

Stochastic models of some endemic infections

Ingemar Nåsell *

Department of Mathematics, The Royal Institute of Technology, S-100 44 Stockholm, Sweden

Received 6 February 2001; received in revised form 30 April 2001; accepted 7 May 2001

Abstract

Stochastic models are established and studied for several endemic infections with demography. Approximations of quasi-stationary distributions and of times to extinction are derived for stochastic versions of SI, SIS, SIR, and SIRS models. The approximations are valid for sufficiently large population sizes. Conditions for validity of the approximations are given for each of the models. These are also conditions for validity of the corresponding deterministic model. It is noted that some deterministic models are unacceptable approximations of the stochastic models for a large range of realistic parameter values. © 2002 Published by Elsevier Science Inc.

Keywords: Quasi-stationarity; Time to extinction; Demographic stochasticity

1. Introduction

Both deterministic and stochastic models are used for epidemiological modelling. It is therefore of importance to understand the relations between these two kinds of models, and between the results that they predict. Both model types are needed, and both have their advantages and weaknesses. The deterministic models often lead to powerful qualitative results with important threshold behaviour. They also lead to simpler mathematical problems than the stochastic ones. Work on deterministic models has therefore dominated strongly over work on stochastic models.

The models considered here are among the simplest possible, where the deterministic model takes the form of a system of ordinary differential equations, and the corresponding stochastic model is a Markov population process with continuous time and discrete state space. It is important to realize that every such deterministic model is an approximation of a corresponding stochastic model. It can be derived from the stochastic model by dividing the original state

* Tel.: +46-87 90 7555; fax: +46-87 23 1788.

E-mail address: ingemar@math.kth.se (I. Nåsell).

variables by a constant measure N of the population size and taking the limit as N approaches infinity. An important difference between the two models is that the stochastic model deals with a finite population size, while the deterministic model only deals with proportions; the population size measure N is an essential parameter in the stochastic model, but a so-called innocent parameter in the deterministic one. This means that if the population size measure N appears at all in the deterministic model, then it can be eliminated completely by a rescaling of the state variables. The stochastic models account for what is called demographic stochasticity.

The deterministic model is clearly an approximation of the stochastic one, since real population sizes are always finite. It is well known that the deterministic model is an acceptable approximation of the stochastic one if the population size is sufficiently large. But this fact is in itself not a guarantee that all deterministic models are acceptable. A mathematical study of the stochastic model is required in order to quantify what is meant by a ‘sufficiently large’ population size.

In this paper I study both deterministic and stochastic models for some endemic infections in closed populations, accounting for both epidemic and demographic forces. Qualitative results are of major importance in mathematical epidemiology. It is therefore important to observe that a qualitative difference exists between the deterministic and stochastic models that we are concerned with. Above a threshold determined by the parameters of the model, the deterministic model predicts that the proportion of infected individuals will approach a positive endemic level as time approaches infinity, while the stochastic model predicts that the infection will become extinct. The time to extinction is a random variable whose distribution clearly depends on the distribution of initial states. If the process has been going on for a long time and has not gone extinct, then the so-called quasi-stationary distribution can be used as an approximation of the distribution of states. The time to extinction from the quasi-stationary distribution is therefore a measure of the persistence of the infection, see [1]. The quasi-stationary distribution is of additional importance, since the expected time to extinction from quasi-stationarity can be expressed in terms of this distribution. An explicit expression for the quasi-stationary distribution cannot be determined. A major goal of the analysis of the stochastic model is therefore to derive an approximation of the quasi-stationary distribution. This derivation is based on a diffusion approximation of the stochastic discrete-state model. The methods used for this are similar to the methods introduced in [2].

The analysis in this paper is concerned with SI, SIS, SIR, and SIRS models. Here, the letters S, I, and R refer to susceptible, infective, and recovered individuals, respectively. The SI model deals with infection without recovery. The SIS model is a model for an infection without immunity, where recovery is possible, and where recovered individuals are immediately susceptible. The SIR model deals with an infection where recovered individuals are completely immune. The SI and SIR models without demography were referred to as ‘the simple epidemic’ and ‘the general epidemic’, respectively, by Bailey in [3]. His discussion deals with both deterministic and stochastic model versions. The SIR model without demography has transients, corresponding to epidemic outbreaks. In sharp contrast, the SIR model with demography admits an almost stationary behaviour, corresponding to endemic infections. (To be precise, the deterministic model has stationary solutions, while the stochastic model has quasi-stationarity.) Accordingly, the recent review paper [4] by Hethcote refers to the SIR model without demography as ‘the classic epidemic model’, while the SIR model with demography is described as ‘the classic endemic model’. The analysis in [4] is limited to deterministic models. The SIRS model deals with a situation where the

immunity is temporary. Our results allow an estimate of the time to extinction and of the conditions for the deterministic model to be an acceptable approximation of the fully stochastic model.

The number of susceptible, infected, and recovered individuals are denoted by S , I , and R , respectively, in all the models studied, where $R = 0$ for the SI and SIS models. All infected hosts are assumed to be infective. We adopt the barbaric idea of using the same notation in both deterministic and stochastic models. Thus, S , I , and R are random variables with discrete state space in the stochastic models and continuous variables in the corresponding deterministic models. The state variables are generally functions of time, although the main interest is focussed on steady-state or quasi-stationary behaviour.

I assume that the population size is determined by an immigration-death process with differential death rate. The immigration rate is put equal to μN and the death rate per individual is μ for susceptible and recovered hosts and $\mu + \mu_1$ for infected hosts, where $\mu_1 \geq 0$. Thus, N is equal to the (steady-state) expected population size if no infection is present in the community. The methods used here do not allow any conclusions to be drawn about the epidemiologically interesting case of an exponentially growing population.

All the models studied are based on the ‘standard incidence’ considered by [5], and homogeneous exposure. To model this I assume a constant contact rate β . The infection rate at time t will in each of the models be expressed as $\beta si/(s + i + r)$ (with $r = 0$ for the SI and the SIS models), where s , i , and r are the values taken by the state variables S , I and R , respectively, at time t . This expression for the infection rate is equal to the product of the contact rate β , the proportion $i/(s + i + r)$ of the population that is infected and the number of susceptible individuals s . Note that this expression for the infection rate deviates from the ‘conventional’ expression $\beta si/N$ by having a denominator that depends on the state variables. This deviation affects the deterministic model time-dependent behaviour if initially the populations of susceptible and infected individuals are not in steady state. It also affects the important steady state behaviour of the deterministic model if there are disease related deaths, i.e. if $\mu_1 > 0$. Note also that if there are no disease related deaths, i.e. if $\mu_1 = 0$, then the steady-state behaviour of the deterministic model is the same for the two infection rate expressions. The difference between hypothesized and conventional infection rates will, however, always have an effect on the predictions of the stochastic model. The deterministic versions of the SIS, SIR, and SIRS models, without differential mortality, and with conventional infection rate, were studied by Hethcote in 1976 [6]. These models are the starting point for our study.

The three models SI, SIS, and SIR can be derived as special cases of the SIRS model. The deterministic SIRS model has been treated in [5]. The treatment of the stochastic version of each of the models is preceded by the formulation and analysis of the corresponding deterministic model. The consideration of the deterministic model allows a reparametrization of the model. Non-dimensional parameters are introduced that are of importance in the analysis of both the deterministic and the stochastic versions of each model.

