University of West Bohemia

Faculty of Applied Sciences

Department of Mathematics

BACHELOR'S THESIS

A comparison of different stochastic epidemiological models of SIS type with regard to duration of an epidemic

Pavel Říšský

Plzeň 2015

Acknowledgement

I wish to express my sincere gratitude to Ing. Jan Pospíšil, Ph.D for the continuous support of my study, for his knowledge and patience. His guidance helped me in all the time of research and writing of this thesis. I also thank my family for the unceasing encouragement and attention.

Statement

I hereby declare that I wrote this bachelor's thesis myself without the use of any means or sources other then indicated.

In Plzeň, 25th May 2015 Pavel Říšský

"Not everything that can be counted counts,
and not everything that counts can be counted."

William Bruce Cameron

Abstract

This thesis deals with the epidemiological SIS model. The main aim of the paper is to describe stochastic SIS model by Markov chain in both discrete-time (DTMC) and continuous-time (CTMC) with regard to duration of an epidemic. Simulations and comparison of outcomes are performed in MATLAB. Analytic solution of the deterministic model and simulated mean value are compared in both cases. Following properties are explored: standard deviation of process, mean persistence time, quasi-stationary distribution and its approximations for both supercritical and sub-critical reproduction number. Both models evince identical limit behaviour. The only difference is the speed of dynamics. While the dynamics of the continuous-time Markov chain is determined solely by corresponding transition rates, the frequency of transitions of the discrete-time Markov chain depends on chosen parameter Δt.

Key words
SIS model; Markov chains; quasi-stationary distribution; mean persistence time of epidemic

Abstrakt

Tato práce se zabývá epidemiologickým SIS modelem. Jejím hlavním cílem je popsat stochastický SIS model Markovskými řetězci v diskrétním (DTMC) a spojitém (CTMC) čase s ohledem na trvání epidemie. Simulace včetně porovnání výsledků jsou provedeny pomocí softwaru MATLAB. Pro obě možnosti je analytické řešení deterministického modelu porovnáno se střední hodnotou simulovaných procesů. Dále jsou vykresleny: směrodatná odchylka procesu, střední čas trvání epidemie, kvazistacionární rozdělení a jeho aproximace pro superkritické a subkritické reprodukční číslo. Oba modely mají shodné limitní chování. Liší se pouze rychlostí dynamiky. Zatímco Markovský řetězec se spojitým časem se řídí jen velikostí intenzit přechodu, frekvence přechodů Markovského řetězce s diskrétním časem je určena voleným parametrem Δt.

Klíčová slova
SIS model; Markovské řetězce; kvazistacionární rozdělení; střední čas trvání epidemie

Glossary of notation

$\mathbb{P}$	Probability
$\mathbf{P}$	Transition matrix
$\mathbf{Q}$	Intensity matrix
$\mathbb{N}_0$	Set of Natural Numbers including zero
$\mathbb{R}$	Set of Real Numbers
$\mathbb{E}$	Expected value
$\mathbb{V}$	Variance
SIS	Susceptible-infectious-susceptible
DTMC	Discrete-time Markov chain
CTMC	Continuous-time Markov chain
T_k	Extinction time for a population with initial size k
τ_k	$\mathbb{E}[T_k]$
$i \leftrightarrow j$	States i and j are mutually accessible

Contents

Chapter 1

Introduction

The simplest way to describe dynamics of a contact-transmitted disease which does not lead to immunity against reinfection is the susceptible-infectious-susceptible model. Since the population under study is divided into so-called compartments, this epidemiological approach is know as compartmental modelling (see Brauer, 2008). Furthermore, assumptions about the nature and time rate of transfer from one compartment to another are made. The name indicates that the passage of individuals is from the susceptible class S to the infectious class I and then back to the susceptible class S.

First one to introduce the deterministic SIS model was Ross (1910). This model deals with the transition of infection in a closed (fixed size) population, i.e. there are no births, deaths, immigration or emigration during the study period.

If the population is large, epidemic dynamics may be approximated well by the deterministic model. But when dealing with smaller population (e.g. school or hospital) it seems reasonable to assume randomness in the final number of infectious. Also, even if the community is large but the outbreak is initiated by only one (or a few) initial infectious it should be possible that, by chance, the epidemic never takes off as was shown by Britton (2010). These situations where stochastic effects cannot be neglected led to the development of stochastic models for discrete populations with either continuous or discrete time (see Chalub and Souza, 2014).

First ones to exercise this approach were Kermack and McKendrick (1927). In 1953, Bailey summed up the motivation to the stochastic approach and described general epidemiological stochastic model. Three years later, Kendall (1956) picked up the threads of Kermack and McKendrick and described the behaviour of stochastic model depending on its parameters. In the 1960's Darroch and Seneta introduced the concept of quasi-stationary distribution.

Stochastic SIS model was first discussed by Weiss and Dishon (1971) and Mollison (1995). (Jacquez and Simon, 1993) bring the susceptible-infectious model with mor-

tality and makes a comparison with the closed-population one. Statistical aspects of estimating reproduction number and its the historical development were reviewed by Heesterbeek and Dietz (1996).

In 2008, Allen published a chapter in the Lecture Notes in Mathematics vol. 1945, where she compared various SIS models with both discrete and continuous state space as well as discrete and continuous time. In the paper of Keeling and Ross (2008), SIS model's Kolmogorov forward equation is explored and results are compared with traditional stochastic simulations. Yaesoubi and Cohen (2011) proposed models which can be effectively used by dynamic optimization methods to select optimal dynamic health policies. Anderson et al. (2014) used it as a model for chemical reactions. Kundan and Joy (2014) considers information epidemics and employs SIS model to find optimal control strategy. Finally, Bartholomew (2015) describes social systems and human behaviour using stochastic SIS model.

In this thesis, the stochastic SIS model is explored with regard to duration of an epidemic. The basic dynamics is represented by two differential equations (section 1.2) and the stochastic behaviour is described in both discrete (Chapter 2) and continuous (Chapter 3) time frame by Markov chain modelling. Probability distribution is investigated and quasi-stationary distribution (conditional on non-extinction) is presented (sections 2.3 and 3.3). Furthermore, comparison of several approximations were made (sections 2.3.2 and 3.3.2).

1.1 Preliminaries

Definition 1.1 (discrete-time Markov chain)**.**
We say that a sequence of integer valued random variables $(X_n)_{n\geqq 0}$ is a *discrete-time Markov chain (DTCM)* with initial distribution $\mathbf{p}(0)$, discrete state space $\mathbf{S} = \{i_0,\ i_1,\ \ldots,\ i_k\}$ and transition matrix $\mathbf{P}$ if $\forall n \in \mathbb{N}_0\ \forall s \in \mathbf{S}$:

$$\mathbb{P}(X_{n+1} = j_{n+1} | X_n = i_n, \ldots, X_0 = i_0) = \mathbb{P}\{X_{n+1} = j_{n+1} | X_n = i_n\} = p_{i,j}(n, n+1),$$

i.e., process satisfies the Markov's property (memorylessness). The process at time $t + \Delta t$ only depends on the state at the previous time step t. Without loss of generality, we can assume $\mathbf{S} = \{0,\ 1,\ \ldots,\ k\}$.

The elements of the transition matrix $\mathbf{P}$ are called transition probabilities. The transition probability $p_{i,j}$ is the conditional probability of going to state j in the next step given that the chain is in state i now. Initial distribution $\mathbf{p_0}$ describes the process at the beginning $\forall i \in \mathbf{S} : \mathbb{P}\{X_0 = i\} = p_i$. We consider reachable states only - i.e., those for which $p_{i,i}(n) > 0\ \forall n \in \mathbb{N}_0$, therefore $\exists n \in \mathbb{N}_0\ \forall i \in \mathbf{S} : \mathbb{P}\{X_n = i\} > 0$. Let us denote probability distribution at time $n \in \mathbb{N}_0$ as $\mathbf{p}(n) = (p_0(n),\ p_1(n),\ \ldots p_k(n))^T$, for which $\sum_{s=0}^{k} p_i = 1$.

Definition 1.2 (stationary distribution)**.**
A probability distribution $\boldsymbol{\pi} = (\pi_0,\ \pi_1,\ \ldots,\ \pi_k)^T$ is said to be a *stationary distribution* of Markov chain $(X_n)_{n\geqq 0}$ with a state space $\mathbf{S} = \{i_0,\ \ldots,\ i_k\}$ and transition matrix $\mathbf{P}$ if

$$\sum_{i=0}^{k} \pi_i p_{i,j} = \pi_j$$

or in matrix form

$$\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}.$$

If the initial distribution $\mathbf{p}(0)$ is equal to $\boldsymbol{\pi}$, then the probability distribution $\mathbf{p}(n)$ of the chain at time n remains equal to $\boldsymbol{\pi}$ for every n (see Häggström, 2002).

Definition 1.3 (continuous-time Markov chain)**.**
A stochastic process $(X_t)_{t\geqq 0}$ with discrete state space $\mathbf{S} = \{i_0,\ \ldots,\ i_k\}$ is called a *continuous-time Markov chain (CTMC)* if $\forall t \geq 0\ \forall s \geq 0, i, j \in \mathbf{S}$:

$$\mathbb{P}(X_t = j | X_s = i, X_{t_n} = i_n, \ldots, X_{t_1} = i_1) = \mathbb{P}(X_t = j | X_s = i) = p_{i,j}(t),$$

where $p_{i,j}(t)$ is the probability that the chain will skip to state j, given it is in state i at time t.

Definition 1.4 (intensity matrix)**.**
For any $i \in \mathbf{S}$ there is a limit

$$\lim_{h\to 0_+} \frac{1 - p_{i,i}(h)}{h} = q_i \leq \infty,$$

and $\forall i, j \in \mathbf{S}, i \neq j$ there are limits

$$\lim_{h\to 0_+} \frac{1 - p_{i,j}(h)}{h} = q_{i,j} \leq \infty.$$

These so-called transition rate coefficients $q_{i,j}$ form an *intensity matrix* (also known as generator matrix) $\mathbf{Q} = \left\{q_{i,j}, i, j \in \mathbf{S}\right\}$, where $q_{i,i} = -q_i$.

