BIT

A three-term recurrence relation for accurate evaluation of transition probabilities of the simple birth-and-death process

Alberto Pessia[1] 🟢 · **Jing Tang[1]** 🟢

Abstract

The simple (linear) birth-and-death process is a widely used stochastic model for describing the dynamics of a population. When the process is observed discretely over time, despite the large amount of literature on the subject, little is known about formal estimator properties. Here we will show that its application to observed data is further complicated by the fact that numerical evaluation of the well-known transition probability is an ill-conditioned problem. To overcome this difficulty we will rewrite the transition probability in terms of a Gaussian hypergeometric function and subsequently obtain a three-term recurrence relation for its accurate evaluation. We will also study the properties of the hypergeometric function as a solution to the three-term recurrence relation. We will then provide formulas for the gradient and Hessian of the log-likelihood function and conclude the article by applying our methods for numerically computing maximum likelihood estimates in both simulated and real dataset.

Keywords Three-term recurrence relation · Hypergeometric function · Birth-death process · Transition probabilities · Maximum likelihood estimation

Mathematics Subject Classification Primary 65Q30; Secondary 60J80 · 33C05 · 62F10 · 62M05

✉ Alberto Pessia
academic@albertopessia.com

[1] Research Program in Systems Oncology (ONCOSYS), Faculty of Medicine, University of Helsinki, P.O.Box 63, 00014 Helsinki, Finland

🖄 Springer

1 Introduction

A birth-and-death process (BDP) is a stochastic model that is commonly employed for describing changes over time of the size of a population. Its first mathematical formulation is due to Feller [11], followed by important mathematical contributions of Arley and Borchsenius [1] and Kendall [18,19]. According to the basic assumptions of the model, when the population size at time t is j, the probability of a single birth occurring during an infinitesimal time interval $(t, t + dt)$ is equal to $\lambda_j dt + o(dt)$ while the probability of a single death is $\mu_j dt + o(dt)$, where $\lambda_j \geq 0$ and $\mu_j \geq 0$ for $j \geq 0$. If $p_j(t)$ is the probability of observing j individuals at time t then

$$p_j(t + dt) = \lambda_{j-1} dt p_{j-1}(t) + \mu_{j+1} dt p_{j+1}(t) + (1 - (\lambda_j + \mu_j) dt) p_j(t) + o(dt)$$

If we subtract $p_j(t)$ from both sides of the equation, divide by dt, and then take the limit of dt to zero, we obtain the well known BDP differential equation

$$p'_j(t) = \lambda_{j-1} p_{j-1}(t) + \mu_{j+1} p_{j+1}(t) - (\lambda_j + \mu_j) p_j(t) \tag{1.1}$$

By assuming that at time zero the size of the population was $i \geq 0$, that is $p_i(0) = 1$, we have the initial condition required to solve the differential equation (1.1).

In this article we will focus on the simple (linear) BDP without migration [19] defined by $\lambda_j = j\lambda$ and $\mu_j = j\mu$. With this particular choice of parameters a starting size of zero implies $\lambda_0 = \mu_0 = 0$, i.e. it remains zero for all $t \geq 0$. The rate of growth does not increase faster than the population size and therefore $\sum_{j=0}^{\infty} p_j(t) = 1$ [12, Chapter 17, Section 4]. When $i > 0$ the population becomes extinct if it reaches the size $j = 0$ at time $t > 0$. Obviously i, j, and t are not allowed to be negative, nor the basic birth and death rates.

What makes this model particularly attractive for real applications is the fact that its transition probability is available in closed form [2, Chapter 8] and we could, in principle, easily evaluate it. By defining

$$\phi(t, \lambda, \mu) = \frac{e^{(\lambda-\mu)t} - 1}{\lambda e^{(\lambda-\mu)t} - \mu}, \quad \alpha(t, \lambda, \mu) = \mu\phi(t, \lambda, \mu), \quad \beta(t, \lambda, \mu) = \lambda\phi(t, \lambda, \mu)$$

$$\gamma(t, \lambda, \mu) = 1 - \alpha(t, \lambda, \mu) - \beta(t, \lambda, \mu) = 1 - (\lambda + \mu)\phi(t, \lambda, \mu)$$

and assuming that at time 0 the size of the population was $i > 0$, the probability $p_j(t)$ of observing j units at time t is

$$
\begin{cases}
\displaystyle\sum_{h=0}^{m}\binom{i}{h}\binom{i+j-h-1}{i-1}\alpha(t,\lambda,\mu)^{i-h}\beta(t,\lambda,\mu)^{j-h}\gamma(t,\lambda,\mu)^{h}, & \mu\neq\lambda \\[4pt]
& \text{(1.2)} \\[4pt]
\displaystyle\sum_{h=0}^{m}\binom{i}{h}\binom{i+j-h-1}{i-1}\left(\frac{\lambda t}{1+\lambda t}\right)^{i+j-2h}\left(\frac{1-\lambda t}{1+\lambda t}\right)^{h}, & \mu=\lambda \quad \text{(1.3)} \\[10pt]
\displaystyle\binom{j-1}{i-1}e^{-j\lambda t}(e^{\lambda t}-1)^{j-i}, & \mu=0,\ j\geq i \qquad \text{(1.4)} \\[10pt]
\displaystyle\binom{i}{j}e^{-i\mu t}(e^{\mu t}-1)^{i-j}, & \lambda=0,\ j\leq i \qquad \text{(1.5)} \\[10pt]
1, & t=0,\ j=i \qquad \text{(1.6)} \\[6pt]
1, & \lambda=\mu=0,\ j=i \qquad \text{(1.7)} \\[6pt]
0, & \text{otherwise} \qquad \text{(1.8)}
\end{cases}
$$

where t, λ, and μ are to be considered strictly positive if not otherwise specified and $m=\min(i,j)$. The probability of the population being extinct at time t is

$$
p_0(t)=\begin{cases}
\displaystyle\left(\frac{\mu e^{(\lambda-\mu)t}-\mu}{\lambda e^{(\lambda-\mu)t}-\mu}\right)^{i}, & \mu\neq\lambda & \text{(1.9)} \\[12pt]
\displaystyle\left(\frac{\lambda t}{1+\lambda t}\right)^{i}, & \mu=\lambda & \text{(1.10)} \\[12pt]
(1-e^{-\mu t})^{i}, & \lambda=0 & \text{(1.11)} \\[6pt]
0, & \mu=0\ \text{or}\ t=0 & \text{(1.12)}
\end{cases}
$$

In the majority of applications direct evaluation of Eqs. (1.2)–(1.12) is sufficient. However, for particular values of process parameters, Eqs. (1.2) and (1.3) are numerically unstable (Fig. 1) and alternative methods are needed [28].

A possible approach might be the algorithm introduced by Crawford and Suchard [5] based on the continued fraction representation of Murphy and O'Donohoe [22], or the saddlepoint approximation method of Davison et al. [8]. For this particular case where we know the exact closed form we will show that a simpler and more efficient method is available.

The remainder of the paper is organized as follows. In Sect. 2 we introduce the problem and find the parameter sets for which it is ill-conditioned. In Sect. 3 we rewrite the transition probability in terms of a Gaussian hypergeometric function and find in Sect. 4 a three-term recurrence relation (TTRR) for its computation. In Sect. 5 we extend the results to the log-likelihood function, its gradient, and its Hessian matrix. In Sect. 6 we apply our method to both simulated and real data. In Sect. 7 we conclude the article with a brief discussion.
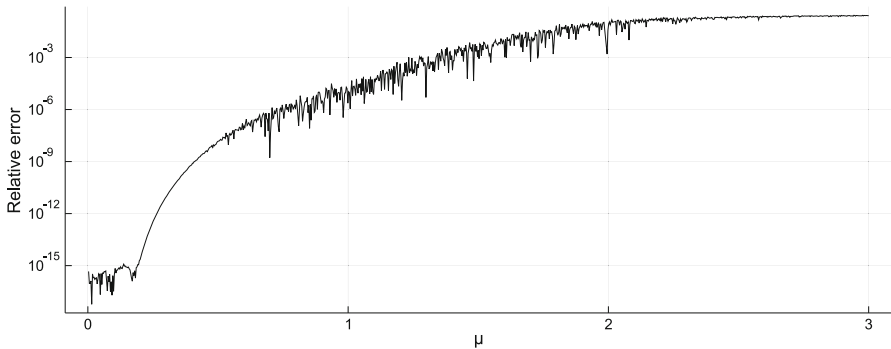
Fig. 1 Numerical relative error of the log-probability as evaluated by direct application of Eqs. (1.2) and (1.3). For this particular example we set $i = 25$, $j = 35$, $t = 2$, $\lambda = 1$ and evaluated the log-probability as a function of μ. We computed correct values with Maple™ 2018.2 [21] using 100 significant digits. Relative error is defined as $|1 - \log \hat{p}_j(t) / \log p_j(t)|$ where $\hat{p}_j(t)$ is the numerically evaluated transition probability

2 Numerical stability

We will always assume that all basic mathematical operations (arithmetic, logarithmic, exponentiation, etc.) are computed with a relative error bounded by a value ϵ that is close to zero and small enough for any practical application [25]. Following this assumption and after taking into consideration floating-point arithmetic [15], Eqs. (1.4)–(1.12) can be considered numerically stable and won't be discussed further. We will instead focus our attention on the series (1.2) and (1.3) assuming all variables to be strictly positive, including j.

