

CONDITIONAL DISTRIBUTIONS AND WAITING TIMES IN MULTITYPE BRANCHING PROCESSES

H. K. ALEXANDER,* *Queen's University and ETH Zürich*

Abstract

In this paper we present novel results for discrete-time and Markovian continuous-time multitype branching processes. As a population develops, we are interested in the waiting time until a particular type of interest (such as an escape mutant) appears, and in how the distribution of individuals depends on whether this type has yet appeared. Specifically, both forward and backward equations for the distribution of type-specific population sizes over time, conditioned on the nonappearance of one or more particular types, are derived. In tandem, equations for the probability that one or more particular types have not yet appeared are also derived. Brief examples illustrate numerical methods and potential applications of these results in evolutionary biology and epidemiology.

Keywords: Multitype branching process; probability generating function; waiting time; escape mutant; evolutionary biology; epidemiology

2010 Mathematics Subject Classification: Primary 60G99

Secondary 92D15; 92D30

1. Introduction

Branching processes are widely used to model a variety of phenomena, including chain reactions [14], [26], cellular proliferation and differentiation [16], [19], and the spread of epidemics [8]. Multitype branching processes allow for several types of replicating entities, thus incorporating variation within a population, and are therefore useful in addressing questions in evolutionary biology. Branching process models are often concerned with finding the probability of extinction of the population. In some applications, however, one would also like to understand the time course of the population's development. This might include the time of appearance of different types, as well as the distribution of individuals present over time, perhaps depending on the types that have or have not yet appeared. For instance, in a model of within-host infection initiated by a wild-type pathogen sensitive to drug treatment, one might investigate the distribution of pathogen population size over time, conditional on whether a drug-resistant mutant has arisen. The main result of this paper is to derive equations for the probability generating function (PGF) of the conditioned type-specific population size distribution, along with equations for the probability that a particular type has not yet appeared.

There appears to be little previous work considering the absence of a particular type in multitype branching processes, with the notable exception of that by M. C. Serra and colleagues [22], [23], [24]. They first considered a two-type Galton–Watson branching process with independent mutations from the subcritical type 1 to the supercritical type 2. A PGF for the number of type-2 individuals with type-1 parents, produced up to any given time, leads to expressions for the distribution of the ‘time to escape’ (i.e. the first generation in which a

Received 5 May 2011; revision received 22 October 2012.

* Postal address: Institute for Integrative Biology, ETH Zürich, CHN H.74, Universitätsstrasse 16, Zürich, CH-8092, Switzerland. Email address: helen.alexander@env.ethz.ch

type 2 whose lineage escapes extinction is produced) [23]. The authors subsequently derived an approximation for the distribution of the waiting time for a successful mutant in the two-type process [24]. Finally, they focused on derivations of the asymptotic structure of the process with an arbitrary number of types, albeit with a particular form of mutation, as the mutation rate approaches 0 [22]. Using the authors' approach of tracking the number of mutants, results on the exact distribution of the waiting time for a particular type can readily be extended to more general scenarios (arbitrary number of types, mutation scheme, etc.) as well as continuous-time models; see [2] for details. However, these results do not delve into the full distribution of population sizes, which depends on whether the mutant has arisen.

The present paper is organized as follows. In Section 2 we define the process under consideration; present a few simple probabilistic results that will be useful later; and recall well-known equations for the PGFs describing the unconditioned type-specific population size distribution in a multitype branching process. In Sections 3–5 we present derivations of novel results concerning a multitype branching process in which a particular type is absent. (Extensions to more general conditions, such as nonextinction, the presence of some type, or the absence of several types, are considered in Appendix A.) Specifically, we obtain coupled equations governing (i) the probability that a particular type has not yet appeared, and (ii) the PGF for the distribution of the number of individuals of each type over time, conditional on the nonappearance of the particular type. As with classical results in branching process theory, we derive both forward and backward equations. These represent two conceptually distinct approaches to the problem, from which we can confirm that the results agree. Furthermore, the two resulting sets of formulae offer two possibilities for numerical implementation. In Section 3 we treat the discrete-time case (Galton–Watson branching processes), while in Section 4 we deal with Markovian (memoryless) processes in continuous time. The interpretation of these results to express the exact distribution of the waiting time until the appearance of a type is outlined in Section 5. Finally, in Section 6 we provide a few examples illustrating how these results can be used to address questions of interest in evolutionary biology and epidemiology.

2. Preliminaries

2.1. Definitions and notation

We consider m -type Markovian branching processes, in both discrete and continuous time. The random vector $X \in \mathbb{N}_0^m$ (where $\mathbb{N}_0 = \{0, 1, 2, \dots\}$) denotes the number of each type of individual in the process, while G denotes the PGF for X . More explicitly, $G_k(s; n)$ (respectively $G_k(s; t)$ in continuous time) denotes the PGF for X at time step n (respectively time t), given that the process starts with one type- k individual. All functions and variables, such as $s = (s_1, \dots, s_m)$, are understood to be m -dimensional vectors where appropriate. Generally, a subscript k on a probability (Pr) or expectation (\mathbb{E}) indicates that the process is conditioned on starting with a single type- k individual.

A discrete-time branching process is defined by $F(s)$, the PGF for the distribution of the number of offspring produced in a time step (with the k th component again corresponding to a type- k parent). In general,

$$F_k(s) = (1 - \eta_k)s_k + \eta_k f_k(s) = s_k + \eta_k(f_k(s) - s_k),$$

where η_k is the probability that a type- k individual's lifetime ends in a given time step. If this is the case, it produces a new generation of offspring according to the PGF $f_k(s)$. Otherwise, the individual's lifetime does not end and it simply places itself (PGF s_k) in the next

time step. Biologically, the motivation for this formulation is to explicitly model overlapping generations. From a mathematical standpoint, however, this distinction is immaterial: one can always consider such a process as a Galton–Watson branching process with nonoverlapping generations, by viewing a surviving parent as one whose lifetime has ended and who has produced a single offspring of the same type.

Analogously, in continuous time, suppose that a type- k individual’s lifespan is exponentially distributed with parameter ξ_k . At the end of its lifetime, it gives birth to offspring according to the PGF $f_k(s)$. Thus, the distribution of individuals produced by a type k in a discrete-time interval of length δt is

$$F_k(s) = s_k + \xi_k(f_k(s) - s_k)\delta t + o(\delta t). \tag{1}$$

Some additional notational conventions are required when conditioning is introduced. For clarity of exposition, in the main text we focus only on the condition that type m has not yet appeared, and leave the more general case to Appendix A. As such, a subscript ‘*’ on a variable or a function will be used to indicate that the m th component is fixed at 0, e.g. $s_* := (s_1, \dots, s_{m-1}, 0)$, $1_* := (1, \dots, 1, 0)$, and $F_*(s_*) := (F_1(s_*), \dots, F_{m-1}(s_*), 0)$. A tilde on the PGF G will indicate that X is conditioned on some event (in the main text this will be the nonappearance of type m). More explicitly, $\tilde{G}_k(s; n)$ is the PGF for $X(n)$ if the process starts with a single type- k progenitor, $k \neq m$, conditional on no type m appearing by time step n . Note that $\tilde{G}_k(s; n) \equiv \tilde{G}_k(s_*; n)$, since conditioning on type m not being present implies that the PGF is independent of s_m . It is assumed throughout that the initial type k is not m , and that no type $i \in \{1, 2, \dots, m - 1\}$ produces type- m offspring with probability 1. (Note that these conditions do not limit us to indecomposable processes, although for some decomposable cases, the results may be trivial, as when type m can never appear.) We use $P_k(n)$ to denote the probability that type m does not appear in the lineage of a type k by time n , which is positive under the preceding conditions.

2.2. Useful PGF results

2.2.1. *Compound distributions.* For many of the derivations that follow, it will be useful to know the PGF for the number of individuals of each type at the present time given the numbers in the previous time step, conditional on some event. Denote by $\tilde{Y}_l^{(ij)}$ the number of type- l individuals placed in the next time step by the j th type- i individual present at time $n - 1$, conditional on event C holding in the entire population at time n . Since all individuals of a given type behave the same way, the $\tilde{Y}^{(ij)}$ are identically distributed copies of a random vector $\tilde{Y}^{(i)}$. The total number of type- l individuals at time n is the sum of the type- l offspring produced by each individual existing at time $n - 1$, and, therefore,

$$\begin{aligned} \mathbb{E} \left[\prod_i s_i^{X_i(n)} \mid X(n - 1); C \right] &= \mathbb{E} \left[s_1^{\sum_{i=1}^m \sum_{j=1}^{X_i(n-1)} \tilde{Y}_1^{(ij)}} \dots s_m^{\sum_{i=1}^m \sum_{j=1}^{X_i(n-1)} \tilde{Y}_m^{(ij)}} \right] \\ &= \mathbb{E} \left[\prod_{i=1}^m \prod_{j=1}^{X_i(n-1)} s_1^{\tilde{Y}_1^{(ij)}} \dots s_m^{\tilde{Y}_m^{(ij)}} \right]. \end{aligned}$$

At this point, we can make no further progress unless all the $\tilde{Y}^{(ij)}$ are independent. Independence holds if there is no condition applied ($C = \Omega$), or if the condition holds in the main process if and only if it holds, independently, in every subprocess (e.g. if C is the absence of some type(s)). With such independence, and assuming that the $\tilde{Y}^{(ij)}$ are also independent of $X(n - 1)$ (i.e. the

number of offspring produced by each individual is independent of the population size), the above expression simplifies as

$$\mathbb{E} \left[\prod_i s_i^{X_i(n)} \mid X(n-1); C \right] = \prod_{i=1}^m \mathbb{E} [s_1^{\tilde{Y}_1^{(i)}} \dots s_m^{\tilde{Y}_m^{(i)}}]^{X_i(n-1)} = \prod_{i=1}^m \tilde{F}_i(s)^{X_i(n-1)}, \tag{2}$$

where $\tilde{F}_i(s)$ is the PGF for the number of offspring placed in the next time step by one type i , conditioned on C holding in the process.

2.2.2. *One-step conditional offspring distribution.* As in (2), we will repeatedly make use of the next-time-step ‘offspring’ distribution, conditioned on no type m appearing among the offspring. The PGF of this distribution is denoted by $\tilde{F}(s)$ and it can be expressed in terms of the corresponding unconditioned PGF $F(s)$ as

$$\tilde{F}_k(s) = \frac{F_k(s_*)}{F_k(1_*)} \tag{3}$$

since $F_k(1_*)$ is the probability that there is no type m among a type- k individual’s offspring.

2.2.3. *Probability of a condition holding in multiple processes.* We will also need to express the probability that a condition C holds in multiple processes. If there are X_i processes started by type- i individuals then, provided that C holds independently in each of these processes, we have

$\Pr(C \text{ holds in all processes over } \nu \text{ time steps} \mid X = (X_1, X_2, \dots, X_m) \text{ progenitors})$

$$= \prod_{i=1}^m (P_i^{(C)}(\nu))^{X_i},$$

where $P_i^{(C)}(\nu)$ is the probability that C holds over ν time steps in a process with a type- i progenitor. If X has a probability distribution given by the PGF $\tilde{G}(s)$ then this expression simplifies as

$$\Pr(C \text{ holds in all processes over } \nu \text{ time steps}) = \tilde{G}(P^{(C)}(\nu)) \tag{4}$$

with $P^{(C)}(\nu) = (P_1^{(C)}(\nu), P_2^{(C)}(\nu), \dots, P_m^{(C)}(\nu))$.

