

ASYMPTOTIC NORMALITY OF THE MAXIMUM LIKELIHOOD ESTIMATOR FOR COOPERATIVE SEQUENTIAL ADSORPTION

MATHEW D. PENROSE,* *University of Bath*

VADIM SHCHERBAKOV,** *Moscow State University*

Abstract

We consider statistical inference for a parametric cooperative sequential adsorption model for spatial time series data, based on maximum likelihood. We establish asymptotic normality of the maximum likelihood estimator in the thermodynamic limit. We also perform and discuss some numerical simulations of the model, which illustrate the procedure for creating confidence intervals for large samples.

Keywords: Cooperative sequential adsorption; time series of spatial locations; spatial random growth; maximum likelihood estimation; asymptotic normality; Fisher information; martingale; thermodynamic limit

2010 Mathematics Subject Classification: Primary 62M30; 60K35
Secondary 60D05

1. Introduction

This paper is concerned with maximum likelihood estimation for the cooperative sequential adsorption (CSA) model for spatial time series. The main peculiarity of CSA is that the likelihood for a point (particle) to appear at a given location in space depends on the number of previous particles of the series nearby; depending on the parameters, the new particle may be attracted to, or repelled by, previous particles. Mathematically, CSA is formulated as a random finite sequential allocation of particles in a bounded region of space (the observation window).

Originally motivated by adsorption processes in physics and chemistry [3], CSA dynamics seem to be relevant to many applications. It was first noticed by physicists (see, e.g. [3, p. 1285]) that this type of model can be used for modelling the spatial-temporal processes similar to the irreversible spread of disease or epidemics, and biological growth was mentioned in [3] as another potential application. These ideas have been recently supported by both experimental and simulation studies of the keratin filament (KF) network formation in biology. KF networks are part of the cell cytoskeleton and they determine the shape and biophysical properties of the cells. Loosely speaking, the KF is an aggregated spatial structure formed by a union of curved finite segments (fibres). Experimental results [12] and simulation studies [1] suggest that the KF can be thought as a result of a sequential spatial growth process with self-organising properties. It is also argued in [10] (see also the references therein) that self-organising processes combined with simple physical constraints seem to have key roles in controlling organelle size, number,

Received 2 August 2010; revision received 10 February 2011.

* Postal address: Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK.

Email address: masmdp@bath.ac.uk

** Postal address: Laboratory of Large Random Systems, Faculty of Mechanics and Mathematics, Moscow State University, Glavnoe Zdanie, Leninskie Gory, Moscow, 119991, Russia. Email address: v.shcherbakov@mech.math.msu.su

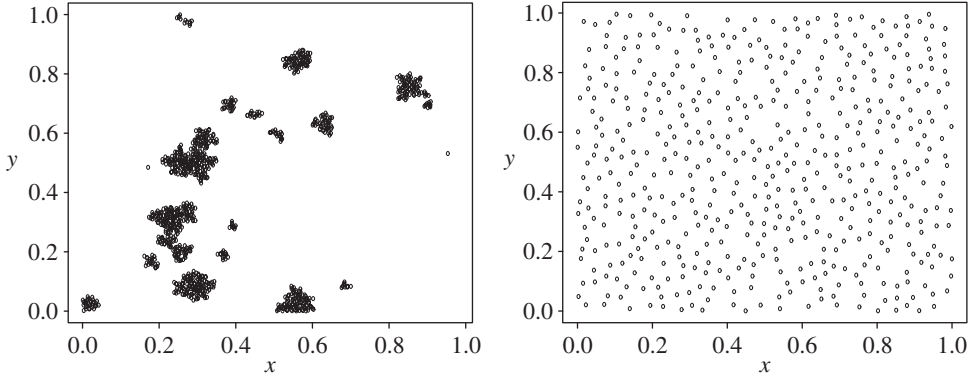


FIGURE 1: CSA simulations in $D = [0, 1]^2$. *Left:* $\ell = 1000$, $R = 0.01$, and $(\beta_i)_{i \geq 0} = (1, 1000, 10\,000, 0, 0, \dots)$. *Right:* $\ell = 500$, $R = 0.03$, and $(\beta_i)_{i \geq 0} = (1, 0, 0, \dots)$, i.e. RSA.

shape, and position, and these factors then combine to produce the overall cell architecture. CSA seems to be useful for modelling spatial random growth with self-organising properties.

Following [9] and [11] we consider a parametric version of CSA formulated as follows. Consider a sequence of points X_1, X_2, \dots, X_ℓ located randomly in a bounded convex region D of Euclidean space \mathbb{R}^d . Let the parameters $R > 0$ (the interaction radius) and $\beta_i \geq 0$ ($i = 0, 1, 2, \dots$) be fixed. Given the first k points X_1, \dots, X_k , let the conditional probability density of X_{k+1} at $x \in D$ be proportional to β_i , if x has i points among X_1, \dots, X_k within distance R of it. The special case of CSA with $\beta_0 = 1$ and $\beta_i = 0$ for $i \geq 1$ is known as random sequential adsorption (RSA).

CSA can be used for modelling both clustered and regular point patterns. A variety of spatial point patterns can be generated by modulating the parameters. See, for example, Figure 1.

We consider statistical inference for the parameters of our model based on maximum likelihood estimation (MLE). In earlier work [9] we showed that the maximum likelihood estimator exists uniquely, and is consistent in the thermodynamic limit, whereby the observation window expands to the whole space and the number of allocated points grows linearly in the volume of the window. The main result of the present paper is asymptotic normality of the maximum likelihood estimator in the same limiting regime, which we prove in Section 3 using martingale techniques. This provides asymptotic justification for the creation of confidence intervals based on the normal distribution, which we illustrate on simulated examples in Section 5. Section 6 contains some concluding remarks.

2. Assumptions and results

Let D , R , and $\{\beta_k, k \geq 0\}$ be as in Section 1. For any $x \in \mathbb{R}^d$ and any finite sequence $y = (y_1, \dots, y_n)$, $n \geq 1$, of points in \mathbb{R}^d , we denote by $\nu(x, y)$ the number of points y_i in the sequence y , such that the distance between x and y_i is at most R . By definition, $\nu(x, \emptyset) = 0$.

