Mathematical Statistics
Stockholm University

Endemic persistence or disease extinction: the effect of separation into subcommunities

Mathias Lindholm
Tom Britton

Research Report 2006:6

Postal address:
Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:
http://www.math.su.se/matstat

Endemic persistence or disease extinction: the effect of separation into subcommunities

Mathias Lindholm[*]
Tom Britton[†]

September 2006

Abstract

Consider an infectious disease which is endemic in a population divided into several large subcommunities that interact. Our aim is to understand how the time to extinction is affected by the level of interaction between communities.

We present two approximations for the expected time to extinction in a population consisting of a small number of large subcommunities. These approximations are derived for an SIR epidemic model, with focus on diseases with short infectious period in relation to life length, such as childhood diseases and influenza. Both approximations are based on Markov jump processes.

Simulations indicate that the time to extinction is increasing in the degree of interaction between communities. This behaviour can also be seen in our approximations in relevant regions of the parameter space.

KEY WORDS: Endemic diseases, SIR-epidemic, Population heterogeneity, Expected time to extinction, Quasi-stationary distribution

---

[*]Postal address: Mathematical Statistics, Stockholm University, SE-106 91 Stockholm, Sweden. E-mail: lindholm@math.su.se.

[†]Postal address: Mathematical Statistics, Stockholm University, SE-106 91 Stockholm, Sweden. E-mail: tomb@math.su.se

1 Introduction

When modelling infectious diseases a simplifying assumption often made is that the social interaction within the population is homogeneous. This assumption becomes less credible as the population size increases, i.e. there is a need for including some kind of population heterogeneity. A way to include population heterogeneity is to divide the population into subcommunities. By doing so we can, in the easiest setting, allow two different levels of social interaction, one level within and one between subcommunities. Under these conditions it is natural to let the social interaction within subcommunities be homogeneous. In the present paper we are interested in the situation when there are k subcommunities each of size n, and to study the effect of this new level of social interaction on the epidemic behaviour as an infectious disease is introduced into the population. Here, typical values of k is $2, \ldots, 5$ and n is 50,000 or larger. Throughout this report we will focus on infectious diseases that have a short infectious period in relation to life length and give rise to life long immunity, e.g. childhood diseases. When an outbreak of such a disease occurs in a community, we have three possible scenarios. The first being that only a few become infected and the time to extinction is short. The second one being that many become infected but the time to extinction is short. We are interested in diseases that behave as in the third scenario, namely when many individuals become infected and the time to extinction is long. When a disease behaves in this way it is called endemic. Dynamically, an endemic disease only has a rather small fraction of the population infectious at each time point during its progression, but the accumulated number of infected individuals may still be large due to that the disease is persistent in the community during a long time period. Usually the fraction of infected individuals at each time point fluctuates around some specific level, the endemic level, until disease extinction.

Whether a disease becomes endemic or not depends on a number of factors. Possibly the most important one, apart from that we need a sufficiently large population, is the relation between social activity and the infectiousness of the disease. This relation is usually quantified as the so called basic reproduction number. The basic reproduction number, R_0, is defined as the expected number of individuals that a single infectious individual infects in a large susceptible community. One can show that the basic reproduction number works as a threshold which determines the dynamics of the disease, see Anderson and May (1992) pp. 13-19. If $R_0 \leq 1$, the disease will go extinct rather quickly. On the other hand, if $R_0 > 1$, the disease has a positive probability to persist in the population over a long time period. Henceforth R_0 is assumed to be greater than one.

When dividing the population into subcommunities, the dynamics of the spread of disease becomes more intricate. Now, some subcommunities may be disease-free, while others contain infected individuals, and infectious contacts between individuals from different subcommunities are also possible. Another important feature in the dynamics is that subcommunities can get reinfected several times. When thinking of this behaviour heuristically, it seems reasonable to expect that the mean time to extinction of an endemic disease that affects the entire population will depend on the social activity between the different communities when keeping everything else fixed. This can be shown to be true, see Hagenaars et al. (2004).

Endemic diseases can be modelled stochastically in several different ways. Depending on

the model, different aspects of the qualitative behaviour of the dynamics of the disease can be studied. Our aim is to study the expected time to extinction of an endemic disease in the situation with a small number of large subcommunities, when each subcommunity starts at the so-called endemic level. The model used is an SIR model with demography in the situation with subcommunities. For the case with only a single large community, much work has been done, see for example van Herwaarden and Grasman (1995), Nåsell (1999, 2005), and Andersson and Britton (2000b). For more on epidemic models in general, see Anderson and May (1992), Andersson and Britton (2000a) and Diekmann and Heesterbeek (2000).

A short heuristic description of the SIR model with demography is that all individuals in the community are equally likely to meet, and that each individual may switch between being Susceptible, Infectious and Recovered (and immune). Thus, switches occur according to S → I → R. Another important property is that susceptible individuals are born into the community and that individuals eventually die, i.e. demographic aspects. This will also give us a non constant community size. The version of this model which we use is from Haagenars et al. (2004), and does not allow for birth of infectious individuals. Important results for the single community case that we will use are from Nåsell (1999, 2005). There approximations for the expected time to extinction when starting at the quasi-stationary level of infection are derived. In Hagenaars et al. (2004) they study the same expected time to extinction as Nåsell but for the case with a small number of subcommunities. They obtain an approximation of this expected time, but the approximation is derived under the assumption of low mixing between communities and that the infectious period is long in relation to life length.

In the present paper we have adopted ideas from both Nåsell (1999) and Hagenaars et al. (2004), trying to find better approximations for the expected time to extinction for the case with subcommunities for diseases with short infectious period with respect to life length. We present two approximations, the first one is based on similar arguments as in Hagenaars et al. and the second one is based on more heuristic arguments motivating the use of an exponential form.

We derive approximations of the expected time to extinction given that the process is started at the quasi-stationary (endemic) level of infection because we can then approximate the *distribution* of T_Q, the time to extinction given that the process is started in the quasi-stationary distribution. One can show that T_Q is exponentially distributed, and by approximating $\tau = E(T_Q)$, we get that $T_Q \sim \text{Exp}(1/\tau)$. Simulations indicate that our approximations are more suitable for situations when there is low mixing between subcommunities and that the second, more heuristically motivated approximation, performs somewhat better.

In sections 2 we define the SIR model with demography for the case with subcommunities and present results needed later on. In sections 3 and 4 we present the approximation from Hagenaars et al. (2004) and we derive our two approximations. Section 5 is devoted to a small simulation study and a numerical evaluation of these approximations. A closing discussion and summary of our results is given in section 6.

