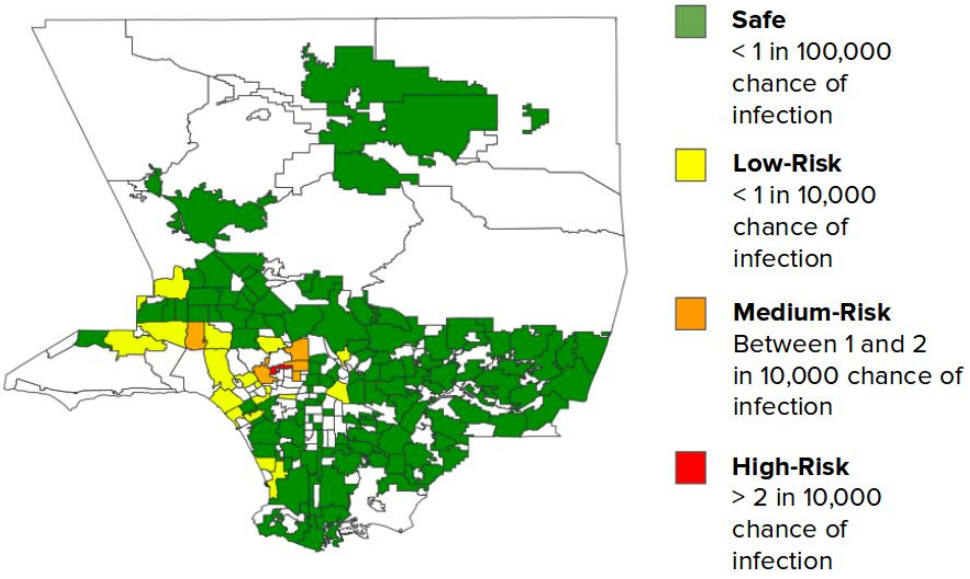
COVID-19 Risk Estimation for Los Angeles County using a Bayesian time-varying SIR-model

Mehrdad Kiamari, Gowri S. Ramachandran, Quynh Nguyen, Bhaskar Krishnamachari
Team Name: USC_ANRG
Autonomous Networks Research Group
Viterbi School of Engineering, University of Southern California
June 16, 2020

Executive Summary: We have developed a rigorous hybrid model-and-data-driven Bayesian approach to risk scoring based on a time-varying SIR epidemic model that yields a simplified color-coded risk level for each community. The risk score corresponds to the probability of someone currently healthy getting infected with COVID-19 in the near future. Code and data from our effort have been open-sourced. We apply our approach to assess the risk of infection in various communities in the County of LA.



At a glance: the result of our algorithm showing color-coded risk levels for different communities within LA County based on the probability of infection

1. Introduction

In the context of the ongoing COVID-19 epidemic, there is a need for a systematic way to assess and present the risk associated with different communities and locations. We present here a methodology and the associated software to assess the current and predicted COVID-19 infection risk for different sub-communities that is based on a novel Bayesian time-varying extension of the well-known SIR mathematical model of infection spreads, and demonstrate its use with LA County data.

The block diagram below illustrates key elements of our system design. Our data parser is able to get the raw data from online data sources, clean them up and store them in machine-friendly (csv and json) formats. Our code for infection risk calculation uses this data in conjunction with a time-varying SIR-based Bayesian mathematical model to obtain risk estimates and prediction for different communities. The results are provided in CSV format and can be used to generate a heatmap-type visualization as well.

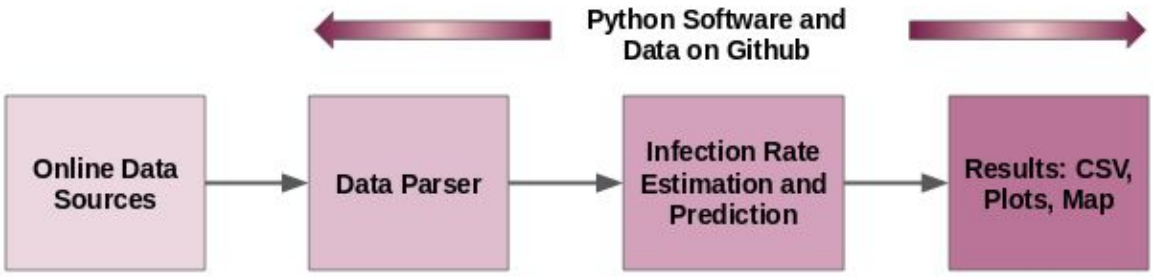


Figure 1: Overview of our system

2. Data Sources

We have acquired COVID-19 case data from the LA County's Department of Public Health using a Python-based data parser, which is made available to the public as open-source software at the following link: https://github.com/ANRGUSC/lacounty_covid19_data. We have been updating this repository regularly with the latest data every day since mid-march and also making available plots of the number of cases, number of fatalities, top 6 communities with the large number of cases, infection rate for the entire LA County, and the top 9 communities with the highest infection rate at the following link: http://anrg.usc.edu/www/covid19.html.

The following data sources are used for the infection rate and prediction:

- The *CoVID-19 case information* was collected through LA County's daily press releases, which can be accessed through the following website: http://publichealth.lacounty.gov/media/Coronavirus/
- The public health department of LA County does not report statistics about recovery. Therefore, we have used the recovery information provided by the World Health Organization.
- The population data collected from LA County's Census website: https://lacounty.gov/government/geography-statistics/cities-and-communities/

3. Methodology

Compartmental mathematical models for epidemic spreads including the well-known SIR model have been used since the work of Kermack and McKendrick in 1927 [1]. In the SIR model, illustrated below, each member of a given population is in one of three states at any time: susceptible, infectious, recovered. Any individual that is susceptible could become infected with some probability when they come into contact with an infected individual. Any individual that is infectious eventually recovers (in the context of COVID-19 when applying the SIR model, note that the category of recovered individuals will also include removed individuals due to deaths, which could be modeled as a constant fraction of all individuals in this category). In the classical SIR model, the number of susceptible individuals that become infected depends on the rate at which infected and susceptible individuals encounter each other and this rate is assumed to be constant. A well-known parameter in the classical SIR model is called R0, the effective reproductive number, which measures the average number of infections caused by infectious individuals at the beginning of the epidemic.

