Mathematical Biology

Partially observed branching processes for stochastic epidemics

Victor M. Panaretos

Abstract   At the offset of a (stochastic) epidemic, it is of importance to have a mathematical model that will assist in the making of an informed judgement on whether the epidemic will *explode*, or will be minor and *die out*. In this paper, we consider probabilistic inferences related to the event of extinction of a discrete time branching process when this cannot be directly observed. Instead, we are able to observe only a random "trace" of the process, which not only *trails* the latter, but also directly affects it (in terms of interventions). A simple model is proposed that provides tractability, preserves a marginal branching property, and gives reasonable closed form expressions.

Keywords   Partial observation · Extinction/explosion · Joint Markov property · Conditional inference · Family name problem

Mathematics Subject Classification (2000)   60J80 · 60J85 · 92D30

1 Introduction

Suppose that the observation of a process describing an epidemic is imperfect. As a result, the observed process differs from the true process according to

V. M. Panaretos (✉)
Department of Statistics, University of California - Berkeley, 367 Evans Hall,
Berkeley, CA 94720, USA
e-mail: victor@stat.berkeley.edu

some random mechanism. The epidemic is assumed to be in its initial stages and may die out with only a few cases or may explode. An objective is to specify when emergency control rules are justified. Such controls can be thought of as actions that can decisively affect the epidemic, but have a serious cost to be avoided if unnecessary (e.g. culling in animal disease, or vaccination/quarantining in human disease, as in the foot and mouth and SARS epidemics, or, most recently, in the bird flu incidents). The decision to implement such control rules will have to take into account the expected future evolution of the epidemic process, in the context of the assumed model, and also the observations on the process up to the present time.

More specifically, provided that a suitable Markovian model is chosen, and its basic parameters are known (perhaps from previous experience), the current number N of infected individuals is at any point predictively sufficient for the future of the process. Further, the probability $\mathbb{P}[\mathcal{E}_0]$ that the epidemic becomes extinct without special intervention can be determined, so that given suitable tests, a possible decision rule for intervention is to do so only if $\mathbb{P}[\mathcal{E}_\infty]/\mathbb{P}[\mathcal{E}_0] > c$ (with $\mathcal{E}_0, \mathcal{E}_\infty$ denoting the events of extinction and of explosion, respectively). Of course, such a decision rule only takes account of the general specifications of the process, and does not use information carried by the current value of the process. If we denote the process by $\{Z_n\}_{n\geq 0}$, then it is desirable to use the probabilities $\mathbb{P}[\mathcal{E}_0|Z_n = N] = 1 - \mathbb{P}[\mathcal{E}_\infty|Z_n = N]$. The reason we condition only on the event $\{Z_n = N\}$ obviously stems from the Markovian property.

However, in order to be more realistic, we have to accept that the values we observe are not the realisations of the actual process $\{Z_n\}$, but only a random subset of them (a point already raised by Kendall [5]). The reality is that not all infected individuals are detected (at least not immediately). We therefore have to admit that observation is imperfect, something that makes conditioning on the event $\{Z_n = N\}$ inappropriate. What we need to condition on is the value of the *observable* process. Of course, it should be noted that while strongly motivated by practical considerations concerning real epidemics, in the absence of very specific details, it is hard to proceed beyond general results, and thus somewhat precarious to make more elaborate suggestions to the scientist making the decisions. Therefore, the approach we adopt is purposely fairly general, in order to allow for flexibility in specific modelling situations. Our main aim is thus to describe the probabilistic dynamics involved in a certain class of models incorporating the element of imperfect observation, and to subsequently set the scene for a number of further problems.

With this decision process in mind, the remainder of this paper deals with the investigation of the interplay between the properties of the observable and the true processes when observation is partial. As it might be expected, the observable process will not necessarily be of a Markovian nature. However, it is important to note that it is "trailing" a Markov process (the true process, $\{Z_n\}$) something that is of interest in itself. We work in the setup of Galton–Watson branching processes.

Branching processes include discretely observed birth–death processes, and have been used extensively to model the evolution of the initial stages of epidemics in discrete time (see [3,5]; also see [1]). In the standard setting, on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, define $\{\zeta_{j,n}\}_{j \geq 1, n \geq 0}$ to be an infinite array of independent and identically distributed nonnegative integer-valued random variables, and N_0 to be a nonnegative integer-valued random variable, independent of the $\zeta_{j,n}$. If X_n is the number of infected individuals at time n, then we describe its evolution in terms of the following branching process

$$X_0 = N_0, \quad X_{n+1} = \sum_{j=1}^{X_n} \zeta_{j,n}, \quad n \geq 1, \tag{1}$$

where an empty sum is taken to be zero. The random variable $\zeta_{j,n}$ corresponds to the number of infections caused by the jth infectious individual at time n. Using the standard terminology, these variables are the offsprings, and their common law $p(k) = \mathbb{P}[\zeta_{j,n} = k]$, is called the offspring distribution.

We consider an often more realistic situation, in which at any time point, it is not possible to observe all infections, but only a random subset of them. In addition, we do not passively witness the evolution of the epidemic, but we intervene in a way depending on our observations. The general issue in such a situation is what kind of *probabilistic inference* we may draw on the behavior of the true process of infectious individuals based on its observable counterpart (for an approach to *statistical inference* see [7]; also see Sect. 7).

To this aim, we propose a model in discrete time, $\mathbb{N} \cup \{0\}$, for which the intuitive idea is as follows. A number (possibly random) of infected individuals Z_0 is inserted among an infinite population of susceptibles at time $n = 0$. Each of these Z_0 infected individuals is overlooked with probability θ, or observed with probability $1 - \theta$, independently for all individuals. If Y_0 is the number of individuals we observe, then $U_0 = Z_0 - Y_0$ individuals remain unobserved. Each of the U_0 unobserved infected individuals independently produces a number of infections, according to the same probability mass function, say Ξ. On the other hand, the observed individuals are isolated at some point between time 0 and time 1, so that they do not produce infections according to the same law. In particular, we assume that each observed individual independently produces a number of infections, according to a probability mass function Υ, where we assume that $F_\Upsilon \geq F_\Xi$ (with F denoting the corresponding cumulative distribution). The sum of the new infections caused by observed and unobserved individuals constitutes Z_1, i.e. the number of infectious individuals at time $n = 1$. The process then continues iteratively, as just described. Thus, the observation procedure does not simply provide us with a "trace" of the initial process, but has a direct impact on the actual process itself, for infected individuals that have been observed present a different behaviour as far as offsprings are concerned.

Therefore, at each time point $n \geq 0$, there is a number Z_n of infected individuals. There is a probability $1 - \theta$ of observing each one of the infected individuals, independently for each individual, so that at time n, $Y_n | Z_n = z_n$

is binomially distributed with parameters z_n and $1 - \theta$ (i.e. we have binomial thinning). Finally, the new infections are given by,

$$Z_{n+1} \overset{d}{=} \sum_{j=1}^{U_n} \xi_{j,n} + \sum_{j=1}^{Y_n} \zeta_{j,n}, \tag{2}$$

where the $\xi_{j,n}$ and $\zeta_{j,n}$ are "independent copies" from the distributions Ξ and Υ, respectively.

