MCMC simulation of Markov point processes

Jenny Andersson

28 februari 2005

1 Introduction

Spatial point process are often hard to simulate directly with the exception of the Poisson process. One example of a wide class of spatial point processes is the class of Markov point processes. These are mainly used to model repulsion between points but attraction is also possible. Limitations of the use of Markov point processes is that they can not model to strong regularity or clustering. The Markov point processes, or Gibbs processes, were first used in statistical physics and later in spatial statistics. In general they have a density with respect to the Poisson process. One problem is that the density is only specified up to a normalisation constant which is usually very difficult to compute. We will consider using MCMC to simulate such processes and in particular a Metropolis–Hastings algorithm. In Section 2 spatial point processes in general will be discussed and in particular the Markov point processes. Some properties of uncountable state space Markov chains are shortly stated in Section 3. In Section 4 we will give the Metropolis–Hastings algorithm and some of its properties. Virtually all the material in this paper can be found in [1] with exception of the example at the end of Section 4. A few additions on Markov point processes are taken from [2].

2 Markov point processes

In this section we will define and give examples of Markov point processes. To make the notation clear, a brief description of spatial point processes, in particular the Poisson point process, will proceed the definition of Markov point processes.

A spatial point process is a countable subset of some space S, which here is assumed to be a subset of $\mathbb{R}^n$. Usually we demand that the point process is locally finite, that is the number of points in bounded sets are finite, and that the points are distinct. A more formal description can be given as follows. Let N_{lf} be the family of all locally finite point configurations. If $x \in N_{lf}$ and $B \subseteq S$ then let $x(B) = x \cap B$ and let $n(x)$ count the number of points in x. Let $\mathcal{N}_{lf}$ be the smallest σ–algebra such that for $x \in N_{lf}$, all mappings from x to $x(B)$ are measurable for all Borel sets B. Then a point process X is a measurable mapping from a probability space $(S, \mathcal{F}, \mathbb{P})$ to $(N_{lf}, \mathcal{N}_{lf})$. For technical reasons, we will use

$$N_f = \{x \subseteq S : n(x) < \infty\},$$

the set of all finite point sequences in S, in place of N_{lf} when discussing Markov point processes. A point configuration will typically be denoted by $x = \{x_1, \ldots, x_n\}$ while points will be denoted ξ or η.

The most important spatial point process is the Poisson process. It is often described to be "completely random" or as having "no interaction" since the number of points in disjoint sets are independent. A homogeneous Poisson process is characterised by the fact that the number of points in a bounded Borel set B is Poisson distributed with expectation $\lambda|B|$ for some constant $\lambda > 0$, where $|\cdot|$ is the Lebesgue measure. It is possible to show that X is a stationary Poisson point process on S with intensity λ if and only if for all $B \subseteq S$ and all $F \subseteq N_{lf}$,

$$\mathbb{P}(X(B) \in F) = \sum_{n=0}^{\infty} \frac{e^{-\lambda|B|}}{n!} \int_B \ldots \int_B 1_{[\{x_1,\ldots,x_n\}\in F]} \lambda^n \mathrm{d}x_1 \cdots \mathrm{d}x_n. \tag{2.1}$$

For an inhomogeneous Poisson process, the number of points in B is instead Poisson distributed with expectation $\Lambda(B)$, where Λ is an intensity measure. Often we think of the case when it has a density with respect to the Lebesgue measure,

$$\Lambda(B) = \int_B \lambda(\xi)\mathrm{d}\xi,$$

where $\lambda(\xi)$ is the intensity of points at ξ. In this way we can get different densities of points in different areas, but still the number of points in disjoint sets are independent. There is no interaction or dependence between them at all.

One way to introduce a dependence between the points and not only dependence on the location is by a Markov point process. Such a process has a certain density with respect to the Poisson process with some extra conditions added that give some sort of spatial Markov property. For a point process X on $S \subseteq \mathbb{R}^d$ with density f with respect to the standard Poisson process, i.e. the Poisson process with constant intensity 1 on S, we have for $F \subseteq S$

$$\mathbb{P}(X \in F) = \sum_{n=0}^{\infty} \frac{e^{-|S|}}{n!} \int_S \ldots \int_S 1_{[\{x_1,\ldots,x_n\}\in F]} f(\{x_1, \ldots, x_n\}) \mathrm{d}x_1 \cdots \mathrm{d}x_n. \tag{2.2}$$

from (2.1). In general, the density f may only be specified as proportional to some known function, that is the normalising constant is unknown.

