Likelihood-based Inference for Matérn Type III Repulsive Point Processes

Version: April 29, 2014

Mark L. Huber
Departments of Mathematics and Statistical Science, Duke University
autotomic@gmail.com

Robert L. Wolpert
Department of Statistical Science and Nicholas School of the Environment, Duke University
wolpert@stat.duke.edu

Abstract

In a repulsive point process, points act as if they are repelling one another, leading to underdispersed configurations when compared to a standard Poisson point process. Such models are useful when competition for resources exists, as in the locations of towns and trees. Bertil Matérn introduced three approaches to modeling repulsive point processes, and used Types I and II. The third he regarded as intractable. In this paper an algorithm is developed that allows for arbitrarily accurate approximation of the likelihood for data modeled by the Matérn Type III process. This method relies on a perfect simulation method that is shown to be fast in practice, generating samples in time that grows nearly linearly in the intensity parameter of the model, while the running times for more naive methods grow exponentially.

Keywords: Coupling from the past; underdispersed; point process; thinning.

AMS Classification 2000: 65C6062M30;68U20

1 Introduction

Spatial data are often more regularly spaced than would be expected under a simple Poisson point process model because of competition for resources among the entities represented by the points of the model. These types of point processes are called *repulsive* because the points act as if they are repelling one another. Examples considered in Section 4 are the locations of Iberian towns, whose competition for farm produce and labor supply leads to apparent repulsion; and the locations of Swedish pine saplings, whose competition for light, water and minerals leads to underdispersion.

Various methods have been introduced to model repulsive or underdispersed point processes, including simple inhibition processes (Matérn, 1986, Chapter 3), thinned processes (Brown, 1979; Stoyan, 1979), the Strauss process (Strauss, 1975), and Markov point processes (Ripley & Kelly, 1977), which model the configuration by assigning a density (to discourage "closeness" of the points) with respect to a homogeneous Poisson point process.

Matérn (1986, *p.* 48) offered a different approach to modeling repulsion. Rather than use a density to describe the model, he gave three algorithms for generating configurations that implicitly (rather than explicitly) build in repulsion.

Section 1 describes this approach in more detail, and presents the central questions that need to be addressed in order to pursue a Bayesian (or other likelihood-based) approach to inference for these problems. Section 2 gives a simple Acceptance/Rejection approach to solve this problem that is suitable for small data sets, but which scales poorly with problem size. Section 3 presents a new perfect simulation method that can handle much larger sets, and develops an arbitrarily accurate product estimator for the likelihood function. In Section 4 this new method is applied to several examples, and comparisons to earlier work are made.

For one-dimensional point sets, a modification of dynamic programming techniques can be used to evaluate the likelihood function. The data sets considered in Section 4 are two-dimensional, so the dynamic programming approach fails; moreover, the techniques developed in this paper work not only for $\mathbb{R}^d$ but also for far more general sets, including any complete separable metric space $(\mathcal{S}, \text{dist})$ equipped with its Borel sets $\mathcal{B}$ and a finite positive

reference measure $\mu(\cdot)$ such as Lebesgue or counting measure.

Matérn (1986, *pp.* 47–48) introduced what are now called his *Type I* and *Type II* thinned point processes (see Figure 1). Begin with an intensity $\lambda > 0$, a sigma-finite Borel reference measure $\mu(\cdot)$ on $\mathbb{R}^d$, and a "primary" Poisson point process with mean $\lambda\,\mu(\cdot)$ that will assign to any region $\mathcal{S}$ of finite measure a Poisson-distributed number of points with expected value $\lambda \cdot \mu(\mathcal{S})$.

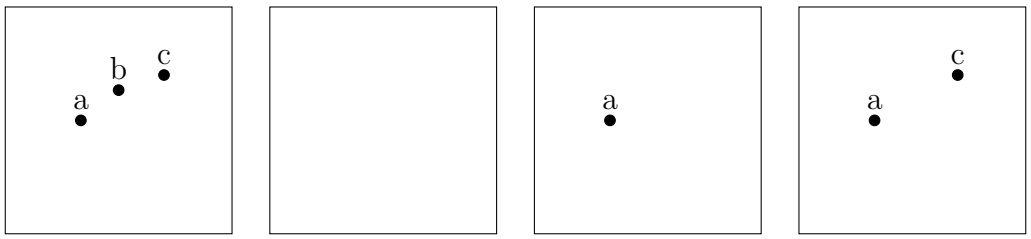
Given a fixed radius $R > 0$, Matérn's Type I process consists of the "secondary" points remaining upon removing all primary points that lie within distance R of any other primary point (so, if a pair lie within R, both are removed).

For the Type II process, Matérn assigns independent time stamps $t_j \overset{\text{iid}}{\sim} \mathsf{Un}(0, 1]$ to each of the primary points. The secondary points are those points whose time stamp is smaller than all neighbors within distance R. In other words, Matérn Type II only removes the later of two points which lie within distance R; for fixed λ and R this will have a higher density of points than the Type I process.

Matérn (1986, *p.* 48) briefly mentions a third point process featuring "dynamic" thinning in which points are only removed if they lie within distance R of an earlier *secondary* point; the remaining points, which we call the Matérn Type III process (or, more briefly, Matérn III), will have still higher density (for fixed λ and R) than Types I or II. It is this third process which is considered in detail below.

Let $\mathcal{S}$ be a finite-measure subset of $\mathbb{R}^d$. All three of these processes yield random sets $x \subseteq \mathcal{S}$ of secondary points with the property that balls of radius $R/2$ centered at each point do not intersect. Let $\#(x)$ denote the number of points in the set x. Call the $\#(x)$ secondary points $x \subseteq X$ that remain in the process *seen points*, and the $N - \#(x)$ points $X \setminus x$ that have been removed *hidden points*. Renumber if necessary so one can write $x = (x_1, \ldots, x_{\#(x)})$ for the seen points and $t = (t_1, \ldots, t_{\#(x)})$ for the associated (unobserved) time stamps.

Note that the time stamps could have been drawn from any continuous distribution on the real numbers. Suppose that instead of uniform distributions, independent draws $(\tilde{t}_i)$ are taken from some other distribution over the real numbers with continuous cumulative distribution function $F(\cdot)$. Then



Primary Poisson process	Matérn I	Matérn II	Matérn III
$R = 0.4$	$\{a, b\}$ eliminated	$t_a < t_b$ so	$t_a < t_b$ so
$\text{dist}(a, b) \approx 0.30$	No points remain.	a eliminates b.	a eliminates b.
$\text{dist}(b, c) \approx 0.23$		$t_b < t_c$ so	Since b gone
$\text{dist}(a, c) \approx 0.53$		b eliminates c.	a does *not* elim. c.
$t_a \approx 0.27,\ t_b \approx 0.52,\ t_c \approx 0.78$		Point a remains.	Points a and c remain.

Figure 1: Example of Matérn repulsive point processes.

$\{t_i := F(\tilde{t}_i)\}$ will be independent and uniformly distributed on $(0, 1]$. Moreover, ties are impossible under both (continuous) distributions and $\tilde{t}_i < \tilde{t}_j$ if and only if $t_i < t_j$, so no generality is lost in assigning uniform distributions on $(0, 1]$ to the time stamps.

