



Birth/birth-death processes and their computable transition probabilities with biological applications

Lam Si Tung Ho¹ · Jason Xu² · Forrest W. Crawford³ · Vladimir N. Minin⁴ · Marc A. Suchard⁵

Received: 3 August 2016 / Revised: 4 April 2017 / Published online: 24 July 2017
© Springer-Verlag GmbH Germany 2017

Abstract Birth-death processes track the size of a univariate population, but many biological systems involve interaction between populations, necessitating models for two or more populations simultaneously. A lack of efficient methods for evaluating finite-time transition probabilities of bivariate processes, however, has restricted statistical inference in these models. Researchers rely on computationally expensive methods such as matrix exponentiation or Monte Carlo approximation, restricting likelihood-based inference to small systems, or indirect methods such as approximate Bayesian computation. In this paper, we introduce the *birth/birth-death process*, a tractable bivariate extension of the birth-death process, where rates are allowed to be nonlinear. We develop an efficient algorithm to calculate its transition probabilities using a continued fraction representation of their Laplace transforms. Next, we identify several exemplary models arising in molecular epidemiology, macro-parasite evolution, and infectious disease modeling that fall within this class, and demonstrate advantages of our proposed method over existing approaches to inference in these models. Notably, the ubiquitous stochastic susceptible-infectious-removed (SIR) model falls within this

This work was partially supported by the National Institutes of Health (R01 HG006139, R01 AI1107034, and U54 GM111274) and the National Science Foundation (IIS 1251151, DMS 1264153, DMS 1606177). We thank Christopher Drovandi, Edwin Michael, and David Denham for access to the Brugia pahangi count data.

✉ Lam Si Tung Ho
lamho86@gmail.com

¹ Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA, USA

² Department of Biomathematics, University of California, Los Angeles, Los Angeles, CA, USA

³ Department of Biostatistics, Yale University, New Haven, CT, USA

⁴ Departments of Statistics and Biology, University of Washington, Seattle, WA, USA

⁵ Departments of Biomathematics, Biostatistics and Human Genetics, University of California, Los Angeles, Los Angeles, WA, USA

class, and we emphasize that computable transition probabilities newly enable direct inference of parameters in the SIR model. We also propose a very fast method for approximating the transition probabilities under the SIR model via a novel branching process simplification, and compare it to the continued fraction representation method with application to the 17th century plague in Eyam. Although the two methods produce similar maximum a posteriori estimates, the branching process approximation fails to capture the correlation structure in the joint posterior distribution.

Keywords Stochastic models · Birth-death process · Infectious disease · SIR model · Transition probabilities

Mathematics Subject Classification 60J27 · 92D30 · 62F15

1 Introduction

Birth-death processes have been used extensively in many applications including evolutionary biology, ecology, population genetics, epidemiology, and queuing theory (see e.g. [Novozhilov et al. 2006](#); [Crawford and Suchard 2012](#); [Doss et al. 2013](#); [Rabier et al. 2014](#); [Crawford et al. 2015](#)). However, establishing analytic and computationally practical formulae for their transition probabilities is usually difficult ([Novozhilov et al. 2006](#)). The state-of-the-art method for computing the transition probabilities of birth-death processes proposed in [Crawford and Suchard \(2012\)](#) enables statistical estimation for general birth-death processes using likelihood-based inference ([Crawford et al. 2014](#)). Unfortunately, birth-death processes inherently only track one population, and extending this technique beyond the univariate case is nontrivial. Many applied models require the consideration of two or more interacting populations simultaneously to model behavior such as competition, predation, or infection. Examples of such bivariate models include epidemic models ([McKendrick 1926](#); [Kermack and McKendrick 1927](#); [Griffiths 1972](#)), predator-prey models ([Hitchcock 1986](#); [Owen et al. 2015](#)), genetic models ([Rosenberg et al. 2003](#); [Xu et al. 2015](#)), and within-host macro-parasite models ([Drovandi and Pettitt 2011](#)).

The most general extensions of birth-death processes to bivariate processes are competition processes ([Reuter 1961](#)). These processes allow not only “birth” and “death” events in each population, but also “transition” events where an individual moves from one population to the other. Unlike birth-death processes, few attempts have been made to compute the transition probabilities of competition processes or their special cases. Hence, researchers usually rely on classical continuous-time Markov chain methods such as matrix exponentiation and diffusion approximation. Unfortunately, these methods fail to leverage the specific structure of competition processes, and have several intrinsic limitations. Matrix exponentiation methods compute the transition probability matrix $\mathbf{P}(t)$ by solving the matrix form of Kolmogorov’s forward equation $\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{Q}$ with initial condition $\mathbf{P}(0) = \mathbf{I}$, where \mathbf{Q} is the instantaneous rate matrix of the process. While this equation admits a unique solution $\mathbf{P}(t) = \exp(\mathbf{Q}t)$ ([Ephraim and Mark 2012](#)), numerical evaluation of the matrix exponential is often troublesome ([Moler and Loan 2003](#)). Its computational cost via eigenvalue decomposition,

for instance, is cubic in the size of the state-space and thus becomes computationally prohibitive even with moderately sized state-spaces (Drovandi and Pettitt 2011; Crawford and Suchard 2012). For example, Keeling and Ross (2008) demonstrate that computing transition probabilities via matrix exponentiation for the simplest epidemic models is practical only when modeling spread of an infectious disease through a very small population (e.g., 100 people). Moreover, matrix exponentiation can introduce serious rounding errors for certain rate matrices even for biologically reasonable values (Schranz et al. 2008; Crawford and Suchard 2012; Crawford et al. 2014). Diffusion approximations, on the other hand, require the state-space to be large in order to justify approximating a discrete process by a continuous-valued diffusion process (Karev et al. 2005; Golightly and Wilkinson 2005), and can often remain inaccurate for simulation even in settings with large state-spaces (Golightly and Wilkinson 2005). Branching processes form another closely related class of processes, and have been used in a likelihood-based framework to study bivariate populations (Xu et al. 2015). Branching processes are at once more general than competition processes, permitting events that increment populations by more than one, and also more restrictive in that linearity is implied by an assumption that particles act independently. The latter assumption is limiting in epidemiological applications, for instance, which commonly feature non-linear interactions between populations.

The lack of a reliable method for computing transition probabilities in bivariate processes forces researchers to apply alternative likelihood-free approaches such as approximate Bayesian computation (ABC) (Blum and Tran 2010; Drovandi and Pettitt 2011; Owen et al. 2015). The ABC approach uses simulated and observed summary statistics to bypass likelihood evaluation. Nonetheless, this is not a panacea approach that can completely replace traditional likelihood-based methods. The ABC method itself has several sources for loss of information such as non-zero tolerance, and non-sufficient summary statistics (Sunnåker et al. 2013). The tolerance is an ad hoc threshold to decide whether ABC accepts a new proposal. If the tolerance is zero and the summary statistics are sufficient, ABC is guaranteed to return the correct posterior distribution. In practice, however, tolerance is always positive which often leads to bias. In the context of counting processes, sufficient summary statistics usually do not exist because the data are observed partially. Thus, credible interval estimates under ABC are potentially inflated due to the loss of information (Csilléry et al. 2010). Also, when sufficient summary statistics are not available, the ABC method can not be trusted in selecting between models (Robert et al. 2011). Because of all these limitations, direct likelihood-based methods are often more favorable.

In this paper, we develop an efficient method to compute the transition probabilities of a subclass of competition processes with two interacting populations of particles, enabling likelihood-based inference. We call this subclass birth(death)/birth-death processes, whose first population is increasing (decreasing). It is worth mentioning that we do not impose linearity condition for the rates of these processes. A rigorous characterization of this class of processes and derivation of recursive formulae to compute their transition probabilities are provided in Sect. 2. Our main tools are the Laplace transform and continued fractions that have been successfully applied for univariate birth-death processes in Crawford and Suchard (2012). These formulae enable accurate and computationally efficient numerical computation of transition probabilities. We implement

this method in the new R package `MultiBD` <https://github.com/msuchard/MultiBD>. In Sect. 3, we discuss multiple scientifically relevant applications of birth(death)/birth-death processes including stochastic susceptible-infectious-removed (SIR) models in epidemiology (McKendrick 1926; Kermack and McKendrick 1927; Raggett 1982), monomolecular reaction systems (Jahnke and Huisinga 2007), a birth-death-shift model for transposable elements (Rosenberg et al. 2003; Xu et al. 2015), and a within-host macro-parasite model (Riley et al. 2003; Drovandi and Pettitt 2011). We examine the accuracy of our method in simulation studies, including comparisons to branching process, matrix exponentiation method, and Monte Carlo approximations. Finally, we apply our method to estimate infection rates and death rates during the plague of Eyam in 1666 within a likelihood-based Bayesian framework in Sect. 4.

Previous work on computing the transition probabilities: Analytic expressions of the transition probabilities have only been found for some special cases such as linear birth-death processes (see e.g. Novozhilov et al. 2006) and monomolecular reaction systems (Jahnke and Huisinga 2007). Therefore, matrix exponentiation is still the most common method for computing the transition probabilities of general Markov processes. The state-of-the-art software package for exponentiating sparse matrices is `Expokit` (Sidje 1998; Moler and Loan 2003), which uses Krylov subspace projection method. Eshof and Hochbruck (2006) propose a modified version using a simple preconditioned transformation to improve the convergence behavior of this method. Although matrix exponentiation has the advantage of generality in that it can be applied to any Markov process, it is not the most efficient method in many scenarios. Recently, Crawford and Suchard (2012) propose an efficient method for evaluating the transition probabilities of general birth-death processes using Laplace transform and continued fraction. However, efficient methods that extend this result to general bivariate birth-death processes have yet to be found.

2 Birth(death)/birth-death processes

2.1 Birth/birth-death processes

A birth/birth-death process is a bivariate continuous-time Markov process $\mathbf{X}(t) = (X_1(t), X_2(t))$, $t \geq 0$, whose state-space is in $\mathbb{N} \times \mathbb{N}$, the Cartesian product of the non-negative integers. We can describe a birth/birth-death process as governing dynamics of a system consisting two types of particles, where one out of four possible events can happen in infinitesimal time: (1) a new **type 1** particle enters the system; (2) a new **type 2** particle enters the system; (3) a **type 2** particle leaves the system; or (4) a **type 2** particle becomes a **type 1** particle. In this system, $X_1(t)$ and $X_2(t)$ track the number of **type 1** and **type 2** particles at time t respectively. Mathematically, there are five possibilities for $\mathbf{X}(t)$ during a small time interval $(t, t + dt)$:

$$\Pr \left\{ \begin{array}{l} X_1(t + dt) = a + 1 \\ X_2(t + dt) = b \end{array} \middle| \begin{array}{l} X_1(t) = a \\ X_2(t) = b \end{array} \right\} = \lambda_{ab}^{(1)} dt + o(dt)$$

$$\Pr \left\{ \begin{array}{l} X_1(t + dt) = a \\ X_2(t + dt) = b + 1 \end{array} \middle| \begin{array}{l} X_1(t) = a \\ X_2(t) = b \end{array} \right\} = \lambda_{ab}^{(2)} dt + o(dt)$$

$$\begin{aligned}
 \Pr \left\{ \begin{array}{l} X_1(t + dt) = a \\ X_2(t + dt) = b - 1 \end{array} \middle| \begin{array}{l} X_1(t) = a \\ X_2(t) = b \end{array} \right\} &= \mu_{ab}^{(2)} dt + o(dt) \\
 \Pr \left\{ \begin{array}{l} X_1(t + dt) = a + 1 \\ X_2(t + dt) = b - 1 \end{array} \middle| \begin{array}{l} X_1(t) = a \\ X_2(t) = b \end{array} \right\} &= \gamma_{ab} dt + o(dt) \\
 \Pr \left\{ \begin{array}{l} X_1(t + dt) = a \\ X_2(t + dt) = b \end{array} \middle| \begin{array}{l} X_1(t) = a \\ X_2(t) = b \end{array} \right\} &= 1 - (\lambda_{ab}^{(1)} + \lambda_{ab}^{(2)} + \mu_{ab}^{(2)} + \gamma_{ab}) dt + o(dt),
 \end{aligned}
 \tag{1}$$

where $a, b \in \mathbb{N}$, $\lambda_{ab}^{(1)} \geq 0$ is the birth rate of **type 1** particles given a **type 1** particles and b **type 2** particles, $\lambda_{ab}^{(2)} \geq 0$ is the equivalent birth rate of **type 2** particles, $\mu_{ab}^{(2)} \geq 0$ is the death rate of **type 2** particles, and γ_{ab} is the transition rate from **type 2** particles to **type 1** particles. We fix $\lambda_{-1,b}^{(1)} = \lambda_{a,-1}^{(2)} = \mu_{a0}^{(2)} = \gamma_{-1,b} = \gamma_{a0} = 0$.

Letting $P_{ab}^{a_0 b_0}(t) = \Pr\{\mathbf{X}(t) = (a, b) \mid \mathbf{X}(0) = (a_0, b_0)\}$, the forward Kolmogorov’s equations for the birth/birth-death process are

$$\begin{aligned}
 \frac{dP_{ab}^{a_0 b_0}(t)}{dt} &= \lambda_{a-1,b}^{(1)} P_{a-1,b}^{a_0 b_0}(t) + \lambda_{a,b-1}^{(2)} P_{a,b-1}^{a_0 b_0}(t) + \mu_{a,b+1}^{(2)} P_{a,b+1}^{a_0 b_0}(t) \\
 &\quad + \gamma_{a-1,b+1} P_{a-1,b+1}^{a_0 b_0}(t) - (\lambda_{ab}^{(1)} + \lambda_{ab}^{(2)} + \mu_{ab}^{(2)} + \gamma_{ab}) P_{ab}^{a_0 b_0}(t), \tag{2}
 \end{aligned}$$

for all (a, b) . In practice, we can usually only observe the process discretely. In this scenario, the likelihood function is the product of transition probabilities between consecutive observations. Therefore, computing $P_{ab}^{a_0 b_0}(t)$ is an important step for any direct likelihood-based analysis.

