



Centrum voor Wiskunde en Informatica

REPORTRAPPORT

PNA

Probability, Networks and Algorithms



Probability, Networks and Algorithms

On a model of sequential point patterns

V. Shcherbakov

REPORT PNA-R0505 NOVEMBER 2005

CWI is the National Research Institute for Mathematics and Computer Science. It is sponsored by the Netherlands Organization for Scientific Research (NWO).

CWI is a founding member of ERCIM, the European Research Consortium for Informatics and Mathematics.

CWI's research has a theme-oriented structure and is grouped into four clusters. Listed below are the names of the clusters and in parentheses their acronyms.

Probability, Networks and Algorithms (PNA)

Software Engineering (SEN)

Modelling, Analysis and Simulation (MAS)

Information Systems (INS)

Copyright © 2005, Stichting Centrum voor Wiskunde en Informatica

P.O. Box 94079, 1090 GB Amsterdam (NL)

Kruislaan 413, 1098 SJ Amsterdam (NL)

Telephone +31 20 592 9333

Telefax +31 20 592 4199

ISSN 1386-3711

On a model of sequential point patterns

ABSTRACT

A sequential point process motivated by the cooperative sequential adsorption is studied. Analytical properties of the process are considered in details. It is shown that the point process is useful for modelling cluster point patterns. The test on the real life data is carried out.

2000 Mathematics Subject Classification: 60G55, 60D05, 62M30

Keywords and Phrases: Markov point process; Hammersley-Clifford factorization; Packale triangle; local stability; cooperative sequential adsorption; desorption, consistency; maximum likelihood estimation; MCMC methods

Note: This research is supported by the Technology Foundation STW, applied science division of NWO, and the technology programme of the Ministry of Economic Affairs (project CWI.6155 'Markov sequential point processes for image analysis and statistical physics')

On a model of sequential point patterns.

V. Shcherbakov,
CWI, Postbus 94079, 1090 GB,
Amsterdam, The Netherlands

Abstract

A sequential point process motivated by the cooperative sequential adsorption is studied. Analytical properties of the process are considered in details. It is shown that the point process is useful for modelling cluster point patterns. The test on the real life data is carried out.

2000 Mathematics Subject Classification: 60G55, 60D05, 62M30.

Keywords and Phrases: Markov point process, Hammersley-Clifford factorization, Packale triangle, local stability, cooperative sequential adsorption, desorption, consistency, maximum likelihood estimation, MCMC methods.

Note: This research is supported by the Technology Foundation STW, applied science division of NWO, and the technology program of the Ministry of Economic Affairs (project CWI.6155 'Markov sequential point processes for image analysis and statistical physics')

1 Introduction

In this paper we study a model of point processes that was motivated by the models of cooperative sequential adsorption (CSA). Those models are widely used in physics and chemistry for modelling of adsorption process like chemisorption on single-crystal surfaces, adsorption in colloidal systems and other similar processes. For the physics-chemistry background and for surveys of the relevant literature we refer to [4], [17], [18] and references therein. To imagine the phenomena one can think of a surface of some material that attracts particles from the space around. The main peculiarity of all of these models is that adsorbed particles change adsorption rates of the surface. The dependence on previously adsorbed particles can be modelled in various ways. We choose the variant where the adsorption rates at a point depend on a number of previously adsorbed molecules in some vicinity of the point. Therefore the point process we propose in this paper is directly related to the lattice model of monomer filling with cooperative effects ([4]).

We treat a point pattern as an ordered set of points, i.e. sequence or vector of points. There are several ways to incorporate sequentiality into a point pattern. For instance, one can consider marked point processes, where a mark means "time". Another possibility is naturally formulated in terms of sequential Markov point processes [11]. Briefly, it means that a non-symmetric neighbourhood relation between points is defined and the conditional intensity of the process depends on directed neighbours of a point. The process state is described by ordered sets of points (sequences) in some bounded subset $D \subset \mathbf{R}^d$. The process distribution is specified by a density with respect to a reference measure. The reference measure corresponds to random sequences formed by a Poisson's number of points, and each point is uniformly distributed in D . We refer to [11] for details and examples (see also [12]) and follow this approach here.

The proposed point process is named as CSA point process (CSA process). We start with studying the analytical properties of the process. Analytical properties include conditions for the model existence, Markov and local stability properties. All these issues are treated exhaustively. It is well known that there are two fundamental characteristics for any Markov point process, namely conditional intensity and clique interaction functions. In the general case the process density might not exist. We prove sufficient conditions for existence of a well-defined (integrable) density and show that in most interesting (non-degenerated) cases these conditions are also necessary. Critical growth of intensities is found. We compute explicitly a clique interaction function of any order. Computation of the clique interaction functions involves such fundamental objects of combinatorial analysis as the famous Stirling numbers and results in explicit formulas in terms of the Pascal triangle.

For simulation of the process the Metropolis-Hastings and spatial birth-death algorithms can be used. For the sequential Markov point processes these algorithms were introduced in [11]. To ensure the convergence of the algorithms a point process should be locally stable. In general the CSA process is not locally stable and we construct an example of that. We prove that if there are only a finite number of different intensities, then the process is locally stable. Moreover, we give a formula for the local stability bound and construct examples showing that the obtained local stability bound cannot be improved in general.

We show that the studied sequential Markov point process provides a simple and flexible choice for modelling both repulsive and attractive effects in point patterns. We give examples of computer simulations confirming theoretical predictions. The mechanism of clusters formation is explained. The way the process is constructed provides an easy and transparent way to control clusters formation. Varying the parameters, i.e. both the interaction radius and intensities, one can intensify or reduce cluster effects. There are though limitations in modelling point clusters by CSA processes. These limitations are explained analytically and illustrated by computer simulations.

In Section 5 we try to fit the CSA model to the real life data. Our analysis is based on the method of Monte-Carlo maximum likelihood estimation (proposed in general setting in [7]). Finally we briefly mention possible generalizations of CSA processes.

2 CSA Markov point processes

Let D be a compact subset of \mathbf{R}^d with positive Lebesgue measure. Fix a number $R \geq 0$ and specify a relation \sim between points as follows

$$x \sim y \Leftrightarrow \|x - y\| \leq R. \quad (1)$$

R is an interaction radius, a point $y \in D$ is called a neighbour of a point x if (1) holds. Let $\{\beta_n, n \geq 0\}$ be a set of non-negative numbers. Denote $F_D = \cup_{n=0}^{\infty} D^n$ the set of finite sequences, $D^0 = \emptyset$, and \mathcal{F}_D the standard σ -algebra on F_D . Let $n(x, \vec{x})$ be the number of neighbours of a point x in the sequence $\vec{x} = (x_1, \dots, x_k)$. Given the sequence of points $\vec{x} = (x_1, \dots, x_k)$, $k \geq 1$ we will denote $\vec{x}_{<k} = (x_1, \dots, x_{k-1})$, for every $k \geq 2$ and $\vec{x}_{<1} = \emptyset$. Define a reference measure ν on (F_D, \mathcal{F}_D) corresponding to random sequences of Poisson length with mean 1 with independent uniformly distributed components in D . Thus,

$$\nu(F) = \sum_{n=0}^{\infty} \frac{e^{-|D|}}{n!} \int_{D^n} 1_{\{(x_1, \dots, x_n) \in F\}} dx_1 \dots dx_n, \quad F \in \mathcal{F}_D. \quad (2)$$

Let us consider a probability distribution on (F_D, \mathcal{F}_D) defined by the following density with respect to the reference measure ν

$$f(x_1, \dots, x_n) = Z^{-1} \prod_{k=1}^n \beta_{n(x_k, \vec{x}_{<k})}, \quad (3)$$

where Z is a normalising constant (partition function)

$$Z = \sum_{n=0}^{\infty} \frac{e^{-|D|}}{n!} \int_{D^n} \prod_{k=1}^n \beta_{n(x_k, \vec{x}_{<k})} dx_1 \dots dx_n.$$

Note that $f(\vec{y}) > 0$ implies $f(\vec{z}) > 0$ for all subsequences $\vec{z} \subset \vec{y}$. For all sequences \vec{y} such that $f(\vec{y}) > 0$, the ratio

$$\frac{f((\vec{y}, u))}{f(\vec{y})} = \lambda(u|\vec{y}) = \beta_{n(u, \vec{y})} \quad (4)$$

depends only on its neighbours in the sequence \vec{y} . Therefore the sequential spatial process X on the set D defined by its density (3) with respect to the reference measure $\nu(\cdot)$ is a sequential Markov process with respect to the relation \sim (see [11] for more details). For convenience we name this sequential Markov point process as CSA point process.

Given $N(D) = n$ the conditional density of the process is the following function

$$p_n(x_1, \dots, x_n) = \frac{\prod_{k=1}^n \beta_{n(x_k, \vec{x}_{<k})}}{\int_{D^n} \prod_{k=1}^n \beta_{n(y_k, \vec{y}_{<k})} dy_k} \quad (5)$$

Remarks. 1) The model of point processes we have just defined is motivated by CSA lattice models of statistical physics ([4]). In these models single lattice sites are filled by particles at rates which depend on the state of the neighbouring sites. Rates might be either enhanced by reacted neighbours or inhibited. The former leads to clustering and the latter leads to repulsion between particles. These properties are inherited by the CSA point process (see Section 4). We choose a case when the rates depend on a number of neighbours, therefore the conditional intensity (4) of the process depends on the number of neighbours of point u . This particular choice of dependence already leads to a very wide class of sequential Markov point processes. Certainly, other forms of dependence can be taken.

It must be noted that the lattice models like cooperative sequential adsorption (CSA) are used to describe *irreversible* dynamic processes. The point configurations obtained by CSA point processes are results of *reversible* dynamics. The probability distribution of the CSA process is an equilibrium distribution for the Markov chains described by Metropolis-Hastings and spatial birth-and-death samplers which can be chosen to be reversible (see [12] for the details). Therefore, it might be better to consider CSA point processes as continuous space analogy of adsorption-desorption models that allow for the possibility of particles desorption ([4]). Epidemic processes when there is recovery should be mentioned as an appropriate biological example.

2) A wide class of CSA point processes that might be interesting in applications is formed by CSA processes which densities are specified by a finite number of different intensities, i.e. β_k , $k \leq N$ and $\beta_k = \gamma \geq 0$, $k > N$ for some $N \geq 0$. Let us introduce some notation. Denote by $t(\vec{x})$ the total number of points in the configuration \vec{x} ,

$$t_k(\vec{x}) = \sum_{x_i \in \vec{x}} 1_{\{n(x_i, \vec{x}_{<i})=k\}}, \quad k = 0, \dots, N. \quad (6)$$

and

$$s(\vec{\mathbf{x}}) = t(\vec{\mathbf{x}}) - \sum_{k=0}^N t_k(\vec{\mathbf{x}}). \quad (7)$$

In words, $t_k(\vec{\mathbf{x}})$ is the number of sequence coordinates that have k , $0 \leq k \leq N$, neighbours in the sequence $\vec{\mathbf{x}}_{<i}$, $s(\vec{\mathbf{x}})$ is the total number of points that have at least $N + 1$ neighbours. In terms of these notation the unnormalized process density is

$$h(x_1, \dots, x_n) \sim \gamma^{s(\vec{\mathbf{x}})} \prod_{k=0}^N \beta_k^{t_k(\vec{\mathbf{x}})}. \quad (8)$$

3) An important subclass of the CSA point processes with a finite number of different intensities is formed by CSA processes with *hard core interaction*. The last means that we consider the case when there is only a finite number of non-zero intensities β_0, \dots, β_N . In this case any point of a sequence $\vec{\mathbf{x}}$ cannot have more than N neighbours. It is obvious that in this case $t(\vec{\mathbf{x}}) = t_0(\vec{\mathbf{x}}) + \dots + t_N(\vec{\mathbf{x}})$, where $t(\vec{\mathbf{x}})$ is the total number of points in the sequence $\vec{\mathbf{x}}$. The unnormalized process density can be written as follows

$$h(\vec{\mathbf{x}}) = h(x_1, \dots, x_n) \sim \prod_{k=1}^n \beta_{n(x_k, \vec{\mathbf{x}}_{<k})} 1_{\{\widehat{N}(\vec{\mathbf{x}}) \leq N\}} = \prod_{k=0}^N \beta_k^{t_k(\vec{\mathbf{x}})} 1_{\{\widehat{N}(\vec{\mathbf{x}}) \leq N\}}, \quad (9)$$

where the statistic

$$\widehat{N}(\vec{\mathbf{x}}) = \max_{x_i \in \vec{\mathbf{x}}} n(x_i, \vec{\mathbf{x}}_{<i}) = \max\{k : t_k(\vec{\mathbf{x}}) > 0\} \quad (10)$$

is a maximal number of neighbours observed in $\vec{\mathbf{x}}$.

