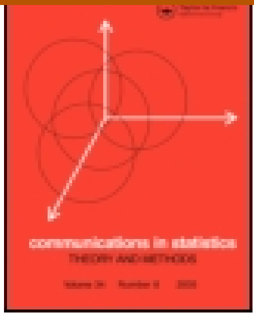

Communications in Statistics - Theory and Methods

ISSN: 0361-0926 (Print) 1532-415X (Online) Journal homepage: http://www.tandfonline.com/loi/lsta20

# The Discretely Observed Immigration-Death Process: Likelihood Inference and Spatio-Temporal Applications

Ottmar Cronie & Jun Yu

**To cite this article:** Ottmar Cronie & Jun Yu (2015): The Discretely Observed Immigration-Death Process: Likelihood Inference and Spatio-Temporal Applications, Communications in Statistics - Theory and Methods, DOI: 10.1080/03610926.2014.942433

**To link to this article:** http://dx.doi.org/10.1080/03610926.2014.942433

Revision:

# The Discretely Observed Immigration-Death Process: Likelihood Inference and Spatio-Temporal Applications

Ottmar Cronie*[1] and Jun Yu†[2]

[1]Stochastics, CWI, The Netherlands

[2]Department of Mathematics and Mathematical Statistics, Umeå University, Sweden

**Corresponding author:**

Ottmar Cronie

**Address corresponding author:**

P.O. Box 94079

1090 GB Amsterdam

The Netherlands

ottmar@alumni.chalmers.se

Tel.: +31 (0)20 592 4276

**Short title:**

The Discretely Observed Immigration-Death Process

*ottmar@alumni.chalmers.se

†jun.yu@math.umu.se





ACCEPTED MANUSCRIPT

**Abstract**

We define the homogeneous spatial immigration-death (HSID) process, a spatial birth-death process with as building blocks i) an immigration-death (ID) process (a continuous time Markov chain), and ii) a probability distribution assigning iid spatial locations to all points. For the ID-process we derive the likelihood function, reduce the likelihood estimation problem to one dimension, and prove consistency and asymptotic normality for the maximum likelihood estimators (MLEs) under a discrete sampling scheme. We additionally prove consistency for the MLEs of HSID-processes. In connection to the growth-interaction process, which has a HSID-process as basis, we also fit HSID-processes to Scots pine data.

**Key words**: Asymptotic normality, Consistency, Homogenous spatial immigration-death process, Maximum Likelihood, Spatial birth-death process, Spatio-temporal growth-interaction process



## 1 Introduction

A classical example of the application of spatial point processes can be found in the area of forestry and there particular focus has been the modelling of forest stands. This is clearly indicated by the extensive literature related to the subject (see e.g. (Illian, Penttinen, Stoyan, & Stoyan, 2008; Stoyan & Penttinen, 2000)). Initially these studies concerned themselves with modelling solely the spatial locations $x_1, \ldots, x_N$ of all the trees found within a spatial study region $W \subseteq \mathbb{R}^2$, at some particular point in time.

Since e.g. a forest stand is a temporally evolving entity, it may be reasonable to additionally consider some time horizon $[0, T)$, $T > 0$, over which we perform our study. It is thus natural to connect a temporal event time $t_i \in [0, T)$, such as a birth time, to each spatial location $x_i \in W$, $i = 1, \ldots, N \equiv N(T, W)$. Such a collection may be referred to as a *spatio-temporal point process* and for natural reasons there has been a growing interest in specific such models.

Often when dealing with spatio-temporal observations of e.g. a forest stand, we do not actually possess data sets which have been observed continuously over $[0, T)$, but rather only a fixed set of snapshots $\mathbf{x}_1, \ldots, \mathbf{x}_n \subseteq \{x_i\}_{i=1}^N$ of the locations, obtained at a fixed set of sample times $T_1, \ldots, T_n \in [0, T)$ (assume that $\mathbf{x}_0 = \emptyset$ is the sample at $T_0 = 0$). As a consequence we do not actually have access to the exact temporal event times $t_i$, but rather only information on which sample interval $(T_{k-1}, T_k]$ some observed $t_i$ belongs to and, possibly, in which interval the point is removed. Furthermore,



ACCEPTED MANUSCRIPT

there likely exist unobserved points (trees) which arrive and disappear within the same unobserved part $(T_{k-1}, T_k) \subseteq [0, T)$, $k = 1, \ldots, n$. Hereby, in order to correctly model such a data set with some given spatio-temporal model, we need to develop a modelling framework specifically for such spatio-temporal samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$.

There are different model structures that can be proposed here and one of them is given by *spatial jump processes* (Berthelsen & Møller, 2002). These can be described as Markov jump processes which have quite general structures imposed as state spaces. One possible such structure is the space of finite point patterns (van Lieshout, 2000; Møller & Waagepetersen, 2004) and here one particular instance is given by the class of *spatial birth-death processes*. These have mainly been used for the purposes of simulating static spatial point patterns $\{x_i\}_{i=1}^N$. However, these processes can also be purposefully employed for direct spatio-temporal modelling attempts.



One particular such model, which is the focus of this study, is what we here will refer to as the *homogeneous spatial immigration-death (HSID) process* $Y(s)$, $s \geq 0$. Here, when points arrive, they are assigned iid spatial locations in $W$ according to some common predefined spatial distribution. Its underlying temporal part, which governs the arrival times of new points $x_i$ to $W$ and the amount of time each point spends in $W$, is given by an *immigration-death (ID) process* $\{N(s)\}_{s\geq 0}$, often also referred to as an $M/M/\infty$-queue (Asmussen, 2003; Gani & Swift, 2011; Gillespie & Renshaw, 2005; Grimmett & Stirzaker, 2001). It is a continuous time Markov chain with state space $\{0, 1, \ldots\}$ and two parameters; an arrival rate $\alpha > 0$ and a death rate $\mu > 0$.

The main objective here is to correctly derive exact ML-estimators for the parameters $\alpha, \mu$ when $\{N(s)\}_{s\geq 0}$ is sampled at discrete times $T_1, \ldots, T_n$. Note that this is different than the scenario treated in (Gibson & Renshaw, 2001), where it is assumed that the actual death times are observed, or the scenario in (Bhat & Adke, 1981) where a continuous sampling scheme is applied to the ID-process. In general, in the case of continuous time Markov chains, the maximum likelihood (ML) theory based on continuous observations of sample paths has been covered quite extensively in the literature (see e.g. (Basawa & Prakasa Rao, 1980; Billingsley, 1961; Keiding, 1975); see (Guttorp, 1991) for inference related to branching processes). However, in the case of ML estimation based on processes sampled according to a discrete sample scheme much less has been done. Regarding the asymptotic properties of such ML-estimators, in recent years some general results have emerged (e.g in the context of discretely sampled Markov jump processes – see e.g. (Dehay & Yao, 2007)) and we may exploit these to establish properties such as strong consistency and asymptotic normality of the ML-estimators, in the context of ID-processes. Having obtained such ML-estimators and related asymptotic results, our further objective here is to extend this discrete sample approach to the full HSID-process. Due to the imposed lack of dependence between the locations, e.g. consistency is obtained as a straight forward corollary of the results for $\{N(s)\}_{s\geq 0}$.

The papers is structured as follows. We start by finding the transition probabilities of the ID-process (Section 2) and the finite dimensional distributions of the HSID-process (Section 3), which in turn are used to define the corresponding likelihood functions. In Section 4 we give results on strong consistency and asymptotic normality of the ML-estimators when the ID-process is sampled at discrete times. Here the consistency is also extended to the HSID-process case and the ID-process ML-estimators are evaluated numerically. In Section 5 we fit a HSID-process to a collection of forest plots. Section 6 contains a brief summary and discussion.



## 2 The immigration-death process

The *immigration-death process* (ID-process), $\{N(s)\}_{s\geq 0}$, is a time-homogeneous irreducible continuous time Markov chain where the possible states for which transitions $i \to j$ are possible are supplied by the state space $E = \mathbb{N} = \{0, 1, \dots\}$. The process is governed by the parameter pair $\theta = (\alpha, \mu)$, which we henceforth assume to take values in some parameter space $\Theta \subseteq \mathbb{R}_+^2 = (0, \infty) \times (0, \infty)$.

One way of viewing the ID-process is to treat it as a special case of a birth-death process. Here, for any $h \geq 0$, the infinitesimal transition probabilities of the birth-death process are given by

$$p_{ij}(t;\theta) := \mathbb{P}\left(N(h+t) = j | N(h) = i\right) = \begin{cases} \lambda_i t + o(t) & \text{if } j = i+1 \\ 1 - (\lambda_i + \mu_i)t + o(t) & \text{if } j = i \\ \mu_i t + o(t) & \text{if } j = i-1 \\ o(t) & \text{if } |j-i| > 1, \end{cases} \tag{2.1}$$

where the birth rates are given by $\lambda_i = \alpha$, $i = 0, 1, \dots$, and the death rates are given by $\mu_i = i\mu$, $i = 0, 1, \dots$, (see (Grimmett & Stirzaker, 2001)). Within this framework the interpretation of $\{N(s)\}_{s\geq 0}$ is the following: By letting the arrivals of new individuals to a population occur according to a Poisson process with intensity $\alpha$ and upon arrival assigning to all individuals independent and



ACCEPTED MANUSCRIPT

exponentially distributed lifetimes with mean $1/\mu$, $N(s)$ gives us the number of individuals alive at time $s$. Another possibility is to consider the equivalently defined $M/M/\infty$ queuing system; each customer (arriving according to a Poisson process with intensity $\alpha$) is being handled by its own server so that its sojourn time in the system is exponential with intensity $\mu$ and independent of all other customers. We note also that $N(s)$ in fact is a branching process, and as such it is possible to derive many of the (statistical) results of this paper (see e.g. (Bhat & Adke, 1981)). However, for reasons related to the temporal Markovianity of HSID-processes, we choose to treat $N(s)$ as a continuous time Markov process.

We here are interested in the finite dimensional distributions of $\{N(s)\}_{s\geq 0}$, mainly since these allow us to construct a likelihood structure for our model.

**Theorem 1.** *The transition probabilities of the ID-process are given by convolutions of Poisson densities and Binomial densities such that*

$$p_{ij}(t;\theta) = \left(f_{Poi(\rho)} * f_{Bin(i,\mathrm{e}^{-\mu t})}\right)(j) = \sum_{k=0}^{j} f_{Poi(\rho)}(k) f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k) \tag{2.2}$$

$$= \sum_{k=0}^{i\wedge j} f_{Poi(\rho)}(j-k) f_{Bin(i,\mathrm{e}^{-\mu t})}(k) = \frac{\mathrm{e}^{-\frac{\alpha}{\mu}\left(1-\mathrm{e}^{-\mu t}\right)}}{j!} \sum_{k=0}^{j} \left(\frac{\alpha}{\mu}\right)^k \binom{j}{k} \frac{\mathrm{e}^{-(j-k)\mu t}}{(1-\mathrm{e}^{-\mu t})^{j-2k-i}} \frac{i!}{(i-(j-k))!},$$

*where* $i, j \in E = \mathbb{N}$, $\theta = (\alpha, \mu) \in \Theta \subseteq \mathbb{R}_+^2$, $f_{Poi(\rho)}(\cdot)$ *is the Poisson density with parameter* $\rho = \frac{\alpha}{\mu}\left(1-\mathrm{e}^{-\mu t}\right)$, *and* $f_{Bin(i,\mathrm{e}^{-\mu t})}(\cdot)$ *is the Binomial density with parameters* $i$ *and* $\mathrm{e}^{-\mu t}$. *Note that* $p_{ij}(t;\theta) = 0$ *if* $i < 0$ *or* $j < 0$. *Moreover, we have that the probability generating function and the first two moments of* $(N(h+t)|N(h) = i)$ *are given by*

$$G_i(z;\theta) = \left(1 + (z-1)\,\mathrm{e}^{-\mu t}\right)^i \mathrm{e}^{\rho(z-1)} \tag{2.3}$$

$$\mathbb{E}[N(h+t)|N(h) = i] = i\,\mathrm{e}^{-\mu t} + \rho$$

$$\mathbb{E}[N^2(h+t)|N(h) = i] = i(i-1)\,\mathrm{e}^{-2\mu t} + (1+2\rho)i\,\mathrm{e}^{-\mu t} + \rho^2 + \rho.$$

*Proof.* Notice first that for any fixed $t > 0$, $N(t)$ is the result of applying so called *p-thinning* (see e.g. (Stoyan, Kendall, & Mecke, 1995)) to a Poisson process with intensity $\alpha$, using thinning probability $1 - p(t)$, given $N(0) = 0$. Since an individual's arrival time, conditioned on the individual's arrival during $(0,t]$, is uniformly distributed on $(0,t]$ and its life-time is $Exp(\mu)$-distributed we get that

$$\begin{aligned} p(t) &= \mathbb{P}\left(\text{An individual arrives during } (0,t] \text{ and survives time } t\right) \\ &= \int_0^t \left(1 - F_{Exp(\mu)}(t-x)\right) f_{Uni(0,t)}(x)dx = \frac{1}{t}\int_0^t \mathrm{e}^{-\mu(t-x)}\, dx = \frac{1-\mathrm{e}^{-\mu t}}{\mu t}. \end{aligned}$$

By the properties of thinned Poisson processes (see e.g. (Stoyan et al., 1995)) we have that $N(t) \sim Poi(\alpha t p(t)) = Poi(\rho)$, $\rho = \frac{\alpha}{\mu}\left(1 - \mathrm{e}^{-\mu t}\right)$.

