

# Simple Stochastic Models for Epidemics

Helen J. Wearing

July 23, 2014

Before we think about stochastic models that are analogous to the continuous-time SIR model with demography, we will develop some intuition about the key differences between stochastic and deterministic models by starting out with the same framework we used on day 1.

## Stochastic Reed-Frost model

Let's think about a simple epidemic that evolves in discrete time-steps (or generations). We need to describe the number of new infections that occur in each generation, and this should be dependent on the number of susceptible and infected individuals in the previous generation. Instead of this being exactly determined (as we saw previously), we're going to consider that this process involves some element of chance. Specifically, we're going to consider each susceptible in turn, flip a weighted coin, if the coin comes up heads then the susceptible individual becomes infected, otherwise they remain susceptible.

Mathematically, in each generation, new infections are binomially distributed with the number of trials equal to the number of susceptibles,  $S_t$ , and the probability of success (infection) equal to  $P_t$ . In probability notation:

$$I_{t+1} \sim \text{binom}(S_t, P_t)$$

A binomial random variable can be thought of as the number of independent "successes" (here infections) in a sequence of weighted coin tosses (here the weight is  $P_t$ , and we flip a coin for every susceptible individual). But what should  $P_t$ , the probability of infection, depend on? Let's consider a single susceptible individual and their risk of infection in a particular generation. Let's denote  $q$  as the probability that a susceptible *does not* get the disease from a given infectious individual. If there are  $I_t$  infectious individuals, and they all behave independently, then the probability

the susceptible does not get the disease is  $q^{I_t}$ . Therefore, the probability that the susceptible gets infected is  $P_t = 1 - q^{I_t}$ . So the full model is

$$\begin{aligned} I_{t+1} &\sim \text{binom}(S_t, 1 - q^{I_t}) \\ S_{t+1} &= S_t - I_{t+1} \\ R_{t+1} &= R_t + I_t \end{aligned}$$

where the number of susceptible individuals is depleted by new infections and the number of recovered individuals increases by the number who were infectious during the previous generation (individuals are infected for a single generation). This model is the stochastic Reed-Frost model, more generally a *chain binomial* model, and is part of a large class of stochastic models known as *Markov chain* models. A Markov chain is defined as a stochastic process with the property that the future state of the system is dependent only on the present state of the system and conditionally independent of all past states. This is known as the “memoryless” or Markov property.

The parameter  $q$  can be rewritten as  $e^{-A}$ , where the parameter  $A$  represents a rate of contracting the disease per infective (whereas  $q$  is a probability), and can be interpreted in the same way as the  $A$  in the models that you’ve already seen. In addition,  $e^{-AI_t}$  can be thought of as the zero term of a Poisson distribution with mean  $AI_t$ .

**\*Try This\***

1. Fix the values of  $S_0 = 2000$ ,  $q = 0.999$  and  $I_0 = 2$ , and simulate the model 10 times for 20 generations. What do you notice?
2. Systematically, vary  $S_0$ ,  $q$  and  $I_0$ . One way of summarizing the model output is, for each parameter set, to construct a histogram of the final size of the epidemic. What patterns do you uncover?
3.  $R_0$  is defined as the average number of secondary infections generated by one infection in a completely susceptible population, what is  $R_0$  for this model? Does an epidemic always take off if  $R_0 > 1$ ? Note that for a binomial random variable  $X \sim \text{binom}(n, p)$ ,  $E(X) = np$ .

Summary of the key differences between this stochastic model and previous deterministic models:

- Chance events can lead to extinction before an epidemic occurs, even if  $R_0 > 1$ .