Definition 1.5 (Kolmogorov differential equations)**.**
Suppose $\forall i \in \mathbf{S} : q_i < \infty$. Then transition probabilities $p_{i,j}(t)$ are differentiable $\forall i, j \in \mathbf{S}$ and $t > 0$ and the following retrospective equation holds

$$\mathbf{P}'(t) = \mathbf{Q}\mathbf{P}(t).$$

If the convergence $\dfrac{1 - p_{i,j}(h)}{h} \to q_{i,j}$ is uniform at i, than we can add the following prospective equation (Prášková and Lachout, 2001)

$$\mathbf{P}'(t) = \mathbf{P}(t)\mathbf{Q}. \tag{1.1}$$

Definition 1.6 (classification of states)**.**
State $i \in \mathbf{S}$ is said to be:

- *absorbing* if $q_i = 0$,
- *stable* if $0 > q_i > \infty$,
- *unstable* if $q_i = \infty$.

Definition 1.7 (persistent and transient states)**.**
State $i \in \mathbf{S}$ is said to be *persistent* if the process leaving the state i will return to the state i in a finite time. That is

$$\mathbb{P}(\tau_i(1) < \infty) = 1,$$

where $\tau_i(1)$ is a time of the first return to the state i. State $i \in \mathbf{S}$ is said to be *transient* if there is a non-zero probability that the process leaving the state i will never return to the state i. That is

$$\mathbb{P}(\tau_i(1) = \infty) > 0.$$

Definition 1.8 (absorbing state)**.**
Set of states C is called *closed class of states* if all states are accessible from each other ($\forall i,\ j \in C;\ i \leftrightarrow j$) but there is no state outside the C accessible from C ($\forall\ i \in C$, $\forall\ j \notin C,\ p_{i,j} = 0$). If this class contains only one state, this state is called *absorbing*.

Definition 1.9 (reducibility)**.**
Process with state space $\mathbf{S} = \{i_0, ..., i_k\}$ is said to be *irreducible* if $i \leftrightarrow j\ \forall\ i, j \in S$. Otherwise we say that the chain is *reducible*.

Lemma 1.1 (probability of absorption)**.**
The probability of leaving the set of transient states T and moving to closed class C is called the probability of absorption *and it is equal to one (see Prášková and Lachout, 2001).*

1.2 Model formulation

SIS epidemic model divides population into two classes according to disease status. Individuals are either susceptible or infectious. These sets are denoted by the variables S and I, respectively.

Figure 1.1: Transitions between compartments

Every susceptible individual may get to contact with infectious one and become infectious. SIS model does not assume developing immunity nor death caused by disease. Therefore, as we can see in Figure 1.1, after recovery, infectious individual return to the susceptible class. Dynamics of the SIS model are described by the following differential equations

$$\begin{aligned} S'(t) &= -rS(t)I(t) + \alpha I(t), \\ I'(t) &= rS(t)I(t) - \alpha I(t), \end{aligned} \tag{1.2}$$

where $r > 0$ is the infection coefficient and $\alpha \geq 0$ is the recovery coefficient. Total size of population is constant: $S(t) + I(t) = N$ and the initial conditions are: $S(0) = S_0$, $I(0) = I_0 = N - S_0$.

1.2.1 Analytic solution

Using equality $S(t) = N - I(t)$ in equation (1.2), we get

$$I'(t) = r(N - I(t))I(t) - \alpha I(t).$$

By substitution

$$\begin{aligned} y &= I(t)^{-1} \\ I(t) &= \frac{1}{y(t)} \\ I'(t) &= -\frac{y(t)'}{y^2(t)} \end{aligned}$$

we obtain

$$-\frac{y'(t)}{y^2(t)} = r\left(N - \frac{1}{y(t)}\right)\frac{1}{y(t)} - \frac{\alpha}{y(t)}.$$

Multiplied by $y^2(t)$,

$$\begin{aligned} -y'(t) &= r(Ny(t) - 1) - \alpha y(t) \\ &= rNy(t) - r - \alpha y(t) \\ &= y(t)(rN - \alpha) - r. \end{aligned}$$

Adjusted for integration

$$y'(t) + (rN - \alpha)y(t) = r. \tag{1.3}$$

We denote a simplifying substitution $(rN - \alpha) = \beta$ and we get the final form of equation to solve

$$y'(t) + \beta y(t) = r.$$

Homogeneous solution
Let us consider equation with a right hand side equal to zero:

$$y'(t) + \beta y(t) = 0.$$

Solution of this equation is

$$y(t) = ce^{-\beta t}, c \in \mathbb{R}.$$

Using back-substitution, our homogeneous solution is

$$y(t) = ce^{-(rN-\alpha)t}, c \in \mathbb{R}.$$

Particular solution
Particular solution can be found using the method of variation of constants. We start with the homogeneous solution and we consider the constant to be a function of time.

$$y(t) = c(t)e^{-(rN-\alpha)t}$$

We make the first derivative

$$y'(t) = c'(t)e^{-(rN-\alpha)t} + c(t)(-(rN-\alpha))e^{-(rN-\alpha)t}.$$

Now, we can substitute y(t) and y(t)' into equation (1.3)

$$\begin{aligned} y'(t) + (rN-\alpha)y(t) = c'(t)e^{-(rN-\alpha)t} &= r \\ c'(t) &= re^{(rN-\alpha)t} \\ c(t) &= \frac{r}{rN-\alpha}e^{(rN-\alpha)t} + d, d \in \mathbb{R}. \end{aligned}$$

Knowing this, we can write down the particular solution:

$$y(t) = \frac{r}{rN-\alpha} + d, d \in \mathbb{R}.$$

General solution
General solution is a sum of homogeneous and particular solution.

$$y(t) = ce^{-(rN-\alpha)t} + \frac{r}{rN-\alpha}e^{-(rN-\alpha)t}, c \in \mathbb{R}$$

Using simplifying substitution $(rN - \alpha) =: \beta$

$$y(t) = \frac{c\beta e^{-\beta t} + r}{\beta}. \tag{1.4}$$

By applying initial conditions, we eliminate constant c.

$$\begin{aligned} I(0) &= I_0 \\ I(t) &= \frac{1}{y(t)} = \frac{\beta}{c\beta e^{-\beta t} + r} \\ I(0) &= \frac{\beta}{c\beta + r} = I_0 \\ I_0(r + \beta c) &= \beta \\ c &= \frac{\beta - I_0 r}{\beta I_0}. \end{aligned}$$

Now, we can substitute formula of constant c back into equation (1.4)

$$I(t) = \frac{\beta}{r + \beta \frac{\beta - I_0 r}{\beta I_0} e^{-\beta t}}.$$

The exact solution for S(t) obtained by using the fact that S(t) + I(t) = N is

$$S(t) = N - \frac{\beta}{r + \beta \frac{\beta - I_0 r}{\beta I_0} e^{-\beta t}}.$$

1.3 SIS limit behaviour

The dynamics of this model is determined by the ratio between parameters r and β. This ratio is called the basic reproduction number and it is the number of secondary infections caused by one infectious individual. We will denote it R_0.

$$R_0 = N\frac{r}{\alpha} = \frac{\beta}{\alpha} - 1$$

Let S(t) and I(t) be a solution.

1) If $R_0 \leq 1$, then

$$\lim_{t\to\infty}(S(t), I(t)) = (N, 0)$$

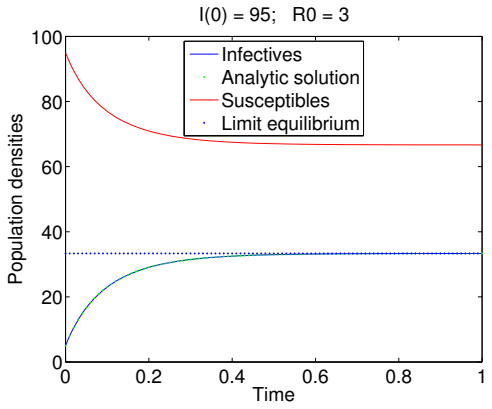
which means *disease-free equilibrium*.

2) If $R_0 > 1$, then

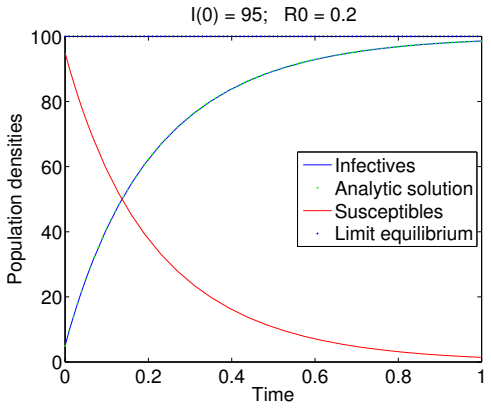
$$\lim_{t\to\infty}(S(t), I(t)) = \left(\frac{N}{R_0}, N - \frac{N}{R_0}\right)$$

which means *endemic equilibrium*.

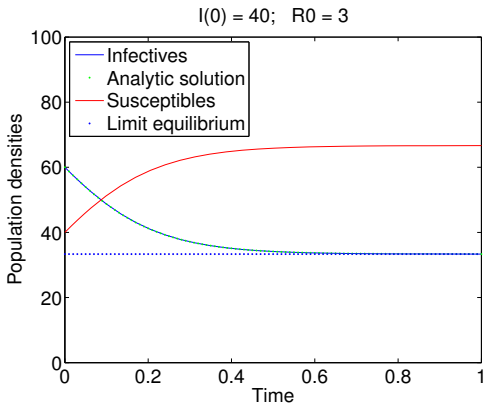
Four examples of dependency of the SIS model on the basic reproduction number and initial conditions are presented in Figures 1.2a, 1.2b, 1.2c and 1.2d. Total size of population is one hundred.



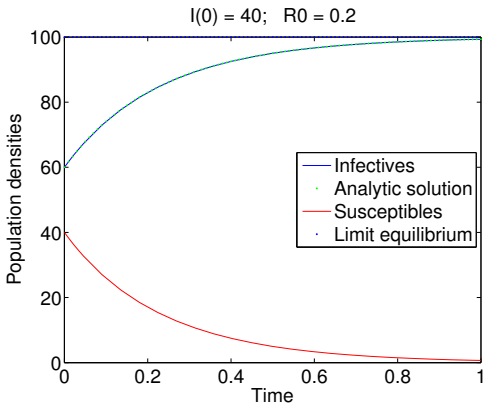
(a) Dynamics of epidemic with the parameters $R_0 = 3$ and $I(0) = 95$.



(b) Dynamics of epidemic with the parameters $R_0 = 0.2$ and $I(0) = 95$



(c) Dynamics of epidemic with the parameters $R_0 = 3$ and $I(0) = 40$.



(d) Dynamics of epidemic with the parameters $R_0 = 0.2$ and $I(0) = 40$.