With the marginal distributions of $\{N(t)\}_{t\geq 0}$ at hand (given $N(0) = 0$) we now proceed to finding $p_{ij}(t;\theta)$. Given that there are $i$ individuals present at a given time $h > 0$, we denote by $X$ the number of these individuals who have survived $(h, h+t]$. Clearly $X$ is $Bin(i, \mathrm{e}^{-\mu t})$-distributed and by denoting by $Y$ the number of new individuals arriving in $(h, h+t]$, which by the previous argument is $Poi(\rho)$-distributed and is independent of $X$, we obtain $p_{ij}(t;\theta)$ as the convolution of the distributions of $X$ and $Y$, i.e.

$$\begin{aligned} p_{ij}(t;\theta) &= \mathbb{P}(X+Y=j) = \sum_{k=0}^{\infty}\mathbb{P}(Y=k)\mathbb{P}(X=j-k) \\ &= \sum_{k=0}^{\infty} f_{Poi(\rho)}(k) f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k) = \sum_{k=0}^{j} f_{Poi(\rho)}(k) f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k). \end{aligned}$$

Since the independent random variables $X$ and $Y$ have probability generating functions $G_X(z;t) = \left(1 + (z-1)\,\mathrm{e}^{-\mu t}\right)^i$ and $G_Y(z;t) = \mathrm{e}^{\rho(z-1)}$, respectively, we find that the probability generating function is given by $G_i(z;\theta) = \mathbb{E}[\mathrm{e}^{zN(h+t)}|N(h)=i] = \mathbb{E}[\mathrm{e}^{z(X+Y)}] = G_X(z;t)G_Y(z;t) = (1 + (z-1)\,\mathrm{e}^{-\mu t})^i\, \mathrm{e}^{\rho(z-1)}$.

In the same spirit, we finally find that the first two moments of $(N(h+t)|N(h)=i)$ are given by $\mathbb{E}[N(h+t)|N(h)=i] = \mathbb{E}[X+Y] = i\,\mathrm{e}^{-\mu t} + \rho$ and $\mathbb{E}[N^2(h+t)|N(h)=i] = \mathbb{E}[X^2] + 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y^2] = i(i-1)\,\mathrm{e}^{-2\mu t} + (1+2\rho)i\,\mathrm{e}^{-\mu t} + \rho^2 + \rho$.

□

It should be pointed out that these results also may be obtained in another (less intuitive) way. From e.g. the infinitesimal probabilities (2.1) we may derive the probability generating function $G_i(z;\theta)$ (see e.g. (Grimmett & Stirzaker, 2001)) and by induction we can derive its $j$th derivative

(Cronie & Yu, 2010)

$$G_i^{(j)}(z) := \frac{\partial^j G_i(z;\theta)}{\partial z^j} = G_i(z) \sum_{k=0}^{j} \rho^k \binom{j}{k} \frac{1}{(\mathrm{e}^{\mu t} - 1 + z)^{j-k}} \frac{i!}{(i-(j-k))!}, \tag{2.4}$$

which in turn allows us to retrieve the transition probabilities (2.2) as $p_{ij}(t;\theta) = G_i^{(j)}(0)/j!$. The first two moments can be obtained through $G_i^{(1)}(1^-)$ and $G_i^{(2)}(1^-)$.

Note that in practice it is often natural to condition on $N(0) = 0$ and under this condition, either from e.g. (Gani & Swift, 2011) or through the proof of Theorem 1, we see that the marginal distribution of $N(s)$ is given by the Poisson distribution with parameter $\rho = \frac{\alpha}{\mu}(1 - \mathrm{e}^{-\mu s})$. Furthermore, we then have that $\mathbb{P}(N(s) \in \cdot) \to \pi_\theta(\cdot) = \mathbb{P}(Poi(\alpha/\mu) \in \cdot)$, as $s \to \infty$, since $\lim_{s\to\infty} \frac{\alpha}{\mu}(1 - \mathrm{e}^{-\mu s}) = \alpha/\mu$. Extending this, the following theorem, given in (Asmussen, 2003), establishes the ergodicity of $\{N(s)\}_{s\geq 0}$ (which together with the irreducibility gives us its positive recurrence) and its invariant distribution.

**Theorem 2.** *The ID-process is ergodic with invariant distribution given by the Poisson distribution with mean $\alpha/\mu$.*

Note that this invariant distribution is unique due to the positive recurrence, and it is also the same as its asymptotic distribution since every asymptotic distribution is an invariant distribution.

A further important characterisation of $\{N(s)\}_{s\geq 0}$ is its representation as a Markov jump process.

**Theorem 3.** *Let $\theta = (\alpha, \mu) \in \Theta \subseteq \mathbb{R}_+^2$. $\{N(s)\}_{s\geq 0}$ is a Markov jump process with state space $E = \mathbb{N}$, jump intensity function $\lambda(\theta; i) = \alpha + \mu i$, $i \in E$, and transition kernel $r(\theta;\cdot) = \{r(\theta; i, j) : i, j \in E\}$, where $r(\theta; i, j) = \frac{1}{\alpha+\mu i}(\alpha \mathbf{1}\{j = i+1\} + \mu i \mathbf{1}\{j = i-1\})$, $i, j \in E$.*

*Proof of Theorem 3.* Let $\{N(s)\}_{s\geq 0}$ be adapted to some filtered probability space $\left(\Omega, \mathcal{F}, \{\mathcal{F}_s\}_{s\geq 0}, \mathbb{P}\right)$. A continuous-time Markov chain is by definition a Markov jump process (Kallenberg, 2002, p. 243). Let $0 = \tau_0 < \tau_1 < \tau_2 < \dots$ ($\lim_{n\to\infty} \tau_n = \infty$) be the jump-times of $N(s) = N(0) + \sum_{k=1}^{\infty} Y_k \mathbf{1}\{\tau_k \leq s\}$, having associated jump-sizes $Y_k = N(\tau_k) - N(\tau_{k-1}) \in \{-1, 1\}$, $k = 1, 2, \dots$ (we consider a right continuous version of $\{N(s)\}_{s\geq 0}$). This is the embedded jump chain of $\{N(s)\}_{s\geq 0}$. Since $\{N(s)\}_{s\geq 0}$ is a Markov jump process, each increment $\tau_k - \tau_{k-1}$ will be independent of $\mathcal{F}_{\tau_{k-1}}$ and, given that





ACCEPTED MANUSCRIPT

$N(\tau_{k-1}) = i$, it holds that $\tau_k - \tau_{k-1}$ is $Exp\,(\lambda(\theta; i))$-distributed. Noticing that the lifetimes $L_1, L_2, \dots$ of all individuals generated by $N(s)$ are iid $Exp(\mu)$-distributed and also that any inter-jump-time $\tau_\alpha$ of the (Poisson) arrival process is $Exp(\alpha)$-distributed we find that $\tau_k - \tau_{k-1} \overset{d}{=} \min\{\tau_\alpha, L_1, \dots, L_i\}$ for $i \in \mathbb{Z}_+$, and clearly $\tau_k - \tau_{k-1} \overset{d}{=} \tau_\alpha$ if $i = 0$. Since the minimum of $n$ independent exponential random variables with parameters $\lambda_1, \dots, \lambda_n$ is exponentially distributed with parameter $\sum_{i=1}^n \lambda_i$, this implies that the jump intensity function is given by $\lambda(\theta; i) = \left(\mathbb{E}_\theta[\tau_k - \tau_{k-1} | N(\tau_{k-1}) = i]\right)^{-1} = \alpha + \mu i$, $i \in E$, where $\mathbb{E}_\theta[\cdot]$ denotes expectation under the parameter pair $\theta = (\alpha, \mu)$. Applying again the arguments above we find that

$$
\begin{aligned}
r\,(\theta; i, i+1) &= \mathbb{P}\left(N(\tau_k) = i+1 | N(\tau_{k-1}) = i\right) = \mathbb{P}\left(\tau_\alpha < \min\left(L_1, \dots, L_i\right) | N(\tau_{k-1}) = i\right) \\
&= \int_0^\infty \left(1 - \mathrm{e}^{-\alpha y}\right) f_{\min(L_1, \dots, L_i) | N(\tau_{k-1})}\,(y|i)\, dy \\
&= 1 - \mathbb{E}\left[\mathrm{e}^{-\alpha \min(L_1, \dots, L_i)} \,\middle|\, N(\tau_{k-1}) = i\right] = 1 - \left(1 + \frac{\alpha}{\mu i}\right)^{-1} = \frac{\alpha}{\alpha + \mu i},
\end{aligned}
$$

since a random variable $X \sim Exp(\gamma)$ has moment generating function $m_X(z) = \mathbb{E}[\mathrm{e}^{zX}] = (1 - z/\gamma)^{-1}$. Therefore, since $Y_k = N(\tau_k) - N(\tau_{k-1}) \in \{-1, 1\}$, $k = 1, 2, \dots$, the transition kernel of the Markov jump process, $r(\theta; \cdot) = \{r(\theta; i, j) : i, j \in E\}$, is determined by

$$
\begin{aligned}
&r(\theta; i, j) = \mathbb{P}\left(N(\tau_k) = j | N(\tau_{k-1}) = i\right) = \mathbf{1}\{j = i+1\} \mathbb{P}\left(N(\tau_k) = i+1 | N(\tau_{k-1}) = i\right) \\
&+ \mathbf{1}\{j = i-1, i > 0\}\left(1 - \mathbb{P}\left(N(\tau_k) = i+1 | N(\tau_{k-1}) = i\right)\right) = \frac{\alpha \mathbf{1}\{j = i+1\} + \mu i \mathbf{1}\{j = i-1\}}{\alpha + \mu i}.
\end{aligned}
$$

□

## 3 The homogeneous spatial immigration-death process

Turning to the spatio-temporal framework, we next define what here will be referred to as a *homogeneous spatial immigration-death process* (HSID-process). We define it as a spatial jump process (Berthelsen & Møller, 2002; Møller & Waagepetersen, 2004) $Y(s) \in N_f$, $s \geq 0$, where $N_f = \{x \subseteq W : |x| < \infty\}$ is the collection of all finite point configurations $x = \{x_1, \dots, x_n\}$ contained in some set $W \subseteq \mathbb{R}^d$ with Borel sets $\mathcal{B}$. Such a process is specified by two components: a Markov jump process $\{N(s)\}_{s \geq 0}$ governing the times $\tau_1 < \tau_2 < \dots$ at which transitions between states take place and a transition kernel $k_x(\cdot)$, $x \in N_f$, which, given $Y(\tau_i) \in N_f$ and the subsequent jump time $\tau_{i+1}$ and state $N(\tau_{i+1})$, determines to which state $Y$ will jump at $\tau_{i+1}$.

**Definition 1.** *Consider some suitable spatial study region $W \subseteq \mathbb{R}^d$ with Borel sets $\mathcal{B}$, some suitable probability measure $P(\cdot) = P(\cdot;\theta_W)$, $\theta_W \in \Theta_W \subseteq \mathbb{R}^m$, $m \geq 0$, on $(W, \mathcal{B})$ and an ID-process $\{N(s)\}_{s\geq 0}$. A homogeneous spatial immigration-death process $\{Y(s)\}_{s\geq 0}$ is a spatial jump process which describes a population where:*

1. *additions and removals of individuals in the population are determined by $N(s)$,*
2. *each individual is assigned a spatial location according to the distribution $P(\cdot)$.*

Here typical examples of spatial study regions include Euclidean hyper-rectangles or balls $W \subseteq \mathbb{R}^d$, $d = 2, 3$. Note that the lifetimes of the individuals are determined by the Poisson process arrival times and the iid $Exp(\mu)$-distributed lifetimes. Moreover, conditionally on the cardinality $|Y(s)| = n$, $Y(s)$ simply gives us a collection of iid (hence the name "homogeneous") $P(\cdot)$-distributed points $X_1, \ldots, X_n \in W$. It can be seen that HSID-processes are special cases of spatial birth-death processes (see e.g. (Berthelsen & Møller, 2002; Møller & Waagepetersen, 2004)). It should be pointed out that, although the spatial locations are independent, the HSID-process may serve as a decent model also when there is spatial dependence present, if we can assume that this dependence is weak or, possibly, if the points are far apart spatially.

Let $Y_1, \ldots, Y_n$ be an arbitrary collection of (non-random) elements of $N_f$, i.e. an arbitrary collection of point configurations. If $P(\cdot)$ has density $f_W(\cdot)$ w.r.t. Lebesgue measure $\nu(\cdot)$ in $\mathbb{R}^d$, we obtain the following result.

