Applied Probability Trust (7 April 2009)

LIKELIHOOD-BASED INFERENCE FOR MATÉRN TYPE III REPULSIVE POINT PROCESSES

MARK L. HUBER,* *Duke University*

ROBERT L. WOLPERT,** *Duke University*

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; point process; thinning; underdispersed.

AMS 2000 Subject Classification: Primary 65C60

Secondary 62M30;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

* Postal address: Departments of Mathematics and Statistical Science, Box 90320, Duke University, Durham, NC 27708-0320 USA. Email address : mhuber@math.duke.edu

** Postal address: Department of Statistical Science and Nicholas School of the Environment, Duke University, Durham, NC, U.S.A. Email address: wolpert@stat.duke.edu

considered in Section 5 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 [19, Chapter 3], thinned processes [4, 31], the Strauss process [35], and Markov point processes [27], which model the configuration by assigning a density (to discourage "closeness" of the points) with respect to a homogeneous Poisson point process.

Matérn [19, *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 2 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 3 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 4 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 5 this new method is applied to several examples, and comparisons to earlier work are made.

2. The Matérn Point Processes

Matérn [19, *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 σ-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 distance R, both are removed).

For the Type II process, Matérn assigns independent marks $t_j \overset{\text{iid}}{\sim} \mathsf{Un}(0,1]$ (called 'times') to each of the primary points. The secondary points are those whose time

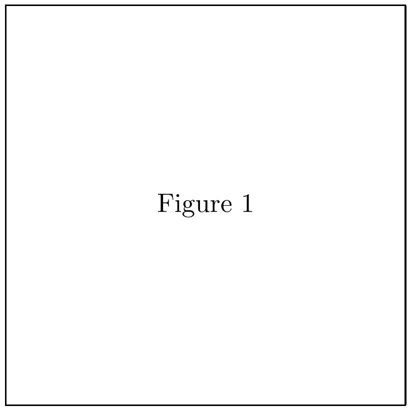


FIGURE 1: Example of Matérn repulsive point processes.

mark 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 [19, *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 μ be a finite measure on a Borel subset $\mathcal{S}$ of $\mathbb{R}^d$. All three Matérn processes begin with a Poisson distributed primary random finite subset $z \sim \mathsf{Po}\big(\lambda\mu(dz)\big)$ of $\mathcal{S}$, and yield random sets $x \subseteq z \subseteq \mathcal{S}$ of secondary points with the property that balls of radius $R/2$ centered at each point do not intersect. Let $\#(s)$ denote the number of points in any set $s \subseteq \mathcal{S}$. Call the $\#(x)$ secondary points $x \subseteq z$ that remain in the process *seen points*, and the $\#(y) = \#(z) - \#(x)$ points $y = z \setminus x$ 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 marks.

Note that the time marks could have been drawn from any continuous distribution on the real numbers [33]. 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 $\{t_i := F(\tilde{t}_i)\}$ 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 marks.

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 $\Lambda_\theta(dz\, dt) := \lambda \cdot \mu(dz)\, m(dt)$, 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 mark. The *seen* points are those for which no other seen point with a lower time mark lies within distance R. Note $\Lambda_\theta\big((\mathcal{S} \times (0, 1]\big) = \lambda \cdot \mu(\mathcal{S})$.

2.1. Connection with RSA

Physicists and chemists have long studied the 'random sequential adsorption' or continuum RSA model for the irreversible binding of proteins to surfaces. In its most common form this model constructs a hard core process as a sequence of points drawn from the uniform distribution on the complements (within some region $\mathcal{S} \subset \mathbb{R}^2$) of the unions of disks of radius R centered at each of the previously-drawn points. When the disks cover $\mathcal{S}$ entirely the "jamming limit" has been reached, whereupon the process halts. Variations include using replacing the disks with squares or other convex shapes, constructing the process in $\mathbb{R}^d$ rather than the plane, employing independent random radii R_i, stopping after a specified number of points have been drawn or attempts have been made, *etc.* For more details and some of the historical development see [6, 8, 9, 24, 28, 29, 36].

If the time marks $\{t_i\}$ are drawn uniformly not from the unit interval but rather from the interval $(0, \lambda(\theta)]$, then the Matérn III process described above may be viewed as the product of thinning (by secondary points) a primary Poisson process on $\mathcal{S} \times (0, \lambda(\theta)]$ with intensity $\mu(dz)\, m(dt)$. For the special case where this intensity is Lebesgue measure on $\mathbb{R}^{d+1}$, the limit as $\lambda(\theta) \to \infty$ yields the RSA process on $\mathbb{R}^d$.

2.2. Likelihood Function for Matern III

Informal estimates of $\{\lambda(\theta), R(\theta)\}$ are simple to find for any of the Matérn processes. The MLE $\hat{R}$ is simply the smallest observed point spacing, $\hat{R} = \min_{i \neq j}\{d(x_i, x_j)\}$, for example, and method of moments estimates are available for λ from empirical estimates of their packing densities (which have simple closed-form relations to λ and R for the

Matérn I and II processes, and a relation available from simulation for Matérn III). Our concern in this paper is quite different— we wish to construct the likelihood function for θ, to support fully coherent inference from the Bayesian or other perspectives [1].

For one-dimensional point sets, a modification of dynamic programming techniques can be used to evaluate the Matérn III likelihood function. The data sets considered in Section 5 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.

For any possible configuration $x = (x_1, \ldots, x_{\#(x)}) \subseteq \mathcal{S}$ of seen points, the vector of time marks $t = (t_1, \ldots, t_{\#(x)})$ 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 mark $t_y > t_i$, since 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 open 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 times t, and $A_\theta(x,t)$ the shadow's *area*, the expected number of hidden points for seen 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 implicitly on R, and $A_\theta(x,t)$ depends on both R and λ.