Analysis of the stochastic version of the SI model with demography and differential death rate and proper infection rate is given in [7]. Analysis of the quasi-stationary distribution and of the time to extinction for the SIS model without demography is contained in [8] and [9]. In these two references, the population size is constant, so the proper infection rate is actually equal to the conventional infection rate. Analysis of an SIR model with demography is given in [2]. This model does not account for differential mortality. It uses a conventional infection rate as an

approximation of the proper infection rate. The analysis in the present paper allows us to comment on the consequences of this approximation.

The deterministic version of either of the models we consider here has a threshold that is described in terms of the so-called basic reproduction ratio R_0 , which is defined for each model in terms of basic parameters. The threshold value $R_0 = 1$ for the deterministic model identifies two parameter regions, namely $R_0 > 1$ and $R_0 < 1$, with qualitatively different behaviours. The qualitative behaviour of the stochastic version of the model is described by identification of three parameter regions with qualitatively different behaviours. The boundaries between these regions depend on the value of N . The ‘extra’ region appearing in the stochastic model is a transition region near $R_0 = 1$. The width of the transition region approaches zero as $N \rightarrow \infty$. This explains why this region is absent from the deterministic version of the model. The quasi-stationary distribution can be expected to be definitely non-normal in the transition region. The problem of finding a useful approximation of this distribution is mathematically challenging. Some progress has been made on this problem for the univariate SIS model in [9], while it remains unsolved in any of the bivariate or multivariate models dealt with in this paper.

The results given in this paper are confined to the parameter region above the transition region. It is therefore of considerable practical interest to know how large R_0 and N must be in each model in order for the results to be valid. Such results are given for each model in terms of the rescaling of R_0 as function of N that is necessary for the study of the model in the transition region.

The analysis of all the models uses a diffusion approximation similar to that introduced in [2]. An important step in this analysis requires one to solve a linear system of equations, with coefficients depending on the parameters of the model. The expressions that appear are in many cases rather complicated. I have used Maple to solve this and related problems. Results are quoted here, while the Maple programs that have been used to derive these results are found at the web address <http://www.math.kth.se/~ingemar/forsk/desto/desto.html>.

2. A stochastic SI model with demography

The mathematical treatments of all models we deal with are rather similar. Some of the details in the treatment of the simplest of the models, the SI model, are given in this section. These details are not repeated in the following sections.

2.1. Model formulation

I introduce four parameters, namely the expected population size N in case there are no infected individuals, the contact rate β , the death rate per susceptible individual μ , and an additional death rate μ_1 such that the death rate per infected individual is $\mu + \mu_1$. Among these, N is a large positive integer, β and μ are positive, and μ_1 is non-negative.

The hypotheses of the model are described by the transition rates in Table 1. The two state variables $S(t)$ (the number of susceptible individuals) and $I(t)$ (the number of infected individuals) take values in the state space $\{(s, i): s = 0, 1, \dots, i = 0, 1, \dots\}$. The states $(s, 0)$ communicate with each other, but not with any state (s, i) with $i \geq 1$. The quasi-stationary distribution is supported on the set of transient states (s, i) with $i \geq 1$. The stationary distribution is supported on the set of states $(s, 0)$, where the infection is extinct.

Table 1
Transition rates for the SI model of Section 2

Event	Transition	Transition rate
Immigration of susceptible	$(s, i) \rightarrow (s + 1, i)$	$\lambda_1(s, i) = \mu N$
Death of susceptible	$(s, i) \rightarrow (s - 1, i)$	$\bar{\mu}_1(s, i) = \mu s$
Death of infected	$(s, i) \rightarrow (s, i - 1)$	$\mu_2(s, i) = (\mu + \mu_1)i$
Infection of susceptible	$(s, i) \rightarrow (s - 1, i + 1)$	$v_2(s, i) = \beta si / (s + i)$

The Kolmogorov forward equations for the state probabilities $p_{si}(t) = P\{S(t) = s, I(t) = i\}$ can be written

$$p'_{si}(t) = \lambda_1(s - 1, i)p_{s-1,i}(t) + \bar{\mu}_1(s + 1, i)p_{s+1,i}(t) + \mu_2(s, i + 1)p_{s,i+1}(t) + v_2(s + 1, i - 1)p_{s+1,i-1}(t) - \kappa(s, i)p_{si}(t),$$

where

$$\kappa(s, i) = \lambda_1(s, i) + \bar{\mu}_1(s, i) + \mu_2(s, i) + v_2(s, i).$$

2.2. The deterministic model and reparametrization

The deterministic version of the model leads to the following system of differential equations:

$$S' = \mu N - \beta \frac{SI}{S + I} - \mu S,$$

$$I' = \beta \frac{SI}{S + I} - (\mu + \mu_1)I.$$

The model is reparametrized by the introduction of the following new parameters:

$$R_0 = \frac{\beta}{\mu + \mu_1}, \tag{2.1}$$

$$\delta_1 = \frac{\mu + \mu_1}{\mu}. \tag{2.2}$$

These parameters are both free of dimension. R_0 is referred to as the basic reproduction ratio. The deterministic version of the model has its threshold at $R_0 = 1$. Among the parameters, R_0 is strictly positive, while δ_1 is larger than or equal to one. Note that δ_1 equals one if there is no differential death rate.

For the derivation of diffusion approximations of the stochastic version of the model it will be convenient to introduce the scaled state variables $x_1 = S/N$ and $x_2 = I/N$. The differential equations for the deterministic version of the model can be written as follows for these state variables, using the reparametrization introduced above:

$$x'_1 = \mu - \mu\delta_1 R_0 \frac{x_1 x_2}{x_1 + x_2} - \mu x_1, \tag{2.3}$$

$$x'_2 = \mu\delta_1 R_0 \frac{x_1 x_2}{x_1 + x_2} - \mu\delta_1 x_2. \tag{2.4}$$

This system of differential equations has two critical points. One of them is at $(1, 0)$, which corresponds to the absence of infection. The other critical point corresponds to an endemic infection level. Its coordinates are $\hat{x} = (\hat{x}_1, \hat{x}_2)$, where

$$\hat{x}_1 = \frac{1}{\delta_1(R_0 - 1) + 1},$$

$$\hat{x}_2 = \frac{R_0 - 1}{\delta_1(R_0 - 1) + 1}.$$

2.3. Analysis of the stochastic model

The probability of having zero infected individuals at time t is of importance both for the derivation of the quasi-stationary distribution and for the study of the time to extinction. A differential equation is derived for this probability. I write $p_{\cdot i}(t) = \sum_{s=0}^{\infty} p_{si}(t) = P\{I(t) = i\}$ for the marginal distribution of the number of infected individuals at time t . By adding the Kolmogorov forward equations for $i = 0$ over all non-negative values of S one obtains the differential equation

$$p'_{\cdot 0}(t) = \mu\delta_1 p_{\cdot 1}(t) \quad (2.5)$$

for the probability $p_{\cdot 0}(t)$ of zero infected individuals at time t .