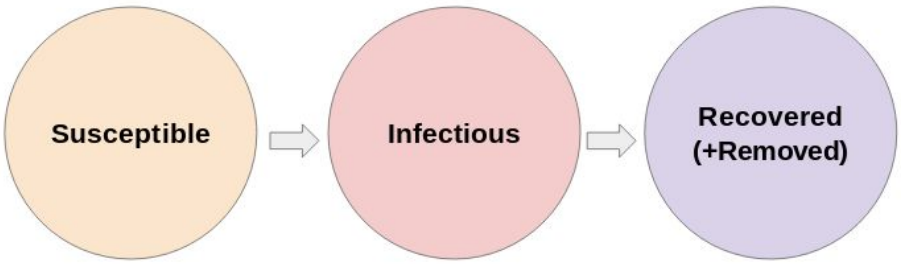


Figure 2: Classical SIR model of Epidemic Spread

3.1 Time-Varying SIR model and R_t: In our work, we have extended the SIR model to a time-varying model, in which the rate of encounters and infection probability between individuals in the population is assumed to be time-varying. This better reflects the reality of our present epidemic where interventions such as stay-at-home have been put in place and relaxed and

various times and compliance with recommendations such as wearing masks and maintaining physical density has also been time-varying. Based on this model, we are able to define and derive a new approach to calculating a time-varying version of the effective reproductive number, which we refer to as R_t.

A particularly innovative aspect of our model is that it is a Bayesian model that allows the incorporation of various sources of uncertainty in the model, including uncertainty in the actual numbers of infected individuals (due to not every infected individual having been tested, as studies [2] have shown), uncertainty in recovery times, and uncertainty in the choice of parameters for de-noising the empirical data. This allows us to generate not only an estimate of R_t, but also quantify confidence in the estimate from a rigorous statistical perspective. For ease of exposition, we have omitted the mathematical details of our approach from the body of this report, but they can be seen in the appendix.

3.2 Risk score calculation: Using the above-model to generate estimates of the effective reproductive number (R_t), we are able to calculate a risk score. We define a novel risk score metric for a given community that is proportional to **the probability of someone in that community becoming infected in the near-term.** This probability can be derived as the average number of people in that community that are likely to get infected by the currently infectious people divided by the current number of susceptible individuals (which is approximately equal to the total population of the community at the current relatively early stage of the epidemic). We normalize this probability by multiplying by 10,000, so that a score or 1 implies a 1 in 10,000 chance of getting infected, a score of 2 implies a 2 in 10,000 chance of getting infected, and so on.

For each community, the risk score is defined as follows:

$$Risk\ Score = \frac{number\ of\ infectious\ people \times effective\ reproductive\ number \times 10{,}000}{population\ size}$$

3.3 Color-coded Risk-levels: To further simplify the presentation of the risk score to a wider audience, we propose to classify the risk levels into four color-coded levels:

1. Safe (Green) - we set this to correspond to risk score < 0.1 i.e. less than 1 in 100,000 chance of getting infected in the near term, in that community
2. Low-Risk (Yellow) - we set this to correspond to risk score between 0.1 and 1, i.e., less than 1 in 10,000 chance of getting infected in the near term in that community
3. Medium-Risk (Orange) - we set this to correspond to a risk score between 1 to 2, i.e. between 1 in 10,000 and 2 in 10,0000 chance of getting infected in the near term in that community.
4. High-Risk (Red) - we set this to corresponding to a risk score greater than 2, i.e. greater than 2 in 10,000 chance of getting infected in the near term in that community.

Our framework allows changing the above threshold based on additional input from health experts and public authorities.

The attached CSV file shows the risk score and risk level for each community in Los Angeles County.

3.4 Prediction: Going beyond assessing the present risk, our methodology also allows the estimation of risk in the future by using a regression-based estimate of the risk score. We apply linear regression on the past 7 days of the risk score to predict the prediction score for the next day. By iterating this approach on a sliding basis, the risk score can be predicted for future days beyond the next day as well.

4. Results

We present below plots from our analysis of LA County community case data in terms of aggregate as well as community-specific effective reproductive number (R_t) and our risk score, which is proportional to the infection probability based on a 14-day moving average applied on the daily number of confirmed cases. Finally, we present maps showing color-coded risk levels for communities in LA County for select dates over the past 3 months.

4.1 Aggregated results for whole of LA County

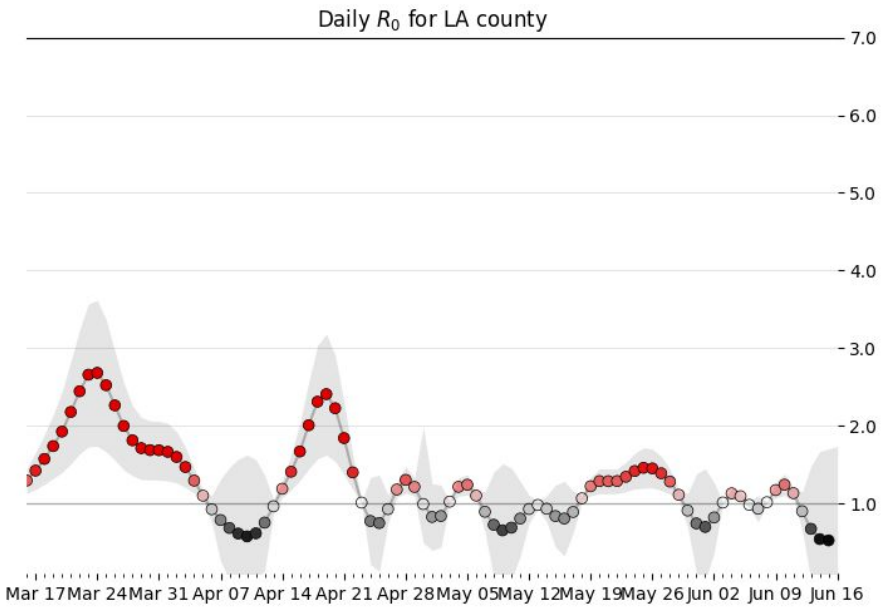


Figure 3: Effective reproductive number (R_t) over time for the entire county of LA

