Bayesian Inference for Pairwise Interacting Point Processes

Matthew A. Bognar
and
Mary Kathryn Cowles

December 2002

Technical Report No. 319

Bayesian inference for pairwise interacting point processes

Matthew A. Bognar*

Department of Statistics and Actuarial Science

University of Iowa, Iowa City, IA 52242, USA

and

Mary Kathryn Cowles†

Department of Statistics and Actuarial Science, Department of Biostatistics

University of Iowa, Iowa City, IA 52242, USA

SUMMARY. Pairwise interacting point processes are commonly used to model spatial point patterns. To perform inference, the established frequentist methods can produce good point estimates when the interaction in the data is moderate, but some methods may produce severely biased estimates when there is strong interaction present in the data. Furthermore, because the sampling distributions of the estimates are unclear, interval estimates are typically obtained by parametric bootstrap methods. In the current setting however, the behavior of such estimates is not well understood. In this article we propose Bayesian methods for obtaining inferences in a pairwise interacting point process. The requisite application of Markov chain Monte Carlo (MCMC) techniques is complicated by an intractable function of the

* *email:* mbognar@stat.uiowa.edu

† *email:* kcowles@stat.uiowa.edu

parameters in the likelihood. The acceptance probability in a Metropolis-Hastings algorithm involves the ratio of two likelihoods evaluated at differing parameter values. The intractable functions do not cancel, and hence an intractable ratio r must be estimated within each iteration of a Metropolis-Hastings sampler. We propose the use of importance sampling techniques within MCMC to address this problem. While r may be estimated by other methods, these, in general, are not readily applied in a Bayesian setting. We demonstrate the validity of our importance sampling approach with a small simulation study. Finally, we analyze the Swedish pine sapling dataset (Strand, 1972) and contrast the results with those in the literature.

KEY WORDS: Bayesian estimation; Metropolis-Hastings algorithm; pairwise interacting point process; grid-based importance sampling; spatial point pattern; Swedish pine sapling data

1. Introduction

The spatial locations of events in some designated region constitute a "spatial point pattern". Figure 1 depicts a well-known example, data collected by Strand (1972) on the locations of 71 Swedish pine saplings in a 10×10 meter plot.

[Figure 1 about here.]

In comparison with a binomial (completely random) process with the same number of points (Figure 2), the sapling data has far fewer pairs of points that are close together.

[Figure 2 about here.]

The more regular spacing of the saplings suggests biological competition for water, light, canopy space, or adequate soil for root growth.

Pairwise interacting point processes (PIPPs) form a very popular and flexible class of models for spatial point patterns in a bounded region. In a PIPP, the interaction (attraction or inhibition) between two points is described by a *pair potential function* – typically a function of the inter-point Euclidean distance. Estimation in PIPPs is notoriously difficult because the likelihood contains an analytically intractable function of the parameters. Because PIPPs are not optimal to model spatial point patterns with spatial attraction (Heikkinen and Penttinen, 1999), we focus on estimation in regular (inhibitory) patterns.

2. Pairwise interacting point processes

Suppose a point pattern $\mathbf{x} = \{x_i : i = 1, \ldots, n\}$ is observed in some bounded region V where $x_i \in V \ \forall i$ (we assume without loss of generality that V is a subset of two-dimensional Euclidean space). Consider a family of pair potential functions $\{\phi_\theta(s) : \theta \in \Theta\}$, indexed by a parameter vector $\theta = (\theta_1, \ldots, \theta_m)$, of Euclidean distance s. A common pair potential function is (Strauss, 1975)

$$\phi_\theta(s) = \begin{cases} h & s \leq b \\ 0 & s > b \end{cases} \tag{1}$$

where $\theta = (b, h)$. In the literature, b is known as the *interaction distance* and h is termed the *Straussian* parameter. The Straussian parameter describes the strength of inhibition (or attraction), and the interaction distance indicates the distance at which pairs of points cease to interact. Patterns realized under a PIPP in which

$\phi_\theta(\cdot) > 0$ tend to exhibit spatial regularity; the inter-point distance tends to be larger than would be expected under a binomial process (i.e. when $\phi_\theta(\cdot) \equiv 0$).

As in Degenhardt (1999), Diggle et al. (1994), Ogata and Tanemura (1981), and Stoyan and Penttinen (2000), we will condition on n because estimation of the *chemical activity* parameter (which regulates the number of points in the pattern) is not of primary concern. Therefore, the probability density function, conditional on n and θ, is

$$\begin{aligned} p(\mathbf{x}|\theta) &= \frac{\exp\left\{-\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\phi_\theta\left(\|x_i - x_j\|\right)\right\}}{Z_n(\theta)} \\ &\stackrel{\text{def}}{=} \frac{g(\mathbf{x}|\theta)}{Z_n(\theta)} \end{aligned}$$

where $\|\cdot\|$ denotes Euclidean distance and

$$Z_n(\theta) \stackrel{\text{def}}{=} \int_{V^n} \exp\left\{-\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\phi_\theta\left(\|x_i - x_j\|\right)\right\} dx_1 \ldots dx_n$$

is a normalization constant which depends on θ. For simplicity, we notationally suppress the conditioning on n.

When $p(\mathbf{x}|\theta)$ is regarded as a probability density function (i.e. $p(\mathbf{x}|\theta)$ is a function of $\mathbf{x}$), then $Z_n(\theta)$ is a normalizing *constant*. On the other hand, when $p(\mathbf{x}|\theta)$ is regarded as a likelihood, say

$$L(\theta) \stackrel{\text{def}}{=} \frac{g(\mathbf{x}|\theta)}{Z_n(\theta)},$$

then $Z_n(\theta)$ is not a constant, but a function of the parameter vector θ. In a PIPP, analytical computation of $Z_n(\theta)$ is generally impossible. Thus, to perform (approximate) likelihood based inferences, $Z_n(\theta)$ must be estimated for different candidate

values of θ. Heikkinen and Penttinen (1999) observe that with conditioning on n, $Z_n(\theta)$ is finite for all pair potential functions $\phi : (0, \infty) \to \{\infty\} \cup \Re$.

