Published in final edited form as:

Stat Probab Lett. 2016 February 1; 109: 30–38. doi:10.1016/j.spl.2015.10.013.

Coupling bounds for approximating birth-death processes by truncation

Forrest W. Crawford¹, Timothy C. Stutz², and Kenneth Lange^{2,3}

Forrest W. Crawford: forrest.crawford@yale.edu

¹Departments of Biostatistics and Ecology & Evolutionary Biology, Yale University, 60 College St, PO Box 208034 New Haven, CT 06510 USA, phone: (203) 785-6125

²Department of Biomathematics, University of California, Los Angeles

³Departments of Human Genetics and Statistics, University of California, Los Angeles

Abstract

Birth-death processes are continuous-time Markov counting processes. Approximate moments can be computed by truncating the transition rate matrix. Using a coupling argument, we derive bounds for the total variation distance between the process and its finite approximation.

1 Introduction

A general birth-death process (BDP) is a continuous-time Markov chain $X(t)$, $t > 0$ taking values on the non-negative integers $\mathbb{N}$ (Karlin and McGregor, 1957). A “birth” from state $k \in \mathbb{N}$ to $k + 1$ occurs with instantaneous non-negative rate λ_k , and a “death” to $k - 1$ occurs with rate μ_k . Typically λ_k and μ_k are functions of k and are time-homogeneous. To conserve probability on $\mathbb{N}$, it is customary to set $\mu_0 = 0$. BDPs are widely used as phenomenological and mechanistic models in ecology, evolution, genetics, and other fields (Novozhilov et al, 2006). For example, the Poisson process results from taking $\lambda_k = \lambda$ and $\mu_k = 0$. Kendall’s simple linear birth-death process has $\lambda_k = k\lambda$ and $\mu_k = k\mu$ (Kendall, 1948, 1949). The Yule (pure-birth) process has $\lambda_k = k\lambda$ and $\mu_k = 0$. The Moran process in population genetics (Moran, 1958), logistic population growth (Tan and Piantadosi, 1991), and the susceptible-infective-susceptible infectious disease model (Andersson and Britton, 2000) have more complicated rates. BDPs are also useful for specification of arbitrary probability distributions on $\mathbb{N}$ with features such as power-law tail behavior (Klar et al, 2010), or over- and under-dispersion relative to the Poisson distribution (Faddy, 1997).

Researchers have explored the theoretical properties of BDPs (Karlin and McGregor, 1957; Guillemin and Pinchon, 1998; Flajolet and Guillemin, 2000). Although many properties of BDPs have been characterized analytically, moments still can be difficult or impossible to compute in practical settings (Novozhilov et al, 2006; Renshaw, 2011). Moments are

important for both prospective modeling to determine the dynamics of a specified process and statistical inference to estimate parameters using realizations from the process. Applied researchers must either use simpler BDPs that have analytic solutions or resort to numerical simulations to determine the mean dynamics of the process. Methods for calculating moments of general BDPs are available when the birth and death rates λ_j and μ_j are linear functions of j , but for more sophisticated models in which analytic information is unavailable, finite-time solutions remain elusive (Novozhilov et al, 2006), except in some special cases (e.g. Näsell, 2003). There also exist approximations to the moments using a polynomial representation (e.g. Engblom and Pender, 2014). Recently some progress has been made on computational routines for evaluating probabilities for arbitrary BDPs, when no analytic information is known *a priori* about the BDP beyond its birth and death rates. Following Murphy and O'Donohoe (1975), Crawford and Suchard (2012) and Crawford et al (2014) provide computational methods to evaluate transition probabilities and perform statistical estimation for BPDs. But there is currently no general method for computing the finite-time conditional moments $m_i^k(t) = E[X^k(t) | X(0) = i]$ of a general BDP with arbitrarily specified birth and death rates.

Although a BDP is fully characterized by its stochastic rate matrix (infinitesimal generator), this matrix is infinite-dimensional when the state space is $\mathbb{N}$. One straightforward approach to computing moments is to truncate the rate matrix at a sufficiently large index. After truncation, the problem of computing moments becomes tractable using established tools from linear algebra. Choice of the truncation index n determines the accuracy of the approximation. But without knowing its behavior in advance, it is unclear how to choose n for a given BDP: if n is chosen too small, then $X(t) > n$ might occur with high probability, resulting in error; on the other hand, if n is too large, matrix computations should be sufficiently precise, but may be unnecessarily slow. Finding the right truncation index without explicitly computing transition probabilities for the process in advance is a major challenge for modelers of BDPs.

