




12-27-2020

## Physical Models of Living Systems 2nd ed, new chapter: Demographic Variation in Epidemic Spread

Philip C. Nelson

University of Pennsylvania, [nelson@physics.upenn.edu](mailto:nelson@physics.upenn.edu)

Follow this and additional works at: [https://repository.upenn.edu/physics\\_papers](https://repository.upenn.edu/physics_papers)

 Part of the [Biological and Chemical Physics Commons](#), [Biophysics Commons](#), [Other Immunology and Infectious Disease Commons](#), and the [Virus Diseases Commons](#)

---

### Recommended Citation

Nelson, P. C. (2020). Physical Models of Living Systems 2nd ed, new chapter: Demographic Variation in Epidemic Spread. *Not yet published*, Retrieved from [https://repository.upenn.edu/physics\\_papers/661](https://repository.upenn.edu/physics_papers/661)

This work was supported by the Center for Engineering MechanoBiology (CEMB), an NSF Science and Technology Center, under grant agreement CMMI-15-48571.

This chapter extends the first edition of Physical Models of Living Systems (WH Freeman 2015). This preliminary version is made freely available as-is in the hope that it will be useful.

This paper is posted at ScholarlyCommons. [https://repository.upenn.edu/physics\\_papers/661](https://repository.upenn.edu/physics_papers/661)  
For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

## Physical Models of Living Systems 2nd ed, new chapter: Demographic Variation in Epidemic Spread

### Abstract

This chapter extends the first edition of Physical Models of Living Systems (WH Freeman 2015). This preliminary version is made freely available as-is in the hope that it will be useful.

### Keywords

epidemiology, physical modeling, dynamical systems, stochastic simulation, Gillespie algorithm, superspreaders

### Disciplines

Biological and Chemical Physics | Biophysics | Other Immunology and Infectious Disease | Physical Sciences and Mathematics | Physics | Virus Diseases

### Comments

This work was supported by the Center for Engineering MechanoBiology (CEMB), an NSF Science and Technology Center, under grant agreement CMMI-15-48571.

This chapter extends the first edition of Physical Models of Living Systems (WH Freeman 2015). This preliminary version is made freely available as-is in the hope that it will be useful.

# CHAPTER 14

## Demographic Variation in Epidemic Spread

*This antique discipline, tenderly severe,  
Renews belief in love yet masters feeling,  
Asking of us a grace in what we bear.  
Form is the ultimate gift that love can offer—  
The vital union of necessity  
With all that we desire, all that we suffer.*  
— Adrienne Rich

### 14.1 SIGNPOST: SUPERSPREADERS

Chapter 10 introduced computer simulation of general “chemical reaction” type models, including their inherently stochastic character. However, we then observed that the predictions of such models sometimes follow those of simpler models based on the continuous, deterministic approximation.<sup>1</sup> Accordingly, Chapters 11–13 then explored dynamical systems in that approximation. Interesting phenomena such as bistability and oscillation appeared when we introduced realistic nonlinearity into our rate equations.

Some of the nonlinear models we studied were related to disease outbreaks (Section 12.3). One of them (the SIR model) displayed a line of fixed points, any of which was a possible endpoint of an outbreak. Which of those endpoints is chosen for given initial conditions, and hence the severity of the outbreak, depended sensitively on a system parameter called the basic reproduction number,  $R_0$ .

However, certainly many living systems of interest to us are not really huge collections of identical “molecules.” In this last Part of the book, we will see that qualitatively new phenomena can appear when we combine nonlinearity with discreteness (and its associated stochasticity). As always, living organisms have evolved to cope with, or even benefit from, such effects—sometimes to the detriment of other host organisms.

This chapter’s Focus Question is:

*Biological question:* Why do some outbreaks of a communicable illness spread explosively, while others, in similar communities, fizzle after the first few cases?

*Physical idea:* A tiny subpopulation of *superspreader* individuals can introduce giant variations in the course of an epidemic.

