Bee The Data S.L. research department,
IESE Business School

Review on multi-region epidemic evolution and optimal control approach.

April 29, 2020
Miguel Ángel Ariño (IESE)

Abstract

In this document we make a review an existing model for multi-region pandemic propagation based on the SIR model (see [6, 5, 2, 3]). We then proceed with identifying three weak points of the reviewed multi-region SIR model that may lead to performance loss when modeling covid-19 propagation. Finally we conclude the document by proposing a novel modification of the multi-region SIR model that solves such weaknesses by shifting from a SIR model to a SEIQRD model, introducing a new intra-region restriction parameter and uses the more reliable number of deceased individuals[1] instead of the number of infected individuals to perform parameter training.

Research field: epidemiology, machine learning.

Key-words: covid-19, multi-region epidemic modeling, SIR, SEIQRD, restriction optimization,

Introduction and goals.

Due to covid-19 outbreak many countries have been forced to implement movement restriction measures on the whole population in order to prevent further propagation of the disease. Such measures however have terrible economic consequences and will probably lead to huge economical crisis not to mention the negative effects a prolonged state of isolation can have on the mental health of the isolated individuals. In order to mitigate these negative effects it is important to be able to predict when and where will it be safe to revoke the imposed restriction measures. In order to do so we must be able to predict exactly how the disease spreading will evolve to different levels of restrictions applied to different regions of a given country. If we where able to do so then we could use the predictions in order to determine the optimal degrees of severity with which restrictions should be applied a t every given point in time in order to minimize the direct negative effects of the epidemic (total number of infected or deceased) as well as the number of restriction imposed and their duration.

There are several existing mathematical models that address this situation however the vast majority of them (see [6, 5, 2, 3]) are based on the SIR model (see [4]) to which are added a term for considering propagation of disease between multiple regions, a term for considering the restriction of interactions between regions and a cost function that, when optimized, leads to the optimal restriction related parameters which should be imposed in order to simultaneously minimize the total number of infected individuals and the amount of imposed restrictions.

The goal of this document is double, on one hand we wish to describe the mentioned variation of the SIR model while on the other hand we want to expand it in order to take into account restriction measures that could have an effect in the propagation of the disease within a region (for instance at home confinement). Our modified model will also take into account other characteristics of the SARS-CoV2 virus that are not considered in [6, 2, 5]. For instance we will consider the incubation period of the virus (as done in [3]), isolation (i.e. hospitalization) of some infected individuals and possibility of deceasing from the infection (see the SEIQRD model described in [1]).

[1] This value is updated daily and can be easily retrieved from ISCIII in the case of Spain.

Current state of the art.

As mentioned earlier most of the currently existing models that attempt to simulate the evolution of an epidemic considering interaction between different regions affected by such epidemic are based on the SIR model.

The SIR model is an acronym for *Susceptible Infected Recovered* which represent the categories in which population is divided in this model. Namely we have that *susceptible* individuals are not infected by the virus and can become infected by it, *infected* individuals are infected by the virus and can infect other individuals and *recovered* individuals are neither infected nor can become infected. The order in which this categories appear in the model name correspond to the full path an individual that contracts the disease will follow. That is at first the individual is susceptible, then it becomes infected and finally it either dies (exiting the contemplated categories) or recovers building an immunity to the disease.

This model can be used in order to predict the total number of individual through time and is usually presented in its differential form where the evolution of each category S, I or R is described by some differential equations. However in most of its variations that introduce multiple regions (see [6, 5, 2, 3]) it appears in its discrete form. In this form the number of susceptible, infected and removed individuals at time $i+1$ (for example at day $i+1$) are obtained from those same numbers at time i according to the following equations

$$S_{i+1} = S_i - \beta I_i S_i - dS_i + bS_i, \tag{1}$$
$$I_{i+1} = I_i + \beta I_i S_i - (\alpha + \gamma + d)\, I_i, \tag{2}$$
$$R_{i+1} = R_i + \gamma I_i - dR_i, \tag{3}$$

where b and d indicate respectively the natural birth and death rates (which are independent of the epidemic), β indicates the infection rate, γ is the probability of recovery from the disease and α is the death rate due to the infection. For short periods of time (a few months or even a year) we can safely consider the birth and death rates to be zero since there will not be relevant for computing population growth. Because of this from now on we will consider $b = d = 0$.

