Change-point analysis to quantify the impact of African government policy interventions to slow the spread of COVID-19: A case study of Senegalese Government COVID-19 policies

The outbreak of the coronavirus 2019 disease (Covid-19) has led to a drastic change in the affairs of men, old forgotten terms like social distancing, use of face mask, ban of large gathering e.t.c. In this case study, we would be conducting a change point analysis to quantify the impact of African government policy interventions to slow the spread of Covid-19 with a keen interest in Senegal and her government. Before we dive into the analysis we will need to familiarize ourselves with some terms which are

- Explain in your words the purpose of the SIR/SEIR model in the current project?
  The SIR/SEIR is a model that is specifically targeted at diseases & epidemics, it helps to know the numbers of people who are susceptible(liable to contact) a disease/epidemic, numbers of people who are infected by particular diseases and the numbers of people who have either died or recovered from a particular disease. Also the SIR/SEIR (susceptible-Infected-Removed or Recovery/ Susceptible-Exposed-Infected-Recovery) model helps scientists predict if a particular pandemic will be full-blown or not. This model is very important to our particular challenge in that, the model can help us to know those who are susceptible to Covid-19, the number of people who are infected with Covid-19 and numbers that have either died or have recovered, with this we can check the numbers before the measures that were taken by the Senegalese government and how effective it is based on the condition in which the spread of the diseases is reduced or not.
- We will use the COVID19 cases data for this challenge, comment on the output of a SIR/SEIR model in relation to this data.
  The outputs of the SIR/SEIR model in relation to this data are the contact ratio(the rate at which susceptible people contract the disease), the reproduction number of the disease, the recovery rate and infective rates.
- List the processes that affect the generation of the COVID19 cases data in your country, and which part of this process is modelled by the SIR/SEIR model.
  (1) Through the help of tracking community transmission
  (2) Through contact tracing.
  (3) Testing people and producing testing kits

The part that is modelled by the SIR/SEIR is the I (Infected numbers and this is gotten through testing)

- Explain the difference between the SIR and SEIR model.
  The main difference between the SIR model and SEIR model is that in SIR we do not account for people that were exposed to infected people but in SEIR model we account for

the numbers of exposed. The exposed number can be of help for an example, some persons might be exposed to an infected COVID-19 patient but might not show any symptoms and can likely pass it across to other susceptible people, which to an extent is not fine.

- Explain the distinct characteristics of an exponential function.
  An exponential function is a function that involves a real and positive constant(let us say b) which has another constant raised to its power or exponent. The distinct nature is that the base constant must be a positive real number and it must not be equal to 1. $y=b^x$. The power variable can take any value.
- Explain the similarities and differences of exponential growth, exponential decay, geometric progression, and logistic growth.
  The similarities between exponential growth, exponential decay, geometric progression & logistic growth is that all growth increase with certain function in which their respective functions are raised to a power, the difference between them is that in exponential growth there is a spontaneous increase in growth with time e.g the rate at which bacterials split with time(the rate is really fast and the result is way bigger than the first), in decay the reverse is the case as the population reduces with time an example is the decay of radioactive elements which follows a decay growth, in geometric progression the growth parameter follows a constant, the constant is called the common ratio in which the division of the next growth number divided by the previous gives a constant values for all the points with time, logistic growth is a modified exponential growth in which a carrying capacity is introduced into the exponential growth, the logistic growth follows the assumption that at first the population growth will follow an exponential manner if the population is few and resources are abundant but as the population increases the resources will become small and smaller and it will get to a point when the resources will not be enough and the growth will stop, but if the population figure continues to increase then the population growth will start to decline.
- Explain what approximation leads the SIR/SEIR model to take an exponential form.
  The assumption that makes a SIR/SEIR model to take an exponential form is when an assumption that the number of susceptibles is constant, this can only occur when the time is zero and with integration it will take an exponential growth.
- Explain how the rate parameters in the SIR/SEIR model are estimated from COVID19 cases data. The rate parameters are
  (1) The rate of change of number of susceptible with respect to time is directly population to the number of infected, which gives rise to -(r*I*S), r the rate of contact,I- number of infected, S- Number of susceptibles. It is negative because it leads to a subtraction from the number of susceptibles, this is the number of people that are infected.
  (2) The rate of change of infected numbers with time, this is the positive product of rate of contact of susceptibles with infected people, the number of infected and the number of susceptibles minus the product of the rate of recovery/death and Number of infected people. Mathematically it is rIS - rI

