Assignment 1

Alberto Allegri, Anna Vastola

Abstract

The spread of the Covid-19 disease was first reported by the Chinese authorities to the World Health Organization on December 31st, 2019 (WHO Timeline – Covid 19). Since then, the world economy has been struggling and, due to difficulties in the vaccines supply, it seems it will not recover anytime soon. For these reasons, it is paramount to understand the behavior of the epidemics, to gain a deep knowledge about duration, spreading and possibly make forecasts, in order to help governments in their actions. In this work, we firstly used a SEIR model to understand the behavior of the epidemics in its first 65/70 days. We then focused on one city (Como), and adapted an SIR model to investigate its trend among age groups. Lastly, we improved our model with possible governmental policies: introduction of masks, the closing of schools and lockdown.

Introduction

Since 11 March 2020, when World Health Organization declared COVID-19 a pandemic, every aspect of life as we know it has been impacted profoundly: schools were closed, curfew was introduced, people were forced to stay at home and respect social distancing. On 16 March 2021, a year later, U.S.A. counts nearly 30 million of cases, with more than half a million deaths, but also smaller nations like Italy, France and England account for more than 3/4 millions of cases each. Governments have been struggling to find a trade-off between shutting down everything for safety reasons, and risking people's lives for the sake of the economy. At first, the majority of world leaders decided to go for the first option: in January 2020, China had already quarantined many cities and shut down a large part of the country. After a few weeks, those measures were implemented in every angle of the world: U.S., but also European and Eastern countries strongly reduced contacts and implemented social distancing. The only exception was the U.K., which delayed the restrictions and after just a couple of months saw its health system collapse.

For these reasons, in the last year epidemiology has become very important, both for political aspects and for medical ones: understanding the behavior of the disease is fundamental to save both people's lives and the economy.

In this study, we propose two of the most common, but effective, models employed by epidemiologists: a S-E-I-R model in the first part, to understand the epidemic trend in 12 cities of Lombardy, and an S-I-R model for the city of Como in the second and third part, to compare the spread of the virus among different age classes. We also evaluated how different policies such as the introduction of masks, the closing of schools and a lockdown, would impact our model.

Data

We used a dataset from the 12 provinces of Lombardy, starting from March 2020, for the duration of 92 days. The dataset included the number of daily cases related to COVID-19. For the second part, we referred to a contact matrix for 16 age-ranges, from 0-4 to 75+, going by steps of 5 years. Moreover, we collected data about population in Como on the Istat website and grouped it by age, to match the steps of the contact matrix. We used Python to deal with the differential equations, the basis of epidemiological models.

SEIR model

Our task consisted in developing a deterministic compartmental model that could simulate the spread of COVID-19 in 12

Provinces of Lombardy. We used a different range of days for every city to capture the exponential growth of the pandemic with the lowest Mean Squared Error. The number of infected oscillated from a few cases to hundreds of cases per day in just 20/30 days, despite government restrictions. It is easy to conclude that the growth would have been much more consistent without the limitations.

For the choice of the model, we relied on previous literature and on experts' knowledge about COVID spread. In particular, in our model, we considered that for a certain number of days, the patient could be infected but not infectious, meaning that he/she was not able to transmit it to someone else. This time-range can be variable and can depend on the viral load of the person, so it is impossible to assert a unique measure to this "latent period". Based on previous literature published on the Annuals of Internal Medicine, we decided to set this latent period to 5 days (Lauer, Stephen A., et al., 2020). The model that we thought could best replicate COVID behavior was a S-E-I-R (Susceptibles – Exposed – Infected – Recovered) model, since it comprehends also a compartment that suited well for people that are incubating the disease. We made 3 assumptions: first, for the sake of simplicity, we set μ (natality rate) equal to ν (mortality rate), equal to 0.001, taking into account ISTAT data for 2019, so before COVID-19. The model could be improved by changing the rate at which both Exposed, but especially Infected, are dying, since COVID had an higher mortality rate than normality at the time of our data (Patrick W Brady et al., 2021). As we mentioned before, we set σ, the parameter related to the people moving from compartment E to compartment I, to $\frac{1}{5}$, as it denoted the reciprocate of the average latent period: $\sigma = \frac{1}{(average\ latent\ period)}$.

