order to get a new configuration with lower energy E. We emphasise that outputs of the stochastic reconstruction algorithm has always the same number of points as the input. The stopping rule should be choosen so that every point of the initial configuration $\zeta^{(0)}$ has a chance to be moved several times.

From the description of the algorithm it is clear that it is in fact an optimization problem. We want to minimize the energy functional $E(\varphi, \zeta)$ as a function of the point configuration ζ, φ is fixed and represents the input of the algorithm, i.e. the observed data. After a number of iteration steps we will end up in a local minimum of $E(\varphi, \bullet)$ and the output will be a point configuration with low energy. Examples of outputs of the stochastic reconstruction algorithm can be seen in Figure 2.2. Note that during the run of the algorithm it is impossible to leave the local minimum, because configurations with higher energy could not be accepted. However, this is not a big limitation. Take $t \in \mathbb{N}$. Suppose that we want to obtain t reconstructions (i.e. outputs ot the algorithm) of the point pattern φ. We will thus run the algorithm t times, each run will start from a different initial configuration $\zeta^{(0)}$ (different realisation of a binomial point process) and different runs may end up in different local minima. Evolution of values of the energy functional during the run of the algorithm can be seen in Figure 2.3.

It is important to realize that $E(\varphi, \varphi) = 0$ and thus there exists at least one global minimum, where the energy functional equals 0. Our aim is not to attain this particular global minimum, because we do not want to get copies of the observed point pattern. It is worth to remember that the stochastic reconstruction approach aims to provides independent replicates of the data, i.e. independent point patterns with low enrgy that come from various local (or possibly global,

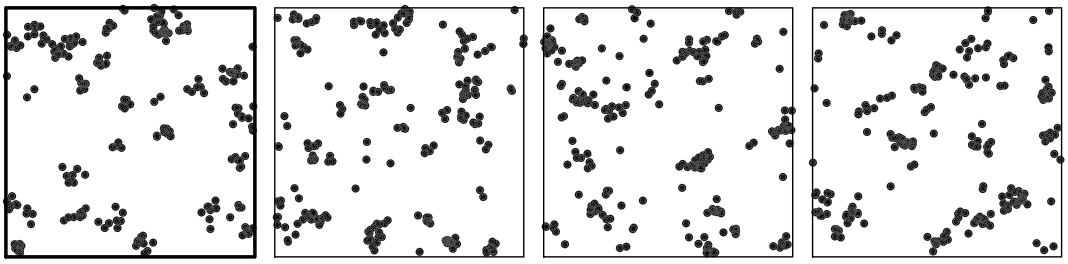

Figure 2.2: Illustration of the stochastic reconstruction approach (from left to right): input of the algorithm (a realization of a planar Thomas point process on the unit square), three different outputs (from three different runs of the algorithm). Energy functional is based only on the pair correlation function.

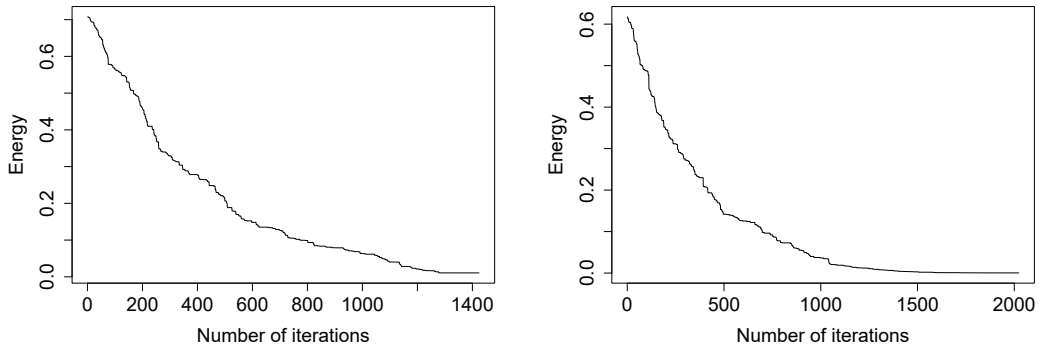

Figure 2.3: Values of the energy functional (based on pair correlation function) during the run of algorithm. Input is the same as in Figure 2.2. We stop the run if the new configuration is not accepted for 100 times in a row. Left picture corresponds to the first output in Figure 2.2, 1423 iterations were executed. Right picture corresponds to the third output in Figure 2.2 and 2022 interation were executed.

but different from φ) minima of $E(\varphi, \bullet)$. Another key point is that the use of this method does not require any explicit assumptions on the theoretical model for observed data. We work only with estimates of selected summary characteristics which depend on the observation window W and the edge correction method.

As we have seen in the previous paragraphs the crucial part of the stochastic reconstruction algorithm is the energy functional. Thus we should pay attention to the choice of summary characteristics used in the equation (2.1). We have already said that the choice should be based on our information about the data, but it is important to realize that some characteristics may contain more information about the data than others. For example, it is known that the K-function (and hence also the L-function) does not determine the distribution of a point process. In other words, there exist a number of point processes with different distributions and the same K-function, see Baddeley and Silverman [1984]. Therefore there could be a large variability among the results of the

reconstruction algorithm with the energy functional based only on L-function. In Tscheschel and Stoyan [2006] the authors suggest to use a combination of the k-th nearest-neighbour distance distribution functions instead. It is known that there is a relationship between K-function and the k-th nearest-neighbour distance distribution functions D_k. If we have a stationary point process with intensity κ, then

$$\kappa K(r) = \sum_{k=1}^{\infty} D_k(r), \quad r > 0,$$

see Section 4.3.1 in Illian et al. [2004]. Thus the sequence of all functions D_k contains more information then the K or L-function.

At the begining of this chapter we have mentioned that the outputs of stochastic reconstruction algorithm may be used instead of simulations when performing simulation-based tests. Equally important is the use of simulations when exploring the sampling variation of estimated summary characteristics. In this context, simulations can be also replaced by the stochastic reconstruction approach in order to construct confidence interval and bands for estimates of various summary charracteristics, even though this application does not appear in the literature. Moreover, Tscheschel and Stoyan [2006] suggest to use the algorithm for constructing neighbours of points lying close to the edges of the observation window. This idea leads to the procedure that Tseschel and Stoyan call conditional simulation - positions of points in some region are fixed while points in the other regions have to be reconstructed. In this text conditional simulation will be called conditional reconstruction. We will list some examples of the practical use of the stochastic reconstruction method in the next section.

2.2 Examples of use of the stochastic reconstruction method

Nowadays the method of stochastic reconstuction is of particular interest in biology and ecology. Note that references to the paper Tscheschel and Stoyan [2006] can be found in many specialized articles and only few of them will be cited in the next subsections.

2.2.1 Testing statistical hypotheses

We have already encountered the first example in the Introduction when testing the independence of populations of two related species of trees in Barro Colorado Island. This experiment is inspired by the paper Getzin et. al. [2014], where authors are interested in testing independence of locations of adult and recruit trees. Another example of use of the stochastic reconstruction algorithm when testing independence of two point patterns can be found in Mundo et al. [2013].

The hypothesis of independence of two point pattern is a typical example of hypothesis that is not specific enough to enable simulations. Another example could be the isotropy hypothesis. Wong and Chiu [2016] introduce a model-free isotropy test based on stochastic reconstruction approach. They affirm that

the null distribution of the test statistic is approximated by the empirical distribution obtained from bootstrap-type samples, i.e. outputs of the stochastic reconstruction algorithm. To ensure that the reconstructions correspond to realisations of an isotropic point process, only characteristics which do not take into account any directional properties are used in the energy functional.

2.2.2 Quasi-plus sampling edge correction method

Last example we would like to mention is the employment of the conditional reconstruction when dealing with edge effects. Tscheschel and Chiu [2008] introduce the quasi-plus samping edge correction method based on conditional reconstruction (we are able to reconstruct positions of points lying outside the observation window W conditioned on the points lying in W) and they compare it with other edge correction methods. The idea of conditional reconstruction outside the observation widnow is also used in applied papers concernig forest ecosystems monitoring. For more detail see Pommerening and Stoyan [2008] and Lilleleht et al. [2014].

3. Stochastic reconstruction of inhomogeneous point patterns

So far we have given a brief introduction to the theory of spatial point processes and point pattern analysis. Also, we have introduced stochastic reconstruction, an algorithmic procedure which can be used to provide independent replicates of point process data. In Chapter 2 we have supposed that we observe a realisation of some stationary point process on a bounded observation window. Main aim of this work is to discuss the possibility of extension of the stochastic reconstruction algorithm to the situation when we observe a realisation of an inhomogeneous point process. In this chapter we consider only point processes whose intensity function exists and it is a non-constant function.

Note that this text is not the first work concerning stochastic reconstruction for point processes with non-constant intensity function. In the paper Wiegand et al. [2013] authors modify the algorithm described in Section 2.1 in the following way:

- initial configuration $\zeta^{(0)}$ is a realisation of an inhomogeneous binomial point process (with probability measure ν whose density with respect to the Lebesgue measure is proportional to the estimated intensity function of the observed point pattern),
- new point y in the configuration ζ^{new} is also generated from the probability distribution ν.

We will show that these two modifications are not sufficient to produce reconstructions which actually correspond to the observed point pattern. Moreover, we will show that the main problem is the intensity function. We will also suggest a solution to this problem and we will discuss how efficient it is for different point process models (clustered, Poisson and regular).

3.1 The algorithm

Let us suppose that we observe a point pattern φ on a bounded observation window $W \in \mathfrak{B}_0^d$, $0 < |W| < \infty$. We assume that φ is a realisation of a spatial point process Φ with non-constant intensity function λ. Let us denote by $\widehat{\lambda}_\varphi$ the estimate (based on φ) of λ on W.

As we have mentioned in the second paragraph of this chapter, there are some obvious modifications of the stochastic reconstruction algorithm for inhomogeneous point patterns. Let ν_φ be a probability measure on W given by

$$\nu_\varphi(A) = \int_A \frac{1}{C_{W,\lambda}} \widehat{\lambda}_\varphi(x)\, \mathrm{d}x, \quad A \in \mathfrak{B}^d,\ A \subseteq W,$$

where

$$C_{W,\lambda} = \int_W \widehat{\lambda}_\varphi(x)\, \mathrm{d}x$$

is the normalising constant.

Remark. The intensity function λ is estimated in such a way that $C_{W,\lambda} < \infty$ for arbitrary φ.

Suppose that $N(\varphi \cap W) = n \in \mathbb{N}$. Then the initial step $\zeta^{(0)}$ of the algorithm from Section 2.1 is now a realisation of a binomial point process on W with n points and the measure ν_φ. It means that points of $\zeta^{(0)}$ are realisations of n independent random elements with probability distribution ν_φ. Suppose that after l iteration steps we have obtained point configuration $\zeta^{(l)}$ with n points. In the iteration step number $l+1$ we:

- randomly (with probability $\frac{1}{n}$) choose one point z from the configuration $\zeta^{(l)}$ to be deleted,
- generate a new point y in W from the probability distribution ν_φ,
- denote by ζ^{new} the configuration $\zeta^{new} = \left(\zeta^{(l)} \setminus \{z\}\right) \cup \{y\}$,
- accept the configuration ζ^{new} to be $\zeta^{(l+1)}$ if and only if

$$E(\varphi, \zeta^{new}) \leq E\left(\varphi, \zeta^{(l)}\right),$$

- otherwise we set $\zeta^{(l+1)} = \zeta^{(l)}$.

We emphasise that these two modifications are exactly the same as in Wiegand et al. [2013]. In the next sections we will show that we have to do more if we want to obtain "reasonable" results. What the word reasnable means will be explained in Section 3.2. For now let us say that these two modifications do not ensure that the estimated intensity function of reconstructions corresponds to the estimated intensity function of the observed point pattern. This algorithm will be sometimes refered to as the improvement-only algortihm.

Now we would like to discuss which summary characteristics are available for the energy functional E. First characteristic that we have in mind when talking about interactions among points of a process is the pair correlation function. Let us thus suppose that for the point process Φ the pair correlation function exists. If we moreover assume that it is translation invariant, we obtain that Φ is a SOIRS point process and we can use the inhomogeneous L-function. In the paper van Lieshout [2010] a definion of J-function for SOIRS point processes is given, but we are not going to use it. The main reason is that the interpretation of J function for SOIRS point processes is very complicated (in contrast to the stationary case) and our experiments with this characteristic in the software `R` showed that the variability of its estimator is too large and thus it is not a suitable characteristic when we want to describe precisely the properties of the observed point pattern that should be preserved during the reconstruction. So next to the inhomogeneous initial configuration and non-uniform distribution of the proposed points we may improve the algorithm by using inhomogeneous pair correlation function and inhomogeneous L-function. Recall that the estimators of these two characteristics take into account the

non-constant intensity function of the investigated point pattern. We emphasise that (until the end of this text) every time we speak about the pair correlation function and the L-function we mean their inhomogeneous counterparts (even if we do not say the word inhomogeneous explicitly).

When working with point process data, it is never guaranteed that the data comes from a model satisfying the SOIRS condition. We are convinced that it is useful to use the inhomogeneous pair correlation function and the inhomogeneous L-function even if the SOIRS property is not guaranteed. In such situations we use these characeristics as empirical and we do not give them the theoretical interpretation from Section 1.3. But they are still useful because they contain information about pairs of points of the observed point pattern with a special property (e.g. pairs of points that are closer together than some distance $r > 0$ for L-function) and moreover the influence of the non-constant intensity function is taken into account.

Recall that in Section 1.3 we have defined the nearest-neighbour distance distribution function D and the spherical contact distribution function F for stationary point processes. In the Section 2.1 we have mentioned that in the stationary context Tscheschel and Stoyan [2006] suggest to use a combination of the k-th nearest-neighbour distance distribution functions D_k in the energy functional. The main question is whether it makes sense to use these distribution functions when working with inhomogeneous point processes.

When dealing with inhomogeneous point patterns we will use the empirical distribution function of distances to the k-th nearest neighbour $\widehat{D}_k$ and the distribution function of the distances from a set of fixed points in the observation window W to the k-th nearest point of the process $\widehat{F}_k$. In the stationary context, $\widehat{D}_k$ is the empirical estimator of D_k. In the inhomogeneous case we use $\widehat{D}_k$ only as a useful empirical characteristic without claiming that it is an estimator of some theoretical characteristic. The same for $\widehat{F}_k$.

However, we think that it is still useful to work with such empirical characteristics because they can contain detailed information about the sructure (geometry) of observed point pattern. Additionally, if we combine only the empirical distribution functions in the energy functional, the choice of weights is very natural (every empirical distribution function will get the same weight). Thus the weights can be chosen before running the algorithm (if we want to combine the L-function or the pair correlation function with other characteristics we have to choose the weights experimentally). On the other hand we should not forget that they do not take into account the non-constant intensity function. In the papers van Lieshout and Baddeley [1996], Baddeley et al. [2000] and van Lieshout [2010] we can find inhomogeneous counterparts of the sperical contact distribution function F and the nearest-neighbour distance distribution function D. We do not use these characteristics because it is not obvious if they can be generalised for $k > 1$ and we do not wan to treat $\widehat{D}_k$ and $\widehat{F}_k$ differently for $k = 1$ and $k > 1$.

To estimate selected characteristics mentioned above we use the estimates given in Section 1.4. For $\widehat{D}_k$ and $\widehat{F}_k$ we use so-called raw estimates, i.e. we do not use any edge correction. If we use some of the classical edge correction methods, some properties of the empirical distribution function (e.g. the monotonicity or maximal value equal to 1) can be violated, see Section 4.2.6 in Illian et al. [2004]. Using uncorrected estimator does not cause any trouble since the observation window does not change during the run of the algorithm. Size and shape of the observation window have impact on the values of the raw estimates, it means that raw estimates of the same process computed on different observation windows may actually differ a lot. For this reason we should use the edge correction if we want to perform the conditional reconstruction. Note that the conditional reconstruction outside the observation window W is not possible (at least without additional assumptions on the intensity function) in the inhomogenous context since we do not have the information about the intensity function outside the observation window.

In the simulation studies, we will consider 8 different versions of the energy funtional:

$$
\begin{aligned}
E_1(\varphi,\zeta) &= \int_0^{R(L)} \left[\widehat{L}(\varphi,r) - \widehat{L}(\zeta,r)\right]^2 \,\mathrm{d}r,\\
E_2(\varphi,\zeta) &= \int_0^{R(g)} \left[\widehat{g}(\varphi,r) - \widehat{g}(\zeta,r)\right]^2 \,\mathrm{d}r,\\
E_3(\varphi,\zeta) &= \sum_{k=1}^{5} \left(\int_0^{R(F_k)} \left[\widehat{F}_k(\varphi,r) - \widehat{F}_k(\zeta,r)\right]^2 \,\mathrm{d}r \right),\\
E_4(\varphi,\zeta) &= \sum_{k=1}^{20} \left(\int_0^{R(F_k)} \left[\widehat{F}_k(\varphi,r) - \widehat{F}_k(\zeta,r)\right]^2 \,\mathrm{d}r \right),\\
E_5(\varphi,\zeta) &= \sum_{k=1}^{5} \left(\int_0^{R(D_k)} \left[\widehat{D}_k(\varphi,r) - \widehat{D}_k(\zeta,r)\right]^2 \,\mathrm{d}r \right),\\
E_6(\varphi,\zeta) &= \sum_{k=1}^{20} \left(\int_0^{R(D_k)} \left[\widehat{D}_k(\varphi,r) - \widehat{D}_k(\zeta,r)\right]^2 \,\mathrm{d}r \right),\\
E_7(\varphi,\zeta) &= \sum_{k=1}^{5} \left(\int_0^{R(F_k)} \left[\widehat{F}_k(\varphi,r) - \widehat{F}_k(\zeta,r)\right]^2 \,\mathrm{d}r \right) +\\
&\quad + \sum_{j=1}^{5} \left(\int_0^{R(D_j)} \left[\widehat{D}_j(\varphi,r) - \widehat{D}_j(\zeta,r)\right]^2 \,\mathrm{d}r \right),\\
E_8(\varphi,\zeta) &= \sum_{k=1}^{20} \left(\int_0^{R(F_k)} \left[\widehat{F}_k(\varphi,r) - \widehat{F}_k(\zeta,r)\right]^2 \,\mathrm{d}r \right) +\\
&\quad + \sum_{j=1}^{20} \left(\int_0^{R(D_j)} \left[\widehat{D}_j(\varphi,r) - \widehat{D}_j(\zeta,r)\right]^2 \,\mathrm{d}r \right).
\end{aligned}
$$

In the functionals E_3 - E_8 we have put all of the weights equal to 1. It is because we want to combine the empirical distribution functions which take values in $[0,1]$. If we give to each function the weight 1, each of them will contribute to the final sum by values on the same scale. Note that the assumption of the existence of

the pair correlation function which is moreover translation invariant is not needed when using variants 3 - 8. For each energy functional the algorithm is stopped if we do not accept the new configuration for 100 times in a row. We emphisize that we do not want to give here any strong recommendation about the choice of characteristics (and their weights) to be combine in the energy functional. Actually, we aim to illustrate the the advantages and disadvantages of use of the stochastic reconstruction method.

3.2 Quality of the reconstructions

Recall that we observe a point pattern φ on a bounded observation window W. We assume that φ is a realisation of a point process Φ for which the intensity function exists and it is a non-constant function. Suppose that we are able to simulate from the distribution of Φ. If we want to use the stochastic reconstruction algorithm, we shoud show that it gives "reasonable" results. If the input φ is a realisation of a particular theoretical model Φ, we want the reconstructions (i.e. the outputs of the algorithm) to correspond to the data and thus to the theoretical model Φ. By this we mean that a simulation based test will not reject the outputs too often. In the next paragraphs we will discuss what kind of simulation-based test will be used.

We will first describe the accumulated persistence function introduced in Biscio and Møller [2017]. This functional characteristic aims to describe geometrical structure of a point pattern by studying topological properties of the union of discs centered at the observed points.

Suppose now that $N(\varphi \cap W) = n \in \mathbb{N}$, i.e. φ is a configuration of n points in W. Thus we can write $\varphi = \{x_1, x_2, \ldots, x_n\}$. Let C_r^i be a disc centered at point x_i with radius $r \geq 0$, $i = 1, 2, \ldots, n$. Denote $C_r = \bigcup_{i=1}^{n} C_r^i$ the union of discs with radius r centered at points of φ. We are interested in how the s-dimensional topological features of C_r vary for different values of parameter r and for $s = 1, 2, \ldots, d-1$. Thus in case of planar point patterns it is only relevant to take $s = 0$ and $s = 1$. If $s = 0$ we focus on how connected components of the union C_r are changing when the parameter r is growing. If $s = 1$ the 1-dimensional holes (loops) are of particular interest. Figure 3.1 shows C_r for a realisation of a planar Thomas point process on unit square (with the intensity of the underlying Poisson process equal to 40 and the mean number of points per cluster equal to 6) and for three different values of the parameter r. We will denote the accumulated persistence function computed for a point patter φ by

$$APF_s(\varphi, m),\ m > 0,\ s = 1, 2, \ldots, d-1.$$

For more details and the general definiton see Appendix A.1 and Biscio and Møller [2017]. Note that the accumulated persistence function is again only an empirical characteristic and we do not give it any theoretical meaning.

Now we aim to test the null hypothesis H_{APF}. Suppose that we have a point pattern φ which is a realisation of a point process Φ. We want φ to be recon-

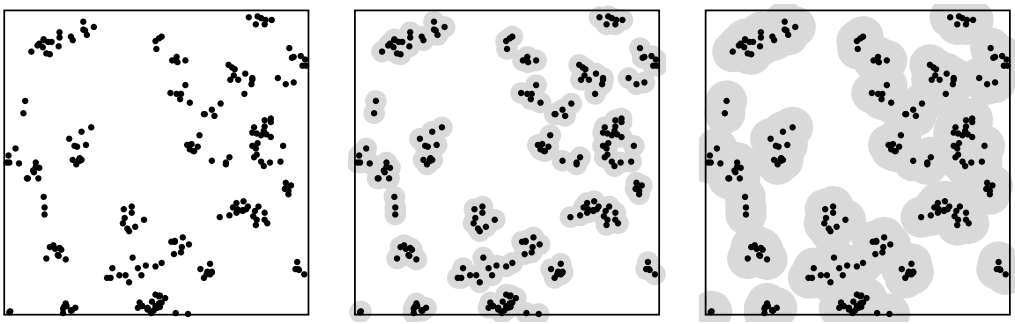

Figure 3.1: A union of discs centered at points of a realisation of planar Thomas point process for r_0 (left), r_1 (middle) and r_2 (right), where $0 = r_0$ and $r_0 < r_1 < r_2 < r_{max}$.

structed. Let us denote as ξ the reconstruction, i.e. the output of the stochastic reconstruction with φ as input. We should realize that ξ is a random object (if we run the algorithm with the same input several times we will obtain different outputs). By $APF_s(\xi, m), m > 0,\ s = 0, 1$, we denote the accumulated persistence function computed for ξ. In Section 3.3 we will work only with planar point processes and thus we consider only $s = 0$ and $s = 1$. Since ξ is a random object, $APF_s(\xi, m)$ is a random function and therefore it has a distribution. Similarly, $APF_s(\Phi, m), m > 0,\ s = 1, 2$, denotes the accumulated persistence function computed for the point process Φ. Again, $APF_s(\Phi, m)$ is a random object and thus it has a distribution. It is important to distinguish between $APF_s(\Phi, m)$ which is a random function and $APF_s(\varphi, m)$ (computed for the concrete realisation φ of Φ) which is a deterministic object. We aim to test the hypothesis

$$H_{APF}: \quad APF_s(\xi, m),\ m > 0,\ s = 0, 1, \text{ has the same distribution as } APF_s(\Phi, m),\ m > 0,\ s = 0, 1.$$

Since we do not know the distribution of $APF_s(\bullet, m),\ m > 0,\ s = 0, 1$ under the null hypothesis, a simulation-based test will be used. Let us now explain the basic principle of such tests.