**Theorem 4.** *Given $0 = T_0 < T_1 < \ldots < T_n$ and some point configuration $Y(0) = Y_0$ in $W$, the joint density of $(Y(T_1), \ldots, Y(T_n))$, evaluated at the vector $(Y_1, \ldots, Y_n)$ of elements in $N_f$, is given by*

$$f_Y((Y_k)_{k=1}^n) = \frac{1}{|\mathcal{P}|} \prod_{\xi \in \bigcup_{k=1}^n Y_k} f_W(\xi) \prod_{k=1}^n p_{N_{k-1}N_k}(\Delta T_{k-1}; \alpha, \mu),$$

*where $|\mathcal{P}| \in \mathbb{Z}_+$, which is given in expression (3.1), depends on $N_1 = |Y_1|, \ldots, N_n = |Y_n|$ and $m = |\bigcup_{k=1}^n Y_k| \in \mathbb{N}$, and $2/mn(n+1) \leq 1/|\mathcal{P}| \leq 1$.*

*Proof.* Let $M = |\bigcup_{k=1}^{n} Y(T_k)| \in \mathbb{N}$ denote the total number of distinct points observed at $T_1, \ldots, T_n$ and label the locations of the $M$ points as $X_1, \ldots, X_M \in W$ (in some suitable fashion). Define additionally the index sets $\Omega(T_k) = \{i \in \mathbb{Z}_+ : X_i \in Y(T_k)\}$, $k = 1, \ldots, n$.

Now, based on the observed sets $Y_1, \ldots, Y_n$, let $m$, $x_1, \ldots, x_m \in W$, and $\omega_k = \{i \in \mathbb{Z}_+ : x_i \in Y_k\}$, $k = 1, \ldots, n$, be the observed counterparts of $M$, $X_1, \ldots, X_M$ and $\Omega(T_k)$, $k = 1, \ldots, n$, respectively.

Since the locations $X_1, \ldots, X_M$ are iid with density $f_W(\cdot)$, clearly we may now write the joint density of $(Y(T_1), \ldots, Y(T_n))$ as

$$f_Y((Y_k)_{k=1}^n) = f_\Omega(\omega_1, \ldots, \omega_n) f_X(x_1, \ldots, x_m) = f_\Omega(\omega_1, \ldots, \omega_n) \prod_{\xi \in \bigcup_{k=1}^n Y_k} f_W(\xi)$$

where $f_\Omega(\omega_1, \ldots, \omega_n)$ is the joint density of $(\Omega(T_1), \ldots, \Omega(T_n))$, evaluated at $(\omega_1, \ldots, \omega_n)$, and $f_X(x_1, \ldots, x_m) = f_Y((Y_k)_{k=1}^n | \Omega(T_1), \ldots, \Omega(T_n))$ is the joint density of $(X_1, \ldots, X_m)$, evaluated at $(x_1, \ldots, x_m)$. Note that the density $f_\Omega(\omega_1, \ldots, \omega_n)$ controls two main pieces of information: (i) the cardinality $N(T_k) = |Y(T_k)| = |\Omega(T_k)|$ (whence it controls $M$) and the labelling/order of the $N(T_k)$ points in each $Y(T_k)$, $k = 1, \ldots, n$ (which makes them distinct). The (conditional) density $f_X(x_1, \ldots, x_m)$, on the other hand, controls only the spatial locations of the $M$ points in $W$.

By denoting the joint density of $(N(T_1), \ldots, N(T_n)) = (|\Omega(T_1)|, \ldots, |\Omega(T_n)|)$, evaluated at $(N_1, \ldots, N_n) = (|\omega_1|, \ldots, |\omega_n|)$, by $f_N((|\omega_k|)_{k=1}^n)$ it is clear that

$$f_\Omega(\omega_1, \ldots, \omega_n) = f_O((\omega_k)_{k=1}^n) f_N((|\omega_k|)_{k=1}^n) = f_O((\omega_k)_{k=1}^n) \prod_{k=1}^{n} p_{N_{k-1}N_k}(\Delta T_{k-1}; \alpha, \mu),$$

where $f_O((\omega_k)_{k=1}^n)$ is a density which determines the order in which the $M$ points arrive to $W$ and $p_{N_{k-1}N_k}(\Delta T_{k-1}; \alpha, \mu)$, $k = 1,,$ is the transition probability of the ID-process.

Since we have already accounted for the number of points present at $T_1, \ldots, T_n$ and the locations $X_1, \ldots, X_M$, it now only remains to determine $f_O((\omega_k)_{k=1}^n)$. It follows that $f_O((\omega_k)_{k=1}^n) = 1/|\mathcal{P}|$, where

$$\mathcal{P} = \Big\{ (\omega_1, \ldots, \omega_n) \ : \ \omega_1, \ldots, \omega_n \subseteq \{1, \ldots, m\};\ |\omega_1| = n_1, \ldots, |\omega_n| = n_n; \quad i \in \omega_{k-1}, i \notin \omega_k \Rightarrow i \notin \omega_{k+1} \text{ for any } k = 2, \ldots, n-1 \Big\}. \tag{3.1}$$



ACCEPTED MANUSCRIPT

This can be seen by considering the matrix $A = [a_{i,k}]_{i=1,\ldots,m;\ k=1,\ldots,n}$, which has entries $a_{i,k} = \mathbf{1}_{\Omega(T_k)}(i)$. When we condition on $(N(T_k))_{k=1}^n = (N_k)_{k=1}^n$ we only specify that the column-sums of $A$ are given by $N_1, \ldots, N_n$, whence we still have to determine what the probability is of $A$ being observed as the matrix $\bar{A}$ with entries $\bar{a}_{i,k} = \mathbf{1}_{\omega_k}(i)$. It may be seen that the sample space of the conditional random matrix $A|\{(N(T_k))_{k=1}^n = (N_k)_{k=1}^n\}$ is given by the $m \times n$-matrices which have 0-1 entries, column-sums $d_1, \ldots, d_n$ and all 1's in each row connected (this follows since an indexed point cannot return again once it has been removed); when $n = 5$, say, a row can be given by e.g. $(0, 1, 1, 0, 0)$ or $(1, 1, 1, 1, 0)$. To obtain a bound for $|\mathcal{P}|$ we find that the number of matrices which have rows with connected 1's is given by $m \sum_{j=1}^n (n - (j-1)) = mn(n+1)/2$, whereby $2/mn(n+1) \leq 1/|\mathcal{P}| \leq 1$.

□

A simple example here is to let $W$ be bounded and $f_W(\cdot) = 1/\nu(W)$, so that $P(B) = \int_B 1/\nu(W)\nu(dx) = \nu(B)/\nu(W)$, i.e. the spatial locations are $Uni(W)$-distributed. This is the form employed in the growth-interaction process (see Section 5).

## 4 Maximum likelihood estimation

Assume now that we sample $\{N(s)\}_{s\geq 0}$ as $N_1, \ldots, N_n$ at the respective times $T_1 < \ldots < T_n$ (we write $0 = T_0 < T_1$). Since the likelihood function for $\theta = (\alpha, \mu) \in \Theta$, $L_n(\theta)$, is given by the joint density of the distribution of $(N(T_1), \ldots, N(T_n))$, by the Markov property of $N(s)$ it can be factorised into a product of transition probabilities, $L_n(\theta) = \mathbb{P}(N(T_1) = N_1) \prod_{k=2}^n p_{N_{k-1}N_k}(\Delta T_{k-1}; \theta)$, where $\Delta T_{k-1} = T_k - T_{k-1}$. Since by assumption we condition on $N(T_0) = 0$, the log-likelihood will be given by

$$l_n(\theta) = l_n(\theta; N_1, \ldots, N_n) = \sum_{k=1}^n \log p_{N_{k-1}N_k}(\Delta T_{k-1}; \theta). \tag{4.1}$$

In the case of equidistant sampling, i.e. $\Delta T_{k-1} = t$ for each $k = 1, \ldots, n$, the log-likelihood takes the form

$$l_n(\theta) = \sum_{i,j \in E} N_n(i,j) \log p_{ij}(t; \theta), \tag{4.2}$$



ACCEPTED MANUSCRIPT

where $N_n(i,j) = \sum_{k=1}^{n} \mathbf{1}\{(N_{k-1}, N_k) = (i,j)\}$. Hereby, for each of the sampling schemes, the likelihood estimator of $\theta = (\alpha, \mu) \in \Theta$ (obtained by replacing $N_k$ by $N(T_k)$, $k = 0, 1, \ldots$, in expressions (4.1) and (4.2)) will be defined as

$$\hat{\theta}_n = (\hat{\alpha}_n, \hat{\mu}_n) = \arg\max_{\theta\in\Theta} l_n(\theta). \tag{4.3}$$

Since the system of equations

$$\begin{cases} 0 = \frac{\partial}{\partial\alpha} l_n(\theta) & = & \sum_{i,j\in E} N_n(i,j) \frac{\partial}{\partial\alpha} \log p_{ij}(t;\theta) = \sum_{i,j\in E} N_n(i,j) \frac{1}{\alpha} \left( \frac{p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)} - \rho \right) \\ 0 = \frac{\partial}{\partial\mu} l_n(\theta) & = & \sum_{i,j\in E} N_n(i,j) \frac{\partial}{\partial\mu} \log p_{ij}(t;\theta) \\ & = & \sum_{i,j\in E} N_n(i,j) \frac{\rho\tau}{(1-\mathrm{e}^{-\mu t})\mu} - \frac{(j - i\,\mathrm{e}^{-\mu t})\mu t}{(1-\mathrm{e}^{-\mu t})\mu} - \frac{\tau - \mu t}{(1-\mathrm{e}^{-\mu t})\mu} \frac{p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)}, \\ p_{ij}(t;\theta)_k & := & \sum_{k=0}^{j} k f_{Poi(\rho)}(k) f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k) = \rho p_{i(j-1)}(t;\theta), \end{cases} \tag{4.4}$$

where $\rho = \frac{\alpha}{\mu}\left(1 - \mathrm{e}^{-\mu t}\right)$ and $\tau = 1 - \mathrm{e}^{-\mu t} - \mu t\,\mathrm{e}^{-\mu t}$, has no known closed form solution, numerical methods have to be employed in order to obtain the ML-estimates. However, as unfortunate as this may be, as we can see from the result below, it is still possible to express the estimator of $\alpha$ as a function of the parameter $\mu$ (and the sample), hence reducing the maximisation to a one dimensional problem.

**Proposition 1.** *The ML-estimator is found by maximising $l_n(\hat{\alpha}_n(\mu), \mu)$ over $\Theta_\mu \subseteq \mathbb{R}_+$ (the projection of $\Theta$ onto the $\mu$-axis); $\hat{\mu}_n = \arg\max_{\mu\in\Theta_2} l_n(\hat{\alpha}(\mu), \mu)$, $\hat{\alpha}_n = \hat{\alpha}_n(\hat{\mu}_n)$, and*

$$\hat{\alpha}_n(\mu) := \frac{\frac{\mu}{(1-\mathrm{e}^{-\mu t})} \frac{1}{n} \sum_{i,j\in E} N_n(i,j)(j - i\,\mathrm{e}^{-\mu t})}{2\left(\frac{1-\mathrm{e}^{-\mu t}}{\mu t} - \mathrm{e}^{-\mu t}\right) - 1} = \frac{\mu \frac{1}{n} \left( \frac{\mathrm{e}^{-\mu t} N_n - N_0}{1-\mathrm{e}^{-\mu t}} + \sum_{k=0}^{n} N_k \right)}{2\left(\frac{1-\mathrm{e}^{-\mu t}}{\mu t} - \mathrm{e}^{-\mu t}\right) - 1}. \tag{4.5}$$

*Proof.* Since $\sum_{i,j\in E} N_n(i,j) = n$, through (4.4) we obtain $\sum_{i,j\in E} N_n(i,j) \frac{p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)} = \frac{\alpha}{\mu}\left(1 - \mathrm{e}^{-\mu t}\right) n$ and $\sum_{i,j\in E} N_n(i,j) \frac{p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)} = \frac{\rho\tau n - \mu t \sum_{i,j\in E} N_n(i,j)(j - i\,\mathrm{e}^{-\mu t})}{\mu t - \tau}$. By combining these two expressions we obtain the desired result.

□



ACCEPTED MANUSCRIPT

Turning to the HSID-process, through Theorem 4 we obtain the log-likelihood function

$$\begin{aligned} l_n^Y(\theta, \theta_W) &= l_n^Y(\theta, \theta_W; Y_1, \ldots, Y_n) = \log f_Y((Y_k)_{k=1}^n) = -\log(|\mathcal{P}|) + l_n^W(\theta_W) + l_n(\theta), \\ l_n^W(\theta_W) &= \sum_{\xi \in \bigcup_{k=1}^n Y_k} \log f_W(\xi; \theta_W), \end{aligned} \tag{4.6}$$

whereby the ML-estimator will be given by

$$\hat{\theta}_n^Y = (\hat{\alpha}_n, \hat{\mu}_n, \hat{\theta}_n^W) = \arg\max_{\theta \in \Theta} l_n^Y(\theta, \theta_W). \tag{4.7}$$

Note here that the full set of ML-estimates $(\hat{\alpha}_n, \hat{\mu}_n, \hat{\theta}_n^W)$ may be obtained by in (4.4) additionally differentiating $l_n^W(\theta_W)$ w.r.t. $\theta_W$ and setting the derivative equal to zero.