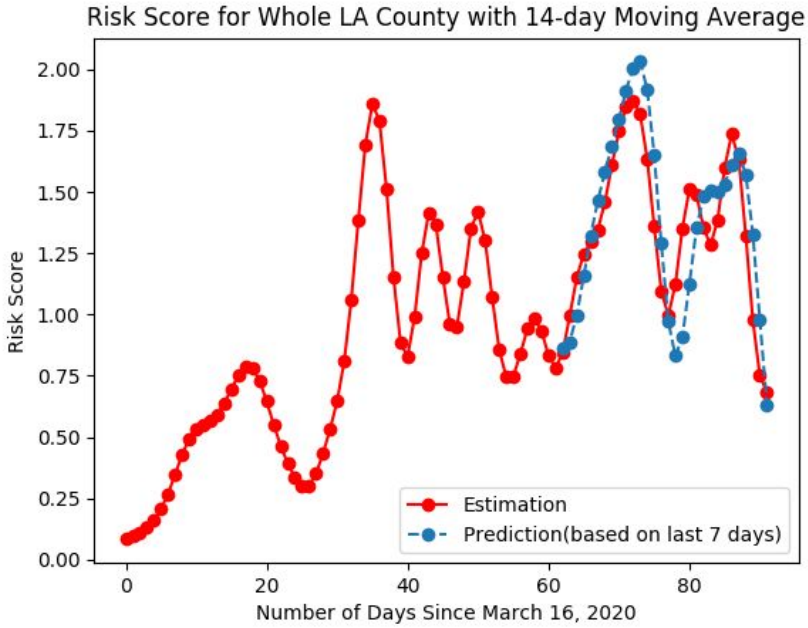


Figure 4: Risk score over time aggregated for the entire county of LA

4.2 Community-specific results

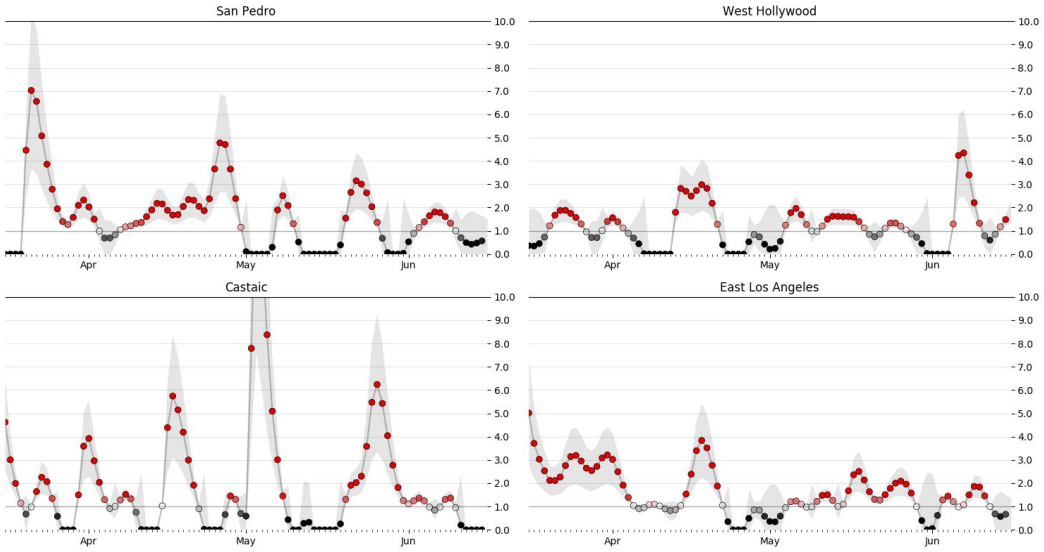


Figure 5: Estimate of effective reproductive number R_t over time for four representative communities in LA County with different profiles: San Pedro, West Hollywood, Castaic, East LA. Our Bayesian approach also yields uncertainty in the estimate, as shown in the form of confidence intervals (in gray).

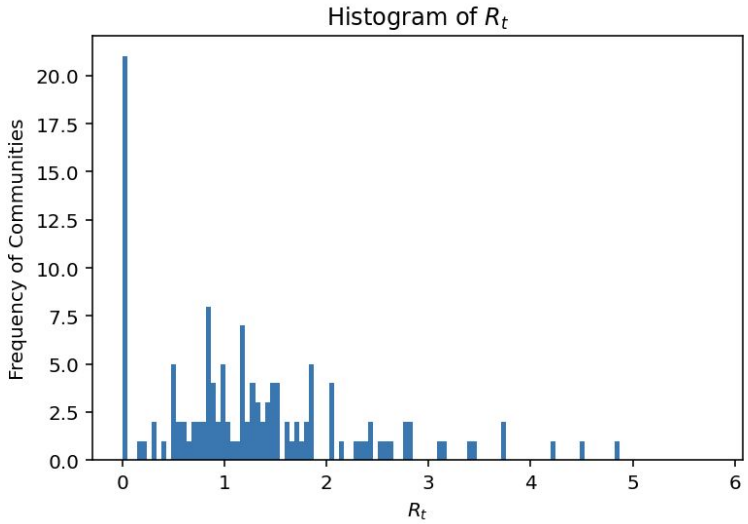


Figure 6: Histogram of R_t across all communities in LA county as of June 15, 2020

