Dynamic Prioritization for Vaccination during Covid-19 pandemic in Michigan, U.S.

– IOE 512 Course Project –

submitted by

Ruixuan Zhang, Isaac Smith, Shreyas Bhat

December 18, 2020

Ruixuan Zhang, Isaac Smith, Shreyas Bhat

ruixuanz@umich.edu, isjsmith@umich.edu, shreyasb@umich.edu

Abstract

The topic of this project is about prioritizing vaccination during the Covid-19 pandemic in Michigan U.S. using dynamic programming. Besides prioritizing vaccination, we also consider the vaccine supply chain issue and combine these two problems together, which makes it a spatial-temporal multi-objective optimization problem with stochastic patterns. In a finite time horizon T, we evaluate the priority among G groups of population (based on age) at each time stage. We divide Michigan into two regions based on the risk level. The GSIR (General Susceptible-Infected-Removal) model is utilized to describe the dynamics of the transmission and the contact matrix is introduced to define the latent risk to infect others. The vaccine supply chain problem is also modelled as stochastic along with time-variant vaccine production ability at each supply node and transportation cost at each edge within the network. In the second part, we implemented the numerical simulation based on Monte Carlo method with up-to-date Covid-19 data in Michigan. The sensitivity analysis, conclusions, and future work are discussed in the end.

Acknowledgments

We would like to thank Siqian Shen and Kati Moug for their continued support through this difficult and evolving semester.

Contents

List of Figures iv

List of Tables v

List of Acronyms vi

1 Introduction 1

1.1 Related Work . . . 1

2 Preliminaries 2

2.1 The Markov decision process . . . 2

2.2 The transmission model: Generalized SIR . . . 2

2.3 Vaccine supply chain . . . 4

3 Details in Model Formulation 4

3.1 State Space . . . 4

3.2 Action Space . . . 5

3.3 State Transition . . . 5

3.4 Cost Functions . . . 7

4 Model Application in Michigan study 8

4.1 Data Preparation . . . 8

4.2 Infection Risk: Contact Matrix in Covid-19 pandemic . . . 9

4.3 Design of Action Space . . . 9

4.4 Numerical Simulation . . . 10

4.5 Algorithms . . . 11

5 Results and Conclusions 13

5.1 Simulation Results . . . 14

5.2 Sensitivity Analysis . . . 14

5.3 Discussions . . . 16

5.4 Workload Summary . . . 17

References 18

List of Figures

1 Transitions in the SIR model 3
2 The eight regions in Michigan d 8
3 Comparison of simulated optimal policy and baseline policy 14
4 6 (left) vs 12 (right) Day Detection Period 15
5 18000 vs 30000 Daily Vaccines 16

List of Tables

1 Daily Contact Matrix for Age Groups 0-19, 20-59, and 60+ 9
2 Model parameters 11
3 Data by age group 11
4 Variation of Parameters 15

List of Acronyms

GSIR Generalized Susceptible, Infected, and Removed

1 Introduction

The novel coronavirus (COVID-19) has been spreading rapidly, posing a great threat to global health and economy. Although vaccines are being developed, the initial production of vaccines would be very limited compared to the number of people in need of these vaccines. The population of a geographical region can be divided into groups based on age. These groups have different risks of getting infected and infecting others. Due to these factors, it becomes important to come up with a prioritization plan to distribute the vaccines among these population groups.

Secondly,these vaccines are produced and distributed from different areas. So, the transportation cost associated with the supply of vaccines to cities within that region also needs to be considered. These costs may be variable depending on conditions such as weather.

Finding the optimal policy of vaccine distribution becomes a challenging problem given these two competing objective functions. In this study, we propose a model that finds the best policy to optimize a weighted sum of these objectives. We model the spread of the disease using a generalized SIR model of infectious diseases and demonstrate our model using data from the state of Michigan [2].

1.1 Related Work

Wan et al. came up with the Generalized SIR model for modeling the spread of the novel coronavirus [8]. Their model determined the optimal policy for implementing various levels of shutdown in the economy considering the economic and infection costs. Matrajt et al. used simulation based search to find the optimal vaccine allocation strategy [5]. However, their model assumed a deterministic spread of the virus. Surveys by Roijers et al. and Liu et al. focus on problems with multiple competing objectives [7, 3]. Wu et al. consider the ethical aspects of a vaccine prioritization plan [9]. They also show how prioritizing the worse-off groups of populations can be ethically feasible and beneficial. Yaesoubi and Cohen also propose a generalized Markov chain model for modeling the spread of an infectious disease [10]. However, their model suffers from the curse of dimensionality when the population size increases beyond a certain value.

2 Preliminaries

2.1 The Markov decision process

For a Markov decision process (MDP), a tuple $M = (S, A, P, \gamma, R)$ is defined to claim a sequential decision making model, where: S is a finite set of **states**; $A = \{a_1, a_2, \cdots, a_k\}$ is a set of $k \geq 2$ **actions**; $P = \{P_{sa}(\cdot)|s \in S, a \in A\}$ are the next-state **transition probabilities**, with $P_{sa}(s')$ giving the probability of transitioning to state s' upon taking action a in state s; $\gamma \in (0, 1]$ is the **discount factor**; and R specifies the reward distributions. In this problem, for simplicity, we will assume rewards are deterministic and bounded, and since it is a multi-objective MDP problem, here $R = \{R_1, \cdots, R_N\}^T$ is a vector of N **reward functions** for N different objectives. The finite horizon is predefined as T, which can be set large enough to represent the infinite horizon case.