Suppose to be interested in the value $s_m = \sum_{h=0}^{m} u_h$ and use a naïve recursive summation algorithm for its computation, that is

$$s_0 = u_0, \quad s_n = s_{n-1} + u_n, \quad n = 1, \ldots, m$$

The relative condition number of this algorithm is [27]

$$\rho_m = \rho_m^A + \rho_m^R = \frac{\sum_{h=0}^{m} |u_h|}{\left| \sum_{h=0}^{m} u_h \right|} + \frac{\sum_{n=1}^{m} |s_n|}{\left| \sum_{h=0}^{m} u_h \right|}$$

where ρ_m^A is associated with perturbations in the the value of the addends while ρ_m^R is associated with rounding errors in the arithmetic operations. Note that when $u_h \geq 0$ for all h, $\rho_m^A = 1$ and the condition number depends only on rounding errors. With a compensated summation algorithm [16] we might significantly reduce the numerical error and evaluate accurately the sum. However, when the addends are alternating in sign, the condition number can be of large magnitude and the problem is numerically unstable even when compensating for rounding errors. In our case it is likely that the magnitude of the binomial coefficients make ρ_m^A a ratio between a very large number and a probability that is instead close to zero. We will now find the conditions under which the sums (1.2) and (1.3) are alternating in sign.

Proposition 2.1 *For all $\lambda \geq 0$ and $\mu \geq 0$ the function*

$$\phi(t, \lambda, \mu) = \frac{e^{(\lambda-\mu)t} - 1}{\lambda e^{(\lambda-\mu)t} - \mu}$$

is zero if and only if $t = 0$. It is always positive otherwise.

Proof If $\mu \neq \lambda$ and $t = 0$ the numerator $e^{(\lambda-\mu)t} - 1$ is equal to zero but the denominator is not. When $\mu = \lambda$ the function becomes

$$\lim_{\mu \to \lambda} \frac{e^{(\lambda-\mu)t} - 1}{\lambda e^{(\lambda-\mu)t} - \mu} = \frac{t}{1 + \lambda t}$$

For all $\lambda \geq 0$, it is equal to zero if $t = 0$ and positive otherwise. Assume now $t > 0$. When $\lambda > \mu$ we have $e^{(\lambda-\mu)t} > 1$ and $\mu/\lambda < 1$ implying that the numerator and the denominator are always positive. When $\lambda < \mu$ the numerator $e^{(\lambda-\mu)t} - 1$ is negative. The denominator is also negative when $e^{(\lambda-\mu)t} < \mu/\lambda$. Since $\lambda < \mu$ the left hand side is less than one while the right hand side is greater than one, proving the proposition. $\square$

Corollary 2.1 *Functions $\alpha(t, \lambda, \mu) = \mu\phi(t, \lambda, \mu)$ and $\beta(t, \lambda, \mu) = \lambda\phi(t, \lambda, \mu)$ are non-negative for all $\lambda \geq 0$, $\mu \geq 0$, and $t \geq 0$.*

Proof This is a direct consequence of Proposition (2.1) and the assumptions $\lambda \geq 0$ and $\mu \geq 0$. $\square$

Proposition 2.2 *Assume $t > 0$, $\lambda > 0$, $\mu > 0$. If $\mu \neq \lambda$, define $\xi = \log(\lambda/\mu)/(\lambda-\mu)$. If $\mu = \lambda$, define instead $\xi = 1/\lambda$. Let $\gamma(t, \lambda, \mu) = 1 - (\lambda + \mu)\phi(t, \lambda, \mu)$. Then*

$$(t, \lambda, \mu) \begin{cases} < 0 & \text{if } t > \xi \\ > 0 & \text{if } t < \xi \\ = 0 & \text{if } t = \xi \end{cases}$$

Proof Rewrite the function $\gamma(t, \lambda, \mu)$ in its expanded form:

$$\gamma(t, \lambda, \mu) = \frac{\lambda - \mu e^{(\lambda-\mu)t}}{\lambda e^{(\lambda-\mu)t} - \mu}$$

Assume $\mu \neq \lambda$. If $t > 0$, we already proved in Proposition (2.1) that the denominator is always positive when $\lambda > \mu$ while it is always negative when $\lambda < \mu$. The numerator is positive when $(\lambda - \mu)t < \log(\lambda/\mu)$, that is when $t < \xi$ and $\lambda > \mu$ or $t > \xi$ and $\lambda < \mu$. It is zero when $t = \xi$. From these results follow the first set of inequalities. When $\mu = \lambda$ the function becomes

$$\lim_{\mu \to \lambda} \frac{\lambda - \mu e^{(\lambda-\mu)t}}{\lambda e^{(\lambda-\mu)t} - \mu} = \frac{1 - \lambda t}{1 + \lambda t}$$

If $t > 0$ and $\lambda > 0$ the denominator is always positive. The numerator is zero if $t = \lambda^{-1}$, it is positive when $t < \lambda^{-1}$, and it is negative otherwise. $\square$

Corollary 2.2 *When $t = \log(\lambda/\mu)/(\lambda - \mu)$ equation (1.2) becomes*

$$p_j(t) = \binom{i + j - 1}{i - 1} \left(\frac{\mu}{\lambda + \mu}\right)^i \left(\frac{\lambda}{\lambda + \mu}\right)^j$$

When $t = \lambda^{-1}$ Eq. (1.3) becomes

$$p_j(t) = \binom{i + j - 1}{i - 1} \left(\frac{1}{2}\right)^{i+j}$$

Proof This is a direct consequence of the fact that $\gamma(t, \lambda, \mu) = 0$ and that 0^h is zero for all $h > 0$ and one for $h = 0$. □

From Corollary 2.2 we have simple closed form solutions when $\gamma(t, \lambda, \mu)$ is zero and therefore we will not consider this case further. We know from Proposition 2.2 the conditions under which Eqs. (1.2) and (1.3) are alternating in sign and might become numerically unstable. Looking at the example shown in Fig. 1, where $t = 2$ and $\lambda = 1$, function $\gamma(t, \lambda, \mu)$ is non-negative when $0 < \mu \lesssim 0.2032$. We clearly see from Fig. 1 that the error steadily increases starting at the value $\mu \approx 0.2032$. Computer algebra systems implementing symbolic computation are obviously a way of evaluating the transition probability since we can, in principle, increase the number of significant digits to an appropriate amount. The computational cost of symbolic computation, however, can be prohibitive for real applications where the probability is usually evaluated multiple times with a large number of addends in the summation. In the next section we will instead find an alternative representation to Eqs. (1.2) and (1.3) that will lead to an algorithm for their accurate and efficient evaluation.

3 Hypergeometric representation

Define

$$\omega(i, j, t, \lambda, \mu) = \binom{i + j - 1}{i - 1} \mu^i \lambda^j \left(\frac{e^{(\lambda-\mu)t} - 1}{\lambda e^{(\lambda-\mu)t} - \mu}\right)^{i+j}$$

$$z(t, \lambda, \mu) = \frac{\gamma(t, \lambda, \mu)}{\alpha(t, \lambda, \mu)\beta(t, \lambda, \mu)} = \frac{(\lambda - \mu e^{(\lambda-\mu)t})(\lambda e^{(\lambda-\mu)t} - \mu)}{\lambda\mu(e^{(\lambda-\mu)t} - 1)^2}$$

Note that $\omega(i, j, t, \lambda, \mu)$ is simply the first term in (1.2). When $\mu = \lambda$ set

$$\omega(i, j, t, \lambda, \lambda) = \lim_{\mu \to \lambda} \omega(i, j, t, \lambda, \mu) = \binom{i + j - 1}{i - 1} \left(\frac{\lambda t}{1 + \lambda t}\right)^{i+j}$$

$$z(t, \lambda, \lambda) = \lim_{\mu \to \lambda} z(t, \lambda, \mu) = \left(\frac{1}{\lambda t}\right)^2 - 1$$

Multiply and divide each term in the series (1.2) by $\omega(i, j, t, \lambda, \mu)$ to get

$$
\begin{aligned}
p_j(t) &= \omega(i, j, t, \lambda, \mu) \sum_{h=0}^{m} \frac{\binom{i}{h}\binom{j}{h}}{\binom{i+j-1}{h}} z(t, \lambda, \mu)^h \\
&= \omega(i, j, t, \lambda, \mu) \sum_{h=0}^{m} \frac{i!}{(i-h)!} \frac{j!}{(j-h)!} \frac{(i+j-1-h)!}{(i+j-1)!} \frac{z(t, \lambda, \mu)^h}{h!} \\
&= \omega(i, j, t, \lambda, \mu) \sum_{h=0}^{m} \frac{(-i)_h(-j)_h}{(-(i+j-1))_h} \frac{(-z(t, \lambda, \mu))^h}{h!} \\
&= \omega(i, j, t, \lambda, \mu) \, {}_2F_1\left[\begin{matrix} -i, \ -j \\ -(i+j-1) \end{matrix} ; -z(t, \ \lambda, \ \mu)\right]
\end{aligned}
\tag{3.1}
$$

where $(q)_h$ is the rising Pochhammer symbol and $_2F_1(a, b; c; y)$ is the Gaussian hypergeometric function [26, Chapter 1]. To evaluate (3.1) is then sufficient to separately compute the functions $\omega(i, j, t, \lambda, \mu)$ and $_2F_1(-i, -j; -(i + j - 1); -z(t, \lambda, \mu))$.