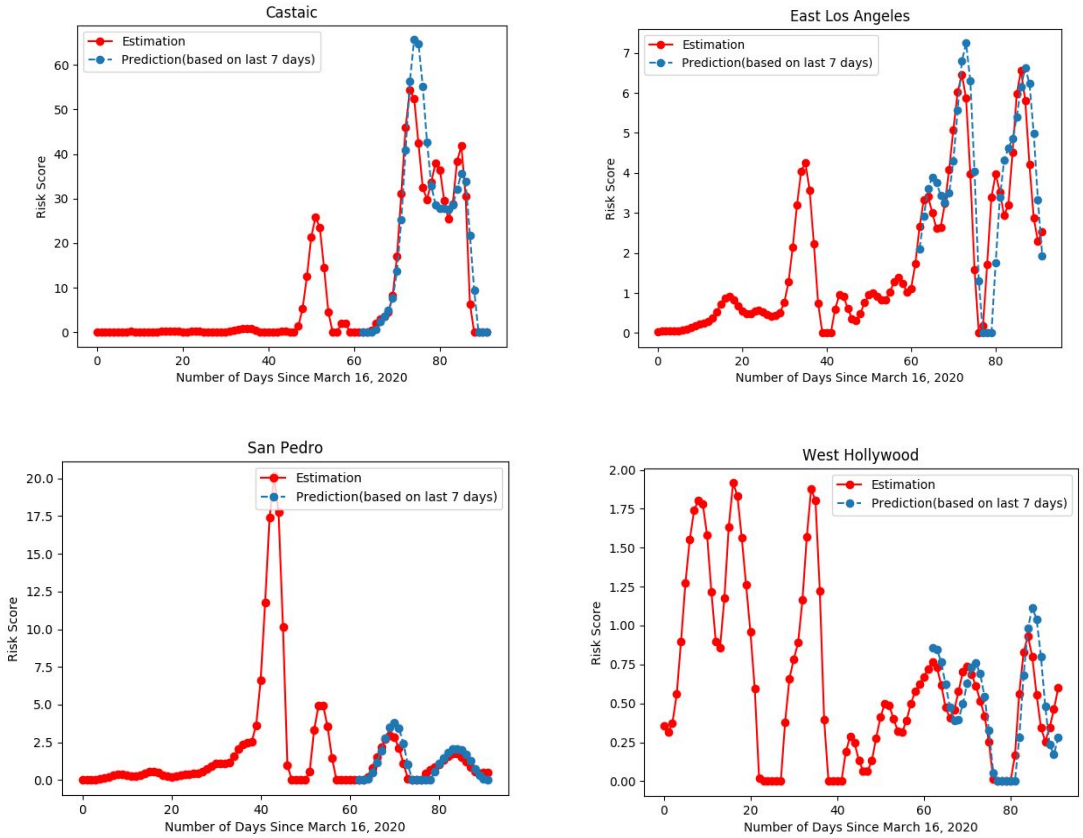


Figure 7: Estimation of Risk Scores for four representative communities in LA County (in red), along with predictions for the last 30 days based on 7-day historical data (in blue).