2 The SIR model with demography and subcommunities

Let's start with a brief look at the SIR model with demography for the single community case from Nåsell (1999) p. 311. As mentioned in the introduction, the letters S, I and R, stand for Susceptible, Infectious, and Recovered (and immune). These are the three possible states an individual can experience during an epidemic. The demographic aspect of the model is that we allow individuals to be born into the population, as susceptibles, and that individuals also may die. Susceptibles are born into the population according to a Poisson process with constant birth intensity μn, and all individuals live for an exponentially distributed time with mean $1/\mu$. A consequence of the constant birth rate is that the population size will fluctuate around n, which is thought of as being large. This is deliberate to avoid that the dynamics of the disease depend on extensive population fluctuations.

A consequence of the fact that we only let individuals be born as susceptibles is that once the population becomes disease-free, it will remain so forever on. In other words, the disase free states are absorbing, and all other states are transient.

When adding population separation to the SIR model with demography, the dynamics of the disease becomes more involved. The idea is that we now have a population that is divided into subcommunities, and individuals in different subcommunities may contact each other at different rates.

The easiest way to model this situation, based on the SIR model with demography, is to let all subcommunities be equally large, having size n, and to let individuals from different subcommunities be equally likely to meet, regardless of which subcommunities they belong to. We are interested in the situation when the number of subcommunities, k, is fixed and small in relation to n. With this model the population structure is symmetric and we only need to add one parameter, ε, which is the proportion of an individuals contacts that are with other subcommunities. This parameter, ε, is defined such that $\varepsilon = 0$ corresponds to having k isolated subcommunities, and $\varepsilon = 1$ corresponds to the case where all k subcommunities act as a single large community of size kn. One can also think of ε as an inverse distance, where $\varepsilon = 0$ corresponds to that all subcommunities lie infinitely far apart and $\varepsilon = 1$ corresponds to the case when they coincide.

A natural way to model the situation with subcommunities is to do so such that the overall infectious pressure in the entire population is kept constant regardless of the value of ε. This also has the advantage that we get the same R_0 as for the SIR model with demography without subcommunities and hence the two models become easier to compare. The basic reproduction number for the single community SIR model with demography can be shown to be

$$R_0 = \frac{\beta}{\mu\alpha}, \tag{2.1}$$

where $\alpha = (\mu + \nu)/\mu$, see Eq. (2.2) on p. 311 in Nåsell (1999). For the case with subcommunities, an infected individual makes contacts with any given individual within its own

subcommunity at rate β'/n, and at rate $\varepsilon\beta'/n$ with a given individual in any of the $k-1$ surrounding subcommunities. Thus, we get that the probability that a contact is within the subcommunity is

$$\frac{n\beta'/n}{n\beta'/n + (k-1)n\varepsilon\beta'/n} = \frac{1}{1 + \varepsilon(k-1)}.$$

If we have a single infected individual in an otherwise susceptible population, she will infect a given individual within her own subcommunity at rate β'/n during an exponentially distributed infectious period with mean $1/(\mu + \nu)$, since the infectious period terminates due to recovery or death, and infect a given individual in any of the $k-1$ neighbouring subcommunities at rate $\varepsilon\beta'/n$, hence the basic reproduction number becomes

$$R_0 = \frac{1}{\mu + \nu}(n\beta'/n + (k-1)n\varepsilon\beta'/n) = \frac{\beta'}{\mu + \nu}(1 + \varepsilon(k-1)).$$

Thus, if we let $\beta' = \beta/(1 + \varepsilon(k-1))$ we arrive at

$$R_0 = \frac{\beta}{\mu\alpha}$$

as desired, since $\mu\alpha = \mu + \nu$.

Hence, $\beta' = \beta/(1 + \varepsilon(k-1))$ is the proper scale in order to keep the infectious pressure constant. The possible transitions and their rates are specified in Table 1, which are the same as in Hagenaars et al. (2004).

As for the dynamics of an endemic disease, we indicated in the introduction that there might exist some kind of equilibrium in terms of number of infected individuals, the quasi-stationary level of infection. Let $(\mathbf{X}(t), \mathbf{Y}(t)), t \geq 0$, denote a $2k$ dimensional Markov jump process, where $X_j(t) = s_j$ and $Y_j(t) = i_j$ denote the number of susceptibles and infectious in subcommunity j at time t, with random transition rates defined in Table 1.

If we instead look at the process of proportions $(\mathbf{X}(t)/n, \mathbf{Y}(t)/n) = (\mathbf{x}, \mathbf{y})$, when n is large, this process can be approximated by the solution of a deterministic system of differential equations corresponding to the transition rates defined in Table 1. This system is given by

$$\frac{dx_i}{dt} = \mu - \frac{\beta}{(1 + \varepsilon(k-1))}x_i\left(y_i + \varepsilon\sum_{j\neq i}y_j\right) - \mu x_i$$

$$\frac{dy_i}{dt} = \frac{\beta}{(1 + \varepsilon(k-1))}x_i\left(y_i + \varepsilon\sum_{j\neq i}y_j\right) - (\mu + \nu)y_i.$$

Setting these equations equal to zero for $i = 1, \ldots, k$ gives us the stationary points, which turn out to be $(1, 0)$, the disease-free state, and

$$(\hat{x}_i, \hat{y}_i) = (\hat{x}, \hat{y}) = \left(\frac{1}{R_0}, \frac{1}{\alpha}\left(1 - \frac{1}{R_0}\right)\right), \tag{2.2}$$

Table 1: SIR model with demography for the case with subcommunities

Rates for community j		
From	To	Rate
(s_j, i_j)	$(s_j + 1, i_j)$	μn
(s_j, i_j)	$(s_j - 1, i_j)$	μs_j
(s_j, i_j)	$(s_j - 1, i_j + 1)$	$\frac{\beta}{n} \frac{1}{(1+\varepsilon(k-1))} s_j \left(i_j + \varepsilon \sum_{u \neq j} i_u \right)$
(s_j, i_j)	$(s_j, i_j - 1)$	$(\mu + \nu) i_j$

which corresponds to the endemic level, and which only exists if $R_0 > 1$. These two stationary points are the same as for the single community case, see Nåsell (1999) p. 312.

The term quasi-stationary level of infection originates from something called the quasi-stationary distribution. A quasi-stationary distribution is defined as the distribution after a long time conditioned on that the process has not been absorbed. The endemic level can be thought of as the mean of this distribution, which the process fluctuates around. The quasi-stationary distribution is important when modelling endemic diseases, since we are interested in the behaviour of the epidemic until it goes extinct. But, quasi-stationary distributions give rise to many difficulties such as questions of uniqueness and existence, see Pollett and Roberts (1990).

