Modelling - Project

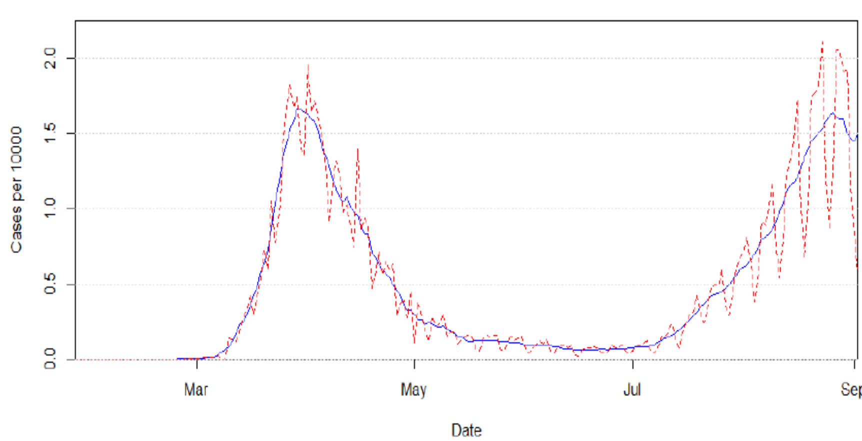
Modelling the influence of vaccinations on the progress of an epidemic using SIR model on a graph.

In this project we want to investigate the impact of vaccination strategies on the spread of a Coronavirus in a population. First we need to build a model of the spread of the virus including measures that are taken to prevent the spread. To achieve this goal we will concentrate on the period between September 2020 and May 2021. We will use a probabilistic model where interactions between individuals in a population will be described using a random graph and the spread of virus via SIR (Susceptible, Infected, Recovered) model associated to the random graph (which is implemented in R under EpiModel package). With the obtained model we will be able to generate data of daily cases and include the impact of vaccinations on the progression of the pandemic. Moreover, having built the model we can study how different policies of e.g. earlier start or larger rate of vaccinations would influence the number of infections.

Background and Data

Coronavirus infections are monitored around the world (in NL by RIVM). Since Sars-Cov-2 is a novel virus the data about the number of infections is unreliable due to e.g. insufficient testing. We also know that there is a percentage of infections that is undetected and does not appear in the number of confirmed cases. Hence we have to rely on mathematical models to monitor the pandemic. We will build a mathematical model and use the available data about the number of infected people in a given time to calibrate parameters of the model. Since the available data is affected by the interventions that have been taken to combat the pandemic (think lock down, wearing masks, etc.) the calibration is quite challenging.

Before the modelling can start we have to preprocess that available data. There are few reasons why this has to be done. First of all, it is known that numbers of new infections reported during weekends were not very accurate. As an example, in the figure below, the number of confirmed cases per 10000 people in Spain from January 26 until September 2, 2020, is shown. The red dotted line shows the number of confirmed cases, which is very variable due to inaccuracy in reporting. One can smooth the curve by taking a seven day rolling window and obtain the blue curve that represents a more realistic progression of the pandemic.



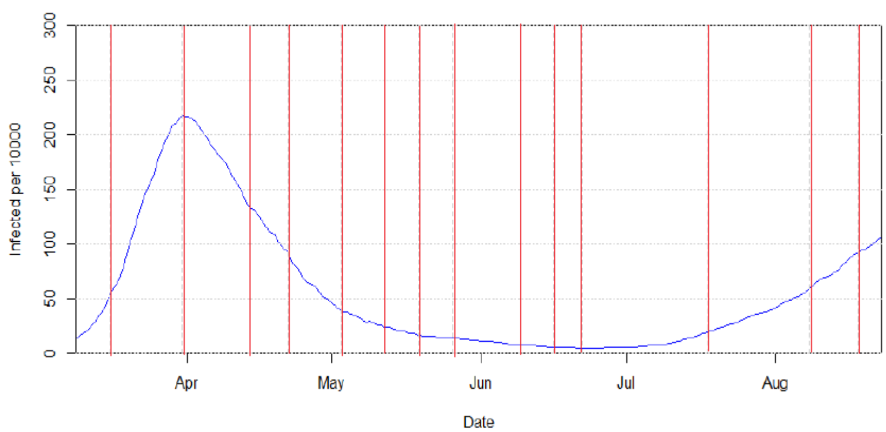
Since we want to know how many infectious people are in the population on each day, the insufficient testing, symptomless cases, duration of the infection as well as time before the infection is confirmed by a positive test have to be included into the data. To take all these issues into account the data of confirmed cases from Spain have been preprocessed by including the following assumptions:

- infection duration is 14 days,
- 6 days from infection to positive test,
- 5% of Spaniards tested between April 27 and May 11, 2020 had antibodies.

Taking all this into account, the data was preprocessed using the following formula for the number of infectious people in day t, I_t:

$$I_t = \sum_{i=t-6}^{i=t+7} f_i C_i.$$

Here C_i denotes the number of confirmed cases on day i (not smoothed) and f_i is a correction factor (to account for unconfirmed cases) which is 10 up to May 15 and then is reduced linearly to 5 on June 23 and stays 5. The choices have to be very well motivated and obviously will differ when the data from other countries and/or another period is modelled. As result the following graph of the number of infected Spaniards per 10000 can be obtained.



In the figure above you can also see the red vertical lines indicating dates where different measures have been taken by the Spanish government to combat the pandemic.

Modelling

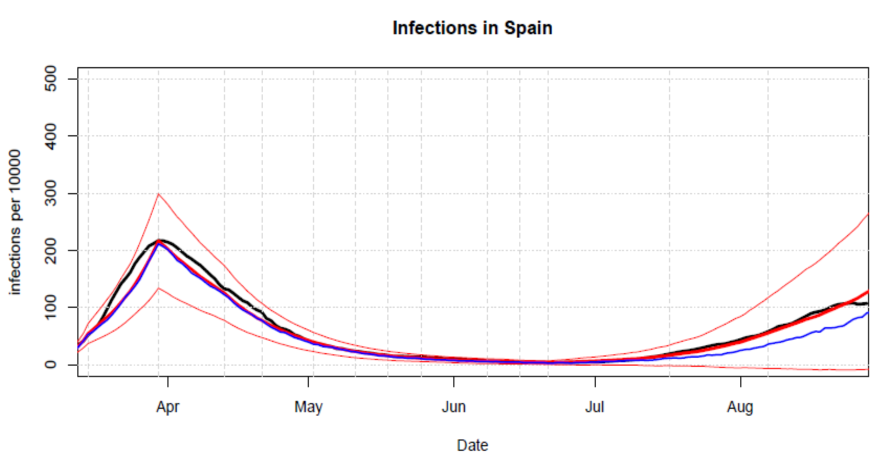
We will use in this project the SIR (Susceptible, Infectious, Recovered) model combined with a random graph. The interactions between individuals in a population will be modeled using a random graph and spread of virus via SIR model. In principle such combined model allows modeling different groups in a population with varying behaviour and allows more realistic results that can be specialized for different areas of a country or different age groups. We will keep our graph model simple (as it is difficult to have realistic numbers about the population dynamics). Moreover, for simplicity we will assume that the number of nodes in the network does not change, hence we do not have births or deaths in the population. Then we will try to calibrate this model (find appropriate parameters) with available information about the observed cases (preprocessed data) and about interventions (shown for the example of Spain in the table below).

Date	Measure
15 Mar	State of alarm, lockdown
30 Mar	Ban on non-essential activities, non-essential workers ordered to stay home
13 Apr	Relaxation of ban on non-essential activities
21 Apr	Children under 14 are allowed to go on short walks with an adult
2 May	Short walks and individual sports allowed, face masks in public transport
11 May	Small shops reopen, terraces to 50% and workplace to one third of capacity
18 May	Further de-escalation
25 May	Half of Spain in phase two
8 June	Half of Spain in phase two and half in phase three
15 June	Three-quarters of the population on phase three
21 June	State of alarm expired, 'new normal' phase entered
17 July	Gatherings of more than ten people forbidden in Catalonia, advice to stay home
7 Aug	Closing of nightclubs, masks in public