On the way to the definition of Markov point processes we first define a neighbourhood of a point $\xi \in S$ and then let the only influence on the point be from its neighbours. Take a reflexive and symmetric relation[1], $\sim$, on S and define the neighbourhood of ξ as $H_\xi = \{\eta \in S : \eta \sim \xi\}$. A common neighbourhood is the R–close–neighbourhood which is specified by a ball with radius R. We are now ready for the definition of a Markov point process.

Definition 2.1 *Let $h : N_f \to [0, \infty)$ be a function such that $h(x) > 0 \Rightarrow h(y) > 0$ for $y \subseteq x$. If for all $x \in N_f$ with $h(x) > 0$ and all $\xi \in S \setminus x$ we have*

$$\frac{h(x \cup \xi)}{h(x)}$$

[1]Reflexive means $\xi \sim \xi$ and symmetric means $\xi \sim \eta \Rightarrow \eta \sim \xi$ for all $\xi, \eta \in S$.

only depending on x through $x \cap H_\xi$, then h is called a Markov function *and a point process with density h with respect to the Poisson process with intensity 1 is called a* Markov point process.

We can interpret $h(x \cup \xi)/h(x)$ as a conditional intensity. Observe that $h(x \cup \xi)/h(x)\mathrm{d}\xi$ is the conditional probability of seeing a point ξ in an infinitesimal region of size $\mathrm{d}\xi$ given that the rest of the process is x. If this conditional intensity only depends on what happens in the neighbourhood of ξ it resembles the usual Markov property and therefore it is called the *local Markov property*.

The following theorem characterises the Markov functions.

Theorem 2.2 *A function $h : N_f \to [0, \infty)$ is a Markov function if and only if there is a function $\phi : N_f \to [0, \infty)$, with the property that $\phi(x) = 1$ whenever there are $\xi, \eta \in x$ with $\xi \nsim \eta$, such that*

$$h(x) = \prod_{y \subseteq x} \phi(y), \quad x \in N_f.$$

The function ϕ is called an interaction function.

A natural question to ask is whether a Markov function is integrable with respect to the Poisson process. To answer we make the following definition.

Definition 2.3 *Suppose $\phi^* : S \to [0, \infty)$ is a function such that $c^* = \int_S \phi^*(\xi)\mathrm{d}\xi$ is finite. A function h is* locally stable *if*

$$h(x \cup \xi) \leq \phi^*(\xi)h(x)$$

for all $x \in N_f$ and $\xi \in S \setminus x$.

Local stability implies integrability of h with respect to the Poisson process with intensity 1 and that

$$h(x) > 0 \Rightarrow h(y) > 0, \quad y \subseteq x.$$

Many Markov point processes are locally stable and it will be an important property when it comes to simulations.

The simplest examples of Markov point processes are the pairwise interaction processes. Their density is proportional to

$$h(x) = \prod_{\xi \in x} \phi(\xi) \prod_{\{\xi,\eta\} \subseteq x} \phi(\{\xi, \eta\}),$$

where ϕ is an interaction function. If $\phi(\xi)$ is the intensity function and $\phi(\{\xi, \eta\}) = 1$ we get the Poisson point process. It is possible to show that h is repulsive if and only if $\phi(\{\xi, \eta\}) \leq 1$, and then the process is locally stable. The attractive case $\phi(\{\xi, \eta\}) \geq 1$ is not in general well defined.

Example 2.1. The simplest example of a pairwise interaction process in turn, is the Strauss process where

$$\phi(\{\xi, \eta\}) = \gamma^{1_{\{\|\xi,\eta\| \leq R\}}}, \quad 0 \leq \gamma \leq 1$$

and $\phi(\xi)$ is a constant. The Markov function can be written as

$$h(x) = \beta^{n(x)} \gamma^{S_R(x)}, \tag{2.3}$$

with

$$S_R(x) = \sum_{\{\xi,\eta\}\subseteq x} 1_{\{|||\{\xi,\eta\}|||\leq R\}}, \tag{2.4}$$

the number of R–close neighbours. The condition $0 \leq \gamma \leq 1$ is to ensure local stability. It also means that points repel each other. □

The Markov point processes can be extended to be infinite, or restricted by conditioning on the number of points in S. Neither of these cases will be covered here.