Figure 1.2: SIS dynamics in both supercritical and sub-critical case

Chapter 2

Discrete-time Markov chain (DTMC) model

Let $S(t)$ and $I(t)$ denote discrete-time random variables for the number of susceptible and infectious individuals at time t, where $t \in \{0,\ \Delta t,\ 2\Delta t,\ ...\}$ and Δt is a sufficiently small time interval. Population size is assumed to be constant

$$S(t) + I(t) = N.$$

Let $\mathbf{p}(t) = (p_0(t),\ p_1(t),\ ...,\ p_N(t))^T$ denote the probability distribution vector associated with $I(t)$. It consists of elements

$$p_i(t) = \mathbb{P}\{I(t) = i\},$$

which describe the probability of the process being at the time t in the state i. The probability of a transition from state $I(t) = i$ to state $I(t + \Delta t) = j,\ i \to j$, in time Δt, is denoted as

$$p_{i,j}(t + \Delta t, t) = \mathbb{P}\{I(t + \Delta t) = j | I(t) = i\}.$$

To reduce the number of transitions in time Δt, we make a assumption that the time step Δt is chosen sufficiently small such that the number of infectious individuals changes by at most by one, that is,

$$i \to i + 1,\ i \to i - 1,\ \text{or}\ i \to i.$$

If there are i infectious individuals, the probability of a new infection in time Δt is $ri(N - i)\Delta t$. The probability of a recovery in time Δt is $\alpha i \Delta t$. And the probability of no change in time Δt is $1 - [ri(N - i) + \alpha i]\,\Delta t$.

$$p_{i,j}(\Delta t) = \begin{cases} ri(N - i)\Delta t, & j = i + 1 \\ \alpha i \Delta t, & j = i - 1 \\ 1 - [ri(N - i) + \alpha i]\,\Delta t, & j = i \\ 0 & j \neq i + 1, i, i - 1. \end{cases}$$

To simplify the notation and to relate the SIS epidemic process to a general birth and death process, the transition probability for a new infection is denoted as b(i)Δt and for a death or a recovery is denoted as d(i)Δt (Allen, 2008). Then,

$$p_{i,j}(\Delta t) = \begin{cases} b(i)\Delta t, & j = i+1 \\ d(i)\Delta t, & j = i-1 \\ 1 - [b(i) + d(i)]\,\Delta t, & j = i \\ 0 & j \neq i+1, i, i-1, \end{cases}$$

where $b(i) = ri(N-i)$ and $d(i) = \alpha i$. To ensure that the sum of these probabilities representing all possible changes in the state i during the time interval Δt equals to one, the time step Δt must be chosen sufficiently small such that

$$\max_{i \in S} \{[b(i) + d(i)]\,\Delta t\} \leq 1.$$

Then the probabilities $p_i(t + \Delta t)$ can be expressed in terms of the probabilities at time t as:

$$p_i(t+\Delta t) = p_{i-1}(t)b(i-1)\Delta t + p_{i+1}(t)d(i+1)\Delta t + p_i(t)(1 - [b(i) + d(i)]\Delta t), \quad (2.1)$$

for $i = 1,\ 2,\ ...,\ N$.

Let us denote the transition matrix as **P**(Δt). Matrix **P**(Δt) is a $(N+1) \times (N+1)$ tridiagonal stochastic matrix consisting of all possible transition probabilities in the state i by rows:

$$\begin{pmatrix} 1 - b_0\Delta t & b_0\Delta t & 0 & 0 & \cdots & 0 & 0 & 0 \\ d_1\Delta t & 1 - (b_1 + d_1)\Delta t & b_1\Delta t & 0 & \cdots & 0 & 0 & 0 \\ 0 & d_2\Delta t & 1 - (b_2 + d_2)\Delta t & b_2\Delta t & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & d_{N-1}\Delta t & 1 - (b_{N-1} + d_{N-1})\Delta t & b_{N-1}\Delta t \\ 0 & 0 & 0 & 0 & \cdots & 0 & d_N & 1 - d_N \end{pmatrix},$$

where b_0 is 0 since the origin is a absorbing state. Probability vector at time Δt is easy to be found as

$$\mathbf{p}(\Delta t) = \mathbf{P}(\Delta t)\mathbf{p}(0),$$

where **p**(0) is a given initial probability vector. Which results in a matrix notation for identity (2.1)

$$\mathbf{p}(t + \Delta t) = \mathbf{P}(\Delta t)\mathbf{p}(t) = \mathbf{P}^{n+1}(\Delta t)\mathbf{p}(0),$$

where $t = n\Delta t$.

2.1 Mean and variance

As follows from (Allen, 2008), difference equations for the mean and the higher order moments of the epidemic process can be obtained directly from the difference equation (2.1). The expected value for $I(t)$ is $E(I(t)) = \sum_{i=0}^{N} i p_i(t)$. Multiplying (2.1) by i and summing on i leads to

$$\begin{aligned}\mathbb{E}(I(t+\Delta t)) =& \sum_{i=0}^{N} i p_i(t+\Delta t) = \sum_{i=1}^{N} i p_{i-1}(t) b(i-1)\Delta t + \sum_{i=0}^{N-1} i p_{i+1}(t) d(i+1)\Delta t \\ &+ \sum_{i=0}^{N} i p_i(t) - \sum_{i=0}^{N} i p_i(t)[b(i) - d(i)]\Delta t.\end{aligned}$$

By substitution of $ri(N-i)$ and αi for $b(i)$ and $d(i)$, respectively, we obtain

$$\begin{aligned}\mathbb{E}(I(t+\Delta t)) =& \mathbb{E}(I(t)) + \sum_{i=1}^{N} p_{i-1}(t) r(i-1)(N-[i-1])\Delta t - \sum_{i=0}^{N-1} p_{i+1}(t)\alpha(i+1)\Delta t \\ =& \mathbb{E}(I(t)) + [rN - \alpha]\Delta t \mathbb{E}(I(t)) - r\Delta t \mathbb{E}(I^2(t)),\end{aligned}$$

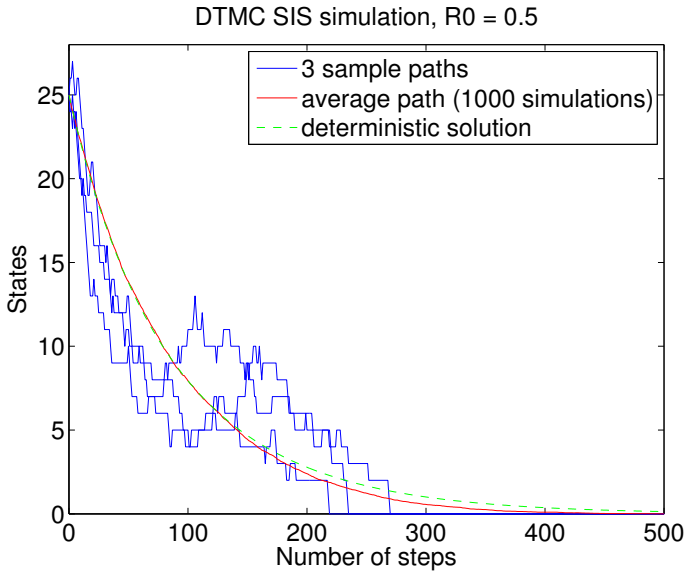
where $\mathbb{E}(I^2(t)) = \sum_{i=0}^{N} i^2 p_i(t)$. The difference equation for the mean depends on the second moment and can not be solved precisely because the difference equations for higher moments depend on even higher order moments. However, we can make some assumptions regarding the second moment, such as $\mathbb{E}(I^2(t)) \geq \mathbb{E}^2(I(t))$, to be able to bound the estimate of mean. From which follows

$$\frac{\mathbb{E}(I(t+\Delta t)) - E(I(t))}{\Delta t} \leq [rN - \alpha]\mathbb{E}(I(t)) - r\mathbb{E}^2(I(t))$$

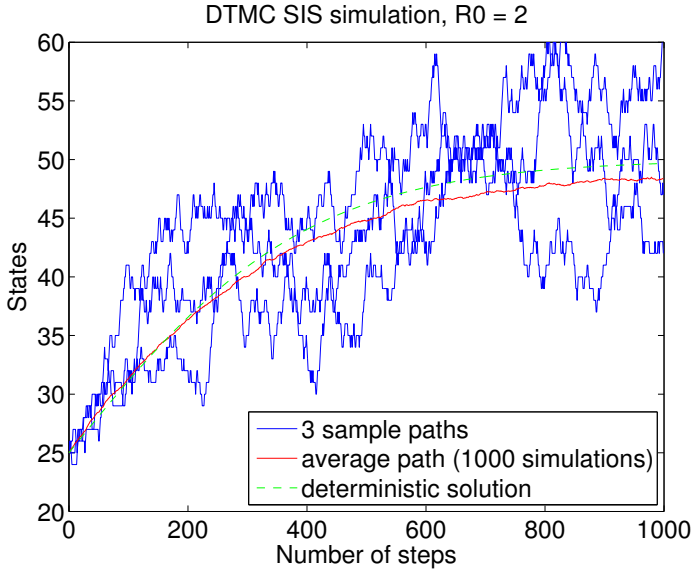
and as $\Delta t \to 0$,

$$\begin{aligned}\frac{d\mathbb{E}(I(t))}{dt} &\leq [rN - \alpha]\mathbb{E}(I(t)) - r\mathbb{E}^2(I(t)) \\ &= r[N - \mathbb{E}(I(t))]E(I(t)) - \alpha\mathbb{E}(I(t)). \qquad (2.2)\end{aligned}$$

The right side of (2.2) is the same as the differential equation for $I'(t)$ in (1.2), if $I(t)$ is considered as its expected value. This differential inequality indicates that the mean of $I(t)$ in the stochastic SIS epidemic process is less than the solution of the deterministic model (see Figures 2.2a and 2.2b). Since the difference equation for $\mathrm{Var}(I(t))$ depends on the third moment, it can not be solved precisely. Therefore, numerical solution only is presented in Figures 2.1a and 2.1b.