Suppose we want to test the hypothesis that an observed point pattern corresponds to a given model. Assume we have a suitable test statistic S and its estimate for the observed pattern $\widehat{S}$. First we generate N independent simulations from the given model and for each simulation we estimate S. Hence we obtain $\widehat{S}_1, \widehat{S}_2, \ldots, \widehat{S}_N$. Now we rank the estimates from the smallest to the largest. We get an ordered sample $\widehat{S}_{(1)} \leq \widehat{S}_{(2)} \leq \cdots \leq \widehat{S}_{(N)}$. If the observed pattern corresponds to the null model we have that $\widehat{S}, \widehat{S}_1, \widehat{S}_2, \ldots, \widehat{S}_N$ are independent and identically distributed. By symmetry, every ranking has the same probability and therefore the probability that $\widehat{S}$ is smaller than $\widehat{S}_{(q)},\ q \in \{1, 2, \ldots, N\}$, equals $\frac{q}{N+1}$. Suppose that we want to test the null hypothesis on the significance level α. Then we find q in such a way that

$$\alpha = \frac{2q}{N+1}.$$

If α is a rational number, we can always choose such N for which we are able to compute q. The null hypothesis (observed pattern corresponds to the null

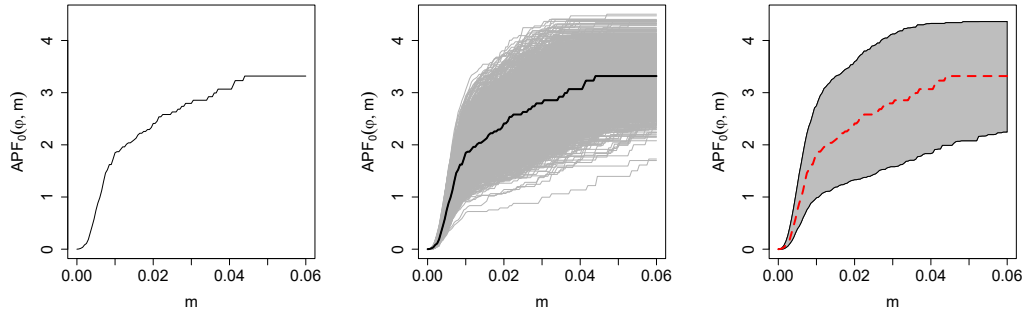

Figure 3.2: Left to right: the accumulated persistence function with $s = 0$ (connected components) for the point pattern from Figure 3.1, the accumulated persistence function with $s = 0$ for 2499 simulations from the theoretical model (Thomas point process with the intensity of the underlying Poisson point process equal to 40 and the mean number of points per cluster equal to 6), black curve is for the observed pattern from Figure 3.1, global rank envelope based on these 2499 simulations, red dashed curve is for the observed pattern from Figure 3.1.

model) is then rejected if $\widehat{S} \notin \left[\widehat{S}_{(q)}, \widehat{S}_{(N-q+1)}\right]$. We will refer to this type of test as the pointwise Monte Carlo test. Note that we have just described a two-sided test. Modification to one-sided test is obvious: q is found in such a way that $\alpha = \frac{q}{N+1}$ and the null hypothesis is rejected if $\widehat{S} < \widehat{S}_{(q)}$ (or $\widehat{S} > \widehat{S}_{(N-q+1)}$).

Note that the test statistic S in the pointwise Monte Carlo test is a single number. But the accumulated persistance function is a function. We could test the hypothesis H_{APF} using the accumulated persistence function for a particular value of s and m, but we do not have any prior information how to choose these parameters. Hence we would like to use another approach. Suppose now that the test statistic S from the last paragraph is now a function, i.e. we consider $S(r)$ on an interval $[r_{min}, r_{max}]$, where $0 < r_{min} < r_{max}$ are prescribed real constants. Again $\widehat{S}(r)$ denotes the estimate of $S(r)$ for the observed pattern and by $\widehat{S}_1(r), \widehat{S}_2(r), \ldots, \widehat{S}_N(r)$ we denote estimates based on the simulations from null model. Now for each value of r we are able to estimate $\widehat{S}^{(q)}(r)$ and $\widehat{S}^{(N-q+1)}(r)$. If we join the values $\widehat{S}^{(q)}(r)$, $r \in [r_{min}, r_{max}]$ we obtain so-called lower envelope and by joining $\widehat{S}^{(N-q+1)}(r)$, $r \in [r_{min}, r_{max}]$ we get the upper envelope. This approach results in pointwise envelopes which can be used for visualizing deviations from the null hypothesis. However, pointwise envelopes should not be used to perform statistical tests because they do not take into account the problem of multiple testing.

We will use the global rank envelope test described in Myllymäki et al. [2017] instead. This approach gives us the recipe how to rank the observed and simulated functional statistics in order to use the basic principle of Monte-Carlo testing described above. It means that we give a rank R_i to each functional statistics $\widehat{S}_i(r)$, $i = 0, 1, \ldots, N$, where by $\widehat{S}_0(r)$ we mean the data curve. It may happend that the same rank is assigned to several functional statistics $\widehat{S}_i(r)$. For this

reason we compute the most liberal and the most conservative p-value as follows

$$p_- = \frac{1}{N+1}\sum_{i=0}^{N} \mathbf{1}\{R_i < R_0\} \quad \text{and} \quad p_+ = \frac{1}{N+1}\sum_{i=0}^{N} \mathbf{1}\{R_i \leq R_0\}.$$

These values provide the lower and upper bounds for the p-value of the test. Also this approach allows to display test results in a graphical manner - based on the simulated functional statistics we are able to draw the global envlope. We reject the null hypothesis if the data curve leaves the envelope anywhere. This is equivalent with the situation that $p_+ \leq \alpha$. If the data curve is fully contained inside the enelope, the hypothesis is not rejected. This is equivalent with $p_- > \alpha$. If the data curve does not leave the envelope but coincides with the bounds of the envelope it is equivalent with $p_- \leq \alpha < p_+$. In this situation it is not clear if we should reject the null hypothesis or not. To make this situation unlikely, the expected width of the p-interval should be small. Myllymäki et al. [2017] claims that using $N = 2499$ (at least) leads to the reasonable small wiidth of the p-interval.

In Figure 3.2 we can see the accumulated persistence function for $s = 0$ and the point pattern from Figure 3.1 (left) and the same for 2499 simulations from the null model (Thomas point process with the intensity of the underlying Poisson process equal to 40 and the mean number of points per cluster equal to 6) and the global rank envelope based on the 2499 simulations (right).

Since we suppose that the observed point pattern φ comes from the theoretical model Φ with a non-constant intensity function λ, we should also control whether the intensity function of the output ξ corresponds to λ. Let us equip the observation window W with a pixel grid V with vertices v. Consider a test statistic

$$D(\xi) = \sum_{v \in V} a(v) \left[\lambda(v) - \widehat{\lambda}_\xi(v)\right]^2.$$

By the symbol $\widehat{\lambda}_\xi$ we meant the estimated intensity function of the output ξ. Since ξ is a random object, $\widehat{\lambda}_\xi$ is random function and for different realisations of ξ it will take different values. Similarly, by $D(\Phi)$ we denote the test statistic D computed for the null model Φ. We aim to test the hypothesis

$$H_\lambda : D(\xi) \text{ has the same distribution as } D(\Phi).$$

Suppose again that we have generated N simulations from the null model. We compute the test statistics

$$D_i = \sum_{v \in V} a(v) \left[\lambda(v) - \widehat{\lambda}_i(v)\right]^2, \quad i = 1, 2, \ldots, N,$$

where $\widehat{\lambda}_i$ is the estimated intensity function computed for the i-th simulation on the observation window W. By the symbol $a(v)$ we mean that we multiply the summands by the area of one pixel. In the same way we compute the value of D for one particular reconstruction of φ (i.e. one particular realisation of ξ). Again we can rank the quantities $D_1, D_2, \ldots, D_N$ from the smallest to the largest.

The null hypothesis H_λ is rejected if $D > D^{(N-q+1)}$, where q is chosen accordingly to the required significance level. We will refer to this test as the deviation test. Note that this test is one-sided and the p-value is computed as

$$p = 1 - \frac{1}{N+1} \sum_{i=1}^{N} \mathbf{1}\{D > D_i\},$$

see Loosmore and Ford [2006].

We emphasise that statistical test used for thesting the hypothesis H_{APF} will be always refered to as the global rank envelope test. Similarly, the test used for thesting the hypothesis H_λ will be always refered to as the deviation test. Note that we are able to perform these tests since we are investigating the properties of the stochastic reconstruction algorithm in an controled experiment. On the other hand, in practical applications we are not able to perform such tests.

3.3 Simulation study

In this section we will describe the simulation study we made in order to compare the quality of outputs of the stochastic reconstruction algorithm based on the 8 different energy functionals described in Section 3.1.

As the theoretical model Φ we will take the thinned Thomas point process, inhomogeneous Poisson process and the transformed Matérn hard-core process of type II respectively. Parameters of these models will be chosen in such a way that all of these three models will have the same intensity function. We have chosen the thinned Thomas process as a typical example of clustered processes that are of particular interest in ecological applications. In spatial statistics Poisson point process often serves as a benchmark to whom we compare other processes. For this reason we have included the inhomogeneous Poisson point process in the simulation study. Transformed Matérn hard-core process was chosen as an example of regular processes. Even though we have not found in literature an application of use of the stochastic reconstruction algorithm on regular data, we want to investigate the properties of the algorithm under these circumstances.

In the simulation study we will consider only planar point patterns and the observation window W will be always the unit square. We emphasise that for all computations we use the software `R` (version 3.4.3) with packages `spatstat` (version 1.55-0), `TDA` (version 1.6) and `GET` (version 0.1). The code for evaluation the accumulated persistence function is not included in any of these packages but it can be found on `http://people.math.aau.dk/~christophe/`.

Let us now make few remarks about the implementation of the stochastic reconstruction algorithm or more precisely the energy functionals from Section 3.1. The estimates of functional characteristics are computed for a finite number of values of r. Hence the integrals are for simplicity of the implementation considered as sums. The constants $R(L)$ and $R(g)$ are set to be 0.2. We have

followed the common recommendation which says that we should not compute the estimtes of L and g for r bigger than one fourth of the shorter side of the observation window (which is $[0,1]^2$). For the empirical distribution functions $\widehat{F}_k$ and $\widehat{D}_k$ we have run some experiments and we have decided to put $R(F_k) = R(D_k) = 0.3$ if we work with $k = 1, 2, \ldots, 5$ and $R(F_k) = R(D_k) = 0.4$ if we work with $k = 1, 2, \ldots, 20$. We use the stopping rule mentioned in Section 3.1 – the algorithm is stopped if we do not accept the new configuration for 100 times in a row.

The aim of this work is not to find "the best" energy functional for some particular observed point pattern. We try to describe the difficulties related with the modification of the algorithm for inhomogeneous point patterns and to suggest some solutions to these problems. For these reasons we do not want to discuss here the choice of all possible parameters when computing estimates of the functional characteristics in the software `R`. Let us just mention that the estimates of the intensity function are computed using the `spatstat` function `density.ppp` with default choice of parameters (Gaussian kernel). Estimates of the pair correlation function are computed using the `spatstat` function `pcfinhom` with the default choice of parameters (Epanechnikov kernel) except the parameter `divisor` which is set to "d". For more details see Chapter II, Section 7 in Baddeley et al. [2015]. Edge corrections were specified in Section 1.4. Illustration of the implementation of the stochastic reconstruction algorithm can be found in Appendix A.6.

The global rank envelope test and also the deviation test will be performed on the significance level $\alpha = 0.05$ and they will be based on 2499 simulations from the null model. We emphasise that for the global rank envelope test we use the conservative test – we reject the null hypothesis if the upper bound of the p-interval p_+ is less than or equal to α. For the 2499 simulations the typical width of the p-interval in our experiments is about 0.02.

3.3.1 Reconstructing thinned Thomas process

In this section we will investigate quality of stochastic reconstruction for thinned Thomas process. We have stated the definition of Thomas point process in Definition 13 in Section 1.2. Now we will define independent thinning – a point process operation when we independently delete certain points of the process.

Definition 25. *Let $f : \mathbb{R}^d \longrightarrow [0,1]$ be a measurable function. Let Φ_{thin} be a cluster point process with independent clusters. If for all $x \in \mathbb{R}^d$ we have*

$$\Theta_x = \begin{cases} \{x\} & \text{with propability } f(x), \\ \emptyset & \text{with probability } 1 - f(x), \end{cases}$$

then we call Φ_{thin} the independent thinning of the parent point process Φ_p.

Remark. We can set $\Phi_{thin} = \{X \in \Phi_p : \ R(X) \leq p(X)\}$, where $\left\{R(x) : \ x \in \mathbb{R}^d\right\}$ is a collection of random variables with the uniform distribution on $[0,1]$ which are mutually independent and independent of Φ_p.

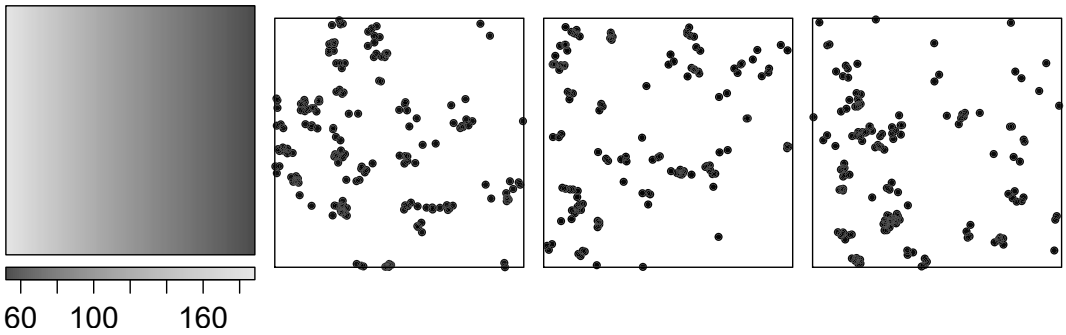

Figure 3.3: Three different realisation of $\Phi_{T,thin}$ on the unit square. For detailed description see Definition 25 and the main text below it. The leftmost picture shows the values of the intensity function λ of $\Phi_{T,thin}$ on the unit square.

If we apply the independent thinning given by the function f to a stationary point process with intensity τ we obtain a second-order intenrity reweighted stationary point process with intensity function λ given by

$$\lambda(x) = \tau f(x), \quad x \in \mathbb{R}^d,$$

see Section 6.2.1 in Illian et al. [2004] and Baddeley et al. [2000].

Now let Φ_T be a stationary planar Thomas point process with the intensity of the underlying Poisson point process $\kappa = 40$ and the expected number of points per cluster $\gamma = 6$. As we have mentioned in Section 1.2, Φ_T has the intensity $\tau = \gamma\kappa = 240$. In order to get our first theoretical model $\Phi_{T,thin}$ we will apply the independent thinning given by the function

$$f_{thin}(x, y) = \min\left\{Ke^{-Ax}, 1\right\}, \quad (x, y) \in \mathbb{R}^2,$$

where $K \approx 0.986$ and $A = 1.3$, on Φ_T. Choice of such constants will be clarified in Subsection 3.3.3. Note that the function f_{thin} will be used only on the observation window $W = [0, 1]^2$. Together with the fact that $K < 1$ we obtain that the term Ke^{-Ax} is less than 1 for all $(x, y) \in [0, 1]^2$. We emphasize that $\Phi_{T,thin}$ is a second-order reweighted stationary point process, see page 331 in Baddeley et al. [2000].

The intensity function of $\Phi_{T,thin}$ is then given by

$$\lambda(x, y) = 240 f_{thin}(x, y), \quad (x, y) \in \mathbb{R}^2. \tag{3.1}$$

Values of the intensity function λ on the unit square can be seen in Figute 3.3 as well as three different realisations of $\Phi_{T,thin}$ on the unit square. Each realisation has about 120 points. Usually, the stochastic reconstruction algorithm stops after 1500 or 2000 iterations, hence each point of the initial configuration has a chance to be moved approximately 10 or more times.

To investigate the quality of stochastic reconstruction for $\Phi_{T,thin}$ we generate 15 different realisations of $\Phi_{T,thin}$ on the observation window $W = [0, 1]^2$.

Energy functional	Summary characteristics
E_1	inhomogeneous L-function
E_2	inhomogeneous pair correlation function
E_3	$\widehat{F}_1$ up to $\widehat{F}_5$
E_4	$\widehat{F}_1$ up to $\widehat{F}_{20}$
E_5	$\widehat{D}_1$ up to $\widehat{D}_5$
E_6	$\widehat{D}_1$ up to $\widehat{D}_{20}$
E_7	$\widehat{F}_1$ up to $\widehat{F}_5$ and $\widehat{D}_1$ up to $\widehat{D}_5$
E_8	$\widehat{F}_1$ up to $\widehat{F}_{20}$ and $\widehat{D}_1$ up to $\widehat{D}_{20}$

Table 3.1: Reminder of the 8 energy functionals used in the simulation study. More details can be found is Section 3.1.

We have then 15 different inputs for the stochastic reconstruction algorithm. For each of these inputs we will generate 20 reconstructions for each of the energy functionals E_1 to E_8. Hence for each energy functional we obtain 300 reconstructions. We aim to compare how many of these 300 reconstructions correspond, for each energy functional, to the null model $\Phi_{T,thin}$. To achieve this we will use the two statistical tests described in Section 3.2.

First we will test the null hypothesis H_{APF}. Recall that we have generated 2499 simulations from the model $\Phi_{T,thin}$ for which we have computed the accumulated persistence function (for both $s = 0$ and $s = 1$, it is possible to construct the envelope from two different curves concatenated into one long curve). Based on these curves we have constructed the envelope. This particular envelope will be used for all of the 300 tests performed. For each of the energy functionals we will check how many of the 300 curves computed from the reconstructions leave the envelope. The significance level α is 0.05, thus we expect to see about 5% of the 300 reconstructions rejected. Results are briefly summarized in Table 3.2, detailed version can be found in Appendix A.2. Reminder of the 8 energy functionals that we are using can be found in Table 3.1.

Before testing the reconstructions, we have first tested the 15 inputs to see how extreme they are with respect to the distribution under the null hypothesis H_{APF}. We have discovered that 3 of the 15 outputs can be qualified as extreme, i.e. the upper bound of the p-interval given by the global rank envelope test is 0.05, 0.05 and 0.091 respectively. See Table A.2 in Appendix A.2.

For E_1 and E_2 we reject more than one half of the 300 reconstructions (see Table 3.2). It seems that the L-function (E_1) and the pair correlation function (E_2) do not contain enough information about the structure (interactions and/or the intensity function) of the investigated patterns. For L-function this result corresponds to the conclusion made in Tscheschel and Stoyan [2006]. Even though Wiegand et al. [2013] claim that the pair correlation function contains the most information about a point pattern (comparing with the other characteristics listed in Section 1.3), the stochastic reconstruction algorithm described in Section 3.1 with the energy functional based only on the pair correlation function

	Percentage	Mean	SD
E_1	56%	11.2	5.92
E_2	68%	13.5	5.21
E_3	25%	5.00	5.06
E_4	77%	15.4	4.00
E_5	25%	4.93	6.70
E_6	16%	3.20	4.16
E_7	6%	1.27	2.25
E_8	13%	2.60	3.36

	Percentage	Mean	SD
E'_1	55%	11.0	5.59
E'_2	40%	7.93	4.76
E'_3	62%	12.4	4.44
E'_4	96%	19.3	1.22
E'_5	7%	1.47	2.92
E'_6	23%	4.53	5.33
E'_7	7%	1.40	2.29
E'_8	23%	4.53	5.59

Table 3.2: Testing the hypothesis H_{APF}. Three different entries are given: percentage of the overall rejected reconstructions, mean number of rejected outputs (per one input) and the standard deviation of the number of rejected outputs (per one input). Left table corresponds to the 8 energy functionals described in Section 3.1, right table corresponds to the situation when we add the term $\widehat{F}_\lambda$ to each of the energy functionals.

(E_2) does not provide reasonable reconstructions in the sense of the quality criterion we use. Wiegand et al. [2013] also encourage to use more than one summary characteristic.

For E_3 and E_5 we reject one fourth of the outputs. If we use E_6 instead of E_5 (i.e. we increase the number k of considered nearest neighbours from 5 to 20) the number of rejected outputs decreases a little (see Table 3.2). This observation corresponds to Tscheschel and Stoyan [2006] where the authors (in the stationary context) suggest using the combination of D_1, D_2 up to D_{41}. On the other hand if we move from E_3 to E_4 we suddenly reject three times more reconstructions than before. The functional E_7 which combines both types of the characteristics based on interpoint distances gives so far the the best results, we reject only 6% of reconstructions. If we add more terms, i.e. we use the functional E_8 we obtain slightly more rejections. Since we are trying to reconstruct clustered point patterns, it is useful to work with the empirical distribution functions $\widehat{D}_k$ and $\widehat{F}_k$ because they can describe the geometry of clusters properly. The number k (i.e. the number of considered nearest neighbours or nearest points of the pattern respectively) shoud be big enough to cover almost all points (or at least a major part of points) in a cluster. In our case $k = 5$ seems to be enough. If we increase k it may happend that the terms with high k no longer describes the geometry of one cluster but they are describing the distances among different clusters. In our model $\Phi_{T,thin}$ the clusters are mutually independent and thus there is no reason to try to reproduce distances among clusters that we have observed for the input. For that reason the functionals E_4, E_6 and E_8 may become less informative and they may give worse results than E_3, E_5 and E_7 respectively.

