Mathematical Statistics
Stockholm University

The time to extinction for an SIS-household-epidemic model

Tom Britton
Peter Neal

Research Report 2009:7

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

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

The time to extinction for an SIS-household-epidemic model

Tom Britton
Peter Neal *

June 2009

Abstract

We analyse a stochastic SIS epidemic amongst a finite population partitioned into households. Since the population is finite, the epidemic will eventually go extinct, i.e. have no more infectives in the population. We study the effects of population size and within household transmission upon the time to extinction. This is done through two approximations. The first approximation is suitable for all levels of within household transmission and is based upon an Ornstein-Uhlenbeck process approximation for the diseases fluctuations about an endemic level relying on a large population. The second approximation is suitable for high levels of within household transmission and approximates the number of infectious households by a simple homogeneously mixing SIS model with the households replaced by individuals. The analysis, supported by a simulation study, shows that the mean time to extinction is minimized by moderate levels of within household transmission.

Key words: SIS epidemics · households model · time to extinction · Ornstein-Uhlenbeck process

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

The time to extinction for an SIS-household-epidemic model

Tom Britton · Peter Neal

Abstract We analyse a stochastic SIS epidemic amongst a finite population partitioned into households. Since the population is finite, the epidemic will eventually go extinct, *i.e.* have no more infectives in the population. We study the effects of population size and within household transmission upon the time to extinction. This is done through two approximations. The first approximation is suitable for all levels of within household transmission and is based upon an Ornstein-Uhlenbeck process approximation for the diseases fluctuations about an endemic level relying on a large population. The second approximation is suitable for high levels of within household transmission and approximates the number of infectious households by a simple homogeneously mixing SIS model with the households replaced by individuals. The analysis, supported by a simulation study, shows that the mean time to extinction is minimized by moderate levels of within household transmission.

Keywords SIS epidemics · households model · time to extinction · Ornstein-Uhlenbeck process

1 Introduction

Epidemic models are widely used for increasing the understanding of infectious disease dynamics (*e.g.* Anderson and May [1]) and for determining suitable preventive measures to reduce or ideally stop disease spread (*e.g.* Wallinga et al. [19], Ferguson et al. [9] and Cauchemez et al. [6]). In the present work we study a class of models for diseases that are endemic. More precisely we look at SIS-epidemic models (Kryscio

T. Britton

Department of Mathematics, Stockholm University, SE-106 91 Stockholm, Sweden.
E-mail: tom.britton@math.su.se

P.Neal

School of Mathematics, Alan Turing Building, University of Manchester, Manchester, M13 9PL, UK.
Tel.: +44-161-3063634
Fax: +44-161-3063669
E-mail: peter.neal@manchester.ac.uk

and Lefèvre [13]) meaning that individuals are either susceptible (S) or infectious (I), and susceptible individuals might get infected and infectious, and after a time in the infectious state individuals recover and return to the susceptible state. The community is considered closed being of size N, but an alternative interpretation of the model is that an individual that stops being infectious becomes immune for the rest of its life (or dies) but is "replaced" by a new susceptible individual when the infectious period stops.

SIS-epidemic models can potentially lead to the disease establishing itself within the population and becoming endemic for a long period of time. However, eventually the disease goes extinct. It is therefore important to study properties of the model that determine whether the disease may become endemic, and if so what the endemic level is and for how long (on average) it will take for the disease to go extinct. In terms of prevention the answers to these questions can give guidance in determining how effective a given preventive measure is, for example: what proportion is necessary to vaccinate in order to make the disease go extinct promptly?

In the present paper we address the question of the time until an endemic disease goes extinct, where we assume the community size N is large and that we start in the endemic level. More precisely we study the situation where individuals are grouped into households and where it is assumed that transmission is much higher between individuals in the same household as compared with individuals in different households. We derive an approximation for the expected time to extinction and compare it with the case that there are no households (homogeneous mixing) for which asymptotic results for the time to extinction are already available (Andersson and Djehiche [4]). This paper is hence concerned with analysing how the time to extinction is affected when the assumption of homogeneous mixing of identical individuals is relaxed by admitting the presence of households. Similar questions have been analysed when relaxing the assumption of homogeneous mixing but in other "directions": Hagenars et al. [11] look at the case with an SIR epidemic where the community consists of a few large sub-communities assuming higher contact rates within the sub-communities, Lindholm [14] studies an epidemic model where he investigates how the time to extinction is affected by individuals being of different types having different susceptibility and infectivity to the disease.

Our main conclusion is that, assuming a fixed endemic level of infectives $\hat{i}$, the time to extinction is smaller if there is moderate or high transmission within households as compared to the case of homogeneous mixing in the limit as $N \to \infty$. For high values of $\hat{i}$ our Ornstein-Uhlenbeck approximation suggests a monotone decrease of the expected time to extinction as a function of the degree of within-household transmission rate λ_L, whereas for small $\hat{i}$-values the expected time to extinction first increases and then decreases as a function of λ_L. In the latter case the longest time to extinction is hence for the case with small (but positive!) λ_L. The Ornstein-Uhlenbeck approximation is an asymptotic result as $N \to \infty$ using a normal approximation for the probability of being close to extinction. We introduce a second approximation which is appropriate when the within-household transmission rate, λ_L, tends to infinity. This approximation shows that for fixed population size N, the mean time to extinction grows polynomially with λ_L as $\lambda_L \to \infty$ with the rate of the polynomial growth depending upon the maximum household size. Thus there are two competing approximations leading in practice to the mean time to extinction being smallest for moderate levels of within household infection. We focus the analytical analysis on the case with households of size 2 with a

simulation study showing that qualitatively similar results hold for households having other and possible different household sizes.