## 4.1 Asymptotic properties of the ID-process ML-estimators

We now wish to establish the consistency and the asymptotic normality of the sequence of estimators (4.3) under an equidistant sampling scheme.

More specifically we assume now that we sample $N(s)$ at the times $T_n = nt$, $n \in \mathbb{N}$, $t > 0$, and from the Markov property of $N(s)$ the *observation chain* $Z = (Z_n)_{n=1}^\infty \equiv (N(T_n))_{n=1}^\infty$ will also be a Markov chain, having transition kernel

$$q(\theta; \cdot) = \{q(\theta; i, j) : i, j \in E\} = \{\mathbb{P}(X(T_n) = j | X(T_{n-1}) = i; \theta) : i, j \in E\}, \quad E = \mathbb{N}.$$

Given $Z_0 = X(0)$, the log-likelihood generated by $(Z_1, \ldots, Z_n)$ is given by

$$l_n(\theta) = \sum_{k=1}^n \log q(\theta; Z_{k-1}, Z_k) = \sum_{i,j \in E} N_n(i,j) \log q(\theta; i, j),$$

where $N_n(i,j) = \sum_{k=1}^n \mathbf{1}\{(Z_{k-1}, Z_k) = (i,j)\}$. Throughout we denote by $\theta_0 = (\alpha_0, \mu_0) \in \Theta$ the actual underlying parameter pair.

In order to prove consistency (Theorem 5) and asymptotic normality (Theorem 6) of $\hat{\theta}_n$, we must ensure that identifiability holds.

**Lemma 1.** *For any $\theta \neq \theta_0$, $q(\theta; \cdot) \neq q(\theta_0; \cdot)$.*

*Proof.* Recall from expression (2.3) the p.g.f. $G_i(z;\theta)$ of $(N(h+t)|N(h)=i)$, $\theta \in \Theta$. It is sufficient to show that

$$1 \;=\; \frac{G_i(z;\theta_0)}{G_i(z;\theta)} = \left(\frac{1+(z-1)\,\mathrm{e}^{-\mu_0 t}}{1+(z-1)\,\mathrm{e}^{-\mu t}}\right)^i \exp\left\{(z-1)\left(\frac{\alpha_0}{\mu_0}(1-\mathrm{e}^{-\mu_0 t})-\frac{\alpha}{\mu}(1-\mathrm{e}^{-\mu t})\right)\right\} = a_1 a_2$$

is cannot hold for $\theta \neq \theta_0$ and $z \neq 1$. We first see that $a_1 = 1$ iff $\mu_0 = \mu$ or $i = 0$. Hence, when $\mu = \mu_0$, $a_2 = \mathrm{e}^{(\alpha_0-\alpha)(z-1)\left(1-\mathrm{e}^{-\mu t}\right)/\mu}$ equals 1 iff $\alpha_0 = \alpha$. When $\mu_0 \neq \mu$ and $\alpha_0 \neq \alpha$, $a_2 = \mathrm{e}^{\alpha t(z-1)(\frac{1-\mathrm{e}^{-\mu_0 t}}{\mu_0 t}-\frac{1-\mathrm{e}^{-\mu t}}{\mu t})}$, which holds iff $\mu_0 = \mu$, since $(1-\mathrm{e}^{-x})/x$ is strictly decreasing. Assuming that $\alpha \neq \alpha_0$ and $\mu \neq \mu_0$, we obtain $a_2 = \mathrm{e}^{(z-1)(\frac{\alpha_0}{\mu_0}(1-\mathrm{e}^{-\mu_0 t})-\frac{\alpha}{\mu}(1-\mathrm{e}^{-\mu t}))}$. If $\frac{\alpha_0}{\mu_0} = \frac{\alpha}{\mu}$, by the monotonicity of $1-\mathrm{e}^{-x}$ we find that $a_2 = 0$ iff $\mu = \mu_0$ and if $1-\mathrm{e}^{-\mu t} = \eta(1-\mathrm{e}^{-\mu_0 t})$, where $\eta = \frac{\alpha_0\mu}{\alpha\mu_0} > 0$, we must require that $\mu = \mu_0$.

□

We now arrive at the consistency result.

**Theorem 5.** *Let $\Theta$ be any compact subset of $\mathbb{R}_+^2$. The maximum likelihood estimator satisfies $(\hat{\alpha}_n, \hat{\mu}_n) \xrightarrow{a.s.} (\alpha_0, \mu_0)$, as $n \to \infty$.*

*Proof.* Treated as a Markov jump process, the consistency may be obtained by showing that the generic conditions for Markov jump processes in (Dehay & Yao, 2007) are satisfied.

Through Theorem 3 we find that the jump intensity is positive and the transition kernel is irreducible. Furthermore, by Theorem 2, under $\theta_0$ the Markov chain $(Z_n)_{n\in\mathbb{N}}$ has a unique invariant probability measure $\pi_{\theta_0} = Poi(\alpha_0/\mu_0)$, with moments of all orders $a \geq 1$, i.e. $\sum_{i\in E} |i|^a \pi_{\theta_0}(i) < \infty$. Furthermore, due to the positive recurrence of $\{N(s)\}_{s\geq 0}$ (provided by Theorem 2), by an ergodic theorem (e.g. (Norris, 1997, Theorem 1.10.2)), for any $\pi_{\theta_0}$-integrable function $\phi : E \to \mathbb{R}$, $\frac{1}{n}\sum_{k=1}^n \phi(Z_k) \xrightarrow{a.s.} \sum_{i\in E}\phi(i)\pi_{\theta_0}(i)$, as $n \to \infty$.

By choosing $C = \max_{i,j\in\{0,1\}} |\log q(\theta_0; i, j)| < \infty$, we can find $a \in \mathbb{N}$ such that $|\log q(\theta_0; i, j)| \leq C(1 + |i|^{a/2} + |j|^{a/2})$ since the free choice of $a \in \mathbb{N}$ allows us to create an arbitrary large bound $(1 + |i|^{a/2} + |j|^{a/2})$ for $|\log q(\theta_0; i, j)|$ when $i, j \in \{2, 3, \ldots\}$.

By the compactness of $\Theta = \Theta_\alpha \times \Theta_\mu \subseteq \mathbb{R}_+^2$ we have that $\alpha_{min} := \inf\Theta_\alpha > 0$, $\alpha_{max} := \sup\Theta_\alpha < \infty$, $\mu_{min} := \inf\Theta_\mu > 0$ and $\mu_{max} := \sup\Theta_\mu < \infty$. Recalling (4.4), it can easily be verified that $|\frac{\partial}{\partial\alpha}\log q(\theta; i, j)| < t + \frac{j}{\alpha_{min}} < \infty$ and $|\frac{\partial}{\partial\mu}\log q(\theta; i, j)| < \frac{\alpha_{max}t^2+(3j+i)t}{1-\mathrm{e}^{-\mu_{min}t}} < \infty$.



ACCEPTED MANUSCRIPT

Furthermore, by the mean value theorem and the Schwarz-inequality, for $\theta, \theta' \in \Theta$, it can be verified that $|\log q(\theta; i, j) - \log q(\theta'; i, j)| < (t + \frac{j}{\alpha_{min}} + \frac{\alpha_{max} t^2 + (3j+i)t}{1 - e^{-\mu_{min} t}})|\theta - \theta'|(1 + |i|^{a/2} + |j|^{a/2})$. By setting $\gamma(x) = (t + \frac{1}{\alpha_{min}} + \frac{\alpha_{max} t^2 + 4t}{1 - e^{-\mu_{min} t}})x$, $x \geq 0$, we find that $\gamma(\cdot)$ is a continuity modulus such that, for all $i, j \in E$ and $\theta, \theta' \in \Theta$, $|\log q(\theta; i, j) - \log q(\theta'; i, j)| \leq \gamma(|\theta - \theta'|)(1 + |i|^{a/2} + |j|^{a/2})$, since we may choose $a \in \mathbb{N}$ freely. Given the above, there exists a strong law of large numbers for $U_n(\theta) = l_n(\theta_0) - l_n(\theta)$.

By recalling Lemma 1, we now conclude, by an appeal to (Dehay & Yao, 2007, Theorem 2), that $(\hat{\alpha}_n, \hat{\mu}_n) \xrightarrow{a.s.} (\alpha_0, \mu_0)$, as $n \to \infty$.

□

Having proved the consistency, we next turn to the asymptotic normality and in order to prove it we need the following technical lemma.

**Lemma 2.** *The transition probabilities satisfy the recursive relation*

$$\frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)} = \frac{(j+1)(e^{\mu t} - 1)}{\rho} \frac{p_{i(j+1)}(t;\theta)}{p_{ij}(t;\theta)} + \frac{j-i}{\rho} - e^{\mu t} + 1,$$

*where* $i, j \in E = \mathbb{N}$ *and* $\rho = \frac{\alpha}{\mu}\left(1 - e^{-\mu t}\right)$. *Note that* $p_{ij}(t;\theta) = 0$ *if* $i < 0$ *or* $j < 0$.

*Proof.* Recalling the probability generating function and (2.4), by utilising that $p_{ij}(t;\theta) = G_i^{(j)}(0)/j!$ and letting $a(z) = e^{\mu t} - 1 + z$, we obtain

$$G_i^{(j+1)}(z) = \left(\frac{i-j}{a(z)} + \rho\right) G_i^{(j)}(z) + \frac{j!}{a(z)} \frac{G_i(z)}{j!} \sum_{k=0}^{j} k\rho^k \binom{j}{k} \frac{1}{a(z)^{j-k}} \frac{i!}{(i-(j-k))!}$$

and through (4.4) we obtain

$$\begin{aligned} \frac{p_{i(j+1)}(t;\theta)}{p_{ij}(t;\theta)} &= \frac{j!}{(j+1)!} \frac{G_i^{(j+1)}(0)}{G_i^{(j)}(0)} = \frac{1}{j+1}\left(\frac{i-j}{e^{\mu t} - 1} + \rho + \frac{j!}{G_i^{(j)}(0)(e^{\mu t} - 1)} p_{ij}(t;\theta)_k\right) \\ &= \frac{\rho}{(j+1)(e^{\mu t} - 1)}\left(\frac{i-j}{\rho} + e^{\mu t} - 1 + \frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)}\right). \end{aligned}$$

□

**Theorem 6.** *Let* $\theta_0 = (\alpha_0, \mu_0)$ *be an interior point of the compact parameter space* $\Theta \subseteq \mathbb{R}_+^2$. *Furthermore, assume that* $(\log(\alpha_0 + \mu_0) - \log(\alpha_0))/\mu_0 \geq 2t$. *Then, as* $n \to \infty$, $\sqrt{n}\,((\hat{\alpha}_n, \hat{\mu}_n) - (\alpha_0, \mu_0))$



ACCEPTED MANUSCRIPT

*converges in distribution to the two-dimensional zero-mean Gaussian distribution with covariance matrix*

$$I(\theta_0)^{-1} = \frac{\mu_0}{t\left((1+\mathrm{e}^{-\mu_0 t})\,\rho_0(\Xi-1)-1\right)} \times \times \begin{pmatrix} \frac{\rho_0\left(2\tau_0-\mu_0 t\left(1-\mathrm{e}^{-\mu_0 t}\right)\right)+\frac{\rho_0^2}{\mu_0 t}(\Xi-1)(\tau_0-\mu_0 t)^2}{\left(1-\mathrm{e}^{-\mu_0 t}\right)^2} & 1+\frac{\rho_0}{\mu_0 t}(\Xi-1)(\tau_0-\mu_0 t) \\ 1+\frac{\rho_0}{\mu_0 t}(\Xi-1)(\tau_0-\mu_0 t) & \frac{1}{\mu_0 t}(\Xi-1)\left(1-\mathrm{e}^{-\mu_0 t}\right)^2 \end{pmatrix}, \tag{4.8}$$

*where* $\Xi = \sum_{i,j\in E} \frac{\left(p_{i(j-1)}(t;\theta_0)\right)^2}{p_{ij}(t;\theta_0)}\pi_{\theta_0}(i)$, $\pi_{\theta_0}(\cdot) = \mathbb{P}(Poi(\alpha_0/\mu_0) \in \cdot)$ *is the invariant distribution evaluated at* $\theta_0$*,* $\rho_0 = \frac{\alpha_0}{\mu_0}(1-\mathrm{e}^{-\mu_0 t})$ *and* $\tau_0 = 1-\mathrm{e}^{-\mu_0 t} - \mu_0 t\,\mathrm{e}^{-\mu_0 t}$*.*

*Proof.* Given the properties shown in Theorem 5, in order to prove the asymptotic normality we show that some additional conditions are satisfied and once again appeal to (Dehay & Yao, 2007). Here we denote the partial derivatives of a function $\psi(\cdot)$ of $\theta$ by $D_u\psi = \partial\psi/\partial\theta_u$ and $D^2_{uv}\psi = \partial^2\psi/\partial\theta_u\partial\theta_v$, $u,v=1,2$, and let $\Lambda_{\theta_0} \subseteq \Theta$ denote some neighbourhood of $\theta_0$.

