Bayesian analysis of Markov point processes

Kasper K. Berthelsen and Jesper Møller

Abstract

Recently Møller, Pettitt, Berthelsen and Reeves [17] introduced a new MCMC methodology for drawing samples from a posterior distribution when the likelihood function is only specified up to a normalising constant. We illustrate the method in the setting of Bayesian inference for Markov point processes; more specifically we consider a likelihood function given by a Strauss point process with priors imposed on the unknown parameters. The method relies on introducing an auxiliary variable specified by a normalised density which approximates the likelihood well. For the Strauss point process we use a partially ordered Markov point process as the auxiliary variable. As the method requires simulation from the "unknown" likelihood, perfect simulation algorithms for spatial point processes become useful.

Keywords: Bayesian inference; Markov chain Monte Carlo; Markov point process; Partially ordered Markov point process; Perfect simulation; Spatial point process; Strauss process.

1 Introduction

Markov point processes [16, 19, 14] are models for point processes with interacting points, and they constitute one of the most important classes of spatial point process models. The basic problem with parametric inference for such point processes is the presence of a normalising constant which cannot be evaluated explicitly, cf. Chapter 9 in [19]. So far most work on parametric inference for Markov point processes have concentrated on parameter estimation based on maximum pseudo likelihood estimation [1, 5, 13] or approximate maximum likelihood estimation using Markov chain Monte Carlo (MCMC) algorithms [9, 10, 18, 19]. Apart from a few papers [3, 12], very little has been done on Bayesian inference for Markov point processes.

In this paper we consider the problem of simulating from a posterior density

$$\pi(\theta|y) \propto \pi(\theta)\pi(y|\theta) \tag{1}$$

when the likelihood

$$\pi(y|\theta) = q_\theta(y)/Z_\theta \tag{2}$$

is given by an unnormalised density $q_\theta(y)$ with an unknown normalising constant (or partition function) Z_θ. By "unknown", we mean that Z_θ is not available analytically and/or that exact computation is not feasible. Indeed this is the case when (2) is a likelihood function for a parametric family of Markov point process models, cf. [16, 19].

For example, consider a Strauss process defined on a region $S \subset \mathbb{R}^2$ of area $|S| \in (0, \infty)$. This has a density

$$\pi(y|\theta) = \frac{1}{Z_\theta}\beta^{n(y)}\gamma^{s_R(y)} \tag{3}$$

with respect to μ, which denotes a homogeneous Poisson point process on S with intensity one. Further, y is a point configuration, i.e. a finite subset of S; $\theta = (\beta, \gamma, R)$, with $\beta > 0$ (known as the chemical activity in statistical physics), $0 < \gamma \leq 1$ (the interaction parameter), and $R > 0$ (the interaction range); $n(y)$ is the cardinality of y; and

$$s_R(y) = \sum_{\{\xi,\eta\}\subseteq y:\xi\neq\eta} \mathbf{1}[\|\eta - \xi\| \leq R]$$

is the number of pairs of points in y within a distance R from each other. Figure 1 shows a realisation y of a Strauss point process, where $s_R(y)$ is given by the number of pairs of overlapping discs with diameter $R/2$ and centred at the points in y. For $\gamma = 1$, we obtain a homogeneous Poisson process on S with intensity β. For $\gamma < 1$, typical realisations look more regular than in the case $\gamma = 1$. This is due to inhibition between the points, and the inhibition gets stronger as γ decreases or R increases. The normalising constant is unknown when $\gamma < 1$, since

$$Z_\theta = \mathrm{e}^{-|S|} + \mathrm{e}^{-|S|}\sum_{n=1}^{\infty}\beta^n \int_S \cdots \int_S \gamma^{s_R(\{y_1,\ldots,y_n\})}\,\mathrm{d}y_1 \cdots \mathrm{d}y_n$$

where the n-fold integrals are unknown, cf. [14].

It is not straightforward to generate samples from (1) by MCMC algorithms: Consider a Metropolis-Hastings algorithm, see e.g. [20]. If θ is the current state of the chain generated by the algorithm, and if a proposal θ' with density

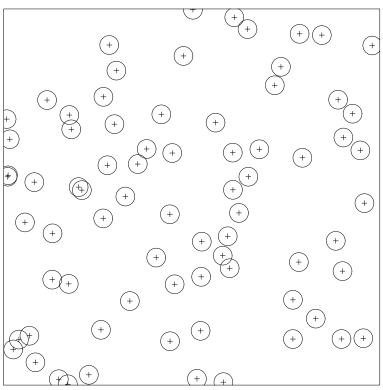

Figure 1: Realisation of a Strauss point process on the unit square, with $(\beta, \gamma, R) = (100, 0.5, 0.05)$, and generated by perfect simulation algorithm (dominated CFTP, see[15]). Circles centred at points have radii 0.025.