In Section 2 we define the SIS-household epidemic model and present some known results for the model (Ball [5], Neal [15]). In Section 3 we derive our two approximations for the expected time to extinction and give a full discussion of the merits of the two approximations. In Section 4 we give some numerical examples and plot the expected time to disease extinction as a function of the degree of transmission within households, and compare these approximations with simulation results. The paper ends with a short discussion about the limitations of the present analysis, and a discussion about interesting related models/questions.

2 The SIS household epidemic model

2.1 Model definition

We now define the Markovian household SIS epidemic model. This a special case of the more general household SIS epidemic model defined and analysed by Ball [5].

Consider a fixed community consisting of n households, all being of the same size k for simplicity of exposition (later we will primarily focus on the case $k = 2$). The community size is hence $N = kn$, and we assume that n and hence N is large. Each individual is at any time either infectious or susceptible. Susceptible individuals recover (and become susceptible again) independently at the constant rate γ. While infectious an individual makes two types of contacts: the individual has *global* "close contacts" at rate λ_G, each time the contacted person is selected uniformly at random from the whole community, and the individual has *local* "close contacts" at rate λ_L with individuals belonging to the same household, here too the individual to be contacted is selected uniformly at random, but now among the individuals belonging to the same household as the person in question. By "close contact" we mean a contact that results in infection if the other person is susceptible – if the other person is infectious the contact is assumed to have no effect. The epidemic is initiated by some given initial condition specifying how many households have j infective individuals, $j = 0, 1, \ldots, k$ infective individuals.

Let $Y_j(t)$ denote the number of households having j infectives (and hence $k - j$ susceptibles) at t. The initial condition is hence specified by the vector $(Y_0(0), \ldots, Y_k(0))$, for any t: $\sum_j Y_j(t) = n$ since the total number of households is n. Let $I(t) = \sum_j j Y_j(t)$ denote the number of infectives (=infectious individual) at t, and similarly $S(t) = \sum_j (k-j) Y_j(t)$ is the number of susceptible individuals.

The process $(Y_0(t), \ldots, Y_k(t))$ is Markovian and there is at most one infection or recovery occurring at a given time point, implying the possible jumps are that one component decreases by 1 and at the same time the component directly to the left (recovery) or directly to the right (infection) increases by 1. From the model, each individual recovers at rate γ, so the rate at which some household having i infectives at present has a recovery (implying that it changes to state $i - 1$) equals $\gamma \cdot i Y_i(t)$. An infection can occur in two different ways. Either an individual is infected globally, or else locally from within the household. The rate of being infected globally is the same for all susceptible individuals: $\lambda_G I(t)/N$, each infective has close contact with a given individual at rate λ_G/N and there are $I(t)$ infectives (in principle the division should be with $N - 1$ since it is not possible to have contact with yourself, but when N is

large this is negligible). Additionally, a susceptible in a household having i infectives at present gets infected locally at rate $i\lambda_L/(k-1)$ since each infective has close contact with a specific household member at rate $\lambda_L/(k-1)$. The overall rate at which some household having i infectives moves to the state $i+1$ is hence

$$(\lambda_G I(t)/N + i\lambda_L/(k-1))(k-i).$$

2.2 Properties of the SIS household epidemic model

From now on we restrict ourselves to the case where all households are of size $k = 2$. The reason for this is to simplify notation and keeping the dimensions of vectors and matrices low – there is nothing harder in principle for larger household sizes. The SIS household epidemic model have been analysed previously be Ball [5], Ghoshal *et al.* [10] and Neal [15], in particular studying how the model behaves when n is large. To this end we define $\bar{Y}_i(t) = Y_i(t)/n$, the proportion of households having i infectives and let $\bar{I}(t) = I(t)/N$, the proportion of the population infectious. In Ball [5] it is shown that, as $n \to \infty$, the process $(\bar{Y}_0(t), \bar{Y}_1(t), \bar{Y}_2(t))$ converges in probability (uniformly on finite intervals) to the deterministic vector process $(y_0(t), y_1(t), y_2(t))$ defined as the solution to the differential equations

$$\begin{aligned}
y_0'(t) &= \gamma y_1(t) - \lambda_G i(t) 2 y_0(t) \\
y_1'(t) &= \lambda_G i(t) 2 y_0(t) + 2\gamma y_2(t) - \gamma y_1(t) - (\lambda_G i(t) + \lambda_L) y_1(t) \\
y_2'(t) &= -2\gamma y_2(t) + (\lambda_G i(t) + \lambda_L) y_1(t),
\end{aligned}$$

where $i(t) = \sum_i i y_i(t)/2$ (the fraction of infectives). The result relies on the initial conditions agreeing and having a positive fraction of infectives: $(y_0(0), y_1(0), y_2(0)) = (y_0, y_1, y_2)$ as well as $(\bar{Y}_0(0), \bar{Y}_1(0), \bar{Y}_2(0)) \to (y_0, y_1, y_2)$ as $n \to \infty$, such that $\sum_j y_j = 1$ (the proportions sum up to 1) and $\sum_j j y_j/2 = i(0) > 0$ (there is a positive initial fraction of infectives).

Ball [5] also derives a threshold parameter $R_* = (\lambda_G/\gamma)(1 + \lambda_L/\gamma)$ such that the deterministic epidemic will go extinct (i.e. $y_0(t) \to 1$ as $t \to \infty$) if $R_* \leq 1$, whereas it will reach an endemic level if $R_* > 1$. In the latter case the endemic level $(\hat{y}_0, \hat{y}_1, \hat{y}_2)$ is given by Ball [5], Theorem 3.1. This gives that the endemic level of infectives $\hat{i} = \sum_j j\hat{y}_j/2$ equals

$$\hat{i} = \frac{2\lambda_L \gamma}{\lambda_G(\sqrt{(\lambda_L + \lambda_G)^2 + 4\lambda_L\gamma} - (\lambda_L + \lambda_G))} - \frac{\gamma + \lambda_L}{\lambda_G}. \tag{1}$$

3 The time to extinction for the SIS household epidemic

In the present section we derive a rough approximation for the time to extinction of the SIS household epidemic model defined in Section 2.1 applicable for all values of λ_L and relying on n being large, and a more precise approximation suitable for large values of λ_L. Before doing this we present the results of Andersson and Djehiche [4] for the time to extinction of the homogeneous SIS epidemic model, i.e. the model above but without having household structure corresponding to the case where $\lambda_L = 0$. Without loss of generality we shall fix $\gamma = 1$ which simply means that we measure time with the average infectious period as the base time unit.