We first show that $\log q(\theta;i,j)$ and its derivatives are sufficiently well-behaved, which is needed to ensure that there exists a strong law of large numbers for functions which involve first/second order derivatives of $g(\theta;i,j) := \log q(\theta_0;i,j) - \log q(\theta;i,j)$. Since the expression for $q(\theta;i,j)$ contains the term $\mathrm{e}^{-\frac{\alpha}{\mu}\left(1-\mathrm{e}^{-\mu t}\right)}$, the mapping $\theta \mapsto g(\theta;i,j)$ is (at least) twice continuously differentiable for all $\theta \in \Lambda_{\theta_0}$. Recalling from the proof of Theorem 5 the definitions of $\alpha_{min}$, $\alpha_{max}$, $\mu_{min}$ and $\mu_{max}$, it can be shown that (see (Cronie & Yu, 2010) for details) $\max_{(i,j)\in\{0,1\}^2} |D_1 \log q(\theta_0;i,j)| < \frac{1}{\alpha_{min}} + t =: C_1 < \infty$, $\max_{(i,j)\in\{0,1\}^2} |D_2 \log q(\theta_0;i,j)| < \frac{\alpha_{max}t^2+4t}{1-\mathrm{e}^{-\mu_{min}t}} =: C_2 < \infty$, $\max_{(i,j)\in\{0,1\}^2} \left|D^2_{11} \log q(\theta;i,j)\right| < \frac{1+2(1+\alpha_{max}t)^2}{\alpha^2_{min}} =: C_{11} < \infty$,

$$\max_{(i,j)\in\{0,1\}^2} \left|D^2_{12}\log q(\theta;i,j)\right| <$$
$$< \frac{2t}{\alpha_{min}} + \alpha_{max}t^3 + \frac{2t}{(1-\mathrm{e}^{-\mu_{min}t})\alpha_{min}} + 2t^2 + \frac{2}{\mu_{min}}t + \frac{(1+\alpha_{max}t)(4+\alpha_{max}t)}{(1-\mathrm{e}^{-\mu_{min}t})\alpha_{min}}t =: C_{12} < \infty,$$
$$\max_{(i,j)\in\{0,1\}^2} \left|D^2_{22}\log q(\theta;i,j)\right| <$$
$$< \left(\frac{\alpha_{max}t^2+4t}{1-\mathrm{e}^{-\mu_{min}t}}\right)^2 + t^2\left(6+10\alpha_{max}t+\alpha^2_{max}t^2+\mu^2_{max}t^2(6+3\alpha_{max}t)\right) =: C_{22} < \infty.$$

By choosing $C = \max\{C_1, C_2, C_{11}, C_{12}, C_{22}\}$, due to the free choice of moment order $a$ of the



ACCEPTED MANUSCRIPT

invariant distribution, we find that $\max\left\{|D_u \log q(\theta_0;i,j)|, |D^2_{uv} \log q(\theta_0;i,j)|\right\} < C(1+|i|^{a/2}+|j|^{a/2})$, for all $u,v=1,2$ and all $(i,j)\in E^2$. Furthermore, by the mean value theorem and the Schwarz-inequality we obtain that

$$\frac{\left|D^2_{uv}\log q(\theta;i,j) - D^2_{uv}\log q(\theta_0;i,j)\right|}{|\theta-\theta_0|} \leq \left|\nabla D^2_{uv}\log q\left((1-c)\theta + c\theta_0;i,j\right)\right|$$
$$\leq |D_1 D^2_{uv}\log q((1-c)\theta + c\theta_0;i,j)| + \ |D_2 D^2_{uv}\log q((1-c)\theta + c\theta_0;i,j)|$$

for some $0<c<1$, where $\theta$ and $\theta_0$ are in some open subset of $\mathbb{R}^2$ (in particular $\theta,\theta_0\in\Lambda_{\theta_0}$). Now, by consulting the bounds in expressions (A.1), (A.2), (A.3), (A.4) and choosing

$$\sigma_{11}(x) = \max_{(i,j)\in\{0,1\}^2}\left(\sup_{\mu\in\Theta_\mu}\sup_{\alpha\in\Theta_\alpha} B_{111}(\alpha,\mu,t,j,i) + \sup_{\mu\in\Theta_\mu}\sup_{\alpha\in\Theta_\alpha} B_{112}(\alpha,\mu,t,j,i)\right)x$$
$$\sigma_{12}(x) = \sigma_{21}(x) = \max_{(i,j)\in\{0,1\}^2}\left(\sup_{\mu\in\Theta_\mu}\sup_{\alpha\in\Theta_\alpha} B_{112}(\alpha,\mu,t,j,i) + \sup_{\mu\in\Theta_\mu}\sup_{\alpha\in\Theta_\alpha} B_{122}(\alpha,\mu,t,j,i)\right)x$$
$$\sigma_{22}(x) = \max_{(i,j)\in\{0,1\}^2}\left(\sup_{\mu\in\Theta_\mu}\sup_{\alpha\in\Theta_\alpha} B_{122}(\alpha,\mu,t,j,i) + \sup_{\mu\in\Theta_\mu}\sup_{\alpha\in\Theta_\alpha} B_{222}(\alpha,\mu,t,j,i)\right)x$$

we conclude that, for each pair $u,v=1,2$, there exists a continuity modulus $\sigma_{uv}(x)$ such that, for $\theta\in\Lambda_{\theta_0}$, $(i,j)\in E^2$, $|D^2_{uv}\log q(\theta_0;i,j) - D^2_{uv}\log q(\theta;i,j)| \leq \sigma_{uv}(|\theta_0-\theta|)(1+|i|^{a/2}+|j|^{a/2})$.

We next turn to regularity conditions related to the Fisher information. Note that since $(D_1\log q(\theta_0;i,j))q(\theta_0;i,j) = \frac{\rho_0(p_{i(j-1)}(t;\theta_0)-p_{ij}(t;\theta_0))}{\alpha_0}$, $(D_2\log q(\theta_0;i,j))q(\theta_0;i,j) = \frac{\rho_0\tau_0(p_{ij}(t;\theta_0)-p_{i(j-1)}(t;\theta_0))}{(1-\mathrm{e}^{-\mu_0 t})\mu_0} - \frac{(j-i\,\mathrm{e}^{-\mu_0 t})t}{1-\mathrm{e}^{-\mu_0 t}}p_{ij}(t;\theta_0) + \frac{\rho_0 t}{1-\mathrm{e}^{-\mu_0 t}}p_{i(j-1)}(t;\theta_0)$ and $\sum_{j=0}^{\infty}p_{ij}(t;\theta_0) = \sum_{j=0}^{\infty}p_{i(j-1)}(t;\theta_0) = \sum_{j=0}^{\infty}p_{i(j-2)}(t;\theta_0) = 1$, it follows that $\sum_{j\in E}\left(D_u\log q(\theta_0;i,j)\right)q(\theta_0;i,j) = 0$ for all $u,v=1,2$ and every $i\in E$. Additionally, through expressions (A.5), (A.6), (A.7) we obtain that $\sum_{j\in E}D^2_{uv}q(\theta_0;i,j) = 0$ for all $u,v=1,2$ and every $i\in E$, which is equivalent to having $-\sum_{j\in E}\left(D^2_{uv}\log q(\theta_0;i,j)\right)q(\theta_0;i,j) = \sum_{j\in E}\left(D_u\log q(\theta_0;i,j)\right)\left(D_v\log q(\theta_0;i,j)\right)q(\theta_0;i,j) = I_{uv}(\theta_0;i)$.

Turning to its structure, it can readily be shown (Cronie & Yu, 2010) that the Fisher information matrix $I(\theta_0;i) = \left(I_{uv}(\theta_0;i)\right)_{u,v=1,2}$ at $\theta_0$, associated with the family of distributions





ACCEPTED MANUSCRIPT

$\{q(\theta;i,\cdot):\theta\in\Lambda_{\theta_0}\}$, is given by $I(\theta_0;i)=A(\theta_0)+B(\theta_0)i+C(\theta_0)(\sum_{j=0}^{\infty}\frac{p_{i(j-1)}(t;\theta)^2}{p_{ij}(t;\theta)}-1)$, where

$$A(\theta_0)=\begin{pmatrix}0 & -\frac{t}{\mu_0}\\ -\frac{t}{\mu_0} & \frac{\alpha_0^2\mu_0 t(2\tau_0-\mu_0 t)}{\rho_0\mu_0^4}\end{pmatrix},\quad B(\theta_0)=\begin{pmatrix}0 & 0\\ 0 & \frac{\alpha_0 t^2\,\mathrm{e}^{-\mu_0 t}}{\mu_0\rho_0}\end{pmatrix},C(\theta_0)=\begin{pmatrix}\frac{\rho_0^2}{\alpha_0^2} & \frac{\rho_0(\mu_0 t-\tau_0)}{\mu_0^2}\\ \frac{\rho_0(\mu_0 t-\tau_0)}{\mu_0^2} & \frac{\alpha_0^2(\tau_0-\mu_0 t)^2}{\mu_0^4}\end{pmatrix}.$$

This implies that the (asymptotic) Fisher information $I(\theta_0)=\sum_{i\in E}I(\theta_0;i)\pi_{\theta_0}(i)$ of $(Z_n)_{n\in\mathbb{N}}$ is

$$\begin{aligned}I(\theta_0) &= A(\theta_0)+B(\theta_0)\sum_{i\in E}i\pi_{\theta_0}(i)+C(\theta_0)\left(\sum_{i,j\in E}\frac{\left(p_{i(j-1)}(t;\theta_0)\right)^2}{p_{ij_0}(t;\theta)}\pi_{\theta_0}(i)-1\right)\\ &= A(\theta_0)+\frac{\alpha_0}{\mu_0}B(\theta_0)+(\Xi-1)C(\theta_0),\end{aligned}$$

where $\Xi=\sum_{i,j\in E}\frac{\left(p_{i(j-1)}(t;\theta_0)\right)^2}{p_{ij}(t;\theta_0)}\pi_{\theta_0}(i)$. The next step is to show that $I(\theta_0)$ is invertible, and this holds iff the determinant $\det(I(\theta_0))=\frac{t^2}{\mu_0^2}\left(\rho_0(1+\mathrm{e}^{-\mu_0 t})(\Xi-1)-1\right)$ is non-zero, which is to say that $\Xi\neq\frac{1+\rho_0(1+\mathrm{e}^{-\mu_0 t})}{\rho_0(1+\mathrm{e}^{-\mu_0 t})}$. Through Lemma 2 we have that

$$\begin{aligned}\Xi &= \sum_{i,j\in E}\left(\frac{(j+1)}{\frac{\alpha_0}{\mu_0}\,\mathrm{e}^{-\mu_0 t}}\frac{p_{i(j+1)}(t;\theta_0)}{p_{ij}(t;\theta_0)}+\frac{j-i}{\rho_0}-(\mathrm{e}^{\mu_0 t}-1)\right)p_{i(j-1)}(t;\theta_0)\pi_{\theta_0}(i)\\ &= \frac{1}{\frac{\alpha_0}{\mu_0}\,\mathrm{e}^{-\mu_0 t}}\sum_{i,j\in E}(j+2)\frac{p_{i(j+2)}(t;\theta_0)}{p_{i(j+1)}(t;\theta_0)}p_{ij}(t;\theta_0)\pi_{\theta_0}(i)+(1-\mathrm{e}^{\mu_0 t})\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}p_{ij}(t;\theta_0)\pi_{\theta_0}(i)\\ &+\frac{1}{\rho_0}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(j+1-i)p_{ij}(t;\theta_0)\pi_{\theta_0}(i)=:S_1+S_2+S_3.\end{aligned}$$

Since $\pi_{\theta_0}(\cdot)=\mathbb{P}(Poi(\alpha_0/\mu_0)\in\cdot)$ is the invariant distribution under $\theta_0$, we have that $S_2=(1-\mathrm{e}^{\mu_0 t})\sum_{j=0}^{\infty}\sum_{i=0}^{\infty}p_{ij}(t;\theta_0)\pi_{\theta_0}(i)=\sum_{j=0}^{\infty}\pi_{\theta_0}(j)=1-\mathrm{e}^{\mu_0 t}$ and $S_3=\frac{1}{\rho_0}(1+\sum_{j=0}^{\infty}j\sum_{i=0}^{\infty}p_{ij}(t;\theta_0)\pi_{\theta_0}(i)-\sum_{i=0}^{\infty}i\pi_{\theta_0}(i)\sum_{j=0}^{\infty}p_{ij}(t;\theta_0))=\frac{1}{\rho_0}$ so that $\Xi=S_1+1-\mathrm{e}^{\mu_0 t}+\frac{1}{\rho_0}=S_1+\frac{1+\mathrm{e}^{-\mu_0 t}+\rho_0(\mathrm{e}^{-\mu_0 t}-\mathrm{e}^{\mu_0 t})}{\rho_0(1+\mathrm{e}^{-\mu_0 t})}$. Hereby the invertibility condition is translated into

$$0\neq S_1-\frac{1+\rho_0(1+\mathrm{e}^{-\mu_0 t})-(1+\mathrm{e}^{-\mu_0 t}+\rho_0(\mathrm{e}^{-\mu_0 t}-\mathrm{e}^{\mu_0 t}))}{\rho_0(1+\mathrm{e}^{-\mu_0 t})}=S_1+\frac{\mathrm{e}^{-\mu_0 t}-\rho_0(1+\mathrm{e}^{\mu_0 t})}{\rho_0(1+\mathrm{e}^{-\mu_0 t})}. \qquad (4.9)$$

