Probabilistic Programming on COVID-19

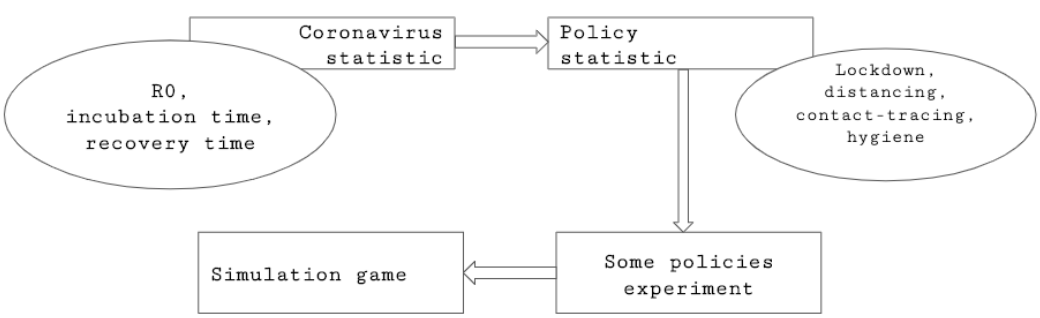
Mai Tung Duong (20180745), Assem Zhunis (20170906)

October 23, 2020

1 Introduction

Coronavirus disease (COVID-19) is an ongoing pandemic all over the world. How dangerous is COVID-19? How should the government respond to it? These are the questions that we are trying to answer in this work. Countries are imposing severe restrictions on their populations in a bid to stop the spread of coronavirus. The impact brought by these policies is very challenging to quantify. We propose a method to evaluate strength of these policies and simulate their impact on the virus spread considering economic trade-off. We aim to infer covid related parameters to simulate the virus, identify the efficacy of the policies applied by countries so far by using available data, and come up with the most optimal strategy to fight the virus. Our approach can provide a direction in improving the country policies in different sectors in order to overcome the pandemic with minimal loss. You can find the the source code of our project in this GitHub link [1].

2 Pipeline



[Fig. 1] Pipeline for the project

3 Dataset

We used COVID-19 Dataset from Kaggle [2]. The dataset includes number of confirmed, death and recovered cases starting from January 22nd 2020, divided by countries, regions, provinces.

4 Analyze important virus statistics by Virus Model

4.1 Mathematical model

We adapted the SEIR model from the epidemiology. The SEIR model assumes people carry lifelong immunity to a disease upon recovery. Here S, E, I, R stand for number of Susceptible, Exposed, Infectious and Recovered people respectively. We assume that population stays constant over time and is equal to N. Here are differential equations to describe the model mathematically:

$$\frac{dS}{dt} = -\frac{\beta IS}{N} + \alpha R \qquad \frac{dE}{dt} = \frac{\beta IS}{N} - \varepsilon E \qquad \frac{dI}{dt} = \varepsilon E - \gamma I \qquad \frac{dR}{dt} = \gamma I - \alpha R \quad N = S + E + I + R$$

4.2 Probabilistic Programming Model

pyro.contrib.epidemiology [3] is a framework for experimenting with a restricted class of stochastic discrete-time discrete-count compartmental models. We used this framework and modified it to fit SEIR model to infer COVID-19 related parameters: reproduction number Ro, recovery time, incubation time, transmission rate and mortality rate. For inference we used MCMC-NUTS algorithm.

4.3 Experiment and result

We run the model on the Sweden data before April 1st. The reason of our choice is that Sweden did not impose strict policies from the beginning. We can assume that virus transmission rate was unaffected and can be used for a simulation experiment with no policies applied.

Table 1: Inference Results

Recovery time	16.33 days	Lock down	0.96
Incubation time	3.83 days	Tracing + Quarantine	0.76
R0	2.40	Distancing	0.60
Mortality rate	2%	Mask + Hygiene	0.34
Recovery rate	27 %		

As we can see, results are consistent with the predictions of World Health Organization [4]. According to its estimations mild cases typically recover within two weeks, the incubation period is on average 5-6 days and R0 is typically around 2. Mortality and recovery rates differ depending on the region and stage of the virus spread. So, results are reasonable enough to use them in our simulations.

