THESIS FOR THE DEGREE OF LICENTIATE OF PHILOSOPHY

Different Aspects of Inference for Spatio-Temporal Point Processes

Ottmar Cronie

CHALMERS | UNIVERSITY OF GOTHENBURG

Department of Mathematical Sciences
Division of Mathematical Statistics
Chalmers University of Technology and University of Gothenburg
Göteborg, Sweden 2010

Different Aspects of Inference for Spatio-Temporal Point Processes
Ottmar Cronie
NO 2010:5
ISSN 1652-9715

Department of Mathematical Sciences
Division of Mathematical Statistics
Chalmers University of Technology and University of Gothenburg
SE-412 96 Göteborg
Sweden
Telephone +46 (0)31 772 1000

Typeset with LaTeX.
Printed in Göteborg, Sweden 2010

Different Aspects of Inference for Spatio-Temporal Point Processes

Ottmar Cronie

Department of Mathematical Sciences
Division of Mathematical Statistics
Chalmers University of Technology and University of Gothenburg

Abstract

This thesis deals with inference problems related to the *Renshaw-Särkkä growth interaction model* (RS-model). It is a continuous time spatio-temporal point process with time dependent interacting marks, in which the *immigration-death process* (a continuous time Markov chain) controls the arrivals of new marked points as well as their potential life-times. The data considered are marked point patterns sampled at fixed time points.

First we propose three edge correction methods for discretely sampled (marked) spatio-temporal point processes. These are all based on the idea of placing an approximated expected behaviour of our process at hand (based on simulated realisations) outside the study region, which in turn interacts with the data during the estimation. We study the methods and evaluate them numerically in the context of the RS-model. The parameters related to the development of the marks are estimated using the least-squares approach.

Secondly, we propose (approximate) maximum likelihood (ML) estimators for the two parameters of the immigration-death process; the arrival intensity and the death rate. The arrival intensity is assumed to be constant and the death rate is assumed to be proportional to a function of the current mark size of a point. The arrival intensity estimator is constructed to compensate for the (unobserved) individuals arriving and dying between two sampled time points.

When assumed that the death rate is constant we can derive the transition probabilities of the immigration-death process. These in turn give us the exact likelihood of its parameter pair. We are able to reduce the likelihood maximisation problem from two dimensions to one dimension. Furthermore, under the condition that the parameter pair lies in some compact subset of the positive part of the real plane, we manage to show the consistency and the asymptotic normality of its ML-estimator under an equidistant sampling scheme. These results are also evaluated numerically.

Keywords: Asymptotic normality, Consistency, Edge correction, Immigration-death process, Least squares estimation, Maximum likelihood estimation, Spatio-temporal marked point process, Transition probability.

Acknowledgments

To start with I would like to thank my advisor Aila Särkkä for her great support, inspiration, friendship and our many nice conversations. I would also like to thank my co-advisor Jun Yu for his support and inspiration. Furthermore, I would like to thank my co-advisor Anastassia Baxevani for her moral support and friendship.

Other people who I would like to show gratitude for ideas and inspiration related to the writing of this thesis include Claudia Redenbach, Eric Renshaw, Gerald van den Boogaart, Kenneth Nyström, Patrik Albin and Serik Sagitov.

Also, a general thank you goes out to all the people at the department of Mathematical Sciences in Gothenburg. A person at the department who holds a special place in my heart is Daniel Ahlberg - Thank you for your friendship and support. Other people at the department who I would like to thank on a personal note include Alexandra Jauhiainen, Carl Lindberg, Dmitrii Zholud, Emilio Bergroth, Erik Broman, Erik Jakobsson, Frank Eriksson, Fredrik Lindgren, Hermann Douanla Yonta, Jan Lennartsson, Marcus Isaksson, Marcus Warfheimer, Mattias Sundén and Sofia Tapani.

To all my friends outside the department (You know who you are!). Unfortunately you are too many to be mentioned here. Thank you! You have helped me in the process of becoming me.

Finally, to the people who mean the most to me; My family. I love you. Thank you for your eternal love and support!

List of Papers

The licentiate thesis includes the following papers.

I. **Cronie, O.** (2010). Some edge correction methods for marked spatio-temporal point process models. *Preprint.*

II. **Cronie, O.**, Yu, J. (2010). Maximum likelihood estimation in a discretely observed immigration-death process. *Research Report 2010:1, Centre of Biostochastics, Swedish University of Agricultural Sciences.*

Contents

Chapter 1

Introduction

In many different instances in our surrounding world we find point patterns of different kinds. Such patterns include galaxy locations, locations of earthquake epicentres, locations of cell centres and locations of trees in a forest stand. In order to help analysing point patterns the field of spatial statistics has lent a helping hand and has simultaneously also been developing through it. The field of spatial statistics incorporates a few different disciplines within the field of stochastic mathematics and in this thesis we will focus on the parts played by *stochastic geometry* (the study of random geometrical objects) and *spatial point processes* (the study of random point structures) (see e.g. [4, 5, 13, 14, 25]).

Sometimes one does not solely record the locations of the points in a point pattern but also some additional feature connected to each point, such as the radii of the trees in a forest stand or the amount of seismic energy in earthquakes. This additional variable, called a *mark*, can often be quite helpful in explaining the behaviour of the point pattern in question. When focusing on the statistical analysis of these point patterns or marked point patterns, we employ *spatial point processes* or *marked spatial point processes*, respectively (see e.g. [4, 7, 13, 25, 26]). However, to a large extent, the field of spatial (marked) point processes has mainly concentrated on treating marked point patterns within a purely spatial framework. In such a setting one fully ignores that the patterns studied, in fact, almost always are results of evolutionary processes in which the changes occurring among the marks are time dependent. Such situations motivate a change of regime to an approach where one instead considers *spatio-temporal marked point processes* (see e.g. [8, 18, 28]). To fully take the evolution of these marked patterns into consideration it is reasonable to demand that the models describing them should incorporate interaction between marks during the development phases.

The application motivating the work presented in this thesis is found in forestry. Treating a forest stand which is recorded at a specific time point as a static entity, thus ignoring the temporal aspects, the literature offers a wide range of statistical tools for analysing and drawing conclusions about its inherent features, whether one includes marks or not (see e.g. [7, 13, 10, 26] to mention a few). However, here we are interested in modelling the development of a forest stand in both space and time. Figure 1.1 illustrates the type of recorded time series of marked point patterns we refer to – a data set of Swedish Scots pines recorded in 1985, 1990 and 1996 where we have scaled the radii (our marks) for more clear visualisation.

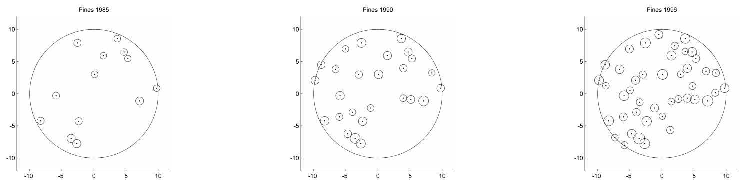

Figure 1.1: Locations and sizes (measured in metres) of Swedish Scots pines recorded in 1985 (left), 1990 (middle) and 1996 (right). The radii of the trees (marks) are scaled by a factor of 10.

A clear risk when formulating the type of spatio-temporal models we are interested in is that the models easily become too involved and we loose both transparency, interpretability and tractability (see e.g. [9]). A spatio-temporal marked point process which manages well to describe this type of spatio-temporal behaviour of a marked population is the so called *Renshaw-Särkkä growth-interaction model* (RS-model) (see [22, 23, 24] or Paper I), which is a combination of stochastic and deterministic components. It has been used to study, among other things, the development of forest stands [24]. Since this model, in spite of being very flexible, is both tractable and easily interpreted it is quite natural to further assess its potential. In the coming chapters we will present and discuss the RS-model together with different statistical tools developed for it (and other models of this type).

In the process of fitting (spatio-temporal) point process models to data sets of the kind presented in Figure 1.1 the following problem emerges. When measurements are made in some bounded study region, the structure of the spatial dependences and interactions existing between points outside and inside the study region remains unobserved. This phenomenon, which in particular concerns those points inside the study region who are close to its boundary, is generally referred to as edge effects. In the context of estimation, if the study region contains a large number of points, the edge effects may not have a large impact on the estimates. However, tree data are often collected in study regions which contain only a small amount of data. In such cases there is a substantial risk that the edge effects generate quite severe biases and we therefore need some type of edge correction method when estimating the model parameters and other summary statistics of interest. In the case of non-temporal analyses a number of methods for edge correction have been devised (see e.g. [13]) but these are, however, not so easily generalized to the spatio-temporal setting. In Paper I we propose three edge correction methods for spatio-temporal marked point processes which all are based on the same idea. By placing an approximated expected behaviour of our spatio-temporal marked point processes outside the study region we let this approximation interact with the data during the estimation. We estimate this expected behaviour by simulating realisations of the process, under a parameter choice based on some non-edge corrected initial estimates, and for each such realisation we generate new edge corrected estimates which we average over to get our edge corrected estimates.

Before we can utilise the edge correction methods developed, we must find estimators which allow us to fit the RS-model. Since this process deals with both space and time we need to be able to fit, not only, the stochastic process controlling the arrivals and deaths of new marked points in time, but also the mechanism controlling the growth of and interaction between the marks. In the RS-model the arrivals and deaths are controlled by a so-called *immigration-death process* – a continuous time Markov chain. In both Paper I and Paper II we will develop estimators for the two parameters of this model and we also present how the growth and interaction parameters of the RS-model are estimated.

In Chapter 2 we introduce the immigration-death process and the RS-model. Then, in Chapter 3, we discuss how we estimate the parameters of the RS-model in an edge corrected setting (see Paper I). In Chapter 3 we also present how maximum likelihood (ML) estimation is carried out in a discretely sampled immigration-death process (Paper II). Finally, in Chapter 4, we look at possible extensions and future work related to the work carried out in Paper I and Paper II.

Chapter 2

The process

We will here define the RS-model, $\mathbb{X}(t) = \{[\mathbf{X}_i, m_i(t)] : i \in \Omega_t\}$, which is a spatio-temporal point process with interacting and size changing marks (see e.g. [24]). It is defined on $[0, \infty)$ in time and spatially we consider it on some region of interest, $W \subseteq \mathbb{R}^d$ (usually $d = 2, 3$), supplied with the Euclidean metric/norm. However, we will start with the immigration-death process, $\{N(t)\}_{t \geq 0}$, since this process is the basis of the RS-model.

2.1 The immigration-death process

The immigration-death process, $\{N(t)\}_{t \geq 0}$, is a time-homogeneous irreducible continuous-time Markov chain (see e.g. [16]) where the possible states for which transitions $i \to j$ are possible are supplied by the state space $E = \{0, 1, \ldots\}$. It is governed by the parameter pair $\theta = (\alpha, \mu)$ which we here assume to take values in some compact parameter space $\Theta \subseteq \mathbb{R}_+^2$.

One way of viewing $\{N(t)\}_{t \geq 0}$ is to treat it as a special case of a birth-death process where the birth rates are given by $\lambda_i = \alpha$, $i = 0, 1, \ldots$, and the death rates are given by $\mu_i = i\mu$, $i = 0, 1, \ldots$, (see [11], p. 268-270). Within this framework the interpretation of $\{N(t)\}_{t \geq 0}$ is the following. By letting the arrivals of new individuals to a population occur according to a Poisson process with intensity α and upon arrival assigning to all individuals independent and exponentially distributed lifetimes with mean $1/\mu$, $N(t)$ gives us the number of individuals alive at time t. Another possibility is to view it as an $M/M/\infty$ queuing system; each customer (arriving according to a Poisson process with intensity α) is being handled by its own server so that its sojourn time in the system is exponential with intensity μ and independent of all other customers.

Being a Markov process, the finite dimensional distributions of $\{N(t)\}_{t\geq 0}$ are controlled by its transition probabilities, $p_{ij}(t;\theta)$ which are given in Paper II.

Proposition 1. *The transition probabilities of the immigration-death process are given by*

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

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 $\rho = \frac{\alpha}{\mu}(1-\mathrm{e}^{-\mu t})$, and $f_{Bin(i,\mathrm{e}^{-\mu t})}(\cdot)$ is the Binomial density with parameters i and $\mathrm{e}^{-\mu t}$. Moreover, we have that

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

The interpretation of $p_{ij}(t;\theta)$ is quite clear. Note that

$$\begin{aligned} f_{Poi(\rho)}(j-k) &= \mathbb{P}(j-k \text{ new arrivals during } (h,h+t)) \\ f_{Bin(i,\mathrm{e}^{-\mu t})}(k) &= \mathbb{P}(k \text{ of the } i \text{ individuals alive at time } h \text{ survive } (h,h+t)), \end{aligned}$$

so that $p_{ij}(t;\theta)$ expresses the sum of the probabilities of all possible ways in which we can decrease i individuals to j individuals. Furthermore, when $i \leq j$, we get that $p_{ij}(t;\theta)$ simply represents the convolution of the $Bin(i,\mathrm{e}^{-\mu t})$-density and the $Poi(\rho)$-density. One can easily show that for the marginal distributions of $\{N(t)\}_{t\geq 0}$ we have that $\mathbb{P}(N(t)=j|N(0)=0) = \mathrm{e}^{-\rho}\rho^j/j!$, i.e. $(N(t)|N(0)=0) \sim Poi\big(\frac{\alpha}{\mu}(1-\mathrm{e}^{-\mu t})\big)$, and that $(N(t)|N(0)=0) \xrightarrow{d} Poi(\alpha/\mu)$ as $t \to \infty$.

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 characterisation of $\{N(t)\}_{t\geq 0}$ which sometimes is useful to exploit is to consider $\{N(t)\}_{t\geq 0}$ as a Markov jump process (see Paper II).

Proposition 2. *Let $\theta = (\alpha,\mu) \in \Theta \subseteq \mathbb{R}_+^2$. $\{N(t)\}_{t\geq 0}$ is a Markov jump process with state space $E = \mathbb{N}$, jump intensity function*

$$\lambda(\theta;i) = \alpha\mathbf{1}\{i=0\} + \min\{\alpha, i\mu\}\mathbf{1}\{i>0\} \quad i \in E,$$

and transition kernel $\{r(\theta;i,j):i,j\in\mathbb{N}\}$*, where*

$$r(\theta;i,j)=\frac{1}{\alpha+\mu i}\left(\alpha\mathbf{1}\{j=i+1\}+\mu i\mathbf{1}\{j=i-1\}\right)\quad i,j\in E.$$

2.2 The RS-model

The process $\mathbb{X}(t)=\{[\mathbf{X}_i,m_i(t)]:i\in\Omega_t\}$ can be described as follows. As time elapses, the arrivals in time of new individuals to $W\subseteq\mathbb{R}^d$ and the time these individuals live in W are governed by an immigration-death process, $N(t)$, having parameter $\theta=(\alpha\nu(W),\mu)\in\Theta$, where $\nu(\cdot)$ denotes volume in $\mathbb{R}^d$ and $\Theta\subseteq\mathbb{R}_+^2$ is compact. We here denote the (Poisson) arrival process by $B(t)$ and the death process by $D(t)$ so that $N(t)=B(t)-D(t)$, where $N(0)=0$. Furthermore, upon arrival at time t_i^0, individual i is assigned a location $\mathbf{X}_i\sim Uni(W)$ (thus far, at each fixed time t this constitutes a spatial Poisson process with intensity $\frac{\alpha}{\mu}(1-\mathrm{e}^{-\mu t})$, restricted to W) together with an initial mark, $m_i(t_i^0)=m_i^0$, which is taken either as some fixed positive value (as will be the case here), or as a value drawn from some suitable distribution ([24] considers $m_i^0\sim Uni(0,\epsilon)$, $\epsilon>0$). When an individual's ($Exp(\mu)$-distributed) life time has expired we say that the individual has suffered a *natural death*.

Once individual i has arrived it starts growing deterministically according to

$$m_i(t) = m_i^0+\int_{t_i^0}^{t}dm_i(s),\quad t_i^0\le t, \tag{2.1}$$

where

$$dm_i(t) = f(m_i(t);\psi)dt-\sum_{\substack{j\in\Omega_t\\ j\neq i}}h\left(m_i(t),m_j(t),\mathbf{X}_i,\mathbf{X}_j;\psi\right)dt.$$

Here $\Omega_t=\{i\in\{1,\dots,B(t)\}:\text{individual } i \text{ is alive at time } t\}$, ψ is a parameter vector, the function $f\left(m_i(t);\psi\right)$ determines the individual growth of mark i in absence of competition with other (neighbouring) individuals and $h\left(m_i(t),m_j(t),\mathbf{X}_i,\mathbf{X}_j;\psi\right)$ is a function handling the individual's spatial interaction with other individuals.

In addition to the natural death, an individual can die *competitively* which we consider to happen as soon as $m_i(t)\le 0$.

The literature offers a wide range of possible choices for the individual growth function, $f(m_i(t);\psi)$ (see e.g. [24]). Two examples are the so-called linear growth function,

$$f(m_i(t);\psi) = \lambda\left(1 - \frac{m_i(t)}{K}\right),$$

and the logistic growth function,

$$f(m_i(t);\psi) = \lambda m_i(t)\left(1 - \frac{m_i(t)}{K}\right).$$

Here $\psi = (\lambda, K, c, r) \in \mathbb{R}_+^2 \times \mathbb{R} \times \mathbb{R}_+$ and the two parameters $\lambda > 0$ and $K > 0$ are, respectively, the growth rate of a mark and its upper bound (carrying capacity). These functions are both special cases of the Von Bertalanffy-Chapman-Richards growth function (see [22]) which has previously been used to model the development of the radii of isolated Scots pines [17]. Since the shape of the logistic growth function resembles the shape of the Von Bertalanffy-Chapman-Richards growth function fitted in [17] we consider it both a good and a tractable candidate for our forestry purposes (see e.g. [22, 24]).

Just as for the individual growth function the possible choices of spatial interaction functions are many (c.f. [15, 22, 24] for examples of interaction functions and related discussions). One example is given by (see [24])

$$h\left(m_i(t), m_j(t), \mathbf{X}_i, \mathbf{X}_j;\psi\right) = c\mathbf{1}\left\{B\left[\mathbf{X}_i, rm_i(t)\right] \cap B\left[\mathbf{X}_j, rm_j(t)\right] \neq \emptyset\right\},$$

where $c \in \mathbb{R}$ is the force of interaction and $r > 0$ is the scale of interaction. Furthermore, $B\left[\mathbf{X}_i, rm_i(t)\right]$ denotes a closed ball with centre $\mathbf{X}_i$ and radius $rm_i(t)$ and it is referred to as the 'influence zone' of individual i. Since competition for resources takes place only within influence zones ([2, 29]), individuals i and j will compete only when their influence zones intersect, i.e. when $B\left[\mathbf{X}_i, rm_i(t)\right] \cap B\left[\mathbf{X}_j, rm_j(t)\right] \neq \emptyset$. This symmetric interaction function has the effect that small individuals have the same impact on large (neighbouring) individuals as the large individuals have on small individuals. Unless our forest stand consists of individuals of similar size this interaction function becomes unrealistic. In order to circumvent this problem we here consider instead the so called area interaction function, given by

$$h\left(m_i(t), m_j(t), \mathbf{X}_i, \mathbf{X}_j;\psi\right) = c\,\frac{\nu\left(B\left[\mathbf{X}_i, rm_i(t)\right] \cap B\left[\mathbf{X}_j, rm_j(t)\right]\right)}{\nu\left(B\left[\mathbf{X}_i, rm_i(t)\right]\right)}, \qquad (2.2)$$

This non-symmetric soft core interaction has the effect that large marks influence small marks more than the other way around, yet allowing the small marks to play their part. This interaction model is more realistic in tree modelling applications than symmetric interaction models (see [22, 24]). Depending on the choice of parameters, this area interaction function has the ability to generate regular as well as aggregated point patterns (despite the underlying uniform distribution of the locations) [21]. Note that the parameter r determines how large the range of interaction is and c mainly determines how regular the point patterns are.

2.2.1 The natural death rate

As previously mentioned the so called natural deaths are governed by the death process, $D(t)$. In situations where it seems plausible that the natural deaths depend on an individual's size we may let the death rate be given by some function $\mu\gamma(\cdot)$, $\mu > 0$, where $\gamma(\cdot)$ is a function of the marks. This means that as time passes the $Exp(\mu\gamma(m_i(t)))$-distributed remaining lifetime of an individual will change with its size. An alternative way of expressing the behaviour of the death process is to say that the conditional probability that an individual i dies naturally during $(t, t + dt)$ given $m_i(t)$ equals $\mu\gamma(m_i(t))\,dt + o(dt)$. Note that if $\gamma(\cdot) \equiv 1$, we retrieve the ordinary immigration death process. In Paper I, we choose to evaluate the RS-model under $\gamma(m_i(t)) = 1/(1 + m_i(t))$ which implies that individuals become more viable as they grow; a choice motivated by our forestry applications. In Paper II, as well as in [21, 22, 23, 24] the model is chosen to have $\gamma(\cdot) \equiv 1$.

2.2.2 Simulation

For clarity, the simulation algorithm given in [24] is presented below, where $W \subseteq \mathbb{R}^2$ is rectangular. Let dt be small, e.g. $dt = 0.01$, fix the final time $T_n > 0$ and let $W = [a, b] \times [c, d]$. Decide on sample time points $0 < T_1 < \ldots < T_n$. Set $k = 0$.

- Cycle over times $t = dt, 2dt, \ldots, \lfloor T_n/dt \rfloor$
    - For all individuals $i \in \Omega_t$ and for $\{Z\}_{i=1,\ldots,|\Omega_t|}$, a sequence of iid $Uni(0, 1)$ random numbers:
        1. `Natural death`: If $Z \leq \mu\gamma(m_i(t))$ remove i from Ω_t, set $m_i(t + dt) = 0$ and let i belong to the set A
        2. `New mark size`: If $i \in \Omega_t$ still, calculate $m_i(t+dt) = m_i(t)+dm_i(t)$ (i.e. calculate (2.1))
        3. `Competitive death`: If $i \in \Omega_t$ still and $m_i(t+dt) \leq 0$ remove i from Ω_t, set $m_i(t + dt) = 0$ and let i belong to the set B

- `Immigration`: Generate a $Poi(\alpha(b-a)(d-c)dt)$-random number N. If $N > 0$:
  Simulate pairs $(x_{k+1}, y_{k+1}), \ldots, (x_{k+N}, y_{k+N})$ where x_i is a random number from $Uni(a, b)$ and y_i is a random number from $Uni(c, d)$. Set $m_{k+1}(t) = m^0_{k+1}, \ldots, m_{k+N}(t) = m^0_{k+N}$ (which will either all be fixed as $m^0_i = \epsilon > 0$ or $Uni(0, \epsilon)$-random numbers).
  Let $\Omega_{t+dt} = \{k+1, \ldots, k+N\} \cup \Omega_t$.
  Update k to $k+N$.
- `Printing`: If $t = T_i$ for some $i = 1, \ldots, n$, print $X_t = \{(i, \mathbf{x}_i, m_i(t))\}_{i \in \Omega_t}$, $X^A_t = \{(i, \mathbf{x}_i, m_i(t))\}_{i \in A}$ and $X^B_t = \{(i, \mathbf{x}_i, m_i(t))\}_{i \in B}$

2.2.3 Remarks about the competitive death

As previously mentioned, one of the possible death occurrences present in the RS-model is the competitive death. Consider the infinitesimal-size interval $(t, t+dt)$ and recall that we classify an individual as having died from competition in $(t, t+dt)$ if $m_i(t) > 0$ and $m_i(t+dt) \leq 0$. Let us call this scenario 1. Consider now an alternative approach, which we call scenario 2, where the individual suffers a competitive death if $m_i(t) > 0$ and $dm_i(t) < 0$. Now a reasonable question emerges, namely, which of the two scenarios should be used to represent competitive/interactive death for tree data. In a tree stand model one could argue that scenario 1 is a more appropriate view than scenario 2 since trees do not disappear immediately after they die. This thus indicates that they should not be removed as soon as $dm_i(t) < 0$, since dead trees occupy the ground where they have been standing some time after their deaths. Also, to some extent, dead trees inhibit the nutrient access and light absorption of other trees close to it. On the other hand it might not seem plausible that a tree, after dying, keeps fighting until it has reached size $m_i(t) = 0$. Furthermore, it is not reasonable that a new tree would end up very close to the centre of a recently deceased one, shortly after the death of the deceased tree. Although a bit artificial in its nature we thus have chosen to use of scenario 1 to represent competitive deaths, just as in [24].

It may also seem troublesome is that the uniform distribution of the locations, $\mathbf{X}_i \sim Uni(W)$, does not prohibit newcomers to end up "within" other individuals, i.e. $B[\mathbf{X}_i, m_i(t)] \subseteq B[\mathbf{X}_j, m_j(t)]$. This however, provided that $c > 0$ is not very small, only causes such newcomers' instantaneous death.

Chapter 3

Parameter estimation

Assume now that we sample the process at times $0 = T_0 < \ldots < T_n = T$. Then, for each $k = 1, \ldots, n$, this gives rise to a sampled marked point configuration $\mathbb{X}_{obs}(T_k) = \left\{[\mathbf{x}_i, m_i(T_k)] : i \in \Omega_{T_k}^{obs}\right\}$ (Figure 1.1 illustrates such a scenario). In this chapter we start by presenting the methods used to estimate the parameters in the RS-model. This includes the presentation of the least squares approach used to estimate the mark related parameters, ψ, and the general idea behind the edge correction methods proposed in Paper I. We then look at ML-estimation in the discretely sampled immigration-death process, these estimators' asymptotic properties, and how they can be applied to the RS-model.

3.1 Estimation of the RS-model parameters

The following least squares approach for estimating the mark related parameters, $\psi = (\lambda, K, c, r) \in \mathbb{R}_+^2 \times \mathbb{R} \times \mathbb{R}_+$, and method for the labelling of naturally dead individuals originally was suggested in [24]. Let $\tilde{\mathbb{X}}_{obs}(T_k) = \left\{\tilde{m}_i\left(T_{k+1}; \psi, \mathbb{X}_{obs}(T_k)\right) : i \in \Omega_{T_k}^{obs}\right\}$ denote the set of predictions of the actual data marks, $\left\{m_i(T_{k+1}) : i \in \Omega_{T_k}^{obs}\right\}$, generated by equation (2.1) under the regime of ψ, based on the configuration $\mathbb{X}_{obs}(T_k)$ (in practise we employ the simulation algorithm presented in [24] in order to create each predicted set $\tilde{\mathbb{X}}_{obs}(T_k)$ from each set $\mathbb{X}_{obs}(T_k)$). Once having produced $\tilde{\mathbb{X}}_{obs}(T_k)$, if the predicted mark indicates that the individual is alive but the the individual is dead in reality, this predicted individual will be treated as having died by natural causes in (T_k, T_{k+1}). The least squares estimates are then found by minimising

$$S(\psi) := \sum_{k=1}^{n-1} \sum_{i \in \Omega_{T_k}^{obs}} \mathbf{1}\{i \in \Omega_{T_{k+1}}^{obs}\} \left[\tilde{m}_i\left(T_{k+1}; \psi, \mathbb{X}_{obs}(T_k)\right) - m_i\left(T_{k+1}\right)\right]^2$$

with respect to $\psi = (\lambda, K, c, r) \in \mathbb{R}_+^2 \times \mathbb{R} \times \mathbb{R}_+$, where $\mathbf{1}\{i \in \Omega_{T_{k+1}}^{obs}\}$ is an indicator function being 1 if the actual data individual i is alive at time T_{k+1}.

In order to minimize $S(\psi)$ some optimization procedure is required. The approach used in [24] is to create a grid of parameter values for each of the parameters in $\psi = (\lambda, K, c, r)$ and then calculate $S(\psi)$ for all combinations of values taken from these grids. One then lets $\hat{\psi} = (\hat{\lambda}, \hat{K}, \hat{c}, \hat{r})$ be given by the combination of grid values which gives rise to the smallest value of $S(\psi)$ and either accepts $\hat{\psi}$ as one's final estimate or one creates a new, finer, grid centred around the estimated parameter values in $\hat{\psi}$ and repeats the procedure a number of times until no change in $\hat{\psi}$ takes place and the grids have all become very dense. This procedure encounters the problem that the actual optimal combination of parameters may fall outside the grids, as the grids are becoming finer, if the initial grid is not chosen correctly. Another approach which is similar in its nature to the grid search, still avoiding the aforementioned problem, is to repeatedly draw parameter values $\psi = (\lambda, K, c, r)$ where $\lambda \sim Uni(\lambda_L, \lambda_U)$, $K \sim Uni(K_L, K_U)$, $c \sim Uni(c_L, c_U)$, $r \sim Uni(r_L, r_U)$ and for each such combination calculate $S(\psi)$, choosing as final estimate the parameter combination giving rise to the smallest $S(\psi)$. This MCMC type of method, however, has the drawback that one needs to make a choice on the upper and lower bounds in the uniform distributions being drawn from. One could handle this by choosing initial intervals on which we sample while successively extending the intervals if candidates near the boundaries are the ones minimizing $S(\psi)$. Note that we do not have to bother too much about the lower bounds since most of the parameters are bounded below by 0.

Paper I adopts an MCMC-type method (see [20]) where we start by choosing initial parameter estimates, i.e. let $\lambda = \lambda_0 > 0$, $K = K_0 > 0$, $c = c_0 \in \mathbb{R}$ and $r = r_0 > 0$, for which we calculate $S(\psi) = S(\lambda, K, c, r)$. We also define the step sizes $\delta_\lambda > 0$, $\delta_K > 0$, $\delta_r > 0$, and $\delta_c > 0$. Now, in each round we

1. randomly choose one of the parameters λ, K, r, c;
2. for our parameter of choice, say λ, let $\lambda' = \lambda + Z$, for Z drawn from $Uni(-\delta_\lambda, \delta_\lambda)$;
3. calculate $S(\psi') = S(\lambda', K, r, c)$;
4. if $S(\psi') < S(\psi)$ let $\lambda = \lambda'$, otherwise let $\lambda = \lambda$;

5. return to step 1.

We continue to run the algorithm until either $S(\psi)$ is less than some predefined minimum value, say, $S_{min} = 10^{-5}$ or until we have not seen any decrease in $S(\psi)$ for a predefined number of consecutive runs, say, $N_{max} = 200$. We let our final estimates $\hat{\psi} = (\hat{\lambda}, \hat{K}, \hat{c}, \hat{r})$ be given by the last ψ obtained in the algorithm above. Note that we here utilize the information obtained in the previous step in order to stepwise get closer to the final estimate.

When minimizing $S(\psi)$, in the case of a simulated data set, it can be seen that $S(\psi)$ may not attain its minimum at the true parameter set but instead at some biased ψ. This 'incorrect' shape of $S(\psi)$ is mainly due to edge effects and dependence between certain parameters. This phenomenon is illustrated in Figure 3.1, a plot of $S(\psi)$ as a function of only λ and K where c and r are kept fixed at their actual values. It is clear from the graph that $S(\psi)$ is decreasing as λ moves away from its actual value 0.2.

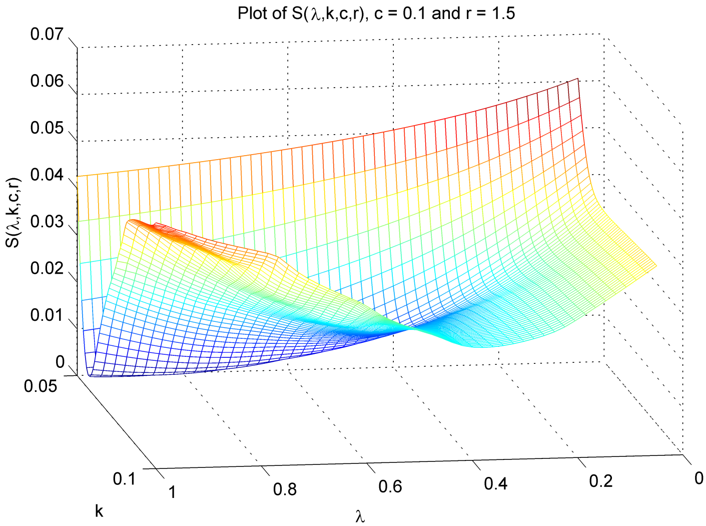

Figure 3.1: Plot of $S(\psi)$ as a function of only λ and K. c and r are kept fixed at their actual values, where $(\lambda, K, r, c) = (0.2, 0.1, 1.5, 0.1)$.

Note that, for instance, two different sets of c and r may result in similar interactions, due to the form of (2.2). In order to control the estimation routine, so that this risk of bias is reduced, the approach of Paper I is to find good starting values, $(\lambda_0, K_0, c_0, r_0)$, (as opposed to arbitrarily chosen ones) and to choose sensible step sizes, δ_λ, δ_K, δ_c, δ_r.

Paper I presents estimators for α and μ when $\gamma(m_i(t)) = 1/(1 + m_i(t))$. The estimator for α partially compensates for the unobserved individuals who arrive and die in the same sample interval, (T_k, T_{k-1}).

When sampling real data, $\{\mathbb{X}_{obs}(T_k)\}_{k=1}^n$, one usually considers all individuals within some region A (in Figure 1.1 circular) which is part of some larger region W. The individuals in A interact with each other but simultaneously also with the individuals present outside A, i.e. the individuals in $B = W \setminus A$. So, if one were to estimate some statistics and/or model parameters in a situation where the interaction among (neighbouring) individuals plays a role, by only taking into consideration the individuals in A the estimators may generate biased estimates since the interaction between the individuals in A and those in B would be neglected. The effects of the absence of the information regarding this interaction are commonly referred to as *edge effects*. The risk that the edge effects generate biases rapidly increases when one deals with small quantities of data in A, as is the case with our tree data set introduced in Figure 1.1. Hence, some type of correction method is needed (see e.g. [7, 13, 30]).

We here give the idea behind the edge correction methods proposed in Paper I. One starts by finding initial (possibly biased) estimates of the model parameters, $\hat{\Theta}_*$, based on our original data set (region A). Then, under the regime of $\hat{\Theta}_*$, we wish to find the expected model behaviour when restricted to region B (possibly conditioned on the actual data in A), $\mathbb{E}_{\hat{\Theta}_*}[\mathbb{X}_{[0,T]}|_B]$. By doing so we wish to establish the expected interaction between $\mathbb{E}_{\hat{\Theta}_*}[\mathbb{X}_{[0,T]}|_B]$ and the individuals in region A. With $\mathbb{E}_{\hat{\Theta}_*}[\mathbb{X}_{[0,T]}|_B]$ at hand we now re-estimate the model parameters from the actual data (region A), however, this time allowing for $\mathbb{E}_{\hat{\Theta}_*}[\mathbb{X}_{[0,T]}|_B]$ to interact with the actual data during the estimation. Once these new estimates have been obtained, we let them replace $\hat{\Theta}_*$ and repeat the above procedure again. By continuing in this fashion we have an iterative procedure which we stop once it has fulfilled a given predefined convergence criterion.

The three edge correction methods presented in Paper I are explained for the RS-model but they may be applied to other spatial and spatio-temporal (marked) point processes as well. In the algorithms presented in Paper I the large rectangular window W will be wrapped onto a torus when we generate the individuals in the outer region, B, (see e.g. [7, 19, 23, 30]).

3.2 Estimation in the immigration-death process

We will here look at estimation of (α, μ) when the immigration-death process, $\{N(t)\}_{t\geq 0}$, is considered as its own entity and as an application we see how this estimation can be applied to the RS-model. The results presented in this section can be found in Paper II.

Assume now that we sample $\{N(t)\}_{t\geq 0}$ as $N_1, \ldots, N_n$ at the respective times $0 = T_0 < T_1 < \ldots < T_n$. 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 $N(t)$ it can be factorised into a product of transition probabilities, i.e. $L_n(\theta) = \mathbb{P}(N(T_1) = N_1)\prod_{k=2}^{n} p_{N_{k-1}N_k}(t;\theta)$. By assumption we condition on $N(T_0) = 0$, so that the log-likelihood will be given by

$$l_n(\theta) = \sum_{k=1}^{n} \log p_{N_{k-1}N_k}(\Delta T_{k-1}; \theta), \tag{3.1}$$

where $\Delta T_{k-1} = T_k - T_{k-1}$. 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

$$l_n(\theta) = \sum_{i,j\in E} N_n(i,j) \log p_{ij}(t;\theta), \tag{3.2}$$

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 the expressions (3.1) and (3.2)) will be defined as

$$(\hat{\alpha}_n, \hat{\mu}_n) = \hat{\theta}_n = \arg\max_{\theta\in\Theta} l_n(\theta). \tag{3.3}$$

3.2.1 The ML-estimators

The ML-estimator for $\theta = (\alpha, \mu)$ is given by solving the system of equations

$$\begin{cases} \frac{\partial}{\partial\alpha} l_n(\theta) & = & \sum_{i,j\in E} N_n(i,j) \frac{\partial}{\partial\alpha} \log p_{ij}(t;\theta) = 0 \\ \frac{\partial}{\partial\mu} l_n(\theta) & = & \sum_{i,j\in E} N_n(i,j) \frac{\partial}{\partial\mu} \log p_{ij}(t;\theta) = 0. \end{cases}$$

As no closed form solution can be found by solving theses likelihood equations, numerical methods have to be employed in order to get ML-estimates. What is possible, however, is to express the estimator of α as a function of both the sample and the parameter μ, hence reducing the maximisation to a one dimensional problem.