In this paper, we show how to compute the moments of a general BDP by dynamic truncation of the state space. We illustrate upper and lower bounding processes for the conditioned process $X(t) | X(0) = i$ that provide insight into its dynamics. To choose the truncation index n we provide coupling bounds for the total variation distance between lower and upper bounding processes and the true process. We discuss methods for numerical computation of the relevant matrices by singular value decomposition and provide numerical illustrations.

2 Background

Let $X(t)$ be the state of a BDP at time $t > 0$, and suppose $X(0) = i \in \mathbb{N}$. Define the transition probability $P_{ij}(t) = \Pr(X(t) = j | X(0) = i)$, where i and j are non-negative integers. These probabilities obey the backward equations

$$\frac{d}{dt} P_{ij}(t) = \lambda_i P_{i+1,j}(t) + \mu_i P_{i-1,j}(t) - (\lambda_i + \mu_i) P_{ij}(t). \quad (1)$$

with initial conditions $P_{ii}(0) = 1$ and $P_{ij}(0) = 0$ for $j \neq i$. This system can be represented as $\frac{d}{dt}\mathbf{P}(t) = \mathbf{A}\mathbf{P}(t)$, where

$$\mathbf{A} = \begin{pmatrix} -\lambda_0 & \lambda_0 & & & 0 \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & & \\ & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & \\ & & \mu_3 & -(\lambda_3 + \mu_3) & \lambda_3 \\ 0 & & & \ddots & \ddots & \ddots \end{pmatrix}$$

is the infinite-dimensional stochastic rate matrix, subject to the initial condition $\mathbf{P}(0) = \mathbf{I}$. The solution may be written informally as $\mathbf{P}(t) = \exp[\mathbf{A}t]$. Karlin and McGregor (1957) show that the transition probability for any general BDP can be expressed as

$$P_{ij}(t) = \pi_j \int_0^\infty e^{-xt} Q_i(x) Q_j(x) \psi(dx), \quad (2)$$

where $\pi_j = (\lambda_0 \cdots \lambda_{j-1}) / (\mu_1 \cdots \mu_j)$ for $n \geq 1$, $\pi_0 = 1$, the functions $Q_j(x)$ are a family of orthogonal polynomials, and ψ is an orthogonalizing spectral measure. Unfortunately, expression (2) can be difficult or impossible to evaluate. Indeed, the orthogonal polynomials $Q_j(x)$ and corresponding measure ψ are only known for a few special BDPs (Novozhilov et al, 2006; Renshaw, 2011). In what follows, we assume that the process is well behaved and the transition probabilities and moments exist.

2.1 Moments

The ability to rapidly prototype a new BDP by studying its moments would be an important benefit to applied researchers who wish to determine whether a given set of rates $\{\lambda_k\}$, $\{\mu_k\}$ provide the desired dynamics on average. Let

$$m_i^k(t) = E[X^k(t) | X(0) = i] = \sum_{j=0}^{\infty} j^k P_{ij}(t) \quad (3)$$

be the k th moment of $X(t)$, conditional on $X(0) = i \in \mathbb{N}$. It is computationally impractical to compute moments directly from transition probabilities in this way. Multiplying the backward equations (1) by j^k and summing from $j = 0$ to ∞ , we obtain

$$\frac{d}{dt} m_i^k(t) = \lambda_i m_{i+1}^k(t) + \mu_i m_{i-1}^k(t) - (\lambda_i + \mu_i) m_i^k(t). \quad (4)$$

The infinite-dimensional matrix representation is therefore $\frac{d}{dt} \mathbf{m}^k(t) = \mathbf{A} \mathbf{m}^k(t)$, with initial condition $\mathbf{m}^k(0) = i^k \mathbf{e}_i$, where $\mathbf{e}_i$ is the infinite vector containing 1 at position i , and zero at all other positions. The solution is $\mathbf{m}^k(t) = e^{t\mathbf{A}} \mathbf{m}^k(0)$. Of course, we cannot evaluate the infinite-dimensional matrix exponential directly. In the following Sections, we outline a method for truncation of the vectors $\mathbf{m}^k(t)$ and the matrix $\mathbf{A}$ to minimize a given error

criterion. This results in a simple and robust heuristic for computing moments of the form (3) for arbitrary BDPs.

