

MAXIMUM LIKELIHOOD ESTIMATION FOR COOPERATIVE SEQUENTIAL ADSORPTION

MATHEW D. PENROSE,* *University of Bath*

VADIM SHCHERBAKOV,** *University of Bath and Moscow State University*

Abstract

We consider a model for a time series of spatial locations, in which points are placed sequentially at random into an initially empty region of \mathbb{R}^d , and given the current configuration of points, the likelihood at location x for the next particle is proportional to a specified function β_k of the current number (k) of points within a specified distance of x . We show that the maximum likelihood estimator of the parameters β_k (assumed to be zero for k exceeding some fixed threshold) is consistent in the thermodynamic limit where the number of points grows in proportion to the size of the region.

Keywords: Cooperative sequential adsorption; space–time point process; maximum likelihood estimation; thermodynamic limit; increasing domain asymptotic; law of large numbers

2000 Mathematics Subject Classification: Primary 62M30; 60K35

Secondary 60D05

1. Introduction

Cooperative sequential adsorption (CSA) models are a class of probabilistic models for sequential packing and deposition (see [11] and the references therein), which are briefly described as follows. Points are placed sequentially at random in a bounded region of Euclidean space. Given the configuration of points placed so far, the likelihood (probability density) for the next point to be placed at location x is a specified function of the current configuration of points near x , divided by a normalising constant which depends on the current global configuration.

CSA is widely used in physics and chemistry for modelling various irreversible adsorption processes, such as chemisorption on single-crystal surfaces or adsorption in colloidal systems (see [3] or [12]). A special case of CSA known as random sequential adsorption (RSA, to be described later) serves as a benchmark in modelling irreversible phenomena in physical chemistry and other applications (see [2] and the references therein). It has been noted in [3, p. 1285] that the lattice CSA model ‘can be useful for describing the spatial features of the irreversible spread of disease epidemics’. The continuum version of CSA considered here seems to be more technically convenient for modelling sequential point patterns in disciplines such as geophysics, biology, and ecology in situations where a data set is presented by a sequential or ordered point pattern, i.e. a collection of *spatial events* which appear sequentially.

For instance, CSA could be used for space–time models of earthquake occurrences in geophysics (see [1], [17], and the references therein for existing alternative models where

Received 6 May 2008; revision received 20 April 2009.

* 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, 119991, Moscow, Russia.

Email address: v.shcherbakov@mech.math.msu.su

the underlying point process is a Poisson process), for spatial spread of an infection in ecology or a material destruction in materials science. While it will not capture all the details of these natural phenomena, CSA can be used as an approximation of spatial spread dynamics; see the simulated images in Section 2 for an illustrative example.

We assume that the likelihood at location x is a function of the number of neighbours, i.e. the current number of points within a certain distance from x , divided by a normalising constant. In statistical physics, the lattice analogue of this model is known as *monomer filling with nearest neighbour cooperative effects* [3] (see also [13], where an asymptotic study of the model in continuum is undertaken under some assumptions). Besides, we assume a *hard-core-type* constraint on interparticle interaction, whereby the likelihood is zero at any location with more than a certain number of neighbours (e.g. in RSA the likelihood is zero at any location with one or more neighbours); one interpretation of this is that usually in practice the ‘defects’ represented by the points have strictly positive size, so that only finitely many defects can be distinguished in a bounded region.

It is well known in the physical literature that lattice CSA models are very flexible for modelling both clustered and regular point patterns (e.g. images on p. 1307 of [3]). It is not surprising that the class of CSA models under consideration inherits the same property (see, e.g. the example in Section 2). See also a discussion of the same property in [14] for the CSA point process, which is closely related to the CSA model (the CSA point process is a particular case of a sequential Markov point process [16] and can also be used for modelling sequential point patterns arising in applications). Also, this class of CSA models is easy to simulate and can be characterised by a finite number of parameters.

Fitting the model to real-life data necessarily requires estimation of the model parameters. The main goal of this paper is to justify statistical inference of these parameters based on maximum likelihood estimation.

As usual, maximum likelihood estimators (MLEs) are defined as maximisers of the model likelihood and can be found by solving MLE equations obtained by equating to zero the log-likelihood derivatives. We prove that, with probability tending to 1, there exists a *unique* solution to the MLE equations as the amount of observed information increases in the following natural sense. Namely, we consider a sequence of target regions (observation windows) expanding to the whole space and assume that the number of observed points in a region is linear in the volume of the region. This limit is known as the *thermodynamic limit* in the statistical physics literature, and known as the *increasing domain asymptotic framework* in the spatial statistics literature.

It turns out that the log-likelihood derivatives behave asymptotically similar to those in the independent and identically distributed (i.i.d.) case, and this allows us to adapt a classical argument from [6] to finish the proof of the existence of MLEs (our Theorem 2.2). We also prove a *consistency* result saying that MLEs converge in probability to the true parameters in the thermodynamic limit.

Our asymptotic analysis of MLEs is based on the observation that the MLE equations are determined by statistics of a special type, namely, sums of locally determined functionals over a configuration of points (see Section 3). This allows us to analyse the asymptotics of the MLE equations, exploiting the limit theory for random sequential packing and deposition developed in [11].

MLEs for the CSA model can be computed numerically by classical Monte Carlo methods, which is in contrast with Markov point process models for which Markov chain Monte Carlo (MCMC) approximation is required and, as a result, the MLE is replaced by the MCMCMLE

(see [4], [5], and [7, Section 9.1]). We are not aware of any results on statistical inference for CSA models, except [15], where estimation of the time horizon for continuous-time RSA is addressed.

2. The model and main results

2.1. The cooperative sequential adsorption model

Let R be a positive constant, and let $(\beta_k, k \geq 0)$ be a sequence of nonnegative numbers. For any point $x \in \mathbb{R}^d$ and any finite sequence $\mathbf{y} = (y_1, \dots, y_n), n \geq 1$, of points in \mathbb{R}^d , we denote by $\nu(x, \mathbf{y})$ the number of points y_i in the sequence \mathbf{y} , such that the Euclidean distance between x and y_i is not greater than R . By definition, $\nu(x, \emptyset) = 0$.

Given a compact convex subset D of \mathbb{R}^d (which we denote the *target region*), let $(Y_i, i \geq 1)$ be a sequence of independent random points uniformly distributed in D , and construct another sequence of random points by accepting each point of the original sequence with a certain probability to be described below, otherwise rejecting that point. Let $X(A(n)) = (X_1, \dots, X_{A(n)})$ be the sequence of accepted points from the finite sequence $Y_i, i = 1, \dots, n$. Here $A(n)$ denotes the number of points in this sequence that are accepted. By definition, $X(0) = \emptyset$.

The point Y_{n+1} is accepted with probability $\beta_{\nu(Y_{n+1}, X(A(n)))}/K$, where K is an arbitrary constant such that $\max_{0 \leq i \leq A(n)} \beta_i \leq K$. Regardless of the particular choice of K , the next accepted point $X_{A(n)+1}$ has the following probability density:

$$\psi_{A(n)+1}(x) = \frac{\beta_{\nu(x, X(A(n)))}}{\int_D \beta_{\nu(u, X(A(n)))} \, du}. \tag{2.1}$$

In other words, given the sequence $X(\ell)$ of the first ℓ accepted points, the next accepted point $X_{\ell+1}$ is sampled from a distribution which is specified by the unnormalised probability density $\beta_{\nu(x, X(\ell))}, x \in D$ (the value of K influences only the number of discarded points Y_i until the next acceptance). The probability distribution of the finite random sequence $X(\ell) = (X_1, \dots, X_\ell)$ is specified by the following probability density:

$$p_{\ell, \beta, D}(x_1, \dots, x_\ell) = \prod_{k=1}^{\ell} \frac{\beta_{\nu(x_k, x^{(k-1)})}}{\int_D \beta_{\nu(u, x^{(k-1)})} \, du}, \quad x_i \in D, i = 1, \dots, \ell, \tag{2.2}$$

where we denote, for short, $x^{(k)} = (x_1, \dots, x_k), k \geq 1$, and $x^{(0)} = \emptyset$ for $k = 0$.

We assume that β_k is zero for all but finitely many k (this is our hard-core-type constraint); in fact, we assume that there exists a fixed positive integer N such that

$$\beta_k > 0, \quad 0 \leq k \leq N; \quad \beta_k = 0, \quad k \geq N + 1.$$

The interaction radius R is assumed to be a *fixed* and *known* constant. Note that the density (2.1) (and, hence, (2.2)) is unaffected by multiplication of all parameters β_k by a constant. Therefore, for the sake of identifiability, we assume that

$$\beta_0 = 1. \tag{2.3}$$

Thus, the model is parametrised by parameters $\beta = (\beta_1, \dots, \beta_N)$, where N is also regarded as a parameter. RSA is the special case with $N = 0$.

Set

$$t_k(x(\ell)) := \sum_{i=1}^{\ell} \mathbf{1}_{\{\nu(x_i, x^{(i-1)})=k\}}, \quad k = 0, \dots, N, \tag{2.4}$$

where $\mathbf{1}_A$ stands for the indicator of set A . Then, for $(x_1, \dots, x_\ell) \in D^\ell$, using assumption

(2.3) and adopting the convention that $0^0 = 1$, we can rewrite (2.2) as

$$p_{\ell, \beta, D}(x_1, \dots, x_\ell) = \frac{\prod_{j=1}^{\ell} \beta_j^{t_j(x(\ell))}}{\prod_{k=1}^{\ell} \int_D \beta_v(u, x(k-1)) \, du}. \tag{2.5}$$

The images in Figures 1 and 2 are of a single simulated realisation of the CSA model in $D = (0, 1)^2$ with parameters $R = 0.04, \beta_0 = 1, \beta_1 = 100, \beta_2 = 300, \beta_3 = 500$, and $\beta_k = 0, k \geq 4$, showing successively the first 50, 100, 200, and 300 points.

Small point clusters are visible in the left image of Figure 1. These clusters might be interpreted as original centres of an infection (for instance). The infection develops in time, with events appearing sequentially, tending to appear in the vicinity of existing points, but they

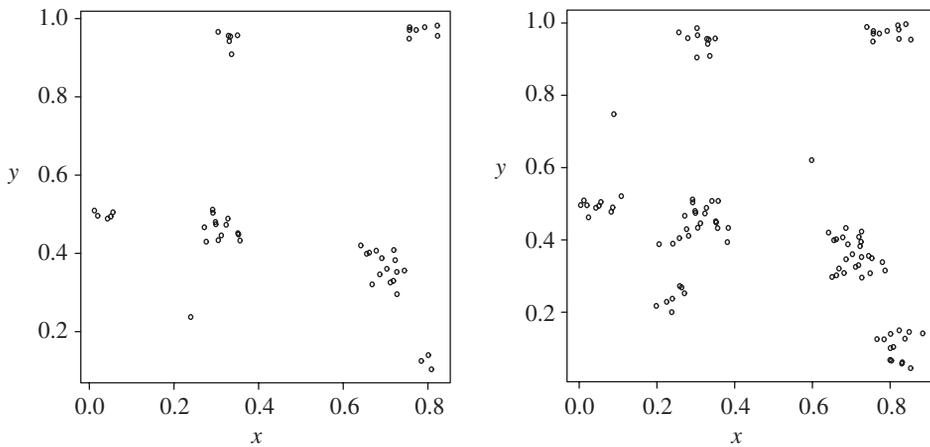


FIGURE 1: Images of the simulated CSA model with $D = (0, 1)^2, R = 0.04, N = 3$, and $\beta = (100, 300, 500)$. *Left:* $\ell = 50$. *Right:* $\ell = 100$.