In general, a birth/birth-death process is a special case of a competition process (Reuter 1961) with rate matrix $\mathbf{Q} = \{q_{ij}\}$ where $i, j \in \mathbb{N} \times \mathbb{N}$ and

j	Competition process	Birth/birth-death
$(a + 1, b)$	$q_{(a,b)(a+1,b)}$	$\lambda_{ab}^{(1)}$
$(a - 1, b)$	$q_{(a,b)(a-1,b)}$	0
$(a, b + 1)$	$q_{(a,b)(a,b+1)}$	$\lambda_{ab}^{(2)}$
$(a, b - 1)$	$q_{(a,b)(a,b-1)}$	$\mu_{ab}^{(2)}$
$(a + 1, b - 1)$	$q_{(a,b)(a+1,b-1)}$	γ_{ab}
$(a - 1, b + 1)$	$q_{(a,b)(a-1,b+1)}$	0
(a, b)	$-\sum_{\substack{k \neq l \\ k, l \in \{-1, 0, 1\}}} q_{(a,b)(a+k,b+l)}$	$-(\lambda_{ab}^{(1)} + \lambda_{ab}^{(2)} + \mu_{ab}^{(2)} + \gamma_{ab})$
other	0	0

for $i = (a, b)$. Competition processes are the most general bivariate Markov processes that only allow transitions between neighboring states. Many practical models in biology are special cases of these processes such as epidemic models (McKendrick 1926; Kermack and McKendrick 1927; Griffiths 1972) and predator-prey models (Hitchcock 1986; Owen et al. 2015).

2.1.1 Sufficient condition for regularity

Definition 1 A birth/birth-death process is regular if there is a unique set of transition probabilities $P_{ab}^{a_0b_0}(t)$ satisfying the system of Eq. (2).

Here, we establish the sufficient condition for regularity of a birth/birth-death process. For $k \in \mathbb{N}$, we denote:

$$D_k = \{(a, b) : a + b = k\} \in \mathbb{N} \times \mathbb{N}, \text{ and}$$

$$\lambda_k = \max_{(a,b) \in D_k} \{\lambda_{ab}^{(1)} + \lambda_{ab}^{(2)}\}. \tag{3}$$

Theorem 1 The sufficient condition for regularity of a general birth/birth-death process is $\sum_{k=1}^{\infty} 1/\lambda_k = \infty$.

Proof We will apply the following Reuter’s condition (Reuter 1957):

Lemma 1 Let $\mathbf{Q} = \{q_{ij}\}$ be a conservative matrix, such that $-q_{ii} = \sum_{j \neq i} q_{ij} < \infty$. A continuous-time Markov chain associated with \mathbf{Q} is regular if and only if for some $\zeta > 0$, the equation $\mathbf{Q}\mathbf{y} = \zeta\mathbf{y}$ subject to $0 \leq y_i \leq 1$ has only trivial solution $\mathbf{y} = \mathbf{0}$.

For a general birth/birth-death process, states i and j are in $\mathbb{N} \times \mathbb{N}$. Let $\{y_{ab}\}_{a,b \in \mathbb{N}}$ be a solution of $\mathbf{Q}\mathbf{y} = \zeta\mathbf{y}$ such that $y_{ab} \in [0, 1]$ for any a and b . Then, we have

$$(\zeta + \lambda_{ab}^{(1)} + \lambda_{ab}^{(2)} + \mu_{ab}^{(2)} + \gamma_{ab})y_{ab} = \lambda_{ab}^{(1)}y_{a+1,b} + \lambda_{ab}^{(2)}y_{a,b+1} + \mu_{ab}^{(2)}y_{a,b-1} + \gamma_{ab}y_{a+1,b-1}. \tag{4}$$

Defining $y_k = \max_{(a,b) \in D_k} \{y_{ab}\}$ and $(a_k, b_k) = \operatorname{argmax}_{(a,b) \in D_k} \{y_{ab}\}$, we deduce that

$$(\zeta + \lambda_{a_k b_k}^{(1)} + \lambda_{a_k b_k}^{(2)} + \mu_{a_k b_k}^{(2)})y_k \leq (\lambda_{a_k b_k}^{(1)} + \lambda_{a_k b_k}^{(2)})y_{k+1} + \mu_{a_k b_k}^{(2)}y_{k-1}, \text{ and}$$

$$\zeta y_k + \mu_{a_k b_k}^{(2)}(y_k - y_{k-1}) \leq (\lambda_{a_k b_k}^{(1)} + \lambda_{a_k b_k}^{(2)})(y_{k+1} - y_k). \tag{5}$$

Since $\mu_{a_{-1}b_{-1}}^{(2)} = 0$, y_k is an increasing sequence. Thus,

$$\frac{\zeta}{\lambda_k} y_k \leq y_{k+1} - y_k. \tag{6}$$

Assuming that there exists k_0 such that $y_{k_0} > 0$, we obtain

$$y_k \geq y_{k_0} + \zeta \sum_{i=k_0}^{k-1} \frac{y_i}{\lambda_i} \geq y_{k_0} \left(1 + \zeta \sum_{i=k_0}^{k-1} \frac{1}{\lambda_i} \right), \tag{7}$$

that is larger than 1 if k is big enough. Hence $y_k = 0$ for every k . Then, the theorem is proved by applying Lemma 1.

Note that the condition in Theorem 1 generalizes the classical regularity condition of a pure birth process (Feller 1968). From now on, we assume that our birth/birth-death processes are regular.

2.1.2 Recursive formula for transition probabilities

In this section, we establish a recursion to calculate the transition probabilities $P_{ab}^{a_0b_0}(t)$ of a birth/birth-death process. Since we assume that our birth/birth-death process is regular, these transition probabilities are unique.

We first note that $P_{ab}^{a_0b_0}(t) = 0$ for all $a < a_0$. Let $f_{ab}(s)$, $s \in \mathbb{C}$, be the Laplace transform of $P_{ab}^{a_0b_0}(t)$, that is

$$f_{ab}(s) = \mathcal{L}[P_{ab}^{a_0b_0}(t)](s) = \int_0^\infty e^{-st} P_{ab}^{a_0b_0}(t) dt. \tag{8}$$

From (2), we have

$$\begin{aligned} sf_{ab}(s) - P_{ab}^{a_0b_0}(0) &= \lambda_{a-1,b}^{(1)} f_{a-1,b}(s) + \lambda_{a,b-1}^{(2)} f_{a,b-1}(s) + \mu_{a,b+1}^{(2)} f_{a,b+1}(s) \\ &\quad + \gamma_{a-1,b+1} f_{a-1,b+1}(s) - (\lambda_{ab}^{(1)} + \lambda_{ab}^{(2)} + \mu_{ab}^{(2)} + \gamma_{ab}) \\ &\quad f_{ab}(s), \quad (a, b) \in \mathbb{N}^2. \end{aligned} \tag{9}$$

Note that $f_{ab}(s)$ is the unique solution of (9) by the uniqueness of $P_{ab}^{a_0b_0}(t)$. We construct the recursive approximation formulae for $f_{ab}(s)$ using continued fractions. ‘‘Appendix A’’ provides necessary background on continued fractions and their convergents. Denote

$$\begin{aligned} x_{a1} &= -\frac{1}{\mu_{a1}^{(2)}}; \quad x_{ab} = -\frac{\lambda_{a,b-2}^{(2)}}{\mu_{ab}^{(2)}}, \quad b \geq 2 \\ y_{ab} &= -\frac{s + \lambda_{a,b-1}^{(1)} + \lambda_{a,b-1}^{(2)} + \mu_{a,b-1}^{(2)} + \gamma_{a,b-1}}{\mu_{ab}^{(2)}}, \quad b \geq 1, \end{aligned} \tag{10}$$

and consider the following continued fraction

$$\phi_{a0}^{(0)}(s) = \frac{x_{a1}}{y_{a1} + \frac{x_{a2}}{y_{a2} + \frac{x_{a3}}{y_{a3} + \dots}}}. \tag{11}$$

We can construct the sequence $\{\phi_{ab}^{(0)}(s)\}_{b=0}^\infty$ (Definition A.3, ‘‘Appendix A’’) as follows:

$$\begin{aligned} (s + \lambda_{a0}^{(1)} + \lambda_{a0}^{(2)})\phi_{a0}^{(0)}(s) - \mu_{a1}^{(2)}\phi_{a1}^{(0)}(s) &= 1, \text{ and} \\ (s + \lambda_{a,b-1}^{(1)} + \lambda_{a,b-1}^{(2)} + \mu_{a,b-1}^{(2)} + \gamma_{a,b-1})\phi_{a,b-1}^{(0)}(s) \\ - \lambda_{a,b-2}^{(2)}\phi_{a,b-2}^{(0)}(s) - \mu_{ab}^{(2)}\phi_{ab}^{(0)}(s) &= 0, \quad b \geq 2. \end{aligned} \tag{12}$$

Comparing the sequences in (12) with (9), we deduce that $\mathcal{L}^{-1}[\phi_{ab}^{(0)}(s)] = P_{ab}^{a_0b_0}(t)$. Since $P_{ab}^{a_0b_0}(t)$ is a probability distribution, we have $\sum_{(a,b) \in \mathbb{N} \times \mathbb{N}} P_{ab}^{a_0b_0}(t) = 1$. Taking

the Laplace transform of the previous equation, we get $\sum_{(a,b) \in \mathbb{N} \times \mathbb{N}} \phi_{ab}^{(0)}(s) = 1/s$. Hence, $\lim_{b \rightarrow \infty} \phi_{a_0b}^{(0)}(s) = 0$ for every $s > 0$. By Lemma A.4 (“Appendix A”), $\phi_{a_0}^{(0)}(s)$ converges for every $s > 0$, and

$$\phi_{ab}^{(0)}(s) = \prod_{i=1}^b x_{ai} \frac{x_{a,b+1}}{Y_{a,b+1} + \frac{x_{a,b+2}Y_{ab}}{y_{a,b+2} + \frac{x_{a,b+3}}{y_{a,b+3} + \frac{x_{a,b+4}}{y_{a,b+4} + \dots}}}} \tag{13}$$

where Y_{ab} is the denominator of the b th convergent of $\phi_{a_0}^{(0)}(s)$.

From (9), we note that

$$(s + \lambda_{a_0b}^{(1)} + \lambda_{a_0b}^{(2)} + \mu_{a_0b}^{(2)} + \gamma_{a_0b})f_{a_0b} - \lambda_{a_0,b-1}^{(2)}f_{a_0,b-1}(s) - \mu_{a_0,b+1}^{(2)}f_{a_0,b+1}(s) = 1_{\{b=b_0\}}, \quad b \in \mathbb{N}. \tag{14}$$

By Lemma A2 (“Appendix A”), $f_{a_0b}(s) = \phi_{a_0b}^{(b_0)}(s)$ where

$$\phi_{ab}^{(m)}(s) = \begin{cases} \frac{(-1)^{m-b+1}Y_{ab}}{\mu_{a,m+1}^{(2)} \prod_{i=1}^{m+1} x_{ai}} \phi_{am}^{(0)}(s), & \text{if } b \leq m \\ \frac{-Y_{am}}{\mu_{a,m+1}^{(2)} \prod_{i=1}^{m+1} x_{ai}} \phi_{ab}^{(0)}(s), & \text{if } b \geq m. \end{cases} \tag{15}$$

Next, we obtain formulae for approximating $f_{ab}(s)$ recursively assuming that we already have evaluated $f_{a-1,b}(s)$. Again, from (2), we have

$$(s + \lambda_{ab}^{(1)} + \lambda_{ab}^{(2)} + \mu_{ab}^{(2)} + \gamma_{ab})f_{ab}(s) - \lambda_{a,b-1}^{(2)}f_{a,b-1}(s) - \mu_{a,b+1}^{(2)}f_{a,b+1}(s) = \lambda_{a-1,b}^{(1)}f_{a-1,b}(s) + \gamma_{a-1,b+1}f_{a-1,b+1}(s), \tag{16}$$

for $b \in \mathbb{N}$. We approximate $f_{ab}(s)$ by solving a truncated version of (16) for $0 \leq b \leq B$, where B is sufficiently large. The intuition of how to choose B follows from the observation that we want $\sum_{a=0}^{\infty} \sum_{b=B+1}^{\infty} P_{ab}^{a_0b_0}(t)$ to be small. By Lemma A2 (“Appendix A”), we have the following approximation:

$$f_{ab}(s) \approx \sum_{m=0}^B \left[\lambda_{a-1,m}^{(1)}f_{a-1,m}(s) + \gamma_{a-1,m+1}f_{a-1,m+1}(s) \right] \phi_{ab}^{(m)}(s). \tag{17}$$

Therefore, the transition probabilities of a birth/birth-death process can be computed recursively using the following Theorem:

Theorem 2 Let $\phi_{ab}^{(m)}(s)$ be defined as in (11), (13), and (15). We have

$$P_{ab}^{a_0b_0}(t) = \begin{cases} 0, & \text{if } a < a_0 \\ \mathcal{L}^{-1}[f_{ab}(s)](t), & \text{if } a \geq a_0, \end{cases} \tag{18}$$

where $f_{a_0b}(s) = \phi_{a_0b}^{(b_0)}(s)$ and

$$f_{ab}(s) \approx \sum_{m=0}^B \left[\lambda_{a-1,m}^{(1)} f_{a-1,m}(s) + \gamma_{a-1,m+1} f_{a-1,m+1}(s) \right] \phi_{ab}^{(m)}(s), \quad a > a_0. \tag{19}$$

Here, $\mathcal{L}^{-1}(\cdot)$ denotes the inverse Laplace transform and B is the truncation level.

