Centrum voor Wiskunde en Informatica

REPORT*RAPPORT*

PNA

Probability, Networks and Algorithms

Probability, Networks and Algorithms

Maximum likelihood estimation for random sequential adsorption

M.N.M. van Lieshout

REPORT PNA-R0503 JULY 2005

CWI is the National Research Institute for Mathematics and Computer Science. It is sponsored by the Netherlands Organization for Scientific Research (NWO).
CWI is a founding member of ERCIM, the European Research Consortium for Informatics and Mathematics.

CWI's research has a theme-oriented structure and is grouped into four clusters. Listed below are the names of the clusters and in parentheses their acronyms.

Probability, Networks and Algorithms (PNA)

Software Engineering (SEN)

Modelling, Analysis and Simulation (MAS)

Information Systems (INS)

P.O. Box 94079, 1090 GB Amsterdam (NL)
Kruislaan 413, 1098 SJ Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

ISSN 1386-3711

Maximum likelihood estimation for random sequential adsorption

ABSTRACT
Existence and uniqueness of a maximum likelihood estimator for the time and range parameters in random sequential adsorption models is established. Nuisance parameters of the reference distribution are estimated by means of profile likelihoods. The approach is illustrated by examples.

2000 Mathematics Subject Classification: 60G55, 60D05, 62M30.
Keywords and Phrases: maximum likelihood estimation, random sequential adsorption, sequential spatial process, simple sequential inhibition.
Note: This research is supported by the Technology Foundation STW, applied science division of NWO, and the technology programme of the Ministry of Economic Affairs (project CWI.6156 `Markov sequential point processes for image analysis and statistical physics').

Maximum Likelihood Estimation for Random Sequential Adsorption

M.N.M. van Lieshout

CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

ABSTRACT
Existence and uniqueness of a maximum likelihood estimator for the time and range parameters in random sequential adsorption models is established. Nuisance parameters of the reference distribution are estimated by means of profile likelihoods. The approach is illustrated by examples.

2000 Mathematics Subject Classification: 60G55, 60D05, 62M30.

Keywords and Phrases: maximum likelihood estimation, random sequential adsorption, sequential spatial process, simple sequential inhibition.

Note: This research is supported by the Technology Foundation STW, applied science division of NWO, and the technology programme of the Ministry of Economic Affairs (project CWI.6156 'Markov sequential point processes for image analysis and statistical physics').

1. Introduction

Packing problems occur in many physical and biological processes. Hard 'particles' such as cars, monomers, or proteins, arrive in some bounded Euclidean window according to a constant rate Poisson process, and select a position for themselves according to some fixed distribution. If there would be no overlap with an established particle, the new one attaches itself to the selected location, otherwise it leaves the system. The process continues either for a given time, until a given number of particles have been adsorbed, or until there is no room for another particle to establish itself without intersecting existing ones.

Complete random packings are those in which the birth process continues until saturation, that is, until no room is left in the window to accommodate another particle. On the line, a rigorous analysis dates back to [14]. Typical results concern the asymptotic coverage fraction of unit length segments in an interval of increasing length, or equivalently the relative vacancy, that is, the length fraction not covered by particles, and central limit theorems for the number of adsorbed particles [6]. The problem is sometimes known as the car parking problem, with the segments representing cars, the interval the kerb. In higher dimensions, analysis is rather harder. Simulation results on the coverage fraction are widespread. For example, simulated complete random packings of discs in two dimensions can be found in [12, 18]. Efficient simulation in $d \geq 2$ dimensions is studied in [5]. From a theoretical point of view, a law of large numbers for the coverage fraction is proved in [13].

Other authors have considered runs of the process over a finite time horizon, or those

that terminate when a certain packing density (i.e. number of adsorbed particles) has been achieved. Alternatively, the number of adsorbed particles could be stochastic, and follow some probability mass function. Such models were dubbed simple sequential inhibition processes in the spatial statistics literature [4], and are known as random sequential adsorption (RSA) amongst physicists [7] to describe the deposition of colloidal particles onto a substrate. For finite time horizons, interest focusses on the evolution of summary statistics, for example mean and variance of the number of particles, the pair correlation function, or the empty space distribution, and especially on the behaviour of the coverage fraction as a function of time, the so-called kinetic law (cf. the Feder conjecture [8]), mostly using series expansions and Monte Carlo techniques.

The reader is referred to [10, Ch. 1.10], to the review papers [7, 16, 17] or the textbooks [1, 3, 19] for overviews and pointers to the literature. For a flavour of current work in statistical physics, we refer to the special issue of *Colloids and Surfaces* (number 165, 2000).

To date, most attention has focussed on probabilistic work rather than inference. In this paper, we take the statistical viewpoint. We place the RSA model in the context of Markov sequential processes [11], derive its Radon–Nikodym derivative, and consider maximum likelihood estimation of its parameters. Note that maximum likelihood is especially convenient in exponential family models, but that RSA exhibits a highly non-linear dependence [11].

The plan of this paper is as follows. In Section 2, we fix notation. In Section 3, some facts from the theory of convolution of exponential distributions are recalled. The next section presents the main results. Section 5 demonstrates the estimation procedure in practice, and the paper concludes with a brief summary.

2. Definitions and notation

The *random sequential adsorption model* [7], also known as *simple sequential inhibition* [3, 4] in some bounded domain of $\mathbb{R}^d$ may be defined as follows. A unit rate stream of particles arrives in the domain. If the arrival location is within a distance $r > 0$ of some existing particle, the particle leaves, otherwise it is adsorbed. The process continues for some fixed time $\theta \geq 0$. Note that in the limit ($\theta \to \infty$), the adsorption process is saturated (jammed), that is, particles are adsorbed until there are no open spaces left with radius at least r.