5 Estimate policy strength by Change-Point Model

5.1 Mathematical model

Consider the case count in log scale, the rate of transmission of the virus is represented by the slope. From the community spread onward, the graph is roughly segmented linear with 2 pieces of different slopes, w_1 and w_2. If the virus statistics remain the same, the policy applied at the changing point can be estimated to have strength of $1 - \frac{w_2}{w_1}$. We assume the model to be correct if changing point lies within the incubation period after the policy establishment.

5.2 Probabilistic Programming Model

This model is adapted from *Jonathan Ramkissoon* [5]. We use piece-wise linear regression with StudentT noise (which is more robust w.r.t outlier than Gaussian noise):

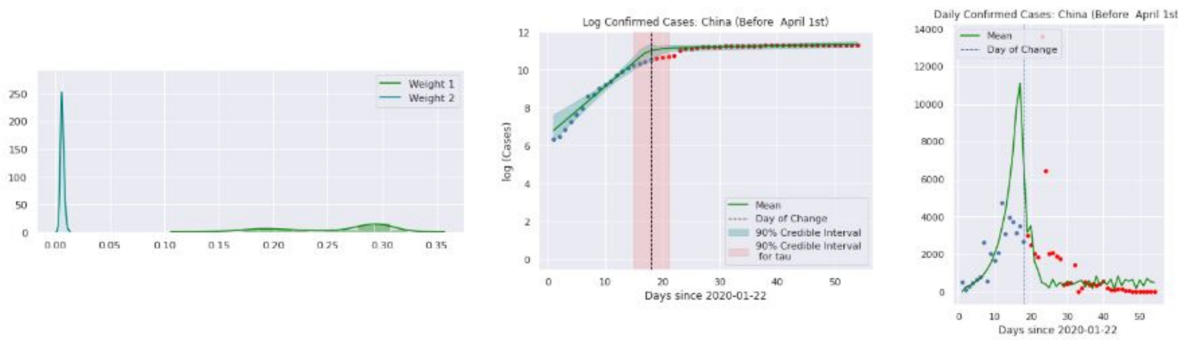
$y = wt + b + \epsilon$ where $\epsilon \sim StudentT(2, 0, \sigma^2)$.

Let τ be the change-point. Then for $t < \tau: \ b = b_1, w = w_1$ and for $t > \tau: \ b = b_2, w = w_2$.

Weights and biases have Normal prior and change-point has Beta prior. To better fit the data we used MCMC-NUTS inference algorithm.

5.3 Experiment and Result

We run the model on the countries that strongly impose a policy. The full efficacy of the policy is assumed to be the policy strength in that country. We considered China and New Zealand for evaluating lockdown strength, South Korea for contact tracing and quarantine and Canada for social distancing. For masks and hygiene, we compared the reproduction number R0 of Japan before policies were applied with the one we obtained from the Sweden analysis. The reason why we chose Japan lies in its cultural practices which lists the culture of wearing masks, very little “skinship” (such as hugging or shaking hands) and not wearing shoes in the house.



[Fig. 2] Posterior for China

All runs converged and the result is consistent with our prior (such as lockdown has very high efficacy). The posteriors also fit well with real data [7]. Result for policy efficacy is in Table 1.

6 Simulation by generative model

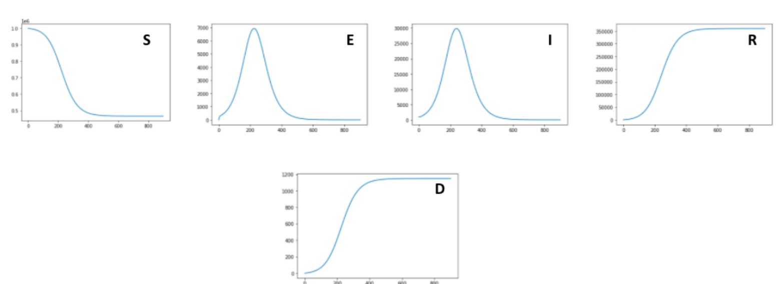
6.1 Model

We used an imaginary country with population 1 million and applied the same SEIR model with extended death toll for virus simulation. In addition to the parameters we inferred from previous results we assumed that hospital capacity is 0.06% of population and critical case rate is 15%. Death rate for critical case is 2% and 90% with and without medical treatment, respectively.