3. Construction of the Metropolis-Hastings algorithm

Suppose a spatial point pattern $\mathbf{x}$ is observed in V. We will let $p(\theta)$ denote the prior distribution on $\theta = (\theta_1, \ldots, \theta_m)$, $L(\theta) = g(\mathbf{x}|\theta)/Z_n(\theta)$ will denote the likelihood, and $p(\theta|\mathbf{x})$ will denote the posterior distribution of θ given $\mathbf{x}$. The Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970) enables sampling from $p(\theta|\mathbf{x})$. The systematic scan algorithm proceeds as follows:

1. Choose an arbitrary starting value for θ, say $\theta^{(t)} = (\theta_1^{(t)}, \ldots, \theta_m^{(t)})$ where $t = 0$. For simplicity we update $\theta_1^{(\cdot)}$, then $\theta_2^{(\cdot)}$, and so on. When all m components have been updated, we repeat the process.

2. Suppose the first $i-1$ components have been updated and we now wish to update $\theta_i^{(t)}$. Letting $\theta^{\text{curr}} = (\theta_1^{(t+1)}, \ldots, \theta_{i-1}^{(t+1)}, \theta_i^{(t)}, \theta_{i+1}^{(t)}, \ldots, \theta_m^{(t)})$ denote the current parameter vector, generate a candidate value θ_i^* from a proposal distribution $q_i(\theta_i^*|\theta^{\text{curr}})$. Let $\theta^{\text{prop}} = (\theta_1^{(t+1)}, \ldots, \theta_{i-1}^{(t+1)}, \theta_i^*, \theta_{i+1}^{(t)}, \ldots, \theta_m^{(t)})$ denote the proposed parameter vector.

3. Accept the candidate value θ_i^*, that is let $\theta_i^{(t+1)} = \theta_i^*$, with probability

$$\begin{aligned}
\alpha &= \min\left[1, \frac{p(\theta^{\text{prop}}|\mathbf{x}) q_i(\theta_i^{(t)}|\theta^{\text{prop}})}{p(\theta^{\text{curr}}|\mathbf{x}) q_i(\theta_i^*|\theta^{\text{curr}})}\right] \\
&= \min\left[1, \frac{L(\theta^{\text{prop}}) p(\theta^{\text{prop}}) q_i(\theta_i^{(t)}|\theta^{\text{prop}})}{L(\theta^{\text{curr}}) p(\theta^{\text{curr}}) q_i(\theta_i^*|\theta^{\text{curr}})}\right] \\
&= \min\left[1, \frac{g(\mathbf{x}|\theta^{\text{prop}}) Z_n(\theta^{\text{curr}}) p(\theta^{\text{prop}}) q_i(\theta_i^{(t)}|\theta^{\text{prop}})}{g(\mathbf{x}|\theta^{\text{curr}}) Z_n(\theta^{\text{prop}}) p(\theta^{\text{curr}}) q_i(\theta_i^*|\theta^{\text{curr}})}\right] \qquad (2)
\end{aligned}$$

otherwise reject the candidate value, i.e. let $\theta_i^{(t+1)} = \theta_i^{(t)}$.

4. Jump to step 2. Repeat mT times to yield T iterations.

Because the intractable $Z_n(\theta^{\text{curr}})$ and $Z_n(\theta^{\text{prop}})$ do *not* cancel in the acceptance probability, they must be approximated within *every* iteration of the sampler. Therefore, the Metropolis-Hastings algorithm will be useful for PIPPs only if the intractable ratio

$$r \stackrel{\text{def}}{=} \frac{Z_n(\theta^{\text{curr}})}{Z_n(\theta^{\text{prop}})}$$

in (2) can be estimated accurately and efficiently.

4. Estimation of an intractable ratio via grid-based importance sampling

Importance sampling (Smith and Gelfand, 1992) allows the approximation of integrals as follows. Suppose, for illustration, it is possible to obtain realizations, say $y_1, \ldots, y_T$, from some univariate density $v(y)$. We are able to approximate an integral, say γ, with respect to some density f having the same support as v by

$$\gamma = \int h(y) f(y)\, dy \simeq \frac{1}{T} \sum_{t=1}^{T} h(y_t) f(y_t)/v(y_t) = \hat{\gamma}. \tag{3}$$

The ratio $f(y_t)/v(y_t)$ is referred to as the importance sampling weight. If $y_1, \ldots, y_T$ are obtained by MCMC, and the chain is ergodic, then $\hat{\gamma} \stackrel{\text{a.s.}}{\to} \gamma$. Higher order integrals can be estimated in a similar manner. If the importance sampling density v is too dissimilar from the target density f, then the importance sampling weights will be highly variable, increasing the number of samples required to approximate γ with any given degree of accuracy.

4.1 *Estimating r via importance sampling*

The key to implementing importance sampling is the ability to generate point patterns from the probability density function $p(\mathbf{x}|\theta')$ given any parameter vector θ'. Ogata and Tanemura (1981) and Geyer and Moller (1994) developed Metropolis-Hastings algorithms that can do this. The realizations, say $\mathbf{x}_1, \ldots, \mathbf{x}_T$, can be obtained *without* knowledge of the normalizing constant $Z_n(\theta')$. Letting $h \equiv 1$, $f = g(\mathbf{x}|\theta)$, and $v = p(\mathbf{x}|\theta')$, we can approximate $Z_n(\theta) = \int_{V^n} g(\mathbf{x}|\theta)\, d\mathbf{x}$ in (2) by

$$\frac{1}{T}\sum_{t=1}^{T} \frac{g(\mathbf{x}_t|\theta)}{p(\mathbf{x}_t|\theta')} = \frac{Z_n(\theta')}{T}\sum_{t=1}^{T} \frac{g(\mathbf{x}_t|\theta)}{g(\mathbf{x}_t|\theta')}.$$

