THE EARLY STAGE BEHAVIOUR OF A STOCHASTIC SIR EPIDEMIC WITH TERM-TIME FORCING

TOM BRITTON [*] AND

MATHIAS LINDHOLM,[**] *Stockholm University*

Abstract

The general stochastic SIR epidemic in a closed population under the influence of a term-time forced environment is considered. An 'environment' in this context is any external factor that influences the contact rate between individuals in the population, but is itself unaffected by the population. Here 'term-time forcing' refers to discontinuous but cyclic changes in the contact rate. The inclusion of such an environment into the model is done by replacing a single contact rate λ with a cyclically alternating renewal process with k different states denoted $\{\Lambda(t)\}_{t \geq 0}$. Threshold conditions in terms of $R_\star$ are obtained, such that $R_\star > 1$ implies that π, the probability of a large outbreak, is strictly positive. Examples are given where π is evaluated numerically from which the impact of the distribution of the time periods that $\Lambda(t)$ spends in its different states is clearly seen.

Keywords: Stochastic epidemic; branching process in a seasonal environment; seasonal forcing; term-time forcing; threshold conditions

2000 Mathematics Subject Classification: Primary 92D30
Secondary 60J80

1. Introduction

In the present paper we focus on infectious diseases of SIR type, where SIR is an abbreviation for susceptible, infectious, and recovered (and immune). These are consequently the only possible states that an individual can belong to when we discuss SIR-type diseases, and the possible transitions between these states follows S $\to$ I $\to$ R.

In the classical general stochastic SIR epidemic (see, e.g. [2, Chapter 2.3, p. 14]) many unrealistic assumptions are made. For example, we assume that the population is closed, that all individuals mix homogeneously, that the contact rate is constant, etc. As a consequence of this, many modifications of this model have been introduced and analysed in the literature, where different assumptions have been relaxed; see, e.g. [2] and the references therein. In the present paper we relax the assumption that the rate of contact between individuals is constant over time, and we are particularly interested in the situation where the contact rate changes following some cyclic pattern. An example of this is how school children interact during school terms and holidays. This effect on the contact rate is quantified as a seasonally varying environment. The term 'environment' in this context is any external factor that influences the rate with which individuals make contact in the population, but is itself independent of

the population. For many viral and bacterial diseases, seasonal incidence patterns have been observed, and measles is a striking example where the seasonal incidence patterns seem to be highly correlated with seasonal contact patterns; see, e.g. [4, pp. 44, 46, 136–137]. The topic of seasonality in connection to epidemics has been addressed in several papers, of which some of the more recent are [13] and [19], and an earlier reference is [22]. These papers are concerned with endemic diseases. In [21] the authors considered a closed population SIR model in a seasonally varying environment using a deterministic approach and gave numerical examples. Related issues can also be found in the area of epizootics; see, e.g. [1] and the references therein. All these references consider seasons following some oscillating pattern, i.e. modelling the seasonality using sinusoids (so-called 'seasonal forcing'; see, e.g. [18]). Another type of seasonality that has been considered is the so-called 'term-time forcing'; see, e.g. [11], [18], [23], and the references therein. In this situation the contact rate changes discontinuously, but cyclically, similar to what is natural to believe happens in the holiday/term example for school children above. In the present paper we focus on a situation similar to term-time forcing. That is, we let the environment change cyclically corresponding to the contact intensities jumping at (non)random time points.

The aim of the present paper is to obtain a better understanding for how term-time forced environments affect a newly introduced disease's ability to take off. We make a simple modification of the general stochastic SIR epidemic that takes environmental effects into account, and derive threshold conditions for this model. These conditions are expressed in terms of $R_\star$, loosely defined as the expected number of individuals that are infectious at the time point when an entire environmental cycle is completed, given that the epidemic was started by a single infective. The quantity $R_\star$ relates to R_0, the so-called basic reproduction number. The basic reproduction number is defined as the average number of susceptible individuals that a typical infectious individual infects in an otherwise disease-free population during its infectious period. More precisely, a large outbreak can only occur when $R_0 > 1$; see, e.g. [2, Theorem 3.1, p. 25]. From our main result, Theorem 1, we show that $R_\star$ possesses the same threshold property.

We consider environments that change cyclically between k different states, and the time that the environment spends in any state may be either random or nonrandom. For the special case when the time periods that the environment spends in its different states are exponentially distributed, we show how this system can be represented as a multitype epidemic, and show, for $k = 2$, that R_0 and $R_\star$ are equivalent in terms of threshold behaviour. The analogy is, however, only valid for the derivation of the threshold and we cannot use the construction to obtain information regarding the probability of a large outbreak. Moreover, we show that if this kind of system is approximated by a general stochastic SIR epidemic with a single contact rate equal to the time-mean environmental effect on the rate of contacts between individuals, the distributions of the time that the environmental process spends in its different states are of importance. More precisely, we show that $R_\star$ of the actual process with environment included is always greater than the corresponding $R_\star$ of the approximated system, except when the time periods that the environment spends in its different states are nonrandom. For this situation, the two thresholds are equal. This is quantified in Corollary 1. The implication is hence somewhat disappointing from an applied point of view, from which we would rather see this kind of approximation yielding an epidemic that takes off easier than the actual one. In fact, in connection with the statement of Theorem 1 in Section 3 we give an example where $R_\star$ of the approximated system is strictly smaller than 1, whereas $R_\star$ of the actual process is strictly greater than 1.

A short outline of our analysis can be formulated as follows. First a coupling between a suitable branching process with a varying environment and the epidemic process with a varying environment is established. Then, by using the fact that the approximating process analysed at the time points when entire environmental cycles are completed behaves as a Galton–Watson branching process, the desired results follow from standard results of branching process theory.

The disposition of this paper is as follows. In Sections 2 and 3 we define the model, and state and discuss our main results. Section 4 is devoted to the coupling between the number of infectious individuals in our epidemic process and the approximating branching process, and the proof of Theorem 1. Section 5 is concerned with some important examples, and we conclude with a discussion in Section 6.

2. The general stochastic SIR epidemic in a term-time forced environment

In the general stochastic SIR epidemic (see [2, Chapter 2.3, pp. 14–15]) a closed homogeneous population is considered, in which all infectious contacts are made according to a Poisson process with intensity λ and infectious individuals recover, independently of everything else, at rate γ. Denote this process by $\{(S^{(n)}(t; \lambda, \gamma), I^{(n)}(t; \lambda, \gamma))\}_{t \geq 0}$, where n refers to the size of the initially susceptible population. Note that $R(t; \lambda, \gamma)$, the number of recovered (and immune) individuals at t, is superfluous since the population is closed, i.e. $S^{(n)}(t; \lambda, \gamma) + I^{(n)}(t; \lambda, \gamma) + R(t; \lambda, \gamma) + 1$. Unless otherwise stated, we will assume throughout that $(S^{(n)}(0; \lambda, \gamma), I^{(n)}(0; \lambda, \gamma)) = (n, 1)$ for notational convenience. As described in the introduction, the environment affects only the intensity with which individuals make contact, and the environmental changes are cyclic. We can hence describe our model as follows. When the environment is in state i, the rate at which individuals make contact will be λ_i. Furthermore, the time periods that the environment spends in state i are distributed as the (non)random variable T_i, and then changes to state $i + 1$. Due to the cyclic behaviour of the environmental process, this will correspond to the contact rates changing according to $\lambda_1 \rightarrow \lambda_2 \rightarrow \cdots \lambda_k \rightarrow \lambda_1 \rightarrow \cdots$, when there are k different environmental states. Thus, by replacing a single λ by a cyclically varying intensity process $\{\Lambda(t)\}_{t \geq 0}$ ($\Lambda(0) = \lambda_1$), where $\Lambda(t) = \lambda_i$ if the environmental process is in state i ($i = 1, \ldots, k$) at t, the general stochastic SIR epidemic can account for environmental effects; see Table 1. That is, our new epidemic process is $\{(S^{(n)}(t; \Lambda(t), \gamma), I^{(n)}(t; \Lambda(t), \gamma))\}_{t \geq 0}$. Henceforth, we will make no distinction between the environmental process and the intensity process $\Lambda(t)$. If we summarise our restrictions on the environmental process, we will assume throughout that

(a) it is piecewise constant, i.e. $\Lambda(t) \in \{\lambda_1, \ldots, \lambda_k\}$,

(b) it changes cyclically between its states,

(c) the time periods that $\Lambda(t)$ spends in a state i are independent and identically distributed (i.i.d.), and distributed as the random variables T_i,

(d) the time periods that $\Lambda(t)$ spends in different states are independent,

(e) $\Lambda(0) = \lambda_1$.

