The SIR and SEIR Epidemiological Models Revisited

John P. Maassen †

Abstract

We review and assess the classic SIR and SEIR epidemiological models regarding possible applications to the COVID-19 pandemic. In spite of numerous more complicated models, we show how the qualitative features of the solution to the SIR and SEIR models continue to provide valuable public health insights in some scenarios. Using estimated COVID-19 data as of this date, the SEIR model shows that if it were possible to reduce R_0 from 2.5 to 1.25 through social distancing and other measures, the maximum fraction of the population that would become infected at any particular time would drop from 17% to 4%, *provided that all of the model assumptions are satisfied*. Finally, we compare the classic SIR model with a recent stochastic model with favorable results. Since this comparison underscores the importance of underlying connectivity assumptions, we conclude with Monte-Carlo simulations with specific connectivity that reproduce the classical SIR model with standard incidence.

Keywords: epidemiological model, basic reproduction number, SIR, SEIR, COVID-19, stochastic model, Monte-Carlo simulation

Introduction

One way to model the spread of infectious diseases is to subdivide the affected population into various compartments. In the *Susceptible-Infected-Recovered (SIR)* model, the entire population is placed in Susceptible, Infected, and Recovered compartments. The model assumes that Recovered individuals cannot become reinfected. The *SEIR* model adds an additional *Exposed* compartment for individuals who will become Infected and contagious after an incubation period. The models are deterministic and move from left to right: A Susceptible individual *may* move into the Exposed compartment, an Exposed individual *will* move into the Infected compartment, and an Infected individual *will* move into the Recovered compartment. These models are the basis of many more complex models that incorporate geography, multiple age groups, etc.

Although the basic SIR and SEIR models can be useful in certain public health situations, they make assumptions about the connectivity of individuals that are frequently inapplicable. To that end, we will look at a recent stochastic model and compare it with the classical SIR model as well as a pair of Monte-Carlo simulation of the SIR model. We provide a link to all of the source code used in this paper. Nevertheless, the qualitative results for the SIR model in the next section are at least as useful as the numerical results and the code itself.

† John P. Maassen, Ph.D., 2010 Eastview Ave., Louisville, KY, 40205, jpmaassen@gmail.com.

Background

Here we will provide some basis background on the SIR and SEIR models. Even so, there will be a closed form solution for the maximum fraction of the population infected I_{max} in terms of the basic reproduction number R_0 for the SIR model, Equations (4), that will provide a starting point for how these models can be useful in public health settings. For a more rigorous mathematical treatment of these models, see [1].

The SIR model was first presented by Kermack and McKendrick in 1927.[2] They used it to successfully model an epidemic on the Island of Bombay in 1905-06. Since then, there have been countless studies of the SIR model and applications of it, both *per se* and as parts of more complicated models.

The SIR model has compartments for *Susceptible (S)*, *Infected (I)*, and *Recovered (R)* individuals and assumes that Recovered individuals cannot become reinfected. Assume $S(t)$, $I(t)$, and $R(t)$ are the numbers of individuals in each compartment so that

$$S(t) + I(t) + R(t) = N, \tag{1}$$

and the total population N is constant. The SIR model defines the rates of change in each of the compartments by three coupled ordinary differential equations:

$$\frac{dS}{dt} = \mu N - \frac{\beta}{N} IS - \mu S$$
$$\frac{dI}{dt} = \frac{\beta}{N} IS - (\gamma + \mu) I \tag{2}$$
$$\frac{dR}{dt} = \gamma I - \mu R.$$

The parameter μ represents the fractional birth or death rates. We take them to be equal so that N will be constant. Adding the three derivatives in Equations (2),

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = \mu N - \mu(S + I + R) = 0 \tag{3}$$

from Equation (1). This shows that N is indeed constant. The model also assumes that all births are into the Susceptible container. In the first equation in (2), μN is the rate of births into the total population, while $-\mu S$ is the rate of natural deaths out of the Susceptible container. Likewise, $-\mu I$ and $-\mu R$ are the rates of natural deaths out of Infected and Recovered containers. When μ is relatively small, Equations (2) simplify to:

$$\frac{dS}{dt} = -\frac{\beta}{N} IS$$
$$\frac{dI}{dt} = \frac{\beta}{N} IS - \gamma I \tag{4}$$
$$\frac{dR}{dt} = \gamma I.$$

Understanding the additional constant parameters β and γ will complete the derivation of Equations (2) and (4). They all have dimensions or 1/time. In the code that we provide, the unit of time will be 1 day.