This estimator can not be applied since $Z_n(\theta')$ is unknown. However, $Z_n(\theta')$ can be estimated by

$$\frac{|V|^n}{\frac{1}{T}\sum_{t=1}^{T} 1/g(\mathbf{x}_t|\theta')} \xrightarrow{\text{a.s.}} \frac{|V|^n}{\int_{V^n} 1/g(\mathbf{x}|\theta')p(\mathbf{x}|\theta')\, d\mathbf{x}} = \frac{Z_n(\theta')|V|^n}{\int_{V^n} 1\, d\mathbf{x}} = Z_n(\theta') \tag{4}$$

where $|V|$ is the area of V. Therefore, an estimate of $Z_n(\theta)$ is

$$\hat{Z}_n(\theta) = \frac{|V|^n}{\sum_{t=1}^{T} 1/g(\mathbf{x}_t|\theta')}\sum_{t=1}^{T} \frac{g(\mathbf{x}_t|\theta)}{g(\mathbf{x}_t|\theta')}. \tag{5}$$

To estimate the intractable ratio r in (2), compute

$$\frac{\hat{Z}_n(\theta^{\text{curr}})}{\hat{Z}_n(\theta^{\text{prop}})} = \frac{\sum_{t=1}^{T} g(\mathbf{x}_t|\theta^{\text{curr}})/g(\mathbf{x}_t|\theta')}{\sum_{t=1}^{T} g(\mathbf{x}_t|\theta^{\text{prop}})/g(\mathbf{x}_t|\theta')}. \tag{6}$$

Note the cancellation of the $|V|^n/\sum_{t=1}^{T} 1/g(\mathbf{x}_t|\theta')$ terms.

4.2 *Grid-based importance sampling*

To use the Metropolis-Hastings algorithm to fit a pairwise interaction model we must be able to accurately estimate r for *any* values of θ^{curr} and θ^{prop} supported in the posterior. It is, however, unlikely that any one importance sampling density $p(\mathbf{x}|\theta')$ will be similar to $p(\mathbf{x}|\theta^{\text{curr}})$ and $p(\mathbf{x}|\theta^{\text{prop}})$ for all θ^{curr} and θ^{prop}. If $p(\mathbf{x}|\theta^{\text{curr}})$ or $p(\mathbf{x}|\theta^{\text{prop}})$ or both are dissimilar to $p(\mathbf{x}|\theta')$, then the estimator (6) is compromised.

Consider a saturation of locations, say $\theta'_1, \ldots, \theta'_l$, which form a grid over the parameter space. Suppose at each location θ'_i, importance samples are realized from the importance sampling density $p(\mathbf{x}|\theta'_i)$. Now, regardless of θ^{curr} and θ^{prop}, we are able to choose an importance sampling density, say $p(\mathbf{x}|\theta'_i)$, which is similar to $p(\mathbf{x}|\theta^{\text{curr}})$ and $p(\mathbf{x}|\theta^{\text{prop}})$. With this well chosen importance sampling density, we can accurately estimate r. While this grid-based technique is quite simple, it is very efficient since the simulations from the l importance sampling densities need to be generated only *once*. The same simulations can be used to estimate r for many θ^{curr} and θ^{prop}.

If θ'_i is the closest importance sampling location to θ^{curr}, and if $\mathbf{x}_1, \ldots, \mathbf{x}_T$ are realized (after burn-in) from the importance sampling density $p(\mathbf{x}|\theta'_i) = g(\mathbf{x}|\theta'_i)/Z_n(\theta'_i)$, then we can estimate r by

$$\hat{r} = \frac{\sum_{t=1}^{T} g(\mathbf{x}_t|\theta^{\text{curr}})/g(\mathbf{x}_t|\theta'_i)}{\sum_{t=1}^{T} g(\mathbf{x}_t|\theta^{\text{prop}})/g(\mathbf{x}_t|\theta'_i)} \xrightarrow{\text{a.s.}} r. \tag{7}$$

In order for the same importance sampling density to be close to both $p(\mathbf{x}|\theta^{curr})$ and $p(\mathbf{x}|\theta^{prop})$, the proposal distribution must propose candidates θ^{prop} that are close to θ^{curr}. Thus, the necessity of accurately estimating r requires that the sampler

move in small steps. Taking large steps causes sampler instability, and has, in some instances, precluded convergence.

It is possible to estimate $Z_n(\theta^{\text{curr}})$ and $Z_n(\theta^{\text{prop}})$ separately via (5) using different importance sampling densities and to obtain an estimate of r by taking the ratio of the two estimates (note that with different importance sampling densities, there is not a cancellation as in (6)). However, estimating r via (7) significantly reduces the variability of the estimates of r. In addition, our grid-based importance sampling estimator (7) yields a notable improvement in the stability and mixing behavior of the Metropolis-Hastings sampler.

Ratio importance sampling (Chen and Shao, 1997), originally described in a non-spatial context, is essentially the same as our grid-based technique without the construction of a grid. Because new importance samples must be generated at every iteration of the overall Metropolis-Hastings sampler, (optimal) ratio importance sampling is computationally costly. However, both techniques are applicable and produce comparable inferences.

5. Choosing the grid coarseness and proposal densities

We now examine 1) how close θ^{curr} and θ^{prop} must be to the grid location θ_i', and 2) how close the proposal θ^{prop} must be to θ^{curr} to get accurate estimates of r. For illustration, we considered patterns with $n = 50$ points in the unit square. We use the Straussian model (1), but with b fixed at 0.10 (thus $\theta = h$). For each θ^{curr} and θ^{prop} listed in Table 1, we estimated r by generating 50,000 unthinned point patterns (after a 5,000 iteration burn-in) from $p(\mathbf{x}|\theta_i' = 1.0)$ and computing (7). The estimation procedure was repeated 100 times, enabling the coefficient of variation

(CV) of r to be approximated.

As expected, the more distant θ^{prop} is from θ^{curr} (see the left side of Table 1), the larger the CV of r.

[Table 1 about here.]