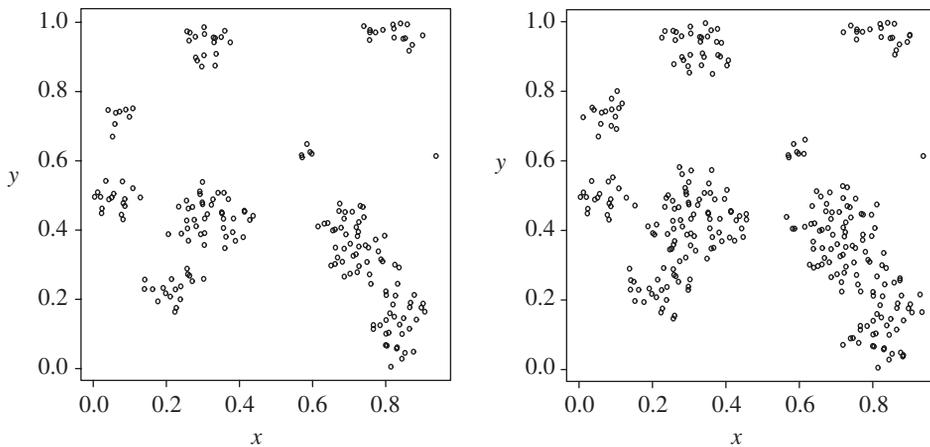


FIGURE 2: Images of the simulated CSA model with $D = (0, 1)^2, R = 0.04, N = 3$, and $\beta = (100, 300, 500)$. *Left:* $\ell = 200$. *Right:* $\ell = 300$.

can appear in unaffected areas as well; see the right image in Figure 1. The images in Figure 2 show further evolution, where the point clusters grow and begin to coalesce. Besides, it is rather visible that the clusters become more saturated (or dense). Therefore, an increasing sequence of parameters results in a combination of spreading and a *strong* clustering effect ('clusters inside clusters'). While the point density cannot increase indefinitely because of the ultimate hard-core constraints, it is easy to imagine that if the parameters increase sufficiently fast then clustering will prevail in the sense that clusters will be internally nearly saturated. For instance, if the parameters increase geometrically then it would correspond to the lattice CSA model with Aarhenius rates [3].

It is interesting to compare the images in Figures 1 and 2 with the images in Figures 3 and 4, which are of a single simulated realisation of the CSA model in $D = (0, 1)^2$ with parameters

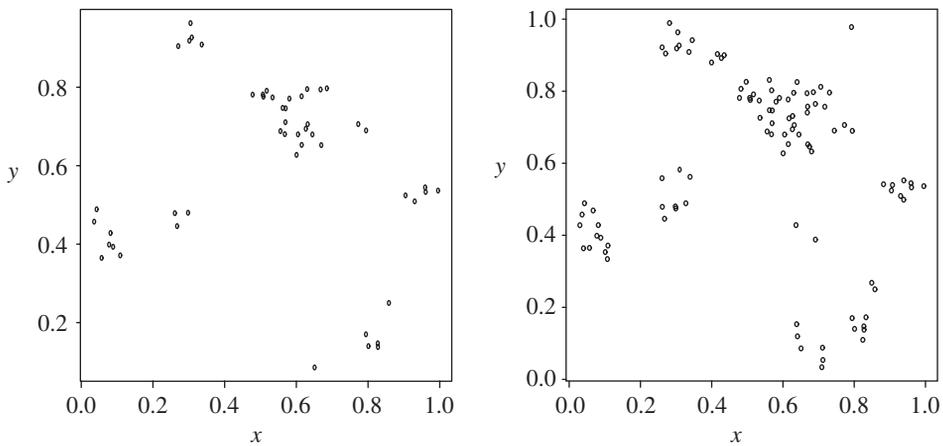


FIGURE 3: Images of the simulated CSA model with $D = (0, 1)^2$, $R = 0.04$, $N = 3$, and $\beta = (100, 100, 100)$. *Left:* $\ell = 50$. *Right:* $\ell = 100$.

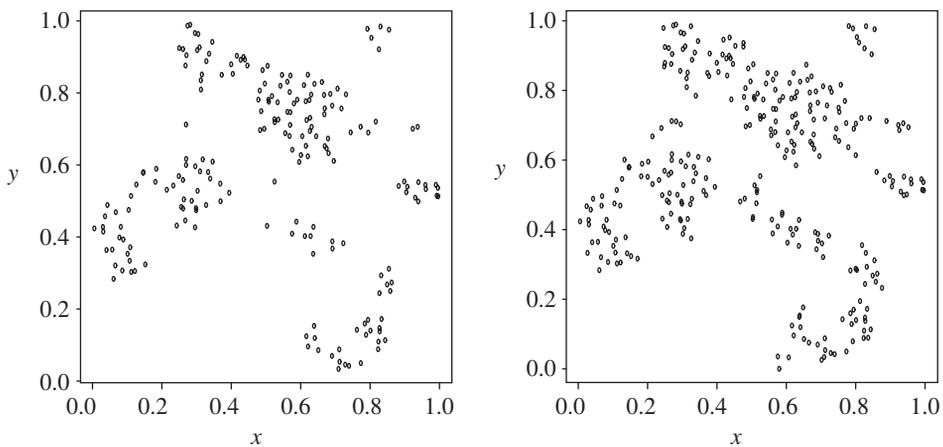


FIGURE 4: Images of the simulated CSA model with $D = (0, 1)^2$, $R = 0.04$, $N = 3$, and $\beta = (100, 100, 100)$. *Left:* $\ell = 200$. *Right:* $\ell = 300$.

$R = 0.04$, $\beta_0 = 1$, $\beta_1 = \beta_2 = \beta_3 = 100$, and $\beta_k = 0$, $k \geq 4$, showing successively the first 50, 100, 200, and 300 points. This choice of a flat sequence of parameters corresponds to the lattice CSA model with so-called Eden rates [3]. It is not surprising that the spatial dynamics are similar to those observed in the previous example. The only essential difference is that now a newly born cluster tends to spread rather than get saturated. We can call this effect *mild* clustering. The distribution of points inside a cluster is more or less regular, since the point distribution is uniform conditioned on being adsorbed in the vicinity of existing points.

The effects that have just been described for the continuous CSA model under consideration are rather straightforward analogues of the phenomena of ‘competition between the birth, growth and coalescence’ (see [3, p. 1307]), which were already well known for lattice CSA models.

2.2. Maximum likelihood estimation

For $\ell \geq 2$, define the log-likelihood function

$$L_D(X(\ell), \beta) := \log(p_{\ell, \beta, D}(X_1, \dots, X_\ell)).$$

For any $0 \leq k \leq \ell - 1$ and $j \geq 0$, set

$$\Gamma_j(X(k)) := \int_D \mathbf{1}_{\{u: v(u, X(k))=j\}} du.$$

Note that $\Gamma_j(X(k))$ also depends on D . It is easy to see that

$$\int_D \beta_{v(u, X(k))} du = \sum_{j=0}^k \beta_j \Gamma_j(X(k)) = \Gamma_0(X(k)) + \sum_{j=1}^{\infty} \beta_j \Gamma_j(X(k)),$$

that $\Gamma_j(X(k)) = 0$ for $k < j$, and that $\Gamma_0(\emptyset)$ is equal to the Lebesgue measure of D . Denote, for short,

$$t_{j, \ell} := t_j(X(\ell)), \quad j \geq 1, \quad \Gamma_{j, k} := \Gamma_j(X(k)), \quad j \geq 0, \quad k = 0, \dots, \ell.$$

Also, define $\widehat{N} = \widehat{N}(X(\ell))$ by

$$\widehat{N}(X(\ell)) := \max_{X_i \in X(\ell)} v(X_i, X(i - 1)) = \max\{j : t_{j, \ell} > 0\}.$$

Then, by (2.5), using the convention that $0 \log 0 = 0$, we have

$$\begin{aligned} L_D(X(\ell), \beta) &= \sum_{j=1}^{\infty} t_{j, \ell} \log(\beta_j) - \sum_{k=1}^{\ell} \log\left(\Gamma_{0, k-1} + \sum_{j=1}^{\infty} \beta_j \Gamma_{j, k-1}\right) \\ &= \sum_{j=1}^{\widehat{N}} t_{j, \ell} \log(\beta_j) - \sum_{k=1}^{\ell} \log\left(\Gamma_{0, k-1} + \sum_{j=1}^{\infty} \beta_j \Gamma_{j, k-1}\right). \end{aligned}$$

This expression is clearly nonincreasing in β_j for $j > \widehat{N}$, and, therefore, we restrict attention to β with $\beta_j = 0$ for all $j > \widehat{N}$. For such β , we have

$$L_D(X(\ell), \beta) = \sum_{j=1}^{\widehat{N}} t_{j, \ell} \log(\beta_j) - \sum_{k=1}^{\ell} \log\left(\Gamma_{0, k-1} + \sum_{j=1}^{\widehat{N}} \beta_j \Gamma_{j, k-1}\right). \tag{2.6}$$

It is easy to see that \widehat{N} is the MLE of parameter N . Given observation $X(\ell)$ and computed \widehat{N} , we define the MLE

$$\widehat{\beta}(X(\ell)) = (\widehat{\beta}_1, \dots, \widehat{\beta}_{\widehat{N}}, 0, 0, \dots)$$

of the true parameter vector $(\beta_1^{(0)}, \dots, \beta_N^{(0)}, 0, 0, \dots)$ as the maximiser of the function $L_D(X(\ell), \beta)$ over vectors of the form $(\beta_1, \dots, \beta_{\widehat{N}}, 0, 0, \dots)$. The latter depends smoothly on $(\beta_1, \dots, \beta_{\widehat{N}})$; therefore, the maximiser $(\widehat{\beta}_1, \dots, \widehat{\beta}_{\widehat{N}})$ (provided it exists) is a solution to the following system of MLE equations:

$$\frac{\partial L_D(X(\ell), \beta)}{\partial \beta_j} = 0, \quad j = 1, \dots, \widehat{N}. \tag{2.7}$$

By (2.6), (2.7) can be rewritten as

$$\frac{t_{j,\ell}}{\beta_j} - \sum_{k=1}^{\ell} \frac{\Gamma_{j,k-1}}{\Gamma_{0,k-1} + \sum_{j=1}^{\widehat{N}} \beta_j \Gamma_{j,k-1}} = 0, \quad j = 1, \dots, \widehat{N}. \tag{2.8}$$

Clearly, $\widehat{N} \leq N$ almost surely, and if $\widehat{N}(X(\ell)) = N' < N$ then $t_{j,\ell} = 0$ for $N' + 1 \leq j \leq N$. It is also possible that $t_{j,\ell} = 0$ for some $j < \widehat{N}(X(\ell))$. Therefore, if an observed point pattern is not a ‘typical’ model pattern then we might not have sufficient information to estimate the full set of parameters. Nevertheless, we shall show that, with probability tending to 1 under the asymptotic regime to be considered in this paper (see Section 2.3), \widehat{N} equals N and all t -statistics are positive, so that the system of likelihood equations (2.8) is nondegenerate. Moreover, there exists a unique positive solution $(\widehat{\beta}_1, \dots, \widehat{\beta}_N)$ of the likelihood equations and this solution is consistent for estimating the true parameter vector $\beta^{(0)} = (\beta_1^{(0)}, \dots, \beta_N^{(0)})$.