3 Markov chains on uncountable state space

The Markov chains used in the MCMC algorithm have uncountable state spaces. The theory of such Markov chains is similar to that of countable state space Markov chains, but there are some differences in notation. Some definitions and a central limit theorem are briefly listed below.

Consider a discrete time Markov chain $\{Y_n\}_0^\infty$ with uncountable state space E. Define the *transition kernel* as

$$P(x, F) = \mathbb{P}(Y_{m+1} \in F | Y_m = x).$$

The *m–step transition probability* is $P^m(x, F) = \mathbb{P}(Y_m \in F | Y_0 = x)$. The chain is said to be *reversible* with respect to a distribution Π on E if for all $F, G \in E$

$$\mathbb{P}(Y_m \in F, Y_{m+1} \in G, Y_m \neq, Y_{m+1}) = \mathbb{P}(Y_m \in G, Y_{m+1} \in F, Y_m \neq, Y_{m+1})$$

when $Y_m \sim \Pi$, meaning that Y_m and Y_{m+1} have the same distribution if Y_m has the stationary distribution. The chain is *irreducible* if it has positive probability of reaching F from x, for all $x \in E$ and $F \subseteq E$ such that $\mu(F) > 0$ for some measure μ on E. Furthermore the chain is *Harris reccurent* if it is irreducible for some μ and $\mathbb{P}(Y_m \in F \text{ for some } m | Y_0 = x) = 1$ for all $x \in E$ and $F \subseteq E$ such that $\mu(F) > 0$. As usual, if a stationary distribution Π exists, irreducibility implies that Π is the unique stationary distribution. The chain is *ergodic* if it is Harris recurrent and aperiodic.

For functions $V : E \to [1, \infty)$ with $\mathbb{E}[V(Y)] < \infty$ if $Y \sim \Pi$, define the *V–norm*

$$||P^m(x, \cdot) - \Pi||_V = \frac{1}{2} \sup_{|k|\leq V} |\mathbb{E}[k(Y_m)|Y_0 = x] - \Pi(k)|.$$

If $V = 1$ this is the total variation norm. The Markov chain is *geometrically ergodic* if it is Harris reccurent with stationary distribution Π and there exist a constant $r > 1$ such that for all $x \in E$ and some V,

$$\sum_{m=1}^{\infty} r^m ||P^m(x, \cdot) - \Pi||_V < \infty.$$

This means that the chain is ergodic and that its rate of convergence to Π is geometric. A special case of geometric ergodicity is *V–uniform ergodicity* which can be expressed as

$$\lim_{m\to\infty} \sup_{x\in E} ||P^m(x, \cdot) - \Pi||_V = 0.$$

A further special case, when $V = 1$, i.e. replace the V–norm with the total variation norm, is called *uniform ergodicity*.

Let Π be a stationary distribution. Then for a real function k and Y distributed according to Π with $\mathbb{E}|k(Y)| < \infty$, let

$$\Pi(k) = \mathbb{E}[k(Y)] \text{ and } \overline{k}_n = \frac{1}{n}\sum_{m=0}^{n-1} k(Y_m).$$

Theorem 3.1 *Let $Y_0, Y_1, \ldots$ be a Markov chain with uncountable state space E. Let k be a function $k : S \to \mathbb{R}$ such that either the chain is reversible and $\mathbb{E}|k(X)|^2 < \infty$ or the chain is V–uniformly ergodic and $k^2 \leq V$. With the assumption that $Y_0 \sim \Pi$, define*

$$\overline{\sigma}^2 = Var(k(Y_0)) + 2\sum_{m=1}^{\infty} Cov(k(Y_0), k(Y_m)).$$

Then

$$\sqrt{n}(\overline{k}_n - \Pi(k)) \text{ converges in distribution to } N(0, \overline{\sigma}^2) \text{ as } n \to \infty.$$

4 Simulation, Metropolis–Hastings algorithms

In this section the so called *Birth–death–move Metropolis Hastings* algorithm will be described. In each step of the algorithm an attempt is made either to add a point, to delete a point or to move a point. The setup is as follows. We want to have a realisation of a point process, X on S with unnormalised density h with respect to the homogeneous Poisson process with unit intensity. Instead of sampling directly from the distribution itself we will use MCMC. We will generate a Markov chain $Y_0, Y_1, \ldots$ whose distribution tends to that of the point process X. First the algorithm to use will be discussed and then what properties are desirable for guaranteeing convergence.