We will start with β. Near the beginning when $S \gg I$, i.e., $S \cong N$ and $I \ll N$, suppose that one infected individual has, on average, κ contacts per time with other individuals in the population. Assume that κ remains constant regardless of the population. The contacted

individuals can be in any of the three compartments, but only those in the susceptible compartment count. Then $\kappa S/N$ is the number of contacts per time that one infected individual has with susceptible individuals. Let τ represent the probability that a contact with a susceptible individual results in a new infection. Then $\kappa \tau S/N$ is the number of susceptible individuals that become infected per time from that one infected individual, and $\kappa \tau S I/N$ is the rate at which susceptible individuals become infected, in other words, $-dS/dt$ in Equations (2). *The product $\beta \equiv \kappa \tau$, is the number of contacts per time from a single infected individual that would result in an infection if the contact(s) were Susceptible.* With this,

$$dS/dt = -\beta SI/N, \tag{5a}$$

which is the first equation in (4).

A contrasting way of getting dS/dt is a law of *mass action*. It carries over from chemistry, where the rate of a chemical reaction is proportional to the product of the concentrations of the reactants. When applied to epidemiology, the term is used when the rate of infection is proportional to the product of the *numbers* of Susceptible and Infected persons in the population, or

$$\frac{dS}{dt} = -rSI, \tag{5b}$$

where r is the constant of proportionality. (Note that this is only roughly analogous to the case in chemistry.) Comparison with the previous result yields $\beta = rN$. This would mean that the rate at which the infection spreads is implicitly proportional to the total population. However, empirical data shows that in many cases the spread of human diseases is only weakly dependent upon N, in which case, Equation (5a) would be the more appropriate model, and it is what we use throughout most of this paper. Equation (5a) is often called *standard incidence*, while (5b) is called *mass action incidence*.

The parameter γ represents the recovery rate, or the fraction of the population per unit time that moves from the infected compartment to the recovered compartment. We assert that it also represents the average time that an individual remains in the infected compartment.

To show the plausibility of this assertion, consider a compartment of infectious individuals that are otherwise isolated from the remainder of the population. The rate at which their fraction drops is simply

$$\frac{dI}{dt} = -\gamma I. \tag{6}$$

Assuming that the initial fraction is I_0, the solution to this equation is

$$I(t) = I_0 e^{-\gamma t}. \tag{7}$$

Substituting Equation (7) back into Equation (6) and normalizing to 1, the *probability per unit time* that an individual *departs* from (note sign change) the infected compartment is

$$P(t) = +\gamma e^{-\gamma t}. \tag{8}$$

With this, the average time an individual spends within the infected compartment is

$$\bar{t} = \int_0^\infty tP(t)dt = \int_0^\infty t\gamma e^{-\gamma t}dt = \frac{1}{\gamma} \tag{9}$$

(integrating by parts). Simply speaking, the average amount of time that an individual is sick and infectious (the same times in the SIR model) is

$$\bar{t} = 1/\gamma. \tag{10}$$

The first two equations dS/dt and dI/dt in (2) do not contain $R(t)$ and can therefore be solved independently of the third equation. Once $S(t)$ and $I(t)$ are known, $R(t)$ can be obtained trivially from $R(t) = N - S(t) - I(t)$. The pair of solutions $[S(t), I(t)]$ can therefore be studied as a dynamical system in the plane. Since time does not appear explicitly in the right sides of Equations (2), they are *autonomous*. This means, for example, that their global stability is guaranteed by the Poincare-Bendixon theorem.[3]

In order for the model to make sense, we look for unique solutions that satisfy $0 \le S(t) \le N$, $0 \le I(t) \le N$, and $0 \le R(t) \le N$. Since $N - S(t) - I(t) = R(t) \ge 0$, we are looking for (S, I) trajectories within the triangle

$$T \equiv \{(S, I)|S \ge 0, I \ge 0, S + I \le N\}. \tag{11}$$

The limits as $t \to \infty$ of the solutions to Equations (4), S_L, I_L, and R_L, exist. Since dS/dt is negative, $0 \le S(t) < S(0) \le N$. By the monotone convergence theorem, the limit S_L exists. Similarly, since $dR/dt \ge 0$, $0 \le R(0) < R(t) \le N$, and the limit R_L exists. Finally, $I_L = 0$, since otherwise the fact that $dR/dt > \gamma I/2$ would contradict the existence of R_L. *In the SIR model, the infection always ends.*

Knowing what β and $1/\gamma$ mean, we can identify the *basic reproduction number*[1]

$$R_0 = \beta \frac{1}{\gamma}. \tag{12}$$

In a fully susceptible population, R_0 is the average number of people that one infected person infects over the course of his infection. R_0 provides important information about the course of an infection.

When $R_0 \le 1$, the infection dies out. Indeed,

$$\frac{dI}{dt} = \left(\frac{\beta}{N}S - \gamma\right)I \le \left[\frac{\beta}{N}S(0) - \gamma\right]I = \gamma\left[R_0\frac{S(0)}{N} - 1\right]I \tag{13}$$

since $S(t)$ decreases monotonically. Since $S(0)/N < 1$, when $R_0 \le 1$, the derivative is negative, and $I(t)$ decays to zero.