Numerical results for the simulated images in Figure 1 were as follows: $\widehat{N} = 3$ and the t -statistics are positive for all images. For the left image (with $\ell = 50$), we obtained $\widehat{\beta} = (88.0, 374.0, 565.0)$ by computing the log-likelihood over the lattice cube $[40.0, 150.0] \times [150.0, 450] \times [400.0, 650.0]$ with mesh 1. The log-likelihood is equal to 131.657 at $\widehat{\beta}$, whereas it takes value 131.306 at the true parameter (the minimal value on the cube is 126.918). Similar computations for the right image (with $\ell = 100$) over the lattice cube $[50.0, 150.0] \times [200.0, 400.0] \times [450.0, 650.0]$ gave $\widehat{\beta} = (110.0, 319.0, 499.0)$. The maximal value taken at $\widehat{\beta}$ equals 247.527, whereas it is 247.494 at the true parameter. It should be noted that the likelihood surface is very flat in a rather large region containing both the maximiser and the true parameter. Therefore, the choice of a parameter is not so critical.

It would certainly be desirable to estimate the precision of the estimates via confidence intervals. This could be done using asymptotic normality of the MLE if one had a proof of this, and we hope to address this issue in future work.

Each of the Γ -statistics (defined by (2.4)) was computed as the proportion of 5×10^5 uniform random points in $(0, 1)^2$ which fall in the region determining the statistic.

Other methods of computing the Γ -statistics might be more accurate; for example, if we update the vector $(\Gamma_j(X(k)))_{j=0}^N$ sequentially in k then each new point X_{k+1} affects only the contribution to each component of this vector from inside the R -neighbourhood of $X(k)$ (cf. (4.14), below). We do not investigate such computational issues in depth here.

2.3. The asymptotic regime

Let D_1 denote the unit cube centred at the origin (though the asymptotic theory below will remain unchanged for any compact convex set $D_1 \subset \mathbb{R}^d$). Consider a sequence of rescaled domains $D_m = m^{1/d} D_1$, $m \in \mathbb{Z}_+$.

The *jamming density* is defined as follows. Given m , consider the CSA process as described in Section 2.1 with target region $D = D_m$. Denote by $A_m(n)$ the (random) number of points accepted out of the first n incoming points (this was called $A(n)$ in Section 2.1, but now it depends on the choice of D , which depends on m). For an *infinite* sequence of points arriving over D_m , the total number of accepted points remains bounded because we assume that $N < \infty$ (so the number of points accepted in any ball of radius $R/2$ is at most N). Therefore, the limit $j_m := \lim_{n \rightarrow \infty} A_m(n)$ exists almost surely, and is a finite random variable. Then it is known [8] that j_m/m converges in probability, as $m \rightarrow \infty$, to a constant (dependent on the parameters $R, N, \beta_1, \dots, \beta_N$), known as the *jamming density* and denoted by $\theta_\infty(\beta)$ in this paper. Actually, Theorem 7.1 of [8] is not quite applicable directly here, since the rule for the acceptance probability takes a slightly different form there. However, the method of proof of Theorem 7.1 of [8] carries through easily enough to the present setting. An alternative characterisation of the jamming density is given later (see Lemma 4.1).

Our asymptotics are as m becomes large, i.e. we consider the model in an *increasing domain asymptotic framework*. Fix $(\ell_m, m \geq 1)$, an arbitrary monotonically increasing sequence of positive numbers, satisfying Assumption 2.1, below. Given N and R , let the state space \mathcal{S}_m be defined by

$$\mathcal{S}_m := \{x(\ell_m) = (x_1, \dots, x_{\ell_m}) : x_i \in D_m, 1 \leq i \leq \ell_m\}.$$

Let $\mathcal{B} := (0, \infty)^N$, the set to which the parameter vector $\beta = (\beta_1, \dots, \beta_N)$ can belong. Given $\beta = (\beta_1, \dots, \beta_N) \in \mathcal{B}$, consider a probability measure $P_{m,\beta}$ (with corresponding expectation $E_{m,\beta}$) on \mathcal{S}_m specified by probability density (2.5) with $\ell = \ell_m$ and $D = D_m$. Note that, given m , if $\ell_m \geq N$ then the family of probability measures $P_{m,\beta'}$, $\beta \in \mathcal{B}$, is identifiable, i.e. the probability measures $P_{m,\beta'}$ and $P_{m,\beta''}$ are distinct, if $\beta' \neq \beta''$.

The true parameter is denoted by $\beta^{(0)} = (\beta_1^{(0)}, \dots, \beta_N^{(0)})$, with $\beta^{(0)} \in \mathcal{B}$. We write $P_m^{(0)}$ for $P_{m,\beta^{(0)}}$ and $E_m^{(0)}$ for $E_{m,\beta^{(0)}}$, and simply θ_∞ for $\theta_\infty(\beta^{(0)})$. If the $\xi_m, m \geq 1$, are random variables with ξ_m defined on the sample space \mathcal{S}_m for each m and $x \in \mathbb{R}$ is a constant, we write

$$\xi_m \xrightarrow{P_{m,\beta}} x \text{ if}$$

$$\lim_{m \rightarrow \infty} P_{m,\beta}\{|\xi_m - x| > \varepsilon\} = 0 \quad \text{for all } \varepsilon > 0.$$

Also, we write $\xrightarrow{P_m^{(0)}}$, for $\xrightarrow{P_{m,\beta^{(0)}}}$.

The following assumption on the sequence ℓ_m is similar to what is known as the *thermodynamic limit condition* in statistical physics and the *increasing domain asymptotic framework* in spatial statistics.

Assumption 2.1. *The number of observed points is asymptotically linear in m with coefficient below the jamming density, that is,*

$$\lim_{m \rightarrow \infty} \left(\frac{\ell_m}{m} \right) = \mu \in (0, \theta_\infty).$$

Given a target region D_m and an observation $X(\ell_m)$, we define

$$L_m(\beta) := L_{D_m}(X(\ell_m), \beta).$$

To emphasise dependence on the target region, we set

$$t_{j,\ell_m}^m := t_j(X(\ell_m)), \quad j = 1, \dots, N, \tag{2.9}$$

$$\Gamma_{j,k}^m := \Gamma_j^m(X(k)), \quad j = 0, \dots, N, k = 0, \dots, \ell_m. \tag{2.10}$$

Thus, (2.6) becomes

$$L_m(\beta) = \sum_{j=1}^{\widehat{N}} t_{j,\ell_m}^m \log(\beta_j) - \sum_{k=1}^{\ell_m} \log\left(\Gamma_{0,k-1}^m + \sum_{j=1}^{\widehat{N}} \beta_j \Gamma_{j,k-1}^m\right), \tag{2.11}$$

and (2.8) in target region D_m can be rewritten as

$$\frac{t_{j,\ell_m}^m}{\beta_j} - \sum_{k=1}^{\ell_m} \frac{\Gamma_{j,k-1}^m}{\Gamma_{0,k-1}^m + \sum_{j=1}^{\widehat{N}} \beta_j \Gamma_{j,k-1}^m} = 0, \quad j = 1, \dots, \widehat{N}. \tag{2.12}$$

2.4. Consistency of the MLE

The next result justifies why statistical inference for the CSA model can be based on the MLE.

Theorem 2.1. *Under Assumption 2.1, there exists a nonnegative function $(L(\beta, \mu), \beta \in \mathcal{B}, \mu \in (0, \theta_\infty))$ such that, for any fixed μ ,*

1. $L(\beta^{(0)}, \mu) = 0$ and $L(\beta, \mu) > 0$ for any $\beta \neq \beta^{(0)}$,
2. $L(\beta, \mu)$ is a continuous function of $\mu \in (0, \theta_\infty)$,
3. for any fixed $\beta \in \mathcal{B}$, as $m \rightarrow \infty$, we have

$$\frac{1}{m} (L_m(\beta^{(0)}) - L_m(\beta)) \xrightarrow{P_m^{(0)}} L(\beta, \mu).$$

Remark 2.1. It can be shown that $L_m(\beta) \sim m \log(m)$ as $m \rightarrow \infty$, whereas in the i.i.d. case the log-likelihood is linear in the number of points. Thus, in contrast with the i.i.d. case, we take the difference $L_m(\beta^{(0)}) - L_m(\beta)$ and prove convergence for this.

Theorem 2.1 immediately implies Corollary 2.1, below, which is analogous to a known result for the classical i.i.d. case (for instance, Theorem 2.1 in Chapter 6 of [6]).

Corollary 2.1. *Let $\beta = (\beta_1, \dots, \beta_N) \neq \beta^{(0)} = (\beta_1^{(0)}, \dots, \beta_N^{(0)})$. Then, under Assumption 2.1,*

$$P_m^{(0)} \left\{ \frac{p_{\ell_m, \beta^{(0)}, D_m}(X_1, \dots, X_{\ell_m})}{p_{\ell_m, \beta, D_m}(X_1, \dots, X_{\ell_m})} > 1 \right\} \rightarrow 1 \quad \text{as } m \rightarrow \infty.$$

The following theorem is the main result of the paper.

Theorem 2.2. *Under Assumption 2.1, with the $P_m^{(0)}$ -probability tending to 1 as $m \rightarrow \infty$,*

1. $\widehat{N} = N$, and
2. there exists a unique positive solution $(\widehat{\beta}_{1,m}, \dots, \widehat{\beta}_{N,m}) \in \mathcal{B}$ of the system of equations (2.12) such that, for $i = 1, \dots, N$,

$$\widehat{\beta}_{i,m} \xrightarrow{P_m^{(0)}} \beta_i^{(0)} \quad \text{as } m \rightarrow \infty.$$

3. Locally determined functionals

In this section we first give some general limit theorems (Theorems 3.1 and 3.2) for functionals defined on finite ordered sequences of points in \mathbb{R}^d by taking sums of so-called *locally determined functionals*. The sequence of points is taken to be a realisation of CSA, and the limit theorems are laws of large numbers as the target region D_m becomes large, which extend in various ways the previous work of Penrose and Yukich [11] on such limit theory.

As an application of the general theorems, in Theorem 3.3, below, we shall then give laws of large numbers for the statistics t_{j,ℓ_m}^m and $\Gamma_{j,k}^m$, which determine the MLE equations (2.12). We shall use Theorem 3.3 to prove Theorems 2.1 and 2.2.

For $x \in \mathbb{R}^d$ and $r > 0$, let $B(x, r)$ be the closed ball of radius r centred at x . A set $\mathcal{X} \subset \mathbb{R}^d$ is *locally finite* if $\mathcal{X} \cap B(x, r)$ is finite for all finite r and all $x \in \mathbb{R}^d$.

Definition 3.1. Let $r > 0$. A *locally determined functional* with range r is a measurable real-valued function $\xi(Y, \mathcal{X})$ defined for all pairs (Y, \mathcal{X}) , where $Y \in \mathbb{R}^d$ and $\mathcal{X} \subset \mathbb{R}^d$ is locally finite, with the property that $\xi(Y, \mathcal{X})$ is determined by the point set $\mathcal{X} \cap B(Y, r)$.

Given $w \in \mathbb{R}^d$ and given $\mathcal{X} \subset \mathbb{R}^d$, we write $w + \mathcal{X}$ for $\{w + x : x \in \mathcal{X}\}$. A locally determined functional ξ is said to be *translation invariant* if, for all $Y \in \mathbb{R}^d$, all locally finite $\mathcal{X} \subset \mathbb{R}^d$, and all $w \in \mathbb{R}^d$, we have $\xi(Y, \mathcal{X}) = \xi(w + Y, w + \mathcal{X})$.

Given a finite sequence $\mathbf{y} = (y_1, \dots, y_n) \in \bigcup_{k=0}^\infty (\mathbb{R}^d)^k$, we write $\langle \mathbf{y} \rangle$ for the corresponding *unordered* set of points $\{y_1, \dots, y_n\}$. Abusing notation, given $x \in \mathbb{R}^d$ and locally finite $\mathcal{X} \subset \mathbb{R}^d$, we write $\nu(x, \mathcal{X})$ for the number of points of \mathcal{X} in $B(x, R)$, where R is the interaction radius of the CSA model, so, in particular, $\nu(x, \langle \mathbf{y} \rangle)$ is the same as $\nu(x, \mathbf{y})$ defined in Section 2.1 (provided \mathbf{y} has no repeats, which is almost surely the case for those \mathbf{y} we consider).

