Time-invariance modeling and estimation for spatial point processes: General theory

Thomas G. Kurtz *
Departments of Mathematics and Statistics
University of Wisconsin - Madison
480 Lincoln Drive
Madison, WI 53706-1388
kurtz@math.wisc.edu

Shun-Hwa Li
15800 Bull Run Road #262
Miami Lakes, FL 33014
shunhwa@yahoo.com

March 16, 2003

Abstract

Models for spatial point processes are characterized as the stationary distributions of spatial birth and death processes. The Markov chain Monte Carlo algorithm determined by the underlying birth and death process immediately gives a method of simulation, and the time-invariance method of estimation proposed by Baddeley (2000) gives a general method for deriving parameter estimates for the models. Typically, in applications of Markov chain Monte Carlo and time-invariance estimation, one begins with the model of interest specified in some other way and then constructs a Markov process having the desired distribution as its stationary distribution; however, specifying the model directly in terms of the Markov process provides an intuitive and flexible approach to modeling.

In time-invariance estimation, the parameter estimates are obtained by equating to zero the generator of the Markov process applied to a suitable collection of statistics. The estimators depend on the choice of the statistics, and the art is to find statistics that give estimators that are easy to compute and have good statistical properties. Statistics are given that are useful for spatial point processes including Gibbs and non-Gibbs models, and a large sample limit theorem is proved for these statistics which enables one to verify consistency of the estimators for particular classes of models.

MSC 2000 subject classifications: Primary: 62M30, 60G55 Secondary: 62F12, 62F10, 60F15

Keywords: spatial point processes, Gibbs point processes, spatial birth-and-death processes, Markov chain Monte Carlo, time-invariance estimation, unbiased estimating equation, consistency, spatial ergodic theory.

*This material is based upon work supported by, or in part by, the U. S. Army Research Laboratory and the U. S. Army Research Office under contract, grant number DAAD19-01-1-0502, and by NSF grant DMS 02-05034

1 Introduction

Statistical models for spatial point processes have been developed and discussed for more than two decades. Many of the models considered are specified in terms of a density function, explicitly given up to a normalizing constant. Densities for Gibbs models, for example, are expressed as exponentials of *interaction potentials* depending on the locations of point pairs, triples, etc. (See Daley and Vere-Jones (2003), Example 5.3c.) Models of this type form a rich class that captures in an intuitive way many of the features observed in spatial point data (e.g. clustering and repulsion); however, even when these densities are of a relatively simple form, the models present many practical challenges. For example, the variability of the density may make the standard rejection method for simulation impractical, and the unknown normalizing constant may make likelihood based inference difficult and unpredictable.

Ripley (1979) addresses the simulation problem by developing a Markov chain Monte Carlo algorithm based on identifying a spatial birth and death process whose stationary distribution gives the desired spatial point process. Baddeley (2000) provides an analogous solution to the parameter estimation problem with the introduction of his time-invariance estimators. Like the more familiar Markov chain Monte Carlo methods, Baddeley's approach to parameter estimation depends on characterizing the distributions of interest $\{\pi_\theta\}$ as stationary distributions of Markov processes, say with generators $\{\mathcal{A}_\theta\}$. Since the stationary distribution anihilates the generator, that is,

$$\int \mathcal{A}_\theta f d\pi_\theta = 0, \quad f \in \mathcal{D}(\mathcal{A}_\theta), \tag{1.1}$$

this characterization gives a large family of unbiased estimating equations that can be used to derive estimators for the parameter θ.

Markov chain Monte Carlo has become the standard method for simulation of spatial point processes, and Baddeley's time-invariance estimation method yields new tractable estimators as well as providing an alternative approach to the derivation of many well-known estimators for Gibbs and other point processes. The success of these methods leads us to consider (1.1) as the primary characterization of a parametric family of models, that is, we specify a parametric family of generators $\{\mathcal{A}_\theta\}$ and verify that each generator uniquely determines a probability distribution π_θ satisfying (1.1). The underlying Markov process then provides a method of simulating samples from π_θ, and Baddeley's method gives parameter estimates.

We identify a finite configuration of points $\{x_i \in S, 1 \le i \le m\}$ in a complete, separable metric space S (typically $\mathbb{R}^2$) with the counting measure $\eta = \sum_{i=1}^m \delta_{x_i}$. The generators $\mathcal{A}_\theta$ in which we are interested are of the form

$$\mathcal{A}_\theta f(\eta) = \int_S (f(\eta+\delta_u) - f(\eta)) b_\theta(\eta, u)\nu(du) + \int_S (f(\eta-\delta_u) - f(\eta)) d_\theta(\eta, u)\eta(du), \tag{1.2}$$

where ν is a fixed, diffuse measure. Note that the first term models the births in our spatial birth and death process and the second, which for $\eta = \sum_{i=1}^m \delta_{x_i}$ could also be written

$$\sum_{i=1}^m d_\theta(\eta, x_i)(f(\eta - \delta_{x_i}) - f(\eta)),$$

models the deaths. Specifiying a model then consists of specifying b_θ and d_θ.

There are several existing methods of parameter estimation for spatial point processes. These include maximum likelihood and pseudo-likelihood methods and the Takacs-Fiksel method. Properties of these estimators and their effectiveness have been investigated by a number of authors. Diggle, Fiksel, Grabarnik, Ogata, Stoyan and Tanemura (1994) review these methods applied to pairwise interaction models, a special class of Gibbs models. Jensen and Møller (1991) prove consistency of pseudo-likelihood estimators for finite range Markov point processes. In addition, the asymptotic normality of estimators for pairwise interaction models has been studied by Jensen (1993) and Jensen and Künsch (1994). However, there seem to be no results available on either the consistency or asymptotic normality of estimators for other classes of models.

In Section 2, we introduce the generators for two general Markov spatial birth and death processes and give conditions under which the processes have unique stationary distributions. These stationary distributions then give models which we will refer to as *time-variance models* for spatial point processes. In Section 3, we derive time-invariance estimators for particular parametric families of time-invariance models, and in Section 4, we discuss consistency, in particular, verifying consistency for the estimators derived in Section 3.

Extensive simulation and data-analytic studies of these estimators have been carried out in Li (1999) and will appear in Li (n.d.). The simulations demonstrate the effectiveness of the estimators. In addition, the modeling and estimation methods are applied to sample data sets. Data sets exhibiting clustering and repulsion are considered, and simple time-invariance models are shown to capture the basic propeties of the data. In particular, goodness-of-fit tests discussed in Diggle (1983) are applied, and the time-invariance models are shown to provide a reasonable fit.

2 Time-invariance models for spatial point processes

We begin by considering models for random configurations including a finite number of points. As noted above, we can identify such a configuration $x = \{x_i \in S : 1 \leq i \leq m\}$ with a counting measure

$$\eta_x = \sum_{i=1}^{m} \delta_{x_i},$$

and we will denote the *size* or total mass of a configuration by $|x| = |\eta_x| = m$. Let $\boldsymbol{M}_p^f$ denote the collection of finite counting measures on S. (We will write $\boldsymbol{M}_p^f(S)$ if there is any ambiguity.) In general, we do not rule out the possibility of points of multiplicity greater than one; however, the assumption that ν in the definition of $\mathcal{A}_\theta$ is diffuse will usually ensure that π_θ is concentrated on configurations without points of multipliciy greater than one.

A configuration without points of multiplicity greater than one will be referred to as *simple,* and $\boldsymbol{M}_s^f \subset \boldsymbol{M}_p^f$ will denote the collection of simple counting measures, that is, $\eta \in \boldsymbol{M}_s^f$ implies $\eta\{x\} \leq 1$, for all $x \in S$. We topologize $\boldsymbol{M}_p^f$ with the weak topology and note that under appropriate metrics, both $\boldsymbol{M}_p^f$ and $\boldsymbol{M}_s^f$ are complete separable metric spaces. (See Appendix A.1.)

By a finite point process, we mean a random variable with values in $\boldsymbol{M}_p^f$, and by a finite, simple point process we mean a random variable with values in $\boldsymbol{M}_s^f$.

2.1 Conditional and unconditional models

Two types of models are considered in this paper. A model will be referred to as *conditional* if the size of the configuration is deterministic. Otherwise, the model is *unconditional.* The terminology reflects the fact that we can obtain a conditional model from an unconditional one by conditioning on the number of points in the configuration. For example, if ξ is a Poisson process with intensity measure ν_0 on S, that is, ξ satisfies

$$P\{\xi(A) = k\} = e^{-\nu_0(A)} \frac{\nu_0(A)^k}{k!}, \quad A \in \mathcal{B}(S), \tag{2.3}$$

then the conditional model obtained by conditioning on $|\xi| = m$ has the same distribution as the configuration $\{\xi_i \in S : 1 \leq i \leq m\}$ in which the $\{\xi_i\}$ are independent identically distributed points with distribution $\nu_0(\cdot)/\nu_0(S)$. Note that ξ is simple, that is, $P\{\xi \in \boldsymbol{M}_s^f\} = 1$, if ν_0 is diffuse.

More generally, we can define conditional and unconditional Gibbs models in terms of functions of the form

$$U_\theta(x) = \sum_{i=1}^{|x|} \sum_{\substack{x_{j_1},\ldots,x_{j_i} \in x \\ j_1 < j_2 \ldots < j_i}} \psi_i^\theta(x_{j_1}, \ldots, x_{j_i}), \tag{2.4}$$

where for each i, ψ_i^θ is a real-valued, symmetric function on S^i and may depend on a vector of parameters θ. $U_\theta(.)$ is called a potential function, and ψ_i^θ is called the ith-order interaction potential function. Let ν_0 be a diffuse, finite measure on S, and let $\pi_0 \in \mathcal{P}(\boldsymbol{M}_s^f)$ be the distribution of the corresponding Poisson process. The unconditional Gibbs model determined by ν_0 and U_θ has a distribution $\pi_\theta \in \mathcal{P}(\boldsymbol{M}_s^f)$ that is absolutely continuous with respect to the Poisson process distribution π_0 with density function p_θ given by

$$p_\theta(x) = \frac{d\pi_\theta}{d\pi_0}(x) = \frac{1}{\ell(\theta)} \exp\{-U_\theta(x)\}\,, \tag{2.5}$$

where $\ell(\theta)$ is a normalizing constant. The conditional Gibbs model $\pi_\theta^m \in \mathcal{P}(\boldsymbol{M}_s^f)$ with size m is absolutely continuous with respect to the distribution $\pi_0^m \in \mathcal{P}(\boldsymbol{M}_s^f)$ corresponding to m independent points in S with distribution $\nu_0(\cdot)/\nu_0(S)$. Its density p_θ satisfies

$$p_\theta^m(x) = \frac{d\pi_\theta^m}{d\pi_0^m}(x) = \frac{1}{\ell_m(\theta)} \exp\{-U_\theta(x)\}\,. \tag{2.6}$$

If $\psi_i^\theta = 0$, for $i \geq 3$, the model is the pairwise interaction model that has been developed and discussed by many authors. (See Ruelle (1969), Preston (1976), Strauss (1975), Lotwick and Silverman (1981), Baddeley and Lieshout (1995).)

2.2 Unconditional time-invariance models

The fact that a finite spatial point process can be represented as the stationary distribution of a spatial birth-and-death process has been exploited by a number of authors (for example, Preston (1977), Ripley (1979), Møller (1989), Geyer and Møller (1994)). We first consider birth and death processes with generators of the form

$$\begin{aligned} \mathcal{A}f(\eta) &= \int_S (f(\eta+\delta_u) - f(\eta))b(u,\eta)\nu(du) \\ &\qquad + \int_S (f(\eta-\delta_u) - f(\eta))d(u,\eta)\eta(du)\,, \end{aligned} \tag{2.7}$$

for a diffuse measure ν on S and nonnegative, measurable b, d defined on $S \times \boldsymbol{M}_p^f$. These models are a subclass of the spatial birth and death processes considered by Preston (1977). $b(u,\eta)$ gives the birth rate for a new point at u if the current configuration is η, and $d(u,\eta)$ gives the death rate for the point at u in the current configuration η. The assumption that ν is diffuse assures that if the process has no multiple points at time zero, then it will never have multiple points.