When $R_0 > 1$, I reaches a maximum value I_{max} and then decays to 0. In this model, I_{max} is easy to calculate.[4] Dividing the first equation by the second and using the chain rule, we obtain the separable differential equation

$$\frac{dS}{dI} = \frac{-\bar{\beta}IS}{\bar{\beta}IS - \gamma I} = \frac{-\bar{\beta}S}{\bar{\beta}S - \gamma}, \tag{14}$$

where $\bar{\beta} \equiv \beta/N$. Therefore,

$$\int \frac{\bar{\beta}S - \gamma}{\bar{\beta}S} dS = -\int dI \Rightarrow I + S - \frac{\gamma}{\bar{\beta}} \ln S = C \tag{15}$$

where C is the constant of integration. Using the initial conditions,

---

[1] R_0 depends upon the model. In the SIR model with demographics, $R_0 = \frac{\beta}{\gamma + \mu}$.

$$I(t) + S(t) - \frac{\gamma}{\beta} \ln S(t) = I(0) + S(0) - \frac{\gamma}{\beta} \ln S(0). \tag{16}$$

Now I_{max} occurs when $dI/dt = 0$, or in other words,

$$\bar{\beta} I_{max} S - \gamma I_{max} = 0 \Rightarrow S = \frac{\gamma}{\beta} = \frac{1}{R_0} N. \tag{17}$$

The right side of Equation (17) gives $S < N$, when $R_0 > 1$, so the existence of I_{max} is consistent with the model. Substituting I_{max} and S from Equation (17) into Equation (16) and solving for I_{max},

$$I_{max} = I(0) + S(0) - \frac{\gamma}{\beta} \ln S(0) - \frac{\gamma}{\beta} + \frac{\gamma}{\beta} \ln \frac{\gamma}{\beta}. \tag{18}$$

Observe that $I(0) + S(0) = N$, since we assume that $R(0) = 0$, i.e., there is no recovered population at the beginning of the simulation. If we seed the model such that $S(0) \cong N$, then $\ln S(0) \cong \ln N$. Finally, substituting $R_0 = \beta/\gamma$, we get

$$\frac{I_{max}}{N} = 1 - \frac{1 + \ln R_0}{R_0}. \tag{19}$$

This is a truly remarkable result, since I_{max}/N depends only upon R_0 and not, e.g., upon $1/\gamma$, the average amount of time that an infected individual can infect others. Finally, having I_{max}/N in closed form provides an additional check on the integrity of the SIR code. (Refer to the Discussion section for details.)

Equation (16) a relationship between S and I parameterized by t. Here is all the information necessary to draw a phase plane portrait of the solution space within the domain T in Equation (11). This phase portrait can look very different when vital dynamics, Equations (2), are included! For example, a sufficiently short lifetime $1/\mu$ can force an attractor above the horizontal axis within the T, wherein the model spirals into endemic equilibrium. See [1] for details.

Harko, Lobo, and Mac (2014) [5] proposed an "exact" solution to a SIR model, our Equations (4) but with mass action incidence. Nevertheless, their solution was also a parameterization, and Equations (20) and (27)-(31) in [5], are easily shown to be equivalent to our Equation (16).

While I_{max} represents the maximum number of individuals infected at any given time, i.e., the maximum population of the Infected compartment, the *total population infected* over the course of the epidemic is $R_L = N - S_L$. Dividing the first equation by the third in (4),

$$\frac{dS}{dR} = -\frac{\beta}{N\gamma} S. \tag{20}$$

This equation is also separable, and with the initial conditions $S(0)$ and $R(0)$, yields the immediate solution

$$\int \frac{1}{S} dS = \int -\frac{\beta}{N\gamma} dR \Rightarrow \ln \frac{S(t)}{S(0)} = -\frac{\beta}{N\gamma} [R(t) - R(0)] \tag{21}$$

or equivalently,

$$S(t) = S(0) \exp \left\{ -\frac{\beta}{N\gamma} [R(t) - R(0)] \right\}. \tag{22}$$

Now take the limit $t \to \infty$. Using $R(0) = 0$, $S(0) = N$, and $R_L = N - S_L$, in Equation (21),

$$ln \frac{S_L}{N} = -\frac{\beta}{N\gamma}(N - S_L) = R_0\left(\frac{S_L}{N} - 1\right). \tag{23}$$

Equation (23) is a transcendental equation for the fraction of the limiting susceptible population.

Many viral infections present an incubation period wherein a person is exposed but not yet contagious. This is apparently true for the SARS-CoV-2 virus associated with the COVID-19 pandemic. For example, the recent UK study by Ferguson et al. (2020) uses an incubation period of 5.1 days with infectiousness beginning at 4.6 days.[6] This motivates adding an *Exposed (E)* compartment to the SIR model.