If the number of **type 2** particles is bounded by B^* , we choose $B = B^*$. In this case, the approximation in Theorem 2 is exact. We prove that the output of our approximation scheme (19) converges to $f_{ab}(s)$ as B goes to infinity in ‘‘Appendix C’’. Further, the transition probability returned by Theorem 2 converges to the true transition probability. This truncation error can be bounded explicitly by extending the coupling argument in Crawford et al. (2016) to multivariate processes. However, we leave it as a subject of future work because a complete treatment is beyond the scope of this paper.

2.1.3 Numerical approximation of the transitions probabilities

To approximate $P_{ab}^{a_0b_0}(t)$ using Theorem 2, we need to compute two quantities: the continued fractions $\phi_{ab}^{(m)}(s)$, and the inverse Laplace transform $\mathcal{L}^{-1}[f_{ab}(s)](t)$. We efficiently evaluate the continued fractions $\phi_{ab}^{(m)}(s)$ through the modified Lentz method (Lentz 1976; Thompson and Barnett 1986); see ‘‘Appendix B’’ for more details. This algorithm enables us to control for and limit truncation error. To approximate the inverse Laplace transform $\mathcal{L}^{-1}[f_{ab}(s)](t)$, we apply the method proposed in Abate and Whitt (1992) using a Riemann sum:

$$\mathcal{L}^{-1}[f_{ab}(s)](t) \approx \frac{e^{H/2}}{2t} \mathcal{R} \left[f_{ab} \left(\frac{H}{2t} \right) \right] + \frac{e^{H/2}}{t} \sum_{k=1}^{\infty} (-1)^k \mathcal{R} \left[f_{ab} \left(\frac{H + 2k\pi i}{2t} \right) \right]. \tag{20}$$

Here $\mathcal{R}[z]$ is the real part of z and H is a positive real number. Abate and Whitt (1992) show that the error that arises in (20) is bounded by $1/(e^H - 1)$. Moreover, we can use the Levin transform (Levin 1973) to improve the rate of convergence because the series in (20) is an alternating series when $\mathcal{R}\{f_{ab}[(H + 2k\pi i)/(2t)]\}$ have the same sign. These numerical methods have been successfully applied by Crawford and Suchard (2012) to compute the transition probabilities of birth-death processes.

In practice, to handle situations where $\mu_{ab}^{(2)}$ can possibly equal to 0 for some (a, b) , we re-parametrize x_{ab} and y_{ab} as follows:

$$x_{a1} = 1; \quad x_{ab} = -\lambda_{a,b-2}^{(2)} \mu_{a,b-1}^{(2)}, \quad b \geq 2, \quad \text{and}$$

$$y_{ab} = s + \lambda_{a,b-1}^{(1)} + \lambda_{a,b-1}^{(2)} + \mu_{a,b-1}^{(2)} + \gamma_{a,b-1}, \quad b \geq 1. \tag{21}$$

With this new parametrization, we obtain

$$\phi_{ab}^{(m)}(s) = \begin{cases} \frac{(\prod_{i=b+1}^m \mu_{ai}^{(2)})Y_{ab}}{Y_{a,m+1} + \frac{x_{a,m+2}Y_{am}}{y_{a,m+2} + \frac{x_{a,m+3}}{y_{a,m+3} + \frac{x_{a,m+4}}{y_{a,m+4} + \dots}}}, & \text{if } b \leq m \\ \frac{(\prod_{i=m+1}^b \lambda_{ai}^{(2)})Y_{am}}{Y_{a,b+1} + \frac{x_{a,b+2}Y_{ab}}{y_{a,b+2} + \frac{x_{a,b+3}}{y_{a,b+3} + \frac{x_{a,b+4}}{y_{a,b+4} + \dots}}}, & \text{if } b \geq m. \end{cases} \tag{22}$$

Our complete algorithm to compute the transition probabilities of birth/birth-death processes is implemented in the function `bbd_prob` in a new R package called `MuLiBD`. The function takes $t, a_0, b_0, \lambda_{ab}^{(1)}, \lambda_{ab}^{(2)}, \mu_{ab}^{(2)}, \gamma_{ab}, A, B$ as inputs and returns the transition probability matrix $\{P_{ab}^{a_0b_0}(t)\}_{a_0 \leq a \leq A, 0 \leq b \leq B}$. Here, there is no requirement for A while B needs to be large enough such that $\sum_{a=a_0}^A \sum_{b=B+1}^\infty P_{ab}^{a_0b_0}(t)$ is small. We can check to see if B is large enough by checking if $\sum_{a=a_0}^A P_{aB}^{a_0b_0}(t)$ is sufficiently small.

In practice, the computational complexity of evaluating each term $(f_{ab}(s))_{a_0 \leq a \leq A, 0 \leq b \leq B}$ is $\mathcal{O}((A - a_0)B^2)$ because the Lentz algorithm terminates quickly. Let K be the number of iterations required by the Levin acceleration method (Levin 1973) to achieve a certain error bound for the Riemann sum in (20). Then, the total complexity of our algorithm is $\mathcal{O}((A - a_0)B^2K)$. However, evaluation of $\{f_{ab}[(H + 2k\pi i)/(2t)]\}_{k=1}^K$ can be efficiently parallelized across different values of k , and we exploit this parallelism via multicore processing, delegating most of the computational work to compiled C++ code.

2.2 Death/birth-death processes

Similar to the birth/birth-death process, a death/birth-death process is also a special case of competition processes. The only difference is that the number of **type 1** particles is decreasing instead of increasing. Mathematically, possible transitions of a death/birth-death process $\mathbf{X}(t) = (X_1(t), X_2(t))$ during $(t, t + dt)$ are:

$$\Pr \left\{ \begin{array}{l} X_1(t + dt) = a - 1 \\ X_2(t + dt) = b \end{array} \middle| \begin{array}{l} X_1(t) = a \\ X_2(t) = b \end{array} \right\} = \mu_{ab}^{(1)} dt + o(dt)$$

$$\Pr \left\{ \begin{array}{l} X_1(t + dt) = a \\ X_2(t + dt) = b + 1 \end{array} \middle| \begin{array}{l} X_1(t) = a \\ X_2(t) = b \end{array} \right\} = \lambda_{ab}^{(2)} dt + o(dt)$$

$$\begin{aligned}
 \Pr \left\{ \begin{array}{l} X_1(t + dt) = a \\ X_2(t + dt) = b - 1 \end{array} \middle| \begin{array}{l} X_1(t) = a \\ X_2(t) = b \end{array} \right\} &= \mu_{ab}^{(2)} dt + o(dt) \\
 \Pr \left\{ \begin{array}{l} X_1(t + dt) = a - 1 \\ X_2(t + dt) = b + 1 \end{array} \middle| \begin{array}{l} X_1(t) = a \\ X_2(t) = b \end{array} \right\} &= \gamma_{ab} dt + o(dt) \\
 \Pr \left\{ \begin{array}{l} X_1(t + dt) = a \\ X_2(t + dt) = b \end{array} \middle| \begin{array}{l} X_1(t) = a \\ X_2(t) = b \end{array} \right\} &= 1 - (\mu_{ab}^{(1)} + \lambda_{ab}^{(2)} + \mu_{ab}^{(2)} + \gamma_{ab}) dt + o(dt),
 \end{aligned}
 \tag{23}$$

where $\mu_{ab}^{(1)} \geq 0$ is the death rate of **type 1** particles given a **type 1** particles and b **type 2** particles, $\lambda_{ab}^{(2)} \geq 0$ is the birth rate of **type 2** particles, $\mu_{ab}^{(2)} \geq 0$ is the death rate of **type 2** particles, and γ_{ab} is the transition rate from **type 1** particles to **type 2** particles. Again, we fix $\mu_{0,b}^{(1)} = \lambda_{a,-1}^{(2)} = \mu_{0,b}^{(2)} = \gamma_{0,b} = \gamma_{a,-1} = 0$.

Following a similar argument as in Sect. 2.1.1, we obtain a sufficient condition for regularity of a death/birth-death process. Denote

$$\begin{aligned}
 D_k &= \{(a, b) : a + b = k, a \leq a_0\} \in \mathbb{N} \times \mathbb{N} \\
 \lambda_k &= \max_{(a,b) \in D_k} \{\lambda_{ab}^{(2)}\} \\
 \mu_k &= \min_{(a,b) \in D_k} \{\mu_{ab}^{(1)} + \mu_{ab}^{(2)}\} \\
 \sigma_0 &= 1, \quad \sigma_k = \frac{\lambda_0 \dots \lambda_{k-1}}{\mu_1 \dots \mu_k},
 \end{aligned}
 \tag{24}$$

where a_0 is the number of **type 1** particles at time $t = 0$. The following Theorem is a direct application of Theorem 1 in Iglehart (1964)

Theorem 3 *A sufficient condition for regularity of a death/birth-death process is*

$$\sum_{k=0}^{\infty} \left(\frac{1}{\lambda_k \sigma_k} \sum_{i=0}^k \sigma_i \right) = \infty.
 \tag{25}$$

We note that if we do a transformation for a death/birth-death process $\mathbf{X}(t) = (X_1(t), X_2(t))$ as follows:

$$\begin{aligned}
 Y_1(t) &= a_0 - X_1(t) \\
 Y_2(t) &= B - X_2(t).
 \end{aligned}
 \tag{26}$$

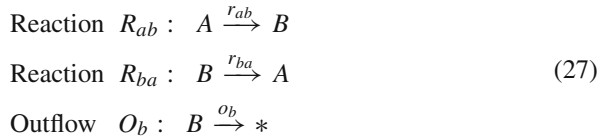
Then, $\mathbf{Y}(t) = (Y_1(t), Y_2(t))$ can be considered as a birth/birth-death process. Therefore, the transition probabilities of a death/birth-death process can also be computed using the R function `bbd_prob` and the transformation (26). Again, we want to choose B such that $\sum_{a=0}^{a_0} \sum_{b=B+1}^{\infty} P_{ab}^{a_0 b_0}(t)$ is small. We implement this procedure in the function `dbd_prob` in our R package `MultiBD`. The function takes $t, a_0, b_0, \mu_{ab}^{(1)}, \lambda_{ab}^{(2)}, \mu_{ab}^{(2)}, \gamma_{ab}, A, B$ as inputs and returns the transition probability matrix $\{P_{ab}^{a_0 b_0}(t)\}_{A \leq a \leq a_0, 0 \leq b \leq B}$. As for birth/birth-death processes, there is no requirement for A .

3 Applications

Birth(death)/birth-death processes are appropriate for modeling two-type populations where the size of the first population is monotonically increasing (decreasing). Here we examine our methods in four applications: a within-host macro-parasite model, a birth-death-shift model for transposable elements, monomolecular reaction systems, and the stochastic SIR epidemiological model. We demonstrate that a birth (death)/birth-death process well captures the dynamics of these common biological problems, and inference using its transition probabilities often outperforms existing approximations. In particular, we emphasize that the birth (death)/birth-death process approach allows us to compute finite-time transition probabilities in the stochastic SIR model that were previously considered unknown or intractable without model simplification (Cauchemez and Ferguson 2008).

3.1 Monomolecular reaction systems

We illustrate the performance of our computational method by considering the following monomolecular reactions:



where r_{ab}, r_{ba} is the reaction rates, and o_b is the outflow rate. Denote

$$Q = \begin{pmatrix} -r_{ab} & r_{ba} \\ r_{ab} & -r_{ba} - o_b \end{pmatrix}, \quad p^{(a)} = e^{Qt} \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad p^{(b)} = e^{Qt} \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

By Theorem 1 in Jahnke and Huisinga (2007), the transition probabilities of the reaction system (27) at time $t > 0$ is

$$P_{ab}^{a_0 b_0}(t) = \mathcal{M}(\cdot, a_0, p^{(a)}) \star \mathcal{M}(\cdot, b_0, p^{(b)})
 \tag{28}$$

where $\mathcal{M}(x, N, p)$ is the multinomial distribution and \star denotes the convolution operator. As analytic expressions for transition probabilities exist for this class of reactions, this example serves as a baseline for comparison to assess the accuracy of our method.

To study these processes in our framework, let $A(t)$ denote the total number of particle A at time t and $L(t)$ be the total number of particle B leaving the system up to t . Then, $\{L(t), A(t)\}$ is a birth/birth-death process with the following possible transitions during $(t, t + dt)$:

$$\Pr \left\{ \begin{array}{l} L(t + dt) = i + 1 \\ A(t + dt) = j \end{array} \middle| \begin{array}{l} L(t) = i \\ A(t) = j \end{array} \right\} = o_b(a_0 + b_0 - i - j)^+ dt + o(dt),$$

$$\begin{aligned}
 & \Pr \left\{ \begin{array}{l} L(t+dt) = i \\ A(t+dt) = j+1 \end{array} \middle| \begin{array}{l} L(t) = i \\ A(t) = j \end{array} \right\} = r_{ba}(a_0 + b_0 - i - j)^+ dt + o(dt), \\
 & \Pr \left\{ \begin{array}{l} L(t+dt) = i \\ A(t+dt) = j-1 \end{array} \middle| \begin{array}{l} L(t) = i \\ A(t) = j \end{array} \right\} = r_{ab}j dt + o(dt), \text{ and} \\
 & \Pr \left\{ \begin{array}{l} L(t+dt) = i \\ A(t+dt) = j \end{array} \middle| \begin{array}{l} L(t) = i \\ A(t) = j \end{array} \right\} \\
 & \quad = 1 - [r_{ab}j + (o_b + r_{ba})(a_0 + b_0 - i - j)^+] dt + o(dt).
 \end{aligned}$$

Here $x^+ = \max(0, x)$. Therefore, $P_{ab}^{a_0 b_0}(t)$ can be computed using our method implemented in the R function `bbd_prob`.