Van Lieshout [37] introduced a density for RSA (and hence for Matérn III) but the formulation there was given as the product of $\#(x)$ factors, each of which could be difficult to evaluate. Here a new result is presented that gives a closed-form expression for the density in terms of a simple expression of the measure of the shadow.

Theorem 2.1. *The density function of the seen points and their times with respect to the Poisson point process with mean $\mu \otimes m$ is*

$$f_{\text{seen points}}(x, t \mid \theta) = \mathbf{1}_{\{\rho(x) > R\}}\, \lambda^{\#(x)}\, e^{\mu(\mathcal{S})\{1-\lambda\}}\, \exp\{A_\theta(x,t)\} \tag{2.1}$$

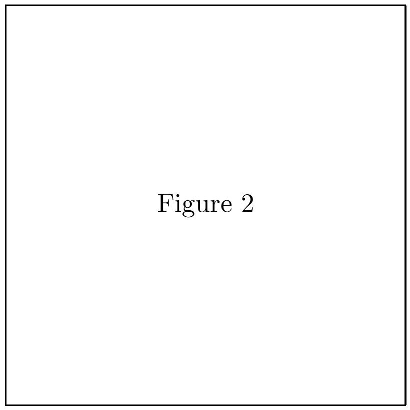


FIGURE 2: Example of a shadow region $D(x,t)$.

where $\rho(x) := \min_{i \neq j} \operatorname{dist}(x_i, x_j)$ *denotes the minimum inter-point distance.*

Proof. There are two ways to create a random draw of the seen points. In the first method, primary points (Z,S) are drawn from the $\mathsf{Po}\big(\lambda(\mu \otimes m)(dz\,ds)\big)$ distribution, and then the seen points are built by thinning as described in Section 2.

Another algorithm is to use acceptance/rejection [38]. Draw a candidate (X,T) from the $\mathsf{Po}\big(\lambda(\mu \otimes m)(dx\,dt)\big)$ distribution and then accept or reject it as a draw from the seen points. If rejection occurs draw again, and continue until a draw is accepted.

To determine if (X,T) should be accepted as a draw from the seen points, use the following procedure. First, draw (Z,S) from $\mathsf{Po}\big(\lambda(\mu \otimes m)(dz\,ds)\big)$ conditioned on the points (X,T) being contained in (Z,S). From a recursive application of Palm distributions [see §4.4 of 32] this can be accomplished by drawing $(Z \backslash X,\ S \backslash T)$ as an independent draw from the $\mathsf{Po}\big(\lambda(\mu \otimes m)(dy\,ds)\big)$ distribution. Next, accept (X,T) if it is exactly the set of seen points in (Z,S)— *i.e.*, if no points of (Z,S) lie outside the shadow of (X,T), an event with probability:

$$\exp(-\lambda(\mu \otimes m)(\mathcal{S} \setminus D(X,T))) = \exp(A_\theta(X,T))\exp(-\mu(\mathcal{S})). \tag{2.2}$$

Also, for (X,T) to be the seen points the minimum distance $\rho(X)$ between any two points in X must exceed R. So the probability of acceptance if $X = x,\ T = t$ is

$$p(x,t) = \mathbf{1}_{\{\rho(x) > R\}} \exp(A_\theta(x,t))\exp(-\lambda\mu(\mathcal{S})). \tag{2.3}$$

The theory of the acceptance/rejection algorithm says the result has a density with respect to the proposal distribution that is proportional to $p(x,t)$:

$$f(x,t \mid \theta) = p(x,t)/p_\theta,$$

where p_θ is the probability that a draw is accepted. So next determine p_θ.

Recall that (Z,S) is the union of two independent draws from $\mathsf{Po}\big(\lambda(\mu\otimes m)(dz\,ds)\big)$. The sum of Poisson random variables is Poisson with parameter equal to the sum of the individual parameters, and the points in (Z,S) are all *i.i.d.*, so $(Z,S) \sim \mathsf{Po}\big(2\lambda(\mu\otimes m)(dz\,ds)\big)$. This has density $2^{\#(z)}\exp(-\lambda\mu(\mathcal{S}))$ with respect to the measure $\mathsf{Po}\big(\lambda(\mu\otimes m)(dz\,ds)\big)$.

Conditioned on the points (Z,S), the points (X,T) are a subset of cardinality $\#(x)$ drawn uniformly at random from (Z,S). This can be checked directly. The probability of drawing any number n of points for (Z,S) from its Poisson point process distribution is $\exp(-2\mu(\mathcal{S}))(2\mu(\mathcal{S}))^n/n!$. This must also be the probability of drawing i seen points (X,T) and $n-i$ hidden points $(Z\backslash X, S\backslash T)$,

$$\frac{\exp(-\mu(\mathcal{S}))\mu(\mathcal{S})^i}{i!} \times \frac{\exp(-\mu(\mathcal{S}))\mu(\mathcal{S})^{n-i}}{(n-i)!} = \frac{\exp(-2\mu(\mathcal{S}))\mu(\mathcal{S})^n}{i!(n-i)!}.$$

Dividing shows the probability of drawing i points from (X,T) conditioned on drawing n points from (Z,S) is $2^{-n}n!/[i!(n-i)!]$, or simply the binomial distribution with parameters n and $1/2$.

This is the same as choosing (X,T) by flipping a fair coin for each point in (Z,S), and putting the points with heads into (X,T). For acceptance to occur, the points chosen to fall into (X,T) must exactly match the points selected from (Z,S) by the Matérn III process. This has probability $1/2^n$ if there are n points in (Z,S). Canceling all the common factors shows that the total probability of acceptance is:

$$p_\theta = \sum_{n=0}^{\infty} \exp(-2\mu(\mathcal{S}))\frac{\mu(\mathcal{S})^n}{n!} = \exp(-\mu(\mathcal{S})), \tag{2.4}$$

thereby showing that the density of (X,T) with respect to $\mathsf{Po}\big(\lambda(\mu\otimes m)(dx\,dt)\big)$ is just

$$f(x,t \mid \theta) = p(x,t)/p_\theta = \mathbf{1}_{\{\rho(x)>R\}}\exp(A_\theta(x,t)). \tag{2.5}$$

Finally, note that a draw (Z,S) from $\mathsf{Po}\big(\lambda(\mu\otimes m)(dz\,ds)\big)$ will have density function $e^{\mu(\mathcal{S})(1-\lambda)}\lambda^{\#(Z)}$ with respect to the $\mathsf{Po}\big((\mu\otimes m)(dz\,ds)\big)$ distribution, which completes the proof.

