Analysis of Influenza-Like Illness Outbreaks at ETSU

Byron Roland and Caleb Shimberg,
East Tennessee State University, Johnson City, TN, 37614

Abstract

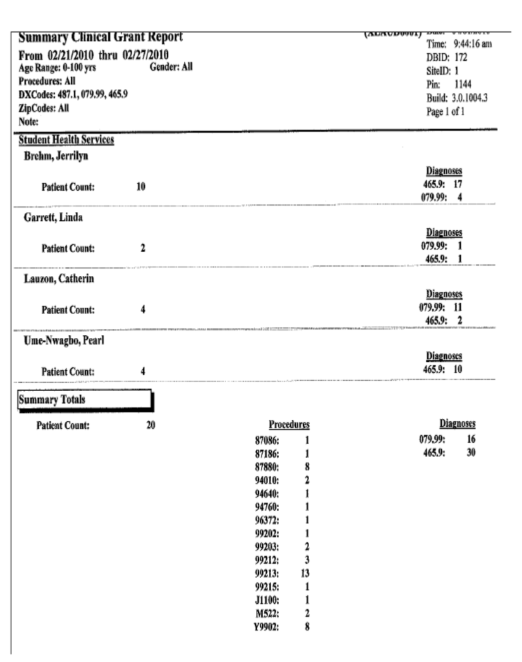
During the course of seven months, data was collected from the student health clinic located on the East Tennessee State University campus. The clinic reported the number of patients with influenza-like illnesses seen by each nurse in the facility. A model for the spread of this influenza-like illness was created using a basic single outbreak SIR model. The differential equations defining the SIR model were solved numerically using a built-in MATLAB function called ode45 (based off an explicit Runge-Kutta(4,5) integration method). In this system of equations a fourth was created to report the incidence rate of the influenza-like illness. This incidence rate is what we are going to fit to our data in order to determine transmission rates, reproductive rates, and recovery rates. In this there are two models, one with constant parameters and another with a time-dependent transmission rate. In the second model, we used another MATLAB function called Pchip (Piecewise Cubic Hermite Interpolating Polynomial) to interpolate the values of transmission rate over the time of the data. This interpolation allows us to vary the number of interpolating values of transmission and to explore subintervals for major shifts.

Introduction

In this poster we will examine the two models and further explain the optimization of each model using the different global optimization routines. Also, the importance of interpolating the beta values and the incidence of the model being fit to the data. Our model (being a non-linear system) does not give us a generic analytic solution, so we must use multiple functions in MATLAB to help us accomplish our goal.

Data

The data collected was taken from an on-campus health clinic. The observance of this data is from October of 2009 up to April of 2010. This clinic is open during regular school hours and is closed during the holidays. In the below graph you can see the time series of the data. This graph shows a large drop in number of infected students sometime around December. We suspect this is because the school is no longer in session during this time so not many people go to this clinic. In the picture on the right an example sheet of what the data came from, these students are not actually infected with the flu but rather have influenza-like symptoms and this is what we are looking at, all the students that have influenza like symptoms.


Summary Clinical Grant Report
From 02/21/2010 thru 02/27/2010
Age Range: 0-100 yrs Gender: All
Procedures: All
DXCodes: 487.1, 079.99, 465.9
ZipCodes: All
Note:
Time: 9:44:16 am
DBID: 172
SiteID: 1
Pin: 1144
Build: 3.0.1004.3
Page 1 of 1