3.1 The time to extinction for the homogeneous SIS model

When there are no households ("the homogeneous case") the process is described by $(S(t), I(t))$, the number of susceptible and the number of infectives, but since $S(t) + I(t) = N$ is fixed it is enough to keep track of $I(t)$ say. When N is large and assuming $R_* = \lambda_G > 1$, $I(t)$ will fluctuate around the endemic level $N\hat{i} = N(1 - 1/\lambda_G)$ but eventually make a big "excursion" down to $I(t) = 0$ when the epidemic goes extinct. Andersson and Djehiche [4] prove that, as N tends to infinity, the time $\tau_N^{(AD)}$ until this happens is exponentially distributed with mean

$$E(\tau_N^{(AD)}) \sim \sqrt{\frac{2\pi}{N} \frac{\lambda_G}{(\lambda_G - 1)^2}} e^{N(\log \lambda_G - 1 + 1/\lambda_G)}. \tag{2}$$

The idea behind the proof is that the process will make many excursions from the endemic level eventually returning to the endemic level. Finally it will make a big excursion down to the absorbing state $I(t) = 0$ when the epidemic goes extinct. It will hence make a geometric number of returns to the endemic level before going extinct, and in the proof they derive the expected duration of an excursion which returns to the endemic level and the (small) probability of making a big excursion to extinction.

3.2 Approximation of the SIS household epidemic

As mentioned in Section 2.2 it was shown by Ball [5] that when N is large the SIS household epidemic may be approximated by a deterministic process. Using theory for Markov population processes (Ethier and Kurtz [8]) it is also possible to show a weak convergence result. Our process has three components, but because the total number of households $\sum_{i=0}^{2} Y_i(t) = n(= N/2)$ is fixed it is really two dimensional. This means it is enough to keep track of two components, e.g. Y_0 and Y_1. Since our prime interest is the total number of infectives, it is natural to follow House and Keeling [12] and parameterise the model in terms of $I(t)$ and $Y_1(t)$. Suppose the process is started from the endemic level, i.e. $(I(0), Y_1(0)) = (N\hat{i}, N\hat{y}_1/2)$, and define the standardized process by

$$\begin{pmatrix} \tilde{I}^{(N)}(t) \\ \tilde{Y}_1^{(N)}(t) \end{pmatrix} := \sqrt{N} \begin{pmatrix} \bar{I}(t) - \hat{i} \\ \bar{Y}_1(t)/2 - \hat{y}_1/2 \end{pmatrix}. \tag{3}$$

Then it follows from Ethier and Kurtz [8] that the vector process $\tilde{Z}^{(N)}(t)$ with components $\tilde{I}^{(N)}(t)$ and $\tilde{Y}_1^{(N)}(t)$ converges weakly to $\tilde{Z}$, where $\tilde{Z}$ is an Ornstein-Uhlenbeck process. The properties of the Ornstein-Uhlenbeck process stem from the system of differential equations (1), where we now have assumed that $\gamma = 1$. These differential equations can in vector form be written as $z'(t) = F(z(t))$, where $z(t) = (i(t), y_1(t)/2)$. Using this notation, $\tilde{Z}$ is defined by the drift matrix

$$B = \partial F = \begin{pmatrix} \lambda_G(1 - 2i(t)) - 1 & \lambda_L \\ \lambda_G(1 - 2i(t)) + 1 - \lambda_G y_1(t)/2 & -1 - \lambda_L, \end{pmatrix}, \tag{4}$$

and local covariance matrix

$$S = \begin{pmatrix} \beta_1(t) + \beta_2(t) + \beta_3(t) + \beta_4(t) & \beta_1(t) - \beta_2(t) - \beta_3(t) + \beta_4(t) \\ \beta_1(t) - \beta_2(t) - \beta_3(t) + \beta_4(t) & \beta_1(t) + \beta_2(t) + \beta_3(t) + \beta_4(t). \end{pmatrix}, \tag{5}$$

where $\beta_1(t) = \lambda_G i(t)(1 - i(t) - y_1(t)/2)$, $\beta_2(t) = (\lambda_G i(t)y_1(t)/2 + \lambda_L y_1(t))$, $\beta_3(t) = 2(i(t) - y_1(t)/2)$ and $\beta_4(t) = y_1(t)$ are the infinitesimal transition rates of an infection within an household with no infectives, an infection within an household with one infective, a recovery within a household with two infectives and a recovery within a household with one infective, respectively. Note that, if $(i(0), y_1(0)) = (\hat{i}, \hat{y}_1)$, then for all t, $(i(t), y_1(t)) = (\hat{i}, \hat{y}_1)$.

This Ornstein-Uhlenbeck has a Gaussian stationary distribution with mean-zero and covariance matrix Σ defined by

$$B\Sigma + \Sigma B^T = -S. \tag{6}$$

This means that, for large t, our original process $Z(t)$ will, conditional upon not having gone extinct, be approximately normal with mean vector $(N\hat{i}, N\hat{y}_1/2)$ and covariance matrix $N\Sigma$.