4) If $R \geq \text{diam}(D)$, then it means that given a current state $\vec{\mathbf{x}}$ the conditional intensity at a point $x \in D$ depends only on the total number of points in the sequence $\vec{\mathbf{x}}$, regardless of their positions. This kind of interaction between points can be named *mean-field interaction*, by analogy with statistical physics. In this setting any point $y \in D$ is a neighbour of point x . If $\beta_k = 0$, $k \geq n$, for some n , then we can observe at most n points of the process. Conditional on the event $N(D) = n$ the probability distribution of n points is the uniform one

$$p_n(x_1, \dots, x_n) = \frac{1}{|D|^n}.$$

5) If $\beta_n \equiv \beta$ for some $\beta > 0$, then the value of the interaction radius is irrelevant and the process distribution is a probability measure on (F_D, \mathcal{F}_D) corresponding to random sequences of Poisson length with mean $\beta|D|$ and independent uniformly distributed components in D .

2.1 Integrability conditions.

We need to check that the partition function is finite to get a well-defined density (3). If, for example, the set of intensities is uniformly bounded

$$\sup_{n \geq 0} \beta_n < \infty,$$

then the partition function is finite. But if the numbers β_n , $n \geq 0$, increase sufficiently fast as $n \rightarrow \infty$, then the partition function might be infinite. Proposition 2.1 shows that if the domain D is not too degenerate then the linear growth of intensities is the critical one separating these two cases.

Proposition 2.1 1) Let the set of intensities form an increasing sequence such that $\beta_n \rightarrow \infty$ as $n \rightarrow \infty$ and assume that the rate of its growth is sublinear: $\beta_n/n \rightarrow 0$ as $n \rightarrow \infty$, then the partition function is finite.

2) Assume that $\beta_n = \beta(n+1)$, $n \geq 0$, for some $\beta > 0$ and the domain D contains a ball $B(y, r)$ of positive radius $r < R/2$. Then the partition function is infinite, if β is sufficiently large.

Proof of the part 1) of Proposition 2.1. Indeed, fix $0 < \varepsilon < |D|^{-1}$ and let $n(\varepsilon)$ be such that $\beta_n/n < \varepsilon$ for any $n \geq n(\varepsilon)$. Represent the partition function as follows

$$Z = \sum_{n < n(\varepsilon)} \frac{e^{-|D|}}{n!} \int_{\dot{D}^n} \prod_{k=1}^n \beta_{n(x_k, \vec{x}_{<k})} dx_1 \dots dx_n + \sum_{n \geq n(\varepsilon)} \frac{e^{-|D|}}{n!} \int_{\dot{D}^n} \prod_{k=1}^n \beta_{n(x_k, \vec{x}_{<k})} dx_1 \dots dx_n.$$

The first term in the preceding display is finite. We are going to bound the second term. Note that for any $i \geq 1$ we have the inequality $n(x_i, \vec{x}_{<i}) \leq i - 1$, therefore $\beta_{n(x_i, \vec{x}_{<i})} \leq \beta_{i-1}$. Also note that by assumption $\beta_k \leq \beta_{n(\varepsilon)}$, if $k \leq n(\varepsilon)$. Thus the product in the integral can be bounded as follows

$$\prod_{k=1}^n \beta_{n(x_k, \vec{x}_{<k})} \leq \frac{\beta_{n(\varepsilon)}^{n(\varepsilon)} \varepsilon^{n-n(\varepsilon)} n!}{n(\varepsilon)!}$$

Using this bound we obtain

$$\sum_{n \geq n(\varepsilon)} \frac{e^{-|D|}}{n!} \int_{\dot{D}^n} \prod_{k=1}^n \beta_{n(x_k, \vec{x}_{<k})} dx_1 \dots dx_n \leq \left(\frac{\beta_{n(\varepsilon)}}{\varepsilon} \right)^{n(\varepsilon)} \frac{e^{-|D|}}{n(\varepsilon)!} \sum_{n \geq n(\varepsilon)} (\varepsilon |D|)^n,$$

and the right hand side is finite by choice of ε . Therefore $Z < \infty$.

Proof of the part 2) of Proposition 2.1. In this case one can bound

$$\begin{aligned} Z &= \sum_{n=0}^{\infty} \frac{e^{-|D|} \beta^n}{n!} \int_{\dot{D}^n} \prod_{k=1}^n n(x_k, \vec{x}_{<k}) dx_1 \dots dx_n \\ &\geq \sum_{n=0}^{\infty} \frac{e^{-|D|} \beta^n}{n!} \int_{B^n(y, r)} \prod_{k=1}^n n(x_k, \vec{x}_{<k}) dx_1 \dots dx_n \\ &\geq e^{-|D|} \sum_{n=0}^{\infty} |B(y, r)|^n \beta^n, \end{aligned}$$

and the last series is infinite provided that $|B(y, r)|\beta \geq 1$. Proposition is proved.

Remark. The effect of non-integrability is similar to the effect observed in the case of Strauss model when the parameter γ responsible for the pairwise interaction is strictly greater than 1 ([10]).

2.2 The Hammersley-Clifford factorisation

In this section we obtain a Hammersley-Clifford factorisation formula for the density (3) in terms of a directed clique interaction function $\varphi : D \times F_D \rightarrow \mathbf{R}_+$ ([11]). We call a sequence \vec{y} an x -clique, $x \in D$, if all its components $y \in \vec{y}$ satisfy $x \sim y$.

Theorem 2.1 *The density (3) is factorised as follows*

$$f(\vec{x}) = \prod_{j=1}^{|\vec{x}|} \prod_{\vec{y} \subseteq \vec{x}_{<j}} \varphi(x_j, \vec{y}),$$

where $\varphi(x_j, \vec{y}) = 1$, if \vec{y} is not an x_j -clique. If a set \vec{y} is an x_j -clique, then

$$\varphi(x_j, \vec{y}) = \prod_{m=0}^{n=n(\vec{y})} \beta_m^{C(n(\vec{y}), m)}, \quad (11)$$

where the coefficients $C(n, m)$ are defined as follows

$$C(n, m) = (-1)^{n+m} \binom{n}{m}, \quad n \geq m \geq 0. \quad (12)$$

Proof of Theorem 2.1. For a set $\vec{y} = \{y_1, \dots, y_n\}$ which is an x -clique we have that

$$\varphi(x, \vec{y}) = \frac{\lambda(x|\vec{y})}{\prod_{\vec{z} \subseteq \vec{y}} \varphi(x, \vec{z})} = \frac{\beta_n}{\prod_{\vec{z} \subseteq \vec{y}} \varphi(x, \vec{z})}, \quad (13)$$

where $n = n(\vec{y})$ is the number of points in the x -clique \vec{y} . It is easy to see that particular location of points in \vec{y} is irrelevant for the value of $\varphi(x, \vec{y})$. Therefore $\varphi(x, \vec{y})$ depends only on $n = n(\vec{y})$ the number of points in \vec{y} . Denote

$$\psi_n = \log(\varphi(x, \{y_1, \dots, y_n\})), \quad n \geq 0.$$

The formula (13) yields the following relations for the coefficients ψ_n , $n \geq 0$,

$$\psi_0 = \log(\beta_0), \quad (14)$$

$$\psi_n = \log(\beta_n) - \sum_{k=0}^{n-1} \binom{n}{k} \psi_k, \quad n \geq 1. \quad (15)$$

Using these relations one can show that ψ_n , $n \geq 1$, is a linear combination of $a_m = \log(\beta_m)$, $m = 0, 1, \dots, n$ with some coefficients $C(n, m)$

$$\psi_n = \sum_{m=0}^n C(n, m) a_m. \quad (16)$$

It is easy to show by induction that

$$C(n, m) = \sum_{k=1}^{n-m} (-1)^k \sum_{n=i_0 > i_1 > i_2 > \dots > i_{k-1} > m} \binom{n}{i_1} \binom{i_1}{i_2} \dots \binom{i_{k-1}}{m}, \quad 0 \leq m \leq n-1, \quad (17)$$

$$C(n, n) = 1, \quad n \geq 1.$$

Indeed, it is easy to see that $\psi_1 = a_1 - a_0$, so the formula (16) holds for $n = 1$. Assume the formula (16) holds for some $n \geq 1$. Using (15) we have that

$$\begin{aligned}
\psi_{n+1} &= a_{n+1} - \sum_{k=0}^n \binom{n+1}{k} \psi_k = a_{n+1} + \sum_{k=0}^n (-1) \binom{n+1}{k} \sum_{m=0}^k C(k, m) a_m \\
&= a_{n+1} + \sum_{m=0}^n a_m \sum_{k=m}^n (-1) \binom{n+1}{k} A(k, m) \\
&= a_{n+1} + \sum_{m=0}^n C(n+1, m) a_m \\
&= \sum_{m=0}^{n+1} C(n+1, m) a_m
\end{aligned}$$

Note now that every term in the inner sum in (17) can be rewritten as follows

$$\binom{n}{i_1} \binom{i_1}{i_2} \cdots \binom{i_{k-1}}{m} = \binom{n}{m} \frac{(n-m)!}{(i_0 - i_1)! \cdots (i_{k-1} - i_k)!}$$

for any numbers $n = i_0 > i_1 > \dots > i_{k-1} > i_k = m, k \geq 1$. Therefore

$$C(n, m) = \binom{n}{m} \sum_{k=1}^{n-m} (-1)^k k! S(n-m, k), \quad (18)$$

where $S(n, m), n \geq 1, 1 \leq m \leq n$, are the so-called Stirling numbers of the second kind, i.e. the number $S(i, j)$ is the number of way to partition a set of i objects into j non-empty groups (see, for example, [3]). To get the representation (18) we used the following formula ([14], problem 2.62)

$$S(n, k) = \frac{1}{k!} \sum_{r_1, \dots, r_k} \frac{n!}{r_1! \cdots r_k!},$$

where the summation is over all sets of positive numbers r_1, \dots, r_k , such that $r_1 + \dots + r_k = n$.

The formula (17) can be explained as follows. Applying the formula (13) recursively we gain the factor β_m or its reciprocal every time when we jump one step down and choose an m -clique. Let us say that we get an m -points subset $\vec{z} = \{z_1, \dots, z_m\} \subset \vec{x} = \{x_1, \dots, x_n\}$ in k steps if we can represent a collection of subsets $\vec{z} \subset \vec{z}(i_k) \subset \vec{z}(i_{k-1}) \subset \dots \subset \vec{z}(i_1) \subset \vec{x}, n(\vec{z}(i_j)) = i_j, j = 0, 1, \dots, k$, such that $n = i_0 > i_1 > i_2 > \dots > i_k = m$. Every such subset, excluding the initial one, is obtained from its predecessor by taking away some (non-zero number) points. So, to compute the coefficient $A(n, m)$ we need to take with minus the number of possible ways to get an m -points subset of \vec{x} in 1 step (it is the number of those terms $\log(\beta_m)$ that we get after applying the recursion (15) to the terms ψ_k in the sum), then to add the number of getting it in 2 steps and so on. The minimal number of steps is 1 and the maximal number of steps is $n - m, 0 \leq m \leq n - 1$.

The last step is to note that for the Stirling numbers of the second type the following identity holds

$$\sum_{k=1}^M (-1)^k k! S(M, k) = (-1)^M, \quad (19)$$

for any $M \geq 1$. This identity is, in fact, the following classical identity (see [1], Section 24.1.4, p.825)

$$\sum_{k=1}^M (-1)^{M-k} k! S(M, k) = 1, \quad (20)$$

written in a slightly different form. Note, that the identity (19) can be proved by induction and the proof is based on the following recurrence relation for the Stirling numbers ([3])

$$S(M, k) = S(M-1, k-1) + kS(M-1, k), \quad M \geq 1$$

with initial conditions $S(0, 0) = 1$ and $S(0, k) = 0, k \neq 0$.

Remarks. 1) We noted at the beginning of the proof that the particular location of points in \vec{y} is irrelevant for the value of $\varphi(x, \vec{y})$, so $\varphi(x, \vec{y})$ depends only on $n = n(\vec{y})$ the number of points in \vec{y} . This is indeed the case since the conditional intensity $\lambda(u|\vec{y})$ of the process depends on the configuration \vec{y} only through the number of directed neighbours $n(u, \vec{y})$.

