

# Computational methods for birth-death processes

Forrest W. Crawford<sup>1</sup>

Lam Si Tung Ho<sup>2</sup>

Marc A. Suchard<sup>3</sup>

<sup>1</sup>Departments of Biostatistics, Ecology & Evolutionary Biology, and School of Management, Yale University

<sup>2</sup>Department of Mathematics and Statistics, Dalhousie University, Halifax, Nova Scotia, Canada

<sup>3</sup>Departments of Biomathematics, Biostatistics and Human Genetics, University of California, Los Angeles

## Abstract

Many important stochastic counting models can be written as general birth-death processes (BDPs). BDPs are continuous-time Markov chains on the non-negative integers in which only jumps to adjacent states are allowed. BDPs can be used to easily parameterize a rich variety of probability distributions on the non-negative integers, and straightforward conditions guarantee that these distributions are proper. BDPs also provide a mechanistic interpretation – birth and death of actual particles or organisms – that has proven useful in evolution, ecology, physics, and chemistry. Although the theoretical properties of general BDPs are well understood, traditionally statistical work on BDPs has been limited to the simple linear (Kendall) process. Aside from a few simple cases, it remains impossible to find analytic expressions for the likelihood of a discretely-observed BDP, and computational difficulties have hindered development of tools for statistical inference. But the gap between BDP theory and practical methods for estimation has narrowed in recent years. There are now robust methods for evaluating likelihoods for realizations of BDPs: finite-time transition, first passage, equilibrium probabilities, and distributions of summary statistics that arise commonly in applications. Recent work has also exploited the connection between continuously- and discretely-observed BDPs to derive EM algorithms for maximum likelihood estimation. Likelihood-based inference for previously intractable BDPs is much easier than previously thought and regression approaches analogous to Poisson regression are straightforward to derive. In this review, we outline the basic mathematical theory for BDPs and demonstrate new tools for statistical inference using data from BDPs.

## GRAPHICAL TABLE OF CONTENTS

**Caption:** Realization of a birth-death process  $X(t)$ .

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1002/wics.1423](https://doi.org/10.1002/wics.1423)

## INTRODUCTION

Birth-death processes (BDPs) are a flexible class of continuous-time Markov chains that model the number of “particles” in a system, where each particle can “give birth” to another particle or “die” (Feller, 1971; Karlin & Taylor, 1975). The rate of births and deaths at any given time depends on how many extant particles there are. When there are  $k$  particles, a birth occurs with instantaneous rate  $\lambda_k$  and a death with instantaneous rate  $\mu_k$ . In the classical “simple linear” BDP,  $\lambda_k = k\lambda$  and  $\mu_k = k\mu$  so that per-particle birth and death rates remain constant. In a “general” BDP,  $\lambda_k$  and  $\mu_k$  can be any function of  $k$  but are time-homogeneous (Kendall, 1948, 1949). Table 1 gives examples of well-known BDPs and their birth and death rates. Figure 1 shows an example realization from a BDP.

The usefulness of BDPs lies in the fact that “particle” can refer to a member of any discrete potentially interacting system in which one only keeps track of the number of objects in existence. BDPs are popular modeling tools in evolution, population biology, genetics, and ecology (Novozhilov, Karev, & Koonin, 2006). For example, if we interpret the particles as species in a macro-evolutionary setting, BDPs can be used to study speciation and extinction over evolutionary timescales (Nee, 2006; Nee, May, & Harvey, 1994). BDPs can also be used to study infectious disease dynamics in a finite population, where the number of individuals infected is the quantity of interest (Andersson & Britton, 2000; N. T. J. Bailey, 1964). In molecular evolution, BDPs can model inserted and deleted nucleotides in a DNA or RNA sequence as part of a probabilistic alignment method (Holmes & Bruno, 2001; Thorne, Kishino, & Felsenstein, 1991), mobile/transposable genetic elements (Rosenberg, Tsolaki, & Tanaka, 2003), gene families (Demuth, De Bie, Stajich, Cristianini, & Hahn, 2006), or even whole chromosomes (Mayrose, Barker, & Otto, 2010). BDPs can model populations of organisms in a resource-limited environment (Renshaw, 1993, 2011; Tan & Piantadosi, 1991). In finite populations, BDPs are commonly used to model quantities of interest in an evolutionary setting, such as mutations, allele frequencies, selection, or coalescence (Fudenberg, Imhof, Nowak, & Taylor, 2004; Kingman, 1982; Krone & Neuhauser, 1997; McFarland, Mirny, & Korolev, 2014; P. A. P. Moran, 1958; Nowak, Michor, & Iwasa, 2003; Nowak, Sasaki, Taylor, & Fudenberg, 2004).

Branching processes are popular tools used in biological modeling that share some of these properties with BDPs (Guttorp, 1991; Kimmel & Axelrod, 2016). Branching processes are specified by assumptions about what happens to individual particles, which typically behave independently: they can die, give birth to or be replaced by offspring of the same or different types. But the nature of this process is individualistic – rules about how individual particles behave in isolation give rise to models for population-level dynamics. In contrast, BDPs are specified at the population level. The conditional rate of particle birth and death given the numbers in existence can depend arbitrarily on the current size of the population.

Many important models in queuing theory can be written as general BDPs (Norris, 1998; Renshaw, 2011; Ross, 1995). In basic Markovian queues, customers arrive into a queue or

buffer as a Poisson process with rate  $\lambda$ , and waiting customers are served (removed from the queue) with per-customer service rate  $\mu$ . In the  $M/M/\infty$  queue, also known as the immigration-death process, there are infinitely many servers, so the arrival and service (birth and death) rates are  $\lambda_k = \lambda$  and  $\mu_k = k\mu$  for  $k > 0$ . In the  $M/M/1$  queue, also known as the immigration-emigration model, there is only a single server, so the rates are  $\lambda_k = \lambda$  and  $\mu_k = \mu$ . In the  $M/M/c$  queue, there are exactly  $c$  servers, so  $\mu_k = \min\{c, k\}\mu$ .

BDPs can also be useful for defining arbitrary probability distributions on the non-negative integers. Crawford and Zelterman (2015) demonstrate that any sum of exchangeable Bernoulli random variables can be exactly represented as a pure-birth BDP. In fact, M. J. Faddy (1997) shows that one can define a pure birth process (a BDP with death rates  $\mu_k = 0$  for all  $k$ ) whose transition probabilities reproduce any discrete distribution on the counting numbers. Klar, Parthasarathy, and Henze (2010) establish a correspondence between several power law distributions and the long-time limit of specially constructed BDPs, providing a time-dependent interpretation that may be useful for modelling mechanistic processes that give rise to power law outcomes. Sometimes this power law behavior is motivated by a mechanistic model: researchers have developed models and statistical methodology for estimating BDP parameters for power law behavior in protein domains (Karev, Wolf, Rzhetsky, Berezovskaya, & Koonin, 2002) and gene and protein family sizes (Reed & Hughes, 2004). Crawford and Suchard (2012) define a BDP to mimic a process of frameshift-aware insertions and deletions in DNA sequences. Lee, Weiss, and Suchard (2011) set the birth and death rates of a BDP to exhibit over-dispersion relative to the Poisson distribution, and Crawford, Weiss, and Suchard (2015) define a BDP to model rounding in counts of sex partners to multiples of 5, 10, 25, or 50 in self-reported counts of sex partners in a public health study.

There is a rich history of theoretical research into the properties of BDPs. Kendall (1948, 1949) introduce the process with constant per-particle birth and death rates and finds the transition probabilities by a generating function argument. In their groundbreaking series of papers, Karlin and McGregor (1957b) analyze properties of BDPs, including stationary distributions, moments, transition probabilities, recurrence and passage times, and other quantities of interest (Karlin & McGregor, 1957a, 1957b). They also explore in-depth applications of this theory to BDPs whose rates depend linearly on  $k$  (Karlin & McGregor, 1958a), and queuing processes (Karlin & McGregor, 1958b).

Beyond the work of Karlin and McGregor (1957b), many authors have discovered extensions and deeper interpretations for the theoretical properties of BDPs. For example, the theory of BDPs is intimately related to properties of continued fractions (Guillemin & Pinchon, 1999). Flajolet and Guillemin (2000) elucidate the relationship between sample trajectories (or state paths) of a BDP and lattice path combinatorics via continued fractions and develop expressions for a variety of recurrence and passage time variables in terms of continued fractions. Lenin and Parthasarathy (2000) and Parthasarathy, Lenin, Schoutens, and Van Assche (1998) discuss further some well-known continued fractions whose connection to BDPs previously went unappreciated. The study of BDPs has benefited from wide interest in the theoretical properties of this class of processes. But their usefulness as flexible tools for statistical inference has been under-appreciated. In this review, we outline basic properties of BDPs and show how to perform principled statistical inference using data from continuous and discrete observation of BDPs.

First, we present the basic time-evolution equations of general BDPs, derive the transition probabilities for the Kendall process (Feller, 1971; Kendall, 1948), and describe the analytic theory developed by Karlin and McGregor (1957a, 1957b) for general BDPs. Then we outline a computational strategy for evaluating BDP transition probabilities using a continued fraction representation of their Laplace transform, which allows routine computation of likelihoods for discretely observed processes (Crawford & Suchard, 2012). We describe a generic class of EM algorithms for maximum likelihood (or maximum *a posteriori*) inference for discretely observed BDPs (Crawford, Minin, & Suchard, 2014). Finally, we derive the distribution of integral summary statistics of BDPs that arise often in applications.

## BACKGROUND

A BDP is a continuous-time Markov chain  $X(t)$  counting the number of particles in a system at time  $t$ , taking values on the non-negative integers  $\mathbb{N}$ . To construct a general BDP in a formal way, we must define the rules according to which the number of particles evolves. We do this by specifying the behavior of the process for a very short time  $dt$ , when there are  $k$  particles in the system. If  $dt$  is very small, the probability of an event during  $(t, t + dt)$  that occurs with rate  $r$  is approximately  $rdt$ . Therefore, the probability of a birth in the interval  $(t, t + dt)$ , given  $X(t) = k$ , is

$$\Pr(X(t + dt) = k + 1 \mid X(t) = k) = \lambda_k dt + o(dt). \quad (1)$$

Intuitively, this means that the probability of more than one birth event in a small time  $dt$  is negligibly small. The probability of a death in  $(t, t + dt)$  is likewise

$$\Pr(X(t + dt) = k - 1 \mid X(t) = k) = \mu_k dt + o(dt), \quad (2)$$

where  $k \geq 1$ . Together, these assumptions imply that the probability of no births or deaths occurring during  $(t, t + dt)$  is

$$\Pr(X(t + dt) = k \mid X(t) = k) = 1 - (\lambda_k + \mu_k)dt + o(dt). \quad (3)$$

## TRANSITION PROBABILITIES

Let  $P_{ab}(t) = \Pr(X(t) = b \mid X(0) = a)$  be the transition probability from state  $X(0) = a$  to  $X(t) = b$ . We can use the above expressions to form a differential equation describing the change in transition probabilities over time. Suppose that  $X(0) = a$ . At the current time  $t$ , we want to know the probability that in the next  $dt$  units of time, the process will reach state  $b$ . We look

into the future by writing the probabilities of three types of events that can take the process to state  $b$ : birth from  $b - 1$ , death from  $b + 1$ , or no change from  $b$ :

$$P_{ab}(t + dt) = \lambda_{b-1}P_{a,b-1}(t)dt + \mu_{b+1}P_{a,b+1}(t)dt + (1 - \lambda_b - \mu_b)P_{ab}(t)dt + o(dt). \quad (4)$$

Subtracting  $P_{ab}(t)$  from both sides, dividing by  $dt$ , and sending  $dt$  to zero, we obtain the Kolmogorov forward equations:

$$\frac{dP_{ab}(t)}{dt} = \lambda_{b-1}P_{a,b-1}(t) + \mu_{b+1}P_{a,b+1}(t) - (\lambda_b + \mu_b)P_{ab}(t), \quad (5)$$

where  $P_{ab}(0) = 1$  if  $a = b$  and zero otherwise. In this article, we always assume  $\mu_0 = \lambda_{-1} = 0$ ; this keeps the process on the non-negative integers. Letting  $\mathbf{P}(t) = \{P_{ab}(t)\}$  in matrix form, (5) becomes

$$\frac{d\mathbf{P}(t)}{dt} = \mathbf{P}(t)\mathbf{A}, \quad (6)$$

where  $\mathbf{A}$  is the infinitesimal generator matrix with entries  $\mathbf{A} = \{a_{ij}\}$ ,  $a_{i,i-1} = \mu_i$ ,  $a_{ii} = -(\lambda_i + \mu_i)$ , and  $a_{i,i+1} = \lambda_i$ . In the matrix case, the initial condition becomes  $\mathbf{P}(0) = \mathbf{I}$ . This infinite sequence of coupled ordinary differential equations can be difficult or impossible to solve for many general BDPs (Novozhilov et al., 2006; Renshaw, 2011).