Solving Σ is straightforward. However, the expression for Σ in terms of λ_G and λ_L is not insightful as can be seen by the expression for Σ_{11}, the variance of the total proportion of infectives, given by,

$$\Sigma_{11} = \frac{(\varsigma_2 - \varsigma_1)}{((\lambda_L + \lambda_G)^2 + 4\lambda_L)\varsigma_3 - \varsigma_4}\hat{i}, \tag{7}$$

where

$$
\begin{aligned}
\varsigma_1 =\ & 6\lambda_L^5 + 22\lambda_G\lambda_L^4 + 36\lambda_L^4 + 52\lambda_L^3 + 30\lambda_G^2\lambda_L^3 + 76\lambda_G\lambda_L^3 + 18\lambda_G^3\lambda_L^2 + 52\lambda_L^2\lambda_G^2 + 8\lambda_L^2 \\
& + 42\lambda_L^2\lambda_G + 10\lambda_G^2\lambda_L + 4\lambda_G^4\lambda_L + 14\lambda_G^3\lambda_L + 2\lambda_G^4 \\
\varsigma_2 =\ & 2\sqrt{(\lambda_L + \lambda_G)^2 + 4\lambda_L}(3\lambda_L^4 + 8\lambda_G\lambda_L^3 + 12\lambda_L^3 + 7\lambda_L^2\lambda_G^2 + 16\lambda_L^2\lambda_G + 8\lambda_L^2 + 2\lambda_G^3\lambda_L \\
& + 3\lambda_L\lambda_G + 6\lambda_G^2\lambda_L + \lambda_G^3)) \\
\varsigma_3 =\ & (5\lambda_L^4 + 12\lambda_G\lambda_L^3 + 23\lambda_L^3 + 9\lambda_L^2\lambda_G^2 + 23\lambda_L^2 + 31\lambda_L^2\lambda_G + 13\lambda_G^2\lambda_L + 2\lambda_L + 2\lambda_G^3\lambda_L \\
& + 14\lambda_L\lambda_G + 2\lambda_G^3 + 2\lambda_G^2) \\
\varsigma_4 =\ & \sqrt{(\lambda_L + \lambda_G)^2 + 4\lambda_L}(15\lambda_G^3\lambda_L + 21\lambda_G^2\lambda_L^3 + 48\lambda_L^2\lambda_G^2 + 59\lambda_L^3 + 33\lambda_L^4 \\
& + 5\lambda_L^5 + 22\lambda_L^2 + 2\lambda_G^3 + 2\lambda_G^4 + 68\lambda_G\lambda_L^3 + 17\lambda_G\lambda_L^4 + 6\lambda_L\lambda_G \\
& + 20\lambda_G^2\lambda_L + 59\lambda_L^2\lambda_G + 11\lambda_G^3\lambda_L^2 + 2\lambda_G^4\lambda_L)).
\end{aligned}
$$

Similar expressions exist for the other components of Σ. For $k > 2$, the corresponding expression to (7) is even more unwieldily. However, studying the behaviour of (7) in the limits as $\lambda_L \downarrow 0$ and $\lambda_L \to \infty$ is informative.

Let $\sigma_i^2(\hat{i}, \lambda_L) = \Sigma_{1,1}$, the variance for the total proportion of infectives in equilibrium, explicitly stating the dependence upon $\hat{i}$ and λ_L. (Note that λ_G can be expressed as a function of $\hat{i}$ and λ_L.) Then for fixed $\hat{i}$, $\sigma_i^2(\hat{i}, 0) = \hat{i}^{-1}$ and $\lim_{\lambda_L \to \infty} \sigma_i^2(\hat{i}, \lambda_L) = 2\hat{i}^{-1}$. That is, for large λ_L the variance of the proportion of infectives is approximately twice the corresponding variance in the homogeneous case. There is a simple explanation for this. In the homogeneous case, $\lambda_L = 0$, we have $N = 2n$ individuals. On the other hand, when $\lambda_L \to \infty$, the two members of a household are effectively paired together, either both susceptible or both infectious. Thus the population in effect consists of n paired individuals resulting in the variance doubling.

This leads onto the question of, whether or not, for fixed $\hat{i}$, $\sigma_i^2(\hat{i}, \lambda_L)$ is monotonically increasing in λ_L. Plots of $\sigma_i^2(\hat{i}, \lambda_L)$, fixing $\hat{i}$ and varying λ_L, suggest that this is the case if $\hat{i} \geq 0.5$, whilst, $\sigma_i^2(\hat{i}, \lambda_L)$, is initially decreasing if $\hat{i} < 0.5$. This is partially

confirmed by studying $\frac{\partial}{\partial \lambda_L} \sigma_i^2(\hat{i}, \lambda_L)\big|_{\lambda_L = 0}$ which is positive, equal to 0 and negative when $\hat{i} > 0.5$, $\hat{i} = 0.5$ and $\hat{i} < 0.5$, respectively. Note that for general choices of k, $\sigma_i^2(\hat{i}, 0) = \hat{i}^{-1}$ and $\lim_{\lambda_L \to \infty} \sigma_i^2(\hat{i}, \lambda_L) = k\hat{i}^{-1}$.

If instead we consider fixed $R_* = \lambda_G(1 + \lambda_L)$, Σ_{11} is increasing as λ_L increase, for all $R_* > 1$. Also $\hat{i}$ is maximised at the extremes $\lambda_L = 0$ and $\lambda_L \to \infty$, where $\hat{i} = 1 - 1/R_*$.

The above analysis gives a good description of the endemic level, we now turn to the question of time to extinction and look to see how the above Ornstein-Uhlenbeck approximation can be used to assist in estimating this quantity.