Partial derivatives of $\log p_j(t)$ are required for computing partial derivatives of the log-likelihood function as we will explicitly show in Sect. 5. The following theorem proves that partial derivatives of $_2F_1(-i, -j; -(i + j - 1); -z(t, \lambda, \mu))$ depend on hypergeometric functions of similar nature.

Theorem 3.1 *Denote with u_x the first-order partial derivative of $u(x, y)$ with respect to x. Similarly, denote with u_{xy} the second-order partial derivative with respect first to x and subsequently y. Then*

$$
\frac{\partial}{\partial x} {}_2F_1\left[\begin{matrix} -i, \ -j \\ -(i+j-1) \end{matrix} ; -u(x, \ y)\right] = u_x \frac{ij}{i+j-1} {}_2F_1\left[\begin{matrix} -a_1, \ -b_1 \\ -(a_1+b_1) \end{matrix} ; -u(x, \ y)\right]
$$

$$
\frac{\partial^2}{\partial x \partial y} {}_2F_1\left[\begin{matrix} -i, \ -j \\ -(i+j-1) \end{matrix} ; -u(x, \ y)\right] = u_{xy} \frac{ij}{i+j-1} {}_2F_1\left[\begin{matrix} -a_1, \ -b_1 \\ -(a_1+b_1) \end{matrix} ; -u(x, \ y)\right]
$$

$$
+ u_x u_y \frac{i(i-1)j(j-1)}{(i+j-1)(i+j-2)} {}_2F_1\left[\begin{matrix} -a_2, \ -b_2 \\ -(a_2+b_2+1) \end{matrix} ; -u(x, \ y)\right]
$$

where $a_1 = i - 1$, $a_2 = i - 2$, $b_1 = j - 1$, and $b_2 = j - 2$.

Proof

$$\frac{\partial}{\partial x}\left(1 + \sum_{h=1}^{m}\frac{i!}{(i-h)!}\frac{j!}{(j-h)!}\frac{(i+j-1-h)!}{(i+j-1)!}\frac{u(x,y)^h}{h!}\right)$$

$$= u_x \sum_{h=1}^{m}\frac{i!}{(i-h)!}\frac{j!}{(j-h)!}\frac{(i+j-1-h)!}{(i+j-1)!}\frac{u(x,y)^{h-1}}{(h-1)!}$$

$$= u_x \frac{ij}{i+j-1}\sum_{h=0}^{m-1}\frac{(i-1)!}{(i-1-h)!}\frac{(j-1)!}{(j-1-h)!}\frac{(i+j-2-h)!}{(i+j-2)!}\frac{u(x,y)^h}{h!}$$

$$= u_x \frac{ij}{i+j-1}{}_2F_1\left[\begin{matrix}-(i-1),\ -(j-1)\\-(i+j-2)\end{matrix}; -u(x,\ y)\right]$$

where $m = \min(i, j)$. Apply the same procedure to obtain the second-order partial derivatives. □

From Theorem 3.1 we see that, in general, we must be able to accurately evaluate the hypergeometric function

$$_2F_1\left[\begin{matrix}-a,\ -b\\-(a+b-k)\end{matrix}; -z\right] = \sum_{h=0}^{\min(a,b)}\frac{\binom{a}{h}\binom{b}{h}}{\binom{a+b-k}{h}}z^h \qquad (3.2)$$

for $a, b \in \mathbb{N}_+, k = 1, 0, -1, -2, \ldots,$ and $z \in \mathbb{R}$.

4 Hypergeometric function $_2F_1(-a, -b; -(a+b-k); -z)$

The following theorem can be considered the main result of this article.

Theorem 4.1 *The hypergeometric function $_2F_1(-a, -b; -(a+b-k); -z)$, as a function of b, is a solution of the three-term recurrence relation (TTRR)*

$$(a+b+1-k)(a+b-k)y_{b+1} - (a+b-k)(a+b+1-k+(a-b)z)y_b +$$
$$-b(b-k)zy_{b-1} = 0 \qquad (3.1)$$

Proof The recursion can be obtained by the method of creative telescoping [24,29]. To prove that it holds, define

$$L_{b,h} = \frac{a!}{(a-h)!}\frac{b!}{(b-h)!}\frac{(a+b-k-h)!}{(a+b-k)!}\frac{z^h}{h!}$$

and let

$$t_h = (a+b+1-k)(a+b-k)L_{b+1,h}$$
$$- (a+b-k)(a+b+1-k+(a-b)z)L_{b,h} + -b(b-k)zL_{b-1,h}$$

Note that $\sum_h t_h$ is the left hand side of equation (3.1) because $y_b = \sum_h L_{b,h}$. Set

$$R_{b,h} = -\frac{(a-k)(a+b-k-h)bh}{(b+1-h)(b-h)}$$

and let

$$u_h = R_{b,h+1}L_{b-1,h+1} - R_{b,h}L_{b-1,h}$$

Sum the previous expression with respect to h to obtain $\sum_h u_h = -R_{b,0}L_{b-1,0} = 0$. We now need to prove that $t_h = u_h$ for all h. Start by dividing t_h by $L_{b-1,h}$ to obtain

$$\frac{(a+b+1-k-h)(a+b-k-h)(b+1)b}{(b+1-h)(b-h)} +$$
$$-\frac{b(a+b-k-h)(a+b+1-k+(a-b)z)}{b-h} - b(b-k)z$$

By expanding the polynomial and collecting the terms with respect to h we get

$$\frac{t_h}{L_{b-1,h}} = -\frac{(a-k)b}{(b+1-h)(b-h)}\left[(1+z)h^2 - (a+b-k+(a+b+1)z)h + a(b+1)z\right]$$

Doing the same with the right hand side we get

$$\frac{u_h}{L_{b-1,h}} = -\frac{(a-k)b}{(b+1-h)(b-h)}\left[(b+1-h)(a-h)z - h(a+b-k-h)\right]$$
$$= -\frac{(a-k)b}{(b+1-h)(b-h)}\left[(1+z)h^2 - (a+b-k+(a+b+1)z)h + a(b+1)z\right]$$

proving the equality. $\qquad\square$

If we divide both sides of equation (3.1) by the coefficient of y_{b+1}, and shift the index by 1, we obtain the *forward* recursion

$$y_b = \left(1 + \frac{(a+1-b)z}{a+b-k}\right)y_{b-1} + \frac{(b-1)(b-1-k)z}{(a+b-k)(a+b-1-k)}y_{b-2} \qquad (3.2)$$

Starting from

$$y_0 = {}_2F_1(-a, 0; -(a-k); -z) = 1$$
$$y_1 = {}_2F_1(-a, -1; -(a+1-k); -z) = 1 + \frac{az}{a+1-k}$$

we can, in principle, obtain all remaining solutions all the way up to the required term. Note that $a \geq 1$ and $k \leq 1$, therefore the denominator $a+1-k$ is strictly positive

and always well defined. Knowing the values of y_{b+2} and y_{b+1}, for large b, we can travel the recursion also in a *backward* manner:

$$y_b = \frac{(a+b+2-k)(a+b+1-k)}{(b+1)(b+1-k)z}\left(y_{b+2} - \left(1 + \frac{(a-b-1)z}{a+b+2-k}\right)y_{b+1}\right)$$
(3.3)

Theorem 4.1 proves that our hypergeometric function is a solution of a TTRR. However, equation (3.1) can also admit a second linearly independent solution.

Definition 4.1 A solution f_b of a TTRR is said to be a *minimal* solution if there exists a linearly independent solution g_b such that

$$\lim_{b\to\infty} \frac{f_b}{g_b} = 0$$

The solution g_b is called a *dominant* solution.

