

Final Report

Group 2: Yujin Liu, Helen Jin, Jiangmeng Li

CAAM37830

University of Chicago

December 9, 2020

1 Abstract

In this report, we investigated how the infectious disease spread throughout the population based on the susceptible-infected-recovered (SIR) epidemic model. We first ran the simulation with both the discrete agent-based and the continuous ODE models. We then proposed and implemented three variations of the model. In the first modification, we aim to understand how the mask use will change the way infectious disease spread, and how the adoption of the masks with different effectiveness influences number of hospitalizations and deaths. Secondly, we used some default parameters to fit the Covid data from Hubei, China to our continuous ODE model with a close attention to the removed and infected rate. Lastly, we also created a visualization of the interactive heat map, interactive time series plot of the confirmed and death for different states. Apart from these variations of the model, we also added the a 2-dimensional spatial component to both the discrete agent-based and the continuous ODE models where population is confined in a box and observe how different initial location conditions change the spread of the disease.

2 Introduction

The susceptible-infected-removed (SIR) model is one of the compartmental models that simplify the mathematical modeling to the spread of infectious disease, where the time dependent variables S , I , R each represent the following populations:

S (susceptible): number of individuals who are not infected but could become infected

I (infected): number of individuals who are already infected and can spread disease

R (removed): number of individuals who are either recovered and immune or have died

Additionally, s , i , r are used to represent the proportion of susceptible, infected and removed individuals among the population. Note that this model assumes the susceptible population S decreases monotonically towards 0, and the population size is fixed, and the duration of infectivity is same as length of the disease.

There are two parameters b and k in the model, where b indicates the number of interactions each day that could spread the disease (per individual) and k indicates the fraction of the infectious population which recovers each day. The model consists of the following system of the nonlinear ordinary differential equations, where t is time:

$$\frac{ds}{dt} = -b * s(t) * i(t)$$

$$\frac{dr}{dt} = k * i(t)$$

$$\frac{di}{dt} = b * s(t) * i(t) - k * i(t)$$

3 Spatial Model and Variations

3.1 spatial agent based model

Firstly, we briefly introduce the agent based model in midterm checkpoint. We assumed that each person contact b people every day. For each infected person, the recover rate is k . We simulated the process of interaction for population and found the region in which all people would be infected. We also attained the phase transition boundary $b = 10 * k - 3$.

Now we would like to introduce the spatial agent based model. This is how we implement the model. We introduce another `Person()` class with the `pos` attribute. For each individual, it will move

and interact with other individuals according to `movpos` function. We also define the `resetcorner` and `resetcenter` functions to set the position of an individual so that we can implement the simulation function in different occasions.

This is the simulated model as we start all the individuals randomly. we set $k = 0.6$, $q = 0.08$, $p = 0.5$. We can see all the people will finally be infected which is reasonable since each individual will move with a large step size and radius $q(0.08)$ is large enough compared with a population size of 20000. We can observe that all susceptible people eventually get infected very quickly.

In this part, we would like to explore the relationship between the total infected individuals and p as infected people initially located at the center of the confined region. We choose the parameter $b = 0.5$, $k = 2$ from the phase transition boundary. We can see that as p increases, the percentage of total infection increases until a peak value of around 0.8 is reached at $p = 0.1$. Then, it remains stable as p changes from 0.1 to 0.2. As p increases from 0.2, the percentage of total infection decreases, especially when $p > 0.7$.

The result of the plot matches our expectation. As p is small ($p = 0.01$), infected individual barely moves, only a small portion of susceptible people will get infected. People who are infected initially will be recovered and the total percentage of infection is small.

As $0.1 < p < 0.4$, all the infected people will move around and contact most of susceptible people. The total percentage of infected people will be large. However, we choose the b, k from transition boundary and we can observe not all the people will finally be infected.

As p is large, some of the infected people may stay at the original position because the step size is large. Thus, the total percentage of infected people will decrease. As p is extremely large, all the infected people will stay at the center. They can only infect the susceptible people around them initially. Only small percentage of population will finally be infected.

