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, 12]; see [9] 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 [8]; see [7] 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 [15], [16] and [4]), which has been used to study, among other things, the development of forest stands in time and space [16]. 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. [10] and [17] for general treatments of spatial point process statistics and e.g. [6], [14], and [18] 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). In the Appendix we give proofs of some results, derivatives of the (log) transition probabilities together with their bounds and the derivation of the Fisher information matrix.

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 $\mu_i = i\mu$, $i = 0, 1, \ldots$, ([8], 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, for which the proof is given in Appendix A.

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

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

We will make use of the following recursive expression for the transition probabilities. Its proof can be found in Appendix A.

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

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

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*

$$\lambda(\theta; i) = \alpha + \mu i, \quad i \in E,$$

and transition kernel $r(\theta; \cdot) = \{r(\theta; i, j) : i, j \in E\}$, where

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

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 ([11], 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 $\tau_k - \tau_{k-1} \stackrel{d}{=} \min\{\tau_\alpha, \xi_1, \ldots, \xi_i\}$ for $i \in \mathbb{Z}_+$, and clearly $\tau_k - \tau_{k-1} \stackrel{d}{=} \tau_\alpha$ if $i = 0$. Since the minimum of n independent exponential random variables with parameters $\lambda_1, \ldots, \lambda_n$ is exponentially distributed with parameter $\sum_{i=1}^{n} \lambda_i$ this implies that the jump intensity function is given by

$$\lambda(\theta; i) = \Big(\mathbb{E}_\theta[\tau_k - \tau_{k-1} | N(\tau_{k-1}) = i]\Big)^{-1} = \alpha + \mu i, \quad i \in E,$$

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

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

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, since $|N(\tau_k) - N(\tau_{k-1})| = 1$ for all $k = 1, 2, \ldots$, the transition kernel of the Markov jump process, $r(\theta; \cdot) = \{r(\theta; i, j) : i, j \in E\}$, is determined by

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

□

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 (B.1) and (B.2) in Appendix B. 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

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

□

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 $q(\theta;\cdot) = \{q(\theta;i,j) : i,j \in E\} = \{\mathbb{P}(X(T_n) = j | X(T_{n-1}) = i;\theta) : i,j \in E\}$. 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 $N_n(i,j) = \sum_{k=1}^{n} \mathbf{1}\left\{(Z_{k-1}, Z_k) = (i,j)\right\}$. 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) $\max\left\{|D_u \log q(\theta_0; i,j)|, |D^2_{uv} \log q(\theta_0; i,j)|\right\} \le C(1 + |i|^{a/2} + |j|^{a/2})$;

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

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

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

(ii)

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

(N3) The matrix $I(\theta_0; i) = (I_{uv}(\theta_0; i))_{u,v=1,\ldots,p}$ 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 $\{N(t)\}_{t \ge 0}$ 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. We have that $(\alpha, \mu) = \theta \in \Theta$ where Θ is a compact subset of $\mathbb{R}_+^2$, hence (G3) holds. 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 $\pi_{\theta_0}(\cdot) = \mathbb{P}(Poi(\alpha_0/\mu_0) \in \cdot)$ 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 [13]) condition (G2) will be fulfilled.

By Proposition 3 the Markov jump process $\{N(t)\}_{t\geq 0}$ has intensity $\lambda(\theta; i) = \alpha + \mu i$, $i \in E$, 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 ([11], 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 (B.3) and (B.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

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

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 $\theta\in\Theta$, $G_i\left(s;\theta\right)$, given by (A.2). If $G_i\left(s;\theta\right)\neq G_i\left(s;\theta_0\right)$, 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 $\mu=\mu_0$:

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

   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 $\mu \neq \mu_0$:

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

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 (B.3), (B.4), (B.6), (B.10) and (B.13) in Appendix B 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 $\Theta_1, \Theta_2, \alpha_{min}, \alpha_{max}, \mu_{min}$ and μ_{max})

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

$$\begin{aligned}\max_{(i,j)\in\{0,1\}^2}|D_{11}^2\log q(\theta_0;i,j)| &< \max_{(i,j)\in\{0,1\}^2}\sup_{\alpha\in\Theta_1}\frac{j+2(j+\alpha t)^2}{\alpha^2}\\ &< \frac{1+2(1+\alpha_{max}t)^2}{\alpha_{min}^2} =: C_{11}<\infty,\end{aligned}$$

