

EXACT SAMPLING FROM CONDITIONAL BOOLEAN MODELS WITH APPLICATIONS TO MAXIMUM LIKELIHOOD INFERENCE

M. N. M. VAN LIESHOUT,* *CWI*

E. W. VAN ZWET,** *University of California, Berkeley*

Abstract

We are interested in estimating the intensity parameter of a Boolean model of discs (the bombing model) from a single realization. To do so, we derive the conditional distribution of the points (germs) of the underlying Poisson process. We demonstrate how to apply coupling from the past to generate samples from this distribution, and use the samples thus obtained to approximate the maximum likelihood estimator of the intensity. We discuss and compare two methods: one based on a Monte Carlo approximation of the likelihood function, the other a stochastic version of the EM algorithm.

Keywords: Boolean model; coupling from the past; Markov chain Monte Carlo simulation; maximum likelihood estimation; stochastic approximation EM algorithm; stochastic EM algorithm

AMS 2000 Subject Classification: Primary 60D05; 62M30

1. Introduction

Many images found in microscopy, materials science and biology can be described by means of random sets. Perhaps the best known model is the *Boolean model* (Matheron (1975)) formalizing a configuration of independent, randomly placed particles. A Boolean model is formed by placing random closed sets at the points of a Poisson process and taking the union of these sets. The points of the Poisson process are sometimes called the *germs*, the associated random sets the *grains* or *particles*. In spite of the strong independence assumptions, inference for Boolean models is far from trivial (Molchanov (1997)). The difficulty is that as we observe the union of the grains, certain grains may be completely covered by the others. In this paper a method is presented to obtain samples from a Boolean model conditioned on certain events. In particular, such an event might be coverage of a given set. We apply our method to estimate the intensity of the Poisson process underlying a Boolean model of balls.

Molchanov (1997) distinguishes between two types of parameters of a Boolean model: aggregate (or macroscopic) and individual (or microscopic). Typical examples of aggregate parameters are the area fraction and the set-covariance. They can easily be estimated by their empirical counterparts. The resulting estimators are unbiased and expressions for the variance can be obtained from Robbins' theorem (cf. Stoyan *et al.* (1995)). Under mild ergodicity assumptions they are strongly consistent (Molchanov (1997)) as the observation window expands to the entire plane. Aggregate functionals such as the contact distribution and pair correlation function are of interest when fitting the Boolean model to data. Usually, estimation

Received 2 November 2000; revision received 1 March 2001.

* Postal address: CWI, PO Box 94079, 1090 GB Amsterdam, The Netherlands. Email address: colette@cwi.nl

** Postal address: University of California, Department of Statistics, 367 Evans Hall # 3860, Berkeley, CA 94720-3860, USA.

is hampered by edge effects, but minus sampling ideas (Ripley (1988), Stoyan *et al.* (1995)) are generally applicable as are Horvitz–Thompson style estimators including the Kaplan–Meier (Baddeley and Gill (1997)) and Hanisch estimators (Hanisch (1984)). Unbiasedness follows from the Campbell–Mecke theorem (Stoyan *et al.* (1995)) and asymptotic results are available (Molchanov (1997)).

Individual parameters, including the intensity of the germ process, are much harder to estimate. Minimum contrast methods (Dupač (1980), Serra (1982)) for the intensity are based on minimizing the distance between an estimated aggregate parameter (e.g. the contact distribution) and an approximation expressed in terms of the intensity. Some asymptotic results are available, but the expressions for the asymptotic variance are too complicated to be useful in practice. An alternative is the method of moments based on coverage fraction, mean area, boundary length and Euler–Poincaré characteristic. This method is computationally easy but leads to biased estimators (Weil (1988)). In the tangent point approach, the Euler–Poincaré characteristic is replaced by the specific connectivity number resulting in easier asymptotics (Molchanov and Stoyan (1994)). Other methods are based on the fully exposed balls only (Ayala *et al.* (1990)). Further details and some other special methods such as Schmitt’s (1991) can be found in Molchanov (1997) and the references therein.

In this paper we shall take a likelihood based approach using Monte Carlo methods to perform the necessary computations. To do so, we need to sample from the conditional distribution of a Boolean model given an observation of the union of its particles. In the next section we show that this distribution is straightforward if the grains are balls with random radius (or indeed many other shapes, see Section 6). Unfortunately, due to an intractable normalizing constant, direct sampling usually is not possible. In Section 3 we use coupling from the past (CFTP) (Propp and Wilson (1996)) to obtain the desired samples and we discuss how our sampling method is related to the work of others, in particular Kendall and Møller (2000).

Section 4 is devoted to two approaches to maximum likelihood estimation through simulation. The first method is based on a Monte Carlo approximation of the likelihood function (Geyer (1999)). The other approach is a stochastic version of the expectation maximization (EM) algorithm (Dempster *et al.* (1977), Celeux and Diebolt (1985)). At each E-step the conditional expectation of the likelihood is replaced by a simulation average. This approximate expectation is optimized in the M-step to obtain a new parameter value. Under certain conditions an ergodic Markov chain on the parameter space is obtained. Coupling from the past can again be applied to obtain exact samples from the stationary distribution of the stochastic EM chain.

In Section 5 we present the results of a modest simulation experiment which we conducted to compare the two methods. Section 6 briefly discusses some generalizations to random set models with stochastic primary grains as well as inter-particle interactions.

2. A conditional Boolean model

Suppose a realization of a Boolean model is observed and we want to estimate the intensity of the underlying point process of germs. Direct maximum likelihood estimation seems impossible because the conditional distribution of the germs given the observed union of particles involves an intractable normalizing constant depending on the intensity parameter of interest (cf. Lemma 1). However, if it were possible to sample from this conditional distribution, Monte Carlo based maximum likelihood estimation would be an option.

In this paper, we focus our attention on the case where the particles are (random) balls. Then the location of a germ is identified upon observation of any part of the boundary of its associated ball. The remaining, not directly identifiable germs turn out to be distributed as

a Poisson process conditioned to satisfy a coverage condition. The remainder of this section makes this claim more precise. We start with a formal definition of the Boolean model of balls.

Definition 1. Let X be a stationary Poisson process with intensity $\lambda > 0$ on \mathbb{R}^d , and $B = B(0, 1)$ the d -dimensional closed unit ball centred at the origin. Then, writing $A \oplus B = \{a + b : a \in A, b \in B\}$ for the Minkowski addition of A and B ,

$$\mathcal{B}(X) = \bigcup_{x \in X} (x \oplus rB)$$

is a Boolean model of balls with radius $r > 0$ on \mathbb{R}^d .

For $d = 2$, the process of Definition 1 is sometimes referred to as the *bombing model* (Kolmogorov (1937), Matérn (1960)). The underlying points X are called the *germs*, the set B is the *primary grain*. More general Boolean models are obtained by letting the germs be scattered according to a non-stationary Poisson process or by allowing the grains to be arbitrary random closed sets (Matheron (1975)). We will return to this briefly in Section 6.

