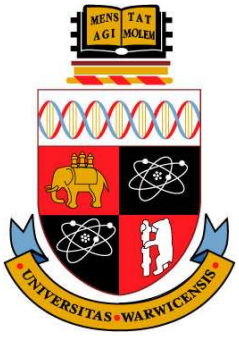

Perfect Simulation of Conditional & Weighted Models

by

Sandeep R. Shah

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Statistics Department

May 2004

THE UNIVERSITY OF
WARWICK

Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank my supervisor Professor Wilfrid S. Kendall for his continued help and guidance throughout my Ph.D. This thesis was funded by an EPSRC studentship (No. 0080183X) and I gratefully acknowledge the financial support. Finally, to my parents and brother for their continued love and support.

Declarations

I declare that this thesis is based on my own research and in accordance with the regulations of the University of Warwick. The work is original except where indicated by specific references in the text. Chapter 2 is based on the following research report and subsequent publication:

Shah S.R. (2003). Stochastic domination and conditional thinning in spatial point processes. Technical Report 412, University of Warwick, UK.

Shah S.R. (2003). A note on stochastic domination and conditional thinning. *Advances in Applied Probability (SGSA)* 5(4), 937–940.

This thesis has not been submitted for examination to any other university.

Abstract

This thesis is about probabilistic simulation techniques. Specifically we consider the *exact* or *perfect* sampling of spatial point process models via the *dominated CFTP* protocol. Fundamental among point process models is the Poisson process, which formalises the notion of complete spatial randomness; synonymous with the Poisson process is the Boolean model. The models treated here are the conditional Boolean model and the area-interaction process. The latter is obtained by weighting a Poisson process according to the area of its associated Boolean model.

A fundamental tool employed in the perfect simulation of point processes are spatial birth-death processes. Perfect sampling algorithms for the conditional Boolean and area-interaction models are described. Birth-death processes are also employed in order to develop an exact *omnithermal* algorithm for the area-interaction process. This enables the simultaneous sampling of the process for a whole range of parameter values using a single realization. A variant of Rejection sampling, namely 2-Stage Rejection, and exact Gibbs samplers for the conditional Boolean and area-interaction processes are also developed here.

A quantitative comparison of the methods employing 2-Stage Rejection, spatial birth-death processes and Gibbs samplers is carried, the performance measured by actual run times of the algorithms. Validation of the perfect simulation algorithms is carried out via χ^2 tests.

Introduction

The huge increase in computing power over the last 25 years has had a profound effect on statistical methodology and applied probabilistic modelling. There have been numerous developments and application of simulation methods in various fields. Realistic models for real-world phenomenon usually involve a wide variety of complexity on high (or even infinite) dimensional state spaces. Mathematical analysis of such models is therefore difficult, if not impossible. However the onset of ever more powerful computers has made it possible to examine such models via *stochastic simulation*: (stochastic) realizations of the models can be obtained and statistical analysis carried out. This has spurred a whole new generation of stochastic simulation algorithms; one in particular, Markov Chain Monte Carlo (MCMC), is now widely used and recognized as a powerful tool in the statistics community. Its origins can be traced back to statistical physics (Metropolis *et al.* 1953), and concerns the simulation of models via Markov chains or processes.

The essence of MCMC involves the construction of a Markov chain which *converges* to a stochastic realization of the model being studied. This class of algorithms has had numerous applications in statistical physics, image analysis, Bayesian statistics and, more recently, spatial modelling and stochastic geometry. However there is a basic set back with MCMC: in practice one has to settle for *approximate* samples, since a priori it is difficult in most cases to determine when the Markov chain will have converged. However it has been discovered that, for some chains, it is possible to modify the MCMC algorithm such that it automatically signals when convergence has been achieved.Such variants, which deliver *exact* samples, have been coined *perfect* or *exact simulation* algorithms.

Indeed a number of models, including the Ising, q-state Potts and random-cluster models from statistical physics and point process models in stochastic geometry have been treated using perfect simulation (Propp & Wilson 1996; Fill 1998; Kendall 1997b; Møller 2001). Stochastic geometry, the roots of which can be traced back to integral geometry and geometrical probability (Stoyan *et al.*

1995), concerns the modelling of phenomena that arise as complicated geometrical shapes or patterns. Furthermore, point processes provide plausible models for collections of individuals or events, such as plants, animals, stars, structure of biological cells and rock sections, disease outbreaks and earthquakes. For example Figure 1 below depicts the locations of the Redwood seedlings data which was originally studied by Strauss (1975), and later by Ripley (1977). It is clear from the figure that there is an underlying 'clustering' mechanism generating this data. Therefore a natural way to model these seedlings would be to assume that there is an *unobserved 'parent' process* which gives rise to the *observed 'daughter' process*. The objective then would be to sample the parent process *given* the observed daughter process.

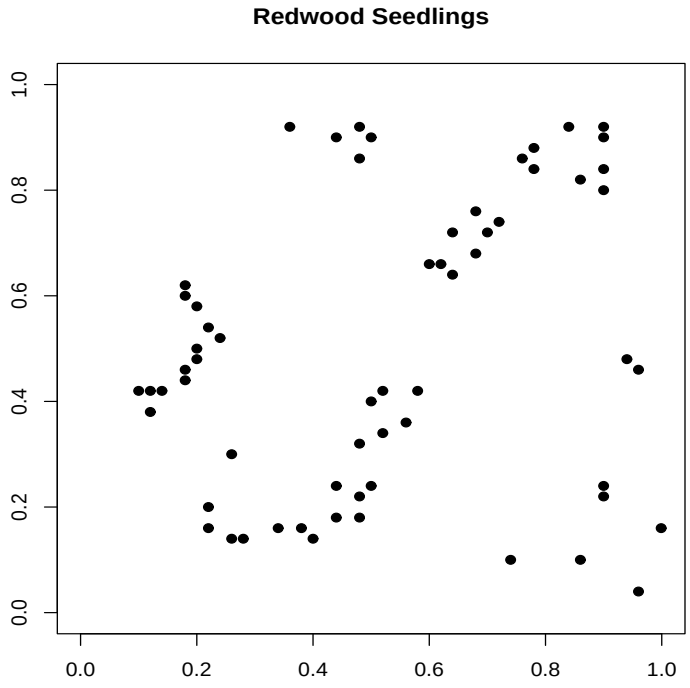

Figure 1: Positions of the Redwood seedlings; the data was obtained by kind permission from Peter Diggle's web page `http://www.maths.lancs.ac.uk/~diggle/`.

Perfect simulation has developed rapidly since its inception. For example *Coupling From The Past (CFTP)* is probably the most popular and widely studied perfect simulation algorithm. It was introduced by Propp & Wilson (1996) for Markov chains on finite state spaces. Within the next few years it was generalized to continuous and even unbounded state spaces (Murdoch & Green 1998; Kendall 1998; Häggström *et al.* 1999; Kendall & Thönnes 1999; Green & Murdoch 1999; Kendall & Møller 2000; Berthelsen & Møller 2002b). It has also been the focus of several doctoral

theses (Wilson 1996; Thönnes 1998; Dimakos 2000; Ambler 2002). While providing solutions to problems of convergence and approximates samples, it also poses more challenges to the researcher. This thesis is another addition to the simulation literature, aiming to introduce and develop more perfect simulation algorithms for point processes.

Overview of Thesis

This thesis comprises five main chapters, a concluding chapter and two appendices. We consider the perfect sampling of the conditional Boolean model, area-interaction and conditional area-interaction process. An investigation of the Van Lieshout & Van Zwet (2001) algorithm for conditional Boolean models, which concludes that their method is actually biased, is first presented. Following this three perfect sampling procedures for the conditional Boolean and conditional area-interaction models are described: a 2-Stage Rejection method, one employing spatial birth-death processes and an exact Gibbs sampler. For the area-interaction process, several existing algorithms are reviewed and the possibility of *omnithermal* sampling for the process is also considered.

The basic theory and foundations for point processes and their simulation is introduced in Chapter 1. A natural way to sample point processes is via spatial birth-death processes, introduced by Preston (1977); indeed a number of CFTP-based perfect algorithms employ birth-death processes (Kendall 1998; Kendall & Thönnes 1999; Berthelsen & Møller 2002b; Fernández *et al.* 2002). A description of how point processes can be simulated via spatial birth-death processes is given and illustrated by simple examples. Chapter 1 also introduces the ideas behind Markov Chain Monte Carlo (MCMC) and perfect simulation, including a description of the Coupling From The Past (CFTP) and Dominated CFTP (domCFTP) protocols.

One of the first simulation algorithms we investigated was that proposed by Van Lieshout & Van Zwet (2001) for sampling point processes conditioned to satisfy some event. Their method was intended to be perfect, ie. producing an unbiased sample in finite time. However our investigation uncovered an error and it was concluded that the method produces biased samples (Shah 2003a; Van Lieshout & Van Zwet 2003). This motivated the consideration of *stochastic domination* in a point process context in order to quantify the nature of the bias. Preston (1977, Theorem 9.1), which uses coupled birth-death processes in order to provide conditions for stochastic dominance, is extended

to the case of *strict* stochastic dominance (Corollary 2.1). This allows the bias of the Van Lieshout & Van Zwet algorithm to be quantified in some cases.

Chapter 3 then considers various other perfect simulation algorithms for conditional point processes. A 2-Stage Rejection algorithm is developed and shown to be more efficient than ordinary Rejection. The Cai & Kendall (2002) algorithm is described in Section 3.4.1. The original formulation employs immigration-death processes on the integers; the description here is phrased, equivalently, in terms of spatial birth-death processes. Gibbs sampling for conditional point processes has not been considered before in the literature; therefore Section 3.5 describes a CFTP-based Gibbs sampler. The coupling construction vaguely resembles that of Häggström *et al.* (1999), as their notion of *quasi-minimal* and *quasi-maximal* elements is employed in order to devise the required coupling of 'bounding' processes. Results of simulation experiments, aimed at comparing the performance of the 2-Stage Rejection, Cai & Kendall and exact Gibbs algorithms, are presented in Section 3.7. The basic conclusion is that 2-Stage Rejection is very competitive for moderate parameter values, whereas the Cai & Kendall algorithm performs well for extreme values. The exact Gibbs sampler, unfortunately, is always outperformed.

The area-interaction process of Baddeley & Van Lieshout (1995) is the focus of Chapter 4. Rejection, Gibbs sampling and two methods (Kendall 1998; Fernández *et al.* 2002), which use birth-death processes, are reviewed. Simulation via spatial birth-death processes involves computation of areas of irregular regions in order to determine appropriate acceptance probabilities. Here, we develop the 'cluster' trick of Kendall (1997a), yielding the first implementation of the Kendall (1998) algorithm which employs this trick. The algorithms considered so far sample the area-interaction process for *fixed* model parameters. The emphasis in Section 4.6 is on *omnithermal sampling*, where the process is sampled for a *whole range* of parameter values simultaneously (cf. Propp & Wilson 1996 and the references therein, Wilson 1996, Dimakos 2000 and Grimmett 1995). The area-interaction process has not been the study of such a sampling procedure; the work here has yielded an exact omnithermal algorithm for this point process.

Chapter 5 combines the ideas of Chapters 3 & 4 and introduces the 'conditional area-interaction process'. We incorporate the Cai & Kendall (2002) and Kendall (1998) algorithms in order to define a domCFTP construction. In addition a 2-Stage Rejection procedure and an exact Gibbs-within-Metropolis Hastings sampler are also developed. A quantitative evaluation of the actual run

times of these three procedures is carried out. The results indicate that the modified Cai & Kendall algorithm performs well for extreme model parameters, while the 2-Stage and Gibbs algorithms are competitive for moderate parameter values. It is not the case now that the Gibbs sampler is always outperformed by the other two; it is as efficient, if not more so, than the 2-Stage algorithm for all parameter values.

Finally, Chapter 6 summarizes what has been achieved in this thesis and possibilities for further work are outlined; in particular we comment on the possibility of omnithermal sampling for point processes in general, as well as multi-parameter omnithermal sampling. The experience with the Van Lieshout & Van Zwet (2001) algorithm emphasizes the necessity to validate any implementation of a perfect simulation algorithm, so as to ensure there are no coding errors or theoretical discrepancies. Therefore the perfect simulation algorithms of Chapters 3 & 5 respectively are validated via χ^2 tests, which are presented in Appendices A & B.

Chapter 1

Mathematical Foundations

1.1 Stochastic Geometry & Point Processes

Stochastic Geometry refers to that area of Mathematical research which provides suitable models and statistical methods for analyzing data representing complicated geometrical patterns. Examples of such patterns include the (spatial) distribution of stars and galaxies, locations of trees in a forest, features of biological tissues, rock sections studied in geology, oil reservoirs and ore deposits. The modern theory of Stochastic Geometry was initiated by D.G. Kendall, K. Krickeberg and R.E. Miles (see Stoyan *et al.* 1995 for a variety of literature on the historical origins and recent advances; D.G. Kendall's foreword provides an anecdotal glimpse into the origins of the subject).

The most basic geometrical objects are points, and collections of random point patterns or 'point processes' play a fundamental role in stochastic geometry. More complicated objects can be built from a collection of points, eg. the Boolean model (Definition 1.10 or 3.1); in addition the theory of line and surface processes generalizes the notion of point patterns to higher dimensions and is useful for modelling applications in geology, stereology and medicine (a rigorous treatment is given in Stoyan *et al.* 1995). The main aim of this thesis is to present new and existing simulation algorithms for some point process models.

A spatial point pattern on some space $\mathcal{X}$ can be viewed as a *finite* unordered set or 'configuration' of points $x = \{\xi_1, \ldots, \xi_n\}$ for $n \in \mathbb{N}$, $\xi_i \in \mathcal{X}$. Point patterns modelling different physical phenomenon could depict the spatial distribution of either objects (trees, stars, etc) or events (outbreak of disease, earthquakes, etc); however elements of a pattern will simply be referred to as

either 'points' or 'individuals'. Point processes can also be viewed as random counting measures or random variables taking values in what is called an *Exponential Space* (Carter & Prenter 1972). Rigorous theoretical treatments of point processes have predominantly been made in terms of the random measure viewpoint (see for example Daley & Vere-Jones 1988, 2003). On the other hand, it is much more convenient to adopt the random variable viewpoint when considering simulation algorithms for point processes (Stoyan *et al.* 1995; Van Lieshout 2000; Møller 1999; Geyer 1999; Ripley 1977). The treatment via random variables or unordered sequences/sets of points arises more naturally when modelling interactions between the individuals.

1.1.1 Point Processes & Exponential Spaces

Let $\mathcal{X}$ be a complete separable metric space (c.s.m.s.). As noted in the previous section a point process X on $\mathcal{X}$ can be thought of as a random set or unordered sequence of points or individuals $\{\xi_n\}$, ie. X is a random variable taking values in the Exponential Space of $\mathcal{X}$, denoted by $\mathcal{X}_e$ (Carter & Prenter 1972). Informally speaking, $\mathcal{X}_e$ is the space of all *finite* (Definition 1.2) point configurations on $\mathcal{X}$. A formal interpretation and construction of $\mathcal{X}_e$ is given in Preston (1977): let $\mathcal{X}_n$ denote the n-fold product of $\mathcal{X}$. Identify those points of $\mathcal{X}_n$ which can be obtained from each other by permutation of the co-ordinates and denote the space obtained by this identification as $\mathcal{S}_n$. Setting $\mathcal{X}_e = \biguplus_n \mathcal{S}_n$ (where the $\mathcal{S}_n$ are disjoint) gives us the Exponential Space of $\mathcal{X}$. Let $\mathcal{B}(\mathcal{X})$ denote the Borel σ-algebra on $\mathcal{X}$, $\mathcal{B}(\mathcal{S}_n)$ the product σ-algebra generated by open sets of $\mathcal{S}_n$ and $\mathcal{B}(\mathcal{X}_e)$ the σ-algebra generated by the $\mathcal{B}(\mathcal{S}_n)$. A point process X on some bounded $W \subseteq \mathcal{X}$ can hence be defined as an W_e-valued random variable:

Definition 1.1. A *point process* X on some bounded $W \subseteq \mathcal{X}$ is a measurable mapping of a probability space $(\Omega, \Im, \mathbb{P})$ to $(W_e, \mathcal{B}(W_e))$.

For a point process X and $A \in \mathcal{B}(\mathcal{X}_e)$ let $X(A)$ be the number of points of X in A. Implicit in the above definition is that X is finite or locally finite.

Definition 1.2. A point process X is *finite* if $X(\mathcal{X}) < \infty$.

Definition 1.3. X is *locally finite* if $X(A) < \infty$ for all bounded $A \in \mathcal{B}(\mathcal{X})$.

Definition 1.4. X is *simple* if $X(\{\xi\}) = 0$ or 1 for all $\xi \in \mathcal{X}$.

1.1.2 Point Processes & Random Measures

In this section we follow the treatment of point processes as random counting measures, as presented in Daley & Vere-Jones (1988, 2003) and Stoyan *et al.* (1995). For a point process $X = \{\xi_n\}$ and $A \in \mathcal{B}(\mathcal{X})$ the mapping $X(A) = \#\{n;\ \xi_n \in A\}$ counts the number of points of X contained in A. If $A = \bigcup_n A_n$, with the A_n disjoint, then $X(A) = \sum_n X(A_n)$ is a countably additive, non-negative, integer-valued function with $X(A) < \infty$ for all bounded A. These are exactly the conditions that make $X(\cdot)$ *counting* measure (ie. a non-negative integer-valued measure) on $\mathcal{B}(\mathcal{X})$. A measure μ on $\mathcal{X}$ is *boundedly finite* if $\mu(A) < \infty$ for all bounded Borel sets A. Let $\mathbf{M}_{\mathcal{X}}$ be the space of all boundedly finite measures on $\mathcal{X}$ and $\mathcal{B}(\mathbf{M}_{\mathcal{X}})$ its corresponding Borel σ-algebra.

Definition 1.5. A *random measure* Y on $\mathcal{X}$ is a random element of $(\mathbf{M}_{\mathcal{X}}, \mathcal{B}(\mathbf{M}_{\mathcal{X}}))$, ie. Y is a measurable mapping of a probability space $(\Omega, \Im, \mathbb{P})$ to $(\mathbf{M}_{\mathcal{X}}, \mathcal{B}(\mathbf{M}_{\mathcal{X}}))$.

A *counting measure* is a boundedly finite and integer-valued measure. A *completely random measure* is a random measure Y such that, for every family of pairwise disjoint bounded Borel sets $A_1, \ldots, A_n$, the random variables $Y(A_1), \ldots, Y(A_n)$ are independent. *Simple counting measures* are such that $Y(\{\xi\}) = 0$ or 1 for all $\xi \in \mathcal{X}$. Let $\mathbf{N}_{\mathcal{X}}$ denote the space of counting measures on $\mathcal{X}$, $\mathcal{B}(\mathbf{N}_{\mathcal{X}})$ its corresponding σ-algebra and $\mathbf{N}^*_{\mathcal{X}}$ the space of simple counting measures on $\mathcal{X}$, ie. $\mathbf{N}^*_{\mathcal{X}}$ constitutes those members of $\mathbf{N}_{\mathcal{X}}$ which are simple.

Definition 1.6. A *point process* X on $\mathcal{X}$ is a measurable mapping of a probability space $(\Omega, \Im, \mathbb{P})$ to $(\mathbf{N}_{\mathcal{X}}, \mathcal{B}(\mathbf{N}_{\mathcal{X}}))$. Furthermore, X is simple if $\mathbb{P}[X \in \mathbf{N}^*_{\mathcal{X}}] = 1$.

"A measurable mapping X from a probability space into $\mathbf{N}_{\mathcal{X}}$ (or $\mathbf{N}^*_{\mathcal{X}}$) is a point process if and only if $X(A)$ is a random variable for each Borel set $A \subseteq \mathcal{X}$" (Daley & Vere-Jones 1988, Proposition 7.1.VIII). A point process X is a random choice of one of the elements of $\mathbf{N}_{\mathcal{X}}$; so it generates a measure $\mathbf{P}$ on $(\mathbf{N}_{\mathcal{X}}, \mathcal{B}(\mathbf{N}_{\mathcal{X}}))$: the distribution of X, defined by $\mathbf{P}(B) = \mathbb{P}[X \in B] = \mathbb{P}[\{\omega;\ X(\omega) \in B\}]$, for $B \in \mathcal{B}(\mathbf{N}_{\mathcal{X}})$. For $x \in \mathcal{X}_e$, $A \in \mathcal{B}(\mathcal{X})$ let $N(A, x) = x(A) = \sum_{\xi \in x} \mathbf{1}_{\{\xi \in A\}}$; then, for fixed A, the mapping N is finite, non-negative, integer-valued and countably additive. Moreover it defines a one-to-one mapping from $\mathcal{X}_e$ to $\mathbf{N}_{\mathcal{X}}$ (Moyal 1962, Theorem 3.1). Thus an element of $\mathcal{X}_e$ can be regarded as a finite counting measure and vice versa.

Definition 1.7. The *finite-dimensional (fidi) distributions* of a point process X are the joint distributions of the random variables $X(A_1), \ldots, X(A_k)$, for all finite collections of bounded Borel sets $A_1, \ldots, A_k$, ie. the family of proper distribution functions

$$\mathbf{P}_k(A_1, \ldots, A_k;\ n_1, \ldots, n_k) = \mathbb{P}[X(A_i) = n_i,\ i = 1, \ldots, k]. \tag{1.1}$$

Definition 1.8. The *void probability* (Stoyan *et al.* 1995) or the *avoidance function* (Daley & Vere-Jones 1988) of a point process X gives the probability that there are no points of the process in a given test set A: $v_0(A) \equiv \mathbb{P}[X(A) = 0]$.

"The distribution of a point process is completely specified by the fidi distribution of $X(A)$ for A in a countable ring generating the Borel sets" (Daley & Vere-Jones 1988, circa Theorem 7.1.XI). Moreover for simple point processes the distribution is determined by the void probabilities/avoidance function v_0 on compact sets (Stoyan *et al.* 1995; Daley & Vere-Jones 1988, Theorem 7.3.II).

1.1.3 Characteristics of Point Processes

1. *Stationarity & Isotropy:* A point process X or its distribution $\mathbf{P}$ is *stationary* if the processes $X = \{\xi_n\}$ and $X_\xi = \{\xi_n + \xi\}$ have the same distributions for all $\xi \in \mathcal{X}$, ie. the distribution is translation invariant: $\mathbf{P}(B) \equiv \mathbb{P}[X \in B] = \mathbb{P}[X_\xi \in B] \equiv \mathbf{P}(B_{-\xi})$ for all $B \in \mathcal{B}(\mathbf{N}_\mathcal{X})$, where $B_\xi = \{Y \in \mathbf{N}_\mathcal{X};\ Y_{-\xi} \in B\}$. The process is *isotropic* if it is rotation invariant: if $\mathbf{r}$ is a rotation around the origin then X and $\mathbf{r}X$ have the same distribution, ie. $\mathbf{P}(B) = \mathbf{P}(\mathbf{r}B)$ for all $B \in \mathcal{B}(\mathbf{N}_\mathcal{X})$, where $\mathbf{r}B = \{Y \in \mathbf{N}_\mathcal{X};\ \mathbf{r}^{-1}Y \in B\}$. Stationarity and isotropy imply motion invariance: the distribution of X is the same as $\mathbf{m}X$ for all Euclidean motions $\mathbf{m}$.

2. *Intensity Measure:* The intensity measure Λ of X is a characteristic analogous to the mean of a real-valued random variable. It is defined as the mean number of points in $A \in \mathcal{B}(\mathcal{X})$:

$$\Lambda(A) = \mathbb{E}[X(A)] = \int x(A)\, \mathbf{P}(dx).$$

   For $\mathcal{X} = \mathbb{R}^d$, if X is translation invariant then $\Lambda(A) = \lambda\, m_d[A]$ where m_d denotes Lebesgue measure and $\lambda > 0$ is referred to as the *intensity* of X. Choosing A to have d-volume 1 shows that λ may be interpreted as the mean number of points per unit d-volume. The *Campbell*

Theorem (Stoyan *et al.* 1995) shows that for any non-negative measurable function f:

$$\mathbb{E}\left[\sum_{\xi\in X} f(\xi)\right] = \int \sum_{\xi\in x} f(\xi)\,\mathbf{P}(dx) = \int\int f(\xi)\,x(d\xi)\,\mathbf{P}(dx) = \int f(\xi)\,\Lambda(d\xi);$$

and for stationary X : $\mathbb{E}\left[\sum_{\xi\in X} f(\xi)\right] = \lambda \int f(\xi)\,d\xi.$

1.1.4 Construction of Point Processes

A convenient and constructive way to model and simulate point processes on bounded regions is by means of a "discrete distribution for the number of points and a family of symmetric densities for the locations" (Van Lieshout 2000; Daley & Vere-Jones 1988, Condition 5.3.I):

- The points are located in a complete separable metric space (c.s.m.s.) $\mathcal{X}$, such as $\mathbb{R}^d$.
- A distribution $\{p_n\}_{n\geq 0}$ is given for determining the total number of points in the population.
- For each integer $n \geq 1$, a symmetric probability density $f_n(\xi_1, \ldots, \xi_n)$ is given that determines the joint distribution of the positions of the n points of the process.

This provides a natural recipe for simulating a point process: first simulate the number of points N according to the distribution $\{p_n\}$; given that $N = n$ generate a random vector $(\xi_1, \ldots, \xi_n)$ according to the density $f_n(\cdot, \ldots, \cdot)$. As noted in Daley & Vere-Jones (1988), point processes are *unordered* sets or sequences of points; therefore it is implicitly assumed that the individuals of the process are indistinguishable. So the joint densities f_n should be indifferent to the order in which the points are listed, ie. they have to be *symmetric* so that equal weight is given to all $n!$ permutations of the points $(\xi_1, \ldots, \xi_n)$.

1.1.5 The Poisson Process

This is mathematically the most-tractable and perhaps the most studied point process (Matheron 1975; Diggle 1983; Ripley 1981; Ripley 1988; Daley & Vere-Jones 1988; Kingman 1993; Stoyan *et al.* 1995). The process typifies what one means by a 'completely random' point pattern and the first step in analyzing a point configuration is to test it against the Poisson process for spatial randomness.

Definition 1.9. Let $\mathcal{X}$ be a c.s.m.s. and Λ a Borel measure such that $\Lambda(\mathcal{X}) > 0$ and $\Lambda(A) < \infty$ for all bounded Borel A; X is a *Poisson process* with intensity measure Λ if

P1 $X(A)$ is a Poisson random variable with mean $\Lambda(A)$ for every bounded Borel set $A \subseteq \mathcal{X}$;

P2 for all disjoint Borel $A_1, \ldots, A_k$, the random variables $X(A_1), \ldots, X(A_k)$ are independent.

If Λ is atom-less (ie. gives zero measure to singleton sets) then the Poisson process is simple (Definition 1.4). **P2** is often interpreted as *complete spatial randomness:* the process in disjoint regions behaves independently. If $\mathcal{X} = \mathbb{R}^d$ and $\Lambda(A) = \lambda\, m_d[A]$ (cf. Section 1.1.3), for m_d Lebesgue measure, then X is referred to as a *homogeneous Poisson process* with intensity $\lambda > 0$, and abbreviated as $X \sim \text{Poisson}(\lambda)$. In order to simulate a Poisson(λ) process on $W \subset \mathbb{R}^d$: draw a Poisson random variable N with mean $\lambda m_d[W]$. If $N = 0$ set $X = \emptyset$; else for each $n \in \{1, \ldots, N\}$ draw $\xi_n \sim \text{Uniform}(W)$ and set $X = \{\xi_1, \ldots, \xi_N\}$.

An *inhomogeneous* Poisson process X' is one whose intensity measure is of the form $\Lambda'(A) = \int_A \lambda'(x)\, dx$. The function $\lambda'(\cdot)$ is referred to as the *intensity function*. The following rejection technique enables one to draw X'. Draw $X \sim \text{Poisson}(\lambda)$ such that $\lambda'(\cdot) \leq \lambda$; then $X' = \left\{\xi \in X;\ \frac{\lambda'(\xi)}{\lambda} \leq \text{Uniform}(0,1)\right\}$, ie. each point of X is retained with probability $\frac{\lambda'(\cdot)}{\lambda}$ to yield an inhomogeneous Poisson process with intensity function $\lambda'(\cdot)$. One can also construct random geometrical shapes/patterns via Poisson processes by defining the *Boolean model* (Stoyan *et al.* 1995) of disks of fixed radius r.

Definition 1.10. Let $X = \{\xi_n\}$ be a Poisson(λ) process on $\mathbb{R}^d$ and $B_r(\xi)$ a d-dimensional ball of radius r and centre ξ. Then the *Boolean model* of balls associated with X is constructed as

$$\mathcal{U}(X) = \bigcup_{n \geq 1} B_r(\xi_n). \tag{1.2}$$

The points of X are called *germs* and the balls $B_r(\cdot)$ *grains*; thus the Boolean model can also be viewed as a *germ-grain random set model*.

A more general definition of Boolean models is given in Chapter 3 (Definition 3.1). However the basic idea of such models is introduced here since Chapter 2 concerns the simulation of conditional Boolean models.

1.1.6 Finite Point Processes Specified by a Density

Poisson processes are a useful tool for building more complex point process models, eg. the *Poisson cluster process*, which is obtained by independently replacing each point of a Poisson process by a finite 'daughter' process. Special cases are the *Neyman & Scott (1958) process*, where the daughters are independent and identically distributed, and the *Matérn (1986) cluster process*, where each parent gives rise to a Poisson number of daughters distributed in a ball of radius r around the parent.

Other useful point process models can be obtained by specifying a density with respect to a 'reference' Poisson process. Let π denote the distribution of a Poisson process on $\mathcal{X}$ with intensity measure Λ, and $f : \mathcal{X}_e \to [0, \infty)$ be a non-negative measurable function on the exponential space (space of locally finite point configurations) $\mathcal{X}_e$. If $\int_{\mathcal{X}_e} f(x)\, \pi(dx) = 1$ then f is a density with respect to π and defines a point process X_f on $\mathcal{X}$. Thus, given a reference process X (usually Poisson(1)) with distribution π, a new process $X_f \sim \pi_f$ can be defined on $(\mathcal{X}_e, \mathcal{B}(\mathcal{X}_e))$ by means of a density f such that $\pi_f = \int f(x)\, \pi(dx)$. Then $f(x)$ is the *likelihood* that X_f takes on the realization x compared with X taking on the same realization. The distribution of the number of points in the new process X_f specified by density f is given by

$$p_n = \frac{e^{-\Lambda(\mathcal{X})}}{n!} \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} f(\{\xi_1, \ldots, \xi_n\})\, \Lambda(d\xi_1) \ldots \Lambda(d\xi_n)$$

(Van Lieshout 2000). Conditional on the event that X_f has n points, the density (with respect to the n-fold product measure Λ^n) of the locations of the points is given by

$$f_n(\xi_1, \ldots, \xi_n) = \frac{f(\{\xi_1, \ldots, \xi_n\})}{\int_{\mathcal{X}} \cdots \int_{\mathcal{X}} f(\{\xi_1, \ldots, \xi_n\})\, \Lambda(d\xi_1) \ldots \Lambda(d\xi_n)}.$$

Example 1.1 (The density of a Poisson process). Let X be a Poisson process on $\mathbb{R}^d$ with intensity measure Λ and distribution π. Let $f_\lambda(x) = \lambda^{n(x)}\, e^{(1-\lambda)}$, where $n(x)$ denotes the number of points in x, and $\pi_\lambda(B) = \int_B f_\lambda(x)\, \pi(dx)$ for $B \in \mathcal{B}(X_e)$. Then π_λ is the distribution of a Poisson process with intensity measure $\lambda\Lambda$ and density f_λ.

Definition 1.11. A point process with density f is *locally stable* if there exists $K > 0$ such that

$$f(x \cup \{\xi\}) \leq K f(x), \qquad \text{for all } x \in \mathcal{X}_e \text{ and } \xi \notin x. \tag{1.3}$$

The density is said to be *Ruelle (1969) stable* if $f(x) \leq C K^{n(x)}$ for some positive constants C, K. A locally stable density is also Ruelle stable, but not the converse.

Definition 1.12. A density is *hereditary* if $f(x) > 0 \Rightarrow f(y) > 0$ for $y \subset x$; conversely f is *anti-hereditary* if $f(y) > 0 \Rightarrow f(x) > 0$ for $y \subset x$.

1.1.7 Marked Point Processes

Suppose that, given the locations of objects or individuals, one wishes to not only model the spatial distribution but also some characteristic of each individual, eg. the diameter or height of trees, shape, colour, weight of animals, etc. Such information can be modelled by *marked point processes*: to each point or individual of the process a 'mark' is attached which represents the feature or characteristic. Thus a marked process X on some space $\mathcal{X}$ is a random unordered sequence of tuples $X = \{(\xi_n, m_n)\}$ such that the $\{\xi_n\}$ themselves form a point process on $\mathcal{X}$ and m_n is the mark attached to the individual ξ_n. The structure of the marks may be complicated and it is assumed that they belong to a given space of marks $\mathbf{M}$. A marked point process X on $\mathcal{X}$ can be viewed as a point process on the product space $\mathcal{X} \times \mathbf{M}$; all the theory of point processes extends to marked point processes (see for example Daley & Vere-Jones 1988 or Stoyan *et al.* 1995).

Definition 1.13. A marked point process X, with positions in $\mathcal{X}$ and marks in $\mathbf{M}$, is a point process on $\mathcal{X} \times \mathbf{M}$ with the additional property that $\{X(A \times \mathbf{M});\ A \in \mathcal{B}(\mathcal{X})\}$, the marginal process of locations, is itself a point process.

1.1.8 Interior & Exterior Conditioning

The conditional distribution of a point process X can be defined under two types of conditioning:

Exterior: This type of conditioning concerns the conditional distribution of the point process at some $\xi \in \mathcal{X}$ given the configuration on $\mathcal{X} \setminus \{\xi\}$, ie. the (conditional) probability that there is a point in an infinitesimally small region $d\xi$ around ξ, given the configuration outside $d\xi$.

Interior: This concerns the conditional distribution of the process on $\mathcal{X} \setminus \{\xi\}$, given a point at ξ.

"The two concepts are dual; exterior conditioning is formalized by the *Papangelou conditional intensity*, interior conditioning by the *Palm distribution*" (Van Lieshout 2000). The Papangelou conditional intensity was introduced by Papangelou (1974) and Palm distributions by Palm (1943).

Palm Distributions

For a point process X with distribution $\mathbf{P}$ a heuristic interpretation of the Palm distribution, $\mathbf{P}_\xi$, is the conditional distribution of X given that $X(\{\xi\}) > 0$, ie. palm probabilities are the conditional probabilities of X given that there is a point at location ξ. For a more rigorous definition, one needs the concept of *Campbell measures* (Stoyan *et al.* 1995):

Definition 1.14. If X is a simple locally finite point process on a c.s.m.s. $\mathcal{X}$ with distribution $\mathbf{P}$ then its *Campbell measure* is the measure on $(\mathcal{X} \times \mathcal{X}_e, \mathcal{B}(\mathcal{X}) \times \mathcal{B}(\mathcal{X}_e))$ such that

$$\int \sum_{\xi \in x} f(\xi, x) \, \mathbf{P}(dx) = \int f(\xi, x) \, \mathbf{C}(d(\xi, x)) \tag{1.4}$$

where f is any non-negative measurable function on $\mathcal{X} \times \mathcal{X}_e$. If $f(\xi, x) = \mathbf{1}_{\{\xi \in A\}}$ and $A \in \mathcal{B}(\mathcal{X})$ then $\mathbf{C}(A \times B) = \mathbb{E}\left[X(A) \, \mathbf{1}_{\{X \in B\}}\right]$, for $B \in \mathcal{B}(\mathcal{X}_e)$.

Definition 1.15. The *Reduced Campbell* measure $\mathbf{C}^!$ is such that

$$\int \sum_{\xi \in x} f(\xi, x \setminus \{\xi\}) \, \mathbf{P}(dx) = \int f(\xi, x \setminus \{\xi\}) \, \mathbf{C}^!(d(\xi, x)) \tag{1.5}$$

Moreover, $\mathbf{C}^!(A \times B) = \mathbb{E}\left[\sum_{\xi \in X \cap A} \mathbf{1}_{\{X \setminus \{\xi\} \in B\}}\right]$.

If the intensity measure Λ of X exists then $\mathbf{C}(\cdot \times B)$ is absolutely continuous with respect to Λ, with Radon-Nikodym derivative $\mathbf{P}_\cdot(B) : \mathcal{X} \to \mathbb{R}$, so that

$$\mathbf{C}(A \times B) = \int_A \mathbf{P}_\xi(B) \, \Lambda(d\xi), \qquad \text{for all } A \in \mathcal{B}(\mathcal{X}). \tag{1.6}$$

For fixed ξ, $\mathbf{P}_\xi$ is a distribution function on $(\mathcal{X}_e, \mathcal{B}(\mathcal{X}_e))$ and is referred to as the *Palm distribution* at ξ. Similarly the *reduced Palm distribution* $\mathbf{P}^!_\xi$ is related to $\mathbf{C}^!$ via

$$\mathbf{C}^!(A \times B) = \int_A \mathbf{P}^!_\xi(B) \, \Lambda(d\xi) \tag{1.7}$$

so $\mathbf{P}^!_0(B) = \mathbb{P}[X \setminus \{0\} \in B \mid 0]$ is the conditional distribution of $X \setminus \{0\}$ given that $X(\{0\}) > 0$.

Papangelou Conditional Intensities

Turning attention to the dual case of exterior conditioning the *Papangelou conditional intensity* ℓ associated with a simple point process X can be interpreted as

$$\ell(\xi; x) \, dx = \mathbb{P}[X(d\xi) = 1 \mid X \cap (d\xi)^c = x \cap (d\xi)^c]$$

the "infinitesimal probability that there is a point in a region $d\xi$ around $\xi \in \mathcal{X}$ given that the point process X agrees with the configuration x outside $d\xi$" (Van Lieshout 2000). A more formal definition requires the concept of the reduced Campbell measure $\mathbf{C}^!$ (Definition 1.15). The reduced Palm distribution (Eq. 1.7) is the Radon-Nikodym derivative of $\mathbf{C}^!(\cdot \times B)$ with respect to the intensity measure Λ of the point process X. If $\mathbf{C}^!(A \times \cdot)$ is absolutely continuous with respect to the distribution $\mathbf{P}$ of X then

$$\mathbf{C}^!(A \times B) = \int_B L(A; x)\, \mathbf{P}(dx), \qquad A \in \mathcal{B}(\mathcal{X}) \tag{1.8}$$

for some measurable function $L(\cdot; x)$, which is referred to as the first order *Papangelou kernel*. Furthermore if $L(\cdot; x)$ admits a density $\ell(\cdot; x)$ with respect to the intensity measure Λ of X, then $\ell(\cdot; \cdot)$ is called the (first order) *Papangelou conditional intensity*. When X admits a density with respect to a Poisson process the Papangelou conditional intensity takes a simple form, and provides a useful way to simulate such processes X (cf. Section 1.3.2).

Theorem 1.1 (Van Lieshout 2000, Theorem 1.6). *For X a finite process specified by density f with respect to a Poisson process the Papangelou conditional intensity is given as*

$$\ell(\xi; x) = \frac{f(x \cup \{\xi\})}{f(x)}, \qquad \textit{for } \xi \notin x \in \mathcal{X}_e \tag{1.9}$$

with convention that $\ell(\xi; x) = 0$ if $f(x) = 0$.

Suppose that a point process X has density f with respect to a Poisson(λ) process, and recall the notions of local stability and hereditary density (Definitions 1.11 & 1.12). The density f is locally stable if there exists a constant $K > 0$ such that $f(x \cup \{\xi\}) \leq K f(x)$ for all x and all $\xi \notin x$; it is hereditary if $f(x) > 0 \Rightarrow f(y) > 0$ for $y \subset x$. It is clear that a locally stable density is hereditary. Furthermore if the Papangelou conditional intensity of f, defined as in Eq. (1.9), is uniformly bounded then local stability and the hereditary condition are equivalent (Kendall & Møller 2000). The density f may also be *attractive* or *repulsive*:

$$\textit{attractive:} \qquad \ell(\xi; x) \geq \ell(\xi; y), \qquad \text{whenever } \xi \notin y \subseteq x; \tag{1.10}$$

$$\textit{repulsive:} \qquad \ell(\xi; x) \leq \ell(\xi; y), \qquad \text{whenever } \xi \notin y \subseteq x. \tag{1.11}$$

1.1.9 Operations on Point processes

In this section three fundamental operations on point processes are presented which produce new processes from old ones. Denote by X_0 the basic or reference process to which the operation is carried out on and X the new process so obtained.

Thinning: Such a operation uses some criterion to delete points of the basic process X_0 and yield a *thinned* process X. The simplest thinning operation is an *independent p-thinning*, where each $\xi \in X_0$ is deleted with probability $1-p$, for some $p \in [0,1]$. Deletion of a point is independent of its location and of other points in X_0. An extension of this is to allow the retention probability p to depend on the location of the point; thus $\xi \in X_0$ is deleted with probability $p(\xi)$, for some deterministic (or even random) function $p(\cdot)$ defined on $\mathcal{X}$, and taking values in $[0,1]$. One could also define *dependent* thinnings where the retention probability depends on the locations of other points as well.

In the case of independent thinning, if the characteristics of the basic process X_0 are known then it is straightforward to calculate the corresponding characteristics of the thinned process X. If X_0 is a Poisson process with intensity measure Λ_0 then an independent $p(\cdot)$-thinning of X_0 is also a Poisson process, with intensity measure $\Lambda(A) = \int_A p(\xi)\,\Lambda_0(d\xi)$.

Clustering: In this case every point ξ of the basic process X_0 is replaced by a cluster Y_ξ of points. X_0 is referred to as the *parent* process and Y_ξ the *daughter* process. The union $\bigcup_\xi Y_\xi$ is the *cluster* process, with the daughters Y_ξ themselves being finite point processes.

Such processes have been used to model many natural phenomenon, eg. locations of galaxies in space, distribution of larvae in fields, the geometry of bombing (cf. Figure 1). If the daughters Y_ξ are independent identically distributed finite point sets then this is called *homogeneous independent clustering*; moreover if the parent process X_0 is Poisson then the resulting daughter process is a *Poisson cluster process*. Important cases are the Neyman & Scott (1958) model, where daughters are independent and identically distributed; and Matérn (1986) cluster process, where daughters are distributed in a ball of radius r around the parent.

Superposition: Here a collection of basic processes $\{X_0^i\}_{i=1}^k$ are combined to form a single process $X = \bigcup_{i=1}^k X_0^i$. If the basic processes are independent then the intensity measure of the super-

imposed process X is just the sum of the intensity measures of the basic processes. If $\{X_0^i\}_{i=1}^k$ are independent Poisson processes (not necessarily with the same intensity measures) then X is also a Poisson process.

1.1.10 Gibbs Point Processes

There are many different classes or models of point processes that have arisen in the literature over the years and this section considers the class referred to as Gibbs point processes. They are a development of the idea in Section 1.1.6: that new point processes can be obtained from old ones by transforming their distributions via a probability density. Following Daley & Vere-Jones (1988) and Stoyan *et al.* (1995) the construction of Gibbs processes is described below. The origins of such processes stem from statistical physics, being related to the so-called *Gibbs distributions* which describe the equilibrium states of subsystems of very large closed physical systems. Such processes are described by means of forces acting on and between the particles. The total potential energy of a given configuration of particles or individuals is assumed to be decomposable into terms representing the interactions between the particles taken in pairs, triples, and so on. First-order terms representing the presence of an external force can also be included. Thus Gibbs processes can be thought of as processes generated by *interaction potentials*.

The fundamental ingredient in specifying a Gibbs process X is an underlying *basic* or *weight* process X_0 with distribution $\mathbf{Q}$ (this is usually taken to be a Poisson process). The distribution $\mathbf{P}$ of a Gibbs process can then be defined by means of a density f:

$$\mathbf{P}(B) = \int_B f(x)\,\mathbf{Q}(dx), \qquad B \in \mathcal{B}(\mathcal{X}_e). \tag{1.12}$$

Rather than specifying a distribution $\mathbf{P}$ for the Gibbs process and then checking absolute continuity $\mathbf{P} \ll \mathbf{Q}$, the usual trick is to do the reverse. An integrable function f (with respect to $\mathbf{Q}$) is specified and the distribution $\mathbf{P}$ is then defined via Eq. (1.12). The form of f is often determined by the field of application: it can be chosen conveniently to model interactions between individuals of the process. This approach is straightforward if the process contains only finitely many points confined to a bounded region W. More generally, for point processes in all of $\mathcal{X}$, the density idea must be applied to conditional distributions confined to bounded regions.

Gibbs Processes in Bounded Regions

In this section we consider how to define Gibbs point processes in a bounded region $W \subset \mathcal{X}$. In order to do so there are two different cases to consider:

Canonical Ensemble: Here the process contains a fixed number n of particles, all contained in W. This is of great practical importance since one often conditions on the number of points observed in some W. The form of f is usually $f(x) = f(\xi_1, \ldots, \xi_n) = \frac{e^{-U(x)}}{Z}$, where Z is a normalizing constant called the *configurational partition function*, and $U : \mathcal{X}_e \to \mathbb{R} \cup \{\infty\}$ is the *energy function*. Frequently U is chosen to have a specialized form: an *interaction potential* $U(x) = \sum_{y \subset x} V(y)$; or a sum of *pair potentials* $U(x) = \sum \sum_{1 \leq i,j \leq n} \theta(\|\xi_i - \xi_j\|)$. The function θ is referred to as the *pair potential*, in homage to the origins of the subject in Physics, and $\|\cdot\|$ is a norm on $\mathcal{X}$.

Grand Canonical Ensemble: Here the total number of points N is random but all assumed to be within W. One approach is to define a sequence of "conditional densities" f_n. The Gibbs process is then obtained by first arranging N to have some distribution and then (conditional on the value of N) using the Canonical Ensemble construction. That is given $N = n$ the n points are distributed in W using the joint density f_n. Another approach is to define a density directly on $\mathcal{B}(W_e)$, where W_e is the exponential space (Section 1.1.1) of W.

Stationary Gibbs Processes

When considering Gibbs processes on unbounded spaces $\mathcal{X}$, a more sophisticated method than just specifying the density on bounded $W \subset \mathcal{X}$ is required. Additionally, the distribution of the process restricted to the observation window W must be *conditioned* on the process outside W.

Formally one considers the family of *local specifications* $\pi_W(\cdot \mid \cdot)$ for bounded Borel W. These represent the probability that the Gibbs process X on W belongs to the set $B \in \mathcal{B}(\mathcal{X}_e)$, given that the process takes on some configuration x outside W, ie. $\pi_W(B \mid x) = \mathbb{P}[X \cap W \in B \mid X \cap W^c = x]$, where $x \in \{y \in \mathcal{X}_e;\ y(B) = 0\}$. Then a point process X with distribution $\mathbf{P}$ is said to be a *Gibbs point process with respect to the specification* π_W if for all bounded Borel W the *DLR-equation*,

after Dobrushin, Lanford & Ruelle (Stoyan *et al.* 1995), holds:

$$\textbf{DLR-Equation:} \qquad \mathbf{P}(B) = \int \pi_W(B \mid x \cap W^c)\, \mathbf{P}(dx)\,. \tag{1.13}$$

Examples of Gibbs Point Processes

Any process with density of the form given in Canonical Ensemble is a Gibbs process; some examples are given below and several more in Section 1.1.11, as examples of Markov point processes.

1. *Hard Core Process:* This is a process where no two points are allowed to be within distance $R > 0$ of each other, so that the density is given by $f(x) \propto \lambda^{n(x)} \prod_{\{\xi,\eta:\, \xi \neq \eta\}} \mathbf{1}_{\{\|\xi-\eta\|>2R\}}$.

2. *Strauss Process:* This model has density with respect to a Poisson(1) given by $f(x) \propto \lambda^{n(x)} e^{\beta\, s_R(x)}$. The parameter $\lambda > 0$ represents the underlying Poisson intensity and $s_R(x)$ denotes the number of pairs $\xi, \eta \in x$ which are closer than distance $R > 0$. This is an example of a *pair-wise* interaction Gibbs process, since the density depends only on the number of R-close pairs. The case when $\beta = 0$ corresponds to a Poisson(λ) process; when $\beta \to -\infty$ the Strauss process converges to the hard core model since $e^{\beta\, s_R(x)}$ will be non-zero only if $s_R(x) = 0$. The Strauss process can be considered as a *soft core* process since the density is weighted by the number of R-close pairs; the case when the weight is non-zero only for zero R-close pairs corresponds to a hard core process.

3. Geyer (1999) remarks that in practical applications, "it is likely that no process model in the existing literature would be of scientific interest and a model specific to the application would be invented". Furthermore, he illustrates the ease with which one can 'invent' new point processes and do statistical inference; the *triplets* and *saturation* processes are described as examples of two new point processes.

1.1.11 Ripley-Kelly Markov Point Processes

These are another class of point processes, introduced by Ripley & Kelly (1977), and especially designed to model inter-point interactions. Let $\sim$ be a symmetric, reflexive relation on $\mathcal{X}$, eg. if $\mathcal{X} = \mathbb{R}^d$, then $\xi \sim \eta \Leftrightarrow \|\xi - \eta\| \leq r$ for some fixed r. Two points ξ, η are said to be *neighbours* with respect to $\sim$ if $\xi \sim \eta$.

Definition 1.16. The *neighbourhood* ∂W of $W \subseteq \mathcal{X}$ with respect to the relation '$\sim$' is defined as $\partial W = \{\xi \in \mathcal{X};\ \xi \sim \eta \text{ for some } \eta \in W\}$.

Definition 1.17. Let $\mathcal{X}$ be a c.s.m.s., Λ a finite non-atomic Borel measure and π_Λ the distribution of a Poisson process on $\mathcal{X}$ with intensity measure Λ. Let X be a point process on $\mathcal{X}$ specified by means of a density f with respect to π_Λ. Then X is a *(Ripley-Kelly) Markov point process* with respect to the symmetric, reflexive relation $\sim$ on $\mathcal{X}$ if for $x \in \mathcal{B}(\mathcal{X}_e)$ such that $f(x) > 0$

1. $f(y) > 0$ for all $y \subseteq x$;
2. for all $\xi \in \mathcal{X}$, $\frac{f(x \cup \{\xi\})}{f(x)}$ depends only on ξ and $\partial\xi \cap x = \{\eta \in x;\ \xi \sim \eta\}$.

Condition 2 above is a *local Markov property* since it concerns the behaviour of at a single point; the following *spatial Markov property* also holds.

Theorem 1.2 (Ripley & Kelly 1977). *Let X be a Markov point process with density f on a c.s.m.s. $\mathcal{X}$ and consider a Borel set $W \subseteq \mathcal{X}$. Then the conditional distribution of $X \cap W$, given $X \cap W^c$, depends only on X restricted to the neighbourhood $\partial W \cap W^c = \{\xi \in \mathcal{X} \setminus W;\ \xi \sim \eta \text{ for some } \eta \in W\}$.*

Ripley & Kelly (1977) show that the density of a Markov point process is of the form

$$f(x) = \frac{1}{Z} \exp\left(-\sum_{y \subset x} V(y)\right), \qquad \text{for } x \in \mathcal{X}_e. \tag{1.14}$$

Here V is an *interaction potential* (cf. Section 1.1.10), so that $V(x) \neq 0$ implies that $y \sim y'$ for all $y, y' \subset x$, for some relation '$\sim$'. In graph-theoretic language V is non-zero for all *cliques* of x. Further specializations of Markov point processes are Markov marked point processes, nearest-neighbour and connected-component Markov point processes (Van Lieshout 2000, Chapter 2).

Examples of Markov Point Processes

Such processes have been intensively studied as models for pair-wise (and possibly higher order) interactions (Van Lieshout 2000; Baddeley & Møller 1989; Ripley & Kelly 1977).

1. Hard/soft core processes.
2. Poisson and Poisson cluster process (Sections 1.1.5 & 1.1.9).

3. *Area/Perimeter/Quermass Interaction processes:* The area-interaction process is a weighted Poisson process where the weight factor is a function of the area covered by the Boolean model (cf. Definition 1.10) associated with the process. The area-interaction process is studied in Chapter 4. The perimeter-interaction process is a weighted Poisson process but with the weight depending on the perimeter of Boolean model. The quermass-interaction process (Kendall *et al.* 1999) is such that the weight is determined by the area, perimeter and the Euler characteristic (cf. Stoyan *et al.* 1995) of the Boolean model. The respective densities are:

$$\begin{aligned}
\text{area-interaction:} \quad & f(x) \propto \lambda^{n(x)} e^{-\beta_1 \, m_2[\mathcal{U}(x)]}. \\
\text{perimeter-interaction:} \quad & f(x) \propto \lambda^{n(x)} e^{-\beta_2 \, p[\mathcal{U}(x)]}. \\
\text{quermass-interaction:} \quad & f(x) \propto \lambda^{n(x)} \exp\left(\beta_1 \, m_2\left[\mathcal{U}(x)\right] + \beta_2 \, p\left[\mathcal{U}(x)\right] + \beta_3 \, \chi\left[\mathcal{U}(x)\right]\right).
\end{aligned}$$

   Here m_2, p, χ denote Lebesgue measure, perimeter length and Euler characteristic respectively, and $\mathcal{U}(x)$ is the Boolean model associated with x. Notice that if $\beta_2 = \beta_3 = 0$ then the quermass-interaction process is just an area-interaction process.

4. *Continuum Random Cluster model:* Suppose $\sim$ denotes a relation where $\xi \sim \eta$ if $\mathcal{U}(\xi) \cap \mathcal{U}(\eta) \neq \emptyset$. Furthermore ξ and ξ' are connected in x if there exist $\xi_0, \ldots, \xi_n$ in x such that $\xi = \xi_0 \sim \ldots \sim \xi_n = \xi'$. Such a connected component $\{\xi_0, \ldots, \xi_n\}$ is called a cluster; let the number of clusters in a point configuration x be $C(x)$. Setting $f(x) \propto \lambda^{n(x)} e^{-\beta \, C(x)}$, for $\lambda > 0$, defines the continuum random cluster model (Møller 1999) with density f.

5. *Penetrable Spheres Model:* This is a bivariate process obtained by conditioning two Poisson processes X_1, X_2 on the event that no point of one process is closer than a distance $R > 0$ to a point of the other. The process $Y = (X_1, X_2)$ is called the Widom & Rowlinson (1970) penetrable spheres model, with density $f(y) \propto \lambda_1^{n(x_1)} \lambda_2^{n(x_2)} \mathbf{1}_{\{d(x_1,x_2)>R\}}$, where $d(x_1, x_2)$ is the smallest distance between a point of x_1 and a point of x_2.

1.2 Markov Processes

Consider a measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, where $\mathcal{B}(\mathcal{X})$ is the Borel σ-algebra on $\mathcal{X}$. The aim of this thesis is to devise sampling schemes for point process models by simulating Markov processes on the

Exponential space $\mathcal{X}_e$. In Section 1.4 a description of how to construct such Markov processes whose *equilibrium* distributions are exactly those one wishes to sample from is given. In this section some basic definitions and results for Markov processes on a general state space $\mathcal{X}$ are presented. The treatment here is made via discrete-time processes; however most of the definitions and results apply to the continuous case as well. Where a distinction is made between the discrete and continuous cases the process will be indexed by $n \in \mathbb{N}$ and $t \in \mathbb{R}$ respectively. For a rigorous study of Markov processes see, for example, Meyn & Tweedie (1993), Nummelin (1984) or Norris (1997).

Definition 1.18. A *Markov or transition kernel* is a mapping $\mathbf{P} : \mathcal{X} \times \mathcal{B}(\mathcal{X}) \to [0,1]$ such that: (i) $\mathbf{P}(\xi, \cdot)$ is a probability measure on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ for each ξ; (ii) $\mathbf{P}(\cdot, A)$ is a non-negative measurable mapping from $\mathcal{X}$ to $[0,1]$ for all A; and (iii) $\mathbf{P}^{n+m} = \mathbf{P}^n \cdot \mathbf{P}^m$ for all $n, m \in \mathbb{N}$.

Definition 1.19. Let $\Phi = \{\Phi(n)\}$ be a sequence of $\mathcal{X}$-valued random variables defined on some probability space $(\Omega, \Im, \mathbb{P})$ and $\{\mathbb{P}_\mu\}$ a family of probability measures, for μ a probability distribution on $\mathcal{B}(\mathcal{X})$. The sequence Φ is a *Markov process* if there exists a Markov kernel $\mathbf{P}$ s.t.

$$\mathbb{P}_\mu[\Phi(n) \in \cdot] = \mu\mathbf{P}^n = \int \mu(d\xi)\, \mathbf{P}^n(\xi, \cdot);$$

$$\textbf{Markov Property:} \qquad \mathbb{P}_\mu[\Phi(n+m) \in \cdot \mid \mathcal{F}_n] = \mathbf{P}^m(\Phi(n), \cdot), \qquad \text{for all } n, m \in \mathbb{N}. \quad (1.15)$$

$\mathbb{P}_\mu$ is the underlying measure for Φ with initial distribution μ and $\mathcal{F}_n = \sigma(\Phi(i);\ 0 \le i \le n)$

Thus $\mathbf{P}^n(\xi, A) = \mathbb{P}[\Phi(n) \in A \mid \Phi(0) = \xi]$ is the probability that the process is in A at time n given that it was in state ξ at time 0. The Markov property in Eq. (1.15) means that the process is *memoryless*: the next transition depends only on the current state and not on the past history, ie. given the 'present' the 'future' is (conditionally) independent of the 'past'. If $\mathbb{P}[\Phi(n) \in \cdot \mid \Phi(n-1)] \equiv \mathbb{P}[\Phi(1) \in \cdot \mid \Phi(0)]$ is independent of the value of n then the Markov process is said to be *time-stationary* or *time-homogeneous*.

Definition 1.20. A Markov process Φ is *irreducible* if there exists a measure φ on $\mathcal{B}(\mathcal{X})$ such that, whenever $\varphi(A) > 0$, $\mathbb{P}[\Phi \text{ ever visits } A] > 0$ (ie. $\mathbf{P}^n(\xi, A) > 0$ for some $n > 0$) for all $\xi \in \mathcal{X}$.

Definition 1.21. A Markov process is *recurrent* if there exists a measure φ on $\mathcal{B}(\mathcal{X})$ such that, for all $A \in \mathcal{B}(\mathcal{X})$ with $\varphi(A) > 0$, $\mathbb{P}[\Phi(n) \in A \text{ infinitely often}] = 1$. This is equivalent to $\mathbb{P}[\tau_A < \infty] = 1$, where $\tau_A = \inf\{n > 0;\ \Phi(n) \in A\}$ is the first return time to A. Also, if $\eta_A = \sum_n \mathbf{1}_{\{\Phi(n) \in A\}}$ is

the occupation time for A and $U(\xi, A) = \sum_n P^n(\xi, A) = \mathbb{E}_\xi[\eta_A]$ the mean occupation time, then Φ is recurrent if $U(\xi, A) \equiv \infty$ for all ξ, A.

Definition 1.22. A σ-finite measure π on $\mathcal{B}(\mathcal{X})$ with

$$\pi(A) = \int_{\mathcal{X}} \mathbf{P}(\xi, A)\, \pi(d\xi), \qquad \text{for all } A \in \mathcal{B}(\mathcal{X}) \tag{1.16}$$

is called an *invariant measure*. If π is totally finite then it may be normalized to a *stationary* probability measure, in which case it is referred to as the *equilibrium distribution* of $\mathbf{P}$.

An irreducible Markov process which admits an equilibrium distribution is recurrent and vice versa (Meyn & Tweedie 1993). Equilibrium distributions are important because they not only define the stationary Markov process Φ but also define its long run or *ergodic* behaviour.

Definition 1.23. A Markov process Φ with equilibrium distribution π is *ergodic* if

$$\lim_{n\to\infty} \|\mathbf{P}^n(\xi, \cdot) - \pi(\cdot)\| = 0 \tag{1.17}$$

where $\|\cdot\|$ is the total variation norm (see Eq. 1.22 below). Furthermore Φ is *geometrically* ergodic if there exists a non-negative function M such that $\mathbb{E}_\pi[M(\Phi)] < \infty$ and a positive constant $0 \leq \rho < 1$ such that $\|\mathbf{P}^n(\xi, \cdot) - \pi\| \leq M(\xi)\rho^n$ for all $\xi \in \mathcal{X}$. The chain is *uniformly ergodic* if there exists a finite $m \geq M(\xi)$ for all $\xi \in \mathcal{X}$.

Definition 1.24. Suppose $\Phi = \{\Phi(n)\}$ is a Markov process with kernel $\mathbf{P}$. For arbitrary but fixed $N > 0$ define the *time-reversed* process $\left\{\tilde{\Phi}(n);\ 0 \leq n \leq N\right\}$ by $\tilde{\Phi}(n) = \Phi(N - n)$. Furthermore if π is a stationary distribution for Φ, then Φ is said to be *time-reversible* (with respect to π) if

$$\int_B \mathbf{P}(\xi, A)\, \pi(d\xi) = \int_A \mathbf{P}(\xi, B)\, \pi(d\xi), \qquad \text{for all } A, B \in \mathcal{B}(\mathcal{X}). \tag{1.18}$$

This means that the probability of the process being in A at time n and in B at time m is equal to the probability of being in B at time n and in A at time m, ie. Φ and $\tilde{\Phi}$ 'look statistically the same'.

Definition 1.25. A Markov process Φ satisfies the *equations of detailed balance* with respect to a probability distribution π on $\mathcal{X}$ if

$$\pi(d\xi)\, \mathbf{P}(\xi, d\eta) = \pi(d\eta)\, \mathbf{P}(\eta, d\xi). \tag{1.19}$$

If detailed balance equations are satisfied for some π then it is the *unique* equilibrium distribution for the process. Moreover if Φ satisfies detailed balance then it is time-reversible; hence Φ is ergodic if and only if it satisfies the detailed balance equations and is irreducible. So in order to ensure that a given Markov process has a unique equilibrium distribution it suffices to check detailed balance equations; if, additionally, it is also irreducible then this guarantees that the process converges to its equilibrium distribution. This is the essence MCMC (cf. Section 1.4): an ergodic Markov process is constructed such that its equilibrium distribution is exactly that which one wishes to sample. Simulating the process long enough will ensure that its distribution converges to the required one.

1.2.1 The Coupling Method

The coupling technique is originally due to Doeblin (1938), and has a wide range of applications such as estimation of total variation distances, establishing inequalities, study of Markov and renewal process asymptotics, Poisson approximations, as well a tool for *perfect simulation* (see Section 1.4.1 on *Coupling From The Past*). Furthermore if some sort of "comparison between probability measures is required then it is often rewarding to construct random variables on a common probability space, with these measures as distributions, so that the comparison may be carried out in terms of the random variables" (Lindvall 1992). Loosely speaking, such a construction is referred to as a *coupling*; a formal definition is given below. Lindvall presents important results for a wide range of discrete- and continuous-time processes such as Markov, renewal, birth-death, Poisson and diffusions; inequalities, domination and monotonicity ideas are also dealt with.

Definition 1.26 (Lindvall 1992). A *coupling* of probability measures $\mathbf{P}, \mathbf{P}'$ on some measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is a probability measure $\widehat{\mathbf{P}}$ on $(\mathcal{X} \times \mathcal{X}, \mathcal{B}(\mathcal{X}) \times \mathcal{B}(\mathcal{X}))$ such that

$$\mathbf{P} = \widehat{\mathbf{P}}\pi^{-1} \qquad \text{and} \qquad \mathbf{P}' = \widehat{\mathbf{P}}\pi'^{-1} \tag{1.20}$$

where $\pi(\xi, \eta) = \xi, \pi'(\xi, \eta) = \eta$ for $(\xi, \eta) \in \mathcal{X} \times \mathcal{X}$.

Definition 1.27. A *coupling* of random variables X, X', defined on underlying probability spaces $(\Omega, \Im, \mathbf{P})$ and $(\Omega', \Im', \mathbf{P}')$ respectively, is a random variable $\left(\widehat{X}, \widehat{X}'\right)$ on a third probability space $\left(\widehat{\Omega}, \widehat{\Im}, \widehat{\mathbf{P}}\right)$, such that the laws

$$\mathcal{L}(X) = \mathcal{L}\left(\widehat{X}\right) \qquad \text{and} \qquad \mathcal{L}(X') = \mathcal{L}\left(\widehat{X}'\right). \tag{1.21}$$

Hence $\widehat{\mathbf{P}}\left(\widehat{X},\widehat{X}'\right)^{-1}$ is a coupling of $\mathbf{P}X^{-1}$ and $\mathbf{P}'X'^{-1}$ in the sense of Definition 1.26.

The couplings considered here will be of a more restrictive nature: two stochastic processes are coupled if their paths coincide (or *coalesce*) after a random time T, which is referred to as the *coupling time*. If $\left(\widehat{\Phi},\widehat{\Phi}'\right)$ is a coupling of two processes $\Phi = \{\Phi(t)\}, \Phi' = \{\Phi'(t)\}$ and $T = \inf_{t\geq 0}\{\Phi(t) = \Phi'(t)\}$ is almost surely finite then the coupling is called *successful*. If both processes have the same equilibrium distribution π and, furthermore if Φ' is started in equilibrium, then we get the following coupling inequality:

$$\begin{aligned}\|\mathbf{P}_t(\xi,\cdot)-\pi\| &= 2\sup_{A\in\mathcal{B}(\mathcal{X})} \mid \mathbb{P}[\Phi(t)\in A]-\mathbb{P}[\Phi'(t)\in A]\mid \\ &= 2\sup_{A\in\mathcal{B}(\mathcal{X})} \mid \mathbb{P}[\Phi(t)\in A,\Phi(t)\neq\Phi'(t)]-\mathbb{P}[\Phi'(t)\in A,\Phi(t)\neq\Phi'(t)]\mid \\ \textbf{coupling inequality:}\quad &\leq 2\,\mathbb{P}[\Phi(t)\neq\Phi'(t)] = 2\,\mathbb{P}[T>t]. \qquad (1.22)\end{aligned}$$

If a coupling of Φ and Φ' is successful then $\mathbb{P}[T>t]\to 0$ as $t\to\infty$; therefore the distribution of $\Phi(t)$ converges to its equilibrium π. The above inequality is useful in showing convergence, ergodicity of Markov processes and computing convergence rates (cf. Lindvall 1992).

1.2.2 Monotone Transition Kernels & Coupling

Definition 1.28. A Markov/transition kernel $\mathbf{P}$ on $\mathcal{X}$, endowed with a partial order $\preccurlyeq$, is *monotone* if, for all $\xi_1\preccurlyeq\xi_2$, $\mathbf{P}(\xi_1,U(\xi_1))\leq\mathbf{P}(\xi_2,U(\xi_2))$ where $U(\xi)=\{\eta:\ \xi\preccurlyeq\eta\}$ is called an 'increasing' set. Conversely $\mathbf{P}$ is *anti-monotone* if $\mathbf{P}(\xi_1,U(\xi_1))\geq\mathbf{P}(\xi_2,U(\xi_2))$.

In terms of *stochastic domination* (Kamae *et al.* 1977), $\mathbf{P}$ monotone means that the probability measures $P_i=\mathbf{P}(\xi_i,\cdot)$ are such that P_2 stochastically dominates P_1 (cf. Remark 1.2 and Section 2.2.2). The result below connects the ideas of monotonicity (or stochastic domination) and coupling. It uses the notion of an *upward kernel*: a kernel K is upward if $K(\xi,U(\xi))=1$ for all ξ.

Theorem 1.3 (Kamae *et al.* 1977, Theorem 1). *For Markov processes Φ,Φ' with the same transition kernel $\mathbf{P}$, '$\preccurlyeq$' a closed partial order and $\Phi(0)=x\preccurlyeq x'=\Phi'(0)$ the following are equivalent.*

(i) $\mathbf{P}$ is monotone;

(ii) the probability measure $\mathbf{P}(x,\cdot)$ is stochastically smaller than $\mathbf{P}(x',\cdot)$;

(iii) there exists a coupling $\left(\widehat{\Phi}, \widehat{\Phi}'\right)$ *such that* $\widehat{\Phi}(t) \preccurlyeq \widehat{\Phi}'(t)$ *for all* t*;*

(iv) there exists an upward kernel K *such that* $\mathbf{P}(x', A) = \int K(y, A)\, \mathbf{P}(x, dy)$.

Statement (iii) means that one can simultaneously produce sample paths of Φ, Φ' such that the path of Φ is always below (with respect to $\preccurlyeq$) that of Φ'. Statement (iv) means that given a sample path of Φ one can construct a sample path of Φ' such that it is always above that of Φ.

1.2.3 Simulation of Markov Processes

The main task in this thesis is to devise simulation algorithms for point process models via the construction of Markov processes whose equilibrium distribution is exactly that of the point process of interest. A natural state space for such processes will be the exponential space $\mathcal{X}_e$, for individuals lying in some space $\mathcal{X}$. The type of Markov processes dealt with here will either be discrete-time 'component' processes or spatial birth-death processes (Preston 1977). A component process with d-components $\Phi = (\Phi_1, \ldots, \Phi_d)$ is just a d-dimensional process. A spatial birth-death process is a special kind of continuous-time Markov *jump* process.