Student Health Services

Brehm, Jerrilyn
Patient Count: 10
Diagnoses
465.9: 17
079.99: 4

Garrett, Linda
Patient Count: 2
Diagnoses
079.99: 1
465.9: 1

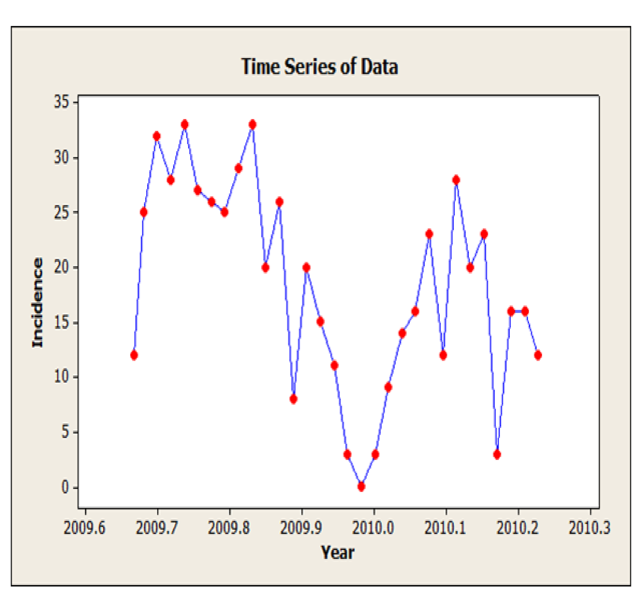
Lauzon, Catherin
Patient Count: 4
Diagnoses
079.99: 11
465.9: 2

Ume-Nwagbo, Pearl
Patient Count: 4
Diagnoses
465.9: 10

Summary Totals
Patient Count: 20

Procedures	
87086:	1
87186:	1
87880:	8
94010:	2
94640:	1
94760:	1
96372:	1
99202:	1
99203:	2
99212:	3
99213:	13
99215:	1
J1100:	1
M522:	2
Y9902:	8

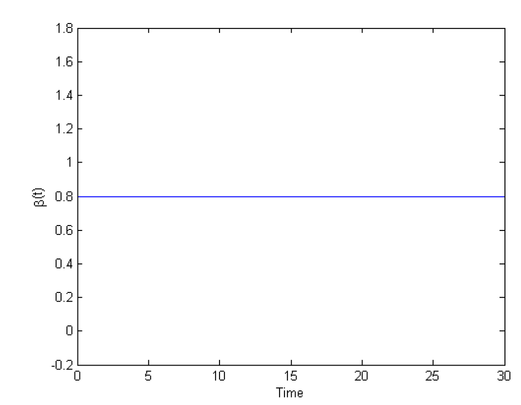
Diagnoses	
079.99:	16
465.9:	30

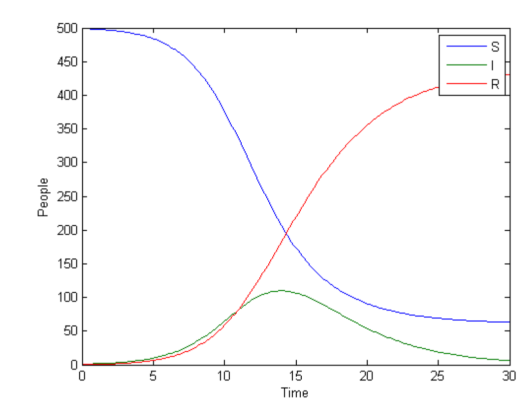


Single Outbreak SIR Model with Constant Parameters

The first model was developed using an SIR model that had parameters not dependent on time.

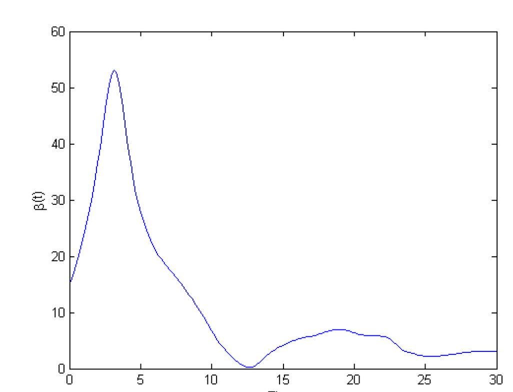
$$dS/dt = -\beta SI/N$$
$$dI/dt = \beta SI/N - \gamma I$$
$$dR/dt = \gamma I$$