The goal of this paper is to estimate the intensity λ based on an observation of the intersection of $\mathcal{B}(X) \cap W$ for some compact sampling window W . By the local knowledge principle (Serra (1982)) and the symmetry of B ,

$$\mathcal{B}(X \cap (W \oplus rB)) \cap W = \mathcal{B}(X) \cap W.$$

In other words, the data $Y = \mathcal{B}(X) \cap W$ depend on X only through $X \cap (W \oplus rB)$.

Since the primary grains are balls, the position of a germ is identified whenever a part of its associated grain's boundary is exposed. Hence, the conditional distribution of $X \cap (W \oplus rB)$ can be decomposed into an 'exposed boundary' part X^b and a stochastic 'interior' part X^i of germs that cannot be uniquely identified. Of course, for given Y the conditional distribution of the exposed points is degenerate at a configuration \mathbf{x}_Y^b . To derive the conditional distribution of the interior germs, set

$$\mathcal{C} = Y \setminus \mathcal{B}(\mathbf{x}_Y^b), \tag{1}$$

$$\mathcal{D} = \{y \in W \oplus rB : (y \oplus rB) \cap W \subseteq Y\}. \tag{2}$$

Clearly, \mathcal{C} and \mathcal{D} depend on Y , but we suppress this dependence on Y in our notation.

Lemma 1. Let X be a stationary Poisson process on \mathbb{R}^d with intensity $\lambda > 0$ and $\mathcal{B}(X)$ be the associated Boolean model of balls with fixed radius $r > 0$ as in Definition 1. Then, conditionally on $Y = \mathcal{B}(X) \cap W$ for some compact window $W \subseteq \mathbb{R}^d$, the interior germ process X^i is distributed as a Poisson process on \mathcal{D} with intensity λ conditional on the event that \mathcal{C} is covered by $\mathcal{B}(X^i)$.

We write $\pi_{\lambda|Y}$ for this conditional distribution of X^i given Y .

Proof. Any configuration \mathbf{y} such that $\mathcal{B}(\mathbf{y}) \cap W = Y$ has to be of the form $\mathbf{y} = \mathbf{x}^b \cup \mathbf{x}^i$ with \mathbf{x}^i a subset of \mathcal{D} such that $\mathcal{B}(\mathbf{x}^i) \supseteq \mathcal{C}$. Moreover, its density (with respect to a unit rate Poisson process on $W \oplus rB$) equals

$$e^{-(1-\lambda)|W \oplus rB|} \lambda^{n(\mathbf{x}^i)} \lambda^{n(\mathbf{x}^b)}.$$

Hence, the conditional density of the interior points is proportional to

$$\lambda^{n(\mathbf{x}^i)} \mathbf{1}\{\mathbf{x}^i \subseteq \mathcal{D}\} \mathbf{1}\{\mathcal{B}(\mathbf{x}^i) \supseteq \mathcal{C}\}.$$

Consequently, the interior germs constitute a Poisson process on \mathcal{D} conditional on the event $\{\mathcal{B}(X^i) \supseteq \mathcal{C}\}$.

The probability of the event $\{\mathcal{B}(X^i) \supseteq \mathcal{C}\}$ is not easy to compute. Hall (1988) gives a lower bound for the probability of coverage of the unit square by a two-dimensional Boolean model with intensity λ of discs of radius r ($\lambda \geq 1$, $0 < r \leq \frac{1}{2}$):

$$1 - 3 \min(1, (1 + \pi r^2 \lambda^2) \exp\{-\pi r^2 \lambda\}).$$

Since a Poisson process of intensity $\kappa > \lambda$ can be written as the independent superposition of Poisson processes with intensity λ and $\kappa - \lambda$, it follows that the probability of a Boolean model with intensity λ covering \mathcal{C} is strictly increasing in λ . Moreover, by Hall's inequality this probability actually increases to one. Hence, for λ large enough, it is feasible to obtain samples from $\pi_{\lambda|Y}$ by rejection sampling. For other values of λ this approach will be too slow. In the next section, we suggest an alternative using coupling from the past (Propp and Wilson (1996)).

3. Coupling from the past for the conditional Boolean model

Recall that the distribution of a homogeneous Poisson point process on a compact set arises as the equilibrium of a spatial birth-and-death process (Feller (1968), Preston (1977)). Starting from any initial configuration, each point is deleted after an exponential lifetime of rate 1. New, uniformly distributed points are added at a rate that is equal to the intensity of the target Poisson process. Lantuéjoul (1997) shows that a similar method can be used to obtain the law of a Poisson process conditioned on an event \mathcal{E} of positive probability. Informally, if whenever a point is added or deleted according to the scheme described above we make sure never to enforce a transition violating \mathcal{E} , the stationary distribution of the resulting process corresponds to the conditional version of the Poisson process.

Kendall and Møller (2000) show how to apply coupling from the past to Lantuéjoul's chains to obtain exact samples from their stationary distribution. However, their method is restricted to events \mathcal{E} such that if a configuration \mathbf{x} satisfies \mathcal{E} then $\mathbf{x} \setminus \{x\}$ also satisfies \mathcal{E} for all $x \in \mathbf{x}$. Coverage of a set is not such an event. Kendall and Thönnnes (1999) do have another CFTP-based algorithm to tackle the conditioning event that a finite set of points be covered. Unfortunately, this method does not seem to extend to our situation: coverage of a non-countable set. We present a CFTP-based method which is more general than the algorithm of Kendall and Thönnnes and in a sense mirrors the algorithm of Kendall and Møller. It will work for events \mathcal{E} such that, if a configuration \mathbf{x} satisfies \mathcal{E} , then, for any x , $\mathbf{x} \cup \{x\}$ also satisfies \mathcal{E} . Many of the ideas of Kendall and Møller (2000) and Kendall and Thönnnes (1999) play a role in our construction also. See Thönnnes (2000) for a comparison between the above-mentioned methods.

Exact simulation, as opposed to approximate sampling by running a Markov chain for a long time (cf. Gilks *et al.* (1996)), was introduced by Propp and Wilson (1996). Let us denote by $X_{s \in [-T, 0]}(\mathbf{x})$ a Markov chain X_s for s in the time-interval $[-T, 0]$, initiated at $X_{-T} = \mathbf{x}$. Now, to briefly explain the basic idea, imagine that coupled processes $X_{s \in [-T, 0]}(\mathbf{x})$ are run in parallel, one for every possible initial state \mathbf{x} . Now suppose that for some $T > 0$ the influence of the initial state wears off before time 0. That is, there is a time $s \in [-T, 0]$ such that $X_s(\cdot) \equiv \tilde{\mathbf{x}}$ for some configuration $\tilde{\mathbf{x}}$. Suppose also that the coupling ensures that all paths remain identical from this time s up to time 0. Then the common path on $[s, 0]$ can be seen as the last part of a process started at time $-T$ in an initial state selected according to the stationary distribution. Hence, X_0 is an exact sample from the stationary distribution. It is very convenient if the state

space admits a partial ordering with minimal and maximal elements and if the sampling process respects this ordering. In that case only two coupled processes need be considered: a *lower process* starting in the minimal state and an *upper process* starting in the maximal state.