Thus, the proposal density should suggest candidates θ^{prop} which are close to θ^{curr}. From the right side of Table 1, we see that a distant importance sampling location from θ^{curr} and θ^{prop} also causes an inflated CV. Here, we can avoid undue variability in the estimate of r by constructing an appropriately fine grid.

To better understand how to choose the grid and proposal densities for the interaction distance, we fixed $h = 1.0$ and estimated the CV for different current and proposed values of the interaction distance $\theta = b$. Here the importance samples were generated from $p(\mathbf{x}|\theta_i' = 0.10)$. Table 2 demonstrates that θ^{prop} must be close to θ^{curr}, and, as before, a close importance sampling location to θ^{curr} and θ^{prop} is essential to prevent undue variability in the estimates of r.

[Table 2 about here.]

To conclude, caution must be taken in the construction of the grid and proposal densities. While general guidelines are difficult to obtain, we hope to motivate the need for proper attention and care.

6. Simulation study: Comparing the frequentist performance of Bayesian and frequentist estimates of a Straussian parameter

We evaluated the performance of our MCMC method by conducting a simulation study using the simplest possible PIPP: a Straussian model with interaction distance

b fixed so that the only unknown parameter θ is the interaction (repulsion) strength h in (1).

6.1 *Setup*

We simulated 100 point patterns of size $n = 50$ in the unit square from each of 10 Straussian models, all with b fixed at 0.15, but with θ set at different values ranging from 0.3 to 2.5.

Before running the Metropolis-Hastings samplers to fit models to the simulated datasets, we generated the needed importance samples. We constructed a grid $\theta'_1, \ldots, \theta'_{10}$ over the parameter space with grid points at intervals of 0.5 from $\theta'_1 = -1.0$ to $\theta'_{10} = 3.5$. We generated 1,000 point patterns from each importance sampling density $p(\mathbf{x}|\theta'_i)$ $i = 1, \ldots, 10$ by running Geyer and Moller's algorithm for 1,000,000 iterations (after burn-in) and retaining every 1,000$^{\text{th}}$ point pattern. Thus the 1,000 importance samples for each θ'_i were approximately independent and contained maximum information. Storage considerations unfortunately limited the possibility of using all 1,000,000 importance samples for estimating r.

We placed a normal prior on θ, $p(\theta) \sim \mathrm{N}(\mu = 0, \sigma^2 = 100)$, which allowed the MCMC sampler to explore spatial repulsion and, if dictated by the posterior, spatial attraction. The large variance ensured only a small bias toward 0 in the posterior means.

In the Metropolis-Hastings samplers for fitting the models, the importance samples from $p(\mathbf{x}|\theta'_i)$ with θ'_i closest to θ^{curr} were used to estimate r. The proposed height θ^* was chosen uniformly in the interval $(\theta^{(t)} - 0.5, \theta^{(t)} + 0.5)$; thus the proposal ratio equals unity. All samplers were started at $\theta^{(0)} \equiv 0$. Geweke's test for convergence indicated no evidence that the chain had failed to converge by the 1,000$^{\text{th}}$ iteration.

For each simulated pattern, the sampler was run for 2,000 iterations following a 1,000 iteration burn-in.

6.2 *Results*

We used the Metropolis-Hastings output to estimate the posterior mean of θ for each simulated dataset. The estimated bias and mean squared error are displayed in Table 3.

[Table 3 about here.]

The bias, in the frequentist sense, is quite negligible for any strength of repulsion. We obtained similar results when estimating the interaction distance b in (1), as well as when estimating b and h simultaneously (results not shown).

To evaluate coverage, five hundred patterns were simulated from $p(\mathbf{x}|\theta = 1.0)$. Table 4 displays the coverage and average width of the 95% highest posterior density (HPD) and equal tail intervals, bias, and mean squared error estimated from the simulated patterns.

[Table 4 about here.]

For comparison, two different proposals were examined: choose θ^* uniformly in either $(\theta^{(t)} - 1.0, \theta^{(t)} + 1.0)$ or $(\theta^{(t)} - 0.5, \theta^{(t)} + 0.5)$.

Diggle et al. (1994) performed a similar simulation study from a frequentist point of view – only in the number of points ($n = 100$) and interaction distance ($\theta = 0.10$) does their study differ from ours. They investigated the approximate maximum likelihood (via virial expansions) (Ripley, 1988), maximum pseudo-likelihood (Ripley, 1988), and Takacs-Fiksel methods (Fiksel, 1984; Fiksel, 1988; Takacs, 1986). The

approximate maximum likelihood method significantly under-estimated the amount of repulsion for strongly repulsive patterns, and over-estimated the amount of repulsion in cases of weak interaction. The maximum pseudo-likelihood method either under or over-estimated (depending on edge correction) the amount of repulsion in patterns with strong inhibition, while the Takacs-Fiksel method generally yielded unbiased estimates, but produced more outlying point estimates than the other methods. Diggle et al. concluded that the Takacs-Fiksel method performs poorly in cases of weak interaction under other (i.e. non-Strauss) pair potential functions.

7. Analysis of the Swedish pine sapling data

Recall that Figure 1 displays the location of $n = 71$ Swedish pine saplings in a 10×10 meter region V. We assume that the pine saplings constitute a realization of a Strauss point process. Therefore, the likelihood is

$$L(\theta) = \frac{\exp\left\{-\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\phi_\theta\left(\|x_i - x_j\|\right)\right\}}{Z_n(\theta)}$$

where $\phi_\theta(\cdot)$ is of the form (1). We will obtain estimates, both point and interval, from the full posterior distribution under diffuse priors: $b \sim \text{Unif}(0, 2.5)$ and $h \sim \text{N}(2, \sigma^2 = 100)$ (independent). We constructed a grid $\theta'_1, \ldots, \theta'_{77}$ over the two-dimensional parameter space with grid points at intervals of 0.25 from 0 to 2.5 for b and at intervals of 1.0 from 0 to 5.0 for h (including -0.5). We generated 1,000,000 point patterns (after burn-in) from each importance sampling density $p(\mathbf{x}|\theta'_i)$ $i = 1, \ldots, 77$ and retained every 1,000$^{\text{th}}$ point pattern.

