Adv. Appl. Probab. **55,** 895–926 (2023)
*doi:*10.1017/apr.2022.58

THE SIZE OF A MARKOVIAN SIR EPIDEMIC GIVEN ONLY REMOVAL DATA

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

Abstract

During an epidemic outbreak, typically only partial information about the outbreak is known. A common scenario is that the infection times of individuals are unknown, but individuals, on displaying symptoms, are identified as infectious and removed from the population. We study the distribution of the number of infectives given only the times of removals in a Markovian susceptible–infectious–removed (SIR) epidemic. Primary interest is in the initial stages of the epidemic process, where a branching (birth–death) process approximation is applicable. We show that the number of individuals alive in a time-inhomogeneous birth–death process at time $t \geq 0$, given only death times up to and including time t, is a mixture of negative binomial distributions, with the number of mixing components depending on the total number of deaths, and the mixing weights depending upon the inter-arrival times of the deaths. We further consider the extension to the case where some deaths are unobserved. We also discuss the application of the results to control measures and statistical inference.

Keywords: Branching processes; time-inhomogeneous birth–death process; negative binomial distribution

2020 Mathematics Subject Classification: Primary 60J80
Secondary 92D30

1. Introduction

A key public health consideration throughout the COVID-19 pandemic has been when and how to act to control the disease. In infectious disease epidemiology there is a balance to be struck between rapid introduction of control measures to limit the size of an epidemic outbreak and the costs (for example, economic, social, and mental health costs) associated with control measures. Given that the probability of a small epidemic outbreak can be close to 1 for a supercritical epidemic (basic reproduction number $R_0 > 1$), it is important to utilise information from the epidemic outbreak to decide whether or not intervention will be cost-effective at any given point in time.

The control of an epidemic outbreak, whether it is at a national or local level, needs to take place early in the outbreak, before the disease has taken significant hold within the population. During the early stages of an epidemic, a branching process approximation can be utilised (see [2], [3], and [19]), with the probability of a small epidemic equated with the extinction

Received 12 July 2021; revision received 26 August 2022.

[*] 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

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

probability of the approximating branching process. Therefore, a significant focus of this paper is on modelling the approximating branching process.

Branching process approximations of epidemic processes are built upon coupling infective individuals to individuals in a branching process, with infectious contacts and removals of infectives respectively corresponding to births and deaths in the branching process. At any point in time, given the current state of the population, we can then compute properties of the branching process (epidemic), such as the probability of extinction. However, in reality we rarely know the current state of the population and instead have a partial observation of the epidemic process; see, for example, [15] and [18]. Typically, we do not know when an individual becomes infected, but we can have information on when they show symptoms. Specifically, we consider susceptible–infectious–removed (SIR) epidemic models where all individuals, except an initial infective, start in the susceptible state. A susceptible individual on receiving an infectious contact becomes infected, and immediately infectious, and is able to infect other individuals until their removal (for example, through recovery, quarantining, or death). Removed individuals play no further role in the epidemic. Moreover, we assume that individuals are potentially detected on removal but nothing is known about infectious individuals prior to removal. This corresponds to observing only some of the deaths of individuals in the approximating branching process.

This paper's focus is the distribution of the number of infectives in an epidemic (the number of individuals alive in a branching process) at any given point in time, $t \geq 0$, based upon only the detected removals (deaths) up to and including time t, with the first detected removal (death) being observed at time $t = 0$. We focus primarily on the case where all removals are detected. We consider a time-inhomogeneous Markovian SIR epidemic model and its approximating branching (birth–death) process. For time-inhomogeneous epidemics, the approximating branching process has a varying environment, and all individuals alive at a given point in time have the same birth rate. The environment can depend upon the number of removals/deaths observed (cf. [14]). We show that the distribution of the number of individuals alive in the approximating branching process at time $t \geq 0$ can be expressed as a mixture of $(k_t + 1)$ negative binomial random variables, where k_t denotes the total number of observed deaths/removals up to and including time t. Given that we allow for time-inhomogeneity in the model, we are able to derive the distribution of the number of infectives in the presence of control measures or seasonal effects which change the infection and/or the removal rates. In particular, given only removal data, we can easily obtain the probability that the epidemic is over, and hence provide a guide for when to lift control measures.

The papers [9] and [18] consider the distribution of the number of individuals alive in a branching process at the first detection, where each individual is detected an exponentially distributed time after their birth (infection) assuming that they are still alive (infectious) at their detection time. Individuals are assumed to have independent and identically distributed lifetime distributions and, whilst alive, to reproduce at the points of a homogeneous Poisson point process with rate α, say. These papers show that the population size at the first detection follows a geometric distribution with support $\mathbb{N}$, and [9] studies further properties of the population at the first detection, such as the ages of individuals and the residual lifetimes. A key tool in [9] is to obtain the (shifted geometric) distribution of the population t units after the birth of the initial individual, conditional upon there being no detections up to time t. We take the same approach in this paper to obtain, in Section 4, Lemma 4.1, a geometric distribution with support $\mathbb{N}$ for the size of the population t units after the birth of the initial individual, conditional upon there being no deaths up to time t. The Markovian models considered in this paper allow for

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

explicit expressions for the parameters of the geometric distribution, in contrast to [9], where the key quantities are expressed in terms of Laplace transforms. Moreover, we allow for time-inhomogeneity in the model in both the birth and death rates. In [11], a birth–death process with detections is considered, and this corresponds to the model of [9] with an exponential lifetime distribution. In [11], the distribution of the number of individuals alive at the *k*th detection is considered, but times between detections are not considered. Finally, in [10] the general stochastic epidemic model (i.e. the Markovian SIR epidemic model with exponentially distributed infectious periods) is considered, with detections possibly occurring at the removal of individuals; the authors investigate the distribution of the number of infectives at the *k*th detection. Both [10] and [11] allow for the parameters of the models to change at the detection times. Since we consider the full time dynamics of the branching (epidemic) process, we can allow for more general temporal behaviour of the parameters, but an important special case is where the parameters are piecewise constant between detection times.

The paper is structured as follows. In Section 2, we introduce the time-inhomogeneous Markovian SIR epidemic model and its branching (birth–death) process approximation. In Section 3, we present the main result, Theorem 3.1, which shows that the distribution of the number of individuals alive in the birth–death process, given only death times, can be expressed as a mixture of negative binomial distributions. This is extended in Theorem 3.2 to the case where only a subset of the death times are detected. The probability that the birth–death process will go extinct and the likelihood of observing a given set of death times, given in Corollary 3.1, follow immediately from Theorem 3.1. In Theorem 3.3 we derive the exact distribution of the number of infectives in the SIR epidemic given only removal times, which is useful for assessing the accuracy of the birth–death process approximations. In Sections 4–6 we provide the proofs of Theorems 3.1–3.3 respectively. In Section 7 we present numerical results using simulations to illustrate the estimation of the number of individuals over time, the implementation of control measures, and comparisons between the number of infectives in the epidemic and the number of individuals alive in the approximating birth–death process. In particular, a time-inhomogeneous birth–death process is shown to give a good approximation of the epidemic process over its full trajectory. Finally, in Section 8 we present concluding remarks regarding how the findings of this paper can be utilised from statistical inference and public health perspectives.

2. Epidemic and branching process models

In this section, we introduce formally the time-inhomogeneous Markovian SIR epidemic model and its approximating branching (birth–death) process.

The time-inhomogeneous Markovian SIR epidemic model is defined as follows. There is assumed to be a closed population of size N with the epidemic initiated by a single infective in an otherwise susceptible population. Whilst infectious, individuals make infectious contacts at the points of a time-inhomogeneous Poisson point process with rate β_t at time t. The individual contacted by an infectious contact is chosen uniformly at random from the whole population, independently of any other infectious contacts. If a susceptible individual is contacted by an infectious contact, the individual becomes infected and is immediately able to transmit the disease to other individuals. Infectious contacts with non-susceptible individuals have no effect on the recipient. Infective individuals recover from the disease and are removed from the epidemic process, playing no further role in the epidemic, at rate γ_t at time t. Let $S(t)$, $I(t)$, and $R(t)$ $(= N - S(t) - I(t))$ denote the total numbers of susceptibles, infectives, and

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

removed individuals, respectively, in the population at time t. Then it suffices to keep track of $\{(S(t), I(t))\}$, and the model satisfies, for $h \geq 0$,

$$\mathbb{P}\left((S(t+h), I(t+h)) = (x, y)\,|(S(t), I(t)) = (s, i)\right)$$
$$= \begin{cases} \dfrac{si}{N}\beta_t h + o(h), & (x, y) = (s-1, i+1) \quad \text{(infection)}, \\ i\gamma_t h + o(h), & (x, y) = (s, i-1) \quad \text{(removal)}, \\ 1 - i\left\{\beta_t \frac{s}{N} + \gamma_t\right\} h + o(h), & (x, y) = (s, i) \quad \text{(no event)}, \end{cases} \tag{2.1}$$

with all other events occurring with probability $o(h)$. We place no restriction on the form of β_t and γ_t, and these could be continuously varying and/or contain discontinuities. The former can be used to model seasonal variability, whilst the latter allows for the implementation and removal of control measures.

In the early stages of the epidemic, whilst the total number of individuals ever infected is small, the epidemic process can be coupled to a branching process. Specifically, each infective in the epidemic has a corresponding individual in the coupled branching process whose lifetime is identical to the infectious period of the infective, so the individual in the branching process dies when the infective is removed from the population. The individual in the branching process gives birth to new individuals whenever the corresponding infective makes infectious contacts in the epidemic process. Thus the two processes can be coupled (see, for example, [3]) until the first time that there is an infectious contact with a previously infected individual. For large N this does not occur until $O(\sqrt{N})$ of the population have been infected; see [3]. Specifically, the time-inhomogeneous Markovian SIR stochastic epidemic model can be approximated by a time-inhomogeneous linear birth–death process [8]. Let $B(t)$ denote the total number of individuals alive in the time-inhomogeneous linear birth–death process at time t. Then for $h \geq 0$, $B(t)$ satisfies

$$\mathbb{P}(B(t+h) = j|B(t) = i) = \begin{cases} i\beta_t h + o(h), & j = i+1 \quad \text{(birth)}, \\ i\gamma_t h + o(h), & j = i-1 \quad \text{(death)}, \\ 1 - i\{\beta_t + \gamma_t\}h + o(h), & j = i \quad \text{(no event)}, \end{cases} \tag{2.2}$$

with all other events occurring with probability $o(h)$.

Suppose that at time t we can estimate the proportion, s_t, of the population who are susceptible in the epidemic process. We can then replace β_t by $\beta_t s_t$ in (2.2) to obtain a birth–death approximation of the epidemic which is valid after the epidemic has taken off. In Section 7 we outline a simple estimate of s_t using only the inter-removal times, up to and including time t, to create a time-inhomogeneous birth–death process approximation to the epidemic process. In Figure 4 we show the excellent agreement between the estimates obtained from these two processes of the mean number of infectives at time t, given only the inter-removal times up to and including time t.

3. Main results

3.1. Overview

In this section, we present the main results along with the necessary notation. Throughout, we set time so that the first death (removal) of the birth–death (epidemic) process occurs at time 0. The time-varying parameters in (2.1) and (2.2) can be adjusted accordingly.

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

In Section 3.2 we present Theorem 3.1, which states that the distribution of the number of individuals alive in a time-inhomogeneous linear birth–death process given only times of deaths follows a mixture of negative binomial distributions. We present an outline of the arguments used to prove Theorem 3.1, with the details of the proof provided in Section 4. We note that Theorem 3.1 allows computation of the probability that the birth–death process will go extinct. In Corollary 3.1 we derive the likelihood of observing the given death times, which can be used in statistical inference for inferring the parameters of the birth–death process given data.

In Section 3.3, we consider the special case of the time-homogeneous linear birth–death process. Lemma 3.1 specialises Theorem 3.1 to the time-homogeneous birth–death process and Lemma 3.2 gives the distribution of the number of individuals alive in the birth–death process immediately after the kth death, conditioning only upon there having been k deaths. The proof of Lemma 3.2 is presented immediately after the lemma and is informative about why, in general, the distribution of the number of individuals alive follows a mixture of negative binomial distributions.

In Section 3.4, we consider the important extension of assuming that not all deaths are detected. In Theorem 3.2, we show that the number of individuals alive given only the times of detected deaths follows a mixture of negative binomial distributions. The overall structure of the proof is similar to that of Theorem 3.1, but there are key differences, as Lemma 4.1, which is a major component in the proof of Theorem 3.1, does not carry over to partial detection of deaths. Therefore we need to resort to an alternative argument based on probability generating functions, whose details are provided in Section 5. In Section 3.5, Theorem 3.3, we derive the number of infectives in the general stochastic epidemic model at time t given only the times of removals. This enables us, in Section 7, to compare the approximate distribution given by the birth–death process for the number of infectives (individuals alive) with the exact distribution given by Theorem 3.3. For succinctness of presentation in Sections 3.4 and 3.5, we assume that the birth (infection) and death (removal) parameters are piecewise constant between observed deaths (removals), although more general piecewise-constant parameter scenarios could easily be considered.

3.2. Time-inhomogeneous linear birth–death process

For $k = 1, 2, \ldots$, let X_k denote the number of individuals alive immediately after the kth death, and for $t \geq 0$, let $Y(t)$ denote the number of individuals alive at time t, with $Y(0) \equiv X_1$. For $k = 2, 3, \ldots$, let T_k denote the inter-arrival time from the $(k-1)$th to the kth death, with the convention that $T_k = \infty$ if fewer than k deaths occur, and let $S_k = \sum_{j=2}^{k} T_j$ denote the time of the kth death, with $S_1 = 0$. Let $\mathbf{T}_{2:k} = (T_2, T_3, \ldots, T_k)$. For $k = 2, 3, \ldots$, let t_k denote the observed inter-arrival time from the $(k-1)$th to the kth death and $s_k = \sum_{j=2}^{k} t_j$, with $s_1 = 0$. For $k = 2, 3, \ldots$, let $\mathbf{t}_{2:k} = (t_2, t_3, \ldots, t_k)$; we consider the distribution of $X_k | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}$. For $t \geq 0$, let k_t denote the number of deaths, and let s_{k_t} denote the time of the last death, up to and including time t. Then for $t \geq 0$, we consider the distribution of $Y(t) | \mathbf{T}_{2:k_t} = \mathbf{t}_{2:k_t}$. Note that $Y(s_k) \equiv X_k$, which is used repeatedly.

Before stating Theorem 3.1 we introduce some notation for probability distributions. Throughout this paper, $G \sim \mathrm{Geom}(p)$ denotes a geometric random variable with parameter p and probability mass function

$$\mathbb{P}(G = x) = (1-p)^x p \qquad (x = 0, 1, \ldots). \tag{3.1}$$

Also, for $r = 0, 1, \dots, V = \sum_{i=1}^{r} G_i \sim \text{NegBin}(r, p)$ denotes a negative binomial random variable, where $G_1, G_2, \dots, G_r$ are independent and identically distributed according to Geom(p), and for $r = 0$, $\mathbb{P}(V = 0) = 1$.

Theorem 3.1 starts with the assumption that there exists $0 < \pi_0 < 1$ such that $X_1 \equiv Y(0) \sim$ Geom(π_0). In the important special case where the birth and death rates are constant prior to the first death—that is, for $t \leq 0$, $\beta_t = \alpha_1$, and $\gamma_t = \mu_1$—it is trivial (cf. the start of the proof of Lemma 3.2) to show that the number of individuals alive immediately after the first death is Geom (π_0) with $\pi_0 = \mu_1/(\alpha_1 + \mu_1)$. (See also [9] and [18].) We discuss how the arguments can be modified to other scenarios in Section 4, after the proof of Theorem 3.1.

Theorem 3.1. *Suppose that* $X_1 \equiv Y(0) \sim \text{Geom}(\pi_0)$ *for some* $0 < \pi_0 < 1$.

For $t \geq 0$, *let* π_t *solve the ordinary differential equation (ODE)*

$$\pi_t' = \gamma_t - (\beta_t + \gamma_t)\pi_t, \tag{3.2}$$

with initial condition $0 < \pi_0 < 1$.