3 Coupling Bounds from Hitting Times

We now bound the distance between $X(t)$ and a finite-state-space approximation truncated at $n \in \mathbb{N}$ by introducing a coupling between random variables (see, e.g. Lange, 2010; Lindvall, 2002). Let $X(t)$ be a BDP whose state space is $\mathbb{N}$ and consider a second BDP $X_n(t)$ that executes the same birth-death chain as $X(t)$ on the states $\{0, 1, \dots, n\}$ except that the birth rate $\lambda_n = 0$. Then $X_n(t)$ and $X(t)$ move as one when $X_n(t) = X(t) < n$. When $X_n(t) = X(t) = n$, only $X(t)$ can jump upward in a birth event from n ; when $X(t) > n$, $X_n(t)$ evolves independently of $X(t)$ until such time as $X(t) = X_n(t)$. These relationships imply a well-defined joint distribution for $(X(t), X_n(t))$ for the coupled processes that will prove useful below. It follows that $X_n(t) \leq X(t)$ for all $t > 0$, so $X_n(t)$ is a lower bound for $X(t)$.

Consequently all moments satisfy $E[X_n^k(t)] \leq E[X^k(t)]$.

We seek to determine how well $X(t)$ is approximated by $X_n(t)$. Define the transition probability

$$P_{ij}^n(t) = \Pr(X_n(t) = j | X_n(0) = i) \quad (5)$$

of the process truncated at n and suppose $X(0) = X_n(0) = i$. Let ν_t be the measure induced by $P_{ij}(t)$ and let $\nu_{n,t}$ be the measure induced by $P_{ij}^n(t)$ with i, t , and n held fixed (we suppress the dependence of ν_t and $\nu_{n,t}$ on i for notational clarity). For a subset $S \subset \mathbb{N}$,

$$\nu_t(S) = \sum_{j \in S} P_{ij}(t) \quad \text{and} \quad \nu_{n,t}(S) = \sum_{j \in S} P_{ij}^n(t). \quad (6)$$

The total variation (TV) distance between the random variables $X(t)$ and $X_n(t)$ is

$$d_{TV}(\nu_t, \nu_{n,t}) = \frac{1}{2} \sum_{j=0}^{\infty} |P_{ij}(t) - P_{ij}^n(t)| = \sup_{S \subset \mathbb{N}} |\Pr(X(t) \in S) - \Pr(X_n(t) \in S)|. \quad (7)$$

3.1 Hitting Times for the Full Process

Let T_{n+1} be the hitting time at which $X(t)$ first reaches state $n+1$ and let $S = \{0, \dots, n\}$.

Conditioning yields

$$\Pr(X(t) \in S) = \Pr(X(t) \in S | T_{n+1} > t) \Pr(T_{n+1} > t) + \Pr(X(t) \in S | T_{n+1} \leq t) \Pr(T_{n+1} \leq t). \quad (8)$$

Given the event $T_{n+1} > t$, $X(t)$ cannot have exceeded $n+1$ before time t . In this setting $X(t)$ has the same state space and the same transition rates as the truncated process $X_n(t)$.

Replacing the first conditional probability produces

$$\begin{aligned}
\Pr(X(t) \in S) &= \Pr(X_n(t) \in S) \Pr(T_{n+1} > t) + \Pr(X(t) \in S | T_{n+1} \leq t) \Pr(T_{n+1} \leq t) \\
&= \Pr(X_n(t) \in S) [1 - \Pr(T_{n+1} \leq t)] + \Pr(X(t) \in S | T_{n+1} \leq t) \Pr(T_{n+1} \leq t) \\
&= \Pr(X_n(t) \in S) + [\Pr(X(t) \in S | T_{n+1} \leq t) - \Pr(X_n(t) \in S)] \Pr(T_{n+1} \leq t).
\end{aligned} \tag{9}$$

Rearranging (9) and applying Markov's inequality give

$$\begin{aligned}
|\Pr(X(t) \in S) - \Pr(X_n(t) \in S)| &\leq |\Pr(X(t) \in S | T_{n+1} \leq t) - \Pr(X_n(t) \in S)| \Pr(T_{n+1} \leq t) \\
&\leq \Pr(T_{n+1} \leq t) \\
&\leq e^{\theta t} \mathbb{E}[e^{-\theta T_{n+1}}]
\end{aligned} \tag{10}$$

for $\theta \geq 0$. The total variation bound $d_{TV}(v_t, v_{n,t}) \leq e^{\theta t} \mathbb{E}(e^{-\theta T_{n+1}})$ is an immediate consequence.

