Using an SI Compartment Model to Approximate Influenza Dynamics

Hunter Ducharme and Aran Bercu

Honors Department, Lone Star College

Fall Semester 2015

1 Abstract

Mathematical epidemiology is the application of mathematics to modeling the spread of infectious diseases. The most fundamental epidemiology compartment model is the SI (susceptible-infectious) model which breaks a population into two health states—those susceptible to being infected and those who are infectious. Using weekly data published by the Center for Disease Control (CDC), an SI compartment model $I(t)$ was created in an attempt to predict the dynamics of seasonal type A influenza in the United States, a disease usually approached with an SIR (susceptible-infectious-recovered) compartment model. A derivative of the SI function was found $I'(t)$ in order to produce a curve identical to that of an SIR infection curve. The parameter values for $I'(t)$ were found using computational techniques and the CDC 2014-2015 data, which resulted in a theoretical SIR function $I'(t)$ that very accurately estimates the annual infection curve for type A influenza for the years 2012 through 2015. This finding suggests there may be possible applications for an SI compartment model in estimating the dynamics of more complex infectious diseases.

2 Introduction

Mathematical epidemiology is the study of of infectious diseases and how they spread throughout a population. The dynamics of a disease can be approximated through the application of a system of differential equations that break a population down into multiple categories of people. Using various assumptions and analyzing data from previous diseases, parameters can be found for each compartment equation to calculate the best estimate of the number of infections at a particular point in time. The most fundamental procedure is the SI (susceptible-infectious) compartment model which breaks a population into two health groups: those susceptible to being infected and those who are already infectious [1].

Multiple variations of this model which include: SIS (susceptible-infectious-susceptible), SIR (susceptible-infectious-recovered), and SEIR (susceptible-exposed-infectious-recovered) compartment models [1]. Depending on the disease in focus, there generally is one compartment model (or a slight variation of it) that models the dynamics of the disease most accurately by taking into account the numerous potential health states that an individual can reside in. With more complex compartment models there arises a degree of difficulty that may be potentially avoidable which can significantly simplify the process of modeling the disease in scope.

3 Methods/Results

3.1 Endemic Model

Our team utilized the SI compartment model in an attempt to estimate the spread of seasonal type A influenza through the United States. Individuals infected with influenza can be seen to reside in one of three different health states at any point in time—susceptible, infectious, or recovered; thus, an SIR or SEIR model would generally be a more accurate approach for a disease such as influenza. However, our team focused on determining the potential application of an SI model when applied to a disease of higher complexity such as influenza.

The SI model divides the total population N into two health states: susceptible S and infectious I. The rate at which susceptible individuals become infected is the number of contacts per unit of time r, multiplied by the probability a contact will become infected by the disease β [1]. The fractional population that is infected is: I/N [1]. This model assumes infectious hosts stay infected for the rest of their lives and no deaths occur during the influenza season.
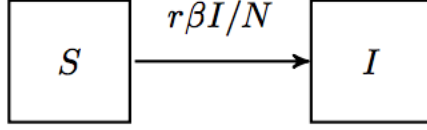
Figure 1: Susceptible-Infectious model. Reprinted from "Introduction to Mathematical Epidemiology," by Author N. Chitnis, 2011, Swiss Tropical and Public Health Institute, 2.

$$\frac{dS}{dt} = -r\beta(N - I)\frac{I}{N} \tag{1a}$$

$$\frac{dI}{dt} = r\beta S\frac{I}{N} \tag{1b}$$

Combining equations (1a) and (1b) with the substitution $S = N - I$ results in a single differential equation,

$$\frac{dI}{dt} = r\beta(N - I)\frac{I}{N}. \tag{2}$$

Equation (2) can be solved by hand with $I(0) = I_0$ to give the number of infections as a function of time $I(t)$,