We use `bbd_prob` to calculate $\{P_{ab}^{20,0}(1)\}_{0 \leq a \leq 20, 0 \leq b \leq 20}$ of the reaction system (27) with $r_{ab} = 2, r_{ba} = 0.5$ and $o_b = 1$. The L_1 distance between our result and the analytic result (28) is less than 4.7×10^{-9} , thus confirming the accuracy of our method compared to explicit analytic solutions.

3.2 Birth-death-shift model for transposable elements

Transposable elements or transposons are genomic sequences that can either duplicate, with a new copy moving to a new genomic location, move to a different genomic location, or be deleted from the genome. [Rosenberg et al. \(2003\)](#) model the number of copies of a particular transposon using a linear birth-death-shift process; a birth is a duplication event, a death is a deletion event, and shift is a switching position event. [Xu et al. \(2015\)](#) propose representing this birth-death-shift process by a linear multi-type branching process $\mathbf{X}(t) = (X_{\text{old}}(t), X_{\text{new}}(t))$ tracking the number of occupied sites where $X_{\text{old}}(t)$ is the number of initially occupied sites and $X_{\text{new}}(t)$ is the number of newly occupied sites. Let $\lambda, \mu,$ and ν be the birth, death, and shift rates respectively. The transitions of $\mathbf{X}(t)$ during a small time interval occur with probabilities

$$\begin{aligned}
 & \Pr \left\{ \begin{array}{l} X_{\text{old}}(t+dt) = x_{\text{old}} - 1 \\ X_{\text{new}}(t+dt) = x_{\text{new}} \end{array} \middle| \begin{array}{l} X_{\text{old}}(t) = x_{\text{old}} \\ X_{\text{new}}(t) = x_{\text{new}} \end{array} \right\} = (\mu x_{\text{old}}) dt + o(dt), \\
 & \Pr \left\{ \begin{array}{l} X_{\text{old}}(t+dt) = x_{\text{old}} \\ X_{\text{new}}(t+dt) = x_{\text{new}} - 1 \end{array} \middle| \begin{array}{l} X_{\text{old}}(t) = x_{\text{old}} \\ X_{\text{new}}(t) = x_{\text{new}} \end{array} \right\} = (\mu x_{\text{new}}) dt + o(dt), \\
 & \Pr \left\{ \begin{array}{l} X_{\text{old}}(t+dt) = x_{\text{old}} \\ X_{\text{new}}(t+dt) = x_{\text{new}} + 1 \end{array} \middle| \begin{array}{l} X_{\text{old}}(t) = x_{\text{old}} \\ X_{\text{new}}(t) = x_{\text{new}} \end{array} \right\} = \lambda(x_{\text{old}} + x_{\text{new}}) dt + o(dt), \\
 & \Pr \left\{ \begin{array}{l} X_{\text{old}}(t+dt) = x_{\text{old}} - 1 \\ X_{\text{new}}(t+dt) = x_{\text{new}} + 1 \end{array} \middle| \begin{array}{l} X_{\text{old}}(t) = x_{\text{old}} \\ X_{\text{new}}(t) = x_{\text{new}} \end{array} \right\} = (\nu x_{\text{old}}) dt + o(dt), \text{ and} \\
 & \Pr \left\{ \begin{array}{l} X_{\text{old}}(t+dt) = x_{\text{old}} \\ X_{\text{new}}(t+dt) = x_{\text{new}} \end{array} \middle| \begin{array}{l} X_{\text{old}}(t) = x_{\text{old}} \\ X_{\text{new}}(t) = x_{\text{new}} \end{array} \right\} \\
 & \quad = 1 - (\mu + \lambda + \nu)x_{\text{old}} - (\mu + \lambda)x_{\text{new}} dt + o(dt). \tag{29}
 \end{aligned}$$

Equivalent to the branching process representation, notice that in this case $\mathbf{X}(t)$ is also a death/birth-death process. Hence, we can effectively compute its transition probabilities. In contrast, [Xu et al. \(2015\)](#) consider the probability generating function

$$\begin{aligned} \Phi_{a_0b_0}(t, s_1, s_2) &= \mathbb{E} \left(s_1^{X_{\text{old}}(t)} s_2^{X_{\text{new}}(t)} \mid X_{\text{old}}(0) = a_0, X_{\text{new}}(0) = b_0 \right) \\ &= \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} P_{ab}^{a_0b_0}(t) s_1^a s_2^b, \end{aligned} \tag{30}$$

where

$$P_{ab}^{a_0b_0}(t) = \Pr \left\{ \begin{array}{l} X_{\text{old}}(t) = a \\ X_{\text{new}}(t) = b \end{array} \mid \begin{array}{l} X_{\text{old}}(0) = a_0 \\ X_{\text{new}}(0) = b_0 \end{array} \right\}. \tag{31}$$

Because of the model-specific linearity in terms of a and b of the birth and death rates, one can evaluate $\Phi_{jk}(t, s_1, s_2)$ by solving an ordinary differential equation. Further transforming $s_1 = e^{2\pi i w_1}$, $s_2 = e^{2\pi i w_2}$, the generating function becomes a Fourier series

$$\Phi_{a_0b_0}(t, e^{2\pi i w_1}, e^{2\pi i w_2}) = \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} P_{ab}^{a_0b_0}(t) e^{2\pi i a w_1} e^{2\pi i b w_2}. \tag{32}$$

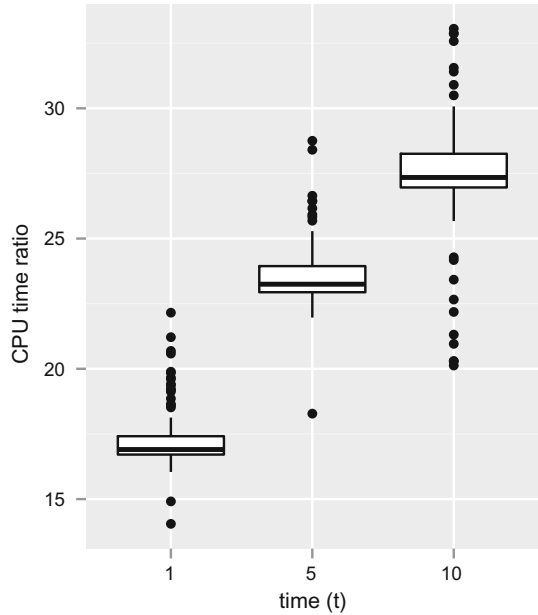
Therefore, [Xu et al. \(2015\)](#) retrieve the transition probabilities through approximating the integral as a Riemann sum

$$\begin{aligned} P_{ab}^{a_0b_0}(t) &= \int_0^1 \int_0^1 \Phi_{a_0b_0}(t, e^{2\pi i w_1}, e^{2\pi i w_2}) e^{-2\pi i a w_1} e^{-2\pi i b w_2} dw_1 dw_2 \\ &\approx \frac{1}{H^2} \sum_{u=0}^{H-1} \sum_{v=0}^{H-1} \Phi_{jk}(t, e^{2\pi i u/H}, e^{2\pi i v/H}) e^{-2\pi i a u/H} e^{-2\pi i b v/H}, \end{aligned} \tag{33}$$

and show that choosing H as the smallest power of 2 greater than $\max(a, b)$ produces accurate estimates of the true transition probabilities of the model. The authors implement this method in the R package `bdsem`. Using their method, evaluating $\{P_{ab}^{a_0b_0}(t)\}_{0 \leq a, b \leq H}$ requires numerically solving H^2 linear ordinary differential equations (ODEs). We perform a simulation to compare the performance between `bdsem` and our function `dbd_prob`. Because [Xu et al. \(2015\)](#) already provide a thorough empirical validation that `bdsem` produces accurate transition probabilities compared to Monte Carlo estimates from the true model, we consider a comparison to their method and omit a complete reproduction of their simulation study. Using both routines to compute the transition probabilities of a birth-death-shift process with rates $\lambda = 0.0188$, $\mu = 0.0147$, $\nu = 0.00268$ (estimated from the *IS6110* data by [Rosenberg et al. \(2003\)](#)) repeatedly over one hundred trials leads to a negligible difference in estimated probabilities. Specifically, we computed $\{P_{ab}^{10,0}(t)\}_{0 \leq a \leq 10, 0 \leq b \leq 50}$ at three different observation period lengths $t = 1, 5, 10$, and found that the L_1 distance between probabilities estimated by each method is less than 4×10^{-8} across all cases. Here, the L_1 distance between two matrices $\mathbf{U} = (u_{ij})$ and $\mathbf{V} = (v_{ij})$ are defined as $\|\mathbf{U} - \mathbf{V}\| = \sum_{i,j} |u_{ij} - v_{ij}|$.

Having validated the accuracy of our approach, we turn to a runtime comparison. The ratios of CPU time required using `bdsem` compared to `dbd_prob` are summarized in [Fig. 1](#), and note that this result is obtained using a single-thread option for `dbd_prob`. We see that `dbd_prob` is about 15 to 30 times faster than the `bdsem` implementation, while producing very similar results.

Fig. 1 CPU compute time ratios of `bdsem` to `dbd_prob` over 100 replications



While there is a large performance difference in wall clock time, we cannot immediately conclude that our method is faster than the method in Xu et al. (2015) because computation time may depend heavily on implementation. Nonetheless, we can make some remarks about the performance of both methods that are platform-independent. Notably, the `bdsem` implementation grows slower as t increases while `dbd_prob` does not. This is expected because solving ODEs is slower when the domain increases. However, it is worth mentioning that we can use the solution paths to get the solutions of these ODEs at other time points in the domain. For example, when we solve the ODEs at $t = 10$, we also get the solutions at $t = 1$ and 5 for free. This point becomes important in applications where we need to compute the transition probabilities at several time points. Another downside of `bdsem` is that it computes $\{P_{ab}^{10,0}(t)\}_{0 \leq a, b \leq 50}$ instead of evaluating $\{P_{ab}^{10,0}(t)\}_{0 \leq a \leq 10, 0 \leq b \leq 50}$ directly as is done by `dbd_prob`.

3.3 Within-host macro-parasite model

Riley et al. (2003) posit a stochastic model to describe a within-host macro-parasite population where *Brugia pahangi* is the parasite and *Felis catus* is the host. *Brugia pahangi* is closely related to *Brugia malayi* which infects millions of people in South and Southeast Asia. The model tracks the number of *B. pahangi* larvae $L(t)$, the number of mature parasites $M(t)$, and hosts experience of infection $I(t)$ at time t . The dynamics of $\{L(t), M(t), I(t)\}$ follow a system of differential equations:

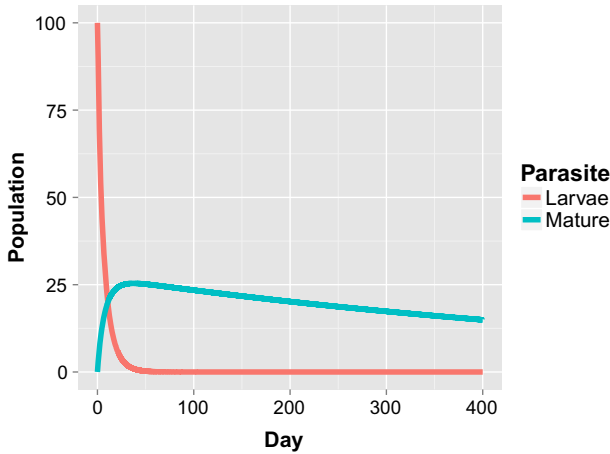


Fig. 2 The dynamic of $\{L(t), M(t)\}$ under the deterministic model (35) with $\mu_L = 0.0682, \mu_M = 0.0015, \eta = 0.0009, \gamma = 0.04$ and $\{L(0), M(0)\} = \{100, 0\}$

$$\begin{aligned} \frac{dL}{dt}(t) &= -\mu_L L(t) - \beta I(t)L(t) - \gamma L(t), \\ \frac{dM}{dt}(t) &= \gamma L(t) - \mu_M M(t), \text{ and} \\ \frac{dI}{dt}(t) &= \nu L(t) - \mu_I I(t) \end{aligned} \tag{34}$$

where μ_L is the natural death rate and γ is the maturation rate of larvae; β is the death rate of larvae due to the immune response from the host; μ_M is the death rate of mature parasites; ν is the acquisition rate and μ_I is the loss rate of immunity.

Drovandi and Pettitt (2011) propose a simplification of this model by applying a pseudoequilibrium assumption for immunity, such that the immunity is constant over time. Under this pseudoequilibrium assumption, the dynamics of $\{L(t), M(t)\}$ becomes

$$\begin{aligned} \frac{dL}{dt}(t) &= -\mu_L L(t) - \eta [L(t)]^2 - \gamma L(t), \text{ and} \\ \frac{dM}{dt}(t) &= \gamma L(t) - \mu_M M(t) \end{aligned} \tag{35}$$

where $\eta = \beta\nu/\mu_I$. We illustrate the dynamic of (35) in Fig. 2. The corresponding stochastic formulation of this model is:

$$\begin{aligned} \Pr \left\{ \begin{matrix} L(t+dt) = i-1 \\ M(t+dt) = j+1 \end{matrix} \middle| \begin{matrix} L(t) = i \\ M(t) = j \end{matrix} \right\} &= (\gamma i)dt + o(dt), \\ \Pr \left\{ \begin{matrix} L(t+dt) = i-1 \\ M(t+dt) = j \end{matrix} \middle| \begin{matrix} L(t) = i \\ M(t) = j \end{matrix} \right\} &= (\mu_L i + \eta i^2)dt + o(dt), \end{aligned}$$

$$\begin{aligned}
 \Pr \left\{ \begin{array}{l} L(t + dt) = i \\ M(t + dt) = j - 1 \end{array} \middle| \begin{array}{l} L(t) = i \\ M(t) = j \end{array} \right\} &= (\mu_M j)dt + o(dt), \text{ and} \\
 \Pr \left\{ \begin{array}{l} L(t + dt) = i \\ M(t + dt) = j \end{array} \middle| \begin{array}{l} L(t) = i \\ M(t) = j \end{array} \right\} &= 1 - (\gamma i + \mu_L i + \eta i^2 + \mu_M j)dt + o(dt).
 \end{aligned}
 \tag{36}$$

Notably, $\{L(t), M(t)\}$ follow a death/birth-death process.