The generator $\mathcal{A}$ determines a Markov process as a solution of a martingale problem (see Section A.4). In order to apply standard results on Markov processes we need to be precise regarding the domain of $\mathcal{A}$. For simplicity, we will assume that there exist positive constants

$$B_m \geq \sup_{|\eta|=m} \int_S b(u,\eta)\nu(du) < \infty, \quad D_m \geq \sup_{|\eta|=m} \int_S d(u,\eta)\eta(du) < \infty. \tag{2.8}$$

Then we can take

$$\mathcal{D}(\mathcal{A}) = \{f \in B(\boldsymbol{M}_p^f) : \text{there exists } c_f \text{ such that } f(\eta) = 0 \text{ if } |\eta| > c_f\}.$$

Note that for $f \in \mathcal{D}(\mathcal{A})$, $\mathcal{A}f(\eta) = 0$ when $|\eta| > c_f + 1$, so

$$\|\mathcal{A}f\| \leq 2\|f\| \sup_{m \leq c_f+1} (B_m + D_m).$$

To avoid having an infinite number of jumps in a finite time interval, additional conditions on b and d must be imposed. Following Preston (1977), we assume that there are positive constants δ_m, $m \geq 1$, such that

$$0 < \delta_m \leq \inf_{|\eta|=m} \int_S d(u,\eta)\eta(du). \tag{2.9}$$

Preston (1977) compares spatial birth and death processes to the simple Markov birth and death process with generator

$$Cg(m) = B_m(g(m+1) - g(m)) + \delta_m(g(m-1) - g(m)) \tag{2.10}$$

for

$$g \in \mathcal{D}(C) = \{g : \text{there exists } c_g \text{ such that } g(m) = 0 \text{ for } m > c_g\}.$$

The martingale problem for C has a unique solution for every initial distribution if and only if no solution hits infinity in finite time. The following lemma is essentially a special case of Proposition 5.1 of Preston (1977).

Lemma 2.1 *Suppose that no solution of the martingale problem for C defined in (2.10) hits infinity in finite time. Then for each initial distribution $\kappa \in \mathcal{P}(\boldsymbol{M}_p^f)$, the martingale problem for $(\mathcal{A}, \kappa)$ has a unique solution.*

Our interest is in characterizing probability distributions in $\mathcal{P}(\boldsymbol{M}_p^f)$ as stationary distributions of spatial birth and death processes. The following result gives simple conditions that ensure $\mathcal{A}$ has a unique stationary distribution.

Theorem 2.2 *Suppose that $\delta_m > 0$, $m \geq 1$, and $B_m > 0$, $m \geq 0$, satisfy (2.8) and (2.9), and that*

$$\text{(a)} \qquad \sum_{m=1}^{\infty} \frac{\delta_1 \cdots \delta_m}{B_1 \cdots B_m} = \infty \,,$$

$$\text{(b)} \qquad \sum_{m=1}^{\infty} \frac{B_0 \cdots B_{m-1}}{\delta_1 \cdots \delta_m} < \infty \,.$$

Then there exists a unique stationary distribution for the spatial birth-and-death process with generator $\mathcal{A}$, and the stationary distribution is the unique $\pi \in \mathcal{P}(\boldsymbol{M}_p^f)$ satisfying

$$\int_S \mathcal{A} f \, d\pi = 0, \qquad f \in \mathcal{D}(\mathcal{A}). \tag{2.11}$$

Proof. Condition (a) implies that every solution of the martingale problem for C is recurrent, so by Lemma 2.1, uniqueness holds for the martingale problem for $\mathcal{A}$. Conditions (a) and (b) together imply that every solution of the martingale problem for C is positive recurrent which, by the coupling argument of Preston (1977), implies that the empty configuration ($\eta = \emptyset$) is a positive recurrent state for any solution of the martingale problem for $\mathcal{A}$. It follows that there is a unique stationary distribution for $\mathcal{A}$. A generalization of Echeverria's theorem (Kurtz and Stockbridge (1998), Theorem 3.1) gives that any $\pi \in \mathcal{P}(\boldsymbol{M}_p^f)$ satisfying (2.11) is a stationary distribution. □

2.3 Conditional time-invariance models

For the conditional case, the generator is

$$\mathcal{A} f(n) = \int_S \int_S (f(\eta + \delta_u - \delta_x) - f(\eta)) d(x, \eta) \eta(dx) b(u, \eta) \nu(du) \qquad \eta \in \boldsymbol{M}_p^m, \tag{2.12}$$

for a fixed positive integer m. (Here, $\boldsymbol{M}_p^m$ denotes the counting measures with total mass m.) Assuming (2.8), $\mathcal{A}$ is a bounded operator, and we can take $\mathcal{D}(\mathcal{A}) = B(\boldsymbol{M}_p^m)$, the bounded, measurable functions on $\boldsymbol{M}_p^m$. Note that each birth occurs simultaneously with a corresponding death, keeping the population size constant, so there is no state that is obviously recurrent. The process will, however, be Harris recurrent if b and d are strictly positive, so under that assumption, uniqueness of the the stationary distribution is assured, and the stationary distribution will be the unique $\pi \in \boldsymbol{M}_p^m$ satisfying (2.11).

3 Time-invariance estimation

Let $(\boldsymbol{M}, \mathcal{M})$ be a measurable space, and let $\{\pi_\theta(\cdot),\ \theta \in \Theta\} \subset \mathcal{P}(\boldsymbol{M})$ be a parametric family of models. Typically, Θ is a subset of $\mathbb{R}^k$, for some k, but that is not necessary. Our problem is to estimate θ from a single observation drawn from π_θ. Baddeley (2000) proposes a new, general method for estimating parameters, when the models π_θ can be characterized as stationary distributions of Markov processes.

3.1 General concept

For each $\theta \in \Theta$, let $\mathcal{A}_\theta$ be the generator of a Markov process with values in $\boldsymbol{M}$ for which π_θ is the unique stationary distribution. Typically, π_θ is characterized as the unique element of $\mathcal{P}(\boldsymbol{M})$ satisfying

$$\int_{\boldsymbol{M}} \mathcal{A}_\theta f d\pi_\theta = 0, \quad f \in \mathcal{D}(\mathcal{A}_\theta). \tag{3.1}$$

The domain $\mathcal{D}(\mathcal{A}_\theta)$ is usually taken to be a subset of $B(\boldsymbol{M})$; however, the definition of $\mathcal{A}_\theta$ frequently extends in a natural way to unbounded f (consider, for example, $\mathcal{A}_\theta$ given by (1.2)) with (3.1) continuing to hold. Consequently, we let $\hat{\mathcal{D}}(\mathcal{A}_\theta) \subset M(\boldsymbol{M})$ be the collection of $f \in M(\boldsymbol{M})$ such that there exist $f_m \in \mathcal{D}(\mathcal{A}_\theta)$ converging to f pointwise and satisfying

$$\lim_{m\to\infty} \int_{\boldsymbol{M}} |\mathcal{A}_\theta f_m - \mathcal{A}_\theta f| d\pi_\theta = 0.$$

It follows that (3.1) holds for $f \in \hat{\mathcal{D}}(\mathcal{A}_\theta)$.

For a 1-dimensional parameter θ, we choose a function $f \in \cap_\theta \hat{\mathcal{D}}(\mathcal{A}_\theta)$, and estimate θ by solving

$$\mathcal{A}_\theta f(\eta) = 0, \tag{3.2}$$

where $\eta \in \boldsymbol{M}$ is the single observation drawn from π_θ. Baddeley calls (3.2) a *time-invariance estimating equation*, and the solution of (3.2) a *time-invariance estimator* for θ. Since (3.1) holds, (3.2) is an unbiased estimating equation.

In general, for a multidimensional parameter θ, we choose a collection f_i, $i = 1, 2, \cdots, k$, where $f_i \in \cap_\theta \hat{\mathcal{D}}(\mathcal{A}_\theta)$ and k is the dimension of θ. The estimator θ then satisfies

$$\mathcal{A}_\theta f_i(\eta) = 0, \quad i = 1, \ldots k. \tag{3.3}$$

We must choose f_i, $i = 1, 2, \cdots, k$ carefully, to avoid inconsistencies among these k equations. If a solution of the system does not exist, Baddeley suggests minimizing

$$\sum_{i=1}^{k} (\mathcal{A}_\theta f_i(\eta))^2,$$

but we do not consider that option here. Note that (3.3) gives a system of unbiased estimating equations for θ.

The time-invariance estimator depends on the choice of $\mathcal{A}_\theta$ (since there may be many Markov processes with stationary distribution given by π_θ) and on the choice of the functions

f_i, $i = 1, \ldots, k$. For example, a spatial point process can be regarded as the stationary distribution of a spatial birth-and-death process in various ways. Baddeley (2000) applies the time-invariance method to a variety of statistical models, including discrete Markov random fields, spatial point processes, and the "dead leaves" model (see Serra (1984)).

We will require that $\mathcal{A}_\theta$ determine a unique, ergodic Markov process. By ergodicity, we mean that there exists a unique stationary distribution π_θ for $\mathcal{A}_\theta$ and that the corresponding process Y^θ converges in distribution to π_θ for every intitial distribution. Under this assumption observations from π_θ can, in principle, be simulated by Markov chain Monte Carlo.

3.2 Applications to spatial point process

For spatial point processes, the relationships between time-invariance and other methods of estimation (pseudo-likelihood and Takacs-Fiksel) have been discussed by Baddeley (2000) (see also Li (1999)). Baddeley gives some general discussion on the choice of the functions f_i; however, he does not provide any systematic discussion of the properties of particular classes of functions. We begin that discussion here.

We consider both unconditional and conditional models, and we begin with a class of models that has two parameters in the unconditional case and one in the essentially equivalent conditional case.

3.3 2-parameter families

Let π_θ be the unique stationary distribution of a spatial birth-and-death process whose generator is given by (1.2) with

$$b_\theta(u, \eta) = c\,\tilde{b}(u, \eta, a) \tag{3.4}$$

and $d_\theta(u, \eta) = 1$. We assume that $\tilde{b}$ is a positive function, and $\theta = (c, a) \in \Theta \subset (0, \infty) \times \mathbb{R}$. The corresponding conditional model is determined by the spatial birth-and-death process whose generator is given by (2.12) with

$$b_\theta(u, \eta) = \tilde{b}(u, \eta, a) \tag{3.5}$$

and $d_\theta(u, n) = 1$ with $\theta = a \in \Theta \subset \mathbb{R}$. Note that we are not claiming that the stationary distribution for the conditional model with $|\eta| = m$ can be obtained from the stationary distribution of the unconditional model, but for an appropriate choice of c, the two models should be qualitatively similar.

Note that there are two parameters for the unconditional model, and one for the conditional model. We will derive time-invariance estimators for both models.

3.3.1 Unconditional model

For the two-parameter unconditional model, a time-invariance estimator of θ is obtained by solving $\mathcal{A}_\theta f_i(\eta) = 0$, $i = 1, 2$, for two suitably selected functions f_1 and f_2. Since $\hat{a}_t$ and

$\hat{c}_t$, the time-invariance estimators of a and c, will depend on f_1 and f_2, we must carefully choose appropriate f_1 and f_2. In this study, we take

$$f_1(\eta) = |\eta| \tag{3.6}$$

and f_2 of the form

$$f_2(\eta) = \frac{1}{|\eta|(|\eta|-1)} \sum_{\substack{i \neq j \\ i,j=1,\dots,|\eta|}} h(x_i, x_j) \tag{3.7}$$

for a symmetric function h, $h(u,v) = h(v,u)$. (Here, $\eta = \sum_{i=1}^{|\eta|} \delta_{x_i}$.) The choice of h depends on the birth rate b_θ.

The choice of f_1 is quite natural. When $\tilde{b}(u,\eta,a) \equiv 1$, π_θ is a spatial Poisson process with mean measure $c\nu$. The time invariance estimator for c obtained using $f_1(\eta) = |\eta|$ is $\hat{c}_t = |\eta|/\nu(S)$ which is also the maximum likelihood estimator.