$$I(t) = \frac{I_0 N}{(N - I)e^{-r\beta t} + I_0} \tag{3}$$

3.2 Data

For the years 2012-2015, we analyzed published data from the Centers for Disease Control (CDC) regarding the weekly number of infections for influenza type A virus strains [2]. Each data set starts in the fall of one year and overlaps into the next year ending 52 weeks later. All data for years 2012-2015 are shown below (Figures 2-5).
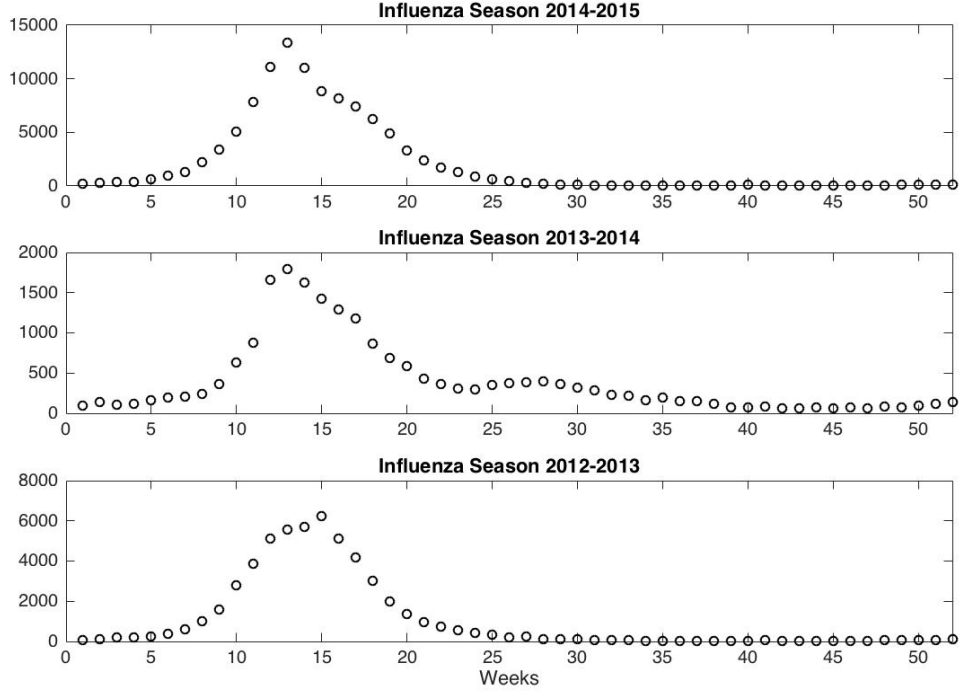
3

Figure 2: A number of infections vs. time plot for each influenza season through the years 2012-2015. Top graph: Influenza season 2014-2015. Middle graph: Influenza season 2013-2014. Bottom graph: Influenza season 2012-2013. Note: All data only includes infections from type A virus strains.

3.3   Determining Parameters for r and β

From the CDC data for each influenza season between 2012 and 2015, we utilized MATLAB to computationally iterate through each data set, fit a curve to the data, and store the resulting smoothing spline function in a MATLAB cell `curve_fits`. We chose the most recent data (2014-2015 influenza season) to generate parameters for r and β in equation (3). Now having a continuous curve `curve_fits{3}` for the 2014-2015 influenza season, we had MATLAB integrate the function `si15` to produce an SI logistic growth curve, which makes sense mathematically (i.e. we found the area under the curve of the 2014-2015 infection data which should generate a converging logistic growth curve).

```matlab
%% Integrate SIR function for 2015 Influenza.
x_data = 1:52;
si15 = integrate(curve_fits{3}, x_data, 0);
```

Next, we fitted equation (3) (with $I_0 = 197$ and $N = 106500$, the values from the 2014-2015 CDC data) on top of `si15` to computationally solve for r and β.

```matlab
%% Fit theoretical SIR to CDC SIR.
[xData, yData] = prepareCurveData( x_data, si15 );

% Set up fittype and options.
ft2 = fittype( '(106500*197)/((106500-197)*exp(-r*b*t)+197);', 'independent',
    't', 'dependent', 'I' );
opts = fitoptions( 'Method', 'NonlinearLeastSquares' );
opts.Display = 'Off';
opts.StartPoint = [6 0.4];

% Fit model to data.
[si15_approx, si15_approx_gof] = fit( xData, yData, ft2, opts );

% Display coefficient values.
r = si15_approx.r;
b = si15_approx.b;
```