For this model, γ and μ_M has been estimated at 0.04 and 0.0015 previously (see [Drovandi and Pettitt 2011](#) for more details). To estimate the remaining parameters, [Drovandi and Pettitt \(2011\)](#) examine the number of mature parasites at host autopsy time (at most 400 days) of those injected with approximately 100 juveniles, assume *a priori* μ_L and η are uniform[0,1) and apply ABC to draw inference because the traditional matrix exponentiation method is computationally prohibitive here. The basic idea of ABC involves sampling from an approximate posterior distribution

$$f(\boldsymbol{\theta}, Y | \rho(Y, Y_s) \leq \epsilon) \propto f(Y_s | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) 1_{\rho(Y, Y_s) \leq \epsilon},
 \tag{37}$$

where $\boldsymbol{\theta}$ is the vector of unknown parameters, $\epsilon > 0$ is an *ad hoc* tolerance, and $\rho(Y, Y_s)$ is a discrepancy measure between summary statistics of the observed data Y and the simulated data Y_s . Because the sufficient statistics are not available for this problem, the authors use a goodness-of-fit statistic. However, the ABC method suffers from loss of information because of non-zero tolerance and non-sufficient summary statistics ([Sunnåker et al. 2013](#)). Therefore, credible intervals obtained by the ABC approach are potentially inflated ([Csilléry et al. 2010](#)).

In contrast, our method makes direct likelihood computation and in turn evaluation of the posterior density feasible. Figure 3 displays a visualization of the posterior density surface of $(\log \mu_L, \log \eta)$ computed using our method, given the collection of numbers of mature parasites $M(t)$ at autopsy under this model (see [Drovandi and Pettitt 2011](#) for more details about the data). Importantly in this example, we are able to efficiently integrate out the unobserved larvae counts $L(t)$ at autopsy. The approximate estimate obtained by [Drovandi and Pettitt \(2011\)](#) using ABC is overlaid on this density surface for comparison, and does not align with the highest density region of our computed posterior. Note that the posterior is flat when η is close to 0, and has an unusual tail toward the region where the ABC estimate lies. This suggests that the previous ABC approach fails to explore the region with high posterior probability well, likely due to loss of information incurred by the method, resulting in a poor estimate from the data.

Finally, we consider this example toward a second runtime comparison between our method and `Expokit`, a state-of-the-art matrix exponentiation package with efficient implementation. In particular, we compute the transition probability matrix $\{P_{ij}^{100,0}(t)\}_{0 \leq i \leq 100, 0 \leq j \leq 100}$ of $\{L(t), M(t)\}$ with $\mu_L = 0.0682, \mu_M = 0.0015, \eta = 0.0009, \gamma = 0.04$ at $t = 100, 200, 400$ using our function `dbd_prob` and the function `expv` in `expokit`, an R-interface to the Fortran package `Expokit`. Both methods produce similar results: the L_1 distance between the two estimated transition probability matrices is less than 3×10^{-9} across all cases. In terms of speed, we see that `dbd_prob` is roughly twice as fast as `expv` when $t = 100, 200$, but about 9-fold

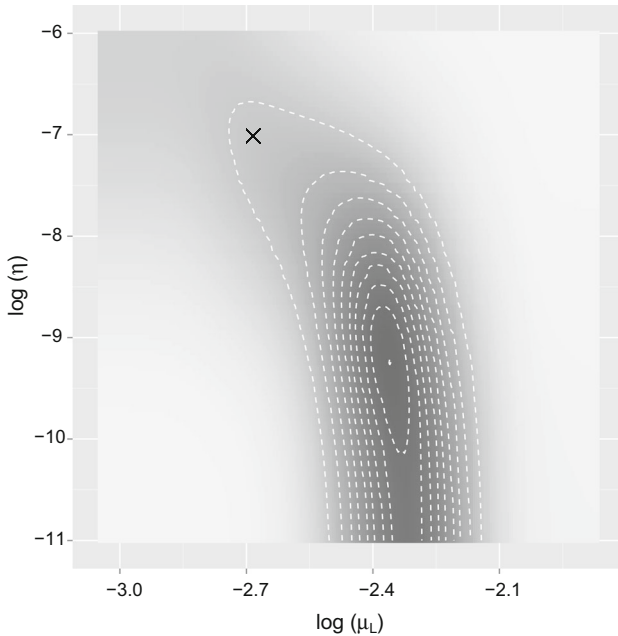


Fig. 3 Posterior density surface of $(\log \mu_L, \log \eta)$ under within-host macro-parasite model. The “x” symbol represents the estimate from [Drovandi and Pettitt \(2011\)](#) using the ABC method

faster when $t = 400$ (Fig. 4). It is worth mentioning that `dbd_prob` can be further accelerated via parallelization.

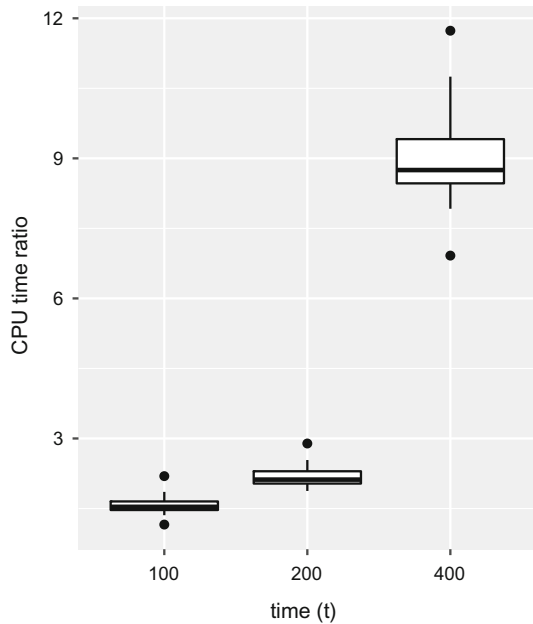
3.4 Stochastic SIR model in epidemiology

[McKendrick \(1926\)](#) models the spread of an infectious disease in a closed population by dividing the population into three categories: susceptible persons (S), infectious persons (I) and removed persons (R). Since the population is closed, the total population size N obeys the conservation equation $N = S(t) + I(t) + R(t)$ for all time t . The deterministic dynamics of these three subpopulations follow a system of nonlinear ordinary differential equations ([Kermack and McKendrick 1927](#)):

$$\begin{aligned}
 \frac{dS}{dt}(t) &= -\beta S(t)I(t), \\
 \frac{dI}{dt}(t) &= \beta S(t)I(t) - \alpha I(t), \text{ and} \\
 \frac{dR}{dt}(t) &= \alpha I(t),
 \end{aligned}
 \tag{38}$$

where $\alpha > 0$ is the removal rate and $\beta > 0$ is the infection rate of the disease. This system of equations cannot be solved analytically, but we can obtain its solution numerically. An important quantity for the SIR model is the basic reproduction number

Fig. 4 CPU compute time ratios of `expv` to `dbd_prob` over 100 replications



$R_0 = \beta N / \alpha$ (Earn 2008). This quantity determines whether a spread of an infectious disease becomes an epidemic. In particular, an epidemic can only occur when $R_0 > 1$.

Unfortunately, the deterministic model is not suitable when the community is small (Britton 2010). In these situations, the original stochastic SIR model (McKendrick 1926) becomes more appropriate. Moreover, Andersson and Britton (2000) argue that stochastic epidemic models are preferable when their analysis is possible because (1) stochastics are the most natural way to describe a spread of diseases, (2) some phenomena do not satisfy the law of large numbers and can only be analyzed in the stochastic setting (for example, the extinction of endemic diseases only occurs when the epidemic process deviates from its expected value), and (3) quantifying the uncertainty in estimates requires stochastic models. Nonetheless, one can bypass Andersson and Britton’s third argument by imposing random sampling errors around the deterministic compartments. Therefore, it is important to distinguish between the deterministic SIR model with sampling errors and the stochastic SIR model.

Without loss of generality, the stochastic SIR model needs only track $S(t)$ and $I(t)$ because $S(t) + I(t) + R(t)$ remains constant. All possible transitions of $\{S(t), I(t)\}$ during a small time interval $(t, t + dt)$ occur with probabilities

$$\begin{aligned}
 \Pr \left\{ \begin{array}{l} S(t + dt) = s \\ I(t + dt) = i - 1 \end{array} \middle| \begin{array}{l} S(t) = s \\ I(t) = i \end{array} \right\} &= (\alpha i)dt + o(dt), \\
 \Pr \left\{ \begin{array}{l} S(t + dt) = s - 1 \\ I(t + dt) = i + 1 \end{array} \middle| \begin{array}{l} S(t) = s \\ I(t) = i \end{array} \right\} &= (\beta si)dt + o(dt), \text{ and} \\
 \Pr \left\{ \begin{array}{l} S(t + dt) = s \\ I(t + dt) = i \end{array} \middle| \begin{array}{l} S(t) = s \\ I(t) = i \end{array} \right\} &= 1 - (\alpha i + \beta si)dt + o(dt). \quad (39)
 \end{aligned}$$

We see that $\{S(t), I(t)\}$ is a death/birth-death process with $\mu_{si}^{(1)} = \lambda_{s,i}^{(2)} = 0, \mu_s^{(2)} = \alpha i, \gamma_{si} = \beta si$.

Due to the interaction between populations and nonlinear nature of the model, mechanistic analysis of the stochastic SIR model is difficult, and the lack of an expression for transition probabilities has been a bottleneck for statistical inference. [Renshaw \(2011\)](#) remarks that while one can write out the Kolmogorov forward equation for the system, the “associated mathematical manipulations required to generate solutions can only be described as heroic.” Instead, the majority of efforts involve either simulation based methods or simplifications and tractable approximations to the SIR model. For instance, the stochastic SIR model can be analyzed using ABC ([McKinley et al. 2009](#)), but we have already mentioned limitations of this approach. Particle filter methods can be used to analyze SIR models within maximum likelihood ([Ionides et al 2006; Ionides et al. 2015](#)) and Bayesian frameworks ([Andrieu et al. 2010; Dukic et al. 2012](#)), but these methods are computationally very demanding and often suffer from convergence problems. When examining large epidemics, to make the likelihood tractable it is reasonable to apply a continuous approximation to the large populations, modeled as a diffusion process with exact solutions ([Cauchemez and Ferguson 2008](#)). However, such an approach is a poor proxy for the SIR model when observed counts are low. When data are collected at regular intervals and coincide with disease generation timescales, it is also possible to study discrete-time epidemic models—the time-series SIR (TSIR) model is one well-known example ([Finkenstädt and Grenfell 2000](#)). However, these simplifications also have their shortcomings, relying on the relatively strong assumption that populations are constant over each interval between observation times.

In the death/birth-death framework, our method enables practical computation of these quantities *without* any simplifying model assumptions. In Sect. 4, we will apply our method to analyze the population of Eyam during the plague of 1666 ([Raggett 1982](#)) to estimate the infection and the death rates of this disease, using the death/birth-death transition probabilities within a Metropolis-Hastings algorithm. Here, we first examine the accuracy of these transition probabilities themselves. We compare the continued fraction method to empirical transition probabilities obtained via simulation from the true model as ground-truth, and to a new two-type branching approximation to the SIR model introduced below. The branching process approximation is appropriate when transition probabilities need to be computed for short time intervals, and its simple expressions for transition probabilities enable much more efficient computation. However, we show that as transition time intervals increase, the branching approximation becomes less accurate, while the transition probabilities computed under the death/birth-death model remain very accurate.

While branching processes fundamentally rely on *independence* of each member of the population, we can nonetheless make a fair approximation by mimicking the interaction effect of infection over short time intervals. In the branching model, let $X_1(t)$ denote the susceptible population and $X_2(t)$ denote the infected population at time t , with details and derivation included in “Appendix D”. Over any time interval $[t_0, t_1)$, we use the initial population $X_2(0)$ as a constant scalar for the instantaneous

rates. This branching process model has instantaneous infection rate $\beta X_2(0)X_1(t)$ and recovery rate $\alpha X_2(t)$ for all $t \in [t_0, t_1]$, closely resembling the true SIR model rates, with the exception of fixing $X_2(0)$ in place of $X_2(t)$ in the rate of infection. This constant initial population fixes a piecewise homogeneous per-particle birth rate to satisfy particle independence while mimicking interactions, but notice that *both* populations can change over the interval, offering much more flexibility than models such as TSIR that assume constant populations and rates between discrete observations.

This branching model admits closed-form solutions to the transition probabilities that can be evaluated quickly and accurately. The transition probabilities of the two-type branching approximation to the SIR model over any time interval of length t are given by

$$\Pr \{ \mathbf{X}(t + \tau) = (k, l) | \mathbf{X}(\tau) = (m, n) \} := P_{kl}^{mn}(t) = \sum_{i=0}^l \binom{l}{i} A(l-i)B(i), \tag{40}$$

where

$$B(i) = 0 \text{ for all } i \geq n, \text{ otherwise,}$$

$$B(i) = \frac{n!}{(n-i)!} (1 - e^{-\alpha t})^{n-i} e^{-i\alpha t} \tag{41}$$

and

$$A(l-i) = 0 \text{ for all } (l-i) \geq (m-k), \text{ otherwise,}$$

$$A(l-i) = \frac{m!}{(m-k-(l-i))!} e^{-k\beta nt}$$

$$\times \left[1 - \frac{\beta n}{\beta n - \alpha} e^{-\alpha t} - \left(1 - \frac{\beta n}{\beta n - \alpha} \right) e^{-\beta nt} \right]^{m-k-(l-i)}$$

$$\times \left[\frac{\beta n}{\beta n - \alpha} (e^{-\alpha t} - e^{-\beta nt}) \right]^{l-i}. \tag{42}$$

The sum over products of expressions (41) and (42) in Eq. (40) may look unwieldy, but this sum is computed extremely quickly with a vectorized implementation, and with high degrees of numerical stability. In settings when such a model is appropriate and $(X_1(t), X_2(t)) \approx (S(t), I(t))$, the branching approximation can offer a much more computationally efficient alternative to the continued fraction method.

