www.simiode.org
SIMIODE A SYSTEMIC INITIATIVE FOR MODELING INVESTIGATIONS & OPPORTUNITIES WITH DIFFERENTIAL EQUATIONS

STUDENT VERSION
Enabling Epidemic Exploration Education

Brian Winkel
Director, SIMIODE
Cornwall NY USA

STATEMENT

While working with the principal author of [3] to prepare the essence [6] of a modeling scenario on the basic SIR model for epidemics for wider use than as a project in the author's class we became aware of several interesting possibilities. Since [3] presents a modeling opportunity with data obtained and cited in [1, 5] we explored the several routes to parameter estimation in the SIR model with respect to the data pictured in [1] and offered in Table 1.

Time (in days)	0	1	2	3	4	5	6	7	8	9	10	11	12	13
# Bedridden Boys	1	3	25	72	222	282	256	233	189	123	70	25	11	4

Table 1. Total Number of Bedridden Boys on Day t.

Further, we move to a new SIR model in which we employ a Michaelis-Menten form of a saturated incidence rate between infectives and susceptibles. We suggest several strategies for estimating parameters in this newer model and challenge the reader to compare our results to the more traditional SIR model efforts.

From [3] we quote

> A boarding school is a somewhat closed community; all the students and teachers tend to live on or near campus, and the students do not regularly interact with people not in the boarding school community. Table 1 gives data for a flu outbreak at a boarding school in England. No one died in this flu outbreak as it was not Spanish Flu. These data were compiled from the Communicable Disease Surveillance Center [1, p. 587] and are given as an example by Murray in his book on mathematical biology [5]. The data values were extracted from the graph by [6].

From [4] we quote

> There are 763 boys in the school. From Table 1 we can see that the initial number of bedridden students is one. Let us assume that the number of boys who are bedridden on day t (the data given) is the number of students infected on day t, i.e. no boys are bedridden for any other reason during this period of time and all boys infected will be bedridden. One may wish to test the effects of this last assumption on the model in the analysis phase. Another, perhaps more realistic assumption, is that the bedridden boys represent a fixed percentage of the infected students. This assumption, which excludes teachers and staff from the population, implies that the number of students infected on day zero is one, i.e. $I(0) = 1$, where $I(t)$ is the number of students infected on day t.

First, we offer a traditional system of differential equations (1) for the spread of disease from susceptible, to infected, to recovered individuals. This can be depicted in the diagram in Figure 1.

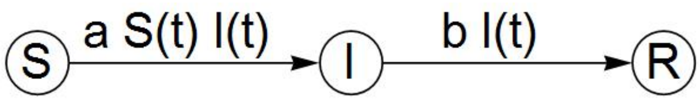

Figure 1. Picture for a traditional SIR model of an epidemic with connecting rates $a \cdot S(t) \cdot I(t)$ and $b \cdot I(t)$.

Our system of differential equations in (1) is referred to as the *SIR model for the spread of disease* or an epidemic model.

$$\begin{aligned} S'(t) &= a \cdot S(t) \cdot I(t) \\ V(t) &= a \cdot S(t) \cdot I(t) - b \cdot I(t) \\ R'(t) &= b \cdot I(t) \end{aligned} \tag{1}$$

where $S(t)$ is the number of susceptible students at time t, $I(t)$ is the number of infected and infectious students at time t, and $R(t)$ is the number of the infected students who have recovered at time t.

Here a is the infection rate or proportion with units 1/((infected number)(time)), while $a \cdot S(t) \cdot I(t)$ represents the number of possible interactions between the susceptible students and the infected students which result in an increase in the infected student population per unit time. b is the recovery rate with units (recovered number) /((infected number)(time)).

Second, we offer a model (2) for the spread of disease in which there is a form of a saturated incidence rate between infectives and susceptibles.

$$\begin{aligned} S'(t) &= -a\frac{S(t)}{K_m + S(t)}I(t) \\ I(t) &= a\frac{S(t)}{K_m + S(t)}I(t) - b \cdot I(t) \\ R'(t) &= b \cdot I(t) \end{aligned} \tag{2}$$

This can be depicted in the diagram in Figure 2 where the variables in (2) are as in (1)..

