Perfect simulation and inference for spatial point processes

Kasper K. Berthelsen and Jesper Møller

Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg, Denmark

Abstract The advantages and limitations of using perfect simulation for simulation-based inference for pairwise interaction point processes are discussed. Various aspects of likelihood inference for the Strauss process with unknown range of interaction are studied. A large part of the paper concerns non-parametric Bayesian inference for the interaction function. Markov chain Monte Carlo methods, particularly path sampling, play an important role. Several empirical results and various datasets are considered.

Keywords: Coupling from the past (CFTP); dominated CFTP; exact simulation; Markov chain Monte Carlo; multiscale process; non-parametric Bayesian smoothing; path sampling; pairwise interaction point process; simulation-based inference; Strauss process.

1 Introduction

Since the seminal work by Propp & Wilson (1996) on perfect (or exact) simulation many new perfect sampling algorithms have been developed, particularly for spatial point processes. Meanwhile the possibilities of using perfect simulation for statistical inference have been less explored. This paper concerns simulation-based inference for spatial point process models using Markov chain Monte Carlo (MCMC) methods, including perfect simulation techniques.

We restrict attention to a pairwise interaction point process defined on a bounded planar region S by a density

$$f_{\beta,\varphi}(x) = \left[\beta^{n(x)} \prod_{\{\xi,\eta\}\subseteq x:\, \xi\neq\eta} \varphi(\|\xi-\eta\|)\right] \Big/ c_{\beta,\varphi} \tag{1}$$

with respect to the unit rate Poisson process on S. Here x denotes any finite subset of S, $\beta > 0$ is a parameter, $n(x)$ is the cardinality of x, $\varphi \geq 0$ is an interaction function, $\|\cdot\|$ denotes Eucledian distance, and $c_{\beta,\varphi}$ is a normalising constant. Some restrict on φ is needed to ensure integrability; we assume that $\varphi \leq 1$. In general a closed form expression of $c_{\beta,\varphi}$ is unknown. Sections 3 and 4 discuss the advantages and the computational limitations of using perfect simulation when the normalising constant is estimated by a useful technique called path sampling (Gelman & Meng 1998).

In many applications φ is of primary interest, and so to get rid of β one sometimes condition on $n(x)$ (the conditional distribution of x given $n(x) = n$ does not depend on β). However, since it is unclear whether $n(x)$ in some sense is an ancillary statistic, we have chosen not to fix $n(x)$; see also the discussion in Gates & Westcott (1986), Ripley (1988), and Geyer & Møller (1994). Incidentally, there is at present no practical way of generating perfect simulations conditional on $n(x)$.

The paper is organised as follows.

Section 2 provides a short description of perfect simulation based on the dominated coupling from past algorithm (dominated CFTP; Kendall 1998, Kendall & Møller 2000), and shows some empirical findings for the running time of the algorithm.

Section 3 discusses how dominated CFTP and path sampling can be combined to obtain a Monte Carlo approximation of the normalising constant when φ belongs to a parametric class of models of low dimension. For specificity we consider a Strauss process given by

$$\varphi(r) = \gamma^{\mathbb{1}[r \leq R]} \tag{2}$$

where $0 < \gamma \leq 1$ is an interaction parameter, $R > 0$ is the range of interaction, and $\mathbb{1}[\cdot]$ denotes indicator function (Strauss 1975, Kelly & Ripley 1976). Often in the literature R is assumed to be known, see e.g. Ripley (1989) and Baddeley & Turner (2000), but we treat all three parameters β, γ, R as unknown parameters when we discuss various aspects of likelihood inference.

Section 4 is the main section. It concerns non-parametric Bayesian analysis for the pairwise interaction point process (1) when φ is approximated by a step function,

$$\varphi(r) = \sum_{i=1}^{p} \gamma_i \mathbb{1}[r_{i-1} < r \leq r_i] + \mathbb{1}[r > r_p] \tag{3}$$

where $p \in \mathbb{N}_0$ is the number of change-points $r_1, \ldots, r_p$ with $r_0 = 0 < r_1 < \ldots < r_p$, and where $0 < \gamma_1 < \ldots < \gamma_p < 1$. This is an obvious extension of (2) called the multiscale process (Penttinen 1984). We impose a prior for (β, ψ) where

$$\psi = \{(r_1, \gamma_1), \ldots, (r_p, \gamma_p)\}$$

is viewed as a marked point process, and discuss how a fully Bayesian MCMC analysis can be performed when the unknown normalising constant from the likelihood term is approximated by path sampling. A similar situation is considered in Heikkinen & Penttinen (1999), but they condition on $n(x)$ in the likelihood, and fix p and $(r_1, \ldots, r_p)$ in the prior. They develop an MCMC root-finding algorithm for the maximum a posteriori estimate of $(\gamma_1, \ldots, \gamma_p)$, but mention at the end of their paper the advantage of estimating φ by the Monte Carlo posterior mean, which is usually a smooth curve (Arjas & Gasbarra 1994). Among other things we consider such estimates.

The suggested methods in Sections 3 and 4 are illustrated on some real and simulated datasets.

2 Perfect simulation using dominated CFTP

Different kinds of perfect samplers have been developed for spatial point processes, cf. the surveys in Møller (2001), Berthelsen & Møller (2001), and the references therein. In general dominated CFTP is the most applicable algorithm for pairwise interaction point processes; for extensions to locally stable point processes, see Kendall & Møller (2000) and Berthelsen & Møller (2001).

The algorithm uses a so-called dominating process $D_i,\ i = 0, -1, -2, \ldots,$ of finite point configurations contained in S. This is a Markov chain, which is generated backwards in time as follows. Let ν_β denote the homogeneous Poisson point process on S with rate $\beta > 0$, and let $\emptyset$ denote the empty point configuration. Then $D_0 \sim \nu_\beta$, and for $i = 0, -1, -2, \ldots$,

- with probability $\beta/(\beta + n(D_i))$ we make a backwards birth: we generate a point η_i which is uniformly distributed on S, and set $D_{i-1} = D_i \cup \{\eta_i\}$;
- else we make a backwards death: we draw randomly uniformly a point $\xi_i \in D_i$, and set $D_{i-1} = D_i \setminus \{\xi_i\}$ (if $D_i = \emptyset$ we set $\{\xi_i\} = \emptyset$); for later use we also draw a "mark" $M_i \sim \text{Uniform}[0,1]$.

