Sir Models and Ensemble-type Algorithms with Applications to COVID-19

Jaciele Oliveira[1], *Raydonal Ospina*[1,2], *Cristiano Ferraz*[1]
[1] Universidade Federal de Pernambuco
[2] Universidade Federal da Bahia

Abbreviated abstract: In January 2020, the world was surprised by the COVID-19 pandemic, making it crucial to use statistical and mathematical tools to describe the course of the epidemic. In this work, we used the SIR compartmentalized model on COVID-19 data from the state of Paraíba and compared it with prevalence results estimated by a serological sample survey. We also use an ensemble approach, which integrates the prediction of several models through a weighted combination.

Related publications: (up to 2 references)

– Anastassopoulou ,C. et al. PLOS ONE 2(2) 2020.
– Silva, V. S. et al, Science and Knowledge in Focus 3 (2), 2020

Universidade Federal de Pernambuco

4th Conference on Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022

Introduction

- In January 2020 the world was surprised by the epidemic of COVID-19, caused by the SARS-CoV-2 virus, a virus with a high degree of lethality and transmissibility.
- In this context, the concern arose that a large part of the population would be infected simultaneously and promote the overloading of the public health system and an increase in the mortality rate from the infection.
- Since it is a new virus, it has become essential to use tools to describe the course of the epidemic, in order to evaluate the effect of restrictive measures, make possible forecasts of scenarios for the spread of the virus, and help governments to implement effective measures against COVID-19.
- The paper aimed to evaluate the infection rates of COVID-19 using compartmentalized SIR model and ensemble model. In addition, we compared results obtained with data from a serological sampling survey conducted in the state of Paraíba, localized in Brazil.

SIR and Ensemble Models

- The SIR model considers a constant population subdivided into three compartments: susceptible, S(t); infected, I(t), and removed, R(t). The goal is to describe the transition between the compartments in the model. The definition of the model is given by the following system of differential equations:

$$\frac{dS}{dt} = -B\frac{I}{N}\mathrm{S}$$

$$\frac{dI}{dt} = B\frac{I}{N}\mathrm{S} - \alpha\mathrm{I}$$

$$\frac{dR}{dt} = \alpha\mathrm{I}$$

- This ensemble method is based on a weighted combination of individual models. Consider I parametric models. From the training data the parameter set and the average incident curve for the i -th model are estimated, i=1,...,I. Based on the quality of the fit of each model, the weight W_i for the i-th model is calculated. Then the ensemble incident curve is given by:

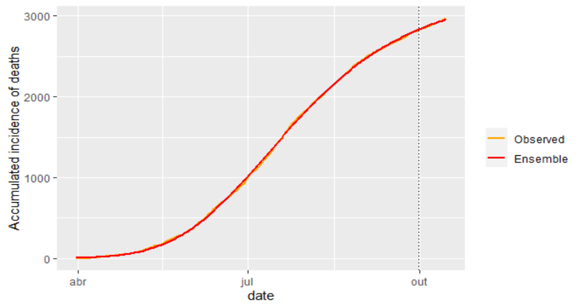
$$f_{ens}(t) = \sum_{i=1}^{I} w_i f_i(t, \hat{\theta})$$

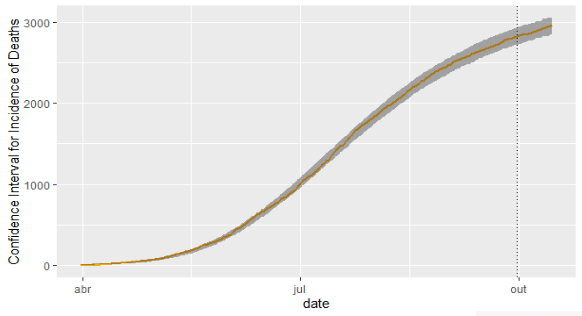
Results and Conclusions

- The **SIR model** was adjusted for nine subdivided periods of the Paraíba COVID-19 dataset, referring to the year 2020. The infection rates estimated by the model were higher in the first two periods, starting to fall thereafter, and rising again at the end of 2020.

- When comparing the results generated by the model with the results obtained by the sample survey carried out in Paraíba, the survey estimates are much higher than the adjusted quantities, since the model is adjusted based on underreported data.

- As an alternative to the compartmentalized model we use the **ensemble model.** Considering that underreporting in counting the number of deaths from COVID-19 is lower than underreporting the number of cases, we used the cumulative curves of deaths in the ensemble fit. The ensemble model fit the data well and generated excellent predictions for up to 30 days ahead as shown in the graphs.





Universidade Federal de Pernambuco

4th Conference on Statistics and Data Science