2) It is easy to see that the coefficients $C(n, m)$ form a table which is nothing else but the alternating Pascale triangle. The first seven rows of the triangle with appropriate signed elements are given below

$$\begin{array}{ccccccc} 1 & & & & & & \\ -1 & 1 & & & & & \\ 1 & -2 & 1 & & & & \\ -1 & 3 & -3 & 1 & & & \\ 1 & -4 & 6 & -4 & 1 & & \\ -1 & 5 & -10 & 10 & -5 & 1 & \\ 1 & -6 & 15 & -20 & 15 & -6 & 1 \end{array}$$

3) Examples of the clique interaction function

$$\begin{aligned} \varphi(x, \emptyset) &= \beta_0, \\ \varphi(x, \{x_1\}) &= \beta_0^{-1} \beta_1, \\ \varphi(x, \{x_1, x_2\}) &= \beta_0 \beta_1^{-2} \beta_2, \\ \varphi(x, \{x_1, x_2, x_3\}) &= \beta_0^{-1} \beta_1^3 \beta_2^{-3} \beta_3, \\ \varphi(x, \{x_1, x_2, x_3, x_4\}) &= \beta_0 \beta_1^{-4} \beta_2^6 \beta_3^{-4} \beta_4, \\ \varphi(x, \{x_1, x_2, x_3, x_4, x_5\}) &= \beta_0^{-1} \beta_1^5 \beta_2^{-10} \beta_3^{10} \beta_4^{-5} \beta_5. \end{aligned}$$

If there are finite number of non-zero intensities, i.e. $\beta_k = 0, k > N$, then there cannot be a clique with number of points larger than N , so the clique interaction functions of order higher than N are trivial, i.e. equal to 1.

4) Assume that the relation (1) between points was replaced by the following one

$$x \sim y \Leftrightarrow \|x - y\| \leq R(x), \quad (21)$$

where $R(x)$ is a positive and bounded function. It is easy to see that the structure of the clique interaction functions would remain the same.

5) In paper [8] a wide class of symmetric Markov point processes was considered. They were named the interacting point processes (INP). One particular case of INP process can be described in the following way. Consider the relation (1) as the symmetric one, the process states are the classic ones, i.e. unordered sets of points, and let the conditional intensity of the process at a point be β_k provided the point has k neighbours. It is easy to see that the clique interaction functions of this Markov point process are given by the formulas (11). Once a clique of a point is determined, i.e. either the number of points in it or its order is known, then the computations and the result is the same as in the proof of the theorem.

3 Local stability and simulation of CSA-processes

To simulate the process we use Metropolis-Hastings and birth-death samplers proposed for the sequential Markov point processes in [11]. For the reader's convenience we briefly recall the basic choices for both of them. Consider a sequential Markov process given by its density $g(\vec{\mathbf{x}})$ with respect to the reference measure ν . The generic variant of the Metropolis-Hastings sampler is the following one. Given the current state $\vec{\mathbf{x}}$ the birth and death are proposed equally likely, with probability 1/2. If birth is proposed, then we choose a position $i \in \{0, \dots, n(\vec{\mathbf{x}})\}$ at random, sample a uniformly distributed point $\xi \in D$ and accept the new state $s_i(\vec{\mathbf{x}}, \xi) = (\dots, x_{i-1}, \xi, x_i, \dots)$ with probability

$$\min \left\{ 1, \frac{g(s_i(\vec{\mathbf{x}}, \xi))}{g(\vec{\mathbf{x}})(n(\vec{\mathbf{x}}) + 1)} \right\}. \quad (22)$$

If death is proposed then a position $i \in \{1, \dots, n(\vec{\mathbf{x}})\}$ is chosen at random and the point x_i is removed with probability

$$\min \left\{ 1, \frac{g(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_{n(\vec{\mathbf{x}})})n(\vec{\mathbf{x}})}{g(\vec{\mathbf{x}})|D|} \right\}. \quad (23)$$

The generic choice of the birth-death sampler is determined by the birth rate

$$b_i(\vec{\mathbf{x}}, \xi) = \frac{g(s_i(\vec{\mathbf{x}}, \xi))}{(n(\vec{\mathbf{x}}) + 1)g(\vec{\mathbf{x}})}, \quad (24)$$

for the birth of a new point ξ at the position i given the current configuration $\vec{\mathbf{x}}$, and $d_i(\vec{\mathbf{x}}) = 1$ for the death rate at position i in the current configuration $\vec{\mathbf{x}}$.

Let us note that the Markov chains generated by both Metropolis-Hastings and birth-death algorithm are reversible with respect to the probability distribution specified by the density $g(\vec{\mathbf{x}})$ with respect to the reference measure ν ([11]). Therefore this distribution is an invariant one for the Markov chains. It is proved in [11] that a sufficient condition for geometric ergodicity for both the Metropolis-Hastings and birth-and-death samplers for a sequential Markov process with density g is the following local stability condition

$$\frac{g(s_i(\vec{\mathbf{x}}, \xi))}{g(\vec{\mathbf{x}})} \leq C, \quad (25)$$

uniformly in $\xi \in D$, $\vec{\mathbf{x}} = (x_1, \dots, x_n)$ and $i = 0, \dots, n$ for some $C > 0$. We assume in (25) that $g(\vec{\mathbf{x}}) > 0$. For more details we refer to [11], where the properties of more general Metropolis-Hastings algorithm for sequential Markov point processes were considered.

3.1 Local stability of CSA processes

We study local stability for the following two important subcases of the process. The first one is when there is only a finite number of non-zero intensities. The second one is when all intensities are positive but they can take only a finite number of different values.

The next theorem provides us with local stability bounds for both cases mentioned above. As we show afterwards (Remark 2) after the proof), in general these bounds cannot be improved.

Theorem 3.1 *Consider a CSA process with a finite number of different intensities.*

- 1) *If $\beta_k > 0$, $k = 0, 1, \dots, N$, $\beta_k = 0$, $k > N$ for some N , then the density (3) is locally stable with a constant*

$$C_1(N, \beta) = \max(\beta_0, \dots, \beta_N) \left(\frac{\max(\beta_0, \dots, \beta_N)}{\min(\beta_0, \dots, \beta_N)} \right)^{T(d)N}, \quad (26)$$

where $T(d)$ is the maximal number of points which can be allocated on d -dimensional unit sphere in such a way that the distance between any two of them is not less than 1.

- 2) *If $\beta_k > 0$, $k = 0, 1, \dots, N$, $\beta_k = \gamma > 0$, $k > N$ for some N and γ , then the density (3) is locally stable with a constant*

$$C_2(N, \beta) = \max(\beta_0, \dots, \beta_N, \gamma) \left(\frac{\max(\beta_0, \dots, \beta_N, \gamma)}{\min(\beta_0, \dots, \beta_N)} \right)^{T(d)N}, \quad (27)$$

where $T(d)$ is the constant defined in the part 1).

Proof of part 1) of Theorem 3.1. We give proof for the case $d = 2$, for higher dimensions the modification is obvious. Note that $T(2) = 6$. Denote for short

$$r_i(\vec{x}, \xi) = \beta_{n(\xi, \vec{x}_{<i})} \prod_{k=i}^n \frac{\beta_{n(x_k, (\vec{x}_{<k}, \xi))}}{\beta_{n(x_k, \vec{x}_{<k})}}. \quad (28)$$

For any ξ and $\vec{x}_{<i}$ we have that $\beta_{n(\xi, \vec{x}_{<i})} \leq \max(\beta_0, \dots, \beta_N)$. Let us bound the product. Note that $n(x_k, (\vec{x}_{<k}, \xi)) - n(x_k, \vec{x}_{<k})$ can be 0 or 1. If $n(x_k, \vec{x}_{<k}) = 0$ and $|\xi - x_k| > R$, then $n(x_k, (\vec{x}_{<k}, \xi)) - n(x_k, \vec{x}_{<k}) = 0$ and the corresponding ratio in the product is

$$\frac{\beta_{n(x_k, (\vec{x}_{<k}, \xi))}}{\beta_{n(x_k, \vec{x}_{<k})}} = 1.$$

If $|\xi - x_k| \leq R$, but $n(x_k, \vec{x}_{<k}) = N$, then $n(x_k, (\vec{x}_{<k}, \xi)) = N + 1$ and the corresponding ratio is

$$\frac{\beta_{N+1}}{\beta_N} = 0,$$

since $\beta_{N+1} = 0$. So, the point ξ can be inserted as the i th coordinate of the vector \vec{x} if and only if there are no points in the configuration $\vec{x} \cap B(\xi, R)$ that already have N neighbours. If $n(x_k, \vec{x}_{<k}) < N$ and $|\xi - x_k| \leq R$, then the product gains the factor $\beta_{n(x_k, \vec{x}_{<k})+1} / \beta_{n(x_k, \vec{x}_{<k})}$ that

we bound by $\max(\beta_0, \dots, \beta_N)/\min(\beta_0, \dots, \beta_N)$. If $\beta_{j+1}/\beta_j \leq 1$ for any $j = 0, \dots, N-1$, then the bound (25) is obvious and the proof is finished. In the general case we bound

$$r_i(\vec{\mathbf{x}}, \xi) \leq \max(\beta_0, \dots, \beta_N) \left(\frac{\max(\beta_0, \dots, \beta_N)}{\min(\beta_0, \dots, \beta_N)} \right)^{m_i(\xi, \vec{\mathbf{x}})},$$

where $m_i(\vec{\mathbf{x}}, \xi)$ is the number of $\vec{\mathbf{x}}$ components x_k , $k \geq i$, such that $n(x_k, \vec{\mathbf{x}}_{<k}) < N$ and $|\xi - x_k| \leq R$. A simple geometric argument shows that we can bound $m_i(\vec{\mathbf{x}}, \xi) \leq 6N$. Indeed, represent the ball $B(\xi, R)$ as a union of 6 equal disjoint sectors. It is easy to see that the distance between any two points belonging to the same sector is not more than R . Therefore there cannot be more than N points of the vector $\vec{\mathbf{x}}$ in any of these sectors. Indeed, assume there was a sector that contains at least $N+1$ points of the vector $\vec{\mathbf{x}} = (x_1, \dots, x_n)$, $n \geq N+1$. Let x_i be the last arrival in this sector, i.e. it is the point appeared there later than the others. Then x_i has already (at least) N neighbours (other vector points belonging to the sector) and it would have at least $N+1$ neighbours if we inserted ξ at the beginning of the vector and this situation is forbidden. So, finally

$$r_i(\vec{\mathbf{x}}, \xi) \leq \max(\beta_0, \dots, \beta_N) \left(\frac{\max(\beta_0, \dots, \beta_N)}{\min(\beta_0, \dots, \beta_N)} \right)^{6N}.$$

The part 1) of the theorem is proved.

Proof of part 2) of Theorem 3.1. Assume again for definiteness and without loss of generality that $d = 2$. For any ξ and $\vec{\mathbf{x}}_{<i}$ one can estimate $\beta_{n(\xi, \vec{\mathbf{x}}_{<i})} \leq \max(\beta_0, \dots, \beta_N, \gamma)$. Bounding the product (28) we distinguish two cases. If $|\xi - x_k| \leq R$ and $n(x_k, \vec{\mathbf{x}}_{<k}) > N$, then the corresponding ratio is trivial

$$\frac{\beta_{n(x_k, \vec{\mathbf{x}}_{<k})}}{\beta_{n(x_k, (\vec{\mathbf{x}}_{<k}, \xi))}} = 1.$$

If $|\xi - x_k| \leq R$, $n(x_k, (\vec{\mathbf{x}}_{<k}, \xi)) - n(x_k, \vec{\mathbf{x}}_{<k}) = 1$ and $n(x_k, \vec{\mathbf{x}}_{<k}) \leq N$, then the product gains the factor that might be not equal to 1 and we bound it as follows

$$\frac{\beta_{n(x_k, \vec{\mathbf{x}}_{<k})+1}}{\beta_{n(x_k, \vec{\mathbf{x}}_{<k})}} \leq \frac{\max(\beta_0, \dots, \beta_N, \gamma)}{\min(\beta_0, \dots, \beta_N)}.$$

So

$$r_i(\vec{\mathbf{x}}, \xi) \leq \max(\beta_0, \dots, \beta_N, \gamma) \left(\frac{\max(\beta_0, \dots, \beta_N, \gamma)}{\min(\beta_0, \dots, \beta_N)} \right)^{l_i(\xi, \vec{\mathbf{x}})},$$

where $l_i(\vec{\mathbf{x}}, \xi)$ is the number of components $x_k \in \vec{\mathbf{x}}$, $k \geq i$, such that $n(x_k, \vec{\mathbf{x}}_{<k}) \leq N$ and $|\xi - x_k| \leq R$. The geometric argument we used in the proof of part 1) gives that $l_i(\vec{\mathbf{x}}, \xi) \leq 6N$ and get the bound (27). The theorem is proved.

Remarks. 1) Similar local stability bounds can be obtained in the same way in the case when the relation (1) between points is replaced by the relation (21) with some positive and bounded function $R(x)$.