2.3. Unconditioned distributions

To facilitate comparison with the later conditioned results, in this section we give the standard forward and backward equations for the PGF G in the unconditioned case. Derivations and further details on these results can be found in many sources, e.g. [6, pp. 183–184, 199–201], [7], [14, pp. 35–36, 113–114], and [20, pp. 8–10].

In discrete time, the forward recursive equation for the PGF $G(s; n)$ is

$$G(s; n) = G(F(s); n - 1) \tag{5}$$

while the backward equation is

$$G(s; n) = F(G(s; n - 1)). \tag{6}$$

Note that the forward equation (5) is ‘decoupled’ in the sense that it can be applied to the k th component of G individually, while the backward equation (6) requires one to solve for

the entire vector G simultaneously [7], [14]. Both recursions can be solved from the initial condition $G(s; 0) = s$ to obtain

$$G(s; n) = F^{(n)}(s),$$

where the superscript ‘ (n) ’ denotes n -fold iteration of the function.

To derive the continuous-time results, one can substitute the offspring distribution PGF from (1) into the forward recursive equation (5), expand in powers of δt , and rearrange, taking the limit $\delta t \rightarrow 0$ to obtain the partial differential equation

$$\frac{\partial G_k(s; t)}{\partial t} = \sum_{i=1}^m \xi_i (f_i(s) - s_i) \frac{\partial G_k(s; t)}{\partial s_i}. \tag{7}$$

Likewise, for the backward equation, we obtain the ordinary differential equation

$$\frac{\partial G_k(s; t)}{\partial t} = \xi_k (f_k(G(s; t)) - G_k(s; t)).$$

Again, both systems of equations have the initial condition $G(s; 0) = s$.

3. Conditioned processes: discrete-time equations

Recall that $\tilde{G}(s; n)$ denotes the PGF for $X(n)$, conditioned on no type m appearing up to time n . Comparably to the unconditioned case, we derive recursive equations for \tilde{G} through both forward and backward approaches. In parallel, we derive recursive equations for the probability $P(n)$ that type m has not appeared by time n . Derivations are given for a process starting from one individual of a given type. However, by the independence of processes, the results extend in a straightforward manner to the case of multiple progenitors. Specifically, if the process starts with x_i progenitors of type i ($i = 1, \dots, m - 1; x_m = 0$) then the overall probability that type m does not appear by time n is

$$\prod_{i=1}^{m-1} P_i(n)^{x_i},$$

while the PGF for the number of individuals of each type in the process at time n , conditioned on no type m appearing up to this time, is

$$\prod_{i=1}^{m-1} \tilde{G}_i(s; n)^{x_i}.$$

3.1. Forward derivations

Forward derivations represent what happens in the last time step of the process. We first derive a decoupled recursion expressing $\tilde{G}(s; n)$ in terms of $\tilde{G}(s; n - 1)$, then write the probability that type m does not appear by a given time in terms of this conditional PGF.

Theorem 1. *The conditional PGF $\tilde{G}_k(s; n)$ ($k \neq m$) can be expressed recursively as*

$$\tilde{G}_k(s; n) = \frac{\tilde{G}_k(F_*(s_*); n - 1)}{\tilde{G}_k(F_*(1_*); n - 1)}$$

with explicit solution

$$\tilde{G}_k(s; n) = \frac{F_k(F_*^{(n-1)}(s_*))}{F_k(F_*^{(n-1)}(1_*))},$$

where $F_*^{(n)}$ denotes the n -fold iteration of F_* .

Proof. By definition,

$$\tilde{G}_k(s; n) = \sum_x \Pr_k(X(n) = x \mid X_m(v) = 0, v \leq n) \prod_i s_i^{x_i}.$$

We first shift the conditioning at time n to go along with the numbers at time n , and then condition on the state at time $n - 1$, summing over all possible such states. The Markov property implies that, given $X(n - 1)$, we can drop all earlier conditions at time n . Thus,

$$\begin{aligned} \tilde{G}_k(s; n) &= \sum_x \frac{\Pr_k(X(n) = x \cap X_m(n) = 0 \mid X_m(v) = 0, v \leq n - 1)}{\Pr_k(X_m(n) = 0 \mid X_m(v) = 0, v \leq n - 1)} \prod_i s_i^{x_i} \\ &= \frac{\sum_{\{x: x_m=0\}} \Pr_k(X(n) = x \mid X_m(v) = 0, v \leq n - 1) \prod_i s_i^{x_i}}{\sum_{\{x: x_m=0\}} \Pr_k(X(n) = x \mid X_m(v) = 0, v \leq n - 1)} \\ &= \frac{\sum_y \Pr_k(X(n-1)=y \mid X_m(v)=0, v \leq n-1) \sum_{\{x: x_m=0\}} \Pr(X(n)=x \mid X(n-1)=y) \prod_i s_i^{x_i}}{\sum_y \Pr_k(X(n-1)=y \mid X_m(v)=0, v \leq n-1) \sum_{\{x: x_m=0\}} \Pr(X(n)=x \mid X(n-1)=y)}. \end{aligned}$$

In the numerator, the sum over x is the PGF for the numbers at time n , given those at time $n - 1$, evaluated at s_* . Applying (2) (with no condition imposed on the offspring), this PGF is $\prod_i F_i(s_*)^{y_i}$. In the denominator, we have the same PGF but evaluated at 1_* . That is,

$$\tilde{G}_k(s; n) = \frac{\sum_y \Pr_k(X(n - 1) = y \mid X_m(v) = 0, v \leq n - 1) \prod_i F_i(s_*)^{y_i}}{\sum_y \Pr_k(X(n - 1) = y \mid X_m(v) = 0, v \leq n - 1) \prod_i F_i(1_*)^{y_i}}.$$

The probabilities in the sums over y are simply the coefficients in $\tilde{G}_k(s; n - 1)$. Thus,

$$\tilde{G}_k(s; n) = \frac{\tilde{G}_k(F_*(s_*); n - 1)}{\tilde{G}_k(F_*(1_*); n - 1)},$$

as given in Theorem 1. Iteration of this recursion from initial conditions $\tilde{G}_*(s; 0) = s_*$ yields the explicit solution.

Theorem 2. *The probability $P_k(n)$ that type m has not yet appeared by time n in the lineage of a type- k progenitor can be expressed via a recursion as*

$$P_k(n) = P_k(n - 1)\tilde{G}_k(P(1); n - 1)$$

for $k \neq m$, with $P_m(n) \equiv 0$; or explicitly as

$$P(n) = F_*^{(n)}(1_*).$$

Proof. Rewrite $P_k(n)$ by conditioning on what has happened up to the previous generation:

$$\begin{aligned} P_k(n) &= \Pr_k(\text{no } m \text{ by } n - 1) \Pr_k(\text{no } m \text{ by } n \mid \text{no } m \text{ by } n - 1) \\ &= P_k(n - 1) \Pr_k(\text{no individual at } n - 1 \text{ produces } m \text{ in next step} \mid \text{no } m \text{ by } n - 1). \end{aligned}$$

Given that no type m has arisen by time step $n - 1$, the distribution of individuals in time step $n - 1$ is given by $\tilde{G}_k(s; n - 1)$, and each type- i individual independently fails to produce type m in the next step with probability $P_i(1)$. Thus, applying (4),

$$P_k(n) = P_k(n - 1)\tilde{G}_k(P(1); n - 1),$$

valid for $k \neq m$, while $P_m(n) \equiv 0$ since starting with type m guarantees that it has been present. We can use the solution for $\tilde{G}_k(s; n - 1)$ (Theorem 1) along with $P(0) = 1_*$ and $P(1) = F_*(1_*)$ to iterate this recursion and obtain $P(n)$ explicitly.

3.2. Backward derivations

Backward derivations proceed by considering what happens in the first time step of the process. In a reversal of the order in the forward derivation, we can now obtain the probability that type m does not appear by time n without first needing the conditional PGF \tilde{G} . We use this result to derive an expression for the offspring distribution conditioned on no type m appearing in the progenitor’s lineage for n time steps to come. Finally, we use these results in the derivation of a coupled backward recursion for $\tilde{G}(s; n)$.

Theorem 3. *The probability $P_k(n)$ that type m has not yet appeared by time n in the lineage of a type- k progenitor satisfies the recursion*

$$P_k(n) = F_k(P(n - 1))$$

for $k \neq m$, with $P_m(n) \equiv 0$; with explicit solution

$$P(n) = F_*^{(n)}(1_*).$$

Proof. We express $P_k(n)$ by conditioning on the first time step:

$$P_k(n) = \Pr_k(\text{no } m \text{ at time } 1)\Pr_k(\text{no } m \text{ by time } n \mid \text{no } m \text{ at time } 1).$$

The second factor is the probability that every individual existing at time 1 independently fails to produce type m within $n - 1$ more time steps. Since the conditional distribution at time 1 is $\tilde{G}_k(s; 1) \equiv \tilde{F}_k(s)$, (4) implies that the probability that type m does not arise in any lineage by time n is $\tilde{F}_k(P(n - 1))$. That is,

$$P_k(n) = P_k(1)\tilde{F}_k(P(n - 1)).$$

Using $\tilde{F}_k(s) = F_k(s_*)/P_k(1)$ ((3)), the expression then simplifies to

$$P_k(n) = F_k(P(n - 1)),$$

valid for $k \neq m$, while $P_m(n) \equiv 0$. The recursion can be iterated, from initial conditions $P(0) = 1_*$, to obtain the explicit solution.

Theorem 4. *Let $\tilde{F}_k(s; n) \equiv \tilde{F}_k(s_*; n)$ denote the PGF of the offspring distribution for a type- $(k \neq m)$ progenitor, conditioned on no type m arising in the first $n \geq 1$ time steps of its lineage to come. Then*

$$\tilde{F}_k(s; n) = \frac{F_k(s_* \cdot P(n - 1))}{P_k(n)},$$

where the dot denotes componentwise multiplication.

Proof. By definition,

$$\tilde{F}_k(s; n) := \sum_y \Pr_k(X(1) = y \mid X_m(v) = 0, 1 \leq v \leq n) \prod_i s_i^{y_i}.$$

We shift the conditioning to write

$$\begin{aligned} \tilde{F}_k(s, n) &= \sum_y \frac{\Pr_k(X(1) = y \cap X_m(v) = 0, 1 \leq v \leq n)}{\Pr_k(X_m(v) = 0, 1 \leq v \leq n)} \prod_i s_i^{y_i} \\ &= \frac{\sum_y \Pr_k(X(1)=y \cap X_m(1)=0) \Pr(X_m(v)=0, 2 \leq v \leq n \mid X(1)=y \cap X_m(1)=0) \prod_i s_i^{y_i}}{\Pr_k(X_m(v)=0, 1 \leq v \leq n)}, \end{aligned}$$

or, since each lineage initiated at time 1 independently fails to produce type m ,

$$\begin{aligned} \tilde{F}_k(s; n) &= \frac{\sum_{\{y: y_m=0\}} \Pr_k(X(1) = y) \prod_{i=1}^{m-1} (P_i(n-1) \times s_i)^{y_i}}{P_k(n)} \\ &= \frac{F_k(s_* \cdot P(n-1))}{P_k(n)}. \end{aligned}$$

Note that this formula simplifies to (3) for $n = 1$.