Proposition 3. *The ML-estimator, $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\mu}_n)$, is found by maximising $l_n(\hat{\alpha}_n(\mu), \mu)$ over $\Theta_2 \subseteq \mathbb{R}_+$ (the projection of Θ onto the μ-axis), i.e.*

$$\begin{aligned}\hat{\mu}_n &= \arg\max_{\mu \in \Theta_2} l_n(\hat{\alpha}(\mu), \mu) \\ \hat{\alpha}_n &= \hat{\alpha}_n\left(\hat{\mu}_n\right),\end{aligned}$$

where

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

3.2.2 Asymptotic properties of the ML-estimators

Assume now that we sample $N(t)$ at the times $T_n = nt$, $n \in \mathbb{N}$, $t > 0$ (equidistant sampling). The following two results show that the ML-estimator (3.3) is strongly consistent (Proposition 4) and asymptotically Gaussian (Proposition 5). We denote by $\theta_0 = (\alpha_0, \mu_0) \in \Theta$ the true parameter pair of the immigration-death process. These results can be found in Paper II. For further discussions on ML-estimation in Markov processes and asymptotic properties thereof, see e.g. [1, 3, 6, 12, 27].

Proposition 4. *Let Θ be any compact subset of $\mathbb{R}_+^2$. Then the maximum likelihood estimator for the immigration-death process satisfies*

$$(\hat{\alpha}_n, \hat{\mu}_n) \xrightarrow{a.s.} (\alpha_0, \mu_0)$$

as $n \to \infty$, where $(\alpha_0, \mu_0) \in \Theta$ is the true parameter pair.

Proposition 5. *Let Θ be any compact subset of $\mathbb{R}_+^2$. Furthermore, assume that $(\log(\alpha_0 + \mu_0) - \log(\alpha_0))/\mu_0 \geq 2t$. Then, as $n \to \infty$, $\sqrt{n}\,((\hat{\alpha}_n, \hat{\mu}_n) - (\alpha_0, \mu_0))$ converges in distribution to the two-dimensional zero-mean Gaussian distribution with covariance matrix, $I(\theta_0)^{-1}$, given by*

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

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

3.2.3 Application to the RS-model

By the definitions of Ω_t and $N(t)$, the number of individuals alive at time t is given by

$$|\Omega_t| = N(t) - C(t) = B(t) - D(t) - C(t), \tag{3.4}$$

where $|A|$ denotes the cardinality of the set A and $C(t) \geq 0$ denotes the interactive death process, i.e. the process counting the total number of individuals who have suffered a competitive death in the time interval $(0,t]$. We will assume that $C(T_0) = 0$ so that $|\Omega_{T_0}| = 0$.

In the minimisation of $S(\psi)$, if $\tilde{m}_i\left(T_{k+1};\psi,\mathbb{X}(T_k)\right) \leq 0$ for an individual $i \in \Omega_{T_k}^{obs}$, it will be labelled as having died from competition in (T_k, T_{k+1}). We let the total number of such individuals be denoted by $(C(T_k) - C(T_{k-1}))_{obs}^{\psi}$ and we use it as an estimate of $C(T_k) - C(T_{k-1})$. Note that by expression (3.4) we can write $N(T_k) = N(T_{k-1}) + |\Omega_{T_k}| - |\Omega_{T_{k-1}}| + C(T_k) - C(T_{k-1})$ where $|\Omega_{T_1}| = C(T_0) = 0$. The observed version of this is given by

$$N_{obs}(T_k) = N_{obs}(T_{k-1}) + |\Omega_{T_k}^{obs}| - |\Omega_{T_{k-1}}^{obs}| + (C(T_k) - C(T_{k-1}))_{obs}^{\psi},$$

where $|\Omega_{T_1}^{obs}| = 0$.

When we here find the estimate $\hat{\theta} = (\hat{\alpha}\nu(W), \hat{\mu})$ with our new likelihood approach we use $(N_{obs}(T_1), \ldots, N_{obs}(T_n))$ as observation of the sampled immigration-death process, $(N(T_1), \ldots, N(T_n))$, and hence the log-likelihood is given by

$$l_n(\theta) = \sum_{k=1}^{n} \log p_{N_{obs}(T_{k-1})N_{obs}(T_k)}\left(T_k - T_{k-1}; \alpha\nu(W), \mu\right).$$

Chapter 4

Future work and extensions

Regarding possible extensions of Paper I, a thorough study of the RS-model's applicability in forestry should be made. Note further that the RS-model here is presented for a single species. However, it can easily be extended to include the scenario where interaction takes place also between different species, living and interacting within the same study region. This extension is made by letting each species be governed by both its unique individual growth function and mark interaction function which can be different within and between species. Hereby the amount an individual is affected by its neighbours depends both on the distance to the neighbours and their sizes and also on these neighbours' species.

The motivation for the work in Paper II comes from the need of improving the estimation of (α, μ) in the RS-model, compared to the estimators given in [24] and one should numerically study the possible improvement achieved. A further extension of the RS-model is given by adding a Brownian noise in the mark growth function of the RS-model, i.e. letting the marks be controlled by $dM_i(t) = dm_i(t) + dB_i(t)$ where the $B_i(t)$'s are independent Brownian motions, so that it incorporates uncertainties in the mark sizes. Having made this extension we hope to find a full likelihood structure for this multivariate diffusion type RS-model, where $L(\alpha, \mu)$ constitutes a part of the likelihood structure. A further improvement of Paper II that possibly can be made is to improve the invertibility condition given in Proposition 5 in Chapter 3 so that asymptotic normality holds for all $(\alpha_0, \mu_0) \in \Theta$. Furthermore, in order to become more realistic in applications, $N(t)$ could be extended by letting the arrival intensity, α, and the death rate, μ, be non-constant functions of time, or in themselves Markov chains (in the latter case $N(t)$ thus becomes a hidden Markov model) whereby, possibly, results similar to the ones found in Paper II can be established and the type of modelling done in Paper I can be developed.

Chapter 5

Summary of Papers

Paper I: Some edge correction methods for marked spatio-temporal point process models

In this paper we consider the RS-model where the death rate of the underlying immigration-death process depends on each individual's mark size, as opposed to the approach used in [24] where the death rate was constant.

We then discuss the estimation of the parameters when the process is sampled discretely in time. Since we let the death rate depend on the size of the individual, a new estimator is derived which takes the size changes of the individuals into consideration. Also a new estimator is suggested for the arrival intensity, which compensates for the unobserved arrivals and deaths of individuals arriving and dying between two consecutive sample time points.

To improve the estimation of the growth and interaction parameters, three edge correction methods for (marked) spatio-temporal point processes are proposed. They are all based on the idea of placing an approximated expected behaviour of the process at hand outside the study region. We then let these simulated realisations outside the study region interact with the data during the estimation. We estimate this expected behaviour by simulating realisations of the process, under a parameter choice based on some non-edge corrected initial estimates, and for each such realisation we generate new estimates which we average over to get our final estimates. By rerunning the whole procedure and using our edge corrected estimates to generate the surrounding realisations, we have created an iterative procedure which we stop once some given stopping criterion is fulfilled. Furthermore, we discuss three different approaches to run this type of edge correction and we present each of them in the context of the RS-model. When we numerically evaluate our edge corrected estimation procedures for the RS-model we see that we manage to reduce the bias substantially, compared to when no edge correction is applied.

Paper II: Maximum likelihood estimation in a discretely observed immigration-death process

In this paper we consider the immigration-death process, $N(t)$, and specifically we treat the ML-estimation of the parameter pair governing it, $\theta = (\alpha, \mu) \in \Theta \subseteq \mathbb{R}_+^2$, when Θ is compact and $N(t)$ is sampled discretely in time; $0 = T_0 < T_1 < \ldots < T_n$, $N(T_0) = 0$.

In order to find the likelihood structure of this continuous time Markov chain we derive its transition probabilities, and further, we manage to reduce the likelihood maximisation from a two dimensional problem to a one dimensional problem, where we maximise the likelihood, $L(\alpha, \mu) = L(\hat{\alpha}_n(\mu), \mu)$, over the projection of Θ onto the μ-axis.

Furthermore, by considering $N(t)$ as a Markov jump process we have shown that, under an equidistant sampling scheme, $T_k = kt$, $t > 0$, $k = 1, \ldots, n$, the sequence of ML-estimators, $\hat{\theta}_n(N(T_1), \ldots N(T_n))$, is consistent and asymptotically Gaussian. The asymptotic normality requires the Fisher information matrix invertability condition $(\log(\alpha_0+\mu_0)-\log(\alpha_0))/\mu_0 \geq 2t$, where (α_0, μ_0) is the underlying parameter pair. These results are further corroborated through simulations. In the simulations we see that the estimates approach the actual parameters and also that the empirical distribution of the estimates show strong indications of Gaussianity, even when the invertability condition is not fulfilled. We discuss how the ML-estimator, $\hat{\theta}_n(N(T_1), \ldots N(T_n))$, could be applied to the RS-model when $N(t)$ controls the arrivals of new marked points, as well as their potential life-times. The motivation for this work comes from the need of improving the estimation of α and μ in the RS-model, compared to the estimators given in [24] and in Paper I.

Bibliography

[1] Basawa, I.V., Prakasa Rao, B.L.S. *Statistical Inference for Stochastic Processes.* Academic press, 1980.

[2] Berger, U., Hildenbrandt, H. A new approach to spatially explicit modelling of forest dynamics: spacing, ageing and neighbourhood competition of mangrove trees. *Ecological Modelling*, 132:287–302, 2000.

[3] P. Billingsley. *Statistical Inference for Markov Processes.* The University of Chicago Press, 1961.

[4] Cressie, N.A.C. *Statistics for Spatial Data.* Wiley Series in Probability and Statistics, revised edition, 1993.

[5] Daley, D.J., Vere-Jones, D. *An Introduction to the Theory of Point Processes.* Springer Series in Statistics, second edition, 2008.

[6] Dehay, D., Yao, J.-F. On likelihood estimation for discretely observed Markov jump processes. *Australian & New Zealand Journal of Statistics*, 49:93–107, 2007.

[7] P. Diggle. *Statistical Analysis of Spatial Point Patterns.* Oxford university press, second edition, 2001.

[8] Finkenstadt, B., Held, L., Isham, V. *Statistical Methods for Spatio-Temporal Systems.* Chapman & Hall/CRC, 2006.

[9] Goreaud, F., Loreau, M., Millier, C. Spatial structure and the survival of an inferior competitor: a theoretical model of neighborhood competition in plants. *Ecological Modelling*, 158:1–19, 2002.

[10] Grabarnik, P., Särkkä, A. Modelling the spatial structure of forest stands by multivariate point processes with hierarchical interactions. *Ecological Modelling*, 220:1232–1240, 2009.

[11] Grimmett, G., Stirzaker, D. *Probability and Random Processes.* Oxford university press, third edition, 2001.

[12] P. Guttorp. *Statistical Inference for Branching Processes.* Wiley Series in Probability and Mathematical Statistics, 1991.

[13] Illian J., Penttinen A., Stoyan H., Stoyan D. *Statistical Analysis and Modelling of Spatial Point Patterns.* Wiley-Interscience, 2008.

[14] Möller, J., Waagepetersen, R.P. *Statistical Inference and Simulation for Spatial Point Processes.* Chapman & Hall/CRC Press, 2004.

[15] Nord-Larsen, T. Modeling individual-tree growth from data with highly irregular measurement intervals. *Forest Science*, 52:198–208, 2006.

[16] J.R. Norris. *Markov Chains.* Cambridge series in statistical and probabilistic mathematics, 1997.

[17] Prévosto, B., Curt, T., Gueugnot, J., Coquillard, P. Modeling mid-elevation Scots pine growth on a volcanic substrate. *Forest Ecology and Management*, 131:223–237, 2000.

[18] S.L. Rathbun. Asymptotic properties of the maximum likelihood estimator for spatio-temporal point processes. *Journal of Statistical Planning and Inference*, 51:55–74, 1996.

[19] E. Renshaw. The linear spatial-temporal interaction process and its relation to $1/\omega$-noise. *J. Roy. Statist. Soc. B*, 56:75–91, 1994.

[20] E. Renshaw. Applying the saddlepoint approximation to bivariate stochastic processes. *Mathematical Biosciences*, 168:57–75, 2000.

[21] Renshaw, E., Comas, C. Space-time generation of high intensity patterns using growth-interaction processes. *Statistics and Computing*, 19:423–437, 2009.

[22] Renshaw, E., Comas, C., Mateu, J. Analysis of forest thinning strategies through the development of space-time growth-interaction simulation models. *Stochastic Environmental Research and Risk Assessment*, 23:275–288, 2009.

[23] Renshaw, E., Särkkä, A. Gibbs point processes for studying the development of spatial-temporal stochastic processes. *Computational Statistics & Data Analysis*, 36:85–105, 2001.

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

[25] Stoyan, D., Kendall, W., Mecke, J. *Stochastic Geometry and its Applications.* John Wiley & sons, second edition, 1995.

[26] Stoyan, D., Penttinen, A. Recent applications of point process methods in forestry statistics. *Statistical Science*, 15:61–78, 2000.

[27] van der Vaart, A.W. *Asymptotic Statistics.* Cambridge series in statistical and probabilistic mathematics, 1998.

[28] Verre-Jones, D. Some models and procedures for space-time point processes. *Environmental and Ecological Statistics*, 16:173–195, 2009.

[29] Weiner, J., Damgaard, C. Size-asymmetric competition and size-asymmetric growth in a spatially explicit zone-of-influence model of plant competition. *Ecological Research*, 21:707–712, 2006.

[30] Williams, M.S., Williams, M.T., Mowrer, H.T. A boundary reconstruction method for circular fixed-area plots in environmental survey. *Journal of agricultural, biological, and environmental statistics*, 6:479–494, 2001.

Paper I

Some edge correction methods for marked spatio-temporal point process models

Ottmar Cronie[1]

[1]Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, 412 96 Göteborg, Sweden
E-mail address: ottmar@chalmers.se

Abstract:

We propose three edge correction methods for (marked) spatio-temporal point processes. They are all based on the idea of placing an approximated expected behaviour of the process at hand (simulated realizations) outside the study region which interacts with the data during the estimation. These methods are applied to the Renshaw-Särkkä growth-interaction model (RS-model) presented in [16]. The specific choices of growth function and interaction function made are purely motivated by the forestry application considered here. A new estimator has been derived for the death rate (since the distribution of the life-time of an individual is allowed to depend on its current size) and, furthermore, we propose a new estimator for the (Poisson process) arrival intensity which compensates for the (unobserved) individuals arriving and dying between two sample time points without having been observed. The parameters related to the development of the marks are estimated using the same least-squares approach as proposed in [16]. Finally, the edge corrected estimation methods, in the context of fitting the RS-model, are applied to a data set of Swedish Scots pines.

Key words: Edge correction, Spatio-temporal marked point process, Least squares estimation, Maximum likelihood estimation

1 Introduction

Many of the spatial point structures, with appurtenant marks, which are encountered in nature and in our surrounding environments, are in fact results of evolutionary processes which have been developing over time. One example of such a process is a forest stand which, from once being an empty piece of land, grows and changes over time to become the full stand observed at a later time point. Often these marked spatial structures are measured only at one specific time point, thus containing no information regarding the temporal aspects of the evolutionary process responsible for the generation of the data. Hence, in situations such as these, tree stands and other marked patterns are treated as realizations of marked point processes (see e.g. [18] and [6]).

However, if one wants a more thorough understanding of the development process and its inherent interaction mechanisms one cannot ignore the collective development of the locations and the marks (sizes) through time. This new scenario makes us to take on a somewhat different approach where one treats recorded time series of marked patterns as outcomes of the development of spatio-temporal marked point processes. This second approach has been less studied, however. As the aspect of time enters the model the level of complexity quickly increases and formulating involved models, which try to cover every aspect of the development, usually has the drawback of creating decrease in tractability, applicability and interpretability (see e.g. [5]). It is therefore necessary to formulate models which are tractable and easily interpreted but yet manage to cover the relevant aspects of spatio-temporal modelling. One such model is what here will be referred to as the Renshaw-Särkkä growth-interaction model (RS-model) which has been studied in a series of papers, most recently in [15], [13], [16] and [14].

When measurements are made in some bounded study region, the structure of the spatial dependences and interactions existing between individuals outside and inside the study region remains unobserved. This phenomenon, which in particular concerns those individuals inside the study region who are close to its boundary, is generally referred to as edge effects. In the context of estimation, if the study region contains a large number of individuals the edge effects may not have a large impact on the estimates. However, it may be the case that we deal with a small study region which contains only a small amount of data, which often is the case with tree data. In such cases there is a substantial risk that the edge effects generate quite severe biases and we therefore need some type of edge correction method when estimating the model parameters and summary statistics of interest. In the case of non-temporal analyses a number of methods for edge correction have been devised (see e.g. [7]) but these are not so easily generalized to the spatio-temporal setting. Hence, our main objective here is to develop methods which correct for these edge effects in the spatio-temporal setting.

We consider three edge correction methods which all, more or less, are based on the same idea. Initially one makes a first estimation (without edge correction) of the parameters of interest, Θ, thereby generating a set of biased parameter estimates, $\hat{\Theta}_*$. Once these estimates have been found one re-estimates the parameters, although, this time placing the "expected behaviour" of our spatio-temporal process, under the regime of $\hat{\Theta}_*$, in a buffer zone which surrounds the study region. During the re-estimation the individuals in this buffer zone have the purpose of interacting with the individuals (trees) at the boundary of the observation window, hence affecting the new estimates. These new edge corrected estimates will now replace $\hat{\Theta}_*$ and will then in turn be used to generate a new expected behaviour of the process. By letting this new expected behaviour take the place of the previous one we re-run the whole procedure, hence producing a new set of estimates. We iteratively continue in this fashion until we see convergence in the estimates. Now the question still remains regarding what is meant by and how to find this so called "expected behaviour" of the spatio-temporal process. The three edge corrections presented in this paper are basically three ways of estimating this expected behaviour and they are all based on successive simulations of an interacting process living outside our study region.

All three edge corrections presented in this paper will be applied to a slightly modified version of the RS-model (see [16]). The model considered here differs from its predecessor in that it allows the (exponential) distribution of each individual's life-time to vary with its size. This slight change of the process has had the consequence that a new maximum likelihood (ML) estimator for the death rate parameter has been derived, which takes into account that the size of an individual influences its viability. Furthermore, a new ML-estimator has been derived for the arrival intensity of the immigration process (Poisson process) governing the arrivals in time of new individuals. This new arrival intensity estimator tries to compensate for the unobserved births and deaths occurring between the time points at which the process is sampled. The parameters related to the growth of the marks and the interaction between the marks, just as in [16], will be estimated separately from the arrival intensity and the death rate. [16] presents an approach where these mark related parameters are estimated using the least-squares method and we here choose to follow the exact same approach.

The paper is set up as follows. In Section 2 we will present the slightly modified version of the RS-model, in which the distributions of the life-times are allowed to vary with the sizes of the individuals. The least squares approach used in the estimation will be presented in Section 3 together with the new death rate estimator and the new arrival intensity estimator mentioned above. Further, in Section 4, we present the data set of Swedish Scots pines considered. In section 5 we describe in detail the three previously mentioned edge correction methods developed for spatio-temporal point processes (with interacting marks). In Section 5 we will also present the results obtained in the evaluation of the methods and once these methods have been presented and evaluated (in the context of the RS-model), they are applied to our Scots pine data set.

2 The model

The spatio-temporal growth-interaction model has recently been studied by Renshaw and Särkkä in [15] and [16] and by Renshaw et al. in [14]. We here investigate the model given in [16], with the modification that the distribution of an individual's lifetime is allowed to depend on its size. The process is defined as follows.

The base of the process can be described as an immigration-death process where the immigration part governs arrivals of new individuals to a region of interest, $W \subseteq \mathbb{R}^2$, and a death part handling the number of 'natural deaths' occurring. Additionally, upon arrival, individuals are assigned locations and appurtenant marks (sizes) which change deterministically over time.

More precisely, individuals enter W randomly in time according to a homogeneous Poisson process with intensity $\alpha\nu(W)$, $\alpha > 0$, where $\nu(W)$ denotes the area of W. As individual i arrives at time t_i^0 it is assigned a location $\mathbf{x}_i \sim Uni(W)$. Together with its location each individual is also given an initial mark (size) $m_i(t_i^0) = m_i^0$ which can be taken as some fixed positive value (suitable when individuals are not observed until they have reached a certain size). Alternatively, one could draw m_i^0 from some distribution, for example the $Uni(0, \epsilon)$-distribution, $\epsilon > 0$, as in [16]. Note that at this stage, at each fixed t, the point process generated by the $\mathbf{x}_i$'s corresponds to a homogeneous spatial Poisson process with intensity αt, observed on W.

Once an individual arrives at W it instantly starts changing its size deterministically according to $m_i(t) = m_i^0 + \int_{t_i^0}^{t} dm_i(s)$, $t \geq t_i^0$, where

$$dm_i(t) \quad = \quad f\left(m_i(t);\theta\right)dt + \sum_{\substack{j\in\Omega_t \\ j\neq i}} h\left(m_i(t), m_j(t), \mathbf{x}_i, \mathbf{x}_j;\theta\right)dt. \qquad (1)$$

Here Ω_t is the index set comprising the individuals alive at time t, $f\left(m_i(t);\theta\right)$ is a function determining the individual growth of mark i in absence of competition with other (neighbouring) individuals and $h\left(m_i(t), m_j(t), \mathbf{x}_i, \mathbf{x}_j;\theta\right)$ is a function handling the individual's spatial interaction with other individuals. Note that it may happen that $m_i(t) \leq 0$ and once this happens we consider an individual to have died 'competitively', just as in [16].

As previously mentioned the so called natural deaths are governed by the death process which is defined as a simple death process having intensity function $\mu\rho(\cdot)$, $\mu > 0$, where $\rho(\cdot)$ is a function of the marks. This means that as time passes an individual's $Exp(\mu\rho(m_i(t)))$ -distributed remaining lifetime will change with its size. An alternative way of expressing the behaviour of the death process is to say that the conditional probability that an individual i dies naturally during $(t, t{+}dt)$ given $m_i(t)$ equals $\mu\rho(m_i(t))\,dt{+}o(dt)$. While [16] uses $\rho(m_i(t)) \equiv 1$ we here consider $\rho(m_i(t)) = 1/(1 + m_i(t))$, implying that individuals become more viable as they grow; a choice motivated by our forestry applications. If, on the contrary, one wishes to consider individuals who become less viable as they grow in size then, for instance, $\rho(m_i(t)) = m_i(t)/(1 + m_i(t))$ would be a better candidate.

The Von Bertalanffy-Chapman-Richards growth function has previously been used to model the development of the radii of isolated Scots pines [10]. This growth function has as special case the logistic growth function [14] and its shape resembles the shape of the Von Bertalanffy-Chapman-Richards growth function fitted in [10]. We therefore consider the logistic growth function, given by

$$f\left(m_i(t); \theta\right) = \lambda\, m_i(t) \left(1 - \frac{m_i(t)}{K}\right), \tag{2}$$

both a good and a tractable candidate for our purposes (see e.g. [14] and [16]). Expression (2) contains the two parameters $\lambda > 0$ and $K > 0$ which, respectively, denote the growth rate of a mark and its upper bound (carrying capacity). If we consider an individual in absence of interacting neighbouring individuals then (1) together with (2) gives rise to the ordinary differential equation $dm_i(t)/dt = \lambda\, m_i(t)\,(1 - m_i(t)/K)$ for which the solution is given by

$$m_i(t) = \frac{K}{1 + \left(K/m_i^0 - 1\right) \mathrm{e}^{-\lambda t}}. \tag{3}$$

Note that (3) (and thereby (2)) requires that $m_i^0 > 0$.

Just as for the individual growth function the possible choices of spatial interaction functions are many (c.f. [8], [14] and [16] for examples of interaction functions and related discussions). Here, we consider the so called area interaction function, given by

$$h\left(m_i(t), m_j(t), \mathbf{x}_i, \mathbf{x}_j; c, r\right) = -c\, \frac{\nu\left(B\left[\mathbf{x}_i, rm_i(t)\right] \cap B\left[\mathbf{x}_j, rm_j(t)\right]\right)}{\nu\left(B\left[\mathbf{x}_i, rm_i(t)\right]\right)}, \tag{4}$$

where $c \in \mathbb{R}$ is the force of interaction and $r > 0$ is the scale of interaction. Furthermore, $B\left[\mathbf{x}_i, rm_i(t)\right]$ is a closed disk centred at $\mathbf{x}_i$ with radius $rm_i(t)$ and it is referred to as the 'influence zone' of the individual. Since competition for resources takes place only within influence zones ([3] and [20]), individuals i and j will compete only when their influence zones overlap, i.e. when $B[\mathbf{x}_i, rm_i(t)] \cap B[\mathbf{x}_j, rm_j(t)] \neq \emptyset$. This non-symmetric soft core interaction has the effect that large marks influence small marks more than the other way around, yet allowing the small marks to play their part. This interaction model is more realistic in tree modelling applications than symmetric interaction models ([14] and [16]). Depending on the choice of parameters, this area interaction function has the ability to generate regular as well as aggregated point patterns (despite the underlying uniform distribution of the locations) [13].

3 Estimation

An expression of the full likelihood function is not known for this model and although likelihood methods are generally highly desirable due to their asymptotic properties, under certain regularity conditions (see e.g. [2] and [19]), other more tractable estimation methods often generate estimates of similar quality. We here follow [16] by estimating $\theta = (\lambda, K, c, r)$ using the least squares approach. The death rate, μ, and the arrival intensity, α, are estimated separately by the ML-method.

Regarding the simulation of the process, [16] presents an algorithm where W is the unit square wrapped onto a torus and $\rho(m_i(t)) \equiv 1$, which is easily modified to suite any choice of $\rho(m_i(t))$ (in particular $\rho(m_i(t)) = 1/(1 + m_i(t))$) and any $W \subseteq \mathbb{R}^2$. When computing $m_i(t)$ it should be noted that one does not have to include all $j \in \Omega_t \setminus \{i\}$ in the sum in expression (1), rather only those within the maximal interaction range, i.e. $j \in \Omega_t \setminus \{i\}$ such that $||\mathbf{x}_i - \mathbf{x}_j|| \leq 2rK$.

Given that T_j and N_{T_j}, $j = 1, \ldots, n$, respectively, denote the jth sample time and the total number of individuals observed by time T_j, we let our data set be represented by
$\mathbb{X} = \{\mathbb{X}(T_j)\}_{j=1}^{n} = \left\{(\mathbf{x}_{ij}, m_{ij}, I_{ij}) : i = 1, \ldots, N_{T_j}\right\}_{j=1}^{n}$, where $\mathbf{x}_{ij} = \mathbf{x}_i(T_j)$, $m_{ij} = m_i(T_j)$ and $I_{ij} = I_i(T_j)$. The functions $I_i(\cdot)$, $i = 1, \ldots, N_{T_n}$, are indicator functions such that $I_i(t) = 1$ if individual i is alive at time t and $I_i(t) = 0$ if the individual is dead at time t. As before $\mathbf{x}_i(\cdot)$ and $m_i(\cdot)$ denote the location and the size of individual i, respectively. Note also that the index set comprising the individuals alive at time t can be written as $\Omega_t = \{i \in \{1, \ldots, N_t\} : I_i(t) = 1\}$.

3.1 Least squares estimation of λ, K, c, and r

Considering a set of parameters $\theta = (\lambda, K, c, r)$ and a configuration $\mathbb{X}(T_j)$, let $\tilde{m}_i(T_{j+1}; \theta, \mathbb{X}(T_j))$, $i \in \Omega_{T_{j+1}}$, denote the prediction of $m_{i(j+1)}$ from $\mathbb{X}(T_j)$, based on calculating equation (1). If an individual has $\tilde{m}_i(T_{j+1}; \theta, \mathbb{X}(T_j)) > 0$ while $I_{i(j+1)} = 0$ it will be treated as having died by natural causes during (T_j, T_{j+1}). Our least squares estimates are then found by minimizing

$$S(\theta) := \sum_{j=1}^{n-1} \sum_{i \in \Omega_{T_j}} I_{i(j+1)} \left[\tilde{m}_i(T_{j+1}; \theta, \mathbb{X}(T_j)) - m_{i(j+1)}\right]^2,$$

with respect to $\theta = (\lambda, K, c, r) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}_+$.

In order to minimize $S(\theta)$ some optimization procedure is required. We here adopt an MCMC-type method (see [12]) where we start by choosing initial parameter estimates, i.e. let $\lambda = \lambda_0 > 0$, $K = K_0 > 0$, $c = c_0 \in \mathbb{R}$ and $r = r_0 > 0$, for which we calculate $S(\theta) = S(\lambda, K, c, r)$. We also define the step sizes $\delta_\lambda > 0$, $\delta_K > 0$, $\delta_r > 0$, and $\delta_c > 0$. Now, in each round we

1. randomly choose one of the parameters λ, K, r, c;
2. for our parameter of choice, say λ, let $\lambda' = \lambda + Z$, for Z drawn from $Uni(-\delta_\lambda, \delta_\lambda)$;
3. calculate $S(\theta') = S(\lambda', K, r, c)$;
4. if $S(\theta') < S(\theta)$ let $\lambda = \lambda'$, otherwise let $\lambda = \lambda$;
5. return to step 1.

We continue to run the algorithm until either $S(\theta)$ is less than some predefined minimum value, say, $S_{min} = 10^{-5}$ or until we have not seen any decrease in $S(\theta)$ for a predefined number of consecutive runs, say, $N_{max} = 200$. We let our final estimates $\hat{\theta} = (\hat{\lambda}, \hat{K}, \hat{c}, \hat{r})$ be given by the last θ obtained in the algorithm above. Note that we here utilize the information obtained in the previous step in order to stepwise get closer to the final estimate.

When minimizing $S(\theta)$, in the case of a simulated data set, it can be seen that $S(\theta)$ may not attain its minimum at the true parameter set but instead at some biased θ. This 'incorrect' shape of $S(\theta)$ is mainly due to edge effects (discussed further in Section 5) and dependence between certain parameters. For instance, two different sets of c and r may result in similar interactions, due to the form of (4). In order to control the estimation routine, so that this risk of bias is reduced, our approach is to find good starting values, $(\lambda_0, K_0, c_0, r_0)$, (as opposed to arbitrarily chosen ones) and to choose sensible step sizes, δ_λ, δ_K, δ_c, δ_r. The exact forms and derivations of these are given in Appendix A.1 and A.2.

3.2 Estimation of μ

Let $f_{L_k}(t_k|\mu)$, $k = 1, \ldots, n_T$, denote the densities of the random lifetimes $L_1, \ldots, L_{n_T}$ (observed as $t_1, \ldots, t_{n_T}$) of the n_T individuals who have died from natural causes by time T (determined during the minimization of $S(\theta)$), given some natural death rate function $\mu\rho(m_i(t))$, and let $t^0_{i(L_1)}, \ldots, t^0_{i(L_{n_T})}$ denote the birth times of the individuals having these life times. Also, under the same natural death rate regime, let $S_1, \ldots, S_{m_T}$ denote the m_T random lifetimes of the individuals who are still alive at time T (observed as $s_1, \ldots, s_{m_T}$). Then the likelihood of the death rate, μ, is (approximately) given by

$$
\begin{aligned}
L(\mu) &= \prod_{k=1}^{n_T} f_{L_k}(t_k|\mu) \prod_{l=1}^{m_T} \mathbb{P}(S_l > s_l|\mu) \\
&= \prod_{k=1}^{n_T} \mu\rho\left(m_{i(L_k)}\left(t^0_{i(L_k)} + t_k\right)\right) \exp\left\{-\mu\rho\left(m_{i(L_k)}\left(t^0_{i(L_k)} + t_k\right)\right) t_k\right\} \\
&\quad \times \prod_{l=1}^{m_T} \mathrm{e}^{-\mu\rho(m_{i(S_l)}(T)) s_l},
\end{aligned}
$$

where $m_{i(L_k)}(t)$ denotes the observed mark, at time t, of the individual having life time L_k. Similarly $m_{i(S_l)}(T)$ denotes the observed mark size at time T, of the individual having lived time S_l at time T. By solving with respect to μ in $\mathrm{d}\log\left(L(\mu)\right)/\mathrm{d}\mu = 0$ we get the ML-estimator

$$\hat{\mu} = n_T \Bigg/ \left(\sum_{k=1}^{n_T} \rho\left(m_{i(L_k)}\left(t^0_{i(L_k)} + t_k\right)\right) t_k + \sum_{l=1}^{m_T} \rho(m_{i(S_l)}(T)) s_l\right) \quad (5)$$

In the case of $\rho(m_i(t)) \equiv 1$ this reduces to the estimator, $\hat{\mu}_0$, found in [16]. Since we sample the process only at $0 = T_0 < T_1 < \ldots < T_n = T$, neither the actual death times, $t^0_{i(L_k)} + t_k$, $k = 1, \ldots, n_T$, nor the sizes at these death times, $m_{i(L_k)}(t^0_{i(L_k)} + t_k)$, $k = 1, \ldots, n_T$, will be known. Recall that we label an individual as naturally dead once the predicted mark $\tilde{m}_i\left(T_{j+1}; \theta, m_i(T_j)\right) > 0$ while $I_{i(j+1)} = 0$, during the calculation of $S(\theta)$. Let $T_{j,i(L_k)}$ be the last sample time at which individual $i(L_k)$ was observed alive and let $\tilde{m}_{i(L_k)}(T_{j,i(L_k)})$ denote the prediction of its mark at $T_{j,i(L_k)}$. This censoring forces us to approximate (5) by

$$\begin{aligned}\hat{\mu}_1 = n_T \Bigg/ \Bigg(&\sum_{k=1}^{n_T} \rho\left(\tilde{m}_{i(L_k)}\left(T_{j,i(L_k)}\right)\right)\left(T_{j,i(L_k)} - t^0_{i(L_k)}\right) \\ &+ \sum_{l=1}^{m_T} \rho(m_{i(S_l)}(T))\left(T - t^0_{i(S_l)}\right)\Bigg).\end{aligned} \quad (6)$$

As pointed out earlier, the process is observed only at the sampled time points $0 = T_0 < T_1 < \ldots < T_n = T$ so that the actual birth times (and death times) of the individuals remain unknown. Conditioned on the number of individuals arriving during $(T_{j-1}, T_j]$ the arrival times of the individuals will be uniformly distributed on (T_{j-1}, T_j) (see e.g. [9]). Thus, when estimating μ, for each interval $(T_{j-1}, T_j]$, we simulate $\Delta N_{T_{j-1}} = N_{T_j} - N_{T_{j-1}}$ birth times having a $Uni(T_{j-1}, T_j)$ distribution, provided that $\Delta N_{T_{j-1}} > 0$, which in turn are assigned to all individuals being observed for the first time at T_j. The question regarding which arrival time to assign to which individual is solved by giving the first arrival time to the individual who is the largest at time T_j, the second arrival time to the individual which is the second largest at time T_j and so forth. This will have the consequence that the life times will be random. Hence, by repeating this procedure a suitable number of times, each time simulating new random birth times, we generate a set of estimates of μ which are used to estimate a standard error for $\hat{\mu}$.

3.3 Estimation of α

Let $B(t) \geq 0$ denote the actual number of immigrants by time t and let $N_{T_j} = \left|\cup_{j=1}^{n} \Omega_{T_j}\right|$, $j = 1, \ldots, n$, denote the number of individuals observed at sample times up to T_j. Concerning the estimation of α, the approach of [16] is to ignore all the unobserved individuals who arrive and die within the same time interval (T_j, T_{j+1}), resulting in the immigration-increments $\Delta B(T_{j-1}) = \Delta N_{T_{j-1}}$, $j = 1, \ldots, n$, where $\Delta B(T_{j-1}) = B(T_j) - B(T_{j-1})$ and $\Delta N_{T_{j-1}} = N_{T_j} - N_{T_{j-1}}$. Since $B(t)$ is a Poisson($\alpha\nu(W)$)-process, its (independent) increments are $Poi\,(\alpha(T_{j+1} - T_j)\nu(W))$-distributed. This being the scenario, an ML-estimator for α (see [16]) is provided by

$$\hat{\alpha}_0 = \frac{N_{T_n}}{T_n \nu(W)}. \tag{7}$$

This estimator is unbiased under the hypothesis that $N^{obs}(t) = B(t)$ since $\mathbb{E}[N_{T_n}/T_n\nu(W)] = \mathbb{E}[N_{T_n}]/T_n\nu(W) = \alpha T_n\nu(W)/T_n\nu(W) = \alpha$. This approach, however, underestimates α since we do not account for the individuals who arrive and die in the same sample interval, (T_k, T_{k-1}), (see [16]).

One possible way of partially compensating for this bias is to add to each increment of the observed process, $\Delta N_{T_{j-1}}$, the expected number of individuals suffering a natural death among the expected number of individuals arriving during (T_{j-1}, T_j). Since the expected number of arrivals during (T_{j-1}, T_j) is unknown it will be replaced by an estimate hereof, provided by expression (7). Regarding the expected number of natural deaths, provided by μ, it will be governed by $\hat{\mu}$, the estimate of μ found in the previous subsection. The estimator takes the form

$$\hat{\alpha} = \underbrace{\frac{N_{T_n}}{T_n\nu(W)}}_{=\hat{\alpha}_0} + \frac{1}{T_n\nu(W)} \sum_{j=1}^{n} \left\lfloor N_{T_n} \frac{\Delta T_{j-1}}{T_n} \left(1 - \mathrm{e}^{-\hat{\mu}\rho(m_i^0)\Delta T_{j-1}}\right) \right\rfloor,$$

where $\lfloor x \rfloor$ denotes the integer part of x and $\Delta T_{j-1} = T_j - T_{j-1}$. The derivation of the estimator as well as some characteristics of it and its relation to $\hat{\alpha}_0$ can be found in Appendix A.3.

4 Data

Before presenting the edge correction methods we will introduce the specific tree data set under consideration. The data set we consider consists of measurements of locations and diameters at breast height (dbh) in a west Swedish Scots pine stand[1]. Recordings have been made in the years $T_1 = 1985$, $T_2 = 1990$, and $T_3 = 1996$ and the approximate age of the stand in 1985 was 22 years, thereby setting $T_0 = 1963$. Note that only the time intervals in which births and deaths occur are known, leaving the actual birth and death times unknown. All measurements have been made on a circular region of radius 10 meters where trees having reached 0.01 m dbh are included in the data set. Figure 1 illustrates plots of the data set with scaled radii (factor 10), for improved visualization, together with the appurtenant radius histograms.