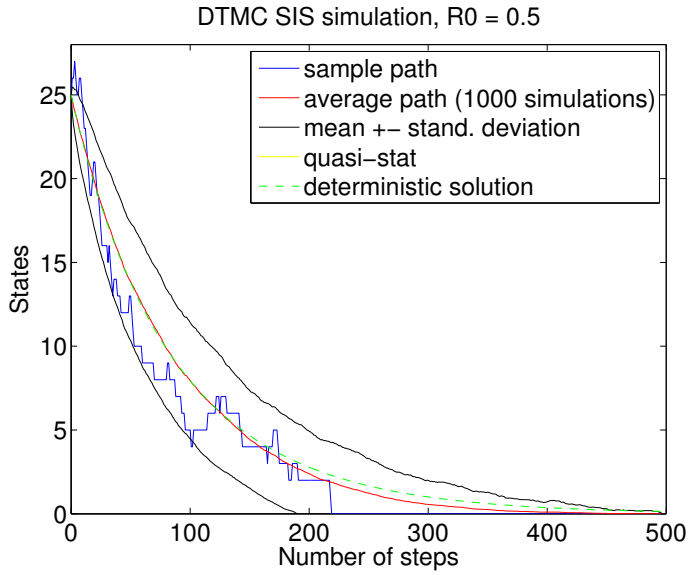


(a) Comparison of the solution of the deterministic model and the average of 1000 simulations for $R_0 = 0.5$

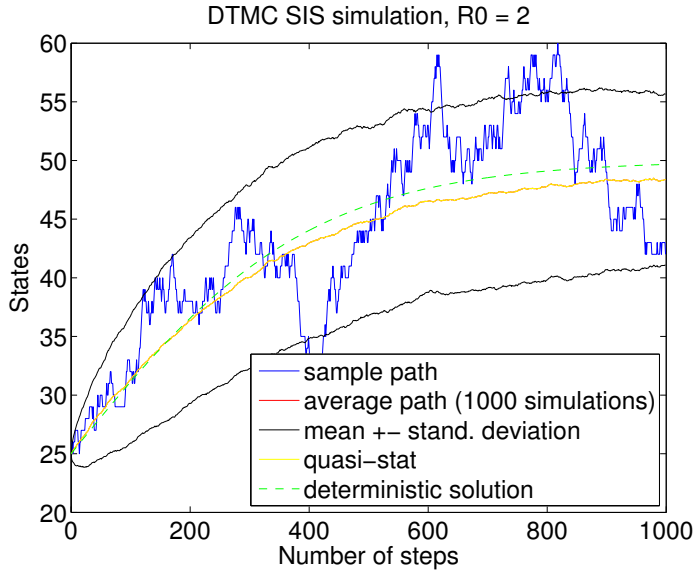


(b) Comparison of the solution of the deterministic model and the average of 1000 simulations for $R_0 = 2$

Figure 2.1: SIS dynamics in both supercritical and sub-critical case



(a) Mean quasi-stationary path and standard deviation for $R_0 = 0.5$



(b) Mean quasi-stationary path and standard deviation for $R_0 = 2$

Figure 2.2: Average non-extinct path in both supercritical and sub-critical case

2.2 Mean persistence time

As can be seen from the transition matrix $\mathbf{P}(\Delta t)$, the state space $\mathbf{S} = \{0, 1, ..., N\}$ is divided into set of transient states, $\{1, ..., N\}$ and the absorbing state, $\{0\}$. Every state in the set $\{1, ..., N\}$ is accessible from any other state in the set and no state can be reached from $\{0\}$, but itself. It can be shown that for any transient state i,

$$\lim_{n\to\infty} p_{i,j}^{(n)} = 0,$$

where $p_{i,j}^{(n)}$ is the (i, j) element of the n-th power of the transition matrix $\mathbf{P}(\Delta t)$. Since the limit of matrix $\mathbf{P}(\Delta t)^n$ as $n \to \infty$ is a stochastic matrix where all collumns are zero except the first one which has all ones, vector of a limit distribution is simply

$$\lim_{n\to\infty} \mathbf{p}(t) = (1, 0, ..., 0)^T,$$

where $t = n\Delta t$. This result implies that the Markov chain SIS model approaches the disease-free equilibrium regardless of the value of the basic reproduction number. Therefore the probability of absorption is always one. However, depending on the initial number of infectious individuals i, the population size N and the value of R_0, the time until absorption can be very short or very long (Allen, 2008).

Let T_k denote the extinction time for a population with initial size k and $\tau_k = \mathbb{E}[T_k | I(0) = k]$ expected time to extinction given the initial condition $I(0) = k$.

$$\tau_k = b(k)\Delta t(\Delta t + \tau_{k+1}) + d(k)\Delta t(\Delta t + \tau_{k-1}) + [1 - (b(k) + d_k)\Delta t]\,(\Delta t + \tau_k),$$

for $k \in \{1, ..., N-1\}$. If we consider $\Delta t = 1$, preceding formula reduces to

$$\tau_k = b(k)(1 + \tau_{k+1}) + d(k)(1 + \tau_{k-1}) + [1 - (b(k) + d(k))]\,(1 + \tau_k),$$

for $k \in \{1, ..., N-1\}$;

$$\begin{aligned}
\tau_1 &= 1 + b(1)\tau_2 + [1 - (b(1) + d(1))]\,\tau_1 \\
&\vdots \\
\tau_k &= 1 + b(k)\tau_{k+1} + d(k)\tau_{k-1} + [1 - (b(k) + d(k)]\,\tau_k \\
&\vdots \\
\tau_N &= 1 + d(N)\tau_{N-1} + (1 - d(N))\tau_N,
\end{aligned}$$

which can be written in matrix form

$$\mathbf{D}\boldsymbol{\tau} = -\mathbf{1},$$

where $\boldsymbol{\tau} = (\tau_1,\ \tau_2,\ \ldots,\ \tau_N)^T, \mathbf{1} = (1,\ \ldots,\ 1)^T$ and $\mathbf{D}$ is defined as

$$\begin{pmatrix} -(b(1)+d(1)) & b(1) & 0 & \cdots & 0 & 0 & 0 \\ d(2) & -(b(2)+d(2)) & b(2) & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & d(N-1) & -(b(N-1)+d(N-1)) & b(N-1) \\ 0 & 0 & 0 & \cdots & 0 & d(N) & -d(N) \end{pmatrix}.$$

Since $\mathbf{D}$ is irreducible diagonally dominant and therefore non-singular, $\mathbf{D}^{-1}$ exists and the mean persistence time is uniquely given by

$$\boldsymbol{\tau} = -\mathbf{D}^{-1}\mathbf{1}.$$

Moreover, similar difference equations apply to the higher order moments τ_k^r:

$$d(k)\tau_{k-1}^r - [b(k)+d(k)]\tau_k^r + b(k)\tau_{k+1}^r = -r\tau_k^{r-1},$$

which can be written in matrix form

$$\mathbf{D}\boldsymbol{\tau}^r = -r\boldsymbol{\tau}^{-1}.$$

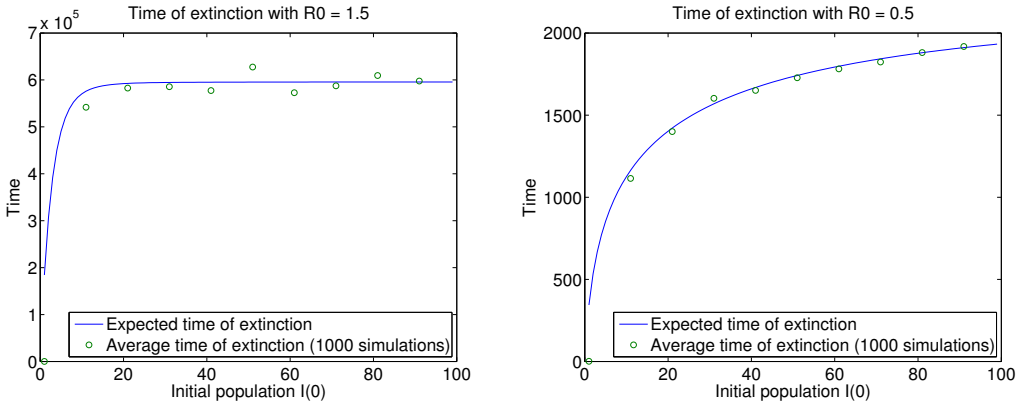
Since D has a special form, using factorisation method for three-diagonal matrices, it is possible to write a recursive formula for τ_k (see Allen and Allen, 2003).

$$\tau_k = \begin{cases} \frac{1}{d(1)} + \sum_{i=2}^{N} \frac{b(1)\ldots b(i-1)}{d(1)\ldots d(i)}, & k = 1, \\ \tau_1 + \sum_{s=1}^{k-1} \left[\frac{d(1)\ldots d(s)}{b(1)\ldots b(s)} \sum_{i=s+1}^{N} \frac{b(1)\ldots b(i-1)}{d(1)\ldots d(i)} \right] & k = 2, \ldots, N. \end{cases}$$

And for the higher moments

$$\tau_k^r = \begin{cases} r\frac{\tau_1^{r-1}}{d(1)} + r\sum_{i=2}^{N} \frac{b(1)\ldots b(i-1)\tau_i^{r-1}}{d(1)\ldots d(i)}, & k = 1, \\ \tau_1^r + r\sum_{s=1}^{k-1} \left[\frac{d(1)\ldots d(s)}{b(1)\ldots b(s)} \sum_{i=s+1}^{N} \frac{b(1)\ldots b(i-1)\tau_i^{r-1}}{d(1)\ldots d(i)} \right] & k = 2, \ldots, N. \end{cases}$$

Two examples of dependency of the expected time of extinction $\mathbb{E}[T_k]$ on initial condition $I(0)$ and reproduction number R_0 follow. In Figure 2.3a, reproduction number is greater than one and the population is in a quasi-stationary outbreak. Figure 2.3b represents expected times of extinction for reproduction number smaller than one, i.e. the disease is receding. Figure 2.4 represents variance of the persistence time in the supercritical case.