## KENDALL PROCESS

In the simple linear BDP, also known as the Kendall process where  $\lambda_k = k\lambda$  and  $\mu_k = k\mu$ , it is possible to solve for these transition probabilities explicitly by finding a generating function solution to the forward equations (N. T. J. Bailey, 1964; Lange, 2010a). To illustrate, let

$G_a(s, t) = \sum_{k=0}^{\infty} s^k P_{ak}(t)$ . Let  $b = k$  in (5), multiply both sides by  $s^k$ , and sum on  $k$  to obtain

$$\begin{aligned} \frac{\partial G_a(s, t)}{\partial t} &= \sum_{k=0}^{\infty} s^k \frac{dP_{ak}(t)}{dt} \\ &= \lambda s^2 \sum_{k=1}^{\infty} (k-1) s^{k-2} P_{a,k-1}(t) + \mu \sum_{k=0}^{\infty} (k+1) s^k P_{a,k+1}(t) \\ &\quad - (\lambda + \mu) s \sum_{k=0}^{\infty} k s^{k-1} P_{ak}(t) \\ &= (\lambda s - \mu)(s-1) \frac{\partial G_a(s, t)}{\partial s}, \end{aligned} \quad (7)$$

with the initial condition  $G_a(s, 0) = s^a$ . The solution is

$$G_a(s, t) = \left( \frac{\mu(1-s) - (\mu - \lambda s)e^{-(\lambda-\mu)t}}{\lambda(1-s) + (\mu - \lambda s)e^{-(\lambda-\mu)t}} \right)^a. \quad (8)$$

Inverting and finding the  $b$ th coefficient of the power series  $G_a(s, t)$ , we find the transition probabilities

$$P_{ab}(t) = \sum_{j=0}^{\min(a,b)} \binom{a}{j} \binom{a+b-j-1}{a-1} \alpha^{a-j} \beta^{b-j} (1-\alpha-\beta)^j, \quad (9)$$

where

$$\alpha(t) = \frac{\mu(e^{(\lambda-\mu)t} - 1)}{\lambda e^{(\lambda-\mu)t} - \mu} \quad \text{and} \quad \beta(t) = \frac{\lambda(e^{(\lambda-\mu)t} - 1)}{\lambda e^{(\lambda-\mu)t} - \mu}. \quad (10)$$

As in the Bienaymé-Galton-Watson branching process, the Kendall process gives rise to a linear-fractional distribution (Athreya & Ney, 2004) when starting from one individual ( $a = 1$ ). Sagitov (2013) recently developed a multi-dimensional linear-fractional distribution to characterize the multi-type Bienaymé-Galton-Watson branching process with countably many types. This generalization may also be applicable for some multivariate extensions of the Kendall process.

## GENERAL BDPS

The problem becomes much more complicated for general BDPS. Karlin and McGregor (1957b) present the definitive treatment of the existence of transition probabilities and other properties of BDPS. They obtain the following integral form for the transition probabilities:

$$P_{ab}(t) = \omega_b \int_0^\infty e^{-xt} Q_a(x) Q_b(x) d\psi(x), \quad (11)$$

where  $\dot{E}_0 = 1$  and  $\dot{E}_k = (\lambda_0 \cdots \lambda_{k-1}) / (\mu_1 \cdots \mu_k)$  for  $k \geq 1$ . Here,  $Q_k(x)$ ,  $k = 0, 1, 2, \dots$  is a system of orthogonal polynomials and  $\psi(x)$  is an orthogonalizing spectral measure that are specific to a particular set of birth and death rates.

This integral representation is intuitively satisfying because the time-dependency of  $P_{ab}(t)$  is contained entirely in the exponential term, and  $P_{ab}(t)$  depends on  $Q_a(x)$  and  $Q_b(x)$  in a simple way. In addition, we have the obvious corollary that

$$\frac{P_{ab}(t)}{P_{ba}(t)} = \frac{\omega_b}{\omega_a}. \quad (12)$$

Beyond these simple results related to the interpretation of (11), the formalism developed by Karlin and McGregor (1957b) makes possible deep analytic insight into the behavior of

general BDPs, including recurrence times and first passage times. Notably, a similar spectral representation for the transition probabilities of time-inhomogeneous linear BDPs has been derived recently (Ohkubo, 2014).

## EQUILIBRIUM PROBABILITIES AND EXPLOSION

Equilibrium solutions are straightforward to obtain (Renshaw, 2011). Setting the left-hand side of the Kolmogorov forward equations (5) to zero and replacing the finite-time transition probabilities  $P_{ab}(t)$  with the equilibrium probabilities  $\hat{A}_b$ , we find that

$$\mu_{b+1}\pi_{b+1} - \lambda_b\pi_b = \mu_b\pi_b - \lambda_{b-1}\pi_{b-1}. \quad (13)$$

Since this is the case for every  $b$ , it is true for  $b = 0$  in particular, and  $\mu_0 = \lambda_{-1} = 0$ , so both sides of (13) are zero for every  $b$  by induction. This gives the detailed balance condition for continuous-time Markov chains,

$$\mu_k\pi_k = \lambda_{k-1}\pi_{k-1} \quad \text{for } k = 1, 2, \dots \quad (14)$$

Therefore every general BDP is a reversible Markov chain. Iterating the recurrence (14), we find that

$$\pi_k = \frac{\lambda_0\lambda_1\cdots\lambda_{k-1}}{\mu_1\mu_2\cdots\mu_k}\pi_0, \quad (15)$$

where we have chosen  $\hat{A}_0$  so that  $\sum_k \pi_k = 1$ . Note that  $\hat{A}_k \propto \omega_k$  for every  $k$ .

The birth and death rates for a general BDP may be such that the process “runs away” to infinity in finite time. This is known as explosive growth. Formally, suppose the process begins at  $X(0) = 0$  and there are no absorbing states. Renshaw (2011) shows that the expected first passage time to infinity  $\tau^\infty$  is

$$\mathbb{E}(\tau^\infty | X(0) = 0) = \sum_{j=0}^{\infty} \pi_j \sum_{n=j}^{\infty} \frac{1}{\mu_n\pi_n}, \quad (16)$$

where  $\hat{A}_1 = 1$  and

$$\pi_n = \prod_{k=1}^n \frac{\lambda_{k-1}}{\mu_k}, \quad (17)$$

for  $i > 1$ . When (16) diverges, the process is non-explosive, and the first passage time from 0 to any finite state  $j$  is almost surely finite. When (16) is finite, the first passage time to infinity is finite with non-zero probability.

One result of special interest to us gives the conditions under which a BDP with a given generator  $\mathbf{A}$  is unique: Karlin and McGregor (1957a) show that there is only one transition probability matrix  $\mathbf{P}(t)$  that satisfies (6) if and only if

$$\sum_{k=0}^{\infty} \left( \omega_k + \frac{1}{\lambda_k \omega_k} \right) = \infty. \quad (18)$$

This property assures that probability is conserved on the non-negative integers. We will always assume this is the case in what follows.

Despite the elegant representation (11) for the transition probabilities, it can be very difficult to find the polynomials  $\{Q_k(x)\}$  (Novozhilov et al., 2006; Renshaw, 2011). Settings in which the expression (11) leads to a tractable analytic representations are rare, even for linear processes. As outlined in Crawford and Suchard (2012), Ismail, Letessier, and Valent (1988) give the polynomials  $Q_k(x)$  and measure for the birth-death-immigration-emigration process with  $\lambda_k = k\lambda + \nu$  and  $\mu_k = k\mu + \gamma$ , but a closed-form expression for transition probabilities remains out of reach. In addition, the problem of finding these polynomials and measure  $\psi(x)$  is a fundamentally analytical task, and is generally not amenable to computational solution. In other words, one cannot simply compute  $P_{ab}(t)$  using a computer for an arbitrary set of birth and death rates  $\{\lambda_k\}$  and  $\{\mu_k\}$  using the formula (11) alone.

Since analytic derivation of  $\dot{E}(x)$  is so complicated, Renshaw (2011, page 111) writes of the need for an alternative approach to solving the forward system in order to find transition probabilities for general BDPs:

“A worthwhile and potentially rewarding challenge would be to develop a simplified and user-friendly version of this technique which would work over a wide range of stochastic processes.”

The next section is devoted to this task.

## TRANSITION PROBABILITIES FOR GENERAL BDPS

We now outline a method, first presented in Crawford and Suchard (2012) and based on work by Murphy and O'donohoe (1975), for numerically computing the transition probabilities for a general BDP with arbitrary birth and death rates. To proceed, denote the Laplace transform of  $P_{ab}(t)$  as

$$f_{ab}(s) = L[P_{ab}(t)](s) = \int_0^{\infty} e^{-st} P_{ab}(t) dt. \quad (19)$$



Now, applying the Laplace transform to (5) with  $a = 0$ , we have

$$\begin{aligned} sf_{00}(s) - P_{00}(0) &= \mu_1 f_{01}(s) - \lambda_0 f_{00}(s), \text{ and} \\ sf_{0b}(s) - P_{0b}(0) &= \lambda_{n-1} f_{0,b-1}(s) + \mu_{b+1} f_{0,b+1}(s) - (\lambda_b + \mu_b) f_{0b}(s) \end{aligned} \quad (20)$$

for  $b \in \mathbb{N}$ . Recalling that  $P_{00}(0) = 1$  and  $P_{0b}(0) = 0$  for  $b \in \mathbb{N}$ , we rearrange (20) to find

$$\begin{aligned} f_{00}(s) &= \frac{1}{s + \lambda_0 - \mu_1 \left( \frac{f_{01}(s)}{f_{00}(s)} \right)}, \text{ and} \\ \frac{f_{0b}(s)}{f_{0,b-1}(s)} &= \frac{\lambda_{b-1}}{s + \mu_b + \lambda_b - \mu_{b+1} \left( \frac{f_{0,b+1}(s)}{f_{0b}(s)} \right)}. \end{aligned} \quad (21)$$

By combining these recurrence relations, we obtain the generalized continued fraction

$$f_{00}(s) = \frac{1}{s + \lambda_0 - \frac{\lambda_0 \mu_1}{s + \lambda_1 + \mu_1 - \frac{\lambda_1 \mu_2}{s + \lambda_2 + \mu_2 - \dots}}}, \quad (22)$$

that is an exact expression for the Laplace transform of the transition probability  $P_{00}(t)$  (Bordes & Roehner, 1983; Flajolet & Guillemin, 2000; Guillemin & Pinchon, 1999; Karlin & McGregor, 1957b). Now define  $a_1 = 1$ ,  $a_n = -\lambda_{n-2} \mu_{n-1}$ ,  $b_1 = s + \lambda_0$  and  $b_n = s + \lambda_{n-1} + \mu_{n-1}$  for  $n \geq 2$ . Then (22) becomes

$$f_{00}(s) = \frac{a_1}{b_1} \frac{a_2}{b_2} \frac{a_3}{b_3} \dots \quad (23)$$

in more concise notation. We denote the  $k$ th convergent of the Laplace transform  $f_{00}(s)$  by

$$f_{00}^{(k)}(s) = \frac{a_1}{b_1} \frac{a_2}{b_2} \dots \frac{a_k}{b_k} = \frac{A_k(s)}{B_k(s)}. \quad (24)$$

The main result of Crawford and Suchard (2012) is the following theorem giving continued fraction expressions for the Laplace transform of the transition probability in a general birth-death process. Theorem. 1 The Laplace transform of the transition probability  $P_{ab}(t)$  is given by

$$f_{ab}(s) = \begin{cases} \left( \prod_{j=b+1}^a \mu_j \right) \frac{B_b(s)}{B_{a+1}(s) +} \frac{B_a(s)a_{a+2}}{b_{a+2} +} \frac{a_{a+3}}{b_{a+3} +} \dots \text{for } b \leq a, \\ \left( \prod_{j=a}^{b-1} \lambda_j \right) \frac{B_a(s)}{B_{b+1}(s) +} \frac{B_b(s)a_{b+2}}{b_{b+2} +} \frac{a_{b+3}}{b_{b+3} +} \dots \text{for } a \leq b, \end{cases} \quad (25)$$

where  $a_n$ ,  $b_n$ , and  $B_n$  are as defined above. The proof of this theorem relies on elementary manipulation of the continued fraction recurrences (21).