3.5 Transition probabilities of the SIR model

Figure 5 provides a comparison between methods of computing transition probabilities. Included are transition probabilities corresponding to the nine pairs of system states $\{(m, n), (k, l)\}_j, j = 1, \dots, 9$, such that $P_{kl}^{mn}(0.5)$ is largest. Fixing these

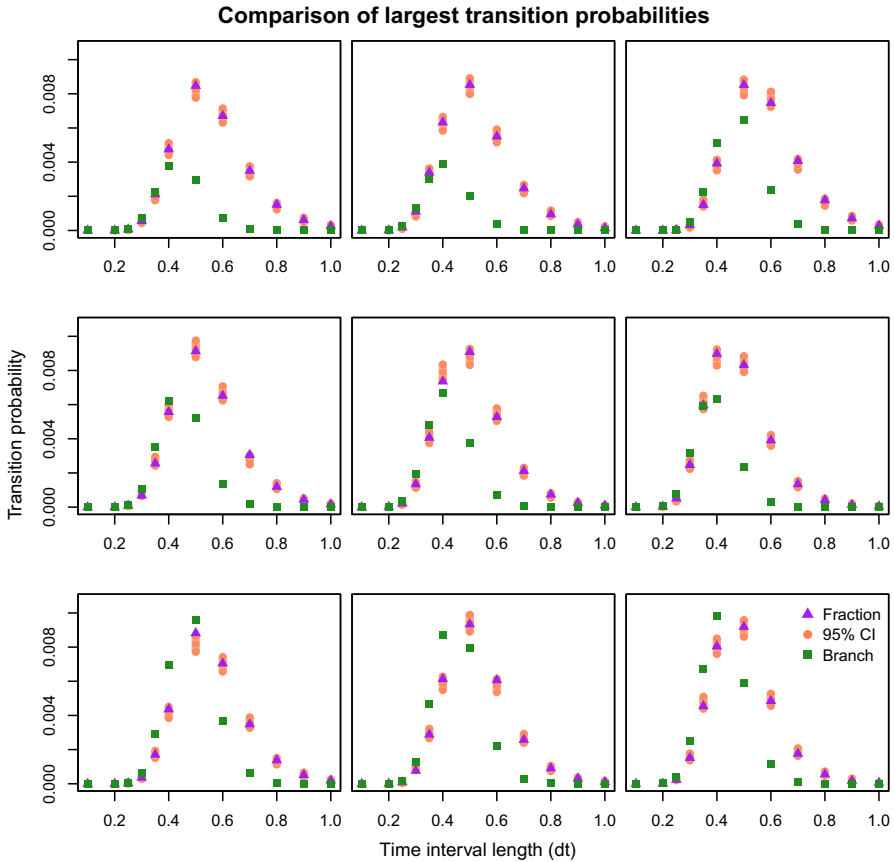


Fig. 5 The plot above displays the values of the nine largest transition probabilities when $t = 0.5$ as we vary t from 0.1, ..., 1.0. Parameters used to generate data are initialized at $I_0 = 15, S_0 = 110, \alpha = 3.2, \beta = 0.025$. Empirical Monte Carlo 95% confidence intervals over 150,000 simulations from the true model are depicted in orange. Probabilities computed using the continued fraction expansion are depicted by purple triangles, while probabilities computed under the branching approximation are denoted by green squares

indices, we plot the set of probabilities $\{P_{kl}^{mn}(t)\}$ while varying t between 0.1 and 1.0. We see that transition probabilities computed using the continued fraction method under the death/birth-death model very closely match those computed empirically via simulation from the model, taken to be the ground truth. Almost all such probabilities in Fig. 5 fall within the 95% confidence interval, while the branching process transitions follow a similar shape over time, but fall outside of the confidence intervals for many observation intervals. An additional heatmap visualization comparing the support of transition probabilities is included in the Appendix, and shows that the branching approximation is accurate with similar support to the empirical transition probabilities for a shorter time interval of length $t = 0.5$, but becomes visibly further from the truth when we increase the observation length to $t = 1.0$.

Table 1 Susceptible and infectious population size in Eyam from June 18th to October 20th, 1666

	Time (months)							
	0	0.5	1	1.5	2	2.5	3	4
Susceptible population	254	235	201	153	121	110	97	83
Infective population	7	14	22	29	20	8	8	0

4 The Plague in Eyam revisited

We revisit the outbreak of plague in Eyam, a village in the Derbyshire Dales district, England, over the period from June 18th to October 20th, 1666. This plague outbreak is widely accepted to originate from the Great Plague of London, that killed about 15% of London’s population at that time. To prevent further spread of the plague after infestation, the Eyam villagers did not escape the village, instead isolating themselves from the outside world. At the end of this horrific event, only 83 people had survived out of an initial population of 350. We summarize data recording the spread of the disease (Raggett 1982) in Table 1. As mentioned in Raggett (1982), this data are obtained by counting the number of deaths from the dead list and estimating the infective population from the list of future deaths assuming a fixed length of illness prior to death. Then, the susceptible population can be computed easily because the the town is isolated.

Raggett (1982) analyzes these data using the stochastic SIR model (39). In this model, α is the unknown death rate of infective people and β is the unknown infection rate of the plague. The author uses a simple approximation method for the forward differential equation and comes up with a point estimate $(\hat{\alpha}, \hat{\beta}) = (3.39, 0.0212)$. We take a Bayesian approach to re-analyze these data.

With n observations $\{(s_k, i_k)\}_{k=1}^n$ at time $\{t_k\}_{k=1}^n$, the log of the likelihood function is:

$$\log l(\alpha, \beta | \{(s_k, i_k)\}_{k=1}^n) = \sum_{k=1}^{n-1} \log \Pr \left\{ \begin{matrix} S(t_{k+1}) = s_{k+1} \\ I(t_{k+1}) = i_{k+1} \end{matrix} \middle| \begin{matrix} S(t_k) = s_k \\ I(t_k) = i_k \end{matrix} \right\}. \tag{43}$$

Because $\{S(t), I(t)\}$ is a death/birth-death process, the individual transition probabilities can be computed efficiently using our continued fraction method. Hence, the log of the likelihood (43) can be computed easily. Since α and β are non-negative, we opt to use $\log \alpha$ and $\log \beta$ as our model parameters and assume *a priori* that $\log \alpha \sim \mathcal{N}(\mu = 0, \sigma = 100)$ and $\log \beta \sim \mathcal{N}(\mu = 0, \sigma = 100)$. We explore the posterior distribution of $(\log \alpha, \log \beta)$ using a random-walk Metropolis algorithm implemented in the R function `MCMCmetrop1R` from package `MCMCpack` (Martin et al. 2011). We start the chain from Raggett’s estimated value $(\log(3.39), \log(0.0212))$ and run it for 100000 iterations. We discard the first 20000 iterations and summarize the posterior distribution of (α, β) using the remaining iterations. We illustrate the density of this posterior distribution in Fig. 6a. The posterior mean of α is 3.22 and the 95% Bayesian credible interval for α lies in (2.69, 3.82). Those corresponding quan-

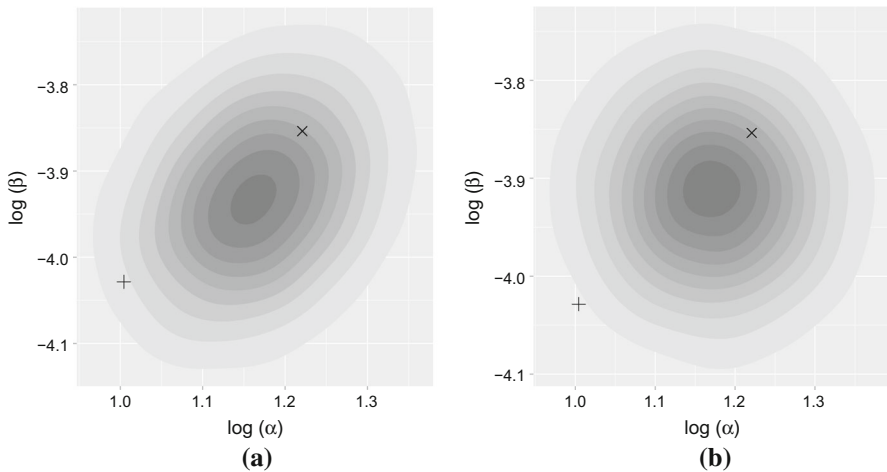


Fig. 6 Posterior distributions (log scale) of the death rate α and the infection rate β during the plague of Eyam in 1666. The “+” symbol represents the estimate from Brauer (2008) using the deterministic SIR model, and the “x” symbol represents the Raggett’s point estimate. **a** Continued fraction method **b** branching approximation method

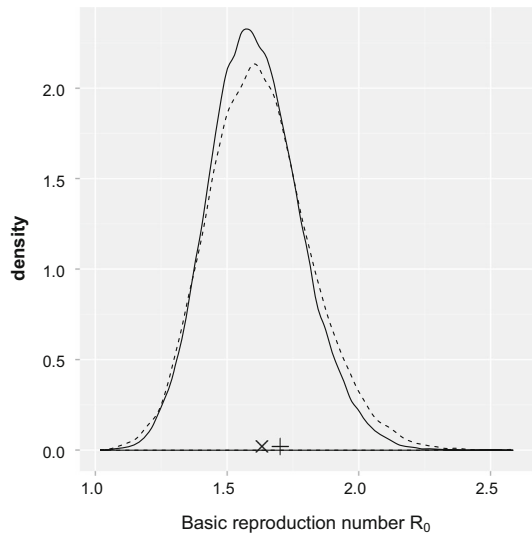
ties for β are 0.0197 and (0.0164, 0.0234). Notice that our credible intervals include the point estimate $(\hat{\alpha}, \hat{\beta}) = (2.73, 0.0178)$ from Brauer (2008) using the deterministic SIR model and Raggett’s point estimate $(\hat{\alpha}, \hat{\beta}) = (3.39, 0.0212)$.

We also apply the two-type branching approximation to compute the log of the likelihood (43). Using the same random-walk Metropolis algorithm as before, we explore the posterior distribution of (α, β) and visualize it in Fig. 6b. The posterior mean of α is 3.237 and the 95% Bayesian credible interval for α is (2.7, 3.84), while those quantities for β are 0.02 and (0.0171, 0.023). Although the posterior means and the 95% Bayesian credible intervals are similar to ones from the continued fraction method, we see in Fig. 6b that this method fails to fully capture the posterior correlation structure between α and β .

The posterior distribution of the basic reproduction number R_0 from the continued fraction method and from the branching approximation method are similar (Fig. 7). The posterior mean of R_0 from the continued fraction method is 1.61 and from the branching approximation method is 1.62. The estimate for R_0 from Brauer (2008) is 1.7, from Raggett (1982) is 1.63. These estimates are similar, and in particular the branching approximation estimate is very close to that under the continued fraction method, offering a very efficient way to provide reasonable estimates of quantities such as R_0 despite being less accurate than the continued fraction approach.

From the results, we can see that estimates of R_0 from different methods are roughly the same while estimates of α and β are different. Although the basic reproduction number R_0 is an important quantity in the SIR model, it is not the only parameter driving the dynamic of the epidemic. Correia-Gomes et al. (2014) demonstrated the important of accurately estimating the transmission parameters between compartments of the SIR model for *Salmonella* Typhimurium in pigs.

Fig. 7 Posterior distribution of the basic reproduction number R_0 (solid line: continued fraction method, *dashed line* branching approximation method). The “+”, and the “x” symbols represent the estimate of R_0 from Brauer (2008), and from Raggett (1982) respectively



5 Discussion

Likelihood-based inference for bivariate continuous-time Markov processes is usually restricted to very small state spaces due to the computational bottleneck of transition probability calculation. In this paper, we provide tools for likelihood-based inference for birth(death)/birth-death processes by developing an efficient method to compute their transition probabilities. We provide a complete implementation of the algorithms to compute these transition probabilities in a new R package called `MultiBD`. Our functions employ sophisticated tools including continued fractions, the modified Lentz method, the method of Abate and Whitt for approximate inverse Laplace transforms, and the Levin acceleration method. Moreover, these methods are naturally amenable to parallelization, and we exploit multicore processing to speed up the algorithm. We remark that birth(death)/birth-death processes remain a limited subclass of general multivariate birth-death processes. For example, many population biology problems require a full bivariate birth-death process including predator-prey models (Hitchcock 1986; Owen et al. 2015) and the SIR model with vital dynamics (Earn 2008). Unfortunately, efficiently computing the transition probabilities of multivariate birth-death processes remains an open problem. Solving this problem will enable numerically stable statistical inference under birth-death processes and will be worth the “heroic” effort (Renshaw 2011).

A Continued fractions

In this section, we give some basic definitions and properties related to continued fractions.

Definition A1 A continued fraction ϕ_0 is a scalar quantity expressed in

$$\phi_0 = \frac{x_1}{y_1 + \frac{x_2}{y_2 + \frac{x_3}{y_3 + \dots}}}, \tag{A.1}$$

where $\{x_i\}_{i=1}^\infty$ and $\{y_i\}_{i=1}^\infty$ are infinite sequences of complex numbers.