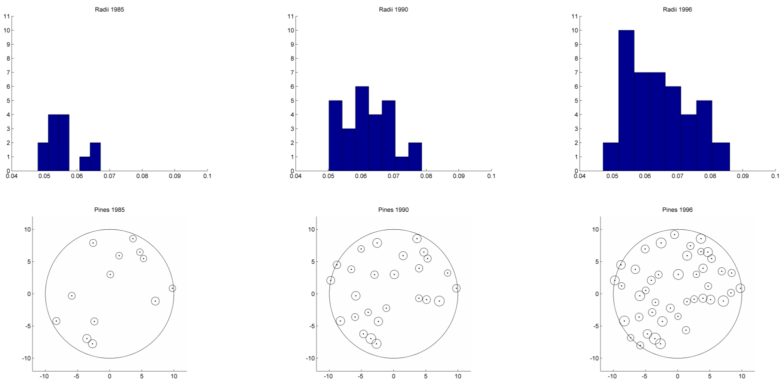

Figure 1: Swedish Scots pines recorded in 1985 (left), 1990 (middle) and 1996 (right). Upper row: Histograms of the radii. Lower row: Locations of the pines with scaled radii (factor 10).

Note how the size histogram tends to change as time elapses, with an increasing number of large trees. This is further confirmed by Table 1.

[1] Area number ("Trakt") 1562, Stand number ("Pålslag") 2060 - The "Lilla Edet" area.

T_j	1984	1990	1996
N_{T_j}	13	26	43
Mean radius	0.0557	0.0619	0.0640
Radius s.d.	0.0050	0.0074	0.0096
$\max_{i \in \Omega_{T_j}} m_{ij}$	0.0645	0.0775	0.0860

Table 1: Total number of trees, estimated mean, estimated standard deviation (s.d.) and maximum of the Scots pine radii at each sample time.

The RS-model has previously been fitted to data sets such as this [16]. However, since the number of trees present at each time point is fairly low it is important to take the edge effects into account, i.e. we have to somehow, for each sample time, estimate the behaviour of the unobserved trees surrounding our region of interest. Given this estimated information one can then correct the estimates such that the unobserved interaction between the region of interest and its surrounding area is compensated for.

5 Spatio-temporal edge correction

When sampling real data, $\mathbb{X}$, one usually considers all individuals within some region A (here circular) which is part of some larger region W. The individuals in A interact with each other but simultaneously also with the individuals present outside A, i.e. the individuals in $B = W \setminus A$. So, if one were to estimate some statistics and/or model parameters in a situation where the interaction among (neighbouring) individuals plays a role, by only taking into consideration the individuals in A the estimators may generate biased estimates since the interaction between the individuals in A and those in B would be neglected. The effects of the absence of the information regarding this interaction are commonly referred to as *edge effects*. The risk that the edge effects generate biases rapidly increases when one deals with small quantities of data in A, as is the case with our tree data set introduced in Section 4. Hence, some type of correction method is needed (see e.g. [4], [7] and [21]).

A simple edge correction method would be the so called minus sampling method (see e.g. [17]). First one finds all individuals who fall within a buffer zone, $C \subseteq A$, consisting of all points $\mathbf{x} \in A$ located less than some distance $d_0 > 0$ from the boundary of A. Then one carries out the estimation based only on the individuals in $A \setminus C$, yet taking into account the locations and marks of the individuals in C. In doing this we let the individuals in C and $A \setminus C$ affect each other, yet basing the computation of the statistic or the parameter estimate in question only on the individuals in $A \setminus C$. However, in situations where there is a limited amount of data in region A, as in our pine data set, removing data is not an option and this method therefore is not applicable.

A more sensible way of doing (spatio-temporal) edge correction in situations where there is little data available is to utilize the features of the parametric model which one attempts to fit to the data. We here give the idea behind the edge correction methods presented in this section. One starts by finding initial (possibly biased) estimates of the model parameters, $\hat{\Theta}_*$, based on our original data set (region A). Then, under the regime of $\hat{\Theta}_*$, we wish to find the expected model behaviour when restricted to region B (possibly conditioned on the actual data in A), $\mathbb{E}_{\hat{\Theta}_*}[\mathbb{X}_{[0,T]}|_B]$. By doing so we wish to establish the expected interaction between $\mathbb{E}_{\hat{\Theta}_*}[\mathbb{X}_{[0,T]}|_B]$ and the individuals in region A. With $\mathbb{E}_{\hat{\Theta}_*}[\mathbb{X}_{[0,T]}|_B]$ at hand we now re-estimate the model parameters from the actual data (region A), however, this time allowing for $\mathbb{E}_{\hat{\Theta}_*}[\mathbb{X}_{[0,T]}|_B]$ to interact with the actual data during the estimation. Once these new estimates have been obtained, we let them replace $\hat{\Theta}_*$ and repeat the above procedure again. By continuing in this fashion we have an iterative procedure which we stop once it has fulfilled a certain predefined convergence criterion.

The question still remains, however, regarding how to find the expected behaviour, $\mathbb{E}_{\hat{\Theta}_*}[\mathbb{X}_{[0,T]}|_B]$. We here suggest three methods based on the idea described above where $\mathbb{E}_{\hat{\Theta}_*}[\mathbb{X}_{[0,T]}|_B]$ is estimated from successive simulations of a (possibly interacting) version of our parametric model, restricted to region B. All three methods are similar to the ideas presented by Geyer in [1] in the sense that they all use simulated data as interacting data in region B. At each iteration step, at the sample times $T_1, \ldots, T_n$, all three methods sample a series of simulated process realisations which all live in region B. Thereafter each such sampled simulated outer realisation is combined with the actual data, $\mathbb{X}$, to form a full data set, $\mathbb{X}^*$, on $W = A \cup B$. From each new data set $\mathbb{X}^*$ we carry out our estimation procedure, however, as opposed to using the full data set $\mathbb{X}^*$ in the estimation we here only include $\mathbb{X}$ (region A) in the calculation of the estimates/statistics while we simultaneously let the (simulated) individuals in B interact with $\mathbb{X}$ (thereby influencing the estimates generated from $\mathbb{X}$). Now, in a given iteration step, by averaging over all estimates generated from each simulated outer realisation we get the final estimates for that specific iteration (this is how the simulated outer regions are considered to create an estimate of $\mathbb{E}_{\hat{\Theta}_*}[\mathbb{X}_{[0,T]}|_B]$ and its interaction with $\mathbb{X}$). This averaged set of estimates now replaces $\hat{\Theta}_*$ and by repeating the whole procedure once again we have executed the next iteration step.

A further question, yet to be explained in detail, is the stopping criterion used in the algorithms. Note that the estimates may be vector-valued. For each of the algorithms, given that we use N simulated outer realisations in each iteration, we will keep running it until the estimates, $\hat{\Theta}_*$, generated in two consecutive iterations differ by at most a distance $\epsilon > 0$. Once this has occurred we save these estimates and run the algorithm for another $M - 1$ iterations and average over the M estimates hereby generated, in order to get our final estimates. Another possible stopping criterion which may be used is the following. We run M iterations of our edge correction, hence generating a set of M estimates, $\Xi_1 = \{\hat{\Theta}_*^1, \ldots, \hat{\Theta}_*^M\}$, for which we estimate the variance, $\hat{\sigma}_1^2 = \widehat{Var(\Xi_1)}$, component wise. By running one more iteration of the edge correction, thus getting a new vector of estimates, $\hat{\Theta}_*^{M+1}$, we create the set $\Xi_2 = \{\hat{\Theta}_*^2, \ldots, \hat{\Theta}_*^M, \hat{\Theta}_*^{M+1}\}$ for which we estimate the variance, $\hat{\sigma}_2^2 = \widehat{Var(\Xi_2)}$. We continue in this fashion, i.e. creating $\Xi_{i+1} = (\Xi_i \setminus \{\hat{\Theta}_*^i\}) \cup \{\hat{\Theta}_*^{M+i}\}$, $i = 2, 3, \ldots$, to get $\hat{\sigma}_{i+1}^2 = \widehat{Var(\Xi_{i+1})}$, until $||\hat{\sigma}_{i+1}^2|| < \epsilon$ for some $\epsilon > 0$, where $||\cdot||$ is the Euclidean norm. Since the second approach considers the variation of a large number of estimates it is generally preferable to the first method. However, the first stopping criterion is less computationally demanding than the second one (since we have to wait M iterations before we can judge whether to stop or not in the second one) and it does a good enough job for the illustrative purposes we have here. Hence, in what follows we choose to apply the first of the two stopping criteria.

We will use the remainder of the section to present, discuss and evaluate the different methods.

5.1 Edge correction methods

The three edge correction methods we will present are explained for the RS-model but they may be applied to other spatial and spatio-temporal (marked) point processes as well. In the algorithms presented here the large rectangular window W will be wrapped onto a torus when we generate the individuals in the outer region, B, (see e.g. [4], [15], [11] or [21]). Recall that we sample the process as $\mathbb{X} = \{\mathbb{X}(T_j)\}_{j=1}^{n} = \left\{\left(\mathbf{x}_{ij}, m_{ij}, I_{ij}\right) : i = 1, \ldots, N_{T_j}\right\}_{j=1}^{n}$, where $\mathbf{x}_{ij} = \mathbf{x}_i(T_j)$, $m_{ij} = m_i(T_j)$ and $I_{ij} = I_i(T_j)$ is an indicator function such that $I_i(t) = 1$ if individual i is alive at time t and $I_i(t) = 0$ if the individual is dead at time t. Also recall that $\tilde{m}_i\left(T_{j+1}; \theta, \mathbb{X}(T_j)\right)$, $i \in \Omega_{T_{j+1}}$, denotes the prediction of $m_{i(j+1)}$ from $\mathbb{X}(T_j)$ generated by equation (1), under $\theta = (\lambda, K, c, r) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}_+$.

For a general process the three edge correction methods would have been presented in such a way that the whole parameter set would have been considered in each iteration. But in the case of the RS-model we may in fact omit the re-estimation of μ and α since their estimates tend not to change significantly between two iterations, despite the fact that we anew label individuals as naturally dead once $S(\theta)$ is evaluated for a new set of parameters θ, possibly leading to other life-times (of the naturally dead individuals) used in the estimator generating the new estimate $\hat{\mu}$ (hence also leading to a new estimate $\hat{\alpha}$). Note that below $||\hat{\theta}_* - \hat{\theta}||$ represents the Euclidean distance between $\hat{\theta}_*$ and $\hat{\theta}$.

5.1.1 Simple simulation of the outer region

We here present the first of the three methods; an algorithm which illustrates the basic idea on which all three methods are based.

1. Choose some small $\epsilon > 0$ and positive integers M and N.

2. Estimate the parameters from the data set $\mathbb{X}$ (region A) to generate a set of (non-edge-corrected) estimates $\hat{\theta}_* = \left(\hat{\lambda}_*, \hat{K}_*, \hat{c}_*, \hat{r}_*\right)$.

3. For $i = 1, \ldots, N$:

   (a) Simulate the process on $W = A \cup B$, based on $\hat{\theta}_*$ and $(\hat{\mu}, \hat{\alpha})$, and sample it at $T_1, \ldots, T_n$ (where W is wrapped onto a torus).

   (b) Create the data set $\mathbb{X}^*$ by removing what has been simulated in region A (for the sample times $T_1, \ldots, T_n$) and then replacing it with the data, $\mathbb{X}$.

   (c) Least squares estimation of $\theta = (\lambda, K, c, r)$ based on $\mathbb{X}^*$: Minimize

$$S(\theta) = \sum_{j=1}^{n-1} \sum_{i \in \Omega_{T_j} \cap \{k \in \mathbb{Z}_+ : \mathbf{x}_k \in A\}} I_{i(j+1)} \left[\tilde{m}_i\left(T_{j+1}; \theta, \mathbb{X}(T_j)\right) - m_{i(j+1)}\right]^2$$

   w.r.t. θ to get the estimates in this iteration, $\hat{\theta}_i = (\lambda_i, K_i, c_i, r_i)$. Note that we include only the individuals in $\mathbb{X}$ (region A) in the sum of squares $S(\theta)$. Also note that we must generate the predictions $\tilde{m}_i\left(T_{j+1}; \theta, \mathbb{X}(T_j)\right)$ for all the individuals in $\mathbb{X}^*$ (the individuals in B in A hereby interact) each time we evaluate $S(\theta)$ for a new θ.

4. Calculate $\hat{\theta} = \left(\frac{1}{N}\sum_{i=1}^{N} \lambda_i, \frac{1}{N}\sum_{i=1}^{N} K_i, \frac{1}{N}\sum_{i=1}^{N} c_i, \frac{1}{N}\sum_{i=1}^{N} r_i\right)$.

5. If $\left|\left|\hat{\theta}_* - \hat{\theta}\right|\right| < \epsilon$ set $\hat{\theta}^{(1)} = \hat{\theta}$ and go to step 6, otherwise go to step 3. Also set $\hat{\theta}_* = \hat{\theta}$.

6. For $j = 1, \ldots, M-1$:

   (a) Repeat steps 3 and 4 to generate the estimates $\hat{\theta}$ and set $\hat{\theta}_* = \hat{\theta}$.

   (b) Denote these estimates by $\hat{\theta}^{(j)} = \left(\lambda^{(j)}, K^{(j)}, c^{(j)}, r^{(j)}\right)$.

7. Let the final estimates be given by

$$\hat{\theta} = \left(\frac{1}{M} \sum_{i=1}^{M} \lambda^{(j)}, \frac{1}{M} \sum_{i=1}^{M} K^{(j)}, \frac{1}{M} \sum_{i=1}^{M} c^{(j)}, \frac{1}{M} \sum_{i=1}^{M} r^{(j)} \right).$$

Since the algorithm averages over all estimates $\hat{\theta}_1, \ldots, \hat{\theta}_N$ in a given iteration, it reduces the risk of having surrounding areas of too artificial nature generating the estimates. For instance, it is possible that some large individual(s) in B, close to the boundary of A, end up within the interaction range of some large individual(s) in A for a given simulated surrounding area. Such a scenario would not be encountered if the two individuals had been interacting naturally with each other throughout time. The algorithm above, through its averaging effect, reduces the strong impact which an extreme situation such as the aforementioned may have on some of the estimates.

5.1.2 Rotations of the outer region

We now consider a modifications of the previous algorithm which differs in the way it generates the surrounding realisations. Instead of simulating several outer realisations at each iteration, the idea here is that we instead use only one simulated outer region which we rotate a number of times, relative to the actual data, $\mathbb{X}$. By combining $\mathbb{X}$ with each rotation of the outer region we get a series of full data sets on W on which we base the estimation.

More specifically we replace step 2 and step 3 in the algorithm presented in Section 5.1.1 by

2*. Estimate the parameters from the data set $\mathbb{X}$ (region A) to generate a set of (non-edge-corrected) estimates $\hat{\theta}_* = \left(\hat{\lambda}_*, \hat{K}_*, \hat{c}_*, \hat{r}_*\right)$.

   Choose the angles $\omega_1 < \ldots < \omega_N$ either according to $\omega_{i+1} - \omega_i = 2\pi/N$ or $\omega_i \sim Uni(0, 2\pi)$. For all $i = 1, \ldots, N$, perform counterclockwise rotations (around the centre of A) of all locations, $\mathbf{x}_k = (x_k, y_k)$, in $\mathbb{X}$:

$$\mathbf{x}_k(\omega_i) = (x_k \cos(\omega_i) - y_k \sin(\omega_i), x_k \sin(\omega_i) + y_k \cos(\omega_i)).$$

We get the rotated data sets $\mathbb{X}_{\omega_1}, \ldots, \mathbb{X}_{\omega_N}$.

3*. Simulate the process on $W = A \cup B$, based on $\hat{\theta}_*$ and $(\hat{\mu}, \hat{\alpha})$, and sample it at $T_1, \ldots, T_n$ (where W is wrapped onto a torus).

For $i = 1, \ldots, N$:

1. Create the data set $\mathbb{X}^*_{\omega_i}$ by removing what has been simulated in region A (for the sample times $T_1, \ldots, T_n$) and then replacing it with the rotated data, $\mathbb{X}_{\omega_i}$.
2. Least squares estimation of $\theta = (\lambda, K, c, r)$ based on $\mathbb{X}^*_{\omega_i}$: Minimize

$$S(\theta) = \sum_{j=1}^{n-1} \sum_{i \in \Omega_{T_j} \cap \{k \in \mathbb{Z}_+ : \mathbf{x}_k(\omega_i) \in A\}} I_{i(j+1)} \left[\tilde{m}_i \left(T_{j+1}; \theta, \mathbb{X}(T_j) \right) - m_{i(j+1)} \right]^2$$

w.r.t. θ to get the estimates in this iteration, $\hat{\theta}_i = (\lambda_i, K_i, c_i, r_i)$. Note that we include only the individuals in $\mathbb{X}$ (region A) in the sum of squares $S(\theta)$. Also note that we must generate the predictions $\tilde{m}_i \left(T_{j+1}; \theta, \mathbb{X}(T_j) \right)$ for all the individuals in $\mathbb{X}^*_{\omega_i}$ (the individuals in B in A hereby interact) each time we evaluate $S(\theta)$ for a new θ.

As mentioned in step 2* one possibility is to use random angles. Although this adds an extra component of randomness to the procedure it has the drawback of allowing for situations where two or more of the angles become nearly the same, hence increasing the risk of the type of extreme estimates mentioned in Section 5.1.1. We therefore choose not to evaluate the version with random angles any further.

5.1.3 Outer region influenced by the growth of the data

Instead of rotating the surrounding area to avoid estimates based on the artificial surroundings described in Section 5.1.1 one may choose to condition on the development of the individuals in $\mathbb{X}$ (region A) when generating the surrounding individuals in region B. Our third edge correction method tries to overcome the problem of these artificial surroundings by letting the actual data individuals enter region A and directly start growing, alongside the simulation of the surrounding individuals which takes place in region B. During this growth the individuals in region A are allowed to influence the development of the individuals in region B but not the other way around. By doing this we try to mimic the actual underlying growth scenario.

Since the actual arrival times and the exact growth patterns remain unknown, an individual will enter A at an arrival time simulated uniformly over the sample time interval in which it was first observed (jumps of a Poisson process are uniformly distributed over time intervals) and then grow linearly between its observed sizes at the sample times so that it (possibly) affects the growth of the simulated surrounding individuals. The exact algorithm is given by replacing steps 2 and 3 in the algorithm of Section 5.1.1 by

2^{**}. Estimate the parameters from the data set $\mathbb{X}$ (region A) to generate a set of (non-edge-corrected) estimates $\hat{\theta}_* = \left(\hat{\lambda}_*, \hat{K}_*, \hat{c}_*, \hat{r}_*\right)$.

For each time interval $(T_{j-1}, T_j]$, $j = 1, \ldots, n$, we observe ΔN_{j-1} new individuals. Simulate $Uni(t_{j-1}, t_j)$-distributed birth times $b_1^j < \ldots < b_{\Delta N_{j-1}}^j$ and assign these to the individuals in $\mathbb{X}$ (region A) who have arrived in $(T_{j-1}, T_j]$ in such an order that the largest individual gets the smallest time, going upwards until the smallest individual has received the largest time.

3^{**}. For $i = 1, \ldots, N$:

(a) Simulate the process, based on $\hat{\theta}_*$ and $(\hat{\mu}, \hat{\alpha})$, but now only on the region $B = W \setminus A$ (where W is wrapped onto a torus) and sample it at $T_1, \ldots, T_n$. Furthermore, during the simulation, let each individual in $\mathbb{X}$ (region A) enter at its simulated birth time, b_k^j, $k = 1, \ldots, \Delta N_{j-1}$, $j = 1, \ldots, n$, and grow linearly between the sample time points until the last sample time point, T_j, it has been observed alive. This will have the consequence that these linearly growing individuals will have their actual (observed) sizes at the sample times. The effect acquired here is that the data, $\mathbb{X}$, will affect the growth of the simulated individuals in B (but not the other way around). Refer to this (partially) simulated data set as $\mathbb{X}^*$. Note that the only individuals in $\mathbb{X}^*$ located in A are the ones found in $\mathbb{X}$.

(b) Least squares estimation of $\theta = (\lambda, K, c, r)$ based on $\mathbb{X}^*$: Minimize

$$S(\theta) = \sum_{j=1}^{n-1} \sum_{i \in \Omega_{T_j} \cap \{k \in \mathbb{Z}_+ : \mathbf{x}_k \in A\}} I_{i(j+1)} \left[\tilde{m}_i\left(T_{j+1}; \theta, \mathbb{X}(T_j)\right) - m_{i(j+1)}\right]^2$$

w.r.t. θ to get the estimates in this iteration, $\hat{\theta}_i = (\lambda_i, K_i, c_i, r_i)$. Note that we include only the individuals in $\mathbb{X}$ (region A) in the sum of squares $S(\theta)$. Also note that we must generate the predictions $\tilde{m}_i\left(T_{j+1}; \theta, \mathbb{X}(T_j)\right)$ for all the individuals in $\mathbb{X}^*$ (the individuals in B in A hereby interact) each time we evaluate $S(\theta)$ for a new θ.

5.2 Evaluation of the estimation methods

In order to be able to evaluate the estimation methods previously presented we simulate what we here will refer to as a 'test set', consisting of a simulated realisation of the process on $W = [0, 30] \times [0, 30]$ (wrapped onto a torus), using step size $dt = 0.01$. We include those individuals alive at the sample times $T_1 = 22$, $T_2 = 27$, $T_3 = 33$ (the age of our pine stand at its sample time points) who are located within the circular region $A = \{\mathbf{y} \in W : ||\mathbf{y} - (15, 15)|| \leq 10\}$. The parameters used are $K = 0.1$, $\lambda = 0.08$, $c = 2$, $r = 2$, $\alpha = 0.007$, $\mu = 0.02$, and $m_i^0 = 0.05$. In order to check the accuracy of our estimation techniques we re-estimate the parameters generating the test set. We do not estimate m_i^0, but instead treat it as known since in forest stands one mostly knows the minimal tree radius from which measurements are being made (see Section A.1 for its estimation). The specific choice of parameters used to generate the test set was made since it generates realisations which resemble our tree data set. However, all methods we here apply have been evaluated for a range of different parameter values and the results obtained have been similar to those obtained for the test set.

If one gradually decreases the size of W in a series of edge corrections the distance on the torus between some of the individuals in B

gradually decreases. This is particularly the case for those individuals located close to and on opposite sides of the boundary of A (the individuals interacting the strongest with $\mathbb{X}$). In this gradual decrease these individuals start interacting more strongly with each other and thereby increasingly inhibit each other's growths, resulting in a gradual decrease in the edge correcting effect. Hence, a small W results in a slow convergence to the final estimates, whereas, a very large W makes the edge corrections computationally demanding. When we edge correct the re-estimation of the test set parameters we have chosen W to be a square region with side length 25, a choice purely based on trials.

Standard error estimates are obtained by re-running the edge correction procedure of choice a large number of times. However, in situations where this is computationally demanding, some resampling technique may be used to obtain the standard error estimates. For each edge correction method we have considered 10 different estimation runs where each of these uses the last $M = 4$ iterations, once $||\hat{\theta}_* - \hat{\theta}|| < \epsilon = 1$, in order to create the averaged final estimates and in each iteration we have considered $N = 3$ simulated surroundings ($N = 3$ angles in the case of the rotation-correction). Furthermore, on the basis of trials we have concluded that for each simulated surrounding it is sufficient to run the edge corrected estimation procedures until no change in $S(\theta)$ has been observed for $N_{max} = 50$ consecutive runs.

Table 2 presents both the initial estimates (see Appendix A.1), the final estimates found when applying no edge correction (stopping criterion for the minimisation of $S(\theta)$, $N_{max} = 3000$) and the estimates found for each of the edge corrections. The estimates of μ and α are given by $\hat{\mu} = 0.0113$ and $\hat{\alpha} = 0.00637$ ($(\hat{\mu} - \mu)/\mu = -43.5\%$ and $(\hat{\alpha} - \alpha)/\alpha = -9.0\%$).

	λ	K	c	r
True value	0.08	0.1	2	2
Initial	0.1006	0.0933	0.007	2.75
Bias	0.0206	-0.0067	-1.9930	0.75
Bias (%)	25.8%	-6.7%	-99.7%	37.5%
Uncorrected	0.0822	0.0995	5.4991	1.8301
Bias	0.0022	-0.0005	3.4991	-0.1699
Bias (%)	2.80%	-0.50%	174.96%	-8.50%
Simple				
Est. mean	0.0822	0.0996	2.7978	1.8694
Bias	0.0022	-0.0004	0.7978	-0.1306
Bias (%)	2.80%	-0.43%	39.89%	-6.53%
Est. s.e.	0.0001	0.0001	0.7755	0.1572
Rotations				
Est. mean	0.0821	0.0995	2.8364	1.7614
Bias	0.0021	-0.0005	0.8364	-0.2386
Bias (%)	2.58%	-0.46%	41.82%	-11.93%
Est. s.e.	0.0004	0.0001	0.5439	0.1416
Influenced growth				
Est. mean	0.0823	0.0996	2.7499	1.7926
Bias	0.0023	-0.0004	0.7499	-0.2074
Bias (%)	2.86%	-0.36%	37.50%	-10.37%
Est. s.e.	0.0005	0.0002	0.5422	0.1437

Table 2: Test set estimates: Initial estimates, non-edge corrected estimates ($N_{max} = 3000$) and estimates obtained through the different edge corrections. We have run each edge correction 10 times in order to get the estimated mean values and standard errors (s.e.). In each run we have used $N = 3$ (simulated surroundings/rotations), $N_{max} = 50$ (stopping criterion for the minimisation of $S(\theta)$), $\epsilon = 1$ (convergence criterion) and $M = 4$ (number of final iterations).

As one can see in the uncorrected estimation, the biases for the estimates of λ, K, r and α are fairly moderate. This, however, cannot be said about c and μ and regarding the under-estimation of μ there is little to be done. The large over-estimation obtained for c in the uncorrected estimation, however, is mainly a result of the edge effects which we correct for. Furthermore, we also see that the small biases of the estimates of

λ and K tend not to change significantly from the uncorrected estimates. The influenced growth correction manages to reduce the bias of $\hat{c}$ slightly more than the other two methods but this comes with a trade off in the form of an increased under-estimation of r, compared to both the uncorrected estimates and the simple correction estimates (the main reason being the strong dependence between c and r). A possible reason that the influenced growth generally performs the best in the estimation of c is that it actually takes into consideration the (approximate) behaviour of the actual data and it therefore restricts the previously mentioned artificial surroundings more than the other two methods. The simple correction is the only one of the three methods which reduces the r-bias but it is also the method giving the highest standard error estimates for c and r. As we see the rotation correction performs slightly worse than the other two methods but it has the advantage of reducing the computational time compared to the other two methods. By increasing the number of rotations one may be able to decrease the bias, but this comes with an increase in computation time. If no edge corrections are used, the points of the (data) point patterns likely will have less close neighbours than in reality. This will result in too large estimates of c, which in turn will result in more regular point patterns since c to a large extent controls the regularity of the point patterns generated by the process.

Table 3 gives us the results obtained after the first iteration. Note that the large bias generated by the uncorrected estimate of c directly is reduced by each of the methods.

Since our main concern is correcting the estimate of c, choosing $\epsilon = 1$ more or less implies that the final M iterations start once $|\hat{c}_* - \hat{c}| < \epsilon = 1$. By increasing ϵ a bit one may think that the final estimates get very different. However, since a substantial reduction takes place already after the first iteration and since we average over the final M iterations, if we were to choose ϵ a bit larger than 1 this in fact does not change the results drastically. Note further that one can start with a given ϵ and then increase it after a couple of iterations if the fluctuations between consecutive iterations are larger than initially believed (i.e. if $||\hat{\theta}_* - \hat{\theta}|| < \epsilon$ does not occur). The average number of iterations that were needed in order to reach $||\hat{\theta}_* - \hat{\theta}|| < \epsilon = 1$ in the 10 runs are 2.6 for the simple correction, 3.2 for the rotation correction and 2.8 for the influenced growth correction.

Iteration 1	λ	K	c	r
True value	0.08	0.1	2	2
Uncorrected	0.0822	0.0995	5.4991	1.8301
Bias	0.0022	-0.0005	3.4991	-0.1699
Bias (%)	2.80%	-0.50%	174.96%	-8.50%
Simple	0.0822	0.0995	3.3362	1.7991
Bias	0.0022	-0.0005	1.3362	-0.2009
Bias (%)	2.81%	-0.46%	66.81%	-10.05%
Rotations	0.0822	0.0995	3.1342	1.8770
Bias	0.0022	-0.0005	1.1342	-0.1230
Bias (%)	2.72%	-0.46%	56.71%	-6.15%
Influenced growth	0.0821	0.0996	3.6874	1.8697
Bias	0.0021	-0.0004	1.6874	-0.1303
Bias (%)	2.64%	-0.44%	84.37%	-6.54%

Table 3: Results obtained for the edge corrected estimation of the test set parameters after the first iteration. We have used $N = 3$ (simulated surroundings/rotations) and $N_{max} = 50$ (stopping criterion for the minimisation of $S(\theta)$).

5.3 Fitting the model to the Scots pines

In Section 4 we introduced our data set, a stand of Swedish Scots pines measured at three time points. In Table 4 we give the estimates found after having run the non-edge corrected estimation procedure, together with the results obtained in the three edge corrected estimation procedures. Just as for the test set, in the uncorrected estimation we have run the estimation until no change in $S(\theta)$ has been observed for $N_{max} = 3000$ consecutive runs whereas in all the corrected ones we have used $N_{max} = 50$. In the edge corrections we have chosen W to be a square region with side length 25 and for each edge correction method we have considered 10 different estimation runs where each of these uses the last $M = 4$ iterations to create its final estimates and in each iteration we have considered $N = 3$ simulated surroundings/rotations. However, here we have chosen the less restrictive value 2 for ϵ. In the uncorrected estimation we found $\hat{\alpha} = 0.004148$ and $\hat{\mu} = 0$ and these will be taken as final estimates for α and μ.

Note that, as expected, for all three methods, the edge corrected estimates are quite close to the uncorrected ones, except for c. The estimated values of c show that the point patterns are less regular than the uncorrected estimate suggests.

6 Discussion

We have recalled the Renshaw-Särkkä growth-interaction model (RS-model) – a spatio-temporal point process with interacting marks. The death rate of the underlying immigration-death process here depends on each individual's mark size, as opposed to the approach used in [16] where the death rate is constant.

We have then discussed the estimation of the parameters of the model when the process is sampled discretely in time. The parameters which control the marks' growth and interaction, λ, K, c, and r, are estimated using the same least-squares approach as proposed in [16]. Related to the least-squares estimation, we specify how we minimise the sum of squares numerically and discuss some issues related to that.

	λ	K	c	r
Initial	0.0350	0.0860	0.0195	8.0
Uncorrected	0.0790	0.0943	6.3314	3.7325
Simple				
Est. mean	0.0781	0.0949	3.1626	4.0680
Est. s.e.	0.0019	0.0017	1.0327	0.6351
Rotations				
Est. mean	0.0794	0.0944	3.1010	3.9396
Est. s.e.	0.0025	0.0015	0.7992	0.3802
Influenced growth				
Est. mean	0.0778	0.0954	3.5054	3.6151
Est. s.e.	0.0026	0.0016	0.7911	0.7229

Table 4: Parameter estimates found for the Scots pines: Initial estimates, non-edge corrected estimates ($N_{max} = 3000$) and estimates obtained through the different edge corrections. We have run each edge correction 10 times in order to get the estimated mean values and standard errors (s.e.). In each run we have used $N = 3$ (simulated surroundings/rotations), $N_{max} = 50$ (stopping criterion for the minimisation of $S(\theta)$), $\epsilon = 2$ (convergence criterion) and $M = 4$ (number of final iterations).

Parallel to this, a new estimator is derived which takes the size changes of the individuals into consideration. Also a new estimator is suggested for the arrival intensity, which compensates for the unobserved arrivals and deaths of individuals arriving and dying between two consecutive sample time points.

We finally propose three edge correction methods for (marked) spatio-temporal point processes which all are based on the idea of placing an approximated expected behaviour of the process at hand outside the study region. We estimate this expected behaviour by simulating realizations of the process, under a parameter choice based on some non-edge corrected initial estimates, and for each such realisation we generate new edge corrected estimates which we average over to get our edge corrected estimates.

We finally fit the RS-model to a data set of Swedish Scots pines. A thorough study of the RS-model's applicability in forestry will be made later. Regarding further developments, note that the RS-model here is presented for a single species. However, it can easily be extended to include the scenario where interaction takes place also between different species, living and interacting within the same study region. This extension is made by letting each species be governed by, on one hand, its own individual growth function and, on the other hand, its own mark interaction function. Hereby the amount an individual is affected by its neighbours depends, not only on the distance to the neighbours and the sizes of these neighbours, but also on the species of the neighbours. Another interesting extension would be to add a (Brownian) noise in the mark growth function of the RS-model, for example by letting the marks be governed by $dM_i(t) = dm_i(t) + dB_i(t)$, where the $B_i(t)$'s are independent Brownian motions, so that it incorporates uncertainties in the mark sizes. We then hope to find a full likelihood structure for this multivariate diffusion type RS-model.

7 Acknowledgements

The author wishes to thank Aila Särkkä (Chalmers University of Technology) for wonderful supervision, ideas and feedback. The author is also grateful for useful comments and suggestions (mainly related to the edge correction methods) from Claudia Redenbach (University of Ulm), Eric Renshaw (University of Strathclyde) and Gerald van den Boogaart (TU Bergakademie). Gratitude also goes out to Kenneth Nyström (Swedish University of Agricultural Sciences) for useful comments and discussions related to forestry. This research has been supported by the Swedish Research Council. Grants have been received from Chalmers vänner and Chalmerska forskningsfonden.

A Appendix

A.1 Initial estimates for λ, K, c and r

Since K represents the carrying capacity, an upper bound of the marks, it is sensible to use the largest observed mark value as starting value K_0.

Having found K_0 we can find an initial estimate of λ. Since the least interaction among individuals takes place at early time points, i.e. $dm_i(t) \approx f(m_i(t))dt$ for small t, by neglecting the interaction term in (1) one ends up with expression (3). By solving w.r.t. λ in (3), where K_0 replaces K and the largest observed individual at the first sample time point, m_{max}, replaces $m_i(t)$, we get as initial estimate of λ

$$\lambda_0 = -\frac{1}{T_1} \log \left(\frac{m_i^0 (1 - \frac{m_{max}}{K_0})}{m_{max}(1 - \frac{m_i^0}{K_0})} \right).$$

Recall that $m_i^0 > 0$ is the initial size which, if unknown, can be estimated by the smallest size of all individuals observed throughout all time points.

In the case of r and c, however, no obvious choices of initial values are present. What is possible, though, is to construct appropriate bounds for r, $r \in [r_l, r_u]$, which control the optimization and then choose the starting value for r to be, say, $r_0 = (r_u + r_l)/2$. Once this is done we choose our starting value for c to be

$$c_0 = \arg\max_{c \in \mathbb{R}} S\left(\lambda_0, K_0, r_0, c\right).$$

There is a natural lower bound for r, namely $r_l = 1$, since two trees cannot grow inside each other. To determine the upper bound, consider the mark correlation function of a stationary marked point process in $\mathbb{R}^2$ (see [7]), defined as

$$k(r) = \frac{\mathbb{E}_{or}[m_i m_j]}{\mu_m^2} \qquad \text{for } r > 0.$$

Here μ_m is the mean mark of the process and $\mathbb{E}_{or}[m_i m_j]$ denotes the conditional expectation of the mark-product of a pair of (marked) points of the process, given the existence of two such points distance r apart. It is a measure of dependence between the marks of two arbitrary points of the process a distance r apart. If, for some r, $k(r) = 1$ then the marks having inter-point distance r are uncorrelated whereas values of $k(r)$ smaller than 1 indicate inhibition (competition) at distance r and $k(r) > 1$ is a sign of mutual stimulation (points benefit from having inter-point distance r). Figure 2 illustrates idealized shapes of $k(r)$.

Denote by r^* the smallest value of $r > 0$ for which $k(r) = 1$. This is the shortest inter-point distance at which there are indications of uncorrelated marks. In the context of the RS-model, for a fixed time t, r^* indicates where the expected influence zone ends, i.e. $\mathbb{E}[rm_i(t)] \le r^*$. Consider now a time point at which the marked point pattern generated by the RS-model has stabilised, here taken as the last sample time point available, T_n. We get that

$$r \le r_u = r^*/\mathbb{E}[m_i(T_n)].$$

We estimate the mean mark at time T_n, $\mathbb{E}[m_i(T_n)]$, by $\bar{m}(T_n)$, the average size of the marks present at time T_n. In the case of our test set (see Section 5.2) the mark correlation plot at $T_3 = 33$ is given by Figure 2. The mean mark size for $T_3 = 33$ is given by $\bar{m}(T_3) = 0.0743$ and, as can be seen in Figure 2, $r^* \approx 1/3$, implying that $r_u = r^*/\bar{m}(T_3) \approx 4.5$ thus leading to $r_0 = 2.75$.