In a recent paper, we showed that a simple sequential inhibition model can be placed in the context of *finite sequential spatial processes* [11]. Realisations of such a process consist of a finite sequence

$$\vec{\mathbf{x}} = (x_1, \ldots, x_n), \qquad n \in \mathbb{N}_0$$

of points in some bounded open subset D of the $\mathbb{R}^d$ with non-empty interior. The family of all such configurations is denoted by N^{f}. The notation $n(\vec{\mathbf{x}})$ shall be used for the length of $\vec{\mathbf{x}}$, and $\vec{\mathbf{x}}_{<i}$ for the subsequence $(x_1, \ldots, x_{i-1})$, $i = 1, \ldots n+1$. Note that for $i = 1$, $\vec{\mathbf{x}}_{<i}$ has length 0.

Suppose that D is equipped with a metric $\rho(\cdot, \cdot)$. The metric defines a topology and a Borel σ-algebra, denoted by $\mathcal{B}$. Write $\rho(z, A) := \inf\{\rho(z, a) : a \in A\}$ for the distance between $z \in D$ and $A \in \mathcal{B}$. The space N^{f} is equipped with the σ-algebra generated by the Borel product σ-algebras. The notation $\mu(A)$ is used for the Lebesgue measure of A.

Definition 1. *The random sequential adsorption model is a jump process, starting from an empty vector, with transition rate*

$$b(z, \vec{\mathbf{x}}) = \mathbf{1}\{\rho(z, \vec{\mathbf{x}}) > r\}\, d\mu(z)$$

for a birth jump from $\vec{\mathbf{x}}$ *to* $(\vec{\mathbf{x}}, z)$. *Here* $r \geq 0$ *is the hard core distance,* $\theta \geq 0$ *the time horizon.*

A variation that allows for spatial heterogeneity is to set

$$b(z, \vec{\mathbf{x}}) = \pi(z)\, \mathbf{1}\{\rho(z, \vec{\mathbf{x}}) > r\}\, dz$$

where $\pi(\cdot)$ is a strictly positive probability density on D. Thus, regions with high π-mass are favoured over those with smaller mass. Realisations tend to have particles with a low index concentrated in high density regions, whereas latecomers have to contend themselves with locations with smaller π-mass. In the classic case of Definition 1, such a competition advantage is not present.

Interaction structures other than hard core repulsion can be modelled by suitable choices of $b(z, \vec{\mathbf{x}})$, see e.g. [7]. Such models are known as *cooperative sequential adsorption*, and can allow for both attraction and repulsion of particles by previously arrived ones.

Example 1. Consider the special case $r = 0$. Then particles arrive according to a unit rate Poisson process, that is, $b(\cdot, \cdot) \equiv 1$ a.e. with respect to Lebesgue measure. The number of particles adsorbed in D up to time θ is Poisson distributed with parameter $\theta\mu(D)$. In other words, the probability of exactly n transitions in $[0, \theta]$ is given by

$$q_n = \frac{e^{-\theta\mu(D)}}{n!}\,(\theta\mu(D))^n$$

for $n = 0, 1, 2, \ldots$, and, given n particles arrived up to time θ, their joint probability distribution has density $p_n \equiv 1/\mu(D)^n$ with respect to μ^n. Since p_n does not depend on θ, the maximum likelihood estimator is readily derived from

$$\frac{\partial}{\partial\theta} \log q_n = -\mu(D) + \frac{n}{\theta}$$

which is 0 if and only if $\hat{\theta} = n/\mu(D)$. A little closer examination reveals $\hat{\theta}$ is indeed the unique optimiser.

In order to be able to study maximum likelihood estimation when the hard core distance is unknown, some results on the convolution of exponential distributions are required, as outlined in the next section.

3. Convolution of exponential distributions

Recall [9] that if X and Y are two independent random variables with values in $\mathbb{R}^+$, absolutely continuous with densities $f_X(\cdot)$ and $f_Y(\cdot)$ respectively, for $t \geq 0$, a density $f_{X+Y}(\cdot)$ of the sum $X + Y$ is given by

$$f_{X+Y}(t) = \int_0^t f_X(s)\, f_Y(t - s)\, ds.$$

Similarly, the cumulative distribution function satisfies

$$F_{X+Y}(t) = \int_0^t f_X(s)\, F_Y(t-s)\, ds.$$

Higher order convolutions can be obtained analogously.

Lemma 1. *Let, for $n \geq 1$, $X_1, \ldots, X_n$ be independent exponentially distributed random variables with distinct rates $\lambda_1, \ldots, \lambda_n$. Then the sum $X_1 + \cdots + X_n$ has probability density*

$$(-1)^{n+1} \left\{ \prod_{i=1}^n \lambda_i \right\} \sum_{i=1}^n \frac{e^{-\lambda_i t}}{\prod_{j \neq i}(\lambda_i - \lambda_j)} = (-1)^{n+1} \sum_{i=1}^n \left\{ \lambda_i\, e^{-\lambda_i t} \prod_{j \neq i} \frac{\lambda_j}{\lambda_i - \lambda_j} \right\}, \quad t \geq 0.$$

The result is presented as an exercise in [9]. Here we give a proof for the sakes of completeness and correction of a misprint in formula (*) of [9, p.40, I.13].

