Infectious Disease Modelling 2 (2017) 128–142

Contents lists available at ScienceDirect

Infectious Disease Modelling

journal homepage: www.keaipublishing.com/idm

A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis

Linda J.S. Allen

Department of Mathematics and Statistics, Texas Tech University, Lubbock, TX 79409-1042, USA

ARTICLE INFO

Article history:
Received 29 December 2016
Received in revised form 5 March 2017
Accepted 7 March 2017
Available online 11 March 2017

2000 MSC:
60H10
60J28
92D30

Keywords:
Branching process
Continuous-time Markov chain
Minor outbreak
Stochastic differential equation

ABSTRACT

Some mathematical methods for formulation and numerical simulation of stochastic epidemic models are presented. Specifically, models are formulated for continuous-time Markov chains and stochastic differential equations. Some well-known examples are used for illustration such as an SIR epidemic model and a host-vector malaria model. Analytical methods for approximating the probability of a disease outbreak are also discussed.

1. Introduction

The intent of this primer is to provide a brief introduction to the formulation, numerical simulation, and analysis of stochastic epidemic models for a newcomer to this field. A background in modeling with ordinary differential equations (ODEs) is assumed. The ODE epidemic models serve as a framework for formulating analogous stochastic models and as a source of comparison with the stochastic models. This primer is restricted to two types of stochastic settings, continuous-time Markov chains (CTMCs) and stochastic differential equations (SDEs). Some well-known examples are used for illustration such as an SIR epidemic model and a host-vector malaria model. For additional examples and information on stochastic epidemic models and stochastic modeling in general, consult the textbooks and papers listed in the references, e.g., (E. Allen 2007; Allen 2008, 2010, 2015; E. Allen, Allen, Arciniega, & Greenwood, 2008; Andersson & Britton, 2000; Bailey, 1975; Britton, 2010; Daley & Gani, 1999; Durrett, 1999; Greenwood et al., 2009; Isham et al., 2005; Jagers, 1975; Karlin & Taylor, 1975, 1981).

Stochastic modeling of epidemics is important when the number of infectious individuals is small or when the variability in transmission, recovery, births, deaths, or the environment impacts the epidemic outcome. The variability associated with individual dynamics such as transmission, recovery, births or deaths is often referred to as demographic variability. The variability associated with the environment such as conditions related to terrestrial or aquatic settings is referred to as environmental variability. Environmental variability is especially important in modeling zoonotic infectious diseases, vector-

E-mail address: linda.j.allen@ttu.edu.
Peer review under responsibility of KeAi Communications Co., Ltd.

http://dx.doi.org/10.1016/j.idm.2017.03.001

borne diseases, and waterborne diseases (e.g., Ebola, avian influenza, malaria, and cholera) (Altizer, Ostfeld, Johnston, Kutz, & Harvell, 2013; Jutla et al., 2013; Wu, Lu, Zhou, Chen, & Xu, 2016). In this primer, the emphasis is on demographic variability.

In CTMCs and SDEs, the time variable is continuous, $t\in[0,\infty)$, but the state variables are either discrete (CTMC) or continuous (SDEs). In the following sections, these two stochastic processes are formulated for the well-known SIR (Susceptible-Infectious-Recovered) epidemic model and the Ross malaria host-vector model. The Gillespie algorithm and the Euler-Maruyama numerical method are described for the two types of stochastic processes. In addition, some analytical methods from branching processes that are related to the CTMC models are used to approximate the probability of an outbreak. In the last section, some stochastic methods for modeling environmental variability are presented.

2. SIR deterministic epidemic model

In the SIR deterministic model, $S(t)$, $I(t)$, and $R(t)$ are the number of susceptible, infectious, and recovered individuals, respectively. In the simplest model, there are no births and deaths, only infection and recovery:

$$\begin{aligned}\frac{dS}{dt}&=-\beta I\frac{S}{N}\\ \frac{dI}{dt}&=\beta I\frac{S}{N}-\gamma I\\ \frac{dR}{dt}&=\gamma I,\end{aligned}\tag{1}$$

where the total population size is constant, $S(t)+I(t)+R(t)=N$. The disease-free equilibrium is $S=N$ and $I=R=0$. The basic reproduction number $\mathcal{R}_0=\beta/\gamma$ which is equal to the ratio of the transmission rate β and the recovery rate γ, determines the epidemic outcome when $S(0)\approx N$. If $I(0)>0$ and $\mathcal{R}_0S(0)/N>1$, then the number of infectious individuals increases, an outbreak, and if $\mathcal{R}_0S(0)/N<1$, the number of infectious individuals decrease. As $R(t)=N-S(t)-I(t)$, system (1) can be simplified to two equations for $S(t)$ and $I(t)$.

The stochastic formulation of the CTMC and SDE models requires defining two random variables for S and I whose dynamics depend on the probabilities of the two events: infection and recovery. For simplicity, the same notation is used in the stochastic and the deterministic formulations.

3. SIR continuous time Markov chain

3.1. Formulation

The discrete random variables for the SIR CTMC model satisfy

$$S(t),I(t)\in\{0,1,2,\ldots,N\},$$

where $t\in[0,\infty)$. The lower case s and i denote the values of the discrete random variables from the set $\{0,1,2,\ldots,N\}$. The transition probabilities associated with the stochastic process are defined for a small period of time $\Delta t>0$:

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

The transition probabilities depend on the time between events Δt but not on the specific time t, a time-homogeneous process. In addition, given the current state of the process at time t, the future state of the process at time $t+\Delta t$, for any $\Delta t>0$, does not depend on times prior to t, known as the Markov property. For comparison purposes, the transition probabilities are defined in terms of the rates in the SIR ODE model:

$$p_{(s,i),(s+k,i+j)}(\Delta t)=\begin{cases}\beta i\frac{s}{N}\Delta t+o(\Delta t), & (k,j)=(-1,+1)\\ \gamma i\Delta t+o(\Delta t), & (k,j)=(0,-1)\\ 1-\left(\beta i\frac{s}{N}+\gamma i\right)\Delta t & \\ \quad+o(\Delta t), & (k,j)=(0,0)\\ o(\Delta t), & \textit{otherwise}.\end{cases}\tag{2}$$

Summarized in Table 1 are the changes, $\Delta S(t)=S(t+\Delta t)-S(t)$ and $\Delta I(t)=I(t+\Delta t)-I(t)$, associated with the two events, infection and recovery.

Given $S(0)=N-i$ and $I(0)=i>0$, the epidemic ends at time t, when $I(t)=0$. The states (S,I), where $I=0$ are referred to as absorbing states; the epidemic stops when an absorbing state is reached. The absorbing states are the states (s,i) with $i=0$.