A.2 Choosing step-lengths

Another issue of importance here are the step-lengths δ_λ, δ_K, δ_c, δ_r. The simplest way of choosing δ_λ, δ_K, and δ_r is to choose $\delta_\lambda = \lambda_0$, $\delta_K = K_0$,

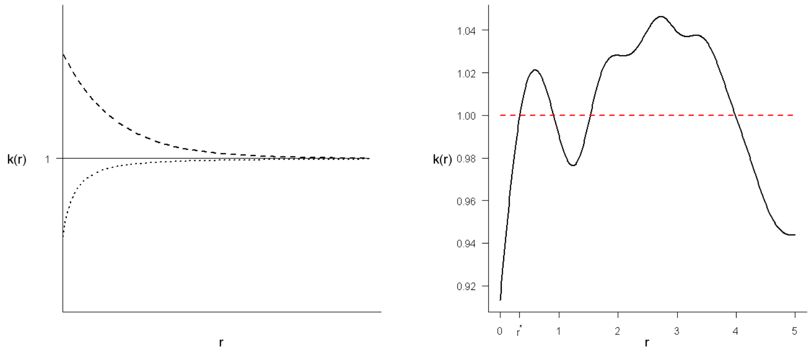

Figure 2: **Left**: Idealized shapes of different mark correlation functions. Mutual stimulation (dashed), uncorrelated marks (solid) and inhibition (dotted). **Right**: Mark correlation plot of our test set (see Section 5.2) at time $T_3 = 33$ ($\lambda = 0.08$, $K = 0.1$, $c = 2$, $r = 2$, $\alpha = 0.007$, $\mu = 0.02$) where $r^* = 1/3$.

and $\delta_r = r_0 - r_l$ since this way we allow for the estimates of these parameters to reach their minimum values. However, since $c \in \mathbb{R}$, choosing $\delta_c = c_0$ is not in any way self-evident. Choosing a too small δ_c would be more or less equivalent to keeping it fixed which certainly is not desirable. Although letting δ_c be too big may result in slower convergence, trials have shown that it does not affect the convergence of the estimation as much as keeping it too small. Since c and r do not have as natural choices of initial estimates as λ and K do and because of the strong dependence between them, new starting values for c and r can be found by starting the minimization, keeping $\lambda = \lambda_0$ and $K = K_0$ fixed, and then run the procedure a few times (say $N_{max} = 50$) with δ_c chosen big. This generates new estimates of c and r which in turn can be used as new starting values, c_0 and r_0, and we can then choose δ_c to be this new c_0, which we keep throughout the remaining estimation procedure (including the edge correction parts).

A.3 The estimator for α

When constructing our α-estimator we wish to somehow compensate for the unobserved individuals who arrive and die during the same interval (T_{j-1}, T_j), $j = 1, \ldots, n$.

For each $j = 1, \ldots, n$, let N_{T_j} be the number of individuals observed at sample times up until T_j, i.e. $N_{T_j} = \left| \cup_{i=1}^{j} \Omega_{T_i} \right|$, where Ω_t consists of the indices of the individuals alive at t and $|A|$ denotes the cardinality of a set A. Further, let $B(t) \geq 0$ denote the number of arrivals to W by time t. Instead of considering $\Delta B(T_{j-1}) = \Delta N_{T_{j-1}}$, where $\Delta B(T_{j-1}) = B(T_j) - B(T_{j-1})$ and $\Delta N_{T_{j-1}} = N_{T_j} - N_{T_{j-1}}$, and let our likelihood be based on these independent $Poi\left(\alpha(T_j - T_{j-1})\right)$-distributed increments, as was done in [16], we here consider

$$\begin{aligned} \Delta B(T_{j-1}) &= \Delta N_{T_{j-1}} \\ &+ \underbrace{\mathbb{E}\left[\sum_{k=1}^{\Delta B(T_{j-1})} \mathbf{1}\left\{\text{Individual } k \text{ dies in } (T_{j-1}, T_j)\right\}\right]}_{I}, \end{aligned} \tag{8}$$

where $\mathbf{1}\{\cdot\}$ is an indicator function. In other words, we add to the observed increments the expected number of individuals arriving and dying during (T_{j-1}, T_j).

Let $\eta_k^{\Delta T_{j-1}}$ denote the lifetime of individual $k \in \{1, \ldots, \Delta B(T_{j-1})\}$ in (8) and recall that m_k^0 is its (deterministic) initial size and $t_i^0 \sim Uni(T_{j-1}, T_j)$ its arrival-time (since the jumps of a Poisson process occurring in a given time interval are uniformly distributed on that interval [9]). By the lack of memory property of the exponential distribution and by Fubini's theorem the expectation in expression (8) can be written as

$$
\begin{aligned}
I &= \mathbb{E}\left[\sum_{k=1}^{\Delta B(T_{j-1})} \mathbf{1}\left\{T_{j-1} < t_k^0 + \eta_k^{\Delta T_{j-1}} < T_j\right\}\right] \qquad (9)\\
&= \mathbb{E}\left[\mathbb{E}\left[\sum_{k=1}^{\Delta B(T_{j-1})} \mathbf{1}\left\{T_{j-1} < t_k^0 + \eta_k^{\Delta T_{j-1}} < T_j\right\} \middle| \Delta B(T_{j-1})\right]\right]\\
&= \mathbb{E}\left[\sum_{k=1}^{\Delta B(T_{j-1})} \frac{1}{\Delta T_{j-1}} \int_{T_{j-1}}^{T_j} \mathbb{E}\left[\mathbf{1}\left\{T_{j-1} < x_k + \eta_k^{\Delta T_{j-1}} < T_j\right\}\right] dx_k\right]\\
&= \mathbb{E}\left[\sum_{k=1}^{\Delta B(T_{j-1})} \frac{1}{\Delta T_{j-1}} \int_{T_{j-1}}^{T_j} P\left(\eta_k^{\Delta T_{j-1}} < T_j - T_{j-1}\right) dx_k\right]\\
&\approx \mathbb{E}\left[\sum_{k=1}^{\Delta B(T_{j-1})} \left(1 - \mathrm{e}^{-\mu\rho(m_i^0)\Delta T_{j-1}}\right)\right]\\
&= \alpha\nu(W)\Delta T_{j-1}\left(1 - \mathrm{e}^{-\mu\rho(m_i^0)\Delta T_{j-1}}\right).
\end{aligned}
$$

Since the actual μ is unknown we will replace it by its estimate, $\hat{\mu}$, found in expression 6. Furthermore, expression (9) also contains α, the parameter we want to estimate. We deal with this by replacing α by an initial estimate, namely, $\hat{\alpha}_0 = N_{T_n}/(T_n\nu(W))$, given by (7).

In order for expression (8) to be treated as an actual Poisson process increment it needs to be integer valued, hence

$$
\Delta B(T_{j-1}) = \Delta N_{T_{j-1}} + \left\lfloor N_{T_n}\frac{\Delta T_{j-1}}{T_n}\left(1 - \mathrm{e}^{-\hat{\mu}\rho(m_i^0)\Delta T_{j-1}}\right)\right\rfloor, \qquad (10)
$$

where $\lfloor x \rfloor$ denotes the integer part of x. For convenience we will denote the right hand side of (10) by $H(\Delta T_{j-1}, \Delta N_{T_{j-1}}, \hat{\mu}, N_{T_n})$. We end up with the likelihood function

$$
\begin{aligned}
L(\alpha) &= \prod_{j=1}^{n} \mathbb{P}\left(\Delta B(T_{j-1}) = H(\Delta T_{j-1}, \Delta N_{T_{j-1}}, \hat{\mu}, N_{T_n})\right)\\
&= \prod_{j=1}^{n} \frac{\mathrm{e}^{-\alpha\nu(W)\Delta T_{j-1}}\left(\alpha\nu(W)\Delta T_{j-1}\right)^{H(\Delta T_{j-1}, \Delta N_{T_{j-1}}, \hat{\mu}, N_{T_n})}}{H(\Delta T_{j-1}, \Delta N_{T_{j-1}}, \hat{\mu}, N_{T_n})!}
\end{aligned}
$$

and by evaluating $\mathrm{d}\log(L(\alpha))/\mathrm{d}\alpha = 0$ we finally arrive at the estimator

$$\hat{\alpha} = \underbrace{\frac{N_{T_n}}{T_n \nu(W)}}_{=\hat{\alpha}_0} + \frac{1}{T_n \nu(W)} \sum_{j=1}^{n} \left\lfloor N_{T_n} \frac{\Delta T_{j-1}}{T_n} \left(1 - \mathrm{e}^{-\hat{\mu}\rho(m_i^0)\Delta T_{j-1}}\right) \right\rfloor. \quad (11)$$

Since $\hat{\mu} > 0$, $\Delta T_{j-1} > 0$ and $\rho(x) > 0$, for all $x > 0$, and since $f(x) = 1 - \mathrm{e}^{-x}$ is strictly increasing and bounded below by 0 and above by 1, for $x > 0$, it is clear that $\hat{\alpha}$ is increasing with $\hat{\mu}$ and

$$\hat{\alpha}_0 = \lim_{\hat{\mu} \to 0} \hat{\alpha}|_{\hat{\mu}} < \hat{\alpha} < \lim_{\hat{\mu} \to \infty} \hat{\alpha}|_{\hat{\mu}} = \hat{\alpha}_0 + \frac{1}{T_n \nu(W)} \sum_{j=1}^{n} \left\lfloor N_{T_n} \frac{\Delta T_{j-1}}{T_n} \right\rfloor.$$

For a random variable $Z = X + Y$ it holds that $\mathrm{Var}(Z) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$. Let now $X = \hat{\alpha}_0$ and let Y be the sum in expression (11). Since X and Y are positively correlated (both contain N_{T_n}) and since $\mathrm{Var}(Y) \geq 0$ it is clear that $\mathrm{Var}(\hat{\alpha}) > \mathrm{Var}(\hat{\alpha}_0)$ for all $\hat{\mu} > 0$. This implies that the trade off for using $\hat{\alpha}$ instead of $\hat{\alpha}_0$ is a higher standard error. Furthermore, as $\hat{\alpha}$ is increasing with $\hat{\mu}$, so is $\mathrm{Var}(\hat{\alpha})$.

Table 5 gives us the estimated means and standard errors (s.e.) of $\hat{\alpha}$ (and $\hat{\alpha}_0$) for a few values of $\hat{\mu}$, based on 30 simulated realisations from the same parameters as the test set (recall that $\alpha = 0.007$; see Section 5.2).

In estimations of μ based on simulated realisations it has been observed that there seems to be no indication of over-estimation of μ. As one can see in Table 5, on average $\hat{\alpha}_0$ under-estimates α more than $\hat{\alpha}$ does when $\hat{\mu} \leq \mu$, in the above scenario indicating that $\hat{\alpha}$ is preferred to $\hat{\alpha}_0$. Note also the smaller standard error of $\hat{\alpha}_0$.

$\alpha = 0.007$	Est. mean	Est. s.e.	Est. bias (%)
$\hat{\alpha}_0 = \lim_{\hat{\mu}\to 0} \hat{\alpha}\|_{\hat{\mu}}$	0.0060	0.0008	-0.00099 (-14%)
$\hat{\alpha}$ ($\hat{\mu} = 0.0002$)	0.0060	0.0008	-0.00099 (-14%)
$\hat{\alpha}$ ($\hat{\mu} = 0.002$)	0.0061	0.0009	-0.00089 (-13%)
$\hat{\alpha}$ ($\hat{\mu} = 0.02$)	0.0074	0.0011	0.00044 (6%)
$\hat{\alpha}$ ($\hat{\mu} = 0.1$)	0.0102	0.0014	0.00320 (46%)
$\hat{\alpha}$ ($\hat{\mu} = 0.2$)	0.0111	0.0016	0.00411 (59%)
$\hat{\alpha}$ ($\hat{\mu} = 5$)	0.0119	0.0017	0.00489 (70%)
$\lim_{\hat{\mu}\to\infty} \hat{\alpha}\|_{\hat{\mu}}$	0.0119	0.0017	0.00489 (70%)

Table 5: Estimated means, standard errors (s.e.), and biases of $\hat{\alpha}$ (and $\hat{\alpha}_0$), based on 30 simulated realisations from the same parameters as the test set (see Section 5.2; recall that $\alpha = 0.007$, $\mu = 0.02$, $\lambda = 0.08$, $K = 0.1$, $c = 2$, and $r = 2$).

References

[1] Barndorff-Nielsen, O.E., Kendall, W.S., van Lieshout, M.N.M. *Stochastic Geometry: Likelihood and Computation.* Monographs on Statistics and Applied Probability 80. Chapman & Hall/CRC, 1999.

[2] Basawa, I.V., Prakasa Rao, B.L.S. *Statistical Inference for Stochastic Processes.* Academic press, 1980.

[3] Berger, U., Hildenbrandt, H. A new approach to spatially explicit modelling of forest dynamics: spacing, ageing and neighbourhood competition of mangrove trees. *Ecological Modelling*, 132:287–302, 2000.

[4] P. Diggle. *Statistical Analysis of Spatial Point Patterns.* Oxford university press, second edition, 2001.

[5] Goreaud, F., Loreau, M., Millier, C. Spatial structure and the survival of an inferior competitor: a theoretical model of neighborhood competition in plants. *Ecological Modelling*, 158:1–19, 2002.

[6] Grabarnik, P., Särkkä, A. Modelling the spatial structure of forest stands by multivariate point processes with hierarchical interactions. *Ecological Modelling*, 220:1232–1240, 2009.

[7] Illian J., Penttinen A., Stoyan H., Stoyan D. *Statistical Analysis and Modelling of Spatial Point Patterns.* Wiley-Interscience, 2008.

[8] Nord-Larsen, T. Modeling individual-tree growth from data with highly irregular measurement intervals. *Forest Science*, 52:198–208, 2006.

[9] J.R. Norris. *Markov Chains.* Cambridge series in statistical and probabilistic mathematics, 1997.

[10] Prévosto, B., Curt, T., Gueugnot, J., Coquillard, P. Modeling mid-elevation Scots pine growth on a volcanic substrate. *Forest Ecology and Management*, 131:223–237, 2000.

[11] E. Renshaw. The linear spatial-temporal interaction process and its relation to $1/\omega$-noise. *J. Roy. Statist. Soc. B*, 56:75–91, 1994.

[12] E. Renshaw. Applying the saddlepoint approximation to bivariate stochastic processes. *Mathematical Biosciences*, 168:57–75, 2000.

[13] Renshaw, E., Comas, C. Space-time generation of high intensity patterns using growth-interaction processes. *Statistics and Computing*, 19:423–437, 2009.

[14] Renshaw, E., Comas, C., Mateu, J. Analysis of forest thinning strategies through the development of space-time growth-interaction simulation models. *Stochastic Environmental Research and Risk Assessment*, 23:275–288, 2009.

[15] Renshaw, E., Särkkä, A. Gibbs point processes for studying the development of spatial-temporal stochastic processes. *Computational Statistics & Data Analysis*, 36:85–105, 2001.

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

[17] Stoyan, D., Kendall, W., Mecke, J. *Stochastic Geometry and its Applications.* John Wiley & sons, second edition, 1995.

[18] Stoyan, D., Penttinen, A. Recent applications of point process methods in forestry statistics. *Statistical Science*, 15:61–78, 2000.

[19] van der Vaart, A.W. *Asymptotic Statistics.* Cambridge series in statistical and probabilistic mathematics, 1998.

[20] Weiner, J., Damgaard, C. Size-asymmetric competition and size-asymmetric growth in a spatially explicit zone-of-influence model of plant competition. *Ecological Research*, 21:707–712, 2006.

[21] Williams, M.S., Williams, M.T., Mowrer, H.T. A boundary reconstruction method for circular fixed-area plots in environmental survey. *Journal of agricultural, biological, and environmental statistics*, 6:479–494, 2001.

Paper II

Maximum likelihood estimation in a discretely observed immigration-death process

Ottmar Cronie and Jun Yu

Research Report
Centre of Biostochastics

Swedish University of Agricultural Sciences

Report 2010:01
ISSN 1651-8543

Maximum likelihood estimation in a discretely observed immigration-death process

Ottmar Cronie[1]

Mathematical Sciences
Chalmers University of Technology and University of Gothenburg, 412 96 Göteborg, Sweden

Jun Yu

Centre of Biostochastics
Swedish University of Agricultural Sciences, 901 83 Umeå, Sweden

Abstract

In order to find the maximum likelihood (ML) estimator of the parameter pair governing the immigration-death process (a continuous time Markov chain) we derive its transition probabilities. The likelihood maximisation problem is reduced from two dimensions to one dimension. We also show the consistency and the asymptotic normality of the ML-estimator under an equidistant sampling scheme, given that the parameter pair lies in some compact subset of the positive part of the real plane. We thereafter evaluate, numerically, the behaviour of the estimator and we finally see how our ML-estimation can be applied to the so called Renshaw-Särkkä growth interaction model; a spatio-temporal point process with time dependent interacting marks in which the immigration-death process controls the arrivals of new marked points as well as their potential life-times.

Keywords: Immigration-death process, $M/M/\infty$-queue, transition probability, likelihood, consistency, asymptotic normality, spatio-temporal marked point process

[1]E-mail address to the corresponding author: ottmar@chalmers.se

1 Introduction

In the case of continuous time Markov chains, the likelihood theory based on continuous observations of sample paths has been covered quite extensively in the literature (see e.g. [2, 3, 13]; see [10] for inference related to branching processes). However, in the case of maximum likelihood (ML) estimation based on processes sampled according to a discrete sampling scheme much less is done. But in later years general results for the asymptotic properties of ML-estimators based on discretely sampled Markov jump processes have emerged (see [5]) and these can be used to establish properties such as strong consistency and asymptotic normality of the ML-estimators for discretely sampled Markov chains.

In this paper we are considering the ML-estimation of the parameters of a particular discretely sampled Markov chain, namely the *immigration-death process* - sometimes also referred to as the $M/M/\infty$-queue (see e.g. [1] or [9]; see [8] for the problem of parameter estimation for immigration-death models when only death times are observed). It is a useful tool which can be used for describing, not only a queue (where the customers arrive according to a Poisson process and get served immediately upon arrival during iid exponential times), but also the dynamics of a population size. Regarding the latter application, one such instance is the role of the immigration-death process in the *Renshaw-Särkkä growth-interaction model* (RS-model) (see [16], [17] and [4]), which has been used to study, among other things, the development of forest stands in time and space [17]. More specifically, the RS-model is a spatio-temporal marked point process, $\mathbb{X}(t) = \{[\mathbf{X}_i, m_i(t)] : i \in \Omega_t\}$, $t \geq 0$, $\mathbf{X}_i \sim Uni(W)$, $W \subseteq \mathbb{R}^d$. Here Ω_t is an index set giving the points present in W at time t and the marks, $m_i(t) \geq 0$, are allowed to interact with each other while growing. The arrivals of new marked points, $[\mathbf{X}_i, m_i(t)]$, and the potential lifetimes of these marked points (they may also die from competition) are governed by an immigration-death process (see e.g. [11] and [18] for general treatments of spatial point process statistics and e.g. [7], [15], and [19] for an overview of spatio-temporal point processes).

We start by finding the transition probabilities of the immigration-death process which give us the likelihood function. Furthermore, we derive its jump intensity function and its transition kernel when viewed as a Markov jump process (Section 2). Treating the process as a Markov jump process, we then proceed to derive the strong consistency and the asymptotic normality of the ML-estimators obtained by sampling the process at equidistant sample times (Section 3). We finally evaluate the ML-estimators numerically (Section 3) and finish off by assessing how these ML-techniques can be used in the RS-model (Section 4).

2 The immigration-death process

The *immigration-death process*, $\{N(t)\}_{t\geq 0}$, 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 $E = \{0, 1, \ldots\}$. It is governed by the parameter pair $\theta = (\alpha, \mu)$ which we henceforth, for technical reasons, assume to take values in some parameter space Θ which is a compact subset of $\mathbb{R}_+^2$. One way of viewing $\{N(t)\}_{t\geq 0}$ is to treat it as a special case of a birth-death process for which the infinitesimal transition probabilities are given by

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

where the birth rates are given by $\lambda_i = \alpha$, $i = 0, 1, \ldots$, and the death rates are given by $\mu_i = i\mu$, $i = 0, 1, \ldots$, ([9], p. 268-270). Within this framework the interpretation of $\{N(t)\}_{t\geq 0}$ is the following. By letting the arrivals of new individuals to a population occur according to a Poisson process with intensity α and upon arrival assigning to all individuals independent and exponentially distributed lifetimes with mean $1/\mu$, $N(t)$ gives us the number of individuals alive at time t. Another possibility is to view it as an $M/M/\infty$ queuing system; each customer (arriving according to a Poisson process with intensity α) is being handled by its own server so that its sojourn time in the system is exponential with intensity μ and independent of all other customers.

Being a Markov process, the finite dimensional distributions of $\{N(t)\}_{t\geq 0}$ are controlled by its transition probabilities, $p_{ij}(t;\theta)$. The exact form of $p_{ij}(t;\theta)$ is given by the following proposition.

Proposition 1. *The transition probabilities of the immigration-death process are given by*

$$\begin{aligned} p_{ij}(t;\theta) &= \frac{\mathrm{e}^{-\frac{\alpha}{\mu}\left(1-\mathrm{e}^{-\mu t}\right)}}{j!} \sum_{k=0}^{j} \left(\frac{\alpha}{\mu}\right)^k \binom{j}{k} \frac{\mathrm{e}^{-(j-k)\mu t}}{(1-\mathrm{e}^{-\mu t})^{j-2k-i}} \frac{i!}{(i-(j-k))!} \\ &= \sum_{k=0}^{j} f_{Poi(\rho)}(k) f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k), \end{aligned} \tag{2.1}$$

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 $\mathrm{e}^{-\mu t}$. Moreover, we have that*

$$\begin{aligned}\mathbb{E}[N(s+t)|N(s)=i] &= i\,\mathrm{e}^{-\mu t}+\rho \qquad (2.2)\\ \mathbb{E}[N^2(s+t)|N(s)=i] &= i(i-1)\,\mathrm{e}^{-2\mu t}+(1+2\rho)i\,\mathrm{e}^{-\mu t}+\rho^2+\rho.\end{aligned}$$

Proof. Given the probability generating function (p.g.f.), $G_X(s)=\mathbb{E}\left[s^X\right]$, of a discrete random variable X it possible to find $P(X=k)$ by evaluating

$$P(X=k)=\frac{1}{k!}\frac{\partial^k}{\partial s^k}G_X(s)\bigg|_{s=0}. \qquad (2.3)$$

Hence, one possible way of finding $p_{ij}(t;\theta)=\mathbb{P}\left(N(h+t)=j|N(h)=i\right)$, $h\geq 0$, is to evaluate expression (2.3) for the p.g.f. of $(N(h+t)|N(h)=i)$, $G(s):=G_{N(h+t)|N(h)=i}(s)$, which is given by ([9], p. 299)

$$\begin{aligned}G(s) &= \left(1+(s-1)\,\mathrm{e}^{-\mu t}\right)^i\exp\left\{(\alpha/\mu)(s-1)\left(1-\mathrm{e}^{-\mu t}\right)\right\}\\ &= \left(1+(s-1)\,\mathrm{e}^{-\mu t}\right)^i\mathrm{e}^{\rho(s-1)}, \qquad (2.4)\end{aligned}$$

where we for convenience have defined $\rho=\frac{\alpha}{\mu}(1-\mathrm{e}^{-\mu t})$.

Considering the first three derivatives $G^{(k)}(s)=\partial^k G(s)/\partial s^k$, $k=1,2,3$, we get

$$\begin{aligned}G^{(1)}(s) &= G(s)\left(\frac{i}{\mathrm{e}^{\mu t}-1+s}+\rho\right) \qquad (2.5)\\ G^{(2)}(s) &= G(s)\left(\frac{i(i-1)}{(\mathrm{e}^{\mu t}-1+s)^2}+2\rho\frac{i}{\mathrm{e}^{\mu t}-1+s}+\rho^2\right)\\ G^{(3)}(s) &= G(s)\left(\frac{i(i-1)(i-2)}{(\mathrm{e}^{\mu t}-1+s)^3}+3\rho\frac{i(i-1)}{(\mathrm{e}^{\mu t}-1+s)^2}+3\rho^2\frac{i}{\mathrm{e}^{\mu t}-1+s}+\rho^3\right).\end{aligned}$$

This suggests that

$$G^{(j)}(s)=G(s)\sum_{k=0}^{j}\rho^k\binom{j}{k}\frac{1}{(\mathrm{e}^{\mu t}-1+s)^{j-k}}\frac{i!}{(i-(j-k))!} \qquad (2.6)$$

and thus

$$\begin{aligned}p_{ij}(t;\theta) &= \frac{G^{(j)}(0)}{j!}\\ &= \frac{\left(1-\mathrm{e}^{-\mu t}\right)^i\mathrm{e}^{-\rho}}{j!}\sum_{k=0}^{j}\rho^k\binom{j}{k}\frac{1}{(\mathrm{e}^{\mu t}-1)^{j-k}}\frac{i!}{(i-(j-k))!}\\ &= \frac{\mathrm{e}^{-\frac{\alpha}{\mu}\left(1-\mathrm{e}^{-\mu t}\right)}}{j!}\sum_{k=0}^{j}\left(\frac{\alpha}{\mu}\right)^k\binom{j}{k}\frac{\mathrm{e}^{-(j-k)\mu t}}{(1-\mathrm{e}^{-\mu t})^{j-2k-i}}\frac{i!}{(i-(j-k))!}.\end{aligned}$$

Now we prove (2.6) by induction. Assume that (2.6) holds for j and let $a(s) = \mathrm{e}^{\mu t} - 1 + s$. It follows from (2.5) and (2.6) that

$$\begin{aligned}
G^{(j+1)}(s) &= G^{(1)}(s) \sum_{k=0}^{j} \rho^k \binom{j}{k} \frac{1}{a(s)^{j-k}} \frac{i!}{(i-(j-k))!} \\
&\quad - G(s) \sum_{k=0}^{j} \rho^k \binom{j}{k} \frac{j-k}{a(s)^{j+1-k}} \frac{i!}{(i-(j-k))!} \\
&= G(s) \left(\frac{i}{a(s)} + \rho \right) \sum_{k=0}^{j} \rho^k \binom{j}{k} \frac{1}{a(s)^{j-k}} \frac{i!}{(i-(j-k))!} \\
&\quad - G(s) \sum_{k=0}^{j} \rho^k \binom{j}{k} \frac{j-k}{a(s)^{j+1-k}} \frac{i!}{(i-(j-k))!}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\frac{G^{(j+1)}(s)}{G(s)} &= \sum_{k=0}^{j} \rho^k \binom{j}{k} \frac{i-(j-k)}{a(s)^{j+1-k}} \frac{i!}{(i-(j-k))!} \\
&\quad + \sum_{k=0}^{j} \rho^{k+1} \binom{j}{k} \frac{1}{a(s)^{j-k}} \frac{i!}{(i-(j-k))!} \\
&= \sum_{k=0}^{j} \rho^k \binom{j}{k} \frac{1}{a(s)^{j+1-k}} \frac{i!}{(i-(j+1-k))!} \\
&\quad + \sum_{k=1}^{j+1} \rho^k \binom{j}{k-1} \frac{1}{a(s)^{j+1-k}} \frac{i!}{(i-(j+1-k))!} \\
&= \sum_{k=0}^{j} \rho^k \binom{j+1}{k} \frac{1}{a(s)^{j+1-k}} \frac{i!}{(i-(j+1-k))!} \frac{j+1-k}{j+1} \\
&\quad + \sum_{k=1}^{j+1} \rho^k \binom{j+1}{k} \frac{1}{a(s)^{j+1-k}} \frac{i!}{(i-(j+1-k))!} \frac{k}{j+1} \\
&= \sum_{k=0}^{j+1} \rho^k \binom{j+1}{k} \frac{1}{a(s)^{j+1-k}} \frac{i!}{(i-(j+1-k))!} \left(\frac{j+1-k}{j+1} + \frac{k}{j+1} \right)
\end{aligned}$$

which implies that (2.6) holds for $j+1$, and therefore completes the proof by induction.

To describe $p_{ij}(t;\theta)$ as a sum of products of Poisson densities and Binomial densities, recall that $\rho = \frac{\alpha}{\mu}(1 - \mathrm{e}^{-\mu t})$ and rewrite $p_{ij}(t;\theta)$ as

$$\begin{aligned} p_{ij}(t;\theta) &= \sum_{k=0}^{j} \frac{\rho^k \mathrm{e}^{-\rho}}{k!} \frac{\mathrm{e}^{-(j-k)\mu t}}{(1-\mathrm{e}^{-\mu t})^{j-k-i}} \frac{k!\binom{j}{k} i!}{j!(i-(j-k))!} \\ &= \sum_{k=0}^{j} \frac{\rho^k \mathrm{e}^{-\rho}}{k!} \binom{i}{j-k} \left(\mathrm{e}^{-\mu t}\right)^{j-k} \left(1-\mathrm{e}^{-\mu t}\right)^{i-(j-k)} \\ &= \sum_{k=0}^{j} f_{Poi(\rho)}(k) f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k) = \sum_{k=0}^{i \wedge j} f_{Poi(\rho)}(j-k) f_{Bin(i,\mathrm{e}^{-\mu t})}(k). \end{aligned}$$

Also, the first two moments of $(N(h+t)|N(h)=i)$ are given by

$$\begin{aligned} \mathbb{E}[N(h+t)|N(h)=i] &= \lim_{s\uparrow 1} G^{(1)}(s) = i\,\mathrm{e}^{-\mu t} + \rho \\ \mathbb{E}[N^2(h+t)|N(h)=i] &= \lim_{s\uparrow 1} [G^{(1)}(s) + G^{(2)}(s)] \\ &= i\,\mathrm{e}^{-\mu t} + \rho + i(i-1)\,\mathrm{e}^{-2\mu t} + 2\rho i\,\mathrm{e}^{-\mu t} + \rho^2 \\ &= i(i-1)\,\mathrm{e}^{-2\mu t} + (1+2\rho) i\,\mathrm{e}^{-\mu t} + \rho^2 + \rho. \end{aligned}$$

□

We will make use of the following recursive expression for the transition probabilities.

Corollary 1. *The transition probabilities can be expressed recursively as*

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

where $i, j \in E = \mathbb{N}$ *and* $\rho = \frac{\alpha}{\mu}\left(1 - \mathrm{e}^{-\mu t}\right)$, *and consequently*

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

Proof. From the proof of Proposition 1 we have that

$$\begin{aligned} G^{(j+1)}(s) &= \left(\frac{i-j}{a(s)} + \rho\right) G^{(j)}(s) \\ &\quad + \frac{j!}{a(s)} \frac{G(s)}{j!} \sum_{k=0}^{j} k\rho^k \binom{j}{k} \frac{1}{a(s)^{j-k}} \frac{i!}{(i-(j-k))!}, \end{aligned}$$

where $a(s) = \mathrm{e}^{\mu t} - 1 + s$, and by noting that

$$\begin{aligned}
p_{ij}(t;\theta)_k &:= \sum_{k=0}^{j} k f_{Poi(\rho)}(k) f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k) \\
&= \sum_{k=1}^{j} k \frac{\rho^k \mathrm{e}^{-\rho}}{k!} \binom{i}{j-k} \left(\mathrm{e}^{-\mu t}\right)^{j-k} \left(1-\mathrm{e}^{-\mu t}\right)^{i-(j-k)} \\
&\overset{l=k-1}{=} \rho \sum_{l=0}^{j-1} \frac{\rho^l \mathrm{e}^{-\rho}}{l!} \binom{i}{j-1-l} \left(\mathrm{e}^{-\mu t}\right)^{j-1-l} \left(1-\mathrm{e}^{-\mu t}\right)^{i-(j-1-l)} \\
&= \rho p_{i(j-1)}(t;\theta).
\end{aligned}$$

we get that

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

□

In practice it is often natural to condition on $N(0) = 0$. In this situation one can easily find that the marginal distribution of $N(t)$ is given by the Poisson distribution with parameter $\rho(t) = \frac{\alpha}{\mu}\left(1 - \mathrm{e}^{-\mu t}\right)$ since $\mathbb{P}\left(N(t) = j\right) = \sum_{i=0}^{\infty} p_{ij}(t;\theta)\mathbb{P}\left(N(0) = i\right) = p_{0j}(t;\theta) = \mathrm{e}^{-\rho(t)} \rho(t)^j / j!$. Furthermore, in this case we get that $N(t) \xrightarrow{d} Poi(\alpha/\mu)$ as $t \to \infty$ since $\lim_{t\to\infty} \rho(t) = \alpha/\mu$. Extending this, the following proposition (see [1]) establishes the ergodicity of $\{N(t)\}_{t\geq 0}$ (which together with the irreducibility gives us its positive recurrence) and its invariant distribution.

Proposition 2. *The immigration-death process is ergodic with invariant distribution given by the Poisson distribution with mean α/μ.*

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.

On the interpretation of $p_{ij}(t;\theta) = \mathbb{P}\left(N(h+t)=j|N(h)=i;\theta\right) = \sum_{k=0}^{i\wedge j} f_{Poi(\rho)}(j-k) f_{Bin(i,\mathrm{e}^{-\mu t})}(k)$, note that

$$\begin{aligned} f_{Poi(\rho)}(j-k) &= \mathbb{P}(j-k \text{ new arrivals during } (h,h+t)) \\ f_{Bin(i,\mathrm{e}^{-\mu t})}(k) &= \mathbb{P}(k \text{ of the } i \text{ individuals alive at time } h \text{ survive } (h,h+t)), \end{aligned}$$

thus implying that $p_{ij}(t;\theta)$ expresses the sum of the probabilities of all possible ways in which we can decrease i individuals to j individuals. Furthermore, when $i \leq j$, we get that $p_{ij}(t;\theta)$ simply represents the convolution of a $Bin(i,\mathrm{e}^{-\mu t})$-density and a $Poi(\rho)$-density, hence expressing the probability that the sum of i iid $Exp(\mathrm{e}^{-\mu t})$-distributed random variables added to a $Poi(\rho)$-distributed random variable takes the value j.

A further characterisation of $\{N(t)\}_{t\geq 0}$ which we will exploit when we establish the asymptotic properties of the ML-estimators is to consider $\{N(t)\}_{t\geq 0}$ as a Markov jump process.

Proposition 3. *Let $\theta = (\alpha,\mu) \in \Theta \subseteq \mathbb{R}_+^2$. $\{N(t)\}_{t\geq 0}$ is a Markov jump process with state space $E=\mathbb{N}$, jump intensity function*

$$\lambda(\theta;i) = \alpha \mathbf{1}\{i=0\} + \min\{\alpha, i\mu\}\mathbf{1}\{i>0\} \quad i \in E,$$

and transition kernel

$$r(\theta;i,j) = \frac{1}{\alpha+\mu i}\left(\alpha\mathbf{1}\{j=i+1\} + \mu i \mathbf{1}\{j=i-1\}\right) \quad i,j \in E.$$

Proof. Let $\{N(t)\}_{t\geq 0}$ be adapted to some suitable filtered probability space $\left(\Omega,\mathcal{F},\{\mathcal{F}_t\}_{t\geq 0},\mathbb{P}\right)$. Since a continuous-time Markov chain by definition is a Markov jump process ([12], p. 243) it holds that $\{N(t)\}_{t\geq 0}$ is a Markov jump process with state space $E=\mathbb{N}$.

Let $0=\tau_0<\tau_1<\tau_2<\ldots$ ($\lim_{n\to\infty}\tau_n=\infty$) be the jump-times of $N(t) = N(0)+\sum_{k=1}^{\infty} Y_k \mathbf{1}\{\tau_k \leq t\}$, having appurtenant jump-sizes $Y_1, Y_2, \ldots$, where $Y_k = N(\tau_k) - N(\tau_{k-1}) \in \{-1,1\}$, $k=1,2,\ldots$ (we consider a right continuous version of $\{N(t)\}_{t\geq 0}$). This is the embedded jump chain of $\{N(t)\}_{t\geq 0}$.

Since $\{N(t)\}_{t\geq 0}$ is a Markov jump process, each increment $\tau_k-\tau_{k-1}$ will be independent of $\mathcal{F}_{\tau_{k-1}}$ and, given that $N(\tau_{k-1})=i$, it holds that $\tau_k - \tau_{k-1}$ is $Exp\left(\lambda(\theta;i)\right)$-distributed. Noticing that the lifetimes of all individuals generated by $N(t)$, $\xi_1,\xi_2,\ldots$, are iid $Exp(\mu)$-distributed and also that an inter-jump-time, τ_α, of the (Poisson) arrival process, $B(t)$, is $Exp(\alpha)$-distributed we get that $\tau_k-\tau_{k-1} \overset{d}{=} \min\{\tau_\alpha,\xi_1,\ldots,\xi_i\}$ 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 $\lambda_1, \ldots, \lambda_n$ is exponentially distributed with parameter $\sum_{i=1}^n \lambda_i$ (see [6]) this implies that the jump intensity function is given by

$$\begin{aligned} \lambda(\theta; i) &= \Big(\mathbb{E}_\theta[\tau_k - \tau_{k-1} | N(\tau_{k-1}) = i]\Big)^{-1} \\ &= \alpha \mathbf{1}\{i = 0\} + \min\{\alpha, i\mu\} \mathbf{1}\{i > 0\}, \quad i \in E, \end{aligned}$$

where $\mathbb{E}_\theta[\cdot]$ denotes expectation under the parameter pair $\theta = (\alpha, \mu)$. Applying again the arguments above we get that

$$\begin{aligned} r(\theta; i, i+1) &= \mathbb{P}(N(\tau_k) = i + 1 | N(\tau_{k-1}) = i) \\ &= \mathbb{P}(\tau_\alpha < \min(\xi_1, \ldots, \xi_i) | N(\tau_{k-1}) = i) \\ &= \int_0^\infty \left(1 - \mathrm{e}^{-\alpha y}\right) f_{\min(\xi_1,\ldots,\xi_i)|N(\tau_{k-1})}(y|i)\, dy \\ &= 1 - \mathbb{E}\left[\mathrm{e}^{-\alpha \min(\xi_1,\ldots,\xi_i)} \,\middle|\, N(\tau_{k-1}) = i\right] \\ &= 1 - \left(1 + \frac{\alpha}{\mu i}\right)^{-1} = \frac{\alpha}{\alpha + \mu i}, \end{aligned}$$

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

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

for all $i, j \in E$ since $|N(\tau_k) - N(\tau_{k-1})| = 1$ for all $k = 1, 2, \ldots$ □

3 Maximum likelihood estimation of α and μ

Assume now that we sample $\{N(t)\}_{t \geq 0}$ as $N_1, \ldots, N_n$ at the respective times $0 = T_0 < T_1 < \ldots < T_n$. 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 $N(t)$ it can be factorised into a product of transition probabilities, i.e. $L_n(\theta) = \mathbb{P}(N(T_1) = N_1) \prod_{k=2}^n p_{N_{k-1} N_k}(t; \theta)$. Since by assumption we condition on $N(T_0) = 0$, the log-likelihood will be given by

$$l_n(\theta) = \sum_{k=1}^n \log p_{N_{k-1} N_k}(\Delta T_{k-1}; \theta), \tag{3.1}$$

where $\Delta T_{k-1} = T_k - T_{k-1}$. 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

$$l_n(\theta) = \sum_{i,j \in E} N_n(i,j) \log p_{ij}(t;\theta), \tag{3.2}$$

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 the expressions (3.1) and (3.2)) will be defined as

$$(\hat{\alpha}_n, \hat{\mu}_n) = \hat{\theta}_n = \arg\max_{\theta \in \Theta} l_n(\theta). \tag{3.3}$$

3.1 The ML-estimators

The ML-estimator for $\theta = (\alpha, \mu)$ is given by solving the system of equations

$$\begin{cases} \frac{\partial}{\partial \alpha} l_n(\theta) &= \sum_{i,j \in E} N_n(i,j) \frac{\partial}{\partial \alpha} \log p_{ij}(t;\theta) = 0 \\ \frac{\partial}{\partial \mu} l_n(\theta) &= \sum_{i,j \in E} N_n(i,j) \frac{\partial}{\partial \mu} \log p_{ij}(t;\theta) = 0. \end{cases} \tag{3.4}$$

As no closed form solution can be found by solving theses likelihood equations, numerical methods have to be employed in order to get ML-estimates. What is possible, however, is to express the estimator of α as a function of both the sample and the parameter μ, hence reducing the maximisation to a one dimensional problem.