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

$$g(x \mid \theta) = \mathbf{1}_{\{\rho(x)>R\}}\, e^{\mu(\mathcal{S})\{1-\lambda\}} \lambda^{\#(x)} \int_t \exp\{A_\theta(x,t)\}\; dt \tag{2.6}$$

where "dt" denotes the Lebesgue measure differential $m(dt_1)\dots m(dt_{\#(x)})$ 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, making $g(x \mid \theta)$ an increasing function of R over $\big[0, \rho(x)\big)$. 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 (2.6) 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 marks T for both seen and hidden points, from their conditional distributions (given x and θ); after this, inference about θ can be accomplished using a variety of methods. For use in Section 4.1, we first compute the conditional density function of the time mark vector $T \in (0,1]^{\#(x)}$ for seen points x. From Equation (2.6), this is simply:

$$h_\theta(t \mid x) = Z_{\theta,x}^{-1} \exp\{A_\theta(x,t)\} \tag{2.7}$$

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

3. Acceptance/Rejection

A simple approach for approximating $g(x \mid \theta)$ is the *Acceptance/Rejection* method of von Neumann [38]. This also provides a method to draw hidden points Y and time marks 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 has a geometric distribution with success

---

Algorithm 3.1 `Acceptance_Rejection`

Input: parameter θ, seen points x

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

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

---

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 acceptance probability p decreases (and the expected number of loop steps increases) geometrically in the number $\#\,(x)$ of seen points. Consider, for example, a rectangular region $\mathcal{S} \subseteq \mathbb{R}^d$ for some $d \in \mathbb{N}$, 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 $\nu_d := \pi^{d/2}/\Gamma(1+d/2)$, a lower bound on the *unshadowed* volume leads to an upper bound on the shadowed volume and hence on the acceptance probability:

$$\begin{aligned} \lambda V - A_\theta(x,t) &\ge \lambda v \textstyle\sum_{j=1}^{\#(x)} t_j \qquad \text{so} \\ p := \int_t \exp\{A_\theta(x,t) - \lambda V\}\,dt &\le \int_t \exp\{-\lambda v \textstyle\sum_{j=1}^{\#(x)} t_j\}\,dt \\ &= \rho^{\#(x)} \end{aligned} \tag{3.1}$$

where $\rho := \{1 - \exp\{-\lambda v\}\}/(\lambda v) < 1$. 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:

$$A_\theta(x,t) \ge \lambda v \textstyle\sum_{j=1}^{\#(x)} (1-t_j)$$

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

$$= \rho^{\#(x)} \exp\{-\lambda\{V - v\,\#\,(x)\}\},$$

so Algorithm 3.1 may be very reasonable for sufficiently small $\#\,(x)$. The geometric decrease of Equation (3.1) makes the basic form of Acceptance/Rejection untenable for large $\#\,(x)$; we now turn to an alternative.

4. 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) [39].

Our CFTP approach employs Form 1 of the bounding chains of Huber [15], closely related to the antimonotonicity approach introduced by Kendall [18] and developed by Møller [21] and Häggström and Nelander [13]. 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 marks 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 4.1. 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:

$$\mathrm{P}(A_{t+1} \in A \mid A_t = a, \mathcal{F}_t) = \mathrm{P}(A_{t+1} \in A \mid A_t = a)$$

$$\mathrm{P}(B_{t+1} \in B \mid B_t = b, \mathcal{F}_t) = \mathrm{P}(B_{t+1} \in B \mid B_t = b).$$

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

Definition 4.2. 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

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

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 is $B_0 = \mathcal{T}^{\mathcal{I}}$, 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 $\#\left(B_{\tilde{t}}\right) = 1$, 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 [39]. For completeness a proof is presented similar in form to Wilson's.

Lemma 4.1. *Let $G \sim \mathsf{Ge}(p)$ for some $0 < p \le 1$ (so $\mathrm{P}(G = j) = p(1-p)^j$ for integers $j \ge 0$). Let Y_0 have any distribution, and run the Markov chain forward where the first block is conditioned to be a success, and the next G blocks are conditioned to be failures. Then $Y_{\tilde{t}(1+G)} \sim \pi$.*

Proof. Consider $X_0 \sim \pi$ 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. Note that $Y_{\tilde{t}(1+G)} \sim$

Y, regardless of the distribution of Y_0. 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:

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

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

The coupling between chain and bounding chain is typically accomplished by use of update functions [see, *e.g.*, 11]. 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:

$$\mathrm{P}\left(\phi(x, U) \in A\right) = \mathrm{P}\left(X_1 \in A \mid X_0 = x\right).$$

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 independent 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

$$\Phi(B_t, U) := \cup_{x \in B_t} \ \phi(x, U). \tag{4.1}$$

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

Lemma 4.1 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,

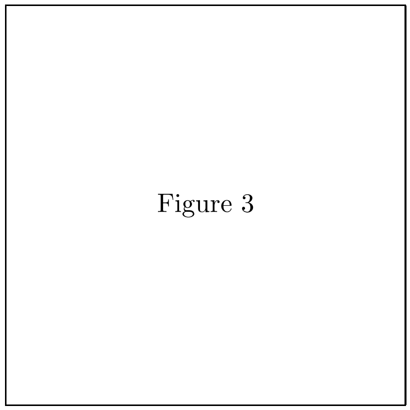


FIGURE 3: Read once coupling from the past.

