Lecture 9: Full Calibration of SIR Models

Radu Balan

Department of Mathematics, NWC
University of Maryland, College Park, MD

Version: February 23, 2023

SIR Model Calibration

Recall the model:

$$\left\{\begin{array}{lll} \frac{dS}{dt} & = & -\beta S \frac{I}{N} \quad , \quad S(0) \\ \frac{dI}{dt} & = & \beta S \frac{I}{N} - \alpha I \quad , \quad I(0) \\ \frac{dR}{dt} & = & \alpha I \quad , \quad R(0) \end{array}\right.$$

with the sub-compartments $X(t) = (1-\gamma)R(t)$ for "recovered" and $Y(t) = \gamma R(t)$ for deaths.

Before making useful predictions (testing), the model has to be calibrated. Last time we analyzed estimators for γ from the time series of *cumulative detected infections*, $\{V(0), \cdots, V(T_{max})\}$, and the time series of *cumulative deaths*, $\{Y(0), \cdots, Y(T_{max})\}$. The idea was to minimize over $\tau \geq 0, \gamma \in [0,1]$ the norm $\|Y(\cdot + \tau) - \gamma V\|_p$, scaled by the number of terms in each sum. Now we integrate these estimators into a scheme that performs full calibration of the SIR model.

Optimization based on Forward Model Simulation

Pre-processing steps

Step 1: t_0. First detect the onset of infections, and resent the time origin to match this onset t_0.

Step 2: $I(t)$. The time series of cumulative detected infections $\{V(0), \cdots, V(T_{T_{max}})\}$ should be turned into a rate of infections. The daily rate may not be relevant to infection transmissions. Instead use a time window to convert the cumulative count into a rate:

$$I(t) = V(t + t_0 + \tau_0) - V(t + t_0 - \tau_0) \quad , \quad t = 0, 1, 2, \cdots, T_{max}$$

where $\tau_0 > 0$ is chosed so that τ_0 accounts for average infection period. For Covid-19 it is somewhere between 5 days and 10 days. A possible value is $\tau_0 = 7$. Thus $I(t)$ measures the number of invections in a 2-week period centered around t.

Optimization based on Forward Model Simulation

The Meta Loop

A natural choice for initial conditions is given by: $S(0) = N$ and $R(0) = 0$, where we assumed $I(0) \ll N$:

$$\left\{ \begin{array}{rcl} \frac{dS}{dt} & = & -\beta S \frac{I}{N} \quad , \quad S(0) = N \\ \frac{dI}{dt} & = & \beta S \frac{I}{N} - \alpha I \quad , \quad I(0) > 0 \\ \frac{dR}{dt} & = & \alpha I \quad , \quad R(0) = 0 \\ Y & = & \gamma R \end{array} \right.$$

At this point, the parameters that need to be estimated are: $\{\alpha, \beta, \gamma, N\}$.
The Forward Model based calibration works like this:
1. Construct a search set Ω of the "free" parameters (α, β, N)
2. For each triple $(\alpha, \beta, N) \in \Omega$: (i) run a SIR simulator using the Euler scheme that produces $(S_{sim}, I_{sim}, R_{sim})$. (ii) Fit $\hat{\gamma}$ to match the observed time series Y. (iii) Compute the value of the objective function $J(\alpha, \beta, N, \hat{\gamma})$
3. Select $(\hat{\alpha}, \hat{\beta}, \hat{N})$ that minimize J.

Optimization based on Forward Model Simulation

Details: The objective function

The objective function J can be chosen to measure a norm of residuals $I - I_{sim}$ and $Y - Y_{sim}$, where $Y_{sim} = \hat{\gamma} R_{sim}$

To do so, you need to choose and fix two meta-parameters p and λ. For $1 \leq p < \infty$ define

$$J(\alpha, \beta, N, \gamma) = \sum_{t=0}^{T_{max}} |I(t) - I_{sim}(t)|^p + \lambda \sum_{t=0}^{T_{max}} |Y(t) - \gamma R_{sim}(t)|^p$$

For $p = \infty$ define

$$J(\alpha, \beta, N, \gamma) = \max_{0 \leq t \leq T_{max}} |I(t) - I_{sim}(t)| + \lambda \max_{0 \leq t \leq T_{max}} |Y(t) - \gamma R_{sim}(t)|$$

The meta-parameter λ weights one term of error measure (the rate of infections mismatch) against the other term (the number of deaths).

Typical choices: $p = 1, 2, \infty$ and $\lambda = 1$.

The Calibration Algorithm for SIR Models

Algorithm (Meta-Algorithm for SIR Calibration)

Inputs: Time series $\{V(0), \cdots, V(T)\}$, $\{Y(0), \cdots, Y(T)\}$. Parameters: V_{min} (default, $V_{min} = 5$), τ_0 (default, $\tau_0 = 7$), $p \in [1, \infty]$ (default, $p = 2$), $\lambda > 0$ (default, $\lambda = 1$), Search set Ω.

1. Detect the onset of the infection t_0 as the first time so that $V(t_0) \geq V_{min}$, reset the time origin, and create the time series of infection rates $I(t) = V(t + t_0 + \tau_0) - V(t + t_0 - \tau_0)$, $0 \leq t \leq T_{max}$.
2. For each $(\alpha, \beta, N) \in \Omega$ repeat:
   1. Simulate a SIR model with parameters (α, β, N) and obtain daily time series $(S_{sim}, I_{sim}, R_{sim})$.
   2. Solve $\hat{\gamma} = argmin_{\gamma} \|Y - \gamma R_{sim}\|_p$.
   3. compute the objective function J.
3. Determine the minimum and the minimizer of J.

Outputs: Estimated $\hat{\alpha}, \hat{\beta}, \hat{N}, \hat{\gamma}$ and minimum value of the objective function J_{min}.