The D_i and (in case of a backwards death) their marks M_i are easy to generate. They are used for generating so-called upper and lower processes $U^j = \{U^j_j, \dots, U^j_0\}$ and $L^j = \{L^j_j, \dots, L^j_0\}$, which are started at times $j = 0, -1, -2, \dots$, and which are generated forwards in time until time 0 as follows. Initially set $U^j_j = D_j$ and $L^j_j = \emptyset$. For $i = j+1, \dots, 0$,

$$D_i = D_{i-1} \setminus \{\eta_i\} \Rightarrow U^j_i = U^j_{i-1} \setminus \{\eta_i\} \quad \text{and} \quad L^j_i = L^j_{i-1} \setminus \{\eta_i\}, \tag{4}$$

and

$$D_i = D_{i-1} \cup \{\xi_i\} \Rightarrow$$

$$U^j_i = \begin{cases} U^j_{i-1} \cup \{\xi_i\} & \text{if } M_i \leq \prod_{\eta \in L^j_{i-1}} \varphi(\|\xi_i - \eta\|) \\ U^j_{i-1} & \text{otherwise} \end{cases} \tag{5}$$

$$\text{and} \quad L^j_i = \begin{cases} L^j_{i-1} \cup \{\xi_i\} & \text{if } M_i \leq \prod_{\eta \in U^j_{i-1}} \varphi(\|\xi_i - \eta\|) \\ L^j_{i-1} & \text{otherwise.} \end{cases} \tag{6}$$

Note that $(U^j, L^j), (U^{j-1}, L^{j-1}), \dots$ are coupled by the same η_i, ξ_i, M_i for $i > j$.

Berthelsen & Møller (2001) propose to use a doubling scheme given by $j_k = -2^k T_{\min},\ k = 0, 1, 2, \dots$, where

$$T_{\min} = \begin{cases} \inf\{-i : D_i \cap D_0 \neq \emptyset, D_{i-1} \cap D_0 = \emptyset\} & \text{if } D_0 \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$$

is the first time a point in D_0 is born (when considering D forwards in time). We use the same scheme and let

$$T = \inf\{-j_k : U^{j_k}_0 = L^{j_k}_0\}$$

denote the first time we have coalescence at time 0 when the upper and lower processes are started at times times $j_0, j_1, j_2, \dots$. We have that $U^{-T}_0 \sim f_{\beta,\varphi}$.

To summarise we use the following dominated CFTP algorithm where we set $j_{-1} = 0$:

1. generate $D_0 \sim \nu_\beta$;
2. repeat the following steps 3–4 for $k = 0, 1, 2, \ldots$ until $U_0^{j_k} = L_0^{j_k}$;
3. generate backwards $D_{j_{k-1}-1}, \ldots, D_{j_k}$ and generate the associated marks $M_i \sim \text{Uniform}[0,1]$ each time $D_i \setminus D_{i-1} \neq \emptyset$, $j_k < i \leq j_{k-1}$;
4. generate forwards $(U_{j_k}^{j_k}, L_{j_k}^{j_k}), \ldots, (U_0^{j_k}, L_0^{j_k})$ as in (4)–(6);
5. return $U_0^{-T} \sim f_{\beta,\varphi}$.

The computer time of the algorithm depends of course on how efficient it is implemented. It is advantageous to exploit the fact that the products in (5) and (6) depend only on local information. Specifically, if S is rectangular and $R = \sup\{r > 0 : \varphi(r) < 1\}$ is the range of interaction, we use a subdivision $S = \cup_k C_k$ of rectangular cells of side lengths $\geq R$, and exploit the fact that if $\xi \in C_k$, then $\prod_{\eta \in x} \varphi(\|\xi - \eta\|)$ depends only on the points from x falling in C_k and in the neighbouring cells to C_k.

Example 1: The dominated CFTP algorithm for the Strauss process (2) gets slower and slower as the interaction parameter γ decreases. Fig. 1 shows the mean of T in the limiting case of (2) as $\gamma \to 0$, i.e. a hard core point process given by $\varphi(r) = \mathbb{1}[r > R]$. For fixed $R = 0.1$, the figure indicates the existence of a "critical value" $\beta_{\text{crit}} \approx 100$, where $\log \mathbb{E}T$ is nearly linear for $\beta < \beta_{\text{crit}}$, and $\log \mathbb{E}T$ grows faster and faster for $\beta > \beta_{\text{crit}}$. For fixed $\beta = 100$, $\log \mathbb{E}T$ depends on R in a similar way but with respect to a critical value $R_{\text{crit}} \approx 0.1$.

3 Perfect simulation and likelihood inference for the Strauss process

In this section we consider a Strauss process with density

$$f_\theta(x) = \beta^{n(x)} \gamma^{s_R(x)} / c_\theta \tag{7}$$

where $\theta = (\beta, \gamma, R) \in (0, \infty) \times (0, 1] \times (0, \infty)$ and

$$s_R(x) = \sum_{\{\xi,\eta\} \subseteq x} \mathbb{1}[0 < \|\xi - \eta\| \leq R]$$

is the number of R-close pairs of points in x.

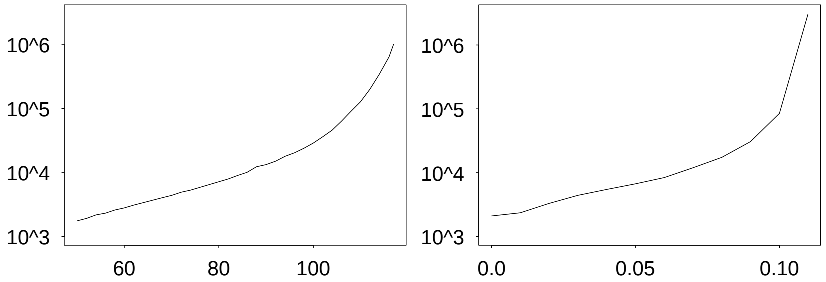

Fig. 1. Mean coalescence times on a log scale for a hard core process on $S = [0,1]^2$. Left plot: $\mathbb{E}T$ versus β when $R = 0.1$. Right plot: $\mathbb{E}T$ versus R when $\beta = 100$.

3.1 Path sampling

The normalising constant of the Strauss process is unknown except in the Poisson case:

$$c_\theta = \exp(\beta) \quad \text{if } \gamma = 1. \tag{8}$$

We discuss below how path sampling can be combined with perfect simulation for estimating ratios of normalising constants.