3.3 The time to extinction for the SIS household epidemic

It seems hard to derive an explicit result corresponding to that of Andersson and Djehiche [4] for the household epidemic, the reason being that the process need not return to the endemic level $(N\hat{i}, N\hat{y}_1/2)$, and also because the trajectory down to extinction is not unique. It seems possible to derive a large deviation result but not to obtain a useful explicit expression for the time to extinction. Instead we have taken the approach first used by Nåsell [17] who applies it to the homogeneous SIR epidemic with demography, also used in Andersson and Britton [3]. We now present this approximation.

Let Q denote the quasi-stationary distribution of the SIS household epidemic, i.e. Q is the stationary distribution of the process $Y(t) = (Y_0(t), Y_1(t), Y_2(t))$ conditioned on not having gone extinct. Hence,

$$(Y(t)|Y(0) \sim Q, I(t) > 0) \sim Q.$$

Starting in the quasi-stationary distribution it follows, because of the memoryless property, that the time to extinction $T_Q = \inf\{t; I(t) = 0|Y(0) \sim Q\}$ is exponentially distributed with intensity equal to the probability of being one step away from extinction multiplied by the rate of moving into absorption. If we denote the quasi-stationary distribution $Q = \{q_{i,y_1}\} = P_Q(I(t) = i, Y_1(t) = y_1)$ we hence have that

$$T_Q \sim \text{Exp}(\gamma q_{1,1}) = \text{Exp}(q_{1,1}), \tag{8}$$

since we have taken $\gamma = 1$. Note that this is an exact result for any N. However, it remains to derive $q_{1,1}$. Since when $I(t) = 1$ we have by necessity that $Y_1(t) = 1$ (and $Y_0(t) = n - 1$), it is enough to look at the marginal distribution $\{q_i\}$ of the number of infectives. An approximation for q_1 is given by the normal approximation of the stationary Gaussian distribution derived in Section 3.2. This approximation consists of computing the normal density for $I(t)$ (which is approximately normal with mean $N\hat{i}$ and variance $N\sigma_i^2(\hat{i}, \lambda_L)$) at the point 1 and conditioning on that $I(t) > 0$. This approximation is of course better in central parts of the distribution (around $I(t) \approx N\hat{i}$) but in the absence of a better approximation we use it also in the tail $I(t) = 1$.

Using the expression for the endemic level $(\hat{i}, \hat{y}_1)$ and the variance $\sigma_i^2(\hat{i}, \lambda_L)$ we get the following approximation for q_1

$$
\begin{aligned}
q_1 &\approx P(I(t) = 1 | I(t) > 0, t \text{ large}) \\
&\approx \frac{1}{\sqrt{2\pi N}\sigma_i(\hat{i}, \lambda_L)} \exp\left(-\frac{1}{2} \times \frac{(1 - N\hat{i})^2}{N\sigma_i^2(\hat{i}, \lambda_L)}\right) \\
&\approx \frac{1}{\sqrt{2\pi N}\sigma_i(\hat{i}, \lambda_L)} \exp\left(-\frac{N\hat{i}^2}{2\sigma_i^2(\hat{i}, \lambda_L)}\right),
\end{aligned}
\tag{9}
$$

giving our first approximation for the time to extinction of the SIS household epidemic

$$E(T_Q) = \frac{1}{q_1} \approx \sqrt{2\pi N} \sigma_i(\hat{i}, \lambda_L) \exp\left(\frac{N\hat{i}^2}{2\sigma_i^2(\hat{i}, \lambda_L)} \right). \qquad (10)$$

The dominant term in our first approximation (10) is $\exp\left(\frac{N}{2} \times \frac{\hat{i}^2}{\sigma_i^2(\hat{i}, \lambda_L)} \right)$. Thus for large N and a fixed value of $\hat{i}$, the time to extinction is determined by $\sigma_i^2(\hat{i}, \lambda_L)$ with the smaller the value of $\sigma_i^2(\hat{i}, \lambda_L)$, the longer the time to extinction is expected to be. From the study of $\sigma_i^2(\hat{i}, \lambda_L)$ our first approximation hence suggests that $E(T_Q)$ is monotonically decreasing in λ_L if $\hat{i}$ is fixed and larger than 0.5. If on the other hand we fix $\hat{i} < 0.5$, $E(T_Q)$ first increases and then decreases with λ_L, and $E(T_Q)$ hence has a (local) maximum for a small but positive λ_L. It was noted in Doering $et\ al.$ [7], that for the homogeneously mixing case the Ornstein-Uhlenbeck approximation is only reasonable for $R_* = 1 + C/N^{\frac{1}{3}}$, for $C > 0$ and this is also likely to be the case for the household model. However, (10) still proves to be useful in gaining an understanding of how λ_L affects the extinction time.

We now consider the case where λ_L is large and derive a second more precise approximation of $E(T_Q)$. As noted in Section 3.2, for large values of λ_L, it is unlikely that there will be just one infective in a household. Consider the infection of a susceptible household making one member of the household infectious and assume for the moment that no further global infections are made with the household. Then after a waiting time of $\text{Exp}(1 + \lambda_L)$ either the second member of the household is infected (with probability $\lambda_L/(1 + \lambda_L)$) or the infective recovers (with probability $1/(1 + \lambda_L)$). If the second member of the household becomes infected the time until one of these individual recovers is $\text{Exp}(2)$. This recovery will shortly be followed by an infection (with probability $\lambda_L/(1 + \lambda_L)$) or a second recovery (with probability $1/(1 + \lambda_L)$). Let $A_1, A_2, \ldots$ be independent and identically distributed according to $A \sim \text{Exp}(2)$ and let $B_0, B_1, \ldots$ be independent and identically distributed according to $B \sim \text{Exp}(1 + \lambda_L)$. Let $G \sim Geom(1/(1 + \lambda_L))$ with support on $\mathbf{Z}^+$. Let R be the time from infection of the household until it recovers and S be the sum of the total time infectious of the two individuals in the household from global infection until recovery. Then

$$R = B_0 + \sum_{j=1}^{G} (A_j + B_j),$$

where the sum is 0 if $G = 0$ and

$$S = \sum_{j=0}^{G} B_j + 2 \sum_{j=1}^{G} A_j.$$

Now $\sum_{j=0}^{G} B_j \sim \text{Exp}(1)$ and

$$\sum_{j=1}^{G} A_j = \begin{cases} 0 & \text{with probability } \frac{1}{1+\lambda_L} \\ \text{Exp}(2/(1 + \lambda_L)) & \text{with probability } \frac{\lambda_L}{1+\lambda_L}. \end{cases}$$