Let $\theta \in \Theta$ index possible parameter vectors $\{\lambda(\theta), R(\theta)\}$. The Matérn process may also be constructed from a Poisson point process on $\mathcal{S} \times (0, 1]$ with product intensity measure $\Lambda_\theta(ds\, dt) := \lambda \cdot \mu(ds)\, m(dt)$, where m denotes Lebesgue (or uniform) measure on $(0, 1]$. The first coordinate is the location in $\mathcal{S}$, while the second coordinate is the time stamp. The *seen* points are those for which no other seen point with a lower time stamp lies within distance R. Note $\Lambda_\theta\big((\mathcal{S} \times (0, 1]\big) = \lambda \cdot \mu(\mathcal{S})$.

For any possible configuration $x = (x_1, \ldots, x_{\#(x)}) \subset \mathcal{S}$ of seen points, the time stamp vector $t = (t_1, \ldots, t_{\#(x)})$ could take any value in the cube $(0, 1]^{\#(x)}$. Any hidden point $y \in \mathcal{S}$ must lie within distance R from some seen point $x_i \in x$ with time stamp $t_y > t_i$— otherwise, (y, t_y) would not have been removed in thinning. For $s \in \mathcal{S}$ and $r > 0$ let $B_r(s) := \{y \in \mathcal{S} :\ \text{dist}(s, y) < r\}$ denote the r-ball in $\mathcal{S}$ centered at s. Then define:

$$\begin{aligned} D(x, t) &:= \big\{(y, t_y) : (\exists x_i \in x)\ \big[\,\text{dist}(x_i, y) < R\big] \wedge [t_i < t_y]\big\} \\ &= \cup_{i=1}^{\#(x)}\ B_R(x_i) \times (t_i, 1] \\ A_\theta(x, t) &:= \Lambda_\theta\big(D(x, t)\big). \end{aligned}$$

Call region $D(x,t)$ the *shadow* of configuration x with time stamps t, and $A_\theta(x,t)$ the shadow's *area*, the expected number of hidden points for the configuration (x,t) and parameter θ. Figure 2 shows an example of the shadow region for three points in $\mathcal{S} = [0,10]$. Note $D(x,t)$ depends on R, and $A_\theta(x,t)$ depends on both R and λ.

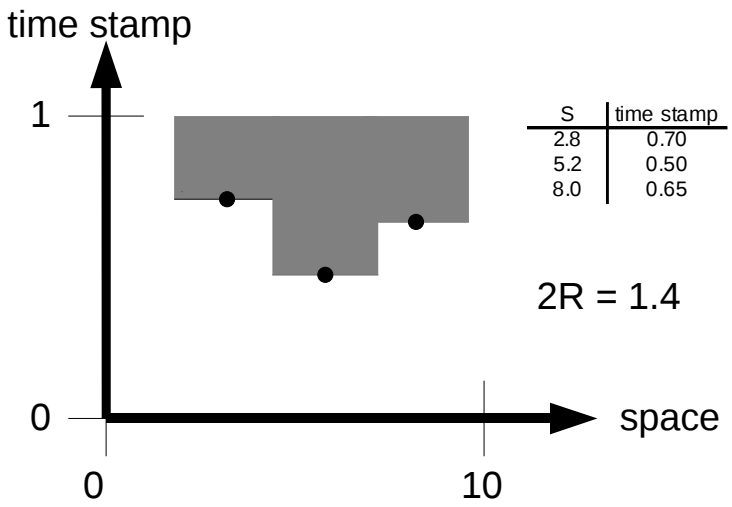


Figure 2: Example of a shadow region $D(x,t)$.

To find the density function for the seen points, that is, the likelihood function for θ, begin by writing the density function for *all* the points (seen and hidden) with respect to a Poisson point process on $\mathcal{S}$ with intensity $\mu(\cdot)$:

$$f_{\text{all points}}(x,t \mid \theta) = e^{\mu(S)(1-\lambda)}\lambda^{\#(x)}. \tag{1}$$

Conditional on the seen points x with time stamps t, the hidden points are a simple Poisson point process on $D(x,t)$ with intensity $\Lambda_\theta = \lambda\,\mu{\times}m$, where m is Lebesgue measure. Integrating away the unobserved hidden points in $D(x,t)$ leaves the marginal density function for the locations and latent time stamps for the seen points (again with respect to a Poisson point process with intensity $\mu(\cdot)$ on $\mathcal{S}$):

$$f_{\text{seen points}}(x,t \mid \theta) = \mathbf{1}_{\{(\forall i \neq j)\,\text{dist}(x_i,x_j)>R\}}\, e^{\mu(\mathcal{S})\{1-\lambda\}}\lambda^{\#(x)}\, \exp\{A_\theta(x,t)\} \tag{2}$$

The marginal density function $g(x \mid \theta)$ for the seen points is then given by integrating the latent time stamps:

$$g(x \mid \theta) = \mathbf{1}_{\{(\forall i \neq j)\, \mathrm{dist}(x_i, x_j) > R\}}\, e^{\mu(\mathcal{S})\{1-\lambda\}} \lambda^{\#(x)} \int_t \exp\{A_\theta(x,t)\}\, dt \qquad (3)$$

where "dt" denotes Lebesgue measure $m(dt_1) \ldots m(dt_{\#(x)})$ on the unit cube $(0,1]^{\#(x)}$. This is also the *likelihood function* for θ, upon observing $X = x$.

Note that $A_\theta(x,t)$ is an increasing function of R. This makes $g(x \mid \theta)$ an increasing function of R over $\big[0, \min_{i,j} \mathrm{dist}(x_i, x_j)\big)$.

In principle, for any prior distribution $\pi(d\theta)$ the posterior distribution for θ can be found as:

$$\pi(d\theta \mid X = x) = Z_x^{-1}\; g(x \mid \theta)\; \pi(d\theta)$$

where $Z_x := \int_\Theta g(x \mid \theta')\pi(d\theta')$ is the appropriate normalizing constant. The awkward integral in Equation (3) is an obstacle for this straightforward approach to inference. Instead, we will follow a data augmentation strategy in which random samples are drawn of the hidden points Y and of the unobserved time stamps T for both seen and hidden points, from their conditional distributions (given x and θ); after this, inference about θ is routine. For use in Section 3.1, we first compute the conditional density function of the time stamp vector $T \in (0,1]^{\#(x)}$ for seen points x. From Equation (3), this is simply:

$$h_\theta(t \mid x) = Z_{\theta,x}^{-1} \exp\{A_\theta(x,t)\}\; \mathbf{1}_{\{t \in (0,1]^{\#(x)}\}}, \qquad (4)$$

with normalizing constant $Z_{\theta,x} := \int_t \exp\{A_\theta(x,t)\}\, dt$.

2 Acceptance/Rejection

A simple approach for approximating $g(x \mid \theta)$ is the *Acceptance/Rejection* method of von Neumann (1951). This also provides a method to draw hidden points Y and time stamps T from their conditional distributions that works well for small examples but (as shown below) it scales badly in the number $\#(x)$ of seen points:

This is an example of a perfect simulation algorithm: the output comes exactly from the desired distribution, but the running time is itself an unbounded random variable. The number of failure steps in the repeat loop

Algorithm 2.1 `Acceptance_Rejection`

Input: parameter θ, seen points x

Output: random draws of hidden points Y, time stamps T

1: **repeat**
2: **draw** hidden points $Y \subseteq \mathcal{S}$ from Poisson process w/intensity $\lambda\,\mu(\cdot)$
3: **draw** time stamps T for x and Y uniformly from $(0,1]^{\#(x)+\#(Y)}$
4: **until** for all $a \in Y$, there exists $b \in x$ with $\mathrm{dist}(a,b) \le R$ and $t_b < t_a$
5: return Y and T

has a geometric distribution with success parameter $p = \int_t \exp\{A_\theta(x,t) - \lambda\mu(\mathcal{S})\}\,dt$ and so the expected number of loop steps is $1/p$.