Theorem 5. *The PGF for the number of individuals in a type- k ($k \neq m$) progenitor’s lineage after n time steps, conditioned on no type m arising during this time, satisfies the recursive equation*

$$\tilde{G}_k(s; n) = \frac{F_k(\tilde{G}_*(s; n-1) \cdot P(n-1))}{P_k(n)}$$

with explicit solution

$$\tilde{G}_k(s; n) = \frac{F_k(F_*^{(n-1)}(s_*))}{F_k(F_*^{(n-1)}(1_*))}.$$

Proof. The PGF for the number of individuals at time 1, conditioned on no type m appearing up to time n , is $\tilde{F}_k(s; n)$. A type- i individual existing at time 1 gives rise to individuals at time n according to the PGF $\tilde{G}_i(s; n-1)$, also conditioned on no type m appearing within this time. Using results for compound distributions analogous to those presented in Section 2.2.1, we have

$$\tilde{G}_k(s; n) = \tilde{F}_k(\tilde{G}(s; n-1); n).$$

Applying Theorem 4 to substitute the n -step conditional offspring distribution yields the recursive formula for \tilde{G} . Note the coupling in this equation requires us to solve for the entire vector \tilde{G}_* simultaneously. Iteration from initial conditions $\tilde{G}_*(s; 0) = s_*$ yields the explicit solution, in agreement with that from the forward approach.

4. Conditioned processes: continuous-time equations

We now consider results for a Markovian branching process in continuous time. Analogously to the discrete case, we would like to find $\tilde{G}(s; t)$, the PGF for the number of individuals of each type, $X(t)$, conditioned on no type m appearing up to time t . We consider $X(t)$ (and its PGF) to be a limit of discrete-time processes, where we take one discrete-time step to be of length δt in continuous time, and take $\delta t \rightarrow 0$ while number of steps $n \rightarrow \infty$ simultaneously,

such that the time point $n\delta t$ is fixed at t . In a slight abuse of notation, we write $\tilde{G}(s; n)$, with the time argument in the number of steps, to be equivalent to $\tilde{G}(s; t)$, with the time argument in continuous time. We ‘translate’ between these representations when the meaning is clear from the context. Analogously to the continuous-time results for the unconditioned case, we derive both a system of decoupled partial differential equations (PDEs) from the forward approach and a system of coupled ordinary differential equations (ODEs) from the backward approach for $\tilde{G}(s; t)$. We also derive two systems of ODEs for $P(t)$, the probability that type m has not appeared by time t : one from the forward approach and one from the backward approach. Although analytical solutions are not generally available, in Appendix B.1 we provide numerical solutions showing agreement of the two solution methods.

4.1. Forward derivations

Theorem 6. *The PGF $\tilde{G}_k(s; t)$ for the number of individuals in a type- $(k \neq m)$ progenitor’s lineage after time t , conditioned on no type m arising during this time, satisfies the equation*

$$\frac{\partial \tilde{G}_k(s; t)}{\partial t} = \sum_{i=1}^{m-1} \xi_i(f_i(s_*) - s_i) \frac{\partial \tilde{G}_k(s; t)}{\partial s_i} + \tilde{G}_k(s; t) \sum_{i=1}^{m-1} \xi_i(1 - f_i(1_*)) \frac{\partial \tilde{G}_k(s; t)}{\partial s_i} \Big|_{s=1}$$

with initial condition $\tilde{G}_k(s; 0) = s_k$.

Proof. We start with the discrete-time difference equation, i.e.

$$\Delta \tilde{G}_k(s; n) := \tilde{G}_k(s; n) - \tilde{G}_k(s; n - 1),$$

and rewrite $\tilde{G}_k(s; n)$ using the forward recursion from Theorem 1. Then substituting (1) for F_* , we have

$$\tilde{G}_k(s; n) = \frac{\tilde{G}_k(s_* + \xi \cdot (f_*(s_*) - s_*)\delta t + o(\delta t); n - 1)}{\tilde{G}_k(1_* + \xi \cdot (f_*(1_*) - 1_*)\delta t + o(\delta t); n - 1)},$$

where the dot denotes componentwise multiplication. We then expand both the numerator and denominator in a Taylor series about s_* and 1_* , respectively. (Recall that $\tilde{G}_k(s_*; n) \equiv \tilde{G}_k(s; n)$, and, in particular, $\tilde{G}_k(1_*; n) = 1$.) This yields

$$\tilde{G}_k(s; n) = \frac{\tilde{G}_k(s; n - 1) + \delta t \sum_{i=1}^{m-1} \xi_i(f_i(s_*) - s_i) \partial \tilde{G}_k(s; n - 1) / \partial s_i + o(\delta t)}{1 + \delta t \sum_{i=1}^{m-1} \xi_i(f_i(1_*) - 1) \partial \tilde{G}_k(s; n - 1) / \partial s_i |_{s=1} + o(\delta t)}.$$

Further simplifying this fraction as a power series and cancelling $\tilde{G}_k(s; n - 1)$ in the difference equation, we have

$$\begin{aligned} \Delta \tilde{G}_k(s; n) &= \delta t \sum_{i=1}^{m-1} \xi_i(f_i(s_*) - s_i) \frac{\partial \tilde{G}_k(s; n - 1)}{\partial s_i} \\ &\quad + \tilde{G}_k(s; n - 1) \delta t \sum_{i=1}^{m-1} \xi_i(1 - f_i(1_*)) \frac{\partial \tilde{G}_k(s; n - 1)}{\partial s_i} \Big|_{s_*=1_*} + o(\delta t). \end{aligned}$$

Dividing through by δt and taking the limit as $\delta t \rightarrow 0$ (and $n \rightarrow \infty$, such that $n\delta t \equiv t$) yields the desired equation. This is technically a functional PDE, due to the dependence of the equation at all points on the derivative of the solution evaluated at a particular point.

Note that if a particular type i is guaranteed not to produce type- m offspring, then $f_i(s_*) = f_i(s)$ and $1 - f_i(1_*) = 0$. Therefore, the second term disappears for any such type. If every type $i \neq m$ is guaranteed not to produce type- m offspring, the equation reduces to the standard unconditioned forward equation (7) for an $(m - 1)$ -type process.

Theorem 7. *The probability $P_k(t)$ that no type m arises in the lineage of a type- $(k \neq m)$ progenitor by time t satisfies the equation*

$$\frac{dP_k(t)}{dt} = -P_k(t) \sum_{i=1}^{m-1} \xi_i (1 - f_i(1_*)) \left. \frac{\partial \tilde{G}_k(s; t)}{\partial s_i} \right|_{s=1}$$

with initial conditions $P(0) = 1_*$; while $P_m(t) \equiv 0$.

Proof. We use the discrete-time forward recursion from Theorem 2, translated to continuous time as

$$P_k(t) = P_k(t - \delta t) \tilde{G}_k(1_* + \delta t \xi \cdot (f_*(1_*) - 1_*) + o(\delta t); t - \delta t).$$

Expanding \tilde{G}_k in a Taylor series about 1_* (noting that $\tilde{G}_k(1_*; t) = 1$), rearranging terms, and taking $\delta t \rightarrow 0$ yields the above ODE.

Note that $\partial \tilde{G}_k(s; t) / \partial s_i |_{s=1}$ is the expected number of type- i individuals in the process at time t (conditioned on no type m having yet appeared), ξ_i is the rate at which a type- i 's lifetime ends, and $1 - f_i(1_*)$ is the probability that a type i produces a type m when its lifetime ends. Thus, the summation gives the expected total rate of production of type m at time t given that type m has not appeared previously; $P_k(t)$ is correspondingly reduced at this rate.

4.2. Backward derivations

Theorem 8. *The probability $P_k(t)$ that no type m arises in the lineage of a type- $(k \neq m)$ progenitor by time t satisfies the equation*

$$\frac{dP_k(t)}{dt} = \xi_k (f_k(P(t)) - P_k(t))$$

with initial conditions $P(0) = 1_*$; while $P_m(t) \equiv 0$.

Proof. To obtain $P(t)$ by a backward approach, we use (1) to expand the discrete-time backward recursion on $P(n)$ given in Theorem 3. This yields the difference equation

$$\Delta P_k(n) := P_k(n) - P_k(n - 1) = \delta t \xi_k (f_k(P(n - 1)) - P_k(n - 1)) + o(\delta t).$$

Dividing through by δt , taking $\delta t \rightarrow 0$ and $n \rightarrow \infty$ yields the above ODE.

Theorem 9. *The PGF $\tilde{G}_k(s; t)$ for the number of individuals in a type- $(k \neq m)$ progenitor's lineage after time t , conditioned on no type m arising during this time, satisfies the equation*

$$\frac{\partial \tilde{G}_k(s; t)}{\partial t} = \frac{\xi_k}{P_k(t)} (f_k(\tilde{G}_*(s; t) \cdot P(t)) - \tilde{G}_k(s; t) f_k(P(t)))$$

with initial condition $\tilde{G}_k(s; 0) = s_k$.

Proof. From the discrete-time backward recursion in Theorem 5, we can write the difference equation

$$\begin{aligned} \Delta \tilde{G}_k(s; n) &:= \tilde{G}_k(s; n) - \tilde{G}_k(s; n - 1) \\ &= \frac{1}{P_k(n)} (F_k(\tilde{G}_*(s; n - 1) \cdot P(n - 1)) - \tilde{G}_k(s; n - 1)P_k(n)). \end{aligned}$$

Substituting (1) for F_k yields

$$\begin{aligned} \Delta \tilde{G}_k(s; n) &= \frac{1}{P_k(n)} (\tilde{G}_k(s; n - 1)P_k(n - 1) \\ &\quad + \xi_k(f_k(\tilde{G}_*(s; n - 1) \cdot P(n - 1)) - \tilde{G}_k(s; n - 1)P_k(n - 1))\delta t \\ &\quad + o(\delta t) - \tilde{G}_k(s; n - 1)P_k(n)), \end{aligned}$$

or, rearranging,

$$\begin{aligned} \frac{\Delta \tilde{G}_k(s; n)}{\delta t} &= \frac{1}{P_k(n)} \left(\xi_k(f_k(\tilde{G}_*(s; n - 1) \cdot P(n - 1)) - \tilde{G}_k(s; n - 1)P_k(n - 1)) \right. \\ &\quad \left. - \tilde{G}_k(s; n - 1) \frac{\Delta P_k(n)}{\delta t} + \frac{o(\delta t)}{\delta t} \right). \end{aligned}$$

Taking $\delta t \rightarrow 0$ and $n \rightarrow \infty$, we have the ODE

$$\frac{\partial \tilde{G}_k(s; t)}{\partial t} = \frac{1}{P_k(t)} \left(\xi_k(f_k(\tilde{G}_*(s; t) \cdot P(t)) - \tilde{G}_k(s; t)P_k(t)) - \tilde{G}_k(s; t) \frac{dP_k(t)}{dt} \right).$$

Substituting $dP_k(t)/dt$ from Theorem 8 yields the desired result. To solve for \tilde{G} , we must couple this ODE with the ODE for $P(t)$ from Theorem 8.