Also, in order to make it more realistic, we apply deterministic policies at each time period t. Given a deterministic policy π, for each objective n, there is one value function $V^{\pi}_{n,t_0}(s)$ corresponding to state s and time period t_0. Then $V^{\pi}_{n,t_0}(s) = E_{\pi}[\sum_{t=t_0}^{T} R_n(s_t, a_t)|s_{t_0} = s]$. For the MDP setting, the objective is typically to find an optimal policy π^* which maximizes the reward or minimizes the cost. Two remarks here: we will define the objective as minimizing the total cost; for multi-objective problems, we will consider a unified weight vector w for the linear combination of N objective value functions, $\{\exists w \in R^N_+ \text{ s.t. } \mathbf{w}^T\mathbf{V}^{\pi}_{t_0}(s) \leq \mathbf{w}^T\mathbf{V}^{\pi'}_{t_0}(s)\}$, where π and π' are two feasible policies, $\mathbf{w}^T\mathbf{1} = 1$ and $\mathbf{V}^{\pi}_{t_0}(s) = (V^{\pi}_{1,t_0}, \cdots, V^{\pi}_{N,t_0})^T$. In fact, the "optimal" policy π^* may not be unique since it will depend on how much weight we would like to select on each individual objective. Therefore, the Pareto-optimal policy could be introduced, that is, a policy can not improve further without sacrificing any of others. But again, the final decision still depend on the rule-maker's subjective consideration, which we will discuss later.

2.2 The transmission model: Generalized SIR

When it comes to model the transmission of a disease, the SIR model has been proved to be one of the most widely applied model to describe the dynamics of infection disease [8].

The SIR model is a type of compartmental model for modeling the spread of infectious diseases. These models assign different compartments to the population depending on their health state. In the SIR model, people are categorized into three compartments: The **S**usceptible group of people who have not yet been infected but have a possibility of

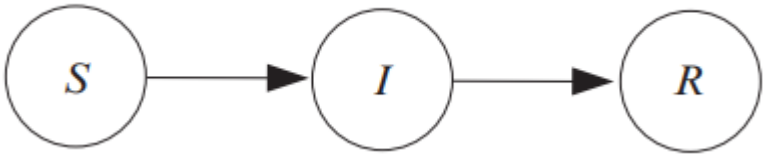

Figure 1: Transitions in the SIR model

getting infected, the **I**nfected group of people are the ones that are currently carrying the disease and can transmit it to the susceptible people, and the **R**emoved group which is the group of people that have either recovered, vaccinated or died and hence, cannot be infected again. The tuple (S, I, R) represents the number of people in each of these three groups at a given time, and thus constitutes the state of disease spread. Each member of the population typically progresses from susceptible to infectious to removed. Two hyper parameters, β and γ, control the rate of spread of the disease. Here, β captures the rate of spread of infection and γ captures the rate at which people transition from the infected group to the removed group.

While SIR model well fits in many cases, there are still at least two deficiencies for simply applying naive SIR model to this problem. Firstly, we can see that the infection rate β and removal rate γ are all constants, which cannot reflect the influence of the actions that we take in different regions. Secondly, the SIR model is deterministic without stochastic patterns, which is not realistic and over simplified. Having considered the above aspects, prompted by relevant researches, the dynamic of getting infected can be modelled as a Poisson process because we can view the future new cases as the number of arrival events during time period t; the dynamic of removal (recovering or death) can be modelled as a Binomial process because during each period t, the infected people are whether still infected or getting recovered/death. Therefore, the generalized SIR model can be shown as below:

$$
\begin{aligned}
X^S_{l,t+1} &= X^S_{l,t} - e^S_{l,t}, \quad e^S_{l,t} \sim \textbf{Poisson}(\beta_{l,t}(A_{l,t} = a_{l,t}) X^S_{l,t} \frac{X^I_{l,t}}{M_l}) \\
X^R_{l,t+1} &= X^R_{l,t} + e^R_{l,t}, \quad e^R_{l,t} \sim \textbf{Binomial}(X^I_{l,t},\ \gamma_t) \\
X^I_{l,t+1} &= M_l - X^S_{l,t} - X^R_{l,t}
\end{aligned}
$$

where $A_{l,t}$ is the action taken in region l at time period t. Note that the actions taken in every regions are mutually dependent at the same time t, and $\beta_{l,t}(A_{l,t} = a_{l,t})$ denotes the infection rate under action $a_{l,t}$, and γ is assumed the same in every region (suppose the medical treatment ability is state-wide equal). Then the parameter vector $\boldsymbol{\theta_t} = (\gamma_t, \beta_{1,t}, \cdots, \beta_{L,t})^T$ is what we need to formulate the generalized SIR model.

2.3 Vaccine supply chain

Mclean et al. wrote a paper studying the global production of seasonal and pandemic influenza [6]. Since there weren't any concrete studies done on the capacity of Covid vaccines, these may be a reasonable approximation. In their paper, they state that in one year that companies could create enough vaccines to vaccinate 86% of the current population. It is possible that this might be an overestimate for the Covid vaccines, but this will be our baseline. To adapt it from the world to Michigan we will take the current population of Michigan multiplied by 86% and the yearly ratio of our time period.

3 Details in Model Formulation

In the model estimation step, Covid-19 data from several adjacent counties are aggregated together as region-based to share information and operate as the basic unit of the whole state. In the decision making step, all eight regions of Michigan take simultaneous actions according to the state-wide distribution plan.

3.1 State Space

For each region l, at each decision time period t, according to the estimated transition model, the current state $\mathbf{S}_{l,t} = (X^S_{l,t}, X^I_{l,t}, X^R_{l,t})$ and the next state $\mathbf{S}_{l,t+1} = (X^S_{l,t+1}, X^I_{l,t+1}, X^R_{l,t+1})$ are related by the conditional density for $S_{l,t+1}$ given $S_{l,t}$ and $A_{l,t}$ in the G-SIR model.

