





THE NUMBER OF INDIVIDUALS ALIVE IN A BRANCHING PROCESS GIVEN ONLY TIMES OF DEATHS

FRANK BALL ,* ** AND
PETER NEAL ,* *** *University of Nottingham*

Abstract

The study of many population growth models is complicated by only partial observation of the underlying stochastic process driving the model. For example, in an epidemic outbreak we might know when individuals show symptoms to a disease and are removed, but not when individuals are infected. Motivated by the above example and the long-established approximation of epidemic processes by branching processes, we explore the number of individuals alive in a time-inhomogeneous branching process with a general phase-type lifetime distribution given only (partial) information on the times of deaths of individuals. Deaths are detected independently with a detection probability that can vary with time and type. We show that the number of individuals alive immediately after the k th detected death can be expressed as the mixture of random variables each of which consists of the sum of k independent zero-modified geometric distributions. Furthermore, in the case of an Erlang lifetime distribution, we derive an easy-to-compute mixture of negative binomial distributions as an approximation of the number of individuals alive immediately after the k th detected death.

Keywords: Phase-type distribution; negative binomial distribution; SIR epidemic models

2010 Mathematics Subject Classification: Primary 60J80
Secondary 92D30

1. Introduction

For many population growth models, both mathematical modelling and statistical inference are complicated by only partial observation of the underlying stochastic process driving the model. For example, in epidemic models we might have information on when individuals show symptoms of the disease and hence are removed from the active population, but rarely have information on when an individual is infected. Therefore, a question of interest at time t , say, is, given information on the times at which individuals exhibited symptoms (and were removed) up to time t , how many infectives are there in the population at time t ? This is the question considered and answered for the Markovian susceptible–infectious–removed (SIR) epidemic in Ball and Neal [2], where the distribution of the number of infectives was derived given only information on removal times assuming an exponential infectious period. In [2]

Received 23 January 2024; accepted 27 October 2024.

* Postal address: School of Mathematical Sciences, University of Nottingham, NG7 2RD, United Kingdom.

** Email address: frank.ball@nottingham.ac.uk

*** Email address: peter.neal@nottingham.ac.uk

© The Author(s), 2025. Published by Cambridge University Press on behalf of Applied Probability Trust.

a time-inhomogeneous birth–death process was considered as an approximation of the epidemic process, with the number of individuals alive in the birth–death process approximating the number of infectives in the epidemic process; cf. Whittle [11]. In particular, it was shown that the distribution of the number of individuals alive in a time-inhomogeneous birth–death process at time t , given the times of deaths of individuals up to and including time t , can be expressed as a mixture of negative binomial distributions.

The time-inhomogeneous birth–death process assumes that the rate μ_t at which an individual dies does not depend on how long the individual has been alive. In the case where $\mu_t = \gamma$ ($t \in \mathbb{R}$), this corresponds to individuals having exponentially distributed lifetimes with mean duration $1/\gamma$. However, exponentially distributed lifetimes (infectious periods) are unrealistic in most biological settings. Therefore phase-type distributions, and in particular, the Erlang distribution, are commonly chosen to model infectious period distributions in SIR epidemic models; see, for example, Walker *et al.* [10], Black and Ross [3], and House *et al.* [5].

In this paper we consider the situation where individuals' lifetimes are independent and identically distributed (i.i.d.) according to an arbitrary, but specified, phase-type distribution L ; see, for example, Chapter III of Asmussen [1]. Phase-type distributions are presented in Section 2, and L can be represented as the distribution of the time to absorption in a continuous-time Markov chain with J ($< \infty$) transient states and one absorbing state 0. Thus the case $J = 1$ corresponds to an exponential distribution. Special cases of the phase-type distribution include the hyper-exponential distribution (mixture of J (> 1) exponential distributions) and the Erlang distribution (sum of J i.i.d. exponential distributions). Moreover, the class of phase-type distributions is dense, so any lifetime distribution on $[0, \infty)$ can be approximated arbitrarily closely by a phase-type distribution and there is essentially no loss of generality in focussing on phase-type distributions.

We study time-inhomogeneous branching processes where all individuals alive at time t have birth rate β_t . We place no restriction on the form β_t can take, and therefore we can use the branching process model to approximate epidemic models with β_t denoting the effective infection rate at time t . For example, in the modelling of an emerging disease such as Covid-19 in a large population (e.g. a country), the infection rate β_t varies over time with the implementation and relaxation of control measures. We assume that individuals are potentially detected (observed) at death (removal in an epidemic context). We allow the detection probability to depend on both the time of death and the individual's type at death, where, during their lifetime, an individual's type is defined by the state j ($j = 1, 2, \dots, J$) of the lifetime Markov chain to which they belong. In an epidemic context we allow for asymptomatic individuals and more general under-reporting of infectious cases, which can fluctuate through time to account for differing levels of surveillance and disease awareness.

Given the times of observed deaths (removals) in the population, we show in Theorem 1 that the number of individuals alive in the population immediately after the k th detected death, X_k^* say ($k = 1, 2, \dots$), can be expressed as the mixture of $(k - 1)!$ random variables, each of which consists of the sum of k independent zero-modified geometric distributions; see Definition 1. Computing the mixture of $(k - 1)!$ random variables to obtain X_k^* using Theorem 1 is only possible for small k . However, under certain conditions it is shown in Corollary 1 that X_k^* can be expressed as a mixture of k negative binomial distributions, although computing the mixing weights is non-trivial. In Section 3.5.2, we present an approximation for X_k^* using a mixture of k negative binomial distributions with a simple-to-compute recursive relationship for X_k^* in terms of X_{k-1}^* and the proportion of individuals of each type at the $(k - 1)$ th detected death in

the spirit of [2, Theorem 3.2]. This approximation is applicable for Erlang lifetime distributions and is successfully applied in a simulation study in Section 7.

As noted above, the current work extends the results of [2]. Previous work in a similar vein but for a slightly different model can be found in Trapman and Bootsma [9], Lambert and Trapman [6], Lefèvre and Picard [7], and Lefèvre *et al.* [8]. In these papers it is assumed that each individual, whilst alive, can be detected at a constant rate, δ say. The papers [6, 9] focus on the distribution of individuals alive at the first detection time, with [6] allowing for a general lifetime distribution. The papers [7, 8] restrict attention to Markovian models but allow for epidemic dynamics and multiple births, and they study the size of the population at the j th detection ($j \geq 1$). Probability generating functions are also a key tool in [7, 8] and in themselves provide no insight as to the distribution of the size of the population (number of infectives) at a given detection time.

The remainder of the paper is structured as follows. In Section 2 we define the time-inhomogeneous phase-type branching process model. In Section 3, we present the main results (Theorem 1 and the approximation of Section 3.5.2), along with an overview of the exploration process (Section 3.2) between two time points t and $t + \tau$, which plays an important role in identifying how the population evolves between detected deaths. We also introduce key notation (Section 3.3). In Section 4 we present a more detailed analysis of the exploration process, including the computation of the key quantities required for Theorem 1. In Section 5 we prove the main result, Theorem 1. In Section 6 we use results from the proof of Theorem 1 to establish the approximation given in Section 3.5.2. Finally, in Section 7 we illustrate the approximation given in Section 3.5.2 through a numerical example. Further examples are provided in the supplementary material.

2. Phase-type branching process model

In this section we define the phase-type branching process model considered in this paper. We consider a time-inhomogeneous branching process where the birth rate, β_t , depends on t but all individuals have i.i.d. lifetimes distributed according to a phase-type distribution L . Examples of L include the exponential, hyper-exponential (mixture of exponentials), and Erlang distributions. We begin by considering the relationship between L and an associated Markov process.

Consider a Markov process with J transient states, labelled $1, 2, \dots, J$, and an absorbing state, labelled state 0. For $j = 1, 2, \dots, J$, let χ_j be the probability that the Markov process starts at time 0 in state j with $\sum_{i=1}^J \chi_i = 1$. The Markov process moves between the states $1, 2, \dots, J$ according to a sub-stochastic transition-rate matrix $\Lambda = (\lambda_{ij})$. That is, for all $j = 1, 2, \dots, J$, we have $\sum_{i \neq j} \lambda_{ji} \leq -\lambda_{jj}$, with strict inequality for at least one j . For $t \geq 0$, let

$$\mathbf{P}(t) = \exp(t\Lambda), \quad (2.1)$$

so $P_{ij}(t)$ is the probability that if the Markov process starts in state i at time 0 then it is in state j at time t . Since the states $1, 2, \dots, J$ are transient, we have that $\mathbf{P}(t) \rightarrow \mathbf{0}_{J \times J}$ ($J \times J$ matrix of zeros) as $t \rightarrow \infty$. The probability the Markov process has entered state 0 (absorption) by time t given that it starts in state i is $P_{i0}(t) = 1 - \sum_{j=1}^J P_{ij}(t)$. The distribution of the time to absorption, L , is a phase-type distribution with cumulative distribution function

$$\mathbb{P}(L \leq t) = \sum_{i=1}^J \chi_i P_{i0}(t) = 1 - \sum_{i=1}^J \chi_i \sum_{j=1}^J P_{ij}(t) \quad (t \geq 0) \quad (2.2)$$

and $\mathbb{P}(L \leq t) = 0$ ($t < 0$).

Let us turn to the branching process. Suppose that there is one initial individual in the population. We define time 0 to be the time at which the first detected death occurs in the population, so the initial individual enters the population sometime before time 0. Note that it is possible for the branching process to die out without any deaths being detected. However, we are interested in situations where at least one death is detected. We assume that there are J types of individual in the population, corresponding to the J states of the Markov process with transition matrix Λ to which L is associated. Therefore, for $t, \tau > 0$, an individual of type j at time t has probability $P_{jl}(\tau)$ of being an individual of type l at time $t + \tau$.

For all $t \in \mathbb{R}$, let β_t denote the birth rate at time t . We assume that there exists $\alpha_1 > 0$ such that $\beta_t = \alpha_1$ ($t \leq 0$). That is, there is a constant, positive birth rate up to and including the time of the first detected death. At time t , we assume that all individuals alive, irrespective of their type, give birth to new individuals at rate β_t . Similarly, for $j = 1, 2, \dots, J$ and $t \in \mathbb{R}$, let $d_{t,j}$ denote the probability that an individual of type j who dies at time t is detected with $\mathbf{d}_t = (d_{t,1}, d_{t,2}, \dots, d_{t,J})$. Therefore we allow the probability of a death being detected to depend on type, although it should be noted that, for $l = 1, 2, \dots, J$, $d_{t,l}$ is only required if $\gamma_l > 0$, where $\gamma_l = -\sum_{i=1}^J \lambda_{li}$ is the death rate of a type- l individual. We require at least one $j = 1, 2, \dots, J$, such that $d_{t,j}\gamma_j > 0$ to ensure that some deaths are detected. The proofs presented in this paper do not extend to the case where the birth rate is type-dependent. For $j = 1, 2, \dots, J$, we assume that there exists $0 \leq \epsilon_j \leq 1$ such that $d_{t,j} = \epsilon_j$ ($t \leq 0$). That is, there is a constant probability of detecting a death of a type- j individual up to, and including, the first detected death. The assumptions of time-homogeneity in the branching process and its observation process prior to the first detected death are in agreement with [2, Theorems 3.1 and 3.2] and ensure that the number of individuals alive immediately following the first detected death follows a geometric distribution.

Given the number of individuals of each type, the above time-inhomogeneous phase-type branching process is a Markov process. For $t \in \mathbb{R}$ and $j = 1, 2, \dots, J$, let $Y_j(t)$ denote the number of individuals of type j in the population at time t , and let $\mathbf{Y}(t) = (Y_1(t), Y_2(t), \dots, Y_J(t))$ and $Y^*(t) = \sum_{j=1}^J Y_j(t)$. Let S_0 denote the (unknown) time at which the initial individual enters the population, with $\mathbf{Y}(S_0)$ denoting the initial state of the population at time S_0 . Then $\mathbb{P}(\mathbf{Y}(S_0) = \mathbf{e}_j) = \chi_j$ ($j = 1, 2, \dots, J$), where \mathbf{e}_j is a vector of length J in which the j th element is equal to 1 and all other elements are 0. For $t > S_0$, the phase-type branching process satisfies

$$\mathbb{P}(\mathbf{Y}(t+h) = \mathbf{y} + \mathbf{v} | \mathbf{Y}(t) = \mathbf{y}) = \begin{cases} \left[\sum_{l=1}^J y_l \right] \beta_t \chi_j h + o(h), & \mathbf{v} = \mathbf{e}_j, \\ y_j \gamma_j h + o(h), & \mathbf{v} = -\mathbf{e}_j, \\ y_l \lambda_{lj} h + o(h), & \mathbf{v} = \mathbf{e}_j - \mathbf{e}_l \ (l \neq j), \\ 1 - \sum_{l=1}^J y_l (\beta_t - \lambda_{ll}) h + o(h), & \mathbf{v} = \mathbf{0}, \end{cases} \tag{2.3}$$

where all other events occur with probability $o(h)$. The events in (2.3) correspond to the birth of a type- j individual, the death of a type- j individual, the transition of an individual from type l to type j , and nothing happening.

Suppose that there exists $\alpha > 0$ such that $\beta_t = \alpha$ for all $t \in \mathbb{R}$. Then we have a time-homogeneous branching process where individuals have i.i.d. lifetimes distributed according to L and reproduce during their lifetime at the points of a homogeneous Poisson point process with rate α . Moreover, if there exists $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_J)$ such that $\mathbf{d}_t = \boldsymbol{\epsilon}$, for all $t \in \mathbb{R}$, then the observation process of the branching process is also time-homogeneous.

Similar to [2], the motivation for the phase-type branching process model is as an approximation for SIR epidemic models. Consider a SIR epidemic in a population of size N with

one initial infective, with infection rate at time t given by α_t (that is, we allow for a time-inhomogeneous infection rate), and with individuals having infectious periods that are i.i.d. according to L . At time t , if we can estimate the proportion of the population that is susceptible, s_t , the dynamics of the SIR epidemic can be approximated by a phase-type branching process with birth rate $\beta_t = \alpha_t s_t$, the rate at which infectious contacts are made with susceptible individuals. The assumption that, prior to the first detected death, the birth (infection) rate and detection probabilities are constant is not unreasonable, at least from an epidemic perspective. For example, until a disease is detected it is unlikely that any control measures will be put in place. In the supplementary material we discuss the scenario where the time at which the initial individual is born (infected) is known and results similar to those obtained in Theorem 1 can be derived without restrictions on β_t or \mathbf{d}_t ($t < 0$).

3. Main results

3.1. Introduction

In this section, we present the main results of the paper. In Theorem 1, presented in Section 3.4, we derive the distribution of the number of individuals of each type immediately following the k th detected death for $k = 1, 2, \dots$. The distribution obtained in Theorem 1 is a mixture of $(k - 1)!$ components, so is impractical for calculating the distribution of the number of individuals of each type for moderate k . However, it is informative about the distribution of the total number of individuals alive immediately following the k th detected death, which can be expressed as sums of zero-modified geometric random variables; see Definition 1 below.

Definition 1. Suppose that X is a non-negative discrete random variable with probability mass function

$$\begin{aligned}\mathbb{P}(X = 0) &= a, \\ \mathbb{P}(X = x) &= (1 - a)(1 - b)^{x-1}b \quad (x = 1, 2, \dots),\end{aligned}$$

for some $0 \leq a, b \leq 1$. Then we say that X has a zero-modified geometric distribution.

In Section 3.2 we present an overview of the exploration process; technical details of, and results for, the exploration process are presented in Section 4. This is followed in Section 3.3 by the introduction of the key notation required to state Theorem 1. The notation is defined in a similar manner to that of [2, Section 3], to allow for comparisons where appropriate. In Section 3.5 we provide an approximation for the distribution of the number of individuals of each type immediately following the k th detected death which consists of only k components.

3.2. Overview of exploration process

A key component in understanding the dynamics of a branching process through partial observation of its death process is to study the behaviour of the branching process between two given time points t and $t + \tau$ ($t \in \mathbb{R}$, $\tau > 0$). In particular, we consider the cases where (i) there are no detected deaths in the interval $(t, t + \tau]$, and (ii) the first detected death after time t is at time $t + \tau$. In this section we focus on (i) and use an approach similar to that of [6, Section 2] to explore the branching process between two time points from a single progenitor. Note that if we have multiple individuals alive at time t then the branching processes evolving from each individual are independent. In Section 4 we present a more technical

description of the exploration process, along with the methodology for calculating the key quantities required for Theorem 1.

Consider a single progenitor of type j , say, at time t . For $\tau > 0$, we start with the *forward* exploration process from time t to time $t + \tau$. Specifically, we are interested in the probability of the joint event that there is at least one individual alive at time $t + \tau$ and there are no detected deaths in the interval $(t, t + \tau]$, and also in the probability of the joint event that there are no individuals alive at time $t + \tau$ and no detected deaths in the interval $(t, t + \tau]$. The forward exploration process is as follows. If the progenitor is still alive at time $t + \tau$ we identify the type of the progenitor at time $t + \tau$. If the progenitor dies, undetected, at time $t + u$ say, we then explore back in time through the progenitor's life from time $t + u$ to reveal offspring. For each offspring revealed, we check whether they or one of their descendants is alive at time $t + \tau$. Otherwise we continue to explore the progenitor's life history to reveal any other offspring born earlier in time and study their descendants. The exploration ends in one of three ways; (i) an individual alive at time $t + \tau$, whose type is identified, (ii) the exploration process being exhausted with no individuals alive at time $t + \tau$, or (iii) a death being detected. For $l = 1, 2, \dots, J$, let $p_{jl}(t, 0, \tau)$ be the probability that the forward exploration process results in an individual of type l alive at time $t + \tau$ with no detected deaths along the way, given that we start from a progenitor of type j . Similarly, let $q_j(t, 0, \tau)$ be the probability of the branching process going extinct, with no detected deaths, by time $t + \tau$ given a progenitor of type j . For $0 < u < \tau$, we give definitions of $p_{jl}(t, u, \tau)$ and $q_j(t, u, \tau)$ in (4.1) and (4.2), respectively, in Section 4.2.

In Figure 1A, we present a *successful* realisation of the forward exploration process which results in an individual alive at time $t + \tau$ with no detected deaths. In the event of the forward exploration process being successful, there is an unexplored path from time t to $t + \tau$, consisting of a single individual alive at each timepoint, along which births can occur. The set of individuals comprising the path may consist of only the progenitor or of the progenitor and some of their descendants. The path presented in Figure 1A consists of the progenitor's lifetime from t to $t + v$ and the lifetime of a child of the progenitor from $t + v$ to $t + \tau$. Given the existence of an unexplored path from time t to $t + \tau$, we initiate the *backward* exploration process, which consists of successive i.i.d. cycles until we encounter a failure. The backward exploration process takes the unexplored path from time t to $t + \tau$ and starting at time $t + \tau$ explores back in time to reveal the births of individuals. Each time we reveal an individual we check whether or not they, or one of their descendants, is alive at time $t + \tau$, and whether there are any detected deaths by time $t + \tau$ of the revealed individual or their offspring. If we reveal an individual alive at time $t + \tau$, without encountering any detected deaths, then the current cycle of the exploration process is over and has been successful. An example of the outcome of a successful cycle in the backward exploration process is given in Figure 1B. A cycle fails if either it reveals a death in the interval $(t, t + \tau]$ or the exploration process hits time t without revealing any additional individuals alive at time $t + \tau$. For $l = 1, 2, \dots, J$, let $\psi_l(t; \tau)$ be the probability that a cycle is successful and results in an individual of type l alive at time $t + \tau$. Similarly, let $\zeta(t; \tau)$ be the probability that a cycle results in failure, with the exploration process hitting time t without revealing any additional individuals alive at time $t + \tau$. A formal construction of $\psi_l(t; \tau)$ and $\zeta(t; \tau)$ is given in Section 4.2; see (4.5) and (4.6), respectively. An example of a realisation of the full exploration process with no detected deaths is given in Figure 1C.