Given a locally determined functional ξ , we define the corresponding additive functional H^ξ on finite sequences $X(k) = (X_1, \dots, X_k) \in (\mathbb{R}^d)^k$ by

$$H^\xi(X(k)) = \sum_{i=1}^k \xi(X_i, \langle X(i-1) \rangle). \tag{3.1}$$

Given m , let (Y_1^m, Y_2^m, \dots) be a sequence of independent random d -vectors uniformly distributed in D_m . Given $\ell \in \mathbb{N}$, let $A_m(\ell)$ be the number of accepted points, according to the acceptance/rejection scheme determining the CSA model, from the first ℓ arrivals $Y^m(\ell) = (Y_1^m, \dots, Y_\ell^m)$. For $0 \leq j \leq A_m(\ell)$, let $X^m(j) = (X_1^m, \dots, X_j^m) \in (\mathbb{R}^d)^j$ be the sequence of the first j accepted points. Note that sometimes (as in earlier sections of this paper) we omit the superscript and write $X(j)$ rather than $X^m(j)$ and X_k rather than X_k^m , when it is clear that the points are in D_m (this is the case when the measure $P_{m,\beta}$ or $P_m^{(0)}$ is mentioned explicitly).

The following law of large numbers for sums of locally determined functionals, given the number of arrivals Y_i^m , generalises Theorem 2.1 of [11].

Theorem 3.1. Let $\beta \in \mathcal{B}$. Let ξ be a bounded, translation-invariant, locally determined functional. Then there exists a function $(G_\beta(\tau), \tau > 0)$ such that if $(M_m, m \geq 1)$ is a sequence of nonrandom integers such that $M_m/m \rightarrow \tau > 0$ as $m \rightarrow \infty$, then, as $m \rightarrow \infty$,

$$\frac{H^\xi(X^m(A_m(M_m)))}{m} \xrightarrow{P_{m,\beta}} G_\beta(\tau). \tag{3.2}$$

In practice, we observe only the accepted points. Hence, we shall use the following law of large numbers for sums of locally determined functionals given the number of accepted points.

Theorem 3.2. *Let $\beta \in \mathcal{B}$. Let ξ be a bounded, translation-invariant, locally determined functional, and let $\beta \in \mathcal{B}$. There exists a continuous function $(V(\lambda), 0 \leq \lambda < \theta_\infty(\beta))$ (also dependent on β and ξ) such that if $\lim_{m \rightarrow \infty} (\ell_m/m) = \lambda \in (0, \theta_\infty(\beta))$ then*

$$\frac{H^\xi(X^m(\ell_m))}{m} \xrightarrow{P_{m,\beta}} V(\lambda) \quad \text{as } m \rightarrow \infty.$$

The following result gives the asymptotic behaviour of the statistics (2.9) and (2.10) determining the MLE.

Theorem 3.3. *For $\beta \in \mathcal{B}$, there exist strictly positive and continuous (in μ) functions $(\rho_j(\mu, \beta), \mu \in (0, \theta_\infty(\beta)))$ (defined for $1 \leq j \leq N$) and $(\gamma_j(\mu, \beta), \mu \in (0, \theta_\infty(\beta)))$ (defined for $0 \leq j \leq N$) such that if $\lim_{m \rightarrow \infty} (\ell_m/m) = \mu \in (0, \theta_\infty(\beta))$ then, as $m \rightarrow \infty$,*

$$\frac{t_{j,\ell_m}^m}{m} \xrightarrow{P_{m,\beta}} \rho_j(\mu, \beta), \quad j = 1, \dots, N, \tag{3.3}$$

and

$$\frac{\Gamma_{j,\ell_m}^m}{m} \xrightarrow{P_{m,\beta}} \gamma_j(\mu, \beta), \quad j = 0, \dots, N. \tag{3.4}$$

Also,

$$\rho_j(\mu, \beta) = \int_0^\mu \frac{\beta_j \gamma_j(\lambda, \beta)}{\gamma_0(\lambda, \beta) + \sum_{i=1}^N \beta_i \gamma_i(\lambda, \beta)} d\lambda, \quad j = 1, \dots, N. \tag{3.5}$$

Set $\gamma_j^{(0)}(\lambda) := \gamma_j(\lambda, \beta^{(0)})$ for $j = 0, \dots, N$ and $\rho_j^{(0)}(\mu) := \rho_j(\mu, \beta^{(0)})$ for $j = 1, \dots, N$. Given $\mu \in (0, \theta_\infty)$, by (3.5), the vector of true parameters $\beta^{(0)} = (\beta_1^{(0)}, \dots, \beta_N^{(0)})$ is a solution of the system of equations

$$\frac{\rho_j^{(0)}(\mu)}{\beta_j} - \int_0^\mu \frac{\gamma_j^{(0)}(\lambda)}{\gamma_0^{(0)}(\lambda) + \sum_{i=1}^N \beta_i \gamma_i^{(0)}(\lambda)} d\lambda = 0, \quad j = 1, \dots, N. \tag{3.6}$$

This is the infinite-volume analogue of the MLE (2.8). By a similar argument to the proof of uniqueness in the proof of Theorem 2.2, below, with the sums in (5.12) replaced by integrals, it can be shown that $\beta = \beta^{(0)}$ is the *unique* solution in \mathcal{B} to (3.6).

4. Proofs of Theorems 3.1, 3.2, and 3.3

4.1. Proof of Theorem 3.1

The aim is to use Theorem 3.2 of [11] or Theorem 2.1 of [9]. Given a finite point set $\mathcal{Y} \subset [0, \lambda] \times [0, 1] \times \mathbb{R}^d$ with I elements, list the elements of \mathcal{Y} as $(t_i, u_i, y_i)_{i=1}^I$ (with $t_i \in [0, \lambda]$, $u_i \in [0, 1]$, and $y \in \mathbb{R}^d$) in lexicographic order (so, in particular, $t_1 \leq t_2 \leq \dots \leq t_I$). Let χ_ℓ (the indicator that the ℓ th point is accepted) be defined recursively for $\ell = 1, \dots, I$ as follows. Suppose that $\chi_1, \dots, \chi_{\ell-1}$ are known; then set $k_{\ell-1} = \sum_{i=1}^{\ell-1} \chi_i$ (with $k_0 = 0$), and set

$$\mathbf{x}(k_\ell) = \{y_i : 1 \leq i \leq \ell, \chi_i = 1\}.$$

Then, with K as in Section 2.1, set $\chi_\ell = 1$ if $u_\ell \leq K^{-1} \beta_{v(y_\ell, \mathbf{x}(k_{\ell-1}))}$, with $\chi_\ell = 0$ otherwise. Also (with some abuse of notation), set

$$\psi(y_\ell; \mathcal{Y}) := \psi((t_\ell, u_\ell, y_\ell); \mathcal{Y}) := \begin{cases} \xi(y_\ell; \mathbf{x}(k_{\ell-1})) & \text{if } \chi_\ell = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then, with (t_ℓ, u_ℓ) viewed as the mark attached to the point $y_\ell \in \mathbb{R}^d$, we assert that $\psi(y, \mathcal{Y})$ is a deterministic and translation invariant functional on marked point sets in \mathbb{R}^d , in the sense of [9] or [11]. Here the mark of each point $y \in \mathbb{R}^d$ is an element (t, u) of $[0, \lambda] \times [0, 1]$, which we write as its prefix. Translation invariance follows because, when we translate each point $y \in \mathbb{R}^d$ by the same $x \in \mathbb{R}^d$ (leaving the marks unchanged), then the lexicographic ordering of points is unchanged, and the acceptance rule is determined by $v(y_\ell, \mathbf{x}(k_\ell - 1))$, which is unaffected by translation of all points by x .

Set $M = M_n$, as given in the statement of the theorem. Let \mathcal{Y}_M be a set of M independent, uniformly distributed random points in $[0, \lambda] \times [0, 1] \times D_m$. List them in lexicographic order as $(t_i, u_i, y_i)_{i=1}^M$. Then, following the recipe described above, it can be seen that the sequence of accepted points x_i follows the CSA model, so that (with ‘ $\stackrel{D}{=}$ ’ denoting equality in distribution)

$$H^\xi(X(A_m(M))) \stackrel{D}{=} \sum_{(t,u,y) \in \mathcal{Y}} \psi(y; \mathcal{Y}_M),$$

and we can apply Theorem 3.2 of [11] to obtain the required result (3.2) for the case with $M_m = \lfloor \lambda m \rfloor$. Indeed, the moment conditions for the result in [11] are trivial here because we assume that ξ is bounded so that all summands are bounded by a constant. Also, for all $\tau > 0$, the functional H is strongly stabilising in the sense of [11, p. 285]. That is, there is an almost-sure random variable \tilde{R}_τ such that the so-called ‘add-one cost’ of an insertion at a randomly marked point at the origin $\mathbf{0} := (0, \dots, 0)$ into a Poisson process of intensity τ in \mathbb{R}^d , restricted to $B(\mathbf{0}, \tau)$, is unaffected by addition of points outside $B(\mathbf{0}, \tilde{R}_\tau)$. All points here have marks uniformly distributed over $[0, \tau] \times [0, 1]$. This property holds by Lemma 4.2 of [11], as in the proof of Theorem 2.1 of [11].

The general case, $M_m/m \rightarrow \lambda$, is covered by Theorem 2.1 of [9], which generalises Theorem 3.2 of [11]. Only the case with $M_m = \lfloor m \rfloor$ will be used in the sequel.

4.2. The function $G_\beta(\tau)$

The preceding proof (using Theorem 3.2 of [11]) provides information about the nature of the limit $G_\beta(\tau)$ in (3.2). To describe this more clearly, let \mathcal{H} denote a unit-rate homogeneous Poisson process in $[0, \infty] \times [0, 1] \times \mathbb{R}^d$. Given $t > 0$ and $r > 0$, let $\mathcal{A}_t(r)$ denote the set of spatial locations of accepted points when the acceptance rule is applied to the restriction of \mathcal{H} to $[0, t] \times [0, 1] \times B(\mathbf{0}, r)$ (a finite set). By arguments in [11], there exists a limiting set \mathcal{A}_t such that, for all $\mathcal{K} > 0$, there exists finite (random) $r(\mathcal{K})$ such that $\mathcal{A}_t \cap B(\mathbf{0}, \mathcal{K}) = \mathcal{A}_t(r) \cap B(\mathbf{0}, \mathcal{K})$ for all $r > r(\mathcal{K})$.

The set \mathcal{A}_t can also be viewed as a set of spatial locations of particles accepted up to time t for the CSA process in \mathbb{R}^d . This is a spatial birth process, in the sense of, e.g. Section 4.1 of [10] (with only immigration events), where particles arrive as a unit-rate Poisson process in space–time $\mathbb{R}^d \times (0, \infty)$ and are accepted with probability equal to β_j/K , with j equal to the number of previously accepted neighbours of the incoming particle, and the initial state at time 0 is the empty set. As discussed in [10], this process turns out to be well defined. The spatial birth process is parametrised by $\beta = (\beta_1, \dots, \beta_N)$, so we write P_β and E_β for probability and expectation, respectively, in this model.