For $k = 2, 3, \dots$, *the distribution* X_k *of the number of individuals alive immediately following the kth death satisfies*

$$\{X_k | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}\} \sim \text{NegBin}(R_k, \pi_{s_k}), \tag{3.3}$$

where R_k *is a random variable with support on* $\{2, 3, \dots, k\}$, *and* $\mathbb{P}(R_k = j) = B_{k,j}$ *with* $\mathbf{B}_k = (B_{k,2}, B_{k,3}, \dots, B_{k,k})$ *given by* (3.11) *below.*

For $t \geq 0$, *the distribution Y(t) of the number of individuals alive at time t satisfies* $Y(t) \equiv X_{k_t}$ *if* $t = s_{k_t}$, *and if* $t > s_{k_t}$,

$$\{Y(t) | \mathbf{T}_{2:k_t} = \mathbf{t}_{2:k_t}\} \sim \text{NegBin}(Z(t), \pi_t), \tag{3.4}$$

where Z(t) is a random variable with support on $\{0, 1, \dots, k_t\}$, *and* $\mathbb{P}(Z(t) = j) = D_{t,j}$ *with* $\mathbf{D}_t = (D_{t,0}, D_{t,1}, \dots, D_{t,k_t})$ *given by* (3.12) *and* (3.13) *below. If* $k_t = 1$, *then* $T_2 > t$ *replaces* $\mathbf{T}_{2:k_t} = \mathbf{t}_{2:k_t}$ *in* (3.4).

We begin by defining the notation required to obtain the distributions of R_k and $Z(t)$ introduced in Theorem 3.1. The process $\{Z(t) : t \geq 0\}$ is clarified in the proof of Corollary 4.1. Briefly, it is an integer-valued stochastic process that decreases in steps of size 1 between deaths until either it reaches 0, in which case the number of individuals alive is 0 and no further deaths occur, or a death occurs, in which case Z(t) increases by 1. The process $\{Z(t) : t \geq 0\}$ is Markovian given also the death times prior to t. We also further explore π_t before presenting an outline of the main steps in the proof of Theorem 3.1, with the details provided in Section 4.

For $t \in \mathbb{R}$, let $p_t = \beta_t/\{\beta_t + \gamma_t\}$ (resp. $q_t = \gamma_t/\{\beta_t + \gamma_t\}$) denote the probability that an event occurring at time t is a birth (resp. death). Then for $t \geq 0$, we can rewrite the ODE (3.2) for π_t as

$$\pi_t' = (\beta_t + \gamma_t)[q_t - \pi_t], \tag{3.5}$$

with initial condition $0 < \pi_0 < 1$. This highlights that π_t is increasing (decreasing) if $q_t > \pi_t$ ($q_t < \pi_t$), with π_0 defined by the birth and death rates prior to the first death.

For t, $\tau \geq 0$, let

$$\phi(t; \tau) = \exp\left(- \int_t^{t+\tau} \{\beta_s + \gamma_s\}\, ds \right), \tag{3.6}$$

the probability that an individual alive at time t does not give birth or die in the interval $(t, t+\tau]$, and let

$$\psi(t;\tau)=\int_t^{t+\tau}\beta_u \exp\left(-\int_u^{t+\tau}\{\beta_s+\gamma_s\}\,ds\right)du, \tag{3.7}$$

the probability that an individual alive at time t has at least one offspring and their first offspring (looking back from time $t+\tau$) survives to time $t+\tau$. We can then rewrite (3.2) in integral form as

$$\begin{aligned}\pi_{t+\tau}&=\pi_t \exp\left(-\int_t^{t+\tau}\{\beta_s+\gamma_s\}\,ds\right)+\int_t^{t+\tau}\gamma_u \exp\left(-\int_u^{t+\tau}\{\beta_s+\gamma_s\}\,ds\right)du\\&=\pi_t\phi(t;\tau)+\{1-\phi(t;\tau)-\psi(t;\tau)\}.\end{aligned} \tag{3.8}$$

We can now define the matrices needed for the probability mass functions of R_k and $Z(t)$, which define the distributions of X_k and $Y(t)$ in Theorem 3.1. For $k=2, 3, \ldots$ and for $t, \tau\geq 0$, let $\mathbf{M}_k(t;\tau)$ be the $(k-1)\times k$ matrix with (i, j)th element

$$[\mathbf{M}_k(t;\tau)]_{i,j}=\begin{cases}(i+1)\binom{i}{j-1}\left\{\dfrac{(1-\pi_t)\pi_t}{(1-\pi_{t+\tau})\pi_{t+\tau}}\phi(t;\tau)\right\}^{j-1}\left\{\dfrac{\pi_t}{1-\pi_{t+\tau}}\psi(t;\tau)\right\}^{i+1-j} & \text{for } j\leq i+1,\\ 0 & \text{otherwise,}\end{cases} \tag{3.9}$$

and let $\mathbf{J}_k(t;\tau)$ be the $(k-1)\times(k+1)$ matrix with (i, j)th element

$$[\mathbf{J}_k(t;\tau)]_{i,j}=\begin{cases}\binom{i+1}{j-1}\left\{\dfrac{(1-\pi_t)\pi_t}{(1-\pi_{t+\tau})\pi_{t+\tau}}\phi(t;\tau)\right\}^{j-1}\left\{\dfrac{\pi_t}{1-\pi_{t+\tau}}\psi(t;\tau)\right\}^{i+2-j} & \text{for } j\leq i+2,\\ 0 & \text{otherwise.}\end{cases} \tag{3.10}$$

Let $\mathbf{B}_2=(1)$, and for $k=3, 4, \ldots$ let $\mathbf{B}_k=(B_{k,2}, B_{k,3}, \ldots, B_{k,k})$ be given by

$$\mathbf{B}_k=\left\{\left(\prod_{j=2}^{k-1}\mathbf{M}_j(s_j;t_{j+1})\right)\cdot \mathbf{1}_{k-1}^{\top}\right\}^{-1}\prod_{j=2}^{k-1}\mathbf{M}_j(s_j;t_{j+1}), \tag{3.11}$$

with $\mathbf{1}_{k-1}$ denoting a row vector of 1s of length $k-1$. Let $\mathbf{D}_t=(D_{t,0}, D_{t,1}, \ldots, D_{t,k_t})$ be given by, for $0\leq t<s_2$,

$$\mathbf{D}_t=\left(\frac{\pi_t\psi(0;t)}{\pi_t\psi(0;t)+(1-\pi_0)\phi(0;t)}, \frac{(1-\pi_0)\phi(0;t)}{\pi_t\psi(0;t)+(1-\pi_0)\phi(0;t)}\right), \tag{3.12}$$

and for $t\geq s_2$,

$$\begin{aligned}\mathbf{D}_t=&\left\{\left(\prod_{j=2}^{k_t-1}\mathbf{M}_j(s_j;t_{j+1})\right)\mathbf{J}_{k_t}(s_{k_t};t-s_{k_t})\cdot\mathbf{1}_{k_t+1}^{\top}\right\}^{-1}\\&\times\left(\prod_{j=2}^{k_t-1}\mathbf{M}_j(s_j;t_{j+1})\right)\mathbf{J}_{k_t}(s_{k_t};t-s_{k_t}),\end{aligned} \tag{3.13}$$

with the convention $\prod_{j=2}^{1}\mathbf{M}_j(s_j;t_{j+1})=\mathbf{I}_1$ (the 1×1 identity matrix).

To prove Theorem 3.1, we start by showing in Lemma 4.1 that the number of living individuals at time $t+\tau$ originating from a single individual at time t, conditional on no deaths in the interval $(t, t+\tau]$, is $1+\mathrm{Geom}(1-\psi(t;\tau))$. This is proved using a probabilistic construction of the process. We can use this to consider how the birth–death process evolves between the first and second death when we have Geom (π_0) rather than one initial individual. In Lemma 4.3, we show that for any given time t between the first and second deaths, the distribution of the number of individuals alive follows a mixture of a geometric distribution, $G_1(\pi_t) \sim \mathrm{Geom}(\pi_t)$, and a point mass at 0. This leads, through size-biased sampling in Lemma 4.4, to the number of individuals alive immediately after the second death (time $s_2 = t_2$) being distributed as a negative binomial distribution, $\sum_{i=1}^{2} G_i(\pi_{s_2}) = \mathrm{NegBin}(2, \pi_{s_2})$.

Finally, the dynamics of the process between the first and second death, along with the Markovian nature of the process, allow us to relatively straightforwardly write down the full dynamics. In particular, given that $\{X_k|\mathbf{T}_{2:k} = \mathbf{t}_{2:k}\} = \sum_{i=1}^{R_k} G_i(\pi_{s_k})$, we can consider the evolution of R_k independent processes each starting from $\mathrm{Geom}(\pi_{s_k})$ individuals. At time s_{k+1}, the $(k+1)$th death will occur, and the death will belong to one of the R_k processes starting with $\mathrm{Geom}(\pi_{s_k})$ individuals. The process with the death will have $\mathrm{NegBin}(2, \pi_{s_{k+1}})$ individuals alive at time s_{k+1} to contribute to X_{k+1}. The remaining $R_k - 1$ processes will have experienced no deaths in the interval $(s_k, s_{k+1}]$, and the number of individuals alive in each of these processes is distributed according to a mixture of $\mathrm{Geom}(\pi_{s_{k+1}})$ and a point mass at 0. This leads to the binomial terms in (3.9) and (3.10) for the number of processes which have $\mathrm{Geom}(\pi_{s_{k+1}})$ individuals alive; see Corollary 4.1. It is then straightforward, using Lemma 4.5, to iteratively compute the distribution of $\{X_k|\mathbf{T}_{2:k} = \mathbf{t}_{2:k}\}$ and complete the proof of Theorem 3.1.

We briefly discuss some interesting results which follow straightforwardly from Theorem 3.1.

Given that Theorem 3.1 provides the probability mass function of $Z(t)$, it is straightforward to compute moments of $Y(t)$, the number of individuals alive at time t. However, the forms of (3.4) and (3.13) do not permit any simplifications for rapid calculations of the moments of $Y(t)$.

From [8, Equation (18)] we have that at time t, the probability of non-extinction of the time-inhomogeneous birth–death process from a single individual is

$$\rho_t = \left\{1 + \int_t^\infty \gamma_s \exp\left(\int_t^s [\gamma_u - \beta_u]\, du\right) ds\right\}^{-1}. \tag{3.14}$$

Therefore, given $\mathbf{t}_{2:k_t}$, at time t we have that the probability that the birth–death process will go extinct is

$$\mathbb{E}\left[(1-\rho_t)^{Y(t)} \middle| \mathbf{T}_{2:k_t} = \mathbf{t}_{2:k_t}\right] = \mathbb{E}\left[\left(\frac{\pi_t}{1-(1-\pi_t)(1-\rho_t)}\right)^{Z(t)} \middle| \mathbf{T}_{2:k_t} = \mathbf{t}_{2:k_t}\right]. \tag{3.15}$$

It is straightforward to compute the right-hand side of (3.15) using the probability mass function of $Z(t)$, which is given by (3.13). Note that if, for all $u \geq t$, $\beta_u = \alpha$ and $\gamma_u = \mu$, then $\rho_t = 1 - \min\{1, \mu/\alpha\}$, and (3.15) gives the probability of extinction if the current birth and death rate were to persist.

We can compute the likelihood of observing inter-death times $\mathbf{t}_{2:k}$ using Corollary 3.1.

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

Corollary 3.1. *For $k = 2, 3, \ldots$ and $s_k < \infty$, we have that the probability density function of $\mathbf{T}_{2:k}$ is given by*

$$f_{\mathbf{T}_{2:k}}(\mathbf{t}_{2:k}) = \prod_{j=2}^{k} \frac{\pi_{s_{j-1}}(1-\pi_{s_{j-1}})\gamma_{s_j}\phi(s_{j-1}; t_j)}{\pi_{s_j}^2} \times \left\{ \left[\prod_{j=2}^{k-1} M_j(s_j; t_{j+1}) \right] \cdot \mathbf{1}_{k-1}^{\top} \right\}, \tag{3.16}$$

with, for $k = 2$, the vacuous latter term set equal to 1.

Corollary 3.1 is the key result in using the findings of this paper for statistical inference and inferring the parameters of the birth–death process given $\mathbf{t}_{2:k}$. The generality of the time-inhomogeneous birth–death process allows for different features such as the periodicity or control measures to be incorporated. We will explore likelihood inference for the birth–death process given only death times in future work.

3.3. Time-homogeneous linear birth–death process

An important special case of Theorem 3.1 is the time-homogeneous model, where, for all $t \in \mathbb{R}$, $\beta_t = \alpha$ and $\gamma_t = \mu$. In this case, for all $t \geq 0$,

$$\pi_t = q_t = \frac{\mu}{\alpha + \mu} = \hat{\pi}, \qquad \text{say,}$$

and for all $\tau \geq 0$, $\phi(t; \tau) = \hat{\phi}(\tau) = \exp(-[\alpha + \mu]\tau)$ and $\psi(t; \tau) = \hat{\psi}(\tau) = \hat{\pi}\{1 - \hat{\phi}(\tau)\}$. For $k = 2, 3, \ldots$, let $\hat{\mathbf{M}}_k(\tau)$ be the $(k-1) \times k$ matrix with (i, j)th element

$$\left[\hat{\mathbf{M}}_k(\tau)\right]_{i,j} = \begin{cases} (i+1)\binom{i}{j-1}\left\{\hat{\phi}(\tau)\right\}^{j-1}\left\{\hat{\pi}(1-\hat{\phi}(\tau))\right\}^{i+1-j} & \text{for } j \leq i+1, \\ 0 & \text{otherwise,} \end{cases} \tag{3.17}$$

and let $\hat{\mathbf{J}}_k(t; \tau)$ be the $(k-1) \times (k+1)$ matrix with (i, j)th element

$$\left[\hat{\mathbf{J}}_k(t; \tau)\right]_{i,j} = \begin{cases} \binom{i+1}{j-1}\left\{\hat{\phi}(\tau)\right\}^{j-1}\left\{\hat{\pi}(1-\hat{\phi}(\tau))\right\}^{i+2-j} & \text{for } j \leq i+2, \\ 0 & \text{otherwise.} \end{cases} \tag{3.18}$$

Lemma 3.1. *For the time-homogeneous birth–death process with birth rate $\beta_t = \alpha$ and death rate $\gamma_t = \mu$, the number of individuals alive immediately following the first death at time $t = 0$ satisfies*

$$X_1 \equiv Y(0) \sim \text{Geom}(\hat{\pi}).$$

For $k = 2, 3, \ldots$, X_k satisfies (3.3) with $\pi_{s_k} = \hat{\pi}$ and $\mathbf{B}_k$ given by (3.11), with $\hat{\mathbf{M}}_k(\tau)$ given in (3.17) replacing $\mathbf{M}_k(t; \tau)$.

Similarly, for $t \geq 0$, $Y(t)$ satisfies (3.4) with $\pi_t = \hat{\pi}$ and $\mathbf{D}_t$ given by (3.12) and (3.13), with $\hat{\mathbf{M}}_k(\tau)$ and $\hat{\mathbf{J}}_k(\tau)$ given in (3.17) and (3.18) replacing $\mathbf{M}_k(t; \tau)$ and $\mathbf{J}_k(t; \tau)$, respectively.

From Lemma 3.1, we observe that $X_k | S_k = \sum_{j=2}^{k} T_j < \infty$ (conditioning on at least k deaths occurring in the birth–death process) is a mixture of negative binomial random variables

$\{Q_j \sim \text{NegBin}(j, \hat{\pi});\ j = 2, 3, \ldots, k\}$. We can look to integrate over $\mathbf{T}_{2:k}$ to obtain the distribution of $X_k|S_k < \infty$, which can be equivalently expressed as $X_k|X_{k-1} > 0$. However, a direct argument, which we present below, yields Lemma 3.2 and gives a straightforward illustration of how the number of individuals alive immediately after a death follows a mixture of negative binomial distributions.