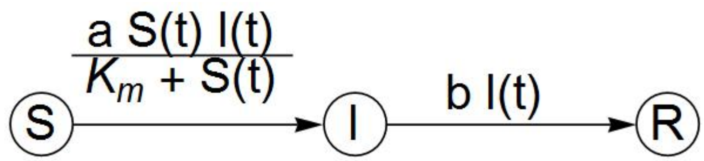

Figure 2. Picture for an SIR model of an epidemic which models saturation or satiation of the incidence rate between infectives and susceptibles with connecting rates $a\frac{S(t)}{K_m+S(t)}I(t)$ and $b \cdot I(t)$.

Here a is the infection rate or proportion with units (susceptible number)/((infected number)(time)), while $a\frac{S(t)}{K_m+S(t)}I(t)$ represents the number of possible interactions between the susceptible students and the infected students which result in an increase in the infected student population per unit time. In this model K_m represents the number of susceptible students at which a given infected student will infect a susceptible student at one half the maximum infection rate, which is a. See Figure 3.

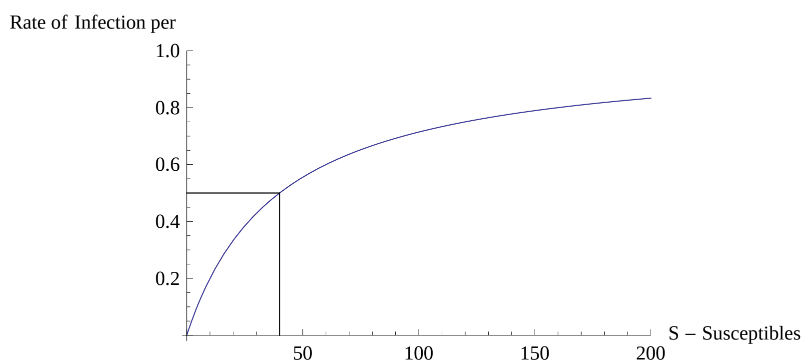

Figure 3. The functional response $\frac{S(t)}{K_m+S(t)}S(t)I(t)$ for $K_m = 40$. The vertical line depicts the susceptible population level and the horizontal line points to the 50% or 0.50 level of infection corresponding to that susceptible population level, exactly one half the maximum infection value of 1 in this case.

Epidemic Modeling

We discuss the traditional SIR model for epidemics and offer a modification, based on Michaelis-Menten kinetics model, to reflect the saturation of infected rates. In both cases we suggest three strategies for estimating parameters on our way to fully assessing the value of each model and each approach. Finally, we compare our results and ask questions.

Weakness in traditional model

In (1) the term $a \cdot S(t) \cdot I(t)$ leads to the following consequences. If we have $I(t) = 1$ infected individuals at time t then the first differential equation for $S(t)$ in (1) suggests that the susceptible pool decreases at a rate proportional $S(t)$, i.e. , $-a \cdot S(t) \cdot I(t) = -a \cdot S(t) \cdot 1 = -a \cdot S(t)$, which means that this one individual will infect the same percentage per unit time no matter how many susceptibles there are available, namely a, of the susceptibles, $S(t)$. Thus, suppose $a = 0.1$. Then for $S(t) = 1,000$ we would have $(0.1) \cdot 1000 = 100$ new infecteds per unit time while for $S(t) = 100,000$ we would have $(0.1) \cdot 100000 = 1000$ new infecteds per unit time. Indeed, if $S(t)$ were very large then this one individual, $I(t) = 1$, could infect the same percentage of this large population of susceptibles, $S(t)$, *no matter how large $S(t)$ is*. This is highly unrealistic, for this one infected individual could not contact or transmit infections to such a large number of susceptibles per unit time.

Strength in modified model with saturation

In the differential equation for $S'(t)$ in (2) we attempt to be more realistic by replacing the term $a \cdot S(t) \cdot I(t)$ with the term $a\frac{S(t)}{K_m+S(t)}I(t)$. The term $K_m + S(t)$ in the predator-prey literature is an example of a *functional response* [2] or rate of prey intake by the predator as a function of the prey levels. This functional response suggests that as $S(t)$ gets larger and larger there is a limiting value of the term representing the rate of infection, indeed, $\lim_{t->\infty} \frac{S(t)}{K_m+S(t)} = 1$. College Joe (predator) can consume only so many pizzas (prey) per unit time even if presented the whole pizza store's supply! In our situation we do not have predator or prey, but rather infected and susceptible, respectively.