Proposition 4.1. *The constant $G_\beta(\tau)$ in (3.2) is given by*

$$G_\beta(\tau) = \int_0^\tau K^{-1} E_\beta[\beta_{v(\mathbf{0}; \mathcal{A}_t)} \xi(\mathbf{0}, \mathcal{A}_t)] dt, \tag{4.1}$$

where $\mathbf{0}$ denotes a particle inserted at the origin (at time t), and where K is as in Section 2.1.

Proof. Let \mathcal{H}_τ be a homogeneous Poisson point process in $[0, \tau] \times [0, 1] \times \mathbb{R}^d$, and let (T, U) be uniformly distributed over $[0, \tau] \times [0, 1]$ (and independent of \mathcal{H}_τ), representing the mark of an inserted point at the origin. According to Theorem 3.2 of [11], as used in the proof of Theorem 3.1, the limit $G_\beta(\tau)$ in (3.2) is given by the expression $\tau E \psi_\infty((T, U, \mathbf{0}); \mathcal{H}_\tau)$, where the limit

$$\psi_\infty((T, U, \mathbf{0}); \mathcal{H}_\tau) := \lim_{L \rightarrow \infty} \psi((T, U, \mathbf{0}); \mathcal{H}_\tau \cap ([0, \tau] \times [0, 1] \times B(\mathbf{0}, L)))$$

is shown to exist in [11]. Integrating out the value of T , we have

$$G_\beta(\tau) = \tau E_\beta[\psi_\infty((T, U, \mathbf{0}); \mathcal{H}_\tau)] = \int_0^\tau E_\beta[\psi_\infty((t, U, \mathbf{0}); \mathcal{H}_\tau)] dt. \tag{4.2}$$

Note that $\psi_\infty((t, U, \mathbf{0}); \mathcal{H}_\tau)$ is unaffected by points of \mathcal{H}_τ with time coordinate greater than t . In fact, $\psi_\infty((t, U, \mathbf{0}); \mathcal{H}_\tau)$ is equal to $\xi(\mathbf{0}; \mathcal{A}_t)$ if $U \leq K^{-1} \beta_{\nu(\mathbf{0}; \mathcal{A}_t)}$ and equal to zero otherwise. Hence, the integrand in (4.2) equals $K^{-1} \beta_{\nu(\mathbf{0}; \mathcal{A}_t)} \xi(\mathbf{0}; \mathcal{A}_t)$, so we obtain (4.1).

4.3. Proof of Theorem 3.2

Theorem 3.2 is proved using Theorem 3.1 and the following relation between the number of incoming Y_i^m and the number of accepted points X_i^m . Let $N_m(k) := \min\{n : A_m(n) \geq k\}$ be the number of input points for the rejection scheme on the cube D_m until at least k are accepted, with $N_m(k) = \infty$, if the total number of accepted points for an infinite input sequence is less than k .

Lemma 4.1. *Given $\beta \in \mathcal{B}$, there exists a strictly positive, continuous, and strictly increasing function $(\theta(\tau), \tau > 0)$ (which also depends on β) such that*

$$\frac{A_m([\tau m])}{m} \xrightarrow{P_{m,\beta}} \theta(\tau) \quad \text{as } m \rightarrow \infty. \tag{4.3}$$

Also, $\lim_{\tau \rightarrow \infty} \theta(\tau) = \theta_\infty(\beta)$, where the jamming density $\theta_\infty(\beta)$ was defined in Section 2.3. Also, with θ^{-1} denoting the inverse function to θ , for any $\lambda \in (0, \theta_\infty(\beta))$, as $m \rightarrow \infty$, we have

$$\frac{N_m(\lambda m)}{m} \xrightarrow{P_{m,\beta}} \theta^{-1}(\lambda). \tag{4.4}$$

Proof. The limit in (4.3) exists as a consequence of Theorem 3.1, where here we set $\xi(x, \mathcal{X}) \equiv 1$. By (4.1) we have

$$\theta(\tau) = \int_0^\tau K^{-1} E_\beta[\beta_{\nu(\mathbf{0}; \mathcal{A}_t)}] dt. \tag{4.5}$$

Hence, θ is continuous. Also, for all $t > 0$, $E_\beta[\beta_{\nu(\mathbf{0}; \mathcal{A}_t)}]$ is strictly positive since there is a nonzero probability that no Poisson points arrive within distance R of the origin before time t , and if this happens then $\beta_{\nu(\mathbf{0}; \mathcal{A}_t)} = \beta_0 = 1$. Hence, θ is monotonically strictly increasing, and is strictly positive.

Since we assume that $N < \infty$, it is not hard to see that θ is bounded and, therefore, since θ is monotonic, $\lim_{\lambda \rightarrow \infty} \theta(\lambda)$ exists. This limit is equal to the jamming density θ_∞ as defined earlier, as was shown in [8] for a slightly different CSA model; the proof carries over to the present case (for our purposes here, one could equally well have defined θ_∞ to be the limit $\lim_{\lambda \rightarrow \infty} \theta(\lambda)$).

Choose $\lambda \in (0, \theta_\infty(\beta))$. Since $\theta(\cdot)$ is continuous and strictly increasing, $\theta^{-1}(\lambda)$ is well defined. Let $b_0 < \theta^{-1}(\lambda) < b_1$. Then $\theta(b_0) < \lambda < \theta(b_1)$, and setting $t = b_i m$ in (4.3) ($i = 0, 1$) yields

$$P_{m,\beta}\{N_m(\lambda m) > b_0 m\} = P_{m,\beta}\left\{\frac{A_m(\lfloor b_0 m \rfloor)}{m} < \lambda\right\} \rightarrow 1$$

and

$$P_{m,\beta}\{N_m(\lambda m) \leq b_1 m\} = P_{m,\beta}\{A_m(\lfloor b_1 m \rfloor) \geq \lambda m\} \rightarrow 1,$$

and (4.4) follows.

We now complete the proof of Theorem 3.2. Set $s = \theta^{-1}(\lambda)$. Then it follows from Theorem 3.1 that

$$\frac{H^\xi(X^m(A_m(\lfloor ms \rfloor)))}{m} \xrightarrow{P_{m,\beta}} G_\beta(s) \tag{4.6}$$

for some deterministic limit $G_\beta(s)$. Also, using (4.3) and the assumption that $\ell_m/m \rightarrow \lambda$, we have

$$\frac{\ell_m - A_m(\lfloor ms \rfloor)}{m} \xrightarrow{P_{m,\beta}} 0,$$

and since ξ is assumed bounded, this implies that

$$\frac{|H^\xi(X^m(\ell_m)) - H^\xi(X^m(A_m(\lfloor ms \rfloor)))|}{m} \xrightarrow{P_{m,\beta}} 0,$$

which combined with (4.6) gives the result, if we equate

$$V(\lambda) = G_\beta(s) = G_\beta(\theta^{-1}(\lambda)). \tag{4.7}$$

To prove that the limit $V(\lambda)$ is a continuous function, first note that function $G_\beta(\cdot)$ is continuous by (4.1) and the assumption that ξ is bounded. The function $\theta^{-1}(\lambda)$ is continuous by Lemma 4.1, so $V(\lambda) = G_\beta(\theta^{-1}(\lambda))$ is continuous. The theorem is proved.

4.4. Proof of Theorem 3.3

We first show that the statistics t_{j,ℓ_m}^m and Γ_{j,ℓ_m}^m , $j = 1, \dots, N$, are sums of locally determined functionals, as in (3.1), modulo some asymptotically negligible boundary effects.

First consider the statistics t_{j,ℓ_m}^m , $j = 1, \dots, N$. Note that

$$t_{j,m}^m = \sum_{i=1}^{\ell_m} \mathbf{1}_{\{v(X_i, (X(i-1)))=j\}}. \tag{4.8}$$

It is easy to see that $\xi(x, \mathcal{X}) = \mathbf{1}_{\{v(x, \mathcal{X})=j\}}$ is a bounded, translation-invariant, locally determined functional with range R .

Now consider the statistics Γ_{j,ℓ_m}^m , $j = 1, \dots, N$, $k = 0, \dots, \ell_m$. It is helpful to define the approximations $\tilde{\Gamma}_{j,k}^m$ to $\Gamma_{j,k}$ as follows:

$$\tilde{\Gamma}_{j,k}^m = \int_{\mathbb{R}^d} \mathbf{1}_{\{u:v(u, X(k))=j\}} \, \mathbf{d}u, \quad j \geq 1, \quad \tilde{\Gamma}_{0,k}^m = m - \sum_{j=1}^{\infty} \tilde{\Gamma}_{j,k}^m.$$

The $\tilde{\Gamma}_{j,k}^m$ are good approximations to $\Gamma_{j,k}^m$ for the following reason. Set $\partial_R(D_m)$ to be the set of points in $\mathbb{R}^d \setminus D_m$ distant at most R from D_m , and let $|\cdot|$ denote the Lebesgue measure. Since R is fixed, it is straightforward to verify that

$$0 \leq \tilde{\Gamma}_{j,k}^m - \Gamma_{j,k}^m \leq |\partial_R(D_m)|, \quad j \geq 1, \tag{4.9}$$

$$0 \leq \Gamma_{0,k}^m - \tilde{\Gamma}_{0,k}^m \leq |\partial_R(D_m)|, \tag{4.10}$$

and

$$|\partial_R(D_m)| = O(m^{(d-1)/d}) \quad \text{as } m \rightarrow \infty. \tag{4.11}$$

For locally finite $\mathcal{X} \subset \mathbb{R}^d$, let $U_j(\mathcal{X}) \subseteq \mathbb{R}^d$ be the set of points with j neighbours in \mathcal{X} , i.e. the set of $y \in \mathbb{R}^d$ such that $\mathcal{X} \cap B(y, R)$ has j elements. For $x \in \mathbb{R}^d$, define

$$\xi_j(x, \mathcal{X}) = |B(x, R) \cap U_{j-1}(\mathcal{X})| - |B(x, R) \cap U_j(\mathcal{X})|, \quad j \geq 1, \tag{4.12}$$

$$\xi_0(x, \mathcal{X}) = |B(x, R) \cap U_0(\mathcal{X})|. \tag{4.13}$$

Then, for each j , $\xi_j(x, \mathcal{X})$ is a bounded, locally determined functional with range $2R$. It is not hard to see that, for $j \geq 1$,

$$\sum_{r=j}^{\infty} \tilde{\Gamma}_{r,k}^m = \sum_{i=1}^k |B(X_i, R) \cap U_{j-1}(\langle X(i-1) \rangle)|,$$

and from this we may deduce that

$$\tilde{\Gamma}_{j,k}^m = \sum_{i=1}^k \xi_j(X_i, \langle X(i-1) \rangle), \quad j \geq 1, \tag{4.14}$$

$$\tilde{\Gamma}_{0,k}^m = m - \sum_{i=1}^k \xi_0(X_i, \langle X(i-1) \rangle). \tag{4.15}$$

For (3.3), let $0 \leq j \leq N$. If we define the locally determined functional $\zeta_j(x, \mathcal{X}) = \mathbf{1}_{\{v(x, \mathcal{X})=j\}}$ then, by (2.9), (3.1), and (4.8), we have

$$t_{j, \ell_m}^m = H^{\zeta_j}(X^m(\ell_m)),$$

and, therefore, by Theorem 3.2, along with (4.1) and (4.7), as $m \rightarrow \infty$, we have

$$\begin{aligned} \frac{t_{j, \ell_m}^m}{m} &\xrightarrow{\mathbb{P}_{m, \beta}} \int_0^{\theta^{-1}(\mu)} \mathbb{E}_{\beta} \left[\frac{\beta_{v(\mathbf{0}; \mathcal{A}_t)}}{K} \zeta_j(\mathbf{0}, \mathcal{A}_t) \right] dt \\ &= \int_0^{\theta^{-1}(\mu)} \frac{\beta_j}{K} \mathbb{P}_{\beta} \{v(\mathbf{0}; \mathcal{A}_t) = j\} dt \\ &:= \rho_j(\mu, \beta). \end{aligned} \tag{4.16}$$

For each $t > 0$, there is a nonzero probability of there being precisely j Poisson arrivals in $B(\mathbf{0}; R)$ up to time t , and no other arrivals in $B(\mathbf{0}; 2R)$ up to time t , and of all the Poisson arrivals in $B(\mathbf{0}; R)$ up to time t being accepted. Thus, the integrand is strictly positive, so ρ_j is strictly positive. Also, $\mathbb{P}_{\beta} \{v(\mathbf{0}; \mathcal{A}_t) = j\}$ is continuous in β , so $\rho_j(t, \beta)$ is continuous in (t, β) .