Our CSA model is defined as follows. Let $X(\ell) = (X_1, \dots, X_\ell) \in D^\ell$ be the vector of first ℓ random points sequentially generated by CSA. Given the sequence $X(k)$ (which can be empty, i.e. $k = 0$), the conditional probability density function of the next point X_{k+1} is

$$\psi_{k+1}(x \mid X(k)) = \frac{\beta_{\nu(x, X(k))}}{\int_D \beta_{\nu(y, X(k))} \, dy}, \quad x \in D.$$

Write $p_{\ell,\beta,D}(x_1, \dots, x_\ell)$ for the joint density function $\prod_{k=1}^\ell \psi_k(x_k \mid x(k-1))$ of the first ℓ points.

As in [9], we assume throughout that the interaction radius R is a *fixed* and *known* constant, that $\beta_0 = 1$ (this entails no loss of generality; see [9]), and that there is a (possibly unknown) constant $N \in \mathbb{N}$ such that $\beta_k > 0$ for $k \leq N$ and $\beta_k = 0$ for $k > N$. One interpretation of the last assumption is that it might be impossible to distinguish a new point that has too many existing points nearby, and, therefore, such new points are not observed. If N is unknown, it has to be estimated before estimating the β s. In a fixed finite volume, N is estimated by \hat{N} , which is defined by Equation (2.9) of [9]. Given \hat{N} , the rest of the estimation procedure is carried out as described in Section 2.2 of [9]. By Theorem 2.2 of [9], \hat{N} converges to N in probability in the *thermodynamic limit* (recalled below). That is, we have an exact value for N with high confidence, but only confidence intervals for the continuous parameters β_k , $k \leq N$; we are interested here in constructing these intervals. Therefore, for simplicity of notation, we assume in the rest of the paper that $\hat{N} = N$.

For asymptotics, it is natural to consider the thermodynamic limit which we now describe. Let D_1 be the unit cube centred at the origin (or some other bounded convex region with unit volume), and consider a sequence of rescaled domains

$$D_m = m^{1/d} D_1, \quad m \in \mathbb{Z}_+.$$

Fix $\{\ell_m, m \geq 1\}$, an arbitrary monotonically increasing \mathbb{N} -valued sequence. Here ℓ_m represents the number of observed points in the domain D_m .

Assumption 1. *The number of observed points is asymptotically linear in m , that is,*

$$\lim_{m \rightarrow \infty} \frac{\ell_m}{m} = \mu \in (0, \theta_\infty), \tag{1}$$

where θ_∞ is the *jamming density* (see [9]).

Given parameters N and $\beta = (\beta_1, \dots, \beta_N)$, consider the probability measure $P_{m,\beta}$ on $D_m^{\ell_m}$ specified by the probability density $p_{\ell,\beta,D}$ with $\ell = \ell_m$ and $D = D_m$. The expectation with respect to this measure is denoted by $E_{m,\beta}$. We assume that $\beta \in \mathcal{B}$, where $\mathcal{B} := (0, \infty)^N$. The true parameter is denoted by $\beta^{(0)} = (\beta_1^{(0)}, \dots, \beta_N^{(0)})$. Also, we define for short $P_m^{(0)} = P_{m,\beta^{(0)}}$ and $E_m^{(0)} = E_{m,\beta^{(0)}}$.

Given m , assume that $\ell_m \geq 2$ and define the log-likelihood function

$$L_m(X^m(\ell_m), \beta) = \log(p_{\ell_m,\beta,D_m}(X_1^m, \dots, X_{\ell_m}^m)),$$

where $X^m(\ell_m) = (X_1^m, \dots, X_{\ell_m}^m)$ is the vector of observed points in D_m . Given observation $X^m(\ell_m)$, we define the maximum likelihood estimators

$$\hat{\beta}(X^m(\ell_m)) = (\hat{\beta}_{1,m}, \dots, \hat{\beta}_{N,m})$$

of parameters $\beta^{(0)} = (\beta_1^{(0)}, \dots, \beta_N^{(0)})$ as maximisers of the function $L_m(X^m(\ell_m), \beta)$, which can be found as a solution of $\nabla L_m(X^m(\ell_m), \beta) = 0$, where we define

$$\nabla L_m(X^m(\ell_m), \beta) = \left(\frac{\partial L_m(X^m(\ell_m), \beta)}{\partial \beta_1}, \dots, \frac{\partial L_m(X^m(\ell_m), \beta)}{\partial \beta_N} \right).$$

The following two results are from [9] (see Theorem 2.2 and Lemma 5.2, part 2, respectively, of [9]).

Lemma 1. (Consistency of MLE.) *Under Assumption 1, with $P_m^{(0)}$ -probability tending to 1 as $m \rightarrow \infty$, there exists a unique positive solution $(\hat{\beta}_{1,m}, \dots, \hat{\beta}_{N,m})$ of the likelihood equations and*

$$(\hat{\beta}_{1,m}, \dots, \hat{\beta}_{N,m}) \rightarrow (\beta_1^{(0)}, \dots, \beta_N^{(0)})$$

in $P_m^{(0)}$ -probability as $m \rightarrow \infty$.

Lemma 2. (Asymptotics of the observed information matrix.) *Consider the matrix*

$$J_m(X^m(\ell_m), \beta) := - \left(\frac{\partial^2 L_m(X^m(\ell_m), \beta)}{\partial \beta_i \partial \beta_j} \right)_{i,j=1}^N.$$

There is a family of $N \times N$ positive definite real matrices $J^{(0)}(\mu)$, defined for $\mu \in (0, \theta_\infty)$, such that, under Assumption 1, if $\beta(m)$ is a sequence of \mathcal{B} -valued random variables converging in probability to $\beta^{(0)}$ as $m \rightarrow \infty$ then

$$\frac{J_m(X^m(\ell_m), \beta(m))}{m} \rightarrow J^{(0)}(\mu) \tag{2}$$

in $P_m^{(0)}$ -probability as $m \rightarrow \infty$.

In [9], we proved (2) for only the special case with $\beta(m) = \beta^{(0)}$, but the general case stated above can be proved in the same manner as that result.