$p(\theta'|\theta)$ is generated, then θ' is accepted as the new state with probability $\alpha(\theta'|\theta) = \min\{1, H(\theta'|\theta)\}$, and otherwise we retain θ. Here

$$H(\theta'|\theta) = \frac{\pi(\theta'|y)p(\theta|\theta')}{\pi(\theta|y)p(\theta'|\theta)}$$

is the Hastings ratio. By (2),

$$H(\theta'|\theta) = \frac{\pi(\theta')q_{\theta'}(y)p(\theta|\theta')}{\pi(\theta)q_{\theta}(y)p(\theta'|\theta)} \Big/ \frac{Z_{\theta'}}{Z_{\theta}} \tag{4}$$

is unknown, since it depends on the ratio of unknown normalising constants $Z_{\theta'}/Z_{\theta}$.

Because of their intractability, earlier Bayesian work on Markov point processes attempted to avoid algorithms involving unknown normalising constants. An example in connection to spatial point processes is Heikkinen and Penttinen [12], who instead of estimating the entire posterior distribution, focused on finding the maximum a posteriori estimate for the interaction function in a Bayesian model where the likelihood function is given by a pairwise interaction point processes (like the Strauss process) and its normalising constant is unknown. Recently, Berthelsen and Møller [3] performed a more detailed Bayesian MCMC analysis, using path sampling [8] or, as it is known in statistical physics, thermodynamic integration, for estimating the ratio of normalising constants.

Section 2 considers the approach introduced by Møller et al. [17] which avoids approximations of (ratios of) normalising constants such as those discussed above. Their approach consists in introducing an auxiliary variable x into a Metropolis-Hastings algorithm for (θ, x) so that ratios of normalising constants no longer appear but the posterior distribution for θ is retained. Access to algorithms for making perfect (or exact) simulations [2, 11, 15, 19] from (2) is an important ingredient as explained later. Section 3 applies this approach to a Bayesian analysis of a Strauss process. This section has earlier been published as a part of the research report [17]. Finally, Section 4 contains some concluding remarks.

2 Auxiliary variable method

Consider the general setting (1) when Z_θ in (2) is unknown. The method described in this section applies for Markov point process models as well as many other statistical models with an unknown normalising constant, cf. [17].

We introduce an auxiliary variable x defined on the same space as the state space of y. Assume that x has a normalised conditional density $f(x|\theta, y)$, so that the joint density of (θ, x, y) is given by

$$\pi(\theta, x, y) = f(x|\theta, y)\pi(y|\theta)\pi(\theta).$$

The posterior density with $\pi(y|\theta)$ given by (2),

$$\pi(\theta, x|y) \propto f(x|\theta, y)\pi(\theta)q_\theta(y)/Z_\theta$$

still involves the unknown Z_θ.

A Metropolis-Hastings algorithm for drawing from $\pi(\theta, x|y)$ has a Hasting ratio given by

$$\begin{aligned} H(\theta', x'|\theta, x) &= \frac{\pi(\theta', x'|y)p(\theta, x|\theta', x')}{\pi(\theta, x|y)p(\theta', x'|\theta, x)} \\ &= \frac{f(x'|\theta', y)\pi(\theta')q_{\theta'}(y)p(\theta, x|\theta', x')}{f(x|\theta, y)\pi(\theta)q_\theta(y)p(\theta', x'|\theta, x)} \bigg/ \frac{Z_{\theta'}}{Z_\theta} \end{aligned}$$

where $p(\theta', x'|\theta, x)$ is the proposal density for (θ', x'). The proposal density can be factorised as

$$p(\theta', x'|\theta, x) = p(x'|\theta', \theta, x)p(\theta'|\theta, x) \tag{5}$$

and the choice of proposal distribution is arbitrary from the point of view of the equilibrium distribution of the chain of θ-values. Hence we may take the proposal density for the auxiliary variable x' to be the same as the likelihood, but depending on θ', rather than θ,

$$p(x'|\theta', \theta, x) = p(x'|\theta') = q_{\theta'}(x')/Z_{\theta'}. \tag{6}$$