In the next section we state and discuss our main result, which describes how a term-time forced environment affects the possibility of a large outbreak. The notion of a large outbreak is here equivalent to the fact that the number of ultimately infected individuals will tend to ∞ when passing to the limit in n.

TABLE 1: The general stochastic SIR epidemic in a term-time forced environment. Here $S^{(n)}(t; \Lambda(t), \gamma) = s$ and $I^{(n)}(t; \Lambda(t), \gamma) = i$ denote the number of susceptible and infectious individuals at time t, respectively, and $\Lambda(t)$, $\Lambda \in \{\lambda_1, \ldots, \lambda_k\}$, is the intensity process that changes state depending on the environment. Unless stated otherwise, we assume that $(S^{(n)}(0; \Lambda(0), \gamma), I^{(n)}(0; \Lambda(0), \gamma)) = (n, 1)$ and $\Lambda(0) = \lambda_1$.

From	To	Rate
(s, i)	$(s - 1, i + 1)$	$\dfrac{\Lambda(t)}{n} si$
(s, i)	$(s, i - 1)$	γi

3. Results

We now define a number of quantities needed in the formulation of our main result, Theorem 1. These quantities will be explained in more detail in Section 4.

Recall from the previous section that T_i is a stochastic variable distributed as the typical time that the environmental process spends in state i. If we denote the moment generating function of T_i (T_i is nonnegative) by

$$\Psi_i(s) := \mathrm{E}[e^{sT_i}] = \int_0^\infty e^{st} f_{T_i}(t)\, \mathrm{d}t, \tag{1}$$

which is assumed to be finite for $s < s_0$, where $s_0 > 0$. Note that the integral in (1) is always convergent for $s < 0$, but that $\Psi_i(s)$ then loses its interpretation as a moment generating function. Throughout, we will use statements with negative s without further comments.

Furthermore, let $\xi_i(s, t)$ denote the generating function of $X(t; \lambda_i, \gamma)$, which is defined as a linear birth-and-death process with birth rate λ_i and death rate γ. Then it can be shown that

$$\xi_i(s, t) := \mathrm{E}[s^{X(t; \lambda_i, \gamma)}] = \frac{\gamma(s - 1) - e^{(\gamma - \lambda_i)t}(\lambda_i s - \gamma)}{\lambda_i(s - 1) - e^{(\gamma - \lambda_i)t}(\lambda_i s - \gamma)}, \qquad |s| \le 1,\ t \ge 0; \tag{2}$$

see, e.g. [5, Chapter III.5, p. 109]. Furthermore, define

$$\chi(s) := \mathrm{E}[\xi_1(\xi_2(\xi_3(\ldots \xi_{k-1}(\xi_k(s, T_k), T_{k-1}) \ldots, T_3), T_2), T_1)], \qquad |s| \le 1 \tag{3}$$

Moreover, let ρ be the smallest nonnegative solution to the equation

$$s = \chi(s).$$

We can now state our main result.

Theorem 1. *Let $\{(S^{(n)}(t; \Lambda(t), \gamma), I^{(n)}(t; \Lambda(t), \gamma))\}_{t \ge 0}$ denote an epidemic process in a term-time forced environment defined by the rates in Table 1, and let everything else be defined as above. Moreover, let $E^{(n)} := n - S^{(n)}(\infty; \Lambda(\infty), \gamma)$. Then $R_\star$ defined by*

$$R_\star = \prod_{i=1}^k \Psi_i(\lambda_i - \gamma) \tag{4}$$

works as a threshold, such that $E^{(n)} \xrightarrow{\mathrm{D}} E$ as $n \to \infty$, and $\pi := \mathrm{P}(E = +\infty)$ satisfies the relation $\pi = 1 - \rho$. In addition, $\pi > 0$ if and only if $R_\star > 1$.

The proof of Theorem 1 is rather lengthy and is given in Section 4.

Remark 1. As will become apparent later, the epidemiological interpretation of $R_\star$ is that it corresponds to the average number of individuals that are *infectious* at the time point when an environmental cycle is completed, given that the epidemic was initiated by a single infectious individual whose infectious period starts at the time point when a new environmental cycle is started. In Remark 3, below, it will become clear why this threshold should be expressed in terms of infectious rather than infected individuals. By inspecting the above theorem we see that it is relatively easy to compute $R_\star$, at least with the aid of a computer. The value of $R_\star$ tells us whether the probability of a large outbreak is strictly greater than 0 or not. Unfortunately, the actual probability π is much harder to compute. In Section 5 we discuss two special cases where it is possible to compute π numerically. Furthermore, if we have only one environmental state then $T = +\infty$, and $R_\star = +\infty$ if $\lambda - \gamma > 0$ and $R_\star = 0$ otherwise. For the general stochastic SIR epidemic with a single environment, $R_0 = \lambda/\gamma$ (see, e.g. [2, Chapter 2.1, p. 12]), and the threshold property of $R_\star$ is hence preserved since $R_0 > 1$ is equivalent to $R_\star = +\infty$. Moreover, from (1), it is clear that $\Psi_i(s) > 1$ if $s > 0$ and $\Psi_i(s) \leq 1$ otherwise. Thus, if $\lambda_i > \gamma$ for $i = 1, \ldots, k$ then $R_\star > 1$, and if $\lambda_i \leq \gamma$ for $i = 1, \ldots, k$ then $R_\star \leq 1$. Consequently, the remaining cases are those where $\lambda_i > \gamma$ holds for some i and $\lambda_i \leq \gamma$ for some other i.

Suppose that we try to approximate this type of system. One way would be to calculate the long-time average environmental effect, $\tilde{\lambda}$, and use this single rate in the general SIR epidemic. Postponing the calculation of $\tilde{\lambda}$ for a moment, assuming that it is possible to obtain, we get

$$\tilde{R}_0 = \frac{\tilde{\lambda}}{\gamma}. \tag{5}$$

As mentioned in Section 2, the intensity process $\Lambda(t)$ is a cyclically varying renewal process. (These processes are often referred to in the literature as alternating renewal processes when $k = 2$.) For the case with k environmental states, we obtain

$$\tilde{\lambda} := \sum_{i=1}^{k} \frac{\mathrm{E}[T_i]}{\mathrm{E}[T_1] + \cdots + \mathrm{E}[T_k]} \lambda_i, \tag{6}$$

using well-known facts from renewal theory (i.e. the factors in front of the λ_is are the long-time proportions of time that $\Lambda(t)$ spends in state i, which follow from the law of large numbers for renewal processes; see, e.g. [10, Chapter 3.4, pp. 202–216]). By using $R_\star$ from Theorem 1 together with $\tilde{\lambda}$ from (6), we can conclude the following.

Corollary 1. *Let $\tilde{R}_\star$ be defined as in Theorem 1, but with all the λ_i ($i = 1, \ldots, k$) replaced with $\tilde{\lambda}$ from (6), the mean intensity of the process $\Lambda(t)$, and let $R_\star$ be defined as in Theorem 1. For $R_\star$, we also assume that $\mathrm{var}(T_i) > 0$ for some i and let $t_i := \mathrm{E}[T_i]$. Furthermore, let $R_\star^\delta$ denote $R_\star$ where all $T_i \in \delta(t_i)$, i.e. $T_i \equiv t_i$, with the corresponding $\tilde{R}_\star^\delta$. Then,*

(i) $R_\star > R_\star^\delta$,

(ii) $R_\star^\delta = \tilde{R}_\star^\delta$,

(iii) $\tilde{R}_0 > 1$ *implies that* $R_\star > 1$.