Since f_1 relates to the overall intensity of the process, f_2 should capture some feature of the relationship among the points in the configuration η, and f_2 of the form (3.7) relates naturally to pairwise interactions among the points. In addition, these choices of f_1 and f_2 simplify the calculation of $\mathcal{A}_\theta f_i(\eta)$. Setting $\mathcal{A}_\theta f_1(\eta) = 0$ yields

$$\begin{aligned} 0 &= \mathcal{A}_\theta f_1(\eta) \\ &= c\int_S ((|\eta|+1) - |\eta|)\tilde{b}(u,\eta,a)\nu(du) + \sum_{k=1}^{|\eta|}((|\eta|-1) - |\eta|) \\ &= c\int_S \tilde{b}(u,\eta,a)\nu(du) - |\eta|, \end{aligned} \tag{3.8}$$

giving

$$c\int_S \tilde{b}(u,\eta,a)\nu(du) = |\eta|\,. \tag{3.9}$$

For f_2 of the form (3.7) and $|\eta| > 2$,

$$\begin{aligned} \int_S (f_2(\eta-\delta_x) - f_2(\eta))\eta(dx) &= \sum_{k=1}^{|\eta|}(f_2(\eta-\delta_{x_k}) - f_2(\eta)) \\ &= \sum_{k=1}^{|\eta|}\Big(\frac{|\eta|(|\eta|-1)f_2(\eta) - 2\sum_{j\neq k} h(x_j,x_k))}{(|\eta|-1)(|\eta|-2)}\Big) - |\eta|\, f_2(\eta) \\ &= \frac{2|\eta|}{|\eta|-2} f_2(\eta) - \frac{2}{(|\eta|-1)(|\eta|-2)}\sum_{k=1}^{|\eta|}\sum_{j\neq k} h(x_j,x_k) \\ &= 0\,, \end{aligned} \tag{3.10}$$

where the next to last equality follows from the symmetry of h. Consequently, requiring $\mathcal{A}_\theta f_2(\eta) = 0$ yields

$$\begin{aligned} 0 &= \mathcal{A}_\theta f_2(\eta) \\ &= c\int_S (f_2(\eta+\delta_u) - f_2(\eta))\tilde{b}(u,\eta,a)\,\nu(du)\,. \end{aligned} \tag{3.11}$$

From (3.11), we have

$$\begin{aligned} 0 &= \mathcal{A}_\theta f_2(\eta) \\ &= c\int_S \left[\frac{|\eta|(|\eta|-1)}{|\eta|(|\eta|+1)} f_2(\eta) + 2\sum_{i=1}^{|\eta|} h(u,x_i) - f_2(\eta)\right] \tilde{b}(u,\eta,a)\nu(du) \\ &= \frac{2c}{|\eta|(|\eta|+1)} \int_S (\sum_{i=1}^{|\eta|} h(u,x_i) - |\eta|\, f_2(\eta))\tilde{b}(u,\eta,a)\nu(du)\,, \end{aligned} \tag{3.12}$$

and (3.12) yields

$$\frac{\int_S (\sum_{i=1}^{|\eta|} h(u,x_j))\tilde{b}(u,\eta,a)\nu(du)}{\int_S \tilde{b}(u,\eta,a)\nu(du)} - |\eta| f_2(\eta) = 0\ . \tag{3.13}$$

Of course, a solution of (3.13) may not exist or may not be unique.

If a solution $\hat{a}_t$ does exists, then the corresponding estimator for c, $\hat{c}_t$, can be obtained by substituting $\hat{a}_t$ in (3.9) to get

$$\hat{c}_t = \frac{|\eta|}{\int_S \tilde{b}(u,\eta,\hat{a}_t)\nu(du)}\ . \tag{3.14}$$

3.3.2 Conditional model

Since there is only one parameter for the conditional model, we need to solve $\mathcal{A}_\theta f = 0$ for a single function f. The generator is given by (2.12) with $b_\theta(u,\,\eta) = \tilde{b}(u,\,\eta,\,a)$ and $d_\theta(u,\,\eta) = 1$, and taking f_2 of the form (3.7), for $|\eta| > 1$, we have

$$\begin{aligned} 0 &= \mathcal{A}_\theta f_2(\eta) = \sum_{i=1}^{|\eta|} \int_S (f_2(\eta - \delta_{x_i} + \delta_u) - f_2(\eta))\tilde{b}(u,\,\eta,\,a)\nu(du) \\ &= \frac{2}{|\eta|(|\eta|-1)} \sum_{i=1}^{|\eta|} \int_S \sum_{\substack{j\neq i \\ j=1,\ldots,|\eta|}} (h(u,\,x_j) - h(x_i,\,x_j))\tilde{b}(u,\,\eta,\,a)\nu(du) \\ &= 2\int_S (\frac{\sum_{i=1}^{|\eta|} h(u,x_i)}{|\eta|} - f_2(\eta))\tilde{b}(u,\,\eta,\,a)\nu(du)\,. \end{aligned}$$

Consequently, $\hat{a}_t$ is the solution of (3.13), and the time-invariance estimator for a is the same regardless of whether we view the data as coming from the conditional or the unconditional model.

3.4 Examples

Example 3.1 Let

$$\tilde{b}_\theta(u,\,\eta) = \exp\{-a\sum_{j=1}^{|\eta|} \psi_2(u,\,x_j)\}, \tag{3.15}$$

where $a \geq 0$ and $\psi_2 \geq 0$. Then the corresponding π_θ gives a *pairwise interaction model*, a particular case of a Gibbs model.

The time-invariance estimators of c and a, $\hat{c}_t$ and $\hat{a}_t$, can be obtained by solving (3.13) and (3.14) for some function h. The solution of (3.13) also gives the time-invariance estimator of a in the conditional case.

For this model, the time-invariance approach gives the same family of estimators as the Takacs-Fiksel approach (see Baddeley (2000), Proposition 2). Taking $h = \psi_2$, that is,

$$f_2(\eta) = \frac{1}{|\eta|(|\eta|-1)} \sum_{\substack{i \neq j \\ i,\, j=1,\ldots,|\eta|}} \tilde{\psi}_2(x_i, x_j)\,,$$

the time-invariance estimator is essentially the same as the maximum pseudo-likelihood estimator. (See Baddeley (2000) and Li (1999) for more details and the results of a simulation study.) □

Example 3.2 Next we consider a family of models with nearest neighbor interactions. Let

$$b_\theta(u,\, \eta) = c_1 + c_2\, \mathbf{1}_{\{\min_{x_j} |u-x_j|<t_0\}}\,. \tag{3.16}$$

for a fixed $t_0 > 0$. Because b_θ is required to be non-negative, c_1 and c_2 must satisfy $c_1 \geq 0$ and $c_1 \geq -c_2$. Note that b_θ is bounded by c_1 when $-c_1 \leq c_2 < 0$ and by $c_1 + c_2$ when $c_2 \geq 0$. Since b_θ is bounded and we can take $\delta_m = m$, Theorem 2.2 implies that the birth and death process has a unique stationary distribution.

We assume that t_0 is fixed and known in the following discussion. That is, there are two parameters in this parametric family. Note that (3.16) can be rewritten as

$$b_\theta(u,\, \eta) = c(1 + a\, \mathbf{1}_{\{\min_{x_j} |u-x_j|<t_0\}}) \tag{3.17}$$

where $c = c_1$ and $a = c_2/c_1 \geq -1$, so that (3.16) has the form (3.4).

Taking $h(u,\, v) = \mathbf{1}_{\{|u-v|<t_0\}}$, that is,

$$f_2(\eta) = \frac{1}{|\eta|(|\eta|-1)} \sum_{\substack{i \neq j \\ i,\, j=1,\ldots,|\eta|}} \mathbf{1}_{\{|x_i-x_j|<t_0\}},$$

the estimating equations become linear and are solvable as long as $\int_S \mathbf{1}_{\{\min_{x_j} |u-x_j|<t_0\}}\, \nu(du) < \nu(S)$. When this inequality holds, the estimates for c_1 and c_2 become

$$\hat{c}_{1t} = \frac{\left(\int_S \sum_{i=1}^{|\eta|} \mathbf{1}_{\{|u-x_i|<t_0\}}\, \nu(du) - |\eta|\, f_2(\eta) \int_S \mathbf{1}_{\{\min_{x_j} |u-x_j|<t_0\}}\, \nu(du)\right) |\eta|}{\left(\int_S \sum_{i=1}^{|\eta|} \mathbf{1}_{\{|u-x_i|<t_0\}}\, \nu(du)\right) \left(\nu(S) - \int_S \mathbf{1}_{\{\min_{x_j} |u-x_j|<t_0\}}\, \nu(du)\right)}\,, \tag{3.18}$$

and

$$\hat{c}_{2t} = \frac{\left(|\eta|\, f_2(\eta)\nu(S) - \int_S \sum_{i=1}^{|\eta|} \mathbf{1}_{\{|u-x_i|<t_0\}}\, \nu(du)\right) |\eta|}{\left(\int_S \sum_{i=1}^{|\eta|} \mathbf{1}_{(|u-x_i|<t_0)}\, \nu(du)\right) \left(\nu(S) - \int_S \mathbf{1}_{\{\min_{x_j} |u-x_j|<t_0\}}\, \nu(du)\right)}\,. \tag{3.19}$$

For the conditional case, the estimator for $a = c_2/c_1$ is

$$\hat{a}_t = \frac{m\, f_2(\eta)\Big(\nu(S) - \int_S \mathbf{1}_{\{\min_j |u-x_j|<t_0\}}\, \nu(du)\Big)}{\int_S \sum_{i=1}^m \mathbf{1}_{\{|u-x_i|<t_0\}}\nu(du) - m f_2(\eta) \int_S \mathbf{1}_{\{\min_{x_j} |u-x_j|<t_0\}}\nu(du)}, \tag{3.20}$$

where m is the number of points in the configuration.

In this model, t_0 should really be treated as a parameter. Since we have not found a clear and direct way to estimate t_0, we have treated t_0 as a known constant; however, it is simple to compute estimates for c_1 and c_2 for several choices of t_0. Goodness-of-fit tests can then be used as a tool for selecting a suitable t_0. For more details see Li (1999). □

Example 3.3 The previous example can be generalized so that the birth rate becomes

$$b(u,\, \eta) = c_1 + c_2\, \rho(\frac{\min_{x_j} |u-x_j|}{t_0})\,, \tag{3.21}$$

where ρ is a non-negative function bounded by a constant k. Since b is required to be non-negative, we assume that $c_1 > 0$ and $c_1 + c_2\, k > 0$. b is bounded by c_1 if $-c_1/k < c_2 < 0$ and by $c_1 + c_2 k$ if $c_2 > 0$. Again, by Theorem 2.2, there exists a unique stationary distribution for the spatial birth-and-death process.

As in the previous example, we evaluate the time-invariance estimator under the assumption that t_0 is a known positive constant. Then a suitable value of t_0 can be selected using goodness-of-fit tests.

In this model, we suggest taking $h(u,\, v) = \rho^*(\frac{|u-v|}{t_0})$, that is,

$$f_2(\eta) = \frac{1}{|\eta|(|\eta|-1)} \sum_{\substack{i\neq j \\ i,\, j=1,\ldots,|\eta|}} \rho^*(\frac{|u-v|}{t_0})\,,$$

where ρ^* is a non-negative function. Setting

$$C_1(\eta) = \int_S \sum_{i=1}^{|\eta|} \rho^*(\frac{|u-x_i|}{t_0})\rho(\frac{\min_{x_j} |u-x_j|}{t_0})\nu(du)$$

and

$$C_2(\eta) = \int_S \sum_{i=1}^{|\eta|} \rho^*(\frac{|u-x_i|}{t_0})\nu(du) \int_S \rho(\frac{\min_{x_j} |u-x_j|}{t_0})\nu(du)$$

For the unconditional model, the time-invariance estimators are

$$\hat{c}_{1t} = \frac{\Big(\int_S \sum_{i=1}^{|\eta|} \rho^*(\frac{|u-x_i|}{t_0})\rho(\frac{\min_{x_j} |u-x_j|}{t_0})\nu(du) - |\eta| f_2(\eta) \int_S \rho(\frac{\min_{x_j} |u-x_j|}{t_0})\nu(du)\Big)\, |\eta|}{\nu(S)C_1(\eta) - C_2(\eta)} \tag{3.22}$$

and

$$\hat{c}_{2t} = \frac{\Big(|\eta| f_2(\eta)\nu(S) - \int_S \sum_{i=1}^{|\eta|} \rho^*(\frac{|u-x_i|}{t_0})\nu(du)\Big)\, |\eta|}{\nu(S)C_1(\eta) - C_2(\eta)}\,. \tag{3.23}$$

provided the denominator is not equal to zero. For the conditional case, the time-invariance estimator of $a = c_2/c_1$ is

$$\hat{a}_t = \frac{\Big(mf_2(\eta)\nu(S) - \int_S\sum_{i=1}^m \rho^*(\frac{|u-x_i|}{t_0})\nu(du)\Big)\, m}{\Big(\int_S\sum_{i=1}^m \rho^*(\frac{|u-x_i|}{t_0})\rho(\frac{\min_{x_j}|u-x_j|}{t_0})\nu(du) - mf_2(\eta)\int_S\rho(\frac{\min_{x_j}|u-x_j|}{t_0})\nu(du)\Big)\, m}, \tag{3.24}$$

where m is the number of points in the configuration. □

Typically the models considered in Examples 3.2 and 3.3 do not have explicit forms for the density function. Consequently, it would be difficult to apply classical methods of estimation to these models; however, the qualitative properties of the models can be readily inferred from the form of the birth rate, Markov chain Monte Carlo gives a simulation method, and the parameters can be estimated using the time-invariance approach.