Now, we would like to discuss what will happen if all the people move from the corner or randomly spread. We choose the parameter $p = 0.05$. We can see that the percentage of total infection will attain the maximum if all the infected people start from the random position in the grid. It will reach minimum if all the infected people start from a single corner of the grid. This is reasonable. As all the infected individuals move from a random position, they will frequently interact with different people among the population and it will be stable with most of people quickly get infected. However, if all the infected people start from a single corner, they will not move too far away from that corner if the step is small. Thus, they will only contact with a small amount of susceptible people and all the infected people will quickly be recovered. Therefore, we can conclude that total percentage of the infected people will be small if they start from a single corner and the step size is small.

Now we consider the case that all people start at the center, with a small step size $p = 0.05$. Since infected people will have more contact with others compared to the situation where they start at the corner, the total percentage of infected people will be larger. However, the step size is small and all the infected people will also quickly be recovered and the total percentage of infected individuals will be smaller than the case that all the infected people are random spreaded out in the grid.

3.2 Spatial PDE Model

Firstly, we briefly introduce the ode model in the midterm checkpoint. The implementation is quite straightforward. The inputs for the derivative are three variables s, i, r and a function for obtaining derivative. Then we ran the ode simulation with the `solve_ivp` function. The result is reasonable and we almost have the same transition phase boundary as what we explored in agent based model. In the following part, we use the parameter $b = 1$ and $k = 0.4$. Not all the people will be susceptible in this case.

Now we would like to explain how we implemented the spatial pde model. We have several functions for generating the desired distribution of population. `generatesum2` is the function for

generating the population with initial infected people randomly spread out in the grid. `generatesum2center` is the function for setting the infected individuals at center initially while `generatesum2corner` is the one setting the infected individuals at corner. We combine the `s,i,r` vector together as input for derivative function, and then calculate the derivative according to the relationship:

$$\partial s(x, t) / \partial t = -b * s(x, t) * i(x, t) + p * Ls(x, t)$$

As for $i(x, t)$ and $r(x, t)$, similar relationship can be established. Since we have the function for obtaining derivative, we can simulate the pde process. The following is the result.

The plot is reasonable. We choose the parameter $b=1$ and $k = 0.4$. We can observe the case that not all the individuals will finally be infected. We choose an appropriate diffusion term $p = 1$ so that all the susceptible in the grids can be potentially infected by the initially infected group located at center.

Now, we analyze how total percentage of infected people i changes with p . We choose the p in list $[1, 2, 3, 4, 5, 10, 20, 30, 40, 50]$. We can see that as p is increasing, the total infected people will increase. The increment is not large.

We can conclude that the percentage of total infection will be large if diffusion term is large since all the people will interact with individuals located in different positions frequently. The population will reach a stable state quickly.

We also find that the increment is not large as p is increasing. This is because we have the diffusion term $p \downarrow = 1$ which guarantees the condition that the infected people can be spread out in the grid before most of the infected people in population are recovered. Thus, as p is increasing, we can observe the increment while it is not large.

In this part, we try to understand what will happen if we place infected people at different locations initially. As $p = 20$, the total infected percentage of population are 0.8926, 0.88, 0.86 as initial infected people are randomly spread out, located from center, located in a corner respectively. The result is reasonable. The total infected people will reach its maximum if people infected initially are distributed randomly in the grid because there is a chance that infected people may contact all the susceptible people in different positions before all the they recover. However, we control the relationship between b and k to ensure that not all the people will finally be infected. As for the case that all the infected people were initially located at center, we can see that the total percentage of infected people will be close to the first case since the diffusion term is large and most of infected people will have a chance to contact susceptible people before they recover. As for the case that all infected people were located at a corner initially, the total percentage of infection will be smaller because the infected people will contact less people due to the starting position before they are recovered.

4 Covid-19 fit

4.1 Introduction and Implementation

In this variation, we will fit SIR model with the Covid-19 dataset in China Hubei Province. Instead of using the SARS dataset claimed in midterm report, we find that Covid-19 dataset in China is complete.