Proof. By noting that

$$\Psi_i(s) := \mathrm{E}[e^{sT_i}] > e^{s\,\mathrm{E}[T_i]}$$

holds by a direct application of Jensen's inequality (except when $T_i \equiv \mathrm{E}[T_i]$, in which case we have equality), the proof of Corollary 1 follows directly from the definition of $R_\star$ in (4) combined with the definition of $\tilde{\lambda}$ in (6).

Recall from Section 1 that a large outbreak in general can only occur when $R_0 > 1$ (or in our situation when $\tilde{R}_0 > 1$). Using this and combining Corollary 1, Remark 1, and (5), it is clear that

$$\tilde{R}_\star \gtreqqless 1 \quad \Longleftrightarrow \quad \tilde{R}_0 \gtreqqless 1,$$

which implies that $\tilde{R}_0 \geq 1$, which in turn implies that $R_\star \geq 1$.

To see why the opposite implication of Corollary 1(iii) does not hold, use the following counterexample. Consider the situation with two environmental states, where T_1 and T_2 are independent $\mathrm{Exp}(\nu_i)$ distributed ($i = 1, 2$) with corresponding moment generating functions

$$\Psi_i(s) = \frac{1}{1 - s/\nu_i},$$

where $s \in [0, \nu_i)$. Furthermore, owing to Remark 1, we assume that $\lambda_1 > \gamma > \lambda_2$, and $R_\star$ from Theorem 1 becomes

$$R_\star = \frac{1}{1 - (\lambda_1 - \gamma)/\nu_1} \frac{1}{1 - (\lambda_2 - \gamma)/\nu_2}. \tag{7}$$

If we, for example, set $\lambda_1 = 2.7, \lambda_2 = 0.8, \gamma = 2$, and $\nu_1 = \nu_2 = 1$, then $\tilde{\lambda} = 1.75$ and $\tilde{R}_0 = 0.875 < 1$ while $R_\star \approx 1.52 > 1$. Thus, we have an example where the approximation cannot take off but the true epidemic can. This could make health authorities incorrectly feel 'safe' when using the approximated process.

In the next section we will prove a coupling between a suitable branching process and our epidemic process, and prove Theorem 1. Once this coupling has been established, we use the fact that this approximating process observed at the time points when environmental cycles are completed behaves as a Galton–Watson process. Known facts from branching process theory then gives us Theorem 1.

4. Branching process approximation

In this section we relate the epidemic process to a suitable branching process in a varying environment and show that the ultimate number of infected individuals in the epidemic process converges in distribution to the ultimate number of individuals born in this approximating process. This is done by using a coupling argument analogous to that used in [6] for the general epidemic. For more on branching processes in varying and random environments, we refer the reader to, e.g. [14, Chapter 5.10, pp. 145–152] and the references therein, and for more mathematical references on branching processes in general, see, e.g. [5], [15], and [16].

Recall that $\Lambda(t)$ denotes the environmental process defined in Section 2, where $\Lambda(t) \in \{\lambda_1, \ldots, \lambda_k\}$. Let $X(t; \Lambda(t), \gamma)$ denote a continuous-time branching process in a varying environment with k different states describing the number of individuals alive at time t, defined analogously to $X(t; \lambda_i, \gamma)$ in Section 3 except with λ_i replaced by $\Lambda(t)$. For this process, individuals give birth at rate λ_i while the environmental process is in state i, and individuals die independently of everything else at rate γ.

To see how we can construct $I^{(n)}(t; \Lambda(t), \gamma)$ using $X(t; \Lambda(t), \gamma)$, we first introduce a sequence of i.i.d. U(0,1)-distributed random variables ζ_i, $i = 1, 2, \ldots$, and label all n initially

susceptible individuals in the epidemic process from 1 to n. Then, whenever an individual in $X(t; \Lambda(t), \gamma)$ is born, an individual in $I^{(n)}(t; \Lambda(t), \gamma)$ becomes infectious. That is, when the ith individual in $X(t; \Lambda(t), \gamma)$ is born, we pick an individual amongst the n initially susceptible individuals labelled $\lfloor \zeta_i n \rfloor + 1$, where $\lfloor \cdot \rfloor$ refers to the integer part, and infect it in $I^{(n)}(t; \Lambda(t), \gamma)$. This infected individual will then have the same characteristics as that of the ith born individual in $X(t; \Lambda(t), \gamma)$. On the other hand, if the picked individual is no longer susceptible, the coupling breaks down and we have encountered what is called in [20] a 'ghost', but until this event occurs the two processes will be identical.

Let $B(t; \Lambda(t), \gamma)$ denote the number of births in $X(t; \Lambda(t), \gamma)$ up to t. If we define $E^{(n)} \colon -S^{(n)}(\infty; \Lambda(\infty), \gamma)$, and $E := B(\infty; \Lambda(\infty))$, we can state the main result of this section.

Theorem 2. *We have $E^{(n)} \xrightarrow{\mathrm{D}} E$ as $n \to \infty$.*

Before we proceed with the proof of Theorem 2, we need to establish the length of time the coupling between $I^{(n)}(t; \Lambda(t), \gamma)$ and $X(t; \Lambda(t), \gamma)$ holds.

Lemma 1. *The processes $I^{(n)}(t; \Lambda(t), \gamma)$ and $X(t; \Lambda(t), \gamma)$ can, with a probability tending to 1 (as n tends to ∞), be coupled up to time $\tau(n) := c \log n$, where $0 < c < (1-2\varepsilon)/(2(\lambda_{\max} - \gamma))$, $\lambda_{\max} := \max_{i=1,\ldots,k} \lambda_i$ ($\Lambda(t) \in \{\lambda_1, \ldots, \lambda_k\}$), and $0 < \varepsilon < \frac{1}{2}$.*

Proof. To prove Lemma 1, we need the following notation. Let $B^{(n)}(t; \Lambda(t), \gamma)$ denote the number of individuals infected up to t in $I^{(n)}(t; \Lambda(t), \gamma)$. Furthermore, let $X(t; \lambda_{\max}, \gamma)$ be a linear birth-and-death process with birth and death intensities $\lambda_{\max}$ and γ, respectively, and let $B(t; \lambda_{\max}, \gamma)$ denote the number of individuals born up to t in $X(t; \lambda_{\max}, \gamma)$. Note that $B^{(n)}(t; \Lambda(t), \gamma) \leq_{\mathrm{D}} B(t; \Lambda(t), \gamma) \leq_{\mathrm{D}} B(t; \lambda_{\max}, \gamma)$ holds for all $t \geq 0$.

From the classical birthday problem, it follows that the coupling will hold until $n^{1/2-\varepsilon}$ individuals have been born with a probability tending to 1 as n tends to ∞, where $0 < \varepsilon < \frac{1}{2}$; see, e.g. [12, Chapter II.3, p. 31]. From the relation $B^{(n)}(t; \Lambda(t), \gamma) \leq_{\mathrm{D}} B(t; \Lambda(t), \gamma) \leq_{\mathrm{D}} B(t; \lambda_{\max}, \gamma)$, it follows that if it takes no more than $\tau(n)$ time units for $B(t; \lambda_{\max}, \gamma)$ to reach $n^{1/2-\varepsilon}$ individuals with a high probability, the coupling between $I^{(n)}(t; \Lambda(t), \gamma)$ and $X(t; \Lambda(t), \gamma)$ will hold for at least as many time units. Thus, we want to show that

$$\mathrm{P}(B(\tau(n); \lambda_{\max}, \gamma) > n^{1/2-\varepsilon}) \to 0 \quad \text{as } n \to \infty$$

for some suitable choice of $\tau(n)$. This we will do by using Markov's inequality:

$$\mathrm{P}(B(t; \lambda_{\max}, \gamma) > n^{1/2-\varepsilon}) \leq \frac{\mathrm{E}[B(t; \lambda_{\max}, \gamma)]}{n^{1/2-\varepsilon}}. \tag{8}$$