7.1 *Sampler details*

In the Metropolis-Hastings sampler, suppose the current and proposed parameter vectors are $\theta^{\text{curr}} = (b^{\text{curr}}, h^{\text{curr}})$ and $\theta^{\text{prop}} = (b^{\text{prop}}, h^{\text{prop}})$ respectively. We tried three different ways of choosing which set of importance samples to use to estimate r: 1) the θ_i' which minimized $||\theta^{\text{curr}} - \theta_i'|| = \sqrt{(b^{\text{curr}} - b_i')^2 + (h^{\text{curr}} - h_i')^2}$; 2) the θ_i' which minimized the average distance from θ^{curr} and θ^{prop} to θ_i'; i.e. $(||\theta^{\text{curr}} - \theta_i'|| + ||\theta^{\text{prop}} - \theta_i'||)/2$; and 3) the θ_i' which minimized the average *squared* distance from θ^{curr} and θ^{prop} to θ_i' – a significant penalty for a distant θ^{curr} or θ^{prop} is levied.

A systematic scan Metropolis-Hastings sampler updated the parameter vector by alternating between a b-move and an h-move. Suppose at the beginning of iteration t, $\theta^{\text{curr}} = (b^{(t)}, h^{(t)})$. For a b-move, a new interaction distance b^* was proposed uniformly in the interval $(b^{(t)} - 0.2, b^{(t)} + 0.2)$; thus $\theta^{\text{prop}} = (b^*, h^{(t)})$ and the proposal ratio equals 1. Note that if $b^* \notin (0, 2.5)$, then the candidate was automatically rejected (via the prior) and $b^{(t+1)} = b^{(t)}$. For the subsequent h-move, when $\theta^{curr} = (b^{(t+1)}, h^{(t)})$, h^* was chosen uniformly in the interval $(h^{(t)} - 0.5, h^{(t)} + 0.5)$. Thus $\theta^{\text{prop}} = (b^{(t+1)}, h^*)$ and the proposal ratio equals 1. The proposals were chosen to balance good Metropolis-Hastings acceptance rates (Gilks et al., 1996) while, at the same time, not proposing too large of change to the current state. The Metropolis-Hastings acceptance rates for a b-move and an h-move were 43.4% and 66.1% respectively.

We computed Gelman and Rubin's convergence diagnostic using the output from three chains with differing initial values: $\theta^{(0)} = (1.0, 0.5)$, $\theta^{(0)} = (1.0, 3.0)$, and $\theta^{(0)} = (2.0, 1.0)$. After 1,000 iterations, the potential scale reduction factors were 1.015696 and 1.001861 for b and h respectively. Because there was no evidence of non-convergence after 1,000 iterations, Metropolis-Hastings sampler burn-in was

taken as such. The time to obtain 25,000 post burn-in iterations was approximately 5 minutes on an HP Visualize B2000 workstation.

7.2 *Results*

The estimated marginal posterior distribution of the interaction distance b is displayed in Figure 3; separate estimates corresponding to the method of choosing the importance sampling density $p(\mathbf{x}|\theta_i')$ are displayed.

[Figure 3 about here.]

The choice of importance sampling density does not have a large impact. The marginal posterior distribution of b is surprisingly complex, with strong bi-modality and numerous local modes. The jagged nature of the plot is apparently due to $\phi_\theta(\cdot)$, and hence $L(\theta)$, being a discontinuous function. The posterior mean of b, when choosing the θ_i' closest to θ^{curr}, is 0.817 meters with Monte Carlo error of 0.002. The 95% HPD interval for b is $(0.654, 1.004)$ meters, and the posterior mode is approximately 0.757. To conclude, saplings at a distance of more than approximately 0.8 meters cease to interact with one another.

Figure 4 displays the marginal posterior density estimate of the Straussian parameter h.

[Figure 4 about here.]

The posterior distribution indicates that the saplings show *moderate* inhibition, which is expressed by having *some* trees within a distance of b. The posterior mean of h is 1.153 with Monte Carlo error of 0.0066. The 95% HPD interval for h is $(0.501, 1.821)$, and the posterior mode is 1.037. The posterior probability that the

pine saplings exhibit inhibition is $P(h > 0|\mathbf{x}) \simeq 1$. Similar inferences for b and h were obtained when substituting Chen and Shao's optimal ratio importance sampling estimator of r for our grid-based importance sampling estimator.

Venables and Ripley (1994) obtained maximum pseudo-likelihood estimates for b and h of 0.7 and 1.897 respectively. In an earlier analysis, Ripley (1981) obtained an estimate of h equal to 1.609. Without conditioning on n, Baddeley and Turner (2000) obtained maximum pseudo-likelihood estimates (depending on quadrature scheme and type of edge correction) for h ranging from 1.234 to 1.609 while holding b fixed at 0.7 meters. Their estimate of b was obtained by maximizing the profile log-pseudo-likelihood.

Baddeley and Turner obtained interval estimates by initially running a Metropolis-Hastings birth-death-shift algorithm (Geyer and Moller, 1994) to generate 500 realizations $\hat{\theta}_1, \ldots, \hat{\theta}_{500}$ from the distribution of the maximum pseudo-likelihood estimator under the fitted model; namely, the unconditional Strauss process with $b = 0.7$ and $h = 1.546$. From the simulations, the mean vector and covariance matrix were estimated. Normal-based 95% confidence intervals were $(0.627, 0.773)$ for b and $(0.998, 2.856)$ for h (the normality of the estimator of b is suspect). Confidence intervals based on the empirical quantiles of the parametric bootstrap procedure were $(0.62, 0.81)$ for b and $(0.942, 2.408)$ for h.

8. Some thoughts on convergence

It is not clear whether a Metropolis-Hastings sampler with a stochastically approximated acceptance probability will converge to the correct target distribution – however, analytical verification is difficult. While our simulation study shows that good

point estimates can be obtained, we sought verification in our real data example, for which, of course, the true parameter values are unknown.