Then

$$H(\theta', x'|\theta, x) = \frac{f(x'|\theta', y)\pi(\theta')q_{\theta'}(y)q_\theta(x)p(\theta|\theta', x')}{f(x|\theta, y)\pi(\theta)q_\theta(y)q_{\theta'}(x')p(\theta'|\theta, x)}. \tag{7}$$

does not depend on $Z_{\theta'}/Z_\theta$, and the marginalisation over x of the equilibrium distribution $\pi(\theta, x|y)$, gives the desired distribution $\pi(\theta|y)$. In contrast to (4) we now have a much simpler problem of finding the ratio of the distributions of the proposed and current auxiliary variable, $f(x'|\theta', y)/f(x|\theta, y)$, the other factors in (7) presenting no difficulty in evaluation.

Henceforth, for simplicity, we assume that

$$p(\theta'|\theta, x) = p(\theta'|\theta) \tag{8}$$

does not depend on x. For simulation from the proposal density (5) we suppose that it is straightforward to make simulations from $p(\theta'|\theta)$ but not necessarily from $p(x'|\theta', \theta, x)$; for $p(x'|\theta', \theta, x)$ given by (6) appropriate perfect simulation algorithms [2, 11, 15, 19] are used to avoid convergence questions of straightforward MCMC algorithms.

A critical design issue for the algorithm is to choose an appropriate auxiliary density $f(x|\theta, y)$ and proposal density $p(\theta'|\theta)$ so that the algorithm has good mixing and convergence properties. Assume for the moment that Z_θ is known and the algorithm based on (4) has good mixing properties. If we let $f(x|\theta, y) = q_\theta(x)/Z_\theta$, then by (8), (7) reduces to (4), and so the mixing and convergence properties of the two Metropolis-Hastings algorithms using (4) and (7) are the same. Furthermore, recommendations on how to tune Metropolis-Hastings algorithms to obtain optimal acceptance probabilities may exist in the case of (4). This suggests that the auxiliary distribution should approximate the distribution given by q_θ,

$$f(x|\theta, y) \approx q_\theta(x)/Z_\theta. \tag{9}$$

It is interesting to notice that if equality holds in (9), then the states from the chain for the auxiliary variable x can be interpreted as posterior predictions. Choices where (9) are satisfied will be discussed in the following. One particular choice is

$$f(x|\theta, y) = q_{\tilde{\theta}}(y)/Z_{\tilde{\theta}}, \tag{10}$$

where $\tilde{\theta}$ is fixed. This choice is expected to work well if the posterior distribution is concentrated around $\tilde{\theta}$.

3 The Strauss process

The Strauss process (3) is an example of a so-called locally stable point process, and in fact most Markov point processes used in applications are locally stable [9, 19]. Locally stable point processes can be simulated perfectly by an extension of the Propp-Wilson CFTP algorithm, called dominated CFTP, see [15]. Maximum likelihood and maximum pseudo likelihood estimation for the Strauss process is well established [1, 3, 5, 9, 10, 13, 18, 19].

3.1 Specification of auxiliary point processes

In Section 3.2 we consider results for three different kinds of auxiliary variables (referred to as auxiliary point processes) with densities $f = f_1, f_2, f_3$ with respect to μ. In the sequel, for simplicity, we fix R, though our method extends to the case of varying interaction radius, but at the expense of further calculations.

The simplest choice of an auxiliary point process is a homogeneous Poisson point process on S. We let its intensity be given by the MLE $n(y)/|S|$ based on the data y. This auxiliary point process has density

$$f_1(x|\theta, y) = e^{|S|-n(y)}(n(y)/|S|)^{n(x)}, \tag{11}$$

see e.g. [19]. We refer to (11) as the fixed Poisson process.

The second choice takes the interaction into account. Its density is given by

$$f_2(x|\theta, y) \propto \hat{\beta}^n(x)\hat{\gamma}^{s_R(x)} \tag{12}$$

where $(\hat{\beta}, \hat{\gamma})$ is the MLE based on y and approximated by MCMC methods (for details, see Section 3 in [3]). We refer to (12) as the fixed Strauss process and to $(\hat{\beta}, \hat{\gamma})$ as the MCMC MLE.

The densities f_1 and f_2 do not depend on the parameters β and γ, and they are both of the type (10). The third choice we consider takes both interaction and parameters into account, but not the data y. Its density is more complicated to present, but it is straightforward to make a simulation in a sequential way: Choose a subdivision C_i, $i = 1, \ldots, m$ of S into, say, square cells C_i of equal size. The simulation is then done in a single sweep, where the cells are visited once in some order. Each visit to a cell involves updating the point configuration within the cell in a way that only depends on the point configuration within the cells already visited.