2) It is well known that the practical implementation of birth-and-death samplers in the locally stable case is based on dependent thinning of transitions of a Poisson process with unit death rate and constant birth rate C that dominates any $r_i(\vec{\mathbf{x}}, \xi)$: $r_i(\vec{\mathbf{x}}, \xi) \leq C$. If C is overestimated, then it leads to very low acceptance probabilities. Therefore it is desirable to get the bound (25) as

small as possible. From this point of view the bounds we have obtained in Theorem 3.1 are not satisfactory. We obviously overestimated the values of $m_i(\xi, \vec{x})$ and $l_i(\xi, \vec{x})$, since the points of any sector might have neighbours from other sectors. But in general both bounds (26) and (27) are close to optimal. Indeed, consider at first the bound (26) in the case when $\beta_0, \beta_1 > 0$ and $\beta_k = 0$ for $k \geq 2$. Then following the proof, one gets that the bound for $m_i(\xi, \vec{x})$ (the maximal number of points of the configuration \vec{x} which can contribute the non-trivial factor in the product (28)) is 6. One would get this number if the configuration shown in Figure 1 occurred, where 6 points of the vector \vec{x} are located at the boundary of the ball $B(\xi, R)$ such that the distance between the nearest of them is R . Obviously, this set of configurations has probability 0, since the ball boundary has Lebesgue measure 0 in dimension 2. But the set of configurations shown in Figure 2 has already a non-zero probability and gives the value $m_i(\xi, \vec{x}) = 5$.

Consider now the process with intensities $\beta_0 > 0, \beta_1 > 0, \beta_k = \gamma > 0, k \geq 2$. It is easy to see that the same "ideal" configuration in Figure 1 gives the bound (27) with $T(2) = 6$ and the configuration in Figure 2 provides the "lower bound" for the bound (27) and it is 5 since $l_i(\xi, \vec{x}) = 5$ in this case. The only difference is that now the points of configuration \vec{x} which we locate near the ball boundary, cannot have more than 1 neighbour outside the ball (otherwise the corresponding factor in the product (28) would be trivial, i.e. equal to 1).

The bounds we obtained in Theorem can be very large even for small N . For instance, if $N = 2$ and $\max(\beta_0, \dots, \beta_N) / \min(\beta_0, \dots, \beta_N) = 10$, then the bound is of order of 10^{12} . Therefore we can conclude that the birth-and-death sampler based on thinning is not a good practical choice for simulating of CSA processes.

3) It is easy to see that the condition (25) holds for any decreasing sequence of intensities $\beta_0 \geq \dots \geq \beta_k \geq \beta_{k+1} \geq \dots, k \geq 0$, which is not necessarily finite, since in this case the factor $\beta_n(x_k, (\vec{x}_{<k}, \xi)) / \beta_n(x_k, \vec{x}_{<k})$ is always bounded by 1.

4) If we put $\beta_k = \beta > 0$ for any $k \geq 0$, then we obtain the sequential Poisson point process. In this case the local bound is just β .

5) If the set of intensities is unbounded, then the density (3) may not be locally stable. Indeed, assume that $\beta_i \rightarrow \infty$ as $i \rightarrow \infty$. Let the area D contain at least one ball $B(y, r)$ with radius $r < R/2$. Consider a configuration $\vec{x} = (x_k, k = 1, \dots, n) \in B(y, r)$. Then, for $\xi \in B(y, r)$

$$r_0(\vec{x}, \xi) = \beta_{n(\xi, 0)} \prod_{k=1}^n \frac{\beta_n(x_k, (\vec{x}_{<k}, \xi))}{\beta_n(x_k, \vec{x}_{<k})} = \beta_0 \prod_{k=0}^{n-1} \frac{\beta_{k+1}}{\beta_k} = \beta_n,$$

which is unbounded in n .

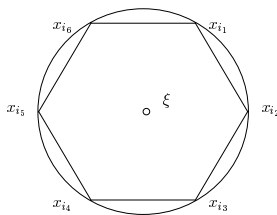


Figure 1: Six points at the surface with the minimal distance R between them, R is radius of the ball. It is assumed that the points do not have any neighbours outside the ball.

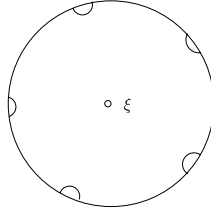


Figure 2: Five domains with positive measures near the ball boundary. The distances between the domains are not less than the radius of the ball. A vector \bar{x} has a point of intersection with each of the domains, these points do not have any neighbours outside the ball.

4 Modelling of repulsion and attraction in sequential point patterns by CSA processes

An intuitive picture that one may expect is the following one. If the intensities are decreasing, i.e. $\beta_0 \geq \dots \geq \beta_k \geq \beta_{k+1} \geq \dots$, then it might result in repulsion between points. If the intensities are increasing, i.e. $\dots \leq \beta_k \leq \beta_{k+1} \leq \dots$, then one expect to observe the aggregation effect in a point pattern. Basically, these predictions hold, but there are certain limitations that will be mentioned.

Consider the simplest non-trivial situation, when there are only two non-zero intensities β_0 and β_1 , i.e. a point can have at most one neighbour. If $\beta_0 \ll \beta_1$, then clustered patterns appear, see Figure 4. The clusters obviously have a *different* structure than clusters formed, for example, by a Poisson cluster process. The CSA-clusters are formed by "directed local chains" of points which approximately have "a local linear structure". Figure 3 illustrates the mechanism of the chain formation. The areas labeled by + are areas where the intensity of point appearance is

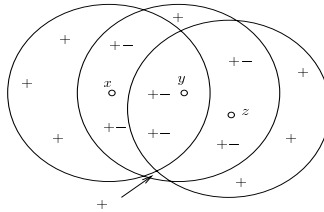


Figure 3: Formation of the "linear cluster".

high, $+ -$ areas are the ones which were active but are blocked at the moment. Point x is the first one, it activates the area in the ball $B(x, R)$. Then point y appears and immediately the area $B(x, R) \cap B(y, R)$ is frozen. Point z is the third one, its appearance blocks the intersection $B(z, R) \cap B(y, R)$. Using the terminology of the theory of cluster processes one can name point x as the parent point and x has a daughter point y . Point y itself can be considered as a parent point for z . It should be noted that if a value of β_1 is very large, then it might spoil the cluster structure resulting in regular patterns with high density of points, even if β_0 is rather small. See, for instance, Figures 7 and 8, where the patterns have a very regular structure. This "regularity effect" is explained analytically in Subsection 4.1, see cases 4) and 5). In the case $\beta_0 \gg \beta_1$ and β_1 is small patterns with repulsion between points are observed. These patterns are similar to patterns produced by the hard core process with the same intensity β_0 (compare the images provided by Figures 5 and 6). The patterns in Figure 6 and the right-hand side patterns in Figures 4 and 5

have visible regular structure since β_0 is relatively large and β_1 is relatively small.

With some obvious modification the same mechanism of cluster formation works in the case of any CSA process with a finite number of non-zero intensities and the same mixed attraction-repulsion effects can be observed. Images of a CSA model with 10 non-zero intensities are provided in Figure 9. The left-hand side image in Figure 9 shows the pattern with a rather regular structure, though clusters with small numbers of points are visible. Apparently, this is stipulated by the moderate growth of intensities. The clusters cannot grow more, because relatively large areas are becoming saturated by relatively small numbers of points. Apparently, this is due to the fact that the interaction radius is sufficiently large. The right-hand side image in Figure 9 has a much more aggregated structure, since the interaction radius is smaller and the growth of intensities is more substantial.

Another easy to handle choice for modelling of cluster structures is provided by CSA processes with a finite number of different intensities in case when all of them are positive. The mechanism of cluster formation is the same one as explained above. The only difference is that there are no restrictions on the number of neighbours in this situation. Density and sizes of clusters can be controlled by making at least one of the intensities smaller or bigger. For examples of images see Figures 10 and 11. By varying the interaction radius (which is a very important factor) one can also reduce or intensify clustering in point patterns. For example, compare the images provided in Figure 11. The same cautions that we have mentioned in the case of processes with hard core interaction should be taken into account when simulating cluster point patterns. The cluster structure can be transformed into the regular one if, say, β_1 and successive intensities are taken too large. For instance, the images in Figure 11 have visible cluster structure, clusters are separated one from each other. In the left image of Figure 12 the clusters begin coalescing though they still can be recognized. The right image in Figure 12 gives an example when clusters have coalesced forming one "connected component". The last makes sense if we draw a ball of small radius around every point.

Remark. Note that the images in Figure 10 look like images of the CSA process with only two non-zero intensities. This similarity is generated, of course, by the choice of parameters $\beta_0 \ll \beta_1$ and $\beta_1 \gg \beta_k$, $k \geq 2$.

4.1 Extreme behaviour of CSA patterns.

To emphasize the idea and for simplicity of notation we are going to consider the limit behavior of CSA patterns when there are only two non-zero intensities β_0 and β_1 . We describe the cases when β_0 and β_1 tend to their extreme values 0 or ∞ . In particular, we give analytical explanation why the clusterisation effects vanish when $\beta_0 < \beta_1$ and β_1 is taken sufficiently large.

Due to the hard-core character of interaction the partition function is in fact a sum of finitely many terms

$$Z = \sum_{m=0}^{C(\beta, R, D)} \frac{e^{-|D|}}{n!} \int_{D^m} \prod_{k=1}^m \beta_{n(x_k, \vec{x}_{<k})} 1_{\{\max\{k:t_k(\vec{x})>0\}=1\}} dx_1 \dots dx_m,$$

for some constant $C(\beta, R, D)$. Let us assume for definiteness, that D is a unit cube. It is easy to see that for a given sequence of points \vec{x}

$$\alpha(\vec{x}) = \int_D \beta_{n(x, \vec{x})} dx = \int_D (\beta_0 1_{\{n(x, \vec{x})=0\}} + \beta_1 1_{\{n(x, \vec{x})=1\}}) dx = \beta_0 |D_0(\vec{x})| + \beta_1 |D_1(\vec{x})|, \quad (29)$$

where $D_i(\vec{x}) = \{x \in D : n(x, \vec{x}) = i\}$, $i = 0, 1$. Therefore integrating backward in

$$\int_{D^m} \prod_{k=1}^m \beta_{n(x_k, \vec{x}_{<k})} dx_1 \dots dx_m,$$

i.e. fix (x_1, \dots, x_{m-1}) , integrate over x_m and so on, and using equation (29) we get that

$$\int_{D^m} \prod_{k=1}^m \beta_{n(x_k, \vec{x}_{<k})} dx_1 \dots dx_m = \sum_{i,j} \beta_0^i \beta_1^j a_m(i, j), \quad (30)$$

with some coefficients $\{a_m(i, j)\}$, where the sum is over all possible integers $i \geq 1, j \geq 0, i + j = m$, and only a finite number of coefficients $a_m(i, j) \geq 0$ in this expansion are non-zero. The $a_m(i, j)$ do not depend on β_0, β_1 , only on the interaction radius R . So, the partition function is just a finite sum of terms $\beta_0^i \beta_1^j, i > 0$. To understand the structure of limit patterns it suffices to know a term determining the partition function asymptotics in a particular extreme regime.

1) $\beta_0 \rightarrow 0$ and β_1 is bounded. The limit process is trivial, its realisation is just an empty configuration, since for any $0 < k \leq C(\beta, R, D)$ we have that $P\{X = k\} \sim \beta_0$ as $\beta_0 \rightarrow 0$.

2) $\beta_0 \rightarrow \infty$ and β_1 is bounded. The terms $\beta_0^{j_{max}} \beta_1^j$ with maximal power of β_0 will determine the asymptotics of the partition function and the limit distribution will concentrate on corresponding configurations. If β_1 is sufficiently small, then the limit process is similar to the pure hardcore process, i.e. $\beta_0 > 0, \beta_1 = 0$ in the limit $\beta_0 \rightarrow \infty$.

3) $\beta_0 \rightarrow \infty$ and $\beta_1 \rightarrow \infty$. In general, the structure of limit configurations depends on the way the pair (β_0, β_1) approach the limit (∞, ∞) .

4) $\beta_0 = const$ and $\beta_1 \rightarrow \infty$. Let j_{max} be a maximal power of β_1 in the expansion for the partition function. Then

$$Z \sim \left(\sum_m \sum_i a_m(i, j_{max}) \beta_0^i \right) \beta_1^{j_{max}}$$

as $\beta_1 \rightarrow \infty$, so the configurations with $t_1(\vec{x}) = j_{max}$ will survive. The limit process distribution is concentrated on the configurations \vec{x} with statistics $t_1(\vec{x}) = j_{max}$. One can imagine that when a point appears (with intensity β_0), then the "maximal chain" with j_{max} points is instantly formed.

5) $\beta_0 \rightarrow 0$ and $\beta_1 \rightarrow \infty$. If $\beta_0 \beta_1^{j_{max}} \rightarrow 0$ as $\beta_0 \rightarrow 0$ and $\beta_1 \rightarrow \infty$, then the limit distribution is trivial. If $\beta_0 \beta_1^{j_{max}} \rightarrow \infty$ (or just stays bounded) and $\beta_0 \beta_1^{j_{max}-1} \rightarrow 0$ as $\beta_0 \rightarrow 0$ and $\beta_1 \rightarrow \infty$, then the limit process exists. It is non-trivial and its distribution is concentrated on the configurations \vec{x} with $t_0(\vec{x}) = 1$. Therefore in the limit $(\beta_0, \beta_1) \rightarrow (0, \infty)$ the limit distribution depends on the way the pair (β_0, β_1) tends to their extreme values.