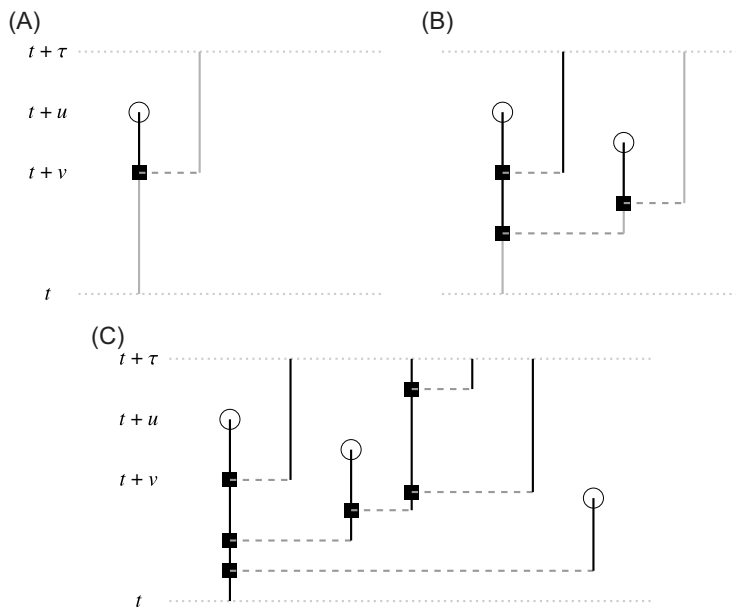


FIGURE 1. Exploration process from time t to $t + \tau$ in three stages, A–C, with one progenitor. Horizontal dotted lines denote start (t) and end ($t + \tau$) points. Vertical lines denote lifetimes of individuals, with black squares denoting explored lifetimes (births revealed) and grey lines denoting unexplored lifetimes. Black squares denote births, with horizontal dashed lines identifying parent–child. Open circles denote undetected deaths. Note that individuals are revealed in the exploration process from left to right.

3.3. Key notation

The description of the exploration process points towards the role of zero-modified geometric random variables in the branching process. Conditional on no detected deaths, the forward exploration process reveals whether there is no individual or at least one individual alive at time $t + \tau$. If there is at least one individual alive we consider the backward process which consists of i.i.d. cycles with binary outcomes (success means an additional individual added to those alive in the branching process at time $t + \tau$; failure means there are no more individuals to add to the branching process at time $t + \tau$). Therefore we start our summary of key notation with a simple yet useful lemma concerning zero-modified geometric random variables.

Lemma 1. For $0 < q < 1$ and $A \geq -(1 - q)$, a random variable X with probability generating function

$$\mathbb{E} [s^X] = \frac{1 + As}{1 + A} \times \frac{q}{1 - (1 - q)s} \left(0 \leq s < \frac{1}{1 - q} \right)$$

is a zero-modified geometric random variable with probability mass function given by $\mathbb{P}(X = 0) = q/(1 + A)$ and $\mathbb{P}(X = k) = [1 - q/(1 + A)](1 - q)^{k-1}q$ ($k = 1, 2, \dots$).

Proof. The result follows trivially from properties of probability generating functions. \square

In Lemma 1, for $A = 0$, $A < 0$, and $A > 0$, X corresponds to a $\text{Geom}(q)$ random variable, a mixture of a $\text{Geom}(q)$ random variable and a point mass at 0, and a sum of a $\text{Geom}(q)$ and an

independent Bernoulli random variable, $\text{Bin}(1, A/(1 + A))$, respectively, where $G \sim \text{Geom}(q)$ has probability mass function $\mathbb{P}(G = k) = (1 - q)^k q$ ($k = 0, 1, \dots$).

For $k = 2, 3, \dots$, let T_k denote the inter-arrival time from the $(k - 1)$ th to the k th detected death, with the convention that $T_k = \infty$ if fewer than k detected deaths occur. Also, let $S_k = \sum_{j=2}^k T_j$ denote the time of the k th detected death, with $S_1 = 0$. Let $\mathbf{T}_{2:k} = (T_2, T_3, \dots, T_k)$. Let t_k and $s_k = \sum_{j=2}^k t_j$ ($k = 2, 3, \dots$) denote the observed inter-arrival time from the $(k - 1)$ th to the k th detected death and the time of the k th detected death, respectively. For $k = 1, 2, \dots$ and $j = 1, 2, \dots, J$, let X_k^j denote the number of individuals alive of type j immediately after the k th detected death, with $X_k^* = \sum_{j=1}^J X_k^j$ denoting the total number of individuals alive. Let $\mathbf{X}_k = (X_k^1, X_k^2, \dots, X_k^J)$. We observe that, for $k = 2, 3, \dots$,

$$\{\mathbf{X}_k | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}\} = \{\mathbf{Y}(s_k) | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}\},$$

with $\mathbf{X}_1 = \mathbf{Y}(0)$.

As noted in Section 2, we assume that prior to the first detected death the branching process (and its observation process) is time-homogeneous with birth rate α_1 and death detection probabilities ϵ . For $t \geq 0$, let $\bar{\psi}_j(t) = \psi_j(-t; t)$, where $\psi_j(\cdot; \cdot)$, defined in Section 3.2, is the probability that the backward exploration process is successful and reveals an individual of type j . Let $\bar{\psi}(t) = \sum_{j=1}^J \bar{\psi}_j(t)$ and set $\pi_0 = 1 - \bar{\psi}(\infty)$. For $j = 1, 2, \dots, J$, let $\eta_j^0 = \bar{\psi}_j(\infty) / \bar{\psi}(\infty)$, the probability that an individual alive at the first detected death is of type j . The distribution of \mathbf{X}_1 is defined in terms of π_0 and $\boldsymbol{\eta}^0 = (\eta_1^0, \eta_2^0, \dots, \eta_J^0)$, and in Lemma 6 we show that

$$X_1^* \sim \text{Geom}(\pi_0),$$

$$\mathbf{X}_1 | X_1^* = x \sim \text{Multinomial}(x, \boldsymbol{\eta}^0).$$

For $n \in \mathbb{N}$ and $\mathbf{q} \in [0, 1]^J$ with $\sum_{i=1}^J q_i = 1$, $\mathbf{Y} \sim \text{Multinomial}(n, \mathbf{q})$ denotes a multinomial distribution with J categories and n trials, having probability mass function

$$\mathbb{P}(\mathbf{Y} = \mathbf{y}) = \frac{n!}{\prod_{j=1}^J y_j!} \prod_{j=1}^J q_j^{y_j} \left(y_j \geq 0 \ (j = 1, 2, \dots, J); \sum_{j=1}^J y_j = n \right).$$

The distribution of the number of individuals alive of each type at time $t \geq 0$ is determined by π_t and $\boldsymbol{\eta}^t$, which are updated from the initial conditions π_0 and $\boldsymbol{\eta}^0$ as follows. Let $\psi(t; \tau) = \sum_{j=1}^J \psi_j(t; \tau)$, and recall that $\zeta(t; \tau)$ is the probability that a cycle of the backward exploration process reveals no further individuals in the branching process at time $t + \tau$. For the forward exploration process (see Section 3.2), from a type- j progenitor, the probability of having at least one individual alive at time $t + \tau$ and the individual being of type l is $p_{jl}(t, \tau)$, and the probability of having no individuals alive at time $t + \tau$ with no detected deaths is $q_j(t, \tau)$. For $t, \tau \geq 0$, we set

$$\pi_{t+\tau} = 1 - \psi(t; \tau) - \frac{(1 - \pi_t)\zeta(t; \tau) \sum_{i=1}^J \sum_{j=1}^J \eta_i^t p_{ij}(t, \tau)}{1 - (1 - \pi_t) \sum_{i=1}^J \eta_i^t q_i(t, \tau)} \tag{3.1}$$

(cf. [2, (3.8) and (3.25)]), and for $j = 1, 2, \dots, J$,

$$\eta_j^{t+\tau} = \frac{\psi_j(t; \tau)}{1 - \pi_{t+\tau}} + \frac{(1 - \pi_t)\zeta(t; \tau) \sum_{i=1}^J \eta_i^t p_{ij}(t, \tau)}{[1 - \pi_{t+\tau}] \left[1 - (1 - \pi_t) \sum_{i=1}^J \eta_i^t q_i(t, \tau) \right]} \tag{3.2}$$

$$= \frac{\psi_j(t; \tau) + \phi_j(t; \tau)}{1 - \pi_{t+\tau}}, \quad \text{say.} \tag{3.3}$$

The distribution of $\mathbf{X}_k | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}$ is given in Theorem 1 in terms of random variables $\mathbf{W}(t, \mathbf{a})$ which are now defined. For $\mathbf{a}, \mathbf{b} \in \mathbb{R}^J$, $\mathbf{a} > \mathbf{b}$ ($\mathbf{a} \geq \mathbf{b}$) if and only if $a_j > b_j$ ($a_j \geq b_j$) ($j = 1, 2, \dots, J$). For $t \geq 0$ and $\mathbf{a} > -(1 - \pi_t)\boldsymbol{\eta}^t$, let $\mathbf{W}(t, \mathbf{a}) = (W_1(t, \mathbf{a}), W_2(t, \mathbf{a}), \dots, W_J(t, \mathbf{a}))$ denote a J -dimensional random variable with, for $\boldsymbol{\theta} \in [0, 1]^J$, probability generating function

$$\begin{aligned} \varphi(\boldsymbol{\theta}; t, \mathbf{a}) &= \mathbb{E} \left[\prod_{j=1}^J \theta_j^{W_j(t, \mathbf{a})} \right] \\ &= \frac{1 + \sum_{j=1}^J a_j \theta_j}{1 + \sum_{j=1}^J a_j} \times \frac{\pi_t}{1 - (1 - \pi_t) \sum_{j=1}^J \eta_j^t \theta_j}. \end{aligned} \tag{3.4}$$

Let $W_*(t, \mathbf{a}) = \sum_{j=1}^J W_j(t, \mathbf{a})$, the sum of the components of $\mathbf{W}(t, \mathbf{a})$. Then it follows immediately from (3.4) that, for $0 \leq \vartheta \leq 1$,

$$\mathbb{E} \left[\vartheta^{W_*(t, \mathbf{a})} \right] = \frac{1 + \vartheta \sum_{j=1}^J a_j}{1 + \sum_{j=1}^J a_j} \times \frac{\pi_t}{1 - (1 - \pi_t)\vartheta},$$

and since $\sum_{j=1}^J a_j > \sum_{j=1}^J \{-(1 - \pi_t)\eta_j^t\} = -(1 - \pi_t)$, by Lemma 1 $W_*(t, \mathbf{a})$ is a zero-modified geometric random variable. Throughout the following we assume that, for $t \geq 0$, $\mathbf{a} > -(1 - \pi_t)\boldsymbol{\eta}^t$.

In the special case $\mathbf{a} = \mathbf{0}$, it follows immediately that

$$W_*(t, \mathbf{0}) \sim \text{Geom}(\pi_t), \tag{3.5}$$

$$\mathbf{W}(t, \mathbf{0}) | W_*(t, \mathbf{0}) = n \sim \text{Multinomial}(n, \boldsymbol{\eta}^t). \tag{3.6}$$

Hence, $\mathbf{X}_1 \stackrel{D}{=} \mathbf{W}(0, \mathbf{0})$.

In determining the distribution of $\mathbf{X}_k | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}$ from the $\{\mathbf{W}(t, \mathbf{a})\}$ random variables, we need to update the values of \mathbf{a} in an iterative manner. The key relationships are defined in Lemma 2.

Lemma 2. For $t, \tau \geq 0$ and $\mathbf{a} \in \mathbb{R}^J$, let $\mathbf{b}(t, \tau; \mathbf{a}), \mathbf{c}(t, \tau; \mathbf{a}) \in \mathbb{R}^J$ satisfy, for $i = 1, 2, \dots, J$,

$$b_i(t, \tau; \mathbf{a}) = \frac{-\psi_i(t; \tau) + \sum_{h=1}^J a_h [\zeta(t; \tau) p_{hi}(t, 0, \tau) - \psi_i(t; \tau) q_h(t, 0, \tau)]}{1 + \sum_{h=1}^J a_h q_h(t, 0, \tau)} \tag{3.7}$$

and

$$c_i(t, \tau; \mathbf{a}) = \frac{(1 - \pi_{t+\tau}) \sum_{j=1}^J d_{t+\tau, j} \gamma_j \left[b_i(t, \tau; \mathbf{a}) \eta_j^{t+\tau} - b_j(t, \tau; \mathbf{a}) \eta_i^{t+\tau} \right]}{\sum_{j=1}^J d_{t+\tau, j} \gamma_j \left[(1 - \pi_{t+\tau}) \eta_j^{t+\tau} + b_j(t, \tau; \mathbf{a}) \right]}. \tag{3.8}$$

Then provided $\mathbf{a} > -(1 - \pi_t)\boldsymbol{\eta}^t$, we have

$$\mathbf{b}(t, \tau; \mathbf{a}), \mathbf{c}(t, \tau; \mathbf{a}) > -(1 - \pi_{t+\tau})\boldsymbol{\eta}^{t+\tau}. \tag{3.9}$$

Proof. The proof of (3.9) is given in Appendix A. □

The coefficients $b_i(t, \tau; \mathbf{a})$ and $c_i(t, \tau; \mathbf{a})$ defined in (3.7) and (3.8), respectively, arise from the two cases noted in Section 3.2: (i) there are no detected deaths in the interval $(t, t + \tau]$, and (ii) the first detected death after time t is at time $t + \tau$.

For $t \in \mathbb{R}$, $\tau \geq 0$, and $\mathbf{a} \in \mathbb{R}^J$, let $\tilde{E}_{t,\tau}^{\mathbf{a}}$ be the event that there are no detected deaths in the interval $(t, t + \tau]$ given that the number of individuals of each type in the branching process at time t is distributed according to $\mathbf{W}(t, \mathbf{a})$. Similarly, let $\tilde{D}_{t,\tau}^{\mathbf{a}}$ be the event that the first detected death after time t is at time $t + \tau$, given that the number of individuals of each type in the branching process at time t is distributed according to $\mathbf{W}(t, \mathbf{a})$. Let

$$U_E(t, \tau; \mathbf{a}) = \mathbb{P}(\tilde{E}_{t,\tau}^{\mathbf{a}}), \tag{3.10}$$

$$U_D(t, \tau; \mathbf{a}) = -\frac{\partial}{\partial \tau} U_E(t, \tau; \mathbf{a}). \tag{3.11}$$

Hence, $U_D(t, \tau; \mathbf{a})$ is the probability density function for the time until the next detected death after time t . We obtain expressions for $U_E(t, \tau; \mathbf{a})$ and $U_D(t, \tau; \mathbf{a})$ in (5.17) and (5.27), respectively.

3.4. Statement and discussion of Theorem 1

We are now in position to state Theorem 1.

Theorem 1 (a) For $k = 1$, $\mathbf{X}_1 \stackrel{D}{=} \mathbf{W}(0, \mathbf{0})$.

(b) For $k = 1, 2, \dots$ and $m = 1, 2, \dots, M_k (= (k - 1)!)$, suppose that

$$\{\mathbf{X}_k | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}\} \stackrel{D}{=} \sum_{l=1}^k \mathbf{W}_{ml}(s_k, \mathbf{a}_{ml}^k) \tag{3.12}$$

with probability w_m^k , where $\{\mathbf{W}_{ml}(s_k, \mathbf{a}_{ml}^k)\}$ are mutually independent and $\mathbf{a}_{ml}^k > -(1 - \pi_{s_k})\boldsymbol{\eta}^{s_k}$. Then

$$\{\mathbf{X}_{k+1} | \mathbf{T}_{2:k+1} = \mathbf{t}_{2:k+1}\} \stackrel{D}{=} \sum_{l=1}^{k+1} \mathbf{W}_{vl}(s_{k+1}, \mathbf{a}_{vl}^{k+1}) \tag{3.13}$$

with probability w_v^{k+1} ($v = 1, 2, \dots, M_{k+1} (= k!)$), where, for $m = 1, 2, \dots, M_k$ and $i = 1, 2, \dots, k$,

$$w_{(m-1)k+i}^{k+1} = \frac{w_m^k U_D(s_k, t_{k+1}; \mathbf{a}_{mi}^k) \prod_{l \neq i} U_E(s_k, t_{k+1}; \mathbf{a}_{ml}^k)}{\sum_{u=1}^{M_k} \sum_{j=1}^k w_u^k U_D(s_k, t_{k+1}; \mathbf{a}_{uj}^k) \prod_{l \neq j} U_E(s_k, t_{k+1}; \mathbf{a}_{ul}^k)}, \tag{3.14}$$

$$\mathbf{a}_{(mk+i)i}^{k+1} = \mathbf{c}_{mi}^k, \quad \mathbf{a}_{(mk+i)j}^{k+1} = \mathbf{b}_{mj}^k \quad (j \neq i), \quad \mathbf{a}_{(mk+i)(k+1)}^{k+1} = \mathbf{0}. \tag{3.15}$$

Here we employ the shorthand notation $\mathbf{b}_{ml}^k (= \mathbf{b}(s_k, t_{k+1}; \mathbf{a}_{ml}^k))$ and $\mathbf{c}_{ml}^k (= \mathbf{c}(s_k, t_{k+1}; \mathbf{a}_{ml}^k))$. Furthermore, $\{\mathbf{W}_{vl}(s_{k+1}, \mathbf{a}_{vl}^{k+1})\}$ are mutually independent and $\mathbf{a}_{vl}^{k+1} > -(1 - \pi_{s_{k+1}})\boldsymbol{\eta}^{s_{k+1}}$.

For $0 < \tau < t_{k+1}$, it is straightforward to show that if $\mathbf{X}_k | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}$ satisfies (3.12), then, for $m = 1, 2, \dots, M_k$,

$$\{\mathbf{Y}(s_k + \tau) | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}\} \stackrel{D}{=} \sum_{l=1}^k \mathbf{W}_{ml}(s_k + \tau, \mathbf{b}(s_k, \tau; \mathbf{a}_{ml}^k))$$

with probability

$$\tilde{w}_m^k = \frac{w_m^k \prod_{l=1}^k U_E(s_k, \tau; \mathbf{a}_{ml}^k)}{\sum_{u=1}^{M_k} w_u^k \prod_{l=1}^k U_E(s_k, \tau; \mathbf{a}_{ul}^k)}.$$

An immediate consequence of Theorem 1 is that $X_k^* | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}$ is the sum of k independent zero-modified geometric random variables. If the components of the $\{\mathbf{a}_{ml}^k\}$ are all less than or equal to 0, we can instead express $X_k^* | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}$ as a random sum of R_k i.i.d. $\text{Geom}(\pi_{s_k})$ random variables, where R_k has support on $\{1, 2, \dots, k\}$.