Finally, we made an assumption also on the recovery rate γ, so the rate at which Infected become Recovered. This coefficient was defined as $\frac{1}{average\ duration\ of\ infectious\ period}$ and we set it to $\frac{1}{5}$, resulting in a final value of 0.2 (Byrne et al., 2020). Also in this case, our choice was motivated by previous literature.

For these reasons, we focused on calibrating the most important value of this equation: β. It expressed the rate of moving from compartment S to compartment E. We implemented an SEIR using these 4 differential equations:

$$\frac{dS}{dt} = \mu N - \nu S - \frac{\beta SI}{N}$$

$$\frac{dE}{dt} = \frac{\beta SI}{N} - \nu E - \sigma E$$

$$\frac{dI}{dt} = \sigma E - \gamma I - \nu I$$

$$\frac{dR}{dt} = \gamma I - \nu R$$

To estimate our parameter, we wrote a function on Python, in order to define β as a function of R0:

$$\beta = \frac{R0}{average\ duration\ of\ infectious\ period}$$

Basically, since γ was given, by testing values of R0 from 1 to 3, we were able to calibrate the best β that minimized the Mean Squared Error with respect to the true data that we had. Results were obviously biased from the fact that we trained our model just on the growing part of the curve for every city. This is the reason why the rest of our simulated curve does not resemble the pattern that we observe in the real data, so our model is fitting quite well the first 60/70 observations, but it is not able to capture the decreasing part of the curve.

The equation we used to compute the MSE was:

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{I}_i - I_i\right)^2}$$

where Î represent the simulated data and I the real data. 12 different βs were found for the different cities, showing that the highest β, so the highest R0s, were related to Milano (R0: 2.53, MSE: 14.96) and Bergamo (R0: 2.64, MSE: 22.83), while the other cities had lower R0s and better MSEs (around 3 or 4).

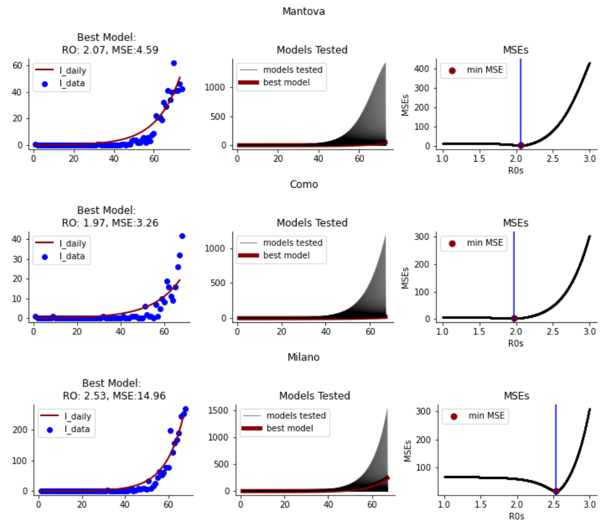

Figure 1: Best Models for Mantova, Como and Milano.

To further validate our results, we wrote another function, in order to exploit Metropolis Hastings power: we computed our function three times, setting the starting β to 0.2, 0.3 and 0.4. Results confirmed our prediction done with MSE, since the mean was always very similar to the β computed before.

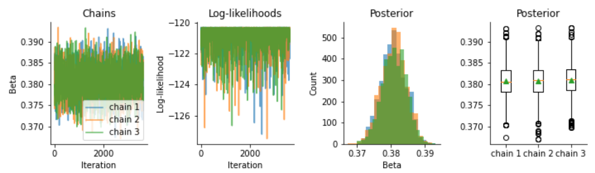

Figure 2: Metropolis Hastings plots for Como.

Even this model was trained just on the growing part of the graph, so for this reason all the models that we calibrated were not able to resemble all the original pattern of the curve.