Definition 1.29. A stochastic process on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is called a *jump* process with *intensity* α and transition kernel K if, given that the process is currently in state $\xi \in \mathcal{X}$, then the waiting time till the next jump has an exponential distribution with rate $\alpha(\xi)$, independent of the past, and the probability that the jump leads to a state in $A \in \mathcal{B}(\mathcal{X})$ is $K(\xi, A)$.

The intensity and transition kernel of the birth-death process will be defined via *birth* and *death* rates. Furthermore if the birth and death rates of the spatial birth-death process satisfy detailed balance for some integrable function f (with respect to a measure μ on $(\mathcal{X}_e, \mathcal{B}(\mathcal{X}_e))$) then the equilibrium distribution of the process is given by $\pi = \int f d\mu$ (see Section 1.3.2 for details). This means that if the distribution of a point process admits a density then it is possible to define birth and death rates such that detailed balance is always satisfied; see Example 1.3.

The dynamics of component process will be defined via one step transition probabilities or kernels for each component. These will be chosen in order to ensure that the equilibrium distribution of the process is exactly that of interest. One way to guarantee this is by updating each component

according to the (equilibrium) conditional distribution of that component given the other components. This is the idea embodying *Gibbs sampling* described in Section 1.4. For discrete-time or continuous-time jump processes one can define the notion of a *transition function*, which essentially determines the next state of the process.

Definition 1.30. A *transition function* for a Markov kernel $\mathbf{P}$ is a measurable function $f : \mathcal{X} \times \cdot \to \mathcal{X}$, where $\cdot$ is the state space of some (auxiliary) random variable U, such that the law $\mathcal{L}(f(\xi, U)) = \mathbf{P}(\xi, \cdot)$ for all $\xi \in \mathcal{X}$. If $\mathcal{X}$ admits a partial order $\preccurlyeq$ then the transition function f is

$$\textit{monotone:} \quad \text{if} \quad f(\eta, U) \preccurlyeq f(\xi, U), \quad \text{whenever } \eta \preccurlyeq \xi; \tag{1.23}$$

$$\textit{anti-monotone:} \quad \text{if} \quad f(\eta, U) \succcurlyeq f(\xi, U), \quad \text{whenever } \eta \preccurlyeq \xi. \tag{1.24}$$

Remark 1.1. If $\mathbf{P}$ has a monotone transition function then $\mathbf{P}$ is monotone; this is referred to as *realizable* monotonicity by Fill & Machida (2001), who also show that the converse is not true.

Thus if $\mathbf{P}$ has equilibrium distribution π then a transition rule is simply a measurable mapping which preserves the distribution, ie. if $\xi \sim \pi$ then $f(\xi, \cdot) \sim \pi$. So any mapping which preserves the equilibrium distribution is a transition function. Therefore simulation of a discrete-time Markov process Φ reduces to sampling its transition function: initialize $\Phi(0)$ at some arbitrary state; for $n \geq 1$ set $\Phi(n) = f(\Phi(n-1), \cdot)$. A similar set up works for jump processes, but with updates occurring at 'jump' times.

Suppose Ψ is the *target* process of interest which we wish to sample, but for which direct sampling is generally difficult. Suppose also that its transition function f' can be obtained as an adapted functional of f, the transition function of some easy-to-simulate process Φ. In this case Ψ can be simulated by *coupling its evolution* to that of Φ, ie. one can produce a coupling $\left(\widehat{\Phi}, \widehat{\Psi}\right)$. The transition function f is used to update $\widehat{\Phi}$ and, conditional on such an update, the $\widehat{\Psi}$ component is updated according to f'. The process $\widehat{\Phi}$ is usually referred to as the *dominating* or *basic* or *free* process since it is the underlying process upon which the transitions of $\widehat{\Psi}$ depend.

Example 1.2. A simple example of such a coupling is when Ψ is a jump process on some $W \subseteq \mathcal{X}$ with intensity $\alpha'(x) = \int b'(\xi, x)\, \mu(d\xi) + D(x)$, where μ is some measure on $\mathcal{X}$. If $\int b'(\xi, x)\, \mu(d\xi)$ is difficult to compute but $b'(\xi, x) \leq b$, for some constant b, then $\alpha(x) = b\mu(W) + D(x) \geq \alpha'(x)$. Defining Φ as a jump process with intensity $\alpha(\cdot)$ then enables one to simulate Ψ by devising a

coupling $\left(\widehat{\Phi}, \widehat{\Psi}\right)$ as follows. The processes are initialized at the same state; suppose that the current state of $\widehat{\Phi}$ is x. The next jump time of $\widehat{\Phi}$ is simulated as an exponentially distributed random variable of rate $\alpha(x)$ where $b\mu(W)$, respectively $D(x)$, represent the total birth, respectively death, rates. With probability $\frac{b\mu(W)}{\alpha(x)}$ a birth ξ is proposed, drawn uniformly on W since the birth rate is constant; else a death $\eta \in x$ is proposed. Hence $\widehat{\Phi}$ has transition function

$$f(x, U) = \begin{cases} x \cup \{\xi\} & \xi \in W,\ \xi \notin x, \qquad \text{if } U \leq \frac{b\mu(W)}{\alpha(x)}; \\ x \setminus \{\eta\} & \eta \in x, \qquad \text{else.} \end{cases}$$

So, conditional on a transition in $\widehat{\Phi}$, the same transition is considered as a *proposed* transition in $\widehat{\Psi}$. Death transitions are always accepted in Ψ, whereas births are accepted with probability $\frac{b'(\xi,x)}{b}$. The transition function f' of $\widehat{\Psi}$ then looks like f above but with $b\mu(\mathcal{X})$ replaced by $\int b'(\xi, x)\,\mu(d\xi)$; hence $\widehat{\Psi}$ has the same transition rates as Ψ, as required.

1.3 Spatial Birth-Death Processes

The usual birth-death process is a continuous-time Markov jump process on the non-negative integers. If the current state is n then the next transitions can only be made to state $n+1$ (a *birth*) or state $n-1$ (a *death*); furthermore the transition rates depend only on the number of individuals alive. For many models it is reasonable to expect that the transition rates depend on the number of individuals *as well as* their locations. Preston (1977) describes a process which takes into account the *positions* of the individuals; hence the name 'spatial birth-death process'. The relevance of these processes lies in the close relationship to Gibbs processes and especially in the way they provide a means to simulate such processes, as suggested by Ripley (1977). When the process is *time-reversible* (Definition 1.24) then one can find *equilibrium distributions* (Definition 1.22) for the process; these actually turn out to be various kinds of Gibbs states or distributions (Preston 1977).

Suppose the individuals $\{\xi\}$ live in some space $\mathcal{X}$ and recall that $\mathcal{X}_e$ denotes the Exponential space (cf. Section 1.1.1) of $\mathcal{X}$. A spatial birth-death process Φ is a continuous-time Markov jump (cf. Definition 1.29) process taking values in $\mathcal{X}_e$. The process has a Markov property: the probability of the next transition depends only on the current configuration. Following Preston (1977) we now describe the construction of a spatial birth-death process Φ, which evolves in time with individuals dying or being born, given by the following recipe:

1. At any time t there are only a finite number of individuals alive.

2. If at time t the configuration of alive individuals is $x = \{\xi_1, \ldots, \xi_n\}$ then there exists a finite measure $B(x, \cdot)$ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ such that the probability of an individual being born in some $W \in \mathcal{B}(\mathcal{X})$ in the interval $[t, t+\delta]$ is $B(x, W)\,\delta + o(\delta)$.

3. If at time t the configuration of alive individuals is $x = \{\xi_1, \ldots, \xi_n, \xi\}$ then there exists a $\mathcal{B}(\mathcal{X})$ measurable function $d(x, \cdot) : \mathcal{X} \to \mathbb{R}_+$ such that the probability of the individual ξ dying in the interval $[t, t+\delta]$ is $d(x \setminus \{\xi\}, \xi)\,\delta + o(\delta)$.

4. The probability that there is more than one transition in $[t, t+\delta]$ is $o(\delta)$.

If $B(x; \cdot)$ is absolutely continuous with respect to some measure μ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ then there exists a positive measurable function $b : \mathcal{X}_e \times \mathcal{X} \to \mathbb{R}_+$ such that $B(x, W) = \int_W b(x, \xi)\,\mu(d\xi)$, $W \in \mathcal{B}(\mathcal{X})$. Refer to b as the *birth rate* and d the (per capita) *death rate*; then $B(x, \cdot)$ denotes the *total birth rate* and $D(x) = \sum_{\xi \in x} d(x \setminus \{\xi\}, \xi)$ the *total death rate*. In the terminology of Fernández *et al.* (2002) a spatial birth-death process whose birth and death rates, $b(x, \cdot)$ and $d(x, \cdot)$ respectively, do not depend on the current configuration x is called a *free process* since then there is no interactions between the individuals. Conversely if the birth/death rates do depend on the current configuration, then such a process is referred to as an *interacting* spatial birth-death process since there is some kind of interaction.

1.3.1 Simulation of Spatial Birth-Death Processes

Suppose Φ is a spatial birth-death process on some bounded $W \in \mathcal{B}(\mathcal{X})$ with birth rate $b(\cdot)$ and death rate $d(\cdot)$, which may depend on the current state of the process. Therefore in order to simulate Φ the state of the birth death process *outside* W must be fixed, and a realization of Φ simulated *conditional* on this fixed state outside W (cf. Section 1.1.10). If $\Phi(t) = x$ then the time till the next 'jump' or transition is exponentially distributed with rate $\alpha(x) = B(x, W) + D(x)$, where $B(x, W) = \int_W b(x, \xi)\,\mu(d\xi)$ and $D(x) = \sum_{\xi \in x} d(x \setminus \{\xi\}, \xi)$. With probability $\frac{B(x,W)}{\alpha(x)}$ the next transition is a birth, else it is a death. Births are drawn from density $\frac{b(x,\cdot)}{B(x,W)}$; deaths are drawn from density $\frac{d(x\setminus\cdot,\cdot)}{D(x)}$. The following describes how to obtain a sample path of Φ on $[0, T]$.

Algorithm 1.1 (Simulation of Φ).

Initialize $\Phi(0)$ (this can be at any arbitrary configuration); set $t = 0$.

while $t \leq T$:

draw $\tau \sim$ Exponential with rate $\alpha\left(\Phi\left(t\right)\right) = B\left(\Phi\left(t\right), W\right) + D\left(\Phi\left(t\right)\right)$.

draw $U \sim$ Uniform$(0, 1)$.

if $U \leq \frac{B(\Phi(t),W)}{\alpha(\Phi(t))}$: draw $\xi \sim \frac{b(\cdot)}{B(\Phi(t),W)}$; set $\Phi\left(t\right) = \Phi\left(t-\right) \cup \{\xi\}$.

else: draw $\xi \sim \frac{d(\cdot)}{D(\Phi(t))}$; set $\Phi\left(t\right) = \Phi\left(t-\right) \setminus \{\xi\}$.

set $t = t + \tau$.

return $\Phi\left(T\right)$.

1.3.2 Simulation of Point Processes via Spatial Birth-Death Processes

In this section we look at how one can employ spatial birth-death processes in order to sample point processes defined by a density, such as Gibbs processes. If a spatial birth-death process is statistically indistinguishable from its time-reversal then it is *time-reversible* (Definition 1.24). A time-reversible process has an equilibrium distribution (Preston 1977, Proposition 8.3), ie. if the process is started with this distribution then its distribution is preserved for all time. The equilibrium distributions of spatial birth-death processes are of particular importance, since they represent the so-called Gibbs states or distributions from statistical physics. Under suitable conditions the spatial birth-death process will converge to a unique equilibrium distribution (Ripley 1977; Preston 1977, Theorem7.1). If f is some μ-integrable function such that

$$b\left(x, \xi\right) f\left(x\right) = d\left(x \setminus \{\xi\}; \xi\right) f\left(x \cup \{\xi\}\right), \qquad \text{for all } \xi \notin x \tag{1.25}$$

then b and d satisfy the equations of *detailed balance* (Definition 1.25) with respect to f. In this case the process with birth and death rates b and d respectively is time-reversible. Moreover $\pi = \int f d\mu$ is the unique equilibrium distribution of the spatial birth-death process (Preston 1977, Lemma 8.2).

Example 1.3. For constant birth rate b and unit death rate per point, detailed balance calculations show that the process converges to a Poisson(b) process on some bounded $W \subset \mathbb{R}^d$. The density of a Poisson(b) process (with respect to the distribution of a Poisson(1) process) is given as $f\left(x\right) =$

$e^{(1-bm_d[W])}b^{n(x)}$ (cf. Example 1.1). Then for any $\xi \notin x$

$$bf(x) = be^{(1-bm_d[W])}b^{n(x)} = e^{(1-bm_d[W])}b^{n(x\cup\{\xi\})} = d(x;\xi)\, f(x \cup \{\xi\}).$$

Now consider a point process X with density f. Suppose that one can define a birth-death process Ψ so that its birth and death rates satisfy the detailed balance condition in Eq. (1.25) with respect to f. Then, provided that Ψ is irreducible, Ψ will be ergodic and hence converge to the point process X. It is easily shown that if the birth rate of Ψ is equal to the Papangelou conditional intensity of X (Eq. 1.9) and the death rate per point is one then detailed balance is satisfied.

For most point processes of interest the Papangelou conditional intensity $\ell(\cdot; x)$ depends on the current configuration x. In this case simulating Ψ with birth rate $b = \ell$ may pose practical problems since calculating the total birth rate $B(x) = \int_W \ell(\xi; x)\, \mu(d\xi)$ may be difficult, if not impossible. However if ℓ is bounded in x so that $\ell(\cdot; x) \leq \ell^*(\cdot)$ for all x then it is possible to simulate Ψ *without* having to compute $B(\cdot)$. The idea is to define another process Φ with birth rate $b^* = \ell^*$ and unit death rate per point, so that simulation of Φ is straight forward. A realization of Ψ is obtained by *coupling* its evolution to that of Φ: transitions of Φ are considered as 'proposed' transitions in Ψ and accepted so as to ensure the correct transition rates for Ψ. Specifically, since the per capita death rate for both processes is one, deaths in Φ are always accepted. The birth rate of Φ is higher, so births need to be censored and accepted with probability $\frac{\ell}{\ell^*}$ (cf. Example 1.2).

Remark 1.2. Suppose X, Y are two point processes with respective distributions π_X, π_Y and Papangelou conditional intensities ℓ_X, ℓ_Y. If $\ell_X \geq \ell_Y$ then it is possible to construct coupled spatial birth-death processes Φ_X, Φ_Y with equilibria π_X, π_Y respectively such that $\Phi_X(t) \supseteq \Phi_Y(t)$ for all t. Existence of such Φ_X and Φ_Y shows that X is stochastically larger than Y; cf. Theorem 1.3.

1.4 Markov Chain Monte Carlo (MCMC)

Markov Chain Monte Carlo (MCMC) refers to a class of probabilistic simulation techniques for sampling complex distributions where direct sampling is not feasible or inefficient. For many distributions it is possible to devise an ergodic (ie. aperiodic, irreducible, positive recurrent; see Definition 1.23) Markov chain/process such that its equilibrium distribution is exactly that which one wishes

to sample from. The essence of an MCMC algorithm is to construct an ergodic Markov chain with the target equilibrium distribution.

Metropolis *et al.* (1953) were the first to present a simulation method based on Markov chains; Hastings (1970) then generalized their methods with a view towards statistical problems. It was not until the 1980's that MCMC finally took off in the statistics community, thanks to Geman & Geman (1984), who introduced the *Gibbs Sampler* to researchers in image analysis. Bayesian statistics finally helped establish MCMC firmly in the statistics community (Gelfand & Smith 1990; Besag & Green 1993); MCMC is nowadays recognized as a powerful and widely used tool.

The MCMC preprint service `http://www.statslab.cam.ac.uk/~mcmc/` provides a list of all registered papers on MCMC methodology currently submitted for publication. Gilks *et al.* (1996) provides a good introduction to MCMC methodology; in addition Geyer (1999) and Møller (1999) describe MCMC methods and statistical inference for spatial point processes and Besag & Green (1993) traces the early developments of MCMC in Bayesian statistics and reviews computational progress in statistical physics. There are various updating schemes or 'samplers' one can employ in order to define transition probabilities/kernels of a Markov chain so that its equilibrium distribution is the desired target distribution. Some of the more common ones are discussed below.

Gibbs Sampling: If one wishes to sample a random vector $(X_1, \ldots, X_d) \sim \pi$ then it is possible to use a Gibbs sampler. A d-dimensional Markov chain $\Phi(n) = (\Phi_1(n), \ldots, \Phi_d(n))$ is initialized arbitrarily. The j-th component Φ_j is updated by drawing from the conditional distribution of X_j, given the rest of the $d-1$ components Φ_{-j}. The updating order can be fixed (*sequential scan*) or random (*random scan*).

Metropolis-Hastings Samplers: This is useful when it is difficult to sample the full conditional distributions. The idea is to first choose some mass function or density q which specifies the transition probabilities of a Markov chain; q should be easy to sample from. If the current state of the chain is x then a new state y is sampled from $q(x, \cdot)$; y is referred to as a 'proposed' state and $q(x, \cdot)$ is referred to as the 'proposal distribution'. The new state y is accepted as the next state of the chain with probability

$$\alpha(x, y) = \min\left\{1, \frac{\pi(y)\, q(y, x)}{\pi(x)\, q(x, y)}\right\}.$$

If y is accepted then the chain moves on to this new state; otherwise it stays at x. If the proposal distribution is actually the full conditional distribution, then the acceptance probability always equals 1 and we get a Gibbs sampler. Another choice for $q(x, \cdot)$ is one which is independent of the current state x so that $q(x, y) = f(y)$ for some probability mass function f; in this case one gets an *Independent Sampler*. Metropolis *et al.* (1953) originally proposed a symmetric q so that $q(x, y) = q(y, x)$, which yields the *Metropolis Sampler*. Yet another type is the *Random Walk Sampler* where $q(x, y) = f(y - x)$ for some probability mass function f. If f is symmetric around zero then we just get a Metropolis sampler. The name Random walk sampler is derived from the fact that the proposals are made according to a random walk, ie. $y = x + z$ where z is drawn from f.

Spatial Birth-Death Processes: Details of this are discussed in Section 1.3.2. This is essentially a kind of Gibbs sampling because the next state of the birth-death process is determined by the Papangelou conditional intensity, which concerns the conditional distribution of the process at a point ξ, given the rest of the configuration (cf. Section 1.1.8).

Thus given a target π, implementing an MCMC algorithm to sample (approximately) from π is relatively straightforward: choose an initial state for the chain and update its states according to the chosen sampling scheme. The difficulty is determining how long the chain should be run so that its distribution can be deemed 'close enough' to π. Three main approaches to this problem are:

1. *Convergence Rate Computations:* This compares the distribution of the chain at time n with the equilibrium π; one can then determine a value M such that the chain is sufficiently close to equilibrium for all time after M. The advantage of this is that it is an exact method; however the computations are often difficult and may yield very large values of M.

2. *Convergence Diagnostics:* Examines output of the chain to check if it has 'settled down'.

3. *Perfect Simulation:* These are a class of MCMC algorithms which can automatically detect when the chain has converged. The obvious advantage is that this is an exact method and the sample has the required distribution. However, it is usually more computational, complicated to implement and not applicable to all MCMC problems.

The purpose of this thesis is to consider existing and develop new perfect simulation algorithms for various point process models. In the following section an ingenious method called *Coupling From The Past (CFTP)*, introduced by Propp & Wilson (1996), is presented. This is a perfect simulation algorithm for sampling Markov chains on finite state spaces. An extension of this to unbounded spaces, *Dominated Coupling From The Past (domCFTP)*, was introduced by Kendall (1998) and generalized in Kendall & Møller (2000).

Another perfect simulation algorithm, Fill's *interruptible algorithm*, was proposed by Fill (1998) for finite state spaces. It has the advantage of being interruptible, ie. long runs can be aborted and a new run can be started without introducing bias in the samples. This is not the case with CFTP, where the state sampled is dependent on the run time of the algorithm. Fill's algorithm is based on accept/reject sampling and requires time-reversibility and monotonicity (Definition 1.28) of the transition kernel for the target process. The idea of the algorithm is to simulate the target process Φ with kernel $\mathbf{P}$ forwards on $\{0, \ldots, n\}$. Monotonicity of $\mathbf{P}$ means that its time-reversal $\tilde{\mathbf{P}}$ is also monotonic, and this is exploited so as to determine whether $\Phi(n)$ can be accepted as an exact sample from its equilibrium distribution. Thönnes (1999) extends Fill's algorithm to point processes. The FMMR algorithm (Fill *et al.* 2000) extends it to general state spaces where no assumptions on monotonicity or discreteness are made. Møller & Schladitz (1999) review the algorithm in terms of stochastic recursive sequences and extend it for general discrete repulsive models.

1.4.1 Coupling From the Past (CFTP)

A main problem with MCMC techniques is determining how long to run the chain/process until it is deemed to be 'near' equilibrium, ie. MCMC algorithms only provide an *approximate* draw from the target distribution. In some cases, however, it is possible to devise a clever algorithm which manages to 'determine by itself' when the chain has converged. In this instance one obtains an *exact* or *perfect* sample, ie. an unbiased sample from the target distribution in finite time; such variants are coined *perfect simulation algorithms*.

Coupling From the Past (CFTP) is one of the more widely studied and used perfect simulation protocols. It is a probabilistic algorithm or 'simulation recipe' for sampling Markov chains perfectly. The technique, introduced by Propp & Wilson (1996), was originally proposed for finite chains. It

has since received much attention in the literature and has been generalized to a variety of chains and unbounded state spaces including spatial point process models (Murdoch & Green 1998; Foss & Tweedie 1998; Häggström *et al.* 1999; Green & Murdoch 1999; Kendall & Møller 2000; Thönnes 2000; Murdoch & Rosenthal 2000; Berthelsen & Møller 2002b). Several primers on CFTP and perfect simulation for point processes are also available (Thönnes 1997; Møller 2001; Berthelsen & Møller 2002a); Propp & Wilson (1998) gives a comprehensive overview on applications of CFTP.

An MCMC algorithm for sampling (approximately) from the target distribution π involves the construction of an ergodic Markov chain whose time-stationary (equilibrium) distribution is π. The idea then is to initialize the chain at some arbitrary state and simulate it for a 'long' time. The MCMC sampler is constructed so as to ensure the transitions of the chain preserve the distribution π. An obvious difficulty with this is how 'long' should the chain be run in order to ensure that it has converged or that its distribution is 'close enough' to π? If the chain could be started at time $-\infty$ then the state of the chain at time 0 would be an exact draw from π, since the chain is ergodic.

An actual simulation from time $-\infty$ is infeasible; however since the interest lies predominantly in the state of the chain at time 0, what if the path of the chain in a *virtual simulation* (Kendall & Thönnes 1999) from time $-\infty$ could be re-constructed for a finite time interval $[-T, 0]$? This would provide a practical way of determining the state of the chain at time 0. This is the essence of CFTP: it intrinsically determines how far back in the past the chain needs to be started from in order to reconstruct its infinite-time simulation path on $[-T, 0]$, for some $T < \infty$. Furthermore the chain at time 0 is guaranteed to have the same value as would have been obtained in a virtual simulation from time $-\infty$. So let Φ^∞ denote the target chain started at time $-\infty$ whose stationary distribution is the target distribution π and let $\{0, \ldots, k\}$ represent the finite state space of the chain.

For $T > 0$ and $-T \leq t \leq 0$, $\Phi^\infty(t)$ must take on one of the values $0, \ldots, k$. So, for $0 \leq i \leq k$, let Φ_i^T denote a chain started at time $-T$ in state i. Suppose $\left\{\Phi_i^T\right\}_{i=0}^k$ and Φ^∞ are coupled and evolved using the same randomness on the interval $[-T, 0]$. If $\Phi_i^T(0) = \Phi_j^T(0)$ for all $i \neq j$ then this is referred to as *coalescence*, ie. all the chains started in each of the $k+1$ possible states have met and take on a single common value at time 0. In this case it must follow that $\Phi^\infty(0)$ also has the same common value since $\Phi^\infty(-T)$ is equal to $\Phi_i^T(-T)$, for some i, and the coupling ensures this equality persists for all $-T \leq t \leq 0$. If coalescence is not achieved by time 0, the whole procedure is started off from time $-S < -T$. Care must be taken to ensure that the same randomness is used

over the interval $[-T, 0]$, as failure to do so will introduce bias (Propp & Wilson 1996). Such a procedure is also called *vertical CFTP*, since the target chains Φ_i^T are started at every state i.

Foss & Tweedie (1998) point out that CFTP can be viewed in the context of "backward couplings in stochastic recursive sequences theory"; such methods have been around for a while and indeed Propp & Wilson (1998) comment that "the conceptual ingredients were up in the air even before the versatility of the method was made clear". Foss & Tweedie thus attempt to embed CFTP in a general probabilistic framework by describing the theoretical framework that leads to forward and backward coupling constructions. Further theory and historical origins can be found in their paper as well as in Propp & Wilson (1998) who identify some precursors to the ideas behind CFTP.

Vertical CFTP can be employed for any finite ergodic Markov chain. However as the state space gets larger it becomes impractical and computationally burdensome to keep track of *all* the Φ_i^T chains. Suppose there exists a partial order $\preccurlyeq$ on the state space with maximal and minimal elements $\mathbf{1}$ and $\mathbf{0}$ respectively. In addition suppose also that the transitions of the target chain Φ preserve the partial ordering, ie. the transition function is monotone (Definition 1.30). In this case CFTP can be implemented more efficiently by keeping track of only two chains: $\Phi_{\mathbf{1}}^T$ and $\Phi_{\mathbf{0}}^T$. This is because coalescence of all the Φ_i^T chains is equivalent to the coalescence of $\Phi_{\mathbf{1}}^T$ and $\Phi_{\mathbf{0}}^T$, since the transitions of the chain maintain the ordering with respect to $\preccurlyeq$. This is referred to as *Monotone CFTP*.

A natural extension is to the case of continuous or even unbounded state spaces. Such a consideration is both natural and important, especially when dealing with point processes since the state space is unbounded. It turns out that there is a neat extension of CFTP called *Dominated CFTP (domCFTP)* developed by Kendall (1998) for the area-interaction point process, and generalized by Kendall & Møller (2000) for any point process with a *locally stable* density (Definition 1.11). Wilson (2000) also use the term *Coupling Into And From The Past (CIAFTP)* for domCFTP.

Häggström & Nelander (1998) consider perfect sampling of anti-monotone systems, such as point processes with repulsive densities (cf. Eq. 1.11). Häggström *et al.* (1999) describe an exact Gibbs sampler for the bivariate Widom & Rowlinson (1970) penetrable spheres model, employing quasi-minimal and quasi-maximal states in order to devise a CFTP-based construction. Another addition, *read-once CFTP* (Wilson 2000), involves running the Markov chain forwards in time and never restarting it in the past. It has the advantage of not needing to store random numbers for reuse, whereas conventional CFTP does. David Wilson's web site contains an annotated bibliography on

perfect simulation: `http://dimacs.rutgers.edu/~dbwilson/exact.html/`.

1.4.2 Dominated Coupling From the Past (domCFTP)

Foss & Tweedie (1998) point out that the CFTP protocol requires uniform ergodicity (Definition 1.23) of the target process. In unbounded state spaces there is typically no maximal element and the (crucial) uniform ergodicity of the target chain is the exception rather than the rule. In this case the standard procedure needs modification. The main idea behind domCFTP is to introduce a *dominating* process which acts as a stochastically varying maximum. Perfect simulation will then be viable if the dominating process can itself be simulated in statistical equilibrium and in reverse-time.

It is assumed that there is a partial order $\preccurlyeq$ and the state space has a minimal element $\mathbf{0}$, but no maximal element. Let Ψ denote the target chain and consider a virtual simulation of Ψ from time $-\infty$, denoted by Ψ^{∞}. Furthermore suppose that there is a process Φ on the same state space such that $\Psi(t) \preccurlyeq \Phi(t)$ for all t; then Φ is the dominating process. If Φ can be sampled in stationarity and in reverse-time, then it is easy to reconstruct a path of Φ^{∞} in a finite interval $[-T, 0]$; refer to this as *extending* the dominating process backwards on $[-T, 0]$. Since Ψ^{∞} is coupled to Φ^{∞}, the path of Ψ^{∞} is always bounded by that of Φ^{∞} on the same interval.

At time $-T$ define *upper-* and *lower-sandwich* processes $\Psi^{T,max}$ and $\Psi^{T,min}$ respectively by $\Psi^{T,max}(-T) = \Phi^{\infty}(-T)$ and $\Psi^{T,min}(-T) = \mathbf{0}$. The purpose of these sandwich processes is to bound the target process Ψ^{∞} on the interval $[-T, 0]$. If these processes coalesce by time 0 then their common value must also be the value of Ψ^{∞}. Dominated CFTP has also been referred to as *horizontal CFTP*, since it involves starting the bounding processes $\Psi^{T,max}, \Psi^{T,min}$ further and further back in time until coalescence at time 0 occurs. The evolution of $\Psi^{T,max}$ and $\Psi^{T,min}$ (forwards in time) is coupled to that of Φ^{∞} on the interval $[-T, 0]$ in such a way so that the following properties hold for all $-S \leq -T \leq s \leq t \leq 0$:

$$\textbf{sandwiching:} \qquad \Psi^{T,min}(t) \preccurlyeq \Psi^{\infty}(t) \preccurlyeq \Psi^{T,max}(t) \preccurlyeq \Phi^{\infty}(t). \tag{1.26}$$

$$\textbf{coalescence:} \qquad \Psi^{T,min}(t) = \Psi^{T,max}(t) \qquad \text{if} \qquad \Psi^{T,min}(s) = \Psi^{T,max}(s). \tag{1.27}$$

$$\textbf{funnelling:} \qquad \Psi^{T,min}(t) \preccurlyeq \Psi^{S,min}(t) \preccurlyeq \Psi^{S,max}(t) \preccurlyeq \Psi^{T,max}(t). \tag{1.28}$$

It is not required that $\Psi^{T,max}$ and $\Psi^{T,min}$ typically evolve like Ψ^{∞} or that they be individually Markovian. However as soon as they coalesce, ie. $\Psi^{T,max} = \Psi^{T,min}$ then the above sandwiching,

coalescence and funnelling properties ensure that their subsequent common evolution follows that of Ψ^{∞}. Furthermore, conditional upon coalescence, the common value of the sandwich processes at time 0 will be equal to $\Psi^{\infty}(0)$, and hence an exact draw from the target distribution (Kendall & Møller 2000, Theorem 2.1).

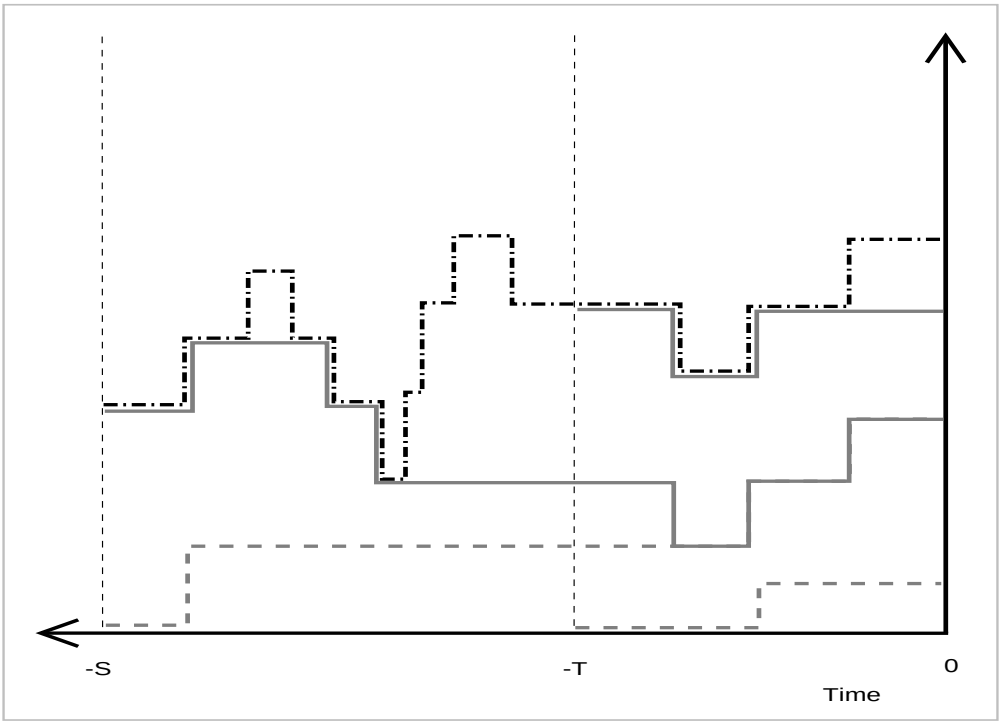

Figure 1.1: Illustration of dominated CFTP. The black dashed-dotted line depicts the evolution a dominating process (eg. immigration-death process on the integers). The solid (resp. dotted) grey line shows the maximal (resp. minimal) processes started at times $-T$ and $-S$. The dominating process is extended backwards till time $-T$. The maximal process is started in the state of the dominating process at time $-T$; the minimal is started in the minimal state. These bounding processes are evolved forwards in time, coupled to the dominating process. Here coalescence of the maximal and minimal processes started at time $-T$ does not occur at time 0; so the dominating process is extended further from time $-T$ to $-S$. The maximal and minimal process are evolved in a similar fashion as previously; coalescence now occurs on $[-T, 0]$. Notice that the same realization of the dominating process on $[-T, 0]$ is used when evolving the bounding processes started at time $-S < -T$. The sandwiching, coalescence and funnelling properties are satisfied by the maximal and minimal processes; hence processes started before time $-S$ will also have coalesced by time 0.

Chapter 2

Stochastic Domination & Point Processes

2.1 Introduction

This chapter is based on the research report Shah (2003b) and subsequent publication Shah (2003a), and concerns the idea of *stochastic domination and ordering* in point processes. The motivation for this work came from the paper by Van Lieshout & Van Zwet (2001), who propose a simulation algorithm (which we denote by VLVZ) for obtaining samples from conditional Boolean models. The algorithm was intended to be *perfect* and concerns a germ-grain random set model (cf. Definition 1.10) where germs are homogeneous Poisson and grains disks of fixed radius. They considered sampling a Poisson process conditioned to belong to a conditioning set $\mathcal{E}$ with the property that "if $x \in \mathcal{E}$, then for any ξ, $x \cup \{\xi\} \in \mathcal{E}$ " Van Lieshout & Van Zwet (2001).

Sets $\mathcal{E}$ satisfying such a property are referred to as *anti-hereditary* (Definition 2.4). Incidentally, Kendall & Møller (2000) present a perfect simulation algorithm for point processes with locally stable densities (Definition 1.11). This means that their method can be employed to sample point processes conditioned to belong to *hereditary* conditioning sets $\mathcal{E}$, ie. if $x \in \mathcal{E}$ then $x \setminus \{\xi\} \in \mathcal{E}$ for all $\xi \in x$. Kendall & Thönnes (1999) and Cai & Kendall (2002) present perfect simulation algorithms for sampling conditional Boolean models where the conditioning event was the coverage of a finite set of points. As these algorithms cannot be generalized to deal with other anti-hereditary conditioning events, this spurred an investigation into the VLVZ algorithm.

We implemented their algorithm for the special case of coverage of a singleton set. Validation of our implementation via χ^2 goodness-of-fit tests uncovered a discrepancy between the expected and

observed counts. A further investigation resulted in the discovery that, unfortunately, there was an error in their method, ie. it produces biased samples (see Shah 2003a and the correction note Van Lieshout & Van Zwet 2003). This chapter explores the nature of this bias.

The issue of bias for VLVZ motivates a question concerning *stochastic domination and ordering* in a point process context since, in some cases, it produces samples which are *strictly stochastically smaller* than the target distribution. The notion of 'stochastic dominance' is made precise via the definition of stochastic orders. Corollary 2.1 extends Theorem 9.1 of Preston (1977), giving conditions when one probability measure *strictly* dominates another. We have discussed a fix with Van Lieshout & Van Zwet; both parties agree that this fix would not be any more efficient than direct Rejection sampling. Thus efficient sampling from Boolean models conditioned to belong to *any* anti-hereditary set is "still an open problem" Van Lieshout & Van Zwet (2003).

This chapter is organized as follows: following Shaked & Shanthikumar (1994) & Kamae *et al.* (1977) the notion of stochastic dominance is introduced in Section 2.2. We first consider stochastic orders on totally ordered spaces and establish some domination results for random variables obtained via accept/reject mechanisms and conditional thinnings (Section 2.2.1). Section 2.2.2 then deals with partially ordered spaces and in particular, *Exponential Spaces* (cf. Section 1.1.1), which are a natural state space for point processes.

Section 2.3.1 describes four algorithms, including VLVZ, that have the objective of sampling from Boolean models conditioned to belong to some conditioning set $\mathcal{E}$. The distributions of the outputs of the various algorithms are compared via conditional expectation and the bias of VLVZ is established. Theorem 2.5 employs Corollary 2.1 to give conditions when VLVZ produces strictly smaller samples; the case when $\mathcal{E}$ consists of all configurations which cover a single point satisfies these conditions. The coverage of two points is also considered; in this case Examples 2.2 & 2.3 illustrate that VLVZ can produce strictly larger and neither smaller nor larger samples respectively.

2.2 Stochastic Orders & Conditional Thinning

The notion of stochastic domination is now made precise via the definition of stochastic orders. There are *numerous* stochastic orders (see for example Shaked & Shanthikumar 1994) which illustrates the many different, often inequivalent, ways in which the idea of stochastic domination can

be defined. The purpose of this section is to derive some domination results for random variables obtained by *conditional thinning* procedures, the idea behind which is as follows.

Consider some space $\mathcal{X}$ endowed with a partial order $\preccurlyeq$. Let $\theta : \mathcal{X} \to \mathcal{X}$ denote a random function which maps $x \in \mathcal{X}$ to some $y \preccurlyeq x$, ie. $\theta(x) \equiv \theta_x \preccurlyeq x$; then θ_x is referred to as a *thinning* of x. Suppose that $\mathcal{E} \subset \mathbb{R}_+$ is some conditioning set, N is a random variable conditioned to belong to $\mathcal{E}$ and X is obtained from a realization of N via the following accept/reject mechanism. The map θ is applied to N; if $\theta_N \in \mathcal{E}$ then $X = \theta_N$. Otherwise another realization of θ_N is generated independently. This iterative procedure is carried on until a $\theta_N \in \mathcal{E}$ is obtained. Then X is referred to as a *conditional thinning* of N since the law of X, $\mathcal{L}(X) \equiv \mathcal{L}\left(\theta_N \mathbf{1}_{\{\theta_N \in \mathcal{E}\}}\right)$. The motivation for such an investigation came from the algorithm proposed by Van Lieshout & Van Zwet (2001) for sampling Boolean models conditioned to belong to some $\mathcal{E}$. The essence of their method is to generate a process with higher intensity $\kappa > \lambda$ and then conditionally thin it to yield a process with intensity λ.

It was discovered that if $\mathcal{E}$ represents those configurations which cover single point, then their algorithm produces samples whose distribution is 'smaller' in some sense than the target. This naturally warranted an investigation as to whether such 'conditional thinning' procedures always produce 'smaller' samples, ie. for all types of conditioning sets $\mathcal{E}$. The results of such an investigation are presented here, starting with some domination results on totally ordered spaces (such as $\mathbb{R}$) and then moving on to partially ordered spaces.

2.2.1 Totally Ordered Spaces

In this section the treatment is made using random variables defined on $\mathbb{R}$, but the ideas and results generalize trivially to any totally ordered space. Two stochastic orders are defined below and some domination results on random variables obtained by some conditional accept/reject mechanism are established. Following Shaked & Shanthikumar (1994) define the *usual stochastic order* between random variables on $\mathbb{R}$ as:

Definition 2.1. If X and X' are two random variables such that $\mathbb{P}[X \geq n] \leq \mathbb{P}[X' \geq n]$ for $n \in (-\infty, \infty)$, then X is defined to be smaller than X' in the *usual stochastic order*, written as $X \preccurlyeq_{st} X'$. Also, X is *strictly* smaller than X', written $X \prec_{st} X'$, if $\mathbb{P}[X \geq n] < \mathbb{P}[X' \geq n]$ for at least one n.

Definition 2.2. Let X and X' be continuous (discrete) random variables with densities (mass functions) f and g, respectively. If $f(u)\,g(v) \geq f(v)\,g(u)$ for all $u \leq v$; or equivalently if $\frac{f(u)}{g(u)}$ decreases over the union of the supports of X and X', then X is smaller than X' in the *likelihood ratio order*, written $X \preccurlyeq_{lr} X'$. For the strict case, $X \prec_{lr} X'$ if $f(u)\,g(v) \geq f(v)\,g(u)$ for all $u \leq v$ with strict inequality for at least one $\{(u,v);\ u < v\}$ with positive measure. In the case when f, g are mass functions it suffices to check the condition $f(u)\,g(v) > f(v)\,g(u)$ for at least one $u < v$ since this will have positive measure.

Remark 2.1. Shaked & Shanthikumar (1994) deal with *weak orders*, ie. $\preccurlyeq_{st}$ & $\preccurlyeq_{lr}$; the strict orders $\prec_{st}$ & $\prec_{lr}$ are defined here since these are required to ensure the existence of a bias. These strict orders are defined as a natural extensions of the weak ones.

From Definition 2.1 it follows that $X \preccurlyeq_{st} (\prec_{st})\, X' \Leftrightarrow \mathbb{E}[g(X)] \leq (<)\, \mathbb{E}[g(X')]$ for all increasing (strictly increasing) measurable functions g for which the expectations exist. Shaked & Shanthikumar show that $X \preccurlyeq_{lr} X' \Rightarrow X \preccurlyeq_{st} X'$ so that the likelihood ratio ordering is a stronger form of stochastic dominance; in fact one can also show that $X \prec_{lr} X' \Rightarrow X \prec_{st} X'$.

Lemma 2.1. *Let X, X' be random variables on $\mathbb{R}$; if $X \prec_{lr} X'$ then $X \prec_{st} X'$.*

Proof. The result holds for both continuous and discrete random variables; however for simplicity suppose that X, X' are discrete random variables on $\{0, 1, \ldots\}$ with mass functions f, g respectively. Fix $n \geq 1$; then

$$\begin{aligned}
\mathbb{P}[X \leq n] - \mathbb{P}[X' \leq n] &= \sum_{i=0}^{n} f_i - \sum_{i=0}^{n} g_i = \sum_{i=0}^{n} (g_0 + g_1 + \ldots)\, f_i - \sum_{i=0}^{n} (f_0 + f_1 + \ldots)\, g_i \\
&= \sum_{j=0}^{n} g_j\,(f_0 + \ldots + f_n) + \sum_{j \geq n+1} g_j\,(f_0 + \ldots + f_n) \\
&\quad - \sum_{j=0}^{n} f_j\,(g_0 + \ldots + g_n) - \sum_{j \geq n+1} f_j\,(g_0 + \ldots + g_n) \\
&= \sum_{j \geq n+1} g_j\,(f_0 + \ldots + f_n) - \sum_{j \geq n+1} f_j\,(g_0 + \ldots + g_n).
\end{aligned}$$

Now $X \prec_{lr} X'$ implies that $f(i)\,g(j) \geq f(j)\,g(i)$ for all $i \leq j$, with a strict inequality for at least one $i < j$. Thus $\mathbb{P}[X \leq n] - \mathbb{P}[X' \leq n] \geq 0$; since n was arbitrary this holds for all $n \geq 0$, with strict inequality for at least one n. Hence $X \prec_{st} X'$ by Definition 2.1. □

Some Domination Results

Having set up the basic definitions of stochastic dominance in Definitions 2.1 & 2.2 we are now ready to look at some domination results for random variables produced by conditional thinning procedures. Suppose $\mathcal{E} \subseteq \mathbb{R}_+$ and N is some random variable on $\mathbb{R}_+$ conditioned to belong to $\mathcal{E}$. Recall that θ is a random map which takes some $x \in \mathbb{R}_+$ and maps it to some $y \leq x$. All the results in this section hold for both discrete and continuous random variables. However, for simplicity, suppose that N is discrete, with mass function q_n; define another discrete random variable N' which is generated via the following mechanism:

Set $i = 1$.

while $i \neq 0$:

draw N and sample θ_N.

if $\theta_N \in \mathcal{E}$: set $i = 0$; **else**: set $i = i + 1$.

return $N' = N$

Let $\mathbb{P}[\theta_N \in \mathcal{E} \mid N] = w_N$, so that $\mathbb{P}[N' = n] \propto q_n w_n$.

Theorem 2.1. *Let N be a discrete random variable with mass function q_n conditioned to belong to some conditioning set $\mathcal{E}$, and let N' be generated by the above procedure. If the sequence of acceptance weights $\{w_n\}$ is increasing (strictly increasing) in n then $N \preccurlyeq_{lr} N'$ ($N \prec_{lr} N'$). Conversely, if the sequence is decreasing (strictly decreasing) then $N \succcurlyeq_{lr} N'$ ($N \succ_{lr} N'$).*

Proof. Fix $u \leq v$; for an increasing sequence $\{w_n\}$ $q_u w_v q_v \geq q_v w_u q_u$ holds. Definition 2.2 is satisfied and hence $N \preccurlyeq_{lr} N'$. For a strictly increasing sequence, if $u < v$ then $q_u w_v q_v > q_v w_u q_u$, and hence $N \prec_{lr} N'$. The converse result for a decreasing (strictly decreasing) sequence of weights follows by reversing the inequalities (strict inequalities) above. □

Suppose that, for each n, P_n denotes the distribution of θ_n on $\{0, 1, \ldots, n\}$ conditioned to belong to $\mathcal{E}$. Define *conditionally thinned* random variables X and X', generated from N and N' respectively as follows: $X \mid N = n \sim P_n$ and $X' \mid N' = n' \sim P_{n'}$. That is, X is a conditional thinning (cf. Section 2.2) of N and X' that of N'.

Theorem 2.2. *Suppose N and N' are two random variables conditioned to belong to some $\mathcal{E}$, with mass functions q_n and q'_n respectively. For each n, let P_n be the distribution of θ_n, on $\{0, \ldots, n\}$,*

also conditioned to belong to $\mathcal{E}$. Generate conditionally thinned random variables X and X' as follows: $X \mid N = n \sim P_n$ and $X' \mid N' = n' \sim P_{n'}$. If $P_n \preccurlyeq_{lr} P_{n'}$, for $n \leq n'$, and $N \preccurlyeq_{lr} N'$ then $X \preccurlyeq_{lr} X'$. Furthermore if $P_n \prec_{lr} P_{n'}$, for $n < n'$, and $N \prec_{lr} N'$ then $X \prec_{lr} X'$.

Proof. Let $p_{n,r}$, for $r \in \{0, \ldots, n\}$, denote the mass function of P_n so that $\mathbb{P}[X = r] = \sum_{n \geq r} q_n p_{n,r}$ and similarly for X'. Fix $u \leq v$; if $\mathbb{P}[X = v]\,\mathbb{P}[X' = u] \leq \mathbb{P}[X' = v]\,\mathbb{P}[X = u]$ then $X \preccurlyeq_{lr} X'$ by Definition 2.2. Since P_n is defined on $\{0, \ldots, n\}$, set $p_{n,r} = 0$ for $r > n$ and consider

$$\begin{aligned}
\mathbb{P}[X = v]\,\mathbb{P}[X' = u] &= (q_v p_{v,v} + \ldots)\left(q'_u p_{u,u} + \ldots + q'_{v-1} p_{v-1,u}\right) + (q_v p_{v,v} + \ldots)\,(q'_v p_{v,u} + \ldots) \\
&= \sum_{i \geq v} \sum_{j=u}^{v-1} q_i p_{i,v} q'_j p_{j,u} + \sum_{i \geq v} \sum_{j \geq v} q_i p_{i,v} q'_j p_{j,u} = \sum_{i \geq v} \sum_{j=u}^{v-1} Q_{ij} + \sum_{i \geq v} \sum_{j \geq v} Q_{ij}. \\
\mathbb{P}[X' = v]\,\mathbb{P}[X = u] &= (q'_v p_{v,v} + \ldots)\,(q_u p_{u,u} + \ldots + q_{v-1} p_{v-1,u}) + (q'_v p_{v,v} + \ldots)\,(q_u p_{v,u} + \ldots) \\
&= \sum_{i \geq v} \sum_{j=u}^{v-1} q'_i p_{i,v} q_j p_{j,u} + \sum_{i \geq v} \sum_{j \geq v} q'_i p_{i,v} q_j p_{j,u} = \sum_{i \geq v} \sum_{j=u}^{v-1} Q'_{ij} + \sum_{i \geq v} \sum_{j \geq v} Q'_{ij}.
\end{aligned}$$

For $n \leq n'$, $N \preccurlyeq_{lr} N'$ and $P_n \preccurlyeq_{lr} P_{n'}$ imply $q_i q'_j \leq q'_i q_j$ and $p_{n,i} p_{n',j} \leq p_{n',i} p_{n,j}$ for $j \leq i \leq n'$ (with strict inequalities for at least one $j < i$ if $N \prec_{lr} N'$ and $P_n \prec_{lr} P_{n'}$ for $n < n'$). Thus for

$$\begin{aligned}
i \geq v \text{ and } u \leq j \leq v-1: \qquad & Q_{ij} - Q'_{ij} = p_{i,v} p_{j,u}\left(q_i q'_j - q'_i q_j\right) \leq 0; \\
i = j \geq v: \qquad & Q_{ij} - Q'_{ij} = 0; \\
i > j \geq v \geq u: \qquad & Q_{ij} + Q_{ji} - Q'_{ij} - Q'_{ji} = p_{i,v} p_{j,u}\left(q_i q'_j - q'_i q_j\right) - p_{j,v} p_{i,u}\left(q_i q'_j - q'_i q_j\right) \\
& \qquad = \left(p_{i,v} p_{j,u} - p_{j,v} p_{i,u}\right)\left(q_i q'_j - q'_i q_j\right) \leq 0
\end{aligned}$$

where the last inequality follows since $i > j \geq v \geq u$ implies $(p_{i,v} p_{j,u} - p_{j,v} p_{i,u}) \geq 0$ and $\left(q_i q'_j - q'_i q_j\right) \leq 0$ (with strict inequality for at least one $j < i$ in the case when $N \prec_{lr} N'$ and $P_n \prec_{lr} P_{n'}$ for $n < n'$). So

$$\mathbb{P}[X = v]\,\mathbb{P}[X' = u] = \sum_{i \geq v} \sum_{j=u}^{v-1} Q_{ij} + \sum_{i=j \geq v} Q_{ij} + \sum_{i > j} \sum_{j \geq v} Q_{ij} + Q_{ji}; \tag{2.1}$$

$$\mathbb{P}[X' = v]\,\mathbb{P}[X = u] = \sum_{i \geq v} \sum_{j=u}^{v-1} Q'_{ij} + \sum_{i=j \geq v} Q'_{ij} + \sum_{i > j} \sum_{j \geq v} Q'_{ij} + Q'_{ji}. \tag{2.2}$$

Therefore $(2.1) - (2.2) \leq 0$; $X \preccurlyeq_{lr} X'$ follows from Definition 2.2. If $N \prec_{lr} N'$ and $P_n \prec_{lr} P_{n'}$, for $n < n'$, then the above inequality will be strict for at least one $u < v$; so $X \prec_{lr} X'$ follows. □

In the case when N, N' and $P_n, P_{n'}$, for $n \leq n'$, are not ordered with respect to the likelihood order but *are* ordered with respect to the usual stochastic order then the thinned random variables X, X' are also ordered with respect to the usual stochastic order.

Theorem 2.3. *Suppose that N, N', P_n and X, X' are defined as in the statement of Theorem 2.2. If the family $\{P_n\}$ is such that $P_n \preccurlyeq_{st} P_{n'}$, for $n \leq n'$, and $N \preccurlyeq_{st} N'$ then $X \preccurlyeq_{st} X'$. Furthermore if $P_n \prec_{st} P_{n'}$, for $n < n'$, and $N \prec_{st} N'$ then $X \prec_{st} X'$.*

Proof. For fixed n let X_n be a random variable with distribution P_n so that $n \leq n'$ implies $X_n \preccurlyeq_{st} X_{n'}$; equivalently, for an increasing function g, $\mathbb{E}\left[g\left(X_n\right)\right] \leq \mathbb{E}\left[g\left(X_{n'}\right)\right]$; so $\mathbb{E}\left[g\left(X_n\right)\right]$ is increasing in n. Now $N \preccurlyeq_{st} N'$ and $\mathbb{E}\left[g\left(X_n\right)\right]$ increasing implies $\mathbb{E}_N\mathbb{E}\left[g\left(X_N\right)\right] \leq \mathbb{E}_{N'}\mathbb{E}\left[g\left(X_{N'}\right)\right]$; moreover

$$\begin{aligned}\mathbb{E}\left[g\left(X\right)\right] = \mathbb{E}_N\mathbb{E}\left[g\left(X\right) \mid N\right] &= \mathbb{E}_N\mathbb{E}\left[g\left(X_N\right)\right] \\ &\leq \mathbb{E}_{N'}\mathbb{E}\left[g\left(X_{N'}\right)\right] = \mathbb{E}_{N'}\mathbb{E}\left[g\left(X'\right) \mid N'\right] = \mathbb{E}\left[g\left(X'\right)\right]\end{aligned}$$

so that $X \preccurlyeq_{st} X'$; $X \prec_{st} X'$ also follows, but with strict inequalities and g strictly increasing. □

So if two random variables N, N' conditioned to belong to some $\mathcal{E}$ are such that $N \prec_{st} N'$ and $P_n \prec_{st} P_{n'}$, for $n < n'$, then their respective conditional thinnings X, X' are also ordered, ie. $X \prec_{st} X'$ for *any* $\mathcal{E}$. Thus if the map θ respects the ordering between N and N' (ie. if $P_n \prec_{st} P_{n'}$ for $n < n'$) then, for totally ordered spaces, "the conditional thinning of a (stochastically) smaller process is always (stochastically) smaller". This is quite an intuitive result which, however, does not always hold in partially ordered spaces. For the analogous result to hold in such spaces an additional *lattice* structure on $\mathcal{E}$ is required (Definition 2.5).

2.2.2 Partially Ordered Spaces

In this section we generalize the definition of stochastic dominance to partially ordered spaces and, in particular, to *Exponential Spaces* (cf. Section 1.1.1). The ultimate objective of this chapter is to investigate conditional thinning simulation algorithms for point processes (Section 2.3.1). The definition of stochastic domination is given below, which is the analogue of that in Definition 2.1. Suppose that $\mathcal{X}$ is some Polish space endowed with a closed partial order '$\preccurlyeq$' and P_1, P_2 are two probability measures on $\mathcal{X}$. Following Kamae *et al.* (1977) define the notion of stochastic domination between the P_i:

Definition 2.3. For $\mathcal{X}$, P_1 and P_2 as defined above P_1 is *stochastically smaller* than P_2, written $P_1 \preccurlyeq_{st} P_2$, if and only if $\int g\, dP_1 \leq \int g\, dP_2$ for all bounded increasing (with respect to $\preccurlyeq$) functions g defined on $\mathcal{X}$. We also define P_1 to be *strictly* stochastically smaller than P_2, written $P_1 \prec_{st} P_2$, if and only if $\int g\, dP_1 \leq \int g\, dP_2$ for all bounded increasing g, with strict inequality for at least one g.

Recall the notion of a 'coupling' from Section 1.2.1. A stochastic ordering between two probability measures is equivalent to the existence of a coupling of random variables which are ordered and have distributions given by the two probability measures. This result of Kamae *et al.* (1977) (a special case of Strassen 1965, Theorem 11) is stated as Theorem 1.3 of Chapter 1.

Exponential Spaces & Point Processes

The notion of an 'Exponential Space' was introduced by Carter & Prenter (1972) as the underlying state space for counting or point processes. For some space $\mathcal{X}$ the construction of its exponential space, $\mathcal{X}_e$, is given in Section 1.1.1, following Preston (1977). A point process on $\mathcal{X}$ is defined to be an $\mathcal{X}_e$-valued random variable (cf. Definition 1.1). Endow $\mathcal{X}_e$ with the inclusion order '$\subseteq$', where $y \subseteq x$ if there exists $z \in \mathcal{X}_e$ such that $y \cup z = x$, ie. $y \subseteq x$ if y is a sub-configuration of x.

Proposition 2.1. *The inclusion order $\subseteq$ is closed with respect to the metric in Carter & Prenter.*

Proof. Carter & Prenter (1972) define a metric δ on $\mathcal{X}_e$ by: for $x = \left\{\xi_1, \ldots, \xi_{n(x)}\right\}, y = \left\{\eta_1, \ldots, \eta_{n(y)}\right\}$

$$\delta(x,y) = \begin{cases} |n(x) - n(y)| & \text{if} \quad n(x) \neq n(y) \text{ or } n(x) = n(y) = 0; \\ \inf_\sigma \sup_i d\left(\xi_i, \eta_{\sigma(i)}\right) & \text{else.} \end{cases}$$

Here d is the metric on $\mathcal{X}$, σ is a permutation on $\{1, \ldots, n\}$ and $n(x)$ is the number of points in x. It is required to show that if $\{x_n\}$ and $\{y_n\}$ are sequences in $\mathcal{X}_e$ tending to x and y respectively, such that $y_n \subseteq x_n$ for each n, then $y \subseteq x$. For each n let the $x_n = \left\{\xi_{n_1}, \ldots, \xi_{n_{n(x)}}\right\}$ and $y_n = \left\{\eta_{n_1}, \ldots, \eta_{n_{n(y)}}\right\}$.

So suppose $y \nsubseteq x$; then $\exists\ \eta_{j^*} \in y$ such that $\eta_{j^*} \notin x$. Let $\varepsilon^* = \inf_i d(\eta_{j^*}, \xi_i) > 0$ since $\eta_{j^*} \notin x$, and choose $\varepsilon < \varepsilon^*$. If $x_n \to x$ and $y_n \to y$ then, for sufficiently large n, it must be the case that $n(x_n) = n(x)$ and $n(y_n) = n(y)$. Thus there exists N_ε^y such that $n \geq N_\varepsilon^y \Rightarrow \delta(y_n, y) < \varepsilon$, ie. there exists some $\eta_{n_{i^*}} \in y_n$ such that $d(\eta_{n_{i^*}}, \eta_{j^*}) < \varepsilon$. Also, $y_n \subseteq x_n$ means that $\eta_{n_{i^*}} \equiv \xi_{n_{i^*}} \in x_n$. Since $x_n \to x$ then there exists N_ε^x such that $n \geq N_\varepsilon^x \Rightarrow$

$\delta(x_n, x) < \varepsilon$. So, for $n \geq N_\varepsilon^x \vee N_\varepsilon^y$, consider $\delta(x_n, x) = \inf_\sigma \sup_i d\left(\xi_{n_i}, \xi_{\sigma(i)}\right)$. For any permutation σ, $d\left(\eta_{j^*}, \xi_{\sigma(i^*)}\right) \leq d\left(\eta_{j^*}, \xi_{n_{i^*}}\right) + d\left(\xi_{n_{i^*}}, \xi_{\sigma(i^*)}\right)$, as d is a metric. From above we have that $d\left(\xi_{n_{i^*}}, \xi_{\sigma(i^*)}\right) \geq d\left(\eta_{j^*}, \xi_{\sigma(i^*)}\right) - d\left(\eta_{j^*}, \xi_{n_{i^*}}\right) > \varepsilon^* - \varepsilon$. But this implies that $\delta(x_n, x) = \inf_\sigma \sup_i d\left(\xi_{n_i}, \xi_{\sigma(i)}\right) \geq \inf_\sigma d\left(\xi_{n_{i^*}}, \xi_{\sigma(i^*)}\right) > \varepsilon^* - \varepsilon > 0$, contradicting $x_n \to x$. Hence the assumption that $y \nsubseteq x$ must be wrong and the result follows. □

Definition 2.3 and Theorem 1.3 hence enable the following equivalence:

(i) $P_1 \preccurlyeq_{st} P_2$, where the P_i are probability measures on $\mathcal{X}_e$.

(ii) There exist $\mathcal{X}_e$-valued random variables (ie. point processes) $X_i \sim P_i$ such that $X_1 \subseteq X_2$ a.s.

2.2.3 Coupled Birth-Death Processes & Stochastic Domination

In light of Theorem 1.3 from Chapter 1 and statements (i) & (ii) above, a natural way to show (stochastic) ordering between two probability measures is to construct Markov processes with these measures as equilibrium distributions such that the processes are ordered for all time. If the two probability measures admit densities with respect to some fixed measure then Preston (1977, Theorem 9.1) gives sufficient conditions for the densities in order to ensure that their corresponding probability measures are stochastically ordered.

The proof of this Theorem employs spatial birth-death processes (Section 1.3) in order to establish the domination result. The purpose of this section is to describe Preston's construction; the reason for presenting it explicitly is because we will refer to the construction in later parts of the thesis and so it proves convenient to state it separately. In addition, conditions guaranteeing *strict* stochastic dominance are also developed here (Corollary 2.1), and used in Section 2.3.2 in order to establish some domination results.

Suppose that $\mathcal{X}$ is some space and $\mathcal{X}_e$ its exponential space, endowed with the inclusion order '$\subseteq$'. For $i = 1, 2$ let μ_i denote a probability measure on $\mathcal{X}_e$ with density f_i with respect to some fixed measure $\tilde{\omega}$ on $\mathcal{X}_e$. Let Φ_i be a spatial birth-death process with birth rate $b_i(x, \xi) = \frac{f_i(x \cup \{\xi\})}{f_i(x)}$ (with convention that $\frac{0}{0} = 0$) and unit death rate per point. Detailed balance equations (Eq. 1.25) show that the equilibrium distribution of Φ_i is μ_i. Furthermore let $\tilde{\Phi}$ be a spatial birth-death process on $\mathcal{X}_e \times \mathcal{X}_e$ with equilibrium distribution $\tilde{\mu}$ such that, for $B \in \mathcal{B}(\mathcal{X}_e)$, $\tilde{\mu}(B \times \mathcal{X}_e) = \mu_1(B)$ and

$\tilde{\mu}(\mathcal{X}_e \times B) = \mu_2(B)$. Denote the transition kernel of $\tilde{\Phi}$ by $\tilde{K}$. Suppose the following condition holds for all $x, y \in \mathcal{X}_e, \xi \in \mathcal{X}$ such that $y \subseteq x$ but $y \cup \{\xi\} \not\subseteq x$:

$$\frac{f_1(x \cup \{\xi\})}{f_1(x)} \geq \frac{f_2(y \cup \{\xi\})}{f_2(y)}, \qquad \text{with convention } \frac{0}{0} = 0. \tag{2.3}$$

That is $b_1(x, \xi) \geq b_2(y, \xi)$ for all $y \subseteq x$. Then, for $y \subseteq x$, $\tilde{K}(x, y; T) = 1$ where $T = \{(x, y);\ y \subseteq x\}$. This means that if $\tilde{\Phi} = \left(\tilde{\Phi}_1, \tilde{\Phi}_2\right)$ and $\tilde{\Phi}_1(0) \supseteq \tilde{\Phi}_2(0)$ then $\tilde{\Phi}_1(t) \supseteq \tilde{\Phi}_2(t)$ for all $t \geq 0$. Furthermore the equilibrium distribution of $\tilde{\Phi}_i$ is μ_i, hence μ_2 is stochastically smaller than μ_1 (Preston 1977, Theorem 9.1)

Remark 2.2. The requirement that $f_1(x \cup \{\xi\}) f_2(y) \geq f_1(x) f_2(y \cup \{\xi\})$ is the partially-ordered-space-analogue of the requirement in Definition 2.2 for likelihood ratio ordering; see also Shaked & Shanthikumar (1994) for a description of the *multivariate likelihood ratio order*.

It is natural to want conditions which guarantee strict dominance between probability measures since only then can one guarantee the existence of a bias. So suppose that, in addition to Eq. (2.3) holding, the set $S' = \left\{(x, y);\ \frac{f_1(x \cup \{\xi\})}{f_1(x)} > \frac{f_2(y \cup \{\xi\})}{f_2(y)}\right\}$ has positive $\tilde{\mu}$-measure, where $\tilde{\mu}$ is the equilibrium distribution of the coupled process $\tilde{\Phi}$. So for $(x, y) \in S'$, we have that $b_1(x, \xi) > b_2(y, \xi)$; hence $\tilde{K}(x, y; T') > 0$ where $T' = \{(x, y);\ y \subset x\}$. Moreover $\tilde{\mu}(T') = \int_{S'} \tilde{K}(x, y; T')\, d\tilde{\mu}(x, y) > 0$; so for all strictly increasing measurable functions $g : \mathcal{X}_e \to \mathbb{R}$

$$\begin{aligned}\int g d\mu_1 &= \int_T g(x)\, d\tilde{\mu}(x, y) = \int_{\{(x,y);x=y\}} g(x)\, d\tilde{\mu}(x, y) + \int_{T'} g(x)\, d\tilde{\mu}(x, y) \\ &> \int_{\{(x,y);x=y\}} g(y)\, d\tilde{\mu}(x, y) + \int_{T'} g(y)\, d\tilde{\mu}(x, y) = \int_T g(y)\, d\tilde{\mu}(x, y) = \int g d\mu_2.\end{aligned}$$

Thus μ_2 is strictly stochastically smaller than μ_1, in the sense of Definition 2.3. This then proves the following Corollary to (Preston 1977, Theorem 9.1).

Corollary 2.1. *Let μ_i be a probability measure on $\mathcal{X}_e$ and f_i its density with respect to some fixed measure $\tilde{\omega}$, for $i = 1, 2$. Suppose that, for all $x, y \in \mathcal{X}_e, \xi \in \mathcal{X}$ such that $y \subseteq x$ but $y \cup \{\xi\} \not\subseteq x$, Eq. (2.3) holds. Suppose also that $\tilde{\mu}$ is a measure on $\mathcal{X}_e \times \mathcal{X}_e$ with marginals $\tilde{\mu}(B \times \mathcal{X}_e) = \mu_1(B)$ and $\tilde{\mu}(\mathcal{X}_e \times B) = \mu_2(B)$. If $\tilde{\mu}$ puts positive measure on the set $S' = \left\{(x, y);\ \frac{f_1(x \cup \{\xi\})}{f_1(x)} > \frac{f_2(y \cup \{\xi\})}{f_2(y)}\right\}$ then for all strictly increasing measurable functions g on $\mathcal{X}_e$ $\int g f_1 d\tilde{\omega} > \int g f_2 d\tilde{\omega}$, ie. $\mu_2 \prec_{st} \mu_1$.*

2.3 Conditional Point Processes

In Section 2.2.2 the notion of an exponential space was introduced and a point process is simply a random variable taking values in an exponential space (Definition 1.1). Thus a point process on some $\mathcal{X}$ is a finite unordered set of points $x = \{\xi_1, \ldots, \xi_n\} \in \mathcal{X}_e$, for $n \in \{0, 1, \ldots\}$ and $\xi_i \in \mathcal{X}$. The task now is to consider the simulation of point processes which are *conditioned* to belong to some set $\mathcal{E}$. Van Lieshout & Van Zwet (2001) proposed a perfect simulation algorithm for sampling a Poisson process conditioned to belong to an *anti-hereditary* conditioning set $\mathcal{E}$ (cf. Definition 2.4).

In what follows, the algorithm of Van Lieshout & Van Zwet, denoted here by VLVZ, is briefly summarized and three other simulation algorithms are also presented. The objective of these algorithms is to sample a conditional Poisson process. A comparison of the distribution of the output of the four procedures is carried out and it is shown that VLVZ generally produces biased samples. The nature of this bias is explored via the notion of stochastic dominance introduced in Section 2.2.2. Theorem 2.5 gives conditions on $\mathcal{E}$ when VLVZ produces *strictly stochastically smaller* samples than the target distribution. Finally, Example 2.2 illustrates a simple case when VLVZ actually outputs *strictly stochastically larger* samples.

2.3.1 Sampling Algorithms for Conditional Poisson Processes

VLVZ was proposed for any *anti-hereditary* conditioning set $\mathcal{E}$, defined as.

Definition 2.4. $\mathcal{E}$ is *anti-hereditary* if $x \in \mathcal{E} \Rightarrow x \cup \{\xi\} \in \mathcal{E}$, for any $\xi \notin x$.

Definition 2.5. $\mathcal{E}$ is *lattice* (or has the lattice property) if $x, y \ \in \mathcal{E} \Rightarrow x \cap y \in \mathcal{E}$ and $x \cup y \in \mathcal{E}$.