### 14.2 STOCHASTIC SIR MODEL

#### 14.2.1 Some outbreaks fizzle out

Figure 12.4b depicts one traditional model of disease progression as a network diagram. After rescaling time, we expressed this physical model as a pair of ordinary differential

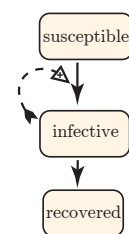


Fig. 12.4b, p. 306

<sup>1</sup>See Idea 10.8 (page 246).

equations:

$$\frac{ds}{d\bar{t}} = -R_0 i s; \quad \frac{di}{d\bar{t}} = (R_0 s - 1)i; \quad i + s \leq 1. \quad [12.7, \text{page 307}]$$

Here  $s$  is the fraction of the population in the susceptible state;  $i$  is the infective fraction;  $\bar{t}$  is dimensionless rescaled time; and  $R_0$  (the basic reproduction number) is a constant. We obtained these equations by making some assumptions:

- Total population size is fixed to some value  $N_{\text{tot}}$  (no births, deaths, arrivals, or departures).
- Susceptible individuals make random encounters with all individuals, and each has probability  $R_0 i$  per unit  $\bar{t}$  to become infected (and hence infective), where  $R_0$  is a constant.
- Each infective individual has probability 1 per unit  $\bar{t}$  to recover permanently.
- Each subpopulation is large enough to use the continuous, deterministic approximation.

Each of the assumptions just listed can be criticized as an oversimplification for any particular illness and population. But certainly the last one is *never* valid at the start of an outbreak, where just one or a handful of infectives are introduced into an otherwise fully  $S$  population. It seems reasonable that some outbreaks will fizzle out, because there is always a chance that the initial infective(s) will simply recover before they have a chance to create any secondary infections. And yet, that behavior is not what we found in Chapter 12. Because we took populations to be continuous variables, not integers, we can consider arbitrarily low initial values of the infective population  $i(0)$ , that is, initial states that approach the lower right extreme on the phase portraits (Figure 12.5a). In that limit, our trajectory's endpoint approaches some definite point on the horizontal axis of the figure—not necessarily  $s(\infty) \approx 1$ , which would correspond to a failed outbreak. Indeed, because the initial growth is exponential for small  $i$ , reducing  $i(0)$  by a factor of  $1/e$  just postpones the peak by an  $e$ -folding time, without changing its eventual height (maximum of  $i$ ) nor its endpoint ( $s(\infty)$ ).

Conversely, the *end* of an outbreak also involves small numbers; unlike in the continuous, deterministic approximation, a real outbreak can end completely as the number of infectives makes a final drop from one to zero cases. In short, the discrete, random character of infection leads to new phenomena collectively called **demographic variation**.

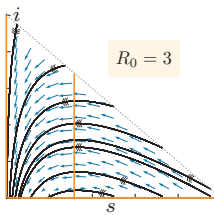


Fig. 12.5a, p. 308

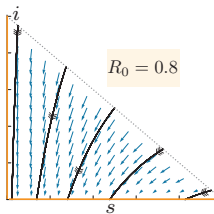


Fig. 12.5b, p. 308

### 14.2.2 Infections attributable to a single individual follow a Geometric distribution in the SIR model

To model demographic variation, we now translate the verbal description in the preceding section into a stochastic simulation, via the Gillespie direct algorithm.<sup>2</sup> Suppose that at some moment there are  $S$  susceptibles,  $I$  infectives, and hence  $N_{\text{tot}} - S - I$  recovered individuals. The total probability per  $\bar{t}$  for any transition to occur is then  $\beta_{\text{tot}} = IS(R_0/N_{\text{tot}}) + I$ , so we draw a waiting time from the appropriate Exponential distribution. We then decide which class of event happened at the chosen time: The probability that one of the infectives recovers is  $I/\beta_{\text{tot}}$ ; otherwise, a susceptible gets infected. We make the appropriate Bernoulli trial, then implement whatever happened by updating  $S$  and  $I$  and repeat.