Now, we would like to discuss how we fit the covid-19 data. We collect the dataset from ... in first 76 days. The covid-19 was controlled in China in around 76 days. Building the specialized hospital helped the government to collect severe patients so that the transmission of covid-19 was controlled. We divide the process of simulation in two phases. Firstly, people were not aware that covid-19 is a kind of coronavirus which can severely influence life. After around 25 days of pandemic period, the government started to control the covid19 through building large hospitals for collecting severe

patients and announce the policy that all the people should stay at home, especially in province Hubei. From then on, the total infected people was stable after around 50 days and we try to simulate the whole process through the basic SIR model.

In the first phase, we assume that we use the default parameters from the source ... $b = 0.35$ and $r = 0.03$. We explore a range of b, k parameters so that in first 25 days, the MSE of the i, r people in Hubei and the simulated i, r values will attain the minimum. We use the function `calcuatemse` to calculate the MSE between the simulated value and practical dataset. We have the conclusion that in first phase, $b = 0.23$, $k = 0.025$.

As for the second phase, we apply the same function to find the most appropriate b, k according to calculating mse between simulated data and real dataset. However, the policy is adjusted here and most of infected people are moved to hospitals, the transmission rate b will be much smaller than original value. We try to find b from the range $(0.01, 0.04)$ and it works. We find $b_2 = 0.02$, $k_2 = 0.06$. The result is reasonable from the article (source). The recovered rate should be around 0.05 per infected person each day. As people recognized the fact that covid-19 is a severe covoriance, most people were wearing masks and stayed at home. Even if they are affected with covid19, the chance for being recovered will be improved.

4.2 Result and Discussion

Following is the result of the simulation of covid19 in Hubei Province.

From the graph, we can see that the observed curve is close to the real dataset with small deviation. There are several reasons that the simulated curve is not very close to the real dataset. First of all, the SIR model is a little bit simplified. To be specific, we don't consider the case that some people may be asymptomatic among the population. The recover rate and hospitalized rate is different for asymptomatic group and symptomatic group. A sir model doesn't take account into these variations.

Apart from that, the limitation comes from the incomplete of statistics. To be specific, we know that for a large province in China, it is hard to ensure the dataset is matched with the real case. Some infected people may not be tested as positive while the tested sample is only part of population.

Finally, there is also a possibility that the factors influencing the covid19 is hard to be identified. For instance, we regard the transimission rate and removed rate as constant in two phases. However, the transimission rate or removed rate may change frequently in several weeks due to the change of climate. We can only have a much accurate result if we can have enough research of covid-19.

To be summarized, we can conclude three factors that may influence our simulation, the simplified sir model, the incomplete statistics and complex factors influencing transmission of covid-19. We can have an accurate simulation if we have a much more complex model taking into account different groups among the population, having a method for collecting the complete sample for removed or confirmed individuals, or recognizing what influenced the spread of covid-19 after some research.

5 Modelling the usage of mask

5.1 Modified SIR model with mask

In this section, we study the effectiveness of mask in the spreading of the novel COVID-19 disease and develop a variation of the susceptible-infected-removed (SIR) model that provides a theoretical framework to investigate its spread within a community. More specifically, we will investigate how the usage of mask will influence the percentage of deaths and hospitalizations with respect to different infectious contact rate. The model is based on the paper "To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic" by Steffen E. Eikenberry et al. In this model, we will introduce seven variables. $S(t)$, $E(t)$, $I(t)$, $H(t)$, $A(t)$, $R(t)$

each denote susceptible, exposed, symptomatic infectious, hospitalized, asymptomatic infectious, and recovered classes, with an assumption that people progress from

$$S \rightarrow E \rightarrow A \rightarrow I \rightarrow H.$$

At each state, infectious people will be able to recover, and only people in severe condition will go to hospital and may die. $D(t)$ is also included to track cumulative deaths. Two sets of ODE equations will be used during simulation to find how use of mask will impact the spread of disease. We first consider a baseline model for the case where no masks are used. In the second set of ODEs, we will introduce more variables $S_U(t)$, $E_U(t)$, $I_U(t)$, $H_U(t)$, $A_U(t)$, $R_U(t)$, $S_M(t)$, $E_M(t)$, $I_M(t)$, $H_M(t)$, $A_M(t)$, $R_M(t)$ to divide the population in two part: the population wearing masks and the one not wearing masks. Here are the two sets of ODEs:

$$\begin{aligned}
 \frac{dS}{dt} &= -\beta(t)(I + \eta A) \frac{S}{N} \\
 \frac{dE}{dt} &= \beta(t)(I + \eta A) \frac{S}{N} - \sigma E \\
 \frac{dI}{dt} &= \alpha \sigma E - \phi I - \gamma_I I \\
 \frac{dA}{dt} &= (1 - \alpha) \sigma E - \gamma_A A \\
 \frac{dH}{dt} &= \phi I - \sigma H - \gamma_H H \\
 \frac{dR}{dt} &= \gamma_I I + \gamma_A A + \gamma_H H \\
 \frac{dD}{dt} &= \sigma H
 \end{aligned}
 \quad
 \begin{aligned}
 \frac{dS_U}{dt} &= -\beta(I_U + \eta A_U) \frac{S_U}{N} - \beta((1 - \epsilon_0)I_M + (1 - \epsilon_0)\eta A_M) \frac{S_U}{N} \\
 \frac{dE_U}{dt} &= \beta(I_U + \eta A_U) \frac{S_U}{N} + \beta((1 - \epsilon_0)I_M + (1 - \epsilon_0)\eta A_M) \frac{S_U}{N} - \sigma E_U \\
 \frac{dI_U}{dt} &= \alpha \sigma E_U - \phi I_U - \gamma_I I_U \\
 \frac{dA_U}{dt} &= (1 - \alpha) \sigma E_U - \gamma_A A_U \\
 \frac{dH_U}{dt} &= \phi I_U - \delta H_U - \gamma_H H_U \\
 \frac{dR_U}{dt} &= \gamma_I I_U + \gamma_A A_U + \gamma_H H_U \\
 \frac{dD_U}{dt} &= \delta H_U \\
 \frac{dS_M}{dt} &= -\beta(1 - \epsilon_i)(I_U + \eta A_U) \frac{S_M}{N} - \beta(1 - \epsilon_i)((1 - \epsilon_0)I_M + (1 - \epsilon_0)\eta A_M) \frac{S_M}{N} \\
 \frac{dE_M}{dt} &= \beta(1 - \epsilon_i)(I_U + \eta A_U) \frac{S_M}{N} + \beta(1 - \epsilon_i)((1 - \epsilon_0)I_M + (1 - \epsilon_0)\eta A_M) \frac{S_M}{N} - \sigma E_M \\
 \frac{dI_M}{dt} &= \alpha \sigma E_M - \phi I_M - \gamma_I I_M \\
 \frac{dA_M}{dt} &= (1 - \alpha) \sigma E_M - \gamma_A A_M \\
 \frac{dH_M}{dt} &= \phi I_M - \delta H_M - \gamma_H H_M \\
 (1) \quad \frac{dR_M}{dt} &= \gamma_I I_M + \gamma_A A_M + \gamma_H H_M \\
 \frac{dD_M}{dt} &= \delta H_M
 \end{aligned}
 \tag{2}$$

5.2 Implementation

We ran the simulation and recorded what happened in several days. Here, β is the infectious contact rate, σ is the transition exposed to infectious, η is the infectiousness factor for asymptomatic carrier, α is the fraction of infections that become symptomatic, ϕ is rate of hospitalization, γ_A is the recovery rate for asymptomatic, γ_I is the recovery rate of symptomatic, γ_H is the recovery rate for hospitalized, σ is the death rate in hospital, ϵ_0 is the outward efficiency of the masks, while ϵ_i is the inward efficiency of the masks. Then we found the default estimated values for all parameters with population masked and not masked from the source [2]. Here, an assumption was made that the inward efficiency and outward efficiency of masks are the same $\epsilon_0 = \epsilon_1$. We drew a 2 dimensional phase diagram to show how the efficiency of mask (homemade mask or surgery mask) and the percentage of masked people will influence the hospitalization as well as the total deaths among the population.