Recall that $X(t; \lambda_{\max}, \gamma)$ is a linear birth-and-death process, and can hence be represented as $X(t; \lambda_{\max}, \gamma) = B(t; \lambda_{\max}, \gamma) - D(t; \gamma) + 1$ (i.e. $X(0; \lambda_{\max}, \gamma) = 1$), where $B(t; \lambda_{\max}, \gamma)$ is a process counting the number of individuals born up to time t and $D(t; \lambda_{\max}, \gamma)$ is a process counting the number of individuals that have died up to time t. For an ordinary linear birth-and-death process $X(t; \lambda, \gamma)$ with probability generating function $\xi(s, t)$ given by (2) where $\lambda_i = \lambda$, we obtain

$$\mathrm{E}[X(t; \lambda, \gamma)] = \left. \frac{\mathrm{d}}{\mathrm{d}s} \xi(s, t) \right|_{s=1} = \mathrm{e}^{(\lambda-\gamma)t}, \qquad t \geq 0. \tag{9}$$

In order to determine $\mathrm{E}[B(t; \lambda_{\max}, \gamma)]$, we condition and obtain

$$\mathrm{E}[B(t + h; \lambda_{\max}, \gamma)] = \mathrm{E}[\mathrm{E}[B(t + h; \lambda_{\max}, \gamma) \mid B(t; \lambda_{\max}, \gamma), X(t; \lambda_{\max}, \gamma)]]$$
$$= \lambda \, \mathrm{E}[X(t; \lambda_{\max}, \gamma) + B(t; \lambda_{\max}, \gamma)]h + o(h),$$

which can be rewritten as

$$\frac{1}{h} \mathrm{E}[B(t+h; \lambda_{\max}, \gamma) - B(t; \lambda_{\max}, \gamma)] = \lambda \mathrm{E}[X(t; \lambda_{\max}, \gamma)] + \frac{o(h)}{h}.$$

If we let $h \to 0$ and integrate with respect to time, and use (9), we arrive at

$$\mathrm{E}[B(t; \lambda_{\max}, \gamma)] = \frac{\lambda_{\max}}{\lambda_{\max} - \gamma} (\mathrm{e}^{(\lambda_{\max} - \gamma)t} - 1), \qquad t \geq 0.$$

This together with (8) gives us $\mathrm{P}(B(\tau(n); \lambda_{\max}, \gamma) > n^{1/2-\varepsilon}) \to 0$ as $n \to \infty$ if $\tau(n) := c \log n$, where $c < (1 - 2\varepsilon)/(2(\lambda_{\max} - \gamma))$ and $0 < \varepsilon < \frac{1}{2}$. This completes the proof.

Remark 2. Note that if $\lambda_{\max} \leq \gamma$ then $\tau(n) = +\infty$.

Proof of Theorem 2. First, note that $E^{(n)} \colon -S^{(n)}(\infty; \Lambda(\infty), \gamma) = B^{(n)}(\infty; \Lambda(\infty), \gamma)$ and recall that $E := B(\infty; \Lambda(\infty), \gamma)$. Hence, in order to show that $E^{(n)} \xrightarrow{\mathrm{D}} E$ as $n \to \infty$, we need to relate $B^{(n)}(t; \Lambda(t), \gamma)$ to $B(t; \Lambda(t), \gamma)$. As before, let $\tau(n)$ denote the time until the coupling breaks down. From Lemma 1, for large enough n, it holds that $\mathrm{P}(\tau(n) \leq c \log n) > 1 - \varepsilon/2$, where $\varepsilon > 0$, and $\{B^{(n)}(t; \Lambda(t), \gamma) \leq k\} = \{B(t; \Lambda(t), \gamma) \leq k\}$ for any fixed $k \leq n^{1/2}$ and $t \leq \tau(n)$. Following the line of proof of Theorem 3.1 of [9], introduce $A_n = \{\tau(n) \leq c \log n\}$. For fixed $t \leq \tau(n)$ and $k \leq \sqrt{n}$, we obtain

$$\begin{aligned}
\mathrm{P}(B^{(n)}(t; \Lambda(t), \gamma) \leq k) &= \mathrm{P}(\{B^{(n)}(t; \Lambda(t), \gamma) \leq k\} \cap A_n) \\
&\quad + \mathrm{P}(\{B^{(n)}(t; \Lambda(t), \gamma) \leq k\} \cap A_n^{\mathrm{c}}) \\
&= \mathrm{P}(\{B(t; \Lambda(t), \gamma) \leq k\} \cap A_n) + \mathrm{P}(\{B^{(n)}(t; \Lambda(t), \gamma) \leq k\} \cap A_n^{\mathrm{c}}) \\
&\leq \mathrm{P}(B(t; \Lambda(t), \gamma) \leq k) + \frac{\varepsilon}{2}.
\end{aligned}$$

Likewise, we obtain

$$\begin{aligned}
\mathrm{P}(B^{(n)}(t; \Lambda(t), \gamma) \leq k) &= \mathrm{P}(\{B^{(n)}(t; \Lambda(t), \gamma) \leq k\} \cap A_n) \\
&\quad + \mathrm{P}(\{B^{(n)}(t; \Lambda(t), \gamma) \leq k\} \cap A_n^{\mathrm{c}}) \\
&\geq \mathrm{P}(\{B(t; \Lambda(t), \gamma) \leq k\} \cap A_n) \\
&= \mathrm{P}(B(t; \Lambda(t), \gamma) \leq k) - \mathrm{P}(\{B(t; \Lambda(t), \gamma) \leq k\} \cap A_n^{\mathrm{c}}) \\
&\geq \mathrm{P}(B(t; \Lambda(t), \gamma) \leq k) - \frac{\varepsilon}{2},
\end{aligned}$$

and, hence,

$$|\mathrm{P}(B^{(n)}(t; \Lambda(t), \gamma) \leq k) - \mathrm{P}(B(t; \Lambda(t), \gamma) \leq k)| < \varepsilon$$

holds. But, in the limit as $n \to \infty$ we have $k \in \mathbb{N}$, $t \in \mathbb{R}^+$, and $\varepsilon > 0$, which were all chosen arbitrarily. Hence, it holds that $B^{(n)}(t; \Lambda(t), \gamma) \xrightarrow{\mathrm{D}} B(t; \Lambda(t), \gamma)$ as $n \to \infty$ for all fixed $t \in \mathbb{R}^+$ and, in particular, it follows that $E^{(n)} \xrightarrow{\mathrm{D}} E$ as $n \to \infty$.

Note that even though we know that $E^{(n)} \to E$ in distribution as $n \to \infty$ we have not yet established any characteristics concerning the distribution of E. For the purposes of the present paper, it is of interest to determine whether E is distributed as a degenerate random variable, with a point mass at ∞ or not, and give conditions for when this occurs. This is effectively what is stated in Theorem 1. By noting that $X(t; \Lambda(t), \gamma)$ observed at the time points when environmental cycles are completed behaves as a certain Galton–Watson process, Theorem 1 follows by applying standard results from branching process theory.

Proof of Theorem 1. As before, let T_i be a random variable distributed as the time that $\Lambda(t)$ spends in state i during a typical visit. We will also use the notation T_{ij} to denote the jth visit to state i, and we will make no distinction between T_i and T_{i1}. Recall that all the $\{T_i\}$ are mutually independent and that, for each $i = 1, \ldots, k$, all $T_{i1}, T_{i2}, \ldots$ are i.i.d. In addition, let U_i denote the time point of the ith completed environmental cycle. This gives us $U_1 = T_1 + \cdots + T_k$, and in general we have $U_i = T_1 + \cdots + T_k + T_{12} + \cdots + T_{ki}$.

Let $\xi_i(s, t)$, $i = 1, \ldots, k$, denote the generating function of $X(t; \lambda_i, \gamma)$ given by (2). Recall that, whenever the environmental process stays constant, the process $X(t; \Lambda(t), \gamma)$ behaves like an ordinary linear birth-and-death process. Thus, by conditioning on the time points $\{U_i\}$ and using the fact that all individuals in $X(t; \Lambda(t), \gamma)$ live for exponentially distributed time periods, we obtain the recursive relation

$$
\mathrm{E}[s^{X(U_1; \Lambda(U_1), \gamma)} \mid \{T_1, \ldots, T_k; X(T_1 + \cdots + T_{k-1}; \Lambda(T_1 + \cdots + T_{k-1}), \gamma)\}]
$$
$$
= \xi_k(s, T_k)^{X(T_1 + \cdots + T_{k-1}; \Lambda(T_1 + \cdots + T_{k-1}), \gamma)}. \tag{10}
$$

Solving this recursive relation backwards from k to 1, it then follows after some further conditioning that

$$
\chi(s) := \mathrm{E}[s^{X(U_1; \Lambda(U_1), \gamma)}] = \mathrm{E}[\xi_1(\xi_2(\xi_3(\ldots \xi_{k-1}(\xi_k(s, T_k), T_{k-1}) \ldots, T_3), T_2), T_1)],
$$

which is the same expression as (3).