In Section 4 we give an extended study of the structure of the limiting information matrix $J^{(0)}(\mu)$.

We can now give the main result of the paper, which states that the MLE is asymptotically normal.

Theorem 1. *Under Assumption 1,*

$$\sqrt{m}(\hat{\beta}(X^m(\ell_m)) - \beta^{(0)}) \rightarrow \mathcal{N}(0, (J^{(0)}(\mu))^{-1})$$

in $P_m^{(0)}$ -distribution as $m \rightarrow \infty$, where $\mathcal{N}(0, (J^{(0)}(\mu))^{-1})$ is the Gaussian vector with zero mean and covariance matrix $(J^{(0)}(\mu))^{-1}$.

Theorem 1 provides asymptotic justification for creating confidence intervals based on the normal distribution, when we have a large number of points in a large region relative to the interaction radius R ; this is the case for the simulated examples in Section 5, which take place on the unit square with small R . We shall describe the creation of confidence intervals in more detail in the context of those examples.

We prove Theorem 1 using the following limiting result for the model score

$$\nabla L_m(X^m(\ell_m), \beta^{(0)}),$$

which we prove in Section 3.

Theorem 2. *Under Assumption 1, $m^{-1/2} \nabla L_m(X^m(\ell_m), \beta^{(0)})$ converges in distribution as $m \rightarrow \infty$ to a Gaussian vector with zero mean and covariance matrix $J^{(0)}(\mu)$.*

Proof of Theorem 1. Choose $\delta > 0$ such that the ball of radius δ centred at $\beta^{(0)}$ is contained in \mathcal{B} . By Lemma 1 we have

$$|\hat{\beta}(X^m(\ell_m)) - \beta^{(0)}| < \delta,$$

with probability $P_m^{(0)}$ close to 1 if m is large enough. With ∂_i denoting differentiation with respect to the i th component of β , we implement a Taylor expansion of $\partial_i(L_m(X^m(\ell_m), \beta))$ about $\beta^{(0)}$:

$$\begin{aligned} 0 &= \partial_i L_m(X^m(\ell_m), \hat{\beta}(X^m(\ell_m))) \\ &= \partial_i L_m(X^m(\ell_m), \beta^{(0)}) + \sum_{j=1}^N \partial_{ij}^2(X^m(\ell_m), \bar{\beta})(\hat{\beta}(X^m(\ell_m)) - \beta^{(0)})_j, \end{aligned}$$

where $\bar{\beta}$ lies on the line segment from $\beta^{(0)}$ to $\hat{\beta}(X^m(\ell_m))$. Rewriting this expression, we obtain

$$\sum_{j=1}^N \frac{-\partial_{ij}^2(X^m(\ell_m), \bar{\beta})}{m} (\sqrt{m}(\hat{\beta}(X^m(\ell_m)) - \beta^{(0)})_j) = \frac{\partial_i L_m(X^m(\ell_m), \beta^{(0)})}{\sqrt{m}}.$$

In the left-hand expression $\bar{\beta}$ depends on i but converges in probability to $\beta^{(0)}$ as $n \rightarrow \infty$ by Lemma 1. By Lemma 2, for each (i, j) , the first factor inside the sum converges in probability to $J_{ij}^{(0)}(\mu)$. Observing that Theorem 2 applies to the right-hand side, we can complete the proof by applying Lemma 6.4.1 of [5].

3. Proof of Theorem 2

The proof of Theorem 2 is based on the observation that each component of $\nabla L_m(X^m(\ell_m), \beta)$ can be expressed as a sum of martingale differences; see (8) below.

Let $X^m(\ell_m) = (X_1^m, \dots, X_{\ell_m}^m)$ be the sequence of observed points X_i^m in D_m . Define the quantities

$$\xi_{k,i}^m := \mathbf{1}_{\{v(X_i^m, X^{m(i-1)})=k\}}, \quad k = 0, \dots, N, \quad i = 1, \dots, \ell_m, \tag{3}$$

$$t_{k,\ell_m}^m := \sum_{i=1}^{\ell_m} \xi_{k,i}^m, \quad k = 0, \dots, N, \tag{4}$$

and

$$\Gamma_{k,i}^m := \int_{D_m} \mathbf{1}_{\{u: v(u, X^{m(i)})=k\}} du, \quad 0 \leq i \leq \ell_m, \quad k \geq 0.$$

Note that $\Gamma_{k,i}^m = 0$ for $i < k$ and that $\Gamma_{0,0}^m$ is equal to m . In terms of t - and Γ -statistics, the k th component of $\nabla L_m(X^m(\ell_m), \beta)$ can be written as follows (see Equation (6.5) of [9]):

$$\frac{\partial L_m(X^m(\ell_m), \beta)}{\partial \beta_k} = \frac{t_{k,\ell_m}^m}{\beta_k} - \sum_{j=1}^{\ell_m} \frac{\Gamma_{k,j-1}^m}{\Gamma_{0,j-1}^m + \sum_{i=1}^N \beta_i \Gamma_{i,j-1}^m} \tag{5}$$

for $j = 1, \dots, N$. Let $\mathcal{F}_j^{(m)} = \sigma\{X_1^m, \dots, X_j^m\}$ be the σ -algebra generated by the first j points observed in D_m . Define, for short,

$$\bar{\xi}_{k,i}^m := E_{m,\beta}(\xi_{k,i}^m \mid \mathcal{F}_{i-1}^{(m)}).$$

It is easy to see that

$$\bar{\xi}_{k,i}^m = \frac{\beta_k \Gamma_{k,i-1}^m}{\Gamma_{0,i-1}^m + \sum_{j=1}^N \beta_j \Gamma_{j,i-1}^m}, \quad k = 1, \dots, N, \quad i = 1, \dots, \ell_m. \tag{6}$$

Given k and m , define the sequence of martingale differences

$$\zeta_{k,i}^m := \frac{1}{\beta_k} (\xi_{k,i}^m - \bar{\xi}_{k,i}^m), \quad 1 \leq i \leq \ell_m. \tag{7}$$

Equation (5) can now be rewritten as

$$\frac{\partial L_m(X^m(\ell_m), \beta)}{\partial \beta_k} = \frac{1}{\beta_k} \sum_{i=1}^{\ell_m} (\xi_{k,i}^m - \bar{\xi}_{k,i}^m) = \sum_{i=1}^{\ell_m} \zeta_{k,i}^m, \quad k = 1, \dots, N. \tag{8}$$

This implies that, for any real vector $\mathbf{a} = (a_1, \dots, a_N)^\top$, the triangle array

$$\left[E_{m,\beta} \left(\frac{1}{\sqrt{m}} \sum_{k=1}^N a_k \frac{\partial L_m(X^m(\ell_m), \beta)}{\partial \beta_k} \mid \mathcal{F}_j^{(m)} \right), \mathcal{F}_j^{(m)} \right]_{j=1}^{\ell_m}, \quad m \geq 2,$$

is a zero-mean square integrable martingale array.