Proposition 4. *The ML-estimator, $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\mu}_n)$, is found by maximising $l_n(\hat{\alpha}_n(\mu), \mu)$ over $\Theta_2 \subseteq \mathbb{R}_+$ (the projection of Θ onto the second dimension of $\mathbb{R}^2$), i.e.*

$$\begin{aligned} \hat{\mu}_n &= \arg\max_{\mu \in \Theta_2} l_n(\hat{\alpha}(\mu), \mu) \\ \hat{\alpha}_n &= \hat{\alpha}_n(\hat{\mu}_n), \end{aligned} \tag{3.5}$$

where $\hat{\alpha}_n(\mu)$ is given by expression (3.6).

Proof. The derivatives $\frac{\partial}{\partial \alpha} \log p_{ij}(t;\theta)$ and $\frac{\partial}{\partial \mu} \log p_{ij}(t;\theta)$ are given, respectively, by (A.1) and (A.2) in Appendix A. Plugging these into the system of equations (3.4) we first get

$$\frac{1}{\alpha} \sum_{i,j \in E} N_n(i,j) \frac{p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)} - \frac{\rho}{\alpha} \overbrace{\sum_{i,j \in E} N_n(i,j)}^{=n} = 0$$

which gives us (recall that $\rho = \frac{\alpha}{\mu}\left(1 - \mathrm{e}^{-\mu t}\right)$)

$$\sum_{i,j\in E} N_n(i,j)\frac{p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)} = \frac{\alpha}{\mu}\left(1 - \mathrm{e}^{-\mu t}\right) n.$$

Furthermore,

$$\begin{aligned} 0 &= \frac{\rho\tau}{(1-\mathrm{e}^{-\mu t})\mu}\overbrace{\sum_{i,j\in E} N_n(i,j)}^{=n} - \frac{\mu t}{(1-\mathrm{e}^{-\mu t})\mu}\sum_{i,j\in E} N_n(i,j)(j - i\,\mathrm{e}^{-\mu t}) \\ &\quad + \frac{\tau - \mu t}{(1-\mathrm{e}^{-\mu t})\mu}\sum_{i,j\in E} N_n(i,j)\frac{p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)} \end{aligned}$$

which gives us (recall that $\tau = 1 - \mathrm{e}^{-\mu t} - \mu t\,\mathrm{e}^{-\mu t}$)

$$\sum_{i,j\in E} N_n(i,j)\frac{p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)} = \frac{\rho\tau n - \mu t\sum_{i,j\in E} N_n(i,j)(j - i\,\mathrm{e}^{-\mu t})}{\mu t - \tau}.$$

By putting these two expressions together we get

$$\begin{aligned} \alpha &= \hat{\alpha}_n(\mu) := \frac{\mu/(1-\mathrm{e}^{-\mu t})}{2\left(\frac{1-\mathrm{e}^{-\mu t}}{\mu t} - \mathrm{e}^{-\mu t}\right) - 1}\frac{1}{n}\sum_{i,j\in E} N_n(i,j)(j - i\,\mathrm{e}^{-\mu t}) \\ &= \frac{\mu}{2\left(\frac{1-\mathrm{e}^{-\mu t}}{\mu t} - \mathrm{e}^{-\mu t}\right) - 1}\frac{1}{n}\left(\frac{\mathrm{e}^{-\mu t} N_n - N_0}{1 - \mathrm{e}^{-\mu t}} + \sum_{k=0}^{n} N_k\right). \qquad (3.6) \end{aligned}$$

□

3.2 Asymptotic properties of the ML-estimators

We now wish to establish the consistency and the asymptotic normality of the sequence of estimators (3.3). We do this by showing that the immigration-death process fulfils the conditions under which the related theorems in [5] hold. We first present the theorems of [5] and then give the results for $\{N(t)\}_{t\geq 0}$ as corollaries to the theorems.

The general setting is the following. Let $X(t)$ be a Markov jump process with countable state space E, having transition kernel $r(\theta;\cdot) = \{r(\theta;i,j) : i,j \in E\}$ and intensity function $\lambda(\theta;i)$, which are controlled by the parameter $\theta = (\theta_1, ..., \theta_p) \in \Theta \subseteq \mathbb{R}^p$. We let θ_0 denote the actual value of the underlying controlling parameter. Assume now that we sample $X(t)$ at the times $T_n = nt$, $n \in \mathbb{N}$, $t > 0$ (equidistant sampling). From the Markov property of $X(t)$ the *observation chain*, $Z = (Z_n)_{n=1}^{\infty} \equiv (X(T_n))_{n=1}^{\infty}$, will also be a Markov chain having transition kernel $q(\theta;\cdot) = \{q(\theta;i,j) : i,j \in E\} = \{\mathbb{P}(X(T_n) = j | X(T_{n-1}) = i) : i,j \in E\}$. The log-likelihood of $(Z_1,\ldots,Z_n)$, given that $Z_0 = X(0) = z$, 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 $N_n(i,j) = \sum_{k=1}^{n} \mathbf{1}\{(Z_{k-1}, Z_k) = (i,j)\}$. The likelihood estimator will be defined as

$$\hat{\theta}_n = \arg\max_{\theta\in\Theta} l_n(\theta).$$

In the sequel we denote the partial derivatives of a function $\psi(\cdot)$ of θ by $D_u\psi = \partial\psi/\partial\theta_u$ and $D^2_{uv}\psi = \partial^2\psi/\partial\theta_u\partial\theta_v, u,v = 1,...,p$.

Consider now the following series of conditions put on $(Z_n)_{n\in\mathbb{N}}$.

General conditions (G):

Call any function $\gamma(\cdot)$ defined on $[0,\infty)$ a continuity modulus if it is increasing and $\lim_{x\to 0}\gamma(x) = \gamma(0) = 0$.

(G1) Under θ_0 the Markov chain $(Z_n)_{n\in\mathbb{N}}$ has a unique invariant probability measure π_{θ_0} having moments of order a, for some $a \geq 1$, i.e. $\sum_{i\in E} |i|^a \pi_{\theta_0}(i) < \infty$.

(G2) For any π_{θ_0}-integrable function $\phi : E \to \mathbb{R}$, the following strong law of large numbers holds:

$$\frac{1}{n}\sum_{k=1}^{n}\phi(Z_k) \xrightarrow{a.s.} \sum_{i\in E}\phi(i)\pi_{\theta_0}(i) \quad \text{as } n\to\infty.$$

(G3) Θ is a compact subset of $\mathbb{R}^p$.

(G4) For all $\theta\in\Theta$, $r(\theta;\cdot)$ is an irreducible kernel and $\lambda(\theta;\cdot)$ is positive.

(G5) For some constant C and for all $i,j\in E$,

$$|\log q(\theta_0; i,j)| \leq C(1 + |i|^{a/2} + |j|^{a/2})$$

(G6) There exists a continuity modulus $\gamma(\cdot)$ such that, for all $i,j\in E$ and $\theta,\theta'\in\Theta$,

$$|\log q(\theta; i,j) - \log q(\theta'; i,j)| \leq \gamma(|\theta-\theta'|)(1 + |i|^{a/2} + |j|^{a/2}).$$

Identifiability condition (I):

(I) For any $\theta \neq \theta_0$, $q(\theta;\cdot) \neq q(\theta_0;\cdot)$.

Normality conditions (N):

Assume that θ_0 is an interior point of Θ and that there is a neighbourhood Λ_{θ_0} of θ_0 such that, for any $\theta \in \Lambda_{\theta_0}$ and for any $(i,j) \in E^2$, the mapping $\theta \mapsto g(\theta;i,j) := \log q(\theta_0;i,j) - \log q(\theta;i,j)$ is twice continuously differentiable and satisfies the following conditions for all $u, v = 1, \ldots, p$:

(N1) (i) $\max\left\{|D_u \log q(\theta_0;i,j)|, |D^2_{uv} \log q(\theta_0;i,j)|\right\} \leq C(1+|i|^{a/2}+|j|^{a/2})$;

(ii) there exists a continuity modulus σ_{uv} such that, for $\theta \in \Lambda_{\theta_0}$, $(i,j) \in E^2$,

$$|D^2_{uv} \log q(\theta_0;i,j) - D^2_{uv} \log q(\theta;i,j)| \leq \sigma_{uv}(|\theta_0-\theta|)(1+|i|^{a/2}+|j|^{a/2});$$

(N2) for every $i \in E$, the family of transition kernels $\{q(\theta;i,\cdot) : \theta \in \Lambda_{\theta_0}\}$ is regular at θ_0, in the sense that

(i) $\displaystyle\sum_{j\in E} \left(D_u \log q(\theta_0;i,j)\right) q(\theta_0;i,j) = 0$;

(ii)

$$\begin{aligned} I_{uv}(\theta_0;i) &= \sum_{j\in E} \left(D_u \log q(\theta_0;i,j)\right)\left(D_v \log q(\theta_0;i,j)\right) q(\theta_0;i,j) \\ &= -\sum_{j\in E} \left(D^2_{uv} \log q(\theta_0;i,j)\right) q(\theta_0;i,j). \end{aligned}$$

(N3) The matrix $I(\theta_0;i) = \left(I_{uv}(\theta_0;i)\right)_{u,v=1,\ldots,p}$ is the Fisher information matrix at θ_0 associated with the family of distributions $\{q(\theta;i,\cdot) : \theta \in \Lambda_{\theta_0}\}$. The (asymptotic) Fisher information of $(Z_n)_{n\in\mathbb{N}}$,

$$I(\theta_0) = \sum_{i\in E} I(\theta_0;i)\pi_{\theta_0}(i),$$

is invertible.

Theorem 1. *Let assumptions (G) and (I) hold. Then the maximum likelihood estimator $\hat{\theta}_n$ is strongly consistent, i.e. $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$ as $n \to \infty$.*

Theorem 2. *Let assumptions (G) and (N) hold. Then $\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)$ converges in distribution to the p-dimensional zero-mean Gaussian distribution with covariance matrix $I(\theta_0)^{-1}$, as $n \to \infty$, for every weakly consistent estimator $\hat{\theta}_n$ of θ_0.*

In the case of $\{N(t)\}_{t\geq 0}$ these theorems translate into the following corollaries. We start with the consistency (Corollary 2) and then show the asymptotic normality (Corollary 3).

Corollary 2. *Let* Θ *be any compact subset of* $\mathbb{R}^2_+$. *Then the maximum likelihood estimator for the immigration-death process satisfies*

$$(\hat{\alpha}_n, \hat{\mu}_n) \xrightarrow{a.s.} (\alpha_0, \mu_0)$$

as $n \to \infty$, *where* $(\alpha_0, \mu_0) \in \Theta$ *is the true parameter pair.*

Corollary 3. *Let* Θ *be any compact subset of* $\mathbb{R}^2_+$. *Furthermore, assume that* $(\log(\alpha_0 + \mu_0) - \log(\alpha_0))/\mu_0 \geq 2t$. *Then, as* $n \to \infty$, $\sqrt{n}\,((\hat{\alpha}_n, \hat{\mu}_n) - (\alpha_0, \mu_0))$ *converges in distribution to the two-dimensional zero-mean Gaussian distribution with covariance matrix,* $I(\theta_0)^{-1}$, *given by expression (3.11).*

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

Regarding the condition given in Corollary 3, $g(\alpha_0, \mu_0) := \frac{\log(\alpha_0+\mu_0)-\log(\alpha_0)}{\mu_0} \geq 2t$, by the mean value theorem we get that $\frac{1}{\alpha_0+\mu_0} < g(\alpha_0, \mu_0) < \frac{1}{\alpha_0}$. This means that the condition will be fulfilled if $2t(\alpha_0 + \mu_0) \leq 1$, which is to say that we may sample the process relatively sparsely when both α_0 and μ_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 α_0 increases, with μ_0 kept fixed, we are required to sample the process more densely in order for the condition to hold ($\lim_{\alpha_0\to\infty} g(\alpha_0, \mu_0) = 0$) and when we decrease α_0, with μ_0 fixed, it is more likely that the condition is fulfilled ($\lim_{\alpha_0\to 0} g(\alpha_0, \mu_0) = \infty$). Furthermore, when we let μ_0 increase while keeping α_0 fixed, we move towards a situation where the condition will not be fulfilled ($\lim_{\mu_0\to\infty} g(\alpha_0, \mu_0) = 0$). When we decrease μ_0, with α_0 fixed, so that $N(t)$ is approaching a Poisson process, we get that $\lim_{\mu_0\to 0} g(\alpha_0, \mu_0) = 1/\alpha_0$ so that the condition will be fulfilled provided that α_0 is not too big (note, however, that when $N(t)$ 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).

Proof of Corollary 2. Let Θ be a compact subset of $\mathbb{R}^2_+$ (hence (G3) holds), where $(\alpha, \mu) = \theta \in \Theta$. Furthermore, consider the observation chain of

$\{N(t)\}_{t\geq 0}$, $(Z_n)_{n\in\mathbb{N}}$, where $Z_n = N(T_n) = N(nt)$, and define $q(\theta; i, j) := p_{ij}(t;\theta)$, $i, j \in E = \mathbb{N}$, which constitute the transition kernel $q(\theta;\cdot)$.

By Proposition 2 the invariant distribution of $\{N(t)\}_{t\geq 0}$ under $\theta_0 = (\alpha_0, \mu_0)$, π_{θ_0}, is given by the $Poi(\alpha_0/\mu_0)$-distribution. Since $\pi_{\theta_0} = \pi_{\theta_0}\mathrm{P}(t)$ for any $t \geq 0$, where $\mathrm{P}(t) = (p_{ij}(t))_{i,j\in\mathbb{N}}$ is the matrix of transition probabilities for the time increment t, we see that $\pi_{\theta_0}(\cdot) = \mathbb{P}\left(Poi(\alpha_0/\mu_0) \in \cdot\right)$ is also the invariant probability measure for $(Z_n)_{n\in\mathbb{N}}$, which has moments of all orders $a \in \mathbb{N}$. Hence, condition (G1) is fulfilled.

Due to the positive recurrence of $\{N(t)\}_{t\geq 0}$ (provided by Proposition 2), by an ergodic theorem (e.g. Theorem 1.10.2 in [14]) condition (G2) will be fulfilled.

By Proposition 3 the Markov jump process $\{N(t)\}_{t\geq 0}$ has intensity $\lambda(\theta; i) = \alpha\mathbf{1}\{i = 0\} + \min\{\alpha, i\mu\}\mathbf{1}\{i > 0\}$ which clearly is positive for all $\theta \in \Theta$. Since $\{N(t)\}_{t\geq 0}$ is irreducible if and only if its embedded jump chain, $(Y_n)_{n\geq 1}$, is irreducible ([12], p. 244) we get that its transition kernel $r(\theta;\cdot) = \{r(\theta; i, j) : i, j \in E\}$, $r(\theta; i, j) = \frac{1}{\mu i+\alpha}\left(\alpha\mathbf{1}\{j = i + 1\} + \mu i\mathbf{1}\{j = i - 1\}\right)$, is irreducible for all $\theta \in \Theta$ and thereby condition (G4) is fulfilled.

Since $q(\theta_0; i, j) > 0$ for all $i, j \in E$ we have that $|\log q(\theta_0; i, j)| < \infty$ for all $i, j \in E$. Furthermore, the free choice of $a \in \mathbb{N}$ allows us to create an arbitrary large bound $(1 + |i|^{a/2} + |j|^{a/2})$, when $i, j \in \{2, 3, \ldots\}$. Hence, by choosing, say, $C = \max_{i,j\in\{0,1\}} |q(\theta_0; i, j)|$ we have shown that condition (G5) holds since there are $a \in \mathbb{N}$ such that $|\log q(\theta_0; i, j)| \leq C(1 + |i|^{a/2} + |j|^{a/2})$.

We now wish to show that there is a continuity modulus, $\gamma(\cdot)$, such that

$$|\log q(\theta; i, j) - \log q(\theta'; i, j)| \leq \gamma(|\theta - \theta'|)(1 + |i|^{a/2} + |j|^{a/2}),$$

for all $\theta, \theta' \in \Theta$ and for all $i, j \in E$. Denoting by Θ_1 and Θ_2 the projections of Θ onto the first and the second dimension, respectively, by the compactness of $\Theta \subseteq \mathbb{R}_+^2$ we have that $\alpha_{min} := \inf\Theta_1 > 0$, $\alpha_{max} := \sup\Theta_1 < \infty$, $\mu_{min} := \inf\Theta_2 > 0$ and $\mu_{max} := \sup\Theta_2 < \infty$. By using the bounds given by expressions (A.3) and (A.4), we get that

$$\begin{aligned} |D_1 \log q(\theta; i, j)| &< t + \frac{j}{\alpha} \leq t + \frac{j}{\alpha_{min}} < \infty \\ |D_2 \log q(\theta; i, j)| &< \frac{\alpha t^2 + (3j + i)t}{1 - \mathrm{e}^{-\mu t}} \leq \frac{\alpha_{max} t^2 + (3j + i)t}{1 - \mathrm{e}^{-\mu_{min} t}} < \infty. \end{aligned}$$

Letting $\Lambda = (\alpha_{min}, \alpha_{max}) \times (\mu_{min}, \mu_{max})$ we have, by the mean value theorem and the Schwarz-inequality, for $\theta, \theta^{'} \in \Theta$ and some $0 < c < 1$, that

$$\begin{aligned}
&\left|\log q(\theta; i, j) - \log q(\theta^{'}; i, j)\right| \qquad (3.7)\\
\leq\ & \left|\theta - \theta^{'}\right| \left|\nabla \log q\left((1-c)\theta + c\theta^{'}; i, j\right)\right| \\
=\ & \left|\theta - \theta^{'}\right| \sqrt{(D_1 \log q((1-c)\theta + c\theta'; i, j))^2 + (D_2 \log q((1-c)\theta + c\theta'; i, j))^2} \\
\leq\ & \left|\theta - \theta^{'}\right| \left(|D_1 \log q((1-c)\theta + c\theta^{'}; i, j)| + |D_2 \log q((1-c)\theta + c\theta^{'}; i, j)|\right) \\
\leq\ & \left|\theta - \theta^{'}\right| \sup_{\theta, \theta' \in \bar{\Lambda}} \left(|D_1 \log q(\theta; i, j)| + |D_2 \log q(\theta^{'}; i, j)|\right) \\
<\ & \left(t + \frac{j}{\alpha_{min}} + \frac{\alpha_{max} t^2 + (3j + i)t}{1 - \mathrm{e}^{-\mu_{min} t}}\right) \left|\theta - \theta^{'}\right| (1 + |i|^{a/2} + |j|^{a/2}),
\end{aligned}$$

where $\bar{\Lambda}$ denotes the closure of Λ. Since the free choice of $a \in \mathbb{N}$ (the order of the moment of π_{θ_0}) allows us to make $(1 + |i|^{a/2} + |j|^{a/2})$ as large as required, provided that $i \geq 2$ and/or $j \geq 2$, we only have to take into consideration the cases where $i, j \in \{0, 1\}$. Since the right hand side of (3.7) is maximised when $i = j = 1$ (given that $i, j \in \{0, 1\}$) we choose as continuity modulus

$$\gamma(|\theta - \theta^{'}|) = \left(t + \frac{1}{\alpha_{min}} + \frac{\alpha_{max} t^2 + 4t}{1 - \mathrm{e}^{-\mu_{min} t}}\right) |\theta - \theta^{'}|$$

and we have shown that condition (G6) holds.

To check the identifiability condition (I) consider the probability generating (p.g.f.) function of $(N(h+t)|N(h) = i)$ under $\theta \in \Theta$, $G_i(s; \theta)$, given by (2.4). If $G_i(s; \theta) \neq G_i(s; \theta_0)$, for $\theta \neq \theta_0$, it follows that $\{p_{ij}(t; \theta) : i, j \in E\} \neq \{p_{ij}(t; \theta_0) : i, j \in E\}$. We check whether the assumption $1 = \frac{G_i(s;\theta_0)}{G_i(s;\theta)}$ contradicts any of the three possible scenarios where $\theta \neq \theta_0$. Note that $G_X(1) = \mathbb{E}[1^X] = 1$ for all random variables X so we assume $s \neq 1$.

1. Assume $\alpha \neq \alpha_0$ and $\mu = \mu_0$:

$$1 = \frac{G_i(s; \theta_0)}{G_i(s; \theta)} = \exp\left\{(\alpha_0 - \alpha)(s-1)\left(1 - \mathrm{e}^{-\mu t}\right)/\mu\right\}$$

   holds iff $\alpha_0 = \alpha$.

2. Assume $\alpha = \alpha_0$ and $\mu \neq \mu_0$:
   Since $(1 - \mathrm{e}^{-x})/x$ is a strictly decreasing function

$$1 = \underbrace{\left(\frac{1 + (s-1)\,\mathrm{e}^{-\mu_0 t}}{1 + (s-1)\,\mathrm{e}^{-\mu t}}\right)^i}_{=1 \text{ iff } \mu_0 = \mu \text{ or } i = 0} \exp\left\{\alpha t(s-1)\left(\frac{1 - \mathrm{e}^{-\mu_0 t}}{\mu_0 t} - \frac{1 - \mathrm{e}^{-\mu t}}{\mu t}\right)\right\}$$

   can hold iff $\mu_0 = \mu$.

3. Assume $\alpha \neq \alpha_0$ and $\mu \neq \mu_0$:

$$1 = \underbrace{\left(\frac{1+(s-1)\,\mathrm{e}^{-\mu_0 t}}{1+(s-1)\,\mathrm{e}^{-\mu t}}\right)^i}_{=1 \text{ iff } \mu_0=\mu \text{ or } i=0} \exp\underbrace{\left\{(s-1)\left(\frac{\alpha_0}{\mu_0}(1-\mathrm{e}^{-\mu_0 t}) - \frac{\alpha}{\mu}(1-\mathrm{e}^{-\mu t})\right)\right\}}_{=(*)}.$$

If $\frac{\alpha_0}{\mu_0} = \frac{\alpha}{\mu}$ we get $(*) = 0$ iff $\mu = \mu_0$ (by the monotonicity of $1 - \mathrm{e}^{-x}$), 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 also must require $\mu = \mu_0$.

Hence, there is a one-to-one correspondence between θ and the kernel $q(\theta; \cdot)$. The corollary hereby follows from Theorem 1. □

Proof of Corollary 3. Let Θ be a compact subset of $\mathbb{R}_+^2$ and let $\theta_0 = (\alpha_0, \mu_0)$ be an interior point of Θ. Furthermore, consider the observation chain of $\{N(t)\}_{t\geq 0}$, $(Z_n)_{n\in\mathbb{N}}$, where $Z_n = N(T_n) = N(nt)$, and define $q(\theta; i, j) := p_{ij}(t; \theta)$, $i, j \in E = \mathbb{N}$. From Corollary 2 we know that the estimators (3.3), $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\mu}_n)$, are strongly consistent and that the general conditions (G) hold.

Since the expression for $q(\theta; i, j)$, given by (2.1), contains the term $\mathrm{e}^{-\rho}$ where $\rho = \frac{\alpha}{\mu}\left(1 - \mathrm{e}^{-\mu t}\right)$, we get that, for all $(i, j) \in E^2$ and for all $\theta \in \Theta$, $\log q(\theta; i, j)$ is infinitely many times continuously differentiable w.r.t. θ. This in particular implies that the mapping $\theta \mapsto g(\theta; i, j) := \log q(\theta_0; i, j) - \log q(\theta; i, j)$ is twice continuously differentiable for all θ in some neighbourhood $\Lambda_{\theta_0} \subseteq \Theta$ of θ_0.

Regarding condition (N1) we only have to be concerned with the cases where $i, j \in \{0, 1\}$ since we may choose a as any positive integer, implying that $(1 + |i|^{a/2} + |j|^{a/2})$ can be made as large as required when $i \geq 2$ and/or $j \geq 2$.

Expressions (A.3), (A.4), (A.6), (A.10) and (A.13) in the appendix give us bounds for $|D_u \log q(\theta_0; i, j)|$ and $|D^2_{uv} \log q(\theta_0; i, j)|$, $u, v = 1, 2$, from which we get (recall from the proof of Corollary 2 the definitions of Θ_1, Θ_2, α_{min},

α_{max}, μ_{min} and μ_{max})

$$\begin{aligned}
\max_{(i,j)\in\{0,1\}^2} |D_1 \log q(\theta_0; i,j)| &< \max_{j\in\{0,1\}} \sup_{\alpha\in\Theta_1} \left(\frac{j}{\alpha} + t\right) = \frac{1}{\alpha_{min}} + t =: C_1 < \infty, \\
\max_{(i,j)\in\{0,1\}^2} |D_2 \log q(\theta_0; i,j)| &< \max_{(i,j)\in\{0,1\}^2} \sup_{\mu\in\Theta_2} \sup_{\alpha\in\Theta_1} \left(\frac{\alpha t^2 + (3j+i)t}{1 - \mathrm{e}^{-\mu t}}\right) \\
&= \frac{\alpha_{\max} t^2 + 4t}{1 - \mathrm{e}^{-\mu_{min} t}} =: C_2 < \infty, \\
\max_{(i,j)\in\{0,1\}^2} |D_{11}^2 \log q(\theta_0; i,j)| &< \max_{(i,j)\in\{0,1\}^2} \sup_{\alpha\in\Theta_1} \frac{j + 2(j+\alpha t)^2}{\alpha^2} \\
&< \frac{1 + 2(1+\alpha_{max} t)^2}{\alpha_{min}^2} =: C_{11} < \infty,
\end{aligned}$$

$$\begin{aligned}
&\max_{(i,j)\in\{0,1\}^2} |D_{12}^2 \log q(\theta_0; i,j)| = \max_{(i,j)\in\{0,1\}^2} |D_{21}^2 \log q(\theta_0; i,j)| \\
&< \max_{(i,j)\in\{0,1\}^2} \sup_{\mu\in\Theta_2} \sup_{\alpha\in\Theta_1} \left(\frac{(j^2+j)t}{\alpha} + \alpha t^3 + \frac{j(j+i)t}{(1-\mathrm{e}^{-\mu t})\alpha} + t^2(1+j) \right. \\
&\quad \left. + \frac{j+i}{\mu} t + \frac{(j+\alpha t)(\alpha t^2 + (3j+i)t)}{(1-\mathrm{e}^{-\mu t})\alpha}\right) \\
&< \frac{2t}{\alpha_{min}} + \alpha_{max} t^3 + \frac{2t}{(1-\mathrm{e}^{-\mu_{min} t})\alpha_{min}} + 2t^2 \\
&\quad + \frac{2}{\mu_{min}} t + \frac{(1+\alpha_{max} t)(4+\alpha_{max} t)}{(1-\mathrm{e}^{-\mu_{min} t})\alpha_{min}} t =: C_{12} < \infty,
\end{aligned}$$

$$\begin{aligned}
&\max_{(i,j)\in\{0,1\}^2} |D_{22}^2 \log q(\theta_0; i,j)| < \\
&< \max_{(i,j)\in\{0,1\}^2} \sup_{\mu\in\Theta_2} \sup_{\alpha\in\Theta_1} \left\{ \left(\frac{\alpha t^2 + (3j+i)t}{1-\mathrm{e}^{-\mu t}}\right)^2 \right. \\
&\quad + t^2 \Big(j^2 + 2(j+i)j + \big(2\alpha t + 1 + 2\alpha t + 2\mu^2 t^2 (1+\alpha t)\big) j \\
&\quad \left. + \alpha^2 t^2 + \alpha t(\mu^2 t^2 + 2) + (j+i)^2 \mu^2 t^2 + 2\alpha t(j+i) + (j+i)\mu^2 t^2 \Big) \right\} \\
&< \left(\frac{\alpha_{max} t^2 + 4t}{1-\mathrm{e}^{-\mu_{min} t}}\right)^2 + t^2 \left(6 + 10\alpha_{max} t + \alpha_{max}^2 t^2 + \mu_{max}^2 t^2 (6 + 3\alpha_{max} t)\right) \\
&=: C_{22} < \infty,
\end{aligned}$$

so that by choosing $C = \max\{C_1, C_2, C_{11}, C_{12}, C_{22}\}$ we have 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 $(i, j) \in E^2$.

By the mean value theorem and the Schwarz-inequality it holds that

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

where θ and θ_0 are in some open subset of $\mathbb{R}^2$ (in particular $\theta, \theta_0 \in \Lambda_{\theta_0}$) and $0 < c < 1$. Since, for all $\theta \in \Theta$, by expressions (A.15), (A.16), (A.17) and (A.18), there are bounds such that (by the compactness of Θ)

$$\begin{aligned}
D^3_{111} \log q(\theta; i, j) &< B_{111}(\alpha, \mu, t, j, i) < \infty \\
D^3_{112} \log q(\theta; i, j) &< B_{112}(\alpha, \mu, t, j, i) < \infty \\
D^3_{122} \log q(\theta; i, j) &< B_{122}(\alpha, \mu, t, j, i) < \infty \\
D^3_{222} \log q(\theta; i, j) &< B_{222}(\alpha, \mu, t, j, i) < \infty,
\end{aligned}$$

by choosing the continuity indices according to

$$\begin{aligned}
\sigma_{11}(z) &= \max_{(i,j)\in\{0,1\}^2} \left(\sup_{\mu\in\Theta_2} \sup_{\alpha\in\Theta_1} B_{111}(\alpha, \mu, t, j, i) + \sup_{\mu\in\Theta_2} \sup_{\alpha\in\Theta_1} B_{112}(\alpha, \mu, t, j, i) \right) z \\
\sigma_{12}(z) &= \sigma_{21}(z) \\
&= \max_{(i,j)\in\{0,1\}^2} \left(\sup_{\mu\in\Theta_2} \sup_{\alpha\in\Theta_1} B_{112}(\alpha, \mu, t, j, i) + \sup_{\mu\in\Theta_2} \sup_{\alpha\in\Theta_1} B_{122}(\alpha, \mu, t, j, i) \right) z \\
\sigma_{22}(z) &= \max_{(i,j)\in\{0,1\}^2} \left(\sup_{\mu\in\Theta_2} \sup_{\alpha\in\Theta_1} B_{122}(\alpha, \mu, t, j, i) + \sup_{\mu\in\Theta_2} \sup_{\alpha\in\Theta_1} B_{222}(\alpha, \mu, t, j, i) \right) z
\end{aligned}$$

we have shown that condition (N1) holds.

Turning now to condition (N2), with $\rho_0 = \frac{\alpha_0}{\mu_0}(1 - \mathrm{e}^{-\mu_0 t})$ and $\tau_0 = 1 - \mathrm{e}^{-\mu_0 t} - \mu_0 t\, \mathrm{e}^{-\mu_0 t}$, we have that

$$(D_1 \log q(\theta_0; i, j))\, q(\theta_0; i, j) = \frac{\rho_0}{\alpha_0} \left(p_{i(j-1)}(t; \theta_0) - p_{ij}(t; \theta_0)\right)$$

and

$$\begin{aligned}
(D_2 \log q(\theta_0; i, j))\, q(\theta_0; i, j) &= \frac{\rho_0 \tau_0}{(1 - \mathrm{e}^{-\mu_0 t})\mu_0} \left(p_{ij}(t; \theta_0) - p_{i(j-1)}(t; \theta_0)\right) \\
&\quad - \frac{\left(j - i\, \mathrm{e}^{-\mu_0 t}\right) t}{1 - \mathrm{e}^{-\mu_0 t}} p_{ij}(t; \theta_0) + \frac{\rho_0 t}{1 - \mathrm{e}^{-\mu_0 t}} p_{i(j-1)}(t; \theta_0)
\end{aligned}$$

so that, by considering expression (2.2) and noticing that

$$\sum_{j=0}^{\infty} p_{ij}(t;\theta_0) = \sum_{j=0}^{\infty} p_{i(j-1)}(t;\theta_0) = \sum_{j=0}^{\infty} p_{i(j-2)}(t;\theta_0) = 1,$$

we find that

$$\sum_{j\in E} \left(D_1 \log q(\theta_0;i,j)\right) q(\theta_0;i,j) = \frac{\rho_0}{\alpha_0}\left(\sum_{j=0}^{\infty} p_{i(j-1)}(t;\theta_0) - \sum_{j=0}^{\infty} p_{ij}(t;\theta_0)\right) = 0$$

and

$$\begin{aligned}
&\sum_{j\in E} \left(D_2 \log q(\theta_0;i,j)\right) q(\theta_0;i,j) = \\
&= \frac{\rho_0\tau_0}{(1-\mathrm{e}^{-\mu_0 t})\mu_0}\left(\sum_{j=0}^{\infty} p_{ij}(t;\theta_0) - \sum_{j=0}^{\infty} p_{i(j-1)}(t;\theta_0)\right) \\
&\quad - \frac{t}{1-\mathrm{e}^{-\mu_0 t}}\Bigg(\underbrace{\sum_{j=0}^{\infty} j p_{ij}(t;\theta_0)}_{\overset{(2.2)}{=}\rho_0 + i\,\mathrm{e}^{-\mu_0 t}} - i\,\mathrm{e}^{-\mu_0 t}\Bigg) + \frac{\rho_0 t}{1-\mathrm{e}^{-\mu_0 t}}\sum_{j=0}^{\infty} p_{i(j-1)}(t;\theta_0) = 0.
\end{aligned}$$

Since

$$D^2_{uv} \log q(\theta_0;i,j) = \frac{D^2_{uv} q(\theta_0;i,j)}{q(\theta_0;i,j)} - \left(D_u \log q(\theta_0;i,j)\right)\left(D_v \log q(\theta_0;i,j)\right),$$

checking the condition

$$\begin{aligned}
I_{uv}(\theta_0;i) &= \sum_{j\in E} \left(D_u \log q(\theta_0;i,j)\right)\left(D_v \log q(\theta_0;i,j)\right) q(\theta_0;i,j) \\
&= -\sum_{j\in E} \left(D^2_{uv} \log q(\theta_0;i,j)\right) q(\theta_0;i,j).
\end{aligned}$$

is equivalent to checking

$$\sum_{j\in E} D^2_{uv} q(\theta_0;i,j) = 0,$$

which, according to expressions (A.7), (A.11) and (A.14), holds for all combinations of $u,v \in \{1,2\}$. Thus condition (N2) holds.

