Stockholm University

Mathematical Statistics
Stockholm University
Bachelor Thesis **2021:20**
http://www.math.su.se

Investigating nowcasting of COVID-19 infected using linear regression

Lukas Fredriksson*

September 2021

Abstract

This thesis aims to investigate the possibility to perform nowcasting of the number individuals infectious with COVID-19. In order to do this a discrete time SIR (Susceptible, Infectious, Removed) with the added Diagnosed state is used to simulate epidemics. This model has three parameters, q which denotes the probability that a susceptible individual becomes infected by a given infectious individual, r which denotes the probability that an infectious individual stops being infectious and p which denotes the probability that an individual who stops being infectious become diagnosed. Four different sets of parameters where q, r and p either are fixed or uniformly randomised are used to simulate four sets of training data. These sets of data are then used to train a linear regression model where the inputs are the number of new daily diagnosed individuals of current time step and four steps before. Only diagnosed individuals are used in the linear regression because that is what we observe in reality.

The theory behind the model used for simulation show that the parameter p is not observed in the available data, while q and r are. This suggests that the cases where p is randomised during simulation should provide less accurate nowcasting, which is also suggested in linear regression. The regression model does however violate some of the assumptions made during linear regression. For this reason, further studies of nowcasting should investigate other regression models.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: lrfson@gmail.com. Supervisor: Pieter Trapman.