Specifically, let $I = \{1, \ldots, m\}$ be the index set for the subdivision and for each $i \in I$ let X_i be a point process on C_i. Furthermore, we introduce a permutation $\rho : I \mapsto I$ of I; we shall later let ρ be random but for the moment we condition on ρ. Then, let $X_{\rho(1)}$ be a homogeneous Poisson point process on $C_{\rho(1)}$ with intensity κ_1 and for $i = 2, \ldots, m$, conditional on $X_{\rho(1)} = x_1, \ldots, X_{\rho(i-1)} = x_{i-1}$, let $X_{\rho(i)}$ be a homogeneous Poisson point process on $C_{\rho(i)}$ with intensity κ_i, where κ_i may depend on $x_1, \ldots, x_{i-1}$ (which is the case below). Then $X = \cup_{i=1}^m X_i$ is a point process which is an example of a so-called partially ordered Markov model (POMM).

POMMs were introduced by Cressie and Davidson [6] who applied POMMs in the analysis of grey scaled digital images. POMMs have the attractive properties that their normalising constants are known (and equal one), and that they can model some degree of interaction. Cressie, Zhu, Baddeley and Nair [7] consider what they call directed Markov point processes (DMPP) as limits of POMM point processes. Such processes are similar to our POMM point process X.

When specifying κ_i, $i \in I$ we want to approximate a Strauss point process. To do so we introduce the following concepts and notation. To each cell C_i, $i \in I$ we associate a reference point $\xi_i \in C_i$. Two cells C_i and C_j, $i \neq j$, are said to be neighbour cells if $\|\xi_i - \xi_j\| \leq R_P$, where $R_P > 0$ is the POMM interaction range (to be specified below). Further, for a given point configuration $x \subset S$, let $n_i(x) = n(x \cap C_{\rho(i)})$ denote the number of points in cell $C_{\rho(i)}$, and let $s_{i,R_P,\rho}(x) = \sum_{j \in I : j < i} n_j(x) \mathbf{1}[\|\xi_{\rho(j)} - \xi_{\rho(i)}\| \leq R_P]$ be the number of points in the cells $C_{\rho(j)}$, $j < i$, which are neighbours to $C_{\rho(i)}$ (setting $s_{1,R_P,\rho}(x) = 0$). Note that we have suppressed the dependence on $\{C_i : i \in I\}$ and $\{\xi_i : i \in I\}$ in the notation. Setting $\kappa_i = \beta_P \gamma_P^{s_{i,R_P,\rho}(x)}$ we have that X is a POMM point process with density

$$f_P(x|\beta_P, \gamma_P, R_P, \rho) = \exp\big(- \beta_P \sum_{i \in I} |C_{\rho(i)}| \gamma_P^{s_{i,R_P,\rho}(x)} \big) \beta_P^{n(x)} \prod_{i \in I} \gamma_P^{n_i(x) s_{i,R_P,\rho}(x)} \tag{13}$$

with respect to μ.

Cressie et al. [7] use a Strauss like DMPP which obviously suffers from directional effects (incidentally this does not show up in the examples they consider). In order to eliminate directional effects in our POMM point process we consider ρ as a random variable uniformly distributed over all permutations of I independent of (θ, y). Moreover, we assume that x given (θ, y, ρ) has density f_3 as specified below. Letting ρ be a random variable requires a slight modification of the auxiliary variable method: each Metropolis-Hastings update consists in first proposing new values of θ and ρ and then conditional on these proposals proposing a new value of x. Using a uniform proposal ρ' the Hastings ratio (7) is modified by replacing $f(x'|\theta', y)/f(x|\theta, y)$ with

$f_3(x'|\theta', \rho', y)/f_3(x|\theta, \rho, y)$ when (θ, x, ρ) is the current state of the chain and (θ', x', ρ') is the proposal; for further details, see Appendix A.