Corollary 1. For $k = 1, 2, \dots$, suppose that the distribution of $X_k | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}$ satisfies the mixture distribution given in Theorem 1(b) with, for all $l = 1, 2, \dots, k$,

$$-(1 - \pi_{s_k}) \boldsymbol{\eta}^{s_k} < \mathbf{a}_{ml}^k \leq \mathbf{0}. \tag{3.16}$$

For $m = 1, 2, \dots, M_k$ and $l = 1, 2, \dots, k$, let

$$A_{ml}^k = \sum_{i=1}^J a_{mli}^k \quad \text{and} \quad Q_{ml}^k = \frac{1 - \pi_{s_k} + A_{ml}^k}{(1 - \pi_{s_k})(1 + A_{ml}^k)},$$

so $-(1 - \pi_{s_k}) < A_{ml}^k \leq 0$ and $0 \leq Q_{ml}^k \leq 1$.

Let $G_1(s_k), G_2(s_k), \dots, G_k(s_k)$ be i.i.d. according to $\text{Geom}(\pi_{s_k})$. Then

$$X_k^* | \mathbf{T}_{2:k} = \mathbf{t}_{2:k} \sim \sum_{i=1}^{R_k^*} G_i(s_k) = \text{NegBin}(R_k^*, \pi_{s_k}),$$

where R_k^* has support $\{1, 2, \dots, k\}$, and for $i = 1, 2, \dots, k$, $\mathcal{V}_{k,i}$ is the set of vectors of length k consisting of i ones and $k - i$ zeros, with

$$\mathbb{P}(R_k^* = i) = \sum_{m=1}^{M_k} \omega_m^k \sum_{\mathbf{v} \in \mathcal{V}_{k,i}} \prod_{l=1}^k (Q_{ml}^k)^{v_l} [1 - Q_{ml}^k]^{1-v_l}. \tag{3.17}$$

Proof. From Lemma 1 it is straightforward to show that, for \mathbf{a}_{ml}^k satisfying (3.16),

$$W^*(s_k; \mathbf{a}_{ml}^k) \sim \begin{cases} \text{Geom}(\pi_{s_k}) & \text{with probability } Q_{ml}^k, \\ 0 & \text{with probability } 1 - Q_{ml}^k. \end{cases} \tag{3.18}$$

From (3.15), for $m = 1, 2, \dots, k$, $\mathbf{a}_{mk}^k = \mathbf{0}$, so $Q_{mk}^k = 1$ and $\mathbb{P}(R_k^* = 0) = 0$. The corollary then follows from (3.18). \square

Corollary 1 is similar to [2, Theorems 3.1 and 3.2] in deriving the total size of the branching process immediately after the k th death as a mixture of negative binomial distributions. The main difference between Corollary 1 and the theorems in [2] is that the probability mass function of R_k^* given by (3.17) is computationally intensive to compute for moderate k , whereas, in [2], R_k^* is closely related to a binomial distribution, which makes computing its probability mass function much quicker. These observations motivate the following derivation of an approximation of $\mathbf{X}_k | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}$, which provides an approximation for (3.17) and a multinomial approximation for the number of individuals of each type.

3.5. Approximation to the distribution of $\mathbf{X}_k | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}$

This section is structured as follows. In Section 3.5.1 we present the additional notation necessary for the approximation. We state the approximation in Section 3.5.2, with the justification for the approximation given in Section 3.5.3.

3.5.1. *Notation.* In order to construct the approximation we require the following additional notation. For $t, \tau \geq 0$, let $\mathbf{C}_{t;\tau} = (C_{t;\tau}^1, C_{t;\tau}^2, \dots, C_{t;\tau}^J) = \mathbf{c}(t, \tau; \mathbf{0})$. Setting $\mathbf{a} = \mathbf{0}$ in Lemma 2, from (3.7) we obtain $b_i(t, \tau; \mathbf{0}) = -\psi_i(t; \tau)$ ($i = 1, 2, \dots, J$). It then follows from (3.8) and (3.3) that

$$C_{t;\tau}^i = \frac{(1 - \pi_{t+\tau}) \sum_{j=1}^J d_{t+\tau,j} \gamma_j \left[\psi_j(t; \tau) \eta_i^{t+\tau} - \psi_i(t; \tau) \eta_j^{t+\tau} \right]}{\sum_{j=1}^J d_{t+\tau,j} \gamma_j \phi_j(t; \tau)}. \tag{3.19}$$

Hence,

$$C_{t;\tau}^* = \sum_{i=1}^J C_{t;\tau}^i = \frac{(1 - \pi_{t+\tau}) \sum_{j=1}^J d_{t+\tau,j} \gamma_j \left[\psi_j(t; \tau) - \psi(t; \tau) \eta_j^{t+\tau} \right]}{\sum_{j=1}^J d_{t+\tau,j} \gamma_j \phi_j(t; \tau)}.$$

Let

$$\tilde{h}(t; \tau) = 1 + \frac{\pi_{t+\tau} C_{t;\tau}^*}{(1 - \pi_{t+\tau})(1 + C_{t;\tau}^*)} = \frac{1 - \pi_{t+\tau} + C_{t;\tau}^*}{(1 - \pi_{t+\tau})(1 + C_{t;\tau}^*)}. \tag{3.20}$$

Then, provided that $C_{t;\tau}^*$ is non-positive, we have from (3.9) that $-(1 - \pi_{t+\tau}) < C_{t;\tau}^* \leq 0$, and hence $0 \leq \tilde{h}(t; \tau) \leq 1$ is a valid probability.

It is difficult to make general statements concerning the form of $C_{t;\tau}^*$, although in the fully detected, time-homogeneous case, where $\beta_t = \alpha$ and $\mathbf{d}_t = \mathbf{1}$ ($t \in \mathbb{R}$), we can show that $C_{t;\tau}^* \leq 0$ when the lifetime L follows an Erlang distribution (i.e. Gamma($J, J\mu$) with $J \geq 2$) and $C_{t;\tau}^* \geq 0$ when L is a mixture of $J (\geq 2)$ exponential distributions. Moreover, in the fully detected, time-homogeneous case, it can be shown that $\{X_2^* | T_2 = \tau\}$ is stochastically decreasing in τ for an Erlang distribution and stochastically increasing in τ for a mixture of exponential distributions. The proofs are straightforward but somewhat tedious, and are therefore given in the supplementary material.

For $k = 1, 2, \dots$ and $t, \tau \geq 0$, let $\mathbf{M}_k(t; \tau)$ be the $k \times k$ matrix with (i, j) th element

$$[\mathbf{M}_k(s_k; \tau)]_{i,j} = \begin{cases} i \binom{i-1}{j-1} \left\{ \frac{\pi_{s_k} \phi(s_k; t_{k+1})}{\pi_{s_{k+1}} (1 - \pi_{s_{k+1}}) L(s_k, \tau)} \right\}^{j-1} \left\{ \frac{\pi_{s_k} \psi(s_k; t_{k+1})}{(1 - \pi_{s_{k+1}}) L(s_k, \tau)} \right\}^{i-j} & \text{for } j \leq i, \\ 0 & \text{otherwise,} \end{cases} \tag{3.21}$$

where $\phi(s_k; t_{k+1}) = \sum_{j=1}^J \phi_j(s_k; t_{k+1})$ and

$$L(s_k, \tau) = 1 - (1 - \pi_{s_k}) \sum_{i=1}^J \eta_i^{s_k} q_i(s_k, 0, \tau). \tag{3.22}$$

We define $\mathbf{M}_k^1(t; \tau)$ and $\mathbf{M}_k^2(t; \tau)$ to be $k \times (k + 1)$ matrices constructed from $\mathbf{M}_k(t; \tau)$ by the addition of a column of zeros before (after) $\mathbf{M}_k(t; \tau)$ in $\mathbf{M}_k^1(t; \tau)$ ($\mathbf{M}_k^2(t; \tau)$). The matrix

$\mathbf{M}_k(s_k; \tau)$ is of a similar form to the matrices defined in [2]. (3.9) and the equation between (3.26) and (3.27) in Theorem 3.2, and is used to derive an approximation for the probability mass function of R_k^* .

For $t \geq 0$ and $\mathbf{q} \in [0, 1]^J$ such that $\sum_{i=1}^J q_i = 1$, let $\hat{\mathbf{W}}(t; \mathbf{q})$ be the J -dimensional random variable with

$$\hat{W}_*(t; \mathbf{q}) \left(= \sum_{j=1}^J \hat{W}_j(t; \mathbf{q}) \right) \sim \text{Geom}(\pi_t), \tag{3.23}$$

$$\hat{\mathbf{W}}(t; \mathbf{q}) | \hat{W}_*(t; \mathbf{q}) = n \sim \text{Multinomial}(n, \mathbf{q}). \tag{3.24}$$

The relationship between the random variables $\hat{\mathbf{W}}(t; \mathbf{q})$ and $\mathbf{W}(t; \mathbf{a})$ is that, for $t \geq 0$,

$$\hat{\mathbf{W}}(t; \boldsymbol{\eta}^t) \stackrel{D}{=} \mathbf{W}(t; \mathbf{0}). \tag{3.25}$$

Therefore it follows from Theorem 1(a) that $\mathbf{X}_1 \stackrel{D}{=} \hat{\mathbf{W}}(0; \boldsymbol{\eta}^0)$, which is the starting point for the approximation presented in Section 3.5.2.

3.5.2. *Statement of approximation.* Suppose that for all $t, \tau \geq 0$, $C_{i;\tau}^i \leq 0$ ($i = 1, 2, \dots, J$).

Set $\hat{\mathbf{X}}_1 = \mathbf{X}_1 \stackrel{D}{=} \mathbf{W}(0; \mathbf{0}) \equiv \hat{\mathbf{W}}(0; \boldsymbol{\eta}^0)$.

For $k \geq 2$, we approximate $\mathbf{X}_k | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}$ by $\hat{\mathbf{X}}_k | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}$ satisfying

$$\left\{ \hat{\mathbf{X}}_k | \mathbf{T}_{2:k} = \mathbf{t}_{2:k} \right\} \stackrel{D}{=} \sum_{i=1}^{\hat{R}_k} \hat{\mathbf{W}}_i(s_k; \hat{\boldsymbol{\xi}}^k), \tag{3.26}$$

where \hat{R}_k has support $\{1, 2, \dots, k\}$ with $\hat{R}_1 \equiv 1$ and probability mass function given by (3.27), $\hat{\boldsymbol{\xi}}^k$ is given by (3.29) with $\sum_{j=1}^J \hat{\xi}_j^k = 1$, and $\hat{\mathbf{W}}_1(s_k; \hat{\boldsymbol{\xi}}^k), \hat{\mathbf{W}}_2(s_k; \hat{\boldsymbol{\xi}}^k), \dots$ are i.i.d. according to $\hat{\mathbf{W}}(t; \hat{\boldsymbol{\xi}}^k)$ defined by (3.23) and (3.24).

Let $\hat{P}_k^m = \mathbb{P}(\hat{R}_k = m)$ ($k = 1, 2, \dots; m = 1, 2, \dots, k$) and $\hat{\mathbf{P}}_k = (\hat{P}_k^1, \hat{P}_k^2, \dots, \hat{P}_k^k)$. Then $\hat{\mathbf{P}} = (1)$ and $\hat{\mathbf{P}}_{k+1}$ can be obtained recursively using

$$\hat{\mathbf{P}}_{k+1} = \frac{1}{K_k} \hat{\mathbf{P}}_k \left[\tilde{h}(s_k; t_{k+1}) \mathbf{M}_k^1(s_k; t_{k+1}) + \{1 - \tilde{h}(s_k; t_{k+1})\} \mathbf{M}_k^2(s_k; t_{k+1}) \right], \tag{3.27}$$

where $\mathbf{M}_k^1(s_k; t_{k+1})$ and $\mathbf{M}_k^2(s_k; t_{k+1})$ are defined after (3.21) and

$$\begin{aligned} K_k &= \sum_{m=1}^k m \left[\frac{\hat{\pi}_k(1 - \psi(s_k; \tau))}{\hat{\pi}_{k+1}} \right]^{m-1} \hat{P}_k^m \\ &= \hat{\mathbf{P}}_k \left[\tilde{h}(s_k; t_{k+1}) \mathbf{M}_k^1(s_k; t_{k+1}) + \{1 - \tilde{h}(s_k; t_{k+1})\} \mathbf{M}_k^2(s_k; t_{k+1}) \right] \cdot \mathbf{1}_k^\top. \end{aligned} \tag{3.28}$$

For $k = 1, 2, \dots$, let $v_k = \mathbb{E}[\hat{R}_k | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}]$. Set $\boldsymbol{\eta}^{s_k} = \hat{\boldsymbol{\xi}}^k$ and compute $\boldsymbol{\psi}(s_k; t_{k+1}) = (\psi_1(s_k; t_{k+1}), \psi_2(s_k; t_{k+1}), \dots, \psi_J(s_k; t_{k+1}))$ and $\boldsymbol{\eta}^{s_{k+1}}$ using (3.2) with $t = s_k$ and $\tau = t_{k+1}$.

For $k = 1, 2, \dots$ and $j = 1, 2, \dots, J$, set

$$\hat{\zeta}_j^{k+1} = (v_k + 1)\eta_j^{s_{k+1}} \frac{1 - \pi_{s_{k+1}}}{\pi_{s_{k+1}}} - (v_k - 1) \frac{\psi_j(s_k; t_{k+1})}{1 - \psi(s_k; t_{k+1})} + \frac{C_{s_k; t_{k+1}}^j}{1 + C_{s_k; t_{k+1}}^*}$$

$$\times \left[(v_k + 1) \frac{1 - \pi_{s_{k+1}}}{\pi_{s_{k+1}}} - (v_k - 1) \frac{\psi(s_k; t_{k+1})}{1 - \psi(s_k; t_{k+1})} + \frac{C_{s_k; t_{k+1}}^*}{1 + C_{s_k; t_{k+1}}^*} \right]^{-1}.$$
(3.29)

For $j = 1, 2, \dots, J$, the marginal distribution of \hat{X}_k^j , satisfies

$$\hat{X}_k^j | \mathbf{T}_{2:k} = \mathbf{t}_{2:k} \sim \text{NegBin} \left(\hat{R}_k, \frac{\pi_{s_k}}{\pi_{s_k} + \hat{\zeta}_k^j - \pi_{s_k} \hat{\zeta}_k^j} \right).$$
(3.30)

3.5.3. *Justification of approximation.* From Section 3.5.2, we note that $\hat{X}_k^* | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}$ takes an almost identical form to the distribution of the number of individuals alive in a birth–death process immediately after the k th detected death given in [2, Theorem 3.2, (3.28)]. The only difference is that in the phase-type distribution case, \hat{R}_k can take the value 1 for any k . The recursive equation for the probability mass function for \hat{R}_k given in (3.27), and utilising (3.21), is similar to [2, Lemma 4.5, (4.17)] and is derived along similar lines.

The approximation in Section 3.5.2 works as follows. If $\{\mathbf{X}_k | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}\} \stackrel{D}{=} \{\hat{\mathbf{X}}_k | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}\}$, then

$$\{\mathbf{X}_{k+1}^* | \mathbf{T}_{2:k+1} = \mathbf{t}_{2:k+1}\} \stackrel{D}{=} \{\hat{\mathbf{X}}_{k+1}^* | \mathbf{T}_{2:k+1} = \mathbf{t}_{2:k+1}\}$$

with $\hat{\zeta}^{k+1} = (\hat{\zeta}_1^{k+1}, \hat{\zeta}_2^{k+1}, \dots, \hat{\zeta}_{k+1}^{k+1})$ satisfying

$$\mathbb{E}[\hat{X}_{k+1}^j | \mathbf{T}_{2:k+1} = \mathbf{t}_{2:k+1}] = \hat{\zeta}_j^{k+1} \mathbb{E}[X_{k+1}^* | \mathbf{T}_{2:k+1} = \mathbf{t}_{2:k+1}] (= \mathbb{E}[\hat{X}_{k+1}^j | \mathbf{T}_{2:k+1} = \mathbf{t}_{2:k+1}]).$$

Details of the proof of the above are provided in Section 6.

4. Exploration process

4.1. Introduction

In this section we show how the key probabilities from the exploration process for Theorem 1 can be computed. We use an approach similar to that of [6, Section 2] to explore the branching process between two time points from a single progenitor. Note that the branching processes studied in [6] are referred to as splitting trees; see Geiger and Kersting [4]. There are minor modifications to the exploration process and analysis in [6, Section 2] to take account of the time-inhomogeneous birth rate and detection probability, differences in the detection process, and the phase-type lifetime distribution. The phase-type lifetime distribution enables us to obtain more explicit results than those presented in [6, Section 2].

In Section 4.2 we define in (4.1)–(4.4) the probabilities of the exploration process. These probabilities are expressed in terms of systems of ordinary differential equations (ODEs) in Lemma 3, which enable analytical solutions in some important special cases; see Sections 4.3 (all deaths detected) and 4.4 (a piecewise time-homogeneous branching process).

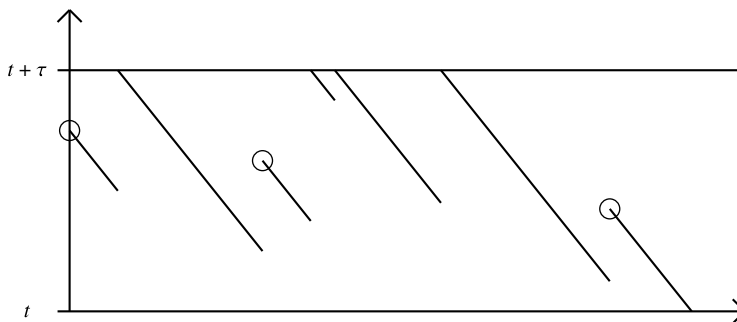


FIGURE 2. The JCCP associated with the branching process in Figure 1C. The open circles denote potential detected deaths which in the given realisation of the branching process are all undetected.

4.2. Exploration process equations

To obtain the key probabilities of the exploration process, it is helpful to use the jump chronological contour process (JCCP) of the branching process, as defined in [6], but with a time shift t in the starting time and time-dependent birth rate, β_{t+u} . Let $\mathcal{X}^{(t,\tau)}$ denote the JCCP, started with one unique progenitor at time t , truncated up to height (time) $t + \tau$ of the birth process. The JCCP is associated with the exploration process between times t and $t + \tau$ as follows. The process $\mathcal{X}^{(t,\tau)}$ starts at $t + (\tilde{L} \wedge \tau)$, where \tilde{L} is the residual lifetime, after time t , of the single progenitor. The JCCP has its own timescale and explores backward in time at rate 1 until it encounters a birth in the branching (exploration) process, at which point the JCCP jumps. If a jump occurs in the process $\mathcal{X}^{(t,\tau)}$ at position $t + v$ ($0 < v < \tau$), then the process jumps to $(t + v + L) \wedge (t + \tau)$, where L is the length of the lifetime of the individual born at time $t + v$. That is, the JCCP jumps to whichever occurs first, the end of the individual's lifetime or time $t + \tau$. The process $\mathcal{X}^{(t,\tau)}$ terminates when it reaches time t , corresponding to the exploration process being over, so $\mathcal{X}^{(t,\tau)}$ remains within the interval $[t, t + \tau]$. The JCCP associated with the realisation of the branching process in Figure 1C is given in Figure 2.

For $x \in \mathbb{R}$, let $\sigma_x = \sigma_{\{x\}}$ denote the first hitting time of x by $\mathcal{X}^{(t,\tau)}$. Let $\sigma = \sigma_t \wedge \sigma_{t+\tau}$, the first time the process hits either endpoint of the interval, corresponding to either the exploration process being over or an individual being identified as alive at time $t + \tau$. Two useful quantities associated with the JCCP at time σ are $D(\sigma)$, the number of detected deaths in $[0, \sigma]$, and $J(\sigma)$, the type of the individual at time σ , when $\sigma = \sigma_{t+\tau}$.