5. Waiting time to first appearance

A key point of interest is to find the distribution of the waiting time for the first appearance of a particular type. Of course, this question is only relevant if it is possible for the type to appear in the process, thus excluding some decomposable processes. We now have all the ingredients for this distribution, and in this section simply make these results explicit. Equations are presented in discrete time, but analogous results hold in continuous time.

Let the random variable N represent the time until the first appearance of type m in a process. We can express the cumulative distribution function of N exactly in terms of the probability that type m has not yet appeared by a given time. We start from one type- k progenitor,

$$\Pr_k(N \leq n) = 1 - P_k(n),$$

where $P_k(n)$ is given by Theorem 2 or 3, and N takes on the value ∞ if type m never appears. This occurs with probability v_k , which is a fixed point of the equation for $P(n)$, i.e. $v = F_*(v)$, in agreement with an earlier result [22].

Provided $v_k < 1$, i.e. it is possible for type m to appear (which may fail in a decomposable process), we can also find the distribution of the waiting time conditioned on type m eventually appearing, i.e. N being finite:

$$\Pr_k(N \leq n \mid N < \infty) = \frac{\Pr_k(N \leq n)}{\Pr_k(N < \infty)} = \frac{1 - P_k(n)}{1 - v_k}.$$

If we start from multiple individuals, the overall waiting time for type m to appear is the minimum of the waiting time in each individual subprocess. If the initial distribution of individuals has PGF $\tilde{G}(s)$, then (4) implies that

$$\Pr_k(N \leq n) = 1 - \tilde{G}(P(n)).$$

In particular, we might consider a process that began with one type k and has not yet produced any type m by time n , and compute the probability that type m still has not arisen after a further ν time steps. In this case, the distribution of $X(n)$ is given by $\tilde{G}_k(s; n)$, and

$$\Pr_k(N > n + \nu \mid N > n) = \tilde{G}_k(P(\nu); n).$$

Alternatively, we can write

$$\Pr_k(N > n + \nu \mid N > n) = \frac{\Pr_k(N > n + \nu)}{\Pr_k(N > n)} = \frac{P_k(n + \nu)}{P_k(n)}.$$

The distribution of the waiting time for a ‘successful’ type m (one whose lineage escapes extinction), as considered previously in [23], can also be obtained through our approach. We must expand the type space to $2m$ types, counting ‘successful’ and ‘unsuccessful’ individuals of each original type separately, and derive offspring distributions of these new types conditioned on the fate of their lineages.

6. Biological examples

In this section we briefly present two biological models, one in discrete time and one in continuous time. These serve to provide specific examples of the sorts of questions one can investigate with the results presented above. The numerical methods are described in detail in Appendix B; all results were generated using MATLAB®.

6.1. Between-host evolutionary epidemiology model (discrete time)

A Galton–Watson branching process model has previously been developed to investigate the emergence of a novel pathogen in a host population with given contact structure [3]. Type m is typically considered to be the sole well-adapted (high fitness) strain. Individuals of interest are infected hosts, identified by the infecting pathogen strain. Briefly, the offspring distribution for a typical (i.e. all but the initial) host infected with pathogen strain i is

$$F_i(s) = \frac{g'(1 - T_i + T_i \sum_{j=1}^m \varepsilon_{ij} s_j)}{g'(1)},$$

where $g(z)$ is the host contact distribution (common to all types), and each contact is treated independently, with T_i the probability that strain i transmits to a contact (‘transmissibility’) and ε_{ij} the probability of i -to- j conversion (‘mutation’) in a transmitted infection.

6.1.1. *Probability distribution evolving over time.* Consider a two-type model with a Poisson distribution of infectious contacts (offspring). In Figure 1 we illustrate how the conditional probability distribution of X_1 , the number of type-1 infectives, changes over time, under various forms of conditioning (see also Appendix A.4). Observe the skewing of the distribution depending on our knowledge of what events have occurred up to the present time: for instance, if the higher-fitness type 2 has appeared at some point, there is a much lower probability that the process is extinct and a higher probability that there is a large number of type-1 individuals.

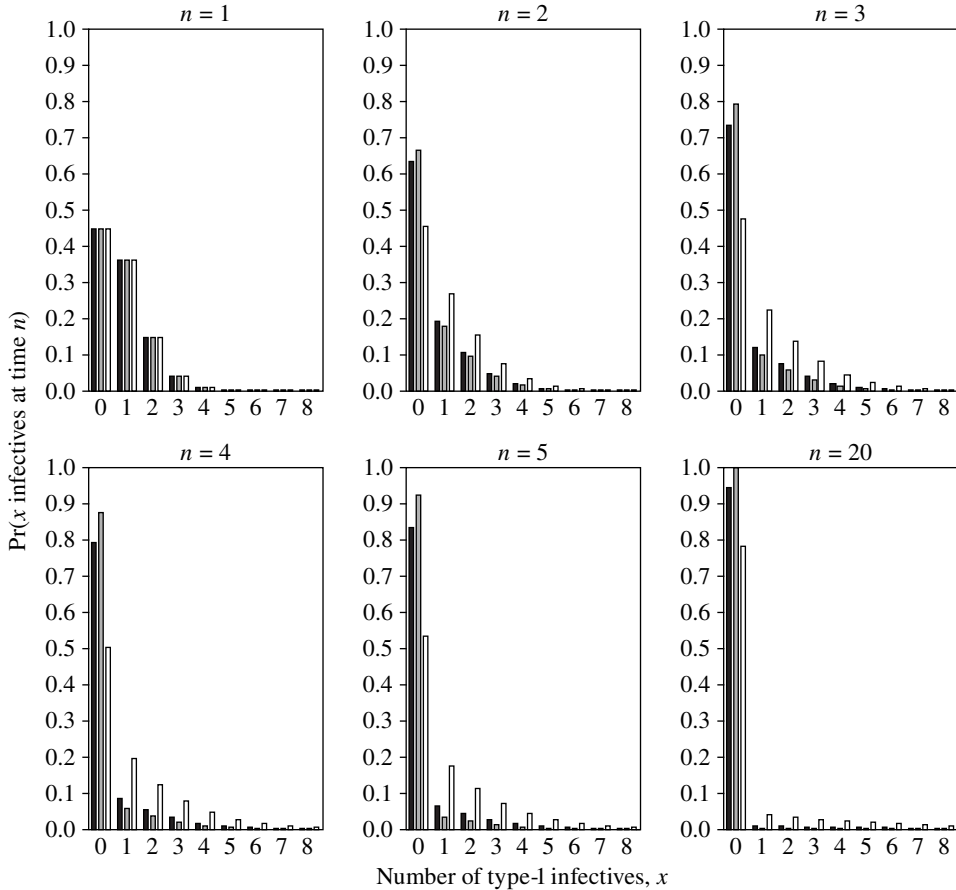


FIGURE 1: The distribution of $X_1(n)$, the number of type-1 infectives at generation n , for various n , and three cases of conditioning: no conditioning (*black bars*); conditioning on type 2 not having appeared by generation n (*grey bars*); or conditioning on type 2 having appeared at some point by generation n (*white bars*). The process is assumed to start from one type-1 individual. Both types have a Poisson offspring distribution (with mean 0.9 for type 1 and 1.1 for type 2), and mutations occur independently with probability 0.1 in each direction. The distribution is obtained by numerical inversion of the PGF $\tilde{G}_1(s; n)$ using a fast Fourier transform (see Appendix B.2).

6.1.2. *Impact of fitness valleys.* A key interest motivating this work is to find the distribution of the waiting time for the well-adapted strain m to appear. One might expect that in a case where emergence is less likely, the appearance of type m must occur more quickly if it is ever to appear. As an example, we compare fitness valleys on the path to strain m that vary in severity, either in the number of intermediate types or in the fitness of the intermediate types. We assume for simplicity that all intermediate strains have the same fitness. The qualitative result turns out to depend on how the fitness valley is made more severe. The deeper the fitness valley (i.e. the lower the intermediate strain fitness for a fixed number of intermediate strains), the less likely the appearance of type m , but the faster it tends to appear if it ever appears. The longer the fitness valley (i.e. the more intermediate strains for a fixed intermediate strain fitness), again the less likely the appearance of type m , but now the slower it tends to appear

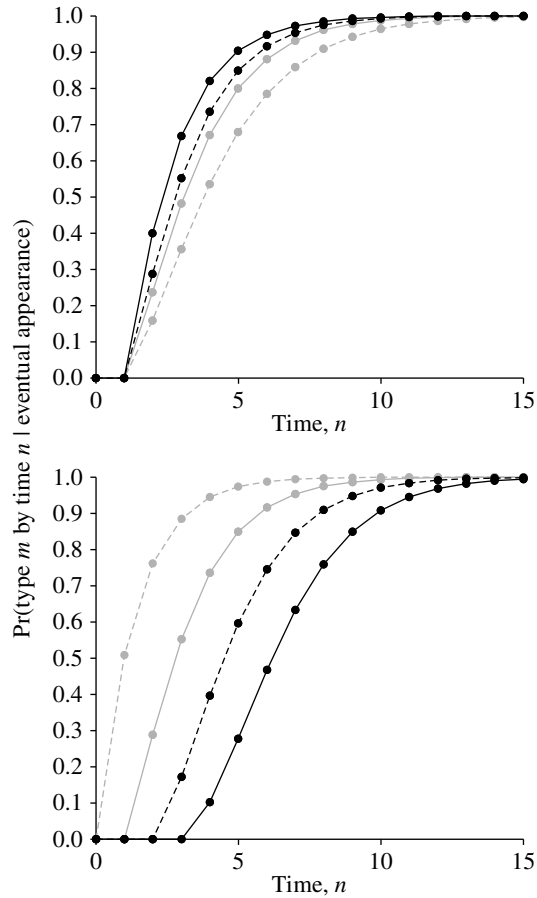


FIGURE 2: The impact of a fitness valley on the distribution of the waiting time for type m to appear, conditioned on eventual appearance. The host contact distribution is Poisson with mean 30. Mutations occur only from strain i to strain $i + 1$, with probability $\varepsilon = 0.1$ in each transmission. The top panel illustrates the impact of the depth of the fitness valley. There are three strains, with the transmissibilities of the first and last fixed ($T_1 = 0.02$ and $T_3 = 0.05$). The intermediate strain has transmissibility T_2 of 0.005 (black solid line), 0.015 (black dashed line), 0.02 (grey solid line), or 0.03 (grey dashed line). (The latter two cases are not fitness valleys in the biological sense, but are included for comparison.) The bottom panel illustrates the impact of length of the fitness valley, for a fixed intermediate strain fitness. Now, the number of intermediate strains varies: 3 (black solid line), 2 (black dashed line), 1 (grey solid line), or 0 (grey dashed line). (The last case, with no fitness valley, is a base case for comparison.) The transmissibilities of the strains are fixed: $T_1 = 0.02$, $T_i = 0.015$ for $i = 2, \dots, m - 1$, and $T_m = 0.05$.

if it does eventually appear. Here, it simply takes longer to step through all the intermediates, although it is conceivable that this trend could be reversed for different parameter ranges. In Figure 2 we compare these two types of fitness valleys.

6.2. Within-host multitype burst-death model (continuous time)

The burst-death model, a Markovian continuous-time process, was originally developed to describe a single-type population of organisms whose life history includes variable generation

times and a constant (possibly large) number of offspring produced at a ‘burst’ or proliferation event [15]. (The standard birth–death, or binary fission, process is identified as a special case with two offspring per burst.) Although previously applied to experimental evolution settings, one can also use this model to describe pathogen replication within a host. The model is readily generalized to include multiple types, with type-specific life history parameters, including an arbitrary distribution of the burst size, and an arbitrary mutation scheme among types; see [2] for details.