The available data to help make decision is the cumulative number of confirmed cases up to time t in region l, denoted as observation $O^I_{l,t}$. Intuitively, $O^I_{l,t}$ can be considered as $X^R_{l,t}$ if we assume the infected people are quarantined until recovery/death and they won't get infected again. As for $X^I_{l,t}$, it is not trivial if time t is not the initial t_0. On the one hand, we can assume that the initial $X^I_{l,t_0} = O^I_{l,t_0} - O^I_{l,t_0-D}$, where D is a fixed number indicating the delay time of from being infected to being quarantined. When $t > t_0$, we will estimate the $X^I_{l,t}$ in the following way: suppose among X^I_{l,t_0}, the proportions of each population group are $\alpha_1, \alpha_2, \cdots, \alpha_K$ s.t. $\sum_{k=1}^K \alpha_k = 1$, and we will utilize the **contact matrix** to describe the potential risk of infecting other people by a specific group of infected population, and then estimate the future $X^I_{l,t+1}$. More details will be explained in the next section. Finally, $X^S_{l,t} = M_l - X^I_{l,t} - X^R_{l,t}$, assuming the total population in region l remains the same regarding the time period t.

3.2 Action Space

Theoretically, the action space in this problem would be of nearly infinite dimensions if we allocate the vaccines completely region-wise. In fact, it is neither realistic nor practicable. Because first we need to satisfy the constraint that the sum of vaccines allocated to each region equal to the vaccine production η_t, which means the allocated vaccines to each region are linearly dependent; to the rule-makers, it is also impossible for them to select a policy from numerous options. Here, we propose the following action space.

As mentioned above, suppose at time t, the total vaccine production is η_t, which is obviously less than the total population in all regions $\sum_{l=1}^{L} M_l$. We firstly allocate the vaccines to each region based on the proportion of the region-wise population and state-wise population, that is $\frac{M_l}{\sum_{l=1}^{L} M_l} \cdot \eta_t$. For a specific region l with G groups of population, the action space contains finite options and these options are designed with the domain knowledge or by experts. For example, the rule-makers can either prioritize the vaccinations by the infected ratio of each group or by the susceptible ratio of each group, or combine these two perspectives together with each weights. In this case, the dimension of the action space for region l at time t is three.

3.3 State Transition

At the initial time period t_0, we need to first obtain the posterior of $\boldsymbol{\theta_t}$ in the G-SIR model with known data. Firstly, γ_t is the intrinsic property of Covid-19, which can be easily obtained with the current data and relative literature, while we can also estimate it dynamically. $\beta_{l,t}$ denotes the influence of taking each kind of actions in region l at time t. Here we use **Maximum Likelihood Estimation** to obtain the approximation of $\beta_{l,t}$:

$$\begin{cases} \hat{\beta}_{l,t} = \arg\max\{L(\beta_{l,t}|D)\} \\ \hat{\gamma}_t = \arg\max\{L(\gamma_t|D)\} \end{cases}$$

The results and proofs are given below.

From the G-SIR model, the changes of $X^S_{l,t}$ come from the Poisson random variable, naturally, for any t' in $[t_0, t]$, we would like to find a $\hat{\beta}_{l,t}$ such that maximizes the likelihood of all $e^S_{l,t'}$ happening accordingly.

$$\begin{aligned} L(\beta_{l,t}|D) &= \log\left(\prod_{t'=t_0}^{t} \frac{(\beta_{l,t} Z_{l,t'})^{e^S_{l,t'}}}{e^S_{l,t'}!} e^{-\beta_{l,t} Z_{l,t'}}\right) \\ &= \sum_{t'=t_0}^{t} \left[e^S_{l,t'} \log(\beta_{l,t} Z_{l,t'}) - \beta_{l,t} Z_{l,t'} - \log(e^S_{l,t'}!) \right] \end{aligned}$$

where $Z_{l,t'} = X^S_{l,t'} \frac{X^I_{l,t'}}{M_l}$
From the optima conditions,

$$\begin{cases} \frac{\partial L}{\partial \hat{\beta_{l,t}}} = \sum_{t'=t_0}^{t} (\frac{e^S_{l,t'}}{\beta_{l,t}} - Z_{l,t'}) = 0 \\ \\ \frac{\partial L^2}{\partial^2 \hat{\beta_{l,t}}} = -\sum_{t'=t_0}^{t} \frac{e^S_{l,t'}}{\beta^2_{l,t}} < 0 \end{cases}$$

Thus,

$$\hat{\beta}_{l,t} = \frac{\sum_{t'=t_0}^{t} e^S_{l,t'}}{\sum_{t'=t_0}^{t} Z_{l,t'}}$$

To estimate γ_t, we have:

$$\begin{cases} e^R_{l,t} \sim \text{Binomial}(X^I_{l,t}, \gamma_t) \\ E[e^R_{l,t} \sim \text{Binomial}|X^I_{l,t}] = \gamma_t X^I_{l,t} \end{cases}$$

Note that, since actions don't influence γ_t, all collected data (X^S, X^I, X^R) can be utilized in estimating γ_t. Given the data set, similarly, we can write the log likelihood function as:

$$\begin{aligned} L(\gamma_t|D) &= \log\left(\prod_{t'=t_0}^{t} \binom{X^I_{l,t'}}{e^R_{l,t'}} \gamma_t^{e^R_{l,t'}} (1-\gamma_t)^{X^I_{l,t'} - e^R_{l,t'}}\right) \\ &= \sum_{t'=t_0}^{t} \left[\log\binom{X^I_{l,t'}}{e^R_{l,t'}} + e^R_{l,t'} \log(\gamma_t) + (X^I_{l,t'} - e^R_{l,t'}) \log(1-\gamma_t) \right] \end{aligned}$$

From the optima conditions,

$$\begin{cases} \frac{\partial L}{\partial \hat{\gamma_t}} = \sum_{t'=t_0}^{t} (\frac{e^R_{l,t'}}{\gamma_t} - \frac{X^I_{l,t'} - e^R_{l,t'}}{1-\gamma_t}) = 0 \\ \\ \frac{\partial L^2}{\partial^2 \hat{\gamma_t}} = -\sum_{t'=t_0}^{t} (\frac{e^R_{l,t'}}{\gamma_t^2} - \frac{X^I_{l,t'} + e^R_{l,t'}}{(1-\gamma_t)^2}) < 0 \end{cases}$$

Thus,

$$\hat{\gamma}_t = \frac{\sum_{t'=t_0}^{t} e_{l,t'}^R}{\sum_{t'=t_0}^{t} X_{l,t'}^I}$$

Having known the parameters, we further denote the conditional transition probability $P(S_{l,t+1}|S_{l,t}, A_{l,t}) = f(\cdot|\cdot,\cdot,\hat{\boldsymbol{\theta}}_{\boldsymbol{t}})$. Again, $\hat{\boldsymbol{\theta}}_{\boldsymbol{t}} = (\hat{\gamma}_t, \hat{\beta}_{1,t}, \cdots, \hat{\beta}_{L,t})^T$.