In our context, upon observation of $Y = \mathcal{B}(X) \cap W$, the state space of interior points consists of all finite subsets of \mathcal{D} . The natural partial ordering is the inclusion ordering, but then there is no maximal state because \mathcal{D} itself is infinite. To overcome this problem, we shall construct a birth-and-death process on a state space consisting of all subsets of a finite (but random) ‘maximal’ set. Note that this set must meet the coverage condition. This process also converges to $\pi_{\lambda|Y}$ but is amenable to coupling from the past. We will use a stochastic, varying ‘minimal’ process.

As noted in Section 2, for large intensity parameters rejection sampling can be used to obtain a realization of $\pi_{\lambda|Y}$. Using this observation, the first step of our algorithm is to generate a sample, say $D = \{z_1, z_2, \dots, z_n\}$, from $\pi_{\kappa|Y}$ for some $\kappa > \lambda$. The pattern D will serve as maximal state. All configurations obtained when running the birth-and-death processes will be subsets of D . The second step is to thin D by independently retaining each point with probability λ/κ .

The key result is the following.

Proposition 1. *Let Y be a realization of a Boolean model of balls of radius r observed in a compact window W , and define \mathcal{C} as in (1). Let D have distribution $\pi_{\kappa|Y}$ (cf. Lemma 1), and suppose $E(0)$ is an independent thinning of D with retention probability λ/κ . Then the conditional distribution of $E(0)$ given that $\mathcal{B}(E(0))$ covers the set \mathcal{C} is $\pi_{\lambda|Y}$.*

Proof. Let X_1, X_2, \dots be independent Poisson point processes on \mathcal{D} with intensity $\kappa > \lambda$, and, independently for each $i = 1, 2, \dots$, let Y_i be an independent thinning of X_i with retention probability λ/κ . Then Y_i is a Poisson process with intensity λ . Define Z to be the first Y_i that covers \mathcal{C} . Then, clearly, Z is distributed according to $\pi_{\lambda|Y}$. Suppose that only those X_i that cover \mathcal{C} are considered, which form a sequence of independent point processes distributed according to $\pi_{\kappa|Y}$. If we had thinned these and waited for the first thinning to cover \mathcal{C} , we would have found exactly the same Z . This proves the claim.

Whereas the (random) maximal state D remains fixed throughout the algorithm, the minimal state varies in time. Conditionally on $D = \{z_1, \dots, z_n\}$ we define the minimal process $E(t)$ ($t \geq 0$) as follows. The state space of $E(\cdot)$ consists of all subsets of D . The process $E(0)$ is an independent thinning of D with retention probability λ/κ as in Lemma 1. The dynamics of the process are such that points are removed from $E(\cdot)$ after an exponentially distributed ‘holding’ time of rate 1. Whenever a point is removed, it is added again after an exponentially distributed holding time of rate $\lambda/(\kappa - \lambda)$. All holding times are independent of each other and all other random variables involved. Points $z_i \in E(0)$ are treated as if they were added at time 0, and points $z_i \notin E(0)$ are treated as if they were removed at time 0.

Lemma 2. *Conditionally on D , the spatial birth-and-death process $E(t)$, $t \geq 0$, is an irreducible, homogeneous, positive recurrent Markov process. Moreover, $E(t)$ is in equilibrium and time-reversible.*

In particular $E(t)$ is distributed as $E(0)$ for all $t \leq 0$.

Proof. As the $E(\cdot)$ process is defined conditionally on $D = \{z_1, z_2, \dots, z_n\}$, it has only a finite number of points at any time. Clearly, the transition rates do not change in time, hence $E(\cdot)$ is homogeneous. Since the state space is finite, explosion is prevented, implying that the

Markov process is well-defined by the given rates. Moreover, any state $\mathbf{x} \subseteq D$ can be reached from any other state \mathbf{x}' (say) by successively deleting the points in \mathbf{x}' followed by addition of the points in \mathbf{x} , and any state is revisited almost surely. Therefore, the birth-and-death process possesses a unique invariant probability distribution (see e.g. Chapter 7 in Parzen (1962)).

Let us call this (conditional) equilibrium distribution μ_D . To determine it, consider the detailed balance equations

$$\frac{\lambda}{\kappa - \lambda} \mu_D(\mathbf{x}) = \mu_D(\mathbf{x} \cup \{z_i\}),$$

where $z_i \notin \mathbf{x} \subseteq D$. Now, since $E(0)$ is obtained from D by independent thinning with retention probability λ/κ , $\mu_D(\mathbf{x}) = P(E(0) = \mathbf{x} \mid D)$ satisfies these equations. Hence $E(t)$, $t \geq 0$ is time-reversible and in equilibrium.

When designing a coupling from the past algorithm, it is helpful to consider a forward version first. Therefore, we will start by defining a spatial birth-and-death process $Z(t)$, $t \geq 0$, in the spirit of Lantuéjoul (1997). The state space of $Z(\cdot)$ consists of the subsets of D . Given an initial state $Z(0)$, the only transitions are births and deaths. Whenever a point is added, say $z \in D$, it *might be* removed again after an exponential ‘holding time’ of rate 1. However, z is *actually* removed only if it does not cause part of \mathcal{C} to become uncovered, that is if

$$B(z, r) \cap \mathcal{C} \subseteq \mathcal{B}(Z(t^-) \setminus \{z\}) \cap \mathcal{C}. \tag{3}$$

If z cannot be deleted, it is granted an additional exponential lifetime of rate 1 after which its removal is re-evaluated. Following the terminology of Kendall and Thönnies (1999) such points will be called *perpetuated*. Whenever a point is removed it is added again after an exponential holding time with rate $\lambda/(\kappa - \lambda)$. All holding times are independent of each other and all other random variables. Points in $Z(0)$ are treated as if they were added at time 0, the points of D that were not in $Z(0)$ are removed at time 0.