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

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

so that by choosing $C=\max\{C_1,C_2,C_{11},C_{12},C_{22}\}$ we have that

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

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

$$\begin{aligned}\frac{\left|D_{uv}^2\log q(\theta;i,j)-D_{uv}^2\log q(\theta_0;i,j)\right|}{|\theta-\theta_0|} &\le \left|\nabla D_{uv}^2\log q\left((1-c)\theta+c\theta_0;i,j\right)\right|\\ &\le |D_1D_{uv}^2\log q((1-c)\theta+c\theta_0;i,j)|\\ &\quad+|D_2D_{uv}^2\log q((1-c)\theta+c\theta_0;i,j)|\end{aligned}$$

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 (B.15), (B.16), (B.17) and (B.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

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

and

$$\sum_{j\in E}\left(D_2\log q(\theta_0;i,j)\right)q(\theta_0;i,j) =$$

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

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

Considering expressions (C.1), (C.2) and (C.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

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

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

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

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.

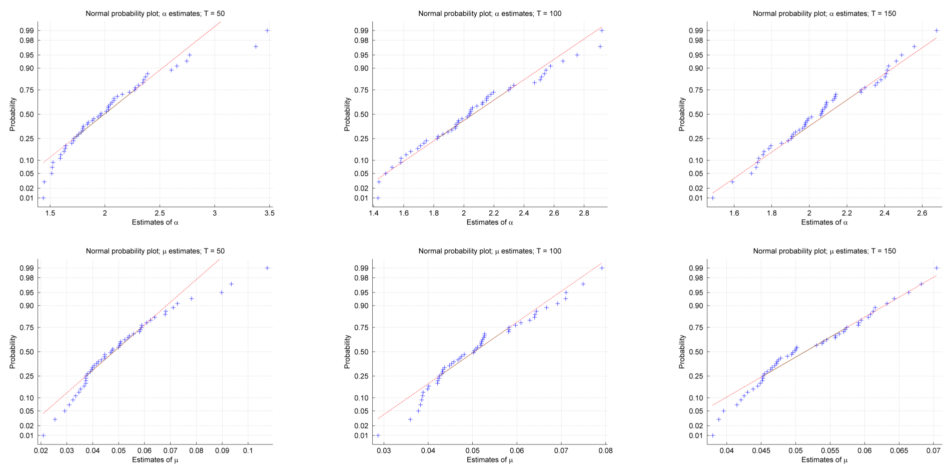

Figure 1: Normal probability plots of the estimates of $(\alpha_0, \mu_0) = (2, 0.05)$ based on 50 sample paths sampled at times $T_k = kt$, $t = 1$, $k = 1, \ldots, T$. Upper row: The estimates of α_0 at final times $T = 50$ (left), $T = 100$ (middle) and $T = 150$ (right). Lower row: The estimates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

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

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.

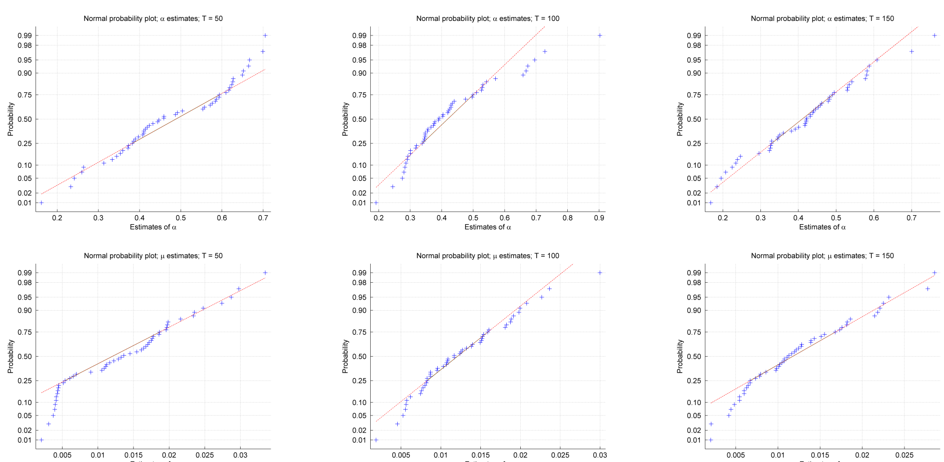

Figure 2: Normal probability plots of the estimates of $(\alpha_0, \mu_0) = (0.4, 0.01)$ based on 50 sample paths sampled at times $T_k = kt$, $t = 1$, $k = 1, \ldots, T$. Upper row: The estimates of α_0 at final times $T = 50$ (left), $T = 100$ (middle) and $T = 150$ (right). Lower row: The estimates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

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

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 $T_k = kt$, $t = 1$, $k = 1, \dots, T$.

