



# INAR approximation of bivariate linear birth and death process

Zezhun Chen<sup>1</sup> · Angelos Dassios<sup>1</sup> · George Tzougas<sup>2</sup>

Received: 19 April 2022 / Accepted: 12 April 2023  
© The Author(s) 2023

## Abstract

In this paper, we propose a new type of univariate and bivariate Integer-valued autoregressive model of order one (INAR(1)) to approximate univariate and bivariate linear birth and death process with constant rates. Under a specific parametric setting, the dynamic of transition probabilities and probability generating function of INAR(1) will converge to that of birth and death process as the length of subintervals goes to 0. Due to the simplicity of Markov structure, maximum likelihood estimation is feasible for INAR(1) model, which is not the case for bivariate and multivariate birth and death process. This means that the statistical inference of bivariate birth and death process can be achieved via the maximum likelihood estimation of a bivariate INAR(1) model.

**Keywords** Bivariate birth and death · Linear birth and death · Integer-valued autoregressive of order one · Convergence in distribution · Discrete approximation

## 1 Introduction

The simple linear birth and death process, which was first introduced by Feller (1939), is a widely used Markov model with applications in population growth, epidemiology, genetics and so on. The basic idea of this process is that the probabilities of any individual giving birth to a new individual, or any individual dying, are constant at any moment in time and all individuals are independent of each other. Many statistical properties, including moments,

---

Angelos Dassios and George Tzougas have contributed equally to this work.

---

✉ Zezhun Chen  
z.chen58@lse.ac.uk

Angelos Dassios  
a.dassios@lse.ac.uk

George Tzougas  
george.tzougas@hw.ac.uk

<sup>1</sup> Department of Statistics, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, England, UK

<sup>2</sup> Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Campus The Avenue, City EH14 4AS, Scotland, UK

distribution function, extinction probability, or some other cumulative distribution of interests, are explicitly derived in the literature; see for example, Kendall (1949). The statistical inference for simple birth and death processes is then developed by Keiding (1975), where maximum likelihood estimators and other asymptotic results are discussed. Since the distribution function of simple birth and death processes is explicit, the construction of the likelihood function is straightforward. However, it is pointed out in the literature that the transition probability is actually cumbersome and numerically unstable when the size of population is large over time. At the same time, a variety of alternative estimation methods have been proposed. For example, quasi- and pseudo - likelihood estimators (Chen and Hyrien 2011; Crawford et al. 2014) addressed it as a missing data problem and apply an EM algorithm to maximize it. Tavaré (2018) found those transition probabilities by numerical inversion of the probability generating function and then applied Bayesian methods to perform estimation. Davison et al. (2021) adopted a saddle point approximation method to further improve the accuracy of transition probabilities.

The bivariate and multivariate birth and death process are developed in Griffiths (1972, 1973). Griffiths (1972) described the transmission of malaria (so called host-vector situation) as a bivariate birth and death process where there is no direct infection between the same type of population. Then the author extended the model to multivariate case (Griffiths 1973) which can be regarded as an approximation of general epidemic with several types of infective. However, due to the intractability of the joint probability generating function, maximum likelihood estimation for parameters is not implementable. One possible way forward is to use integer-valued time series to approximate the continuous birth and death process and maximum likelihood estimation would then be feasible.

In recent years, there has been a growing interest in modelling integer-valued time series due to the presence of count data from different scientific fields such as social science, healthcare, insurance, economic and the financial industry. In particular, regarding to the univariate case, Al-Osh and Alzaid (1987) and McKenzie (1985) were the first to consider an INAR(1) model based on the so-called binomial thinning operator. The idea here is to manipulate the operation between coefficients and variables as well as the innovation terms in a way that the values are always integer. One can apply different discrete random variables to describe this operation. For more details, the interested reader can refer to Weiß (2018), Davis et al. (2016), Scotto et al. (2015), Weiß (2008) among many more.

In this paper, we propose an integer-valued autoregressive model of order one (INAR(1)) to approximate continuous birth and death process. In this way, the continuous process is approximated by a discrete Markov chain so that transition probabilities as well as likelihood function can be written down explicitly. As the birth and death process in our setting does not consider any immigrant, the innovation term is dropped in the proposed INAR(1) model. Similar to Nelson (1990), Kirchner (2016), where they find out the relationship between discrete models and their continuous counterparts, we also first need to make sure the our proposed discrete INAR(1) model would converge to birth and death process in weak convergence sense. Then we will explore how our proposed model would help facilitate the statistical inference. According to the probability generating function of the simple birth and death process, the death part can be described by binomial random variable while the birth part corresponds to a negative binomial. Then one can construct a bivariate INAR model based on these random variables to describe the bivariate birth and death process and even the multivariate one. As the transition probabilities and likelihood function of bivariate birth and death process cannot be written down explicitly, the main contribution is that the proposed bivariate INAR(1) model would provide a feasible way to estimate the parameters of bivariate birth and death process (Maximum likelihood estimation).

The paper is organized as follows: Sect. 2 reviews some main results of univariate and bivariate birth and death processes with constant rates. Section 3 introduces Integer-valued autoregressive models as well as some distributional properties. Section 4 constructs the discrete semimartingale using the proposed INAR models and proves the weak convergence between constructed semimartingale and birth and death processes. A simulation study is carried out in Sect. 5 to illustrate the estimation method via proposed INAR models an their corresponding properties of estimators. Some concluding remarks are in Sect. 6.

## 2 Univariate and bivariate birth and death processes

In this section, we will review the essential elements of simple birth-and-death processes, including moments and other distributional properties. These are well known and extensively discussed in the literature. Then, we will discuss the bivariate case where analytic expressions of the distribution function are not available.

### 2.1 Simple univariate birth-and-death process

Suppose that we have a population whose total number is evolved as a simple birth and death process  $Z_t$ , with constant birth rate  $\lambda \geq 0$ , death rate  $\mu \geq 0$  and initial population  $Z_0 \in \mathbb{N}$ . In other words, the probability that any individual gives birth in time  $\Delta$  is  $\lambda\Delta$ , and the probability that any individual dies in time  $\Delta$  is  $\mu\Delta$ . Individuals are independent of each other. Let  $P_n(t) = \Pr(Z_t = n)$  be the probability that the total population is  $n$  at time  $t$ . Then the transition probability of the simple birth and death process is characterized by the following ordinary differential equation (ODE)

$$\begin{cases} \frac{dP_n(t)}{dt} = \lambda(n-1)P_{n-1}(t) + \mu(n+1)P_{n+1}(t) - (\lambda + \mu)nP_n(t), & n \geq 1 \\ P_{Z_0}(0) = 1 \end{cases} \tag{1}$$

Applying a liner transform  $\sum_n \theta^n$  on both sides and defining  $\varphi(t, \theta) = \sum_n \theta^n P_n(t)$ , we can get a partial differential equation whose solution  $\varphi$  is the probability generating function of  $Z_t^{(a)}$ .

$$\begin{aligned} \frac{\partial \varphi}{\partial t} &= \lambda \theta^2 \frac{\partial \varphi}{\partial \theta} + \mu \frac{\partial \varphi}{\partial \theta} - (\lambda + \mu) \theta \frac{\partial \varphi}{\partial \theta} \\ &= (\lambda \theta - \mu)(\theta - 1) \frac{\partial \varphi}{\partial \theta} \\ \varphi(0, \theta) &= \theta^a \end{aligned} \tag{2}$$

This linear PDE can be solved explicitly

$$\begin{aligned} \varphi(t, \theta) &= \left( 1 - \alpha(t) + \alpha(t) \frac{\beta(t)\theta}{1 - (1 - \beta(t))\theta} \right)^{Z_0} \\ \alpha(t) &= \frac{(\lambda - \mu)e^{(\lambda - \mu)t}}{\lambda e^{(\lambda - \mu)t} - \mu}, \quad \beta(t) = \frac{\lambda - \mu}{\lambda e^{(\lambda - \mu)t} - \mu} \end{aligned} \tag{3}$$

This probability generating function clearly gives the construction of  $Z_t$  given  $Z_0$ , i.e. the sum of i.i.d zero-modified geometric random variables

$$Z_t \sim \sum_{i=1}^{Z_0} B_i(\alpha(t))G_i(\beta(t)), \tag{4}$$

where  $B_i$  are i.i.d Bernoulli random variables and  $G_i$  are i.i.d Geometric random variables with mean  $\alpha(t)$  and  $\frac{1}{\beta(t)}$ , respectively. Furthermore, from the definition of transition probability, the linear birth and death process is a pure-jump semimartingale with following characteristic triplet:

$$Ch(Z_t) = \begin{cases} B_t = 0 \\ C_t = 0 \\ v(Z_t; dt, dx) = dt K(Z_t, dx) = dt(\lambda Z_t \delta_1(dx) + \mu Z_t \delta_{-1}(dx)) \end{cases} \quad (5)$$

$$\int_{\mathbb{R}} (x^2 \wedge 1) K(Z_t, dx) = (\lambda + \mu)Z_t < \infty, \quad \text{given that } Z_t \text{ is finite}$$

With the help of piece-wise deterministic Markov process theory in Davis (1984), the infinitesimal generator of the simple birth and death process  $Z_t$  acting on a function  $f(t, Z)$  within its domain  $\Omega(\mathcal{A})$  is given by

$$\mathcal{A}f(t, Z) = \frac{\partial f}{\partial t} + \lambda Z(f(t, Z + 1) - f(t, Z)) + \mu Z(f(t, Z - 1) - f(t, Z)), \quad (6)$$

where  $\Omega(\mathcal{A})$  is the domain for the generator  $\mathcal{A}$  such that  $f(t, Z)$  is differentiable with respect to  $t$  for all  $t, Z$ , and

$$\begin{aligned} |f(t, Z + 1) - f(t, Z)| < \infty \\ |f(t, Z - 1) - f(t, Z)| < \infty. \end{aligned} \quad (7)$$

The first and second moments can be derived by applying infinitesimal generator to the functions  $f(t, z) = Z, Z^2$  such that

$$\begin{aligned} AZ &= \lambda Z(Z + 1 - Z) + \mu Z(Z - 1 - Z) \\ AZ^2 &= \lambda Z((Z + 1)^2 - Z^2) + \mu Z((Z - 1)^2 - Z^2), \end{aligned} \quad (8)$$

which leads to two ODEs,

$$\begin{aligned} \frac{d\mathbb{E}[Z_t]}{dt} &= (\lambda - \mu)\mathbb{E}[Z_t] \\ \frac{d\mathbb{E}[Z_t^2]}{dt} &= 2(\lambda - \mu)\mathbb{E}[Z_t^2] + (\lambda + \mu)\mathbb{E}[Z_t] \end{aligned} \quad (9)$$

Then, we can solve them explicitly

$$\begin{aligned} \mathbb{E}[Z_t] &= Z_0 e^{(\lambda - \mu)t} \\ \mathbb{E}[Z_t^2] &= Z_0^2 e^{2(\lambda - \mu)t} + \frac{Z_0(\lambda + \mu)}{(\lambda - \mu)} e^{(\lambda - \mu)t} (e^{(\lambda - \mu)t} - 1) \\ Var(Z_t) &= \frac{Z_0(\lambda + \mu)}{(\lambda - \mu)} e^{(\lambda - \mu)t} (e^{(\lambda - \mu)t} - 1) \end{aligned} \quad (10)$$

According to the analytic expression of the first moment, it is clear that the population is bound to become extinct if  $\lambda < \mu$ .

### 2.2 Bivariate birth-and-death process

Suppose there are two populations  $\mathbf{M} = (M_1, M_2)^T$  with initial population  $\mathbf{M}_0 \in \mathbb{N}_+^2$ . The rate with which the population  $M_1$  increases by one is  $\lambda_{21}M_2 + \lambda_{11}M_1$  while the same for the population  $M_2$  would be  $\lambda_{12}M_1 + \lambda_{22}M_2$ . The subscript  $\lambda_{i,j}$  means that the rate is

from population  $i$  contributed to population  $j$ . The death rate for two populations would be  $\mu_1, \mu_2$  respectively. The two population is not independent as long as the cross birth rates  $\lambda_{i,j} \neq 0, i \neq j$ . Then denote  $P_{mn}(t) = \Pr(M_{1,t} = m, M_{2,t} = n)$ . This satisfies the following ODE

$$\begin{cases} \frac{dP_{m,n}}{dt} = (\lambda_{11}(m-1) + \lambda_{21}n) P_{m-1,n} + \mu_1(m+1) P_{m+1,n} \\ \quad + (\lambda_{12}m + \lambda_{22}(n-1)) P_{m,n-1} + \mu_2(n+1) P_{m,n+1} \\ \quad - ((\lambda_{11} + \lambda_{12} + \mu_1)m + (\lambda_{21} + \lambda_{22} + \mu_2)n) P_{m,n} \\ P_{M_0}(0) = 1, \quad M_{1,0}, M_{2,0} \in \mathbb{N}_+ \end{cases} \tag{11}$$

Griffiths (1972) introduced this bivariate birth death process ( $\lambda_{11} = \lambda_{22} = 0$ ) to describe the host-vector epidemic situation where the birth probability of two population depends on the size the other population only, e.g. transmission of malaria. To get the joint probability generating function of  $\Psi(t, \theta, \phi) = \sum_m \sum_n \theta^m \phi^n P_{mn}(t)$ , we can apply a linear transform  $\sum_m \sum_n \theta^m \phi^n$  on both sides of the ODE. The resulting PDE is

$$\begin{aligned} \frac{\partial \Psi}{\partial t} &= \lambda_{11} \theta^2 \frac{\partial \Psi}{\partial \theta} + \lambda_{21} \theta \phi \frac{\partial \Psi}{\partial \phi} + \mu_1 \frac{\partial \Psi}{\partial \theta} + \lambda_{12} \theta \phi \frac{\partial \Psi}{\partial \theta} + \lambda_{22} \phi^2 \frac{\partial \Psi}{\partial \phi} + \mu_2 \frac{\partial \Psi}{\partial \phi} \\ &\quad - \theta(\lambda_{11} + \lambda_{12} + \mu_1) \frac{\partial \Psi}{\partial \theta} - \phi(\lambda_{21} + \lambda_{22} + \mu_2) \frac{\partial \Psi}{\partial \phi} \\ &= (\lambda_{11} \theta^2 + \lambda_{12} \theta \phi + \mu_1 - \theta(\lambda_{11} + \lambda_{12} + \mu_1)) \frac{\partial \Psi}{\partial \theta} \\ &\quad + (\lambda_{22} \phi^2 + \lambda_{21} \theta \phi + \mu_2 - \phi(\lambda_{21} + \lambda_{22} + \mu_2)) \frac{\partial \Psi}{\partial \phi} \end{aligned} \tag{12}$$

$$\Psi(0, \theta, \phi) = \theta^{M_{1,0}} \phi^{M_{2,0}}$$

This is a semi-linear PDE. The subsidiary equations are defined as

$$\begin{aligned} \frac{d\Psi}{0} = \frac{dt}{1} &= \frac{-d\theta}{\lambda_{11} \theta^2 + \lambda_{12} \theta \phi + \mu_1 - \theta(\lambda_{11} + \lambda_{12} + \mu_1)} \\ &= \frac{-d\phi}{\lambda_{22} \phi^2 + \lambda_{21} \theta \phi + \mu_2 - \phi(\lambda_{21} + \lambda_{22} + \mu_2)} \end{aligned} \tag{13}$$

The first fraction does not mean divide  $d\Psi$  by 0 and combining with the second fraction  $\frac{dt}{1}$  infers that  $\Psi = \text{constant}$ , according to chapter 8 of Bailey (1991). Matching the third and fourth differentials above, we have

$$\frac{d\theta}{d\phi} = \frac{\lambda_{11} \theta^2 + \lambda_{12} \theta \phi + \mu_1 - \theta(\lambda_{11} + \lambda_{12} + \mu_1)}{\lambda_{22} \phi^2 + \lambda_{21} \theta \phi + \mu_2 - \phi(\lambda_{21} + \lambda_{22} + \mu_2)} \tag{14}$$

It seems that there is no way to solve this non-linear ODE and therefore no explicit solution is available for this PDE. However, it can be shown that this PDE gives a unique solution by Existence-Uniqueness Theorem for Quasilinear First-Order Equations. With regard to its characteristic, similar to the univariate case, this process is a pure-jump semimartingale with

following characteristic triplets:

$$Ch(\mathbf{M}_t) = \begin{cases} B_t = 0 \\ C_t = 0 \\ v(\mathbf{M}_t; dt, dx) = dt K(\mathbf{M}_t, dx) = \\ dt(\tilde{\lambda}_1 \delta_{(1,0)}(dx) + \tilde{\lambda}_2 \delta_{(0,1)}(dx) + \tilde{\mu}_1 \delta_{(-1,0)}(dx) + \tilde{\mu}_2 \delta_{(0,-1)}(dx)) \mathbf{M}_{t-} \\ \int_R (x^2 \wedge 1) K(\mathbf{M}_t, dx) = (\tilde{\lambda}_1 + \tilde{\lambda}_2 + \tilde{\mu}_1 + \tilde{\mu}_2) \mathbf{M}_{t-} < \infty, \text{ given that } \mathbf{M}_{t-} \text{ is finite,} \\ \text{where} \\ \tilde{\lambda}_1 = (\lambda_{11}, \lambda_{21}), \quad \tilde{\lambda}_2 = (\lambda_{21}, \lambda_{22}), \quad \tilde{\mu}_1 = (\mu_1, 0), \quad \tilde{\mu}_2 = (0, \mu_2) \end{cases} \quad (15)$$

The moments of this bivariate process can be derived by applying again infinitesimal generator.