By the Cramér–Wold device (see, for example, [2]), Theorem 2 follows from the following fact.

Lemma 3. *Under Assumption 1, for any real vector $\mathbf{a} = (a_1, \dots, a_N)^\top$,*

$$\frac{1}{\sqrt{m}} \sum_{k=1}^N a_k \frac{\partial L_m(X^m(\ell_m), \beta^{(0)})}{\partial \beta_k} \rightarrow \mathcal{N}(0, \sigma_{\mathbf{a}}^2)$$

in distribution as $m \rightarrow \infty$, where

$$\sigma_{\mathbf{a}}^2 = \mathbf{a}^\top J^{(0)}(\mu) \mathbf{a},$$

$J^{(0)}(\mu)$ is the matrix arising in Lemma 2, and $\mathcal{N}(0, \sigma_{\mathbf{a}}^2)$ is the Gaussian vector with zero mean and variance $\sigma_{\mathbf{a}}^2$.

In proving Lemma 3 we shall repeatedly use the following fact which is simple enough for us to omit its proof.

Lemma 4. *Let $\xi_n, n \geq 1$, and $\eta_n, n \geq 1$, be two sequences of random variables, and let $C > 0, a$, and b be some constants. Suppose that $|\xi_n| < C, |\eta_n| < C, \xi_n \rightarrow a$ in probability as $n \rightarrow \infty$, and $E(\eta_n) \rightarrow b$ as $n \rightarrow \infty$. Then $E(\xi_n \eta_n) \rightarrow ab$ as $n \rightarrow \infty$.*

Proof of Lemma 3. By (8), for any $\beta \in \mathcal{B}$, we have

$$\frac{1}{\sqrt{m}} \sum_{k=1}^N a_k \frac{\partial L_m(X^m(\ell_m), \beta)}{\partial \beta_k} = \frac{1}{\sqrt{m}} \sum_{i=2}^{\ell_m} \eta_i^m,$$

where

$$\eta_i^m = \sum_{k=1}^N a_k \zeta_{k,i}^m$$

and $\zeta_{k,i}^m$ are the quantities defined by (7). It is easy to see that

$$\frac{1}{\sqrt{m}} \max_i |\eta_i^m| \leq \frac{2N}{\sqrt{m}} \max_{k=1, \dots, N} \frac{a_k}{\beta_k} \rightarrow 0 \quad \text{as } m \rightarrow \infty \tag{9}$$

and

$$\frac{1}{m} E_{m,\beta}(\max_i (\eta_i^m)^2) \leq \frac{4N^2}{m} \max_{k=1,\dots,N} \left(\frac{a_k}{\beta_k}\right)^2 \rightarrow 0 \quad \text{as } m \rightarrow \infty. \tag{10}$$

By Propositions 1 and 2 below, we also have, under Assumption 1,

$$\frac{1}{m} \sum_{i=2}^{\ell_m} (\eta_i^m)^2 \rightarrow \mathbf{a}^\top J^{(0)}(\mu) \mathbf{a} \tag{11}$$

in $P_m^{(0)}$ -probability as $m \rightarrow \infty$.

Using (9), (10), and (11), we can then apply the central limit theorem for martingale difference arrays (see Theorem (2.3) of [7]) to complete the proof of Lemma 3.

Proposition 1. *Under Assumption 1,*

$$\lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=2}^{\ell_m} E_m^{(0)}((\eta_i^m)^2) = \mathbf{a}^\top J^{(0)}(\mu) \mathbf{a}.$$

Proof. By Theorem 2.2 of [9], under assumption (1), we have, as $m \rightarrow \infty$,

$$\frac{\Gamma_{j,\ell_m}^m}{m} \xrightarrow{P_m^{(0)}} \gamma_j^{(0)}(\mu), \quad j = 1, \dots, N, \tag{12}$$

where the functions $(\gamma_j^{(0)}(\mu), \mu \in (0, \theta_\infty(\beta^{(0)})), 0 \leq j \leq N)$, are strictly positive and continuous in μ .

It was shown in Section 6.2 of [9] that the limit of the scaled Hessian in Lemma 2 evaluated at the true parameter has the integral representation

$$J^{(0)}(\mu) = J(\beta^{(0)}, \mu) = \int_0^\mu Q^{(0)}(\lambda) d\lambda, \tag{13}$$

where

$$Q^{(0)}(\lambda) = \left[\frac{\gamma_i^{(0)}(\lambda)}{\beta_i^{(0)} Z(\beta^{(0)}, \lambda)} \delta_{ij} - \frac{\gamma_i^{(0)}(\lambda) \gamma_j^{(0)}(\lambda)}{Z^2(\beta^{(0)}, \lambda)} \right]_{i,j=1}^N, \tag{14}$$

δ_{ij} is the Kroneker symbol, and

$$Z(\beta, \lambda) = \gamma_0^{(0)}(\lambda) + \sum_{i=1}^N \beta_i \gamma_i^{(0)}(\lambda).$$