The quasi-stationary distribution is a stationary distribution, conditioned on non-extinction. The state probabilities conditioned on non-extinction at time t are denoted $\tilde{q}_{si}(t)$ and given by

$$\tilde{q}_{si}(t) = P\{S(t) = s, I(t) = i | I(t) > 0\} = \frac{p_{si}(t)}{1 - p_{\cdot 0}(t)}, \quad s = 0, 1, \dots, \quad i = 1, 2, \dots$$

By differentiating this expression with respect to t and applying the differential Eq. (2.5) above for the function $p_{\cdot 0}(t)$ one obtains the differential equation

$$\tilde{q}'_{si}(t) = \frac{p'_{si}(t)}{1 - p_{\cdot 0}(t)} + \mu\delta_1 \tilde{q}_{s-1,i}(t) \frac{p_{si}(t)}{1 - p_{\cdot 0}(t)}. \quad (2.6)$$

Here, $\tilde{q}_{\cdot i}(t) = \sum_{s=0}^{\infty} \tilde{q}_{si}(t)$ denotes the marginal distribution of the number of infected individuals at time t , conditioned on $I(t) > 0$. A system of differential equations for the probabilities $\tilde{q}_{si}(t)$ can be derived from this expression by using the Kolmogorov forward equations for the probabilities $p_{si}(t)$. The quasi-stationary distribution q_{si} is the stationary solution of this system of equations.

The distribution of the time to extinction τ can be determined from the probability $p_{\cdot 0}(t)$, since the event that τ is less than or equal to t is equal to the event that $I(t) = 0$. Hence we get

$$P\{\tau \leq t\} = P\{I(t) = 0\} = p_{\cdot 0}(t).$$

Thus the probability that the time to extinction is less than or equal to t equals the probability that the number of infected individuals at time t equals 0.

The distribution of the time to extinction is especially simple when the initial distribution is equal to the quasi-stationary distribution, i.e. when $p_{si}(0) = q_{si}$ for $i \geq 1$ and $p_{s0}(0) = 0$. The time to extinction with this initial distribution is denoted by τ_Q . I show that τ_Q has an exponential distribution and that its expected value is equal to

$$E(\tau_Q) = \frac{1}{\mu\delta_1 q_{\cdot 1}}.$$

To derive this result we put $\tilde{q}'_{si}(t) = 0$ in (2.6). Thus I am led to the initial value problems

$$p'_{si}(t) = -\mu\delta_1 q_{\cdot 1} p_{si}(t), \quad p_{si}(0) = q_{si}, \quad s = 0, 1, \dots, \quad i = 1, 2, \dots,$$

with solutions

$$p_{si}(t) = q_{si} \exp(-\mu\delta_1 q_{\cdot 1} t), \quad s = 0, 1, \dots, \quad i = 1, 2, \dots$$

This says that probability is leaking from all states with $i \geq 1$ with the same rate. By adding these expressions for $p_{si}(t)$ over all values of s one gets

$$p_i(t) = q_{\cdot i} \exp(-\mu\delta_1 q_{\cdot 1} t), \quad i = 1, 2, \dots$$

The differential equation for $p_0(t)$ in (2.5) can now be solved since its right-hand side is known. Using the initial value $p_0(0) = 0$, I get

$$p_0(t) = 1 - \exp(-\mu\delta_1 q_{\cdot 1} t).$$

This establishes that τ_Q has an exponential distribution with expectation equal to $1/(\mu\delta_1 q_{\cdot 1})$.

Approximating normal distributions are derived of the quasi-stationary distribution in the Sections 2.4 and 2.5. This means that the marginal distribution of the number of infected individuals in quasi-stationarity is normal. The normal distribution is modified by a truncation at 0.5, to account for the fact that the number of infected individuals in quasi-stationarity is larger than or equal to one. With this modification the probability $q_{\cdot 0}$ is approximated as follows:

$$q_{\cdot 0} \approx \frac{1}{\sigma_I} \frac{\varphi((\mu_I - 1)/\sigma_I)}{\Phi((\mu_I - 0.5)/\sigma_I)},$$

where μ_I and σ_I are the mean and standard deviation, respectively, of the marginal distribution of the number of infected individuals in quasi-stationarity, and φ and Φ denote the normal density and the normal distribution function.

This approximation of the expected time to extinction from quasi-stationarity is coarse. The reason is that the normal approximation of the marginal distribution of I in quasi-stationarity is valid only in the parameter region formally defined by the condition that $R_0 > 1$ as $N \rightarrow \infty$, and there it is only valid in the body of the distribution. However, the state $I = 1$ will then lie in the left tail of the marginal distribution of I , and one can expect that the distribution will there deviate from normality. On the other hand, if the condition $R_0 > 1$ as $N \rightarrow \infty$ is not satisfied, then the state $I = 1$ will belong to the body of the marginal distribution of I , but the body of the distribution is then not normal.

2.4. Diffusion approximation for $\mu_1 = 0$

The analysis of the process formulated above has as its main goal a derivation of the quasi-stationary distribution when R_0 is distinctly larger than one. This is done by the aid of a diffusion approximation. The diffusion approximation has a continuous state space, in contrast with the discrete state space of the original process. It is derived under the restrictions that N is large and that R_0 is strictly larger than one. The main result is that the quasi-stationary distribution is then approximated by a bivariate normal distribution.

The case when susceptible and infected individuals have the same death rate, i.e. when $\mu_1 = 0$, will be treated in the present section, while I defer the treatment of the case $\mu_1 > 0$ to Section 2.5. The method for derivation of diffusion approximations is common for all the models treated in

this paper. Some of the details in the method are spelled out in the present section, and are not repeated in the following sections.

The right-hand side of the system of differential equations (2.3) and (2.4) is denoted $b(x)$, and the Jacobian matrix of the vector $b(x)$ with respect to $x = (x_1, x_2)$ is denoted $B(x)$:

$$B(x) = \frac{\delta b(x)}{\delta x} = \mu \begin{pmatrix} A - Bx_2 - 1 & A - Bx_1 \\ -A + Bx_2 & -A + Bx_1 - 1 \end{pmatrix},$$

where

$$A = \frac{R_0 x_1 x_2}{(x_1 + x_2)^2},$$

$$B = \frac{R_0}{x_1 + x_2}.$$

We approximate the matrix $B(x)$ by evaluating it at the critical point $\hat{x} = (1/R_0, (R_0 - 1)/R_0)$:

$$B(\hat{x}) = \frac{\mu}{R_0} \begin{pmatrix} -(R_0 - 1)^2 - R_0 & -1 \\ (R_0 - 1)^2 & -(R_0 - 1) \end{pmatrix}.$$

Next I determine the covariance matrix of the vector of changes in the state variables x_1 and x_2 during the time interval $(t, t + \Delta t)$. I write $\Delta x_i = x_i(t + \Delta t) - x_i(t)$, $i = 1, 2$. By applying the hypotheses of the model I get

$$\text{cov} \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \end{pmatrix} = \frac{\mu}{N} \begin{pmatrix} 1 + x_1 + Bx_1x_2 & -Bx_1x_2 \\ -Bx_1x_2 & Bx_1x_2 + x_2 \end{pmatrix} \Delta t + o(\Delta t) = \frac{S(x)}{N} + o(\Delta t).$$

The matrix $S(x)$ is approximated by evaluating it at the critical point \hat{x} . The result is

$$S(\hat{x}) = \frac{\mu}{R_0} \begin{pmatrix} 2R_0 & -(R_0 - 1) \\ -(R_0 - 1) & 2(R_0 - 1) \end{pmatrix}.$$