Proof: Proceed by induction. It is readily verified that the formula holds for $f_{X_1+\cdots+X_n}(t)$ if $n = 1$ or 2. Suppose then Lemma 1 is true up to some $n \geq 2$ and consider $X_1 + \cdots + X_{n+1}$. By the convolution formula, a density is given by

$$\int_0^t (-1)^{n+1} \sum_{i=1}^n \left[\left\{ \prod_{n \geq j \neq i} \frac{\lambda_j}{\lambda_i - \lambda_j} \right\} \lambda_i e^{-\lambda_i s}\, \lambda_{n+1} e^{-\lambda_{n+1}(t-s)} \right] ds =$$

$$(-1)^{n+1} \sum_{i=1}^n \left[\prod_{n \geq j \neq i} \frac{\lambda_j}{\lambda_i - \lambda_j} \int_0^t \lambda_i e^{-\lambda_i s}\, \lambda_{n+1} e^{-\lambda_{n+1}(t-s)}\, ds \right].$$

Use the convolution formula for $n = 2$ inside the sum to obtain

$$(-1)^{n+1} \sum_{i=1}^n \left[\prod_{n \geq j \neq i} \frac{\lambda_j}{\lambda_i - \lambda_j} \left\{ \frac{-\lambda_i \lambda_{n+1}}{\lambda_i - \lambda_{n+1}} e^{-\lambda_i t} + \frac{-\lambda_i \lambda_{n+1}}{\lambda_{n+1} - \lambda_i} e^{-\lambda_{n+1} t} \right\} \right] =$$

$$(-1)^{n+2} \sum_{i=1}^n \left\{ \lambda_i\, e^{-\lambda_i t} \prod_{n+1 \geq j \neq i} \frac{\lambda_j}{\lambda_i - \lambda_j} \right\} +$$

$$+ (-1)^{n+2} e^{-\lambda_{n+1} t} \left\{ \prod_{i=1}^{n+1} \lambda_i \right\} \left\{ \sum_{j=1}^n \frac{-1}{\prod_{n+1 \geq k \neq j}(\lambda_j - \lambda_k)} \right\}.$$

By a symmetry argument (interchange X_1 and X_{n+1}), the sum in the last term after $e^{-\lambda_{n+1} t}$ is equal to $\prod_{k \neq n+1}(\lambda_{n+1} - \lambda_k)^{-1}$, and the proof is completed. □

4. Maximum likelihood estimation

The random sequential adsorption model has two parameters: r and θ. The purpose of this section is to study maximum likelihood estimation. In order to do so, we derive a Radon–Nikodym derivative (Theorem 1) with respect to the distribution of a sequence of Poisson length (mean $\mu(D)$) with independently, μ-uniformly distributed components, thus placing RSA into the framework of Markov sequential spatial processes [11]. We proceed to establish existence and uniqueness of

$$(\hat{r}, \hat{\theta}) := \text{ argsup } \{f(\vec{\mathbf{x}}; r, \theta) : r, \theta \geq 0\}$$

cf. Proposition 1 and Theorem 2.

Note that the Radon–Nikodym derivative $f(\cdot; \cdot, \cdot)$ can be expressed as

$$f(\vec{\mathbf{x}}; r, \theta) = e^{\mu(D)}\, n(\vec{\mathbf{x}})!\, j_n(\vec{\mathbf{x}}; \theta)$$

where $j_n(\cdot; r, \theta)\, d\mu^n$ is a Janossy density [2, 11] that may be interpreted as the probability of a sequence of exactly n points, successively at infinitesimal regions centered at x_1 up to x_n. Clearly, $j_n(\cdot; r, \theta)$ is sufficient for (r, θ). The next theorem gives an explicit expression of this Janossy density in terms of the birth rates.

Theorem 1. *Consider the model of Definition 1, and write*

$$B(\vec{\mathbf{x}}) = B(\vec{\mathbf{x}}; r) := \int_D b(z, \vec{\mathbf{x}})\, d\mu(z)$$

for the total birth rate, suppressing the dependence of $b(\cdot, \cdot)$ on r for notational convenience. The sufficient statistic $j_n(\vec{\mathbf{x}}; \theta)$ is zero if $B(\vec{\mathbf{x}}_{<n}) = 0$. If $B(\vec{\mathbf{x}}) > 0$, it can be written as

$$j_n(\vec{\mathbf{x}}; \theta) = \frac{g_{n+1}(x_1, \ldots, x_n; \theta)}{B(x_1, \ldots, x_n)} \prod_{i=1}^{n} \frac{b(x_i, \vec{\mathbf{x}}_{<i})}{B(\vec{\mathbf{x}}_{<i})}$$

for $n \geq 1$. Here $g_{n+1}(x_1, \ldots, x_n; \cdot)$ is the probability density (Lemma 1) of the sum of $n+1$ independent random variables that are exponentially distributed with rate $B(\vec{\mathbf{x}}_{<i})$, $i = 1, \ldots, n+1$. If $B(\vec{\mathbf{x}}_{<n}) > 0 = B(\vec{\mathbf{x}})$,

$$j_n(\vec{\mathbf{x}}; \theta) = G_n(x_1, \ldots, x_{n-1}; \theta) \prod_{i=1}^{n} \frac{b(x_i, \vec{\mathbf{x}}_{<i})}{B(\vec{\mathbf{x}}_{<i})},$$

where $G_n(x_1, \ldots, x_{n-1}; \cdot)$ is the cumulative probability distribution function of the sum of n independent random variables that are exponentially distributed with rate $B(\vec{\mathbf{x}}_{<i})$, $i = 1, \ldots, n$. For $n = 0$,

$$j_0(\emptyset; \theta) = \exp\left[-B(\emptyset)\, \theta\right].$$

Proof: Note that the total birth rates $B(\vec{\mathbf{x}}_{<i})$ are strictly decreasing in $i = 1, \ldots, n$ under the assumption that $B(\vec{\mathbf{x}}_{<n}) > 0$. Hence, under the latter assumption, all $B(\vec{\mathbf{x}}_{<i})$, $i < n$, are strictly positive, and the ratios in the posed Janossy densities well defined.

Since $j_0(\emptyset;\theta) = q_0(\theta)$, the probability that no transition occurs up to time θ, the formula for $n = 0$ holds true.

From now on, let $n \geq 1$. If $B(\vec{\mathbf{x}}_{<n}) = 0$, since $D \setminus \cup_{i<n}\{z : \rho(z, x_i) \leq r\}$ is open, the set is either empty or contains an open ball of positive radius. The latter cannot be the case, as then $B(\vec{\mathbf{x}}_{<n})$ would exceed the area of the open ball and hence be strictly positive. In the first case, $\rho(x_n, \vec{\mathbf{x}}_{<n}) \leq r$, so that the configuration $\vec{\mathbf{x}}$ almost surely will not occur and one may set its Janossy density to zero.