The cases 4) and 5) show that for any $\beta_0 > 0$, regardless how small it is, if β_1 is sufficiently large, then regular high density (the density depends on interaction radius) patterns should appear. Images in Fig.7 and 8 certainly confirm these theoretical predictions. The images demonstrate visible regular high density structure.

Note, that in all cases when there exists a non-trivial limit process, i.e. the probability of non-empty configuration is strictly less than 1, the limit process is not a Markov process, since its density with respect to the reference measure becomes non-hereditary.

The images in Figures 4-11 were obtained by the Metropolis-Hastings algorithm.

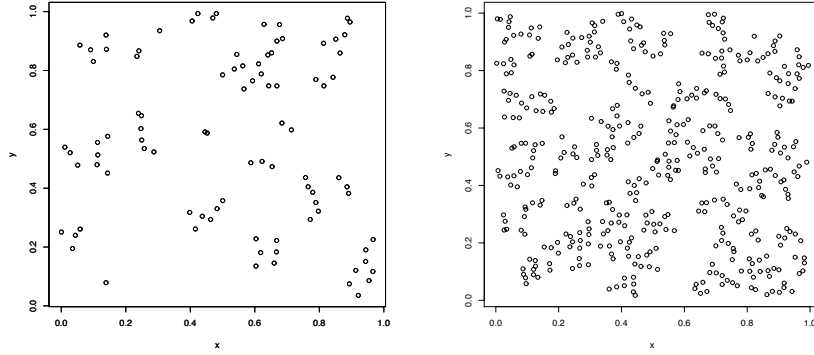


Figure 4: CSA patterns with two non-zero intensities. Left: $R = 0.05$ and $(\beta_0 = 5, \beta_1 = 500)$. Right: $R = 0.03$ and $(\beta_0 = 10, \beta_1 = 2000)$. Statistics are $(20, 72)$ and $(76, 367)$ respectively.

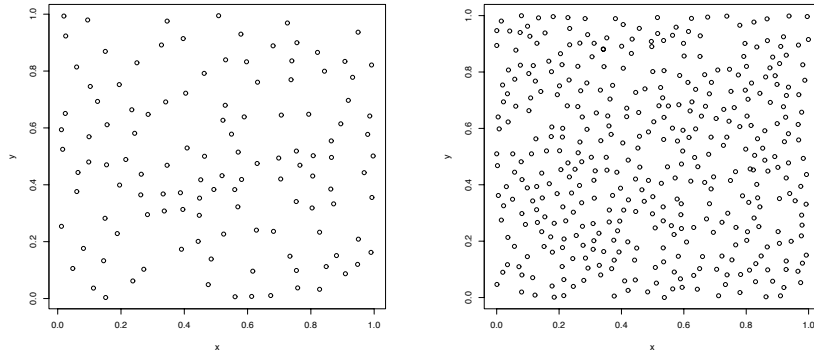


Figure 5: CSA patterns with two non-zero intensities. Left: $R = 0.05$ and $(\beta_0 = 500, \beta_1 = 5)$. Right: $R = 0.03$ and $(\beta_0 = 2000, \beta_1 = 10)$. Statistics are $(121, 1)$ and $(391, 4)$ respectively.

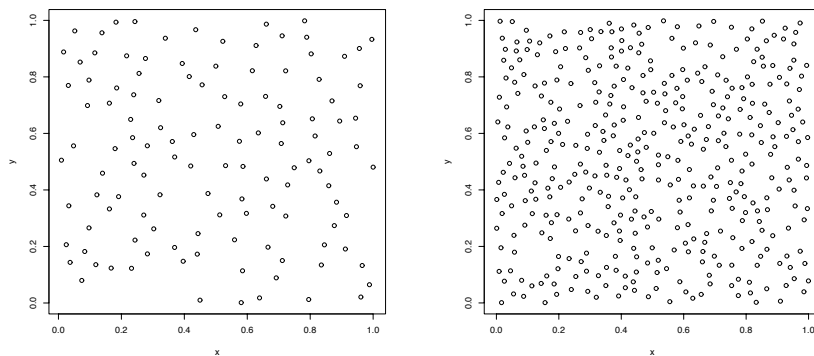


Figure 6: Sequential hard-core patterns, $\beta_k = 0, k \geq 1$. Left: $R = 0.05$ and $\beta_0 = 500$. Right: $R = 0.03$ and $\beta_0 = 2000$. The total numbers of points are 121 and 408 respectively.

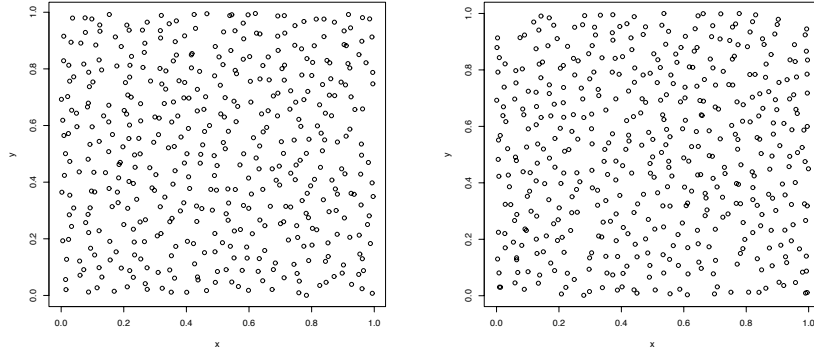


Figure 7: CSA patterns with two non-zero intensities, $R = 0.05$. Left: $(\beta_0 = 2, \beta_1 = 10002)$. Right: $(\beta_0 = 0.5, \beta_1 = 20000)$. Statistics are $(24, 419)$ and $(12, 451)$ respectively.

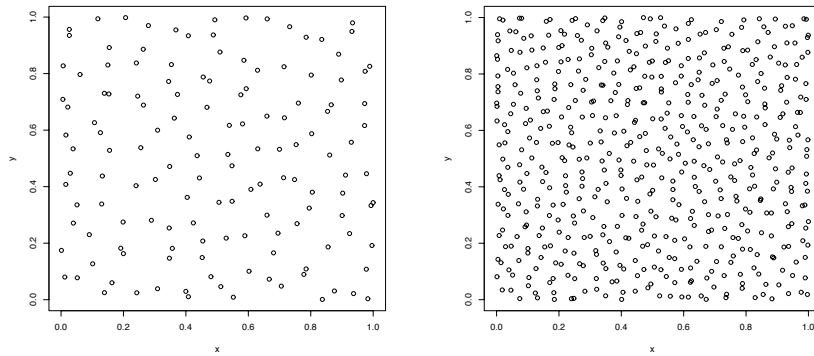


Figure 8: CSA patterns with two non-zero intensities. Left: $R = 0.1$ and $(\beta_0 = 0.0005, \beta_1 = 500000)$, statistics $(1, 139)$. Right: $R = 0.05$ and $(\beta_0 = 0.0005, \beta_1 = 500000)$, statistics $(1, 544)$.

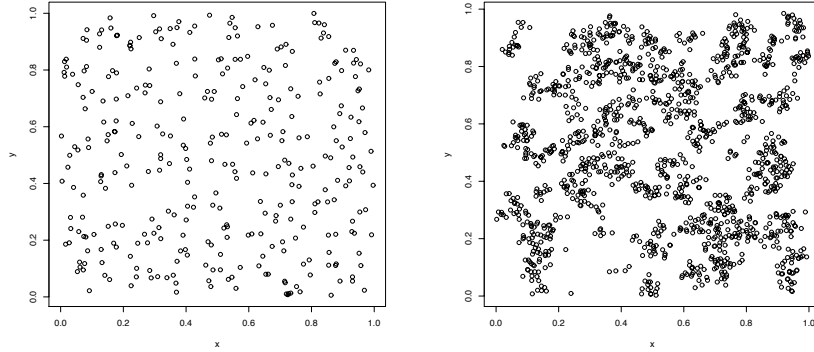


Figure 9: CSA patterns with 10 non-zero intensities. Left: $R = 0.1$ and $\beta_0 = 10, \beta_1 = 110, \beta_2 = 210, \beta_3 = 310, \beta_4 = 410, \beta_5 = 510, \beta_6 = 610, \beta_7 = 710, \beta_8 = 810, \beta_9 = 910$. Right: $R = 0.03$ and $\beta_0 = 10, \beta_1 = 510, \beta_2 = 1010, \beta_3 = 1510, \beta_4 = 2010, \beta_5 = 2510, \beta_6 = 3010, \beta_7 = 3510, \beta_8 = 4010, \beta_9 = 4510$. Statistics are $(16, 46, 29, 35, 42, 40, 40, 31, 22, 29)$ and $(79, 254, 236, 179, 146, 134, 111, 86, 72, 43)$ respectively.

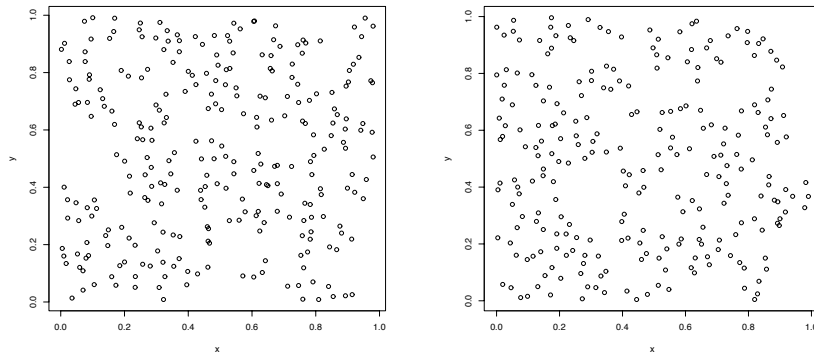


Figure 10: CSA patterns with 3 different intensities, $R = 0.05$. Left: $\beta_0 = 10, \beta_1 = 1010, \beta_k = 100, k \geq 2$. Right: $\beta_0 = 10, \beta_1 = 1010, \beta_k = 10, k \geq 2$. Statistics are $(42, 211, 25)$ and $(36, 213, 1)$ respectively.

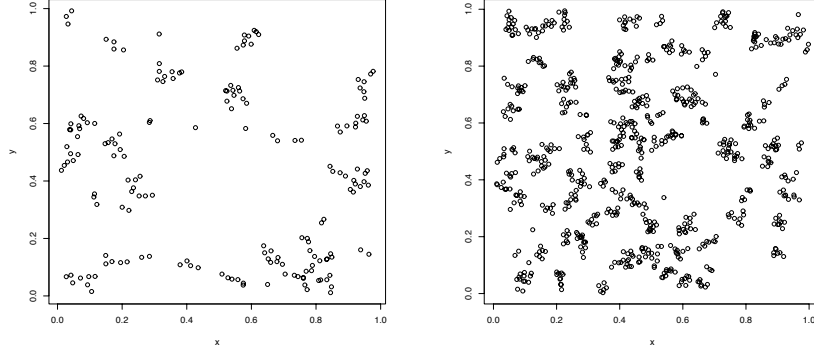


Figure 11: CSA patterns with 3 (left) and 2 (right) different intensities. Left: $R = 0.03$ and $\beta_0 = 5, \beta_1 = 1000, \beta_k = 500, k \geq 2$, and statistics (35, 105, 37). Right: $R = 0.02$ and $\beta_0 = 10, \beta_k = 2000, k \geq 1$, and statistics (93, 661).

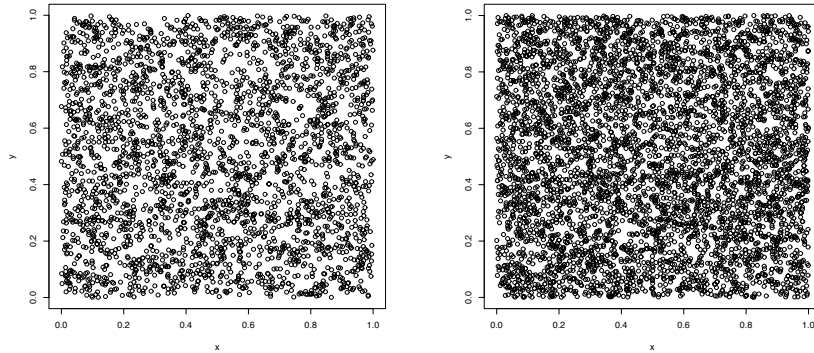


Figure 12: CSA patterns with 2 different intensities. Left: $R = 0.03$ and $\beta_0 = 5, \beta_k = 3000, k \geq 2$, the total number of points 2971, the maximal number of neighbours 18. Right: $R = 0.07$ and $\beta_0 = 5, \beta_k = 5000, k \geq 1$, the total number of points 5057, the maximal number of neighbours 97.

5 Application of the CSA processes for the analysis of the real life data

In this section we try to fit the CSA model to the real life data provided by Drs. R. Huele, Leiden University, Institute of Environmental Sciences (CML). Our main aim here is to illustrate by a concrete example how CSA processes can be used in applications.