For the purposes of the example to follow, we consider a particular version of this general model. For each type i , denote the constant rate at which bursts occur by λ_i and the constant rate at which deaths occur by μ_i . Thus, a type- i individual has an exponentially distributed lifetime with total rate $\xi_i = \lambda_i + \mu_i$. With probability $\mu_i/(\lambda_i + \mu_i)$, the individual dies before generating a burst, and thus produces zero offspring. With probability $\lambda_i/(\lambda_i + \mu_i)$, a burst occurs first, and the individual produces a constant number B_i (the burst size) of offspring. Furthermore, suppose that mutation from type i to j occurs with probability ε_{ij} , and any mutation event affects *all* offspring. This model of mutation would apply well, for instance, to HIV: the most error-prone step in the replication cycle is the reverse transcription of its RNA into DNA when it first infects a cell. Progeny virions are then produced by high-fidelity transcription from the DNA and are thus likely to carry the same mutations (J. Heffernan and C. Beauchemin, personal communications). In this case, the offspring distribution has PGF

$$f_i(s) = \frac{\mu_i}{\lambda_i + \mu_i} + \frac{\lambda_i}{\lambda_i + \mu_i} \sum_{j=1}^m \varepsilon_{ij} s_j^{B_i}.$$

The Malthusian fitness of a type i , i.e. the exponential growth rate of the expected population size in a type- i process without mutation, can be defined as follows [15]:

$$\gamma_i = \lambda_i(B_i - 1) - \mu_i.$$

This measure is used below for comparing processes initiated by strains with different life histories, before any mutants have arisen.

6.2.1. Impact of life history. Previous work with the single-type burst-death model has shown that consistent differences in eventual extinction probability arise in populations with different life histories [4], [15]. In particular, for the same positive Malthusian fitness and burst size, a population with higher burst and death rates is more likely eventually to go extinct than one with lower burst and death rates. This finding motivates an interest in exploring life history differences in a multitype model, and attempting to better understand the probability of extinction based on the probability of evolution of new types.

Here, we begin investigating the effect of life history on the time course of extinction and evolution in a two-type model. Burst size is identical for both types, and mutations occur from type 1 to 2 only. Type 2 has high fitness due to a high burst rate and low death rate. We compare results for a fast-turnover type 1 (increased death rate relative to type 2) and a slow-turnover type 1 (reduced burst rate) with the same Malthusian fitness. This comparison is made at two fitness values for type 1, either greater than 0 (i.e. the type-1 population, in expectation, grows exponentially in the absence of mutation) or less than 0 (i.e. the type-1 population, in expectation, declines exponentially, and is guaranteed to go extinct in the absence of mutation).

One can gain insights into the time course of the process by plotting key probabilities changing over time (Figure 3). In general, the slower turnover strains are, not surprisingly,

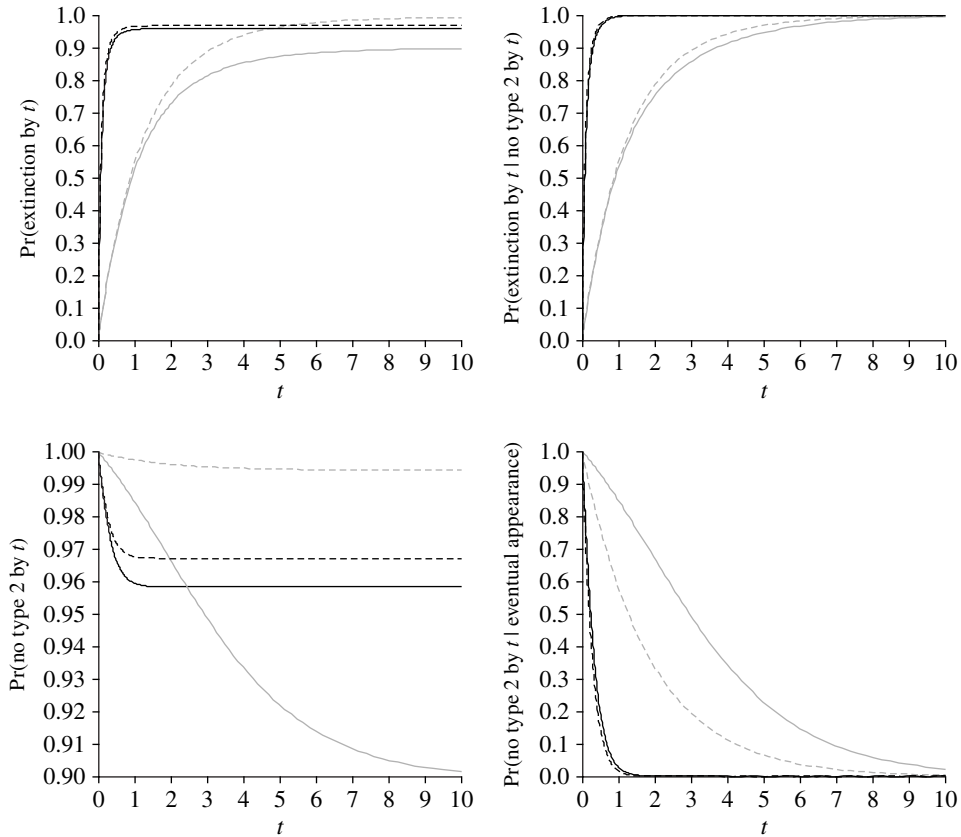


FIGURE 3: A comparison of key probabilities over time for different life histories in the two-type burst-death model. Burst size is $B = 11$ in all cases. Type 2 has fixed life history parameters: $\lambda_2 = 1$ and $\mu_2 = 0.8$ (thus, $\gamma_2 = 9.2$). We consider two values of type-1 fitness, negative (*dashed lines*; $\gamma_1 = -0.5$) and positive (*solid lines*; $\gamma_1 = 0.5$); and two versions of type 1 with different life history parameters: fast-turnover (*black lines*; $\lambda_1 = 1, \mu_1 = 10.5$ for low fitness or $\mu_1 = 9.5$ for high fitness) and slow-turnover (*grey lines*; $\mu_1 = 0.8, \lambda_1 = 0.03$ for low fitness or $\lambda_1 = 0.13$ for high fitness). Mutations (affecting all progeny) from type 1 to 2 occur with probability $\varepsilon = 0.1$ in all cases.

associated with a longer time course to extinction (top-left panel of Figure 3). However, the strain that ultimately approaches a higher extinction probability depends on the Malthusian fitness of type 1. If fitness is positive, the faster turnover strain is more likely eventually to go extinct, as in the single-type case. However, if fitness is negative, the pattern is reversed. It appears that when type 1 is not fit enough to survive without mutation, a faster turnover strain gains an advantage through increased production of type 2 that outweighs its greater propensity to die out. Indeed, if one ‘controls’ for this advantage by conditioning on type 2 not yet having arisen in the process (top-right panel), the slower turnover strain shows a lower extinction probability over time, regardless of the fitness value. The bottom panels show the distribution of time until type 2 appears, either unconditioned (left) or conditioned on eventual appearance (right). The probability that type 2 has not yet appeared initially drops off more quickly for faster turnover strains, but also levels out sooner due to the faster time course to extinction.

7. Discussion

The key focus of this paper has been to present the mathematical theory behind two closely related concepts in a multitype branching process: the distribution of population sizes conditioned on the nonappearance of a particular type, and the distribution of the waiting time for a particular type to appear. Although the latter question has been considered previously by Serra *et al.* [22], [23], [24], we took a different approach that provided an alternative conceptual framework. (It can be shown that the results of the two approaches, where comparable, agree; see [2].) The approach of [23] tracked the number of a particular ‘mutant’ type in the process, whereas we treated the appearance of the type of interest in a binary fashion. Furthermore, we linked the probability of nonappearance to the conditional distribution of type-specific population sizes.

To the author’s knowledge, the results of this work for the conditional PGFs for the population sizes are entirely novel. We have given both forward and backward derivations of these results, providing two conceptual approaches and a check on the results. One observes complementary patterns in the coupling versus decoupling of type-specific equations, and in the dependence of the conditional population size PGF on the probability of nonappearance or vice versa. Besides its mathematical interest, this dual approach may have practical implications, as each set of resulting formulae might yield more efficient computations in different circumstances.

An important point is that the equations presented here are exact. Although they can generally be solved only numerically, it is possible to obtain high accuracy under quite general conditions and for arbitrary parameter values. On the other hand, analytical approximations (such as those made in [22] and [24]), while providing some important insights into behaviour, may be valid only with certain model restrictions, and may show significant deviation from exact results for parameter values not sufficiently close to the limit (see, e.g. [24]).

A key direction for further work could be to apply these theoretical results to questions of interest in evolutionary biology and other fields. A few such questions were suggested in Section 6. One might also investigate the emergence of drug resistant or immune escape mutants, for instance, the waiting time until the appearance of a mutant, and the extent of disease spread dependent on its appearance or absence. In Section 6 we also illustrated some practical numerical approaches for applying the results. In particular, the technique of inverting PGF solutions to recover probability distributions is not widely used in biological modelling (see, however, [9], [10], [17], and [18]), but can be very useful in applications that consider the time course of a branching process and not only its long-term behaviour.

The present results are also promising for further theoretical development in connection with statistical inference methods, elucidating how to link particular branching process models with data [13]. For instance, estimates of model parameters or the time since a process began may be skewed by conditioning on particular events known to have occurred. One might also wish to infer the occurrence or timing of events: are there ‘red flags’ in observations of population numbers that suggest particular events have occurred? (See [5] and [11] for discussions of such warning signs.)

The theoretical results derived here thus represent a starting point for potentially widespread applications.

Appendix A. Generalizing the condition

In the main text, all results were derived for a particular condition, the absence of type m . To what extent can this condition be generalized? We present results only in discrete time; similar methods as used in Section 4 could be applied to extend the results to continuous time.

We initially suppose that the condition could be any restriction on the prevalence of type(s) and use the following notation: $C(\nu_1, \nu_2)$ denotes the event that condition C holds in the process in all time steps from ν_1 to ν_2 (inclusive), C_ν indicates the condition holding precisely at time ν , and \mathbb{N}_C denotes the subset of the state space \mathbb{N}_0^m on which condition C holds. We require that C_0 occurs with probability 1, i.e. the condition is guaranteed to hold initially, and that $C(1, n)$ occurs with positive probability, i.e. it is possible that the condition holds at every subsequent time step in the interval under consideration. Under these requirements, and supposing that the process starts with one type k , the probability that $C(1, n)$ holds is denoted by $P_k^{(C)}(n)$ and the PGF for $X(n)$ conditioned on $C(1, n)$ is denoted by $\tilde{G}_k^{(C)}(s; n)$.