Unfortunately, for simple choices of reference measure such as Lebesgue, the probability p decreases (and the expected number of loop steps increases) geometrically in the number $\#\,(x)$ of seen points. Consider a rectangular region $\mathcal{S} \subseteq \mathbb{R}^d$ for some $d \in \mathbb{N}$, for example, with sides of length at least R and with periodic boundary conditions (to simplify the argument by eliminating edge effects). With Lebesgue reference measure $\mu(\cdot)$, the total volume of the region is $V = \mu(\mathcal{S})$, the product of the lengths of the sides.

Since balls of diameter R centered at the seen points x are disjoint, each with volume $v := \nu_d(R/2)^d$ in $\mathbb{R}^d$ for $\nu_d = \pi^{d/2}/\Gamma(1+d/2)$bound on the *unshadowed* volume leads to an upper bound on the shadowed volume and hence on the acceptance probability:

$$\begin{aligned} \lambda V - A_\theta(x,t) \ge \lambda v \textstyle\sum_{j=1}^{\#(x)} t_j \qquad \text{so} \\ p := \int_t \exp\{A_\theta(x,t) - \lambda V\}\,dt \le \int_t \exp\{-\lambda v \textstyle\sum_{j=1}^{\#(x)} t_j\}\,dt \\ = \rho^{\#(x)} \end{aligned} \tag{5}$$

where $\rho := \big\{1 - \exp\{-\lambda v\}\big\}/(\lambda v) < 1$. In $d = 1$ dimension one has $\nu_1 = 2$ and $p \le \rho^{\#(x)}$ with $\rho = \{1 - \exp\{-\lambda\,R\}\}/(\lambda\,R)$, for example, while $\nu_2 = \pi$ and $\rho = [1 - \exp\{-\lambda\,\pi R^2/4\}]/\{\lambda\,\pi R^2/4\}$ in $d = 2$ dimensions.

A similar argument gives a lower bound for the shadow volume, giving a

lower bound for p:

$$A_\theta(x,t) \geq \lambda v \sum_{j=1}^{\#(x)}(1-t_j)$$
$$p := \int_t \exp\{A_\theta(x,t) - \lambda V\}\, dt \geq \int_t \exp\{\lambda\{v\sum_{j=1}^{\#(x)}(1-t_j) - V\}\}\, dt$$
$$= \rho^{\#(x)} \exp\{-\lambda\{V - v\,\#\,(x)\,\}\},$$

so Algorithm 2.1 may be very reasonable for sufficiently small $\#\,(x)$. The geometric decrease of Equation (5) makes the basic form of Acceptance/ Rejection untenable for large $\#\,(x)$; we now turn to an alternative.

3 Coupling from the past

Acceptance/Rejection works well for small sets of seen points, but quickly becomes impractical as $\#\,(x)$ grows. In this section a new Markov chain for this problem is created, that allows for perfect sampling when used with read once coupling from the past (ROCFTP) (Wilson, 2000).

Our CFTP approach employs Form 1 of the bounding chains of Huber (2004), closely related to the antimonotonicity approach of Møller (1999) and Häggström & Nelander (1999). The bounding chain view is as follows: Suppose that the state space for a Markov chain (X_t) is of the form $\Omega_X = \mathcal{T}^{\mathcal{I}}$ for a set of "labels" $\mathcal{T}$ and a set of "indices" $\mathcal{I}$, that is, each state is an assignment of labels to the indices. The state space for the time stamps is $(0,1]^{\{1,\ldots,\#(x)\}}$, so here $\mathcal{T} = (0,1]$ and $\mathcal{I} = \{1,\ldots,\#\,(x)\}$. Now consider a new state space $\Omega_B = (2^{\mathcal{T}})^{\mathcal{I}}$ so that each index $i \in \mathcal{I}$ is assigned a *subset* of labels rather than a single label. The bounding chain method runs a Markov chain (B_t) on this larger space that bounds the possible values for the original chain (X_t).

Definition 1. Suppose that (A_t) and (B_t) are Markov chains, both adapted to the same filtration $\mathcal{F}_t$. Say that (A_t, B_t) is a *coupling* of the processes if for all measurable A, B and states a, b:

$$\mathbf{P}(A_{t+1} \in A \mid A_t = a, \mathcal{F}_t) = \mathbf{P}(A_{t+1} \in A \mid A_t = a)$$
$$\mathbf{P}(B_{t+1} \in B \mid B_t = b, \mathcal{F}_t) = \mathbf{P}(B_{t+1} \in B \mid B_t = b).$$

A coupling is simply a construction of both Markov chains on the same probability space with the same filtration.

Definition 2. A Markov chain (B_t) on state space $(2^{\mathcal{T}})^{\mathcal{I}}$ is a *bounding chain* for Markov chain (X_t) on state space $\mathcal{T}^{\mathcal{I}}$ if there exists a coupling (X_t, B_t) such that

$$(\forall i \in \mathcal{I})\ X_t(i) \in B_t(i) \quad \Longrightarrow \quad (\forall i \in \mathcal{I})\ X_{t+1}(i) \in B_{t+1}(i).$$

For any $x \in \mathcal{T}^{\mathcal{I}}$ and $B \in (2^{\mathcal{T}})^{\mathcal{I}}$, write "$x \in B$" if $x(i) \in B(i)$ for all $i \in \mathcal{I}$. Now CFTP with bounding chains can be described. Suppose that the initial state of the chain is drawn from its stationary distribution, $X_0 \sim \pi$, and that the initial state of the bounding chain $B_0 = \mathcal{T}^{\mathcal{I}}$, so $B_0(i) = \mathcal{T}$ for every $i \in \mathcal{I}$ and hence $X_0 \in B_0$. Then after a block of any fixed number of steps $\tilde{t}$, by stationarity and boundedness one also has $X_{\tilde{t}} \sim \pi$ and $X_{\tilde{t}} \in B_{\tilde{t}}$.

After taking such a block of length $\tilde{t}$ starting in states $X_0 \sim \pi$ and $B_0 = \mathcal{T}^{\mathcal{I}}$, there are two possibilities. Either $B_{\tilde{t}}$ contains exactly one state (namely $X_{\tilde{t}}$), or it contains more than one state. When $\#(B_{\tilde{t}}) = 1$, call the block a *success*, otherwise call it a *failure*. Each block succeeds or fails independently, and the probability p that any particular block is a success is the same for all blocks. The *read once CFTP* method of Wilson takes advantage of the following fact, first shown in different form by Wilson (2000).

Lemma 1. *Let $G \sim \mathsf{Ge}(p)$ for some $0 < p \leq 1$ (so $\mathbf{P}(G = j) = p(1-p)^j$ for integers $j \geq 0$). Let $X_0 \sim \pi$, where π is the stationary distribution of the Markov chain (X_t), and run the Markov chain forward through one success block followed by a series of G failure blocks, each block of length $\tilde{t}$ steps beginning with $B_{j\tilde{t}} = \mathcal{T}^{\mathcal{I}}$. Then $X_{\tilde{t}(1+G)} \sim \pi$.*

Proof. Draw $X_0 \sim \pi$ and, for some large number $k \in \mathbb{N}$ to be specified later, take $k\tilde{t}$ steps in the Markov chain. Divide these steps into k blocks of length $\tilde{t}$. Since π is stationary, $X_{k\tilde{t}} \sim \pi$ as well.