The Susceptible-Exposed-Infected-Recovered (SEIR) model is

$$\frac{dS}{dt} = \mu - \beta IS - \mu S$$

$$\frac{dE}{dt} = \beta IS - (\sigma + \mu)E \tag{24}$$

$$\frac{dI}{dt} = \sigma E - (\gamma + \mu)I$$

$$\frac{dR}{dt} = \gamma I - \mu R.$$

There is an additional parameter σ that represents the fraction of the population leaving the Exposed compartment per unit time. It can be shown that in the SEIR model

$$R_0 = \frac{\beta \sigma}{(\gamma + \mu)(\sigma + \mu)}. \tag{25}$$

Code[2]

Although much can be inferred analytically for the SEIR model as well, when SIR and SEIR are used as building blocks for new models, it is more expeditious to solve them numerically. Consider the first member of Equations (2). We have approximately

$$\frac{dS}{dt} = f[S(t)] \approx \frac{\Delta S}{\Delta t}. \tag{26}$$

(Of course, f also depends upon $I(t)$). By far the simplest approach (called Euler's method) is to use

$$\Delta S = f(S)\Delta t. \tag{27}$$

Starting with the initial condition $S = S_0$, increment S after time Δt with

$$S_1 = S_0 + (\Delta S)_0 = S_0 + f(S_0)\Delta t, \tag{28}$$

and repeat this process throughout the duration of the epidemic. When Δt is sufficiently small, this will provide an adequate approximation for $S(t)$. However, the convergence is slower than necessary, because in each step, we are sampling the function $S(t)$ at the beginning of the time increment Δt. The Runga-Kutta method improves the rate of

[2] Source code for the SIR and SEIR models, the Monte Carlo simulation of SIR used in this paper are publicly available. The link is https://www.dropbox.com/sh/pfkgnc424p9dol9/AADQNkyaf25wixFvvAvyevjva?dl=0.

convergence by iteratively sampling $S(t)$ at several points through Δt. Since t does not enter *explicitly* into this equation, the fourth-order Runga-Kutta formulae are [7]

$$k_1 = \Delta t f(S_n)$$

$$k_2 = \Delta t f\left(S_n + \frac{k_1}{2}\right)$$

$$k_3 = \Delta t f\left(S_n + \frac{k_2}{2}\right) \tag{29}$$

$$k_4 = \Delta t f(S_n + k_3)$$

$$S_{n+1} = S_n + \frac{k_1}{6} + \frac{k_2}{3} + \frac{k_3}{3} + \frac{k_4}{6}.$$

This method will be used for each of the seven equations in the SIR and SEIR code in this paper.

The code solves equivalent sets of the equations for fractions of the total population. With

$$\bar{S} \equiv \frac{S}{N}, \bar{I} \equiv \frac{I}{N}, and \ \bar{R} \equiv \frac{R}{N}, \tag{30}$$

Equations (2) become

$$\frac{d\bar{S}}{dt} = \mu - \beta\bar{I}\bar{S} - \mu\bar{S}$$

$$\frac{d\bar{I}}{dt} = \beta\bar{I}\bar{S} - (\gamma + \mu)\bar{I} \tag{31}$$

$$\frac{d\bar{R}}{dt} = \gamma\bar{I} - \mu\bar{R}.$$

Equations (24) for the SEIR model transform similarly. N drops out of the equations, because we have chosen a model, Equations (5a), with standard incidence in which the spread of the infection does not depend upon the total population. Additionally, this keeps the values of $S(t)$, $E(t)$, $I(t)$, and $R(t)$ between zero and one which is good from an arithmetical perspective.

The reason for providing code for both SIR.c and SEIR.c is to show how the SEIR model, a natural extension of the SIR model, can be coded with just a few extra lines. It is anticipated that some readers will want to extend the SEIR model to specific applications, and the transition from SIR to SEIR provides a pattern for accomplishing this. To this end, both models are coded as similarly as possible. Nevertheless, we have commented the SIR model code more thoroughly, since it provides the foundation on which the SEIR model code stands.

The SIR and SEIR models are systems of three and four ordinary differential equations. The cores of both programs are (a) codices describing these equations and (b) a Runga-Kutta function to integrate them. Consider the first equation in the SIR model, Equations (31):

$$\frac{dS}{dt} = \mu - \beta SI - \mu S. \tag{32}$$

In SIR.c, we code this as

```
double dSdt(double Svar)
{
    return(mu - beta*Svar*I - mu*Svar);
}
```

There are two model variables S and I here. We designate S by Svar in the code, because S is the function being differentiated in Equation (32) that the Runga-Kutta function will need to increment. We do the same thing for each of the seven equations in both models. There is a specialized fourth-order Runga-Kutta for autonomous equations where the independent variable (in our case, time) does not appear explicitly. It accepts as an argument the function that defines the differential equation. During each time step, each of the functions on the right in Equations (31) are integrated by one time step. The solutions are not incremented until after all the differences ($\Delta's$) are computed in order to keep the updates to the members of each coupled set of equations in sequence. Since all of the derivative functions are passed to another function, the model parameters are global in order to shorten the functions passed as arguments and to make the code more readable.

Our goal is for both SIR.c and SEIR.c to be ISO C and POSIX compliant. They will compile without modification under most flavors of Linux (including the new Windows Subsystem for Linux) and Unix (including macOS). Some environments may lack *unistd.h* and *getopt()* for parsing command line arguments. For this reason, we have included versions of getopt.c and getopt.h in the distribution directory. Please refer to the *Readme* file for details.