We found that the values for r and β were approximately 1.5912 and 0.2749 respectively. More accurately, the coefficient value of t in equation (3) should be approximately 0.4374 (the product of r and β). The product of the two numbers is a much more accurate statement because there exists an infinite amount of numbers that can take the place of r and β yet still produce the same product value. This result suggests the number of contacts an individual has per day is 1.5912 and the probability a contact will be infected is 0.4374%. Figure 6 below shows the theoretical SI function $I(t)$ (top left corner) with a coefficient value of 1, the SI function produced from integrating the 2014-2015 CDC data `si15` (top right corner), and the theoretical SI function with corrected parameters on top of the CDC data (lower half of the figure).
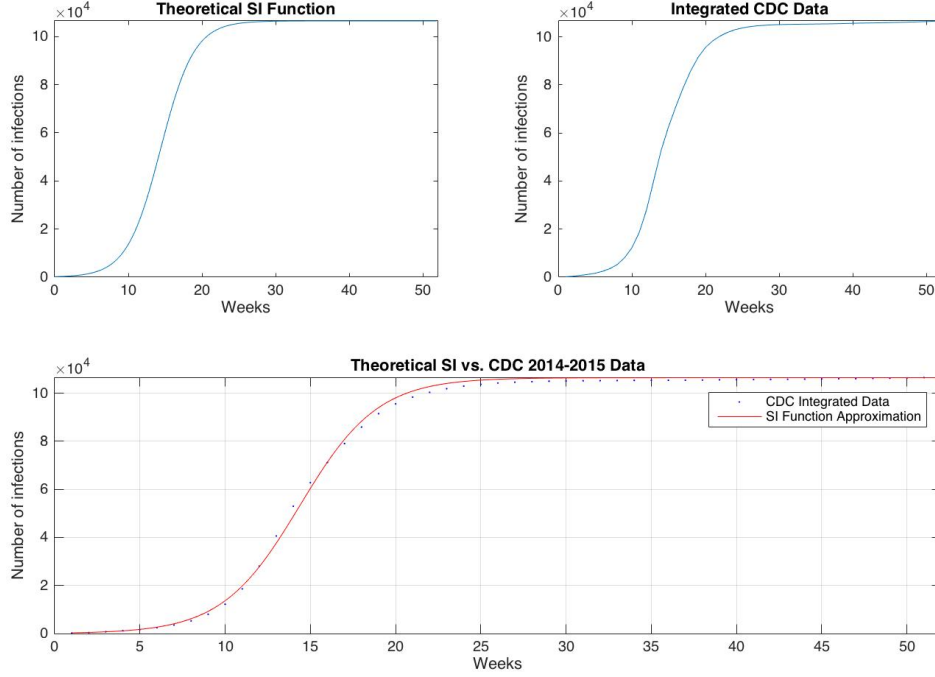
5

Figure 3: Top left corner: the graph of the theoretical SI function represented by equation (3). Top right corner: the graph produced by integrating the fitted curve for the 2014-2015 data (i.e. the function `curve_fits3`). Lower half: the theoretical SI function (equation (3)) with corrected parameter values plotted against the CDC data for the 2014-2015 influenza season.

3.4   Results

After attaining parameter values for r and β (with $>95\%$ confidence), we analytically differentiated equation (3),

$$I(t) = \frac{NI_0}{(N - I_0)e^{-r\beta t} + I_0} \tag{3}$$

to produce a theoretical SIR function,

$$I'(t) = \frac{dI}{dt} = \frac{NI_0(N - I_0)r\beta e^{-r\beta t}}{\left((N - I_0)e^{-r\beta t} + I_0\right)^2}. \tag{4}$$

In order to test the hypothesis that an SI model can be applied to estimating the dynamics of an SIR disease, such as influenza, equation (4) was plotted against the CDC data representing the three previous influenza seasons (years 2012-2015). Equation (4) was normalized and plotted against each respective data set using parameter values found above, represented by figure 6 below.
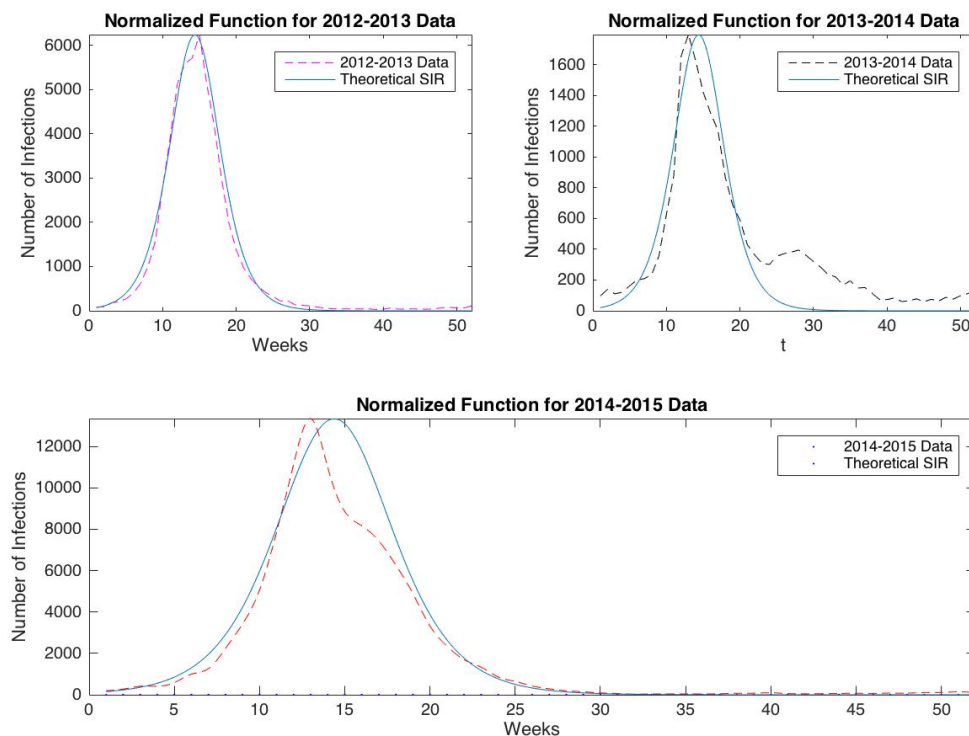