We use a simple version of path sampling; other versions are discussed in Gelman & Meng (1998). We start by letting (β, R) be fixed, and notice the following identity: for $\theta_i = (\beta, \gamma_i, R),\ i = 1, 2$, with $0 < \gamma_1 < \gamma_2 \leq 1$,

$$\ln(c_{\theta_2}/c_{\theta_1}) = \int_{\gamma_1}^{\gamma_2} \mathbb{E}_\theta s_R/\gamma \, \mathrm{d}\gamma \tag{9}$$

where $\theta = (\beta, \gamma, R)$ varies with γ in the integral, and where $\mathbb{E}_\theta$ denotes expectation with respect to (7). Taking $\gamma_2 = 1$ we know c_{θ_2} by (8). Let $\gamma^{(j)} = \gamma_1 + j\delta,\ j = 0, \ldots, k$, be a grid of γ-values where $k \geq 1$ and $\delta = (\gamma_2 - \gamma_1)/k$. Estimating $\mathbb{E}_\theta s_R$ for $\theta = \theta^{(j)} = (\beta, \gamma^{(j)}, R),\ j = 0, \ldots, k-1$, as described below, we approximate the integral in (9) by a Riemann sum,

$$\ln(c_{\theta_2}/c_{\theta_1}) \approx \delta \left[\mathbb{E}_{\theta^{(0)}} s_R/(2\gamma^{(0)}) + \mathbb{E}_{\theta^{(k)}} s_R/(2\gamma^{(k)}) + \sum_{j=1}^{k-1} \mathbb{E}_{\theta^{(j)}} s_R/\gamma^{(j)} \right].$$

Similarly, for $j = 0, \ldots, k-1$,

$$\ln(c_{\theta_2}/c_{\theta^{(j)}}) \approx \delta\left[\mathbb{E}_{\theta^{(j)}} s_R/(2\gamma^{(j)}) + \mathbb{E}_{\theta^{(k)}} s_R/(2\gamma^{(k)}) + \sum_{l=j+1}^{k-1} \mathbb{E}_{\theta^{(l)}} s_R/\gamma^{(l)}\right]. \tag{10}$$

Thereby we obtain estimates of $c_{\theta^{(j)}},\ j = 0, \ldots, k-1$. Finally, combining the grid of γ-values with a grid of (β, R)-values, all normalising constants c_θ can first be estimated for θ in the corresponding 3D-grid, and next by interpolation and extrapolation over a region of the parameter space.

For a given value of θ, we estimate $\mathbb{E}_\theta s_R$ by

$$\sum_{i=0}^{m} s_R(X_i)/(m+1) \tag{11}$$

where $X_0 \sim f_\theta$ is generated by the dominated CFTP algorithm, and where $X_1, \ldots, X_m$ is a sample of length $m \geq 0$ obtained by the Metropolis-Hastings algorithm in Geyer & Møller (1994). We find this attractive for several reasons: If the initial state is not in equilibrium, we need to determine an appropriate burn-in before we start sampling. As the rate of convergence for the Metropolis-Hastings algorithm depends much on the value of θ, so should the burn-in. At least to our knowledge there are no analytical results (Møller 1999) or automatic methods (apart from perfect simulation) for determining an appropriate burn-in. Hence we have to perform a thorough output analysis for each value of θ in the 3D-grid. But this can be rather time-consuming, and we can not be assured if the burn-in is appropriately determined. However, when X_0 is generated by the dominated CFTP algorithm, we let the machine do all the work so that we are certain that the sample $X_i \sim f_\theta,\ i = 0, \ldots, m$, is in equilibrium. In particular, the estimate (11) is unbiased. Finally, perfect simulation allow us to check a method based on non-perfect simulation; we illustrate this point in Section 4.3.

It remains to specify k and m. For (10) to be a "good" approximation we need a sufficiently large k. As k increases, the precision of the unbiased estimator (11) becomes less and less important. We can even take $m = 0$ for sufficiently large values of k. On the other hand, the perfect simulation of X_0 is computationally more demanding than a Metropolis-Hastings update $X_i \to X_{i+1}$. So the "right" choice of (k, m) requires some experimentation as in Example 2 below.

3.2 Empirical results

In the following two examples, $S = [0, 1]^2$ is the unit square.

Example 2: In Fig. 2 we fix $(\beta, R) = (100, 0.1)$ and consider 6 different values of (k, m). For each choice of (k, m), Fig. 2 shows the empirical mean and a 95% confidence region when the curve $\ln(c_{(\beta,1,R)}/c_{(\beta,\gamma,R)})$, $0.1 \le \gamma \le 1$, has been estimated 100 times (independently of each others) by path sampling as described above. The computer time relative to the fastest case (the upper left plot where $(k, m) = (16, 0)$) is shown in each plot. The plots indicate that it suffices to have a small value of k but a large value of m. Reasonable good and fast estimates are obtained when $(k, m) = (16, 1000)$ and $(k, m) = (16, 10000)$.

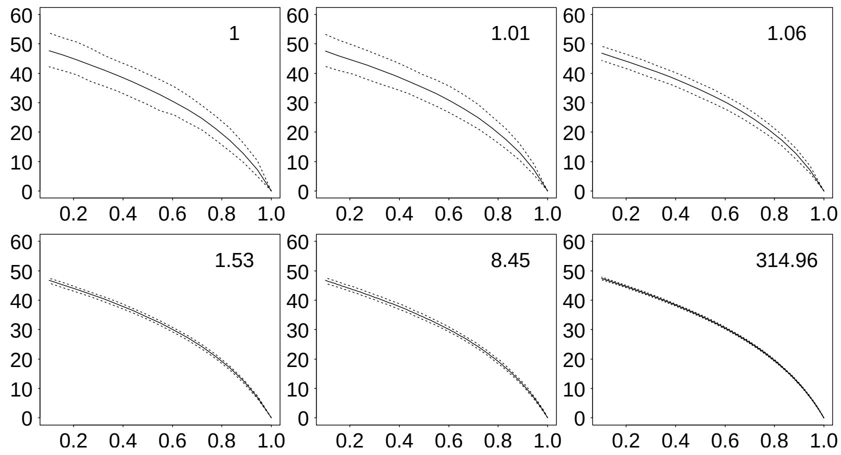

Fig. 2. Empirical means of estimated curves $\gamma \to \ln(c_{(\beta,1,R)}/c_{(\beta,\gamma,R)})$ (full lines) and 97.5% and 2.5% quantiles (broken lines) when $(\beta, R) = (100, 0.1)$. The figures show the computer times relative to the fastest case. First row, from left to right: $(k, m) = (16, 0)$, $(16, 100)$, $(16, 1000)$. Second row, from left to right: $(k, m) = (16, 10000)$, $(128, 1000)$, $(4096, 0)$.

Example 3: Fig. 3 shows a perfect simulation x of a Strauss process on the unit square when $(\beta, \gamma, R) = (100, 0.5, 0.05)$. We use this as our data in the following. Note that both $n(x) = 68$ and $s_{0.05}(x) = 3$ are rather small.

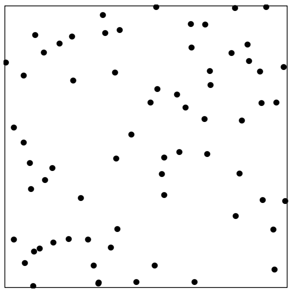

Fig. 3. A perfect simulation of the Strauss process on the unit square when $(\beta, \gamma, R) = (100, 0.5, 0.05)$.

The estimated likelihood function is rather flat with respect to β as illustrated in Fig. 4. Here a rectangular 3D-grid of $76 \times 17 \times 10 = 12920$ points for (β, γ, R) in $[35, 110] \times [0.1, 1] \times [0.01, 0.1]$ has been used. We obtain an approximate MLE $(\hat{\beta}, \hat{\gamma}, \hat{R}) = (103, 0.16, 0.05)$, which except for $\hat{\gamma}$ is close to the true value; the low value of $\hat{\gamma}$ is caused by the low value of $s_{0.05}(x)$.