Let us show that if $i = i_m$ is such that $i/m \rightarrow \lambda \in (0, \mu)$ as $m \rightarrow \infty$ then

$$E_m^{(0)}((\eta_i^m)^2) \rightarrow \mathbf{a}^\top Q^{(0)}(\lambda) \mathbf{a}$$

as $m \rightarrow \infty$. Indeed,

$$\begin{aligned} E_m^{(0)}((\eta_i^m)^2) &= \sum_{k,j=1}^N a_k a_j E_m^{(0)}(\xi_{k,i}^m \xi_{j,i}^m) \\ &= \sum_{k,j=1}^N \frac{a_k a_j}{\beta_k^{(0)} \beta_j^{(0)}} E_m^{(0)}((\xi_{k,i}^m - \bar{\xi}_{k,i}^m)(\xi_{j,i}^m - \bar{\xi}_{j,i}^m)) \\ &= \sum_{k,j=1}^N \frac{a_k a_j}{\beta_k^{(0)} \beta_j^{(0)}} E_m^{(0)}(\xi_{k,i}^m \xi_{j,i}^m - \xi_{k,i}^m \bar{\xi}_{j,i}^m - \xi_{j,i}^m \bar{\xi}_{k,i}^m + \bar{\xi}_{k,i}^m \bar{\xi}_{j,i}^m). \end{aligned}$$

Note that

$$E_m^{(0)}(\xi_{k,i}^m \xi_{j,i}^m) = E_m^{(0)}(\xi_{k,i}^m) \delta_{kj} = E_m^{(0)}(\bar{\xi}_{k,i}^m) \delta_{kj},$$

where δ_{ij} is the Kroneker symbol and

$$E_m^{(0)}(\xi_{k,i}^m \bar{\xi}_{j,i}^m) = E_m^{(0)}(\xi_{j,i}^m \bar{\xi}_{k,i}^m) = E_m^{(0)}(\bar{\xi}_{k,i}^m \bar{\xi}_{j,i}^m).$$

Therefore,

$$E_m^{(0)}((\eta_i^m)^2) = \sum_{k,j=1}^N \frac{a_k a_j}{\beta_k^{(0)} \beta_j^{(0)}} [E_m^{(0)}(\bar{\xi}_{k,i}^m) \delta_{kj} - E_m^{(0)}(\bar{\xi}_{k,i}^m \bar{\xi}_{j,i}^m)].$$

By (6) and (12),

$$\bar{\xi}_{r,i}^m = \frac{\beta_r^{(0)} \Gamma_{r,i-1}^m}{\Gamma_{0,i-1}^m + \sum_{j=1}^N \beta_j^{(0)} \Gamma_{j,i-1}^m} \rightarrow \frac{\beta_r^{(0)} \gamma_r^{(0)}(\lambda)}{\gamma_0^{(0)}(\lambda) + \sum_{j=1}^N \beta_j^{(0)} \gamma_j^{(0)}(\lambda)} = \frac{\beta_r^{(0)} \gamma_r^{(0)}(\lambda)}{Z(\beta^{(0)}, \lambda)} \tag{15}$$

in $P_m^{(0)}$ -probability as $i/m \rightarrow \lambda$ for any $r = 0, \dots, N$. This fact along with Lemma 4 yields

$$\frac{1}{\beta_k^{(0)} \beta_j^{(0)}} [E_m^{(0)}(\bar{\xi}_{k,i}^m) \delta_{kj} - E_m^{(0)}(\bar{\xi}_{k,i}^m \bar{\xi}_{j,i}^m)] \rightarrow Q_{kj}^{(0)}(\lambda)$$

as $i/m \rightarrow \lambda$. We can then complete the proof of Proposition 1 by applying the dominated convergence theorem to show that the sum converges to the integral (see Section 5.2 of [9] for a similar argument).

Proposition 2. *Under Assumption 1,*

$$\lim_{m \rightarrow \infty} \frac{1}{m^2} \text{var} \left(\sum_{i=2}^{\ell_m} (\eta_i^m)^2 \right) = 0,$$

where the expectation is taken with respect to the measure $P_m^{(0)}$.

Proof. To simplify notation, we assume in the proof that $N = 1$; modifications for the multivariate case are obvious. Also, we omit the upper index in the notation for the η , ζ , and ξ variables. So, in the rest of the proof we define $\beta = \beta_1$, $a = \mathbf{a} \in \mathbb{R}$, $\eta_i = \eta_i^m$, $\zeta_i = \zeta_{1,i}^m$, $\xi_i = \xi_{1,i}^m$, $\bar{\xi}_i = \bar{\xi}_{1,i}^m$, and $\mathcal{F}_j = \mathcal{F}_j^{(m)}$. Besides, we write E instead of $E_m^{(0)}$.

It suffices to show that, under Assumption 1,

$$\text{cov}(\eta_i^2, \eta_j^2) \rightarrow 0$$

for any pair of sequences $i = i_m$ and $j = j_m$ such that $i \neq j$ and $i/m \rightarrow \lambda'$, $j/m \rightarrow \lambda''$ as $m \rightarrow \infty$, where λ' can coincide with λ'' . This suffices because the contribution from terms with $i = j$, divided by m^2 , is asymptotically negligible since the η_i are uniformly bounded.

Recall that $\eta_i = a(\xi_i - \bar{\xi}_i)/\beta$, where $\bar{\xi}_i = E(\xi_i | \mathcal{F}_{i-1})$. Therefore, we need to prove that

$$\text{cov}((\xi_i - \bar{\xi}_i)^2, (\xi_j - \bar{\xi}_j)^2) \rightarrow 0 \tag{16}$$

under the same assumptions about the index sequences. Assuming for definiteness that $i < j$, we have the following identities:

$$\xi_i^2 = \xi_i, \tag{17}$$

$$\begin{aligned} E(\xi_i) &= E(E(\xi_i \mid \mathcal{F}_{i-1})) = E(\bar{\xi}_i), \\ E(\xi_i \bar{\xi}_i) &= E(\bar{\xi}_i E(\xi_i \mid \mathcal{F}_{i-1})) = E(\bar{\xi}_i^2), \end{aligned} \tag{18}$$

$$E(f(\xi_i, \bar{\xi}_i, \bar{\xi}_j) \xi_j) = E(f(\xi_i, \bar{\xi}_i, \bar{\xi}_j) E(\xi_j \mid \mathcal{F}_{j-1})) = E(f(\xi_i, \bar{\xi}_i, \bar{\xi}_j) \bar{\xi}_j). \tag{19}$$