## COMPUTATION OF TRANSITION PROBABILITIES

The Laplace transforms (25) usually cannot be inverted analytically to obtain time-domain transition probabilities. However, the representation (25) has several desirable properties for computational inversion. First, infinite continued fraction representations often converge much faster than their corresponding power series (Bankier & Leighton, 1942; Wall, 1948). Second, highly efficient numerical methods exist for evaluating continued fractions to a specified depth (Blanch, 1964; Lange, 2010b; Lorentzen & Waadeland, 1992; Wallis, 1972). Third, stable numerical inversion of Laplace transforms is well established, and methods exist for bounding the discretization and approximation error (Abate & Whitt, 1999; Cravittot, Jones, & Thron, 1993; Cuyt, Petersen, Verdonk, Waadeland, & Jones, 2008). For these reasons, Crawford and Suchard (2012) obtain time-domain transition probabilities  $P_{ab}(t)$  from (33) by numerically inverting the Laplace transforms. Error bounds for numerical Laplace inversion are derived in Crawford and Suchard (2012) and Crawford et al. (2014). Alternatively, when the state space of a BDP is finite, numerical computation of transition probabilities can sometimes be accomplished by matrix methods. For a BDP that takes values on  $\{0, \dots, N\}$ , consider the  $N+1 \times N+1$  stochastic transition rate matrix  $Q$  whose elements are

$$Q = \begin{pmatrix} -\lambda_0 & \lambda_0 & & & & \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & & & \\ & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \mu_N & -(\lambda_N + \mu_N) & \lambda_N \end{pmatrix}. \quad (26)$$

When the eigendecomposition  $Q = UDU^{-1}$  exists, where  $U$  is orthogonal and  $D$  is diagonal, the matrix of transition probabilities  $P(t)$  satisfies

$$\frac{dP}{dt} = PQ \quad (27)$$

with  $P(0) = I$ . Then the time-domain transition probabilities can be obtained by matrix exponentiation,

$$P(t) = \exp[Qt] = Ue^{tD}U^{-1}. \quad (28)$$

When the state space is large or infinite, it is sometimes possible to truncate the state space at a suitably large index  $N$  and compute transition probabilities using (28). When the decomposition exists and can be found in a numerically stable way, Crawford, Stutz, and Lange (2016) give coupling bounds for finding an appropriate truncation index to control the truncation error.

## FIRST PASSAGE TIMES

Now consider the time of first arrival of a BDP  $X(t)$  into an arbitrary set  $S$  of taboo states, and suppose  $X(0) = i \in \mathbb{N} \setminus S$ . This first passage time is defined formally as

$$\tau_i = \inf \{t : X(t) \in S \mid X(0) = i\}. \quad (29)$$

To find the relationship between first passage times and the expressions for transition probabilities discussed above, construct a new process  $Y(t)$  identical to  $X(t)$  except that  $\mu_j = 0$  for every  $j \in S$ , so every state in  $S$  is absorbing. Then for this modified process, with  $P_{ij}(t) = \Pr(Y(t) = j \mid Y(0) = i)$ ,

$$\Pr(\tau_i < t) = \sum_{j \in S} P_{ij}(t). \quad (30)$$

The intuitive reason for this equality is the absorbing nature of the states in  $S$ : if  $Y$  reaches an absorbing state  $j \in S$  at any time before  $t$ , we must have  $Y(t) = j$ . Furthermore,  $Y$  cannot visit more than one state in  $S$ , so the absorption events are mutually exclusive and the probability of absorption is simply the sum of the individual absorption probabilities. Therefore the cumulative distribution function of the first passage time into  $S$  is given by the sum of the transition probabilities from  $i$  to every taboo state in  $S$  for the modified process  $Y(t)$ .

## LIKELIHOODS

One factor hindering more widespread adoption of BDPs by applied researchers is the difficulty in performing statistical estimation of the unknown parameters in a BDP using real-world data (Doss, Suchard, Holmes, Kato-Maeda, & Minin, 2013; Holmes & Bruno, 2001). Typically efforts in estimation for BDPs have been limited to continuous observation of the process (Anscombe, 1953; Darwin, 1956; P. Moran, 1951, 1953; Reynolds, 1973; Wolff, 1965). In addition, much work to date has focused on the simple linear BDP because it is analytically tractable (Dauxois, 2004; Keiding, 1975; Rosenberg et al., 2003; Thorne et al., 1991), though important progress has been made in analysis of nonlinear processes (Karev et al., 2002; Klar et al., 2010; Reed & Hughes, 2004). In practice researchers often observe data from BDPs only at

discrete times through longitudinal sampling. In addition, the simple linear BDP may be unappealing because it fails to capture more complicated dynamics of population growth and decay that arise when particles do not behave independently. To learn from discretely-observed general BDPs, we will need more advanced statistical tools.

## LIKELIHOOD FOR THE CONTINUOUSLY-OBSERVED PROCESS

In a discretely-observed general BDP, the likelihood cannot be written in closed form, making analytic maximum likelihood estimation impossible. However, the likelihood of a continuously-observed BDP is straightforward to express (Keiding, 1975; Reynolds, 1973). To develop the likelihood for continuously-observed data from a general BDP, we note the following important fact: the exponentially distributed waiting time of a continuous-time Markov process in a certain state is independent of the destination of the next jump (Lange, 2010a). Recall that the waiting time  $W$  for the first event to occur from state  $k$  is exponentially distributed with rate  $\lambda_k + \mu_k$ . If the waiting time in the current state  $k$  is  $W = \tau$ , and the next change is a birth,

$$\begin{aligned} \Pr(W = \tau, \text{birth} | X(0) = k) &= \Pr(W = \tau | X(0) = k) \Pr(\text{birth} | X(0) = k) \\ &= (\lambda_k + \mu_k) e^{-(\lambda_k + \mu_k)\tau} \left( \frac{\lambda_k}{\lambda_k + \mu_k} \right) \\ &= \lambda_k e^{-(\lambda_k + \mu_k)\tau}. \end{aligned} \quad (31)$$

Likewise, the probability of a waiting time  $W = \tau$  followed by a death is

$$\Pr(W = \tau, \text{death} | X(0) = k) = \mu_k e^{-(\lambda_k + \mu_k)\tau}. \quad (32)$$

Since we can only observe the process for a finite time  $t$ , the last observation will be the waiting time in some state  $k$  from the time of the jump to  $k$  to the end of observation. Using the same reasoning,

$$\Pr(W \geq \tau | \text{no births or deaths}, X(0) = k) = e^{-(\lambda_k + \mu_k)\tau}. \quad (33)$$

To write the likelihood of a continuously-observed BDP from time 0 to  $t$ , we introduce some notation to ease our presentation. Suppose we observe  $n$  jumps in the time interval  $(0, t)$ , and label the jumps  $i = 1, \dots, n$ . Let  $W_i$  be the waiting time in the current state just before the  $i$ th jump. Define the indicator  $B_i = 1$  if the  $i$ th jump is a birth, and  $B_i = 0$  if the  $i$ th jump is a death. Let  $t_1, \dots, t_n$  be the times of the  $n$  jumps, with  $t_0 = 0$  and  $t_n < t$ . Then the likelihood of a sequence of observations  $\mathbf{Y} = \{X(\check{A}), 0 < \check{A} < t\}$  is

$$\begin{aligned}
 L &= \prod_{i=1}^n \left[ \Pr(W_i = t_i - t_{i-1} \mid X(t_{i-1})) \right. \\
 &\quad \times \Pr(\text{birth} \mid X(t_{i-1}))^{B_i} \Pr(\text{death} \mid X(t_{i-1}))^{1-B_i} \left. \right] \\
 &\quad \times \Pr(W_{n+1} = t - t_n \mid \text{no births or deaths}, X(t_n)) \\
 &= \prod_{i=1}^n (\lambda_{X(t_{i-1})} + \mu_{X(t_{i-1})}) \exp[-(\lambda_{X(t_{i-1})} + \mu_{X(t_{i-1})})(t_i - t_{i-1})] \\
 &\quad \times \left( \frac{\lambda_{X(t_{i-1})}^{B_i} \mu_{X(t_{i-1})}^{1-B_i}}{\lambda_{X(t_{i-1})} + \mu_{X(t_{i-1})}} \right) \times \exp[-(\lambda_{X(t_n)} + \mu_{X(t_n)})(t - t_n)] \\
 &= \prod_{i=1}^n \lambda_{X(t_{i-1})}^{B_i} \mu_{X(t_{i-1})}^{1-B_i} \exp[-(\lambda_{X(t_{i-1})} + \mu_{X(t_{i-1})})(t_i - t_{i-1})] \\
 &\quad \times \exp[-(\lambda_{X(t_n)} + \mu_{X(t_n)})(t - t_n)], \tag{34}
 \end{aligned}$$

where  $X(t_{i-1})$  is the state just before the  $i$ th jump. This cumbersome notation can be eliminated if we instead keep track of the total waiting time in each state and the number of births and deaths from each state. Define  $\{E\}$  to be the indicator of an event  $E$ , and let

$$T_k = \sum_{i=1}^n (t_i - t_{i-1}) \{X(t_{i-1}) = k\} \tag{35}$$

be the total time spent in state  $k$  over all visits to  $k$ . Then let

$$U_k = \sum_{i=1}^n \{X(t_{i-1}) = k, B_i = 1\} \tag{36}$$

be the number of up steps (births) from state  $k$ , and let

$$D_k = \sum_{i=1}^n \{X(t_{i-1}) = k, B_i = 0\} \tag{37}$$

be the number of down steps (deaths) from state  $k$ . Then we can re-write the likelihood (34) in much simpler and more transparent form as

$$L = \prod_{k=0}^{\infty} \lambda_k^{U_k} \mu_k^{D_k} \exp[-(\lambda_k + \mu_k)T_k]. \tag{38}$$

Of course, in a BDP observed continuously for a finite time (for which (18) holds), there are only finitely many jumps observed, so the product above is not really infinite in practice.

Equation (38) also reveals that the likelihood for a continuously-observed BDP is a member of the exponential family, where  $\{U_k\}$ ,  $\{D_k\}$ , and  $\{T_k\}$  for  $k = 0, 1, \dots$  are the sufficient statistics of the continuously-observed BDP likelihood. In other words, one only needs to know the total number of up and down steps from, and time spent in, each state  $k$  visited by the process in order to compute the likelihood.

## EXAMPLE: CONTINUOUSLY-OBSERVED KENDALL PROCESS

Maximum likelihood estimation for continuously-observed BDPs is often straightforward. Consider the simple linear BDP with birth rate  $\lambda_k = \lambda$  and death rate  $\mu_k = \mu$ . The likelihood (38) of a single observation, up to a normalizing constant, becomes

$$L \propto \lambda^U \mu^D \exp \left[ -(\lambda + \mu) \int_0^t X(\tau) d\tau \right], \quad (39)$$

where  $U = \sum_k U_k$  is the total number of up steps (births),  $D = \sum_k D_k$  is the total number of down steps (deaths) during the interval  $(0, t)$ , and

$$\int_0^t X(\tau) d\tau = \sum_{k=0}^{\infty} kT_k \quad (40)$$

is the “total particle time” or total time lived by every particle that existed during the interval  $(0, t)$ . Maximizing (39) with respect to the unknown parameters  $\lambda$  and  $\mu$ , we obtain the maximum likelihood estimators

$$\hat{\lambda} = \frac{U}{\int_0^t X(\tau) d\tau} \quad \text{and} \quad \hat{\mu} = \frac{D}{\int_0^t X(\tau) d\tau}, \quad (41)$$

first given by Reynolds (1973). Although the estimators provided by (41) involve an integral over the state path of the process, the integrand is simply a step function that is fully observed over  $(0, t)$ .