4 Consistency of time-invariance estimators

There are at least two ways to formulate consistency results for these estimators. First, one could conceive of observing multiple realizations, $\eta_1, \dots, \eta_n$ (say independent), of the same point process in a fixed region, and then obtaining the estimates by solving

$$\frac{1}{n}\sum_{j=1}^n \mathcal{A}_\theta f_i(\eta_j) = 0, \quad i = 1, \dots, k.$$

If the true value of the parameter is θ_0, then as $n \to \infty$, the left side converges to

$$\int_S \mathcal{A}_\theta f_i(\eta)\pi_{\theta_0}(d\eta), \quad i = 1, \dots, k, \tag{4.1}$$

which, of course, vanishes at $\theta = \theta_0$. Assuming that θ_0 is the unique value of θ for which (4.1) vanishes, consistency would follow under appropriate continuity and compactness conditions.

The second approach to consistency, and the one in which we are interested, assumes that the data comes by observing the point process in a finite subregion of an infinite region. The consistency question then becomes whether or not the parameter estimates converge to the correct value as data is collected from larger and larger subregions.

The birth rates specifying Examples 3.2 and 3.3 have a translation invariance property that will be made precise below, and Example 3.1 does also, provided $\psi_2(u, v) = \tilde{\psi}(u - v)$ for some function $\tilde{\psi}$. Intuitively, at least, the birth and death processes make sense on any $S \subset \mathbb{R}^d$ including $S = \mathbb{R}^d$. Consequently, we can consider analogous models on larger and larger regions. The simulation results given by Li (1999) show increasing accuracy with increasing size of the region, so consistency, in some sense, is plausible.

In the next subsection, we review (spatial) ergodicity for spatial point processes. These results imply almost sure convergence for a large class of appropriately normalized statistics which in turn implies consistency for a variety of time-invariance estimators.

4.1 Ergodicity for spatial processes

Following Nguyen and Zessin (1979), we review results on the ergodicity of spatial processes in $\mathbb{R}^d$. Let $\mathcal{B}_b(\mathbb{R}^d)$ be the family of bounded Borel subsets in $\mathbb{R}^d$, and ν be Lebesgue measure on $\mathbb{R}^d$. (We will write du rather than $\nu(du)$.) Let $\mathcal{K}$ be the collection of convex and bounded subsets of $\mathbb{R}^d$, and for $G \in \mathcal{K}$, let $r_0(G)$ denote the supremum of the radii of spheres contained in G. Let $\mathbb{Z}^d$ denote the set of points in $\mathbb{R}^d$ with integer coordinates. For $G \in \mathcal{B}(\mathbb{R}^d)$, let $G + v$ denote the translate of G by v, that is, $G + v = \{(x + v) : x \in G\}$.

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $\mathcal{G} = \{T_u : u \in \mathbb{R}^d\}$ be a family of $\mathcal{F}$-measurable, 1-1, measure preserving transformations $T_u : \Omega \to \Omega$ (that is, $P(T_v^{-1}E) = P(E)$ for all $E \in \mathcal{F}$, $v \in \mathbb{R}^d$) satisfying $T_{u+v} = T_u \circ T_v$, $u, v \in \mathbb{R}^d$, so $T_u^{-1} = T_{-u}$ and T_0 is the identity mapping. Let $\mathcal{I}$ denote the σ-algebra of (almost surely) $\mathcal{G}$-invariant events, that is

$$\mathcal{I} = \{E \in \mathcal{F} : P(E \triangle T_v^{-1}E) = 0, v \in \mathbb{R}^d\}.$$

$\mathcal{I}$ is *trivial* if $P(E) = 0$ or 1 for each $E \in \mathcal{I}$. If $\mathcal{I}$ is trivial, then $\mathcal{G}$ is called *ergodic*. Note also that if $\mathcal{I}$ is trivial (that is, $\mathcal{G}$ is ergodic),

$$E[Z|\mathcal{I}] = E[Z] \quad a.s.$$

for all $Z \in L^1(P)$.

A *spatial process* is a family of random variables indexed by $\mathcal{B}_b(\mathbb{R}^d)$, $\{X_G : G \in \mathcal{B}_b(\mathbb{R}^d)\}$, with $X_\emptyset = 0$. A random variable Y is $\mathcal{G}$*-invariant* (or just invariant if the context is clear) if $Y \circ T_u = Y$ a.s., for all $u \in \mathbb{R}^d$, and a spatial process $\{X_G : G \in \mathcal{B}_b(\mathbb{R}^d)\}$ is $\mathcal{G}$*-covariant* if

$$X_{G+u} \circ T_u = X_G \quad a.s., \quad \forall u \in \mathbb{R}^d.$$

Y is $\mathcal{G}$-invariant if and only if Y is $\mathcal{I}$ measurable.

Let $F_0 = \left[-\frac{1}{2}, \frac{1}{2}\right)^d$. Order $\mathbb{Z}^d$ lexicographically, that is $u = (u_1, \ldots, u_d) < v = (v_1, \ldots, v_d)$ if $u_1 < v_1$ or if $u_1 = v_1$ and $u_2 < v_2$ or if $u_1 = v_1$, $u_2 = v_2$, and $u_3 < v_3$, etc. Define

$$F_+ = \cup\{F_0 + u : u \in \mathbb{Z}^d, u < 0\}.$$

Note that for $v \in \mathbb{Z}^d$

$$F_+ + v = \cup\{F_0 + u : u \in \mathbb{Z}^d, u < v\}$$

and that $F_0 \cap F_+ = \emptyset$, so

$$(F_+ \cup F_0 + v) - (F_+ + v) = F_0 + v.$$

Observing that $\{F_0 + v : v \in \mathbb{Z}^d\}$ is a partition of $\mathbb{R}^d$, if $\{X_G : G \in \mathcal{B}_b(\mathbb{R}^d)\}$ is covariant, then

$$\begin{aligned} X_G &= \sum_{u \in \mathbb{Z}^d, (F_0+u)\cap G \neq \emptyset} \left(X_{G\cap(F_+\cup F_0+u)} - X_{G\cap(F_++u)}\right) \\ &= \sum_{u \in \mathbb{Z}^d, F_0\cap(G-u) \neq \emptyset} \left(X_{(G-u)\cap(F_+\cup F_0)} \circ T_{-u} - X_{(G-u)\cap F_+} \circ T_{-u}\right). \end{aligned} \tag{4.2}$$

A spatial process is *additive* if $X_{G_1 \cup G_2} = X_{G_1} + X_{G_2}$ for $G_1, G_2 \in \mathcal{B}_b(\mathbb{R}^d)$, $G_1 \cap G_2 = \emptyset$. If X_G is additive and covariant, then (4.2) is just

$$\begin{aligned} X_G &= \sum_{u \in \mathbb{Z}^d, (F_0+u)\cap G \neq \emptyset} X_{G \cap (F_0+u)} \\ &= \sum_{u \in \mathbb{Z}^d, F_0 \cap (G-u) \neq \emptyset} X_{(G-u)\cap F_0} \circ T_{-u} \; . \end{aligned} \tag{4.3}$$

The following theorem is essentially Theorem 4.10 of Nguyen and Zessin (1979).

Theorem 4.1 *Let $\{X_G, \, G \in \mathcal{B}_b(\mathbb{R}^d)\}$ be a covariant spatial process. Suppose that there is a nonnegative, integrable random variable Y such that*

$$|X_{G\cup\Lambda} - X_\Lambda| \leq Y \quad a.s. \tag{4.4}$$

for each $\Lambda \in \mathcal{B}_b(\mathbb{R}^d)$, with $\Lambda \subset F_+$ and each $G \in \mathcal{K}$ with $G \subset F_0$. Assume there exists $Z \in L^1(P)$ such that

$$\lim_{\substack{\Lambda \to F_+ \\ \Lambda \in \mathcal{B}_b(\mathbb{R}^d), \Lambda \subset F_+}} X_{\Lambda \cup F_0} - X_\Lambda = Z \quad a.s., \tag{4.5}$$

in the sense that for almost all $\omega \in \Omega$ and $\epsilon > 0$, there is an integer $m = m(\omega, \epsilon)$ such that $|X_{\Lambda\cup F_0}(\omega) - X_\Lambda(\omega) - Z(\omega)| < \epsilon$, holds for $\Lambda \in \mathcal{B}_b(\mathbb{R}^d)$, with $[-\frac{m+1}{2}, \frac{m+1}{2})^d \cap F_+ \subset \Lambda \subset F_+$.

Let $\{K_t, t \geq 0\} \subset \mathcal{K}$ satisfy $K_{t_1} \subset K_{t_2}$, for $t_2 > t_1$, and $\lim_{t\to\infty} r_0(K_t) = \infty$. Then

$$\lim_{t\to\infty} \frac{X_{K_t}}{\nu(K_t)} = E[Z|\mathcal{I}] \quad a.s. \text{ and in } L^1. \tag{4.6}$$

□

Remark 4.2 Each term on the right of (4.2) is bounded by $Y \circ T_{-u}$, so

$$|X_G| \leq \sum_{u \in \mathbb{Z}^d, F_0 \cap (G-u) \neq \emptyset} Y \circ T_{-u}.$$

This inequality implies uniform integrability for $\{X_K/\nu(K) : K \in \mathcal{K}, \nu(K) > 1\}$, so almost sure convergence in (4.6) implies convergence in L^1. □

If X_G is additive, the statement of the result simplifies. (The following is essentially Corollary 4.20 of Nguyen and Zessin (1979).)

Corollary 4.3 *Let $\{X_G, G \in \mathcal{B}_b(\mathbb{R}^d)\}$ be an additive, covariant spatial process. Suppose that there is a nonnegative, integrable random variable Y such that*

$$|X_G| \leq Y \quad a.s.$$

for each each $G \in \mathcal{K}$ with $G \subset F_0$. Then for $\{K_t\}$ as above

$$\lim_{t\to\infty} \frac{X_{K_t}}{\nu(K_t)} = E[X_{F_0} \,|\, \mathcal{I}] \quad a.s. \tag{4.7}$$

In particular, if Z is an integrable random variable, then

$$\lim_{t\to\infty} \frac{1}{\nu(K_t)} \int_{K_t} Z(T_{-u})\nu(du) = E[Z|\mathcal{I}] \ a.s. \text{ and in } L^1. \tag{4.8}$$

□

For $\eta = \sum_i \delta_{x_i} \in \boldsymbol{M}_p$ and $u \in \mathbb{R}^d$, define

$$T_u\eta = \sum_i \delta_{x_i+u}, \tag{4.9}$$

or equivalently, by setting $T_u\eta(A) = \eta(A-u)$ for $A \in \mathcal{B}(\mathbb{R}^d)$. For $G \in \mathcal{B}_b(\mathbb{R}^d)$, let $f_G \in M(\boldsymbol{M}_p)$. We will call $\{f_G : G \in \mathcal{B}_b(\mathbb{R}^d)\}$ *covariant* if $f_G(\eta) = f_{G+u}(T_u\eta)$ for all $G \in \mathcal{B}_b(\mathbb{R}^d)$, $\eta \in \boldsymbol{M}_p$, and $u \in \mathbb{R}^d$. A set $A \in \mathcal{B}(\boldsymbol{M}_p)$ is *invariant* if $\eta \in A$ implies $T_u\eta \in A$ for all $u \in \mathbb{R}^d$. Let $\mathcal{I}(\boldsymbol{M}_p)$ denote the collection (a σ-algebra) of invariant sets.

A point process ξ on $\mathbb{R}^d$ is stationary if for all choices of $G_1, \ldots, G_k \in \mathcal{B}_b(\mathbb{R}^d)$, the distribution of $(\xi(G_1+u), \ldots, \xi(G_k+u))$ does not depend on u, or equivalently, the distribution of $T_u\xi$ does not depend on u. If $\xi(G)$ is integrable, then $E[\xi(G)] = \lambda\nu(G)$, for some $\lambda \geq 0$, and λ is called the *mean intensity* for the point process. We say that ξ is *ergodic* if for each $A \in \mathcal{I}(\boldsymbol{M}_p)$, $P\{\xi \in A\}$ is 0 or 1.

We can restate Theorem 4.1 for covariant functions of stationary point processes.

