Technical Documentation

COVID-19 Bayesian Three-stage ODE Model for U.S. Counties Forecasting

Gary Lin[1], Yupeng Yang[2], Eili Klein[1,2]

[1] Department of Emergency Medicine, Johns Hopkins University, Baltimore, Maryland
[2] Center for Disease Dynamics, Economics & Policy, Silver Springs, Maryland

Table of Contents

Model Specifications 2
Bayesian Inference of Model Parameters 3
Deriving the Next Generation Matrix and Calculating R_0 3

Model Specifications

Using the classic SIR model formulation from Kermack and McKendrik, we utilized the classical epidemiological dynamics to reflect viral transmission patterns of the COVID-19. We defined five compartmental states: susceptible (*S*), exposed (*E*), asymptomatic and mild symptomatic infections (*C*), moderate to severe symptomatic infections (I_N), hospitalized (I_H), and recovered (*R*), which are described in the main text. The model can be formulated as a set of ordinary differential equations (ODE).

$$
\begin{aligned}
\dot{S} &= -\alpha\beta(t)\frac{SC}{N} - \beta(t)\frac{SI}{N} \\
\dot{E} &= \alpha\beta(t)\frac{SC}{N} + \beta(t)\frac{SI}{N} - \mu E \\
\dot{C} &= \mu E(1-\theta) - \gamma_1 C \\
\dot{I_N} &= \mu\theta E(1-\varphi) - \gamma_2 I_N \\
\dot{I_H} &= \varphi\mu\theta E - \gamma_3 I_H \\
\dot{R} &= \gamma_1 C + \gamma_2 I_N + \gamma_3 I_H
\end{aligned}
\qquad (1)
$$

Transmission rate due to symptomatic population, β, is assumed to be reduced by α for asymptomatic transmissions. The detected symptomatic rate, θ, determines the fraction of exposed individuals that transition to the infected state with moderate and severe symptoms. The inverse of the incubation period is defined by μ, and the inverse of the infectiousness periods for infected individuals with no mild or no symptoms *C*, moderate to severe symptoms I_N, hospitalizations I_H group with mild or no symptoms are indicated by γ_1, γ_2, and γ_3, respectively. Finally, the parameter, φ, is the hospitalization rate.

In our model, we assumed that interventions, such as lockdown and post-lockdown social distancing policies will impact the transmission of SARS-CoV-2 depending on the stage of the lockdown. There are three stages in our model: (1) Pre-lockdown, (2) lockdown, and (3) post-lockdown with social distancing. For the lockdown intervention, we assumed that the interventions start and end, respectively, at t_{start}, and t_{end}. Hence,

$$
\beta(t) = \begin{cases} \beta_{pre}, & t < t_{start} \\ \beta_{lock}, & t_{start} > t \geq t_{end} \\ \beta_{post}, & t > t_{end} \end{cases}
$$

The initial susceptible population is assumed to be the population size of the country based on population estimates from the U.S. Census Bureau[1]. The cumulative detected infections, defined as the sum of I_N and I_H, was fitted to data on confirmed cases reported by the Center for System Science and Engineering at Johns Hopkins University[2]. We initialize the model by assuming the first-time step of the simulation is the same day as the first incidence of confirmed cases for each county. The initial conditions assumed that seeded infections coincided with the number of observed cases and an additional five exposed individuals for each seeded infection. The set of ordinary differential equations (ODEs) were solved using the LSODA algorithm [3], and implemented in Python 3.8.

Bayesian Inference of Model Parameters

In order to understand the credibility of parameters, Θ, we used Monte Carlo Markov Chain (MCMC) methods to estimate the $P(\Theta \mid \hat{X})$, the posterior probability of the parameters, in which we are estimating based on the observed confirmed cases, $\hat{X}$. From the observed data, we were able to estimate the posterior distribution using the following Bayesian framework,

$$P(\Theta|\hat{X}) = \frac{P(\hat{X}|\Theta)P(\Theta)}{P(\hat{X})} \propto \mathcal{L}(\Theta)P(\Theta)$$

where $\mathcal{L}(\Theta) = P(\hat{X}|\Theta)$ is the likelihood function and P(Θ) is prior on our belief of Θ. The prior assumptions are assumed to uniform with an upper and lower bound. Since we are assuming that the error follows a normal distribution $\mathcal{N}(0, \sigma^2)$ with known variance σ^2, the loglikelihood can be defined as