3.4 Cost Functions

Since it is a multi-objective problem, we will discuss the individual cost function one by one. First, we want to minimize the impact by the Covid-19 after we prioritize the vaccination. Epidemiological cost $C_{l,t}^E$ is naturally defined as the number of new cases $X_{l,t}^S - X_{l,t+1}^S$. Second, we also want to minimize the transportation cost of delivering the vaccines. Suppose there are F factories which can produce the vaccines and they are located in the Michigan state. The unit transportation cost $p_{f,l}$ from each factory f to each region l is different and stochastic due to the external factors, for example, the weather. Suppose the total number of vaccines is N_v, and the corresponding production at each factory is N_f s.t. $\sum_{f=1}^{F} N_f = N_v$. Then the total transportation cost can be formulated as:

$$\begin{aligned}
\min \quad & C_t^T = \sum_{f=1}^{F}\sum_{l=1}^{L} N_{f,l} p_{f,l} \\
\text{s.t.} \quad & \sum_{l=1}^{L} N_{f,l} = N_f && \forall f \in F \\
& \sum_{f=1}^{F} N_{f,l} = N_v \cdot \frac{M_l}{\sum_{l=1}^{L} M_l} && \forall l \in L \\
& p_{f,l} \sim \text{Gaussian}(\mu_{f,l}, \sigma_{f,l}) && \forall f \in F,\ \forall l \in L
\end{aligned}$$

With the defined transportation cost, we can then combine the epidemiological cost and the transportation cost together with some predefined weight vector $\boldsymbol{w}$. Next, by choosing the wight vector, the agent can learn a deterministic policy $\hat{\pi}_{t_0,w}$ through implementing $A_{t_0} = \hat{\pi}_{t_0,w}(S_{t_0}; w, t_0)$. The following optimization problem is what the agent is going to solve to obtain $\hat{\pi}_{t0}(\cdot; w, t_0)$:

$$\hat{\pi}_{t0}(\cdot; w, t_0) = \arg\min_{\pi} E_{\pi,\boldsymbol{\theta_t}}\big[\sum_{t=t_0}^{T}\big(\sum_{l=1}^{L} C_{l,t}^E + w_t C_t^T\big)\big]$$

In this forward calculation, we denote the $E_{\pi,\boldsymbol{\theta_t}}$ as the total expected cost of $C_t^E = \sum_{l=1}^{L} C_{l,t}^E$ and C_t^T; the transition probability $P(S_{l,t+1} = s_{l,t_{t+1}}|S_{l,t} = s_{l,t}, A_{l,t} = a_{l,t}) =$

$f(s_{l,t+1}|s_{l,t}, a_{l,t;\boldsymbol{\theta_t}})$. Note that the transportation cost is only dependent on time and independent on the actions.

4 Model Application in Michigan study

In this section, we will apply the model proposed above to the real-world case. Some explanations and theoretic supplements are introduced. The numerical simulation is studied with the real-world data.

4.1 Data Preparation

For simplicity, all counties in Michigan are aggregated into eight regions and each region represents one corresponding basic operation unit. The regions are shown in Figure 2. To reduce the complexity of decision making, we broke the the regions into two groups: high and low risk. This division was done based on whether the daily cases in that region were more or less than the average daily cases for the whole state. This division then takes the number of action decisions from eight to two.

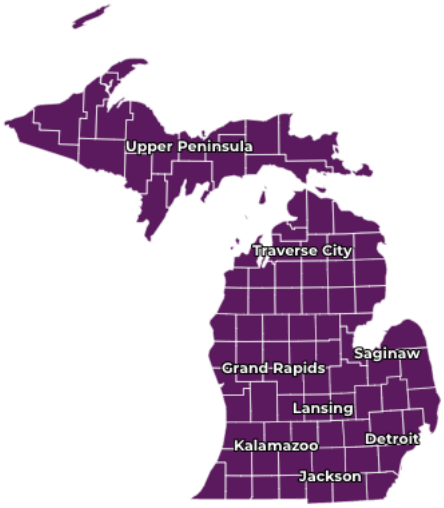

Figure 2: The eight regions in Michigan d

High risk	Saginaw	Grand Rapids	Kalamazoo	Jackson
Low risk	Upper Peninsula	Traverse City	Lansing	Detroit

Table 1: Daily Contact Matrix for Age Groups 0-19, 20-59, and 60+

Age	**0-19**	**20-59**	**60+**
0-19	0.792	2.238	0.444
20-59	2.238	1.365	1.397
60+	0.444	1.397	0.315

4.2 Infection Risk: Contact Matrix in Covid-19 pandemic

In Table 1, it shows the daily number of individuals that a person will contact of each age group. These were estimated using contact data from Wuhan and Shanghai by Zhang et al., while also considering contact intensities between age groups [11]. That intensity factor between groups was calculated in a paper written by Liu et al. [4]. It is important to note that we are assuming that the contact data between age groups in China is representative of Michigan. To address the uncertainty of that assumption, there will be sensitivity analyses done on the values in the matrix.

After trying to incorporate the contact matrix into our model, we noticed that the contact matrix values from China did not match well with the GSIR dynamics of infection spread. Hence, we then decided not to use the contact matrix in our final simulations. However, we believe that our framework should work well if the contact matrix for Michigan was found.