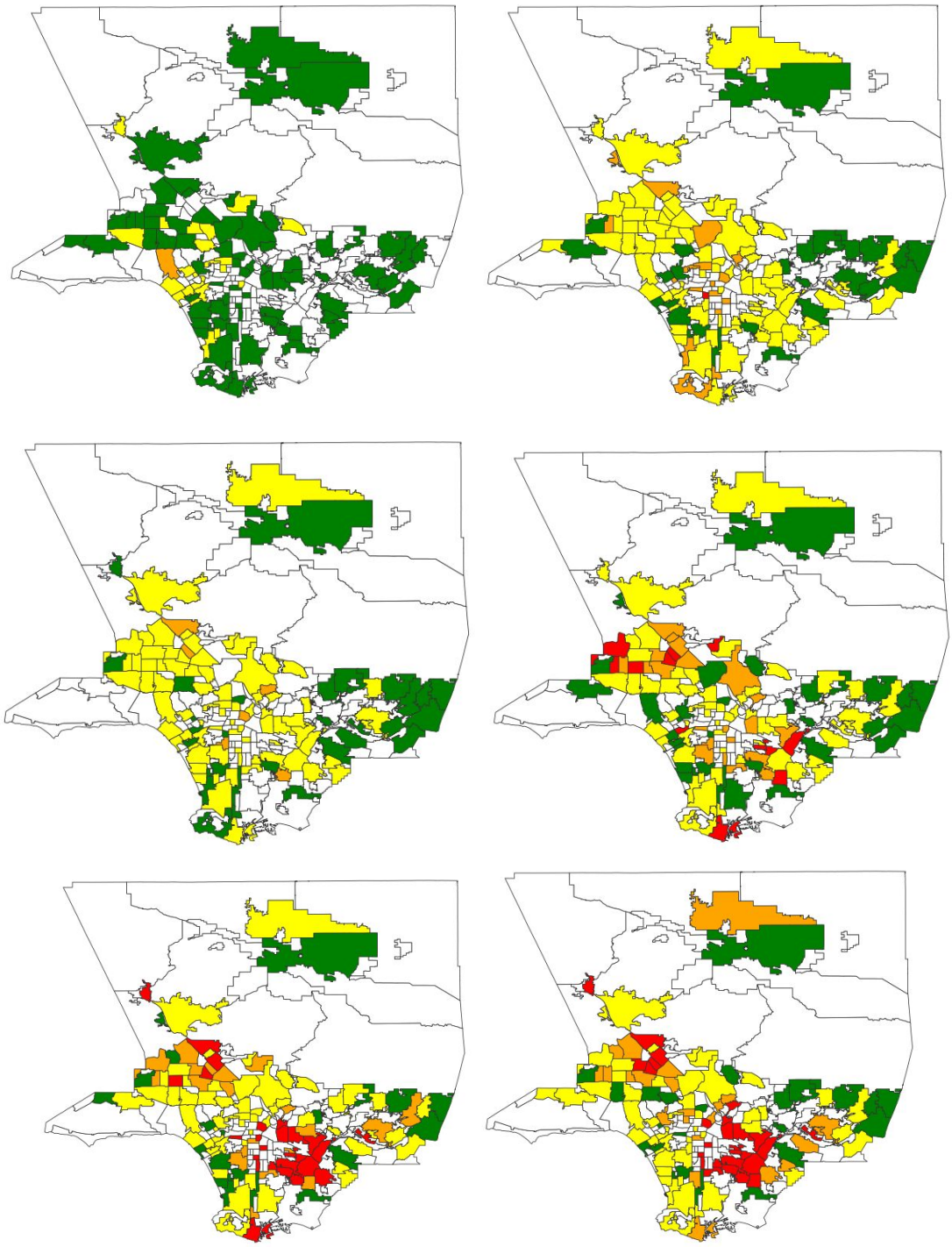


Figure 8: Maps showing the estimated risk score for different LA County Communities on different dates since mid-March 2020. Top row: 3/23/20, 4/6/20; Middle row: 4/20/20, 5/4/20; Bottom row: 5/18/20, 6/15/20

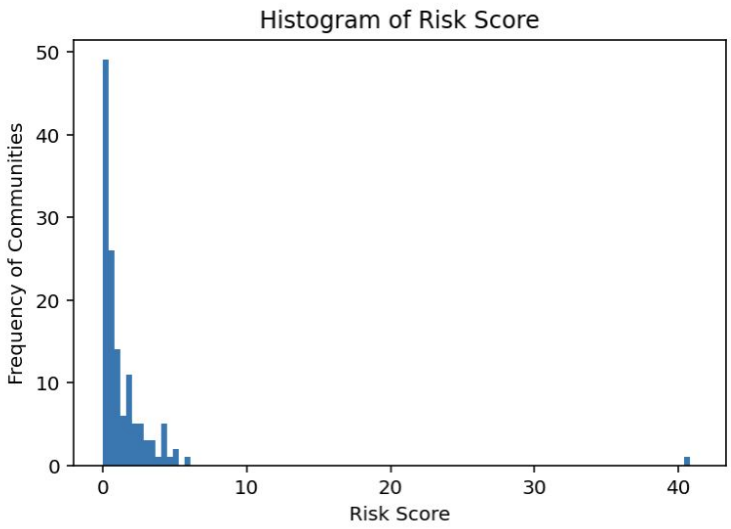


Figure 9: Histogram of Risk Scores for communities in LA County as of June 15, 2020

5. Implementation

The software for data collection, infection rate estimation and prediction has already been implemented and made available as open-source software in the following repositories:

- https://github.com/ANRGUSC/covid19_risk_estimation
- https://github.com/ANRGUSC/lacounty_covid19_data

All our software is written in Python using standard data processing libraries such as NumPy and SciPy. For creating the animated map showing the infection rate for the past two months, we have used geopandas to generate the maps, and an online service (https://ezgif.com/maker), which converts maps from multiple days into an animated GIF file. We believe that it is easy to use our software for daily infection rate estimation and prediction, with minimal effort.

5.1 Extensions and Future Work: To further improve the system, methodology and approach presented in this work, we have a number of possible directions:

- Our software is currently a backend system that needs to be run daily to produce meaningful outputs - it should be connected to a live front-end website or mobile app for better end-user experience. Such a web/mobile app could also be extended to provide support for the querying of infection rates based on community name or ZIP code.
- Clean up ambiguities between data sets we have on population counts for LA county communities and the data on case counts from LA County public health
- Improve the model by explicitly making use of not only the daily new case data but also the daily mortality data.

- Incorporate data on inter-community mobility of people such as may be obtained from cellular or mobility data providers
- Incorporate incubation period in the model by extending to the SEIR model
- Separate out high-risk populations within the same community: this would require more fine-grained data on numbers of individuals in prisons, industries involving dense-interactions, old-age homes, large companies, schools, etc.
- Correlate measured data with and improve risk prediction using news articles and other sources of information to help policy makers and community members understand root causes of virus spread and take quick actions.

6. Risk Mitigation Recommendations

Based on our analysis and understanding of the dynamic and fine-grained nature of risk levels, we make the following recommendations:

- It is important to focus on community-specific risk assessment as there are clearly significant differences in the risk score across different communities within LA County
- Put “boots on the ground” and news sources to gain a deeper understanding of exactly what is happening in each community, particularly those with higher risk levels.
- Communicate the risk levels to the broader public within each community with appropriate contextual news and information about what is happening within their localities, particularly in communities with higher risk-levels
- Implement targeted interventions in communities with high risk levels. For example, more strict stay at home or business closure requirements in certain cities or communities that show consistently high risk levels over a period of time.
- Implement a regime of manual contact tracing, allocating greater resources and attention to communities with greater risk levels.
- Prioritize allocations to ensure that adequate testing and healthcare resources are available in communities with greater risk level

7. Conclusion:

In conclusion, we would like to summarize some of the merits of our approach.

- **Innovation:** Our algorithms are novel, based on a new synthesis of time-varying extension of the classic SIR model and a Bayesian data-driven approach. Our risk score metric is novel and intuitive, pertaining to the risk of infection.
- **Impact and Usability:** We believe the use of color-coded levels will make it easy to communicate risk levels to citizens, and the community-specific risk estimation and prediction will enable more targeted interventions.
- **Methodology Validity and Reproducibility:** We have confidence in our methodology given our team’s significant prior expertise with mathematical modeling, statistical inference and machine learning. Our mathematical derivations and algorithms will be

posted online initially as an arxiv.org preprint, and soon as a submission to a suitable peer-reviewed venue. Our code and data are made available as open source for other experts to review and build upon.

