Predicting COVID-19 Evolution during Mid-March Crisis

Pedro Furtado
University of Coimbra
Polo II, Coimbra, Portugal
+35123700000
pnf@dei.uc.pt

ABSTRACT

The corona virus responsible for COVID-19 has come into our lives with huge stampede. Every human activity has been seriously hurt and millions were confined to their homes. As of March, people in Europe wonder whether the confinement, closures and no-flights policies are effective or how effective they are, in spite of the positive previous example of China. In this paper we present our analysis specifically focused at detecting whether the new daily cases curves are in a stabilization route or exploding. This required a set of steps for data processing and analysis that we describe in detail. The conclusion is that, as of 22 March, the curves were in a trajectory of stabilization and possible decrease soon. We show why, also finding a most probable correlation with confinement and other government policies.

CCS Concepts

• Applied computing → Health informatics

Keywords

Epidemiology; Data analysis; Analysis of variation

1. INTRODUCTION

In January the world woke up to the news that China had a new outbreak of a deadly virus. The novel coronavirus, identified by Chinese authorities on January 7 and since named SARS-CoV-2, is a new strain that had not been previously identified in humans. Signs of infection include fever, cough, shortness of breath and breathing difficulties. In more severe cases, it can lead to pneumonia, multiple organ failure and even death, affecting particularly older people and people with other health-related conditions. Covid-19 death rate was estimated to be around 3% and the reproduction rate somewhere between 2 and 3. At those initial times, the virus was already a potential threat to the rest of the world, however governments and health officials in most countries were slow to act because there was no idea that it would spread as it did.

A new virus may be especially dangerous to humans if it is simultaneously sufficiently virulent and deadly and there is a lack of immunity. Immunization eventually happens as the virus infects a vast majority of the population or when the population is

inoculated by vaccination, but until one or the other happens this experience has proven that a pandemic such as Covid-19 can cause mayhem.

Estimates of the incubation period - the time between infection and the onset of symptoms - range from 1 to 14 days, most infected people showing symptoms within five to six days. One of the problems of Covid-19 was that asymptomatic carriers infect others before they even realize they are infected, and since the virus spreads easily and there is no immunity in humans to that new strain. This way the virus has been able to spread all over the world and by mid-February the situation started to feel out of control.

1.1 The Meaning of r_o

The reproduction number (r_o) is a single number with a significant meaning. It helps health organizations determine if the outbreak will spread. If r_o is greater than 1, then the disease will probably spread. It is the average number of people who will be infected by a single infected person. As an example, a r_o of 2 means each infected person will in average infect two other persons. In spite of being so relevant, r_o is hard to determine. Several groups used different methods to try to determine r_o for SARS-CoV-2, their estimates varying mostly within the range of 2 to 3.

At around the beginning of February, patients started testing positive for Covid-19 in Italy, and in just one month and a half the virus spread quickly to other places in Italy, other European countries and to the rest of the world, and the daily numbers of new cases and deaths followed an exponential growth curve in several places.

Initially, as the virus started spreading, governments and health authorities were slow to react, probably due to the potential huge damage that action would bring to the economy. In particular, flights continued unabated and some countries had no significant checks or quarantine requirements. Even testing for the disease was restricted, initially. Of course, this all changed as soon as the numbers skyrocketed and as everyone watched on TV hospitals overcrowded in Italy, health staff completely desperate and double daily death figures such as more than 500 just in Italy.

One after the other, countries started closing schools, bars, restaurants, non-essential shops, social gatherings, flights and other activities, until finally states of exception were raised and people confined to their homes as much as possible.

Our current work was motivated precisely by the danger posed by the exponential growth of the SARS-CoV-2 virus among the population. Our goal was to analyze the available data to understand whether the current government actions are going to have relevant effects. This is important since the number of daily cases and the number of total cases is seen increasing, therefore people who are confined may think that the measures are not being well-succeeded, and that in turn can make them less careful

This work is licensed under a Creative Commons Attribution International 4.0 License

ICBBT 2020, May 22–24, 2020, Xi'an, China

© 2020 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-7571-9/20/05.

DOI: <https://doi.org/10.1145/3405758.3405771>

about confinement and other important actions such as washing their hands frequently after arriving home and so on. We take a data-driven approach to study the initial exponential growth of the SARS-CoV-2 virus in the helpless society, and the effect of social isolation and other mechanisms in turning that terrible evolution into a sigmoid or logarithmic curve.