5.1 Data

The data is a sequence of geographical coordinates of the North Sea points where appearances of harbour porpoises were recorded. We decided on the following observation window: the interior of a polygon with vertices

$$A_1(2.00, 51.26), A_2(2.00, 56.00), A_3(8.20, 56.00), A_4(8.20, 54.00), A_5(4.47, 52.55), A_6(3.20, 51.26),$$

where the first coordinate is the geographical longitude and the second one is the geographical latitude. The line A_6A_5 approximates the North Sea coast of the Netherlands from Den Helder to a point near Knokke-Heist the place where the Belgium-Dutch border touches the North Sea. The line A_5A_4 approximates the North Sea coast from Den Helder to point A_4 in the North Sea near Cuxhaven in Denmark. There is a shallow water territory in the North Sea to the right of the line A_5A_4 and below the line A_6A_5 , but we discard these regions. The area of the shallow water regions is negligible and does not influence the analysis. We transform the observation window as follows $x \rightarrow \tilde{x} = (x - 2.0)/6.2$ (the vertical coordinate) and $y \rightarrow \tilde{y} = (y - 51.26)/4.74$ (the horizontal coordinate), so that the transformed observation window is now a subset of the unit square. Denote the images of the vertices by the same letters A_i , $i = 1, \dots, A_6$, then the transformed window is bounded by the segment A_1A_2 connecting the points $(0.0000, 0.0000)$ and $(0.0000, 1.0000)$, the segment A_2A_3 connecting the points $(0, 1)$ and $(1.0000, 1.0000)$, the segment A_3A_4 connecting the points $(1.0000, 1.0000)$ and $(1.0000, 0.5781)$, the segment A_4A_5 connecting the points $(1.0000, 0.5781)$ and $(0.3984, 0.2722)$ (the line equation $y = 0.5085x + 0.0700$), the segment A_4A_5 connecting $(0.3984, 0.2722)$ and $(0.1935, 0.0000)$ (the line equation $y = 1.3600x - 0.2721$) and by the segment A_5A_6 connecting the point $(0.1935, 0.0000)$ with the origin $(0.0000, 0.0000)$. The image of the data in the transformed window is plotted in Fig. 13. The records, initial and transformed, are stored in Table 1.

We denote by \vec{x} the sequence of points with transformed coordinates \tilde{x} and \tilde{y} . The point pattern in Fig. 13 is a low density point pattern. There are visible pairs of points, that can be interpreted as a parent and a daughter point in a CSA cluster. There are no visible clusters with more than 2 points. Our experience based on many computer experiments says that such a pattern is typical for CSA models with two non-zero intensities, where $\beta_0 \ll \beta_1$, β_0 itself is not large, and the interaction radius is sufficiently small in comparison with the size of the observation window. So the subjective impression suggests that a CSA model with two non-zero intensities could be a possible candidate. Though other models might fit as well. For definiteness we restrict our analysis to CSA models with hard core interaction where the interaction radius R is treated as an unknown parameter.

The main component of our analysis is the method of MCMC maximum likelihood estimation (MCMCMLE) applied to a CSA model with a finite number of non-zero intensities and known radius. To proceed further we need to establish applicability of this general method. We do it in the next section. For the reader convenience we recall also the key steps of MCMCMLE. For

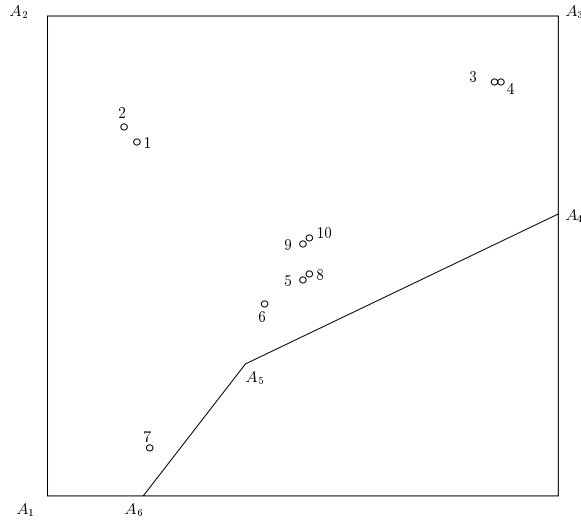


Figure 13: The observation window and data

	Longitude, x	Latitude, y	\tilde{x}	\tilde{y}
1	3.0875	54.7700	0.1754	0.7398
2	2.9404	54.9400	0.1517	0.7773
3	7.4750	55.3167	0.8831	0.8558
4	7.5117	55.3167	0.8890	0.8558
5	5.1433	53.4100	0.5070	0.4536
6	4.6933	53.1200	0.4344	0.3924
7	3.4000	51.7817	0.2258	0.1101
8	5.2500	53.4400	0.5242	0.4599
9	5.0583	53.8267	0.4933	0.5415
10	5.2150	53.8583	0.5185	0.5482

Table 1: Data

detailed description, discussions and applications of the approach see [5], [6], [7], [9] and references therein.

5.1.1 MCMC maximum likelihood estimation for the CSA model with hard core interaction

The formula (9) means that given R , the interaction radius, and $N + 1$, the number of non-zero intensities, the point process forms an exponential family. We denote by θ the set of parameters $\{\theta_k = \log(\beta_k), k \leq N\}$ and canonical statistics in this case are $t_k(\vec{\mathbf{x}})$, $k \leq N$. The unnormalized density determined by parameters θ will be denoted by h_θ . Also, denote by \mathbf{P}_θ and \mathbf{E}_θ the probability distribution of the process with parameter intensities θ and expectation with respect to \mathbf{P}_θ .

Given observation $\vec{\mathbf{x}}$ we are looking for a set of numbers $\hat{\theta}_k$, $k = 0, \dots, N$, that maximize the log-likelihood

$$L(\theta, \vec{\mathbf{x}}) = \sum_{k=0}^N \theta_k t_k(\vec{\mathbf{x}}) - \log(Z(\theta)), \quad (31)$$

$Z(\theta)$ is the partition function corresponding to the parameters θ . Note that the partition function $Z(\theta)$ depends smoothly on θ and can be differentiated under the integral sign since it is formed by a sum of a finite number of terms smoothly depending on θ . To use the MCMCMLE method we also need to prove positive definiteness of $\Phi(\theta) = -\nabla^2 L(\theta, \vec{\mathbf{x}})$ the Fisher information matrix of the model. The proof of this fact is given in Section 7.

Since the log-likelihood is a concave function of θ there exists either a unique maximum in the interior of the parameter space or a unique maximum on the boundary. If the maximum is achieved in an interior point $\hat{\theta}$ of Θ , then $\hat{\theta}$ can be found as a unique solution of the maximum log-likelihood equation

$$\nabla L(\theta, \vec{\mathbf{x}}) = 0,$$

which is the following set of equations

$$t_k(\vec{\mathbf{x}}) = \mathbf{E}_\theta t_k(X), \quad k = 0, 1, \dots, N, \quad (32)$$

We cannot solve the system of equations (32) directly, since the partition function $Z(\theta)$ in (31) cannot be computed. Therefore MCMC approximation is used.

Note that any two densities h_θ and h_ψ that are determined by different sets of intensities and with the same value of the interaction radius have the common support, i.e. the set of configurations where they are non-zero. Therefore, fixing a reference parameter $\psi = (\psi_0, \dots, \psi_N)$, we can consider the log-likelihood ratio against the parameter ψ

$$l_\psi(\theta, \vec{\mathbf{x}}) = L(\theta, \vec{\mathbf{x}}) - L(\psi, \vec{\mathbf{x}}) = \log \left(\frac{h_\theta(\vec{\mathbf{x}})}{h_\psi(\vec{\mathbf{x}})} \right) - \log \left(\frac{Z(\theta)}{Z(\psi)} \right). \quad (33)$$

Obviously, if $\hat{\theta}$ maximizes $L(\theta, \vec{\mathbf{x}})$, then it also maximizes $l_\psi(\theta, \vec{\mathbf{x}})$ and vice versa. The Monte Carlo approximation to the true log-likelihood ratio with respect to the parameter ψ is defined as

$$l_{\psi,m}(\theta, \vec{\mathbf{x}}) = \sum_{k=0}^N (\theta_k - \psi_k) t_k(\vec{\mathbf{x}}) - \log \left(\frac{1}{m} \sum_{i=1}^m e^{\sum_{k=0}^N (\theta_k - \psi_k) t_k(X_i)} \right), \quad (34)$$

where X_1, \dots, X_m is a sample from a Markov chain having P_ψ as its stationary distribution. We generate the sample X_1, \dots, X_m using the Metropolis-Hastings algorithm described in [11] and where it is proved that the corresponding Markov chain is reversible and geometrically ergodic. Moreover, in the case of a CSA process with a finite number of non-zero intensities this Markov chain is even uniformly ergodic due to the hard-core character of interaction.

Sampling analogues of the maximum likelihood equations are obtained by differentiating $l_{\psi,m}(\theta, \vec{\mathbf{x}})$ with respect to θ_k , $k = 0, 1, \dots, N$

$$t_k(\vec{\mathbf{x}}) = \frac{\sum_{i=1}^m t_k(X_i) e^{\sum_{k=0}^N (\theta_k - \psi_k) t_k(X_i)}}{\sum_{i=1}^m e^{\sum_{k=0}^N (\theta_k - \psi_k) t_k(X_i)}}, \quad k = 0, \dots, N \quad (35)$$

If the maximizer of $l_{\psi,m}(\theta, \vec{\mathbf{x}})$ is an interior point of Θ , then it can be found as a solution of the equation (35).

The random variable $h_\theta(X)/h_\psi(X)$ is integrable with respect to the measure P_ψ (it is even bounded here, as one can see). Therefore by the ergodic theorem for Markov chains we get that

$$\frac{1}{m} \sum_{i=1}^m \frac{h_\theta(X_i)}{h_\psi(X_i)} \rightarrow \mathbb{E}_\psi \frac{h_\theta(X)}{h_\psi(X)} = \frac{Z(\theta)}{Z(\psi)} \quad (36)$$

almost surely as $m \rightarrow \infty$. Hence, we get that for any fixed observation $\vec{\mathbf{x}}$ the Monte Carlo log-likelihood ratio $l_{\psi,m}(\theta, \vec{\mathbf{x}})$ converges almost surely to $l_\psi(\theta, \vec{\mathbf{x}})$ as $m \rightarrow \infty$. If $\hat{\theta} = (\hat{\theta}_0, \dots, \hat{\theta}_N)$ are the true maximum likelihood estimators and $\hat{\theta}_m = (\hat{\theta}_{m,0}, \dots, \hat{\theta}_{m,N})$ are the maximizers of the Monte Carlo log-likelihood $l_m(\theta, \vec{\mathbf{x}})$, then by the general results ([5]) for exponential families, $\hat{\theta}_m \rightarrow \hat{\theta}$ almost surely as $m \rightarrow \infty$.

Following [6] we can estimate the variance of error $\hat{\theta}_m - \hat{\theta}$, where $\hat{\theta}$ is the true likelihood estimator. The estimation is based on the fact that by Theorem 7 of [6] the normalized MC score function $m^{1/2}(\hat{\theta}_m - \hat{\theta})$ converges weakly to a normal distribution $N(0, \Phi_{\hat{\theta}}^{-1} A \Phi_{\hat{\theta}}^{-1})$ as $m \rightarrow \infty$, where $\Phi_{\hat{\theta}}$ is the Fisher information matrix at the true maximum likelihood estimator $\hat{\theta} = (\hat{\theta}_0, \dots, \hat{\theta}_N)$, the matrix A is the limiting covariance matrix of $m^{1/2} \nabla l_m(\hat{\theta}_m)$ and is defined as follows

$$A = \left(\frac{Z(\psi)}{Z(\theta)} \right)^2 \sigma_g^2, \quad (37)$$

where

$$\sigma_g^2 = \sum_{t=-\infty}^{\infty} \gamma_g(t),$$

$\gamma_g(t)$ is an $M \times M$ -matrix with elements $\gamma_{ij}(t) = Cov(g_i(X_0), g_j(X_t))$, $i, j = 1, \dots, M$, where

$$g(X) = (g_0(X), \dots, g_N(X)) = \left((t_k(\vec{\mathbf{x}}) - t_k(X)) \frac{h_\theta(X)}{h_\psi(X)}, k = 0, \dots, N \right).$$

The covariance matrix σ_g^2 is well defined in the sense that all its elements are finite and this is due to the boundedness of the functional $g(X)$ with respect to the measure P_ψ .