- **Software Deployment:** Our backend software is easy to deploy and can be extended readily to other communities beyond LA County by changing the data sources accordingly. REST-based web service and mobile app front ends could be built easily based on our software for easier visualization and communication.
- **Ethical and Fair Use of Data:** Our modeling is based on data released by the LA County public health department since mid-March, and we are not using any privacy-sensitive data in our work.
- **Inclusivity and Diversity:** Our team consists of members from multiple ethnicities and is also gender-diverse.

Acknowledgment

We thank the LA County Public Health department for releasing community-specific COVID-19 case data which has been essential for our modeling and analysis. Our approach to estimate the effective number R_t based on a time-varying SIR model is quite different from the scheme described by Kevin Systrom (http://systrom.com/topic/modeling/ , the basis for https://rt.live/) but we were motivated by his work and have used Python code provided by him for plotting our R_t curves.

References

[1] Kermack, W. O.; McKendrick, A. G. (1927). "A Contribution to the Mathematical Theory of Epidemics". Proceedings of the Royal Society A. 115 (772): 700–721.

[2] Neeraj Sood, Paul Simon Peggy Ebner et al, Seroprevalence of SARS-CoV-2–Specific Antibodies Among Adults in Los Angeles County, California, on April 10-11, 2020, JAMA, published online May 18, 2020. https://jamanetwork.com/journals/jama/fullarticle/2766367

Appendix: Technical Details of Estimating Effective Reproductive Number R_t using a Bayesian Time-Varying SIR Model

I. Introduction

Epidemics caused by infectious viruses such as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-COV-2) which causes COVID-19 can have a significant negative impact on public health. Such epidemics force governments and public authorities to employ stringent measures [1], [2], including lockdowns, to contain the spread. When making such decisions, policymakers require tools to understand in "real-time" how the virus is spreading in the community.

One metric that is useful for authorities to assess the level of containment over time is the effective reproduction number [3]. The effective reproduction number, R_t, indicates on average how many currently susceptible persons can be infected by a currently infected individual. The epidemic grows if this measure is above one. Authorities desire to keep this value as far below below one as possible over time in order to contain and eventually, hopefully, eliminate the virus from the community.

One challenging aspect of COVID-19 has been that there are significant numbers of asymptomatic infections [4]. In other words, individuals with asymptomatic infections, who are not aware of their status can go on to infect other persons they come into contact with. The typical symptoms of COVID-19 consist of shortness of breath, cough, fever, and diarrhea. Based on a recent report from WHO [4], around 12% and 32% of patients do not have a fever and dry cough, respectively, which indicates that a considerable amount of patients are not aware that they are contagious and this can severely imperil the health of the community.

Since hospital capacity is limited, it is crucial to flatten the curve that represents the number of confirmed cases over time. Otherwise, the number of fatalities significantly increases because some patients who get into critical conditions cannot be admitted at hospitals that are already full of patients. To contain the outbreak, countries around the world have resorted to strict measures such as monitoring traffic through the city, issuing stay-at-home orders, or locking down the entire city. However, as there is an economic cost to these measures, there is a desire to track and understand the ongoing rate of infection to decide when to ease these containment measures. We therefore focus on the important problem of estimating the daily effective reproduction number R_t.

One of the well-known models to analyze any infectious diseases is the Susceptible Infected Recovered (SIR) model, which consists of three states, namely the susceptible state, the infected state, and the recovered state. While traditionally, this model is assumed to have a interaction rate / infection rate parameter that is constant, one recent work has used a time-varying SIR model to recover the time-varying effective reproduction number [5]. However, that work doesn't provide any approach to derive a confidence interval on R_t, which is a key focus of our work. Further [5] considers a strong assumptions on the number of susceptible individuals by approximating it as a constant factor of the entire population. This assumption may not be accurate when the number of infected individuals are high compared to the total population of a community.

Another recent work has presented a Bayesian prediction approach to obtain confidence intervals for R_t [6]. However, the underlying idea of [6] is based on [7] where the definition of infection rate R_t is not based on a time-variant contact rate of the SIR model. In other words, Bettencourt *et al.* consider a

fixed contact rate β which is not changing over time. As we know, this assumption may not be the case due to the dependency of interaction rate on time. Furthermore, [7] pursues a probabilistic approach of estimating infection rate from number of new cases rather than utilizing the inherent partial equations of the SIR model.

The main contribution of this work is to show how to obtain the daily effective reproduction number R_t with a time-varying SIR model as well as the corresponding Confidence Interval (CI). The confidence interval reflects uncertainty in both the parameter of the underlying model, and uncertainty in the data itself. We present an application of our approach to the daily COVID-19 confirmed case data released by the Los Angeles (LA) county public health department.

II. Mathematical Analysis of R_t

In this section, we elaborate upon the SIR model in detail. The SIR model is one of the simplest and the most well-known epidemic model [8] where each person belongs to one of the following three states: the susceptible state, the infected state, and the recovered state. Regarding the susceptible state, individuals have not had the virus yet. However, they may get infected in case of being exposed to an infected individual. As far as the infected state is concerned, a susceptible person have the virus after being exposed to infected individuals. Finally, a person enters the recovered state in case of either the individual get healed or being dead. One important point about this model is that a recovered person will not be a susceptible one anymore.

The SIR model follows the following differential equations:

$$\begin{aligned} \frac{dS(t)}{dt} &= -\beta \frac{S(t)I(t)}{N} \\ \frac{dI(t)}{dt} &= \beta \frac{S(t)I(t)}{N} - \sigma I(t) \\ \frac{dR(t)}{dt} &= \sigma I(t) \end{aligned} \tag{1}$$

where $S(t)$, $I(t)$, and $R(t)$ respectively represent the number of susceptible, infected, and recovered people in a population size of N at time t. Regarding the parameter σ, it is the recovery rate after being infected and is equal to $\frac{1}{D_I}$ where D_I represents the average infectious days. Parameter β is known as the effective contact rate, i.e. the average number of contacts an individual have with others is β.