4.3 Design of Action Space

Recall that the target of this project is to prioritize the vaccination among different groups of population, which here is based on age. Theoretically, the number of possible actions are nearly infinite as given a number of vaccines allocated to one region, suppose the available number of vaccines for region l at time t is $\eta_{l,t}$, and there are G groups of population, then any action which satisfies the following constraints is feasible:

$$\begin{cases} \sum_{g=1}^{G} \eta_{l,t}^{g} = \eta_{l,t} \\ \eta_{l,t}^{g} \geq 0 \qquad\qquad \forall g \in G \end{cases}$$

where $\eta_{l,t}^{g}$ represents the number of vaccines allocated to group g population in region l at time t.

If we set the action space in this way, it is neither realistic nor computationally efficient. As the rule-makers, they don't have to decide exactly how many vaccines are allocated to

each groups, and in fact, many actions in this action space are redundant and they can be aggregated somehow. Therefore, we propose three basic but intuitive action options for rule-makers to choose from at each region l:

- Prioritize the vaccination based on the number of **possible** new cases among each group at time t, that is $\eta^g_{l,t} = \eta_{l,t} \cdot \frac{X^{I,g}_{l,t}}{X^I_{l,t}}$, where $X^{I,g}_{l,t}$ represents the new cases in group g in region l at time t;
- Prioritize the vaccination based on the ratio of susceptible population and infected population of groups, that is $\eta^g_{l,t} = \eta_{l,t} \cdot \frac{r^g_{l,t}}{\sum_{g=1}^G r^g_{l,t}}$, where $r^g_{l,t} = \frac{X^{I,g}_{l,t}}{X^{S,g}_{l,t}}$;
- Prioritize the vaccination based on the number of susceptible population of each group at time t, that is $\eta^g_{l,t} = \eta_{l,t} \cdot \frac{X^{S,g}_{l,t}}{X^S_{l,t}}$
- No prioritization and the vaccines will be equally distributed based on the ration of group non-removal population and total non-removal population, that is, $\eta^g_{l,t} = \eta_{l,t} \cdot \frac{X^{S,g}_{l,t} + X^{I,g}_{l,t}}{M_l - X^R_{l,t}}$

Here the fourth action can be tested as the benchmark without prioritization. There are also several assumptions needed in order to make the model coherent. Firstly, we suppose that the action is selected and conducted at the beginning of each time period but its effectiveness will not be observed until at the beginning of next time period, which means there is a one stage delay for vaccines to take into effects; Secondly, we suppose that the vaccines are absolutely successful, which means if $\eta^g_{l,t}$ number of vaccines are applied to group g, then the number of removal population of group g in the next time period will at least increase $\eta^g_{l,t}$, more precisely speaking, $X^{R,g}_{l,t+1} \geq \eta^g_{l,t}$. An accompanying consequence is $X^R_{l,t+1} - X^R_{l,t} \geq \eta_{l,t} + e^R_{l,t}$.

With the contact matrix, further inference about **possible infected** people in the next time period $t+1$ can be revealed by utilizing the contact patterns within and between each group of the population.

4.4 Numerical Simulation

The data we used was collected from the publicly available source of the Michigan State website, mainly the daily number of newly detected cases [2]. From the data which was collected, details were introduced with respect to the eight regions represented in figure 2. All the parameters used as the default setting in the model are listed in table 2

The population data by age was obtained through the United States Census Bureau and it was aggregated to reflect the age groups and regions mentioned above [1]. It is presented in table 3.

Table 2: Model parameters

Parameters	**Value**
Region L	2
Episodes	1,000
Simulation Period T/Day	29
Available Vaccines η/Day	24,000
Discounting factor α	1.0
Forecasting Window H	4
Detection Delay D/Day	9
Discounting Factor ϵ	0

Table 3: Data by age group

Region	Age group	Total Population	Confirmed Cases	Deaths
Upper Peninsula	0-19	62923	1488	<10
	20-59	144071	6845	<10
	60+	91857	3389	212
Traverse City	0-19	92056	755	<10
	20-59	204152	5045	15
	60+	148951	2849	156
Grand Rapids	0-19	397424	8492	<10
	20-59	793761	41414	71
	60+	343930	13222	712
Saginaw	0-19	136620	3591	<10
	20-59	289341	21813	94
	60+	180358	9830	827
Kalamazoo	0-19	240272	4212	<10
	20-59	485254	21578	54
	60+	238676	8210	496
Lansing	0-19	143153	2781	<10
	20-59	314538	11279	25
	60+	133411	3315	216
Detroit	0-19	1263276	17662	<10
	20-59	2739909	95673	742
	60+	1240358	38505	5095
Jackson	0-19	71966	1075	<10
	20-59	151356	5524	19
	60+	79244	2544	149

4.5 Algorithms

In the simulation, the main working flow is described as the follows: Firstly, the baseline results for comparisons are necessary. Recall that we have proposed four action options before, all the first three actions will be considered as the possible action taken at any

time period, and the last action will serve as the baseline action as it seems more like a commonsense to many people. Besides that, a random policy will also be tested and served as a baseline result; Secondly, to obtain the results with the randomness the model has, we sampled paths using Monte Carlo method with applying a certain policy and found using the transition function discussed in section 3.4. The prediction period was set to be $T = 4$ days. We assume that each of the regions receive a fixed number of vaccines proportional to the total population of that region. Our policy then decides the proportion of vaccines to be given to each age group at each decision epoch according to the available actions in the action space. The results are discussed in the next section.

For the Algorithm 1, it is used to find the best action for the current state. Acknowledging the randomness of the model and the large dimension of the state space, the prediction window was defined as 4 days ahead, which means the current best action would be selected by minimizing the accumulated cost in the following 5 days. Note that even though some actions are based on ratios, the similar logic will still apply: we have a fixed total number of vaccines for each region, and we can calculation how many vaccines to distribute for each age group with the given ratios.