To verify that our grid-based algorithm is converging to the correct distribution, we will utilize a deterministic approximation of $Z_n(\theta)$, denoted by $\tilde{Z}_n(\theta)$, in the sampler. Now, noting that

$$p(\theta|\mathbf{x}) \propto \frac{g(\mathbf{x}|\theta)}{Z_n(\theta)}p(\theta) \simeq \frac{g(\mathbf{x}|\theta)}{Z_n(\theta)}\frac{Z_n(\theta)}{\tilde{Z}_n(\theta)}p(\theta) = \frac{g(\mathbf{x}|\theta)}{\tilde{Z}_n(\theta)}p(\theta),$$

we can sample from $\tilde{p}(\theta|\mathbf{x}) = [g(\mathbf{x}|\theta)/\tilde{Z}_n(\theta)]p(\theta)$ which is the posterior corresponding to a slightly different prior

$$\tilde{p}(\theta) \propto \frac{Z_n(\theta)}{\tilde{Z}_n(\theta)}p(\theta)$$

than $p(\theta)$. Thus, if our grid-based sampler converges to $\tilde{p}(\theta|\mathbf{x})$, this would indicate that it is converging to the correct distribution.

In the Swedish pine sapling data, we use, for simplicity, a Straussian model (1) with interaction distance fixed at 0.7 meters – thus, θ is simply the Straussian parameter. To estimate $\tilde{Z}_n(\theta)$, at 13 evenly spaced values of θ' ranging from 0 to 3, we obtained ten million updates (after a 50,000 iteration burn-in) from $p(\mathbf{x}|\theta')$ via Geyer and Moller's algorithm. We then used the realizations to obtain 13 respective estimates of $Z_n(\theta')$ via (4). A cubic regression of $\log(Z_n(\theta'))$ on θ' allowed $\log(Z_n(\theta))$, and hence $Z_n(\theta)$, to be approximated for $\theta \in (0,3)$ (the log-transformed regression enabled a better fit through the otherwise very small $Z_n(\theta)$ values). Our parametric estimate $Z_n(\theta)$ was $\tilde{Z}_n(\theta) = \exp\{0.037 - 35.337\theta + 13.089\theta^2 - 1.536\theta^3\}$ for $\theta \in (0,3)$.

On the interval $(0,3)$, we placed a truncated $\mathrm{N}(2, \sigma^2 = 100)$ prior on θ. Using the same proposal density as in our earlier analysis, we obtained 20,000 updates

(following a 5,000 iteration burn-in) via MCMC. The density estimate of $\tilde{p}(\theta|\mathbf{x})$ is displayed in Figure 5.

[Figure 5 about here.]

The posterior mean is approximately 1.127, while the 95% HPD interval is $(0.619, 1.601)$. The shorter interval estimate is largely a consequence of fixing the interaction distance during the analysis, and possibly due to estimating r via deterministic means. Additionally, we performed a similar analysis with $\theta = (b, h)$ in (1). The results were again consistent with the grid-based analysis. Because the grid-based sampler converges to a distribution close to $\tilde{p}(\theta|\mathbf{x})$, this supports its convergence to the proper target distribution.

9. Discussion

In this article we have outlined a procedure to obtain Bayesian inferences in pairwise interacting point processes. Our new implementation allows for a wide variety of inferences from the full posterior distribution such as point and interval estimates; the probability that there exists spatial regularity is also easily approximated. While we have only explored the simple Straussian model, the ideas are easily generalized to more complex models such as marked pairwise interacting point processes. In addition, we have shown that importance sampling can be implemented within a *reversible jump* MCMC framework.

With current computing technology, grid-based importance sampling is feasible only for models in which the parameter space is of low dimension. If the parameter space had high dimension, there would be too many locations θ_i' at which importance samples would need to be generated (imagine constructing a grid in $\Re^1$, then $\Re^2$, then

$\Re^3$, etc.). Future advances in computing and storage technology will make for the routine application of grid-based importance sampling within higher-dimensional spaces. However, we have demonstrated that the dimensionality problem can be avoided if importance samples are generated on the fly, for example via (optimal) ratio importance sampling. Unfortunately such an implementation is computationally costly.

References

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

Chen, M.-H. and Shao, Q.-M. (1997). On Monte Carlo methods for estimating ratios of normalizing constants. *The Annals of Statistics* **25**, 1563–1594.

Degenhardt, A. (1999). Description of tree distribution patterns and their development through marked Gibbs processes. *Biometrika* **74**, 763–770.

Diggle, P. J., Fiksel, T., Grabarnik, P., Ogata, Y., Stoyan, D. and Tanemura, M. (1994). On parameter estimation for pairwise interaction point processes. *International Statistical Review* **62**, 99–117.

Fiksel, T. (1984). Estimation of parameterized pair potentials of marked and unmarked Gibbsian point processes. *Elektron. Inform. Kybernet.* **20**, 270–278.

Fiksel, T. (1988). Estimation of interaction potentials of Gibbsian point processes. *Math. Operationsf. Statist. Ser. Statist.* **19**, 77–86.

Geyer, C. J. and Moller, J. (1994). Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics* **21**, 359–373.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996). *Markov Chain Monte*

Carlo in Practice. Chapman and Hall, London.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

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

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087–1091.

Ogata, Y. and Tanemura, M. (1981). Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure. *Annals of the Institute of Statistical Mathematics* **33**, 315–338.

Ripley, B. D. (1981). *Spatial Statistics.* John Wiley and Sons, New York.

Ripley, B. D. (1988). *Statistical Inference for Spatial Point Processes.* Cambridge University Press, Cambridge.

Smith, A. F. M. and Gelfand, A. E. (1992). Bayesian statistics without tears: A sampling resampling perspective. *The American Statistician* **46**, 84–88.

Stoyan, D. and Penttinen, A. (2000). Recent applications of point process methods in forestry statistics. *Statistical Science* **15**, 61–78.