	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

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, 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 [16]. 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 ([16] 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

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

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 $\Omega_t = \{i \in \{1, \ldots, B(t)\} : \text{individual } i \text{ is alive at time } t\}$, 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 [16] 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 $B\left[\mathbf{X}_i, rm_i(t)\right] \cap B\left[\mathbf{X}_j, rm_j(t)\right] \neq \emptyset$.

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 [16] 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 [16] 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

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

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 [16] 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 [16] is given by

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

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 [16] 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 [16] 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 [16]). 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 $\tilde{m}_i\left(T_{k+1}; \psi, \mathbb{X}(T_k)\right) \leq 0$ for an individual $i \in \Omega_{T_k}^{obs}$, it will be labelled as having died from competition in (T_k, T_{k+1}) 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 [16]) 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

The authors would like to thank Aila Särkkä (Chalmers University of Technology), Bo Ranneby and Lennart Norell (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] Finkenstadt, B., Held, L., Isham, V. *Statistical Methods for Spatio-Temporal Systems.* Chapman & Hall/CRC, 2006.

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

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

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

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

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

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

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

[14] 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.

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

Appendix

A Proofs

Below we give two different proofs of Proposition 1 and then the proof of Corollary 1.

Proof of Proposition 1. 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

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

Hence, one possible way of finding $p_{ij}(t;\theta) = \mathbb{P}(N(h+t) = j|N(h) = i)$, $h \geq 0$, is to evaluate expression (A.1) 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 (see [16] or [8], p. 299)

$$\begin{aligned} G(s) &= \left(1 + (s-1)\,\mathrm{e}^{-\mu t}\right)^i \exp\left\{(\alpha/\mu)(s-1)\left(1 - \mathrm{e}^{-\mu t}\right)\right\} \\ &= \left(1 + (s-1)\,\mathrm{e}^{-\mu t}\right)^i \mathrm{e}^{\rho(s-1)}, \end{aligned} \tag{A.2}$$

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

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

This suggests that

$$G^{(j)}(s) = G(s)\sum_{k=0}^{j} \rho^k \binom{j}{k} \frac{1}{(\mathrm{e}^{\mu t} - 1 + s)^{j-k}} \frac{i!}{(i - (j-k))!} \tag{A.4}$$

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 (A.4) by induction. Assume that (A.4) holds for j and let $a(s) = \mathrm{e}^{\mu t} - 1 + s$. It follows from (A.3) and (A.4) 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 (A.4) 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

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

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

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

□

*Proof * of Proposition 1.*

Notice first that for any fixed $t > 0$, $N(t)$ is the result of applying so called *p-thinning* (see e.g. [17]) to a Poisson process with intensity α, using thinning probability $1 - p(t)$, given $N(0) = 0$. Since an individual's arrival time, conditioned on the individual's arrival during $(0, t]$, is uniformly distributed on $(0, t]$ and its life-time is $Exp(\mu)$-distributed we get that

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

By the properties of thinned Poisson processes (see e.g. [17]) we have that $N(t) \sim Poi(\alpha t p(t)) = Poi(\rho)$, $\rho = \frac{\alpha}{\mu}\left(1 - \mathrm{e}^{-\mu t}\right)$.

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

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

In the same spirit we finally get that the first two moments of $(N(h +$

$t)|N(h)=i)$ are given by

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

Since the independent random variables X and Y have probability generating functions $G_X(t) = \left(1+(s-1)\,\mathrm{e}^{-\mu t}\right)^i$ and $G_Y(t) = \mathrm{e}^{\rho(s-1)}$, respectively, we find the probability generating function to be

$$\begin{aligned}
G_i\,(s;\theta) &= \mathbb{E}\left[\mathrm{e}^{sN(h+t)}\,\middle|\,N(h)=i\right] = \mathbb{E}\left[\mathrm{e}^{s(X+Y)}\right] = G_X(t)G_Y(t) \\
&= \left(1+(s-1)\,\mathrm{e}^{-\mu t}\right)^i \mathrm{e}^{\rho(s-1)}.
\end{aligned}$$