$$\log \mathcal{L}(\Theta) = -\log(2\pi\sigma^2) - \frac{SSR(\Theta)}{2\sigma^2}$$

where SSR is the sum of squared residuals between the simulated data X_{it} and observed data $\hat{X}_{it}$. For Italy, Spain, South Korea, and Chicago, we fitted the model to confirmed cases. For New York City, we fitted the model to confirmed cases and cumulative hospitalizations, which means there are two times series that the MCMC is using to determine the goodness of fit. The goodness of fit is based on the sum of squared residuals (SSR) which is commonly used in time series analysis and can be calculated as

$$SSR = \sum_{i \in M} \sum_{t=0}^{T} \left(\log(X_{it}(\Theta) + 1) - \log(\hat{X}_{it} + 1)\right)^2$$

where the subscripts $i \in M$ represents a fitted series (e.g. cumulative infections), and *t* represents a time point. The inverse sum of squared residuals, SSR^{-1}, are shown in Tables S1 and S2. We used the log value of the simulated and real data values to better capture the exponential increase. Using the Metropolis-Hasting (M-H) Algorithm, the posterior marginal distributions for each parameter in Θ were generated assuming uniform priors. Normally distributed proposal densities were considered to generate candidate samples for the acceptance-rejection step of the M-H algorithm. In order to obtain independent samples of the posterior distribution, a burn-in period and thinning was conducted with appropriate intervals for each case. The simulations were performed on Amazon Web Services (AWS).

Deriving the Next Generation Matrix and Calculating R_0

To assess the biological relevance of the parameters, we calculated the R_0 of the model, based on the dominant eigenvalue of the *next generation matrix*. The infection subsystem of the model can be described by equations (2).

$$\begin{aligned} \dot{E} &= \alpha\beta\frac{SC}{N} + \beta\frac{SI}{N} - \mu E \\ \dot{C} &= \mu E(1-\theta) - \gamma_1 C \\ \dot{I} &= \theta\mu E - \gamma_2 I \end{aligned} \quad (2)$$

We use a linearization of the infection subsystem, similar to the formulation outlined by Diekmann et al. [4],

(4) $$\dot{\mathbf{x}} = (\mathbf{T} + \mathbf{M})\mathbf{x}$$

where **x** = [E, C, I], T represents the transmission matrix that captures the number of transmissions from the susceptible compartments, and the matrix M represents the transition between compartments within the infection subsystem. Specifically, T is equal to

$$\mathbf{T} = \begin{bmatrix} 0 & \alpha\beta & \beta \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and M is equal to

$$\mathbf{M} = \begin{bmatrix} -\mu & 0 & 0 \\ (1-\theta)\mu & -\gamma_1 & 0 \\ \theta\mu & 0 & -\gamma_2 \end{bmatrix}.$$

From this formulation, we can mathematically construct the next generation matrix W**,** which is defined as

$$\mathbf{W} = -\mathbf{T}\mathbf{M}^{-1},$$

which would equate to

$$\mathbf{W} = -\begin{bmatrix} 0 & \alpha\beta & \beta \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} -\frac{1}{\mu} & 0 & 0 \\ -\frac{(1-\theta)}{\gamma_1} & -\frac{1}{\gamma_1} & 0 \\ -\frac{\theta}{\gamma_2} & 0 & -\frac{1}{\gamma_2} \end{bmatrix}$$

$$= \begin{bmatrix} \alpha\beta\frac{(1-\theta)}{\gamma_1} + \beta\frac{\theta}{\gamma_2} & \alpha\beta\frac{1}{\gamma_1} & \beta\frac{1}{\gamma_2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

R_0 is equal to the dominant eigenvalue of **W** (highlighted in red), which is equal to

$$R_0 = \alpha\beta\frac{(1-\theta)}{\gamma_1} + \beta\frac{\theta}{\gamma_2}.$$

References

1 US Census Bureau. Data. Census.gov. https://www.census.gov/data (accessed July 9, 2020).

2 Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases* 2020; **0**. DOI:10.1016/S1473-3099(20)30120-1.

3 Petzold L. Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations. *SIAM J Sci and Stat Comput* 1983; **4**: 136–48.

4 Diekmann O, Heesterbeek JAP, Roberts MG. The construction of next-generation matrices for compartmental epidemic models. *J R Soc Interface* 2010; **7**: 873–85.