Lemma 3.2. *For* $k = 2, 3, \ldots$, *let* $\hat{R}_k$ *be a discrete random variable with support* $\{2, 3, \ldots, k\}$ *and probability mass function*

$$\mathbb{P}\big(\hat{R}_k = j|X_{k-1} > 0\big) = \frac{c_{k-1,k-j}\hat{\pi}^{k-j}}{\sum_{l=2}^{k} c_{k-1,k-l}\hat{\pi}^{k-l}},$$

where $\{c_{m,j}\}$ *are the entries of Catalan's triangle (see, for example,* [17]*), a triangular array satisfying the following: for* $j = 0, 1, \ldots$ *and* $m = 1, 2, \ldots$, $c_{m,0} = 1$, $c_{m,j} = 0$ $(j \geq m)$, *and*

$$c_{m,j} = \sum_{i=0}^{j} c_{m-1,i} = c_{m-1,j} + c_{m,j-1} = \frac{m-j}{m+j}\binom{m+j}{m}. \tag{3.19}$$

Then for $k = 2, 3, \ldots$,

$$\{X_k|X_{k-1} > 0\} \sim \text{NegBin}(\hat{R}_k, \hat{\pi}).$$

Proof. Given that the outcome (birth/death) of each event is independent, the number of births which take place between each pair of deaths are independent and are distributed according to $\hat{G} \sim \text{Geom}(\hat{\pi})$. Therefore, for $k = 1, 2, \ldots$,

$$X_k \overset{D}{=} X_{k-1} + \hat{G} - 1, \tag{3.20}$$

subject to $X_{k-1} > 0$. Given $X_0 = 1$ (the birth of the initial individual), it immediately follows from (3.20) that $X_1 \sim \hat{G}$.

To study X_k $(k > 1)$, we consider a Markov chain for the evolution of the birth–death process until the kth individual is born (which guarantees at least k deaths occur) or the process goes extinct. Let (a, d) denote the state in the Markov chain corresponding to a births and d deaths having occurred. Then for $k = 2, 3, \ldots$, the possible states are

$$\{(a, d);\ a = 1, 2, \ldots, k-1,\ d = 0, 1, \ldots, a\} \cup \{(k, d);\ d = 0, 1, \ldots, k-2\},$$

with a total of $L_k = (k+4)(k-1)/2$ possible states. The Markov chain starts in state (1,0), the birth of the initial individual, and has absorbing states $\mathcal{E} = \{(x, x);\ x = 1, 2, \ldots, k-1\}$ (the birth–death process dies out before the kth birth) and $\mathcal{I} = \{(k, d);\ d = 0, 1, \ldots, k-2\}$ (there are at least k births in the birth–death process). For $(a, d) \notin \mathcal{E} \cup \mathcal{I}$, we have that the probability of transiting from state (a, d) to state $(a+1, d)$ (birth) is $1 - \hat{\pi}$ and the probability of transiting from state (a, d) to state $(a, d+1)$ (death) is $\hat{\pi}$. Note that the process reaches an absorbing state in at most $2k - 3$ transitions, and if the birth–death process reaches $a = k$, it does so through a birth, so that there are at least two individuals alive $(a - d \geq 2)$.

Let V_k denote the final (absorbing) state of the Markov chain. For $d = 0, 1, \ldots, k-2$,

$$\mathbb{P}(V_k = (k, d)) = c_{k-1,d}(1 - \hat{\pi})^{k-1}\hat{\pi}^d, \tag{3.21}$$

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

where $\{c_{k-1,d}\}$ are the Catalan numbers given in (3.19). The derivation of (3.21) is as follows. Any path from (1,0) to (k, d) must include $k-1$ births and d deaths, so has probability $(1-\hat{\pi})^{k-1}\hat{\pi}^d$ of occurring. The admissible paths are those for which, at any point on the path, the number of births (including the initial ancestor) is always greater than the number of deaths. Therefore, the number of admissible paths is equivalent to the number of ways of counting the votes in the ballot theorem with candidates A and B having k and d votes, respectively, such that after the first vote, candidate A always has more votes than candidate B, with there being $c_{k-1,d}$ such paths. Hence, for $j = 2, 3, \ldots, k$,

$$\begin{aligned}\mathbb{P}(V_k = (k, k-j) | V_k \in \mathcal{I}) &= \frac{c_{k-1,k-j}(1-\hat{\pi})^{k-1}\hat{\pi}^{k-j}}{\sum_{l=2}^{k} c_{k-1,k-l}(1-\hat{\pi})^{k-1}\hat{\pi}^{k-l}} \\ &= \frac{c_{k-1,k-j}\hat{\pi}^{k-j}}{\sum_{l=2}^{k} c_{k-1,k-l}\hat{\pi}^{k-l}}. \end{aligned} \tag{3.22}$$

Conditioned upon the birth–death process reaching state $(k, k-j)$ $(j = 2, 3, \ldots, k)$, the number of additional births until the kth death is the sum of j independent and identically distributed $\hat{G}$ random variables. Hence, the number of individuals alive immediately following the kth death is distributed according to

$$\sum_{i=1}^{j} \hat{G}_i \sim \text{NegBin}(j, \hat{\pi}), \tag{3.23}$$

and the lemma follows immediately from (3.22) and (3.23). □

3.4. Partial observations of deaths

We turn our attention to the case where not every death is detected. Suppose δ_t denotes the probability that a death at time t is detected and that the detection or otherwise of a death is independent of all other events. The epidemic model considered in [10] can be constructed in this manner. We are in a situation similar to that studied in [9] and [11], although in those papers (i) the death and detection processes are independent and (ii) individuals do not die upon detection.

We modify the notation slightly to take account of partial detection of the death process. For $k = 2, 3, \ldots$, let $\tilde{T}_k$ denote the length of time between the $(k-1)$th and kth detected deaths, with the convention that $\tilde{T}_1$ denotes the time from the birth of the initial individual to the first detected death (with $\tilde{T}_1 = \infty$ if no death is ever detected). Given that $\tilde{T}_1 < \infty$, i.e. that a death is detected, we set time 0 to be the time at which the first death is detected. Let $\tilde{X}_k$ denote the number of individuals alive immediately after the kth detected death. In Theorem 3.2 we show that $\tilde{X}_k$ is a mixture of negative binomial random variables with mixture weights depending on $\tilde{T}_{2:k} = \tilde{t}_{2:k}$, similarly to Theorem 3.1. Note that we set $\tilde{s}_k = \sum_{j=2}^{k} \tilde{t}_j$.

As stated in Section 3.1, we assume that the birth and death rates are piecewise constant between detected deaths. That is, setting $\tilde{s}_0 = -\infty$, we assume for $\tilde{s}_{k-1} < t \leq \tilde{s}_k$ that $(\beta_t, \gamma_t, \delta_t) = (\tilde{\alpha}_k, \tilde{\mu}_k, \tilde{d}_k)$. Let $\tilde{q}_k = \tilde{\mu}_k/(\tilde{\alpha}_k + \tilde{\mu}_k)$ and $\tilde{p}_k = 1 - \tilde{q}_k$. Note that for $\tilde{s}_{k-1} < t \leq \tilde{s}_k$, $q_t = \tilde{q}_k$ and $p_t = \tilde{p}_k$. Let

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

$$
\begin{aligned}
u_k &= \sqrt{1-4\tilde{p}_k\tilde{q}_k(1-\tilde{d}_k)}, & \lambda_k &= \frac{1+u_k-2\tilde{p}_k}{1+u_k},\\
\nu_k &= \frac{1-u_k}{2\tilde{p}_k}, & \zeta_k &= \frac{1+u_k}{2\tilde{p}_k} = \frac{1}{1-\lambda_k},\\
\tilde{\phi}_k(\tau) &= \exp(-[\tilde{\alpha}_k+\tilde{\mu}_k]u_k\tau), & \tilde{\psi}_k(\tau) &= \frac{(1-\lambda_k)(1-\tilde{\phi}_k(\tau))}{1-\nu_k(1-\lambda_k)\tilde{\phi}_k(\tau)}.
\end{aligned}
\tag{3.24}
$$

Note that if $\tilde{d}_k = 1$, i.e. all deaths are detected, then the above simplify to $u_k = 1$, $\lambda_k = \tilde{q}_k$, $\nu_k = 0$, $\zeta_k = 1/\tilde{p}_k$, $\tilde{\phi}_k(\tau) = \exp(-[\tilde{\alpha}_k+\tilde{\mu}_k]\tau)$, and $\tilde{\psi}_k(\tau) = \tilde{p}_k[1-\exp(-[\tilde{\alpha}_k+\tilde{\mu}_k]\tau)]$. Thus λ_k, $\tilde{\phi}_k(\tau)$, and $\tilde{\psi}_k(\tau)$ take the roles of q_t, $\phi(t;\tau)$, and $\psi(t;\tau)$ in Section 3.2.

Let $\tilde{\pi}_1 = \lambda_1$, and for $k = 2, 3, \ldots$ let

$$
\tilde{\pi}_k = \frac{\lambda_k[1-\nu_k(1-\tilde{\pi}_{k-1})] - (1-\nu_k)[\lambda_k-\tilde{\pi}_{k-1}]\tilde{\phi}_k(\tilde{t}_k)}{1-\nu_k(1-\tilde{\pi}_{k-1}) + \nu_k[\lambda_k-\tilde{\pi}_{k-1}]\tilde{\phi}_k(\tilde{t}_k)}. \tag{3.25}
$$

Thus $\tilde{\pi}_k = \pi_{\tilde{s}_k}$, the success probability of the geometric at the kth detected death, and we can modify (3.25) to obtain π_t $(t \geq 0)$; see (3.29).

Theorem 3.2. *For the piecewise time-homogeneous linear birth–death process with parameters* $(\beta_t, \gamma_t, \delta_t) = (\tilde{\alpha}_k, \tilde{\mu}_k, \tilde{d}_k)$ $(\tilde{s}_{k-1} < t \leq \tilde{s}_k)$, *we have the following:*

1. $\tilde{X}_1 | \tilde{T}_1 < \infty \sim \text{Geom}(\tilde{\pi}_1)$.
2. *Let* $\tilde{\mathbf{B}}_2 = (1)$, *and for* $k = 3, 4, \ldots$, *let* $\tilde{\mathbf{B}}_k = (\tilde{B}_{k,2}, \tilde{B}_{k,3}, \ldots, \tilde{B}_{k,k})$ *be given by*

$$
\tilde{\mathbf{B}}_k = \left\{\left(\prod_{j=2}^{k-1}\tilde{\mathbf{M}}_j(\tilde{t}_{j+1})\right)\cdot \mathbf{1}_{k-2}^{\top}\right\}^{-1}\prod_{j=2}^{k-1}\tilde{\mathbf{M}}_j(\tilde{t}_{j+1}), \tag{3.26}
$$

where, for $\tau \geq 0$, $\tilde{\mathbf{M}}_k(\tau)$ *is the* $(k-1)\times k$ *matrix with (i,j)th element*

$$
\left[\tilde{\mathbf{M}}_k(\tau)\right]_{i,j} = \begin{cases} (i+1)\binom{i}{j-1}\tilde{h}_{k+1}(\tau)^{j-1}\left[1-\tilde{h}_{k+1}(\tau)\right]^{i+1-j}\tilde{r}_{k+1}(\tau)^i & \textit{for } j \leq i+1,\\ 0 & \textit{otherwise,}\end{cases}
$$

where

$$
\tilde{h}_{k+1}(\tau) = \frac{1-\tilde{\pi}_{k+1}-\tilde{\psi}_{k+1}(\tau)}{\left(1-\tilde{\pi}_{k+1}\right)\left(1-\tilde{\psi}_{k+1}(\tau)\right)} \tag{3.27}
$$

and

$$
\tilde{r}_{k+1}(\tau) = \frac{\tilde{\pi}_k\left[\lambda_{k+1}+(1-\lambda_{k+1})(1-\nu_{k+1})\tilde{\phi}_{k+1}(\tau)\right]}{\lambda_{k+1}\left[1-\nu_{k+1}\left(1-\tilde{\pi}_k\right)\right]-(1-\nu_{k+1})\left[\lambda_{k+1}-\tilde{\pi}_k\right]\tilde{\phi}_{k+1}(\tau)}.
$$

Then for $k = 2, 3, \ldots$, $\tilde{X}_k$, *the number of individuals alive immediately following the kth detected death, satisfies*

$$
\left\{\tilde{X}_k | \tilde{\mathbf{T}}_{2:k} = \tilde{\mathbf{t}}_{2:k}\right\} \sim \text{NegBin}\left(\tilde{R}_k, \tilde{\pi}_k\right), \tag{3.28}
$$

where $\tilde{R}_k$ *has support* $\{2, 3, \ldots, k\}$ *and* $\mathbb{P}(\tilde{R}_k = j) = \tilde{B}_{k,j}$ $(j = 2, 3, \ldots, k)$.

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

3. *For* $k = 2, 3, \ldots$, *let* $\tilde{\mathbf{J}}_k(\tau)$ *be the* $(k-1) \times (k+1)$ *matrix with (i,j)th entry*

$$\left[\tilde{\mathbf{J}}_k(\tau)\right]_{i,j} = \begin{cases} \binom{i+1}{j-1}\tilde{h}_{k+1}(\tau)^{j-1}\left[1-\tilde{h}_{k+1}(\tau)\right]^{i+2-j}\tilde{r}_{k+1}(\tau)^{i+1} & \textit{for } j \leq i+2, \\ 0 & \textit{otherwise.} \end{cases}$$

For $t \geq 0$, *let* $\tilde{k}_t$ *and* $\tilde{s}_{\tilde{k}_t}$ *denote the number of detected deaths and the time of the last detected death, respectively, up to and including time t. Let* $\tilde{\mathbf{D}}_t = (\tilde{D}_{t,0}, \tilde{D}_{t,1}, \ldots, \tilde{D}_{t,k_t})$ *satisfy, for* $t < \tilde{s}_2$,

$$\tilde{\mathbf{D}}_t = \left(\frac{\tilde{\pi}_t\tilde{\psi}_2(t)}{(1-\tilde{\pi}_t)(1-\tilde{\psi}_2(t))}, \frac{1-\tilde{\pi}_t-\tilde{\psi}_2(t)}{(1-\tilde{\pi}_t)(1-\tilde{\psi}_2(t))}\right),$$

where

$$\tilde{\pi}_t = \frac{\lambda_{\tilde{k}_t}\left[1-\nu_{\tilde{k}_t}\left(1-\tilde{\pi}_{\tilde{k}_t-1}\right)\right]-\left(1-\nu_{\tilde{k}_t}\right)\left[\lambda_{\tilde{k}_t}-\tilde{\pi}_{\tilde{k}_t-1}\right]\tilde{\phi}_{\tilde{k}_t}\left(t-\tilde{s}_{\tilde{k}_t}\right)}{1-\nu_{\tilde{k}_t}\left(1-\tilde{\pi}_{\tilde{k}_t-1}\right)+\nu_{\tilde{k}_t}\left[\lambda_{\tilde{k}_t}-\tilde{\pi}_{\tilde{k}_t-1}\right]\tilde{\phi}_{\tilde{k}_t}\left(t-\tilde{s}_{\tilde{k}_t}\right)}, \tag{3.29}$$

and for $t \geq \tilde{s}_2$,

$$\tilde{\mathbf{D}}_t = \left\{\left(\prod_{j=2}^{k_t-1}\tilde{\mathbf{M}}_j(t_{j+1})\right)\tilde{\mathbf{J}}_{k_t}\left(t-\tilde{s}_{\tilde{k}_t}\right)\cdot\mathbf{1}_{k-1}^{\top}\right\}^{-1}\left(\prod_{j=2}^{k_t-1}\tilde{\mathbf{M}}_j(t_{j+1})\right)\tilde{\mathbf{J}}_{k_t}\left(t-\tilde{s}_{\tilde{k}_t}\right).$$

Then Y(t) satisfies $Y(t) \equiv \tilde{X}_{\tilde{k}_t}$, *if* $t = \tilde{s}_{\tilde{k}_t}$, *and*

$$\left\{Y(t)|\tilde{\mathbf{T}}_{2:\tilde{k}_t} = \tilde{\mathbf{t}}_{2:\tilde{k}_t}\right\} \sim \text{NegBin}\left(\tilde{Z}(t), \pi_t\right),$$

where $\tilde{Z}(t)$ *has support* $\{0, 1, \ldots, k_t\}$ *and* $\mathbb{P}(\tilde{Z}(t) = j) = \tilde{D}_{t,j}$ $(j = 0, 1, \ldots, \tilde{k}_t)$.