Consequently, for large λ_L, $R \approx \tilde{R} \sim \text{Exp}(2/(1 + \lambda_L))$ and whilst both members of the household are infectious the household is generating global infections at the points of a homogeneous Poisson point process with rate $2\lambda_G$. Since λ_L is large, it is highly

unlikely that a global infectious contact (λ_G is necessarily small) with a household with at least one infective in it will be with a susceptible individual. Therefore for large λ_L, the total number of infectious households (approximately the total number of infectives divided by 2) can be approximated by an SIS epidemic in a homogeneously mixing population of size $N/2$ with infection rate $2\lambda_G$ and recovery rate $2/(1 + \lambda_L)$. From (2), this gives our second approximation for the time to extinction

$$E(T_Q) \approx \frac{1 + \lambda_L}{2} \times \sqrt{\frac{2\pi}{N/2}} \frac{R_*}{(R_* - 1)^2} e^{\frac{N}{2}(\log R_* - 1 + 1/R_*)}, \qquad (11)$$

where $R_* = \lambda_G(1 + \lambda_L)$. Thus the time to extinction depends upon λ_L and N. This suggests that the mean time to extinction will approximately grow linearly in λ_L, for fixed $\hat{i}$, as $\lambda_L \to \infty$. (For large λ_L, $R_* \approx 1/(1 - \hat{i})$.)

This result can be extended to households where $k > 2$. In general, for large values of λ_L, the number of infectious households can be approximated by a homogenously mixing SIS epidemic model with infection rate $k\lambda_G$ and recovery rate $k \prod_{j=1}^{k-1} \frac{k-1}{k-1+\lambda_L(k-j)}$. For fixed N, the mean extinction time will behave like $C_N \lambda_L^{k-1}$ as $\lambda_L \to \infty$ for some $C_N > 0$. Also $\lim_{\lambda_L \to \infty} \sigma_i^2(\hat{i}, \lambda_L) = k\hat{i}^{-1}$. Thus the effect of λ_L (households) on mean time to extinction is more marked for larger values of k. A similar result holds for unequal sized households.

There is an apparent contradiction between the two approximations with the Ornstein-Uhlenbeck approximation (10) stating that $E(T_Q)$ *decreases* in λ_L for large enough λ_L and the second approximation (11) stating that $E(T_Q)$ *increases* in λ_L as $\lambda_L \to \infty$. The explanation for this is that there are two asymptotic regimes considered: $N \to \infty$ (Ornstein-Uhlenbeck approximation) and $\lambda_L \to \infty$ (second approximation). For fixed population size, N, there is a cross-over from the Ornstein-Uhlenbeck approximation to the second approximation as λ_L increases. Hence this suggests that as λ_L changes from 0 to ∞ and when $\hat{i} < 0.5$ (the most common situation) the mean time to extinction should first increase, then decrease and, when moving over to the second approximation, eventually start increasing again. If we instead consider the situation where N increases, the transition between the two approximations occurs at increasing values of λ_L. Thus as $N \to \infty$, the Ornstein-Uhlenbeck approximation dominates for all values of λ_L.

Equations (2) and (11) give approximate mean times to extinction in the cases where λ_L is close to 0 and λ_L is large, respectively. The question remains of estimating the time to extinction for moderate values of λ_L since as noted by Doering *et al.* [7] the Ornstein-Uhlenbeck estimation of q_1 can be several orders of magnitude too small even for moderate N. However, the Ornstein-Uhlenbeck approximation can be used for a qualitative assessment of how the mean time to extinction changes when departing from the homogeneous case. That is, we can compare the Ornstein-Uhlenbeck approximation for q_1 under the assumption of homogeneously mixing and for moderate values of λ_L for a given value of $\hat{i}$. For example, for $\hat{i} < 0.5$, does the mean extinction time increase for small values of λ_L before decreasing as suggested by the Ornstein-Uhlenbeck approximation? Furthermore, what are the competing influences of N and λ_L through the two approximations on the mean time to extinction? These questions are addressed in the following section.
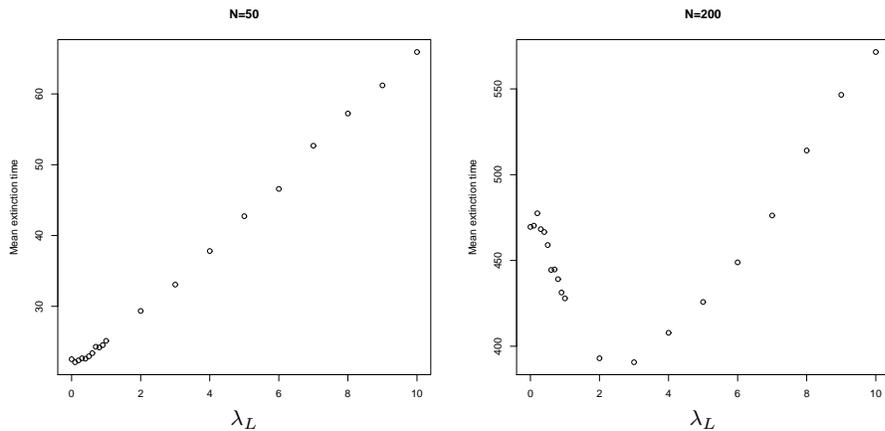
Figure 1: Mean extinction times from 10000 simulations: Left figure $N = 50$ and right figure $N = 200$.