By Theorem 7 of [6] the matrix $\Phi_{\hat{\theta}}$ can be approximated by $\hat{\Phi}_m(\hat{\theta}_m) = -\nabla^2 l_m(\theta, \vec{\mathbf{x}})|_{\theta=\hat{\theta}_m}$. The ratio $Z(\psi)/Z(\theta)$ is estimated by $(\mathbb{E}_{m,\psi} h_\theta(X)/h_\psi(X))^{-1}$ (see the equation (36)). We estimate the

R	Statistics $t_k = t_k(R), k = 0, \dots$	$\widehat{N}(R) + 1$
0.001, ..., 0.005	(10, 0)	1
0.006, ..., 0.018	(9, 1)	2
0.019, ..., 0.026	(8, 2)	2
0.027, ..., 0.044	(7, 3)	2
0.045, ..., 0.087	(6, 4)	2
0.088	(5, 5)	2
0.089, ..., 0.094	(5, 3, 2)	3
0.095	(4, 4, 2)	3
0.096, ..., 0.1	(4, 4, 1, 1)	4

Table 2: Statistics and the number of non-zero intensities for given radius

matrix σ_g^2 using the method of batch means, which is the standard method in the time series analysis for estimating variances of sample means (see, for instance, [15]).

Remark. The weak convergence of the normalized MC score function to a normally distributed vector $N(0, A)$ holds here by the central limit theorem for square integrable functionals of reversible, geometrically ergodic Markov chains ([19]).

5.2 Back to the data analysis

The number of non-zero intensities is unknown and $\widehat{N}(\vec{\mathbf{x}})$ can be taken as an estimator of N . It is obvious, that $\widehat{N}(\vec{\mathbf{x}})$ is a biased estimator of N , since $\widehat{N}(\vec{\mathbf{x}}) \leq N$ for any realisation $\vec{\mathbf{x}}$. Since the statistic \widehat{N} underestimates the true value of N , different CSA models with $N \geq \widehat{N}$ non-zero intensities could be tested for the same values of the interaction radius. In fact, the following observation reduces the choice in the model selection if we look for the maximum likelihood estimators. Namely, let the interaction radius R be known and the observed point pattern $\vec{\mathbf{x}}$ be such that $\widehat{N}(\vec{\mathbf{x}}) = N$. Assume that we try to fit the model with $N + n - 1$ non-zero intensities, where $n \geq 2$, and let $\widehat{\theta}_k(\vec{\mathbf{x}}), k = 0, 1, \dots, N + n - 1$, be the solutions of the maximum likelihood equations

$$t_k(\vec{\mathbf{x}}) = \mathbb{E}_{\theta} t_k(X), \quad k = 0, 1, \dots, N. \quad (38)$$

Then necessarily $\widehat{\theta}_k(\vec{\mathbf{x}}) = -\infty$, or equivalently, $\widehat{\beta}_k = 0$ for $k = N + 2, \dots, N + n$. Indeed, if $\theta_k > -\infty$ ($\beta_k > 0$), then it is easy to show that $\mathbb{E}_{\theta} t_k(X) > 0$. Therefore, if for a given value of the interaction radius we observe $\widehat{N}(\vec{\mathbf{x}}) = N$, then the only choice among the models with hard core interaction is the model with $N + 1$ non-zero intensities. For instance, in the particular case of our data set there is no need to test the hard core model with three non-zero intensities if the true value of the interaction radius is in the range $[0.006, 0.088]$ (see Table 2). It is easy to see that given the intensities β 's the likelihood function is a piecewise constant function of R . Therefore we use a profile likelihood approach (see, for instance, [16], p.163). It means that we first compute maximum profile likelihood estimates

$$\theta(\vec{\mathbf{x}}, R_k) = \arg \max_{\theta} L(\theta, \vec{\mathbf{x}}, R_k),$$

over some range of possible values of the interaction radius. Note that we now stress dependence of L on the interaction radius R . We decided that the true value of the radius is somewhere in the range

[0.001, 0.1] and we approximate this continuous set by the discrete set $R_k = 0.001 \cdot k$, $k = 1, \dots, 100$, that covers the segment [0.001, 0.1] sufficiently dense. The table below gives the values of canonical statistics and the estimated number of non-zero intensities for a given value of the interaction radius.

For a fixed value $R = R_k$ one can approximate the parameters $\theta(\vec{x}, R_k)$ by the points $\theta_m(\vec{x}, R_k)$ that maximize the function $l_{\psi, m}(\theta, \vec{x})$ defined by the formula (34). The sample X_i , $i = 1, \dots, m$ is in the formula (34) is taken from a reference distribution P_{ψ} corresponding to the hard core model with $R = R_k$, with $\widehat{N}(R_k) + 1$ and some non-zero intensities $\psi_0, \dots, \psi_{\widehat{N}(R_k)}$. We took different reference parameters for different pairs (N, R) . For example, for $R = 0.49$ we have $\widehat{N} = 1$ and take $\beta_0 = 5$ and $\beta_1 = 180$ as reference intensities. Concerning this choice of the reference model see Remark 1) at the end of the section.

In all cases we run the Metropolis-Hastings algorithm for $25 \cdot 10^6$ steps sampling at every 10000th step, so 2500 reference samples were simulated. We computed the values of the function (34) for some discrete subset of a box $\Theta = \prod_{i=0}^N [a_i, b_i] \subseteq (\mathbf{R}_+ \setminus \{0\})^{N+1}$, where the values of bounds a_i, b_i , $i = 0, \dots, N$, were taken such that this box certainly contained the point $\theta(\vec{x}, R_k)$. In cases when we got the maximum at the boundary of the cube we took a larger box and repeated the computations. In terms of the parameters β 's we computed the function (34) at points

- 1) $\beta_0 \in \{0.5k, k = 1, \dots, 40\}$, if $\widehat{N} = 0$,
- 2) $\beta_0 \in \{0.5k, k = 1, \dots, 40\}$, $\beta_1 \in \{0.5 \cdot k, k = 1, \dots, 2500\}$, if $\widehat{N} = 1$, in this case it is enough to take $\beta_1 \in \{k, k = 1, \dots, 500\}$ for $R \geq 0.012$,
- 3) $\beta_0 \in \{0.5k, k = 1, \dots, 30\}$, $\beta_1 \in \{0.5k, k = 1, \dots, 5000\}$, $\beta_2 \in \{k, k = 10, \dots, 500\}$, if $\widehat{N} = 2$,
- 4) $\beta_0 \in \{0.5k, k = 1, \dots, 30\}$, $\beta_1 \in \{0.5k, k = 1, \dots, 5000\}$, $\beta_2 \in \{k, k = 10, \dots, 500\}$, $\beta_3 \in \{k, k = 10, \dots, 500\}$, if $\widehat{N} = 3$.

Therefore, as one can see, the points $\theta_m(\vec{x}, R_k)$, $k = 1, \dots, 100$ are themselves approximated by some points from these discrete sets. Certainly, the finer grid approximation can be taken to get better approximation. We decided on these sizes of the mesh, compensating it by taking rather fine discrete approximation for the values of the interaction radius.

To find (= to approximate) the point of the global maximum we need to compare the values of the function (31) at all points (approximations to them) of local (profile) maximums. We do it by computing approximations (34) to the function (33) and at this stage the common reference model has to be taken. The Poisson model with intensity 20 has been taken as the reference one. The parameter 20 has been taken since we wanted to have the reference samples with the number of points comparable with $t(\vec{x}) = 10$ and to avoid high density samples (to avoid "loosing" some of reference samples, see Remark 1) later on).

We computed $m = 3000$ samples from the reference model drawing a sample every 10000 steps in a long run of Metropolis-Hastings algorithm. They were used to compute $l_{\psi, m}(\theta(R_k), R_k)$, $k = 1, \dots, 100$. We plot these values in Figure 14. This function has clear jumps at those values of the radius where the statistics \widehat{N} and t'_k 's change their values. For the reader convenience we show the table below that contains the values of R and $\beta(R)$ such that $\theta(R) = \log(\beta(R))$ maximize the function $l(\theta, R)$. For convenience we included in the table the values of sufficient statistics. Also the table contains the same information for those values of R , where the statistics values change, these rows are labeled by a star.

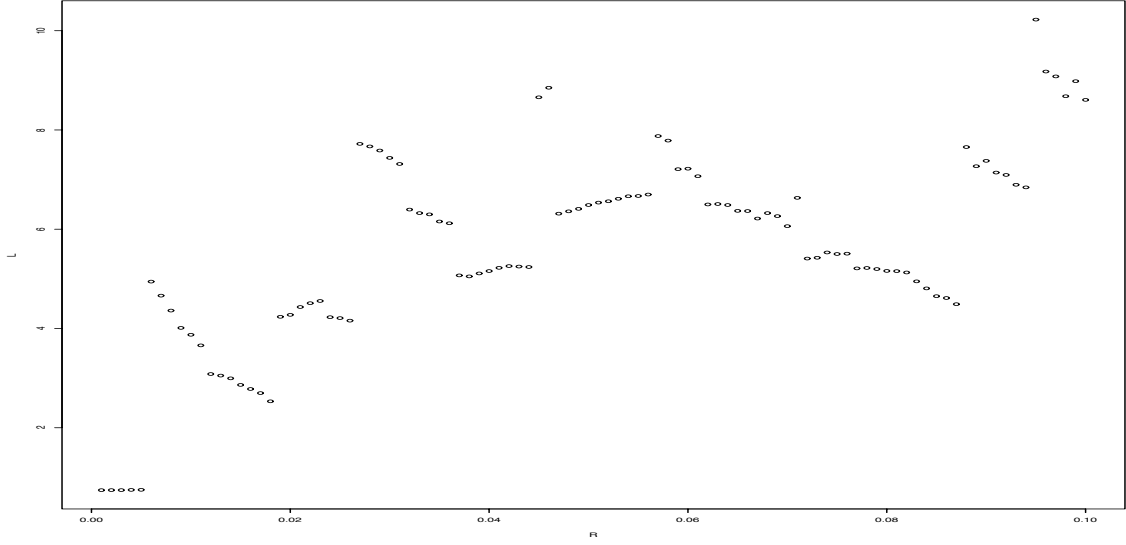


Figure 14: Plot of $L(R_k) = l_{\psi, m}(\theta(R_k), R_k)$ for $k = 1, \dots, 100$.

R	statistics $t_k, k = 0, \dots$	$\hat{N} + 1$	β	$l_m(\theta, R)$
0.005	$t_0 = 10$	1	$\beta_0 = 14$	0.750
*0.006	$(t_0, t_1) = (9, 1)$	2	$(\beta_0, \beta_1) = (11, 2010)$	4.944
*0.019	$(t_0, t_1) = (8, 2)$	2	$(\beta_0, \beta_1) = (8.5, 441)$	4.235
*0.022	$(t_0, t_1) = (8, 2)$	2	$(\beta_0, \beta_1) = (9, 323)$	4.511
*0.027	$(t_0, t_1) = (7, 3)$	2	$(\beta_0, \beta_1) = (6.5, 367)$	7.720
0.046	$(t_0, t_1) = (6, 4)$	2	$(\beta_0, \beta_1) = (5, 187)$	8.851
*0.088	$(t_0, t_1) = (5, 5)$	2	$(\beta_0, \beta_1) = (3.5, 83.5)$	7.655
*0.09	$(t_0, t_1, t_2) = (5, 3, 2)$	3	$(\beta_0, \beta_1, \beta_2) = (4, 35, 130)$	7.378
0.095	$(t_0, t_1, t_2) = (4, 4, 2)$	3	$(\beta_0, \beta_1, \beta_2) = (2.5, 50.5, 105)$	10.224
*0.096	$(t_0, t_1, t_2, t_3) = (4, 4, 1, 1)$	4	$(\beta_0, \beta_1, \beta_2, \beta_3) = (2.5, 51, 40, 158)$	9.177

The maximal value is when $R = 0.095$ and $\hat{N} = 2$, the intensities estimators are $\beta_0 = 2.5, \beta_1 = 50.5$ and $\beta_2 = 105$.

Variations of errors. To estimate the variances of errors $\hat{\theta}_m - \hat{\theta}$ we have drawn 50000 samples from the model with $R = 0.095$ and $\beta_0 = 3, \beta_1 = 50, \beta_2 = 100$, 250 batches of length 200 were formed. The variances of errors are (0.0328, 0.0207, 0.0386).

Finally, we estimated the mean values of canonical statistics for the model with $R = 0.095$ and $\beta_0 = 2.5, \beta_1 = 50.5, \beta_2 = 105$. For that we simulated 1000 samples from the model. The sample means of statistics (t_0, t_1, t_2) are (3.937, 3.975, 2.006), so the observed values (4, 4, 2) practically coincide with the sample means. The sample variances of (t_0, t_1, t_2) are (1.7706, 2.56171, 1.9063), therefore any reasonable confidence intervals cover the observed values of statistics (t_0, t_1, t_2) .

This suggests that the hard core model with parameters $R = 0.095$ and $\beta_0 = 2.5, \beta_1 = 50.5, \beta_2 = 105$ fits the data very well.