We are now in position to define the probabilities of the exploration process required to obtain $p_{ij}(t, 0, \tau)$, $q_i(t, 0, \tau)$, $\psi_j(t; \tau)$, and $\zeta(t; \tau)$ as described in Section 3.2.

For $t \in \mathbb{R}$, $0 \leq u \leq \tau$, and $i, j = 1, 2, \dots, J$, let

$$p_{ij}(t, u, \tau) = \mathbb{P}_{u,i}(\sigma_{t+\tau} < \sigma_t, J(\sigma) = j, D(\sigma) = 0) \tag{4.1}$$

and

$$q_i(t, u, \tau) = \mathbb{P}_{u,i}(\sigma_t < \sigma_{t+\tau}, D(\sigma) = 0), \tag{4.2}$$

where $\mathbb{P}_{u,i}$ denotes the probability conditional upon $\tilde{L} \geq u$ and the progenitor being of type i at time $t + u$. The equations (4.1) and (4.2) are associated with the forward exploration

process, and the special case $u = 0$ yields the $p_{ij}(t, 0, \tau)$ and $q_i(t, 0, \tau)$ required for Theorem 1.

For $t \in \mathbb{R}$, $0 \leq u \leq \tau$, and $j = 1, 2, \dots, J$, let

$$p_j(t, u, \tau) = \mathbb{P}_u(\sigma_{t+\tau} < \sigma_t, J(\sigma) = j, D(\sigma) = 0) \tag{4.3}$$

and

$$q(t, u, \tau) = \mathbb{P}_u(\sigma_t < \sigma_{t+\tau}, D(\sigma) = 0), \tag{4.4}$$

where \mathbb{P}_u denotes probability conditional upon $\tilde{L} = u$ and the death of the progenitor being unobserved.

In contrast to (4.1), the probabilities in (4.3) are obtained by starting at time $t + u$ and exploring backward in time for births which deliver an undetected pathway to an individual of type j alive at time $t + \tau$. Since an individual’s type at birth does not depend on the type of their parent, we do not need to specify the type of the progenitor in (4.3) and (4.4). Then $\psi_j(t; \tau)$ ($j = 1, 2, \dots, J$) and $\zeta(t; \tau)$ can be defined as special cases of (4.3) and (4.4), respectively, with

$$\psi_j(t; \tau) = \lim_{u \uparrow \tau} p_j(t, u, \tau), \tag{4.5}$$

the probability that a cycle of the backward exploration process is successful and produces an individual of type j at time $t + \tau$, and

$$\zeta(t; \tau) = \lim_{u \uparrow \tau} q(t, u, \tau), \tag{4.6}$$

the probability that a cycle of the backward exploration process reveals no further individuals alive at time $t + \tau$ with no detected deaths.

The main reason for giving the probabilities of the exploration process in the form of (4.1)–(4.4) is that we can express the probabilities via systems of ODEs. These ODEs can be solved exactly in terms of matrix exponentials in some special cases (see, for example, Section 4.4), or otherwise solved numerically in an efficient manner using an ODE solver.

Lemma 3. Fix $t \in \mathbb{R}$ and $\tau > 0$.

For $j = 1, 2, \dots, J$, let $\mathbf{p}_{\cdot j}(t, u, \tau) = (p_{1j}(t, u, \tau), p_{2j}(t, u, \tau), \dots, p_{Jj}(t, u, \tau))$.

Then $\{(\mathbf{p}_{\cdot j}(t, u, \tau), p_j(t, u, \tau)); 0 \leq u \leq \tau\}$ solves the system of $J + 1$ ODEs with

$$p'_{ij}(t, u, \tau) = - \sum_{l=1}^J \lambda_{il} p_{lj}(t, u, \tau) - \gamma_i(1 - d_{t+u,i}) p_j(t, u, \tau), \tag{4.7}$$

$$p'_j(t, u, \tau) = \beta_{t+u} \left\{ \sum_{l=1}^J \chi_l p_{lj}(t, u, \tau) - p_j(t, u, \tau) \right\}, \tag{4.8}$$

and boundary conditions $\mathbf{p}_{\cdot j}(t, \tau, \tau) = \mathbf{e}_j$ and $p_j(t, 0, \tau) = 0$.

Similarly, let $\mathbf{q}(t, u, \tau) = (q_1(t, u, \tau), q_2(t, u, \tau), \dots, q_J(t, u, \tau))$.

Then $\{(\mathbf{q}(t, u, \tau), q(t, u, \tau)); 0 \leq u \leq \tau\}$ solves the system of $J + 1$ ODEs with

$$q'_i(t, u, \tau) = - \sum_{l=1}^J \lambda_{il} q_l(t, u, \tau) - \gamma_i(1 - d_{t+u,i}) q(t, u, \tau), \tag{4.9}$$

$$q'(t, u, \tau) = \beta_{t+u} \left\{ \sum_{l=1}^J \chi_l q_l(t, u, \tau) - q(t, u, \tau) \right\}, \tag{4.10}$$

with boundary conditions $\mathbf{q}(t, \tau, \tau) = \mathbf{0}$ and $q(t, 0, \tau) = 1$.

Proof. Fix $j = 1, 2, \dots, J$. The boundary conditions $\mathbf{p}_j(t, \tau, \tau) = \mathbf{e}_j$ and $p_j(t, 0, \tau) = 0$ follow immediately from the construction of the JCCP.

For $i = 1, 2, \dots, J$, conditioning on the time the progenitor remains in state i , we have

$$p_{ij}(t, u, \tau) = \exp(\lambda_{ii}[\tau - u])\delta_{ij} + \int_u^\tau -\lambda_{ii} \exp(\lambda_{ii}[z - u]) \times \left\{ \sum_{l \neq i} \frac{\lambda_{il}}{-\lambda_{ii}} p_{lj}(t, z, \tau) + \frac{\gamma_i}{-\lambda_{ii}} (1 - d_{t+z,i}) p_j(t, z, \tau) \right\} dz, \tag{4.11}$$

where δ_{ij} denotes the Kronecker delta with $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise. Differentiating (4.11) with respect to u and substituting in $p_{ij}(t, u, \tau)$, we have that

$$\begin{aligned} p'_{ij}(t, u, \tau) &= -\lambda_{ii} \exp(\lambda_{ii}[\tau - u])\delta_{ij} - \lambda_{ii} \int_u^\tau -\lambda_{ii} \exp(\lambda_{ii}[z - u]) \\ &\quad \times \left\{ \sum_{l \neq i} \frac{\lambda_{il}}{-\lambda_{ii}} p_{lj}(t, z, \tau) + \frac{\gamma_i}{-\lambda_{ii}} (1 - d_{t+z,i}) p_j(t, z, \tau) \right\} dz \\ &\quad - \sum_{l \neq i} \lambda_{il} p_{lj}(t, u, \tau) - \gamma_i (1 - d_{t+u,i}) p_j(t, u, \tau) \\ &= - \sum_{l=1}^J \lambda_{il} p_{lj}(t, u, \tau) - \gamma_i (1 - d_{t+u,i}) p_j(t, u, \tau), \end{aligned} \tag{4.12}$$

which agrees with (4.7).

Similarly, by conditioning on the time of the first birth when exploring back from u , we have

$$p_j(t, u, \tau) = \int_0^u \exp\left(-\int_z^u \beta_{t+s} ds\right) \beta_{t+z} \sum_{l=1}^J \chi_l p_{lj}(t, z, \tau) dz. \tag{4.13}$$

Differentiating (4.13) with respect to u yields

$$\begin{aligned} p'_j(t, u, \tau) &= \beta_{t+u} \sum_{l=1}^J \chi_l p_{lj}(t, u, \tau) - \int_0^u \beta_{t+u} \exp\left(-\int_z^u \beta_{t+s} ds\right) \beta_{t+z} \sum_{l=1}^J \chi_l p_{lj}(t, z, \tau) dz \\ &= \beta_{t+u} \left\{ \sum_{l=1}^J \chi_l p_{lj}(t, u, \tau) - p_j(t, u, \tau) \right\}, \end{aligned}$$

which agrees with (4.8).

Turning to $\{(\mathbf{q}(t, u, \tau), q(t, u, \tau)); 0 \leq u \leq \tau\}$, we see that the boundary conditions $\mathbf{q}(t, \tau, \tau) = \mathbf{0}$ and $q(t, 0, \tau) = 1$ follow immediately from the construction of the JCCP.

For $i = 1, 2, \dots, J$, conditioning on the time the progenitor remains in state i , we have

$$q_i(t, u, \tau) = \int_u^\tau -\lambda_{ii} \exp(\lambda_{ii}[z - u]) \times \left\{ \sum_{l \neq i} \frac{\lambda_{il}}{-\lambda_{ii}} q_l(t, z, \tau) + \frac{\gamma_i}{-\lambda_{ii}} (1 - d_{t+z,i}) q(t, z, \tau) \right\} dz.$$

By an argument similar to the derivation of (4.12), it is straightforward to show that $q'_i(t, u, \tau)$ satisfies (4.9). Also, we have that

$$q(t, u, \tau) = \int_0^u \exp\left(-\int_z^u \beta_{t+s} ds\right) \beta_{t+z} \sum_{l=1}^J \chi_l q_l(t, z, \tau) dz,$$

from which it is straightforward to show that $q'(t, u, \tau)$ satisfies (4.10), completing the proof of the lemma. □

4.3. All deaths detected

The first special case we consider is the one in which all deaths are detected, $\mathbf{d}_{t+u} = \mathbf{1}$ ($0 \leq u \leq \tau$). In this case $p_{ij}(t, u, \tau) = P_{ij}(\tau - u)$, $q_i(t, u, \tau) = 0$, and in (4.3) and (4.4) the conditioning event in \mathbb{P}_u has probability 0. However, the exploration process is simpler if all deaths are detected, since any individual born in the process will either be alive at time $t + \tau$ or have been detected at their time of death. We can slightly modify the definitions of $\psi_j(t; \tau)$ and $\zeta(t; \tau)$ given in (4.5) and (4.6), respectively, and provide expressions for these quantities.

Let $\psi_j(t; \tau)$ be the probability that the backward exploration process from time $t + \tau$ to time t reveals a birth with the individual who is born being alive at time $t + \tau$ and of type j . Then, conditioning on the time of the birth and the type of individual at birth, we have by the theorem of total probability that

$$\psi_j(t; \tau) = \int_t^{t+\tau} \beta_u \exp\left(-\int_u^{t+\tau} \beta_s ds\right) \left[\sum_{i=1}^J \chi_i P_{ij}(t + \tau - u) \right] du. \tag{4.14}$$

Similarly, let $\zeta(t; \tau)$ be the probability that the backward exploration process from time $t + \tau$ to time t reveals no births. Then we simply have that

$$\zeta(t; \tau) = \exp\left(-\int_t^{t+\tau} \beta_s ds\right). \tag{4.15}$$

4.4. Time-homogeneous equations

An important special case is when the branching process and its observation process are time-homogeneous on the interval $(t, t + \tau]$ with $\beta_{t+u} = \alpha$ and $\mathbf{d}_{t+u} = \boldsymbol{\epsilon}$ ($0 \leq u \leq \tau$). In this case, explicit solutions to the systems of ODEs given in Lemma 3 exist, and since we assume that the branching process is time-homogeneous prior to the first detected death, we utilise the explicit solution in deriving the size of the branching process at the first detected death in Lemma 6.

For $\alpha, \tau > 0$ and $\mathbf{0} \leq \boldsymbol{\epsilon} \leq \mathbf{1}$, let $\Omega (= \Omega(\alpha, \boldsymbol{\epsilon}))$ and $\Sigma (= \Sigma(\alpha, \boldsymbol{\epsilon}, \tau))$ denote the $(J + 1) \times (J + 1)$ matrices given by

$$\Omega (= \Omega(\alpha, \boldsymbol{\epsilon})) = \begin{pmatrix} \Lambda & \boldsymbol{\rho}^T \\ -\alpha\boldsymbol{\chi} & \alpha \end{pmatrix}, \tag{4.16}$$

and

$$\Sigma (= \Sigma(\alpha, \boldsymbol{\epsilon}, \tau)) = \exp[\tau\Omega], \tag{4.17}$$

where $\boldsymbol{\rho} = ([1 - \epsilon_1]\gamma_1, [1 - \epsilon_2]\gamma_2, \dots, [1 - \epsilon_J]\gamma_J)$ and $\boldsymbol{\chi} = (\chi_1, \chi_2, \dots, \chi_J)$ are row vectors of length J . Note that for $j = 1, 2, \dots, J$, $\rho_j = [1 - \epsilon_j]\gamma_j$ is the rate at which a type- j individual dies undetected. In Lemma 4, we show that $\mathbf{p}_{\cdot j}(t, 0, \tau)$, $q_j(t, 0, \tau)$, $\psi_j(t; \tau)$ ($j = 1, 2, \dots, J$), and $\zeta(t; \tau)$ can be obtained from Σ given in (4.17).

Lemma 4. Fix $t \in \mathbb{R}$ and $\tau > 0$. For a branching process with $\beta_{t+u} = \alpha$ and $\mathbf{d}_{t+u} = \boldsymbol{\epsilon}$ ($0 \leq u \leq \tau$), we have that $\zeta(t; \tau) = \Sigma_{J+1, J+1}^{-1}$, $\psi_j(t; \tau) = -\Sigma_{J+1, j} / \Sigma_{J+1, J+1}$,

$$\mathbf{p}_{\cdot j}(t, 0, \tau) = \Sigma_{1:J, j} + \psi_j(t; \tau)\Sigma_{1:J, J+1}, \tag{4.18}$$

and

$$\mathbf{q}(t, 0, \tau) = \zeta(t; \tau)\Sigma_{1:J, J+1}. \tag{4.19}$$

Proof. For $i, j = 1, 2, \dots, J$ and $0 \leq u \leq \tau$, let $\tilde{p}_{ij}(t, u, \tau) = p_{ij}(t, \tau - u, \tau)$, $\tilde{p}_j(t, u, \tau) = p_j(t, \tau - u, \tau)$, $\tilde{q}_j(t, u, \tau) = q_j(t, \tau - u, \tau)$, and $\tilde{q}(t, u, \tau) = q(t, \tau - u, \tau)$ with $\tilde{\mathbf{p}}_{\cdot j}(t, u, \tau) = \mathbf{p}_{\cdot j}(t, \tau - u, \tau)$ and $\tilde{\mathbf{q}}(t, u, \tau) = \mathbf{q}(t, \tau - u, \tau)$. Thus the tilde probabilities are time-reversed versions of the probabilities defined in (4.1)–(4.4), starting at time $t + \tau$ when $u = 0$ and finishing at time t when $u = \tau$.

Using (4.7) and (4.8), it is straightforward to show that

$$\tilde{p}'_{ij}(t, u, \tau) = -p'_{ij}(t, \tau - u, \tau) = \sum_{l=1}^J \lambda_{il}\tilde{p}_{lj}(t, u, \tau) + [1 - \epsilon_i]\gamma_i\tilde{p}_j(t, u, \tau), \tag{4.20}$$

$$\tilde{p}'_j(t, u, \tau) = -p'_j(t, \tau - u, \tau) = -\alpha \sum_{l=1}^J \chi_l\tilde{p}_{lj}(t, u, \tau) + \alpha\tilde{p}_j(t, u, \tau), \tag{4.21}$$

with boundary conditions $\tilde{\mathbf{p}}_{\cdot j}(t, 0, \tau) = \mathbf{e}_j$ and $\tilde{p}_j(t, \tau, \tau) = 0$. Therefore, for $j = 1, 2, \dots, J$, the system of $J + 1$ ODEs $\{\tilde{\mathbf{p}}_{\cdot j}(t, u, \tau), \tilde{p}_j(t, u, \tau)\}$ has J initial conditions and one final boundary condition. Similarly, using (4.9) and (4.10), we have a system of $J + 1$ ODEs satisfying

$$\tilde{q}'_i(t, u, \tau) = -q'_i(t, \tau - u, \tau) = \sum_{l=1}^J \lambda_{il}\tilde{q}_l(t, u, \tau) + [1 - \epsilon_i]\gamma_i\tilde{q}(t, u, \tau), \tag{4.22}$$

$$\tilde{q}'(t, u, \tau) = -q'(t, \tau - u, \tau) = -\alpha \sum_{l=1}^J \chi_l\tilde{q}_l(t, u, \tau) + \alpha\tilde{q}(t, u, \tau), \tag{4.23}$$

with initial conditions $\tilde{\mathbf{q}}(t, 0, \tau) = \mathbf{0}$ and final boundary condition $\tilde{q}(t, \tau, \tau) = 1$.

From (4.5) and (4.6) we have that $\psi_j(t; \tau) = \lim_{u \downarrow 0} \tilde{p}_j(t, u, \tau)$ and $\zeta(t; \tau) = \lim_{u \downarrow 0} \tilde{q}(t, u, \tau)$. Therefore it follows from (4.20)–(4.23) and (4.17) that

$$\begin{pmatrix} \tilde{\mathbf{p}}_j(t, \tau, \tau) \\ 0 \end{pmatrix} = \Sigma \begin{pmatrix} \mathbf{e}_j \\ \tilde{p}_j(t, 0, \tau) \end{pmatrix} = \Sigma \begin{pmatrix} \mathbf{e}_j \\ \psi_j(t; \tau) \end{pmatrix} \quad (j = 1, 2, \dots, J) \tag{4.24}$$

and

$$\begin{pmatrix} \tilde{\mathbf{q}}(t, \tau, \tau) \\ 1 \end{pmatrix} = \Sigma \begin{pmatrix} \mathbf{0} \\ \tilde{q}(t, 0, \tau) \end{pmatrix} = \Sigma \begin{pmatrix} \mathbf{0} \\ \zeta(t; \tau) \end{pmatrix}. \tag{4.25}$$

The lemma follows immediately from (4.24) and (4.25). □

5. Proof of Theorem 1

The outline of the proof of Theorem 1 is as follows. A key step is to study the distribution of the number of individuals alive at time $t + \tau$, and their types, conditioning on the number of individuals of each type alive in the population at time t . As noted in Section 4, we focus on two scenarios: (i) no detected deaths in the time interval $(t, t + \tau]$, and (ii) precisely one detected death in the time interval $(t, t + \tau]$, at time $t + \tau$. The probability generating functions of the required distributions are given in Lemma 5 and Corollary 2. This enables us to show in Lemma 6 that $\mathbf{X}_1 = \mathbf{W}(0, \mathbf{0})$, provided that the branching process and its observation process are time-homogeneous before the first detected death. This provides the initial conditions, \mathbf{X}_1 , for Theorem 1. Thus $X_1^* \sim \text{Geom}(\pi_0)$, which complements results in [6, 9], where individuals have independent exponential clocks for detection. In Lemmas 7 and 8, we establish the distribution of the number of individuals alive at time $t + \tau$ in the two scenarios, given that the number of individuals alive at time t follows a $\mathbf{W}(t, \mathbf{a})$ distribution defined by its probability generating function (3.4). Then, given that $\mathbf{X}_k | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}$ is the mixture of sums of k independent $\mathbf{W}(s_k, \cdot)$ random variables, it is straightforward to conclude the proof of Theorem 1 using induction and Lemmas 7 and 8.

