

## STAT440 Final Project Proposal - Group 42

Coronavirus Disease 2019 (COVID-19) is a respiratory illness that can spread from person to person. The novel coronavirus was first identified during an investigation into an outbreak in Wuhan, China. As of March 3, 2020, the mortality rate estimated by the World Health Organization (WHO) is 3.4%. Up to March 12, there were 125,048 confirmed cases worldwide and 80,981 confirmed cases in China. Given the severity of this epidemic, it is crucial to estimate the theoretical number of people infected with COVID-19 in a closed population over time.

SIR model and its variations have been used to predict the progression of a virus. In this project, we will predict the progression of COVID-19 from deterministic SIR model, stochastic SIR model and SEIR model and analyze their advantages and disadvantages.

**Deterministic SIR** is a commonly used compartmental model in epidemiology, which computes the change in the population of each of the three compartments (susceptible, infected, removed) in terms of two parameters, effective contact rate ( $\beta$ ) and removed rate ( $\gamma$ ). Although deterministic SIR provides the basis for studying and tracking an epidemic, it assumes that individual characteristics of immunity, susceptibility, and ability to recover, are essentially the same for all members of the population. Moreover, it excludes the unconfirmed infected cases and fails to distinguish the recovered and dead individuals. Last but not least, when the reproductive number ( $R_0$ ) is greater than 1, the model converges to an endemic equilibrium that violates the assumption that the disease will eventually disappear.

To make the model more realistic, we propose two variations of the deterministic model introduced above:

1. An SEIR model that distinguishes between the recovery and the death individuals, as well as factoring in unconfirmed cases through adding additional measurable variables
2. A Stochastic SIR model that generates measurable variables that follows a linear birth and death process (CTMC)

These modifications allow the new model to compute a more precise recovery rate and fatal rate respectively.

The **SEIR model** proposed by S. Cao et al. keeps susceptible group ( $S$ ) and recovered ( $R$ ) as the original model, but it and added new factors such as hospitalized ( $H$ ) and exposed ( $E$ ). The population is further stratified to include quarantined susceptible ( $S_q$ ), isolated exposed ( $E_q$ ) and isolated infected ( $I_q$ ) compartments. The SEIR model is undoubtedly a more precise depiction of the situation in mainland China from January to March, 2020; however, it may not be generalizable for other countries.

The SIR model implies that the changes in infected, susceptible or recovered are continuous. In reality, however, the changes are discrete. We then propose the **stochastic SIR** model to reveal this state transition feature. We will use Gillespie's direct method to simulate data.

We'd like to note that the solutions from stochastic SIR must converge to the disease-free state, though the time until disease extinction may be very long in some cases, which satisfies the assumption. Prior to the absorption, the process appears to follow quasi-stationary probability distribution. The stochastic SIR model has managed to compute the probability of epidemic outbreak fairly accurately. Compared to the SEIR model introduced above, the stochastic SIR model can be a more general model.

After the COVID-19 outbreak took place in China, the number of the susceptible, infected and death cases has been recorded relatively accurately. We will compare the simulation with the actual cases to evaluate how well the model performs in mainland China versus the rest of the world.