## LIKELIHOOD FOR THE DISCRETELY-OBSERVED PROCESS

Suppose now that the process  $X(t)$  is observed only discretely, once at time 0 and again at time  $t$ , without loss of generality owing to the Markov assumption. Let us label the state of the BDP at these times as  $X(0) = a$  and  $X(t) = b$ . Then given that  $X(0) = a$ , the probability that  $X(t) = b$  is the transition probability  $P_{ab}(t)$ . Above we outlined a method for numerically computing this probability for any general BDP. If we regard the transition probability  $P_{ab}(t)$  as a function of some unknown parameters  $\theta$ , which control the birth and death rates, writing  $P_{ab}(t | \theta)$ , then we have the *likelihood* of our observation,

$$L(\theta) = P_{ab}(t | \theta). \quad (42)$$

In principle, we could numerically maximize the likelihood for discrete observations to find an estimate of  $\theta$ . However, as the number of parameters increases, naïve numerical optimization often suffers from poor convergence (Doss et al., 2013). The difficulty in writing or computing the likelihood for discrete observations from BDPs has limited the usefulness of BDPs in applications.

In contrast to the appealing analytic characterization (38) of the continuously-observed process likelihood, the discretely-observed process is hard to characterize. To bridge this gap, it is helpful to view computation of the likelihood in the discretely-observed process as a missing data problem. When a BDP is observed discretely, we do not know the sufficient statistics  $\{U_k, D_k, T_k\}_{k=0}^{\infty}$ . This perspective suggests that we exploit analytic information about these statistics, conditional on the start and end states of the observed process.

## EM ALGORITHMS FOR MAXIMUM LIKELIHOOD ESTIMATION

In this section, we review the estimation machinery developed by Crawford et al. (2014) for maximum likelihood or maximum *a posteriori* estimation in BDPs. When a BDP is discretely sampled,  $U_k$ ,  $D_k$ , and  $T_k$  are unobserved for every  $k$ ; we cannot maximize the likelihood without knowing these statistics. We therefore appeal to the expectation-maximization (EM) algorithm for iterative maximum likelihood estimation with missing data (Dempster, Laird, & Rubin, 1977). When the incomplete data likelihood is intractable but the complete data likelihood has a simple form, the EM algorithm operates by replacing each missing datum by a conditional expectation as follows. If  $X$  is the complete (unobserved) data,  $Y$  represents the incomplete (observed) data, and  $\ell(\theta | X)$  is the complete data log-likelihood, we form a surrogate function  $Q$  as the expectation of the complete data likelihood, conditional on the observed data  $Y$  and the current ( $m$ th) parameter iterate:

$$Q(\theta | \theta^{(m)}) = \mathbb{E}(\ell(\theta | X) | Y = y, \theta^{(m)}). \quad (43)$$

This is the E-step of the EM algorithm, and it accomplishes a minorization of  $\ell(\theta | X)$  at  $\theta^{(m)}$ . The M-step maximizes (or takes a step toward the maximum of)  $Q$ . By alternating these steps

— minorizing  $Q$  by  $Q$ , then finding  $\theta$  that increases  $Q$  — the EM algorithm drives succeeding iterates toward the MLE.

Taking the expectation of the logarithm of (38), conditional on the observed data  $Y = (X(0) = a, X(t) = b, t)$  and the current parameter estimate  $\theta^{(m)}$ , we write the surrogate function for the BDP as follows:

$$\begin{aligned} Q(\theta | \theta^{(m)}) &= \mathbb{E}[\ell(\theta | Y, \theta^{(m)})] \\ &= \sum_{k=0}^{\infty} \mathbb{E}(U_k | Y) \log[\lambda_k(\theta)] + \mathbb{E}(D_k | Y) \log[\mu_k(\theta)] \quad (44) \\ &\quad - \mathbb{E}(T_k | Y) [\lambda_k(\theta) + \mu_k(\theta)], \end{aligned}$$

In the above equation and many that follow, we omit the dependence of the conditional expectations on  $\theta^{(m)}$  from the  $m$ th iterate for visual clarity.

To calculate the conditional expectations necessary for the E-step of the EM algorithm, we appeal to the following integral expressions

$$\mathbb{E}(U_k | Y) = \frac{\int_0^t P_{ak}(\tau) \lambda_k P_{k+1,b}(t - \tau) d\tau}{P_{ab}(t)}, \quad (45a)$$

$$\mathbb{E}(D_k | Y) = \frac{\int_0^t P_{ak}(\tau) \mu_k P_{k-1,b}(t - \tau) d\tau}{P_{ab}(t)}, \quad \text{and } (45b)$$

$$\mathbb{E}(T_k | Y) = \frac{\int_0^t P_{ak}(\tau) P_{kb}(t - \tau) d\tau}{P_{ab}(t)}. \quad (45c)$$

These expressions have appeared repeatedly in literature on inference for discretely-observed continuous-time Markov chains (Bladt & Sorensen, 2005; Hobolth & Jensen, 2005; Holmes & Rubin, 2002; Lange, 1995; Metzner, Dittmer, Jahnke, & Schütte, 2007). When the process takes only finitely many states, matrix solutions are possible using the uniformization method (Neuts, 1995). Hobolth and Stone (2009) develop efficient Monte Carlo methods using simulation conditioned on the start and end points of the discrete observation  $Y$ . Finally, Doss et al. (2013) study a linear BDP on an infinite state space and derive the expectations analytically using a generating function argument. None of the exact methods is a general approach for arbitrary BDPs on an infinite state space. The Monte Carlo approaches, while not reliant on a particular parameterization of the process, can suffer from poor performance when observed realizations occur with low probability. The lack of a reliable method for computing the E-step of the EM algorithm for discretely-observed BDPs has hindered progress on statistical inference for these processes.



An alternative approach introduced by Crawford et al. (2014) takes advantage of the Laplace transforms  $f_{ab}(s)$  of the transition probabilities (25). The numerators in (45) are time-domain convolutions of transition probabilities. The functional form of these expressions suggests using the Laplace convolution property to obtain

$$\mathbb{E}(U_k | Y) = \lambda_k \frac{\mathcal{L}^{-1}[f_{ak}(s) f_{k+1,b}(s)](t)}{P_{ab}(t)}, \quad (46a)$$

$$\mathbb{E}(D_k | Y) = \mu_k \frac{\mathcal{L}^{-1}[f_{ak}(s) f_{k-1,b}(s)](t)}{P_{ab}(t)}, \quad \text{and} \quad (46b)$$

$$\mathbb{E}(T_k | Y) = \frac{\mathcal{L}^{-1}[f_{ak}(s) f_{kb}(s)](t)}{P_{ab}(t)}, \quad (46c)$$

where  $\mathcal{L}^{-1}[\cdot]$  denotes inverse Laplace transformation. These expressions are formally equivalent to (45), but they offer substantial computational time savings over numerical integration of (45), and make possible efficient computation of conditional expectations for EM algorithms for any BDP (Crawford et al., 2014).

We now show how to complete the M-step for several BDP models. The first two, variations on the simple linear (Kendall) process, were given in Crawford et al. (2014). The others are novel, yet remarkably easy to derive and implement computationally. In each case, we describe the surrogate likelihood function  $Q(\theta | \theta^{(m)})$  and give the M-step updates for each unknown parameter.

## EXAMPLE: DISCRETELY-OBSERVED KENDALL PROCESS

In the simple linear BDP, births and deaths happen at constant per-particle rates, so  $\lambda_k = \lambda$  and  $\mu_k = k\mu$ . The unknown is  $\lambda, \mu$ . The surrogate function  $Q$  becomes

$$Q(\theta) = \sum_{k=0}^{\infty} \mathbb{E}(U_k | Y) \log[k\lambda] + \mathbb{E}(D_k | Y) \log[k\mu] - \mathbb{E}(T_k | Y) k(\lambda + \mu). \quad (47)$$

Maximizing (47) with respect to the  $\lambda, \mu$  yields the updates:

$$\lambda^{(m+1)} = \frac{\mathbb{E}(U | Y)}{\mathbb{E}(T_{\text{particle}} | Y)} \quad \text{and} \quad (48a)$$

$$\mu^{(m+1)} = \frac{\mathbb{E}(D | Y)}{\mathbb{E}(T_{\text{particle}} | Y)}, \quad (48b)$$

where

$$T_{\text{particle}} = \int_0^t X(\tau) d\tau, \quad (49)$$

and we have again suppressed the dependence of the conditional expectations on  $\theta^{(m)}$  for typographic clarity. These expressions are identical in form to the estimators given in (41), but are instead iterative updates in the EM algorithm.

## EXAMPLE: LINEAR BDP WITH IMMIGRATION

The linear BDP with immigration is similar to the simple linear BDP, but there is a source of new arrivals whose rate is constant and does not depend on the number of particles already in existence. This yields the birth and death rates  $\lambda_k = k\lambda + \nu$  and  $\mu_k = k\mu$ . The log-likelihood becomes

$$\ell(\theta) = \sum_{k=0}^{\infty} U_k \log(k\lambda + \nu) + D_k \log(\mu) - T_k[k(\lambda + \mu) + \nu]. \quad (50)$$

Unfortunately, it is difficult to maximize the resulting surrogate function analytically. But since each term in the sum is a concave function of the unknown parameters, we can separate them in a second minorizing function  $H$  such that for all  $\theta$ ,  $H(\theta | \theta^{(m)}) \leq \ell(\theta)$  and

$H(\theta^{(m)} | \theta^{(m)}) = \ell(\theta^{(m)})$ . To accomplish the minorization, note that

$$\begin{aligned} \log(k\lambda + \nu) &\geq \frac{k\lambda^{(m)}}{k\lambda^{(m)} + \nu^{(m)}} \log\left[\frac{k\lambda^{(m)} + \nu^{(m)}}{\lambda^{(m)}} \lambda\right] \\ &\quad + \frac{\nu^{(m)}}{k\lambda^{(m)} + \nu^{(m)}} \log\left[\frac{k\lambda^{(m)} + \nu^{(m)}}{\nu^{(m)}} \nu\right]. \end{aligned} \quad (51)$$

We form a minorizing log-likelihood function  $H$  as follows:

$$\begin{aligned} H(\theta | \theta^{(m)}) &\propto \sum_{k=0}^{\infty} U_k \left[ p_k^{(m)} \log(\lambda) + (1 - p_k^{(m)}) \log(\nu) \right] + D_k \log(\mu) \\ &\quad - [k(\lambda + \mu) + \nu] T_k, \end{aligned} \quad (52)$$

where

$$p_k^{(m)} = \frac{k\lambda^{(m)}}{k\lambda^{(m)} + v^{(m)}}. \quad (53)$$

Exploiting this surrogate function and maximizing with respect to the unknown sufficient statistics gives the updates

$$\lambda^{(m+1)} = \frac{\sum_{k=0}^{\infty} p_k \mathbb{E}(U_k | Y)}{\mathbb{E}(T_{\text{particle}} | Y)}, \text{ and} \quad (54a)$$

$$v^{(m+1)} = \frac{\sum_{k=0}^{\infty} (1 - p_k) \mathbb{E}(U_k | Y)}{t}. \quad (54b)$$

The update for  $\mu$  is the same as (48b).