**Proposition 1** *The first and second moments of the bivariate birth and death process  $\mathbf{M}_t = (M_{1,t}, M_{2,t})$  defined in (11) are given by*

$$\begin{aligned} \mathbb{E}[M_{1,t}] &= M_{1,0} \left( \frac{\lambda_{12}c}{2\lambda_{12}c + \kappa_1 - \kappa_2} e^{(\lambda_{12}c - \kappa_2)t} + \frac{\lambda_{12}c + \kappa_1 - \kappa_2}{2\lambda_{12}c + \kappa_1 - \kappa_2} e^{-(\lambda_{12}c + \kappa_1)t} \right) \\ &\quad + M_{2,0} \frac{\lambda_{21}}{2\lambda_{12}c + \kappa_1 - \kappa_2} \left( e^{(\lambda_{12}c - \kappa_2)t} - e^{-(\lambda_{12}c + \kappa_1)t} \right) \\ \mathbb{E}[M_{2,t}] &= M_{1,0} \frac{\lambda_{12}}{2\lambda_{12}c + \kappa_1 - \kappa_2} \left( e^{(\lambda_{12}c - \kappa_2)t} - e^{-(\lambda_{12}c + \kappa_1)t} \right) \\ &\quad + M_{2,0} \left( \frac{\lambda_{12}c + \kappa_1 - \kappa_2}{2\lambda_{12}c + \kappa_1 - \kappa_2} e^{(\lambda_{12}c - \kappa_2)t} + \frac{\lambda_{12}c}{2\lambda_{12}c + \kappa_1 - \kappa_2} e^{-(\lambda_{12}c + \kappa_1)t} \right), \end{aligned} \quad (16)$$

where

$$\kappa_1 = \mu_1 - \lambda_{11}, \quad \kappa_2 = \mu_2 - \lambda_{22}, \quad c = \frac{\kappa_2 - \kappa_1 + \sqrt{(\kappa_1 - \kappa_2)^2 + 4\lambda_{21}\lambda_{12}}}{2\lambda_{12}}.$$

The second moments  $\mathbb{E}[M_{1,t}^2]$ ,  $\mathbb{E}[M_{2,t}^2]$  and  $\mathbb{E}[M_{1,t}M_{2,t}]$  are determined by the following system of ODE,

$$\begin{aligned} \frac{d}{dt} \mathbb{E}[M_{1,t}^2] &= -2\kappa_1 \mathbb{E}[M_{1,t}^2] + 2\lambda_{21} \mathbb{E}[M_{1,t}M_{2,t}] + \lambda_{21} \mathbb{E}[M_{2,t}] + \mu_1 \mathbb{E}[M_{1,t}] \\ \frac{d}{dt} \mathbb{E}[M_{2,t}^2] &= -2\kappa_2 \mathbb{E}[M_{2,t}^2] + 2\lambda_{12} \mathbb{E}[M_{1,t}M_{2,t}] + \lambda_{12} \mathbb{E}[M_{1,t}] + \mu_2 \mathbb{E}[M_{2,t}] \\ \frac{d}{dt} \mathbb{E}[M_{1,t}M_{2,t}] &= -(\kappa_1 + \kappa_2) \mathbb{E}[M_{1,t}M_{2,t}] + \lambda_{21} \mathbb{E}[M_{2,t}^2] + \lambda_{12} \mathbb{E}[M_{1,t}^2] \end{aligned} \quad (17)$$

**Proof** See Appendix A.1. □

Note that to ensure the bivariate process becomes extinct with probability one, we need the (necessary and sufficient condition)  $(\mu_1 - \lambda_{11})(\mu_2 - \lambda_{22}) > \lambda_{12}\lambda_{21}$  according to Griffiths (1973). Many interesting properties of the process have been investigated by Griffiths (1972, 1973). In general, this bivariate birth and death process is not straightforward to apply in practice because there are no explicit solutions to the above PDE, and the second moments have to be evaluated by numerical methods. The discrete integer-value model proposed in the next section would be a possible solution.

### 3 Univariate and bivariate INAR models

In this section, we will introduce integer-valued autoregressive models which will serve as discrete approximations for continuous counterparts discussed in the last section. The derivation of this approximation will demonstrate how to parameterize the bivariate INAR case.

#### 3.1 Univariate INAR model

The classical integer-value autoregressive (INAR) model is introduced by defining a so-called binomial thinning operator  $\circ$  such that  $\alpha \circ X$  is the sum of  $X$  i.i.d Bernoulli random variable with success probability  $\alpha$ . i.e.

$$\alpha \circ X = \sum_{i=1}^X b_i, \quad b_i \stackrel{i.i.d}{\sim} \text{Bernoulli}(\alpha) \tag{18}$$

A well-known Poisson INAR(1) model  $X_t$  is given by

$$X_t = \alpha \circ X_{t-1} + R_t, \tag{19}$$

where  $\{R_i\}_{i=1, \dots, t}$  are i.i.d Poisson variables with parameter  $\rho$ . The key idea of the integer-value model would be the operator  $\circ$ . One can choose different discrete random variables to construct different integer-valued models. Indicated by the transition probability of continuous birth and death process, i.e. the sum of i.i.d zero-modified geometric random variables shown in Eq. (4), INAR model can be a good approximation by combining  $\circ$  and geometric operator as defined below.

**Definition 1** A birth and death INAR(1) model with survival probability  $\alpha \in [0, 1]$  and birth probability  $p \in [0, 1]$  is defined as

$$X_t = p *_1 \alpha \circ X_{t-1}, \tag{20}$$

where

- $\circ$  is the binomial operator
- $*_1$  is a geometric (reproduction) operator such that  $p *_1 X = \sum_{i=1}^X g_i^{(1)}$  with  $g_i^{(1)}$  being i.i.d geometric random variable with success probability  $p$  whose probability mass function is given by

$$P(g_i^{(1)} = k) = p(1 - p)^{k-1}, \quad k = 1, 2, \dots,$$

- $p *_1 \alpha \circ X = \sum_{i=1}^{\alpha \circ X} g_i^{(1)}$

**Remark** The innovation is dropped as there is no independent immigrant process in the birth and death process investigated.

**Proposition 2** *The birth and death INAR(1) model has the following statistical properties*

1. *The probability generating function of  $X_t$  can be iterated backwardly such that*

$$\begin{aligned} \varphi^{(I)}(t, \theta) &= \mathbb{E}[\theta^{X_t}] = \mathbb{E} \left[ \left( 1 - \alpha + \frac{\alpha p \theta}{1 - (1 - p)\theta} \right)^{X_{t-1}} \right] \\ &= \mathbb{E} \left[ \left( 1 - \alpha_i + \frac{\alpha_i p_i \theta}{1 - (1 - p_i)\theta} \right)^{X_{t-i}} \right], \quad i = 1, \dots, t \end{aligned} \tag{21}$$

where

$$\begin{aligned}
 p_i &= \frac{p^i}{d_{i-1}} \quad \alpha_i = \frac{\alpha^i}{d_{i-1}} \\
 d_i &= p^i \left( 1 + (1-p) \frac{\frac{\alpha}{p} - \left(\frac{\alpha}{p}\right)^{i+1}}{1 - \frac{\alpha}{p}} \right)
 \end{aligned}
 \tag{22}$$

In other words, the birth and death operator  $p *_1 \alpha \circ$  as a whole is iterable.

$$X_t = p_1 *_1 \alpha_1 \circ X_{t-1} = p_2 *_1 \alpha_2 \circ X_{t-2} = \dots = p_t *_1 \alpha_t \circ X_0
 \tag{23}$$

2. Then the mean, variance and covariance are given by

$$\begin{aligned}
 \mathbb{E}[X_t] &= \frac{\alpha_i}{p_i} \mathbb{E}[X_{t-i}] \\
 Var(X_t) &= \left( \frac{\alpha_i(1-p_i)}{p_i^2} + \frac{\alpha_i(1-\alpha_i)}{p_i^2} \right) \mathbb{E}[X_{t-i}] + \frac{\alpha_i^2}{p_i^2} Var(X_{t-i}) \\
 Cov(X_t, X_{t-i}) &= \frac{\alpha_i}{p_i} Var(X_{t-i})
 \end{aligned}
 \tag{24}$$

**Proof** See Appendix A.2. □

Note that if  $\alpha/p < 1$ , the process  $X_t$  will become extinct eventually. It is obvious that the continuous birth and death process can be approximated by this discrete INAR(1) model by directly matching the probability generating function  $\varphi^{(I)}$  to the one  $\varphi$  in Eq. (3) as the  $p *_1 \alpha \circ X$  is the sum of  $X$  i.i.d zero-modified geometric random variables.

### 3.2 Bivariate INAR model

Discrete approximation for univariate birth and death process is somehow simple because the PDE(2) has an explicit solution and hence the distribution is already known. In the case where the dynamic of two populations are characterized by (11), no explicit solution for its PDE (12). However, from the birth and death INAR(1) model, it is clear that birth and death probability are closely related to binomial and negative binomial random variables. Based on the dynamic (11) and linear form of the first moment (16), a bivariate INAR(1) model is proposed as follows.

**Definition 2** A bivariate birth and death INAR(1) model  $\mathbf{Y}_t = (Y_{1,t}, Y_{2,t})^T$  with survival probability  $\alpha_1, \alpha_2 \in [0, 1]$  and birth probability  $\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22} \in [0, 1]$  is defined as

$$\begin{aligned}
 Y_{1,t} &= \beta_{11} *_1 \alpha_1 \circ Y_{1,t-1} + \beta_{21} *_2 Y_{2,t-1} \\
 Y_{2,t} &= \beta_{12} *_2 Y_{1,t-1} + \beta_{22} *_1 \alpha_2 \circ Y_{2,t-1},
 \end{aligned}
 \tag{25}$$

where

- $\circ$  is the binomial operator
- $*_2$  is another geometric (reproduction) operator different from  $*_1$  such that  $\beta *_2 X = \sum_{i=1}^X g_i^{(2)}$  with  $g_i^{(2)}$  being i.i.d geometric random variable whose success probability is  $\beta$ . The probability mass function is given by

$$P(g_i^{(2)} = k) = \beta(1-\beta)^k, \quad k = 0, 1, 2, \dots,$$



- Conditional on  $Y_{t-1}$ , the random variables  $\beta_{11} * \alpha_1 \circ Y_{1,t-1}$ ,  $\beta_{21} * Y_{2,t-1}$ ,  $\beta_{12} * Y_{1,t-1}$  and  $\beta_{22} * \alpha_2 \circ Y_{1,t-1}$  are all independent of each other.

Now it seems that the structure of bivariate INAR(1) matches the the dynamics of (11), i.e. the birth probability depends on the size of both populations while death probability depends on the size of its own population. We adopt another geometric random variable  $g^{(2)}$  which is slightly different from  $g^{(1)}$  because for example, if we use  $g^{(1)}$ ,  $Y_{1,t} \geq Y_{2,t-1} \forall t$  which is not reasonable when  $Y_{1,t-1} < Y_{2,t-1}$  for a population.

**Proposition 3** *The first and second moments of the bivariate INAR(1) defined above are characterized by the following recursive formulas*

$$\begin{aligned}
 \mathbb{E}[Y_{1,t}] &= \frac{\alpha_1}{\beta_{11}} \mathbb{E}[Y_{1,t-1}] + \frac{1 - \beta_{21}}{\beta_{21}} \mathbb{E}[Y_{2,t-1}] \\
 \mathbb{E}[Y_{2,t}] &= \frac{1 - \beta_{12}}{\beta_{12}} \mathbb{E}[Y_{1,t-1}] + \frac{\alpha_2}{\beta_{22}} \mathbb{E}[Y_{2,t-1}] \\
 Var(Y_{1,t}) &= \frac{\alpha_1^2}{\beta_{11}^2} Var(Y_{1,t-1}) + \frac{\alpha_1(2 - \beta_{11} - \alpha_1)}{\beta_{11}^2} \mathbb{E}[Y_{1,t-1}] + \left(\frac{1 - \beta_{21}}{\beta_{21}}\right)^2 Var(Y_{2,t-1}) \\
 &\quad + \frac{1 - \beta_{21}}{\beta_{21}^2} \mathbb{E}[Y_{2,t-1}] + 2 \frac{\alpha_1(1 - \beta_{21})}{\beta_{11}\beta_{21}} Cov(Y_{1,t-1}, Y_{2,t-1}) \\
 Var(Y_{2,t}) &= \left(\frac{1 - \beta_{12}}{\beta_{12}}\right)^2 Var(Y_{1,t-1}) + \frac{1 - \beta_{12}}{\beta_{12}^2} \mathbb{E}[Y_{1,t-1}] + \frac{\alpha_2^2}{\beta_{22}^2} Var(Y_{2,t-1}) \\
 &\quad + \frac{\alpha_2(2 - \beta_{22} - \alpha_2)}{\beta_{22}^2} \mathbb{E}[Y_{2,t-1}] + 2 \frac{\alpha_2(1 - \beta_{12})}{\beta_{12}\beta_{22}} Cov(Y_{1,t-1}, Y_{2,t-1}) \\
 Cov(Y_{1,t}, Y_{2,t}) &= \left(\frac{\alpha_1\alpha_2}{\beta_{11}\beta_{22}} + \frac{(1 - \beta_{21})(1 - \beta_{12})}{\beta_{12}\beta_{21}}\right) Cov(Y_{1,t-1}, Y_{2,t-1}) \\
 &\quad + \frac{\alpha_1(1 - \beta_{12})}{\beta_{11}\beta_{12}} Var(Y_{1,t-1}) + \frac{\alpha_2(1 - \beta_{21})}{\beta_{21}\beta_{22}} Var(Y_{2,t-1})
 \end{aligned} \tag{26}$$

**Proof** Similar to Proposition 2, the moments can be derived by conditional expectation. The first and second moment for random variable  $g_i^{(2)}$  with parameter  $\beta$  are  $\frac{1-\beta}{\beta}$  and  $\frac{1-\beta}{\beta^2}$ . Then the first moment for  $X_t$  are

$$\begin{aligned}
 \mathbb{E}[Y_{1,t}|Y_{t-1}] &= \mathbb{E}[\beta_{11} * \alpha_1 \circ Y_{1,t-1}|Y_{1,t-1}] + \mathbb{E}[\beta_{21} * Y_{2,t-1}|Y_{2,t-1}] \\
 &= \frac{\alpha_1}{\beta_{11}} Y_{1,t-1} + \frac{1 - \beta_{21}}{\beta_{21}} Y_{2,t-1}
 \end{aligned}$$

The second moments are given by

$$\begin{aligned}
 \text{Var}(Y_{1,t} | \mathbf{Y}_t) &= \text{Var}(\beta_{11} * \alpha_1 \circ Y_{1,t-1} | Y_{1,t-1}) + \text{Var}(\beta_{21} * Y_{2,t-1} | Y_{2,t-1}) \\
 &= \frac{\alpha_1(2 - \beta_{11} - \alpha_1)}{\beta_{11}^2} Y_{1,t-1} + \frac{1 - \beta_{21}}{\beta_{21}^2} Y_{2,t-1} \\
 \text{Var}(Y_{1,t}) &= \text{Var}(\mathbb{E}[Y_{1,t-1} | \mathbf{Y}_{t-1}]) + \mathbb{E}[\text{Var}(Y_{1,t} | \mathbf{Y}_{t-1})] \\
 &\quad + 2\text{Cov}(\mathbb{E}[\beta_{11} * \alpha_1 \circ Y_{1,t-1} | Y_{1,t-1}], \mathbb{E}[\beta_{21} * Y_{2,t-1} | Y_{2,t-1}]) \\
 &\quad + 2\mathbb{E}[\text{Cov}(\beta_{11} * \alpha_1 \circ Y_{1,t-1}, \beta_{21} * Y_{2,t-1} | \mathbf{Y}_{t-1})] \\
 &= \text{Var}(\mathbb{E}[Y_{1,t} | \mathbf{Y}_{t-1}]) + \mathbb{E}[\text{Var}(Y_{1,t} | \mathbf{Y}_{t-1})] \\
 &\quad + \frac{\alpha_1(1 - \beta_{21})}{\beta_{11}\beta_{21}} \text{Cov}(Y_{1,t-1}, Y_{2,t-1}) \\
 \text{Cov}(Y_{1,t}, Y_{2,t}) &= \text{Cov}(\beta_{11} * \alpha_1 \circ Y_{1,t-1}, \beta_{12} * Y_{1,t-1}) \\
 &\quad + \text{Cov}(\beta_{11} * \alpha_1 \circ Y_{1,t-1}, \beta_{22} * \alpha_2 \circ Y_{2,t-1}) \\
 &\quad + \text{Cov}(\beta_{21} * Y_{2,t-1}, \beta_{12} * Y_{1,t-1}) + \text{Cov}(\beta_{21} * Y_{2,t-1}, \beta_{22} * \alpha_2 \circ Y_{2,t-1}) \\
 &= \frac{\alpha_1(1 - \beta_{12})}{\beta_{11}\beta_{12}} \text{Var}(Y_{1,t-1}) + \frac{\alpha_1\alpha_2}{\beta_{11}\beta_{22}} \text{Cov}(Y_{1,t-1}, Y_{2,t-1}) \\
 &\quad + \frac{(1 - \beta_{12})(1 - \beta_{21})}{\beta_{12}\beta_{21}} \text{Cov}(Y_{2,t-1}, Y_{1,t-1}) + \frac{(1 - \beta_{21}\alpha_2)}{\beta_{21}\beta_{22}} \text{Var}(Y_{2,t-1})
 \end{aligned}$$

The first and second moments of  $Y_{2,t}$  can be derived in a similar way.  $\square$

**Proposition 4** *If the eigen-values  $\eta_1, \eta_2$  of the following matrix*

$$A = \begin{bmatrix} \frac{\alpha_1}{\beta_{11}} & \frac{1 - \beta_{21}}{\beta_{21}} \\ \frac{1 - \beta_{12}}{\beta_{12}} & \frac{\alpha_2}{\beta_{22}} \end{bmatrix} \quad (27)$$

*lie in the interval  $[-1, 1]$ , then the bivariate population  $X_t, Y_t$  will become extinct eventually.*

**Proof** The first moment can be expressed in a matrix form

$$\mathbb{E}[\mathbf{Y}_t] = A\mathbb{E}[\mathbf{Y}_{t-1}] = A^t \mathbb{E}[\mathbf{Y}_0] \quad (28)$$

The  $t$ th power of a matrix here is defined as  $t$  times matrix multiplication. By eigen-decomposition, power of a matrix can be expressed as

$$A^t = Q \text{diag}(\{\eta_1^t, \eta_2^t\}) Q^{-1}, \quad (29)$$

where  $Q = (v_1, v_2)$  is eigen vector matrix with  $v_1, v_2$  as eigen vectors for  $\eta_1, \eta_2$ . Now, it is clear that  $\mathbb{E}[\mathbf{Y}_t]$  is decreasing in  $t$  when  $\eta_1, \eta_2 \in [-1, 1]$ .  $\square$

## 4 Weak convergence to continuous birth and death process

In this section, we will construct two continuous processes from the above proposed INAR models. These processes, under a certain parametrization, will converge weakly to the aforementioned continuous birth and death processes when the length of sub-interval goes to 0.

### 4.1 Construction of continuous processes

Since the continuous birth and death processes are clearly semimartingale defined in non-negative state spaces, to apply limit theorem of locally bounded semimartingales, we need to construct 'continuous' processes on a dense subsets of  $\mathbb{R}_+$  (will take  $t \in [0, 1]$  for convenience) and compute their characteristic triplets from the discrete INAR models. Finally, when everything is set up nicely, we can apply weak convergence of semimartingale theorem to prove the result. The construction mainly follows from Jacod and Shiryaev (2013, Chapter II, section 3).