Strand, L. (1972). A model for stand growth. In *IUFRO Third Conference Advisory Group of Forest Statisticians*, pages 207–216. Institut National de la Recherche Agronomique, Paris.

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

Takacs, R. (1986). Estimator for the pair-potential of a Gibbsian point process. *Math. Operationsf. Statist. Ser. Statist.* **17**, 429–433.

Venables, W. N. and Ripley, B. D. (1994). *Modern Applied Statistics with S-Plus.* Springer-Verlag, New York.

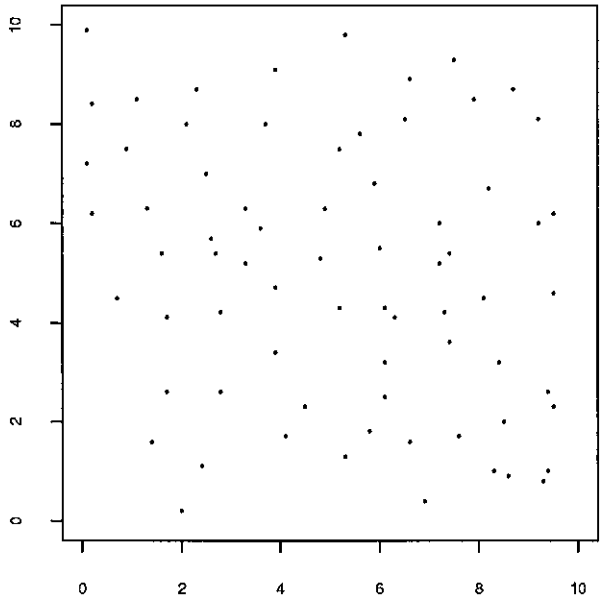

Figure 1. Location of 71 pine saplings within a 10×10 meter square region of forest. Data obtained from the MASS library accompanying Venables and Ripley (1994).

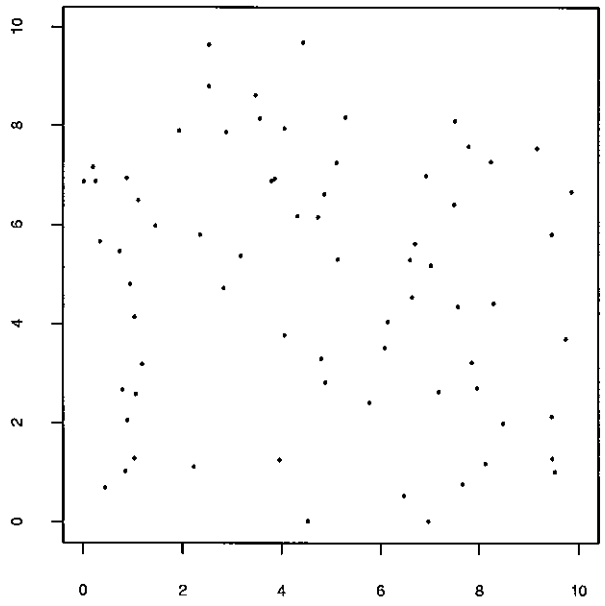

Figure 2. Realization of binomial process with 71 points.

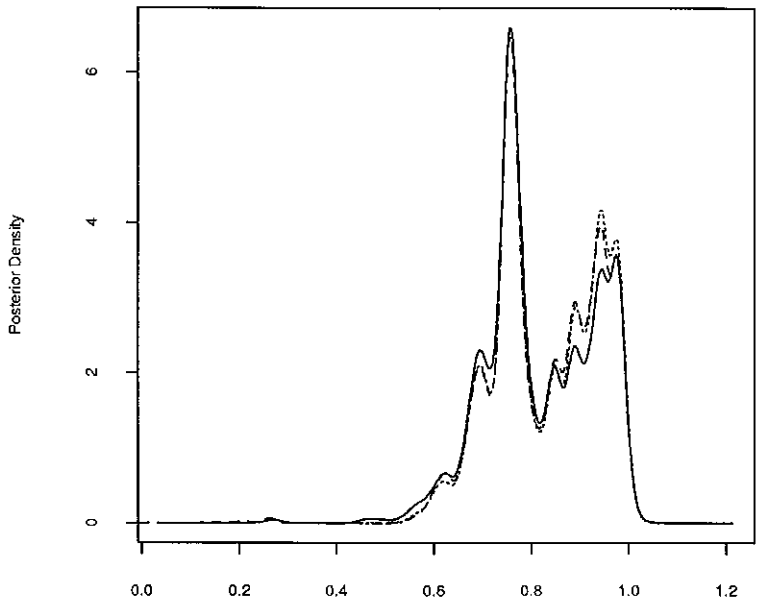

Figure 3. Marginal posterior density estimate of the interaction distance b. Solid line depicts the marginal posterior density estimate when choosing the θ_i' closest to θ^{curr}, dashed line when choosing the θ_i' which minimizes $(||\theta^{\text{curr}} - \theta_i'|| + ||\theta^{\text{prop}} - \theta_i'||)/2$, dotted line when choosing the θ_i' which minimizes $(||\theta^{\text{curr}} - \theta_i'||^2 + ||\theta^{\text{prop}} - \theta_i'||^2)/2$.