It seems that (except E_4) the energy functionals based on interpoint distance characteristics work in this case better than functionals based on L or g. Moreover, energy functional E_7 gives in the context of the APF reasonable result and it seem that no more modifications are needed. Examples of outputs

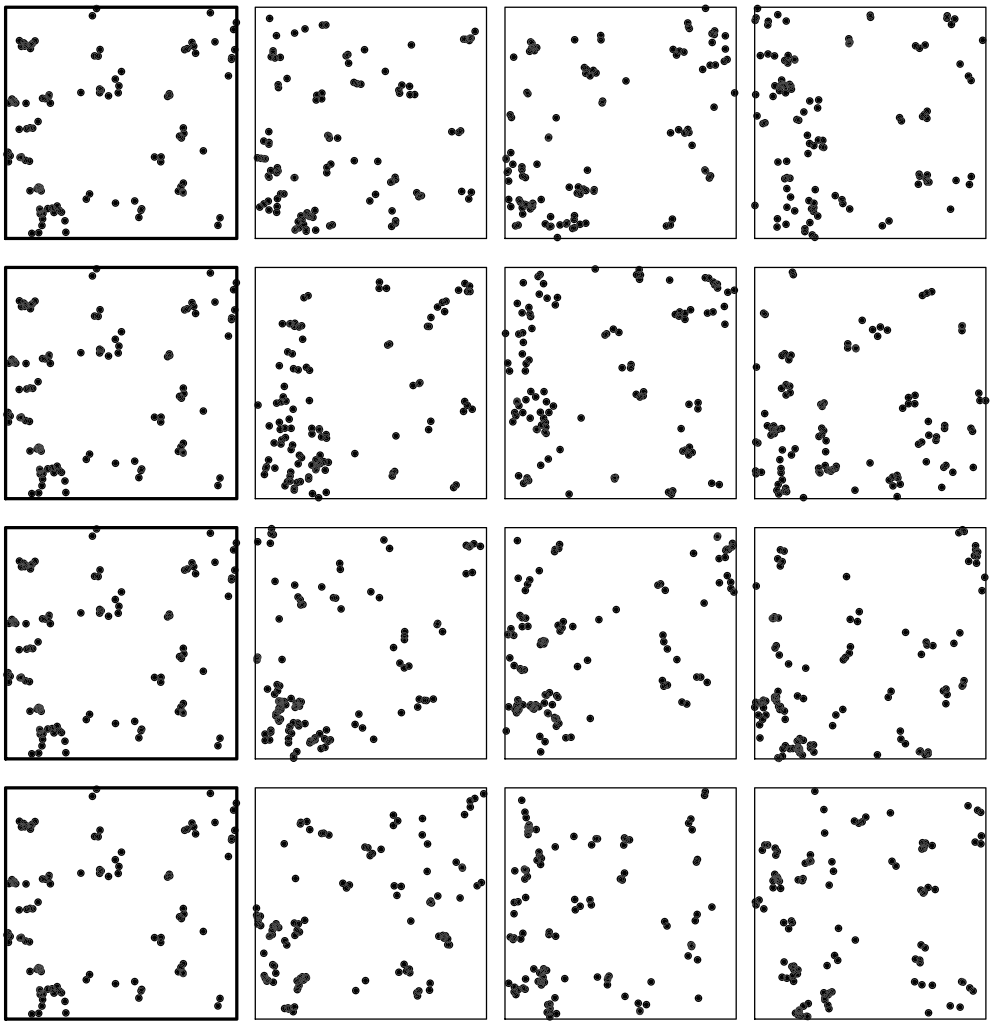

Figure 3.4: Input (on the left, same for all the rows) and three different outputs of the stochastic reconstruction algorithm based on the energy functional E_1 (first row), E_2 (second row), E_5 (third row) and E_7 (last row). Functionals E_1 and E_2 do not contain enough information about the point pattern to reproduce successfully the size and the shape of clusters. In contrary, for E_5 and E_7 the structure of clusters of the outputs corresponds more or less to the structure observed for the input. Note that in this moment we say nothing about the intensity function.

of the stochastic reconstruction algorithm based on E_1, E_2, E_5 and E_7 can be seen in Figure 3.4. We have chosen the functionals E_1 and E_2 for which we reject more than one half of the outputs and the functionals E_5 (one fourth of the rejected outputs) and E_7 (best results) in order to illustrate the differences between the quality of the reconstructions. Since the APF describes the geometrical structure of a point pattern (e.g. the shape and size of clusters), there should be a visible difference between the outputs based on E_1 and E_7. Indeed, in Figure 3.4 it can be seen that for the L-function (E_1) or the pair correlation (E_2) the structure of clusters of the outputs does not correspond to the structure we have observed for the input. in contrast, for E_5 and E_7 the shape and the size of clusters is reproduced in the right way. However, in

	Percentage	Mean	SD
E_1	22%	4.33	5.97
E_2	39%	7.73	7.26
E_3	27%	5.33	7.86
E_4	20%	4.07	7.23
E_5	65%	13.0	7.46
E_6	27%	5.33	8.11
E_7	33%	6.53	8.83
E_8	18%	3.60	7.18

	Percentage	Mean	SD
E'_1	7%	1.47	4.91
E'_2	9%	1.87	5.19
E'_3	14%	2.73	5.81
E'_4	17%	3.47	6.66
E'_5	15%	3.00	6.33
E'_6	12%	2.33	5.70
E'_7	14%	2.87	5.90
E'_8	13%	2.60	5.79

Table 3.3: Testing the hypothesis H_λ. Three different entries are given: percentage of the overall rejected reconstructions, mean number of rejected outputs (per one input) and the standard deviation of the number of rejected outputs (per one input). Left table corresponds to the 8 energy functionals described in Section 3.1, right table corresponds to the situation when we add the term $\widehat{F}_\lambda$ to each of the energy functionals.

this moment we say nothing about the intensity function. We must check also results of the second test.

We will now test the hypothesis H_λ. Again, before testing the reconstructions, we will first consider the extremeness of the inputs. This time, 4 of the 15 inputs can be qualified as extreme, i.e. the p-values given by the deviation test are 0.058, 0.085, 0.090 and 0.027. Detailed results of the deviation test can be found in Table A.2 in Appendix A.2.

After testing the inputs, we can start with testing the reconstructions. Results are briefly summarized in Table 3.3, detailed version can be found in Appendix A.2. For the pair correlation function (E_2) we reject almost two times more outputs than for the L-function (E_1). If we compare the estimators (1.3) for L-function and (1.2) for the pair correlation function we can see that the non-constant intensity function is treated similarly in both of them. It may be interesting to examine if the choice of the bandwidth b of the kernel funtion u_b in the estimator of the pair correlation function (function `pcfinhom` in `spatstat`) influences the number of rejected outputs. But in this text we do not perform such experiments. For the L-function (E_1) the major part of the rejected outputs corresponds to the 4 extreme inputs. For the pair correlation function (E_2) also the non-extreme inputs cause many rejections, see Table A.2 in Appendix A.2.

Using E_3 and E_4 we reject about one fourth of the 300 outputs. At the same time we rejcet almost exclusively the outputs corresponding to the 4 extreme inputs. Even though the information about the geometrical sturcture of the input contained in E_4 may be misleading (and hence we reject more than three fourth of the outputs when testing H_{APF}) it seems that the intensity function of outputs roughly corresponds to λ.

Concerning the deviation test, for E_5 we reject the biggest number of outputs

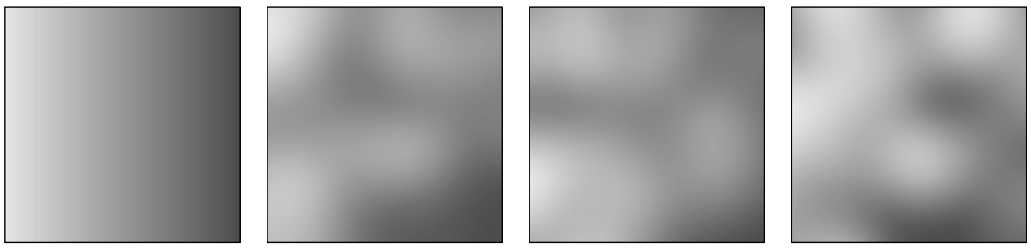

Figure 3.5: Values of the theoretical intensity function λ on the observation window W (leftmost). Values of the estimated intensity function on the observation window W for inputs 1, 4 and 14.

among all the energy functionals, see Table 3.3. In this case we reject too much outputs even for the non-extreme inputs. If we increase k from 5 to 20, the percentage of rejected outputs decreases. For the last two energy functionals only the outputs corresponding to the 4 extreme inputs are rejected. In this moment with the functional E_8 we achieve the best reconstructions regarding both, the APF and the intensity function. But we think we are able do better.

Even though there are four extreme patterns among our 15 inputs, the deviation test showed that some of the energy functionals (e.g. E_2 or E_4) produce point patterns whose estimaded intensity function deviates (on the observation window W) a lot from λ. Also for some of the extreme inputs we suppose that it is possible to improve the results. We have discovered that all of the energy functionals we use do not contain enough information about the intensity function. Thus we want to add some information about the intensity function to each of them. For more details see the text below. In this moment we can conclude that neither the two obvious modifications of the stochastic reconstruction algorithm mentioned in Section 3.1 nor the use of inhomogeneous characteristics in the energy functional really ensure that the estimated intensity function of outputs corresponds to λ.

Let us now take again a pixel grid V in the observation window W. For every point pattern on the observation window W we are able to compute values of the estimated intensity function in vertices of V. Once we have these values, it is easy to compute their empirical distribution function. We will denote this empirical characteristic by $\widehat{F}_\lambda$ and we will add it to each of our 8 energy functionals. This characteristic gives us information about the values of the estimated intensity function of the investigated point pattern, but it does not tell us where we have observed such values. It means that we have some idea about how big or low the values of the estimated intensity function should be and how large are the ares with high/low values, but we know nothing about how to place the areas with high or low intensity function in the observation window W.

By adding the term $\widehat{F}_\lambda$ to the energy functionals we force the outputs to have the right values of the estimated intensity function. What we do not enforce is the right placement of the right values in the observation window. In other words we do not enforce that the surface (e.i. right values plus placement) of the

estimated intensity function on W corresponds to the surface of the theoretical intensity function of $\Phi_{T,thin}$ on W. The main reason why we do not want to control the whole surface is that such characteristic would not be an empirical distribution function and hence the advantage of choosing weights naturally when combining empirical distribution functions would be lost. Moreover, it would be useless to perform the deviation test (testing hypothesis H_λ), because this test itself checks the whole surface of the estimated intensity function of outputs. Values of the theoretical intensity function λ on the observation window W and values of the estimated intensity function for inputs 1, 4 and 14 can be seen in Figure 3.5. We emphasise that these inputs are not qualified as extreme.

We have empirically chosen weights for the first two energy functionals and we have obtained the following

$$\begin{aligned} E_1' &= 10E_1 + \widehat{F}_\lambda, \\ E_2' &= E_2 + 100\widehat{F}_\lambda. \end{aligned}$$

What we mean by the phrase "empirically chosen" is described in Section 2.1. When testing the null hypothesis H_{APF}, we reject slightly less outputs for E_1' than for E_1. For E_2' the decrease of the number of rejected outputs is even bigger, see Table 3.2. Performing the second test, i.e. the deviation test of the null hypothesis H_λ, the percentage of rejected outputs is satisfyingly small for both of these functionals, see Table 3.3. So adding the term $\widehat{F}_\lambda$ to E_1 and E_2 really assures the right form of the intensity function. It may be surprising because $\widehat{F}_\lambda$ assure the right values of the estimated intensity function but not the right placement of the right values in the observation window W. In contrary, the deviation test is testing the right values and the right placement. In this situation the right placement is guaranted by the inhomogeneous initial configuration and the non-uniform distribution of the proposed points. The fact the estimated intensity function of the reconstructions corresponds to the theoretical one influences positively results of the global rank envelope test based on the APF. But still neither the L-function nor the pair correlation function contain enough information to reproduce the interactions among points and the geometry of clusters properly.

For the energy functionals based on the empirical distribution functions we have in mind the following choice of weights – the weight of the term $\widehat{F}_\lambda$ will be the same as the sum of weights of the other terms. For example if we work with the empirical distribution functions of the distance to the k-th nearest neighbour for $k = 1, 2, \ldots, 5$ we combine 5 emprirical distribution functions. Each of them has the weight 1. To make the information about the intensity function equally importat as the information contained in the combination of the 5 empirical distribution functions, we will give to the term $\widehat{F}_\lambda$ the weight $1+1+1+1+1 = 5$.

We thus obtain

$$\begin{aligned} E_3' &= E_3 + 5\widehat{F}_\lambda, \\ E_4' &= E_4 + 20\widehat{F}_\lambda, \\ E_5' &= E_5 + 5\widehat{F}_\lambda, \\ E_6' &= E_6 + 20\widehat{F}_\lambda, \\ E_7' &= E_7 + 10\widehat{F}_\lambda, \\ E_8' &= E_8 + 40\widehat{F}_\lambda. \end{aligned}$$

We are aware that this choice of weights is arbitrary and in practical applications there amy be reasons to choose the weights differently, possibly based on prior knowledge about the problem.

For E_3' and E_4' the global rank envelope test rejects more outputs than before, see Table 3.2. So it seems that in these cases adding the new term $\widehat{F}_\lambda$ does not bring any benefits. The natural choice of weights may be wrong when combining with $\widehat{F}_k$. On the other hand, the deviation test rejects slightly less outputs, see Table 3.5.

For E_5' the global rank envelope test (hypothesis H_{APF}) now rejects only 7% of outputs. That is a nice improvement compared to E_5. So adding the term $\widehat{F}_\lambda$ to the energy functional E_5 has a positive impact on the accumulated persistence function of the outputs. Concerning the deviation test (hypothesis H_λ), the enhancementand is even more evident – for E_5' we reject only 15% of outputs. For E_5 it was 65%. We should remember that testing the hypothesis H_λ, 4 of the 15 inputs were classified as extreme. For the combination of $\widehat{D}_1$ up to $\widehat{D}_5$ with $\widehat{F}_\lambda$ we reject almost no outputs for the non-extreme inputs and hence we think that if we had less extreme inputs we would see the desired 5% (or less) of rejected outputs. Thus we can say that this energy functional provides reconstructions with the correct form of the APF and the intensity function. So far E_5' gives the best results regarding both of the two statistical tests we have performed. Not only for E_5' but also for E_7' we get satisfying results. The global rank envelope test rejects 7% and the deviation test rejects 14% of outputs.

To conclude this section, let us say that the stochastic reconstruction algorithm based on the energy functional E_7 produces outputs with the right form of the accumulated persistence function but the estimated intensity function of outputs deviates from the theoretical one. If we add the term $\widehat{F}_\lambda$ to all considered energy functionals, we enforce relatively small number of rejected outputs in the deviation test (for all of the energy functionals). Energy functionals based on inhomogeneous L-function (E_1) and the pair correlation function (E_2) does not contain enough information to reproduce the geometrical properties of clusters sucesfully, even though the term $\widehat{F}_\lambda$ enforces the right form of the estimated intensity function of the reconstrutions. For the energy functionals based on the empirical distribution functions $\widehat{D_k}$ and $\widehat{F}_k$ we have achieved the best results for E_5' and E_7'. These two energy functionals provide reconstructions that really corresponds

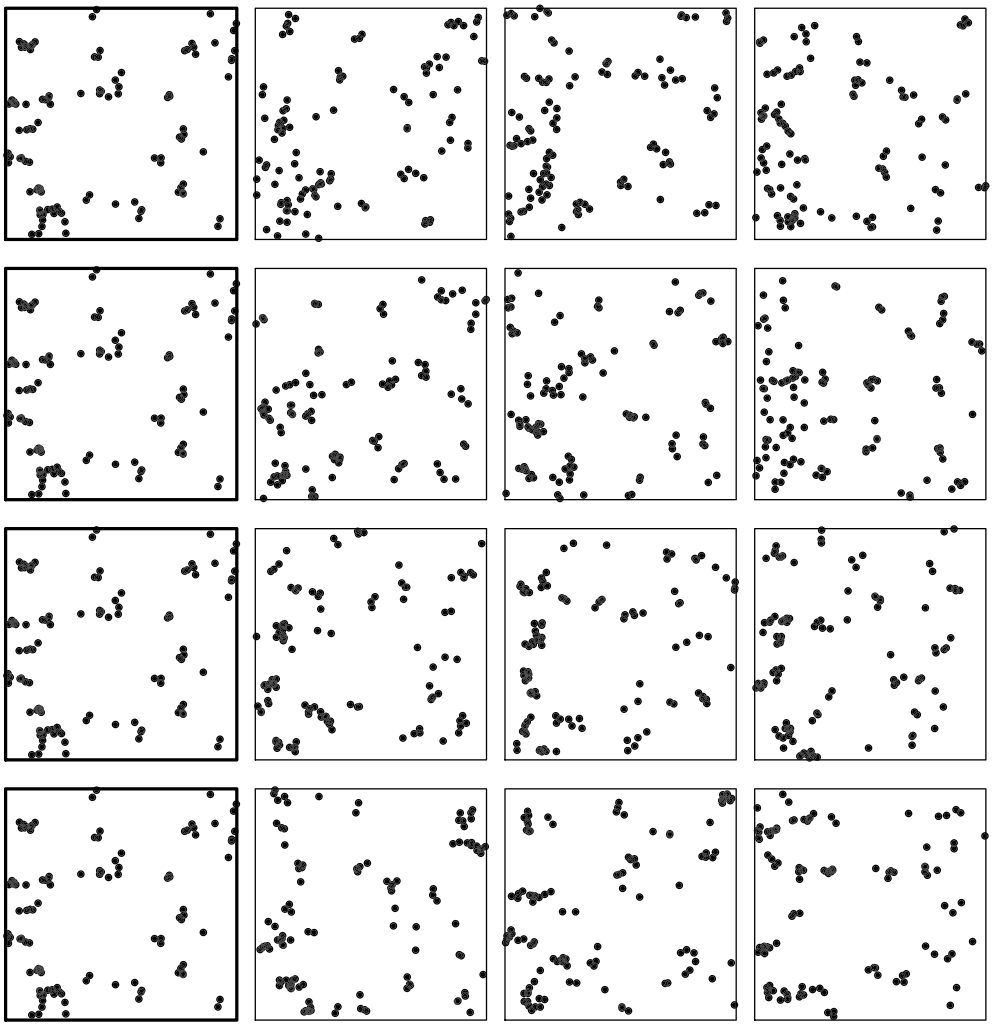

Figure 3.6: Input (on the left, same for all the rows) and three different outputs of the stochastic reconstruction algorithm based on the energy functional E_1' (first row), E_2' (second row), E_5' (third row) and E_7' (last row). The correct form of the estimated intensity function of the outputs is enforced by adding the term $\widehat{F}_\lambda$ to the energy functionals. But still, E_1' and E_2' do not contain enough information to reproduce the geometric properties of the clusters properly. On the other hand for functionals E_5' and E_7' the outputs really correspond to the input (in the context of testing hypotheses H_{APF} and H_λ).

to the observed data. Examples of outputs of the stochastic reconstruction algorithm based on E_1', E_2', E_5' and E_7' can be seen in Figure 3.6. We have chosen these functional in order to compare the outpusts with Figure 3.4.

3.3.2 Reconstructing inhomogeneous Poisson point process

In this subsection we will focus on reconstruction of an inhomogeneous Poisson process. We have already stated the definition of Poisson point process, see Definition 9 in Section 1.2. So our second theoretical model will be inhomogeneous planar Poisson point process Φ_{Poiss} with the intenity measure given

	Percentage	Mean	SD
E_1	11%	2.13	4.16
E_2	8%	1.67	3.79
E_3	5%	1.07	2.34
E_4	6%	1.13	2.07
E_5	6%	1.27	2.28
E_6	5%	1.00	1.89
E_7	4%	0.80	2.24
E_8	5%	0.93	2.46

	Percentage	Mean	SD
E'_1	4%	0.87	2.64
E'_2	9%	1.80	3.93
E'_3	8%	1.67	3.85
E'_4	5%	0.93	2.34
E'_5	3%	0.53	1.55
E'_6	5%	1.00	2.10
E'_7	4%	0.87	1.92
E'_8	3%	0.60	1.68

Table 3.4: Testing the hypothesis H_{APF}. Three different entries are given: percentage of the overall rejected reconstructions, mean number of rejected outputs (per one input) and the standard deviation of the number of rejected outputs (per one input). Left table corresponds to the 8 energy functionals described in Section 3.1, right table corresponds to the situation when we add the term $\widehat{F}_\lambda$ to each of the energy functionals.

by the intensity function λ from Equation (3.1). So the theoretical intensity function of Φ_{Poiss} is the same as for the thinned Thomas process $\Phi_{T,thin}$. Also, we recall that inhomogeneous Poisson process is a SOIRS process, see Section 1.3.

Again we generate 15 different realisations of Φ_{Poiss} on the observation window $W = [0,1]^2$ which will serve as inputs of the stochastic reconstruction algorithm. For each of these inputs and each of the 16 energy functionals mentioned in Subsection 3.3.1 we will generate 20 reconstructions. For each energy functional we will investigate how many of the 300 reconstructions do not contradict the null hypotheses H_{APF} and H_λ.

	Percentage	Mean	SD
E_1	34%	6.73	5.82
E_2	39%	7.73	5.52
E_3	32%	6.47	6.17
E_4	28%	5.67	5.89
E_5	36%	7.27	6.30
E_6	34%	6.73	5.79
E_7	34%	6.87	5.97
E_8	28%	5.67	6.13

	Percentage	Mean	SD
E'_1	30%	6.00	6.43
E'_2	35%	7.00	6.01
E'_3	32%	6.47	6.81
E'_4	25%	5.00	6.02
E'_5	36%	7.20	7.21
E'_6	30%	5.93	6.54
E'_7	32%	6.40	6.80
E'_8	29%	5.80	6.85

Table 3.5: Testing the hypothesis H_λ. Three different entries are given: percentage of the overall rejected reconstructions, mean number of rejected outputs (per one input) and the standard deviation of the number of rejected outputs (per one input). Left table corresponds to the 8 energy functionals described in Section 3.1, right table corresponds to the situation when we add the term $\widehat{F}_\lambda$ to each of the energy functionals.

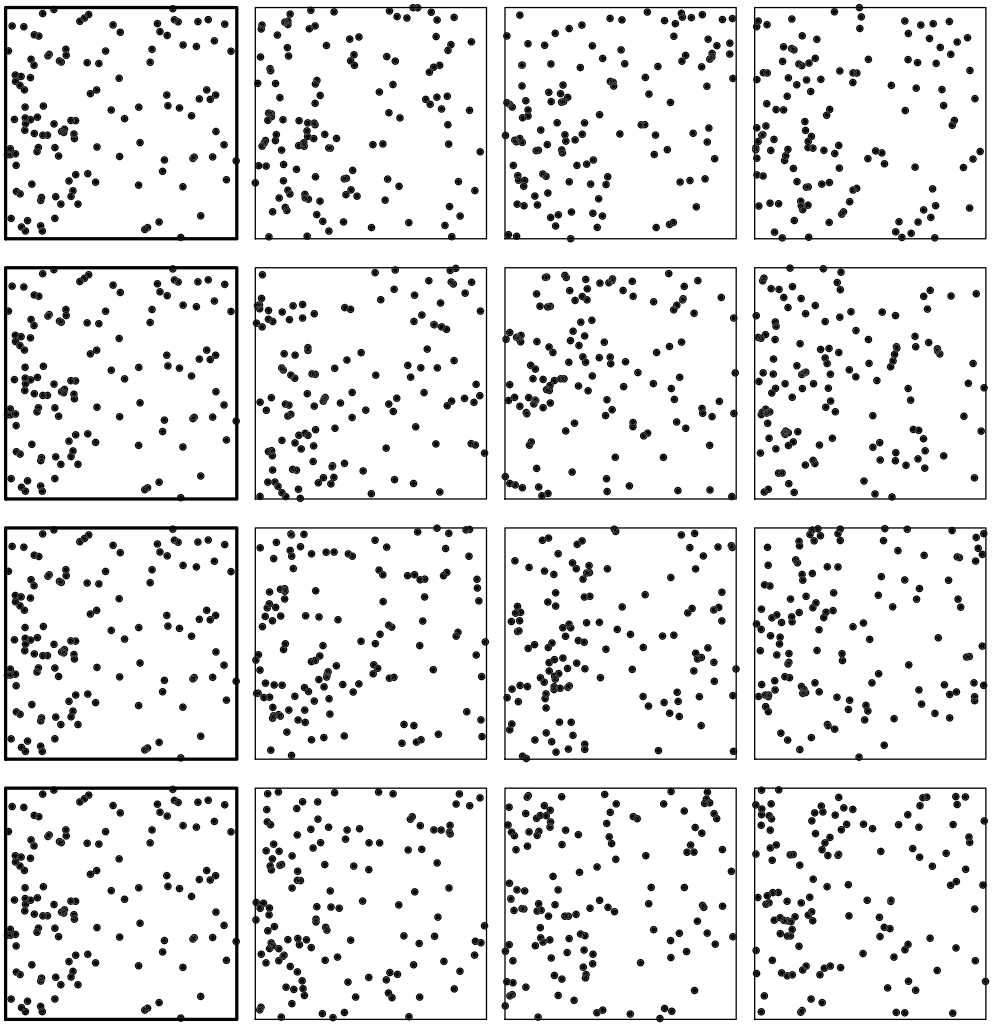

Figure 3.7: Input (on the left, same for all the rows) and three different outputs of the stochastic reconstruction algorithm based on the energy functional E_1' (first row), E_2' (second row), E_5' (third row) and E_8' (last row). For all of the rows the estimated intensity function of the outputs deviates a lot from the theoretical intensity function λ. For the energy functional E_2' (second row) the global rank envelope test based on the APF rejects more than 5% outputs, for the other rows we reject less than 5% outputs.