Let $x = \{x_1, x_2, \ldots, x_n\}$ denote a set of points in S. First, we can make either a move step or a birth–death step. Specify $0 \leq q < 1$ as the probability of making a move step. Also for each x, with $n(x) = n$, and each $i \in \{1, \ldots, n\}$ let $q_i(x, \cdot)$ be a given density on S, called the proposal density, i.e. a density for the location of a new point given that x_i is proposed to be replaced. Define the *Hastings ratio* as

$$r_i(x, \xi) = \frac{h((x \setminus x_i) \cup \xi) q_i((x \setminus x_i) \cup \xi, x_i)}{h(x) q_i(x, \xi)}$$

If we are to do a replacement a point is selected uniformly at random among the points in x and a proposal location ξ is chosen according to q_i. The move is accepted with probability

$$\alpha_i(x, \xi) = \min\{1, r_i(x, \xi)\}.$$

With probability $1 - q$ an attempt is made either to add a new point or remove a current point. Let $p(x)$ be a given probability for giving birth to a point if x is the state of the chain. The location of a proposed new point is given by the density function $q_b(x, \cdot)$ on S. With probability $1 - p(x)$ an attempt will be made to remove a point unless $x = \emptyset$ in which case nothing is done. The point to remove is given by the discrete density $q_d(x, \cdot)$ on x. The acceptance probability for the birth of the point ξ is

$$\alpha_b(x, \xi) = \min\{1, r_b(x, \xi)\},$$

with Hastings ratio

$$r_b(x,\xi) = \frac{h(x\cup\xi)(1-p(x\cup\xi))q_d(x\cup\xi,\xi)}{h(x)p(x)q_b(x,\xi)}.$$

The acceptance probability for the death of a point η is

$$\alpha_d(x,\eta) = \min\{1, r_d(x,\eta)\},$$

with Hastings ratio

$$r_d(x,\eta) = \frac{h(x\setminus\eta)(p(x\setminus\eta))q_b(x\setminus\eta,\eta)}{h(x)(1-p(x))q_d(x,\eta)}.$$

Now to the algorithm in more detail. It generates a Markov chain $Y_0, Y_1, \ldots$, where Y_0 is assumed to be given. For example $Y_0 = \emptyset$ or Y_0 could be a realisation of a Poisson process or chosen so that $h(Y_0) > 0$.

Algorithm 4.1 *Let $0 \le q < 1$ and suppose $Y_m = x = \{x_1, \ldots, x_n\}$ is given. Generate $R_m \sim U[0,1]$.*

- *If $R_m \le q$, make a move step as follows. Take $R'_m \sim U[0,1]$ and $I_m \sim U(\{1,\ldots,n\})$. Generate ξ_m according to $q_i(x,\cdot)$ given $I_m = i$. Then let*

$$Y_{m+1} = \begin{cases} (x\setminus x_i)\cup\xi_m & \text{if } R'_m \le r_i(x,\xi_m) \\ x & \text{otherwise} \end{cases}$$

- *If $R'_m > q$, make a birth–death step as follows Take $R''_m \sim U[0,1]$ and $R'''_m \sim U[0,1]$.*
  - *If $R''_m \le p(x)$ then make a birth step by taking ξ_m according to $q_b(x,\cdot)$. Then let*

$$Y_{m+1} = \begin{cases} x\cup\xi_m & \text{if } R'''_m \le r_b(x,\xi_m) \\ x & \text{otherwise.} \end{cases}$$

  - *If $R''_m > p(x)$, make a death step.*
    - *If $x = \emptyset$*

$$Y_{m+1} = x.$$

    - *Otherwise generate η_m according to $q_d(x,\cdot)$. Then let*

$$Y_{m+1} = \begin{cases} x\setminus\eta_m & \text{if } R'''_m \le r_d(x,\eta_m) \\ x & \text{otherwise.} \end{cases}$$

The random variables $R_m, R'_m, R''_m, R'''_m, I_m, \xi_m$ and η_m are mutually independent and also independent of $Y_0, \ldots Y_m$.

If in the move step $q_i(x,\cdot)$ only depends on x_i and is symmetric we have the Metropolis algorithm. If on the other hand $q_i(x,\cdot)$ depends of everything except x_i we have the Gibbs sampler.

Reversibility with respect to Π and irreducibility ensures that if the chain has a limit distribution it must be Π. Furthermore irreducibility is needed to establish convergence. The only property we show for the Markov chain in the algorithm is the reversibility. The proof is similar as in the case of a Markov chain with countable state space, but the notation is a bit messier.