It is well known that, regardless of the starting values, forward evaluation of a TTRR converges to the dominant solution while backward evaluation converges instead to the minimal solution [14, Chapter 4]. We now need to find the conditions under which our hypergeometric function is either the dominant or the minimal solution.

Lemma 4.1 (Poincaré–Perron) *Let $y_{b+1} + v_b y_b + u_b y_{b-1} = 0$ and suppose that v_b and u_b are different from zero for all $b > 0$. Suppose also that $\lim_{b\to\infty} v_b = v$ and $\lim_{b\to\infty} u_b = u$. Denote with ϕ_1 and ϕ_2 the (not necessarily distinct) roots of the characteristic equation $\phi^2 + v\phi + u = 0$. If f_b and g_b are the linearly independent non-trivial solutions of the difference equation, then*

$$\limsup_{b\to\infty} \sqrt[b]{|f_b|} = |\phi_1|, \quad \limsup_{b\to\infty} \sqrt[b]{|g_b|} = |\phi_2|$$

If $|\phi_1| < |\phi_2|$ it is also

$$\lim_{b\to\infty} \frac{f_{b+1}}{f_b} = \phi_1, \quad \lim_{b\to\infty} \frac{g_{b+1}}{g_b} = \phi_2$$

and f_b is the minimal solution while g_b is the dominant solution. If $|\phi_1| = |\phi_2|$ the lemma is inconclusive about the existence of a minimal solution.

Proof See Chapter 8 of Elaydi [10]. □

Using Lemma 4.1 we can study the nature of our hypergeometric function as a solution of the TTRR.

Theorem 4.2 $_2F_1(-a, -b; -(a+b-k); -z)$ *is a dominant solution of Eq. (3.1) when $|z| < 1$. It is a minimal solution when $|z| > 1$. The nature of the solution is unknown when $|z| = 1$.*

Proof Our TTRR is

$$y_{b+1} - \left(1 + \frac{(a-b)z}{a+b+1-k}\right) y_b - \frac{b(b-k)z}{(a+b+1-k)(a+b-k)} y_{b-1} = 0$$

Take the limit of the coefficients

$$\lim_{b\to\infty} -\left(1 + \frac{a-b}{a+b+1-k}z\right) = -(1-z), \quad \lim_{b\to\infty} -\frac{b(b-k)z}{(a+b+1-k)(a+b-k)} = -z$$

The characteristic equation is $\phi^2 - (1-z)\phi - z = 0$ with solutions $\phi_1 = -z$ and $\phi_2 = 1$. When $|z| < 1$ the solution associated with ϕ_1 is minimal and the one associated with ϕ_2 is dominant. The opposite is true when $|z| > 1$. We will now prove that $_2F_1(-a, -b; -(a+b-k); -z)$ is associated with the characteristic root $\phi_2 = 1$. The summation index h in Eq. (3.2) depends on b since the upper bound of the series is the minimum between a and b. Note, however, that variable a is considered known and fixed to a finite value. When b goes to infinity the summation index h in (3.2) does not depend on b any more and the series is always finite, so that we can safely exchange the limit of the sum with the sum of the limits:

$$\lim_{b\to\infty} \sum_{h=0}^{\min(a,b)} \frac{(-a)_h(-b)_h}{(-(a+b-k))_h} \frac{(-z)^h}{h!} = \sum_{h=0}^{a}(-a)_h \frac{(-z)^h}{h!} \lim_{b\to\infty} \frac{(-b)_h}{(-(a+b-k))_h}$$

Using Stirling's approximation $n! \sim \sqrt{2\pi n}(n/e)^n$ for large n, we obtain

$$\lim_{b\to\infty} \frac{(-b)_h}{(-(a+b-k))_h} = \lim_{b\to\infty} \frac{b^{b+1/2}(a+b-k-h)^{a+b-k-h+1/2}}{(b-h)^{b-h+1/2}(a+b-k)^{a+b-k+1/2}}$$

$$= \lim_{b\to\infty} \left(\frac{b}{b-h}\right)^{b+1/2} \left(\frac{a+b-k-h}{a+b-k}\right)^{b-k+1/2} = e^h e^{-h} = 1$$

from which follows that

$$\lim_{b\to\infty} {}_2F_1\left[\begin{matrix} -a, \ -b \\ -(a+b-k) \end{matrix}; -z\right] = \sum_{h=0}^{a}(-a)_h \frac{(-z)^h}{h!} = \sum_{h=0}^{a}\binom{a}{h}(-z)^h = (1-z)^a$$

and

$$\lim_{b\to\infty} \frac{{}_2F_1\left[\begin{matrix} -a, \ -(b+1) \\ -(a+b+1-k) \end{matrix}; -z\right]}{{}_2F_1\left[\begin{matrix} -a, \ -b \\ -(a+b-k) \end{matrix}; -z\right]} = \frac{(1-z)^a}{(1-z)^a} = 1$$

The solution is therefore dominant when $|z| < 1$ and minimal for $|z| > 1$. When $|z| = 1$ Lemma 4.1 is inconclusive about the nature of the solution. $\square$

Since Lemma 4.1 refers to asymptotic results, Theorem 4.2 is always valid for large values of b. For small values of b, instead, there is a possibility of anomalous behaviour as described by Deaño and Segura [9]. By Definition 4.1 we would expect that the sequence of ratios between a minimal and a dominant solution would be monotonically decreasing to zero for all b. There are cases, however, in which this is not necessarily true. A minimal solution might behave as a dominant solution up to a finite value b^* and then switch to its asymptotic minimal nature only starting at $b^* + 1$.

Definition 4.2 Let f_b and g_b be respectively the minimal and dominant solution of a TTRR as $b \to \infty$. f_b is said to be *pseudo-dominant* for all $b \le b^*$ if the sequence $\{R_b = |f_b/g_b|\}$ is increasing for $b \le b^*$ but decreasing for $b > b^*$.

Lemma 4.2 (Deaño–Segura) *Let $y_{b+1} + v_b y_b + u_b y_{b-1} = 0$ be a recurrence such that, for $b \ge b^-$, $u_b < 0$ and v_b changes sign at $b^* > b^- + 1$. Suppose that there exists a solution f_b with fixed pattern of signs for all $b \ge b^-$, the pattern being alternating if $v_b < 0$ for large b or with constant sign if $v_b > 0$ for large b (f_b may be minimal). Let g_b be any solution (not minimal) such that*

$$\frac{g_{b^*+1}}{g_{b^*}} = -\psi \frac{f_{b^*+1}}{f_{b^*}}, \qquad \psi > 0,$$

and let $R_b = |f_b/g_b|$, then for $b \ge b^-$ the following holds depending on the value ψ:

(1) *if $\psi > 1$, then $R_b < R_{b^*}$ if $b \ne b^*$.*
(2) *if $\psi < 1$, then $R_b < R_{b^*+1}$ if $b \ne b^* + 1$.*
(3) *if $\psi = 1$, then $R_b < R_{b^*} = R_{b^*+1}$ if $b \ne b^*, b^* + 1$.*

Proof See Deaño and Segura [9].   □

According to Lemma 4.2, if u_b is negative and v_b changes sign at index b^*, then our asymptotic minimal solution behaves as a dominant solution up to $b^* - 1$ or b^*. We must then study the sign of the two coefficients
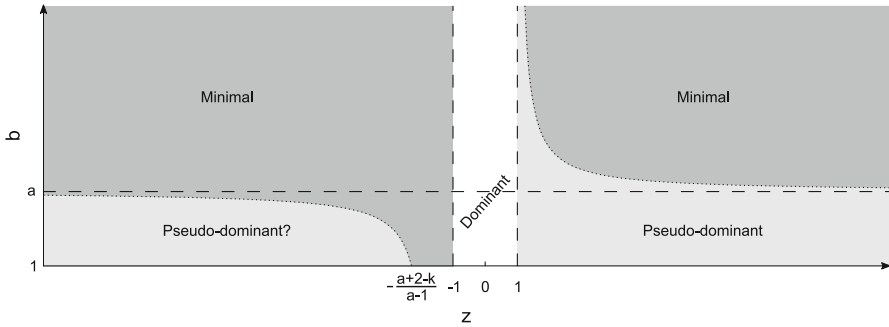
$$u_b = -\frac{b(b-k)z}{(a+b+1-k)(a+b-k)}$$
$$v_b = -\frac{(a+b+1-k) + (a-b)z}{a+b+1-k}$$

with $b \ge 1$, $a \ge 1$ and $k \le 1$. Since the denominators are strictly positive, we can simply study the signs of the associated quantities

$$u_b' = -b(b-k)z$$
$$v_b' = -(a+b+1-k) - (a-b)z$$

u_b' is negative when $z > 0$, positive when $z < 0$, and zero when $b = k = 1$. Define $b^* = (z-1)^{-1}((z+1)a + 1 - k)$. v_b' is negative when $z < 1$ and $b > b^*$ or when $z > 1$ and $b < b^*$. It is obviously positive in the complementary set. The point b^* is

Fig. 2 Nature of the hypergeometric function $_2F_1(-a, -b; -(a+b-k); -z)$ as a solution of TTRR (3.1). The minimum value admissible for b is one. When $b = 0$ we can simply use the numerically stable Eqs. (1.9)–(1.12). The dotted curve is given by equation $b^* = (z-1)^{-1}((z+1)a + 1 - k)$. It is dotted to represent the fact that we don't know if the solution becomes minimal at b^* or $b^* + 1$. The curve has an horizontal asymptote at $b = a$

the delimiter at which the coefficient v_b switches from positive sign to negative sign or vice versa.