## EXAMPLE: PURE-BIRTH AND GENERALIZED POISSON PROCESSES

Recall that the Poisson process with arrival rate  $\lambda$  is a BDP with  $\lambda_k = \lambda$ ,  $\mu_k = 0$  for all  $k$ . Many researchers have found that real-world count data are sometimes over- or under-dispersed relative to the Poisson distribution. Statisticians seeking a more flexible distribution for count outcomes that can accommodate over- and under-dispersion have arrived at several alternative distributions. A notable example that fits neatly into the BDP framework is the general pure-birth process with arbitrary birth rates  $\lambda_k$ ,  $k = 0, 1, \dots$ , and  $\mu_k = 0$  for all  $k$ . This class of processes has an appealing property: it can recover any discrete probability distribution on the counting numbers by appropriately setting the birth rates (M. Faddy & Bosch, 2001; M. J. Faddy, 1997). Crawford and Zelteman (2015) show that any such pure-birth process can be represented as a sum of exchangeable Bernoulli random variables, a result that connects BDPs with phenomenological models often used for dependent outcomes in toxicology and epidemiology. Renshaw (2011, page 65 gives an analytic form for these transition probabilities

$$P_{ab}(t) = \left( \prod_{k=a}^{b-1} \lambda_k \right) \sum_{k=a}^b \left( \prod_{\ell \neq k} (\lambda_{\ell} - \lambda_k) \right)^{-1} \exp[-\lambda_k t] \quad (55)$$

for  $0 \leq a \leq b$  and  $t > 0$  provided that  $\lambda_i \neq \lambda_j$  for all  $i$  and  $j$ . While (55) has an appealing form, it depends on none of the birth rates being equal. Another potentially serious drawback is that it can be numerically troublesome to compute; the summands may be alternating in sign and the product of small differences in the denominators can lead to serious roundoff error. In many scenarios, especially when some observed counts are large and some  $\lambda_k$ 's are nearly or exactly equal, (55) provides an unappealing way to compute the likelihood. Exactly equal  $\lambda_k$  may arise, for example, when entertaining a Bayesian non-parametric prior. Fortunately, the

EM framework does not require use of (55). We now provide an example of a pure birth process intended to generalize the Poisson distribution to accommodate over- and under-dispersion.

M. J. Faddy (1997) describes a class of pure-birth BDPs with  $\lambda_k = \lambda(\gamma+k)^c$  and  $\mu_k = 0$ , where  $c = 0$  corresponds to a Poisson process with rate  $\lambda$ ,  $c > 0$  results in overdispersion relative to Poisson, and  $c < 0$  results in underdispersion. The log-likelihood for the continuously-observed process beginning at  $X(0) = a$  and ending at  $X(t) = b$  is

$$\ell(\theta) = \sum_{k=a}^b 1\{k < b\} [\log(\lambda) + c \log(\gamma + k)] - \lambda(\gamma + k)^c T_k. \quad (56)$$

Letting  $\theta = (\lambda, \gamma, c)$ , the surrogate function is

$$Q(\theta) = \sum_{k=a}^b 1\{k < b\} [\log(\lambda) + c \log(\gamma + k)] - \lambda(\gamma + k)^c \mathbb{E}(T_k | \mathbf{Y}). \quad (57)$$

The update for  $\lambda$  is given by

$$\lambda^{(m+1)} = \frac{b-a-1}{\sum_{k=a}^b (\gamma + k)^c \mathbb{E}(T_k | \mathbf{Y})}, \quad (58)$$

but the updates for  $\gamma$  and  $c$  are not available in closed form. However, Lange (1995) shows that one step of a gradient ascent algorithm suffices to preserve the ascent property of the EM algorithm. Therefore a Newton-Raphson update can be derived, and

$$\begin{pmatrix} \gamma^{(m+1)} \\ c^{(m+1)} \end{pmatrix} = \mathbf{d}^2 Q(\gamma^{(m)}, c^{(m)})^{-1} \nabla Q(\gamma^{(m)}, c^{(m)}), \quad (59)$$

where  $\nabla Q$  and  $\mathbf{d}^2 Q$  are the gradient and Hessian of  $Q$  with respect to  $\gamma$  and  $c$  respectively.

## EXAMPLE: MORAN MODEL

The Moran process models genetic drift in a finite population by keeping track of the number of alleles of a certain type at a biallelic locus in a haploid population of constant size  $N < \infty$ . Call the two alleles  $A$  and  $B$ , and suppose we wish to keep track of the number of  $A$  carriers in the population. In the Moran model with selection, carriers of  $A$  have fitness  $\pm$ , and carriers of  $B$  have fitness  $^2$ . For the sake of identifiability in a statistical setting, we specify  $^2 = 1$  and let  $\pm$  denote the relative fitness of  $A$  carriers over  $B$  carriers. Furthermore,  $A$  mutates to  $B$  in one

generation with probability  $u$ , and *vice versa* with probability  $v$ . When an existing individual dies, a new allele is drawn at random. The birth and death rates are

$$\begin{aligned}\lambda_n &= \frac{N-n}{N} \left[ \alpha \frac{n}{N} (1-u) + \frac{N-n}{N} v \right], \text{ and} \\ \mu_n &= \frac{n}{N} \left[ \frac{N-n}{N} (1-v) + \alpha \frac{n}{N} u \right]\end{aligned}\quad (60)$$

for  $n = 0, \dots, N$ . Forming the surrogate function from (44), we see that maximizing the log-likelihood with respect to the unknowns  $\alpha$ ,  $u$ , and  $v$  is difficult. However, we can again construct a minorizing function to separate the parameters in the logarithm terms. We minorize the birth rate as

$$\begin{aligned}\log(\lambda_n) &\propto \log[n\alpha(1-u) + (N-n)v] \\ &\geq p_n^{(m)} \log\left(\frac{n\alpha^{(m)}(1-u^{(m)}) + (N-n)v^{(m)}}{\alpha^{(m)}(1-u^{(m)})} \alpha(1-u)\right) \\ &\quad + (1-p_n^{(m)}) \log\left(\frac{n\alpha^{(m)}(1-u^{(m)}) + (N-n)v^{(m)}}{v^{(m)}} v\right) \\ &\propto p_n^{(m)} (\log(\alpha) + \log(1-u)) + (1-p_n^{(m)}) \log(v),\end{aligned}\quad (61)$$

where

$$p_n^{(m)} = \frac{n\alpha^{(m)}(1-u^{(m)})}{n\alpha^{(m)}(1-u^{(m)}) + (N-n)v^{(m)}}. \quad (62)$$

Although (61) and (62) may appear complicated, this minorization has the effect of separating the parameters  $\alpha$  and  $u$  in the surrogate function, allowing closed-form updates. In a similar way, we minorize the log-death rate  $\log(\mu_n)$  as

$$q_n^{(m)} \log(1-v) + (1-q_n^{(m)}) (\log(\alpha) + \log(u)), \quad (63)$$

where

$$q_n^{(m)} = \frac{(N-n)(1-v^{(m)})}{(N-n)(1-v^{(m)}) + n\alpha^{(m)}u^{(m)}}. \quad (64)$$

We form the complete minorizing function  $H$  as

$$\begin{aligned}
 H(\theta) = & \sum_{k=0}^N B_k \left[ p_k^{(m)} (\log(\alpha) + \log(1-u)) + (1-p_k^{(m)}) \log(v) \right] \\
 & + D_k \left[ q_k^{(m)} \log(1-v) + (1-q_k^{(m)}) (\log(\alpha) + \log(u)) \right] \\
 & - \frac{T_k}{N^2} \left[ (N-k)k\alpha(1-u) + (N-k)^2v + (N-k)k(1-v) + k^2\alpha u \right],
 \end{aligned} \tag{65}$$

and the surrogate function is  $Q(\theta) = \mathbb{E}(H(\theta) | Y, \theta^{(m)})$ . A simple way to proceed is to find updates for each of the unknowns individually, conditional on the previous ( $m$ th) estimate of the others, giving a cyclic coordinate ascent algorithm. The update for  $\alpha$  is

$$\alpha^{(m+1)} = \frac{\sum_{k=0}^N p_k^{(m)} B_k + (1-q_k^{(m)}) D_k}{\frac{1}{N^2} \sum_{k=0}^N T_k \left[ (N-k)k(1-u^{(m)}) + k^2 u^{(m)} \right]}. \tag{66}$$

The update for  $u$  is the positive solution of the quadratic equation

$$\begin{aligned}
 0 = & \sum_{k=0}^N -u B_k p_k^{(m)} + (1-u) D_k (1-q_k^{(m)}) \\
 & - u(1-u) \frac{T_k}{N^2} \left[ k^2 \alpha^{(m)} - (N-k)k \alpha^{(m)} \right],
 \end{aligned} \tag{67}$$

when  $0 < u < 1$ . The update for  $v$  is obtained by similar manipulations.

## EXAMPLE: MAXIMUM *A POSTERIORI* ESTIMATION FOR THE KENDALL PROCESS

In a Bayesian setting, a prior distribution  $f(\cdot)$  on the unknown parameters  $\cdot$  is given, and we seek to maximize the log-posterior distribution of the parameters, given the data,

$\Pr(\theta | Y) \propto \Pr(Y | \theta) f(\theta)$  to obtain the maximum *a posteriori* (MAP) estimate of  $\cdot$ . Here the surrogate function becomes  $Q(\theta | \theta^{(m)}) = \mathbb{E}(\ell(\theta) | Y, \theta^{(m)}) + \log[f(\theta)]$ . To illustrate, suppose that independent observations from a BDP follow the simple linear model, and we believe that  $\lambda$  and  $\mu$  are *a priori* independent and are Gamma-distributed:

$$\lambda \sim \text{Gamma}(k_\lambda, \beta_\lambda) \quad \text{and} \quad \mu \sim \text{Gamma}(k_\mu, \beta_\mu). \tag{68}$$

Then the unknowns are  $\cdot = (\lambda, \mu)$  and the log-prior for  $\cdot$  is

$$\log f(\theta) \propto (k_\lambda - 1) \log(\lambda) + (k_\mu - 1) \log(\mu) - \beta_\lambda \lambda - \beta_\mu \mu \tag{69}$$

Ignoring irrelevant terms, the surrogate function becomes

$$Q(\theta | \theta^{(m)}) = \mathbb{E}(U | Y) \log(\lambda) + \mathbb{E}(D | Y) \log(\mu) - \mathbb{E}(T_{\text{particle}} | Y)(\lambda + \mu) \\ + (k_\lambda - 1) \log(\lambda) + (k_\mu - 1) \log(\mu) - \beta_\lambda \lambda - \beta_\mu \mu \quad (70)$$

The MAP updates are

$$\lambda^{(m+1)} = \frac{\mathbb{E}(U | Y) + k_\lambda - 1}{\mathbb{E}(T_{\text{particle}} | Y) + \beta_\lambda}, \quad \text{and} \quad (71a)$$

$$\mu^{(m+1)} = \frac{\mathbb{E}(D | Y) + k_\mu - 1}{\mathbb{E}(T_{\text{particle}} | Y) + \beta_\mu}. \quad (71b)$$

## EXAMPLE: REGRESSION FOR COUNT DATA

Perhaps the most interesting use of EM algorithms for BDP inference is to provide a unified framework for regression estimation. To illustrate, consider a collection of  $n$  independent BDPs,  $X^i(t)$  with  $\lambda_k^i = \exp[\mathbf{Z}_i \beta]$  and  $\mu_k^i = 0$  for  $i = 1, \dots, n$ , where  $\mathbf{Z}_i$  is a  $d \times 1$  vector of covariates and  $\beta$  is a covariate vector of corresponding dimension and  $\mu_k = 0$  for all  $k$ . Then letting  $X^i(0) = 0$  and  $X^i(1) = x_i$  for each  $i$ , the log-likelihood becomes

$$\ell(\beta) = \sum_{i=1}^n x_i \mathbf{Z}_i^T \beta - \exp[\mathbf{Z}_i^T \beta]. \quad (72)$$

This is the log-likelihood for classical Poisson regression, and updates are found using a Newton-Raphson step (Dobson, 2001).

It is possible to formulate an analogous model for the Kendall process. Let  $\lambda_k^i = k \exp[\mathbf{Z}_i^T \beta]$  and let  $\mu_k^i = k \exp[\mathbf{W}_i^T \gamma]$  be the birth and death rates of a BDP  $X^i(t)$ . The log-likelihood is

$$\ell(\beta, \gamma) \propto \sum_{i=1}^n \sum_{k=0}^{\infty} U_k^i \mathbf{Z}_i^T \beta + D_k^i \mathbf{W}_i^T \gamma - k T_k^i (\exp[\mathbf{Z}_i^T \beta] + \exp[\mathbf{W}_i^T \gamma]), \quad (73)$$

where the statistics  $U_k^i$ ,  $D_k^i$ , and  $T_k^i$  correspond to observation  $i$ . When the process is discretely-observed, we form the surrogate as before, and find the gradient vector

$$\nabla_{\beta} Q = \sum_{i=1}^n \mathbb{E}[U^i | Y_i] \mathbf{Z}_i - \mathbb{E}[T_{\text{particle}}^i | Y_i] \mathbf{Z}_i \exp[\mathbf{Z}_i^T \beta] \quad (74)$$

for  $\beta$ . The Hessian matrix is

$$d_{\beta}^2 Q = - \sum_{i=1}^n \mathbb{E}[T_{\text{particle}}^i | Y_i] \mathbf{Z}_i \mathbf{Z}_i^T \exp[\mathbf{Z}_i^T \boldsymbol{\beta}]. \quad (75)$$

Then, the Newton-Raphson update for  $\boldsymbol{\beta}$  becomes

$$\boldsymbol{\beta}^{(m)} = \boldsymbol{\beta}^{(m-1)} - (d_{\beta}^2 Q)^{-1} \nabla_{\beta} Q. \quad (76)$$

A similar update is available for  $\boldsymbol{\beta}$ . We contrast the simplicity of the update expressions (76) with the formula for the Kendall process transition probability (9).