A.1. Forward derivations

We derive a forward recursion for the conditional PGF $\tilde{G}_k^{(C)}(s; n)$ much as in Theorem 1:

$$\begin{aligned} \tilde{G}_k^{(C)}(s; n) &= \sum_{x \in \mathbb{N}_0^m} \Pr_k(X(n) = x \mid C(1, n)) \prod_i s_i^{x_i} \\ &= \sum_{x \in \mathbb{N}_0^m} \frac{\Pr_k(X(n) = x \cap C_n \mid C(1, n - 1))}{\Pr_k(C_n \mid C(1, n - 1))} \prod_i s_i^{x_i} \\ &= \frac{\sum_{x \in \mathbb{N}_C} \Pr_k(X(n) = x \mid C(1, n - 1)) \prod_i s_i^{x_i}}{\sum_{x \in \mathbb{N}_C} \Pr_k(X(n) = x \mid C(1, n - 1))} \\ &= \frac{\sum_{y \in \mathbb{N}_C} \Pr_k(X(n-1)=y \mid C(1, n-1)) \sum_{x \in \mathbb{N}_C} \Pr(X(n)=x \mid X(n-1)=y) \prod_i s_i^{x_i}}{\sum_{y \in \mathbb{N}_C} \Pr_k(X(n-1)=y \mid C(1, n-1)) \sum_{x \in \mathbb{N}_C} \Pr(X(n)=x \mid X(n-1)=y)}. \end{aligned}$$

The summation over x is the PGF for $X(n)$ given $X(n - 1)$, with no conditioning at time n , evaluated at some point(s) according to \mathbb{N}_C . The summation over y puts both the numerator and denominator in terms of the PGF $\tilde{G}_k^{(C)}(s; n - 1)$, achieving a recursive formula. At this point, the way in which we simplify the above expression depends on the condition C . However, C can theoretically be any condition on the prevalence of type(s), since we have separated out the conditioning on the final time step. In the absence of conditioning at time n , all individuals at time $n - 1$ place offspring at time n independently; we simply count only those outcomes that fall within \mathbb{N}_C . The preliminary result in (2) (without any conditioning applied to the offspring distribution) is useful here. Some examples are as follows.

- C is the absence of all types except type 1 ($\mathbb{N}_C = \mathbb{N}_0 \times \{0\}^{m-1}$). Then the summation over $x \in \mathbb{N}_C$ is the PGF for $X(n)$ given $X(n - 1)$, evaluated at $(s_1, 0, \dots, 0)$. Thus,

$$\begin{aligned} \tilde{G}_k^{(C)}(s; n) &= \frac{\sum_{y \in \mathbb{N}_C} \Pr_k(X(n - 1) = y \mid C(1, n - 1)) \prod_{i=1}^m (F_i(s_1, 0, \dots, 0))^{y_i}}{\sum_{y \in \mathbb{N}_C} \Pr_k(X(n - 1) = y \mid C(1, n - 1)) \prod_{i=1}^m (F_i(1, 0, \dots, 0))^{y_i}} \\ &= \frac{\tilde{G}_k^{(C)}(F_{1*}(s_{1*}); n - 1)}{\tilde{G}_k^{(C)}(F_{1*}(1_{1*}); n - 1)}, \end{aligned}$$

where s_{1*} denotes $(s_1, 0, \dots, 0)$ and F_{1*} denotes $(F_1, 0, \dots, 0)$.

- C is nonextinction ($\mathbb{N}_C = \mathbb{N}_0^m - \{0\}^m$). Then the summation over $x \in \mathbb{N}_C$ is best rewritten as the sum over all x minus the point at $x = 0$, yielding

$$\begin{aligned} &\tilde{G}_k^{(C)}(s; n) \\ &= \frac{\sum_{y \in \mathbb{N}_C} \Pr_k(X(n-1) = y \mid C(1, n-1)) (\prod_{i=1}^m (F_i(s))^{y_i} - \prod_{i=1}^m (F_i(0))^{y_i})}{\sum_{y \in \mathbb{N}_C} \Pr_k(X(n-1) = y \mid C(1, n-1)) (1 - \prod_{i=1}^m (F_i(0))^{y_i})} \\ &= \frac{\tilde{G}_k^{(C)}(F(s); n-1) - \tilde{G}_k^{(C)}(F(0); n-1)}{1 - \tilde{G}_k^{(C)}(F(0); n-1)}. \end{aligned}$$

- C is the condition that no more than a fixed number a of type- m individuals are present ($\mathbb{N}_C = \mathbb{N}_0^{m-1} \times \{0, 1, \dots, a\}$). This condition does not give such a tidy simplification of the recursive formula, but the formula could, in principle, still be used.

We can extend the forward derivation from Theorem 2 of $P^{(C)}(n) := \Pr(C(1, n))$, the probability that condition C holds at all time steps up to n , to arbitrary C , provided we can write an expression for $\Pr(C_n \mid C(1, n-1))$. (Note that a condition on the overall process at time n does not necessarily translate into an independent condition on each subprocess from time $n-1$ to n .) In general,

$$P_k^{(C)}(n) = P_k^{(C)}(n-1) \Pr_k(C_n \mid C(1, n-1)).$$

Some examples are as follows.

- For $C = \{\text{nonextinction}\}$, rewrite $\Pr_k(C_n \mid C(1, n-1))$ as

$$1 - \Pr_k(X(n) = 0 \mid X(n-1) \neq 0) = 1 - \tilde{G}_k^{(C)}(F(0); n-1).$$

- For $C = \{\text{nonextinction} \cap \text{no type } m\}$, rewrite $\Pr_k(C_n \mid C(1, n-1))$ as

$$\begin{aligned} &\Pr_k(X_m(n) = 0 \mid C(1, n-1)) - \Pr_k(X(n) = 0 \mid C(1, n-1)) \\ &= \tilde{G}_k^{(C)}(F_*(1_*); n-1) - \tilde{G}_k^{(C)}(F_*(0); n-1), \end{aligned}$$

i.e. the probability that none of the individuals present at time $n-1$ give rise to type m , minus the probability that they all have zero offspring. We could also have taken a shortcut using the probability $P_k(n)$ of having no type m by time n and the PGF $\tilde{G}_k(s, n)$ conditioned only on the absence of type m , to write

$$P_k^{(C)}(n) = P_k(n) (1 - \tilde{G}_k(0; n)),$$

i.e. the probability that type m has not arisen reduced by those cases in which the process has gone extinct.

- For $C = \{\text{no more than } a \text{ type-}m \text{ individuals}\}$, the expression for $\Pr_k(C_n \mid C(1, n-1))$ is not simple to write out for arbitrary a , since there is dependence among subprocesses in deciding whether C is satisfied.

A.2. Backward derivations

A backward derivation does not work out for a completely arbitrary condition C . To derive the PGF at time n given the numbers at time 1, we proceed similarly to Section 2.2.1 to obtain the expression

$$\mathbb{E} \left[\prod_i s_i^{X_i(n)} \mid X(1); C(1, n) \right] = \mathbb{E} \left[s_1^{\sum_{i=1}^m \sum_{j=1}^{X_i(1)} \tilde{Y}_1^{(ij)}} \dots s_m^{\sum_{i=1}^m \sum_{j=1}^{X_i(1)} \tilde{Y}_m^{(ij)}} \right],$$

where $\tilde{Y}_l^{(ij)}$ denotes the number of type- l individuals at time n in the lineage of the j th type- i individual present at time 1. However, depending on C , the $\tilde{Y}^{(ij)}$ may or may not be independent. For instance, given that the overall process is not extinct at time n , the event that individual A has no descendants is not independent of the event that individual B has no descendants at time n . Similarly, the derivation of the probability that C holds up to time n (cf. Theorem 3) does not hold for arbitrary C . However, if C is such that the $\tilde{Y}^{(ij)}$ are independent then the backward derivations proceed much as in the main text. This would be the case for the absence of any subset of types, since these types are absent in the main process if and only if every subprocess fails, independently, to produce these types. The results in the main text for $\tilde{G}(s; n)$ and $P(n)$ then carry over with simple redefinitions of s_* and F_* to replace the i th component with 0 if and only if type i is in the subset of interest. This extension is, of course, also subject to extending the restrictions on the process correspondingly: the initial type must not be in the subset of interest and no type outside the subset may produce offspring of types in the subset with probability 1.

A.3. The case of nonextinction

Conditioning on nonextinction is a special case with a simpler alternative derivation. This is because the presence of individual(s) at a particular time implies that individual(s) have been present at all previous times, so $C(1, n) = \bigcap_{v=1}^n C_v = C_n$. This allows us to express the distribution conditioned on nonextinction quite easily in terms of the distribution without this condition.

If we want to condition on C' , both the absence of type m and nonextinction, then we can shift the condition on nonextinction to rewrite the conditional probabilities:

$$\begin{aligned} \tilde{G}_k^{(C')}(s; n) &= \sum_x \Pr_k(X(n) = x \mid X(n) \neq 0; X_m(v) = 0, v \leq n) \prod_i s_i^{x_i} \\ &= \sum_x \frac{\Pr_k(X(n) = x \cap X(n) \neq 0 \mid X_m(v) = 0, v \leq n)}{\Pr_k(X(n) \neq 0 \mid X_m(v) = 0, v \leq n)} \prod_i s_i^{x_i} \\ &= \frac{\sum_x \Pr_k(X(n) = x \mid X_m(v) = 0, v \leq n) \prod_i s_i^{x_i} - \Pr_k(X(n) = 0 \mid X_m(v) = 0, v \leq n)}{1 - \Pr_k(X(n) = 0 \mid X_m(v) = 0, v \leq n)} \\ &= \frac{\tilde{G}_k(s; n) - \tilde{G}_k(0; n)}{1 - \tilde{G}_k(0; n)}. \end{aligned}$$

Here $\tilde{G}_k(s; n)$ is the PGF conditioned only on the absence of type m , which was derived in the main text. If we want to condition on nonextinction alone (C), we simply drop the conditioning on the absence of type m in the above derivation; then

$$\tilde{G}_k^{(C)}(s; n) = \frac{G_k(s; n) - G_k(0; n)}{1 - G_k(0; n)},$$

where $G_k(s; n)$ is the unconditioned PGF for $X(n)$.

A.4. Conditioning on the presence of a type

It is complicated (forward approach) or impossible (backward approach) to directly derive a PGF conditioned on the presence of a particular type at *all* times up to the present, due to the nonindependence of the condition among subprocesses. However, we can obtain the PGFs

$G(s; n)$ for $X(n)$ without any conditioning, and $\tilde{G}(s; n)$ for $X(n)$ conditioned on the absence of type m thus far. From these PGFs, we can indirectly obtain the PGF for $X(n)$ conditioned on the presence of type m at *some* time by n , say $\hat{G}(s; n)$. We do so by partitioning the set of all realizations into those in which type m appears by n and those in which it does not, and decomposing the expected value in the unconditioned case in terms of the contribution from each part:

$$\begin{aligned} G_k(s; n) &:= \mathbb{E}_k \left[\prod_i s_i^{X_i(n)} \right] \\ &= \Pr_k(X_m(v) = 0 \text{ for all } v \leq n) \mathbb{E}_k \left[\prod_i s_i^{X_i(n)} \mid X_m(v) = 0 \text{ for all } v \leq n \right] \\ &\quad + \Pr_k(X_m(v) > 0 \text{ for some } v \leq n) \mathbb{E}_k \left[\prod_i s_i^{X_i(n)} \mid X_m(v) > 0 \text{ for some } v \leq n \right] \\ &= P_k(n) \tilde{G}_k(s; n) + (1 - P_k(n)) \hat{G}_k(s; n). \end{aligned}$$

Rearranging,

$$\hat{G}_k(s; n) = \frac{G_k(s; n) - P_k(n) \tilde{G}_k(s; n)}{1 - P_k(n)}$$

provided $P_k(n) \neq 1$. (If $P_k(n) = 1$, type m is guaranteed not to appear by time n , and the PGF conditioned on its presence is not defined.) This formula is useful for comparing the distribution of population sizes if a particular type either has or has not arisen by a given time, as in Section 6.1.1.