In case of modelling the data from Spain the following values of parameters were obtained after an intensive calibration process:

Date	Infection prob	Number of edges	Duration
8 Mar	0.05	40000	10
15 Mar	0.045	22000	15
30 Mar	0.02	10000	30
13 Apr	0.005	15000	30
21 Apr	0.004	15000	30
2 May	0.004	18000	20
11 May	0.006	18000	20
18 May	0.007	20000	20
25 May	0.008	20000	20
8 June	0.009	22000	15
15 June	0.032	25000	15
21 June	0.032	25000	15
17 July	0.032	22000	15
7 Aug	0.03	20000	20

The model with these parameters led to the results shown in the figure below. The black curve represents the preprocessed data of number of cases per 10000 Spaniards and the blue curve is the median (thick red curve is the mean) of 100 simulations. Moreover the two red thin curves represent uncertainty bounds (lower = mean - stdev, upper = mean + stdev). We see that the model performs quite well as the data sits snugly between the uncertainty lines.



Project

The main focus of this project is to investigate the impact of vaccination strategies on the spread of a Coronavirus in a population. We will concentrate on the period from September 2020 until May 2021. Even though vaccination efforts started around December 2020, January 2021 depending on the country, it is necessary to start the calibration of the model earlier to take into account the discovery of a new variant of Coronavirus around November 2020 which was followed by lockdowns. To make this project more interesting we will allow you to choose a country (should be not to large with good data about the confirmed cases, measures and vaccination strategies available). Hence, we propose to pick one of the following countries: Netherlands, Israel, Ireland or Lithuania. The following websites contain useful information: `https://news.google.com/covid19/map?hl=nl&mid=%2Fm%2F059j2&gl=NL&ceid=NL%3Anl`

First preprocess the data to obtain the daily number of infected people. Then calibrate your model on the part of the data between September and December/January (before vaccination). Next, propose a way to include vaccination strategies into the model and then calibrate your model together with vaccinations.

The second part of the project concerns the application of your model. Since we have build a model that describes the development of new cases we can now perform "what if" analysis. At this stage you can show your creativity and propose many scenarios that allow us to see what would have happened if for instance the vaccinations were delayed, were introduced earlier or at a different rate.

Schedule of work

Week 1: Prepare a plan of group work. Familiarize yourself with the literature about random graphs and SIR model (the most important literature is available on BrightSpace). Install the EpiModel package and prepare a few figures of a random network of, say 10 nodes, with some chosen values of parameters for number of connections, duration of edges, probabilities of formation of new edges, dissolutions of edges etc... Choose which country you want to work with and collect the necessary data for the chosen country (number of confirmed cases, measures taken by the government, vaccination strategy).

Week 2-4: The goal in these weeks is to get experience with EpiModel package and theory behind it. Preprocess the data about the number of infections and then calibrate the model. You should calibrate the required parameters in the model to be able to generate the evolution of the number of infections similar to the preprocessed data. Propose a method to include the impact of vaccinations.

Week 5: Deliver the first version of the report.
The report has to contain theory required to understand how the EpiModel is built, what are its parameters and inputs of this model. Explain how are the parameters estimated from the inputs. This does not mean that you need to rewrite the theory included in the literature. You have to show that you understand concepts and main issues of estimation. Include formulas that are necessary to explain the topics. Include figures of evolution of network as well as evolution of the number of infected people for different choices of parameters. Present how you have preprocessed the data and how you have calibrated the model.

Week 6-7: Include the impact of vaccinations into the developed model. Consider what would have happened if the vaccinations started earlier or later (obviously we do not know if the measures that governments taken would have been the same if different vaccination strategy was present).

As minimum you should compare one extra scenario. Be creative and do more to improve your grade.

Final report and Movie

Deliverables in this project are:

- Final report (about 20-25 pages). It should be clearly written and does not need to contain the copy of the computer code.
- Computer code that you have used to generate figures and calculations in the report.
- Movie presenting what you have done in this project (max 5 minutes). Be creative. We will judge the movies based on: a) content b) clarity c) entertainment and creativity.