Here $f(\xi_i, \bar{\xi}_i, \bar{\xi}_j)$ is a polynomial function, e.g. $\xi_i \bar{\xi}_i^2$ etc. (Note that (19) fails for $i = j$.) We can write $\text{cov}((\xi_i - \bar{\xi}_i)^2, (\xi_j - \bar{\xi}_j)^2)$ as a linear combination of terms of the form

$$E(\bar{\xi}_i^p \xi_i^{2-p} \bar{\xi}_j^q \xi_j^{2-q}) - E(\bar{\xi}_i^p \xi_i^{2-p}) E(\bar{\xi}_j^q \xi_j^{2-q}), \tag{20}$$

where $p \in \{0, 1, 2\}$ and $q \in \{0, 1, 2\}$. As mentioned before (see (15)), we have

$$\bar{\xi}_r \xrightarrow{P_m^{(0)}} b(\lambda) := \frac{\beta_1^{(0)} \gamma_1^{(0)}(\lambda)}{Z(\beta^{(0)}, \lambda)} \tag{21}$$

as $r/m \rightarrow \lambda$, and, also, $E(\xi_r) \rightarrow b(\lambda)$ as $n \rightarrow \infty$. Since ξ_i and ξ_j are bounded, we have $E(\bar{\xi}_i^2) \rightarrow b^2(\lambda')$ and (using (18)) $E(\xi_i \bar{\xi}_i) \rightarrow b^2(\lambda')$, while (using (17)) $E(\xi_i^2) \rightarrow b(\lambda')$, and likewise for j . Therefore,

$$E(\bar{\xi}_i^p \xi_i^{2-p}) E(\bar{\xi}_j^q \xi_j^{2-q}) \rightarrow b^{1+\min(p,1)}(\lambda') b^{1+\min(q,1)}(\lambda''). \tag{22}$$

But using (19), (21), and Lemma 4, we also have

$$E(\bar{\xi}_i^p \xi_i^{1-p} \bar{\xi}_j \xi_j) = E(\bar{\xi}_i^p \xi_i^{1-p} \bar{\xi}_j^2) \rightarrow b^{1+\min(p,1)}(\lambda') b^2(\lambda''), \tag{23}$$

and using (17) for j , (19), (21), and Lemma 4, we also have

$$E(\bar{\xi}_i^p \xi_i^{1-p} \xi_j^2) = E(\bar{\xi}_i^p \xi_i^{1-p} \bar{\xi}_j) \rightarrow b^{1+\min(p,1)}(\lambda') b(\lambda''). \tag{24}$$

Combining (23) and (24) shows that $E(\bar{\xi}_i^p \xi_i^{1-p} \bar{\xi}_j^q \xi_j^{1-q})$ converges to the same limit as the expression in (22). Hence, each expression of the form in (20) tends to 0, and we have established (16). Hence, Proposition 2 is proved.

4. Structure of the limit information matrix

It is easy to see from (13) that positive definiteness of the matrix $Q^{(0)}(\lambda) = Q(\beta^{(0)}, \lambda)$ for any fixed $\lambda \in (0, \theta_\infty)$ implies positive definiteness of the limit matrix $J^{(0)}(\mu)$. Positive definiteness of the matrix $Q^{(0)}(\lambda)$ was shown in Lemma 5.2 of [9]. Here we give another proof by studying the matrix structure in more detail.

It can be seen from (14) that the matrix principal minor formed by the intersection of the first k rows and k columns is

$$\begin{aligned} & D_{N,k}(\beta^{(0)}, \lambda) \\ &= \begin{pmatrix} \frac{\gamma_1^{(0)}(\lambda)(Z(\beta^{(0)}, \lambda) - \gamma_1^{(0)}(\lambda)\beta_1^{(0)})}{\beta_1^{(0)} Z^2(\beta^{(0)}, \lambda)} & \cdots & -\frac{\gamma_1^{(0)}(\lambda)\gamma_k^{(0)}(\lambda)}{Z^2(\beta^{(0)}, \lambda)} \\ \vdots & \cdots & \vdots \\ -\frac{\gamma_1^{(0)}(\lambda)\gamma_k^{(0)}(\lambda)}{Z^2(\beta^{(0)}, \lambda)} & \cdots & \frac{\gamma_k^{(0)}(\lambda)(Z(\beta^{(0)}, \lambda) - \gamma_k^{(0)}(\lambda)\beta_k^{(0)})}{\beta_k^{(0)} Z^2(\beta^{(0)}, \lambda)} \end{pmatrix}. \end{aligned}$$

It is easy to see that the determinant of $D_{N,k}(\beta^{(0)}, \lambda)$ is

$$|D_{N,k}(\beta^{(0)}, \lambda)| = \frac{(-1)^k}{Z^{2k}(\beta^{(0)}, \lambda)} \prod_{i=1}^k \frac{\gamma_i^{(0)}(\lambda)}{\beta_i^{(0)}} |A_k - Z(\beta^{(0)}, \lambda)E_k|,$$

where $|A_k - Z(\beta^{(0)}, \lambda)E_k|$ is the determinant of the matrix $A_k - Z(\beta^{(0)}, \lambda)E_k$, with the matrix A_k defined as

$$A_k = (\beta_1^{(0)}, \dots, \beta_k^{(0)})(\gamma_1^{(0)}(\lambda), \dots, \gamma_k^{(0)}(\lambda))^T$$

and E_k a $k \times k$ unit matrix. By definition, $|A_k - Z(\beta^{(0)}, \lambda)E_k|$ is the characteristic polynomial of A_k evaluated at the point $Z(\beta^{(0)}, \lambda)$. It can be shown (we omit the proof) that if $a, b \in \mathbb{C}^n$ are nonzero complex vectors, such that $a^T b \neq 0$, then a quadratic matrix $M = ab^T$ has the only nonzero eigenvalue $a^T b$ of multiplicity 1; 0 is the other matrix eigenvalue of multiplicity $n - 1$ and the matrix characteristic polynomial is

$$|M - uE_n| = (-1)^n u^{n-1}(u - a^T b), \quad u \in \mathbb{C}^n.$$

Hence,

$$|A_k - Z(\beta^{(0)}, \lambda)E_k| = (-1)^k Z^{k-1}(\beta^{(0)}, \lambda) \left(\gamma_0^{(0)}(\lambda) + \sum_{i=k+1}^N \beta_i^{(0)} \gamma_i^{(0)}(\lambda) \right)$$

and

$$|D_{N,k}(\beta^{(0)}, \lambda)| = \frac{(\gamma_0^{(0)}(\lambda) + \sum_{i=k+1}^N \beta_i^{(0)} \gamma_i^{(0)}(\lambda))}{Z^{k+1}(\beta^{(0)}, \lambda)} \prod_{i=1}^k \frac{\gamma_i^{(0)}(\lambda)}{\beta_i^{(0)}}.$$