This lattice structure becomes crucial when employing Corollary 2.1 in order to establish strict stochastic dominance (cf. Theorem 2.4 & Lemma 2.6). Let θ be a random map $\theta : \mathbb{R}_e^2 \to \mathbb{R}_e^2$ which takes $x \in \mathbb{R}_e^2$ and returns some (random) sub-configuration of x. For $x \in \mathbb{R}_e^2$ refer to $\theta_x \equiv \theta(x)$ as a *thinning* of x since $\theta_x \subseteq x$. Henceforth θ will represent an *independent p-thinning* (cf. Section 1.1.9), so that θ_x is the configuration obtained by independently retaining each point of x with probability p. Let $n(x)$ be the number of points in x and $\pi_\lambda^{\mathcal{E}}$ the distribution of a Poisson(λ) which is conditioned to belong to some $\mathcal{E}$.

A number of algorithms are now presented that have the objective of sampling from $\pi_\lambda^{\mathcal{E}}$, for some $\lambda > 0$ and $\mathcal{E}$ a lattice anti-hereditary conditioning set. The basis of all the algorithms is the generation first of a *dominating* process of a higher intensity $\kappa > \lambda$, which is then thinned in some way to obtain a process of lower intensity. Processes of higher intensity are more likely to belong to $\mathcal{E}$ (see the discussion in Van Lieshout & Van Zwet 2001); hence let κ be large enough so that it is feasible to use Rejection sampling to draw from $\pi_\kappa^{\mathcal{E}}$ and set $p = \frac{\lambda}{\kappa}$. Define two dominating processes D and D' by: $D \sim \pi_\kappa^{\mathcal{E}}$ and D' is obtained via:

Set $i = 1$.

while $i \neq 0$:

draw $D \sim \pi_\kappa^{\mathcal{E}}$ and sample θ_D.

if $\theta_D \in \mathcal{E}$: set $i = 0$; **else**: set $i = i + 1$.

return $D' = D$.

Thus the distribution of D' is a weighted version of $\pi_\kappa^{\mathcal{E}}$ since we accept $D' = D \sim \pi_\kappa^{\mathcal{E}}$ with probability $\mathbb{P}\left[\theta_D \in \mathcal{E} \mid D\right]$, which is given by:

$$\mathbb{P}\left[\theta_D \in \mathcal{E} \mid D\right] = \sum_{x \subseteq d} \mathbb{P}\left[\theta_D = x \mid D = d\right] \mathbf{1}_{\{x \in \mathcal{E}\}} = \sum_{x \subseteq d} p^{n(x)} \left(1-p\right)^{n(d)-n(x)} \mathbf{1}_{\{x \in \mathcal{E}\}}$$

since θ is an independent p-thinning. Three algorithms, which are variants of the same basic idea (a two-step rejection sampling), and the VLVZ algorithm are now described below:

Algorithm 2.1 (Rejection/Thinning (RT)).

Set $i = 1$.

while $i \neq 0$:

draw $D \sim \pi_\kappa^{\mathcal{E}}$ and sample θ_D.

if $\theta_D \in \mathcal{E}$: set $i = 0$; **else**: set $i = i + 1$.

return $X^{RT} = \theta_D$.

Algorithm 2.2 (Iterated Rejection/Thinning (IRT)).

Set $i = 1$; draw $D \sim \pi_\kappa^{\mathcal{E}}$.

while $i \neq 0$:

sample θ_D.

if $\theta_D \in \mathcal{E}$: set $i = 0$; **else**: set $i = i + 1$.

return $X^{IRT} = \theta_D$.

Algorithm 2.3 (Modified Iterated Rejection/Thinning (Mod)).

Set $i = 1$; draw D'.

while $i \neq 0$:

sample $\theta_{D'}$.

if $\theta_{D'} \in \mathcal{E}$: set $i = 0$; **else**: set $i = i + 1$.

return $X^{Mod} = \theta_{D'}$.

Algorithm 2.4 (VLVZ).

Draw $D \sim \pi_\kappa^{\mathcal{E}}$.

Define a conditional spatial birth-and-death process $\tilde{\Phi}$ with the following dynamics:

the birth rate is $\frac{p}{1-p}$ for each point in $D \setminus \tilde{\Phi}$;

unit death rate per point in $\tilde{\Phi}$;

births are generated from $D \setminus \tilde{\Phi}$ and always accepted;

deaths are accepted only if $\mathcal{E}$ is not violated.

Apply CFTP to obtain an exact sample from the equilibrium distribution of $\tilde{\Phi}$.

return X^{VLVZ}, the output of the CFTP procedure.

Remark 2.3. Observe the analogy between the random variables N, N' (X, X') of Section 2.2.1 and the processes D, D' $\left(X^{IRT}, X^{Mod}\right)$. In Section 2.2.1 N' is obtained from N via an accept reject mechanism where the acceptance probability is $\mathbb{P}\left[\theta_N \in \mathcal{E} \mid N\right]$; here D' is similarly obtained from D with acceptance probability $\mathbb{P}\left[\theta_D \in \mathcal{E} \mid D\right]$. Moreover X, X' are conditional thinnings of N, N' and X^{IRT}, X^{Mod} are those of D, D' respectively.

The difference between IRT and modified IRT is in the generation of the dominating processes D and D' respectively. Except for trivial $\mathcal{E}$ implementation of modified IRT will be computationally burdensome; however the principle of the algorithm enables derivation of some distributional results. The difference between RT and IRT is that in RT a *new* dominating process is generated at every iteration, while in IRT θ is applied to *same* dominating process. Thus the output of IRT is a *conditional thinning* of the dominating process D. For VLVZ the equilibrium distribution of the

conditional process $\tilde{\Phi}$ depends on D and detailed balance calculations (cf. Eq. 1.25) show that the equilibrium (conditional) distribution of $\tilde{\Phi}$, given a realization of D, is that of θ_D but conditioned to belong to $\mathcal{E}$, ie. $\mathbb{P}\left[X^{VLVZ} = x \mid D\right] \propto \mathbb{P}\left[\theta_D = x \mid D\right] \mathbf{1}_{\{x \in \mathcal{E}\}}$.

Remark 2.4. Only the RT and modified IRT algorithms actually produce samples from the required target distribution $\pi_\lambda^{\mathcal{E}}$; the other two will generally not, as the following Lemmas show.

Lemma 2.2. *The modified IRT algorithm is equivalent to the RT algorithm, in that they output samples with the same distribution.*

Proof. This can be seen via conditional expectation; suppose g is a measurable function. For RT D is the dominating process, θ_D the process produced by applying θ to D and X^{RT} the output. Then

$$\mathbb{E}\left[g\left(X^{RT}\right)\right] = \frac{\mathbb{E}\left[g\left(\theta_D\right)\mathbf{1}_{\{\theta_D \in \mathcal{E}\}}\right]}{\mathbb{P}\left[\theta_D \in \mathcal{E}\right]} \tag{2.4}$$

where $\mathbb{P}\left[\theta_D \in \mathcal{E}\right] = \int_d \mathbb{P}\left[\theta_D \in \mathcal{E} \mid D = d\right] d\pi_\kappa^{\mathcal{E}}\left(d\right)$. If f and f' are the densities (with respect to a unit rate Poisson) of D and D' respectively, then $f' \propto f\, \mathbb{P}\left[\theta_D \in \mathcal{E} \mid D\right]$, as the distribution of D' is that of D but weighted by $\mathbb{P}\left[\theta_D \in \mathcal{E} \mid D\right]$. Thus $\mathbb{E}\left[\mathbb{E}\left[g\left(\theta_{D'}\right) \mid D'\right]\right] = \mathbb{E}\left[\mathbb{E}\left[g\left(\theta_D\right) \mid D\right] \times \frac{\mathbb{P}[\theta_D \in \mathcal{E} \mid D]}{\mathbb{P}[\theta_D \in \mathcal{E}]}\right]$.

Since the output of modified IRT, X^{Mod}, is a conditional thinning of D', obtained by iteratively sampling $\theta_{D'}$ until it satisfies $\mathcal{E}$:

$$\begin{aligned}
\mathbb{P}\left[X^{Mod} = x \mid D'\right] &= \frac{\mathbb{P}\left[\theta_{D'} = x \mid D'\right]}{\mathbb{P}\left[\theta_{D'} \in \mathcal{E} \mid D'\right]} \mathbf{1}_{\{x \in \mathcal{E}\}}; \\
\text{so that} \qquad \mathbb{E}\left[g\left(X^{Mod}\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[g\left(X^{Mod}\right) \mid D'\right]\right] = \mathbb{E}\left[\frac{\mathbb{E}\left[g\left(\theta_{D'}\right)\mathbf{1}_{\{\theta_{D'} \in \mathcal{E}\}} \mid D'\right]}{\mathbb{P}\left[\theta_{D'} \in \mathcal{E} \mid D'\right]}\right] \\
&= \mathbb{E}\left[\frac{\mathbb{E}\left[g\left(\theta_D\right)\mathbf{1}_{\{\theta_D \in \mathcal{E}\}} \mid D\right]}{\mathbb{P}\left[\theta_D \in \mathcal{E} \mid D\right]} \times \frac{\mathbb{P}\left[\theta_D \in \mathcal{E} \mid D\right]}{\mathbb{P}\left[\theta_D \in \mathcal{E}\right]}\right] \\
&= \frac{\mathbb{E}\left[\mathbb{E}\left[g\left(\theta_D\right)\mathbf{1}_{\{\theta_D \in \mathcal{E}\}} \mid D\right]\right]}{\mathbb{P}\left[\theta_D \in \mathcal{E}\right]} = \frac{\mathbb{E}\left[g\left(\theta_D\right)\mathbf{1}_{\{\theta_D \in \mathcal{E}\}}\right]}{\mathbb{P}\left[\theta_D \in \mathcal{E}\right]}.
\end{aligned} \tag{2.5}$$

The third equality follows as D' is a weighted realization of D. Now compare Eqs. $(2.4\ \&\ 2.5)$. □

Lemma 2.3. *The IRT and RT do not, in general, produce samples from the same distribution.*

Proof. This can also be seen via conditional expectation. If X^{IRT} is the output of IRT, then the distribution of X^{IRT} given D is that of θ_D but weighted so that

$$\begin{aligned}
\mathbb{P}\left[X^{IRT} = x \mid D\right] &= \frac{\mathbb{P}\left[\theta_D = x \mid D\right]}{\mathbb{P}\left[\theta_D \in \mathcal{E} \mid D\right]} \mathbf{1}_{\{x \in \mathcal{E}\}} \\
\text{thus} \qquad \mathbb{E}\left[g\left(X^{IRT}\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[g\left(X^{IRT}\right) \mid D\right]\right] = \mathbb{E}\left[\frac{\mathbb{E}\left[g\left(\theta_D\right)\mathbf{1}_{\{\theta_D \in \mathcal{E}\}} \mid D\right]}{\mathbb{P}\left[\theta_D \in \mathcal{E} \mid D\right]}\right].
\end{aligned} \tag{2.6}$$

Comparing Eqs. $(2.4 \,\&\, 2.6)$ one can see that the difference lies in the outermost expectation: in IRT one takes the outer expectation of the ratio while in RT it is the ratio of the outer expectations. The two expressions will only be equal if $\mathbb{P}\left[\theta_D \in \mathcal{E} \mid D\right]$ does *not* depend on D; but $D \in \mathcal{E}$ by construction, so that $\mathbb{P}\left[\theta_D \in \mathcal{E} \mid D\right]$ *will* generally depend on D (unless $p = 1$ in which case it would be a very trivial thinning!). □

Lemma 2.4. *The VLVZ algorithm is equivalent to the IRT algorithm.*

Proof. Let X^{VLVZ} denote the output from VLVZ and X^{IRT} that from IRT. The dominating process for either is D, and θ_D is obtained by applying θ to D. Detailed balance calculations (Van Lieshout & Van Zwet 2001) show that the distribution of X^{VLVZ}, given D, is simply that of θ_D conditioned to belong to $\mathcal{E}$. The distribution of X^{IRT}, given D, is that of θ_D but conditioned to belong to $\mathcal{E}$. Since $X^{VLVZ} \mid D$ and $X^{IRT} \mid D$ agree in distribution, so will X^{VLVZ} and X^{IRT}. □

Thus VLVZ produces biased samples as X^{VLVZ} and X^{RT} do have the same distribution.

2.3.2 Domination Results for the VLVZ Algorithm

The previous section showed that the VLVZ algorithm generally outputs biased samples; however the nature of the bias is not clear. In this section the ideas of stochastic dominance and conditions yielding such ordering, as presented in Section 2.2.2, are employed in order to explore this bias. Corollary 2.1 enables one to conclude that if the conditioning event has the lattice property (Definition 2.5) then VLVZ produces strictly stochastically smaller samples than RT.

Recall the procedures RT, IRT, modified IRT & VLVZ (Algorithms 2.1, 2.2, 2.3 & 2.4 respectively). Their objective is to sample from $\pi_\lambda^{\mathcal{E}}$, the distribution of a Poisson(λ) conditioned to belong to some anti-hereditary $\mathcal{E}$. The basis of RT, IRT & modified IRT is the generation of a dominating process of higher intensity which is then conditionally thinned in order to produce a process of lower intensity. In a totally ordered space it was shown that the analogous conditional thinning procedures to IRT and modified IRT produced random variables X and X' respectively such that $X \prec_{st} X'$ *for all* $\mathcal{E}$ (Theorem 2.3). The question whether this also holds in a partially ordered space is now answered here. The dominating process of RT & IRT is $D \sim \pi_\kappa^{\mathcal{E}}$, for some $\kappa > \lambda$. That of modified IRT, D', is obtained via an accept/reject mechanism from D with the acceptance probability being $\mathbb{P}\left[\theta_D \in \mathcal{E} \mid D\right]$ (where θ represents an independent p-thinning with $p = \frac{\lambda}{\kappa}$).

Lemma 2.5. *Let $\mathcal{E}$ be some anti-hereditary conditioning event such that $\emptyset \notin \mathcal{E}$, where $\emptyset$ is the configuration with no points. If θ_D represents an independent p-thinning of D then $\mathbb{P}[\theta_D \in \mathcal{E} \mid D]$ is strictly increasing in D, in the sense $\mathbb{P}[\theta_D \in \mathcal{E} \mid D = d \cup \{\xi\}] > \mathbb{P}[\theta_D \in \mathcal{E} \mid D = d]$, for $\xi \notin d$.*

Proof. $\mathbb{P}[\theta_D \in \mathcal{E} \mid D = d] = \sum_{x \subseteq d} \mathbb{P}[\theta_d = x] \mathbf{1}_{\{x \in \mathcal{E}\}}$; hence

$$\begin{aligned}
\mathbb{P}[\theta_D \in \mathcal{E} \mid D = d \cup \{\xi\}] &= \sum_{y \subseteq d \cup \{\xi\}} \mathbb{P}\left[\theta_{d\cup\{\xi\}} = y\right] \mathbf{1}_{\{y \in \mathcal{E}\}} \\
&= \sum_{x \subseteq d} \mathbb{P}\left[\theta_{d\cup\{\xi\}} = x\right] \mathbf{1}_{\{x \in \mathcal{E}\}} + \sum_{x \subseteq d} \mathbb{P}\left[\theta_{d\cup\{\xi\}} = x \cup \{\xi\}\right] \mathbf{1}_{\{x\cup\{\xi\} \in \mathcal{E}\}} \\
&= (1-p) \sum_{x \subseteq d} \mathbb{P}[\theta_d = x] \mathbf{1}_{\{x \in \mathcal{E}\}} + p \sum_{x \subseteq d} \mathbb{P}[\theta_d = x] \mathbf{1}_{\{x\cup\{\xi\} \in \mathcal{E}\}} \\
&= (1-p)\, \mathbb{P}[\theta_D \in \mathcal{E} \mid D = d] + p \sum_{x \subseteq d} \mathbb{P}[\theta_d = x] \mathbf{1}_{\{x\cup\{\xi\} \in \mathcal{E}\}} \\
&> \mathbb{P}[\theta_D \in \mathcal{E} \mid D = d]\,.
\end{aligned}$$

Anti-hereditary $\mathcal{E}$ implies $\mathbf{1}_{\{x\cup\{\xi\} \in \mathcal{E}\}} \geq \mathbf{1}_{\{x \in \mathcal{E}\}}$ for all $x \subseteq d$; moreover since $\emptyset \notin \mathcal{E}$ implies that there must exist $x \subset d$ such that $x \notin \mathcal{E}$ but $x \cup \{\xi\} \in \mathcal{E}$. Hence the inequality follows. □

Theorem 2.4. *If D and D' are the dominating processes in IRT (Algorithm 2.2) and modified IRT (Algorithm 2.3) respectively and $\mathcal{E}$ is lattice anti-hereditary then $D \prec_{st} D'$.*

Proof. Let f, f' denote the densities (with respect to a unit rate Poisson) of D and D' respectively. Since D is a conditioned Poisson process of intensity κ, $f(d) \propto \kappa^{n(d)} \mathbf{1}_{\{d \in \mathcal{E}\}}$; and $f'(d) \propto \kappa^{n(d)}\, \mathbb{P}[\theta_d \in \mathcal{E}]\, \mathbf{1}_{\{d \in \mathcal{E}\}}$ from the definition of D'. Let $y \subseteq x$ and ξ be such that $y \cup \{\xi\} \not\subseteq x$; consider

$$\begin{aligned}
f'(x \cup \{\xi\})\, f(y) &= \alpha\, \alpha'\, \kappa^{n(x\cup\{\xi\})} \kappa^{n(y)}\, \mathbb{P}\left[\theta_{x\cup\{\xi\}} \in \mathcal{E}\right] \mathbf{1}_{\{x\cup\{\xi\} \in \mathcal{E}\}} \mathbf{1}_{\{y \in \mathcal{E}\}} \\
f'(x)\, f(y \cup \{\xi\}) &= \alpha\, \alpha'\, \kappa^{n(x)} \kappa^{n(y\cup\{\xi\})}\, \mathbb{P}[\theta_x \in \mathcal{E}]\, \mathbf{1}_{\{x \in \mathcal{E}\}} \mathbf{1}_{\{y\cup\{\xi\} \in \mathcal{E}\}}
\end{aligned}$$

where α, α' are the respective normalizing constants for f and f'. If $y \notin \mathcal{E}$, then either $x \in \mathcal{E}$ and $y \cup \{\xi\} \notin \mathcal{E}$ or $x \notin \mathcal{E}$, since otherwise we would have $y = x \cap (y \cup \{\xi\}) \in \mathcal{E}$, as $\mathcal{E}$ is lattice. Conversely if both $x \in \mathcal{E}$ and $y \cup \{\xi\} \in \mathcal{E}$ then both $x \cup \{\xi\}$ and y must also belong to $\mathcal{E}$. Thus $f(y) = 0$ (ie. if $y \notin \mathcal{E}$) implies that either $f'(x)$ or $f(y \cup \{\xi\})$ must also be zero; and $f'(x)\, f(y \cup \{\xi\}) \neq 0$ implies that $f'(x \cup \{\xi\})\, f(y)$ is also non-zero. Finally, Lemma 2.5 yields $\mathbb{P}\left[\theta_{x\cup\{\xi\}} \in \mathcal{E}\right] > \mathbb{P}[\theta_x \in \mathcal{E}]$.

Thus Eq. (2.3) of Section 2.2.3 is satisfied. Denote the distributions corresponding to f, f' by μ, μ' respectively, and let $\tilde{\mu}$ be the measure with respective marginals μ' and μ. The construction of Section 2.2.3 shows that $\tilde{\mu}(T) = 1$, where $T = \{(x, y)\,;\, y \subseteq x\}$. Now let $S' = \left\{(x, y)\,;\, \frac{f'(x\cup\{\xi\})}{f'(x)} > \frac{f(y\cup\{\xi\})}{f(y)}\right\}$; it needs to be shown that $\tilde{\mu}(S') > 0$. Notice that if, for some ξ, $y \subseteq x$ and $x, y \in \mathcal{E}$ then either $y \cup \{\xi\} \in \mathcal{E}$ ($\Rightarrow x \cup \{\xi\} = x \cup (y \cup \{\xi\}) \in \mathcal{E}$), or $x \cup \{\xi\} \in \mathcal{E}$.

Following the coupling construction in Section 2.2.3, let $\tilde{\Phi}$ be a spatial birth-death process with equilibrium $\tilde{\mu}$. If $y \subseteq x$ then the only ξ that will have non-zero birth-rates in $\tilde{\Phi}$ are those for which either $x \cup \{\xi\} \in \mathcal{E}$ or $y \cup \{\xi\} \in \mathcal{E}$. Hence any (x, y) with $y \subseteq x$ is contained in S', so the set S' has positive $\tilde{\mu}$-measure. The conditions of Corollary 2.1 are satisfied and the result follows. □

Denote by X^{IRT} the output of IRT and X^{Mod} that of modified IRT. X^{IRT} is obtained as a conditional thinning of D and X^{Mod} as that of D'. Section 2.3.1 established that $\mathcal{L}\left(X^{IRT}\right) \neq \mathcal{L}\left(X^{Mod}\right)$, where $\mathcal{L}(X)$ denotes the law of X. Theorem 2.5 now shows that, for lattice $\mathcal{E}$, $X^{IRT} \prec_{st} X^{Mod}$.

Lemma 2.6. *Let $\mathcal{E}$ be lattice anti-hereditary and θ an independent p-thinning. For a configuration d let X_d denotes its conditional thinning, ie. the random configuration obtained by iteratively generating θ_d until $\theta_d \in \mathcal{E}$, and P_d its respective distribution. Then, for $d \subset d'$, $X_d \prec_{st} X_{d'}$.*

Proof. $\mathbb{P}[X_d = x] = p^{n(x)}(1-p)^{n(d)-n(x)}\mathbf{1}_{\{x\subseteq d\}}\mathbf{1}_{\{x\in\mathcal{E}\}}$. Thus for $y \subseteq x$ such that $y \cup \{\xi\} \nsubseteq x$:

$$\mathbb{P}[X_{d'} = x\cup\{\xi\}]\,\mathbb{P}[X_d = y] = p^{n_1}(1-p)^{n_2}\mathbf{1}_{\{x\cup\{\xi\}\subseteq d';\, x\cup\{\xi\}\in\mathcal{E}\}}\mathbf{1}_{\{y\subseteq d;\, y\in\mathcal{E}\}}; \quad (2.7)$$

$$\mathbb{P}[X_{d'} = x]\,\mathbb{P}[X_d = y\cup\{\xi\}] = p^{n_1}(1-p)^{n_2}\ \mathbf{1}_{\{x\subseteq d';\, x\in\mathcal{E}\}}\mathbf{1}_{\{y\cup\{\xi\}\subseteq d;\, y\cup\{\xi\}\in\mathcal{E}\}}; \quad (2.8)$$

$$\text{where } n_1 = n(x\cup\{\xi\}) + n(y) = n(x) + n(y\cup\{\xi\});$$

$$\text{and } n_2 = n(d') - n(x\cup\{\xi\}) + n(d) - n(y) = n(d') - n(x) + n(d) - n(y\cup\{\xi\}).$$

Since $\mathcal{E}$ is lattice, $y \notin \mathcal{E}$ means that either $x \notin \mathcal{E}$ or $y \cup \{\xi\} \notin \mathcal{E}$. If $x \cup \{\xi\} \nsubseteq d'$ then either $x \subseteq d'$ and $\xi \notin d' \supseteq d$ (which implies that $y \cup \{\xi\} \nsubseteq d$) or $x \nsubseteq d'$. Finally, $y \nsubseteq d$ implies $y \cup \{\xi\} \nsubseteq d$; therefore $(2.7) = 0 \Rightarrow (2.8) = 0$. Conversely, if $y \cup \{\xi\} \nsubseteq d$ then it does not necessarily follow that either $y \nsubseteq d$ or $x \cup \{\xi\} \nsubseteq d'$; so $(2.8) = 0 \not\Rightarrow (2.7) = 0$. This yields $(2.7) \geq (2.8)$.

X_d is defined only on those subsets of d which belong to $\mathcal{E}$; if $\tilde{\mu}$ (cf. Section 2.2.3) denotes the measure with marginals $P_{d'}$ and P_d, then its support $\{(x, y)\,;\, y \subseteq d, x \subseteq d', y \subseteq x \text{ and } x, y \in \mathcal{E}\}$ is finite. Now, for any $\xi \notin x$ such that $\xi \in d' \setminus d$, (2.7) is positive but (2.8) is always zero. Thus the set

$S' = \{(x,y)\,;\, \mathbb{P}[X_{d'} = x \cup \{\xi\}]\,\mathbb{P}[X_d = y] > \mathbb{P}[X_{d'} = x]\,\mathbb{P}[X_d = y \cup \{\xi\}]\}$ must have positive measure, since $\tilde{\mu}$ has finite support and $d \subset d'$. Corollary 2.1 now completes the proof. □

Theorem 2.5. *Let U and U' be any processes which are conditioned to belong to some lattice antihereditary $\mathcal{E}$ and θ be an independent p-thinning. Denote the conditional thinning of U by X and that of U' by X'. If $U \prec_{st} U'$ then $X \prec_{st} X'$.*

Proof. For a configuration u let $X_u \sim P_u$ be its conditional thinning; from Lemma 2.6 $u \subset u'$ implies $P_u \prec_{st} P_{u'}$. For g a strictly increasing function $\mathbb{E}[g(X_u)] < \mathbb{E}[g(X_{u'})]$, hence $\mathbb{E}[g(X_u)]$ is also strictly increasing. Thus, for $U \prec_{st} U'$, $\mathbb{E}_U\mathbb{E}[g(X_U)] < \mathbb{E}_{U'}\mathbb{E}[g(X_{U'})]$ (Definition 2.3), and

$$\begin{aligned}\mathbb{E}[g(X)] = \mathbb{E}_U\mathbb{E}[g(X) \mid U] &= \mathbb{E}_U\mathbb{E}[g(X_U)] \\ &< \mathbb{E}_{U'}\mathbb{E}[g(X_{U'})] = \mathbb{E}_{U'}\mathbb{E}[g(X') \mid U'] = \mathbb{E}[g(X')].\end{aligned}$$

That $X \prec_{st} X'$ now follows from Definition 2.3. □

So for a lattice $\mathcal{E}$, Theorem 2.4 yields $D \prec_{st} D'$ and Theorem 2.5 yields $X^{IRT} \prec_{st} X^{Mod}$.

Coverage of a Single Point

Suppose that, for $c \in \mathbb{R}^2$, $B_r(c)$ denotes a disk with centre c and fixed radius r. If $\mathcal{E}$ is the set of all point configurations on $B_r(c)$ whose Boolean model covers $\{c\}$ then any $x \in \mathcal{E}$ has at least one point in $B_r(c)$ and vice versa. So if X is a homogenous Poisson process on $B_r(c)$, then conditioning it to belong to $\mathcal{E}$ is equivalent to conditioning X to have a positive number of points.

Recall that D, D' are the dominating processes, and X^{VLVZ}, X^{Mod} the output, of the VLVZ and modified IRT algorithms respectively; and let $D(B_r(c))$ denote the number of points of D contained in $B_r(c)$. The results of Section 2.2.1 enable one to conclude that $D(B_r(c)) \prec_{st} D'(B_r(c))$ and $X^{IRT}(B_r(c)) \prec_{st} X^{Mod(B_r(c))}$. This is because D is a Poisson(κ) process on $B_r(c)$ conditioned to belong to $\mathcal{E}$, and so $D(B_r(c))$ is a Poisson random variable conditioned to be positive. D' is obtained from D via an accept/reject mechanism where the acceptance probability is $\mathbb{P}[\theta_D \in \mathcal{E} \mid D] \equiv \mathbb{P}[\theta_D(B_r(c)) > 0 \mid D]$. The map θ is an independent p-thinning; therefore the distribution of $\theta_D(B_r(c))$, given D, is a Binomial$(D(B_r(c)), p)$ random variable. The acceptance weight $\mathbb{P}[\theta_D(B_r(c)) > 0 \mid D]$ thus becomes $\mathbb{P}[\text{Binomial}(D(B_r(c)), p) > 0 \mid D]$, which is

strictly increasing in $D(B_r(c))$. The outputs $X^{VLVZ}(B_r(c))$ and $X^{Mod}(B_r(c))$ are conditional thinnings of $D(B_r(c))$ and $D'(B_r(c))$ respectively. Moreover, since $\theta_D(B_r(c))$ is a Binomial random variable, the conditional thinning procedure respects the ordering between $D(B_r(c))$ and $D'(B_r(c))$. Theorems 2.1 & 2.3 hence show that $D(B_r(c)) \prec_{st} D'(B_r(c))$ and $X^{IRT}(B_r(c)) \prec_{st} X^{Mod}(B_r(c))$.

Since the distributions of the processes D, D', X^{VLVZ} and X^{Mod} depend only on the number of points and not their locations within $B_r(c)$, samples can be generated by first sampling the number of points and then distributing them uniformly on $B_r(c)$. Therefore $D(B_r(c)) \prec_{st} D'(B_r(c))$ implies that there exist $\left(\widehat{D}, \widehat{D}'\right)$ such that: $\mathcal{L}\left(\widehat{D}\right) = \mathcal{L}(D)$, $\mathcal{L}\left(\widehat{D}'\right) = \mathcal{L}(D')$ and $\widehat{D} \subseteq \widehat{D}'$ almost surely. The inclusion is strict with positive probability; hence $D \prec_{st} D'$. A similar argument establishes $X^{VLVZ} \prec_{st} X^{Mod}$.

Coverage of Two Points

Suppose now $\mathcal{E}$ is the set of all configurations whose Boolean models cover two points $\{c_1, c_2\}$, with $B_r(c_1) \cap B_r(c_2) \neq \emptyset$; then $\mathcal{E}$ is not lattice. Examples 2.1 & 2.2 show, via explicit calculations, that in certain cases the ordering between D & D' and X^{IRT} & X^{Mod} is *reversed*, in the sense $D(W) \succ_{st} D'(W)$ and $X^{IRT}(W) \succ_{st} X^{Mod}(W)$, for some region W. Furthermore Example 2.3 also illustrates that in some cases neither X^{IRT} nor X^{Mod} dominates the other.

Let $A = B_r(c_1) \setminus B_r(c_2), B = B_r(c_1) \cap B_r(c_2), C = B_r(c_2) \setminus B_r(c_1)$; so a point in A (C) covers only $\{c_1\}$ $(\{c_2\})$ while a point in B covers $\{c_1, c_2\}$. For m_2 denoting Lebesgue measure, let $\beta = \frac{m_2[B]}{m_2[A \cup B \cup C]} > 0$ denote the 'overlap proportion' and $m_2[A] = m_2[C] = \frac{(1-\beta)}{2} m_2[A \cup B \cup C]$. Suppose N is a Poisson(κ) and $N(W)$ the number of points falling in some bounded region W. Let

$$p_n = \mathbb{P}[N(A \cup B \cup C) = n] = \frac{(\kappa m_2[A \cup B \cup C])^n}{n!} e^{-\kappa m_2[A \cup B \cup C]};$$

$$\begin{aligned} p_N(n_A, n_B, n_C) &= \mathbb{P}[N(A) = n_A, N(B) = n_B, N(C) = n_C] \equiv \mathbb{P}[N(A, B, C) = (n_A, n_B, n_C)] \\ &= p_n \times \frac{n!}{n_A! n_B! n_C!} \beta^{n_B} \left(\frac{1-\beta}{2}\right)^{n_A + n_C} \qquad \text{where } n = n_A + n_B + n_C. \end{aligned} \tag{2.9}$$

D, the dominating process of IRT is Poisson process of intensity κ conditioned to belong to $\mathcal{E}$; so

$$p_D(n_A, n_B, n_C) = \frac{p_N(n_A, n_B, n_C)}{\mathbb{P}[N \in \mathcal{E}]} \mathbf{1}_{\{n_A+n_B\geq 1,\ n_B+n_C\geq 1\}}; \tag{2.10}$$

$$\begin{aligned}\text{where } \mathbb{P}[N \in \mathcal{E}] &= \sum_{n\geq 1} \mathbb{P}[N(A\cup B\cup C) = n, N(A)+N(B)\geq 1, N(B)+N(C)\geq 1] \\ &= \sum_{n\geq 1} \sum_{\substack{n_A,n_B,n_C;\\ n_A+n_B+n_C=n}} p_N(n_A, n_B, n_C)\, \mathbf{1}_{\{n_A+n_B\geq 1,\ n_B+n_C\geq 1\}}. \end{aligned} \tag{2.11}$$

For D' the dominating process of modified IRT:

$$\begin{aligned} p_{D'}(n_A, n_B, n_C) &= \mathbb{P}[D(A,B,C) = (n_A, n_B, n_C) \mid \text{ accept } D] \\ &= \frac{p_D(n_A, n_B, n_C)\, \mathbb{P}[\theta_D \in \mathcal{E} \mid D]}{\mathbb{P}[\theta_D \in \mathcal{E}]} \qquad (2.12) \\ &= \frac{p_N(n_A, n_B, n_C)\, \mathbb{P}[\theta_N \in \mathcal{E} \mid N]}{\mathbb{P}[\theta_N \in \mathcal{E}]} \mathbf{1}_{\{N\in\mathcal{E}\}}; \qquad (2.13) \end{aligned}$$

$$\text{where } \mathbb{P}[\theta_N \in \mathcal{E}] = \sum_{n_A,n_B,n_C} \mathbb{P}[\theta_N \in \mathcal{E} \mid N]\, p_N(n_A, n_B, n_C)$$

$$\begin{aligned}\text{and } \mathbb{P}[\theta_N \in \mathcal{E} \mid N] &\equiv \mathbb{P}[\theta_N \in \mathcal{E} \mid N(A,B,C) = (n_A, n_B, n_C)] \\ &= \sum_{i_A\leq n_A} \sum_{i_B\leq n_B} \sum_{i_C\leq n_C} \mathbb{P}[\theta_N(A,B,C) = (i_A, i_B, i_C) \mid N]\, \mathbf{1}_{\{i_A+i_B\geq 1,\ i_B+i_C\geq 1\}} \\ &= \sum_{i_A\leq n_A} \sum_{i_B\leq n_B} \sum_{i_C\leq n_C} \binom{n_A}{i_A}\binom{n_B}{i_B}\binom{n_C}{i_C} p^i (1-p)^{n-i}\, \mathbf{1}_{\{i_A+i_B\geq 1,\ i_B+i_C\geq 1\}}. \end{aligned} \tag{2.14}$$

Here $i = i_A + i_B + i_C$ and $n = n_A + n_B + n_C$ for each term in the summand. We now consider the respective probabilities for X^{IRT} and X^{Mod}:

$$\begin{aligned}\mathbb{P}\left[X^{IRT}(A,B,C) = (i_A, i_B, i_C) \mid D\right] &= \mathbb{P}[\theta_D(A,B,C) = (i_A, i_B, i_C) \mid D, \theta_D \in \mathcal{E}] \\ &= \frac{\mathbb{P}[\theta_D(A,B,C) = (i_A, i_B, i_C) \mid D]}{\mathbb{P}[\theta_D \in \mathcal{E} \mid D]} \mathbf{1}_{\{i_A+i_B\geq 1, i_B+i_C\geq 1\}} \\ &= \binom{n_A}{i_A}\binom{n_B}{i_B}\binom{n_C}{i_C} \frac{p^i(1-p)^{n-i}}{\mathbb{P}[\theta_D \in \mathcal{E} \mid D]} \mathbf{1}_{\{i_A+i_B\geq 1, i_B+i_C\geq 1\}} \end{aligned} \tag{2.15}$$

where we implicitly have $D(A,B,C) = (n_A, n_B, n_C)$, $n = n_A + n_B + n_C$ and $i = i_A + i_B + i_C$; a similar expression holds for X^{Mod} but with D' instead of D. Thus the marginal for X^{IRT} is

$$\begin{aligned}&\mathbb{P}\left[X^{IRT}(A,B,C) = (i_A, i_B, i_C)\right] \\ &= \sum_{\substack{n_A\geq i_A\\ n_B\geq i_B\\ n_C\geq i_C}} \mathbb{P}\left[X^{IRT}(A,B,C) = (i_A, i_B, i_C) \mid D\right] p_D(n_A, n_B, n_C)\, \mathbf{1}_{\{i_A+i_B\geq 1, i_B+i_C\geq 1\}}. \end{aligned} \tag{2.16}$$

The marginal of X^{Mod} is given by a similar expression but with D' replacing D. For a lattice $\mathcal{E}$, D is always strictly stochastically smaller than D' (Theorem 2.4), which in turn implies that X^{IRT} is also always strictly stochastically smaller than X^{Mod} (Theorem 2.5). However the coverage of two points is not a lattice event and the following examples show that when the dominating intensity κ is low then the above ordering can either be reversed or not hold at all.

Example 2.1. For the overlap proportion $\beta \in \left(0, \sqrt{2}-1\right)$ and the dominating Poisson intensity κ small enough so that $p_n = \mathbb{P}\left[N\left(A \cup B \cup C\right) = n\right] = o\left(\kappa^{n-1}\right)$ for $n \geq 3$, ie. for low enough intensity κ so that the probability that N has three or more points in $A \cup B \cup C$ is negligible, $D\left(A \cup B \cup C\right) \succ_{st} D'\left(A \cup B \cup C\right)$. Furthermore, for $\beta \in \left(\sqrt{2}-1, 1\right)$ and κ as above $D\left(A \cup B \cup C\right) \prec_{st} D'\left(A \cup B \cup C\right)$.

This is seen as follows. The requirement on κ yields Eq. (2.11) as

$$\begin{aligned}
\mathbb{P}\left[N \in \mathcal{E}\right] &= \sum_{n=1}^{2} \mathbb{P}\left[N\left(A \cup B \cup C\right) = n, N\left(A\right) + N\left(B\right) \geq 1, N\left(B\right) + N\left(C\right) \geq 1\right] + o\left(\kappa^2\right) \\
&= p_N\left(0,1,0\right) + p_N\left(1,0,1\right) + p_N\left(1,1,0\right) + p_N\left(0,1,1\right) + p_N\left(0,2,0\right) + o\left(\kappa^2\right) \\
&= p_1\beta + p_2\,\frac{\left(1-\beta\right)^2}{2} + 2p_2\beta\left(1-\beta\right) + p_2\beta^2 + o\left(\kappa^2\right) = K \qquad (2.17)
\end{aligned}$$

since for $n = n_A + n_B + n_C = 1$ the only triple (n_A, n_B, n_C) which satisfies $n_A + n_B \geq 1$ and $n_B + n_C \geq 1$ is $(0,1,0)$; while for $n = 2$ there are four which belong to the conditioning set: $(1,0,1)$, $(1,1,0)$, $(0,1,1)$ and $(0,2,0)$. Using Eq. (2.10), the corresponding probabilities for D are

$$\begin{aligned}
&p_D\left(0,1,0\right) = \tfrac{p_1\beta}{K}; \qquad p_D\left(1,0,1\right) = \tfrac{p_2\,\frac{(1-\beta)^2}{2}}{K}; \quad p_D\left(1,1,0\right) = \tfrac{p_2\beta(1-\beta)}{K}; \\
&p_D\left(0,1,1\right) = \tfrac{p_2\beta(1-\beta)}{K}; \quad p_D\left(0,2,0\right) = \tfrac{p_2\beta^2}{K}; \qquad p_D\left(n_A,n_B,n_C\right) = o\left(\kappa^2\right),\ n_A+n_B+n_C \geq 3.
\end{aligned}$$

Similarly, for the D' we have

$$\begin{aligned}
\mathbb{P}\left[\theta_N \in \mathcal{E}\right] &= \sum_{\substack{n_A,n_B,n_C; \\ n_A+n_B+n_C \geq 1}} \mathbb{P}\left[\theta_N \in \mathcal{E} \mid N\right] p_N\left(n_A,n_B,n_C\right) \\
&= \sum_{\substack{n_A,n_B,n_C; \\ 1 \leq n_A+n_B+n_C \leq 2}} \mathbb{P}\left[\theta_N \in \mathcal{E} \mid N\right] p_N\left(n_A,n_B,n_C\right) + o\left(\kappa^2\right) \\
&= p_1\beta p + p_2\,\frac{\left(1-\beta\right)^2}{2}p^2 + 2p_2\beta\left(1-\beta\right)p + p_2\beta^2 p\left(2-p\right) + o\left(\kappa^2\right) = K'. \qquad (2.18)
\end{aligned}$$

Eqs. (2.14 & 2.9) are used to get the last line; Eq. (2.13) then yields:

$$\begin{aligned}
&p_{D'}\left(0,1,0\right) = \tfrac{p_1\beta p}{K'}; \qquad p_{D'}\left(1,0,1\right) = \tfrac{p_2\,\frac{(1-\beta)^2}{2}p^2}{K'}; \quad p_{D'}\left(1,1,0\right) = \tfrac{p_2\beta(1-\beta)p}{K'}; \\
&p_{D'}\left(0,1,1\right) = \tfrac{p_2\beta(1-\beta)p}{K'}; \quad p_{D'}\left(0,2,0\right) = \tfrac{p_2\beta^2 p(2-p)}{K'}; \quad p_{D'}\left(n_A,n_B,n_C\right) = o\left(\kappa^2\right),\ n_A+n_B+n_C \geq 3.
\end{aligned}$$

Now consider $\mathbb{P}[D(A \cup B \cup C) = 1] - \mathbb{P}[D'(A \cup B \cup C) = 1] = \frac{p_1 \beta}{KK'}(K' - pK) + o(\kappa^2)$, where

$$\begin{aligned} K' - pK &= p_2 \frac{(1-\beta)^2}{2}\left(p^2 - p\right) + p_2\beta^2\left(p(2-p) - p\right) + o\left(\kappa^2\right) \\ &= p_2 p(1-p)\left\{\beta^2 - \frac{(1-\beta)^2}{2}\right\} = p_2 p(1-p)\left\{\frac{\beta^2 + 2\beta - 1}{2}\right\} + o\left(\kappa^2\right). \end{aligned}$$

The roots of $\beta^2 + 2\beta - 1 = 0$ are $-1 \pm \sqrt{2}$; so $\beta \in \left(0, \sqrt{2} - 1\right)$ implies $\mathbb{P}[D(A \cup B \cup C) = 1] < \mathbb{P}[D'(A \cup B \cup C) = 1]$, since the contribution from $o(\kappa^2)$ is negligible; but this means that

$$\mathbb{P}[D(A \cup B \cup C) \geq 2] > \mathbb{P}[D'(A \cup B \cup C) \geq 2].$$

Now $\mathbb{P}[D(A \cup B \cup C) \geq 1] = \mathbb{P}[D'(A \cup B \cup C) \geq 1] = 1$; for $n \geq 3$, $\mathbb{P}[D(A \cup B \cup C) \geq n] = \mathbb{P}[D'(A \cup B \cup C) \geq n] = o(\kappa^2)$. Thus for $\beta \in \left(0, \sqrt{2} - 1\right)$ and κ small enough we have $D(A \cup B \cup C) \succ_{st} D'(A \cup B \cup C)$. Conversely, $\beta \in \left(\sqrt{2} - 1, 1\right) \Rightarrow \beta^2 + 2\beta - 1 > 0 \Rightarrow \mathbb{P}[D(A \cup B \cup C) \geq 2] < \mathbb{P}[D'(A \cup B \cup C) \geq 2]$; hence $D(A \cup B \cup C) \prec_{st} D'(A \cup B \cup C)$.

Intuitively, a small value of the overlap proportion β means that there is a low probability of getting a configuration with a point in B; thus for the 2-point configurations, $(1, 0, 1)$ is more likely than the others while $(0, 1, 0)$ is the only 1-point configuration. This then reduces to choosing between these two configurations when generating D' from D. Since since the acceptance probability for $(1, 0, 1)$ is p^2 and that for $(0, 1, 0)$ is p, the 1-point configuration is more likely to be accepted than the 2-point one. The stochastic ordering between D and D' is hence reversed.

Remark 2.5. The case when $\beta = 0$ means that no single point can simultaneously cover both conditioning points and the issue reduces to requiring that $D(A) \geq 1$ and $D(C) \geq 1$. The problem then simplifies to the case of coverage of a single point and the results of the previous section yield $D(A) \prec_{st} D'(A)$ and $D(C) \prec_{st} D'(C)$. Since A and C are disjoint (ie. $\beta = 0$) we have $D(A \cup C) \prec_{st} D'(A \cup C)$.

Example 2.2. For the overlap proportion $\beta \in \left(0, \sqrt{2} - 1\right)$ and the dominating intensity κ small enough so that $p_n = \mathbb{P}[N(A \cup B \cup C) = n] = o(\kappa^{n-1})$, for $n \geq 3$, then $X^{IRT}(A \cup B \cup C) \succ_{st} X^{Mod}(A \cup B \cup C)$.

This can be seen in a similar manner to Example 2.1. The thinned processes will have negligible

probability of having three or more points given the restriction on κ; thus Eqs. $(2.16\ \&\ 2.15)$ yield

$$\begin{aligned}
\mathbb{P}\left[X^{IRT}(A,B,C)=(0,1,0)\right] &= \frac{p_1\beta+2p_2\beta(1-\beta)(1-p)+2p_2\beta^2\frac{1-p}{2-p}}{K}+o\left(\kappa^2\right);\\
\mathbb{P}\left[X^{IRT}(A,B,C)=(1,0,1)\right] &= \frac{p_2\frac{(1-\beta)^2}{2}}{K}+o\left(\kappa^2\right);\\
\mathbb{P}\left[X^{IRT}(A,B,C)=(1,1,0)\right] &= \frac{p_2\beta(1-\beta)p}{K}+o\left(\kappa^2\right);\\
\mathbb{P}\left[X^{IRT}(A,B,C)=(0,1,1)\right] &= \frac{p_2\beta(1-\beta)p}{K}+o\left(\kappa^2\right);\\
\mathbb{P}\left[X^{IRT}(A,B,C)=(0,2,0)\right] &= \frac{p_2\beta^2\frac{p}{2-p}}{K}+o\left(\kappa^2\right);\\
\mathbb{P}\left[X^{IRT}(A,B,C)=(i_A,i_B,i_C)\right] &= o\left(\kappa^2\right), \qquad \text{for } i_A+i_B+i_C\geq 3.
\end{aligned}$$

Similarly, for X^{Mod}:

$$\begin{aligned}
\mathbb{P}\left[X^{Mod}(A,B,C)=(0,1,0)\right] &= \frac{p_1\beta p+2p_2\beta(1-\beta)p(1-p)+2p_2\beta^2p(2-p)\frac{1-p}{2-p}}{K'}+o\left(\kappa^2\right);\\
\mathbb{P}\left[X^{Mod}(A,B,C)=(1,0,1)\right] &= \frac{p_2\frac{(1-\beta)^2}{2}p^2}{K}+o\left(\kappa^2\right);\\
\mathbb{P}\left[X^{Mod}(A,B,C)=(1,1,0)\right] &= \frac{p_2\beta(1-\beta)p^2}{K'}+o\left(\kappa^2\right);\\
\mathbb{P}\left[X^{Mod}(A,B,C)=(0,1,1)\right] &= \frac{p_2\beta(1-\beta)p^2}{K'}+o\left(\kappa^2\right);\\
\mathbb{P}\left[X^{Mod}(A,B,C)=(0,2,0)\right] &= \frac{p_2\beta^2(2-p)\frac{p^2}{2-p}}{K'}+o\left(\kappa^2\right);\\
\mathbb{P}\left[X^{Mod}(A,B,C)=(i_A,i_B,i_C)\right] &= o\left(\kappa^2\right), \qquad \text{for } i_A+i_B+i_C\geq 3.
\end{aligned}$$

Here K and K' are given in Eqs. $(2.17\ \&\ 2.18)$ respectively. Now consider

$$\begin{aligned}
&\mathbb{P}\left[X^{IRT}(A\cup B\cup C)=1\right]-\ \mathbb{P}\left[X^{Mod}(A\cup B\cup C)=1\right]\\
&=\frac{K'-pK}{KK'}\left\{p_1\beta+2p_2\beta(1-\beta)(1-p)\right\}+\frac{K'-p(2-p)K}{KK'}\left\{\frac{2p_2\beta^2(1-p)}{2-p}\right\}+o\left(\kappa^2\right).
\end{aligned} \tag{2.19}$$

For $\beta\in\left(0,\sqrt{2}-1\right)$, Example 2.1 shows $K'<pK$ and, since $2-p\geq 1$, $K'<p(2-p)K$. Thus for $\beta\in\left(0,\sqrt{2}-1\right)$ and κ small enough $\mathbb{P}\left[X^{IRT}(A\cup B\cup C)=1\right]<\mathbb{P}\left[X^{Mod}(A\cup B\cup C)=1\right]$ (and hence $\mathbb{P}\left[X^{IRT}(A\cup B\cup C)\geq 2\right]>\mathbb{P}\left[X^{Mod}(A\cup B\cup C)\geq 2\right]$) since the contribution from $o(\kappa^2)$ is negligible. For $n\geq 3$ the probability that X^{IRT} or X^{Mod} has n points in $A\cup B\cup C$ is negligible. The desired result $X^{IRT}(A\cup B\cup C)\succ_{st}X^{Mod}(A\cup B\cup C)$ follows.

So for κ, β small $D(A \cup B \cup C) \succ_{st} D'(A \cup B \cup C)$, $X^{IRT}(A \cup B \cup C) \succ_{st} X^{Mod}(A \cup B \cup C)$; for β large $D(A \cup B \cup C) \prec_{st} D'(A \cup B \cup C)$. The example below shows that $X^{IRT}(A \cup B \cup C)$ need not always stochastically dominate or be dominated by $X^{Mod}(A \cup B \cup C)$.

Example 2.3. For the set up as in Example 2.2 but with $\beta > \sqrt{2} - 1 \simeq 0.4$ then $X^{IRT}(A \cup B \cup C)$ need not always stochastically dominate or be dominated by $X^{Mod}(A \cup B \cup C)$.

Since p_i is a Poisson probability with mean κ, we have that $p_1 = 2\frac{p_2}{\kappa}$. Substituting this and explicit expressions for $K' - pK$ and $K' - p(1-p)K$ into Eq. (2.19), we have

$$\mathbb{P}\left[X^{IRT}(A \cup B \cup C) = 1\right] - \mathbb{P}\left[X^{Mod}(A \cup B \cup C) = 1\right] = \frac{p_2^2 \beta p (1-p)}{KK'} \tilde{M} + o\left(\kappa^2\right);$$
$$\tilde{M} = \frac{(3p-2)\beta^2 + 2(2-p)\beta - (2-p)}{\kappa(2-p)} + (1-\beta)(1-p)\frac{-p\beta^2 + 2(1-p)\beta - (2-p)}{2-p}.$$

For $p = \beta = 0.5$, $\tilde{M}$ is strictly negative; so in this case $X^{IRT}(A \cup B \cup C) \succ_{st} X^{Mod}(A \cup B \cup C)$ (arguing as in Example 2.2). However, for $p = 0.5$ and $\beta = 0.9$

$$(3p-2)\beta^2 + 2(2-p)\beta - (2-p) + (1-\beta)(1-p)\left\{-p\beta^2 + 2(1-p)\beta - (2-p)\right\} > 0.$$

Since we require κ to be small enough so that $p_n = o(\kappa^2)$, for $n \geq 3$, κ must be less than 1. This yields $\mathbb{P}\left[X^{IRT}(A \cup B \cup C) = 1\right] > \mathbb{P}\left[X^{Mod}(A \cup B \cup C) = 1\right]$; so that for this p and β the ordering between $X^{IRT}(A \cup B \cup C)$ and $X^{Mod}(A \cup B \cup C)$ is reversed, ie. $X^{IRT}(A \cup B \cup C) \prec_{st} X^{Mod}(A \cup B \cup C)$.

Remark 2.6. From Example 2.1 $\beta > \sqrt{2} - 1 \simeq 0.4$ implies $D(A \cup B \cup C) \prec_{st} D'(A \cup B \cup C)$; thus even though for we have the above ordering between the dominating processes, the same is not necessarily true for the conditionally thinned processes. Only under conditioning by a lattice $\mathcal{E}$ may one guarantee that the conditional thinning of a stochastically larger process is stochastically larger.

In this chapter it is shown that the Van Lieshout & Van Zwet (2001) algorithm (VLVZ) generally outputs biased samples; in some cases strictly smaller while in others strictly larger samples. The comment by Van Lieshout & Van Zwet (2003) that efficient sampling of conditional Boolean models is still an open problem hence motivates further exploration of perfect sampling for such models. In the following chapter we introduce other simulation algorithms for Boolean models conditioned to cover a finite set; in addition the relative performance of the perfect simulation algorithms is evaluated in order to determine the most efficient algorithm for such Boolean models.

Chapter 3

Conditional Boolean Models

3.1 Introduction

Applications arising in microscopy, materials science, mining and oil industries, geology, forestry, ecology, physics and biology often involve modelling collections of randomly distributed compact sets in the plane. Examples include complicated geometric patterns depicting vegetation cover, oil/ore deposits, collection of cells and particles. Effective models for such purposes are 'Boolean Models'. This is a basic model in stochastic geometry and stereology; it has received much attention as a natural model for systems of randomly distributed particles and irregular patterns observed in nature (Matheron 1975; Hall 1988; Stoyan *et al.* 1995; Serra 1982; Molchanov 1996). A Boolean model is the union of independent (ie. non-interacting), randomly placed, non-empty, compact subsets of $\mathbb{R}^d$. The location of each such set is determined by the points of a Poisson process; thus a Boolean model formalizes the notion of 'independent and randomly distributed sets':

Definition 3.1. Let $X = \{\xi_n\}_{n\geq 0}$ be a Poisson(λ) process on $\mathbb{R}^d$ and $\Xi_0, \Xi_1, \ldots$ independent identically distributed random compact sets in $\mathbb{R}^d$ satisfying $\mathbb{E}\left[m_d\left[\Xi_o \oplus K\right]\right] < \infty$ for all compact K. Here m_d denotes Lebesgue measure on $\mathbb{R}^d$ and $\Xi_0 \oplus K = \{x + y;\ x \in \Xi_0,\ y \in K\}$ Minkowski-addition (Stoyan *et al.* 1995). The *Boolean model* $\mathcal{U}(X)$ of *germs* $\{\xi_n\}$ and *grains* $\{\Xi_n\}$ is constructed as

$$\mathcal{U}(X) = \bigcup_{n\geq 1} (\xi_n \oplus \Xi_n). \tag{3.1}$$

The underlying germ process of a Boolean model is a Poisson point process, which is mathematically quite tractable and has been immensely studied (Section 1.1.5). Nevertheless inference for the Boolean model is far from trivial (Molchanov 1996). The occlusion caused by observing only the *union* of the grains rather than the individual germ-grain pairs poses difficulties in trying to estimate both macroscopic and microscopic features. Such characteristics include the volume or perimeter of the grains, number of overlapping/connected grains, contact distributions and intensity of the underlying Poisson process. Hence it may be worthwhile trying a Monte Carlo simulation procedure. In many instances experimental data is often available for the physical phenomenon being studied. For example, one may know that certain locations are covered (or uncovered) by the grains, or that a given set of points are connected by the grains, etc. "Simulations that respect these experimental data are desirable because they will have a reduced variability, which may considerably speed up the Monte Carlo procedure. Such simulations are called *conditional simulations*" (Lantuéjoul 1997).

In this chapter the exact sampling of a conditional Boolean model is considered for the case when the germs are homogeneous Poisson, grains are circular disks of fixed radius and the conditioning event is the coverage of a *finite* set; each element of the conditioning set being referred to as a node (cf. Section 3.2). Suppose that the unconditional Boolean model can be generated as the equilibrium state of an ergodic, time-reversible Markov jump process. If the transitions of the process which violate the conditioning are not allowed then the restricted process will still satisfy detailed balance equations (Eq. 1.19). Furthermore if the restricted process is also irreducible then its equilibrium distribution will be that of the conditional Boolean model (Lantuéjoul 1997). For example, if a Metropolis-Hastings sampler (Section 1.4) is employed to sample the conditional Boolean model, then any transition violating the conditioning will have zero acceptance probability. The sampler is still irreducible, so the process converges to the conditional Boolean model.

This idea was implemented by Kendall & Thönnes (1999) who use spatial birth-death processes to construct the first *perfect simulation* algorithm for conditional Boolean models. They employ the method of dominated Coupling form the Past (domCFTP) which was developed by Kendall (1998), following on from the seminal work of Propp & Wilson (1996) who first proposed Coupling From The Past (CFTP) as a tool for perfect simulation. CFTP and domCFTP are introduced in Sections 1.4.1 & 1.4.2 respectively. An alternative perfect simulation algorithm for conditional Boolean models was developed by Cai & Kendall (2002); here the set up of the algorithm is in terms of

random variables rather than spatial point processes. Their simulation algorithm is a generalization of dominated CFTP, which they call *Extended State-Space CFTP*. Thönnes (2000) also presents a review of other simulation algorithms for this model.

Section 3.3 briefly presents the simplest simulation procedure for conditional Boolean models, namely Rejection sampling. The efficiency of the Rejection algorithm is inversely proportional to the acceptance probability. For extreme model parameters (such as low intensity λ or large conditioning set) the acceptance probability can be very low, and Rejection sampling may prove inefficient. In Section 3.3.1 we introduce a variant of Rejection Sampling, namely *2-Stage Rejection*. Simple yet logical modifications are made to the usual Rejection algorithm in order to produce the 2-Stage Rejection algorithm. It turns out that this variant is much more efficient than straightforward Rejection (Lemma 3.1). Section 3.4 then introduces the idea of using spatial birth-death processes to simulate conditional Boolean models; the algorithm of Cai & Kendall (2002) is described in Section 3.4.1. In Section 3.5 yet another CFTP-based simulation algorithm, which employs a Gibbs updating scheme, is presented. The exact Gibbs sampler described uses the ideas of Häggström *et al.* (1999), who consider perfect simulation of the penetrable spheres model and the area-interaction point process (see Section 1.1.11 and Chapter 4 for details on these point process models).

For any simulation algorithm it is important to validate it in order to ensure that there are no coding errors, or indeed even theoretical inconsistencies. Therefore the Cai & Kendall and exact Gibbs algorithms are validated via χ^2 tests in Appendix A. The final part of this chapter presents some results of simulation experiments carried out in order to quantitatively evaluate the efficiency of the 2-Stage, Cai & Kendall and exact Gibbs algorithms. The performance is measured in terms of actual run times (in seconds). At first glance the results are quite striking: Figures 3.3 & 3.4 depict how the run times of Rejection sampling compare with the above three algorithms. The 2-Stage algorithm is seen to have much lower run times than Rejection, making the former much more competitive to the Cai & Kendall and exact Gibbs algorithms.

More rigorous simulation experiments (Sections 3.7.3 & 3.7.4) also suggest that the 2-Stage method is very competitive for moderate intensity λ and number of conditioning nodes k. As λ decreases or k increases the relative performance of the Cai & Kendall method is better than 2-Stage Rejection and exact Gibbs. The exact Gibbs algorithm is, unfortunately, always outperformed by either of the other two. This is however not the case when interactions are introduced; see Chapter

5 where a 2-Stage Rejection, modified Cai & Kendall and exact Gibbs algorithm for the conditional area-interaction process are compared in terms of relative efficiency.

3.2 Notation

Let $\mathcal{C} = \{c_1, \ldots, c_k\}$ denote the finite conditioning set and $\mathcal{U}(x)$ the Boolean model of disks associated with a point configuration x, defined by the union $\mathcal{U}(x) = \bigcup_{\xi \in x} B_r(\xi)$, where $B_r(\xi)$ is a disk of radius r and centre ξ (Definition 3.1). For $\lambda > 0$ let $\pi_\lambda^{\mathcal{C}}$ denote the distribution of a Poisson(λ) process which is conditioned so that its associated Boolean model covers the set $\mathcal{C}$, and read "$\mathcal{U}(x) \supset \mathcal{C}$" as "$\mathcal{U}(x)$ covers $\mathcal{C}$".

The set $\mathcal{C}$ can be viewed as a graph with nodes $\{c_1, \ldots, c_k\}$ and edges determined as follows. Two nodes c_i and c_j are *connected* (ie. there is an edge between them) if $\mathcal{U}(c_i) \cap \mathcal{U}(c_j) \neq \emptyset$, written $c_i \sim_r c_j$ since the radius r essentially governs whether two points are connected. Similarly, c_i and c_j are *disconnected* if $\mathcal{U}(c_i) \cap \mathcal{U}(c_j) = \emptyset$. To avoid confusion between the points ξ_n of X and conditioning points c_i of $\mathcal{C}$, the $\{\xi_n\}$ will be referred to as *germs*, $\{\mathcal{U}(\xi_n)\}$ as *grains* and $\{c_i\}$ as *(conditioning) nodes*. Figure 3.1 below shows a realization of a Boolean model conditioned to cover a part of the Redwood seedlings data (Figure 1).

3.3 Rejection Sampling

This is the simplest sampling algorithm for sampling from $\pi_\lambda^{\mathcal{C}}$. The idea is straightforward: draw an unconditional Boolean model. If it covers the set $\mathcal{C}$ then it has the desired distribution; else discard it and draw another one independently. Continue in this fashion until the required sample is obtained.

Algorithm 3.1 (Rejection: Conditional Boolean Model).

Set $T = 1$.

while $T > 0$:

draw $X \sim$ Poisson(λ).

if $\mathcal{U}(X) \supset \mathcal{C}$: set $T = -1$.

else: set $T = T + 1$.

return X.

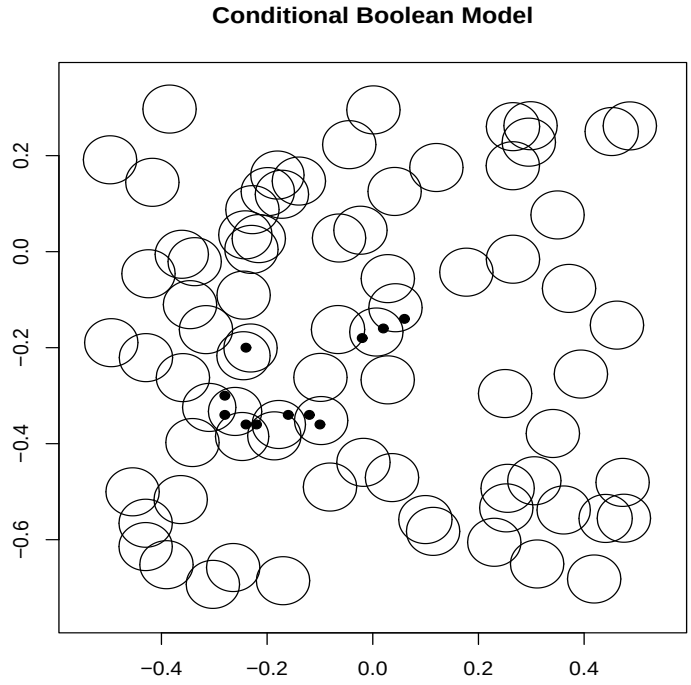

Figure 3.1: Sample of a conditional Boolean model. The conditioning nodes (solid dots) are a part of the Redwood seedlings data set (Figure 1). The underlying Poisson intensity $\lambda = 40$ and disk radius $r = 0.05$.

Algorithm 3.1 samples a Poisson(λ) process restricted to those realizations whose associated Boolean model covers $\mathcal{C}$, so the distribution of the output is $\pi_\lambda^{\mathcal{C}}$. The expected number of iterations of the Rejection algorithm is equal to the inverse of the probability that the Boolean model associated with a Poisson(λ) covers $\mathcal{C}$. This probability becomes very low as λ decreases or k increases, hence Rejection becomes inefficient. Figures 3.3 & 3.4 in Section 3.7, compare the empirical run times of Rejection sampling against the algorithms described in the following sections. It is seen that the Rejection algorithm does slow down as λ decreases or k increases. In the next section a more efficient variant of Rejection, namely the 2-Stage Rejection, is introduced. The results of Section 3.7 indicate that 2-Stage Rejection is very competitive for moderate parameter values to the relatively more complicated algorithms presented in Sections 3.4.1 & 3.5.1.

3.3.1 2-Stage Rejection

The Rejection Algorithm 3.1 of Section 3.3 can be viewed as a one-off accept/reject mechanism for drawing from $\pi_\lambda^{\mathcal{C}}$, the distribution of a Poisson(λ) process conditioned to cover $\mathcal{C}$. An unconditional Poisson(λ) process is drawn; it is accepted if its associated Boolean model covers $\mathcal{C}$; else it is

discarded and another one is independently drawn. This is carried out until a sample is accepted.

When the conditioning set $\mathcal{C}$ gets large Rejection sampling can be quite inefficient, in terms of either the run-time (in seconds) or the number of iterations required before a sample is accepted (see Figures 3.3 & 3.4 for an illustration). Indeed the acceptance probability decreases monotonically as more and more nodes are added to $\mathcal{C}$: if $X \sim \text{Poisson}(\lambda)$ process then $\mathbb{P}\left[\mathcal{U}\left(X\right) \supset \mathcal{C}\right] \geq \mathbb{P}\left[\mathcal{U}\left(X\right) \supset \mathcal{C} \cup \{\xi\}\right]$, for $\xi \notin \mathcal{C}$. Thus, for $A \subseteq \mathcal{C}$, the number of iterations required to draw from π_λ^A will be (stochastically) less than those required to draw from $\pi_\lambda^{\mathcal{C}}$; moreover, if $Y \sim \pi_\lambda^A$ then $\mathbb{P}\left[\mathcal{U}\left(Y\right) \supset \mathcal{C}\right] > 0$. Therefore, in order to draw from $\pi_\lambda^{\mathcal{C}}$ it will be just as efficient, if not more so (in terms of number of iterations), to sample from π_λ^A and then check if the sample also covers $\mathcal{C}$. This is the idea behind the 2-Stage Rejection algorithm.

Recall, from Section 3.2, that $\mathcal{C}$ can be viewed as a graph with nodes $\{c_1, \ldots, c_k\}$. Two conditioning nodes c_i and c_j are connected, written $c_i \sim_r c_j$ if $\mathcal{U}\left(c_i\right) \cap \mathcal{U}\left(c_j\right) \neq \emptyset$, where $\mathcal{U}\left(c\right)$ is a disk of fixed radius r centred at c. Similarly c_i and c_j are *disconnected* if $\mathcal{U}\left(c_i\right) \cap \mathcal{U}\left(c_j\right) = \emptyset$.

Definition 3.2. Define A^* to be a subset of $\mathcal{C}$ that has the maximum number of *disconnected* nodes.

Let $A^* \subseteq \mathcal{C}$ be as defined above and X an unconditional Poisson(λ) process on $\mathcal{U}\left(\mathcal{C}\right)$. If X_c denotes the process X restricted to $\mathcal{U}\left(c\right)$ then the $\{X_c;\ c \in A^*\}$ are independent Poisson(λ) processes since nodes in A^* are disconnected. Moreover, if $\tilde{X} \sim \pi_\lambda^{A^*}$ is a Poisson(λ) process whose associated Boolean model covers A^*, then $\left\{\tilde{X}_c;\ c \in A^*\right\}$ are independent Poisson(λ) processes, each containing a positive number of germs. A Rejection algorithm to sample from $\pi_\lambda^{A^*}$ would involve drawing an unconditional Poisson process and checking if the number of germs in $\mathcal{U}\left(c\right)$ is positive for all $c \in A^*$. The expected number of iterations required such a procedure is $\left(1 - e^{-\lambda m_2[\mathcal{U}(c)]}\right)^{-|A^*|}$, where $|A^*|$ is the number of nodes in A.

However the observation that $\left\{\tilde{X}_c;\ c \in A^*\right\}$ are *independent* Poisson processes provides a more efficient way to sample from $\pi_\lambda^{A^*}$. For each $c \in A^*$ sample $\tilde{X}_c \sim \pi_\lambda^{\{c\}}$ *independently*; this is straightforward since $\tilde{X}_c$ is a Poisson(λ) process on $\mathcal{U}\left(c\right)$ conditioned to have at least one germ. Setting $\tilde{X} = \bigcup_{\{c \in A^*\}} \tilde{X}_c$ yields the desired sample with distribution $\pi_\lambda^{A^*}$. The expected number of iterations for this is $|A^*| \left(1 - e^{-\lambda m_2[\mathcal{U}(c)]}\right)^{-1}$. For reasonably sized A^* this value will be much lower than the corresponding one for Rejection.

Algorithm 3.2 (2-Stage Rejection: Conditional Boolean Model).

Choose $A^* \subseteq \mathcal{C}$ as in Definition 3.2; set $T = 1$.

while $T > 0$:

for each $c \in A^*$: draw $\tilde{X}_c \sim \pi_\lambda^{\{c\}}$.

draw $X \sim$ Poisson(λ) process on $\mathcal{U}(\mathcal{C}) \setminus \mathcal{U}(A^*)$.

set $\tilde{X} = \bigcup_{\{c \in A^*\}} \tilde{X}_c$.

if $\mathcal{U}\left(\tilde{X} \cup X\right) \supset \mathcal{C}$: set $T = -1$.

else: set $T = T + 1$.

return $\tilde{X} \cup X$.