Proposition 4.2 *The Markov chain generated by Algorithm 4.1 is reversible with respect to h.*

Proof. If P_m is the transition kernel for a Markov chain with only move steps and P_{bd} is the transition kernel of a chain with only birth and death steps, the transition kernel of the Markov chain in the algorithm can be written as

$$P(x, F) = qP_m(x, F) + (1 - q)P_{bd}(x, F). \tag{4.1}$$

If the move chain and the birth–death chain are both reversible with respect to h then (4.1) shows that the birth–death–move chain is reversible. First we show that the birth–death chain is reversible with respect to h. Suppose that Y_0 is such that $h(Y_0) > 0$ and define $E_n = \{x \in N_f : h(x) > 0, \quad n(x) = n\}$[2]. To prove that the birth–death chain is reversible we only need to consider events $F \subseteq E_n$ and $G \subseteq E_{n+1}$, since in each step either a point is added or deleted. We want to show

$$\mathbb{P}(Y_{m+1} \in G, Y_m \in F) = \mathbb{P}(Y_{m+1} \in F, Y_m \in G) \tag{4.2}$$

provided $Y_m \sim h$. Let the unknown normalising constant be denoted c. Then

$$\begin{aligned}&\mathbb{P}(Y_{m+1} \in G, Y_m \in F)\\ &\quad = c \int_B \dots \int_B 1_{\{x \in F\}} \mathbb{P}(Y_{m+1} \in G | Y_m \in F) h(x) \frac{e^{-|S|}}{n!} \mathrm{d}x_1 \dots \mathrm{d}x_n\end{aligned}$$

by (2.2). Since the chain makes a birth step with probability $p(x)$ if it is sitting in x, the density for the new born point is $q_b(x, \xi)$ and the new point is accepted with probability $\alpha_b(x, \xi)$,

$$\begin{aligned}&\mathbb{P}(Y_{m+1} \in G, Y_m \in F)\\ &\quad = c \int_B \dots \int_B \int_B 1_{\{x \in F, x \cup \xi \in G\}} p(x) q_b(x, \xi) \alpha_b(x, \xi) h(x) \frac{e^{-|S|}}{n!} \mathrm{d}\xi \mathrm{d}x_1 \dots \mathrm{d}x_n.\end{aligned}$$

Writing out $\alpha_b(x, \xi)$ this is

$$\begin{aligned}&\mathbb{P}(Y_{m+1} \in G, Y_m \in F)\\ &\quad = c \int_B \dots \int_B \int_B 1_{\{x \in F, x \cup \xi \in G\}} p(x) q_b(x, \xi)\\ &\quad \min\left\{1, \frac{h(x \cup \xi)(1 - p(x \cup \xi)) q_d(x \cup \xi, \xi)}{h(x) p(x) q_b(x, \xi)}\right\} h(x) \frac{e^{-|S|}}{n!} \mathrm{d}\xi \mathrm{d}x_1 \dots \mathrm{d}x_n.\end{aligned} \tag{4.3}$$

Similarly the right hand side of (4.2) can be expressed as

$$\begin{aligned}&\mathbb{P}(Y_{m+1} \in G, Y_m \in F)\\ &\quad = c \int_B \dots \int_B \int_B \sum_{i=1}^{n+1} 1_{\{y \setminus y_i \in F, y \in G\}} (1 - p(y)) q_d(y, y_i)\\ &\quad = \alpha_d(y, y_i) h(y) \frac{e^{-|S|}}{(n+1)!} \mathrm{d}y_1 \dots \mathrm{d}y_{n+1}.\end{aligned}$$

[2] If Y_0 is such that $h(Y_0) > 0$ the state space of the chain is $E = \bigcup_n E_n$, otherwise the chain might eventually end up in E. To avoid technicalities it is assumed that $h(Y_0) > 0$.