The right-hand side of the preceding display is positive because the functions $\gamma_i, i = 1, \dots, N$, are positive. Thus, any principal minor of matrix (14) is positive and, by Sylvester’s criterion, this matrix is positive definite.

5. Numerical example

This section contains a simulated example illustrating the effectiveness of MLE in distinguishing between CSAs which might generate rather similar patterns.

We consider two single realisations of CSA on the unit square $[0, 1]^2$, one with increasing rates and one with flat rates. In both cases, we take the interaction radius to be $R = 0.02$, and we take $N = 2$ and $\beta_0 = 1$. The parameters (β_1, β_2) are given by

$$(\beta_1, \beta_2) = (300, 500) \quad \text{for increasing rates,} \quad (\beta_1, \beta_2) = (100, 100) \quad \text{for flat rates.}$$

We shall give confidence intervals for the estimated parameters with $\ell = 200, 500, 1000, 2000$, and 3000 points, respectively. In the cases $\ell = 500$ and $\ell = 2000$ we also provide images in Figures 2 and 3. (Images for the other cases of ℓ are included in a preprint version of this paper, available at <http://arxiv.org/abs/1005.2335v1>.) Can one tell apart these two sets of parameters given the images provided?

In the images with 500 points, it is slightly visible that the right pattern is more dispersed than the left pattern. This effect becomes somewhat more visible for the pair of images showing further evolution and containing 2000 points. Though the main basic feature of *clustering* is common to both choices of parameters, the clustering effect is more visible in the images

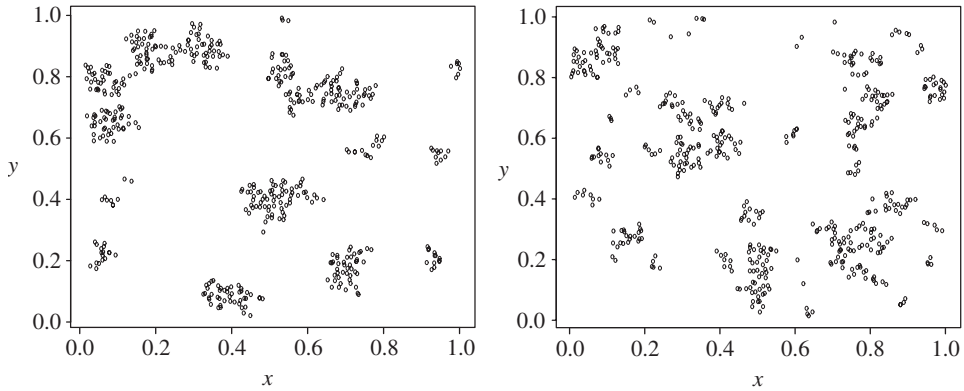


FIGURE 2: Realisations generated with $\ell = 500$ points: increasing rates (left); flat rates (right).

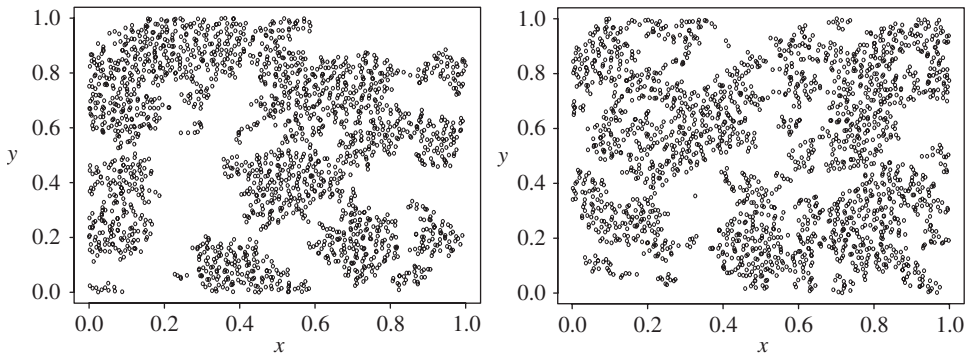


FIGURE 3: Realisations generated with $\ell = 2000$ points: increasing rates (left); flat rates (right).

produced by the model with increasing rates. It seems that the right realisation spreads faster in comparison to the left realisation. This is called *mild clustering*; the distribution of points inside a cluster is more or less regular, since a new point distribution is uniform conditioned on being adsorbed in the vicinity of existing points.

Tables 1 and 2 contain MLEs for both sets of parameters along with corresponding approximate confidence intervals (any computed value is rounded to its nearest integer). The 95% confidence intervals (CIs) are computed by *formally assuming* normality of $\hat{\beta}$. The variances of the estimates are estimated, as usual, by the corresponding diagonal elements of the matrix inverse to the observed information matrix, divided by m . To get this matrix, we differentiate (5) and then evaluate at $\hat{\beta}$ using the observed t - and Γ -statistics. We provide values of the t -statistics in the tables as well. Arrays of Γ -statistics were computed as described in [9] (see Section 2.2 therein).

The resulting information matrices turned out to be nondegenerate for all values of ℓ considered. The variances of the estimates decrease as the number of observed points increase. As a result, the CIs become narrower. This observed reduction of variances is intuitively expected, and consistent with the limiting situation where the effect is clearly implied by the integral representation (13) for the information matrix. For instance, in the case of a single parameter ($N = 1$) this representation implies that m times the variance of the estimate $\hat{\beta}_1$

TABLE 1: MLEs, CIs, and t -statistics for a set of increasing rates ($\beta_1 = 300$ and $\beta_2 = 500$).