## INTEGRAL FUNCTIONALS OF BDPS

Many important real-life applications of BDPS can be characterized as questions about the distribution of summary statistics. A common feature of stochastic processes in decision-making contexts is that the parameters estimated by the statistical inference procedure are not always the ones of interest in the application. Often the quantity of interest is a summary statistic related to the time-integral of the process. To illustrate, let  $g: \mathbb{R}^+ \rightarrow [0, \infty)$  be a function and let  $S$  be a set of “taboo” or prohibited states. Suppose the initial state of the BDP is  $X(0) = i \in \mathbb{R}^+ \setminus S$ . Define the functional

$$C_i = \int_0^{\tau_i} g(X(t)) dt, \quad (77)$$

where the upper limit of integration is the first passage time

$$\tau_i = \inf \{t : X(t) \in S \mid X(0) = i\}. \quad (78)$$

Here,  $C_i$  is a functional because it maps a realization of the stochastic process  $g(X(t))$  to its integral. Figure 2 shows an example realization of a BDP and its integral  $C_i$  with  $S = \{0\}$ . The left-hand side shows a BDP beginning at  $X(0) = 1$ , and ending at  $X(\tilde{A}) = 0$ . The right-hand plot shows  $g(X(t))$  over the same time interval, and the area under the trajectory is  $C_i$ .

Expressions like (77) arise often in applied work. For example, epidemiologists usually estimate the parameters (contact/infection rate and recovery rate) of an epidemic process from data, but their objective is to make inference of the predictive distribution of the *cost* of the epidemic in the future. Operations researchers may estimate the arrival rate  $\lambda$  and service rate  $\mu$  in a queuing process, but the object of inference is the distribution of *customer-hours* waited. Traffic engineers may be interested in the number of *vehicle-hours* waited in models for highway accident delays (Gaver, 1969).



To illustrate the role of integral summaries of BDPs in statistical prediction, let  $p(c|\theta)$  be the density of  $C_i$  given  $\theta$ . The posterior predictive uncertainty about the statistic is the marginal distribution

$$\begin{aligned} p(c|Y) &= \int_{\Theta} p(c|\theta)p(\theta|Y) d\theta \\ &\propto \int_{\Theta} p(c|\theta)p(Y|\theta)p(\theta) d\theta, \end{aligned} \quad (79)$$

where  $p(\theta|Y)$  is the sampling distribution of  $\theta$ , given the realized data  $Y$ . In a Bayesian context,  $p(\theta|Y)$  is a posterior distribution, and we might estimate  $p(c|Y)$  by a Monte Carlo approximation involving  $N$  draws  $\theta_j \sim p(\theta|Y)$  via

$$p(c|Y) \approx \frac{1}{N} \sum_{j=1}^N p(c|\theta_j). \quad (80)$$

## BACKGROUND ON INTEGRALS OF BDPS

Karlin and McGregor (1957a, 1957b) provided the first theoretical tools for working with integral functionals of general BDPs. P. Puri (1966); P. S. Puri (1968) derives the characteristic function for the joint distribution of simple linear BDP and its integral and gives expressions for moments and limiting distributions (P. Puri, 1972; P. S. Puri, 1971, 1972). McNeil (1970) gives the first results for general BDPs, Gani and McNeil (1971) derive expressions for the joint distribution of a general BDP and its integral, and Kaplan (1974) provides limit theorems for integrals of simple BDPs with immigration. Straightforward methods for moments of integrals of general BDPs using Laplace transforms are also available (Gani & Swift, 2008; Hernandez-Suarez & Castillo-Chavez, 1999; Pollett, 2003; Pollett & Stefanov, 2003). However, most analyses of integral functionals of general BDPs are limited to simple analytically tractable models or focused on moments.

Now we consider the problem of computing the distribution of (77). Our emphasis on first-passage times as the upper limit of integration in (77) has two benefits. First, our analyses need not be conditional on an arbitrary time in the future. Second, first passage times allow us to exploit powerful analytic tools that establish a correspondence between transition probabilities and first-passage times, enabling analytic progress on integrals for arbitrary well-behaved processes. Our presentation follows the outline given by McNeil (1970). Let  $c_i(s) = \mathbb{E}[e^{-sC_i}]$  be the Laplace transform of  $C_i$ . Note that if  $X(0) = i \in S$  then  $\tau_i = 0$ ,  $C_i = 0$ , and so  $c_i(s) = 1$ . Now by an analogous conditioning argument for  $X(0) = i \notin S$ , we re-write the Laplace transform as

$$\begin{aligned}
 c_i(s) &= \int_0^\infty \mathbb{E} \left[ e^{-s(C_{i+1} + ug(i))} \right] \Pr(\text{birth}, U = u \mid X(0) = i) \, du \\
 &\quad + \int_0^\infty \mathbb{E} \left[ e^{-s(C_{i-1} + ug(i))} \right] \Pr(\text{death}, U = u \mid X(0) = i) \, du \\
 &= \mathbb{E} \left[ e^{-sC_{i+1}} \right] \int_0^\infty e^{-sug(i)} \lambda_i e^{-(\lambda_i + \mu_i)u} \, du \\
 &\quad + \mathbb{E} \left[ e^{-sC_{i-1}} \right] \int_0^\infty e^{-sug(i)} \mu_i e^{-(\lambda_i + \mu_i)u} \, du \\
 &= c_{i+1}(s) \lambda_i \int_0^\infty e^{-u(sg(i) + \lambda_i + \mu_i)} \, du \\
 &\quad + c_{i-1}(s) \mu_i \int_0^\infty e^{-u(sg(i) + \lambda_i + \mu_i)} \, du,
 \end{aligned} \tag{81}$$

that gives

$$(sg(i) + \lambda_i + \mu_i) c_i(s) = \lambda_i c_{i+1}(s) + \mu_i c_{i-1}(s). \tag{82}$$

Now dividing both sides of the above by  $g(i)$ , we find that

$$(s + \lambda_i^* + \mu_i^*) c_i(s) = \lambda_i^* c_{i+1}(s) + \mu_i^* c_{i-1}(s), \tag{83}$$

where  $\lambda_i^* = \lambda_i / g(i)$  and  $\mu_i^* = \mu_i / g(i)$ . Therefore, we see that (127) is simply the backward equation for a modified process with birth and death rates  $\lambda_i^*$  and  $\mu_i^*$  for  $i \in \mathbb{N}$ . The forward equation for the cumulative distribution function of  $C_i$  is therefore equivalent to (5) with the modified birth and death rates.

Pollett (2003) gives the conditions, analogous to those for (16), under which this modified process explodes. We note that differentiation of solutions of (83) yields the moments of  $C_i$ , as noted by McNeil (1970) and subsequently refined by Hernandez-Suarez and Castillo-Chavez (1999), Stefanov and Wang (2000), and Pollett (2003). We refer interested readers to those papers and focus here on results for the distribution of  $C_i$  that are more useful in statistical and decision applications.

To take advantage of (83), we modify (29) as follows. Fix  $S \subset \mathbb{N}$  and suppose  $X(t)$  is a general BDP with rates  $\{\lambda_n\}$  and  $\{\mu_n\}$  with starting state  $X(0) = i \in \mathbb{N} \setminus S$ . Suppose  $g(n)$  is a positive function defined for all  $n \in \mathbb{N}$ . Let  $Y(t)$  be a general BDP with rates  $\lambda_n^* = \lambda_n / g(n)$  and  $\mu_n^* = \mu_n / g(n)$  for all  $n \in \mathbb{N} \setminus S$ , and  $\lambda_n^* = \mu_n^* = 0$  for every  $n \in S$ . Then let  $P_{ij}^*(t) = \Pr(Y(t) = j \mid Y(0) = i)$ . We then have

$$H(c) = \Pr(C_i < c) = \sum_{j \in S} P_{ij}^*(c). \tag{84}$$

If instead of the cumulative distribution function  $H(c)$  of  $C_i$ , we wish to have the probability density, we could numerically differentiate (84). However, using the properties of the Laplace transform,

$$\begin{aligned} h(c) &= \frac{d}{dc} \Pr(C_i < c) \\ &= \sum_{j \in S} \frac{d}{dc} P_{ij}^*(c) \\ &= \sum_{j \in S} \mathcal{L}^{-1} [s f_{ij}^*(s) - P_{ij}^*(0)](c) \\ &= \sum_{j \in S} \mathcal{L}^{-1} [s f_{ij}^*(s)](c) \end{aligned} \quad (85)$$

where  $f_{ij}^*(s)$  is the Laplace transform of  $P_{ij}^*(t)$ ,  $\mathcal{L}^{-1}[\cdot]$  denotes Laplace inversion, and  $P_{ij}^*(0) = 0$  for all  $j \in S$  since we have assumed  $i \notin S$ .

## EXAMPLE: PROBABILISTIC CONTROL OF AN EPIDEMIC

In infectious disease epidemiology, stochastic modeling can give valuable insight into both disease dynamics and optimal intervention strategies (Ball, 1986; Wickwire, 1977). The total cost of an infectious disease epidemic is proportional to the area under the time trajectory of the number of infected people (Gani & Jerwood, 1972; Jerwood, 1970). To illustrate, we model the number of infected persons in a homogeneously mixing population as a type of general BDP. This simple model, called the susceptible-infected-susceptible (SIS) model, keeps track of the number of infected in a finite population of size  $N$  (N. T. Bailey, 1975). If there are currently  $n < N$  infected persons in the population, the rate of new infections is proportional to the product of the number infected  $n$  and susceptible  $N - n$ . The contact/transmission rate between infected and susceptible persons is  $\lambda$ . Infected persons recover and revert to susceptible status with constant per-person rate  $\mu$ . For a SIS process  $X(t)$ , the addition and removal rates are

$$\lambda_n = \lambda n(N - n) \quad \text{and} \quad \mu_n = n(\mu + \epsilon), \quad (86)$$

where  $\epsilon$  is a positive control parameter related to vaccination or some other public health intervention strategy. Suppose the initial number of infected is  $X(0) = i \leq N$  and we are interested in the total cost of the epidemic until its eventual extinction, so  $S = \{0\}$ . Let the cost of managing the epidemic per unit time be  $a\epsilon$ . Additionally, let the cost per infected person per unit time be  $b > 0$ , so the cost function becomes  $g(n) = a\epsilon + bn$ . Then the total cost is

$$C_i = \int_0^{\tau_i} [a\epsilon + bX(t)] dt = a\epsilon\tau_i + b \int_0^{\tau_i} X(t) dt, \quad (87)$$

where  $\tau_i$  is the time to extinction of the epidemic.

Most optimal control models seek a policy that minimizes the expected total cost, corresponding to the expectation of (77) under certain conditions on the intervention and cost functions (Cai & Luo, 1994; Clancy, 1999; Guo & Hernandez-Lerma, 2009; Lefevre, 1981). The availability of probability distributions for the total cost allows us to seek the minimal intervention policy that guarantees that the total cost of the epidemic is small with high probability. Let  $X(t)$  be the process with rates given by (86) for a certain control setting  $\tilde{\delta}$ . Then we wish to find the smallest  $\tilde{\delta}$  such that

$$\Pr(C_i < c) > 1 - \alpha, \quad (88)$$

where  $c$  is a desired bound on the total cost, and  $0 < \alpha < 1$  is a small probability. Assuming this probability is continuous and increases monotonically with  $\tilde{\delta}$  near  $1 - \alpha$ , it is straightforward to find the smallest  $\tilde{\delta}$  that satisfies (88).

