Lecture 8: Calibration of SIR Models

Radu Balan

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

Version: February 21, 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. For calibration and testing we are using two pieces of measured data: the *cumulative detected infections*, $\{V(0), \cdots, V(T_{max})\}$, and the time series of *cumulative deaths*, $\{Y(0), \cdots, Y(T_{max})\}$. The cumulative detected infections will have to be converted into infection rates $\{I(0), \cdots, I(T_{max})\}$. Note that, if we know γ and N we can compute $R(0) = \frac{Y(0)}{\gamma}$ and $S(0) = N - I(0) - R(0)$. At the onset of an infectious disease it is the likely the case that $Y(0) = 0$ and $I(0)$ can be neglected in which case, $S(0) = N$ (regardless of γ). This approximation may hold for a certain interval of

The LSE

The least-squares estimator (LSE) tries to find parameters α, β, γ, and N that minimize:

$$\begin{aligned} \text{minimize}_{\substack{N \in \mathbb{N} \\ \alpha,\beta,\gamma \geq 0 \\ \gamma \leq 1}} \; J(\alpha,\beta,\gamma;N) \;,\; J(\alpha,\beta,\gamma;N) := \sum_{t=0}^{T_{max}} (I(t) - I_{sim}(t))^2 + (Y(t) - \gamma R_{sim}(t))^2 \end{aligned}$$

where $(S_{sim}(t), I_{sim}(t), R_{sim}(t))$ are given by a numerical solver of the SRI model with parameters (α, β, γ) and total population N initialized at $(S(0), I(0), R(0))$.

The LSE

The least-squares estimator (LSE) tries to find parameters α, β, γ, and N that minimize:

$$\begin{array}{c} \text{minimize} \\ N \in \mathbb{N} \\ \alpha, \beta, \gamma \geq 0 \\ \gamma \leq 1 \end{array} \; J(\alpha, \beta, \gamma; N) \; , \; J(\alpha, \beta, \gamma; N) := \sum_{t=0}^{T_{max}} (I(t) - I_{sim}(t))^2 + (Y(t) - \gamma R_{sim}(t))^2$$

where $(S_{sim}(t), I_{sim}(t), R_{sim}(t))$ are given by a numerical solver of the SRI model with parameters (α, β, γ) and total population N initialized at $(S(0), I(0), R(0))$. However ... this is a hard optimization problem ... and potentially not useful!

We shall analyze different approaches to calibrate the SIR model.

First, the initial time $t_0 = 0$ is chosen once a significant number of infections occured.

γ estimators

We start with the "simpler" problem of estimating γ.
Assume $\{V(0), V(1), \cdots, V(T_{max})\}$ denotes the cumulative number of detected infections, and $\{Y(0), Y(1), \cdots, Y(T_{max})\}$ denote the time series of virus related deaths. It is necessary that
$0 \leq Y(t) \leq V(t) \leq V(T_{max})$ for every $0 \leq t \leq T_{max}$.
Since all infected individuals eventually transit into the "removed" state, $R(t')$, **for calibration purposes** we make the assumption that $V(t) \approx R(t+\tau)$ for some $\tau > 0$. In fact, τ should be close to $\frac{1}{\alpha}$. In this case we obtain: $Y(t+\tau) \approx \gamma V(t)$. A natural optimization problem is to minimize a norm of the difference between $\gamma V(t)$ and $Y(t+\tau)$. Consider $1 \leq p < \infty$ and define

$$F(\gamma, \tau; p) := \frac{1}{T_{max} - \tau + 1} \sum_{t=0}^{T_{max}-\tau} |Y(t+\tau) - \gamma V(t)|^p$$

l^p estimators for γ

For $p = \infty$ adjust the definition:

$$F(\gamma, \tau; \infty) := \max_{0 \leq t \leq T_{max} - \tau} |Y(t + \tau) - \gamma V(t)|$$

Then consider the optimization problem:

$$\begin{array}{cc} \text{minimize} & F(\gamma, \tau; p) \\ \tau, \gamma \geq 0 \quad , \quad \gamma \leq 1 & \end{array}$$

for the given calibration data set. In the following we analyze the cases $p = 1, 2, \infty$. In each case, the optimization problem minimizes an l^p norm of the form $\|Y(\cdot + \tau) - \gamma V\|_p$, scaled by the number of terms in each sum.

Good news: The optimization problem is convex. The bad news: Given τ, except for $p = 2$, in the other cases the optimization problem does not have a closed form solution, but can be easily solved.

l^p estimators for γ

The general optimization problem is solved by an iterative algorithm:

Algorithm (Meta-Algorithm for γ estimation)

Inputs: *Time series* $\{V(0),\cdots,V(T_{max})\}$, $\{Y(0),\cdots,Y(T_{max})\}$.

Parameters: $p\in[1,\infty]$, τ_{max}.