Now let $w_i(\theta) = \mathbb{E}(e^{-\theta T_{n+1}} | X(0) = i)$. We can deduce a system of recurrence relationships for the w_i 's by applying the Laplace transform to the backwards equation (1). Conditioning on the first event of the birth-death process and dividing by $\mu_i + \lambda_i + \theta$ gives

$$w_0(\theta) = \frac{\lambda_0}{\lambda_0 + \theta} w_1(\theta) \quad \text{and} \quad w_n(\theta) = \frac{\mu_n}{\mu_n + \lambda_n + \theta} w_{n-1}(\theta) + \frac{\lambda_n}{\mu_n + \lambda_n + \theta} \tag{11}$$

and, for $1 \leq i \leq n-1$,

$$\begin{aligned}
w_i(\theta) &= \frac{\mu_i}{\mu_i + \lambda_i} \frac{\mu_i + \lambda_i}{\mu_i + \lambda_i + \theta} w_{i-1}(\theta) + \frac{\lambda_i}{\mu_i + \lambda_i} \frac{\mu_i + \lambda_i}{\mu_i + \lambda_i + \theta} w_{i+1}(\theta) \\
&= \frac{\mu_i}{\mu_i + \lambda_i + \theta} w_{i-1}(\theta) + \frac{\lambda_i}{\mu_i + \lambda_i + \theta} w_{i+1}(\theta).
\end{aligned} \tag{12}$$

These can be written as the matrix-vector equation

$$\begin{pmatrix} 1 & a_0 & & & 0 \\ b_1 & 1 & a_1 & & \\ & b_2 & 1 & & \\ & & \ddots & \ddots & \\ 0 & & & b_n & 1 \end{pmatrix} \begin{pmatrix} w_0(\theta) \\ \vdots \\ \vdots \\ \vdots \\ w_n(\theta) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ \vdots \\ 0 \\ \frac{\lambda_n}{\mu_n + \lambda_n + \theta} \end{pmatrix} \tag{13}$$

where $a_i = -\frac{\lambda_i}{\mu_i + \lambda_i + \theta}$ and $b_i = -\frac{\mu_i}{\mu_i + \lambda_i + \theta}$. There are well established numerical methods for solving the above tridiagonal system for the expectations $w_i(\theta)$.

According to Chernoff's method, for fixed n one achieves the tightest bound on $d_{TV}(v_t, v_{n,t})$ by choosing θ to minimize $e^{\theta t} \mathbb{E}(e^{-\theta T_{n+1}})$. The function $f(\theta) = \mathbb{E}[e^{\theta(t - T_{n+1})}]$ is convex in θ with derivative $f'(\theta) = \mathbb{E}[(t - T_{n+1})e^{\theta(t - T_{n+1})}]$. Since $f'(0) = t - \mathbb{E}(T_{n+1})$, the minimum on $[0, \infty)$ occurs at 0 where $f(0) = 1$ whenever $t \leq \mathbb{E}(T_{n+1})$. Otherwise, the minimum occurs on the interior of $(0, \infty)$. Minimizing $f(\theta)$ for $\theta \geq 0$ can be achieved via Newton's method. The derivatives $w'_i(\theta)$ are calculated via a tridiagonal system similar to equation (13) derived by

differentiating equations (11) and (12). For fixed n , these maneuvers allow one to calculate the bound

$$d_{TV}(\nu_t, \nu_{n,t}) \leq \min_{\theta \geq 0} e^{\theta t} E(e^{-\theta T_{n+1}}). \quad (14)$$

As $n \rightarrow \infty$, $E(e^{-\theta T_{n+1}}) \rightarrow 0$ for any $\theta > 0$. Therefore increasing n will decrease the bound until a desired error $d_{TV}(\nu_t, \nu_{n,t}) < \varepsilon$ is achieved. While this does not provide an explicit analytic expression, the right-hand side of (14) is straightforward to compute for fixed n and a given set of birth and death rates. We now present a simpler approach that does yield explicit bounds.

3.2 Hitting Times for the Pure-Birth Process

Calculation of hitting probabilities via equation (13) is appealing, but can be computationally burdensome. A simpler procedure that results in a looser bound arises by considering a pure-birth process $Y(t)$ that shares its birth rates λ_j with $X(t)$ but has $\mu_j = 0$ for all $j \in \mathbb{N}$. Let $X(t)$ and $Y(t)$ move independently when $X(t) < Y(t)$. When $X(t) = Y(t)$, they move as one at a jump corresponding to a birth. At a jump corresponding to a death, $X(t)$ takes the downward jump while $Y(t)$ does not move. It is clear that $X(t) \leq Y(t)$ for all $t > 0$, so $Y(t)$ is an upper bound for $X(t)$. We assume the process $Y(t)$ is non-explosive, so its first-passage time to infinity is infinite, i.e. $\sum_{j=0}^{\infty} \lambda_j^{-1} = \infty$. Then $Y(t)$ is finite with probability 1 for all $t > 0$. In view of the coupling between $Y(t)$ and $X(t)$, it is clear that $X(t)$ enjoys the same property. We now address approximation of the process $X(t)$ by the truncated processes $X_n(t)$.