Before testing the outputs we should test the inputs to see whether there are some extreme configurations or not. If we perform the global rank envelope test based on the accumulated persistence function, two of the inputs can be qualified as extreme (upper bound of the p-interval equals 0.089 and 0.066 respectively). If we perform the deviation test, 4 of the 15 inputs can be qualified as extreme (p-values are equal to 0.094, 0.004, 0.072 and 0.066 respectively). We should realize that we expect to see 0.1 outputs with p-value smaller than 0.1. The four extreme inputs may cause that we will see a lot of outputs rejected while testing the hypothesis H_λ, see Table A.3 in Appendix A.3.

If we test the hypothesis H_{APF}, for almost all of the energy functionals we reject about 5% of the outputs, see Table 3.4. For the energy functional based on

L-function (E_1) the number of rejected outputs is higher, but after adding the term $\widehat{F}_\lambda$ it decreases. For both of the energy functionals based on the pair correlation function we reject more than 5% of reconstructions. It seems that E_1' and all of the functionals based on the empirical distribution functions provide reconstructions with the APF corresponding to the APF computed on simulations from the null model. For more details see Table A.3 in Appendix A.3.

The results of the deviation test which is used to test the hypothesis H_λ shows that for all of the energy functionals we reject more than one fourth of the outputs, see Table 3.5. Even though there are 4 inputs that can be qualified as extreme, the number of rejected outputs is too big. So it seems that the intensity function of the outputs does not correspond to the theoretical intensity function of the null model no matter what energy funtional we use. Adding the term $\widehat{F}_\lambda$, which controls the values of the estimated intensity function of the outputs but not the right placement of the values in the observation window does not improve in this case the quality of the reconstructions.

This observation is in contrary to the results of the previous section. While estimating intensit function for different realisations of thinned Thomas process, which is clustered, there is a large variability among the estimates. But this is no longer true for Poisson point process – here the variability of estimates from different realisations is very small and thus we should control the whole surface of the estimated intensity function of the investigated point pattern (even though we would loose the possibility of testing the hypothesis H_λ). Examples of outputs of the stochastic reconstruction algorithm based on E_1', E_2', E_5' and E_8' can be seen in Figure 3.6. We have chosen the functionals E_1' and E_2' as examples of the energy functionals based on inhomogeneous summary characteristics. E_5' and E_8' were chosen because they reject the smallest number of outputs regarding the global rank envelope test based on the accumulated persistence function.

To conclude this subsection we can say that neither the two obvious modifications of the stochastic reconstruction algorithm mentioned in Section 3.1 nor adding the term $\widehat{F}_\lambda$ suffice to produce reconstructions with the correct form of the intensity function. Also in this case it seems that the wrong form of the estimated intensity function of the outputs does not influence the results of the global rank envelope test based on the APF. Regarding this test, almost all energy functionals provide reconstructions with the correct form of the APF. Detailed results can be found in Apprendix A.3.

3.3.3 Reconstructing transformed Matérn process of type II

We have already examined the difficulties that may arise while reconstructing clustered data or a realisation of inhomogeneous Poisson process. Now we will focus on the stochastic reconstruction approach for regular data. Namely, we will try to reconstruct realisations of transformed Matérn process of type II. The definition of Matérn hard-core process of type II were stated in Definiton 14 in

Section 1.2. We will now explain how to transform this process into an inhomogeneous one.

Let Φ be a Matérn hard-core process of type II in $\mathbb{R}^2$ with the intensity of the underlying Poisson process $\kappa = 300$ and the hard-core distance $r = 0.05$. We recall that Φ is a stationary planar point process with the intensity

$$\tau = \frac{1 - e^{-\kappa\omega_2 r^2}}{\omega_2 r^2},$$

where ω_2 is the volume of the unit ball in $\mathbb{R}^2$.

Let us now have an exponential transformation function

$$c\left((x,y)\right) = \frac{1}{C_0}\left(e^{Ax} - 1\right), \quad (x,y) \in \mathbb{R}^2,$$

where $A = 1.3$ is the inhomogeneity parameter and $C_0 = e^A - 1$. Transformed point process $\Phi_{M,trans}$ is obtained by applying the transformation function c on the process Φ, i.e. $\Phi_{M,trans} = \{c(X) : X \in \Phi\}$. All of the realisations of $\Phi_{M,trans}$ will be observed on the observation window $W = [0,1]^2$. Subtracting 1 from the exponential and multiplying the formula by $\frac{1}{C_0}$ ensures that after the transformation we will have a point pattern on W. Note that Φ is a point process, i.e. a locally finite random configuration of points. If we transform the first coordinate of the points with the transformation function c, we will still have a locally finite configuration of points. Thus $\Phi_{M,trans}$ is a point process.

We are interested in the intensity function λ of the transformed process $\Phi_{M,trans}$. Let us take some $B \in \mathfrak{B}^2$. Then the intensity measure Λ of $\Phi_{M,trans}$ can be expressed as follows:

$$\begin{aligned}
\Lambda(B) &= \mathbb{E}N_{\Phi_{M,trans}}(B) \\
&= \mathbb{E}\sum_{Z\in\Phi_{M,trans}} \mathbf{1}\left\{Z \in B\right\} \\
&= \mathbb{E}\sum_{Y\in\Phi} \mathbf{1}\left\{c(Y) \in B\right\} \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbf{1}\left\{c(u,v) \in B\right\} \tau \,\mathrm{d}u \,\mathrm{d}v \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbf{1}\left\{\left(\frac{1}{C_0}\left(e^{Au} - 1\right), v\right) \in B\right\} \tau \,\mathrm{d}u \,\mathrm{d}v \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbf{1}\left\{(x,y) \in B\right\} \; \tau\frac{C_0}{Ae^{Ax}}\mathbf{1}\left\{x > -\frac{1}{C_0}\right\} \,\mathrm{d}x \,\mathrm{d}y.
\end{aligned}$$

For the fourth equation we have used the Campbell theorem, see Proposition 4.1 from Section 4.1.1 of the book Møller and Waagepetersen [2004]. In the last equation we have used the substitution $\frac{1}{C_0}\left(e^{Au} - 1\right) = x$ and $v = y$. It can be seen that the intensity function of $\Phi_{M,trans}$ has the form

$$\lambda(x,y) = \frac{\tau C_0}{Ae^{Ax}}\mathbf{1}\left\{x > -\frac{1}{C_0}\right\}, \quad (x,y) \in \mathbb{R}^2.$$

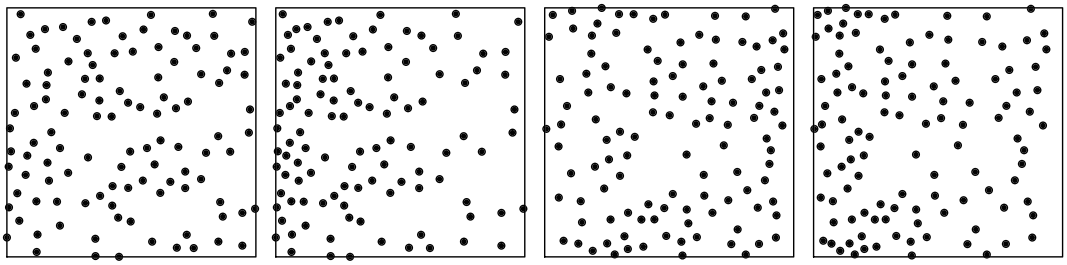

Figure 3.8: Illustration of the transformation of Matérn hard-core process of type II. First two panels: realisation φ of the Matérn hard-core process Φ on the unit square (left), φ after applying the trasformation fucntion c (right). The same situation is for the other two panels.

Set $K := \frac{\tau C_0}{240A}$. Then $K \approx 0.986$ and the intensity function of $\Phi_{M,trans}$ can be expressed as

$$\lambda(x, y) = 240Ke^{-Ax}\left\{x > -\frac{1}{C_0}\right\}, \quad (x, y) \in \mathbb{R}^2.$$

On the observation window $W = [0, 1]^2$, the indicator function takes only the value 1 and hence the intensity function λ is the same as for the thinned Thomas process from Subsection 3.3.1, see Equation (3.1). Two realisations of $\Phi_{M,trans}$ on the unit square can be seen in Figure 3.8. Note that the hard-core distance is changing in dependence on the position of a pair of points in the observation window.

As in the previous subsections we will generate 15 realisations of $\Phi_{M,trans}$ and we will use these realisations as inputs of the stochastic reconstruction algorithm. For each of the 16 energy functionals listed in Subsection 3.3.1 and each of the inputs we will produce 20 reconstructions and we will test the hypothesis H_{APF} and H_λ. Detailed results can be see in Appendix A.4.

Before testing the reconstructions let us have a look on the inputs. Concerning the global rank envelope test based on the APF, none of the inputs can be classified as extreme, see Table **??** in Appendix A.4. For the deviation test which tests the hypothesis H_λ one of the 15 inputs can be classified as extreme, the p-value is 0.047, see Table A.4 in Appendix A.4. So this time we do not have to be worried that the results of our experiment will be distorted by the influence of extreme inputs.

If we focus on the hypothesis H_{APF}, we will discover that almost all of the enrgy functionals provide reconstructions whose accumulated persistence function deviates a lot from the accumulated persistence function computed on the simulations from our null model $\Phi_{M,trans}$, see Table 3.6. The global rank envelope test rejects almost all of the 300 outputs no matter what energy functional we use. There are only two exceptions – energy functional based on L-function without the term $\widehat{F}_\lambda$ (E_1) and functional based on the pair correlation function without the term $\widehat{F}_\lambda$ (E_2). For the L-function we reject 10% of the reconstructions, for the

	Percentage	Mean	SD
E_1	10%	2.000	2.80
E_2	45%	9.067	3.28
E_3	100%	19.93	0.26
E_4	100%	20.00	0.00
E_5	71%	14.27	3.15
E_6	95%	19.00	1.20
E_7	72%	14.47	2.56
E_8	97%	19.40	1.06

	Percentage	Mean	SD
E'_1	79%	15.80	2.21
E'_2	99%	19.87	0.35
E'_3	100%	20.00	0.00
E'_4	100%	20.00	0.00
E'_5	87%	17.33	1.50
E'_6	100%	20.00	0.00
E'_7	95%	18.93	1.22
E'_8	100%	20.00	0.00

Table 3.6: Testing the hypothesis H_{APF}. Three different entries are given: percentage of the overall rejected reconstructions, mean number of rejected outputs (per one input) and the standard deviation of the number of rejected outputs (per one input). Left table corresponds to the 8 energy functionals described in Section 3.1, right table corresponds to the situation when we add the term $\widehat{F}_\lambda$ to each of the energy functionals.

pair correlation function it is 45%. Comparing to the results stated in Subsection 3.3.1, it is fairly opposite situation. The inhomogeneous characteristics give better results than the characteristics based on the interpoint dostances.

This fact can be caused by the changing hard-core distance r. The smallest hard-core distance occures in the parts of the observation window with highest values of the intensity function while the largest hadcore-distance can be observed in areas with lowest values of the intensity function. Since the characteristics based on the interpoint distances do not take into account the non-constant intensity function, they are not able to distinguish that pairs of points with small interpoint dostance are typical in some part of the observation window but they cannot occur in other parts, see Figue 3.9. It means that energy functionals based on these characteristics produce point configurations where pairs of points with small interpoint distance occures at the parts of the observation window with the largest hard-core distance. The accumulated persistence function which describes the geometrical structure of the investigated point pattern cointains the information about the changing hard-core distance and thus most of the outputs provided by the energy functionals based on the interpoint distances are rejected. Adding the term $\widehat{F}_\lambda$ to these energy functionals does not bring any improvement, the number of rejected outputs actually increases. This may be caused by the fact that we are not controling the surface (right values plus right positions) of the estimated intensity function of the outputs but only the right values that can be placed completly wrong in the observation window. Also, the choice of weights may not be righ – it is possible that by adding the term $\widehat{F}_\lambda$ we weaken the information about the geometry of the investigated point pattern contained in the energy functionals.

On the other hand the estimators of L-function and the pair correlation function deal with the non-constant intensity function and thus the energy functionals based on these characteristics are able to distinguish that in some part of the

	Percentage	Mean	SD
E_1	50%	9.93	5.15
E_2	57%	11.3	4.42
E_3	54%	10.9	5.85
E_4	27%	5.47	5.24
E_5	64%	12.8	5.62
E_6	48%	9.6	5.51
E_7	58%	11.5	5.50
E_8	30%	5.93	5.04

	Percentage	Mean	SD
E'_1	36%	7.13	5.24
E'_2	41%	8.20	5.36
E'_3	26%	5.20	4.66
E'_4	23%	4.67	4.34
E'_5	28%	5.60	5.60
E'_6	24%	4.73	4.86
E'_7	29%	5.87	4.84
E'_8	26%	5.13	4.44

Table 3.7: Testing the hypothesis H_λ. Three different entries are given: percentage of the overall rejected reconstructions, mean number of rejected outputs (per one input) and the standard deviation of the number of rejected outputs (per one input). Left table corresponds to the 8 energy functionals described in Section 3.1, right table corresponds to the situation when we add the term $\widehat{F}_\lambda$ to each of the energy functionals.

observation window smallest distances between pairs of points are typical while in other parts the hard-core distance is bigger and thus small distances between pairs of points are not allowed. If we add to these characteristics the term $\widehat{F}_\lambda$, we obtain worse results, see Table 3.6. It may be caused by the wrong choice of the weights – it may happen that we disable the effect of the pair correlation function or the L-function too much and we loose the advantage of capturing the changing hard-core distance.

Concerning the second test, i.e. the null hypothesis H_λ, we can say that all of the energy functionals produce reconstructions whose estimated intensity function does not correspond to the theoretical intensity function λ, see Table 3.7. This may be again due to the fact that there is very small variability in the estimated intensity function among simulations from the null model and even a small deviation from the theoretical intensity function may cause the rejection. We should also note that after adding the term $\widehat{F}_\lambda$ to the energy functionals the number of rejected output slightely decreases, but it is still too high. In contrast to the conclusion made in Subsection 3.3.1, while reconstructing transformed Matérn process of type II it is not enough to control the values of the estimated intensity function and not the positions of these values. Similarly to the inhomogeneous Poisson process, we should control the whole surface of the estimated intensity function on the observation window W.

Examples of the outputs of the stochastic reconstruction algorithm based on the energy functionals E_1, E_2, E'_4 and E_7 can be found in Figure 3.9. We have chosen the first two functionals because they give the best results in the global rank envelope test (nul hypothesis H_{APF}) and to illustrate that the use of inhomogeneous characteristics takes into account the non-constant hard-core distance. Functional E'_4 was chosen because it gives the best results while testing the hypothesis H_λ, even though the accumulated persistence function of the outputs does not correspond to the accumulated persistence function computed

on simulations from the null model. The last chosen functional E_7 gives the best result among the functionals based on the empirical distribution functions regarding the global rank envelope test.

To conclude this subsection we have seen that the two obvious modifications described in Section 3.1 are not enough to produce reconstructions of the transformed Matérn process with the right form of the accumulated persistence function and the estimated intensity function. The main problem is dealing with the non-constant hard-core distance and the fact that even small deviations of the surface of the estimated intensity function of outputs from the theoretical intensity function λ cause rejections of the null hypothesis. Energy functionals based on

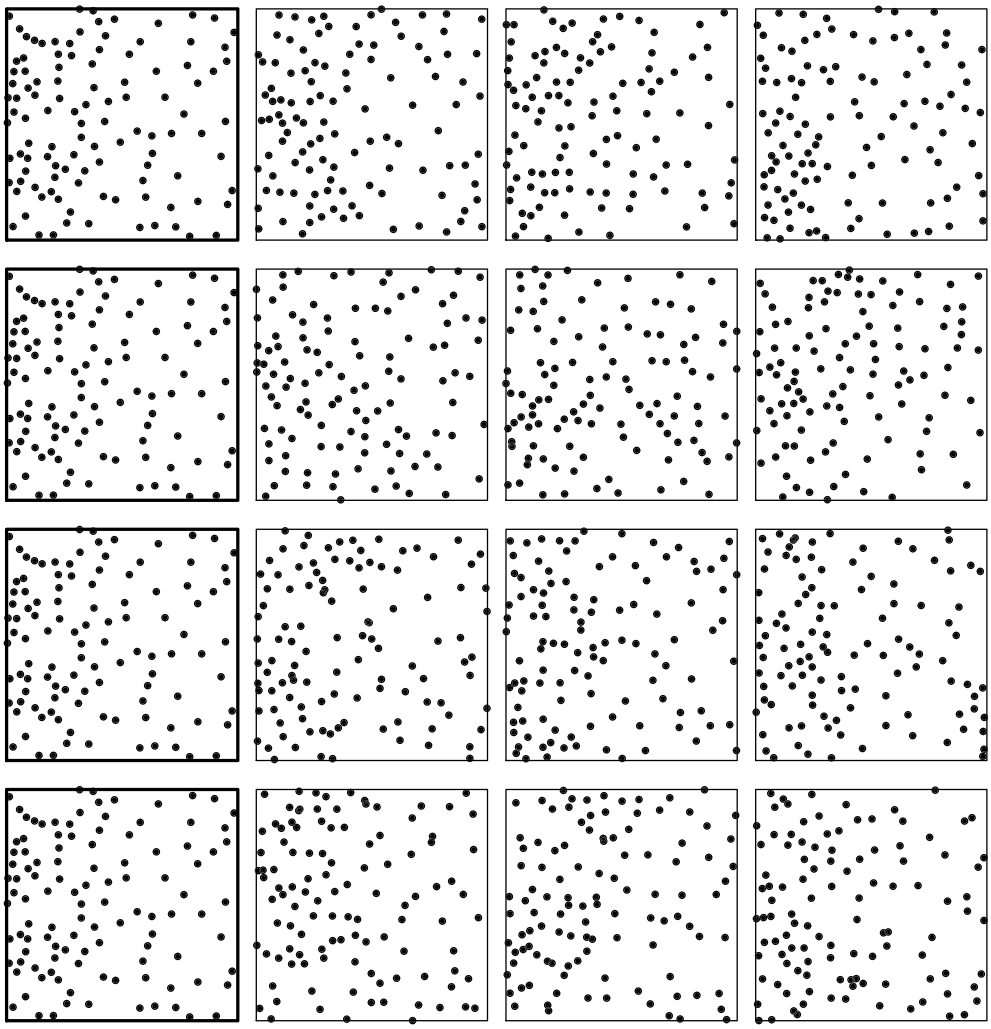

Figure 3.9: Input (on the left, same for all the rows) and three different outputs of the stochastic reconstruction algorithm based on the energy functional E_1 (first row), E_2 (second row), E_4' (third row) and E_7 (last row). For the firs two rows we should not see (in any part of the observation window) pair of points that are closer together than the non-constant hard-core distance. For the two last rows pairs of points that are very close together may appear even in the parts of the observation window where the hard-core distance is large (rightmost third of the observation window).

the inhomogeneous L-function (E_1) and the pair correlation function (E_2) give better results than the rest of the functionals. It is due to the fact that they are able to distinguish that in some parts of the observation window smallest distances between pairs of points are typical while in other parts such small distances between pairs of points are not allowed. In this situation it is needed to control the whole surface of the estimated intensity function, not only the values as it does the term $\widehat{F}_\lambda$. It is also possible that the inhomogenous binomial point process is not an appropriate initial configuration for the transforemed Matérn process and some other possibilities should be examined, e.g. a pattern uniformly randomly producing one point in each cell of the Voronoi tesstelation generated by the observed point pattern.

3.4 Summary

In this chapter we have discused the difficulties that may arise when transforming the stochastic reconstruction algorithm described in Section 2.1 for inhomogenenous point patterns.

We have stated two modifications of the algorithm that can be found in litterature, namely in Wiegand et al. [2013]. In the course of the simulation study we have compared the quality of reconstructions for 8 different energy functionals. We have chosen three theoretical models with the same intensity function – thinned Thomas process, inhomogeneous Poisson process and transformed Matérn hard-core process of type II. Realisations of these models were used as inputs of the stochastic reconstruction algorithm and we have checked whether the outputs corresponds to the theoretical models or not. More precisely, we have tested two hypotheses H_{APF} (distribution of the accumulated persistence function of output corresponds to the distribution of this characteristic under the null model) and H_λ (intensity function of output corresponds to the theoretical intensity function λ on the observation window W) using simulation-based tests.

Results of the tests have shown that the two modifications described in Section 3.1 are not sufficient, i.e. they do not assure the correspondence between outputs of the stochastic reconstruction algorithm and the observed data. For the thinned Thomas process the main problem was the intensity function. We have improved the results by adding the term $\widehat{F}_\lambda$ to each of the energy functionals. In contrast, adding the term $\widehat{F}_\lambda$ to the energy functional does not assure the correct form of the intensity function while reconstructing Poisson point process or Matérn hard-core process of type II.

It is due to the fact that the statistical test we perform control the deviation of the estimated intensity function of the outputs from the theoretical intensity function over the whole surface of the observation window W. The term $\widehat{F}_\lambda$ controls only the values of the intensity function, not their placement in the observation window. While estimating intensity function of thinned Thomas process, there is a large variability among estimates for different realisations from the model (it is due to the fact that the patterns are clustered and we use

kernel smoothing) and thus it is sufficient to control only the values and not the placement. But this is not true anymore for Poisson process or transformed Matérn hard-core process of type II – here the variability of estimates from different realisations is very small and thus we should control the whole surface of the estimated intensity function.

Additionally, for the transformed Matérn hard-core process of type II we have encountered the problem with non-constant hard-core distance. Energy functionals based on empirical distribution functions of interpoint distances do not contain enough information to distinguish that in some parts of the observation window smallest distances between pairs of points are typical while in other parts the hard-core distance is bigger and thus small distances between pairs of points are not allowed. On the other hand, the inhomegeneous L-function and the pair correlation function are able to deal with this problem.

In Section 2.1 we have claimed that one of the biggest advantages of the stochastic reconstruction approach is that we do not need to specify any theoretical model for the observed point pattern. But from Section 3.3 it is obvious that the implementation of the stochastic reconstruction algorithm requires a lot of parameters chosen by the user, e.g. the weights in the energy functional, the choice of the constants $R(\bullet)$ or parameters of some `spatstat` functions. Choice of these parameters should be based on prior knowledge about the data (including the shape and size of the observation window etc). Also the choice of the characteristics to be combined in the energy functional should be induced from some preliminary analysis of the point process data, e.g. for clustered data we should use another characteristics than for regular one.