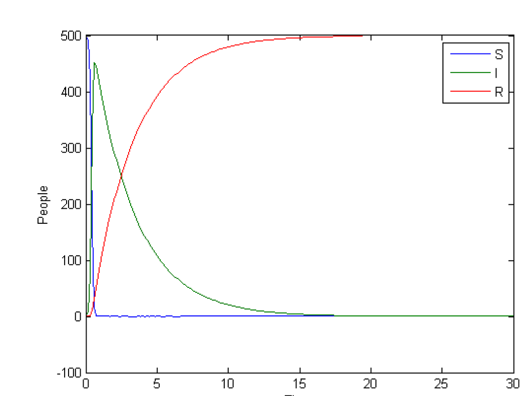




Single Outbreak SIR Model with Time-Dependent Transmission Rate

The second model was developed using beta values that were dependent on time, each time point corresponded to a different beta value (see figure below).





Global Optimization Routines

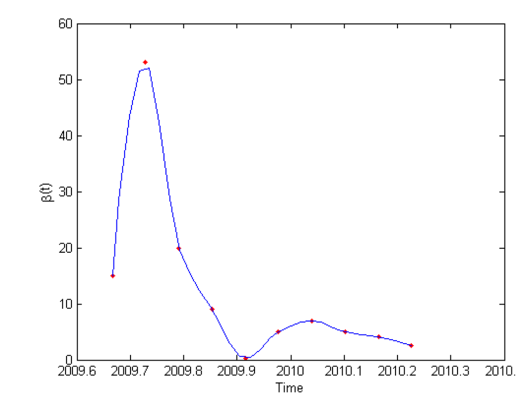
These routines were specifically designed to optimize non-linear systems using constants and/or time-dependent parameters.
PyMinuit – Python optimization for minimizing general N-dimensional functions [1]
Genetic Algorithm – MATLAB optimization mimicking the principles of biological evolution [2]

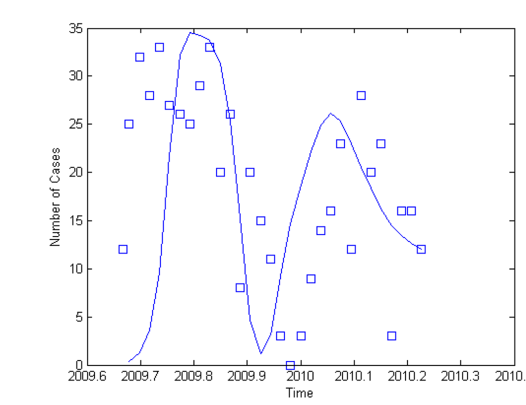
Optimization Problem

In this we used the Global Optimization Toolbox function GA to solve for θ_{OLS}, which is the argmin over θ for $\sum_{i=1}^{n}[Y_i - I(t_i,\theta)]^2$. Below are graphs of the running optimization.

Best Fit Curve

This is the solution to the best fit of the optimization of the ETSU health dataset. See "Ongoing Research" for what is to come later.





Ongoing Research

As an ongoing part of our research we plan on learning more about a Python library called PyMinuit that has a built in optimization algorithm similar to the Global Optimization toolbox in MATLAB. We plan on generating a synthetic dataset as a placebo to further our knowledge in Python then use a random number generator from a normal distribution to create "noise" in our dataset so as to be similar to a real dataset. Then use this PyMinuit in Python to optimize this synthetic dataset. We then plan on using the PyMinuit library to optimize this ETSU dataset, specifically optimizing the beta values (in the time-dependent transmission rate SIR model) and the number of tau subintervals. We will also address uncertainty estimation. Also, effective reproductive number estimation.

References

[1] "Pyminuit - Minuit Numerical Function Minimization in Python - Google Project Hosting."*Google Code*. Google, 2011. Web. 22 July 2011. <http://code.google.com/p/pyminuit/>.
[2] "Genetic Algorithm Solver - Global Optimization Toolbox for MATLAB & Simulink."*MathWorks - MATLAB and Simulink for Technical Computing*. The Mathworks Inc., 2011. Web. 22 July 2011. <http://www.mathworks.com/products/global-optimization/description4.html>.

Acknowledgements

B. Roland was funded by ETSU Honors College through a Research Discovery Student position. The data collection was funded by an RDC Small Grant RDC 10-020sm 212665.