Hence assume $B(\vec{\mathbf{x}}_{<n}) > 0$ and consider the probability of exactly n transitions in $[0,\theta]$, the first jump to the infinitesimal region $d\mu(x_1)$ around x_1, the second to $d\mu(x_2)$, up to the last jump involving $d\mu(x_n)$. Then the first exponentially distributed waiting time (with rate $B(\emptyset)$) has to be less than θ, the first point has to be selected at x_1, all subsequent cumulative waiting times must fall before θ until the last point x_n is selected, and no further transition may take place within the time horizon θ. Integration with respect to the waiting times yields

$$\int_0^\theta B(\emptyset)e^{-s_1 B(\emptyset)} ds_1 \frac{b(x_1,\emptyset)}{B(\emptyset)} d\mu(x_1) \int_0^{\theta-s_1} B(x_1)e^{-s_2 B(x_1)} ds_2 \frac{b(x_2,x_1)}{B(x_1)} d\mu(x_2) \times \cdots$$

$$\cdots \times \int_0^{\theta-(s_1+\cdots+s_{n-1})} B(\vec{\mathbf{x}}_{<n})e^{-s_n B(\vec{\mathbf{x}}_{<n})} \frac{b(x_n,\vec{\mathbf{x}}_{<n})}{B(\vec{\mathbf{x}}_{<n})} e^{-(\theta-\sum_{i=1}^n s_i)B(\vec{\mathbf{x}}_{\leq n})} ds_n\, d\mu(x_n) =$$

$$\prod_{i=1}^n \frac{b(x_i,\vec{\mathbf{x}}_{<i})}{B(\vec{\mathbf{x}}_{<i})} \int_0^\theta \int_0^{\theta-s_1} \cdots \int_0^{\theta-\sum_{i=1}^{n-1} s_i} \prod_{i=1}^n \left\{ B(\vec{\mathbf{x}}_{<i})e^{-s_i B(\vec{\mathbf{x}}_{<i})} \right\} e^{-(\theta-\sum_{i=1}^n s_i)B(\vec{\mathbf{x}}_{\leq n})}$$

$$ds_1 \ldots ds_n\, d\mu(x_1) \ldots d\mu(x_n) =$$

$$h_n(\vec{\mathbf{x}}_{\leq n}) \int_0^\theta \int_0^{\theta-s_1} \cdots \int_0^{\theta-\sum_{i=1}^{n-1} s_i} \prod_{i=1}^n f_{B(\vec{\mathbf{x}}_{<i})}(s_i) \left(1 - F_{B(\vec{\mathbf{x}}_{\leq n})}(\theta - \sum_{i=1}^n s_i)\right)$$

$$ds_1 \ldots ds_n\, d\mu(x_1) \ldots d\mu(x_n) \qquad (4.1)$$

with the notation $f_\lambda(\cdot)$ for the density of an exponential distribution of rate λ, $F_\lambda(\cdot)$ for the corresponding cumulative distribution function, and

$$h_n(\vec{\mathbf{x}}_{\leq n}) = \prod_{i=1}^n \frac{b(x_i,\vec{\mathbf{x}}_{<i})}{B(\vec{\mathbf{x}}_{<i})}.$$

First consider the case that $B(\vec{\mathbf{x}}) > 0$. To prove that the inner integrand in (4.1) is equal to $g_{n+1}(x_1,\ldots,x_n;\theta)/B(x_1,\ldots,x_n)$, use induction. For $n = 1$

$$\int_0^\theta f_{B(\emptyset)}(s)(1-F_{B(x)}(\theta-s))\, ds = \int_0^\theta f_{B(\emptyset)}(s) f_{B(x)}(\theta-s) \frac{(1-F_{B(x)}(\theta-s))}{f_{B(x)}(\theta-s)} ds = \frac{g_2(x;\theta)}{B(x)}$$

as an exponential distribution has constant hazard rate equal to its rate. Next, suppose (4.1) holds for n and consider $n+1$. Then

$$\int_0^\theta f_{B(\emptyset)}(s_1)\, ds_1 \int_0^{\theta-s_1} \cdots \int_0^{\theta-s_1-\sum_{i=2}^n s_i} \left\{ \prod_{i=2}^{n+1} f_{B(\vec{\mathbf{x}}_{<i})}(s_i) \right\} \left(1 - F_{B(\vec{\mathbf{x}}_{\leq n+1})}(\theta - s_1 - \sum_{i=2}^{n+1} s_i)\right)$$

$$ds_2 \dots ds_{n+1} =$$

$$\int_0^\theta f_{B(\emptyset)}(s_1) \frac{g_{n+1}(x_2, \dots, x_{n+1}; \theta - s_1)}{B(\vec{\mathbf{x}}_{\le n+1})} ds_1 = \frac{g_{n+2}(x_1, \dots, x_{n+1}; \theta)}{B(\vec{\mathbf{x}}_{n+1})}$$

by the induction assumption.

To conclude the proof, note that if $B(\vec{\mathbf{x}}) = 0$, the inner integral of (4.1) is equal to $G_n(x_1, \dots, x_{n-1}; \theta)$ as the survival probability $1 - F_0(t) = 1$ for $0 \le t < \infty$. □

Next, turn to statistical inference.

Example 2. Upon observation of the empty set,

$$f(\emptyset; r, \theta) = \exp\left[(1 - \theta)\, \mu(D)\right]$$

is strictly decreasing in θ and does not depend on r. Hence, we cannot carry out inference on the interaction distance, whereas the maximum likelihood estimator of the time horizon is $\hat{\theta}(\emptyset) = 0$.

As for the classic hard core point process, r may be estimated by the minimum inter-point distance [15].