By independence, $\tilde{X} \cup X$ is a Poisson(λ) process on $\mathcal{U}(\mathcal{C})$ conditioned to cover A^*; the above algorithm therefore outputs a Poisson(λ) process restricted to those realizations which cover $\mathcal{C}$, as required. For $A \subseteq \mathcal{C}$, let $\tau_{\text{rej}}(A)$ denote the (random) number of iterations in the Rejection Algorithm 3.1 required to obtain a sample from π_λ^A. Then $\tau_{\text{rej}}(A)$ can take values in $\{1, 2, \ldots\}$ and has a Geometric distribution with success probability $p(A) = \mathbb{P}[\mathcal{U}(X) \supset A]$ and expectation $\mathbb{E}[\tau_{\text{rej}}(A)] = p(A)^{-1}$. Set $p = \mathbb{P}[\mathcal{U}(X) \supset \{c\}] = \left(1 - e^{-\lambda m_2[\mathcal{U}(c)]}\right)$, the probability that $\mathcal{U}(X)$ covers a single conditioning node c. Define also $p(\mathcal{C} \mid A) = \mathbb{P}[\mathcal{U}(X) \supset \mathcal{C} \mid \mathcal{U}(X) \supset A]$, as the (conditional) probability that $\mathcal{U}(X)$ covers $\mathcal{C}$ given that it covers some $A \subseteq \mathcal{C}$, and $\tau(\mathcal{C} \mid A)$ a Geometric random variable with success probability $p(\mathcal{C}|A)$.

Suppose A^* is as in Definition 3.2; in order to sample from $\pi_\lambda^{\mathcal{C}}$ the 2-Stage algorithm involves two stages of iteration steps. One implicit in the rejection step where a sample from $\pi_\lambda^{A^*}$ is drawn; the other explicitly in the number of times a sample from $\pi_\lambda^{A^*}$ must be drawn before it also covers $\mathcal{C}$. Denote by $\tau_{\text{2Stg}}(A^*)$ the total number of iterations required by the 2-Stage algorithm. Let $\tau'(A^*)$ be a sum of $|A^*|$ independent and identically distributed Geometric random variables each with success probability p. Then $\mathbb{E}[\tau'(A^*)] = |A^*|p^{-1}$; furthermore if $\tau_i'(A^*)$ are a sequence of independent and identically distributed random variables then

$$\tau_{\text{2Stg}}(A^*) = \sum_{i=1}^{\tau(\mathcal{C}|A^*)} \tau_i'(A^*). \tag{3.2}$$

$$\mathbb{E}[\tau_{\text{2Stg}}(A^*)] = \mathbb{E}\left[\mathbb{E}\left[\sum_{i=1}^{\tau(\mathcal{C}|A^*)} \tau_i'(A^*)\right] \middle| \tau(\mathcal{C} \mid A^*)\right] = \mathbb{E}[\tau(\mathcal{C} \mid A^*)] \frac{|A^*|}{p}. \tag{3.3}$$

Lemma 3.1. *Given A^* (definition 3.2) let $\tau_{2Stg}(A^*)$ be defined by Eq. (3.2) above and $\tau_{rej}(\mathcal{C})$ the number of iterations of the Rejection Algorithm 3.1. Set $p = \left(1 - e^{-\lambda m_2[\mathcal{U}(c)]}\right)$ and let $|A^*|$ denote the number of conditioning nodes in A^*. If $\frac{|A^*|}{p} < \frac{1}{p^{|A^*|}}$ then $\mathbb{E}[\tau_{2Stg}(A^*)] < \mathbb{E}[\tau_{rej}(\mathcal{C})]$.*

Proof. From Eqs. (3.2 & 3.3) we have that $\mathbb{E}[\tau_{2\text{Stg}}(A^*)] = \mathbb{E}[\tau(\mathcal{C} \mid A^*)]\frac{|A^*|}{p}$. Now $\tau_{\text{rej}}(\mathcal{C})$ is a Geometric random variable with success probability $p(\mathcal{C}) = \mathbb{P}[\mathcal{U}(X) \supset \mathcal{C}]$. Given A^*, this probability can be decomposed as $\mathbb{P}[\mathcal{U}(X) \supset \mathcal{C}] = \mathbb{P}[\mathcal{U}(X) \supset \mathcal{C} \mid \mathcal{U}(X) \supset A^*] \times \mathbb{P}[\mathcal{U}(X) \supset A^*]$; thus $p(\mathcal{C}) = p(\mathcal{C} \mid A^*)\, p(A^*)$. Since A^* contains only disconnected nodes $p(A^*) = p^{|A^*|}$. Thus $p(\mathcal{C}) = p(\mathcal{C} \mid A^*)\, p^{|A^*|}$; so $\mathbb{E}[\tau_{\text{rej}}(\mathcal{C})] = p(\mathcal{C})^{-1} = \left(p(\mathcal{C} \mid A^*)\, p^{|A^*|}\right)^{-1}$. But $p(\mathcal{C} \mid A^*)$ is the success probability of the Geometric random variable $\tau(\mathcal{C} \mid A^*)$; therefore $\mathbb{E}[\tau_{\text{rej}}(\mathcal{C})] = \mathbb{E}[\tau(\mathcal{C} \mid A^*)]\frac{1}{p^{|A^*|}}$. The result hence follows if $\frac{|A^*|}{p} < \frac{1}{p^{|A^*|}}$. □

This Lemma thus proves that the 2-Stage Rejection algorithm is more efficient than its Rejection counterpart. In Section 3.7 results of simulation experiments are presented, with the objective of quantitatively comparing the performance of the algorithms described in this chapter. One of the experiments involves comparing the actual run times (in seconds) of the Rejection and 2-Stage algorithms. The increase in efficiency by adopting the 2-stage sampling procedure is quite striking. It is seen that for low λ and/or large $\mathcal{C}$ the Rejection Algorithm 3.1 becomes quite inefficient. However the run times of the 2-Stage Algorithm 3.2 are much lower than those of Rejection; moreover it also fares well against the more complex perfect simulation algorithms presented in the following sections. More details and discussions are postponed till Section 3.7.

3.4 Simulation via Spatial Birth-Death Processes

A spatial birth-death process (Section 1.3) is a continuous-time Markov jump process taking values in an *Exponential Space* (space of locally finite point configurations; see Section 1.1.1). The only transitions are births (where a germ or individual is added to the current configuration) or deaths (where a germ is deleted). Birth incidents are controlled by the birth rate b and deaths by the per capita death rate d. For a constant birth rate b and unit per capita death rate the equilibrium distribution of the spatial birth-death process is that of a Poisson(b) process (Example 1.3).

Lantuéjoul (1997, 2002) presents an algorithm of Matheron which employs this result: that the

stationary distribution of a spatial birth-death process is an unconditional Poisson process. He shows that forbidding those transitions (ie. births or deaths) that violate the given conditioning (eg. coverage, connectivity) yields a conditional process whose limiting distribution is that of a conditional Poisson process. This provides a dynamic sampling algorithm for the conditional Boolean model via a spatial birth-death process Φ:

Algorithm 3.3 (Lantuéjoul (1997)).

Set $\Phi(0) = \emptyset$, $t = 0$.

while $t \leq T$:

simulate the next incident time τ of Φ; set $t = t + \tau$.

if next incident is a birth ξ:

if addition of ξ does not disrupt the conditioning: set $\Phi(t) = \Phi(t-) \cup \{\xi\}$.

else: set $\Phi(t) = \Phi(t-)$.

else if next incident is a death ξ:

if removal of ξ does not disrupt the conditioning: set $\Phi(t) = \Phi(t-) \setminus \{\xi\}$.

else: set $\Phi(t) = \Phi(t-)$.

return $\Phi(T)$.

This algorithm yields an approximate sample from a conditional Boolean model, where T is set by the user. However, after the inception of Coupling From the Past (CFTP) by Propp & Wilson (1996), there has been much research on the possibility of perfectly simulating point processes (Kendall 1998; Kendall & Møller 2000; Häggström *et al.* 1999; Fernández *et al.* 2002; Thönnes 2000). Indeed, Kendall & Thönnes (1999) were the first to present a perfect simulation algorithm for the conditional Boolean model, where the conditioning event considered was the coverage of a finite set of nodes. Subsequent algorithms have also been presented and we now describe the sampling scheme proposed by Cai & Kendall (2002).

3.4.1 The Cai & Kendall Algorithm

The Cai & Kendall procedure perfectly samples a collection of correlated Poisson random variables conditioned to be positive. Such a collection can then be converted into a sample of a conditional

Boolean model. Their perfect simulation algorithm is a variant of dominated CFTP, namely *Extended State-Space* CFTP.

Let $\mathcal{C} = \{c_1, \ldots, c_k\}$ be the set of conditioning nodes and X an unconditional Poisson(λ) process on $\mathbb{R}^2$. Denote a disk of fixed radius r and centre ξ by $B_r(\xi)$; the Boolean model (Definition 3.1) of disks associated with X is then given by $\mathcal{U}(X) = \bigcup_{\xi \in X} B_r(\xi)$. For each non-empty $A \subseteq \mathcal{C}$ define

$$E_A = \left(\bigcap_{x \in A} B_r(x) \right) \setminus \left(\bigcup_{y \in \mathcal{C};\ y \notin A} B_r(y) \right). \tag{3.4}$$

So a germ in E_A will cover exactly those conditioning nodes in A and no others. Note that even if $A \neq \emptyset$ but $E_A = \emptyset$, the dynamics of the algorithm are unaffected by this; see the comment at the end of this section. An illustration of these regions is given by Figure 3.2 in Section 3.7.

Let X_A denote the restriction of X to E_A and, for $c \in \mathcal{C}$, let $X_{\{c\}} \equiv X_c = \bigcup_{\{A;\ c \in A\}} X_A$ represent those germs of X which cover the conditioning node $\{c\}$, ie. those germs contained in $B_r(c)$. Then $\{X_{c_i}\}_{i=1}^k$ is a collection of *correlated Poisson processes*. Equivalently, since the number of germs of a Poisson process covering a given node is a Poisson random variable, $\{n(X_{c_i})\}_{i=1}^k$ is a collection of correlated Poisson random variables, where $n(X)$ denotes the number of germs in X. Conditioning X to cover $\mathcal{C}$ is equivalent to conditioning the processes $\{X_{c_i}\}_{i=1}^k$ each to have at least one germ; thus "conditional Boolean models can be viewed as arising from correlated conditioned Poisson random variables" (Cai & Kendall 2002).

Let $\left\{ \tilde{X}_{c_i} \right\}_{i=1}^k$ be a collection of correlated Poisson processes each conditioned to have at least one germ; then $\tilde{X} = \bigcup_{i=1}^k \tilde{X}_{c_i}$ is a sample of a Poisson process which covers $\mathcal{C}$, with distribution denoted by $\pi_\lambda^{\mathcal{C}}$. Cai & Kendall employ immigration-death processes in order to carry out MCMC for a collection of correlated Poisson random variables conditioned to be positive; such a sample can then be converted into a conditional Boolean model. In the spirit of their construction, described below is a conditional simulation algorithm via spatial birth-death processes for sampling directly from $\pi_\lambda^{\mathcal{C}}$. The essence of the construction below is very similar to (and indeed motivated by) that of Cai & Kendall. Differences in certain aspects of the formulation arise due to the fact that they use immigration-death processes on $\mathbb{N}$ whereas spatial birth-death processes on $\mathbb{R}_e^2$ are used here.

Let Φ denote a spatial birth death process with birth rate λ and unit death rate per germ. Detailed balance calculations show that the equilibrium distribution of Φ is that of a Poisson(λ) process (cf. Example 1.3). So Φ_A, the restriction of Φ to E_A, converges to a Poisson(λ) process on E_A; and

$\Phi_{c_i}(t) = \bigcup_{\{A;\ c_i \in A\}} \Phi_A(t)$ converges to X_{c_i} defined above. This enables one to sample the collection of correlated Poisson processes $\{X_{c_i}\}_{i=1}^k$ given a realization of the $\{\Phi_A;\ A \subseteq \mathcal{C}\}$ processes. Births are generated uniformly and assigned to the relevant Φ_A process according to which region E_A the new germ falls in; deaths are chosen uniformly at random from the set of alive individuals since each has unit death rate.

There are more efficient ways of simulating unconditional Poisson processes than via spatial birth-death processes (eg. direct sampling); however the dynamic nature of spatial birth-death processes enables one to modify the construction in order to incorporate conditioning. The conditioning event considered is the coverage of the finite set $\mathcal{C}$. The birth of a germ never disrupts the conditioning, so there are no changes required for birth transitions. On the other hand, suppose that one of the Φ_A processes experiences a death transition. If deletion of this germ would uncover one or more of the conditioning nodes in A, ie. causes at least one of the Φ_{c_i} processes, for $c_i \in A$, to become empty, then this transition is disallowed. Such a germ is referred to as a *perpetuated* germ (Kendall & Thönnes 1999; Cai & Kendall 2002); it is granted an additional unit-rate Exponentially distributed lifetime, after which its deletion decision is re-evaluated.

In order to prescribe *replacement* death times for potential death times which are prevented from becoming actual, introduce independent unit-rate Poisson processes Z_A, one for each non-empty $A \subseteq \mathcal{C}$; define also a virtual process V_A which will be used to keep track of perpetuated germs. So, if a death in some Φ_A is disallowed, the perpetuated germ is deleted from Φ_A and stored in V_A. The next incident time $\hat{t} > t$ of Z_A determines a new potential death time for the perpetuated germ. Refer to such incidences as *perpetuation incidents* and $\hat{t}$ as the *perpetuation time* of V_A.

Each virtual process V_A is initialized so as to contain a single germ in the respective region E_A. If, for some $A \neq \emptyset$, $E_A = \emptyset$ then this will not affect the dynamics of the sampling procedure. The reason being that E_A empty implies that there exists $B \supset A$ with non-empty E_B. So as soon as a germ is born in E_B then both A and B are covered. The initializing and updating scheme ensures that $\tilde{\Phi}_{c_i} = \bigcup_{\{A;\ c_i \in A\}} \Phi_A \cup V_A$ always contains at least one germ for each i. Thus the virtual processes essentially supply extra individuals to ensure that the conditioning is always satisfied; furthermore each V_A process contains *at most one* perpetuated germ. Detailed balance shows that the equilibrium distribution of $\left\{\tilde{\Phi}_{c_i}\right\}_{i=1}^k$ is that of $\left\{\tilde{X}_{c_i}\right\}_{i=1}^k$ (Cai & Kendall 2002), where $\bigcup_{i=1}^k \tilde{X}_{c_i} \sim \pi_\lambda^{\mathcal{C}}$ is a Poisson(λ) process conditioned to cover $\mathcal{C}$.

3.4.2 Dominated CFTP Construction

In this section a perfect variant of the MCMC algorithm described above is presented. The target distribution is $\pi_\lambda^{\mathcal{C}}$, the distribution of a Poisson(λ) process conditioned to cover the finite set $\mathcal{C} = \{c_1, \ldots, c_k\}$. The essence of any CFTP-based algorithm is the "construction of interleaved sequences of upper- and lower-sandwich processes, bounding intermediate Markov processes which actually have the required equilibrium distribution" (Cai & Kendall 2002). The basis of the construction described below is, to a large extent, borrowed from the *Extended State-Space CFTP* construction of Cai & Kendall (2002). In their method the original un-ordered state-space is inserted as a subset of a bigger ordered space, thereby allowing a description of the sandwich processes which bound (in the larger ordered space) the target process.

A modification of their formulation yields an equivalent coupling construction where there is no need to extend the state space; this alternative formulation is therefore described here. Recall that the target processes of the MCMC procedure are $\left\{\tilde{\Phi}_{c_i}\right\}_{i=1}^{k}$, defined in terms of the spatial birth-death processes $\{\Phi_A\}$ and virtual processes $\{V_A\}$. In order to define upper and lower bounding processes, let '$\subseteq$' denote the usual inclusion ordering between two processes so that $y \subseteq x$ means that every germ in y is also contained in x. For $T > 0$ and non-empty A let $V_A^{T,max}(t)$, $V_A^{T,min}(t)$ denote the maximal, respectively, minimal virtual processes started at time $-T$, for $t \in [-T, 0]$. Define also

$$\tilde{\Phi}_c^{T,min}(t) = \bigcup_{\{A;\ c \in A\}} \Phi_A(t) \cup V_A^{T,min}(t). \tag{3.5}$$

$$\tilde{\Phi}_c^{T,max}(t) = \bigcup_{\{A;\ c \in A\}} \Phi_A(t) \cup V_A^{T,max}(t). \tag{3.6}$$

The recursive updating scheme has to maintain the monotonicity between the maximal and minimal processes with respect to the partial order '$\subseteq$'. So initialize the bounding processes such that $V_A^{T,min}(-T) \subseteq V_A^{T,max}(-T)$. Employing the 'cross-over' trick described in Kendall (1998) ensures that the required ordering between these two processes is maintained: the upper process is updated according to the state of the lower one and vice versa. Furthermore recall the unit-rate Poisson processes $\{Z_A\}$ from the previous section which supply additional death-times for perpetuated individuals stored in the virtual processes. Simulation of $\{\Phi_A\}$ and $\{Z_A\}$ backwards in time is straightforward since their equilibrium distributions are Poissonian. Let $\lfloor \frac{T}{2} \rfloor$ denote the integer part of $\frac{T}{2}$; the domCFTP construction can be summarized as follows:

Algorithm 3.4 (domCFTP: Conditional Boolean Model).

Fix $T > 0$.

while $T > 0$:

Extend $\{\Phi_A;\ A \subseteq \mathcal{C}\}$, $\{Z_A;\ A \subseteq \mathcal{C}\}$ backwards on the interval $\left[-T, -\lfloor \frac{T}{2} \rfloor\right)$.

Initialize $V_A^{T,min}(-T) = \emptyset$, $V_A^{T,max}(-T) = \{\xi\}$, where $\xi \in E_A$, for each $A \subseteq \mathcal{C}$.

for $t \in [-T, 0]$:

if t is the birth time of $\xi \in E_A$: set $\Phi_A(t) = \Phi_A(t-) \cup \xi$.

else if t is the death time of $\xi \in \Phi_A(\cdot)$:

set $\Phi_A(t) = \Phi_A(t-) \setminus \xi$.

if $\tilde{\Phi}_{c_i}^{T,min}(t-) = \{\xi\}$ for some $c_i \in A$: set $V_A^{T,max}(t) = \{\xi\}$.

if $\tilde{\Phi}_{c_i}^{T,max}(t-) = \{\xi\}$ for some $c_i \in A$: set $V_A^{T,min}(t) = \{\xi\}$.

else if t is a perpetuation time for $V_A(\cdot)$ process (ie. incident time of Z_A):

if $\tilde{\Phi}_{c_i}^{T,min}(t-) \supset V_A^{T,min}(t-)$ for all $c_i \in A$: set $V_A^{T,max}(t) = \emptyset$.

if $\tilde{\Phi}_{c_i}^{T,max}(t-) \supset V_A^{T,max}(t-)$ for all $c_i \in A$: set $V_A^{T,min}(t) = \emptyset$.

if $V_A^{T,min}(0) = V_A^{T,max}(0)$ for all A: set $T = -1$.

else: set $T = 2T$.

return $\left\{\tilde{\Phi}_{c_i}^{T,min}(0)\right\}_{i=1}^{k}$.

Note that when extending the $\{\Phi_A\}$ and $\{Z_A\}$ processes backwards over some interval in one iteration of the algorithm, it is vital to re-use the same realizations over the same interval for all subsequent iterations. Otherwise a bias is introduced; see the original formulation of CFTP in Propp & Wilson (1996).

Properties of the domCFTP Algorithm

Cai & Kendall (2002) show that the above Algorithm 3.4 satisfies the following Lemmas; brief proofs are given here since the arguments of the proofs will be referred to later in Chapter 5. The Lemmas outline the relationships which the maximal and minimal virtual processes must satisfy in order to ensure that the output of the algorithm has the required target distribution (Theorem 3.1). Kendall & Møller (2000) show that any CFTP-based construction satisfying the sandwiching, coalescence and funnelling properties, introduced in Section 1.4.2, must output samples from the

required target distribution. Thus these relationships between the maximal and minimal processes serve as a basis for validating any CFTP-based construction. Analogous results for the Exact Gibbs sampler (Algorithm 3.7) of Section 3.5 also hold, ensuring its correctness.

Lemma 3.2 (Sandwiching). *The virtual processes $V_A^{T,min}$ and $V_A^{T,max}$ of Algorithm 3.4 satisfy*

$$V_A^{T,min}(t) \subseteq V_A^{T,max}(t), \text{ for all } A \subseteq \mathcal{C}, -T \leq t \leq 0. \tag{3.7}$$

Proof. The key argument of the proof is that the 'cross-over' trick ensures that $V_A^{T,min}(t)$ will be non-empty only if $V_A^{T,max}(t)$ also is, following a death or perpetuation incident. The initial values of the virtual processes satisfies Eq. (3.7); the result then follows by induction along the sequence of incidents of over the time interval $[-T, 0]$. □

Lemma 3.3 (Coalescence). *If $V^{T,min}(t^*) = V^{T,max}(t^*)$ then $V^{T,min}(t) = V^{T,max}(t)$ for all $t \geq t^*$.*

Proof. At time t^* the maximal and minimal virtual processes coalesce, and hence so will $\tilde{\Phi}_{c_i}^{T,min}(t^*)$ and $\tilde{\Phi}_{c_i}^{T,max}(t^*)$ for all $c_i \in \mathcal{C}$. There is no longer any distinction between the minimal and maximal processes; thus from time t^* the transitions for $V_A^{T,min}$ and $V_A^{T,max}$ will be the same for all A. □

Lemma 3.4 (Funnelling). *For all $-S \leq -T \leq t \leq 0$ and $A \subseteq \mathcal{C}$*

$$V_A^{T,min}(t) \subseteq V_A^{S,min}(t) \subseteq V_A^{S,max}(t) \subseteq V_A^{T,max}(t). \tag{3.8}$$

Proof. The key point here is that the 'cross-over' trick ensures that if Eq. (3.8) holds at time $-T$ then it will persist for all $t \in [-T, 0]$. So suppose the relationship holds up to some time $t > -T$; then $\tilde{\Phi}_{c_i}^{T,min}(t-) \subseteq \tilde{\Phi}_{c_i}^{S,min}(t-) \subseteq \tilde{\Phi}_{c_i}^{S,max}(t-) \subseteq \tilde{\Phi}_{c_i}^{T,max}(t-)$ for all $c_i \in \mathcal{C}$. Following a death or perpetuation event at time t, $V_A^{T,min}(t)$ will only be non-empty if $V_A^{S,min}(t)$ also is, and $V_A^{S,max}(t)$ will only be non-empty if $V_A^{T,max}(t)$ also is. Induction along the sequence of incidents on $[-T, 0]$ completes the proof. □

Lemma 3.5. *For sufficiently large T, the virtual processes $V^{T,min}$ and $V^{T,max}$ coalesce on $[-T, 0]$.*

Proof. For $i = 1, \ldots, k$ let $D_i = \left\{\tilde{\Phi}_{c_i}^{T,max} \text{ experiences a birth in the time interval } [t, t+\delta]\right\}$. Cai & Kendall (2002) show that, conditional on $\bigcap_{i=1}^{k} D_i$ occurring at time $t + \delta$, coalescence of $V_A^{T,min}$ and $V_A^{T,max}$ will occur with a conditional probability (given $\bigcap_i D_i$) greater than

$$\left(1 - \mathbb{P}\left[\text{Poisson}(\delta) = 0\right]\right)^{\left(2^k - 1\right)} \times \mathbb{P}\left[\text{Exponential}\left(2^k - 1\right) > 2\delta\right] = \rho_1 > 0.$$

Let C_t be the event that coalescence occurs on $[t, t+2\delta]$. The second Borel-Cantelli Lemma shows that $C_{2s\delta}$ occurs for infinitely many s, hence coalescence is almost sure to happen. □

Theorem 3.1. *Algorithm 3.4 terminates in finite time, and the output of the algorithm has distribution $\pi_\lambda^{\mathcal{C}}$, the distribution of a Poisson process of intensity λ conditioned to cover the set $\mathcal{C}$.*

Proof. Lemma 3.5 establishes that Algorithm 3.4 terminates almost surely in finite time. Let the target processes be denoted by $\left\{\tilde{\Phi}_{c_i}\right\}_{i=1}^{k}$ and consider a virtual simulation of these processes from time $-\infty$, with $\tilde{\Phi}_{c_i}^{T}$ denoting the process on $[-T, 0]$. Furthermore, for a given T, couple the evolution of $\left\{\tilde{\Phi}_{c_i}^{T}\right\}_{i=1}^{k}$ to the realizations of $\{\Phi_A\}$ and $\{Z_A\}$ on the interval $[-T, 0]$.

By construction of the lower and upper virtual processes $V_A^{T,min}$ and $V_A^{T,max}$, the following holds at time T for all i: $\tilde{\Phi}_{c_i}^{T,min}(-T) \subseteq \tilde{\Phi}_{c_i}^{T}(-T) \subseteq \tilde{\Phi}_{c_i}^{T,max}(-T)$. Lemmas 3.2 & 3.4 then ensure that

$$\tilde{\Phi}_{c_i}^{T,min}(t) \subseteq \tilde{\Phi}_{c_i}^{S,min}(t) \subseteq \tilde{\Phi}_{c_i}^{T}(t) \subseteq \tilde{\Phi}_{c_i}^{S,max}(t) \subseteq \tilde{\Phi}_{c_i}^{T,min}(t)$$

holds for all $-S \leq -T \leq t \leq 0$ and i.

A virtual simulation of $\left\{\tilde{\Phi}_{c_i}\right\}_{i=1}^{k}$ from time $-\infty$ must mean that at time 0 the processes are in equilibrium. Set $T^* = \inf\left\{T \geq 1;\ \tilde{\Phi}_{c_i}^{T,min}(0) = \tilde{\Phi}_{c_i}^{T,max}(0) \text{ for all } i\right\}$; hence, for $T \geq T^*$, $\tilde{\Phi}^{T}(0) = \tilde{\Phi}^{T^*,min}(0) = \tilde{\Phi}^{T^*,max}(0)$. Thus $\tilde{\Phi}^{T^*,min}(0) = \lim_{T\to\infty} \tilde{\Phi}^{T}(0)$ exists almost surely and has required distribution, since $\tilde{\Phi}^{T}(0)$ converges to the target distribution $\pi_\lambda^{\mathcal{C}}$ as $T \to \infty$. □

3.5 Gibbs Sampling

In the previous section we saw how continuous-time spatial birth-death processes can be employed to sample conditional Boolean models. It turns out that one can also use a discrete-time Gibbs sampler for the same sampling problem. To the best of our knowledge a Gibbs sampler for conditional Boolean models has not yet been considered in the simulation literature. The exact coupling construction presented here bears a loose resemblance to the Häggström *et al.* (1999) algorithm for the bivariate penetrable spheres model. Their method (Algorithm 4.5) is described in Section 4.4, and can be used to perfectly sample the area-interaction process. For point processes there is atypically no maximal state; however they introduce *quasi-minimal* and *quasi-maximal* elements in order to carry out perfect sampling via Monotone CFTP.

Recall that $\mathcal{C} = \{c_1, \ldots, c_k\}$ denotes the finite set of conditioning nodes and for each non-empty $A \subseteq \mathcal{C}$ the region E_A, defined by Eq. (3.4), is such that a germ in E_A covers only those conditioning nodes in A and no others. The (conditional) distribution of the conditional Boolean model on some E_A, given the process in the other regions $\{E_B;\ B \neq A\}$, is just Poissonian. Hence devising Gibbs sampler for the conditional Boolean model is straightforward. Due to the sequential nature of a Gibbs updating scheme it is convenient to consider the set $\{A;\ A \subseteq \mathcal{C},\ A \neq \emptyset\}$ as an ordered set $\{A_1, \ldots, A_N\}$ so that each $i \in \{1, \ldots, N\}$ corresponds uniquely to some A, where N is the number of non-empty $A \subseteq \mathcal{C}$.

If $X \sim$ Poisson(λ), X_i its restriction to E_{A_i} and $\tilde{X} = \left(\tilde{X}_1, \ldots, \tilde{X}_N\right)$ a collection of Poisson processes with $\mathcal{U}\left(\tilde{X}\right) \supset \mathcal{C}$, then $\tilde{X} \sim \pi_\lambda^{\mathcal{C}}$. Consider a discrete-time component process $\tilde{\Phi} = \left(\tilde{\Phi}_0, \ldots, \tilde{\Phi}_N\right)$ with equilibrium distribution $\pi_\lambda^{\mathcal{C}}$. Let the updating order of the components $\tilde{\Phi}_i$ be $\{1, \ldots, N\}$ and define

$$\tilde{\Phi}_{-i}(n) = \bigcup_{j<i} \tilde{\Phi}_j(n) \bigcup_{j>i} \tilde{\Phi}_j(n-1); \qquad \chi_i(n) = \mathbf{1}_{\left\{\mathcal{U}\left(\tilde{\Phi}_{-i}(n)\right) \supset A_i\right\}}.$$

Denote the distribution of an unconditional Poisson(λ) process by π_λ and that conditioned to have at least one germ by π'_λ. The density of a Poisson(λ) process conditioned to cover $\mathcal{C}$ is given by $f(x) \propto \lambda^{n(x)} \mathbf{1}_{\{\mathcal{U}(x) \supset \mathcal{C}\}}$. So the density of the process restricted to the 'cell' E_{A_i} is given by $f_i(x) \propto \lambda^{n(x)} \mathbf{1}_{\{\mathcal{U}(x) \supset A_i\}}$, since a germ in E_{A_i} only covers A_i and the processes in distinct regions E_{A_i} and E_{A_j} are independent. Therefore, given $\left\{\tilde{\Phi}_{-i}\right\}$, the conditional density of $\tilde{\Phi}_i$ is given by

$$f_i\left(\cdot \mid \tilde{\Phi}_{-i}\right) \propto \begin{cases} \lambda^{n(\cdot)} & \text{if } \chi_i = \mathbf{1}_{\left\{\mathcal{U}\left(\tilde{\Phi}_{-i}\right) \supset A_i\right\}} = 1; \\ \lambda^{n(\cdot)} \mathbf{1}_{\{n(\cdot)>0\}} & \text{if } \chi_i = 0. \end{cases} \tag{3.9}$$

Thus, given $\tilde{\Phi}_{-i}$, the next update of $\tilde{\Phi}_i$ is (i) $X_i \sim \pi_\lambda$ on E_{A_i} if $\chi_i = 1$; or (ii) $X'_i \sim \pi'_\lambda$ if $\chi_i = 0$. The Gibbs sampler for the $\left\{\tilde{\Phi}_i\right\}_{i=1}^N$ with updating order $\{1, \ldots, N\}$ is summarized as follows.

Algorithm 3.5 (Gibbs: Conditional Boolean Model).

for $i = 1, \ldots, N$: initialize $\tilde{\Phi}_i(0) = \{\xi_i\}$, where $\xi_i \in E_{A_i}$.

for $n = 1, 2, \ldots$:

for $i = 1, \ldots, N$:

if $\chi_i(n) = 1$: draw $X_i \sim \pi_\lambda$ on E_{A_i}; set $\tilde{\Phi}_i(n) = X_i$.

else if $\chi_i(n) = 0$: draw $X'_i \sim \pi'_\lambda$ on E_{A_i}; set $\tilde{\Phi}_i(n) = X'_i$.

3.5.1 Exact Gibbs Sampler

In this section a perfect variant of the Gibbs sampler (Algorithm 3.5) is presented. The construction here has been motivated by that of Häggström *et al.* (1999) who present an exact 2-component Gibbs sampler for the bivariate penetrable spheres model. In order to devise a CFTP-based procedure one requires maximal and minimal elements of the state space to exist. For point processes a maximal element does not exist since the state space is uncountable. However Häggström *et al.* (1999) go round this problem by defining *quasi-minimal* and *quasi-maximal* elements in order to carry out perfect sampling via Monotone CFTP. In the spirit of this we define such quasi-minimal and -maximal elements for each of the N regions or cells E_{A_i}, and devise a CFTP-based construction that outputs samples with the required target distribution.

Consider an N-tuple of spatial point configurations $x = (x_1, \ldots, x_N)$, where x_i is a configuration on cell E_{A_i}. Define $I_x = \left(\mathbf{1}_{\{x_1 \neq \emptyset\}}, \ldots, \mathbf{1}_{\{x_N \neq \emptyset\}}\right)$ and a relation $\preccurlyeq$ by $x \preccurlyeq y$ if $I_x \leq I_y$, where the inequality is interpreted component-wise. Let $\underline{x} = (\emptyset, \ldots, \emptyset)$ be an N-tuple of empty configurations, so that $I_{\underline{x}} = (0, \ldots, 0)$; and $\overline{x} = (\{\xi_1\}, \ldots, \{\xi_N\})$, with $\xi_i \in E_{A_i}$, so that $I_{\overline{x}} = (1, \ldots, 1)$. Refer to $\underline{x}$ and $\overline{x}$ as *quasi-minimal* and *quasi-maximal* elements respectively. Denote by $\tilde{\Phi}^{T,min} = \left(\tilde{\Phi}_1^{T,min}, \ldots, \tilde{\Phi}_N^{T,min}\right)$ and $\tilde{\Phi}^{T,max} = \left(\tilde{\Phi}_1^{T,max}, \ldots, \tilde{\Phi_N}^{T,max}\right)$ the minimal, respectively, maximal processes started at time $-T$; set

$$\tilde{\Phi}_{-i}^{T,min}(n) = \bigcup_{j<i} \tilde{\Phi}_j^{T,min}(n) \bigcup_{j>i} \tilde{\Phi}_j^{T,min}(n-1); \qquad \chi_i^{T,min}(n) = \mathbf{1}_{\left\{\mathcal{U}\left(\tilde{\Phi}_{-i}^{T,min}(n)\right) \supset A_i\right\}}. \tag{3.10}$$

$$\tilde{\Phi}_{-i}^{T,max}(n) = \bigcup_{j<i} \tilde{\Phi}_j^{T,max}(n) \bigcup_{j>i} \tilde{\Phi}_j^{T,max}(n-1); \qquad \chi_i^{T,max}(n) = \mathbf{1}_{\left\{\mathcal{U}\left(\tilde{\Phi}_{-i}^{T,max}(n)\right) \supset A_i\right\}}. \tag{3.11}$$

Furthermore for $n = 0, -1, -2, \ldots$ and $i = 1, \ldots, N$ let $(X(n,i), X'(n,i))$ denote a tuple of Poisson processes on E_{A_i} such that $X'(n,i)$ is conditioned to contain at least one germ and $X(n,i) \subseteq X'(n,i)$. It is relatively straightforward to simulate such a tuple:

Algorithm 3.6 (Simulation of $(X(n,i), X'(n,i))$).

Set $t = 1$; draw $Y_i \sim$ Poisson(λ) on E_{A_i}.

if $Y_i \neq \emptyset$: set $X(n,i) = X'(n,i) = Y_i$; $t = 0$

else: set $X(n,i) = \emptyset$.

while $t \neq 0$:

draw $Y_i \sim$ Poisson(λ) on E_{A_i}.

if $Y_i \neq \emptyset$: set $X'(n,i) = Y_i$; $t = 0$.

else: set $t = t + 1$.

return $(X(n,i), X'(n,i))$.

Denoting the integer part of $\frac{T}{2}$ by $\lfloor \frac{T}{2} \rfloor$, the exact Gibbs sampler is summarized below.

Algorithm 3.7 (Exact Gibbs: Conditional Boolean Model).

Fix $T > 0$.

while $T > 0$:

for $i = 1, \ldots, N$:

construct realizations of $(X(n,i), X'(n,i))$ for $n \in \left\{-T+1, \ldots, -\lfloor \frac{T}{2} \rfloor\right\}$.

Initialize $\tilde{\Phi}^{T,min}(-T) = \underline{x}$, $\tilde{\Phi}^{T,max}(-T) = \overline{x}$.

for $n \in \{-T+1, \ldots, 0\}$:

for $i \in \{1, \ldots, N\}$:

if $\chi_i^{T,max}(n) = 1$: set $\tilde{\Phi}_i^{T,min}(n) = X(n,i)$.

else: set $\tilde{\Phi}_i^{T,min}(n) = X'(n,i)$.

if $\chi_i^{T,min}(n) = 1$: set $\tilde{\Phi}_i^{T,max}(n) = X(n,i)$.

else: set $\tilde{\Phi}_i^{T,max}(n) = X'(n,i)$.

if $\tilde{\Phi}_i^{T,min}(0) = \tilde{\Phi}_i^{T,max}(0)$ for all $i \in \{1, \ldots, N\}$ (ie. coalescence): set $T = -1$.

else: set $T = 2T$.

return $\tilde{\Phi}^{T,min}(0)$

As noted in Algorithm 3.4, when extending the $(X(n,i), X'(n,i))$ backwards over some interval in one iteration of the algorithm, it is vital to re-use the same realizations for all subsequent iterations.

Properties of the Exact Gibbs Sampler

Algorithm 3.7 is now shown to satisfy the following relationships, which are analogues of the results for the Cai & Kendall Algorithm 3.4. The following Lemmas will be used to prove that its output is indeed a conditional Boolean model (Theorem 3.2).

Lemma 3.6 (Sandwiching). *The processes $\tilde{\Phi}^{T,min}$ and $\tilde{\Phi}^{T,max}$ of Algorithm 3.7 satisfy*

$$\tilde{\Phi}_i^{T,min}(n) \preccurlyeq \tilde{\Phi}_i^{T,max}(n), \qquad \text{for all } i \in \{1, \ldots, N\}, n \in \{-T, \ldots, 0\}. \tag{3.12}$$

Proof. The initialization procedure ensures that Eq. (3.12) holds at time $-T$. Suppose that this relationship holds at some time $n-1 > -T$. The tuple $(X(n,i), X'(n,i))$ is such that $X(n,i) \subseteq X'(n,i)$ for all $i = 1, \ldots, N$ and $n = 0, -1, \ldots$ (cf. Algorithm 3.6). The update at time n for the upper process is determined by the current state of the lower process and vice versa; this 'cross-over' ensures that: $\tilde{\Phi}_i^{T,max}$ is set equal to $X(n,i)$ only if $\tilde{\Phi}_i^{T,min}$ is; and $\tilde{\Phi}_i^{T,min}$ is set equal to $X'(n,i)$ only if $\tilde{\Phi}_i^{T,max}$ is. The result then follows via induction along the sequence of times $\{-T, \ldots, 0\}$, since the maximal and minimal processes are initialized so as to satisfy Eq. (3.12). □

Lemma 3.7 (Coalescence). *If* $\tilde{\Phi}^{T,min}(n^*) = \tilde{\Phi}^{T,max}(n^*)$ *then* $\tilde{\Phi}^{T,min}(n) = \tilde{\Phi}^{T,max}(n)$ *for all* $n \in \{n^*, \ldots, 0\}$.

Proof. At time n^* there is no longer any distinction between the upper and lower processes. So from time n^* the updates for $\tilde{\Phi}^{T,min}$ and $\tilde{\Phi}^{T,max}$ will be identical for each of the N components. □

Lemma 3.8 (Funnelling). *For all* $-S \leq -T \leq n \leq 0$ *and* $i = 1, \ldots, N$

$$\tilde{\Phi}_i^{T,min}(n) \preccurlyeq \tilde{\Phi}_i^{S,min}(n) \preccurlyeq \tilde{\Phi}_i^{S,max}(n) \preccurlyeq \tilde{\Phi}_i^{T,max}(n). \tag{3.13}$$

Proof. By definition of '$\preccurlyeq$', if $x = (x_1, \ldots, x_N)$, $y = (y_1, \ldots, y_N)$ and $x_i \subseteq y_i$ for all $i = 1, \ldots, N$ then $x \preccurlyeq y$. The initializing procedure ensures that Eq. (3.13) holds at time $-T$. Suppose that this relationship holds up time $n-1 > -T$; consider updating the first component at time n.

By virtue of Eq. (3.13) holding at time $n-1$, $\tilde{\Phi}_i^{T,min}(n-1) \subseteq \tilde{\Phi}_i^{S,min}(n-1)$. Therefore if $\chi_1^{T,min}(n) = 1$ then $\chi_1^{S,min}(n) = 1$, hence the 'cross-over' trick ensures that $\tilde{\Phi}_1^{T,max}(n)$ will be set equal to $X(n,1)$ only if $\tilde{\Phi}_1^{S,max}(n)$ also is. Similarly, $\tilde{\Phi}_1^{T,min}(n)$ is set equal to $X'(n,1)$ only if $\tilde{\Phi}_1^{S,min}(n)$ also is. Either way

$$\tilde{\Phi}_1^{T,min}(n) \subseteq \tilde{\Phi}_1^{S,min}(n) \subseteq \tilde{\Phi}_1^{S,max}(n) \subseteq \tilde{\Phi}_1^{T,max}(n)$$

where $\tilde{\Phi}_1^{T,min}(n) \subseteq \tilde{\Phi}_1^{T,max}(n)$ is established by Lemma 3.6. Induction along $i = 1, \ldots, N$ gives

$$\tilde{\Phi}_i^{T,min}(n) \subseteq \tilde{\Phi}_i^{S,min}(n) \subseteq \tilde{\Phi}_i^{S,max}(n) \subseteq \tilde{\Phi}_i^{T,max}(n), \qquad \text{for all } i = 1, \ldots, N.$$

The results follows by induction along the sequence of times $\{-T, \ldots, 0\}$. □

Lemma 3.9. *For all sufficiently large* T, $\tilde{\Phi}^{T,min}$ *and* $\tilde{\Phi}^{T,max}$ *coalesce on* $\{-T, \ldots, 0\}$.

Proof. If $X(n,i) = X'(n,i)$, for all i and some $n < 0$, then, for $T > -n$, $\tilde{\Phi}_i^{T,min}(n) = \tilde{\Phi}_i^{T,max}(n)$; moreover $\tilde{\Phi}_i^{T,min}(0) = \tilde{\Phi}_i^{T,max}(0)$ for all i and hence the algorithm terminates. The event that $X(n,i) = X'(n,i)$ has probability $\left(1 - e^{-\lambda m_2[E_{A_i}]}\right)$ (cf. Algorithm 3.6), with m_2 denoting Lebesgue measure. Thus

$$\mathbb{P}\left[X(n,i) = X'(n,i) \text{ for all } i\right] = \prod_i \left(1 - e^{-\lambda m_2[E_{A_i}]}\right) > 0$$

and independent of n. The event $\{X(n,i) = X'(n,i) \text{ for all } i\}$ is independent of events in the past. So, applying the second Borel-Cantelli Lemma, $\{X(n,i) = X'(n,i) \text{ for all } i\}$ happens almost surely for infinitely many n. The algorithm hence terminates in finite time. □

Theorem 3.2. *Algorithm 3.7 terminates in finite time, and the output of the algorithm has distribution $\pi_\lambda^{\mathcal{C}}$, the distribution of a Poisson process of intensity λ conditioned to cover the set $\mathcal{C}$.*

Proof. Lemma 3.9 establishes the first part. Lemmas 3.6, 3.7 & 3.8 and the arguments of Theorem 3.1 then completes the proof. □

3.6 Implementational Issues

In light of the simulation results presented in Section 3.7 the exact Gibbs Algorithm 3.7 is always outperformed by either the 2-Stage Algorithm 3.2 or the Cai & Kendall Algorithm 3.4. This poses a natural question as to whether the implementation of the Gibbs algorithm could be made more efficient. That is, is the relative poor performance due to the way it is implemented or is the algorithm intrinsically inefficient? The purpose of this exposition is to describe the implementation of the Gibbs algorithm, and to show that any avoidable computational costs have been dealt with, hence suggesting that its performance is intrinsically poor. Recall the details of Algorithm 3.7 from Section 3.5. The main stages of the algorithm can be summarized as follows.

Given $T > 0$ the first stage involves extending the basic driving processes $(X(n,i), X'(n,i))$ for $n \in \left\{-T+1, \ldots, -\lfloor \frac{T}{2} \rfloor\right\}$ and $i = 1, \ldots, N$. Here $X(n,i)$ is an unconditional Poisson(λ) process on E_{A_i} and $X'(n,i)$ is one conditioned to contain a positive number of germs, such that $X(n,i) \subseteq X'(n,i)$. Simulation of the $(X(n,i), X'(n,i))$ is unavoidable since they are the basic driving processes. The easiest way to sample $\{X(n,i)\}_{i=1}^N$ is to sample a Poisson(λ) process X on

$\mathcal{U}(\mathcal{C})$ and set $X(n,i) = \{\xi \in X;\ \xi \in E_{A_i}\}$. If $X(n,i) \neq \emptyset$ then setting $X'(n,i) = X(n,i)$ yields the required tuple. Conversely if $X(n,i) = \emptyset$ then the objective is to sample a non-empty Poisson(λ) process on E_{A_i}. Rather than doing this (as it is much likely to be computationally burdensome) set $X'(n,i)$ equal to a 'dummy' non-empty configuration. The dynamics of the algorithm will not be affected by this, since all that matters is that the *number* of germs in $X'(n,i)$ is positive rather than their locations.

The next stage involves initializing the maximal and minimal processes $\Phi^{T,max}, \Phi^{T,min}$ and evolving them forwards in time, coupled to realizations of the $(X(n,i), X'(n,i))$ processes. For each n and i the algorithm has to determine which of the processes $X(n,i)$ or $X'(n,i)$ will be the next update for the maximal and minimal processes. This stage is going to be computationally quite expensive; however this cost is unavoidable since the $\Phi^{T,max}$ and $\Phi^{T,min}$ must be updated sequentially up to time 0. Finally, after coalescence has occurred, it needs to be determined which of the processes $\Phi_i^{T,min}(0)$ (or equivalently, $\Phi_i^{T,max}(0)$ since coalescence has occurred) take on the 'dummy' configurations. For all such processes realizations of non-empty Poisson(λ) processes must be simulated. However this is a one-off expense which will not affect the run times significantly. Therefore, stage-by-stage, any computational costs have been minimized and the current implementation of Algorithm 3.7 is as efficient as possible.

3.7 Simulation Results

In this section we present some simulation results, the objective of which is to compare the performance of the three perfect simulation algorithms: 2-Stage Rejection (Algorithm 3.2), the Cai & Kendall (Algorithm 3.4) and Exact Gibbs (Algorithm 3.7). The performance will be measured by the actual run times (in seconds) of the algorithms. A series of simulation experiments is carried out and the log mean empirical run times are plotted against the parameters λ (intensity) and k (number of conditioning nodes).

It is vital to carefully design the experiments in order to ensure that the comparisons made are meaningful. For example, when comparing how run times vary with the intensity λ, the 'structure' of the conditioning set $\mathcal{C}$ will play an important role. By 'structure' of $\mathcal{C}$ we refer to, among other aspects, the spatial distribution or 'spread' of the conditioning nodes, eg. randomly/regularly dis-

tributed, close/far apart, etc. In order to formalize this structure some notation is first introduced; description of the experiments will then use this to enable a realistic evaluation of the algorithms.

The simulations were carried out on a PC (Pentium 4 2.67GHz, 248MB RAM) running Windows XP; the implementations of the algorithms were programmed in `Python` (version 2.3).

3.7.1 Notation

The conditioning set $\mathcal{C} = \{c_1, \ldots, c_k\}$ consists of k nodes in $\mathbb{R}^2$. The Boolean model $\mathcal{U}(x)$ associated with a point configuration x on $\mathbb{R}^2$ is obtained as $\mathcal{U}(x) = \bigcup_{\xi \in x} B_r(\xi)$ the union of disks of radius r centred at each point $\xi \in x$. Let $c_i = (c_i^1, c_i^2)$ and $[w_{lo}^1, w_{hi}^1] \times [w_{lo}^2, w_{hi}^2]$ be the smallest rectangle that covers $\mathcal{C}$; then

$$w_{lo}^j = \min_{1 \le i \le k} \left\{c_i^j\right\} \qquad \text{and} \qquad w_{hi}^j = \max_{1 \le i \le k} \left\{c_i^j\right\}.$$

Define $W_\delta = [w_{lo}^1 - \delta, w_{hi}^1 + \delta] \times [w_{lo}^2 - \delta, w_{hi}^2 + \delta]$. It follows that W_r is the smallest rectangle that covers $\mathcal{U}(\mathcal{C})$. In all the simulations, the sampling window is taken to be W_r. As noted in Section 3.2, for fixed radius r, the set $\mathcal{C}$ can be viewed as a graph with nodes c_i and edges determined as:

Definition 3.3. Two conditioning nodes c_i and c_j are *connected*, ie. there is an edge between nodes c_i and c_j if $\mathcal{U}(c_i) \cap \mathcal{U}(c_j) \neq \emptyset$, written $c_i \sim_r c_j$; c_i and c_j are *disconnected* if $\mathcal{U}(c_i) \cap \mathcal{U}(c_j) = \emptyset$.

Definition 3.4. A *cluster* or *connected component* of $\mathcal{C}$ is a subset $\mathbf{C} \subseteq \mathcal{C}$ such that every node in $\mathbf{C}$ is connected to at least one other node in $\mathbf{C}$.

If $\mathbf{C}_i$ and $\mathbf{C}_j$ are two distinct clusters of $\mathcal{C}$ then $\mathcal{U}(\mathbf{C}_i) \cap \mathcal{U}(\mathbf{C}_j) = \emptyset$. Moreover if $\mathcal{C}$ is partitioned into clusters $\mathbf{C}_1, \ldots, \mathbf{C}_n$ then $\mathcal{U}(\mathcal{C})$ can be partitioned into $\mathcal{U}(\mathbf{C}_1), \ldots, \mathcal{U}(\mathbf{C}_n)$. Suppose $X = \{\xi_n\}$ is a Poisson process on $\mathcal{U}(\mathcal{C})$ and $\mathcal{U}(X)$ its associated Boolean model. If $\mathbf{C}_i$ and $\mathbf{C}_j$ are two connected components of $\mathcal{C}$ then the germs of X (or equivalently, the grains of $\mathcal{U}(X)$) will be independent in either of the disjoint regions $\mathcal{U}(\mathbf{C}_i)$ and $\mathcal{U}(\mathbf{C}_j)$. Moreover, if $\mathcal{C} = \mathbf{C}_1 \cup \ldots \cup \mathbf{C}_n$ and $\tilde{X}$ is a Poisson process whose associated Boolean model covers $\mathcal{C}$, then $\mathcal{U}\left(\tilde{X}\right)$ can be partitioned into an irreducible collection of Boolean models $\mathcal{U}\left(\tilde{X}_i\right)$, each covering the respective cluster $\mathbf{C}_i$ of $\mathcal{C}$. This means that if $\mathcal{C}$ consists of several clusters then it would be logical and more efficient to draw a conditional Boolean model for each cluster *independently* and then take the union in order to obtain

a conditional Boolean model for $\mathcal{C}$ (cf. Section 3.3.1). Therefore it suffices to restrict attention to the case when $\mathcal{C}$ constitutes a single cluster.

Definition 3.5. A *clique* of the graph $\mathcal{C}$ is a subset $A \subseteq \mathcal{C}$ containing the maximum number of nodes which are all connected to each other. Furthermore, if $|A|$ denotes the number of nodes in A, then define $I(\mathcal{C}) = \max_{\{\text{cliques } A \subseteq \mathcal{C}\}} |A|$ to be the *maximum clique size*.

Abbreviate the 2-Stage Rejection Algorithm 3.2 by '2Stg', the Cai & Kendall Algorithm 3.4 by 'CK' and the Exact Gibbs Algorithm 3.7 by 'Gibbs' for convenience. The dynamics of both CK and Gibbs depend on the regions $\{E_{A_1}, \ldots, E_{A_N}\}$, since these determine the number of upper and lower processes in either algorithm. High values of N will affect coalescence time, and hence the empirical run times. Moreover the maximum clique size $I(\mathcal{C})$ affects the value of N; the higher $I(\mathcal{C})$ is the larger the value of N will be and vice versa (cf. Figure 3.2 below). More is said about this in Experiment 2 (Section 3.7.4).

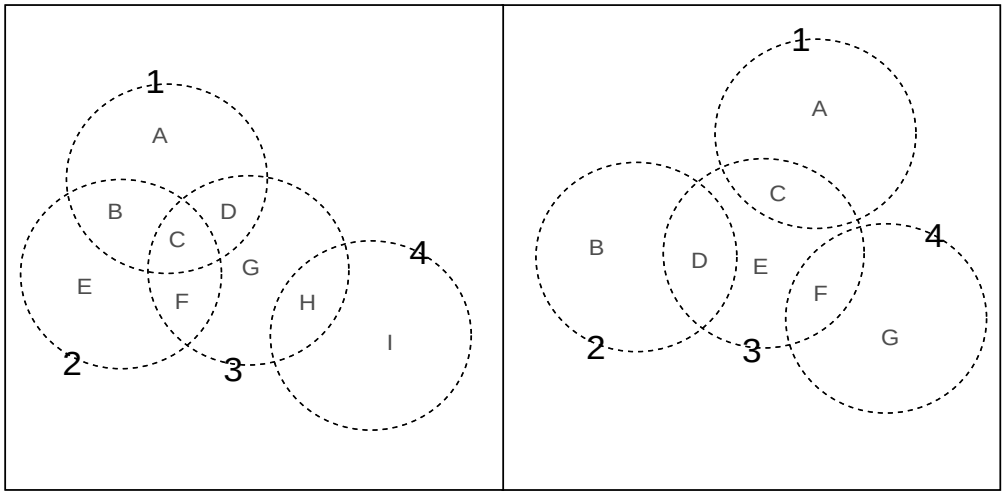

Figure 3.2: Illustration of the maximum clique size and how it influences the value of N. Suppose that $\mathcal{C} = \{c_1, c_2, c_3, c_4\}$, where c_j is the centre of the disk labelled j, for $j = 1, 2, 3, 4$. *Left:* Nodes $\{c_1, c_2, c_3\}$ are all connected to each other and so form a clique; similarly $\{c_3, c_4\}$ form another clique. Therefore $I(\mathcal{C}) = 3$. The regions $\{A, \ldots, I\}$ represent the cells $\{E_{A_1}, \ldots, E_{A_N}\}$. For example the cell $E_{\{c_2,c_3\}}$ is represented by F and $E_{\{c_1,c_2,c_3\}}$ by C. In this case $N = 9$. *Right:* Here $\mathcal{C}$ still consists of $k = 4$ conditioning nodes. The cliques in this case are $\{c_1, c_3\}$, $\{c_2, c_3\}$ and $\{c_3, c_4\}$; so $I(\mathcal{C}) = 2$ and $N = 7$. Thus, for fixed number of conditioning nodes k, a higher value of $I(\mathcal{C})$ corresponds to a higher value of N and vice versa.

3.7.2 Simulation Experiments

In Section 3.3.1 it was commented that 2Stg was much more efficient that ordinary Rejection sampling (Algorithm 3.1). Therefore a simple illustration of this is first presented below; it is hoped to convince the reader that the modifications made to the Rejection algorithm which yield 2Stg do indeed improve the run times. Figures 3.3 & 3.4 below depict the evolution of the run times for 2Stg, Rejection, CK and Gibbs as the intensity λ and number of nodes k vary. Rejection performs quite poorly compared to CK and Gibbs as λ decrease and k increases. However 2Stg remains competitive to both CK and Gibbs. The basic conclusions of these two illustrations is that 2Stg is a great improvement on Rejection sampling; therefore an investigation on the extent to which 2Stg remains comparable to CK and Gibbs now follows.

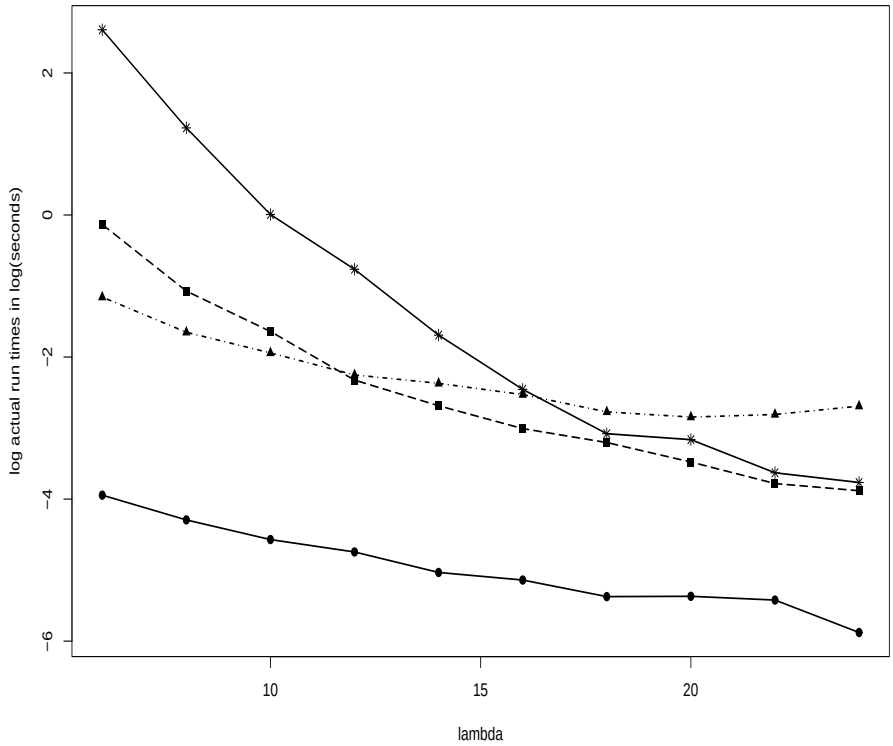

Figure 3.3: Empirical log mean run times for Rejection (*), 2Stg (solid dots), CK (solid triangles) and Gibbs (solid squares). The conditioning set $\mathcal{C} = \{(0,0),(0.15,0),(0.3,0),(0.45,0),(0.6,0),(0.75,0),(0.9,0),(1.05,0),(1.2,0)\}$, radius $r = 0.1$. The x-axis represents the intensity λ.

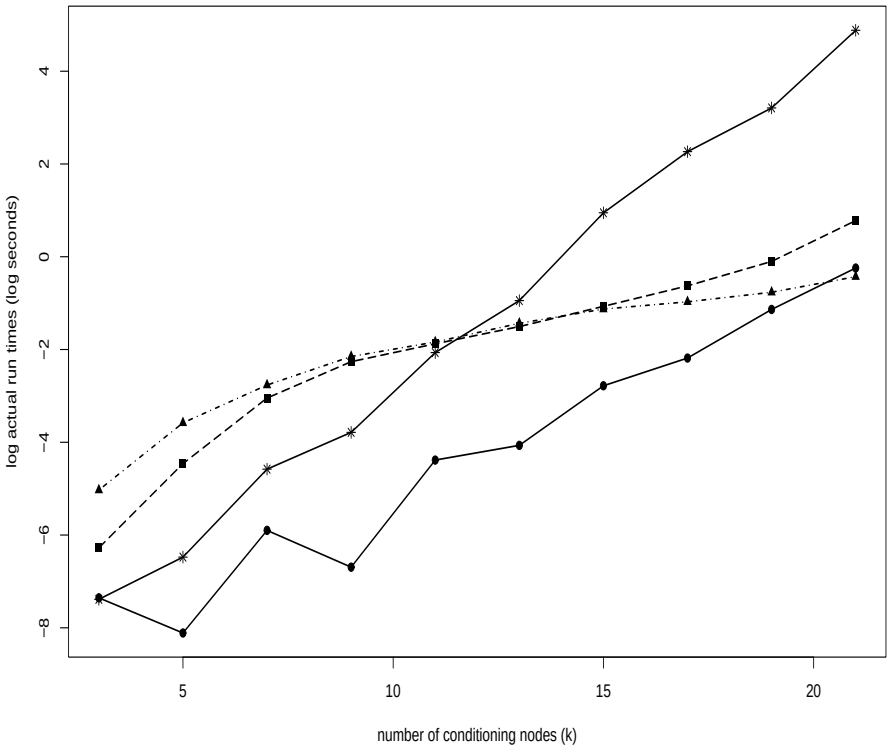

Figure 3.4: Empirical log mean run times for Rejection (*), 2Stg (solid dots), CK (solid triangles) and Gibbs (solid squares). The conditioning set $\mathcal{C}$ comprises of a single cluster of k randomly placed nodes. The intensity $\lambda = 20$ and radius $r = 0.1$. The x-axis represents k, the size of $\mathcal{C}$.

Remark 3.1. In Figure 3.4 the run times for 2Stg do not necessarily increase with k. The set A^*, consisting of the maximum number of disconnected nodes, is determined for each k. Therefore as new conditioning nodes are added to $\mathcal{C}$ the size of A^* may also increase. This in turn may actually lead to lower run times for 2Stg since the probability that Boolean model covers $\mathcal{C}$, given that it covers A^*, is not necessarily decreasing in k.

Remark 3.2. CK is a *local change* algorithm, in the sense that the only transitions are births or deaths; so the current pattern only changes 'locally' at each transition. On the other hand 2Stg and Gibbs are *global change* algorithms, since a completely different pattern on the whole sampling window (or some sub-region) is proposed at each iteration. Intuitively one would expect a local change algorithm to outperform global change ones for very extreme model parameters. The reason being that for extreme parameter values (low λ and/or large k) the density of the conditional Boolean model becomes highly concentrated at certain configurations. Thus the acceptance probability for

global changes is likely to be very low. Conversely, CK involves addition/deletion of single germs only. The birth and death rates are balanced and such local changes are more likely to be accepted. Indeed the following empirical results support this: the CK algorithm performs better than the other two for extreme parameter values. For moderate model parameters, 2Stg does better than CK and Gibbs. The results also indicate that Gibbs is outperformed by either 2Stg or CK for all parameter values. This is however not the case for the conditional area-interaction process, where the Gibbs sampler is just as efficient as a 2-Stage Rejection procedure (cf. Chapter 5, Section 5.7).

3.7.3 Experiment 1: Run Times Versus Intensity λ

In this experiment the conditioning set $\mathcal{C}$ is fixed and the empirical run times of the three algorithms 2Stg, CK and Gibbs are compared as the underlying Poissonian intensity λ changes. Two cases are considered: when (i) $\mathcal{C}$ is a lattice of k regularly spaced nodes; and (ii) $\mathcal{C}$ consists of k randomly distributed nodes. The radius r of the disks consisting the Boolean model is fixed and the sampling window is W_r (cf. Section 3.7.1). Recall the discussion about considering only those $\mathcal{C}$ which consists of a single cluster of k nodes. For the lattice case, the distance between each node is fixed so that the whole grid is a single cluster. In the case of randomly placed conditioning nodes, an accept/reject mechanism produces k uniformly drawn nodes which form a single cluster. For large λ an unconditional Poisson(λ) process is quite likely to cover $\mathcal{C}$, hence 2Stg should have relatively lower run times. As λ decreases the conditioning becomes increasingly stringent, and therefore CK should do better as λ decreases.

Figures 3.5 & 3.6 show the evolution of run times when $\mathcal{C}$ is a lattice of $k = 16$ and $k = 36$ nodes respectively, while Figure 3.7 depicts the case when $\mathcal{C}$ is a single cluster of $k = 15$ randomly distributed nodes. The CK algorithm is fairly complex and the majority of its run time is accounted for by this complexity. However this means that the extra computational effort incurred as λ decreases is likely to be lower for CK than 2Stg or Gibbs. So as λ decreases this added complexity of CK becomes worthwhile to implement. Notice also that Gibbs is always outperformed by the other two algorithms. In fact in all three Figures 3.5, 3.6 & 3.7 the Gibbs algorithm failed to output a single sample in around 7-10 hours for some values of λ. Therefore the respective run times of Gibbs are truncated at some λ due to lack of sufficient samples.

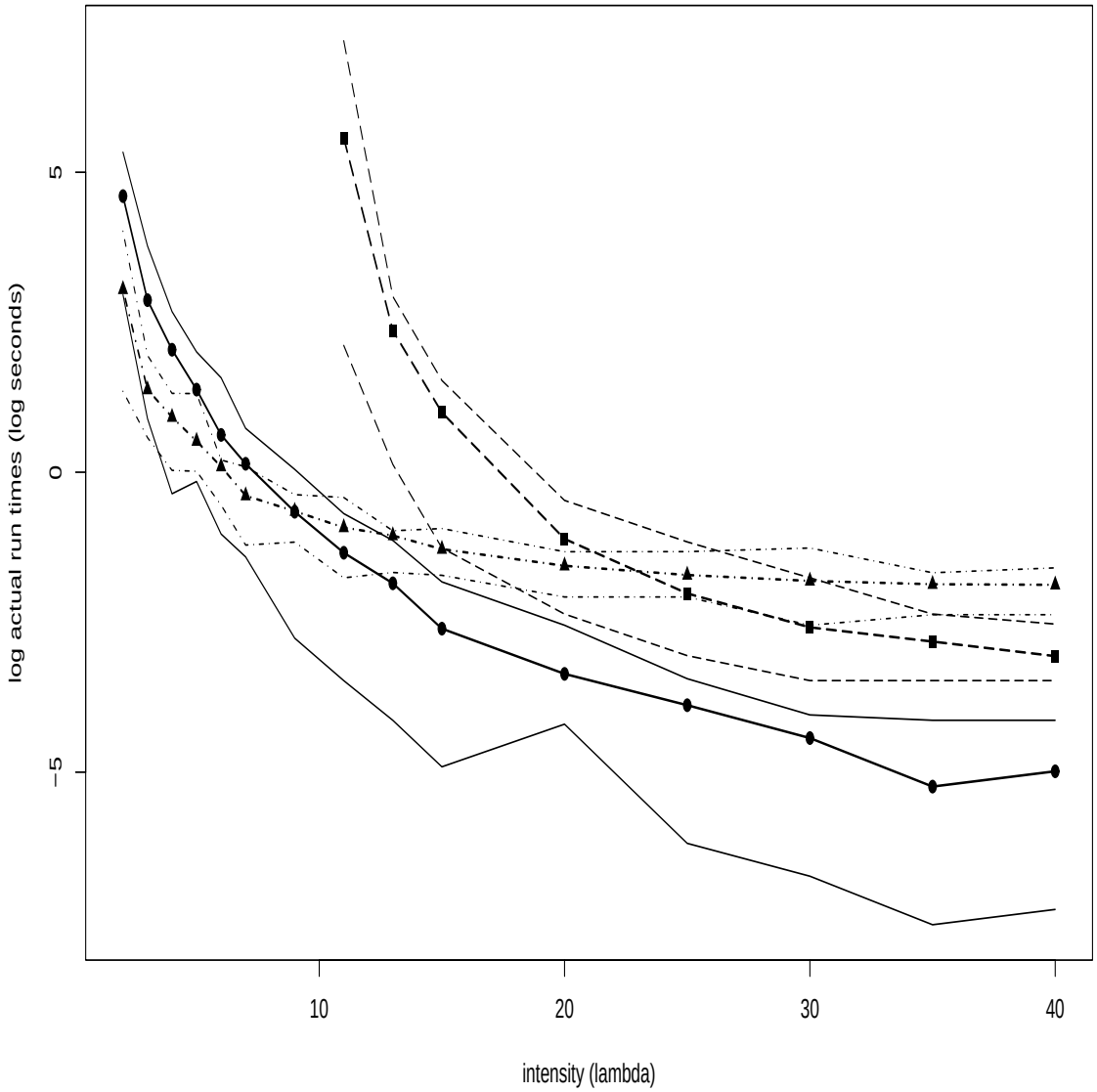

Figure 3.5: Empirical log mean run times for 2Stg (solid dots), CK (solid triangles) and Gibbs (solid squares). The solid (2Stg), dashed-dotted (CK) and dashed (Gibbs) lines represent the upper and lower simulation envelopes computed at the 90% and 10% quantiles respectively. The conditioning set $\mathcal{C}$ is a grid of $k = 16$ regularly placed nodes and the radius $r = 0.12$. The x-axis represents the intensity λ. It took the Gibbs algorithm about 7 hours to output a single sample for $\lambda = 10$; therefore the run times corresponding to Gibbs are plotted only for values down to $\lambda = 11$.