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.

Algorithm 4.1 `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` $\in \{\text{true}, \text{false}\}$

$B \leftarrow \mathcal{T}^{\mathcal{I}}$
for t from 1 to $\tilde{t}$ **do**
  draw $U \leftarrow \mathsf{Un}(0, 1]$
  $X \leftarrow \phi(X, U)$
  $B \leftarrow \Phi(B, U) := \cup_{x \in B}\ \phi(x, U)$
end for
`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 4.2 `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

4.1. The update function for Matérn time marks

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 marks, given the seen points x, was given in Equation (2.7) for $t \in (0,1]^{\#(x)}$ as $h_\theta(t \mid x) = Z_{\theta,x}^{-1} \exp\{A_\theta(x,t)\}$, where (see Equation (2.1)) $A_\theta(x,t) := \lambda\mu\big(D(x,t)\big)$ is the expected number of hidden points in the shadow of configuration x with time marks 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 [20, 14] 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 4.3 `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:     $T \leftarrow S$
6: **end if**

Our present goal in making inference about the Matérn Type III process is to draw samples of the time marks 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 marks (see Equation (2.7)) may be replaced by the simpler unnormalized version

$$f(t) = \exp\{A_\theta(x,t)\}. \tag{4.2}$$

Consider a proposal chain in which an index $i \in \{1,\ldots,\#(x)\}$ is chosen uniformly and then a new time mark 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 marks t_{prop}. By symmetry, the Metropolis-Hastings ratio becomes simply:

$$r = \exp\{A_\theta(x,t_{\text{prop}}) - A_\theta(x,t)\}. \tag{4.3}$$

Calculating the change in shadow area under the new time mark 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 4.3.

In the case that the proposed new time-mark is smaller, $t_{\text{prop}} \leq t$, the shadow of x will grow ($A_\theta(x, t_{\text{prop}}) \geq A_\theta(x, t)$) 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 $t_{\text{prop}} > t$, the Hastings ratio r of Equation (4.3) is *exactly* the probability that a Poisson point process with intensity Λ_θ will assign zero points to the newly unshadowed region $\Delta := D(x, t) \setminus D(x, t_{\text{prop}})$. Thus, in lieu of lines 2–3, generate a Poisson point process $W \sim \mathsf{Po}(\Lambda_\theta)$ on Δ, and set $B \leftarrow \mathbf{1}_{\{0=\#(W)\}}$. 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 mark 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. This idea of using an auxiliary Poisson process to generate a Bernoulli without the need to calculate the parameter exactly also appears in [3], where it was used in the very different context of generating diffusion sample paths.

For later use in Section 4.2 we present as Algorithm 4.4 a method for drawing a Metropolis-Hastings update of the vector T of time marks for a configuration x of seen points.

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

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

Pseudocode for the bounding chain update is presented in Algorithm 4.5, followed by a brief explanation of why it works.

Algorithm 4.4 `Matern_III_Metropolis_Hastings`

Input: seen points x with current state of time marks T

Output: next state of time marks 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

Algorithm 4.5 `Matern_III_Bounding_Chain`

Input: seen points x, current bounds $B = [T^{\text{lo}}, T^{\text{hi}}]$

Output: next state of time mark 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 $t_i \le T_i^{\text{lo}}$ **then** {Case I}
$T_i^{\text{lo}} \leftarrow t_i, T_i^{\text{hi}} \leftarrow t_i$
else if $T_i^{\text{lo}} < t_i \le T_i^{\text{hi}}$ **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 $\#\big(W \cap \{D(x, t^1) \setminus D(x, t^3)\}\big) = 0$ **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 $\#\big(W \cap \{D(x, t^1) \setminus D(x, t^3)\}\big) = 0$ **then**
$T_i^{\text{lo}} \leftarrow t_i$
end if
if $\#\big(W \cap \{D(x, t^2) \setminus D(x, t^4)\}\big) = 0$ **then**
$T_i^{\text{hi}} \leftarrow t_i$
end if
end if

To summarize what is happening: when the proposed time mark 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 marks are as small as possible (*i.e.*, equal to T_i^{lo}), and for all $j \neq i$ the time marks 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 4.1. *Algorithm 4.5 is a valid update for the bounding chain.*

Proof. Suppose that the call

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

is used for the bounding chain and

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

for the time marks, 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 4.5, 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:

$$\Delta = \{D(x, T) \setminus D(x, t)\} = \big\{B_R(x_i) \times (T_i, t_i]\big\} \cap D(x, t)^c. \tag{4.4}$$

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: $T_i^{\mathrm{hi}} < t_i$**.** The case for $T'^{\,\mathrm{lo}}$ is the same as in Case II. It remains to show that the move for $T'^{\,\mathrm{hi}}$ is valid. Use representation (4.4) above for Δ. This is as small as possible when $(T_i, t_i]$ is as small as possible, so $T_i = T_i^{\mathrm{hi}}$. Similarly, $D(x,t)^c$ is as small as possible when $D(x,t)$ is as large as possible, which happens when $t_j = T_j^{\mathrm{lo}}$ 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 $\#\big(W \cap \{D(x,t^2) \setminus D(x,t^4)\}\big) > 0$, then $\#\left(W \cap \{D(x,T) \setminus D(x,t)\}\right) > 0$ for all T satisfying $T^{\mathrm{lo}} \leq T \leq T^{\mathrm{hi}}$, so the proposal t will be rejected and $T_i'^{\,\mathrm{hi}} = T_i^{\mathrm{hi}} \geq T_i' = T_i$ is fine. Otherwise, there might be some T that accepts the move, and so $T_i'^{\,\mathrm{hi}} = t_i \geq T_i'$ as in line 21.

This completes the proof.

4.2. 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 [10, §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 $A_{\theta_k} = k A_{\theta_1} = k\,\gamma\,(\mu{\otimes}m)\big\{D(x,t)\big\}$.