When $z > 1$ we are under the conditions of Lemma 4.2, therefore the solution is surely minimal for $b > b^* + 1$. It is pseudo-dominant for all $b < b^*$. Not knowing the shape of the linearly independent solution g_b, we don't know if the solution becomes minimal at b^* or $b^* + 1$. Interestingly, when $z < -(a + 2 - k)/(a - 1)$, we have the opposite behaviour of a positive u_b and v_b changing sign from positive to negative at the same index b^*. The regions are highlighted in Fig. 2.

Lemma 4.2 does not consider the case of a positive u_b but we conjecture that it might be applied to this case as well. Nevertheless, as shown by the following proposition, we can simply ignore the problem altogether.

Proposition 4.1 *For all finite $\lambda > 0$, $\mu > 0$, and $t > 0$, function $z(t, \lambda, \mu)$ is always greater than -1. It is positive when $\mu \neq \lambda$ and $t < \log(\lambda/\mu)/(\lambda - \mu)$ or when $\mu = \lambda$ and $t < \lambda^{-1}$.*

Proof Rewrite the function $z(t, \lambda, \mu)$ as

$$z(t, \lambda, \mu) = \frac{\gamma(t, \lambda, \mu)}{\alpha(t, \lambda, \mu)\beta(t, \lambda, \mu)} = \frac{(\frac{\lambda}{\mu} + \frac{\mu}{\lambda})e^{(\lambda-\mu)t} - e^{2(\lambda-\mu)t} - 1}{e^{2(\lambda-\mu)t} - 2e^{(\lambda-\mu)t} + 1}$$

It is straightforward to show that the function converges to -1 when λ, μ, or t go to infinity. The limit is never attained for finite λ, μ, or t. When any of the parameters approaches zero, instead, the function approaches positive infinity. We know from Corollary 2.1 that the denominator $\alpha(t, \lambda, \mu)\beta(t, \lambda, \mu)$ is always positive. The sign of the function $z(t, \lambda, \mu)$ is therefore equal to the sign of $\gamma(t, \lambda, \mu)$, which is given in Proposition 2.2. Same results apply when $\mu = \lambda$. $\square$

Note that for $|z| > 1$, as clearly shown in Fig. 2, we have to use either the forward or backward recursion depending on the value of b that we wish to evaluate. We can

simplify our computations and compute the hypergeometric function for all values of $z \in \mathbb{R}$ by applying the well known symmetric property $_2F_1(-a, -b; -(a + b - k); -z) = \, _2F_1(-b, -a; -(b + a - k); -z)$. If $z > 1$ and $b > a$, swap the two variables to transform a minimal solution into a pseudo-dominant one. If $z < -1$ and $b < a$, swap the two variables to transform a (possibly) pseudo-dominant solution into a minimal one. When $|z| = 1$ the two solutions are neither dominant nor minimal and recursion in both directions is not unstable [13].

All previous results are summarized in Algorithm 1 in Supplementary Material S1. Assuming a constant time for arithmetic operations the time complexity is simply $O(m)$, where $m = \min(a, b)$, that is the total number of iterations required. Note that we only use basic arithmetic operations, saving computational time when compared to the more expensive functions found in Eqs. (1.2)–(1.3), such as the Binomial function. Using the TTRR approach is better, from a computational point of view, also when the problem is well-behaved.

5 Likelihood function

Let $\boldsymbol{t} = (t_0, \ldots, t_S)^T$ be the vector of observation times with $t_S \leq t$, $\boldsymbol{n} = (n_0, \ldots, n_S)^T$ be the corresponding observed population sizes, and $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_S)^T = (t_1 - t_0, \ldots, t_S - t_{S-1})^T$ be the vector of inter-arrival times. When the process is observed continuously the log-likelihood function is [7, Equation (25)]

$$\log \mathscr{L}(\lambda, \mu | \boldsymbol{t}, \boldsymbol{n}) = B_t \log \lambda + D_t \log \mu - (\lambda + \mu)X_t + \sum_{s=0}^{S-1} \log n_s \qquad (5.1)$$

where B_t and D_t are respectively the total number of births and deaths recorded during the time interval $[0, t]$ while $X_t = \sum_{s=0}^{S} n_s \tau_{s+1}$ is the total time lived in the population during $[0, t]$. By convention we set $\tau_{S+1} = t - t_S$. From (5.1) we obtain the maximum likelihood estimators (MLEs) of λ and μ as

$$\hat{\lambda} = \frac{B_t}{X_t}, \qquad \hat{\mu} = \frac{D_t}{X_t} \qquad (5.2)$$

from which follows that the MLE of the growth rate $\theta = \lambda - \mu$ is $\hat{\theta} = \hat{\lambda} - \hat{\mu} = (B_t - D_t)/X_t$. A more challenging situation is encountered when the BDP is observed discretely at fixed time points. Rewrite the probability of transitioning from i to j in t time with birth rate λ and death rate μ as $p(j|i, t, \lambda, \mu)$. Since the BDP is a continuous time Markov chain [19] we can write the likelihood function as

$$\mathscr{L}(\lambda, \mu | \boldsymbol{t}, \boldsymbol{n}) = \prod_{s=1}^{S} p(n_s | n_{s-1}, \tau_s, \lambda, \mu)$$

Note that the joint likelihood of M observations of stochastically independent processes, having the same birth and death rates, is simply the product of the M likelihoods associated with each process. To the best of our knowledge, no known closed form solutions for $\hat{\lambda}$ and $\hat{\mu}$ are currently available for the discrete case. However, in the case of equidistant sampling where $\tau_s = \tau$ for all s, we know that [17]

$$\hat{\theta} = \frac{1}{\tau} \log \left(\frac{n_1 + \cdots + n_S}{n_0 + \cdots + n_{S-1}} \right) \tag{5.3}$$

It is easy to show that the first moment of $\hat{\theta}$ does not exist. Starting with $S = 1$ we have

$$\mathbb{E}[\hat{\theta}] = \frac{1}{\tau} \sum_{j=0}^{\infty} \log \left(\frac{j}{n_0} \right) p(j|n_0, t, \lambda, \mu)$$

The first term in the summation is not defined because the probability of extinction is strictly positive, unless the process is a pure birth process (see Eqs. (1.9)–(1.12)). For a simple birth-and-death process without migration the population stays extinct once its size reaches a value of zero, therefore the previous result can be extended to any value $S > 1$. To estimate $\hat{\lambda}$, $\hat{\mu}$, and $\hat{\theta}$ we must consider only observations in which the population is not immediately extinct at time point $s = 1$.

To find the maximum likelihood estimators we will use a numerical approach, that is the Newton–Raphson method [4, Chapter 4] applied to the log-likelihood function. To proceed we need its gradient and Hessian matrix, that are

$$\nabla l(\lambda, \mu|t, n) = \nabla \log \mathscr{L}(\lambda, \mu|t, n) = \sum_{s=1}^{S} \begin{pmatrix} \frac{\partial}{\partial \lambda} \log p(n_s|n_{s-1}, \tau_s, \lambda, \mu) \\ \frac{\partial}{\partial \mu} \log p(n_s|n_{s-1}, \tau_s, \lambda, \mu) \end{pmatrix} \tag{5.4}$$

$$H(\lambda, \mu|t, n) = \sum_{s=1}^{S} \begin{pmatrix} \frac{\partial^2}{\partial \lambda^2} \log p(n_s|n_{s-1}, \tau_s, \lambda, \mu) & \frac{\partial^2}{\partial \lambda \partial \mu} \log p(n_s|n_{s-1}, \tau_s, \lambda, \mu) \\ \frac{\partial^2}{\partial \mu \partial \lambda} \log p(n_s|n_{s-1}, \tau_s, \lambda, \mu) & \frac{\partial^2}{\partial \mu^2} \log p(n_s|n_{s-1}, \tau_s, \lambda, \mu) \end{pmatrix}$$
$$\tag{5.5}$$

with closed form solutions of partial derivatives of log-probabilities appearing in (5.4) and (5.5) given in Supplementary Material S2. They can be evaluated with our proposed TTRR approach. Note that (5.4) and (5.5) are sums of piecewise functions with subdomains inherited from Eqs. (1.2)–(1.8).
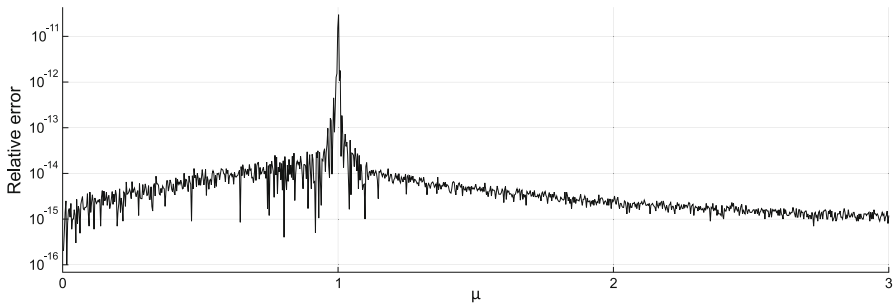
Fig. 3 Numerical relative error of the log-probability evaluated using the hypergeometric representation and the TTRR approach. Parameters are the same as in Fig. 1, that is $i = 25$, $j = 35$, $t = 2$, and $\lambda = 1$. Relative error is always less than 10^{-10}