Clearly $S_1>0$ and since $\rho_0(1+\mathrm{e}^{-\mu_0 t})>0$ we have that the right hand side of (4.9) is positive if $\mathrm{e}^{-\mu_0 t}\geq\rho_0(1+\mathrm{e}^{\mu_0 t})=\frac{\alpha_0}{\mu_0}(\mathrm{e}^{\mu_0 t}-\mathrm{e}^{-\mu_0 t})$, which can be expressed as $\mathrm{e}^{-2\mu_0 t}(\alpha_0+\mu_0)\geq\alpha_0$. By taking logarithms on both sides of the latter inequality we end up with $(\log(\alpha_0+\mu_0)-\log(\alpha_0))/\mu_0\geq 2t$,





ACCEPTED MANUSCRIPT

which holds by assumption. This implies that $I(\theta_0)$ is invertible, with inverse given by (4.8), and by an appeal to (Dehay & Yao, 2007) we obtain $\sqrt{n}\left((\hat{\alpha}_n, \hat{\mu}_n) - (\alpha_0, \mu_0)\right) \xrightarrow{d} N\left(\mathbf{0}, I(\theta_0)^{-1}\right)$, as $n \to \infty$.

□

Regarding the invertibility condition $C(\alpha_0, \mu_0) := (\log(\alpha_0 + \mu_0) - \log(\alpha_0))/\mu_0 \geq 2t$ in Theorem 6, by the mean value theorem we have that $1/(\alpha_0 + \mu_0) < C(\alpha_0, \mu_0) < 1/\alpha_0$. This means that the condition will be satisfied if $2t(\alpha_0 + \mu_0) \leq 1$, which is to say that we may sample the process relatively sparsely when both $\alpha_0$ and $\mu_0$ are small and, conversely, we have to follow a tight sampling scheme when $\max(\alpha_0, \mu_0)$ becomes large. In other words, if there is a lot of activity going on in the process we need to monitor it more frequently, compared to when arrivals and deaths occur rarely, in order to ascertain that the condition is fulfilled. Note further that when $\alpha_0$ increases, with $\mu_0$ kept fixed, we are required to sample the process more densely in order for the condition to hold ($\lim_{\alpha_0 \to \infty} C(\alpha_0, \mu_0) = 0$) and when we decrease $\alpha_0$, with $\mu_0$ fixed, it is more likely that the condition is fulfilled ($\lim_{\alpha_0 \to 0} C(\alpha_0, \mu_0) = \infty$). Furthermore, when we let $\mu_0$ increase while keeping $\alpha_0$ fixed, we move towards a situation where the condition will not be fulfilled ($\lim_{\mu_0 \to \infty} C(\alpha_0, \mu_0) = 0$). When we decrease $\mu_0$, with $\alpha_0$ fixed, so that $N(s)$ is approaching a Poisson process, we get that $\lim_{\mu_0 \to 0} C(\alpha_0, \mu_0) = 1/\alpha_0$ so that the condition will be fulfilled provided that $\alpha_0$ is not too big (note, however, that when $N(s)$ is a Poisson process, by exploiting its Lévy process properties and the central limit theorem, one can easily show that the ML-estimator $\hat{\alpha}_n$ is asymptotically Gaussian).

Note that the results in these theorems still may hold for $N(s)$ under a different sampling scheme than equidistant sampling, although the approach used to prove the results may be different.

#### 4.1.1 Numerical evaluations

Consider now two different sets of parameter pairs, $(\alpha_0, \mu_0) = (2, 0.05)$ and $(\alpha_0, \mu_0) = (0.4, 0.01)$, each from which we simulate 50 independent sample paths of $N(s)$ on $[0, T]$, $T = 150$, $N(0) = 0$. Thereafter we sample each at times $T_k = kt$, $t = 1$, $k = 1, \ldots, 150$, and based on these discrete observations, for each we estimate $(\alpha_0, \mu_0)$ three times; up to time 50, up to time 100 and up to time 150. Table 1 and Table 2 display the estimated means, biases, standard errors (S.E.), skewnesses (the skewness of a normal distribution is 0) and kurtoses (the kurtosis of a normal distribution is 3) for each parameter pair, $(\alpha_0, \mu_0)$, based on its 50 discretely sampled sample paths.





ACCEPTED MANUSCRIPT

Table 1: Estimated moments of the estimator for $(\alpha_0, \mu_0) = (2, 0.05)$, based on the 50 sample paths sampled at times $T_k = kt$, $t = 1$, $k = 1, \ldots, T$.

| | Mean | Bias (%) | S.E. | Skewness | Kurtosis |
|---|---|---|---|---|---|
| $T = 50$: $\hat{\alpha}_T$ | 2.0305 | 1.5 | 0.4406 | 1.3284 | 5.0738 |
| $\hat{\mu}_T$ | 0.0503 | 0.6 | 0.0175 | 1.1350 | 4.4391 |
| $T = 100$: $\hat{\alpha}_T$ | 2.0605 | 3.0 | 0.3729 | 0.4076 | 2.6461 |
| $\hat{\mu}_T$ | 0.0511 | 2.2 | 0.0112 | 0.5632 | 2.6832 |
| $T = 150$: $\hat{\alpha}_T$ | 2.0640 | 3.2 | 0.2667 | 0.1881 | 2.4832 |
| $\hat{\mu}_T$ | 0.0517 | 3.4 | 0.0081 | 0.4088 | 2.2849 |

We can see how the skewness of the data goes through a stepwise reduction for every additional 50 time units we utilise in the estimation (Table 1). As a measure of the heaviness of the tails we consider the kurtosis estimates given in Table 1; we see a strong reduction after the first 50 time units, going from something fairly heavy tailed to something a bit more light tailed than a Gaussian distribution (note that there are robustness issues with kurtosis estimators based on sample fourth moment estimators). From Table 1 we also see that already after 50 sampled time units the biases are quite small. Hence, the consistency of the estimator $(\hat{\alpha}_n, \hat{\mu}_n)$ becomes clear quite quickly and although the parameter pair $(\alpha_0, \mu_0) = (2, 0.05)$ does not fulfil the invertibility condition $(\log(\alpha_0 + \mu_0) - \log(\alpha_0))/\mu_0 \geq 2t = 2$ of Theorem 6, it asymptotically seems to behave Gaussian, thus indicating that the invertibility condition $(\log(\alpha_0 + \mu_0) - \log(\alpha_0))/\mu_0 \geq 2t$ may be improved.

Table 2: Estimated moments of the estimator for $(\alpha_0, \mu_0) = (0.4, 0.01)$, based on the 50 sample paths sampled at times $T_k = kt$, $t = 1$, $k = 1, \ldots, T$.

| | Mean | Bias (%) | S.E. | Skewness | Kurtosis |
|---|---|---|---|---|---|
| $T = 50$: $\hat{\alpha}_T$ | 0.4751 | 18.8 | 0.1372 | -0.1604 | 2.1189 |
| $\hat{\mu}_T$ | 0.0137 | 37.0 | 0.0080 | 0.4021 | 2.3971 |
| $T = 100$: $\hat{\alpha}_T$ | 0.4251 | 5.4 | 0.1412 | 1.1873 | 4.4208 |
| $\hat{\mu}_T$ | 0.0126 | 26.0 | 0.0057 | 0.6537 | 3.2866 |
| $T = 150$: $\hat{\alpha}_T$ | 0.4166 | 4.2 | 0.1314 | 0.1742 | 2.9146 |
| $\hat{\mu}_T$ | 0.0123 | 23.0 | 0.0064 | 0.6493 | 2.8343 |

As opposed to the previous choice of parameters, $(\alpha_0, \mu_0) = (0.4, 0.01)$ does fulfil the invertibility condition of Theorem 6. From the estimated means/biases and kurtoses in Table 2 we can verify that also here the tails of the empirical distribution approach those of a normal distribution. Regarding the skewness of the estimates, we see from Table 2 that we end up at values fairly close to 0, i.e. close to that of a Gaussian distribution. Hence, as expected, also here we see that $(\hat{\alpha}_n, \hat{\mu}_n)$ approaches the actual parameter pair and at $T = 150$ we have strong indications of approximate Gaussianity of $(\hat{\alpha}_n, \hat{\mu}_n)$.

### 4.2 Asymptotic properties of the HSID-process ML-estimators

Turning now to the HSID-process $Y(s)$, consider sampling it at times $T_n = nt$, $n \in \mathbb{N}$, $t > 0$, and estimating its parameters with the ML-estimator (4.7). We denote by $\theta_W^0 \in \Theta_W^0 \subseteq \mathbb{R}^m$, $m \geq 1$, the parameter of the underlying spatial distribution $P_W(\cdot)$ and by $\mathbb{E}_{\theta_W^0}[\cdot]$ expectation under $\theta_W^0$. Here consistency is readily obtained if we impose certain conditions on $f_W(\cdot;\theta_W)$.

**Corollary 1.** *Assume that both $\Theta$ and $\Theta_W$ are compact. Assume further that i) $f_W(\cdot;\theta_W)$ is identifiable, ii) $\log f_W(\xi;\theta_W)$ is upper semi-continuous in $\theta_W$ for all $\xi \in W$, iii) there exists a function $K(\xi)$ such that $\mathbb{E}_{\theta_W^0}[|K(\xi)|] < \infty$ (for a random spatial location $\xi \sim P_W(\cdot)$) and $\log f_W(\xi;\theta_W) - \log f_W(\xi;\theta_W^0) \leq K(\xi)$ for all $\xi \in W$ and $\theta_W \in \Theta_W$, and iv) for all $\theta_W \in \Theta_W$ and $\rho > 0$ sufficiently small, $\sup_{\theta_W:|\theta_W - \theta_W^0|<\rho} f_W(\xi;\theta_W)$ is measurable in $\xi$. Then the estimator (4.7) is strongly consistent, i.e. $(\hat{\alpha}_n, \hat{\mu}_n, \hat{\theta}_n^W) \xrightarrow{a.s.} (\alpha_0, \mu_0, \theta_W^0)$, as $n \to \infty$.*

*Proof.* Since $N(s)$ is a positively recurrent Markov process, the total number $M(n) = |\bigcup_{k=1}^n Y(T_k)|$ of distinct points $\xi$ observed in $W$ over the sample grid $T_1, \ldots, T_n$ a.s. tends to infinity as $n \to \infty$. Under the conditions imposed on $f_W$, the log-likelihood function $l_n^W(\theta)$ in (4.6) gives rise to a strongly consistent estimator, i.e. $\hat{\theta}_{M(n)}^W \xrightarrow{a.s.} \theta_W^0$, as $n \to \infty$ (see e.g. (Ferguson, 1996, Theorem 17)). We combine this convergence with the a.s. convergence obtained in Theorem 5 by appealing to Slutsky's theorem (Ferguson, 1996, Theorem 6').

□

## 5 Forest stand data

Given some spatio-temporal point process, we may further assign a mark (additional feature) $m_i$ to each of its points. This results in a *marked spatio-temporal point process*. One particular marked spatio-temporal point process, which is the underlying motivation for this study, is the so-called *growth-interaction process* (Cronie & Särkkä, 2011; Renshaw, Comas, & Mateu, 2009; Särkkä & Renshaw, 2006). It has mainly been used for spatio-temporal modelling of forest stands and it is constructed by assigning a set of dynamic marks $m_i = \{m_i(s)\}_{s\geq 0}$, $i = 1, \ldots, N$, to the points of a HSID-process $Y(s)$ with $P(\cdot) = \mathbb{P}(Uni(W) \in \cdot)$. Here the purpose of $m_i(s)$ is to describe the radius of the $i$th tree at time $s$ or, equivalently, the space $B[x_i, m_i(s)] = \{x \in \mathbb{R}^2 : \|x - x_i\| \leq m_i(s)\}$ occupied by the $i$th tree at time $s$.