Proposition 2. *Let Y be a realization of a Boolean model of balls of radius r observed in a compact window W and define \mathcal{C} as in (1). Let D have distribution $\pi_{\kappa|Y}$ (cf. Lemma 1). Conditionally on D , the spatial birth-and-death process $Z(t)$, $t \geq 0$, is homogeneous and has a single positive recurrent class consisting of those subsets z of D for which $\mathcal{B}(z)$ covers \mathcal{C} . Furthermore, $Z(t)$ tends in distribution to $\pi_{\lambda|Y}$ as $t \rightarrow \infty$.*

Proof. We work conditionally on $D = \{z_1, z_2, \dots, z_n\}$. First note that almost surely $\mathcal{B}(Z(t))$ will cover \mathcal{C} for some $t \geq 0$. The transition mechanism then ensures that $\mathcal{B}(Z(s)) \supseteq \mathcal{C}$ for all $s \geq t$. Moreover, the class \mathcal{C} of configurations $z \subseteq D$ whose associated Boolean model covers \mathcal{C} is irreducible, since any state z can be reached from any other state z' (say) by successively adding all points of $D \setminus z'$, then deleting those of $D \setminus z$. Thus $Z(\cdot)$ is well-defined, with a unique invariant distribution concentrated on \mathcal{C} (Parzen (1962)). Since the state space is finite, not all stationary probabilities can be zero, hence they are all positive, and the class \mathcal{C} is positive recurrent.

Next, restrict the process $Z(\cdot)$ to \mathcal{C} . Clearly, the birth rate of $Z(\cdot)$ is identical to that of $E(\cdot)$ and the death rate for both perpetuated and non-perpetuated points is 1, as it is for $E(\cdot)$, provided the coverage condition is not violated. Thus, the detailed balance conditions for $E(\cdot)$ and $Z(\cdot)$ coincide on \mathcal{C} . Consequently, the limit distribution of $Z(t)$ is that of $E(\cdot)$ restricted to \mathcal{C} . By Lemmas 1 and 2 the result follows.

We shall now describe how to apply coupling from the past to obtain a sample from $\pi_{\lambda|Y}$, the conditional distribution of interior points given a realization Y of our Boolean model in the

sampling window W . Fix a time $-T < 0$. Let $D(t) \equiv D$ for all $-T \leq t \leq 0$. By Lemma 2, $E(\cdot)$ is time-reversible and hence it can easily be extended backwards from $E(0)$ until time $-T$ to serve as a stochastically varying minimal process. Next, introduce an upper process $U_{-T}(\cdot)$ initialized in the maximum $U_{-T}(-T) = D$ and a lower process $L_{-T}(\cdot)$ starting in the current minimum $E(-T)$. The dynamics of these upper and lower processes are similar to those of $Z(\cdot)$, except for the fact that in order to ensure that the inclusion

$$E(t) \subseteq L_{-T}(t) \subseteq U_{-T}(t) \subseteq D \quad (4)$$

is preserved for all $t \in [-T, 0]$ we have to apply the ‘cross-over’ trick (Kendall (1997), Häggström and Nelander (1998)). The appropriate condition for deletion of z_i from the lower process at time t is

$$B(z_i, r) \cap \mathcal{C} \subseteq \mathcal{B}(U_{-T}(t^-) \setminus \{z_i\}) \cap \mathcal{C}, \quad (5)$$

while removal of z_i from $U_{-T}(t^-)$ is enforced only if

$$B(z_i, r) \cap \mathcal{C} \subseteq \mathcal{B}(L_{-T}(t^-) \setminus \{z_i\}) \cap \mathcal{C}. \quad (6)$$

To make sure that (4) is always respected, the times of the births, deaths and deaths after perpetuation must be coordinated. The births and deaths of $L(\cdot)$ and $U(\cdot)$ are governed by those of $E(\cdot)$. To regulate the deaths after perpetuation, we associate with every point $z_i \in D$ a unit rate Poisson process Ξ_i on $\{t : z_i \notin E(t)\}$.

Summarizing, we propose the following algorithm.

Algorithm 1. Generate a random sample D from $\pi_{\kappa|Y}$ for some $\kappa > \lambda$ and delete each point independently with probability $1 - (\lambda/\kappa)$ to obtain $E(0)$. Set $T = 1$ and write $\lfloor T/2 \rfloor$ for the integer part of $T/2$.

- Extend $E(\cdot)$ backwards on $[-T, -\lfloor T/2 \rfloor]$ with birth rate $\lambda/(\kappa - \lambda)$ and death rate 1.
- Extend independent unit rate Poisson processes Ξ_i backwards on $\{-T \leq t < \lfloor T/2 \rfloor : z_i \notin E(t)\}$.
- Set $L_{-T}(-T) = E(-T)$ and $U_{-T}(-T) = D$.
- At a birth transition $E(t^+) = E(t) \cup \{z_i\}$, add z_i to $L_{-T}(t)$ and $U_{-T}(t)$.
- At a death transition $E(t^+) = E(t) \setminus \{z_i\}$ or a jump of Ξ_i :
 - (i) delete (if present) z_i from $L_{-T}(t)$ provided that does not cause the Boolean model associated with $U_{-T}(t)$ to uncover part of \mathcal{C} , that is, z_i may be deleted only if (5) holds;
 - (ii) delete (if present) z_i from $U_{-T}(t)$ provided that does not cause the Boolean model associated with $L_{-T}(t)$ to uncover part of \mathcal{C} , that is, z_i may be deleted only if (6) holds.
- If $L_{-T}(0) = U_{-T}(0)$ exit; otherwise double T and repeat.

Algorithm 1 is designed in such a way that several inclusion relations hold (Kendall and Møller (2000)).

Lemma 3. *The processes $E(t)$, $L_{-T}(t)$, $U_{-T}(t)$ ($t \leq 0$) and D satisfy the following relations:*

- (i) (‘sandwiching’) $E(t) \subseteq L_{-T}(t) \subseteq U_{-T}(t) \subseteq D$, for all $-T \leq t \leq 0$;
- (ii) (‘funneling’) $L_{-T}(t) \subseteq L_{-S}(t) \subseteq U_{-S}(t) \subseteq U_{-T}(t)$, for all $-S \leq -T \leq t \leq 0$;
- (iii) (‘coalescing’) if, for some s , $L_{-T}(s) = U_{-T}(s)$, then $L_{-T}(t) = U_{-T}(t)$, for all $t \geq s$.

Proof. By definition, $E(-T) = L_{-T}(-T) \subseteq U_{-T}(-T) = D$, hence the sandwiching property holds for $t = -T$. Also D contains all other sets. Since births in $E(\cdot)$ are mimicked in $L_{-T}(\cdot)$ and $U_{-T}(\cdot)$, the inclusion relationship is preserved under birth transitions. Next, consider a death at some time $t \in [-T, 0]$, say $E(t^+) = E(t) \setminus \{z_i\}$ or a jump at time t of Ξ_i . Suppose that $E(t) \subseteq L_{-T}(t) \subseteq U_{-T}(t)$. Since Ξ_i is restricted to the set $\{t \leq 0 : z_i \notin E(t)\}$, $E(t^+)$ is a subset of $L_{-T}(t^+)$ and $U_{-T}(t^+)$. Furthermore, if z_i dies in the upper process,

$$\begin{aligned} B(z_i, r) \cap \mathcal{C} &\subseteq \mathcal{B}(L_{-T}(t) \setminus \{z_i\}) \cap \mathcal{C} \\ &\subseteq \mathcal{B}(U_{-T}(t) \setminus \{z_i\}) \cap \mathcal{C} \end{aligned}$$

and consequently z_i also dies in the lower process.