It remains to specify f_3 and (β_P, γ_P, R_P) in terms of $\theta = (\beta, \gamma, R)$. Let $(\beta_P, \gamma_P, R_P) = g(\theta) \equiv (g_1(\theta), g_2(\theta), g_3(\theta))$ where $g : (0, \infty) \times (0, 1] \times (0, \infty) \mapsto (0, \infty) \times (0, 1] \times (0, \infty)$ is a function specified as follows. Conditional on (θ, ρ, y), the POMM auxiliary point process has density

$$f_3(x|\theta, \rho, y) = f_P(x|g(\theta), \rho). \tag{14}$$

When specifying g we note that for point configurations x (except for a null set with respect to a homogeneous Poisson process), $\sum_{i \in I} s_{i,R_P,\rho}(x)$ tends to $s_{R_P}(x)$ as $m \to \infty$. This motivates setting $g_3(\theta) = R$ when the cell size is small compared to R. We would like that

$$(g_1(\theta), g_2(\theta)) = \mathbb{E}[\text{argmax}_{(\tilde{\beta}, \tilde{\gamma})} f_P(Y|\tilde{\beta}, \tilde{\gamma}, R, \rho)] \tag{15}$$

where Y is a Strauss process with parameter $\theta = (\beta, \gamma, R)$ and ρ is uniformly distributed and independent of Y. As this expectation is unknown to us, it is approximated as explained in Appendix B. In Table 1, Section 3.2, we refer to (15) as the "MLE". For comparison, we also consider the identity mapping $g(\theta) = \theta$ in Section 3.2, where we in Table 1 refer to this case as the "identity".

3.2 Results for the auxiliary variable method

In our simulation study, the data y is given by the perfect simulation in Figure 1, where $S = [0, 1]^2$, $\beta = 100$, $\gamma = 0.5$, $R = 0.05$, $n(y) = 75$, and $s_R(y) = 10$. For the MCMC MLE, we obtained $\hat{\beta} = 108$ and $\hat{\gamma} = 0.4$. A priori we assume that $R = 0.05$ is known and β and γ are independent and uniformly distributed on $(0, 150]$ and $(0, 1]$, respectively; perfect simulations for $\beta > 150$ can be slow [2, 3]. For the POMM point process we divide S into $m = N^2$ square cells of side length $1/N$. Below we consider the values $N = 50$, 100, 200, or in comparison with $R = 0.05$, $1/N = 0.02$, 0.01, 0.005. Further details on the auxiliary variable method can be found in Appendix A.

The results are summarised in Table 1 for the different auxiliary processes, and in the POMM case, for different choices of N, the function g in (14), and proposal distributions. Experiments with the algorithm for the fixed Poisson and Strauss and the POMM processes with smaller values of N showed that trace plots of $n(x)$ and $s_R(x)$ (not shown here) may exhibit seemingly satisfactory mixing properties for several million updates and then

Aux.proc.	g	Prop σ_β	Prop σ_γ	MAcP	Extr	c_β	c_γ
Fixed Poisson		2	0.05	0.128	0.151	0.88	0.53
POMM (N=100)	identity	2	0.05	0.171	0.127	0.86	0.54
POMM (N=200)	identity	2	0.05	0.213	0.064	0.85	0.47
POMM (N=50)	MLE	2	0.05	0.246	0.055	0.85	0.46
Fixed Strauss		2	0.05	0.393	0.031	0.79	0.46
POMM (N=100)	MLE	4	0.1	0.298	0.030	0.52	0.21
POMM (N=200)	MLE	4	0.1	0.366	0.014	0.41	0.14
POMM (N=100)	MLE	2	0.05	0.321	0.013	0.79	0.38
POMM (N=200)	MLE	2	0.05	0.406	0.002	0.75	0.33

Table 1: Empirical results: For each auxiliary process considered, one million updates were generated. “Aux.Proc.” is the type of auxiliary process used; g is the type of mapping used for each POMM point process (see the end of Section 3.1); “Prop σ_β” and “Prop σ_γ” are the proposal standard deviations for β and γ; “MAcP” is the mean acceptance probability; “Extr” is the fraction of acceptance ratios below $\exp(-10)$; c_β and c_γ are the lag 100 autocorrelation for β and γ.

get stuck — sometimes for more than 100,000 updates. Therefore we consider the fraction of acceptance probabilities below $\exp(-10)$ as an indicator for the mixing properties of the chain. Table 1 also shows the mean acceptance probability and the lag 100 autocorrelation of β and γ.