(3) The rate of removal (recovery or death) which is the gain from people that were infected and recovered. Mathematically it means +rI

- Explain which probability distributions will be used to model the SIR/SEIR rate parameters. Why? We will use a normal distribution to model the rate infection spread, the recovery rate and use a logarithm distribution to model our number of infected and the forecast.
- Describe the expected outputs of the modelling phase of this challenge. The number of reproductions, the rate of spread of the infection and the maximum numbers that are infected.
- How are predictions to future dates, for example, one week from the date of the last date of the training data, is done? We can make predictions by using forecast.
- Describe how the effectiveness of government COVID19 non-pharmaceutical interventions policies is evaluated in this challenge. We will use the spreading rate and recovery rate prior to the non-pharmaceutical interventions and after the

1. Scenario: public interest in corruption
    a. **Data**: number of daily tweets containing the word corruption from a given country for a year.
    We will split the groups into
        (1) Those that saw the tweets(susceptibles), those that interacted with the tweets(liked, retweeted) as infected and those that saw but did not interact as recovery
    b. **Business Objective**: disentangle the relative interest of people in corruption issues because of the following triggers:
        i. Reports emerged at the beginning of February, that an important politician was taking bribes.
        People will be more interested since it involves an important figure
        ii. This politician was convicted on the 15th of August. People will be less interested
        iii. The slow rise during this year of every-day bribery (police, clerks, etc.), very passive since it occurs virtually every day.

Task 2

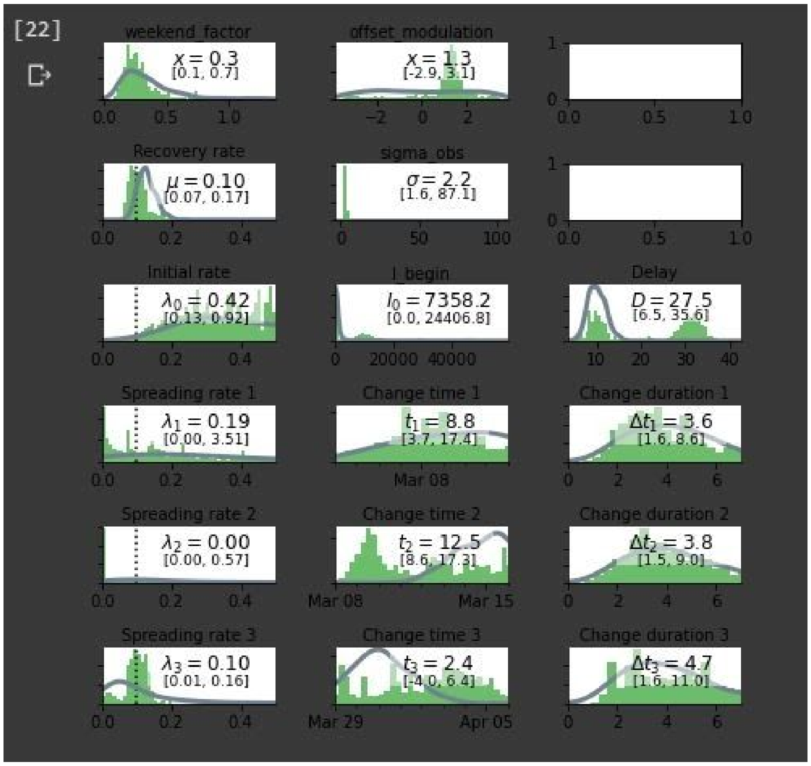
2.1(g)

(1) The banning of major gatherings occurred on the 14th of March 2020
(2) The closure of schools occurred on 14tho of march 2020
(3) Strong social distancing was effected on 31st of March 2020
(4) Mask wearing was made mandatory on the 20th of April 2020