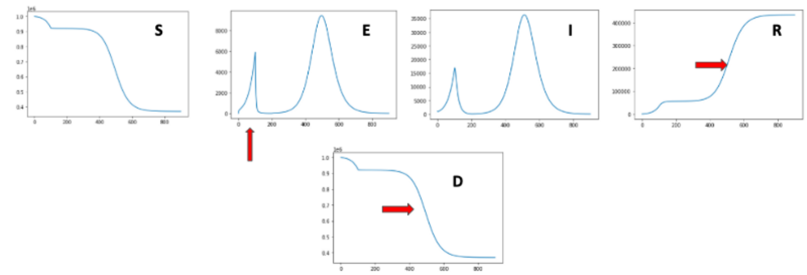
6.2 Experiment and Result

Run the model (1) without any policy and (2) with lockdown applied from days 100 to 200.

(1)



(2)



As you can see from the graphs, applying lockdown for 100 days just postpones the virus spread. Another problem is that it significantly hits the economy of country, so cannot be applied for a long time. Thus, even though lockdown is estimated to have the highest efficiency (0.98) it might not be the best policy to apply. So further experiments are required to identify how, when and for how long the policies should be applied.

6.3 Simulation game by generative model

To compute how much government strategy against virus costs to country, we propose to run the generative model on our imaginary world. We add the costs of each policy, treatment for infected and loss of death cases. Game mission: Derive a policy in 3 month period with 1-month interval policy. Each policy can be applied fully, partially (50% efficacy) or not applied at all. The game tries to minimize the cost.
Result: Best policy identified so far is full contact tracing + quarantine in all 3 month with loss of 67,411,086$. If we do nothing, the loss in our world is 3,355,503,642$ (scale up for the USA: 4 trillion $ per year).

7 What we learned?

About COVID-19 policy: From the results obtained, we can conclude that trade-off between virus prevention and economy should be considered more accurately. Decisions that government make regarding the policies affect the trajectory of the virus spread. So it is crucial to identify efficiencies and costs of each policy and estimate what would be the best time and strength to impose them. So far, we have found that quarantine and contact tracing are the most efficient policies, even though their strength is lower than that of lockdown. Indeed, we can see that countries that are most successful in controlling virus cases, eg. South Korea, use quarantine and contact tracing as their main policies.
About Probabilistic Programming (PP): We realized that Probabilistic Programming is an effective tool in simulating and inferring many real world problems. Even a complex problems like COVID-19 can be simulated quite well with our model and give reasonable predictions on how virus would spread given particular policies. It has simple interface that allows user to model different statistical or probabilistic systems without worrying much about background settings.
About Probabilistic Programming Language: We realized that understanding probabilistic language design is crucial in constructing the models. We found out that building the game model in Pyro is hard to do without understanding background gradient calculations. We need to study it more in order to finish our last task. We also noticed that inference algorithm really affects the efficiency and result of the problem. Knowledge of correctly setting the program hyper-parameters can greatly improve the accuracy and save time.

8 Future Work

We need to complete the game model by including more rules and restrictions on policy applications. To do so, further investigations of real countries strategies are needed. After that, we plan to solve the simulation game with continuous policy efficacy with generative modeling in Pyro.

9 Appendix

9.1 Results obtained by Virus Statistics Model:

	mean	std	median	5.0%	95.0%	n_eff	r_hat
rec_time	16.33	0.00	16.33	16.33	16.33	2.78	2.28
incub_time	3.83	0.00	3.83	3.83	3.83	7.60	1.42
R0	2.40	0.00	2.40	2.40	2.40	7.69	1.09
rho	0.54	0.03	0.52	0.51	0.58	2.39	4.38
mort_rate	0.02	0.00	0.02	0.01	0.02	2.38	5.53
rec_rate	0.27	0.00	0.27	0.27	0.27	2.38	4.84