Considering expressions (B.1), (B.2) and (B.3), we get that the Fisher information matrix at θ_0 associated with $\{q(\theta; i, \cdot) : \theta \in \Lambda_{\theta_0}\}$ is given by

$$\begin{aligned} I(\theta_0; i) &= \begin{pmatrix} I_{11}(\theta_0; i) & I_{12}(\theta_0; i) \\ I_{21}(\theta_0; i) & I_{22}(\theta_0; i) \end{pmatrix} \\ &= A(\theta_0) + B(\theta_0) i + C(\theta_0) \left(\sum_{j=0}^{\infty} \frac{\left(p_{i(j-1)}(t;\theta)\right)^2}{p_{ij}(t;\theta)} - 1 \right) \end{aligned}$$

where

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

$$C(\theta_0) = \begin{pmatrix} \frac{\rho_0^2}{\alpha_0^2} & \frac{\rho_0(\mu_0 t - \tau_0)}{\mu_0^2} \\ \frac{\rho_0(\mu_0 t - \tau_0)}{\mu_0^2} & \frac{\alpha_0^2 (\tau_0 - \mu_0 t)^2}{\mu_0^4} \end{pmatrix},$$

which implies that the (asymptotic) Fisher information is given by

$$\begin{aligned} I(\theta_0) &= A(\theta_0) + B(\theta_0) \sum_{i \in E} i \pi_{\theta_0}(i) + C(\theta_0) \left(\sum_{i,j \in E} \frac{\left(p_{i(j-1)}(t;\theta_0)\right)^2}{p_{ij_0}(t;\theta)} \pi_{\theta_0}(i) - 1 \right) \\ &= A(\theta_0) + \frac{\alpha_0}{\mu_0} B(\theta_0) + (\Xi - 1) C(\theta_0), \end{aligned} \tag{3.8}$$

where $\Xi = \sum_{i,j \in E} \frac{\left(p_{i(j-1)}(t;\theta_0)\right)^2}{p_{ij}(t;\theta_0)} \pi_{\theta_0}(i)$. It holds that $I(\theta_0)$ is invertible iff

$$\det(I(\theta_0)) = \frac{t^2}{\mu_0^2} \left(\rho_0 (1 + \mathrm{e}^{-\mu_0 t}) (\Xi - 1) - 1 \right) \neq 0,$$

which is to say

$$\Xi \neq \frac{1 + \rho_0 (1 + \mathrm{e}^{-\mu_0 t})}{\rho_0 (1 + \mathrm{e}^{-\mu_0 t})}. \tag{3.9}$$

By Corollary 1 we get that

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

Since $\pi_{\theta_0}(\cdot) = \mathbb{P}(Poi(\alpha_0/\mu_0) \in \cdot)$ is the invariant distribution under θ_0 we have that

$$S_2 = (1 - \mathrm{e}^{\mu_0 t}) \sum_{j=0}^{\infty} \underbrace{\sum_{i=0}^{\infty} p_{ij}(t;\theta_0)\pi_{\theta_0}(i)}_{=\pi_{\theta_0}(j)} = 1 - \mathrm{e}^{\mu_0 t}$$

and

$$S_3 = \frac{1}{\rho_0}\left(1 + \overbrace{\sum_{j=0}^{\infty} j \underbrace{\sum_{i=0}^{\infty} p_{ij}(t;\theta_0)\pi_{\theta_0}(i)}_{=\pi_{\theta_0}(j)}}^{=\alpha_0/\mu_0} - \overbrace{\sum_{i=0}^{\infty} i\pi_{\theta_0}(i) \underbrace{\sum_{j=0}^{\infty} p_{ij}(t;\theta_0)}_{=1}}^{=\alpha_0/\mu_0}\right) = \frac{1}{\rho_0}$$

so that

$$\Xi = S_1 + 1 - \mathrm{e}^{\mu_0 t} + \frac{1}{\rho_0} = S_1 + \frac{1 + \mathrm{e}^{-\mu_0 t} + \rho_0(\mathrm{e}^{-\mu_0 t} - \mathrm{e}^{\mu_0 t})}{\rho_0(1 + \mathrm{e}^{-\mu_0 t})},$$

whereby condition (3.9) is translated into

$$\begin{aligned} 0 &\neq S_1 - \frac{1 + \rho_0(1 + \mathrm{e}^{-\mu_0 t}) - (1 + \mathrm{e}^{-\mu_0 t} + \rho_0(\mathrm{e}^{-\mu_0 t} - \mathrm{e}^{\mu_0 t}))}{\rho_0(1 + \mathrm{e}^{-\mu_0 t})} \\ &= S_1 + \frac{\mathrm{e}^{-\mu_0 t} - \rho_0(1 + \mathrm{e}^{\mu_0 t})}{\rho_0(1 + \mathrm{e}^{-\mu_0 t})}. \end{aligned} \tag{3.10}$$

Clearly $S_1 > 0$ and since $\rho_0(1 + \mathrm{e}^{-\mu_0 t}) > 0$ we get that the right hand side of (3.10) 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 $\mathrm{e}^{-2\mu_0 t}(\alpha_0 + \mu_0) \geq \alpha_0$. Taking logarithms on both sides of the latter inequality we end up with $(\log(\alpha_0 + \mu_0) - \log(\alpha_0))/\mu_0 \geq 2t$, which holds by assumption. This implies that $I(\theta_0)$ is invertible and we conclude that condition (N3) is fulfilled. Its inverse is given by

$$\begin{aligned} I(\theta_0)^{-1} &= \frac{\mu_0}{t\left((1 + \mathrm{e}^{-\mu_0 t})\,\rho_0(\Xi - 1) - 1\right)} \\ &\times \begin{pmatrix} \frac{\rho_0\left(2\tau_0 - \mu_0 t\left(1 - \mathrm{e}^{-\mu_0 t}\right)\right) + \frac{\rho_0^2}{\mu_0 t}(\Xi - 1)(\tau_0 - \mu_0 t)^2}{\left(1 - \mathrm{e}^{-\mu_0 t}\right)^2} & 1 + \frac{\rho_0}{\mu_0 t}(\Xi - 1)(\tau_0 - \mu_0 t) \\ 1 + \frac{\rho_0}{\mu_0 t}(\Xi - 1)(\tau_0 - \mu_0 t) & \frac{1}{\mu_0 t}(\Xi - 1)\left(1 - \mathrm{e}^{-\mu_0 t}\right)^2 \end{pmatrix} \end{aligned} \tag{3.11}$$

so that $\sqrt{n}\,((\hat{\alpha}_n, \hat{\mu}_n) - (\alpha_0, \mu_0)) \xrightarrow{d} N\left(\mathbf{0}, I(\theta_0)^{-1}\right)$, as $n \to \infty$. □

3.3 Numerical evaluations

We here consider 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 the immigration-death process, $N(t)$, on $[0, T]$, $T = 150$, $N(0) = 0$. Thereafter each sample path is sampled at times $T_k = kt$, $t = 1$, $k = 1, \ldots, 150$. For each sample path, based on these discrete observations, we estimate (α_0, μ_0) three times; up to time 50, up to time 100 and up to time 150. Figures 1 and 2 give us normal probability plots of the estimates of our two sets of parameter pairs based on the simulated trajectories. Furthermore, Table 1 and Table 2 display the estimated means, biases, standard errors (s.e.), covariances, skewness (the skewness of a normal distribution is 0) and kurtosis (the kurtosis of a normal distribution is 3) for each parameter pair, (α_0, μ_0), based on its 50 discretely sampled sample paths.

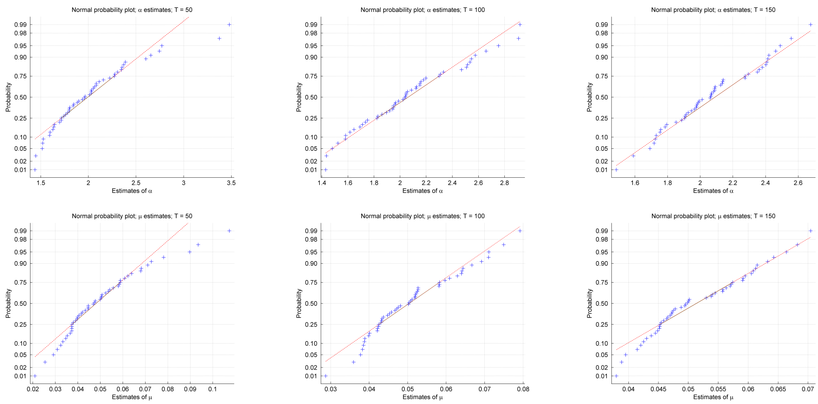

Figure 1: Normal probability plots of the estimates of $(\alpha_0, \mu_0) = (2, 0.05)$ based on 50 sample paths sampled at times $T_k = kt$, $t = 1$, $k = 1, \ldots, T$. Upper row: The estimates of α_0 at final times $T = 50$ (left), $T = 100$ (middle) and $T = 150$ (right). Lower row: The estimates of μ_0 at final times $T = 50$ (left), $T = 100$ (middle) and $T = 150$ (right).

From Figure 1 we can see, not only that the empirical distributions more or less are centred around the actual parameter values, but also how the tails stepwise become lighter, approaching the behaviour of a normal distribution. We can also see how the skewness of the data goes through a stepwise reduction for every additional 50 time units we utilise in the estimation, which further is also verified in Table 1. As a measure of the heaviness of the tails we consider

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 (%)	Std error	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
$T = 100$: $\hat{\alpha}_T$	2.0640	3.2	0.2667	0.1881	2.4832
$\hat{\mu}_T$	0.0517	3.4	0.0081	0.4088	2.2849

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 of Corollary 3, $(\log(\alpha_0 + \mu_0) - \log(\alpha_0))/\mu_0 \geq 2t = 2$, it asymptotically seems to behave Gaussian, thus indicating that the condition 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 (%)	Std error	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, the choice $(\alpha_0, \mu_0) = (0.4, 0.01)$ does fulfil the invertibility condition of Corollary 3. In Figure 2, just as in Figure 1, we can see that each empirical distribution centres around the actual parameter value and the tails approach those of a normal distribution (further verified by the estimated means/biases and kurtoses in Table 2). 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,

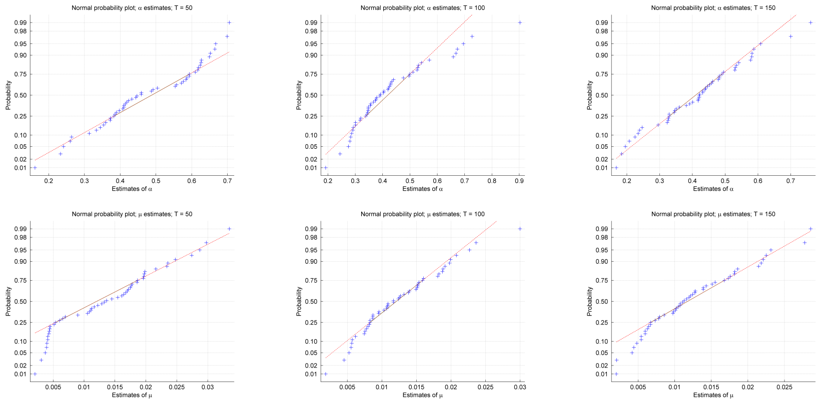

Figure 2: Normal probability plots of the estimates of $(\alpha_0, \mu_0) = (0.4, 0.01)$ based on 50 sample paths sampled at times $T_k = kt$, $t = 1$, $k = 1, \dots, T$. Upper row: The estimates of α_0 at final times $T = 50$ (left), $T = 100$ (middle) and $T = 150$ (right). Lower row: The estimates of μ_0 at final times $T = 50$ (left), $T = 100$ (middle) and $T = 150$ (right).

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 Application: The RS-model

We now turn our focus to a spatio-temporal point process with interacting and size changing marks which here is defined in accordance with [17]. It is defined on $[0, \infty)$ in time and spatially we consider it on some region of interest, $W \subseteq \mathbb{R}^d$, supplied with the Euclidean metric/norm.

More specifically, the process $\mathbb{X}(t) = \{[\mathbf{X}_i, m_i(t)] : i \in \Omega_t\}$ can be described as follows. As time elapses, the arrivals in time of new individuals to W and the time these individuals live in W are governed by an immigration-death process, $N(t)$, having parameter $\theta = (\alpha\nu(W), \mu) \in \Theta$, where $\nu(\cdot)$ denotes volume in $\mathbb{R}^d$ and $\Theta \subseteq \mathbb{R}_+^2$ is compact. We here denote the (Poisson) arrival process by $B(t)$ and the death process by $D(t)$ so that $N(t) = B(t) - D(t)$, where $N(0) = 0$. Furthermore, upon arrival at time t_i^0, individual i is assigned a location $\mathbf{X}_i \sim Uni(W)$ (thus far, at each fixed time t this constitutes a spatial Poisson process with intensity $\frac{\alpha}{\mu}(1 - \mathrm{e}^{-\mu t})$, restricted to W) together with an initial mark, $m_i(t_i^0) = m_i^0$, which is taken either as some fixed positive value (as will be the case here), or as a value drawn from some suitable distribution ([17] considers $m_i^0 \sim Uni(0,\epsilon)$, $\epsilon > 0$). When an individual's ($Exp(\mu)$-distributed) life time has expired we say that the individual has suffered a *natural death.*

Once individual i has received its initial mark it starts growing deterministically according to

$$m_i(t) \quad = \quad m_i^0 + \int_{t_i^0}^{t} dm_i(s), \quad t_i^0 \leq t, \tag{4.1}$$

where

$$dm_i(t) \quad = \quad f(m_i(t);\psi)dt - \sum_{\substack{j\in\Omega_t \\ j\neq i}} h\left(m_i(t), m_j(t), \mathbf{X}_i, \mathbf{X}_j; \psi\right) dt.$$

Here $\Omega_t = \{i \in \{1,\ldots,B(t)\} : \text{individual } i \text{ is alive at time } t\}$, the function $f\left(m_i(t);\psi\right)$ determines the individual growth of mark i in absence of competition with other (neighbouring) individuals and $h\left(m_i(t), m_j(t), \mathbf{X}_i, \mathbf{X}_j; \psi\right)$ is a function handling the individual's spatial interaction with other individuals.

In addition to the natural death, an individual can die *competitively* which we consider to happen as soon as $m_i(t) \leq 0$.

Numerous candidates can be thought of for the individual growth function and the spatial interaction function , depending on the application in question (see [17] for some examples), and here, motivated by the model's forestry applications (see [4]), we will focus on the logistic individual growth function,

$$f(m_i(t);\psi) = \lambda m_i(t)\left(1 - \frac{m_i(t)}{K}\right), \tag{4.2}$$

where $\psi = (\lambda, K, c, r) \in \mathbb{R}_+^2 \times \mathbb{R} \times \mathbb{R}_+$, λ is the growth rate and K is the upper bound (carrying capacity) of the individual's mark size. Further, we choose to consider the so called area interaction function,

$$h\left(m_i(t), m_j(t), \mathbf{X}_i, \mathbf{X}_j; \psi\right) = c\,\frac{\nu\left(B\left[\mathbf{X}_i, rm_i(t)\right] \cap B\left[\mathbf{X}_j, rm_j(t)\right]\right)}{\nu\left(B\left[\mathbf{X}_i, rm_i(t)\right]\right)}, \tag{4.3}$$

where $B\left[\mathbf{x},\epsilon\right]$ denotes a closed ball in $\mathbb{R}^d$ with center $\mathbf{x}$ and radius $\epsilon > 0$. This non-symmetric soft core interaction function has the effect that smaller individuals affect larger individuals less than the other way around. Note that $r \geq 1$ implies that the marks are not allowed to intersect whereas $r < 1$ implies that some intersection between the marks will be allowed before interaction takes place. $c < 0$ implies that individuals gain in size from being close to each other and $c > 0$ has the effect that individuals inhibit each other's growths once $B[\mathbf{X}_i, rm_i(t)] \cap B[\mathbf{X}_j, rm_j(t)] \neq \emptyset$.

By the definitions of Ω_t and $N(t)$, the number of individuals alive at time t is given by

$$|\Omega_t| = N(t) - C(t) = B(t) - D(t) - C(t), \tag{4.4}$$

where $|A|$ denotes the cardinality of the set A and $C(t) \geq 0$ denotes the interactive death process, i.e. the process counting the total number of individuals who have suffered a competitive death in the time interval $(0, t]$. We will assume that $C(T_0) = 0$ so that $|\Omega_{T_0}| = 0$.

4.1 Estimation

Assume now that we sample the process at times $0 = T_0 < \ldots < T_n = T$. Then, for each $k = 1, \ldots, n$, this gives rise to a sampled marked point configuration $\mathbb{X}_{obs}(T_k) = \left\{[\mathbf{x}_i, m_i(T_k)] : i \in \Omega_{T_k}^{obs}\right\}$.

For clarity we here present the least squares approach which we employ for the estimation of $\psi = (\lambda, K, c, r) \in \mathbb{R}_+^2 \times \mathbb{R} \times \mathbb{R}_+$ and also, connected to it, the way in which we label individuals as naturally dead. This approach was originally suggested in [17] wherein it was shown to generate estimates of ψ of good quality.

Let $\tilde{\mathbb{X}}_{obs}(T_k) = \left\{\tilde{m}_i\left(T_{k+1}; \psi, \mathbb{X}_{obs}(T_k)\right) : i \in \Omega_{T_k}^{obs}\right\}$ denote the set of predictions of the actual data marks, $\left\{m_i(T_{k+1}) : i \in \Omega_{T_k}^{obs}\right\}$, generated by equation (4.1) under the regime of ψ, based on the configuration $\mathbb{X}_{obs}(T_k)$ (in practise we employ the simulation algorithm presented in [17] in order to create each predicted set $\tilde{\mathbb{X}}_{obs}(T_k)$ from each set $\mathbb{X}_{obs}(T_k)$). Once having produced $\tilde{\mathbb{X}}_{obs}(T_k)$, if $\tilde{m}_i\left(T_{k+1}; \psi, \mathbb{X}(T_k)\right) > 0$ for an individual $i \in \Omega_{T_k}^{obs}$ but yet $i \notin \Omega_{T_{k+1}}^{obs}$, this predicted individual will be treated as having died by natural causes in (T_k, T_{k+1}). Our least squares estimates are then found by minimising

$$S(\psi) := \sum_{k=1}^{n-1} \sum_{i \in \Omega_{T_k}^{obs}} \mathbf{1}\{i \in \Omega_{T_{k+1}}^{obs}\} \left[\tilde{m}_i\left(T_{k+1}; \psi, \mathbb{X}_{obs}(T_k)\right) - m_i\left(T_{k+1}\right)\right]^2 \tag{4.5}$$

with respect to $\psi = (\lambda, K, c, r) \in \mathbb{R}_+^2 \times \mathbb{R} \times \mathbb{R}_+$, where $\mathbf{1}\{i \in \Omega_{T_{k+1}}^{obs}\}$ is an indicator function being 1 if the actual data individual i is alive at time T_{k+1}.

Regarding the possible edge effects encountered, [4] suggests some edge correction methods which manage to reduce biases generated in the estimation of ψ. Furthermore, [4] also deals with numerical issues related to the minimisation of $S(\psi)$.

The way in which [17] estimates α and μ is to estimate them separately by approximate ML-estimators which we present here for the purpose of comparison. The ML-estimator used to estimate μ in [17] is given by

$$\hat{\mu}_0 = n_T \,/\, \left(\sum_{i=1}^{n_T} t_i + \sum_{j=1}^{m_T} s_j \right), \tag{4.6}$$

where $t_1, \ldots, t_{n_T}$ and $s_1, \ldots, s_{m_T}$ denote, respectively, the lifetimes of the n_T individuals who have been labelled as dead from natural causes by time T and the m_T individuals who are still alive at time T. Since the exact arrival times and death times of the individuals remain unknown, with the only information available being the intervals in which arrivals and deaths occur, the exact lifetimes will remain unknown. The way [17] deals with this is to independently draw each birth time occurring in (T_{k-1}, T_k) from the $Uni(T_{k-1}, T_k)$-distribution while considering the death of an individual to occur at the last sample time at which the individual has been observed.

Note that when estimating α we actually need only to consider the case $\nu(W) = 1$ since we can write α as $\alpha' = \alpha\nu(W)$, find the estimate $\widehat{\alpha'}$ and then get the estimate of α by considering $\hat{\alpha} = \widehat{\alpha'}/\nu(W)$. The approach of [17] is to ignore all deaths occurring by setting $C(T_k) = D(T_k) = 0$, thereby generating the following ML-estimator

$$\hat{\alpha}_0 = \frac{\left| \bigcup_{k=0}^{n} \Omega_{T_k}^{obs} \right|}{T_n}. \tag{4.7}$$

However, using this approach has the consequence that we ignore the interplay between $B(t)$ and $C(t)$ and underestimate α and μ (see [17]). In the case of α this comes from paying no regard to the deaths, which will reduce the number of observed individuals.

A more correct, and thus more sensible, way of estimating μ and α, as opposed to the above approach, is to incorporate the interplay between the deaths and the arrivals of individuals in the estimation by utilising the actual multivariate distribution of $(N(T_1), \ldots, N(T_n))$ in the ML-estimation, i.e. using the likelihood approach developed in the previous sections.

In the minimisation of $S(\psi)$, if $\tilde{m}_i\,(T_{k+1}; \psi, \mathbb{X}(T_k)) \leq 0$ for an individual $i \in \Omega_{T_k}^{obs}$, it will be labelled as having died from competition in (T_k, T_{k+1}) and the total number of such individuals is denoted by $(C(T_k) - C(T_{k-1}))_{obs}^{\psi}$ and is used as an estimate of $C(T_k) - C(T_{k-1})$. Note that by expression (4.4) we can write $N(T_k) = N(T_{k-1}) + |\Omega_{T_k}| - |\Omega_{T_{k-1}}| + C(T_k) - C(T_{k-1})$ where $|\Omega_{T_1}| = C(T_0) = 0$. The observed version of this is given by

$$N_{obs}(T_k) = N_{obs}(T_{k-1}) + |\Omega_{T_k}^{obs}| - |\Omega_{T_{k-1}}^{obs}| + (C(T_k) - C(T_{k-1}))_{obs}^{\psi},$$

where $|\Omega_{T_1}^{obs}| = 0$.

When we here estimate $\theta = (\alpha\nu(W), \mu) \in \Theta$ with our new likelihood approach we use $(N_{obs}(T_1), \ldots, N_{obs}(T_n))$ as observation of the sampled immigration-death process, $(N(T_1), \ldots, N(T_n))$, and hence the log-likelihood is given by

$$l_n(\theta) = \sum_{k=1}^{n} \log p_{N_{obs}(T_{k-1})N_{obs}(T_k)}\left(T_k - T_{k-1}; \alpha\nu(W), \mu\right).$$

5 Discussion

In this paper we have considered the immigration-death process, $N(t)$, and specifically we have treated the ML-estimation of the parameter pair governing it, $\theta = (\alpha, \mu) \in \Theta \subseteq \mathbb{R}_+^2$, when Θ is compact and $N(t)$ is sampled discretely in time; $0 = T_0 < T_1 < \ldots < T_n$, and $N(T_0) = 0$. In order to find the likelihood structure of this Markov process we have derived its transition probabilities, and further, we have managed to reduce the likelihood maximisation from a two-dimensional problem to a one-dimensional problem, where we maximise the likelihood, $L(\alpha, \mu) = L(\hat{\alpha}_n(\mu), \mu)$, over the projection of Θ onto the second dimension of $\mathbb{R}^2$ (μ-axis). Furthermore, by considering $N(t)$ as a Markov jump process we have managed to show that, under an equidistant sampling scheme, $T_k = kt$, $t > 0$, $k = 1, \ldots, n$, the sequence of estimators, $\hat{\theta}_n(N(T_1), \ldots N(T_n))$, is consistent and asymptotically Gaussian. The asymptotic normality requires the invertibility condition $(\log(\alpha_0 + \mu_0) - \log(\alpha_0))/\mu_0 \geq 2t$, where (α_0, μ_0) is the underlying parameter pair. These results have been further corroborated through simulations which also indicate that the estimates approach the actual parameter pair. Furthermore, we see that the empirical distribution of the estimates show strong indications of Gaussianity, even when the invertibility condition of Corollary 3 is not fulfilled. An interesting application for the immigration-death process is the so called RS-model – a spatio-temporal point process with time dependent interacting marks in which $N(t)$ controls the arrivals of new marked points to our region of interest, $W \subseteq \mathbb{R}^d$, as well as their potential life-times – and we discuss how the ML-estimator, $\hat{\theta}_n(N(T_1), \ldots N(T_n))$, could be applied to the RS-model.

The motivation for this work comes from the need of improving the estimation of (α, μ) in the RS-model (compared to the estimators given in [17]) and, as a note on future work, one should numerically study the possible improvement achieved. A further extension is given by adding a Brownian noise in the mark growth function of the RS-model (i.e. letting the marks be controlled by $dM_i(t) = dm_i(t) + dB_i(t)$ where the $B_i(t)$'s are independent Brownian motions) so that it incorporates uncertainties in the mark sizes.

Having made this extension we hope to find a full likelihood structure for this SDE-driven RS-model, where $L(\alpha, \mu)$ constitutes a part of the likelihood structure. A further improvement that possibly can be made is to improve the invertibility condition given in Corollary 3 so that asymptotic normality holds for all $(\alpha_0, \mu_0) \in \Theta$. Furthermore, in order to become more realistic in applications, $N(t)$ could be extended by letting the arrival intensity, α, and the death rate, μ, be non-constant functions of time or in themselves Markov chains (in the latter case $N(t)$ thus becomes a hidden Markov model) whereby, possibly, results similar to the ones found in this paper can be established.

Acknowledgements

Authors would like to thank Aila Särkkä of Chalmers University of Technology and Bo Ranneby of the Swedish University of Agricultural Sciences for useful comments and feedback. This research has been supported by the Swedish Research Council.

References

[1] S. Asmussen. *Applied Probability and Queues.* Springer, second edition, 2003.

[2] Basawa, I.V., Prakasa Rao, B.L.S. *Statistical Inference for Stochastic Processes.* Academic press, 1980.

[3] P. Billingsley. *Statistical Inference for Markov Processes.* The University of Chicago Press, 1961.

[4] O. Cronie. Some edge correction methods for marked spatio-temporal point process models. Preprint, 2009.

[5] Dehay, D., Yao, J.-F. On likelihood estimation for discretely observed Markov jump processes. *Australian & New Zealand Journal of Statistics*, 49:93–107, 2007.

[6] Enger, J., Grandell, J. *Markovprocesser och kteori (Lecture notes in Swedish).* 2000.

[7] Finkenstadt, B., Held, L., Isham, V. *Statistical Methods for Spatio-Temporal Systems.* Chapman & Hall/CRC, 2006.

[8] Gibson, G.J., Renshaw, E. Inference for immigration-death processes with single and paired immigrants. *Inverse Problems*, 17:455–466, 2001.

[9] Grimmett, G., Stirzaker, D. *Probability and Random Processes.* Oxford university press, third edition, 2001.

[10] P. Guttorp. *Statistical Inference for Branching Processes.* Wiley Series in Probability and Mathematical Statistics, 1991.

[11] Illian J., Penttinen A., Stoyan H., Stoyan D. *Statistical Analysis and Modelling of Spatial Point Patterns.* Wiley-Interscience, 2008.

[12] O. Kallenberg. *Foundations of Modern Probability.* Springer, second edition, 2002.

[13] N. Keiding. Maximum likelihood estimation in the birth-and-death process. *Annals of Statistics*, 3:363–372, 1975.

[14] J.R. Norris. *Markov Chains.* Cambridge series in statistical and probabilistic mathematics, 1997.

[15] S.L. Rathbun. Asymptotic properties of the maximum likelihood estimator for spatio-temporal point processes. *Journal of Statistical Planning and Inference*, 51:55–74, 1996.

[16] Renshaw, E., Srkk, A. Gibbs point processes for studying the development of spatial-temporal stochastic processes. *Computational Statistics & Data Analysis*, 36:85–105, 2001.

[17] Srkk, A., Renshaw, E. The analysis of marked point patterns evolving through space and time. *Computational Statistics & Data Analysis*, 51:1698–1718, 2006.

[18] Stoyan, D., Kendall, W., Mecke, J. *Stochastic Geometry and Its Applications.* John Wiley & sons, second edition, 1995.

[19] Verre-Jones, D. Some models and procedures for space-time point processes. *Environmental and Ecological Statistics*, 16:173–195, 2009.

Appendix

A Derivatives

Recall that

$$
\begin{aligned}
p_{ij}(t;\theta) &= \frac{\mathrm{e}^{-\frac{\alpha}{\mu}\left(1-\mathrm{e}^{-\mu t}\right)}}{j!}\sum_{k=0}^{j}\left(\frac{\alpha}{\mu}\right)^{k}\binom{j}{k}\frac{\mathrm{e}^{-(j-k)\mu t}}{(1-\mathrm{e}^{-\mu t})^{j-2k-i}}\frac{i!}{(i-(j-k))!} \\
&= \sum_{k=0}^{j} f_{Poi(\rho)}(k) f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k),
\end{aligned}
$$

where $i, j \in E = \mathbb{N}$, $f_{Poi(\rho)}(\cdot)$ is a 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 a Binomial density with parameters i and $\mathrm{e}^{-\mu t}$. Note further that

$$
\begin{aligned}
p_{ij}(t;\theta)_{k^2} &:= \sum_{k=0}^{j} k^2 f_{Poi(\rho)}(k) f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k) \\
&= \sum_{k=1}^{j} k^2 \frac{\rho^k \mathrm{e}^{-\rho}}{k!}\binom{i}{j-k}\left(\mathrm{e}^{-\mu t}\right)^{j-k}\left(1-\mathrm{e}^{-\mu t}\right)^{i-(j-k)} \\
&\overset{l=k-1}{=} \rho\sum_{l=0}^{j-1}(1+l)\frac{\rho^l \mathrm{e}^{-\rho}}{l!}\binom{i}{j-1-l}\left(\mathrm{e}^{-\mu t}\right)^{j-1-l}\left(1-\mathrm{e}^{-\mu t}\right)^{i-(j-1-l)} \\
&= \rho p_{i(j-1)}(t;\theta) + \rho\sum_{l=1}^{j-1} l\frac{\rho^l \mathrm{e}^{-\rho}}{l!}\binom{i}{j-1-l}\left(\mathrm{e}^{-\mu t}\right)^{j-1-l}\left(1-\mathrm{e}^{-\mu t}\right)^{i-(j-1-l)} \\
&\overset{k=l-1}{=} \rho p_{i(j-1)}(t;\theta) + \rho^2\sum_{k=0}^{j-2}\frac{\rho^k \mathrm{e}^{-\rho}}{k!}\binom{i}{j-2-k}\left(\mathrm{e}^{-\mu t}\right)^{j-2-k}\left(1-\mathrm{e}^{-\mu t}\right)^{i-(j-2-k)} \\
&= \rho p_{i(j-1)}(t;\theta) + \rho^2 p_{i(j-2)}(t;\theta)
\end{aligned}
$$

from which we see that

$$
p_{ij}(t;\theta)_k := \sum_{k=0}^{j} k f_{Poi(\rho)}(k) f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k) = \rho p_{i(j-1)}(t;\theta).
$$

With $\tau = (1 - \mathrm{e}^{-\mu t} - \mu t\,\mathrm{e}^{-\mu t})$ we get that

$$\begin{aligned}
\frac{\partial}{\partial\alpha} f_{Poi(\rho)}(k) &= \frac{k-\rho}{\alpha} f_{Poi(\rho)}(k) \\
\frac{\partial}{\partial\mu} f_{Poi(\rho)}(k) &= \frac{\tau(\rho-k)}{(1-\mathrm{e}^{-\mu t})\mu} f_{Poi(\rho)}(k) \\
\frac{\partial^2}{\partial\alpha\partial\mu} f_{Poi(\rho)}(k) &= \frac{\tau(\rho-(k-\rho)^2)}{(1-\mathrm{e}^{-\mu t})\alpha\mu} f_{Poi(\rho)}(k) \\
\frac{\partial^2}{\partial\alpha^2} f_{Poi(\rho)}(k) &= \frac{\rho^2+k^2-k(1+2\rho)}{\alpha^2} f_{Poi(\rho)}(k) \\
\frac{\partial^2}{\partial\mu^2} f_{Poi(\rho)}(k) &= \Bigg(\frac{-2\rho\tau(1-\mathrm{e}^{-\mu t}) + \rho(1-\mathrm{e}^{-\mu t})\mu^2 t^2\,\mathrm{e}^{-\mu t}}{(1-\mathrm{e}^{-\mu t})^2\mu^2} \\
&\quad + \frac{\rho^2\tau^2}{\left(1-\mathrm{e}^{-\mu t}\right)^2\mu^2} + k^2\frac{\tau^2}{(1-\mathrm{e}^{-\mu t})^2\mu^2} \\
&\quad + k\frac{-2\rho\tau^2 + (1-\mathrm{e}^{-\mu t})^2 - \mu^2 t^2\,\mathrm{e}^{-\mu t}}{\left(1-\mathrm{e}^{-\mu t}\right)^2\mu^2} \Bigg) f_{Poi(\rho)}(k) \\
\frac{\partial}{\partial\mu} f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k) &= \frac{-(j-k-i\,\mathrm{e}^{-\mu t})\mu t}{(1-\mathrm{e}^{-\mu t})\mu} f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k) \\
\frac{\partial^2}{\partial\mu^2} f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k) &= \frac{((j-k)-i\,\mathrm{e}^{-\mu t})^2\mu^2 t^2 + ((j-k)-i)\mu^2 t^2\,\mathrm{e}^{-\mu t}}{(1-\mathrm{e}^{-\mu t})^2\mu^2} \\
&\quad \times f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k).
\end{aligned}$$

Below we will make use of expression (2.2),

$$\sum_{j=0}^{\infty} p_{i(j-2)}(t;\theta) = \sum_{j=0}^{\infty} p_{i(j-1)}(t;\theta) = \sum_{j=0}^{\infty} p_{ij}(t;\theta) = 1$$

and (by expression (2.2))

$$\sum_{j=0}^{\infty} j p_{i(j-1)}(t;\theta) = \sum_{j=0}^{\infty}(j+1)p_{ij}(t;\theta) = \mathbb{E}[N(s+t)|N(s)=i] + 1 = i\,\mathrm{e}^{-\mu t} + \rho + 1.$$

A.1 First order derivatives of $p_{ij}(t;\theta)$ and $\log p_{ij}(t;\theta)$ with bounds

$$
\begin{aligned}
\frac{\partial p_{ij}(t;\theta)}{\partial\alpha} &= \sum_{k=0}^{j} \frac{\partial f_{Poi(\rho)}(k)}{\partial\alpha} f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k) = \sum_{k=0}^{j} \frac{k-\rho}{\alpha} f_{Poi(\rho)}(k) f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k) \\
&= \frac{p_{ij}(t;\theta)_k - \rho p_{ij}(t;\theta)}{\alpha}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \log p_{ij}(t;\theta)}{\partial\alpha} &= \frac{1}{p_{ij}(t;\theta)} \frac{\partial p_{ij}(t;\theta)}{\partial\alpha} = \frac{1}{\alpha}\left(\frac{p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)} - \rho\right) \\
&= \frac{\rho}{\alpha}\left(\frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)} - 1\right)
\end{aligned} \tag{A.1}
$$

$$
\begin{aligned}
\frac{\partial p_{ij}(t;\theta)}{\partial\mu} &= \sum_{k=0}^{j} \frac{\partial f_{Poi(\rho)}(k)}{\partial\mu} f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k) + f_{Poi(\rho)}(k) \frac{\partial f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k)}{\partial\mu} \\
&= \sum_{k=0}^{j} \left(\frac{\rho\tau}{(1-\mathrm{e}^{-\mu t})\mu} - \frac{(j - i\,\mathrm{e}^{-\mu t})\,\mu t}{(1-\mathrm{e}^{-\mu t})\mu} - k\frac{\tau-\mu t}{(1-\mathrm{e}^{-\mu t})\mu} \right) \\
&\quad \times f_{Poi(\rho)}(k) f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k) \\
&= \frac{\rho\tau}{(1-\mathrm{e}^{-\mu t})\mu} p_{ij}(t;\theta) - \frac{(j - i\,\mathrm{e}^{-\mu t})\,\mu t}{(1-\mathrm{e}^{-\mu t})\mu} p_{ij}(t;\theta) - \frac{\tau-\mu t}{(1-\mathrm{e}^{-\mu t})\mu} p_{ij}(t;\theta)_k
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \log p_{ij}(t;\theta)}{\partial\mu} &= \frac{1}{p_{ij}(t;\theta)} \frac{\partial p_{ij}(t;\theta)}{\partial\mu} \\
&= \frac{\rho\tau}{(1-\mathrm{e}^{-\mu t})\mu} - \frac{(j - i\,\mathrm{e}^{-\mu t})\,\mu t}{(1-\mathrm{e}^{-\mu t})\mu} - \frac{\tau-\mu t}{(1-\mathrm{e}^{-\mu t})\mu} \frac{p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)} \\
&= \frac{\rho\tau}{(1-\mathrm{e}^{-\mu t})\mu} - \frac{(j - i\,\mathrm{e}^{-\mu t})\,\mu t}{(1-\mathrm{e}^{-\mu t})\mu} - \frac{\rho(\tau-\mu t)}{(1-\mathrm{e}^{-\mu t})\mu} \frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)}
\end{aligned} \tag{A.2}
$$

Note first that $\rho = \alpha t \frac{1-\mathrm{e}^{-\mu t}}{\mu t} < \alpha t$, $\tau < \mu t$, $\tau < \mu^2 t^2$, $0 < \tau < 1$, $p_{ij}(t;\theta)_k \leq j$ and $p_{ij}(t;\theta)_{k^2} \leq j^2$ since $k \leq j$ for all $k = 0, \ldots, j$. Using the triangle inequality and that $\alpha, \mu, t, i, j > 0$ together with these bounds we get that

$$
\left| \frac{\partial p_{ij}(t;\theta)}{\partial\alpha} \right| \leq \frac{j+\rho}{\alpha} < \frac{j}{\alpha} + t
$$

$$
\left| \frac{\partial \log p_{ij}(t;\theta)}{\partial\alpha} \right| = \frac{1}{\alpha} \left| \frac{p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)} - \rho \right| \leq \frac{j+\rho}{\alpha} < \frac{j}{\alpha} + t \tag{A.3}
$$

$$\left|\frac{\partial p_{ij}(t;\theta)}{\partial\mu}\right| < \frac{\rho\tau + |j - i\,\mathrm{e}^{-\mu t}|\mu t + \rho(\mu t - \tau)}{(1-\mathrm{e}^{-\mu t})\mu} < \frac{(j+i+\rho)t}{1-\mathrm{e}^{-\mu t}} = \frac{(j+i)t}{1-\mathrm{e}^{-\mu t}} + \frac{\alpha}{\mu}$$

$$\begin{aligned}\left|\frac{\partial \log p_{ij}(t;\theta)}{\partial\mu}\right| &< \alpha t^2 \overbrace{\frac{\tau}{(\mu t)^2}}^{<1} + \frac{|i\,\mathrm{e}^{-\mu t} - j|}{1-\mathrm{e}^{-\mu t}} t + \frac{t\,\overbrace{\tau/\mu t}^{<1} + t}{1-\mathrm{e}^{-\mu t}} j \frac{p_{ij}(t;\theta)}{p_{ij}(t;\theta)} \\ &< \frac{\alpha t^2 \overbrace{(1-\mathrm{e}^{-\mu t})}^{<1} + (i+j)t + 2jt}{1-\mathrm{e}^{-\mu t}} < \frac{\alpha t^2 + (3j+i)t}{1-\mathrm{e}^{-\mu t}} \end{aligned} \tag{A.4}$$

A.2 Second order derivatives of $p_{ij}(t;\theta)$ and $\log p_{ij}(t;\theta)$ with bounds