6 Applications

We developed a free Julia [3] package called "SimpleBirthDeathProcess" to implement all results presented so far.[1]

Returning to the example shown in Fig. 1, we can see from Fig. 3 that our method improves significantly the accuracy of the computations. Interestingly, although not entirely unexpected, the algorithm has a higher numerical error in the neighbourhood of the special point $\mu = \lambda$, that is the removable singularity of Eq. (1.2). Note that relative errors for this particular example are always less than 10^{-10} and small enough for any practical application. In Fig. 4 we can see a more general example where points near the line $\mu = \lambda$ are again associated with higher relative errors. Also in this case they are very small and always less than 10^{-13}.

6.1 Simulated data

We will now study some properties of the maximum likelihood estimator of the birth rate λ, death rate μ, and growth rate $\theta = \lambda - \mu$. We will use our software package to perform simulations and apply standard Monte Carlo integration to approximate the bias and root mean square error (RMSE) of MLEs. The total number of simulations is set to 10^5 in each of the following synthetic experiments.

Constant growth rate

The first study mimics a situation in which both rate parameters are strictly positive. For simplicity we fix the total observation time to $t = 10$ and assume the process to be observed at S equidistant time points, that is every $\tau = t/S$ amount of time. To reduce the amount of possible combinations to test we choose birth and death rates so that the expected population size and standard deviation at time t is approximately proportional to the initial population size. In what follows we will always condition our estimators only to populations that are not immediately extinct, as explained in

---

[1] Download available at https://github.com/albertopessia/SimpleBirthDeathProcess.jl.
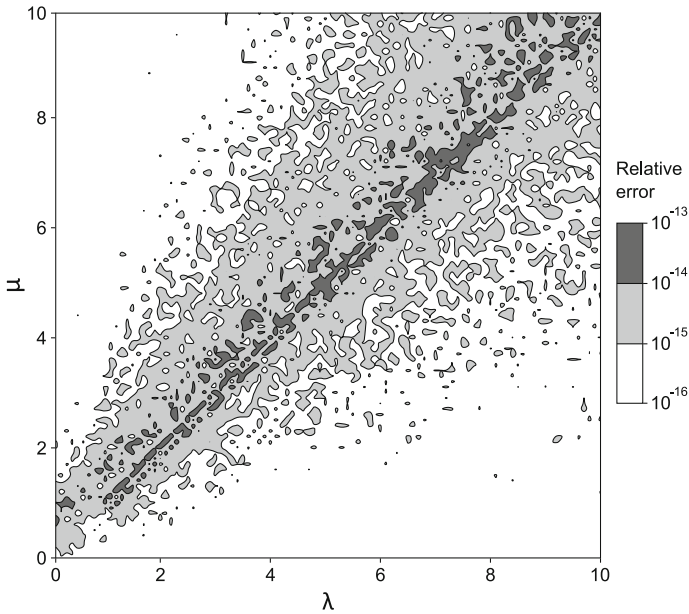
Fig. 4 Numerical relative error of the log-probability evaluated using the hypergeometric representation and the TTRR approach. Parameters for this example are $i = 200$, $j = 100$, and $t = 1$. Relative error is always less than 10^{-13}

Sect. 5. Results of the simulations when $\lambda > \mu$ are shown in Table 1 while results of the simulations when $\lambda < \mu$ are shown in Table 2.

Estimators are generally negatively biased but we also observe situations when $\lambda < \mu$ in which the bias is positive. The magnitude of the bias of $\hat{\lambda}$ and $\hat{\mu}$ is very large when only one time point is used, for which we have $|\text{Bias}(\hat{\lambda})| \approx \text{RMSE}(\hat{\lambda})$ and $|\text{Bias}(\hat{\mu})| \approx \text{RMSE}(\hat{\mu})$. Increasing the number of time points S help reducing both the bias and RMSE of $\hat{\lambda}$ and $\hat{\mu}$. All estimators obviously perform worse when the standard deviation of the stochastic process is high. What is surprising to us is the observation that $\hat{\theta}$ has approximately the same performance regardless of the initial sample size. Increasing the number of time points has also the counter-intuitive effect of making the estimation worse.

Technical replicates

Following the results from the previous section we want to investigate the performance of the estimators when the stochastic process is observed more than once. As an example, this is a standard setting in dose-response drug screening experiments where cell counts are observed after a period of incubation and (usually) 3 to 5 technical replicates are produced under the same experimental conditions [23]. For a fair comparison we will use the same simulation parameters from the previous simulation experiment with the only difference of now having three technical replicates instead

Table 1 Monte Carlo estimates from 10^5 simulations of a simple BDP where $\lambda > \mu$. Growth rate $\theta = \lambda - \mu = 0.0693$ for each row, i.e. the expected population size at time $t = 10$ is set to be two times the initial population size n_0. For each number of time points S, the three rows correspond respectively to a standard deviation of 1.25, 1.5, and 2.0 times the initial population size n_0

n_0	S	λ			μ			θ	
		Truth	Bias	RMSE	Truth	Bias	RMSE	Bias	RMSE
10	1	0.305	−0.244	0.249	0.236	−0.222	0.226	−0.022	0.078
		0.425	−0.362	0.366	0.355	−0.335	0.339	−0.027	0.092
		0.728	−0.659	0.662	0.658	−0.634	0.636	−0.025	0.105
	8	0.305	−0.045	0.141	0.236	−0.019	0.138	−0.026	0.086
		0.425	−0.069	0.205	0.355	−0.028	0.203	−0.041	0.123
		0.728	−0.132	0.391	0.658	−0.042	0.390	−0.090	0.226
100	1	2.742	−2.681	2.681	2.673	−2.658	2.658	−0.023	0.083
		3.934	−3.872	3.872	3.864	−3.841	3.842	−0.030	0.101
		6.966	−6.898	6.898	6.897	−6.867	6.867	−0.031	0.118
	8	2.742	−0.382	1.292	2.673	−0.357	1.285	−0.026	0.086
		3.934	−0.565	1.853	3.864	−0.524	1.841	−0.041	0.121
		6.966	−1.018	3.430	6.897	−0.928	3.413	−0.089	0.226
1000	1	27.111	−27.050	27.050	27.041	−27.026	27.026	−0.024	0.085
		39.024	−38.962	38.962	38.955	−38.932	38.932	−0.031	0.102
		69.349	−69.281	69.281	69.280	−69.249	69.249	−0.032	0.121
	8	27.111	−3.700	12.914	27.041	−3.675	12.906	−0.025	0.085
		39.024	−5.522	18.489	38.955	−5.481	18.478	−0.041	0.122
		69.349	−9.961	33.635	69.280	−9.871	33.616	−0.090	0.226

of one. Results of the simulations when $\lambda > \mu$ are shown in Table 3 while results of the simulations when $\lambda < \mu$ are shown in Table 4.

As expected, we see a decrease in both bias magnitude and RMSE for $\hat{\lambda}$ and $\hat{\mu}$. A small improvement is obtained also for $\hat{\theta}$. Again, increasing the number of time points allow for a better estimation of λ and μ but make the estimation of θ worse. When increasing the number of time points S, the loss (gain) of performance is lower (higher) that in the single observation case of the previous section.