4 Numerical examples and simulations

For the numerical examples and simulations we focus attention upon $\hat{i} = 0.2$. In the homogeneous case this corresponds to $\lambda_G = 1.25$. Thus $\sigma^2(0.2, 0) = 5$ and $\lim_{\lambda_L \to \infty} \sigma^2(0.2, \lambda_L) = 10$. A simulation study involving $N = 50$ and $N = 200$ and $\lambda_L = 0, 0.1, 0.2, \ldots, 1, 2, \ldots, 10$ was run, with for each set of parameter values the mean time to extinction estimated from 10000 simulations starting at the endemic level (see Figure 1). For $N = 50$, the mean time to extinction was found to grow linearly in λ_L. The correlation between λ_L and mean time to extinction was found to be 0.9992. The case $\lambda_L = 50$ was also tested and found to satisfy the trend found for smaller values of λ_L. We conclude that the Ornstein-Uhlenbeck process (relying on N to be large) is not applicable for this case, and the second approximation is better for the whole range of λ_L. However, as N increased a different story emerged. First, for very small λ_L $E(T_Q)$ seems to increase slightly and then decreases up until $\lambda_L \approx 2.5$. Hence this part agrees with the behaviour suggested by the first approximation. After this, i.e. for $\lambda_L > 2.5$, $E(T_Q)$ starts growing close to linearly with λ_L as suggested by the second approximation (11).

Similar results can be obtained where all households are of size $k > 2$. In such cases, as $\lambda_L \to \infty$, the mean extinction time for fixed N and $\hat{i}$ increases polynomially in λ_L. For moderate values of λ_L and $\hat{i} < 0.5$, the mean time to extinction is maximized by small (but positive) values of λ_L as with households of size 2. This is illustrated in Figure 2 below where households of size $k = 5$ are considered, and where the mean extinction time has been simulated for $N = 50$ and $N = 200$. The results are similar with those for $k = 2$, only more markedly. When $N = 50$ the second approximation is better for all λ_L (suggesting polynomial growth of $E(T_Q)$ as a function of λ_L) whereas when $N = 200$ the first approximation (Ornstein-Uhlenbeck), suggesting that $E(T_Q)$ should first increase and then decrease, works for small and moderate λ_L, and after that, the second approximation suggesting polynomial growth in λ_L performs better.

In Figure 3, the ratio of the mean time to extinction with $N = 200$ divided by the mean time to extinction with $N = 50$ is plotted. By (10), this ratio should be
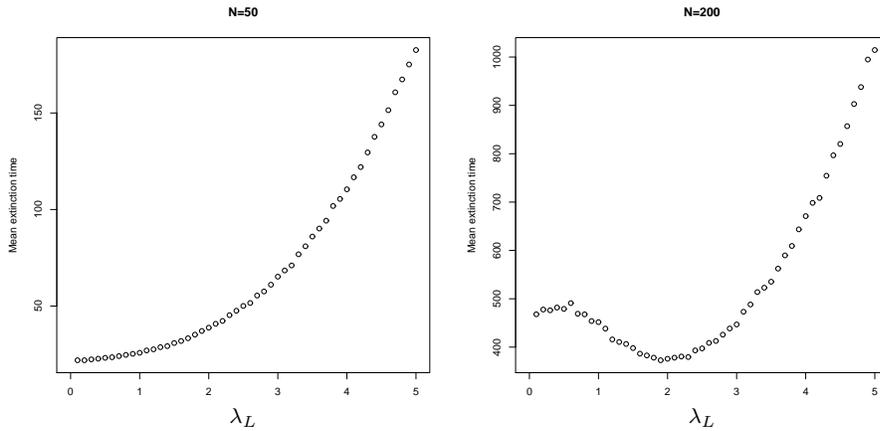
Figure 2: *Mean extinction times from 10000 simulations in the case $k = 5$: Left figure $N = 50$ and right figure $N = 200$.*
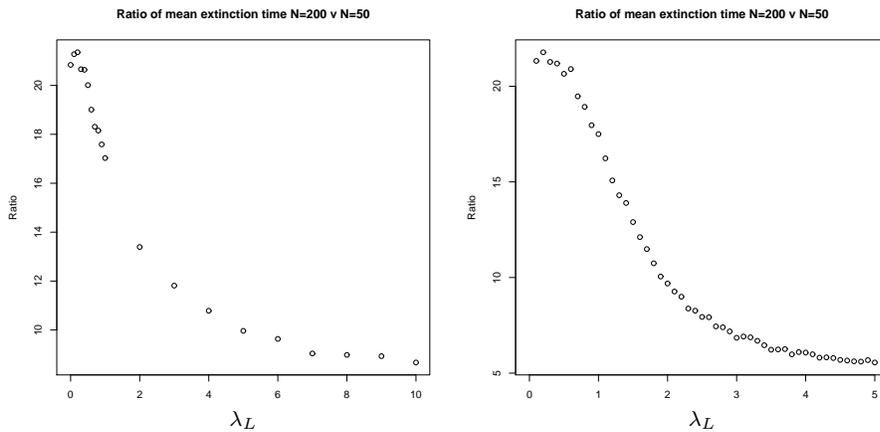

Figure 3: *Ratio of mean extinction times from 10000 simulations for $N = 200$ and $N = 50$. Left figure $k = 2$ and right figure $k = 5$*

approximately,

$$\frac{\sqrt{2\pi \times 200}\sigma(\hat{i}, \lambda_L)\exp(200\hat{i}^2/2\sigma^2(\hat{i}, \lambda_L))}{\sqrt{2\pi \times 50}\sigma(\hat{i}, \lambda_L)\exp(50\hat{i}^2/2\sigma^2(\hat{i}, \lambda_L))} = 2\exp(3/\sigma^2(\hat{i}, \lambda_L)). \qquad (12)$$

Thus (12) states we should expect to see the ratio between the extinction times initially increase before decreasing as λ_L goes from 0 to infinity, corresponding to $\sigma^2(\hat{i}, \lambda_L)$ initially decreasing before increasing as $\lambda_L \to \infty$. Figure 3 is consistent, at least qualitatively, with the Ornstein-Uhlenbeck approximation. Similar results are observed with $N = 500$ with the Ornstein-Uhlenbeck approximation 'valid' for larger values of λ_L.