Proposition 1. *Suppose the pattern $\vec{\mathbf{x}} = (x_1, \dots, x_n)$, $n \ge 1$, is observed. Then $\hat{r}(\vec{\mathbf{x}}) = r_{\min}(\vec{\mathbf{x}}) := \min_{i<j} \rho(x_i, x_j)$ is a maximum likelihood estimator for the hard core distance for any value of $\theta \ge 0$.*

The argsup is not attained (i.e. it is not an argmax), as indicated by the proof.

Proof: Note that $j_n(\vec{\mathbf{x}}; r, \theta) = 0$ unless $\rho(x_i, \vec{\mathbf{x}}_{<i}) > r$ for all $i = 1, \dots, n$ (which implies $B(\vec{\mathbf{x}}_{<n}; r) > 0$). Therefore, $\hat{r}(\vec{\mathbf{x}}) \le r_{\min}(\vec{\mathbf{x}})$.

If $r_1 < r_2 < r_{\min}(\vec{\mathbf{x}})$, then $b(x_i, \vec{\mathbf{x}}_{<i}; r_1) = b(x_i, \vec{\mathbf{x}}_{<i}; r_2)$, while $B(\mathbf{x}_{<i}; r_1) > B(\mathbf{x}_{<i}; r_2)$ for all $i = 1, \dots, n+1$. Hence the function

$$I(s_1, \dots, s_n, \theta; r) := \exp\left[-\sum_{i=1}^n s_i B(\vec{\mathbf{x}}_{<i}; r) - \left(\theta - \sum_{i=1}^n s_i\right) B(\vec{\mathbf{x}}; r)\right]$$

that is the crucial ingredient of the proof of Theorem 1 is increasing for $r < r_{\min}(\vec{\mathbf{x}})$, hence $I(s_1, \dots, s_n, \theta; r_1) < I(s_1, \dots, s_n, \theta; r_2)$ which in turn implies the maximum likelihood estimator is as stated. □

It remains to estimate the time horizon θ.

Example 3. As we saw in Example 2, for $n = 0$ and fixed r, $\hat{\theta} = 0$. For $n \ge 1$, consider the case where for given r a sequence $\vec{\mathbf{x}}$ of length n is observed such that $j_n(\vec{\mathbf{x}}; \theta)$ is strictly positive, but $B(\vec{\mathbf{x}}) = 0$. To optimise $j_n(\vec{\mathbf{x}}; \theta)$ over θ, by Theorem 1, one has to optimise $G_n(x_1, \dots, x_{n-1}; \theta)$, which, being a cumulative distribution function, is monotonically increasing in θ. Furthermore, by the convolution formula, it is easily seen that $G_n(x_1, \dots, x_{n-1}; t) < 1$ for any finite $t \ge 0$. Hence, $\hat{\theta}(\vec{\mathbf{x}}) = \infty$.

Lemma 2. *For fixed r, suppose a pattern $\vec{\mathbf{x}} = (x_1, \ldots, x_n)$, $n \geq 1$, is observed for which $B(\vec{\mathbf{x}}) > 0$. Then the score function for θ is given by*

$$\frac{\partial}{\partial\theta} g_{n+1}(x_1, \ldots, x_n; \theta) = B(x_1, \ldots, x_n) \left[g_n(x_1, \ldots, x_{n-1}; \theta) - g_{n+1}(x_1, \ldots, x_n; \theta) \right].$$

Proof: Write $f_\lambda(t) = \lambda e^{-\lambda t}$ for $t \geq 0$. By the convolution formula (Section 3)

$$g_{n+1}(x_1, \ldots, x_n; \theta) = \int_0^\theta g_n(x_1, \ldots, x_{n-1}; s)\, f_{B(\vec{\mathbf{x}}_{\leq n})}(\theta - s)\, ds.$$

The function $(e^{-\lambda h} - 1)/h$ is uniformly bounded in absolute value by λ. Moreover, $f_\lambda(\cdot)$ and $g_n(x_1, \ldots, x_{n-1}; \cdot)$ are uniformly bounded, and continuously differentiable, so by dominated convergence

$$\lim_{h \downarrow 0} \frac{1}{h} \int_0^\theta g_n(x_1, \ldots, x_{n-1}; s) \left[f_{B(\vec{\mathbf{x}}_{\leq n})}(\theta + h - s) - f_{B(\vec{\mathbf{x}}_{\leq n})}(\theta - s) \right] ds =$$

$$\int_0^\theta g_n(x_1, \ldots, x_{n-1}; s)\, f'_{B(\vec{\mathbf{x}}_{\leq n})}(\theta - s)\, ds.$$

Furthermore, the continuity of $f_\lambda(\cdot)$ and the fundamental theorem of analysis imply

$$\lim_{h \downarrow 0} \frac{1}{h} \int_\theta^{\theta+h} g_n(x_1, \ldots, x_{n-1}; s)\, f_{B(\vec{\mathbf{x}}_{\leq n})}(\theta + h - s)\, ds =$$

$$g_n(x_1, \ldots, x_{n-1}; \theta)\, f_{B(\vec{\mathbf{x}}_{\leq n})}(0).$$

Hence the derivative of $g_{n+1}(x_1, \ldots, x_n; \theta)$ can be written as

$$-B(\vec{\mathbf{x}}_{\leq n}) \int_0^\theta g_n(x_1, \ldots, x_{n-1}; s)\, f_{B(\vec{\mathbf{x}}_{\leq n})}(\theta - s)\, ds + B(\vec{\mathbf{x}}_{\leq n})\, g_n(x_1, \ldots, x_{n-1}; \theta) =$$

$$B(\vec{\mathbf{x}}_{\leq n}) \left[g_n(x_1, \ldots, x_{n-1}; \theta) - g_{n+1}(x_1, \ldots, x_n; \theta) \right].$$

A similar argument for the limit $h \uparrow 0$ completes the proof. □

Upon observation of $(x_1, \ldots, x_n)$, consider the only non-trivial optimisation problem, that is for $n \geq 1$ and $B(x_1, \ldots, x_n) > 0$. We have to find out whether

$$\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n) := \text{ argsup }_{\theta \geq 0}\, g_{n+1}(x_1, \ldots, x_n; \theta)$$

exists and is unique.