2 Definition of the partially observed branching process

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let N_0 be a fixed (or possibly random) non-negative integer. Let $\{\xi_{i,j}\}_{i \geq 1, j \geq 0}$ and $\{\zeta_{i,j}\}_{i \geq 1, j \geq 0}$ be independent infinite arrays of iid nonnegative integer-valued random variables, with common probability mass function (pmf) Ξ and Υ, respectively. Let $\{B_{i,j}\}_{i \geq 1, j \geq 0}$ be an infinite array of iid Bernoulli random variables on the same space, with probability of success $1 - \theta$, where $\theta \in (0, 1)$. Let N_0, $\{\xi_{i,j}\}_{i \geq 1, j \geq 0}$, $\{\zeta_{i,j}\}_{i \geq 1, j \geq 0}$ and $\{B_{i,j}\}_{i \geq 1, j \geq 0}$ be independent. We define a stochastic process $\{(Z_n, Y_n)\}_{n \geq 0}$ as follows:

$$\begin{cases} Z_0 = N_0, \\ Z_{n+1} = \sum_{i=1}^{Z_n} \xi_{i,n}(1 - B_{i,n}) + \sum_{i=1}^{Z_n} \zeta_{i,n} B_{i,n}, & n \geq 0 \\ Y_n := \sum_{i=1}^{Z_n} B_{i,n}, & n \geq 0 \end{cases} \tag{3}$$

An empty sum is zero. We also put $U_n := Z_n - Y_n$, for $n \geq 0$. As noted in the introduction, Z_n represents the infectious individuals at time n, Y_n the observed infected individuals at time n, and U_n the unobserved infected individuals at time n. We call $\xi_{i,j}$ the offsprings of the unobserved infected individual i at time j, and thus Ξ corresponds to the pmf of the offspring distribution (similarly for the observed individuals). In the sequel, we denote the respective probability generating functions (pgf) by G_Ξ and G_Υ.

It is immediate from the iterative definition of Z_n, that the process $\{Z_n\}_{n \geq 1}$ of infectious individuals at time n is a Galton–Watson branching process, with an offspring distribution that is a mixture of Ξ and Υ, with mixing proportions $1 - \theta$ and θ, respectively (set $\psi_{i,j} := \xi_{i,j}(1 - B_{i,j}) + \zeta_{i,j} B_{i,j}$). We shall denote the common pmf of the $\{\psi_{i,j}\}$ by $\Psi(\cdot)$, and the corresponding pgf by G_Ψ.

Our attention will be focused on the interplay between the Z_n and Y_n processes: their joint behaviour, as well as the conditional behaviour of the true process Z_n which is unobservable given observations on its observable counterpart, Y_n.

In order to study this interplay, we first consider a simpler model than model (3). The intuitive idea is to break down the analysis of this process into

separately analysing the contributions made from unobserved and observed individuals, respectively. In mathematical terms, "separately" should translate into taking advantage of a lurking stochastic independence:

Lemma 1 *For any $n \geq 0$, the contributions made by the unobserved individuals are conditionally independent of the contributions made by the observed individuals, given the number of observed individuals, i.e.*

$$\sum_{i=1}^{Z_n} \xi_{i,n}(1 - B_{i,n}) \bigg| Y_n \quad \coprod \quad \sum_{i=1}^{Z_n} \zeta_{i,n} B_{i,n} \bigg| Y_n \tag{4}$$

Proof This is immediate from the assumptions on model (3).

Since the observed individuals are known at any point in time, any probability measure of interest is conditioned on Y_n. Therefore, we may proceed in our analysis by first studying in depth the extreme scenario in which observed individuals produce *no offsprings at all* (that is, Υ is such that $\mathbb{P}[\zeta_{i,j} = 0] = 1$). The results can then be extended to cover the full model, by our conditional independence argument (Lemma 1), modifying the results according to the appropriate convolutions of the probability measures involved.

3 Analysis of a simplified model

According to the approach set out in the previous paragraph, we begin by studying the extreme case in which observed individuals are unable to produce offsprings. We thus modify model (3) so that

$$Z_{n+1} = \sum_{i=1}^{Z_n} \xi_{i,n}(1 - B_{i,n}) \stackrel{d}{=} \sum_{i=1}^{Z_n - Y_n} \xi'_{i,n} \tag{5}$$

It is this formulation that we shall follow throughout Sects. 3, 4, and 5. The results are extended to the general case [Eq. (3)] in Sect. 6.

Remark 1 In this simplified situation, one may note a simple analogy to the original "family name problem" of branching processes (see [6]). Suppose that we are interested in the survival of a family name. At each time point $n \geq 1$, the family size Z_n, consists of the sum of Y_n females and U_n males. We assume that the society is patriarchical, so only the males produce offsprings that hold the family name. If for each generation we only observe the number of female family members, what can we infer about the survival of the family name? In the family name context, each time point represents a generation: if someone is part of generation n, they are obviously not part of generation $n + 1$. Nevertheless, in the epidemic context, we make no such assumption. An individual that is infected at time n may continue to be so at the next time point (as an "offspring" of himself).

Interestingly, in this simplified model, it can be seen that the process of unobserved individuals $\{U_n\}$ admits a weak representation as a Galton–Watson branching process:

Proposition 1 *There exists a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ on which we may define a Galton–Watson branching process $\{U'_n\}_{n \geq 0}$, such that $(U'_{n_1}, \ldots, U'_{n_k}) \overset{d}{=} (U_{n_1}, \ldots, U_{n_k})$, for any finite collection of non-negative integers $\{n_1, \ldots, n_k\}$.*

Proof It suffices to show that the law of $\{U_n\}_{n \geq 0}$ is that of a branching process started at $U_0 = Z_0 - Y_0$. This follows immediately upon noting that

$$U_{n+1} = Z_{n+1} - Y_{n+1} \overset{d}{=} \sum_{i=1}^{U_n} \xi'_{i,n} - \sum_{i=1}^{U_n} \sum_{k=1}^{\xi'_{i,n}} B'_{k,n} = \sum_{i=1}^{U_n} \left\{ \xi'_{i,n} - \sum_{k=1}^{\xi'_{i,n}} B'_{k,n} \right\}, \quad (6)$$

where $\{\xi'_{i,j}\}_{i \geq 1, j \geq 0}$ is an infinite array of iid-$\varXi$ random variables on the space $(\Omega', \mathcal{F}', \mathbb{P}')$, and $\{B'_{ij}\}_{i \geq 1, j \geq 0}$ is an infinite array of iid Bernoulli$(1 - \theta)$ random variables on the same space, independent of $\{\xi'_{i,j}\}_{i \geq 1, j \geq 0}$.

Although $\{U_n\}$ admits the branching property, and hence has a straightforward behaviour, it is not of particular interest, since, in any practical situation, it is, by definition, unobservable.

Now, the marginal process of infected individuals $\{Z_n\}_{n \geq 0}$ is a Galton–Watson branching process, and as such, it is a discrete time Markov chain with respect to the filtration $\{\mathcal{Z}_n\}_{n \geq 0} = \{\sigma(Z_0, \ldots, Z_n)\}_{n \geq 0}$. On the other hand, knowledge of the history of $\{Y_n\}_{n \geq 0}$ up to any order, is predictively insufficient in the Markov sense for $\{Y_n\}_{n \geq 0}$ in itself. Such an insufficiency is inherent from the very definition of the process $\{Y_n\}_{n \geq 0}$. The explicit dependence of Y_n on Z_n suggests the consideration of the joint process $\{(Z_n, Y_n)\}_{n \geq 0}$ as a process on the first quadrant of the lattice $\mathbb{Z}^2$. By the iterative nature of the process, it is straightforward to see that the joint process constitutes a Markov chain.

3.1 One-step transition probabilities

To derive an explicit form for the one step transition probabilities, we notice that by definition, Y_{n+1} is conditionally independent both of Z_n and of Y_n, given Z_{n+1} (this is clear by the definition of Y_{n+1}). This enables us to obtain a particularly simple closed form for one-step transition probabilities. Let the k-fold convolution of $\varXi$ with itself be $\varXi_k^*$. Then,

$$\begin{aligned}
&\mathbb{P}[Z_1 = z_1, Y_1 = y_1 | Z_0 = z_0, Y_0 = y_0] \\
&\quad = \mathbb{P}[Z_1 = z_1 | Z_0 = z_0, Y_0 = y_0] \, \mathbb{P}[Y_1 = y_1 | Z_1 = z_1, Z_0 = z_0, Y_0 = y_0] \\
&\quad = \varXi_{z_0 - y_0}^*(z_1) \binom{z_1}{y_1} (1 - \theta)^{y_1} \theta^{z_1 - y_1}.
\end{aligned}$$

If one is to use standard offspring distributions such as Ξ being Poisson with parameter λ or Geometric with parameter ν, then one obtains a Poisson with parameter $\lambda(z_0 - y_0)$ or a Negative Binomial with parameters ν and $z_0 - y_0$, respectively, for $\Xi_{z_0-y_0}^*$.