Let $Q = \{q_{\mathbf{x},\mathbf{y}}\}$ denote the quasi-stationary distribution, where $q_{\mathbf{x},\mathbf{y}}$ is the probability that the process $(\mathbf{X}(t), \mathbf{Y}(t))$ is in $(\mathbf{x}, \mathbf{y})$ as $t \to \infty$, conditioned on not having gone extinct. Remembering that the lack of memory property implies an exponential distribution, we have

$$
\begin{aligned}
P(T_Q > t + s \mid T_Q > t, \ (\mathbf{X}(0), \mathbf{Y}(0)) \sim Q) \\
= P(T_Q > t + s \mid T_Q > t, \ (\mathbf{X}(t), \mathbf{Y}(t)) \sim Q) \\
= P(T_Q > s \mid (\mathbf{X}(0), \mathbf{Y}(0)) \sim Q),
\end{aligned}
$$

which establishes that T_Q is exponentially distributed. The rate parameter for this exponential distribution is the intensity with which the process leaves the set of transient states. For the case with subcommunities the set of states from which the process can be absorbed is $\mathbf{X}(t), \mathbf{Y}(t) = \{(\mathbf{x}, \mathbf{y}); \mathbf{y} = \mathbf{e}_i, i = 1, \ldots, k\}$.

Proposition 2.1 *The time to extinction given that the process is started in the quasi-stationary distribution, T_Q, is exponentially distributed with mean*

$$
\tau = \frac{1}{\mu \alpha q_{\cdot,1}}, \tag{2.3}
$$

where

$$
q_{\cdot,1} = \sum_{\mathbf{x}} \sum_{i=1}^{k} q_{\mathbf{x}, \mathbf{e}_i}, \tag{2.4}
$$

6

and where $\mathbf{e}_i$ is the i'th unit vector.

The reasoning here is the same as in the proof of Proposition 4.1 in Andersson and Britton (2000b), but this result was first derived for the single community case in Nåsell (1999), and if we set $k = 1$ in Proposition 2.1 we get the single community case. In Nåsell (1999) he approximates $q_{.,1}$ with the stationary distribution of the diffusion approximation of $(X(t), Y(t))$, conditioning on non extinction. The resulting diffusion approximation is of Ornstein-Uhlenbeck type, and thus the stationary distribution is Gaussian. By doing so he was able to approximate τ, and thereby determine the distribution of T_Q. More precisely, when the quasi-stationary distribution is approximated with a truncated normal distribution, T_Q is exponentially distributed with mean

$$\tau_n = \frac{\sigma_Y}{\mu\alpha} \frac{\Phi((\mu_Y - 0.5)/\sigma_Y)}{\varphi((\mu_Y - 1)/\sigma_Y)} \tag{2.5}$$

where $\Phi(\cdot)$ and $\varphi(\cdot)$ are the standard normal distribution function and density function, and

$$\begin{aligned} \sigma_Y &= \frac{\sqrt{n}}{R_0}\sqrt{R_0 - 1 + R_0^2/\alpha} \\ \mu_Y &= n\frac{R_0 - 1}{\alpha R_0}. \end{aligned} \tag{2.6}$$

This corresponds to Eqs. (2.13) and (2.10) in Nåsell (1999). From here on τ_n refers to the single community case with population size n, and all other types of references to τ are for the case with subcommunities unless otherwise stated.

However, when the average life length is long in relation to average infectious period, Nåsell (2005) shows that (2.5) is a too crude approximation when n is only moderately large, smaller than say a couple of million individuals. In Nåsell (2005) he instead proposes that the quasi-stationary distribution of the number of infected individuals could be approximated with a geometric distribution with $p = 1/\mu_Y$ where μ_Y is from (2.6). As a consequence of that $Y \sim Ge(p)$ and $E(Y) = 1/p = \mu_Y$ together with Proposition 2.1 with $k = 1$ yields the following: When the quasi-stationary distribution is approximated with a $Ge(1/\mu_Y)$ distribution with mean μ_Y from (2.6), then T_Q is exponentially distributed with mean

$$\tau_n = n\frac{R_0 - 1}{\mu\alpha^2 R_0}. \tag{2.7}$$

This corresponds to Eqs. (8.3) and (9.2) in Nåsell (2005).

Returning to the case with subcommunities again, we proceed as in Nåsell (1999) or (2005) to approximate the quasi-stationary distribution, but due to symmetry, the resulting Ornstein-Uhlenbeck diffusion process approximated in the endemic level is independent of ε, and hence not of much help. One sees this when replacing all s_j and i_j in the rates defined in Table 1, by $n\hat{x}$ and $n\hat{y}$. Due to this, the second approach will also give us an approximation of the quasi-stationary distribution which is independent of ε, since it

was a geometric distribution with parameter $p = 1/\mu_y$, where μ_y is as (2.6), but with n replaced with kn.

However, we can still say something about the two extreme cases, $\varepsilon = 0$ and $\varepsilon = 1$, with the results from the single community case. When $\varepsilon = 0$ all k subcommunities are isolated. When all k subcommunities start at the endemic level of infection, the expected time until one of the k infected subcommunities recovers is τ_n/k, due to independence and that the expected duration of an epidemic within a subcommunity is exponentially distributed with mean τ_n, where τ_n is from one of Eqs. (2.5) or (2.7). Due to the Markov property and that a disease-free community never can be reinfected when $\varepsilon = 0$, the expected time until one of the $k - 1$ remaining communities recovers is $\tau_n/(k-1)$. Repeating this argument gives us

$$\tau(0) = \tau_n \sum_{i=1}^{k} \frac{1}{i}, \tag{2.8}$$

where τ_n is either of Nåsell's approximations of the expected time to extinction for the single community case, Eqs. (2.5) or (2.7). Note that $\tau(\varepsilon)$ in fact is a function of the form $\tau(\varepsilon) = \tau(\varepsilon, n, k, \mu, \alpha, R_0)$ and denotes the expected time to extinction when each of the k subcommunities are started at the quasi-stationary level of infection, each having size n, and having a proportion ε of the contacts being with other subcommunities.

On the other hand, when $\varepsilon = 1$, all k communities behave as one large community of size kn, and we can again make use of (2.5) with n replaced by kn, i.e.

$$\tau(1) = \tau_{kn}. \tag{2.9}$$

For sufficiently large n, we have the following relation between the two

$$\tau(0) < \tau(1),$$

see Hagenaars et al. (2004).