Suppose that the κ th detected death occurs at time t . The precise value of κ is not important in the following arguments. Let $E_{t,\tau}^y$ be the event that there are no detected deaths in the interval $(t, t + \tau)$ given that $\mathbf{Y}(t) = \mathbf{y}$. Similarly, let $D_{t,\tau}^y$ be the event that the first detected death in the interval $(t, t + \tau]$ is at time $t + \tau$ given that $\mathbf{Y}(t) = \mathbf{y}$. Then for $t, \tau \geq 0$, $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_J) \in [0, 1]^J$, and $\mathbf{y} = (y_1, y_2, \dots, y_J) \in \mathbb{Z}_+^J$, let

$$H_E(\boldsymbol{\theta}; t, \tau; \mathbf{y}) = \mathbb{E} \left[\prod_{j=1}^J \theta_j^{y_j(t+\tau)} 1_{\{E_{t,\tau}^y\}} \mid \mathbf{Y}(t) = \mathbf{y} \right]$$

and

$$H_D(\boldsymbol{\theta}; t, \tau; \mathbf{y}) = \mathbb{E} \left[\prod_{j=1}^J \theta_j^{X_{\kappa+1,j}} \mid \mathbf{X}_\kappa = \mathbf{y}, T_{\kappa+1} = \tau \right] f_{T_{\kappa+1}}(\tau | \mathbf{X}_\kappa = \mathbf{y}).$$

Note that $f_{T_{\kappa+1}}(\tau | \mathbf{X}_\kappa = \mathbf{y})$ satisfies

$$f_{T_{\kappa+1}}(\tau | \mathbf{X}_\kappa = \mathbf{y}) = H_D(\mathbf{1}; t, \tau; \mathbf{y}). \tag{5.1}$$

Lemma 5. For any $\tau > 0$, $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_J) \in [0, 1]^J$, and $\mathbf{y} = (y_1, y_2, \dots, y_J) \in \mathbb{Z}_+^J$,

$$H_E(\boldsymbol{\theta}; t, \tau; \mathbf{y}) = \prod_{i=1}^J \left(q_i(t, 0, \tau) + \frac{\zeta(t, \tau) [\sum_{j=1}^J p_{ij}(t, 0, \tau) \theta_j]}{1 - \sum_{j=1}^J \psi_j(t; \tau) \theta_j} \right)^{y_i}. \tag{5.2}$$

Proof. For $i = 1, 2, \dots, J$, let \mathbf{e}_i denote the vector of length J in which the i th entry is 1 and all other entries are equal to 0. Since the branching process evolves independently from each of the \mathbf{y} individuals alive in the population at time 0, we have that

$$H_E(\boldsymbol{\theta}; t, \tau; \mathbf{y}) = \prod_{i=1}^J H_E(\boldsymbol{\theta}; t, \tau; \mathbf{e}_i)^{y_i}. \tag{5.3}$$

Using the exploration process described in Section 4, the joint probability of the number of individuals alive at time $t + \tau$ from a single progenitor of type i at time t , $\hat{Y}_i^*(t; \tau)$, and the event of no detected deaths in the interval $(t, t + \tau]$, denoted by $\hat{E}_{t,\tau}^i$, are given by

$$\begin{aligned} \mathbb{P}(\hat{Y}_i^*(t; \tau) = 0, \hat{E}_{t,\tau}^i) &= q_i(t, 0, \tau), \\ \mathbb{P}(\hat{Y}_i^*(t, \tau) = n, \hat{E}_{t,\tau}^i) &= \left[\sum_{j=1}^J p_{ij}(t, 0, \tau) \right] \psi(t; \tau)^{n-1} \zeta(t; \tau) \quad (n = 1, 2, \dots). \end{aligned}$$

Note that if all deaths are detected then $q_i(t, 0, \tau) = 0$.

Given that $\hat{Y}_i^*(t, \tau) = n > 0$, there is a single individual, either the progenitor or its descendant derived from the forward pathway, whose probability of being of type l at time $t + \tau$ is $p_{il}(t, 0, \tau) / \sum_{i=1}^J p_{il}(t, 0, \tau)$. By the strong Markov property, each of the other $n - 1$ individuals alive at time $t + \tau$ are independently of type j with probability $\psi_j(t; \tau) / \psi(t; \tau)$. It is then straightforward to show that

$$\begin{aligned} H_E(\boldsymbol{\theta}; t, \tau; \mathbf{e}_i) &= q_i(t, 0, \tau) + \sum_{n=1}^{\infty} \zeta(t; \tau) \left[\sum_{j=1}^J p_{ij}(t, 0, \tau) \theta_j \right] \left[\sum_{j=1}^J \psi_j(t; \tau) \theta_j \right]^{n-1} \\ &= q_i(t, 0, \tau) + \frac{\zeta(t; \tau) \left[\sum_{j=1}^J p_{ij}(t, 0, \tau) \theta_j \right]}{1 - \sum_{j=1}^J \psi_j(t; \tau) \theta_j}. \end{aligned} \tag{5.4}$$

Finally, (5.2) follows from substituting (5.4) into (5.3). □

We now derive the relationship between $H_D(\boldsymbol{\theta}; t, \tau; \mathbf{y})$ and $H_E(\boldsymbol{\theta}; t, \tau; \mathbf{y})$. Recall that, for $i = 1, 2, \dots, J$, $d_{t,i}$ denotes the probability that a death of a type- i individual at time t is detected.

Corollary 2. For any $t, \tau > 0$, $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_J) \in [0, 1]^J$, and $\mathbf{y} = (y_1, y_2, \dots, y_J) \in \mathbb{Z}_+^J$,

$$H_D(\boldsymbol{\theta}; t, \tau; \mathbf{y}) = \sum_{i=1}^J d_{t+\tau,i} \gamma_i \frac{\partial}{\partial \theta_i} H_E(\boldsymbol{\theta}; t, \tau; \mathbf{y}). \tag{5.5}$$

Proof. The proof follows arguments similar to those used for [2, Lemma 4.4]. Let $f_{T_{\kappa+1}}^i(\cdot)$ denote the joint density of $T_{\kappa+1}$ and the $(\kappa + 1)$ th detected death being of type i . Hence, for $t \geq 0$, $f_{T_{\kappa+1}}(t) = \sum_{i=1}^J f_{T_{\kappa+1}}^i(t)$.

We observe that

$$\begin{aligned} f(\mathbf{X}_{\kappa+1} = \mathbf{x}, T_{\kappa+1} = \tau | \mathbf{X}_{\kappa} = \mathbf{y}) \\ = \lim_{\sigma \uparrow \tau} \sum_{i=1}^J f_{T_{\kappa+1}}^i(\tau | \mathbf{Y}(t + \sigma) = \mathbf{x} + \mathbf{e}_i, T_{\kappa+1} > \sigma) \end{aligned}$$

$$\begin{aligned} & \times \mathbb{P}(\mathbf{Y}(t + \sigma) = \mathbf{x} + \mathbf{e}_i, T_{k+1} > \sigma | \mathbf{Y}(t) = \mathbf{y}) \\ & = \sum_{i=1}^J d_{t+\tau, i} \gamma_i (x_i + 1) \mathbb{P}(\mathbf{Y}(t + \tau) = \mathbf{x} + \mathbf{e}_i, T_{k+1} > \tau | \mathbf{Y}(t) = \mathbf{y}). \end{aligned}$$

Therefore, letting $\mathbf{v}^{(i)} = \mathbf{x} + \mathbf{e}_i$, we have that

$$\begin{aligned} H_D(\boldsymbol{\theta}; t, \tau; \mathbf{y}) &= \sum_{i=1}^J d_{t+\tau, i} \gamma_i \sum_{\mathbf{x}} (x_i + 1) \prod_{j=1}^J \theta_j^{x_j} \mathbb{P}(\mathbf{Y}(t + \tau) = \mathbf{x} + \mathbf{e}_i, T_{k+1} > \tau | \mathbf{Y}(t) = X_k = \mathbf{y}) \\ &= \sum_{i=1}^J d_{t+\tau, i} \gamma_i \sum_{\mathbf{v}^{(i)}} v_i^{(i)} \theta_i^{v_i^{(i)} - 1} \prod_{j \neq i}^J \theta_j^{v_j^{(i)}} \mathbb{P}(\mathbf{Y}(t + \tau) = \mathbf{v}^{(i)}, T_{k+1} > \tau | \mathbf{Y}(t) = X_k = \mathbf{y}) \\ &= \sum_{i=1}^J d_{t+\tau, i} \gamma_i \frac{\partial}{\partial \theta_i} H_E(\boldsymbol{\theta}; t, \tau; \mathbf{y}), \end{aligned}$$

as required. □

Suppose that we let \mathbf{X}_0 denote the state of the population when the initial individual enters the population. For $i = 1, 2, \dots, J$, $\mathbb{P}(\mathbf{X}_0 = \mathbf{e}_i) = \chi_i$. We can use Lemma 5 and Corollary 2 to obtain \mathbf{X}_1 by integrating over T_1 , the inter-arrival time from the initial individual entering the population to the first detected death.

Lemma 6. *Given a single initial individual who starts in type l with probability χ_l , the joint probability generating function of \mathbf{X}_1 is given by*

$$\mathbb{E} \left[\prod_{j=1}^J \theta_j^{X_1^j} \middle| T_1 < \infty \right] = \frac{\pi_0}{1 - (1 - \pi_0) \sum_{j=1}^J \eta_j^0 \theta_j}. \tag{5.6}$$

Therefore, $\mathbf{X}_1 \sim \mathbf{W}(0, \mathbf{0})$ as defined in (3.5) and (3.6), so

$$\begin{aligned} X_1^* &= \sum_{j=1}^J X_1^j \sim \text{Geom}(\pi_0), \\ \{\mathbf{X}_1 | X_1^* = x^*\} &\sim \text{Multinomial}(x^*, \boldsymbol{\eta}^0), \end{aligned}$$

where $\boldsymbol{\eta}^0 = (\eta_1^0, \eta_2^0, \dots, \eta_J^0)$, and for $j = 1, 2, \dots, J$, $X_1^j \sim \text{Geom}(\pi_0 / \{\pi_0 + \eta_j^0 - \eta_j^0 \pi_0\})$.

Proof. By conditioning upon T_1 and the type of the initial individual, we have, for any $\boldsymbol{\theta} \in [0, 1]^J$, that

$$\mathbb{E} \left[\prod_{j=1}^J \theta_j^{X_1^j} \middle| T_1 < \infty \right] = \frac{\sum_{i=1}^J \chi_i \int_0^\infty H_D(\boldsymbol{\theta}; -t, t; \mathbf{e}_i) dt}{\sum_{i=1}^J \chi_i \int_0^\infty H_D(\mathbf{1}; -t, t; \mathbf{e}_i) dt}. \tag{5.7}$$

Using $\mathbf{d}_t = \boldsymbol{\epsilon}$ ($t \leq 0$), we have that

$$\sum_{i=1}^J \chi_i \int_0^\infty H_D(\boldsymbol{\theta}; -t, t; \mathbf{e}_i) dt = \sum_{j=1}^J \gamma_j \epsilon_j \frac{\partial}{\partial \theta_j} \left\{ \sum_{i=1}^J \chi_i \int_0^\infty H_E(\boldsymbol{\theta}; -t, t; \mathbf{e}_i) dt \right\}. \tag{5.8}$$

Since the branching (and detection) process is time-homogeneous prior to the first detected death at time $t = 0$, it follows from Lemma 5 and Section 4 that

$$H_E(\boldsymbol{\theta}; -t, t; \mathbf{e}_i) = q_i(-t, 0, t) + \frac{\bar{\zeta}(t) \left[\sum_{j=1}^J p_{ij}(-t, 0, t) \theta_j \right]}{1 - \sum_{j=1}^J \bar{\psi}_j(t) \theta_j}, \tag{5.9}$$

where $\bar{\zeta}(t) = \zeta(-t; t)$ and $\bar{\psi}_j(t) = \psi_j(-t; t)$; cf. Section 4.2.

Let $\bar{\Omega} = \Omega(\alpha_1, \boldsymbol{\epsilon})$ and $\bar{\Sigma}(t) = \Sigma(\alpha_1, \boldsymbol{\epsilon}, t)$ ($t \geq 0$), where $\Omega(\cdot, \cdot)$ and $\Sigma(\cdot, \cdot, \cdot)$ are defined in (4.16) and (4.17), respectively. Then

$$\frac{d}{dt} \bar{\Sigma}(t) = \bar{\Omega} \bar{\Sigma}(t).$$

Hence, for $i, j = 1, 2, \dots, J + 1$,

$$\frac{d}{dt} \bar{\Sigma}_{ij}(t) = \sum_{l=1}^{J+1} \bar{\Omega}_{il} \bar{\Sigma}_{lj}(t). \tag{5.10}$$

It follows from Lemma 4 that $\bar{\psi}_j(t) = -\bar{\Sigma}_{J+1,j}(t) / \bar{\Sigma}_{J+1,J+1}(t)$ and $\bar{\zeta}(t) = 1 / \bar{\Sigma}_{J+1,J+1}(t)$. Therefore, from (5.10), we have that

$$\begin{aligned} \frac{d}{dt} \bar{\psi}_j(t) &= \frac{-\bar{\Sigma}_{J+1,J+1}(t) \sum_{l=1}^{J+1} \bar{\Omega}_{J+1,l} \bar{\Sigma}_{lj}(t) + \bar{\Sigma}_{J+1,j}(t) \sum_{l=1}^{J+1} \bar{\Omega}_{J+1,l} \bar{\Sigma}_{l,J+1}(t)}{\bar{\Sigma}_{J+1,J+1}(t)^2} \\ &= - \left(\frac{\sum_{l=1}^{J+1} \bar{\Omega}_{J+1,l} \left[\bar{\Sigma}_{J+1,J+1}(t) \bar{\Sigma}_{lj}(t) - \bar{\Sigma}_{J+1,j}(t) \bar{\Sigma}_{l,J+1}(t) \right]}{\bar{\Sigma}_{J+1,J+1}(t)^2} \right). \end{aligned} \tag{5.11}$$

From (4.18), $p_{ij}(-t, 0, t) = \bar{\Sigma}_{lj}(t) + \bar{\psi}_j(t) \bar{\Sigma}_{l,J+1}(t)$ ($j, l = 1, 2, \dots, J$). Then, since the sum on the right-hand side of (5.11) is 0 for $l = J + 1$ and $\bar{\Omega}_{J+1,l} = -\alpha_1 \chi_l$ ($l = 1, 2, \dots, J$), we have that

$$\frac{d}{dt} \bar{\psi}_j(t) = \alpha_1 \bar{\zeta}(t) \sum_{l=1}^J \chi_l p_{lj}(-t, 0, t). \tag{5.12}$$

Therefore, substituting (5.9) and (5.12) into the integral on the right-hand side of (5.8), we have that

$$\begin{aligned} &\sum_{i=1}^J \chi_i \int_0^\infty H_E(\boldsymbol{\theta}; -t, t; \mathbf{e}_i) dt \\ &= \sum_{i=1}^J \chi_i \int_0^\infty \left\{ q_i(-t, 0, t) + \frac{\bar{\zeta}(t) \left[\sum_{j=1}^J p_{ij}(-t, 0, t) \theta_j \right]}{1 - \sum_{j=1}^J \bar{\psi}_j(t) \theta_j} \right\} dt \\ &= \int_0^\infty \left\{ \sum_{i=1}^J \chi_i q_i(-t, 0, t) \right\} dt + \int_0^\infty \frac{\bar{\zeta}(t) \sum_{j=1}^J \theta_j \left[\sum_{i=1}^J \chi_i p_{ij}(-t, 0, t) \right]}{1 - \sum_{j=1}^J \theta_j \bar{\psi}_j(t)} dt \\ &= K + \frac{1}{\alpha_1} \int_0^\infty \frac{\frac{d}{dt} \sum_{j=1}^J \theta_j \bar{\psi}_j(t)}{1 - \sum_{j=1}^J \theta_j \bar{\psi}_j(t)} dt, \end{aligned} \tag{5.13}$$

where $K = \sum_{i=1}^J \chi_i \int_0^\infty q_i(-t, 0, t) dt$ does not depend on θ .

Applying a change of variable $z = \sum_{j=1}^J \theta_j \bar{\psi}_j(t)$ in (5.13) and noting that $\bar{\psi}_j(\infty) = (1 - \pi_0)\eta_j^0$, we have that

$$\begin{aligned} \sum_{i=1}^J \chi_i \int_0^\infty H_E(\theta; -t, t; \mathbf{e}_i) dt &= K + \frac{1}{\alpha_1} \int_0^{\sum_{j=1}^J \theta_j(1-\pi_0)\eta_j^0} \frac{1}{1-z} dz \\ &= K - \frac{1}{\alpha_1} \log \left(1 - (1 - \pi_0) \sum_{j=1}^J \theta_j \eta_j^0 \right). \end{aligned} \tag{5.14}$$

Substituting (5.14) into (5.8) and taking partial derivatives yields

$$\sum_{i=1}^J \chi_i \int_0^\infty H_D(\theta; -t, t; \mathbf{e}_i) dt = \frac{\sum_{i=1}^J \gamma_i \epsilon_i (1 - \pi_0) \eta_i^0}{\alpha_1 (1 - (1 - \pi_0) \sum_{j=1}^J \eta_j^0 \theta_j)}. \tag{5.15}$$

It follows from (5.15) and (5.7) that

$$\begin{aligned} \mathbb{E} \left[\prod_{j=1}^J \theta_j^{X_1^j} \middle| T_1 < \infty \right] &= \frac{\sum_{i=1}^J \gamma_i \epsilon_i (1 - \pi_0) \eta_i^0}{\alpha_1 (1 - (1 - \pi_0) \sum_{j=1}^J \eta_j^0 \theta_j)} \times \frac{\alpha_1 \pi_0}{\sum_{i=1}^J \gamma_i \epsilon_i (1 - \pi_0) \eta_i^0} \\ &= \frac{\pi_0}{1 - (1 - \pi_0) \sum_{j=1}^J \eta_j^0 \theta_j}, \end{aligned}$$

as required.

The distributions of X_1^* , $\mathbf{X}_1|X_1^*$, and X_1^j can all be obtained straightforwardly from (5.6). \square

The assumption that the birth rate and detection probabilities are constant prior to the first detected death are required in Lemma 6 to show that $\mathbf{X}_1 \sim \mathbf{W}(0, \mathbf{0})$. In the supplementary material we show that the distribution of \mathbf{X}_1 can also be obtained in the case where the birth time of the initial individual, S_0 , is known and equal to $-t_0$, say.

We proceed by studying the dynamics of a branching process from time t onwards with

$$\mathbf{X}_\kappa = \mathbf{Y}(t) = \mathbf{W}(t, \mathbf{a}). \tag{5.16}$$

Remember that (5.16) implies that $X_\kappa^* = Y^*(t) = W_*(t; \mathbf{a})$ is a zero-modified geometric random variable with success probability π_t . In Lemma 7 we consider the distribution of $\mathbf{Y}(t + \tau)$ given that there are no detected deaths in the interval $(t, t + \tau]$, the event $\tilde{E}_{t;\tau}^{\mathbf{a}}$ defined in Section 3, and in Lemma 8 we consider the distribution of $X_{\kappa+1}$ given that $T_{\kappa+1} = \tau$.

In the first instance, Lemmas 7 and 8 are useful in the case $t = 0 (= s_1)$ and $\tau = t_2 (= s_2)$, with $\mathbf{X}_1 = \mathbf{Y}(0) = \mathbf{W}(0, \mathbf{0})$, where we consider the behaviour of the branching process between the first and second detected death and at the second detected death.