Figure 14.1a shows the resulting simulated time courses for three runs, each of which starts with  $I(0) = 3$  sick individuals in a population of the same size as the one shown in Figure 12.6. Comparing the two simulated results to each other, and to the data for the

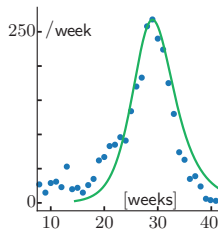


Fig. 12.6, p. 308

<sup>2</sup>Section 10.3.3 (page 244) introduced this approach.

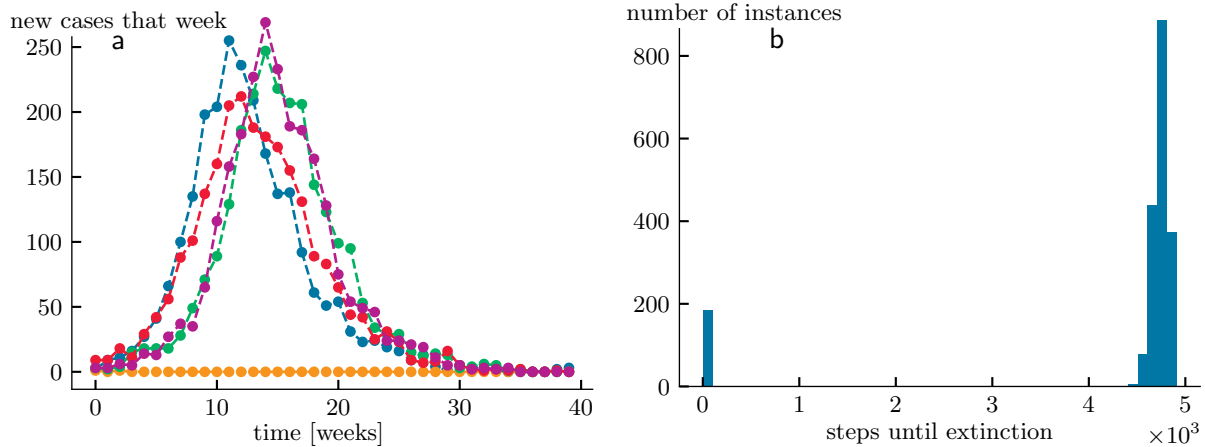


Figure 14.1: [Numerical simulations.] **Stochastic SIR model.** (a) Five typical time courses for the same model as in Figure 12.6, but this time simulated via Gillespie’s direct algorithm. Thus,  $R_0 = 2.2$  in a community of size 2816 starting with three infected individuals. One of these outbreaks went extinct almost immediately (orange). (b) Durations of 2000 simulated outbreaks in the same model.

measles outbreak, shows that in the improved treatment:

- Many outbreaks indeed fizzle out when we account for discreteness of populations (Figure 14.1b).
- There is some randomness in the weekly case load when we account for stochasticity, and it is roughly of the same magnitude as was observed in the real data.
- There is also some randomness in timing.
- But if an outbreak exceeds a threshold, its time course and overall severity are then roughly as predicted by the continuous, deterministic approximation.

To gain more insight into the nature of this model, we can ask our simulation to report a quantity that is difficult to measure in the clinic: For each infected individual, we ask, how many further infections are attributable to *that individual*? Of course the answer is a random variable, so as usual we instead ask for the *distribution* of the attributed infection count  $\ell_{\text{inf}}$ . Near the start of an outbreak, the susceptible population has not declined significantly from its initial value of nearly 100%. An infective individual thus has a constant probability per rescaled time to encounter and infect susceptibles ( $R_0 S$  is approximately constant), so for the duration of the infective period, that individual generates new infectives in a Poisson process. If the duration of each infection were fixed, then Idea 9.6 (page 217) would then imply that the total number of new infectives attributable to this individual would be Poisson distributed.