We have also investigated the empirical distribution for $(\hat{\beta}, \hat{\gamma}, \hat{R})$ obtained from 1000 independent perfect simulations with $(\beta, \gamma, R) = (100, 0.5, 0.05)$ (not shown here). The empirical marginal distributions have distinct modes about the true parameter values, and empirical means 100.88, 0.50, 0.057, respectively. The correlation is most pronounced between $\hat{\gamma}$ and $\hat{R}$: the empirical correlations are 0.17 for $(\hat{\beta}, \hat{\gamma})$, 0.24 for $(\hat{\beta}, \hat{R})$, and 0.80 for $(\hat{\gamma}, \hat{R})$.

The left plot in Fig. 5 shows the empirical distribution of the log likelihood ratio statistic $-2 \ln Q$ for the Poisson hypothesis $\gamma = 1$. It is obtained from 10000 independent simulations under the Poisson model with $\beta = 68$ (the MLE when $\gamma = 1$). The degrees of freedom between the Strauss and the Poisson model is $3 - 1 = 2$. As seen in the figure, $-2 \ln Q$ is better described by a $\chi^2(4)$-distribution than a $\chi^2(2)$-distribution.

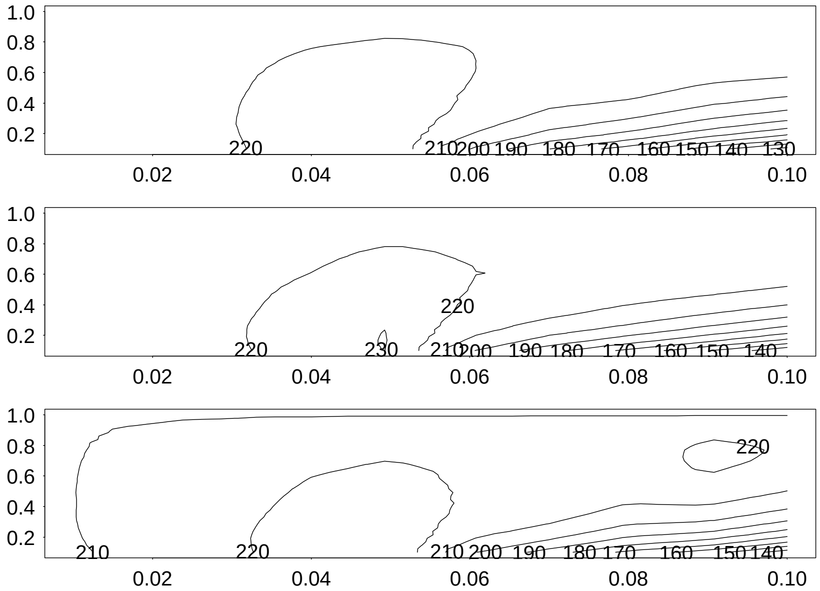

Fig. 4. Estimated profile log likelihood functions for (R, γ) when $\beta = 90,\ 103,\ 110$ (from top to bottom).

The Strauss model is only an exponential family model if we fix R; then the degrees of freedom between the Strauss and the Poisson model is $2 - 1 = 1$. For fixed $R = 0.05$, approximating the distribution of $-2\ln Q$ by a $\chi^2(1)$-distribution seems rather satisfactory, cf. the right plot in Fig. 5.

4 Non-parametric Bayesian MCMC inference for pairwise interaction point processes

In this section we discuss non-parametric Bayesian MCMC inference when the likelihood term is given by the pairwise interaction point process (1). In Section 4.1 we specify a prior for (β, φ), derive the posterior given an observation x, and point out the similarities and differences with the approach used in Heikkinen & Penttinen (1999). In Section 4.2 we propose an MCMC algorithm for the posterior. As ratios of normalising constants $c_{\beta,\varphi}$ has to be estimated in each update, path sampling is used during the simulations. Finally, in Section 4.3 we consider some simulated and real datasets, and

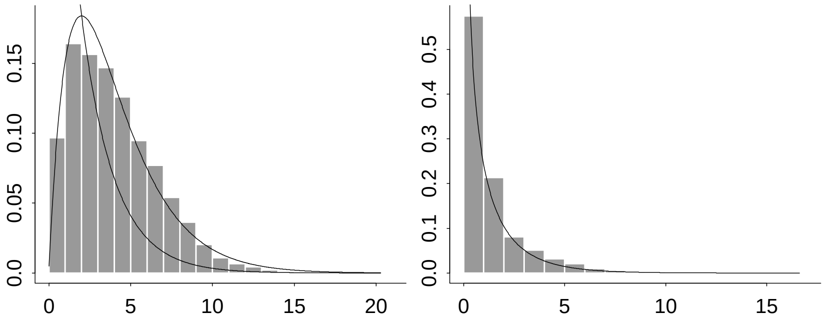

Fig. 5. The empirical distribution of $-2\ln Q$ under the Poisson hypothesis $\gamma = 1$. Left plot: when all three parameters in the Strauss process are unknown; the densities of $\chi^2(2)$ and $\chi^2(4)$ are also shown. Right plot: fixed $R = 0.05$, and the density of $\chi^2(1)$.

discuss the results obtained by either perfect or non-perfect MCMC simulation when the normalising constants are estimated.

4.1 Specification of prior and posterior

Assuming that φ is non-decreasing, we approximate φ by the step function (3); for a discussion on whether a step function or another approximation such as splines should be used, see Heikkinen & Penttinen (1999). Recall that the step function is given by the marked point process $\psi = \{(r_1, \gamma_1), \ldots, (r_p, \gamma_p)\}$, where the change-points are ordered $0 < r_1 < \ldots < r_p$ and their associated marks satisfy the constraint $0 < \gamma_1 < \ldots < \gamma_p < 1$. Note that r_p is the range of interaction. Below we let the number of change points $p \geq 0$ be random, and set $r_0 = 0$.

Set $\gamma_0 = 0$, $\gamma_{p+1} = 1$, and $\delta_i = (\gamma_i - \gamma_{i-1})$, $i = 1, \ldots, p+1$. For p fixed, the state space of $(\delta_1, \ldots, \delta_{p+1})$ is the simplex in $\mathbb{R}^{p+1}$. It is convenient to transform $(\delta_1, \ldots, \delta_{p+1})$ into $(\zeta_1, \ldots, \zeta_p) = (\ln(\delta_2/\delta_1), \ldots, \ln(\delta_{p+1}/\delta_p))$ with state space $\mathbb{R}^p$. We assume a priori that ψ and β are independent, and

(a) $r_1 < \ldots < r_p$ are the events of a homogeneous Poisson process on $[0, r_{\max}]$ of rate κ;
(b) conditionally on $(r_1, \ldots, r_p)$, we have that $\zeta_p, \ldots, \zeta_1$ is a Markov chain with $\zeta_p \sim \mathrm{N}(0, \sigma_p^2)$ and $\zeta_i | \zeta_{i+1} \sim \mathrm{N}(\zeta_{i+1}, \sigma_i^2)$, $i = p-1, \ldots, 1$;
(c) $\beta \sim \mathrm{Uniform}[\beta_{\min}, \beta_{\max}]$.