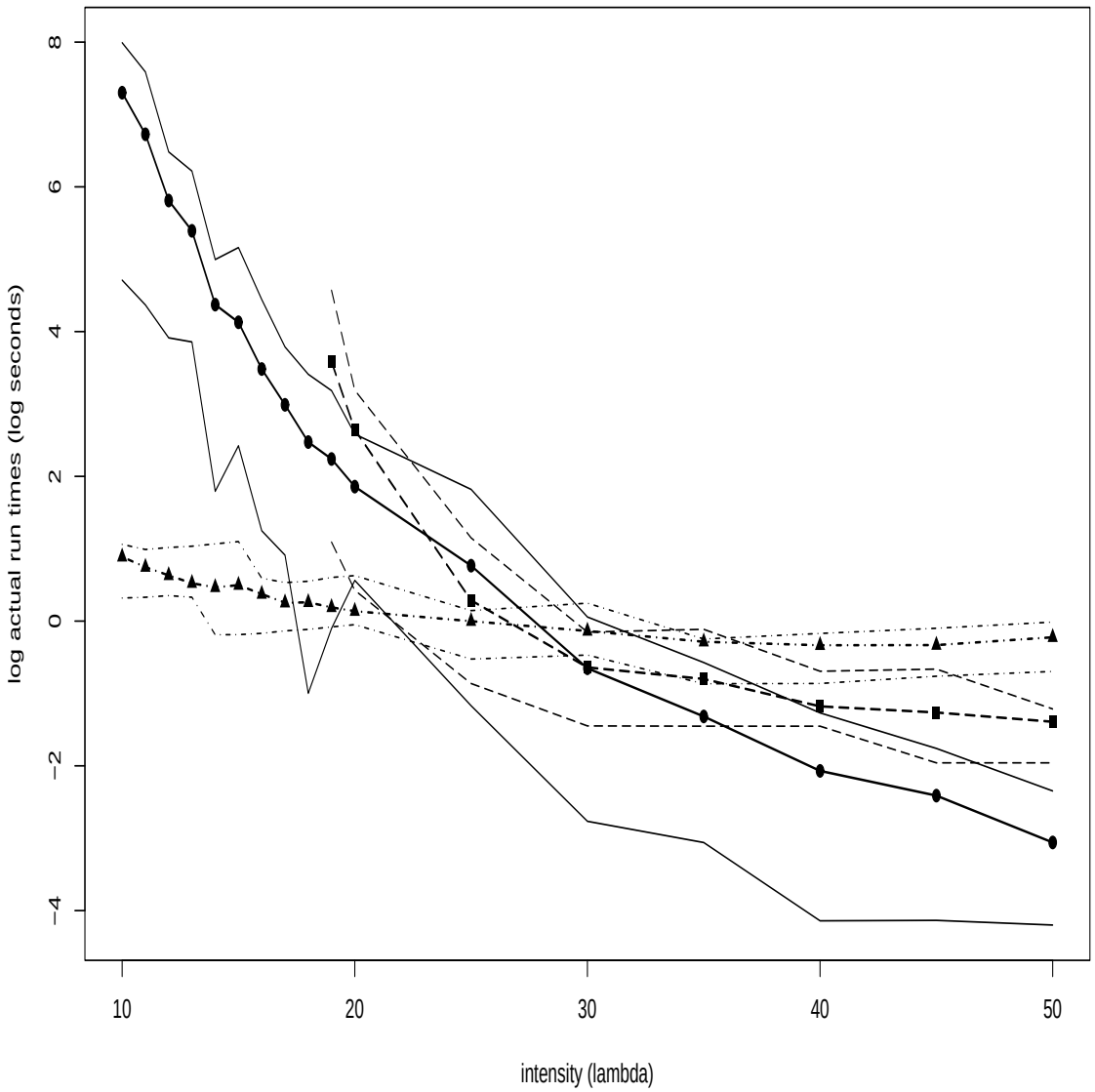

Figure 3.6: Empirical log mean run times for 2Stg (solid dots), CK (solid triangles) and Gibbs (solid squares). The solid (2Stg), dashed-dotted (CK) and dashed (Gibbs) lines represent the upper and lower simulation envelopes computed at the 90% and 10% quantiles respectively. The conditioning set $\mathcal{C}$ is a grid of $k = 36$ regularly placed nodes and the radius $r = 0.12$. The x-axis represents the intensity λ. It took the Gibbs algorithm about 10 hours to output a single sample for $\lambda = 18$; therefore the run times corresponding to Gibbs are plotted only for values down to $\lambda = 19$.

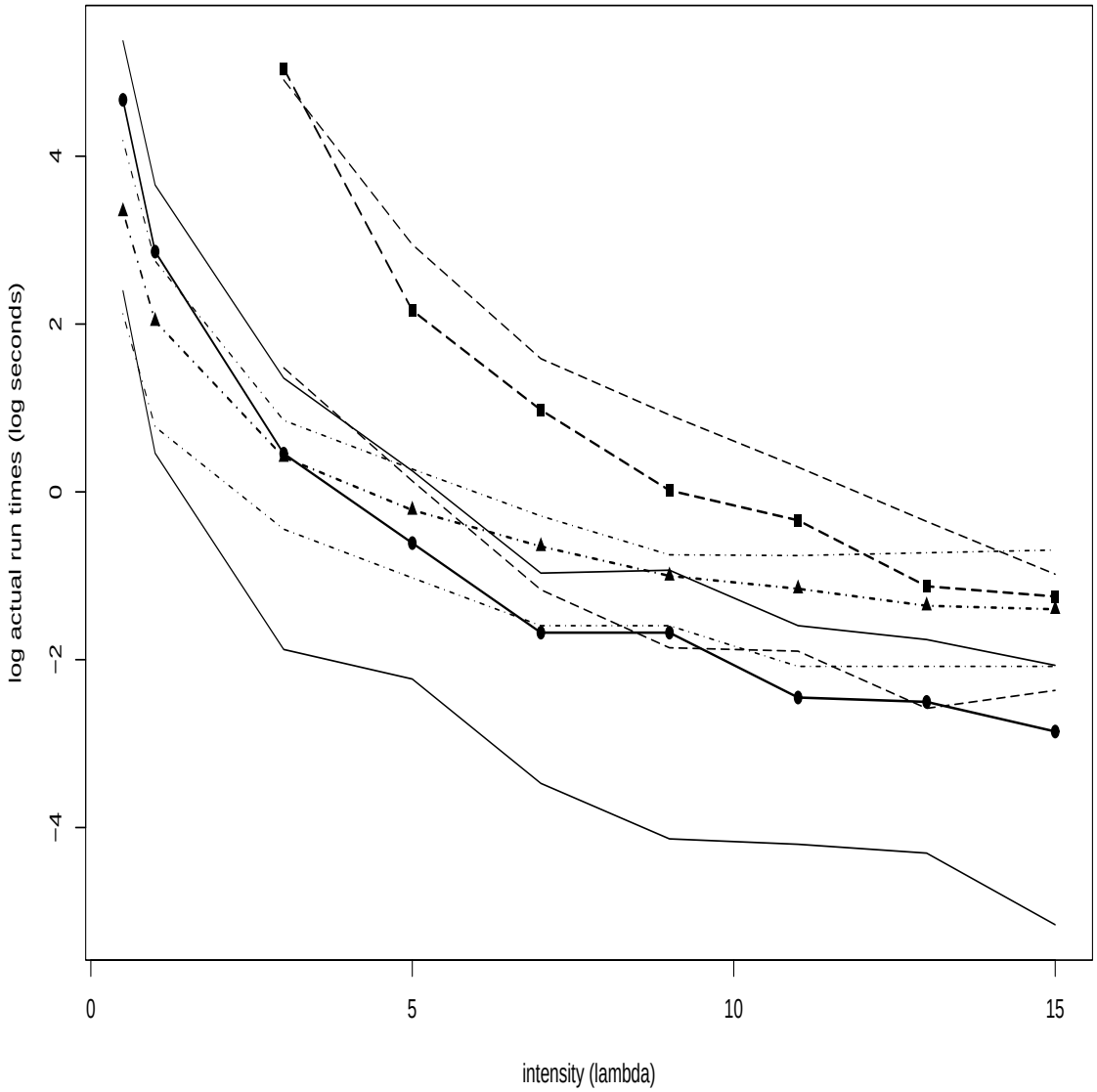

Figure 3.7: Empirical log mean run times for 2Stg (solid line), Gibbs (dashed line) and CK (dashed-dotted line). The solid (2Stg), dashed-dotted (CK) and dashed (Gibbs) lines represent the upper and lower simulation envelopes computed at the 90% and 10% quantiles respectively. The conditioning set $\mathcal{C}$ comprises of a single connected component of $k = 15$ randomly placed nodes; the radius $r = 0.1$. The x-axis represents the intensity λ. It took the Gibbs algorithm about 10 hours to output one sample for $\lambda = 1$, so the run times for Gibbs are plotted only for values down to $\lambda = 3$.

3.7.4 Experiment 2: Run Times Versus k

In this experiment we look at how run times change with k the size of $\mathcal{C}$. Recall the discussion in Section 3.7.1 that the run times of CK and Gibbs are dependent on N, the number of the E_A regions, where E_A (defined by Eq. 3.4) is that region where a germ covers only those nodes in A and no others. The value of N is in turn affected by $I(\mathcal{C})$, the maximum clique size (Definition 3.5, cf. Figure 3.2). It is expected that, for fixed k, the run times for CK and Gibbs should be higher for those $\mathcal{C}$ with higher values of $I(\mathcal{C})$.

For a more realistic comparison evaluation of the run times is restricted, as k varies, to those $\mathcal{C}$ which have the same value of $I(\mathcal{C})$. The value of $I(\mathcal{C})$ is fixed at I^* for all values of k; as k increases randomly drawn nodes are added so as to maintain $I(\mathcal{C}) = I^*$. As k increases the coverage of $\mathcal{C}$ becomes increasingly rare and so the run times of 2Stg should steadily increase with k. Hence 2Stg should perform well for low k whereas CK should fare well as k increases. Figures 3.8 & 3.9 illustrate the run times for $I^* = 2 \,\&\, 3$ respectively. For small k 2Stg is much more competitive than CK and Gibbs and for larger k CK does better than the other two. In addition observe also that the run times of Gibbs are always higher than either 2Stg or CK.

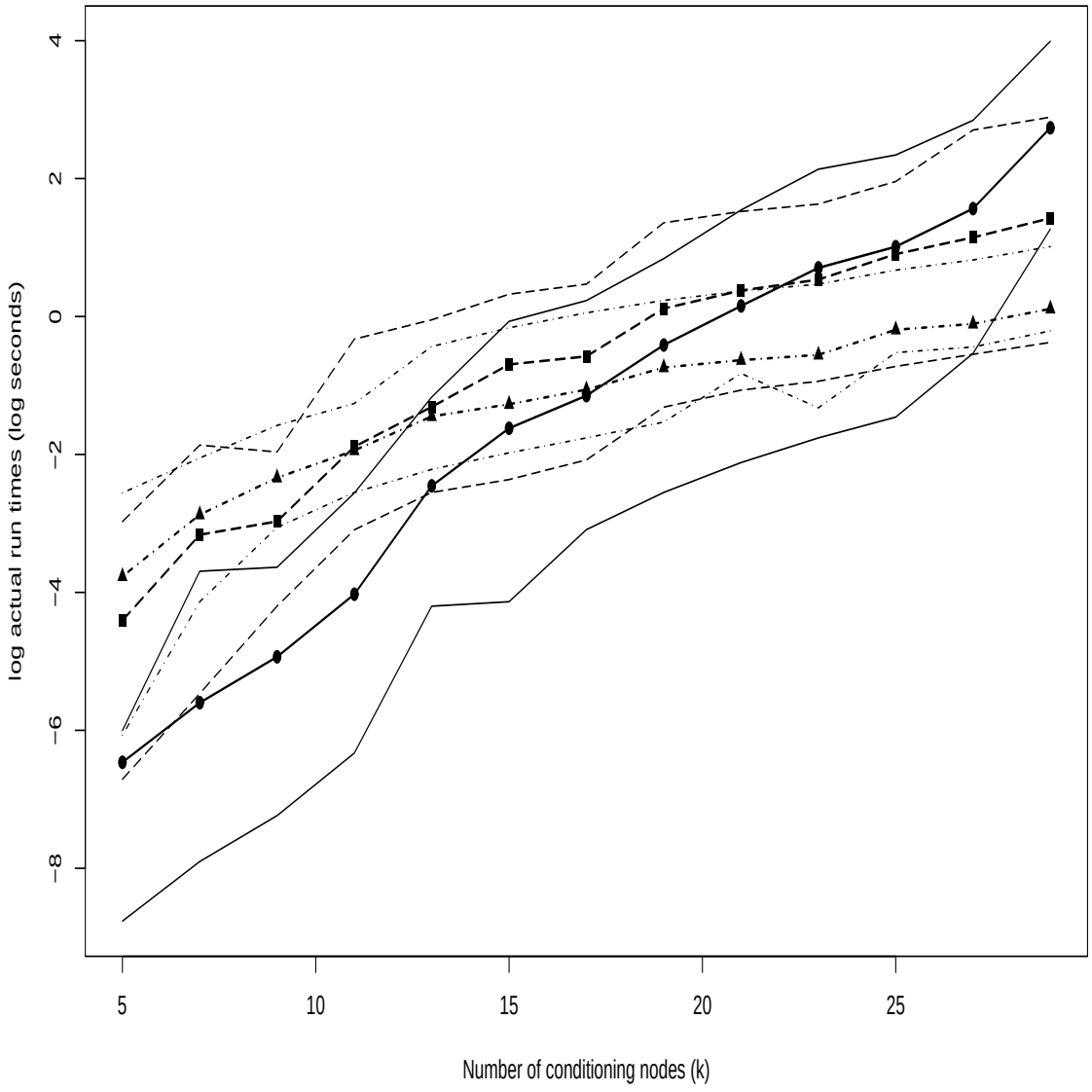

Figure 3.8: Empirical log mean run times for 2Stg (solid line), Gibbs (dashed line) and CK (dashed-dotted line). The solid (2Stg), dashed-dotted (CK) and dashed (Gibbs) lines represent the upper and lower simulation envelopes computed at the 90% and 10% quantiles respectively. The conditioning set $\mathcal{C}$ consists of a single cluster of k nodes, with $I(\mathcal{C})$ fixed for all k at $I^* = 2$. The radius $r = 0.08$ and intensity $\lambda = 20$. The x-axis represents k, the size of $\mathcal{C}$.

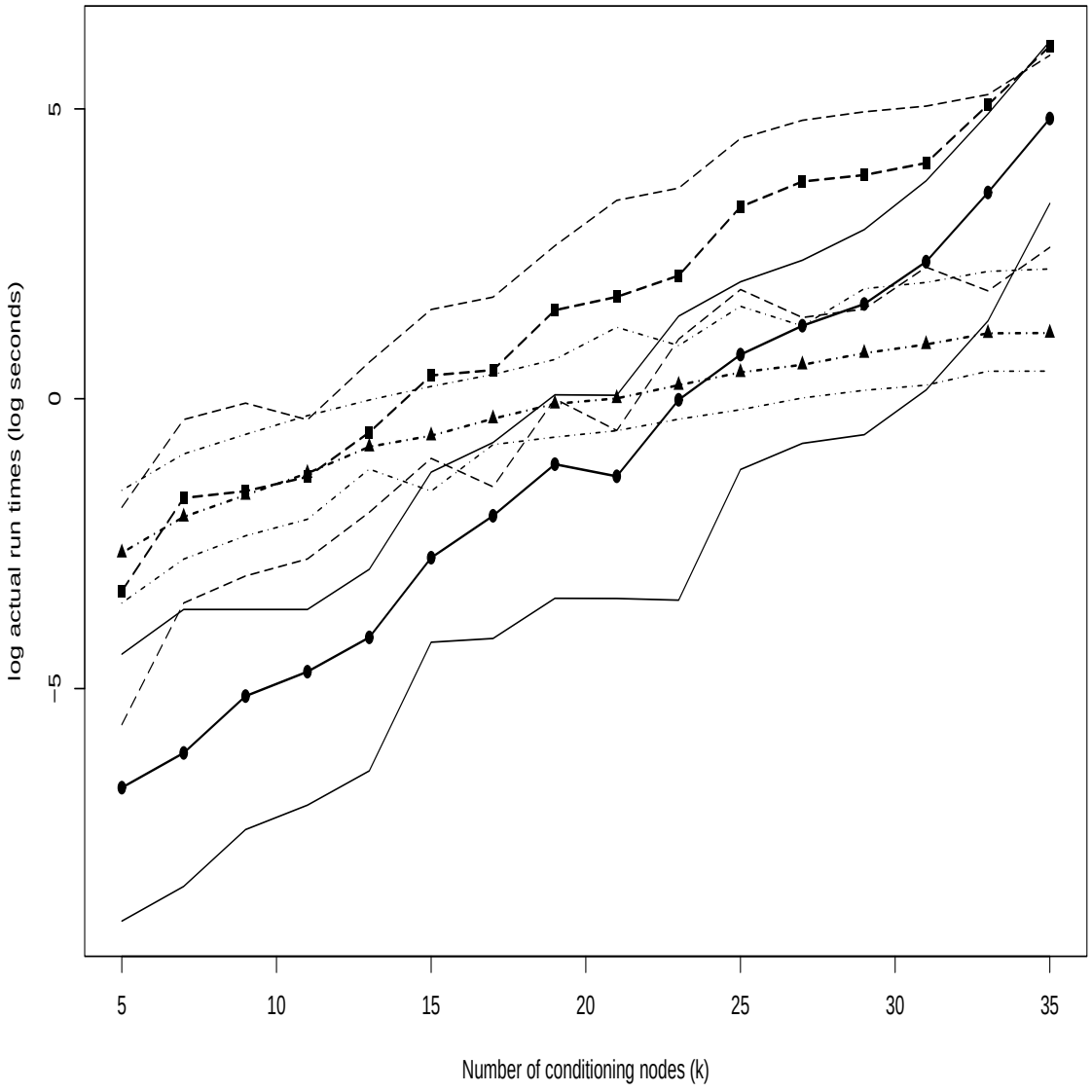

Figure 3.9: Empirical log mean run times for 2Stg (solid line), Gibbs (dashed line) and CK (dashed-dotted line). The solid (2Stg), dashed-dotted (CK) and dashed (Gibbs) lines represent the upper and lower simulation envelopes computed at the 90% and 10% quantiles respectively. The conditioning set $\mathcal{C}$ consists of a single cluster of k nodes, with $I(\mathcal{C})$ fixed for all k at $I^* = 3$. The radius $r = 0.08$ and intensity $\lambda = 20$. The x-axis represents k, the size of $\mathcal{C}$.

3.8 Conclusions & Further Work

This chapter has considered the perfect simulation of Boolean models conditioned to cover a finite set $\mathcal{C}$. Three algorithms for sampling from $\pi_\lambda^{\mathcal{C}}$ have been described: the 2-Stage Rejection Algorithm 3.2, the Cai & Kendall Algorithm 3.4 and the exact Gibbs Algorithm 3.7. 2-Stage Rejection is a more efficient variant of ordinary Rejection sampling (Lemma 3.1), and illustrated by Figures 3.3 & 3.4. The description of the Cai & Kendall method is given here in terms of spatial birth-death processes rather than in its original formulation in terms of immigration-death processes on the natural numbers. The Gibbs algorithm was developed in this chapter, where the idea of quasi-minimal and -maximal elements (Häggström *et al.* 1999) permit an exact CFTP-based construction. An evaluation of the three algorithms is carried out by comparing the respective run times against λ or k. The results suggest that for extreme values (low λ, large $\mathcal{C}$) the Cai & Kendall method fares best, while for moderate values 2-Stage Rejection proves most competitive.

An extension to the work here is sampling from $\pi_\lambda^{\mathcal{C}}$ for *uncountable* $\mathcal{C}$. The dynamics of the Cai & Kendall and Gibbs algorithms rely on partitioning $\mathcal{U}(\mathcal{C})$ into the disjoint regions $\{E_A;\ A \subseteq \mathcal{C}\}$, which in turn enables the construction of appropriate upper and lower sandwich processes that bound the target process. For uncountable $\mathcal{C}$, this is not possible; however the principle of the coupling construction still works for such $\mathcal{C}$. The challenge then is to cleverly define/devise maximal and minimal processes that would efficiently bound the target process. The 2-Stage Rejection method also has potential here. The idea would be to choose a finite set A and sample from π_λ^A; if the sample also covers $\mathcal{C}$, then it has the required distribution. Nevertheless this also poses some serious questions such as the 'best' or most efficient choice of A.

An even more difficult task would be the sampling of Boolean models conditioned to satisfy some *connectivity* (Lantuéjoul 1997) or *disconnectivity* constraints. Such conditioning arises when modelling oil reservoirs, where preliminary drilling indicates which sites are covered and also connected by the oil deposit. Finding a feasible and efficient way of devising the bounding processes in order to deliver perfect samples poses quite a challenge, since a priori there is no bound on the size of a connected component of the Boolean model. In case of disconnectivity constraints the density of the Boolean model is locally stable and so one can employ the standard dominated CFTP protocol. Nevertheless implementing an *efficient* algorithm may not be as straightforward.

Chapter 4

Area-Interaction Point Process

4.1 Introduction

The area-interaction process was briefly introduced in Sections 1.1.10 & 1.1.11 as an example of a Gibbs and Markov point process. It was introduced as a model for both clustering and orderliness by Baddeley & Van Lieshout (1995). However it is related to the penetrable spheres model of Widom & Rowlinson (1970) which was introduced much earlier (Section 4.4 discusses the relation between the two models). Pairwise-interaction models have received much attention since they provide a "large variety of complex patterns" (Mase 1990). Nevertheless they do not seem to produce sufficient variety of clustering. The Strauss (1975) model (Section 1.1.10) was introduced as a model for clustering; however Kelly & Ripley (1976) showed that this model is non-integrable for parameter values that correspond to the desired clustering. Moreover simulations by Møller (1993) suggest that the Strauss model does not exhibit 'moderate' clustering; there is an abrupt transition from 'Poisson-like' patterns to tightly clustered ones. The area-interaction process exhibits both moderate clustering (*attraction*) and orderliness (*repulsion*). Baddeley & Van Lieshout (1995) show that it is a Ripley-Kelly Markov point process (cf. Section 1.1.11) with interactions of infinite order.

Markov point processes provide useful models for individuals or particles exhibiting some sort of interaction. If, in addition, the interaction between individuals depends not only on the location but also on some region *around* the individual then the area-interaction process is a plausible model for such instances. This region, 'the region of influence', can be a disk of fixed radius centred on the location of the individual. For example, competition for food and resources may cause a repulsion

effect between individual animals or plants as each tries to maximize its region of influence (the "selfish herd", Hamilton 1971). Conversely if the individuals are being hunted by a predator then they would try and minimize their 'region of vulnerability'; so an attractive area-interaction process is a plausible model (Baddeley & Van Lieshout 1995; Kendall 1997a).

This Chapter describes how the area-interaction process can be simulated on some bounded region $W \subset \mathbb{R}^2$. The process is defined via specifying its density (Eq. 4.1) with respect to a unit rate Poisson process. The density of a realization depends on the positions of the germs, hence the process in disjoint regions is not independent. Therefore a realization of the process on some bounded W is given *conditionally* on the process outside W, ie. conditional on the 'boundary conditions' (see Section 4.2.1). In Section 4.3 Rejection sampling is outlined and some comments about its limitations made. The two-component Gibbs sampler of Häggström *et al.* (1999) for the penetrable spheres model is described in Section 4.4. The penetrable spheres model is a bivariate point process model and it is seen how a realization of this model can be used to obtain an area-interaction process.

Section 4.5 introduces two methods which employ spatial birth-death process: the dominated CFTP algorithm of Kendall (1998) and the 'Clan of Ancestors' algorithm of Fernández *et al.* (2002). The description in Fernández *et al.* (2002), though general, requires further work in order to practically implement. We present a specialization of their Clan algorithm to the area-interaction process, but with explicit details on practical implementation. The implementation involves censoring births in order to generate the required interaction, where the acceptance probability depends on the area of irregular regions. Kendall (1997a) describes a 'cluster' trick which allows correct censoring without having to explicitly compute the acceptance probabilities. This trick has not been implemented before; therefore it is developed here and incorporated into the set up.

These procedures so far sample the area-interaction process for *fixed* values of the model parameters. In many instances it may be worthwhile to consider a procedure that simultaneously samples the process for a whole *range* of parameter values, ie. an *omnithermal* sampling scheme. To the best of our knowledge such an algorithm for the area-interaction process has not appeared in the literature; therefore in Section 4.6 we consider such an algorithm. The Clan algorithm of Fernández *et al.* (2002) is extended in order to define an exact omnithermal algorithm. Finally some concluding remarks and directions for further work are discussed in Section 4.7.

4.2 Definition of the Area-Interaction Process

As is usual with Gibbs point processes, the area-interaction process is specified by a density with respect to a unit rate Poisson process on a bounded region $W \subset \mathbb{R}^2$. The process on bounded regions is referred to as the 'standard area-interaction' process (Baddeley & Van Lieshout 1995). They also show that the standard area-interaction process can be "considered as the restriction to a bounded sampling window of a stationary point process on the whole of $\mathbb{R}^d$". So the standard process is a realization of a stationary process on a bounded window $W \subset \mathbb{R}^d$, with specified *boundary conditions* representing the stationary process on $W^c = \mathbb{R}^d \setminus W$.

Definition 4.1. The *standard area-interaction* point process in a compact region $W \subset \mathbb{R}^d$ is a point process specified via the (conditional) density with respect to a unit rate Poisson:

$$f(x \mid y) = \alpha_\beta \lambda^{n(x)} e^{-\beta\, m_2[\mathcal{U}(x)\setminus\mathcal{U}(y)]}. \tag{4.1}$$

where y represents the state of the process outside W, ie. the boundary conditions. The parameter $\lambda > 0$ is the underlying Poissonian intensity, β is the interaction parameter, α_β the normalizing constant, m_d Lebesgue measure and $\mathcal{U}(x) = \bigcup_{\xi\in x} B_r(\xi)$ the Boolean model (cf. Definition 3.1) of d-dimensional disks of fixed radius, $B_r(\cdot)$, associated with x.

Thus the area-interaction process can be thought of as a *weighted Poisson process* (Kendall 1997a). Henceforth refer to a process that has density of the form in Eq. (4.1) as an area-interaction process with underlying intensity λ and parameter β. The above definition can be generalized by replacing Lebesgue measure with any "totally finite, Borel regular measure on a locally compact complete separable metric space", and $\mathcal{U}(\cdot)$ replaced by a "myopically continuous function" (Baddeley & Van Lieshout 1995). For the purposes here, attention is restricted to the standard case on $\mathbb{R}^2$. Eq. (4.1) is integrable for all values of λ and β; if $\beta < 0$ the process exhibits *orderliness* or *repulsion*; if $\beta = 0$ the process is just an unconditional Poisson and for $\beta > 0$ it exhibits *clustering* or *attraction*. The density (Eq. 4.1) is attractive for $\beta > 0$ and repulsive for $\beta < 0$ (cf. Eqs. 1.10 & 1.11):

$$\text{attractive:} \qquad \ell(\xi; x) \geq \ell(\xi; y), \qquad \text{whenever } \xi \notin y \subseteq x; \tag{4.2}$$

$$\text{repulsive:} \qquad \ell(\xi; x) \leq \ell(\xi; y), \qquad \text{whenever } \xi \notin y \subseteq x. \tag{4.3}$$

The Papangelou conditional intensity ℓ is defined as $\ell(\xi; x) = \frac{f(x\cup\xi)}{f(x)}$ (cf. Section 1.1.8, Eq. 1.9).

4.2.1 Boundary Conditions

In this chapter the simulation of the standard area-interaction process on some bounded window $W \subset \mathbb{R}^2$ is considered. Let X denote a stationary area-interaction process on $\mathbb{R}^2$ and X_W its restriction to W. As is the case with Gibbs processes in bounded regions, the distribution of X_W must be defined conditional on the configuration outside W. Thus the density of a realization of X_W given X_{W^c} is then given by $f\left(x \mid X_{W^c}\right) \propto \lambda^{n(x)} e^{-\beta\ \psi(x|X_{W^c})}$, where $\psi\left(x \mid y\right) = m_2\left[\mathcal{U}\left(x\right) \setminus \mathcal{U}\left(y\right)\right]$; $e^{-\beta\psi}$ will be referred to as the 'weight factor'. There are various 'boundary conditions' one can assume when simulating such a process:

Zero ('minimal') boundary conditions: The stationary process on W^c is empty, ie. $X_{W^c} = \emptyset$.

Rectangular ('maximal') boundary conditions: X_{W^c} is such that $\left(W \oplus B_r\left(0\right) \setminus W\right) \subseteq \mathcal{U}\left(X_{W^c}\right)$, where $W \oplus B_r\left(0\right) = \{\xi \in \mathbb{R}^2;\ B_r\left(\xi\right) \cap W \neq \emptyset\}$ is the 'dilation' of W.

Toroidal/Periodic boundary conditions: Here W is assumed to be the centre of a 3×3 grid of identical sampling windows. Given a configuration on W, the process on W^c is the union of the same configuration in each of the other eight sampling windows.

The simulation algorithms considered here will assume *zero boundary* conditions and X_W will be denoted simply as X; non-trivial boundary conditions can be dealt with by simple modifications.

4.3 Rejection Sampling

An area-interaction process is a weighted Poisson process with weight $e^{-\beta\psi}$. This suggests a simple accept/reject procedure for sampling such a weighted process. The parameter β denotes the interaction parameter of an area-interaction process (Eq. 4.1); $\beta > 0$ produces clustered or attractive point patterns while $\beta < 0$ yields ordered or repulsive patterns. The Rejection algorithm for simulating a realization of an area-interaction process X on a bounded region $W \subset \mathbb{R}^2$ involves drawing an unconditional Poisson process, which is then accepted with probability proportional to the weight factor $e^{-\beta\psi}$. Computation of ψ may be difficult; so the following algorithm enables one to correctly determine whether a realization of a Poisson(λ) process can be accepted with the correct probability. Let $X \sim$ Poisson(λ) and $Y \sim$ Poisson($|\beta|$) processes on W.

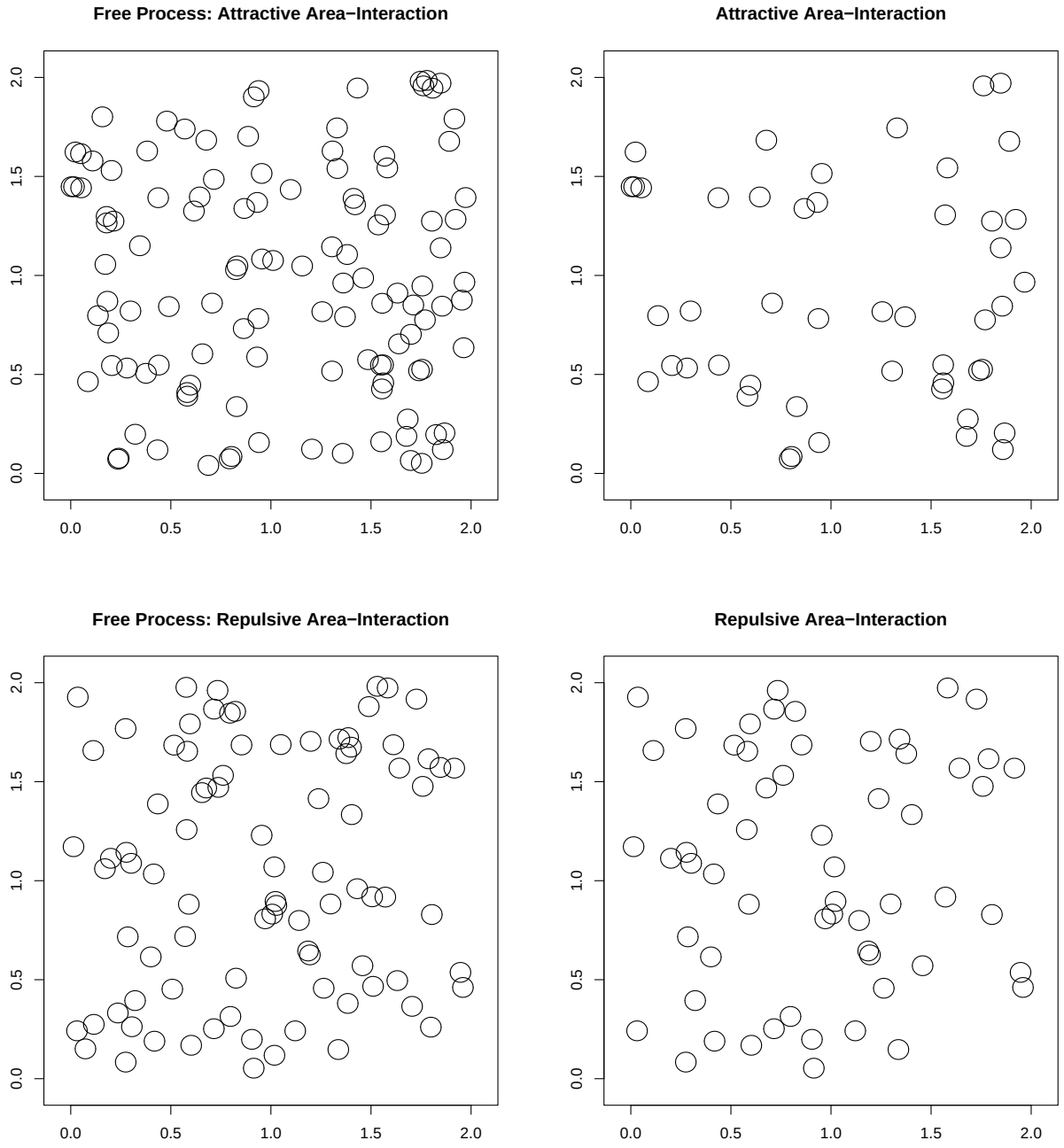

Figure 4.1: Sample of an attractive and repulsive area-interaction process. The perfect simulation procedure of Section 4.5.4 was used to produce these realizations. As discussed in that section, the area-interaction process can be thought of as a *dependent thinning* of an underlying Poisson process (referred to as the 'free' process). The pictures on the left show realizations of the free process and those on the right their respective dependent thinnings, ie. the area-interaction processes. *Top:* Attractive case; the underlying Poisson intensity $\lambda = 30$, interaction parameter $\beta = 100$, radius $r = 0.07$ and sampling window $W = [0, 2]^2$. *Bottom:* Repulsive case; $\lambda = 5$, $\beta = -100$, $r = 0.07$ and $W = [0, 2]^2$.

Algorithm 4.1 (Rejection: Attractive Case).

Set $T = 1$.

while $T > 0$:

Draw X and Y independently.

if $Y \cap \mathcal{U}(X) = \emptyset$: set $T = 0$.

else: set $T = T + 1$.

return X.

Algorithm 4.2 (Rejection: Repulsive Case).

Set $T = 1$.

while $T > 0$:

Draw X and Y independently.

if $Y \subseteq \mathcal{U}(X)$: set $T = 0$.

else: set $T = T + 1$.

return X.

Lemma 4.1 (Baddeley & Van Lieshout 1995, Lemma 5). *Let X, Y be as above. For $\beta > 0$ ($\beta < 0$) the conditional distribution of X, given $\{Y \cap \mathcal{U}(X) = \emptyset\}$ ($\{Y \subseteq \mathcal{U}(X)\}$), is that of an attractive (repulsive) area-interaction process with parameter β.*

The performance of Rejection sampling will be directly influenced by factors affecting the acceptance probability. As the underlying Poissonian intensity λ and/or the size of the sampling window W increases the acceptance probability becomes very low, and so Rejection is likely to be inefficient. One solution when W is large is to split it into smaller sub-windows and use a Gibbs sampling approach; Section 5.5 describes such a procedure for the conditional area-interaction process. In the next section a different type of Gibbs sampling approach is described.

4.4 Gibbs Sampling

In this section the two-component Gibbs sampler presented in Häggström *et al.* (1999) for the area-interaction process is described. They consider a bivariate process (X, Y) where X and Y refer to two different types of point configurations on W. This mixture model is related to the Widom & Rowlinson (1970) *penetrable spheres model* introduced in Section 1.1.11. The density of (X, Y) with respect to the product measure of two independent unit rate Poisson processes is given by

$$f(x, y) \propto \lambda_1^{n(x)} \lambda_2^{n(y)} \mathbf{1}_{\{d(x,y) > R\}}. \tag{4.4}$$

Here $\lambda_i, R > 0$ are model parameters and $d(x, y)$ is the shortest distance between x and y. Thus the interaction is only between points of different types and is such that a point in X is not allowed to be within distance R of a point in Y.

Häggström *et al.* (1999) show that the marginal distribution of X (Y) is an attractive area-interaction process with underlying intensity λ_1 (λ_2) and interaction parameter λ_2 (λ_1). Furthermore, the conditional distribution of X (Y) given Y (X) is a Poisson process of intensity λ_1 (λ_2) on $W \setminus \mathcal{U}(Y)$ $(W \setminus \mathcal{U}(X))$. These conditional distributions are easy to sample from; therefore a Gibbs sampler can be employed to sample (X, Y), and hence an area-interaction process.

4.4.1 Gibbs Sampling: Attractive Case

Let (Φ, Ψ) be a discrete-time bivariate Markov process. The following Gibbs sampler updates each component so as to ensure that the limiting distribution of (Φ, Ψ) has density given by Eq. (4.4).

Algorithm 4.3 (Gibbs: Attractive Area-Interaction).

Initialize $\Phi(0) \sim$ Poisson(λ_1), $\Psi(0) \sim$ Poisson(λ_2) on W.

for $n = 1, 2, \ldots$ update:

$\Phi(n) \sim$ Poisson(λ_1) on $W \setminus \mathcal{U}(\Psi(n-1))$; $\Psi(n) \sim$ Poisson(λ_2) on $W \setminus \mathcal{U}(\Phi(n))$.

4.4.2 Gibbs Sampling: Repulsive Case

Häggström *et al.* (1999) also define a bivariate process (X', Y') such that the marginal distribution of X' is a repulsive area-interaction process. Let the joint density at (x, y) be given by

$$f'(x, y) \propto \lambda_1^{n(x)} \lambda_2^{n(y)} \mathbf{1}_{\{y \subseteq \mathcal{U}(x)\}}. \tag{4.5}$$

In Section 4.4.1 the marginal distribution of either components with joint density given by Eq. (4.4) was that of an attractive area-interaction process. For the process with joint density given by Eq. (4.5), the marginal distribution of only the X'-component is a repulsive area-interaction process with intensity λ_1 and parameter λ_2. The conditional distribution of X', given Y', is that of a Poisson(λ_1) process on W conditioned so that $\mathcal{U}(X') \supseteq Y'$; that of Y' given X' is a Poisson(λ_2) on $\mathcal{U}(X')$.

Algorithm 4.4 (Gibbs: Repulsive Area-Interaction).

Initialize $\Phi'(0) \sim$ Poisson(λ_1), $\Psi'(0) \sim$ Poisson(λ_2) on W.

for $n = 1, 2, \ldots$ update:

$\Phi'(n) \sim$ Poisson(λ_1) conditioned so that $\mathcal{U}(\Phi'(n)) \supseteq \Psi'(n-1)$.

$\Psi'(n) \sim$ Poisson(λ_2) on $\mathcal{U}(\Phi'(n))$.

4.4.3 Exact Gibbs Sampler: Attractive Case

Häggström *et al.* (1999) combine the Coupling From The Past (CFTP) idea of Propp & Wilson (1996) with the two-component Gibbs sampler of Algorithm 4.3 in order to devise an exact simulation algorithm for the (attractive) area-interaction process. The idea of Propp & Wilson (1996) requires that the state space of the Markov process be finite, and that there is a maximal and minimal state (cf. Section 1.4.1). The state space of point processes is unbounded and there is no maximal element; however Häggström *et al.* (1999) point out that the fact that the state space is uncountable is "inconsequential" and introduce the notion of a *quasi-maximal* and *quasi-minimal* elements in order to construct the required coupling.

Let $\preccurlyeq$ be a partial order on the space of mixed configurations defined by $(x, y) \preccurlyeq (x', y')$ if $x \subseteq x'$ and $y \supseteq y'$. Call an element $\overline{(x, y)}$ *quasi-maximal* if $\mathcal{U}(x) \supseteq W$ and $y = \emptyset$; similarly $\underline{(x, y)}$ is *quasi-minimal* if $x = \emptyset$ and $\mathcal{U}(y) \supseteq W$. A random element (X, Y) is *stochastically dominated* by (X', Y'), written $(X, Y) \preccurlyeq_{st} (X', Y')$, if there exist couplings $\left(\tilde{X}, \tilde{Y}\right)$ and $\left(\tilde{X}', \tilde{Y}'\right)$ of (X, Y) and (X', Y') respectively such that $\left(\tilde{X}, \tilde{Y}\right) \preccurlyeq \left(\tilde{X}', \tilde{Y}'\right)$ almost surely (cf. Section 1.2.1 for more on coupling and stochastic domination).

Lemma 4.2 (Häggström *et al.* 1999, Lemma 2). *Fix $(x, y) \preccurlyeq (x', y')$. Let, for $n = 0, 1, \ldots$, $(\Phi(n), \Psi(n))$ be the random element obtained by starting with $(\Phi(0), \Psi(0)) = (x, y)$ and running n iterations of the two-component Gibbs sampler in Algorithm 4.3. Similarly define $(\Phi'(n), \Psi'(n))$, with $(\Phi'(0), \Psi'(0)) = (x', y')$. Then $(\Phi(n), \Psi(n)) \preccurlyeq (\Phi'(n), \Psi'(n))$ for all n.*

Lemma 3 in Häggström *et al.* (1999) then shows that if (Φ, Ψ) is initialized in a *quasi- maximal* state and (Φ', Ψ') in an arbitrary state, then $(\Phi(n), \Psi(n)) \succcurlyeq (\Phi'(n), \Psi'(n))$ for all n; conversely if (Φ, Ψ) is initialized in a *quasi-minimal* state and (Φ', Ψ') in an arbitrary state, then $(\Phi(n), \Psi(n)) \preccurlyeq (\Phi'(n), \Psi'(n))$ for all n. Thus the Gibbs sampler has the required monotonicity properties necessary in order to employ a Monotone CFTP construction (cf. Section 1.4.1).

Algorithm 4.5 (Exact Gibbs: Attractive Area-Interaction).

For $i = 0, 1, \ldots$, let $Z_1(-i) \sim \text{Poisson}(\lambda_1)$ and $Z_2(-i) \sim \text{Poisson}(\lambda_2)$ on W.

Choose an increasing sequence of positive integers $\{k_i;\ i = 0, 1, \ldots\}$; set $I = -1$, $i = 0$.

Let $\overline{(x, y)}$ and $\underline{(x, y)}$ be quasi-maximal and quasi-minimal elements respectively.

while $I < 0$:

$$\text{Initialize: } \begin{cases} \left(\Phi^{k_i,max}\left(-k_i\right), \Psi^{k_i,max}\left(-k_i\right)\right) = & \overline{(x,y)}. \\ \left(\Phi^{k_i,min}\left(-k_i\right), \Psi^{k_i,min}\left(-k_i\right)\right) = & \underline{(x,y)}. \end{cases}$$

for $n \in \{-k_i + 1, \ldots, 0\}$:

$\Phi^{k_i,max}(n) = Z_1(n) \setminus \mathcal{U}\left(\Psi^{k_i,max}(n-1)\right)$.

$\Psi^{k_i,max}(n) = Z_2(n) \setminus \mathcal{U}\left(\Phi^{k_i,max}(n)\right)$.

$\Phi^{k_i,min}(n) = Z_1(n) \setminus \mathcal{U}\left(\Psi^{k_i,min}(n-1)\right)$.

$\Psi^{k_i,min}(n) = Z_2(n) \setminus \mathcal{U}\left(\Phi^{k_i,min}(n)\right)$.

if $\left(\Phi^{k_i,max}(0), \Psi^{k_i,max}(0)\right) = \left(\Phi^{k_i,min}(0), \Psi^{k_i,min}(0)\right)$: set $i = I$.

else: set $i = i + 1$.

return $\left(\Phi^{k_I,min}(0), \Psi^{k_I,min}(0)\right)$.

$I = \inf\left\{i \geq 0;\ \left(\Phi^{k_i,max}(0), \Psi^{k_i,max}(0)\right) = \left(\Phi^{k_i,min}(0), \Psi^{k_i,min}(0)\right)\right\} < \infty$ a.s., ie. Algorithm 4.5 terminates in finite time, and the distribution of the output $\left(\Phi^{k_I,min}(0), \Psi^{k_I,min}(0)\right)$ has density given by Eq. (4.4) (Häggström *et al.* 1999, Theorem 3). Thus $\Phi^{k_I,min}(0)$ is an exact draw from an attractive area-interaction process with underlying intensity λ_1 and parameter λ_2. The Gibbs sampler for the repulsive case (Algorithm 4.4) does not preserve the partial ordering between two elements. This monotonicity is vital in order to employ CFTP on unbounded state spaces; thus an exact Gibbs sampling scheme for the repulsive does not seem possible.

4.5 Simulation via Spatial Birth-Death Processes

The area-interaction process is specified via the density in Eq. (4.1) with respect to a unit rate Poisson process. This section details how spatial birth-death processes can be used to sample the area-interaction process on $W \subset \mathbb{R}^2$. The main idea is to devise a coupling of spatial birth-death processes, where one process can be simulated relatively easily and a realization of the other can be obtained by coupling its evolution to that of the first (see Example 1.2 for a simple illustration).

Kendall (1998) first used birth-death processes in order to describe a dominated CFTP algorithm for both the attractive and repulsive area-interaction processes. Fernández *et al.* (2002) later on proposed a perfect simulation algorithm which could be used to sample any process whose distribu-

tion is absolutely continuous with respect to that of a Poisson process. In this section we describe the construction of a birth-death process whose equilibrium distribution is that of an area-interaction process. The birth rate of the process involves calculating areas of complicated geometrical patterns. Kendall (1998) uses a discretized approximation of these areas while the description in Fernández *et al.* (2002) lacks details on practical implementation. Here we develop a 'cluster' trick where the idea is to construct an observable event which would correctly *implement* the birth rate.

A spatial birth-death process (Section 1.3) on $W \subset \mathbb{R}^2$ is a continuous-time Markov jump process, taking values in the *Exponential Space* (Section 1.1.1) of W. The only transitions that occur are *births* and *deaths*. The birth times $\{t_\xi\}$ are determined by the birth rate b and the lifetimes $\{s_\xi\}$ by the death rate d. Refer to a spatial birth-death process as *free* (Fernández *et al.* 2002) if its birth and death rates do not depend on the current configuration, ie. there is no *interaction* between the individuals ξ, which therefore has a Poissonian equilibrium distribution. Conversely if the birth or death rate does depend on the current configuration, then we have an *interacting* spatial birth-death process since there is some kind of interaction between the individuals.

If the birth rate b and death rate d satisfy *detailed balance* (Eq. 1.25) for some density f:

$$f(x)\, b(x,\xi) = f(x \cup \{\xi\})\, d(x,\xi) \tag{4.6}$$

then the spatial birth-death process is time-reversible and f is the density of the unique stationary distribution of the birth-death process (Preston 1977; Kendall & Møller 2000). Setting $b(x,\xi) = \ell(\xi; x)$ and $d(x,\xi) = 1$ shows that detailed balance is trivially satisfied; process with these birth and death rates will converge to a unique stationary distribution with density f. The density of an area-interaction process is attractive for $\beta > 0$ and repulsive for $\beta < 0$ (cf. Eqs. 4.2 & 4.3). Its Papangelou conditional intensity, given by $\ell(\xi; x) = \frac{f(x \cup \{\xi\})}{f(x)}$ (cf. Eq. 1.9), is uniformly bounded:

$$\frac{f(x \cup \{\xi\})}{f(x)} = \lambda\, e^{-\beta\, m_2[\mathcal{U}(\xi) \setminus \mathcal{U}(x)]} \leq \lambda, \qquad \text{for } \beta > 0; \tag{4.7}$$

$$\frac{f(x \cup \{\xi\})}{f(x)} = \lambda\, e^{-\beta\, m_2[\mathcal{U}(\xi) \setminus \mathcal{U}(x)]} \leq \lambda\, e^{-\beta\, m_2[\mathcal{U}(\xi)]}, \qquad \text{for } \beta < 0 \tag{4.8}$$

since $m_2[\mathcal{U}(\xi)]$ is constant for all ξ. Eq. (4.7) shows that the conditional intensity of an attractive area-interaction process is uniformly bounded by λ; thus a Poisson(λ) *stochastically dominates* an attractive area-interaction process. Eq. (4.8) shows that a Poisson$\big(\lambda\, e^{-\beta\, m_2[\mathcal{U}(\xi)]}\big)$ *stochastically*

dominates a repulsive area-interaction process (cf. Remark 1.2). We use these observations to construct an attractive (respectively, repulsive) interacting birth-death process by coupling its transitions to those of a free process with birth rate λ (respectively, $\lambda\, e^{-\beta\, m_2[\mathcal{U}(\xi)]}$).

4.5.1 The Free Process

A free spatial birth-death process on W with birth rate b experiences births $\{\xi\}$ at random times $\{t_\xi\}$ which are Exponentially$(b\, m_2\,[W])$ distributed; unit per capita death rate yields Exponentially(1) distributed lifetimes $\{s_\xi\}$. The number of transitions in a finite time interval are almost surely finite. The countable family $\{(\xi, t_\xi, s_\xi)\}$ constitute the *free process*, which can thus be regarded as a Poisson process on $W \times \mathbb{R} \times (0, \infty)$ with intensity measure

$$\left(be^{-s}\, \mathbf{1}_{\{x\, \in\, W\}}\right) m_2\,(dx)\, m_1\,(dt)\, m_1\,(ds)$$

where m_d denotes Lebesgue measure on $\mathbb{R}^d$. The intensity measure factorizes as a product of the uniform measure of intensity b on $W \times \mathbb{R}$ and the Exponential(1) distribution for the lifetime s (Kendall 1998). Recall that $\mathcal{U}\,(x)$ denotes the Boolean model of disks of fixed radius associated with x, ie. $\mathcal{U}\,(x) = \bigcup_{\xi \in x} B_r\,(\xi)$, where $B_r\,(\xi)$ is a disk of fixed radius r centred on ξ. The construction in Kendall (1998) and Fernández *et al.* (2002) is based on the free *cylinder* process $\{(\mathcal{U}\,(\xi) \times [t_\xi, t_\xi + s_\xi])\}$; however we will work with the free process $\{(\xi, t_\xi, s_\xi)\}$ itself.

The birth rate of an interacting spatial birth-death process is of the form $b\rho\,(\xi \mid x)$, where x denotes the current configuration. "The first factor (b) is the basic birth rate due to an "internal" Poissonian clock, and ρ acts as an un-normalized probability for the individual ξ to be actually born once the internal clock has rung" (Fernández *et al.* 2002). For the area-interaction process:

$$\text{attractive } (\beta > 0): \qquad b = \lambda; \qquad \rho\,(\xi \mid x) = e^{-\beta\, m_2[\mathcal{U}(\xi) \setminus \mathcal{U}(x)]} \leq 1. \tag{4.9}$$

$$\text{repulsive } (\beta < 0): \qquad b = \lambda\, e^{-\beta\, m_2[\mathcal{U}(\xi)]}; \qquad \rho\,(\xi \mid x) = e^{\beta\, m_2[\mathcal{U}(\xi) \cap \mathcal{U}(x)]} \leq 1. \tag{4.10}$$

In order to implement the correct birth rates for the interacting process, an additional piece of randomness is required. The construction in Kendall (1998) and Fernández *et al.* (2002) employ a Uniform$(0, 1)$ random variable Z_ξ to do this: a birth ξ in the free process is accepted in the interacting process if $Z_\xi < \rho\,(\xi \mid x)$.

A more complicated piece of randomness is employed here, the reason being twofold: (i) it will enable implementation of the correct birth rates for the interacting process *without* having to compute the birth-acceptance probabilities in Eqs. $(4.9\ \&\ 4.10)$ (the 'cluster' trick described in Section 4.5.2 below); (ii) it will enable an omnithermal construction for the area-interaction process (Section 4.6). Specifically, to each triplet (ξ, t_ξ, s_ξ) attach a Poisson process Z_ξ on $\mathcal{U}(\xi) \times (0,1)$ with intensity measure $|\beta|\, m_2(dy)\, m_1(du)$ (cf. Algorithm 4.6 below). The construction of the attractive (respectively, repulsive) interacting process will be based on the *marked free process* Φ with birth rate b given by Eq. (4.9) (respectively, Eq. 4.10):

$$\Phi = \{(\xi, t_\xi, s_\xi, Z_\xi)\}; \tag{4.11}$$

$$\Phi(t) = \{(\xi, t_\xi, s_\xi, Z_\xi);\ t_\xi \le t \le t_\xi + s_\xi\}. \tag{4.12}$$

Thus $\Phi(t)$ has a Poissonian distribution for all t. For an individual ξ let $\tilde{\xi} = (\xi, t_\xi, s_\xi, Z_\xi)$ and use the notation "$\left\{\xi;\ \tilde{\xi} \in \Phi(t)\right\}$" to mean "$\{\xi;\ (\xi, t_\xi, s_\xi, Z_\xi) \in \Phi(t)\}$".

Simulation of Φ

A free process Φ on W with birth rate b is such that the collection of individuals $\left\{\xi;\ \tilde{\xi} \in \Phi(t)\right\}$ alive at time t are a realization of a Poisson(b) process on W. It is straightforward to draw a Poisson(b) process on W (cf. Section 1.1.5). The mark process Z_ξ is constructed as follows:

Algorithm 4.6 (Construction of Z_ξ).

Draw $Y_\xi \sim$ Poisson$(|\beta|)$ process on $\mathcal{U}(\xi)$; set $U_\xi = \emptyset$.

for each $y_\xi^k \in Y_\xi$: draw $U_\xi^k \sim$ Uniform$(0,1)$; set $U_\xi = U_\xi \cup \left\{U_\xi^k\right\}$.

return $Z_\xi = (Y_\xi, U_\xi) = \left\{\left(y_\xi^k, U_\xi^k\right);\ k = 0, \ldots, n(Y_\xi)\right\}$.

Algorithm 4.7 (Simulation of $\Phi(0)$).

Set $\Phi(0) = \emptyset$; draw $X \sim$ Poisson(b) on W.

for $\xi \in X$:

set $t_\xi = 0$; draw $s_\xi \sim$ Exponential(1); construct Z_ξ as in Algorithm 4.6.

set $\tilde{\xi} = (\xi, t_\xi, s_\xi, Z_\xi)$; $\Phi(0) = \Phi(0) \cup \left\{\tilde{\xi}\right\}$.

return $\Phi(0)$.

A realization of Φ on a finite time interval, say $[0, T]$, is now described below.

Algorithm 4.8 (Forward Simulation of Φ).

Draw $\Phi(0)$; set $t = 0$.

while $t \leq T$:

Draw $\tau_{\text{birth}} \sim \text{Exponential}(b\, m_2\, [W])$; this yields the next birth time.

Set $\tau = \min\left\{\tau_{\text{birth}},\ \inf\left\{t_\xi + s_\xi;\ \tilde{\xi} \in \Phi\,(t)\right\}\right\}$; this yields the next incident time of Φ.

if $\tau = \tau_{\text{birth}}$:

draw ξ Uniformly on W; $s_\xi \sim \text{Exponential}(1)$; construct Z_ξ; set $t_\xi = t + \tau$.

update $\Phi\,(t+\tau) = \Phi\,(t) \cup \left\{\tilde{\xi}\right\}$, where $\tilde{\xi} = (\xi, t_\xi, s_\xi, Z_\xi)$.

else if $\tau = t_\xi + s_\xi$ for some $\tilde{\xi} \in \Phi\,(t)$:

update $\Phi\,(t+\tau) = \Phi\,(t) \setminus \left\{\tilde{\xi}\right\}$.

update $t = t + \tau$.

return $\Phi(T)$.

It is worthwhile at this point to mention that Φ is time-reversible so one can simulate it *forwards* as well as *backwards* in time. For an individual ξ, t_ξ represents its *forward* birth time. So in order to simulate Φ backwards from time 0 to $-T$ it suffices to simulate it forwards on $[0, T]$ and then reflect in the time axis at $t = 0$. Thus a forward birth becomes a backward death and vice versa. If a forward birth ξ occurs at time t_ξ and dies at time $t_\xi + s_\xi$ then in the backwards evolution ξ is born at time $-\,(t_\xi + s_\xi)$ and dies at time $-t_\xi$, since forward births become backward deaths and vice versa.

4.5.2 The Interacting Process

Let Ψ denote an interacting spatial birth-death process on W such that the equilibrium distribution of Ψ is an area-interaction process with parameter β. Simulation of Ψ is carried out by coupling its evolution to that of Φ. Deaths in Φ are always accepted in Ψ, if present, so that each individual has unit death rate. At the birth time t_ξ of individual ξ in the free process Φ, its corresponding birth rate in the Ψ process is given by $b\, \rho\,(\xi \mid \Psi\,(t_\xi -))$, where b and ρ are given by Eq. (4.9) if $\beta > 0$ and by Eq. (4.10) if $\beta < 0$. Detailed balance (Eq. 4.6) calculations show that accepting the *proposed* birth ξ in $\Psi(t_\xi -)$ with probability $\rho\,(\xi \mid \Psi\,(t_\xi -))$ ensures that the limiting distribution of $\Psi(t)$ is an area-interaction process with parameter β.

The idea behind the 'cluster' trick (Kendall 1997a) is to construct an observable event which has

occurrence probability ρ; the proposed birth is then accepted if the event occurs and not otherwise. For the attractive case ($\beta > 0$) the event $\{Y_\xi \cap (\mathcal{U}(\xi) \setminus \mathcal{U}(\Psi(t_\xi-))) = \emptyset\}$ occurs with probability

$$e^{-\beta\, m_2\left[\mathcal{U}(\xi)\setminus\mathcal{U}\left(\Psi\left(t_\xi-\right)\right)\right]} = \rho(\xi \mid \Psi(t_\xi-)) \qquad \text{in Eq. } (4.9)$$

since $Y_\xi \sim$ Possion($|\beta|$). For $\beta < 0$, $\{Y_\xi \cap (\mathcal{U}(\xi) \cap \mathcal{U}(\Psi(t_\xi-))) = \emptyset\}$ occurs with probability

$$e^{-|\beta|\, m_2\left[\mathcal{U}(\xi)\cap\mathcal{U}\left(\Psi\left(t_\xi-\right)\right)\right]} = e^{\beta\, m_2\left[\mathcal{U}(\xi)\cap\mathcal{U}\left(\Psi\left(t_\xi-\right)\right)\right]} = \rho(\xi \mid \Psi(t_\xi-)) \qquad \text{in Eq. } (4.10).$$

It is easy to check whether any points of Y_ξ fall in the appropriate region. Furthermore, if the set of neighbours of ξ in $\Psi(t_\xi-)$ is denoted by $N_\xi^\beta = \{\eta \in \Psi(t_\xi-)\,;\, \mathcal{U}(\eta) \cap \mathcal{U}(\xi) \neq \emptyset\}$, then

$$\{Y_\xi \cap (\mathcal{U}(\xi) \setminus \mathcal{U}(\Psi(t_\xi-))) = \emptyset\} \equiv \left\{Y_\xi \cap \left(\mathcal{U}(\xi) \setminus \mathcal{U}\left(N_\xi^\beta\right)\right) = \emptyset\right\}. \tag{4.13}$$

$$\{Y_\xi \cap (\mathcal{U}(\xi) \cap \mathcal{U}(\Psi(t_\xi-))) = \emptyset\} \equiv \left\{Y_\xi \cap \left(\mathcal{U}(\xi) \cap \mathcal{U}\left(N_\xi^\beta\right)\right) = \emptyset\right\}. \tag{4.14}$$

So acceptance of ξ depends only on $Z_\xi = (Y_\xi, U_\xi)$ and N_ξ^β; Figure 4.2 below illustrates these rules.

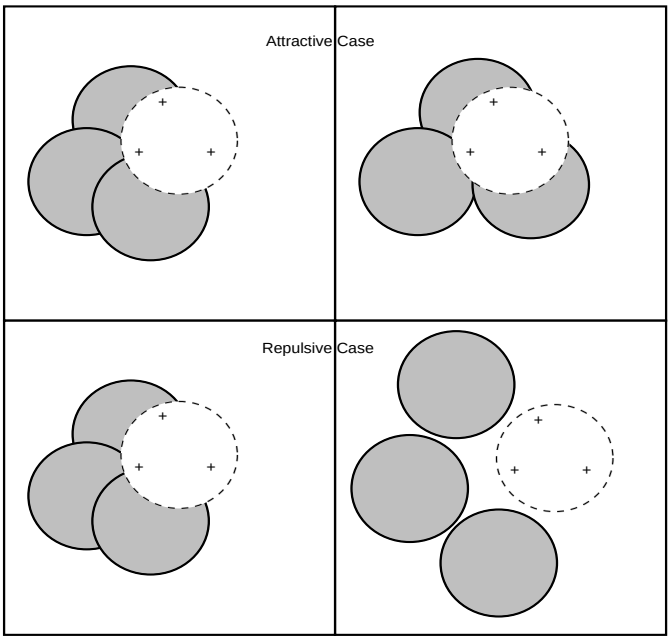

Figure 4.2: Illustration of the 'cluster' trick. The birth-acceptance events are given by Eqs. (4.13 & 4.14). The unfilled disk represents a birth ξ in the free process Φ, the filled disks its neighbours N_ξ^β in the interacting process Ψ. The crosses (+) represent the mark process Y_ξ. *Top (Attractive case):* ξ will be accepted if none of the crosses fall outside the filled region. *Bottom (Repulsive case):* ξ will be accepted if none of the crosses fall within the shaded region. In either case ξ will favour the pattern of filled grains on the right to those on the left.

Algorithm 4.9 (Simulation of Ψ: Finite Time Construction).

Draw $\Phi(0)$; set the birth rate b to λ (if $\beta > 0$) or $\lambda\, e^{-\beta\, m_2[\mathcal{U}(\xi)]}$ (if $\beta < 0$).

Initialize $\Psi(0) \subseteq \left\{\xi;\ \tilde{\xi} \in \Phi(0)\right\}$; set $t = 0$.

while $t \leq T$:

Simulate the next incident time τ of Φ; update $t = t + \tau$.

if τ is the birth time of individual ξ:

update $\Phi(t) = \Phi(t-) \cup \{\xi\}$.

for $\beta > 0$ (attractive case):

if $Y_\xi \cap \left(\mathcal{U}(\xi) \setminus \mathcal{U}\left(N_\xi^\beta\right)\right) = \emptyset$: update $\Psi(t) = \Psi(t-) \cup \{\xi\}$.

else: update $\Psi(t) = \Psi(t-)$.

for $\beta < 0$ (repulsive case):

if $Y_\xi \cap \left(\mathcal{U}(\xi) \cap \mathcal{U}\left(N_\xi^\beta\right)\right) = \emptyset$: update $\Psi(t) = \Psi(t-) \cup \{\xi\}$.

else: update $\Psi(t) = \Psi(t-)$.

else if τ is the death time of ξ:

update $\Phi(t) = \Phi(t-) \setminus \{\xi\}$; $\Psi(t) = \Psi(t-) \setminus \{\xi\}$.

return $\Psi(T)$.

Theorem 4.1. *The limiting distribution of $\Psi(T)$ as $T \to \infty$ is that of an attractive area-interaction process with parameter β.*

Proof. The birth rate $b(x, \xi)$ of Ψ is $\lambda\, \rho(\xi \mid x)$ (ρ is defined in Eqs. 4.9 & 4.10) and death rate is one per point. Detailed balance (Eq. 4.6) holds with respect to the density of an area-interaction process. Furthermore Ψ is irreducible since the process can go from one configuration to any other via the addition and deletion of individuals. Thus the process converges to an area-interaction process. □

4.5.3 Simulation of Ψ: Time Stationary Construction

In the following sections two perfect simulation algorithms for sampling exactly from the time-stationary distribution of Ψ are discussed. A virtual simulation of Ψ started from time $-\infty$ would result in $\Psi(0)$ being an exact draw from the required distribution, ie. an area-interaction point process with parameter β. Simulation from time $-\infty$ is sufficient but not necessary in order to

obtain $\Psi(0)$. This is because Algorithm 4.9 ensures that $\Psi(t) \subseteq \left\{\xi;\ \tilde{\xi} \in \Phi(t)\right\}$ for all t (where Φ has birth rate λ if $\beta > 0$ and $\lambda\, e^{-\beta\, m_2[\mathcal{U}(\xi)]}$ if $\beta < 0$). So one needs only to determine which individuals in the set $\left\{\xi;\ \tilde{\xi} \in \Phi(0)\right\}$ would also be alive in $\Psi(0)$.

Kendall (1998) describes a *dominated CFTP* algorithm involving the construction of upper- and lower-sandwich processes which coalesce in finite time and whose common value at time 0 is guaranteed to be a perfect draw from an area-interaction process. An alternative perfect simulation algorithm is presented in Fernández *et al.* (2002), namely the *Clan of Ancestors* algorithm. Since $\Psi(0) \subseteq \left\{\xi;\ \tilde{\xi} \in \Phi(0)\right\}$, it follows that $\Psi(0)$ can be obtained by a *dependent thinning* of the larger set of individuals (cf. Figure 4.1). Their algorithm is based on the notion of *ancestors*, which are those individuals in Φ born before time 0 that would influence the birth rates of those alive at time 0. Once all the relevant information about the ancestors is obtained, it is then possible to determine which of the individuals in $\left\{\xi;\ \tilde{\xi} \in \Phi(0)\right\}$ would also be alive in $\Psi(0)$.

4.5.4 Dominated CFTP Construction

Kendall & Møller (2000) present a general recipe, Dominated Coupling From The Past (domCFTP), for perfect simulation of locally stable point processes (Definition 1.11). A full description of domCFTP is given in Section 1.4.2. In this section the specific construction relating to the area-interaction process (as described in Kendall 1998) is discussed. The essence of dominated CFTP is the construction of upper- and lower-sandwich processes which bound the target process. As the name suggests, one requires a *dominating* process which serves as a stochastically varying maximal process; the evolution of the upper and lower bounding processes is then coupled to that of the dominating process. If coalescence of the bounding processes occurs in finite time, then their common value will also be the value of the target process. If one can arrange such a construction it would then be possible to obtain a perfect sample from the target distribution.

We first describe the construction for the attractive case $(\beta > 0)$; the repulsive case $(\beta < 0)$ can be dealt with by a simple modification, which is discussed at the end of this section. Let the target process Ψ be an interacting birth-death process whose stationary distribution is an attractive area-interaction process. Algorithm 4.9 shows that $\Psi(t) \subseteq \left\{\xi;\ \tilde{\xi} \in \Phi(t)\right\}$ for all t. The free process Φ thus serves as the 'dominating' process. Consider a virtual simulation of the free process

Φ backwards from time 0 to time $-\infty$. If Ψ is initialized at time $-\infty$ and coupled to the forwards evolution of the free process then Ψ would be in equilibrium at time 0. So, for $T > 0$, let $\Psi^{T,max}$ and $\Psi^{T,min}$ denote the upper and lower sandwich process started at time $-T$, initialized as:

$$\Psi^{T,max}(-T) = \left\{\xi;\ \tilde{\xi} \in \Phi(-T)\right\}; \qquad \Psi^{T,min}(-T) = \emptyset. \tag{4.15}$$

These sandwich processes are coupled the free process on $[-T, 0]$ such that the *sandwiching*, *coalescence* and *funnelling* properties (cf. Eqs. 1.26, 1.27 & 1.28) hold:

$$\Psi^{T,min}(t) \subseteq \Psi(t) \subseteq \Psi^{T,max}(t) \subseteq \left\{\xi;\ \tilde{\xi} \in \Phi(t)\right\}, \qquad -T \leq t \leq 0. \tag{4.16}$$

$$\Psi^{T,min}(t) = \Psi^{T,max}(t) \ \text{ if } \ \Psi^{T,min}(s) = \Psi^{T,max}(s), \qquad -T \leq s \leq t \leq 0. \tag{4.17}$$

$$\Psi^{T,min}(t) \subseteq \Psi^{S,min}(t) \subseteq \Psi(t) \subseteq \Psi^{S,max}(t) \subseteq \Psi^{T,max}(t), \qquad -S \leq -T \leq t \leq 0. \tag{4.18}$$

These are necessary conditions to ensure that once the bounding processes have coalesced their common value is also the value of the target process Ψ (Kendall & Møller 2000, Theorem 2.1), so that their common value at time 0 has the required distribution. The following algorithm describes how to update these bounding processes in order to ensure Eqs. $(4.16, 4.17\ \&\ 4.18)$ hold.

Algorithm 4.10 (domCFTP: Attractive Area-Interaction).

Fix $T = 1$.

while $T > 0$:

Extend Φ backwards on $\left[-T, -\lfloor \frac{T}{2} \rfloor\right)$ with birth rate λ.

Initialize $\Psi^{T,max}(-T)$ and $\Psi^{T,min}(-T)$ as in Eq. (4.15).

for $t \in [-T, 0]$:

if t is the birth time of ξ in Φ:

set $\Psi^{T,max}(t) = \Psi^{T,max}(t-) \cup \{\xi\}$ if $Y_\xi \cap \left(\mathcal{U}(\xi) \setminus \mathcal{U}\left(\Psi^{T,max}(t-)\right)\right) = \emptyset$.

set $\Psi^{T,min}(t) = \Psi^{T,min}(t-) \cup \{\xi\}$ if $Y_\xi \cap \left(\mathcal{U}(\xi) \setminus \mathcal{U}\left(\Psi^{T,min}(t-)\right)\right) = \emptyset$.

else if t is the death time of ξ in Φ:

set $\Psi^{T,max}(t) = \Psi^{T,max}(t-) \setminus \{\xi\}$; $\Psi^{T,min}(t) = \Psi^{T,min}(t-) \setminus \{\xi\}$.

if $\Psi^{T,max}(0) = \Psi^{T,min}(0)$: set $T = -1$.

else: set $T = 2T$.

return $\Psi^{T,min}(0)$.