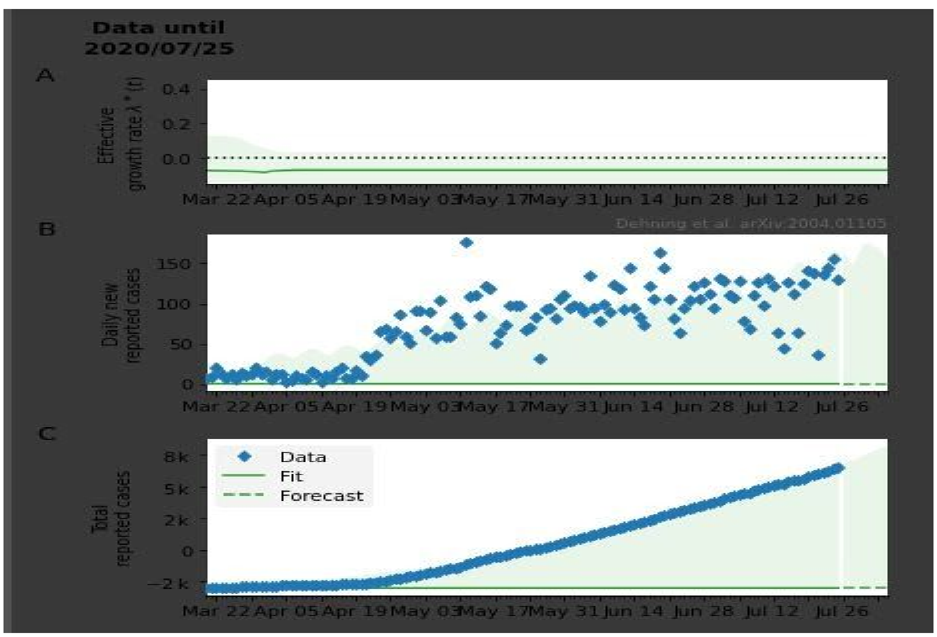
Theses rules are yet to be relaxed.

2.1(h) On 23rd March 2020 state of emergency was introduced to limit the movement of people from 6pm to 8am, after about 207 cases. Although this has been relaxed on 30th of June 2020.

Below are the training data plotted against samples from the posterior SIR model



Fig(1) A plot of the distribution of different parameters in the model with respect to change points



Choice of Model and overview

In other to generate and visualize the rate of change and recovery rate at every change point we are constrained to use the SIR/SEIR model which helps us in getting our objective, we did not opt for linear regression models or clustering models because the linear regression model or logistic regression model we only tell us the number of result for the future but will not provide us with the changes at each change points, clustering will only cluster our algorithm, which is not what we need. This model is built on the naive Bayes theorem in which we are looking at the probability of having values within a particular range, with that we set a prior probability or value for our model based on domain knowledge, the model then calculator the likelihood and get the posterior, after the posterior is gotten, the prior value is updated with the new posterior. This posterior helps the model parameter to be predicted based on given data or unseen data, this is our end goal to check if the model can make good predictions based on new datasets. Also the posterior is sampled using Markov Chain Monte Carlo chain, in which there is a random walk and the next step is based on the previous steps. The model that was used was gotten from the model that was used by Jonas Dehning etal(2020) in writing the research article "*Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions"* , the model is built on the PYMC3 which is a python package for Bayesian statistical modelling & probabilistic machine learning, which calculates probability by taking into considerations the prior and the likelihood. To get our result we factored in the spreading rate, the recovery rate, the change points.

Looking at the distribution of the rate of spread of the disease, we could see that before the first change point, the rate of spread was high, but after the first change point of banning religious activities on the 10th of March 2020, the rate of spread reduced to about 0.19 which is fine, and after the second change point which is the restrictions on travelling out and into the country, the rate of spread reduced to zero, probably because people are not coming from more infected places but after the third change point which is the implementation of the state of emergency which was on the 31st of March 2002, the numbers increased to 0.10 which is still lower than the rate of spread after the first intervention

Conclusion

In conclusion from the results that our model gave and the data, we could see that the 3 different interventions that were implemented by the Senegalese government to curb the coronavirus spread were effective and were important at reducing the spread of the disease.