Lemma 7. For $t \geq 0$ and $\mathbf{a} > -(1 - \pi_t)\boldsymbol{\eta}^t$, suppose that $\mathbf{Y}(t) \sim \mathbf{W}(t, \mathbf{a})$. In addition, for $\tau \geq 0$, let $\mathbf{b} (= \mathbf{b}(t, \tau; \mathbf{a}))$ satisfy (3.7).

Then $U_E(t, \tau; \mathbf{a}) = \mathbb{P}(\tilde{E}_{t;\tau}^{\mathbf{a}})$ (see (3.10)) satisfies

$$U_E(t, \tau; \mathbf{a}) = \frac{\pi_t \times \left[1 + \sum_{i=1}^J a_i q_i(t, 0, \tau) \right] \times \left[1 + \sum_{j=1}^J b_j \right]}{\pi_{t+\tau} \times \left[1 - (1 - \pi_t) \sum_{i=1}^J \eta_i^t q_i(t, 0, \tau) \right] \times \left[1 + \sum_{j=1}^J a_j \right]}, \tag{5.17}$$

$$\mathbb{E} [H_E(\boldsymbol{\theta}; t, \tau; \mathbf{W}(t, \mathbf{a}))] = U_E(t, \tau; \mathbf{a})\varphi(\boldsymbol{\theta}; t + \tau, \mathbf{b}), \tag{5.18}$$

and

$$\mathbb{E} \left[\prod_{j=1}^J \theta_j^{Y_j^{t+\tau}} \middle| \tilde{E}_{t,\tau}^{\mathbf{a}}, \mathbf{Y}(t) \sim \mathbf{W}(t, \mathbf{a}) \right] = \varphi(\boldsymbol{\theta}; t + \tau, \mathbf{b}). \tag{5.19}$$

Thus, $\{\mathbf{Y}(t + \tau) | \tilde{E}_{t,\tau}^{\mathbf{a}}, \mathbf{Y}(t) \sim \mathbf{W}(t, \mathbf{a})\} \sim \mathbf{W}(t + \tau, \mathbf{b})$.

Proof. For $t, \tau \geq 0$ and $\boldsymbol{\theta} \in [0, 1]^J$, let

$$\xi_i(\boldsymbol{\theta}; t, \tau) = q_i(t, 0, \tau) + \frac{\zeta(t; \tau) \sum_{j=1}^J p_{ij}(t, 0, \tau)\theta_j}{1 - \sum_{j=1}^J \psi_j(t; \tau)\theta_j}.$$

Therefore it follows from Lemma 5 and (3.4) that, for $\boldsymbol{\theta} \in [0, 1]^J$,

$$\begin{aligned} \mathbb{E} [H_E(\boldsymbol{\theta}; t, \tau; \mathbf{W}(t, \mathbf{a}))] &= \mathbb{E} \left[\prod_{j=1}^J \xi_j(\boldsymbol{\theta}; t, \tau)^{W_j(t, \mathbf{a})} \right] \\ &= \frac{1 + \sum_{i=1}^J a_i \xi_i(\boldsymbol{\theta}; t, \tau)}{1 + \sum_{i=1}^J a_i} \times \frac{\pi_t}{1 - (1 - \pi_t) \sum_{j=1}^J \eta_j^t \xi_j(\boldsymbol{\theta}; t, \tau)}. \end{aligned} \tag{5.20}$$

Note that

$$\begin{aligned} &\frac{1 + \sum_{i=1}^J a_i \xi_i(\boldsymbol{\theta}; t, \tau)}{1 - (1 - \pi_t) \sum_{i=1}^J \eta_i^t \xi_i(\boldsymbol{\theta}; t, \tau)} \\ &= \frac{1 - \sum_{i=1}^J \psi_i(t; \tau)\theta_i + \sum_{i=1}^J a_i \left(q_i(t, 0, \tau)[1 - \sum_{j=1}^J \psi_j(t; \tau)\theta_j] + \zeta(t; \tau) \left[\sum_{j=1}^J \theta_j p_{ij}(t, 0, \tau) \right] \right)}{1 - \sum_{i=1}^J \psi_i(t; \tau)\theta_i - (1 - \pi_t) \sum_{i=1}^J \eta_i^t \left(q_i(t, 0, \tau)[1 - \sum_{j=1}^J \psi_j(t; \tau)\theta_j] + \zeta(t; \tau) \left[\sum_{j=1}^J \theta_j p_{ij}(t, 0, \tau) \right] \right)}. \end{aligned} \tag{5.21}$$

Using (3.2), we have that the denominator in (5.21) satisfies

$$\begin{aligned} &\left[1 - (1 - \pi_t) \sum_{j=1}^J \eta_j^t q_j(t, 0, \tau) \right] \left\{ 1 - \sum_{i=1}^J \theta_i \left[\psi_i(t; \tau) + \frac{(1 - \pi_t)\zeta(t; \tau) \sum_{j=1}^J \eta_j^t p_{ji}(t, 0, \tau)}{1 - (1 - \pi_t) \sum_{j=1}^J \eta_j^t q_j(t, 0, \tau)} \right] \right\} \\ &= \left[1 - (1 - \pi_t) \sum_{j=1}^J \eta_j^t q_j(t, 0, \tau) \right] \times \left[1 - (1 - \pi_{t+\tau}) \sum_{l=1}^J \theta_l \eta_l^{t+\tau} \right]. \end{aligned} \tag{5.22}$$

Thus, substituting (5.22) into (5.21) and using (3.7) yields

$$\begin{aligned} &\frac{1 + \sum_{i=1}^J a_i \xi_i(\boldsymbol{\theta}; t, \tau)}{1 - (1 - \pi_t) \sum_{i=1}^J \eta_i^t \xi_i(\boldsymbol{\theta}; t, \tau)} \\ &= \frac{1 + \sum_{i=1}^J a_i q_i(t, 0, \tau) + \sum_{j=1}^J \theta_j \left[-\psi_j(t; \tau) \left\{ 1 + \sum_{i=1}^J a_i q_i(t, 0, \tau) \right\} + \zeta(t; \tau) \sum_{i=1}^J a_i p_{ij}(t, 0, \tau) \right]}{\left[1 - (1 - \pi_t) \sum_{j=1}^J \eta_j^t q_j(t, 0, \tau) \right] \times \left[1 - (1 - \pi_{t+\tau}) \sum_{l=1}^J \theta_l \eta_l^{t+\tau} \right]} \\ &= \frac{\left[1 + \sum_{i=1}^J a_i q_i(t, 0, \tau) \right] \times \left[1 + \sum_{j=1}^J b_j \theta_j \right]}{\left[1 - (1 - \pi_t) \sum_{j=1}^J \eta_j^t q_j(t, 0, \tau) \right] \times \left[1 - (1 - \pi_{t+\tau}) \sum_{l=1}^J \theta_l \eta_l^{t+\tau} \right]}. \end{aligned} \tag{5.23}$$

Then (5.18) follows from (5.20) and (5.23) by straightforward algebraic manipulation.

By setting $\boldsymbol{\theta} = \mathbf{1}$ in (5.18), we obtain (5.17).

Finally, (5.19) follows immediately from

$$\begin{aligned} \varphi(\boldsymbol{\theta}; t + \tau, \mathbf{b})U_E(t, \tau; \mathbf{a}) &= \mathbb{E} [H_E(\boldsymbol{\theta}; t, \tau; \mathbf{W}(t, \mathbf{a}))] \\ &= \mathbb{E} \left[\prod_{j=1}^J \theta_j^{Y_j^{(t+\tau)}} \mathbf{1}_{\{\tilde{E}_{t,\tau}^{\mathbf{a}}\}} \middle| \mathbf{Y}(t) \sim \mathbf{W}(t, \mathbf{a}) \right] \\ &= \mathbb{E} \left[\prod_{j=1}^J \theta_j^{Y_j^{(t+\tau)}} \middle| \tilde{E}_{t,\tau}^{\mathbf{a}}, \mathbf{Y}(t) \sim \mathbf{W}(t, \mathbf{a}) \right] \mathbb{P}(\tilde{E}_{t,\tau}^{\mathbf{a}} | \mathbf{Y}(t) \sim \mathbf{W}(t, \mathbf{a})) \\ &= \mathbb{E} \left[\prod_{j=1}^J \theta_j^{Y_j^{(t+\tau)}} \middle| \tilde{E}_{t,\tau}^{\mathbf{a}}, \mathbf{Y}(t) \sim \mathbf{W}(t, \mathbf{a}) \right] U_E(t, \tau; \mathbf{a}). \quad \square \end{aligned}$$

The following observations, which are consequences of Lemma 7, prove useful in the derivation of the approximation in Section 3.5.2. For $t \geq 0$, let $G_t \sim \text{Geom}(\pi_t)$. Then if $\mathbf{Y}(t) \sim \mathbf{W}(t, \mathbf{0})$, we have that $Y^*(t) \sim G_t$. As noted prior to (3.19), $b_i(t, \tau; \mathbf{0}) = -\psi_i(t; \tau)$ ($i = 1, 2, \dots, J$), so it follows from (5.19) that, for $0 \leq \vartheta \leq 1$,

$$\mathbb{E} \left[\vartheta^{Y^*(t+\tau)} \middle| \tilde{E}_{t,\tau}^{\mathbf{0}}, \mathbf{Y}(t) \sim \mathbf{W}(t; \mathbf{0}) \right] = \frac{1 - \psi(t; \tau)\vartheta}{1 - \psi(t; \tau)} \times \frac{\pi_{t+\tau}}{1 - (1 - \pi_{t+\tau})\vartheta}. \tag{5.24}$$

Since $0 < \psi(t; \tau) < 1 - \pi_{t+\tau}$, (see (3.1)) it follows from (5.24) and Lemma 1 that

$$\left\{ Y^*(t + \tau) \middle| E_{t,\tau}^{\mathbf{Y}(t)}, Y^*(t) \sim G_t \right\} \sim \begin{cases} G_{t+\tau} & \text{with probability } h(t; \tau), \\ 0 & \text{with probability } 1 - h(t; \tau), \end{cases} \tag{5.25}$$

where, for $t, \tau \geq 0$,

$$h(t; \tau) = \frac{1 - \pi_{t+\tau} - \psi(t; \tau)}{(1 - \pi_{t+\tau})(1 - \psi(t; \tau))}. \tag{5.26}$$

We turn to the distribution of $\mathbf{X}_{\kappa+1}$ given that $\mathbf{X}_{\kappa} \sim \mathbf{W}(t, \mathbf{a})$. Recall from (3.11) that $U_D(t, \tau; \mathbf{a}) = -\frac{\partial}{\partial \tau} U_E(t, \tau; \mathbf{a})$, where $U_E(t, \tau; \mathbf{a})$ is defined in (3.10) and derived in (5.17).

Lemma 8. *For some $\kappa = 1, 2, \dots$, suppose that $s_{\kappa} = t$ with $\mathbf{X}_{\kappa} \sim \mathbf{W}(t, \mathbf{a})$. For $\tau \geq 0$, let $\mathbf{b} (= \mathbf{b}(t, \tau; \mathbf{a}))$ and $\mathbf{c} (= \mathbf{c}(t, \tau; \mathbf{a}))$ satisfy (3.7) and (3.8). Then*

$$\begin{aligned} f_{T_{\kappa+1}}(\tau | S_{\kappa} = t, \mathbf{X}_{\kappa} \sim \mathbf{W}(t, \mathbf{a})) &= \frac{U_E(t, \tau; \mathbf{a}) \sum_{j=1}^J d_{t+\tau, j} \gamma_j \left[b_j \pi_{t+\tau} + (1 - \pi_{t+\tau}) \eta_j^{t+\tau} (1 + \sum_{i=1}^J b_i) \right]}{\pi_{t+\tau} (1 + \sum_{i=1}^J b_i)} \\ &= -\frac{\partial}{\partial \tau} U_E(t, \tau; \mathbf{a}) = U_D(t, \tau; \mathbf{a}), \end{aligned} \tag{5.27}$$

and

$$\mathbb{E} \left[\prod_{j=1}^J \theta_j^{X_j^{\kappa+1}} \middle| T_{\kappa+1} = \tau, \mathbf{X}_{\kappa} \sim \mathbf{W}(t, \mathbf{a}) \right] = \varphi(t + \tau, \mathbf{0})\varphi(t + \tau, \mathbf{c}). \tag{5.28}$$

Thus,

$$\{\mathbf{X}_{\kappa+1} | S_{\kappa} = t, T_{\kappa+1} = \tau, \mathbf{X}_{\kappa} \sim \mathbf{W}(t, \mathbf{a})\} \sim \mathbf{W}(t + \tau, \mathbf{c}) + \mathbf{W}(t + \tau, \mathbf{0}), \tag{5.29}$$

where $\mathbf{W}(t + \tau, \mathbf{c})$ and $\mathbf{W}(t + \tau, \mathbf{0})$ are independent random variables.

Proof. For $\theta \in [0, 1]^J$, it follows from Corollary 2, (5.5), and Lemma 7, (5.18), that

$$\begin{aligned} \mathbb{E}[H_D(\theta; t, \tau; \mathbf{W}(t, \mathbf{a}))] &= \sum_{j=1}^J \gamma_j d_{t+\tau, j} \frac{\partial}{\partial \theta_j} \mathbb{E}[H_E(\theta; t, \tau; \mathbf{W}(t, \mathbf{a}))] \\ &= \sum_{j=1}^J \gamma_j d_{t+\tau, j} U_E(t, \tau; \mathbf{a}) \frac{\partial}{\partial \theta_j} \varphi(t + \tau; \mathbf{b}). \end{aligned} \tag{5.30}$$

The partial derivative of $\varphi(t + \tau; \mathbf{b})$ with respect to θ_j ($j = 1, 2, \dots, J$) is given by

$$\begin{aligned} \frac{\partial}{\partial \theta_j} \varphi(t + \tau; \mathbf{b}) &= \frac{\pi_{t+\tau}}{1 + \sum_{i=1}^J b_i} \\ &\times \frac{b_j \left[1 - (1 - \pi_{t+\tau}) \sum_{l=1}^J \theta_l \eta_l^{t+\tau} \right] + (1 - \pi_{t+\tau}) \eta_j^{t+\tau} \left(1 + \sum_{i=1}^J b_i \theta_i \right)}{\left[1 - (1 - \pi_{t+\tau}) \sum_{l=1}^J \theta_l \eta_l^{t+\tau} \right]^2}. \end{aligned} \tag{5.31}$$

By setting $\theta = \mathbf{1}$ in (5.31) and using (5.1) and (5.30), we obtain (5.27).

Since

$$\mathbb{E} \left[\prod_{j=1}^J \theta_j^{X_{\kappa+1}^j} \middle| T_{\kappa+1} = \tau, \mathbf{X}_{\kappa} \sim \mathbf{W}(t, \mathbf{a}) \right] = \frac{\mathbb{E}[H_D(\theta; t, \tau; \mathbf{W}(t, \mathbf{a}))]}{f_{T_{\kappa+1}}(\tau | S_{\kappa} = t, \mathbf{X}_{\kappa} \sim \mathbf{W}(t, \mathbf{a}))}$$

(cf. (5.1)), it follows that

$$\begin{aligned} &\mathbb{E} \left[\prod_{j=1}^J \theta_j^{X_{\kappa+1}^j} \middle| T_{\kappa+1} = \tau, \mathbf{X}_{\kappa} \sim \mathbf{W}(t, \mathbf{a}) \right] \\ &= \frac{\pi_{t+\tau}^2}{\left[1 - (1 - \pi_{t+\tau}) \sum_{l=1}^J \theta_l \eta_l^{t+\tau} \right]^2} \\ &\times \frac{\sum_{j=1}^J d_{t+\tau, j} \gamma_j \left(b_j \left[1 - (1 - \pi_{t+\tau}) \sum_{l=1}^J \theta_l \eta_l^{t+\tau} \right] + (1 - \pi_{t+\tau}) \eta_j^{t+\tau} \left(1 + \sum_{i=1}^J b_i \theta_i \right) \right)}{\sum_{j=1}^J d_{t+\tau, j} \gamma_j \left[b_j \pi_{t+\tau} + (1 - \pi_{t+\tau}) \eta_j^{t+\tau} \left(1 + \sum_{i=1}^J b_i \theta_i \right) \right]}. \end{aligned} \tag{5.32}$$

The second fraction on the right-hand side of (5.32) simplifies to

$$\begin{aligned} & \frac{\sum_{j=1}^J d_{t+\tau,j} \gamma_j \left(b_j \left[1 - (1 - \pi_{t+\tau}) \sum_{l=1}^J \theta_l \eta_l^{t+\tau} \right] + (1 - \pi_{t+\tau}) \eta_j^{t+\tau} \left(1 + \sum_{i=1}^J b_i \theta_i \right) \right)}{\sum_{j=1}^J d_{t+\tau,j} \gamma_j \left[b_j \pi_{t+\tau} + (1 - \pi_{t+\tau}) \eta_j^{t+\tau} (1 + \sum_{i=1}^J b_i) \right]} \\ &= \frac{\sum_{j=1}^J d_{t+\tau,j} \gamma_j \left[(1 - \pi_{t+\tau}) \eta_j^{t+\tau} + b_j \right] + \sum_{l=1}^J \theta_l (1 - \pi_{t+\tau}) \sum_{j=1}^J d_{t+\tau,j} \gamma_j \left(b_l \eta_j^{t+\tau} - b_j \eta_l^{t+\tau} \right)}{\sum_{j=1}^J d_{t+\tau,j} \gamma_j \left[(1 - \pi_{t+\tau}) \eta_j^{t+\tau} + b_j \right] + \sum_{l=1}^J (1 - \pi_{t+\tau}) \sum_{j=1}^J d_{t+\tau,j} \gamma_j \left(b_l \eta_j^{t+\tau} - b_j \eta_l^{t+\tau} \right)} \\ &= \frac{\left(\sum_{j=1}^J d_{t+\tau,j} \gamma_j \left[(1 - \pi_{t+\tau}) \eta_j^{t+\tau} + b_j \right] \right) \times \left(1 + \sum_{l=1}^J c_l \theta_l \right)}{\left(\sum_{j=1}^J d_{t+\tau,j} \gamma_j \left[(1 - \pi_{t+\tau}) \eta_j^{t+\tau} + b_j \right] \right) \times \left(1 + \sum_{l=1}^J c_l \right)} = \frac{1 + \sum_{l=1}^J c_l \theta_l}{1 + \sum_{l=1}^J c_l}. \end{aligned} \tag{5.33}$$

Since $c > -(1 - \pi_{t+\tau}) \eta^{t+\tau}$, we can combine (5.32) and (5.33) to obtain (5.28), completing the proof of the lemma. \square

The following observations, which are consequences of Lemma 8, prove useful in the derivation of Section 3.5.2. The distribution of the size of the population immediately following the $(\kappa + 1)$ th detected death, given that we have $\mathbf{X}_\kappa \sim \mathbf{W}(s_\kappa; \mathbf{0})$, depends on the sign of

$$C_{s_\kappa; t_{\kappa+1}}^* = \sum_{i=1}^J C_{s_\kappa; t_{\kappa+1}}^i,$$

where $C_{s_\kappa; t_{\kappa+1}}^i$ ($i = 1, 2, \dots, J$) are given by (3.19). In particular, it follows from (5.29) and Lemma 1 that if $C_{s_\kappa; t_{\kappa+1}}^* \leq 0$ with $s_{\kappa+1} = s_\kappa + t_{\kappa+1}$, then

$$\left\{ \mathbf{X}_{\kappa+1}^* \mid \mathbf{X}_\kappa \sim \mathbf{W}(s_\kappa; \mathbf{0}) \right\} \sim \begin{cases} G_{s_{\kappa+1}}^1 + G_{s_{\kappa+1}}^2 & \text{with probability } \tilde{h}(s_\kappa, t_{\kappa+1}), \\ G_{s_{\kappa+1}} & \text{with probability } 1 - \tilde{h}(s_\kappa, t_{\kappa+1}), \end{cases} \tag{5.34}$$

where $\tilde{h}(s_\kappa, t_{\kappa+1})$ is defined in (3.20) and $G_{s_{\kappa+1}}^1$ and $G_{s_{\kappa+1}}^2$ are i.i.d. copies of $G_{s_{\kappa+1}}$.