Appendix B. Numerical methods

All numerical results were generated using MATLAB. In the discrete case, either the forward or backward recursions or the expressions for their solutions are easily iterated numerically from a given starting point. In the continuous case, the backward ordinary differential equations are also readily solved using the built-in MATLAB solver `ode45`, with tolerance reduced when necessary to obtain reasonably smooth solutions. A numerical solver was coded for the forward continuous-time equations; see Section B.1 below for details. We use this solver to demonstrate agreement between the forward and backward solutions, but all results in the main text use the backward method.

Derivations throughout this paper have dealt with the PGF \tilde{G} for the number of individuals X . However, in practice, the probability distribution of X (i.e. the coefficients in the power series representation of \tilde{G}) may be more readily interpreted or useful for applications. We therefore also consider the problem of numerical inversion of a PGF. This step is not widely used in connection with branching process problems, and we thus describe the approach in Section B.2.

B.1. Solutions of the continuous-time equations

On a theoretical basis, we expect corresponding solutions of both the forward and backward equations to agree. This agreement was confirmed by an explicit solution to both recursions in the discrete-time case, but in the continuous-time case, the differential equations are not generally analytically solvable. Thus, we must use numerical solution methods to check for agreement in specific test cases. For the backward PGF equation (Theorem 9), we implemented the built-in MATLAB ODE solver `ode45`. The form of the forward equation (Theorem 6), as

a functional PDE, limits the use of standard built-in PDE solvers, so we wrote a code using a ‘leapfrog’ scheme [25]. For simplicity, we restrict the comparison to the $m = 2$ case, meaning that one need only solve for $\tilde{G}_1(s_*; t) \equiv \tilde{G}_1(s_1, t)$ as a function of two scalar variables, s_1 and t (where $s_* \equiv (s_1, 0)$). The numerical solver for the forward equation approximates $\tilde{G}_1(s_1, t)$ at (s_1, t) grid points respectively spaced Δs and Δt apart. The derivatives with respect to both t and s_1 are replaced by central differences (approximations of second-order accuracy), except at boundary points: $\tilde{G}_1(s_1, \Delta t)$ is obtained from the initial conditions $\tilde{G}_1(s_1, 0) = s_1$ with a first-order forward approximation for $\partial\tilde{G}_1/\partial t$, and $\tilde{G}_1(0, t)$ is obtained with a first-order forward approximation for $\partial\tilde{G}_1/\partial s_1$. Furthermore, the numerical boundary condition $\tilde{G}_1(1, t) \equiv 1$ (which we know to be true for a PGF) is used, and the derivative at the $s_1 = 1$ boundary is approximated using a first-order backward difference. This scheme appears to yield numerically stable solutions, in which numerical error, although accumulating as time progresses, is reduced by decreasing the step size (Δs and Δt) [25]. We expect that a more sophisticated numerical PDE solver could further improve performance if one wished to use the forward equation alone in solving $\tilde{G}(s; t)$ for applications. However, for the present purpose, the simple leapfrog scheme is sufficient to show agreement between the forward and backward solutions in test cases.

All tests were run using the burst-death model with constant burst size (B) and all-or-none mutations (see Section 6.2). In Figures 4 and 5 we illustrate two examples, both for $B = 11$, one with burst rate (λ) larger than the death rate (μ) and the other vice versa. Observe that the

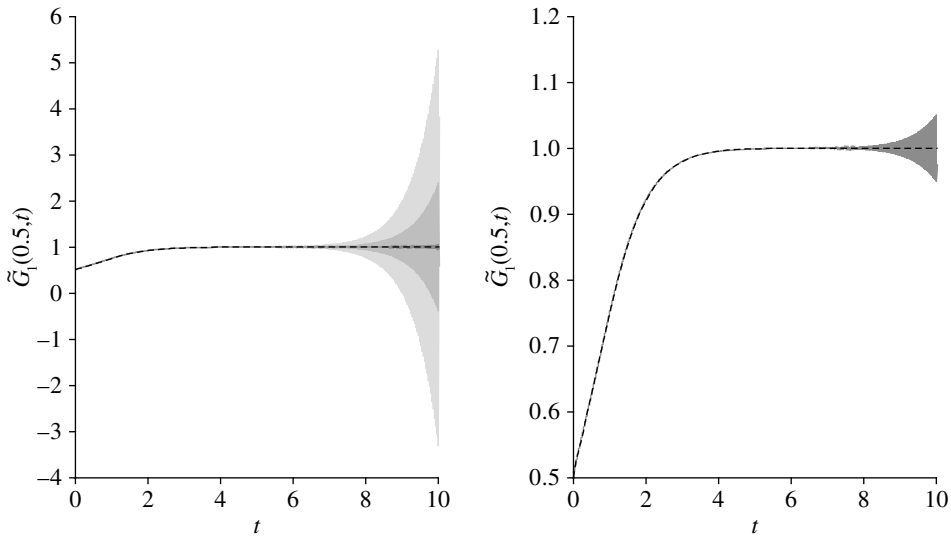


FIGURE 4: Comparison of forward and backward solutions over time t for the conditioned PGF $\tilde{G}_1(s_1, t)$, evaluated at $s_1 = 0.5$, in the two-type burst-death model. Type 1 has burst rate $\lambda_1 = 0.5$, death rate $\mu_1 = 1$, and constant burst size $B_1 = 11$; and all-or-none mutations occur with probability $\varepsilon = 0.1$ from type 1 to 2. The dashed black line indicates the numerical solution to the backward equation obtained with `ode45`, while the solid grey lines indicate numerical solutions to the forward equation obtained with the leapfrog scheme, darker grey corresponding to a smaller step size ($\Delta s = 0.01$ and $\Delta t = 0.005$; $\Delta s = 0.005$ and $\Delta t = 0.0025$; $\Delta s = 0.001$ and $\Delta t = 0.0005$). On the left, all step sizes are included to illustrate improving accuracy as the grid becomes finer; on the right, only the smallest step size is included, to more clearly illustrate the close agreement to the backward solution.

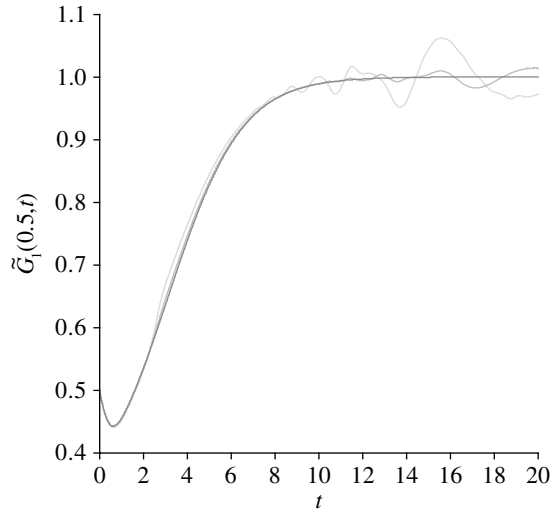


FIGURE 5: Comparison of forward and backward solutions over time t for the conditioned PGF $\tilde{G}_1(s_1, t)$, evaluated at $s_1 = 0.5$, in the two-type burst-death model. Parameters are $\lambda_1 = 0.5$, $\mu_1 = 0.1$, $B_1 = 11$, and $\varepsilon = 0.1$. Again, the dashed black line indicates the backward solution and the solid grey lines indicate forward solutions with the same step sizes as in the previous figure. Note that the smallest step size forward solution is almost indistinguishable from the backward solution.

magnitude of deviations from the backward solution increases as time goes on, but is reduced by using a finer grid. Solutions were generally more accurate for smaller values of λ and μ (not shown). Increasing B to 101 produced somewhat larger errors. Results are plotted at a fixed intermediate value of $s_1 = 0.5$, but broadly similar patterns (i.e. general agreement with some numerical errors accumulating over time; reduced step size yielding improved accuracy) appear at other values of s_1 . For $B = 11$, errors are smaller with larger s_1 , but this is not the case for $B = 101$.

B.2. Probability distributions via numerical inversion

When we are interested in the time course of a process, we may want to know the probability of having a given number of individuals at some finite time. Given an explicit probability generating function, one can hypothetically obtain these probabilities by differentiation; however, these expressions can quickly become messy. Furthermore, for many applications, we have only an equation that the PGF satisfies, not an analytical expression. That is, we can find only a numerical value of $G(s; t)$ at any given point $(s; t)$. Extracting the coefficients of the PGF, i.e. the probability distribution of $X(t)$, is a problem generally known as numerical inversion of transforms. It has been considered previously in the literature, as described in [1], which presents several methods as well as an error analysis. The idea is also described in the context of branching processes in [10]. However, as this method does not appear to be widely used, we describe the basic principle below.

For simplicity, we drop the time argument, but keep in mind that the random variable of interest can be time dependent. We restrict the exposition to a single-type PGF, which is applicable not only in a single-type process, but also to find the marginal distribution for the number of individuals of any given type in a multitype process.

Comparing the definition of a PGF for the distribution of a nonnegative, discrete random variable X ,

$$G(s) := \mathbb{E}[s^X]$$

with the definition of a characteristic function (CF) [12]

$$\phi(\theta) := \mathbb{E}[e^{i\theta X}],$$

we see that $\phi(\theta) = G(e^{i\theta})$. That is, the CF of the distribution of X is simply its PGF evaluated at $s = e^{i\theta}$. Furthermore, we can express the CF as

$$\phi(\theta) = \sum_{n=0}^{\infty} c_n e^{in\theta},$$

which is the form of a Fourier series [21], where the coefficient c_n is equal to $\Pr(X = n)$. To extract the desired probability distribution, one can take the Fourier transform of $\phi(\theta)$. Extensive work on Fourier transforms has already yielded effective numerical methods for doing so, even when the function ϕ is only known numerically. We make use of the fast Fourier transform (FFT), as described in [21], and implemented with the built-in MATLAB command `fft` (see the online MATLAB documentation at <http://www.mathworks.com/access/helpdesk/help/techdoc/>).

The basic procedure to apply the FFT method of numerical inversion is as follows.

1. Define the number of points to be used for the FFT, $N = 2^r$ for some integer r ($r = 12$ in numerical results presented here). A larger number of points improves accuracy but also increases the computations required [21].
2. Define θ as a vector of N points spaced $2\pi/N$ apart.
3. Define the CF $\phi(\theta)$ at each θ point.
4. Take the Fourier transform of ϕ using the command, ‘`Pr = fft(phi)/N`’, which returns the vector ‘Pr’ containing the desired coefficients, c_0 through c_{N-1} . (Note that numerical error may result in small residual imaginary parts in the coefficients; we ignored these imaginary parts, on the order of 10^{-16} or smaller in the tests, for plotting.)