Theorem 2. *Let $n \geq 1$, $\lambda_1 > \lambda_2 > \cdots > \lambda_{n+1} > 0$. Then the function*

$$f(t) := (-1)^{n+2} \sum_{i=1}^{n+1} \left[\lambda_i e^{-\lambda_i t} \prod_{1 = j \neq i}^{n+1} \frac{\lambda_j}{\lambda_i - \lambda_j} \right]$$

is unimodal with a unique maximum and no local extrema (including flats).

Note that $\lambda_i = B(\vec{\mathbf{x}}_{<i})$, is a strictly decreasing sequence, and for this choice the function $f(t)$ in the above theorem is the likelihood function for the time horizon (cf. Lemma 1 and Theorem 1).

Proof: By the convolution formula of Section 3, a minimum $f(0) = 0$ is attained at $t = 0$. Also the limit for $t \to \infty$ is zero. As $f(t)$ is a probability density that integrates to unity, there is a t_0 such that $f(t_0) > 0$. Hence, a $t_\infty > 0$ can be found such that for $t > t_\infty$, $f(t) < f(t_0)$. Clearly, $f(\cdot)$ is continuous as a sum of continuous functions, so on the compact interval $[0, t_\infty]$, $f(\cdot)$ attains its maximum too.

It remains to establish uniqueness of maximum and minimum. For $n = 1$,

$$f(t) = \frac{\lambda_1 \lambda_2}{\lambda_1 - \lambda_2} \left[e^{-\lambda_2 t} - e^{-\lambda_1 t} \right] .$$

The constant in front is positive and can be ignored. The derivative of the term in brackets is

$$\lambda_1 e^{-\lambda_1 t} - \lambda_2 e^{-\lambda_2 t} = e^{-\lambda_1 t} \left[\lambda_1 - \lambda_2 e^{(\lambda_1 - \lambda_2)t} \right]$$

which has a unique zero at $t^* = (\lambda_1 - \lambda_2)^{-1}(\log \lambda_1 - \log \lambda_2)$ with positive values for smaller t, negative ones for $t > t^*$. We conclude that for $n = 1$, the function $f(\cdot)$ increases from a boundary minimum $f(0) = 0$ at $t = 0$ to a unique maximum, then decreases to an asymptote level 0 (cf. Figure 1). In particular, there are no local extrema (zeroes of the derivative).

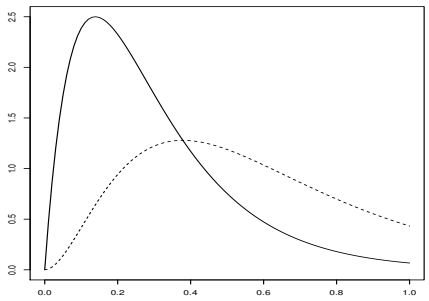

Figure 1: Graph of $f(t)$ for $\lambda_1 = 10$, $\lambda_2 = 5$ (solid line) and $\lambda_3 = 3$ (dotted line).

For $n \geq 2$, the boundary minimum at 0 corresponds, in contrast to the case $n = 1$, to a zero score (Lemma 2).

Next, proceed by induction. Suppose that $f(\cdot)$ is unimodal for $\lambda_1 > \cdots > \lambda_{n+1}$ with no local extrema, and consider the convolution with another exponential distribution with parameter $0 < \lambda_{n+2} < \lambda_{n+1}$. To distinguish between them, write $f_{n+1}(\cdot)$ respectively $f_{n+2}(\cdot)$.

Suppose the maximum is attained at two points $\theta_1 \neq \theta_2$, without loss of generality $\theta_1 < \theta_2$. Then

$$f_{n+2}(\theta_1) = f_{n+2}(\theta_2) \geq f_{n+2}(t) \text{ for all } t \geq 0$$

and, since $f_{n+2}(\cdot)$ is differentiable everywhere and the maximum cannot be at the boundary $t = 0$, the derivative in both points is zero, that is

$$f_{n+1}(\theta_1) - f_{n+2}(\theta_1) = 0 = f_{n+1}(\theta_2) - f_{n+2}(\theta_2)$$

by Lemma 2. Hence $f_{n+1}(\theta_1) = f_{n+1}(\theta_2)$, so the induction hypothesis implies that θ_1 and θ_2 must lie on either side of the top of f_{n+1}. In particular, $f_{n+1}(t) > f_{n+1}(\theta_1)$ for all $t \in (\theta_1, \theta_2)$. For such t,

$$f_{n+1}(t) - f_{n+2}(t) > f_{n+1}(\theta_1) - f_{n+2}(t) \geq f_{n+1}(\theta_1) - f_{n+2}(\theta_1) = 0.$$

Consequently, $f_{n+2}(t)$ is increasing on (θ_1, θ_2), so the endpoints cannot be global maxima. Hence, there is a unique maximum, which will be denoted by θ.

To complete the proof, we show that $f_{n+2}(t)$ is strictly increasing for $0 < t < \theta$, strictly decreasing for $t > \theta$. First, suppose $f_{n+2}(t) \equiv c$ is flat for $t \in (\theta_0, \theta_0')$. Then the derivative is zero at these t, hence $f_{n+1}(t)$ is flat on (θ_0, θ_0') as well, a contradiction with the induction hypothesis.