	$\ell = 200$	$\ell = 500$	$\ell = 1000$	$\ell = 2000$	$\ell = 3000$
$\hat{\beta}_1$	401	377	334	320	318
CI	(176, 626)	(214, 540)	(213, 455)	(218, 422)	(223, 413)
$\hat{\beta}_2$	695	594	566	546	521
CI	(298, 1091)	(335, 853)	(360, 772)	(371, 721)	(364, 678)
(t_0, t_1, t_2)	(16, 93, 91)	(25, 233, 242)	(34, 434, 532)	(43, 825, 1132)	(47, 1190, 1763)

TABLE 2: MLEs, CIs, and t -statistics for a set of flat rates ($\beta_1 = 100$ and $\beta_2 = 100$).

	$\ell = 200$	$\ell = 500$	$\ell = 1000$	$\ell = 2000$	$\ell = 3000$
$\hat{\beta}_1$	89	98	96	104	101
CI	(55, 123)	(69, 127)	(73, 119)	(81, 127)	(80, 122)
$\hat{\beta}_2$	106	97	88	100	99
CI	(61, 151)	(67, 127)	(66, 110)	(78, 122)	(78, 120)
(t_0, t_1, t_2)	(43, 100, 57)	(62, 272, 166)	(84, 552, 364)	(95, 1048, 857)	(106, 1473, 1421)

converges, as $m \rightarrow \infty$ and $\ell_m/m \rightarrow \mu$, to $(\int_0^\mu Q^{(0)}(\lambda) d\lambda)^{-1}$, where $Q^{(0)}(\lambda) > 0$ (the matrix is just a number here) in representation (13). This is consistent with the expected ‘reduction of variances’ effect, if the density of points, i.e. μ , increases (a lower bound for the variance of the estimate in the single parameter case is given by the integral with $\mu = \theta_\infty$, where θ_∞ is the jamming density).

Finally, it should be noted that MLEs effectively capture the correct magnitude of the parameters and this is why the two sets of parameters considered in the example (producing sometimes quite similar images) can be effectively distinguished. For the sake of completeness, consider also the left image in Figure 1. It has been generated by CSA with interaction radius 0.01, $N = 2$, and $(\beta_0, \beta_1, \beta_2) = (1, 1000, 10000)$. The image contains $\ell = 1000$ points and the t -statistics are $t_0 = 23$, $t_1 = 149$, and $t_2 = 828$. The MLEs for β_1 and β_2 are 1105.0 and 10 510.0, respectively.

6. Final remarks

It should be noted that estimation of both N and the β s requires the interaction radius R to be known or estimated. If the radius is unknown then it can be estimated by the method of profile likelihood estimation [8, p. 163]. It should also be noted that estimation of the interaction radius is a particular case of a problem of estimation of the interaction range in models with spatial interaction (see [4] and [6]).

Recall that the observation window D is assumed to be convex. This assumption is not important and made mostly for definiteness and simplicity of presentation. At the same time it is easy to imagine an application where this assumption cannot be made. For instance, if CSA is applied for modelling the spread of infection in a nonconvex geographical region. In such a case the model dynamics should be modified as follows. Given a nonconvex region D , call two points of D *geographical* neighbours if they are within distance R of each other and the straight

line segment connecting them belongs entirely to D . It would be plausible to assume that the probability distribution of a new point is specified by a function depending on the number of geographical neighbours. Note that this alteration of the model does not change the proposed statistical analysis of the model.

Finally, note that the main proofs here and in [9] can be carried out in the same way for any spatial-temporal model which has a likelihood structure similar to that of CSA. Namely, the likelihood must depend on a point configuration through statistics which are locally determined functionals [9]. For example, one could assume that the sequence (β_k) is eventually constant, rather than eventually 0 as here.

Acknowledgements

This research was supported by the Royal Society of London, Travel for Collaboration Grant, award reference code RC-MA1038.

The authors would like to thank the anonymous referees for helpful remarks and suggestions.

References

- [1] BEIL, M. *et al.* (2009). Simulating the formation of keratin filament networks by a piecewise-deterministic Markov process. *J. Theoret. Biol.*, **256**, 518–532.
- [2] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. John Wiley, New York.
- [3] EVANS, J. W. (1993). Random and cooperative sequential adsorption. *Rev. Mod. Phys.* **65**, 1281–1329.
- [4] GALVES, A., ORLANDI, E. AND TAKAHASHI, D. Y. (2010). Identifying interacting pairs of sites in infinite range Ising models. Preprint. Available at <http://arxiv.org/abs/1006.0272v2>.
- [5] LEHMANN, E. L. (1983). *Theory of Point Estimation*. John Wiley, New York.
- [6] LÖCHERBACH, E. AND ORLANDI, E. (2011). Neighborhood radius estimation for variable-neighborhood random fields. Preprint. Available at <http://arxiv.org/abs/1002.4850v5>.
- [7] MCLEISH, D. L. (1974). Dependent central limit theorems and invariance principles. *Ann. Prob.* **2**, 620–628.
- [8] MØLLER, J. AND WAAGEPETERSEN, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton, FL.
- [9] PENROSE, M. D. AND SHCHERBAKOV, V. (2009). Maximum likelihood estimation for cooperative sequential adsorption. *Adv. Appl. Prob.* **41**, 978–1001.
- [10] RAFELSKI, S. M. AND MARSHALL, W. F. (2008). Building the cell: design principles of cellular architecture. *Nature Rev. Mol. Cell Biol.* **9**, 593–602.
- [11] SHCHERBAKOV, V. (2006). Limit theorems for random point measures generated by cooperative sequential adsorption. *J. Statist. Phys.* **124**, 1425–1441.
- [12] WINDOFFER, R., WÖLL, S., STRNAD, P. AND LEUBE, R. E. (2004). Identification of novel principles of keratin filament network turnover in living cells. *Mol. Biol. Cell* **15**, 2436–2448.