From Equation (2.6) 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 4.4 (or 3.1), draw time marks $T_1, \ldots, T_n \overset{\mathrm{iid}}{\sim}$

$h_{\theta_i}(t \mid x)\, dt$ (see Equation (2.7)) for intensity $\lambda_i = i\gamma$. For each $1 \le j \le n$, draw a Poisson point process W_j with intensity $\gamma(\mu \otimes m)(ds\, dt)$ over the shadow $D(x, T_j)$. The independent Bernoulli random variables $B_j = \mathbf{1}_{\{\#(W_j)=0\}}$ have conditional expectation $\mathrm{E}(B_j \mid T_j) = \mathrm{P}\{\#\,(W_j) = 0 \mid T_j\} = \exp\{-A_{\theta_1}(x, T_j)\}$, so

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

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 $\mathrm{E}(B_j) = p_i > e^{-1}$ 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 $\mathrm{E}(\bar{B}_i) = (a_{i-1}/a_i)$ and $\mathrm{cov}(\bar{B}_i)^2 \le (e-1)/n$. Let $\hat{a}_i := 1/\prod_{j=1}^{i} \bar{B}_j$ be an estimate for a_i. Dyer and Frieze [7, *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

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

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 $\epsilon \le 0.1$:

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

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 [5] can be used to show that for $n' \ge 10.4 \ln \delta^{-1}$:

$$\mathrm{P}\left(|\hat{m} - a_i| > \epsilon\, a_i\right) \le \delta.$$

The resulting method is presented in pseudocode as Algorithm 4.6 where, as usual, $\lceil x \rceil$ denotes the greatest integer less than or equal to a real number $x \in \mathbb{R}$.

Two notes:

- 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 $1+\epsilon$

Algorithm 4.6 `Matern_III_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 marks T_j using $\lambda = i\gamma$
      draw Poisson point proc. W_j w/ intensity $\gamma \cdot (\mu \otimes m)$ 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 $\{m_1, \ldots, m_{n'}\}$

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 $\exp(-1)$.

5. Applications

5.1. Spanish towns

Glass and Tobler [12, Fig. 1] employ a Gibbs hard core process to study the spatial distributions of 69 towns in a $40\,\mathrm{mi} \times 40\,\mathrm{mi}$ 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 [25, 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 26, *p.* 150 or the Appendix of 25), finding the data consistent with an "interaction model with a distance of $R \approx 3.5\,\text{mi}$" and a density of 0.5 towns/ mi^2, and by Illian *et al.* [16, Example 3.14 on *p.* 168] who "estimate visually" the interaction distance to be $R \approx 3.5\,\text{mi}$ from empirical plots of the K function, and find $R \approx 2.16$ from their "minimum contrast method."

Figure 4 shows the likelihood for these data under the Matérn III model. As noted in Section 2, 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 $40\,\text{mi} \times 40\,\text{mi}$ 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 $\hat{R} = 0.98\,\text{mi}$ 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 $\hat{R} = 0.98\,\text{mi}$ is simply the minimum distance between two cities. Arguably the parameter R has a different meaning for a hard core process such as ours than for a soft core (e.g. Gibbs) approach.

The purpose of the results presented here is not model selection, but rather to show how likelihood based inference is possible for Matérn III. That being said, it is helpful in testing the algorithm to know whether or not the model is reasonable for this data. Figure 5 shows the estimated Besag's L-function [2] (a simple transformation of Ripley's K) for the towns data, as well as a draw from the model at the values of the parameters that maximize the likelihood [see 16, *pp.* 231–232 for details of the estimation]. The fit is poor for small values of R, as expected for a hard-core model, but improves and matches the data almost perfectly for $R > 0.1$ (in scaled units, or about 4 mi).

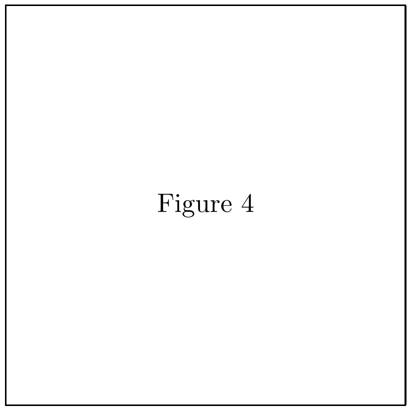


FIGURE 4: Slices of likelihood function for Spanish town data.

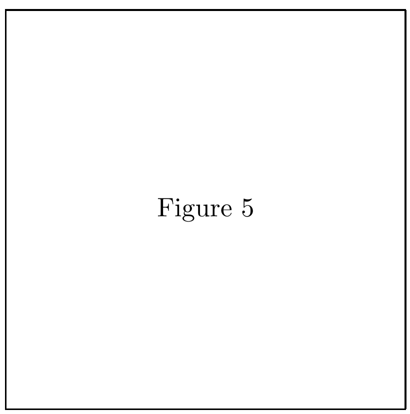


FIGURE 5: L-function for Spanish town data and model at MLE.

5.2. Swedish pines

Ripley [26, *pp.* 172, 175] studies the locations of 71 Swedish pine saplings in a $10\,\text{m} \times 10\,\text{m}$ square, using data reported by Strand [34]. He finds the data to be adequately described by a Strauss process [17, 35] with inhibition distance of $R \approx 0.70\,\text{m}$ and proximity penalty of $(0.20)^t$, where t denotes the number of pairs of trees within distance R.

Figure 6 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 $\hat{R} = 0.22\,\text{m}$ and $\hat{\lambda} = 0.731\,\text{m}^{-2}$. Again repulsion increases the estimated value of λ above that of

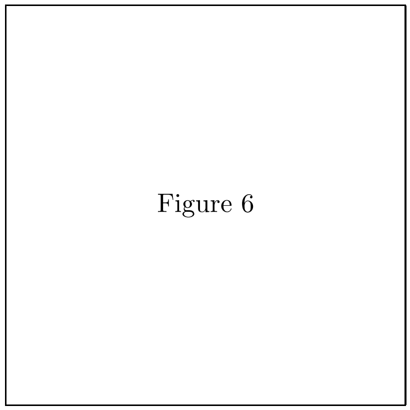


FIGURE 6: Slices of likelihood function for Swedish pines.