(a) Mean persistence time for $R_0 = 1.5$ (b) Mean persistence time for $R_0 = 0.5$

Figure 2.3: Mean persistence time in both supercritical and sub-critical case

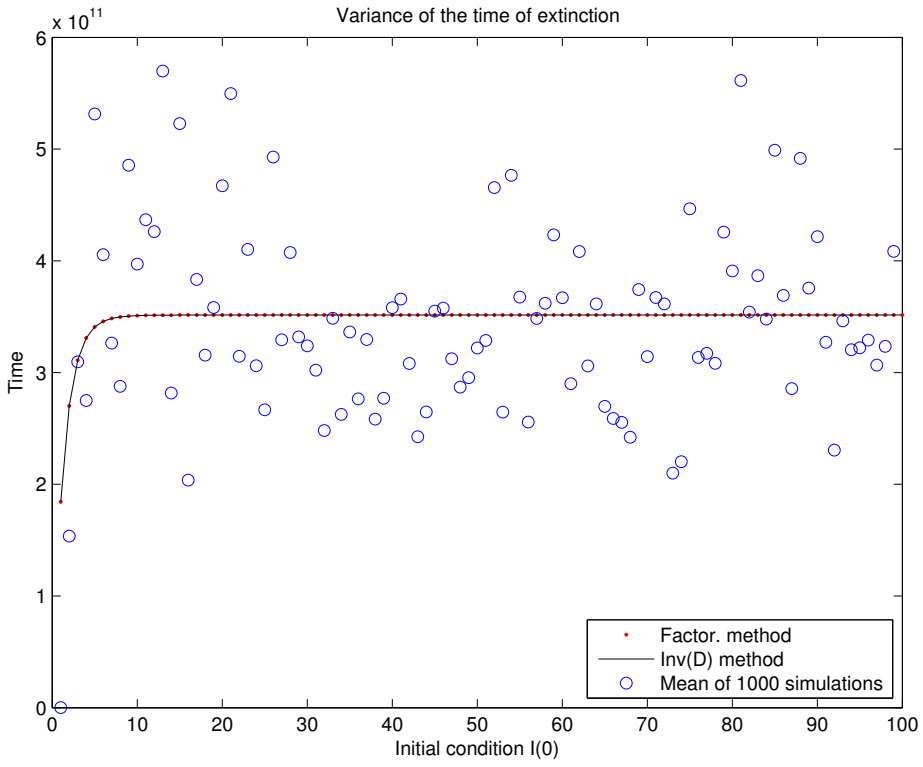


Figure 2.4: Variance of the persistence time for $R_0 = 1.5$

2.3 Quasi-stationary distribution

Knowing that the extinction is inevitable, let us consider $(Q(n))$ to be a population size at time n of the given process $(I(n))$, conditional on non-extinction.

$$\begin{aligned} p_i(n) &= \mathbb{P}\{I(n) = i\} \\ p_0(n) &= \mathbb{P}\{I(n) = 0\} \\ \mathbb{P}\{I(n) \neq 0\} &= 1 - p_0(n) \\ q_i(n) &= \mathbb{P}\{I(n) = i | \overbrace{I(j) \neq 0, j = 0,\ 1,\ ...,\ N-1}^{\text{condition on non-extinction}}\} \\ &= \frac{p_i(n)}{1 - p_0(n)} \end{aligned} \tag{2.3}$$

Lemma 2.1. *$q_i(n) = \mathbb{P}\{Q(n) = i\}$ is a probability distribution on $\{1,\ ...,\ N\}$.*

Proof.

$$\sum_{i=1}^{N} q_i(n) = \frac{\sum\limits_{i=1}^{N} p_i(n)}{1 - p_0(n)} = \frac{1 - p_0(n)}{1 - p_0(n)} = 1.$$

□

Let $\mathbf{q}^*$ denote the stationary distribution of (Q_n). It is also the quasi-stationary distribution of (I_n).

2.3.1 Exact solution

Using assumption that $\Delta t = 1$, the equation (2.1) for $p_i(t + \Delta t)$ results in

$$p_i(n+1) = p_{i-1}(n)b(i-1)\Delta t + p_{i+1}(n)d(i+1) + p_i(n)(1 - [b(i) + d(i)]). \tag{2.4}$$

Based on preceding formula (2.3) for $q_i(n)$, we make the formula for $q_i(n+1)$

$$q_i(n+1) = \frac{p_i(n+1)}{1 - p_0(n+1)}$$

which after multiplying by $\frac{1-p_0(n)}{1-p_0(n)}$ leads to

$$= \frac{p_i(n+1)}{1 - p_0(n)} \frac{1 - p_0(n)}{1 - p_0(n+1)}.$$

By substitution of $p_0(n+1)$ from (2.4), we get

$$= \frac{p_i(n+1)}{1 - p_0(n)} \frac{1 - p_0(n)}{1 - d_1 p_1(n)}.$$

Finally, substitution of $p_i(n+1)$ from (2.3) gives us iterative scheme for $q_i(n+1)$

$$q_i(n+1) = q_{i-1}(n)\frac{b_{i-1}}{1-d_1q_1(n)} + q_i(n)\frac{1-(b_i+d_i)}{1-d_1q_1(n)} + q_{i+1}(n)\frac{d_{i+1}}{1-d_1q_1(n)},$$

in matrix form as a non-linear equation (it involves also an unknown element q_1^*)

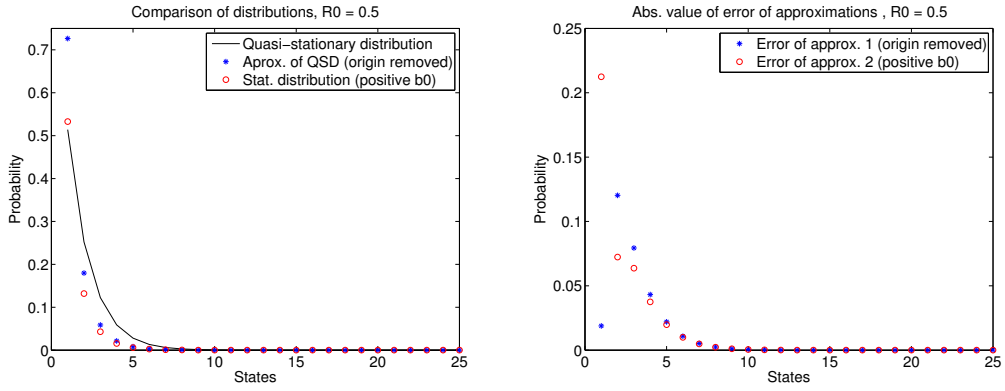
$$\mathbf{D}\mathbf{q}^* = -d_1q_1^*\mathbf{q}^*, \tag{2.5}$$

where $\mathbf{D}$ is the submatrix of $\mathbf{P}$ defined in section 2.2. Initial value of $\mathbf{q}^*$ for the iteration is obtained from (2.3). Subsequent iterations are found using matrix $\mathbf{D}$ in (2.5). Computation is terminated when the norm of difference between last two iterations is reduced so that it meets the termination condition, e.g. 10^{-6}.

2.3.2 Approximation

There are two basic concepts of approximating quasi-stationary distribution. Both are based on omitting the absorbing state, so that the new process remains irreducible. Than, we can approximate the original quasi-stationary distribution by stationary distribution of the new irreducible process. It is known, that any irreducible and aperiodic Markov chain has exactly one stationary distribution (Häggström, 2002).

First way to discard the absorbing state is to remove the first column and row of the transition matrix and consider the reduced state space $\mathbf{S}$ to be the set of former transient states $\{i_1, ..., i_k\}$. Other option is to keep the original state space $\mathbf{S} = \{i_0, ..., i_k\}$ and remove the property of reducibility only. This can be done by assigning a small positive value to the term $b(0)$, so that the the state i_1 becomes accessible from the state i_0. Thus, i_0 ceases to be absorbing.

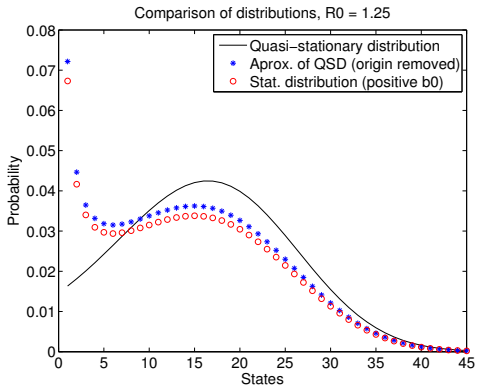


(a) Quasi-stationary distribution and its approximations

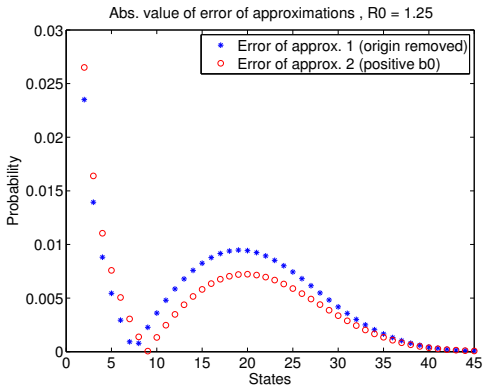
(b) Absolute value of errors of approximations

Figure 2.5: Quasi-stationary distribution for $R_0 = 0.5$

In the Figures 2.5a, 2.6a and 2.7a, the quasi-stationary distribution and its approximations for three different values of R_0 can be seen. Figures 2.5b, 2.6b and 2.7b feature errors in absolute values. Total size of population is $N = 100$.



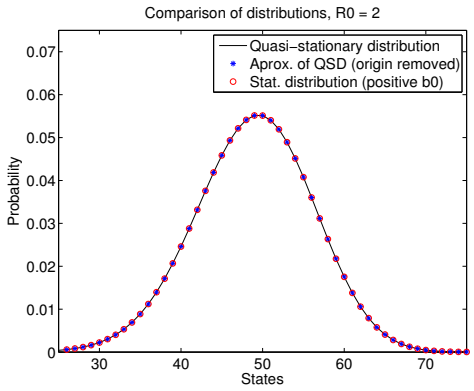
(a) Quasi-stationary distribution and its approximations



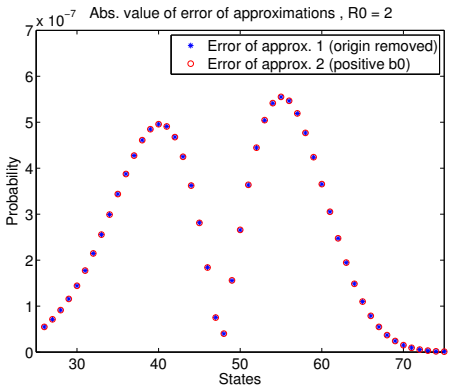
(b) Absolute value of errors of approximations

Figure 2.6: Quasi-stationary distribution for $R_0 = 1.25$

Figures 2.6a and 2.6b show the quasi-stationary distribution of a population with the reproduction number R_0 close to 1. It can be seen that the approximations are not precise in this so-called transition region.