The process $\sqrt{N}(x(t) - \hat{x})$ is approximated for large N and $R_0 > 1$ by a bivariate Ornstein–Uhlenbeck process with local drift matrix $B(\hat{x})$ and local covariance matrix $S(\hat{x})$. Its stationary distribution approximates the quasi-stationary distribution. This approximation is bivariate normal with mean 0 and covariance matrix Σ . The matrix Σ is determined from the matrices $B(\hat{x})$ and $S(\hat{x})$ through the equation

$$B(\hat{x})\Sigma + \Sigma B^T(\hat{x}) = -S(\hat{x}), \tag{2.7}$$

where the superscript T is used to denote transpose. By solving this equation with the above expressions for $B(\hat{x})$ and $S(\hat{x})$ I get

$$\Sigma = \frac{1}{R_0^2} \begin{pmatrix} (R_0 - 1)^2 + R_0 & -1 \\ -1 & R_0 + 1 \end{pmatrix}.$$

These results allow the conclusion that the quasi-stationary distribution is approximated by a bivariate normal distribution when $R_0 > 1$ and N is large. The expected values of the marginal distributions of the number of susceptible and the number of infected individuals in quasi-stationarity agree with the corresponding steady-state values from the deterministic version of the model. In particular, we find from above the following expressions for the mean and the variance of the marginal distributions in quasi-stationarity for the number of susceptible and infected individuals:

$$\begin{aligned} \mu_S &\approx \frac{1}{R_0}N, & \sigma_S^2 &\approx \frac{R_0 + 1}{R_0^2}N, \\ \mu_I &\approx \frac{R_0 - 1}{R_0}N, & \sigma_I^2 &\approx \frac{(R_0 - 1)^2 + R_0}{R_0^2}N. \end{aligned}$$

Furthermore, the covariance of S and I in quasi-stationarity is

$$\sigma_{SI} \approx -\frac{1}{R_0^2}N.$$

2.5. Diffusion approximation for $\mu_1 > 0$

For the derivation of the diffusion approximation one starts by determining the local drift matrix $B(\hat{x})$ and the local covariance matrix $S(\hat{x})$. They can be written as follows:

$$B(\hat{x}) = \frac{\mu}{R_0} \begin{pmatrix} -\delta_1(R_0 - 1)^2 - R_0 & -\delta_1 \\ \delta_1(R_0 - 1)^2 & -\delta_1(R_0 - 1) \end{pmatrix}$$

and

$$S(\hat{x}) = \frac{\mu}{\delta_1(R_0 - 1) + 1} \begin{pmatrix} 2[\delta_1(R_0 - 1) + 1] & -\delta_1(R_0 - 1) \\ -\delta_1(R_0 - 1) & 2\delta_1(R_0 - 1) \end{pmatrix}.$$

The covariance matrix Σ is determined as the solution of the Eq. (2.7), with the above expressions for $B(\hat{x})$ and $S(\hat{x})$. The solution derived with the aid of Maple leads to the following approximations for the means and variances of S and I in quasi-stationarity:

$$\begin{aligned} \mu_S &\approx \frac{1}{\delta_1(R_0 - 1) + 1}N, \\ \sigma_S^2 &\approx \frac{(\delta_1 R_0 + 1)R_0[\delta_1(R_0 - 1) + 1] + \delta_1(\delta_1 - 1)}{[\delta_1(R_0 - 1) + 1]^3 R_0}N, \\ \mu_I &\approx \frac{R_0 - 1}{\delta_1(R_0 - 1) + 1}N, \\ \sigma_I^2 &\approx \frac{[\delta_1(R_0 - 1)^3 + R_0^2][\delta_1(R_0 - 1) + 1] + \delta_1^2(R_0 - 1)^2 R_0}{[\delta_1(R_0 - 1) + 1]^3 R_0}N. \end{aligned}$$

The results of Section 2.4 are recovered when $\delta_1 = 1$.

2.6. Qualitative results

The deterministic version of the model contains the powerful qualitative result that the infection will go extinct if $R_0 < 1$, while it will establish itself at a positive endemic level if $R_0 > 1$. There is no direct correspondence to this result in the stochastic version of the model, since the infection will ultimately go extinct for all positive values of R_0 . Note that the deterministic model qualitative result takes the form that the one-dimensional parameter space where R_0 takes its values is partitioned into two subsets with qualitatively different behaviours in state space. A similar partitioning of the parameter space can be established for the stochastic version of the model. Here it is important to observe that the parameter space for the stochastic version of the model is

two-dimensional, containing both the basic reproduction ratio R_0 and the population size measure N . It is shown for the SIS model without demography studied in [8] that one can identify three regions in this two-dimensional parameter space where the two important indicators of the infection, namely the quasi-stationary distribution and the time to extinction, show qualitatively different behaviours. Thus the stochastic version result contains an extra region that is absent from the result that holds for the deterministic version of the model. This additional region is a transition region where R_0 is close to one. The boundaries of the transition region depend on N . The transition region becomes narrower as N increases, and it becomes empty as $N \rightarrow \infty$. Thus, the description of the qualitative results of the stochastic version of the model can be seen as a generalization of the well-known threshold result for the deterministic version of the model. The deterministic model is not valid in the transition region.

The different qualitative results for the stochastic version of the model in the three parameter regions become evident by deriving asymptotic approximations of the quasi-stationary distribution and of the expected time to extinction from the quasi-stationary distribution as $N \rightarrow \infty$. The analysis in the transition region requires a reparametrization that makes R_0 approach one as $N \rightarrow \infty$. For the SIS model without demography this reparametrization takes the form $R_0 = 1 + \rho/\sqrt{N}$. Asymptotic analysis in the transition region is made by requiring the new parameter ρ to be fixed as $N \rightarrow \infty$. In the remaining two regions, R_0 is kept fixed as $N \rightarrow \infty$. I refer to the region where R_0 is fixed at a value larger than one (smaller than one) by saying that R_0 is distinctly larger than one (distinctly smaller than one).

Useful asymptotic approximations of the quasi-stationary distribution and of the time to extinction in the transition region can be derived with this approach for the univariate SIS model, as shown in [9], but so far no progress has been made on the corresponding problem for any of the multivariate models considered in this paper. The reparametrization for the transition region is, however, still useful for another purpose, namely the one of identifying the boundary between the transition region and the region where R_0 is distinctly larger than one. This boundary depends on N . The width of the transition region is a measure of the width of the parameter region where the deterministic model is not valid. The deterministic approximation is acceptable, as far as it goes, in the region where R_0 is distinctly larger than one, but not in the transition region.

The parameter ρ used for the reparametrization of R_0 in the transition region can be identified with the ratio of the mean to the standard deviation in the marginal distribution of the number of infected individuals in quasi-stationarity. Furthermore, the boundary between the transition region and the region where R_0 is distinctly larger than one can be chosen where $\rho = 3$ or 4. The choice of ρ -value here is dictated by the condition that the probability for a normal random variable to deviate from its mean by less than three or four standard deviations is close to one.

To derive the reparametrization of R_0 in the transition region I put provisionally ρ equal to the approximation of the ratio of μ_I to σ_I . After that, I derive the asymptotic approximation of this provisional value as $R_0 \rightarrow 1$. The resulting value of this ratio is defined as the new parameter ρ . The result can be written

$$\rho = (R_0 - 1)\sqrt{N}.$$

This is formally the same as the expression for ρ for the SIS model without demography.