Definition A2 The n^{th} convergent of ϕ_0 is

$$\frac{X_n}{Y_n} = \frac{x_1}{y_1 + \frac{x_2}{y_2 + \frac{x_3}{y_3 + \dots + \frac{x_n}{y_n}}}}. \tag{A.2}$$

Definition A3 We define the corresponding sequence $\{\phi_n\}_{n=0}^\infty$ of a continued fraction (A.1) by the following recurrence formulae

$$\begin{aligned} \phi_1 &= x_1 - y_1\phi_0, \text{ and} \\ \phi_n &= x_n\phi_{n-2} - y_n\phi_{n-1} \text{ for } n \geq 2. \end{aligned} \tag{A.3}$$

Murphy and O’Donohoe (1975) provided the following sufficient condition for the convergence of (A.1):

Lemma A1 Assume that there exists N such that $\inf_{n>N} |Y_n| > 0$ and $\lim_{n \rightarrow \infty} \phi_n = 0$. Then, the continued fraction (A.1) is convergent. Moreover;

$$\phi_n = \prod_{i=1}^n x_i \frac{x_{n+1}}{Y_{n+1} + \frac{x_{n+2}Y_n}{y_{n+2} + \frac{x_{n+3}}{y_{n+3} + \frac{x_{n+4}}{y_{n+4} + \dots}}}}. \tag{A.4}$$

Now, if we consider a more general recurrence formulae

$$\begin{aligned} \phi_1^{(m)} &= -y_1\phi_0^{(m)} + k_1 1_{\{m=0\}} \\ \phi_n^{(m)} &= x_n\phi_{n-2}^{(m)} - y_n\phi_{n-1}^{(m)} + k_{m+1} 1_{\{m=n-1\}} \text{ for } n \geq 2, \end{aligned} \tag{A.5}$$

then under the assumption of Lemma A.4, we have the following lemma:

Lemma A2 *The solution for (A.5) is*

$$\phi_n^{(m)} = \begin{cases} \frac{(-1)^{m-n} k_{m+1}}{\prod_{i=1}^{m+1} x_i} Y_n \phi_m, & \text{if } n \leq m \\ \frac{k_{m+1}}{\prod_{i=1}^{m+1} x_i} Y_m \phi_n, & \text{if } n \geq m. \end{cases} \tag{A.6}$$

B Modified Lentz method

Modified Lentz method (Lentz 1976; Thompson and Barnett 1986) is an efficient algorithm to finitely approximate the infinite expression of the continued fraction ϕ_0 in (A.1) to within a prescribed error tolerance. Let $\phi_0^{(n)}$ be the n^{th} convergence of ϕ_0 , that is $\phi_0^{(n)} = X_n/Y_n$. The main idea of Lentz’s algorithm lies in using the ratios

$$A_n = \frac{X_n}{X_{n-1}} \text{ and } B_n = \frac{Y_{n-1}}{Y_n} \tag{B.1}$$

to stabilize the computation of $\phi_0^{(n)}$. We can calculate A_n, B_n , and $\phi_0^{(n)}$ recursively as follows:

$$\begin{aligned} A_n &= y_n + \frac{x_n}{A_{n-1}} \\ B_n &= \frac{1}{y_n + x_n B_{n-1}} \\ \phi_0^{(n)} &= \phi_0^{(n-1)} A_n B_n. \end{aligned} \tag{B.2}$$

If $\phi_0^{(n)}$ converges to ϕ_0 , then Craviotto et al. (1993) show that

$$\left| \phi_0^{(n)} - \phi_0 \right| \leq \frac{|Y_n/Y_{n-1}|}{\mathcal{I}[Y_n/Y_{n-1}]} \left| \phi_0^{(n)} - \phi_0^{(n-1)} \right| = \frac{|1/B_n|}{\mathcal{I}[1/B_n]} \left| \phi_0^{(n)} - \phi_0^{(n-1)} \right|, \tag{B.3}$$

where $\mathcal{I}[Y_n/Y_{n-1}]$ is the imaginary part of Y_n/Y_{n-1} and is assumed to be non-zero. Hence, the Lentz’s algorithm terminates when

$$\frac{|1/B_n|}{\mathcal{I}[1/B_n]} \left| \phi_0^{(n)} - \phi_0^{(n-1)} \right| \tag{B.4}$$

is small enough. However, A_n and B_n can equal zero themselves and cause problem. Hence, Thompson and Barnett (1986) propose a modification for Lentz’s algorithm by setting A_n and B_n to a very small number, such as 10^{-16} , whenever they equal zero. In practice, the algorithm often terminates after small number of iterations. However, in some rare cases where the numerical computation is unstable, it might take too long before the algorithm terminates. So, we set a predefined maximum number of iterations H as a fallback for these cases.

C Convergence results of increasing the truncation level

Let $f_{ab}^{(B)}(s)$ be the output of the approximation scheme (19) in Theorem 2. In this section, we prove that $f_{ab}^{(B)}(s)$ converges to $f_{ab}(s)$ as B goes to infinity. To do so, let us consider a truncated birth/birth-death process $\mathbf{X}^{(B)}(t) = (X_1^{(B)}(t), X_2^{(B)}(t))$ at truncation level B such that it executes the same process as $\mathbf{X}(t)$ on the state $\{a_0, a_0 + 1, a_0 + 2, \dots\} \times \{0, 1, 2, \dots, B\}$ except that $\lambda_{aB}^{(2)} = 0$. Define $P_{ab}^{a_0b_0, (B)}(t)$ be the transition probabilities of $\mathbf{X}^{(B)}(t)$ and T_B be the hitting time at which $X_2(t)$ first reach state $B + 1$. For any set $S \subset \mathbb{N}^2$, we have

$$\begin{aligned} \Pr(\mathbf{X}(t) \in S) &= \Pr(\mathbf{X}(t) \in S | T_B > t) \Pr(T > t) + \Pr(\mathbf{X}(t) \in S | T_B \leq t) \Pr(T_B \leq t) \\ &= \Pr(\mathbf{X}^{(B)}(t) \in S) \Pr(T_B > t) + \Pr(\mathbf{X}(t) \in S | T_B \leq t) \Pr(T_B \leq t) \\ &= \Pr(\mathbf{X}^{(B)}(t) \in S) + [\Pr(\mathbf{X}(t) \in S | T_B \leq t) \\ &\quad - \Pr(\mathbf{X}^{(B)}(t) \in S)] \Pr(T_B \leq t) \end{aligned}$$

Therefore $|\Pr(\mathbf{X}(t) \in S) - \Pr(\mathbf{X}^{(B)}(t) \in S)| \leq \Pr(T_B \leq t)$. Note that $f_{ab}^{(B)}(s)$ is the Laplace transform of $P_{ab}^{a_0b_0, (B)}(t)$. Hence

$$\begin{aligned} |f_{ab}^{(B)}(s) - f_{ab}(s)| &\leq \int_0^\infty |P_{ab}^{a_0b_0, (B)}(t) - P_{ab}^{a_0b_0}(t)| e^{-st} dt \\ &\leq \int_0^\infty \Pr(T_B \leq t) e^{-st} dt \end{aligned} \tag{C.1}$$

By Dominated convergence theorem and the fact that $\lim_{B \rightarrow \infty} \Pr(T_B \leq t) = 0$, we deduce that $\lim_{B \rightarrow \infty} f_{ab}^{(B)}(s) = f_{ab}(s)$.

D Branching SIR approximation

Here we derive and solve the Kolmogorov backward equations of the two-type branching process necessary for evaluating the probability generating functions (PGFs) whose coefficients yield transition probabilities.

D.1 Deriving the PGF

Our two-type branching process is represented by a vector $(X_1(t), X_2(t))$ that denotes the numbers of particles of two types at time t . Let the quantities $a_1(k, l)$ denote the rates of producing k type 1 particles and l type 2 particles, starting with one type 1 particle, and $a_2(k, l)$ be analogously defined but beginning with one type 2 particle. Given a two-type branching process defined by instantaneous rates $a_i(k, l)$, denote the following *pseudo-generating* functions for $i = 1, 2$ as

$$u_i(s_1, s_2) = \sum_k \sum_l a_i(k, l) s_1^k s_2^l. \tag{D.1}$$

We may expand the probability generating functions in the following form:

$$\begin{aligned} \phi_{10}(t, s_1, s_2) &= E(s_1^{X_1(t)} s_2^{X_2(t)} | X_1(0) = 1, X_2(0) = 0) \\ &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} P_{1,0}^{kl}(t) s_1^k s_2^l \\ &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} (\mathbf{1}_{k=1, l=0} + a_1(k, l)t + o(t)) s_1^k s_2^l \\ &= s_1 + u_1(s_1, s_2)t + o(t). \end{aligned} \tag{D.2}$$

We have an analogous expression for $\phi_{01}(t, s_1, s_2)$ beginning with one particle of type 2 instead of type 1. For short, we will write $\phi_{10} := \phi_1, \phi_{01} := \phi_2$. Thus, we have the following relation between the functions ϕ and u :

$$\begin{aligned} \frac{d\phi_1}{dt}(t, s_1, s_2)|_{t=0} &= u_1(s_1, s_2) \text{ and} \\ \frac{d\phi_2}{dt}(t, s_1, s_2)|_{t=0} &= u_2(s_1, s_2). \end{aligned} \tag{D.3}$$

To derive the backwards and forward equations, Chapman–Kolmogorov arguments yield the symmetric relations

$$\begin{aligned} \phi_1(t + h, s_1, s_2) &= \phi_1(t, \phi_1(h, s_1, s_2), \phi_2(h, s_1, s_2)) \\ &= \phi_1(h, \phi_1(t, s_1, s_2), \phi_2(t, s_1, s_2)). \end{aligned} \tag{D.4}$$

First, we derive the backward equations by expanding around t and applying (D.3):

$$\begin{aligned} \phi_1(t + h, s_1, s_2) &= \phi_1(t, s_1, s_2) + \frac{d\phi_1}{dh}(t + h, s_1, s_2)|_{h=0}h + o(h) \\ &= \phi_1(t, s_1, s_2) + \frac{d\phi_1}{dh}(h, \phi_1(t, s_1, s_2), \phi_2(t, s_1, s_2))|_{h=0}h + o(h) \\ &= \phi_1(t, s_1, s_2) + u_1(\phi_1(t, s_1, s_2), \phi_2(t, s_1, s_2))h + o(h). \end{aligned} \tag{D.5}$$

Since an analogous argument applies for ϕ_2 , we arrive at the system

$$\begin{aligned} \frac{d}{dt}\phi_1(t, s_1, s_2) &= u_1(\phi_1(t, s_1, s_2), \phi_2(t, s_1, s_2)) \text{ and} \\ \frac{d}{dt}\phi_2(t, s_1, s_2) &= u_2(\phi_1(t, s_1, s_2), \phi_2(t, s_1, s_2)), \end{aligned} \tag{D.6}$$

with initial conditions $\phi_1(0, s_1, s_2) = s_1, \phi_2(0, s_1, s_2) = s_2$.

Recall in our SIR approximation, we use the initial population $X_2(0)$ as a constant that scales the instantaneous rates over any time interval $[t_0, t_1)$. The only nonzero rates specifying this proposed model, in the notation above, are

$$a_1(0, 1) = \beta X_2(0), \quad a_1(1, 0) = -\beta X_2(0), \quad a_2(0, 1) = -\alpha, \quad a_2(0, 0) = \alpha. \tag{D.7}$$

For simplicity, call $X_2(0) := I_0$, the constant representing the infected population at the beginning of the time interval. Thus, the corresponding pseudo-generating functions have a simple form:

$$\begin{aligned} u_1(s_1, s_2) &= \beta I_0 s_2 - \beta I_0 s_1 \text{ and} \\ u_2(s_1, s_2) &= \alpha - \alpha s_2 = \alpha(1 - s_2). \end{aligned} \tag{D.8}$$

Plugging into the backward equations, we obtain

$$\begin{aligned} \frac{d}{dt} \phi_1(t, s_1, s_2) &= \beta I_0 (\phi_2(t, s_1, s_2) - \phi_1(t, s_1, s_2)) \text{ and} \\ \frac{d}{dt} \phi_2(t, s_1, s_2) &= \alpha - \alpha \phi_2(t, s_1, s_2). \end{aligned} \tag{D.9}$$

The ϕ_2 differential equation corresponds to a pure death process and is immediately solvable; suppressing the arguments of ϕ_2 for notational convenience, we obtain

$$\begin{aligned} \frac{d}{dt} \phi_2 &= \alpha - \alpha \phi_2 \\ \frac{d}{dt} \phi_2 \left(\frac{1}{1 - \phi_2} \right) &= \alpha \\ \ln(1 - \phi_2) &= -\alpha t + C \\ \phi_2 &= 1 - \exp(-\alpha t + C). \end{aligned} \tag{D.10}$$

Plugging in $\phi_2(0, s_1, s_2) = s_2$, we obtain $C = \ln(1 - s_2)$, and we arrive at

$$\phi_2(t, s_1, s_2) = 1 + (s_2 - 1) \exp(-\alpha t) \tag{D.11}$$

Substituting this solution into the first differential equation and applying the integrating factor method provides

$$\begin{aligned} \phi_1 e^{\beta I_0 t} &= \int \beta I_0 e^{\beta I_0 t} \left(1 + \frac{s_2 - 1}{e^{\alpha t}} \right) dt = e^{\beta I_0 t} + \beta I_0 (s_2 - 1) \int e^{(\beta I_0 - \alpha)t} dt \\ &= e^{\beta I_0 t} + \beta I_0 (s_2 - 1) \frac{e^{(\beta I_0 - \alpha)t}}{\beta I_0 - \alpha} + C. \end{aligned} \tag{D.12}$$

Plugging in the initial condition $\phi_1(0, s_1, s_2) = s_1$ and rearranging yields

$$\phi_1 = 1 + \frac{\beta I_0(s_2 - 1)}{\beta I_0 - \alpha} e^{-\alpha t} + e^{-\beta I_0 t} (s_1 - 1 - \frac{\beta I_0(s_2 - 1)}{\beta I_0 - \alpha}). \tag{D.13}$$