Proposition 3.1—*Consider a BDP $X(t)$ with $X(0) = i$, and assume the corresponding pure birth process $Y(t)$ is nonexplosive. Then the truncated processes $X_n(t)$ are increasing in n and converge to X_t as $n \rightarrow \infty$. Furthermore, the conditional moments $E[X_n^k(t) | X_n(0) = i]$ converge monotonically to $E[X^k(t) | X(0) = i]$, and the transition probabilities $\Pr(X_n(t) = j | X_n(0) = i)$ converge to $\Pr(X(t) = j | X(0) = i)$.*

Proof: If the increasing sequence $X_n(t)$ converges to $X(t)$, then the limit relation

$$\lim_{n \rightarrow \infty} E[X_n^k(t) | X_n(0) = i] = E[X^k(t) | X(0) = i] \quad (15)$$

follows from the monotone convergence theorem. Likewise, the limit relation

$$\lim_{n \rightarrow \infty} \Pr(X_n(t) = j | X_n(0) = i) = \Pr(X(t) = j | X(0) = i) \quad (16)$$

follows from the bounded convergence theorem since each transition probability in question is the expectation of an indicator random variable and as such is bounded by the constant 1.

To show that the truncated processes are increasing, imagine that the $X_n(t)$ are increasing and occur in blocks. Thus, $X_m(t)$ and $X_n(t)$ occupy the same block whenever $X_m(t) = X_n(t)$.

Different blocks behave independently. Within a block the processes move as one except when $X_n(t)$ in a block takes its maximal value n . In this case a proposed birth is declined and the process $X_n(t)$ stays in place. Monotonicity is preserved by this protocol for updating the participating processes. Note the total number of blocks at any time $s \leq t$ is bounded above by the finite number $Y(s) \leq Y(t)$. Finally, convergence can be proved by considering a typical realization of the combined processes. The state $Y(t)$ represents the outermost limit that $X(s)$ or any $X_n(s)$ can reach during $[0, t]$. Hence, $X_n(t) = X(t)$ as soon as n exceeds $Y(t)$. In other words, $X_n(t)$ converges monotonically to $X(t)$.

4 Rate of Convergence

We now establish a bound for the TV distance between the original and truncated processes.

Proposition 4.1—Suppose $X(0) = X_n(0) = Y(0) = i < n$ and $t < \sum_{m=i}^n \lambda_m^{-1}$. Then the total variation distance between the $X(t)$ and $X_n(t)$ is bounded above by

$$d_{TV}(\nu_t, \nu_{n,t}) \leq e^{st} \prod_{m=i}^n \frac{\lambda_m}{\lambda_m + s}, \quad (17)$$

where $s > 0$ satisfies $t = \sum_{m=i}^n (\lambda_m + s)^{-1}$.

Proof: First, the bound

$$d_{TV}(\nu_t, \nu_{n,t}) \leq \Pr(X(t) \neq X_n(t) | X(0) = X_n(0) = i) \leq \Pr(Y(t) > n | Y(0) = i) \quad (18)$$

applies because the event $X(t) \neq X_n(t)$ is contained in the event $Y(t) > n$. Furthermore, $Y(t) > n$ is equivalent to the event $W_i + \dots + W_n \leq t$, where W_j is the waiting time from the arrival of the pure-birth process $Y(t)$ in state j until it moves to $j + 1$. Applying the Chernoff bound, we have

$$\Pr(W_i + \dots + W_n \leq t) \leq \min_{s \geq 0} e^{st} \prod_{m=i}^n \mathbb{E}[e^{-sW_m}] \quad (19)$$

The Laplace transform of W_i is $\mathbb{E}[e^{-sW_i}] = \lambda_m / (\lambda_m + s)$, and it follows that

$$\Pr(W_i + \dots + W_n \leq t) \leq \min_{s \geq 0} e^{st} \prod_{m=i}^n \frac{\lambda_m}{\lambda_m + s}, \quad (20)$$

and the optimal s for a given n minimizes the convex function

$$f_n(s) = st - \sum_{m=i}^n \log(\lambda_m + s). \quad (21)$$

Differentiating $f_n(s)$ with respect to s , we find that the optimal s satisfies

$$t = \sum_{m=i}^n \frac{1}{\lambda_m + s} \quad (22)$$

for $t < \sum_{m=i}^n \lambda_m^{-1}$, as claimed.

The condition $t < \sum_{m=i}^n \lambda_m^{-1}$ guarantees that t is less than the expected first passage time of the pure-birth process from i to n . Clearly if this were not true, we would expect $Y(t) > n$ with high probability, and hence approximation of $X(t)$ by $X_n(t)$ would be poor.