The transition region is narrow for the SI model. Indeed, it suffices for N to satisfy the inequality $N > \rho^2/(R_0 - 1)^2$ to conclude (with $R_0 > 1$) that we are outside the transition region.

With $\rho = 3$ we find that this inequality is satisfied by any N larger than nine if R_0 is larger than two.

3. A stochastic SIS model with demography

The SIS model without demography is a univariate model whose deterministic approximation is given by the well-known logistic equation. The stochastic version of this model is a finite-state birth-death process with an absorbing state at the origin. The goal of the mathematical analysis of this model is to find approximations of the quasi-stationary distribution and of the time to extinction. Methods have been developed for such an analysis; results are given by [8] and [9]. The SIS model without demography is based on the implicit assumption that individual hosts live forever. This assumption is quite unsatisfactory, especially since the results show that the time to extinction can in some cases become very long. It is therefore natural to study a SIS model with demography, allowing birth and death of individual hosts.

In addition to the four parameters N , β , μ , and μ_1 introduced for the SI model of the previous section, I use γ_1 to denote the recovery rate per infected individual. The SIS model takes the form of a bivariate Markov population process whose transition rates are given in Table 2.

The Kolmogorov forward equations for the state probabilities $p_{si}(t) = P\{S(t) = s, I(t) = i\}$ can be written

$$p'_{si}(t) = \lambda_1(s - 1, i)p_{s-1,i}(t) + \bar{\mu}_1(s + 1, i)p_{s+1,i}(t) + \mu_2(s, i + 1)p_{s,i+1}(t) + v_1(s - 1, i + 1)p_{s-1,i+1}(t) + v_2(s + 1, i - 1)p_{s+1,i-1}(t) - \kappa(s, i)p_{si}(t),$$

where

$$\kappa(s, i) = \lambda_1(s, i) + \bar{\mu}_1(s, i) + \mu_2(s, i) + v_1(s, i) + v_2(s, i).$$

The deterministic version of the model leads to the following system of differential equations:

$$S' = \mu N - \beta \frac{SI}{S + I} - \mu S + \gamma_1 I,$$

$$I' = \beta \frac{SI}{S + I} - (\gamma_1 + \mu + \mu_1) I.$$

I reparametrize the model by introducing the following new parameters:

$$R_0 = \frac{\beta}{\gamma_1 + \mu + \mu_1}, \tag{3.1}$$

Table 2
Transition rates for the SIS model of Section 3

Event	Transition	Transition rate
Immigration of susceptible	$(s, i) \rightarrow (s + 1, i)$	$\lambda_1(s, i) = \mu N$
Death of susceptible	$(s, i) \rightarrow (s - 1, i)$	$\bar{\mu}_1(s, i) = \mu s$
Death of infected	$(s, i) \rightarrow (s, i - 1)$	$\mu_2(s, i) = (\mu + \mu_1) i$
Recovery of of infected	$(s, i) \rightarrow (s + 1, i - 1)$	$v_1(s, i) = \gamma_1 i$
Infection of susceptible	$(s, i) \rightarrow (s - 1, i + 1)$	$v_2(s, i) = \beta si / (s + i)$

$$\alpha_1 = \frac{\gamma_1 + \mu + \mu_1}{\mu}, \quad (3.2)$$

$$\delta_1 = \frac{\mu + \mu_1}{\mu}. \quad (3.3)$$

The deterministic version of the model has its threshold at $R_0 = 1$. Among the parameters, R_0 is strictly positive and α_1 is strictly larger than one, while δ_1 is larger than or equal to one. In most cases of interest the average time infected, $1/(\gamma_1 + \mu + \mu_1)$, is much shorter than the average life length $1/(\mu + \mu_1)$ of infected individuals, which in turn is less than or equal to the average life length $1/\mu$ of susceptible individuals. This implies that α_1 is much larger than one. The analysis in this and later sections will concentrate on this case.

The differential equations for the deterministic version of the model with the scaled state variables $x_1 = S/N$ and $x_2 = I/N$ has two critical points. One of them is at $(1, 0)$, which corresponds to the absence of infection. The other critical point corresponds to an endemic infection level. Its coordinates are $\hat{x} = (\hat{x}_1, \hat{x}_2)$, where

$$\hat{x}_1 = \frac{1}{\delta_1(R_0 - 1) + 1},$$

$$\hat{x}_2 = \frac{R_0 - 1}{\delta_1(R_0 - 1) + 1}.$$

By using arguments similar to those in Section 2.3 one finds that the probability of having zero infected individuals at time t obeys the differential equation

$$p'_0(t) = \mu\alpha_1 p_1(t). \quad (3.4)$$

It follows that the time to extinction τ_Q from quasi-stationarity has an exponential distribution with the expectation

$$E(\tau_Q) = \frac{1}{\mu\alpha_1 q_1}.$$

The details in the derivation of the diffusion approximation are given in the Maple program that can be found on the web-address cited above. The first step is a derivation of the local drift matrix $B(\hat{x})$ and the local covariance matrix $S(\hat{x})$. The covariance matrix Σ is then determined as the solution of the Eq. (2.7), with these expressions for $B(\hat{x})$ and $S(\hat{x})$. An explicit expression can be found for Σ , but it turns out to be rather complicated. Each entry in Σ is a ratio of two polynomials in the three parameters R_0 , α_1 , and δ_1 . I give therefore only the asymptotic approximations of the elements of the matrix Σ that result when $\alpha_1 \rightarrow \infty$. The results are therefore approximations in the interesting case when the average life length is much longer than the average time of being infected.

The moments for the marginal distributions of S and I in quasi-stationarity can then be written as follows:

$$\mu_S \approx \frac{1}{\delta_1(R_0 - 1) + 1} N, \quad \sigma_S^2 \approx \frac{R_0 + 1}{R_0[\delta_1(R_0 - 1) + 1]} N,$$

$$\mu_I \approx \frac{R_0 - 1}{\delta_1(R_0 - 1) + 1} N, \quad \sigma_I^2 \approx \frac{R_0^2 - R_0 + 1}{R_0[\delta_1(R_0 - 1) + 1]} N.$$

An increase of the additional death rate for infected individuals μ_1 causes δ_1 to increase. This causes all the four moments above to decrease. It is easy to verify that the coefficients of variation of both S and I are increasing functions of δ_1 .

The influence of the demographic forces on the moments of S and I in quasi-stationarity can be understood by a comparison between the results without differential mortality ($\delta_1 = 1$) and the results of the analysis of the SIS model without demography.