Figure 4: Top left corner: a normalized theoretical SI function plotted against the CDC data curve for the 2012-2013 influenza season. Top right corner: normalized theoretical SI function plotted against the CDC data curve for the 2013-2014 influenza season. Lower half: a normalized theoretical SI function plotted against the CDC data curve for the 2014-2015 influenza season.

4 Discussion

Our findings suggest there is more applications for a simple SI model than simply just modeling a disease that only breaks a population into a susceptible and infectious group.

The results shown in figure 6 clearly demonstrates that equation (4) is a very good approximation to the CDC infection curves from the past three years of seasonal type A influenza data. As the United States enters the influenza season for 2015-2016, we predict our theoretical SIR model should offer a reliable estimation for the number of type A influenza infections as a function of time $I'(t)$, given the initial amount of infected people at time zero $I(0)$ and the total population size expected to be influenced by the disease N.

Our team attempted to find the parameter values for r and β by using what is known as the "reproduction number." The reproduction number for a disease is the total number of secondary cases a single infectious host will produce [3]. It is a product between the number of secondary cases from a host per unit of time, and the amount of time for which an individual is infected. Mathematically, the reproduction number R_0 can be represented by $R_0 = (r\beta)(T)$ where T is the length of time a host remains infected. Using the CDC's estimate of eight days for T [4], and influenza's average reproduction number of 1.28 [5], we solved for the product of r and β. This gave us a product value of 0.16 which is significantly lower than the computationally produced product value of 0.4374. A product value of 0.16 would generate an infection curve with an extremely delayed growth when compared to the data published by the CDC. Contrary to solving for $r \cdot \beta$, we also tried solving for the mean infectious period using our previously found product value from section 3.2 and the average reproduction number for influenza. This produced a mean infectious period of 2.93 days, a 63.38% error from the CDC's data. This error can most likely be attributed to the fact that the product between the two parameters are being used in an SI model without having been adjusted from the SIR model of which they were obtained. Further evaluation should be performed to determine how the parameter values can adjusted in order to fit the compartment model being used.

Although our SI model was able to accurately approximate the infection curve for type A influenza through the years 2012-2015, further research should be performed to revise the model as influenza continuously evolves over time. In addition, the limitations of an SI model's application to diseases with multiple health states is not quite clear from our findings, and the extent at which an SI model can be applied should be further studied.

5 References

[1] Chitnis, Nakul. *Introduction to Mathematical Epidemiology: Deterministic Compartmental Models*, Swiss Tropical and Public Health Institute. 1 (2011): 12.

[2] Centers for Disease Control and Prevention, 2015. Public-use data file and documentation. http://gis.cdc.gov/grasp/fluview/fluportaldashboard.html.

[3] Chitnis, Nakul. *Introduction to Mathematical Epidemiology: The Basic Reproduction Number*, Swiss Tropical and Public Health Institute. 1 (2011): 5.

[4] Centers for Disease Control and Prevention. N.D.. Centers for Disease Control: N.P.; [2015 Nov 20; 2015 Nov 20]. http://www.cdc.gov/flu/about/disease/spread.htm.

[5] G., Chowewll, et al. *Seasonal Influenza in the United States, France, and Australia: transmission and prospects for control.* 1 (2007): 13.