Starting with a discrete basis  $\mathcal{B} = (\cdot, \mathbf{F}, (\mathcal{F}_n)_{n \in \mathbb{N}}, \mathbf{P})$ , assume that the INAR models  $X_n$  and  $\mathbf{Y}_n$  defined above are adapted to this discrete stochastic basis and so as the increment processes

$$\begin{aligned} U_k &= X_k - X_{k-1}, \quad U_0 = X_0 \\ \mathbf{V}_k &= \mathbf{Y}_k - \mathbf{Y}_{k-1}, \quad \mathbf{V}_0 = \mathbf{Y}_0, \quad k = 0, 1, 2, \dots \end{aligned} \tag{30}$$

then we can construct 'continuous' processes via time change.

**Definition 3** Given a fixed time interval  $[0, 1]$ , one can define a equal-length grid with size  $n$  such that each subinterval with length  $\Delta = \frac{1}{n}$ . The following the processes:

$$Z_t^{(n)} = \sum_{k=0}^{\sigma_t} U_k, \quad \mathbf{M}_t^{(n)} = \sum_{k=0}^{\sigma_t} \mathbf{V}_k, \tag{31}$$

where  $\sigma_t = \lfloor tn \rfloor$ , are adapted to the continuous-time basis  $\tilde{\mathcal{B}} = (\cdot, \mathbf{F}, G = (g_t)_{t \geq 0}, \mathbf{P})$ . The parameters setting for  $Z_t^{(n)}$  are

$$\alpha = \frac{(\lambda - \mu)e^{(\lambda - \mu)\Delta}}{\lambda e^{(\lambda - \mu)\Delta} - \mu}, \quad p = \frac{\lambda - \mu}{\lambda e^{(\lambda - \mu)\Delta} - \mu}. \tag{32}$$

The parameters setting for  $M_t^{(n)}$  are

$$\begin{aligned} \alpha_1 &= \frac{(\lambda_{11} - \mu_1)\omega_1(\Delta)}{\lambda_{11}\omega_1(\Delta) - \mu_1}, \quad \alpha_2 = \frac{(\lambda_{22} - \mu_2)\omega_2(\Delta)}{\lambda_{22}\omega_2(\Delta) - \mu_2} \\ \beta_{11} &= \frac{\lambda_{11} - \mu_1}{\lambda_{11}\omega_1(\Delta) - \mu_1}, \quad \beta_{22} = \frac{\lambda_{22} - \mu_2}{\lambda_{22}\omega_2(\Delta) - \mu_2} \\ \beta_{21} &= (1 + C_{\beta_1}(e^{u_1\Delta} - e^{u_2\Delta}))^{-1}, \quad \beta_{12} = (1 + C_{\beta_2}(e^{u_1\Delta} - e^{u_2\Delta}))^{-1}, \end{aligned} \tag{33}$$

where

$$\begin{aligned} \omega_1(\Delta) &= C_\alpha e^{u_1\Delta} + (1 - C_\alpha)e^{u_2\Delta}, \quad \omega_2(\Delta) = (1 - C_\alpha)e^{u_1\Delta} + C_\alpha e^{u_2\Delta} \\ C_\alpha &= \frac{\lambda_{12}c}{2\lambda_{12}c + \mu_1 - \mu_2}, \quad C_{\beta_1} = \frac{\lambda_{21}}{2\lambda_{12}c + \kappa_1 - \kappa_2}, \quad C_{\beta_2} = \frac{\lambda_{12}}{2\lambda_{12}c + \kappa_1 - \kappa_2} \\ u_1 &= \lambda_{12}c - \kappa_2, \quad u_2 = -(\lambda_{12}c + \kappa_1), \quad \kappa_i = \mu_i - \lambda_{ii}, \quad i = 1, 2 \\ c &= \frac{\kappa_2 - \kappa_1 + \sqrt{(\kappa_1 - \kappa_2)^2 + 4\lambda_{21}\lambda_{12}}}{2\lambda_{12}}. \end{aligned}$$

It is straightforward to derive the parameter setting for univariate case since we only need to match the parameter via probability generating function between  $Z_t^{(n)}$  and  $Z_t$ . However, in the other case where the closed form probability generating function for  $\mathbf{M}_t$  is not available, we need to seek other ways to set up  $\alpha_i$  and  $\beta_{i,j}$  in terms of  $\lambda$  and  $\mu$ . The direct approach

would be to match the first and second order moments to see whether it works. It is clear that we can match moment equations (26) to (16) and find out the mapping of  $\beta_{12}, \beta_{21}$  in terms of  $\lambda_{i,j}, \mu_i, i, j \in \{1, 2\}$ . Unfortunately, only the ratio  $\alpha_i/\beta_{ii}$  is known. Nevertheless, the parameter setting in univariate case shows us the way to distribute the ratio  $\alpha/p$  to  $\alpha$  and  $p$ . Then  $\alpha_i, \beta_{ii}$  can be set up in a similar way.

**Proposition 5** *With the above parameters setting and any non-negative integer  $m$ , the transition probabilities for  $Z_t^{(n)}$  conditional on  $Z_{t-\Delta}^{(n)} = k$  are*

$$\begin{aligned} \Pr(Z_t^{(n)} = k + m | Z_{t-\Delta}^{(n)} = k) &= \binom{k + m - 1}{k - 1} (\lambda \Delta)^m + o(\Delta^m) \\ \Pr(Z_t^{(n)} = k - m | Z_{t-\Delta}^{(n)} = k) &= \binom{k}{k - m} (\mu \Delta)^m + o(\Delta^m) \end{aligned} \tag{34}$$

The above probabilities can be simplified as,

$$\begin{aligned} \Pr(Z_t^{(n)} = k + 1 | Z_{t-\Delta}^{(n)} = k) &= \lambda k \Delta + o(\Delta) \\ \Pr(Z_t^{(n)} = k - 1 | Z_{t-\Delta}^{(n)} = k) &= \mu k \Delta + o(\Delta) \\ \Pr(|Z_t^{(n)} - k| \geq 2 | Z_{t-\Delta}^{(n)} = k) &= o(\Delta) \end{aligned} \tag{35}$$

On the other hand, the transition probabilities for  $\mathbf{M}_t^{(n)}$  conditional on  $\mathbf{M}_{t-\Delta}^{(n)} = \mathbf{k} = (k_1, k_2)$  given by

$$\begin{aligned} \Pr(M_{i,t}^{(n)} = k_i + m | \mathbf{M}_{t-\Delta}^{(n)} = \mathbf{k}) &= \sum_{j=k_i}^{k_i+m} \binom{j-1}{k_i-1} \binom{k_1+k_2+m-j-1}{k_{i'}-1} (\lambda_{ii} \Delta)^{j-k_i} (\lambda_{i',i} \Delta)^{k_i+m-j} + o(\Delta^m) \\ \Pr(M_{i,t}^{(n)} = k_i - m | \mathbf{M}_{t-\Delta}^{(n)} = \mathbf{k}) &= \binom{k_i}{k_i-m} (\mu_i \Delta)^m + o(\Delta^m), \end{aligned} \tag{36}$$

where  $i \in \{1, 2\}$  and  $i' = 3 - i$ . Due to the conditional independence of bivariate INAR models, the joint transition probabilities for  $\mathbf{M}_t^{(n)}$  conditional on  $\mathbf{M}_{t-\Delta}^{(n)}$  are

$$\begin{aligned} \Pr(M_{1,t}^{(n)} = k_1 \pm m_1, M_{2,t}^{(n)} = k_2 \pm m_2 | \mathbf{M}_{t-\Delta}^{(n)} = \mathbf{k}) &= \Pr(M_{1,t}^{(n)} = k_1 \pm m_1 | \mathbf{M}_{t-\Delta}^{(n)} = \mathbf{k}) \Pr(M_{2,t}^{(n)} = k_2 \pm m_2 | \mathbf{M}_{t-\Delta}^{(n)} = \mathbf{k}) \end{aligned} \tag{37}$$

Similarly, the above probabilities can be simplified as

$$\begin{aligned} \Pr(M_{i,t}^{(n)} = k_i + 1 | \mathbf{M}_{t-\Delta}^{(n)} = \mathbf{k}) &= \lambda_{ii} k_i \Delta + \lambda_{i',i} k_2 \Delta + o(\Delta) \\ \Pr(M_{i,t}^{(n)} = k_i - 1 | \mathbf{M}_{t-\Delta}^{(n)} = \mathbf{k}) &= \mu_i k_i \Delta + o(\Delta) \\ \Pr(|M_{i,t}^{(n)} - k_i| \geq 2 | \mathbf{M}_{t-\Delta}^{(n)} = \mathbf{k}) &= o(\Delta) \end{aligned} \tag{38}$$

**Proof** See Appendix A.3. □

It is obvious that the above transition probabilities have exactly the same form as continuous counterparts when  $m = 1$ . Consequently, the Lévy measures of  $Z_t^{(n)}$  and  $\mathbf{M}_t^{(n)}$  have similar structure to their continuous counterparts.

**Proposition 6** *The continuous processes  $Z_t^{(n)}$  and  $\mathbf{M}_t^{(n)}$  defined above are semimartingales with following characteristics triplets.*

$$\begin{aligned}
 Ch(Z_t^{(n)}) &= \begin{cases} B_t = 0 \\ C_t = 0 \\ \nu([0, t] \times g) = \sum_{k=1}^{\sigma_t} (g(1)\lambda + g(-1)\mu)X_{k-1}\Delta + O(\Delta) \end{cases} \\
 Ch(\mathbf{M}_t^{(n)}) &= \begin{cases} \mathbf{B}_t = 0 \\ \mathbf{C}_t = 0 \\ \nu([0, t] \times g) = \sum_{k=1}^{\sigma_t} \left( g(1, 0)\tilde{\lambda}_1 + g(-1, 0)\tilde{\mu}_1 \right) \mathbf{Y}_{k-1}\Delta \\ \quad + \left( g(0, 1)\tilde{\lambda}_2 + g(0, -1)\tilde{\mu}_2 \right) \mathbf{Y}_{k-1}\Delta \\ \quad + O(\Delta), \end{cases} \tag{39}
 \end{aligned}$$

where the  $g$  is a continuous, non-negative, bounded Borel function vanishing near 0 and  $\mathbf{M}_t^{(n)}$  respectively, the truncation function is  $h = |x|\mathbf{1}_{\{|x|<1\}}$  and

$$\tilde{\lambda}_1 = (\lambda_{11}, \lambda_{21}), \quad \tilde{\lambda}_2 = (\lambda_{21}, \lambda_{22}), \quad \tilde{\mu}_1 = (\mu_1, 0), \quad \tilde{\mu}_2 = (0, \mu_2)$$

**Proof** See Appendix A.4. □

**Theorem 7** *With the the definition and the parametrization above, and the initial distribution condition:*

$$Z_0^{(n)} = Z_0, \quad \mathbf{M}_0^{(n)} = \mathbf{M}_0, \tag{40}$$

the processes  $Z_t^{(n)}$  and  $\mathbf{M}_t^{(n)}$  converge weakly to the continuous birth and death processes  $Z_t$  and  $\mathbf{M}_t$ ,

$$\begin{aligned}
 \lim_{n \rightarrow \infty} Z_t^{(n)} &\xrightarrow{w} Z_t \\
 \lim_{n \rightarrow \infty} \mathbf{M}_t^{(n)} &\xrightarrow{w} \mathbf{M}_t, \tag{41}
 \end{aligned}$$

when the size of subinterval  $\Delta$  goes to 0 or equivalently,  $n \rightarrow \infty$ .

**Proof** Here we simply apply Theorem 3.39 from Jacod and Shiryaev (2013, chapter IX, section 3), the limit theorem of semimartingales for the locally bounded case.

- i The local strong Majorization Hypothesis: For both cases  $Z_t$  and  $\mathbf{M}_t$ , the first two terms of the characteristic triplets are 0 and stochastic integrals with respect to the function is clearly finite on  $[0, 1]$
- ii Local Conditions on big jumps: For both cases  $Z_t$  and  $\mathbf{M}_t$ , there is no jump with absolute size greater than 1.
- iii The local uniqueness: for every choices of initial distributions for  $Z_0$  and  $\mathbf{M}_0$ , their Lévy measures are uniquely characterized by their (joint) probability distribution functions.
- iv Continuity Condition, the characteristic triplets  $B_t(\omega)$ ,  $C_t(\omega)$ ,  $\nu(\omega; dt, dx)$  of  $Z_t$  and  $\mathbf{M}_t$  are continuous with respect to  $\omega$ .
- v Weak convergence of initial distribution. This is stated at the beginning of this theorem.
- vi Convergence of characteristic triplet of discrete processes to that of their continuous counterparts. This can be proved by showing the uniform convergence of Lévy measures. For every  $a > 0$ , define a stopping time for the population process:

$$S_a(X) = \inf \{t : |X_t| > a, \text{ or } |X_{t-}| > a\} \tag{42}$$

For the univariate case, the stochastic integral with respect to  $g * \nu$  for any Borel function  $g$  is given by

$$\begin{aligned}
 (g * \nu_{t \wedge S_a}) \circ Z^{(n)} &= g * \nu(Z^{(n)}; [0, t \wedge S_a(Z^{(n)})], R) \\
 &= \int_0^{t \wedge S_a(Z^{(n)})} \int_R g(x)(\lambda \delta_1(dx) + \mu \delta_{-1}(dx)) Z_{s^-}^{(n)} ds \\
 &= \int_0^{t \wedge S_a(Z^{(n)})} (g(1)\lambda + g(-1)\mu) Z_{s^-}^{(n)} ds \\
 &= \sum_{k=1}^{\sigma_{t \wedge S_a(Z^{(n)})}} (g(1)\lambda + g(-1)\mu) Z_{k-1}^{(n)} \Delta \\
 &\quad + (g(1)\lambda + g(-1)\mu) Z_{\sigma_{t \wedge S_a(Z^{(n)})}}^{(n)} \left( t \wedge S_a(Z^{(n)}) - \sigma_{t \wedge S_a(Z^{(n)})} \Delta \right)
 \end{aligned} \tag{43}$$

and the absolute difference of two stochastic integrals is given by,

$$\begin{aligned}
 &|g * \nu_{t \wedge S_a}^n - (g * \nu_{t \wedge S_a}) \circ Z^{(n)}| \\
 &= \left| O(\Delta) + (g(1)\lambda + g(-1)\mu) Z_{\sigma_{t \wedge S_a(Z^{(n)})}} \left( t \wedge S_a(Z^{(n)}) - \sigma_{t \wedge S_a(Z^{(n)})} \Delta \right) \right| \\
 &\leq O(\Delta) + |g(1)\lambda + g(-1)\mu| Z_{\sigma_{t \wedge S_a(Z^{(n)})}} \left( t \wedge S_a(Z^{(n)}) - \sigma_{t \wedge S_a(Z^{(n)})} \Delta \right)
 \end{aligned} \tag{44}$$

It is clear that all the quantity inside  $|\cdot|$  are finite and for every  $\xi > 0$ , and then there exists a natural number  $N$  such that for  $n > N$ , we have

$$|g * \nu_{t \wedge S_a}^n - (g * \nu_{t \wedge S_a}) \circ Z^{(n)}| < \xi \tag{45}$$

and hence we have the uniform convergence for  $g * \nu_{t \wedge S_a}^n$  to  $(g * \nu_{t \wedge S_a}) \circ Z^{(n)}$ . For the bivariate case, the stochastic integral  $g * \nu$ , where  $\nu$  is the Lévy measure of  $M$ , for any Borel function  $g$  is given by

$$\begin{aligned}
 (g * \nu_{t \wedge S_a}) \circ \mathbf{M}^{(n)} &= g * \nu(\mathbf{M}^{(n)}; [0, t \wedge S_a(\mathbf{M}^{(n)})], R) \\
 &= \int_0^{t \wedge S_a(\mathbf{M}^{(n)})} \int_R g(x)(\tilde{\lambda}_1 \delta_{(1,0)}(dx) + \tilde{\lambda}_2 \delta_{(0,1)}(dx) \\
 &\quad + \tilde{\mu}_1 \delta_{(0,-1)}(dx) + \tilde{\mu}_2 \delta_{(-1,0)}(dx)) \mathbf{M}_{s^-}^{(n)} ds \\
 &= \int_0^{t \wedge S_a(\mathbf{M}^{(n)})} \left( g(1,0)\tilde{\lambda}_1 + g(0,1)\tilde{\lambda}_2 + g(-1,0)\tilde{\mu}_1 + g(0,-1)\tilde{\mu}_2 \right) \mathbf{M}_{s^-}^{(n)} ds \\
 &= \sum_{k=1}^{t \wedge S_a(\mathbf{M}^{(n)})} \left( g(1,0)\tilde{\lambda}_1 + g(0,1)\tilde{\lambda}_2 + g(-1,0)\tilde{\mu}_1 + g(0,-1)\tilde{\mu}_2 \right) \mathbf{M}_{k-1}^{(n)} \Delta \\
 &\quad + \left( g(1,0)\tilde{\lambda}_1 + g(0,1)\tilde{\lambda}_2 + g(-1,0)\tilde{\mu}_1 + g(0,-1)\tilde{\mu}_2 \right) \\
 &\quad \times \mathbf{M}_{\sigma_{t \wedge S_a(\mathbf{M}^{(n)})}}^{(n)} \left( t \wedge S_a(\mathbf{M}^{(n)}) - \sigma_{t \wedge S_a(\mathbf{M}^{(n)})} \Delta \right)
 \end{aligned} \tag{46}$$

Then the absolute difference of two stochastic integrals is given by

$$\begin{aligned}
 & |g * \nu_{t \wedge S_a}(\mathbf{M}^{(n)}) - (g * \nu_{t \wedge S_a}) \circ \mathbf{M}^{(n)}| \\
 & \leq O(\Delta) + \left| g(1, 0)\tilde{\lambda}_1 + g(0, 1)\tilde{\lambda}_2 + g(-1, 0)\tilde{\mu}_1 + g(0, -1)\tilde{\mu}_2 \right| \\
 & \times \mathbf{M}_{\sigma_{t \wedge S_a}(\mathbf{M}^{(n)})}^{(n)} \left( t \wedge S_a(\mathbf{M}^{(n)}) - \sigma_{t \wedge S_a}(\mathbf{M}^{(n)}) \right)
 \end{aligned} \tag{47}$$

Hence the uniform convergence holds using similar argument as in the univariate case. Finally, the  $Z_t^{(n)}$ ,  $M_t^{(n)}$  converge weakly to  $Z_t$  and  $M_t$  respectively.  $\square$

### 5 Simulation study

In this section, we outline the simulation algorithm for bivariate birth and death processes. Then estimation method, properties of estimators are investigated in the simulation study.