4. Metropolis-Hastings algorithm

So far we have considered the stochastic reconstruction as an optimization problem, i.e. we have tried to find some local minima of the energy functional E defined in Section 2.1. In this chapter another approach will be discussed. We will focuse on the situation when we accept (with some probability) even the proposed configurations with higher energy. To be more concrete, the possibility of using a special case of Metropolis-Hastings algorithm for the stochastic reconstruction will be discussed. In Møller and Waagepetersen [2004] it is mentioned that it is possible to modify the improvement-only algorithm described in Section 2.1 into a special case of Metropolis-Hastings algorithm, but no futher details are given.

First basic definitions and some essential results from the theory of Markov chains will be stated. Then we will describe a special case of the Metropolis-Hastings algorithm stated in Section 7.1.1 in Møller and Waagepetersen [2004]. Also application of this algorithm to the stochastic reconstruction context will be specified including the discussion about ergodicity and the convergence properties of the Metropolis-Hastings algorithm. Subsection 4.2.2 then focuses on a short simulation study – this new approach will be used to reconstruct thinned Thomas proces.

4.1 Basic definitons and results

In the next sections we will talk about the Metropolis-Hastings algorithm. It as an algorithmic procedure that generates a Markov chain so that after a sufficiently large number of iteration steps we obtain approximately a sample from prescribed probability distribution. To understand the algorithm and its properties we have to state some basic definitions and results from the theory of Markov chains on a general state space. We emphasise that all of the definitions and theorems are taken from four different sources, namely Møller and Waagepetersen [2004], Meyn and Tweedie [1993], Roberts and Rosenthal [2004] and Tierney [1994].

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let S be a general set and denote by $\mathfrak{B}(S)$ a countably generated σ-algebra on S. We will first define a transition probability kernel.

Definition 26. *Let us have a mapping $\mathcal{P} : S \times \mathfrak{B}(S) \longrightarrow [0,1]$ such that*

- *for each $A \in \mathfrak{B}(S)$ $\mathcal{P}(\bullet, A)$ is a non-negative measurable function on S,*
- *for each $s \in S$ $\mathcal{P}(s, \bullet)$ is a probability measure on $\mathfrak{B}(S)$.*

Then we call $\mathcal{P}$ a transition probability kernel on $(S, \mathfrak{B}(S))$.

Let us now denote by $\mathbb{S}$ the product space $\mathbb{S} = \prod_{i=1}^{\infty} S$, where the symbol $\prod$ represents the Cartesian product. Also we will denote as $\mathfrak{S}$ the product σ-algebra on $\mathcal{S}$, i.e. $\mathfrak{S} = \bigotimes_{i=1}^{\infty} \mathfrak{B}(S)$.

Theorem 1. *For any probability measure μ on $\mathfrak{B}(S)$ and any transition probability kernel $\mathcal{P}$ on $(S, \mathfrak{B}(S))$, there exists a stochastic process $\mathbf{X} = \{X_0, X_1, X_2, \dots\}$ on $\mathbb{S}$ (measurable with respect to $\mathfrak{S}$) such that*

$$\begin{aligned}&\mathbb{P}\left[X_0 \in A_0, X_1 \in A_1, \dots, X_n \in A_n\right] = \\ &= \int_{A_0} \cdots \int_{A_{n-1}} \mathcal{P}(y_{n-1}, A_n)\mathcal{P}(y_{n-2}, \mathrm{d}y_{n-1}) \dots \mathcal{P}(y_0, \mathrm{d}y_1)\mu(\mathrm{d}y_0) \qquad (4.1)\end{aligned}$$

for all $A_0, A_1, \dots, A_n \in \mathfrak{B}(S)$ and $n \in \mathbb{N} \cup \{0\}$.

Proof. See Theorem 3.4.1 in Meyn and Tweedie [1993]. □

Now we have prepared the needed background to state the formal definition of a Markov chain on a general state space S.

Definition 27. *Stochastic proces $\mathbf{X}$ with a general state space S is called a time-homogeneous Markov chain with transition probability kernel $\mathcal{P}$ and initial distribution μ, if the finite dimensional distributions of $\mathbf{X}$ satisfy the Equation (4.1) for every $n \in \mathbb{N} \cup \{0\}$.*

The next step is to define the n-step transition probability kernel and the invariant measure π of a Markov chain $\mathbf{X}$.

Definition 28. *We set $\mathcal{P}^0(s, A) = \delta_s(A),\ s \in S,\ A \in \mathfrak{B}(S)$, where δ_s represents the Dirac measure. For $n \geq 1$ we define the n-th step probability kernel $\mathcal{P}^n$ inductively*

$$\mathcal{P}^n(s, A) = \int_S \mathcal{P}^{n-1}(y, A)\mathcal{P}(s, \mathrm{d}y), \quad s \in S,\ A \in \mathfrak{B}(S). \qquad (4.2)$$

Definition 29. *A σ-finite measure π on $\mathfrak{B}(S)$ is said to be invariant if*

$$\pi(A) = \int_S \mathcal{P}(s, A)\pi(\mathrm{d}s), \quad A \in \mathfrak{B}(S).$$

If π is a probability measure, then it will be called invariant (or stationary) distribution of the chain $\mathbf{X}$ with the transition probability kernel $\mathcal{P}$.

Definition 30. *The invariant distribution π is called a limiting distribution of the Markov chain $\mathbf{X}$ with the trasition probability kernel $\mathcal{P}$ if for π-almost all $s \in S$*

$$\lim_{n\to\infty} \mathcal{P}^n(s, A) = \pi(A),\ \textit{for all}\ A \in \mathfrak{B}(S).$$

A key notion is reversibility because it gives us information about invariant distributions of the chain $\mathbf{X}$.

Definition 31. *A Markov chain $\mathbf{X}$ on a state space S is reversible with respect to a probability distribution π on $\mathfrak{B}(S)$, if*

$$\pi(\mathrm{d}x)\mathcal{P}(x, \mathrm{d}y) = \pi(\mathrm{d}y)\mathcal{P}(y, \mathrm{d}x), \quad x, y \in S.$$

Theorem 2. *If the Markov chain $\mathbf{X}$ is reversible with respect to π, then π is its invariant distribution.*

Proof. See Proposition 1 in Roberts and Rosenthal [2004]. □

Another important notion in the theory of Markov chains with the general state space is the concept of irreducibility.

Definition 32. *Let η be a probability measure on $\mathfrak{B}(S)$. We say that the Markov chain $\mathbf{X}$ is η-irreducible if for any $s \in S$ and $A \in \mathfrak{B}(S)$ such that $\eta(A) > 0$ we have $\mathcal{P}^n(s, A) > 0$ for some $n \in \mathbb{N}$.*

Theorem 3. *Let η be a probability measure on $\mathfrak{B}(S)$. If an invariant distribution π exists, η-irreducibility of the chain $\mathbf{X}$ implies the following:*

- *$\mathbf{X}$ is π-irreducible,*
- *π is the unique invariant distribution.*

Proof. See Proposition 7.2 in Møller and Waagepetersen [2004]. □

Definition 33. *Suppose that the chain $\mathbf{X}$ is η-irreducible and admits an invariant distribution π. Then $\boldsymbol{X}$ is called positive.*

In order to study convergence of the Markov chain $\mathbf{X}$ we have to state the definition of Harris recurrence and aperiodicity.

Definition 34. *We say that the Markov chain $\mathbf{X}$ is Harris recurrent if it is η-irreducible (for some probability measure η) and for all $s \in S$ and $A \in \mathfrak{B}(S)$ such that $\eta(A) > 0$ we have*

$$\mathbb{P}\left[X_m \in A \text{ for some m} \mid X_0 = s\right] = 1.$$

It is known that for an ψ-irreducible chain the whole space S can be decomposed into sets A, $D_0, D_1, \ldots, D_{d-1}$ such that $\mathcal{P}(s, D_j) = 1$ for $s \in D_i$ and $j = i + 1 \mod d$ and $\psi(A) = 0$. The proof can be found in Meyn and Tweedie [1993], Theorem 5.4.4. We will denote this decomposition as $\mathcal{D}$.

Definition 35. *If there exists the decomposition $\mathcal{D}$ of the state space S with $d > 1$ then we say that the chain $\mathbf{X}$ is periodic. Otherwise it is called aperiodic.*

Definition 36. *A Markov chain $\mathbf{X}$ is said to be ergodic if it is positive Harris recurrent and aperiodic.*

Once we have defined ergodicity, we are able to formilize the convergence of the chain to its invariant distribution. But first we have to define a distance of two probability measures on $(S, \mathfrak{B}(S))$.

Definition 37. *Let μ and ν be two probability measures on $(S, \mathfrak{B}(S))$. The total variation norm si given by*

$$||\mu - \nu||_{TV} = \sup_{A \in \mathfrak{B}(S)} |\mu(A) - \nu(A)|.$$

Remark. Note that $||\mu - \nu||_{TV} \leq 1$, $||\mu - \nu||_{TV} = 0$ if $\mu = \nu$ and $||\mu - \nu||_{TV} = 1$ if μ and ν have disjoint supports, see Section 7.2.2 in Møller and Waagepetersen [2004].

Theorem 4. *Let* $\mathbf{X}$ *be an ergodic Markov chain with invariant distribution* π*. Then for all* $s \in S$ *we have*

$$||\mathcal{P}^n(s, \bullet) - \pi(\bullet)||_{TV} \longrightarrow 0, \ n \to \infty. \tag{4.3}$$

Proof. See Theorem 1 in Tierney [1994] or Theorem 13.3.3 in Meyn and Tweedie [1993]. □

Theorem 4 says that an ergodic Markov chain has its invariant distribution (if it exists) as a limiting distribution.

In order to have some information about the speed of convergence (and hence we would be able to state Central limit theorem for Markov chains) we will need a stronger property than ergodicity.

Definition 38. *A Markov chain* $\mathbf{X}$ *with invariant distribution* π *is called uniformly ergodic if* $||\mathcal{P}^n(s, \bullet) - \pi(\bullet)||_{TV}$ *converges to* 0 *uniformly in* s *for* $n \to \infty$*, e.i.*

$$\sup_{s \in S} ||\mathcal{P}^n(s, \bullet) - \pi(\bullet)||_{TV} \longrightarrow 0, \ n \to \infty.$$

Definition 39. *A Markov chain* $\mathbf{X}$ *with invariant distribution* π *is geometrically ergodic if there exists* $0 \leq \varrho < 1$ *and a real-valued function* M *on* S *such that* $M(s) < \infty$ *for* π*-almost all* $s \in S$ *and*

$$||\mathcal{P}^n(s, \bullet) - \pi(\bullet)||_{TV} \leq M(s)\varrho^n, \quad n = 1, 2, 3, \ldots$$

It is known that uniform ergodicity implies geometrical ergodicity and geometrical ergodicity implies ergodicity, see p. 1741 in Tierney [1994]. To state the necessary and sufficient conditions for uniform ergodicity, we will need the definition of a small set.

Definition 40. *A set* $C \in \mathfrak{B}(S)$ *is said to be* (m, μ)*-small if there exists* $m \in \mathbb{N}$ *and a non-zero measure* μ *on* $\mathfrak{B}(S)$ *such that*

$$\mathcal{P}^m(s, A) > \mu(A) \ \textit{for all} \ s \in C \ \textit{and} \ A \in \mathfrak{B}(S).$$

Theorem 5. *Suppose that a Markov chain* $\mathbf{X}$ *has invariant distribution* π*. Then uniform ergodicity is equivalent to that* S *is a* (m, μ)*-small set (for some* $m \in \mathbb{N}$ *and a non-zero measure* μ *on* $\mathfrak{B}(S)$*). Moreover it holds that*

$$||\mathcal{P}^n(s, \bullet) - \pi(\bullet)||_{TV} \leq (1 - \mu(S))^{n/m}, \quad n = 1, 2, 3, \ldots$$

Proof. See Proposition 7.8 in Møller and Waagepetersen [2004] and Theorem 16.2.2 in Meyn and Tweedie [1993]. □

4.2 Stochastic reconstruction based on Metropolis-Hastings algorithm

This section will be devoted to a special case of Metropolis-Hastings algorithm described in Section 7.1.1 of the book Møller and Waagepetersen [2004] and its application to the stochastic reconstruction context. Metropolis-Hastings algorithm is one of the so-called MCMC (Markov chain Monte Carlo) algorithms which permets simulations of complex stochastic systems. We have already encountered the expression "Monte-Carlo" in Section 3.2 – we have described the pointwise Monte-Carlo test.

An MCMC algorithm is a reciept for generation of a Markov chain with prescribed invariant distribution. It can be used to simulate spatial point processes defined by an unnormalised density, e.g. Gibbs point processes. MCMC algorithms can be also used to compute complex (typically multidimensional) integrals, which may arise when analysing theoretical model that was chosen for observed data. For more details see Chapter 8 in Møller and Waagepetersen [2004].

Before we start describing the algorithm we have to state a definition of a finite point process with a density. Suppose that Φ_p is a Poisson point process with finite intensity measure Λ, e.i. $\Lambda(\mathbb{R}^d) < \infty$. Recall that point process is a random element with values in measurable space $(\mathcal{N}_{lf}, \mathfrak{N}_{lf})$. Take $\mathcal{U} \in \mathfrak{N}_{lf}$. The distribution Π of a finite Poisson point process Φ_p can be expressed as follows:

$$\begin{aligned}&\Pi(\mathcal{U}) = \mathbb{P}\left[\Phi_p \in \mathcal{U}\right] = \\ &= e^{-\Lambda(\mathbb{R}^d)} \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \mathbf{1}\left\{\{x_1, \ldots, x_n\} \in \mathcal{U}\right\} \, \Lambda(\mathrm{d}x_1) \ldots \Lambda(\mathrm{d}x_n),\end{aligned}$$

see Proposition 3.1 in Møller and Waagepetersen [2004].

Assume we have $W \subset \mathbb{R}^d$ such that $|W| < \infty$. We will denote the set of finite point configurations contained in W as

$$\mathcal{N}_f^W = \left\{\zeta \subset W : \; N(\zeta) < \infty\right\}.$$

We equip $\mathcal{N}_f^W$ with a σ-algebra $\mathfrak{N}_f^W = \left\{\mathcal{U} \cap \mathcal{N}_f^W : \; \mathcal{U} \in \mathfrak{N}_{lf}\right\}$. In Definition 9 in Section 1.2 we have stated the definition of Poisson point process on $\mathbb{R}^d$. By Poisson point process on W we mean Poisson point process on $\mathbb{R}^d$ whose intensity measure Λ is concentrated on W, i.e. the support of Λ is the set W.

Definition 41. *Let us have $f : \mathcal{N}_f^W \longrightarrow [0, \infty)$ a measurable function and Φ_p be a homogeneous Poisson point process on W with the intensity $\kappa = 1$. We will say that Φ is a finite point process with density f with respect to the distribution Π of the Poisson point process Φ_p if for all $\mathcal{U} \in \mathfrak{N}_f^W$ we have*

$$\begin{aligned}&\mathbb{P}\left[\Phi \in \mathcal{U}\right] = \\ &= e^{-|W|} \sum_{n=0}^{\infty} \frac{1}{n!} \int_W \cdots \int_W \mathbf{1}\left\{\{x_1, \ldots, x_n\} \in \mathcal{U}\right\} f\left(\{x_1, \ldots, x_n\}\right) \, \mathrm{d}x_1 \ldots \mathrm{d}x_n.\end{aligned}$$

Let us have Φ a finite point process with a density. Denote by W^n the set of all n-points configurations in W, i.e.

$$W^n = \{\{x_1, x_2, \ldots, x_n\} : \ x_i \in W, \ i = 1, 2, \ldots, n, \ x_i \neq x_j \text{ whenever } i \neq j\}.$$

We will equip W^n with the σ-algebra $\mathfrak{N}_n^W = \left\{\mathcal{U} \cap W^n : \ \mathcal{U} \in \mathfrak{N}_f^W\right\}$. Suppose now that we have conditioned on $N_\Phi(W) = n$ for some integer n. Assume that under this condition, the distribution of Φ is given by

$$\begin{aligned}&\mathbb{P}\left[\Phi \in \mathcal{U} \mid N_\Phi(W) = n\right] = \\ &= \int_W \cdots \int_W \mathbf{1}\left\{\{x_1, \ldots, x_n\} \in \mathcal{U}\right\} f_n\left(\{x_1, \ldots, x_n\}\right) \, \mathrm{d}x_1 \ldots \mathrm{d}x_n,\end{aligned}$$

where $\mathcal{U} \in \mathfrak{N}_f^W$ and $f_n : \ W^n \longrightarrow [0, \infty)$ is a measurable function such that $\int_W \cdots \int_W f_n\left(\{x_1, \ldots, x_n\}\right) \, \mathrm{d}x_1 \ldots \mathrm{d}x_n = 1$. We will denote the conditional distribution of Φ as π, i.e.

$$\pi(\mathcal{U}) = \mathbb{P}\left[\Phi \in \mathcal{U} \mid N_\Phi(W) = n\right], \ \mathcal{U} \in \mathfrak{N}_f^W. \tag{4.4}$$

The distribution π will be alse refered to as target distribution.

In this moment assume that we have observed a point pattern φ on W such that $N_\varphi(W) = n$. We would like to reconstruct the patter φ using the special case of Metropolis-Hastings algorithm, namely the Algorithm 7.1, described in Section 7.1.1 in Møller and Waagepetersen [2004]. Por this purpose set

$$f_n(\zeta) = \frac{1}{G_n^W} \exp\left\{-KE(\varphi, \zeta)\right\}, \quad \zeta \in W^n, \tag{4.5}$$

where E is the energy functional defined in Section 2.1, $K > 0$ is a constant, $\frac{1}{G_n^W}$ is the normalising constant. Hence the distribution π is chosen in such a way that the configurations with low energy are more likely to occur than configurations with high energy. Our aim is to use the Metropolis-Hastings algorithm to produce an ergodic Markov chain $\mathbf{X}$ with invariant distribution π. After a sufficiently large number of iteration steps T we will obtain realisations of a point process whose distribution is approximately π.

Since we are mainly interested in stochastic reconstruction for inhomogeneous point processes, we will modify the improvement-only algorithm described in Section 3.1. Suppose that the observed point pattern φ is a realisation of an inhomogeneous point process such that its intensity function λ exists. Recall that by $\widehat{\lambda}_\varphi$ we denote the estimated intensity fuction computed for the configuration φ and $C_{W,\lambda}$ we denote the normalising constant, i.e. $C_{W,\lambda} = \int_W \widehat{\lambda}_\varphi(x) \, \mathrm{d}x$.

Now we will prepare the background for the Metropolis-Hastings algorithm from Section 7.1.1 in Møller and Waagepetersen [2004]:

- for all configurations $\zeta \in W^n$ and $i \in \{1, 2, \ldots, n\}$ set $q_i(\zeta, \bullet) = \dfrac{1}{C_{W,\lambda}} \widehat{\lambda}_\varphi(\bullet)$,

- for all $\zeta = \{x_1, x_2, \ldots, x_n\} \in W^n$, $y \in W$ and $i \in \{1, 2, \ldots, n\}$ set

$$r_i(\zeta, y) = \frac{f_n\left(\left(\zeta \setminus \{x_i\}\right) \cup \{y\}\right) q_i\left(\left(\zeta \setminus \{x_i\}\right) \cup \{y\}, x_i\right)}{f_n(\zeta) q_i(\zeta, y)},$$

  if the denominator equals 0 then we set $r_i(\zeta, y) = 1$,

- for all $\zeta \in W^n$, $y \in W$ and $i \in \{1, 2, \ldots, n\}$ set $\alpha_i(\zeta, y) = \min\{1, r_i(\zeta, y)\}$.

Remark. In the previous text we abuse a bit the notation and we write $y \in W$ even though we consider only $y \in W$ such that the configuration $(\zeta \setminus \{x_i\}) \cup \{y\}$ has exactly n-points.

In Section 7.1.1 in Møller and Waagepetersen [2004] the ratio r_i is called the Hastings ratio, q_i is called the proposal density and α_i is the acceptance probability. We will now generate the Markov chain $X_0, X_1, \ldots$ in the following way. Let X_0 be a realisation of inhomogeneous binomial point process with n points and a measure ν_φ given by the density function $\frac{1}{C_{W,\lambda}} \widehat{\lambda}_\varphi$. Suppose that after l-th iteration step we have a configuration $X_l = \{x_l^1, x_l^2, \ldots, x_l^n\}$. In the iteration step number $l + 1$ we

- with probability $\frac{1}{n}$ choose a point x_l^i of the configuration X_l to be deleted,
- generate a new point y in W with probability distribution given by the probability density $q_i(X_l, \bullet)$,
- with probability $\alpha_i(X_l, y)$ set $X_{l+1} = (X_l \setminus \{x_l^i\}) \cup \{y\}$, otherwise we set $X_{l+1} = X_l$ with probability $1 - \alpha_i(X_l, y)$.

Individual parts of each iteration step are executed independently and echach iteration step is independet on the previous steps. The algorithm is almost the same as the one from Section 3.1 – in each iteration step we choose (with equal probabilities) one point of a current configuration an move it into another part of the observation window W. The only difference is that in this case also configurations with higher energies may be accepted. In the algorithm defined in Section 3.1 it was impossible to accept a configuration with higher energy.

4.2.1 Convergence properties of the algorithm

In this Subsection we will focus on the convergence properties of the Markov chain generated by the Metropolis-Hastings algorithm that has been described above.

If we consider the stochastic reconstruction as an optimization problem (see Section 3.1), we do not ensure the convergence of the intermediate states of the algoritm to an optimal solution (local minimum). On the other hand, if we use the Metropolis-Hastings algorithm, we are able to verify whether the target distribution π is the limiting distribution. Moreover, we are able to say something about the rate of convergence.

The Markov chain produced by the Metropolis-Hastings algorithm will be refered to as the Metropolis-Hastings chain. We will first check that the Metropolis-Hastings chain is reversible with respect to the target distribution π, see Equation (4.4). Then we will show that the chain is not only ergodic but also uniformly ergodic. To prove these statements we will check the assumptions of Proposition 7.11 in Møller and Waagepetersen [2004]. We emphasise that all of the statements will be proven for a special choice of the summary characteristics used in the energy functionals.

Let us denote by $\mathcal{X}$ the natural state space of the Metropolis-Hastings chain, i.e. $\mathcal{X} = \{\{x_1, x_2, \ldots, x_n\} \subset W : \ f_n(\{x_1, x_2, \ldots, x_n\}) > 0\}$. For further details see Remark 7.1 in Møller and Waagepetersen [2004]. From Equation (4.5) it can be seen that $\mathcal{X} = W^n$. Denote by $\mathfrak{X}$ the σ-algebra on $\mathcal{X}$. Since $\mathcal{X} = W^n$ we have that

$$\mathfrak{X} = \mathfrak{N}_n^W = \left\{\mathcal{U} \cap W^n : \ \mathcal{U} \in \mathfrak{N}_f^W\right\}.$$

In Section 4.1 we have supposed that state space is equiped by countably generated σ-algebra. We should verify this assumption for $\mathfrak{X}$. In Section 1.1 we have discussed that $\mathfrak{N}_{lf}$ is countably generated. Thus $\mathfrak{N}_f^W$ is countably generated and so is $\mathfrak{X}$.