In contrast to our model, Heikkinen & Penttinen (1999) fix p and $r_1, \ldots, r_p$ in the prior, and they condition on $n(x) = n$ in the likelihood. They assume that $\gamma_p, \ldots, \gamma_1$ is a Markov chain with $\ln \gamma_i | \ln \gamma_{i+1} \sim N(\ln \gamma_{i+1}, \sigma_i^2)$, $i = p, \ldots, 1$, where $\gamma_{p+1} = 1$ and the σ_i^2 are determined so that

$$\mathbb{V}\mathrm{ar}(\gamma_1 | \gamma_2) = \ldots = \mathbb{V}\mathrm{ar}(\gamma_p | \gamma_{p+1}).$$

As noticed in Heikkinen & Penttinen (1999), some constraints such as $0 < \gamma_i \leq 1,\ i = 1, \ldots, p$, are needed in their prior model in order to obtain a well defined likelihood in the unconditional case.

The hyperparameters $r_{\max} > 0$, $\kappa > 0$, etc. in (a)–(c) should be chosen in accordance to background knowledge on (β, φ). The choice of $(r_{\max}, \kappa)$ specifies the "resolution", since $\mathbb{E}p = \kappa r_{\max}$. Often a crude estimate $\hat{R}$ of the range of interaction can be obtained (van Lieshout & Baddeley 1996). Then $r_{\max}$ should be chosen to be substantially larger than $\hat{R}$, so that $\varphi(r)$ is expected to be rather stable and close to 1 for large values of $r \leq r_{\max}$ (hence it is natural in (b) to generate the Markov chain backwards rather than forwards). The choice of variances in (b) may depend on the change-points, but in the following $\sigma_p^2 = \sigma^2$ is chosen not to depend on these, and for $i < p$, we allow σ_i^2 to depend only on ζ_{i+1} and (i, p); two specific models are studied in (b1) and (b2) below. Finally, in order to determine the hyperparameters $0 \leq \beta_{\min} < \beta_{\max}$, it may be used that $n(x)/|S|$ is an upper bound on the MLE of β, where $|S|$ is the area of S.

The means for the Markov chain in (b) are specified so that $(\delta_2/\delta_1, \ldots, \delta_{p+1}/\delta_p)$ and $(\delta_1/\delta_2, \ldots, \delta_p/\delta_{p+1})$ are identically distributed. We have studied two models for the variances in detail:

(b1) $\sigma_1^2 = \ldots = \sigma_p^2 = \sigma^2$.

(b2) $\sigma_p^2 = \sigma^2$ and $\sigma_i^2 = \ln[1/2 + \{(\sigma\delta_{i+1}/\delta_{i+2})^2 + 1/4\}^{1/2}]$, $i = 1, \ldots, p-1$.

Here $\sigma > 0$ is a tuning parameter. Under (b1) the conditional distribution of $(\zeta_1, \ldots, \zeta_p)$ given $(r_1, \ldots, r_p)$ is a multivariate normal distribution with mean zero, variances $\mathbb{V}\mathrm{ar}(\zeta_i) = (p - i + 1)\sigma^2$, $1 \leq i \leq p$, and correlations $\mathbb{C}\mathrm{orr}(\zeta_{i-j}, \zeta_i) = ((p - i + 1)/(p - i + j + 1))^{1/2}$, $1 \leq j < i \leq p$. Thus $(\delta_1, \ldots, \delta_{p+1})$ follows an additive logistic normal distribution (Aitchison 1986). The conditional distribution of

$(\zeta_1, \ldots, \zeta_p)$ is more complicated under (b2). Here the σ_i^2 are determined so that

$$\mathbb{V}\mathrm{ar}(\delta_2/\delta_1|\delta_3/\delta_2) = \ldots = \mathbb{V}\mathrm{ar}(\delta_p/\delta_{p-1}|\delta_{p+1}/\delta_p) = \mathbb{V}\mathrm{ar}(\delta_{p+1}/\delta_p).$$

Simulated results for the prior distribution of φ look rather similar when different values of σ are used in the two models (b1) and (b2), so the choice of σ is important while it is less important which model is used. Fig. 6 shows the smoothing effect of σ in (b1) when $\kappa r_{\max} = 5$.

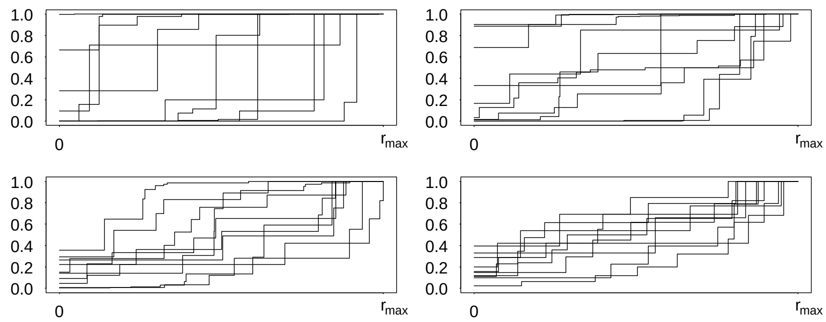

Fig. 6. Each plot shows ten independent realizations of φ under (b1). From top left to bottom right: $\sigma^2 = 10, 1, 0.1, 0.01$.

Finally, we derive the posterior density. The Jacobian of the transformation $(\zeta_1, \ldots, \zeta_p) \to (\gamma_1, \ldots, \gamma_p)$ is

$$|\partial(\gamma_1, \ldots, \gamma_p)/\partial(\zeta_1, \ldots, \zeta_p)| = \delta_1 \times \cdots \times \delta_{p+1}.$$

Let μ denote the unit rate Poisson process on $[0, r_{\max}] \times [0, 1]$. Then under (a)–(b), $\psi = \{(r_1, \gamma_1), \ldots, (r_p, \gamma_p)\}$ with $r_1 < \ldots < r_p$ has prior density

$$\pi(\psi) \propto \kappa^p \mathbb{1}[0 < \gamma_1 < \ldots < \gamma_p < 1]/(\delta_1 \times \cdots \times \delta_{p+1}) \times \prod_{i=1}^{p} \left(2\pi\sigma_i^2\right)^{-1/2} \exp\left(-\left(\zeta_i - \zeta_{i+1}\right)^2/\left(2\sigma_i^2\right)\right) \qquad (12)$$

with respect to μ, where we set $\zeta_{p+1} = 1$. Setting

$$s_i = \sum_{\{\xi,\eta\}\subseteq x} \mathbb{1}[r_{i-1} < \|\xi - \eta\| \leq r_i], \quad i = 1, \ldots, p,$$

the posterior density of $\theta = (\beta, \psi)$ is

$$\pi(\theta|x) \propto \pi(\psi)\beta^{n(x)}\gamma_1^{s_1} \times \cdots \times \gamma_p^{s_p}/c_\theta \tag{13}$$

with respect to the product measure of the Lebesgue measure on $[\beta_{\min}, \beta_{\max}]$ and μ, where $c_\theta = c_{\beta,\varphi}$ denotes the normalising constant of the likelihood.