Accordingly, the generating function for the one-step transition probabilities assumes a very convenient form. Let the n-step transition probability generating function be

$$P_n(r,s) = \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} r^a s^b \mathbb{P}[Z_n = a, Y_n = b | Z_0 = z_0, Y_0 = y_0]. \qquad (7)$$

Using the explicit expression derived for the one step transition probability, we may write

$$P_1(r,s) = \sum_{a=0}^{\infty} r^a \Xi_{z_0-y_0}^*(a) \sum_{b=0}^{\infty} \binom{a}{b} [s(1-\theta)]^b \theta^{a-b} = \sum_{a=0}^{\infty} r^a [\theta + s(1-\theta)]^a \Xi_{z_0-y_0}^*(a)$$

$$= \{G_\Xi[rs(1-\theta) + r\theta]\}^{z_0-y_0}.$$

Notice that, were it not for the "quarantine assumption", the joint process $\{(Z_n, Y_n)\}$ would constitute an "infinite" hidden Markov process. However, the removal of the observed infected individuals from the population implies that

$$\text{dist}\{Z_{n+1}, Y_{n+1} | Z_n, Y_n\} \neq \text{dist}\{Z_{n+1}, Y_{n+1} | Z_n\}.$$

3.2 Higher order transition probabilities

In the case of higher order transitions, it is a non-trivial task to obtain the exact form for the conditional probability mass function. However, it is feasible to obtain the corresponding probability generating function.

Theorem 1 (*n-step transitions*) *Let $P_n(r,s)$ be the probability generating function corresponding to the mass function $\mathbb{P}[Z_n = \cdot, Y_n = \cdot | Z_0 = z_0, Y_0 = y_0]$. Then,*

$$P_n(r,s) = \{G_\Xi[\gamma_{n-1}(r\theta + rs(1-\theta))]\}^{z_0-y_0}$$

$$= \left\{ \frac{1}{\theta} \gamma_n(r\theta + rs(1-\theta)) - \frac{1-\theta}{\theta} \right\}^{z_0-y_0},$$

where $\gamma_k(\cdot)$ stands for the probability generating function of the kth generation of a Galton–Watson branching process started from a single individual and evolving as $\{Z_n\}$ evolves marginally.

Proof As in the case of the one-step transitions, we may write

$$\mathbb{P}[Z_n = z_n, Y_n = y_n | Z_0 = z_0, Y_0 = y_0]$$
$$= \mathbb{P}[Z_n = z_n | Z_0 = z_0, Y_0 = y_0] \, \mathbb{P}[Y_n = y_n | Z_n = z_n, Z_0 = z_0, Y_0 = y_0].$$

Conditionally on Z_n, Y_n is independent of both Z_0, and Y_0, so that

$$\mathbb{P}[Y_n = y_n | Z_n = z_n, Z_0 = z_0, Y_0 = y_0] = \mathbb{P}[Y_n = y_n | Z_n = z_n] = \mathcal{B}_{(1-\theta, z_n)}(y_n),$$

where $\mathcal{B}_{(1-\theta, z_n)}$ is the binomial probability mass function, with parameters z_n and $1-\theta$. Furthermore, the definition of the $\{Z_n\}$ process, along with the Markov property yield the following expression:

$$\mathbb{P}[Z_n = z_n | Z_0 = z_0, Y_0 = y_0]$$
$$= \sum_{z_1=0}^{\infty} \mathbb{P}[Z_n = z_n | Z_1 = z_1, Z_0 = z_0, Y_0 = y_0] \, \mathbb{P}[Z_1 = z_1 | Z_0 = z_0, Y_0 = y_0]$$
$$= \sum_{z_1=0}^{\infty} \mathbb{P}[Z_n = z_n | Z_1 = z_1] \Xi^*_{z_0-y_0}(z_1).$$

The above observations allow us to determine the generating function for the n-step transition probabilities. We have

$$P_n(r,s) = \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} r^a s^b \mathbb{P}[Z_n = a, Y_n = b | Z_0 = z_0, Y_0 = y_0]$$
$$= \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} r^a s^b \mathbb{P}[Z_n = a | Z_0 = z_0, Y_0 = y_0] \mathcal{B}_{(1-\theta, a)}[b]$$
$$= \sum_{a=0}^{\infty} r^a \mathbb{P}[Z_n = a | Z_0 = z_0, Y_0 = y_0] \sum_{b=0}^{\infty} \binom{a}{b} [s(1-\theta)]^b \theta^{a-b}$$
$$= \sum_{a=0}^{\infty} [r\theta + rs(1-\theta)]^a \mathbb{P}[Z_n = a | Z_0 = z_0, Y_0 = y_0]$$
$$= \sum_{a=0}^{\infty} [r\theta + rs(1-\theta)]^a \sum_{z_1=0}^{\infty} \mathbb{P}[Z_n = a | Z_1 = z_1] \Xi^*_{z_0-y_0}(z_1)$$
$$= \sum_{z_1=0}^{\infty} \Xi^*_{z_0-y_0}(z_1) \sum_{a=0}^{\infty} [r\theta + rs(1-\theta)]^a \mathbb{P}[Z_n = a | Z_1 = z_1].$$

Now let $\gamma_k(\cdot)$ be as in the assumptions of the theorem. We will have

$$
\begin{aligned}
P_n(r,s) &= \sum_{z_1=0}^{\infty} \Xi^*_{z_0-y_0}(z_1) \sum_{a=0}^{\infty} [r\theta + rs(1-\theta)]^a \mathbb{P}[Z_n = a|Z_1 = z_1] \\
&= \sum_{z_1=0}^{\infty} \Xi^*_{z_0-y_0}(z_1)[\gamma_{n-1}(r\theta + rs(1-\theta))]^{z_1} \\
&= \{G_\Xi[\gamma_{n-1}(r\theta + rs(1-\theta))]\}^{z_0-y_0}.
\end{aligned}
$$

Finally, the second form of the generating function is a consequence of the fact that the distribution Ψ is a modified version of Ξ, so that

$$
G_\Psi(s) = \theta G_\Xi(s) + (1-\theta).
$$

This completes the proof.

3.3 Predictions based on the observable component

Since at any point in time n it is only the history $\mathcal{Y}_n = \sigma(Y_1, \ldots, Y_n)$ of Y that is available for any stochastic inference concerning the evolution of the true epidemic, we may wish to assign probabilities to events concerning the future of the true process, conditioning on $\mathcal{Y}_n$. In general however, the probability measure $\mathbb{P}[\cdot|Y_0 = y_0, \ldots, Y_n = y_n]$ will not induce distributions that may be represented in a simple closed form. For example, consider the distribution $\mathbb{P}[Z_{n+1} = z_{n+1}|Y_0 = y_0, \ldots, Y_{n-k} = y_{n-k}]$. Assume that the epidemic begins with $Z_0 = 1$ infections, If one conditions on the past of Z_n in order to take advantage of the Markov property of $\{Z_n, Y_n\}$, then one has

$$
\begin{aligned}
&\mathbb{P}[Z_{n+1} = z_{n+1}|Y_0 = y_0, \ldots, Y_{n-k} = y_{n-k}] \\
&= \frac{\sum_{z_n=0}^{\infty} \Xi^*_{z_n-y_n}(z_{n+1}) w(z_n, y_{n-k}, \ldots, y_n)}{\sum_{z_n=0}^{\infty} w(z_n, y_{n-k}, \ldots, y_n)},
\end{aligned}
$$