Algorithm 1 Action Selection (for risk level L)

Input: S_t, Contact Matrix $\mathbf{\Omega}$, Infected probability $\boldsymbol{P_t}$, Prediction Window H
Output: A_t
Initialize Cost = 0
for $j = 1, 2, 3$ **do**
  Action[j], given $\mathrm{S}_t[X_t^I]$
  if $S_t[X_t^I] \neq 0$ **then**
    Explore all the possible sample paths from now till the next five time periods
    for $h = 0, 1, \cdots, H-1$ **do**
      for $i = 1, 2, \cdots, i \in G$ **do**
        Obtain infected people in each group $X_{t+h}^{I,i}$
        Estimate new cases by i $\hat{X}_{t+1+h}^{I|i}(j) = X_{t+h}^{I,i}\Omega[i,:]P_{t+h}(j)[:,i]$
      end
      Estimation of new cases $\hat{X}_{t+1+h}^{I}(j) = \sum_{i=1} \hat{X}_{t+1+h}^{I|i}(j)$
      Cost[j] += $\alpha^h X_{t+h}^{I,i}$
    end
  end
end
$\mathrm{A}_t = \arg\min_j \mathrm{Cost}$
return A_t

For the Algorithm 2, it is used to obtain the total accumulated cost by taking the action selected from the Algorithm 1. In order to mitigate the influence of high variance, we calculated the expected value from many episodes. Since the cost from the transportation

is not relevant, in our simulation, we just neglected this part and only solved the problem with respect to the epidemiological perspective.

Algorithm 2 Entire Time Horizon Cost Calculation

Input: S_{t_0}, T, Contact Matrix $\mathbf{\Omega}$, Infected probability $\boldsymbol{P}_{t_0}$
Output: $C_{t_0:T}$
Set $C = 0$, $C_l = 0\ \forall \ell \in L$
for $t = t_0; t \leq T-1$ **do**
  for $\ell = 1; \ell \leq L$ **do**
    $A_{\ell,t}$ = ActionSelection($S_{\ell,t}$,Ω,P_t)
    Update $\hat{\boldsymbol{\theta}}_{\boldsymbol{t}} = (\hat{\gamma}_t, \hat{\beta}_{1,t}, \cdots, \hat{\beta}_{\ell,t})^T$
    $S_{\ell,t+1} \leftarrow (S_{\ell,t}, A_{\ell,t})$, $P(S_{\ell,t+1}|S_{\ell,t}, A_{\ell,t}) = f(\cdot|\cdot,\cdot,\hat{\boldsymbol{\theta}}_{\boldsymbol{t}})$
    Update $\boldsymbol{P}_t$
    $C_\ell = C_\ell + S_{\ell,t}[X^S] - S_{\ell,t+1}[X^S]$
  end
  $C_{t_0:t} = w_t^E \sum_{\ell=1}^{L} C_\ell + w_t^T \sum_t C_t^T$
end
return $C_{t_0:T}$

For the Algorithm 3, it is used to search for the optimal policy. The Monte Carlo method was used to test the given trial policy (such as always taking action 4 during the whole time horizon) and obtain the expected costs. Then we updated the current policy if the new policy could outperform the existed one.

Algorithm 3 Policy Search

Input: S_{t_0}, T, Contact Matrix $\mathbf{\Omega}$, Infected probability $\boldsymbol{P}_{t_0}$, Episodes R
Output: $\hat{\pi}^*_{t_0}(\cdot; w_{t_0}, t_0)$
for $r = 1, \cdots, R$ **do**
  Obtain $C^r_{t_0:T}$ by applying Monte Carlo method (Initial Trial Policy)
end
$\hat{V}(S_{t_0}) = E_r\left[C^r_{t_0:T}\right]$, estimated value function of S_{t_0}, Set $\hat{V}(S_{t_0})$ as the baseline value
for $\hat{\pi}_{t_0}(i)$ **do**
  Apply the Monte Carlo method to obtain the expected costs $\hat{C}_{t_0:T}(i)$
  if $\hat{C}_{t_0:T}(i) < \hat{V}(S_{t_0})$ **then**
    $\pi^*_{t_0} = \hat{\pi}_{t_0}(i)$
  end
end
return $\pi^*_{t_0}$

5 Results and Conclusions

In this section, the results of the simulations will be presented. The simulation is run on Google Colab and the averaged running time is 19.39 minutes per scenario. Then we

conducted sensitivity analysis. The discussion and future work will also be addressed in this section.

5.1 Simulation Results

With the default parameter setting in the Table 2, we tested our model and compared the results of the optimal policy we found with the predefined baseline policy, as shown in the Figure 3.