4.2 Simulation of the posterior

We use a hybrid Metropolis-Hastings algorithm for the posterior (13), where we alternate between updating β and ψ. In the description below of both types of updates we let (β, ψ) denote the current state of the algorithm.

For a β update, we use a Metropolis random walk algorithm: generate $\beta' \sim \text{Uniform}\big([\max\{\beta_{\min}, \beta - \epsilon\}, \min\{\beta_{\max}, \beta + \epsilon\}]\big)$, and replace β by β' with probability $\min\{1, \alpha_1(\beta, \beta', \psi)\}$, where

$$\alpha_1(\beta, \beta', \psi)\} = \frac{c_{\beta,\psi}}{c_{\beta',\psi}} \left(\frac{\beta'}{\beta}\right)^{n(x)} \frac{\min\{\beta_{\max}, \beta + \epsilon\} - \max\{\beta_{\min}, \beta - \epsilon\}}{\min\{\beta_{\max}, \beta' + \epsilon\} - \max\{\beta_{\min}, \beta' - \epsilon\}} \tag{14}$$

where $\epsilon > 0$ is a tuning parameter; retain β otherwise.

For a ψ update, we use a Metropolis-Hastings algorithm of the type studied in Geyer & Møller (1994): if $\psi = \{(r_1, \gamma_1), \ldots, (r_p, \gamma_p)\}$, then

- with probability 1/2 we make a birth proposal: generate $r' \sim \text{Uniform}[0, r_{\max}]$; find the interval where $r_{i-1} < r' \le r_i$ (setting $r_{p+1} = r_{\max}$); generate $\gamma' \sim \text{Uniform}[\gamma_{i-1}, \gamma_i]$; replace ψ by $\psi' = \psi \cup \{(r', \gamma'\})$ with probability $\min\{1, \alpha_2(\beta, \psi, \psi')\}$, where

  $$\alpha_2(\beta, \psi, \psi') = \frac{c_{\beta,\psi}}{c_{\beta,\psi'}} \left(\frac{\gamma'}{\gamma_i}\right)^{s'} \frac{\pi(\psi')}{\pi(\psi)} \frac{r_{\max}(\gamma_i - \gamma_{i-1})}{p+1} \tag{15}$$

  with

  $$s' = \sum_{\{\xi,\eta\} \subseteq x} \mathbb{1}[r_{i-1} < \|\xi - \eta\| \le r'];$$

  retain ψ otherwise;

- else we generate a death proposal: generate randomly uniformly $(r_i, \gamma_i) \in \psi$ (if $\psi = \emptyset$ we set $\{r_i, \gamma_i)\} = \emptyset$); replace ψ by $\psi' = \psi \backslash \{(r_i, \gamma_i)\}$ with probability $\min\{1, \alpha_2(\beta, \psi', \psi)^{-1}\}$; retain ψ otherwise.

Under (b1), $\pi(\psi')/\pi(\psi)$ in (15) depends only on $\gamma_{i-2}, \gamma_{i-1}, \gamma', \gamma_i, \gamma_{i+1}$ (with obvious adjustments at the boundaries $i = 1$ and $i = p$), cf. (12). For a β-update, if for example $\beta < \beta'$, the term

$$\ln(c_{\beta',\psi}/c_{\beta,\psi}) = \int_{\beta}^{\beta'} \mathbb{E}_{\tilde{\beta},\psi} n/\tilde{\beta}\, \mathrm{d}\tilde{\beta} \tag{16}$$

in (14) has to be estimated; and for a ψ-update, the term

$$\ln(c_{\beta,\psi\cup(r',\gamma')}/c_{\beta,\psi}) = \int_{\gamma_{i-1}}^{\gamma'} \mathbb{E}_{\beta,\psi\cup(r',\gamma')}(s_i - s')/\gamma\, \mathrm{d}\gamma \tag{17}$$

in (15) has to be estimated. We can do this by path sampling as discussed below.

4.3 Empirical results

We illustrate now how our method applies for two datasets when the prior for (φ, β) is specified by (a)–(c) and (b1).

Example 5 concerns the locations of 152 displaced amacrine cells within a rectangular $1070 \times 600\,\mu m^2$ region, see Fig. 7. This dataset has been analysed in many papers, including Diggle & Gratton (1984) and Heikkinen & Penttinen (1999). Diggle & Gratton (1984) modelled the data by the interaction function

$$\varphi_{\mathrm{DG}}(r) = \begin{cases} 0, & r < \sigma_{\mathrm{DG}} \\ ((r - \sigma_{\mathrm{DG}})/(\rho_{\mathrm{DG}} - \sigma_{\mathrm{DG}}))^{\beta_{\mathrm{DG}}}, & \sigma_{\mathrm{DG}} \le r \le \rho_{\mathrm{DG}} \\ 1, & r > \rho_{\mathrm{DG}}. \end{cases} \tag{18}$$

Note that σ_{DG} is a hard core parameter and ρ_{DG} specifies the range of interaction. Diggle & Gratton (1984) obtained the estimate

$$(\sigma_{\mathrm{DG}}, \rho_{\mathrm{DG}}, \beta_{\mathrm{DG}}) = (19, 76, 1.67)$$

by an ad hoc method.

To examine whether perfect simulation is feasible under the estimated Diggle-Gratton model, we have considered a plot of $\mathbb{E}T$ versus β (not shown here). It is similar to the left plot in Fig. 1, but with a critical value for β around $200/(1070 \times 600)$. Simulated point patterns for $\beta = 200/(1070 \times 600)$, $(\sigma_{\mathrm{DG}}, \rho_{\mathrm{DG}}, \beta_{\mathrm{DG}}) = (19, 76, 1.67)$, and $S = [0, 1070] \times [0, 600]$ contain on average about 65 points. This is far below the 152 points in the data, so perfect simulation is not feasible.

When using path sampling we have to estimate the means of n and $s_i - s'$ from (16) and (17). We do this along similar lines as in Section 3.1 but let the initial state be given by the result after an appropriate burn-in (again the sample is generated by the Metropolis-Hastings algorithm in Geyer & Møller (1994) but with a multiscale process as equilibrium distribution).