2. RELATED WORK

The mathematics of infectious diseases is reviewed in work [7], and references [3], [1], [4] and [2] are textbooks on the subject. The dynamics of an epidemic can be modelled as a set of differential equations, and one of the most famous initial models is SIR, by Kermack and McKendrick [8]. The mnemonic SIR stands for Susceptible (S), Infectious (I) and Recovered (R). Given the three possible states (S, I and R), they are organized sequentially (S->I->R) such that individuals transition between those states, and the differential equations describe the change in the stock of each state per unit of time. Naturally, the sum of changes of S, I and R is 0, since the individuals transition between states (one more individual in one state corresponds to one less in the previous state. Between S and I the transition rate is βI , where β is the average number of contacts per person per time, multiplied by the probability of disease transmission in a contact between a susceptible and an infectious subject. Between I and R the transition rate is γ (number of recovered or dead during one time unit divided by the total number of infected on that same time interval). If the duration of the infection is denoted by D, then $\gamma = 1/D$, since an individual experiences one recovery in D units of time. It is assumed that the permanence of each single subject in the epidemic states is a random variable with exponential distribution. With these differential equations it is possible to simulate the dynamics of a specific epidemic. In [5] the authors studied analytical solutions to the SIR model with equal death and birth rates. SIR has given rise to numerous other derived models. As an example, in [6] the authors present and analyze three basic epidemiological models.

SIR and other related models can be applied to Covid-19 and simulation ensues however the work we present here focuses solely on a data-driven analysis of the rates of growth, analyzing the countries and trying to understand whether they are evolving positively in terms of containing the current outbreak. Consequently, a simpler derivative model analyzing only the rate of change of new cases is sufficient for our purposes.

Another concept very much related to our work is the concept of exponential and sigmoid growth. Exponential growth describes a possible way that a variable may increase over time. As an example of use, in [9] the authors study how cells can grow exponentially at a constant rate while remodeling their metabolism and gene expression. In exponential growth the rate of change is proportional to the value itself. The function that models exponential growth is an exponential function of time (i.e. time is the exponent). Exponential growth is important in this work because a virus typically will spread exponentially if no artificial immunization is available and the reproduction rate is high enough, as each infected person can infect multiple other persons. Our current work was motivated precisely by the danger posed by the exponential growth of the Covid-19 virus among the population. We take a data-driven approach to study the initial exponential growth of the Covid virus in the helpless society, and we analyze curves trying to find whether the measures and public alarm contribute to ease the spread. In short, what is the potential effect of social isolation and other mechanisms in turning that terrible evolution into a sigmoid or logarithmic curve.

3. MODELING GROWTH AND RELIEF

3.1 Exponential and Sigmoid Growth

Mathematical models can be used to accurately describe changes in a population and to predict future changes as well. In population growth theory, r (per-capita rate of increase or rate of growth) is used to define population growth as being either exponential, if r is constant, or logistic (a.k.a. logarithmic) if r decreases as the population grows. Given N , the population size, equation 1 describes exponential growth, while equation 2 describes logistic growth,

$$dN/dt = r N \quad (1)$$

$$dN/dt = r (k - N) N \quad (2)$$

In (2) k is the carrying capacity, the maximum population size that a specific environment can support. The idea is that, in order to reproduce, a population needs resources, such as energy and other nutrients, therefore the growth will be restricted by such carrying capacity.

Figure 1 illustrates the “daily rate of change” (dN/dt) and the population size (N) of an exponential growth distribution with growth rate factor r of 3. Figure 2 shows a logistic growth with the same r and k , the carrying capacity, equal to 25000. The meaning of this 25000 is that the environment can only provide for as many as 25000 specimens. Note that in Figure 2 the curve initially looks like exponential growth but, as the population size nears k (25000) the rate or increase decreases significantly.

Note also that the equation for the total population size at any time in exponential growth is,

$$N = n_0(r+1)^t \quad (3)$$

In this formula n_0 is an initial number of cases.

Figure 1. Exponential growth example ($r=3$).

Figure 2. Logistic growth example ($r=3$).