Sweden before April 1st

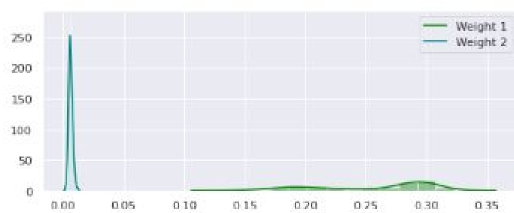
	mean	std	median	5.0%	95.0%	n_eff	r_hat
rec_time	16.00	0.00	16.00	15.99	16.00	3.09	1.92
incub_time	6.53	0.00	6.54	6.53	6.54	7.69	1.00
R0	1.58	0.01	1.57	1.56	1.60	2.48	2.73
rho	0.29	0.01	0.29	0.28	0.30	2.58	2.46
mort_rate	0.02	0.00	0.02	0.02	0.02	2.50	2.79
rec_rate	0.14	0.01	0.14	0.13	0.14	2.57	2.51

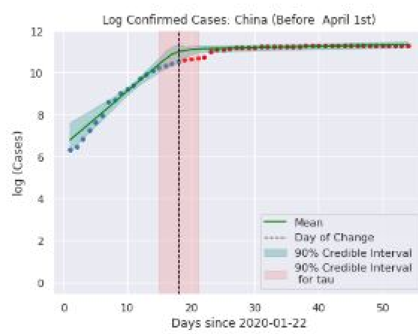
Japan before April 1st

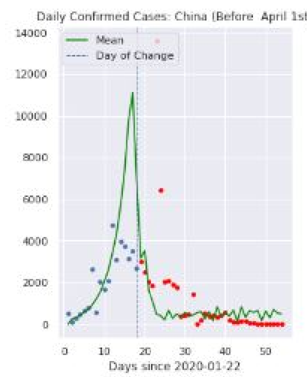
Calculated strength of hygiene and masks: 1 - 1.58/2.40 = 0.34

9.2 Results obtained by Change Point Model:

9.2.1 China

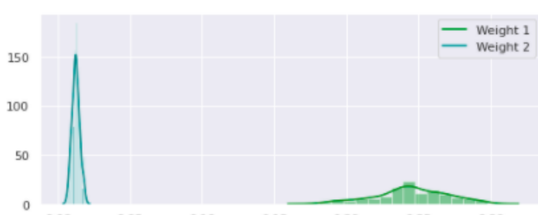


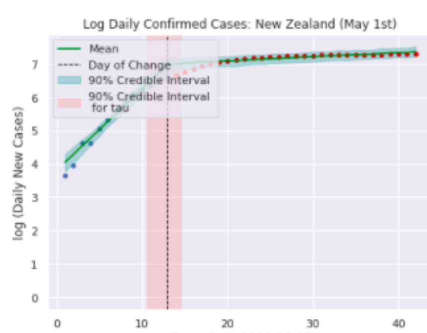




Changing point for China: 2020-02-09.
Lockdown applied: 2020-01-23.
Lockdown strength: $1 - 0.01/0.26 = 0.96$