SIRno.c (no command line) is an alternative to SIR.c. Without the command line processing and other features not directly related to the model, it is only about half the length. It is the better place to start for understanding what the code does and to extend it to other scenarios.

Both models contain a demographic parameter $\mu = 3 \times 10^{-5} day^{-1}$ as a composite birth and death rate for the United States. It is small in comparison with the COVID-19 numbers. Nevertheless, this code may have other applications in the future. The presence of μ allows the code to "breathe" by providing an example of input and output. With a few additional lines, one could add another four differential equations for a second SEIR model and suitably couple the two sets of equations with an additional I/O parameter similar to μ. This could be the beginning of an age-dependent model.[1] Our purpose in releasing this code is to make epidemiological model building as fast and as intuitive as possible.

Discussion

We begin with the SIR code to verify that it is working correctly and then compare the results with the SEIR code. Figure 1 is the well-known SIR plot. We used $R_0 = 2.5$, the initial condition $I(0) = 0.05$, and $\Delta t = 6/day$. To compare with theory we used $\Delta t = 24/day$ and $I_0 = 0.001$. The code gives $I_{max}/N = 0.235$, while Equation (19) yields $I_{max}/N = 0.233$. The peak occurs at 69 days in the code, at which point, $S = 0.40$, while Equation (17) yields $S/N = 1/R_0 = 1/2.5 = 0.40$. Finally, the total number infected over

the course of the epidemic is $R_L/N = 0.891$, while solving Equation (23) numerically gives $R_L/N = 0.893$. Results from the code that differ in the third decimal place can be brought into even better agreement by further reducing $S(0)$ and Δt.

To see how the SIR model might be applied in a public health scenario, once a set of β, γ, and R_o have been found that fit empirical data, deconstruct I_{max}/N in Equation (19) as follows:

$$\frac{I_{max}}{N} = 1 - \frac{1 + ln(\kappa\tau/\gamma)}{\kappa\tau/\gamma}. \tag{33}$$

Although we may not have individual measurements of κ and τ, we do know how changing either one would affect their product β. Social distancing would reduce κ while frequent hand washing and wearing masks would reduce τ. An antiviral agent could be developed to reduce $1/\gamma$. With a closed-form result like Equation (33), it is possible to map small changes in the parameters to changes in I_{max}. Similar considerations would apply to Equation (23).

Having successfully compared the output of SIR.c with theory, we move on to SEIR.c. We coded SEIR.c identically to SIR.c but with the addition of Exposed compartment (one additional ordinary differential equation and one additional parameter σ). Thus, we have verified SEIR.c informally "by analogy." At least this is consistent with using these core models as building blocks for more complex models: We have used SIR.c as a building block for SEIR.c, even though there is considerable additional theory to support the SEIR model itself.

Figures 2 and 3 show some sample runs from the SEIR model. R_0 is 2.5 in Figure 2 and 1.25 in Figure 5. $R_0 = 2.5$ is consistent with COVID-19 data as of the date of this paper. The Imperial College Study [6] used $R_0 = 2.4$ and a mean generation time of 6.5 days, which we assigned to $1/\sigma$. Based upon the WHO-China study, we used $1/\gamma = 14$ days as the time an average patient is contagious.[8] This is likely the most uncertain parameter. (For example, sick patients are more isolated.) We started the model at a relatively mature point with 5% of the population exposed. With $R_0 = 2.5$, the infection peaks at $I_{max} = 17\%$ (Figure 2). For $R_0 = 1.25$, $I_{max} = 4\%$ (Figure 3), only 24% of the previous result.[3] Given our premise that models like SEIR most reliably map the effects of small changes in the input, perhaps a more reliable result is that if R_0 drops from 2.5 to 2.25, I_{max} drops to 0.15, 86% of its value at $R_0 = 2.5$. Table 1 shows the journey from $R_0 = 2.5$ to 1.25 for these estimated COVID-19 parameters.

[3] In the SIR model, Equation (19) yields $I_{max}/N = 0.24$ and 0.02 respectively when R_0 drops from 2.5 to 1.25. However, Equation (29) requires that $S(0) \cong 1$.

R_0	I_{Max}	$t_{max}\ (days)$
2.50	0.171	59
2.25	0.147	63
2.00	0.121	68
1.75	0.094	72
1.50	0.066	74
1.25	0.043	59

Table 1 - Imax versus R0 for E0 = 0.05, 1/γ =14 days, and 1/σ=6.5 days.

In Table 1, it is interesting how t_{max} shifts from 59 days to 74 days and then back to 59 days. In our analysis of the SIR model, we were not able to produce a closed form expression for t_{max}. By dividing one differential equation by another, Equations (14) and (20), we obtained separable differential equations. Their solutions gave relationships between $S(t)$ and $I(t)$ and $S(t)$ and $R(t)$ respectively, where time t entered implicitly as a parameter. It is not possible to invert these equations for t, because we do not have independent solutions for $S(t)$, $I(t)$, and $R(t)$, but only parametric relations between them. This underscores the fact that there is no known exact solution to the SIR model in Equations (4). On the other hand, it is easy to locate t_{max} by integrating the equations stepwise in time.