To start, we tested this method on a few simple probability distributions known exactly. Thus, one can write the CF explicitly for use in step 3 above. Trials with uniform, binomial, and Poisson distributions show excellent agreement with the exact values of the desired probabilities (Figure 6). A slightly more complicated, but still analytically solvable, case is the single-type binary fission process (see Section 6.2; exact PGF for $X(t)$ given in [15]). In Figure 7 we show trials at particular times t , with a comparison to the first two coefficients, $c_0 = \Pr(X(t) = 0)$ and $c_1 = \Pr(X(t) = 1)$, computed from their exact expressions as derivatives of the PGF. Again, the FFT method matches the exact values very closely.

Finally, we consider numerical inversion when the PGF/CF is not known analytically, and can only be solved numerically. This makes step 3 of the FFT procedure more involved. In a discrete-time branching process, the PGF is found by numerically iterating a recursion, while in a continuous-time branching process, the PGF is found by the numerical solution of a differential equation. Using the backward equation, we obtain the CF by solving the ODE from the initial conditions of $e^{i\theta}$, repeated for each value of θ defined in step 2. Using the forward

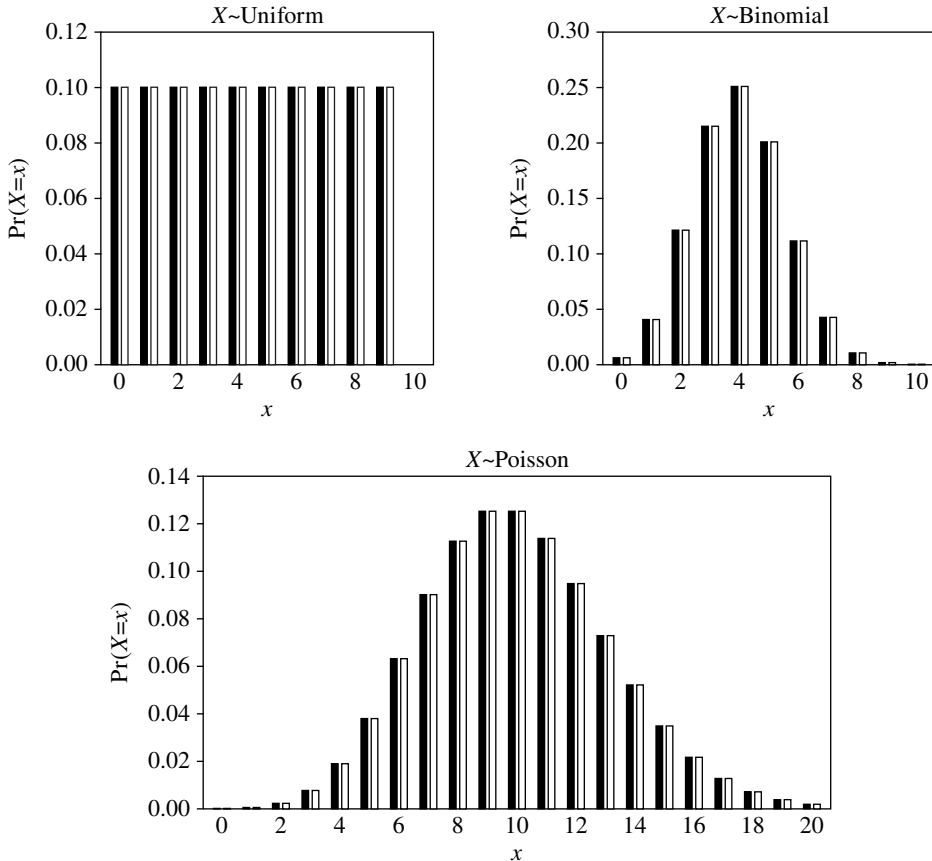


FIGURE 6: The results of the FFT method (*black bars*) in recovering a probability distribution from an explicit characteristic function, along with the exact probability distribution (*white bars*), for the following distributions: uniform on $\{0, 1, \dots, 9\}$; binomial with 10 trials and 0.4 probability of success; and Poisson with mean 10.

equation (not done here), one would first transform variables to write the PDE in terms of θ instead of s , then solve over all θ simultaneously. Though significantly more computationally intensive than the cases where we know the PGF/CF analytically, the solution process is still feasible. An example is given in Figure 1.

We repeated these tests using our code that implements a different method of inversion, given by Equation 5.38 of [1]. The results were visually indistinguishable from those using the FFT method.

Details of error analysis and computational efficiency for numerical inversions of transforms are discussed in [1], and should be taken into account for future use of this method. As also pointed out in [10], the FFT method is most effective when most probability mass is at small numbers, and breaks down as the chance of having a large population grows, i.e. as time progresses in a supercritical process. The FFT method can be extended to joint distributions, to consider the number of each type of individual in a multitype process (see [17] for an example). The built-in MATLAB command `fft` (n -dimensional Fourier transform) should be useful here.

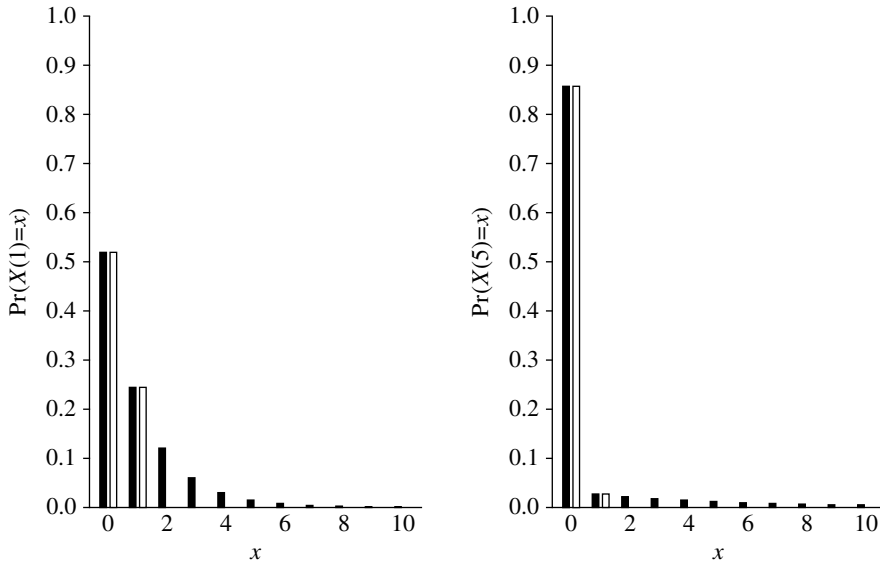


FIGURE 7: The results of the FFT method in recovering the probability distribution of $X(t)$, the number of individuals in a single-type binary fission process at time t . Parameters are birth rate $\lambda = 1$ and death rate $\mu = 1.05$. *Left*: $t = 1$; *right*: $t = 5$. The white bars show the exact values of the first two coefficients evaluated from their analytical expressions.

Acknowledgements

I wish to thank Troy Day for valuable advice and suggestions throughout the development of this research. Roland Regös, Tanja Stadler, and Glen Takahara provided additional helpful comments on drafts of this paper. I also thank an anonymous referee for further suggestions which improved this paper. I was supported in part by a Natural Sciences and Engineering Research Council of Canada (NSERC) postgraduate scholarship, and have conducted this work in both the Department of Mathematics and Statistics at Queen's University and the Theoretical Biology group at ETH Zürich.

References

- [1] ABATE, J. AND WHITT, W. (1992). The Fourier-series method for inverting transforms of probability distributions. *Queueing Systems* **10**, 5–87.
- [2] ALEXANDER, H. K. (2010). Modelling pathogen evolution with branching processes. Masters Thesis, Queen's University Kingston, Canada. Available at <http://hdl.handle.net/1974/5947>.
- [3] ALEXANDER, H. K. AND DAY, T. (2010). Risk factors for the evolutionary emergence of pathogens. *J. R. Soc. Interface* **7**, 1455–1474.
- [4] ALEXANDER, H. K. AND WAHL, L. M. (2008). Fixation probabilities depend on life history: fecundity, generation time and survival in a burst-death model. *Evolution* **62**, 1600–1609.
- [5] ARINAMINPATHY, N. AND MCLEAN, A. R. (2009). Evolution and emergence of novel human infections. *Proc. R. Soc. B* **276**, 3937–3943.
- [6] ATHREYA, K. B. AND NEY, P. E. (1972). *Branching Processes*. Springer, New York.
- [7] BARTLETT, M. S. (1951). The dual recurrence relation for multiplicative processes. *Math. Proc. Camb. Phil. Soc.* **47**, 821–825.
- [8] BECKER, N. (1977). Estimation for discrete time branching processes with application to epidemics. *Biometrics* **33**, 515–522.
- [9] CONWAY, J. M. AND COOMBS, D. (2011). A stochastic model of latently infected cell reactivation and viral blip generation in treated HIV patients. *PLoS Comput. Biol.* **7**, e1002033.

- [10] DORMAN, K. S., SINSHEIMER, J. S. AND LANGE, K. (2004). In the garden of branching processes. *SIAM Rev.* **46**, 202–229.
- [11] FERGUSON, N. M. *et al.* (2004). Public health risk from the avian H5N1 influenza epidemic. *Science* **304**, 968–969.
- [12] GRIMMETT, G. R. AND STIRZAKER, D. R. (1992). *Probability and Random Processes*, 2nd edn. Oxford University Press.
- [13] GUTTORP, P. (1991). *Statistical Inference for Branching Processes*. John Wiley, New York.
- [14] HARRIS, T. E. (1963). *The Theory of Branching Processes*. Springer, Berlin.
- [15] HUBBARDE, J. E., WILD, G. AND WAHL, L. M. (2007). Fixation probabilities when generation times are variable: the burst-death model. *Genetics* **176**, 1703–1712.
- [16] KIMMEL, M. AND AXELROD, D. E. (2002). *Branching Processes in Biology*. Springer, New York.
- [17] LANGE, K. (1982). Calculation of the equilibrium distribution for a deleterious gene by the finite Fourier transform. *Biometrics* **38**, 79–86.
- [18] LANGE, K. AND FAN, R. Z. (1997). Branching process models for mutant genes in nonstationary populations. *Theoret. Pop. Biol.* **51**, 118–133.
- [19] MACKEN, C. A. AND PERELSON, A. S. (1988). *Stem Cell Proliferation and Differentiation: A Multitype Branching Process Model* (Lecture Notes Biomath. **76**). Springer, Berlin.
- [20] MODE, C. J. (1971). *Multitype Branching Processes. Theory and Applications* (Modern Anal. Comput. Methods Sci. Math. **34**). American Elsevier Publishing, New York.
- [21] SAFF, E. B. AND SNIDER, A. D. (2003). *Fundamentals of Complex Analysis: with Applications to Engineering and Science*, 3rd edn. Pearson Education, Upper Saddle River, NJ.
- [22] SAGITOV, S. AND SERRA, M. C. (2009). Multitype Bienaymé–Galton–Watson processes escaping extinction. *Adv. Appl. Prob.* **41**, 225–246.
- [23] SERRA, M. C. (2006). On the waiting time to escape. *J. Appl. Prob.* **43**, 296–302.
- [24] SERRA, M. C. AND HACCOU, P. (2007). Dynamics of escape mutants. *Theoret. Pop. Biol.* **72**, 167–178.
- [25] STRIKWERDA, J. C. (2004). *Finite Difference Schemes and Partial Differential Equations*, 2nd edn. Society for Industrial and Applied Mathematics, Philadelphia, PA.
- [26] SUN, F. (1995). The polymerase chain reaction and branching processes. *J. Comput. Biol.* **2**, 63–86.