3 Approximation using a recovered (and immune) state

As we have seen, it is hard to find approximations of the quasi-stationary distribution which depend on ε. But, if we rely on Proposition 2.1, that T_Q is exponentially distributed, we can approximate $\tau = E(T_Q)$ directly, instead of going via approximations of the quasi-stationary distribution.

In Hagenaars et al. (2004) they look at the case when $0 < \varepsilon \ll 1$ and α is thought of as small, such as $\alpha = 2$ or 160. An example of a disease with small α is scrapie among sheep, see Hagenaars et al. (2004). For scrapie the average incubation period is

a few years which is of the same order of magnitude as the average life length of sheep. Hence, for diseases with small α one can assume that when an individual recovers from infection, she will likely be removed due to death within a relatively short time period. This motivates that we can look at the system from a subcommunity view, classifying each subcommunity as either endemic or susceptible. That a subcommunity is endemic here means that the subcommunity on average has a fraction of infected individuals corresponding to the endemic level $\hat{y}$. A subcommunity that is susceptible only contains susceptible individuals. Hagenaars et al. (2004) further assume that switches between these two states occur instantaneously. This is a reasonable approximation, since the time it takes from that a single individual becomes infected until the endemic level of infection is reached is short in relation to the time it takes for an endemic subcommunity to become disease-free.

When defining the rate with which susceptible subcommunities becomes endemic, it is natural to think that this rate depends on the infectious pressure generated by the endemic subcommunities. But, we are only interested in those infectious contacts between subcommunities that result in a disease invasion and not those that fade out by chance, so we must take this fact into account. We will give a short heuristic derivation of this probability for the case when a subcommunity has a fraction x susceptibles and $1 - x$ recovered (and immune) individuals, because this more general result is needed later on.

Suppose a subcommunity with a fraction x susceptibles and $1-x$ recovered (and immune) individuals has just been reinfected, i.e. a single susceptible becoming infected. In the early stages of an epidemic it behaves approximately as a branching process. In our model the infectious period is exponentially distributed. When there is only one infected individual in a population with a fraction x susceptible individuals, the effective reproduction number in this situation becomes xR_0. If we let D denote the number of children of this one infected individual, $D \sim Geo(1/(1 + xR_0))$, we get that the probability that the epidemic started by this single infected individual will not fade out by chance, p, is the solution to the following equation:

$$1 - p = E((1 - p)^D),\tag{3.1}$$

see Andersson and Britton (2000a) pp. 22-25. Solving this gives us the solution

$$p = 1 - \frac{1}{xR_0}.\tag{3.2}$$

From this we get that the probability that disease invasion will not fade out by chance in a fully susceptible population is $1 - 1/R_0$.

One individual contacts a given individual in a different subcommunity at rate $\varepsilon\beta'/n$, and hence contacts a whole subcommunity at rate $\varepsilon\beta'$. Consequently, a subcommunity at the endemic level, having $\hat{y}n$ infectives, infects a given susceptible subcommunity at rate $\varepsilon\beta'\hat{y}n = \varepsilon\beta\hat{y}n/(1 + \varepsilon(k - 1))$. Thus, each endemic subcommunity generates an infectious pressure $\varepsilon\beta\hat{y}n/(1 + \varepsilon(k - 1))$ on each of the surrounding susceptible subcommunities, where each infectious contact has the probability $1 - 1/R_0$ that the introduced disease will become endemic.

9

If we again look at the rate with which subcommunities becomes disease-free, this is thought of occuring independently of everything else, i.e. the time to disease extinction in a subcommunity is exponentially distributed with mean parameter τ. From this we can define a birth and death process of number of endemic subcommunities, with transition rates

$$\begin{cases} \zeta_j &= (k-j)j\varepsilon n \frac{\mu R_0}{1+\varepsilon(k-1)} \left(1 - \frac{1}{R_0}\right)^2 \\ \eta_j &= j/\tau_n \end{cases} \tag{3.3}$$

where ζ_j is the rate for a transition from j to $j+1$ endemic subcommunities, and η_j is the rate for a transition from j to $j-1$ endemic subcommunities.

Since ε is small, the probability of reinfection will also be small. Based on this fact Hagenaars et al. (2004) assume that the probability of more than one reinfection during the epidemic is negligible. Their approximation can be described as the expected time to absorption of a birth and death process for the number of endemic subcommunities, with rates as in (3.3), which only allow one birth, or more formally:

Approximation (Hagenaars et al. (2004)) *The expected time to extinction given that the process is started at the endemic level can be approximated by*

$$\tau_{\mathrm{SI}}(\varepsilon) = \tau(0) + \varepsilon\left((k+1)\sum_{j=1}^{k}\frac{1}{j} - 2k\right)\tau_n^2 \mu R_0 \left(1 - \frac{1}{R_0}\right)^2 + O(\varepsilon^2) \tag{3.4}$$

where $\tau(0)$ is from (2.8) and τ_n is any approximation for a single community of size n, e.g. (2.5) or (2.7).

This corresponds to Eq. (6) in Hagenaars et al. (2004). Here SI in τ_{SI} is used to emphasise that they only use the two subcommunity states, Susceptible and Infected, in their approximation.

Remember that our aim is to find approximations for the expected time to extinction when α is large, corresponding to an average infectious period of one to two weeks for humans, as opposed to Hagenaars et al. (2004), and then their approach is not completly feasible. This is because when α is large, the approximation that an endemic subcommunity that becomes disease-free instantaneously becomes susceptible is not reasonable. One way to avoid this problem is to add a recovered (and immune) state to our approximating subcommunity Markov process. A subcommunity is defined as being recovered (and immune) when it is disease-free but not possible to infect. The difference between this state and the susceptible state is that, when a subcommunity is recovered (and immune) there is on average a fraction $\hat{x}$ susceptible and $1 - \hat{x}$ immune individuals, as opposed to the susceptible state which only contain susceptible individuals.

By introducing this type of transitions for the subcommunities we have a communication between the states of subcommunities that can be described as $S \to I \to R \to S$, so what we need to define is the rate, ξ, with which a community makes a transition from R to S, since (3.3) can be used for the other transitions.
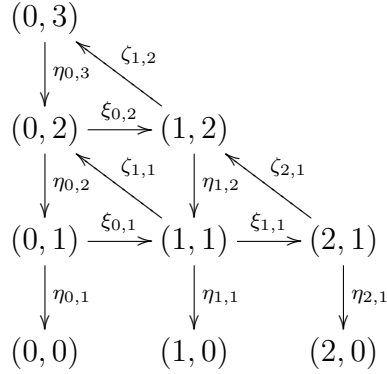
Figure 1: Schematic graph of the dynamics in our approximating SIR Markov jump process for $k = 3$ subcommunities, where each node is (no. of susceptible subcommunities, no. of endemic subcommunities) and the rates are from (3.5).