Let us now take $\zeta = \{x_1, x_2, \ldots, x_n\} \in \mathcal{X}$ and $\mathcal{U} \in \mathfrak{X}$. The transition probability kernel $\mathcal{P}$ of the Metropolis-Hastings chain then has the form

$$\mathcal{P}(\zeta, \mathcal{U}) = \frac{1}{n}\sum_{i=1}^{n} \int_W \mathbf{1}\left\{(\zeta \setminus \{x_i\}) \cup \{y\} \in \mathcal{U}\right\} \ q_i(\zeta, y) \ \alpha_i(\zeta, y) \ \mathrm{d}y \ + \ r(\zeta)\delta_\zeta(\mathcal{U}),$$

where δ is the Dirac measure and

$$r(\zeta) = \frac{1}{n}\sum_{i=1}^{n} \int_W \mathbf{1}\left\{(\zeta \setminus \{x_i\}) \cup \{y\} \in \mathcal{U}\right\} \ q_i(\zeta, y) \ (1 - \alpha_i(\zeta, y)) \ \mathrm{d}y$$

is the probability that we do not leave the configuration ζ within one iteration step. We will consider another Markov chain with transition probability kernel $\mathcal{Q}$ given by

$$\mathcal{Q}(\zeta, \mathcal{U}) = \frac{1}{n}\sum_{i=1}^{n} \int_W \mathbf{1}\left\{(\zeta \setminus \{x_i\}) \cup \{y\} \in \mathcal{U}\right\} \ q_i(\zeta, y) \ \mathrm{d}y.$$

This chain will be refered to as the proposal chain. This transition probability kernel corresponds to the situation when we always accept the proposed configuration. Now we have prepared the background to state the theorem about the convergence of the Metropolis-Hastings chain.

Theorem 6. *The following properties hold for the Metropolis-Hastings algorithm described in Section 4.2.*

1. *The Metropolis-Hastings chain produced by the algorithm described in Section 4.2 is reversible with respect to the target distribution π.*

2. *Suppose that $\mathcal{X} = W^n$ and the proposal chain with transition probability kernel $\mathcal{Q}$ is η-irreducible for some η probability measure on $\mathfrak{X}$. Assume that for all $\zeta = \{x_1, x_2, \ldots, x_n\} \in W^n$, $y \in W$ and $i \in \{1, 2, \ldots, n\}$ it holds that*

$$q_i(\zeta, y) > 0 \implies q_i\left((\zeta \setminus \{x_i\}) \cup \{y\}, x_i\right) > 0.$$

Then the Metropolis-Hastings chain is η-irreducible.

3. *Suppose that the Metropolis-Hastings chain is η-irreducible (for some η probability measure on $\mathcal{X}$) and that there exists $\epsilon > 0$, $x_2, \ldots, x_n \in W$ and $D \subseteq W$ such that $|D| > 0$ and for all $x_1, y \in D$ we have that $f_n(\{x_1, x_2, \ldots, x_n\}) > 0$ and*

$$\min\left\{q_i\left(\{x_1, \ldots, x_n\}, y\right), \frac{f_n\left(\{y, x_2, \ldots, x_n\}\right) q_i\left(\{y, x_2, \ldots, x_n\}, x_1\right)}{f_n\left(\{x_1, x_2, \ldots, x_n\}\right)}\right\} \geq \epsilon. \tag{4.6}$$

Then $C = \{\{x_1, x_2, \ldots, x_n\} : x_1 \in D\}$ is a small set and the Metropolis-Hastings chain is aperiodic.

4. *Suppose that $\mathcal{X} = W^n$ and that there exists $\epsilon > 0$ such that for all $\zeta = \{x_1, x_2, \ldots, x_n\} \in W^n$, $i \in \{1, 2, \ldots, n\}$ and $y \in W$ we have*

$$\min\left\{q_i(\zeta, y), \frac{f_n\left((\zeta \setminus \{x_i\}) \cup \{y\}\right) q_i\left((\zeta \setminus \{x_i\}) \cup \{y\}, x_i\right)}{f_n(\zeta)}\right\} \geq \epsilon.$$

Then the Metropolis-Hastings chain is uniformly ergodic and

$$||\mathcal{P}^m(\zeta, \bullet) - \pi(\bullet)||_{TV} \leq \left(1 - n!\left(\frac{\epsilon|W|}{n}\right)^n\right)^{\frac{m}{n}}, \quad m = 1, 2, \ldots$$

Remark. As it was mentionde in Section 4.2 we suppose that all of the configurations mentioned in the statements of Theorem 6 have n points, even though it is not explicitely written.

Proof. See Proposition 7.11 in Section 7.3.1 of the book Møller and Waagepetersen [2004]. □

First claim of the Theorem 6 says (without any assumptions) that the Metropolis-Hastings chain generated by the algorithm described in Section 4.2 is reversible with respect to the target distribution π given by the Equation 4.5. It means that π is the invariant distribution of the Metropolis-Hastings chain.

Now we will check the assuptions of the second claim of Theorem 6.We have already discussed that $\mathcal{X} = W^n$. It is due to the choice of f_n, see Equation (4.5). Let us now take η the distribution of a binomial point process on W with n points and the d-dimensional Lebesgue measure. It means that for any $\mathcal{U} \in \mathfrak{X}$ we have

$$\eta(\mathcal{U}) = \int_W \cdots \int_W \mathbf{1}\left\{\{x_1, \ldots, x_n\} \in \mathcal{U}\right\} \, \mathrm{d}x_1 \ldots \mathrm{d}x_n.$$

We aim to prove that for any $\zeta \in \mathcal{X}$ and $\mathcal{U} \in \mathfrak{X}$ such that $\eta(\mathcal{U}) > 0$ we have $\mathcal{Q}^m(\zeta, \mathcal{U}) > 0$ for some $m \in \mathbb{N}$.

It is important to realize how the n-th probability kernel $\mathcal{Q}^m$ looks like. Let us compute $\mathcal{Q}^2$. Take $\zeta = \{x_1, x_2, \ldots, x_n\} \in \mathcal{X}$ and $\mathcal{U} \in \mathfrak{X}$.

$$\mathcal{Q}^2(\zeta,\mathcal{U}) = \int_{\mathcal{X}} \mathcal{Q}(\xi,\mathcal{U})\mathcal{Q}(\zeta,\mathrm{d}\xi) = \frac{1}{n^2}\sum_{i=1}^{n}\int_W\int_W A(s,z)\; q_i(\zeta,y)q_i(\xi,z)\;\mathrm{d}y\;\mathrm{d}z,$$

where

$$A(s,z) = \sum_{s\in(\zeta\setminus\{x_i\})\cup\{y\},s\neq y} \mathbf{1}\left\{(\xi\setminus\{s\})\cup\{z\}\in\mathcal{U},\; \xi = (\zeta\setminus\{x_i\})\cup\{y\}\right\}.$$

It is important to realize that $Q(\zeta,\mathcal{U})$ is the probability that the chain moves in one step from configuration ζ to some configuration from $\mathcal{U}$. $Q^2(\zeta,\mathcal{U})$ is then the probability that the chain moves in two steps from the configuration ζ to some configuration from $\mathcal{U}$, i.e. we first move one point of the configuration ζ to a new location in W and then we ask if it is possible to move from the new configuration to some configuration in $\mathcal{U}$.

If we have the configuration ζ we can move 1 or 2 or even more points in order to obtain a configuration from $\mathcal{U}$. Moreover, if we move all of the n points of the configuration ζ, we will end up with a completly new configuration of n points in W. Therefore, we are always able to move some of the points of the configuration ζ (or all of them) so that we obtain a configuration from $\mathcal{U}$. Thus we will always find $m\in\mathbb{N}$ so that $\mathcal{Q}^m(\zeta,\mathcal{U})>0$. We have thus proved that the proposal chain is η-irreducible.

It remains to check the condition on the proposal density q_i. Recall that for all $\zeta=\{x_1,\ldots,x_n\}\in\mathcal{X}$ and all $i\in\{1,2,\ldots,n\}$ we have $q_i(\zeta,\bullet) = \dfrac{1}{C_{W,\lambda}}\widehat{\lambda}_\varphi(\bullet)$. Assume that for all $y\in W$ we have $\widehat{\lambda}_\varphi(y)>0$ (if we use the Gaussian kernel to estimate the intensity , which is the standard choice, this condition will be fulfilled for $n\geq 1$). Than the condition

$$q_i(\zeta,y)>0 \implies q_i\left((\zeta\setminus\{x_i\})\cup\{y\},x_i\right)>0$$

is trivially fullfiled. We have thus shown that the Metropolis-Hastings chain is η-irreducible.

Now we aim to verify the assumptions of the third statement of Theorem 6. Our choice of the density f_n ensures that the condition $f_n\left(\{x_1,x_2,\ldots,x_n\}\right)>0$ is trivialy fullfiled for all n-point configurations in W. The observed pattern φ consists of n points, thus we can write $\varphi=\{s_1,s_2,\ldots,s_n\}$. We can take $x_i=s_i,\; i\in\{2,3,\ldots,n\}$ and $D=W$.

We estimete the intensity function λ using the Gaussian probability kernel (which is the most often choice) and thus there exist a finite constants $M_\lambda>0$ so that for all configurations $\zeta'\in W^n$ and for all $x\in W$

$$\widehat{\lambda}_{\zeta'}(x)\geq M_\lambda.$$

Thus $\widehat{\lambda}_\varphi$ is bounded away from 0 on W. We can thus write

$$q_i\left(\{x_1,s_2,\ldots,s_n\},y\right) = \frac{1}{C_{W,\lambda}}\widehat{\lambda}_\varphi(y)\geq\frac{1}{C_{W,\lambda}}M_\lambda>0.$$

In the same way we can estimate the term $q_i\left(\{y, s_2, \ldots, s_n\}, x_1\right)$.

We will now focus on the ratio

$$\frac{f_n\left(\{y, s_2, \ldots, s_n\}\right)}{f_n\left(\{x_1, s_2, \ldots, s_n\}\right)}.$$

If we use the Equation (4.5), we obtain

$$\frac{f_n\left(\{y, s_2, \ldots, s_n\}\right)}{f_n\left(\{x_1, s_2, \ldots, s_n\}\right)} = \frac{\exp\left\{-KE\left(\varphi, \{y, s_2, \ldots, s_n\}\right)\right\}}{\exp\left\{-KE\left(\varphi, \{x_1, s_2, \ldots, s_n\}\right)\right\}}. \tag{4.7}$$

We can rewrite the ratio from Equation 4.7 to the form

$$\exp\left\{-K\left[E\left(\varphi, \{y, s_2, \ldots, s_n\}\right) - E\left(\varphi, \{x_1, s_2, \ldots, s_n\}\right)\right]\right\}.$$

Let us now estimate the difference

$$B = E\left(\varphi, \{y, s_2, \ldots, s_n\}\right) - E\left(\varphi, \{x_1, s_2, \ldots, s_n\}\right).$$

from above in order to estimate the ratio from Equation 4.7 from below. Suppose that the energy functional is based on a functional summary characteristic $S(r),\ r \in (0, R(S)]$. We also suppose that the estimator $\widehat{S}(\bullet, r)$ is a non-negative function. This assumption involve all of the characteristics we have considered in Section 3.1. But it excludes some characteristics, for example the centerd L-function $L(r) - r,\ r > 0$. Using the Equation (2.1) we obtain

$$\begin{aligned}
B &= \int_0^{R(S)} \left[\widehat{S}\left(\varphi, r\right) - \widehat{S}\left(\{y, s_2, \ldots, s_n\}, r\right)\right]^2 - \\
&\quad - \left[\widehat{S}\left(\varphi, r\right) - \widehat{S}\left(\{x_1, s_2, \ldots, s_n\}, r\right)\right]^2 \mathrm{d}r \\
&= \int_0^{R(S)} \left[2\widehat{S}\left(\varphi, r\right) - \widehat{S}\left(\{y, s_2, \ldots, s_n\}, r\right) - \widehat{S}\left(\{x_1, s_2, \ldots, s_n\}, r\right)\right] \times \\
&\quad \times \left[\widehat{S}\left(\{x_1, s_2, \ldots, s_n\}, r\right) - \widehat{S}\left(\{y, s_2, \ldots, s_n\}, r\right)\right]\ \mathrm{d}r \\
&\leq \int_0^{R(S)} \left[2\widehat{S}\left(\varphi, r\right)\right] \left[\widehat{S}\left(\{x_1, s_2, \ldots, s_n\}, r\right) - \widehat{S}\left(\{y, s_2, \ldots, s_n\}, r\right)\right]\ \mathrm{d}r
\end{aligned}$$

In the second equation we have used that $a^2 - b^2 = (a+b)(a-b)$. We will now estimate the difference $\left[\widehat{S}\left(\{x_1, s_2, \ldots, s_n\}, r\right) - \widehat{S}\left(\{y, s_2, \ldots, s_n\}, r\right)\right]$. We will consider three different situations, that correspond to the characteristics we will use in Subsection 4.2.2. For other characteristics a different proceeding may be required. Recall that φ is one particular realisation from the theoretical model Φ and we suppose that $\widehat{S}(\varphi, r),\ r \in (0, R(S)]$ is bounded.

First suppose that $\widehat{S}(\bullet, r)$ is an empirical distribution function. Then $\left[\widehat{S}\left(\{x_1, s_2, \ldots, s_n\}, r\right) - \widehat{S}\left(\{y, s_2, \ldots, s_n\}, r\right)\right] \leq 1$. Therefore

$$B \leq \int_0^{R(S)} 2\widehat{S}\left(\varphi, r\right)\ \mathrm{d}r = B_{edf}.$$

Here we integrate a multiple of an empirical distribution function which is bounded over a bounded interval, hence the constant B_{edf} is finite.

Consider now that $\widehat{S}(\bullet, r)$ is the estimator of the inhomogeneous pair correlation function. Denote by ζ the configuration $\{x_1, s_2, \ldots, s_n\}$ and by ξ the configuration $\{y, s_2, \ldots, s_n\}$. We will now estimate the difference $\widehat{S}(\zeta, r) - \widehat{S}(\xi, r)$. Using the Equation (1.2) we obtain

$$\begin{aligned} A &= \widehat{S}(\zeta, r) - \widehat{S}(\xi, r) = \\ &= \sum_{x,z\in\zeta}^{\neq} \frac{u_b(r - \|x - z\|)}{\widehat{\lambda}_\zeta(x)\widehat{\lambda}_\zeta(z)\sigma_d r^{d-1}|W|} e_{W,r}(x, z) - \sum_{v,t\in\xi}^{\neq} \frac{u_b(r - \|v - t\|)}{\widehat{\lambda}_\xi(v)\widehat{\lambda}_\xi(t)\sigma_d r^{d-1}|W|} e_{W,r}(v, t), \end{aligned}$$

where u_b is the kernel function with bandwidth b and $e_{W,r}$ is the edge correction factor. First note that both of the sums are non-negative, thus we can estimate the difference from above by summing up these two terms. We should realize that we work only with such kernel fuctions that are bounded. It means that there exists a finite number $M^{u_b}_{upp} > 0$ such that for all $w_1, w_2 \in W$ and for all values of r

$$u_b(r - ||w_1 - w_2||) \leq M^{u_b}_{upp}.$$

Recall that for the pair correlation function we work with a kernel function with bounded support (e.g. the Epanechnikov kernel). We use the pair correlation function to investigate the interactions among pairs of points. So while using this characteristic, we should have some idea about the maximal range of interactions we want to investigate. Based on this information, we choose the size of the observation window (it should be few times bigger than the maximal range of interactions) and the constant $R(S)$. Also, while choosing the bandwidth b of the kernel function u_b the information about the range of interactions is taken into account. So if we observe a pair of points with distance bigger than the maximal range of interactions (such pairs of points are not of particular interest, if we wanted to explore so large range of interactions, we would take bigger observation window and we would choose the constant $R(S)$ differently), the kernel function u_b will be zero. It is due to the choice of the bandwidth b and the bounded support of u_b.

Now focus on the term $e_{W,r}(\bullet, \bullet)$. Imagine for simplicity that we have a translation edge correction. If we observe a pair of points (w_1, w_2) with a very large distance (e.g. they lie close to the different edges of the observation window), the ege correction factor

$$\frac{|W|}{|W \cap (W + w_1 - w_2)|}$$

may be close to infinity. But for such a pair of points the kernel function u_b is zero. On the other hand, if we have a pair of points (w_1, w_2) for which u_b is not zero, than the edge correction factor $e_{W,r}(w_1, w_2)$ will be bounded. Therefore, there exists a finite number $M_{upp} > 0$ such that for all $w_1, w_2 \in W$ and for all $r \in (0, R(S)]$ the producht $u_b(r - ||w_1 - w_2||)\, e_{W,r}(w_1, w_2)$ is bounded, i.e.

$$u_b(r - ||w_1 - w_2||)\, e_{W,r}(w_1, w_2) \leq M_{upp}.$$

We can thus write

$$A \leq \frac{M_{upp}}{\sigma_d r^{d-1}|W|}\left[\sum_{x,z\in\zeta}^{\neq}\frac{1}{\widehat{\lambda}_\zeta(x)\widehat{\lambda}_\zeta(z)}+\sum_{v,t\in\xi}^{\neq}\frac{1}{\widehat{\lambda}_\xi(v)\widehat{\lambda}_\xi(t)}\right].$$

We have already mentioned that we estimate the intensity function in such a way that $\widehat{\lambda}'_\zeta$ is bounded away form 0 on W irrespective the configuration ζ'. Thus we can write

$$A \leq \frac{2mM_{upp}}{M_\lambda^2\sigma_d r^{d-1}|W|},$$

where $m = \binom{n}{2}$ is the number of pairs of points in an n-point configuration. Finally we obtain

$$B \leq \int_0^{R(S)} \frac{2mM_{upp}}{M_\lambda^2\sigma_d r^{d-1}|W|} 2\widehat{S}\left(\varphi, r\right)\ \mathrm{d}r = B_g.$$

Note that we integrate a multiple of a bounded function $\widehat{S}(\varphi, r)$ over a bounded interval, hence the constant B_g is finite.

Finally, suppose that $\widehat{S}(\bullet, r)$ is the estimator of the inhomogeneous L-function. Assume that $d = 2$ (in Subsection 4.2.2 we will work only with planar point processes). Thus we have to estimate

$$\begin{aligned}\widehat{S}(\zeta, r) - \widehat{S}(\xi, r) &= \left(\frac{\widehat{K}(\zeta, r)}{\omega_2}\right)^{\frac{1}{2}} - \left(\frac{\widehat{K}(\xi, r)}{\omega_2}\right)^{\frac{1}{2}} \\ &\leq \left(\frac{\widehat{K}(\zeta, r)}{\omega_2}\right)^{\frac{1}{2}} + \left(\frac{\widehat{K}(\xi, r)}{\omega_2}\right)^{\frac{1}{2}}.\end{aligned}$$

We can use the second inequation since $\widehat{K}(\zeta, r)$ and $\widehat{K}(\xi, r)$ are non-negative. The sum can be estimated using

$$\begin{aligned}&1 \leq \sqrt{x} \leq x, \quad x \geq 1,\\ &0 \leq \sqrt{x} \leq 1, \quad 0 \leq x < 1.\end{aligned}$$

The terms $\widehat{K}(\zeta, r)$ or $\widehat{K}(\xi, r)$ then can be estimated in a very similar way as for the pair correlation function. In this case we do not have to deal with the kernel function u_b, we have just the indicator function instead.

So we have proved that the difference $E\left(\varphi, \{y, s_2, \ldots, s_n\}\right) - E\left(\varphi, \{x_1, s_2, \ldots, s_n\}\right)$ is (at least for the three different cases discussed above) bounded from above by some constant which will be denoted by $B_\bullet$. Thus

$$\frac{f_n\left(\{y, s_2, \ldots, s_n\}\right)}{f_n\left(\{x_1, s_2, \ldots, s_n\}\right)} = \frac{\exp\left\{-KE\left(\varphi, \{y, s_2, \ldots, s_n\}\right)\right\}}{\exp\left\{-KE\left(\varphi, \{x_1, s_2, \ldots, s_n\}\right)\right\}} \geq \exp\left\{-KB_\bullet\right\}.$$

Thus

$$\frac{f_n\left(\{y, x_2, \ldots, x_n\}\right) q_i\left(\{y, x_2, \ldots, x_n\}, x_1\right)}{f_n\left(\{x_1, x_2, \ldots, x_n\}\right)} \geq \frac{1}{C_{W,\lambda}} \min_{z\in W} \widehat{\lambda}_\varphi(z) \exp\left\{-KB_\bullet\right\}.$$

If we realize that $0 < \exp\{-KB_\bullet\} \leq 1$ we can put

$$\epsilon = \frac{1}{C_{W,\lambda}} \min_{z \in W} \widehat{\lambda}_\varphi(z) \exp\{-KB_\bullet\}$$

and we will have that the minimum from Equation (4.6) is always greater or equal to ϵ. Hence we have proven that $C = \{\{x_1, s_2, \ldots, s_n\} : \ x_1 \in D\}$ is a small set and the Metropolis-Hastings chain is aperiodic.

It is important to realize that in course of the proof of the third statement of Theorem 6 we have not used the fact that $x_2, \ldots, x_n$ from the definition of the set C were chosen as $s_2, \ldots, s_n$, i.e. the points of the configuration φ. Moreover, the set D was chosen as the whole observation window W. Also, q_i does not depend on the choice of i. Thus the fourth statement of Theorem 6 which says that there exists $\epsilon > 0$ such that for all $\zeta = \{x_1, x_2, \ldots, x_n\} \in W^n$ and for all $i \in \{1, 2, \ldots, n\}$ and $y \in W$ we have

$$\min\left\{q_i(\zeta, y), \frac{f_n((\zeta \setminus \{x_i\}) \cup \{y\})\, q_i((\zeta \setminus \{x_i\}) \cup \{y\}, x_i)}{f_n(\zeta)}\right\} \geq \epsilon$$

can be proven using the same ideas. Thus we obtain that the Metropolis-Hastings chain is uniformly ergodic and

$$||\mathcal{P}^m(\zeta, \bullet) - \pi(\bullet)||_{TV} \leq \left(1 - n!\left(\frac{\epsilon|W|}{n}\right)^n\right)^{\frac{m}{n}}, \quad m = 1, 2, \ldots$$

While proving the uniform ergodicity, we should try to find the constant ϵ in such a way that we get the best upper bound of the rate of convergence. Above we have just given some general ideas how to deal with some of the terms that have to be estimated. Better results can be obtained while working with particular kernel function and the observaion window W. To conclude, we have given the sketch of the proof (for a special choice of the summary characteristics) that the Metropolis-Hastings algorithm described in Section 4.2 converges to the target distribution π which prefers the configurations with low energy. Moreover, we have some idea about the rate of convergence.

4.2.2 Example – reconstructing thinned Thomas process

In this subsection we will demonstrate the stochastic reconstruction approach using the Matropolis-Hastings algorithm on a theoretical example – reconstruction of thinned Thomas process. Thinned Thomas process was defined in Subsection 3.3.1. So assume that we have $\Phi_{T,thin}$ thinned Thomas process with non-constant intensity function λ given by the Equation (3.1).