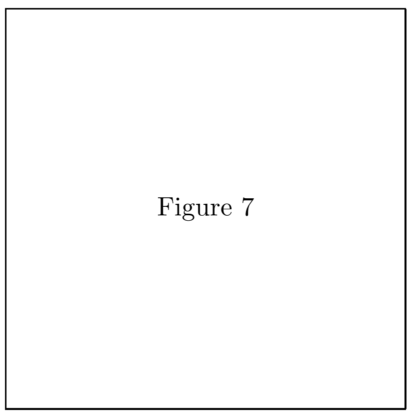


FIGURE 7: L-function for Swedish pines data and MLE.

a pure Poisson process.

As with the towns data, it is helpful in estimating the behavior of the algorithm to know if the data are reasonably consistent with the model. Figure 7 shows how well the estimated L-function matches that of a draw from the model evaluated at the MLE. Here the overall shape of the L-function is captured by the Matérn III model, but there is noticeable deviation from the curve for both small and large values of R.

5.3. Running Times

Figure 8 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

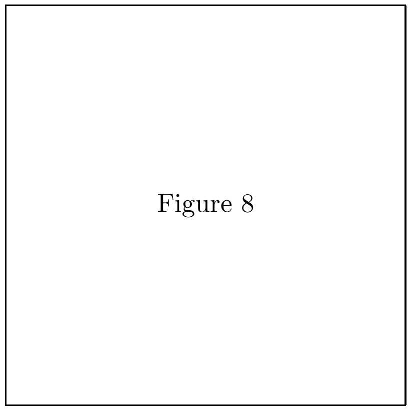


Figure 8: Running times for town and trees datasets.

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 [17], which tend to have phase transitions that make sampling difficult for high values of λ. This is in part a direct result of features of the model: for simulation based 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 the Markov chain and only the latent time marks 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.

6. Discussion

Despite notable recent progress in the study of Gibbs random fields [22, 23], the problem of making likelihood-based inference for repulsive point processes remains challenging. In the absence of calculable likelihood functions, the common approach to inference is to simulate random samples from a specified class of processes and compute summary statistics such as the first-order packing density and second-order K, L, or g functions, and to compare these with similar statistics computed from data. Similar values of these statistics (or similar features of their plots) are regarded as evidence of model fit, while the inability to reconstruct features of the observed data is regarded as evidence against the suitability of a particular model family. Our goal has been to support quantitative, likelihood based inference for some family of repulsive point processes to enable investigators to generate both Bayesian and Maximum Likelihood estimators of model parameters, or to compute Bayesian predictive distributions of as yet unobserved quantities, within either of the present dominant paradigms of statistical inference.

A host of interesting models have been proposed in the past half century for repulsive point processes (simple inhibition processes, thinned processes, area interaction processes, the Strauss process, Markov point processes, *etc.*), but most resist our efforts to calculate likelihood functions. One notable exception is the Matérn III process, which has proven to be an effective tool for modeling repulsive point process data.

In this work, we have developed and presented an algorithm for the perfect simulation of latent time marks for data under the Matérn III model, 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 basis functions with unknown coefficients.
- Improved proposal distribution for time marks. The Metropolis proposals for time marks used in Section 4.1 were uniformly distributed on the unit interval. Since the marginal distributions of time-marks for seen points are skewed, typically smaller than those for the hidden points, it may be more efficient to sample using a proposal 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.

Acknowledgements

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

[1] Berger, J. O. and Wolpert, R. L. (1988). *The Likelihood Principle: A Review, Generalizations, and Statistical Implications (with discussion)* second ed. vol. 6 of *IMS Lecture Notes-Monograph Series*. Inst. Math. Statist., Hayward, CA.

[2] Besag, J. E. (1977). Discussion of Dr. Ripley's paper [25]. *J. Roy. Stat. Soc.* B **39,** 193–195.

[3] Beskos, A., Papaspiliopoulos, O. and Roberts, G. O. (2006). Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli* **12,** 1077–1098.

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

[5] 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.

[6] Döge, G. (2001). Perfect simulation for random sequential adsorption of d-dimensional spheres with random radii. *Journal of Statistical Computation and Simulation* **69,** 141–156.

[7] Dyer, M. E. and 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*. ed. B. Bollobás. vol. 44

of *Proceedings of Symposia in Applied Mathematics.* American Mathematical Society, Providence, RI. pp. 123–170.

[8] Feder, J. (1980). Random sequential adsorption. *Journal of Theoretical Biology* **87,** 273–254.

[9] Finegold, L. and Donnell, J. T. (1979). Maximum density of random placing of membrane particles. *Nature* **278,** 443–445.

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

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

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

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

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

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

[16] Illian, J., Penttinen, A., Stoyan, H. and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns.* Statistics in Practice. John Wiley & Sons, Chichester, West Sussex, UK.

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

[18] Kendall, W. S. (1991). Perfect simulation for the area-interaction point process. In *Probability Towards 2000.* ed. L. Accardi and C. C. Heyde. vol. 128 of *Lecture Notes in Statistics.* Springer-Verlag, New York, NY pp. 218–234. Symposium volume from 1995 workshop at Center for Applied Probability, Columbia University.

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

[20] Metropolis, N. C., 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,** 1087–1092.

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

[22] Møller, J., Pettitt, A. N., Berthelsen, K. K. and Reeves, R. W. (2006). An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. *Biometrika* **93,** 451–458.