(a) Quasi-stationary distribution and its approximations



(b) Absolute value of errors of approximations

Figure 2.7: Quasi-stationary distribution for $R_0 = 2$

Chapter 3

Continuous-time Markov chain (CTMC) model

The CTMC SIS process is defined on a continuous time scale, $t \in [0, \infty)$. Population is represented by discrete-valued continuous-time process $I(t)$ and it's probability vector $\mathbf{p}(t) = (p_0(t),\ p_1(t),\ ...,\ p_N(t))$, where

$$p_i(t) = \mathbb{P}\{I(t) = i\}.$$

In the CTMC model, the transition probabilities are called infinitesimal transition probabilities. Again, Δt is considered to be sufficiently small, so that there is at most by-one change of state possible. In the definition, there is included the term $o(\Delta t)$ and $\lim_{t\to\infty}(o(\Delta t)/\Delta t) = 0$. The infinitesimal transition probabilities are defined (Allen, 2008)

$$p_{i,j}(\Delta t) = \begin{cases} ri(N-i)\Delta t + o(\Delta t), & j = i+1 \\ \alpha i \Delta t + o(\Delta t), & j = i-1 \\ 1 - [ri(N-i) + \alpha i]\,\Delta t + o(\Delta t), & j = i \\ o(\Delta t), & j \neq i+1, i, i-1. \end{cases}$$

Using the notation from DTMC SIS, we obtain

$$p_{i,j}(\Delta t) = \begin{cases} b(i)\Delta t + o(\Delta t), & j = i+1 \\ d(i)\Delta t + o(\Delta t), & j = i-1 \\ 1 - [b(i) + d(i)]\,\Delta t + o(\Delta t), & j = i \\ o(\Delta t), & j \neq i+1, i, i-1, \end{cases}$$

where $b(i) = ri(N-i)$ and $d(i) = \alpha i$. When dealing with a continuous time scale, instead of a system of difference equations, we get a system of differential equations

$$p_i(t+\Delta t) = p_{i-1}(t)b(i-1)\Delta t + p_{i+1}(t)d(i+1)\Delta t + p_i(t)(1 - [b(i) + d(i)]\Delta t + o(\Delta t)).$$

Subtracting $p_i(t)$, dividing by Δt, and letting $\Delta t \to 0$, leads to so called forward Kolmogorov differential equations

$$\frac{dp_i}{dt} = p_{i-1}b(i-1) + p_{i+1}d(i+1) + p_i[b(i) + d(i)],$$

for $i = 1,\ 2,\ ...,\ N$ and

$$\frac{dp_0}{dt} = p_1 d(1),$$

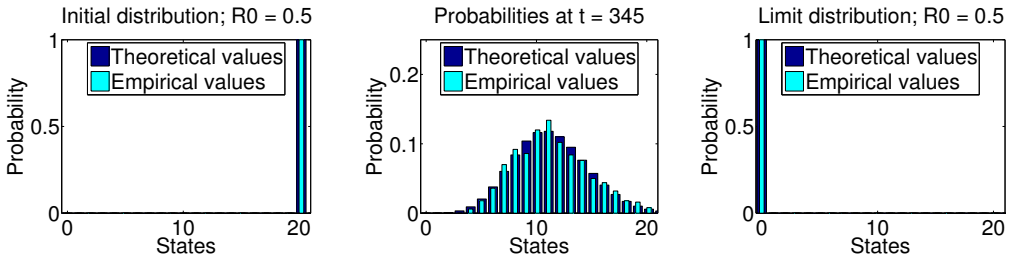
for the absorbing state. This system can be gracefully expressed in matrix notation as

$$\frac{d\mathbf{p}}{dt} = \mathbf{Q}\mathbf{p}, \tag{3.1}$$

where $\mathbf{p}(t) = (p_0(t),\ p_1(t),\ ...,\ p_N(t))$ and $\mathbf{Q}$ is known as infinitesimal generator matrix, defined as

$$\mathbf{Q} = \begin{pmatrix} -b_0 & b_0 & 0 & 0 & \cdots & 0 & 0 \\ d_1 & -(b_1+d_1) & b_1 & 0 & \cdots & 0 & 0 \\ 0 & d_2 & -(b_2+d_2) & b_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & d_N & -d_N \end{pmatrix}.$$

Following three figures represent progress of the epidemic in the sub-critical case. Figure 3.1a features initial distribution, Figure 3.1b displays probability distribution during the progress. Limit distribution can be seen in the Figure 3.1c.



(a) Initial distribution ($I_0 = 20$)

(b) Probability distribution at time $t = 345$

(c) Limit distribution - absorption

Figure 3.1: Probability distribution at three times; $R_0 = 0.5$

3.1 Mean and variance

Differential equation for the mean of $I(t)$ can be derived from the equation (3.1) using a probability generating function or moments generating function. The probability generating function for $I(t)$ is defined as

$$\pi(s,t) = E(s^{I(t)}) = \sum_{i=0}^{\infty} p_i(t)s^i,$$

with its partial derivative with respect to t denoted by

$$\frac{\partial \pi(s,t)}{\partial t} = \sum_{i=0}^{\infty} p_i'(t)s^i,$$

partial derivative with respect to s denoted by

$$\frac{\partial \pi(s,t)}{\partial s} = \sum_{i=1}^{\infty} p_i(t)s^{i-1}i,$$

and its second partial derivative with respect to s denoted by

$$\frac{\partial^2 \pi(s,t)}{\partial s^2} = \sum_{i=2}^{\infty} p_i(t)s^{i-2}i^2.$$

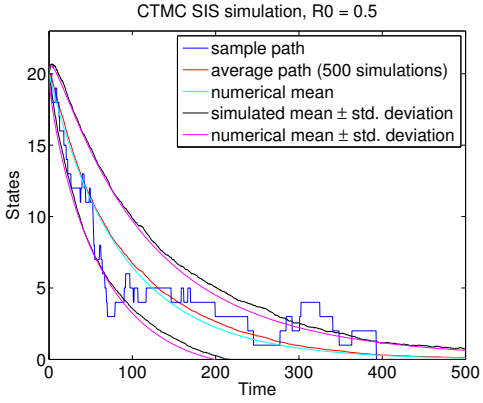
With respect to the general birth and death process, let us substitute $d(j) = j(d_1 + jd_2)$ and $b(j) = j(b_1 + jb_2)$, where b_1, b_2, d_1 and d_2 are general non-negative constants. Based on preceding formulas and equation (3.1), we obtain the following second order partial differential equation for the probability generating function

$$\frac{\partial \pi(s,t)}{\partial t} = [b_1(s^2 - s) + d_1(1 - s)]\frac{\partial \pi(s,t)}{\partial s} + [b_2(s^3 - s^2) + d_2(s - s^2)]\frac{\partial^2 \pi(s,t)}{\partial s^2}.$$

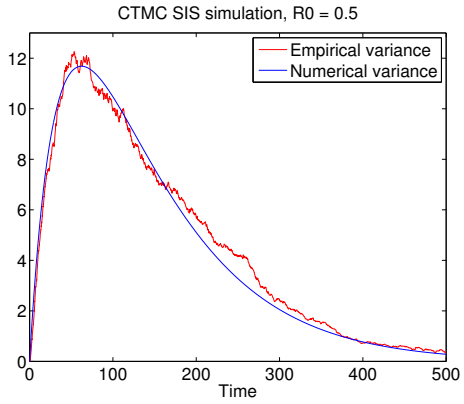
PDE for the probability generating function seems to be rather difficult to solve, but as shown by (Allen, 2008), the PDE for the moment generating function can be used to obtain an ordinary differential equation satisfied by the mean of $I(t)$. As well as in the discrete time case, the equation for the mean depends on the second moment and therefore, it cannot be solved directly:

$$\frac{d\mathbb{E}(I(t))}{dt} = (rN - \alpha)\mathbb{E}(I(t)) - r\mathbb{E}(I^2(t)).$$

This equation can be solved by making some assumptions regarding higher order moments to give an approximation. Similarly to DTMC, the mean of the stochastic SIS epidemic model is less than solution of the deterministic one. In Figures 3.2a and 3.3a we can see simulated epidemic for parameter $R_0 = 0.5$ and $R_0 = 2$ respectively. Variance is provided here from simulations only (see Figures 3.2b and 3.3b).

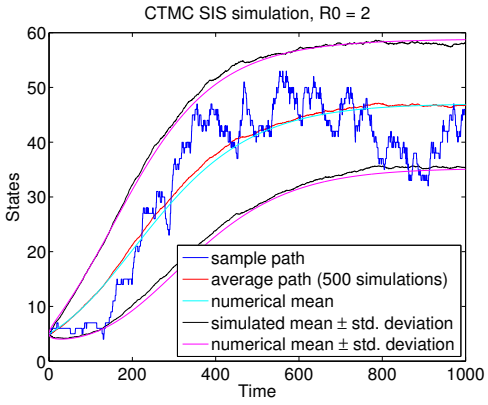


(a) Population approaching disease-free equilibrium ($I_0 = 25$)

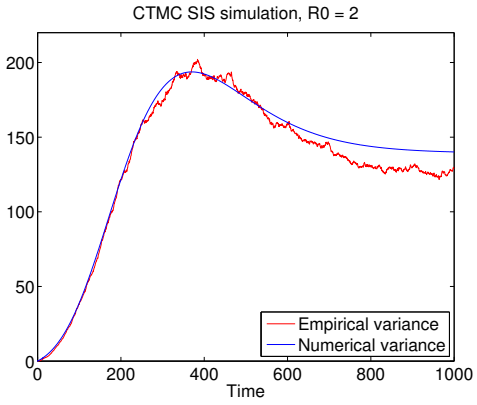


(b) Variance of the simulated $I(t)$

Figure 3.2: Simulation for $R_0 = 0.5$



(a) Population approaching endemic equilibrium ($I_0 = 5$)



(b) Variance of the simulated $I(t)$

Figure 3.3: Simulation for $R_0 = 2$

3.2 Mean persistence time

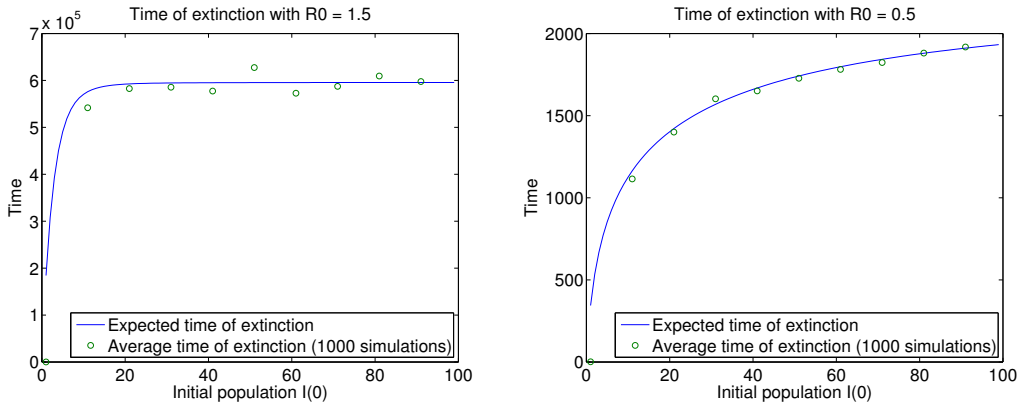
As it was shown on DTMC, SIS stochastic model approaches absorption regardless of the value of the basic reproduction number. CTMC does not evince otherwise. Let T_k denote the extinction time for a population with initial size k and $\tau_k = \mathbb{E}[T_k | I(0) = k]$ expected time to extinction given the initial condition $I(0) = k$. Since the matrix $\mathbf{D}$ is tridiagonal, an explicit solution for the τ_k and τ_k^r can be found as was shown by (Richter-Dyn and Goel, 1972).