Our objective in the case of Covid-19 is that the rate of increase, initially an exponential-shaped curve, becomes a sigmoid. In our

case this is achieved by social isolation, confinement and closures, to break the chain of transmission and therefore decrease the environment conditions favorable to the spread.

3.2 Turning the Tide

The first objective of epidemic containment should be to turn the exponential growth into a sigmoid or logarithmic curve as soon as possible, through either immunization, which is currently unavailable for the Covid-19 pandemic (as of March 2019), or by resorting to the only viable existing alternative, which is social isolation and confinement. Social isolation and confinement intends to decrease the r (per-capita transmission rate) and is achieved by closing most activities, schools and industry included, except those that are absolutely necessary, such as hospitals, pharmacies, food production and supermarkets. If r is less than 1, the epidemic is contained and we can see a decrease of those infected along time. Figure 3 shows the three moments, (1) exponential growth, (2) logarithmic growth, and (3) the decrease in those infected.

Figure 3. Turning exponential growth curve into sigmoid.

3.3 Data-driven Approach

Our data-driven approach intends to evaluate whether containment is being well-succeeded by analyzing the data. Data-driven approaches are possible today mainly because data is updated almost in real-time. In our case we have daily data on the number of new cases (infected), deaths and recoveries all over the world, organized by country. Our methodology consisted on the following phases:

1. Data collection: data collection involves accessing the sources and extracting the relevant daily data about Covid-19, e.g. through web-scraping;
2. Organizing the data: the data is organized as time series along the days since the start of the pandemic. We have detailed data from each country that was affected;
3. Pre-processing of the data: in this step we apply a set of functions to clean the data and to prepare it for analysis. Note that this is one of the most important steps, because there are a number of uncertainties in the way the various countries handle the process. In particular, countries may do more or less testing of suspect infections, and we notice that some days there are many more cases reported than others, therefore we need to clean the data;
4. Data transformation: in the data transformation step we transform the input data into actionable knowledge. In this case we concentrate in a single detail, the curbing of the exponential growth. Therefore, we planned the appropriate key analysis to evaluate this factor, taking into consideration that if we can detect

that a sigmoid curve is happening anytime soon, it means the pandemic is probably going to recede, at least partially. This simple analysis needs to be done for each country, which we did, however in this work we only illustrate for a few countries for size reasons;

5. Analysis of potential correlation of government measures (e.g. confinements, closures, other) with evolution of the daily rate.

Using an experimental approach and the steps described above, we tested several data possible sequences of data transformations and concluded for a specific sequence of transformations over the series of daily new cases, which is our choice of main input for the analysis:

- a. Smoothing daily new cases – the raw data is not smooth, because of some factors. As an example, in many countries some days have many more tests than others, and sometimes there is a delay in reporting (which means that specific daily quantities may be accounted for in posterior days). As a result, we observed that the data has oscillations that need to be smoothed by pre-processing. Moving averages over the daily numbers allows these effects to be smoothed;
- b. Obtaining the series of daily rates of change – we obtain the series of daily rates of change (dC/dt) from the moving averages, where C is the number of cases in each day. This way we are able to analyze how the daily rate of change is evolving. The dC/dt is obtained as $C_i - C_{i-1} / C_{i-1}$ for each day, and results in a sequence of rates of change. Our objective is to observe whether rates or change are decreasing as governments impose measures;
- c. Cleaning the series of daily rates of change per country - since C_{i-1} is 0 before the first case happens, the first non-zero daily rate of change would be infinite (divided by $C_{i-1}=0$). Additionally, while there are too few cases (e.g. 3) the daily rates of change are very high (e.g. another 3 new infected persons is a rate of change of 100%, even though it represents only 3 new patients). Since we are analyzing the rates of changes, it is important to not consider these initial huge changes by filtering them out;
- d. Further smoothing of the daily rates of change prior to analysis – even after smoothing the initial series of daily new cases (a), the calculated daily rates of change (b) still had some degree of oscillation that we smoothed by applying moving averages on that series. This removed abrupt inter-day variations are smoothed to ease the job of analysis further;
- e. Analysis of the curves – In this step we analyzed the daily rate of change series for each country. In this work we provide a few examples and show how we did it, also concluding for those countries.