3.5. General stochastic epidemic

Finally, we turn our attention to the time-inhomogeneous general stochastic epidemic, defined by (2.1) in Section 2, with one initial infective in an otherwise susceptible population of size N. Let $\{(S(t), I(t))\}$ denote the process of the numbers of susceptibles and infectives at time t; as with the birth–death process, we employ the convention that the first removal takes place at time 0. In a natural change of terminology for moving from the birth–death process to the epidemic process, let T_k denote the inter-arrival time between the $(k-1)$th and kth removals, and let S_k denote the time of the kth removal, with the convention that $S_1 = 0$.

The Markovian nature of the general stochastic epidemic model allows us to model the evolution of $\{(S(t), I(t))\}$ using continuous-time Markov chains; see, for example, [5] and [12]. To facilitate analysis of the model, we assume that the infection and removal parameters are piecewise constant, and for succinctness of presentation we assume, as with the birth–death process with partial detection of deaths, that the parameters are piecewise constant between removals. That is, we assume that for $s_{k-1} < t \leq s_k$, between the $(k-1)$th and kth removal, $\beta_t = \alpha_k$ and $\gamma_t = \mu_k$ in (2.1).

For $k = 1, 2, \ldots, N$ and $s_k \leq t < s_{k+1}$, we derive the distribution of $\{I(t)|\mathbf{T}_{2:k} = \mathbf{t}_{2:k}\}$ in Theorem 3.3 as a product of matrices which determine the transition in the number of infectives between removal events and the transition at a removal event. We continue by defining the required matrices before stating Theorem 3.3.

For $k = 0, 1, \ldots, N$, let

$$\Omega_k = \{(N-k-i, i) : i = 1, 2, \ldots, N-k\}$$

be the set of states of $\{(S(t), I(t))\}$ in which the epidemic is still going (i.e. there is at least one infective) and precisely k removals have occurred. Give the states in Ω_k the labels $k_1, k_2, \ldots, k_{N-k}$, where the state $(N-k-i, i)$ has label k_i $(i = 1, 2, \ldots, N-k)$. For $k = 0, 1, \ldots, N$, let $\mathbf{Q}_{k,k} = [q_{k_i,k_j}]$ be the $(N-k) \times (N-k)$ transition-rate matrix for transitions of $\{(S(t), I(t))\}$ *within* Ω_k. Then, using (2.1), with $\beta_t = \alpha_{k+1}$ and $\gamma_t = \mu_{k+1}$,

$$q_{k_i,k_j} = \begin{cases} -\left(\dfrac{\alpha_{k+1}}{N}(N-k-i)i + \mu_{k+1} i\right) & \text{if } j = i, \\ \dfrac{\alpha_{k+1}}{N}(N-k-i)i & \text{if } j = i+1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.30}$$

For $k = 0, 1, \ldots, N-1$, let $\mathbf{Q}_{k,k+1} = [q_{k_i,(k+1)_j}]$ be the $(N-k) \times (N-k-1)$ transition-rate matrix for transitions of $\{(S(t), I(t))\}$ from Ω_k to Ω_{k+1} (a removal), so that, using (2.1),

$$q_{k_i,(k+1)_j} = \begin{cases} i\mu_{k+1} & \text{if } i \in \{2, 3, \ldots, N-k\} \text{ and } j = i-1, \\ 0 & \text{otherwise.} \end{cases}$$

The above partitioning of the state space differs from that of both [5] and [12], where the state space is partitioned on the basis of the number of susceptibles remaining and the focus is on the final size (number of removals) of the epidemic.

In a similar vein, for $k = 0, 1, \ldots, N$, let

$$\tilde{\Omega}_k = \{(N-k-j, j) : j = 0, 1, \ldots, N-k\}$$

be the set of *all* states of $\{(S(t), I(t))\}$ in which precisely k removals have occurred. Thus $\tilde{\Omega}_k = \Omega_k \cup \{(N-k, 0)\}$. Let $\tilde{\mathbf{Q}}_{k,k}$ be the $(N-k+1) \times (N-k+1)$ transition-rate matrix for transitions of $\{(S(t), I(t))\}$ within $\tilde{\Omega}_k$. The elements of $\tilde{\mathbf{Q}}_{k,k}$ are given by (3.30), except that the indices now run from 0 to $N-k$. Note that the first row of $\tilde{\mathbf{Q}}_{k,k}$ comprises all zeros as the process has been absorbed. Finally, for $k = 0, 1, \ldots, N$, let $\tilde{\mathbf{Q}}_{k,k+1}$ be the $(N-k) \times (N-k)$ transition-rate matrix for transitions of $\{(S(t), I(t))\}$ from Ω_k to $\tilde{\Omega}_{k+1}$. Note that $\tilde{\mathbf{Q}}_{k,k+1}$ is the diagonal matrix with successive diagonal elements $\mu_{k+1}, 2\mu_{k+1}, \ldots, (N-k)\mu_{k+1}$.

Theorem 3.3. *Consider an epidemic satisfying* (2.1) *with one initial infective in an otherwise susceptible population of size N, and let $s_1 = 0$, so the first removal occurs at time 0.*

For $k = 2, 3, \ldots, N$, $i = 0, 1, \ldots, N-k$, and $\tau \geq 0$, let

$$v_{k,i}(\tau|\mathbf{t}_{2:k}) = \mathbb{P}(I(s_k + \tau) = i | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}, T_{k+1} > \tau),$$

with $v_{1,i}(\tau) = \mathbb{P}(I(\tau) = i | T_2 > \tau)$, and let

$$\mathbf{v}_k(\tau|\mathbf{t}_{2:k}) = (v_{k,0}(\tau|\mathbf{t}_{2:k}), v_{k,1}(\tau|\mathbf{t}_{2:k}), \ldots, v_{k,N-k}(\tau|\mathbf{t}_{2:k})).$$

Then

$$\mathbf{v}_k(\tau|\mathbf{t}_{2:k}) = \frac{1}{c_k(\tau)} \mathbf{u}_1 \mathbf{Q}_{0,0}^{-1} \left(\prod_{i=1}^{k-1} \mathbf{Q}_{i-1,i} \exp(\mathbf{Q}_{i,i} t_{i+1}) \right) \tilde{\mathbf{Q}}_{k-1,k} \exp\big(\tilde{\mathbf{Q}}_{k,k} \tau\big), \tag{3.31}$$

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

where $\mathbf{u}_1$ *is the row vector of length N whose first element is 1 and all other elements are 0,*

$$c_k(\tau) = \mathbf{u}_1\mathbf{Q}_{0,0}^{-1}\left(\prod_{i=1}^{k-1} \mathbf{Q}_{i-1,i}\exp(\mathbf{Q}_{i,i}t_{i+1})\right)\tilde{\mathbf{Q}}_{k-1,k}\exp(\tilde{\mathbf{Q}}_{k,k}\tau)\cdot \mathbf{1}_{N+1-k}^{\top}, \tag{3.32}$$

and the product in (3.31) *is equal to* $\mathbf{I}_{N-1}$ *if* $k=1$.

Also setting $\tau=0$ *in* (3.31) *yields the conditional distribution of the number infected immediately after the kth removal, given only the removal times. Note that* $\exp(\tilde{\mathbf{Q}}_{k,k}0) = \mathbf{I}_{N-k+1}$.

The key difference from the approach for the birth–death process is the effect of the finite size of the population, which means that infectives do not behave independently, but that for moderate N, matrix multiplication directly yields the distribution of $I(t)$ given only removal times. To obtain the distribution of $I(t)$ we require matrix exponentials and matrix multiplication with the initial matrices of size $N\times N$. By comparison, the computation of the distribution of $Z(t)$ involves successive vector–matrix multiplication with the largest matrices of size $(k_t-1)\times(k_t+1)$. Therefore, it is much faster to compute the distribution of $Z(t)$, and subsequently the distribution of $Y(t)$, than the distribution of $I(t)$.

4. Time-inhomogeneous Markov model

In this section, we study the time-inhomogeneous linear birth–death process, with the main aim of proving Theorem 3.1. We begin by assuming that at time $t=0$ (the first death), $X_1 \sim \mathrm{Geom}(\pi_0)$ for some $0<\pi_0<1$. For example, if the birth and death rates are constants, α_1 and μ_1 respectively, prior to the first death, then we can follow Section 3.3 and obtain $\pi_0 = \alpha_1/(\alpha_1+\mu_1)$. We address the distribution of X_1 ($Y(0)$) more fully at the end of this section, after the proof of Theorem 3.1.

The first step to proving Theorem 3.1 is to show how the birth–death process progresses between death events. In Lemma 4.1 we show that if we start with one initial individual at time t, then given that no deaths have been observed by time $t+\tau$, the number of offspring originating from the initial individual follows a geometric distribution. Before stating and proving Lemma 4.1 we introduce some notation.

For a non-negative, integer random variable W, $t\geq 0$, and $\tau>0$, let $E_{t,\tau}^{W}$ and $D_{t,\tau}^{W}$ respectively denote the events that there are no deaths in the interval $[t, t+\tau)$ and that the first death in the birth–death process after time t is at time $t+\tau$, given that there are W individuals in the population at time t. For $w\in\mathbb{N}$, if $\mathbb{P}(W=w)=1$, we employ the convention $E_{t,\tau}^{w}$.

Lemma 4.1. *For* $t\geq 0$ *and* $\tau>0$, *let* $Y^{(t)}(\tau)$ *denote the number of individuals alive at time* $t+\tau$, *given that one individual is alive at time t. Then for* $n\in\mathbb{N}$,

$$\mathbb{P}\left(Y^{(t)}(\tau)=n,\, E_{t,\tau}^{1}\right) = \psi(t;\tau)^{n-1}\phi(t;\tau), \tag{4.1}$$

where $\phi(t;\tau) = \exp\left(-\int_t^{t+\tau}(\beta_s+\gamma_s)\,ds\right)$ *and* $\psi(t;\tau) = \int_t^{t+\tau}\beta_u\exp\left(-\int_u^{t+\tau}(\beta_s+\gamma_s)ds\right)du$ *(see* (3.6) *and* (3.7)*). Therefore, we have that*

$$\mathbb{P}\left(E_{t,\tau}^{1}\right) = \frac{\phi(t;\tau)}{1-\psi(t;\tau)}, \tag{4.2}$$

and

$$Y^{(t)}(\tau)\,|\,E_{t,\tau}^{1} \sim 1+\mathrm{Geom}(1-\psi(t;\tau)). \tag{4.3}$$

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

Proof. The proof is similar to the exploration process in [9, Section 2].

First, we note that the probability that the initial individual survives to time $t+\tau$ is $\exp\Big(-\int_t^{t+\tau}\gamma_s\,ds\Big)$. We then start at time $t+\tau$ and explore the lifeline of the initial individual back from time $t+\tau$ until we either discover an offspring or reach time t. The probability that we reach time t, i.e. that the initial individual has no offspring, is $\exp\Big(-\int_t^{t+\tau}\beta_s\,ds\Big)$, with $\exp\Big(-\int_t^{t+\tau}\gamma_s\,ds\Big)\times\exp\Big(-\int_t^{t+\tau}\beta_s\,ds\Big)=\phi(t;\tau)$.

The (defective) probability density function for the time of the first offspring, looking back from time $t+\tau$, is $\beta_u\exp\Big(-\int_{t+u}^{t+\tau}\beta_s\,ds\Big)$. Therefore the probability that the initial individual has at least one offspring and their first offspring (looking back from time $t+\tau$) survives to time $t+\tau$ is given by $\psi(t;\tau)$. If the initial individual has at least one offspring and their first offspring survives to time $t+\tau$, then we repeat the above process of exploring lifelines back from time $t+\tau$ until we either discover an offspring or reach time t. This will start with the offspring's lifeline, and if they have no offspring, we will continue with the unexplored lifeline of the initial individual. The total length of the lifeline to explore is again of length τ, and therefore, as above, the probability of no additional offspring is $\exp\Big(-\int_t^{t+\tau}\beta_s\,ds\Big)$, while the probability of at least one offspring and that the first offspring discovered survives until time $t+\tau$ is $\psi(t;\tau)$. We can repeat this process by at each stage considering the combined unexplored lifelines of length τ and whether or not an offspring is discovered, and if an offspring is discovered whether or not it survives to time $t+\tau$. This yields (4.1).

By summing over n in (4.1), we have (4.2), and (4.3) follows immediately. □

Before we consider the birth–death process evolving from a random number of individuals at time t, and specifically a geometrically distributed number of individuals in Lemma 4.3, we give the following elementary lemma concerning the sums of geometric random variables, which will prove useful throughout the remainder of the paper.

Lemma 4.2. *Let $X, Y_1, Y_2, \ldots$ be independent, with $X\sim\mathrm{Geom}(q_1)$ and $Y_i\sim 1+\mathrm{Geom}(q_2)$ $(i=1,2,\ldots)$. Let*

$$Z=\sum_{i=1}^{X}Y_i,$$

where $Z=0$ if $X=0$. Then

$$Z\overset{D}{=}\begin{cases}\tilde{X} & \textit{with probability } \dfrac{1-q_1}{1-q_1q_2},\\[2ex] 0 & \textit{with probability } \dfrac{q_1(1-q_2)}{1-q_1q_2},\end{cases}$$

where $\tilde{X}\sim\mathrm{Geom}(q_1q_2)$.

Proof. This is elementary using probability generating functions. An alternative, more constructive, proof is available by noting that if $X'\sim 1+\mathrm{Geom}(q_1)$ and $Y_i\sim 1+\mathrm{Geom}(q_2)$ $(i=1,2,\ldots)$ are independent then $\sum_{i=1}^{X'}Y_i\sim 1+\mathrm{Geom}(q_1q_2)$. □

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

For $t \geq 0$, let

$$G_t \sim \text{Geom}(\pi_t), \tag{4.4}$$

where π_t is given by (3.2) with initial condition π_0 satisfying $0 < \pi_0 < 1$.

Lemma 4.3. *For $t \geq 0$, suppose that the number of individuals alive in a birth–death process at time t is distributed according to G_t. For $\tau > 0$, let $W_t(\tau)$ denote the number of individuals alive at time $t + \tau$. Then*

$$W_t(\tau) | E_{t;\,\tau}^{G_t} \overset{D}{=} \begin{cases} G_{t+\tau} & \textit{with probability } h(t;\tau) = \dfrac{(1-\pi_t)\phi(t;\tau)}{(1-\pi_{t+\tau})[1-\psi(t;\tau)]}, \\ 0 & \textit{with probability } 1-h(t;\tau) = \dfrac{\pi_{t+\tau}\psi(t;\tau)}{(1-\pi_{t+\tau})[1-\psi(t;\tau)]}. \end{cases} \tag{4.5}$$

Proof. Firstly, using Lemma 4.1, it is easily shown that

$$\left\{G_t | E_{t;\,\tau}^{G_t}\right\} \overset{D}{=} \tilde{X} \sim \text{Geom}\left(1 - \frac{(1-\pi_t)\phi(t;\tau)}{1-\psi(t;\tau)}\right).$$

The birth–death processes from each individual alive at time t proceed independently, so by Lemma 4.1,

$$\left\{W_t(\tau) | E_{t;\,\tau}^{G_t}\right\} = \sum_{i=1}^{\tilde{X}} \tilde{Y}_i, \tag{4.6}$$

where the $\tilde{Y}_i$s are independent and identically distributed according to $\tilde{Y} \sim 1 + \text{Geom}(1 - \psi(t;\tau))$. From (3.8), we have that

$$\left(1 - \frac{(1-\pi_t)\phi(t;\tau)}{1-\psi(t;\tau)}\right)\{1-\psi(t;\tau)\} = \pi_{t+\tau}. \tag{4.7}$$

Then the lemma follows immediately from (4.6) and (4.7), using Lemma 4.2. □

An immediate consequence of Lemma 4.3 is that if $T_2 > t$, then

$$Y(t) \overset{D}{=} \left\{W_0(t) | E_{0;\,t}^{G_0}\right\},$$

thus proving (3.4) for $0 \leq t < s_2$.

We are now in a position to show that if a second death is observed in the birth–death process, then the distribution of X_2 only depends on $T_2 = t_2$ through π_{t_2}.