Turning to the funnelling property, we have to show that $L_{-T}(t) \subseteq L_{-S}(t)$ and that $U_{-S}(t) \subseteq U_{-T}(t)$. Now, by definition $E(-S) = L_{-S}(-S)$. Since the dynamics of Algorithm 1 preserve inclusion, it follows that $L_{-T}(-T) = E(-T) \subseteq L_{-S}(-T)$ and, more generally, $L_{-T}(t) \subseteq L_{-S}(t)$ for any $t \geq -T$. Regarding the upper process, $U_{-T}(-T) = D \supseteq U_{-S}(-T)$. Applying once more the fact that the algorithm preserves the inclusion ordering yields $U_{-T}(t) \supseteq U_{-S}(t)$ for any $t \geq -T$.

Finally, suppose that the upper and lower processes meet at some time $s \leq 0$. Then, as they are coupled by the same realizations of the $E(\cdot)$ and the Ξ_i , they proceed as one.

We are now ready to state the main result of this section.

Theorem 1. *Let Y be a realization of a Boolean model of balls of radius r with intensity λ , observed in a compact window $W \subseteq \mathbb{R}^d$, and define \mathcal{C} as in (1). Then Algorithm 1 almost surely terminates in finite time. Its output is distributed according to $\pi_{\lambda|Y}$, the conditional distribution of interior points given Y (cf. Lemma 1).*

Proof. Note that $P(E(0) = D \mid D) = (\lambda/\kappa)^{n(D)}$, where $n(D)$ is the number of points in D . Hence

$$P(E(0) = D) = E \left[\left(\frac{\lambda}{\kappa} \right)^{n(D)} \right] > 0,$$

so that the event $\{E(0) = D\}$ has strictly positive probability. By Lemma 2, $E(-T) = D$ will occur for some T almost surely. Hence, by Lemma 3, the algorithm terminates almost surely in finite time.

To show that the output has the required distribution, consider a process $Z_{-T}(\cdot)$ initialized at any A such that $E(-T) \subseteq A \subseteq D$ which evolves similar to the upper and lower processes, except that a death at time t is implemented only if it respects the coverage condition (3). Thus $Z_{-T}(\cdot)$ has the same stochastic properties as the process $Z(\cdot)$ considered in the beginning of this section, and in particular has stationary distribution $\pi_{\lambda|Y}$ (cf. Lemma 2). Moreover, as Algorithm 1 preserves the inclusion ordering, $L_{-T}(t) \subseteq Z_{-T}(t) \subseteq U_{-T}(t)$ for all $-T \leq t \leq 0$. The claim now follows from Kendall and Møller (2000, Theorem 3.1).

We conclude this section with a practical observation. Generally, Algorithm 1 will take a relatively long time to terminate when sampling from $\pi_{\lambda|Y}$ for small values of λ . We then expect that a κ for which we can reasonably perform rejection sampling to obtain D will be much bigger than λ . In our experience, it is efficient to choose a κ_1 which is not too much smaller than κ and sample from $\pi_{\kappa_1|Y}$. This sample can be used as a dominating pattern to sample from $\pi_{\kappa_2|Y}$ where we choose κ_2 not too much smaller than κ_1 . In this manner we gradually descend until we reach λ .

4. Maximum likelihood estimation

4.1. Monte Carlo approach

A single realization y of a Boolean model Y of balls with radius $r > 0$ (Definition 1) has been observed in a compact window $W \subseteq \mathbb{R}^d$, and the aim is to estimate the intensity λ of the underlying germ process X . Let Po_λ be the probability measure corresponding to the relevant part of the germ process, i.e. the Poisson process of intensity λ on the set $\tilde{W} = W \oplus rB$. Let $f_\lambda = d\text{Po}_\lambda/d\text{Po}_1$. We have

$$f_\lambda(\mathbf{x}) = e^{(1-\lambda)|\tilde{W}|} \lambda^{n(\mathbf{x})}.$$

By Theorem 1 of van Lieshout (1997), the density of our Boolean model with intensity λ with respect to a Boolean model with intensity κ ($\kappa \neq 0$) is

$$\begin{aligned} p_\kappa(y; \lambda) &= e^{(\kappa-\lambda)|\tilde{W}|} \mathbb{E}_\kappa \left[\left(\frac{\lambda}{\kappa} \right)^{n(X)} \mid \mathcal{B}(X) \cap W = y \right] \\ &= e^{(\kappa-\lambda)|\tilde{W}|} \left(\frac{\lambda}{\kappa} \right)^{n(x^b)} \int \left(\frac{\lambda}{\kappa} \right)^{n(x)} d\pi_{\kappa|y}(x). \end{aligned} \tag{7}$$

It can be proved that there exists a maximum likelihood estimator of λ , but proof of its uniqueness has eluded us. Assuming there is a unique MLE, computing it does not seem possible because the integral in the above expression includes an intractable normalizing constant. The idea behind Monte Carlo maximum likelihood (Thompson and Guo (1991), Geyer (1994), (1999)) is to approximate (7) by

$$\hat{p}_\kappa(y; \lambda) = e^{(\kappa-\lambda)|\tilde{W}|} \left(\frac{\lambda}{\kappa} \right)^{n(x^b)} \frac{1}{n} \sum_{i=1}^n \left(\frac{\lambda}{\kappa} \right)^{n(X_i)}, \tag{8}$$

where X_1, \dots, X_n is an independent sample from $\pi_{\kappa|y}$ for some choice of $\kappa > 0$. We can now use (8) to approximate the score function, the maximum likelihood estimator, the Fisher information and anything else we might be interested in.

Let us denote the true MLE by $\hat{\lambda}$, its Monte Carlo approximation by $\hat{\lambda}_n$ and the Monte Carlo approximation to the score function by s_n . Provided the likelihood (7) is concave—as it appears to be for the example in Section 5; cf. Figure 2—the Monte Carlo maximum likelihood estimator is consistent. Moreover, $\sqrt{n}s_n(\hat{\lambda})$ converges in distribution to a normal distribution with mean zero and a certain variance, say, σ^2 . By Geyer (1994, Theorem 7) it follows that

$$\sqrt{n}(\hat{\lambda}_n - \hat{\lambda}) \rightarrow^{\mathcal{D}_\kappa} \mathcal{N}\left(0, \frac{\sigma^2}{I(\hat{\lambda})^2}\right),$$

where $\rightarrow^{\mathcal{D}_\kappa}$ denotes convergence in distribution with respect to the parameter κ . Now (8) is a good approximation to (7) only if λ is not too far from κ . Algorithm 1 allows us to sample from $\pi_{\kappa|y}$ for any κ we like. In practice, we are mostly interested in $p_\kappa(y; \lambda)$ in the vicinity of $\hat{\lambda}$. Therefore, we will use a pilot estimate of λ as our choice for κ .