The expressions related to $\frac{\partial^2}{\partial\alpha^2}$:

$$\begin{aligned}\frac{\partial^2 p_{ij}(t;\theta)}{\partial\alpha^2} &= \sum_{k=0}^{j} \frac{\rho^2 + k^2 - k(1+2\rho)}{\alpha^2} f_{Poi(\rho)}(k) f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k) \\ &= \frac{\rho^2 p_{ij}(t;\theta) + p_{ij}(t;\theta)_{k^2} - p_{ij}(t;\theta)_k(1+2\rho)}{\alpha^2} \\ &= \frac{\rho^2}{\alpha^2}\left(p_{i(j-2)}(t;\theta) - 2p_{i(j-1)}(t;\theta) + p_{ij}(t;\theta)\right)\end{aligned}$$

$$\left(\frac{\partial \log p_{ij}(t;\theta)}{\partial\alpha}\right)^2 = \frac{\rho^2}{\alpha^2}\left(\frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)} - 1\right)^2 \tag{A.5}$$

$$\frac{\partial^2 \log p_{ij}(t;\theta)}{\partial\alpha^2} = \frac{1}{p_{ij}(t;\theta)}\frac{\partial^2 p_{ij}(t;\theta)}{\partial\alpha^2} - \left(\frac{\partial \log p_{ij}(t;\theta)}{\partial\alpha}\right)^2$$

$$\left|\frac{\partial^2 p_{ij}(t;\theta)}{\partial\alpha^2}\right| < 2\frac{\rho^2}{\alpha^2} = 2\left(\frac{1-\mathrm{e}^{-\mu t}}{\mu t}\right)^2 t^2 < 2t^2$$

$$\frac{1}{p_{ij}(t;\theta)}\left|\frac{\partial^2 p_{ij}(t;\theta)}{\partial\alpha^2}\right| \leq \frac{\rho^2 + j^2 + j(1+2\rho)}{\alpha^2} < \frac{j}{\alpha^2} + \left(\frac{j}{\alpha} + t\right)^2$$

$$\begin{aligned}\left|\frac{\partial^2 \log p_{ij}(t;\theta)}{\partial\alpha^2}\right| &\leq \frac{1}{p_{ij}(t;\theta)}\left|\frac{\partial^2 p_{ij}(t;\theta)}{\partial\alpha^2}\right| + \left(\frac{\partial \log p_{ij}(t;\theta)}{\partial\alpha}\right)^2 \\ &< \frac{j}{\alpha^2} + 2\left(\frac{j}{\alpha} + t\right)^2\end{aligned} \tag{A.6}$$

$$\sum_{j=0}^{\infty}\frac{\partial^2 p_{ij}(t;\theta)}{\partial\alpha^2} = \frac{\rho^2}{\alpha^2}\left(\sum_{j=0}^{\infty}p_{i(j-2)}(t;\theta) - 2\sum_{j=0}^{\infty}p_{i(j-1)}(t;\theta) + \sum_{j=0}^{\infty}p_{ij}(t;\theta)\right) = 0 \tag{A.7}$$

The expressions related to $\frac{\partial^2}{\partial\alpha\partial\mu}$:

$$\begin{aligned}
\frac{\partial^2 p_{ij}(t;\theta)}{\partial\alpha\partial\mu} &= \sum_{k=0}^{j}\Bigg(- k^2\frac{\tau-\mu t}{(1-\mathrm{e}^{-\mu t})\alpha\mu} + k\frac{2\rho\tau-\rho\mu t-(j-i\,\mathrm{e}^{-\mu t})\mu t}{(1-\mathrm{e}^{-\mu t})\alpha\mu} \\
&\quad + \frac{-\rho^2\tau+\rho\tau}{(1-\mathrm{e}^{-\mu t})\alpha\mu} + \frac{\rho\,(j-i\,\mathrm{e}^{-\mu t})\,\mu t}{(1-\mathrm{e}^{-\mu t})\alpha\mu}\Bigg) f_{Poi(\rho)}(k) f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k) \\
&= -\frac{\tau-\mu t}{(1-\mathrm{e}^{-\mu t})\alpha\mu}p_{ij}(t;\theta)_{k^2} + \frac{2\rho\tau-\rho\mu t-(j-i\,\mathrm{e}^{-\mu t})\mu t}{(1-\mathrm{e}^{-\mu t})\alpha\mu}p_{ij}(t;\theta)_k \\
&\quad + \frac{-\rho^2\tau+\rho\tau}{(1-\mathrm{e}^{-\mu t})\alpha\mu}p_{ij}(t;\theta) + \frac{\rho\,(j-i\,\mathrm{e}^{-\mu t})\,\mu t}{(1-\mathrm{e}^{-\mu t})\alpha\mu}p_{ij}(t;\theta) \\
&= \frac{\rho(\tau-\mu t)}{(1-\mathrm{e}^{-\mu t})\alpha\mu}\left(p_{ij}(t;\theta) - p_{i(j-1)}(t;\theta) - \rho p_{i(j-2)}(t;\theta) + \rho p_{i(j-1)}(t;\theta)\right) \\
&\quad + \frac{\rho^2\tau}{(1-\mathrm{e}^{-\mu t})\alpha\mu}(p_{i(j-1)}(t;\theta) - p_{ij}(t;\theta)) \\
&\quad + \frac{\rho\mu t}{(1-\mathrm{e}^{-\mu t})\alpha\mu}p_{ij}(t;\theta) - \frac{\rho\mu t i\,\mathrm{e}^{-\mu t}}{(1-\mathrm{e}^{-\mu t})\alpha\mu}(p_{ij}(t;\theta) - p_{i(j-1)}(t;\theta)) \\
&\quad + \frac{\rho\mu t}{(1-\mathrm{e}^{-\mu t})\alpha\mu}(jp_{ij}(t;\theta) - jp_{i(j-1)}(t;\theta))
\end{aligned}$$

$$\begin{aligned}
&\frac{\partial\log p_{ij}(t;\theta)}{\partial\alpha}\frac{\partial\log p_{ij}(t;\theta)}{\partial\mu} \\
&= \frac{\rho}{\alpha}\left(\frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)} - 1\right)\left(\frac{\rho\tau}{(1-\mathrm{e}^{-\mu t})\mu} - \frac{(j-i\,\mathrm{e}^{-\mu t})\,\mu t}{(1-\mathrm{e}^{-\mu t})\mu} - \frac{\rho(\tau-\mu t)}{(1-\mathrm{e}^{-\mu t})\mu}\frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)}\right) \\
&= \frac{\rho^2\left(\frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)} - 1\right)}{\frac{\alpha}{\mu}(1-\mathrm{e}^{-\mu t})\mu^2}\left(\mu t\left(\frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)} - \frac{(j-i\,\mathrm{e}^{-\mu t})}{\rho}\right) - \tau\left(\frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)} - 1\right)\right) \\
&= \frac{\rho t}{\mu}\left(\frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)} - 1\right)\left(\frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)} - \frac{(j-i\,\mathrm{e}^{-\mu t})}{\rho}\right) - \frac{\rho\tau}{\mu^2}\left(\frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)} - 1\right)^2 \\
&= \frac{t}{\rho\mu}\left(\frac{p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)} - \rho\right)\left(\frac{p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)} - \left(j-i\,\mathrm{e}^{-\mu t}\right)\right) - \frac{t^2}{\rho}\frac{\tau}{(\mu t)^2}\left(\frac{p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)} - \rho\right)^2 \quad \text{(A.8)}
\end{aligned}$$

$$\frac{\partial^2\log p_{ij}(t;\theta)}{\partial\alpha\partial\mu} = \frac{1}{p_{ij}(t;\theta)}\frac{\partial^2 p_{ij}(t;\theta)}{\partial\alpha\partial\mu} - \frac{\partial\log p_{ij}(t;\theta)}{\partial\alpha}\frac{\partial\log p_{ij}(t;\theta)}{\partial\mu}$$

$$\begin{aligned}
\left|\frac{\partial^2 p_{ij}(t;\theta)}{\partial\alpha\partial\mu}\right| &< \frac{\mu t-\tau}{(\mu t)^2}(1+\rho)\,t^2+\rho t^2\frac{\tau}{(\mu t)^2}+\frac{t}{\mu}+\frac{ti}{\mu}+\frac{t}{\mu}j \\
&< \left(\frac{(1+\rho)}{\mu t}-1\right)t^2+\frac{(1+j+i)t}{\mu} \\
&< \frac{t}{\mu}(1+\rho-\mu t+1+j+i) \\
&< \frac{t}{\mu}(2+\alpha t-\mu t+j+i) \qquad \text{(A.9)}
\end{aligned}$$

$$\left|\frac{\partial\log p_{ij}(t;\theta)}{\partial\alpha}\frac{\partial\log p_{ij}(t;\theta)}{\partial\mu}\right| < \frac{j+\alpha t}{\alpha}\frac{\alpha t^2+(3j+i)t}{1-\mathrm{e}^{-\mu t}}$$

$$\begin{aligned}
\frac{1}{p_{ij}(t;\theta)}\left|\frac{\partial^2 p_{ij}(t;\theta)}{\partial\alpha\partial\mu}\right| &< \frac{\mu t-\tau}{(1-\mathrm{e}^{-\mu t})\alpha\mu}\left|\frac{p_{ij}(t;\theta)_{k^2}-p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)}\right|+\left|-\frac{\rho\tau}{(\mu t)^2}t^2\right| \\
&\quad+\frac{|-(j-i\,\mathrm{e}^{-\mu t})|\mu t}{(1-\mathrm{e}^{-\mu t})\alpha\mu}\frac{p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)}+\frac{\tau}{(\mu t)^2}t^2\left(1+\frac{p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)}\right) \\
&\quad+\frac{|j-i\,\mathrm{e}^{-\mu t}|}{\mu}t \\
&< \frac{(j^2+j)t}{\alpha}+\alpha t^3+\frac{j(j+i)t}{(1-\mathrm{e}^{-\mu t})\alpha}+t^2(1+j)+\frac{j+i}{\mu}t \\
&\quad+\frac{(j+\alpha t)(\alpha t^2+(3j+i)t)}{(1-\mathrm{e}^{-\mu t})\alpha}
\end{aligned}$$

$$\begin{aligned}
\left|\frac{\partial^2\log p_{ij}(t;\theta)}{\partial\alpha\partial\mu}\right| &\le \frac{1}{p_{ij}(t;\theta)}\left|\frac{\partial^2 p_{ij}(t;\theta)}{\partial\alpha\partial\mu}\right|+\left|\frac{\partial\log p_{ij}(t;\theta)}{\partial\alpha}\frac{\partial\log p_{ij}(t;\theta)}{\partial\mu}\right| \\
&< \frac{(j^2+j)t}{\alpha}+\alpha t^3+\frac{j(j+i)t}{(1-\mathrm{e}^{-\mu t})\alpha}+t^2(1+j)+\frac{j+i}{\mu}t \\
&\quad+\frac{(j+\alpha t)(\alpha t^2+(3j+i)t)}{(1-\mathrm{e}^{-\mu t})\alpha} \qquad \text{(A.10)}
\end{aligned}$$

$$
\begin{aligned}
\sum_{j=0}^{\infty} \frac{\partial^2 p_{ij}(t;\theta)}{\partial\alpha\partial\mu} &= \frac{\rho(\tau-\mu t)}{(1-\mathrm{e}^{-\mu t})\alpha\mu} \left(\sum_{j=0}^{\infty} p_{ij}(t;\theta) - \sum_{j=0}^{\infty} p_{i(j-1)}(t;\theta) \right) \\
&+ \frac{\rho^2(\tau-\mu t)}{(1-\mathrm{e}^{-\mu t})\alpha\mu} \left(\sum_{j=0}^{\infty} p_{i(j-1)}(t;\theta) - \sum_{j=0}^{\infty} p_{i(j-2)}(t;\theta) \right) \\
&+ \frac{\rho^2\tau}{(1-\mathrm{e}^{-\mu t})\alpha\mu} \left(\sum_{j=0}^{\infty} p_{i(j-1)}(t;\theta) - \sum_{j=0}^{\infty} p_{ij}(t;\theta) \right) \\
&- \frac{\rho\mu t i\,\mathrm{e}^{-\mu t}}{(1-\mathrm{e}^{-\mu t})\alpha\mu} \left(\sum_{j=0}^{\infty} p_{ij}(t;\theta) - \sum_{j=0}^{\infty} p_{i(j-1)}(t;\theta) \right) \\
&+ \frac{\rho\mu t}{(1-\mathrm{e}^{-\mu t})\alpha\mu} \sum_{j=0}^{\infty} p_{ij}(t;\theta) \\
&+ \frac{\rho\mu t}{(1-\mathrm{e}^{-\mu t})\alpha\mu} \left(\underbrace{\sum_{j=0}^{\infty} j p_{ij}(t;\theta)}_{=\mathbb{E}[N(s+t)|N(s)=i]} - \underbrace{\sum_{j=0}^{\infty} j p_{i(j-1)}(t;\theta)}_{=1+\mathbb{E}[N(s+t)|N(s)=i]} \right) \\
&= 0 \qquad \text{(A.11)}
\end{aligned}
$$

The expressions related to $\frac{\partial^2}{\partial\mu^2}$:

$$
\begin{aligned}
\frac{\partial^2 p_{ij}(t;\theta)}{\partial\mu^2} &= \frac{\rho^2}{\alpha^2} \sum_{k=0}^{j} f_{Poi(\rho)}(k) f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k) \Bigg(k^2 (\tau-\mu t)^2 \\
&+ k \left(2\mu t\tau(j - i\,\mathrm{e}^{-\mu t}) - 2\mu^2 t^2 (j - i\,\mathrm{e}^{-\mu t})\right) \\
&+ k \left(-2\rho\tau^2 + (1-\mathrm{e}^{-\mu t})^2 + 2\rho\mu t(1-\mathrm{e}^{-\mu t}) - 2\mu^2 t^2\,\mathrm{e}^{-\mu t}(1+\rho)\right) \\
&+ \left(\rho^2\tau^2 + \rho(1-\mathrm{e}^{-\mu t})(\mu^2 t^2\,\mathrm{e}^{-\mu t} - 2\tau)\right) \\
&+ \left(\left(j - i\,\mathrm{e}^{-\mu t}\right)^2 \mu^2 t^2 - 2\rho\mu t\tau \left(j - i\,\mathrm{e}^{-\mu t}\right) + (j-i)\mu^2 t^2\,\mathrm{e}^{-\mu t} \right) \Bigg) \\
&= \frac{\rho^2}{\alpha^2} \Bigg((\tau-\mu t)^2 p_{ij}(t;\theta)_{k^2} + 2\mu t (\tau-\mu t)(j - i\,\mathrm{e}^{-\mu t}) p_{ij}(t;\theta)_k \\
&+ \left(-2\rho\tau^2 + (1-\mathrm{e}^{-\mu t})^2 + 2\rho\mu t(1-\mathrm{e}^{-\mu t}) - 2\mu^2 t^2\,\mathrm{e}^{-\mu t}(1+\rho)\right) p_{ij}(t;\theta)_k \\
&+ \left(\rho^2\tau^2 + \rho(1-\mathrm{e}^{-\mu t})(\mu^2 t^2\,\mathrm{e}^{-\mu t} - 2\tau)\right) p_{ij}(t;\theta) \\
&+ \left(\left(j - i\,\mathrm{e}^{-\mu t}\right)^2 \mu^2 t^2 - 2\rho\tau\mu t \left(j - i\,\mathrm{e}^{-\mu t}\right) + (j-i)\mu^2 t^2\,\mathrm{e}^{-\mu t} \right) p_{ij}(t;\theta) \Bigg)
\end{aligned}
$$

$$\left(\frac{\partial \log p_{ij}(t;\theta)}{\partial \mu}\right)^2 = \frac{\rho^2\left(\tau - (j - i\,\mathrm{e}^{-\mu t})\,\mu t/\rho - (\tau - \mu t)\frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)}\right)^2}{(1-\mathrm{e}^{-\mu t})^2\mu^2}$$

$$= \frac{\rho^2\left(\tau\left(1 - \frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)}\right) + \mu t\left(\frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)} - \frac{j - i\,\mathrm{e}^{-\mu t}}{\rho}\right)\right)^2}{(1-\mathrm{e}^{-\mu t})^2\mu^2}$$

$$= \frac{\rho^2\tau^2\left(1 - \frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)}\right)^2}{(1-\mathrm{e}^{-\mu t})^2\mu^2} + \frac{\rho^2(\mu t)^2\left(\frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)} - \frac{j - i\,\mathrm{e}^{-\mu t}}{\rho}\right)^2}{(1-\mathrm{e}^{-\mu t})^2\mu^2}$$

$$+ \frac{2\rho^2\tau\mu t\left(1 - \frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)}\right)\left(\frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)} - \frac{j - i\,\mathrm{e}^{-\mu t}}{\rho}\right)}{(1-\mathrm{e}^{-\mu t})^2\mu^2} \qquad \text{(A.12)}$$

$$\frac{\partial^2 \log p_{ij}(t;\theta)}{\partial \mu^2} = \frac{1}{p_{ij}(t;\theta)}\frac{\partial^2 p_{ij}(t;\theta)}{\partial \mu^2} - \left(\frac{\partial \log p_{ij}(t;\theta)}{\partial \mu}\right)^2$$

$$\left|\frac{\partial^2 p_{ij}(t;\theta)}{\partial \mu^2}\right| < t\frac{(1-\mathrm{e}^{-\mu t})^2}{(\mu t)^2}\Bigg(\mu^2 t^2 j^2 \overbrace{\frac{(\mu t - \tau)^2}{(\mu t)^2}}^{<1} + 2\mu^2 t^2 j(j+i)\overbrace{\frac{\mu t - \tau}{\mu t}}^{<1}$$

$$+ \Big(\overbrace{2\rho\tau^2}^{<\alpha t} + \overbrace{(1-\mathrm{e}^{-\mu t})^2}^{<1} + \overbrace{2\rho\mu t(1-\mathrm{e}^{-\mu t})}^{<2\alpha t} + \overbrace{2\mu^2 t^2\,\mathrm{e}^{-\mu t}(1+\rho)}^{<2(1+\alpha t)}\Big) j$$

$$+ \overbrace{\rho^2\tau^2}^{<1} + \overbrace{\rho(1-\mathrm{e}^{-\mu t})}^{<\alpha t}\overbrace{(2\tau - \mu^2 t^2\,\mathrm{e}^{-\mu t})}^{<2}$$

$$+ (j+i)^2\mu^2 t^2 + 2(j+i)\overbrace{\rho\tau\mu t}^{<\alpha t} + (j+i)\overbrace{\mu^2 t^2\,\mathrm{e}^{-\mu t}}^{<1}\Bigg)$$

$$< t\Big(\mu^2 t^2\left(j^2 + 2j(j+i) + (j+i)^2\right) + \alpha t\,(7j + 2i + 2) + 4j + i + 1\Big)$$

$$\frac{1}{p_{ij}(t;\theta)}\left|\frac{\partial^2 p_{ij}(t;\theta)}{\partial \mu^2}\right| < t^2\Big(j^2 + 2(j+i)j + \left(2\alpha t + 1 + 2\alpha t + 2\mu^2 t^2(1+\alpha t)\right) j$$

$$+ \alpha^2 t^2 + \alpha t(\mu^2 t^2 + 2) + (j+i)^2\mu^2 t^2 + 2\alpha t(j+i) + (j+i)\mu^2 t^2\Big)$$

$$
\begin{aligned}
\left|\frac{\partial^2 \log p_{ij}(t;\theta)}{\partial\mu^2}\right| &\leq \frac{1}{p_{ij}(t;\theta)}\left|\frac{\partial^2 p_{ij}(t;\theta)}{\partial\mu^2}\right| + \left(\frac{\partial \log p_{ij}(t;\theta)}{\partial\mu}\right)^2 \\
&< t^2\Bigg(j^2 + 2(j+i)j + \big(2\alpha t + 1 + 2\alpha t + 2\mu^2 t^2(1+\alpha t)\big)\, j \\
&\quad + \alpha^2 t^2 + \alpha t(\mu^2 t^2 + 2) + (j+i)^2\mu^2 t^2 + 2\alpha t(j+i) + (j+i)\mu^2 t^2\Bigg) \\
&\quad + \left(\frac{\alpha t^2 + (3j+i)t}{1 - \mathrm{e}^{-\mu t}}\right)^2 \qquad \text{(A.13)}
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\alpha^2}{\rho^2}\sum_{j=0}^{\infty}\frac{\partial^2 p_{ij}(t;\theta)}{\partial\mu^2} = \\
= \quad & (\tau-\mu t)^2\sum_{j=0}^{\infty}p_{ij}(t;\theta)_{k^2} + 2\mu t\,(\tau-\mu t)\sum_{j=0}^{\infty}(j-i\,\mathrm{e}^{-\mu t})\underbrace{p_{ij}(t;\theta)_k}_{=\rho p_{i(j-1)}(t;\theta)} \\
& + \left((1-\mathrm{e}^{-\mu t})^2 - 2\rho\tau^2 + 2\rho\mu t(1-\mathrm{e}^{-\mu t}) - 2\mu^2t^2\,\mathrm{e}^{-\mu t}(1+\rho)\right)\sum_{j=0}^{\infty}p_{ij}(t;\theta)_k \\
& + \left(\rho^2\tau^2 + \rho(1-\mathrm{e}^{-\mu t})(\mu^2t^2\,\mathrm{e}^{-\mu t}-2\tau)\right)\sum_{j=0}^{\infty}p_{ij}(t;\theta) \\
& + \sum_{j=0}^{\infty}\left(\left(j-i\,\mathrm{e}^{-\mu t}\right)^2\mu^2t^2 - 2\rho\mu t\tau\left(j-i\,\mathrm{e}^{-\mu t}\right) + (j-i)\mu^2t^2\,\mathrm{e}^{-\mu t}\right)p_{ij}(t;\theta) \\
= \quad & (\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) \\
& + 2\mu t\,(\tau-\mu t)\underbrace{\rho\left(\mathbb{E}[N(s+t)|N(s)=i] + 1 - i\,\mathrm{e}^{-\mu t}\right)}_{\overset{(2.2)}{=}\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^2t^2\,\mathrm{e}^{-\mu t}(1+\rho)\right) \\
& + \rho^2\tau^2 + \rho(1-\mathrm{e}^{-\mu t})(\mu^2t^2\,\mathrm{e}^{-\mu t}-2\tau) \\
& + \mu^2t^2\underbrace{\mathbb{E}\left[N(s+t)^2 - 2x\,\mathrm{e}^{-\mu t}N(s+t) + i^2\,\mathrm{e}^{-2\mu t}|N(s)=i\right]}_{\substack{\overset{(2.2)}{=}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(s+t)|N(s)=i] - i\,\mathrm{e}^{-\mu t}\right)}_{\overset{(2.2)}{=}i\,\mathrm{e}^{-\mu t}+\rho-i\,\mathrm{e}^{-\mu t}=\rho} \\
& + \mu^2t^2\,\mathrm{e}^{-\mu t}\underbrace{\left(\mathbb{E}[N(s+t)|N(s)=i] - i\right)}_{\overset{(2.2)}{=}i\,\mathrm{e}^{-\mu t}+\rho-i=\rho-i(1-\mathrm{e}^{-\mu t})} \\
= \quad & (\tau-\mu t)^2\rho(\rho+1) + 2\mu t\,(\tau-\mu t)\,\rho(\rho+1) \\
& - 2\rho^2\tau^2 + \rho(1-\mathrm{e}^{-\mu t})^2 + 2\rho^2\mu t(1-\mathrm{e}^{-\mu t}) - 2\mu^2t^2\,\mathrm{e}^{-\mu t}\rho(1+\rho) \\
& + \rho^2\tau^2 + \rho(1-\mathrm{e}^{-\mu t})(\mu^2t^2\,\mathrm{e}^{-\mu t}-2\tau) + \mu^2t^2\,\mathrm{e}^{-\mu t}(\rho - i(1-\mathrm{e}^{-\mu t})) \\
& + \mu^2t^2((1-\mathrm{e}^{-\mu t})i\,\mathrm{e}^{-\mu t} + \rho(\rho+1)) - 2\rho^2\mu t\tau \\
= \quad & \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) \\
= \quad & \rho\left(\tau+2\mu t\left(\rho+\mathrm{e}^{-\mu t}\right)-\tau\right)(\tau-\tau) = 0 \qquad \text{(A.14)}
\end{aligned}
$$

A.3 Third order derivatives of $p_{ij}(t;\theta)$ and $\log p_{ij}(t;\theta)$ with bounds

The expressions related to $\frac{\partial^3}{\partial\alpha^3}$:

Since $\frac{\partial p_{ij}(t;\theta)}{\partial\alpha} = \frac{\rho}{\alpha}\left(p_{i(j-1)}(t;\theta) - p_{ij}(t;\theta)\right)$, we get

$$\begin{aligned}
\frac{\partial^3 p_{ij}(t;\theta)}{\partial\alpha^3} &= \frac{\rho^2}{\alpha^2}\left(\frac{\partial}{\partial\alpha}p_{i(j-2)}(t;\theta) - 2\frac{\partial}{\partial\alpha}p_{i(j-1)}(t;\theta) + \frac{\partial}{\partial\alpha}p_{ij}(t;\theta)\right)\\
&= \frac{\rho^3}{\alpha^3}\left(p_{i(j-3)}(t;\theta) - 3p_{i(j-2)}(t;\theta) + 3p_{i(j-1)}(t;\theta) - p_{ij}(t;\theta)\right)\\
&= \frac{\rho^3}{\alpha^3}\left(p_{i(j-3)}(t;\theta) - 2p_{i(j-2)}(t;\theta) + p_{i(j-1)}(t;\theta)\right)\\
&\quad -\frac{\rho^3}{\alpha^3}\left(p_{i(j-2)}(t;\theta) - 2p_{i(j-1)}(t;\theta) + p_{ij}(t;\theta)\right)\\
&= \frac{\rho}{\alpha}\left(\frac{\partial^2 p_{ij}(t;\theta)}{\partial\alpha^2} - \frac{\partial^2 p_{i(j-1)}(t;\theta)}{\partial\alpha^2}\right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^3 \log p_{ij}(t;\theta)}{\partial\alpha^3} &= \frac{1}{p_{ij}(t;\theta)}\frac{\partial^3 p_{ij}(t;\theta)}{\partial\alpha^3} - 3\frac{1}{p_{ij}(t;\theta)}\frac{\partial^2 p_{ij}(t;\theta)}{\partial\alpha^2}\frac{\partial \log p_{ij}(t;\theta)}{\partial\alpha}\\
&\quad + \left(\frac{\partial \log p_{ij}(t;\theta)}{\partial\alpha}\right)^3
\end{aligned}$$

$$\begin{aligned}
&\frac{1}{p_{ij}(t;\theta)}\left|\frac{\partial^3 p_{ij}(t;\theta)}{\partial\alpha^3}\right| \le\\
&\le \frac{\rho}{\alpha}\frac{1}{p_{ij}(t;\theta)}\left|\frac{\partial^2 p_{ij}(t;\theta)}{\partial\alpha^2}\right| + \frac{\rho}{\alpha}\frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)}\frac{1}{p_{i(j-1)}(t;\theta)}\left|\frac{\partial^2 p_{i(j-1)}(t;\theta)}{\partial\alpha^2}\right|\\
&< \alpha t\left(\frac{j}{\alpha^2} + \left(\frac{j}{\alpha} + t\right)^2\right) + \frac{\rho}{\alpha}\frac{1}{\rho}\frac{p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)}\left(\frac{j-1}{\alpha^2} + \left(\frac{j-1}{\alpha} + t\right)^2\right)\\
&< \frac{j + (j+\alpha t)^2}{\alpha}t + \frac{j}{\alpha}\left(\frac{j-1}{\alpha^2} + \left(\frac{j-1}{\alpha} + t\right)^2\right)
\end{aligned}$$

$$\begin{aligned}
&\left|\frac{\partial^3 \log p_{ij}(t;\theta)}{\partial\alpha^3}\right| \leq \\
&\leq \quad \frac{1}{p_{ij}(t;\theta)}\left|\frac{\partial^3 p_{ij}(t;\theta)}{\partial\alpha^3}\right| + 3\frac{1}{p_{ij}(t;\theta)}\left|\frac{\partial^2 p_{ij}(t;\theta)}{\partial\alpha^2}\right|\left|\frac{\partial \log p_{ij}(t;\theta)}{\partial\alpha}\right| \\
&\quad + \left|\frac{\partial \log p_{ij}(t;\theta)}{\partial\alpha}\right|^3 \\
&< \quad \frac{j+(j+\alpha t)^2}{\alpha}t + \frac{j}{\alpha}\left(\frac{j-1}{\alpha^2} + \left(\frac{j-1}{\alpha}+t\right)^2\right) \\
&\quad + 3\frac{j+(j+\alpha t)^2}{\alpha^2}\left(\frac{j}{\alpha}+t\right) + \left(\frac{j}{\alpha}+t\right)^3 \\
&=: \quad B_{111}(\alpha,\mu,t,j,i) \qquad \text{(A.15)}
\end{aligned}$$

The expressions related to $\frac{\partial^3}{\partial\alpha^2\partial\mu}$:

Since

$$\begin{aligned}
\frac{\partial^2 p_{ij}(t;\theta)}{\partial\alpha^2} &= \frac{\rho^2 p_{ij}(t;\theta) + p_{ij}(t;\theta)_{k^2} - p_{ij}(t;\theta)_k(1+2\rho)}{\alpha^2} \\
&= \frac{\rho^2}{\alpha^2}\left(p_{ij}(t;\theta) + \frac{1}{\rho^2}p_{ij}(t;\theta)_{k^2} - \frac{1+2\rho}{\rho^2}p_{ij}(t;\theta)_k\right) \\
&= \frac{\rho^2}{\alpha^2}\left(p_{i(j-2)}(t;\theta) - 2p_{i(j-1)}(t;\theta) + p_{ij}(t;\theta)\right)
\end{aligned}$$

and $\frac{\partial}{\partial\mu}\frac{\rho^2}{\alpha^2} = -2\frac{\rho\tau}{\alpha\mu^2}$ we get

$$\begin{aligned}
\frac{\partial^3 p_{ij}(t;\theta)}{\partial\alpha^2\partial\mu} &= \frac{\partial}{\partial\mu}\frac{\partial^2 p_{ij}(t;\theta)}{\partial\alpha^2} \\
&= -2\frac{\rho\tau}{\alpha\mu^2}\left(p_{ij}(t;\theta) + \frac{1}{\rho^2}p_{ij}(t;\theta)_{k^2} - \frac{1+2\rho}{\rho^2}p_{ij}(t;\theta)_k\right) \\
&\quad + \frac{\rho^2}{\alpha^2}\left(\frac{\partial}{\partial\mu}p_{i(j-2)}(t;\theta) - 2\frac{\partial}{\partial\mu}p_{i(j-1)}(t;\theta) + \frac{\partial}{\partial\mu}p_{ij}(t;\theta)\right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^3 \log p_{ij}(t;\theta)}{\partial\alpha^2\partial\mu} &= \frac{1}{p_{ij}(t;\theta)}\frac{\partial^3 p_{ij}(t;\theta)}{\partial\alpha^2\partial\mu} - 2\frac{1}{p_{ij}(t;\theta)}\frac{\partial^2 p_{ij}(t;\theta)}{\partial\alpha\partial\mu}\frac{\partial \log p_{ij}(t;\theta)}{\partial\alpha} \\
&\quad + \frac{\partial \log p_{ij}(t;\theta)}{\partial\mu}\left(\left(\frac{\partial \log p_{ij}(t;\theta)}{\partial\alpha}\right)^2 - \frac{\partial^2 \log p_{ij}(t;\theta)}{\partial\alpha^2}\right)
\end{aligned}$$

$$
\begin{aligned}
\frac{1}{p_{ij}(t;\theta)}\left|\frac{\partial p_{ij}(t;\theta)}{\partial\mu}\right| &= \left|\frac{\rho\tau}{(1-\mathrm{e}^{-\mu t})\mu} - \frac{(j-i\,\mathrm{e}^{-\mu t})\,\mu t}{(1-\mathrm{e}^{-\mu t})\mu} - \frac{\tau-\mu t}{(1-\mathrm{e}^{-\mu t})\mu}\frac{p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)}\right| \\
&< \frac{\rho\tau}{(1-\mathrm{e}^{-\mu t})\mu} + \frac{(j+i)\mu t}{(1-\mathrm{e}^{-\mu t})\mu} + \frac{\mu t+\tau}{(1-\mathrm{e}^{-\mu t})\mu}j \\
&< \alpha t + \frac{(j+i+1)t}{1-\mathrm{e}^{-\mu t}} + j
\end{aligned}
$$

$$
\begin{aligned}
&\frac{1}{p_{ij}(t;\theta)}\frac{\partial^3 p_{ij}(t;\theta)}{\partial\alpha^2\partial\mu} = \\
&= -2\frac{\tau}{\alpha\mu^2}\left(\rho + \frac{1}{\rho}\frac{p_{ij}(t;\theta)_{k^2}}{p_{ij}(t;\theta)} - \frac{1+2\rho}{\rho}\frac{p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)}\right) + \frac{\rho^2}{\alpha^2}\frac{1}{p_{ij}(t;\theta)}\frac{\partial}{\partial\mu}p_{ij}(t;\theta) \\
&\quad + \frac{\rho^2}{\alpha^2}\frac{p_{i(j-2)}(t;\theta)}{p_{ij}(t;\theta)}\frac{1}{p_{i(j-2)}(t;\theta)}\frac{\partial}{\partial\mu}p_{i(j-2)}(t;\theta) \\
&\quad - 2\frac{\rho^2}{\alpha^2}\frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)}\frac{1}{p_{i(j-1)}(t;\theta)}\frac{\partial}{\partial\mu}p_{i(j-1)}(t;\theta) \\
&= -2\frac{\tau}{\alpha\mu^2}\left(\rho + \frac{1}{\rho}\frac{p_{ij}(t;\theta)_{k^2} - p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)} - 2\frac{p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)}\right) + \frac{\rho^2}{\alpha^2}\frac{1}{p_{ij}(t;\theta)}\frac{\partial}{\partial\mu}p_{ij}(t;\theta) \\
&\quad + \frac{1}{\alpha^2}\frac{p_{ij}(t;\theta)_{k^2} - p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)}\frac{1}{p_{i(j-2)}(t;\theta)}\frac{\partial}{\partial\mu}p_{i(j-2)}(t;\theta) \\
&\quad - 2\frac{\rho}{\alpha^2}\frac{p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)}\frac{1}{p_{i(j-1)}(t;\theta)}\frac{\partial}{\partial\mu}p_{i(j-1)}(t;\theta)
\end{aligned}
$$

$$
\begin{aligned}
&\frac{1}{p_{ij}(t;\theta)}\left|\frac{\partial^3 p_{ij}(t;\theta)}{\partial\alpha^2\partial\mu}\right| < \\
&< 2\frac{t^2}{\alpha}\left(\alpha t + \frac{\mu(j^2+j)}{\alpha(1-\mathrm{e}^{-\mu t})} + 2j\right) + t^2\left(\alpha t + \frac{(j+i+1)t}{1-\mathrm{e}^{-\mu t}} + j\right) \\
&\quad + \frac{j^2+j}{\alpha^2}\left(\alpha t + \frac{(j+i-1)t}{1-\mathrm{e}^{-\mu t}} + j - 2\right) + 2t\frac{j}{\alpha}\left(\alpha t + \frac{(j+i)t}{1-\mathrm{e}^{-\mu t}} + j - 1\right)
\end{aligned}
$$

$$
\begin{aligned}
&\left|\frac{\partial^3 \log p_{ij}(t;\theta)}{\partial\alpha^2\partial\mu}\right| \leq \\
\leq\quad & \frac{1}{p_{ij}(t;\theta)}\left|\frac{\partial^3 p_{ij}(t;\theta)}{\partial\alpha^2\partial\mu}\right| + 2\frac{1}{p_{ij}(t;\theta)}\left|\frac{\partial^2 p_{ij}(t;\theta)}{\partial\alpha\partial\mu}\right|\left|\frac{\partial \log p_{ij}(t;\theta)}{\partial\alpha}\right| \\
& + \left|\frac{\partial \log p_{ij}(t;\theta)}{\partial\mu}\right|\left|\left(\frac{\partial \log p_{ij}(t;\theta)}{\partial\alpha}\right)^2 - \frac{\partial^2 \log p_{ij}(t;\theta)}{\partial\alpha^2}\right| \\
<\quad & 2\frac{t^2}{\alpha}\left(\alpha t + \frac{\mu(j^2+j)}{\alpha(1-\mathrm{e}^{-\mu t})} + 2j\right) + t^2\left(\alpha t + \frac{(j+i+1)t}{1-\mathrm{e}^{-\mu t}} + j\right) \\
& + \frac{j^2+j}{\alpha^2}\left(\alpha t + \frac{(j+i-1)t}{1-\mathrm{e}^{-\mu t}} + j - 2\right) + 2t\frac{j}{\alpha}\left(\alpha t + \frac{(j+i)t}{1-\mathrm{e}^{-\mu t}} + j - 1\right) \\
& + 2\left(\frac{j}{\alpha} + t\right)\left(\frac{(j^2+j)t}{\alpha} + \alpha t^3 + \frac{j(j+i)t + (j+\alpha t)(\alpha t^2 + (3j+i)t)}{(1-\mathrm{e}^{-\mu t})\alpha}\right. \\
& \left. + t^2(1+j) + \frac{j+i}{\mu}t\right) + \frac{\alpha t^2 + (3j+i)t}{1-\mathrm{e}^{-\mu t}}\left(\frac{j}{\alpha^2} + 2\left(\frac{j}{\alpha} + t\right)^2\right) \\
=:\quad & B_{112}(\alpha,\mu,t,j,i) \qquad \text{(A.16)}
\end{aligned}
$$

The expressions related to $\frac{\partial^3}{\partial\alpha\partial\mu^2}$:

Since

$$
\frac{\partial p_{ij}(t;\theta)}{\partial\alpha} = \frac{p_{ij}(t;\theta)_k - \rho p_{ij}(t;\theta)}{\alpha} = \frac{\rho}{\alpha}\left(p_{i(j-1)}(t;\theta) - p_{ij}(t;\theta)\right)
$$

and

$$
\frac{\partial^2}{\partial\mu^2}\frac{\rho}{\alpha} = \frac{\rho}{\alpha}\frac{2\tau - (\mu t)^2\,\mathrm{e}^{-\mu t}}{(1-\mathrm{e}^{-\mu t})\mu^2}
$$

we get

$$
\begin{aligned}
\frac{\partial^3 p_{ij}(t;\theta)}{\partial\alpha\partial\mu^2} &= \left(\frac{\partial^2}{\partial\mu^2}\frac{\rho}{\alpha}\right)\left(p_{i(j-1)}(t;\theta) - p_{ij}(t;\theta)\right) + \frac{\rho}{\alpha}\left(\frac{\partial^2}{\partial\mu^2}p_{i(j-1)}(t;\theta) - \frac{\partial^2}{\partial\mu^2}p_{ij}(t;\theta)\right) \\
&= \frac{2\tau - (\mu t)^2\,\mathrm{e}^{-\mu t}}{(1-\mathrm{e}^{-\mu t})\mu^2}\frac{\partial p_{ij}(t;\theta)}{\partial\alpha} + \frac{\rho}{\alpha}\left(\frac{\partial^2}{\partial\mu^2}p_{i(j-1)}(t;\theta) - \frac{\partial^2}{\partial\mu^2}p_{ij}(t;\theta)\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^3 \log p_{ij}(t;\theta)}{\partial\alpha\partial\mu^2} &= \frac{1}{p_{ij}(t;\theta)}\frac{\partial^3 p_{ij}(t;\theta)}{\partial\alpha\partial\mu^2} - 2\frac{1}{p_{ij}(t;\theta)}\frac{\partial^2 p_{ij}(t;\theta)}{\partial\alpha\partial\mu}\frac{\partial \log p_{ij}(t;\theta)}{\partial\mu} \\
&\quad + \frac{\partial \log p_{ij}(t;\theta)}{\partial\alpha}\left(\left(\frac{\partial \log p_{ij}(t;\theta)}{\partial\mu}\right)^2 - \frac{\partial^2 \log p_{ij}(t;\theta)}{\partial\mu^2}\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{p_{ij}(t;\theta)}\frac{\partial^3 p_{ij}(t;\theta)}{\partial\alpha\partial\mu^2} \quad = \quad & \frac{2\tau-(\mu t)^2\,\mathrm{e}^{-\mu t}}{(1-\mathrm{e}^{-\mu t})\mu^2}\,\frac{1}{\alpha}\left(\frac{p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)}-\rho\right)\\
& + \frac{1}{\alpha}\frac{p_{ij}(t;\theta)_k}{p_{ij}(t;\theta)}\frac{1}{p_{i(j-1)}(t;\theta)}\frac{\partial^2}{\partial\mu^2}p_{i(j-1)}(t;\theta)\\
& - \frac{\rho}{\alpha}\frac{1}{p_{ij}(t;\theta)}\frac{\partial^2}{\partial\mu^2}p_{ij}(t;\theta)
\end{aligned}
$$

$$
\begin{aligned}
& \frac{1}{p_{ij}(t;\theta)}\left|\frac{\partial^3 p_{ij}(t;\theta)}{\partial\alpha\partial\mu^2}\right| < \frac{2+\mathrm{e}^{-\mu t}}{1-\mathrm{e}^{-\mu t}}t^2\left(\frac{j}{\alpha}+t\right)\\
& \quad + \frac{j}{\alpha}t^2\Bigg((j-1)^2+2(j+i-1)(j-1)+\big(2\alpha t+1+2\alpha t+2\mu^2t^2(1+\alpha t)\big)(j-1)\\
& \quad + \alpha^2t^2+\alpha t(\mu^2t^2+2)+(j+i-1)^2\mu^2t^2+2\alpha t(j+i-1)+(j+i-1)\mu^2t^2\Bigg)\\
& \quad + t^3\Bigg(j^2+2(j+i)j+\big(2\alpha t+1+2\alpha t+2\mu^2t^2(1+\alpha t)\big)j\\
& \quad + \alpha^2t^2+\alpha t(\mu^2t^2+2)+(j+i)^2\mu^2t^2+2\alpha t(j+i)+(j+i)\mu^2t^2\Bigg)
\end{aligned}
$$

$$
\begin{aligned}
\left|\frac{\partial^3 \log p_{ij}(t;\theta)}{\partial\alpha\partial\mu^2}\right| \leq & \frac{1}{p_{ij}(t;\theta)}\left|\frac{\partial^3 p_{ij}(t;\theta)}{\partial\alpha\partial\mu^2}\right| + 2\frac{1}{p_{ij}(t;\theta)}\left|\frac{\partial^2 p_{ij}(t;\theta)}{\partial\alpha\partial\mu}\right|\left|\frac{\partial \log p_{ij}(t;\theta)}{\partial\mu}\right| \\
& + \left|\frac{\partial \log p_{ij}(t;\theta)}{\partial\alpha}\right|\left(\left(\frac{\partial \log p_{ij}(t;\theta)}{\partial\mu}\right)^2 + \left|\frac{\partial^2 \log p_{ij}(t;\theta)}{\partial\mu^2}\right|\right) \\
< \quad & \frac{2+\mathrm{e}^{-\mu t}}{1-\mathrm{e}^{-\mu t}} t^2\left(\frac{j}{\alpha}+t\right) \\
& + \frac{j}{\alpha}t^2\Bigg((j-1)^2 + 2(j+i-1)(j-1) + \left(2\alpha t + 1 + 2\alpha t + 2\mu^2 t^2(1+\alpha t)\right)(j-1) \\
& + \alpha^2 t^2 + \alpha t(\mu^2 t^2 + 2) + (j+i-1)^2\mu^2 t^2 + 2\alpha t(j+i-1) + (j+i-1)\mu^2 t^2\Bigg) \\
& + 2\frac{\alpha t^2 + (3j+i)t}{1-\mathrm{e}^{-\mu t}}\Bigg(\frac{(j^2+j)t}{\alpha} + \alpha t^3 + \frac{j(j+i)t}{(1-\mathrm{e}^{-\mu t})\alpha} + t^2(1+j) + \frac{j+i}{\mu}t \\
& + \frac{(j+\alpha t)(\alpha t^2 + (3j+i)t)}{(1-\mathrm{e}^{-\mu t})\alpha}\Bigg) + 2\left(\frac{j}{\alpha}+t\right)\left(\frac{\alpha t^2 + (3j+i)t}{1-\mathrm{e}^{-\mu t}}\right)^2 \\
& + t^2\left(\frac{j}{\alpha}+2t\right)\Bigg(j^2 + 2(j+i)j + \left(2\alpha t + 1 + 2\alpha t + 2\mu^2 t^2(1+\alpha t)\right)j \\
& + \alpha^2 t^2 + \alpha t(\mu^2 t^2 + 2) + (j+i)^2\mu^2 t^2 + 2\alpha t(j+i) + (j+i)\mu^2 t^2\Bigg) \\
=: \quad & B_{122}(\alpha,\mu,t,j,i) \qquad \text{(A.17)}
\end{aligned}
$$

The expressions related to $\frac{\partial^3}{\partial\mu^3}$:

Since

$$
\frac{\partial}{\partial\mu}\frac{\rho^2}{\alpha^2} = \frac{\rho^2}{\alpha^2}\left(\frac{6}{\mu^2} - \frac{8\tau\,\mathrm{e}^{-\mu t}}{(1-\mathrm{e}^{-\mu t})^2\mu}t - \frac{2(1+2\,\mathrm{e}^{-\mu t})\,\mathrm{e}^{-\mu t}}{(1-\mathrm{e}^{-\mu t})^2}t^2\right)
$$

$$
\begin{aligned}
p_{ij}(t;\theta)_{k^3} \quad := & \quad \sum_{k=0}^{j} k^3 f_{Poi(\rho)}(k) f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k) \\
= & \quad \sum_{k=1}^{j} k^3 \frac{\rho^k\,\mathrm{e}^{-\rho}}{k!}\binom{i}{j-k}\left(\mathrm{e}^{-\mu t}\right)^{j-k}\left(1-\mathrm{e}^{-\mu t}\right)^{i-(j-k)} \\
\overset{l=k-1}{=} & \quad \rho\sum_{l=0}^{j-1}(1+2l+l^2)\frac{\rho^l\,\mathrm{e}^{-\rho}}{l!}\binom{i}{j-1-l}\left(\mathrm{e}^{-\mu t}\right)^{j-1-l}\left(1-\mathrm{e}^{-\mu t}\right)^{i-(j-1-l)} \\
= & \quad \rho p_{i(j-1)}(t;\theta) + 2\rho^2 p_{i(j-2)}(t;\theta) + \rho p_{i(j-1)}(t;\theta)_{k^2} \\
= & \quad \rho p_{i(j-1)}(t;\theta) + 3\rho^2 p_{i(j-2)}(t;\theta) + \rho^3 p_{i(j-3)}(t;\theta)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial p_{ij}(t;\theta)_{k^m}}{\partial \mu} &= \sum_{k=0}^{j} k^m \frac{\partial}{\partial \mu}\Big(f_{Poi(\rho)}(k) f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k)\Big) \\
&= \sum_{k=0}^{j} k^m \left(\frac{\rho\tau - (j - i\,\mathrm{e}^{-\mu t})\,\mu t}{(1-\mathrm{e}^{-\mu t})\mu} - k\frac{\tau-\mu t}{(1-\mathrm{e}^{-\mu t})\mu}\right) f_{Poi(\rho)}(k) f_{Bin(i,\mathrm{e}^{-\mu t})}(j-k) \\
&= \frac{\rho\tau - (j - i\,\mathrm{e}^{-\mu t})\,\mu t}{(1-\mathrm{e}^{-\mu t})\mu} p_{ij}(t;\theta)_{k^m} - \frac{\tau-\mu t}{(1-\mathrm{e}^{-\mu t})\mu} p_{ij}(t;\theta)_{k^{m+1}}
\end{aligned}
$$

$$
\left|\frac{1}{p_{ij}(t;\theta)}\frac{\partial p_{ij}(t;\theta)_{k^m}}{\partial \mu}\right| < \left(\alpha t^2 + \frac{(3j+i)t}{1-\mathrm{e}^{-\mu t}}\right) j^m,
$$

where $m = 0, 1, 2, \ldots$ and $p_{ij}(t;\theta)_{k^0} = p_{ij}(t;\theta)$, we get

$$
\begin{aligned}
&\frac{\partial^3 p_{ij}(t;\theta)}{\partial \mu^3} = \frac{\partial}{\partial \mu}\frac{\partial^2 p_{ij}(t;\theta)}{\partial \mu^2} \\
&= \left(\frac{6}{\mu^2} - \frac{8\tau\,\mathrm{e}^{-\mu t}}{(1-\mathrm{e}^{-\mu t})^2\mu} t - \frac{2(1+2\,\mathrm{e}^{-\mu t})\,\mathrm{e}^{-\mu t}}{(1-\mathrm{e}^{-\mu t})^2} t^2\right)\frac{\partial^2 p_{ij}(t;\theta)}{\partial \mu^2} \\
&\quad + \frac{\rho^2}{\alpha^2}\Bigg(p_{ij}(t;\theta)_{k^2}\frac{\partial}{\partial \mu}(\tau-\mu t)^2 + (\tau-\mu t)^2\frac{\partial}{\partial \mu} p_{ij}(t;\theta)_{k^2} \\
&\quad + \frac{\partial}{\partial \mu} p_{ij}(t;\theta)_k \Big(2\mu t\,(\tau-\mu t)\,(j - i\,\mathrm{e}^{-\mu t}) \\
&\quad - 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)\Big) \\
&\quad + p_{ij}(t;\theta)_k \frac{\partial}{\partial \mu}\Big(2\mu t\,(\tau-\mu t)\,(j - i\,\mathrm{e}^{-\mu t}) \\
&\quad - 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)\Big) \\
&\quad + \frac{\partial}{\partial \mu} p_{ij}(t;\theta)\Big(\left(j - i\,\mathrm{e}^{-\mu t}\right)^2 \mu^2 t^2 - 2\rho\tau\mu t\left(j - i\,\mathrm{e}^{-\mu t}\right) + (j-i)\mu^2 t^2\,\mathrm{e}^{-\mu t} \\
&\quad + \rho^2\tau^2 + \rho(1-\mathrm{e}^{-\mu t})(\mu^2 t^2\,\mathrm{e}^{-\mu t} - 2\tau)\Big) \\
&\quad + p_{ij}(t;\theta)\frac{\partial}{\partial \mu}\Big(\left(j - i\,\mathrm{e}^{-\mu t}\right)^2 \mu^2 t^2 - 2\rho\tau\mu t\left(j - i\,\mathrm{e}^{-\mu t}\right) + (j-i)\mu^2 t^2\,\mathrm{e}^{-\mu t} \\
&\quad + \rho^2\tau^2 + \rho(1-\mathrm{e}^{-\mu t})(\mu^2 t^2\,\mathrm{e}^{-\mu t} - 2\tau)\Big)\Bigg) \\
&= A_{\frac{\partial^2 p_{ij}(t;\theta)}{\partial \mu^2}}\frac{\partial^2 p_{ij}(t;\theta)}{\partial \mu^2} \\
&\quad + A_{p_{ij}(t;\theta)_{k^2}} p_{ij}(t;\theta)_{k^2} + A_{\frac{\partial}{\partial \mu} p_{ij}(t;\theta)_{k^2}}\frac{\partial}{\partial \mu} p_{ij}(t;\theta)_{k^2} + A_{\frac{\partial}{\partial \mu} p_{ij}(t;\theta)_k}\frac{\partial}{\partial \mu} p_{ij}(t;\theta)_k \\
&\quad + A_{p_{ij}(t;\theta)_k} p_{ij}(t;\theta)_k + A_{\frac{\partial}{\partial \mu} p_{ij}(t;\theta)}\frac{\partial}{\partial \mu} p_{ij}(t;\theta) + A_{p_{ij}(t;\theta)} p_{ij}(t;\theta)
\end{aligned}
$$

where

$$
\begin{aligned}
A_{\frac{\partial^2 p_{ij}(t;\theta)}{\partial \mu^2}} &= \left(\frac{6}{\mu^2} - \frac{8\tau\,\mathrm{e}^{-\mu t}}{(1-\mathrm{e}^{-\mu t})^2\mu}t - \frac{2(1+2\,\mathrm{e}^{-\mu t})\,\mathrm{e}^{-\mu t}}{(1-\mathrm{e}^{-\mu t})^2}t^2\right) \\
A_{p_{ij}(t;\theta)_{k^2}} &= \frac{\rho^2}{\alpha^2}2t(\tau-\mathrm{e}^{-\mu t})(\mu t-\tau) \\
A_{\frac{\partial}{\partial\mu}p_{ij}(t;\theta)_{k^2}} &= \frac{\rho^2}{\alpha^2}\left(\tau-\mu t\right)^2 \\
A_{\frac{\partial}{\partial\mu}p_{ij}(t;\theta)_{k}} &= \frac{\rho^2}{\alpha^2}\Big(2\mu t\left(\tau-\mu t\right)\left(j-i\,\mathrm{e}^{-\mu t}\right) \\
&\quad -2\rho\tau^2+(1-\mathrm{e}^{-\mu t})^2+2\rho\mu t(1-\mathrm{e}^{-\mu t})-2\mu^2t^2\,\mathrm{e}^{-\mu t}(1+\rho)\Big) \\
A_{p_{ij}(t;\theta)_{k}} &= 2\frac{\rho^2}{\alpha^2}\Bigg(t\left((j-i\,\mathrm{e}^{-\mu t})(\tau-\mu t-\mu t(\tau+e^{-\mu t}))+i\,\mathrm{e}^{-\mu t}\,\mu t(\tau-\mu t)\right) \\
&\quad + t\,\mathrm{e}^{-\mu t}\left(\tau-\mu t-\tau\mu t+\mu^2t^2(1-\mathrm{e}^{-\mu t})\right) \\
&\quad + \rho\frac{\left(\tau\left(\tau^2+(\mu t)^2\,\mathrm{e}^{-\mu t}(-1+2\,\mathrm{e}^{-\mu t})\right)+(\mu t)^3\,\mathrm{e}^{-\mu t}(1-\mathrm{e}^{-\mu t})\right)}{(1-\mathrm{e}^{-\mu t})\mu}\Bigg) \\
A_{\frac{\partial}{\partial\mu}p_{ij}(t;\theta)} &= \frac{\rho^2}{\alpha^2}\Big(\left(j-i\,\mathrm{e}^{-\mu t}\right)^2\mu^2t^2-2\rho\tau\mu t\left(j-i\,\mathrm{e}^{-\mu t}\right)+(j-i)\mu^2t^2\,\mathrm{e}^{-\mu t} \\
&\quad +\rho^2\tau^2+\rho(1-\mathrm{e}^{-\mu t})(\mu^2t^2\,\mathrm{e}^{-\mu t}-2\tau)\Big) \\
A_{p_{ij}(t;\theta)} &= \frac{\rho^2}{\alpha^2}\Bigg(2t(\mu t-\mathrm{e}^{-\mu t}(\alpha t\tau+\rho(\mu t)^2))(j-i\,\mathrm{e}^{-\mu t})+2\mu t^2 i\,\mathrm{e}^{-\mu t}(\mu t-\rho\tau-1) \\
&\quad + j(2-\mu t)\mu t^2\,\mathrm{e}^{-\mu t}+2\tau^3\frac{\rho(1-\rho)}{(1-\mathrm{e}^{-\mu t})\mu} \\
&\quad -\rho(\mu t)^2\,\mathrm{e}^{-\mu t}\,\frac{\tau^2(1-2\rho+\mu t)+2\tau\,\mathrm{e}^{-\mu t}(1+(\mu t)^2)}{(1-\mathrm{e}^{-\mu t})\mu} \\
&\quad +\rho(\mu t)^2\,\mathrm{e}^{-\mu t}\,\frac{(\mu t\,\mathrm{e}^{-\mu t})^2+\mu t\,\mathrm{e}^{-\mu t}(2\rho\tau-(\mu t)^2\,\mathrm{e}^{-\mu t})}{(1-\mathrm{e}^{-\mu t})\mu}\Bigg)
\end{aligned}
$$

$$
\frac{\partial^3\log p_{ij}(t;\theta)}{\partial\mu^3} = \frac{1}{p_{ij}(t;\theta)}\frac{\partial^3 p_{ij}(t;\theta)}{\partial\mu^3} - \frac{\partial\log p_{ij}(t;\theta)}{\partial\mu}\left(\left(\frac{\partial\log p_{ij}(t;\theta)}{\partial\mu}\right)^2+3\frac{\partial^2\log p_{ij}(t;\theta)}{\partial\mu^2}\right)
$$

$$
\begin{aligned}
\left|A_{\frac{\partial^2 p_{ij}(t;\theta)}{\partial\mu^2}}\right| &\leq \frac{6}{\mu^2}+\frac{8t^2}{\mu t}\frac{\tau\,\mathrm{e}^{-\mu t}}{(1-\mathrm{e}^{-\mu t})^2}+2(1+2\,\mathrm{e}^{-\mu t})\,\mathrm{e}^{-\mu t}\,t^2\frac{1}{(1-\mathrm{e}^{-\mu t})^2}\\
&< \frac{6}{\mu^2}+\frac{8t}{\mu}+\frac{6t^2}{(1-\mathrm{e}^{-\mu t})^2}\\
\left|A_{p_{ij}(t;\theta)_{k^2}}\right| &\leq 2t\,\overbrace{\frac{\rho^2}{\alpha^2}}^{<t^2}\,\overbrace{|\tau-\mathrm{e}^{-\mu t}|}^{<1}(\mu t+\overbrace{\tau}^{<1})<2t^3(\mu t+1)\\
\left|A_{\frac{\partial}{\partial\mu}p_{ij}(t;\theta)_{k^2}}\right| &= \overbrace{\frac{\rho^2}{\alpha^2}}^{<t^2}\,\overbrace{(\tau-\mu t)^2}^{<(1+\mu t)^2}<t^2(1+\mu t)^2\\
\left|A_{\frac{\partial}{\partial\mu}p_{ij}(t;\theta)_{k}}\right| &\leq \overbrace{\frac{\rho^2}{\alpha^2}}^{<t^2}\Big(2\mu t\,\overbrace{|\tau-\mu t|}^{<1+\mu t}\overbrace{|j-i\,\mathrm{e}^{-\mu t}|}^{<j+i}+2\,\overbrace{\rho\tau^2}^{<\alpha t}+\overbrace{(1-\mathrm{e}^{-\mu t})^2}^{<1}\\
&\quad +2\,\underbrace{\rho\mu t(1-\mathrm{e}^{-\mu t})}_{<\alpha t}+2\mu^2t^2\,\underbrace{\mathrm{e}^{-\mu t}(1+\rho)}_{<1+\alpha t}\Big)\\
&< t^2\left(2\mu t(j+i)+4\alpha t+1+2\mu^2t^2(1+\alpha t+j+i)\right)\\
\left|A_{p_{ij}(t;\theta)_k}\right| &< 2t^3\left((j+i)(1+3\mu t)+i\mu t(1+\mu t)+(1+\mu t)^2+\alpha\left(1+(\mu t)^2+\mu^2t^3\right)\right)\\
\left|A_{\frac{\partial}{\partial\mu}p_{ij}(t;\theta)}\right| &< t^2\left((j+i)^2\mu^2t^2+2\alpha t(j+i)+(j+i)\mu^2t^2+\alpha^2+\alpha t(\mu^2t^2+2)\right)\\
\left|A_{p_{ij}(t;\theta)}\right| &< t^3\Bigg(2(\mu t+(\alpha t+\alpha t(\mu t)^2))(j+i)+2\mu t i(\mu t+\alpha t+1)+j(2+\mu t)\mu t\\
&\quad +2\frac{\alpha(1+\alpha t)}{\mu^2 t}+\alpha\mu t^2\left(3+4\alpha t+2\mu t+3(\mu t)^2\right)\Bigg)
\end{aligned}
$$

$$
\begin{aligned}
&\frac{1}{p_{ij}(t;\theta)}\left|\frac{\partial^3 p_{ij}(t;\theta)}{\partial\mu^3}\right| \leq \\
\leq\quad & \frac{1}{p_{ij}(t;\theta)}\Bigg(\left|A_{\frac{\partial^2 p_{ij}(t;\theta)}{\partial\mu^2}}\right|\left|\frac{\partial^2 p_{ij}(t;\theta)}{\partial\mu^2}\right| + \left|A_{p_{ij}(t;\theta)_{k^2}}\right|\left|p_{ij}(t;\theta)_{k^2}\right| \\
& + \left|A_{\frac{\partial}{\partial\mu}p_{ij}(t;\theta)_{k^2}}\right|\left|\frac{\partial}{\partial\mu}p_{ij}(t;\theta)_{k^2}\right| + \left|A_{\frac{\partial}{\partial\mu}p_{ij}(t;\theta)_{k}}\right|\left|\frac{\partial}{\partial\mu}p_{ij}(t;\theta)_{k}\right| \\
& + \left|A_{p_{ij}(t;\theta)_{k}}\right|\left|p_{ij}(t;\theta)_{k}\right| + \left|A_{\frac{\partial}{\partial\mu}p_{ij}(t;\theta)}\right|\left|\frac{\partial}{\partial\mu}p_{ij}(t;\theta)\right| + \left|A_{p_{ij}(t;\theta)}\right|\left|p_{ij}(t;\theta)\right|\Bigg) \\
<\quad & \left(\frac{6}{\mu^2}+\frac{8t}{\mu}+\frac{6t^2}{(1-\mathrm{e}^{-\mu t})^2}\right)t^2\Big(j^2+2(j+i)j+\left(4\alpha t+1+2\mu^2t^2(1+\alpha t)\right)j \\
& +\alpha^2t^2+\alpha t(\mu^2t^2+2)+(j+i)^2\mu^2t^2+2\alpha t(j+i)+(j+i)\mu^2t^2\Big) \\
& +2t^3(\mu t+1)j^2 \\
& +t^2(1+\mu t)^2\left(\alpha t^2+\frac{(3j+i)t}{1-\mathrm{e}^{-\mu t}}\right)j^2 \\
& +t^2\left(2\mu t(j+i)+4\alpha t+1+2\mu^2t^2(1+\alpha t+j+i)\right)\left(\alpha t^2+\frac{(3j+i)t}{1-\mathrm{e}^{-\mu t}}\right)j \\
& +2t^3\left((j+i)(1+3\mu t)+i\mu t(1+\mu t)+(1+\mu t)^2+\alpha\left(1+(\mu t)^2+\mu^2t^3\right)\right)j \\
& +t^2\left((j+i)^2\mu^2t^2+2\alpha t(j+i)+(j+i)\mu^2t^2+\alpha^2+\alpha t(\mu^2t^2+2)\right)\left(\alpha t^2+\frac{(3j+i)t}{1-\mathrm{e}^{-\mu t}}\right) \\
& +t^3\Bigg(2(\mu t+(\alpha t+\alpha t(\mu t)^2))(j+i)+2\mu t i(\mu t+\alpha t+1)+j(2+\mu t)\mu t \\
& +2\frac{\alpha(1+\alpha t)}{\mu^2 t}+\alpha\mu t^2\left(3+4\alpha t+2\mu t+3(\mu t)^2\right)\Bigg) \\
=:\quad & A_{\frac{1}{p_{ij}(t;\theta)}\left|\frac{\partial^3 p_{ij}(t;\theta)}{\partial\mu^3}\right|}
\end{aligned}
$$

$$
\begin{aligned}
&\left|\frac{\partial^3 \log p_{ij}(t;\theta)}{\partial \mu^3}\right| \leq \\
&\leq \quad \frac{1}{p_{ij}(t;\theta)}\left|\frac{\partial^3 p_{ij}(t;\theta)}{\partial \mu^3}\right| + \left|\frac{\partial \log p_{ij}(t;\theta)}{\partial \mu}\right| \left(\left(\frac{\partial \log p_{ij}(t;\theta)}{\partial \mu}\right)^2 + 3\left|\frac{\partial^2 \log p_{ij}(t;\theta)}{\partial \mu^2}\right|\right) \\
&< \quad A_{\frac{1}{p_{ij}(t;\theta)}\left|\frac{\partial^3 p_{ij}(t;\theta)}{\partial \mu^3}\right|} + 4\left(\frac{\alpha t^2 + (3j+i)t}{1-\mathrm{e}^{-\mu t}}\right)^3 \\
&\quad + 3t^2\left(\frac{\alpha t^2 + (3j+i)t}{1-\mathrm{e}^{-\mu t}}\right)\Big(j^2 + 2(j+i)j + \big(2\alpha t + 1 + 2\alpha t + 2\mu^2 t^2(1+\alpha t)\big)\, j \\
&\quad + \alpha^2 t^2 + \alpha t(\mu^2 t^2 + 2) + (j+i)^2 \mu^2 t^2 + 2\alpha t(j+i) + (j+i)\mu^2 t^2\Big) \\
&=: \quad B_{222}(\alpha,\mu,t,j,i) \qquad \text{(A.18)}
\end{aligned}
$$

B Derivation of the Fisher information matrix

See Section 3.2 for definitions.

The Fisher information matrix at θ_0 associated with $\{q(\theta; i, \cdot) : \theta \in \Lambda_{\theta_0}\}$ is given by

$$
I(\theta_0; i) = \begin{pmatrix} I_{11}(\theta_0; i) & I_{12}(\theta_0; i) \\ I_{21}(\theta_0; i) & I_{22}(\theta_0; i) \end{pmatrix}.
$$

By expression (2.2) we have that

$$
\begin{aligned}
\sum_{j=0}^{\infty}(j - i\,\mathrm{e}^{-\mu t})p_{i(j-1)}(t;\theta_0) &= 1 + \mathbb{E}_{\theta_0}[N(s+t)|N(s)=i] - i\,\mathrm{e}^{-\mu_0 t} = 1 + \rho_0 \\
\sum_{j=0}^{\infty}(j - i\,\mathrm{e}^{-\mu t})p_{ij}(t;\theta_0) &= \mathbb{E}_{\theta_0}[N(s+t)|N(s)=i] - i\,\mathrm{e}^{-\mu_0 t} = \rho_0 \\
\sum_{j=0}^{\infty}(j - i\,\mathrm{e}^{-\mu t})^2 p_{ij}(t;\theta_0) &= \mathbb{E}_{\theta_0}[N(s+t)^2|N(s)=i] + i^2\,\mathrm{e}^{-2\mu_0 t} \\
&\quad -2i\,\mathrm{e}^{-\mu_0 t}\,\mathbb{E}_{\theta_0}[N(s+t)|N(s)=i] \\
&= (1-\mathrm{e}^{-\mu_0 t})i\,\mathrm{e}^{-\mu_0 t} + \rho_0^2 + \rho_0,
\end{aligned}
$$

where $\mathbb{E}_{\theta_0}[\cdot]$ denotes expected value under $\theta_0 = (\alpha_0, \mu_0)$. Using these results and by considering expressions (A.5), (A.8) and (A.12), we get that the entries of $I(\theta_0; i)$ are given by

$$
\begin{aligned}
I_{11}(\theta_0;i) &= \sum_{j\in E} \left(D_1 \log q(\theta_0;i,j)\right)^2 q(\theta_0;i,j) \qquad \text{(B.1)}\\
&= \sum_{j=0}^{\infty} \frac{\rho_0^2}{\alpha_0^2}\left(\frac{p_{i(j-1)}(t;\theta_0)}{p_{ij}(t;\theta_0)} - 1\right)^2 p_{ij}(t;\theta_0)\\
&= \frac{\rho_0^2}{\alpha_0^2}\left(\sum_{j=0}^{\infty}\frac{\left(p_{i(j-1)}(t;\theta_0)\right)^2}{p_{ij}(t;\theta_0)} - 2\sum_{j=0}^{\infty} p_{i(j-1)}(t;\theta_0) + \sum_{j=0}^{\infty} p_{ij}(t;\theta_0)\right)\\
&= \frac{\rho_0^2}{\alpha_0^2}\left(\sum_{j=0}^{\infty}\frac{\left(p_{i(j-1)}(t;\theta_0)\right)^2}{p_{ij}(t;\theta_0)} - 1\right),
\end{aligned}
$$

$$
\begin{aligned}
I_{12}(\theta_0;i) &= I_{21}(\theta_0;i) = \sum_{j\in E}\left(D_1 \log q(\theta_0;i,j)\right)\left(D_2 \log q(\theta_0;i,j)\right) q(\theta_0;i,j) \qquad \text{(B.2)}\\
&= \frac{\rho_0 t}{\mu_0}\sum_{j=0}^{\infty}\left(\frac{p_{i(j-1)}(t;\theta_0)}{p_{ij}(t;\theta_0)} - 1\right)\left(p_{i(j-1)}(t;\theta_0) - \frac{\left(j - i\,\mathrm{e}^{-\mu_0 t}\right)}{\rho_0} p_{ij}(t;\theta_0)\right)\\
&\quad - \frac{\rho_0\tau_0}{\mu_0^2}\sum_{j=0}^{\infty}\left(\frac{p_{i(j-1)}(t;\theta_0)}{p_{ij}(t;\theta_0)} - 1\right)^2 p_{ij}(t;\theta_0)\\
&= \frac{\rho_0 t}{\mu_0}\left(\sum_{j=0}^{\infty}\frac{(p_{i(j-1)}(t;\theta_0))^2}{p_{ij}(t;\theta_0)} - \sum_{j=0}^{\infty} p_{i(j-1)}(t;\theta_0)\right)\\
&\quad - \frac{t}{\mu_0}\sum_{j=0}^{\infty}\left(j - i\,\mathrm{e}^{-\mu_0 t}\right) p_{i(j-1)}(t;\theta_0)\\
&\quad + \frac{t}{\mu_0}\sum_{j=0}^{\infty}\left(j - i\,\mathrm{e}^{-\mu_0 t}\right) p_{ij}(t;\theta) + \frac{\rho_0\tau_0}{\mu_0^2}\sum_{j=0}^{\infty}\left(p_{i(j-1)}(t;\theta_0) - p_{ij}(t;\theta_0)\right)\\
&\quad - \frac{\rho_0\tau_0}{\mu_0^2}\left(\sum_{j=0}^{\infty}\frac{(p_{i(j-1)}(t;\theta_0))^2}{p_{ij}(t;\theta_0)} - \sum_{j=0}^{\infty} p_{i(j-1)}(t;\theta_0)\right)\\
&= \left(\frac{\rho_0 t}{\mu_0} - \frac{\rho_0\tau_0}{\mu_0^2}\right)\left(\sum_{j=0}^{\infty}\frac{(p_{i(j-1)}(t;\theta_0))^2}{p_{ij}(t;\theta_0)} - 1\right)\\
&\quad + \frac{t}{\mu_0}\left(\mathbb{E}_{\theta_0}[N(s+t)|N(s)=i] - i\,\mathrm{e}^{-\mu_0 t}\right)\\
&\quad - \frac{t}{\mu_0}\left(1 + \mathbb{E}_{\theta_0}[N(s+t)|N(s)=i] - i\,\mathrm{e}^{-\mu_0 t}\right)\\
&= \frac{\rho_0(\mu_0 t - \tau_0)}{\mu_0^2}\left(\sum_{j=0}^{\infty}\frac{(p_{i(j-1)}(t;\theta_0))^2}{p_{ij}(t;\theta_0)} - 1\right) - \frac{t}{\mu_0},
\end{aligned}
$$

$$I_{22}(\theta_0;i) = \sum_{j\in E} \left(D_2 \log q(\theta_0;i,j)\right)^2 q(\theta_0;i,j) \tag{B.3}$$

$$\begin{aligned}
&= \frac{\rho_0^2\tau_0^2}{(1-\mathrm{e}^{-\mu_0 t})^2\mu_0^2}\sum_{j=0}^{\infty}\left(\frac{p_{i(j-1)}(t;\theta_0)}{p_{ij}(t;\theta_0)}-1\right)^2 p_{ij}(t;\theta_0) \\
&\quad + \frac{\rho_0^2(\mu_0 t)^2}{(1-\mathrm{e}^{-\mu_0 t})^2\mu_0^2}\sum_{j=0}^{\infty}\left(\frac{p_{i(j-1)}(t;\theta_0)}{p_{ij}(t;\theta_0)}-\frac{j-i\,\mathrm{e}^{-\mu_0 t}}{\rho_0}\right)^2 p_{ij}(t;\theta_0) \\
&\quad - \frac{2\rho_0^2\tau_0\mu_0 t}{(1-\mathrm{e}^{-\mu_0 t})^2\mu_0^2}\sum_{j=0}^{\infty}\left(\frac{p_{i(j-1)}(t;\theta_0)}{p_{ij}(t;\theta_0)}-1\right)\left(\frac{p_{i(j-1)}(t;\theta_0)}{p_{ij}(t;\theta_0)}-\frac{j-i\,\mathrm{e}^{-\mu_0 t}}{\rho_0}\right) p_{ij}(t;\theta_0) \\
&= \frac{\alpha_0^2\tau_0^2}{\mu_0^4}\left(\sum_{j=0}^{\infty}\frac{(p_{i(j-1)}(t;\theta_0))^2}{p_{ij}(t;\theta_0)}-1\right) \\
&\quad + \frac{\alpha_0^2(\mu_0 t)^2}{\mu_0^4}\left(\sum_{j=0}^{\infty}\frac{(p_{i(j-1)}(t;\theta_0))^2}{p_{ij}(t;\theta_0)} - \underbrace{\frac{2}{\rho_0}\sum_{j=0}^{\infty}(j-i\,\mathrm{e}^{-\mu_0 t})p_{i(j-1)}(t;\theta_0)}_{=1+(1+2/\rho_0)}\right) \\
&\quad + \frac{\alpha_0^2(\mu_0 t)^2}{\mu_0^4}\frac{1}{\rho_0^2}\sum_{j=0}^{\infty}(j-i\,\mathrm{e}^{-\mu_0 t})^2 p_{ij}(t;\theta_0) \\
&\quad - \frac{2\alpha_0^2\tau_0\mu_0 t}{\mu_0^4}\left(\sum_{j=0}^{\infty}\frac{(p_{i(j-1)}(t;\theta_0))^2}{p_{ij}(t;\theta_0)}-1\right) \\
&\quad + \frac{2\alpha_0^2\tau_0\mu_0 t}{\mu_0^4}\frac{1}{\rho_0}\left(\sum_{j=0}^{\infty}(j-i\,\mathrm{e}^{-\mu_0 t})p_{i(j-1)}(t;\theta_0) - \sum_{j=0}^{\infty}(j-i\,\mathrm{e}^{-\mu_0 t})p_{ij}(t;\theta_0)\right) \\
&= \frac{\alpha_0^2\left(\tau_0^2-2\tau_0\mu_0 t+(\mu_0 t)^2\right)}{\mu_0^4}\left(\sum_{j=0}^{\infty}\frac{(p_{i(j-1)}(t;\theta_0))^2}{p_{ij}(t;\theta_0)}-1\right) + \frac{2\alpha_0^2\tau_0\mu_0 t}{\mu_0^4}\frac{1}{\rho_0} \\
&\quad + \frac{\alpha_0^2(\mu_0 t)^2}{\mu_0^4}\left(\frac{(1-\mathrm{e}^{-\mu_0 t})i\,\mathrm{e}^{-\mu_0 t}+\rho_0(\rho_0+1)}{\rho_0^2}-1-\frac{2}{\rho_0}\right) \\
&= \frac{\alpha_0^2\left(\tau_0-\mu_0 t\right)^2}{\mu_0^4}\left(\sum_{j=0}^{\infty}\frac{(p_{i(j-1)}(t;\theta_0))^2}{p_{ij}(t;\theta_0)}-1\right) + \frac{\alpha_0 t^2\,\mathrm{e}^{-\mu_0 t}}{\mu_0\rho_0}i + \frac{\alpha_0^2\mu_0 t(2\tau_0-\mu_0 t)}{\rho_0\mu_0^4}.
\end{aligned}$$