Real data

As an example application we will use real data from a cancer drug combination experiment originally performed and analysed by Liu et al. [20]. Briefly, two monoclonal antibodies were combined together at a concentration ratio of 1:1 to form a mixture. Tested concentrations of the mixture were 0 (no antibody), 0.025, 0.25, 2.5, and 10 µ g/ml. Living cell counts were subsequently measured with a fluorescence microscopy at 1, 2, and 3 days. For each time point they performed six technical replicates for concentrations greater than zero and twelve replicates for the control dose of zero. Since the initial number of cells was not available, they estimated it from the data

Table 2 Monte Carlo estimates from 10^5 simulations of a simple BDP where $\lambda < \mu$. Growth rate $\theta = \lambda - \mu = -0.0693$ for each row, i.e. the expected population size at time $t = 10$ is set to be half the initial population size n_0. For each number of time points S, the three rows correspond respectively to a standard deviation of 0.25, 0.5, and 1.0 times the initial population size n_0

n_0	S	λ			μ			θ	
		Truth	Bias	RMSE	Truth	Bias	RMSE	Bias	RMSE
10	1	0.052	−0.052	0.052	0.121	−0.039	0.067	−0.012	0.057
		0.312	−0.306	0.306	0.381	−0.295	0.304	−0.011	0.085
		1.352	−1.318	1.319	1.421	−1.370	1.372	0.051	0.116
	8	0.052	−0.007	0.046	0.121	0.005	0.062	−0.012	0.057
		0.312	−0.053	0.208	0.381	0.000	0.225	−0.053	0.152
		1.352	−0.279	0.970	1.421	−0.035	0.905	−0.245	0.476
100	1	0.832	−0.831	0.831	0.901	−0.816	0.819	−0.015	0.061
		3.431	−3.425	3.425	3.500	−3.396	3.398	−0.029	0.112
		13.828	−13.796	13.797	13.898	−13.835	13.835	0.038	0.128
	8	0.832	−0.116	0.419	0.901	−0.105	0.416	−0.011	0.054
		3.431	−0.461	1.702	3.500	−0.409	1.695	−0.052	0.146
		13.828	−1.821	8.481	13.898	−1.556	8.365	−0.265	0.537
1000	1	8.630	−8.629	8.629	8.699	−8.615	8.615	−0.015	0.061
		34.623	−34.617	34.617	34.692	−34.585	34.586	−0.032	0.119
		138.595	−138.563	138.563	138.664	−138.600	138.600	0.037	0.132
	8	8.630	−1.107	4.163	8.699	−1.096	4.160	−0.011	0.054
		34.623	−4.462	16.635	34.692	−4.411	16.627	−0.052	0.145
		138.595	−16.466	83.001	138.664	−16.198	82.886	−0.267	0.545

Table 3 Monte Carlo estimates from 10^5 simulations of three simple BDPs where $\lambda > \mu$. Growth rate $\theta = \lambda - \mu = 0.0693$ for each row, i.e. the expected population size at time $t = 10$ is set to be two times the initial population size n_0. For each number of time points S, the three rows correspond respectively to a standard deviation of 1.25, 1.5, and 2.0 times the initial population size n_0

| n_0	k	λ			μ			θ	
		Truth	Bias	RMSE	Truth	Bias	RMSE	Bias	RMSE
10	1	0.305	−0.107	0.170	0.236	−0.102	0.164	−0.006	0.038
		0.425	−0.185	0.242	0.355	−0.179	0.234	−0.005	0.045
		0.728	−0.427	0.469	0.658	−0.430	0.466	0.003	0.054
	8	0.305	−0.014	0.082	0.236	−0.006	0.080	−0.008	0.039
		0.425	−0.020	0.117	0.355	−0.009	0.116	−0.011	0.049
		0.728	−0.033	0.209	0.658	−0.011	0.208	−0.022	0.075
100	1	2.742	−1.064	1.712	2.673	−1.057	1.707	−0.006	0.039
		3.934	−1.793	2.390	3.864	−1.787	2.382	−0.006	0.046
		6.966	−4.201	4.610	6.897	−4.202	4.606	0.001	0.054
	8	2.742	−0.126	0.765	2.673	−0.119	0.763	−0.007	0.039
		3.934	−0.176	1.099	3.864	−0.165	1.097	−0.011	0.049
		6.966	−0.270	1.986	6.897	−0.248	1.984	−0.021	0.075
1000	1	27.111	−10.559	17.127	27.041	−10.553	17.122	−0.006	0.039
		39.024	−17.871	23.869	38.955	−17.865	23.860	−0.006	0.046
		69.349	−41.816	45.913	69.280	−41.818	45.909	0.001	0.055
	8	27.111	−1.275	7.621	27.041	−1.268	7.620	−0.007	0.039
		39.024	−1.673	10.999	38.955	−1.662	10.997	−0.011	0.049
		69.349	−2.647	19.663	69.280	−2.625	19.661	−0.021	0.075

to be on average approximately equal to 23. Following previous studies [6] we will fix for each and every observation an initial cell count of 23 as if it were known in advance. The complete dataset is visually represented in Fig. 5.

It is important to note that the dataset is made of 108 independent observations, i.e. counts referring to the same concentration at different time points are not part of the same time series but are, instead, independent realizations of the same stochastic process observed at different times. In our notation, $S = 1$ and $\tau = t/S = t$ for each of the 108 measurements. The basic datum x_i, $i = 1, \ldots, 108$, is a vector $(c_i, t_i, n_i(0), n_i(t_i))^T$ where c_i is the tested antibody concentration, t_i is the time in days, $n_i(0)$ is the initial population size set to 23 for each and every observation, and $n_i(t_i)$ is the final cancer cell counts for observation i. For further details about the study and the experimental design we refer to the original article of Liu et al. [20]. To model the data we use a similar approach to that of Crawford et al. [6], that is a linear model on the logarithm scale of the basic process rates. Formally we define

$$\begin{cases} \log(\lambda_i) = \alpha_\lambda + \beta_\lambda \log(1 + c_i) \\ \log(\mu_i) = \alpha_\mu + \beta_\mu \log(1 + c_i) \end{cases}, \quad \text{for all } i = 1, \ldots, 108 \qquad (6.1)$$

Table 4 Monte Carlo estimates from 10^5 simulations of three simple BDPs where $\lambda < \mu$. Growth rate $\theta = \lambda - \mu = -0.0693$ for each row, i.e. the expected population size at time $t = 10$ is set to be half the initial population size n_0. For each number of time points S, the three rows correspond respectively to a standard deviation of 0.25, 0.5, and 1.0 times the initial population size n_0

n_0	k	λ			μ			θ	
		Truth	Bias	RMSE	Truth	Bias	RMSE	Bias	RMSE
10	1	0.052	−0.024	0.050	0.121	−0.021	0.052	−0.003	0.030
		0.312	−0.215	0.237	0.381	−0.224	0.240	0.009	0.051
		1.352	−1.198	1.209	1.421	−1.269	1.277	0.071	0.106
	8	0.052	−0.002	0.027	0.121	0.001	0.033	−0.004	0.030
		0.312	−0.016	0.111	0.381	−0.001	0.113	−0.015	0.066
		1.352	−0.090	0.468	1.421	−0.022	0.454	−0.068	0.178
100	1	0.832	−0.305	0.588	0.901	−0.301	0.584	−0.004	0.030
		3.431	−2.115	2.339	3.500	−2.117	2.333	0.002	0.055
		13.828	−12.336	12.447	13.898	−12.399	12.507	0.063	0.108
	8	0.832	−0.038	0.246	0.901	−0.035	0.246	−0.004	0.029
		3.431	−0.136	0.986	3.500	−0.121	0.985	−0.015	0.065
		13.828	−0.498	4.210	13.898	−0.427	4.196	−0.071	0.185
1000	1	8.630	−3.055	5.873	8.699	−3.050	5.870	−0.004	0.030
		34.623	−20.741	22.973	34.692	−20.742	22.967	0.001	0.054
		138.595	−115.124	116.248	138.664	−115.200	116.320	0.076	0.076
	8	8.630	−0.371	2.462	8.699	−0.367	2.461	−0.004	0.029
		34.623	−1.310	9.754	34.692	−1.295	9.753	−0.016	0.066
		138.595	−4.562	41.566	138.664	−4.490	41.555	−0.071	0.185
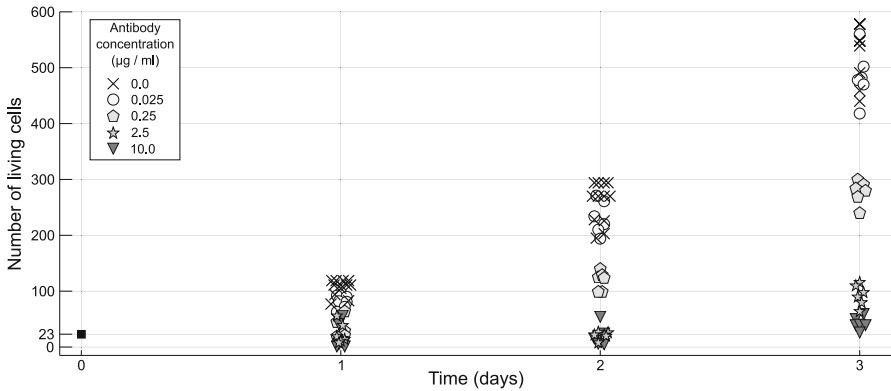
Fig. 5 Antibody dataset by Liu et al. [20]. All the observed counts are assumed to be originated from the same number of cells $n_0 = 23$. Increasing the antibody concentration reduces the growth rate of cancer cells

Table 5 Maximum likelihood estimates of model (6.1) based on the antibody dataset

Dose (μ g/ml)	λ		μ		$\theta = \lambda - \mu$	
	Estimate	SE	Estimate	SE	Estimate	SE
0	4.0344	0.4844	2.9572	0.4835	1.0772	0.0292
0.025	4.0238	0.4806	2.9595	0.4806	1.0644	0.0285
0.25	3.9397	0.4548	2.9774	0.4595	0.9623	0.0283
2.5	3.5304	0.4476	3.0721	0.4366	0.4583	0.0535
10	3.1249	0.5778	3.1810	0.5962	−0.0561	0.0740

Maximum likelihood estimates and their corresponding standard errors are shown in Table 5. We obtained estimates by numerically maximizing the log-likelihood function with the BFGS algorithm [4, Chapter 4]. We applied the delta method to the observed Fisher information matrix in order to compute the standard error of all parameters.