SIR model

In the second part, we developed an S-I-R model for one of the cities that we analyzed during the first part, introducing age-structure. To do this, we considered a 16x16 matrix where the position c_{i_j} represented the average number of contacts between individual i and j. In the equations we included $\beta_{i_j} = q * c_{i_j}$, where q is the infection-specific parameter to calibrate, in order to be able to resemble the pattern of real data for the first 68 days. The only assumption we made concerned γ, so the recovery rate. The three differential equations were:

$$\frac{dS}{dt} = -S_i \sum_{j=1}^{16} \beta_{i_j} \frac{I_j}{N_j}$$

$$\frac{dI}{dt} = S_i \sum_{j=1}^{16} \beta_{i_j} \frac{I_j}{N_j} - \gamma I_i$$

$$\frac{dR}{dt} = \gamma I_i$$

In our function in Python, we tested again R0 from 1 to 3, and we defined q as:

$$q = \frac{R0}{(average\ duration\ of\ infectious\ period)\rho(C)}$$

where ρ(C) is the is the spectral radius of matrix C, that we computed as 20.89464. Finally, we made the infection start from the oldest age group, as it is what indeed happened in reality: the first case in Como is dated February 27th 2020.

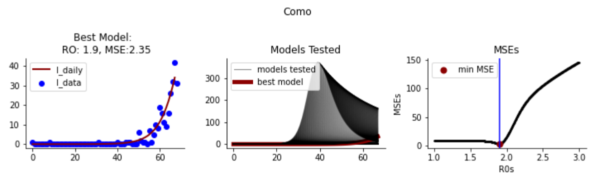

Figure 3: Best Model for S-I-R with age structure

The best q that we calibrated resulted from the lowest MSE, was 0.0181577. Results were also better than the previous part, since the MSE that we got was 2.35, while before it was 3.26. On the contrary, R0 was lower than before (1.90 instead of 1.97).

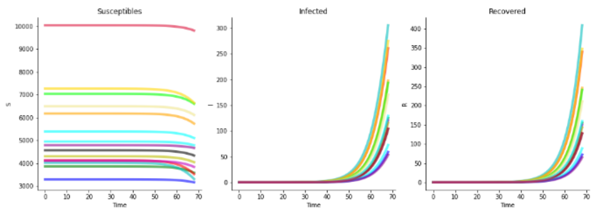

Figure 4: Behavior of the three compartments over 68 days

Taking a look at the curves of the different compartments in the figure 4, we can see that Infected are just increasing from day 40 on, with no signal of a possible stop. Also Recovered are increasing, with a good speed.

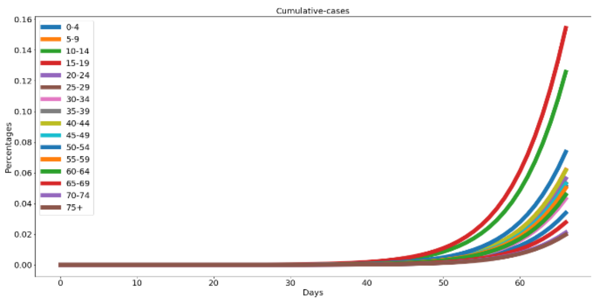

Figure 5: Cumulative cases at 68th day

We also computed the cumulative cases until the 68th day, to understand which of the groups got infected the most. Result was that the first two groups for cumulative cases were 15-19 (17.75%) and 10-14 (14.53%), mostly due to the fact that the value of the contact matrix is way higher than the others for these two groups. The value for the group that started the infection, 75+, is very low (2.31%), but also this is due to the fact that interactions are a few among these people.

Model improvements

In the third and final part, we extended the age-stratified model, adding one or more of the restrictive measures that were introduced by the government after increase of cases back in March. In dealing with this task, we took into account the timing of the policies introduction, and we decided to have two different approaches: one deterministic, and the other stochastic.

a) Deterministic approach

The first consisted in splitting the time-range in 4 steps: 0-60, 61-70, 71-80, 81-92. We left the differential equations free for the first 60 days with the q that we calibrated before, in order to have a quick increase in Infected, and then to introduce closing of schools, by multiplying the summation with a 16x16 matrix all made by ones, except for the top-left box (from 0-4 to 20-24) which was substituted by a factor that we wanted to calibrate, in order to resemble the real pattern of data from t = 60 to t = 69.

$$\sigma_{i_j} \sum_{j=1}^{16} \beta_{i_j} \frac{I_j}{N_j}$$

The terms inside the summation were the same, but outside we added σ_{i_j}, the "school matrix". The calibration was done by keeping the value that minimized the MSEs. The result that we got for this box was a value of 1.224. The explanation behind this value is that we trained our model on data that is still increasing, but this is what happens in real world: results of an interventions can be seen after some days, and not immediately (Soltesz et al., 2020).