where

$$
\begin{aligned}
&w(z_n, y_{n-k}, \ldots, y_n) \\
&= \sum_{z_n=0}^{\infty} \cdots \sum_{z_{n-k}=0}^{\infty} \left\{ \prod_{m=n-k}^{n-1} \Xi^*_{z_m-y_m}(z_{m+1}) \binom{z_{m+1}}{y_{m+1}} \left(\frac{1-\theta}{\theta}\right)^{y_{m+1}} \theta^{z_{m+1}} \right\} \\
&\quad \times \binom{z_{n-k}}{y_{n-k}} (1-\theta)^{y_{n-k}} \theta^{z_{n-k}-y_{n-k}} \frac{\gamma_{n-k}^{(z_{n-k})}(0)}{z_{n-k}!}
\end{aligned}
$$

are non-negative weights depending on the history of the process Y. If one does not assume that Ξ (and hence Ψ) belongs to a specific family of distributions,

then one may hardly proceed fruitfully with the above form. It is intuitively obvious that conditioning beyond some point in the past history of Y will not contribute appreciable additional information. However, the complicated form of the above distribution does not allow one to quantify such a statement (e.g. in terms of the Kullback–Leibler divergence between two measures conditioned on different lengths of the history of Y). We shall see in the next subsections that it is possible to get simple closed form expressions when conditioning only on the present value of the observable process, rather than on its whole history.

3.4 One step predictions

What may we infer about the number of infected individuals tomorrow based on the number we observe today? Alternatively, in the family name context, suppose that we are told that the number of female family members at some generation n is y_n. What can one infer about the family size at generation $n+1$? We therefore wish to determine the distribution $\mathbb{P}[Z_{n+1} = z_{n+1} | Y_n = y_n]$. Throughout the remainder of this section, we assume that $Z_0 = N_0$ and $Y_0 = 0$. That is, we assume that $\mathbb{P}$ is in fact $\mathbb{P}_\mu$, with the initial distribution μ, that assigns unit mass to the event $\{Z_0 = N_0, Y_0 = 0\}$. The choice of μ as initial distribution is not restrictive. For suppose that at some time (which we call time $n = 1$) we observe a number of infections for the first time. Then, without loss of generality, we may assume that the epidemic had begun from time $n = 0$ (the previous time point) from a (possibly random) number of individuals, all of whom were (by definition) unobserved.

We will proceed as before, using generating functions. Recall that by definition we have

$$Z_{n+1} \overset{d}{=} \sum_{i=1}^{U_n} \xi_{i,n+1} \implies Z_{n+1} | \{Y_n = y_n\} \overset{d}{=} \sum_{i=1}^{Z_n - y_n} \xi_{i,n+1} \Bigg| \{Y_n = y_n\}.$$

Hence, using standard notation for the generating functions, we will have

$$
\begin{aligned}
G_{Z_{n+1}|Y_n=y_n}(s) &= G_{\xi_{1,n}+\cdots+\xi_{Z_n-y_n,n}|Y_n=y_n}(s) \\
&= \mathbb{E}[s^{\xi_{1,n}+\cdots+\xi_{Z_n-y_n,n}}|Y_n = y_n] \\
&= G_{Z_n-y_n|Y_n=y_n}(G_\mathcal{E}(s)) \\
&= \mathbb{E}[\{G_\mathcal{E}(s)\}^{Z_n-y_n}|Y_n = y_n] \\
&= \frac{\mathbb{E}[\{G_\mathcal{E}(s)\}^{Z_n}|Y_n = y_n]}{\{G_\mathcal{E}(s)\}^{y_n}} \\
\implies G_{Z_{n+1}|Y_n=y_n}(s) &= \frac{G_{Z_n|Y_n=y_n}(G_\mathcal{E}(s))}{\{G_\mathcal{E}(s)\}^{y_n}}
\end{aligned}
\tag{8}
$$

This reduces the problem to the specification of the generating function of the conditional distribution of Z_n given Y_n. We have,

Proposition 2 *Suppose that the initial distribution of $\{Z_n, Y_n\}$ is μ. Then, the generating function for the conditional probability distribution of Z_n given Y_n is given by,*

$$G_{Z_n|Y_n=y_n}(r) = \frac{r^{y_n} \lambda_n^{(y_n)}(r\theta)}{\lambda_n^{(y_n)}(\theta)},$$

where $\lambda_k(\cdot) = G_{N_0}\{G_\Xi[\gamma_{k-1}(\cdot)]\}$ is the generating function of the $(k-1)$th generation of a Galton–Watson branching process started with initial distribution $\mathrm{dist}\{\sum_{i=1}^{N_0} \xi_{i,0}\}$ and evolving as $\{Z_n\}$ evolves marginally.

Proof Manipulating the definition of a conditional pgf, results in

$$G_{Z_n|Y_n=y}(r) = \left\{ \left.\frac{\partial^y}{\partial s^y} P_n(r,s) \right|_{(r,0)} \right\} \Big/ \left\{ \left.\frac{\partial^y}{\partial s^y} P_n(r,s) \right|_{(1,0)} \right\}$$

When the initial distribution is μ,

$$
\begin{aligned}
P_n(r,s) &= \sum_{x=0}^{\infty} \sum_{y=0}^{\infty} r^x s^y \mathbb{P}_\mu[Z_n = x, Y_n = y] \\
&= \sum_{x=0}^{\infty} r^x \mathbb{P}_\mu[Z_n = x] \sum_{y=0}^{\infty} s^y \mathbb{P}_\mu[Y_n = y | Z_n = x] \\
&= \lambda_n(rs(1-\theta) + r\theta),
\end{aligned}
$$

Therefore,

$$\frac{\partial^y}{\partial s^y} P_n(r,s) = [r^y(1-\theta)^y] \lambda_n^{(y)}(rs(1-\theta) + r\theta),$$

and the result follows by substitution. $\qquad\blacksquare$

Corollary 1 *Suppose that the initial distribution of $\{Z_n, Y_n\}$ is μ. The generating function for the one-step prediction distribution may be expressed as:*

$$G_{Z_{n+1}|Y_n=y_n}(r) = \frac{\lambda_n^{(y_n)}(\theta G_\Xi(r))}{\lambda_n^{(y_n)}(\theta)}.$$

3.5 Higher order predictions

We now address the determination of the distribution $\mathbb{Q}_k(z|y) := \mathbb{P}_\mu[Z_{n+k} = z | Y_n = y]$, i.e. the k-step ahead "prediction" distribution. First, we recall a

fundamental yet straightforward "self-similarity" property of the Galton–Watson branching process. Thinking about the process as a random tree, if one is to isolate any branch of the tree and consider it as a tree of its own, then this is again a Galton–Watson tree (process).

The actual process $\{Z_n\}$ is unobservable, and at time n, we do not know the number of truly infected individuals. Had we known them, then the distribution of Z_{n+k} would just be that of the nth generation of a Galton–Watson branching process started from Z_n individuals. All that is known is the value of Y_n, and we use this as a basis for prediction.