We here consider four plots $j = 1, \ldots, 4$ from the Swedish National Forest Inventory, all measured in the years 1985, 1990, 1996 and 2005 within a circular region $W$ of radius 10 m. Although the years of measurement are the same for all of the plots, the ages (sample times) $T_{j,1}$, $T_{j,2}$, $T_{j,3}$, $T_{j,4}$, $j = 1, \ldots, 4$, of the plots at the measurement times 1985, 1990, 1996, 2005 are different. Each plot consists of at least 90% Scots pines (we have chosen to not remove the non-Scots pine trees in the data set) and only trees with a radius at breast height (1.3 m above ground), rbh, of at least 0.05 m are considered. Given some sample time $T_{j,k}$, $k = 1, \ldots, 4$, the two features measured are the location $x_i$ and the rbh $m_{ik}$ of each tree present at $T_{j,k}$. As a result, each data set is represented by $\mathbb{X}_j = (\mathbb{X}_{j,k})_{k=1}^4 = (\{(x_i, m_{ik}) : i \in \Omega_{j,k}\})_{k=1}^4$, $\Omega_{j,k} = \{i : \text{tree } i \text{ is alive/visible at time } T_{j,k}\}$, where $n_{T_{j,k}} = |\Omega_{j,k}|$ is the number of trees present at time $T_{j,k}$.

Assuming that we want to model each $\mathbb{X}_j$ by means of the growth-interaction process, we need to fit to each a HSID-process with $P(\cdot) = \mathbb{P}(Uni(W) \in \cdot)$. Since $W$ is known, we may ignore the spatial uniform distribution part, whereby this amounts to fitting ID-processes to the data in Table 3. Note that for each of these processes we have that $N(T_{j,0}) = 0$. For each $j = 1, \ldots, 4$, we here choose to generate estimates of $\alpha$ and $\mu$ (see Table 3) based on $T_{j,1}$, $T_{j,2}$, $T_{j,3}$ and then use the data from $T_{j,4}$ as reference, in order to evaluate whether the parameter estimates may be used to predict the actual population sizes $n_{T_{j,4}}$ (Table 4).

Table 3: Information about the data sets $\mathbb{X}_j$ $(j = 1, \ldots, 4)$: $T_{j,k}$ $(k = 1, 2, 3)$ is the $k$th inventory time (stand age) and $n_{T_{j,k}}$ is the number of distinct trees observed at $T_{j,k}$. $\hat{\alpha}$ and $\hat{\mu}$ are estimated arrival- and death rates.

| $j$ | $T_{j,1}$ | $T_{j,2}$ | $T_{j,3}$ | $n_{T_{j,1}}$ | $n_{T_{j,2}}$ | $n_{T_{j,3}}$ | $\hat{\alpha}$ | $\hat{\mu}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 22 | 27 | 33 | 13 | 26 | 43 | 0.1944 | 2.3488 |
| 2 | 32 | 37 | 43 | 24 | 36 | 48 | 0.4958 | 4.3264 |
| 3 | 23 | 28 | 34 | 40 | 51 | 53 | 0.4191 | 2.5816 |
| 4 | 45 | 50 | 56 | 9 | 15 | 15 | 0.1344 | 2.8145 |

Regarding the predictions in Table 4, both $T_{j,4}$ and $\Delta T_{j,4} = T_{j,4} - T_{j,3}$ are fairly large for all of the data sets. Hereby, by exploiting equation (2.3), we may predict the number of trees present at time $T_{j,4}$ by $\hat{N}(T_{j,4}) = \hat{\mathbb{E}}[N(T_{j,4}) \mid N(T_{j,3}) = n_{T_{j,3}}] = n_{T_{j,3}} \mathrm{e}^{-\hat{\mu}\Delta T_{j,4}} + \hat{\rho} \approx \hat{\rho} = \hat{\mathbb{E}}[N(T_{j,4})]$. The related prediction error (PE) and standard deviation (SD) are given by PE $= \hat{N}(T_{j,4}) - n_{T_{j,4}}$, SD $= \widehat{\mathrm{Var}}(N(T_{j,4}) \mid N(T_{j,3}) = n_{T_{j,3}}) = n_{T_{j,3}}(n_{T_{j,3}} - 1)\,\mathrm{e}^{-2\hat{\mu}\Delta T_{j,4}} + (1 + 2\hat{\rho}) n_{T_{j,3}} \mathrm{e}^{-\hat{\mu}\Delta T_{j,4}} + \hat{\rho}^2 + \hat{\rho} - (n_{T_{j,3}} \mathrm{e}^{-\hat{\mu}\Delta T_{j,4}} + \hat{\rho})^2 \approx \hat{\rho} = \widehat{\mathrm{Var}}(N(T_{j,4}))$. Also SD is a consequence of equation (2.3). The obtained results may be found in Table 4. Note that $\widehat{\mathbb{E}}[N(T_{j,4})] = \hat{\alpha}\nu(W)(1 - \mathrm{e}^{-\hat{\mu}T_{j,4}})/\hat{\mu} \approx \hat{\alpha}\nu(W)/\hat{\mu}$, where $\nu(W) = 10^2\pi \approx 314$ is the size of the study region.

Table 4: Observed and expected number (with PE and SD) of alive trees, as well as stand ages, for the plots at the fourth inventory occasion.

| $j$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $T_{j,4}$ | 42 | 52 | 43 | 65 |
| $n_{T_{j,4}}$ | 52 | 54 | 36 | 16 |
| $\hat{N}(T_{j,4})$ | 26.6569 | 35.9840 | 50.9751 | 14.9944 |
| PE (%) | -25.3432 (51.26%) | -18.0160 (33.36%) | 14.9751 (41.60%) | -1.0057 (-6.29%) |
| SD | 5.1630 | 5.9987 | 7.1397 | 3.8723 |

From Table 4 we see that the very few sample times give rise to a PE which, in most cases, becomes quite extensive. Note, however, that the large sample times for $j = 4$ seem to decrease PE substantially. Also SD seems to be decreased by larger sample times. It is clear here that, in general, one should have access to data sets which have been been sampled more extensively.

## 6 Discussion

In this paper we have considered the ML-estimation for the ID-process and the HSID-process. In particular, under an equidistant sampling scheme, $T_k = kt$, $t > 0$, $k = 1, \dots, n$, for the ID-process both consistency and asymptotic normality have been proved and for the HSID-process consistency has been obtained. The empirical distribution of the estimates of the ID-process show strong indications of Gaussianity, even when the imposed invertibility condition of Theorem 6 is not fulfilled. In addition, in connection to the growth-interaction process, we fit a HSID-process (with uniform spatial distribution kernel) to four different forest stands, primarily consisting of Scots pines.

Regarding possible future work, our main interests are related to the HSID-process. Note that one could consider some more sophisticated spatial distribution $P(\cdot)$ and obtain a more general definition of a *spatial immigration-death process*. For instance, we could let the spatial distribution





ACCEPTED MANUSCRIPT

of a newcomer depend on the locations of the individuals already present. If we additionally put a spatial Markov structure (see e.g. (van Lieshout, 2000)) on the distribution of a newcomer (through a symmetric, reflexive neighbourhood relation on $W$), we would obtain one possible definition of a *Markov spatial immigration-death process*. Such processes naturally have interesting properties and spatial statistical applications, as they may be useful tools for modelling many different kinds of spatio-temporal dynamics (e.g. the spatio-temporal development of a forest stand). Additionally we may also consider marked versions of these models (see e.g. (van Lieshout, 2000)) and we may for instance construct spatio-temporal Boolean models by marking each point with a disk/ball (see e.g. (Stoyan et al., 1995)). One further aim here is to further develop a (pseudo)likelihood estimation scheme based on discretely sampled HSID-process with general spatial kernels, in particular, and for discretely sampled general spatial birth-death processes in general.

## Acknowledgements

The authors would like to thank Aila Särkkä (Chalmers University of Technology), Bo Ranneby and the late Lennart Norell (Swedish University of Agricultural Sciences) for useful comments and feedback. This research has been supported by the Swedish Research Council and the Swedish Foundation for Strategic Research.

## References

Asmussen, S. (2003). *Applied probability and queues* (2nd ed.). New York: Springer.

Basawa, I., & Prakasa Rao, B. (1980). *Statistical inference for stochastic processes*. London: Academic press.

Berthelsen, K., & Møller, J. (2002). Spatial jump processes and perfect simulation. In K. Mecke & D. Stoyan (Eds.), *Morphology of condensed matter: Physics and geometry of spatially complex systems* (pp. 391–417). Berlin: Springer.

Bhat, B. R., & Adke, S. R. (1981). Maximum likelihood estimation for branching processes with immigration. *Advances in Applied Probability*, *13*, 498–509.

Billingsley, P. (1961). *Statistical inference for markov processes*. Chicago: The University of Chicago Press.

Cronie, O., & Särkkä, A. (2011). Some edge correction methods for marked spatio-temporal point process models. *Computational Statistics & Data Analysis*, *55*, 2209–2220.

Cronie, O., & Yu, J. (2010). *Maximum likelihood estimation in a discretely observed immigration-death process* (Tech. Rep. No. 2010:1). Swedish University of Agricultural Sciences. (http://biostochastics.slu.se/publikationer/dokument/Report2010_01.pdf)

Dehay, D., & Yao, J.-F. (2007). On likelihood estimation for discretely observed markov jump processes. *Australian & New Zealand Journal of Statistics*, *49*, 93–107.

Ferguson, T. (1996). *A course in large sample theory*. Chapman & Hall.

Gani, J., & Swift, R. (2011). An immigration-death process with a time-dependent delay. *Communications in Statistics – Theory and Methods*, *40*, 1236–1240.

Gibson, G., & Renshaw, E. (2001). Inference for immigration-death processes with single and paired immigrants. *Inverse Problems*, *17*, 455–466.

Gillespie, C., & Renshaw, E. (2005). The evolution of a single-paired immigration death process. *Proc. R. Soc. A*, *461*, 1563–1581.

Grimmett, G., & Stirzaker, D. (2001). *Probability and random processes* (3rd ed.). Oxford: Oxford university press.

Guttorp, P. (1991). *Statistical inference for branching processes*. New York: Wiley Series in Probability and Mathematical Statistics.

Illian, J., Penttinen, A., Stoyan, H., & Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*. West Sussex: Wiley-Interscience.

Kallenberg, O. (2002). *Foundations of modern probability* (2nd ed.). New York: Springer.

Keiding, N. (1975). Maximum likelihood estimation in the birth-and-death process. *The Annals of Statistics*, *3*, 363–372.

van Lieshout, M. (2000). *Markov point processes and their applications*. London: Imperial College Press.

Møller, J., & Waagepetersen, R. (2004). *Statistical inference and simulation for spatial point processes*. Boca Raton: Chapman & Hall/CRC Press.

Norris, J. (1997). *Markov chains*. Cambridge: Cambridge University Press.

Renshaw, E., Comas, C., & Mateu, J. (2009). Analysis of forest thinning strategies through the development of space-time growth-interaction simulation models. *Stochastic Environmental*





ACCEPTED MANUSCRIPT

*Research and Risk Assessment*, *23*, 275–288.

Särkkä, A., & Renshaw, E. (2006). The analysis of marked point patterns evolving through space and time. *Computational Statistics & Data Analysis*, *51*, 1698–1718.

Stoyan, D., Kendall, W., & Mecke, J. (1995). *Stochastic geometry and its applications* (2nd ed.). West Sussex: John Wiley & sons.

Stoyan, D., & Penttinen, A. (2000). Recent applications of point process methods in forestry statistics. *Statistical Science*, *15*, 61–78.