4. CLEANING AND TRANSFORMATIONS

The countries outbreak evolution daily raw data is frequently dirty, in the sense that the number of new cases can vary widely from one day to the next. Figure 4 shows this effect. Here we show daily cases per country in February and March for a subset of countries.

4.1 Moving Average of Daily New Cases

Figure 5 shows the result of applying moving averages over the countries data. To obtain this result we applied a moving average $d'_i = [d_{i-1} + d_i + d_{i+1}] / 3$, where d is the original series and d' is the new series. From comparison with Figure 4 we can conclude that some

sudden variations that are actually an artifact of the test and report process were smoothed.

Figure 4. Daily cases for a set of countries.

Figure 5. M-Averaged daily cases for countries.

5. COMPUTING THE DAILY RATES OF CHANGE

Given the series of daily new cases of a country d , the daily rates of change is a series given by $(d_i' = d_i - d_{i-1})$. Figure 6 shows the daily rates of change for the same countries of Figure 5. Note that this result is not ready for analysis of rate of change because, for at least some countries, there is an initial huge value for the rate of change. This is an expected artifact of computing a relative value since, as we explained previously, a change from 0 to 3 is an infinite increase (because the initial value was zero), and a change from 3 to 6 is a 100% increase.

Figure 6. Daily rates of change for countries.

6. REMOVING TRAILING RATES OF CHANGE

This step, which we denoted as truncate (TRUNC), removes the first peak of the series of daily rates of change. Given a series d where i_0 is the index of the first non-zero case, the algorithm zeroes all values from i_0 to the first index i_1 with both more than n_D daily cases and rate of change below a specified threshold Thr (these two parameters must be given). Figure 7 shows the new version of the series of Figure 6 after applying the algorithm with $n_D=25$ and $Thr=1$ (these parameter mean to zero elements before the first instance of more than 25 daily cases occurs and then after the daily rate of change decreased below 100%). Note that in Figure 7 the maximum rate of change is 2 (200%), whereas in Figure 6 it was 17.5 (1750%!).

Figure 7. Truncated daily rates of change for countries.

7. ANALYSIS OF COUNTRIES

After the previous steps the dataset became ready for country analysis. It is not possible to show here the analysis for all 20 considered countries, instead we concentrate in a few illustrating examples (e.g. we have chosen Italy because it one of the epicenters). Figure 8, Figure 9 and Figure 10 show the analysis for Italy, Spain and USA respectively. The next section will correlate the evolutions with measures taken by the governments and conclude.

In most of the countries analyzed in our study, after one or more peaks, the daily rate of change is already decreasing or inverting its tendency and starting to decrease, but in most cases it is still above 0. The observed rate decrease can possibly be correlated with social distancing, since the disease spread has triggered a huge alarm and authorities have been very vocal regarding the danger and the need to take shelter. Posteriorly, authorities acted also by closing non-essential institutions and making social closure and confinement mandatory. The decrease in daily rate of change is also a sign of sigmoid behavior of new daily cases, an important improvement of the conditions. It means that, looking at Figure 3, we are in the sigmoid phase for most countries. In the future, as soon as this rate turns below 0, it means that the total number of cases starts decreasing, the third phase in Figure 3. According to these values, as of March 22, the USA still had a rate of daily increase of new cases of around 25%, as is visible in Figure 10, while Italy was around 8% as visible in Figure 8 and Spain was around 6%, as visible in Figure 9.

Figure 8. Daily rates of change for Italy.

The daily rates of change for China, the first country affected, is also shown in Figure 11.

8. CORRELATING WITH EVENTS

In this section we resort to analysis of policy milestone events to help analyze the evolution further. In Figure 12 we can see that, in the case of Italy, February 25 marked the closure of 11 towns in the north of Italy, which were the main focus of the outbreak (confining 50K people). The rate of change still increased from 30 to 35% for 2 days, but then inverted its evolution to decrease at a reasonably fast rate. Later on, when the curve was flattening at around 20% further political actions, first the north confinement (March 08) and then confinement of the whole country (March 11) are the most probable cause for further and faster decrease to the current 8% daily change rate. As of March 22, Italy had a daily rate of new cases above 5000, and still increasing (the 8% rate of increase), but this rate of increase was slowing and, in a trajectory, to soon cross the 0 axis and start decreasing the number of new daily cases).

Figure 9. Daily rates of change for Spain.