If we consider the full term $a\frac{S(t)}{K_m+S(t)}I(t)$ in the first differential equation for $S'(t)$ in (2) and let $S(t)$ get larger and larger, i.e. the available susceptible pool level increases indefinitely, we see that $a\frac{S(t)}{K_m+S(t)}I(t) \to a \cdot I(t)$, which means that the rate of infection or decrease in susceptibles, depends only on the number of infected individuals. Indeed, a can be considered as the maximum infection rate of infecting susceptible individuals attainable per individual infected member of the population. The number K_m signifies the number of susceptibles available to the infected population at which infection takes place at exactly half this maximum rate, a. We illustrate this in Figure 3 in which $K_m = 40$. We see that when the susceptible population level is at $S = 40$ then the functional response level or maximum infection rate is at 50% or 0.50.

Parameter Estimation Efforts

Now that we have two models for the spread of the disease, (1) and (2) , we seek to ascertain if either or both render good models for the boarding school data in Table 1 by estimating the parameters for both and fitting the resulting models to the data in Table 1. We proceed in three ways, based on minimization of the sum of square errors between the model prediction of infected individuals and the observed number of infected individuals over the entire 14 day period of the epidemic.

First, in the case of (1) we consider the parameter surface, ParSpace $= (a, b, SSE(ab))$ in three dimensional space. formed by using the parameters a and b to numerically solve the nonlinear system of differential equations in (1). We then compute the sum of square errors, $SSE(a, b)$ between the model prediction of infected individuals and the observed number of infected individuals over the entire 14 day period of the epidemic for the parameter values of a and b. The totality of these points, $(a, b, SSE(a, b))$, form our sum of square errors function or parameter surface in this parameter space. We seek to determine the parameters a and b which minimize $SSE(a, b)$.

We can do this minimization in several ways.

1. Select reasonable initial estimates of a and b from a range of values, say $a_{\min} \leq a \leq a_{\max}$ and $b_{\min} \leq b \leq b_{\max}$. Then use trial and error by numerically obtaining solutions of (1) and comparing the resulting model plot with the data plot as seen in Figure 4.

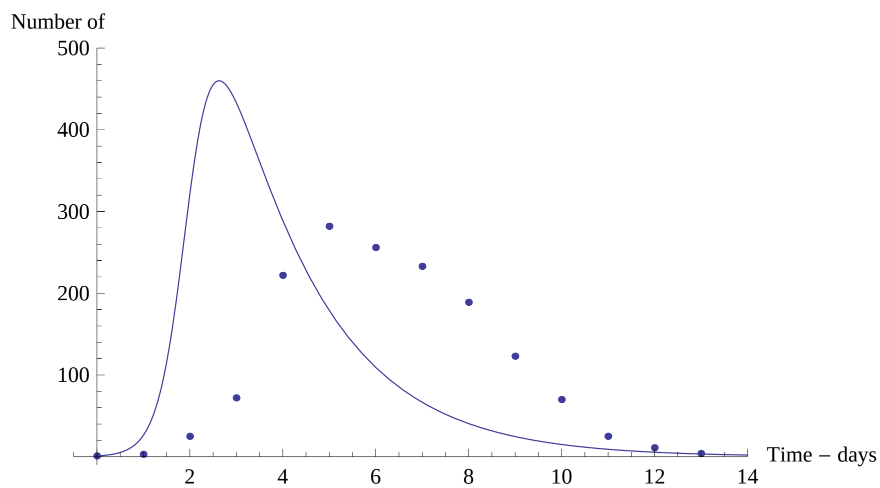

Figure 4. An early estimate for the model in (1) using $a = 0.005$ and $b = 0.5$, resulting in a sum of quares error of $SSE(0.005, 0.5) = 318,089$. This is not a particularly good estimate!