$$\tau_k = \begin{cases} \frac{1}{d(1)} + \sum_{i=2}^{N} \frac{b(1)...b(i-1)}{d(1)...d(i)}, & k = 1, \\ \tau_1 + \sum_{s=1}^{k-1} \left[\frac{d(1)...d(s)}{b(1)...b(s)} \sum_{i=s+1}^{N} \frac{b(1)...b(i-1)}{d(1)...d(i)} \right] & k = 2, \ ..., \ N. \end{cases}$$

And for the higher moments

$$\tau_k^r = \begin{cases} r\frac{\tau_1^{r-1}}{d(1)} + r \sum_{i=2}^{N} \frac{b(1)...b(i-1)\tau_i^{r-1}}{d(1)...d(i)}, & k = 1, \\ \tau_1^r + r \sum_{s=1}^{k-1} \left[\frac{d(1)...d(s)}{b(1)...b(s)} \sum_{i=s+1}^{N} \frac{b(1)...b(i-1)\tau_i^{r-1}}{d(1)...d(i)} \right] & k = 2, \ ..., \ N. \end{cases}$$

Two examples of dependency of the expected time of extinction $\mathbb{E}[T_k]$ on initial condition $I(0)$ and reproduction number R_0 follow. In the first case, reproduction number is greater than one and the population is in a quasi-stationary outbreak (see Figure 3.4a). The second graph represents expected times of extinction for reproductive number smaller than one i.e. the disease is receding (see Figure 3.4b).



(a) Mean persistence time for $R_0 = 1.5$ (b) Mean persistence time for $R_0 = 0.5$

Figure 3.4: Mean times of extinction

3.3 Quasi-stationary distribution

Analogically to DTMC SIS, we denote $q_i(t)$ the probability that process is in the state i at the time t, conditional on non-extinction.

3.3.1 Exact solution

$$q_i(t) = \mathbb{P}\{I(t) = i | I(s) > 0; t > s\},$$

for $i = 1, ..., N$. Since the state $\{0\}$ is absorbing,

$$q_1(t) = \mathbb{P}\{I(s) > 0 | t > s\} = 1 - p_0(t).$$

And for $i = 1, ..., N$

$$q_i(t) = \frac{p_i(t)}{1 - p_0(t)}. \tag{3.2}$$

From forward Kolmogorov equations (1.1) we obtain the formula for $p_i'(t)$:

$$\begin{aligned}\mathbf{p}'(t) &= \mathbf{p}(t)\mathbf{Q} \\ p_i'(t) &= p_{i-1}(t)b(i-1) - p_i(t)[b(i) + d(i)] + p_{i+1}d(i+1),\end{aligned}$$

for $i = 1, ..., N$ and

$$p_0'(t) = d(1)p_1(t).$$

Substitution of (3.2) results in formula for $q_i'(t)$:

$$\begin{aligned}q_i'(t) &= \frac{p_i'(t)(1 - p_0(t)) + p_i(t)p_0'(t)}{[1 - p_0(t)]^2} \\ &= \frac{p_i'(t)}{1 - p_0(t)} + \frac{p_i(t)}{1 - p_0(t)} + \frac{d(1)p_1(t)}{1 - p_0(t)} \\ &= \frac{1}{1 - p_0(t)}\left[p_{i-1}(t)b(i-1) - p_i(t)[b(i) + d(i)] + p_{i+1}d(i+1)\right] + d(1)q_1(t)q_i(t) \\ &= b(i-1)q_{i-1}(t) - [b(i) + d(i)]q_i(t) + d(i+1)q_{i+1}(t) + d(1)q_1(t)q_i(t).\end{aligned}$$

Which can be expressed in matrix form as

$$\mathbf{q}_i'(t) = \mathbf{q}(t)\tilde{\mathbf{Q}} + d(1)q_1(t)\mathbf{q}(t),$$

where $\tilde{\mathbf{Q}}$ represents matrix $\mathbf{Q}$ with the first row and column deleted. This formula gives us iterative scheme for computing the quasi-stationary distribution $\mathbf{q}^*$ as:

$$0 = \mathbf{q}^*\tilde{Q} + d(1)q_1^*\mathbf{q}^*.$$

Unfortunately, matrix $\tilde{\mathbf{Q}}$ is ill-conditioned with eigenvalues close to zero. Therefore, an alternative approach was presented by Nåsell (1999), who exploits the fact, that the quasi-stationary distribution $\mathbf{q}^*$ is a left eigenvector of the matrix $\mathbf{Q}$ corresponding to the eigenvalue $-d(1)q_1$. He derived the following scheme:

$$q_j^* = \gamma(j)\alpha(j)R_0^{j-1}q_1,$$

for $j = 1, \ldots, N$, where

$$\begin{aligned}
\gamma(j) &= \frac{1}{j}\sum_{k=1}^{j}\delta(k), \\
\delta(k) &= \frac{1-\sum_{l=1}^{k-1}q_l}{\alpha(k)R_0^{k-1}}, \\
\alpha(j) &= \frac{N!}{(N-j)!N^j} \\
q_1 &= \frac{1}{S} \\
S &= \sum_{j=1}^{N}\gamma(j)\alpha(j)R_0^{j-1}.
\end{aligned} \tag{3.3}$$

3.3.2 Approximation

Analogically to DTMC SIS, we consider two approximations of quasi-stationary distribution. In both cases, state space is reduced by the absorbing state $\{0\}$, thus there exist non-degenerate stationary distributions that can be found explicitly.

In the first modified model, there is one permanently infectious individual, so that the population will not extinct. Recovery rates $d(i) = \alpha i$ are replaced by $d^{(1)}(i) = \alpha(i-1)$ and infection rates remain unchanged. Let us denote the stationary distribution of the first approximation of the SIS model $\mathbf{p}^{(1)} = (p_1^{(1)}, p_2^{(1)}, \ldots, p_N^{(1)})^T$. This distribution satisfies the explicit relation

$$p_j^{(1)} = \alpha(j)R_0^{j-1}p_1^1,$$

j = 1, 2, . . . , N, where $\alpha(j)$ is defined in eq. (3.3) and

$$p_1^{(1)} = \frac{1}{S^{(1)}},$$

with

$$S^{(1)} = \sum_{j=1}^{N}\alpha(j)R_0^{j-1}.$$

The second approximation represents the SIS model with the origin removed. In this approximation the recovery rate $d(1)$ is replaced by 0. All other transition rates remain unchanged. Let us denote the stationary distribution $\mathbf{p}^{(0)} = (p_1^{(0)}, p_2^{(0)}, \ldots, p_N^{(0)})^T$. Again, the distribution is found directly as

$$p_j^{(0)} = \frac{1}{j}\alpha(j)R_0^{j-1}p_1^0,$$

$j = 1, 2, \ldots, N$, where $\alpha(j)$ is defined in eq. (3.3) and

$$p_1^{(0)} = \frac{1}{S^{(0)}},$$

with

$$S^{(0)} = \sum_{j=1}^{N} \frac{1}{j}\alpha(j)R_0^{j-1}.$$

As was shown by (Kryscio and Lefévre, 1989) the original quasi-stationary distribution $\mathbf{q}^*$ is well approximated by the distribution $\mathbf{p}^{(0)}$ when R_0 is distinctly larger than 1 and by the distribution $\mathbf{p}^{(1)}$ when R_0 is distinctly smaller than 1. As the value of R_0 passes one, the quasi-stationary distribution makes a transition from close to $\mathbf{p}^{(1)}$ to close to $\mathbf{p}^{(0)}$ (Nåsell, 1999).

Moreover, Clancy and Mendy (2011) introduced several approximations of both $\mathbf{p}^{(1)}$ and $\mathbf{p}^{(0)}$, out of which the most accurate were the beta-binomial distribution and the geometric distribution.

In the supercritical case, if $R_0 \gg 1$, beta-binomial distribution is suitable to approximate $\mathbf{p}^{(0)}$. It is obtained from the Binomial(n, p) distribution by allowing the success probability p itself to be a random variable distributed according to a beta distribution with parameters a, b. Probability mass function is defined by following equality:

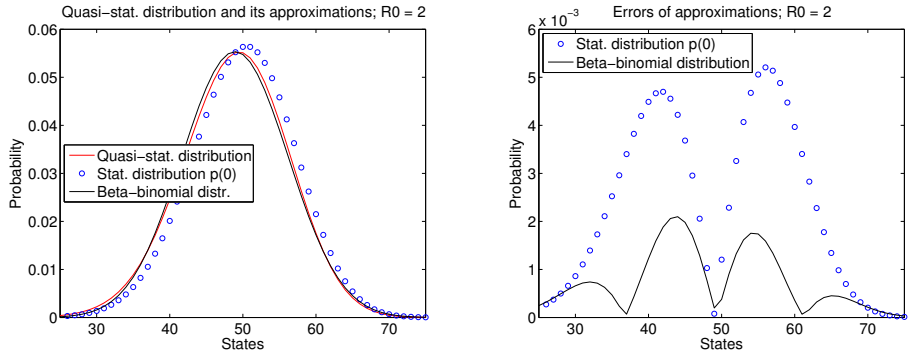
$$p(i) = \frac{B(i+a, N-i+b)}{B(N,b)}\binom{N}{i}$$

for $i = 1, \ldots, N$, where $B(\cdot,\cdot)$ represents the beta function,

$$a = \frac{(N-1)(N-2)R_0^2 - 2N(N-1)R_0 + N^2}{NR_0}$$

and

$$b = N(1 - \frac{1}{R_0}) - 2.$$



(a) Quasi-stationary distribution and its approximations

(b) Absolute value of errors of approximations

Figure 3.5: Quasi-stationary distribution for $R_0 = 2$

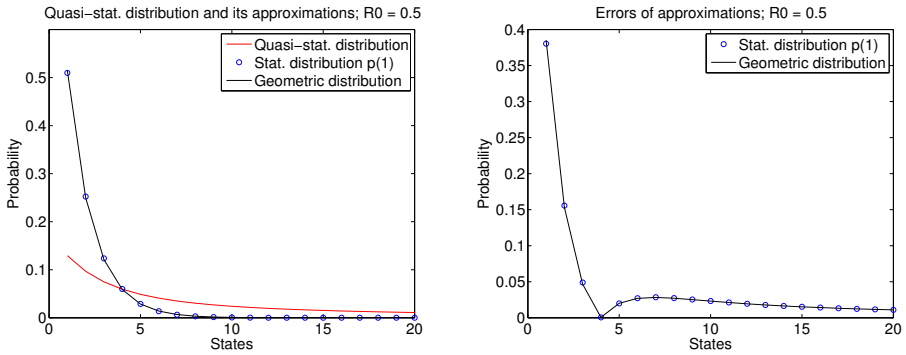
Figures 3.6a and 3.6b show approximations of the quasi-stationary distribution and its error in the sub-critical case. Figures 3.5a and 3.5b show approximations of the quasi-stationary distribution and its error in the supercritical case.