Moreover, from the definition of $\Lambda(t)$ we know that the sums $T_{1i} + \cdots + T_{ki}$, $i = 1, 2, \ldots$, are i.i.d., which together with the fact that all individuals live for an exponential life length in the process $X(t; \Lambda(t), \gamma)$ gives us that the process restarts at the time points $\{U_i\}$. Consequently, if we introduce $\tilde{X}_1, \tilde{X}_2, \ldots$ i.i.d. random variables all distributed as $X(U_1; \Lambda(U_1), \gamma)$, we obtain

$$
\mathrm{E}[s^{X(U_2; \Lambda(U_2), \gamma)}]
$$
$$
= \mathrm{E}[\mathrm{E}[s^{X(U_2; \Lambda(U_2), \gamma)} \mid \{T_1, \ldots, T_{k2}; X(T_1 + \cdots + T_k; \Lambda(T_1 + \cdots + T_k), \gamma)\}]]
$$
$$
= \mathrm{E}[\mathrm{E}[s^{X(U_2; \Lambda(U_2), \gamma)} \mid \{T_1, \ldots, T_{k2}; X(U_1; \Lambda(U_1), \gamma)\}]]
$$
$$
= \mathrm{E}[\mathrm{E}[s^{\sum_{i=1}^{X(U_1; \Lambda(U_1), \gamma)} \tilde{X}_i} \mid X(U_1; \Lambda(U_1), \gamma)]]
$$
$$
= \mathrm{E}[\mathrm{E}[s^{\tilde{X}}]^{X(U_1; \Lambda(U_1), \gamma)}]
$$
$$
= \chi(\chi(s)),
$$

and, hence, $X(t; \Lambda(t), \gamma)$ observed at the time points $\{U_i\}$ behaves as a Galton–Watson process with offspring distribution determined by $\chi(s)$ from (3).

Note that if $X(t; \Lambda(t), \gamma)$ observed at the time points $\{U_i\}$ has died out then, obviously, $X(t; \Lambda(t), \gamma)$ has died out as well. For our purposes, the time point when this event occurs is of no interest, and if we introduce $E_d := \sum_{i=0}^{\infty} X(U_i; \Lambda(U_i), \gamma)$, it hence holds that $E_d = +\infty$ if and only if $E = +\infty$. From standard references in branching process theory (see, e.g. [16, Chapter 2.11, pp. 39–42]), it is known that $\mathrm{P}(E_d = +\infty) > 0$ if and only if $\mathrm{E}[X(U_1; \Lambda(U_1), \gamma)] > 1$. But, from the definition of $\chi(s)$ (see (3)), it follows that $\chi'(1) = \mathrm{E}[X(U_1; \Lambda(U_1), \gamma)]$. Note that (10) is a conditional generating function. Hence, we obtain the corresponding conditional expectations by differentiating both sides of (10) with respect to s, and then setting $s = 1$ and using (9). Thus, by solving the recursive relation (10) backwards and using the fact that we can extract all the conditional expectations along the way, we obtain $\chi'(1) = R_\star$, where $R_\star$ is given in (4). Thus, $\mathrm{P}(E_d = +\infty) > 0$ and, hence,

$\pi := \mathrm{P}(E = +\infty) > 0$ if and only if $R_\star > 1$. Moreover, π is given by $\pi := 1 - \rho$, where ρ is the smallest nonnegative solution to $s = \chi(s)$; see, e.g. [16, Theorem 2.3.1, p. 22]. Thus, Theorem 1 is proved.

Remark 3. In the proof of Theorem 1 we made use of an approximating Galton–Watson process. For this process, generations correspond to completed environmental cycles, and the particles correspond to infectious individuals. Thus, a natural interpretation of $R_\star$ is as the average number of individuals that are infectious at the time point when the first environmental cycle is completed, given that the epidemic was initiated by a single infectious individual.

5. Examples

In this section we consider two examples, both with $k = 2$ environmental states. In the first example the time periods spent in different environmental states are exponentially distributed, and in the second example the time periods are nonrandom. Before we proceed with the computation of the probabilities of a large outbreak, $\pi := 1 - \rho$ from Theorem 1, we describe how the situation with exponentially distributed time periods relates to a certain multitype epidemic.

5.1. On the relation to multitype epidemics

Consider the situation where the environment may change between two states, corresponding to the intensities λ_1 and λ_2, and assume that $T_i \in \mathrm{Exp}(\nu_i)$. The situation is now equivalent to that of the counterexample that was treated after the statement of Theorem 1 in Section 3, and $R_\star$ is thus given by (7). Let $r_1 := \lambda_1/\gamma$ and $r_2 := \lambda_2/\gamma$. The interpretation of r_i is that, when $r_i > 1$, the epidemic process in its early stages under the influence of environment i behaves like a supercritical branching process, when $r_i < 1$, it behaves like a subcritical branching process, and when $r_i = 1$, it behaves like a critical branching process. As a consequence of Remark 1, the interesting case is when $r_1 > 1 > r_2$. By straightforward algebraic manipulation of (7) we obtain

$$\frac{r_1 - 1}{\nu_1} \left(\frac{\nu_2}{1 - r_2} + \gamma \right) \gtreqqless 1 \quad \Longleftrightarrow \quad R_\star \gtreqqless 1,$$

and since these expressions are equivalent, we set

$$R_\star := \frac{r_1 - 1}{\nu_1} \left(\frac{\nu_2}{1 - r_2} + \gamma \right). \tag{11}$$

By analysing $R_\star$ termwise we see that $(r_1 - 1)/\nu_1$ is the average excess of infectious individuals that is generated during a typical time period when the environment is in state 1, and $(1 - r_2)/\nu_2$ is the corresponding shortage of infectious individuals. Thus, the heuristic interpretation of this threshold condition is that, for the epidemic to take off, there must be a sufficient excess of infectious individuals from the time spent in the 'favourable' (from the disease's perspective) state of the environment in order to compensate for the time spent in the 'unfavourable' state of the environment.

We now relate this model to a two-type epidemic, i.e. an epidemic model where there is no change of environment but where there are two types of individuals, and individuals can infect others of both types. Let Z_{ij} denote the number of type-j individuals that a single type-i individual infects in an otherwise disease-free population during its infectious period in this two-type epidemic. Here a type-i individual corresponds to an individual that becomes infectious during a time period where the environment is in state i. Let $\tilde{T}$ denote the infectious

period of any individual, $\tilde{T} \in \text{Exp}(\gamma)$, and let T_i denote the time that the environment spends in state i as before. Here $i, j = 1, 2$, and we will, for ease of notation, henceforth assume that $i \neq j$.