As inputs of the algorithm we will take the 15 realisations of $\Phi_{T,thin}$ on the observation window $W = [0, 1]^2$ from Subsection 3.3.1. We will use three different energy functionals to produce the reconstructions. In the Subsection 3.3.1 we have discovered that inhomogeneous initial configuration and non-uniform distribution of the proposed new points is not enough to provide reconstructions that

corresponds to the observed data. The main problem was the intensity function. For thinned Thomas process it was sufficient to add the term $\widehat{F}_\lambda$, which is the empirical distribution function of values of the estimated intensity function computed in vertices of a pixel grid in the observarion window W, to energy functional. Thus we have chosen the energy functional only from the 8 variants which contain the term $\widehat{F}_\lambda$. To compare the use of inhomogeneous summary characteristics and the empirical distribution function of distances to k-th nearest neighbours (which contain detailed information about the geometry of the clusters but do not take into account the non-constant intensity function), we have chosen the following functionals:

$$\begin{aligned} E_1' &\dots \text{ inhomogeneous } L\text{-function} + \widehat{F}_\lambda \\ E_2' &\dots \text{ inhomogeneous pair correlationfunction} + \widehat{F}_\lambda \\ E_5' &\dots \widehat{D}_1 \text{ up to } \widehat{D}_5 + \widehat{F}_\lambda \end{aligned}$$

Note that there is a big difference between the improvement-only algorithm from Section 3.1 and the Metropolis-Hastings algorithm described in Section 4.2. While using the improvemet-only approach we aim to reach different local minima of the energy functional. We need to run the algorithm 20 times to obtain 20 outputs and we suppose that these 20 outputs come from different local minima of the energy functional. On the other hand, we have proved that (for energy functionals based on summary statistics mentioned in Subsection 4.2.1) the outputs of the Metropolis-Hastings algorithm approximate (after a sufficiently large number of iteration steps) sample from the invariant distribution π. The density f_n of the target distribution π have been chosen so that configurations with low energy are more likely to occur. This density function is, however, positive on the set of all n-point configurations in the observation window W. The choice of the density f_n and the constant K from Equation 4.5 influences the dynamic on the chain. We will now investigate whether the samples we obtain by running the Metropolis-Hastings algorithm corresponds to the theoretical model $\Phi_{T,thin}$ in the sens of the two statistical tests from Section 3.2.

For each of the 15 inputs and each of the selected energy functionals we will generate 20 reconstruction. Then we will test the hypotheses H_{APF} (global rank envelope test) and H_λ (deviation test), that were described in detail in Section 3.2. Roughly speaking, we will investigate two things. First we will ask whether the distribution of the accumulated persistence function of output of the stochastic reconstruction algorithm (with realisations of $\Phi_{T,thin}$ as input) corresponds to the distribution under the null model $\Phi_{T,thin}$. Then we will investigate whether the intensity function of the output corresponds to the theoretical intensity function of the null model $\Phi_{T,thin}$. The extremness of the inputs themselves was discussed in Subcection 3.3.1. Moreover, results of the tests for inputs can be found in Appendix A.5.

While using the Metropolis-Hastings algorithm the choice of the constant K in the formula for the probability density of target distribution π (see Equation 4.5) is crucial. The bigger the constant K is, the more we force the correspondence

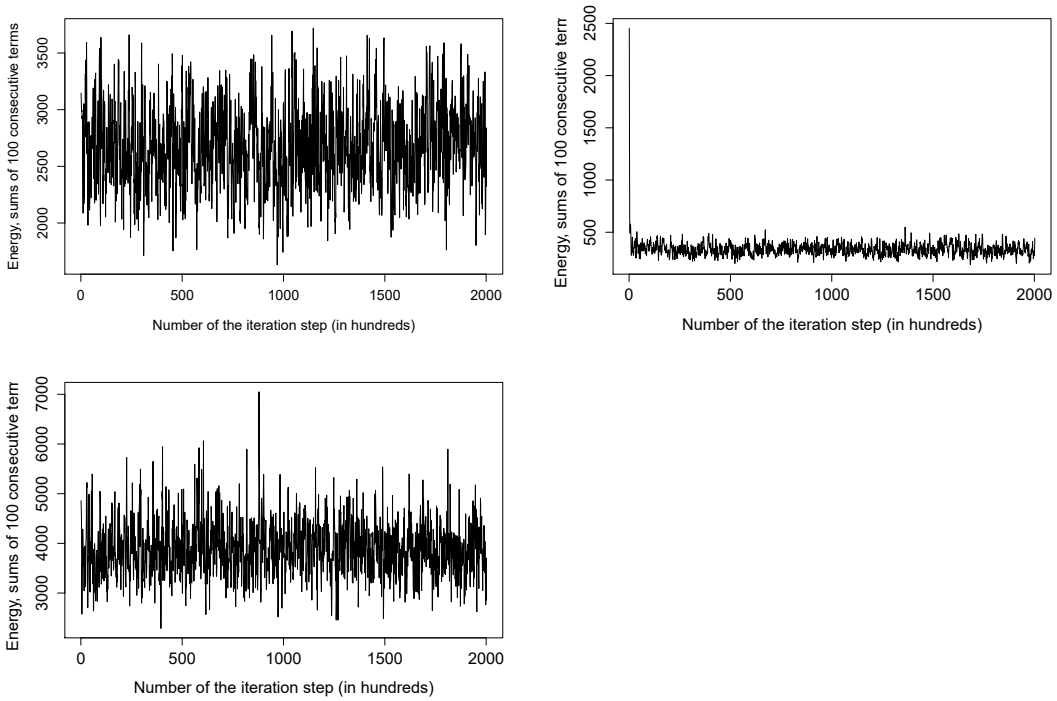

Figure 4.1: Evolution of the values of the energy functional E_5' in dependence on the choice of the constant K (sums of values of the energy functional for 100 consecutive terms are plotted, 200000 iteration steps were executed): $K = 1$ (top left corner), $K = 10$ (top right corner) and $K = 50$ (bottom left corner). For $K = 1$ and $K = 50$ the chain is exploring a region of the state space $\mathcal{X}$ with rather high values of the enrgy functional and in the 200000 executed iteration steps it does not move into a region with lower energy. For $K = 10$ the values of the energy functional decreases rapidly at the begining of the run and then they oscilates around relatively small numbers.

between the estimated summary characteristics of input and outputs. Also, the bigger K is, the smaller the values of the energy functional are. We will illustrate this fact on the energy functional E_5'. We have executed severel run of the algorithm with 200000 iterations.

If we take $K = 1$, the algorithm explores a region of the state space $\mathcal{X}$ with rather high values of the enrgy functional and it seems to be impossible to reach (in the 200000 executed itaration steps) regions with low energy. If we take $K = 10$ the energy decreases rapidly in the first few hundred of iteration steps and then oscilates around relatively small values. But if we take $K = 50$, the same situation as for $K = 1$ occurs. Different choices of K are illustrated in Figure 4.1. If we have some prior knowledge about the desired values of the energy functional (here we have some experience from the improvement-only algorithm from Subsection 3.3.1, we know how small the energy of outputs was), we can run the algorithm wiht different values of K and controle if the algorithm explores also the states with reasonably small energy. But this experimental approach is inapplicable when reconstructing real data, since we have no prior information about the values of the energy functional. So the choice of the constant K in the

	Percentage	Mean	SD
E'_1	98%	19.6	0.83
E'_2	93%	18.5	3.31
E'_5	29%	5.73	6.69

	Percentage	Mean	SD
E'_1	25%	5.00	6.16
E'_2	33%	6.67	6.62
E'_5	44%	8.80	6.81

Table 4.1: Testing the hypotheses H_{APF} and H_λ. Three different entries are given: percentage of the overall rejected reconstructions, mean number of rejected outputs (per one input) and the standard deviation of the number of rejected outputs (per one input). Left table corresponds to the global rank envelope test and hence the hypothesis H_{APF}, right table corresponds to the deviation test and thus the hypothesis H_λ.

target distribution of Metropolis-Hastings algorithm makes this approach very inconvenient for practical use.

After running some experiments we have chosen to take $K = 10$ for E'_5, $K = 0.9$ for E'_2 and $K = 150$ for E'_1. For all of these choices the values of the energy functional decrease rapidly at the begining of the run of the algorithm. After that the energy oscillates between relatively small values. We have thus decided to take the first output after running only 10000 iteration steps and then we sample each 10000-th iteration step. After running 200000 iterations we will thus obtain 20 samples.

Results of the two statistical tests we have performed are briefly summarised in Table 4.1. Detailed results can be found in Appendix A.5, see Table A.7 and A.8.

First we will focus on the energy functional E'_5. If we test the hypothesis H_{APF} this functional gives better results than E'_1 and E'_2. But if we compare this approach and the improvement-only algorithm (see Table 3.2, E'_5 gives one of the best results) we can see that for the Metropolis-Hastings algorithm we reject more reconstructions. The same situation occurs while testing the hypothesis H_λ (see Table 3.3 to compare). Moreover, if we test the hypothesis H_λ, E'_5 gives the worst results from all of the considered energy functionals, see Table 4.1. Recall that we have set the contant $K = 10$. After several hundereds of iteration steps the values of the energy functional stabilise around 1.5. This number cannot be deduced from Figure 4.1 where the sums of 100 consecutive terms are plotted. Plotting all the terms would make the figures more obscure. During our experiments we have examined the evolution of the values of the energy functional for each of the 200000 itaration steps and thus we were able to find out on what level the values of the energy functional stabilise. In the Subsection3.3.1 we have seen that the energy of outputs was less than 0.5. For some inputs (e.g. 1, 2, 10 or 14 – see Table A.7 in Appendix A.5) the Metropolis-Hastings algorithm provides outputs that more or less correspond to the input. The main question is whether the algorithm explores in the 200000 iteration steps a major part of the state space or it stays for a long time in one particular region.

Based on our knowledge from Subsection 3.3.1 we have expected that for E_1' and E_2' will see about 50% of rejected outputs when testing the hypothesis H_{APF} (see Table 3.2). But if we look at the Table 4.1, we can see that for both of the functionals we reject almost all outputs. This may indicate slow mixing properties of the algorithm – the algorithm stays for a long time in one particular region and thus the 200000 iteration step may not be enough to explore the major part of the state space. Recall that we sample each 10000-th iteration step. If the algorithm really stays for a long time in one region, it is possible taht we take several samples from this one region. If we then reject on of these samples, we will probably reject all of them. On the other hand we have seen in Section 3.3.1 that neither the inhomogeneous L-function nor the pair correlation function contain enough information to reproduce successfully the geometric structure of clusters. This may cause the big number of rejections as well.

Recall that for these two energy functionals we have chosen K to be 150 for E_1' and 0.9 for E_2'. Values of the energy functional then stabilise on the level 0.1 for E_1' and 15 for E_2'. We can compare these numbers with the resuls of the experiments made in Subsection 3.3.1. We have observed that the values of the energy functional of outputs of the improvement-only algorithm are lower than 0.1 for E_1' and around 5 for E_2'.

When testing the hypothesis H_λ, energy functionals E_1' and E_2' gives better results that E_1'. But the numbers of rejected outputs are bigger than for improvement-only algorithm, see Table 3.3. So it seems that some of the outputs has the correct form of the intensity function, but for almost all outputs the interactions among pairs of points are not well reproduced. Examples of outputs of all of the three energy functionals can be seen in Figure 4.2.

To conclude this Section, let us say that we are aware that the number of executed iteration steps presented here may be insufficient to make relevant conclusions about the behavior of the algorithm. Also the experiments were done in a limited manner. Our priority was to show that even though Tscheschel and Stoyan [2006] suggest to use the Metropolis-Hastings algorithm in the context of stochastic reconstruction and we are able to prove (for a concrete choice of summary characteristics used in the energy functional) the uniform ergodicity, the practical use the algorithm is not straitforward and it is rather complicated. The constant K, that influence the mixing properties of the chain (i.e. how long it would take to move from one part of the state space to another), must be chosen experimentally and also we have to decide how many iteration steps will be executed. Also evaluating the upper bound of the rate of the convergence form the fourth statement of Theorem 6 was not the main aim of this work.

From our experiments it seems that even if for our choices of K the values of the energy functional decreases rapidly at the beginning of the run of the algorithm, then we may stay for a long time in one part of the state stpace. Also, the 200000 iterations may not be sufficient to reach the invariant distribution with some given accuracy. But even if we approximate the invarinat distribution in a satisfactory manner, it only means that the outputs will be more likely configu-

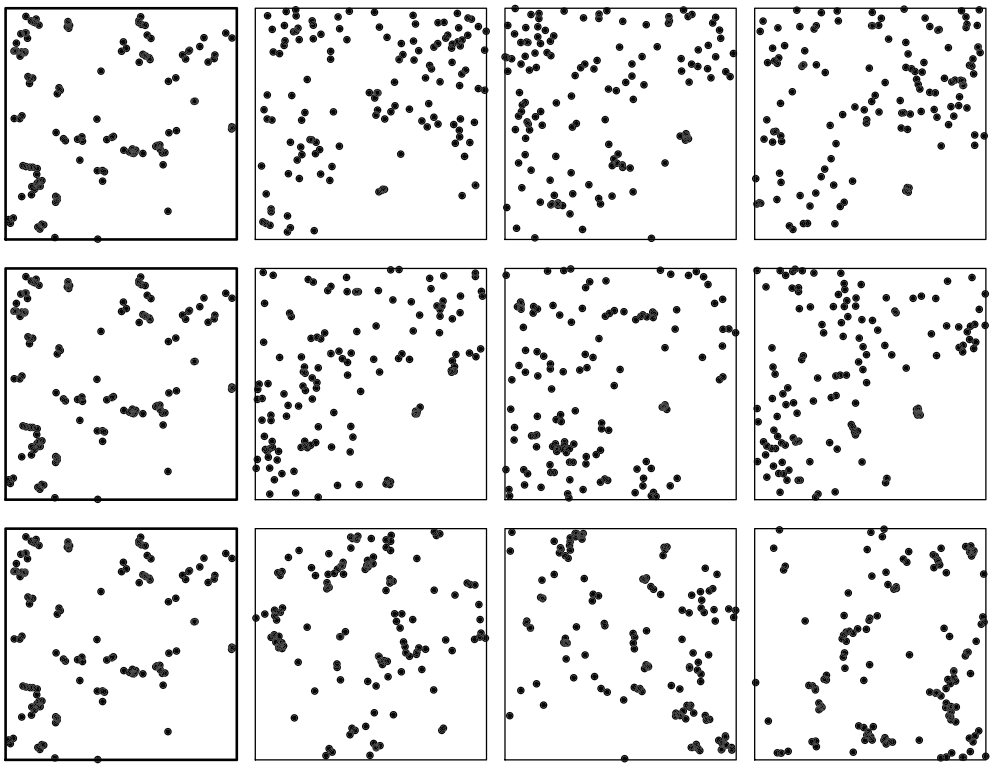

Figure 4.2: Input (on the left, same for all the rows) and three outputs (after 10000, 100000 and 150000 iteration steps) of the stochastic reconstruction algorithm based on the energy functional E_1' and $K = 150$ (first row), E_2' and $K = 0.9$ (second row) and E_5' and $K = 10$ (last row). For E_1' and E_2' the geometric structure of clusters is not well reproduced. For E_5' the size and shape of the clusters of the outputs is more or less correct but there is a problem with the intensity function. Moreover, we are not able to decide whether the samples produced by the Metropolis-Hastings algorithm comes from a different parts of the state space or the mixing properties of the algorithm are slow and we stay in one part of the state space for a long time.

rations with low energy. We do not claim that the outputs will come from local minima of the enrgy functional as in the improvement-only case. On the other hand, we have seen that for the energy functional E_5' we were able to produce outputs that do not contradict the hypotheses H_{APF} and H_λ. Another key point is that it is not obvious (except the visual inspection of the chain) how to decide whether the outputs we have obtained comes from different parts of the state space or the mixing properties are slow and we have explore (during the executed iteration steps) only a small part of the statespace.

Conclusion

It this text we have discussed the possibility of extension of the stochastic reconstruction approach for inhomogeneous point processes.

In the first chapter we have stated basic definitions for spatial point processes, we have discussed the difference between a homogeneous and an inhomogeneous point process and we have mentioned some examples of point process models. We have also listed selected summary characteristics for spatial point processes, e.g. the pair correlation function or the L-function. Most of the definitions have been taken from Møller and Waagepetersen [2004].

In the second chapter we have described the stochastic reconstruction algorithm from Tscheschel and Stoyan [2006]. Then we have stated two obvious modifications for the inhomogenous case, namely the inhomogeneous initial configuration and the non-uniform distribution of the proposed new points. These modifications are also described in the litterature, see Wiegand et al. [2013].

The main contribution of this work is the simulation study we have made – stochastic reconstruction of three different planar point process models, namely the thinned Thomas process, inhomogeneous Poisson process and transformed Matérn hard-core process of type II. We have suggested two statistical tests which may be used to control the quality of outputs of the modified stochastic reconstruction algorithm. These tests were used to compare quality of reconstructions based on 8 different variants of the energy functional. We have discovered that two obvious modifications of the algorithm mentioned above do not assure that the intensity function of output of the stochastic reconstruction algorithm (with a realisation from the corresponding theoretical model as input) does not deviate a lot from the theoretical intensity function on the observation window. Hence the approach from Wiegand et al. [2013] may not be sufficient to produce point patterns that actually correspond to the observed data. Based on this result, we have decided to add a new term to each of the energy functionals. This term contains information about how big or low the values of the estimated intensity function should be and how large are the areas with small/large values, but tells nothing about how to place the areas with high or low intensity function in the observation window. This modification may be useful while reconstructing clusterd data. For regular data or the Poisson point process more sophisticated method for controlling the intensity function is needed. Also for the transformed Marétn hard-core process of type II another choices of the initial configuration should be examined.

Second and third chapter have treated the stochastic reconstruction algorithm as an optimization problem – we have tried to minimize the energy functional E in order to obtain outputs that corresponds to the observed point pattern. The last chapter is devoted to the situation when also the configurations with higher energy are accepted (with some probability). First some basic definitions and re-

sults from the theory of Markov chains on a general state space are given. Then a special case of the Metropolis-Hastings algorithm described in Section 7.1.1 in Møller and Waagepetersen [2004] is adapted to the stochastic reconstruction context. This approach is suggested in Tscheschel and Stoyan [2006] but without further detais. The proof of the uniform ergodicity of the Markov chain produced by this version of Metropolis-Hastings algorithm for a special choice of the summary characteristics used in the energy functional is given in Subsection 4.2.1. We have also given a brief example of use of the Metropolis-Hastings algorithm while reconstructing thinned Thomas process. Since there are no further details about the use of the Metropolis-Hasting algorithm in the stochastic reconstruction context, we have aimed to present the theoretical background of this approach. Concerning the practical application we wanted to show that this approach is less straightforward than the improvement-only algorithm and we wanted to mention the difficulties arising when running the Metropolis-Hastings algorithm.

Bibliography

Baddeley, Adrian and Silverman, Bernard Walter. "A cautionary example on the use of second-order methods for analyzing point patterns." *Biometrics* 40, (1984). 1089–1094.

Baddeley, Adrian, Moller, Jesper and Waagepetersen, Rasmus Plenge. "Non-and semiparametric estimation of interaction in inhomogeneous point patterns." *Statistica Neerlandica* 54, (2000). 329–350.

Baddeley, Adrian, Rubak, Ege and Turner, Rolf. *Spatial Point Patterns: Methodology and Applications with R*. Boca Raton: Chapman & Hall/CRC, 2015. ISBN 9781482210200.

Biscio, Christophe and Møller, Jesper. "The accumulated persistence function, a new useful functional summary statistic for topological data analysis, with a view to brain artery trees and spatial point process applications." *Submitted to Journal of Computational and Graphical Statistics*. Available on arXiv: 1611.00630 [cited 20. 4. 2018]. Last version: June 2017.

Chiu, Sung Nok, Stoyan, Dietrich, Kendall, Wilfrid S. and Mecke, Joseph. *Stochastic Geometry and Its Applications*. Third edition. Chichester: John Wiley & Sons Ltd, 2013. ISBN 978-0-470-66481-0.

Condit, Richard. *Tropical Forest Census Plots*. Berlin: Springer-Verlag; Georgetown: R. G. Landes Company, 1998. ISBN 978-3-540-64144-5.

Daley, Daryl and Vere-Jones, David. *An Introduction to the Theory of Point Processes. Volume II: General Theory and Structure*. Second edition. New York: Springer-Verlag, 2008. ISBN 978-0-387-21337-8.

Getzin, Stephan, Wiegand, Thorsten and Hubbell, Stephen P. "Stochastically driven adult-recruit associations of tree species on Barro Colorado Island." *Proceedings of the Royal Society: Series B* 281, (2014).

Hubbell, Stephen P., Foster, Robin B., O'Brien, Sean T., Harms, Kyle E., Condit, Richard, Wechsler, Beat, Wright, S. Joseph and Loo de Lao, Suzanne. "Light gap disturbances, recruitment limitation, and tree diversity in a neotropical forest." *Science* 283, (1999). 554-557.

Hubbell, Stephen P., Condit, Richard and Foster, Robin B. "Barro Colorado Forest Census Plot Data." (2005). Accessed March 14, 2018. `http://ctfs.si.edu/webatlas/datasets/bci`

Illian, Janine, Penttinen, Antti, Stoyan, Helga and Stoyan, Dietrich. *Statistical Analysis and Modelling of Spatial Point Patterns*. Chichester: John Wiley & Sons Ltd, 2004. ISBN 978-0-470-01491-2.

Kallenberg, Olav. *Foundations of Modern Probability*. Second edition. New York: Springer-Verlag, 2002. ISBN 978-0-387-95313-7.

Lilleleht, Ando, Sims, Allan and Pommerening, Arne. "Spatial forest structure reconstruction as a strategy for mitigating edge-bias in circular monitoring plots." *Forest Ecology and Management* 316, (2014). 47-53.

Loosmore, Bert N. and Ford, David E. "Statistical inference using the g or K point pattern spatial statistics." *Ecology* 87, (2006). 1925-1931.

Meyn, Sean P. and Tweedie, Richard L. *Markov Chains and Stochastic Stability*. New York: Springer-Verlag, 1993. ISBN 978-1-4471-3267-7.

Møller, Jesper and Waagepetersen, Rasmus Plenge. *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton: Chapman & Hall/CRC, 2004. ISBN 1-58488-265-4.

Mundo, Ignacio, Wiegand, Thorsten, Kanagaraj, Rajapandian and Kitzberger, Thomas. "Environmental drivers and spatial dependency in wildfire ignition patterns of northwestern Patagonia." *Journal of Environmental Management* 123, (2013). 77-87.

Myllymäki, Mari, Mrkvička, Tomáš, Grabarnik, Pavel, Seijo, Henri and Hahn, Ute. "Global envelope tests for spatial processes." *Journal of the Royal Statistical Society: Series B* 79, (2017). 381–404.

Neyman, Jerzy and Scott, Elizabeth L. "Statistical approach to problems of cosmology." *Journal of the Royal Statistical Society: Series B* 20, (1958). 1-43.

Pommerening, Arne. "Evaluating structural indices by reversing forest structural analysis." *Forest Ecology and Management* 224, (2006). 266–277.

Pommerening, Arne and Stoyan, Dietrich. "Reconstructing spatial tree point patterns from nearest neighbour summary statistics measured in small subwindows." *Canadian Journal of Forest Research* 38, (2008). 1110-1122.

Ripley, Brian David. "The second-order analysis of stationary point processes." *Journal of Applied Probability* 13, (1976). 255–266.

Roberts, Gareth O. and Rosenthal, Jeffrey S. "General state space Markov chains and MCMC algorithms." *Probability Surveys* 1, (2004). 20-71.

Tierney, Luke. "Markov chains for exploring posterior distribution." *The Annals of Statistics* 22, (1994). 1701-1762.