In the sub-critical case, if $R_0 < 1$, $\mathbf{p}^{(1)}$ is approximated by the geometric distribution. Its probability mass function is defined by

$$p(i) = (1-k)^{i-1}k,$$

where

$$k = 4\left(1 - N\left(\frac{1}{R_0} - 1\right) + \sqrt{\left[1 - N\left(\frac{1}{R_0} - 1\right)\right]^2 + \frac{8N}{R_0}}\right)^{-1}.$$



(a) Quasi-stationary distribution and its approximations

(b) Absolute value of errors of approximations

Figure 3.6: Quasi-stationary distribution for $R_0 = 0.5$

Chapter 4

Conclusion

The first one of the primary objectives of this thesis was to understand the stochastic epidemiological modelling and to study the recommended literature. The first chapter begins with the historical overview of the subject and presents a summary of the current state of the art.

Secondly, this paper dealt with the SIS model described by continuous-time and discrete-time Markov chains. Main properties of these processes were examined including expected value, variance, mean persistence time, quasi-stationary distribution and its approximations for both supercritical and sub-critical reproduction number.

Models were implemented in MATLAB. Corresponding codes are included on the attached CD ROM.

Dependence of the epidemic behaviour on chosen parameters was explored for discrete-time and continuous-time Markov chains. Analytical results of models were compared with numerical outcomes. Expected values of both processes are bounded by the solution of deterministic model. The only difference between the models is the transition rate. While the continuous-time Markov chain is determined by given parameters, speed of dynamics of the discrete-time Markov chain is assigned by chosen parameter Δt.

In this paper the properties of deterministic SIS model and stochastic Markov chain model were reviewed only. Further possibility is to add comparison with the SIS model described by stochastic differential equation.

Chapter 5

Appendix

Few examples of MATLAB codes for simulation of Markov chains follow:

Script deterministic.m

```
% DETERMINISTIC shows limit behaviour of the deterministic SIS model
clear all
clc
global N alfa r;
N       = 100;              % population size (constant)
I0      = 95;               % initial condition
p0      = [N-I0 I0];        % initial population
tspan   = [0,100];          % time
time    = (0:.01:.99);      % time sampling
r       = 0.12;             % infection rate
alfa    = 4;                % recovery rate
beta    = r*N-alfa;         % substitution
T       = (0:.01:5);        % sampling
R0      = N*r/alfa;         % reproduction number

% analytic soution
s = N - beta./(r + beta*(beta-p0(2)*r)./(beta*p0(2))*exp(-beta*(T)));

% limit behaviour
lim = zeros(1,tspan(2));
if R0 > 1                   % supercritical case
    for i=1:tspan(2)
        lim(i) = alfa/r;
    end
else                        % sub-critical case
    for i=1:tspan(2)
        lim(i) = N;
    end
end

% solution of the deterministic model
[t,x]=ode45(@SIS,tspan,p0);
```

Function SIS.m

```
function model = SIS(t,x);
% SIS captures the dynamics of deterministic model
global N alfa r;
model=[-r*x(1)*x(2) + alfa*x(2); r*(N-x(2))*x(2) - alfa*x(2)];
end
```

Initiation of the transition matrix and the intensity matrix

```
b1  = beta; b2=-beta/N; d1=b+gamma; d2=0; % parameters corresponding to GBDP
j   = [0:1:N];                           % vector of states
bi  = b1.*j+b2.*j.^2; bi(N+1)=0;         % infection rates
di  = d1.*j+d2.*j.^2;                    % recovery rates
P   = zeros(N+1,N+1);                    % P is the transition matrix
Q   = zeros(N+1,N+1);                    % Q is the intensity matrix
% P filled by diagonals
P   = diag(1-(bi+di)*dtt,0)+diag(bi(1:N)*dtt,1)+diag(di(2:N+1)*dtt,-1);
% Q filled by diagonals
Q   = diag(-bi-di,0)+diag(bi(1:N),1)+diag(di(2:N+1),-1);
```

Function kolmogorov.m

```
function dy = kolmogorov(t,y)
% DY contains matrix Q for FKE
global Q
dy = Q'*y;
end
```

Solution of the forward Kolmogorov differential equations

```
T       = 1000;                                  % simulation-time
Options = odeset('RelTol',1e-8,'AbsTol',1e-8);
[t,Y]   = ode45(@kolmogorov,[0 T],p0,Options);   % solution of FKDE
```

Function update.m

```
function S = update(s,P)
% UPDATE generates particular row of matrix P based on given state
% and calls function initiation
p = P(s,:);             % row of matrix P
S = initiation(p);      % updated state
end
```

Function initiation.m

```
function S=initiation(p)
% INITIATION generates the updated state based on the given row
% of matrix P
x = rand(1);            % random variable from [0,1]
l = length(p);          % length of vector p
a = 0;                  % interval <a,b)
b = 0;
% calculation of interval and comparison of boundaries
for i = 1:l
    a = b;
    b = p(i)+a;
    if x >= a && x < b
       S = i-1;
       break
    end
end
end
```

Function dtmc.m

```
function X = dtmc(P,I0,T)
% DTMC generates one path of the discrete-time Markov chain
X   = zeros(1,T);
X(1)= I0;                           % initial condition
for i = 1:T-1
    if X(i) > 1                     % I(t) != 0
    X(1,i+1)  = update(X(i)+1,P);   % update of state
    end
end
```

Bibliography

ALLEN, L. An Introduction to Stochastic Epidemic Models. In BRAUER, F., DRIESSCHE, P. and WU, J. (Ed.) *Mathematical Epidemiology*, 1945 / *Lecture Notes in Mathematics*. Springer Berlin Heidelberg, 2008. s. 81–130.

ALLEN, L. and ALLEN, E. A comparison of three different stochastic population models with regard to persistence time. *Theoretical Population Biology*. 2003, 64, 4, s. 439 – 449. ISSN 0040-5809.

ANDERSON, D., ENCISO, G. and JOHNSTON, M. Stochastic analysis of biochemical reaction networks with absolute concentration robustness. *Journal of The Royal Society Interface*. 2014, 11, 93. ISSN 1742-5689.

BAILEY, N. The total size of a general stochastic epidemic. *Biometrika*. 1953, 40, 1-2, s. 177–185.

BARTHOLOMEW, D. J. *Stochastic Models*. Elsevier, second edition edition, 2015. ISBN 978-0-08-097087-5.

BRAUER, F. Compartmental Models in Epidemiology. In BRAUER, F., DRIESSCHE, P. and WU, J. (Ed.) *Mathematical Epidemiology*, 1945 / *Lecture Notes in Mathematics*. Springer Berlin Heidelberg, 2008. s. 19–79. ISBN 978-3-540-78910-9.

BRITTON, T. Stochastic epidemic models: a survey. *Mathematical biosciences*. 2010, 225, 1, s. 24–35.

CHALUB, F. and SOUZA, M. Discrete and continuous SIS epidemic models: A unifying approach. *Ecological Complexity*. 2014, 18, s. 83–95.

CLANCY, D. and MENDY, S. Approximating the Quasi-stationary Distribution of the SIS Model for Endemic Infection. *Methodology and Computing in Applied Probability*. 2011, 13, 3, s. 603–618. ISSN 1387-5841.

DARROCH, J. and SENETA, E. On quasi-stationary distributions in absorbing continuous-time finite Markov chains. *Journal of Applied Probability*. 1967, 4, 1, s. 192–196.

HEESTERBEEK, H. and DIETZ, K. The concept of Ro in epidemic theory. *Statistica Neerlandica*. 1996, 50, 1, s. 89–110. ISSN 1467-9574.

HÄGGSTRÖM, O. *Finite Markov Chains and Algorithmic Applications*. Cambridge University Press, 2002. ISBN 0-521-89001-2.

JACQUEZ, J. and SIMON, C. The stochastic {SI} model with recruitment and deaths I. comparison with the closed {SIS} model. *Mathematical Biosciences*. 1993, 117, 1–2, s. 77 – 125. ISSN 0025-5564.

KEELING, M. and ROSS, J. On methods for studying stochastic disease dynamics. *Journal of The Royal Society Interface*. 2008, 5, 19, s. 171–181. ISSN 1742-5689.

KENDALL, D. G. Deterministic and Stochastic Epidemics in Closed Populations. 1956, s. 149–165.

KERMACK, W. O. and MCKENDRICK, A. G. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. 1927, 115, 772, s. 700–721. ISSN 0950-1207.

KRYSCIO, R. J. and LEFÉVRE, C. On the extinction of the SIS stochastic logistic epidemic. *Journal of Applied Probability*. 1989, s. 685–694.

KUNDAN, K. and JOY, K. How to Run a Campaign: Optimal Control of SIS and SIR Information Epidemics. *CoRR*. 2014, abs/1401.6702.

MOLLISON, D. *Epidemic models: their structure and relation to data*. 5. Cambridge University Press, 1995.

NÅSELL, I. On the quasi-stationary distribution of the stochastic logistic epidemic. *Mathematical Biosciences*. 1999, 156, 1–2, s. 21 – 40. ISSN 0025-5564.

PRÁŠKOVÁ, Z. and LACHOUT, P. *Základy náhodných procesů*. Karolinum, 2001. ISBN 80-7184-688-0.

RICHTER-DYN, N. and GOEL, N. S. On the extinction of a colonizing species. *Theoretical Population Biology*. 1972, 3, 4, s. 406–433.

ROSS, R. *The prevention of malaria*. Dutton, 1910.

WEISS, G. and DISHON, M. On the asymptotic behavior of the stochastic and deterministic models of an epidemic. *Mathematical Biosciences*. 1971, 11, 3–4, s. 261 – 265. ISSN 0025-5564.

YAESOUBI, R. and COHEN, T. Generalized Markov models of infectious disease spread: A novel framework for developing dynamic health policies. *European Journal of Operational Research*. 2011, 215, 3, s. 679 – 687. ISSN 0377-2217.