One way to do so is to assume that a subcommunity stays immune for an exponentially distributed time with mean τ_{R}, which is equivalent to that an immune subcommunity becomes susceptible again independently of everything else. We will will return to the definition of τ_{R} later.

Let s be the number of susceptible subcommunities and i be the number of endemic subcommunities out of a total of k subcommunities, so that $k - (s+i)$ are recovered (and immune), then the transition rates become

$$\begin{cases} \zeta_{s,i} &= si\varepsilon \frac{\mu n R_0}{1+\varepsilon(k-1)}\left(1 - \frac{1}{R_0}\right)^2 \\ \eta_{s,i} &= \frac{s}{\tau_n} \\ \xi_{s,i} &= \frac{k-(s+i)}{\tau_{\mathrm{R}}}. \end{cases} \tag{3.5}$$

There are $k(k+1)/2 + k$ possible states, and k of them are disease-free and hence an absorbing class of states. For a schematic graph of the dynamics of this process, see Fig. 1.

Based on the rates (3.5) we are able to set up a difference equation $\tilde{t}_{s,i}$, the expected time to extinction when starting with i endemic and s susceptible subcommunities out of k possible, by conditioning on the first transition. We then get the following relation

$$\tilde{t}_{s,i} = \frac{1}{\zeta_{s,i} + \eta_{s,i} + \xi_{s,i}} + \frac{\zeta_{s,i}}{\zeta_{s,i} + \eta_{s,i} + \xi_{s,i}}\tilde{t}_{s-1,i+1}$$
$$+ \frac{\eta_{s,i}}{\zeta_{s,i} + \eta_{s,i} + \xi_{s,i}}\tilde{t}_{s,i-1} + \frac{\xi_{s,i}}{\zeta_{s,i} + \eta_{s,i} + \xi_{s,i}}\tilde{t}_{s+1,i}. \tag{3.6}$$

In general this system has no closed form solution. But, by looking at the transition rates (3.5) and the relation (3.6), we see that we can write this as an equation system of the

11

form

$$\tilde{\mathbf{t}} = \mathbf{v} + \mathbf{A}\tilde{\mathbf{t}}, \tag{3.7}$$

where $\tilde{\mathbf{t}} = (\tilde{t}_{0,k}, \tilde{t}_{0,k-1}, \ldots)'$, $\mathbf{v} = (v_{0,k}, v_{0,k-1}, \ldots)$, $v_{s,i} = 1/(\zeta_{s,i} + \eta_{s,i} + \xi_{s,i})$, and $\mathbf{A}$ is the matrix with transition probabilities defined by the rates (3.5). A general solution to (3.7) is given by

$$\tilde{\mathbf{t}} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{v}, \tag{3.8}$$

which can be solved numerically, where $\mathbf{I}$ is the identity matrix of the same dimension as $\mathbf{A}$. As before, we are mainly interested in the case when all k subcommunities are initially endemic corresponding to $\tilde{t}_{0,k}$, corresponding to the expected time to extinction starting at the endemic level. This defines our first approximation for the expected time to extinction.

Approximation 1 *The expected time to extinction given that the process is started at the endemic level can be approximated by $\tau_{SIR}(\varepsilon) = \tilde{t}_{0,k}$, where $\tilde{\mathbf{t}}$ solves (3.7).*

In general an explicit expression for τ_{SIR} is not attainable, but for the case $k = 2$ we have the following explicit expression

$$\tau_{\mathrm{SIR}}(\varepsilon) = \tau(0) + \frac{\varepsilon \mu n R_0 (1 - 1/R_0)^2 \tau_{\mathrm{R}} \tau_n^3}{2(\varepsilon \mu n R_0 (1 - 1/R_0)^2 \tau_{\mathrm{R}} \tau_n + \tau_{\mathrm{R}} + \tau_n)}, \tag{3.9}$$

where $\tau(0)$ is from (2.8). From (3.9) one sees that τ_{SIR} is increasing in ε, and that if $\varepsilon = 0$ then $\tau_{\mathrm{SIR}} = \tau_{\mathrm{SI}}$. For larger values of k the calculations becomes more tedious, since the number of unknown equations increases rapidly.

We now return to the derivation of τ_{R}, the expected time which a subcommunity is recovered (and immune). When a subcommunity becomes recovered (and immune), there is approximately a fraction $\hat{x} = 1/R_0$ susceptible individuals and a fraction $1 - \hat{x}$ immune individuals. The problem now is that the probability that disease invasion will be able to persist depends on the fraction of susceptibles in the subcommunity.

If we look at the probability that disease invasion will take off, (3.2), we see that this probability is zero when we have a proportion of susceptibles corresponding to the endemic level, and we know that this probability is $1 - 1/R_0$ when a subcommunity is fully susceptible. Thus, we can define the expected time which a subcommunity is kept recovered (and immune) in terms of the average time it takes until a fraction $\tilde{x} > \hat{x}$ becomes susceptible in a subcommunity such that disease invasion will persist in the population with a pre-specified probability. A natural, but somewhat arbitrary, choice of this probability is $(1 - 1/R_0)/2$, i.e. half way between 0 and $1 - 1/R_0$.

This gives us that the fraction of susceptibles $\tilde{x}$ is the solution to

$$1 - \frac{1}{\tilde{x}R_0} = \frac{1}{2}\left(1 - \frac{1}{R_0}\right),$$

which is $\tilde{x} = 2/(R_0 + 1) \approx 2\hat{x}$ when R_0 is fairly large.

While a subcommunity is treated as recovered (and immune), no infectious contacts may occur, and the expected fraction of susceptibles $x(t)$ at a certain time point t after becoming disease-free is given by the solution to the differential equation

$$\begin{cases} \frac{dx}{dt} &= \mu(1 - x) \\ x(0) &= \hat{x} \end{cases} \tag{3.10}$$

Solving this equation gives us the relation

$$x(t) = 1 - (1 - \hat{x}) \exp(-\mu t). \tag{3.11}$$

If we set $x(t) = \tilde{x} = 2/(R_0 + 1)$ and solve (3.11) in terms of t, we get

$$t = \tau_{\mathrm{R}} = \frac{1}{\mu} \log\left(\frac{R_0 + 1}{R_0}\right). \tag{3.12}$$

Note that the longer we treat a subcommunity as recovered (and immune), the harder it gets for the infection to persist in the rest of the population. If τ_{R} is close to zero, we loose the effect of the recovered (and immune) state and the approximation resembles that of Hagenaars et al. (2004), and if τ_{R} tend to infinity it is the same as removing a subcommunity which becomes disease-free. Our suggestion of an approximation of τ_{R}, (3.12), will give relatively small values. However, as said before, it is hard to find a natural definition of this quantity.