However, in this model the duration of each infection is *itself* a random variable; indeed, it is Exponentially distributed because we assumed constant probability per time to recover. So really,  $\ell_{\text{inf}}$  follows a *mixture* of Poisson distributions.<sup>3</sup> You found in Problem 5.23 that the net effect is that the quantity  $1 + \ell_{\text{inf}}$  is predicted to be Geometrically distributed. Extracting this information from our simulation indeed confirms that expectation (Figure 14.2b).

<sup>3</sup>Section 5.2.5 (page 106) introduced this notion.

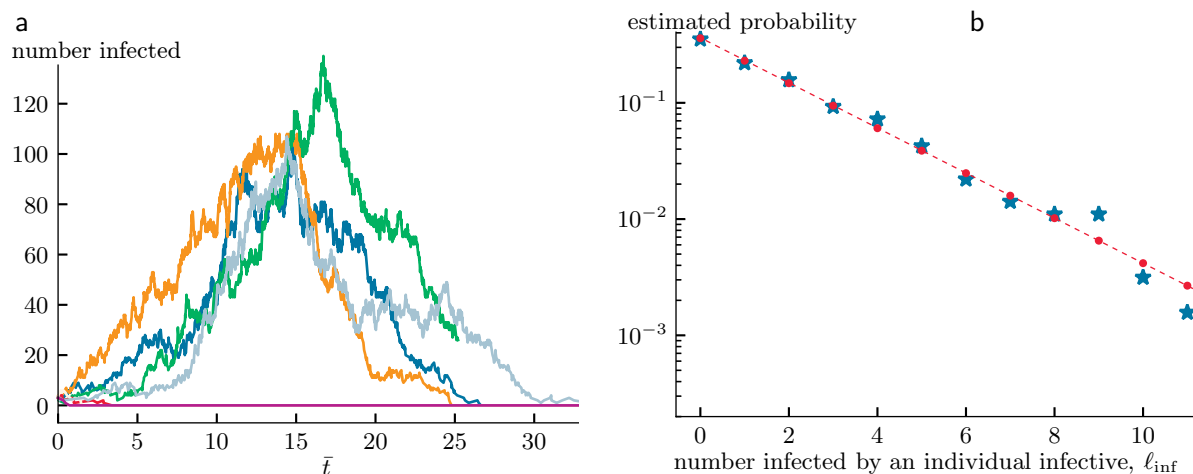


Figure 14.2: [Numerical simulations.] **Stochastic SIR model with  $R_0 = 1.3$ .** (a) Seven representative time courses, of which three went extinct almost immediately. The total population is again 2816, and again initially three individuals are infected. (b) Stars: Semilog histogram of the number of infections attributable to a particular infective early in the outbreak, based on 2000 simulated outbreaks. The sample mean is 1.8; the estimated variance is 4.3, in rough agreement with a Geometric distribution. Dots: A Geometric distribution with the same expectation, for comparison.

## 14.3 SUPERSPREADING VIA OVERDISPERSION

### 14.3.1 For some illnesses, infectivity is heterogeneous

The preceding section outlined a way to improve the realism of the SIR model. But some pandemics, for example SARS COV-1 in the 2000s, were found to be poorly described with models of this type. Certainly there are many idealizations implicit in our model. For example, some illnesses have a latent “exposed” state between contact with an infective and the onset of infectiousness. Individuals also have varying social networks, varying geographic mobility, and a host of other confounding factors. Introducing any such effects into a model also introduces new unknown parameters that must be fit to data (reducing predictive power) or else measured somehow (not always possible). Of all the many improvements we might entertain, which one should we try *first*?