#### 5.1 Simulation of bivariate birth and death process

The simulation algorithm of bivariate birth and death process  $\mathbf{M}_t$  can be derived straightforwardly according to its ODE (11). Given the current population  $\mathbf{M}_t$ , the waiting time that an event (birth or death in either population) will happen follows exponential distribution with rate

$$\rho_t = (\lambda_{11} + \lambda_{12} + \mu_1)M_{1,t} + (\lambda_{21} + \lambda_{22} + \mu_2)M_{2,t}$$

Then the probability that this event will happen in population  $M_{1,t}$  is

$$p_1 = \frac{\lambda_{21}M_{2,t} + (\lambda_{11} + \mu_1)M_{1,t}}{\rho_t} \tag{48}$$

The probability that this event will happen in population  $M_{2,t}$  would simply be  $p_2 = 1 - p_1$ . Suppose now an event happens in population  $M_{1,t}$ , the probability that there is a new individual would be

$$p_1^b = \frac{\lambda_{11}M_{1,t} + \lambda_{21}M_{2,t}}{\lambda_{21}M_{2,t} + (\lambda_{11} + \mu_1)M_{1,t}}, \tag{49}$$

and the probability that an individual dies is  $p_1^d = 1 - p_1^b$ . Likewise, if the event happens in the population  $M_{2,t}$ , the birth probability would be

$$p_2^b = \frac{\lambda_{12}M_{1,t} + \lambda_{22}M_{2,t}}{\lambda_{12}M_{1,t} + (\lambda_{22} + \mu_2)M_{2,t}} \tag{50}$$

and death probability  $p_n^d = 1 - p_n^b$ . Overall, the simulation algorithm is shown in the following Algorithm 1.

On the other hand, the simulation procedure of bivariate INAR(1) model is straightforward because the distribution of  $\mathbf{Y}_t$  are indicated by the operator  $(\circ, *_1, *_2)$  given  $\mathbf{Y}_{t-1}$ .

**Algorithm 1** Simulation of bivariate birth and death process with rates  $\{\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22}, \mu_1, \mu_2\}$ , initial population  $M_{1,0}, M_{2,0}$ , a vector of cumulative time  $t_c$  where  $t_c[1] = 0$ , a counter  $i$  and terminal time  $T$

1. Simulate a waiting time  $t_w \sim \text{Exp}(\rho_t)$  and two independent uniform random variable  $U_1, U_2 \sim U(0, 1)$
2. if  $U_1 \leq p_1$ , and  $U_2 \leq p_1^b$ ,  $M_{1,t+t_w} = M_{1,t} + 1$  and  $M_{2,t+t_w} = M_{2,t}$
3. if  $U_1 \leq p_1$ , and  $U_2 > p_1^b$ ,  $M_{1,t+t_w} = M_{1,t} - 1$  and  $M_{2,t+t_w} = M_{2,t}$
4. if  $U_1 > p_1$ , and  $U_2 \leq p_2^b$ ,  $M_{2,t+t_w} = M_{2,t} + 1$  and  $M_{1,t+t_w} = M_{1,t}$
5. if  $U_1 > p_1$ , and  $U_2 > p_2^b$ ,  $M_{2,t+t_w} = M_{2,t} - 1$  and  $M_{1,t+t_w} = M_{1,t}$
6. Append a new element to  $t_c$ ,  $t_c[i + 1] = t_c[i] + t_w$  and update counter  $i = i + 1$
7. Repeat all the steps above until  $t_c[i] > T$  or  $M_{1,t_c[i]} = M_{2,t_c[i]} = 0$
8. Set  $M_{1,T} = M_{1,t_c[i-1]}$  and  $M_{2,T} = M_{2,t_c[i-1]}$  and return the trajectory  $\mathbf{M}$  at each element of  $t_c$

## 5.2 Statistical inference of univariate and bivariate birth and death process

### 5.2.1 Quasi-MLE for bivariate LBD

In the univariate case, parameters estimation and their asymptotic properties are available in Keiding (1975). Suppose now we have the full information of the sample path, the exact inter-arrival times for each birth and death events  $\{\tau_i\}_{i=0,1,2,\dots}$  on the sampling interval  $[0, T]$  where  $\tau_0 = 0$ , the maximum likelihood estimators for  $Z_t$  are

$$\hat{\lambda} = \frac{B_T}{X_T}, \quad \hat{\mu} = \frac{D_T}{X_T}, \quad X_T = \sum_{k=1}^{B_T+D_T} \tau_k Z_{\tau_{k-1}} + \left( T - \sum_{i=1}^n \tau_i \right) Z_T, \quad (51)$$

where  $B_T, D_T$  are total number of birth and death events respectively. The asymptotic properties are given by fixed  $T$  and large population

$$\lim_{Z_0 \rightarrow \infty} \left( \frac{Z_0(e^{(\lambda-\mu)T} - 1)}{\lambda - \mu} \right)^{\frac{1}{2}} (\hat{\lambda} - \lambda) \xrightarrow{D} \mathbf{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix} \right) \quad (52)$$

In practice, one may not have exact information of inter-arrival time of the events. Instead, we have records for populations sampling over a fixed-length interval  $\Delta$  such that  $Z_0, Z_{\Delta}, Z_{2\Delta} \dots Z_{n\Delta}$  are available. Then to estimate the parameters  $\lambda, \mu$ , one can numerically maximize the Quasi log-likelihood function from the proposed INAR(1) model  $X_k = Z_{k\Delta}, k = 0, 1, \dots, n$ . The log likelihood function is given by,

$$\begin{aligned} \ell(\alpha, p) &= \sum_{k=1}^n \log \Pr(X_{k-1}, X_k) \\ \Pr(X_{k-1}, X_k) &= \begin{cases} 1, & X_{k-1} = X_k = 0 \\ (1 - \alpha)^{X_{k-1}}, & X_k = 0 \\ \sum_{j=1}^{\min\{X_{k-1}, X_k\}} f_b(j; X_{k-1}, \alpha) f_{nb}(X_k - j; j, p), & X_{k-1} > 0 \ \& \ X_k > 0, \end{cases} \end{aligned} \quad (53)$$

where  $f_b$  and  $f_{nb}$  are probability mass function of binomial and negative binomial random variables

$$f_b(x; n, \alpha) = \binom{n}{x} \alpha^x (1 - \alpha)^{n-x} \quad f_{nb}(x; n, \beta) = \binom{n+x-1}{n-1} \beta^n (1 - \beta)^x$$



**Table 1** Parameter setting for univariate case

Parameters	$\lambda$	$\mu$	$Z_0$	$T$
Values	1.2	1	100	1

**Table 2** Properties of different maximum likelihood estimators

	Bias $\lambda$	MSE $\lambda$	Bias $\mu$	MSE $\mu$	Time (s)
Theoretical	0	0.000126	0	0.000116	–
MLE	–0.000701	0.000124	0.015898	0.000396	0.09
$\Delta = 0.1$	–0.133845	0.075696	–0.116943	0.072754	49
$\Delta = 0.05$	–0.086435	0.023935	–0.069513	0.021836	98.6
$\Delta = 0.025$	–0.051234	0.007711	–0.034323	0.006557	201.4
$\Delta = 0.01$	–0.014178	0.001378	0.002721	0.001159	522.2

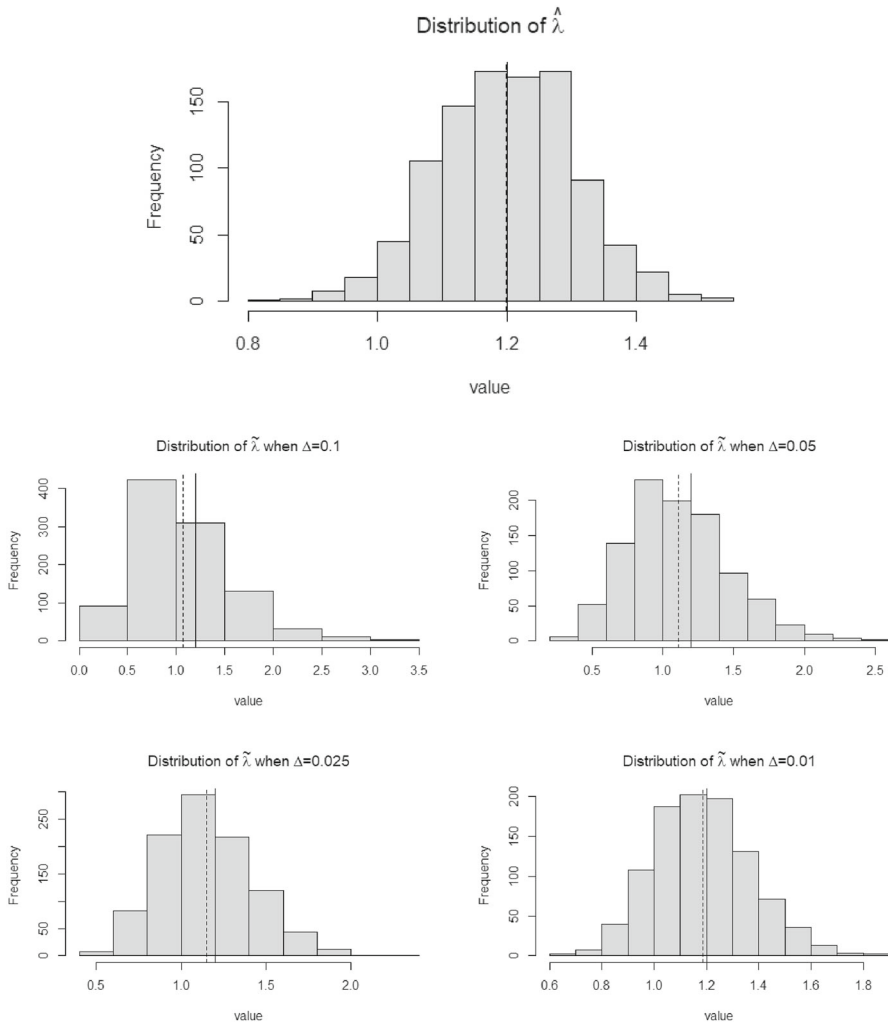
The time column is the total time of estimating 1000 sample paths

The simulation is conducted as follow: we generate 1000 sample paths of  $Z_t$  using the parameters settings in Table 1. Since  $Z_t$  are continuous sample paths, we set up an equal-distance grid with sampling interval  $\Delta$ . Then the equal-distance observations  $X_t$  are obtained by counting the total number of population up to each discrete time  $(0, \Delta, 2\Delta, \dots, n\Delta)$  where  $n = \frac{T}{\Delta}$ . The log likelihood function is then maximized by 'optim' function with method = 'BFGS' in R programming. Finally, we can recover the rate estimates by inverting the parametrization in equation (32) such that

$$\tilde{\lambda} = \frac{1-\hat{p}}{\frac{\hat{\alpha}}{\hat{p}} - 1} \log \frac{\hat{\alpha}}{\hat{p}}, \quad \tilde{\mu} = \tilde{\lambda} - \frac{1}{\Delta} \log \frac{\hat{\alpha}}{\hat{p}} \tag{54}$$

In the following, we will first explore how the size of  $\Delta$  would affect properties estimators, i.e. bias and mean square error (MSE), and how much more computational time we need compared to true MLE method. Four different size of sampling intervals  $\Delta = \{0.1, 0.05, 0.025, 0.01\}$  is chosen and the results are presented in Table 2. The theoretical row shows the biased and MSE computed through equation (52). There is no surprise that the True MLE method from Eq. (51) performs the best, with lowest MSE and computational time. The Quasi-MLE method by constructing INAR model, on the other hand, becomes better as we decreasing the size of sampling interval  $\Delta$  but it still performs no better than the true MLE method and require much more computational time. The empirical distribution of these estimators are illustrated in Fig. 1 and since the general shape of distribution of  $\tilde{\lambda}$  and  $\tilde{\mu}$  has little difference, we will only show the distribution of  $\tilde{\lambda}$ . It is clear that only the case  $\Delta = 0.01$  has satisfactory normal shape compared to all other cases.

To achieve asymptotic normality for Quasi-MLE method from INAR model, one need not only large initial population, but also a small sampling interval  $\Delta$ . In the following simulation, we would fix the sampling interval  $\Delta = 0.01$  and investigate how the size of initial population would affect the asymptotic distribution of estimators and the computational time for estimation procedure. To explore the effect of  $Z_0$  for asymptotic distribution, we choose  $Z_0 \in \{5, 10, 30, 50\}$  and it seems from Fig. 2 that to ensure asymptotic normality for both estimators, one need at least  $Z_0 = 30$ , which is a large sample size in statistical sense.



**Fig. 1** The empirical distribution of estimated parameters. The top panel is the MLE from 51 and the rest of plots are MLE from INAR model. The solid lines are the true values of the parameters listed in Table 1 and the dash lines stand for empirical means

The computational time with respect to  $Z_0 \in \{10, 50, 100, 150, \dots, 500\}$  clearly shows a linear trend in Fig. 3. This is reasonable as the number of summation involved in Eq. (53) increases linearly with respect to  $Z_0$

In summary, the Quasi-MLE method constructed from INAR model can reach moderate level of estimation accuracy and asymptotic normality with large initial population  $Z_0 \geq 30$  and small sampling interval  $\Delta \leq 0.01$ . However, it would require much more computational time than the true MLE method. This method should only be used in the case where we have no information on inter-arrival time of birth and death events.

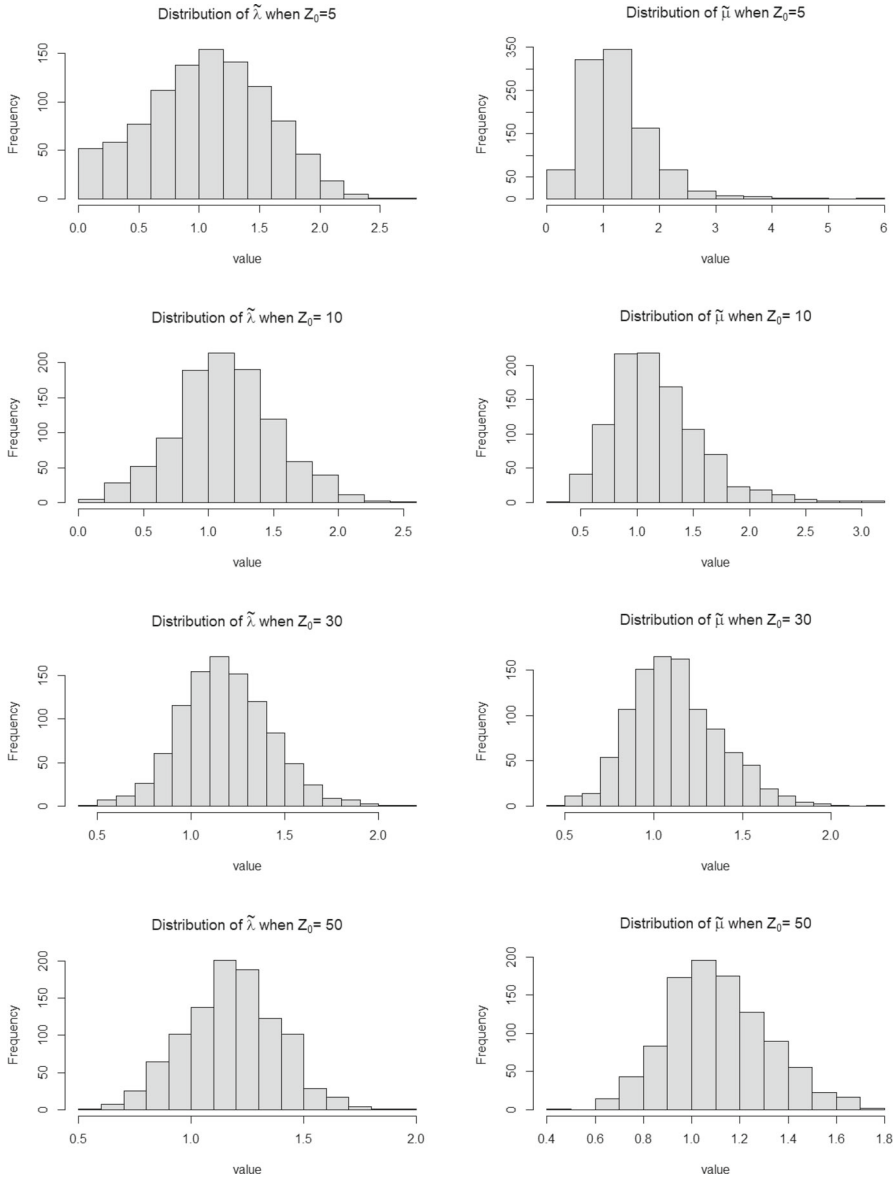


Fig. 2 Asymptotic distribution of  $\tilde{\lambda}$ ,  $\tilde{\mu}$  with different  $Z_0$

### 5.2.2 Quasi-MLE for bivariate LBD

Since the bivariate INAR(1) model is a bivariate Markov Chain, the log likelihood function can be written as the sum of logarithm of transition probabilities. Denote  $\Theta = \{\alpha_1, \alpha_2, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}\}$  as the parameter space of bivariate INAR(1) model, then the