Lemma 4.4. *For any* $0 < t_2 < \infty$,

$$\{X_2 | T_2 = t_2\} \sim Q_2(t_2) = \text{NegBin}(2, \pi_{t_2}).$$

Proof. Given that $X_1 \overset{D}{=} G_0$ and $T_2 = t_2$, for any $0 \leq \tau < t_2$, we have that $Y(\tau) \overset{D}{=} W_0(\tau) | E_{0;\,\tau}^{G_0}$, given in Lemma 4.3, (4.5). Therefore, for $x = 0, 1, \ldots,$

$$\begin{aligned} \mathbb{P}(X_2 = x | T_2 = t_2) &= \lim_{\tau \uparrow t_2} \mathbb{P}(Y(\tau) = x+1 | T_2 = t_2) \\ &= \lim_{\tau \uparrow t_2} \frac{f_{T_2}(t_2 | Y(\tau) = x+1,\, T_2 > \tau)\mathbb{P}(Y(\tau) = x+1 | T_2 > \tau)\mathbb{P}(T_2 > \tau)}{f_{T_2}(t_2)}. \end{aligned} \tag{4.8}$$

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

We consider the four terms on the right-hand side of (4.8). The first two terms are

$$\lim_{\tau\uparrow t_2} f_{T_2}(t_2|Y(\tau)=x+1,\, T_2>\tau)=(x+1)\gamma_{t_2} \tag{4.9}$$

and, using Lemma 4.3,

$$\lim_{\tau\uparrow t_2} \mathbb{P}(Y(\tau)=x+1|T_2>\tau)=\frac{(1-\pi_0)\phi(0;t_2)}{(1-\pi_{t_2})[1-\psi(0;t_2)]}\times(1-\pi_{t_2})^{x+1}\pi_{t_2}. \tag{4.10}$$

Since $\mathbb{P}(T_2>\tau)=\mathbb{P}(E^{G_0}_{0;\tau})$, it is straightforward, using (4.2) and (3.8), to show that

$$\lim_{\tau\uparrow t_2}\mathbb{P}(T_2>\tau)=\lim_{\tau\uparrow t_2}\frac{\pi_0[1-\psi(0;\tau)]}{\pi_\tau}=\frac{\pi_0[1-\psi(0;t_2)]}{\pi_{t_2}}. \tag{4.11}$$

Finally, for $t\geq 0$,

$$f_{T_2}(t)=-\frac{d}{dt}\mathbb{P}(T_2>t).$$

Since $\pi_t=1-\psi(0;t)-(1-\pi_t)\phi(0;t)$, with $\phi'(0,t)=-\{\beta_t+\gamma_t\}\phi(0,t)$ and

$$\psi'(0,t)=-\{\beta_t+\gamma_t\}\psi(0;t)+\beta_t,$$

it follows by the quotient rule that

$$\begin{aligned} f_{T_2}(t)&= \\ &-\pi_0\frac{\psi'(0,t)\{1-\psi(0;t)-(1-\pi_0)\phi(0;t)\}-[1-\psi(0,t)]\{-\psi'(0,t)-(1-\pi_0)\phi'(0;t)\}}{\pi_t^2} \\ &=-\pi_0(1-\pi_0)\phi(0;t)\frac{\psi'(0,t)-\beta_t-\gamma_t+(\beta_t+\gamma_t)\psi(0,t)}{\pi_t^2} \\ &=\frac{\pi_0(1-\pi_0)\phi(0;t)\gamma_t}{\pi_t^2}. \end{aligned} \tag{4.12}$$

Substituting (4.9)–(4.12) into (4.8), we obtain, for $x=0,1,\ldots$,

$$\begin{aligned}\mathbb{P}(X_2=x|T_2=t_2)&=(x+1)\gamma_{t_2}\times(1-\pi_{t_2})^{x+1}\pi_{t_2}\frac{(1-\pi_0)\phi(0;t_2)}{(1-\pi_{t_2})(1-\psi(0;t_2))} \\ &\quad\times\frac{\pi_0(1-\psi(0;t_2))}{\pi_{t_2}}\Bigg/\frac{\pi_0(1-\pi_0)\phi(0;t_2)\gamma_{t_2}}{\pi_{t_2}^2} \\ &=(x+1)(1-\pi_{t_2})^x\pi_{t_2}^2,\end{aligned}$$

as required. □

An immediate consequence of Lemma 4.4 is that the distribution of $X_k|\mathbf{T}_{2:k}=\mathbf{t}_{2:k}$ given by (3.3) holds for $k=2$ with $\mathbb{P}(R_2=2)=1$. We proceed by building upon Lemmas 4.3 and 4.4 to derive $Z(t)$ for $t>s_2$, and $R_k=Z(s_k)$ $(k=3,4,\ldots)$.

Corollary 4.1. *Suppose that at time $t\geq 0$ there exists $j\in\mathbb{N}$ such that $Z(t)=j$; that is, there are $Q_j(t)\sim$ NegBin(j,π_t) individuals in the population. Then*

$$\left\{Z(t+\tau)|Z(t)=j,\,E^{Q_j(t)}_{t,t+\tau}\right\}\sim \text{Bin}(j,h(t;\tau)) \tag{4.13}$$

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

and

$$\left\{Z(t+\tau)|Z(t)=j, D_{t,t+\tau}^{Q_j(t)}\right\} \sim 2+\text{Bin}(j-1, h(t;\tau)), \tag{4.14}$$

where $h(t;\tau)$ is defined in Lemma 4.3, (4.5).

Similarly, for $j, k=1, 2, \ldots$ we have, for $s_k, t_{k+1} \geq 0$,

$$\{R_{k+1}|R_k=j, S_k=s_k, T_{k+1}=t_{k+1}\} \sim 2+\text{Bin}(j-1, h(s_k;t_{k+1})). \tag{4.15}$$

Proof. For $t \geq 0$, suppose that at time t a family group comprises a random number of individuals distributed according to Geom(π_t). Then $Z(t)$ denotes the number of family groups at time t; note that a family group can contain 0 individuals. The corollary follows immediately from Lemmas 4.3 and 4.4, since the j family groups present at time t evolve independently. If there are no deaths in the interval $[t, t+\tau)$, then all j family groups, independently, behave according to (4.5), giving (4.13). On the other hand, if the first death after time t is at time $t+\tau$, one family group must be responsible for the death, and the size of that family group following the death is NegBin($2, \pi_{t+\tau}$), by a modification of the arguments in Lemma 4.4 with a time shift of t. That is, the family group responsible for the death splits (or gives birth to a family group) to become two family groups. The other $j-1$ family groups have experienced no deaths in an interval of length τ, and the sizes of these family groups are independently distributed according to (4.5), giving (4.14). An identical argument gives (4.15). □

We continue by studying $\{(R_j, T_j);\ j=1, 2, \ldots\}$ in detail, with $Z(t)$, and consequently X_k and $Y(t)$, following trivially. Before proceeding we note that by the Markov property, $\{(R_j, T_j); j=(k+1), (k+2), \ldots\}$ depends on $\{(R_l, T_l); l=2, 3, \ldots, k\}$ through R_k only. Also, if $R_1 \equiv 1$, which is the case in the statement of Theorem 3.1 as we can write $X_1 \sim$ NegBin($1, \pi_0$), it follows that R_k only takes values in the range $\{2, 3, \ldots, k\}$. The process $\{(R_j, T_j);\ j=1, 2, \ldots\}$ is a (possibly terminating) semi-Markov sequence (see, for example, [16] or [7]). Lemma 4.5 gives a recursive relationship expressing $\mathbf{B}_k$ (the probability mass function of R_k) in terms of $\mathbf{B}_{k-1}$ (the probability mass function of R_{k-1}) and $T_k=t_k$, after which we will be in a position to complete the proof of Theorem 3.1.

Lemma 4.5. *For $k=3, 4, \ldots$ and $j=2, 3, \ldots, k$, $B_{k,j}$ satisfies*

$$B_{k,j} = \frac{\sum_{l=j-1}^{k-1} \binom{l-1}{j-2} \left\{\frac{(1-\pi_{s_{k-1}})\pi_{s_{k-1}}}{(1-\pi_{s_k})\pi_{s_k}} \phi(s_{k-1};t_k)\right\}^{j-2} \left\{\frac{\pi_{s_{k-1}}}{1-\pi_{s_k}} \psi(s_{k-1};t_k)\right\}^{l+1-j} lB_{k,l}}{\sum_{m=2}^{k} \left\{\frac{\pi_{s_k}}{\pi_{s_{k+1}}}(1-\psi(s_k;t_{k+1}))\right\}^{m-1} mB_{k,m}}. \tag{4.16}$$

Hence,

$$\mathbf{B}_k = \frac{1}{C_{k-1}} \mathbf{B}_{k-1}\mathbf{M}_{k-1}(s_{k-1};t_k), \tag{4.17}$$

where, for $\tau \geq 0$, $\mathbf{M}_{k-1}(t;\tau)$ is a $(k-2) \times (k-1)$ matrix given in (3.9), and

$$C_{k-1} = \sum_{m=2}^{k-1} \left\{\frac{\pi_{s_{k-1}}(1-\psi(s_{k-1};t_k))}{\pi_{s_k}}\right\}^{m-1} mB_{k-1,m} = \mathbf{B}_{k-1}\mathbf{M}_{k-1}(s_{k-1};t_k) \cdot \mathbf{1}_{k-1}^{\top}. \tag{4.18}$$

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

Proof. By conditioning upon R_{k-1}, we have that

$$\begin{aligned} B_{k,j} &= \mathbb{P}(R_k = j | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}) \\ &= \sum_{l=2}^{k-1} \mathbb{P}(R_k = j, R_{k-1} = l | \mathbf{T}_{2:k} = \mathbf{t}_{2:k}) \\ &= \sum_{l=2}^{k-1} \frac{\mathbb{P}(R_k = j | T_k = t_k, R_{k-1} = l) f_{T_k}(t_k | \mathbf{T}_{2:k-1} = \mathbf{t}_{2:k-1}, R_{k-1} = l) B_{k-1,l}}{f_{T_k}(t_k | \mathbf{T}_{2:k-1} = \mathbf{t}_{2:k-1})}. \end{aligned} \tag{4.19}$$

The denominator in (4.19) satisfies

$$f_{T_k}(t_k | \mathbf{T}_{2:k-1} = \mathbf{t}_{2:k-1}) = \sum_{m=2}^{k-1} f_{T_k}(t_k | \mathbf{T}_{2:k-1} = \mathbf{t}_{2:k-1}, R_{k-1} = m) B_{k-1,m}, \tag{4.20}$$

and by considering the $R_{k-1} = j$ independent family groups, it follows by using (4.11) and (4.12) that for $\tau > 0$,

$$\begin{aligned} f_{T_k}(\tau | R_{k-1} = j, \mathbf{T}_{2:k-1} = \mathbf{t}_{2:k-1}) = j \left\{ \frac{\pi_{s_{k-1}} (1 - \psi(s_{k-1}; \tau))}{\pi_{s_{k-1}+\tau}} \right\}^{j-1} \\ \times \frac{\pi_{s_{k-1}} (1 - \pi_{s_{k-1}}) \phi(s_{k-1}; \tau) \gamma_{s_{k-1}+\tau}}{\pi^2_{s_{k-1}+\tau}}. \end{aligned} \tag{4.21}$$

Let $L_{k-1} = \{\pi_{s_{k-1}} (1 - \pi_{s_{k-1}}) \phi(s_{k-1}; t_k) \gamma_{s_k}\} / \pi^2_{s_k}$; then it follows from (4.20) and (4.21) that

$$f_{T_k}(t_k | \mathbf{T}_{2:k-1} = \mathbf{t}_{2:k-1}) = L_{k-1} \sum_{m=2}^{k-1} m \left\{ \frac{\pi_{s_{k-1}} (1 - \psi(s_{k-1}; t_k))}{\pi_{s_k}} \right\}^{m-1} B_{k-1,m}. \tag{4.22}$$

Given that $R_k \le R_{k-1} + 1$, and from (3.8), $\pi_{s_k} = 1 - \psi(s_{k-1}; t_k) - (1 - \pi_{s_{k-1}}) \phi(s_{k-1}; t_k)$, it follows that

$$\begin{aligned} &\sum_{l=2}^{k-1} \mathbb{P}(R_k = j | T_k = t_k, R_{k-1} = l) f_{T_k}(t_k | \mathbf{T}_{2:k-1} = \mathbf{t}_{2:k-1}, R_{k-1} = l) B_{k-1,l} \\ &= \sum_{l=j-1}^{k-1} \binom{l-1}{j-2} h(s_{k-1}; t_k)^{j-2} \{1 - h(s_{k-1}; t_k)\}^{l+1-j} \\ &\quad \times l \left\{ \frac{\pi_{s_{k-1}} (1 - \psi(s_{k-1}; t_k))}{\pi_{s_k}} \right\}^{l-1} \frac{\pi_{s_{k-1}} (1 - \pi_{s_{k-1}}) \phi(s_{k-1}; t_k) \gamma_{s_k}}{\pi^2_{s_k}} \times B_{k-1,l} \\ &= L_{k-1} \sum_{l=j-1}^{k-1} \binom{l-1}{j-2} \left\{ \frac{(1 - \pi_{s_{k-1}}) \phi(s_{k-1}; t_k)}{(1 - \pi_{s_k})(1 - \psi(s_{k-1}; t_k))} \right\}^{j-2} \left\{ 1 - \frac{(1 - \pi_{s_{k-1}}) \phi(s_{k-1}; t_k)}{(1 - \pi_{s_k})(1 - \psi(s_{k-1}; t_k))} \right\}^{l+1-j} \\ &\quad \times l \left\{ \frac{\pi_{s_{k-1}} (1 - \psi(s_{k-1}; t_k))}{\pi_{s_k}} \right\}^{l-1} \times B_{k-1,l} \end{aligned}$$

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

$$
\begin{aligned}
&= L_{k-1} \sum_{l=j-1}^{k-1} \binom{l-1}{j-2} \left\{ \frac{\pi_{s_{k-1}}(1-\pi_{s_{k-1}})\phi(s_{k-1};t_k)}{(1-\pi_{s_k})\pi_{s_k}} \right\}^{j-2} \\
&\quad \times \left\{ \frac{\pi_{s_{k-1}}}{(1-\pi_{s_k})\pi_{s_k}} (\pi_{s_k}(1-\psi(s_{k-1};t_k)) - (1-\pi_{s_{k-1}})\phi(s_{k-1};t_k)) \right\}^{l+1-j} \times lB_{k-1,l} \\
&= L_{k-1} \sum_{l=j-1}^{k-1} \binom{l-1}{j-2} \left\{ \frac{\pi_{s_{k-1}}(1-\pi_{s_{k-1}})}{(1-\pi_{s_k})\pi_{s_k}} \phi(s_{k-1};t_k) \right\}^{j-2} \left\{ \frac{\pi_{s_{k-1}}}{(1-\pi_{s_k})} \psi(s_{k-1};t_k) \right\}^{l+1-j} lB_{k-1,l}.
\end{aligned}
\tag{4.23}
$$

Therefore (4.16) follows from substituting (4.22) and (4.23) into (4.19).

It is straightforward to check that $C_{k-1}^{-1}\mathbf{B}_{k-1}\mathbf{M}_{k-1}(s_{k-1};t_k)$ gives $\mathbf{B}_k$ satisfying (4.17), and the lemma is proved. □

Proof of Theorem 3.1. By noting that $\mathbf{B}_2 = \mathbf{I}_1$ (the 1×1 identity matrix) and iterating from k to 2, it follows from Lemma 4.5, (4.17), that $\mathbf{B}_k$ satisfies (3.11), and (3.3) is proved.