One big clue came when contact tracing data became available for SARS COV-1. Figure 14.3a shows that the distribution of new infections attributable to one individual ( $\ell_{\text{inf}}$ ) looks *nothing like* a Geometric distribution.<sup>4</sup> Instead,  $\mathcal{P}_{\ell_{\text{inf}}}$  has a long tail: A small fraction of the population are **superspreaders**. One way to quantify this statement is to note that<sup>5</sup>

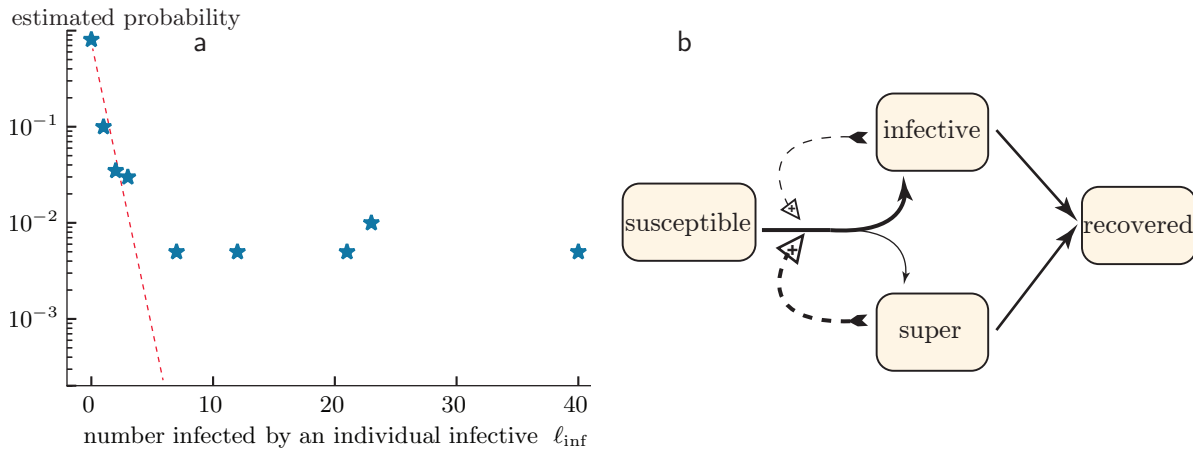
$$\text{var} = \langle \ell_{\text{inf}} \rangle (1 + \langle \ell_{\text{inf}} \rangle).$$

For an outbreak of SARS COV-1 in Singapore, the data in Figure 14.3a gave  $\langle \ell_{\text{inf}} \rangle \approx 0.9$  with estimated variance  $\approx 16$ , so the distribution was far from being Geometric.<sup>6</sup> Later,

<sup>4</sup>It looks even less like a Poisson distribution, which we would predict if each recovery took exactly the same time.

<sup>5</sup>Page 57 expresses expectation and variance of any Geometric distribution in terms of a common parameter  $\xi$ . Eliminating  $\xi$ , and recalling that  $\ell_{\text{inf}} + 1$  is Geometrically distributed, establishes this relationship.

<sup>6</sup>These numbers are likely underestimates due to incomplete contact tracing. A Poisson distribution would have to have  $\text{var} = \langle \ell_{\text{inf}} \rangle$  (Chapter 4), which is even further from observed values.



**Figure 14.3: Evidence for variation in individual reproductive number.** (a) [Public health data.] Stars: Transmission data from the SARS COV-1 outbreak in Singapore in 2003. This semilog plot shows the observed frequency of the number of individuals infected by each case. [Data from Leo et al., 2003; see also Dataset 22.] The line shows an imagined Geometric distribution for comparison. (b) [Network diagram.] Model accounting for superspreaders. A small fraction of new infections are highly infective.

in 2020 the SARS COV-2 virus was also found to be characterized by an “overdispersed” distribution of  $\ell_{inf}$ .

What is the origin of this distribution’s long tail?