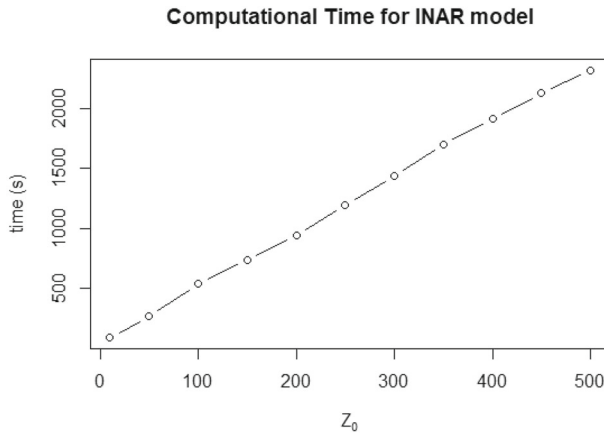


Fig. 3 The computational time for INAR models of 1000 sample paths

likelihood function can be written as

$$\begin{aligned}
 \ell(\Theta) &= \sum_{t=1}^n \log \Pr(X_t, Y_t | X_{t-1}, Y_{t-1}) \\
 &= \sum_{t=1}^n (\log \Pr(X_t | X_{t-1}, Y_{t-1}) + \log \Pr(Y_t | X_{t-1}, Y_{t-1})) \\
 &= \ell_x(\Theta_x) + \ell_y(\Theta_y),
 \end{aligned}
 \tag{55}$$

where  $\Theta_x = \{\alpha_1, \beta_{11}, \beta_{21}\}$  and  $\Theta_y = \{\alpha_2, \beta_{12}, \beta_{22}\}$ . Because  $X_t$  and  $Y_t$  are independent of each other given the last state  $(X_{t-1}, Y_{t-1})$ , the likelihood function can be separated into two parts,  $\ell_x$  and  $\ell_y$  respectively. Then transition probability for  $X_t$  is given by

$$\begin{aligned}
 &\Pr(X_t = z_1 | X_{t-1} = x, Y_{t-1} = y) \\
 &= \begin{cases} 1 & z_1 = x = y = 0 \\ (1 - \alpha_1)^x \beta_{21}^y & z_1 = 0 \\ f_{nb}(z_1; y, \beta_{21}) & x = 0 \text{ \& } y > 0 \\ \sum_{i=1}^{\min\{x, z_1\}} f_b(i; x, \alpha_1) f_{nb}(z_1 - i; i, \beta_{11}) & x > 0 \text{ \& } y = 0 \\ \sum_{j=1}^{z_1} \sum_{i=1}^{\min\{x, j\}} f_b(i; x, \alpha_1) f_{nb}(z_1 - i; i, \beta_{11}) f_{nb}(z_1 - j; y, \beta_{21}) & \\ +(1 - \alpha_1)^x f_{nb}(z_1; y, \beta_{21}) & x > 0 \text{ \& } y > 0 \end{cases}
 \end{aligned}$$

The one for  $Y_t$  is

**Table 3** Parameter setting for simulation

Parameters	$\lambda_{11}$	$\lambda_{12}$	$\lambda_{21}$	$\lambda_{22}$	$\mu_1$	$\mu_2$	$T$	$\mathbf{M}_0$
Values	0.3	1.2	1.3	0.4	1.1	1.2	1	(40, 50)

$$\Pr(Y_t = z_2 | X_{t-1} = x, Y_{t-1} = y) = \begin{cases} 1 & z_2 = x = y = 0 \\ (1 - \alpha_2)^y \beta_{12}^x & z_2 = 0 \\ f_{nb}(z_2; x, \beta_{12}) & x > 0 \ \& \ y = 0 \\ \sum_{i=1}^{\min\{y, z_2\}} f_b(i; y, \alpha_2) f_{nb}(z_2 - i; i, \beta_{22}) & x = 0 \ \& \ y > 0 \\ \sum_{j=1}^{z_2} \sum_{i=1}^{\min\{y, j\}} f_b(i; y, \alpha_2) f_{nb}(z_2 - i; i, \beta_{22}) f_{nb}(z_2 - j; y, \beta_{12}) & \\ +(1 - \alpha_2)^y f_{nb}(z_2; x, \beta_{12}) & x > 0 \ \& \ y > 0 \end{cases}$$

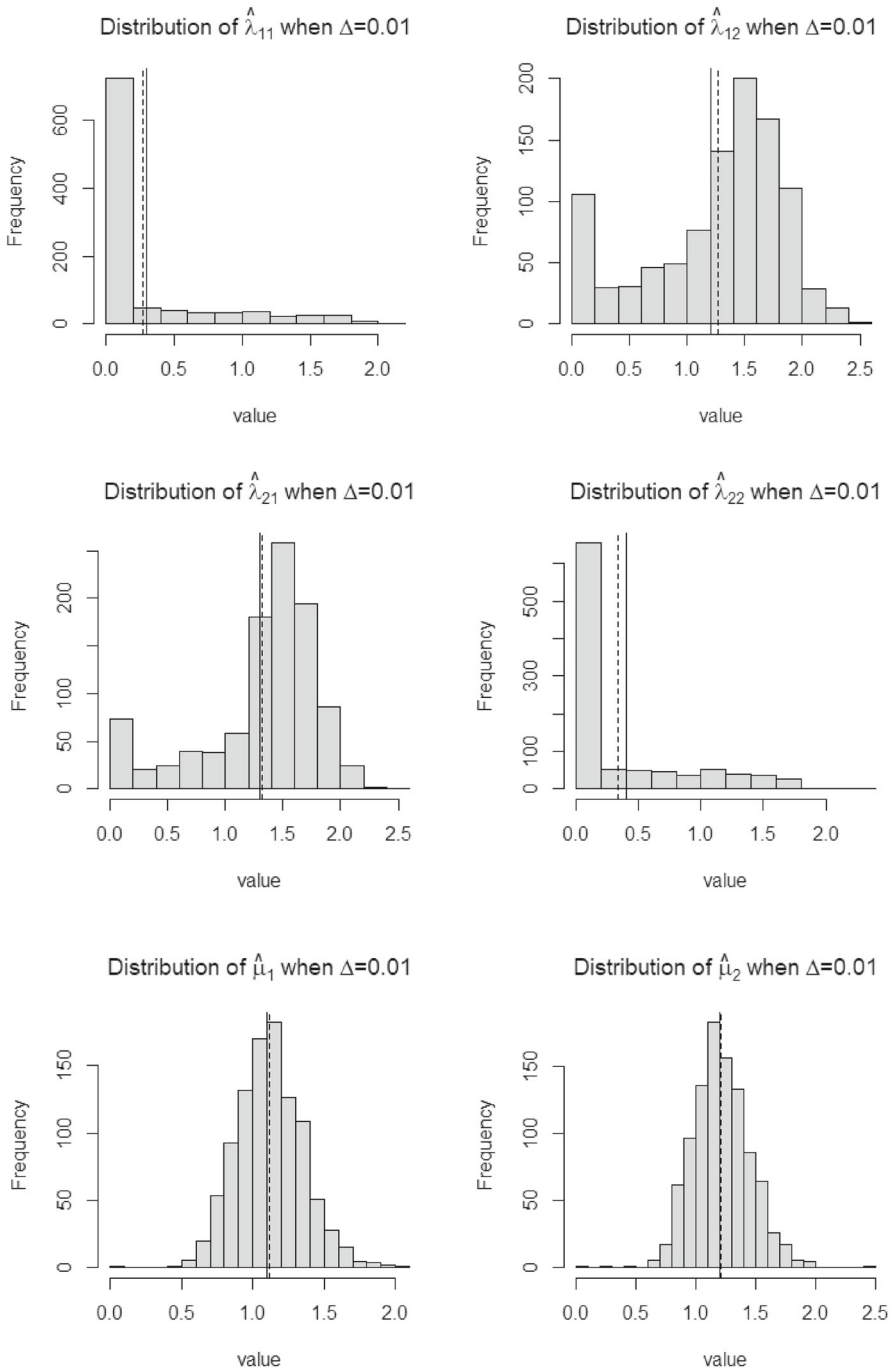
One can then numerically maximize the log likelihood function  $\ell_x, \ell_y$  given the random samples  $\{(X_0, Y_0), (X_1, Y_1), \dots, (X_n, Y_n)\}$ . From the estimated parameters  $\hat{\Theta}$ , we can solve the following system of equations to get the estimates  $\Theta_{bd} = \{\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22}, \mu_1, \mu_2\}$  for bivariate birth and death process.

$$\begin{cases} \alpha_1(\Theta_{bd}, \Delta) - \hat{\alpha}_1 = 0 \\ \alpha_2(\Theta_{bd}, \Delta) - \hat{\alpha}_2 = 0 \\ \beta_{11}(\Theta_{bd}, \Delta) - \hat{\beta}_{11} = 0 \\ \beta_{12}(\Theta_{bd}, \Delta) - \hat{\beta}_{12} = 0 \\ \beta_{21}(\Theta_{bd}, \Delta) - \hat{\beta}_{21} = 0 \\ \beta_{22}(\Theta_{bd}, \Delta) - \hat{\beta}_{22} = 0, \end{cases} \tag{56}$$

where the parametrization function  $\cdot(\Theta_{bd}, \Delta)$  are given in equation 33 and  $\Delta$  is chosen based on the interpretation of birth and death rates. For example, when the random samples are collected on daily basis over a year  $t = 1$ , one can define  $\Delta = t/365$ . Then these parameters  $\Theta_{bd}$  are interpreted on an annual scale.

In the following, we will simulate the  $r_2 = 1000$  sample paths of  $\mathbf{M}_t$  based on the pre-specific parameters in Table 3. Then equal-distance grid with sampling interval  $\Delta$  is set up and random samples  $(\mathbf{Y}_0, \mathbf{Y}_1, \dots, \mathbf{Y}_n)$  are obtained, like the way mentioned in the univariate case. Then the likelihood functions  $\ell_x, \ell_y$  are maximized by 'optim' in **R** with method being specified as 'BFGS' and the maximum likelihood estimators  $\hat{\Theta}$  are obtained. Finally, we can obtain the estimators  $\hat{\Theta}_{bd}$  by numerically solving the system of equations (56) via a root-finding algorithm (e.g. Newton–Raphson method). Referring to the estimation results in univariate case, we focus on the choices of  $\Delta \in \{0.02, 0.01, 0.005\}$  as well as large initial population (40, 50), and hopefully we can obtain asymptotic normality for each estimator. The empirical distribution of these estimators  $\Theta_{bd}$  are illustrated in Fig. 4 and their properties are summarized in Table 4.

The bias and MSE of most estimators are decreasing with respect to  $\Delta$  as expected. However, the MSE of birth rates are much larger than the estimators of death rates. Except the estimators for death rates, all other estimators for birth rate are skewed to different directions and clearly non-normal distributed. This may caused by some of non-normal estimators for proposed INAR model illustrated in Fig. 5. In the classical setting where the innovation term is included, one need stationary condition to ensure asymptotic normality for all estimators

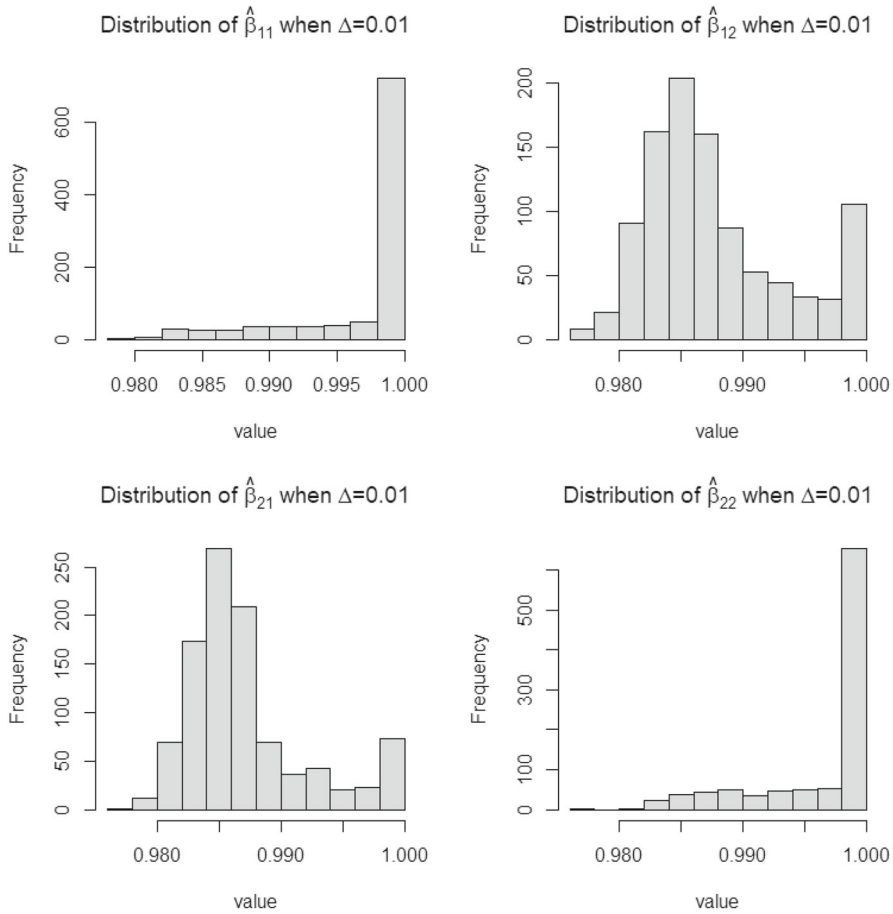


**Fig. 4** Empirical distribution estimators from bivariate INAR model. The solid lines are the true values of the parameters listed in Table 3 and the dash lines stand for empirical means

**Table 4** Properties of different maximum likelihood estimators

Bias	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\mu}_1$	$\hat{\mu}_2$	Time (s)
$\Delta = 0.02$	-0.031393	0.069314	0.031804	-0.072310	0.025835	0.012040	25.46
$\Delta = 0.01$	-0.026080	0.061711	0.019728	-0.064421	0.018005	0.012375	46.7
$\Delta = 0.005$	0.002760	0.044764	-0.004686	-0.049816	0.019450	0.010779	82.32
MSE	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\mu}_1$	$\hat{\mu}_2$	
$\Delta = 0.02$	0.266179	0.401756	0.301175	0.296895	0.104222	0.104958	
$\Delta = 0.01$	0.249248	0.359153	0.263573	0.270960	0.055740	0.058685	
$\Delta = 0.005$	0.266925	0.351415	0.268431	0.286881	0.037742	0.036170	

The time column is the total time of estimating 1 sample paths



**Fig. 5** Empirical distribution estimators from bivariate INAR model

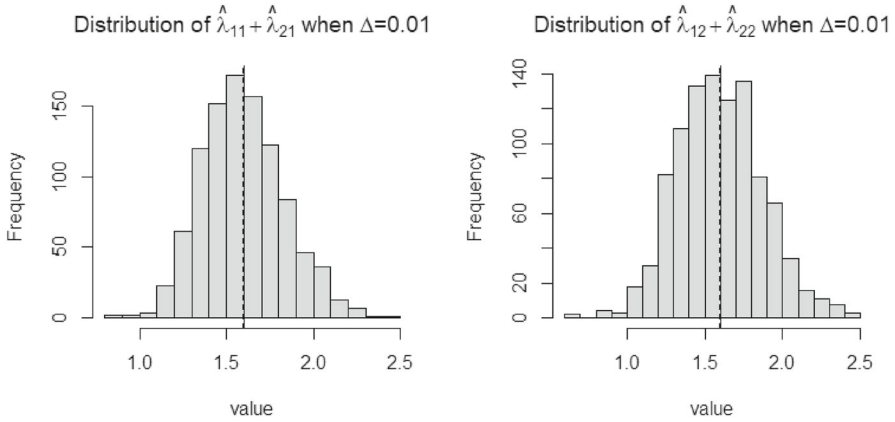
of parameters, see Bu et al. (2008). And in our case, INAR model itself is not stationary and hence some of the estimate can be skewed.

Notice that the pair of birth rates that contributed to the same population,  $(\lambda_{11}, \lambda_{21})$  and  $(\lambda_{12}, \lambda_{22})$  are skewed in opposite directions. It is then worthwhile to see whether the sum of these pair estimators has desired asymptotic properties and the results in Fig. 6 confirms our conjecture. Combining the simulation procedure of bivariate birth and death processes, Quasi-MLE method may not be able to distinguish the pair of birth rates contributed to the same population. Instead, it would provide good estimators for the scale of total birth rates  $\bar{\lambda}_1 = \hat{\lambda}_{11}r_m + \hat{\lambda}_{21}(1 - r_m)$  and  $\bar{\lambda}_2 = \hat{\lambda}_{12}r_m + \hat{\lambda}_{22}(1 - r_m)$  where  $r_m = \frac{\mathbb{E}[M_{1,t}]}{\mathbb{E}[M_{1,t} + M_{2,t}]}$ . Furthermore, according to the proof A.1, the relationship between first moment of two population is given by

$$\mathbb{E}[M_{1,t}] = c\mathbb{E}[M_{2,t}] + (M_{1,0} - cM_{2,0})e^{-(\lambda_{12}c - \kappa_2)t}. \tag{57}$$

As long as the whole process is not extinct with probability one, i.e.  $\kappa_1\kappa_2 < \lambda_{12}\lambda_{21}$ , the exponential power  $(\lambda_{12}c - \kappa_2)$  will always be positive and hence  $\mathbb{E}[M_{1,t}] \approx c\mathbb{E}[M_{2,t}]$  when  $t$  is large. In other words, the ratio

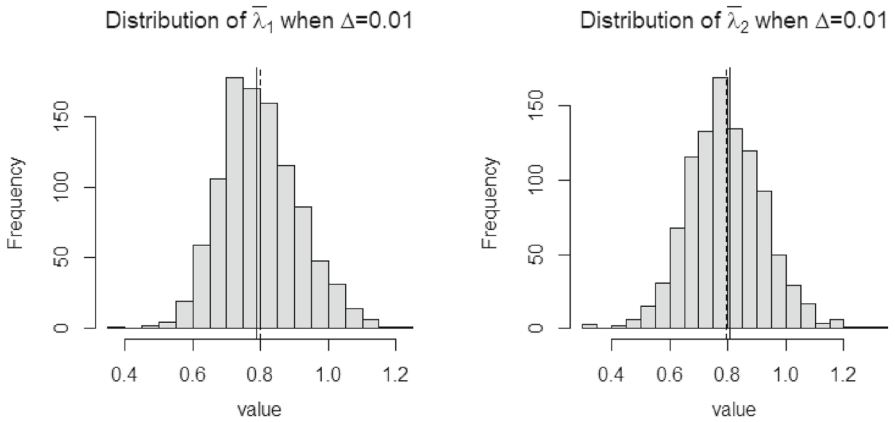




**Fig. 6** Empirical distribution for total birth rates. The solid lines are the true values of the parameters listed in Table 3 and the dash lines stand for empirical means

**Table 5** Properties for total birth rates estimators

	Bias $\bar{\lambda}_1$	MSE $\bar{\lambda}_1$	Bias $\bar{\lambda}_2$	MSE $\bar{\lambda}_2$
$\Delta = 0.02$	0.012623	0.024232	-0.015039	0.028914
$\Delta = 0.01$	0.008386	0.013068	-0.014407	0.017563
$\Delta = 0.005$	0.009380	0.009286	-0.015613	0.011202



**Fig. 7** Empirical distribution for total birth rates. The solid lines are the true values of the parameters listed in Table 3 and the dash lines stand for empirical means

**Table 6** Parameter setting for simulation

Parameters	$\lambda_{11}$	$\lambda_{12}$	$\lambda_{21}$	$\lambda_{22}$	$\mu_1$	$\mu_2$	$T$	$\mathbf{M}_0$
Values	0.3	0.5	1.3	0.4	1.1	1.2	1	(30, 60)