To monitor whether the pandemic has been contained, it is extremely crucial to obtain parameter β for daily bases. In other words, it is essential to consider a time-dependant SIR model where contact rate is a function of time, i.e. β_t.

We next show that how we can derive β_t from the the differential equations of the time-dependant SIR model.

A. Obtaining β_t for the time-dependent SIR Model

In the time-dependent SIR model, we can express the number of susceptible individuals in terms of population size and the number of infected persons as $S(t) \approx N - I(t)$. By replacing $S(t)$ with $N - I(t)$ in the second differential equation of (1), we would have

$$\frac{dI(t)}{dt} = \beta_t \frac{(N - I(t))I(t)}{N} - \sigma I(t). \tag{2}$$

We can rewrite (2) as follows:

$$\frac{dI(t)}{(\beta_t - \sigma)I(t) - \frac{\beta_t}{N} I^2(t)} = dt. \tag{3}$$

By taking definite integral from time t_1 to t_2 and assuming β_t to be constant in this time interval, we would have

$$\int_{t_1}^{t_2} \frac{dI(t)}{(\beta_t - \sigma)I(t) - \frac{\beta_t}{N} I^2(t)} = \int_{t_1}^{t_2} dt \tag{4}$$

which leads to

$$\frac{1}{\beta_t - \sigma}\left(\log \frac{I(t_2)}{\beta_t - \sigma - \frac{\beta}{N} I(t_2)} - \log \frac{I(t_1)}{\beta_t - \sigma - \frac{\beta}{N} I(t_1)}\right) = t_2 - t_1 \tag{5}$$

One can easily check (5) has a unique solution for β_t due to the fact that term $\frac{1}{\beta_t - \sigma}$ and log term have monotonic behaviors.

An epidemic happens in case of increase in the number of infected individuals, i.e. $\frac{dI(t)}{dt} > 0$, or consequently

$$\beta_t \frac{(N - I(t))I(t)}{N} - \sigma I(t) > 0. \tag{6}$$

In the early stage of an epidemic, almost everyone are susceptible except very few cases. Therefore, $N - I(t) \approx N$ and as a result, condition (6) would turn into

$$\frac{\beta_t}{\sigma} > 1. \tag{7}$$

$R_t \triangleq \frac{\beta_t}{\sigma}$ is defined as the *effective reproduction number*. It is a useful metric to determine epidemic growth. In case of having $R_t > 1$, the epidemic is growing exponentially while $R_t < 1$ indicates the epidemic is contained and will decline and die out eventually.

For discrete-time cases such as daily reporting on number of infected cases, the time-variant effective contact rate β_t, which represents the contact rate for time slot t can be derived by solving the following equation:

$$\frac{1}{\beta_t - \sigma}\left(\log \frac{I(t+1)}{\beta_t - \sigma - \frac{\beta_t}{N} I(t+1)} - \log \frac{I(t)}{\beta_t - \sigma - \frac{\beta_t}{N} I(t)}\right) = 1 \ \forall t. \tag{8}$$

Since it is difficult to write a closed form solution for β_t in (8), we take a simpler approximation to β_t by considering the following which is based on (2)

$$\beta_t \approx \frac{\sigma I(t) + (I(t+1) - I(t))}{(1 - \frac{I(t)}{N})I(t)}. \tag{9}$$

Since there is uncertainty about parameter D_I (or equivalently σ) and number of confirmed cases $I(t)$, we now provide the derivation of confidence interval for parameter R_t. To do so, we need to first find the marginal distribution of R_t. By considering $f_D(d)$ and $f_K(k)$ as the probability distribution function (pdf) for parameters D and k, respectively, the joint pdf of these parameters would be

$$f_{D,K}(d,k) = f_D(d) f_K(k) \tag{10}$$

due to the Independence of D and k. We can derive the probability distribution function of R_t by performing the following transformation on parameters d and k:

$$\begin{aligned} Z &\triangleq K, \\ R_t &= \frac{1}{1 - \frac{K I_{rep}(t)}{N}}\left(1 + D\frac{I_{rep}(t+1) - I_{rep}(t)}{I_{rep}(t)}\right). \end{aligned} \tag{11}$$

Since the transformation of (Z, R_t) to (D, K) is one-to-one, we have

$$\begin{aligned} K &= Z, \\ D &= \frac{R_t(1 - Za_t) - 1}{b_t} \end{aligned} \tag{12}$$

where

$$\begin{aligned} a_t &\triangleq \frac{I_{rep}(t)}{N}, \\ b_t &\triangleq \frac{I_{rep}(t+1) - I_{rep}(t)}{I_{rep}(t)} \end{aligned} \tag{13}$$

and the joint pdf of Z and R_t can be found as follows

$$f_{Z,R_t}(z, r) = |J| f_{D,K}(d, k) \tag{14}$$

where

$$J \triangleq \begin{bmatrix} \frac{\partial d}{\partial z} & \frac{\partial d}{\partial r} \\ \frac{\partial k}{\partial z} & \frac{\partial k}{\partial r} \end{bmatrix}. \tag{15}$$

By substituting the corresponding values of parameters and the Jacobin, (14) can be written as

$$f_{Z,R_t}(z, r) = |\frac{1 - za_t}{b_t}| f_D(\frac{r(1 - za_t) - 1}{b_t}) f_K(z). \tag{16}$$

The marginal pdf of R_t can be obtained by taking integral of (16) over parameter z, i.e.

$$\begin{aligned} f_{R_t}(r) &= \int f_{Z,R_t}(z, r) dz \\ &= \int |\frac{1 - za_t}{b_t}| f_D(\frac{r(1 - za_t) - 1}{b_t}) f_K(z) dz. \end{aligned} \tag{17}$$