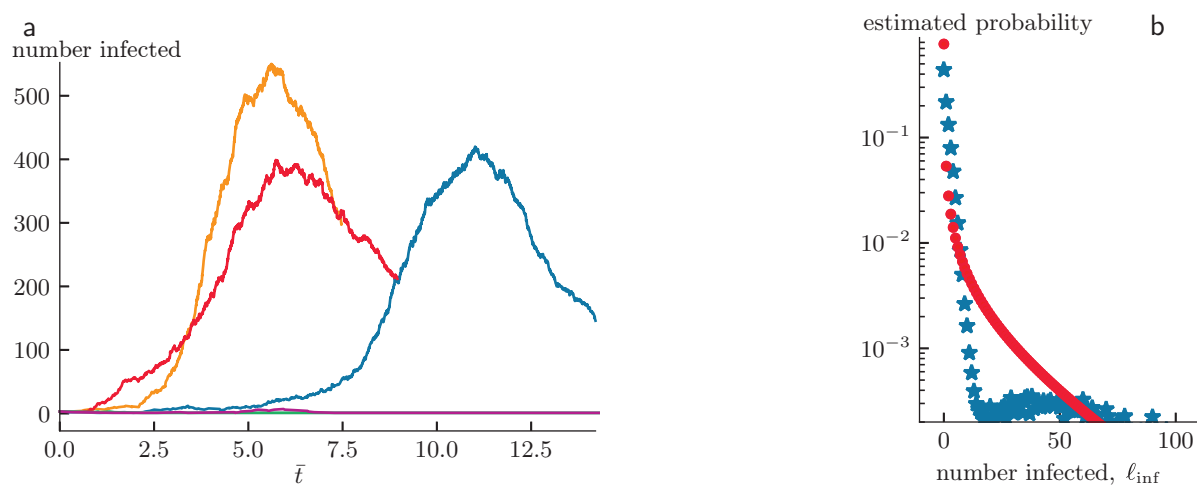
- *Intrinsic variability of infectivity:* Certainly some infected persons sneeze more than others (for example, due to another co-infection). Also, infectivity may be dependent on an individual’s age, general health, and so on.
- *Variability of contacts:* Some choose to go to crowded places, are required to do so for their work, or are confined to nursing homes or prisons.
- *Variability of behavior:* Some attend events where they and others shout or sing. On the other hand, some persons use better public health practices, knowing that they may be infective but not yet (or not ever) symptomatic.

It would be a nightmare to try to model every one of these population heterogeneities mathematically. But we can already reap important insights just by exploring the simplest possible realization of the idea of superspreaders.

### 14.3.2 Even a tiny minority of superspreaders has a big impact

Figure 14.3b shows an extension of the SIR model with two classes of infectives. Each infection event randomly assigns an individual, usually to the regular bin (small  $R_0$ ) but occasionally to the superspreader bin (big  $R_0$ ). For illustration we assume that *just 2% of cases are superspreaders*, but their value of  $R_0$  is 20 times greater than the other group’s.

Figure 14.4a shows that the tiny minority of superspreaders has profound effects on the courses of outbreaks, which are much more severe, and more variable, than in the original SIR model.



**Figure 14.4:** [Computer simulations.] **Effect of superspreaders.** Results from the model represented by Figure 14.3b. All parameters are the same as in Figure 14.2 except that two percent had  $R_0 = 25$ , as suggested by Figure 14.3a. (a) Seven representative time courses, of which 4 went extinct almost immediately. The tiny minority of superspreaders greatly increased both the mean size and the variability of successful outbreaks. (b) *Stars*: Semilog histogram of the number of infections attributable to a particular infective early in the outbreak. The Geometric distribution arising in the SIR model (Figure 14.2b) has now been replaced by something with a much longer tail. The sample mean is 2.6; the estimated variance is 100. Epidemiologists sometimes model the corresponding data from a real population by using a family of distributions called Negative Binomial; the one shown here as *dots* was chosen to have the same expectation and variance as the simulation data.

## THE BIG PICTURE

We have found that because outbreaks always begin with just one or a few infective individuals, the discrete, stochastic character of transmission has a large effect on outbreak dynamics. Thus, a community that is lucky to get only a mild outbreak in the first instance must not become complacent, imagining themselves to be somehow protected: Always *some* outbreaks fizzle, but any such instance is just as likely to be followed by a severe outbreak on a later introduction as in any other community.