From the above results we get the following expressions for the means and the variances of the marginal distributions in quasi-stationarity for the number of susceptible and infected individuals when $\delta_1 = 1$:

$$\begin{aligned} \mu_S &\approx \frac{1}{R_0}N, & \sigma_S^2 &\approx \frac{R_0 + 1}{R_0^2}N, \\ \mu_I &\approx \frac{R_0 - 1}{R_0}N, & \sigma_I^2 &\approx \frac{R_0^2 - R_0 + 1}{R_0^2}N. \end{aligned}$$

The SIS model without demography is considerably simpler than the one considered here, since the total population size $S + I$ is constant. This has two consequences, namely that the model is univariate, and that the denominator in the expression for the infection rate is constant and equal to N . From [8] we find for $R_0 > 1$ the following expressions for the means and variances of S and I in quasi-stationarity:

$$\begin{aligned} \mu_S &\approx \frac{1}{R_0}N, & \sigma_S^2 &\approx \frac{1}{R_0}N, \\ \mu_I &\approx \frac{R_0 - 1}{R_0}N, & \sigma_I^2 &\approx \frac{1}{R_0}N. \end{aligned}$$

A comparison between these two sets of results shows that the approximations of the means of S and I in quasi-stationarity are the same with and without demography, and also that the variances are larger with demography than when demography is ignored. This difference is related to the fact that the sum of the random variables S and I is equal to the total population size in both cases. This sum is constant and equal to N in the model without demography, while it is a random variable in the model that accounts for demography. This random variable has (at least asymptotically for large times) a Poisson distribution with both mean and variance equal to N . Thus the variance of $S + I$ equals 0 in the model without demography, while it equals N in the model of the present subsection.

By using the same arguments as in Section 2.6 we find that the reparametrization of R_0 in the transition region can be written

$$\rho = (R_0 - 1)\sqrt{N}.$$

This relation has the same form as we found in the previous section for the SI model. Thus we conclude that the transition region, where the deterministic model is not valid, is narrow.

4. A stochastic SIR model with demography

I proceed to formulate a stochastic SIR model with demography. The model is a Markov population process with three state variables, S , I , and R , standing for the number of susceptible,

infected, and recovered individuals, respectively. The recovered individuals are assumed permanently immune to additional infections. I use the same five basic parameters as for the SIS model of the previous section, namely the expected population size N if all individuals are susceptible, the death rate per susceptible or removed individual μ , the additional death rate μ_1 per infected individual, the contact rate β , and the recovery rate per infected individual γ_1 . The hypotheses of the model are given in terms of the transition rates in Table 3.

The state probabilities are defined by $p_{sir}(t) = P\{S(t) = s, I(t) = i, R(t) = r\}$. The Kolmogorov forward equations for these probabilities can be written

$$p'_{sir}(t) = \lambda_1(s - 1, i, r)p_{s-1,i,r}(t) + \bar{\mu}_1(s + 1, i, r)p_{s+1,i,r}(t) + \mu_2(s, i + 1, r)p_{s,i+1,r}(t) + \mu_3(s, i, r + 1)p_{s,i,r+1}(t) + v_2(s + 1, i - 1, r)p_{s+1,i-1,r}(t) + v_3(s, i + 1, r - 1)p_{s,i+1,r-1}(t) - \kappa(s, i, r)p_{sir}(t),$$

where

$$\kappa(s, i, r) = \lambda_1(s, i, r) + \bar{\mu}_1(s, i, r) + \mu_2(s, i, r) + \mu_3(s, i, r) + v_2(s, i, r) + v_3(s, i, r).$$

The deterministic version of the model leads to the following system of differential equations:

$$\begin{aligned} S' &= \mu N - \beta \frac{SI}{S + I + R} - \mu S, \\ I' &= \beta \frac{SI}{S + I + R} - (\gamma_1 + \mu + \mu_1)I, \\ R' &= \gamma_1 I - \mu R. \end{aligned}$$

I reparametrize the model in exactly the same way as for the SIS model by introducing the derived parameters R_0 , α_1 , and δ_1 , defined in terms of the basic parameters in (3.1)–(3.3).

The system of differential equations for the deterministic version of the model with the scaled state variables $x_1 = S/N$, $x_2 = I/N$, $x_3 = R/N$ has two critical points. One of them is $(1, 0, 0)$ and corresponds to the absence of infection. The other one corresponds to an endemic infection level. Its coordinates are $\hat{x} = (\hat{x}_1, \hat{x}_2, \hat{x}_3)$, given by

$$\begin{aligned} \hat{x}_1 &= \frac{\alpha_1 - (\delta_1 - 1)}{R_0\alpha_1 - (\delta_1 - 1)}, \\ \hat{x}_2 &= \frac{R_0 - 1}{R_0\alpha_1 - (\delta_1 - 1)}, \\ \hat{x}_3 &= \frac{(R_0 - 1)(\alpha_1 - \delta_1)}{R_0\alpha_1 - (\delta_1 - 1)}. \end{aligned}$$

Table 3
Transition rates for the SIR model with demography

Event	Transition	Transition rate
Immigration of susceptible	$(s, i, r) \rightarrow (s + 1, i, r)$	$\lambda_1(s, i, r) = \mu N$
Death of susceptible	$(s, i, r) \rightarrow (s - 1, i, r)$	$\bar{\mu}_1(s, i, r) = \mu s$
Death of infected	$(s, i, r) \rightarrow (s, i - 1, r)$	$\mu_2(s, i, r) = (\mu + \mu_1)i$
Death of removed	$(s, i, r) \rightarrow (s, i, r - 1)$	$\mu_3(s, i, r) = \mu r$
Infection of susceptible	$(s, i, r) \rightarrow (s - 1, i + 1, r)$	$v_2(s, i, r) = \beta si / (s + i + r)$
Recovery of infected	$(s, i, r) \rightarrow (s, i - 1, r + 1)$	$v_3(s, i, r) = \gamma_1 i$

The probability of having zero infected individuals at time t obeys the differential equation

$$p'_{0.}(t) = \mu\alpha_1 p_{1.}(t).$$

It follows that the time to extinction τ_Q from quasi-stationarity has an exponential distribution with the expectation

$$E(\tau_Q) = \frac{1}{\mu\alpha_1 q_{1.}}.$$

As in Section 3, I find that the covariance matrix Σ that solves (2.7) with the expressions for $B(\hat{x})$ and $S(\hat{x})$ for this model is too complicated to be useful. I therefore limit myself to a consideration of the approximations of the moments of S , I , and R in quasi-stationarity that apply when the ratio of average life length to average time infected is large. For the expectations of S , I , and R I derive the first terms in the asymptotic approximations of $N\hat{x}_1$, $N\hat{x}_2$, and $N\hat{x}_3$, respectively, as $\alpha_1 \rightarrow \infty$, and for the variances I evaluate the first terms in asymptotic approximations of the diagonal elements of Σ , also as $\alpha_1 \rightarrow \infty$, followed by a multiplication by N . The results are

$$\begin{aligned} \mu_S &\approx \frac{1}{R_0}N, & \sigma_S^2 &\approx \frac{\alpha_1}{R_0^2}N, \\ \mu_I &\approx \frac{R_0 - 1}{R_0\alpha_1}N, & \sigma_I^2 &\approx \frac{R_0 - 1}{R_0^2}N, \\ \mu_R &\approx \frac{R_0 - 1}{R_0}N, & \sigma_R^2 &\approx \frac{\alpha_1}{R_0^2}N. \end{aligned}$$

Note that these results are not influenced by the parameter δ_1 , which measures the additional mortality suffered by infected individuals. This is in sharp contrast to the situation for the SIS model, where a clear influence of the differential mortality could be observed.