4 Approximation using an exponential form

When we introduced the SIR model with demography for several communities, we derived the expected time to extinction both for the case when all communities are isolated and the case when they are mixing as one large homogeneous community, corresponding to $\varepsilon = 0$ and $\varepsilon = 1$ respectively. We have also mentioned that these two approximations cannot be improved along the present lines without improving Nåsell's approximations for the single community SIR model with demography, Eqs. (2.5) and (2.7).

For $0 < \varepsilon < 1$ we now introduce a new approximation, $\tau_{\mathrm{Exp}}(\varepsilon)$, by simply fitting an exponential curve having $\tau(0)$ as starting point and approximately $\tau(1)$ as end point such that $\tau'_{\mathrm{Exp}}(0) = \tau'_{\mathrm{SIR}}(0)$, i.e. we let the shape of τ_{Exp} be determined by the behaviour of τ_{SIR} in the ε region where we expect it to work satisfactory. We propose the following approximation:

Approximation 2 *The expected time to extinction given that the process is started in quasi-stationarity can be approximated by*

$$\tau_{Exp}(\varepsilon) = \tau(1) - (\tau(1) - \tau(0)) \exp\left(-\frac{\tau'_{SIR}(0)}{\tau(1) - \tau(0)}\varepsilon\right), \tag{4.1}$$

where $\tau'_{SIR}(\cdot)$ is the first derivative of (3.8) with respect to ε, and $\tau(0)$ and $\tau(1)$ are from Eqs. (2.8) and (2.9) respectively.

One can easily verify that $\tau_{\text{Exp}}(0) = \tau_{\text{SIR}}(0)$ and that $\tau'_{\text{Exp}}(0) = \tau'_{\text{SIR}}(0)$, and one also sees that when $\tau'_{\text{SIR}}(0) \gg \tau(1) - \tau(0)$ then $\tau_{\text{Exp}}(1) \approx \tau(1)$, as desired. To see that this is reasonable, look at the expoenent of (4.1), $-\tau'_{\text{SIR}}(0)\varepsilon/(\tau(1) - \tau(0))$, when $k = 2$ and use τ_{SIR} from (3.9). We then get that $\tau'_{\text{SIR}}(0) = \mu n(R_0 - 1)^2 \tau_\text{R} \tau_n^3/(2R_0(\tau_\text{R} + \tau_n))$ and a first order expansion of τ_R around 1 gives us that $\tau_\text{R} \approx 1/(\mu R_0)$ which togheter with Nåsell's geometric approximation of τ_n yields

$$-\frac{\tau'_{\text{SIR}}(0)}{\tau(1) - \tau(0)}\varepsilon \approx -n^3 \frac{(R_0 - 1)^4}{R_0^3 \mu \alpha^2 (\alpha^2 + n(R_0 - 1))}\varepsilon, \qquad (4.2)$$

which is a very small number for reasonable parameter values and choices of n. We illustrate this with a numerical example: Suppose that we have a population which is separated into two equally large subcommunities of size $n = 50,000$. Suppose further that the average infectious period is one week and a typical individual lives for ca. 70 years, i.e. $\alpha \approx 3500$. This togheter with $R_0 = 14$ and $\varepsilon = 1$ gives us that the exponent (4.2) is approximately -475, and $\exp(-475) \approx 0$, thus $\tau_{\text{SIR}}(1) \approx \tau(1)$.

5   Examples and simulations

We have made comparisons of our two approximations with simulations for some different parameter values and number of subcommunities. First, we recollect some vital assumptions that we have made. For both approximations we make use of Nåsell's approximations for the expected time to extinction for a single large community given that it initially has a number of infected individuals corresponding to the endemic level, i.e. the quasi-stationary level of infection, Eqs. (2.5) and (2.7). In order for the concept of quasi-stationarity to have any meaning, the quasi-stationary level of infection should at least correspond to that 10-20 individuals are infectious in each subcommunity, otherwise the disease will die out too quickly and the disease cannot be regarded as endemic.

Turning to our examples. For childhood diseases and diseases such as influenza the average infectious period is about one to two weeks, see pp. 81-86 in Anderson and May (1992). This togheter with the assumption that the average life length among individuals in the population is 70 years, gives us α values between 1,800 and 3,500. Usually, these kind of diseases have values of R_0 around 10 or higher. We have chosen to set R_0 to 14 in compliance with Nåsell (2005). These are the parameter values which we will use. As for the number of subcommunities we have chosen $k = 3$ and 5.

All simulations have been done using Monte Carlo simulation and the routines where written in the C-programing language and all graphics have been made using MATLAB. We estimate the expected time to extinction when starting in quasi-stationarity from the simulations as follows. Initially, 500 epidemics were started at the endemic level, which is the mean in the limiting quasi-stationary distribution. Then the epidemics were simulated

long enough for 100 of them to go extinct, and at this time point, the clock for the remaining 400 simulations was started. These starting points will be approximately from the quasi-stationary distribution since the epidemics have been started at the endemic level together with the fact that they had not gone extinct for some time.

Depending on parameter region we have alternated between using Nåsell's truncated normal and geometric approximation of the quasi-stationary distribution. For more on the apropriate choice of approximation in different parts of the parameter region see Nåsell (2005), Fig. 3 on p. 210. Roughly one can say that for the parameter values used here, the geometric approximation, Eq. (2.7), is to prefer if n (n_{tot}) is smaller than ca. 5-600,000 and when n (n_{tot}) is greater than ca. 2,000,000 one should use the truncated normal approximation, Eq. (2.5). For values of n (n_{tot}) in between, neither of the approximations work that well. From here onwards all references of type 'Fig. (A)' are refering to subfigure (A) in Fig. 2. In Fig. (A) we have a total population size of $n_{tot} = 150,000$ and $k = 3$, i.e. $n = 50,000$, and in this parameter region we have used Nåsell's geometric approximation of the quasi-stationary distribution. We see that both approximations work well for small values of ε, but that τ_{Exp} also gives a good fit for intermediate values of ε as well. But, notice that the expected time to extinction is too short to say that the disease is endemic, since the average time to extinction is ≈ 0.01 life lengths, i.e. less than a year. Worth noticing is also that the simulation of the expected time to extinction when starting at the endemic level indicates that it is increasing in the degree of social interaction between subcommunities, and that already for small values of ε the expected time to extinction is close to the case when the population is mixing homogeneously. This behaviour is also seen in both our approximations. In Fig. (B) we have used the same parameter values as in Fig. (A), but now with $k = 5$. Note that since we keep n_{tot} fixed the subcommunity size is smaller now. We again see that τ_{Exp} performs better than τ_{SIR}.