We now turn to $Z(t)$. As noted after Lemma 4.3, for $0 \le t < s_2$ the probability mass function of $Z(t)$ is given by $\mathbf{D}_t$ defined in (3.12). For $t \ge s_2$, $k_t \ge 2$, and let $\sigma_t = t - s_{k_t}$ denote the time since the last death, up to and including time t. It follows, by arguments similar to those in the proof of Lemma 4.5, that for $k_t = 2, 3, \ldots$ and $j = 0, 1, \ldots, k_t$,

$$
\begin{aligned}
D_{t,j} &= \sum_{l=2}^{k_t} \mathbb{P}\big(Z(t) = j, Z\big(s_{k_t}\big) = R_{k_t} = l \big| \mathbf{T}_{2:k_t} = \mathbf{t}_{2:k_t}\big) \\
&= \frac{\sum_{l=2}^{k_t} \mathbb{P}\big(Z(t) = j \,\big|Z\big(s_{k_t}\big) = l, \mathbf{T}_{2:k_t} = \mathbf{t}_{2:k_t}\ \big)\mathbb{P}\big(T_{k_t+1} > \sigma_t | Z\big(s_{k_t}\big) = l\big) B_{k_t,l}}{\sum_{m=2}^{k_t} \mathbb{P}\big(T_{k_t+1} > \sigma_t | Z\big(s_{k_t}\big) = m\big) B_{k_t,m}} \\
&= \frac{\sum_{l=j}^{k_t} \binom{l}{j} h(s_{k_t};\sigma_t)^j \{1 - h(s_{k_t};\sigma_t)\}^{l-j} \times \left\{ \frac{\pi_{s_{k_t}}}{\pi_t} \big(1 - \psi(s_{k_t};\sigma_t)\big) \right\}^l \times B_{k_t,l}}{\sum_{m=2}^{k_t} \left\{ \frac{\pi_{s_{k_t}}}{\pi_t} \big(1 - \psi(s_{k_t};\sigma_t)\big) \right\}^m B_{k_t,m}} \\
&= \frac{\sum_{l=j}^{k_t} \binom{l}{j} \left\{ \frac{(1-\pi_{s_{k_t}})\pi_{s_{k_t}}}{(1-\pi_t)\pi_t} \phi(s_k;\sigma_t) \right\}^j \left\{ \frac{\pi_{s_{k_t}}}{1-\pi_t} \psi(s_k;\sigma_t) \right\}^{l-j} B_{k_t,l}}{\sum_{m=2}^{k_t} \left\{ \frac{\pi_{s_{k_t}}}{\pi_t} \big(1 - \psi(s_k;\sigma_t)\big) \right\}^m B_{k_t,m}}.
\end{aligned}
\tag{4.24}
$$

It is straightforward to combine (4.24) with Lemma 4.5 to show that $\mathbf{D}_t$ can be expressed in matrix form as (3.13), completing the proof of Theorem 3.1. □

We return to the distribution of X_1. As already noted, if, for $t < 0$, $\beta_t = \alpha_1$ and $\gamma_t = \mu_1$, then $X_1 \sim \text{Geom}\,(\pi_0)$ with $\pi_0 = \mu_1/(\alpha_1 + \mu_1)$. Suppose that $T_1 = t_1$ is known; for example, it might be known, or found by contact tracing, when the introductory individual entered the population. Then the initial individual is born at time $-t_1$, and it follows from Lemma 4.1 and arguing along similar lines to Lemma 4.4 that, for $x = 0, 1, \ldots$,

$$
\mathbb{P}(X_1 = x | T_1 = t_1) = (x+1)\,(\psi(-t_1;t_1))^x\,(1 - \psi(-t_1;t_1))^2\,,
$$

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

and therefore $X_1|T_1 = t_1 \sim \text{NegBin}(2, 1 - \psi(-t_1; t_1))$. It is straightforward to adjust the above arguments by setting $\pi_0 = 1 - \psi(-t_1; t_1)$ and using (3.8) to obtain the distribution $X_k|\mathbf{T}_{1:k} = \mathbf{t}_{1:k}$ and $Y(t)|\mathbf{T}_{1:k} = \mathbf{t}_{1:k_t}$. Specifically, for $k = 1, 2, \ldots$, the development of the birth–death process with known introductory time after the kth death will mirror the development of the birth–death process with unknown introductory time after the $(k+1)$th death. Consequently, for the case of known introductory time, the distribution of the size of the population at time t will be a mixture of $\{\text{NegBin}(j, \pi_t);\ j = 0, 1, \ldots, k_t + 1\}$.

We conclude this section with the proof of Corollary 3.1.

Proof of Corollary 3.1. In the case $k = 2$, (3.16) follows immediately from (4.12) in the proof of Lemma 4.4.

To prove (3.16) for $k > 2$, we can use induction. From (4.22) and (4.18), we have that

$$\begin{aligned} f_{T_k}(t_k|\mathbf{T}_{2:k-1} = \mathbf{t}_{2:k-1}) &= L_{k-1}C_{k-1} \\ &= \frac{\pi_{s_{k-1}}(1-\pi_{s_{k-1}})\phi(s_{k-1}; t_k)\gamma_{s_k}}{\pi_{s_k}^2} \times \left[\mathbf{B}_{k-1}M_{k-1}(s_{k-1}; t_k)\cdot \mathbf{1}_{k-1}^\top\right]. \end{aligned}$$

Using (3.11), we have that

$$\begin{aligned} &\mathbf{B}_{k-1}M_{k-1}(s_{k-1}; t_k)\cdot \mathbf{1}_{k-1}^\top \\ &= \left[\left\{\prod_{j=2}^{k-2} M_j(s_j; t_{j+1})\cdot \mathbf{1}_{k-2}^\top\right\}^{-1} \prod_{j=2}^{k-2} M_j(s_j; t_{j+1})\right] M_{k-1}(s_{k-1}; t_k)\cdot \mathbf{1}_{k-1}^\top \\ &= \left\{\prod_{j=2}^{k-2} M_j(s_j; t_{j+1})\cdot \mathbf{1}_{k-2}^\top\right\}^{-1}\left\{\prod_{j=2}^{k-1} M_j(s_j; t_{j+1})\cdot \mathbf{1}_{k-1}^\top\right\}. \end{aligned} \tag{4.25}$$

Therefore,

$$\begin{aligned} f_{\mathbf{T}_{2:k}}(\mathbf{t}_{2:k}) &= f_{T_k}(t_k|\mathbf{T}_{2:k-1} = \mathbf{t}_{2:k-1}) f_{\mathbf{T}_{2:k-1}}(\mathbf{t}_{2:k-1}) \\ &= \frac{\pi_{s_{k-1}}(1-\pi_{s_{k-1}})\phi(s_{k-1}; t_k)\gamma_{s_k}}{\pi_{s_k}^2}\left\{\prod_{j=2}^{k-2} M_j(s_j; t_{j+1})\cdot \mathbf{1}_{k-2}^\top\right\}^{-1}\left\{\prod_{j=2}^{k-1} M_j(s_j; t_{j+1})\cdot \mathbf{1}_{k-1}^\top\right\} \\ &\quad \times \prod_{j=2}^{k-1}\frac{\pi_{s_{j-1}}(1-\pi_{s_{j-1}})\gamma_{s_j}\phi(s_{j-1}; t_j)}{\pi_{s_j}^2} \times \left\{\left[\prod_{j=2}^{k-2} M_j(s_j; t_{j+1})\right]\cdot \mathbf{1}_{k-2}^\top\right\} \\ &= \prod_{j=2}^{k}\frac{\pi_{s_{j-1}}(1-\pi_{s_{j-1}})\gamma_{s_j}\phi(s_{j-1}; t_j)}{\pi_{s_j}^2} \times \left\{\left[\prod_{j=2}^{k-1} M_j(s_j; t_{j+1})\right]\cdot \mathbf{1}_{k-1}^\top\right\}, \end{aligned} \tag{4.26}$$

as required. □

5. Partial detection of deaths

In this section, we prove Theorem 3.2 with piecewise constant birth and death rates between the detected deaths. The overall structure of the proof of Theorem 3.2 is similar to that of

Theorem 3.1, but Lemma 4.1 does not extend to the case where some deaths are not detected. It is straightforward to modify [9, Section 2] to show that, given that there is one individual at time t and that there has been no detection by time $t+\tau$, the number of individuals alive, $\tilde{Y}^{(t)}(\tau)$, at time $t+\tau$ is a shifted geometric random variable. Specifically, if $\tilde{E}_{t,\tau}$ is the event that there are no detections in the interval $[t, t+\tau]$, then there exist $a(\tau)$ and $b(\tau)$ such that $\mathbb{P}(\tilde{Y}^{(t)}(\tau)=0|\tilde{E}_{t,\tau})=a(\tau)$ and $\mathbb{P}(\tilde{Y}^{(t)}(\tau)=n+1|\tilde{Y}^{(t)}(\tau)\neq 0, \tilde{E}_{t,\tau})=(1-b(\tau))^n b(\tau)$ $(n=0, 1, \ldots)$. However, this approach does not yield explicit expressions for $a(\tau)$ and $b(\tau)$. It is more constructive to use probability generating functions and in particular [6], which gives the joint probability generating function of the number of individuals alive and the total number of deaths up to time t in a time-homogeneous birth–death process with one initial individual at time 0. The result analogous to Lemma 4.1 is given in Lemma 5.1. A consequence of Lemma 5.1 is that the number of individuals alive immediately following the first detected death satisfies Theorem 3.2(1), i.e. $\tilde{X}_1|\tilde{T}_1<\infty\sim\text{Geom}(\tilde{\pi}_1)$, where $\tilde{\pi}_1=\lambda_1$ is defined in (3.24). Then in Lemma 5.2, we prove results analogous to Lemmas 4.3 and 4.4, stating that if the number of individuals alive at time t follows a geometric distribution and there are no detected deaths in the interval $[t, t+\tau)$, then the number of individuals alive at time $t+\tau$ is a mixture of a geometric random variable and a point mass at 0, and if the first detected death after time t is at time $t+\tau$, then the number of individuals alive at time $t+\tau$ follows a negative binomial distribution with shape parameter 2. The proof of Lemma 5.2 utilises probability generating functions in a similar manner to the proof of Lemma 5.1. The remainder of the proof of Theorem 3.2 follows straightforwardly as Corollary 4.1 and Lemma 4.5 hold with minor modifications.

Before proving Theorem 3.2, we show how we can use [6] to combine the probability generating function of the number of individuals alive at time $\tau>0$ with the events of no deaths being detected in the interval $[0, \tau)$ and the first detected death after time 0 being at time τ. Let $\omega=(\alpha, \mu, d)$ denote a generic triple of birth rate, death rate, and detection probability, with $\tilde{\omega}_k=(\tilde{\alpha}_k, \tilde{\mu}_k, \tilde{d}_k)$ denoting the triple of birth rate, death rate, and detection probability between the $(k-1)$th and kth detections of death. For $\tau>0$, let $Y_\omega(\tau)$, $V_\omega(\tau)$, and $U_\omega(\tau)$ denote the number of individuals alive, the number of deaths, and the number of detected deaths, respectively, in a time-homogeneous birth–death process with parameters ω at time τ given a single initial individual at time 0. Let

$$H^E_\omega(\tau,\theta)=\mathbb{E}\Big[\theta^{Y_\omega(\tau)}1_{\{\tilde{T}>\tau\}}\Big]=\mathbb{E}\Big[\theta^{Y_\omega(\tau)}1_{\{U_\omega(\tau)=0\}}\Big],$$

where $\tilde{T}$ denotes the time of the first detected death after time t. Since $U_\omega(\tau)|V_\omega(\tau)\sim \text{Bin}(V_\omega(\tau), d)$, it follows that

$$H^E_\omega(\tau,\theta)=\mathbb{E}\Big[\theta^{Y_\omega(\tau)}(1-d)^{V_\omega(\tau)}\Big].$$

In line with the notation prior to Theorem 3.2, we set $p=\alpha/(\mu+\alpha)$, $q=\mu/(\mu+\alpha)$,

$$\begin{aligned}
\bar{u} &= \sqrt{1-4pq(1-d)}, & \bar{\lambda} &= \frac{1+\bar{u}-2p}{1+\bar{u}},\\
\bar{\nu} &= \frac{1-\bar{u}}{2p}, & \bar{\zeta} &= \frac{1+\bar{u}}{2p}=\frac{1}{1-\bar{\lambda}},\\
\bar{\phi}(\tau) &= \exp(-[\alpha+\mu]\bar{u}\tau)\,, & \bar{\psi}(\tau) &= \frac{(1-\bar{\lambda})(1-\bar{\phi}(\tau))}{1-\bar{\nu}(1-\bar{\lambda})\bar{\phi}(\tau)}.
\end{aligned}$$

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

Then, using the equation before (3) in [6], we have, in our notation and after a minor rearrangement, that

$$
\begin{aligned}
H^E_\omega(\tau,\theta) &= \frac{1}{2p} - \frac{\bar{u}}{2p}\left[\frac{\theta-\bar{\zeta}+(\theta-\bar{\nu})\bar{\phi}(\tau)}{\theta-\bar{\zeta}-(\theta-\bar{\nu})\bar{\phi}(\tau)}\right] \\
&= \frac{(\theta-\bar{\zeta})\bar{\nu}-(\theta-\bar{\nu})\bar{\zeta}\bar{\phi}(\tau)}{\theta-\bar{\zeta}-(\theta-\bar{\nu})\bar{\phi}(\tau)}. \qquad (5.1)
\end{aligned}
$$

Similarly, we can define

$$
\begin{aligned}
H^D_\omega(\tau,\theta) &= \mathbb{E}\Big[\theta^{W_\omega(\tau)}|\tilde{T}=\tau\Big]f_{\tilde{T}}(\tau) \\
&= \frac{\partial}{\partial\theta}H^E_\omega(\tau,\theta) \\
&= \frac{\bar{u}^2\bar{\phi}(\tau)}{p^2[\theta-\bar{\zeta}-(\theta-\bar{\nu})\bar{\phi}(\tau)]^2} = \frac{[1-4pq(1-d)]\bar{\phi}(\tau)}{p^2[\theta-\bar{\zeta}-(\theta-\bar{\nu})\bar{\phi}(\tau)]^2}. \qquad (5.2)
\end{aligned}
$$

Lemma 5.1. *Suppose that for* $t<0$, $(\beta_t,\gamma_t,\delta_t)=(\tilde{\alpha}_1,\tilde{\mu}_1,\tilde{d}_1)$. *Then*

$$\{Y(0)|\tilde{T}<\infty\}\sim \text{Geom}\,(\tilde{\pi}_1)\,. \qquad (5.3)$$

Proof. Given that a death is detected ($\tilde{T}<\infty$) and the parameters prior to the first detected death are $\tilde{\omega}_1$, we have that $\{Y(0)|\tilde{T}<\infty\}$ has probability generating function

$$\mathbb{E}\Big[\theta^{Y(0)}\Big|\,\tilde{T}<\infty\Big]=\mathbb{E}\Big[\,\theta^{W_{\tilde{\omega}_1}(\tilde{T})}\Big|\,\tilde{T}<\infty\Big]=\frac{\int_0^\infty H^D_{\tilde{\omega}_1}(\tau,\theta)\,d\tau}{\int_0^\infty H^D_{\tilde{\omega}_1}(\tau,1)\,d\tau}. \qquad (5.4)$$

It is then straightforward, applying a change of variable $y=\theta-\tilde{\zeta}_1-[\theta-\tilde{\nu}_1]\tilde{\phi}_1(\tau)$ in the integrals in (5.4), to show that

$$\mathbb{E}\Big[\theta^{Y(0)}\Big|\,\tilde{T}<\infty\Big]=\frac{\tilde{\pi}_1}{1-(1-\tilde{\pi}_1)\theta},$$

whence (5.3) follows. □

Lemma 5.2. *Suppose that we have a time-homogeneous linear birth–death process with parameters* $\omega=(\alpha,\mu,d)$. *For* $t\geq 0$, *let* $\bar{Y}_\omega(t)$ *denote the number of individuals alive at time t, with* $\bar{Y}_\omega(0)\sim\text{Geom}(\pi^*)$ *for some* $0<\pi^*<1$.