These results for the moments of S and I in quasi-stationarity are compared with the corresponding results in the similar but somewhat simpler SIR model with demography dealt with in [2]. This model differs from the present one in three ways. First, it does not allow for differential mortality. The finding above prepares us for the result that this will not influence the means and variances of S and I in quasi-stationarity. Second, this simpler model is bivariate instead of the trivariate model of the present section, since the hypotheses of the model lead to the consequence that the state variable R does not influence S or I . Thirdly, and most importantly, this simpler model uses the conventional infection rate $\beta SI/N$ instead of $\beta SI/(S + I + R)$. This represents an unjustified simplification, since stochastic variability in the denominator of the expression for the infection rate is not accounted for. It is therefore rather surprising to find that the two models show agreement between the first-order terms in the asymptotic approximations as $\alpha_1 \rightarrow \infty$ of the means and variances of S and I in quasi-stationarity. One consequence of this is that the approximation of the expected time to extinction from quasi-stationarity that is given in [2] applies also to the SIR model with differential mortality and proper infection rate treated in this section.

The reparametrization of R_0 in the transition region can for this model be written

$$\rho = \frac{\sqrt{R_0 - 1}\sqrt{N}}{\alpha_1}.$$

The transition region is considerably wider for the SIR model than for the SI and SIS models treated in the two previous sections. The condition $N > \alpha_1^2 \rho^2 / (R_0 - 1)$ must be satisfied in order to wind up in the region where R_0 is distinctly larger than one, and where the deterministic model is an acceptable approximation of the stochastic one. With $\alpha_1 = 3000$ and $\rho = 3$ we find that N must be larger than 81 million if $R_0 = 2$, and larger than 4.2 million if $R_0 = 20$.

5. A stochastic SIRS model with demography

This section is devoted to a study of an SIRS model with demography. The model differs from the SIR model of the previous section only by allowing a loss of immunity. In addition to the five basic parameters used in Sections 3 and 4, I introduce γ_2 to denote the rate of loss of immunity. The hypotheses of the model are spelled out in terms of the transition rates in Table 4.

The state probabilities at time t are defined by $p_{sir}(t) = P\{S(t) = s, I(t) = i, R(t) = r\}$. The Kolmogorov forward equations for these probabilities can be written

$$\begin{aligned}
 p'_{sir}(t) = & \lambda_1(s - 1, i, r)p_{s-1,i,r}(t) + \bar{\mu}_1(s + 1, i, r)p_{s+1,i,r}(t) \\
 & + \mu_2(s, i + 1, r)p_{s,i+1,r}(t) + \mu_3(s, i, r + 1)p_{s,i,r+1}(t) \\
 & + v_1(s - 1, i, r + 1)p_{s-1,i,r+1}(t) + v_2(s + 1, i - 1, r)p_{s+1,i-1,r}(t) \\
 & + v_3(s, i + 1, r - 1)p_{s,i+1,r-1}(t) - \kappa(s, i, r)p_{sir}(t),
 \end{aligned}$$

where

$$\kappa(s, i, r) = \lambda_1(s, i, r) + \bar{\mu}_1(s, i, r) + \mu_2(s, i, r) + \mu_3(s, i, r) + v_1(s, i, r) + v_2(s, i, r) + v_3(s, i, r).$$

The deterministic version of the model leads to the following system of differential equations:

$$\begin{aligned}
 S' &= \mu N - \beta \frac{SI}{S + I + R} - \mu S + \gamma_2 R, \\
 I' &= \beta \frac{SI}{S + I + R} - (\gamma_1 + \mu + \mu_1) I, \\
 R' &= \gamma_1 I - (\mu + \gamma_2) R.
 \end{aligned}$$

I reparametrize by introducing the parameters R_0 , α_1 and δ_1 as in (3.1)–(3.3), and in addition the parameter α_2 defined by

$$\alpha_2 = \frac{\mu + \gamma_2}{\mu}.$$

Table 4
Transition rates for the SIRS model with demography

Event	Transition	Transition rate
Immigration of susceptible	$(s, i, r) \rightarrow (s + 1, i, r)$	$\lambda_1(s, i, r) = \mu N$
Death of susceptible	$(s, i, r) \rightarrow (s - 1, i, r)$	$\bar{\mu}_1(s, i, r) = \mu s$
Death of infected	$(s, i, r) \rightarrow (s, i - 1, r)$	$\mu_2(s, i, r) = (\mu + \mu_1) i$
Death of removed	$(s, i, r) \rightarrow (s, i, r - 1)$	$\mu_3(s, i, r) = \mu r$
Loss of immunity	$(s, i, r) \rightarrow (s + 1, i, r - 1)$	$v_1(s, i, r) = \gamma_2 r$
Infection of susceptible	$(s, i, r) \rightarrow (s - 1, i + 1, r)$	$v_2(s, i, r) = \beta si / (s + i + r)$
Recovery of infected	$(s, i, r) \rightarrow (s, i - 1, r + 1)$	$v_3(s, i, r) = \gamma_1 i$

As in the previous section we introduce the scaled state variables $x_1 = S/N$, $x_2 = I/N$, and $x_3 = R/N$. The system of differential equations for the scaled state variables has two critical points. One of them is $(1, 0, 0)$, corresponding to the absence of infection. The second one corresponds to an endemic infection level. It is equal to $\hat{x} = (\hat{x}_1, \hat{x}_2, \hat{x}_3)$, where

$$\begin{aligned} \hat{x}_1 &= \frac{\alpha_1 + \alpha_2 - \delta_1}{R_0\alpha_1 + ((R_0 - 1)\delta_1 + 1)\alpha_2 - \delta_1R_0}, \\ \hat{x}_2 &= \frac{(R_0 - 1)\alpha_2}{R_0\alpha_1 + ((R_0 - 1)\delta_1 + 1)\alpha_2 - \delta_1R_0}, \\ \hat{x}_3 &= \frac{(R_0 - 1)(\alpha_1 - \delta_1)}{R_0\alpha_1 + ((R_0 - 1)\delta_1 + 1)\alpha_2 - \delta_1R_0}. \end{aligned}$$

The probability of having zero infected individuals at time t obeys the same differential equation as in the SIR model:

$$p'_{0.}(t) = \mu\alpha_1 p_{1.}(t).$$

It follows that the time to extinction τ_Q from quasi-stationarity has an exponential distribution with the expectation

$$E(\tau_Q) = \frac{1}{\mu\alpha_1 q_{1.}}.$$

As in the case for the SIR model one finds that the covariance matrix Σ that solves (2.7) with the expressions for $B(\hat{x})$ and $S(\hat{x})$ that hold for this model is too complicated to be useful. I therefore derive the asymptotic approximations of the moments of S , I , and R in quasi-stationarity when the ratio of average life length to average time infected is large, i.e. when $\alpha_1 \rightarrow \infty$. Here, I get different results when α_2 is kept finite as compared to when α_2 grows with α_1 . Note that the SIR model is a special case corresponding to $\alpha_2 = 1$, and that the SIS model is recovered as $\alpha_2 \rightarrow \infty$.