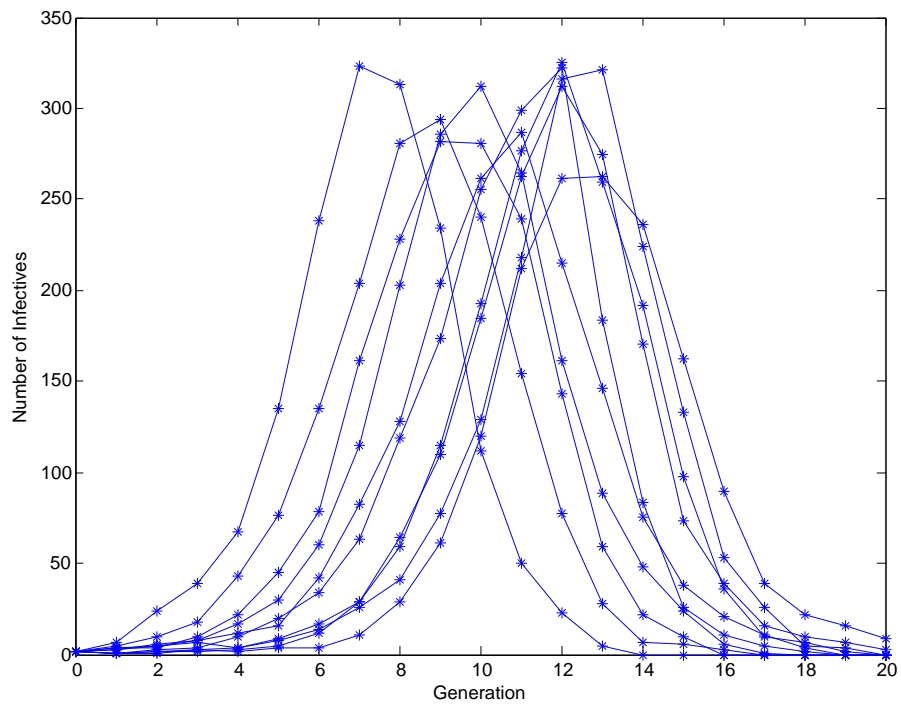


Figure 1: 10 simulations of the Reed-Frost model for parameter values given in (1.) above.

- The states  $(S, I, R)$  are integer-valued rather than real-valued variables, i.e. the state-space is discrete rather than continuous.

This is a stochastic version of the discrete-time deterministic model we saw on Tuesday. In fact, if we calculate the average value of  $I_{t+1}$

$$E(I_{t+1}) = S_t(1 - \exp(-AI_t))$$

this is just the deterministic equation for  $I_{t+1}$ .

So, how do we construct a stochastic model that is related to the continuous-time *SIR* model? There's actually more than one answer to this question - it depends on certain assumptions that we make. What we'll discuss here is how to think about the processes that contribute to the *SIR* model in an alternative way - as events that lead to discrete changes in the state of the system.

## Event-driven methods and Gillespie's algorithm

The state of the system is defined by the *integer* number of individuals in each subpopulation and changes discretely whenever an event (such as a birth, death or transition from one class to another) occurs. For the *SIR* model (with demography), we need to consider the six events that can occur and how the numbers in each class change as a result of each event:

1. Births occur at the rate  $\mu N$ , and the result is  $S \rightarrow S + 1$ .
2. Transmission occurs at the rate  $\beta \frac{I}{N} S$ , and the result is  $S \rightarrow S - 1, I \rightarrow I + 1$ .
3. Recovery occurs at the rate  $\gamma I$ , and the result is  $I \rightarrow I - 1, R \rightarrow R + 1$ .
4. Deaths of  $S$  occur at the rate  $\mu S$ , and the result is  $S \rightarrow S - 1$ .
5. Deaths of  $I$  occur at the rate  $\mu I$ , and the result is  $I \rightarrow I - 1$ .
6. Deaths of  $R$  occur at the rate  $\mu R$ , and the result is  $R \rightarrow R - 1$ .

There are different ways of implementing this framework, but we'll consider Gillespie's Direct Method (Gillespie 1977).

## Gillespie's Direct Method

For the system in a given state, Gillespie's direct method asks two questions:

1. When does the next event occur? The time to the next event ( $\tau$ ) is exponentially distributed, with rate equal to the sum of the rates over all possible events. The probability density function is given by

$$f(\tau) = \left( \sum_i a_i \right) \exp(-\tau \sum_i a_i). \quad (1)$$

2. Which event occurs next? We convert event rates into probabilities, and randomly select one of these events according to

$$P(\text{Event} = v) = a_v / \sum_i a_i \quad (2)$$

where the  $a_i$  are the event rates listed above. (The numbers in each class are then updated according to which event is selected.)