[23] Murray, I., Ghahramani, Z. and MacKay, D. J. C. (2006). MCMC for doubly-intractable distributions. In *Proceedings of the* 22^{nd} *Annual Conference on Uncertainty in Artificial Intelligence (UAI-06).* ed. R. Dechter and T. Richardson. AUAI Press, Arlington, VA.

[24] Pálasti, I. (1960). On some random space filling problems. *Publ. Math. Inst. Hungar. Acad. Sci.* **5,** 353–359.

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

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

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

[28] Schaaf, P., Voegel, J.-C. and Senger, B. (1998). Irreversible deposition/adsorption processes on solid surfaces. *Annales de Physique* **23,** 1–89.

[29] Solomon, H. (1967). Random packing density. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability.* ed. L. M. Le Cam

and J. Neyman. vol. 3: Physical Sciences. University of California Press, Berkeley, CA pp. 119–134.

[30] Štefankovič, D., Vempala, S. and 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, Los Alamitos, CA. pp. 183–193.

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

[32] Stoyan, D., Kendall, W. S. and Mecke, J. (1995). *Stochastic Geometry and Its Applications* second ed. John Wiley & Sons, Chichester, West Sussex, UK.

[33] Stoyan, D. and Stoyan, H. (1985). On one of Matérn's hard-core point process models. *Mathematische Nachrichten* **122,** 205–214.

[34] 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.* INRA: Institut National de la Recherche Agronomique, Paris, FR. pp. 207–216. Jouy en Josas, 7–11 Septembre 1970.

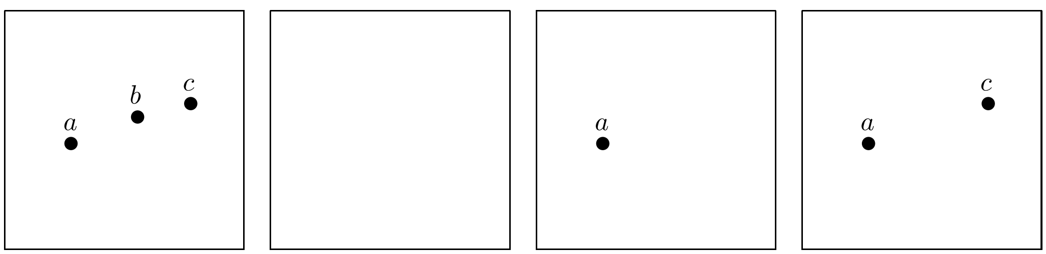
[35] Strauss, D. J. (1975). A model for clustering. *Biometrika* **62,** 467–476.

[36] Tanemura, M. (1979). On random complete packing by discs. *Annals of the Institute of Statistical Mathematics B* **31,** 351–365.

[37] van Lieshout, M.-C. N. M. (2006). Maximum likelihood estimation for random sequential adsorption. *Adv. Appl. Prob.* **38,** 889–898.

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

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



Primary Poisson process	Matérn I	Matérn II	Matérn III
$R = 0.4$	pair a and b eliminated;	$t_a < t_b$: a eliminates b;	$t_a < t_b$: a eliminates b;
$\mathrm{dist}(a, b) \approx 0.30$	pair b and c eliminated;	$t_b < t_c$: b eliminates c;	since b eliminated, it does
$\mathrm{dist}(b, c) \approx 0.23$	No points remain.	Only point a remains.	not eliminate c.
$\mathrm{dist}(a, c) \approx 0.53$			Points a and c remain.
$t_a \approx 0.27,\ t_b \approx 0.52\ t_c \approx 0.78$			

FIGURE 1: Example of Matérn repulsive point processes.

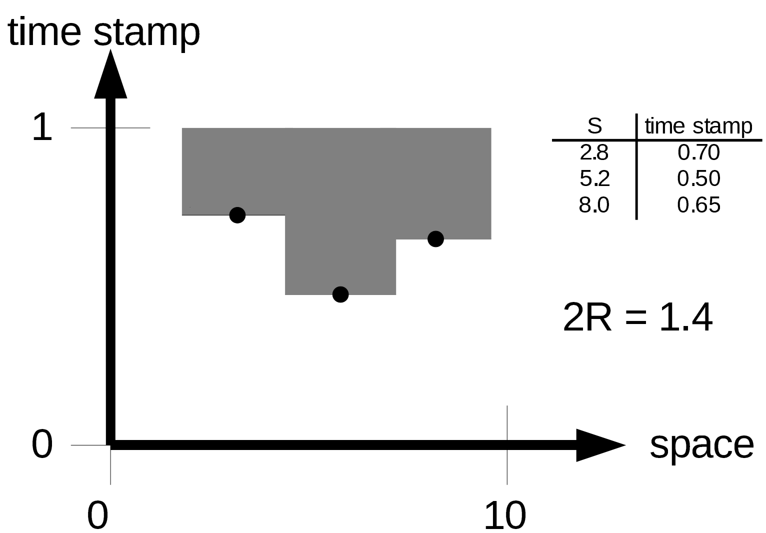


FIGURE 2: Example of a shadow region $D(x, t)$.