There are many ways to improve the realism of the SIR model, but we focused on just one: the well documented fact that some illnesses have *superspreader* individuals. The implications are profound. Although Figure 14.4a is frightening, the fact is that such time courses can be replaced by the milder ones in Figure 14.2 by promptly identifying and quarantining just a few percent of the infected population. For example, backward contact tracing seeks to identify contacts of each sick individual who may have been the source of that person's infection. When multiple backward trails point to the same person, that person may be a superspreader.

## KEY FORMULAS

[Not ready yet.]



## FURTHER READING

*Semipopular:*

Kucharski, 2020.

Tufekci, 2020=

[www.theatlantic.com/health/archive/2020/09/k-overlooked-variable-driving-pandemic/616548/](http://www.theatlantic.com/health/archive/2020/09/k-overlooked-variable-driving-pandemic/616548/) .

*Intermediate:*

Allen, 2011; Rock et al., 2014; Andersson & Britton, 2000.

*Technical:*

Superspreaders: Lloyd-Smith et al., 2005; Laxminarayan et al., 2020; Adam et al., 2020.

Maximum likelihood estimation of overdispersion: Lloyd-Smith, 2007.

Epidemic modeling on a network: Miller & Ting, 2019.

PROBLEMS
----------

14.1 .  
[Not ready yet.]


14.2 .  
[Not ready yet.]

14.3 .  
[Not ready yet.]

14.4 *Peak rate distribution*

Implement the stochastic SIR model via Gillespie's direct algorithm, using parameter values  $R_0 = 1.5$ ,  $\gamma = 1/(20 \text{ day})$ . One key descriptor of an outbreak is the peak rate  $X$  of new cases, that is, the highest value in any one week of  $S(t) - S(t - 1 \text{ week})$ . This quantity determines whether health-care facilities will become overloaded. It is a random variable, so its *distribution*  $\mathcal{P}(X)$  is of interest.

Run your simulation many times, always starting with 2 infective individuals, zero recovered, and 2814 susceptible. There will be many events, occurring at an ascending sequence of randomly chosen times  $t_i$ . Simulate enough such steps to always catch the peak of each simulated outbreak.

- a. Make graphs showing the number of infected individuals  $I_i$  versus  $t_i$  for three simulated outbreaks.
- b. Extend your code to do more after each run is finished: For each integer  $j$ , have your computer find the step number  $i_j$  at which  $t_{i_j}/(7 \text{ day})$  passes  $j$ . Then find the differences  $Y_j = S_{i_{j-1}} - S_{i_j}$ , which give the number of new infections in week  $j$ . The largest of these  $Y_j$  values is an instance of  $X$ , so save it. Then start the next new run. After you have enough runs, estimate  $\mathcal{P}(X)$ .
- c.  Create an animation of a 3-bar chart, showing S, I, and R populations versus (actual) time for a single representative run in the first model (SIR).


14.5 *Peak rate distribution, II*

First work Problem 14.4, then modify it as follows:

- Implement the model outlined in Section 14.3.2, with parameter values

$$R_0 = 1.5 \text{ for } 98\% \text{ of new infections or } 25 \text{ for the other } 2\% .$$

Again use  $\gamma = 1/(20 \text{ day})$ , and always start with 2 infective individuals (neither of whom is a superspreader), zero recovered, and 2814 susceptible. As in the preceding question, estimate the distribution of peak caseloads  $\mathcal{P}(X)$ . Compare/contrast with the corresponding result of the SIR model (preceding question).

-  Create an animation of a 4-bar chart, showing S, I1, I2, and R populations versus (real) time for a single representative run in the second model. If you see qualitatively new behavior in this presentation, describe it.

14.6 [Not ready yet.]... Dataset 22