Remarks. 1) In general, it is possible to find and choose the common reference model for the whole range of hard core models that we deal with in this particular situation. It suffices to require

the reference density support be larger, than the support of a density of any of these models ([16], p. 142). It is easy to see that the hard core model with the minimal radius and the maximal number of non-zero intensities can be taken as a common reference model. Certainly, any Poisson density can be taken as a reference one for any hard core density h_θ since its support coincides with the whole space of finite configurations. Assume that the support of the reference density h_ψ is larger than the support of the density h_θ . It yields that when computing the function (34) we have to discard all reference samples X_i such that $h_\theta(X_i) = 0, h_\psi(X_i) \neq 0$. By the ergodic theorem the more samples we use the better approximation we have. Therefore we have to be sure that sufficiently many reference samples are used for the computation. It is not convenient to check it in practice. It is much more convenient to use the reference model with the same radius and the same number of non-zero intensities. In this case all reference samples are in "use" and the approximation routine is very robust. This is why we take the hard core model with $R = R_k$ and $N = \hat{N}(R_k)$ while approximating the maximizers $\theta(\vec{x}, R_k)$ of the cross-section $R = R_k$ of the likelihood surface. In this case we "do not loose" any reference sample.

2) We have tested MCMCMLE method on several artificially simulated data sets. The data were sampled from the CSA models with one, two and three non-zero intensities with fixed interaction radius. The computations show that the method works in these cases reasonably well.

3) If the true value of the interaction radius is in the range $[0.006, 0.088]$, then we have to stick with the CSA model with two non-zero intensities. We have found that in this case the estimator $\hat{\beta}_0(R)$ is stable as we vary the interaction radius in such a way that does not change the values of the canonical statistics. The value of the estimator $\hat{\beta}_1(R)$ increases as the value of the interaction radius decreases and vice versa. This effect can be explained as follows. Indeed, if we decrease the interaction radius, then the random area where the daughter points can be generated is shrinking. So, in order to obtain the patterns with "the same statistics" we need to increase the intensity β_1 . And if the radius is increased, then the area covered by the balls centered at the points with zero number of neighbours also increases on average, therefore the intensity β_1 can be reduced and we still get the "same patterns".

4) It has been noted above that the statistics \hat{N} underestimates the parameter N . We can prove (but do not stop on this here) that the probability $P\{\hat{N} = N\}$ is bounded from below by some constant that does not depend on the size and the shape of the observation window.

6 Possible generalizations of CSA processes

Following the standard recipes of the theory of Markov point processes ([13]) a variety of sequential Markov processes can be proposed.

Inhomogeneous CSA processes. The first obvious generalization of CSA process can be obtained if we consider the interaction radius and the intensities as functions depending on a point. Namely, fix a positive bounded function $R : D \rightarrow \mathbf{R}_+$ and name a point $y \in D$ a neighbour of point x if the following condition holds

$$x \sim y \Leftrightarrow \|x - y\| \leq R(x).$$

Then, let $f(x, \vec{x}) : D \times F_D \rightarrow \mathbf{R}_+$ be a non-negative function, such that $f(x, \emptyset) > 0$. Define the intensity of adsorbing a particle at point x given a configuration \vec{x} as follows $\beta(x, \vec{x}) = f(x, \vec{x} \cap$

$B(x, R(x))$). A particular model when $R > \text{diam}(D)$ and

$$f(x, \vec{x}) = \sum_{k=1}^{n(\vec{x})} \|x - x_k\|^a,$$

where $a \in \mathbf{R}$, was proposed in [2].

CSA point process with random radius of interaction. It is also possible to consider the situation when the radius of interaction is a *random* variable. Let ξ_k , $k \in \mathbf{Z}_+$ be independent random variables. Denote by μ_k the probability distribution on \mathbf{R}_+ corresponding to the random variable ξ_k .

The CSA-process in this case is a sequential marked point process with positions in D and marks in \mathbf{R}_+ . It is a sequential marked Markov process ([11]) with respect to the following non-symmetric relation

$$(x, \xi) \sim (y, \eta) \Leftrightarrow \|x - y\| \leq \xi. \quad (39)$$

Let us directly specify the probability density of the process on $\cup_{n=0}^{\infty} (D \times \mathbf{R}_+)^n$ with respect to the sequential Poisson point process on D with independent marks ξ_k , $k = 1, \dots$

$$f((x_1, \xi_1), \dots, (x_n, \xi_n)) = Z_{\beta, \xi}^{-1} \prod_{k=1}^n \beta_{n((x_k, \xi_k), x_{<k})}(x_k) d\mu_k(\xi_k),$$

where $n((x_k, \xi_k), x_{<k})$ is the number of points y in the configuration $x_{<k}$ such that $\|x_k - y\| \leq \xi_k$.

CSA point process with two types of particles (points) with interaction between the types can be considered as a model for spatial conflict between two types of particles adsorbed in the region D . Let X_1, X_2 be two independent *CSA*-point processes conditioned on the event that distance between any two points of different types is not less than r . The *bivariate* sequential process (X_1, X_2) is absolutely continuous with respect to the product $\nu \times \nu$ on the space $(F_D \times F_D, \mathcal{F}_D \times \mathcal{F}_D)$ with density

$$f(\vec{x}, \vec{y}) = Z^{-1} \left(\prod_{k=1}^{n(\vec{x})} \beta_{n(x_k, x_{<k})}(x_k) \prod_{j=1}^{n(\vec{y})} \gamma_{n(y_j, y_{<j})}(y_j) \right) 1_{\{d(\vec{x}, \vec{y}) > r\}},$$

where $\{\beta_k(u), u \in D, k \geq 0\}$ and $\{\gamma_k(u), u \in D, k \geq 0\}$ are two sets of intensity functions. The process (X_1, X_2) is a sequential *bivariate* Markov point process with respect to the relation

$$(x, l) \sim (y, k) \Leftrightarrow \{k \neq l \text{ and } d(x, y) \leq r\} \text{ or } \{k = l \text{ and } d(x, y) \leq R\}$$

CSA point process with area dependent cooperative effects. Let us define another CSA process where the intensity of getting a new point depends on a current configuration in a different way. Fix a positive number $r < R$, where R is the radius of interaction. Given a configuration $\vec{x} = (x_1, \dots, x_n)$ consider for any point $z \in D$ a set

$$V(z, \vec{x}) = \bigcup_{z \sim x_i} B(x_i, r),$$

the area covered by balls with radius r and centered at the neighbours of z . Define the intensity of birth at point z as $\beta(|V(z, \vec{x})|)$, where $\beta : \mathbf{R}_+ \rightarrow \mathbf{R}_+$ is some positive function.

The process density with respect to the reference measure ν is

$$f(x_1, \dots, x_n) = Z_{\beta, R, r}^{-1} \prod_{k=1}^n \beta(|V(x_k, \vec{x}_{<k})|), \quad (40)$$

where $Z_{\beta, R, r}$ is a normalising constant (partition function)

$$Z_{\beta, R, r} = \sum_{n=0}^{\infty} \frac{e^{-|D|}}{n!} \int_{D^n} \prod_{k=1}^n \beta(|V(x_k, \vec{x}_{<k})|) dx_1 \dots dx_n.$$

If $\beta(u)$ is a bounded function, then $Z_{\beta, R, r} < \infty$ and the density is well defined. If $\beta(u)$ increases too fast as u increases, then the density might become non-integrable.

7 Appendix. Positive definiteness of the Fisher information matrix for the hard core model

We are going to show that for any set of positive numbers $\theta = (\theta_0, \dots, \theta_N)$ the Fisher information matrix $\Phi(\theta)$ is positive definite. Since the model forms the exponential family the matrix $\Phi(\theta)$ is a covariance matrix of the vector of sufficient statistics, so we immediately get that $\Phi(\theta)$ is nonnegative definite, i.e.

$$z^T \Phi(\theta) z \geq 0,$$

for any vector z . We have to show that for any non-zero vector z

$$z^T \Phi(\theta) z > 0.$$

Assume that there exists a vector $z = (z_0, \dots, z_N) \in \mathbf{R}^{N+1}$ such that $z \neq 0$ and

$$z^T \Phi(\theta) z = \mathbf{E}_{\theta} \left(\sum_{k=0}^N z_k (t_k(X) - \mathbf{E}_{\theta} t_k(X)) \right)^2 = 0.$$

Then it implies that with probability one (with respect to the measure \mathbf{P}_{θ})

$$\sum_{k=0}^N z_k t_k(X) = \sum_{k=0}^N z_k \mathbf{E}_{\theta} t_k(X).$$

In particular, it means that if we fix values of N of them, then the statistic left would take just one possible value with probability one. Let us show that it is not so. Assume for definiteness that the dimension $d = 2$, for higher dimensions the reasonings below are the same. Consider an event A such that $t_0(X) = \dots = t_{N-1}(X) = 1$. It is easy to see that \mathbf{P}_{θ} -probability of this event is positive. If the random variables t_k , $k = 0, \dots, N$ are linearly dependent, then conditioned on this event the statistics $t_N(X)$ can take just one value with probability 1. But it cannot be so. Indeed, let the first N points x_i , $i = 1, \dots, N$ be in the ball of sufficiently small radius ε centered at a point $y \in D$. Let B be the ball of radius R centered in the same point y (we assume that $B(y, R) \subset D$,

the reasonings can be easily modified for the case when R is sufficiently large and a ball of radius R cannot be inserted in D). Locate 5 points in the ball B such that the distances between any two of them is not less than R and distance from each of them to the center of B is $R - 3/2\varepsilon$. Draw balls of radius ε around the points. Consider an event $A' \subset A$ such that each of these balls contains at most one point. It is easy to see that A' is formed by the events of positive probability that do not intersect with each other, such that $t_0(X) = \dots = t_{N-1}(X) = 1$ and $t_N(X)$ can take values 0, 1, 2, 3, 4, 5. Therefore we get the result.

Acknowledgements

The author would like to thank M.N.M. van Lieshout for helpful discussions and valuable comments on the present paper, A.G.Steenbeek for programming assistance and R.Huele for the real life data.

References

- [1] M.Abramowitz and I.Stegun (1972) Handbook of mathematical functions with formulas, graphs and mathematical tables. New York, Dover.
- [2] S.K.Bhaskar, A.Rosenfeld and A.Wu (1989) Models for neighbour dependency in planar point patterns. Pattern Recognition, v.22, N5, 533-559.
- [3] L.Comtet (1974) Advanced combinatorics. The art of finite and infinite expansions. Dordrecht, Netherlands, Reidel.
- [4] J.W.Evans (1993) Random and cooperative sequential adsorption. Reviews of modern physics, v.65, N4, 1281-1329.
- [5] C.Geyer and E.Thompson(1992) Constrained Monte Carlo likelihood for dependent data. J.R.Statist.Soc.B, 54, N3, pp.657-699.
- [6] C.Geyer (1994) On the convergence of Monte Carlo maximum likelihood calculations. J.R.Statist.Soc.B, 56, N1, pp.261-274.
- [7] C.Geyer (1999) Likelihood inference for spatial point processes. In: O.Barndorff-Nielsen, W.S. Kendall and M.N.M. van Lieshout (Eds.), Stochastic geometry, likelihood and computation, Chapman and Hall/CRC, Boca Raton, 79-140.
- [8] P.Grabarnik and A.Sarkka (2001) Interacting neighbour point processes: some models for clustering. J. Statist. Comput. Simul. v.68, 103-125.
- [9] C.Geyer and J.Moller (1994) Simulation and likelihood inference for spatial point processes. Scandinavian Journal of Statistics, 21, pp.359-373.
- [10] F.P.Kelly and B.D.Ripley (1976) A note on Strauss model for clustering. Biometrika, 63, 357-360.
- [11] M.N.M. van Lieshout (2005) Markovianity in space and time. IMS Lecture Notes-Monograph Series, vol. 1-14

- [12] M.N.M. van Lieshout (2005) Maximum likelihood estimation for random sequential adsorption. Research Report PNA-R0503, CWI, Amsterdam.
- [13] M.N.M. van Lieshout (2000) Markov point processes and their applications. Imperial College Press.
- [14] M.Menshikov, A.Revjakin, A.Kopylova, Y.Makarov and B.Stechkin (1982) Kombinatornyi analiz [Combinatorial analysis]: zadachi i uprazhneniya [problems and exercises] Eds. K.Rybnikova, Moscow, Nauka (in Russian).
- [15] M.Meketon and B.Schmeiser (1984) Overlapping batch means: something for nothing? In Proceedings of the 1984 Winter Simulation Conference, pp.227-230.
- [16] J.Moller and R.Waagepetersen (2004) Statistical inference and simulation for spatial point processes. Chapman&Hall.
- [17] Nonequilibrium statistical mechanics in one dimension. Edited by V.Privman. Cambridge University Press, Cambridge 1997, 470 pp.
- [18] V.Privman, ed. (2000) A special issue of Colloids and Surfaces A, 165.
- [19] G.Roberts and J.Rosenthal (1997) Geometric ergodicity and hybrid Markov chains. Elect. Comm. in Probab. 2, pp.13-25.