We are now in position to prove Theorem 1.

Proof of Theorem 1 The theorem is proved by induction.

By Lemma 6, we have that $\mathbf{X}_1 \sim \mathbf{W}(0, \mathbf{0})$, and by Lemma 8, we have that $\mathbf{X}_2 | T_2 = t_2 \sim \mathbf{W}(s_2, \mathbf{C}_{0;t_2}) + \mathbf{W}(s_2, \mathbf{0})$, where the components of $\mathbf{C}_{0;t_2}$ are given by (3.19) and $\mathbf{W}(s_2, \mathbf{C}_{0;t_2})$ and $\mathbf{W}(s_2, \mathbf{0})$ are independent.

For $k = 2, 3, \dots$, we have that $\mathbf{X}_k | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}$ satisfies (3.12). Therefore suppose that $\mathbf{X}_k | \mathbf{T}_{2:k} = \mathbf{t}_{2:k} \sim \sum_{l=1}^k \mathbf{W}_{ml}(s_k; \mathbf{a}_{ml}^k)$ for some $m = 1, 2, \dots, M_k (= (k - 1)!)$ and note that this occurs with probability w_m^k . We say that the population immediately after the k th detected death consists of k independent family groups, where the number of individuals of each type in the l th family group is given by $\mathbf{W}_{ml}(s_k; \mathbf{a}_{ml}^k)$. Note that a family group can contain 0 individuals. Each family group evolves independently with births of new individuals, individuals changing type, and deaths of individuals. Given that the first detected death after time s_k is at time $s_{k+1} = s_k + t_{k+1}$, one of the family groups is responsible for the death, whilst the other $k - 1$ family groups do not experience a detected death in the interval $(s_k, s_{k+1}]$. Hence, it follows that the joint probability density for $T_{k+1} = t_{k+1}$ and the i th family group being responsible for the death is

$$U_D(s_k, t_{k+1}; \mathbf{a}_{mi}^k) \prod_{l \neq i} U_E(s_k, t_{k+1}; \mathbf{a}_{ml}^k). \tag{5.35}$$

Since $\mathbf{X}_k | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}$ is a mixture of $\{ \sum_{l=1}^k \mathbf{W}_{ml}(s_k; \mathbf{a}_{ml}) \}$ with weights w_m^k , we have that

$$f_{T_{k+1}}(t_{k+1} | S_k = s_k, \mathbf{X}_k) = \sum_{m=1}^{M_k} w_m^k \sum_{i=1}^k U_D(s_k, t_{k+1}; \mathbf{a}_{mi}^k) \prod_{l \neq i} U_E(s_k, t_{k+1}; \mathbf{a}_{ml}^k). \tag{5.36}$$

It follows from (5.35) and (5.36) that, given $T_{k+1} = t_{k+1}$, the probability of the detected death coming from the i th family group in the m th mixture distribution is w_{mk+i}^{k+1} as defined in (3.14).

Given that the i th family group in the m th mixture distribution is responsible for the $(k + 1)$ th detected death at time s_{k+1} , we have $k + 1$ family groups at time s_{k+1} . The $k - 1$ family groups not responsible for the detected deaths will have evolved by Lemma 7 into family groups having different distributions. Specifically, the l th family group ($l \neq i$) which had numbers of individuals of each type distributed according to $\mathbf{W}(s_k, \mathbf{a}_{ml}^k)$ will have, given no detected death in $(s_k, s_{k+1}]$, numbers of individuals of each type distributed according to $\mathbf{W}(s_{k+1}, \mathbf{b}_{ml}^k)$. By Lemma 8, we can split the family group i , in which the detected death occurs at s_{k+1} , into two independent family groups, with the numbers of individuals of each type in the two family groups distributed according to $\mathbf{W}(s_{k+1}, \mathbf{c}_{ml}^k)$ and $\mathbf{W}(s_{k+1}, \mathbf{0})$, respectively. The labellings of $\{\mathbf{a}_{(mk+i)j}^{k+1}\}$ given in (3.15) follow.

Hence, (3.13) is proved and the theorem follows. □

The description of family groups given in Theorem 1 differs from the description in [2, Theorem 3.2]. In [2], the number of family groups immediately after the k th detected death is random with support $\{2, 3, \dots, k\}$ for $k \geq 2$, and its distribution is dependent on the times of the detected deaths. The size of each family group is i.i.d. according to $\text{Geom}(\pi_{s_k})$. In Theorem 1, there are k family groups immediately after the k th detected death, with the sizes of the family groups being independent but not identically distributed, and their distributions depending on the times of the detected deaths.

It is possible to obtain a result similar to Theorem 1 if the time S_0 at which the first individual is born is known and is equal to $-t_0$, say; the details are provided in the supplementary material. In that case it can be shown using Corollary 2 and (5.2) that the distribution of \mathbf{X}_1 is the sum of two independent zero-modified geometric random variables with $\pi_0 = 1 - \psi(-t_0; t_0)$; cf. the discussion after the proof of [2, Theorem 3.1]. The distribution of $\mathbf{X}_{2:k} | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}$ can then be obtained as above, except that it consists of a mixture of $k!$ random variables, each consisting of $k + 1$ independent zero-modified geometric random variables.

6. Derivation of the approximation in Section 3.5.2

In this section we construct the approximation $\hat{\mathbf{X}}_k$ and $\hat{X}_k^* = \sum_{i=1}^k \hat{X}_k^i$ to \mathbf{X}_k and X_k^* , where \hat{X}_k^* is a mixture of negative binomial distributions and $\hat{\mathbf{X}}_k | X_k^*$ follows a multinomial distribution. As stated in Section 3.5.2, we set $\hat{\mathbf{X}}_1 = \mathbf{X}_1 \sim \hat{\mathbf{W}}(0; \boldsymbol{\eta}^0)$.

For $k = 1, 2, \dots$, we outline the construction of $\hat{X}_{k+1}^* | \mathbf{T}_{2:k+1} = \mathbf{t}_{2:k+1}$ and $\hat{\mathbf{X}}_{k+1} | \mathbf{T}_{2:k+1} = \mathbf{t}_{2:k+1}$ given that $\hat{\mathbf{X}}_k | \mathbf{T}_{2:k} = \mathbf{t}_{2:k} \sim \sum_{i=1}^{\hat{R}_k} \hat{\mathbf{W}}_i(s_k; \hat{\boldsymbol{\xi}}^k)$, where \hat{R}_k has support $\{1, 2, \dots, k\}$, $\hat{P}_k^m = \mathbb{P}(\hat{R}_k = m)$ ($k = 1, 2, \dots; m = 1, 2, \dots, k$), and the $\hat{\mathbf{W}}_i(s_k; \hat{\boldsymbol{\xi}}^k)$ are i.i.d. according to $\hat{\mathbf{W}}(s_k; \hat{\boldsymbol{\xi}}^k)$ given by (3.23) and (3.24). That is,

$$\hat{X}_k^* | \mathbf{T}_{2:k} = \mathbf{t}_{2:k} \sim \text{NegBin}(\hat{R}_k, \pi_{s_k}), \tag{6.1}$$

$$\hat{\mathbf{X}}_k | \hat{X}_k^* = x^*, \mathbf{T}_{2:k} = \mathbf{t}_{2:k} \sim \text{Multinomial}(x^*, \hat{\boldsymbol{\xi}}^k). \tag{6.2}$$

Let $\tilde{\mathbf{X}}_{k+1} | \mathbf{T}_{2:k+1} = \mathbf{t}_{2:k+1}$ denote the distribution of the number of individuals alive of each type immediately after the $(k + 1)$ th detected death, given that immediately after the k th detected death, at time s_k , the distribution of the number of individuals alive of each type is given by $\hat{\mathbf{X}}_k | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}$ defined in (6.1) and (6.2). We obtain the distribution of $\tilde{\mathbf{X}}_{k+1} | \mathbf{T}_{2:k+1} = \mathbf{t}_{2:k+1}$ using minor modifications of Lemmas 7 and 8, by arguments along similar lines to those used in the proof of Theorem 1 and [2, Section 4].

Suppose that $\hat{R}_k = j$ for some $j = 1, 2, \dots, k$. Then at time s_k , the population consists of j i.i.d. family groups, where the number of individuals of each type within a family group is distributed according to $\hat{\mathbf{W}}(s_k; \hat{\boldsymbol{\zeta}}^k)$. (As in the proof of Theorem 1, a family group may contain no individuals.) One of these j family groups will be responsible for the next detected death in the population, at time s_{k+1} , and the distribution of the number of individuals alive in that family group and their types are obtained from Lemma 8 by setting $\mathbf{a} = \mathbf{0}$ and $\boldsymbol{\eta}^{s_k} = \hat{\boldsymbol{\zeta}}^k$. The other $j - 1$ family groups will not experience a detected death in the interval $(s_k, s_{k+1}]$, and the distribution of the number of individuals alive in these family groups and their types are obtained by setting $\mathbf{a} = \mathbf{0}$ and $\boldsymbol{\eta}^{s_k} = \hat{\boldsymbol{\zeta}}^k$ in Lemma 7.

For each of the family groups which do not experience a detected death in the interval $(s_k, s_{k+1}]$, the distribution of the number of individuals alive at time s_{k+1} is given by (5.25) with $t = s_k$ and $\tau = t_{k+1}$, a mixture distribution of $G_{s_{k+1}}$ with probability $h(s_k; t_{k+1})$ and a point mass at 0 with probability $1 - h(s_k; t_{k+1})$, where $h(t; \tau)$ is defined in (5.26). Similarly, for the family group which does experience the death, the distribution of the number of individuals alive at time s_{k+1} , immediately after the death, is, by (5.34), a mixture of the sum of two independent copies of $G_{s_{k+1}}$ with probability $\tilde{h}(s_k; t_{k+1})$ and $G_{s_{k+1}}$ with probability $1 - \tilde{h}(s_k; t_{k+1})$. Hence,

$$\tilde{\mathbf{X}}_{k+1}^* = \sum_{j=1}^J X_{k+1}^j \left| \hat{R}_k = l, T_{k+1} = t_{k+1} \sim \text{NegBin}(\tilde{R}_{k+1}, \pi_{s_{k+1}}), \right.$$

where

$$\tilde{R}_{k+1} | \hat{R}_k = l, T_{k+1} = t_{k+1} \sim 1 + \text{Bin}(1, \tilde{h}(s_k; t_{k+1})) + \text{Bin}(l - 1, h(s_k; t_{k+1})) \tag{6.3}$$

and the two binomial random variables are independent. We observe that (6.3) is similar to [2, (4.15)], and hence, we follow [2, Lemma 4.5] in the derivation of the probability mass function of $\{\tilde{R}_{k+1} | \hat{R}_k = l, T_{k+1} = t_{k+1}\}$.

Remember that we set $\boldsymbol{\eta}^{s_k} = \hat{\boldsymbol{\zeta}}^k$ and $L(s_k, \tau)$ is defined in (3.22). Therefore, from (5.17) and (5.27), with $\mathbf{a} = \mathbf{0}$ and noting from (3.7) that $b_i(t, \tau; \mathbf{0}) = -\psi_i(t; \tau)$, we have, for $\tau \geq 0$, that

$$U_E(s_k, \tau; \mathbf{0}) = \frac{\pi_{s_k}(1 - \psi(s_k; \tau))}{\pi_{s_k+\tau}L(s_k, \tau)}, \tag{6.4}$$

and (also using (3.3))

$$U_D(s_k, \tau; \mathbf{0}) = \frac{\pi_{s_k} \left[\sum_{j=1}^J d_{s_k+\tau,j} \gamma_j \phi_j(s_k; \tau) + (1 - \pi_{s_k+\tau}) \sum_{j=1}^J d_{s_k+\tau,j} \gamma_j \left(\psi_j(s_k; \tau) - \eta_j^{s_k+\tau} \psi(s_k; \tau) \right) \right]}{\pi_{s_k+\tau}^2 L(s_k, \tau)}. \tag{6.5}$$

Therefore, it follows from (6.4) and (6.5) that the probability density function of T_{k+1} , given $\hat{\mathbf{X}}_k$ satisfies (3.26), is

$$\hat{f}_{T_{k+1}}(\tau | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}, \hat{R}_k = m) = m \left[\frac{\pi_{s_k}(1 - \psi(s_k; \tau))}{\pi_{s_{k+1}}L(s_k, \tau)} \right]^{m-1} U_D(s_k, \tau; \mathbf{0}) \quad (\tau \geq 0). \quad (6.6)$$

We use (6.3) and (6.6) to derive $\tilde{P}_{k+1}^j = \mathbb{P}(\tilde{R}_{k+1} = j | \mathbf{T}_{2:k+1} = \mathbf{t}_{2:k+1})$, which is given by

$$\begin{aligned} \tilde{P}_{k+1}^j &= \mathbb{P}(\tilde{R}_{k+1} = j | \mathbf{T}_{2:k+1} = \mathbf{t}_{2:k+1}) \\ &= \sum_{l=1}^k \mathbb{P}(\hat{R}_{k+1} = j, \hat{R}_k = l | \mathbf{T}_{2:k+1} = \mathbf{t}_{2:k+1}) \\ &= \sum_{l=1}^k \frac{\mathbb{P}(\hat{R}_{k+1} = j | \hat{R}_k = l, T_{k+1} = t_{k+1}) \hat{f}_{T_{k+1}}(t_{k+1} | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}, \hat{R}_k = l) \hat{P}_k^l}{\hat{f}_{T_{k+1}}(t_{k+1} | \mathbf{T}_{2:k} = \mathbf{t}_{2:k})}. \end{aligned} \quad (6.7)$$

Using (6.6), we have that the denominator in (6.7) is given by

$$\begin{aligned} \hat{f}_{T_{k+1}}(t_{k+1} | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}) &= \sum_{m=1}^k \hat{f}_{T_{k+1}}(t_{k+1} | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}, \hat{R}_k = m) \hat{P}_k^m \\ &= \sum_{m=1}^k m \left[\frac{\pi_{s_k}(1 - \psi(s_k; t_{k+1}))}{\pi_{s_{k+1}}L(s_k, t_{k+1})} \right]^{m-1} U_D(s_k, t_{k+1}; \mathbf{0}) \hat{P}_k^m \\ &= U_D(s_k, t_{k+1}; \mathbf{0}) \sum_{m=1}^k m \left[\frac{\pi_{s_k}(1 - \psi(s_k; t_{k+1}))}{\pi_{s_{k+1}}L(s_k, t_{k+1})} \right]^{m-1} \hat{P}_k^m. \end{aligned} \quad (6.8)$$

Similarly, the numerator in (6.7) is given by

$$\begin{aligned} &\sum_{l=1}^k \mathbb{P}(\hat{R}_{k+1} = j | \hat{R}_k = l, T_{k+1} = t_{k+1}) \hat{f}_{T_{k+1}}(t_{k+1} | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}, \hat{R}_k = l) \hat{P}_k^l \\ &= \sum_{l=1}^k \left[\tilde{h}(s_k; t_{k+1}) \binom{l-1}{j-2} h(s_k; t_{k+1})^{j-2} \{1 - h(s_k; t_{k+1})\}^{l+1-j} \right. \\ &\quad \left. + \{1 - \tilde{h}(s_k; t_{k+1})\} \binom{l-1}{j-1} h(s_k; t_{k+1})^{j-1} \{1 - h(s_k; t_{k+1})\}^{l-j} \right] \\ &\quad \times l \left[\frac{\pi_{s_k}(1 - \psi(s_k; t_{k+1}))}{\pi_{s_{k+1}}L(s_k, t_{k+1})} \right]^{l-1} U_D(s_k, t_{k+1}; \mathbf{0}) \hat{P}_k^l. \end{aligned} \quad (6.9)$$

We note that the term $U_D(s_k, t_{k+1}; \mathbf{0})$ appears both in the numerator, (6.9), and the denominator, (6.8), and hence cancels in the calculation of \tilde{P}_{k+1}^j .

Using (5.26) and (3.3), summed over j , we have

$$h(s_k; t_{k+1}) \frac{\pi_{s_k}(1 - \psi(s_k; t_{k+1}))}{\pi_{s_{k+1}}L(s_k, t_{k+1})} = \frac{\pi_{s_k} \phi(s_k; t_{k+1})}{\pi_{s_{k+1}}(1 - \pi_{s_{k+1}})L(s_k, t_{k+1})},$$

where $\phi(s_k; t_{k+1}) = \sum_{j=1}^J \phi_j(s_k; t_{k+1})$, and

$$(1 - h(s_k; t_{k+1})) \frac{\pi_{s_k}(1 - \psi(s_k; t_{k+1}))}{\pi_{s_{k+1}}L(s_k, t_{k+1})} = \frac{\pi_{s_k} \psi(s_k; t_{k+1})}{(1 - \pi_{s_{k+1}})L(s_k, t_{k+1})}.$$

It then follows from (6.7), by (6.8) and (6.9), given $\hat{\mathbf{P}}^k$, that $\tilde{\mathbf{P}}^{k+1} = (\tilde{P}_{k+1}^1, \tilde{P}_{k+1}^2, \dots, \tilde{P}_{k+1}^{k+1})$ satisfies

$$\tilde{\mathbf{P}}_{k+1} = \frac{1}{K_k} \hat{\mathbf{P}}_k \left[\tilde{h}(s_k; t_{k+1}) \mathbf{M}_k^1(s_k; t_{k+1}) + \{1 - \tilde{h}(s_k; t_{k+1})\} \mathbf{M}_k^2(s_k; t_{k+1}) \right], \tag{6.10}$$

where K_k is the normalising constant given by (3.28). We note that the right-hand side of (6.10) is identical to the right-hand side of (3.27), and hence $\{\tilde{X}_{k+1}^* | \mathbf{T}_{2:k+1} = \mathbf{t}_{2:k+1}\} \stackrel{D}{=} \{\hat{X}_{k+1}^* | \mathbf{T}_{2:k+1} = \mathbf{t}_{2:k+1}\}$.

We complete the construction of $\hat{\mathbf{X}}_{k+1}$ by deriving $\hat{\zeta}^{k+1}$. For $j = 1, 2, \dots, J$, we set

$$\hat{\zeta}_j^{k+1} = \frac{\mathbb{E}[\tilde{X}_{k+1}^j | \mathbf{T}_{2:k+1} = \mathbf{t}_{2:k+1}]}{\mathbb{E}[\tilde{X}_{k+1}^* | \mathbf{T}_{2:k+1} = \mathbf{t}_{2:k+1}]} \tag{6.11}$$

and show that $\hat{\zeta}_j^{k+1}$ satisfies (3.29).