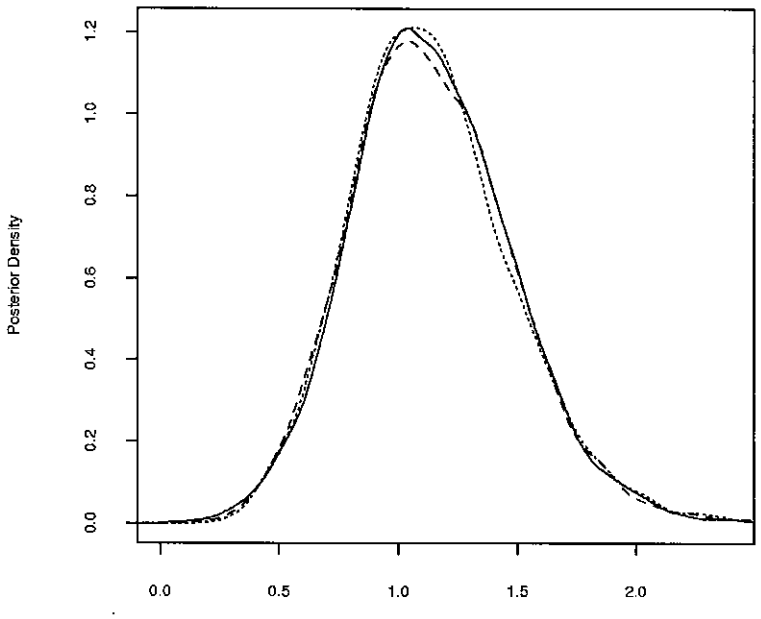

Figure 4. Marginal posterior density estimate of the Straussian parameter h. Solid line depicts the marginal posterior density estimate when choosing the θ_i' closest to θ^{curr}, dashed line when choosing the θ_i' which minimizes $(||\theta^{\text{curr}} - \theta_i'|| + ||\theta^{\text{prop}} - \theta_i'||)/2$, dotted line when choosing the θ_i' which minimizes $(||\theta^{\text{curr}} - \theta_i'||^2 + ||\theta^{\text{prop}} - \theta_i'||^2)/2$.

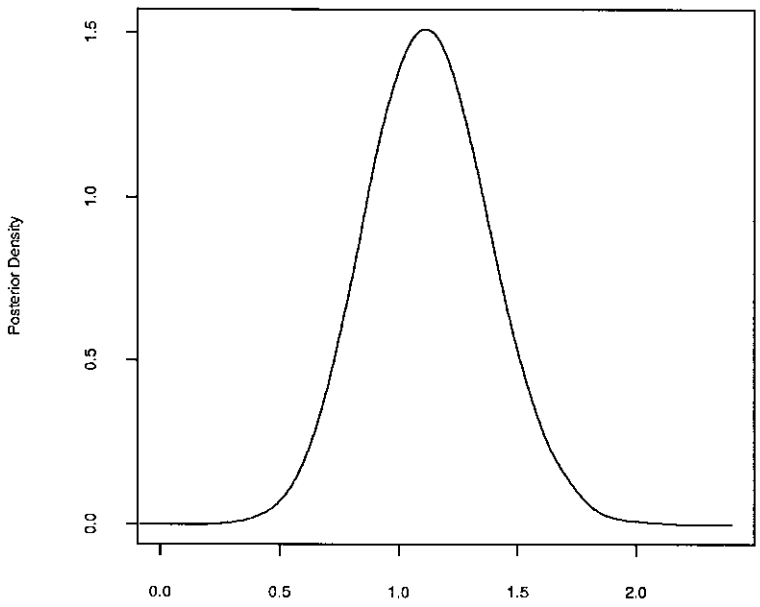

Figure 5. Density estimate of $\tilde{p}(\theta|\mathbf{x})$, the posterior distribution of the Straussian parameter θ under the prior $\tilde{p}(\theta)$.

Table 1

Estimated coefficient of variation (CV) of r for various values of the current θ^{curr} and proposed θ^{prop} Straussian parameter. The importance sampling location was $\theta'_i = 1.0$.

θ^{prop}	θ^{curr}	CV	θ^{prop}	θ^{curr}	CV
1.125	0.875	0.032	1.25	0.75	0.076
1.250	0.750	0.076	1.50	1.00	0.087
1.375	0.625	0.124	1.75	1.25	0.155
1.500	0.500	0.206	2.00	1.50	0.240
1.750	0.250	0.626	2.50	2.00	0.423
2.000	0.000	5.530	3.00	2.50	0.462

Table 2

Estimated coefficient of variation (CV) of r for various values of the current θ^{curr} and proposed θ^{prop} interaction distance. The importance sampling location was $\theta_i' = 0.10$.

θ^{prop}	θ^{curr}	CV	θ^{prop}	θ^{curr}	CV
0.1025	0.0975	0.068	0.105	0.095	0.130
0.1050	0.0950	0.130	0.110	0.100	0.153
0.1075	0.0925	0.231	0.115	0.105	0.234
0.1100	0.0900	0.298	0.120	0.110	0.332
0.1125	0.0875	0.326	0.130	0.120	0.569
0.1200	0.0800	0.522	0.140	0.130	0.915
0.1300	0.0700	1.122	0.150	0.140	1.287

Table 3

Estimated bias and mean squared error resulting from the estimation of the Straussian repulsion parameter from a Bayesian analysis of 100 simulated datasets from $p(\mathbf{x}|\theta)$.

Truth	Bias	MSE	Truth	Bias	MSE
$\theta = 0.3$	0.042	0.030	$\theta = 1.5$	−0.025	0.085
$\theta = 0.5$	−0.029	0.054	$\theta = 1.7$	0.042	0.097
$\theta = 0.7$	0.024	0.043	$\theta = 2.0$	0.018	0.121
$\theta = 1.0$	−0.003	0.054	$\theta = 2.3$	0.069	0.136
$\theta = 1.3$	0.005	0.072	$\theta = 2.5$	0.121	0.128

Table 4

*Estimated coverage (*s.e.*) and average interval width for 95% HPD and equal tail intervals, bias and mean squared error resulting from the Bayesian estimation of the Straussian parameter in 500 simulated datasets generated from* $p(\mathbf{x}|\theta = 1.0)$.

	$\theta^{\text{prop}} \in (\theta^{\text{curr}} \pm 1.0)$	$\theta^{\text{prop}} \in (\theta^{\text{curr}} \pm 0.5)$
HPD Coverage	95.0% (0.97%)	93.2% (1.13%)
HPD Ave. Width	1.012	0.932
Eq. Tail Coverage	95.6% (0.92%)	94.2% (1.05%)
Eq. Tail Ave. Width	1.029	0.948
Bias	0.0074	0.0240
MSE	0.0652	0.0620