Schütz et al. (2008) [9] have recently constructed a stochastic model based upon a linear chain in which each node is either Susceptible, Infected, or Recovered, and each node interacts only with its two nearest neighbors. (This model was improved by Williams et al. (2018) [10] by considering the different cases where a Susceptible node is adjacent to either one or two Infected nodes.) The version of the SIR model that they start with corresponds to our Equations (4) but with mass-action incidence. They refer to this as the "mean-field" model. They describe their stochastic model (linear chain, nearest neighbor interactions only) as "low connectivity." Most interestingly, they discovered that they could better correlate their stochastic model with the mean-field model if they "renormalized" $\beta_{mf} = \beta/N$, thus recovering standard incidence.

With this, it is possible to compare our numerical solution to SIR with their exact solution Equations (26), (29), (32), and (33) in [9]. The results are shown in Figure 4 with their parameters $\beta = 0.9$, $\gamma = 0.1$ (corresponding to $R_0 = 9.0$)4, and $I(0) = 0.2$. The agreement is qualitatively good, with the differences attributable to the connectivity assumptions in their model. Moreover, with their model, it is indeed possible to solve for t_{max}. By setting $dn_I/dt = 0$ in their Equation (32), one obtains a transcendental equation that is easily solved numerically with $t_{max} = 4.1$ days.

---

4 In a linear model where each node interacts only with its two nearest neighbors, this would challenge the conventional interpretation of R_0.

Finally, Schütz et al. compared their exact solution with a Monte-Carlo simulation. Such simulations are relatively easy to code and run—although they run slowly. Stochastic models in general and Monte-Carlo simulations in particular provide information about fluctuations and can be applicable in scenarios like low population and low and non-uniform population densities that stress the deterministic mean-field SIR and SEIR models.

To this end, Figure 5 is a Monte-Carlo simulation of the SIR model with $N = 1000$, $R_0 = 2.5$, and $\gamma = 0.07/day$. In terms of connectivity, this simulation tests all 499,500 Susceptible-Infected node pairs and moves Susceptible nodes to the Infected compartment with a probability $\beta \Delta t / N$. The two versions of the code for this are included with the distribution. One version enforces independently updating the Infected and Recovered compartments at each time step like the model code for SIR and SEIR that integrates the system of equations. This version comes at the cost of more complicated code. Both versions run in quadratic time $O(n^2)$ where n is the number of nodes (corresponding the total population N). Both simulations and the integrated SIR model converge to the same output as N becomes large. We therefore conjecture that a stochastic model like references [9] and [10] would need to consider *all* $N(N-1)/2$ node pairs in order to agree with SIR in the limit $N \rightarrow \infty$.

May 2, 2020

References

1. Hethcote, Herbert W. "The basic epidemiology models: models, expressions for R_0, parameter estimation, and applications," 1-61; "Epidemiology models with variable population size," 63-89; "Age-Structured Epidemiology Models and Expressions For R_0," 91-128; all in *Mathematical Understanding Of Infectious Disease Dynamics*. 2009.
2. Kermack, William Ogilvy, and Anderson G. McKendrick. "A contribution to the mathematical theory of epidemics." *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character* 115.772 (1927): 700-721.
3. Coddington, Earl A., and Norman Levinson. *Theory of ordinary differential equations*. Tata McGraw-Hill Education, 1955.
4. Waltman, Paul. Deterministic threshold models in the theory of epidemics. Vol. 1. Springer Science & Business Media, 2013.
5. Harko, Tiberiu, Francisco SN Lobo, and M. K. Mak. "Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates." *Applied Mathematics and Computation* 236 (2014): 184-194.
6. Ferguson, Neil, et al. "Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand." (2020).
7. Fröberg, Carl-Erik. *Introduction to numerical analysis*. Reading Massachusetts: Addison-Wesley Publishing Company, 1969.

8. World Health Organization. " Report of the WHO-China Joint Mission on Coronavirus Disease 2019." (2020).

9. Schütz, Gunter M., Marian Brandaut, and Steffen Trimper. "Exact solution of a stochastic susceptible-infectious-recovered model." Physical Review E 78.6 (2008): 061132.