Corollary 4.4 *Let ξ be a stationary point process on $\mathbb{R}^d$, and let $\{f_G \in M(\boldsymbol{M}_p) : G \in \mathcal{B}_b(\mathbb{R}^d)\}$ be a covariant family of functions. Define $X_G \equiv f_G(\xi)$ and $\mathcal{I}$ to be the completion of the σ-algebra $\{\{\xi \in A\} : A \in \mathcal{I}(\boldsymbol{M}_p)\}$. If (4.4) and (4.5) hold, then (4.6) holds.*

In particular, if ξ has finite mean intensity, then for K_t as above

$$\lim_{t\to\infty} \frac{\xi(K_t)}{\nu(K_t)} = E[N(F_0)\,|\,\mathcal{I}] \quad a.s. \tag{4.10}$$

□

Remark 4.5 The point of Corollary 4.4 is that T_u is now defined on $\boldsymbol{M}_p$ rather than Ω. The corollary follows simply by defining a new probability space $(\boldsymbol{M}_p, \mathcal{B}(\boldsymbol{M}_p), P_\xi)$, where P_ξ is the distribution of ξ. □

The proof of the next corollary requires the following simple lemma.

Lemma 4.6 *Let $0 \leq \rho(r) \leq C < \infty$ and $\lim_{r\to\infty} \rho(r) = 0$. Then for $\{K_t\}$ as above,*

$$\lim_{t\to\infty} \frac{1}{\nu(K_t)} \int_{K_t} \rho(d(u, K_t^c))du = 0.$$

Proof. The lemma follows from the fact that

$$\lim_{t\to\infty} \frac{\nu(\{u \in K_t : d(u, K_t^c) < r\})}{\nu(K_t)} = 0.$$

□

Corollary 4.7 *Let ξ be a stationary point process on $\mathbb{R}^d$, and define ξ^G by $\xi^G(A) = \xi(A \cap G)$. Let $H : \mathbb{R}^d \times \boldsymbol{M}_p \to \mathbb{R}$ satisfy $H(u, \eta) = H(0, T_{-u}\eta)$, and suppose that there exists $\gamma : \mathbb{R}^d \times \boldsymbol{M}_p \to [0, \infty)$ such that*

$$|H(u, \eta + \delta_v) - H(u, \eta)| \leq \gamma(v - u, T_{-u}\eta),$$

$\eta_1 \subset \eta_2$ implies $\gamma(u, \eta_1) \leq \gamma(u, \eta_2)$, and

$$E[\int_{\mathbb{R}^d} \gamma(v, \xi)\xi(dv)] + E[\int_{F_+} \int_{F_0} \gamma(v - u, T_{-u}\xi)\xi(dv)du] < \infty. \tag{4.11}$$

Then for $\{K_t\}$ as above,

$$\begin{aligned}
&\lim_{t\to\infty} \frac{1}{\nu(K_t)} \int_{K_t} H(u, \xi^{K_t})du \qquad (4.12)\\
&= \lim_{t\to\infty} \frac{1}{\nu(K_t)} \int_{K_t} H(u, \xi)du \\
&= E[\int_{F_0} H(u, \xi^{F_+ \cup F_0})du|\mathcal{I}] + E[\int_{F_+} (H(u, \xi^{F_+ \cup F_0}) - H(u, \xi^{F_+}))du|\mathcal{I}] \\
&= E[H(0, \xi)|\mathcal{I}].
\end{aligned}$$

Proof. Setting

$$X_G = \int_G H(u, \xi^G)du,$$

for $\Lambda \subset F_+$ and $G \subset F_0$, we have

$$\begin{aligned}
|X_{G\cup\Lambda} - X_\Lambda| &\leq \int_G |H(u, \xi^{G\cup\Lambda})|du + \int_\Lambda |H(u, \xi^{G\cup\Lambda}) - H(u, \xi^\Lambda)|du \\
&\leq |H(0, \emptyset)|\nu(G) + \int_G \int_{G\cup\Lambda} \gamma(v - u, T_{-u}\xi^{G\cup\Lambda})\xi(dv)du \\
&\qquad + \int_\Lambda \int_G \gamma(v - u, T_{-u}\xi^{G\cup\Lambda})\xi(dv)du \\
&\leq |H(0, \emptyset)| + \int_{F_0} \int_{\mathbb{R}^d} \gamma(v, T_{-u}\xi)T_{-u}\xi(dv)du + \int_{F_+} \int_{F_0} \gamma(v - u, T_{-u}\xi)\xi(dv)du,
\end{aligned}$$

which, by (4.11), gives (4.4). Then the first expression in (4.12) equals the third expression by Theorem 4.1. Similarly, the second equals the fourth by Corollary 4.3. It remains only to show that the two limits are the same. But

$$|\int_{K_t} (H(u, \xi) - H(u, \xi^{K_t}))du| \leq \int_{K_t} |H(u, \xi) - H(u, \xi^{K_t})|du \leq \int_{K_t} \int_{K_t^c} \gamma(v - u, T_{-u}\xi)\xi(dv)du,$$

and

$$\begin{aligned}
E[\frac{1}{\nu(K_t)} \int_{K_t} \int_{K_t^c} \gamma(v - u, T_{-u}\xi)\xi(dv)du] &= \frac{1}{\nu(K_t)} \int_{K_t} E[\int_{T_{-u}K_t^c} \gamma(v, \xi)\xi(dv)]du \\
&\leq \frac{1}{\nu(K_t)} \int_{K_t} \rho(d(u, K_t^c))du, \qquad (4.13)
\end{aligned}$$

where

$$\rho(r) = E[\int_{B_r(0)^c} \gamma(v,\xi)\xi(dv)]$$

and $B_r(0)$ is the ball of radius r centered at 0. Since, by (4.11) and the dominated covergence theorem, $\lim_{r\to\infty}\rho(r) = 0$, the right side of (4.13) goes to zero by Lemma 4.6. $\square$

The proof of the following corollary is similar.

Corollary 4.8 *Let ξ be a stationary point process on $\mathbb{R}^d$, and define ξ^G by $\xi^G(A) = \xi(A\cap G)$. Let $H : \mathbb{R}^d \times \boldsymbol{M}_p \to \mathbb{R}$ satisfy $H(u,\eta) = H(0, T_{-u}\eta)$, and suppose that there exists $\gamma : \mathbb{R}^d \times \boldsymbol{M}_p \to [0,\infty)$ such that*

$$|H(u,\eta+\delta_v) - H(u,\eta)| \le \gamma(v-u, T_{-u}\eta),$$

$\eta_1 \subset \eta_2$ implies $\gamma(u,\eta_1) \le \gamma(u,\eta_2)$,

$$\begin{aligned} E[\xi(F_0)] &+ E[\int_{F_0}\int_{F_0} \gamma(v-u, T_{-u}\xi)\xi(dv)\xi(du)] \qquad (4.14)\\ &+E[\int_{F_0}\int_{F_+} \gamma(v-u, T_{-u}\xi)\xi(dv)\xi(du)] + E[(\int_{F_+}\int_{F_0} \gamma(v-u, T_{-u}\xi)\xi(dv)\xi(du)] < \infty, \end{aligned}$$

and for $\{K_t\}$ as above

$$\lim_{t\to\infty} \frac{1}{\nu(K_t)} E[\int_{K_t}\int_{K_t^c} \gamma(v-u, T_{-u}\xi)\xi(dv)\xi(du)] = 0. \qquad (4.15)$$

Then

$$\begin{aligned} &\lim_{t\to\infty} \frac{1}{\nu(K_t)} \int_{K_t} H(u,\xi^{K_t})\xi(du) \qquad (4.16)\\ &\quad = \lim_{t\to\infty} \frac{1}{\nu(K_t)} \int_{K_t} H(u,\xi)\xi(du)\\ &\quad = E[\int_{F_0} H(u,\xi^{F_+\cup F_0})\xi(du)|\mathcal{I}] + E[\int_{F_+} (H(u,\xi^{F_+\cup F_0}) - H(u,\xi^{F_+}))\xi(du)|\mathcal{I}]\\ &\quad = E[\int_{F_0} H(u,\xi)\xi(du)|\mathcal{I}]. \end{aligned}$$

Proof. Setting

$$X_G = \int_G H(u,\xi^G)\xi(du),$$

for $\Lambda \subset F_+$ and $G \subset F_0$, we have

$$\begin{aligned} &|X_{G\cup\Lambda} - X_\Lambda|\\ &\quad \le X_G + \int_G |H(u,\xi^{G\cup\Lambda}) - H(u,\xi^G)|\xi(du) + \int_\Lambda |H(u,\xi^{G\cup\Lambda}) - H(u,\xi^\Lambda)|\xi(du)\\ &\quad \le |H(0,\emptyset)|\xi(G) + \int_G\int_G \gamma(v-u, T_{-u}\xi^G)\xi(dv)\xi(du) \end{aligned}$$

$$+\int_G\int_\Lambda \gamma(v-u,T_{-u}\xi^{G\cup\Lambda})\xi(dv)\xi(du)+\int_\Lambda\int_G \gamma(v-u,T_{-u}\xi^{G\cup\Lambda})\xi(dv)\xi(du)$$

$$\leq |H(0,\emptyset)|\xi(F_0)+\int_{F_0}\int_{F_0}\gamma(v-u,T_{-u}\xi^{F_0})\xi(dv)\xi(du)$$

$$+\int_{F_0}\int_{F_+}\gamma(v-u,T_{-u}\xi)\xi(dv)\xi(du)+\int_{F_+}\int_{F_0}\gamma(v-u,T_{-u}\xi)\xi(dv)\xi(du),$$

which, by (4.14), gives (4.4).

The first equality on the right of (4.16) follows from (4.15), and the other equalities follow by Theorem 4.1 and Corollary 4.3. □

4.2 Relationship between finite and infinite space models

Let S be locally compact, and let ν be a Radon measure on S, that is, $\nu(K)<\infty$ for each compact $K\subset S$. Let the generator $\mathcal{A}$ to be of the form

$$\mathcal{A}f(\eta)=\int_S (f(\eta+\delta_u)-f(\eta))b(u,\,\eta)\,\nu(du)+\int_S (f(\eta-\delta_u)-f(\eta))\eta(du) \tag{4.17}$$

for f in an appropriate domain $\mathcal{D}(\mathcal{A})$. We assume that b is bounded by a constant, although this assumption could be weakened.

To be precise, assume that $\mathcal{D}(\mathcal{A})$ consists of functions of the form

$$f(\eta)=g_1(\eta)\exp^{-\int_{\mathbb{R}^d} g_2(x)\eta(dx)},$$

where $g_1:\boldsymbol{M}_p\to\mathbb{R}$ and $g_2:S\to[0,\infty)$ are bounded and continuous and there exist compact sets $K_1\subset K_2\subset S$ such that

$$g_1(\eta)=g_1(\eta^{K_1}),\quad \inf_{x\in K_1} g_2(x)=\epsilon_1>0,\quad \mathrm{supp} g_2\subset K_2. \tag{4.18}$$

We then have

$$\mathcal{A}f(\eta)=\int_{K_2}(f(\eta+\delta_u)-f(\eta))b(u,\eta)\nu(du)+\int_{K_2}(g_1(\eta-\delta_u)e^{g_2(u)}-g_1(\eta))e^{-\int_{K_2} g(x)\eta(dx)}\eta(du),$$

and we see that

$$\begin{aligned}|\mathcal{A}f(\eta)| \;\leq\; & 2\|g_1\|_\infty\int_{K_2} b(u,\eta)du+2\|g_1\|_\infty\eta(K_1)e^{-\epsilon_1\eta(K_1)}\\ & +\|g_1\|_\infty e^{\|g_2\|_\infty}\int_{K_2} g_2(x)\eta(dx)e^{-\int_{K_2} g_2(x)\eta(dx)},\end{aligned}$$

so $\mathcal{A}f$ is a bounded function. Note also that $\mathcal{D}(\mathcal{A})$ is a subset of $\bar{C}(\boldsymbol{M}_p)$ and is closed under multiplication (that is, if $f_1,\,f_2\in\mathcal{D}(\mathcal{A})$, $f_1f_2\in\mathcal{D}(\mathcal{A})$).