If we interchange the order of integration and summation and see that we only get a sum of $n+1$ equal terms,

$$\begin{aligned}&\mathbb{P}(Y_{m+1} \in G, Y_m \in F)\\ &= c \int_B \dots \int_B \int_B 1_{\{y \setminus y_1 \in F, y \in G\}} (1 - p(y)) q_d(y, y_1)\\ &\quad \alpha_d(y, y_1) h(y) \frac{e^{-|S|}}{n!} \mathrm{d}y_1 \dots \mathrm{d}y_{n+1}.\end{aligned}$$

Now let $x = y \setminus y_1$ and $\xi = y_1$,

$$\begin{aligned}&\mathbb{P}(Y_{m+1} \in G, Y_m \in F)\\ &= c \int_B \dots \int_B \int_B 1_{\{x \in F, x \cup \xi \in G\}} (1 - p(x \cup \xi)) q_d(x \cup \xi, \xi)\\ &\alpha_d(x \cup \xi, \xi) h(x \cup \xi) \frac{e^{-|S|}}{n!} \mathrm{d}x_1 \dots \mathrm{d}x_n \mathrm{d}\xi.\end{aligned}$$

But

$$\alpha_d(x \cup \xi, \xi) = \min\left\{1, \frac{h(x) p(x) q_b(x, \xi)}{h(x \cup \xi)(1 - p(x \cup \xi)) q_b(x \cup \xi, \xi)}\right\}$$

which is identical to (4.3) giving (4.2).

For the move chain we can consider two states x and y differing only at one point, say $x_i \neq y_i$. It is easy to see that

$$h(x) \frac{1}{n} q_i(x, y_i) \alpha_i(x, y_i) = h(y) \frac{1}{n} q_i(y, x_i) \alpha_i(y, x_i),$$

showing that the chain is reversible with respect to h. ■

Proposition 4.3 *If Y_0 is such that $h(Y_0) > 0, p(\emptyset) < 1$ and for all x with $h(x) > 0$ and $x \neq \emptyset$ there exists $\eta \in x$ such that*

$$(1 - p(x)) q_d(x, \eta) > 0$$

and

$$h(x \setminus \eta) p(x \setminus \eta) q_b(x \setminus \eta, \eta) > 0$$

then the chain generated by Algorithm 4.1 is irreducible and aperiodic.

Proposition 4.4 *For any $\beta > 1$ let $V(x) = \beta^{n(x)}$. If the Algorithm 4.1 is given by a locally stable version and , then*

(a) the chain is V–uniformly ergodic.

(b) the chain is uniformly ergodic if and only if there exists an n_0 such that $n(x) \leq n_0$ whenever $h(x) > 0$ and the upper bound on the total variation distance to the distribution h with normalisation constant c, is given by

$$\|P^m(x, \cdot) - h(x)/c\| = (1 - ((1 - q) \min\{1 - p, p/c^*\})^{n_0})^{m/n_0}.$$

The upper bound on the total variation distance is of limited use for simulation purposes, moreover it is only hard–core processes, defined as having a minimal interpoint distance, that are uniformly ergodic among the Markov point processes. Nearly all Markov point processes used in practise are locally stable however, which means they are V-uniformly ergodic and then there is the Central Limit Theorem given by Theorem 3.1.

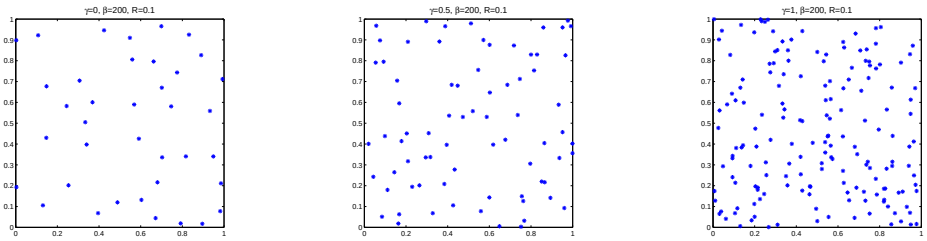
In practise there is a lot of freedom when choosing q, p, q_i, q_d and q_b. The choices might affect the convergence rate and the so called mixing properties of the chain and optimal ones are found by trial and error. A Markov chain is well mixing if the dependence between Y_m and Y_{m+j} dies out quickly. Mixing properties are important for sampling reasons and might be assessed using plots of autocorrelations. Furthermore the initial distribution should be chosen in some way as discussed before.

When has the chain "converged enough´´ so that it is possible to take samples? The Central Limit Theorem is not useful when deciding if the chain is in equilibrium, but more for knowing that Monte Carlo estimates based on the simulations are asymptotically normal. Trace plots are one way of determining if the chain has converged. Plot $g(Y_m)$ for real functions g, often given by some sufficient statistic. If trace plots for two differently started chains does not look similar for some iteration number, n, then at least one of the chains has not reached stationarity. One extreme starting value is the empty set and another, for locally stable processes, is a simulation of a spatial Poisson process. Another way is to exploit the fact that if the chain has converged then reversibility means that $n(Y_{m+1}) - n(Y_m)$ is symmetrically distributed around 0, being either -1,0 or 1. These convergence criteria should be used with some caution since the process might reach some states which are stable in some sense and stay there for a long time although the process is not yet stationary.