According to our model, increasing the antibody concentration has the double effect of reducing the birth rate and raising the death rate while maintaining the overall rate $\lambda + \mu$ approximately the same. When the dose of the treatment increases the global growth rate θ decreases as a consequence, reaching a negative value at the maximum tested concentration. Interestingly, Crawford et al. obtained values that are slightly different from ours but still very close. In particular, the maximum absolute difference between our estimates of θ and theirs is just 0.054. Since their R package *birth.death* is not available for download any more we could not replicate the analysis and investigate the discrepancies more. We believe, however, that the observed differences are simply due to numerical errors or to a chosen solution that is a local optimum rather than a global.

7 Concluding remarks

Maximum likelihood estimators for the basic rates of a simple (linear) birth-and-process are available in closed form only when the process is observed continuously over time. Numerical methods are currently the only option to draw inferences for discretely observed processes. However, we showed that direct application of the well-known transition probability might be subject to large numerical error. We rewrote the probability in terms of a Gaussian hypergeometric function and found a three-term recurrence relation for its evaluation. Not only our approach led to very accurate approximations but also to a computational efficient algorithm when compared to the naïve direct summation method.

By means of simulation we observed that MLEs $\hat{\lambda}$ and $\hat{\mu}$ are largely negatively biased. We confirmed the intuition that to obtain better estimates it is important to employ a large initial population size, multiple time points, and multiple technical replicates. The actual values, as one would expect, depend on the magnitude of the basic rates, i.e. the process standard deviation. If only the growth parameter $\theta = \lambda - \mu$ is of interest then multiple technical replicates with (surprisingly) only one time point provide the best results. Interestingly, the initial population size seems not to affect the bias nor the root mean square error of $\hat{\theta}$.

We also released a free Julia package called "SimpleBirthDeathProcess". With the help of our tool it is possible to simulate, fit, or just evaluate the likelihood function of a simple BDP. Accurate evaluation of the log-likelihood function will create opportunities for future research, such as implementation of MCMC algorithms for Bayesian inference. As a final note, it might be worth investigating our conjecture that Lemma 4.2 can be extended to TTRRs with a positive coefficient.

References

1. Arley, N., Borchsenius, V.: On the theory of infinite systems of differential equations and their application to the theory of stochastic processes and the perturbation theory of quantum mechanics. Acta Mathematica **76**(3), 261–322 (1944)
2. Bailey, N.T.J.: The Elements of Stochastic Processes with Applications to the Natural Sciences. Wiley, New York (1964). ISBN: 0-471-04165-3
3. Bezanson, J., Edelman, A., Karpinski, S., Shah, V.B.: Julia: A fresh approach to numerical computing. SIAM Rev. **59**(1), 65–98 (2017)
4. Bonnans, J.-F., Gilbert, J.C., Lemarechal, C., Sagastizábal, C.A.: Numerical Optimization: Theoretical and Practical Aspects, 2nd edn. Universitext. Springer, Berlin (2006). ISBN 978-3-540-35445-1
5. Crawford, F.W., Suchard, M.A.: Transition probabilities for general birth-death processes with applications in ecology, genetics, and evolution. J. Math. Biol. **65**(3), 553–580 (2012)
6. Crawford, F.W., Minin, V.N., Suchard, M.A.: Estimation for general birth–death processes. J. Am. Stat. Assoc. **109**(506), 730–747 (2014)
7. Darwin, J.H.: The behaviour of an estimator for a simple birth and death process. Biometrika **43**(1/2), 23–31 (1956)
8. Davison, A.C., Hautphenne, S., Kraus, A.: Parameter estimation for discretely-observed linear birth-and-death processes. arXiv, page arXiv:1802.05015 (2018)
9. Deaño, A., Segura, J.: Transitory minimal solutions of hypergeometric recursions and pseudoconvergence of associated continued fractions. Math. Comput. **76**(258), 879–901 (2007)
10. Elaydi, S.: An Introduction to Difference Equations. Undergraduate Texts in Mathematics, 3rd edn. Springer, New York (2005). ISBN 978-0-387-23059-7
11. Feller, W.: Die grundlagen der Volterraschen theorie des kampfes ums dasein in wahrscheinlichkeitstheoretischer behandlung. Acta Biotheoretica 11–40 (1939)
12. Feller, W.: An Introduction to Probability Theory and Its Applications, Vol. I. Wiley Series in Probability and Mathematical Statistics, 3rd edn. Wiley, New York (1968). ISBN: 978-0-471-25708-0
13. Gil, A., Segura, J., Temme, N.M.: The ABC of hyper recursions. J. Comput. Appl. Math. **190**(1), 270–286 (2006)
14. Gil, A., Segura, J., Temme, N.M.: Numerical Methods for Special Functions. Society for Industrial and Applied Mathematics, Philadelphia (2007). ISBN: 978-0-89871-634-4
15. Goldberg, D.: What every computer scientist should know about floating-point arithmetic. ACM Comput. Surv. **23**(1), 5–48 (1991)
16. Higham, N.J.: Accuracy and Stability of Numerical Algorithms, volume 80 of Other Titles in Applied Mathematics, 2nd edn. Society for Industrial and Applied Mathematics, Philadelphia (2002). ISBN 978-0-89871-802-7
17. Keiding, N.: Maximum likelihood estimation in the birth-and-death process. Ann. Stat. **3**(2), 363–372 (1975)
18. Kendall, D.G.: On the generalized birth-and-death process. Ann. Math. Stat. **19**(1), 1–15 (1948)
19. Kendall, D.G.: Stochastic processes and population growth. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **11**(2), 230–282 (1949)
20. Liu, H., Beckett, L.A., DeNardo, G.L.: On the analysis of count data of birth-and-death process type: with application to molecularly targeted cancer therapy. Stat. Med. **26**(5), 1114–1135 (2007)
21. Monagan, M.B., Geddes, K.O., Heal, K.M., Labahn, G., Vorkoetter, S.M., McCarron, J., DeMarco, P.: Maple 10 Programming Guide. Maplesoft, Waterloo (2005)
22. Murphy, J.A., O'Donohoe, M.R.: Some properties of continued fractions with applications in Markov processes. IMA J. Appl. Math. **16**(1), 57–71 (1975)
23. O'Neil, J., Benita, Y., Feldman, I., Chenard, M., Roberts, B., Liu, Y., Li, J., Kral, A., Lejnine, S., Loboda, A., Arthur, W., Cristescu, R., Haines, B.B., Winter, C., Zhang, T., Bloecher, A., Shumway, S.D.: An unbiased oncology compound screen to identify novel combination strategies. Mol. Cancer Therapeut. **15**(6), 1155 (2016)
24. Petkovšek, M., Wilf, H., Zeilberger, D.: A = B. A K Peters/CRC Press, Natick (1996). ISBN: 978-1-56881-063-8
25. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes: The Art of Scientific Computing, 3rd edn. Cambridge University Press, Cambridge (2007). ISBN: 978-0-521-88068-8
26. Slater, L.J.: Generalized Hypergeometric Functions. Cambridge University Press, Cambridge (1966). ISBN: 978-0-521-06483-5

27. Stummel, F.: Rounding error analysis of elementary numerical algorithms. In: Fundamentals of Numerical Computation (Computer-Oriented Numerical Analysis), volume 2 of Computing Supplementa, pp. 169–195. Springer, Vienna (1980). ISBN: 978-3-7091-8577-3
28. Tavaré, S.: The linear birth-death process: an inferential retrospective. Adv. Appl. Probab. **50**(A), 253–269 (2018)
29. Zeilberger, D.: The method of creative telescoping. J. Symb. Comput. **11**(3), 195–204 (1991)