**Table 7** Properties of different maximum likelihood estimators

	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\mu}_1$	$\hat{\mu}_2$
<i>Bias</i>						
$\Delta = 0.02$	-0.091964	1.633439	0.064837	-0.891418	0.050226	0.008661
$\Delta = 0.01$	-0.092042	1.613498	0.056800	-0.863226	0.034508	0.02260
$\Delta = 0.005$	-0.063956	1.513447	0.036706	-0.818343	0.024054	0.016939
<i>MSE</i>						
$\Delta = 0.02$	0.243023	4.669929	0.108730	1.315634	0.115385	0.120301
$\Delta = 0.01$	0.225980	4.535412	0.090212	1.259920	0.062935	0.06831
$\Delta = 0.005$	0.243969	4.186109	0.087121	1.177241	0.042955	0.036328

**Table 8** Properties for total birth rates estimators

	Bias $\bar{\lambda}_1$	MSE $\bar{\lambda}_1$	Bias $\bar{\lambda}_2$	MSE $\bar{\lambda}_2$
$\Delta = 0.02$	0.013970	0.017193	0.007765	0.062907
$\Delta = 0.01$	0.008500	0.009268	0.017503	0.038616
$\Delta = 0.005$	0.004425	0.006884	0.012662	0.023027

$$r_m = \frac{\mathbb{E}[M_{1,t}]}{\mathbb{E}[M_{1,t} + M_{2,t}]} \rightarrow \frac{c}{1 + c}, \tag{58}$$

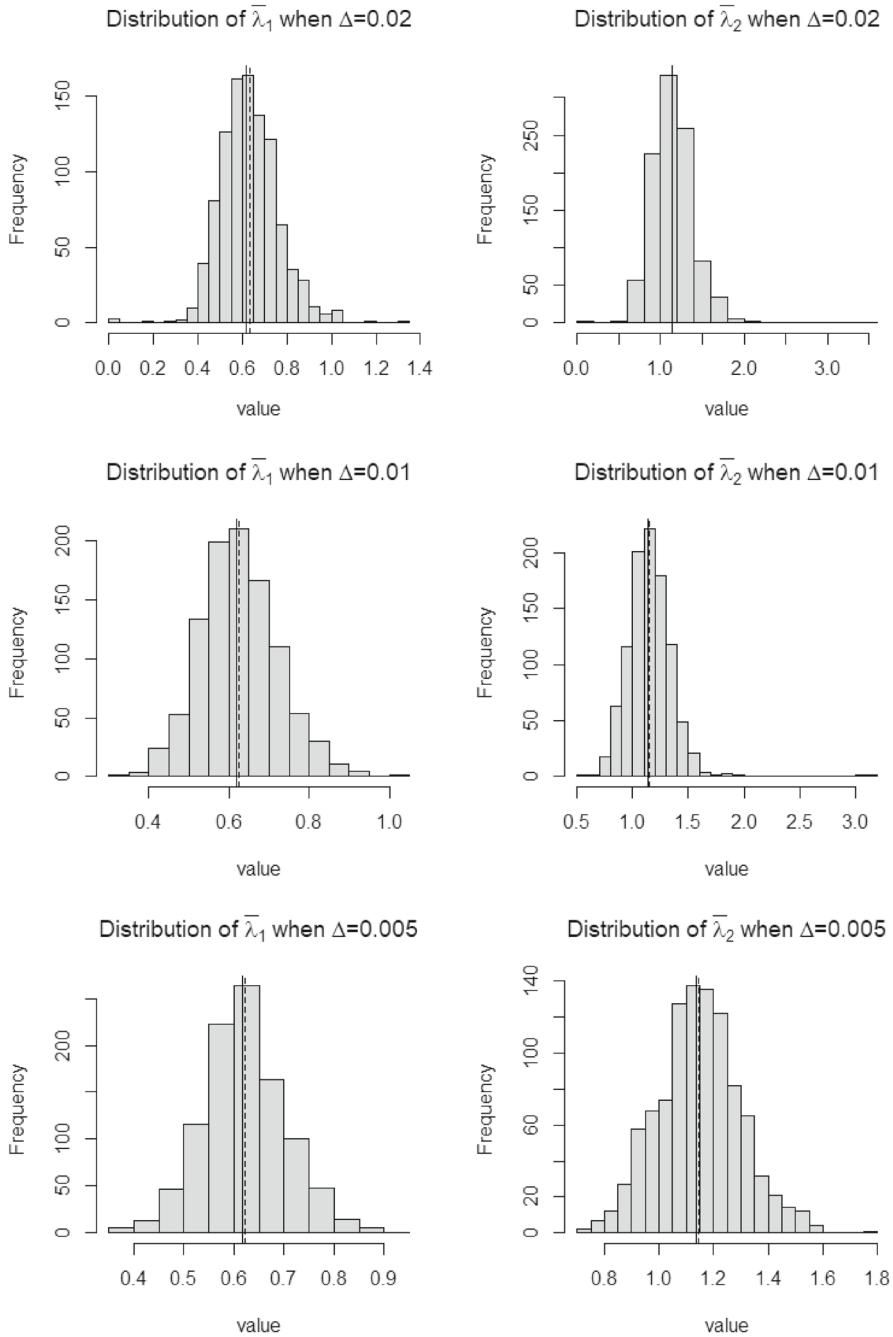
becomes a constant eventually. For the parameter setting in Table 3,  $c = 1.040833$ ,  $r_m \approx \frac{1}{2}$  and hence  $\hat{\lambda}_{11} + \hat{\lambda}_{21}$  serves as an estimator for the total birth rate of  $M_{1,t}$ . In practice, the  $c$  is unknown as true parameters need to be estimated. Then we can use the values at the end of sampling period to approximate  $r_m$ , i.e.

$$r_m \approx \frac{M_{1,T}}{M_{1,T} + M_{2,T}} \tag{59}$$

The properties of  $\bar{\lambda}_1, \bar{\lambda}_2$  and their empirical distribution are shown in Table 5 and Fig. 7. These new estimators benefits from nice properties, low bias and MSE and they decreases as  $\Delta$  decreases. Most importantly, they are not skewed anymore and asymptotic normal.

Let us try another parameter setting in Table 6 to verify this conjecture. Same simulation and estimation process as previous case and the results are shown in Tables 7, 8 and Fig. 8. This time the constant  $c$  is 0.576306 and  $r_m = 0.365605$ . Similar to the last setting, the estimators for all birth rates are skewed and some of them have large bias and MSE. The estimators for total birth rate, on the other hand, are of low bias and MSE and they are again asymptotic normal.

Let us finally try another parameter setting in Table 9 where the  $M_t$  is going to be extinct eventually. It means that the exponential function in Eq. (57) can no longer be omitted. The results are illustrated in Table 10 and Fig. 9 and they look similar to the results of the first case. Nice properties for death rates' estimators but skewed and non-normal for birth rates' estimators.



**Fig. 8** Empirical distribution for total birth rates. The solid lines are the true values of the parameters listed in Table 3 and the dash lines stand for empirical means

**Table 9** Parameter setting for simulation

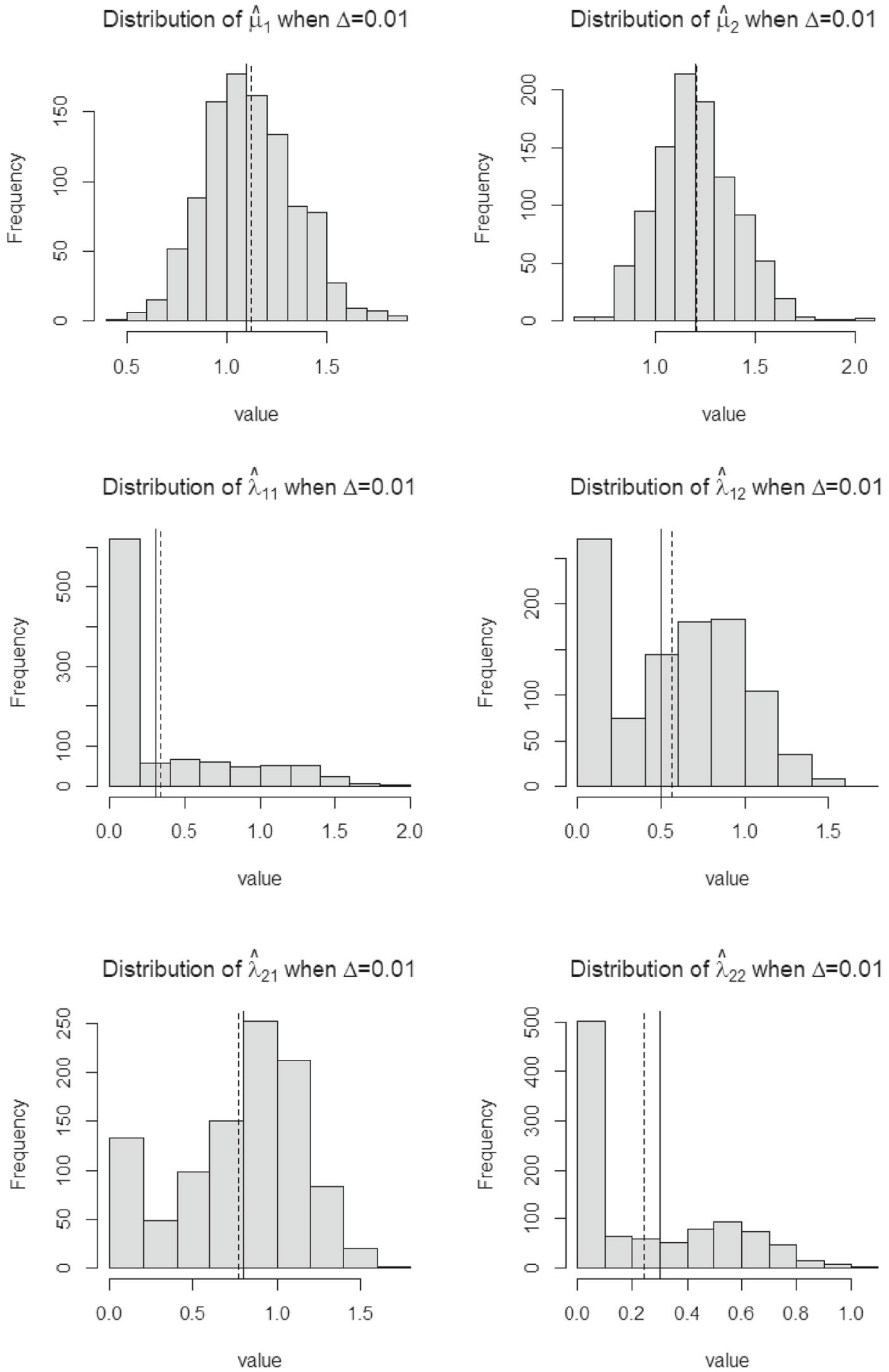
Parameters	$\lambda_{11}$	$\lambda_{12}$	$\lambda_{21}$	$\lambda_{22}$	$\mu_1$	$\mu_2$	$T$	$\mathbf{M}_0$
Values	0.3	0.5	0.8	0.3	1.1	1.2	1	(30, 60)

**Table 10** Properties of different maximum likelihood estimators

	$\hat{\lambda}_{11}$	$\hat{\lambda}_{12}$	$\hat{\lambda}_{21}$	$\hat{\lambda}_{22}$	$\hat{\mu}_1$	$\hat{\mu}_2$
<i>Bias</i>						
$\Delta = 0.02$	0.026802	0.071221	-0.020855	-0.064399	0.015807	0.007330
$\Delta = 0.01$	0.033741	0.063121	-0.024328	-0.056261	0.019261	0.008451
$\Delta = 0.005$	0.032433	0.061202	-0.022955	-0.046396	0.016784	0.011859
<i>MSE</i>						
$\Delta = 0.02$	0.245305	0.184363	0.165970	0.091334	0.088017	0.059102
$\Delta = 0.01$	0.226033	0.167176	0.150903	0.081478	0.052522	0.039980
$\Delta = 0.005$	0.211617	0.155803	0.139283	0.078867	0.042573	0.035270

## 6 Concluding remarks

In this paper, we propose an integer-valued autoregressive model INAR(1) to approximate the continuous birth-and-death process. In univariate case, we propose a birth-death operator  $p *_{1} \alpha \circ X$  which is the sum of zero-modified geometric random variable. The parametrization of  $p$  and  $\alpha$  can be determined by matching the first and second moment of continuous process. Then we propose an bivariate INAR(1) model to approximate bivariate birth and death process where birth probabilities will also depend on the size of the other population. The parametrization of this model can be obtained in a similar way. The convergence from discrete process to continuous process is proved by apply weak convergence theorem of locally bounded semimartingales. Due to the simple Markov structure of INAR(1) model, maximum likelihood estimation would be feasible. It is however not the case for bivariate and multivariate birth and death process. Basically, one can extend the result here to multivariate case, i.e. we can approximate multivariate birth and death process in Griffiths (1973) by multivariate INAR(1) model using the these operators  $*_{1}, *_{2}, \circ$  only as well as adding an immigrant process. However, the difficulty of expressing the parameters of INAR(1) model in terms of the parameters of multivariate birth and death process would be increasing and as we need to find out the first moment of birth and process explicitly.



**Fig. 9** Empirical distribution for individual birth and death rates. The solid lines are the true values of the parameters listed in Table 9 and the dash lines stand for empirical means

**Acknowledgements** The authors would like to thank the anonymous referees for their very helpful comments and suggestions which have significantly improved this article.

## Declarations

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix A: Proofs

### A.1 Proof of proposition 1

Similar to univariate case defined in (6), the infinitesimal generator of the bivariate birth and death process  $(M_{1,t}, M_{2,t})$  acting on a function  $f(t, M_1, M_2)$  within its domain  $\Omega(\mathcal{A})$  is given by

$$\begin{aligned} \mathcal{A}f(t, M_1, M_2) = & \frac{\partial f}{\partial t} + (\lambda_{11}M_1 + \lambda_{21}M_2)(f(t, M_1 + 1, M_2) - f(t, M_1, M_2)) \\ & + \mu_1M_1(f(t, M_1 - 1, M_2) - f(t, M_1, M_2)) \\ & + (\lambda_{12}M_1 + \lambda_{22}M_2)(f(t, M_1, M_2 + 1) - f(t, M_1, M_2)) \\ & + \mu_2M_2(f(t, M_1, M_2 - 1) - f(t, M_1, M_2)), \end{aligned}$$

where  $\Omega(\mathcal{A})$  is the domain for the generator  $\mathcal{A}$  such that  $f(t, M_1, M_2)$  is differentiable with respect to  $t$  for all  $t, M_1, M_2$  and

$$\begin{aligned} |f(t, M_1 + 1, M_2) - f(t, M_1, M_2)| < \infty, & \quad |f(t, M_1 - 1, M_2) - f(t, M_1, M_2)| < \infty \\ |f(t, M_1, M_2 + 1) - f(t, M_1, M_2)| < \infty, & \quad |f(t, M_1, M_2 - 1) - f(t, M_1, M_2)| < \infty \end{aligned}$$

Apply infinitesimal generator  $\mathcal{A}$  to functions  $f(t, M_1, M_2) = M_{1,t}, M_{2,t}, M_{1,t}^2, M_{2,t}^2$  and  $M_{1,t}M_{2,t}$  respectively, we have

$$\begin{aligned} \mathcal{A}M &= (\lambda_{11}M_1 + \lambda_{21}M_2)(M_1 + 1 - M_1) + \mu_1M_1(M_1 - 1 - M_1) \\ \mathcal{A}N &= (\lambda_{12}M_1 + \lambda_{22}M_2)(M_2 + 1 - M_2) + \mu_2M_2(M_2 - 1 - M_2) \\ \mathcal{A}M_1^2 &= (\lambda_{11}M_1 + \lambda_{21}M_2)((M_1 + 1)^2 - M_1^2) + \mu_1M_1((M_1 - 1)^2 - M_1^2) \\ \mathcal{A}M_2^2 &= (\lambda_{12}M_1 + \lambda_{22}M_2)((M_2 + 1)^2 - M_2^2) + \mu_2M_2((M_2 - 1)^2 - M_2^2) \\ \mathcal{A}M_1M_2 &= (\lambda_{11}M_1 + \lambda_{21}M_2)((M_1 + 1)N - M_1M_2) \\ & \quad + (\lambda_{12}M_1 + \lambda_{22}M_2)(M_1(M_2 + 1) - M_1M_2) \\ & \quad + \mu_1M_1((M_1 - 1)M_2 - M_1M_2) + \mu_2(M_1(M_2 - 1) - M_1M_2) \end{aligned}$$

The first two result in the following system of ODE

$$\begin{aligned} \frac{d}{dt} \mathbb{E}[M_{1,t}] &= \lambda_{21} \mathbb{E}[M_{2,t}] - \kappa_1 \mathbb{E}[M_{1,t}] \\ \frac{d}{dt} \mathbb{E}[M_{2,t}] &= \lambda_{12} \mathbb{E}[M_{1,t}] - \kappa_2 \mathbb{E}[M_{2,t}] \end{aligned} \tag{A1}$$

The other three equations become (17), which is hard to solve explicitly as we need to solve an inhomogeneous ordinary differential equation system. To solve the system (A1), we can first assume a linear relationship between  $\mathbb{E}[M_{1,t}]$  and  $\mathbb{E}[M_{2,t}]$  such that  $\mathbb{E}[M_{1,t}] = c\mathbb{E}[M_{2,t}] + g(t)$  for some constant  $c$  and a real value function  $g$ . Applying this substitution, the first ODE in (A1) becomes

$$c \frac{d}{dt} \mathbb{E}[M_{2,t}] + g'(t) = \lambda_{21} \mathbb{E}[M_{2,t}] - \kappa_1 (c\mathbb{E}[M_{2,t}] + g(t))$$

Make use of the second the ODE in, the first ODE can be rearranged into a ordinary equation.

$$\begin{aligned} c(\lambda_{12} \mathbb{E}[M_{1,t}] - \kappa_2 \mathbb{E}[M_{2,t}]) + g'(t) &= \lambda_{21} \mathbb{E}[M_{2,t}] - \kappa_1 \mathbb{E}[M_{1,t}] \\ (\lambda_{12}c + \kappa_1) \mathbb{E}[M_{1,t}] - (c\kappa_2 + \lambda_{21}) \mathbb{E}[M_{2,t}] + g'(t) &= 0 \\ (\lambda_{12}c + \kappa_1)(c\mathbb{E}[M_{2,t}] + g(t)) - (c\kappa_2 + \lambda_{21}) \mathbb{E}[M_{2,t}] + g'(t) &= 0 \\ (\lambda_{12}c^2 + (\kappa_1 - \kappa_2)c - \lambda_{21}) \mathbb{E}[M_{2,t}] + g'(t) + (\lambda_{12} + \kappa_1)g(t) &= 0 \end{aligned}$$

Then  $c$  is the solution of the quadratic equation

$$\begin{aligned} \lambda_{12}c^2 + (\kappa_1 - \kappa_2)c - \lambda_{21} &= 0 \\ c &= \frac{\kappa_2 - \kappa_1 \pm \sqrt{(\kappa_1 - \kappa_2)^2 + 4\lambda_{21}\lambda_{12}}}{2\lambda_{12}} \end{aligned}$$

Both roots would result in the same moments, so just take the positive root. The function  $g$  would be the solution of following ODE

$$\begin{aligned} g'(t) + (\lambda_{12} + \kappa_1)g(t) &= 0 \\ g(t) &= g(0)e^{-(\lambda_{12}c - \kappa_2)t} \\ g(0) &= \mathbb{E}[M_{1,0}] - c\mathbb{E}[M_{2,0}] = (a - cb) \end{aligned}$$

Then  $\mathbb{E}[M_{2,t}]$  is determined by the following ODE

$$\frac{d}{dt} \mathbb{E}[M_{2,t}] = \lambda_{12} \mathbb{E}[M_{1,t}] - \kappa_2 \mathbb{E}[M_{2,t}] = (\lambda_{12}c - \kappa_2) \mathbb{E}[M_{2,t}] + \lambda_{12}g(t)$$

and  $\mathbb{E}[M_{1,t}] = c\mathbb{E}[M_{2,t}] + g(t)$  □

### A.2 Proof of proposition 2

For the first property, we can verify in the following way.