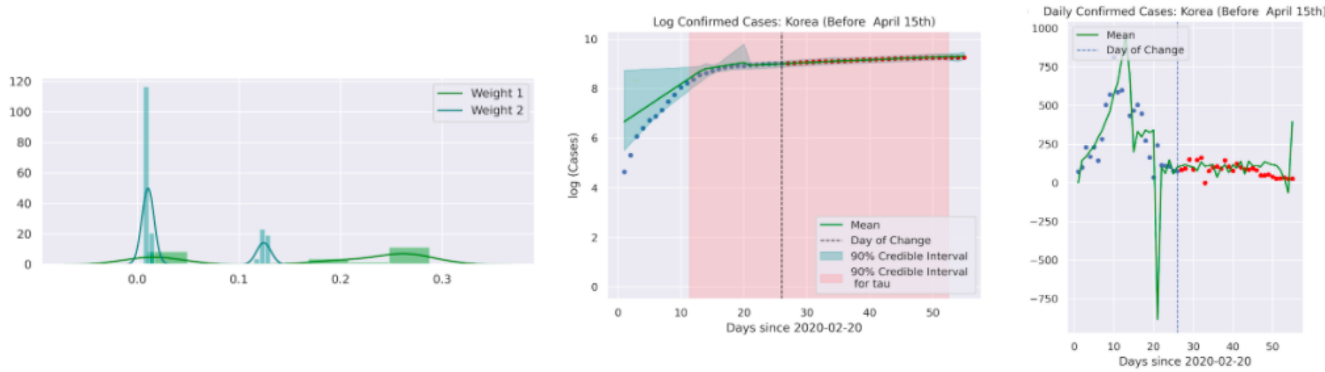
9.2.2 New Zealand





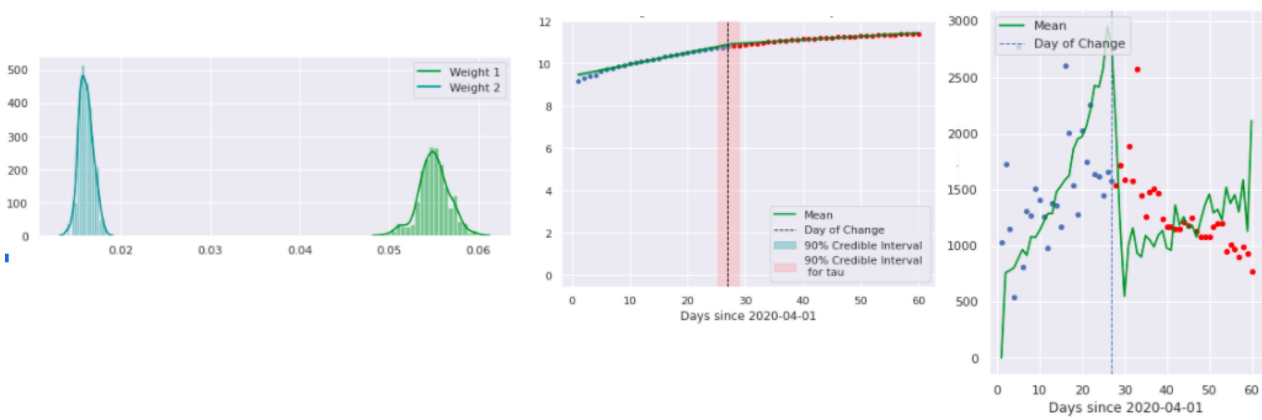
Changing point for New Zealand: 2020-04-02.
Lockdown applied: 2020-03-25.
Lockdown strength: $1 - 0.01/0.25 = 0.96$

9.2.3 South Korea



Changing point for South Korea: 2020-03-17.
Policy applied: 2020-02-25.
Contact tracing strength: $1 - 0.04/0.17 = 0.76$

9.2.4 Canada



Changing point for Canada: 2020-04-28.
Policy applied: March-April.
Distancing strength: $1 - 0.02/0.05 = 0.60$

9.3 Current Simulation Game Rules

We estimate the economic cost for each policy and give negative rewards:
Lockdown: 10% of GDP per year (GDP per capita: 30k\$).
Distancing: 5% of GDP per year.
Hygiene: 2\$ per day per capita.
Quarantine: 100\$ per day per infected.
Infection: 300\$ per infection per day (until recovered).
Death: 4.9 million\$ per death.
Further research is needed to improve these rules.

References

[1] GitHub link to the project repository
https://github.com/assemzh/ProbProg-COVID-19

[2] Data set for COVID-19 cases can be found here.

[3] *Epidemiology Model* http://docs.pyro.ai/en/dev/contrib.epidemiology.html

[4] World Health Organization website
https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-c

[5] Jonathan Ramkissoon, *Detecting Changes in COVID-19 cases with Bayesian models* https://towardsdatascience.com/detecting-changes-in-covid-19-cases-with-bayesian-models-1b6282

[6] Japanese culture effect on COVID-19 spread article
https://www.japantimes.co.jp/community/2020/04/01/voices/japan-culture-low-covid-19-numbers/.X

[7] Marcel Salathé and Nicky Case , *What Happens Next? COVID-19 Futures, Explained With Playable Simulations*, available at https://ncase.me/covid-19/.

[8] Tim Churches, *Simulating COVID-19 interventions with R*, available at
https://rviews.rstudio.com/2020/03/19/simulating-covid-19-interventions-with-r/.