While the environmental process is in state i, infectious contacts are made according to a Poisson process with intensity λ_i. Furthermore, if an individual becomes infectious while the environmental process is in state i, the time until either this individual recovers or the environment changes state is $\min\{T_i, \tilde{T}\} \in \text{Exp}(\nu_i + \gamma)$. Combining these two facts we find that the number of i individuals that a newly infected i individual will infect up to the time of recovery or until the environment changes state is $\text{Po}(\lambda_i \min\{T_i, \tilde{T}\})$ distributed, conditional on $\min\{T_i, \tilde{T}\}$. Let the corresponding unconditional number of infected i individuals be denoted by $\tilde{Z}_i$, and it hence holds that $\tilde{Z}_i \in \text{Geo}(p_i)$, $p_i := (\nu_i + \gamma)/(\nu_i + \gamma + \lambda_i)$, and $\text{E}[\tilde{Z}_i] := (1 - p_i)/p_i = \lambda_i/(\nu_i + \gamma)$. By conditioning on whether an individual recovers before the environment changes state or not, together with the fact that the system is Markovian, gives us

(i) $[Z_{ii} \mid T_i > \tilde{T}] \overset{\text{D}}{=} \tilde{Z}_i + 0,$

(ii) $[Z_{ii} \mid T_i \leq \tilde{T}] \overset{\text{D}}{=} \tilde{Z}_i + Z_{ji},$

(iii) $[Z_{ij} \mid T_i > \tilde{T}] \equiv 0,$

(iv) $[Z_{ij} \mid T_i \leq \tilde{T}] \overset{\text{D}}{=} Z_{jj}.$

Using (i)–(iv) yields

$$
\begin{aligned}
\mu_{ii} &:= \text{E}[Z_{ii} \mid T_i > \tilde{T}] \, \text{P}(T_i > \tilde{T}) + \text{E}[Z_{ii} \mid T_i \leq \tilde{T}] \, \text{P}(T_i \leq \tilde{T}) \\
&= \text{E}[\tilde{Z}_i] \frac{\gamma}{\nu_i + \gamma} + \text{E}[\tilde{Z}_i + Z_{ji}] \frac{\nu_i}{\nu_i + \gamma} \\
&= \text{E}[\tilde{Z}_i] + \frac{\nu_i}{\nu_i + \gamma} \text{E}[Z_{ji}] \\
&= \frac{\lambda_i}{\nu_i + \gamma} + \frac{\nu_i}{\nu_i + \gamma} \mu_{ji}
\end{aligned}
$$

and

$$
\mu_{ij} := \frac{\nu_i}{\nu_i + \gamma} \mu_{jj},
$$

which gives us

$$
\mu_{11} = \frac{\lambda_1}{\nu_1 + \gamma} + \mu_{21} \frac{\nu_1}{\nu_1 + \gamma}, \qquad \mu_{21} = \mu_{11} \frac{\nu_2}{\nu_2 + \gamma},
$$

and μ_{22} and μ_{12} follow by symmetry. Combining these relations implies that

$$
\mu_{ii} = \frac{\lambda_i (\nu_j + \gamma)}{\gamma (\nu_1 + \nu_2 + \gamma)}, \qquad \mu_{ij} = \frac{\lambda_j \nu_i}{\gamma (\nu_1 + \nu_2 + \gamma)}.
$$

The basic reproduction number, R_0, is then defined as the largest eigenvalue for $\boldsymbol{M} = \{\mu_{ij}\}_{i,j=1,2}$, i.e. the largest κ such that $\det(\boldsymbol{M} - \kappa \boldsymbol{I}) = 0$; see, e.g. [7, p. 730]. For this model, R_0 becomes

$$
R_0 = \frac{\mu_{11} + \mu_{22}}{2} + \sqrt{\left(\frac{\mu_{11} + \mu_{22}}{2} \right)^2 + \mu_{21}\mu_{12} - \mu_{11}\mu_{22}} =: w + \sqrt{w^2 + v}. \tag{12}
$$

If we consider conditions for $R_0 > 1$, this is equivalent to $w + \sqrt{w^2 + v} > 1$, which in turn is equivalent to $2w + v > 1$. Inserting v and w from (12) and simplifying then gives us

$$\frac{r_1(v_2 + \gamma) + r_2(v_1 + \gamma) - r_1 r_2 \gamma}{v_1 + v_2 + \gamma} > 1 \quad \Longrightarrow \quad \frac{r_1 - 1}{v_1}\left(\frac{v_2}{1 - r_2} + \gamma\right) > 1,$$

i.e. $R_\star > 1$, where $R_\star$ is given in (11). Thus,

$$R_\star > 1 \quad \Longleftrightarrow \quad R_0 > 1,$$

and, consequently, the same relation holds for $R_0 \leq 1$; hence, the two thresholds are equivalent. We believe that the conclusions are valid for general k, but it is not clear to us how to prove this.

Above we described how to approximate an epidemic in an alternating environment with two states, where the time periods that the environment spends in its different states are exponentially distributed, as a two-type epidemic. This could also be done for the case where the time periods follow some arbitrary distribution (with finite expectation). For this situation, however, we must keep track of the *time points* where individuals become infected, since this system is no longer Markovian. In terms of multitype epidemics this is equivalent to having a continuum of types, where the types now include information about the time point of infection.

The two-type epidemic described above is not identical to the alternating environment epidemic. The reason for this is that in the alternating environment model the number of individuals that two infectious individuals infect during their infectious periods are *dependent* if they are infectious during (partly) the same period. The reason for this dependence comes from the fact that the numbers of people they infect depend both on the environmental states and on how long they happen to last. In the corresponding two-type epidemic these numbers are by definition *independent*. However, when computing the basic reproduction number, R_0, the two models give the same result since this only concerns the mean values, and taking the mean of a sum is a linear operator even when the terms are dependent.

5.2. Probability of a large outbreak: two examples

In this section we give two examples in which it is possible to compute π, the probability of a large outbreak, numerically. The first example is a continuation of the situation treated in Section 5.1 when the time periods that the environment spends in its different states are exponentially distributed. In the second example the time periods that the environment spends in its different states are nonrandom. For both examples, π is obtained directly by numerically finding ρ, the smallest nonnegative solution to the equation $s = \chi(s)$, where $\chi(s)$ is given in (3), and using the relation $\pi := 1 - \rho$.

5.2.1. *Exponentially distributed time periods.* In Theorem 1 it was stated that the probability of a large outbreak is given as one minus the smallest nonnegative solution to the equation $s = \chi(s)$, where $\chi(s)$ is given in (3). As mentioned before, $\chi(s)$ is the generating function defining a certain Galton–Watson process which can be evaluated by using numerical integration.

In order to compare π, the probability of a large outbreak, for different parameter settings, we first keep $v_1 = v_2 = 1$ and set $\gamma = 1$ for simplicity, and vary λ_1 and λ_2 in such a way that $R_\star$ is kept fixed. One such parametrisation, assuming that $\lambda_1 > 1 > \lambda_2$, is

$$\lambda_1 = 1 + \delta \quad \text{and} \quad \lambda_2 = 2 - \frac{1}{R_\star(1 - \delta)}, \qquad 1 - \frac{1}{R_\star} < \delta < 1 - \frac{1}{2R_\star}, \qquad (13)$$

where $R_\star$ is given in (7). For a numerical illustration of how π varies with λ_1 and λ_2, see Figure 1, where it is seen that the probability of a large outbreak decreases as δ increases. On
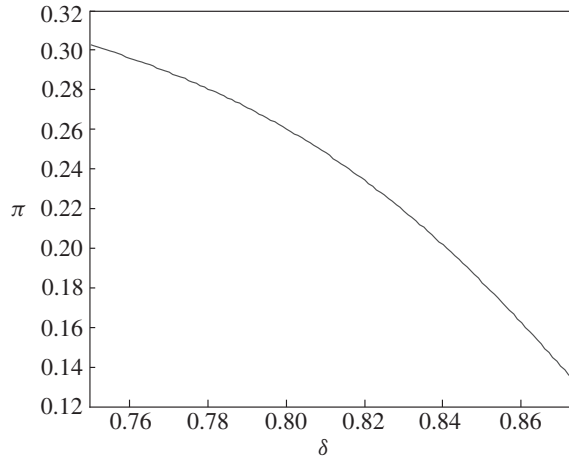
FIGURE 1: Exponentially distributed time periods. A numerical illustration of how π from Theorem 1 varies when $\lambda_1(\delta)$ and $\lambda_2(\delta)$ from (13) vary. Here $\gamma = 1$, $\nu_1 = \nu_2 = 1$, and $R_\star = 4$. On the y-axis we have π, and on the x-axis we have δ from (13).
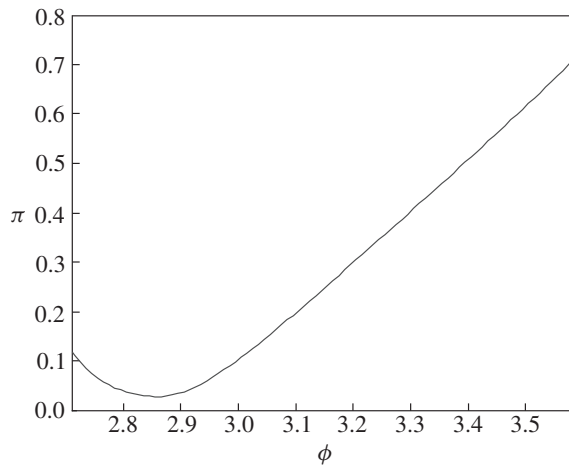