Note that when extending the free process Φ on $\left[-T, -\lfloor \frac{T}{2} \rfloor\right)$ and coupling the bounding processes to the realization of Φ on $[-T, 0]$, it is vital to re-use the same realization of Φ on $\left[-\lfloor \frac{T}{2} \rfloor, 0\right]$ as that obtained in the previous iteration of the algorithm. Algorithm 4.10 does ensure that $\Psi^{\cdot,max}$ and $\Psi^{\cdot,min}$ satisfy the sandwiching, coalescence and funnelling properties in Eqs. $(4.16, 4.17 \ \& \ 4.18)$, and so the distribution of the output is that of an attractive area-interaction with parameter β (Kendall 1998; Kendall & Møller 2000). The key observation from which these properties follow is that the acceptance probability for a given birth is monotonic in the current configuration since the density of the process is attractive (Eq. 4.2). Hence any birth that is accepted in $\Psi^{T,min}$ will also be accepted in $\Psi^{T,max}$ whereas the converse does not necessarily hold. Moreover the algorithm terminates almost surely, because whenever the dominating process Φ hits the empty configuration $\emptyset$, $\Psi^{\cdot,max}$ and $\Psi^{\cdot,min}$ coalesce and the algorithm consequently terminates. Since $\emptyset$ is an ergodic atom for Φ, the process hits $\emptyset$ inifinitely often with probability one.

For a repulsive area-interaction process, the density is repulsive (Eq. 4.3) and the above construction does not ensure the required properties in Eqs. $(4.16, 4.17 \ \& \ 4.18)$ hold. However "there is a remarkably simple modification of the algorithm which is effective in determining equilibrium in the repulsive case" (Kendall 1998). The only change to implement in Algorithm 4.10 is to set the birth rate of Φ to $\lambda \, e^{-\beta \, m_2[\mathcal{U}(\xi)]}$ (cf. Eq. 4.10) and at the birth of an individual ξ at time t_ξ:

$$\text{set} \quad \Psi^{T,max}(t_\xi) = \Psi^{T,max}(t_\xi-) \cup \{\xi\} \quad \text{if} \quad Y_\xi \cap \left(\mathcal{U}(\xi) \cap \mathcal{U}\left(\Psi^{T,min}(t_\xi-)\right)\right) = \emptyset.$$

$$\text{set} \quad \Psi^{T,min}(t_\xi) = \Psi^{T,min}(t_\xi-) \cup \{\xi\} \quad \text{if} \quad Y_\xi \cap \left(\mathcal{U}(\xi) \cap \mathcal{U}\left(\Psi^{T,max}(t_\xi-)\right)\right) = \emptyset.$$

This 'cross-over' trick now ensures that whenever a birth is accepted in the minimal process it is also accepted in the maximal process. This is because the birth-acceptance probability $\rho(\xi \mid x)$ in Eq. (4.10) is *anti-monotonic* in x, ie. $\rho(\xi \mid x) \leq \rho(\xi \mid y)$ for $y \subseteq x$. These changes ensure that births maintain the inclusion ordering between the maximal and minimal processes. The output of the modified algorithm is a perfect sample from a repulsive area-interaction process with parameter β. The perfect sampling of such anti-monotone systems has also been studied by Häggström & Nelander (1998) who present a general protocol for *anti-monotone CFTP* based on the 'cross-over' trick of Kendall (1998).

4.5.5 The Clan of Ancestors Algorithm

In this section we consider another perfect variant of Algorithm 4.9. In Section 4.5.3 it was seen that a realization of an area-interaction process can be obtained as a dependent thinning of a Poisson process. Dominated CFTP (Algorithm 4.10) uses the idea of going back further and further in time in order to obtain such a dependent thinning. The method in this section is based on the notion of ancestors, which are simply those births in the past that would affect the birth rates of the individuals in the present. This perfect simulation algorithm, introduced by Fernández *et al.* (2002), "is applicable, in principle, to any distribution that is absolutely continuous with respect to a Poisson process and which can be obtained as the equilibrium distribution of an interacting spatial birth-and-death process". Examples considered by Fernández *et al.* (2002) include the area- and perimeter-interaction process, low temperature Ising model, the random cluster model and equilibrium measures of loss networks. Their description is very general and conceptual, but practical implementation requires further work. The procedure described below is specialized to the area-interaction process, but with an explicit description of how to implement it in practice.

Recall from Section 4.5.3 that a virtual simulation of the interacting birth-death process Ψ yields $\Psi(0) \subseteq \left\{\xi;\ \tilde{\xi} \in \Phi(0)\right\}$ (where Φ has birth rate λ if $\beta > 0$ and $\lambda\, e^{-\beta\, m_2[\mathcal{U}(\xi)]}$ if $\beta < 0$). So $\Psi(0)$ can be obtained as a dependent thinning of the individuals in $\Phi(0)$. Given all the relevant information about those births that could influence the birth rates of the individuals alive in $\Phi(0)$, it is then possible to determine which individuals in $\left\{\xi;\ \tilde{\xi} \in \Phi(0)\right\}$ would also be alive in $\Psi(0)$.

Definition 4.2. For an individual ξ, let N_ξ denote its *set of neighbours* in Φ, defined as
$N_\xi = \{\eta;\ \tilde{\eta} \in \Phi(t_\xi -) \text{ and } \mathcal{U}(\eta) \cap \mathcal{U}(\xi) \neq \emptyset\} = \{\eta;\ (\eta \times [t_\eta, t_\eta + s_\eta]) \cap (\xi \times [t_\xi, t_\xi + s_\xi]) \neq \emptyset\}$.

Definition 4.3. For an individual ξ, let A^1_ξ denote its set of *first generation ancestors*, defined as $A^1_\xi = \{\eta \in N_\xi;\ Y_\xi \cap \mathcal{U}(\eta) \neq \emptyset\}$, where Y_ξ is a Poisson$(|\beta|)$ process on $\mathcal{U}(\xi)$ attached to ξ (circa Eq. 4.12 and Algorithm 4.6). Individuals in A^1_ξ are the first generation ancestors of ξ, and ξ is a first generation descendant of the individuals in A^1_ξ.

Definition 4.4. An individual ξ is *trivial* if $Y_\xi = \emptyset$; else it is *non-trivial*.

Remark 4.1. A trivial individual ξ is *always* accepted in Ψ, regardless of whether $\beta > 0$ or $\beta < 0$ since the events in Eqs. $(4.13\ \&\ 4.14)$ occur trivially. Furthermore, by definition, ξ will have no

first generation ancestors. A non-trivial individual, on the other hand, has positive (free-process) probability of having first generation ancestors; hence the decision to accept such an individual *will* also depend on which of its neighbours are alive in Ψ.

Remark 4.2. An individual ξ will have no first generation ancestors if it is trivial or it is non-trivial and $N_\xi = \emptyset$. The decision to accept an individual with no first generation ancestors is then:

if $\beta > 0$**:** accept if it is trivial; reject if it is non-trivial.

if $\beta < 0$**:** accept in either case, since the event $\{Y_\xi \cap (\mathcal{U}(\xi) \cap \mathcal{U}(\Psi(t_\xi -))) = \emptyset\}$ always occurs.

To determine whether a non-trivial ξ is accepted in $\Psi(t_\xi -)$ one requires the set of first generation ancestors of ξ which are alive in $\Psi(t_\xi -)$, $K_\xi^1 = \left\{\eta \in A_\xi^1;\ \eta \in \Psi(t_\xi -)\right\}$. Once this is known then:

if $\beta > 0$**:** ξ is accepted if $Y_\xi \cap \left(\mathcal{U}(\xi) \setminus \mathcal{U}\left(K_\xi^1\right)\right) = \emptyset$, and not otherwise.

if $\beta < 0$**:** ξ is accepted if $Y_\xi \cap \left(\mathcal{U}(\xi) \cap \mathcal{U}\left(K_\xi^1\right)\right) = \emptyset$, and not otherwise.

However determining K_ξ^1 may require information about the first generation ancestors of the individuals in A_ξ^1, ie. the *second generation* ancestors of ξ. Working recursively, one has to deal with all generations of ancestors of ξ, ie. the 'Clan of Ancestors' of ξ, denoted by A_ξ. Implicit in the clan algorithm is the requirement that *all* the clans A_ξ "are finite with (free-process) probability one" (Fernández *et al.* 2002). A sufficient condition for this is presented in Section 4.5.6.

Constructing the Clan of Ancestors

If the conditions set out in Section 4.5.6 hold, so that the clan of any individual is almost surely finite, then the following explicitly details how to construct the clan of a given individual. Suppose Algorithm 4.9 is started from time $-\infty$ and Ψ coupled to the forward evolution of Φ till time 0 (bearing in mind that the birth rate is λ for $\beta > 0$ and $\lambda\, e^{-\beta\, m_2[\mathcal{U}(\xi)]}$ for $\beta < 0$). If ξ is an individual alive at time 0, its ancestors are those individuals born (in the forward evolution of Φ) before time 0 which influence the birth rate of ξ. Time reversibility of Φ means that one could evolve the process *backwards* from time 0; forward births become backward deaths and vice versa. The first generation ancestors of ξ are determined at its (forwards) birth incident; therefore in the backwards simulation the ancestors will be determined at its death incident.

Algorithm 4.11 (Clan Construction).

Set $A_\xi = \emptyset;\ B_\xi = \{\xi\}$

while $B_\xi \neq \emptyset$:

Simulate the next backwards incident time of Φ.

if a death is proposed:

if the individual to be deleted is in B_ξ:

if it has any first generation ancestors in Φ: add them to A_ξ and B_ξ.

delete the individual from Φ and B_ξ if present.

return A_ξ.

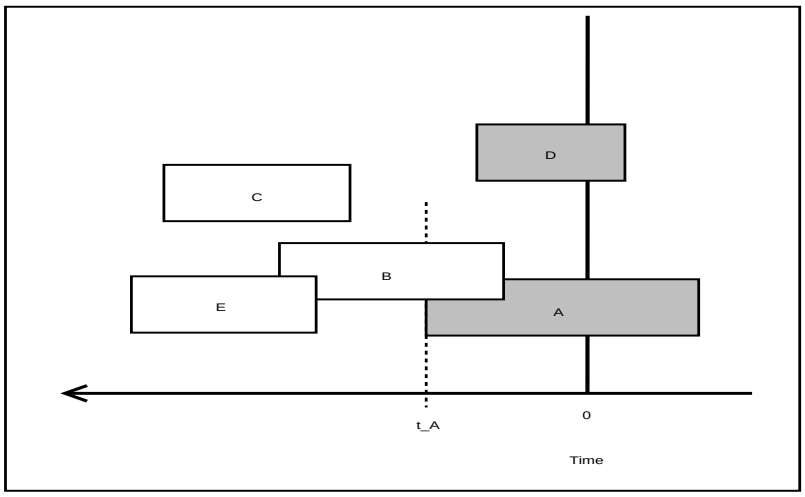

Figure 4.3: An illustration of the clan construction (Algorithm 4.11). The figure represents the evolution of the free process Φ with the rectangles representing the individuals. The end points of the rectangles denote the birth and death times; for example t_A is the forward birth (or equivalently, backward death) time of individual A. The shaded rectangles $\{A, D\}$ are those individuals alive in $\Phi(0)$; thus one needs to determine their respective clans. Individual A has only one neighbour, B. If B is not a first generation ancestor of A (cf. Definition 4.3) then the procedure stops there and the clan of A is $\{A\}$; else B is added to the clan and the procedure propagates further backwards. B has one neighbour, E; if E is a first generation ancestor of B then it is also added to the clan of A. If not, procedure stops and the clan of A is $\{A, B\}$. Since D, E have no neighbours then neither has any ancestors. Therefore the whole clan construction ends there and the Cleaning Algorithm 4.12 determines which of $\{A, D\}$ will be alive in the area-interaction process.

If A_ξ is finite then the set $\{\eta \in A_\xi;\ \eta$ has no first generation ancestors$\}$ is non-empty and the decision to accept any η in this set only depends on its mark process Y_η (cf. Remark 4.2). The following *Cleaning Algorithm* (Fernández *et al.* 2002) now describes how, given this set, to determine exactly

which of the other individuals in A_ξ would also be alive in Ψ.

Algorithm 4.12 (Cleaning Algorithm).

Set $K_\xi = \emptyset$ and $H_\xi = A_\xi \cup \{\xi\}$.

for each individual in $\{\eta \in A_\xi;\ \eta$ has no first generation ancestors$\}$:

if η is allowed to be born in Ψ (as determined by Remark 4.2): update $K_\xi = K_\xi \cup \{\eta\}$.

update $H_\xi = H_\xi \setminus \{\eta\}$.

Order the remaining individuals in H_ξ by time of birth, with the oldest first.

while $H_\xi \neq \emptyset$:

let the first individual in H_ξ be ζ, and $A_\zeta^1 \subseteq A_\xi$ its first generation ancestors.

compute its set of alive first generation ancestors $K_\zeta^1 = \left\{\eta \in A_\zeta^1;\ \eta \in K_\xi\right\}$.

if $\beta > 0$:

if $Y_\zeta \cap \left(\mathcal{U}\left(\zeta\right) \setminus \mathcal{U}\left(K_\zeta^1\right)\right) = \emptyset$: update $K_\xi = K_\xi \cup \{\zeta\}$.

else if $\beta < 0$:

if $Y_\zeta \cap \left(\mathcal{U}\left(\zeta\right) \cap \mathcal{U}\left(K_\zeta^1\right)\right) = \emptyset$: update $K_\xi = K_\xi \cup \{\zeta\}$.

update $H_\xi = H_\xi \setminus \{\zeta\}$.

return K_ξ.

So, for each ξ such that $\tilde{\xi} \in \Phi\left(0\right)$, applying the above procedure yields the set K_ξ of individuals in $A_\xi \cup \{\xi\}$ that are allowed to be born in Ψ. The set $\left\{\xi;\ \tilde{\xi} \in \Phi\left(0\right),\ \xi \in K_\xi\right\} \equiv \Psi\left(0\right)$ must hence be a perfect draw from an attractive area-interaction process with parameter β.

4.5.6 A Space-Time Percolation Problem

Fernández *et al.* (2002) point out that "the relation "being an ancestor of" gives rise to a model of *oriented-percolation*", or more specifically "*backward oriented-percolation*" since the clans are defined only by looking into the past. Such a percolation is the space-time analogue of the 'Snails on a Lily Pond' model discussed in Grimmett (1989, Section 10.5); see also the references therein for results on percolation in space and a general theory of coverage processes (eg. Hall 1988).

A description of the percolation event runs as follows, following the basic notation of Grimmett (1989). For $\tilde{\xi} = (\xi, t_\xi, s_\xi, Z_\xi)$ define $\tilde{\xi}$ to be *adjacent* to $\tilde{\eta}$, written $\tilde{\xi} \sim \tilde{\eta}$, if η is a first generation ancestor of ξ. Furthermore $\tilde{\xi}$ is connected to $\tilde{\eta}$, written $\tilde{\xi} \leftrightarrow \tilde{\eta}$, if there exists a sequence

$\left\{\tilde{\xi}_k;\ 1 \leq k \leq n\right\}$ such that $\tilde{\xi}_1 = \tilde{\xi}$, $\tilde{\xi}_n = \tilde{\eta}$ and $\tilde{\xi}_{k+1} \sim \tilde{\xi}_k$ for all $k \in \{1, \ldots, n-1\}$. A *cluster* is a set $\left\{\tilde{\xi}_i;\ 1 \leq i \leq n\right\}$ which is maximal with the property that $\tilde{\xi}_i \leftrightarrow \tilde{\xi}_j$ for all $i, j \in \{1, \ldots, n\}$. Let $\mathbf{C}_0$ denote the cluster containing some $\tilde{\xi} \in \Phi(0)$; $\mathbf{C}_0$ *extends or propagates backwards* to time $t < 0$ if $t > \inf\left\{t_\xi;\ \tilde{\xi} \in \mathbf{C}_0\right\}$. The validity of time-stationary clan construction for the area-interaction process requires "the absence of a cluster extending back to time $-\infty$" (Fernández *et al.* 2002). We use branching process theory in order to get a sufficient condition which guarantees this.

Consider individual ξ, which by space-time invariance of the free process, can be assumed to be born at time 0 at the origin. Looking backwards into the past of the free process Φ, this individual will have (or 'give rise to') a random number of first generation ancestors. Each of these first generation ancestors will, in turn, give rise to a random number of first generation ancestors and so on. From Remark 4.1 it is seen that only non-trivial individuals have positive probability of having first generation ancestors themselves; if an individual is trivial (so that it will have no first generation ancestors) then that branch of the process will not propagate any further and will die out at that trivial individual. Therefore it suffices to consider the branching process of non-trivial first generation ancestors, since the difference will involve at most a finite number of trivial individuals.

A sufficient condition to ensure certain extinction is that the branching process is sub-critical, ie. the expected number of non-trivial first generation ancestors is less that one. For an individual ξ let $\partial\mathcal{U}(\xi) = \{\eta \in \mathbb{R}^2;\ \mathcal{U}(\eta) \cap \mathcal{U}(\xi) \neq \emptyset\} = B_{2r}(\xi)$, where $\mathcal{U}(\xi) = B_r(\xi)$ is a disk of fixed radius r centred at ξ. The expected non-trivial first generation ancestors of ξ is then

$$\int_{B_{2r}(\xi)} \lambda\left(1 - e^{-\beta\ m_2[\mathcal{U}(\xi)]}\right) d\xi \int_{-\infty}^{0} dt \int_{t}^{\infty} e^{-s} ds = \lambda\left(1 - e^{-\beta\ m_2[\mathcal{U}(\xi)]}\right)\ m_2\left[B_{2r}(\xi)\right] \qquad (4.19)$$

where m_2 is Lebesgue measure on $\mathbb{R}^2$. So for $\lambda\left(1 - e^{-\beta\ m_2[\mathcal{U}(\xi)]}\right)\ m_2\left[B_{2r}(\xi)\right] < 1$ the almost sure finiteness of the clan of ancestors (of any individual) is guaranteed.

4.6 Omnithermal Sampling

The simulation algorithms considered so far sample the area-interaction process for *fixed* values of the model parameters. In this section a different type of sampling scheme is considered. An *omnithermal* sampling scheme is whereby a model or system is sampled simultaneously for a *whole range* of parameter values. That is, a *single* simulation which depicts the state of the system simul-

taneously at a range of parameter values. Such a procedure would prove very useful in estimating certain model/system characteristics or critical parameter values, since only a single (omnithermal) sample would be required as opposed to a "large number of samples at different values of the parameters" (Propp & Wilson 1996).

For example, an omnithermal sample of an Ising system was used by Propp & Wilson in order to illustrate how the internal energy and spontaneous magnetization of the system varies with temperature. This then enabled the estimation of the 'critical temperature'; "other macroscopically-defined quantities may also be graphed as a monotone function of the temperature". This idea of simultaneously sampling a system has been used in the theory of random graphs and percolation; see for example Propp & Wilson (1996) (and the references therein), Dimakos (2000) and Grimmett (1989). The work of Dimakos (2000) on omnithermal sampling for the random-cluster and Ising models proved very helpful and informative in understanding the concept. To the best of our knowledge an omnithermal algorithm for the area-interaction process has not appeared in the literature; therefore such a procedure is presented here.

Specifically, we consider sampling the process for a range of values of the interaction parameter β. It turns out that omnithermal sampling is only possible for the attractive case ($\beta > 0$) (see Remark 4.4 below). Let $\overline{\beta}$ be some large positive value and, for $p \in [0,1]$, let Ψ_p denote an (attractive) interacting spatial birth-death process which converges to an attractive area-interaction process with parameter $\beta = p\overline{\beta} > 0$. Thus Ψ_p is viewed as a *function* of p and the objective then is to simulate a realization of Ψ_p for *all values* of p given a single realization of the free process Φ. Algorithm 4.9 describes how the simulation of Ψ is carried out by coupling its evolution to that of the free process Φ. A birth in Φ is considered a *proposed* birth in Ψ and an additional (stochastic) test determines whether it is accepted, so as to ensure the correct birth rates for the interacting process. In an omnithermal sampling scheme, the idea is to keep track of the range of values of p for which the birth would have been accepted in the corresponding Ψ_p process.

Before proceeding, some modifications to the construction of the free process Φ (cf. Eqs. 4.11 & 4.12) have to be made. Recall from Section 4.5.1 that (ξ, t_ξ, s_ξ) represents an individual ξ, its birth time t_ξ and its lifetime s_ξ. Each such triplet is marked with a Poisson process $Z_\xi = (Y_\xi, U_\xi)$ on $\mathcal{U}(\xi) \times (0,1)$, where $\mathcal{U}(\xi)$ is a disk of fixed radius centred at ξ. The intensity measure of Z_ξ is $|\beta|\ m_2(dy)\, m_1(du)$. This thus enables the simulation of an area-interaction process with

parameter β. For an omnithermal construction the intensity measure of Z_ξ must be changed to $\overline{\beta}\, m_2(dy)\, m_1(du)$, since the idea is to simulate the process for all $\beta \in \left[0, \overline{\beta}\right]$. The mark Z_ξ can be constructed as described in Algorithm 4.6.

Let θ_p denote a random mapping which takes a point configuration x and maps it to some subset $y \subseteq x$ by independently retaining each point of x with probability p. So for the Poisson process $Z_\xi = (Y_\xi, U_\xi)$ and $p = \frac{\beta}{\overline{\beta}}$, $\theta_p(Y_\xi) \sim \text{Poisson}(\beta)$ using the thinning property of Poisson processes. Such a thinning of Y_ξ is obtained by $\theta_p(Y_\xi) = \left\{y_\xi^k \in Y_\xi;\ U_\xi^k < p\right\}$. For an individual ξ its neighbours in Φ (Definition 4.2) are $N_\xi = \{\eta;\ \tilde{\eta} \in \Phi(t_\xi -),\ \mathcal{U}(\eta) \cap \mathcal{U}(\xi) \neq \emptyset\}$; for a given p let $N_\xi^p = \{\eta \in \Psi_p(t_\xi -);\ \mathcal{U}(\eta) \cap \mathcal{U}(\xi) \neq \emptyset\}$ denote the neighbours of ξ in the corresponding Ψ_p process. The following Lemma establishes a monotonicity result between the Ψ_p processes for different values of p, which is a fundamental requirement for omnithermal sampling (Remark 4.3).

Lemma 4.3. *Let $0 \leq p \leq p' \leq 1$ and initialize $\Psi_p(0) = \Psi_{p'}(0)$. Couple the two processes to the same realization of the free process Φ as in Algorithm 4.9. Then $\Psi_p(t) \supseteq \Psi_{p'}(t)$ for all $t \geq 0$.*

Proof. The result follows by induction on the sequence of transitions in the free process. Death transitions are accepted in all the Ψ_p processes, so it needs to be shown that birth transitions maintain the inclusion ordering. Initially $\Psi_p(0) = \Psi_{p'}(0)$; let ξ_1 denote the first birth in Φ after time 0. Then $\Psi_p(t_{\xi_1} -) = \Psi_{p'}(t_{\xi_1} -)$ and hence $N_{\xi_1}^p = N_{\xi_1}^{p'}$. The random function θ_p is monotonic in p: $\emptyset = \theta_0(Y_{\xi_1}) \subseteq \theta_p(Y_{\xi_1}) \subseteq \theta_{p'}(Y_{\xi_1}) \subseteq \theta_1(Y_{\xi_1}) = Y_{\xi_1}$, $0 \leq p \leq p' \leq 1$; so

$$\begin{aligned} &\left\{\theta_p(Y_{\xi_1}) \cap \left(\mathcal{U}(\xi_1) \setminus \mathcal{U}\left(N_{\xi_1}^p\right)\right)\right\} \subseteq \left\{\theta_{p'}(Y_{\xi_1}) \cap \left(\mathcal{U}(\xi_1) \setminus \mathcal{U}\left(N_{\xi_1}^p\right)\right)\right\} \\ \Rightarrow &\left\{\theta_p(Y_{\xi_1}) \cap \left(\mathcal{U}(\xi_1) \setminus \mathcal{U}\left(N_{\xi_1}^p\right)\right) = \emptyset\right\} \supseteq \left\{\theta_{p'}(Y_{\xi_1}) \cap \left(\mathcal{U}(\xi_1) \setminus \mathcal{U}\left(N_{\xi_1}^{p'}\right)\right) = \emptyset\right\} \end{aligned}$$

since $N_{\xi_1}^p = N_{\xi_1}^{p'}$. So ξ_1 is accepted in $\Psi_p(t_{\xi_1} -)$ whenever it is accepted in $\Psi_{p'}(t_{\xi_1} -)$, but not necessarily the converse; hence $\Psi_p(t_{\xi_1}) \supseteq \Psi_{p'}(t_{\xi_1})$.

Now suppose that after the n-th birth ξ_n, $\Psi_p(t_{\xi_n}) \supseteq \Psi_{p'}(t_{\xi_n})$. At the next birth ξ_{n+1}, $N_{\xi_{n+1}}^p \supseteq N_{\xi_{n+1}}^{p'}$ implying $\left(\mathcal{U}(\xi_{n+1}) \setminus \mathcal{U}\left(N_{\xi_{n+1}}^p\right)\right) \subseteq \left(\mathcal{U}(\xi_{n+1}) \setminus \mathcal{U}\left(N_{\xi_{n+1}}^{p'}\right)\right)$. Again, monotonicity of θ_p then implies that ξ_{n+1} will be accepted in Ψ_p whenever it is accepted in $\Psi_{p'}$, but not necessarily the converse; so $\Psi_p\left(t_{\xi_{n+1}}\right) \supseteq \Psi_{p'}\left(t_{\xi_{n+1}}\right)$. The induction goes through since only finitely many births are proposed in finite time, and the ordering between Ψ_p and $\Psi_{p'}$ persists after every birth in Φ. $\square$

Remark 4.3. Lemma 4.3 above implies that, for any ξ, $\{p \in [0,1]\,;\ \xi \text{ is accepted in } \Psi_p\}$ is an interval of the form $[0, r_\xi] \subseteq [0,1]$. The upper limit r_ξ is referred to as the *omnithermal threshold* for ξ, ie. the real number in $[0,1]$ such that ξ is alive in Ψ_p process for all $p \leq r_\xi$ and not for $p > r_\xi$.

Remark 4.4. If we consider sampling Ψ_p for the repulsive case, with parameter $\beta = p\overline{\beta} < 0$, then the monotonicity between Ψ_p and $\Psi_{p'}$, for $p \leq p'$, does not always hold: consider a birth proposal ξ. It will be accepted in $\Psi_p(t_\xi -)$ if $\theta_p(Y_\xi) \cap (\mathcal{U}(\xi) \cap \mathcal{U}(\Psi_p(t_\xi -))) = \emptyset$ (cf. Eq. 4.14). This event is monotonic in p but anti-monotonic in the configurations: for $y \subseteq x$ and $p \leq p'$

$$\begin{aligned}
\{\theta_p(Y_\xi) \cap (\mathcal{U}(\xi) \cap \mathcal{U}(x))\} &\subseteq \{\theta_{p'}(Y_\xi) \cap (\mathcal{U}(\xi) \cap \mathcal{U}(x))\}\,; \\
\{\theta_p(Y_\xi) \cap (\mathcal{U}(\xi) \cap \mathcal{U}(x))\} &\supseteq \{\theta_p(Y_\xi) \cap (\mathcal{U}(\xi) \cap \mathcal{U}(y))\}\,.
\end{aligned}$$

So it is not possible to determine the inclusion ordering between $\{\theta_p(Y_\xi) \cap (\mathcal{U}(\xi) \cap \mathcal{U}(x))\}$ and $\{\theta_{p'}(Y_\xi) \cap (\mathcal{U}(\xi) \cap \mathcal{U}(y))\}$. Thus the analogous result to Lemma 4.3 may not necessarily hold for all $-1 \leq -p' \leq -p \leq 0$ in the repulsive case. For the attractive case the acceptance event (Eq. 4.13) is such that

$$\begin{aligned}
\{\theta_p(Y_\xi) \cap (\mathcal{U}(\xi) \setminus \mathcal{U}(x))\} &\subseteq \{\theta_{p'}(Y_\xi) \cap (\mathcal{U}(\xi) \setminus \mathcal{U}(x))\}\,; \\
\{\theta_p(Y_\xi) \cap (\mathcal{U}(\xi) \setminus \mathcal{U}(x))\} &\subseteq \{\theta_p(Y_\xi) \cap (\mathcal{U}(\xi) \setminus \mathcal{U}(y))\}\,.
\end{aligned}$$

This hence gives $\{\theta_p(Y_\xi) \cap (\mathcal{U}(\xi) \cap \mathcal{U}(x))\} \subseteq \{\theta_{p'}(Y_\xi) \cap (\mathcal{U}(\xi) \cap \mathcal{U}(y))\}$.

4.6.1 Computing the Omnithermal Threshold

Recall individual ξ is marked by a Poisson process $Z_\xi = (Y_\xi, U_\xi)$ on $\mathcal{U}(\xi) \times (0,1)$, where $Y_\xi \sim \text{Poisson}(\overline{\beta})$ on $\mathcal{U}(\xi)$ and $U_\xi = \{U_\xi^k;\ y_\xi^k \in Y_\xi\}$ is the collection of Uniform$(0,1)$ marks attached to each point y_ξ^k of Y_ξ. For a subset $M \subseteq N_\xi$ of the neighbours of ξ in Φ define

$$E_\xi(M) = \{p;\ \text{all } \eta \in M \text{ are alive in } \Psi_p\} \bigcap \{p;\ \theta_p(Y_\xi) \cap (\mathcal{U}(\xi) \setminus \mathcal{U}(M)) = \emptyset\}\,. \tag{4.20}$$

$$w_\xi(M) = \inf\{U_\xi^k;\ y_\xi^k \in (\mathcal{U}(\xi) \setminus \mathcal{U}(M))\}\,, \text{ for non-empty } M \subseteq N_\xi. \tag{4.21}$$

$$w_\xi(\emptyset) = \inf\{U_\xi^k;\ y_\xi^k \in Y_\xi\}\,.$$

$$v_\xi(M) = \inf\{r_\eta;\ \eta \in M\}\,, \text{ with the convention that } \inf\{\emptyset\} = 1. \tag{4.22}$$

If omnithermal thresholds of the neighbours of ξ are $\{r_\eta;\ \eta \in N_\xi\}$ then, for $M \subseteq N_\xi$:

$$\{p:\ \text{all } \eta \in M \text{ are alive in } \Psi_p\} = \bigcap_{\eta \in M} [0, r_\eta] = [0, v_\xi(M)]; \tag{4.23}$$

$$\text{and} \qquad \{p;\ \theta_p(Y_\xi) \cap (\mathcal{U}(\xi) \setminus \mathcal{U}(M)) = \emptyset\} = [0, w_\xi(M)]. \tag{4.24}$$

Eq. (4.24) holds since $\theta_p(Y_\xi) = \left\{y_\xi^k \in Y_\xi;\ U_\xi^k < p\right\}$; so for $p \leq w_\xi(M)$ none of the points of $\theta_p(Y_\xi)$ will fall in $\mathcal{U}(\xi) \setminus \mathcal{U}(M)$. If $p > w_\xi(M)$ then at least one point of Y_ξ must fall in $\mathcal{U}(\xi) \setminus \mathcal{U}(M)$ (else $w_\xi(M)$ would be 1!); furthermore at least one of these points will be retained in $\theta_p(Y_\xi)$. Eqs. (4.23 & 4.24) imply that $E_\xi(M) = [0, \min\{v_\xi(M), w_\xi(M)\}]$.

Theorem 4.2. *Suppose that at the birth time of individual ξ in Φ the thresholds for all its neighbours $\eta \in N_\xi$ are $\{r_\eta;\ \eta \in N_\xi\}$. Then the threshold for ξ is given by*

$$r_\xi = \max_{M \subseteq N_\xi} \{\min\{v_\xi(M), w_\xi(M)\}\}. \tag{4.25}$$

Proof. r_ξ is the upper limit of the set $\{p:\ \xi \text{ is accepted in } \Psi_p\}$; this is equal to

$$\begin{aligned} &\{p;\ \exists \text{ at least one } M \subseteq N_\xi \text{ with all } \eta \in M \text{ alive in } \Psi_p \text{ and } \theta_p(Y_\xi) \cap (\mathcal{U}(\xi) \setminus \mathcal{U}(M)) = \emptyset\} \\ = &\bigcup_{M \subseteq N_\xi} E_\xi(M) = \bigcup_{M \subseteq N_\xi} [0, \min\{v_\xi(M), w_\xi(M)\}] = \left[0, \max_{M \subseteq N_\xi} \{\min\{v_\xi(M), w_\xi(M)\}\}\right]. \end{aligned}$$

□

For implementation purposes Eq. (4.25) may be computationally burdensome as the maximum has to be taken over all subsets of N_ξ; however the following Corollary shows that one needs only to take the maximum over a much smaller set of subsets.

Corollary 4.1. *For an individual ξ let N_ξ denote its neighbours in the free process Φ, A_ξ^1 its first generation ancestors (Definition 4.3) and $\{r_\eta;\ \eta \in N_\xi\}$ the thresholds for neighbours of ξ. Order the individuals in A_ξ^1 as $\{\eta_1, \ldots, \eta_n\}$ so that $r_{\eta_1} \leq r_{\eta_2} \leq \ldots \leq r_{\eta_n}$. For $1 \leq i \leq n$ define $M_i \subseteq A_\xi^1$ by $M_i = \{\eta_n, \eta_{n-1}, \ldots, \eta_i\}$ and $M_0 = \emptyset$. Then*

$$\begin{aligned} r_\xi &= \max_{M \subseteq N_\xi} \{\min\{v_\xi(M), w_\xi(M)\}\} = \max_{M \subseteq A_\xi^1} \{\min\{v_\xi(M), w_\xi(M)\}\} \\ &= \max_{0 \leq i \leq n} \{\min\{v_\xi(M_i), w_\xi(M_i)\}\}. \end{aligned} \tag{4.26}$$

Proof. Any $M \subseteq N_\xi$ can be expressed as a disjoint union $M = Q_1 \cup Q_2$ where $Q_1 \subseteq A^1_\xi$ and $Q_2 \subseteq N_\xi \setminus A^1_\xi$. By Definition 4.3 any $\eta \in N_\xi \setminus A^1_\xi$ must have $Y_\xi \cap \mathcal{U}(\eta) = \emptyset$, where Y_ξ is the Poisson$\left(\overline{\beta}\right)$ process on $\mathcal{U}(\xi)$ attached to the individual ξ. Furthermore $\mathcal{U}(\xi) \setminus \mathcal{U}(M) = \mathcal{U}(\xi) \setminus \mathcal{U}(Q_1 \cup Q_2) = (\mathcal{U}(\xi) \setminus \mathcal{U}(Q_1)) \cap (\mathcal{U}(\xi) \setminus \mathcal{U}(Q_2))$. Since $Q_2 \subseteq N_\xi \setminus A^1_\xi$ then $Y_\xi \cap \mathcal{U}(Q_2) = \emptyset$, or equivalently, $Y_\xi \cap (\mathcal{U}(\xi) \setminus \mathcal{U}(Q_2)) = Y_\xi$. Therefore $Y_\xi \cap (\mathcal{U}(\xi) \setminus \mathcal{U}(M)) = Y_\xi \cap (\mathcal{U}(\xi) \setminus \mathcal{U}(Q_1))$ and Eq. (4.21) then yields $w_\xi(M) = w_\xi(Q_1)$. Since $M \supseteq Q_1$, Eq. (4.22) implies $v_\xi(M) \leq v_\xi(Q_1)$; therefore

$$\begin{aligned} &\max\{\min\{v_\xi(M), w_\xi(M)\}, \min\{v_\xi(Q_1), w_\xi(Q_1)\}\} \\ =\ &\max\{\min\{v_\xi(M), w_\xi(Q_1)\}, \min\{v_\xi(Q_1), w_\xi(Q_1)\}\} = \min\{v_\xi(Q_1), w_\xi(Q_1)\}. \end{aligned}$$

Since this holds for any $M \subseteq N_\xi$ it follows that

$$\max_{M \subseteq N_\xi} \{\min\{v_\xi(M), w_\xi(M)\}\} = \max_{Q \subseteq A^1_\xi} \{\min\{v_\xi(Q), w_\xi(Q)\}\}.$$

Now suppose $Q \subseteq A^1_\xi$, where $A^1_\xi = \{\eta_1, \ldots, \eta_n\}$ is ordered such that $r_{\eta_1} \leq r_{\eta_2} \leq \ldots \leq r_{\eta_n}$. Let the individual in Q with the lowest threshold be denoted by η_{i^*}. By definition, $Q \subseteq M_{i^*}$ where M_i is defined above; hence $v_\xi(Q) = v_\xi(M_{i^*})$ and $w_\xi(Q) \leq w_\xi(M_{i^*})$ (cf. Eq. 4.21). Therefore

$$\begin{aligned} &\max\{\min\{v_\xi(Q), w_\xi(Q)\}, \min\{v_\xi(M_{i^*}), w_\xi(M_{i^*})\}\} \\ =\ &\max\{\min\{v_\xi(M_{i^*}), w_\xi(Q)\}, \min\{v_\xi(M_{i^*}), w_\xi(M_{i^*})\}\} = \min\{v_\xi(M_{i^*}), w_\xi(M_{i^*})\}. \end{aligned}$$

The choice of Q was arbitrary, hence the above holds for all $Q \subseteq A^1_\xi$ and Eq. (4.26) follows. □

Corollary 4.2. *If $w_\xi(\emptyset) = w_\xi(N_\xi)$ (cf. Eq. 4.21) then $r_\xi = w_\xi(\emptyset)$.*

Proof. By definition, $w_\xi(\emptyset) \leq w_\xi(M) \leq w_\xi(N_\xi)$ for any $M \subseteq N_\xi$. Thus if $w_\xi(\emptyset) = w_\xi(N_\xi)$ then $w_\xi(\emptyset) = w_\xi(M)$ for all M. The value of the threshold r_ξ is given by Eq. (4.25); since $w_\xi(M) = w_\xi(\emptyset)$ for all M then $r_\xi = \max_{M \subseteq N_\xi}\{\min\{v_\xi(M), w_\xi(\emptyset)\}\}$. Now $v_\xi(\emptyset) = 1$ from Eq. (4.22); hence $\min\{v_\xi(\emptyset), w_\xi(\emptyset)\} = w_\xi(\emptyset)$. Moreover, for any non-empty $M \subseteq N_\xi$, $\max\{w_\xi(\emptyset), \min\{v_\xi(M), w_\xi(\emptyset)\}\} = w_\xi(\emptyset)$; so $r_\xi = w_\xi(\emptyset)$ as required. □

4.6.2 The Omnithermal Process

Let Υ denote the omnithermal process which keeps track of the omnithermal thresholds of the individuals in the free process Φ, ie. $\Upsilon(t) = \left\{r_{\tilde{\xi}};\ \tilde{\xi} \in \Phi(t)\right\}$. Following Dimakos (2000, Chapter 5) we

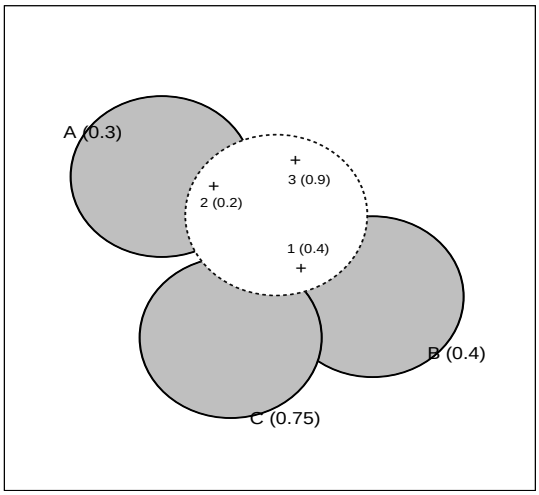

Figure 4.4: Computing the omnithermal threshold. Suppose the unfilled disk represents the individual ξ whose omnithermal threshold r_ξ is required. The neighbours of ξ are $N_\xi = \{A, B, C\}$, with the numbers in the brackets denoting their respective thresholds. $Z_\xi = (Y_\xi, U_\xi)$ is given as follows: the three crosses represent $Y_\xi = \left\{y_\xi^1, y_\xi^2, y_\xi^3\right\}$ and the numbers in the brackets $U_\xi = \left\{U_\xi^1, U_\xi^2, U_\xi^3\right\}$. Eq. (4.25) or (4.26) gives $r_\xi = 0.3$.

describe some properties Υ must satisfy. For fixed p let $\Psi_p^{\text{omni}}(t) = \{\xi;\ r_\xi \in \Upsilon(t) \text{ and } p \leq r_\xi\}$ be the collection of individuals obtained by *thresholding* $\Upsilon(t)$. If we initialize $\Upsilon(0)$ at some 'valid' omnithermal state (ie. at a collection of thresholds so that the configuration $\Psi_p^{\text{omni}}(0)$, for any $p \in [0,1]$, has positive probability under the corresponding area-interaction distribution), then $\Psi_p^{\text{omni}}(t) \equiv \Psi_p(t)$ must hold for all $t \geq 0$ and $p \in [0,1]$. Thus thresholding Υ at some p must yield the same process as that obtained by simulating the Ψ_p process initialized at $\Psi_p^{\text{omni}}(0)$ and using the same realization of Φ. This ensures that the omnithermal updates produce a proper Markov process; furthermore since, for any p, Ψ_p converges to the right distribution then Υ must also converge.

4.6.3 Omnithermal Sampling: Finite Time Construction

In order to simulate the omnithermal process Υ on a finite time interval, say $[0,T]$, simulate a realization of the free process Φ on the same interval and update the Υ process at each transition accordingly. Choose a 'valid' starting state for the omnithermal process, eg. let $\Upsilon(0) = \left\{u_\xi;\ \tilde{\xi} \in \Phi(0)\right\}$ where the $u_\xi \sim \text{Uniform}(0,1)$. Such a collection is a valid initial value since, for any $p \in [0,1]$, $\Psi_p^{\text{omni}}(0) = \{\xi;\ r_\xi \in \Upsilon(0) \text{ and } p \leq u_\xi\}$ is an independent p-thinning of $\left\{\xi;\ \tilde{\xi} \in \Phi(0)\right\}$, which has

a positive probability under the corresponding area-interaction process.

Algorithm 4.13 (Omnithermal Sampling: Finite Time Construction).

Draw $\Phi(0)$; initialize $\Upsilon(0) = \left\{u_\xi;\ \tilde{\xi} \in \Phi(0)\right\}$, where $u_\xi \sim \text{Uniform}(0,1)$; set $t = 0$.

while $t \leq T$:

Simulate the next incident time τ of Φ.

if next incident is a birth ξ: compute r_ξ; update $\Upsilon(\tau) = \Upsilon(\tau-) \cup \{r_\xi\}$.

else if next incident is a death ξ: update $\Upsilon(\tau) = \Upsilon(\tau-) \setminus \{r_\xi\}$.

update $t = t + \tau$.

return $\Upsilon(T)$.

The omnithermal construction Algorithm 4.13 is non-trivial. The omnithermal process Υ must be consistent with the construction in Algorithm 4.9. That is, thresholding $\Upsilon(T)$ at some $p \in [0,1]$ must yield the *same* collection of individuals as would have been obtained via Algorithm 4.9 with $\beta = p\overline{\beta}$ and using the same realization of the free process Φ. The key point to note is that Theorem 4.2 establishes an equivalence between r_ξ and $\{p;\ \xi \text{ is accepted in } \Psi_p\}$: $\xi \in \Psi_p(t_\xi)$ if and only if $p \leq r_\xi$. This ensures that Algorithm 4.13 above produces a proper Markov process Υ which converges to the required target distribution.

Theorem 4.3. *Suppose that $\Upsilon(T)$ is obtained by Algorithm 4.13 via a single realization of Φ. Let $\Psi_p(T)$, for $p \in [0,1]$, be obtained via Algorithm 4.9, with $\Psi_p(0) = \Psi_p^{omni}(0)$ and using the same realization of Φ as in the construction of Υ. Then $\Psi_p^{omni}(T) \equiv \Psi_p(T)$, for all $T \geq 0$. Furthermore $\Psi_p^{omni}(T)$ converges to an (attractive) area-interaction process with parameter $\beta = p\overline{\beta}$.*

Proof. For fixed p, $\Psi_p(0) = \Psi_p^{\text{omni}}(0)$ by construction. If, prior to a birth ξ, $\Psi_p^{\text{omni}}(t_\xi-) = \Psi_p(t_\xi-)$ then Theorem 4.2 ensures $\Psi_p^{\text{omni}}(t_\xi) = \{\eta;\ \tilde{\eta} \in \Phi(t_\xi),\ p \leq r_\eta\} \equiv \Psi_p(t_\xi)$. The same argument shows that this relation is maintained after all consecutive birth incidents. If there is a death incident in Φ, then the appropriate individual is deleted from Ψ_p, while its corresponding threshold is deleted from Υ. Thus deaths do not change the relation between Ψ_p^{omni} and Ψ_p; hence $\Psi_p^{\text{omni}}(t) = \Psi_p(t)$ for all $t \geq 0$, and in particular for $t = T$. Theorem 4.1 shows that $\Psi_p(T)$ converges, as $T \to \infty$, to an area-interaction process with parameter $\beta = p\overline{\beta}$; thus $\Psi_p^{\text{omni}}(T) = \Psi_p(T)$ must also converge to the same limit. Since the choice of p was arbitrary, the above holds for any $p \in [0,1]$. □

4.6.4 Omnithermal Sampling: Time Stationary Construction

In this section exact sampling of the omnithermal process Υ is tackled, via the Clan Algorithm (Section 4.5.5) of Fernández *et al.* (2002). If, for some individual ξ, the conditions of Corollary 4.2 are satisfied, ie. $w_\xi(\emptyset) = w_\xi(N_\xi)$, then in order to compute r_ξ one does not require the respective thresholds for the neighbours of ξ. So for omnithermal sampling, one can interpret ξ as having no first generation ancestor since it is possible to exactly determine r_ξ only from the associated marks $\left\{y_\xi^k, U_\xi^k\right\}$ of ξ (compare to Remark 4.2).

Thus the only alteration in computing the clan of ancestors (via Algorithm 4.11) is that the definition of a first generation ancestor is re-expressed in terms of the omnithermal notation. Hence, given a set of individuals $\left\{\xi;\ \tilde{\xi} \in \Phi(0)\right\}$, one obtains the clan A_ξ for each individual via Algorithm 4.11. If, for each individual ξ, A_ξ is finite then the set $\{\eta \in A_\xi;\ \eta \text{ has no first generation ancestors}\}$ is non-empty and the thresholds for the individuals in this set can be obtained by Corollary 4.2. A sufficient condition for finiteness of A_ξ in the omnithermal setting is given by Eq. (4.19) with $\beta = \overline{\beta}$. Furthermore, given $\{\eta \in A_\xi;\ \eta \text{ has no first order ancestors}\}$, it is then possible to determine the thresholds for all the other individuals in A_ξ via Eq. (4.25). In order to determine $\Upsilon(0) = \left\{r_\xi;\ \tilde{\xi} \in \Phi(0)\right\}$ a variant of the Cleaning Algorithm 4.12 needs to be employed.

Algorithm 4.14 (Omnithermal Thresholding Algorithm).

Set $R_\xi = \emptyset$ and $H_\xi = A_\xi \cup \{\xi\}$.

for each individual in $\{\eta \in A_\xi;\ \eta \text{ has no first generation ancestors}\}$:

compute its corresponding threshold r_η; update $R_\xi = R_\xi \cup \{r_\eta\}$; $H_\xi = H_\xi \setminus \{\eta\}$.

Order the remaining individuals in H_ξ by time of birth, with the oldest first.

while $H_\xi \neq \emptyset$:

let the first individual in H_ξ be ζ and A_ζ^1 its set of first generation ancestors.

(by construction, R_ξ contains the thresholds for all individuals in A_ζ^1).

compute r_ζ via Eq. $(4.25 \text{ or } 4.26)$; update $R_\xi = R_\xi \cup \{r_\zeta\}$; $H_\xi = H_\xi \setminus \{\zeta\}$.

return R_ξ.

The thresholds of the individuals in $\{\eta \in A_\xi;\ \eta \text{ has no first generation ancestors}\}$ are determined exactly by Corollary 4.2. Given these, Theorem 4.2 ensures that the use of Eq. (4.25) gives the correct thresholds for any descendant. So $\Upsilon(0) = \left\{r_\xi \in R_\xi;\ \tilde{\xi} \in \Phi(0)\right\}$ must be an exact draw

from the equilibrium distribution of Υ. Alternatively the distribution of $\Psi_p^{\text{omni}}(0)$, for any p, must be that of an area-interaction process with parameter $\beta = p\overline{\beta}$.

4.7 Conclusions & Further Work

This chapter concerned the perfect simulation of the area-interaction process, introduced by Baddeley & Van Lieshout (1995) as a model for both clustering (attraction) and orderliness (repulsion). Various algorithms for this process are described; in the case of the Kendall (1998) Algorithm 4.10 and Fernández *et al.* (2002) Clan Algorithm 4.11 explicit details on practical implementation are also discussed. An important contribution of this chapter is the *omnithermal* sampling of the area-interaction process. It is seen how spatial birth-death processes allow sampling the process, not only at fixed parameters, but also simultaneously for a whole range of values of the interaction parameter β. Moreover Lemma 4.3 and subsequent Remark 4.4 show that omnithermal sampling is possible only for the attractive case ($\beta > 0$). A variant of the Clan Algorithm 4.11 enabled an exact omnithermal construction (Algorithm 4.14). The following section discuss various possibilities for extending and generalizing the ideas in this chapter.

4.7.1 Perfection in Space

The perfect simulation algorithms introduced in this chapter sample the 'standard' area-interaction process on some bounded region $W \subset \mathbb{R}^2$, *conditional* on the assumed boundary conditions (cf. Section 4.2.1). Baddeley & Van Lieshout (1995) show that the standard process can be viewed as the restriction to W of a stationary point process on the whole of $\mathbb{R}^2$. A natural extension of the work presented here would be to consider sampling the stationary process on W. Kendall (1997b) calls this "perfect simulation in space"; Fernández *et al.* (2002) use the term "sampling the infinite-volume measure". Indeed the Clan algorithm, as presented in Fernández *et al.* (2002), can be employed to sample the infinite volume measure directly.

The description given by Kendall (1997b) involves the construction of upper and lower processes which are extended not only backwards in time but in *space* as well. The basic idea is as follows. Choose a sequence of ever increasing sampling windows $W_n \supset W$ for all $n \geq 0$. For

$n = 0, 1, \ldots$ sample the standard area-interaction process on W_n subject to (i) zero ('minimal'); and (ii) rectangular ('maximal') boundary conditions (Section 4.2.1), but using the same realization of the dominating marked free process (cf. Section 4.5). If both samples agree then the common configuration is a sample of the stationary process. If not set $n = n + 1$ and carry out the procedure again. This procedure will terminate almost surely in finite time if the underlying marked free process of space-time Boolean model "does not percolate to infinity, since then for large enough n the boundary conditions cannot affect the pattern in W" (Kendall 1997a; Kendall 1997b).

It would be interesting not only to carry out such an implementation of 'perfection in space (and time)' but also to combine it with the omnithermal sampling ideas of Section 4.6 in order to draw omnithermally from the stationary process. Indeed it may be possible to incorporate the dynamics of the Omnithermal Thresholding Algorithm 4.13 with the Clan algorithm, as presented in Fernández *et al.* (2002), in order to do so.

4.7.2 Omnithermal Sampling

Although explicit details of implementing the exact omnithermal algorithm of Section 4.6.4 are given, practical implementation has not been carried out. Therefore this presents another viable direction for further work. In addition it seems to us that it may also be possible to devise a CFTP-based construction for the same sampling procedure. Another generalization would be to sample the process simultaneously for any single (or even multiple) parameter(s). The parameter considered here was the interaction parameter β; it seems relatively straightforward to do this for different values of the underlying intensity λ and, possibly, the radius r of the disks which define the Boolean model. Some additional thoughts on the possibility of omnithermal sampling for two or more parameters are discussed in Section 6.1.

Chapter 5

Conditional Area-Interaction Point Process

5.1 Introduction

In Chapters 3 & 4 perfect simulation algorithms for conditional Boolean models and the area-interaction point process were presented. In this chapter we consider a natural extension of the two models: the *conditional area-interaction point process*. The only other reference to an area-interaction process required to cover a given set of nodes is found in McKeague & Loizeaux (2002). They consider Bayesian cluster models for disease outbreaks. It turns out that for a class of cluster models the posterior distribution is an area-interaction process, conditioned to cover the locations of disease outbreak. Their simulation method involved straightforward Rejection sampling; moreover they comment on the inefficiency of such an accept/reject mechanism based on their simulations. We have found no other literature mentioning the conditional area-interaction process; therefore the task of developing perfect simulation algorithms for this model is tackled here.

There are numerous applications for both the conditional Boolean model and area-interaction process. Hence it is reasonable to expect that the conditional area-interaction process itself has an equally broad range of applicability. For example the area-interaction process is a plausible model for the "selfish herd" (cf. Section 4.1). Water particles in clouds may be attracted to regions of higher concentration of cloud cover, hence an attractive area-interaction process may serve as a useful model. As noted in Section 3.1 there is often experimental data available for the phenomenon being studied. This may include information on which regions are covered/uncovered or connected/disconnected by the regions of influence of the individuals. For example regions where

food/water or other resources are quite likely to be covered by the "selfish herd" above. It is both natural and logical to incorporate such information, which thus leads to *conditional models*. For the purposes here we consider the area-interaction process conditioned to cover a finite set of nodes.

This chapter concerns the perfect simulation of the conditional area-interaction process on a bounded window $W \subset \mathbb{R}^2$. The aim here is two-fold: firstly, to introduce three perfect simulation algorithms for this model; and secondly to evaluate the algorithms by comparing their empirical run times. Rejection sampling for this process is briefly introduced; a 2-Stage Rejection algorithm is described for the conditional area-interaction process. In Section 5.4 Algorithm 5.1, which employs spatial birth-death processes, is presented. The method is a hybrid between the Cai & Kendall (2002) and Kendall (1998) procedures.

In Section 5.5 a Gibbs-within-Metropolis Hastings sampler (Algorithm 5.4) is proposed and it is shown that its output is a perfect sample of a conditional area-interaction process. As noted in Section 3.1 it is important to validate any simulation algorithm, even though *in theory* it outputs samples with the required distribution. There may be programming bugs which need to be checked; in addition theoretical discrepancies, if any, may also be uncovered. In Appendix B the modified Cai & Kendall and exact Gibbs algorithms are validated via χ^2 tests.

In Section 5.7 a quantitative evaluation between the 2-Stage Rejection, modified Cai & Kendall and exact Gibbs algorithms is carried. The objective is the same as in the experiments of Section 3.7: the actual run times of the three algorithms are compared for various model parameters. The basic conclusion is that the modified Cai & Kendall procedure is more efficient for extreme parameter values (high λ, β, large window and conditioning set size). The other two algorithms fare well for moderate values; the exact Gibbs sampler seems just as efficient as, if not more efficient than 2-Stage Rejection.

5.2 Notation

Let $W \subset \mathbb{R}^2$ be a bounded sampling window and $\mathcal{C} = \{c_1, \ldots, c_k\}$ be the finite conditioning set such that $\mathcal{U}(\mathcal{C}) \subset W$. Here $\mathcal{U}(\mathcal{C}) = \bigcup_{\{c \in \mathcal{C}\}} \mathcal{U}(c)$, where $\mathcal{U}(\xi)$ denotes a disk of fixed radius r and centre ξ (Definition 3.1). Let X denote an area-interaction process on W conditioned to cover the set $\mathcal{C}$. As noted in Section 3.2 points ξ_n of X will be referred to as *germs*, $\{\mathcal{U}(\xi_n)\}$ as *grains* and

$\{c_i\}$ as *(conditioning) nodes*. As with the area-interaction process (and, in general, Gibbs processes on bounded regions), the distribution of X on W must be specified *conditional on the configuration outside* W (cf. Sections 1.1.10 & 4.2.1). Let y denote the point configuration outside W; the density of X with respect to a unit rate Poisson process, given the boundary condition y, is

$$f(x \mid y) \propto \lambda^{n(x)} e^{-\beta\, \psi(x|y)} \mathbf{1}_{\{\mathcal{U}(x) \supset \mathcal{C}\}}, \qquad \text{where } \psi(x \mid y) = m_2 \left[\mathcal{U}(x) \setminus \mathcal{U}(y)\right]. \tag{5.1}$$

It is assumed here that the process y on $W^c = \mathbb{R}^2 \setminus W$ is empty, ie. 'zero' boundary conditions (cf. Section 4.2.1). Thus Eq. (5.1) simplifies to $f(x) \propto \lambda^{n(x)} e^{-\beta\, m_2[\mathcal{U}(x)]} \mathbf{1}_{\{\mathcal{U}(x) \supset \mathcal{C}\}}$. The parameter β is referred to as the *area-interaction* parameter; $\beta = 0$ corresponds to a Poisson(λ) process while $\beta > 0$ ($\beta < 0$) produces clustered (regular) patterns. Figure 5.1 below shows a realization of a conditional attractive area-interaction process. The conditioning set (part of Redwood seedlings data, Figure 1) is the same as that used in Figure 3.1.

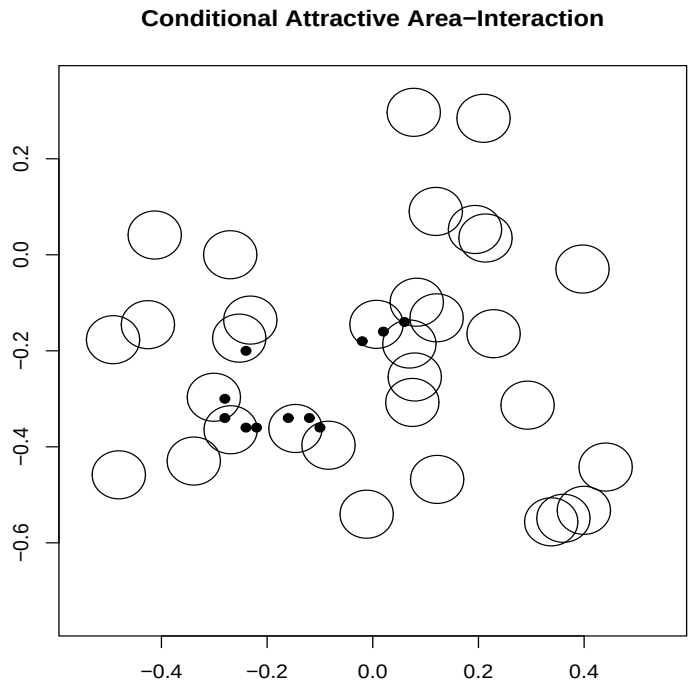

Figure 5.1: Sample of a conditional attractive area-interaction process. The conditioning nodes (solid dots) are a part of the Redwood seedlings data set (Figure 1). The underlying Poisson intensity $\lambda = 40$, interaction parameter $\beta = 25$, disk radius $r = 0.05$ and $W = [-0.5, 0.5] \times [-0.7, 0.3]$.

5.3 Rejection Sampling

Rejection sampling for this model is straightforward. One can employ the Rejection procedures (Algorithms 4.1 & 4.2) in order to draw a sample of an area-interaction process. If it covers $\mathcal{C}$ then it has the required distribution; otherwise the sample is discarded and another is drawn independently. This procedure is carried out until one obtains the desired sample. Alternatively, a Poisson process conditioned to cover $\mathcal{C}$ can first be drawn and then accepted with the correct weight; algorithms 4.1 & 4.2 describe how to obtain such a weighted Poisson process. Moreover the sample from the conditional Poisson process can be obtained via the 2-Stage procedure (Algorithm 3.2), since this is going to be much more efficient that ordinary Rejection sampling. For purposes of practical implementation, we choose this latter variant of Rejection sampling: the 2-Stage Rejection algorithm is employed to sample conditional Boolean model, which is then accepted with the correct probability.

5.4 Simulation via Spatial Birth-Death Processes

In Sections 3.4 & 4.5.4 spatial birth-death processes were used to sample the conditional Boolean model and area-interaction process respectively. The idea being to set the birth rate equal to the Papangelou conditional intensity of the point process and unit death rate per point. These birth and death rates automatically satisfy detailed balance (Eq. 1.25) with respect to the density of the point process; the birth-death process hence converges to the required point process.

Simulation of the conditional Boolean model (Section 3.4) involves forbidding deaths which violate the conditioning; the constraint so imposed applies only to death incidences. Simulation of the area-interaction process (Section 4.5) involves censoring birth transitions in order to generate the required birth rates; the constraint so imposed applies only to birth incidences. Therefore there is no conflict between the respective censoring rules generating the two models. This suggests that censoring deaths and births accordingly would produce an interacting spatial birth-death process which converges to the conditional area-interaction process. The Cai & Kendall (2002) and Kendall (1998) constructions are briefly summarized below and Section 5.4.1 then explicitly details how to obtain a perfect simulation recipe for the conditional area-interaction process (Algorithm 5.1).

Review of Perfect Simulation: Conditional Boolean Model

Algorithm 3.4 of Cai & Kendall (2002), as presented in Section 3.4.1, is reviewed here. The set $\mathcal{C} = \{c_1, \ldots, c_k\}$ denotes the conditioning nodes; for non-empty $A \subseteq \mathcal{C}$, E_A defined by Eq. (3.4) represents the region such that any germ in E_A covers all the nodes in A and no others. Let Φ denote a spatial birth-death process with birth rate λ and unit death rate per germ. Denote by Φ_A the restriction of Φ to E_A; then the basic processes which will drive the simulation are $\{\Phi_A;\ A \subseteq \mathcal{C}\}$, which converge to Poisson(λ) processes on E_A. Death transitions in the Φ_A which uncover any conditioning node are disallowed. Such transitions are called *perpetuated germs* and stored in a respective *virtual* process V_A; hence the virtual processes essentially supply extra individuals to ensure that the conditioning is always maintained. The exact coupling construction defines upper and lower virtual processes $\left\{V_A^{T,max}\right\}$ and $\left\{V_A^{T,min}\right\}$. Given transitions of the driving processes $\{\Phi_A\}$, these bounding virtual processes are updated accordingly in order to ensure that they satisfy the sandwiching, coalescence and funnelling properties (Lemmas 3.2, 3.3 & 3.4).

Review of Perfect Simulation: Area-Interaction

The density of the area-interaction process is locally stable (see Section 4.5.4) and therefore the dominated CFTP recipe (Section 1.4.2) of Kendall & Møller (2000) can be used to sample the process perfectly. An unconditional Poisson process *stochastically dominates* an area-interaction process. Hence it is possible to define coupled spatial birth-death processes (Φ, Ψ) such that $\Phi(t)$ is a Poisson(λ) process, $\Psi(t)$ an area-interaction process and $\Phi(t) \supseteq \Psi(t)$ for all t. Births of the 'basic' process Φ have appropriate marks attached which allow the correct censoring of births in the 'interacting' process Ψ. Deaths in Φ are always accepted in Ψ, if present. A perfect variant of this method is devised via dominated CFTP (cf. Section 1.4.2).

5.4.1 Dominated CFTP Construction

The algorithms reviewed above employ spatial birth-death processes in order to sample the respective distributions. In either case, transitions of the underlying driving process (an unconditional spatial birth-death process with Poissonian equilibrium) were censored accordingly in order to generate the appropriate equilibrium. As noted in Section 5.4 there is no conflict between the censoring

rules which generate the conditional Boolean model and the area-interaction process. Therefore accepting births with the correct probability and restricting those deaths which uncover the conditioning set will yield an interacting spatial birth-death process that converges to an area-interaction process conditioned to cover $\mathcal{C}$.

Recall from Eq. (5.1) that the density of a conditional area-interaction process X on W, with zero boundary conditions, is given by $f(x \mid \emptyset) \equiv f(x) \propto \lambda^{n(x)} e^{-\beta\psi(x|\emptyset)} \mathbf{1}_{\{\mathcal{U}(x)\supset\mathcal{C}\}}$, where $\psi(x \mid y) = m_2[\mathcal{U}(x) \setminus \mathcal{U}(y)]$. The Papangelou conditional intensity of X is

$$\ell(\xi; x) = \lambda\, e^{-\beta\ \psi(\xi|x)} \mathbf{1}_{\{\mathcal{U}(x)\supset\mathcal{C}\}} = \lambda\, e^{-\beta\ m_2[\mathcal{U}(\xi)\setminus\mathcal{U}(x)]} \mathbf{1}_{\{\mathcal{U}(x)\supset\mathcal{C}\}}.$$

In Section 1.3.2 it was seen that spatial birth-death processes can be employed to sample point process specified by density with respect to a Poisson process. Therefore a birth-death process $\tilde{\Psi}$ with birth rate $b(x, \xi) = \lambda\ e^{-\beta\ \psi(\xi|x)} \mathbf{1}_{\{\mathcal{U}(x)\supset\mathcal{C}\}}$ and unit death rate per point satisfies detailed balance Eq. (1.25) with respect to the density f. If the process is irreducible then $\tilde{\Psi}$ converges to a conditional area-interaction process. Irreducibility follows since the state space of $\tilde{\Psi}$ is all configurations which cover $\mathcal{C}$ and the process can go from any configuration x to another y via birth incidents of germs in y followed by death incidents of germs in x.

In order to devise an exact coupling we combine the constructions in Algorithms 3.4 & 4.10 as follows. Attention here is restricted to the attractive conditional area-interaction process $(\beta > 0)$; minor changes required for the repulsive case $(\beta < 0)$ are discussed in Remark 5.1. A simple modification of Algorithm 3.4 will yield exactly what we want. Specifically, censoring births with the correct acceptance probability will ensure the correct transition rates for the target process. Births need to be accepted with some non-trivial probability which involves computing the area of complicated regions. Sections 4.5 & 4.5.4 describe how to implement the required censoring via observable events that have the same occurrence probability.

Other modifications to Algorithm 3.4 involve defining extra maximal and minimal processes and initializing these. For a given $A \subseteq \mathcal{C}$ recall that $E_A \subseteq W$ is such that any germ in this region covers only those conditioning nodes in A and no others. Let $\preccurlyeq$ be a partial order on the space of point patterns defined by $x \preccurlyeq x'$ if $\mathcal{U}(x) \subseteq \mathcal{U}(x')$. Define $\overline{x}_A$ as a collection of germs such that $\mathcal{U}(\overline{x}_A) \supset \mathcal{U}(E_A)$, where $\mathcal{U}(E_A) = \{\eta \in \mathbb{R}^2;\ \eta \in \mathcal{U}(\xi)\ \text{for some}\ \xi \in E_A\}$. This implies that for any configuration y in E_A: $y \preccurlyeq \overline{x}_A$; moreover $\bigcup_{\{A\subseteq\mathcal{C}\}} \overline{x}_A$ covers $\mathcal{C}$. The driving processes in

Algorithm 3.4 are the spatial birth-death processes $\{\Phi_A\}$ (birth rate λ, unit per capita death rate) and the unit-rate Poisson processes $\{Z_A\}$ on $\mathbb{R}_+$. Let $\Psi_A^{T,max}$, $\Psi_A^{T,min}$ denote the bounding *interacting* and $V_A^{T,max}$, $V_A^{T,min}$ the bounding *virtual* processes. Define the *target* processes by:

$$\tilde{\Psi}_c^{T,min}(t) = \bigcup_{\{A;\ c\in A\}} \tilde{\Psi}_A^{T,min}(t) = \bigcup_{\{A;\ c\in A\}} \Psi_A^{T,min}(t) \cup V_A^{T,min}(t). \tag{5.2}$$

$$\tilde{\Psi}_c^{T,max}(t) = \bigcup_{\{A;\ c\in A\}} \tilde{\Psi}_A^{T,max}(t) = \bigcup_{\{A;\ c\in A\}} \Psi_A^{T,max}(t) \cup V_A^{T,max}(t). \tag{5.3}$$

Algorithm 5.1 (domCFTP: Attractive Conditional Area-Interaction).

Fix $T > 0$.

while $T > 0$:

Extend $\{\Phi_A;\ A \subseteq \mathcal{C}\}$, $\{Z_A;\ A \subseteq \mathcal{C}\}$ backwards on the interval $\left[-T, -\lfloor \frac{T}{2} \rfloor\right)$.

for each non-empty A initialize:

$\Psi_A^{T,min}(-T) = V_A^{T,min}(-T) = \emptyset$.