Next, focus on $(0, \theta)$, and suppose $f_{n+2}(\cdot)$ has a strict local maximum, say $\theta_0 < \theta$. Then the derivative in θ_0 is zero, hence

$$f_{n+1}(\theta_0) = f_{n+2}(\theta_0) < f_{n+2}(\theta) = f_{n+1}(\theta). \tag{4.2}$$

By the induction hypothesis on the shape of $f_{n+1}(\cdot)$, it follows that $f_{n+1}(t)$ increases for t in a neighbourhood of θ_0. By the assumption that θ_0 is locally maximal, $f_{n+2}(t)$ decreases in an interval to the right of θ_0. Hence,

$$f_{n+1}(t) - f_{n+2}(t) > f_{n+1}(\theta_0) - f_{n+2}(\theta_0) = 0$$

which implies that $f_{n+2}(\cdot)$ is increasing for t in a right-neighbourhood of θ_0, a contradiction. The existence of a strict local minimum for $f_{n+2}(\cdot)$ at some $\theta_0 < \theta$ implies the existence of a local maximum (or flat) at some smaller t (as $f(0) = 0$), for which one may apply the arguments above to derive a contradiction.

For $t > \theta$, suppose there is a strict local minimum, say $\theta_0 > \theta$. Then (4.2) holds. By the induction hypothesis, $f_{n+1}(\cdot)$ decreases near θ_0, whereas $f_{n+2}(\cdot)$ first decreases, then increases. Hence, for t in a right-neighbourhood of θ_0,

$$f_{n+1}(t) - f_{n+2}(t) < f_{n+1}(\theta_0) - f_{n+2}(\theta_0) = 0,$$

that is, $f_{n+2}(\cdot)$ is decreasing for t in a right-neighbourhood of θ_0, a contradiction. Finally, note that the existence of a strict local maximum for $f_{n+2}(\cdot)$ at some $\theta_0 > \theta$ implies the existence of a local minimum or plateau in (θ, θ_0), which both lead to a contradiction as seen above, an observation that completes the proof. □

5. Examples

We evaluated the behaviour of the maximum likelihood approach for the classic random sequential adsorption model of Definition 1 by means of the data given in Figure 2 (sampled with true values $r = 0.050$ and $\theta = 100.0$). The minimum inter-point distance is $\hat{r} = 0.052$, the estimated time horizon $\hat{\theta} = 99.9$. The convolution densities were computed (with the aid of the R-software available at `www.cran.r-project.org`) by kernel estimation using sample size $10,000$, areas by the Monte Carlo method based on the same sample size.

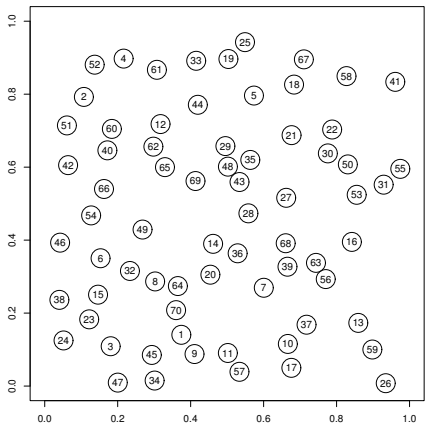

Figure 2: Sequence of 70 points in the planar unit square.

Heterogeneous adsorption

As an example of a spatially heterogeneous model with competition advantage, consider the metric $\rho(x, y) = \max\{|x_1 - y_1|, |x_2 - y_2|\}$ for $x = (x_1, x_2)$, $y = (y_1, y_2)$ in the plane $\mathbb{R}^2$, so that the ρ-ball of radius r centered at x is a square with side length $2r$ and midpoint x, and let

$$\pi(z_1, z_2) = \frac{\lambda^2}{4} \exp\left[-\lambda\left(|z_1 - \frac{1}{2}| + |z_2 - \frac{1}{2}|\right)\right] \tag{5.1}$$

be the joint density of two independent Laplacian distributed components, if necessary restricted to the unit square D. A realisation of the model with 63 points and a time horizon of $1,000.0$, a hard core distance $r = 0.030$ and $\lambda = 25.0$ is given in Figure 3 (left). The hard squares have side r, their colours indicate the position in the sequence: dark indicates an

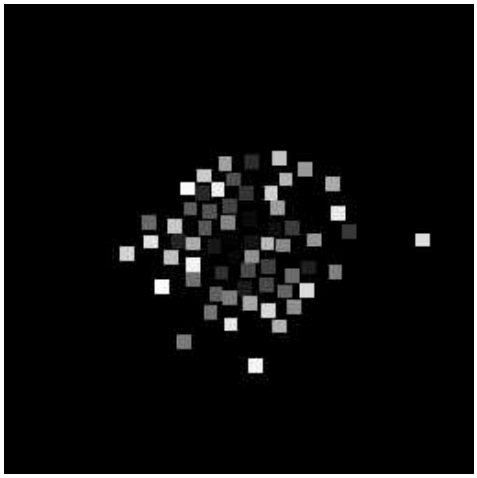

Figure 3: Sample from a sequential adsorption process with Laplacian location selection with dispersion parameter $\lambda = 25.0$, time horizon $1,000.0$, and hard core distance $r = 0.030$ in both coordinates.

early arrival or low position index, light corresponds to a late arrival. Note that dark discs dominate the centre of the picture, where (5.1) is high, and that light discs are relatively often found on the outskirts of the point cloud.

A straightforward computation yields

$$\int_a^b \frac{\lambda}{2} \exp\left[-\lambda|z-\frac{1}{2}|\right] dz = \begin{cases} \frac{1}{2}e^{-\lambda/2}\left(e^{\lambda b} - e^{\lambda a}\right) & \text{if } b \leq \frac{1}{2} \\ \frac{1}{2}e^{\lambda/2}\left(e^{-\lambda a} - e^{-\lambda b}\right) & \text{if } a \geq \frac{1}{2} \\ 1 - \frac{1}{2}e^{-\lambda/2}e^{\lambda a} - \frac{1}{2}e^{\lambda/2}e^{-\lambda b} & \text{if } a < \frac{1}{2} \text{ and } b > \frac{1}{2} \end{cases}$$

for any $0 \leq a < b \leq 1$, so that the birth rates are easily evaluated. The convolution densities were computed by kernel estimates using sample size $10,000$ as before. We obtained the estimates $\hat{r} = 0.030$ and $\hat{\theta} = 1003.9$ for the hard core distance and time horizon.