For (3.4), observe first that, for $j \geq 1$, we have, from (3.1) and (4.14), $\tilde{\Gamma}_{j,\ell_m}^m = H^{\xi_j}(X^m(\ell_m))$, so that, by Theorem 3.1, along with (4.1) and (4.7), as $m \rightarrow \infty$, we have

$$\frac{\tilde{\Gamma}_{j,\ell_m}^m}{m} \xrightarrow{P_{m,\beta}} \int_0^{\theta^{-1}(\mu)} E_\beta \left[\frac{\beta_{v(\mathbf{0}; \mathcal{A}_t)}}{K} \xi_j(\mathbf{0}, \mathcal{A}_t) \right] dt := \gamma_j(\mu, \beta), \quad j \geq 1. \tag{4.17}$$

Also, by a similar argument starting from (4.15),

$$\frac{\tilde{\Gamma}_{0,\ell_m}}{m} \rightarrow 1 - \int_0^{\theta^{-1}(\mu)} E_\beta \left[\frac{\beta_{v(\mathbf{0}; \mathcal{A}_t)}}{K} \xi_0(\mathbf{0}, \mathcal{A}_t) \right] dt := \gamma_0(\mu, \beta). \tag{4.18}$$

Moreover, in (4.17) and (4.18) the convergence still holds with $\tilde{\Gamma}_{j,\ell_m}^m$ replaced by Γ_{j,ℓ_m}^m , by (4.9), (4.10), and (4.11).

Next we simplify the expressions for $\gamma_j(\lambda, \beta)$. First consider the case with $j \geq 1$. By definition (4.12), the integrand on the right-hand side of (4.17) satisfies

$$\begin{aligned} E_\beta \left[\frac{\beta_{v(\mathbf{0}; \mathcal{A}_t)}}{K} \xi_j(\mathbf{0}, \mathcal{A}_t) \right] &= E_\beta \left[\frac{\beta_{v(\mathbf{0}; \mathcal{A}_t)}}{K} |B(\mathbf{0}, R) \cap U_{j-1}(\mathcal{A}_t)| \right] \\ &\quad - E_\beta \left[\frac{\beta_{v(\mathbf{0}; \mathcal{A}_t)}}{K} |B(\mathbf{0}, R) \cap U_j(\mathcal{A}_t)| \right] \\ &= E_\beta \sum_{k=0}^\infty \frac{\beta_k}{K} \int_{B(\mathbf{0}, R)} \mathbf{1}_{\{v(\mathbf{0}; \mathcal{A}_t)=k\}} \mathbf{1}_{\{v(y; \mathcal{A}_t)=j-1\}} dy \\ &\quad - E_\beta \sum_{k=0}^\infty \frac{\beta_k}{K} \int_{B(\mathbf{0}, R)} \mathbf{1}_{\{v(\mathbf{0}; \mathcal{A}_t)=k\}} \mathbf{1}_{\{v(y; \mathcal{A}_t)=j\}} dy \\ &= \sum_{k=0}^\infty \frac{\beta_k}{K} \int_{B(\mathbf{0}, R)} E_\beta [\mathbf{1}_{\{v(y; \mathcal{A}_t)=k\}} \mathbf{1}_{\{v(\mathbf{0}; \mathcal{A}_t)=j-1\}}] dy \\ &\quad - \sum_{k=0}^\infty \frac{\beta_k}{K} \int_{B(\mathbf{0}, R)} E_\beta [\mathbf{1}_{\{v(y; \mathcal{A}_t)=k\}} \mathbf{1}_{\{v(\mathbf{0}; \mathcal{A}_t)=j\}}] dy \\ &= \frac{d}{dt} P_\beta\{v(\mathbf{0}; \mathcal{A}_t) = j\}, \end{aligned}$$

where, for the penultimate line, we either exploit the invariance of the distribution of \mathcal{A}_t under reflection in the mediator of $\mathbf{0}$ and y , or we make a change of variable $y \mapsto -y$ and exploit the translation invariance of the distribution of \mathcal{A}_t , and the last line comes from the Kolmogorov forward equation.

Since $P_\beta\{v(\mathbf{0}; \mathcal{A}_0) = j\} = 0$, it follows from (4.17) and the fundamental theorem of calculus that, for $j \geq 1$,

$$\gamma_j(\mu, \beta) = P_\beta\{v(\mathbf{0}; \mathcal{A}_{\theta^{-1}(\mu)}) = j\}. \tag{4.19}$$

Similarly, using (4.13), the integrand in (4.18) is

$$\begin{aligned} E_\beta \left[\frac{\beta_{v(\mathbf{0}; \mathcal{A}_t)} \xi_0(\mathbf{0}, \mathcal{A}_t)}{K} \right] dt &= E_\beta \left[\frac{\beta_{v(\mathbf{0}; \mathcal{A}_t)} | B(\mathbf{0}, R) \cap U_0(\mathcal{A}_t) |}{K} \right] \\ &= E_\beta \sum_{k=0}^\infty \frac{\beta_k}{K} \int_{B(\mathbf{0}, R)} \mathbf{1}_{\{v(\mathbf{0}; \mathcal{A}_t)=k\}} \mathbf{1}_{\{v(y; \mathcal{A}_t)=0\}} dy \\ &= \sum_{k=0}^\infty \frac{\beta_k}{K} \int_{B(\mathbf{0}, R)} E_\beta [\mathbf{1}_{\{v(y; \mathcal{A}_t)=k\}} \mathbf{1}_{\{v(\mathbf{0}; \mathcal{A}_t)=0\}}] dy \\ &= -\frac{d}{dt} P_\beta \{v(\mathbf{0}; \mathcal{A}_t) = 0\}, \end{aligned}$$

and since $P_\beta \{v(\mathbf{0}; \mathcal{A}_0) = 0\} = 1$, it follows from (4.18) and the fundamental theorem of calculus that (4.19) is valid for $j = 0$ too.

As was argued just after (4.16), for all $t > 0$, the value of $P_\beta \{v(\mathbf{0}; \mathcal{A}_t) = j\}$ is strictly positive, and this probability is also continuous in t . Hence, by (4.19), $\gamma_j(\mu, \beta)$ is strictly positive and continuous in β .

Next we prove (3.5). By (4.5), setting $\tau = \theta^{-1}(\lambda)$, we have

$$\frac{d\lambda}{d\tau} = K^{-1} E_\beta [\beta_{v(\mathbf{0}; \mathcal{A}_\tau)],$$

and, hence, by (4.16) and (4.19)

$$\frac{\partial}{\partial \lambda} \rho_j(\lambda, \beta) = K^{-1} \beta_j \gamma_j(\lambda, \beta) \frac{d\tau}{d\lambda} = \frac{\beta_j \gamma_j(\lambda, \beta)}{E_\beta [\beta_{v(\mathbf{0}; \mathcal{A}_{\theta^{-1}(\lambda)})}]} = \frac{\beta_j \gamma_j(\lambda, \beta)}{\sum_{i=0}^N \beta_i \gamma_i(\lambda, \beta)}, \tag{4.20}$$

and since $\rho_j(0, \beta) = 0$ by (4.16), integrating (4.20) from $\lambda = 0$ to $\lambda = \mu$ yields (3.5) by the fundamental theorem of calculus.

5. Proofs of results in Section 2

We start by introducing some notation and formulating two auxiliary lemmas that are proved in Section 6. Given $\beta^{(0)} \in \mathcal{B}$, define the function $\varphi: \mathcal{B} \times (0, \theta_\infty) \rightarrow \mathbb{R}$ by

$$\varphi(\beta, \lambda) = \sum_{j=1}^N \log \left(\frac{\beta_j^{(0)}}{\beta_j} \right) \frac{\beta_j^{(0)} \gamma_j^{(0)}(\lambda)}{Z(\beta^{(0)}, \lambda)} + \log \left(\frac{Z(\beta, \lambda)}{Z(\beta^{(0)}, \lambda)} \right), \tag{5.1}$$

where

$$Z(\beta, \lambda) = \gamma_0^{(0)}(\lambda) + \sum_{i=1}^N \beta_i \gamma_i^{(0)}(\lambda). \tag{5.2}$$

Lemma 5.1. *The function $\varphi(\beta, \lambda)$ is continuous in λ and, for any fixed $\lambda \in (0, \theta_\infty)$,*

$$\varphi(\beta^{(0)}, \lambda) = 0, \tag{5.3}$$

$$\varphi(\beta, \lambda) > 0 \text{ if } \beta \neq \beta^{(0)}. \tag{5.4}$$

The next lemma describes the asymptotic behaviour of the model’s log-likelihood derivatives.

Lemma 5.2. *Under Assumption 2.1,*

- for any $j = 1, \dots, N$,

$$\frac{1}{m} \frac{\partial L_m(\beta)}{\partial \beta_j} \Big|_{\beta=\beta^{(0)}} \xrightarrow{P_m^{(0)}} 0,$$

- for all $\beta \in \mathcal{B}$, there exists a symmetric matrix $J(\beta, \mu) = (J_{ij}(\beta, \mu))_{i,j=1}^N$ such that, for $1 \leq i, j \leq N$,

$$-\frac{1}{m} \frac{\partial^2 L_m(\beta)}{\partial \beta_i \partial \beta_j} \xrightarrow{P_m^{(0)}} J_{ij}(\beta, \mu) \text{ as } m \rightarrow \infty;$$

also, the matrix $J^{(0)}(\mu) := (J_{ij}(\beta^{(0)}, \mu))_{i,j=1}^N$ is positive definite,

- there exist an open neighbourhood \mathcal{U} of $\beta^{(0)}$ and a constant $C > 0$, such that, for all m ,

$$\left| \frac{1}{m} \frac{\partial^3 L_m(\beta)}{\partial \beta_j \partial \beta_i \partial \beta_p} \right| \leq C \text{ almost surely}$$

for any $1 \leq j, i, p \leq N$, any m , and all $\beta \in \mathcal{U}$.

5.1. Proof of Theorem 2.1

Let $\beta = (\beta_1, \dots, \beta_N) \in \mathcal{B}$, and set $\beta_0 = \beta_0^{(0)} = 1$. By (2.9), (2.10), and (2.11),

$$L_m(\beta^{(0)}) - L_m(\beta) = \sum_{j=1}^N t_{j,\ell_m}^m \log\left(\frac{\beta_j^{(0)}}{\beta_j}\right) + \sum_{k=1}^{\ell_m} \log\left(\frac{\sum_{i=0}^N \beta_i \Gamma_{i,k}^m}{\sum_{i=0}^N \beta_i^{(0)} \Gamma_{i,k}^m}\right). \tag{5.5}$$

By (3.4), for $0 \leq j \leq N$ and any $\lambda \in (0, \mu]$,

$$\frac{\Gamma_{j, \lceil \lambda m \rceil}^m}{m} \xrightarrow{P_m^{(0)}} \gamma_j^{(0)}(\lambda) \text{ as } m \rightarrow \infty. \tag{5.6}$$

By Theorem 3.3, we can (and now do) choose a constant η with $0 < \eta < \sum_{i=1}^N \gamma_j^{(0)}(\mu)$, and using (3.4), if we define the event

$$F_m := \left\{ \sum_{j=1}^N \Gamma_{j, \lceil \mu m \rceil}^m > \eta m \right\},$$

then we have $\lim_{m \rightarrow \infty} P_{m, \beta^{(0)}}(F_m) = 1$.