$\Psi_A^{T,max}(-T) = \{\xi; \xi \in \Phi_A(-T)\}$; $V_A^{T,max}(-T) = \overline{x}_A$.

for $t \in [-T, 0]$:

if t is the birth time of $\xi \in E_A$:

set $\Phi_A(t) = \Phi_A(t-) \cup \{\xi\}$.

set $\Psi_A^{T,min}(t) = \Psi_A^{T,min}(t-) \cup \{\xi\}$ with probability $e^{-\beta\ \psi\left(\xi|\tilde{\Psi}_A^{T,min}(t-)\right)}$.

set $\Psi_A^{T,max}(t) = \Psi_A^{T,max}(t-) \cup \{\xi\}$ with probability $e^{-\beta\ \psi\left(\xi|\tilde{\Psi}_A^{T,max}(t-)\right)}$.

else if t is a death time of $\xi \in \Phi_A(\cdot)$:

set $\Phi_A(t) = \Phi_A(t-) \setminus \{\xi\}$.

set $\Psi_A^{T,min}(t) = \Psi_A^{T,min}(t-) \setminus \{\xi\}$; $\Psi_A^{T,max}(t) = \Psi_A^{T,max}(t-) \setminus \{\xi\}$.

if $\tilde{\Psi}_{c_i}^{T,min}(t-) = \{\xi\}$ for some $c_i \in A$: set $V_A^{T,max}(t) = \{\xi\}$.

if $\tilde{\Psi}_{c_i}^{T,max}(t-) = \{\xi\}$ for some $c_i \in A$: set $V_A^{T,min}(t) = \{\xi\}$.

else if t is a perpetuation time for $V_A(\cdot)$ (ie. incident time of Z_A):

if $\tilde{\Psi}_{c_i}^{T,min}(t-) \supset V_A^{T,min}(t-)$ for all $c_i \in A$: set $V_A^{T,max}(t) = \emptyset$.

if $\tilde{\Psi}_{c_i}^{T,max}(t-) \supset V_A^{T,max}(t-)$ for all $c_i \in A$: set $V_A^{T,min}(t) = \emptyset$.

if $\tilde{\Psi}_{c_i}^{T,min}(0) = \tilde{\Psi}_{c_i}^{T,max}(0)$ for all $i = 1, \ldots, k$: set $T = -1$.

else: set $T = 2T$.

return $\left\{\tilde{\Psi}_{c_i}^{T,min}(0)\right\}_{i=1}^{k}$.

Lemma 5.1 (Sandwiching). *For non-empty $A \subseteq \mathcal{C}$ let $\tilde{\Psi}_A = \Psi_A \cup V_A$ denote a process started at time $-\infty$ and, for each $T > 0$, coupled to the evolution of $\{\Phi_A\}$ and $\{Z_A\}$ on $[-T, 0]$. The equilibrium distribution of the target process $\tilde{\Psi} = \bigcup_{\{c_i \in \mathcal{C}\}} \tilde{\Psi}_{c_i}$ is an area-interaction process, with parameters λ and $\beta > 0$, conditioned to cover $\mathcal{C}$. Let the bounding interacting processes $\Psi_A^{T,min}$, $\Psi_A^{T,max}$ and virtual processes $V_A^{T,min}$, $V_A^{T,max}$ be defined recursively as in Algorithm 5.1. Furthermore if, for all $T > 0$, the evolution of $\left\{\tilde{\Psi}_A\right\}$, $\left\{\Psi_A^{T,min}, \Psi_A^{T,max}\right\}$ and $\left\{V_A^{T,min}, V_A^{T,max}\right\}$ is coupled to the same realizations of $\{\Phi_A\}$ and $\{Z_A\}$ on $[-T, 0]$ then, for all A and $-T \leq t \leq 0$:*

$$\Psi_A^{T,min}(t) \subseteq \Psi_A(t) \subseteq \Psi_A^{T,max}(t) \subseteq \Phi_A(t); \tag{5.4}$$

$$V_A^{T,min}(t) \subseteq V_A(t) \subseteq V_A^{T,max}(t); \tag{5.5}$$

and hence
$$\tilde{\Psi}_A^{T,min}(t) \subseteq \tilde{\Psi}_A(t) \subseteq \tilde{\Psi}_A^{T,max}(t). \tag{5.6}$$

Proof. The result is shown by induction along the sequence of birth, death and perpetuation incidents on the time interval $[-T, 0]$. By construction, $\emptyset = \Psi_A^{T,min}(-T) \subseteq \Psi_A(-T) \subseteq \Psi_A^{T,max}(-T) = \Phi_A(-T)$ and $\emptyset = V_A^{T,min}(-T) \subseteq V_A(-T) \subseteq V_A^{T,max}(-T) = \overline{x}_A$ for all A. Note the $V_A(-T)$, by definition, can have at most one perpetuated germ (cf. Section 3.4.1). Now suppose that Eqs. $(5.4 \,\&\, 5.5)$ hold up to some $t > -T$.

Suppose that time t is a birth time in some Φ_A of individual ξ. The virtual processes are unaffected by births so Eq. (5.5) persists at time t. Since $\beta > 0$ the acceptance probability for ξ in a configuration x is monotonic in x, ie. $e^{-\beta\psi(\xi|y)} \leq e^{-\beta\psi(\xi|x)}$ for $y \subseteq x$, hence Eq. (5.4) will also hold at time t. Death transitions and perpetuation incidents (ie. incidents of the Z_A processes) affect only the virtual processes. The 'cross-over' trick (cf. Lemma 3.2) ensures that both death and perpetuation incidents maintain the ordering between the virtual processes. Thus the result follows by induction along the sequence of transitions on $[-T, 0]$. □

Lemma 5.2 (Coalescence). *If $\Psi^{T,min}(t^*) = \Psi^{T,max}(t^*)$ and $V^{T,min}(t^*) = V^{T,max}(t^*)$, for some $t^* > -T$, then these equalities persist over all $[t^*, 0]$.*

Proof. At time t^* the upper and lower processes all coalesce so that there is no longer any distinction between upper and lower processes at the next birth, death or perpetuation incident. Consequently the transitions for both the upper and lower processes will be identical. □

Lemma 5.3 (Funnelling). *Let $\Psi_A^{T,min}, \Psi_A^{T,max}$ and $V_A^{T,min}, V_A^{T,max}$ be defined as in Lemma 5.1 and updated according to Algorithm 5.1; then for all A and $-S \leq -T \leq t \leq 0$*

$$\Psi_A^{T,min}(t) \subseteq \Psi_A^{S,min}(t) \subseteq \Psi_A^{S,max}(t) \subseteq \Psi_A^{T,max}(t); \tag{5.7}$$

$$V_A^{T,min}(t) \subseteq V_A^{S,min}(t) \subseteq V_A^{S,max}(t) \subseteq V_A^{T,max}(t). \tag{5.8}$$

Proof. The proof follows by induction along the sequence of transitions on $[-T, 0]$. By construction the relationships hold at time $t = -T$. So suppose Eqs. $(5.7\ \&\ 5.8)$ hold up to some time $t > -T$. If t is the birth time of ξ in some Φ_A, then by virtue of the monotonicity of the acceptance probability (cf. proof of Lemma 5.1) the ordering in Eq. (5.7) persists at time t. Furthermore, the arguments in the proof of Lemma 3.4 ensure that deaths or perpetuation incidents also maintain the ordering in Eq. (5.8). The result hence follows. □

Lemma 5.4. *For all sufficiently large T the upper and lower target processes $\tilde{\Psi}_{c_i}^{T,min}$ and $\tilde{\Psi}_{c_i}^{T,max}$ coalesce for all $c_i \in \mathcal{C}$ at some time in the interval $[-T, 0]$.*

Proof. The result follows, by virtue of Eqs. $(5.2\ \&\ 5.3)$, if *both* the bounding interacting $\Psi_A^{T,max}$, $\Psi_A^{T,min}$ *and* virtual processes $V_A^{T,max}$, $V_A^{T,min}$ coalesce for sufficiently large T. The argument follows similar lines as the proof of Lemma 3.5: it is shown that coalescence of the two sets of upper and lower processes occurs with a fixed positive probability; the second Borel-Cantelli Lemma then shows that coalescence happens infinitely often.

Fix $-T \leq t \leq 0$ and $\delta > 0$; let $B = \{\text{all } \Phi_A \text{ processes hit } \emptyset \text{ on } [t, t+\delta]\}$. Since $\emptyset$ is an ergodic atom of each Φ_A processes, which are restricted to a bounded region $E_A \subseteq W$, it follows that each Φ_A hits $\emptyset$ infinitely often. Thus $\mathbb{P}[B] = \rho_1 > 0$, independent of t. Births in each Φ_A process are censored in the corresponding $\Psi_A^{T,max}$ and $\Psi_A^{T,min}$ processes so that

$$\Psi_A^{T,min}(t) \subseteq \Psi_A^{T,max}(t) \subseteq \Phi_A(t), \qquad \text{for all } A, t.$$

Therefore, conditional on B occurring at time $t+\delta$, $\Psi_A^{T,min}(t+\delta) = \Psi_A^{T,max}(t+\delta) = \emptyset$ for all A. Set $D_i = \left\{\Phi_{c_i} \text{ experiences a birth in } [t+\delta, t+2\delta] \text{ which is accepted in } \Psi_{c_i}^{T,min} \text{ and } \Psi_{c_i}^{T,max}\right\}$. The birth-acceptance probability of ξ in a given pattern x is $e^{-\beta\psi(\xi|x)} \geq e^{-\beta\psi(\xi|\emptyset)} = e^{-\beta\, m_2[\xi]}$ since $\beta > 0$. Therefore $\mathbb{P}[D_i] \geq e^{-\beta\, m_2[\xi]} \times \mathbb{P}[\text{all the } \Phi_A \text{ experience a birth in } [t+\delta, t+2\delta]]$. The arguments in the proof of Lemma 3.5 thus show that $\mathbb{P}\left[\bigcap_{i=1}^k D_i\right] = \rho_2 > 0$, independent of t.

Now, conditional on $\bigcap_{i=1}^{k} D_i$ occurring at time $t + 2\delta$, the $V_A^{T,max}$ processes will vanish if all the Z_A processes experience an incident *and* none of the Φ_A experience a death incident in the time interval $[t + 2\delta, t + 3\delta]$. This is because all the $\Psi_{c_i}^{T,min}$ and $\Psi_{c_i}^{T,max}$ will then be non-empty when the incidents of the Z_A processes occur. Again the arguments in the proof of Lemma 3.5 show that (given $\bigcap_i D_i$) the $V_A^{T,max}$ all vanish with some probability $\rho_3 > 0$. Putting all this together, if $C_t = B \bigcap_i D_i \bigcap \left\{\text{all the } V_A^{T,max} \text{ vanish}\right\}$, then $\mathbb{P}[C_t] > 0$. Moreover C_t is independent of events that occur before time t. The second Borel-Cantelli Lemma then ensures that infinitely many of the $C_{3s\delta}$ (for $s = 1, 2, \ldots$) occur. Hence coalescence of the target processes $\tilde{\Psi}_A^{T,min}$ and $\tilde{\Psi}_A^{T,max}$ is almost sure to happen in finite time. □

Theorem 5.1. *Algorithm 5.1 terminates almost surely in finite time and the distribution of the output is that of an (attractive) area-interaction processes conditioned to cover the set $\mathcal{C}$.*

Proof. The Theorem follows as a consequence of Lemmas 5.1, 5.2, 5.3 & 5.4. By virtue of Eq. (5.6), conditional upon coalescence, the common value of $\tilde{\Psi}_A^{T,min}(0)$ and $\tilde{\Psi}_A^{T,max}(0)$ must be equal to that of $\tilde{\Psi}_A(0)$ for all A. But $\left\{\tilde{\Psi}_{c_i}\right\}_{i=1}^{k}$ are the target processes started at time $-\infty$; hence $\bigcup_{i=1}^{k} \tilde{\Psi}_{c_i}(0)$ must be a perfect draw from an attractive area-interaction process conditioned to cover $\mathcal{C}$. □

Remark 5.1. For the repulsive case ($\beta < 0$) the birth rate of the Φ_A has to be changed to $\lambda\, e^{-\beta m_2[\mathcal{U}(\xi)]}$. Birth acceptance decisions for the lower process are made according to the state of the upper and vice versa (cf. Section 4.5.4). This 'cross-over' trick ensures the sandwiching, coalescence and funnelling properties (cf. Lemmas 5.1, 5.4 & 5.3) hold for the repulsive case.

5.5 Gibbs Sampling

In the previous section continuous-time spatial birth-death processes were employed to sample a conditional area-interaction process. It is also possible to devise a Gibbs sampler for this model; it can be thought of as a Gibbs-within-Metropolis Hastings sampler. Each component, given the others, is updated from the conditional distribution; sampling the conditional distribution itself involves a proposal and an accept/reject step.

The set of conditioning nodes to be covered is denoted by $\mathcal{C} = \{c_1, \ldots, c_k\}$. For each $A \subseteq \mathcal{C}$ the region E_A defined in Eq.(3.4) is such that a germ in E_A covers only those conditioning nodes in

A and no others. Due the sequential nature of the Gibbs updating scheme, it is convenient to consider the set $\{A;\ A \subseteq \mathcal{C}\}$ as an ordered set $\{A_0, \ldots, A_N\}$ so that each $i \in \{0, \ldots, N\}$ corresponds uniquely to some $A \subseteq \mathcal{C}$, with $A_0 = \emptyset$. Here N is the number of non-empty subsets of $\mathcal{C}$. Consider a discrete-time random process $\tilde{\Psi} = \left(\tilde{\Psi}_0, \ldots, \tilde{\Psi}_N\right)$, where $\tilde{\Psi}_i$ is a random point configuration on cell E_{A_i}. Let the stationary distribution of $\tilde{\Psi}$ be that of an attractive area-interaction point process conditioned to cover the set $\mathcal{C}$. The underlying Poisson intensity is λ, the area-interaction parameter is β and the density of $\tilde{\Psi}$ on some bounded $W \subset \mathbb{R}^2$ is given by Eq. (5.1). It is assumed that the pattern outside W is empty. We first consider simulation of the attractive process $(\beta > 0)$. Minor adjustments are required for the repulsive case $(\beta < 0)$; these are discussed in section 5.5.2.

Consider a virtual simulation of the $\tilde{\Psi}$ started at time $-\infty$ via Gibbs sampling. For each $n \in \mathbb{Z}_-$ and $i \in \{0, \ldots, N\}$, $\tilde{\Psi}_i(n)$ is updated by drawing from the conditional distribution of the process on cell i given the processes in the other cells, $\tilde{\Psi}_{-i}(n) = \bigcup_{j<i} \tilde{\Psi}_j(n) \ \bigcup_{j>i} \tilde{\Psi}_j(n-1)$. The conditional density f_i of $\tilde{\Psi}_i$, given $\tilde{\Psi}_{-i}(n) = x$, is just the density of an area-interaction process on E_{A_i} conditioned to cover $A_i \subseteq \mathcal{C}$ with boundary conditions x (cf. Eq. 5.1); hence

$$f_i(\cdot \mid x) \propto \begin{cases} \lambda^{n(\cdot)} e^{-\beta\ \psi(\cdot|x)} & \text{if } \mathcal{U}(x) \supset A_i; \\ \lambda^{n(\cdot)} e^{-\beta\ \psi(\cdot|x)} \mathbf{1}_{\{n(\cdot)>0\}} & \text{else;} \end{cases} \tag{5.9}$$

$$\text{where} \qquad \psi(\cdot \mid x) = m_2\left[\mathcal{U}(\cdot) \setminus \mathcal{U}(x)\right].$$

For $j = 1, 2, \ldots$ let $\left(X_j(n,i), X'_j(n,i)\right)$ denote a tuple of marked Poisson(λ) processes at time n on E_{A_i} such that $X_j(n,i) \subseteq X'_j(n,i)$ and $n\left(X'_j(n,i)\right) > 0$. The mark of $X_j(n,i)$, $Y_j(n,i)$, is a Poisson(β) process on $\mathcal{U}(X_j(n,i))$. Similarly $X'_j(n,i)$ is marked by a Poisson(β) process on $\mathcal{U}\left(X'_j(n,i)\right)$, denoted by $Y'_j(n,i)$; furthermore $Y_j(n,i) \subseteq Y'_j(n,i)$. Notice that

$$\begin{aligned} \mathbb{P}\left[Y_j(n,i) \bigcap \left(\mathcal{U}(X_j(n,i)) \setminus \mathcal{U}\left(\tilde{\Psi}_{-i}(n)\right)\right) = \emptyset\right] &= e^{-\beta m_2\left[\mathcal{U}(X_j(n,i)) \setminus \mathcal{U}\left(\tilde{\Psi}_{-i}(n)\right)\right]} \\ &= e^{-\beta\ \psi\left(X_j(n,i) \mid \tilde{\Psi}_{-i}(n)\right)} \end{aligned}$$

and similarly for $Y'_j(n,i)$. Algorithm 3.6 can be employed to generate the sequence of tuples $\left(X_j(n,i), X'_j(n,i)\right)_{j \geq 1}$; we use these marked processes in order to draw from the conditional density $f_i(\cdot \mid x)$ (cf. the 'cluster' trick, Section 4.5.2).

Algorithm 5.2 (Sampling f_i: Attractive Case).

Set $j = 1$; $j_{t,i} = 0$; compute $\chi_i(n) = \mathbf{1}_{\left\{\mathcal{U}\left(\tilde{\Psi}_{-i}(n)\right) \supset A_i\right\}}$.

if $\chi_i(n) = 1$:

  while $j_{t,i} \neq j$:

    if $Y_j(n,i) \bigcap \left(\mathcal{U}\left(X_j(n,i)\right) \setminus \mathcal{U}(x)\right) = \emptyset$ (ie. if $X_j(n,i) \sim f_i(\cdot \mid x)$): set $j_{t,i} = j$.

    else: set $j = j + 1$.

  return $X_{j_{t,i}}(n,i)$.

else if $\chi_i(n) = 0$:

  while $j_{t,i} \neq j$:

    if $Y'_j(n,i) \bigcap \left(\mathcal{U}\left(X'_j(n,i)\right) \setminus \mathcal{U}(x)\right) = \emptyset$ (ie. if $X'_j(n,i) \sim f_i(\cdot \mid x)$): set $j_{t,i} = j$.

    else: set $j = j + 1$.

  return $X'_{j_{t,i}}(n,i)$.

Thus, given $\tilde{\Psi}_{-i}(n) = x$, if $\chi_i(n) = 1$ then $\mathcal{U}(x) \supset A_i$; so the conditional distribution of the i-th component $\tilde{\Psi}_i(n)$ is a weighted Poisson(λ) process with weight $e^{-\beta\ \psi(\cdot|x)}$. The probability that $Y_j(n,i)$ has no germs contained in $\mathcal{U}\left(X_j(n,i)\right) \setminus \mathcal{U}(x)$ is $e^{-\beta\ \psi(X_j(n,i)|x)}$. Conversely if $\chi_i(n) = 0$ then $\mathcal{U}(x)$ does not cover A_i. So the conditional distribution of $\tilde{\Psi}_i(n)$ is a weighted Poisson process conditioned to contain at least one germ. Since any germ in E_{A_i} covers all the nodes in A_i, requiring that $\tilde{\Psi}_i(n)$ has a positive number of germs ensures that A_i is covered after the update of the i-th component. The acceptance probability for $X'_j(n,i)$ is also equal to the weight, therefore the output of the algorithm has the correct weighted Poissonian distribution.

So the relevant process (as determined by $\chi_i(n)$) in each tuple $\left(X_j(n,i), X'_j(n,i)\right)$ is *proposed* as a draw from the conditional distribution, and accepted with the correct probability. If the proposal is rejected then the relevant process in the next tuple $\left(X_{j+1}(n,i), X'_{j+1}(n,i)\right)$ is proposed. This procedure is continued sequentially in j until a proposal is accepted. So, given n and i, *none* of the $\left(X_j(n,i), X'_j(n,i)\right)$, for $j < j_{t,i}$, are a draw from $f_i\left(\cdot \mid \tilde{\Psi}_{-i}(n)\right)$; $j_{t,i}$ is the ***first*** index such that the relevant of the two (as determined by $\chi_i(n)$) is a draw from the desired conditional density $f_i\left(\cdot \mid \tilde{\Psi}_{-i}(n)\right)$. Observe that the use of the mark processes $\left(Y_j(n,i), Y'_j(n,i)\right)$ provides a means to check whether the relevant process in the corresponding tuple $\left(X_j(n,i), X'_j(n,i)\right)$ is a draw from f_i. Thus given a sequence of Poisson processes $\left(X_j(n,i), X'_j(n,i)\right)_{j \geq 1}$, the objective of an exact

coupling construction is to 'pick' the indices $j_{0,i}$ so as to sample an area-interaction point process conditioned to cover the finite set $\mathcal{C}$.

5.5.1 Exact Gibbs Sampler: Attractive Case

In the previous section a virtual simulation of the target process $\left(\tilde{\Psi}_0, \ldots, \tilde{\Psi}_N\right)$ from the indefinite past was considered. Furthermore Algorithm 5.2 described the Gibbs update for each component, along with explicit details on how to sample the conditional distribution. In this section we consider how to produce an exact coupling for this Gibbs sampler.

A priori we do not know what $\tilde{\Psi}$ looks like; thus there could be a whole *range/continuum* of processes from which to simulate. Therefore the sandwich processes must be defined so as to bound the whole range of processes which are a candidate draw from the required distribution. To formalize the notion of 'sandwich' or 'bounding' processes, introduce a partial order $\preccurlyeq$ on the space of point configurations defined as: $x \preccurlyeq y$ if $\mathcal{U}(x) \subseteq \mathcal{U}(y)$. For a given cell E_{A_i}, define a *quasi-maximal* (Häggström *et al.* 1999) element $\overline{x}_i$ if $\mathcal{U}(\overline{x}_i) \supset \mathcal{U}(E_{A_i})$. Here $\mathcal{U}(E_{A_i}) = \bigcup_{\xi \in E_{A_i}} \mathcal{U}(\xi) = \{\eta \in \mathbb{R}^2;\ \eta \in \mathcal{U}(\xi) \text{ for some } \xi \in E_{A_i}\}$. Similarly, $\underline{x}_i$ is *quasi-minimal* if $\mathcal{U}(\underline{x}_i) = \emptyset$, ie. $\underline{x}_i = \emptyset$. Thus for *any* point configuration x on cell E_{A_i}: $\underline{x}_i \preccurlyeq x \preccurlyeq \overline{x}_i$.

Lemma 5.5. *Let X be a Poisson(λ) process on cell E_{A_i}, $\varphi_{-i}^{min} = \bigcup_{j \neq i} \underline{x}_j$ and $\varphi_{-i}^{max} = \bigcup_{j \neq i} \overline{x}_j$ where $\underline{x}_j$ and $\overline{x}_j$ are quasi-minimal, respectively, quasi-maximal elements on cell E_{A_j}. If $X \sim f_i\left(\cdot \mid \varphi_{-i}^{min}\right)$ and $X \sim f_i\left(\cdot \mid \varphi_{-i}^{max}\right)$ then $X \sim f_i(\cdot \mid \varphi_{-i})$ for any point configuration φ_{-i} outside E_{A_i}.*

Proof. For a given configuration φ_{-i} outside E_{A_i}, a Poisson(λ) process X on E_{A_i} will be accepted as a draw from the conditional density $f_i(\cdot \mid \varphi_{-i})$ with probability $e^{-\beta\, \psi(X|\varphi_{-i})}$ (cf. Eq. 5.9 and Algorithm 5.2). Since $\beta > 0$ the acceptance probability is monotonic in φ_{-i}, so that $x \preccurlyeq y$ implies $e^{-\beta\, \psi(X|x)} \leq e^{-\beta\, \psi(X|y)}$. By definition of φ_{-i}^{min} and φ_{-i}^{max} we have $\varphi_{-i}^{min} \preccurlyeq \varphi_{-i} \preccurlyeq \varphi_{-i}^{max}$; hence $e^{-\beta\, \psi\left(X|\varphi_{-i}^{min}\right)} \leq e^{-\beta\, \psi(X|\varphi_{-i})} \leq e^{-\beta\, \psi\left(X|\varphi_{-i}^{max}\right)}$. Therefore if X is accepted as a draw from both $f_i\left(\cdot \mid \varphi_{-i}^{min}\right)$ and $f_i\left(\cdot \mid \varphi_{-i}^{max}\right)$ then it must also be accepted as a draw from $f_i(\cdot \mid \varphi_{-i})$. □

Corollary 5.1. *For the setup as in Lemma 5.5 above, if $X \nsim f_i\left(\cdot \mid \varphi_{-i}^{max}\right)$ then $X \nsim f_i(\cdot \mid \varphi_{-i})$.*

Proof. This follows directly from the monotonicity of the acceptance probability. Since $\varphi_{-i} \preccurlyeq \varphi_{-i}^{max}$ it follows that $e^{-\beta\, \psi(X|\varphi_{-i})} \leq e^{-\beta\, \psi\left(X|\varphi_{-i}^{max}\right)}$. Thus if X is not accepted as a draw from $f_i\left(\cdot \mid \varphi_{-i}^{max}\right)$ then it will not be accepted as one from $f_i(\cdot \mid \varphi_{-i})$. □

The above results provide conditions to determine whether or not a proposed Poisson process is a draw from the required conditional distribution. There could be a whole range of processes that are a candidate for the required sample; the lemma signals when a process *has* the required distribution while the corollary when it *does not*. For $T > 0$ let $\tilde{\Psi}_i^{T,max}$ and $\tilde{\Psi}_i^{T,min}$ denote the maximal and minimal *target* processes respectively started at time $-T$. Define

$$\tilde{\Psi}_{-i}^{T,min}(n) = \bigcup_{l<i} \tilde{\Psi}_l^{T,min}(n) \bigcup_{l>i} \tilde{\Psi}_l^{T,min}(n-1) \qquad \text{and} \qquad \chi_i^{T,min}(n) = \mathbf{1}_{\left\{\mathcal{U}\left(\tilde{\Psi}_{-i}^{T,min}(n)\right) \supset A_i\right\}}.$$

$$\tilde{\Psi}_{-i}^{T,max}(n) = \bigcup_{l<i} \tilde{\Psi}_l^{T,max}(n) \bigcup_{l>i} \tilde{\Psi}_l^{T,max}(n-1) \qquad \text{and} \qquad \chi_i^{T,max}(n) = \mathbf{1}_{\left\{\mathcal{U}\left(\tilde{\Psi}_{-i}^{T,max}(n)\right) \supset A_i\right\}}.$$

Furthermore assume that for each $n \in \{-T, \ldots, 0\}$ and $i \in \{0, \ldots, N\}$ the sequence of marked Poisson processes $\left(X_j(n,i), X'_j(n,i)\right)_{j \geq 1}$ can be generated as needed (cf. Algorithm 3.6). Then the exact Gibbs coupling construction for the conditional area-interaction process proceeds as follows.

The basic idea is that the sandwich processes $\tilde{\Psi}^{T,min}$, $\tilde{\Psi}^{T,max}$ are defined in such a way that they bound (with respect to $\preccurlyeq$) any and all configurations in the corresponding cell. For each n and i the procedure first determines $\chi_i^{T,min}(n)$ and $\chi_i^{T,max}(n)$. Depending on this, it then checks whether the appropriate proposed process in the tuple $\left(X_j(n,i), X'_j(n,i)\right)$ is a sample from $f\left(\cdot \mid \tilde{\Psi}_{-i}^{T,max}(n)\right)$ *and* $f\left(\cdot \mid \tilde{\Psi}_{-i}^{T,min}(n)\right)$. If so then Lemma 5.5 ensures that process has the desired conditional distribution. If not then either it is a sample from $f\left(\cdot \mid \tilde{\Psi}_{-i}^{T,max}(n)\right)$ only or not. In the latter case Corollary 5.1 allows the procedure to set $j = j + 1$. In the former case it is not certain whether the process has the required conditional distribution; so the sandwich processes are updated to the quasi-minimal and -maximal elements respectively. The `next_update` procedure takes arguments i, the component to be updated, the current index $J_i(n)$ of the tuple of marked Poisson processes and the minimal and maximal components in the other cells $\tilde{\Psi}_{-i}^{T,min}(n), \tilde{\Psi}_{-i}^{T,max}(n)$. It returns the updates for the minimal and maximal components and index $J_i(n)$.

Algorithm 5.3 (`next_update`$\left(i, J_i(n), \tilde{\Psi}_{-i}^{T,min}(n), \tilde{\Psi}_{-i}^{T,max}(n)\right)$).

Set $I = 1$; $j = J_i(n)$.

Compute $\left(\chi_i^{T,min}(n), \chi_i^{T,max}(n)\right)$.

while $I > 0$:

if $\left(\chi_i^{T,min}(n), \chi_i^{T,max}(n)\right) = (0,1)$:

if $X_j(n,i) = X'_j(n,i)$:

if $X_j(n,i) \sim f_i\left(\cdot \left| \tilde{\Psi}_{-i}^{T,max}(n)\right.\right)$:

if $X_j(n,i) \sim f_i\left(\cdot \left| \tilde{\Psi}_{-i}^{T,min}(n)\right.\right)$:

set $\texttt{min\_update} = \texttt{max\_update} = X_j(n,i)$; $\texttt{j\_update} = j$; $I = -1$.

else: set $\texttt{min\_update} = \underline{x}_i$; $\texttt{max\_update} = \overline{x}_i$; $\texttt{j\_update} = j$; $I = -1$.

else: set $j = j + 1$.

else: set $\texttt{min\_update} = \underline{x}_i$; $\texttt{max\_update} = \overline{x}_i$; $\texttt{j\_update} = j$; $I = -1$.

else if $\left(\chi_i^{T,min}(n), \chi_i^{T,max}(n)\right) = (1,1)$:

if $X_j(n,i) \sim f_i\left(\cdot \left| \tilde{\Psi}_{-i}^{T,max}(n)\right.\right)$:

if $X_j(n,i) \sim f_i\left(\cdot \left| \tilde{\Psi}_{-i}^{T,min}(n)\right.\right)$:

set $\texttt{min\_update} = \texttt{max\_update} = X_j(n,i)$; $\texttt{j\_update} = j$; $I = -1$.

else: $\texttt{min\_update} = \underline{x}_i$; $\texttt{max\_update} = \overline{x}_i$; $\texttt{j\_update} = j$; $I = -1$.

else: set $j = j + 1$.

else if $\left(\chi_i^{T,min}(n), \chi_i^{T,max}(n)\right) = (0,0)$:

if $X'_j(n,i) \sim f_i\left(\cdot \left| \tilde{\Psi}_{-i}^{T,max}(n)\right.\right)$:

if $X'_j(n,i) \sim f_i\left(\cdot \left| \tilde{\Psi}_{-i}^{T,min}(n)\right.\right)$:

set $\texttt{min\_update} = \texttt{max\_update} = X'_j(n,i)$; $\texttt{j\_update} = j$; $I = -1$.

else: $\texttt{min\_update} = \underline{x}_i$; $\texttt{max\_update} = \overline{x}_i$; $\texttt{j\_update} = j$; $I = -1$.

else: set $j = j + 1$.

return $(\texttt{min\_update}, \texttt{max\_update}, \texttt{j\_update})$.

Algorithm 5.4 (Exact Gibbs: Conditional Area-Interaction: Attractive Case).

Set $T = 1$.

while $T > 0$:

for $n \in \left\{-T+1, \ldots, -\lfloor \frac{T}{2} \rfloor\right\}$ and $i \in \{0, \ldots, N\}$: set $J_i(n) = 1$.

Initialize $\tilde{\Psi}^{T,max}(-T) = (\overline{x}_0, \ldots, \overline{x}_N)$; $\tilde{\Psi}^{T,min}(-T) = (\underline{x}_0, \ldots, \underline{x}_N)$.

for $n \in \{-T+1, \ldots, 0\}$:

for $i \in \{0, \ldots, N\}$:

compute $\texttt{next\_update}\left(i, J_i(n), \tilde{\Psi}_{-i}^{T,min}(n), \tilde{\Psi}_{-i}^{T,max}(n)\right)$.

set $\tilde{\Psi}_i^{T,min}(n) = \texttt{min\_update}$; $\tilde{\Psi}_i^{T,max}(n) = \texttt{max\_update}$.

reset $J_i(n) = \texttt{j\_update}$.

if $\tilde{\Psi}^{T,max}(0) = \tilde{\Psi}^{T,min}(0)$: set $T = -1$; **else**: set $T = 2T$.

return $\tilde{\Psi}^{T,min}(0) = \left(\tilde{\Psi}_0^{T,min}(0), \ldots, \tilde{\Psi}_N^{T,min}(0)\right)$.

Lemma 5.6 (Sandwiching). *Suppose* $\tilde{\Psi} = \left(\tilde{\Psi}_0, \ldots, \tilde{\Psi}_N\right)$ *is the target process started from time* $-\infty$ *and evolved so that its equilibrium distribution is that of an area-interaction process with parameters* λ *and* $\beta > 0$ *and conditioned to cover* $\mathcal{C}$*. Let the bounding processes* $\tilde{\Psi}^{T,min} = \left(\tilde{\Psi}_0^{T,min}, \ldots, \tilde{\Psi}_N^{T,min}\right)$ *and* $\tilde{\Psi}^{T,max} = \left(\tilde{\Psi}_0^{T,max}, \ldots, \tilde{\Psi}_N^{T,max}\right)$ *be defined recursively as in Algorithm 5.4. Furthermore, for each* n *and* i*, let* $\tilde{\Psi}_i(n)$*,* $\tilde{\Psi}_i^{T,min}(n)$ *and* $\tilde{\Psi}_i^{T,max}(n)$ *be updated by the same realization of the sequence of tuples* $\left(X_j(n,i), X'_j(n,i)\right)_{j\geq 1}$*. Then for all* n *and* $0 \leq i \leq N$

$$\Psi_i^{T,min}(n) \preccurlyeq \Psi_i(n) \preccurlyeq \Psi_i^{T,max}(n). \tag{5.10}$$

Proof. The relationship holds at time $-T$ by construction. Suppose that it also holds up to some time $n-1 > -T$ and consider the update at time n. For any i there are only two possibilities for the values the sandwich processes can take:

$$\text{(i)} \qquad \tilde{\Psi}_i^{T,min}(n) = \underline{x}_i, \qquad \tilde{\Psi}_i^{T,max}(n) = \overline{x}_i; \qquad \text{or}$$

$$\text{(ii)} \qquad \tilde{\Psi}_i^{T,min}(n) = \tilde{\Psi}_i^{T,max}(n) = \begin{cases} X_j(n,i) & \begin{cases} \text{if } \chi_i^{min}(n) = \chi_i^{max}(n) = 1 \text{ or} \\ X_j(n,i) = X'_j(n,i); \end{cases} \\ X'_j(n,i) & \text{if } \chi_i^{min}(n) = \chi_i^{max}(n) = 0. \end{cases}$$

In case (i) Eq. (5.10) still persists; in case (ii) Lemma 5.5 ensures that $\tilde{\Psi}_i(n)$ must also be updated to the same value as the bounding processes, hence Eq. (5.10) is maintained. The result now follows by induction along the sequence of times $n \in \{-T, \ldots, 0\}$. □

Lemma 5.7. *Define*

$$j_{t,i} = \inf\left\{ j \geq 1; \begin{array}{ll} X_j(n,i) \sim f_i\left(\cdot \mid \tilde{\Psi}_{-i}(n)\right) & \textit{if} \quad \chi_i(n) = \mathbf{1}_{\left\{\mathcal{U}\left(\tilde{\Psi}_{-i}(n)\right) \supset A_i\right\}} = 1; \\ X'_j(n,i) \sim f_i\left(\cdot \mid \tilde{\Psi}_{-i}(n)\right) & \textit{else}. \end{array}\right.$$

For all n *and* i *let* $J_i(n)$ *be defined recursively as in Algorithm 5.4. Then* $J_i(n) \leq j_{t,i}$*; moreover if* $\tilde{\Psi}_i^{T,min}(n) = \tilde{\Psi}_i^{T,max}(n)$ *then* $J_i(n) = j_{t,i}$*.*

Proof. For a given n and i the initial value of $J_i(n)$ is 1; so $J_i(n) \leq j_{t,i}$ initially. Suppose that before the update of the i-th component at time n, $J_i(n) \leq j_{t,i}$. In order to determine the update for the bounding processes the algorithm sets $j = J_i(n) \leq j_{t,i}$ and computes $\left(\chi_i^{T,min}(n), \chi_i^{T,max}(n)\right)$.

Suppose this is $(1,1)$; the algorithm then checks whether $X_j(n,i) \sim f_i\left(\cdot \mid \tilde{\Psi}_{-i}^{T,max}(n)\right)$. If not then, since $\tilde{\Psi}_{-i}(n) \preccurlyeq \tilde{\Psi}_{-i}^{T,max}(n)$, it follows by Corollary 5.1 that $X_j(n,i)$ is also not a draw from $f_i\left(\cdot \mid \tilde{\Psi}_{-i}(n)\right)$. So it cannot be the case that $j = j_{t,i}$; hence $j < j_{t,i}$. The value of j is increased by one and the above procedure repeated. On the other hand if $X_j(n,i)$ is a draw from $f_i\left(\cdot \mid \tilde{\Psi}_{-i}^{T,max}(n)\right)$ then either (i) it is also a draw from $f_i\left(\cdot \mid \tilde{\Psi}_{-i}^{T,min}(n)\right)$; or (ii) not. In case (i) Lemma 5.5 yields that $X_j(n,i)$ must also be a draw from $f_i\left(\cdot \mid \tilde{\Psi}_{-i}(n)\right)$ and hence $j = j_{t,i}$. In case (ii) the algorithm leaves the value of j unchanged. In any case $j \leq j_{t,i}$; since $J_i(n)$ is reset to j it follows that $J_i(n) \leq j_{t,i}$ after the update of the i-th component, with equality only if (i) is true.

Now if $\left(\chi_i^{T,min}(n), \chi_i^{T,max}(n)\right) = (0,0)$ then the same argument as above (but with $X_j(n,i)$ replaced by $X'_j(n,i)$) shows that the inequality is maintained. Finally if $\left(\chi_i^{T,min}(n), \chi_i^{T,max}(n)\right) = (0,1)$ then the algorithm only proceeds if $X_j(n,i) = X'_j(n,i)$, and similar arguments as above establish the inequality. Thus $J_i(n) \leq j_{t,i}$ after the update of the i-th component, with equality if and only if the bounding processes take on the same value. Induction along the sequence of times n and components i completes the proof. □

Lemma 5.8 (Coalescence). *For the setup as in Lemma 5.6 the bounding processes $\tilde{\Psi}^{T,max}, \tilde{\Psi}^{T,min}$ coalesce for sufficiently large T.*

Proof. Suppose that, for some $n > -T$, $X_1(n,i) = X'_1(n,i)$ and their respective mark processes $Y_1(n,i) = Y'_1(n,i) = \emptyset$ for all i. In this case $X_1(n,i)$ will be a draw from both $f_i\left(\cdot \mid \tilde{\Psi}_{-i}^{T,max}(n)\right)$ and $f_i\left(\cdot \mid \tilde{\Psi}_{-i}^{T,min}(n)\right)$ for all i, since its mark process contains no germs. Thus $\tilde{\Psi}_i^{T,max}(n) = \tilde{\Psi}_i^{T,min}(n)$ for all i and hence coalescence of the upper and lower processes occurs. Since the processes are Poisson, $\mathbb{P}[X_1(n,i) = X'_1(n,i)] = \mathbb{P}[X_1(n,i) \neq \emptyset] > 0$ and $\mathbb{P}[Y_1(n,i) = \emptyset] > 0$ for all i; moreover they are independent of n, and $\{X_1(n,i) = X'_1(n,i) \text{ and } Y_1(n,i) = \emptyset\}$ are independent for each i. Thus $\bigcap_i \{X_1(n,i) = X'_1(n,i) \text{ and } Y_1(n,i) = \emptyset\}$ occurs infinitely often (second Borel-Cantelli Lemma); so coalescence is almost sure to happen. □

Lemma 5.9 (Funnelling). *For the setup as in Lemma 5.6, for all $-S \leq -T \leq t \leq 0$*

$$\tilde{\Psi}_i^{T,min}(n) \preccurlyeq \tilde{\Psi}_i^{S,min}(n) \preccurlyeq \tilde{\Psi}_i^{S,max}(n) \preccurlyeq \tilde{\Psi}_i^{T,max}(n). \tag{5.11}$$

Proof. By construction the relationship holds at time $-T$; assume it also holds up to time $n-1 > -T$ for all i and consider the update of the first component $i = 0$. Now there are two possibilities for

the values $\tilde{\Psi}_0^{T,min}(n), \tilde{\Psi}_0^{T,max}(n)$ can take: either (i) $\tilde{\Psi}_0^{T,min}(n) = \underline{x}_0$ and $\tilde{\Psi}_0^{T,max}(n) = \overline{x}_0$; or (ii) $\tilde{\Psi}_0^{T,min}(n) = \tilde{\Psi}_0^{T,max}(n)$. In case (i) Eq. (5.11) still persists. For case (ii) it needs to be shown that $\tilde{\Psi}_0^{S,min}(n) = \tilde{\Psi}_0^{S,max}(n) = \tilde{\Psi}_0^{T,min}(n)$. By assumption $\tilde{\Psi}_{-0}^{T,min}(n) \preccurlyeq \tilde{\Psi}_{-0}^{S,min}(n) \preccurlyeq \tilde{\Psi}_{-0}^{S,max}(n) \preccurlyeq \tilde{\Psi}_{-0}^{T,max}(n)$ holds. Therefore if $\tilde{\Psi}_0^{T,min}(n) = \tilde{\Psi}_0^{T,max}(n)$ then Lemma 5.5 ensures that $\tilde{\Psi}_0^{S,min}(n)$ and $\tilde{\Psi}_0^{S,max}(n)$ must also take the same value as $\tilde{\Psi}_0^{T,min}(n)$. Eq. (5.11) persists and the result now follows by induction along the sequence of components i and times n. □

Theorem 5.2. *Suppose that the processes defined in Lemma 5.6 are evolved as in Algorithm 5.4. Then the algorithm terminates almost surely in finite time and the distribution of the output is a perfect draw of an attractive area-interaction point process conditioned to cover $\mathcal{C}$.*

Proof. Lemma 5.8 shows that the algorithm terminates in almost sure finite time. Consider a virtual simulation of the target process $\tilde{\Psi} = \left(\tilde{\Psi}_0, \ldots, \tilde{\Psi}_N\right)$ from time $-\infty$ which, for each T, is coupled to the same random sequences $\left(X_j(n,i), X'_j(n,i)\right)_{j\geq 1}$ as the bounding processes on $\{-T, \ldots, 0\}$. The equilibrium distribution of $\tilde{\Psi}$ is that of an attractive conditional area-interaction process and $\tilde{\Psi}(0)$ is a draw from equilibrium. Now Lemmas 5.6, 5.7 & 5.9 ensure that if the upper and lower processes have coalesced then their common value must also be that of the target process, ie. conditional on coalescence, $\tilde{\Psi}^{T,min} = \tilde{\Psi}^{T,max} = \tilde{\Psi}$. Thus $\tilde{\Psi}^{T,min}(0) = \tilde{\Psi}(0)$ has the required distribution. □

5.5.2 Exact Gibbs Sampler: Repulsive Case

In the previous section a Gibbs sampler for the attractive conditional area-interaction process was introduced. Sampling the desired conditional distribution (Algorithm 5.2) and the exact coupling construction (Algorithm 5.4) were described. In this section a Gibbs sampler for the repulsive process $(\beta < 0)$ is considered. The sampler is similar to that for the attractive case; however some modifications to Algorithms 5.2 & 5.4 need to be made.

First consider sampling from the conditional distribution. Recall that $\tilde{\Psi} = \left(\tilde{\Psi}_0, \ldots, \tilde{\Psi}_N\right)$ denotes the target process and in order to update the i-th component the Gibbs sampler draws from the conditional density f_i given the process outside E_{A_i}. Suppose that, for some n and i, $\tilde{\Psi}_{-i}(n)$ is the collection of processes outside cell E_{A_i}. Then, given $\tilde{\Psi}_{-i}(n) = x$, the form of the conditional density on cell E_{A_i} is given by Eq. (5.9): $f_i(\cdot \mid x) \propto \lambda^{n(\cdot)} e^{-\beta\psi(\cdot|x)} \mathbf{1}_{\{\mathcal{U}(x\cup\cdot)\supset A_i\}}$ where

$\psi(\cdot \mid x) = m_2[\mathcal{U}(\cdot) \setminus \mathcal{U}(x)]$. For $\beta > 0$ the weight factor $e^{-\beta\psi}$ is bounded by one; for the repulsive case, $\beta < 0$, so the weight factor is no longer bounded by one.

However in order to draw from $f_i(\cdot \mid x)$ it suffices to draw Poisson(λ) process X and accept it with probability *proportional* to $e^{-\beta\psi(X|x)}\mathbf{1}_{\{\mathcal{U}(x\cup X)\supset A_i\}}$. The alternative to computing this expression is to construct an *observable* event that has a probability of occurring which is proportional to this expression. So consider a Poisson($|\beta|$) process Y on $\mathcal{U}(E_{A_i})$; then, letting $A = \mathcal{U}(X) \setminus \mathcal{U}(x)$

$$\mathbb{P}[Y \subseteq A] = e^{-|\beta|m_2\left[\mathcal{U}\left(E_{A_i}\right)\setminus A\right]} = e^{\beta m_2\left[\mathcal{U}\left(E_{A_i}\right)\setminus A\right]} \propto e^{-\beta m_2[A]} = e^{-\beta m_2[\mathcal{U}(X)\setminus\mathcal{U}(x)]}.$$

Thus in order to sample from f_i: compute $\chi_i(n) = \mathbf{1}_{\left\{\mathcal{U}\left(\tilde{\Psi}_{-i}(n)\right)\supset A_i\right\}}$. If $\chi_i(n) = 1$ then draw $X \sim$ Poisson(λ) process on E_{A_i} and its associated mark $Y \sim$Poison($|\beta|$) process on $\mathcal{U}(E_{A_i})$. Accept X if Y is contained in $\mathcal{U}(X) \setminus \mathcal{U}\left(\tilde{\Psi}_{-i}\right)$. If $\chi_i(n) = 0$ draw $X' \sim$ Poisson(λ) conditioned to contain at least one germ and its associated mark Y'. Accept X' if Y' is contained in $\mathcal{U}(X') \setminus \mathcal{U}\left(\tilde{\Psi}_{-i}\right)$.

We now consider modifications required to devise an exact coupling for the repulsive process. In the attractive case the weight factor is monotonic in $\tilde{\Psi}_{-i}$; for the repulsive case it is anti-monotonic: $x \preccurlyeq y$ implies $e^{-\beta\psi(\cdot|x)} \geq e^{-\beta\psi(\cdot|y)}$ since $\beta < 0$. Therefore, for any configurations $x_{-i} \preccurlyeq y_{-i}$ outside E_{A_i}, if a given Poison process X is *not* a draw from $f_i(\cdot \mid x_{-i})$ then it is also *not* a draw from $f_i(\cdot \mid y_{-i})$. This is seen by considering the mark process $Y \sim$ Poisson($|\beta|$). Since $X \nsim f_i(\cdot \mid x_{-i})$, then $Y \nsubseteq (\mathcal{U}(X) \setminus \mathcal{U}(x_{-i}))$ and so $Y \nsubseteq (\mathcal{U}(X) \setminus \mathcal{U}(y_{-i}))$. In light of this anti-monotonicity modifications also need to be made to the coupling construction. Specifically, Algorithm 5.4 *first* checks whether the relevant candidate process $X_j(n,i)$ or $X'_j(n,i)$ is a draw from $f_i\left(\cdot \mid \tilde{\Psi}_{-i}^{T,max}(n)\right)$ and *then* from $f_i\left(\cdot \mid \tilde{\Psi}_{-i}^{T,min}(n)\right)$. For the repulsive case this order needs to be *reversed*, ie. first check whether the candidate process is a draw from $f_i\left(\cdot \mid \tilde{\Psi}_{-i}^{T,min}(n)\right)$ and then from $f_i\left(\cdot \mid \tilde{\Psi}_{-i}^{T,max}(n)\right)$. The results of Lemmas 5.5, 5.6, 5.7, 5.8 & 5.9 still hold true, so that the modified coupling construction yields a perfect sample of a repulsive area-interaction process conditioned to cover $\mathcal{C}$.

5.6 Implementational Issues

5.6.1 Large Sampling Window

If the sampling window W is large, so that the region $E_{A_0} \equiv E_\emptyset$ is also large, then Algorithm 5.2 may be quite inefficient since the acceptance probability will be low (cf. Section 4.3). A solution

is to further divide E_{A_0} into smaller cells $E^1_{A_0}, \ldots, E^m_{A_0}$ and carry out the Gibbs sampling on the cells $\left(E^1_{A_0}, \ldots, E^m_{A_0}, E_{A_1}, \ldots, E_{A_N}\right)$. This should improve the efficiency of the algorithm without increasing the complexity, since additional cells can be easily incorporated into the sampler.

5.6.2 Implementation via 2-Stage Procedure

When $\mathcal{C}$ consists of connected nodes (cf. Section 3.7.1) the regions $\{E_{A_i};\ i = 1, \ldots, N\}$ may be geometrically quite irregular. Therefore even though the conditional distribution on each cell is relatively straightforward (density given by f_i in Eq. 5.9), the irregular shape of the cells may pose implementational difficulties. Drawing from the density f_i via Algorithm 5.2 would require additional (geometric) tests, which may be both difficult to implement and computationally burdensome.

On the other hand if the regions E_{A_i} are disks of radius r, then $\mathcal{U}\left(E_{A_i}\right)$ is just a disk of radius $2r$. In this case it is relatively straightforward to sample from f_i. Therefore the practical implementation of the exact Gibbs Algorithm 5.4 is carried out via a 2-Stage procedure: a subset $A^* \subseteq \mathcal{C}$ (cf. Definition 3.2) which contains the maximum number of disconnected nodes is chosen. Algorithm 5.4 is applied to A^*; if the output also covers $\mathcal{C}$ then this sample is returned. If not then another sample is drawn independently via Algorithm 5.4 applied to A^*. This is continued until a sample which covers $\mathcal{C}$ is returned. For similar implementational reasons the implementation of the modified Cai & Kendall Algorithm 5.1 is also carried out via the above 2-Stage procedure.

5.7 Simulation Results

In this section the simulation results for the conditional (attractive) area-interaction process are presented. The experiments aim to evaluate the performance (in terms of actual run times in seconds) of the three perfect algorithms: 2-Stage Rejection (cf. Section 5.3), the modified Cai & Kendall Algorithm 5.1, and the Gibbs Algorithm 5.4. Recall from Section 3.7.1 that $\mathcal{C}$ can be viewed as a graph where two nodes c_i and c_j are *connected* if $\mathcal{U}\left(c_i\right) \cap \mathcal{U}\left(c_j\right) \neq \emptyset$. The sampling window is $W_{\delta r}$, for some δ, and A^* is as in Definition 3.2. Recall from Section 5.6.2 above that all three algorithms are implemented via a 2-Stage procedure. For convenience abbreviate the 2-Stage Rejection algorithm by '2Stg', the modified Cai & Kendall algorithm by 'CK' and the exact Gibbs algorithm by 'Gibbs'.

In the experiments for the conditional Boolean model (Section 3.7.2) the sampling window was always taken to be W_r. The reason being that the germs of the Boolean model are independent in disjoint regions. Therefore, as long as $\mathcal{U}(\mathcal{C}) \subset W$, the size of the sampling window W does not matter. However for the area-interaction model the presence of interactions between the germs violates this independence property. In this case the size of W influences the strength of the interaction. Therefore run times are compared against size of W, underlying Poissonian intensity λ, interaction parameter β and size of the conditioning set k.

Remark 3.2 in Section 3.7.2 comments about the notion of *local change* and *global change* algorithms. CK is a local change algorithm since it employs birth-death processes while 2Stg and Gibbs are global change algorithms. The pros and cons of either type of algorithm are also discussed in Remark 3.2. For extreme model parameters (eg. high λ, β, k or a large W) local change algorithms are expected to perform better; for moderate parameters global ones are. The simulation results support this intuition. As parameter values get extreme, CK performs better than the other two, while for moderate parameters 2Stg and Gibbs do well. Notice also that Gibbs here is very competitive to 2Stg for all parameter values, unlike in the case of the conditional Boolean model (cf. Section 3.7).

The simulations were carried out on a PC (Pentium 4 2.67GHz, 248MB RAM) running Windows XP; the implementations of the algorithms were programmed in `Python` (version 2.3).

5.7.1 Experiment 1: Run Times Versus Intensity λ

The evolution of run times as the underlying intensity λ varies is first explored. Holding the other model parameters fixed, as λ decreases the coverage conditioning becomes increasingly stringent. On the other hand as λ decreases so does the mean number of germ-grain pairs; for 2Stg and Gibbs this means that global moves are more likely to be accepted since the acceptance weight $e^{-\beta\psi}$ increases as λ decreases. The computational burden of 2Stg and Gibbs should nevertheless increase as λ decreases since coverage of $\mathcal{C}$ will still be a rare event. The results of Experiment 3.7.3 suggest that the Cai & Kendall algorithm performs better than 2Stg and Gibbs as λ decreases. Whether similar results should hold here requires some careful thought. The modified Cai & Kendall algorithm (CK) involves, additionally, censoring birth transitions. Moreover there is an extra computational burden: that of simulating the marks Z_ξ attached to each birth ξ which enable the correct censoring

of births. A priori it was difficult to predict the relative behaviour of the respective run times as λ gets smaller, except that those for 2Stg and Gibbs should increase.

As λ increases the coverage conditioning will be easily satisfied. On the other hand the area-interaction between the germs will increase and so global moves in 2Stg and Gibbs are less likely to be accepted. Therefore their respective run times will also increase with λ; thus the graphs for these two should look somewhat like parabolas. Moreover single births would be readily accepted in CK since the acceptance probability increases with the number of germ-grain pairs. Thus CK should have lower run times for high λ.

Figures 5.2 & 5.3 depict the run times for two different simulations. The results do illustrate parabolic-like curves for 2Stg and Gibbs and support the expectation that CK is more efficient for large λ. For small λ, it appears that the added complexity of CK (as compared to the Cai & Kendall algorithm for conditional Boolean models) means that its runs times get very large as λ decreases.

5.7.2 Experiment 2: Run Times Versus Intensity Window Size

This experiment explores the evolution of run times as the size of the sampling window W changes. The size of the regions $\{E_A;\ A \neq \emptyset\}$ remains constant as W gets larger; hence $E_\emptyset$ also gets larger. Rejection sampling for the area-interaction process becomes inefficient as the window size increases (cf. Section 4.3) since global changes will have low acceptance probabilities. On the other hand the dynamics of a local change algorithm will be less affected by changes in the size of W. This is because the births and deaths are proposed at rates which satisfy the equations of detailed balance (Eq. 1.25); therefore local changes are more readily accepted. Hence CK should to do well for large W while 2Stg/Gibbs for small to moderate W. Given $\mathcal{C}$, the sampling window W is taken to be $W_{\delta r}$, defined in Section 3.7.1. In Figure 5.4 the run times of the three algorithms are plotted against various values of δ: for small δ 2Stg and Gibbs do well while for large δ CK does better.

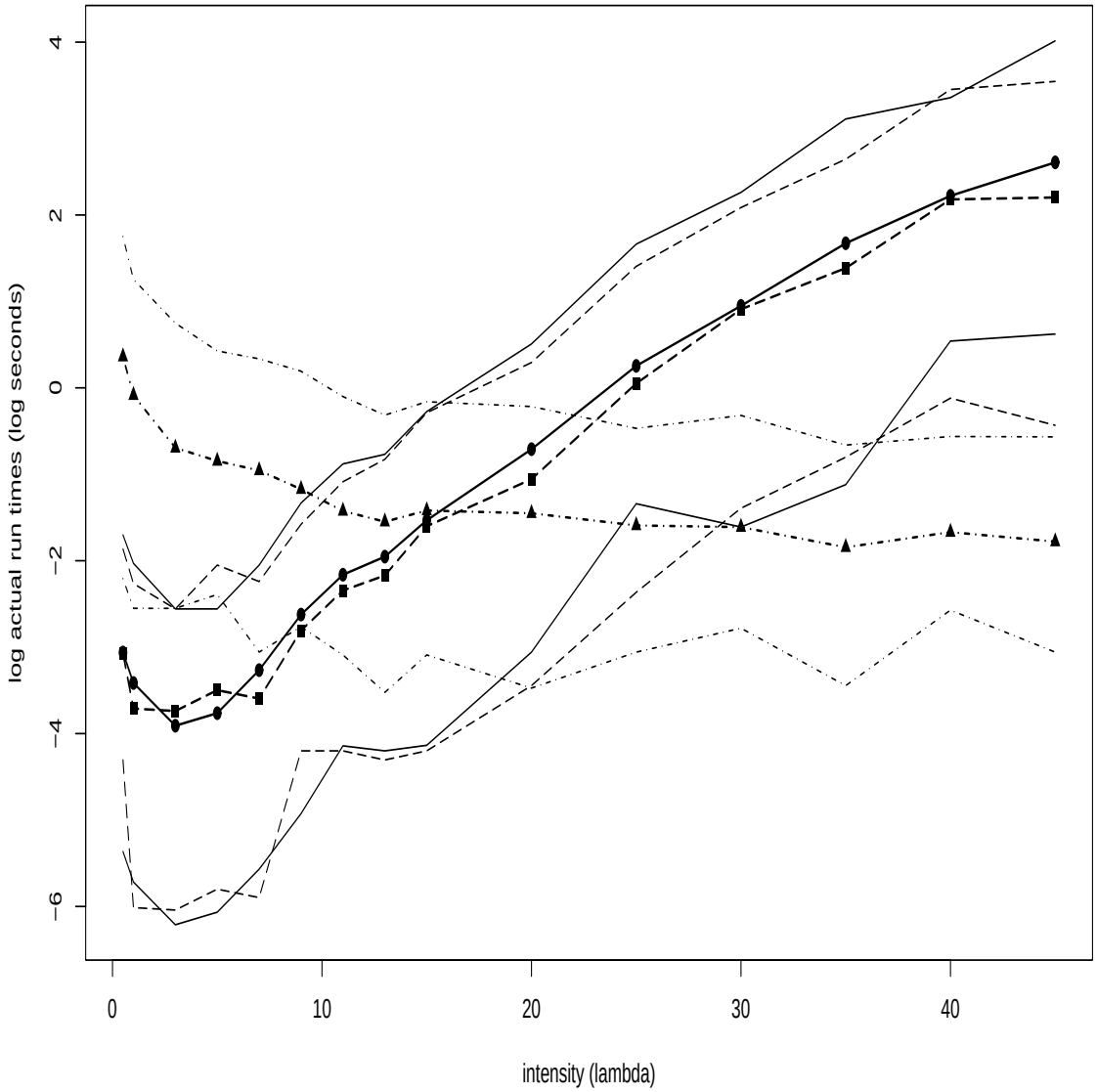

Figure 5.2: Empirical log mean run times for 2Stg (solid dots), CK (solid triangles) and Gibbs (solid squares). The solid (2Stg), dashed-dotted (CK) and dashed (Gibbs) lines represent the upper and lower simulation envelopes computed at the 90% and 10% quantiles respectively. The conditioning set $\mathcal{C}$ consists of a single cluster of $k = 5$ randomly placed nodes, radius $r = 0.1$, interaction parameter $\beta = 15$ and sampling window $W = W_{3r}$. The x-axis represents the intensity λ.

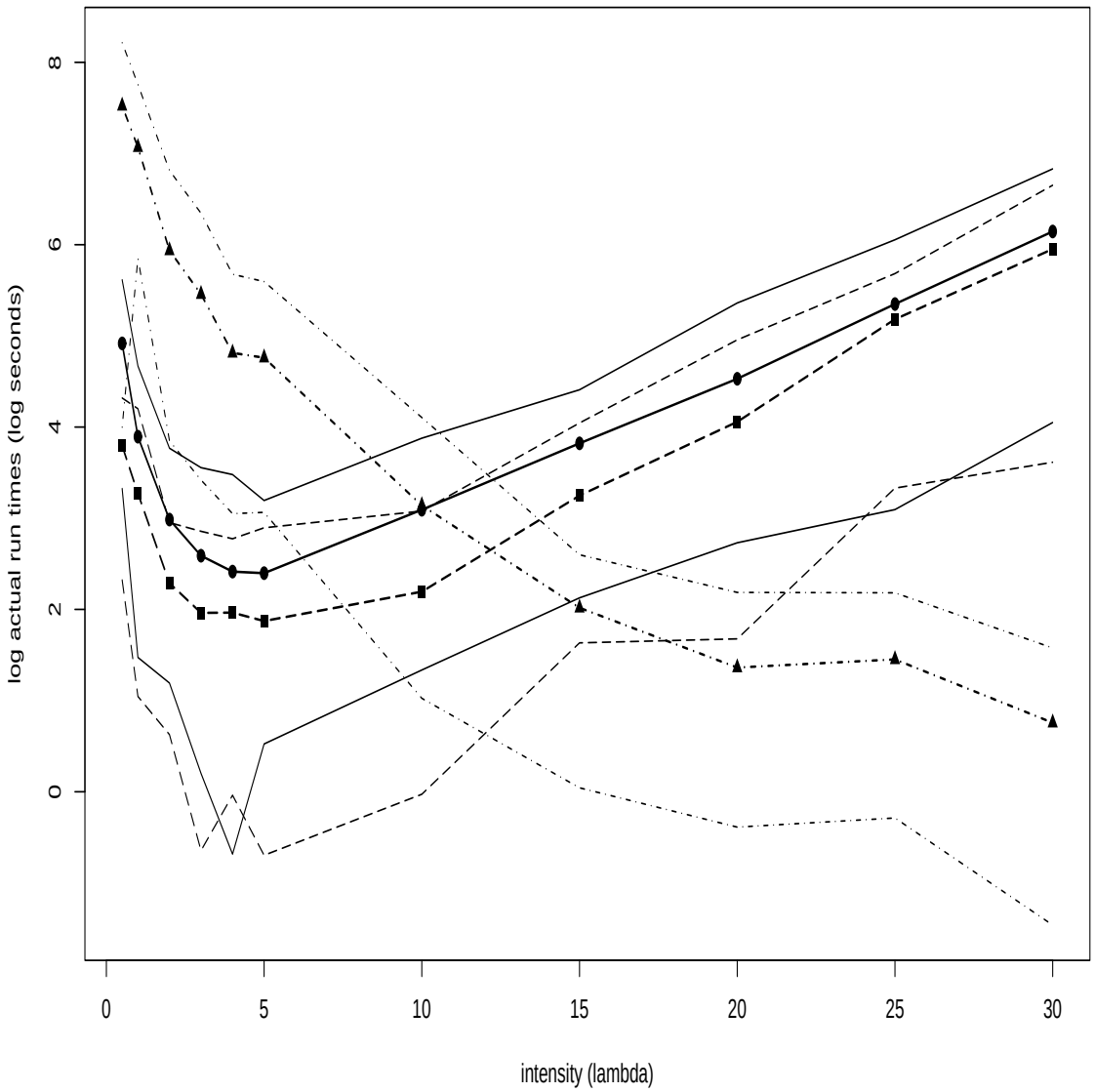

Figure 5.3: Empirical log mean run times for 2Stg (solid dots), CK (solid triangles) and Gibbs (solid squares). The solid (2Stg), dashed-dotted (CK) and dashed (Gibbs) lines represent the upper and lower simulation envelopes computed at the 90% and 10% quantiles respectively. The conditioning set $\mathcal{C}$ consists of a single cluster of $k = 10$ randomly placed nodes, radius $r = 0.1$, interaction parameter $\beta = 10$ and sampling window $W = W_{3r}$. The x-axis represents the intensity λ.

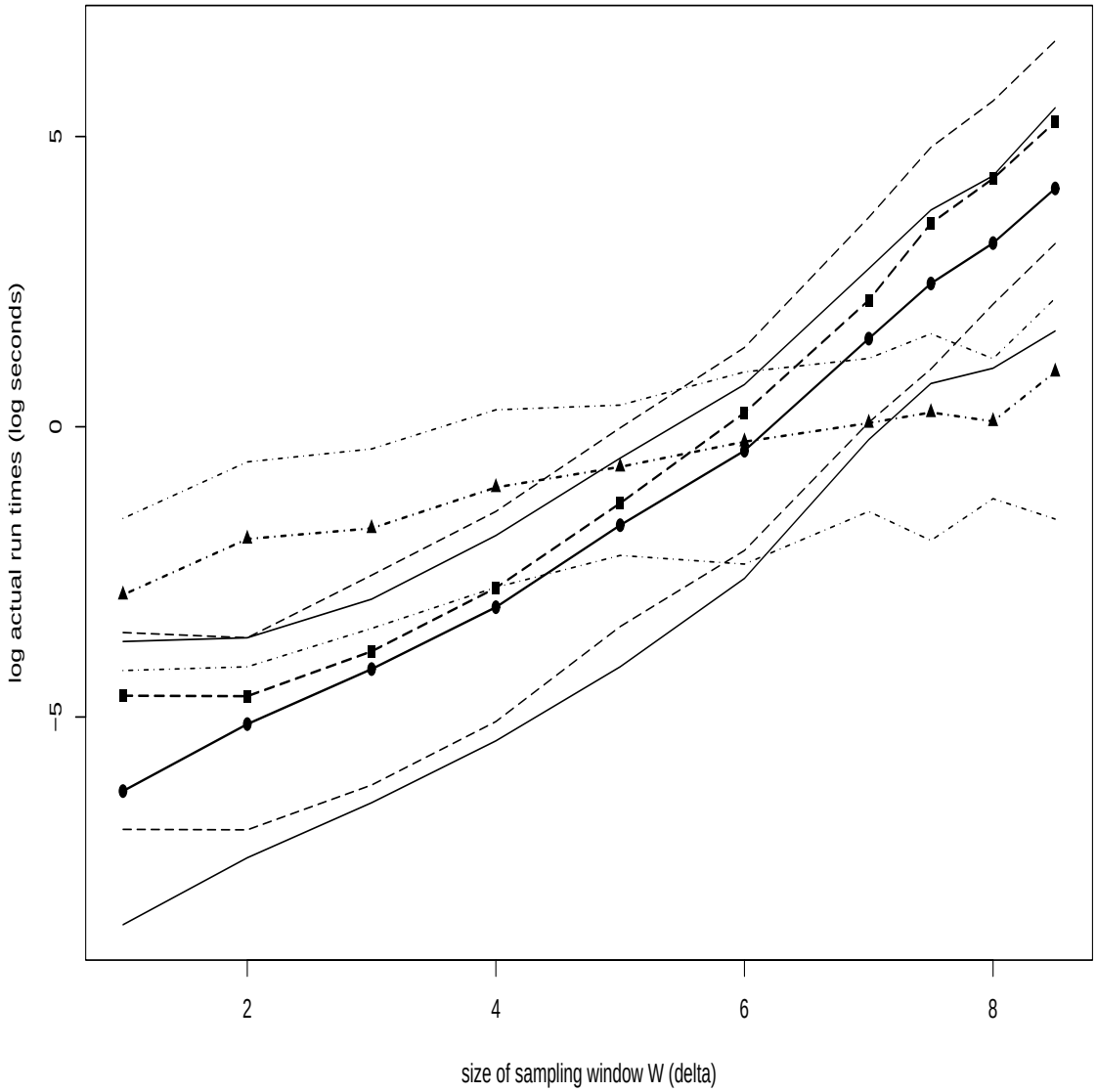

Figure 5.4: Empirical log mean run times for 2Stg (solid dots), CK (solid triangles) and Gibbs (solid squares). The solid (2Stg), dashed-dotted (CK) and dashed (Gibbs) lines represent the upper and lower simulation envelopes computed at the 90% and 10% quantiles respectively. The conditioning set $\mathcal{C}$ consists of a single cluster of $k = 5$ randomly placed nodes, radius $r = 0.1$, interaction parameter $\beta = 10$ and intensity $\lambda = 10$. The sampling window W is taken to be $W_{\delta r}$, with the x-axis representing different values of δ.

5.7.3 Experiment 3: Run Times Versus Interaction Parameter β

In this experiment the interaction parameter β is varied. The experiment considers only the case when $\beta > 0$, ie. an attractive area-interaction process. The value of β controls the interaction between the germs. The acceptance probability for both local and global moves monotonically decreases as β increases; however acceptance of global moves will be more dependent on changes in β than that of local moves. Thus one expects a similar behaviour in run times as in the previous experiments: 2Stg and Gibbs should perform better for low β, while CK becomes more competitive as β increases. The respective run times are depicted by Figure 5.5.

5.7.4 Experiment 4: Run Times Versus k

The number of upper and lower bounding processes in both CK and Gibbs is determined by the number of non-empty regions E_A. This in turn is affected by $I(\mathcal{C})$, the maximum clique size (cf. Definition 3.5 & Figure 3.2). Therefore, for similar reasons as in Section 3.7.4, $I(\mathcal{C})$is fixed at some I^* for all k. As the size of $\mathcal{C}$ increases, both the coverage requirement and the area-interaction become stringent. The reason being that the probability of coverage decreases as k increases. Moreover as $\mathcal{C}$ gets larger, the mean number of germ-grain pairs also increases; this then increases the area-interaction between them. So CK should do relatively better as k increases. Figures 5.6 & 5.7 depict the respective graphs of the run times. The difference in the graphs is not as clear cut for this experiment. The mean run times do evolve as expected: for small k the graphs for 2Stg and Gibbs are lower than those for CK; as k increases, 2Stg becomes less competitive compared to Gibbs and CK. The performance of Gibbs and CK is much closer; however the figures do suggest that CK becomes relatively more efficient as k increases. It took about 30-60 hours to obtain the required samples for the largest value of k in both Figure 5.6 & 5.7.