10. Williams, H. Thomas, Irina Mazilu, and D. A. Mazilu. "Stochastic epidemic-type model with enhanced connectivity: exact solution." Journal of Statistical Mechanics: Theory and Experiment 2012.01 (2012): P01017.
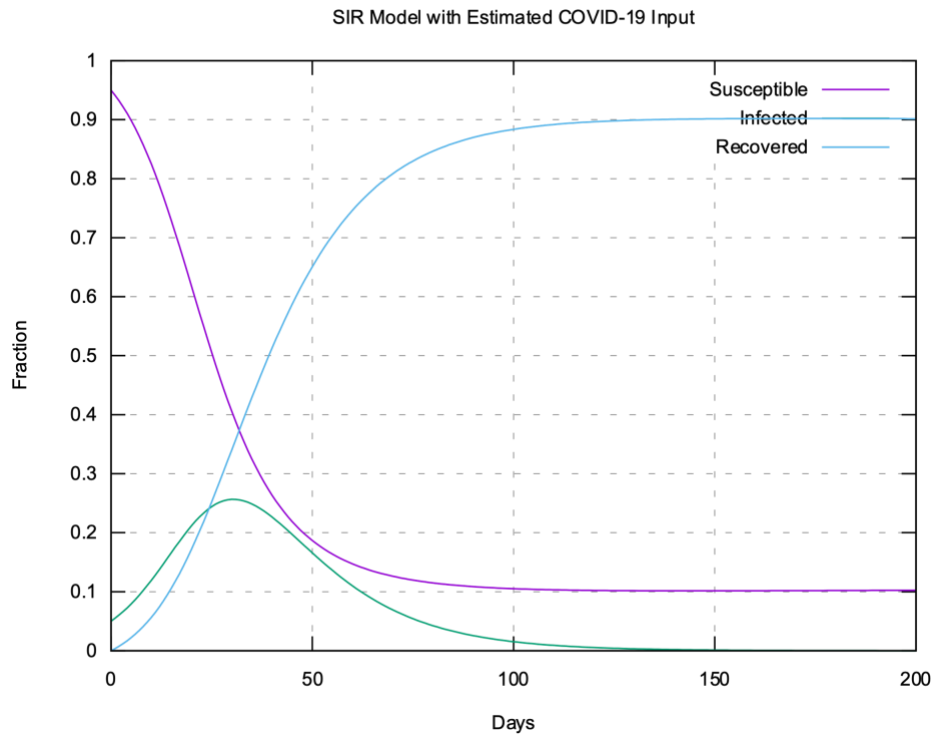
Figure 1 - SIR model with $R_0 = 2.5$ and $I_0 = 0.05$. I_{max} is 0.257 and occurs at 30 days.

Figure 2 - SEIR simulation of COVID-19 with $R_0 = 2.5$ and $E(0) = 0.05$. The percentage infected over the duration of the epidemic is 17%.
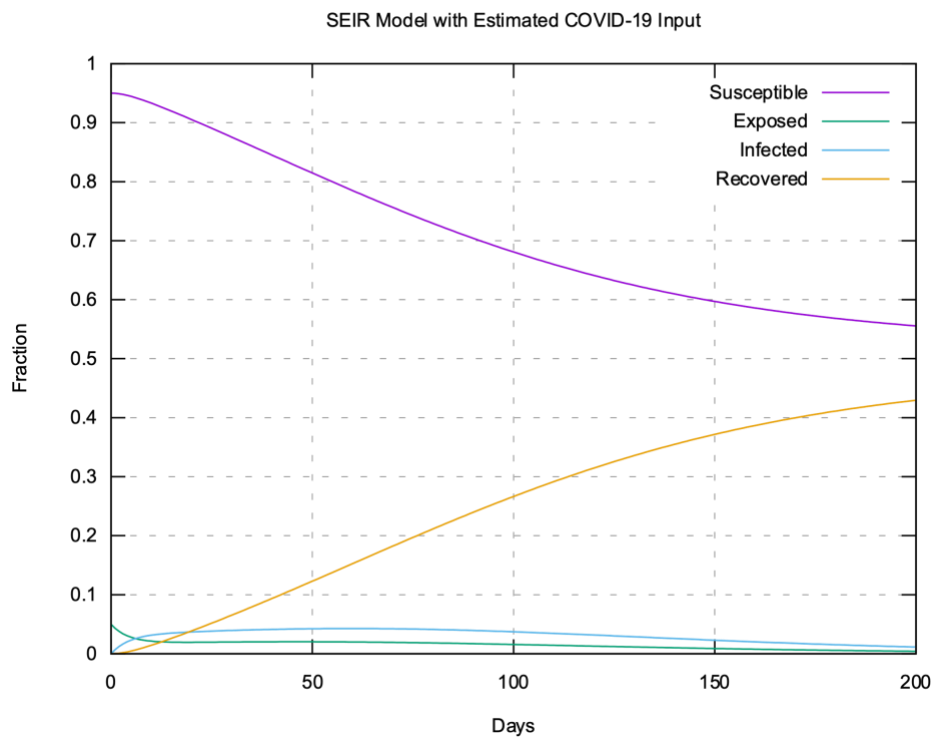

Figure 3 - SEIR simulation of COVID-19 with $R_0 = 1.25$ and $E(0) = 0.05$. The percentage infected over the duration of the epidemic is 4%.