Define $\beta_{\min} = \min\{\beta_0, \dots, \beta_N\}$ and $\beta_{\max} = \max\{\beta_0, \dots, \beta_N\}$. Since

$$\beta_{\min} \sum_{i=0}^N \Gamma_{i,k}^m \leq \sum_{i=0}^N \beta_i \Gamma_{i,k}^m \leq \beta_{\max} \sum_{i=0}^N \Gamma_{i,k}^m,$$

and since $\sum_{i=0}^N \Gamma_{i,k}^m$ is monotonically nonincreasing in k and bounded above by m for all k , on the event F_m we have, for all $\lambda \in [0, \mu]$, the bounds

$$\beta_{\min} \eta \leq \frac{\sum_{i=1}^N \beta_i \Gamma_{i, \lceil \lambda m \rceil}^m}{m} \leq \beta_{\max}. \tag{5.7}$$

Using (5.7) and the mean value theorem, then using (5.6) and (5.7) again, we find that

$$\begin{aligned} & \mathbb{E}_m^{(0)} \left[\left| \log \left(\sum_{i=0}^N \beta_i \frac{\Gamma_{i, \lceil \lambda m \rceil}^m}{m} \right) - \log \left(\sum_{i=0}^N \beta_i \gamma_i^{(0)}(\lambda) \right) \right| \mathbf{1}_{F_m} \right] \\ & \leq \frac{1}{\beta_{\min} \eta} \mathbb{E}_m^{(0)} \left[\left| \sum_{i=1}^N \beta_i \left(\frac{\Gamma_{i, \lceil \lambda m \rceil}}{m} - \gamma_i^{(0)}(\lambda) \right) \right| \mathbf{1}_{F_m} \right] \\ & \rightarrow 0 \end{aligned} \tag{5.8}$$

as $m \rightarrow \infty$. Using (5.2), (5.7) again, and (5.8), and using Assumption 2.1, we have

$$\begin{aligned} & \mathbb{E}_m^{(0)} \left[\left| \left(\frac{1}{m} \sum_{k=1}^{\ell_m} \log \left(\sum_{i=0}^N \beta_i \frac{\Gamma_{i,k}^m}{m} \right) \right) - \int_0^\mu \log(Z(\beta, \lambda)) \, d\lambda \right| \mathbf{1}_{F_m} \right] \\ & = \mathbb{E}_m^{(0)} \left[\left| \left(\int_0^{\ell_m/m} \log \left(\sum_{i=0}^N \beta_i \frac{\Gamma_{i, \lceil \lambda m \rceil}^m}{m} \right) \, d\lambda \right) - \int_0^\mu \log(Z(\beta, \lambda)) \, d\lambda \right| \mathbf{1}_{F_m} \right] \\ & \rightarrow 0 \end{aligned}$$

as $m \rightarrow \infty$. Hence, since $\mathbb{P}_{m, \beta^{(0)}}(F_m^c) \rightarrow 0$, by (3.3), (3.4), and (5.5), as $m \rightarrow \infty$,

$$\frac{1}{m} (L_m(\beta^{(0)}) - L_m(\beta)) \xrightarrow{\mathbb{P}_m^{(0)}} L(\beta, \mu),$$

where

$$L(\beta, \mu) = \sum_{j=1}^N \rho_j^{(0)}(\mu) \log \left(\frac{\beta_j^{(0)}}{\beta_j} \right) + \int_0^\mu \log \left(\frac{Z(\beta, \lambda)}{Z(\beta^{(0)}, \lambda)} \right) \, d\lambda.$$

By (3.5), the function $L(\beta, \mu)$ in the preceding display satisfies

$$L(\beta, \mu) = \int_0^\mu \varphi(\beta, \lambda) \, d\lambda, \tag{5.9}$$

where the function $\varphi(\beta, \lambda)$ is defined by (5.1). By (5.9) and Lemma 5.1, $L(\beta^{(0)}, \mu) = 0$ and $L(\beta, \mu) > 0$ for any $\beta \neq \beta^{(0)}$. Continuity in μ of $L(\beta, \mu)$ follows from (5.9), so the theorem is proved.

5.2. Proof of Theorem 2.2

By (3.3),

$$\mathbb{P}_m^{(0)} \{ \widehat{N}(X(\ell_m)) = N \} = \mathbb{P}_m^{(0)} \{ t_{N, \ell_m}^m > 0 \} \rightarrow 1$$

as $m \rightarrow \infty$; hence, the first assertion of the theorem is proved and without loss of generality we may assume in the rest of the proof that $\widehat{N} = N$.

If $N = 1$ then the second assertion of the theorem is easy to prove by using monotonicity in β of the left-hand side of the only equation in system (2.12) and by noting that

$$\mathbb{P}_m^{(0)} \{ 0 < t_{1, \ell_m}^m < \ell_m - 1 \} \rightarrow 1$$

as $m \rightarrow \infty$ by Theorem 3.3.

We now show for the general case that, with the $P_m^{(0)}$ -probability tending to 1 as $m \rightarrow \infty$, the MLE equations cannot have more than one positive solution. Assume that there are two distinct positive solutions, $\beta = \beta' = (\beta'_1, \dots, \beta'_N)$ and $\beta = \beta'' = (\beta''_1, \dots, \beta''_N)$, of system (2.12), so that

$$t_{j,\ell_m}^m = \beta'_j \sum_{k=1}^{l_m} \frac{\Gamma_{j,k-1}^m}{\Gamma_{0,k-1}^m + \sum_{i=1}^N \beta'_i \Gamma_{i,k-1}^m}, \quad j = 1, \dots, N, \tag{5.10}$$

and

$$t_{j,\ell_m}^m = \beta''_j \sum_{k=1}^{l_m} \frac{\Gamma_{j,k-1}^m}{\Gamma_{0,k-1}^m + \sum_{i=1}^N \beta''_i \Gamma_{i,k-1}^m}, \quad j = 1, \dots, N. \tag{5.11}$$

Set $\beta'_0 = \beta''_0 = 1$, and set $Z'_k := \sum_{i=0}^N \beta'_i \Gamma_{i,k-1}^m$, which is just the denominator inside the sum in (5.10). Similarly, set $Z''_k := \sum_{i=0}^N \beta''_i \Gamma_{i,k-1}^m$, which is the denominator inside the sum in (5.11).

Since $\beta' \neq \beta''$, without loss of generality, we assume that there exists j such that $\beta'_j < \beta''_j$. Setting $w_j = \beta'_j / \beta''_j$, we have, for some j , $w_j > 1$.

Choose j_0 such that $w_{j_0} = \max(w_1, \dots, w_N)$. By the preceding discussion, $w_{j_0} > 1$. By (5.10) and (5.11), for each $j \in \{1, \dots, N\}$, we have

$$0 = \sum_{k=1}^{l_m} \frac{\beta'_j \Gamma_{j,k-1}^m Z''_k - \beta''_j \Gamma_{j,k-1}^m Z'_k}{Z'_k Z''_k} = \sum_{i=0}^N \beta'_i \beta''_i (w_j - w_i) A_{ij}, \tag{5.12}$$

where

$$A_{ij} := \sum_{k=1}^{l_m} \frac{\Gamma_{i,k-1}^m \Gamma_{j,k-1}^m}{Z'_k Z''_k}.$$

By Theorem 3.3, for all i, j , we have $A_{ij} > 0$ with the $P_m^{(0)}$ -probability tending to 1 as $m \rightarrow \infty$.

Consider the last line of (5.12) in the case in which $j = j_0$. By the choice of j_0 , for $j = j_0$, each term in the sum in (5.12) is nonnegative, and the term for $i = 0$ is strictly positive since $w_0 = 1$, so that (5.12) must fail and we have a contradiction. Thus, the solution must be unique.

We complete the proof as follows. By Taylor's theorem, and part 3 of Lemma 5.2, there exists $\delta_0 > 0$ such that, for all β with $\|\beta - \beta^{(0)}\| \leq \delta_0$,

$$\begin{aligned} L_m(\beta) &= L_m(\beta^{(0)}) + (\beta - \beta^{(0)}) \cdot \nabla L_m(\beta^{(0)}) \\ &\quad - (\beta - \beta^{(0)})' \left[\frac{\partial^2 L_m(\beta)}{\partial^2 \beta} \right]_{\beta^{(0)}} (\beta - \beta^{(0)}) + r_m(\beta), \end{aligned}$$

where $r_m(\beta) = r_m(X(l_m), \beta)$ is such that

$$|r_m(\beta)| \leq Cm \|\beta - \beta^{(0)}\|^3.$$

Then, by parts 1 and 2 of Lemma 5.2, there exist strictly positive constants C_1 and C_2 , and a further constant $\delta_1 \in (0, \delta_0)$ such that, for all $\delta \in (0, \delta_1)$,

$$P_m^{(0)} \left\{ \inf_{\|\beta - \beta^{(0)}\| = \delta} \frac{1}{m} (L_m(\beta^{(0)}) - L_m(\beta)) > C_1 \delta^2 - C_2 \delta^3 > 0 \right\} \rightarrow 1$$

as $m \rightarrow \infty$, and this implies that there exists $\delta_2 \in (0, \delta_1)$ such that, for any $\delta \in (0, \delta_2)$, with the $P_m^{(0)}$ -probability tending to 1 as $m \rightarrow \infty$, there exists a positive root $(\beta_{1,m}^{(0)}, \dots, \beta_{N,m}^{(0)}) \in B(\beta^{(0)}, \delta)$ of the MLE equation.

Since, for arbitrarily small $\delta \in (0, \delta_2)$, with the $P_m^{(0)}$ -probability tending to 1, (i) the solution is unique and (ii) there exists a solution in $B(\beta^{(0)}, \delta)$, we can conclude that there exists a sequence of unique positive solutions $(\hat{\beta}_{1,m}, \dots, \hat{\beta}_{N,m})$ of MLE equations converging to the true parameter in $P_m^{(0)}$ -probability. The theorem is proved.