4.1 Closed-form Bounds

The representation

$$e^{st} \prod_{m=i}^n \frac{\lambda_m}{\lambda_m + s} = \exp \left[st - \sum_{m=i}^n \log \left(1 + \frac{s}{\lambda_m} \right) \right] \quad (23)$$

suggests exploiting the logarithmic inequality $-\log(1+u) \leq -u + u^2/2$ for $u \in [0, \infty)$. It follows that

$$e^{st} \prod_{m=i}^n \frac{\lambda_m}{\lambda_m + s} \leq \exp \left[st + \sum_{m=i}^n \left[-\frac{s}{\lambda_m} + \frac{1}{2} \left(\frac{s}{\lambda_m} \right)^2 \right] \right]. \quad (24)$$

The right-hand side of this inequality is minimized by the choice

$$s = \frac{\sum_{m=i}^n \lambda_m^{-1} - t}{\sum_{m=i}^n \lambda_m^{-2}}. \quad (25)$$

Inserting this argument in the previous inequality produces the upper bound

$$d_{TV}(\nu_t, \nu_{n,t}) \leq e^{-\frac{a_n^2}{2b_n}}, \quad (26)$$

where $a_n = \sum_{m=i}^n \lambda_m^{-1} - t$ and $b_n = \sum_{m=i}^n \lambda_m^{-2}$. If the pure-birth process is nonexplosive, $\sum_{n=1}^{\infty} \lambda_n^{-1} = \infty$, and so $\lim_{n \rightarrow \infty} a_n = \infty$ as well. When the condition

$$\lim_{n \rightarrow \infty} \frac{(\sum_{m=1}^n \lambda_m^{-1})^2}{\sum_{m=i}^n \lambda_m^{-2}} = \infty \quad (27)$$

holds, (24) can be made as small as desired, so one can easily find a sufficiently large n that renders $d_{TV}(\nu_t, \nu_{n,t})$ as small as desired.

5 Matrix Exponentiation for the Coupled Process

One approach to calculating moments and probabilities is to substitute the truncated process just described for the actual process. Given the tridiagonal nature of the infinitesimal generator, this will deliver accurate estimates at a computational cost of $O(n^2)$ (Dhillon and Parlett, 2004). Let us assume that the original process possesses a well-behaved equilibrium distribution π . The first task is to approximate π accurately. Detailed balance gives

$$\pi_0 = \frac{1}{1 + \sum_{k=1}^{\infty} \prod_{j=1}^k \frac{\lambda_{j-1}}{\mu_j}} \quad \text{and} \quad \pi_k = \pi_0 \prod_{j=1}^k \frac{\lambda_{j-1}}{\mu_j} \quad \text{for } k > 0. \quad (28)$$

The implied products and sums can be computed in $O(n)$ operations. The point at which the approximation to π_0 stabilizes gives a trial value of n .

Now consider the process $X_n(t)$ with truncated equilibrium distribution $\tilde{\pi}$ and truncated infinitesimal generator $\tilde{A}$ with $\lambda_n = 0$. Owing to detailed balance, the matrix $B = \sqrt{\tilde{\pi}} \tilde{A} \sqrt{\tilde{\pi}}^{-1}$ is symmetric and tridiagonal, where $\sqrt{\tilde{\pi}}$ denotes the diagonal matrix with k th diagonal entry $\sqrt{\pi_k}$. The spectral decomposition

$$B = V D V^t = \sum_{k=1}^n d_k v_k v_k^t$$

of B can be computed very efficiently (Demmel et al, 2008). Furthermore, the matrix exponential

$$e^{sB} = \sum_{k=1}^n e^{s d_k} v_k v_k^t$$

follows directly and gives

$$e^{s\tilde{A}} u = \sum_{k=1}^n e^{s d_k} (\sqrt{\tilde{\pi}} v_k)^t u (\sqrt{\tilde{\pi}}^{-1} v_k)$$

for any vector u . Thus, all truncated transition probabilities and moments are available at little additional cost.

6 Examples

6.1 Kendall Process

In the Kendall process, $\lambda_j = j\lambda$, and $\mu_j = j\mu$. Since the transition probabilities and first few moments are straightforward to calculate analytically, the Kendall process provides a good

illustration of the bounds in comparison to exact values. In this case, the upper bounding process $Y(t)$ is the Yule process with $\lambda_j = j\lambda$ and $\mu_j = 0$ for all $j \in \mathbb{N}$. Applying (26), we find that

$$d_{TV}(\nu_t, \nu_{n,t}) \leq \exp \left[-\frac{(\lambda^{-1} \sum_{m=i}^n m^{-1} - t)^2}{2\lambda^{-2} \sum_{m=i}^n m^{-2}} \right]. \quad (29)$$