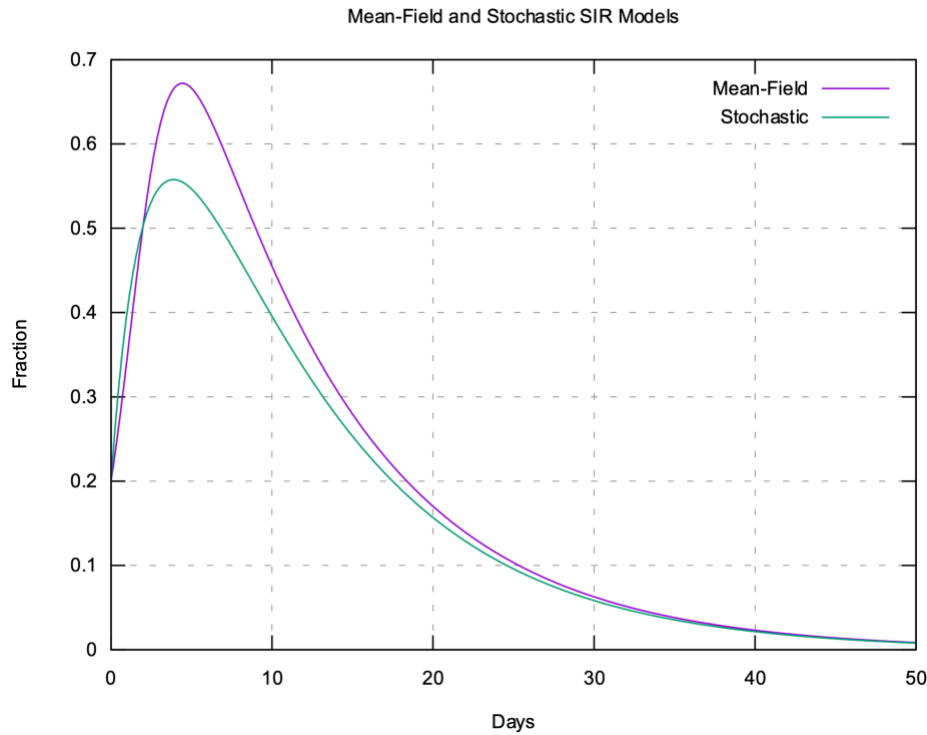Mean-Field and Stochastic SIR Models

Figure 4 - Comparison of Mean-Field (deterministic) SIR model with corresponding stochastic model of Schütz et al. (2008) (Infected container only) with input $\gamma = 0.1$ and $R_0 = 9.0$.
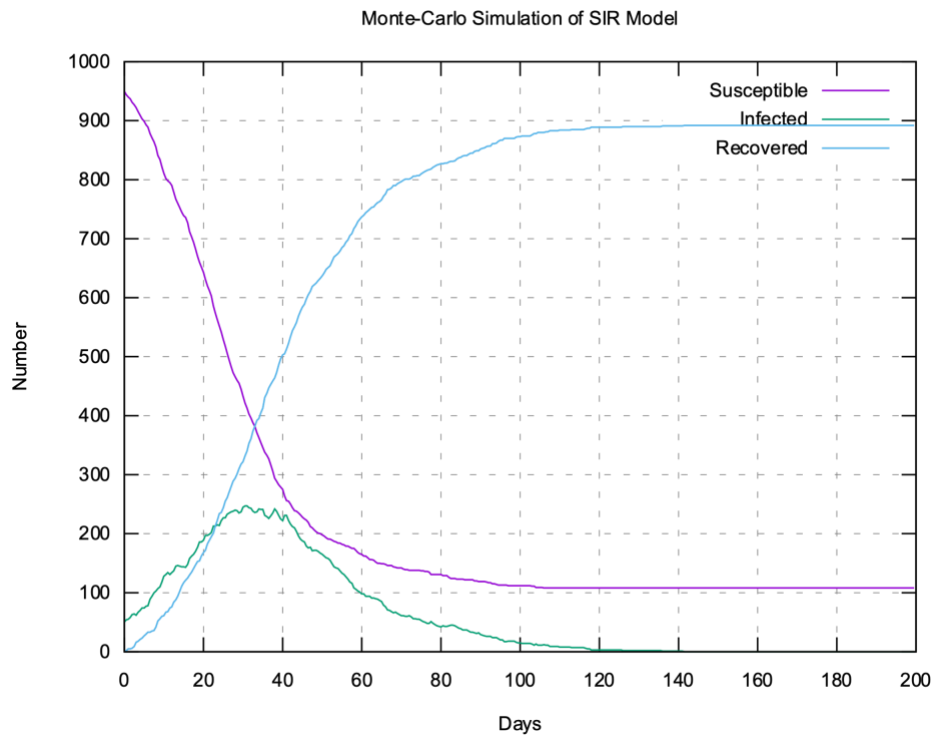
Monte-Carlo Simulation of SIR Model

Figure 5 - Monte-Carlo simulation of SIR model with population N = 1000, $R_0 = 2.5$, and $\gamma = 0.07$/day.