Numerical results (1)

Analysis of DC data for 2020

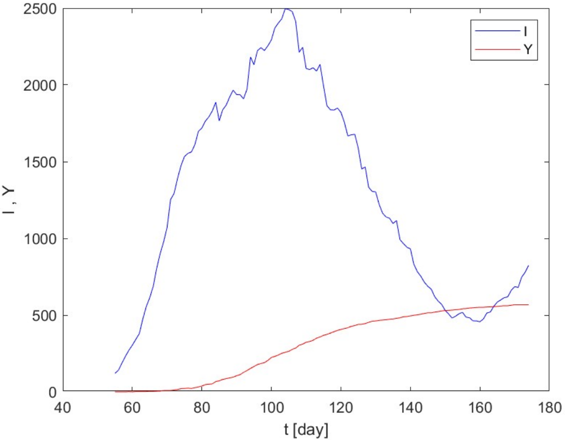
$V_{min} = 5$, $\tau_0 = 7$ [days]. Onset $t_0 = 55$ (Monday: March 16, 2020), $T_{max} = 119$.

Notable dates:

Local dips: day=152 (Sunday: 6/21/2020), day=159 (Sunday: 6/28/2020).

Memorial weekend: Sunday May 24, 2020: day = 124 .

4th of July: Saturday July 4, 2020: day = 165.



Numerical results (2)

Analysis of DC data for 2020. $p = 1$

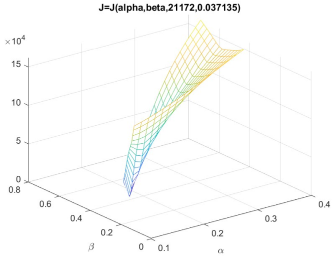
Results:
$\hat{\alpha} = 0.12$, $\hat{\beta} = 0.21$,

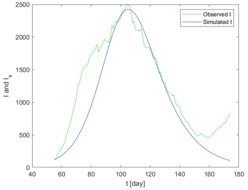
$\hat{R}_0 = 1.75$,

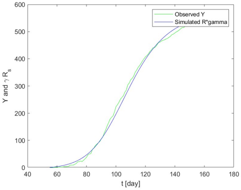
$\hat{N} = 21172$,
$\frac{\hat{N}}{CityPopulation} = 3\%$,

$CityPopulation = 705749$.

$\hat{\gamma} = 3.7\%$.







Numerical results (3)

Analysis of DC data for 2020. $p = 2$

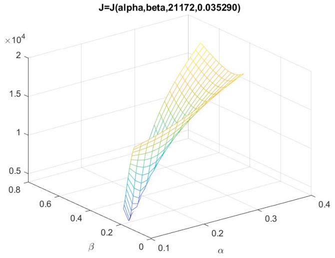
Results:
$\hat{\alpha} = 0.12$, $\hat{\beta} = 0.216$,

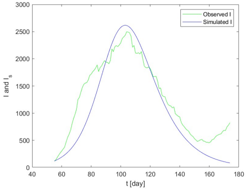
$\hat{R}_0 = 1.8$,

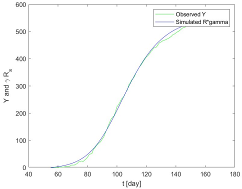
$\hat{N} = 21172$, $\frac{\hat{N}}{CityPopulation} = 3\%$,

$CityPopulation = 705749$.

$\hat{\gamma} = 3.5\%$.







Numerical results (4)

Analysis of DC data for 2020. $p = \infty$

Results:

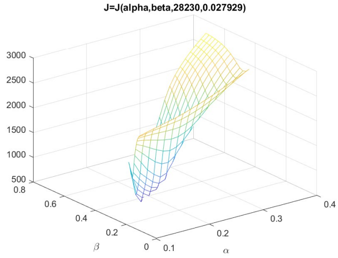
$\hat{\alpha} = 0.14$, $\hat{\beta} = 0.238$,

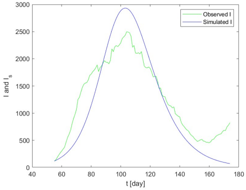
$\hat{R}_0 = 1.7$,

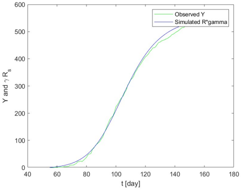
$\hat{N} = 28230$, $\frac{\hat{N}}{CityPopulation} = 4\%$,

$CityPopulation = 705749$.

$\hat{\gamma} = 2.8\%$.







Next Model: SEIR

The Susceptible-Exposed-Infected-Removed (SEIR) model is obtained from SIR by introducing a compartment between Susceptible and Infected of population that has been exposed to the virus but are not yet contagious:

$$\left\{ \begin{array}{rcl} \frac{dS}{dt} & = & -\beta\frac{SI}{N} \quad , \quad S(0) \\ \frac{dE}{dt} & = & \beta\frac{SI}{N} - \delta E \quad , \quad E(0) \\ \frac{dI}{dt} & = & \delta E - \alpha I \quad , \quad I(0) \\ \frac{dR}{dt} & = & \alpha I \quad , \quad R(0) \end{array} \right. \qquad \text{(SEIR Model)}$$

where $\delta \geq 0$ is the rate of transition from exposed to infected. Its reciprocal $1/\delta$ represents the *average incubation period*.
If data is selected from the onset of infections, a natural initial condition is: $R(0) = N \gg I(0)$, $R(0) = 0$. The initial exposed population $E(0)$ may be set to $I(0)$, or can be fine tuned to fit the data. Assumin $E(0) = I(0)$ is known, the parameters that need to be calibrated are: $\alpha, \beta, \gamma, \delta, N$.