The upper bounding process has transition probability

$$P_{ij}^Y(t) = \binom{j-1}{i-1} e^{-\lambda i t} (1 - e^{-\lambda t})^{j-i}$$

so we can calculate the probability that the upper bounding process travels beyond n explicitly as

$$\Pr(Y(t) > n) = 1 - \sum_{j=i}^n P_{ij}^Y(t). \quad (30)$$

Then we take the smallest n that achieves $d_{TV}(\nu_t, \nu_{n,t}) \leq \varepsilon$ as the truncation index, where $\varepsilon > 0$ is the maximum allowable total variation distance between the truncated and original processes. Figure 1 illustrates three calculated bounds: one using the Yule process probability (30), one using inequality (29), and one using inequality (14).

6.2 Kendall Process with Immigration and Emigration

Define $\lambda_j = j\lambda + \alpha$ for $j \geq 0$ and $\mu_j = j\mu + \gamma$ for $j \geq 1$ with $\mu_0 = 0$. When the immigration rate α and emigration rate γ are nonzero, the orthogonal polynomials and measure are available (Ismail et al, 1988). Applying the bound (26), we have

$$d(\nu_t, \nu_{n,t}) \leq \exp \left\{ -\frac{[\sum_{m=i}^n (m\lambda + \alpha)^{-1} - t]^2}{2\sum_{m=i}^n (m\lambda + \alpha)^{-2}} \right\}. \quad (31)$$

Figure 2 illustrates the Laplace and Chernoff bounds.

6.3 Logistic Growth

Crawford and Suchard (2012) give a BDP for logistic population growth,

$$\lambda_j = j^2 \lambda \frac{e^{-\alpha j}}{1 + e^{\beta(j-M)}} \quad \text{and} \quad \mu_j = j\mu$$

where M is the “carrying capacity” of the environment and α and β are non-negative parameters. This model seems to have no analytic solution, but we can easily find bounds for matrix truncation:

$$d_{TV}(\nu_t, \nu_{n,t}) \leq \exp \left\{ - \frac{\left(\sum_{m=i}^n [m^2 \lambda e^{-\alpha m} / (1 + e^{\beta(m-M)})]^{-1} - t \right)^2}{2 \sum_{m=i}^n [m^2 \lambda e^{-\alpha m} / (1 + e^{\beta(m-M)})]^{-2}} \right\}. \quad (32)$$

The birth rates in this example do not satisfy the condition (27) for the Chernoff bound to converge. Figure 3 illustrates the superior performance of the Laplace bound.

6.4 Faddy Distributions

Pure-birth processes are a special case of BDPs with $\mu_j = 0$ for all $j \in \mathbb{N}$. Faddy (1997) proposes a generalization of the Poisson distribution with $\lambda_j = \lambda(\gamma + j)^c$. Setting $c = 0$ recovers the Poisson process, $c > 0$ indicates over-dispersion, and $c < 0$ indicates under-dispersion relative to the Poisson distribution. Faddy (1997) employs a diffusion approximation to compute moments of the process, but the bounds derived in this paper apply directly. The Chernoff bound is

$$d_{TV}(\nu_t, \nu_{n,t}) \leq \exp \left\{ - \frac{[\sum_{m=i}^n (\gamma + m)^{-c} - \lambda t]^2}{2 \sum_{m=i}^n (\gamma + m)^{-2c}} \right\}. \quad (33)$$

Figure 4 shows examples of the bounds.

7 Discussion

In this paper, we have demonstrated bounds for the total variation distance between a truncated and untruncated BDPs. The Laplace approach (14) generally provides the tightest bound, but at a higher computational cost. We have shown that the pure-birth process $Y(t)$ is an upper bound for $X(t)$, and the total variation distance between the truncated process $X_n(t)$ and $X(t)$ is bounded above by the probability that $Y(t) > n$. The Chernoff bound (17) may be quite conservative if the death rates μ_m are large compared to the birth rates, but it can be very easy to compute.

Acknowledgments

FWC was supported by NIH grant KL2 TR000140. TCS was supported by the National Institute of General Medical Sciences of the NIH (T32GM008185). KL was supported by NIH grants from the National Human Genome Research Institute (HG006139) and the National Institute of General Medical Sciences (GM053275). Lam S. Ho and Marc A. Suchard provided helpful comments.

References

Andersson, H.; Britton, T. Lecture notes in statistics. Springer; New York: 2000. Stochastic Epidemic Models and their Statistical Analysis.