The different cases of auxiliary processes in Table 1 are ordered by the values of “Extr” (the fraction of extremely low acceptance probabilities). Seemingly the results for the autocorrelations depend predominantly on the choice of proposal standard deviations for β and γ. Using the POMM point process with $N = 200$ and $g =$ MLE appears to give the best mixing. Figure 2 shows the marginal and joint posterior distributions for β and γ when using the POMM process with $N = 200$, $g =$ MLE, and proposal standard deviations for β and γ equal to 2 and 0.05. From Figure 2 it can be seen that the MCMC MLE $(\hat{\gamma}, \hat{\beta}) = (0.4, 108)$ is not far from the approximative posterior mode obtained by simulation. This is of course to be expected since we have a uniform prior for (γ, β). The marginal posterior modes are close to the posterior mode, since the posterior has nearly elliptical contours.

Despite a seemingly fair number of points in the data, Figure 2 shows a rather large degree of posterior uncertainty about β and γ. The posterior distribution of β suggests that the upper bound of 150 on β should be slightly increased, however we do not expect that increasing this bound would affect

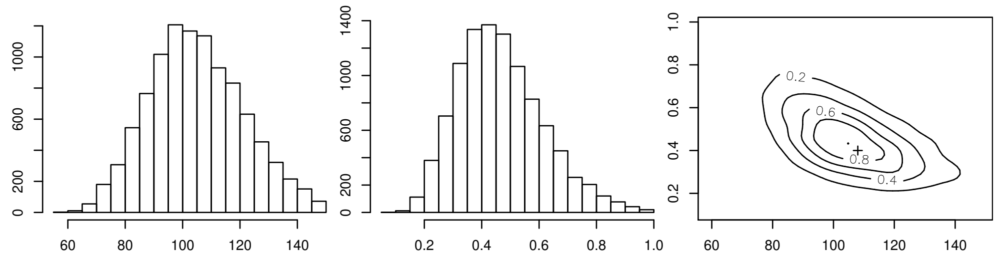

Figure 2: Empirical marginal posterior distributions of β (left plot) and γ (centre plot) generated using a POMM auxiliary process with $N = 200$ and $g =$ MLE. Empirical joint posterior distribution of (β, γ) (right plot) where "." denotes the approximate posterior mode and "+" denotes the approximate MLE.

the overall picture.

In conclusion, to obtain a significant improvement by using a POMM auxiliary process with $g =$ MLE compared to using a fixed Strauss process, a cell side length less than about $R/10$ is needed. Computer times show that using the POMM with $N = 100$ are not much slower than using the fixed Strauss process. For $N = 200$ the POMM takes twice as long as for $N = 100$.

4 Concluding remarks

The technique used in this paper adds significantly to the ability of simulation-based Bayesian inference for Markov point processes, which previously have been subject to one or another approximate analysis. By using the auxiliary variable method presented here in conjunction with perfect sampling, we remove the need for estimating ratios of normalising constants.

We have demonstrated that a workable auxiliary variable distribution has the attribute of closely matching the unnormalised likelihood, while not requiring the computation of a normalising constant. Perhaps the most important consequence of this is that the proposal for the auxiliary variable is then very similar to its full conditional density, which we expect to promote good mixing. For the simulation study in Section 3 a POMM is a more appropriate choice of auxiliary variable than an auxiliary variable density based on the unnormalised likelihood evaluated at the MLE.

To the best of our knowledge, prior specification for Markov point processes has so far not been discussed much in the literature (however, see [12] and

[3]). Here we have chosen uniform priors to keep things simple as our main purpose is to illustrate the auxiliary variable method for Markov point processes. Choosing another prior, our choice of proposal density $p(\theta'|\theta)$ may be different, but otherwise the method is the same.

In [4] we use the auxiliary variable method for a semi-parametric inhomogeneous Markov point process, using again a POMM auxiliary point process.

Appendix A

We now give details for the auxiliary variable method considered in Sections 3.1 and 3.2.

Consider first the Metropolis-Hastings algorithm for (θ, x) updates using either a fixed Poisson or a fixed Strauss auxiliary variable distribution, see (11) and (12). Recall that $\theta = (\beta, \gamma, R)$ where $R = 0.05$ is fixed. As initial values we choose $\theta = (n(y), 1, 0.05)$ and x is a realisation of a Poisson point process on $S = [0,1]^2$ with intensity $n(y)$. Then, if (θ, x) comprises the current state of the Metropolis-Hastings algorithm with $\theta = (\beta, \gamma, R)$, the next state is generated as follows with f in step 3 replaced by either f_1 (fixed Poisson case) or f_2 (fixed Strauss case).