Nuisance parameters

The location distribution (5.1) contains a nuisance parameter λ. In practice, λ may be known or estimated separately, e.g. from observable covariates. In the absence of such information, λ can be estimated using the profile likelihood

$$L(\lambda; \vec{\mathbf{x}}) = f_\lambda(\vec{\mathbf{x}}; \hat{r}_\lambda(\vec{\mathbf{x}}), \hat{\theta}_\lambda(\vec{\mathbf{x}})).$$

Thus, discretise λ in a grid, for each value, find the maximum likelihood estimators $\hat{r}_\lambda(\vec{\mathbf{x}})$ and $\hat{\theta}_\lambda(\vec{\mathbf{x}})$, and optimise the profile likelihood over λ to obtain λ^*, which gives final estimates λ^*, $\hat{r}_{\lambda^*}(\vec{\mathbf{x}})$, and $\hat{\theta}_{\lambda^*}(\vec{\mathbf{x}})$. The result is given in Figure 4 which plots $\log L(\lambda; \vec{\mathbf{x}})$ up to terms that do not depend on λ for a grid mesh 0.1. The optimal value is found at $(\hat{\lambda}, \hat{r}, \hat{\theta}) = (26.2, 0.030, 1136.1)$.

6. Summary

In this paper, we proved that the maximum likelihood method can be applied to estimate the parameters in random sequential adsorption models, and showed that the idea works in practice. It is important to note that the method is not restricted to RSA, but can be applied in principle to estimate the time horizon in other cooperative sequential inhibition processes [7]. In particular, the sequential spatial process framework applies so that densities can be derived as in Theorem 1, provided proper care is taken of any zero rates that might arise.

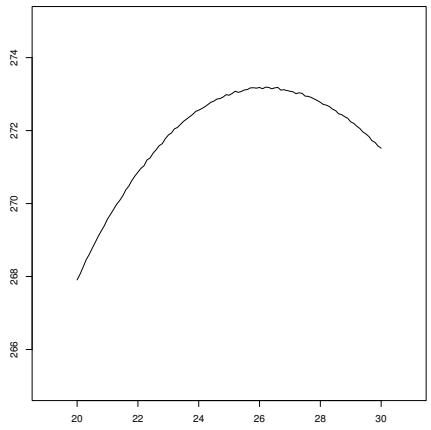

Figure 4: Graph of $\log L(\lambda; \vec{\mathbf{x}})$, up to terms that do not depend on λ, versus λ for a sequential adsorption model with locations chosen according to a Laplacian distribution with dispersion parameter λ. The sequence $\vec{\mathbf{x}}$ is as depicted in Figure 3, a grid mesh 0.1 was used.

References

1. N.A.C. Cressie. *Statistics for spatial data.* John Wiley & Sons, New York, 1991.
2. D.J. Daley and D. Vere-Jones. *An introduction to the theory of point processes.* Springer Verlag, New York, 1988.
3. P.J. Diggle. *Statistical analysis of spatial point patterns.* Academic Press, London, 1983.
4. P.J. Diggle, J. Besag, and J.T. Gleaves. Statistical analysis of spatial point patterns by means of distance methods. *Biometrics*, 32:659–667, 1976.
5. G. Döge. Perfect simulation for random sequential adsorption of d-dimensional spheres with random radii. *Journal of Statistical Computation and Simulation*, 69:141–156, 2001.

6. A. Dvoretzky and H. Robbins. On the 'parking' problem. *Publications of the Mathematical Research Institute of the Hungarian Academy of Sciences*, 9:209–225, 1964.

7. J.W. Evans. Random and cooperative sequential adsorption. *Reviews of Modern Physics*, 65:1281–1329, 1993.

8. J. Feder. Random sequential adsorption. *Journal of Theoretical Biology*, 87:237–254, 1980.

9. W. Feller. *An introduction to probability theory and its applications*. Volume II, Second Edition. John Wiley & Sons, New York, 1970.

10. P.G. Hall. *Introduction to the theory of coverage processes*. John Wiley & Sons, New York, 1988.

11. M.N.M. van Lieshout. Markovianity in space and time. CWI Report PNA-R0501, CWI, Amsterdam, February 2005. To appear in Festschrift Mike Keane, IMS, 2005.

12. H.W. Lotwick. Simulations of some spatial hard core models, and the complete packing problem. *Journal of Statistical Computation and Simulation*, 15:295–314, 1982.

13. M.D. Penrose. Random parking, sequential adsorption, and the jamming limit. *Communications in Mathematical Physics*, 218:153–176, 2001.

14. A. Rényi. On a one-dimensional problem concerned with random space-filling (in Hungarian, 1958). Reprinted in English in *Selected papers of Alfréd Rényi*, Volume 2. P. Turán (Ed.) Akadémiai Kiadó, Budapest, 1976.

15. B.D. Ripley. *Statistical inference for spatial processes*. Cambridge University Press, Cambridge, 1988.

16. B. Senger, J.-C. Voegel and P. Schaaf. Irreversible adsorption of colloidal particles on solid substrates. *Colloids and Surfaces A: Physiochemical and Engineering Aspects*, 165, 255–285, 2000.

17. D. Stoyan. Random sets: Models and statistics. *International Statistical Review*, 66:1–27, 1998.

18. M. Tanemura. On random complete packing by disks. *Annals of the Institute of Statistical Mathematics*, 31:351–365, 1979.

19. G.J.G. Upton and B. Fingleton. *Spatial data analysis by example*, Vol. I. Point pattern and quantitative data. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Chichester, 1985.