Let $\bar{E}_\tau$ *and* $\bar{D}_\tau$ *denote the events that there are no detected deaths on the interval* $[0,\tau)$ *and that the first detected death after time 0 is at time* τ, *respectively. Then*

$$\bar{Y}_\omega(\tau)|\bar{E}_\tau\sim\begin{cases}\text{Geom}(\check{\pi}_\tau) & \textit{with probability } \bar{h}(\tau)=\dfrac{1-\check{\pi}_\tau-\bar{\psi}(\tau)}{(1-\check{\pi}_\tau)(1-\bar{\psi}(\tau))}, \\[2ex] 0 & \textit{with probability } 1-\bar{h}(\tau)=\dfrac{\check{\pi}_\tau\bar{\psi}(\tau)}{(1-\check{\pi}_\tau)(1-\bar{\psi}(\tau))},\end{cases} \qquad (5.5)$$

and

$$\bar{Y}_\omega(\tau)|\bar{D}_\tau\sim\text{NegBin}(2,\check{\pi}_\tau), \qquad (5.6)$$

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

where

$$\check{\pi}_\tau = \frac{\bar{\lambda}[1-\bar{\nu}(1-\pi^*)] - (1-\bar{\nu})[\bar{\lambda}-\pi^*]\bar{\phi}(\tau)}{1-\bar{\nu}(1-\pi^*)+\bar{\nu}[\bar{\lambda}-\pi^*]\bar{\phi}(\tau)}. \tag{5.7}$$

Proof. To prove (5.5), we first note that the joint probability generating function of $\bar{Y}_\omega(\tau)$ and no death detected in the interval $[0, \tau)$ can be written, for $0 \le \theta \le 1$, as

$$\mathbb{E}\left[\theta^{\bar{Y}_\omega(\tau)} 1_{\{\bar{E}_\tau\}}\right] = \sum_{j=0}^{\infty} H_\omega^E(\tau, \theta)^j (1-\pi^*)^j \pi^*. \tag{5.8}$$

It follows from (5.1) and $\bar{\zeta} = 1/(1-\bar{\lambda})$ that

$$\begin{aligned}
&\sum_{j=0}^{\infty} \left\{H_\omega^E(\tau, \theta)(1-\pi^*)\right\}^j \\
&\quad = \frac{\theta - \bar{\zeta} - (\theta-\bar{\nu})\bar{\phi}(\tau)}{[\theta - \bar{\zeta} - (\theta-\bar{\nu})\bar{\phi}(\tau)] - (1-\pi^*)[(\theta-\bar{\zeta})\bar{\nu} - (\theta-\bar{\nu})\bar{\zeta}\bar{\phi}(\tau)]} \\
&\quad = \frac{1-\bar{\nu}(1-\bar{\lambda})\bar{\phi}(\tau) - (1-\bar{\lambda})(1-\bar{\phi}(\tau))\theta}{[1-\bar{\nu}(1-\pi^*)+\bar{\nu}(\bar{\lambda}-\pi^*)\bar{\phi}(\tau)] - [(1-\bar{\lambda})(1-\bar{\nu}(1-\pi^*)) + (\bar{\lambda}-\pi^*)\bar{\phi}(\tau)]\theta}.
\end{aligned} \tag{5.9}$$

Setting $\theta = 1$ in (5.9) and substituting into (5.8) yields

$$\mathbb{P}(\bar{E}_\tau) = \frac{\pi^*[\bar{\lambda} + (1-\bar{\lambda})(1-\bar{\nu})\bar{\phi}(\tau)]}{\bar{\lambda}[1-(1-\pi^*)\bar{\nu}] - (1-\bar{\nu})[\bar{\lambda}-\pi^*]\bar{\phi}(\tau)}. \tag{5.10}$$

Using (5.8)–(5.10) and the definitions of $\check{\pi}_t$ and $\bar{\psi}(\tau)$, we have that

$$\begin{aligned}
&\mathbb{E}\left[\theta^{\bar{Y}_\omega(\tau)} \middle| \bar{E}_\tau\right] \\
&\quad = \frac{\mathbb{E}\left[\theta^{\bar{Y}_\omega(\tau)} 1_{\{\bar{E}_\tau\}}\right]}{\mathbb{P}(\bar{E}_\tau)} \\
&\quad = \frac{\bar{\lambda}[1-(1-\pi^*)\bar{\nu}] - (1-\bar{\nu})[\bar{\lambda}-\pi^*]\bar{\phi}(\tau)}{\bar{\lambda} + (1-\bar{\lambda})(1-\bar{\nu})\bar{\phi}(\tau)} \\
&\qquad \times \frac{1-\bar{\nu}(1-\bar{\lambda})\bar{\phi}(\tau) - (1-\bar{\lambda})(1-\bar{\phi}(\tau))\theta}{[1-\bar{\nu}(1-\pi^*)+\bar{\nu}(\bar{\lambda}-\pi^*)\bar{\phi}(\tau)] - [(1-\bar{\lambda})(1-\bar{\nu}(1-\pi^*)) + (\bar{\lambda}-\pi^*)\bar{\phi}(\tau)]\theta} \\
&\quad = \frac{1-\bar{\nu}(1-\bar{\lambda})\bar{\phi}(\tau) - (1-\bar{\lambda})(1-\bar{\phi}(\tau))\theta}{\bar{\lambda} + (1-\bar{\lambda})(1-\bar{\nu})\bar{\phi}(\tau)} \times \frac{\check{\pi}_\tau}{1-[1-\check{\pi}_\tau]\theta} \\
&\quad = \frac{1-\bar{\psi}(\tau)\theta}{1-\bar{\psi}(\tau)} \times \frac{\check{\pi}_\tau}{1-[1-\check{\pi}_\tau]\theta} \\
&\quad = \frac{\check{\pi}_\tau \bar{\psi}(\tau)}{(1-\check{\pi}_\tau)(1-\bar{\psi}(\tau))} + \frac{1-\check{\pi}_\tau-\bar{\psi}(\tau)}{(1-\check{\pi}_\tau)(1-\bar{\psi}(\tau))} \times \frac{\check{\pi}_\tau}{1-[1-\check{\pi}_\tau]\theta}.
\end{aligned} \tag{5.11}$$

Then (5.5) follows immediately from (5.11).

Finally, let

$$L(\tau;\theta)=\frac{\partial}{\partial\theta}\mathbb{E}\Big[\,\theta^{Y_\omega(\tau)}\Big|\,\bar{E}_\tau\Big]$$
$$=\frac{\check{\pi}_\tau[1-\check{\pi}_\tau-\bar{\psi}(\tau)]}{[1-\bar{\psi}(\tau)][1-(1-\check{\pi}_\tau)\theta]^2}.$$

Then, given that $f_{\bar{T}}(\tau)=L(\tau,\,1)$, it is straightforward to show that for $0\le\theta\le 1$,

$$\mathbb{E}\Big[\,\theta^{\bar{Y}_\omega(\tau)}\Big|\,\bar{D}_\tau\Big]=\frac{L(\tau;\theta)}{L(\tau;1)}=\frac{\check{\pi}_\tau^2}{[1-(1-\check{\pi}_\tau)\theta]^2},$$

which yields (5.6) and completes the proof of Lemma 5.2. □

Proof of Theorem 3.2. Part 1 is proved in Lemma 5.1; we now provide the details of the proof of Part 2, the distribution of the number of individuals alive immediately following the kth detected death. The proof of Part 3, the distribution of the number of individuals alive at time t, can then be obtained along similar lines to those used in the proof of Theorem 3.1, using the distribution of the number of individuals alive at, and the time since, the most recent detected death.

Let $\tilde{R}_k$ $(k=1,\,2,\,\ldots)$ be defined to satisfy

$$\{\tilde{X}_k|\tilde{\mathbf{T}}_{2:k}\}\sim \text{NegBin}(\tilde{R}_k,\,\tilde{\pi}_k).$$

Then it follows from Lemma 5.2, using arguments virtually identical to those in the proof of Corollary 4.1, that for $k=2,\,3,\,\ldots$ and $j=1,\,2,\,\ldots$,

$$\{\tilde{R}_k|\tilde{R}_{k-1}=j,\,\tilde{T}_k=\tilde{t}_k\}\sim 2+\text{Bin}(j-1,\,\tilde{h}_k(\tilde{t}_k)),$$

where $\tilde{h}_k(\tilde{t}_k)$ is given in (3.27). It is then straightforward, following the proof of Lemma 4.5, to show that $\mathbb{P}(R_k=j|\tilde{\mathbf{T}}_{2:k})=\tilde{B}_{k,j}$, where $\tilde{\mathbf{B}}_k=(\tilde{B}_{k,2},\,\tilde{B}_{k,3},\,\ldots,\,\tilde{B}_{k,k})$ satisfies (3.26), completing the proof of Part 2. □

6. Process of infectives in general stochastic epidemic given removal times

In this section, we prove Theorem 3.3. We begin by outlining some results from the theory of aggregated continuous-time Markov chains from which the proof of Theorem 3.3 follows straightforwardly.

Let $\{W(t)\}=\{W(t):t\ge 0\}$ be a homogeneous continuous-time Markov chain having finite state space $E=\{1,2,\ldots,n\}$ and transition-rate matrix $\mathbf{Q}=[q_{ij}]$. Thus $q_{ij}=\lim_{t\downarrow 0}t^{-1}\mathbb{P}(W(t)=j|W(0)=i)$ $(i\ne j)$ and $q_{ii}=-\sum_{j\ne i}q_{ij}$. The state space is partitioned as $E=A\cup B$, where $A=\{1,2,\ldots,n_A\}$ and $B=\{n_A+1,\,n_A+2,\,\ldots,\,n\}$. Let $n_B=n-n_A$ denote the cardinality of B. Partition $\mathbf{Q}$ into

$$\mathbf{Q}=\begin{bmatrix}\mathbf{Q}_{AA} & \mathbf{Q}_{AB}\\ \mathbf{Q}_{BA} & \mathbf{Q}_{BB}\end{bmatrix},$$

in the obvious fashion.

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

Suppose that $W(0) \in A$, and let $T = \inf\{t > 0 : W(t) \in B\}$ be the time when $\{W(t)\}$ first visits B. Suppose that $\mathbb{P}(T < \infty | W(0) = i) = 1$ for all $i \in A$. Let $\mathbf{F}(t) = [F_{ij}(t)]$ be defined elementwise by

$$F_{ij}(t) = \mathbb{P}(T \le t \text{ and } W(T) = n_A + j | W(0) = i) \qquad (t \ge 0; i = 1, 2, \ldots, n_A; j = 1, 2, \ldots, n_B),$$

and let $\mathbf{f}(t) = \mathbf{F}'(t)$, where the differentiation is elementwise. Then (see for example [4])

$$\mathbf{f}(t) = \exp(\mathbf{Q}_{AA}t)\mathbf{Q}_{AB} \qquad (t \ge 0). \tag{6.1}$$

Further, let $\mathbf{P}_{AB}$ be the $n_A \times n_B$ matrix with (i, j)th element given by $\mathbb{P}(W(T) = n_A + j | W(0) = i)$. Then

$$\mathbf{P}_{AB} = \int_0^\infty \mathbf{f}(t)\, \mathrm{d}t = -\mathbf{Q}_{AA}^{-1}\mathbf{Q}_{AB}. \tag{6.2}$$

Note that $\mathbf{Q}_{AA}$ is nonsingular since A is a transient class, so by [1, p. 77], all eigenvalues of $\mathbf{Q}_{AA}$ have strictly negative real parts.

Proof of Theorem 3.3. For $k = 2, 3, \ldots, N+1$ and $i = 0, 1, \ldots, N+1-k$, let $f_{k,i}(\mathbf{t}_{2:k})$ be the probability density of the event $\{\mathbf{T}_{2:k} = \mathbf{t}_{2:k} \text{ and } I(S_k) = i\}$, and let

$$\mathbf{f}_k(\mathbf{t}_{2:k}) = (f_{k,0}(\mathbf{t}_{2:k}), f_{k,1}(\mathbf{t}_{2:k}), \ldots, f_{k,N+1-k}(\mathbf{t}_{2:k})).$$

Exploiting the conditional independence along the sample paths of $\{(S(t), I(t))\}$, application of (6.2) followed by repeated application of (6.1) yields

$$\mathbf{f}_k(\mathbf{t}_{2:k}) = -\mathbf{u}_1\mathbf{Q}_{0,0}^{-1}\mathbf{Q}_{0,1}\left(\prod_{i=1}^{k-2}\exp(\mathbf{Q}_{i,i}t_{i+1})\mathbf{Q}_{i,i+1}\right)\exp(\mathbf{Q}_{k-1,k-1}t_k)\tilde{\mathbf{Q}}_{k-1,k}, \tag{6.3}$$

where $\mathbf{u}_1$ is the row vector of length N whose first element is 1 and all of whose other elements are 0, and the product is given by the identity matrix $\mathbf{I}_{N-1}$ when $k = 2$.

Let

$$\mathbf{w}_k(\mathbf{t}_{2:k}) = (w_{k,0}(\mathbf{t}_{2:k}), w_{k,1}(\mathbf{t}_{2:k}), \ldots, w_{k,N-k}(\mathbf{t}_{2:k})),$$

where $w_{k,i}(\mathbf{t}_{2:k}) = \mathbb{P}(I(S_k) = i | \mathbf{T}_{2:k} = \mathbf{t}_{2:k})$. Then

$$\mathbf{w}_k(\mathbf{t}_{2:k}) = \mathbf{f}_k(\mathbf{t}_{2:k})/(\mathbf{f}_k(\mathbf{t}_{2:k}) \cdot \mathbf{1}_{N+1-k}^\top).$$

Then, since $\tilde{\mathbf{Q}}_{k,k}$ is the transition-rate matrix for transitions of $\{(S(t), I(t))\}$ within $\tilde{\Omega}_k$, for $0 \le \tau < t_{k+1}$,

$$\mathbf{v}_k(\tau | \mathbf{t}_{2:k}) = \mathbf{w}_k(\mathbf{t}_{2:k}) \exp(\tilde{\mathbf{Q}}_{k,k}\tau).$$

Using (6.3), we have that (3.31) and (3.32) follow immediately. □

7. Numerical results

In this section, we illustrate briefly the practical usefulness of the results of this paper.

We simulated a time-homogeneous linear birth–death process with $\beta_t = 1.5$ and $\gamma_t = 1.0$ up to the 200th death and estimated the size of the population over time based upon (partial)

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

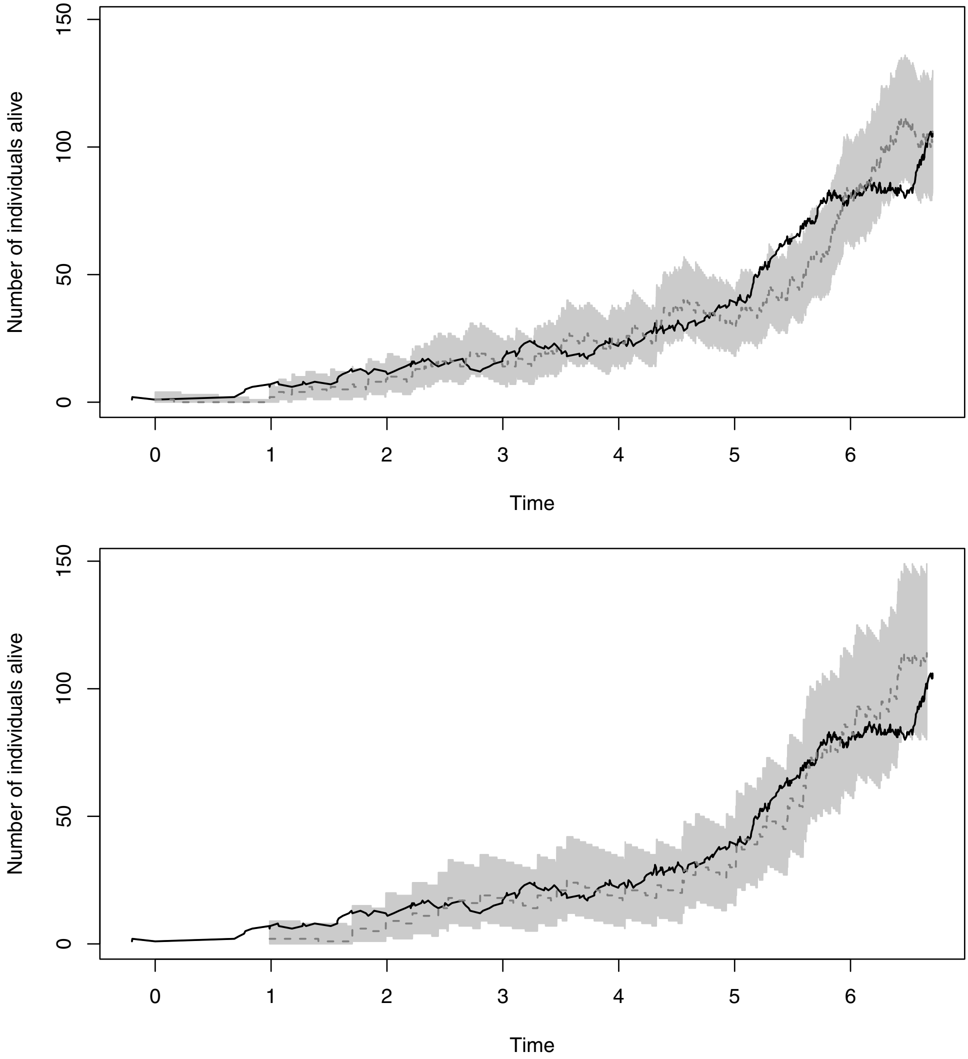

FIGURE 1. Number of individuals alive (solid line) and median of $Y(t)|\mathbf{t}_{2:k_t}$ (dashed line) up to the 200th death, with $\beta_t = 1.5$ and $\gamma_t = 1.0$. The shaded area represents the probability mass between the 10% and 90% quantiles of $Y(t)$. Top: $d = 1$. Bottom: $d = 0.25$.

observation of the death process. We considered the cases where the detection probability of a death was $d = 1$ (all deaths are detected) and $d = 0.25$. In Figure 1, we plot the number of individuals alive against time, along with the median of $Y(t)$ calculated using Theorem 3.1 ($d = 1$) and Theorem 3.2 ($d = 0.25$). The plot also includes the 10% and 90% quantiles of $Y(t)$, $l(t)$, and $h(t)$, respectively, with $[l(t), h(t)]$ shaded for all $t \geq 0$. We set time $t = 0$ to be the time of the first death and note that for the case $d = 0.25$ the first detected death is not until $t = 0.9875$, at which point the estimation of the number of individuals alive starts. The estimation of $Y(t)$ is similar in both cases, although for $d = 0.25$ the loss of information from detecting only a quarter of the deaths is observed in a larger quantile range for $Y(t)$.