1. Draw proposals β' and γ' from independent normal distributions with means β and γ.
2. Generate a realisation x' from a Strauss process specified by $\theta' = (\beta', \gamma', R)$ and using dominated CFTP.
3. With probability

$$\min\Big\{1, \mathbf{1}[0 < \beta' \le 150, 0 < \gamma' \le 1] \times \left(\frac{\beta'}{\beta}\right)^{n(y)} \left(\frac{\gamma'}{\gamma}\right)^{s_R(y)} \frac{f(x'|y,\theta')}{f(x|y,\theta)} \frac{\beta^{n(x)}\gamma^{s_R(x)}}{\beta'^{n(x')}\gamma'^{s_R(x')}}\Big\}$$

   set $\theta = \theta'$ and $x = x'$, otherwise do nothing.

The standard deviations of the normal distributions in step 1 can be adjusted to give the best mixing of the chain.

Consider next using a POMM auxiliary process. Then an extra auxiliary variable, the random permutation ρ, and an additional step is required in the update above. If the current state consists of (β, γ), ρ, and x, then steps 1 and 2 above are followed by

3. Generate a uniform random permutation ρ'.
4. With probability

$$\min\Big\{1, \mathbf{1}[0 < \beta' \leq 150, 0 < \gamma' \leq 1] \times \left(\frac{\beta'}{\beta}\right)^{n(y)} \left(\frac{\gamma'}{\gamma}\right)^{s_R(y)} \frac{f_3(x'|y,\theta',\rho')}{f_3(x|y,\theta,\rho)} \frac{\beta^{n(x)}\gamma^{s_R(x)}}{\beta'^{n(x')}\gamma'^{s_R(x')}}\Big\}$$

set $(\theta, \rho, x) = (\theta', \rho', x')$, otherwise do nothing.

Here f_3 is given by (14).

Appendix B

When the mapping g in Sections 3.1 and 3.2 is not the identity, it is specified as follows.

Based on the range of the empirical posterior distributions in the fixed Strauss case (not shown here) we define a grid $G = \{50, 52, \ldots, 150\} \times \{0.1, 0.2, \ldots, 1.0\} \times \{0.05\}$. For each grid point $\theta = (\beta, \gamma, R) \in G$, using dominated CFTP, we generate 10 independent realisations $x^{(1)}, \ldots, x^{(10)}$ of a Strauss point process with parameter θ together with the generation of 10 independent random permutations $\rho^{(1)}, \ldots, \rho^{(10)}$. For $\theta \in G$, $g(\theta)$ is given by

$$(g_1(\theta), g_2(\theta)) = \frac{1}{10} \sum_{i=1}^{10} \operatorname{argmax}_{(\tilde{\beta}, \tilde{\gamma})} f_P(x^{(i)} | \tilde{\beta}, \tilde{\gamma}, R, \rho^{(i)}),$$

and $g_3(\theta) = R$. For $(\beta, \gamma, 0.05) \notin G$, we set $g(\beta, \gamma, 0.05) = g(\tilde{\beta}, \tilde{\gamma}, 0.05)$ where $(\tilde{\beta}, \tilde{\gamma}, 0.05) \in G$ is the grid point closest to $(\beta, \gamma, 0.05)$.

Figure 3 shows $g_1(\beta, \gamma, R) - \beta$ and $g_2(\beta, \gamma, R) - \gamma$ for a range of β and γ values when $N = 200$. Results for $N = 50$ and $N = 100$ are almost identical to those for $N = 200$. In cases of strong interaction, i.e. for combinations of low values of γ and high values of β, the parameters $\beta_P = g_1(\beta, \gamma, R)$ and $\gamma_P = g_2(\beta, \gamma, R)$ in the POMM process are much smaller than β and γ in the Strauss process. This is explained by the fact that the interaction in the POMM auxiliary process is weaker than in the Strauss process when $(\beta_P, \gamma_P, R_P) = (\beta, \gamma, R)$.