Let G be the number of failure blocks following the *last* success block, if any, among the first k blocks, and otherwise set $G = k$. Then G is distributed as the minimum of k and a $\mathsf{Ge}(p)$ random variable. Let Y denote a random variable chosen by simulating a success block (which returns a single state in the bounding chain) followed by G failure blocks. Let Z be a random variable found by starting with $Z_0 \sim \pi$, and then running k failure blocks. Now consider $X_{k\tilde{t}}$. Conditioned on $G < k$ one has $X_{k\tilde{t}} \sim [Y \mid G < k]$, while conditioned on $G = k$, $X_{k\tilde{t}} \sim Z$. In other words, for measurable A:

$$\mathbf{P}(X_{k\tilde{t}} \in A) = \mathbf{P}(Z_{k\tilde{t}} \in A)(1-p)^k + \mathbf{P}(Y \in A \mid G < k)[1 - (1-p)^k].$$

Since π is stationary, $\mathbf{P}(X_{k\tilde{t}} \in A) = \pi(A)$ for all k. By assumption $p > 0$, so as $k \to \infty$ the right hand side converges to $\mathbf{P}(Y \in A)$, completing the proof. □

The coupling between chain and bounding chain is typically accomplished by use of update functions (see, *e.g.*, Foss & Tweedie, 1998). Let (X_t) be a time homogeneous Markov chain on state space Ω. Let U be a uniform random number on $(0, 1]$. Call $\phi : \Omega \times (0, 1] \to \Omega$ an *update function* if for all $x \in \Omega$ and measurable A:

$$\mathbf{P}\big(\phi(x, U) \in A\big) = \mathbf{P}\big(X_1 \in A \mid X_0 = x\big).$$

The chain can be constructed recursively from a countable set $(U_t) \overset{\text{iid}}{\sim} \mathsf{Un}(0, 1]$ by the recipe $X_0 = x$ and, for $t > 0$, $X_t = \phi(X_{t-1}, U_t)$.

A single uniform variate may be regarded an infinite sequence of uniform $\{0, 1\}$ random bits, which in turn can be used to construct a countable number of new uniform random variates on $(0, 1]$. Therefore, the update function described here is very general, and includes any method for updating that uses at most a countable number of random draws. This generality will be useful when a Markov chain step requires a draw from a Poisson point process, which (as typically implemented) requires a Poisson-distributed number of uniform draws.

Any update function for a Markov chain will induce one for the bounding chain, and a coupling of the chains. Suppose that (B_t) is a bounding chain for process (X_t) with update function ϕ. Then to update B_t given $U_{t+1} \sim \mathsf{Un}(0, 1]$ use $B_{t+1} = \Phi(B_t, U_{t+1})$, with update function

$$\Phi(B_t, U) := \cup_{x \in B_t} \, \phi(x, U). \tag{6}$$

The chains X_t and B_t are coupled, since both are adapted to $\mathcal{F}_t := \sigma\{U_s : s \leq t\}$.

Lemma 1 says that a sample may be generated from π by running a success block followed by a geometrically distributed number G of failure blocks. To determine G, run blocks successively. Once a success block has occurred, the state just prior to each subsequent success block will be a draw from the stationary distribution (see Figure 3). The following pseudocode will run the Markov chain through $\tilde{t}$ steps to form a single block and indicate its success or failure.

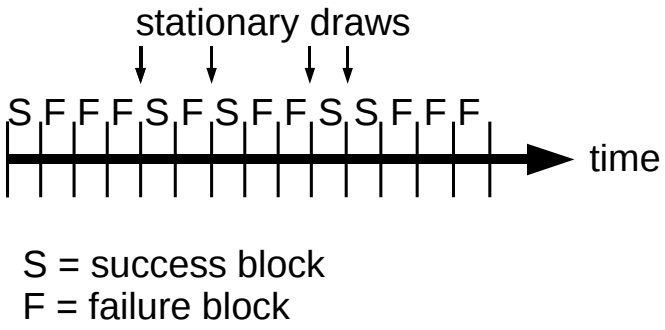


Figure 3: Read once coupling from the past

Algorithm 3.1 `Evaluate_block`

Input: current state X, update function ϕ, block length $\tilde{t}$, state space $\mathcal{T}^{\mathcal{I}}$
Output: X output of block, `success_block` $\in \{\text{true}, \text{false}\}$

1: $B \leftarrow \mathcal{T}^{\mathcal{I}}$
2: **for** t from 1 to $\tilde{t}$ **do**
3: **draw** $U \leftarrow \mathsf{Un}(0, 1]$
4: $X \leftarrow \phi(X, U)$
5: $B \leftarrow \Phi(B, U) := \cup_{x \in B}\ \phi(x, U)$
6: **end for**
7: `success_block` $\leftarrow (\#\,(B) = 1)$

With the block evaluation routine in hand, the following pseudocode runs read once coupling from the past. Begin by running blocks until a successful block is found. Since it is successful, the output of the block is a single state X. Save X in variable Y, and run forward through another block. If that block was a success, then Y was the state immediately preceding the success block, and so Y is a draw from π. Otherwise, continue taking blocks until a success is found.

The time needed to generate n draws with this procedure will be the time needed to generate $n + 1$ successful blocks, with mean $(n + 1)/p$ where p denotes the success probability.

Algorithm 3.2 `ROCFTP` (Read once coupling from the past)

Input: update fcn ϕ, number of samples n, block length $\tilde{t}$, state space $\mathcal{T}^{\mathcal{I}}$
Output: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \pi$

repeat
  $(X, \texttt{success\_block}) \leftarrow \texttt{Evaluate\_block}(\emptyset, \phi, \tilde{t}, \mathcal{T}^{\mathcal{I}})$
until `success_block`
for i from 1 to n **do**
  repeat
    $Y \leftarrow X$
    $(X, \texttt{success\_block}) \leftarrow \texttt{Evaluate\_block}(X, \phi, \tilde{t}, \mathcal{T}^{\mathcal{I}})$
  until `success_block`
  $X_i \leftarrow Y$
end for

3.1 The update function for Matérn time stamps

The key to success using read once coupling from the past is the construction of an update function $\phi(\cdot)$ for which it is easy to evaluate the induced update function $\Phi(\cdot)$ for the bounding chain.

The conditional density for the time stamps, given the seen points x, was given in Equation (4) for $t \in (0,1]^{\#(x)}$ as $h_\theta(t \mid x) = Z_{\theta,x}^{-1} \exp\{A_\theta(x,t)\}$, where (see Equation (1)) $A_\theta(x,t) := \lambda\mu\big(D(x,t)\big)$ is the expected number of hidden points in the shadow of configuration x with time stamps t.

The shadow $D(x,t)$ is a union of overlapping cylinders, making exact computation of $A_\theta(x,t)$ very difficult even for Lebesgue measure $\mu(\cdot)$ in only two dimensions. A Metropolis-Hastings approach will be used to avoid the explicit computation of $A_\theta(x,t)$.

The basic Metropolis-Hastings procedure (Metropolis et al., 1953; Hastings, 1970) converts a proposal Markov chain into one whose stationary distribution is exactly the target distribution. Suppose that the target distribution has a density function $f(t)$ with respect to a reference measure $\eta(dt)$ and, when at point T, the proposal chain proposes a point S using density $q(T,s)$ with respect to $\eta(ds)$.

Further, suppose that $f(t)q(t,s) = 0$ if and only if $f(s)q(s,t) = 0$. Then the following procedure takes one step in the Metropolis-Hastings chain. Here $\mathsf{Be}(r)$ denotes the Bernoulli distribution giving one with probability $r \wedge 1$ and otherwise zero.