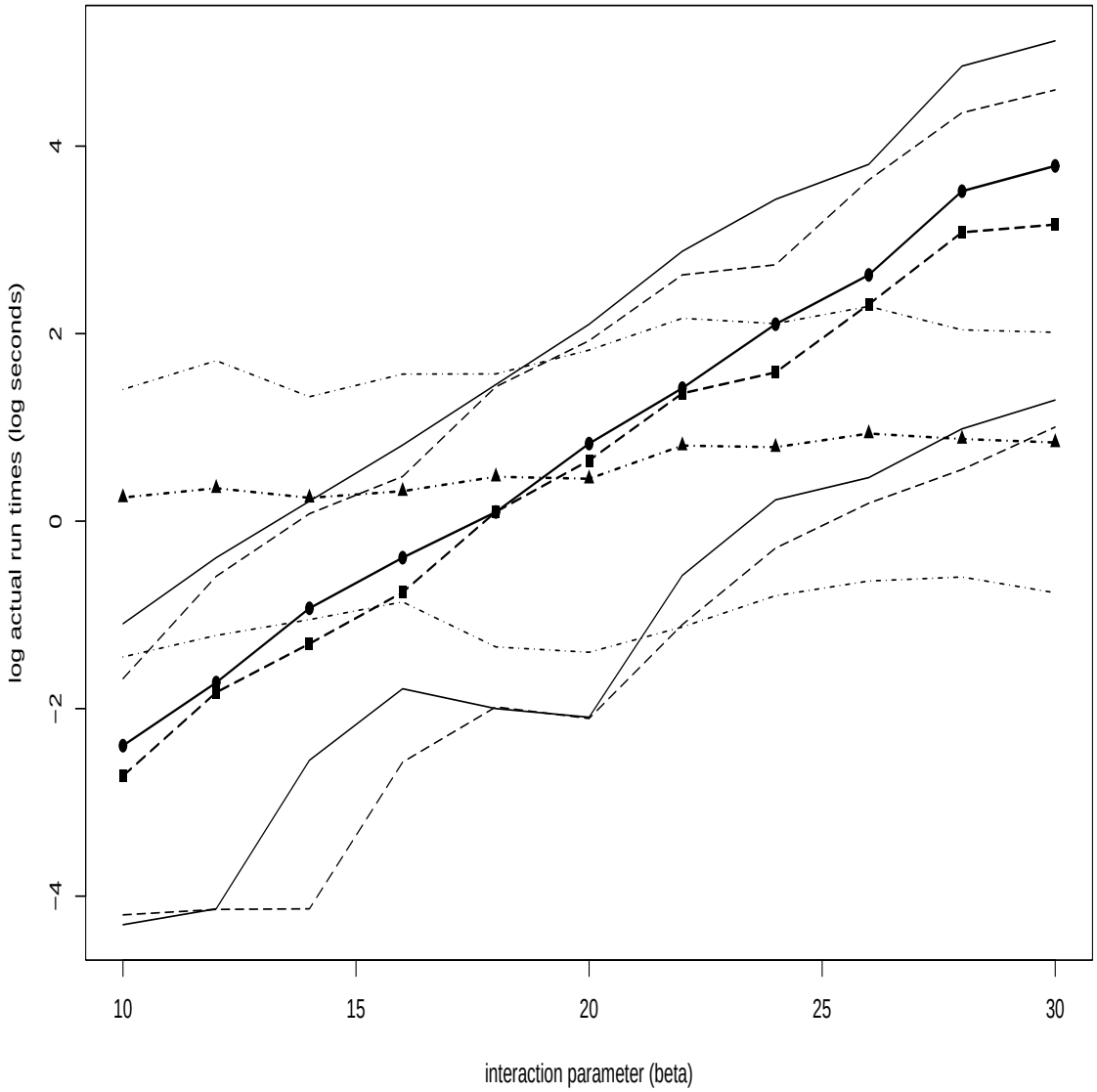

Figure 5.5: Empirical log mean run times for 2Stg (solid dots), CK (solid triangles) and Gibbs (solid squares). The solid (2Stg), dashed-dotted (CK) and dashed (Gibbs) lines represent the upper and lower simulation envelopes computed at the 90% and 10% quantiles respectively. The conditioning set $\mathcal{C}$ consists of a single cluster of $k = 5$ randomly placed nodes, radius $r = 0.1$, the intensity $\lambda = 20$ and the sampling window $W = W_{3r}$. The x-axis representing different values of the interaction parameter β.

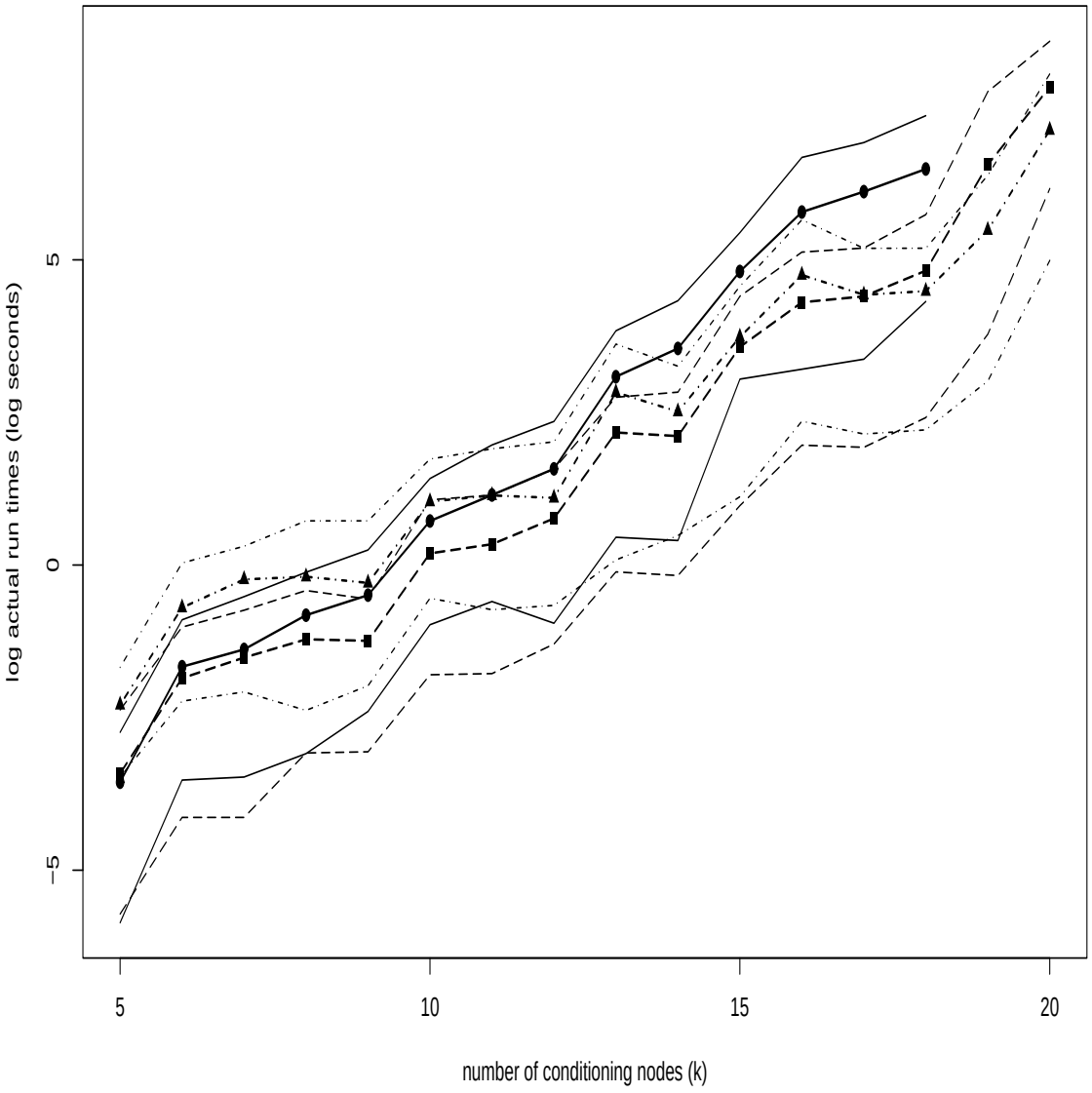

Figure 5.6: Empirical log mean run times for 2Stg (solid dots), CK (solid triangles) and Gibbs (solid squares). The solid (2Stg), dashed-dotted (CK) and dashed (Gibbs) lines represent the upper and lower simulation envelopes computed at the 90% and 10% quantiles respectively. The conditioning set $\mathcal{C}$ consists of a single cluster of k nodes, with $I(\mathcal{C})$ fixed for all k at $I^* = 2$. The radius $r = 0.1$, the interaction parameter $\beta = 10$, intensity $\lambda = 10$ and the sampling window $W = W_{4r}$. The x-axis represents k, the size of $\mathcal{C}$. The 2Stg algorithm failed to output suitable number of samples for $k = 19$; therefore the run times corresponding to 2Stg are plotted only for values up to $k = 18$.

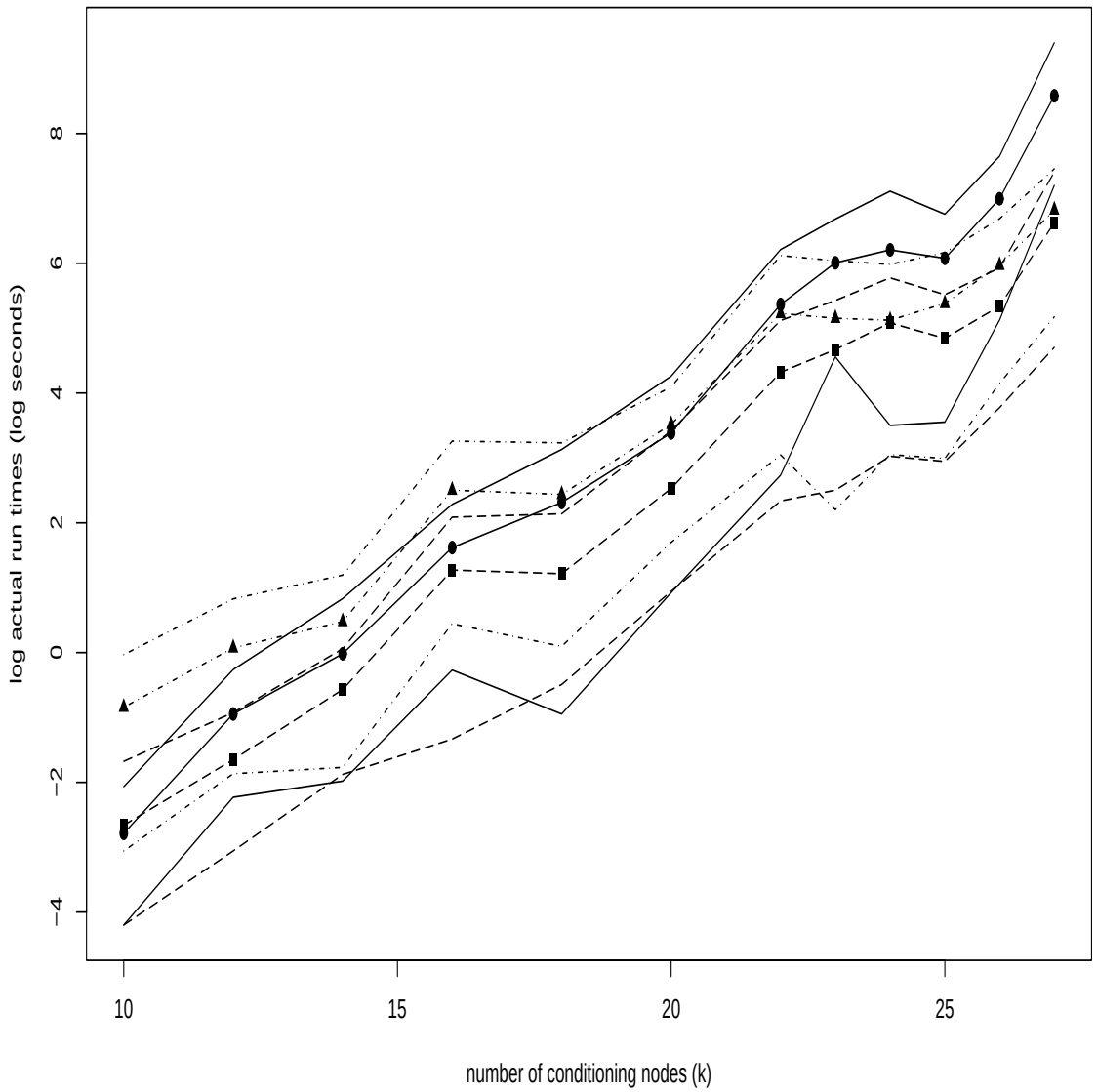

Figure 5.7: Empirical log mean run times for 2Stg (solid dots), CK (solid triangles) and Gibbs (solid squares). The solid (2Stg), dashed-dotted (CK) and dashed (Gibbs) lines represent the upper and lower simulation envelopes computed at the 90% and 10% quantiles respectively. The conditioning set $\mathcal{C}$ consists of a single cluster of k nodes, with $I^* = 2$. The radius $r = 0.1$, the interaction parameter $\beta = 5$, intensity $\lambda = 20$ and the sampling window $W = W_{2r}$. The x-axis represents k, the size of $\mathcal{C}$.

5.8 Conclusions & Further Work

In this chapter the conditional area-interaction process was introduced and various perfect simulation algorithms described. The 2-Stage Rejection method (Section 5.3) is likely to be a more efficient variant of Rejection sampling for this process. It was also seen how the Cai & Kendall (2002) Algorithm 3.4 for conditional Boolean models and the Kendall (1998) Algorithm 4.10 for the area-interaction process can be combined in order to obtain a perfect simulation algorithm for the conditional area-interaction process. The exact Gibbs-within-Metropolis Hastings Algorithm 5.4 was developed here for the same sampling task. A comparison of these three algorithms in Section 5.7 indicates that the modified Cai & Kendall algorithm is favourable for extreme model parameters (high λ, β, large W and $\mathcal{C}$), while 2-Stage Rejection or the Gibbs-Metropolis sampler are preferable for moderate values. The latter seems just as efficient, if not more so, than 2-Stage Rejection. However the results presented here are only for the attractive case. Therefore a complete evaluation of the algorithms calls for simulations results for the repulsive case as well

5.8.1 Omnithermal Sampling

Thönnes (2000) describes an implementation of the Clan algorithm of Fernández *et al.* (2002) for the conditional Boolean model. In Section 4.6 an exact omnithermal sampling algorithm for the area-interaction process using the Clan construction of Fernández *et al.* (2002) was described. It may hence be worthwhile exploring the possibility of using the clan construction in order to devise an omnithermal algorithm for the conditional area-interaction process.

5.8.2 Bayesian Cluster Modelling

McKeague & Loizeaux (2002) consider the perfect simulation of Bayesian cluster models, introduced by Baddeley & Van Lieshout (1993). Cluster models (cf. Sections 1.1.6 & 1.1.9) assume that the 'observed' pattern is generated as the *daughter process* of an 'unobserved' *parent process*. A Bayesian treatment defines the posterior density of the unobserved parents in terms of the prior density of the parents and the likelihood. The cluster processes considered are the Neyman & Scott (1958) process and *pure silhouette* models. The latter require that the observed daughter process

must be covered by the posterior parent process. Moreover under certain priors the posterior distribution turns out to be an area-interaction process (conditioned to cover the observed data points).

They apply the pure silhouette model to a test data set and use a combination of the Kendall & Møller (2000) algorithm and Rejection sampling in order to draw from the posterior distribution. As an illustration of the applicability of the algorithms developed in this chapter we sample an area-interaction process considered to cover a similar test data set. The 2-Stage Rejection algorithm is used to produce 2000 exact samples. The results are shown in Figure 5.8 below, which is in the same format as presented by McKeague & Loizeaux (2002): histograms of the number of individuals in the posterior process are illustrated on the left; contour plots of the posterior intensity on the right.

McKeague & Loizeaux (2002) report that they obtained 500 exact samples in 72 hours on a Silicon Graphics workstation. In addition they comment that their simulations took long because of the "difficulty of covering the data by the posterior", since they employed Rejection sampling in order to obtain coverage. No rigorous comparison of the run times can be made here. However the 2000 exact samples used to produce Figure 5.8 above took under 1 minute, on a PC (Pentium 4 2.67GHz, 248MB RAM) running Windows XP. The implementations of the algorithms were programmed in `Python` (version 2.3). Thus the algorithms in this chapter can be employed for such Bayesian cluster modelling. The Redwood seedlings data (Figure 1) was also studied by McKeague & Loizeaux (2002) using the Neyman & Scott model. A treatment via the pure silhouette model can also me be made using the algorithms developed here. Indeed Figure 5.1 depicts a realization of an area-interaction process conditioned to cover part of the seedlings data.

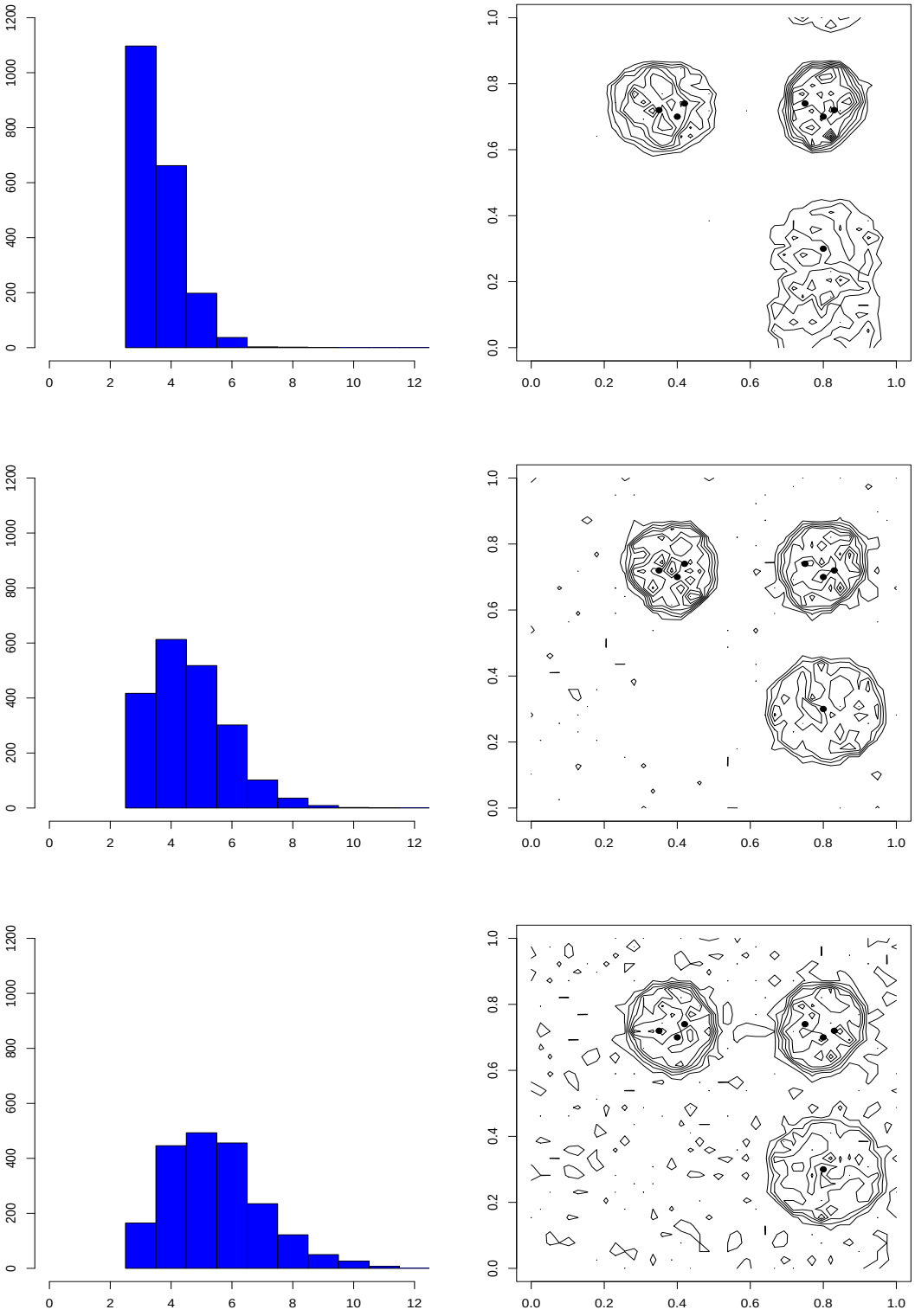

Figure 5.8: Features of the posterior (conditional area-interaction). *Left:* histograms of the number of individuals in the posterior process; *Right:* contour plots of the posterior intensity. Model parameters: $\beta = 3$, $r = 0.15$, $W = [0,1]^2$ and the test data set is represented by the solid dots. *Top:* $\lambda = 0.693$; *middle:* $\lambda = 2$; *Bottom:* $\lambda = 3$. The plots are based on 2000 exact samples using 2-Stage Rejection for the conditional area-interaction.

Chapter 6

Conclusions & Further Work

Our objective has been to introduce the ideas of stochastic simulation and explore how this can be used to examine mathematically intractable models. The emphasis here was on point process models and their 'perfect' simulation, so that an unbiased sample can be obtained in finite time. An important point to note is the necessity of *validating* any perfect simulation algorithm in order to protect against coding errors as well as theoretical discrepancies (cf. Chapter 2). An obvious direction for further work is *using* such perfect samples in order to carry out statistical inference; Murdoch & Rosenthal (2000) record some progress towards that end.

It is hoped that the importance of spatial birth-death processes has been highlighted here, since they provide a very useful way of simulating point processes. Moreover the general dominated CFTP recipe (Kendall & Møller 2000) allows, at least in principle, the perfect simulation of locally stable point processes. For point processes which are not locally stable (such as the conditional Boolean model) exact sampling may still be possible: Cai & Kendall (2002) define *virtual* processes in order to construct an exact algorithm via Monotone CFTP.

The class of point processes specified by a density with respect to the Poisson process are particularly amenable to simulation via birth-death processes, as detailed by Section 1.3.2. For such processes the Papangelou conditional intensity (defined by Eq. 1.9) plays a significant role, since it provides a means to construct a birth-death process that converges to the point process in question. If the Papangelou conditional intensity is bounded and has the necessary monotonicity properties then perfect sampling of the point process may be feasible.

Obviously whether a *practical implementation* is possibly will depend on the context. Inevitably

one may require auxiliary simulation strategies ('clever tricks') which would enable implementation. For example the 'cluster' trick (Section 4.5.2) involving the construction of observable events provides a means to implement perfect simulation of the area-interaction process. The 'cross-over' trick (Sections 3.4.2 & 4.5.4) allows the perfect sampling of anti-monotone systems such as the repulsive area-interaction process. The use of quasi-minimal and -maximal elements (Sections 4.4.3 & 5.5.1) enabled implementation of exact Gibbs samplers. We hope that such ideas will promote and inspire practical implementations of other processes which, in principle, can be perfectly simulated.

The work on omnithermal sampling (Section 4.6) also suggests that, under suitable monotonicity conditions on the Papangelou conditional intensity, such sampling may be feasible. The omnithermal results of Propp & Wilson (1996) show that a single omnithermal sample may provide more information than a number of samples at different parameter values. Thus omnithermal sampling for various models is a possibility worth considering.

Thus birth-death processes are very useful in simulating processes absolutely continuous with respect to a Poisson process. This is not to undermine the importance of discrete-time Gibbs (and more generally Metropolis-Hastings) samplers. However the key point is that spatial birth-death processes provide both a natural and powerful means for simulating point process models (including omnithermal sampling). Moreover the simulation results of Sections 3.7 & 5.7 suggest that algorithms employing birth-death processes may be more efficient for extreme model parameters than those employing a Gibbs sampler (cf. Remark 3.2).

6.1 Further Work: Omnithermal Sampling

In Section 4.7 some comments on further possibilities for omnithermal sampling were given. This section records some further thoughts.

Indeed one can phrase the whole omnithermal description for any locally stable point process X, specified by a density with respect to, say, the unit rate Poisson process. Let $\ell(\xi; x)$ denote the Papangelou conditional intensity of X; furthermore write $\ell(\xi; x) = h(\theta; \xi, x)$ as some functional of model parameters θ. For example in the case of the area-interaction process, $\theta = (\lambda, \beta, r)$ and $h(\theta) = \lambda e^{-\beta m_2[\mathcal{U}(\xi)\setminus\mathcal{U}(x)]}$. If ℓ is uniformly bounded then it is possible to define coupled spatial birth-death processes (Φ, Ψ) such that the equilibrium distribution of Ψ is that of X. The process

Φ has a Poissonian distribution with intensity equal to the uniform bound on ℓ (cf. Section 1.3.2). If omnithermal sampling with respect to parameter θ_i is to be possible, $h(\theta_1, \ldots, \theta_n; \xi, x)$ must be monotonic in θ_i and anti-monotonic in x (or vice versa, cf. Lemma 4.3):

$$h(\theta_i; \xi, x) \leq (\geq)\, h(\theta_i'; \xi, x), \qquad \text{for } \theta_i' \leq \theta_i; \tag{6.1}$$

$$\text{and} \qquad h(\theta_i; \xi, x) \geq (\leq)\, h(\theta_i; \xi, y), \qquad \text{for } y \subseteq x. \tag{6.2}$$

This then enables a comparison of $h(\theta_i; \xi, y)$ and $h(\theta_i'; \xi, x)$: $h(\theta_i; \xi, y) \leq (\geq)\, h(\theta_i'; \xi, x)$. In turn this ensures that if a birth ξ is accepted in a process with parameter θ_i then it would also be accepted in any process with parameter less (greater) than θ_i. Such monotonicity would hence guarantee the existence of an *omnithermal threshold* for the parameter θ_i (cf. Remark 4.3).

As an illustrative simple example, consider sampling the attractive area-interaction process for a range of intensities and interaction parameter values. Fix $\overline{\lambda}, \overline{\beta} > 0$ and suppose that Φ is a spatial birth-death process with birth rate $\overline{\lambda}$ and unit per capita death rate. Let $\Psi_{\lambda,\beta}$ denote an interacting birth-death process with stationary distribution an area-interaction process with intensity λ and interaction parameter β. For $0 \leq \lambda' \leq \lambda \leq \overline{\lambda}$ and $0 \leq \beta' \leq \beta \leq \overline{\beta}$, if the processes $\Psi_{\lambda',\beta}, \Psi_{\lambda,\beta'}$ are coupled to the same realization of Φ then the analogue of Lemma 4.3 would be that $\Psi_{\lambda',\beta}(t) \subseteq \Psi_{\lambda,\beta'}(t)$ for all t. Here $h(\lambda, \beta; \xi, x) = \lambda e^{-\beta m_2[\mathcal{U}(\xi)\setminus\mathcal{U}(x)]}$; so for $y \subseteq x$, $\beta' \leq \beta$ and $\lambda' \leq \lambda$:

$$h(\lambda, \beta; \xi, x) \leq h(\lambda, \beta'; \xi, x); \qquad h(\lambda, \beta; \xi, x) \geq h(\lambda, \beta; \xi, y).$$
$$h(\lambda, \beta; \xi, x) \geq h(\lambda', \beta; \xi, x).$$

Therefore Eqs. (6.1 & 6.2) are satisfied and the required inclusion $\Psi_{\lambda',\beta}(t) \subseteq \Psi_{\lambda,\beta'}(t)$ for all t follows. Thus there exists an *omnithermal region* $R_\xi \subset \left[0, \overline{\lambda}\right] \times \left[0, \overline{\beta}\right]$ in (λ, β)-space such that ξ is accepted in $\Psi_{\lambda,\beta}$ for any $(\lambda, \beta) \in R_\xi$. As a concluding note we conjecture that R_ξ is given as follows. Recall from Section 4.6 that N_ξ denotes the neighbours of ξ in Φ and, for $M \subseteq N_\xi$, the quantity $w_\xi(M)$ is computed via Eq. (4.21). Suppose the respective regions R_η for $\eta \in N_\xi$ are known; if $R_M = \bigcap_{\eta \in M} R_\eta$ and $l_\xi \sim \text{Uniform}\left(0, \overline{\lambda}\right)$ then

$$R_\xi = \bigcup_{M \subseteq N_\xi} \left(\left[l_\xi, \overline{\lambda}\right] \times [0, w_\xi(M)] \cap R_M\right). \tag{6.3}$$

Appendix A

χ^2 Tests: Conditional Boolean Model

This appendix presents the results of χ^2 goodness-of-fit tests, the objective of which is to validate the Cai & Kendall Algorithm 3.4 and exact Gibbs Algorithm 3.7 for conditional Boolean models. Theorems 3.1 & 3.2 ensure that the output of the two algorithms respectively does indeed have the required distribution. However it is still important to validate these procedures in order to guard against programming errors.

The conditioning set $\mathcal{C} = \{c_1, \ldots, c_k\}$ contains k nodes and let X_{c_i} denote the number of germ-grain pairs of the Boolean model which cover the conditioning node c_i. Given $\mathcal{C}$, the idea behind a χ^2 goodness of fit test is to compare the bin counts for $(X_{c_1}, \ldots, X_{c_k})$, where the bins represent the possible joint values $(X_{c_1}, \ldots, X_{c_k})$ can take. In the following cases a fixed number of samples are drawn for each of the two algorithms. We abbreviate the Cai & Kendall algorithm by 'CK' and the exact Gibbs algorithm by 'Gibbs'. The respective bin counts of CK and Gibbs are compared with the expected counts from the target distribution. The χ^2 test statistic s is computed as follows.

$$s = \sum_{\text{bin}} \frac{(E_{\text{bin}} - O_{\text{bin}})^2}{E_{\text{bin}}} \sim \chi^2_{C-1} \tag{A.1}$$

where the sum is taken over the bins, E_{bin} denotes the expected bin counts, O_{bin} the observed ones and C denotes the number of bins. In cases where expected counts are too difficult to compute the counts obtained from the Rejection Algorithm 3.1, which we denote by Rej_{bin}, will be used instead. Denote the respective bin counts for CK and Gibbs by CK_{bin} and $\text{Gibbs}_{\text{bin}}$. In order to test the null hypothesis that the observed counts from Rejection and, say, CK have the same distribution we use the theory of χ^2 *Tests of Homogeneity* (DeGroot & Schervish 2002) to compute the test statistic.

χ^2 Tests of Homogeneity

Suppose that random samples are taken from R different populations, each observation being classified into C different categories or bins. For $i = 1, \ldots, R$ and $j = 1, \ldots, C$ let N_{ij} denote the observed counts from the i-th population and j-th bin; furthermore let N_{i+} denote the total number of counts from the i-th population and N_{+j} that from the j-th bin. Then the test statistic

$$s = \sum_{i=1}^{R} \sum_{j=1}^{C} \frac{\left(N_{ij} - \hat{E}_{ij}\right)^2}{\hat{E}_{ij}} \sim \chi^2_{(R-1)(C-1)}; \tag{A.2}$$

$$\text{where} \qquad \hat{E}_{ij} = \frac{N_{i+} N_{+j}}{n} \qquad \text{and} \qquad n = \sum_i N_{i+}.$$

For the purposes here $R = 2$, since the objective is to test whether the counts from the Rejection and either the CK or Gibbs algorithms have the same distribution. Since the number of samples for each algorithm are the same, $n = 2N_{i+}$; denoting N_{1j} by Rej_{bin} and N_{2j} by O_{bin} we have

$$s = \sum_{\text{bin}} \frac{\left(\text{Rej}_{\text{bin}} - \text{O}_{\text{bin}}\right)^2}{\text{Rej}_{\text{bin}} + \text{O}_{\text{bin}}} \sim \chi^2_{C-1}. \tag{A.3}$$

In order to compute the statistic the observed counts O_{bin} are taken to be either CK_{bin} or $\text{Gibbs}_{\text{bin}}$. The null hypothesis is that the test statistic $s \sim \chi^2_{C-1}$ has a Chi-squared distribution on $C - 1$ degrees of freedom, where C is the total number of bins. We consider four cases, with different model parameters and determine whether any of the test statistic values are significant, in which case the null hypothesis is rejected and one can conclude that the observed counts do not have the same distribution as those from the expected (or Rejection) counts. None of the values here are significant at the 5% level, therefore the null hypothesis is accepted and the CK and Gibbs algorithms validated. The observed counts and test statistics are given in the following tables, with model parameters stated in the caption.

$(n(X_{c_1}), n(X_{c_2}), n(X_{c_3}))$	Expected	CK	Gibbs	$\frac{(\text{Expected}-\text{CK})^2}{\text{Expected}}$	$\frac{(\text{Expected}-\text{Gibbs})^2}{\text{Expected}}$
$(\leq 1, \leq 1, \leq 1)$	135	145	149	1.452	0.741
$(\leq 1, \leq 1, 2)$	107	112	106	0.009	0.234
$(\leq 1, \leq 1, \geq 3)$	74	70	84	1.351	0.216
$(\leq 1, 2, \leq 1)$	107	115	95	1.346	0.598
$(\leq 1, 2, 2)$	105	94	88	2.752	1.152
$(\leq 1, 2, \geq 3)$	94	87	96	0.043	0.521
$(\leq 1, \geq 3, \leq 1)$	74	58	89	3.041	3.459
$(\leq 1, \geq 3, 2)$	94	105	108	2.085	1.287
$(\leq 1, \geq 3, \geq 3)$	124	124	119	0.202	0.000
$(2, \leq 1, \leq 1)$	107	98	100	0.458	0.757
$(2, \leq 1, 2)$	105	105	120	2.143	0.000
$(2, \leq 1, \geq 3)$	94	86	104	1.064	0.681
$(2, 2, \leq 1)$	105	111	103	0.038	0.343
$(2, 2, 2)$	127	135	137	0.787	0.504
$(2, 2, \geq 3)$	144	148	154	0.694	0.111
$(2, \geq 3, \leq 1)$	94	85	76	3.447	0.862
$(2, \geq 3, 2)$	144	148	135	0.562	0.111
$(2, \geq 3, \geq 3)$	250	247	246	0.064	0.036
$(\geq 3, \leq 1, \leq 1)$	74	77	79	0.338	0.122
$(\geq 3, \leq 1, 2)$	94	91	101	0.521	0.096
$(\geq 3, \leq 1, \geq 3)$	124	123	137	1.363	0.008
$(\geq 3, 2, \leq 1)$	94	69	93	0.011	6.649
$(\geq 3, 2, 2)$	144	158	126	2.250	1.361
$(\geq 3, 2, \geq 3)$	250	258	238	0.576	0.256
$(\geq 3, \geq 3, \leq 1)$	124	108	123	0.008	2.065
$(\geq 3, \geq 3, 2)$	250	255	255	0.100	0.100
$(\geq 3, \geq 3, \geq 3)$	761	788	739	0.636	0.958
s				27.34114	23.22761

Table A.1: Bin counts for the CK and Gibbs algorithms compared against the expected. Model parameters are: $\lambda = 8$, $r = 0.3$ and $\mathcal{C} = \{c_1, c_2, c_3\} = \left\{(0.5, 0.375), \left(0.5 - \frac{r}{2}, 0.375 + \sqrt{3}\frac{r}{2}\right), \left(0.5 + \frac{r}{2}, 0.375 + \sqrt{3}\frac{r}{2}\right)\right\}$. The null hypothesis is that the test statistics $s \sim \chi^2_{26}$. The respective p-values for CK and Gibbs are 0.392 and 0.620; therefore neither test statistic is significant at the 5% level.

$(n(X_{c_1}), n(X_{c_2}), n(X_{c_3}), n(X_{c_4}))$	Rej	CK	Gibbs	Uncond
$(\leq 1, \leq 1, \leq 1, \leq 1)$	494	496	527	1322
$(\leq 1, \leq 1, \leq 1, \geq 2)$	164	177	174	157
$(\leq 1, \leq 1, \geq 2, \leq 1)$	166	174	169	126
$(\leq 1, \leq 1, \geq 2, \geq 2)$	51	66	61	17
$(\leq 1, \geq 2, \leq 1, \leq 1)$	227	215	233	126
$(\leq 1, \geq 2, \leq 1, \geq 2)$	111	100	92	14
$(\leq 1, \geq 2, \geq 2, \leq 1)$	119	129	103	10
$(\leq 1, \geq 2, \geq 2, \geq 2)$	41	45	37	0
$(\geq 2, \leq 1, \leq 1, \leq 1)$	162	165	174	95
$(\geq 2, \leq 1, \leq 1, \geq 2)$	59	56	54	33
$(\geq 2, \leq 1, \geq 2, \leq 1)$	61	49	50	25
$(\geq 2, \leq 1, \geq 2, \geq 2)$	12	16	16	8
$(\geq 2, \geq 2, \leq 1, \leq 1)$	158	145	163	42
$(\geq 2, \geq 2, \leq 1, \geq 2)$	85	70	66	10
$(\geq 2, \geq 2, \geq 2, \leq 1)$	72	72	59	13
$(\geq 2, \geq 2, \geq 2, \geq 2)$	18	25	22	2
s		9.234685	11.96706	

Table A.2: Bin counts for the CK and Gibbs algorithms compared against those from the Rejection algorithm (Rej). Model parameters are: $\lambda = 5$, $r = 0.19$, $\mathcal{C} = \{(0,0), (0,-0.2), (0.2,0.2), (-0.2,0.2)\}$. The null hypothesis is that the test statistics $s \sim \chi^2_{15}$. The respective p-values for CK and Gibbs are 0.865 and 0.682; therefore neither test statistic is significant at the 5% level. The observed counts from an unconditional Poisson process (Uncond) have been included to illustrate that these are markedly different to those from the conditional process.

$(n(X_{c_1}), n(X_{c_2}), n(X_{c_3}), n(X_{c_4}))$	Rej	CK	Gibbs	Uncond
$(\leq 1, \leq 1, \leq 1, \leq 1)$	267	244	256	446
$(\leq 1, \leq 1, \leq 1, \geq 2)$	197	189	167	223
$(\leq 1, \leq 1, \geq 2, \leq 1)$	115	118	118	133
$(\leq 1, \leq 1, \geq 2, \geq 2)$	72	75	70	67
$(\leq 1, \geq 2, \leq 1, \leq 1)$	110	131	131	129
$(\leq 1, \geq 2, \leq 1, \geq 2)$	77	70	66	66
$(\leq 1, \geq 2, \geq 2, \leq 1)$	94	114	113	116
$(\leq 1, \geq 2, \geq 2, \geq 2)$	63	55	54	36
$(\geq 2, \leq 1, \leq 1, \leq 1)$	86	102	82	91
$(\geq 2, \leq 1, \leq 1, \geq 2)$	108	122	140	124
$(\geq 2, \leq 1, \geq 2, \leq 1)$	118	91	111	81
$(\geq 2, \leq 1, \geq 2, \geq 2)$	104	107	106	81
$(\geq 2, \geq 2, \leq 1, \leq 1)$	90	75	97	55
$(\geq 2, \geq 2, \leq 1, \geq 2)$	111	116	112	64
$(\geq 2, \geq 2, \geq 2, \leq 1)$	185	198	198	162
$(\geq 2, \geq 2, \geq 2, \geq 2)$	203	193	179	126
s		13.84165	14.55589	

Table A.3: Bin counts for the CK and Gibbs algorithms compared against those from the Rejection algorithm (Rej). Model parameters are: $\lambda = 8$, $r = 0.3$ and $\mathcal{C} = \left\{(0.5, 0.375), \left(0.5 - \frac{r}{2}, 0.375 + \sqrt{3}\frac{r}{2}\right), \left(0.5 + \frac{r}{2}, 0.375 + \sqrt{3}\frac{r}{2}\right), (0.91, 0.91)\right\}$. The null hypothesis is that the test statistics $s \sim \chi^2_{15}$. The respective p-values for CK and Gibbs are 0.538 and 0.484; therefore neither test statistic is significant at the 5% level. The observed counts from an unconditional Poisson process (Uncond) have been included to illustrate that these are markedly different to those from the conditional process.

$(n(X_{c_1}), n(X_{c_2}), n(X_{c_3}), n(X_{c_4}), n(X_{c_5}))$	Rej	CK	Gibbs	Uncond
$(\leq 1, \leq 1, \leq 1, \leq 1, \leq 1)$	39	47	47	177
$(\leq 1, \leq 1, \leq 1, \leq 1, \geq 2)$	41	34	30	70
$(\leq 1, \leq 1, \leq 1, \geq 2, \leq 1)$	24	27	27	67
$(\leq 1, \leq 1, \leq 1, \geq 2, \geq 2)$	37	43	31	46
$(\leq 1, \leq 1, \geq 2, \leq 1, \leq 1)$	27	34	26	73
$(\leq 1, \leq 1, \geq 2, \leq 1, \geq 2)$	31	25	28	41
$(\leq 1, \leq 1, \geq 2, \geq 2, \leq 1)$	18	22	23	18
$(\leq 1, \leq 1, \geq 2, \geq 2, \geq 2)$	29	31	34	14
$(\leq 1, \geq 2, \leq 1, \leq 1, \leq 1)$	30	33	32	69
$(\leq 1, \geq 2, \leq 1, \leq 1, \geq 2)$	25	18	30	30
$(\leq 1, \geq 2, \leq 1, \geq 2, \leq 1)$	38	38	33	18
$(\leq 1, \geq 2, \leq 1, \geq 2, \geq 2)$	38	34	31	17
$(\leq 1, \geq 2, \geq 2, \leq 1, \leq 1)$	38	29	38	18
$(\leq 1, \geq 2, \geq 2, \leq 1, \geq 2)$	32	19	29	8
$(\leq 1, \geq 2, \geq 2, \geq 2, \leq 1)$	37	23	33	1
$(\leq 1, \geq 2, \geq 2, \geq 2, \geq 2)$	24	27	32	7
$(\geq 2, \leq 1, \leq 1, \leq 1, \leq 1)$	26	32	38	38
$(\geq 2, \leq 1, \leq 1, \leq 1, \geq 2)$	23	21	21	10
$(\geq 2, \leq 1, \leq 1, \geq 2, \leq 1)$	34	28	16	30
$(\geq 2, \leq 1, \leq 1, \geq 2, \geq 2)$	18	20	27	11
$(\geq 2, \leq 1, \geq 2, \leq 1, \leq 1)$	30	48	37	34
$(\geq 2, \leq 1, \geq 2, \leq 1, \geq 2)$	41	41	44	17
$(\geq 2, \leq 1, \geq 2, \geq 2, \leq 1)$	25	20	34	19
$(\geq 2, \leq 1, \geq 2, \geq 2, \geq 2)$	35	37	34	14
$(\geq 2, \geq 2, \leq 1, \leq 1, \leq 1)$	18	28	20	50
$(\geq 2, \geq 2, \leq 1, \leq 1, \geq 2)$	11	19	7	21
$(\geq 2, \geq 2, \leq 1, \geq 2, \leq 1)$	24	30	19	29
$(\geq 2, \geq 2, \leq 1, \geq 2, \geq 2)$	19	17	23	6
$(\geq 2, \geq 2, \geq 2, \leq 1, \leq 1)$	44	38	32	21
$(\geq 2, \geq 2, \geq 2, \leq 1, \geq 2)$	46	38	31	12
$(\geq 2, \geq 2, \geq 2, \geq 2, \leq 1)$	54	43	57	7
$(\geq 2, \geq 2, \geq 2, \geq 2, \geq 2)$	44	56	56	7
s		28.54307	28.57248	

Table A.4: Bin counts for the CK and Gibbs algorithms compared against those from the Rejection algorithm (Rej). Model parameters are: $\lambda = 10$, $r = 0.19$ and $\mathcal{C} = \{(0,0), (0,-0.2), (0.2,0.2), (-0.2,0.2), (0,0.4)\}$. The null hypothesis is that the test statistics $s \sim \chi^2_{31}$. The respective p-values for CK and Gibbs are 0.593 and 0.591; therefore neither test statistic is significant at the 5% level. The observed counts from an unconditional Poisson process (Uncond) have been included to illustrate that these are markedly different to those from the conditional process.

Appendix B

χ^2 Tests: Conditional Area-Interaction

This appendix presents the results of χ^2 goodness-of-fit tests, the objective of which is to validate the modified Cai & Kendall Algorithm 5.1 and the exact Gibbs Algorithm 5.4 for the conditional area-interaction process. The goodness of fit tests were carried out only for the attractive conditional area-interaction process $(\beta > 0)$. As noted in Appendix A validation of these perfect simulation algorithms is important and necessary in order to protect against coding errors. The test statistic is computed as in Eq. $(A.3)$, since calculating the expected counts is not possible. The Rejection algorithm described in Section 5.3 will be used as a basis for validating the two algorithms. For a conditioning node $c_i \in \mathcal{C}$ recall that X_{c_i} denotes the number of germ-grain pairs covering the node c_i; for the conditional area-interaction process we also include the counts for the germs covering none of the conditioning nodes, denoted by $X_0 \equiv X_\emptyset$, since the process does not have the independence property. We abbreviate the modified Cai & Kendall algorithm as 'CK' and the exact Gibbs algorithm as 'Gibbs'. The same notation for the bins and bin counts set up in Appendix A will be used here. In addition the observed bin counts from the conditional Boolean model (denoted by 'CondBM') are presented in order to illustrate that the conditional area-interaction does not 'look' like the conditional Boolean model.

$(n(X_0), n(X_{c_1}), n(X_{c_2}), n(X_{c_3}))$	Rej	CK	Gibbs	CondBM
$(\leq 1, \leq 1, \leq 1, \leq 1)$	482	487	491	122
$(\leq 1, \leq 1, \leq 1, \geq 2)$	112	105	98	57
$(\leq 1, \leq 1, \geq 2, \leq 1)$	115	113	112	45
$(\leq 1, \leq 1, \geq 2, \geq 2)$	25	26	22	11
$(\leq 1, \geq 2, \leq 1, \leq 1)$	103	100	111	63
$(\leq 1, \geq 2, \leq 1, \geq 2)$	32	25	21	23
$(\leq 1, \geq 2, \geq 2, \leq 1)$	23	24	28	24
$(\leq 1, \geq 2, \geq 2, \geq 2)$	7	5	8	7
$(\geq 2, \leq 1, \leq 1, \leq 1)$	559	587	593	628
$(\geq 2, \leq 1, \leq 1, \geq 2)$	147	147	137	226
$(\geq 2, \leq 1, \geq 2, \leq 1)$	150	144	137	234
$(\geq 2, \leq 1, \geq 2, \geq 2)$	33	32	41	103
$(\geq 2, \geq 2, \leq 1, \leq 1)$	154	125	117	216
$(\geq 2, \geq 2, \leq 1, \geq 2)$	32	34	41	96
$(\geq 2, \geq 2, \geq 2, \leq 1)$	19	38	32	105
$(\geq 2, \geq 2, \geq 2, \geq 2)$	7	8	11	40
s		11.84425	17.55983	

Table B.1: Bin counts for the CK and Gibbs algorithms compared against those from the Rejection algorithm (Rej). Model parameters are: $\lambda = 5$, $\beta = 10$, $r = 0.19$, $W = [0,1]^2$ and $\mathcal{C} = \{(0.25, 0.25), (0.75, 0.25), (0.7, 0.7)\}$. The null hypothesis is that the test statistics $s \sim \chi^2_{15}$. The respective p-values for CK and Gibbs are 0.691 and 0.287; therefore neither test statistic is significant at the 5% level. The observed counts from the conditional Boolean model (CondBM) have been included to illustrate that these are markedly different to those from the conditional area-interaction.

$(n(X_0), n(X_{c_1}), n(X_{c_2}), n(X_{c_3}))$	Rej	CK	Gibbs	CondBM
$(\leq 11, \leq 1, \leq 1, \leq 1)$	149	183	183	96
$(\leq 11, \leq 1, \leq 1, \geq 2)$	151	131	149	109
$(\leq 11, \leq 1, \geq 2, \leq 1)$	166	140	131	94
$(\leq 11, \leq 1, \geq 2, \geq 2)$	115	115	129	109
$(\leq 11, \geq 2, \leq 1, \leq 1)$	140	147	132	97
$(\leq 11, \geq 2, \leq 1, \geq 2)$	110	136	107	103
$(\leq 11, \geq 2, \geq 2, \leq 1)$	125	102	134	115
$(\leq 11, \geq 2, \geq 2, \geq 2)$	87	81	73	105
$(\geq 12, \leq 1, \leq 1, \leq 1)$	153	133	146	123
$(\geq 12, \leq 1, \leq 1, \geq 2)$	117	120	133	158
$(\geq 12, \leq 1, \geq 2, \leq 1)$	112	127	115	140
$(\geq 12, \leq 1, \geq 2, \geq 2)$	90	117	105	132
$(\geq 12, \geq 2, \leq 1, \leq 1)$	125	124	118	165
$(\geq 12, \geq 2, \leq 1, \geq 2)$	123	123	126	145
$(\geq 12, \geq 2, \geq 2, \leq 1)$	138	108	111	160
$(\geq 12, \geq 2, \geq 2, \geq 2)$	99	113	108	149
χ^2_{15}		23.05972	16.17581	

Table B.2: Bin counts for the CK and Gibbs algorithms compared against those from the Rejection algorithm (Rej). Model parameters are: $\lambda = 20$, $\beta = 6$, $r = 0.2$, $W = [0,1]^2$ and $\mathcal{C} = \{(0.25, 0.25), (0.75, 0.25), (0.7, 0.7)\}$. The null hypothesis is that the test statistics $s \sim \chi^2_{15}$. The respective p-values for CK and Gibbs are 0.083 and 0.370; therefore neither test statistic is significant at the 5% level. The observed counts from the conditional Boolean model (CondBM) have been included to illustrate that these are markedly different to those from the conditional area-interaction.

$(n(X_0), n(X_{c_1}), n(X_{c_2}), n(X_{c_3}))$	Rej	CK	Gibbs	CondBM
$(\leq 3, \leq 2, \leq 2, \leq 2)$	116	128	137	46
$(\leq 3, \leq 2, \leq 2, \geq 3)$	53	55	57	39
$(\leq 3, \leq 2, \geq 3, \leq 2)$	48	52	53	33
$(\leq 3, \leq 2, \geq 3, \geq 3)$	54	63	49	43
$(\leq 3, \geq 3, \leq 2, \leq 2)$	53	60	53	36
$(\leq 3, \geq 3, \leq 2, \geq 3)$	59	54	62	45
$(\leq 3, \geq 3, \geq 3, \leq 2)$	47	40	56	39
$(\leq 3, \geq 3, \geq 3, \geq 3)$	114	104	116	129
$(\geq 4, \leq 2, \leq 2, \leq 2)$	76	88	67	70
$(\geq 4, \leq 2, \leq 2, \geq 3)$	35	30	44	46
$(\geq 4, \leq 2, \geq 3, \leq 2)$	38	39	45	37
$(\geq 4, \leq 2, \geq 3, \geq 3)$	55	47	45	45
$(\geq 4, \geq 3, \leq 2, \leq 2)$	50	46	38	70
$(\geq 4, \geq 3, \leq 2, \geq 3)$	38	45	39	69
$(\geq 4, \geq 3, \geq 3, \leq 2)$	50	41	44	62
$(\geq 4, \geq 3, \geq 3, \geq 3)$	114	108	95	191
χ^2_{15}		6.868712	10.19867	

Table B.3: Bin counts for the CK and Gibbs algorithms compared against those from the Rejection algorithm (Rej). Model parameters are: $\lambda = 10$, $\beta = 7$, $r = 0.3$, $W = [0,1]^2$ and $\mathcal{C} = \left\{(0.5, 0.375), \left(0.5 - \frac{r}{2}, 0.375 + \sqrt{3}\frac{r}{2}\right), \left(0.5 + \frac{r}{2}, 0.375 + \sqrt{3}\frac{r}{2}\right)\right\}$. The null hypothesis is that the test statistics $s \sim \chi^2_{15}$. The respective p-values for CK and Gibbs are 0.961 and 0.8074; therefore neither test statistic is significant at the 5% level. The observed counts from the conditional Boolean model (CondBM) have been included to illustrate that these are markedly different to those from the conditional area-interaction.

$(n(X_0), n(X_{c_1}), n(X_{c_2}), n(X_{c_3}), n(X_{c_4}))$	Rej	CK	Gibbs	CondBM
$(\leq 6, \leq 1, \leq 1, \leq 1, \leq 1)$	159	176	178	29
$(\leq 6, \leq 1, \leq 1, \leq 1, \geq 2)$	36	19	17	6
$(\leq 6, \leq 1, \leq 1, \geq 2, \leq 1)$	45	47	41	20
$(\leq 6, \leq 1, \leq 1, \geq 2, \geq 2)$	15	14	15	2
$(\leq 6, \leq 1, \geq 2, \leq 1, \leq 1)$	53	38	37	11
$(\leq 6, \leq 1, \geq 2, \leq 1, \geq 2)$	6	5	4	6
$(\leq 6, \leq 1, \geq 2, \geq 2, \leq 1)$	14	14	23	3
$(\leq 6, \leq 1, \geq 2, \geq 2, \geq 2)$	7	4	5	3
$(\leq 6, \geq 2, \leq 1, \leq 1, \leq 1)$	58	56	54	15
$(\leq 6, \geq 2, \leq 1, \leq 1, \geq 2)$	40	23	35	10
$(\leq 6, \geq 2, \leq 1, \geq 2, \leq 1)$	74	66	57	18
$(\leq 6, \geq 2, \leq 1, \geq 2, \geq 2)$	51	53	55	18
$(\leq 6, \geq 2, \geq 2, \leq 1, \leq 1)$	26	18	17	6
$(\leq 6, \geq 2, \geq 2, \leq 1, \geq 2)$	8	10	11	4
$(\leq 6, \geq 2, \geq 2, \geq 2, \leq 1)$	30	32	25	9
$(\leq 6, \geq 2, \geq 2, \geq 2, \geq 2)$	24	35	23	11
$(\geq 7, \leq 1, \leq 1, \leq 1, \leq 1)$	104	109	109	197
$(\geq 7, \leq 1, \leq 1, \leq 1, \geq 2)$	11	20	14	64
$(\geq 7, \leq 1, \leq 1, \geq 2, \leq 1)$	28	27	35	49
$(\geq 7, \leq 1, \leq 1, \geq 2, \geq 2)$	3	7	11	16
$(\geq 7, \leq 1, \geq 2, \leq 1, \leq 1)$	29	26	33	68
$(\geq 7, \leq 1, \geq 2, \leq 1, \geq 2)$	6	9	10	18
$(\geq 7, \leq 1, \geq 2, \geq 2, \leq 1)$	11	8	13	28
$(\geq 7, \leq 1, \geq 2, \geq 2, \geq 2)$	5	0	2	8
$(\geq 7, \geq 2, \leq 1, \leq 1, \leq 1)$	21	24	25	53
$(\geq 7, \geq 2, \leq 1, \leq 1, \geq 2)$	25	21	17	52
$(\geq 7, \geq 2, \leq 1, \geq 2, \leq 1)$	29	39	39	63
$(\geq 7, \geq 2, \leq 1, \geq 2, \geq 2)$	35	35	40	83
$(\geq 7, \geq 2, \geq 2, \leq 1, \leq 1)$	7	18	16	21
$(\geq 7, \geq 2, \geq 2, \leq 1, \geq 2)$	5	6	6	23
$(\geq 7, \geq 2, \geq 2, \geq 2, \leq 1)$	20	29	16	43
$(\geq 7, \geq 2, \geq 2, \geq 2, \geq 2)$	15	12	17	43
χ^2_{31}		38.00843	35.89669	

Table B.4: Bin counts for the CK and Gibbs algorithms compared against those from the Rejection algorithm (Rej). Model parameters are: $\lambda = 5$, $\beta = 7$, $r = 0.2$, $W = [-0.75, 0.75]^2$ and $\mathcal{C} = \{(0,0), (0.1,-0.1), (-0.1,-0.1), (0.25, 0.12)\}$. The null hypothesis is that the test statistics $s \sim \chi^2_{31}$. The respective p-values for CK and Gibbs are 0.180 and 0.250; therefore neither test statistic is significant at the 5% level. The observed counts from the conditional Boolean model (CondBM) have been included to illustrate that these are markedly different to those from the conditional area-interaction.

Bibliography

Ambler G.K. (2002). *Dominated Coupling From The Past and Some Extensions of the Area-Interaction Process*. Ph. D. thesis, Department of Mathematics, University of Bristol, UK.

Baddeley A.J. (1999). A crash course in stochastic geometry. In Barndorff-Nielsen O.E., Kendall W.S. & Van Lieshout M.N.M. (Eds.), *Stochastic Geometry: Likelihood and Computation*, Number 80 in Monographs in Statistics & Applied Probability, Chapter 1, pp. 1–35.

Baddeley A.J. & Møller J. (1989). Nearest-neighbour Markov point processes and random sets. *International Statistical Review* 57, 89–121.

Baddeley A.J. & Van Lieshout M.N.M. (1993). Stochastic geometry models in high-level vision. In Mardia K.V. & Kanji G.K. (Eds.), *Statistics and Images: I*, Advances in Applied Statistics, a supplement to Journal of Applied Statistics 20, pp. 233–258. Carfax Publishing, Abingdon.

Baddeley A.J. & Van Lieshout M.N.M. (1995). Area-interaction point processes. *Annals of the Institute of Statistical Mathematics* 47, 601–619.

Berthelsen K.K. & Møller J. (2002a). A primer on perfect simulation for spatial point processes. *Bulletin of the Brazilian Mathematical Society* 33, 351–367.

Berthelsen K.K. & Møller J. (2002b). Spatial jump processes and perfect simulation. In Mecke K. & Stoyan D. (Eds.), *Morphology of Condensed Matter*, Volume 600 of *Lecture Notes in Physics*, pp. 391–417. Springer-Verlag.

Besag J. & Green P.J. (1993). Spatial statistics and bayesian computation. *Journal of the Royal Statistical Society, Series B* 55, 25–37.

Cai Y. & Kendall W.S. (2002). Perfect simulation for correlated Poisson random variables conditioned to be positive. *Statistics and Computing* 12(3), 229–243.

Carter D.S. & Prenter P.M. (1972). Exponential spaces and counting processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 21, 1–19.

Daley D.J. & Vere-Jones D. (1988). *An Introduction to the Theory of Point Processes*. Springer series in Statistics. Springer-Verlag, New York.

Daley D.J. & Vere-Jones D. (2003). *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory & Methods* (2nd ed.). Probability & Its Applications. Springer-Verlag.

DeGroot M.H. & Schervish M.J. (2002). *Probability and Statistics* (3rd ed.). Addison-Wesley.

Diggle P.J. (1983). *Statistical Analysis of Spatial Point Patterns*. Mathematics in Biology. Academic Press Inc., London.

Dimakos X.K. (2000). *Topics in Computer Intensive Statistical Inference*. Ph. D. thesis, Department of Mathematics, University of Oslo, Norway.

Doeblin W. (1938). Exposé de la théorie des chaînes simples constantes de markov à un nombre fini d'états. *Revue Mathematique de l'Union Interbalkanique* 2, 77–105.

Fernández R., Ferrari P. & Garcia N. (2002). Perfect simulation for interacting point processes, loss networks and ising models. *Stochastic Processes & their Applications* 102(1), 63–88.

Fill J.A. (1998). An algorithm, unbiased for user impatience, for exact sampling via markov chains. *Annals of Applied Probability* 8, 131–162.

Fill J.A., Machida M., Murdoch D.J. & Rosenthal J.S. (2000). Extension of fill's perfect rejection sampling algorithm to general chains. *Random Structures and Algorithms* 17, 290–316.

Fill J.A. & Machida M. (2001). Stochastic monotonicity and realizable monotonicity. *Annals of Probability* 29, 938–978.

Foss S.G. & Tweedie R.L. (1998). Perfect simulation and backward coupling. *Stochastic Models* 14(1-2), 187–203.

Gamerman D. (1997). *MCMC: Stochastic Simulation For Bayesian Inference*. Texts in Statistical Science. Chapman & Hall.

Gelfand A.E. & Smith A.F.M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85, 398–409.

Geman S. & Geman D. (1984). Stochastic relaxation, Gibbs distribution & the Bayesian restoration of images. *IEEE Transactions in Pattern Analysis & Machine Intelligence* 6(6), 721–741.

Geyer C.J. (1999). Likelihood inference for spatial point processes. In Barndorff-Nielsen O.E., Kendall W.S. & Van Lieshout M.N.M. (Eds.), *Stochastic Geometry: Likelihood and Computation*, Number 80 in Monographs in Statistics & Applied Probability, Chapter 3, pp. 79–140.

Gilks W.R., Richardson S. & Spiegelhalter D.J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

Green P.J. & Murdoch D.J. (1999). Exact sampling for Bayesian inference: Towards general purpose algorithms (with discussion). In Bernardo J.M., Berger J.O., Dawid A.P. & Smith A.F. M. (Eds.), *Bayesian Statistics*, Volume 6, pp. 301–321. Oxford University Press.

Grimmett G. (1989). *Percolation*. Springer-Verlag, New York.

Grimmett G. (1995). The stochastic random-cluster process and the uniqueness of random-cluster measures. *The Annals of Probability* 23(4), 1461–1510. Special invited paper.

Häggström O. & Nelander K. (1998). Exact sampling from anti-monotone systems. *Statistica Neerlandica* 52, 360–380.

Häggström O., Van Lieshout M.N.M. & Møller J. (1999). Characterization results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes. *Bernoulli* 5(4), 641–658.

Hall P. (1988). *An Introduction to the Theory of Coverage Processes*. Wiley series in Probability & Statistics. John Wiley & Sons, New York.

Hamilton W.D. (1971). Geometry for the selfish herd. *J. Theoretical Biology* 31, 295–311.

Hastings W. (1970). Monte Carlo sampling methods using Markov chains & their applications. *Biometrika* 57(1), 97–109.

Kallenberg O. (1986). *Random Measures* (4th ed.). Akademie-Verlag, Berlin.

Kamae T., Krengel U. & O'Brien G.L. (1977). Stochastic inequalities on partially ordered spaces. *The Annals of Probability* 5(6), 899–912.

Kelly F.P. (1979). *Reversibility and Stochastic Networks*. Wiley Series in Probability & Mathematical Statistics. John Wiley & Sons, Chichester.

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

Kendall W.S. (1997a). On some weighted Boolean models. In Jeulin D. (Ed.), *Advances in the Theory and Applications of Random Sets*, pp. 105–120. World Scientific Publishing Company.

Kendall W.S. (1997b). Perfect simulation for spatial point processes. In *Bulletin ISI, 51st Session Proceedings*, Volume 3, pp. 163–166. Istanbul (August 1997).

Kendall W.S (1998). Perfect simulation for the area-interaction point process. In Acardi L. & Heyde C. (Eds.), *Probability Towards 2000*, Number 128 in Lecture Notes in Statistics, pp. 218–234. Springer Velrag, New York.

Kendall W.S. & Møller J. (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Advances in Applied Probability (SGSA)* 32(3), 844–865.

Kendall W.S. & Thönnes E. (1999). Perfect simulation in stochastic geometry. *Pattern Recognition* 32(9), 1569–1586.

Kendall W.S., Van Lieshout M.N.M. & Baddeley A.J. (1999). Quermass-interaction processes: Conditions for stability. *Advances in Applied Probability (SGSA)* 31(2), 315–342.

Kingman J.F.C. (1993). *Poisson Processes*. Clarendon Press, Oxford.

Lantuéjoul C. (1997). Conditional simulation of object-based models. In Jeulin D. (Ed.), *Advances in Theory and Applications of Random Sets*, pp. 271–288. World Scientific Publishing.

Lantuéjoul C. (2002). *Geostatistical Simulation: Models and Algorithms*. Springer Velrag.

Lindvall T. (1992). *Lectures on the Coupling Method*. Wiley Series in Probability & Mathematical Statistics. John Wiley & Sons, New York.

Mase S. (1990). Mean characteristics of Gibbsian point processes. *Annals of the Institute of Statistical Mathematics* 42, 203–220.

Matérn B. (1986). *Spatial Variation* (2nd ed.). Springer-Verlag, New York.

Matheron G. (1975). *Random Sets and Integral Geometry*. John Wiley & Sons.

McKeague I.W. & Loizeaux M. (2002). Perfect sampling for point process cluster modelling. In Lawson A.B. & Denison D. (Eds.), *Spatial Cluster Modelling*, Chapter 5, pp. 87–107.

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

Meyn S.P. & Tweedie R.L. (1993). *Markov Chains and Stochastic Stability*. Springer Velrag.

Molchanov I.S. (1996). *Statistics of the Boolean Model for Practitioners and Mathematicians*. John Wiley & Sons, Chichester.

Møller J. (1993). Spatial point processes and Markov Chain Monte Carlo methods. Lecture at Conference on Stochastic Processes & their Applications, Amsterdam.

Møller J. (1999). MCMC & spatial point processes. In Barndorff-Nielsen O.E., Kendall W.S. & Van Lieshout M.N.M. (Eds.), *Stochastic Geometry: Likelihood and Computation*, Number 80 in Monographs in Statistics & Applied Probability, Chapter 4, pp. 141–172.

Møller J. (2001). A review of perfect simulation in stochastic geometry. In Basawa I.V., Heyde C.C. & Taylor R.L. (Eds.), *Selected Proceedings of the Symposium on Inference for Stochastic Processes*, Volume 37 of *IMS Lecture Notes & Monographs Series*, pp. 333–355.

Møller J. & Schladitz K. (1999). Extensions of fill's algorithm for perfect simulation. *Journal of the Royal Statistical Society, Series B* 61(4), 955–969.

Moyal J.E. (1962). The general theory of stochastic population processes. *Acta Math.* 108, 1–31.

Murdoch D.J. & Green P.J. (1998). Exact sampling from a continuous state space. *Scandinavian Journal of Statistics* 25(3), 483–502.

Murdoch D.J. & Rosenthal J.S. (2000). Efficient use of exact samples. *Statistics and Computing* 10, 237–243.

Neyman J. & Scott E.L. (1958). Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society, Series B* 20(2), 1–43.

Norris J.R. (1997). *Markov Chains*. Cambridge series in Statistical & Probabilistic Mathematics. Cambridge University Press.

Nummelin E. (1984). *General Irreducible Markov Chains and Non-negative Operators*. Cambridge University Press.

Palm C. (1943). Intensitätsschwankungen im fernsprechverkehr. *Ericsson Techniks* 44, 1–89.

Papangelou F. (1974). The conditional intensity of general point processes and an application to line processes. *Z. Wahrscheinlichkeitstheorie und Verwandte Gebiete* 28, 207–226.

Preston C. (1977). Spatial birth-and-death processes. *Bulletin of the International Statistical Institute* 46(2), 371–391.

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

Propp J.G. & Wilson D.B. (1998). Coupling From The Past: A user's guide. In Aldous D. & Propp J. (Eds.), *Microsurveys in Discrete Probability*, Volume 41 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pp. 181–192.

Ripley B. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society, Series B* 39(2), 172–192.

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

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

Ripley B.D. & Kelly F.P. (1977). Markov point processes. *Journal of the London Mathematical Society* 15, 188–192.

Ruelle D. (1969). *Statistical Mechanics*. John Wiley & Sons, New York.

Serra J. (1982). *Image Analysis and Mathematical Morphology*. Academic Press, London.

Shah S.R. (2003a). A note on stochastic domination and conditional thinning. *Advances in Applied Probability (SGSA)* 5(4), 937–940.

Shah S.R. (2003b). Stochastic domination and conditional thinning in spatial point processes. Technical Report 412, University of Warwick, UK.

Shaked M. & Shanthikumar J.G. (1994). *Stochastic Orders and their Applications*. Academic Press Inc., San Diego.

Stoyan D., Kendall W.S. & Mecke J. (1995). *Stochastic Geometry and its Applications* (2nd ed.). Wiley series in Probability & Statistics. John Wiley & Sons, Chichester.

Strassen V. (1965). The existence of probability measures with given marginals. *The Annals of Mathematical Statistics* 36(2), 423–439.

Strauss D.J. (1975). A model for clustering. *Biometrika* 62(2), 467–475.

Thönnes E. (1997). A primer on perfect simulation. Technical report, Chalmers University.

Thönnes E. (1998). *Perfect and Imperfect Simulations in Stochastic Geometry*. Ph. D. thesis, Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK.

Thönnes E. (1999). Perfect simulation of some point processes for the impatient user. *Advances in Applied Probability* 31(1), 69–87.

Thönnes E. (2000). The conditional Boolean model revisited. Technical Report 369, University of Warwick.

Van Lieshout M.N.M (2000). *Markov Point Processes and their Applications*. Imperial College Press, London.

Van Lieshout M.N.M. & Van Zwet E.W. (2001). Exact sampling from conditional Boolean models with applications to maximum likelihood inference. *Advances in Applied Probability (SGSA)* 33(2), 339–353.

Van Lieshout M. N. M. & Van Zwet E. W. (2003). Correction to: Exact sampling from conditional Boolean models with applications to maximum likelihood inference. *Advances in Applied Probability (SGSA)* 35(2), 362–362.

Widom B. & Rowlinson J.S. (1970). New model for the study of liquid-vapor phase transition. *Journal of Chemical Physics* 52, 1670–1684.

Wilson D.B. (1996). *Exact Sampling with Markov Chains*. Ph. D. thesis, Department of Mathematics, Massachusetts Institute of Technology, USA.

Wilson D.B. (2000). How to couple from the past using a read-once source of randomness. *Random Structures and Algorithms* 16(1), 85–113.