Figure 10. Daily rates of change for USA.

Figure 11. Daily rates of change for China.

The case of Spain showed in Figure 13 shows significant increase until around the 11th March. The public alarm that the rapid increase in cases and deaths brought, alerts by the government and health officials and the start of government actions to restrict and close are very plausible causes for the start of the decrease in the rate of change from 11th March on. Then on March 14 the Spain government declared state of alarm and confinement. The Figure shows that the rate of daily change decreased very significantly until the March 22 value of 6%, and was expected to continue decreasing as of March 22.

Figure 12. daily rates change correlated to events-Italy.

Figure 13. daily rates change correlated to events-Spain.

In both cases discussed above (Figure 12 and Figure 13) it may seem surprising that the curves are already starting to go down when the most stringent government measures kick-in. However, one must consider that, some days before commanding closures and lockdowns, the governments were already asking people to stay at home, quarantining people and other strong sanitary actions were being introduced on a daily basis, also very quickly raising public awareness. In essence, people also got scared and decreased their social contacts and actions drastically even before lockdowns.

The case of China, shown in Figure 11, is more difficult to analyze because, at a certain point in time, there were some adjustments in the number of infected. We can assume that, if that number is moved earlier, the causality between lockdown and improvement also becomes apparent.

9. BRIEF ON ANALYSIS OF OTHER COUNTRIES

We have analyzed more than 20 countries using the same approach shown for the illustrating examples, most evidence pointing also to additional confirmation of the relevant link between government measures, also public alarm and significant reduction of the rate of increase, and later on actual decrease of the daily rate. Our future work on the issue will include forecasting the future curve to determine the expected eradication of the disease in each country.

Finally, note that the analysis is only as good as the raw data that is collected. There are a lot of factors that impact the veracity of the raw data itself, e.g. degree of availability of testing for Covid-19 in a certain country, wrong diagnosis, delays in reporting and many other details.

10. CONCLUSIONS

In this work we proposed a data-wise approach to analyze the evolution of Covid-19 cases in the current pandemic in several countries. We have shown how we prepared and analyzed the data and, most importantly, a degree of correlation of the data with important government measures such as mandatory confinement and closures. The results provide important evidence to the relevance of those measures, the slowing of the exponential growth and the future decay trajectory of the daily new cases.

11. REFERENCES

- [1] Altizer, S., Nunn, C. (2006). Infectious diseases in primates: behavior, ecology and evolution. Oxford Series in Ecology and Evolution. Oxford: Oxford University Press. ISBN 0-19-856585-2.
- [2] Anderson, R. M. (1982). Population Dynamics of Infectious Diseases: Theory and Applications. London-New York: Chapman and Hall. ISBN 0-412-21610-8.
- [3] Bailey, N. T. (1975). The mathematical theory of infectious diseases and its applications (2nd ed.). London: Griffin. ISBN 0-85264-231-8.
- [4] Brauer, F.; Castillo-Chávez, C. (2001). Mathematical Models in Population Biology and Epidemiology. NY: Springer. ISBN 0-387-98902-1.
- [5] Harko, T., Lobo, F. S., Mak, M. K. (2014). Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. Applied Mathematics and Computation. 236: 184–194. doi:10.1016/j.amc.2014.03.030.
- [6] Hethcote, H. W. (1989). "Three Basic Epidemiological Models". Applied Mathematical Ecology. Biomathematics. 18. Berlin: Springer. pp. 119–144. doi:10.1007/978-3-642-61317-3_5. ISBN 3-540-19465-7.
- [7] Hethcote H (2000). "The Mathematics of Infectious Diseases". SIAM Review. 42 (4): 599–653. doi:10.1137/s0036144500371907.
- [8] Kermack, W. O.; McKendrick, A. G. (1927). A Contribution to the Mathematical Theory of Epidemics. Proceedings of the Royal Society A. 115 (772): 700–721. doi:10.1098/rspa.1927.0118.
- [9] Slavov, N., Budnik, B. A. Schwab, D., Airoidi, E. M., van Oudenaarden, A. (2014). Constant Growth Rate Can Be Supported by Decreasing Energy Flux and Increasing Aerobic Glycolysis. Cell Reports. 7 (3): 705–714. doi:10.1016/j.celrep.2014.03.057. ISSN 2211-1247.