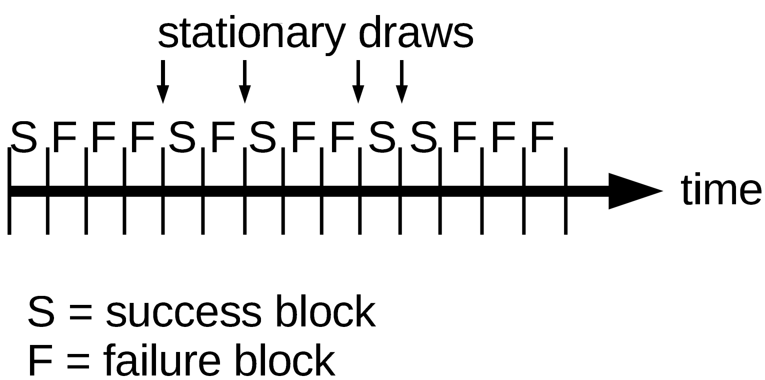


Figure 3: Read once coupling from the past

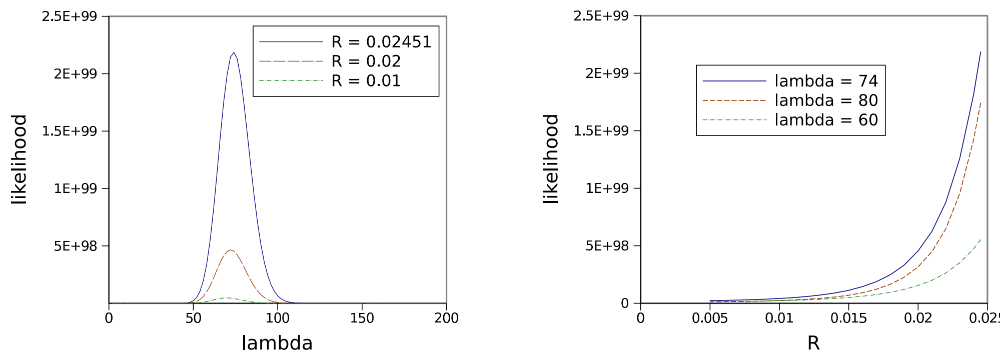


FIGURE 4: Slices of likelihood function for Spanish town data

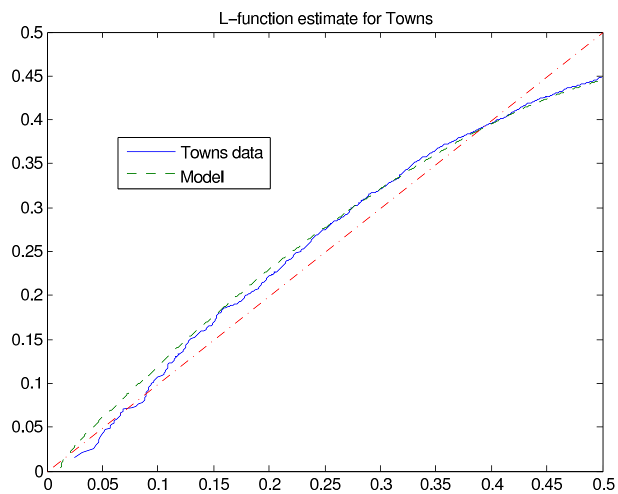


FIGURE 5: L-function for Spanish town data and model at MLE

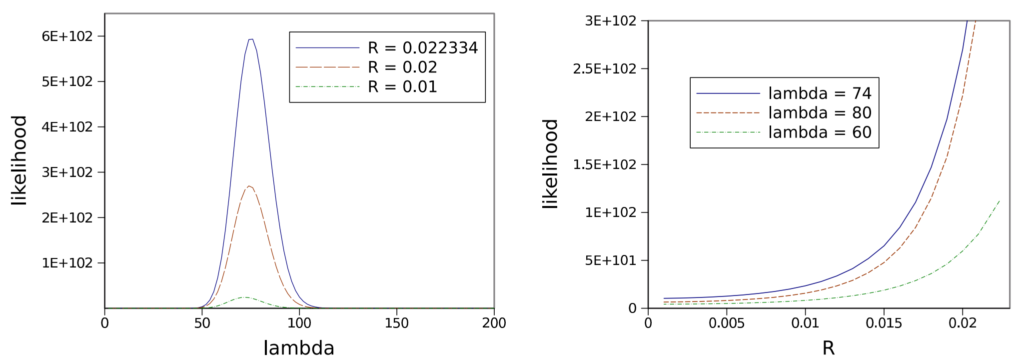


Figure 6: Slices of likelihood function for Swedish pines

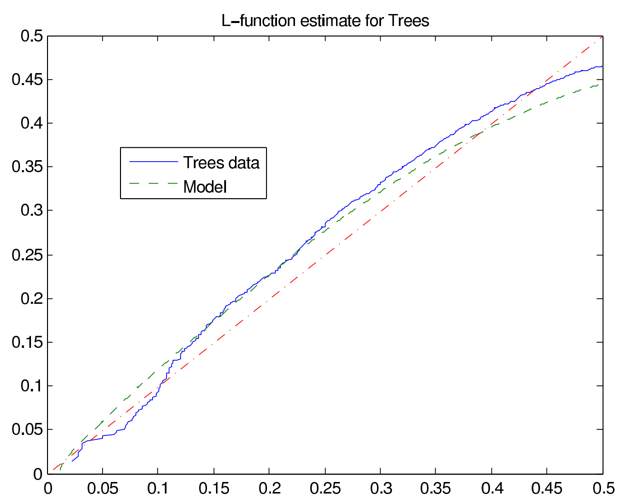


FIGURE 7: L-function for Swedish pines and model at MLE

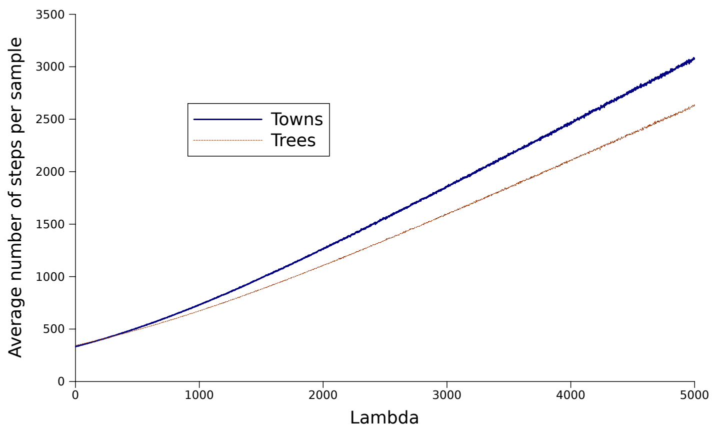


FIGURE 8: Running times for town and trees datasets