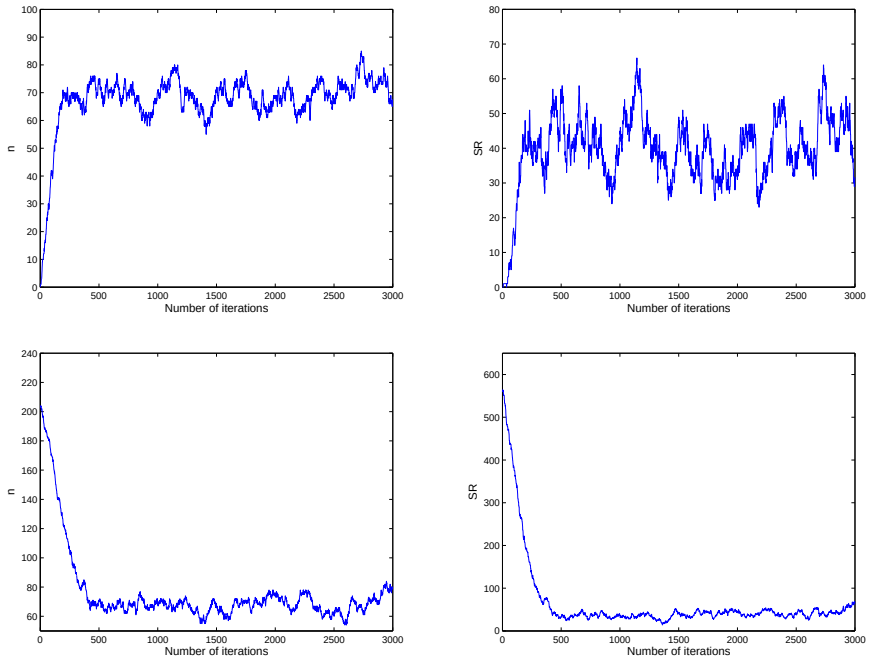
Example 4.2. In Figure 1 there are three realisations of the Strauss process on the unit square. The parameters are $\beta = 200$, $R = 0.1$ and γ is either 0, 0.5 or 1. The plots were taken after 6000 iterations starting in a realisation of a Poisson process with intensity 200. The parameters in the Metropolis Hastings algorithm were as follows. First q was put to 0 since it did not seem to matter for the convergence rate if move steps were made or not. Birth steps and death steps were given equal probability so $p = 0.5$. The distribution of birth proposals was uniform over the square and a death proposal was chosen among the points with equal probability. The chain seemed to converge in about 1000 iterations when γ was 0 or 0.5 but in about 2000 when γ was 1. Trace plots like those in Figure 2 were used to decide when the chain had converged. For the Strauss process the sufficient statistics are the number of points and the number of R–close neighbours. As seen in the plots the statistics seem to settle at the same level when starting in two extreme start configurations after less than 1000 iterations for $\gamma = 0.5$.

When $\gamma = 1$ the Strauss process is a Poisson process with intensity β. When γ decreases towards 0, there is more and more repulsion between the points. At $\gamma = 0$, we have a hard–core process with minimal interpoint distance R. In Figure 1 the leftmost plot has quite regularly spaced points going to more irregular spacings as we look to the right.

□



Figur 1: Simulation of the Strauss model with $\beta = 200$, $R = 0.1$ and $\gamma = 0, 0.5$ and 1 in the unit square.



Figur 2: Convergence diagnostics, above starting with $Y_0 = \emptyset$ and below starting with a realisation of a Poisson process for the Strauss model with $\beta = 200$, $R = 0.1$ and $\gamma = 0.5$. To the left the number of points and to the right the number of R–close neighbours.

Referenser

[1] Möller, J., Waagepetersen, R.P. (2004), *Statistical Inference and Simulation for Spatial Point Processes*, Chapman & Hall/CRC

[2] Stoyan, D., Kendall, S.K., Mecke, J. (1995), *Stochastic Geometry and its Applications.* 2nd edition. Wiley.