Suppose that $\hat{R}_k = l$, i.e. there are l family groups immediately after the k th detected death. We have that $l - 1$ of these family groups will not produce the $(k + 1)$ th detected death. For these family groups, by Lemma 7, the distribution of the number of individuals of each type alive immediately after the $(k + 1)$ th detected death at time s_{k+1} is given by $\mathbf{W}(s_{k+1}; -\boldsymbol{\psi}(s_k, t_{k+1}))$. By differentiating (3.4) with respect to θ_j and setting $\boldsymbol{\theta} = \mathbf{1}$, we have, for $\mathbf{a} \in \mathbb{R}^J$, that

$$\begin{aligned} \mathbb{E}[W_j(s_{k+1}; \mathbf{a})] &= \frac{\partial}{\partial \theta_j} \left[\frac{1 + \sum_{i=1}^J \theta_i \tilde{a}_i}{1 + \sum_{i=1}^J \tilde{a}_i} \times \frac{\pi_{s_{k+1}}}{1 - (1 - \pi_{s_{k+1}}) \sum_{l=1}^J \theta_l \eta_l^{s_{k+1}}} \right] \Bigg|_{\boldsymbol{\theta}=\mathbf{1}} \\ &= \left[\frac{\tilde{a}_j \pi_{s_{k+1}}}{(1 + \sum_{i=1}^J \tilde{a}_i)(1 - (1 - \pi_{s_{k+1}}) \sum_{l=1}^J \theta_l \eta_l^{s_{k+1}})} \right. \\ &\quad \left. + \frac{1 + \sum_{i=1}^J \theta_i \tilde{a}_i}{1 + \sum_{i=1}^J \tilde{a}_i} \times \frac{\eta_j^{s_{k+1}} \pi_{s_{k+1}} (-1)^2 (1 - \pi_{s_{k+1}})}{[1 - (1 - \pi_{s_{k+1}}) \sum_{l=1}^J \theta_l \eta_l^{s_{k+1}}]^2} \right] \Bigg|_{\boldsymbol{\theta}=\mathbf{1}} \\ &= \eta_j^{s_{k+1}} \frac{1 - \pi_{s_{k+1}}}{\pi_{s_{k+1}}} + \frac{\tilde{a}_j}{1 + \sum_{i=1}^J \tilde{a}_i}. \end{aligned} \tag{6.12}$$

Hence, it follows that

$$\mathbb{E}[W_j(s_{k+1}; -\boldsymbol{\psi}(s_k, t_{k+1}))] = \eta_j^{s_{k+1}} \frac{1 - \pi_{s_{k+1}}}{\pi_{s_{k+1}}} - \frac{\psi_j(s_k; t_{k+1})}{1 - \psi(s_k; t_{k+1})}. \tag{6.13}$$

Similarly, we can derive the expected number of individuals of type j immediately after the $(k + 1)$ th detected death at time s_{k+1} in the family group which experienced the death, using

(5.29) in Lemma 8 with $\mathbf{c} = \mathbf{C}_{s_k; t_{k+1}}$ defined in (3.19). By (6.12), it follows that

$$\mathbb{E}[W_j(s_{k+1}; \mathbf{0})] + \mathbb{E}[W_j(s_{k+1}; \mathbf{C}_{s_k; t_{k+1}})] = 2\eta_j^{s_{k+1}} \frac{1 - \pi_{s_{k+1}}}{\pi_{s_{k+1}}} + \frac{C_{s_k, t_{k+1}}^j}{1 + C_{s_k, t_{k+1}}^*}. \tag{6.14}$$

It follows from (6.13), (6.14), and $v_k = \mathbb{E}[\hat{R}_k | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}]$ that

$$\mathbb{E}[\tilde{X}_{k+1}^j | \mathbf{T}_{2:k+1} = \mathbf{t}_{2:k+1}] = (v_k + 1)\eta_j^{s_{k+1}} \frac{1 - \pi_{s_{k+1}}}{\pi_{s_{k+1}}} - (v_k - 1) \frac{\psi_j(s_k; t_{k+1})}{1 - \psi(s_k; t_{k+1})} + \frac{C_{s_k; t_{k+1}}^j}{1 + C_{s_k; t_{k+1}}^*}. \tag{6.15}$$

By summing (6.15) over j , we have that

$$\mathbb{E}[\tilde{X}_{k+1}^* | \mathbf{T}_{2:k+1} = \mathbf{t}_{2:k+1}] = (v_k + 1) \frac{1 - \pi_{s_{k+1}}}{\pi_{s_{k+1}}} - (v_k - 1) \frac{\psi(s_k; t_{k+1})}{1 - \psi(s_k; t_{k+1})} + \frac{C_{s_k; t_{k+1}}^*}{1 + C_{s_k; t_{k+1}}^*}, \tag{6.16}$$

and (3.29) follows immediately from inserting (6.15) and (6.16) into (6.11).

7. Numerical results

In this section we demonstrate the usefulness of the approximate distribution given in Section 3.5.2 using an example parameter set with time-changing parameters, birth rate, and probability of detecting a death. We simulated branching processes with $L \sim \text{Gamma}(3, 3)$ up until the 500th detected death. For $k = 1, 2, \dots, 500$, let α_k and ε_k denote the birth rate and the probability of detecting the death of a type-3 individual, respectively, between the $(k - 1)$ th and k th detected death. (Since $\gamma_1 = \gamma_2 = 0$ it suffices to state the probability of detecting a death of type-3 individuals.) The birth rate and detection probability change after every 100 detected deaths, with $(\alpha_k, \varepsilon_k) = (2.5, 0.3)$, $(\alpha_{100+k}, \varepsilon_{100+k}) = (1.5, 0.4)$, $(\alpha_{200+k}, \varepsilon_{200+k}) = (1.0, 0.5)$, $(\alpha_{300+k}, \varepsilon_{300+k}) = (0.6, 0.6)$, and $(\alpha_{400+k}, \varepsilon_{400+k}) = (0.4, 0.7)$ ($k = 1, 2, \dots, 100$). Thus the birth rate is decreasing over time, with the branching process going from being supercritical to subcritical with the detection probability increasing. (Note that since $\mathbb{E}[L] = 1$, the basic reproduction number is equal to the birth rate.) In Figure 3, we plot the number of individuals alive, of each type and the total number, immediately after a detected death against the number of detected deaths from a single realisation of the branching process, along with the median ($\hat{X}_k^j, j = 1, 2, 3; \hat{X}_k^*$) of the approximate distribution derived in Section 3.5.2. We also include the 5% and 95% quantiles of the approximate distribution, denoted by l_k and u_k , with $[l_k, u_k]$ shaded for $k = 1, 2, \dots, 500$. Throughout, we observe good agreement between the median of the approximate distribution and the actual number of individuals alive of each type.

We now turn to assessing the performance of the approximate distribution based on 100 realisations of the branching process described above. For $i = 1, 2, \dots, 100$, let $\mathbf{t}_{2:500}^i$ denote the inter-arrival times of the detected deaths in simulation i , and let $x_{i,k,j}$ ($k = 1, 2, \dots, 500; j = 1, 2, 3, *$) denote the number of individuals of type j alive (with $x_{i,k,*} = \sum_{j=1}^3 x_{i,k,j}$) immediately after the k th detected death in the i th simulation. (Note that we include only simulations which reach at least 500 detected deaths.) For $i = 1, 2, \dots, 100, k = 1, 2, \dots, 500$, and $j = 1, 2, 3, *$, let

$$u_{i,k,j} = 0.5[\mathbb{P}(\hat{X}_k^j < x_{i,k,j} | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}^i) + \mathbb{P}(\hat{X}_k^j \leq x_{i,k,j} | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}^i)], \tag{7.1}$$

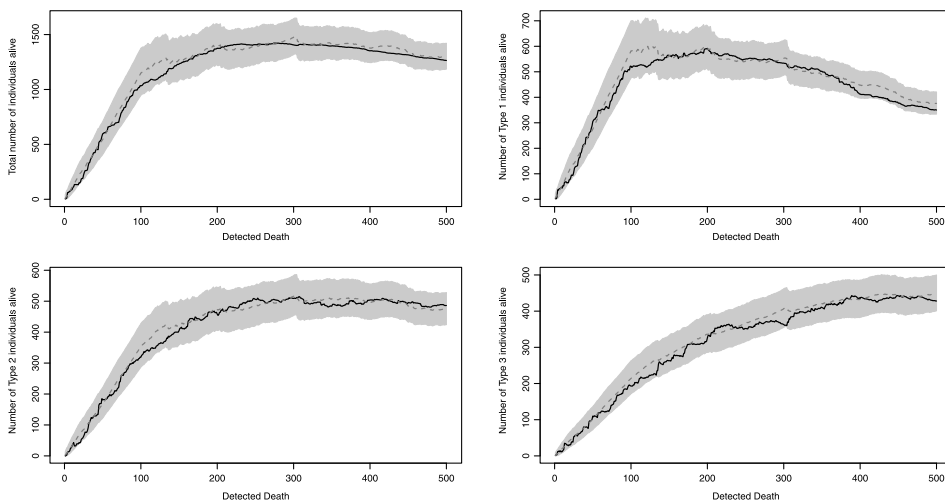


FIGURE 3. Number of individuals alive (solid line) and median of approximate distribution $\hat{X}_k^z | \mathbf{t}_{2:k}$ ($z = *, 1, 2, 3$) (dashed line) up to the 500th detected death with $L \sim \text{Gamma}(3, 3)$, $(\alpha_k, \varepsilon_k) = (2.5, 0.3)$, $(\alpha_{100+k}, \varepsilon_{100+k}) = (1.5, 0.4)$, $(\alpha_{200+k}, \varepsilon_{200+k}) = (1.0, 0.5)$, $(\alpha_{300+k}, \varepsilon_{300+k}) = (0.6, 0.6)$, and $(\alpha_{400+k}, \varepsilon_{400+k}) = (0.4, 0.7)$ ($k = 1, 2, \dots, 100$). The shaded area represents the probability mass between the 5% and 95% quantiles of $\hat{X}_k^z | \mathbf{t}_{2:k}$. Top left: total size of the population. Top right: number of Type 1 individuals. Bottom left: number of Type 2 individuals. Bottom right: number of Type 3 individuals.

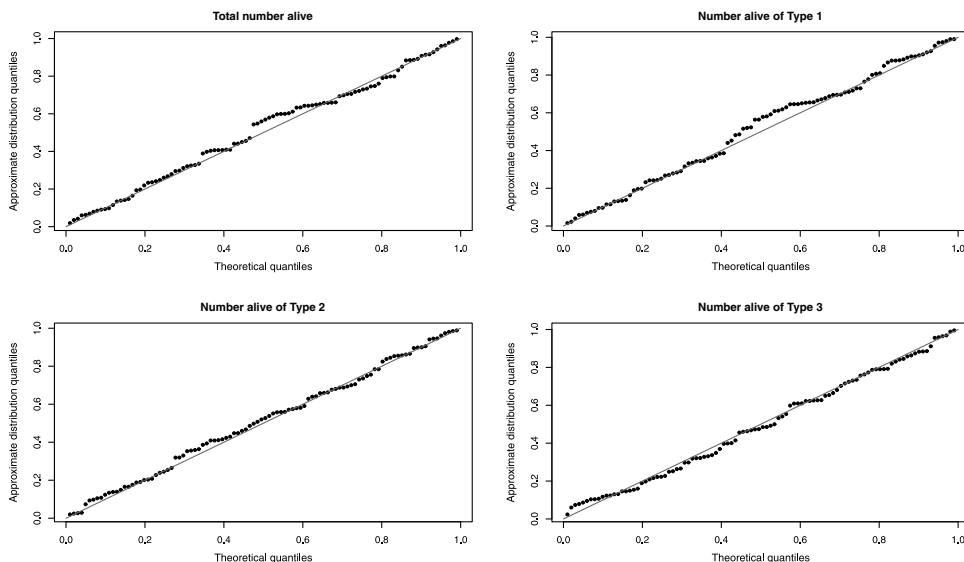


FIGURE 4. P–P plots based on 100 simulations of the ordered quantiles of $\tilde{\mathbf{u}}_{\cdot,500,j}$ ($j = *, 1, 2, 3$), where $L \sim \text{Gamma}(3, 3)$, $(\alpha_k, \varepsilon_k) = (2.5, 0.3)$, $(\alpha_{100+k}, \varepsilon_{100+k}) = (1.5, 0.4)$, $(\alpha_{200+k}, \varepsilon_{200+k}) = (1.0, 0.5)$, $(\alpha_{300+k}, \varepsilon_{300+k}) = (0.6, 0.6)$, and $(\alpha_{400+k}, \varepsilon_{400+k}) = (0.4, 0.7)$ ($k = 1, 2, \dots, 100$). Top left: total number of individuals alive. Top right: number of Type 1 individuals alive. Bottom left: number of Type 2 individuals alive. Bottom right: number of Type 3 individuals alive.

the midpoint of the left- and right-hand limits of the cumulative distribution function of $\hat{X}_k^j | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}^i$ evaluated at $x_{i,k,j}$. Let $\mathbf{u}_{\cdot,k,j} = (u_{1,500,j}, u_{2,k,j}, \dots, u_{100,k,j})$, and let $\tilde{u}_{i,k,j}$ denote the i th-smallest element of $\mathbf{u}_{\cdot,k,j}$ with $\tilde{\mathbf{u}}_{\cdot,k,j} = (\tilde{u}_{1,500,j}, \tilde{u}_{2,k,j}, \dots, \tilde{u}_{100,k,j})$. Under Section 3.5.2, the order statistics $\tilde{\mathbf{u}}_{\cdot,k,j}$ are distributed according to the order statistics of $U \sim U(0, 1)$. In Figure 4, we present P–P plots for $\tilde{\mathbf{u}}_{\cdot,500,j}$ ($j = 1, 2, 3, *$) against the quantiles of $U \sim U(0, 1)$. The P–P plots demonstrate excellent performance of the approximate distribution for number of each type alive and the total number alive after the 500th detected death, with similar performance observed for other values of k .

In the supplementary material two more examples are provided, including an epidemic example where the birth rate in the branching process is adjusted to take account of the expected number of non-susceptible individuals in the population in a similar manner to [2, Section 7]. The results of the supplementary material suggest that the approximate distribution tends to become worse as the number of phases in the Erlang distribution increases and the more dramatic the changes in parameters, especially the detection probability.

Appendix A. Proof of Lemma 2.

Note from (3.7) that for $i = 1, 2, \dots, J$, $b_i (= b_i(t, \tau; \mathbf{a})) > -(1 - \pi_{t+\tau})\eta_i^{t+\tau}$ is equivalent to

$$\begin{aligned}
 & -\psi_i(t; \tau) + \sum_{h=1}^J a_h [\zeta(t; \tau)p_{hi}(t, 0, \tau) - \psi_i(t; \tau)q_h(t, 0, \tau)] \\
 & > -(1 - \pi_{t+\tau})\eta_i^{t+\tau} \left[1 + \sum_{h=1}^J a_h q_h(t, 0, \tau) \right]. \tag{A.1}
 \end{aligned}$$

By (3.3), $\phi_i(t; \tau) = (1 - \pi_{t+\tau})\eta_i^{t+\tau} - \psi_i(t; \tau)$, so we have that (A.1) is equivalent to

$$\phi_i(t; \tau) > \sum_{h=1}^J a_h [-\zeta(t; \tau)p_{hi}(t, 0, \tau) - \phi_i(t; \tau)q_h(t, 0, \tau)]. \tag{A.2}$$

Since $-\zeta(t; \tau)p_{hi}(t, 0, \tau) - \phi_i(t; \tau)q_h(t, 0, \tau) < 0$ ($h = 1, 2, \dots, J$) and $a_h > -(1 - \pi_t)\eta_h^t$, we have that

$$\begin{aligned}
 & \sum_{h=1}^J a_h [-\zeta(t; \tau)p_{hi}(t, 0, \tau) - \phi_i(t; \tau)q_h(t, 0, \tau)] \\
 & < \sum_{h=1}^J [-(1 - \pi_t)\eta_h^t] [-\zeta(t; \tau)p_{hi}(t, 0, \tau) - \phi_i(t; \tau)q_h(t, 0, \tau)] \\
 & = (1 - \pi_t)\zeta(t; \tau) \sum_{h=1}^J \eta_h^t p_{hi}(t, 0, \tau) + \phi_i(t; \tau)(1 - \pi_t) \sum_{h=1}^J \eta_h^t q_h(t, 0, \tau). \tag{A.3}
 \end{aligned}$$

Using (3.2) and (3.3), it is straightforward to show that the right-hand side of (A.3) is equal to $\phi_i(t; \tau)$. Therefore (A.2), and consequently $b_i > -(1 - \pi_{t+\tau})\eta_i^{t+\tau}$ ($i = 1, 2, \dots, J$), both hold.

Given that $b_j > -(1 - \pi_{t+\tau})\eta_j^{t+\tau}$ ($j = 1, 2, \dots, J$), the denominator in (3.8) is positive. Using (3.8), a little algebra shows that for $i = 1, 2, \dots, J$, the inequality $c_i > -(1 - \pi_{t+\tau})\eta_i^{t+\tau}$ is equivalent to

$$(b_i + (1 - \pi_{i+\tau})\eta_i^{t+\tau}) \sum_{j=1}^J d_{t+\tau,j} \gamma_j \eta_j^{t+\tau} > 0,$$

which clearly holds.

Funding information

There are no funding bodies to thank in relation to the creation of this article.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

Supplementary material

The supplementary material for this article can be found at <https://doi.org/10.1017/apr.2024.65>.

References

- [1] ASMUSSEN, S. (2003). *Applied Probability and Queues*, 2nd edn. Springer, New York.
- [2] BALL, F. AND NEAL, P. (2023). The size of a Markovian SIR epidemic given only removal data. *Adv. Appl. Prob.* **55**, 895–926.
- [3] BLACK, A. J. AND ROSS, J. V. (2015). Computation of epidemic final size distributions. *J. Theoret. Biol.* **367**, 159–165.
- [4] GEIGER, J. AND KERSTING, G. (1997). Depth-first search of random trees, and Poisson point processes. In *Classical and Modern Branching Processes*, Springer, New York, pp. 111–126.
- [5] HOUSE, T., ROSS, J. V. AND SIRL, D. (2013). How big is an outbreak likely to be? Methods for epidemic final-size calculation. *Proc. R. Soc. A* **469**, article no. 20120436, 22 pp.
- [6] LAMBERT, A. AND TRAPMAN, P. (2013). Splitting trees stopped when the first clock rings and Vervaat’s transformation. *J. Appl. Prob.* **50**, 208–227.
- [7] LEFÈVRE, C. AND PICARD, P. (2017). On the outcome of epidemics with detections. *J. Appl. Prob.* **54**, 890–904.
- [8] LEFÈVRE, C., PICARD, P. AND UTEV, S. (2020). On branching models with alarm triggerings. *J. Appl. Prob.* **57**, 734–759.
- [9] TRAPMAN, P. AND BOOTSMA, M. C. J. (2009). A useful relationship between epidemiology and queueing theory: the distribution of the number of infectives at the moment of the first detection. *Math Biosci.* **219**, 15–22.
- [10] WALKER, J. N., BLACK, A. J. AND ROSS, J. V. (2019). Bayesian model discrimination for partially-observed epidemic models. *Math Biosci.* **317**, article no. 108266, 13 pp.
- [11] WHITTLE, P. (1955). The outcome of a stochastic epidemic—a note on Bailey’s paper. *Biometrika* **42**, 116–122.