2. We can compute points on the three dimensional surface, $\mathcal{S}$ defined by (3):

$$\mathcal{S} = \{(a, b, SSE(a, b)|a_{\min} \leq a \leq a_{\max},\ b_{\min} \leq b \leq b_{\max}\}. \tag{3}$$

We do this over a grid with a small mesh size, perhaps with thousands of points and hence thousands of solutions of our differential equations system (1) for each paramter pair (a, b), along with an accompanying computation of the sum of square errors betweeen model values and data values. From this set of values, $\mathcal{S}$, we extract the triple $(a, b, SSE(a, b))$ whose $SSE(a, b)$ is smallest and offer up the corresponding parameters a and b as our "best" model parameters. We can then compare our model using these optimal parameter values to our data for a visual confirmation of our modeling process.

3. We can use a gradient approach of steepest descent on the surface, $\mathcal{S}$, to determine values of a and b which render a minimum $SSE(a,b)$. We compute the gradient by estimating the respective partial derivatives in each direction, a and b, using symmetric differences, and forming a unit vector in the opposite direction of the resulting unit gradient vector, for the gradient vector points in the direction of steepest ascent and we want to descend to a minimum; hence we go in the opposite direction, i.e. *steepest descent*

   We compute the partial derivatives numerically using a symmetric difference approach, in (4). For example, we compute $\frac{\partial SSE(a,b)}{\partial a}$ as:

$$\frac{\frac{SSE(a+\Delta a,b)-SSE(a,b)}{\Delta a}+\frac{SSE(a,b)-SSE(a-\Delta a,b)}{\Delta a}}{2}=\frac{SSE(a+\Delta a,b)-SSE(a-\Delta a,b)}{2\Delta a}\,. \tag{4}$$

   It should be pointed out that each evaluation of $SSE(a,b)$ involves placing the parameters a and b into the system of differential equations (1), solving the system numerically, and computing the sum of square errors between the resulting model for the number of infected individuals with this particular parameter set a and b and the actual data values for the number of infected individuals. Thus, each step is computational intense. Moreover,we may change the step size as we near the minimum value of $SSE(a,b)$.

   We can also do something comparable for a grid and a gradient search for the saturation model, (2), except that we are then in a four-dimensional world and our "surface" is defined by (5):

$$\mathcal{SK}=\{(a,b,SSE(a,b)|a_{\min}\le a\le a_{\max},\ b_{\min}\le b\le b_{\max},\ K_{m,\min}\le K_m\le K_{m,\max}\}\,. \tag{5}$$

   Nevertheless, (1) we can adopt a trial and error approach - more computational choices now as there are three parameters a, b, and K_m, but we may have a sense of their value and meaning to proceed intelligently, not just blindly; (2) proceed to form a three dimensional grid as in the first model with only two paramteres, a and b, bu in this model three parameters, a, b, and K_m, and search for a minimum $SSE(a,b,K_m)$ in this case; and (3) compute gradients and use the method of steepest descent to move down the surface, $\mathcal{SK}$ (in four dimensional space!), efficiently to a minimum SSE value and hence a best fitting set of parameters for our model in (2).

ACTIVITY

For both models (1) and (2) determine the best model using one or more of the strategies:

1. Trial and error approach in an attempt to guess optimal parameter values for modeling the data.
2. Form a grid with a reasonable range for the parameters and search for a minimum SSE as a function of these parameter values.

3. Compute gradients and use the method of steepest descent to move down the surface, whose coordinates are the parameters and the sum of squared errrors, efficiently to a minimum SSE value.

REFERENCES

[1] Communicable Disease Surveillance Center. 1978. News and Notes: Influenza in a Boarding School. *British Medical Journal*. 1(6112): 586-590. `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1603269/pdf/brmedj00115-0064.pdf`. Accesses 15 September 2018.

[2] Holling, C. S. 1959. The components of predation as revealed by a study of small-mammal predation of the European pine sawfly. *The Canadian Entomologist*. 91(5): 293-320.

[3] Miller, S. and J. Helms. 2010. The Disease that Infected Half the World: Mathematics, Biology, and History. *PRIMUS - Problems, Resources, and Issues in Mathematics Undergraduate Studies*. 20(3): 245-260.

[4] Miller, S. 2014. Modeling a Nonlethal Influenza Epidemic. 6-001-S-Epidemic. *https://www.simiode.org/resources/572*. Accessed 15 September 2018.

[5] Murray, J. 1993. *Mathematical Biology, Second Corrected Edition*. New York: Springer-Verlag.

[6] Switkes, J. 2003. A modified discrete SIR model. *The College Mathematics Journal*. 34(5): 399-402.