Algorithm 3.3 `General_Metropolis_Hastings`

Input: current state T
Output: next state T
1: **draw** S from distribution $q(T, s)\,\eta(ds)$
2: $r \leftarrow \{f(S)q(S,T)\}/\{f(T)q(T,S)\}$
3: **draw** $B \leftarrow \mathsf{Be}(r)$
4: **if** $B = 1$ **then**
5: $\quad T \leftarrow S$
6: **end if**

Our present goal in making inference about the Matérn Type III process is to draw samples of the time stamps T, conditionally on the observed values of the seen points x. Notice that the Metropolis-Hastings ratio $r = \{f(S)q(S,T)\}/\{f(T)q(T,S)\}$ depends on the target density $f(\cdot)$ only up to a scale factor, so the conditional density $h_\theta(t \mid x)$ of the time stamps (see Equation (4)) may be replaced by the simpler unnormalized version

$$f(t) = \exp\{A_\theta(x,t)\}. \tag{7}$$

Consider a proposal chain in which an index $i \in \{1, \ldots, \#(x)\}$ is chosen uniformly and then a new time stamp t_{prop_i} is proposed for x_i from some distribution with symmetric density $q(s,t) = q(t,s)$. Call the resulting proposed set of time stamps t_{prop}. By symmetry, the Metropolis-Hastings ratio becomes simply:

$$r = \exp\{A_\theta(x, t_{\mathrm{prop}}) - A_\theta(x,t)\}. \tag{8}$$

Calculating the change in shadow area under the new time stamp is a difficult task even in only two dimensions. Therefore, it is necessary to construct an approach that does not require direct computation of r in line 2 of Algorithm 3.3.

In the case that the proposed new time-stamp is smaller, $t_{\mathrm{prop}} \leq t$, the shadow of x will grow ($A_\theta(x, t_{\mathrm{prop}}) \geq A_\theta(x,t)$) and the Metropolis-Hastings ratio $r \geq 1$ will exceed unity, so the proposal will always be accepted and one could replace lines 2–3 with "Set $B = 1$."

In the case that $t_{\mathrm{prop}} > t$, the Hastings ratio r of Equation (8) is *exactly* the probability that a Poisson point process with intensity Λ_θ will assign zero points to the newly unshadowed region $\Delta := D(x,t) \setminus D(x, t_{\mathrm{prop}})$. Thus, in lieu of lines 2–3, generate a Poisson point process $W \sim \mathsf{Po}(\Lambda_\theta)$ on Δ, and

set $B \leftarrow \{0 = \#(W)\}$. The set Δ never needs to be evaluated explicitly—generating a Poisson point process on Δ is a simple two-step process, as follows.

Each point in Δ lies within distance R of x_i and has a time stamp between t_i and t_{prop_i}, so

$$\Delta := D(x,t) \setminus D(x,t_{\text{prop}}) \subseteq B_R(x_i) \times (t_i, t_{\text{prop}_i}].$$

Thus the Poisson point process W can be generated by first generating such a process over $B_R(x_i) \times (t_i, t_{\text{prop}_i}]$, and then thinning it by retaining only points outside $D(x, t_{\text{prop}})$. If there are no such points then $B = 1$ and the proposal is accepted; otherwise $B = 0$ and the previous state is retained.

Algorithm 3.4 `Matern_II_Metropolis_Hastings`

Input: seen points x with current state of time stamps T
Output: next state of time stamps T

draw i uniformly from $\{1, \ldots, \#(x)\}$
$t \leftarrow T$
draw $t_i \leftarrow \mathsf{Un}(0,1]$
if $t_i \leq T_i$ **then**
  $T_i \leftarrow t_i$
else
  draw Poisson point process W on $B_R(x_i) \times (T_i, t_i]$ w/intensity Λ_θ
  if $\#(W \cap D(x,t)^c) = 0$ **then**
    $T_i \leftarrow t_i$
  end if
end if

Now consider the question of how to use this update function for a single state to update an entire bounding chain state. The bounding states for this chain will be intervals $B_t = [T_t^{\text{lo}}, T_t^{\text{hi}}]$, where

$$B_t := [T_t^{\text{lo}}, T_t^{\text{hi}}] = \left\{T : (\forall i)(T_t^{\text{lo}}(i) \leq T(i) \leq T_t^{\text{hi}}(i)\right\}.$$

Pseudocode for the bounding chain update follows.

Algorithm 3.5 `Matern_II_Bounding_Chain`

Input: seen points x, current bounds $B = [T^{\text{lo}}, T^{\text{hi}}]$
Output: next state of time stamp bounds $B = [T^{\text{lo}}, T^{\text{hi}}]$

draw i uniformly from $\{1, \ldots, \#(x)\}$
$t^1 \leftarrow T^{\text{hi}}$, $t_i^1 \leftarrow T_i^{\text{lo}}$
$t^2 \leftarrow T^{\text{lo}}$, $t_i^2 \leftarrow T_i^{\text{hi}}$
draw $t_i \leftarrow \mathsf{Un}(0, 1]$
$t^3 \leftarrow T^{\text{hi}}$, $t_i^3 \leftarrow t_i$
$t^4 \leftarrow T^{\text{lo}}$, $t_i^4 \leftarrow t_i$
if $t_i \leq T_i^{\text{lo}}$ **then** {Case I}
  $T_i^{\text{lo}} \leftarrow t_i, T_i^{\text{hi}} \leftarrow t_i$
else if $T_i^{\text{lo}} < t_i \leq T_i^{\text{hi}}$ **then** {Case II}
  $T_i^{\text{hi}} \leftarrow t_i$
  draw Poisson point process W on $B_R(x_i) \times (T_i, t_i]$ w/intensity Λ_θ
  if $\#(W \cap \{D(x, t^1) \setminus D(x, t^3)\}) = 0$ **then**
    $T_i^{\text{lo}} \leftarrow t_i$
  end if
else if $T_i^{\text{hi}} < t_i$ **then** {Case III}
  draw Poisson point process W on $B_R(x_i) \times (T_i, t_i]$ w/intensity Λ_θ
  if $\#(W \cap \{D(x, t^1) \setminus D(x, t^3)\}) = 0$ **then**
    $T_i^{\text{lo}} \leftarrow t_i$
  end if
  if $\#(W \cap \{D(x, t^2) \setminus D(x, t^4)\}) = 0$ **then**
    $T_i^{\text{hi}} \leftarrow t_i$
  end if
end if

To summarize what is happening: when the proposed time stamp t_i is smaller than the existing one, always accept the proposal and collapse $[T_i^{\text{lo}}, T_i^{\text{hi}}]$ to (t_i). When the proposed t_i is higher, the worst case for acceptance is when the original time stamps are as small as possible (i.e., equal to T_i^{lo}), and for all $j \neq i$ the time stamps t_j are as high as possible (i.e., equal to T_j^{hi}). This is shown in the proof to the following theorem.

Theorem 1. *Algorithm 3.5 is a valid update for the bounding chain.*

Proof. Suppose that the call

$$(T'^{\,\text{lo}}, T'^{\,\text{hi}}) \leftarrow \texttt{Matern\_II\_Bounding\_Chain}(x, T^{\text{lo}}, T^{\text{hi}}),$$

is used for the bounding chain and

$$T' \leftarrow \texttt{Matern\_II\_Metropolis\_Hastings}(x, T)$$

for the time stamps, where T satisfies $T^{\text{lo}} \leq T \leq T^{\text{hi}}$. Then with the same choices for i, t_i, and W as in Algorithm 3.5, we must show that $T'^{\,\text{lo}} \leq T' \leq T'^{\,\text{hi}}$. Three cases arise:

Case I: $t_i \leq T_i^{\text{lo}}$**.** Then $t_i \leq T_i^{\text{lo}} \leq T_i$ and hence $T_i' = t_i$, so setting $T_i^{\text{lo}} = T_i^{\text{hi}} = t_i$ still satisfies $T^{\text{lo}} \leq T' \leq T^{\text{hi}}$.