When  $i = 1$

$$\begin{aligned} d_0 &= p^0 \left( 1 + (1 - p) \frac{\frac{\alpha}{p} - \left(\frac{\alpha}{p}\right)}{1 - \frac{\alpha}{p}} \right) = 1 \\ p_1 &= p, \quad \alpha_1 = \alpha \end{aligned}$$

Suppose equation (22) holds for  $i = k$ . Then for  $i = k + 1$ , we have

$$\begin{aligned} \phi^{(I)}(t, \theta) &= \mathbb{E} \left[ \left( 1 - \alpha_k + \frac{\alpha_k p_k \theta}{1 - (1 - p_k)\theta} \right)^{X_{t-k}} \right] \\ &= \mathbb{E} \left[ \left( \frac{1 - \alpha_k - (1 - \alpha_k - p_k)\theta}{1 - (1 - p_k)\theta} \right)^{X_{t-k}} \right] \\ &= \mathbb{E} \left[ \left( \frac{1 - \alpha - (1 - \alpha - p) \frac{1 - \alpha_k - (1 - \alpha_k - p_k)\theta}{1 - (1 - p_k)\theta}}{1 - (1 - p) \frac{1 - \alpha_k - (1 - \alpha_k - p_k)\theta}{1 - (1 - p_k)\theta}} \right)^{X_{t-k-1}} \right] \\ &= \mathbb{E} \left[ \left( \frac{\frac{1 - \alpha - (1 - \alpha - p)(1 - \alpha_k)}{1 - (1 - p)(1 - \alpha_k)} - \left( \frac{(1 - p_k + (1 - p)(\alpha_k - p_k - 1))}{1 - (1 - p)(1 - \alpha_k)} - \frac{\alpha \alpha_k}{1 - (1 - p)(1 - \alpha_k)} \right) \theta}{1 - \frac{(1 - p_k + (1 - p)(\alpha_k - p_k - 1))}{1 - (1 - p)(1 - \alpha_k)} \theta} \right)^{X_{t-k-1}} \right] \end{aligned}$$

It is then clear that

$$\begin{aligned} 1 - (1 - p)(1 - \alpha_k) &= \frac{d_{k-1} - (1 - p)(d_{k-1} - \alpha^k)}{d_{k-1}} = \frac{d_k}{d_{k-1}} \\ \alpha_{k+1} &= \frac{\alpha \alpha_k}{1 - (1 - p)(1 - \alpha_k)} = \frac{\alpha^{k+1}}{d_k} \\ p_{k+1} &= 1 - \frac{(1 - p_k + (1 - p)(\alpha_k - p_k - 1))}{1 - (1 - p)(1 - \alpha_k)} \\ &= 1 - \frac{1 - (1 - p)(1 - \alpha_k) - pp_k}{1 - (1 - p)(1 - \alpha_k)} = \frac{pp_k}{1 - (1 - p)(1 - \alpha_k)} = \frac{p^{k+1}}{d_k} \end{aligned}$$

So Eq. (22) holds for all  $i = 1, 2, \dots$ . For the second property, the moments can be found by conditional expectation such that

$$\begin{aligned} &\mathbb{E}[p * \alpha \circ X | X] \\ &= \mathbb{E} \left[ \mathbb{E} \left[ \sum_{j=1}^{\alpha \circ X} g_j^{(j)} | \alpha \circ X \right] | X \right] = \frac{1}{p} \mathbb{E}[(\alpha \circ X) | X] = \frac{1}{p} \mathbb{E} \left[ \sum_{j=1}^X b_j | X \right] \\ &= \frac{\alpha}{p} X \\ &Var(p * \alpha \circ X | X) \\ &= \mathbb{E} \left[ Var \left( \sum_{j=1}^{\alpha \circ X} g_j^{(j)} | \alpha \circ X \right) | X \right] + Var \left( \mathbb{E} \left[ \sum_{j=1}^{\alpha \circ X} g_j^{(j)} | \alpha \circ X \right] | X \right) \\ &= \frac{1 - p}{p^2} \mathbb{E}[\alpha \circ X | X] + \frac{1}{p^2} Var(\alpha \circ X | X) \\ &= \left( \frac{\alpha(1 - p)}{p^2} + \frac{\alpha(1 - \alpha)}{p^2} \right) X \\ Var(p * \alpha \circ X) &= \mathbb{E}[Var(p * \alpha \circ X | X)] + Var(\mathbb{E}[p * \alpha \circ X | X]) \\ Cov(X_t, X_{t-i}) &= Cov(p_i * \alpha_i \circ X_{t-i}, X_{t-i}) \\ &= Cov(\mathbb{E}[p_i * \alpha_i \circ X_{t-i} | X_{t-i}], \mathbb{E}[X_{t-i} | X_{t-i}]) \\ &\quad + \mathbb{E}[Cov(p_i * \alpha_i \circ X_{t-i}, X_{t-i} | X_{t-i})] \end{aligned}$$



$$= \frac{\alpha_i}{p_i} Cov(X_{t-i}, X_{t-i}) + 0$$

□

### A.3 Proof of Proposition 5

The transition probability of giving out  $m (m > 0)$  birth of from the process  $Z_t^{(\Delta)}$  given  $Z_{t-\Delta}^{(\Delta)}$  during an infinitesimal time interval  $\Delta$  is given by

$$\begin{aligned} \Pr(Z_t^{(\Delta)} = k + m | Z_{t-\Delta}^{(\Delta)} = k) &= \Pr(X_s = k + m | X_{s-1} = k) \\ &= \sum_{j=1}^{\min\{k, k+m\}} \binom{k}{j} \alpha^j (1 - \alpha)^{k-j} \binom{k+m-1}{j-1} p^j (1-p)^{k+m-j} \\ &= \sum_{j=1}^k \binom{k}{j} \binom{k+m-1}{j-1} e^{(\lambda-\mu)j\Delta} \lambda^{k+m-j} \mu^{k-j} (e^{(\lambda-\mu)\Delta} - 1)^{2k-2j+m} (\lambda - \mu)^{2j} \\ &\times (\lambda e^{(\lambda-\mu)\Delta} - \mu)^{-(2k+m)} \\ &= \sum_{j=1}^{k-1} \binom{k}{j} \binom{k+m-1}{j-1} \lambda^{k-j} \mu^{k-j} (\lambda - \mu)^{2j} ((\lambda - \mu)\Delta + o(\Delta))^{2k-2j+m} \\ &\times (1 + (\lambda - \mu)\Delta + o(\Delta))^j \left( \frac{1}{\lambda - \mu} - \frac{\lambda}{\lambda - \mu} \Delta + o(\Delta) \right)^{2k+m}, \end{aligned}$$

where all the exponential function are expressed as their corresponding Taylor expansion at  $\Delta = 0$ . To make comparison with the continuous birth and death process, we are interested in the coefficients in front of  $\Delta$ . First we need to check the lowest order of  $\Delta$  in above probability. That is, we would like to minimize the sum

$$\min_{1 \leq j \leq k} 2k - 2j + m = m$$

So for the transition probability is rearranged in the following way

$$\begin{aligned} \Pr(Z_t^{(\Delta)} = k + m | Z_{t-\Delta}^{(\Delta)} = k) &= \binom{k}{k} \binom{k+m-1}{k-1} \lambda^m (\lambda - \mu)^{2k} ((\lambda - \mu)\Delta + o(\Delta))^m (1 + (\lambda - \mu)\Delta + o(\Delta))^k \\ &\times \left( \frac{1}{\lambda - \mu} - \frac{\lambda}{\lambda - \mu} \Delta + o(\Delta) \right)^{2k+m} + o(\Delta^m) \\ &= \binom{k+m-1}{k-1} (\lambda\Delta)^m + o(\Delta^m), \end{aligned}$$

Then it is clear that the there is no first order term in the case where  $m \geq 2$  and the probability that giving out exactly one birth is

$$\Pr(Z_t^{(\Delta)} = k + 1 | Z_{t-\Delta}^{(\Delta)} = k) = \lambda k \Delta + o(\Delta)$$

On the other hand, we can derive the probability that  $m (1 \leq m \leq k)$  individuals die within infinitesimal time  $\Delta$  in a similar way

$$\begin{aligned}
 \Pr(Z_t^{(\Delta)} = k - m | Z_{t-\Delta}^{(\Delta)} = k) &= \Pr(X_s = k - m | X_{s-1} = k) \\
 &= \sum_{j=1}^{\min\{k-m, k\}} \binom{k}{j} \alpha^j (1 - \alpha)^{k-j} \binom{k-m-1}{j-1} p^j (1 - p)^{k-m-j} \\
 &= \sum_{j=1}^{k-m} \binom{k}{j} \binom{k-m-1}{j-1} e^{(\lambda-\mu)j\Delta} \lambda^{k-m-j} \mu^{k-j} (e^{(\lambda-\mu)\Delta} - 1)^{2k-2j-m} \\
 &\quad \times (\lambda - \mu)^{2j} (\lambda e^{(\lambda-\mu)\Delta} - \mu)^{-(2k-m)} \\
 &= \sum_{j=1}^{k-m} \binom{k}{j} \binom{k-m-1}{j-1} \lambda^{k-m-j} \mu^{k-j} (\lambda - \mu)^{2j} ((\lambda - \mu)\Delta + o(\Delta))^{2k-2j-m} \\
 &\quad \times (1 + (\lambda - \mu)\Delta + o(\Delta))^j \left( \frac{1}{\lambda - \mu} - \frac{\lambda}{\lambda - \mu} \Delta + o(\Delta) \right)^{2k-m}
 \end{aligned}$$

The minimum order of  $\Delta$  is determined by

$$\min_{1 \leq j \leq k-m} 2k - 2j - m = m \quad (\text{A2})$$

Then the probability is reduced to

$$\begin{aligned}
 \Pr(Z_t^{(\Delta)} = k - m | Z_{t-\Delta}^{(\Delta)} = k) &= \binom{k}{k-m} \binom{k-m-1}{k-m-1} \mu^m (\lambda - \mu)^{2(k-m)} ((\lambda - \mu)\Delta + o(\Delta))^m \\
 &= (1 + (\lambda - \mu)\Delta + o(\Delta))^{k-m} \\
 &\quad \times \left( \frac{1}{\lambda - \mu} - \frac{\lambda}{\lambda - \mu} \Delta + o(\Delta) \right)^{2k-m} + o(\Delta^m) \\
 &= \binom{k}{k-m} (\mu\Delta)^m + o(\Delta^m)
 \end{aligned} \quad (\text{A3})$$

The transition probability that only one individual dies is

$$\Pr(Z_t^{(\Delta)} = k - 1 | Z_{t-\Delta}^{(\Delta)} = k) = \mu k \Delta + o(\Delta) \quad (\text{A4})$$

the probability that more than one individuals die are  $o(\Delta)$ . As the birth rate and death ( $\lambda, \mu$ ) are time-homogeneous, so as the parameters  $\alpha, p$ , the transition probabilities stay the same for all time  $t \in [t_1, t_2]$ . This means that the discrete birth and death INAR(1) model would result in the the same dynamic (1) of simple birth and death process when  $\Delta$  is small enough.

Similar to the univariate case. It is necessary to find out the transition probabilities before proceeding to the weak convergence. The transition probability of giving out  $m$  ( $m \geq 1$ ) births of population  $M_t^{(\Delta)}$  given  $M_{t-\Delta}^{(\Delta)} = k_1, N_{t-\Delta}^{(\Delta)} = k_2$  during an infinitesimal time  $\Delta$  is given by

$$\begin{aligned}
 \Pr(M_t^{(\Delta)} = k_1 + m | M_{t-\Delta}^{(\Delta)} = k_1, N_{t-\Delta}^{(\Delta)} = k_2) &= \Pr(X_s = k_1 + m | X_{s-1} = k_1, Y_{s-1} = k_2) \\
 &= \sum_{j=0}^{k_1+m} \Pr(\beta_{11} * \alpha_1 \circ X_{s-1} = j | X_{s-1} = k_1) \Pr(\beta_{21} * \alpha_2 Y_{s-1} = k_1 + m - j | Y_{s-1} = k_2) \\
 &= \sum_{j=1}^{k_1+m} \left( \sum_{i=1}^{\min\{k_1, j\}} \binom{k_1}{i} \alpha_1^i (1 - \alpha_1)^{k_1-i} \binom{j-1}{i-1} \beta_{11}^i (1 - \beta_{11})^{j-i} \right) \binom{k_2 + k_1 + m - j - 1}{k_2 - 1} \\
 &\quad \times \beta_{21}^{k_2} (1 - \beta_{21})^{k_1+m-j} + (1 - \alpha_1)^{k_1} \binom{k_2 + k_1 + m - 1}{k_2 - 1} \beta_{21}^{k_2} (1 - \beta_{21})^{k_1+m}
 \end{aligned}$$

$$\begin{aligned}
 &= \sum_{j=1}^{k_1+m} \sum_{i=1}^{\min\{k_1, j\}} \binom{k_1}{i} \binom{j-1}{i-1} \binom{k_2+k_1+m-j-1}{k_2-1} \left( \frac{(\lambda_{11}-\mu_1)\omega_1(\Delta)}{\lambda_{11}\omega_1(\Delta)-\mu_1} \right)^i \\
 &\quad \times \left( \frac{\mu_1(\omega_1(\Delta)-1)}{\lambda_{11}\omega_1(\Delta)-\mu_1} \right)^{k_1-i} \left( \frac{\lambda_{11}-\mu_1}{\lambda_{11}\omega_1(\Delta)-\mu_1} \right)^i \left( \frac{\lambda_{11}(\omega_1(\Delta)-1)}{\lambda_{11}\omega_1(\Delta)-\mu_1} \right)^{j-i} \\
 &\quad \times \left( \frac{1}{1+C_{\beta_1}(e^{u_1\Delta}-e^{u_2\Delta})} \right)^{k_2} \left( \frac{C_{\beta_1}(e^{u_1\Delta}-e^{u_2\Delta})}{1+C_{\beta_1}(e^{u_1\Delta}-e^{u_2\Delta})} \right)^{k_1+m-j} + o(\Delta^{2k_1+m-1}) \\
 &= \sum_{j=1}^{k_1+m} \sum_{i=1}^{\min\{k_1, j\}} \binom{k_1}{i} \binom{j-1}{i-1} \binom{k_2+k_1+m-j-1}{k_2-1} (\lambda_{11}-\mu_1)^i (1+(\lambda_{11}-\mu_1)\Delta+o(\Delta))^i \\
 &\quad \times \mu_1^{k_1-i} ((\lambda_{11}-\mu_1)\Delta+o(\Delta))^{k_1-i} (\lambda_{11}-\mu_1)^i \lambda_{11}^{j-i} ((\lambda_{11}-\mu_1)\Delta+o(\Delta))^{j-i} \\
 &\quad \times \left( \frac{1}{\lambda_{11}-\mu_1} - \frac{\lambda_{11}}{\lambda_{11}-\mu_1} \Delta + o(\Delta) \right)^{k_1+j} \\
 &\quad \times (1-C_{\beta_1}(u_1-u_2)\Delta+o(\Delta))^{k_2} (C_{\beta_1}(u_1-u_2)\Delta+o(\Delta))^{k_1+m-j} + o(\Delta^{2k_1+m-1}) \\
 &= \sum_{j=1}^{k_1+m} \sum_{i=1}^{\min\{k_1, j\}} \binom{k_1}{i} \binom{j-1}{i-1} \binom{k_2+k_1+m-j-1}{k_2-1} (\lambda_{11}-\mu_1)^{2i} \lambda_{11}^{j-i} \mu_1^{k_1-i} \\
 &\quad \times ((\lambda_{11}-\mu_1)\Delta+o(\Delta))^{k_1+j-2i} (1+(\lambda_{11}-\mu_1)\Delta+o(\Delta))^i \\
 &\quad \times \left( \frac{1}{\lambda_{11}-\mu_1} - \frac{\lambda_{11}}{\lambda_{11}-\mu_1} \Delta + o(\Delta) \right)^{k_1+j} \\
 &\quad \times (1-C_{\beta_1}(u_1-u_2)\Delta+o(\Delta))^{k_2} (C_{\beta_1}(u_1-u_2)\Delta+o(\Delta))^{k_1+m-j} + o(\Delta^{2k_1+m-1}),
 \end{aligned}$$

where all the exponential function are expressed in their corresponding Taylor expansion at  $\Delta = 0$ . The lowest order of  $\Delta$  is determined by the power of  $((\lambda_{11}-\mu_1)+o(\Delta))$  and  $(C_{\beta_1}(u_1-u_2)\Delta+o(\Delta))$ ,

$$\min_{1 \leq i \leq \min\{k_1, j\}} k_1 + j - 2i + k_1 + m - j = \min_{1 \leq i \leq \min\{k_1, j\}} 2k_1 - 2i + m = m,$$

where  $j \in \{1, \dots, k_1 + m\}$ . This leads to  $j = k_1, i = k_1 - 1$  and  $j = k_1 - 1, i = k_1 - 1$ , respectively. Then the transition probability reduces to

$$\begin{aligned}
 \Pr(M_t^{(\Delta)} = k_1 + m | M_{t-\Delta}^{(\Delta)} = k_1, N_{t-\Delta}^{(\Delta)} = k_2) &= \sum_{j=k_1}^{k_1+m} \binom{k_1}{k_1} \binom{j-1}{k_1-1} \binom{k_2+k_1+m-j-1}{k_2-1} (\lambda_{11}-\mu_1)^{2k_1} \lambda_{11}^{j-k_1} \mu_1^{k_1-k_1} \\
 &\quad \times ((\lambda_{11}-\mu_1)\Delta+o(\Delta))^{k_1+j-2k_1} (1+(\lambda_{11}-\mu_1)\Delta+o(\Delta))^{k_1} \\
 &\quad \times \left( \frac{1}{\lambda_{11}-\mu_1} - \frac{\lambda_{11}}{\lambda_{11}-\mu_1} \Delta + o(\Delta) \right)^{k_1+j} \\
 &\quad \times (1-C_{\beta_1}(u_1-u_2)\Delta+o(\Delta))^{k_2} (C_{\beta_1}(u_1-u_2)\Delta+o(\Delta))^{k_1+m-j} + o(\Delta^{2k_1+m-1}), \\
 &= \sum_{j=k_1}^{k_1+m} \binom{j-1}{k_1-1} \binom{k_2+k_1+m-j-1}{k_2-1} (\lambda_{11}\Delta)^{j-k_1} (C_{\beta_1}(u_1-u_2)\Delta+o(\Delta))^{k_1+m-j} \\
 &\quad + o(\Delta^{2k_1+m-1}), \\
 &= \sum_{j=k_1}^{k_1+m} \binom{j-1}{k_1-1} \binom{k_2+k_1+m-j-1}{k_2-1} (\lambda_{11}\Delta)^{j-k_1} (\lambda_{21}\Delta)^{k_1+m-j} + o(\Delta^m),
 \end{aligned}$$