An idea is that the probability measure $\mathbb{P}_\mu[Z_{n+k} = z|Y_n = y]$ should be "similar" to the measure $\mathbb{P}_{\mu_n(y)}[Z_k = z]$, where $\mu_n(y) := \mathrm{dist}\{Z_n|Y_n = y\}$. Intuitively, this is saying that probabilistically, the information provided by Y_n allows us to think about the k-step ahead prediction distribution, as the distribution of a branching process at time k with initial distribution $\mathrm{dist}\{Z_n|Y_n = y\}$. However, one must be careful in that the value of Y_n contains information about the thinning that occurred in the step from time n to time $n + 1$. Thus we cannot regard the situation as exactly one of a branching process that has started with initial distribution $\mu_n(y)$. We overcome this by conditioning on the first step. The random variables $\{Z_m\}_{m=n+2}^{\infty}$ are conditionally independent of Y_n given Z_{n+1}. Hence, if we treat the first step carefully, we can then reduce the problem to that of determining the $(k-1)$th step distribution of a branching process that evolves as $\{Z_n\}$ with initial distribution $\mathrm{dist}\{Z_{n+1}|Y_n = y\}$ (the latter distribution has been determined in the previous paragraph). This can be expressed easily through the composition of corresponding probability generating functions, leading to the following result:

Theorem 2 *Let μ be the initial distribution of $\{Z_n, Y_n\}$. The generating function for the k-step prediction distribution, $\mathbb{Q}_k(z|y) = \mathbb{P}_\mu[Z_{n+k} = z|Y_n = y]$, is given by*

$$G_\mathbb{Q}(r) = G_{Z_{n+1}|Y_n=y}(\gamma_{k-1}(r)) = \frac{\lambda_n^{(y)}(\theta G_\Xi(\gamma_{k-1}(r)))}{\lambda_n^{(y)}(\theta)}$$

where $G_{Z_{n+1}|Y_n=y}(\cdot)$ is the generating function of the distribution $\mathrm{dist}\{Z_{n+1}| Y_n = y\}$, and $\lambda_m(\cdot), \gamma_m(\cdot)$ are as before.

Proof We begin by noticing that

$$\mathbb{Q}_k(z_{n+k}|y) = \mathbb{P}_\mu[Z_{n+k} = z_{n+k}|Y_n = y]$$

$$= \sum_{z_{n+1}=0}^{\infty} \mathbb{P}_\mu[Z_{n+k} = z_{n+k}|Z_{n+1} = z_{n+1}, Y_n = y]\mathbb{P}_\mu[Z_{n+1} = z_{n+1}|Y_n = y]$$

so that,

$$
\begin{aligned}
G_{\mathbb{Q}}(r) &= \sum_{z_{n+k}=0}^{\infty} \sum_{z_{n+1}=0}^{\infty} r^{z_{n+k}} \mathbb{P}_{\mu}[Z_{n+k} = z_{n+k} | Z_{n+1} = z_{n+1}, Y_n = y] \\
&\quad \times \mathbb{P}_{\mu}[Z_{n+1} = z_{n+1} | Y_n = y] \\
&= \sum_{z_{n+1}=0}^{\infty} \mathbb{P}_{\mu}[Z_{n+1} = z_{n+1} | Y_n = y] \\
&\quad \times \sum_{z_{n+k}=0}^{\infty} r^{z_{n+k}} \mathbb{P}_{\mu}[Z_{n+k} = z_{n+k} | Z_{n+1} = z_{n+1}, Y_n = y]
\end{aligned}
$$

and by conditional independence,

$$
\begin{aligned}
G_{\mathbb{Q}}(r) &= \sum_{z_{n+1}=0}^{\infty} \mathbb{P}_{\mu}[Z_{n+1}=z_{n+1}|Y_n=y] \sum_{z_{n+k}=0}^{\infty} r^{z_{n+k}} \mathbb{P}_{\mu}[Z_{n+k}=z_{n+k}|Z_{n+1}=z_{n+1}] \\
&= \sum_{z_{n+1}=0}^{\infty} \mathbb{P}_{\mu}[Z_{n+1} = z_{n+1} | Y_n = y] (\gamma_{k-1}(r))^{z_{n+1}} \\
&= G_{Z_{n+1}|Y_n=y}(\gamma_{k-1}(r)).
\end{aligned}
$$

The result follows from Corollary 1.

Remark 2 Although writing down an expression for the prediction distribution when conditioning on the whole history of $\{Y_n\}$ is not feasible without making specific distributional assumptions, we may obtain relationships that enable us to obtain an iterative scheme that allows one to obtain expressions when making such specific assumptions. To this aim, notice that mimicking the steps leading to Eq. (8), we may write

$$
G_{Z_{n+1}|\{Y_n=y_n,\dots,Y_1=y_1\}}(s) = \frac{G_{Z_n|\{Y_n=y_n,\dots,Y_1=y_1\}}(G_{\Xi}(s))}{\{G_{\Xi}(s)\}^{y_n}} \tag{9}
$$

On the other hand, we may apply Bayes' rule to obtain a connection between dist$\{Z_n|Y_n,\dots,Y_1\}$ and dist$\{Z_n|Y_{n-1},\dots,Y_1\}$:

$$
\begin{aligned}
&\mathbb{P}[Z_n = z_n | Y_n = y_n, \dots, Y_1 = y_1] \\
&\quad = \frac{\mathbb{P}[Y_n = y_n | Z_n = z_n] \mathbb{P}[Z_n = z_n | Y_{n-1} = y_{n-1}, \dots, Y_1 = y_1]}{\sum_{z_n=0}^{\infty} \mathbb{P}[Y_n = y_n | Z_n = z_n] \mathbb{P}[Z_n = z_n | Y_{n-1} = y_{n-1}, \dots, Y_1 = y_1]} \\
&\implies G_{Z_n=z_n|\{Y_n=y_n,\dots,Y_1=y_1\}}(s) \\
&\quad = \frac{\sum_{z_n=0}^{\infty} s^{z_n} \binom{z_n}{y_n} (1-\theta)^{y_n} \theta^{z_n-y_n} \mathbb{P}[Z_n = z_n | Y_{n-1} = y_{n-1}, \dots, Y_1 = y_1]}{\sum_{z_n=0}^{\infty} \binom{z_n}{y_n} (1-\theta)^{y_n} \theta^{z_n-y_n} \mathbb{P}[Z_n = z_n | Y_{n-1} = y_{n-1}, \dots, Y_1 = y_1]}
\end{aligned}
$$

$$\implies G_{Z_n=z_n|\{Y_n=y_n,\ldots,Y_1=y_1\}}(s) = \frac{G^{(y_n)}_{Z_n=z_n|\{Y_{n-1}=y_{n-1},\ldots,Y_1=y_1\}}(s\theta)}{G^{(y_n)}_{Z_n=z_n|\{Y_{n-1}=y_{n-1},\ldots,Y_1=y_1\}}(\theta)} \quad (10)$$

Starting out from the first time points of the process we can thus bootstrap ourselves upward in time by iteratively switching between Eqs. (9) and (10), according to the following pattern:

$$G_{Z_2|Y_1} \overset{(10)}{\to} G_{Z_2|Y_2,Y_1} \overset{(9)}{\to} G_{Z_3|Y_2,Y_1} \overset{(10)}{\to} G_{Z_3|Y_3,Y_2,Y_1} \overset{(9)}{\to} G_{Z_4|Y_3,Y_2,Y_1} \overset{(10)}{\to} \cdots$$

One may then exploit the connection between the one-step ahead prediction distribution and the *k*-step ahead prediction distribution to obtain the probability generating function of the latter. Of course, this sort of iterative scheme requires "heavy" calculations, since it involves complicated compositions of high order derivatives. However, given a specific form for the offspring distributions, it is feasible (though considerably tedious) to obtain an exact expression. Moreover, as discussed earlier, conditioning beyond some point in the past is not expected to yield considerable additional information, so that one would not have to go very far back. In addition, since it is the initial phase of the epidemic that we are modelling, one can usually expect relatively small values for the $\{Y_n\}$ process (in fact this should be the most interesting scenario mathematically).

4 Questions related to extinction in the simplified model

Naturally, in the context of a partially observed stochastic epidemic, one wishes to make inferences related to the event of extinction of the actual process, based on the information provided by its observable counterpart. Such inferences could be the basis for making a decision on adopting emergency control measures or not. We address this issue by considering the conditional probability of extinction and the conditional distribution of the time to extinction, given the present value of the observable process. In this section, we deal with the simplified model (5), obtaining results that will serve as the basis for the results given in Sect. 6 for the general model (3).