The following results are found for the moments of the three state variables as $\alpha_1 \rightarrow \infty$ and $\alpha_2 = O(1)$:

$$\begin{aligned} \mu_S &\approx \frac{1}{R_0}N, & \sigma_S^2 &\approx \frac{\alpha_1}{\alpha_2 R_0^2}N, \\ \mu_I &\approx \frac{\alpha_2(R_0 - 1)}{\alpha_1 R_0}N, & \sigma_I^2 &\approx \frac{R_0 - 1}{R_0^2}N, \\ \mu_R &\approx \frac{R_0 - 1}{R_0}N, & \sigma_R^2 &\approx \frac{\alpha_1}{\alpha_2 R_0^2}N. \end{aligned}$$

The results for the SIR model in Section 4 are recovered for $\alpha_2 = 1$. The reparametrization of R_0 in the transition region can be written

$$\rho = \frac{\alpha_2}{\alpha_1} \sqrt{R_0 - 1} \sqrt{N}.$$

The approximating expressions for the moments of the three state variables are more complicated when one considers the situation where α_1 and α_2 both grow toward infinity. To study this case I put $\alpha_2 = \theta\alpha_1$ and let $\alpha_1 \rightarrow \infty$ with $\theta = O(1)$. The following approximate expressions are then found for the means of the three state variables in quasi-stationarity:

$$\mu_S \approx \frac{\theta + 1}{\delta_1(R_0 - 1)\theta + \theta + R_0} N,$$

$$\mu_I \approx \frac{(R_0 - 1)\theta}{\delta_1(R_0 - 1)\theta + \theta + R_0} N,$$

$$\mu_R \approx \frac{R_0 - 1}{\delta_1(R_0 - 1)\theta + \theta + R_0} N,$$

The variances of the same state variables in quasi-stationarity are under the same condition approximated as follows:

$$\sigma_S^2 \approx \frac{(\theta + 1)[(R_0 + 1)\theta^2 + R_0(R_0 + 2)\theta + R_0]}{R_0\theta(\theta + R_0)[\delta_1(R_0 - 1)\theta + \theta + R_0]} N,$$

$$\sigma_I^2 \approx \frac{(R_0^2 - R_0 + 1)\theta^3 + (R_0^3 + R_0)\theta^2 + (R_0^3 + R_0^2 - R_0)\theta + R_0^2 - R_0}{R_0(\theta + R_0)(\theta + 1)[\delta_1(R_0 - 1)\theta + \theta + R_0]} N,$$

$$\sigma_R^2 \approx \frac{R_0(R_0 - 1)\theta^3 + (R_0^3 - R_0 + 1)\theta^2 + R_0(R_0^2 - 2R_0 + 3)\theta + R_0}{R_0\theta(\theta + R_0)(\theta + 1)[\delta_1(R_0 - 1)\theta + \theta + R_0]} N.$$

It is straightforward to verify that the results for the SIS model are recovered as $\theta \rightarrow \infty$.

The reparametrization of R_0 in the transition region can in this case be written

$$\rho = \sqrt{\frac{\theta}{\theta + 1}}(R_0 - 1)\sqrt{N}.$$

The width of the transition region for the SIRS model depends strongly on the value of the parameter α_2 . One can expect that the transition region is narrow if α_2 is large, and that it is wide if α_2 is small. The reason for these expectations is that the SIS model, with its narrow transition region, is recovered as a special case as $\alpha_2 \rightarrow \infty$, and that the SIR model, with its wide transition region, is recovered for $\alpha_2 = 1$. These conclusions are valid in the case where the average life length is much longer than the average time of being infected.

6. Concluding comments

The fully stochastic models discussed here are too complicated to allow explicit solutions. The results that have been given are all based on approximations. One would need to make comparisons with numerical evaluations in order to be able to judge the adequacy of these approximations. In the absence of such numerical work, I can use my experience with detailed studies of SIS and SIR models to make two conjectures. Both of them apply when the condition for being above the transition region is satisfied. The first conjecture is that the normal approximation of the body of the quasi-stationary distribution provides a reasonably good approximation. The second conjecture concerns the adequacy of the approximation of the expected time to extinction from quasi-stationarity. The estimate of this quantity is based on an extension to the tail of the normal distribution that is valid only in the body. The second conjecture is that this approximation is crude, but that it gives correct order of magnitude.

The conditions for validity of the approximations of the quasi-stationary distributions and of the expected time to extinction from quasi-stationarity are also conditions for validity of the corresponding deterministic models. It is noteworthy that there are some models (SIR and SIRS) where huge population sizes are necessary for validity of the deterministic versions of the models. This means that deterministic models for such situations are based on unacceptable mathematical approximations for a large range of parameter values. This fact raises questions about the usefulness of deterministic models in this area, and shows that demographic stochasticity cannot be ignored.

The importance of demographic stochasticity for understanding the phenomena of recurrence and extinction in childhood infections is emphasized in [10]. This reference analyses the fully stochastic counterpart of chaotic deterministic models for recurrent epidemics, and shows that the mechanism that drives chaos in these models is an unjustified mathematical approximation introduced in going from the stochastic to the deterministic formulation.

Deterministic models of endemic infections have advanced considerably beyond those considered here, to include such phenomena as age structure, heterogeneity, and spatial structure. Examples are given in Hethcote's review paper [4]. It is clearly desirable to extend the analysis of stochastic models in a similar way. For work in this direction the reader is referred to Andersson and Britton [11], where a stochastic SEIR model with demography is treated. Both the life length and the waiting times in the E and I states are generalized from exponential distributions to gamma distributions. Emphasis is placed on analyzing approximations of the quasi-stationary distribution and of the expected time to extinction from quasi-stationarity.

A similar goal is pursued by Allen and Burgin [12]. They establish stochastic SIS and SIR models in discrete time as approximations of the corresponding continuous-time models.

References

- [1] I. Nåsell, The threshold concept in stochastic epidemic and endemic models, in: D. Mollison (Ed.), *Epidemic Models: Their Structure and Relation to Data*, Cambridge University, Cambridge, 1995.
- [2] I. Nåsell, On the time to extinction in recurrent epidemics, *J. Roy. Statist. Soc. B*, Part 2 61 (1999) 309.
- [3] N.T.J. Bailey, *The Mathematical Theory of Infectious Diseases and Its Applications*, Griffin, London, 1975.
- [4] H.W. Hethcote, The mathematics of infectious diseases, *SIAM Rev.* 42 (4) (2001) 599.
- [5] J. Mena-Lorca, H.W. Hethcote, Dynamic models of infectious diseases as regulators of population sizes, *J. Math. Biol.* 30 (1992) 693.
- [6] H.W. Hethcote, Qualitative analysis of communicable disease models, *Math. Biosci.* 28 (1976) 335.
- [7] J.A. Jacquez, C.A. Simon, The stochastic SI model with recruitment and deaths. I. Comparison with the closed SIS model, *Math. Biosci.* 117 (1993) 77.
- [8] I. Nåsell, The quasi-stationary distribution of the closed endemic SIS model, *Adv. Appl. Prob.* 28 (1996) 895.
- [9] I. Nåsell, On the quasi-stationary distribution of the stochastic logistic epidemic, *Math. Biosci.* 156 (1999) 21.
- [10] I. Nåsell, Measles outbreaks are not chaotic, *IMA Volumes in Mathematics and its Applications, Mathematical Approaches for Emerging and Reemerging Infectious Diseases, Part II: Models*, vol. 126, Methods and Theory, 2002, Springer, in press.
- [11] H. Andersson, T. Britton, Stochastic epidemics in dynamic populations: quasi-stationarity and extinction, *J. Math. Biol.* 41 (6) (2000) 559.
- [12] L.J.S. Allen, A.M. Burgin, Comparison of deterministic and stochastic SIS and SIR models in discrete time, *Math. Biosci.* 163 (2000) 11.