Figure 3 shows how to find the minimal  $\tilde{\delta}$  for a SIS process with  $N = 100$  individuals,  $X(0) = 50$ , infectivity  $\beta = 0.1$ , recovery rate  $\mu = 8$ , control cost  $a = 0.1$ , and per-infected cost  $b = 0.3$  per unit time. The top traces show the cumulative distribution function of the total cost for  $\tilde{\delta} = 0, 0.5, 1, 1.5, 2$ . The vertical gray line shows  $C_i = 7$ , and we wish to keep the total cost less than 7 with probability  $1 - \alpha = 0.95$ . The bottom trace shows  $\Pr(C_i < 7)$  as a function of  $\tilde{\delta}$ .

The horizontal gray dashed line shows 0.95 probability, and the vertical gray dashed line shows the smallest value of  $\tilde{\delta}$  ( $\tilde{\delta} \approx 3.4$ ) that achieves this bound.

## DISCUSSION: COMPUTATION AND LIKELIHOOD-BASED INFERENCE FOR BDPS

BDPs are vital tools for modeling stochastic counting processes in epidemiology, evolution, ecology, chemistry, physics, and other fields. Modeling with BDPs is often straightforward; by considering rates of addition of new particles and removal of existing particles, conditional on the number already present, researchers can specify the birth and death rates  $\{\lambda_k, \mu_k\}_{k=0}^{\infty}$ . The ease of modeling with BDPs stands in stark contrast to the computational difficulty of inference using stochastic realizations of BDPs. Routine use of BDPs in statistical settings has been thwarted by intractable likelihoods and burdensome computations. A unified perspective on BDPs with arbitrary birth and death rates has remained elusive, until recently.

Laplace transforms of transition probabilities provide the essential analytic tools for bridging this gap in practice. Our approach for computing transition probabilities (likelihoods) in (25) and conditional expectations in the E-step (46) is general, robust, and computationally efficient. Laplace transforms of first-passage times also play an important role in finding the distribution of integral functionals of BDPs in applications. As a theoretic tool, this Laplace-perspective is not new; Karlin and McGregor (1957a, 1958a, 1957b) discuss the fundamental importance of Laplace transforms for analysis of BDPs. More recent results related to combinatorial properties of BDPs also rely on Laplace transforms (Flajolet & Guillemin, 2000; Guillemin & Pinchon, 1998, 1999). Extensions to the general approach presented here have been developed for multivariate BDPs, where progress has been slower and analytic approaches to reducing computation are less readily available. Analytic formulae of transition probabilities for the simplest multivariate Kendall-like processes – called monomolecular reaction systems – have only recently been derived (Jahnke & Huisinga, 2007). Notably, Xu, Guttorp, Kato-Maeda, and Minin (2015) propose a fast algorithm to compute the transition probabilities of multi-type branching processes using a generating function approach. However, this method is only applicable for processes whose transition rates are linear. The first result for non-linear multivariate BDPs has recently been established by Ho, Xu, Crawford, Minin, and Suchard (2017), who consider a subclass of bivariate BDPs called birth/birth-death processes and develop an efficient computation method for the transition probabilities by deriving recursion formulae for their Laplace transforms. Nonetheless, the problem of computing the transition probabilities of general multivariate BDPs remains open and will be an exciting research direction for the future.

In this chapter, we have outlined new tools for practical likelihood-based analysis inference of BDP parameters under discrete and continuous observation of the process. In particular, BDP generalizations of Poisson regression yield more flexible and easy-to-fit models for count data. We have intentionally limited our discussion to basic computation of likelihoods, algorithms for maximum likelihood estimation, and finding the distribution of integral summary statistics for general BDPs. But these are only the first steps toward a comprehensive theory of estimation for BDPs. Ideally, we would like to see an analysis of identifiability, consistency and other statistical properties, like the groundbreaking work of Guttorp (1991) for Galton-Watson branching processes. We hope this review will stimulate statistical research related to BDPs with a view to bringing this rich class of stochastic models into wider use by applied scientists.

## ACKNOWLEDGEMENTS

FWC was supported by NIH grants DP2 OD022614 and T32GM008185. LSTH was supported by startup funds from the Dalhousie University and the Canada Research Chairs program. MAS was supported by NIH grants R01 AI107034 and R01 HG006139, and NSF grant DMS 1264153 and IIS 1251151. We thank Yiyi Liu for helpful comments.

## References

- Abate, J. , & Whitt, W. (1999). Computing Laplace transforms for numerical inversion via continued fractions. *INFORMS J Comput*, 11(4), 394-405.
- Andersson, H. , & Britton, T. (2000). *Stochastic Epidemic Models and their Statistical Analysis*. Springer New York.
- Anscombe, F. J. (1953). Sequential estimation. *Journal of the Royal Statistical Society B*, 15(1), 1-29.
- Athreya, K. B. , & Ney, P. E. (2004). *Branching processes*. Courier Corporation.
- Bailey, N. T. (1975). *The mathematical theory of infectious diseases and its applications* (2nd ed.). Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE.
- Bailey, N. T. J. (1964). *The Elements of Stochastic Processes with Applications to the Natural Sciences*. Wiley New York.
- Ball, F. (1986). A unified approach to the distribution of total size and total area under the trajectory of infectives in epidemic models. *Advances in Applied Probability*, 289-310.
- Bankier, J. D. , & Leighton, W. (1942). Numerical continued fractions. *Am J Math*, 64(1), 653-668.
- Bladt, M. , & Sorensen, M. (2005). Statistical inference for discretely observed Markov jump processes. *Journal of the Royal Statistical Society B*, 67(3), 395-410.
- Blanch, G. (1964). Numerical evaluation of continued fractions. *SIAM Rev*, 6(4), 383-421.
- Bordes, G. , & Roehner, B. (1983). Application of Stieltjes theory for S-fractions to birth and death processes. *Advances in Applied Probability*, 15(3), 507-530.
- Cai, H. , & Luo, X. (1994). Stochastic control of an epidemic process. *International Journal of Systems Science*, 25(4), 821-828.
- Clancy, D. (1999). Optimal intervention for epidemic models with general infection and removal rate functions. *Journal of Mathematical Biology*, 39(4), 309-331.
- Craviotto, C. , Jones, W. B. , & Thron, W. J. (1993). A survey of truncation error analysis for Padé and continued fraction approximants. *Acta Appl Math*, 33, 211-272.
- Crawford, F. W. , Minin, V. N. , & Suchard, M. A. (2014). Estimation for general birth-death processes. *Journal of the American Statistical Association*, 109(506), 730-747.

- Crawford, F. W. , Stutz, T. C. , & Lange, K. (2016). Coupling bounds for approximating birth–death processes by truncation. *Statistics & probability letters*, 109, 30–38.
- Crawford, F. W. , & Suchard, M. A. (2012). Transition probabilities for general birth-death processes with applications in ecology, genetics, and evolution. *Journal of Mathematical Biology*, 65, 553-580.
- Crawford, F. W. , Weiss, R. E. , & Suchard, M. A. (2015). Sex, lies, and self-reported counts: Bayesian mixture models for longitudinal heaped count data via birth-death processes. *The Annals of Applied Statistics*, 9(2), 572-596.
- Crawford, F. W. , & Zelterman, D. (2015). Markov counting models for correlated binary responses. *Biostatistics*, 16(3), 427–440.
- Cuyt, A. , Petersen, V. , Verdonk, B. , Waadeland, H. , & Jones, W. (2008). *Handbook of Continued Fractions for Special Functions*. Springer Berlin / Heidelberg.
- Darwin, J. H. (1956). The behaviour of an estimator for a simple birth and death process. *Biometrika*, 43(1), 23-31.
- Dauxois, J.-Y. (2004). Bayesian inference for linear growth birth and death processes. *Journal of Statistical Planning and Inference*, 121(1), 1–19.
- Dempster, A. P. , Laird, N. M. , & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B*, 1–38.
- Demuth, J. P. , De Bie, T. , Stajich, J. E. , Cristianini, N. , & Hahn, M. W. (2006). The evolution of mammalian gene families. *PloS One*, 1(1), e85.
- Dobson, A. J. (2001). *An introduction to generalized linear models*. CRC press.
- Doss, C. R. , Suchard, M. A. , Holmes, I. , Kato-Maeda, M. , & Minin, V. N. (2013). Fitting birth-death processes to panel data with applications to bacterial DNA fingerprinting. *The Annals of Applied Statistics*, 7(4), 2315.
- Faddy, M. , & Bosch, R. (2001). Likelihood-based modeling and analysis of data underdispersed relative to the Poisson distribution. *Biometrics*, 57(2), 620–624.
- Faddy, M. J. (1997). Extended poisson process modelling and analysis of count data. *Biometrical Journal*, 39(4), 431–440.
- Feller, W. (1971). *An Introduction to Probability Theory and its Applications*. Wiley New York.

- Flajolet, P. , & Guillemin, F. (2000). The formal theory of birth-and-death processes, lattice path combinatorics and continued fractions. *Advances in Applied Probability*, 32(3), 750–778.
- Fudenberg, D. , Imhof, L. , Nowak, M. A. , & Taylor, C. (2004). Stochastic evolution as a generalized moran process. *Unpublished manuscript*.
- Gani, J. , & Jerwood, D. (1972). The cost of a general stochastic epidemic. *Journal of Applied Probability*, 9(2), 257–269.
- Gani, J. , & McNeil, D. R. (1971). Joint distributions of random variables and their integrals for certain birth-death and diffusion processes. *Advances in Applied Probability*, 3(2), 339–352.
- Gani, J. , & Swift, R. (2008). A simple approach to the integrals under three stochastic processes. *Journal of Statistical Theory and Practice*, 2(4), 559–568.
- Gaver, D. (1969). Highway delays resulting from flow-stopping incidents. *Journal of Applied Probability*, 6(1), 137–153.
- Guillemin, F. , & Pinchon, D. (1998). Continued fraction analysis of the duration of an excursion in an  $M/M/\infty$  system. *Journal of Applied Probability*, 35(1), 165–183.
- Guillemin, F. , & Pinchon, D. (1999). Excursions of birth and death processes, orthogonal polynomials, and continued fractions. *Journal of Applied Probability*, 36(3), 752–770.
- Guo, X. , & Hernández-Lerma, O. (2009). *Continuous-time Markov decision processes*. Springer.
- Guttorp, P. (1991). *Statistical inference for branching processes*. Wiley-Interscience.
- Hernández-Suárez, C. , & Castillo-Chavez, C. (1999). A basic result on the integral for birth-death Markov processes. *Mathematical Biosciences*, 161(1), 95–104.
- Ho, L. S. T. , Xu, J. , Crawford, F. W. , Minin, V. N. , & Suchard, M. A. (2017). Birth/birth-death processes and their computable transition probabilities with biological applications. *Journal of Mathematical Biology*, in press.
- Hobolth, A. , & Jensen, J. L. (2005). Statistical inference in evolutionary models of DNA sequences via the EM algorithm. *Statistical Applications in Genetics and Molecular Biology*, 4(1).



- Hobolth, A. , & Stone, E. A. (2009). Simulation from endpoint-conditioned, continuous-time Markov chains on a finite state space, with applications to molecular evolution. *Annals of Applied Statistics*, 3(3), 1024-1231.
- Holmes, I. , & Bruno, W. J. (2001, September). Evolutionary HMMs: a Bayesian approach to multiple alignment. *Bioinformatics*, 17(9), 803-820.
- Holmes, I. , & Rubin, G. (2002). An expectation maximization algorithm for training hidden substitution models. *Journal of Molecular Biology*, 317(5), 753-764.
- Ismail, M. E. H. , Letessier, J. , & Valent, G. (1988). Linear birth and death models and associated Laguerre and Meixner polynomials. *J Approx Theory*, 55(3), 337-348.
- Jahnke, T. , & Huisinga, W. (2007). Solving the chemical master equation for monomolecular reaction systems analytically. *Journal of Mathematical Biology*, 54(1), 1-26.
- Jerwood, D. (1970). A note on the cost of the simple epidemic. *Journal of Applied Probability*, 7(2), 440-443.
- Kaplan, N. (1974). Limit theorems for the integral of a population process with immigration. *Stochastic Processes and their Applications*, 2(3), 281-294.
- Karev, G. P. , Wolf, Y. I. , Rzhetsky, A. Y. , Berezovskaya, F. S. , & Koonin, E. V. (2002). Birth and death of protein domains: a simple model of evolution explains power law behavior. *BMC Evolutionary Biology*, 2(1), 18.
- Karlin, S. , & McGregor, J. (1957a). The classification of birth and death processes. *Transactions of the American Mathematical Society*, 86(2), 366-400.
- Karlin, S. , & McGregor, J. (1958a). Linear growth, birth and death processes. *Journal of Mathematics and Mechanics*, 643-662.
- Karlin, S. , & McGregor, J. (1958b). Many server queueing processes with poisson input and exponential service times. *Pacific Journal of Mathematics*, 8(1), 87-118.
- Karlin, S. , & McGregor, J. L. (1957b). The differential equations of birth-and-death processes, and the Stieltjes moment problem. *Transactions of the American Mathematical Society*, 85(2), 489-546.
- Karlin, S. , & Taylor, H. M. (1975). *A First Course in Stochastic Processes*. Academic Press.
- Keiding, N. (1975). Maximum likelihood estimation in the birth-and-death process. *Annals of Statistics*, 3(2), 363-372.