Let $K_1\subset K_2\subset\cdots$ be open sets with compact closure in S satisfying $\cup_m K_m=S$, and let $\mathcal{A}_m$ be the generator for the birth and death process with the birth rate b on the region

K_m. Thus,

$$\mathcal{A}_m f(\eta) = \int_{K_m} (f(\eta+\delta_u) - f(\eta)) b(u,\eta)\,\nu(du) + \int_{K_m} (f(\eta-\delta_u) - f(\eta))\eta(du)\,. \tag{4.19}$$

We can take $\mathcal{D}(\mathcal{A}_m)$ to be the collection of functions on $\boldsymbol{M}_p^f(K_m)$ obtained by restricting the functions in $\mathcal{D}(\mathcal{A})$. (This domain is different from the one used in Section 2.2; however, it is simple to see that a solution of the martingale problem for the operator with one of the domains will be a solution of the martingale problem for the operator with the other domain. Consequently, the two formulations determine the same Markov process.)

Let π_m be the stationary distribution for $\mathcal{A}_m$. By Theorem 2.2, π_m exists and is unique and satisfies $\int \mathcal{A}_m f\, d\pi_m = 0$ for all $f \in \mathcal{D}(\mathcal{A}_m)$.

Lemma 4.9 *If b is bounded by a constant C and the mapping $\eta \to b(u,\eta)$ is continuous for u, then $\{\pi_m\}$ is relatively compact and any limit point is a stationary distribution for $\mathcal{A}$.*

Proof. Let N_m be a point process with distribution π_m. For $A \in \mathcal{B}_b(\mathbb{R}^d)$, let $f(\eta) = \eta(A)$. Approximating f by functions in $\mathcal{D}(\mathcal{A}_m)$, we see that $\int \mathcal{A}_m f\, d\pi_m = 0$. It follows that

$$E[N_m(A)] = \int \eta(A)\,\pi_m(d\eta) = \int_{K_m\cap A} b(u,\eta)\nu(du)\pi_m(d\eta) \le C\nu(A)\,, \tag{4.20}$$

and hence

$$P\{N_m(A) > t\} \le \frac{E[N_m(A)]}{t} \le \frac{C\nu(A)}{t}\,. \tag{4.21}$$

We can view N_m as taking values in $\boldsymbol{M_p}$ by defining $N_m(A) = 0$ for all $A \in \mathcal{B}(S)$ with $A \cap K_m = \emptyset$. By Lemma A.2, $\{N_m\}$ is relatively compact in distribution in the vague topology; hence, $\{\pi_m\}$ considered as a sequence in $\mathcal{P}(\boldsymbol{M_p})\}$ is relatively compact.

Let $\pi \in \mathcal{P}(\boldsymbol{M_p})$ be a limit point of $\{\pi_m\}$, and for simplicity, assume $\pi_m \Rightarrow \pi$. For $f \in \mathcal{D}(\mathcal{A})$, let K_1 and K_2 be compact sets as in (4.18). Then

$$\int_{\boldsymbol{M}_p^f} |\mathcal{A}f - \mathcal{A}_m f| d\pi_m = \int_{\boldsymbol{M}_p^f} |\int_{K_2\cap K_m^c} (f(\eta+\delta_u) - f(\eta)) b(u,\eta)\nu(du)| \pi_m(d\eta),$$

and for m sufficiently large, the right side is zero. The boundedness and continuity assumptions on b imply $\mathcal{A}f \in \bar{C}(\boldsymbol{M}_p)$, and it follows that

$$\int_{\boldsymbol{M}_p} \mathcal{A}f d\pi = \lim_{m\to\infty} \int_{\boldsymbol{M}_p} \mathcal{A}_m f d\pi_m = 0.$$

Consequently, by Theorem 3.1 of Kurtz and Stockbridge (1998), π is a stationary distribution for $\mathcal{A}$. □

4.3 Consistency for spatially stationary models

We now take $S = \mathbb{R}^d$ and ν to be Lebesgue measure. We assume that b_θ, $\theta \in \Theta$, is translation invariant in the sense that $b_\theta(u,\eta) = b_\theta(0, T_{-u}\eta) \equiv b_\theta(T_{-u}\eta)$. Let θ_0 be the "true" value of the parameter, and let π be a stationary distribution for $\mathcal{A}_{\theta_0}$. Let N be a spatial point process with distribution π. We assume that N is (spatially) stationary and ergodic. The stationary distribution of $\mathcal{A}_{\theta_0}$ is not necessarily unique, and there may exist stationary distributions that do not satisfy this assumption. Conditions for the existence of a spatially ergodic stationary distribution are discussed in Garcia and Kurtz (n.d.).

Let $\{K_m\}$ be an increasing sequence of convex, open sets with compact closure satisfying $\cup_m K_m = \mathbb{R}^d$. Let N_m be the restriction of N to the set K_m, that is, $N_m(A) = N(A \cap K_m)$, $A \in \mathcal{B}(\mathbb{R}^d)$. The notion of consistency we will consider is as follows: For each m, we obtain a time-invariance estimator $\hat{\theta}_m$ for the parameter as the solution of a system of equations of the form

$$\mathcal{A}_{\theta,m} f_i(N_m) = 0, \quad i = 1, \dots, \kappa, \tag{4.22}$$

where $\mathcal{A}_{\theta,m}$ has the form (4.19) with b replaced by b_θ. Note that the distribution of N_m will, in general, not be the stationary distribution for $\mathcal{A}_{\theta_0,m}$, so the estimating equations (4.22) are not necessarily unbiased; however, by the results discussed in Section 4.2, they should, in some sense, be asymptotically unbiased.

We want to show that for a class of functions f, $\mathcal{A}_{\theta,m} f(N_m)$, appropriately normalized, converges to a deterministic funtion of θ and that these deterministic functions set equal to 0 give a system of equations having θ_0 as the unique solution. We then need to show that the solution of (4.22) converges to the solution of the limiting equations.

First consider $f_1(\eta) = |\eta|$. Then, as in (3.8),

$$\begin{aligned}
\mathcal{A}_{\theta,m} f_1(N_m) &= \int_{K_m} b_\theta(u, N_m)du - |N_m| \\
&= \int_{K_m} b_\theta(T_{-u}N_m)du - N(K_m) \\
&= \int_{K_m} (b_\theta(T_{-u}N_m) - b_\theta(T_{-u}N))du + \int_{K_m} b_\theta(T_{-u}N)du - N(K_m).
\end{aligned}$$

Lemma 4.10 *Suppose that b_θ is bounded and that there exists $\gamma_\theta \in L^1(\mathbb{R}^d)$ such that for each $\eta \in \boldsymbol{M}_p$ and $v \in \mathbb{R}^d$,*

$$|b_\theta(u,\eta) - b_\theta(u, \eta + \delta_v)| \le \gamma_\theta(u - v). \tag{4.23}$$

Then

$$\lim_{m\to\infty} \frac{1}{|K_m|} \mathcal{A}_{\theta,m} f_1(N_m) = \int_{\boldsymbol{M}_p} (b_\theta(\eta) - \eta(F_0))\pi(d\eta), \tag{4.24}$$

almost surely and in L^1.

Remark 4.11 If we set $g_1(\eta) = \eta(F_0)$, then $\mathcal{A}_\theta g_1(\eta) = \int_{F_0} b_\theta(T_{-u}\eta)du - \eta(F_0)$, so the right side of (4.24) is just $\int \mathcal{A}_\theta g_1 d\pi$.

Proof. Let N^G denote N restricted to G (so $N_m = N^{K_m}$). Define

$$X_G = \int_G b_\theta(u, N^G)du,$$

and let $\bar{b}_\theta = \sup_\eta b_\theta(\eta)$. Then

$$\begin{aligned} |X_{G\cup\Lambda} - X_\Lambda| &\leq \int_G b_\theta(u, N^{G\cup\Lambda})du + \int_\Lambda |b_\theta(u, N^{G\cup\Lambda}) - b_\theta(u, N^\Lambda)|du \\ &\leq \bar{b}_\theta \nu(G) + \int_\Lambda \int_G \gamma_\theta(u-v)N(dv)du \\ &\leq \bar{b}_\theta \nu(G) + \|\gamma_\theta\|_1 N(G), \end{aligned}$$

and hence for $G \subset F_0$,

$$|X_{G\cup\Lambda} - X_\Lambda| \leq \bar{b}_\theta + \|\gamma_\theta\|_1 N(F_0),$$

giving (4.4). In addition,

$$\begin{aligned} Z &\equiv \lim_{\Lambda\to F_+} (X_{F_0\cup\Lambda} - X_\Lambda) \\ &= \lim_{\Lambda\to F_+} \int_{F_0} b_\theta(u, N^{F_0\cup\Lambda})du + \lim_{\Lambda\to F_+} \int_\Lambda (b_\theta(u, N^{F_0\cup\Lambda}) - b_\theta(u, N^\Lambda))du \\ &= \int_{F_0} b_\theta(u, N^{F_0\cup F_+})du + \int_{F_+} (b_\theta(u, N^{F_0\cup F_+}) - b_\theta(u, N^{F_+}))du, \end{aligned}$$

where the second integrand is integrable by (4.23). Consequently, Corollary 4.4 and the ergodicity of N give

$$\lim_{m\to\infty} \frac{1}{K_m} \int_{K_m} b_\theta(u, N_m)du = E[Z] = \lim_{m\to\infty} \frac{1}{K_m} \int_{K_m} b_\theta(u, N)du = \int b_\theta(\eta)\pi(d\eta).$$

□

Now, for $\eta \in \boldsymbol{M}_s$ and $h : (\mathbb{R}^d)^k \to \mathbb{R}$ symmetric and translation invariant, let

$$f_h(\eta) = \frac{1}{\binom{|\eta|}{k}} \sum_{\{x_1,\ldots,x_k\}\in\eta} h(x_1, \ldots, x_k), \tag{4.25}$$

where the sum is over all subsets of k points in the support of η, and

$$f_h^0(\eta) = \frac{1}{\binom{|\eta|}{k-1}} \sum_{\{x_1,\ldots,x_{k-1}\}\in\eta} h(0, x_1, \ldots, x_{k-1}).$$

As in the case $k = 2$, the "death term" in the generator applied to f_h vanishes, and

$$\begin{aligned} \mathcal{A}_{\theta,m} f_h(N_m) &= \int_{K_m} \Bigg(\left(\frac{(|N_m| - k + 1)}{(|N_m| + 1)} - 1 \right) f_h(N_m) \\ &\quad + \frac{1}{\binom{|N_m|+1}{k}} \sum_{\{x_1,\ldots,x_{k-1}\}\in N_m} h(u, x_1, \ldots, x_{k-1}) \Bigg) b_\theta(T_{-u}N_m)du) \\ &= \int_{K_m} \left(\frac{k}{|N_m| + 1} (f_h^0(T_{-u}N_m) - f_h(N_m)) \right) b_\theta(T_{-u}N_m)du\ . \end{aligned}$$

Lemma 4.12 *Let $h : (\mathbb{R}^d)^k \to \mathbb{R}$ be bounded, symmetric and translation invariant, that is, $h(x_1, \ldots, x_k) = h(x_{\sigma_1}, \ldots, x_{\sigma_k})$ for all permutations $\sigma = (\sigma_1, \ldots, \sigma_k)$ of $(1, \ldots, k)$ and $h(x_1 + u, \cdots, x_k + u) = h(x_1, \ldots, x_k)$ for $u \in \mathbb{R}^d$, and assume that there exists C_h such that $\max_{1\le i<j\le k} |x_i - x_j| > C_h$ implies $h(x_1, \ldots, x_k) = 0$. Then*

a)

$$\lim_{m\to\infty} \frac{1}{\nu(K_m)} \sum_{\{x_1,\ldots,x_k\}\in N_m} h(x_1, \ldots, x_k) = E[Z_h] = \frac{1}{k} E[\int_{F_0} H(u, N)N(du)],$$

where

$$Z_h = \sum_{l=1}^{k} \sum_{\{y_1,\ldots,y_l\}\in N^{F_0}, \{x_1,\ldots,x_{k-l}\}\in N^{F_+}} h(y_1, \ldots, y_l, x_1, \ldots, x_{k-l}).$$

(Note that the summands in the definition of Z_h are zero unless $d(x_i, F_0) \le C_h$, $i = 1, \ldots, k - l$.)

b)

$$\lim_{m\to\infty} \frac{1}{\nu(K_m)} \binom{|N_m|}{k} \mathcal{A}_{\theta,m} f_h(N_m) \tag{4.26}$$
$$= E[H(0, N)b_\theta(N)] - E[\int_{F_0} H(u, N)N(du)] \frac{E[b_\theta(N)]}{E[N(F_0)]}$$

where

$$H(u, \eta) = \sum_{\{x_1,\ldots,x_{k-1}\}\in\eta} h(u, x_1, \ldots, x_{k-1}) \prod_{i=1}^{k-1} \mathbf{1}_{\{x_i\ne u\}}.$$

Proof. Let

$$X_G = \sum_{\{x_1,\ldots,x_k\}\in N^G} h(x_1, \ldots, x_k) = \frac{1}{k} \int_G H(u, N^G)N(du).$$

Since

$$|H(u, \eta+\delta_v) - H(u, \eta)| \le \mathbf{1}_{\{|u-v|\le C_h\}} \eta(B_{C_h}(u))^{k-2} \|h\|_\infty = \mathbf{1}_{\{|u-v|\le C_h\}} T_{-u}\eta(B_{C_h}(0))^{k-2} \|h\|_\infty,$$

we can take $\gamma(z, \eta) = \mathbf{1}_{\{|z|\le C_h\}} \eta(B_{C_h}(0))^{k-1} \|h\|_\infty$, and since $\int \eta(K)^l \pi(d\eta) < \infty$ for each $K \in \mathcal{B}_b(\mathbb{R}^d)$ and each l, (4.14) is satisfied.