4.2. Stochastic EM algorithm

It is useful to think of our estimation problem as a missing data problem. The complete data is the germ process X on \tilde{W} and the observed data is $Y = Y(X) = \mathcal{B}(X) \cap W$. Again, suppose that we observe a single realization $Y = y$. The EM algorithm (Dempster *et al.* (1977)) is an

iterative technique to approximately solve likelihood equations for missing data problems. In our case the iteration boils down to

$$\lambda(k + 1) = E_{\lambda(k)} \left[\frac{n(X)}{|\tilde{W}|} \mid \mathcal{B}(X) \cap W = y \right] = \frac{n(\mathbf{x}_y^b) + \int n(\mathbf{x}) d\pi_{\lambda(k)|y}(\mathbf{x})}{|\tilde{W}|}. \tag{9}$$

Dempster *et al.* (1977, p. 7) show that at each step the likelihood under the parameter $\lambda(k)$ of the observed image y increases. Hence, the EM algorithm converges to a (local) maximum of the likelihood function.

Unfortunately, it does not seem possible to compute (9) explicitly. However, samples from the conditional distribution $\pi_{\lambda(k)|y}$ of interior points can be obtained by the method of Section 3. If at the k th iteration step m samples from $\pi_{\lambda(k)|y}$ are available, their average cardinality can be used instead of $\int n(\mathbf{x}) d\pi_{\lambda(k)|y}(\mathbf{x})$. This algorithm is known as the *Monte Carlo EM* (Tanner and Wei (1990)) or, if $m = 1$, the *stochastic EM algorithm* (StEM) (Celeux and Diebolt (1985)). Some large-sample results are available (e.g. Nielsen (2000)) but as we have only one image to work with, those asymptotics do not apply to our situation.

From now on, we specialize to the case $m = 1$. If at each iteration step we use a new sample which is independent of the previous samples, then the iterates of the StEM algorithm form a Markov chain on the parameter space. In principle, this chain must be run until it reaches equilibrium. Here, we propose a modification based on coupling from the past to obtain samples that are guaranteed to be from the equilibrium distribution in a finite number of steps.

As we pointed out earlier, it is useful if the state space of the Markov chain is partly ordered with a minimal and a maximal element. Since our parameter space is the positive reals without such elements, we must restrict it. An appropriate minimal element is $\underline{\lambda} = (|\mathcal{C}|/(\pi r^2) + n(\mathbf{x}_y^b))/|\tilde{W}|$ because any realization from the conditional Boolean model needs at least $|\mathcal{C}|/(\pi r^2)$ interior points to cover \mathcal{C} . Unfortunately, we must choose a maximal element in an *ad hoc* way. We choose $\bar{\lambda}$ so big that we are confident that the true parameter value is less. For technical reasons we choose it such that $\bar{\lambda}|\tilde{W}|$ is an integer. From now on we consider $[\underline{\lambda}, \bar{\lambda}]$ to be our parameter space. Finally, we note that an update of the EM algorithm with the restricted state space is given by the minimum of (9) and $\bar{\lambda}$. This also holds for its stochastic version.

The stochastic EM algorithm is a Markov chain on the finite state space $S = \{\underline{\lambda}, \underline{\lambda} + 1/|\tilde{W}|, \dots, \bar{\lambda}\}$. Now, if $\pi_{\lambda|y}(n(X) = n) > 0$ for some λ , the same holds for any $\lambda > 0$. Hence, the StEM algorithm has a single positive recurrent, aperiodic class.

Our coupling from the past scheme is as follows. Fix an integer $T > 0$. Using the sampling scheme described at the end of Section 3, with the κ_i separated by $1/|\tilde{W}|$, we can obtain nested samples $X_\lambda(t)$ from $\pi_{\lambda|y}$ for all λ in the state space S . Now set $U_{-T}(-T) = \bar{\lambda}$, $L_{-T}(-T) = \underline{\lambda}$ and for $t = -T + 1, -T + 2, \dots, 0$ set $U_{-T}(t) = M(U_{-T}(t - 1), t)$ and $L_{-T}(t) = M(L_{-T}(t - 1), t)$, where $M(\cdot)$ is the modified maximizer

$$M(\lambda, t) = \frac{n(\mathbf{x}_y^b) + n(X_\lambda(t))}{|\tilde{W}|} \wedge \bar{\lambda}.$$

Since for fixed t the $X_\lambda(t)$ are nested, $M(\cdot)$ is an increasing function of λ . Hence, the algorithm is monotone and it suffices to check if $U_{-T}(0) = L_{-T}(0)$. If this is indeed so, then stop. If not, repeat the above procedure starting at time $-2T$, re-using the previous realizations from time $-T$ to 0. To see that the algorithm terminates almost surely, note that with positive probability $X_{\underline{\lambda}}(-T + 1) = X_{\bar{\lambda}}(-T + 1)$, in which case $U_{-T}(\cdot)$ and $L_{-T}(\cdot)$ meet after just one step.

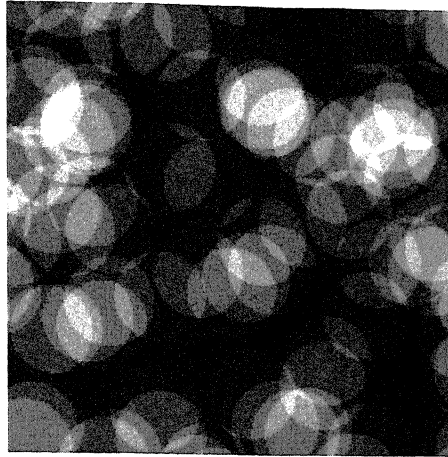


FIGURE 1: Coverage function of a realization of a Poisson process with intensity 75.0 marked by discs of radius 0.1 and clipped in the unit square.

5. An example

We now illustrate the Monte Carlo maximum likelihood and stochastic EM approaches. Figure 1 depicts the coverage function of a realization of a Poisson process with intensity $\lambda = 75.0$ marked by discs of radius $r = 0.1$ and clipped within the unit square $W = [0, 1]^2$. The complete data maximum likelihood estimator is $\hat{\lambda} = 70.14$.

From the picture, we can easily extract the 56 boundary points. For computational convenience we shall use $\tilde{W} = [-r, 1 + r]^2$ instead of $W \oplus rB$.