D.2 Transition probability expressions

Transition probabilities are related to the PGF via repeated partial differentiation; note that

$$\begin{aligned} P_{kl}^{mn}(t) &= \frac{1}{k!} \frac{1}{l!} \frac{\partial^k}{\partial s_1^k} \frac{\partial^l}{\partial s_2^l} \phi_{mn}(t, s_1, s_2) \Big|_{s_1=s_2=0} \\ &= \frac{1}{k!} \frac{1}{l!} \frac{\partial^k}{\partial s_1^k} \frac{\partial^l}{\partial s_2^l} \phi_1^m(t, s_1, s_2) \phi_2^n(t, s_1, s_2) \Big|_{s_1=s_2=0} \\ &= \frac{\partial^l}{\partial s_2^l} \sum_{i=0}^k \binom{k}{i} \frac{\partial^{k-i}}{\partial s_1^{k-i}} \phi_1^m(t, s_1, s_2) \frac{\partial^i}{\partial s_1^i} \phi_2^n(t, s_1, s_2) \Big|_{s_1=s_2=0}. \end{aligned} \tag{D.14}$$

This expression is generally unwieldy, but notice $\frac{\partial^i}{\partial s_1^i} \phi_2^n(t, s_1, s_2) \Big|_{s_1=0} = 0$ for all $i > 0$ in our model. Remarkably, this allows us to further simplify and ultimately arrive at closed-form expressions. Continuing, we see

$$\begin{aligned} P_{kl}^{mn}(t) &= \frac{\partial^l}{\partial s_2^l} \left[\binom{k}{0} \phi_2^n(t, s_1, s_2) \frac{\partial^k}{\partial s_1^k} \phi_1^m(t, s_1, s_2) \right] \Big|_{s_1=s_2=0} \\ &= \frac{\partial^l}{\partial s_2^l} \left\{ \phi_2^n(t, s_1, s_2) \cdot \frac{m!}{(m-k)!} e^{-k\beta I_0 t} \left[1 + \frac{\beta I_0(s_2 - 1)}{\beta I_0 - \alpha} e^{-\alpha t} \right. \right. \\ &\quad \left. \left. - e^{-\beta I_0 t} \left(1 + \frac{\beta I_0(s_2 - 1)}{\beta I_0 - \alpha} \right) \right]^{m-k} \right\} \Big|_{s_1=s_2=0} \\ &:= \frac{\partial^l}{\partial s_2^l} [\phi_2^n(t, s_1, s_2) \cdot h(t, s_1, s_2)] \Big|_{s_1=s_2=0} \\ &= \sum_{i=0}^l \binom{l}{i} \frac{\partial^{l-i}}{\partial s_2^{l-i}} h(t, s_1, s_2) \frac{\partial^i}{\partial s_2^i} \phi_2^n(t, s_1, s_2) \\ &:= \sum_{i=0}^l \binom{l}{i} A(l-i) B(i). \end{aligned} \tag{D.15}$$

From here, it is straightforward to take partial derivatives of $h(t, s_1, s_2)$ and our closed-form expression of $\phi_2^n(t, s_1, s_2)$ to arrive at Conditions (41) and (42). A heatmap

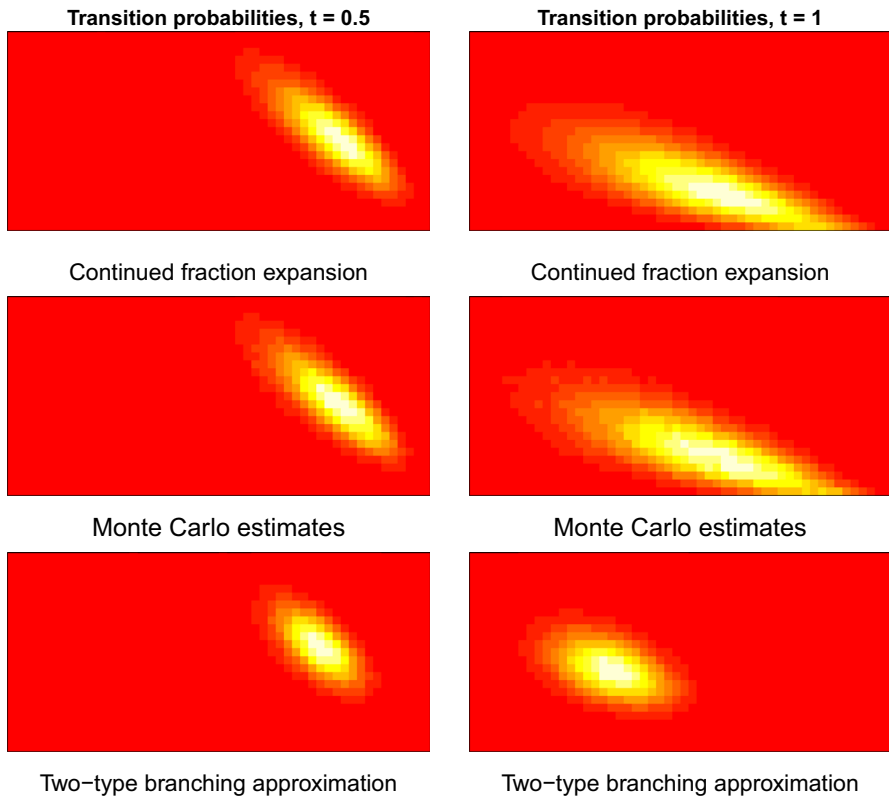


Fig. 8 Heatmap visualizations of transition probabilities near the region of support across methods for $t = 0.5, 1$. We see that the branching approximation is noticeably different from the Monte Carlo ground truth when we increase t to 1, while the continued fraction approach remains accurate

visualization of the difference between transition probabilities under the branching approximation and those computed using the continued fraction method for the SIR model is included below (Fig. 8).

References

- Abate J, Whitt W (1992) The Fourier-series method for inverting transforms of probability distributions. *Queueing Syst* 10(1–2):5–87
- Andersson H, Britton T (2000) *Stochastic epidemic models and their statistical analysis*, vol 4. Springer, New York
- Andrieu C, Doucet A, Holenstein R (2010) Particle Markov chain Monte Carlo methods. *J R Stat Soc Ser B (Stat Methodol)* 72(3):269–342
- Blum MG, Tran VC (2010) HIV with contact tracing: a case study in approximate Bayesian computation. *Biostatistics* 11(4):644–660
- Brauer F (2008) *Compartmental models in epidemiology*. Mathematical epidemiology. Springer, Berlin, pp 19–79
- Britton T (2010) Stochastic epidemic models: a survey. *Math Biosci* 225(1):24–35

- Cauchemez S, Ferguson N (2008) Likelihood-based estimation of continuous-time epidemic models from time-series data: application to measles transmission in London. *J R Soc Interface* 5(25):885–897
- Correia-Gomes C, Economou T, Bailey T, Brazdil P, Alban L, Niza-Ribeiro J (2014) Transmission parameters estimated for salmonella typhimurium in swine using susceptible-infectious-resistant models and a bayesian approach. *BMC Vet Res* 10(1):101
- Craviotto C, Jones WB, Thron W (1993) A survey of truncation error analysis for Padé and continued fraction approximants. *Acta Appl Math* 33(2–3):211–272
- Crawford FW, Suchard MA (2012) Transition probabilities for general birth-death processes with applications in ecology, genetics, and evolution. *J Math Biol* 65(3):553–580
- Crawford FW, Minin VN, Suchard MA (2014) Estimation for general birth-death processes. *J Am Stat Assoc* 109(506):730–747
- Crawford FW, Weiss RE, Suchard MA (2015) Sex, lies, and self-reported counts: Bayesian mixture models for longitudinal heaped count data via birth-death processes. *Ann Appl Stat* 9:572–596
- Crawford FW, Stutz TC, Lange K (2016) Coupling bounds for approximating birth-death processes by truncation. *Stat Prob Lett* 109:30–38
- Csilléry K, Blum MG, Gaggiotti OE, François O (2010) Approximate Bayesian computation (ABC) in practice. *Trends Ecol Evol* 25(7):410–418
- Doss CR, Suchard MA, Holmes I, Kato-Maeda M, Minin VN (2013) Fitting birth-death processes to panel data with applications to bacterial DNA fingerprinting. *Ann Appl Stat* 7(4):2315–2335
- Drovandi CC, Pettitt AN (2011) Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. *Biometrics* 67(1):225–233
- Dukic V, Lopes HF, Polson NG (2012) Tracking epidemics with Google flu trends data and a state-space SEIR model. *J Am Stat Assoc* 107(500):1410–1426
- Earn DJ (2008) A light introduction to modelling recurrent epidemics. *Mathematical epidemiology*. Springer, Berlin, pp 3–17
- Ephraim Y, Mark BL (2012) Bivariate Markov processes and their estimation. *Found Trends Signal Process* 6(1):1–95
- Feller W (1968) An introduction to probability theory and its applications, vol 1. Wiley, Hoboken
- Finkenstädt B, Grenfell B (2000) Time series modelling of childhood diseases: a dynamical systems approach. *J R Stat Soc Ser C (Appl Stat)* 49(2):187–205
- Golightly A, Wilkinson DJ (2005) Bayesian inference for stochastic kinetic models using a diffusion approximation. *Biometrics* 61(3):781–788
- 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
- Hitchcock S (1986) Extinction probabilities in predator-prey models. *J Appl Probab* 23(1):1–13
- Iglehart DL (1964) Multivariate competition processes. *Ann Math Stat* 35(1):350–361
- Ionides E, Bretó C, King A (2006) Inference for nonlinear dynamical systems. *Proc Natl Acad Sci USA* 103(49):18,438–18,443
- Ionides EL, Nguyen D, Atchadé Y, Stoev S, King AA (2015) Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. *Proce Natl Acad Sci USA* 112(3):719–724
- Jahnke T, Huisinga W (2007) Solving the chemical master equation for monomolecular reaction systems analytically. *J Math Biol* 54(1):1–26
- Karev GP, Berezhovskaya FS, Koonin EV (2005) Modeling genome evolution with a diffusion approximation of a birth-and-death process. *Bioinformatics* 21(Suppl 3):iii12–iii19
- Keeling M, Ross J (2008) On methods for studying stochastic disease dynamics. *J R Soc Interface* 5(19):171–181
- Kermack W, McKendrick A (1927) A contribution to the mathematical theory of epidemics. *Proc R Soc Lond Ser A* 115(772):700–721
- Lentz WJ (1976) Generating Bessel functions in Mie scattering calculations using continued fractions. *Appl Opt* 15(3):668–671
- Levin D (1973) Development of non-linear transformations for improving convergence of sequences. *Int J Comput Math* 3(1–4):371–388
- Martin AD, Quinn KM, Park JH (2011) MCMCpack: Markov chain Monte Carlo in R. *J Stat Softw* 42(9):22. <http://www.jstatsoft.org/v42/i09/>
- McKendrick A (1926) Applications of mathematics to medical problems. *Proce Edinb Math Soc* 44:98–130
- McKinley T, Cook AR, Deardon R (2009) Inference in epidemic models without likelihoods. *Int J Biostat* 5(1):1557–4679

- Moler C, Loan C (2003) Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Rev* 45:3–49
- Murphy J, O'Donohoe M (1975) Some properties of continued fractions with applications in Markov processes. *IMA J Appl Math* 16(1):57–71
- Novozhilov AS, Karev GP, Koonin EV (2006) Biological applications of the theory of birth-and-death processes. *Brief Bioinform* 7(1):70–85
- Owen J, Wilkinson DJ, Gillespie CS (2015) Scalable inference for Markov processes with intractable likelihoods. *Stat Comput* 25(1):145–156
- Rabier CE, Ta T, Ané C (2014) Detecting and locating whole genome duplications on a phylogeny: a probabilistic approach. *Mol Biol Evol* 31(3):750–762
- Raggett G (1982) A stochastic model of the Eyam plague. *J Appl Stat* 9(2):212–225
- Renshaw E (2011) Stochastic population processes: analysis, approximations, simulations. Oxford University Press, Oxford
- Reuter GEH (1957) Denumerable Markov processes and the associated contraction semigroups on l_1 . *Acta Math* 97(1):1–46
- Reuter GEH (1961) Competition processes. In: Proc. 4th Berkeley Symp. Math. Statist. Prob, vol 2, pp 421–430
- Riley S, Donnelly CA, Ferguson NM (2003) Robust parameter estimation techniques for stochastic within-host macroparasite models. *J Theor Biol* 225(4):419–430
- Robert CP, Cornuet JM, Marin JM, Pillai NS (2011) Lack of confidence in approximate Bayesian computation model choice. *Proc Natl Acad Sci* 108(37):15,112–15,117
- Rosenberg NA, Tsolaki AG, Tanaka MM (2003) Estimating change rates of genetic markers using serial samples: applications to the transposon IS6110 in *Mycobacterium tuberculosis*. *Theor Popul Biol* 63(4):347–363
- Schranz HW, Yap VB, Easteal S, Knight R, Huttley GA (2008) Pathological rate matrices: from primates to pathogens. *BMC Bioinform* 9(1):550
- Sidje RB (1998) Expokit: a software package for computing matrix exponentials. *ACM Trans Math Softw (TOMS)* 24(1):130–156
- Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, Dessimoz C (2013) Approximate Bayesian computation. *PLoS Comput Biol* 9(1):e1002,803
- Thompson I, Barnett A (1986) Coulomb and Bessel functions of complex arguments and order. *J Comput Phys* 64(2):490–509
- van den Eshof J, Hochbruck M (2006) Preconditioning lanczos approximations to the matrix exponential. *SIAM J Sci Comput* 27(4):1438–1457
- Xu J, Guttorp P, Kato-Maeda M, Minin VN (2015) Likelihood-based inference for discretely observed birth-death-shift processes, with applications to evolution of mobile genetic elements. *Biometrics* 71(4):1009–1021