1. *For each* $\tau=0,1,2,...,\tau_{max}$ *repeat:*
   1. *Solve* $[Fmin,\gamma_{min}]=min_{\gamma\in[0,1]}F(\gamma,\tau;p)$.
   2. *Save* $vectorF(\tau)=F(\gamma_{min},\tau;p)$, $optGamma(\tau)=\gamma_{min}$.
2. *Determine the minimum and the minimizer* $[minF,\hat{\tau}]=min(vectorF)$
3. *Assign* $\hat{\gamma}=optGamma(\hat{\tau})$, $F(\hat{\gamma},\hat{\tau};p)=minF$.

Outputs: *Estimated* $\hat{\gamma},\hat{\tau}$ *and minimum value of the objective function* $F(\hat{\gamma},\hat{\tau};p)$.

Next we analyze the Step 1.1.

The case $p = 2$

The case $p = 2$ is the easiest: it is solved by the least-squares fit with a linear model. Solution of

$$\underset{\gamma}{\text{minimize}} \sum_{t=0}^{T_{max}-\tau} |Y(t+\tau) - \gamma V(t)|^2$$

is given by:

$$\gamma_c = \frac{\sum_{t=0}^{t=T_{max}-\tau} Y(t+\tau)V(t)}{\sum_{t=0}^{T_{max}-\tau} |V(t)|^2}$$

If the above expression does not belong to $[0,1]$, the adjust the value to the closest end point:

$$\gamma_{min} = \begin{cases} 0 & if \quad \gamma_c < 0 \\ \gamma & if \quad \gamma_c \in [0,1] \\ 1 & if \quad \gamma_c > 1 \end{cases}$$

The case $p = 1$

Solution of optimization problem $\min_{\gamma} \sum_{t=0}^{T_{max}-\tau} |Y(t+\tau) - \gamma V(t)|$:

Algorithm (The l^1 estimator for γ)

1. *For each* $k = 0, 1, \cdots, T_{max} - \tau$ *repeat:*
    1. *Compute* $r(k) = \frac{Y(k+\tau)}{V(k)}$.
    2. *If* $r(k) \notin [0, 1]$ *then discard this value and proceed to the next* k.
    3. *Compute:* $f(k) = \sum_{t=0}^{T_{max}-\tau} |Y(t+\tau) - r(k)V(t)|$
2. *Find the minimum and the index* $[minf, indexMin] = min(f)$.
3. *Assign:* $\gamma_{min} = r(indexMin)$.

Independent problem: Try writing it as a linear program!

The case $p = \infty$

Solution of $\text{minimize}_{\gamma} \ \max_{0 \leq t \leq T_{max}-\tau} |Y(t+\tau) - \gamma V(t)|$ is given by the following linear program:

$$\begin{array}{cc} \textit{minimize} & s \\ -s \leq Y(t+\tau) - \gamma V(t) \leq s \ , \ 0 \leq t \leq T_{max} - \tau & \end{array}$$

It can be rewritten into a standard form with vector $x = [s; \gamma]$, matrix A, vectors b, $f = [1; 0]$, lower bound $\mathbf{0} = [0; 0]$ and upper bound $u_\infty = [\infty; 1]$:

$$\begin{array}{c} \textit{minimize} \quad f^T x \\ Ax \leq b \\ \mathbf{0} \leq x \leq u_\infty \end{array}$$

where:

$$A = \begin{bmatrix} -1 & -V(0) \\ -1 & V(0) \\ \vdots & \vdots \\ -1 & -V(T_{max}-\tau) \\ -1 & V(T_{max}-\tau) \end{bmatrix} \quad , \quad b = \begin{bmatrix} -Y(\tau) \\ Y(\tau) \\ \vdots \\ -Y(T_{max}) \\ Y(T_{max}) \end{bmatrix} .$$

Note: A is a matrix of size $2(T_{max} - \tau + 1)x2$ and b is vector of length $2(T_{max} - \tau + 1)$.

SIR Model with Vitals

A simple modification of the SIR vanilla model is to consider vital signals, such as births and deaths at separate processes. In normalized form this becomes:

$$\left\{ \begin{array}{lcl} \frac{ds}{dt} & = & \frac{\Lambda}{N} - \beta s i - \mu s \quad , \quad s(0) = \frac{S_0}{N} \\ \frac{di}{dt} & = & \beta s i - \alpha i - \mu i \quad , \quad i(0) = \frac{I_0}{N} \\ \frac{dr}{dt} & = & \alpha i - \mu r \quad , \quad r(0) = \frac{R_0}{N} \end{array} \right. \qquad \text{(SIR Model)}$$

where $\Lambda \geq 0$ is the constant source of births (=number of births/day) and $\mu \geq 0$ is the natural death rate (i.e., in the absence of this virus). Its reciprocal $1/\mu$ represents the average life expectancy.