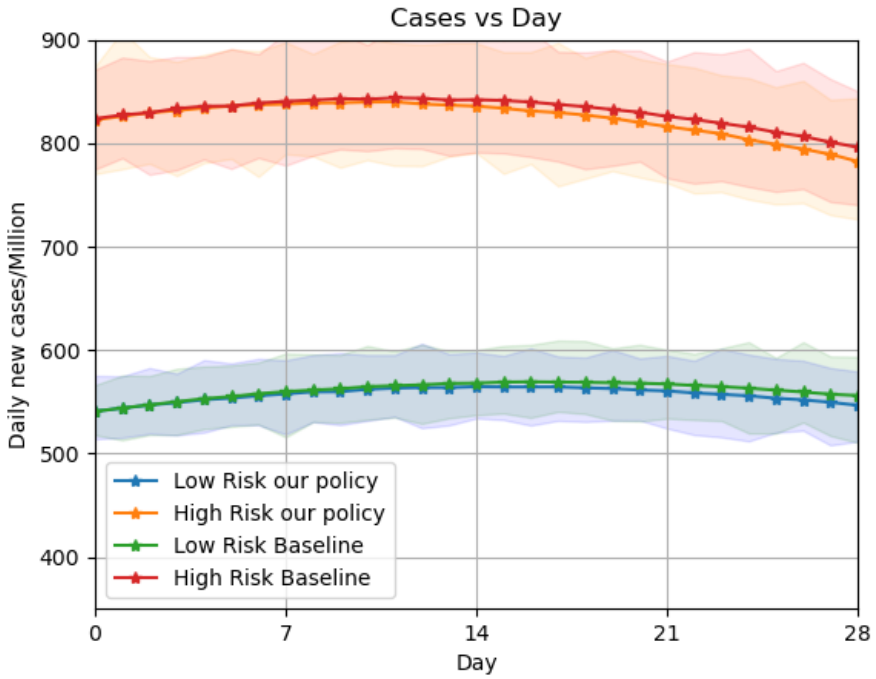

Figure 3: Comparison of simulated optimal policy and baseline policy

As indicated in the plot, we can see that the results of our optimal policy are better than the baseline policy which always takes action 4. The y-axis means the daily new cases per million. It takes quite a long time for our policy to be better than the baseline, which may be due to the limited action space or because of the limitations of our model.

5.2 Sensitivity Analysis

Table 4 shows the parameters that were changed in our sensitivity analysis. We wanted to test the validity of our model by changing key inputs. Based on those, we will be able to tell if it reacts correctly to adjusting scenarios.

In figure 4, the delay until an individual are detected infected is changed. This plot shows that the longer someone takes to know they are infected, the more daily cases there will

Table 4: Variation of Parameters

Parameters	Lower Bound	Default	Upper Bound
Detection Delay D	6	9	12
Vaccine Supply η/Day	18000	24000	30000

be as they could infect other people before they get quarantined. When they find out they are infected at an earlier time, the daily cases are reduced. Also, by changing the detection delay D, the estimation of transition probability will also be varied because the initial values of β and γ are dependent on D. With different initial β and γ, the following parameters of our model are altered throughout. The takeaway is that the government needs to take more tests for the public to decrease the detection delay D as much as possible.

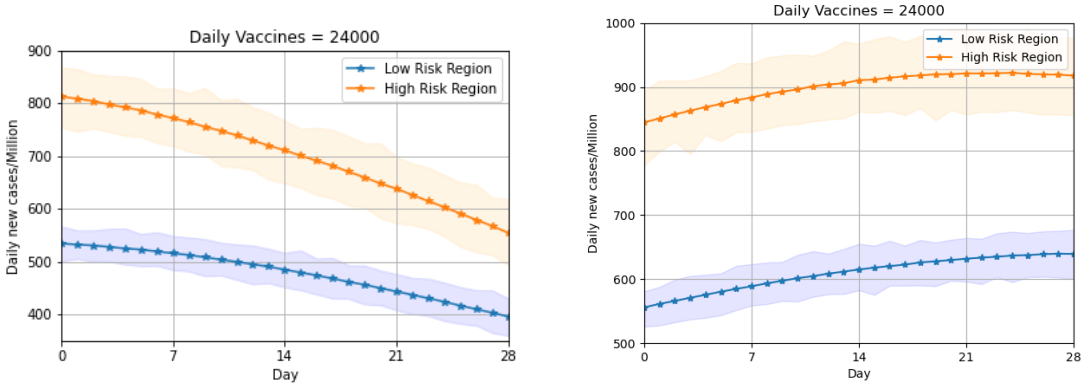

Figure 4: 6 (left) vs 12 (right) Day Detection Period

In figure 5, we compare the model using 18,000 and 30,000 vaccines supplied daily. We can see that when we raise the number of vaccines, the ending number of daily new cases is lowered. When we decrease the number of vaccines, the opposite happens. This was to be expected, but it shows that our model reacts correctly to changes in the parameters. Even though we increased the number of vaccines from the first day, the effect cannot been seen until the four to five days later because considering the detection delay is relatively large, the infected people can still infect others but as time goes, more accumulated benefit can be found.