For $\hat{i} > 0.5$, we would expect to see that the extinction time decreases as λ_L starts increasing from 0, in the limit as $N \to \infty$. Analysis of this case is not presented for two reasons. Firstly, in realistic situations $\hat{i}$ is likely to be small and at the very least less than 0.5. Secondly, the mean extinction time is very large even for small N.

For example, for $\hat{i} = 0.8$ and $N = 50$, in the homogeneous mixing case $E[\tau_N^{(AD)}] = 4.18 \times 10^{16}$ compared with 22.55 when $\hat{i} = 0.2$.

5 Discussion

We present two approximations for obtaining the mean time to extinction in an SIS household epidemic model. Approximating for large values of λ_L the SIS household epidemic model by a homogeneously mixing epidemic model with the households treated as individuals is informative and explains the trend observed for the mean time to extinction in the simulation study as λ_L increases. However, it is only really useful when λ_L is several orders of magnitude larger than λ_G. The most interesting case is when λ_G and λ_L are of the same order of magnitude. In that case we have resorted to the crude Ornstein-Uhlenbeck approximation to gain an understanding in the mean time to extinction. Although this approximation can severely over-estimate the mean time to extinction it does perform very well in giving a qualitative assessment of how the mean time to extinction varies with λ_L.

As mentioned in the introduction, better approximations are in principle available using large deviations, see for example, Shwartz and Weiss [18]. However, in practice it is difficult, if indeed possible, to get an explicit expression for the large deviations calculations.

The SIS epidemic model is the simplest epidemic model which exhibits endemic behaviour. It would be interesting to extend the above analysis to more realistic epidemic models with household structure. A prime example would be an SIR epidemic model with demography (births of susceptible individuals) extending the work of Andersson and Britton [3] to include household structure.

It would also be interesting to consider the mean time to extinction in other SIS epidemic models with heterogeneous mixing of individuals. Examples include the great circle model (Neal [16]) and epidemics upon random graphs (*e.g.* Andersson [2]).

References

1. Anderson, R. and May, R. (1991) *Infectious Diseases of Humans: Dynamics and Control.* Oxford University Press, Oxford.
2. Andersson, H. (1999) Epidemic models and social networks. *Math. Scientist* **24,** 128–147.
3. Andersson, H. and Britton, T. (2000) Stochastic epidemics in dynamic populations: quasi-stationarity and extinction. *J. Math. Biol.* **41**, 559–580.
4. Andersson, H. and Djehiche, B. (1998) A threshold limit theorem for the stochastic logistic epidemic. *J. Appl. Prob.* **35**, 662–670.
5. Ball, F. (1999). Stochastic and deterministic models for SIS epidemics among a population partitioned into households. *Math. Biosci.* **156**, 41–67.
6. Cauchemez, S., Valleron, A.J., Bolle, P.Y., Flahault, A. and Ferguson, N.M. (2008). Estimating the impact of school closure on influenza transmission from Sentinel data. *Nature* **452**, 750–754.
7. Doering, C., Sargsyan, K. and Sander, L. (2005) Extinction times for birth-death processes: exact results, continuum asymptotics, and the failure of the Fokker-Planck approximation. *Multiscale Model. Simul.* **3**, 283–299.

8. Ethier, S. and Kurtz, T. (1986) *Markov Processes: Characterization and Convergence*. Wiley, New York.

9. Ferguson, N.M., Cummings, D.A., Fraser, C., Cajka, J.C., Cooley, P.C. and Burke, D.S. (2006). Strategies for mitigating an influenza pandemic. *Nature.* **442**, 448–452.

10. Ghoshal, G. Sander, L. and Sokolov, I. (2004). SIS epidemics with household structure: the self-consistent field method. *Math. Biosci.* **190**, 71–85.

11. Hagenaars, T., Donnelly, C. and Ferguson, N. (2004) Spatial heterogeneity and the persistence of infectious diseases. *J. Theoret. Biol.* **229**, 349–359.

12. House, T. and Keeling, M. (2008). Deterministic epidemic models with explicit household structure. *Math. Biosci.* **213**, 29-39.

13. Kryscio, R. and Lefèrve, C. (1989). On the extinction of the SIS stochastic logistic epidemic. *J. Appl. Prob.* **26**, 685–694.

14. Lindholm, M. (2008) On the time to extinction for a two-type version of Bartlett's epidemic model. *Math. Biosci.* **212**, 99–108.

15. Neal, P. (2006) Stochastic and deterministic analysis of SIS household epidemics. *Adv. Appl. Prob.* **38**, 943–968.

16. Neal P. (2008) The SIS Great Circle Epidemic model. *J. Appl. Prob.* **45**, 513–530.

17. Nåsell, I. (1999) On the time to extinction in recurrent epidemics. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61**, 309–330.

18. Shwartz, A. and Weiss, A. (1995) *Large deviations for performance analysis.* Chapman and Hall, London.

19. Wallinga, J. and Teunis, P. (2004) Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *Am. J. Epidemiol.* **160**, 509–516.