For comparison we use both perfect and non-perfect simulations in Example 4. The dataset in Example 4 is a perfect simulation of the Diggle-Gratton model but with $\beta = 100$, $(\sigma_{\mathrm{DG}}, \rho_{\mathrm{DG}}, \beta_{\mathrm{DG}}) = (0.025, 0.1, 1.67)$, and $S = [0, 1]^2$, see Fig. 7. This point pattern contains only 48 points, and compared to the amacrine data it is less regular.

Example 4: For the simulated data we let $(k, m) = (10, 1000)$. Compared to the values of k used in Fig. 2, it suffices to use the present smaller value of k, since we usually integrate over a shorter interval in (17). Furthermore, we chose $r_{\max} = 0.14$, $\kappa r_{\max} = 5$, $\sigma^2 = 1$, $\beta_{\min} = 50$, $\beta_{\max} = 115$, and $\epsilon = 2.5$. Time series plots for different statistics (not shown here) show that the algorithm for posterior simulations is mixing well, with an appropriate burn-in of only a few thousands updates, using an initial state of ψ where the r_i are generated from the Poisson prior and the δ_i are all equal.

Fig. 8 shows some results based on a sample of 150000 posterior realizations of (ψ, β).

The top left plot in Fig. 8 shows a clear difference between the posterior and prior mean of $\varphi(r)$, $0 < r < r_{\max}$. The posterior mean of $\varphi(r)$ is rather close to $\varphi_{\mathrm{DG}}(r)$ except for intermediate values of r where it clearly overestimates the true interaction function. However, the credibility interval for $\varphi(r)$ is rather wide, especially for intermediate values of r.

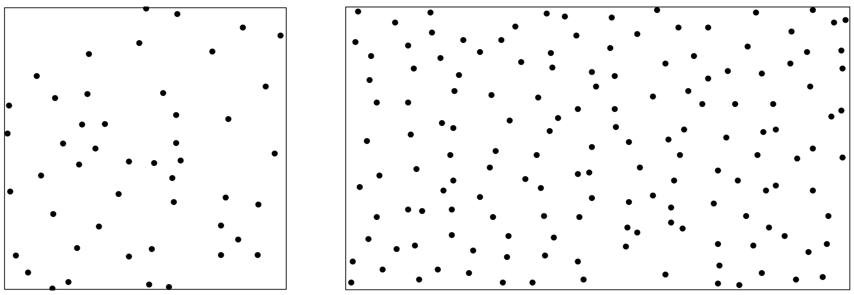

Fig. 7. Left: simulated data. Right: amacrine data.

The top right plot in Fig. 8 shows that the posterior density of r_1 is close to the prior density $\pi(r_1) \propto \exp(-\kappa r_1)$ $(0 < r_1 < r_{\max})$ at least for small and modest values of r_1. The posterior and prior mean of r_1 are 0.026, and 0.027, which are both close to the hard core $\sigma_{DG} = 0.025$.

The bottom left plot in Fig. 8 shows a pronounced difference between the posterior density of r_p and the prior density $\pi(r_p) \propto \exp(-\kappa(r_{\max} - r_p))$ $(0 < r_p < r_{\max})$. The posterior and prior mean of r_p are 0.087 and 0.113, i.e. they are equally far from the true interaction range $\rho_{DG} = 0.1$.

The bottom right plot in Fig. 8 shows that the prior and the posterior distribution for p are rather close. The posterior and prior mean are 5.12 and 5.

As mentioned we have for comparison also used non-perfect simulation when estimating ratios of normalising constants. With a burn-in of 250 iterations the results are very close to those in Fig. 8. This is illustrated in Fig. 9.

Example 5: For the amacrine data we let $(k, m) = (10, 2500)$, $r_{\max} = 120$, $\kappa r_{\max} = 5$, $\sigma^2 = 1$, $\beta_{\min} = 0.005$, $\beta_{\max} = 0.05$, and $\epsilon = 0.0005$. As mentioned we have to use non-perfect simulation when estimating ratios of normalising constants; for this we use a burn-in of 2500 iterations. For instance, when simulating a multi-scale process which approximates the estimated Diggle-Gratto, the Metropolis-Hastings chain seems to be stabilised after 1500 iterations when started in a realization from a homogeneous Poisson process on $[0, 1070] \times [0, 600]$ with mean number of points equal to 152.

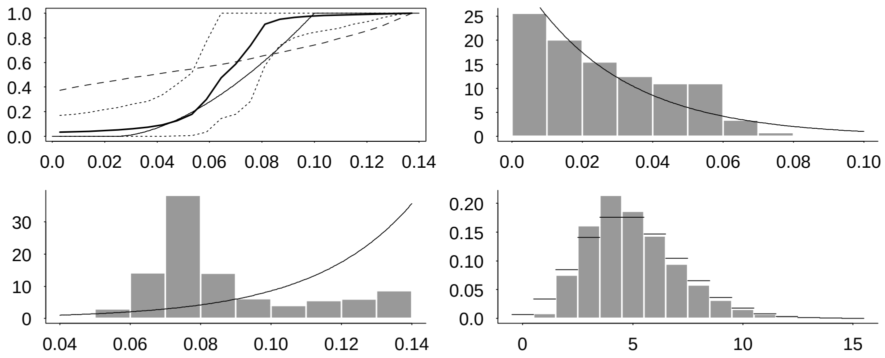

Fig. 8. Results for the simulated data. Top left: posterior mean (thick line) together with a 95% credibility interval (dotted lines), prior mean (dashed line), and true interaction function (thin line). Top right, bottom left, and bottom right: posterior and prior density (full line) of r_1, r_p, and p.

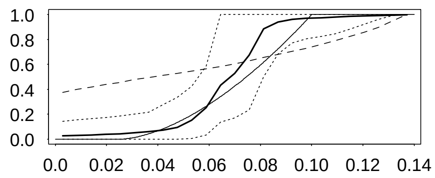

Fig. 9. Same as the top left plot in Fig. 8, but using non-perfect simulation for path sampling.

For the algorithm for posterior simulations, time series plots (not shown here) for different statistics show that the algorithm converges quickly into equilibrium (using a similar initial state as in Example 4). The mixing properties of this algorithm are sensitive to the choice of σ; it is slowly mixing for $\sigma = 0.1$, while it is faster mixing for $\sigma = 1$ and $\sigma = 10$. The posterior results for $\sigma = 10$ are very close to the following results for $\sigma = 1$.

Fig. 10 is similar to Fig. 8 but for the amacrine data and 250000 posterior updates.

In the top left plot in Fig. 10, since we do not know the truth, we have shown the estimated Diggle-Gratton interaction function instead. Again there is a clear difference between the posterior and prior mean of $\varphi(r)$. The plot may be compared with Fig. 3 in Heikkinen & Penttinen (1999) which shows their maximum a posteriori estimate of φ and several other estimated interaction functions from the literature. Most but not all of these estimated interaction functions are within the 95% credibility interval in Fig. 10; the maximum a posteriori estimate is close to $\varphi_{DG}(r)$ when $r \leq 76$, and larger than 1 for $r > 76$.