4.1 The event of extinction

We consider the question of determining the probability of extinction. Let $\mathcal{E}_0$ denote the event of extinction

$$\mathcal{E}_0 := \bigcup_{n\geq 0}\{Z_n = 0\}.$$

In the unconditional case, the matter is settled by the martingale property of $Z_n/(\theta\rho)^n$, where $\rho = \mathbb{E}\xi$. Hence, in the critical and subcritical cases $\theta\rho \leq 1$,

extinction occurs almost surely, 'exponentially fast' (in the sense that $\mathbb{P}[X_n > 0] \leq \mathbb{E}[X_n] = \theta^n \rho^n$). In the case $\theta\rho > 1$ we recall the following fact (e.g. [2]):

Theorem 3 *Let $\varphi(\cdot)$ be the offspring distribution generating function, for a Galton–Watson process $\{X_n\}_{n\geq 1}$ started from a single individual. If the mean $m = \varphi'(1)$ of the offspring distribution is such that $m > 1$, then there exists a unique fixed point s_0 of $\varphi(\cdot)$ that is less than 1, and satisfies*

$$s_0 = \mathbb{P}[\mathcal{E}_0 | X_0 = 1].$$

Obviously, the observation of the process $\{Y_n\}$ provides information about the possibility of extinction or explosion, which we wish to incorporate in our "prediction", through the probability distribution $\mathbb{P}[\mathcal{E}_0 | Y_n = y]$. In order to specify this distribution, we use the fact that each individual of a certain generation of a Galton–Watson process gives rise to a new independent Galton–Watson process with the same probabilistic properties. Thus, at time n, the existence of $Y_n = y$ observed individuals, implies that there is an (unknown) random number of unobserved individuals $U_n = Z_n - Y_n$, that will give rise to U_n new independent Galton–Watson processes, each with the same law, namely that of $\{Z_n\}$. Each of these U_n Galton–Watson processes dies independently with probability s_0, where s_0 is the unique point that satisfies $G_\Psi(s) = s$ and $s \neq 1$.

Assuming that the initial distribution of the process is μ, we may follow an analysis taking special care at the first step (from time n to time $n+1$) to see that

$$\mathbb{P}_\mu[\mathcal{E}_0 | Y_n = y] = \sum_{z=0}^{\infty} \mathbb{P}_\mu[\mathcal{E}_0 | Z_{n+1} = z, Y_n = y] \mathbb{P}_\mu[Z_{n+1} = z | Y_n = y]$$

$$= \sum_{z=0}^{\infty} (s_0)^z \mathbb{P}_\mu[Z_{n+1} = z | Y_n = y]$$

$$= G_{Z_{n+1} | Y_n = y}(s_0)$$

$$= \frac{\lambda_n^{(y)}(\theta G_\Xi(s_0))}{\lambda_n^{(y)}(\theta)}.$$

Recapitulating, the problem of conditional extinction is trivial in the subcritical and critical cases, and in the case $\theta\rho > 1$, the probability of extinction given the present value of the observable process has been seen to have a simple expression, $\mathbb{P}_\mu[\mathcal{E}_0 | Y_n = y] = G_{Z_{n+1} | Y_n = y}(s_0)$.

4.2 Time to extinction

Consider the problem of predicting the time to extinction on the basis of information on the process Y_n at the present time. Let T_0 denote the time to extinction,

$$T_0 := \inf\{n \geq 0 : Z_n = 0\}.$$

We wish to determine the distribution

$$\tau_n(k|y) := \mathbb{P}_\mu[T_0 \leq n + k | Y_n = y].$$

We first note that

$$\{Z_k = 0\} = \{T_0 \leq k\} \quad \text{a.s.}$$

Therefore, the problem of determining the distribution of the time to extinction is a special case of that of making n-step "predictions", in the sense that it suffices to derive the probability distribution $\mathbb{Q}_k(z|y) = \mathbb{P}_\mu[Z_{n+k} = z|Y_n = y]$. If $G_\mathbb{Q}$ is the generating function corresponding to this probability distribution (suppressing dependence on k and y for tidiness), then

$$\mathbb{P}_\mu[T_0 \leq n + k|Y_n = y] = \mathbb{Q}_k(0|y) = G_\mathbb{Q}(0) = \frac{\lambda_n^{(y)}(\theta G_\Xi(\gamma_{k-1}(0)))}{\lambda_n^{(y_n)}(\theta)}.$$

5 An interesting special case

In this section, we consider an instance of the simplified model (5), which is seen to have certain attractive properties. As seen in the previous paragraphs, the determination of n-step transition probabilities as well as prediction distributions depends on the marginal distribution of Z_n. In general, the family of these distributions changes over time, making their determination impractical (although theoretically feasible). Nevertheless, there is a choice of initial distribution and offspring distribution Ξ, such that the marginal distributions of both Z_n and Y_n remain invariant in time, in the sense that they remain within the same family, with different defining parameters.

This special distribution, as might be anticipated, is the Geometric distribution. Suppose that Ξ is the Geometric pmf with probability of success p and mean $\rho =: (1 - p)/p$. In addition, suppose either that $Z_0 = 1$ or that Z_0 has a Geometric distribution. Recalling the definition of Z_1,

$$Z_1 = \sum_{i=1}^{Z_0} \psi_{i,0} = \sum_{i=1}^{Z_0} \xi_{i,0}(1 - B_{i,0}),$$

we see that Z_1 has a modified Geometric distribution (i.e. the result of taking a geometric distribution and shifting an amount of mass at zero, while re-normalizing the mass at the positive integers to maintain unit total mass). In the current section, we assume for simplicity that $Z_0 = 1$ (so $Y_0 = B_{1,0}$), inducing an

initial distribution $\mu_0 = \text{dist}\{Z_0, Y_0\}$. Since a modified Geometric stopped sum of iid modified Geometric random variables, has itself a modified Geometric distribution, then by the iterative definition of Z_n,

$$Z_n = \sum_{i=1}^{Z_{n-1}} \xi_{i,n-1}(1 - B_{i,n-1}),$$

we see that Z_n has a modified Geometric distribution for every $n \geq 1$. In fact, the generating function at time n, admits a simple representation:

$$G_{Z_n}(r) = 1 - v_n + v_n r \frac{p_n}{1 - (1 - p_n)r}, \tag{11}$$

with

$$v_n = (\rho\theta)^n \frac{1 - s_0}{(\rho\theta)^n - s_0}, \quad p_n = \frac{1 - s_0}{(\rho\theta)^n - s_0}, \quad \text{and} \quad s_0 = \frac{1 - \theta q}{q},$$

the latter being the non-negative root of the equation $s = 1 - \theta + (\theta p)/(1 - (1 - p)s)$, different from 1.

What is particularly attractive with such a formulation, is that Y_n is also distributionally temporally invariant (up to defining parameters) and has the modified Geometric distribution.

Proposition 3 *Let Ξ be a geometric distribution with mean $\rho = (1 - p)/p$. Let $\{Z_n, Y_n\}$ have initial distribution μ_0 and evolve as before. Then, for all $n \geq 1$, the marginal distributions induced by μ_0 for both Z_n and Y_n are of a modified Geometric type.*

Proof This is immediate from the fact that an integer-valued random variable X has the modified Geometric distribution if and only if its probability generating function admits the representation (e.g. [4]):

$$G_X(s) = \frac{\alpha + \beta s}{\alpha' + \beta' s}, \quad \alpha\beta' \neq \alpha'\beta.$$

Equation (11) shows that the pgf of the marginal distribution of Z_n admits such a representation and, by binomial thinning, we have that

$$G_{Y_n}(s) = G_{Z_n}(s(1 - \theta) + \theta)$$
$$= 1 - v_n + v_n \frac{p_n\theta + (1 - \theta)p_n s}{1 - q_n\theta - q_n(1 - \theta)s},$$

where v_n and $p_n = 1 - q_n$ are as before.