The Monte Carlo maximum likelihood approach needs a reference parameter κ which is not too far from the maximum likelihood estimator we are trying to approximate. We performed ten steps of the stochastic EM algorithm initialized with $\hat{\lambda} = 40.25$ to obtain $\kappa = 68.06$. To compute the Monte Carlo likelihood (8), 100 independent samples from $\pi_{68.06|\lambda}$ were generated using the modified coupling from the past algorithm described at the end of Section 3 with step size 5.0. To find the dominating pattern D , the first attempt was at intensity 100.0. Fifty rejection sampling steps were performed before increasing the intensity by $10/(1 + 2r)^2$. Usually, it was not necessary to increase the intensity more than once, if at all. The logarithm of the Monte Carlo likelihood we obtained is plotted in Figure 2 for $\lambda \in [50, 100]$. Optimizing over λ yields $\hat{\lambda}_{100} = 68.70$. The Monte Carlo variance is 0.23; cf. Section 4. For comparison, the estimated inverse Fisher information at $\hat{\lambda}$ is 64.60.

Turning to the perfect stochastic EM algorithm, Figure 3 depicts the upper and lower processes of one run. In this case, the two paths coalesce after taking $T = 2, 4$ and 8. As a crude indication of the variability, the sample path over a further 100 forward steps is shown as well. We repeated the process of Figure 3 twenty-five times. The sample mean and variance of the 25 independent samples at time 0 were 69.14 and 19.01, respectively. The variance can be reduced by averaging over a few further steps of the EM algorithm. If 10 such steps are used, the sample mean and variance were 68.09 and 2.82. Using 25 steps after time 0 yielded 68.11 and 1.40.

Comparing the two approaches, Monte Carlo maximum likelihood estimation is overwhelmingly more fast than the perfect stochastic EM algorithm. The latter is particularly slow, because

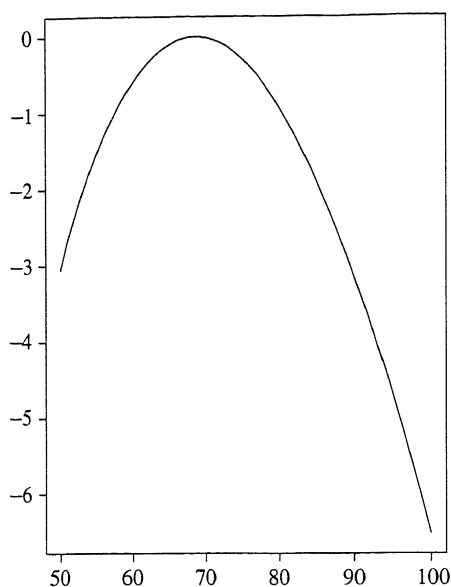


FIGURE 2: Estimated log likelihood with respect to the Boolean model with intensity $\kappa = 68.06$ as a function of λ for the data in Figure 1.

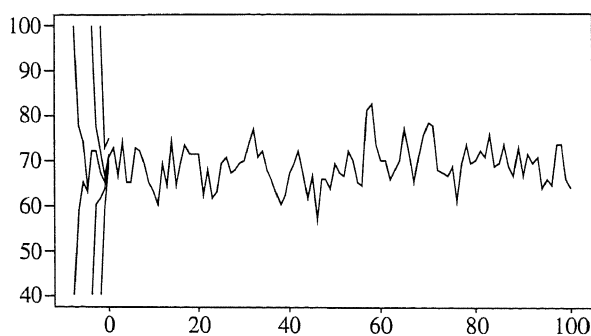


FIGURE 3: The upper and lower processes in the exact stochastic EM algorithm.

each run requires one or more samples from $\pi_{\lambda|y}$. The second major advantage of the Monte Carlo approach is that it produces an estimate of the full likelihood. This immediately translates to estimates of moments of functionals, probabilities of certain events, likelihood ratio statistics, the Fisher information as well as the standard error with respect to the true maximum likelihood estimator.

6. Discussion and loose ends

In the Boolean model of discs, no interaction is present between the underlying germs. This assumption can be relaxed. For instance, by Thönnies (1998, Theorem 3.5), the conditional

distribution of the interior part of a standard area-interaction process (Baddeley and van Lieshout (1995)) given its union set is that of a Poisson process. Hence, the theory of the present paper applies. The same is true for any germ grain process whose density with respect to a unit rate Poisson process is of the form

$$\alpha(\lambda, \theta) \lambda^{n(\mathbf{x})} g_{\theta}(\mathcal{B}(\mathbf{x}) \cap W),$$

where θ is a parameter vector.

Another generalization is to allow grains that have the property that the location of their centre is identified by any part of their boundary. Examples of such grains are ellipses or discs with random radii. The assumption ensures that the distribution of the boundary points is degenerate. If this distribution is not degenerate (for instance when the grains are unit squares) it appears to be very complicated.

Finally, recall that the stochastic EM algorithm did not compare favourably to the Monte Carlo approach (cf. end of Section 5). In a recent paper, Delyon *et al.* (1999) overcame these problems by combining the EM algorithm with stochastic approximation (Penttinen (1984), Younes (1988), Moyeed and Baddeley (1991)). Suppose we have observed a single realization $Y = y$. If the current approximation to $E_{\lambda(k)}[\log f_{\lambda}(X) | y]$ is denoted by $\hat{Q}_k(\lambda)$, Delyon *et al.* generate a sample \mathbf{x}_{k+1} from $\pi_{\lambda(k)|y}$ and set

$$\hat{Q}_{k+1}(\lambda) = \hat{Q}_k(\lambda) + \gamma_{k+1}[\log f_{\lambda}(\mathbf{x}_y^b \cup \mathbf{x}_{k+1}) - \hat{Q}_k(\lambda)],$$

where $\{\gamma_k\}_{k \geq 1}$ is a sequence of positive step sizes. The function $\hat{Q}_{k+1}(\lambda)$ is optimized over $\lambda > 0$ to yield $\lambda(k+1)$. In contrast to the stochastic EM algorithm, all previous samples contribute to $\hat{Q}_{k+1}(\lambda)$. This results in a more efficient use of the simulations at the expense of losing the Markov property. The method does allow estimation of some functionals of interest, such as the Fisher information $I(\hat{\lambda})$ at $\hat{\lambda}$. Moreover, by Delyon *et al.* (1999, Theorem 4), under conditions on the speed of the discount factors γ_k ,

$$\sqrt{n}(\bar{\lambda}_n - \hat{\lambda}) \rightarrow \mathcal{N}\left(0, \frac{\sigma^2}{I(\hat{\lambda})^2}\right),$$

where $\bar{\lambda}_n$ is the mean intensity over the first n iterations and σ^2 is the asymptotic variance of the mean score. Hence, the computational load of the stochastic approximation EM algorithm is comparable to that of the Monte Carlo maximum likelihood approach. However, because the latter method does not involve careful tuning of discount factors, we have preferred it over the stochastic approximation EM algorithm.