Case II: $T_i^{\text{lo}} < t_i \leq T_i^{\text{hi}}$**.** When $T_i \geq t_i$, then $T_i' = t_i$, and when $T_i < t_i$, then $T_i' \in \{t_i, T_i\} \leq t_i$, so $T_i'^{\,\text{hi}} = t_i \geq T_i'$ is a valid move.

Now consider $T'^{\,\text{lo}}$. When $t_i \leq T_i$ again $T_i' = t_i$, and so $T'^{\,\text{lo}} = t_i \leq T_i$ is a valid move, so suppose $t_i > T_i$.

The shadow change upon increasing T_i to t_i can also be written in the form:

$$\Delta = \{D(x,T) \setminus D(x,t)\} = \big\{B_R(x_i) \times (T_i, t_i]\big\} \cap D(x,t)^c. \qquad (9)$$

The set $B_R(x_i) \times (T_i, t_i]$ is as large as possible when T_i is as small as possible. The smallest it can be is T_i^{lo}.

The set $D(x,t)^c$ is as large as possible when $D(x,t)$ is as small as possible, which happens when for all $j \neq i$, $t_j = T_j^{\text{hi}}$. Combining these, $\{D(x,T) \setminus D(x,t)\}$ is as large as possible when $T = t^1$ and $t = t^3$. In other words:

$$D(x,T) \setminus D(x,t) \quad \subseteq \quad D(x,t^1) \setminus D(x,t^3).$$

Now suppose that $\#\left(W \cap \left\{D(x,t^1) \setminus D(x,t^3)\right\}\right) = 0$. Then necessarily

$$\#\left(W \cap \{D(x,T) \setminus D(x,t)\}\right) = 0$$

as well for every $T \in [T^{\text{lo}}, T^{\text{hi}}]$, so the proposal $T' = t$ will be accepted and $T_i'^{\,\text{lo}} = t_i \leq T_i'$ is a valid move in the bounding chain.

Case III: $T_i^{\text{hi}} < t_i$**.** The case for $T'^{\,\text{lo}}$ is the same as in Case II. It remains to show that the move for $T'^{\,\text{hi}}$ is valid. Use representation (9) above for Δ. This is as small as possible when $(T_i, t_i]$ is as small as possible,

so $T_i = T_i^{\text{hi}}$. Similarly, $D(x,t)^c$ is as small as possible when $D(x,t)$ is as large as possible, which happens when $t_j = T_j^{\text{lo}}$ for all $j \neq i$.

Therefore

$$D(x,t^2) \setminus D(x,t^4) \quad \subseteq \quad D(x,T) \setminus D(x,t).$$

If $\#\left(W \cap \{D(x,t^2) \setminus D(x,t^4)\}\right) > 0$, then $\#\left(W \cap \{D(x,T) \setminus D(x,t)\}\right) > 0$ for all T satisfying $T^{\text{lo}} \leq T \leq T^{\text{hi}}$, so the proposal t will be rejected and $T_i'^{\,\text{hi}} = T_i^{\text{hi}} \geq T_i' = T_i$ is fine. Otherwise, there might be some T that accepts the move, and so $T_i'^{\,\text{hi}} = t_i \geq T_i'$ as in line 21.

This completes the proof. □

3.2 A product estimator of the likelihood

Given the ability to draw samples from π, there exist many Monte Carlo methods for estimating the likelihood $g(x \mid \theta)$. Here a version the product estimator is presented (see Fishman (1996, §5.31) for more details of this approach, and Štefankovič et al. (2007) for a more advanced variant). The product estimator algorithm has the advantage of giving approximations to the likelihood that are immune to the effects of multimodality.

Begin by fixing x and R and consider a range of possible values of λ. Fix $0 < \gamma \leq 1/\mu(\mathcal{S})$ and, for integers $k \geq 0$, fix $\theta_k \in \Theta$ such that $\lambda(\theta_k) = \lambda_k := \gamma k$ and $R(\theta_k) = R$. Note that $A_{\theta_k} = kA_{\theta_1} = k\,\gamma\,(\mu \times m)\big\{D(x,t)\big\}$.

From Equation (3) one has $g(x \mid \theta_0) = \mathbf{1}_{\{x=\emptyset\}}$ and, for $k \geq 1$, $g(x \mid \theta_k) = a_k\,\lambda_k^{\#(x)}\,\exp\{\mu(S)[1-\lambda_k]\}$ where

$$a_k := \int_t \exp\{A_{\theta_k}(x,t)\}\,dt = \int_t \exp\{k\,A_{\theta_1}(x,t)\}\,dt$$

Evidently $a_0 = 1$; for $k \geq 1$, estimate a_k by first estimating each ratio in the telescoping product

$$\frac{a_0}{a_1}\,\frac{a_1}{a_2}\,\frac{a_2}{a_3}\,\cdots\,\frac{a_{k-1}}{a_k} = \frac{1}{\int_t \exp\{A_{\theta_k}(x,t)\}\,dt}.$$

Estimate each term (a_{i-1}/a_i) as follows. First, given the seen points x and radius R, fix n (to be chosen later) and, using Algorithm 3.4 (or 2.1), draw time stamps $T_1, \ldots, T_n \overset{\text{iid}}{\sim} h_{\theta_i}(t \mid x)\,dt$ (see Equation (4)) for intensity $\lambda_i = i\gamma$. For each $1 \leq j \leq n$, draw a Poisson point process W_j

with intensity $\gamma(\mu\times m)(ds\,dt)$ over the shadow $D(x,T_j)$. The independent Bernoulli random variables $B_j = \mathbf{1}_{\{\#(W_j)=0\}}$ have conditional expectation $\mathbf{E}(B_j \mid T_j) = \mathbf{P}\{\#(W_j) = 0 \mid T_j\} = \exp\{-A_{\theta_1}(x,T_j)\}$, so

$$\mathbf{E}(B_j) \;=\; \frac{\int_t \exp\{-A_{\theta_1}(x,t) + i\,A_{\theta_1}(x,t)\}\,dt}{\int_t \exp\{i\,A_{\theta_1}(x,t)\}\,dt} = \frac{a_{i-1}}{a_i}.$$

The probability $p_i = (a_{i-1}/a_i) = \exp\{-\gamma\mu(\mathcal{S})\}$ that W is empty is at least e^{-1}, since $\gamma \le 1/\mu(\mathcal{S})$, so the Bernoulli variables (B_j) have means $\mathbf{E}(B_j) = p_i > e^{-1}$ and coefficients of variation

$$\mathrm{Cov}(B_j) = \frac{\sqrt{p_i(1-p_i)}}{p_i} = \sqrt{(1-p_i)/p_i} \le \sqrt{e-1}.$$