- Kendall, D. G. (1948). On the generalized" birth-and-death" process. *The Annals of Mathematical Statistics*, 1–15.
- Kendall, D. G. (1949). Stochastic processes and population growth. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(2), 230–282.
- Kimmel, M. , & Axelrod, D. E. (2016). *Branching processes in biology*. Springer Publishing Company, Incorporated.
- Kingman, J. F. (1982). On the genealogy of large populations. *Journal of Applied Probability*, 19(A), 27–43.
- Klar, B. , Parthasarathy, P. , & Henze, N. (2010). Zipf and lersch limit of birth and death processes. *Probability in the Engineering and Informational Sciences*, 24(1), 129–144.
- Krone, S. M. , & Neuhauser, C. (1997). Ancestral processes with selection. *Theoretical Population Biology*, 51, 210-237.
- Lange, K. (1995). A gradient algorithm locally equivalent to the EM algorithm. *Journal of the Royal Statistical Society B*, 57(2), 425-437.
- Lange, K. (2010a). *Applied Probability* (2nd ed.). Springer New York.
- Lange, K. (2010b). *Numerical analysis for statisticians* (2nd ed.). Springer New York. Hardcover.
- Lee, J. , Weiss, R. E. , & Suchard, M. A. (2011). Using a birth-death process to account for reporting errors in longitudinal self-reported counts of behavior. *Unpublished UCLA Biostatistics Technical report*.
- Lefèvre, C. (1981). Optimal control of a birth and death epidemic process. *Operations Research*, 29(5), 971-982.
- Lenin, R. , & Parthasarathy, P. (2000). A birth-death process suggested by a chain sequence. *Computers & Mathematics with Applications*, 40(2-3), 239–247.
- Lorentzen, L. , & Waadeland, H. (1992). *Continued Fractions with Applications*. North-Holland, Amsterdam.
- Mayrose, I. , Barker, M. S. , & Otto, S. P. (2010). Probabilistic models of chromosome number evolution and the inference of polyploidy. *Systematic Biology*, 59(2), 132-144.

- McFarland, C. D. , Mirny, L. A. , & Korolev, K. S. (2014). Tug-of-war between driver and passenger mutations in cancer and other adaptive processes. *Proceedings of the National Academy of Sciences*, 111(42), 15138–15143.
- McNeil, D. (1970). Integral functionals of birth and death processes and related limiting distributions. *Annals of Mathematical Statistics*, 480-485.
- Metzner, P. , Dittmer, E. , Jahnke, T. , & Schütte, C. (2007). Generator estimation of Markov jump processes. *Journal of Computational Physics*, 227(1), 353–375.
- Moran, P. (1951). Estimation methods for evolutive processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 141–146.
- Moran, P. (1953). The estimation of the parameters of a birth and death process. *Journal of the Royal Statistical Society. Series B (Methodological)*, 241–245.
- Moran, P. A. P. (1958). Random processes in genetics. In *Mathematical proceedings of the cambridge philosophical society* (Vol. 54, pp. 60–71).
- Murphy, J. , & O'donohoe, M. (1975). Some properties of continued fractions with applications in markov processes. *IMA Journal of Applied Mathematics*, 16(1), 57–71.
- Nee, S. (2006). Birth-death models in macroevolution. *Annual Review of Ecology, Evolution, and Systematics*, 37, 1–17.
- Nee, S. , May, R. M. , & Harvey, P. H. (1994). The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 344(1309), 305–311.
- Neuts, M. F. (1995). *Algorithmic probability: A collection of problems (stochastic modeling series)*. Chapman and Hall/CRC.
- Norris, J. R. (1998). *Markov chains* (No. 2008). Cambridge university press.
- Novozhilov, A. S. , Karev, G. P. , & Koonin, E. V. (2006). Biological applications of the theory of birth-and-death processes. *Briefings in Bioinformatics*, 7(1), 70-85.
- Nowak, M. A. , Michor, F. , & Iwasa, Y. (2003). The linear process of somatic evolution. *Proceedings of the national academy of sciences*, 100(25), 14966–14969.
- Nowak, M. A. , Sasaki, A. , Taylor, C. , & Fudenberg, D. (2004). Emergence of cooperation and evolutionary stability in finite populations. *Nature*, 428(6983), 646.

- Ohkubo, J. (2014). Karlin–McGregor-like formula in a simple time-inhomogeneous birth–death process. *Journal of Physics A: Mathematical and Theoretical*, 47(40), 405001.
- Parthasarathy, P. , Lenin, R. , Schoutens, W. , & Van Assche, W. (1998). A birth and death process related to the Rogers–Ramanujan continued fraction. *Journal of Mathematical Analysis and Applications*, 224(2), 297–315.
- Pollett, P. (2003). Integrals for continuous-time markov chains. *Mathematical Biosciences*, 182(2), 213–225.
- Pollett, P. , & Stefanov, V. (2003). A method for evaluating the distribution of the total cost of a random process over its lifetime. In *International congress on modelling and simulation* (Vol. 4, p. 1863–1867).
- Puri, P. (1966). On the homogeneous birth-and-death process and its integral. *Biometrika*, 53(1-2), 61–71.
- Puri, P. (1972). A method for studying the integral functional of stochastic processes with applications II. sojourn time distributions for Markov chains. *Probab Theory Rel*, 23(2), 85–96.
- Puri, P. S. (1968). Some further results on the birth-and-death process and its integral. In *Mathematical proceedings of the cambridge philosophical society* (Vol. 64, pp. 141–154).
- Puri, P. S. (1971). A method for studying the integral functionals of stochastic processes with applications: I. Markov chain case. *Journal of Applied Probability*, 8(2), 331–343.
- Puri, P. S. (1972). A method for studying the integral functionals of stochastic processes with applications III. In *Proceedings of the sixth Berkeley symposium on mathematical statistics and probility* (Vol. 3, p. 481–500). University of California Press.
- Reed, W. J. , & Hughes, B. D. (2004). A model explaining the size distribution of gene and protein families. *Mathematical Biosciences*, 189(1), 97–102.
- Renshaw, E. (1993). *Modelling Biological Populations in Space and Time*. Cambridge University Press.
- Renshaw, E. (2011). *Stochastic Population Processes: Analysis, Approximations, Simulations*. Oxord University Press.
- Reynolds, J. F. (1973). On estimating the parameters of a birth-death process. *Australian & New Zealand Journal of Statistics*, 15(1), 35–43.

- Rosenberg, N. A. , Tsohaki, A. G. , & Tanaka, M. M. (2003). Estimating change rates of genetic markers using serial samples: applications to the transposon IS6110 in *Mycobacterium tuberculosis*. *Theoretical Population Biology*, 63(4), 347-363.
- Ross, S. M. (1995). *Stochastic processes* (2nd ed.). Wiley.
- Sagitov, S. (2013). Linear-fractional branching processes with countably many types. *Stochastic Processes and their Applications*, 123(8), 2940–2956.
- Stefanov, V. , & Wang, S. (2000). A note on integrals for birth-death processes. *Mathematical Biosciences*, 168(2), 161-165.
- Tan, W. Y. , & Piantadosi, S. (1991). On stochastic growth processes with application to stochastic logistic growth. *Statistica Sinica*, 1, 527-540.
- Thorne, J. , Kishino, H. , & Felsenstein, J. (1991, August). An evolutionary model for maximum likelihood alignment of DNA sequences. *Journal of Molecular Evolution*, 33(2), 114-124.
- Wall, H. S. (1948). *Analytic Theory of Continued Fractions*. D. Van Nostrand Company, Inc. New York.
- Wallis, J. (1972). *Opera mathematica volume 1. oxoniae e theatro shedoniano*. reprinted by Georg Olms Verlag, Hildesheim, New York.
- Wickwire, K. (1977). Mathematical models for the control of pests and infectious diseases: a survey. *Theoretical Population Biology*, 11(2), 182-238.
- Wolff, R. W. (1965). Problems of statistical inference for birth and death queuing models. *Operations Research*, 13(3), 343-357.
- Xu, J. , Guttorp, P. , Kato-Maeda, M. , & Minin, V. N. (2015). Likelihood-based inference for discretely observed birth–death-shift processes, with applications to evolution of mobile genetic elements. *Biometrics*, 71(4), 1009–1021.

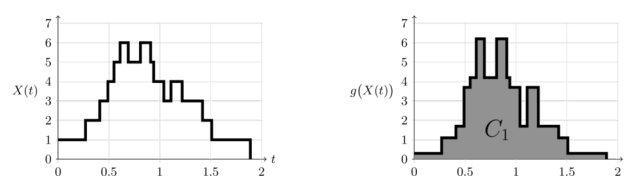
Figure 1. Stochastic simulation of a BDP starting at  $X(0) = 1$  on the interval  $0 < t < 2$ .

Figure 2. Illustration of the integral of a functional of a general birth-death process (BDP). On the left, a BDP begins at  $X(0) = 1$  and ends when the process reaches the absorbing state 0 just before time  $t = 2$ . On the right,  $C_1 = \int_0^{\tau_1} g(X(t)) dt$  is the area under the trajectory of  $g(X(t))$ , where  $g: \mathbb{N} \rightarrow [0, \infty)$  is an arbitrary positive “reward” or “cost” function. The upper limit of integration  $\tau_1$  is the first passage time to zero, beginning at  $X(0) = 1$ .

Figure 3. Probabilistic control of a stochastic SIS epidemic. At top, the distribution of total epidemic cost  $C_i$  for different values of a control parameter  $\delta$ . The dashed gray vertical line is at  $w = 7$ , and we wish to keep  $C_i < 7$  with high probability. At bottom, the probability that  $C_i < 7$  as a function of the control parameter  $\delta$ . The horizontal gray dashed line denotes 0.95, and the vertical dashed line is the smallest epsilon that achieves  $\Pr(C_i < 7) > 0.95$ ; this yields  $\delta_{H3.4}$ . In this way, we can easily find the smallest value of a control parameter that bounds the probability that the epidemic will exceed a certain threshold.

Table 1. Some well-known BDPs with birth and death rates  $\lambda_k$  and  $\mu_k$ . The SI and SIS models refers to the susceptible-infectious(-susceptible) process in epidemiology in which there are  $k$  infectious individuals in a finite population of size  $N$ . The Moran/Ehrenfest process models the change in the numbers of particles of two types, where transitions between types occur at a rate proportional to the number of potential contacts between members each type in a finite population of size  $N$ .

Model	$\lambda_k$	$\mu_k$
Poisson	$\lambda$	0
Yule/Pure birth	$k\lambda$	0
Survival/Pure death	0	$k\mu$
Kendall	$k\lambda$	$k\mu$
Kendall + immigration	$k\lambda + \alpha$	$k\mu$
$M/M/1$ queue	$\lambda$	$\mu$
$M/M/c$ queue	$\lambda$	$\min(k, c)\mu$
$M/M/\infty$ queue	$\lambda$	$k\mu$
SI/Logistic	$k(N - k)\lambda$	0
SIS	$k(N - k)\lambda$	$k\mu$
Moran/Ehrenfest	$k(N - k)\lambda$	$k(N - k)\mu$



bdintegralpath.tif