Since

$$\lim_{\substack{\Lambda\to F_+ \\ \Lambda\in\mathcal{B}_b(\mathbb{R}^d),\Lambda\subset F_+}} X_{\Lambda\cup F_0} - X_\Lambda = Z_h \quad a.s., \tag{4.27}$$

Corollary 4.8 gives Part (a). Since

$$H(u, \eta) = H(0, T_{-u}\eta) = \binom{|\eta|}{k-1} f_h^0(T_{-u}\eta)$$

for almost every u,

$$\int H(u,\eta)\eta(du) = k\binom{|\eta|}{k} f_h(\eta)$$

and

$$\lim_{m\to\infty} \frac{1}{|N_m|+1}\int_{K_m} b_\theta(T_{-u}N_m)du = \frac{E[b_\theta(N)]}{E[N(F_0)]},$$

Part (b) follows from Part (a) and Corollary 4.7. □

Lemma 4.13 *Let* $F_m = [-\frac{m}{2}, \frac{m}{2}]^d$*, and define*

$$g_m(\eta) = \sum_{\{x_1,\ldots,x_k\}\subset\eta} h(x_1,\ldots,x_k)\prod_{i=1}^{k}\mathbf{1}_{F_m}(x_i).$$

Then

$$\begin{aligned}\mathcal{A}_\theta g_m(\eta) \;&=\; \int_{F_m}\sum_{\{x_1,\ldots,x_{k-1}\}\subset\eta} h(u,x_1,\ldots,x_{k-1})\prod_{i=1}^{k-1}\mathbf{1}_{F_m}(x_i)b_\theta(T_{-u}\eta)du \\ &\quad -\int_{F_m}\sum_{\{x_1,\ldots,x_{k-1}\}\subset\eta} h(u,x_1,\ldots,x_{k-1})\prod_{i=1}^{k-1}\mathbf{1}_{F_m-\{u\}}(x_i)\eta(du),\end{aligned} \tag{4.28}$$

and

$$\lim_{m\to\infty}\frac{1}{m^d}\mathcal{A}_\theta g_m(N) = \lim_{m\to\infty}\frac{1}{m^d}\int_{\boldsymbol{M}_p}\mathcal{A}_\theta g_m d\pi = E[H(0,N)b_\theta(N)] - E[\int_{F_0} H(u,N)N(du)] \; a.s.$$

Proof. The convergence of the first expression on the right of (4.28) normalized by m^d follows by observing that

$$\begin{aligned}&\frac{1}{m^d}\int_{F_m}\sum_{\{x_1,\ldots,x_{k-1}\}\subset\eta} h(u,x_1,\ldots,x_{k-1})b_\theta(T_{-u}\eta)du \\ &\quad -\frac{1}{m^d}\int_{F_m}\sum_{\{x_1,\ldots,x_{k-1}\}\subset\eta} h(u,x_1,\ldots,x_{k-1})\prod_{i=1}^{k-1}\mathbf{1}_{F_m}(x_i)b_\theta(T_{-u}\eta)du\end{aligned}$$

converges to zero, is bounded by

$$\frac{1}{m^d}\int_{F_m}\sum_{\{x_1,\ldots,x_{k-1}\}\subset\eta} |h(u,x_1,\ldots,x_{k-1})|b_\theta(T_{-u}\eta)du,$$

hence uniformly integrable with respect π, and

$$\lim_{m\to\infty}\frac{1}{m^d}\int_{F_m}\sum_{\{x_1,\ldots,x_{k-1}\}\subset N} h(u,x_1,\ldots,x_{k-1})b_\theta(T_{-u}N)du = E[H(0,N)b_\theta(N)] \; a.s$$

The second expression in (4.28), normalized by m^d, converges to $E[\int_{F_0} H(u,N)N(du)]$ by Part (a) of Lemma 4.12. □

Combining the lemmas, we have a theorem that implies consistency for many models and appropriately selected time-invariance estimators obtained as solutions of systems of the form

$$\begin{aligned} \int_{K_m} b_{\hat{\theta}_m}(u, N_m)du &= |N_m| \\ \mathcal{A}_{\hat{\theta}_m,m} f_{h_i}(N_m) &= 0, \quad i = 2, \ldots, \kappa \ , \end{aligned}$$

where the $h_i : (\mathbb{R}^d)^{k_i} \to \mathbb{R}$ are bounded, symmetric, and translation invariant. We do not give general conditions for consistency, but a central condition would be that the true parameter is the unique solution of

$$\begin{aligned} E[b_\theta(N)] &= E[N(F_0)] \\ E[H_i(0,N)b_\theta(N)] &= E[\int_{F_0} H_i(u,N)N(du)], \quad i = 2, \ldots, \kappa \ , \end{aligned} \tag{4.29}$$

where

$$H_i(u, \eta) = \sum_{\{x_1,\ldots,x_{k_i-1}\}\in\eta} h_i(u, x_1, \ldots, x_{k_i-1}) \prod_{j=1}^{k-1} \mathbf{1}_{\{x_j \neq u\}}.$$

Of course, the true parameter will satisfy (4.29), so the issue is the uniqueness of the solution.

Theorem 4.14 *Let $h : (\mathbb{R}^d)^k \to \mathbb{R}$ be bounded, symmetric and translation invariant, and assume that there exists C_h such that $\max_{1\le i<j\le k} |x_i - x_j| > C_h$ implies $h(x_1, \ldots, x_k) = 0$, and let f_h be given by (4.25). Let π be a spatially ergodic stationary distribution for $\mathcal{A}_{\theta_0}$ (so $\int_{\boldsymbol{M}_p} \mathcal{A}_{\theta_0} g d\pi = 0$, $g \in \mathcal{D}(\mathcal{A}_{\theta_0})$). Then for K_m as above*

$$\begin{aligned} \lim_{m\to\infty} \frac{1}{\nu(K_m)} \left(\int_{K_m} b_{\theta_0}(u, N_m)du - |N_m| \right) &= \lim_{m\to\infty} \frac{1}{\nu(K_m)} E[\int_{K_m} b_{\theta_0}(u, N_m)du - |N_m|] \\ &= E[b_{\theta_0}(N)] - E[N(F_0)] \\ &= 0 \end{aligned}$$

and

$$\lim_{m\to\infty} \frac{1}{\nu(K_m)} \binom{|N_m|}{k} \mathcal{A}_{\theta_0,m} f_h(N_m) = 0.$$

4.4 Two-parameter example

We apply the results of Section 4.3 to prove consistency of the estimators for a family of 2-parameter models. Let $\rho : \mathbb{R} \longrightarrow [0, \infty)$ be a non-negative function bounded by a constant $\bar{\rho}$, and let

$$b_\theta(u, \eta) = c_1 + c_2 \, \rho(\frac{\min_{x\in\eta} |u - x|}{t_0}), \tag{4.30}$$

where $c_1 > 0$, $c_1 + c_2 \, k > 0$, and $t_0 > 0$. t_0 is assumed known, so $\theta = (c_1, c_2)$.

We discussed this model in Example 3.3, and obtained the time-invariance estimator for c_1 and c_2, taking

$$f_1(\eta) = |\eta|, \quad f_2(\eta) = \frac{1}{\binom{|\eta|}{2}} \sum_{\{x_1,x_2\}} \rho^*(\frac{x_1 - x_2}{t_0}) \; .$$

To simplify notation, let

$$\tilde{\rho}(u,\eta) = \rho(\frac{\min_{x\in\eta}|u-x|}{t_0}).$$

Then the estimator $\hat{\theta}^m = (\hat{c}_1^m, \hat{c}_2^m)$ satisfies

$$\mathcal{A}_{\theta,m} f_1(N_m) = \int_{K_m} (c_1 + c_2\,\tilde{\rho}(u,N_m))du - |N_m| = 0,$$

and, observing that

$$f_1^0(\eta) = \frac{1}{|\eta|}\int \rho^*(\frac{x}{t_0})\eta(dx),$$

$$\begin{aligned}
&\mathcal{A}_{\theta,m} f_2(N_m) \\
&= \int_{K_m} \left(\frac{2}{|N_m|+1}(\frac{1}{|N_m|}\int \rho^*(\frac{u-x}{t_0})N_m(dx) - f_2(N_m))\right)(c_1 + c_2\,\tilde{\rho}(u,N_m))du \\
&= 0\;.
\end{aligned}$$

Normalizing and passing to the limit, we have

$$\lim_{m\to\infty}\frac{1}{\nu(K_m)}\mathcal{A}_{\theta,m} f_1(N_m) = \int_{\boldsymbol{M}_p} \mathcal{A}_\theta g_1 d\pi = \int_{\boldsymbol{M}_p} (c_1 + c_2\,\tilde{\rho}(0,\eta) - \eta(F_0))\pi(d\eta),$$

where $g_1(\eta) = \eta(F_0)$, and

$$\begin{aligned}
\lim_{m\to\infty}\frac{|N_m|(|N_m|-1)}{2\nu(K_m)}\mathcal{A}_{\theta,m} f_2(N_m) &= \int_{\boldsymbol{M}_p}\int_{F_0}\int_{\mathbb{R}^d} \rho^*(\frac{u-x}{t_0})\eta(dx)(c_1 + c_2\tilde{\rho}(u,\eta))du\pi(d\eta) \\
&\quad - \int_{\boldsymbol{M}_p}\int_{F_0}\int_{\mathbb{R}^d} \rho^*(\frac{u-x}{t_0})\eta(dx)\eta(du)\pi(d\eta)\,.
\end{aligned}$$

Consequently, if the limiting system is nonsingular, the solutions must converge, and we have the following theorem.

Theorem 4.15 *Assume that*

$$\begin{aligned}
&\int_{\boldsymbol{M}_p}\int_{F_0}\int_{\mathbb{R}^d} \rho^*(\frac{u-x}{t_0})\eta(dx)\tilde{\rho}(u,\eta)du\pi(d\eta) \\
&\quad - \int_{\boldsymbol{M}_p} \tilde{\rho}(0,\eta)\pi(d\eta) \times \int_{\boldsymbol{M}_p}\int_{F_0}\int_{\mathbb{R}^d} \rho^*(\frac{u-x}{t_0})\eta(dx)du\pi(d\eta) \neq 0.
\end{aligned}$$

Then the true parameter $\theta_0 = (c_1^0, c_2^0)$ is the unique solution of the system

$$\begin{aligned} \int_{\boldsymbol{M}_p} (c_1 + c_2\,\tilde{\rho}(0,\eta) - \eta(F_0))\pi(d\eta) &= 0 \\ \int_{\boldsymbol{M}_p}\int_{F_0}\int_{\mathbb{R}^d} \rho^*(\frac{u-x}{t_0})\eta(dx)(c_1 + c_2\tilde{\rho}(u,\eta))du\pi(d\eta) & \\ - \int_{\boldsymbol{M}_p}\int_{F_0}\int_{\mathbb{R}^d} \rho^*(\frac{u-x}{t_0})\eta(dx)\eta(du)\pi(d\eta) &= 0, \end{aligned}$$

and $(\hat{c}_1^m, \hat{c}_2^m) \to (c_1^0, c_2^0)$ almost surely.

A Appendix

A.1 Point configurations and counting measures

Let (S, r_S) be a complete, separable metric space, $\mathcal{B}(S)$ be the σ-algebra of Borel sets in S, and $\mathcal{B}_b(S)$ be the family of all relatively compact sets in $\mathcal{B}(S)$. $\bar{C}(S)$ will denote the space of bounded continous functions and $B(S)$ the space of bounded Borel-measurable functions on S.