In Figs. (C) and (D), we have increased the total population size, n_{tot}, to 900,000. Now we are in a part of the parameter region where neither of Nåsell's approximations of the quasi-stationary distribution is working that well. We have choosen to use his geometric approximation, Eq. (2.7). Here we see in Figs. (C) and Fig. (D) that both our approximations gives a rather poor fit unless ε is very small, and note that both $\tau_{\text{Exp}}(0)$ and $\tau_{\text{Exp}}(1)$ are quite far away from the corresponding simulated values. If we instead use the simulated values for the cases $\varepsilon = 0$ and $\varepsilon = 1$ to approximate $\tau(0)$, $\tau(1)$ and τ_n, things look different. We have done this in Fig. (E) when the parameter values are as before and when $n_{tot} = 900,000$ and $k = 3$. Again we see that both τ_{SIR} and τ_{Exp} work well for small values of ε, and that τ_{Exp} also gives a good fit for somewhat higher values of ε. This indicates that the functional form of τ_{Exp} works rather well, but that both τ_{SIR} and τ_{Exp} are sensitive to the intital approximations of τ_n, $\tau(0)$ and $\tau(1)$.

In Fig. (F) we have five subcommunities and a total population size of 2,500,000, and in this situation we should use Nåsell's truncated normal approximation of the quasi-stationary distribution, Eq. (2.5), for $\tau(1)$, and his geometric approximation, Eq. (2.7), for τ_n and $\tau(0)$. Once again we see the same behaviour as before: both τ_{SIR} and τ_{Exp} gives a rather good fit for small values of ε, but τ_{Exp} also gives a good fit for high values of ε, but not particularly good fit for intermediate values. We also see that the functional form of τ_{Exp} seems to work rather well.

To conclude, of our two proposed approximations τ_{Exp} seems to be the best. But, re-

member that we still need τ_{SIR}, since we make use of $\tau'_{\mathrm{SIR}}(0)$ when defining τ_{Exp}, see Section 4. Further, we see that the functional form of τ_{Exp} seems to work rather well, but both our approximations are sensitive to the initial approximations of τ_n, $\tau(0)$ and $\tau(1)$, obtained from the single community model. These are also, to our knowledge, the only approximations of the expected time to extinction when α is large.

6   Discussion

In the present paper we have been concerned with approximations of τ, the expected time to extinction for an SIR model with demography for the case with subcommunities, when each subcommunity is started at the endemic level of infection. Our aim has been to understand the effect of population subdivision on the time to extinction. We have mainly focused on endemic diseases which are highly infectious and with a short average infectious period in relation to average life length, such as childhood diseases.

Our first approximation, τ_{SIR} (Approximation 1), is based on a method presented in a paper by Hagenaars et al. (2004). In Hagenaars et al. (2004) they are mainly interested in diseases with rather small α, such as scrapie among sheep, in situations where the social activity between subcommunities is low, i.e. $0 < \varepsilon \ll 1$. Under these circumstances they argue that the underlying SIR model can be analysed from a subcommunity view, where each subcommunity is classified as either fully susceptible or endemic, and they approximate the dynamics in the population with a birth and death process for the number of endemic subcommunities, which only allow for one reinfection. The expected time to extinction for this process is then a reasonable approximation of τ, (3.4). We are, as said before, mainly interested in childhood diseases. Based on similar arguments as those made in Hagenaars et al. (2004), we argue that we in this situation also need to introduce a recovered (and immune) state when classifying subcommunities in order to avoid over-estimation of τ. Using these states we approximate the underlying SIR model with a Markov jump process for the number of endemic subcommunities, see Fig. 1, and estimate τ with the expected time to extinction for this process. We present a general solution form for an arbitrary number of k subcommunities in Eq. (3.8), which can be solved numerically, and we present an explicit expression for the case when $k = 2$ in (3.9).

Simulations indicate that the expected time to extinction is increasing in the degree of social interaction between subcommunities, which also can be seen in τ_{SIR}. Further, we see that τ_{SIR} is more suitable to use when the degree of social activity between subcommunities is very low. One crucial part with this approximation is that it is difficult to find a natural way of defining the time which we let a subcommunity stay recovered (and immune). If one could improve this part of the approximation, it is possible that τ_{SIR} could work better when the degree of social activity between subcommunities is somewhat higher as well. One other thing worth noticing is that τ_{SIR} is rather sensitive to the initial approximations of τ_n, $\tau(0)$ and $\tau(1)$. Depending on the part of the parameter region we are in we must be careful with which of Nåsell's approximations of the quasi-stationary distribution we use, i.e. the truncated normal or geometric distribution, Eqs. (2.5) and (2.7) respectively.