FIGURE 2: Exponentially distributed time periods. A numerical illustration of how π from Theorem 1 varies when $\nu_1(\phi)$ and $\nu_2(\phi)$ from (14) vary. Here $\gamma = 1$, $\nu = 1$, $\lambda_1 = 3.7$, $\lambda_2 = 0.4$, and $R_\star = 4$. On the y-axis we have π, and on the x-axis we have ϕ from (14).

the other hand, we can instead fix λ_1 and λ_2 and set $\gamma = 1$, and vary ν_1 and ν_2 in such a way that $R_\star$ is kept fixed. One such parametrisation, assuming that $\lambda_1 > 1 > \lambda_2$, is

$$\nu_1 = \nu\phi \quad \text{and} \quad \nu_2 = \frac{(1 - \lambda_2) R_\star (1 - (\lambda_1 - 1)/\nu\phi)}{1 - R_\star (1 - (\lambda_1 - 1)/\nu\phi)}, \qquad \frac{\lambda_1 - 1}{\nu} < \phi < \frac{\lambda_1 - 1}{\nu(1 - 1/R_\star)}.$$
$$(14)$$

A numerical illustration is given in Figure 2, from which it is seen that the behaviour of π is not necessarily monotonically increasing with increasing difference between ν_1 and ν_2.

5.2.2. *Nonrandom time periods.* For the situation when the time periods that the environmental process spends in its different states are nonrandom, the probability π is obtained by directly
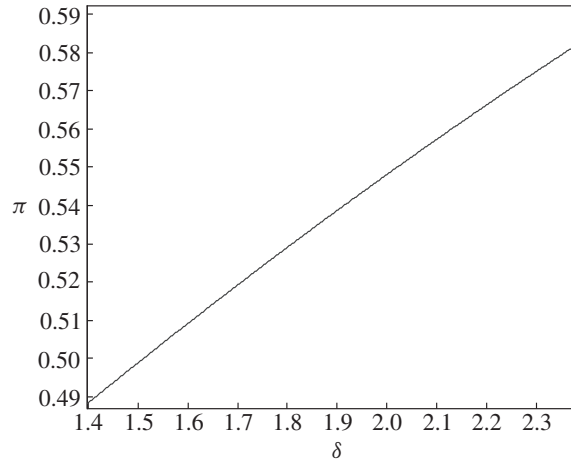
FIGURE 3: Nonrandom time periods. A numerical illustration of how π from Theorem 1 varies when $\lambda_1(\delta)$ and $\lambda_2(\delta)$ from (16) vary. Here $\gamma = 1$, $\nu_1 = \nu_2 = 1$, and $R_\star = 4$. On the y-axis we have π, and on the x-axis we have δ from (16).

solving the equation $s = \chi(s)$, where $\chi(s)$ is given in (3). As for the situation with exponentially distributed time periods, it is not possible to solve this equation analytically.

As in the previous subsection, we first want to vary λ_1 and λ_2 while all other parameters are fixed, including $R_\star$, and then vary ν_1 and ν_2 while everything else is kept fixed. First, from Theorem 1 we obtain, for $k = 2$,

$$R_\star := \exp\left\{\frac{\lambda_1 - \gamma}{\nu_1} + \frac{\lambda_2 - \gamma}{\nu_2}\right\}, \tag{15}$$

if $T_i \equiv 1/\nu_1$. Using this choice of T_i, we have the same expectation as for the exponential case treated in the previous subsection. As before, set $\nu_1 = \nu_2 = 1$ and $\gamma = 1$, and parametrise $\lambda_1 > 1 > \lambda_2$ according to

$$\lambda_1 = 1 + \delta \quad \text{and} \quad \lambda_2 = 1 - \delta + \log R_\star, \qquad \log R_\star < \delta < 1 + \log R_\star, \tag{16}$$

in order to keep $R_\star$ from (15) fixed. For a numerical illustration of how π varies with varying λ_1 and λ_2, see Figure 3, where it is seen that π increases with δ. Recall that in Figure 1 π instead was decreasing in δ. Thus, the distribution of the time periods that the environment spends in its different states may have a great impact on the behaviour of π.

If we instead fix $\lambda_1 > 1 > \lambda_2$ and set $\gamma = 1$, we can parametrise ν_1 and ν_2 as

$$\nu_1 = \nu\phi \quad \text{and} \quad \nu_2 = \frac{1 - \lambda_2}{(\lambda_1 - 1)/\nu\phi - \log R_\star}, \qquad 0 < \phi < \frac{\lambda_1 - 1}{\nu \log R_\star}, \tag{17}$$

keeping $R_\star$ from (15) fixed. In Figure 4 an illustration of how π varies with ν_1 and ν_2 is given, and it seen that π increases monotonically with increasing ϕ. In comparison with Figure 2, where nonmonotonic behaviour was observed, we again see significant differences in the behaviour of π depending on the distribution of the time that the environment spends in its different states.
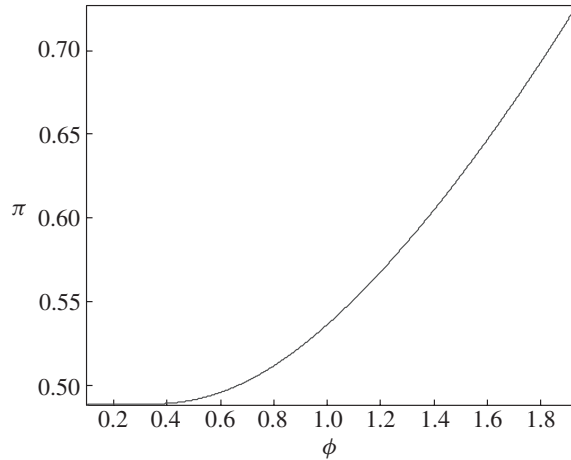
FIGURE 4: Nonrandom time periods. A numerical illustration of how π from Theorem 1 varies when $\nu_1(\phi)$ and $\nu_2(\phi)$ from (14) vary. Here $\gamma = 1$, $\nu = 1$, $\lambda_1 = 3.7$, $\lambda_2 = 0.4$, and $R_\star = 4$. On the y-axis we have π, and on the x-axis we have ϕ from (17).

To conclude, comparing Figures 1 and 2 with Figures 3 and 4, respectively, it is seen that there are dramatic differences in how π varies depending on the distribution of the time that the environmental process spends in its different states. As an example, by inspecting Figure 1 and Figure 3, we see that in the former figure π decreases monotonically, whereas in the latter figure π instead increases monotonically.

It is interesting to note that except for the nonmonotonic part of Figure 2, Figures 2 and 4 show similar behaviour. This could be an indication that changes in contact rate are more vulnerable to different distributions of the time periods than is the case for changes in the rate with which the environment switches state.

6. Discussion

In this paper we examined the threshold behaviour of an SIR model in a term-time forced environment. An 'environment' in this context is any external factor that influences the population, but itself is unaffected by the population. Here 'term-time forcing' refers to discontinuous but cyclic changes in the contact rate. We described this kind of system using the general SIR epidemic, but the traditionally single contact rate, λ, was replaced with a cyclically varying renewal process with k states, denoted by $\Lambda(t)$. For this model, we derived threshold conditions, denoted by $R_\star$, that could be expressed as products of the moment generating functions of the time periods that $\Lambda(t)$ spends in its different states. This was stated more precisely in Theorem 1. Hence, the quantity $R_\star$ can be stated explicitly and works as a threshold, such that a large outbreak can only occur if and only if $R_\star > 1$. Thus, $R_\star$ relates to the so-called basic reproduction number, R_0, in both of these aspects. The epidemiological interpretation of $R_\star$ is more involved than that of R_0, but from the derivation of $R_\star$ it was seen that it corresponded to the expected number of individuals that are infectious at the time point when the first environmental cycle is completed, given that the epidemic was initiated by a single infectious individual whose infectious period starts when a new environmental cycle is started.