With these distributions, the algorithm is implemented as follows:

1. Set initial population numbers. Set  $t \mapsto 0$ .
2. Calculate the  $a_i$  for all  $i$ .
3. Choose  $\tau$  from an exponential distribution with parameter  $\sum_i a_i$  as in (1).
4. Choose the event  $v$  according to the distribution in (2).
5. Change the number of individuals to reflect the event,  $v$ . Set  $t \mapsto t + \tau$ .
6. Go to step 2.

### **\*Try This\***

1. Using the code `stochasticSIR.m`, reproduce a figure similar to Figure 2. Why can't you reproduce the exact same figure? Experiment with the initial number of infectives,  $I(0)$ . What happens as you increase the initial size of the total population? Both dynamically and computationally.
2. Now initialize the stochastic model with the endemic equilibrium values from the deterministic model. How does this affect the solutions?

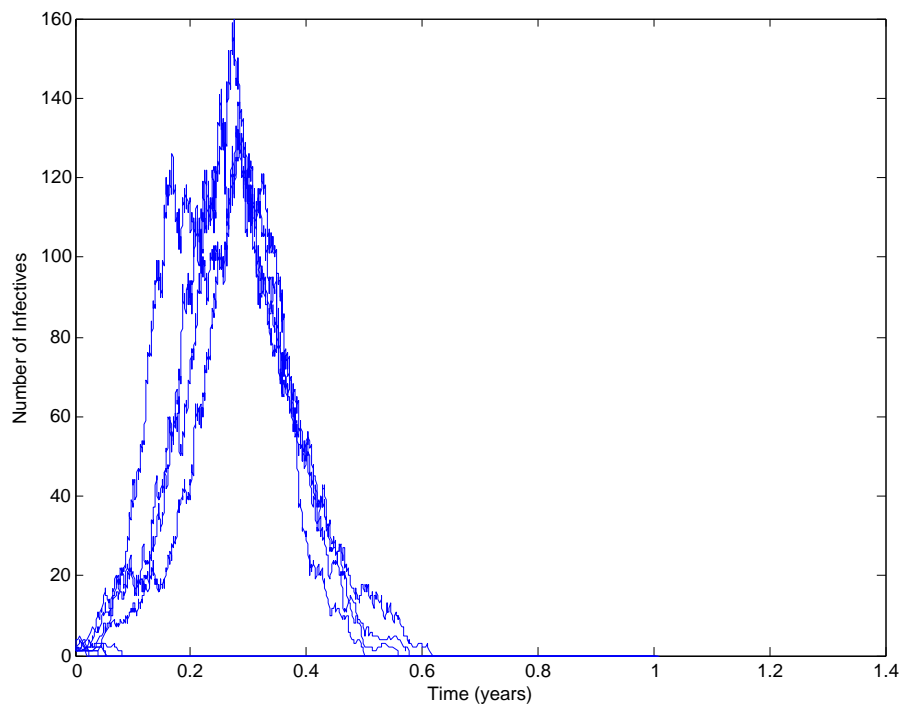


Figure 2: 10 simulations of the Stochastic SIR model with  $R_0 = 10$ ,  $\gamma = 365/10$  year<sup>-1</sup>,  $\mu = 1/70$  year<sup>-1</sup> and  $\beta = R_0(\gamma + \mu)$  and  $S(0) = 1500$ ,  $I(0) = 1$ , and  $R(0) = 8499$ .

This algorithm simulates stochastic realizations of the exact process described by what's known as the master equation. The master equation describes the evolution of all possible states of the system (probability of being in state  $(S, I, R)$  at time  $t$ ) but usually leads to a computationally intractable system of differential equations:

### Master Equation

Let  $p_{SIR}(t)$  be the probability of being in state  $(S, I, R)$  at time  $t$ , and  $N = S + I + R$ . Then the following equation describes how this probability distribution evolves over time:

$$\begin{aligned} \frac{dp_{SIR}(t)}{dt} = & p_{S-1,I,R}[\mu(N-1)] + p_{S+1,I,R}[\mu(S+1)] + p_{S+1,I-1,R}[\beta \frac{(I-1)}{N}(S+1)] \\ & + p_{S,I+1,R-1}[\gamma(I+1)] + p_{S,I+1,R}[\mu(I+1)] + p_{S,I,R+1}[\mu(R+1)] \\ & - p_{S,I,R}[\mu N + \mu S + \beta \frac{I}{N} S + \gamma I + \mu I + \mu R]. \end{aligned}$$