The top right, lower left, and lower right plots in Fig. 10 are very different from those in Fig. 8. Now the posterior distribution of r_1 is rather flat for $r_1 \leq 20$ and has a mode for $r_1 > 20$, while the prior density for r_1 is strictly deceasing. The lower left plot in Fig. 10 shows a large dispersion in the posterior distribution for r_p, indicating difficulties in estimating the range of interaction; the value $\rho_{DG} = 76$ obtained by Diggle & Gratton (1984) is in the centre of the posterior distribution. The lower right plot in Fig. 10 shows clearly that the posterior and prior distribution for p are different. Compared to the plot in Fig. 8, we need now a larger number of change-points due to the increased amount of data, but the number needed is much smaller compared to the 30 change-points used in Fig. 3 in Heikkinen & Penttinen (1999). The posterior and prior means of p are 6.62 and 5.

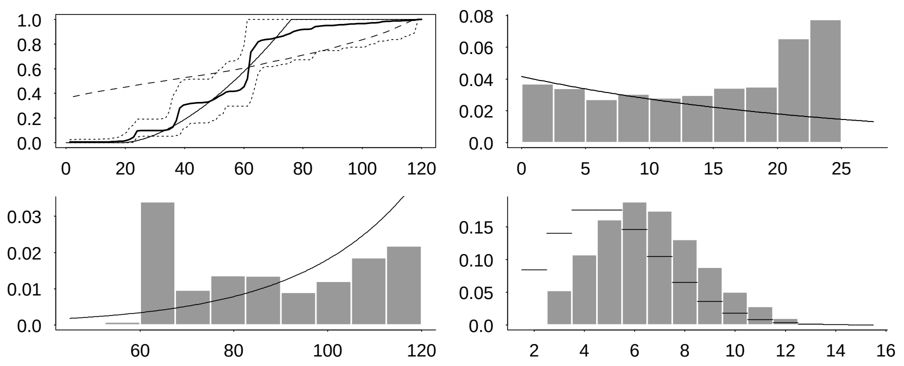

Fig. 10. As Fig. 8 but for the amacrine data. Now the thin line in the top left plot shows the estimated interaction function by Diggle & Gratton (1984).

Finally, Fig. 11 shows the posterior distribution of $\beta \times 1070 \times 600$, with a mode around 4000, a mean of about 4589, and a large dispersion. In order to obtain about 152 points in average, simulation of the estimated Diggle-Gratton model requires $\beta \times 1070 \times 600 \approx$

6500, which is larger than the posterior mean but still not in the tail of the posterior distribution.

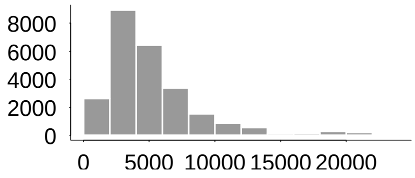

Fig. 11. Posterior distribution of $\beta \times 1070 \times 600$.

Acknowledgement

This research will be a part of KKB's PhD dissertation. JM was supported by the European Union's research network "Statistical and Computational Methods for the Analysis of Spatial Data, ERB-FMRX-CT96-0096", by the Centre for Mathematical Physics and Stochastics (MaPhySto), funded by a grant from the Danish National Research Foundation, and by the Danish Natural Science Research Council.

References

Aitchison, J. (1986). *The statistical analysis of compositional data.* Monographs on Statistics and Applied Probability. Chapman and Hall, London.

Arjas, E. & Gasbarra, D. (1994). Nonparametric bayesian inference from right censored survival data, using the Gibbs sampler. *Statist. Sinica* **4**, 505–524.

Baddeley, A. & Turner, R. (2000). Practical maximum pseudolikelihood for spatial point patterns. *Australian and New Zealand Journal of Statistics* **42**, 283–322.

Berthelsen, K. K. & Møller, J. (2001). A primer on perfect simulation for spatial point processes. Submitted.

Diggle, P. & Gratton, R. (1984). Monte Carlo methods of inference for implicit statistical model (with discussion). *J. Roy. Statist. Soc. Ser. B* **46**, 193–227.

Gates, D. J. & Westcott, M. (1986). Clustering estimates for spatial point distributions with unstable potentials. *Ann. Inst. Statis. Math.* **A 38**, 123–135.

Gelman, A. & Meng, X.-L. (1998). Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statist. Sci.* **13**, 163–185.

Geyer, C. J. & Møller, J. (1994). Simulation procedures and likelihood inference for spatial point processes. *Scand. J. Statist.* **21**, 359–373.

Heikkinen, J. & Penttinen, A. (1999). Bayesian smoothing in the eastimation of the pair potential function of Gibbs point processes. *Bernoulli* **5**, 1119–1136.

Kelly, F. P. & Ripley, B. D. (1976). A note on Strausss model for clustering. *Biometrika* **63**, 357–360.

Kendall, W. S. (1998). Perfect simulation for the area-interaction point process. In *Probability Towards 2000* (eds L. Accardi & C. Heyde). Springer, New York, 218–234.

Kendall, W. S. & Møller, J. (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Adv. Appl. Prob.* **32**, 844–865.

Lieshout, M. N. M. van & Baddeley, A. J. (1996). A nonparametric measure of spatial interaction in point patterns. *Statist. Neerlandica* **50**, 344–361.

Møller, J. (1999). Markov chain Monte Carlo and spatial point processes. In *Stochastic geometry, likelihood and computation* (eds O. E. Barndorff-Nielsen, W. S. Kendall & M. N. M. van Lieshout). Chapman & Hall, 141–172.

Møller, J. (2001). A review of perfect simulation in stochastic geometry. In *Selected proceedings of the symposium on inference for stochastic processes* (eds C. C. H. I. V. Basawa & R. L. Taylor), Vol. 37. IMS Lecture Notes & Monographs Series, 333–355.

Penttinen, A. (1984). *Modelling interaction in spatial point patterns: parameter estimation by the maximum likelihood method.* Number 7 in Jyväskylä Studies in Computer Science, Economics, and Statistics.

Propp, J. G. & Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms* **9**, 223–252.

Ripley, B. D. (1988). *Statistical inference for spatial processes.* Cambridge University Press, Cambridge.

Ripley, B. D. (1989). Gibbsian interaction models. In *Spatial statistics: past, present and future* (ed D. A. Griffiths). Image, New York, 1–19.

Strauss, D. J. (1975). A model for clustering. *Biometrika* **62**, 467–475.

Wilson, D. B. (2000). Layered multishift coupling for use in perfect sampling algorithms (with a primer on CFTP). In *Monte Carlo methods* (ed N. Madras), Vol. 26 of *Fields Institute Communications.* Amer. Stat. Soc., 141–176.