A finite *configuration* in S will be a finite collection of points $x = \{x_i \in S, 1 \leq i \leq m\}$, where $|x| = m$ is the *size* of the configuration. We do not rule out the possibility that $x_i = x_j$ for some $i \neq j$. The indexing is only for convenience and we identify two configurations that contain the same points with the same multiplicities. It follows that we can represent any configuration as an integer-valued measure

$$\eta_x = \sum_{i=1}^{m} \delta_{x_i},$$

where δ_{x_i} denotes the measure that places mass one at the point $x_i \in S$. $\boldsymbol{M}_p^f$ will denote the space of finite, integer-valued measures on S which we identify with the space of finite configurations in S. $\boldsymbol{M}_p$ will denote the space of (possibly infinite) integer-valued measures on S satisfying $\eta(A) < \infty$ for all $A \in \mathcal{B}_b(S)\}$. A configuration will be called *simple* if no point has multiplicity greater than one. The collection of simple configurations can be identified with

$$\boldsymbol{M}_s = \{\eta \in \boldsymbol{M}_p : \eta(\{x\}) \leq 1 \text{ for all } x \in S\},$$

with $\boldsymbol{M}_s^f$ defined analogously.

A natural metric for the space of finite configurations is given by

$$d_p(x, \tilde{x}) = \left\{ \begin{array}{cc} 1 & |x| \neq |\tilde{x}| \\ 1 \wedge \inf_{\sigma \in \Sigma_{|x|}} \sum_{i=1}^{m} r_S(x_i, \tilde{x}_{\sigma_i}) & |x| = |\tilde{x}| \end{array} \right. ,$$

where Σ_m is the collection of all permutations of $\{1, \ldots, m\}$. (Note that the value of $d_p(x, \tilde{x})$ does not depend on the indexing of x. The metric d_p determines the weak topology on $\boldsymbol{M}_p^f$, that is, $d_p(x^n, x) \to 0$ if and only if

$$\int_S f d\eta_{x^n} \to \int_S f d\eta_x, \qquad f \in \bar{C}(S).$$

Note that under d_p, $\boldsymbol{M}_p^f$ is a complete separable metric space, but $\boldsymbol{M}_s^f$ is not complete. With this last observation in mind, for $x, \tilde{x} \in \boldsymbol{M}_s^f$, define

$$d_s(x,\tilde{x}) = d_p(x,\tilde{x}) + \left| \frac{1}{\inf_{1\le i\ne j\le |x|} r_S(x_i,x_j)} - \frac{1}{\inf_{1\le i\ne j\le |\tilde{x}|} r_S(\tilde{x}_i,\tilde{x}_j)} \right|.$$

Then, under d_s, $\boldsymbol{M}_s^f$ is complete, but the topology determined by d_s on $\boldsymbol{M}_s^f$ is still the weak topology.

A.2 The vague topology

Assume that S is locally compact. (We are primarily interested in $S \subset \mathbb{R}^d$.) Let $\hat{C}(S)$ denote the space of continuous functions vanishing at infinite and $C_c(S)$, the space of continuous functions with compact support. A sequence $\{\eta_n\} \in \boldsymbol{M}_p$ converges in the *vague topology* if and only if

$$\int_S f d\eta_n \to \int_S f d\eta, \qquad f \in C_c(S).$$

$\boldsymbol{M}_p$ with the vague topology is Polish, that is, there exists a metric d_v giving the topology under which $\boldsymbol{M}_p$ is complete and separable. (See Kallenberg (1983).)

Note that if we restrict the vague topology to $\boldsymbol{M}_p^f$, we do *not* get the weak topology. For example, for $S = \mathbb{R}$, $\delta_0 + \delta_n$ converges to δ_0 in the vague topology but not in the weak topology.

A.3 Convergence in distribution

Definition A.1 *Let N, N_1, N_2, ... be point processes on S. Then $\{N_m\}$ converges in distribution (in the vague topology) to N, written $N_m \Rightarrow N$, if and only if $E[g(N_m)] \longrightarrow E[g(N)]$ for every bounded, vaguely continuous function $g : \boldsymbol{M_p} \longrightarrow \mathbb{R}$.* □

Note that if π_m is the distribution of N_m, then $\{N_m\}$ converges in distribution to N if and only if π_m converges weakly to the distribution of N.

Prohorov's theorem states that a sequence of random variables $\{X_n\}$ in a complete, separable metric space is relatively compact for convergence in distribution if and only if it is *tight*, that is, for each $\epsilon > 0$, there exists a compact subset K_ϵ such that

$$\inf_n P\{X_n \in K_\epsilon\} \ge 1 - \epsilon.$$

In particular, a sequence of $\mathbb{R}$-valued random varibles is tight if and only if

$$\lim_{r\to\infty} \limsup_{n\to\infty} P\{|X_n| > r\} = 0. \tag{A.1}$$

Lemma A.2 *Let $\{N_m\}$ be point processes in a locally compact space S considered as random variables in $(\boldsymbol{M}, d_v)$. Then the following statements are equivalent.*
(i) $\{N_m\}$ is tight.
(ii) $\{N_m(f)\}$ is tight for each $f \in C_c(S)$.
(iii) $\{N_m(A)\}$ is tight for each $A \in \mathcal{B}_b(S)$.

Proof. See Kallenberg (1983), Lemma 4.5. □

A.4 Markov processes and martingale problems

Let (E, r) be a metric space, $M(E)$ be the collection of all real-valued, Borel measurable functions on E, and $B(E) \subseteq M(E)$ be the Banach space of bounded functions with $||f|| = \sup_{x \in E} |f(x)|$. In addition, let $\mathcal{P}(E)$ be the collection of probability measures on E.

Let $\mathcal{A} : \mathcal{D}(A) \subset B(E) \to B(E)$ be a linear operator. An E-valued stochastic process $Y \equiv \{Y(t), t \geq 0\}$ is a solution of the *martingale problem* for $\mathcal{A}$ if and only if

$$f(Y(t)) - \int_0^t \mathcal{A}f(Y(s))ds$$

is a $\{\mathcal{F}_t^Y\}$-martingale for each $f \in \mathcal{D}(\mathcal{A})$. (Here, $\{\mathcal{F}_t^Y\}$ is the filtration determined by Y, that is, $\mathcal{F}_t^Y = \sigma(Y(s); s \leq t)$.) Y has *initial distribution* $\mu \in \mathcal{P}(E)$ if $Y(0)$ has distribution μ. *Uniqueness* in distribution holds for the martingale problem if any two solutions with the same initial distribution have the same finite dimensional distributions. If uniqueness holds, then any solution of the martingale problem is a *Markov process*, that is,

$$P(Y(t+s) \in \Gamma \mid \mathcal{F}_t^Y) = P(Y(t+s) \in \Gamma \mid Y(t)) \,, \tag{A.2}$$

for all $s, t \geq 0$ and $\Gamma \in \mathcal{B}(E)$.

Specifying the generator essentially determines the short-time behavior of the process, since the martingale property implies

$$E[f(Y(t + \Delta t))|\mathcal{F}_t^Y] \approx f(Y(t)) + \mathcal{A}f(Y(t))\Delta t.$$

Uniqueness implies that the short-time behavior determines the global behavior of the process.

For example, the generator for a pure-jump, Markov process has the form

$$\mathcal{A}f(x) = \lambda(x) \int_{\mathcal{E}} (f(y) - f(x))\, \kappa(x, dy) \,, \tag{A.3}$$

for some nonnegative function λ on E and a transition function κ on $E \times \mathcal{B}(E)$. If $Y(t) = x$, then the probability that the process jumps before time $t + \Delta t$ is approximately $\lambda(x)\Delta t$, and if it jumps, $\kappa(x, \cdot)$ is the distribution of the new value.

π is a *stationary distribution* for $\mathcal{A}$ if and only if there exists a solution of the martingale problem for $\mathcal{A}$ with initial distribution π such that

$$P\{Y(t+s_1) \in \Gamma_1, Y(t+s_2) \in \Gamma_2, \cdots, Y(t+s_k) \in \Gamma_k)\}$$

is independent of $t \geq 0$, for all $k \geq 1$, $0 \leq s_1 < s_2 < \cdots < s_k$ and $\Gamma_1, \Gamma_2, \cdots, \Gamma_k \in \mathcal{B}(E)$. Of course, if uniqueness holds there is only one such solution.

If π is a stationary distribution for $\mathcal{A}$, the martingale property implies

$$E[\mathcal{A}f(Y(t))] = \int_E \mathcal{A}f d\pi = 0.$$

Under mild conditions on the domain (most importantly, that the domain is closed under multiplication) and the operator (that it satisfies a form of the postive maximum principle), the converse also holds. The converse is essentially due to Echeverría (1982) for locally convex E (see Ethier and Kurtz (1986), Theorem 4.9.17). Bhatt and Karandikar (1993) extended the results to general complete, separable metric spaces, and Kurtz and Stockbridge (1998) removed continuity assumptions on the range of $\mathcal{A}$.

References

Baddeley, A. and Lieshout, M. N. M. V. (1995). Area-interaction point processes, *Ann. Inst. Statist. Math.* **47**: 601–619.

Baddeley, A. J. (2000). Time-invariance estimating equations, *Bernoulli* **6**(5): 783–808.

Bhatt, A. G. and Karandikar, R. L. (1993). Invariant measures and evolution equations for Markov processes characterized via martingale problems, *Ann. Probab.* **21**(4): 2246–2268.

Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processess. Vol. I*, Springer-Verlag, New York.

Diggle, P. J. (1983). *Statistical Analysis of Spatial Point Patterns*, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London.

Diggle, P. J., Fiksel, T., Grabarnik, P., Ogata, Y., Stoyan, D. and Tanemura, M. (1994). On parameter estimation for pairwise interaction point processes, *Int. Statist. Rev.* **62**: 99–117.

Echeverría, P. (1982). A criterion for invariant measures of Markov processes, *Z. Wahrsch. Verw. Gebiete* **61**(1): 1–16.

Ethier, S. N. and Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence*, John Wiley & Sons, New York.

Garcia, N. L. and Kurtz, T. G. (n.d.). Spatial birth and death processes as solutions of stochastic equations.

Geyer, C. J. and Møller, J. (1994). Simulation procedures and likelihood inference for spatial point processes, *Scand. J. Statist.* pp. 359–373.

Jensen, J. L. (1993). Asymptotic normality of estimates in spatial point processes, *Scand, J. Statist.* **20**: 97–109.

Jensen, J. L. and Künsch, H. R. (1994). On asymptotic normality of pseudo-likelihood estimates for pairwise interaction processes, *Ann. Inst. Statist. Math.* **46**(3): 475–486.

Jensen, J. L. and Møller, J. (1991). Pseudo-likelihood for exponential family models of spatial point processes, *Ann. Appl. Probab.* **1**: 445–461.

Kallenberg, O. (1983). *Random Measures*, Akademie-Verlag, Berlin. and Academic Press, London.

Kurtz, T. G. and Stockbridge, R. H. (1998). Existence of Markov controls and characterization of optimal Markov controls, *SIAM J. Control Optim.* **36**(2): 609–653 (electronic).

Li, S.-H. (1999). *Stationary Distributions of Markov Processes as Statistical Models: Baddeley's Time-Invariance Method of Estimation*, PhD thesis, University of Wisconsin - Madison.

Li, S.-H. (n.d.). Title to be determined, In preparation.

Lotwick, H. W. and Silverman, B. W. (1981). Convergence of spatial birth-and-death processes, *Math. Proc. Camb. Phil. Soc.* **90**: 155–165.

Møller, J. (1989). On the rate of convergence of spatial birth-and-death processes, *Ann. Inst. Statist. Math.* **41**: 565–581.

Nguyen, X. X. and Zessin, H. (1979). Ergodic theorems for spatial processes, *Z. Wahr. Verw. Gebiete* **48**: 133–158.

Preston, C. J. (1976). *Random Fields*, Spring-Verlag, Berlin. Lecture Notes in Math. 534.

Preston, C. J. (1977). Spatial birth-and-death processes, *Bull. Int. Statist. Inst.* **46**(2): 371–391.

Ripley, B. D. (1979). [algorithm as 137] Simulating spatial patterns: Dependent samples from a multivariate density, *Appl. Statist.* **28**: 109–112.

Ruelle, D. (1969). *Statistical Mechanics*, Wiley, New York.

Serra, J. (1984). *Image analysis and mathematical morphology*, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London. English version revised by Noel Cressie.

Strauss, D. J. (1975). A model for clustering, *Biometrika* **62**: 467–476.