Math 420, Spring 2024
First Team Homework: SI and SIR Modeling

Consider the dataset associated to your team. The problems are stated with Matlab indexing convention: from 1 to the length of the vector. The algorithms presented in class use the natural convention, where time $t_0 = 0$ may be indexed by 0. Be careful to implement algorithms correctly, regardless of (internal to your IDE) indexing convention.

Exercise 1. This problem asks you to fit a SI model to the time series of the *cumulative number* of detected infections. Let $T_{max} = 119$.

(1) Load your assigned data as a vector $v = (v(t))$. Identify the first time (date) t_0 when $v(t_0) \geq 5$, that is, the number of detected cases is at least 5. That time represents the starting time in your simulation. Print t_0.

(2) Let $I(t) = v(t+t_0)$, for $0 \leq t \leq T_{max}$. In otherwords create a vector of length $T_{max}+1$ days of daily infection rates starting from the date at least 5 infections hae been detected. Use 'Population' column from your data set (or https://www.census.gov/data/tables/time-series/demo/popest/2020s-counties-total.html) to determine the total population of the county/city associated to your assigned data set. Let N_{max} denote this county population. Denote by $N_{min} = 1 + I(T_{max})$ the maximum infected population based your data.

(a) Implement Algorithm "SI Alg 1 - known N". Run this algorithm for $N = N_{min}$ and $N = N_{max}$, and for each of the two values of N:
(i) Print the estimated $\beta = \hat{\beta}$ and the value of objective function $J(\beta, N)$ defined in (**), slide 'How to Calibrate SI Models (2)'.
(ii) Plot in the same figure $I(t)$ and the predicted value of the infection count based on the SI model, i.e. $NI(0)/(I(0) + (N - I(0))e^{-\beta t})$.

(b) Implement Algorithm "SI Alg 2 - Unknown N". Run on this data and:
(i) Save intermediate values of $J(N)$ computed at step 2.1. Plot the graph $J = J(N)$ of these intermedate results.
(ii) Print the estimates $N = \hat{N}$ and $\beta = \hat{\beta}$ as wel as the value of objective function $J(\beta, N)$ defined in (**), slide 'How to Calibrate SI Models (2)'.
(iii) Plot in the same figure $I(t)$ and the predicted value of the infection count based on the SI model, i.e. $NI(0)/(I(0) + (N - I(0))e^{-\beta t})$ at the stopping value of N.
(iv) Can you run step 2.1 for all values of N from N_{min} to N_{max} estimated before? If so, plot $J = J(N)$ and determine the global minimum on this interval. Does it differ from part (b.ii)?

(c) Implement the following algorithm: For each value of N considered at part (2.b), compute the optimal $\beta(N) = \hat{\beta}$ according to Algorithm SI Alg 1 - Known N, and then compute the "ideal" objective function $I(N, \beta(N))$ displayed at the bottom of slide "How to Calibrate SI Models". Plot the function $N \mapsto I(N, \beta(N))$ and determine its minimum. Does it match your findings at part (b.ii)?

Exercise 2. This problem asks you to fit a SIR model to the time series of *detected infections* and the *cumulative deaths*. For this purpose the unknown parameters are $\alpha, \beta, \gamma, \rho$. Use two different cost parameters, (c_I, c_Y). In one case choose $(c_I, c_Y) = (0, 1)$, in the other case choose $(c_I, c_Y) = (1, 1)$.

The vector $I(t)$ in this problem represents the sequence of *detected infections*, and the vector $Y(t)$ is the sequence of cumulative deaths which you extract from the same data file. In previous problem the same notation is used for the cumulative number of detected infections.

First load the cumulative number of detected infections $v(t)$ and detect the first time t_0 when $v(t_0) \geq 5$ as in Exercise 1. Let $T_{max} = 119$.

(1) Create the *rates of detected infections* $I(t)$ using a difference formula: $I(t) = v(t + t_0 + \tau) - v(t + t_0 - \tau)$, for $0 \leq t \leq T_{max}$. The parameter τ is related to incubation and infection period. Set $\tau = 7$ days for now (the assumption is that the infection lasts up to 14 days). Plot $I = I(t)$, the rate of detected infections. Create the time series $Y(t)$ by extracting the third row from your data set and align with the first time t_0. Plot the cumulative deaths $Y(t)$ as well.

(2) Implement an Euler scheme for solving the SIR Model with step size $h = 0.01$ (i.e., 100 steps per day). Denote by $(S_{sim}(t), I_{sim}(t), R_{sim}(t))$ the numerical solution. Use initialization $S(0) = N$, $I(0)$ from the data set, $R(0) = 0$. For this part, the unknown parameters are α, β.

(3) For each combination (α, β) in the set Ω described below repeat :
   (a) Run your numerical solver and produce $I_{sim} = (I_{sim}(t))$ and $R_{sim} = (R_{sim}(t))$.
   (b) Find the least-square fit for γ and ρ in optimization problems:

$$\hat{\rho} = argmin_{\rho} \sum_{t=0}^{T_{max}} |I(t) - \rho I_{sim}(t)|^2 \ , \ \hat{\gamma} = argmin_{\gamma} \sum_{t=0}^{T_{max}} |Y(t) - \gamma R_{sim}(t)|^2$$

   (c) Compute the l^2-norm squared of the residuals and save it in an array indexed by the two parameters:

$$J(\alpha, \beta) = c_I \sum_{t=0}^{T_{max}} |I(t) - \hat{\rho} I_{sim}(t)|^2 + c_Y \sum_{t=0}^{T_{max}} |Y(t) - \hat{\gamma} R_{sim}(t)|^2$$

(4) Visualize the function J by plotting it as a two-dimensional surface. Determine where the minimum of this function occurs (over the finite set of values considered above).

For this problem, construct the set Ω as follows:

- For Covid-19, the infectious period ranges from 5 to 20 days. Hence α should be in the range $[\frac{1}{20}, \frac{1}{5}]$. Generate a vector $(0.05 : 0.01 : 0.2)$
- For each value of α, choose $\beta = R_0 \alpha$, where R_0 is the reproduction ratio (or the contact number). Consider values of R_0, in the range $(0.8, 2.2)$ with a stepsize 0.05, $R_0 \in (0.8 : 0.05 : 2.2)$.

This gives a set Ω of about 450 pairs (α, β).

Depending upon your findings, you may consider refining a certain range of parameters (divide-and-conquer strategy).

Run steps 3 and 4 for the two scenarios, $(c_I, c_Y) = (0, 1)$ and $(c_I, c_Y) = (1, 1)$.

Can you adjust $I(0)$, the starting value of the numerical scheme, to get a better fit?