Our second approximation, τ_{Exp} (4.1), is motivated in a slightly different way. At the end
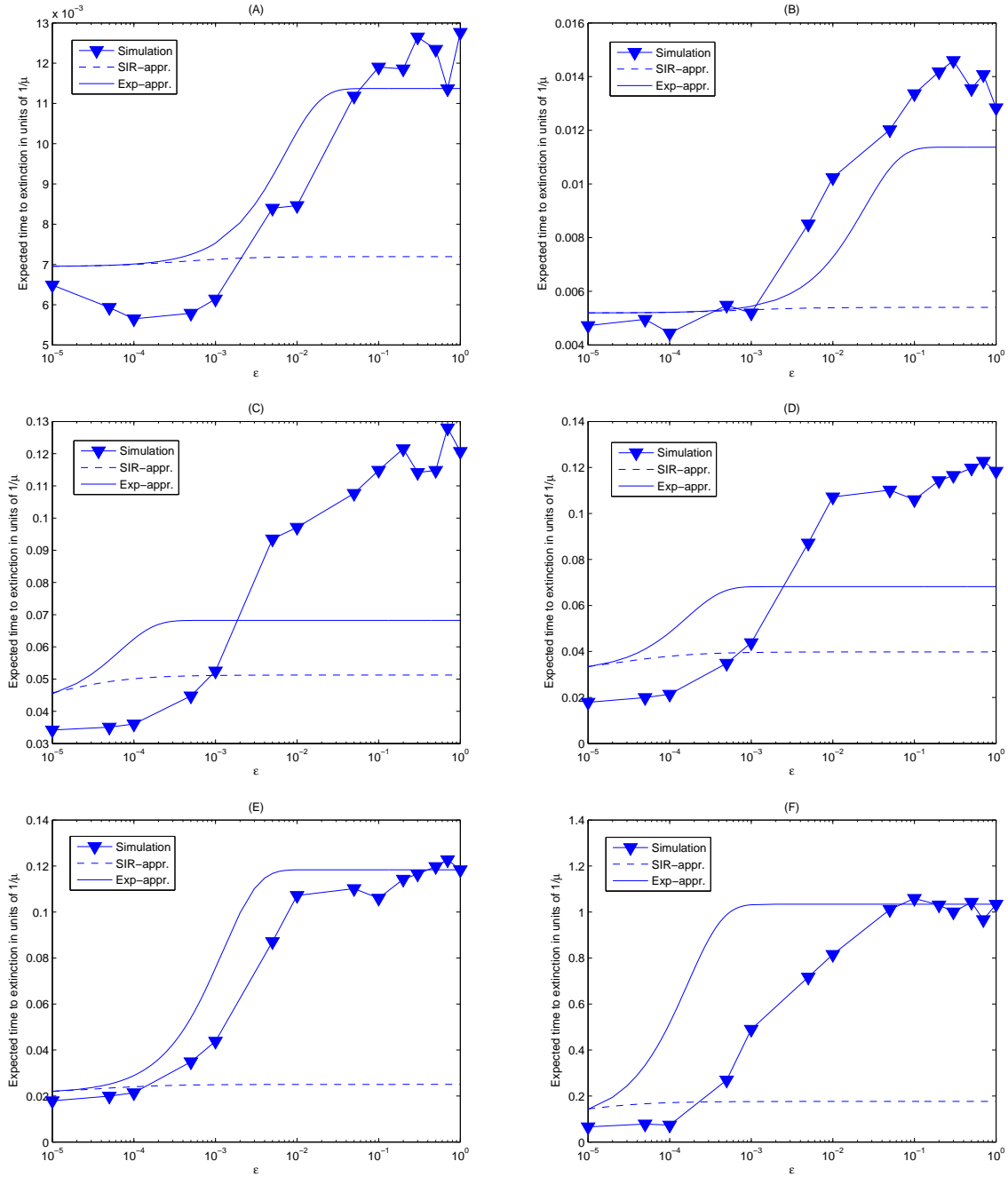
Figure 2: In all (A)-(F) we have used $\alpha = 3,500$ and $R_0 = 14$, and time is measured in units of life lengths. In (A)-(B) we have a total population size of $n_{tot} = 150,000$, but in (A) there are $k = 3$ subcommunities where as in (B) $k = 5$. In (C) and (D) $n_{tot} = 900,000$ and $k = 3$ and $k = 5$ respectively. Further, (E) is the same as (D), but here $\tau(0)$ and $\tau(1)$ are approximated with the values corresponding to $\varepsilon = 0$ and $\varepsilon = 1$ from the simulation, and τ_n can be obtained from $\tau(0)$. In (F) $n_{tot} = 2,500,000$ and $k = 5$.

17

of section 3 we talked about the expected time to extinction for the two extreme cases of social activity, when all subcommunities are totally isolated from each other, $\varepsilon = 0$, and when the social contact between subcommunities is so high that all subcommunities mix as one large homogeneous community, $\varepsilon = 1$. These two approximations cannot be improved without improving results for the single community SIR model with demography. As we just discussed when talking about τ_{SIR}, we said that it is reasonable to use when the social activity between subcommunities is very low. With these two things in mind, we approximate the expected time to extinction with an exponential curve with start point corresponding to $\tau(0)$, Eq. (2.8), and has approximately $\tau(1)$, Eq. (2.9), as end point in such a way that $\tau'_{\mathrm{Exp}}(0) = \tau'_{\mathrm{SIR}}(0)$.

When comparing τ_{Exp} with simulations we see, as for τ_{SIR}, that τ_{Exp} works satisfactory when the degree of social activity is low, and that it is increasing in the degree of social activity. But, for τ_{Exp} we also get a better fit for higher degrees of social activity as well. We have also seen that the exponential form of τ_{Exp} gives a rather good description of the expected time to extinction. Further, since τ_{Exp} is defined using $\tau'_{\mathrm{SIR}}(0)$, it is also sensitive to the initial approximations of τ_n, $\tau(0)$ and $\tau(1)$.

To conclude, both approximations work rather well for small values of ε when we are in the parts of the parameter region where one of Nåsell's approximations of the quasi-stationary distribution is good. But, in these situations τ_{Exp} also gives a rather good fit for intermediate values of ε as well. Thus, τ_{Exp} is the best approximation of the two. These approximations are, to our knowledge, the only ones at hand which deal with the expected time to extinction when α is large.

We have also shown that the time to extinction given that the epidemic process is started in the quasi-stationary distribution T_Q, is exponentially distributed, see Proposition 2.1. This result is important when talking about other quantities of interest such as critical community size.

7 Acknowledgements

References

Anderson R.M., May R.M., 1992. Infectious Diseases of Humans - Dynamics and Control. Oxford University Press, pp. 13-19, 81-86.

Andersson H., Britton T., 2000a. Stochastic Epidemic Models and Their Statistical Analysis. Lecture Notes in Statistics, 151, Springer-Verlag, New York, pp. 22-25.

Andersson H., Britton T., 2000b. Stochastic epidemics in dynamic populations: quasi-

stationarity and extinction. J. Math. Biol. 41, 559-580.

Diekmann O., Heesterbeek J.A.P., 2000. Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation. Chichester : Wiley, cop.

Hagenaars T.J., Donnelly C.A., Ferguson N.M., 2004. Spatial heterogenity and the persistence of infectious diseases. J. Theo. Biol., 229, 349-359.

van Herwaarden O.A., Grasman J., 1995. Stochastic epidemics: major outbreaks and the duration of the endemic period. J. Math. Biol., 33, 581-601.

Nåsell I., 1999. On the time to extinction in recurrent epidemics. J. R. Statist. B, 61, 309-330.

Nåsell I. 2005. A new look at the critical community size for childhood infections. Theo. Pop. Biol., 67, 203-216.

Pollett P.K., Roberts A.J., 1990. A Description of the long-term behaviour of absorbing continuous-time markov chains using a centre manifold. Adv. Appl. Prob., 22, 111-128.