Then the probability when  $m = 1$  is given by

$$\begin{aligned} \Pr(M_t^{(\Delta)} = k_1 + 1 | M_{t-\Delta}^{(\Delta)} = k_1, N_{t-\Delta}^{(\Delta)} = k_2) &= \sum_{j=k_1}^{k_1+1} \binom{j-1}{k_1-1} \binom{k_2+k_1-j}{k_2-1} (\lambda_{11}\Delta)^{j-k_1} (\lambda_{21}\Delta)^{k_1+1-j} + o(\Delta), \\ &= k_2\lambda_{21}\Delta + k_1\lambda_{11}\Delta + o(\Delta) \end{aligned} \tag{A5}$$

By symmetry, the birth transition probability for the other population is given by

$$\begin{aligned} \Pr(N_t^{(\Delta)} = k_2 + m | M_{t-\Delta}^{(\Delta)} = k_1, N_{t-\Delta}^{(\Delta)} = k_2) &= P(Y_s = k_2 | X_{s-1} = k_1, Y_{s-1} = k_2) \\ &= \sum_{j=k_2}^{k_2+m} \binom{j-1}{k_2-1} \binom{k_1+k_2+m-j-1}{k_1-1} (\lambda_{22}\Delta)^{j-k_2} (\lambda_{12}\Delta)^{k_1+m-j} + o(\Delta^m). \end{aligned}$$

On the other hand, the probability that  $m$  individual die in population  $M_t^{(\Delta)}$  given that  $M_{t-\Delta}^{(\Delta)} = k_1, N_{t-\Delta}^{(\Delta)} = k_2$  during an infinitesimal time interval  $\Delta$  is

$$\begin{aligned} \Pr(M_t^{(\Delta)} = k_1 - m | M_{t-\Delta}^{(\Delta)} = k_1, N_{t-\Delta}^{(\Delta)} = k_2) &= \Pr(X_s = k_1 - m | X_{s-1} = k_1, Y_{s-1} = k_2) \\ &= \sum_{j=1}^{k_1-m} \left( \sum_{i=1}^{\min\{k_1, j\}} \binom{k_1}{i} \alpha_1^i (1-\alpha_1)^{k_1-i} \binom{j-1}{i-1} \beta_{11}^i (1-\beta_{11})^{j-i} \right) \\ &\quad \times \binom{k_2+k_1-m-j-1}{k_2-1} \beta_{21}^{k_2} (1-\beta_{21})^{k_1-j} + (1-\alpha_1)^{k_1} \binom{k_2+k_1-m-1}{k_2-1} \beta_{21}^{k_2} (1-\beta_{21})^{k_1} \\ &= \sum_{j=1}^{k_1-m} \sum_{i=1}^j \binom{k_1}{i} \binom{j-1}{i-1} \binom{k_2+k_1-m-j-1}{k_2-1} (\lambda_{11}-\mu_1)^{2i} \lambda_{11}^{j-i} \mu_1^{k_1-i} \\ &\quad \times ((\lambda_{11}-\mu_1)\Delta + o(\Delta))^{k_1+j-2i} \\ &\quad \times (1 + (\lambda_{11}-\mu_1)\Delta + o(\Delta))^j \left( \frac{1}{\lambda_{11}-\mu_1} - \frac{\lambda_{11}}{\lambda_{11}-\mu_1} \Delta + o(\Delta) \right)^{k_1+j} \\ &\quad \times (1 - C_{\beta_1}(u_1 - u_2)\Delta + o(\Delta))^{k_2} (C_{\beta_1}(u_1 - u_2)\Delta + o(\Delta))^{k_1-m-j} + o(\Delta^{2k_1+k_2-1}) \\ &= \binom{k_1+1}{k_1} \binom{k_1-1}{k_1-1} \binom{k_2-1}{k_2-1} (\lambda_{11}-\mu_1)^{2k_1} \mu_1 ((\lambda_{11}-\mu_1)\Delta + o(\Delta)) \\ &\quad \times (1 + (\lambda_{11}-\mu_1)\Delta + o(\Delta))^{k_1} \\ &\quad \times \left( \frac{1}{\lambda_{11}-\mu_1} - \frac{\lambda_{11}}{\lambda_{11}-\mu_1} \Delta + o(\Delta) \right)^{2k_1+1} (1 - C_{\beta_1}(u_1 - u_2)\Delta + o(\Delta))^{k_2} + o(\Delta) \\ &= \mu_1(k_1+1)\Delta + o(\Delta) \end{aligned}$$

The lowest order of  $\Delta$  is determined by the power of  $((\lambda_{11}-\mu_1) + o(\Delta))$  and  $(C_{\beta_1}(u_1 - u_2)\Delta + o(\Delta))$ , which is

$$\min_{1 \leq i \leq j} k_1 + j - 2i + k_1 - m - j = \min_{1 \leq i \leq j} 2k_1 - 2i - m = m,$$

where  $j \in \{1, \dots, k_1 - m\}$ . So the above probability reduces to

$$\begin{aligned} &\Pr(M_t^{(\Delta)} = k_1 - m | M_{t-\Delta}^{(\Delta)} = k_1, N_{t-\Delta}^{(\Delta)} = k_2) \\ &= \binom{k_1}{k_1 - m} \binom{k_1 - m - 1}{k_1 - m - 1} \binom{k_2 - 1}{k_2 - 1} (\lambda_{11} - \mu_1)^{2(k_1 - m)} \mu_1^m \\ &\quad \times ((\lambda_{11} - \mu_1)\Delta + o(\Delta))^m \\ &\quad \times (1 + (\lambda_{11} - \mu_1)\Delta + o(\Delta))^{k_1 - m} \left( \frac{1}{\lambda_{11} - \mu_1} - \frac{\lambda_{11}}{\lambda_{11} - \mu_1} \Delta + o(\Delta) \right)^{2k_1 - m} \\ &\quad \times (1 - C\beta_1(u_1 - u_2)\Delta + o(\Delta))^{k_2} + o(\Delta^m) \\ &= \binom{k_1}{k_1 - m} (\mu_1 \Delta)^m + o(\Delta^m) \end{aligned}$$

It is not surprise that the transition probability only depends on its own size of population from the bivariate INAR construction. Then the death transition probability for the other population is

$$\Pr(N_t^{(\Delta)} = k_2 - m | M_{t-\Delta}^{(\Delta)} = k_1, N_{t-\Delta}^{(\Delta)} = k_2) = \binom{k_2}{k_2 - m} (\mu_2 \Delta)^m + o(\Delta^m) \tag{A6}$$

The it is clear that the probabilities for both population that there is only one death would have the same form as in the univariate case. By conditional independence of bivariate INAR model, the joint transition probability would be the product of any of two transition probabilities shown above. For example, for any two integers  $m_1, m_2 \in \mathbb{Z}$ ,

$$\begin{aligned} &\Pr(M_t^{(\Delta)} = k_1 + m_1, N_t^{(\Delta)} = k_2 + m_2 | M_{t-\Delta}^{(\Delta)} = k_1, N_{t-\Delta}^{(\Delta)} = k_2) \\ &= \Pr(M_t^{(\Delta)} = k_1 + m_1 | M_{t-\Delta}^{(\Delta)} = k_1, N_{t-\Delta}^{(\Delta)} = k_2) \Pr(N_t^{(\Delta)} = k_2 + m_2 | M_{t-\Delta}^{(\Delta)} = k_1, N_{t-\Delta}^{(\Delta)} = k_2) \\ &= \Pr(M_t^{(\Delta)} = k_1 + m_1, N_t^{(\Delta)} = k_2 + m_2 | M_{t-\Delta}^{(\Delta)} = k_1, N_{t-\Delta}^{(\Delta)} = k_2) \end{aligned} \tag{A7}$$

Then it is straightforward to show that, to have a first order  $\Delta$  term, the only possible combinations of  $(m_1, m_2)$  are  $\{(1, 0), (0, 1), (-1, 0), (0, -1)\}$ , which under the proposed parametrization, the joint process only allow one jump during infinitesimal time which coincide with the bivariate continuous birth and death process.

As the birth rates  $\lambda_{i,j}$  and death rate are  $\mu_i$  time homogeneous, so as the parameters  $\alpha_i, \beta_{i,j}, i, j \in \{1, 2\}$ , the transition probabilities stay the same for all time  $t \in [0, 1]$ . This means that this bivariate birth and death INAR(1) model would result in the the same dynamic (11) when  $\Delta$  is small enough. □

### A.4 Proof of Proposition 6

According to Theorem 3.11, Chapter 2 in Jacod and Shiryaev (2013). We need to make sure the following two sum is finite for any truncation function  $h$  before constructing their discrete Lévy measures.

$$\begin{aligned} &\sum_{k=1}^{\sigma_t} |\mathbb{E}[h(U_k) | \mathcal{F}_{k-1}]| < \infty \\ &\sum_{k=1}^{\sigma_t} \mathbb{E}[|U_k^2 \wedge 1| | \mathcal{F}_{k-1}] < \infty, \end{aligned} \tag{A8}$$

where  $U_k$  are increments of any underlying processes. This can be shown straightforwardly as there are only finite number of terms for summation  $\sigma_t \leq n$  and truncation functions are bounded. Then by Theorem 3.11, the Lévy triplets for  $Z_t^{(n)}$  is

$$Ch(Z_t^{(n)}) = \begin{cases} B_t = \sum_{k=1}^{\sigma_t} \mathbb{E}[h(U_k)|\mathcal{F}_{k-1}] \\ C_t = 0 \\ v(Z_t^{(n)}; [0, t] \times g) = \sum_{k=1}^{\sigma_t} \mathbb{E}[g(U_k)1_{\{U_k \neq 0\}}|\mathcal{F}_{k-1}] \end{cases} \tag{A9}$$

We can choose the truncation function  $h(x) = |x|1_{\{|x| < 1\}}$  such that  $B_t$  is always 0 as there is no jump with size smaller than 1 in  $Z_t^{(n)}$ . Finally for the discrete stochastic integral:

$$\begin{aligned} & \mathbb{E}[g(U_k)1_{\{U_k \neq 0\}}|\mathcal{F}_{k-1}] \\ &= \mathbb{E}[g(1)1_{\{U_k=1\}}|\mathcal{F}_{k-1}] + \mathbb{E}[g(-1)1_{\{U_k=-1\}}|\mathcal{F}_{k-1}] \\ &+ \sum_{\eta=2}^{\infty} \mathbb{E}[g(\eta)1_{\{U_k=\eta\}} + g(-\eta)1_{\{U_k=-\eta\}}|\mathcal{F}_{k-1}] \\ &= g(1)\lambda X_{k-1}\Delta + g(-1)\mu X_{k-1}\Delta + o(\Delta) + \frac{o(\Delta^2)}{1-\Delta} \\ &= g(1)\lambda X_{k-1}\Delta + g(-1)\mu X_{k-1}\Delta + o(\Delta) \end{aligned} \tag{A10}$$

Then it is clear that

$$v(Z_t^{(n)}; [0, t] \times g) = \sum_{k=1}^{\sigma_t} (g(1)\lambda + g(-1)\mu)X_{k-1}\Delta + O(\Delta)$$

For the bivariate case, the proof is similar, the conditional expectation

$$\begin{aligned} & \mathbb{E}[g(\mathbf{V}_k)1_{\{\mathbf{V}_k \neq (0,0)^T\}}|\mathcal{F}_{k-1}] \\ &= \mathbb{E}[g(1, 0)1_{\{\mathbf{V}_k=(1,0)^T\}}|\mathcal{F}_{k-1}] + \mathbb{E}[g(0, 1)1_{\{\mathbf{V}_k=(0,1)^T\}}|\mathcal{F}_{k-1}] \\ &+ \mathbb{E}[g(-1, 0)1_{\{\mathbf{V}_k=(-1,0)^T\}}|\mathcal{F}_{k-1}] + \mathbb{E}[g(0, -1)1_{\{\mathbf{V}_k=(0,-1)^T\}}|\mathcal{F}_{k-1}] \\ &+ \sum_{|i|+|j|>1} \mathbb{E}[g(i, j)1_{\{\mathbf{V}_k=(i,j)^T\}}|\mathcal{F}_{k-1}] + \mathbb{E}[g(i, -j)1_{\{\mathbf{V}_k=(i,-j)^T\}}|\mathcal{F}_{k-1}] \\ &+ \sum_{|i|+|j|>1} \mathbb{E}[g(-i, j)1_{\{\mathbf{V}_k=(-i,j)^T\}}|\mathcal{F}_{k-1}] + \mathbb{E}[g(-i, -j)1_{\{\mathbf{V}_k=(-i,-j)^T\}}|\mathcal{F}_{k-1}] \\ &= \left( g(1, 0)\tilde{\lambda}_1 + g(-1, 0)\tilde{\mu}_1 + g(0, 1)\tilde{\lambda}_2 + g(0, -1)\tilde{\mu}_2 \right) \mathbf{Y}_{k-1}\Delta + o(\Delta) + \frac{o(\Delta^2)}{(1-\Delta)^2} \\ &= \left( g(1, 0)\tilde{\lambda}_1 + g(-1, 0)\tilde{\mu}_1 + g(0, 1)\tilde{\lambda}_2 + g(0, -1)\tilde{\mu}_2 \right) \mathbf{Y}_{k-1}\Delta + o(\Delta) \end{aligned} \tag{A11}$$

Finally the discrete stochastic integral is given by

$$\sum_{k=1}^{\sigma_t} \left( g(1, 0)\tilde{\lambda}_1 + g(-1, 0)\tilde{\mu}_1 + g(0, 1)\tilde{\lambda}_2 + g(0, -1)\tilde{\mu}_2 \right) \mathbf{Y}_{k-1}\Delta + O(\Delta) \tag{A12}$$

□

## References

- Al-Osh M, Alzaid AA (1987) First-order integer-valued autoregressive (INAR (1)) process. *J Time Ser Anal* 8(3):261–275
- Bailey NT (1991) *The elements of stochastic processes with applications to the natural sciences*, vol 25
- Bu R, McCabe B, Hadri K (2008) Maximum likelihood estimation of higher-order integer-valued autoregressive processes. *J Time Ser Anal* 29(6):973–994
- Chen R, Hyrien O (2011) Quasi-and pseudo-maximum likelihood estimators for discretely observed continuous-time Markov branching processes. *J Stat Plan Inference* 141(7):2209–2227
- Crawford FW, Minin VN, Suchard MA (2014) Estimation for general birth-death processes. *J Am Stat Assoc* 109(506):730–747
- Davis MH (1984) Piecewise-deterministic Markov processes: a general class of non-diffusion stochastic models. *J R Stat Soc Ser B Methodol* 46(3):353–376
- Davis RA, Holan SH, Lund R, Ravishanker N (2016) *Handbook of discrete-valued time series*
- Davison AC, Hautphenne S, Kraus A (2021) Parameter estimation for discretely observed linear birth-and-death processes. *Biometrics* 77(1):186–196
- Feller W (1939) Die grundlagen der vollterraschen theorie des kampfes ums dasein in wahrscheinlichkeitstheoretischer behandlung. *Acta Biotheor* 5:441–470
- Griffiths D (1972) A bivariate birth-death process which approximates to the spread of a disease involving a vector. *J Appl Probab* 9(1):65–75
- Griffiths D (1973) Multivariate birth-and-death processes as approximations to epidemic processes. *J Appl Probab* 10(1):15–26
- Jacod J, Shiryaev A (2013) *Limit theorems for stochastic processes*, vol 288
- Keiding N (1975) Maximum likelihood estimation in the birth-and-death process. *Ann Stat* 3(2):363–372
- Kendall DG (1949) Stochastic processes and population growth. *J R Stat Soc Ser B Methodol* 11(2):230–282
- Kirchner M (2016) Hawkes and INAR( $\infty$ ) processes. *Stoch Process Their Appl* 126(8):2494–2525
- McKenzie E (1985) Some simple models for discrete variate time series. *J Am Water Resour Assoc* 21(4):645–650
- Nelson DB (1990) Arch models as diffusion approximations. *J Econom* 45(1–2):7–38
- Scotto MG, Weiß CH, Gouveia S (2015) Thinning-based models in the analysis of integer-valued time series: a review. *Stat Model* 15(6):590–618
- Tavaré S (2018) The linear birth-death process: an inferential retrospective. *Adv Appl Probab* 50(A):253–269
- Weiß CH (2008) Thinning operations for modeling time series of counts—a survey. *Adv Stat Anal* 92(3):319–341
- Weiß CH (2018) *An introduction to discrete-valued time series*

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.