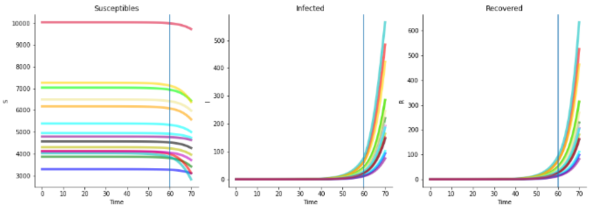

Figure 6: Closing schools intervention

After this, we introduced at time 70 the quarantine, by multiplying all the sum term that we computed before by another coefficient.

$$\alpha\sigma_{i_j}\sum_{j=1}^{16}\beta_{i_j}\frac{I_j}{N_j}$$

In this sum, the α term is introduced to capture the decreasing pattern of data. Even here, we decided to calibrate the term by minimizing MSE. The resulting term was 0.609. Indeed, by plotting the graph, we noticed an immediate flexion in all curves of Infected people. The pattern that we wanted to resemble by adding this term was from t = 70 to t = 79.

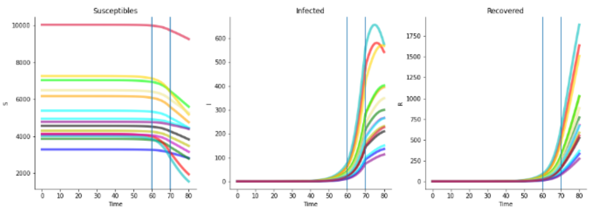

Figure 7: Quarantine intervention

Last but not least, we decided at day 80 to introduce the fact that masks were compulsory, with another term acting on the reduction of the summation term.

$$\phi\alpha\sigma_{i_j}\sum_{j=1}^{16}\beta_{i_j}\frac{I_j}{N_j}$$

This time, the ϕ term is introduced again to capture the final step of the curve, with a strong flexion. We know masks are key to prevent the COVID infection, so we expected a low coefficient and, indeed, the term resulted to be 0.42 after our calibration. So, by adding three terms at time t = 60, t = 70 and t = 80, and keeping them for the rest of the time, we were able to reproduce the real pattern.

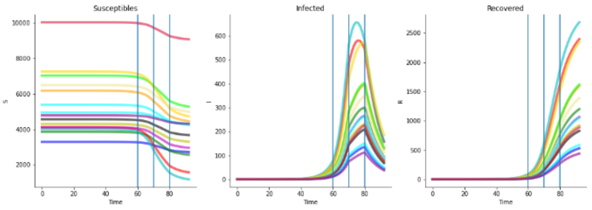

Figure 8: Compulsory masks intervention

As we can easily see from the plot, first quarantine and then compulsory masks were very important and decisive in the reduction of Infected people. As a final test, we decided to compute the MSE between the simulated data that we got by splitting the curve in 4 parts and introducing step by step a new coefficient and real data. The resulting MSE was 5.15, so by calibrating term by term the introduction of multiple subsequential interventions, the estimated final curve is particularly good in predicting what really happened.

b) Stochastic approach

The second approach was the stochastic one: this time, we didn't use the *odeint* pre-built function, but we made the coefficient (lambda) changing over time, by introducing their dependancy on S, I and R.

$$\lambda = \sum_{j=1}^{16}\beta_{i_j}\frac{I[-1]_j}{N_j}$$

This time, we introduced a function (ComputeTransition), which returns a list of samples from a binomial distribution based on the inputted parameters. Here, the parameters that we introduce are the S and the I at every time step, and the function returns a list of samples based on the value of lambda at that time. That value was re-computed at each iteration by taking into account the last value computed of I, always with the ComputeTransition function (for this reason, in the equation is $I[-1]$. At each iteration, we stored the value that we computed before, in order to get a final list of S, I and R over time. In order to improve the stochastic model, we introduced two measures: closing of schools from time t = 50 on, and quarantine from time t = 69 on. For the first one, we acted again on the β matrix, by multiplying it for the same matrix we used in the deterministic approach, in order to have an effect just on the first 5 age-ranges:

$$\lambda = \sigma_{i_j} \sum_{j=1}^{16} \beta_{i_j} \frac{I[-1]_j}{N_j}$$

For the second one, we multiplied the λ computed before with a coefficient, in order to reduce the moving average from S to I.

$$\lambda = \alpha\sigma_{i_j} \sum_{j=1}^{16} \beta_{i_j} \frac{I[-1]_j}{N_j}$$

After computing 100 simulation, we have a tensor, so 3 dimensions. For this reason, we decided to compute the mean over the 100 simulation, so to have an average of the results we got for each simulation. In this way, we had three resulting matrix of shape 92x16, so 92 rows representing the days, and 16 values per row, representing the 16 age ranges. Those values inside the matrix are the mean of the 100 simulations. Then, we decided to plot the results.