Acknowledgements

The authors are grateful to Richard Gill, Andries Lenstra and Ronald Meester for helpful comments, and to Adri Steenbeek for programming assistance. This work was carried out when the second author was at the Mathematical Institute of the University of Utrecht and during a visit of both authors to the Department of Mathematics of the University of Western Australia. It is a pleasure to acknowledge the hospitality and financial support of these departments. Van Zwet also gratefully acknowledges an NWO travel grant SIR 13-4457. Van Lieshout's research was funded by NWO grant CIMS 613.003.100.

References

- AYALA, G., FERRÁNDIZ, J. AND MONTES, F. (1990). Boolean models: maximum likelihood estimation from circular clumps. *Biometrical J.* **32**, 73–78.
- BADDELEY, A. J. AND GILL, R. D. (1997). Kaplan–Meier estimators for interpoint distance distributions of spatial point processes. *Ann. Statist.* **25**, 263–292.
- BADDELEY, A. J. AND VAN LIESHOUT, M. N. M. (1995). Area-interaction point processes. *Ann. Inst. Statist. Math.* **47**, 601–619.
- CELEUX, G. AND DIEBOLT, Y. (1985). The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Comput. Statist. Quart.* **2**, 73–82.
- DELYON, B., LAVIELLE, M. AND MOULINES, E. (1999). Convergence of a stochastic approximation version of the EM algorithm. *Ann. Statist.* **27**, 94–128.
- DEMPSTER, A. P., LAIRD, N. M. AND RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. B* **39**, 2–22.
- DUPAČ, V. (1980). Parameter estimation in the Poisson field of discs. *Biometrika* **67**, 187–190.
- FELLER, W. (1968). *An Introduction to Probability Theory and its Applications*, Vol. 2. John Wiley, New York.
- GEYER, C. J. (1994). On the convergence of Monte Carlo maximum likelihood calculations. *J. R. Statist. Soc. B* **56**, 261–274.
- GEYER, C. J. (1999). Likelihood inference for spatial point processes. In *Proc. Sem. Eur. Statist., Stoch. Geometry, Likelihood, Comput.*, eds O. Barndorff-Nielsen, W. S. Kendall and M. N. M. van Lieshout. CRC Press/Chapman and Hall, London.
- GILKS, W. R., RICHARDSON, S. AND SPIEGELHALTER, D. J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.
- HÄGGSTRÖM, O. AND NELANDER, K. (1998). Exact sampling from anti-monotone systems. *Statist. Neerlandica* **52**, 360–380.
- HALL, P. (1988). *Introduction to the Theory of Coverage Processes*. John Wiley, New York.
- HANISCH, K. H. (1984). Some remarks on estimators of the distribution function of nearest neighbour distance in stationary spatial point processes. *Math. Operationsforsch. Statist. Ser. Statist.* **15**, 409–412.
- KENDALL, W. S. (1997). On some weighted Boolean models. In *Proc. Int. Symp. Adv. Theory Appl. Random Sets*, ed. D. Jeulin. World Scientific Publishing, Singapore, pp. 105–120.
- KENDALL, W. S. AND MØLLER, J. (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Adv. Appl. Prob.* **32**, 844–865.
- KENDALL, W. S. AND THÖNNES, E. (1999). Perfect simulation in stochastic geometry. *Pattern Recognition* **32**, 1569–1586.
- KOLMOGOROV, A. N. (1937). On the statistical theory of metal crystallisation. *Izv. Acad. Sci. USSR Ser. Math.* **3**, 355–360.
- LANTUÉJOU, C. (1997). Conditional simulation of object-based models. In *Proc. Int. Symp. Adv. Theory Appl. Random Sets*, ed. D. Jeulin. World Scientific Publishing, Singapore, pp. 271–288.
- MATÉRN, B. (1960). *Spatial Variation* (Lecture Notes Statist. **36**). Springer, Berlin.
- MATHERON, G. (1975). *Random Sets and Integral Geometry*. John Wiley, Chichester.
- MOLCHANOV, I. S. (1997). *Statistics of the Boolean Model for Practitioners and Mathematicians*. John Wiley, Chichester.
- MOLCHANOV, I. S. AND STOYAN, D. (1994). Asymptotic properties of estimators for parameters of the Boolean model. *Adv. Appl. Prob.* **26**, 301–323.
- MOYEED, R. A. AND BADDELEY, A. J. (1991). Stochastic approximation of the MLE for a spatial point pattern. *Scand. J. Statist.* **18**, 39–50.
- NIELSEN, S. F. (2000). The stochastic EM algorithm: estimation and asymptotic results. *Bernoulli* **6**, 457–489.
- PARZEN, E. (1962). *Stochastic Processes*. Holden-Day, San Francisco.
- PENTTINEN, A. (1984). *Modelling Interaction in Spatial Point Patterns: Parameter Estimation by the Maximum Likelihood Method* (Jyväskylä Studies Comput. Sci., Econ., Statist. **7**). University of Jyväskylä.
- PRESTON, C. J. (1977). Spatial birth-and-death processes. *Bull. Int. Statist. Inst.* **46**, 371–391.
- PROPP, J. G. AND WILSON, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures Algorithms* **9**, 223–252.
- RIPLEY, B. D. (1988). *Statistical Inference for Spatial Processes*. Cambridge University Press.
- SCHMITT, M. (1991). Estimation of the density in a stationary Boolean model. *J. Appl. Prob.* **28**, 702–709.
- SERRA, J. (1982). *Image Analysis and Mathematical Morphology*. Academic Press, London.
- STOYAN, D., KENDALL, W. S. AND MECKE, J. (1995). *Stochastic Geometry and its Applications*, 2nd edn. John Wiley, Chichester.
- TANNER, M. A. AND WEI, G. C. G. (1990). A Monte Carlo implementation of the EM-algorithm and the poor man's data augmentation algorithms. *J. Amer. Statist. Assoc.* **85**, 699–704.

- THOMPSON, E. A. AND GUO, S. W. (1991). Evaluation of likelihood ratios for complex genetic models. *IMA J. Math. Appl. Med. Biol.* **8**, 149–169.
- THÖNNES, E. (1998). Perfect and imperfect simulations in stochastic geometry. Doctoral Thesis, University of Warwick.
- THÖNNES, E. (2000). The conditional Boolean model revisited. Tech. Rept 369, Department of Statistics, University of Warwick.
- VAN LIESHOUT, M. N. M. (1997). On likelihoods for Markov random sets and Boolean models. In *Proc. Int. Symp. Adv. Theory Appl. Random Sets*, ed. D. Jeulin. World Scientific Publishing, Singapore, pp. 121–135.
- WEIL, W. (1988). Expectation formulas and isoperimetric properties for non-isotropic Boolean models. *J. Microscopy* **151**, 235–245.
- YOUNES, L. (1988). Estimation and annealing for Gibbsian fields. *Ann. Inst. H. Poincaré Prob. Statist.* **24**, 269–294.