- Crawford FW, Suchard MA. Transition probabilities for general birth-death processes with applications in ecology, genetics, and evolution. *Journal of Mathematical Biology*. 2012; 65:553–580. [PubMed: 21984359]
- Crawford FW, Minin VN, Suchard MA. Estimation for general birth-death processes. *Journal of the American Statistical Association*. 2014; 109(506):730–747. [PubMed: 25328261]
- Demmel JW, Marques OA, Parlett BN, Vömel C. Performance and accuracy of LAPACK's symmetric tridiagonal eigensolvers. *SIAM Journal on Scientific Computing*. 2008; 30(3):1508–1526.
- Dhillon IS, Parlett BN. Multiple representations to compute orthogonal eigenvectors of symmetric tridiagonal matrices. *Linear Algebra and its Applications*. 2004; 387:1–28.
- Engblom S, Pender J. Approximations for the moments of nonstationary and state dependent birth-death queues. 2014 arXiv preprint arXiv:14066164.
- Faddy M. Extended Poisson process modelling and analysis of count data. *Biometrical Journal*. 1997; 39(4):431–440.
- Flajolet P, Guillemin F. The formal theory of birth-and-death processes, lattice path combinatorics and continued fractions. *Advances in Applied Probability*. 2000; 32(3):750–778.
- Guillemin F, Pinchon D. Continued fraction analysis of the duration of an excursion in an $M/M/\infty$ system. *Journal of Applied Probability*. 1998; 35(1):165–183.
- Ismail MEH, Letessier J, Valent G. Linear birth and death models and associated Laguerre and Meixner polynomials. *Journal of Approximation Theory*. 1988; 55(3):337–348.
- Karlin S, McGregor J. The differential equations of birth-and-death processes, and the Stieltjes moment problem. *Transactions of the American Mathematical Society*. 1957; 85(2):589–646.
- Kendall DG. On the generalized "birth-and-death" process. *Annals of Mathematical Statistics*. 1948; 19(1):1–15.
- Kendall DG. Stochastic processes and population growth. *Journal of the Royal Statistical Society B*. 1949; 11(2):230–282.
- Klar B, Parthasarathy PR, Henze N. Zipf and Lerch limit of birth and death processes. *Probability in the Engineering and Informational Sciences*. 2010; 24(01):129–144.
- Lange, K. Springer texts in statistics. 2. Springer; New York: 2010. Applied Probability.
- Lindvall, T. Lectures on the Coupling Method. Courier Corporation; 2002.
- Moran PAP. Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society*. 1958; 54(01):60–71.
- Murphy JA, O'Donohoe MR. Some properties of continued fractions with applications in Markov processes. *IMA Journal of Applied Mathematics*. 1975; 16(1):57–71.
- Nåsell I. Moment closure and the stochastic logistic model. *Theoretical Population Biology*. 2003; 63(2):159–168. [PubMed: 12615498]
- Novozhilov AS, Karev GP, Koonin EV. Biological applications of the theory of birth-and-death processes. *Briefings in Bioinformatics*. 2006; 7(1):70–85. [PubMed: 16761366]
- Renshaw, E. Stochastic Population Processes: Analysis, Approximations, Simulations. Oxford University Press; 2011.
- Tan WY, Piantadosi S. On stochastic growth processes with application to stochastic logistic growth. *Statistica Sinica*. 1991; 1:527–540.

Figure 1.

Probability bounds on the log scale for the total variation distance obtained between the Kendall process truncated at index n and the untruncated process, where $X(0) = i = 4$, $t = 1$, $\mu = 1$, and $\lambda = 1$. The solid line shows the derived bound (29). The dotted and dashed line shows $\Pr(Y(t) > n/Y(0) = 4)$ from (30). The dashed line shows the bounds obtained from the Laplace transforms (14). The dotted line shows the error tolerance $\varepsilon = 10^{-4}$.

Figure 2.

Probability bounds for the Kendall immigration-emigration process, where $X(0) = i = 4$, $t = 1$, $\alpha = 2$, $\gamma = 2$, $\mu = 1$, and $\lambda = 1$. The solid line shows the derived bound (29). The dashed line shows the bounds from the Laplace transforms (14).

Figure 3.

Probability bounds for the Logistic process truncated at index n , where $X(0) = i = 4$, $t = 1$, $\alpha = 0.2$, $\beta = 0.3$, $M = 20$, $\mu = 1$, and $\lambda = 1$. The solid line shows the derived bound (32). The dashed line shows the bounds obtained from the Laplace transforms (14).

Figure 4. Probability bounds for the Faddy process, where $X(0) = i = 4$, $t = 1$, $\gamma = 1.1$, $\mu = 1$, and $\lambda = 1$. The solid line shows the derived bound (33). The dashed line shows the bounds obtained from the Laplace transforms (14).