6. Proofs of Lemmas 5.1 and 5.2

6.1. Proof of Lemma 5.1

Continuity of $\varphi(\beta, \lambda)$ in λ is implied by definitions (5.1) and (5.2), and Theorem 3.3. Assertion (5.3) is obvious. To prove (5.4), we shall show that $\beta^{(0)}$ is the only minimum point of the function $\varphi(\beta, \lambda)$, $\beta \in \mathbb{R}_+^N$, for any fixed λ . Consider the system of equations determining the extremal points of this function:

$$\frac{\partial \varphi(\beta, \lambda)}{\partial \beta_j} = -\frac{1}{\beta_j} \frac{\beta_j^{(0)} \gamma_j^{(0)}(\lambda)}{Z(\beta^{(0)}, \lambda)} + \frac{\gamma_j^{(0)}(\lambda)}{Z(\beta, \lambda)} = 0, \quad j = 1, \dots, N. \tag{6.1}$$

All the β s are assumed to be positive, and setting $\beta_0 = \beta_0^{(0)} = 1$, $Z = Z(\beta, \lambda)$ as defined in (5.2), and $Z^{(0)} = Z(\beta^{(0)}, \lambda)$, we can rewrite (6.1) (for each j) as

$$\begin{aligned} 0 &= \frac{\beta_j \gamma_j^{(0)}}{Z} - \frac{\beta_j^{(0)} \gamma_j^{(0)} Z^{(0)}}{Z^{(0)}} \\ &= \frac{\beta_j \gamma_j^{(0)} Z^{(0)} - \beta_j^{(0)} \gamma_j^{(0)} Z}{Z^{(0)} Z} \\ &= \frac{\gamma_j^{(0)}}{Z^{(0)} Z} \left(\beta_j \sum_{i=0}^N \beta_i^{(0)} \gamma_i^{(0)} - \beta_j^{(0)} \sum_{i=0}^N \beta_i \gamma_i^{(0)} \right) \\ &= \frac{\gamma_j^{(0)}}{Z^{(0)} Z} \left(\sum_{i=0}^N \gamma_i^{(0)} (\beta_j \beta_i^{(0)} - \beta_i \beta_j^{(0)}) \right), \end{aligned}$$

and setting $v_i = \beta_i / \beta_i^{(0)}$ we find that

$$0 = \frac{\gamma_j^{(0)}}{Z^{(0)} Z} \left(\sum_{i=0}^N \gamma_i^{(0)} \beta_j^{(0)} \beta_i^{(0)} (v_j - v_i) \right). \tag{6.2}$$

Note that $v_0 = 1$. If $\max_i v_i > 1$ then, by choosing $j = j_1$ to maximise v_j , setting $j = j_1$ in (6.2), we obtain a contradiction. If $\min_i v_i < 1$ then, by choosing j_2 to minimise v_j , setting $j = j_2$ in (6.2), we obtain a contradiction. Hence, $v_i = 1$ for all i , i.e. $\beta = \beta^{(0)}$.

To show that the critical point $\beta^{(0)}$ is a minimum point of the function $\varphi(\beta, \lambda)$, $\beta \in \mathbb{R}_+^N$, we need to check that, for any fixed λ , the matrix $Q(\beta^{(0)}, \lambda) = (Q_{ij}(\beta^{(0)}, \lambda))_{i,j=1}^N$ given by

$$Q_{ij}(\beta^{(0)}, \lambda) = \frac{\partial^2 \varphi(\beta, \lambda)}{\partial \beta_i \partial \beta_j} \Big|_{\beta=\beta^{(0)}} \tag{6.3}$$

is positive definite. By differentiating (6.1) with respect to β_i , and considering separately the cases $i = j$ and $i \neq j$, we find that

$$Q_{ij}(\beta^{(0)}, \lambda) = \left(\frac{\gamma_i^{(0)}}{\beta_i^{(0)} Z^{(0)}} \right) \delta_{ij} - \frac{\gamma_i^{(0)} \gamma_j^{(0)}}{(Z^{(0)})^2},$$

where δ_{ij} is the Kronecker symbol. Hence, for any N -vector $\mathbf{a} = (a_1, \dots, a_N)'$, we have

$$\mathbf{a}'((Z^{(0)})^2 Q(\beta^{(0)}, \lambda))\mathbf{a} = \sum_{i=1}^N \frac{a_i^2 \gamma_i^{(0)} Z^{(0)}}{\beta_i^{(0)}} - \sum_{i=1}^N \sum_{j=1}^N a_i a_j \gamma_i^{(0)} \gamma_j^{(0)}, \tag{6.4}$$

but, by the Cauchy–Schwarz inequality, if \mathbf{a} is nonzero then

$$\begin{aligned} \left(\sum_{i=1}^N a_i \gamma_i^{(0)} \right)^2 &= \left(\sum_{i=1}^N (a_i \beta_i^{(0)}) \left(\frac{\gamma_i^{(0)}}{\beta_i^{(0)}} \right) \right)^2 \\ &\leq \left(\sum_{i=1}^N a_i^2 \left(\frac{\gamma_i^{(0)}}{\beta_i^{(0)}} \right) \right) \left(\sum_{i=1}^N (\beta_i^{(0)})^2 \left(\frac{\gamma_i^{(0)}}{\beta_i^{(0)}} \right) \right) \\ &= \left(\sum_{i=1}^N a_i^2 \left(\frac{\gamma_i^{(0)}}{\beta_i^{(0)}} \right) \right) (Z^{(0)} - \gamma_0^{(0)}) \\ &< \left(\sum_{i=1}^N a_i^2 \left(\frac{\gamma_i^{(0)}}{\beta_i^{(0)}} \right) \right) Z^{(0)}, \end{aligned}$$

so that (6.4) is strictly positive, and the positive definiteness of matrix (6.3) is established. Hence, for any fixed λ , the point $\beta^{(0)}$ is the only minimum point of the function $\varphi(\beta, \lambda)$, $\beta \in \mathbb{R}_+^N$, and the lemma is proved.

6.2. Proof of Lemma 5.2

Part 1. Note from (2.11) that

$$\frac{\partial L_m(\beta)}{\partial \beta_j} = \frac{t_{j,\ell_m}^m}{\beta_j} - \sum_{k=1}^{\ell_m} \frac{\Gamma_{j,k-1}^m}{\Gamma_{0,k-1}^m + \sum_{i=1}^N \beta_i \Gamma_{i,k-1}^m}. \tag{6.5}$$

Dividing both the numerator and denominator in the fraction by m and using (3.3) and (3.4), we may deduce that, for any $\beta \in \mathcal{B}$, as $m \rightarrow \infty$, we have

$$m^{-1} \frac{\partial L_m(\beta)}{\partial \beta_j} \xrightarrow{P_m^{(0)}} \frac{\rho_j^{(0)}(\mu)}{\beta_j} - \int_0^\mu \frac{\gamma_j^{(0)}(\lambda)}{Z(\beta, \lambda)} d\lambda,$$

and if $\beta = \beta^{(0)}$ then this is zero by part 3 of Theorem 3.3. Thus, part 1 holds.

Part 2. By (3.3), (3.4), and (6.5), for any $\beta \in \mathcal{B}$,

$$-\frac{1}{m} \frac{\partial^2 L_m(\beta)}{\partial \beta_i \partial \beta_j} \xrightarrow{P_m^{(0)}} \frac{\rho_i^{(0)}(\mu)}{\beta_i^2} \delta_{ij} - \int_0^\mu \left(\frac{\gamma_i^{(0)}(\lambda) \gamma_j^{(0)}(\lambda)}{Z(\beta, \lambda)^2} \right) d\lambda,$$

where δ_{ij} is the Kroneker symbol. By (3.5) and (5.1),

$$\frac{\rho_i^{(0)}(\mu)}{(\beta_i^{(0)})^2} \delta_{ij} - \int_0^\mu \left(\frac{\gamma_i^{(0)} \gamma_j^{(0)}(\lambda)}{Z(\beta^{(0)}, \lambda)^2} \right) d\lambda = \int_0^\mu Q_{ij}(\beta^{(0)}, \lambda) d\lambda,$$

so

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

where $Q(\beta^{(0)}, \lambda)$ is the matrix defined by (6.3). We proved in the proof of Lemma 5.1 that, for any fixed λ , this matrix is positive definite, and, therefore, so is the matrix $J^{(0)}(\mu)$.

Part 3. Assume, for instance, that $j = i = p$. Then a direct computation from (6.5) gives

$$\frac{1}{m} \frac{\partial^3 L_m(\beta)}{\partial \beta_j^3} = \frac{2t_{j,\ell_m}^m}{m\beta_j^3} + \frac{2}{m} \sum_{k=1}^{\ell_m} \frac{(\Gamma_{j,k-1}^m)^3}{(\Gamma_{0,k-1}^m + \sum_{i=1}^N \beta_i \Gamma_{i,k-1}^m)^3} \leq \frac{4\ell_m}{m\beta_j^3}, \tag{6.6}$$

where the last inequality follows since $t_{j,\ell_m}^m \leq \ell_m$ by (2.4) and (2.9).

By Assumption 2.1, ℓ_m/m is bounded by a constant. Also, if \mathcal{U} is a compact neighbourhood of $\beta^{(0)}$ with $\mathcal{U} \subset \mathcal{B}$ then, for $\beta \in \mathcal{U}$, we have $\beta_j^{-3} \leq (\inf_{\beta \in \mathcal{U}} \beta_j)^{-3}$, which is finite since \mathcal{U} is compact and contained in \mathcal{B} . Hence, (6.6) is bounded uniformly in $m \geq 1, \beta \in \mathcal{U}$. When all indices are different or just two of them coincide, similar bounds can be obtained in the same way; we omit the details. Lemma 5.2 is proved.

References

- [1] BRILLINGER, D. R., ROBINSON, E. A. AND SCHOENBERG, F. P. (eds) (2004). *Time Series Analysis and Applications to Geophysical Systems* (IMA Vol. Math. Appl. **139**). Springer, New York.
- [2] CADILHE, A., ARAUJO, N. A. M. AND PRIVMAN, V. (2007). Random sequential adsorption: from continuum to lattice and pre-patterned substrates. *J. Phys. Cond. Matter* **19**, 065124, 12 pp.
- [3] EVANS, J. W. (1993). Random and cooperative sequential adsorption. *Rev. Modern Phys.* **65**, 1281–1329.
- [4] GEYER, C. J. (1999). Likelihood inference for spatial point processes. In *Stochastic Geometry* (Toulouse, 1996; Monogr. Statist. Appl. Prob. **80**), eds O. Barndorff-Nielsen, W. S. Kendall and M. N. M. van Lieshout, Chapman and Hall/CRC, Boca Raton, FL, pp. 79–140.
- [5] GEYER, C. J. AND MØLLER, J. (1994). Simulation procedures and likelihood inference for spatial point processes. *Scand. J. Statist.* **21**, 359–373.
- [6] LEHMANN, E. L. AND CASELLA, G. (1998). *Theory of Point Estimation*. Springer, New York.
- [7] MØLLER, J. AND WAAGEPETERSEN, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton, FL.
- [8] PENROSE, M. D. (2001). Random parking, sequential adsorption, and the jamming limit. *Commun. Math. Phys.* **218**, 153–176.
- [9] PENROSE, M. D. (2007). Laws of large numbers in stochastic geometry with statistical applications. *Bernoulli* **13**, 1124–1150.
- [10] PENROSE, M. D. (2008). Existence and spatial limit theorems for lattice and continuum particle systems. *Prob. Surveys* **5**, 1–36.
- [11] PENROSE, M. D. AND YUKICH, J. E. (2002). Limit theory for random sequential packing and deposition. *Ann. Appl. Prob.* **12**, 272–301.
- [12] PRIVMAN, V. (ed.) (2000). *Adhesion of Submicron Particles on Solid Surfaces* (Colloids and Surfaces A Spec. Vol. **165**), Elsevier BV.
- [13] SHCHERBAKOV, V. (2006). Limit theorems for random point measures generated by cooperative sequential adsorption. *J. Statist. Phys.* **124**, 1425–1441.
- [14] SHCHERBAKOV, V. (2009). On a model of sequential point patterns. *Ann. Inst. Statist. Math.* **61**, 371–390.

- [15] VAN LEISHOUT, M. N. M. (2006). Maximum likelihood estimation for random sequential adsorption. *Adv. Appl. Prob.* **38**, 889–898.
- [16] VAN LIESHOUT, M. N. M. (2006). Markovianity in space and time. In *Dynamics and Stochastics* (Lecture Notes Monogr. Ser. **48**), eds D. Denteneer, F. den Hollander and E. Verbitskiy, Institute for Mathematical Statistics, Beachwood, OH, pp. 154–168.
- [17] ZHUANG, J., OGATA, Y. AND VERE-JONES, D. (2002). Stochastic declustering of space-time earthquake occurrences. *J. Amer. Statist. Soc.* **97**, 369–380.