Set $\bar{B}_i := (B_1 + \cdots + B_n)/n$. Then $\mathbf{E}(\bar{B}_i) = (a_{i-1}/a_i)$ and $\mathrm{Cov}(\bar{B}_i)^2 \le (e-1)/n$. Let $\hat{a}_i := 1\big/\prod_{j=1}^{i}\bar{B}_j$ be an estimate for a_i. Dyer & Frieze (1991, *pp.* 135–136) showed that under these conditions, for any $\epsilon > 0$ the condition $n \ge 16(e-1)i/\epsilon^2$ is sufficient to guarantee that

$$\mathbf{P}\big(|1/\bar{a}_i - 1/a_i| > \epsilon/a_i\big) \le 1/4.$$

The event $(1-\epsilon)/a_i \le 1/\hat{a}_i \le (1+\epsilon)/a_i$ is equivalent to $a_i(1+\epsilon)^{-1} \le \hat{a}_i \le a_i(1-\epsilon)^{-1}$. For $\epsilon < 0.1$, $(1-\epsilon/1.1)^{-1} \le 1+\epsilon$ and always $(1-\epsilon) \le (1+\epsilon)^{-1}$. Therefore, after $n \ge 16(e-1)i(1.1/\epsilon)^2 = 19.36(e-1)i/\epsilon^2$ steps, when $\epsilon \le 0.1$:

$$\mathbf{P}\big(|\hat{a}_i - a_i| > \epsilon\,a_i\big) \le 1/4.$$

The upper bound of 1/4 can be made arbitrarily small by the standard method of repeating the process n' times and taking the median result, as follows. Let $\hat{m}$ denote this median. For $1/4 > \delta > 0$, Chernoff bounds (Chernoff, 1952) can be used to show that for $n' \ge 10.4\ln\delta^{-1}$:

$$\mathbf{P}\big(|\hat{m} - a_i| > \epsilon\,a_i\big) \le \delta.$$

The resulting method is presented in pseudocode as Algorithm 3.6 where, as usual, $\lceil x \rceil$ denotes the greatest integer less than or equal to a real number $x \in \mathbb{R}$.
Two notes:

Algorithm 3.6 `Matern_II_product_estimator`

Input: rel. accuracy $\epsilon \in (0, 0.1)$, failure prob. δ, steps k, step size γ
Output: estimate $\hat{m}$ of a_k

$n \leftarrow \lceil 19.36(e-1)k/\epsilon^2 \rceil, n' \leftarrow \lceil 10.4 \ln \delta^{-1} \rceil$
for ℓ from 1 to n' **do**
  for i from 1 to k **do**
    for j from 1 to n **do**
      draw time stamps T_j using $\lambda = i\gamma$
      draw Poisson point proc. W_j w/ intensity $\gamma \cdot (\mu \times m)$ over $\mathcal{S} \times (0, 1]$
    end for
    $\bar{B}_\ell \leftarrow (1/n) \sum_{j=1}^{n} \mathbf{1}_{\{\#(W_j \cap D(x,T_j))=0\}}$.
  end for
  $m_\ell \leftarrow \prod_{i=1}^{k} \bar{B}_\ell^{-1}$
end for
$\hat{m} \leftarrow$ median of $\{m_1, \ldots, m_{n'}\}$

- Upon replacing δ by $\delta' = \delta/k$, the Bonferroni bound ensures that the probability that *all* of the estimates a_j for j from $1 \ldots k$ are correct to within a factor of $1+\epsilon$ will be at least $1-\delta$. Typically the goal is to develop a profile of the posterior across values of λ, which is why all of these products are being estimated rather than using the more efficient method for a single estimate of Štefankovič et al. (2007) .

- The user interested in a specific value of $\lambda > 0$ may either set $\gamma = \lambda/k$ for $k = \lceil \lambda \mu(\mathcal{S}) \rceil$, or may take $\gamma = 1/\mu(\mathcal{S})$ and separately estimate a_λ / a_{k-1} by the same methods, still with expectation at least $\exp(-1)$.

4 Applications

4.1 Spanish towns

Glass & Tobler (1971, Fig. 1) study the spatial distributions of 69 towns in a $40\,\text{mi} \times 40\,\text{mi}$ square in the Spanish Plateau south-east of Madrid, taken from 1944 U.S. Army Map Service topographic sheets. They construct a histogram estimate $\hat{g}(r)$ of the *radial distribution function* $g(r)$, the ratio of the density of points at distance r from an observed point to the overall point density, and compare it to the theoretical radial distribution function for a hard core

model of liquids at equilibrium; they find closest agreement with a density of 0.50 towns per square mile and a hard-core radius of 3.46 mi, but suggest that a soft-core model (where closer towns are unlikely but possible) would be more appropriate. The data were reanalyzed by Ripley (1977, Fig 13 on *p.* 185), who argued that sampling variation made the histogram estimate $\hat{g}(r)$ "too rough to be useful" and instead estimated his reduced second moment measure or K *function* (see Ripley, 1981, *p.* 150 or the Appendix of Ripley, 1977), finding the data consistent with an "interaction model with a distance of $R \approx 3.5$ mi" and a density of 0.5 towns/ mi^2.

Figure 4 shows the likelihood for these data under the Matérn III model. As noted in Section 1, the likelihood function increases monotonically in R up to the minimum inter-town distance, making it simple to evaluate the maximum likelihood estimates. In scaled units (where the 40 mi $\times$ 40 mi square is represented by the unit square) we have $\hat{R} = 0.02451$ and $\hat{\lambda} = 73.7$; as expected for a hard core model, the MLE $\hat{\lambda}$ is larger than would be the case for a pure Poisson process. In the original units, the estimates are $\hat{R} = 0.98$ mi and $\hat{\lambda} = 0.05\,\text{mi}^{-2}$. This value of R is much smaller than that the Ripley's estimate of 3.5 mi because he was matching moments rather than finding likelihoods: our $\hat{R} = 0.98$ mi is simply the minimum distance between two cities.

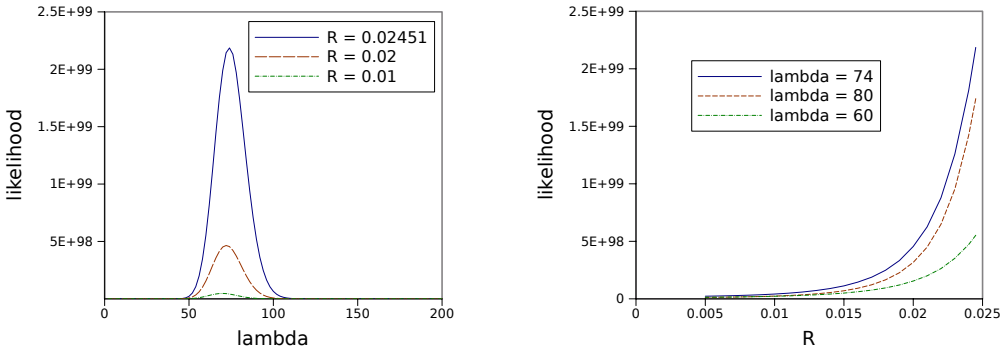


Figure 4: Slices of likelihood function for Spanish town data

4.2 Swedish pines

Ripley (1981, *pp.* 172, 175) studies the locations of 71 Swedish pine saplings

in a $10\,\mathrm{m} \times 10\,\mathrm{m}$ square, using data reported by Strand (1972). He finds the data to be adequately described by a Strauss process (Kelly & Ripley, 1976; Strauss, 1975) with inhibition distance of $R \approx 0.70\,\mathrm{m}$ and proximity penalty of 0.20^t, where t denotes the number of pairs of trees within distance R.

Figure 5 shows the likelihood function for these data under the Matérn III process, which shows the same behavior noted earlier for the trees model. The maximum likelihood estimate occurs at $R = 0.022334$, $\lambda = 73.1$ for the unit square, or $\hat{R} = 0.22\,\mathrm{m}$ and $\hat{\lambda} = 0.731\,\mathrm{m}^{-2}$. Again repulsion increases the estimated value of λ above that of a pure Poisson process.

