Chapter 2
An Exploratory Analysis and Predictive SIR Model for the Early Onset of COVID-19 in Tamil Nadu, India

Chandan Tanvi Mandapati
Lovely Professional University, India

ABSTRACT

The growth of COVID-19 (SARS-CoV-2) in India has been rampant. Despite having a relatively small value of R_0, the spread of disease increases exponentially every consecutive day. This chapter aims to analyze and conduct a concise study for the southern state of Tamil Nadu in India and build non-linear predictive models that evaluate the transmission of coronavirus amongst locals. A logistic regression and SIR model are deployed to understand the potential spread of disease. Through descriptive analysis on theoretical segmented portions, districts in Tamil Nadu with a higher number of confirmed cases are identified. Computation of crude mortality rate, infection fatality rate, predictive models, illustrations, and their results are discussed analytically.

INTRODUCTION

The earliest cases of Covid-19 surfaced in January 2020, which posed to be a dire threat to India; exposure to an Influenza-like disease could prove to be fatal to a country inhabited by 1.3 billion. Predicting the probable transmission of an airborne pathogen (like SARS-CoV-2) is a laborious but necessary task in excessively populated countries, and failing to anticipate the contagion of diseases like Covid-19 could cause epidemic-control bodies, such as Epidemic Intelligence Service (EIS) of India, to harbour a handicap; in turn, leading to inadequacies and risking the lives of millions.

Epidemic models, for contagious and non-contagious diseases, have been employed to predict and analyse the spread of infections to curb the escalation of disease among humans since the dawn of mathematical epidemic models. An epidemic model that relies on documented data might not consider

DOI: 10.4018/978-1-7998-7188-0.ch002

plausible interactions an infected person has had. In cases of a newly-emerged disease (like Covid-19), data might be scarce and of smaller size; a dataset of such population is insufficient for a machine-learning algorithm to grasp patterns from. Parameters that incorporate transmissibility (i.e., rate at which the infection is spreading) are principal aspects for building an appropriate model which considers probable secondary transmitted cases. The primary goals of this paper are defined below.

1. Build a predictive model that works with deterministic parameters (factors that are estimable through mathematical, i.e., differential methods) for Tamil Nadu (from January 2020 to August 2020)
2. Analyse the spread of Covid-19 in Tamil Nadu
3. Obtain an estimate of the Crude Mortality Rate and Infection Fatality Rate.

For objective (a), a SIR (Susceptible-Infected-Recovered) model, built using Covid-19 data from Tamil Nadu (an Indian state whose confirmed reports were higher than the other states), exemplifies the use of a predictive model using differential processes. Non-linear mathematical epidemiological models are popularly utilized for diseased environments in various situations using simulations (such as Monte Carlo Simulation) for foresights on the spread of disease. The SIR model is a mathematical epidemic model that works with constants such as population, the number of confirmed, recovered cases (deceased are clubbed with recovered, as they no longer belong to the susceptible pool), and principal parameters – R_0 and gamma. It is a theoretical model in nature, and unlike most machine learning models, the SIR model doesn't rely on past data. Additionally, a logistic regression model is deployed between confirmed and recovered cases, where confirmed cases are the independent (predictor) variables and recovered cases are dependent (predicted) variables.

(b) To examine the outspread of disease in Tamil Nadu, the state has been segmented into four clusters and visualized through facet grid; the grids revealed that the majority of confirmed cases arose from the cluster of districts situated in the northern part of the state. The final objective calls for estimation of crude mortality rate and infection fatality rate till the mid of August 2020 in Tamil Nadu, which is computed through Python. This paper examines the contagion of Covid-19 through visual analysis and predictive modelling, which helps in understanding the spread of this disease, and could aid in the prevention of further contamination.

BACKGROUND

The Basic SIR Model

The SIR model (Kermack & McKendrick, 1927) predicts the rise/decline in diseased cases over a specified period, utilizing a function of constants and parameters. Differential equations are integral tools for implementing the model; they instigate changes in susceptible and diseased pools with respect to time over a specified period. The model presumes that at a given time *t*, an individual of a population exists in one of the three states – susceptible state, infected state, or recovered state. The sum of susceptible, infected, and the recovered populace is assumed to be equal to the entire population under study, *N*.

The 'Susceptible' populace is at risk of either contracting the virus or being asymptotic carriers. A portion of the susceptible population that becomes infected with the disease advances to the 'Infected' compartment. The infected from the prevalence pool either recover with immunity or meet with un-