B. **Confidence Interval for** R_t

We can present the uncertainty about the actual number of infected cases as a factor of reported ones, i.e. $I_{rep}(t) \triangleq \frac{1}{k} I(t)$, and k is a constant greater than 1. The main intuition behind this factor is due to take into account the following two phenomena, namely lack of sufficient number of tests (specially in the beginning of the pandemic) and asymptomatic cases (mild infections which might not even be noticed). Therefore, (9) can be rewritten as

$$\beta_t \approx \frac{1}{1 - k\frac{I_{rep}(t)}{N}} \sigma + \frac{(I_{rep}(t+1) - I_{rep}(t))}{(1 - k\frac{I_{rep}(t)}{N}) I_{rep}(t)}. \tag{18}$$

Based on the proof presented in Appendix, the marginal probability distribution function (pdf) of parameter R_t is as follows:

$$f_{R_t}(r) = \int |\frac{1 - za_t}{b_t}| f_D(\frac{r(1 - za_t) - 1}{b_t}) f_K(z) dz. \tag{19}$$

Remark 1: one reasonable assumption regarding the pdf of parameters D and K is that both of them have Gaussian distributions. By considering $D \sim \mathcal{N}(\mu_D, \sigma_D^2)$ and $K \sim \mathcal{N}(\mu_K, \sigma_K^2)$, (19) can be simplified as

$$\begin{aligned} f_{R_t}(r) &= \int_{-\infty}^{\frac{1}{a_t}} \frac{1 - za_t}{b_t} f_D(\frac{r(1 - za_t) - 1}{b_t}) f_K(z) + \int_{\frac{1}{a_t}}^{\infty} (-1) \frac{1 - za_t}{b_t} f_D(\frac{r(1 - za_t) - 1}{b_t}) f_K(z) \\ &= \int_{-\infty}^{\frac{1}{a_t}} (\beta_0 + \beta_1 z) C \sqrt{2\pi\sigma_c^2} \phi_{\mu_c, \sigma_c^2}(z) dz + \int_{\frac{1}{a_t}}^{\infty} (-\beta_0 - \beta_1 z) C \sqrt{2\pi\sigma_c^2} \phi_{\mu_c, \sigma_c^2}(z) dz, \end{aligned} \tag{20}$$

where $\phi_{\mu_c,\sigma_c^2}(.)$ indicates the PDF of a normal distribution with mean μ_c and variance σ_c^2 while

$$\beta_0 \triangleq \frac{1}{b_t} \quad , \quad \beta_1 \triangleq \frac{-a_t}{b_t},$$
$$\alpha_0 \triangleq \frac{(\frac{r-1}{b_t}-\mu_D)^2}{2\sigma_D^2}+\frac{\mu_K^2}{2\sigma_K^2} \quad , \quad \alpha_0 \triangleq \frac{(-\frac{ra_t}{b_t})(\frac{r-1}{b_t}-\mu_D)}{\sigma_D^2}-\frac{\mu_K}{\sigma_K^2} \quad , \quad \alpha_2 \triangleq \frac{(\frac{ra_t}{b_t})^2}{2\sigma_D^2}+\frac{1}{2\sigma_K^2}, \tag{21}$$
$$\mu_c \triangleq \frac{-\alpha_1}{2\alpha_2} \quad , \quad \sigma_c^2 \triangleq \frac{1}{2\alpha_2} \quad , \quad C \triangleq \frac{e^{-(\alpha_0-\frac{\alpha_1}{4\alpha_2})}}{2\pi\sigma_D\sigma_K}.$$

By taking integral through using change of parameters, (20) can be rewritten as follows

$$f_{R_t}(r) = -2C\beta_1\sigma_c^2 e^{-\frac{(\frac{1}{a_t}-\mu_c)^2}{2\sigma_c^2}} + C\sqrt{2\pi\sigma_c^2}(\beta_1\mu_c+\beta_0)\Phi_{\mu_c,\sigma_c^2}(\frac{1}{a_t}) + C\sqrt{2\pi\sigma_c^2}(-\beta_1\mu_c-\beta_0)(1-\Phi_{\mu_c,\sigma_c^2}(\frac{1}{a_t})) \tag{22}$$

where $\Phi_{\mu_c,\sigma_c^2}(.)$ represents the Cumulative Distribution Function (CDF) of a normal distribution with mean μ_c and variance σ_c^2.

The confidence interval would belong to $(\bar{R}_t-\delta, \bar{R}_t+\delta)$ where

$$\bar{R}_t \triangleq \mathbb{E}[R_t] = \int r f_{R_t}(r) dr. \tag{23}$$

and δ can be derived by satisfying the following

$$\mathbb{P}(|R_t-\bar{R}_t| \leq \delta) = \int_{\bar{R}_t-\delta}^{\bar{R}_t+\delta} f_{R_t}(x)dx = 1-\epsilon \tag{24}$$

for some small $\epsilon > 0$.

REFERENCES

[1] A. Tobías, "Evaluation of the lockdowns for the sars-cov-2 epidemic in italy and spain after one month follow up," *Science of The Total Environment*, p. 138539, 2020.

[2] K. Kupferschmidt, "The lockdowns worked—but what comes next?" 2020.

[3] Y. Liu, A. A. Gayle, A. Wilder-Smith, and J. Rocklöv, "The reproductive number of covid-19 is higher compared to sars coronavirus," *Journal of travel medicine*, 2020.

[4] W. H. O. (WHO), "Coronavirus disease (covid-19) outbreak," 2020. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019

[5] Y. chun Chen, P.-E. Lu, and C.-S. Chang, "A time-dependent sir model for covid-19," *ArXiv*, vol. abs/2003.00122, 2020.

[6] K. Systrom, "The metric we need to manage covid-19," 2020. [Online]. Available: http://systrom.com/blog/the-metric-we-need-to-manage-covid-19/

[7] L. Bettencourt and R. Ribeiro, "Real time bayesian estimation of the epidemic potential of emerging infectious diseases," *PloS one*, vol. 3, p. e2185, 02 2008.

[8] M. Newman, "Networks: An introduction," in *Oxford University Press*, 2010.