Torquato, Salvatore. *Random Heterogeneous Materials. Microstructure and Macroscopic Properties*. New York: Springer-Verlag, 2002. ISBN 978-1-4757-6355-3.

Tscheschel, André. *Reconstruction of random porous media*. Diploma dissertation, TU Bergakademie Freiberg, Germany, 2001.

Tscheschel, André and Stoyan, Dietrich. "Statistical reconstruction of random point patterns." *Computational Statistics & Data Analysis* 51, (2006). 859-871.

Tscheschel, André and Chiu, Sung Nok. "Quasi-plus sampling edge correction for spatial point patterns." *Computational Statistics & Data Analysis* 52, (2008). 5287-5295.

van Lieshout, Marie-Colette and Baddeley, Adrian. "A nonparametric measure of spatial interaction in point patterns." *Statistica Neerlandica* 50, (1996). 344–361.

van Lieshout, Marie-Colette. "A J-function for inhomogeneous point processes." *Statistica Neerlandica* 65, (2010). 183–201.

Wiegand, Thorsten, He, Fangliang and Hubbell, Stephen P. "A systematic comparison of summary characteristics for quantifying point patterns in ecology." *Ecography* 36, (2013). 92–103.

Wong, Ka Yiu and Chiu, Sung Nok. "Isotropy test for spatial point processes using stochastic reconstruction." *Spatial Statistics* 15, (2016). 56-69.

List of Figures

List of Tables

A. Attachments

A.1 Accumulated persistence function

Let us now give a more detailed description of the accumulated persistence function mentioned in Section 3.2. Recall tha φ is a point pattern such that $N(\varphi \cap W) = n \in \mathbb{N}$. Thus we can write $\varphi = \{x_1, x_2, \ldots, x_n\}$. By $C_r = \bigcup_{i=1}^{n} C_r^i$ we have denoted the union of discs with radius r centered at points of φ. We are interested in how the s-dimensional topological features of C_r vary for different values of parameter r and for $s = 1, 2, \ldots, d-1$. In case of planar point patterns it is only relevant to take $s = 0$ and $s = 1$.

For the accumulated persistence function, birth and death times of k-dimensional topological features of C_r, $r \in [0, r_{max}]$ are of particular interest. Let us now reduce the problem only on $k = 0$, i.e. on the connected components. If $r = 0$, C_0 is in fact just the union of points $x_1, x_2, \ldots, x_n$. So for $r = 0$ we have n connected components. As the value of r grows, some of the discs starts overlaping each other and hence some of the components will disappear. Suppose we have two connected components A and B that are disjoint up to some time r_0. At the time r_0 their intersection is no longer empty, so we choose one of these components, for example A, and we set its death time to be r_0. The other component is from now defined as the union of A and B. It can be seen that under this conditions all of the birth times are equal to zero. We suppose that coordinates of all points are unique, so the death times of different components are unique too. For sufficiently large r there will be just one connected component and its death time is set to be infinity. The situation is more or less the same for $k = 1$, but as the value of r grows, some of the loops may disappear and new loops may appear. Thus for each loop we have a birth time that is not necessairly zero.

Denote the birth time of i-th connected component by b_i and its death time by d_i. Its lifetime is then given by $l_i = d_i - b_i$. We can also compute a meanage $m_i = \frac{b_i + d_i}{2}$. The accumulated persistence function for connected components is then defined as

$$APF_0(\varphi, m) = \sum_{j=1}^{M} l_j \mathbf{1}\{m_j \leq m\}, \ m \geq 0,$$

where M is the overall number of connected components (for large values of r there is just one component with an infinite lifetime which represents the whole space, but we ignore this infinite lifetime since it does not give us any important information).

A.2 Reconstructing thinned Thomas process

Input	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Mean	SD	Percentage
P-value	0.679	0.600	0.050	0.322	0.222	0.125	0.740	0.691	0.050	0.654	0.806	0.516	0.297	0.960	0.091			
E_1	1	4	3	5	14	9	12	15	19	8	12	16	15	15	20	11. 2	5.92	56%
E_1'	2	5	6	10	14	6	14	15	19	6	8	18	13	9	20	11.0	5.59	55%
E_2	5	5	10	9	18	15	14	16	19	8	15	20	18	11	20	13.5	5.21	68%
E_2'	0	3	4	6	8	7	7	12	13	8	6	10	10	5	20	7.93	4.76	40%
E_3	2	3	11	0	3	6	1	1	16	8	4	4	2	0	14	5.00	5.06	25%
E_3'	12	10	15	8	16	9	8	9	20	14	13	11	16	5	20	12.4	4.44	62%
E_4	12	18	20	14	17	17	9	14	20	16	15	17	16	6	20	15.4	4.00	77%
E_4'	20	18	20	18	20	18	19	20	20	20	20	20	20	16	20	19.3	1.22	96%
E_5	0	0	20	0	3	4	1	9	17	0	0	5	2	0	13	4.93	6.70	25%
E_5'	0	0	11	0	1	0	0	0	3	0	1	2	0	0	4	1.47	2.92	7%
E_6	2	2	7	1	1	0	0	0	12	3	0	6	2	0	12	3.20	4.16	16%
E_6'	2	0	10	1	3	2	1	1	9	3	1	14	1	3	17	4.53	5.33	23%
E_7	0	0	7	0	0	0	0	0	5	2	0	0	1	0	4	1.27	2.25	6%
E_7'	0	0	8	0	0	0	0	0	4	2	0	3	1	0	3	1.40	2.29	7%
E_8	1	4	5	0	0	1	1	0	8	2	0	5	1	0	11	2.60	3.36	13%
E_8'	4	0	7	0	2	2	1	1	9	8	1	11	0	2	20	4.53	5.59	23%

Table A.1: Testing the hypothesis H_{APF}. Five different entries are given: upper bound of the p-interval for the inputs, number of rejected outputs for each input, mean number of rejected outputs (per one input), standard deviation of the number of rejected outputs (per one input) and percentage of the overall rejected reconstructions. The global rank envelope test based on 2499 simulations of thinned Thomas process was used.

Input	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Mean	SD	Percentage
P-value	0.983	0.860	0.146	0.838	0.058	0.085	0.655	0.279	0.090	0.729	0.224	0.792	0.424	0.697	0.027			
E_1	0	0	3	1	15	8	3	2	10	0	0	0	2	2	19	4.33	5.97	22%
E'_1	0	0	0	0	3	0	0	0	0	0	0	0	0	0	19	1.47	4.91	7%
E_2	1	0	3	5	20	12	13	4	18	0	3	1	10	6	20	7.73	7.26	39%
E'_2	0	0	1	0	5	0	0	0	2	0	0	0	0	0	20	1.87	5.19	9%
E_3	0	0	1	0	17	19	0	0	14	0	6	0	3	0	20	5.33	7.86	27%
E'_3	0	0	0	0	10	2	0	0	9	0	0	0	0	0	20	2.73	5.81	14%
E_4	0	0	0	0	15	17	0	0	8	0	0	0	0	1	20	4.07	7.23	20%
E'_4	0	0	0	0	17	8	0	0	7	0	0	0	0	0	20	3.47	6.66	17%
E_5	2	3	17	2	20	20	15	19	20	5	19	7	18	8	20	13.0	7.46	65%
E'_5	0	0	0	0	16	4	0	0	5	0	0	0	0	0	20	3.00	6.33	15%
E_6	0	0	0	0	20	20	0	3	3	0	11	0	3	0	20	5.33	8.11	27%
E'_6	0	0	0	0	11	4	0	0	0	0	0	0	0	0	20	2.33	5.70	12%
E_7	0	0	0	0	19	20	0	1	20	0	12	1	5	0	20	6.53	8.83	33%
E'_7	0	0	0	0	11	3	0	0	9	0	0	0	0	0	20	2.87	5.90	14%
E_8	0	0	0	0	16	16	0	0	2	0	0	0	0	0	20	3.60	7.18	18%
E'_8	0	0	0	0	11	7	0	0	1	0	0	0	0	0	20	2.60	5.79	13%

Table A.2: Testing the hypothesis H_λ. Five different entries are given: p-value of the inputs, number of rejected outputs for each input, mean number of rejected outputs (per one input), standard deviation of the number of rejected outputs (per one input) and percentage of the overall rejected reconstructions. The deviation test based on 2499 simulations of thinned Thomas process was used.

A.3 Reconstructing inhomogeneous Poisson process

Input	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Mean	SD	Percentage
P-value	0.704	0.977	0.266	0.527	0.459	0.156	0.794	0.910	0.627	0.654	0.089	0.066	0.924	0.769	0.865			
E_1	0	0	2	0	0	10	0	0	1	2	0	14	2	0	1	2.13	4.16	11%
E_1'	0	0	0	0	0	3	0	0	0	0	0	10	0	0	0	0.87	2.64	4%
E_2	0	0	2	1	0	3	0	1	0	1	1	15	0	1	0	1.67	3.79	5%
E_2'	0	0	1	0	0	8	0	1	1	0	1	14	1	0	0	1.80	3.93	9%
E_3	0	0	0	0	0	5	0	0	2	0	0	8	0	0	1	1.07	2.34	5%
E_3'	0	0	2	0	0	10	0	0	0	0	1	12	0	0	0	1.67	3.85	8%
E_4	0	0	0	0	0	6	1	0	0	2	1	6	0	0	1	1.13	2.07	6%
E_4'	0	1	0	0	0	5	0	0	0	0	0	8	0	0	0	0.93	2.34	5%
E_5	0	0	1	0	0	8	1	0	0	1	2	5	0	0	1	1.27	2.28	6%
E_5'	0	0	0	0	0	1	0	0	0	0	0	6	0	0	1	0.53	1.55	3%
E_6	1	0	1	0	0	6	0	0	0	1	1	5	0	0	0	1.00	1.89	5%
E_6'	0	0	1	1	0	7	0	0	0	0	1	5	0	0	0	1.00	2.10	5%
E_7	0	0	0	0	0	8	0	0	0	0	0	4	0	0	0	0.80	2.24	4%
E_7'	0	0	0	0	0	6	0	0	0	1	0	5	1	0	0	0.87	1.92	4%
E_8	0	0	0	0	0	4	0	0	1	0	0	9	0	0	0	0.93	2.46	5%
E_8'	0	0	0	0	0	6	0	0	0	0	0	3	0	0	0	0.60	1.68	3%

Table A.3: Testing the hypothesis H_{APF}. Five different entries are given: upper bound of the p-interval for the inputs, number of rejected outputs for each input, mean number of rejected outputs (per one input), standard deviation of the number of rejected outputs (per one input) and percentage of the overall rejected reconstructions. The global rank envelope test based on 2499 simulations of thinned Thomas process was used.

Input	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Mean	SD	Percentage
P-value	0.094	0.263	0.173	0.488	0.465	0.004	0.733	0.445	0.072	0.536	0.664	0.066	0.572	0.823	0.983			
E_1	8	7	4	2	8	20	3	2	9	7	3	18	9	0	1	6.73	5.82	34%
E_1'	16	3	6	3	6	20	1	2	10	2	1	16	1	2	1	6	6.43	30%
E_2	10	9	8	4	7	20	0	6	15	7	7	15	5	1	2	7.73	5.52	39%
E_2'	14	9	6	2	4	20	1	2	13	5	2	16	6	3	2	7	6.01	35%
E_3	6	6	14	0	4	19	0	2	17	6	5	12	4	1	1	6.47	6.17	32%
E_3'	14	5	8	0	9	20	0	4	16	3	1	15	2	0	0	6.47	6.81	32%
E_4	7	5	9	2	5	20	0	4	16	2	1	10	1	2	1	5.67	5.89	28%
E_4'	7	3	9	0	6	20	0	3	14	1	1	10	1	0	0	5	6.02	25%
E_5	13	3	12	3	6	20	5	3	14	2	7	17	2	2	0	7.27	6.30	36%
E_5'	14	3	10	1	10	20	0	3	16	3	0	19	8	0	1	7.20	7.21	36%
E_6	10	7	11	5	8	19	0	4	14	2	2	14	3	2	0	6.73	5.79	34%
E_6'	12	4	13	1	5	20	0	1	10	0	2	16	3	1	1	5.93	6.54	30%
E_7	14	4	10	2	3	19	0	4	14	4	4	15	7	1	2	6.87	5.97	34%
E_7'	14	6	11	0	5	20	0	2	14	1	1	16	5	1	0	6.40	6.80	32%
E_8	13	4	12	2	2	20	0	4	9	0	5	12	2	0	0	5.67	6.13	32%
E_8'	13	5	11	0	7	20	0	1	13	0	1	15	1	0	0	5.80	6.85	29%

Table A.4: Testing the hypothesis H_λ. Five different entries are given: p-value of the inputs, number of rejected outputs for each input, mean number of rejected outputs (per one input), standard deviation of the number of rejected outputs (per one input) and percentage of the overall rejected reconstructions. The deviation test based on 2499 simulations of thinned Thomas process was used.

A.4 Reconstructing transformed Matérn hard-core process of type II

Input	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Mean	SD	Percentage
P-value	0.845	0.852	0.175	0.304	0.943	0.894	0.551	0.497	0.210	0.452	0.419	0.372	0.352	0.778	0.113			
E_1	0	1	11	2	1	1	3	2	3	0	2	0	0	0	4	2.00	2.80	10%
E_1'	14	17	19	18	12	13	16	15	18	16	18	13	18	16	14	15.8	2.21	79%
E_2	7	5	19	10	6	7	9	8	10	8	10	8	11	7	11	9.07	3.28	45%
E_2'	20	19	20	20	20	20	20	19	20	20	20	20	20	20	20	19.86	0.35	99%
E_3	20	20	20	20	20	20	20	20	19	20	20	20	20	20	20	19.93	0.26	100%
E_3'	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20.0	0.00	100%
E_4	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20.0	0.00	100%
E_4'	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20.0	0.00	100%
E_5	15	12	20	18	10	10	14	12	15	12	18	12	17	12	17	14.26	3.15	71%
E_5'	16	19	20	17	19	17	18	17	15	19	17	17	18	15	16	17.33	1.50	87%
E_6	20	20	20	20	19	16	19	19	20	19	20	19	19	18	17	19.0	1.20	95%
E_6'	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20.0	0.00	100%
E_7	15	14	17	15	16	9	14	15	18	15	18	12	15	10	14	14.47	2.56	72%
E_7'	17	19	20	20	18	16	19	20	18	20	20	19	19	19	20	18.93	1.22	95%
E_8	20	19	20	20	19	17	20	20	19	20	20	20	20	17	20	19.4	1.06	97%
E_8'	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20.0	0.00	100%

Table A.5: Testing the hypothesis H_{APF}. Five different entries are given: upper bound of the p-interval for the inputs, number of rejected outputs for each input, mean number of rejected outputs (per one input), standard deviation of the number of rejected outputs (per one input) and percentage of the overall rejected reconstructions. The global rank envelope test based on 2499 simulations of thinned Thomas process was used.

Input	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Mean	SD	Percentage
P-value	0.558	0.043	0.994	0.289	0.829	0.423	0.125	0.298	0.650	0.496	0.935	0.469	0.666	0.328	0.557			
E_1	10	19	1	9	6	15	17	12	4	10	6	11	4	15	10	9.93	5.15	50%
E'_1	9	16	2	7	3	10	20	8	3	4	3	5	3	4	10	7.13	5.24	36%
E_2	13	19	3	11	7	15	17	14	7	12	7	11	7	15	12	11.33	4.42	57%
E'_2	9	15	1	5	3	12	20	9	5	5	0	9	7	11	12	8.2	5.36	41%
E_3	14	20	2	11	7	17	18	16	4	6	6	11	3	13	15	10.87	5.85	54%
E'_3	1	14	1	5	4	7	16	3	2	2	1	2	4	8	8	5.20	4.66	26%
E_4	2	17	1	3	2	7	17	6	3	1	2	8	1	6	6	5.47	5.24	27%
E'_4	4	10	0	5	1	6	16	4	4	1	0	9	1	5	4	4.67	4.34	23%
E_5	10	19	1	15	11	19	19	18	8	6	11	14	7	18	16	12.8	5.62	64%
E'_5	6	18	0	4	3	6	19	7	3	3	1	6	2	2	4	5.6	5.60	28%
E_6	11	20	7	7	5	12	19	16	6	2	5	11	3	8	12	9.6	5.51	48%
E'_6	3	16	0	4	2	4	16	7	2	2	2	2	2	4	5	4.73	4.86	24%
E_7	12	19	2	10	11	16	17	17	11	5	4	14	4	16	15	11.53	5.50	58%
E'_7	4	12	2	7	0	6	18	9	2	7	0	4	2	6	9	5.87	4.84	29%
E_8	3	17	0	7	4	6	15	6	2	1	2	5	2	8	11	5.93	5.04	30%
E'_8	3	13	2	5	1	8	15	7	3	1	1	9	1	4	4	5.13	4.44	26%

Table A.6: Testing the hypothesis H_λ. Five different entries are given: p-value of the inputs, number of rejected outputs for each input, mean number of rejected outputs (per one input), standard deviation of the number of rejected outputs (per one input) and percentage of the overall rejected reconstructions. The deviation test based on 2499 simulations of thinned Thomas process was used.

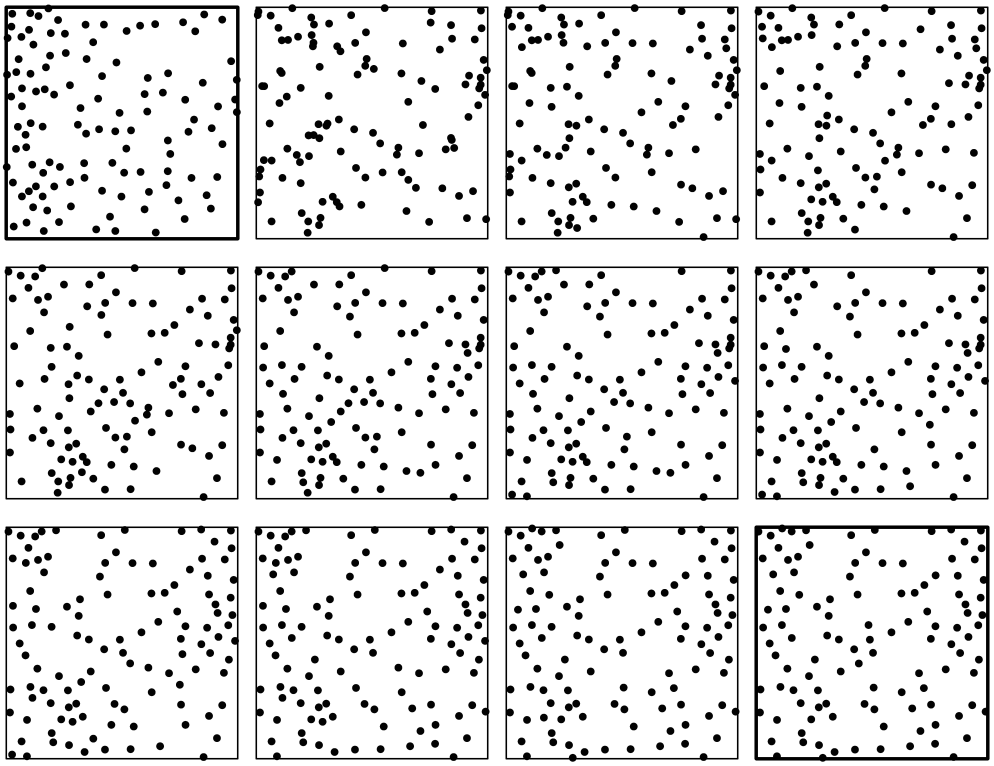

Figure A.1: Reconstructing transformed Matérn hard-core process of type II. Observed data to be reconstructed (top left corner), initial configuration, intermediate states of the stochastic reconstruction algorithm after 50, 100, 150, 200, 250, 300, 400, 500, and 600 iteration steps (from left to right and from top to bottom), output of the stochastic reconstruction algorithm (bottom right corner). Evolution of values of the energy functional (we have used the energy functional E_1) can be seen in Figure A.2.

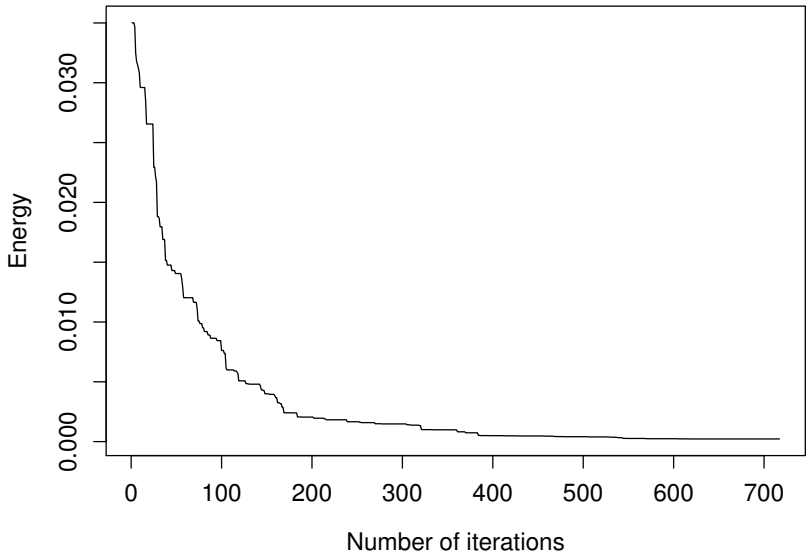

Figure A.2: Values of the energy functional E_1 during the run of the stochastic reconstruction algorithm from Figure A.1.

A.5 Reconstructing thinned Thomas process using Metropolis-Hastings algorithm

Input	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Mean	SD	Percentage
P-value	0.679	0.600	0.050	0.322	0.222	0.125	0.740	0.691	0.050	0.654	0.806	0.516	0.297	0.960	0.091			
E_1'	18	18	20	20	20	20	20	20	20	18	20	20	20	20	20	19.6	0.83	98%
E_2'	9	13	20	20	20	20	20	20	20	16	20	20	20	20	20	18.5	3.31	93%
E_5'	1	0	3	0	5	2	4	7	19	3	1	14	7	0	20	5.73	6.69	29%

Table A.7: Testing the hypothesis H_{APF}. Five different entries are given: upper bound of the p-interval for the inputs, number of rejected outputs for each input, mean number of rejected outputs (per one input), standard deviation of the number of rejected outputs (per one input) and percentage of the overall rejected reconstructions. The global rank envelope test based on 2499 simulations of thinned Thomas process was used.

Input	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Mean	SD	Percentage
P-value	0.983	0.860	0.146	0.838	0.058	0.085	0.655	0.279	0.090	0.729	0.224	0.792	0.424	0.697	0.027			
E_1'	0	0	12	0	6	2	8	0	8	0	1	13	5	0	20	5.00	6.16	25%
E_2'	0	0	15	0	10	10	9	17	4	4	0	4	1	6	20	6.67	6.62	33%
E_5'	0	5	4	16	8	14	13	1	19	2	8	5	15	2	20	8.80	6.81	44%

Table A.8: Testing the hypothesis H_λ. Five different entries are given: p-value for the inputs, number of rejected outputs for each input, mean number of rejected outputs (per one input), standard deviation of the number of rejected outputs (per one input) and percentage of the overall rejected reconstructions. The deviation test based on 2499 simulations of thinned Thomas process was used.

A.6 Electronic attachments

The electronic version of this text contains also the electronic attachement – the `R` scripts illustrating the implementation of the stochastic reconstruction algorithm for inhomogeneous point processes.