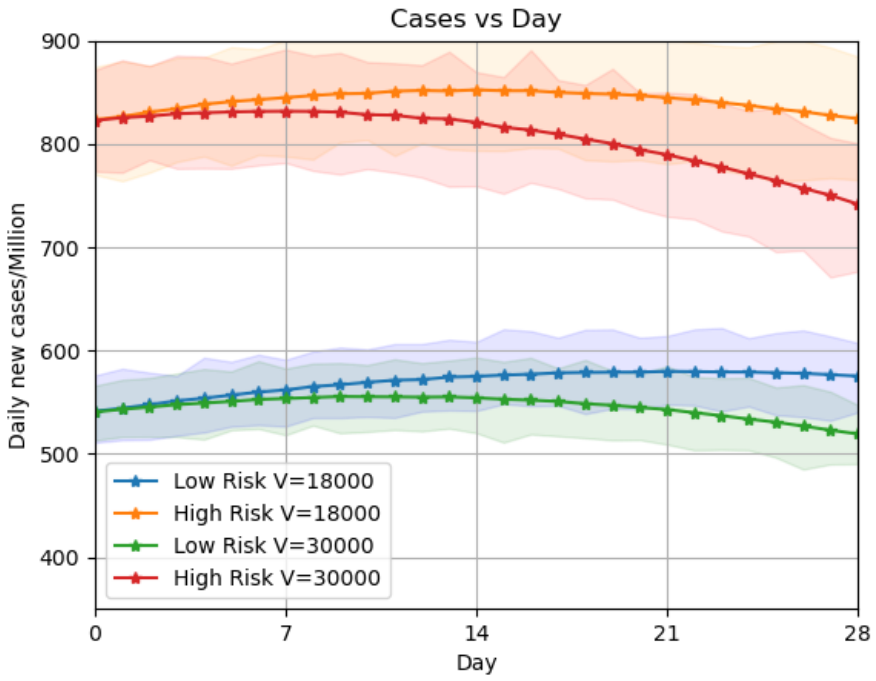

Figure 5: 18000 vs 30000 Daily Vaccines

5.3 Discussions

In this project, we proposed a Generalized SIR model to depict the dynamics of Covid-19 spread in Michigan State involving two random variables. How to prioritize the vaccination is somehow controversial especially when it needs to take every social and economic perspective into considerations. Our solution is to based on the age group of people in the state since there is an obvious age-dependent characteristic revealed according to the current data. We divided the whole population into three age groups and designed the heuristic action space. The results show that the costs obtained by applying our policy are lower than the baseline, which may be a commonsense policy to many people. Having said that, there is no doubt that the randomness of this model makes it vulnerable to many external disturbance. Also, the way we divide the population could be modified as the age cannot fully depict the intrinsic dynamics of disease spreading, but groups divided by different occupations could provide more inspection of how people will interact with others. While not many concrete researches have been conducted so far, some guidelines of how to prioritize vaccines have been posted online, where most of them suggest to prioritize the vaccines based on the occupations. One problem we encountered during the simulation was that the contact matrix we utilized was not accurate enough to serve as the ground truth. Similarly, the contact matrix for different occupations needs to be explored and that would be the key to this problem. Our model is simplified in terms of many aspects due to the computational limitation and available data. One of the

major assumptions we made was that the infection rate remained the same across each age group. This assumption was made because of data limitations. We believed that the difference in the total population of each age group should be enough to have different infection numbers in each age group. Having separate infection rates for different age groups would definitely improve this model.

Acknowledging the large dimension of the state space and the state-action pairs are one to many mapping, the best way may be using the Approximated Dynamic Programming (ADP) or using Reinforcement Learning method to deal with this issue. In the future, it is possible to apply Double Deep Q Networks (DDQN) since the action space is discrete and the state space is of a high dimension. One fact is that the randomness will defect the prediction and thus influence the action to take.

5.4 Workload Summary

Shreyas Bhat worked on data pre-processing and initial estimation of model parameters and coding.

Ruixuan Zhang worked on the design of the model and coding.

Isaac Smith worked on model design, document preparation, and code debugging.

Everyone contributes equally in this project.

References

[1] *2010-2019 Census Data*. 2020. URL: https://www2.census.gov/programs-surveys/popest/datasets/2010-2019/counties/asrh/ (visited on 12/17/2020).

[2] *Coronavirus-Michigan Data*. 2020. URL: https://www.michigan.gov/coronavirus/0,9753,7-406-98163_98173---,00.html (visited on 12/17/2020).

[3] C. Liu, X. Xu, and D. Hu. "Multiobjective Reinforcement Learning: A Comprehensive Overview". In: *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 45.3 (2015), pp. 385–398. DOI: 10.1109/TSMC.2014.2358639.

[4] Y. Liu, Z. Gu, S. Xia, B. Shi, X.-N. Zhou, Y. Shi, and J. Liu. "What are the underlying transmission patterns of COVID-19 outbreak? An age-specific social contact characterization". In: *EClinicalMedicine* 22 (2020), p. 100354. ISSN: 2589-5370. DOI: https://doi.org/10.1016/j.eclinm.2020.100354. URL: http://www.sciencedirect.com/science/article/pii/S2589537020300985.

[5] L. Matrajt, J. Eaton, T. Leung, and E. R. Brown. "Vaccine optimization for COVID-19: who to vaccinate first?" In: *medRxiv* (2020). DOI: 10.1101/2020.08.14.20175257. eprint: https://www.medrxiv.org/content/early/2020/12/15/2020.08.14.20175257.full.pdf. URL: https://www.medrxiv.org/content/early/2020/12/15/2020.08.14.20175257.

[6] K. A. McLean, S. Goldin, C. Nannei, E. Sparrow, and G. Torelli. "The 2015 global production capacity of seasonal and pandemic influenza vaccine". In: *Vaccine* 34.45 (2016). Ten Years of the Global Action Plan for Influenza Vaccines (GAP) - 2006 to 2016, pp. 5410–5413. ISSN: 0264-410X. DOI: https://doi.org/10.1016/j.vaccine.2016.08.019. URL: http://www.sciencedirect.com/science/article/pii/S0264410X16306764.

[7] D. M. Roijers, P. Vamplew, S. Whiteson, and R. Dazeley. "A Survey of Multi-Objective Sequential Decision-Making". In: *Journal of Artificial Intelligence Research* 48 (2013). DOI: https://doi.org/10.1613/jair.3987.

[8] R. Wan, X. Zhang, and R. Song. "Multi-Objective Reinforcement Learning for Infectious Disease Control with Application to COVID-19 Spread". In: *arXiv e-prints*, arXiv:2009.04607 (Sept. 2020), arXiv:2009.04607. arXiv: 2009.04607 [cs.LG].

[9] J. H. Wu, S. D. John, and E. Y. Adashi. "Allocating Vaccines in a Pandemic: The Ethical Dimension". In: *The American Journal of Medicine* 133.11 (2020), pp. 1241–1242. ISSN: 0002-9343. DOI: https://doi.org/10.1016/j.amjmed.2020.06.007. URL: http://www.sciencedirect.com/science/article/pii/S0002934320305544.

[10] R. Yaesoubi and T. Cohen. "Generalized Markov models of infectious disease spread: A novel framework for developing dynamic health policies". In: *European Journal of Operational Research* 215.3 (2011), pp. 679–687. ISSN: 0377-2217. DOI: https://doi.org/10.1016/j.ejor.2011.07.016. URL: http://www.sciencedirect.com/science/article/pii/S0377221711006187.

[11] J. Zhang et al. "Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China". In: *Science* 368.6498 (2020), pp. 1481–1486. ISSN: 0036-8075. DOI: 10.1126/science.abb8001. eprint: https://science.sciencemag.org/content/368/6498/1481.full.pdf. URL: https://science.sciencemag.org/content/368/6498/1481.