By checking whether $R_\star$ is greater than 1 or not, we easily determined whether a large outbreak could occur or not. That is, $R_\star$ determines whether the probability of a large outbreak,

π, is strictly greater than 0 or not. The probability π is, however, much harder to compute than $R_\star$, since it is defined as 1 minus the smallest nonnegative solution to $s = \chi(s)$, where $\chi(s)$ is given in (3). In Section 5 we gave two examples in which we computed π using numerical integration. In the first example all time periods that $\Lambda(t)$ spends in its different states were exponentially distributed. This was done in detail for the case in which $k = 2$, and we also showed that R_0 and $R_\star$ are equivalent in terms of their threshold behaviour. We believe that this is true for general k. In the second example the time periods that the environmental process spends in its different states were nonrandom. By comparing the numerically evaluated values of π for the two cases (see Figures 1–4), it was apparent that the distribution of the time that the environmental process spends in its different states is of importance.

In Corollary 1 we gave some qualitative relations between the behaviour of the general SIR epidemic in a term-time forced environment and the general SIR epidemic with the single contact rate $\tilde{\lambda}$ from (6) corresponding to the long-time average environmental effect. By denoting $R_\star$ of the latter process by $\tilde{R}_\star$, we showed that $R_\star$ of the epidemic, with the environment accounted for, is always greater than $\tilde{R}_\star$, except when all time periods that $\Lambda(t)$ spends in its different states are nonrandom, when $R_\star = \tilde{R}_\star$ holds.

The derivation of these results relied on the theory of branching processes, and the coupling between such processes and epidemic processes. The strongest form of coupling between epidemics and branching processes is given in [8, Theorem 2.1, p. 4], but, for our purposes, it was sufficient to obtain a weaker coupling. More precisely, we only needed convergence in distribution between the ultimate number of infected individuals in the epidemic process and the ultimate number of individuals born in a certain branching process. This was stated in Theorem 2.

Throughout, we made repeated use of the fact that the system is conditionally Markovian in its nature. Consequently, the methods used in the present paper are hard to adapt directly to situations with nonexponential infectious periods, since the Markov property will then be lost. One possible generalisation could be to allow gamma distributed infectious periods. This could possibly be done by subdividing the infectious period into a number of shorter exponentially distributed infectious periods, similarly to what was done in [3]. In this way the Markov property of the system would be preserved. For many real-life applications, the flexibility of the gamma distribution will suffice from a modelling perspective. But, as soon as the system becomes non-Markovian, the approximating branching processes will be of Crump–Mode–Jagers type, and, hence, depend on the *distribution* of the time points of renewal. It might, however, be possible to derive some threshold conditions using embedding techniques, but it is not clear to us how to incorporate the environmental effects into such embeddings.

Assuming that the Markov property is fulfilled, it is possible to use these methods when the environment depends on the population. Perhaps the simplest example corresponds to a naive model for social awareness in which the entire population first becomes aware that a disease is present when some given fraction of the population has become infected. In this situation, the process would still be conditionally Markovian, and it is possible to derive closed, but implicit, threshold conditions. But, these expressions would instead depend on the distribution of the time it takes until this fraction of the population has become infected, which in general is hard to obtain. In simple situations it might be possible to go via generating functions and relations between such functions and orthogonal polynomials similar to those used in [17].

Recall that it was assumed that the environment starts with environment 1, i.e. $\Lambda(0) = \lambda_1$. Of course, if the environment instead starts in λ_i, we can simply relabel the states. A question of interest however is how the probability π of a large outbreak depends on which environment

comes first. If there are only two environments, it is not hard to realise that π is bigger if the epidemic is started in the more socially active environment (the environment with larger λ_i). If there are more than two environments, we cannot make any similar general statements; the answer will depend on how different the λs are, the time distributions of the environments and their ordering. The general trend, that π is bigger if the socially active environments come first, still holds true.

Even though we have only treated nonendemic situations in the present paper, we can still determine whether or not a disease may become endemic, by using $R_\star$ from Theorem 1. A natural continuation of the present work would be to analyse endemic diseases under the influence of varying environments in more detail, both in terms of endemic levels, fluctuations, and the time to disease extinction.

Acknowledgements

References

[1] ALLEN, L. J. S. AND CORMIER, P. J. (1996). Environmentally driven epizootics. *Math. Biosci.* **131,** 51–80.

[2] ANDERSSON, H. AND BRITTON, T. (2000). *Stochastic Epidemic Models and Their Statistical Analysis* (Lecture Notes Statist. **151**). Springer, New York.

[3] ANDERSSON, H. AND BRITTON, T. (2000). Stochastic epidemics in dynamic populations: quasi-stationarity and extinction. *J. Math. Biol.* **41,** 559–580.

[4] ANDERSON, R. M. AND MAY, R. M. (1991). *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press.

[5] ATHREYA, K. B. AND NEY, P. E. (1972). *Branching Processes*. Springer, New York.

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

[7] BALL, F. AND CLANCY, D. (1993). The final size and severity of a generalised stochastic multitype epidemic model. *Adv. Appl. Prob.* **25,** 721–736.

[8] BALL, F. AND DONNELY, P. (1995). Strong approximations for epidemic models. *Stoch. Process. Appl.* **55,** 1–21.

[9] BRITTON, T., DEIJFEN, M., LAGERÅS, A. N. AND LINDHOLM, M. (2008). Epidemics on random graphs with tunable clustering. *J. Appl. Prob.* **45,** 743–756.

[10] DURRETT, R. (2004). *Probability: Theory and Examples*, 3rd edn. Duxbury Press, Belmont, CA.

[11] EARN, D. J. D., ROHANI, P., BOLKER, B. M. AND GRENFELL, B. T. (2000). A simple model for complex dynamical transitions in epidemics. *Science* **287,** 667–670.

[12] FELLER, W. (1957). *An Introduction to Probability Theory and Its Applications*, Vol. I, 2nd edn. John Wiley, New York.

[13] FRANKE, J. E. AND YAKUBU, A.-Z. (2006). Discrete-time SIS epidemic model in a seasonal environment. *SIAM J. Appl. Math.* **66,** 1563–1587.

[14] HACCOU, P., JAGERS, P. AND VATUTIN, V. A. (2005). *Branching Processes: Variation, Growth, and Extinction of Populations*. Cambridge University Press.

[15] HARRIS, T. E. (1963). *The Theory of Branching Processes*. Springer, Berlin.

[16] JAGERS, P. (1975). *Branching Processes with Biological Applications*. John Wiley, London.

[17] KARLIN, S. AND MCGREGOR, J. (1958). Linear growth, birth and death processes. *J. Math. Mech.* **7,** 643–662.

[18] KEELING, M. J., ROHANI, P. AND GRENFELL, B. T. (2001). Seasonally forced dynamics explored as switching between attractors. *Physica D* **148,** 317–335.

[19] KUSKE, R., GORDILLO, L. F. AND GREENWOOD, P. (2007). Sustained oscillations via coherence resonance in SIR. *J. Theoret. Biol.* **245,** 459–469.

[20] MOLLISON, D. (1977). Spatial contact models for ecological and epidemic spread. *J. R. Statist. Soc. B* **39,** 283–326.

[21] PIYAWONG, W., TWIZELL, E. H. AND GUMEL, A. B. (2003). An unconditionally convergent finite-difference scheme for the SIR model. *Appl. Math. Comput.* **146,** 611–625.

[22] PRAJNESHU, GUPTA, C. K. AND SHARMA, U. (1986). A stochastic epidemic model with seasonal variations in infection rate. *Biometrical J.* **28,** 889–895.

[23] STONE, L., OLINKY, R. AND HUPPERT, A. (2007). Seasonal dynamics of recurrent epidemics. *Nature* **446,** 533–536.