Corollary 2 *The marginal probability mass function of Y_n induced by μ_0 is given by,*

$$\mathbb{P}_{\mu_0}[Y_n = y] = \begin{cases} 1 - \nu_n + \nu_n \dfrac{p_n\theta}{1 - q_n\theta} & \text{if } y = 0; \\[3mm] \nu_n p_n q_n^{y-1} \dfrac{(1-\theta)^y}{(1-q_n\theta)^{y+1}} & \text{if } y > 0. \end{cases}$$

5.1 Exact expressions when branching is geometric

Throughout this section, we assume that $\varXi$ is the Geometric distribution with mean $\rho =: (1-p)/p$. The temporal distributional invariance (up to families) and the corresponding closed form expression for the generating function of the process of infected individuals provide us with simple closed forms for the distributions discussed in the previous section. In this paragraph we provide the explicit expressions.

5.1.1 Transition probabilities

The generating function for the transition probabilities of the process $\{(Z_n, Y_n)\}$ can be determined by substitution, from the general form obtained in Theorem 1. This will be

$$P_n(r,s) = \left\{ G_{\varXi} \left[G_{Z_{n-1}}(r\theta + rs(1-\theta)) \right] \right\}^{z_0 - y_0}$$

$$= p^{z_0 - y_0} \left[1 - q \left(1 - \nu_n + \frac{p_n \nu_n (r\theta + rs(1-\theta))}{1 - (1-p_n)(r\theta + rs(1-\theta))} \right) \right]^{y_0 - z_0}.$$

In what follows, we additionally assume that $Z_0 = 1$ (i.e. that the initial distribution of the joint process is μ_0, as earlier).

5.1.2 Predictions based on the observable component

First we determine the probability generating function of the distribution $\mathbb{P}[Z_n = z | Y_n = y]$, which was seen to play a major role in our analysis:

$$G_{Z_n|Y_n=y}(r) = \begin{cases} \dfrac{1 - \nu_n + \nu_n p_n r\theta/(1 - q_n r\theta)}{1 - \nu_n + \nu_n p_n \theta/(1 - q_n\theta)} & \text{if } y = 0; \\[4mm] r^y \left(\dfrac{1 - q_n\theta}{1 - q_n\theta r} \right)^{y+1} & \text{if } y > 0. \end{cases}$$

The second branch of the distribution can be seen to be a y-shifted negative binomial distribution, with success probability $1 - q_n\theta$ and $y + 1$ number of successes.

For the generating function of the k-step ahead prediction distribution we distinguish two cases.

In the case $y = 0$, we have

$$G_{Z_{n+k}|Y_n=0}(r)$$
$$= c_n \left(1 - v_n + v_n \frac{\theta p p_n (1 - q_{k-1} r)}{p + q v_{k-1} - q_n \theta p - (p q_{k-1} + q v_{k-1} - \theta p q_n q_{k-1}) r} \right),$$

whereas if $y > 0$,

$$G_{Z_{n+k}|Y_n=y}(r) = (1 - q_n \theta)^{y+1} \left(1 - \frac{q_n \theta p (1 - q_{k-1} r)}{p + q v_{k-1} - (p q_{k-1} + q v_{k-1}) r} \right)^{-y-1},$$

where

$$c_n = \left[1 - v_n + v_n p_n \theta / (1 - q_n \theta) \right]^{-1}$$

and $v_n, p_n = 1 - q_n$ are as before.

5.1.3 Time to extinction and probability of extinction

The distribution $\tau_n(k|y) = \mathbb{P}_{\mu_0}[T_0 \le n + k | Y_n = y]$ of the time to extinction assumes the form

$$\tau_n(k|y) = G_{Z_{n+k}|Y_n=y}(0) = \begin{cases} c_n \left(1 - v_n + v_n \dfrac{\theta p_n p}{p + q v_{k-1} - q_n \theta p} \right) & \text{if } y = 0; \\[4mm] (1 - q_n \theta)^{y+1} \left(1 - \dfrac{q_n \theta p}{p + q v_{k-1}} \right)^{-y-1} & \text{if } y > 0. \end{cases}$$

As for the conditional distribution of eventual extinction, this can be seen to be

$$\mathbb{P}_{\mu_0}[\mathcal{E}_0 | Y_n = y] = \begin{cases} c_n \left(1 - v_n + v_n \dfrac{p_n (p/q)}{1 - q_n (p/q)} \right) & \text{if } y = 0; \\[4mm] \left(\dfrac{1 - q_n \theta}{1 - q_n (p/q)} \right)^{y+1} & \text{if } y > 0. \end{cases}$$

6 Extension to the general partially observed branching process

So far we have considered a situation where an infected individual that is observed at time n is instantaneously removed, and so is unable to produce offsprings:

$$Z_{n+1} = \sum_{i=1}^{Z_n} \xi_{i,n}(1 - B_{i,n}) \stackrel{d}{=} \sum_{i=1}^{Z_n - Y_n} \xi'_{i,n} \qquad (12)$$

We now use our results as a pivot to obtain answers in the general scenario of model (3), where

$$Z_{n+1} = \sum_{i=1}^{Z_n} \xi_{i,n}(1 - B_{i,n}) + \sum_{i=1}^{Z_n} \zeta_{i,n} B_{i,n} \qquad (13)$$

Since the observed individuals are known at any point in time, any probability measure of interest is conditioned on Y_n. Therefore, Lemma 1 allows us to obtain these measures by convolving the measures obtained in the previous paragraphs, with an appropriate measure arising from the contribution of the observed individuals.

6.1 Transition probabilities

As before, we shall consider the Markov chain $\{(Z_n, Y_n)\}_{n \geq 0}$, whose state-space is the first quadrant of the lattice $\mathbb{Z}^2$. The one-step transition probabilities are given by

$$\mathbb{P}[Z_1 = z_1, Y_1 = y_1 | Z_0 = z_0, Y_0 = y_0] = \Xi^*_{z_0 - y_0} * \Upsilon^*_{y_0}(z_1) \binom{z_1}{y_1} (1 - \theta)^{y_1} \theta^{z_1 - y_1},$$

where Υ^*_k stands for the k-fold convolution of Υ with itself. The generating function for the n-step transition distribution is given in the following theorem.

Theorem 4 *Let $P_n(r, s)$ be the probability generating function corresponding to the mass function $\mathbb{P}[Z_n = \cdot, Y_n = \cdot | Z_0 = z_0, Y_0 = y_0]$, where Z_n evolves as in (13). Then,*

$$P_n(r, s) = \left\{ G_\Xi \left[\phi_{n-1}(r\theta + rs(1 - \theta)) \right] \right\}^{z_0 - y_0} \left\{ G_\Upsilon \left[\phi_{n-1}(r\theta + rs(1 - \theta)) \right] \right\}^{y_0},$$

where $\phi_k(\cdot)$ stands for the probability generating function of the kth generation of a Galton–Watson branching process started from a single individual and evolving as $\{Z_n\}$ evolves marginally.

Proof Mimicking the steps for the proof of Theorem 1, we have

$$
P_n(r, s) = \sum_{a=0}^{\infty} [r\theta + rs(1 - \theta)]^a \mathbb{P}[Z_n = a | Z_0 = z_0, Y_0 = y_0]
$$

$$
= \sum_{a=0}^{\infty} [r\theta + rs(1 - \theta)]^a \sum_{z_1=0}^{\infty} \mathbb{P}[Z_n = a | Z_1 = z_1] \Xi_{z_0 - y_0}^* * \Upsilon_{y_0}^*(z_1)
$$

$$
= \sum_{z_1=0}^{\infty} \Xi_{z_0 - y_0}^* * \Upsilon_{y_0}^*(z_1) \sum_{a=0}^{\infty} [r\theta + rs(1 - \theta)]^a \mathbb{P}[Z_n = a | Z_1 = z_1].
$$

And, by the definition of $\phi_k(\cdot)$,

$$
P_n(r, s) = \sum_{z_1=0}^{\infty} \Xi_{z_0 - y_0}^* * \Upsilon_{y_0}^*(z_1) [\phi_{n-1}(r\theta + rs(1 - \theta))]^{z_1}
$$

$$
= \{ G_\Xi [\phi_{n-1}(r\theta + rs(1 - \theta))] \}^{z_0 - y_0} \{ G_\Upsilon [\phi_{n-1}(r\theta + rs(1 - \theta))] \}^{y_0},
$$

completing the proof.