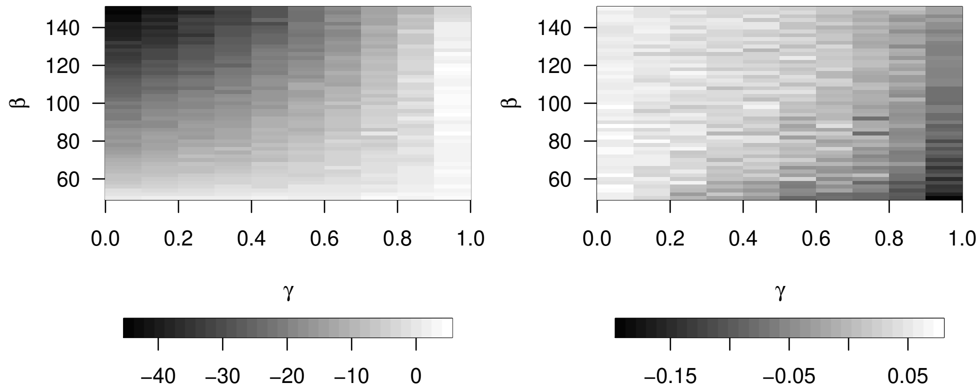

Figure 3: Plot of difference between $g(\theta)$ and θ for $\theta \in G$: $g_1(\beta, \gamma, R) - \beta$ (left) and $g_2(\beta, \gamma, R) - \gamma$ (right).

Acknowledgement

The research of J. Møller was supported by the Danish Natural Science Research Council and the Network in Mathematical Physics and Stochastics (MaPhySto), founded by grants from the Danish National Research Foundation.

References

[1] A. Baddeley and R. Turner. Practical maximum pseudolikelihood for spatial point patterns. *Australian and New Zealand Journal of Statistics*, 42:283–322, 2000.

[2] K. K. Berthelsen and J. Møller. A primer on perfect simulation for spatial point processes. *Bulletin of the Brazilian Mathematical Society*, 33:351–367, 2002.

[3] K. K. Berthelsen and J. Møller. Likelihood and non-parametric Bayesian MCMC inference for spatial point processes based on perfect simulation and path sampling. *Scandinavian Journal of Statistics*, 30:549–564, 2003.

[4] K. K. Berthelsen and J. Møller. Semi-parametric Bayesian inference for inhomogeneous Markov point processes. In preparation.

[5] J. Besag. Some methods of statistical analysis for spatial data. *Bulletin of the International Statistical Institute*, 47:77–92, 1977.

[6] N. Cressie and J. L. Davidson. Image analysis with partially ordered Markov models. *Computational Statistics and Data Analysis*, 29(1):1–26, 1998.

[7] N. Cressie, J. Zhu, A. J. Baddeley, and M. G. Nair. Directed Markov point processes as limits of partially ordered Markov models. *Methodology and Computing in Applied Probability*, 2:5–21, 2000.

[8] A. Gelman and X.-L. Meng. Simulating normalizing contants: From importance sampling to bridge sampling to path sampling. *Statistical Science*, 13:163–185, 1998.

[9] C. J. Geyer. Likelihood inference for spatial point processes. In O. E. Barndorff-Nielsen, W. S. Kendall, and M. N. M. van Lieshout, editors, *Stochastic Geometry: Likelihood and Computation*, pages 79–140, Boca Raton, Florida, 1999. Chapman & Hall/CRC.

[10] C. J. Geyer and J. Møller. Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics*, 21:359–373, 1994.

[11] O. Häggström, M. N. M. Van Lieshout, and J. Møller. Characterisation results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes. *Bernoulli*, 1998. To appear.

[12] J. Heikkinen and A. Penttinen. Bayesian smoothing in the estimation of the pair potential function of Gibbs point processes. *Bernoulli*, 5:1119–1136, 1999.

[13] J. L. Jensen and J. Møller. Pseudolikelihood for exponential family models of spatial point processes. *Annals of Applied Probability*, 3:445–461, 1991.

[14] F. P. Kelly and B. D. Ripley. A note on Strauss' model for clustering. *Biometrika*, 63:357–360, 1976.

[15] W. S. Kendall and J. Møller. Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Advances in Applied Probability*, 32:844–865, 2000.

[16] M. N. M. van Lieshout. *Markov Point Processes and their Applications*. Imperial College Press, London, 2000.

[17] J. Møller, A. N. Pettitt, K. K. Berthelsen, and R. W. Reeves. An efficient MCMC method for distributions with intractable normalising constants. Research report r-2004-02, Department of Mathematical Sciences, Aalborg University, 2004.

[18] J. Møller and R. P. Waagepetersen. An introduction to simulation-based inference for spatial point processes. In J. Møller, editor, *Spatial Statistics and Computational Methods*, Lecture Notes in Statistics 173, pages 143–198. Springer-Verlag, New York, 2003.

[19] J. Møller and R. P. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes.* Chapman and Hall/CRC, 2003.

[20] L. Tierney. Markov chains for exploring posterior distributions. *Annals of Statistics*, 22:1701–1728, 1994.