# Appendices

## A Bounds and conditions

The following bounds and regularity conditions (derived in (Cronie & Yu, 2010)) are exploited in Theorem 6. Note that here the order of differentiation is irrelevant, i.e. for example $D^3_{112}\log q(\theta; i, j) = D^3_{121}\log q(\theta; i, j) = D^3_{211}\log q(\theta; i, j)$.

$$\begin{aligned} \left|D^3_{111}\log q(\theta;i,j)\right| &\leq \frac{\left|D^3_{111}q(\theta;i,j)\right|}{q(\theta;i,j)} + \frac{3\left|D^2_{11}q(\theta;i,j)\right|\left|D_1\log q(\theta;i,j)\right|}{q(\theta;i,j)} \\ &\quad + \left|D_1\log q(\theta;i,j)\right|^3 \\ &< \frac{j+(j+\alpha t)^2}{\alpha}t + \frac{j}{\alpha}\left(\frac{j-1}{\alpha^2} + \left(\frac{j-1}{\alpha}+t\right)^2\right) \\ &\quad + 3\frac{j+(j+\alpha t)^2}{\alpha^2}\left(\frac{j}{\alpha}+t\right) + \left(\frac{j}{\alpha}+t\right)^3 =: B_{111}(\alpha,\mu,t,j,i), \end{aligned} \tag{A.1}$$



ACCEPTED MANUSCRIPT

$$
\begin{aligned}
& \left|D^3_{112}\log q(\theta;i,j)\right| \leq \qquad \text{(A.2)} \\
& \leq \frac{\left|D^3_{112} q(\theta;i,j)\right|}{q(\theta;i,j)} + 2\frac{\left|D^2_{12} q(\theta;i,j)\right| \left|D_1 \log q(\theta;i,j)\right|}{q(\theta;i,j)} \\
& + \left|D_2 \log q(\theta;i,j)\right| \left|(D_1 \log q(\theta;i,j))^2 - D^2_{11}\log q(\theta;i,j)\right| \\
< \; & 2\frac{t^2}{\alpha}\left(\alpha t + \frac{\mu(j^2+j)}{\alpha(1-\mathrm{e}^{-\mu t})} + 2j\right) + t^2\left(\alpha t + \frac{(j+i+1)t}{1-\mathrm{e}^{-\mu t}} + j\right) \\
& + \frac{j^2+j}{\alpha^2}\left(\alpha t + \frac{(j+i-1)t}{1-\mathrm{e}^{-\mu t}} + j - 2\right) + 2t\frac{j}{\alpha}\left(\alpha t + \frac{(j+i)t}{1-\mathrm{e}^{-\mu t}} + j - 1\right) \\
& + 2\left(\frac{j}{\alpha} + t\right)\left(\frac{(j^2+j)t}{\alpha} + \alpha t^3 + \frac{j(j+i)t + (j+\alpha t)(\alpha t^2 + (3j+i)t)}{(1-\mathrm{e}^{-\mu t})\alpha}\right. \\
& \left. + t^2(1+j) + \frac{j+i}{\mu}t\right) + \frac{\alpha t^2 + (3j+i)t}{1-\mathrm{e}^{-\mu t}}\left(\frac{j}{\alpha^2} + 2\left(\frac{j}{\alpha}+t\right)^2\right) \\
=: \; & B_{112}(\alpha,\mu,t,j,i)
\end{aligned}
$$





ACCEPTED MANUSCRIPT

$$
\begin{aligned}
&\left|D^3_{122}\log q(\theta;i,j)\right| \leq && \text{(A.3)}\\
\leq\quad & \frac{\left|D^3_{122}q(\theta;i,j)\right|}{q(\theta;i,j)} + 2\frac{\left|D^2_{12}q(\theta;i,j)\right|\left|D_2\log q(\theta;i,j)\right|}{q(\theta;i,j)}\\
& + \left|D_1\log q(\theta;i,j)\right|\left((D_2\log q(\theta;i,j))^2 + D^2_{22}\log q(\theta;i,j)\right)\\
<\quad & \frac{2+\mathrm{e}^{-\mu t}}{1-\mathrm{e}^{-\mu t}}t^2\left(\frac{j}{\alpha}+t\right)\\
& + \frac{j}{\alpha}t^2\Bigg((j-1)^2 + 2(j+i-1)(j-1) + \left(2\alpha t + 1 + 2\alpha t + 2\mu^2t^2(1+\alpha t)\right)(j-1)\\
& + \alpha^2t^2 + \alpha t(\mu^2t^2+2) + (j+i-1)^2\mu^2t^2 + 2\alpha t(j+i-1) + (j+i-1)\mu^2t^2\Bigg)\\
& + 2\frac{\alpha t^2 + (3j+i)t}{1-\mathrm{e}^{-\mu t}}\Bigg(\frac{(j^2+j)t}{\alpha} + \alpha t^3 + \frac{j(j+i)t}{(1-\mathrm{e}^{-\mu t})\alpha} + t^2(1+j) + \frac{j+i}{\mu}t\\
& + \frac{(j+\alpha t)(\alpha t^2+(3j+i)t)}{(1-\mathrm{e}^{-\mu t})\alpha}\Bigg) + 2\left(\frac{j}{\alpha}+t\right)\left(\frac{\alpha t^2+(3j+i)t}{1-\mathrm{e}^{-\mu t}}\right)^2\\
& + t^2\left(\frac{j}{\alpha}+2t\right)\Bigg(j^2 + 2(j+i)j + \left(2\alpha t + 1 + 2\alpha t + 2\mu^2t^2(1+\alpha t)\right)j\\
& + \alpha^2t^2 + \alpha t(\mu^2t^2+2) + (j+i)^2\mu^2t^2 + 2\alpha t(j+i) + (j+i)\mu^2t^2\Bigg)\\
=:\quad & B_{122}(\alpha,\mu,t,j,i),
\end{aligned}
$$

$$
\begin{aligned}
&\left|D^3_{222}\log q(\theta;i,j)\right| \leq && \text{(A.4)}\\
\leq\quad & \frac{\left|D^3_{222}q(\theta;i,j)\right|}{q(\theta;i,j)} + \left|D_2\log q(\theta;i,j)\right|\left((D_2\log q(\theta;i,j))^2 + 3\left|D^2_{22}\log q(\theta;i,j)\right|\right)\\
<\quad & A_{\frac{1}{p_{ij}(t;\theta)}\left|\frac{\partial^3 p_{ij}(t;\theta)}{\partial\mu^3}\right|} + 4\left(\frac{\alpha t^2+(3j+i)t}{1-\mathrm{e}^{-\mu t}}\right)^3 + \ 3t^2\left(\frac{\alpha t^2+(3j+i)t}{1-\mathrm{e}^{-\mu t}}\right)\times\\
& \times\Bigg(j^2 + 2(j+i)j + \left(2\alpha t + 1 + 2\alpha t + 2\mu^2t^2(1+\alpha t)\right)j + \alpha^2t^2 + \alpha t(\mu^2t^2+2)\\
& + (j+i)^2\mu^2t^2 + 2\alpha t(j+i) + (j+i)\mu^2t^2\Bigg) =: B_{222}(\alpha,\mu,t,j,i),
\end{aligned}
$$

where

$$
\begin{aligned}
&\frac{\left|D^3_{222}q(\theta;i,j)\right|}{q(\theta;i,j)} < \\
&< \left[\frac{6}{\mu^2}+\frac{8t}{\mu}+\frac{6t^2}{(1-\mathrm{e}^{-\mu t})^2}\right]\Big[j^2+2(j+i)j+\left(4\alpha t+1+2\mu^2t^2(1+\alpha t)\right)j \\
&+\alpha^2t^2+\alpha t(\mu^2t^2+2)+(j+i)^2\mu^2t^2+2\alpha t(j+i)+(j+i)\mu^2t^2\Big]t^2 \\
&+\left[2t^3(\mu t+1)+t^2(1+\mu t)^2\left(\alpha t^2+\frac{(3j+i)t}{1-\mathrm{e}^{-\mu t}}\right)\right]j^2 \\
&+t^2\left[2\mu t(j+i)+4\alpha t+1+2\mu^2t^2(1+\alpha t+j+i)\right]\left[\alpha t^2+\frac{(3j+i)t}{1-\mathrm{e}^{-\mu t}}\right]j \\
&+2t^3\left[(j+i)(1+3\mu t)+i\mu t(1+\mu t)+(1+\mu t)^2+\alpha\left(1+(\mu t)^2+\mu^2t^3\right)\right]j \\
&+t^2\left[(j+i)^2\mu^2t^2+2\alpha t(j+i)+(j+i)\mu^2t^2+\alpha^2+\alpha t(\mu^2t^2+2)\right]\left[\alpha t^2+\frac{(3j+i)t}{1-\mathrm{e}^{-\mu t}}\right] \\
&+t^3\Bigg(2(\mu t+(\alpha t+\alpha t(\mu t)^2))(j+i)+2\mu ti(\mu t+\alpha t+1)+j(2+\mu t)\mu t \\
&+2\frac{\alpha(1+\alpha t)}{\mu^2 t}+\alpha\mu t^2\left(3+4\alpha t+2\mu t+3(\mu t)^2\right)\Bigg) =: A_{\frac{1}{p_{ij}(t;\theta)}\left|\frac{\partial^3 p_{ij}(t;\theta)}{\partial\mu^3}\right|}.
\end{aligned}
$$

$$
\sum_{j=0}^{\infty} D^2_{11}q(\theta_0;i,j)=\frac{\rho^2}{\alpha^2}\left(\sum_{j=0}^{\infty}p_{i(j-2)}(t;\theta)-2\sum_{j=0}^{\infty}p_{i(j-1)}(t;\theta)+\sum_{j=0}^{\infty}p_{ij}(t;\theta)\right)=0 \tag{A.5}
$$





ACCEPTED MANUSCRIPT

$$\sum_{j=0}^{\infty} D_{12}^2 q(\theta_0; i, j) = \frac{\rho(\tau-\mu t)}{(1-\mathrm{e}^{-\mu t})\alpha\mu}\left(\sum_{j=0}^{\infty} p_{ij}(t;\theta) - \sum_{j=0}^{\infty} p_{i(j-1)}(t;\theta)\right) \tag{A.6}$$
$$+\frac{\rho^2(\tau-\mu t)}{(1-\mathrm{e}^{-\mu t})\alpha\mu}\left(\sum_{j=0}^{\infty} p_{i(j-1)}(t;\theta) - \sum_{j=0}^{\infty} p_{i(j-2)}(t;\theta)\right)$$
$$+\frac{\rho^2\tau}{(1-\mathrm{e}^{-\mu t})\alpha\mu}\left(\sum_{j=0}^{\infty} p_{i(j-1)}(t;\theta) - \sum_{j=0}^{\infty} p_{ij}(t;\theta)\right)$$
$$-\frac{\rho\mu t i\,\mathrm{e}^{-\mu t}}{(1-\mathrm{e}^{-\mu t})\alpha\mu}\left(\sum_{j=0}^{\infty} p_{ij}(t;\theta) - \sum_{j=0}^{\infty} p_{i(j-1)}(t;\theta)\right)$$
$$+\frac{\rho\mu t}{(1-\mathrm{e}^{-\mu t})\alpha\mu}\sum_{j=0}^{\infty} p_{ij}(t;\theta)$$
$$+\frac{\rho\mu t}{(1-\mathrm{e}^{-\mu t})\alpha\mu}\left(\underbrace{\sum_{j=0}^{\infty} j p_{ij}(t;\theta)}_{=\mathbb{E}[N(h+t)|N(h)=i]} - \underbrace{\sum_{j=0}^{\infty} j p_{i(j-1)}(t;\theta)}_{=1+\mathbb{E}[N(h+t)|N(h)=i]}\right) = 0,$$

$$\frac{\alpha^2}{\rho^2}\sum_{j=0}^{\infty} D_{22}^2 q(\theta_0; i, j) = (\tau-\mu t)^2 \sum_{j=0}^{\infty}\left(\rho p_{i(j-1)}(t;\theta) + \rho^2 p_{i(j-2)}(t;\theta)\right) \tag{A.7}$$
$$+ 2\mu t\,(\tau-\mu t)\,\underbrace{\rho\left(\mathbb{E}[N(h+t)|N(h)=i] + 1 - i\,\mathrm{e}^{-\mu t}\right)}_{\overset{(2.3)}{=}\rho(i\,\mathrm{e}^{-\mu t}+\rho+1-i\,\mathrm{e}^{-\mu t})=\rho(\rho+1)}$$
$$+ \rho\left(-2\rho\tau^2 + (1-\mathrm{e}^{-\mu t})^2 + 2\rho\mu t(1-\mathrm{e}^{-\mu t}) - 2\mu^2 t^2\,\mathrm{e}^{-\mu t}(1+\rho)\right)$$
$$+ \rho^2\tau^2 + \rho(1-\mathrm{e}^{-\mu t})(\mu^2 t^2\,\mathrm{e}^{-\mu t} - 2\tau)$$
$$+ \mu^2 t^2\,\underbrace{\mathbb{E}\left[N(h+t)^2 - 2x\,\mathrm{e}^{-\mu t} N(h+t) + i^2\,\mathrm{e}^{-2\mu t}\,|N(h)=i\right]}_{\substack{\overset{(2.3)}{=} i(i-1)\,\mathrm{e}^{-2\mu t}+(1+2\rho)i\,\mathrm{e}^{-\mu t}+\rho^2+\rho-2x\,\mathrm{e}^{-\mu t}(i\,\mathrm{e}^{-\mu t}+\rho)+i^2\,\mathrm{e}^{-2\mu t} \\ =(1-\mathrm{e}^{-\mu t})i\,\mathrm{e}^{-\mu t}+\rho(\rho+1)}}$$
$$+ \;-2\rho\mu t\tau\,\underbrace{\left(\mathbb{E}[N(h+t)|N(h)=i] - i\,\mathrm{e}^{-\mu t}\right)}_{\overset{(2.3)}{=} i\,\mathrm{e}^{-\mu t}+\rho-i\,\mathrm{e}^{-\mu t}=\rho} + \mu^2 t^2\,\mathrm{e}^{-\mu t}\underbrace{\left(\mathbb{E}[N(h+t)|N(h)=i] - i\right)}_{\overset{(2.3)}{=} i\,\mathrm{e}^{-\mu t}+\rho-i=\rho-i(1-\mathrm{e}^{-\mu t})}$$
$$= \rho\left(1-\mathrm{e}^{-\mu t} - \mu t\,\mathrm{e}^{-\mu t} + 2\mu t\left(\rho+\mathrm{e}^{-\mu t}\right) - \tau\right)\left(1-\mathrm{e}^{-\mu t} - \mu t\,\mathrm{e}^{-\mu t} - \tau\right)$$
$$= \rho\left(\tau + 2\mu t\left(\rho+\mathrm{e}^{-\mu t}\right) - \tau\right)(\tau-\tau) = 0.$$