6.2 Predictions based on the observable component

Theorem 5 *Suppose that the initial distribution of $\{Z_n, Y_n\}$ is μ. The generating function for the one-step prediction distribution, $\mathbb{P}_\mu[Z_{n+1} = z | Y_n = y]$, may be expressed as:*

$$
G_{Z_{n+1} | Y_n = y_n}(r) = \frac{\eta_n^{(y_n)}(\theta G_\Xi(r))}{\eta_n^{(y_n)}(\theta)} [G_\Upsilon(r)]^{y_n},
$$

where $\eta_k(\cdot) = G_{N_0}\{G_\Xi[\phi_{k-1}(\cdot)]\}$ is the probability generating function of the $(k-1)$th generation of a Galton–Watson branching process started with initial distribution dist $\{\sum_{i=1}^{N_0} \xi_{i,0}\}$ *and evolving as $\{Z_n\}$ evolves marginally. More generally, the generating function for the k-step prediction distribution, $\mathbb{P}_\mu[Z_{n+k} = z | Y_n = y]$, is given by*

$$
G(r) = \frac{\eta_n^{(y_n)}(\theta G_\Xi(\phi_{k-1}(r)))}{\eta_n^{(y_n)}(\theta)} [G_\Upsilon(\phi_{k-1}(r))]^{y_n}
$$

where $\eta_m(\cdot)$ is as before.

Proof We have that

$$
Z_{n+1} = A + B, \tag{14}
$$

where A are the offsprings produced by the unobserved individuals and B are the offsprings produced by the observed individuals. Lemma 1 then shows that

A is independent of *B* conditional on Y_n, and the result follows immediately. For the *k*-step generating function, we simply imitate the proof of Theorem 2 replacing $\lambda_k(\cdot)$ by $\eta_k(\cdot)$.

Corollary 3 *Suppose that the initial distribution of* $\{Z_n, Y_n\}$ *is* μ. *The probability of extinction conditional on the current observations on the process is given by*

$$\mathbb{P}_\mu[\mathcal{E}_0|Y_n = y] = \frac{\eta_n^{(y)}(\theta G_\Xi(s_0))}{\eta_n^{(y)}(\theta)}[G_\Upsilon(s_0)]^y, \tag{15}$$

where s_0 *is the unique point that satisfies* $\phi_1(s) = s$ *and* $s \neq 1$. *Furthermore, the time to extinction conditional on the current observations on the process is given by*

$$\mathbb{P}_\mu[T_0 < n + k|Y_n = y] = \frac{\eta_n^{(y)}(\theta G_\Xi(\phi_{k-1}(0)))}{\eta_n^{(y)}(\theta)}[G_\Upsilon(\phi_{k-1}(0))]^y \tag{16}$$

Remark 3 The exact results obtained when the offspring distribution is in the Geometric family are extended easily in the context of the general partially observed branching process—one may assume a geometric distribution with a different parameter for the offsprings of the observed individuals, and modify the results according to the results of this paragraph.

7 Discussion

Incorporating a random mechanism that leads to the partial observation of a branching process will often lead to more realistic models for stochastic epidemics, yet will also usually lead to major complications. Especially in the case when observation of subsets of the process leads to modifications of the actual process of infections, even simple such mechanisms can lead to intractable situations.

In this paper, we have suggested a model that introduces partial observation through binomial thinning, and interventions through what we have called the *quarantine assumption*. We have shown that such a model preserves the marginal branching property of the process of infected individuals, while making the joint process of infected and observed individuals a Markov chain on $\mathbb{Z}_+^2 \cup \{\mathbf{0}\}$. We have specified (up to generating functions) the transition probabilities, and also the conditional distributions for the true process given the present of the observable process.

In addition, we have seen how the probability of extinction and the distribution of the time to extinction are altered when we take into account that what we observe is not the true process. All these distributions are described entirely through the law of a branching process evolving as the true process evolves marginally. Finally, we have seen how the above formulation gives particularly simple expressions in a special case, that of Geometric branching.

These results describe the probabilistic behaviour of the process, and the modifications needed to be made in certain fundamental quantities when modelling the initial stages of an epidemic according to a branching process. These

modifications can be important when considering the implementation of control measures given observations of a small time period of the epidemic. Of course, this requires knowledge (at least to some approximation) of the relevant parameters.

When the parameters involved are unknown, we have the very interesting problem of conducting statistical inferences on the parameter values given the observation of a finite sample path of the observable process $\{Y_n\}$. Statistical inference for branching processes is already a very interesting topic even for processes that are completely observed [4]. A starting point would be to try to write down a likelihood function (which could also lead to a Bayesian setup) given values $\mathbf{y} = \{y_1, \ldots, y_n\}$. This could be rather complicated, and it may be more practical to write down a pseudolikelihood, as if $\{Y_n\}$ possessed the Markov property:

$$\ell\{\mathbf{y}\} := \mathbb{P}_\mu[Y_n = y_n | Y_{n-1} = y_{n-1}] \times \cdots \times \mathbb{P}_\mu[Y_2 = y_2 | Y_1 = y_1]\mathbb{P}_\mu[Y_1 = y_1],$$
(17)

for some initial measure μ. One may immediately obtain expressions for these "transition probabilities":

$$\mathbb{P}[Y_n = y_n | Y_{n-1} = y_{n-1}] = \sum_{z_n=0}^{\infty} \mathbb{P}[Y_n = y_n | Z_n = z_n]\mathbb{P}[Z_n = z_n | Y_{n-1} = y_{n-1}]$$

$$= \sum_{z_n=0}^{\infty} \frac{z_n!}{y_n!(z_n - y_n)!}(1-\theta)^{y_n}\theta^{z_n - y_n}$$

$$\times \mathbb{P}[Z_n = z_n | Y_{n-1} = y_{n-1}]$$

$$= \frac{(1-\theta)^{y_n}}{y_n!} G^{(y_n)}_{Z_n | Y_{n-1} = y_{n-1}}(\theta)$$

Some applied work in this direction has been done by Meester et al. [7] in connection to swine fever data, under the simplifying assumption of a specific finitely supported offspring distribution. In addition to actually performing inference, another interesting question is that of the comparison of the inferential results from assuming perfect observation with those obtained assuming partial observation and how these may affect control decisions.

A further issue raised in this paper, that is immediately connected with the appropriateness of pseudolikelihood estimation, is that of "relevant information". One would think that using the whole past of the observable process should not make much difference in terms of information as compared to using a few steps. How few steps one may use, and what kind of relative entropy difference this would lead to are relevant questions in this context.

Finally, I would like to thank the editor and two referees for their constructive comments and suggestions.

References

1. Andersson, H., Britton, T.: Stochastic epidemic models and their statistical analysis. In: Lecture Notes in Statistics, vol. 151. Springer, Berlin Heidelberg New York (2000)
2. Athreya, K.B., Ney, P.E.: Branching Processes. Springer, Berlin Heidelberg New York (1972)
3. Bartlett, M.S.: An Introduction to Stochastic Processes. Camridge University Press, London (1955)
4. Guttorp, P.: Statistical Inference for Branching Processes. Wiley, New York (1991)
5. Kendall, D.G.: Deterministic and stochastic epidemics in closed populations. In: Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability, vol. 4, pp. 149–165 (1956)
6. Kendall, D.G.: Branching processes since 1873. J. Lond. Math. Soc. **41**, 385–406 (1966)
7. Meester, R., de Koning, J., de Jong, M., Diekmann, O.: Modeling and real-time prediction of classical swine fever epidemics. Biometrics **58**, 178–184 (2002)
8. Panaretos, V.M.: Imperfectly observed Markov processes for stochastic epidemics. In: Technical Report, Department of Statistics, University of Oxford (2004)