□

Proof of Corollary 1. 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

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

□

B 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

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

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))

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

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

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

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

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

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

Note first that $\rho = \alpha t \frac{1-\mathrm{e}^{-\mu t}}{\mu t} < \alpha t$, $\tau < \mu t$, $\tau < \mu^2 t^2$, $0 < \tau < 1$, $p_{ij}(t;\theta)_k \le j$ and $p_{ij}(t;\theta)_{k^2} \le j^2$ since $k \le j$ 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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$
\begin{aligned}
\left(\frac{\partial \log p_{ij}(t;\theta)}{\partial \mu}\right)^2 &= \frac{\rho^2\left(\tau - (j - i\,\mathrm{e}^{-\mu t})\,\mu t/\rho - (\tau - \mu t)\frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)}\right)^2}{(1-\mathrm{e}^{-\mu t})^2\mu^2} \\
&= \frac{\rho^2\left(\tau\left(1 - \frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)}\right) + \mu t\left(\frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)} - \frac{j-i\,\mathrm{e}^{-\mu t}}{\rho}\right)\right)^2}{(1-\mathrm{e}^{-\mu t})^2\mu^2} \\
&= \frac{\rho^2\tau^2\left(1 - \frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)}\right)^2}{(1-\mathrm{e}^{-\mu t})^2\mu^2} + \frac{\rho^2(\mu t)^2\left(\frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)} - \frac{j-i\,\mathrm{e}^{-\mu t}}{\rho}\right)^2}{(1-\mathrm{e}^{-\mu t})^2\mu^2} \\
&\quad + \frac{2\rho^2\tau\mu t\left(1 - \frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)}\right)\left(\frac{p_{i(j-1)}(t;\theta)}{p_{ij}(t;\theta)} - \frac{j-i\,\mathrm{e}^{-\mu t}}{\rho}\right)}{(1-\mathrm{e}^{-\mu t})^2\mu^2} \qquad \text{(B.12)}
\end{aligned}
$$

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

$$
\begin{aligned}
\left|\frac{\partial^2 p_{ij}(t;\theta)}{\partial \mu^2}\right| &< t\frac{(1-\mathrm{e}^{-\mu t})^2}{(\mu t)^2}\Bigg(\mu^2 t^2 j^2 \overbrace{\frac{(\mu t - \tau)^2}{(\mu t)^2}}^{<1} + 2\mu^2 t^2 j(j+i)\overbrace{\frac{\mu t - \tau}{\mu t}}^{<1} \\
&\quad + \Big(\overbrace{2\rho\tau^2}^{<\alpha t} + \overbrace{(1-\mathrm{e}^{-\mu t})^2}^{<1} + \overbrace{2\rho\mu t(1-\mathrm{e}^{-\mu t})}^{<2\alpha t} + \overbrace{2\mu^2 t^2\,\mathrm{e}^{-\mu t}(1+\rho)}^{<2(1+\alpha t)}\Big) j \\
&\quad + \overbrace{\rho^2\tau^2}^{<1} + \overbrace{\rho(1-\mathrm{e}^{-\mu t})}^{<\alpha t}\overbrace{(2\tau - \mu^2 t^2\,\mathrm{e}^{-\mu t})}^{<2} \\
&\quad + (j+i)^2\mu^2 t^2 + 2(j+i)\overbrace{\rho\tau\mu t}^{<\alpha t} + (j+i)\overbrace{\mu^2 t^2\,\mathrm{e}^{-\mu t}}^{<1}\Bigg) \\
&< t\Big(\mu^2 t^2\left(j^2 + 2j(j+i) + (j+i)^2\right) + \alpha t\,(7j + 2i + 2) + 4j + i + 1\Big)
\end{aligned}
$$

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

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

$$
\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}\Big(\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}\Big)\,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
\end{aligned}
\tag{B.14}
$$

B.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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Since

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

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

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

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

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}
$$

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

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

$$
\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 + \big(4\alpha t + 1 + 2\mu^2t^2(1+\alpha t)\big)\,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}
$$

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

C 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 (B.5), (B.8) and (B.12), we get that the entries of $I(\theta_0;i)$ are given by

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

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

$$
\begin{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) \qquad \text{(C.3)}\\
&= \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}
$$