Math 420, Spring 2023
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 pf 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 consdered 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 the *active populations* of detected infections. For this purpose the unknown parameters are $\alpha, \beta, N, S(0), I(0)$. NOTE: The vector $I(t)$ in this problem represents the sequence of *active detected infections*!

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 active infection* $I(t)$ 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 rates of active infection.

(2) Implement an Euler scheme for solving the SIR Model with step size $h = 0.01$. 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 problem, the unknown parameters are α, β, N.

(3) For each combination (α, β, N) in the set Ω described below repeat :
   (a) Run your numerical solver and produce $I_{sim} = (I_{sim}(t))$.
   (b) Compute the l^2-norm squared of the residuals and save it in an array indexed by the three parameters:

$$J(\alpha, \beta, N) = \sum_{t=0}^{T_{max}} |I(t) - I_{sim}(t)|^2$$

(4) Visualize the function J by plotting two-dimensional surfaces $(\beta, N) \mapsto J(\alpha, \beta, N)$ for each value of α. In particular 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 10 days. Hence α should be in the set $[\frac{1}{10}, \frac{1}{5}]$. Consider five values: $\{\frac{1}{10}, \frac{1}{9}, \frac{1}{8}, \frac{1}{7}, \frac{1}{6}, \frac{1}{5}\}$.
- For each value of α, choose $\beta = R_0\alpha$, where R_0 is the reproduction ratio (or the contact number). Consider using 12 values of R_0, $R_0 \in \{0.8, 0.9, 0.95, 1, 1.05, 1.1, 1.15, 1.2, 1.3, 1.4, 1.5, 1.6\}$.
- For N use fractions of the total population of the county associated to the data set assigned to your team. For instance, if N_{max} denotes this county population, try 10 values for $\frac{N}{N_{max}}$: $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$ (in other words, all multiples of $N_{max}/10$).

This gives a set Ω of about 600 triplets (α, β, N).

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