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

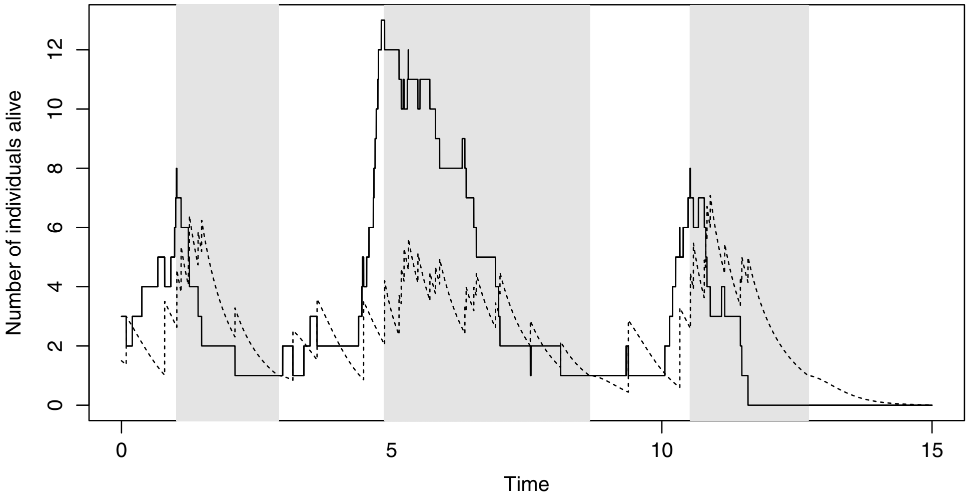

FIGURE 2. Number of individuals alive (solid line) and $\mathbb{E}[Y(t)|\mathbf{t}_{2:k_t}]$ (dashed line) for a population with control measures and $\gamma = 1$. Parameters: $\alpha_N = 1.5$ (no control measures), $\alpha_C = 0.5$ (control measures), $\vartheta_U = 4$ (control measures introduced), $\vartheta_L = 1$ (control measures lifted). Shaded area denotes control measures in place.

We demonstrate the use of Theorem 3.1 for implementing control measures. Suppose that the birth rate is $\alpha_N > 1$ when there are no control measures and $\alpha_C < 1$ when control measures are in place. Let $\gamma_t \equiv 1$, so that the birth–death process is super-critical in the absence of control measures and sub-critical when control measures are in place. We assume that initially the population evolves without control measures until an upper threshold, ϑ_U, is hit, at which point control measures are introduced. The population remains in control measures until a lower threshold, ϑ_L, is reached, at which point control measures are removed. We assume that the population can enter and leave control measures multiple times. We consider control measures based upon $\mathbb{E}[Y(t)|\mathbf{t}_{2:k_t}]$, which can be implemented in real time, although alternatives based on the median of $Y(t)$ or the probability of extinction in the absence of control measures could easily be used. We have that control measures are introduced for the first time at $u_1 = \min_t\{\mathbb{E}[Y(t)|\mathbf{t}_{2:k_t}] \geq \vartheta_U\}$ and are lifted for the first time at $l_1 = \min_{t>u_1}\{\mathbb{E}[Y(t)|\mathbf{t}_{2:k_t}] \leq \vartheta_L\}$. Note that it follows from Theorem 3.1 that $\mathbb{E}[Y(t)|\mathbf{t}_{2:k_t}]$ jumps up at death times and decreases continuously between death times. Therefore the introduction of control measures immediately follows a death, with subsequent lifting when no death has occurred for a sufficiently long time interval. An illustration of a simulation with control measures is given in Figure 2, in which $\alpha_N = 1.5$, $\alpha_C = 0.5$, $\vartheta_U = 4$, and $\vartheta_L = 1$. In the example, three episodes of control measures are required. A plot of π_t given by the ODE in (3.2) is presented in Figure 3, which shows rapid changes after the introduction and removal of control measures.

Finally, we apply the birth–death process calculations to an epidemic outbreak. We consider two different birth–death process approximations of the general stochastic epidemic model. At the kth removal times we compare the mean number of individuals alive in the birth–death process approximation calculated using Theorem 3.1 with the corresponding mean number of infectives in the general stochastic epidemic model calculated using Theorem 3.3.

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

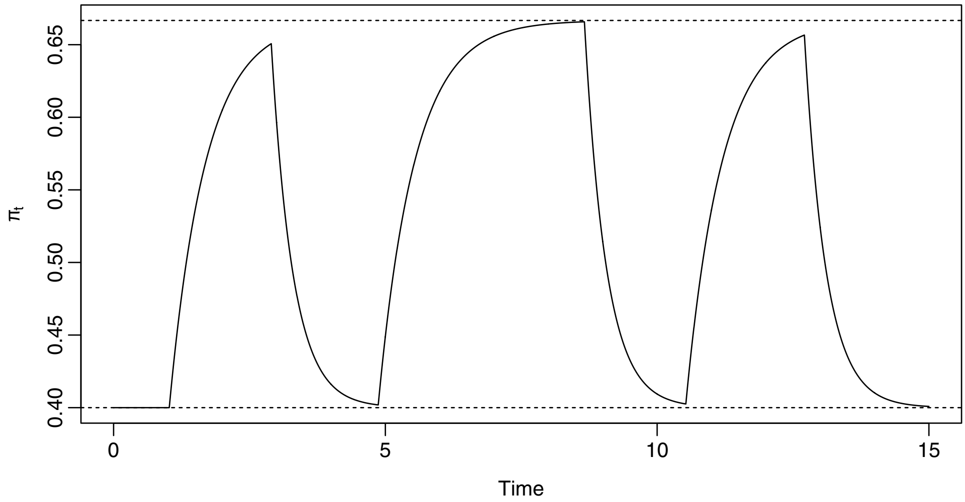

FIGURE 3. Plot of π_t against time for the population with control measures in Figure 2. Parameters: $\alpha_N = 1.5$ (no control measures), $\alpha_C = 0.5$ (control measures), $\gamma = 1$ (death rate), $\vartheta_U = 4$ (control measures introduced), $\vartheta_L = 1$ (control measures lifted). Horizontal dashed lines at $q_C = 1/(1 + \alpha_C) = 2/3$ (upper) and $q_N = 1/(1 + \alpha_N) = 2/5$ (lower) correspond to the probability that an event is a death under a constant regime of control measures (upper) and no control measures (lower).

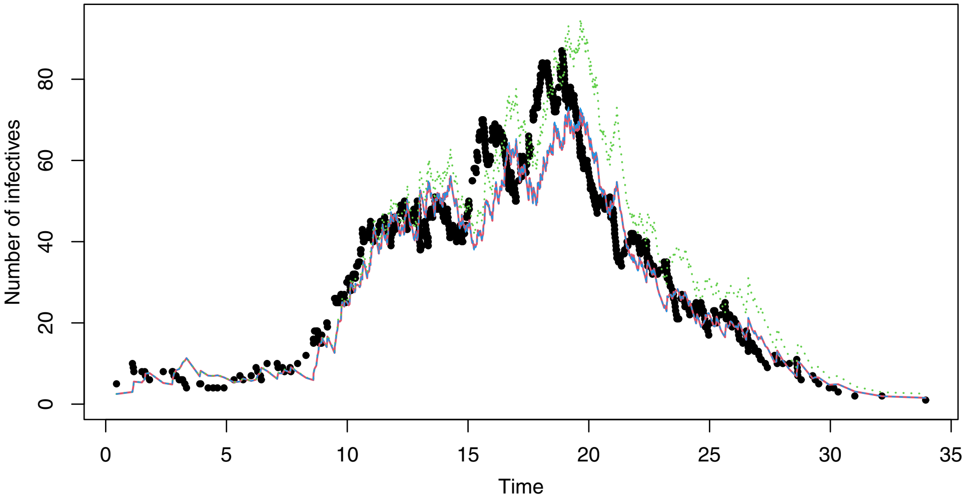

FIGURE 4. Plot of the number of infectives at removal times against time (black dots) for an epidemic of final size 829 in a population of size $N = 2000$, generated with $\beta_t(= \alpha) = 1.25$ and $\gamma_t(= \mu) = 1$. The mean number of infectives estimated using only removal times is given using the general stochastic epidemic model (red line), time-homogeneous birth–death process (green, dotted line), and time-inhomogeneous birth–death process with non-susceptibles accounted for (blue, dashed line).

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

In Figure 4, an epidemic outbreak with final size 829 in a population of size $N = 2000$ generated with $\beta_t(=\alpha) = 1.25$ and $\gamma_t(=\mu) = 1$ is plotted, with the mean number of infectives (individuals alive) after each removal (death) calculated using Theorem 3.3 (Theorem 3.1). The two birth–death process approximations we consider are the time-homogeneous birth–death process with $\alpha = 1.25$ and $\mu = 1$, which does not take account of the removal of infectives, and the time-inhomogeneous birth–death process with piecewise constant birth rate $\alpha_k = 1.25(N - \{k - 1 + \mathbb{E}[Y(s_{k-1})|\mathbf{T}_{2:k-1} = t_{2:k-1}]\})/N$ between the $(k-1)$th and kth removal and death rate $\mu = 1$, which reduces the birth rate by the estimated proportion of the population who are not susceptible ($k - 1$ removed individuals and $\mathbb{E}[Y(s_{k-1})|\mathbf{T}_{2:k-1} = t_{2:k-1}]$, the estimated number of infectives). We observe that both approximations give a reasonable estimate of the mean number of infectives, especially in the early stages of the epidemic, when the birth–death process approximation is most appropriate. By taking account of the number of non-susceptible individuals we obtain a very good approximation in which the expected number of infectives differs by at most 0.9157 from that given by the general stochastic epidemic model. It should be noted that the computation of the distribution of the birth–death approximation is much faster than that of the general stochastic epidemic model, as the former involves vector–matrix multiplication while the latter involves matrix exponentials and matrix multiplication.

8. Concluding remarks

Explicit formulae for the distribution of the number of infectives (individuals alive) in a general stochastic epidemic (branching process), given only partial information, have the potential to assist with many areas of disease management. Firstly, from a statistical perspective, we are able to calculate the likelihood of the observed removal times (Corollary 3.1) without the need for computationally intensive data augmentation (cf. [13, 15]). This allows rapid computation of the likelihood, enabling the use of likelihood-based statistical methods to maximise the likelihood or obtain estimates from the posterior distribution of the parameters, and allowing for greater understanding as we do not need to integrate over augmented data. Moreover, the ability to include time-varying parameters allows the estimation of infection rates before and after control measures are introduced. We will present a summary of statistical methodology, including partial observations of the death (removal) process, elsewhere. Secondly, from a public health perspective, we can easily obtain epidemic quantities of interest, such as the mean number of infectives or the probability that the epidemic is, or will go, extinct. This enables the introduction and lifting of control measures in a scientifically informed manner, extending this work beyond the numerical illustrations given in Section 7.

In this paper we have focused on the Markovian SIR epidemic model and its approximating branching (birth–death) process. We can extend the model to allow for a more general infectious period (lifetime) distribution. It is straightforward using [9] to show that the number of individuals alive immediately after the first death in a branching process, where individuals have a constant birth rate and independent and identically distributed lifetimes, follows a geometric distribution. The distribution of residual lifetime of individuals alive at the first death time means that the arguments used in Section 4 do not extend beyond exponential lifetime distributions. Progress can be made for phase-type lifetime distributions, and we will show elsewhere that the number of individuals alive in the approximating branching process at time $t \geq 0$ can be expressed as a sum of k_t (the total number of observed deaths/removals up to and including time t) independent random variables. Each of the random variables in the sum satisfies one of the three distributions, based on (a) a geometric random variable, (b) a mixture

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press

of a geometric random variable and a point mass at 0, and (c) a sum of an independent geometric random variable and a Bernoulli random variable. Note that for birth–death processes considered in this paper, only the random variables (a) and (b) feature in the distribution of the number of individuals alive at a given point in time.

Acknowledgement

We would like to thank the associate editor and two reviewers for their helpful comments, which have significantly improved the presentation of the paper.

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.

References

[1] Asmussen, S. (1987). *Applied Probability and Queues*. John Wiley, Chichester.

[2] Ball, F. G. (1983). The threshold behaviour of epidemic models. *J. Appl. Prob.* **20**, 227–241.

[3] Ball, F. G. and Donnelly, P. (1995). Strong approximations for epidemic models. *Stoch. Process. Appl.* **55**, 1–21.

[4] Colquhoun, D. and Hawkes, A. G. (1977). Relaxation and fluctuations of membrane currents through drug-operated channels.. *Proc. R. Soc. London B* **199**, 231–262.

[5] Gani, J. and Purdue, P. (1984). Matrix-geometric methods for the general stochastic epidemic. *IMA J. Math. Appl. Med. Biol.* **1**, 333–342.

[6] Gani, J. and Swift, R. J. (2017). Distribution of deaths in a birth–death process. *Math. Scientist* **42**, 111–114.

[7] Hunter, J. J. (1969). On the moments of Markov renewal processes. *Adv. Appl. Prob.* **1**, 188–210.

[8] Kendall, D. G. (1948). On the generalized 'birth-and-death' process. *Ann. Math. Statist.* **19**, 1–15.

[9] Lambert, A. and Trapman, P. (2013). Splitting trees stopped when the first clock rings and Vervaat's transformation. *J. Appl. Prob.* **50**, 208–227.

[10] Lefèvre, C. and Picard, P. (2017) On the outcome of epidemics with detections. *J. Appl. Prob.* **54**, 890–904.

[11] Lefèvre, C., Picard, P. and Utev, S. (2020). On branching models with alarm triggerings. *J. Appl. Prob.* **57**, 734–759.

[12] Lefèvre, C. and Simon, M. (2020). SIR-type epidemic models as block-structured Markov processes. *Methodology Comput. Appl. Prob.* **22**, 433–453.

[13] Neal, P. J. and Roberts, G. O. (2005). A case study in non-centering for data augmentation: stochastic epidemics. *Statist. Comput.* **15**, 315–327.

[14] O'Neill, P. D. (1997). An epidemic model with removal-dependent infection rate. *Ann. Appl. Prob.* **7**, 90–109.

[15] O'Neill, P. D. and Roberts, G. O. (1999). Bayesian inference for partially observed stochastic epidemics. *J. R. Statist. Soc. A* **162**, 121–129.

[16] Pyke, R. (1961). Markov renewal processes: definitions and preliminary properties. *Ann. Math. Statist.* **32**, 1231–1242.

[17] Shapiro, L. W. (1976). A Catalan triangle. *Discrete Math.* **14**, 83–90.

[18] 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.

[19] Whittle, P. (1955). The outcome of a stochastic epidemic—a note on Bailey's paper. *Biometrika* **42**, 116–122.

https://doi.org/10.1017/apr.2022.58 Published online by Cambridge University Press