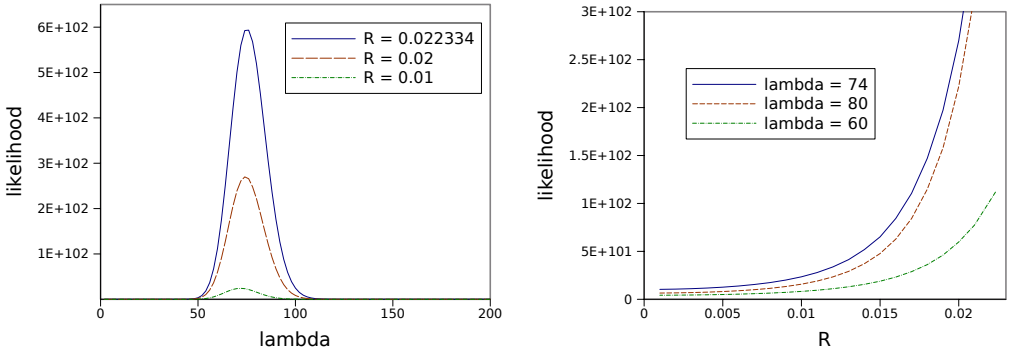


Figure 5: Slices of likelihood function for Swedish pines

4.3 Running Times

Figure 6 shows the average number of Markov chain steps evaluated per sample for the towns and trees datasets, based on 10 000 generated samples. As can be seen, both running times are very close for small values of λ, and then increase nearly linearly for large values of λ, albeit with slightly different slopes.

This illustrates another advantage of modeling with Matérn III processes: the Markov chain appears to be rapidly mixing for all values of λ, in sharp contrast to methods based on pair repulsion models (Kelly & Ripley, 1976), which tend to have phase transitions that make sampling difficult for high values of λ. Partially this is a direct result of features of the model: to make inference with pair repulsion models it is necessary to move point locations around. This is not needed with Matérn III, where the locations are fixed in

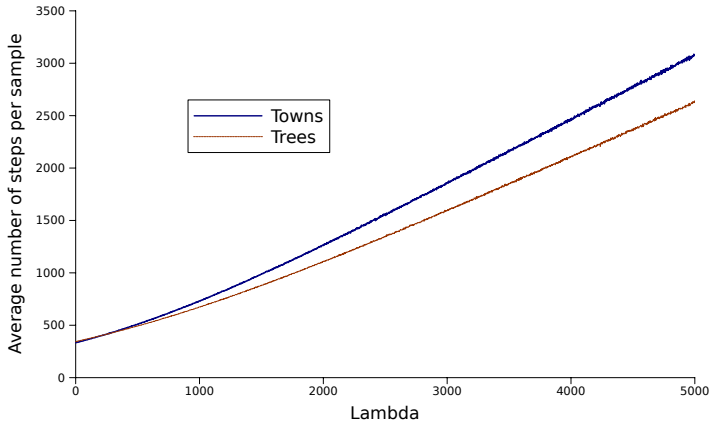


Figure 6: Running times for town and trees datasets

the Markov chain and only the time stamps move. Because the state space is of a completely different type, the issues associated with phase transitions and slow mixing of the Markov chain do not arise.

5 Discussion

The Matérn III process has proven to be an effective tool for modeling repulsive point process data. In this work, we have developed an algorithm for the exact generation of latent time stamps for data, and an algorithm for using these estimates to construct a product estimator for approximating the likelihood function to arbitrarily high accuracy.

There remain several extensions of this method to be considered, including:

- Soft core processes. Here R was constant for all points, but the model becomes far more flexible when R is allowed to vary from point to point.
- In a similar vein, λ can be allowed to vary across the space of interest. This can be accomplished (for instance) by treating the intensity as a linear combination of some basis functions with unknown coefficients.
- Improved proposal distribution for time stamps. In the Markov chain above, the proposed moves for time stamps were uniform. Since the marginal distributions of time-stamps for seen points are skewed, typically smaller than those for the hidden points, it may be more efficient

to sample using a distribution that places more weight towards lower times. Better proposals might improve the mixing properties of the Markov chain, while having the deleterious effect of making the perfect simulation algorithm more complex.

Acknowledgment The authors would like to thank Katja Ickstadt for early conversations about inhibitory processes and Emery Brown, Rob Kass, Fabio Rigat, Jianfeng Feng, and Enrico Rossoni for helpful discussions about spike train data. This work was supported by U.S. National Science Foundation grants DMS–0112069, DMS–0548153, DMS–0635449, DMS–0757549. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

References

Brown, T. (1979). Position dependent and stochastic thinning of point processes. *Stoch. Proc. Appl.* **9**, 189–193.

Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23**, 493–509.

Dyer, M. E. & Frieze, A. M. (1991). Computing the volume of a convex body: A case where randomness provably helps. In *Proceedings of AMS Symposium on Probabilistic Combinatorics and Its Applications*, B. Bollobás, ed., vol. 44 of *Proceedings of Symposia in Applied Mathematics*. Providence, RI: American Mathematical Society.

Fishman, G. S. (1996). *Monte Carlo: Concepts, Algorithms, and Applications*. New York, NY: Springer-Verlag.

Foss, S. & Tweedie, R. L. (1998). Perfect simulation and backward coupling. *Stochastic Models* **14**, 187–204.

Glass, L. & Tobler, W. R. (1971). Uniform distribution of objects in a homogeneous field: Cities on a plain. *Nature* **233**, 67–68.

Häggström, O. & Nelander, K. (1999). On exact simulation from Markov random fields using coupling from the past. *Scand. J. Stat.* **26**, 395–411.

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

Huber, M. (2004). Perfect sampling using bounding chains. *Ann. Appl. Probab.* **14**, 734–753.

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

Matérn, B. (1986). *Spatial Variation*, vol. 36 of *Lecture Notes in Statistics*. New York, NY: Springer-Verlag, 2nd ed. (first edition published 1960 by Statens Skogsforsningsinstitut, Stockholm).

Metropolis, N. C., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087–1092.

Møller, J. (1999). Perfect simulation of conditionally specified models. *J. Roy. Stat. Soc.* B **61**, 251–264.

Ripley, B. D. (1977). Modelling spatial patterns (with discussion). *J. Roy. Stat. Soc.* B **39**, 172–212.

Ripley, B. D. (1981). *Spatial Statistics*. Wiley Series in Probability and Statistics. New York, NY: John Wiley & Sons.

Ripley, B. D. & Kelly, F. P. (1977). Markov point processes. *J. Lond. Math Soc* **15**, 188–192.

Štefankovič, D., Vempala, S. & Vigoda, E. (2007). Adaptive simulated annealing: A near-optimal connection between sampling and counting. In *FOCS 2007: Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science— October 20–23, 2007, Providence, RI*. IEEE.

Stoyan, D. (1979). Interrupted point processes. *Biometrical J.* **21**, 607–610.

Strand, L. (1972). A model for stand growth. In *IUFRO Third Conference Advisory Group of Forest Statisticians*, vol. 72–3 of *I.N.R.A. Publications*. Paris, FR: INRA: Institut National de la Recherche Agronomique. Jouy en Josas, 7–11 Septembre 1970.

Strauss, D. J. (1975). A model for clustering. *Biometrika* **62**, 467–476.

von Neumann, J. (1951). Various techniques used in connection with random digits. In *Monte Carlo Method*, A. S. Householder, G. E. Forsythe & H. H. Germond, eds., vol. 12 of *Applied Mathematics Series*. Washington, D.C.: National Bureau of Standards.

Wilson, D. B. (2000). How to couple from the past using a read-once source of randomness. *Random Struct. Algor.* **16**, 85–113.