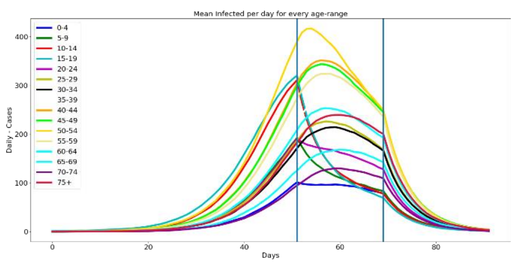

Figure 9: Plot of Infected change over time

As we can easily see, from time t = 50 on, some groups already show a flexion, due to closing of schools. The most interesting thing is that now, with the closing of schools, infections are reducing rapidly for the two age-ranges that were greatly hit by the virus, and now the highest peak is reached by the 50-54 age-range. This is what actually happened: the target of COVID-19 changed over time. From hitting hard mostly older people, now the largest share of the population who has to suffer is people in their fifties. After t = 68, it's interesting to see that, within all age-ranges, infected people are rapidly decaying over 0.

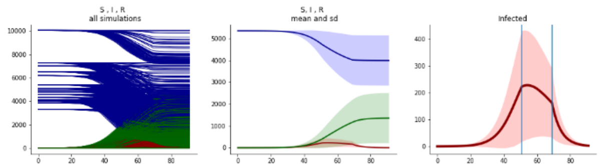

Figure 10: Triple plots with means and standard deviations

Also from these plots we can denote the fact that the average of all age-ranges behavior is showing the action both of closing of schools and of quarantine.

Conclusions

In this paper, we showed how important the timing of interventions is when dealing with pandemics like COVID-19. In the first part, we saw that, without restrictions, the curve can reach peaks that could be dangerous for the sanitary systems of most countries in the world. After that, we saw that groups that are most in contact are the ones that are most likely to have a higher number of infected. Finally, we learnt that taking actions in the right moment could be fundamental both to save people's lives and to keep the economy going: if peaks are reached, introducing quarantine when the curve is low could just cause economic problems, while if it is introduced before reaching the peak, it could save lots of lives. In the future, it could be interesting to develop a similar age-structured model, to understand the behavior of the different categories with the introduction of the vaccine already during the spread of the disease.

References

“Archived: WHO Timeline - COVID-19.” *World Health Organization*, World Health Organization, 2020, www.who.int/news/item/27-04-2020-who-timeline---covid-19.

Patrick W Brady, MD, et al. “Trends in COVID-19 Risk-Adjusted Mortality Rates.” *Journal of Hospital Medicine*, 17 Mar. 2021, www.journalofhospitalmedicine.com/jhospmed/article/230561/hospital-medicine/trends-covid-19-risk-adjusted-mortality-rates.

Bilancio Demografico Nazionale, 20 July 2020, www.istat.it/it/archivio/245466#:~:text=Al%2031%20dicembre%202019%20la%20popolazione%20residente%20in%20Italia%20ammonta,caratterizzato%20gli%20ultimi%20cinque%20anni.

Byrne, Andrew William, et al. “Inferred Duration of Infectious Period of SARS-CoV-2: Rapid Scoping Review and Analysis of Available Evidence for Asymptomatic and Symptomatic COVID-19 Cases.” *BMJ Open*, British Medical Journal Publishing Group, 1 Aug. 2020, bmjopen.bmj.com/content/10/8/e039856.

Istat. *Statistiche Istat*, dati.istat.it/Index.aspx.

Lauer, Stephen A., et al. “The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application.” *Annals of Internal Medicine*, 22 May 2020, www.acpjournals.org/doi/10.7326/M20-0504.

Soltesz, Kristian, et al. “The Effect of Interventions on COVID-19.” *Nature News*, Nature Publishing Group, 23 Dec. 2020, www.nature.com/articles/s41586-020-3025-y.