When introducing multiple regions, equations 1, 2 and 3 become slightly more complex. First of all population is separated into multiple regions, that we will denote by Ω, and for each region $r \in \Omega$ the population of that region is divided as before into the susceptible, infected and recovered categories. Analogously to earlier these region dependent categories will be denoted by S^r, I^r and R^r for every $r \in \Omega$. Moreover, for each region $r \in \Omega$ a set $V_r \subseteq \Omega \backslash \{r\}$ of neighbors is established (i.e. V_r is the set of all those regions that can interact with r) and for each neighbor $s \in V_r$ the inter-region infection rate $\beta_{r,s} = \beta_{s,r}$ is determined. Finally, for every region $r \in \Omega$ equations 1, 2 and 3 are modified in the following way

$$S^r_{i+1} = S^r_i - \beta_r I^r_i S^r_i - \sum_{s \in V_r} u^{r,s}_i \beta_{r,s} I^s_i S^r_i, \tag{4}$$
$$I^r_{i+1} = I^r_i + \beta_r I^r_i S^r_i + \sum_{s \in V_r} u^{r,s}_i \beta_{r,s} I^s_i S^r_i - (\alpha + \gamma)\, I^r_i, \tag{5}$$
$$R^r_{i+1} = R^r_i + \gamma I^r_i, \tag{6}$$

where β_r is the intra-region infection rate and $u^{r,s}_i = u^{s,r}_i \in [0,1]$ indicate the imposed restrictions for traveling between regions r and s at time i. Remember that we are considering $b = d = 0$ that is why some terms ave disappeared in equations 4, 5 and 6.

Finally the optimal values for the parameters $u^{r,s}_i$ in a given period of time $i \in \{0, \dots, N\}$ can be found by minimizing the following loss function

$$L\left(u^{r,s}_i, \dots\right) = \sum_{r \in \Omega} \left(w^r_I I^r_N + \sum_{i=0}^{N-1} \left(w^r_I I^r_i + \sum_{s \in V_r} \frac{w_{r,s}}{2} \left(1 - u^{r,s}_i\right)^2 \right) \right), \tag{7}$$

where both w^r_I and $w_{r,s} = w_{s,r}$ are positive and represent the severity weights associated to respectively the total number of infected individuals in region r and the applied traveling restrictions restrictions between

regions r and s. As we can observe from equation 7, once the desired weight values are set, minimizing the loss function will result in a simultaneous minimization of both the total number of infected individuals and the severity of the imposed restrictions. The resulting optimal parameters $u_i^{r,s}$ tell us exactly how severe restrictions of travels between regions must be at each point in order to obtain these optimal results which is exactly what we are interested in.

Proposed modification.

The above described model stands at the base of many multi-region epidemic evolution description models such as [6, 5, 2, 3]. It is a powerful model that allows us to obtain time dependent parameters indicating the restrictions that should be imposed to any two regions at any point in time in order to minimize both the total number of infected individuals and and the severity of the imposed restrictions. Moreover, as proven in [6, 5, 2, 3], the function defined in equation 7 has in fact , under certain reasonable conditions, a global minimal value which justifies its use as a loss function.

This model however presents a few inconveniences that do not combine correctly with the current covid-19 epidemic:

1. When infected individuals become immediately infectious without any incubation period. This problem is solved in [3] by switching the base model from a SIR model to a SEIRS model which considers both incubation period and possibility of re-infection (in our case this last one is in fact unneeded).

2. Restrictions on mobility within the same region are not considered. This can be solved by simply introducing time-dependent intra-region restriction parameters $u_i^r \in [0, 1]$ besides the already existing inter-region restriction parameters $u_i^{r,s}$.

3. The model optimization is based on the number of infected individuals at all given moments in time, however, since covid-19 is asymptomatic in many cases, it is practically impossible to determine the exact number of individuals that are infected at a given point in time. Thus, in practice, there will be no ground truth value for the number of infected individuals making it impossible to determine if the rest of the parameters of the model (particularly the infection rate β) are correctly set. This problem can be solved in conjunction with that presented in item 1 by switching from the simple SIR model used until now to the more complete SEIQRD model (see [1]). This model introduces the extra classes of *Exposed* individuals (which have been infected but are still in incubation period and cannot infect), *Quarantined* individuals (which have been identified as having contracted the disease and have been isolated due to this) and *Deceased* individuals (which have died from the disease). The number of individuals in this last category is usually much more reliable than the number of infected individuals and can be used in its place in the loss function defined in equation 7. A reliable and daily updated source for retrieving such number of daily deceased individuals is given, in the case of Spain by the ISCIII (Instituto de Salud Carlos III) which provides the daily number of deceased through time and for each region in Spain.

By applying the above described changes equations 4, 5 and 6 can be re-written, and completed with the introduction of the new Exposed, Quarantined and Deceased classes, as follows

$$S_{i+1}^r = S_i^r - u_i^r \beta_r I_i^r S_i^r - \sum_{s \in V_r} u_i^{r,s} \beta_{r,s} I_i^s S_i^r, \tag{8}$$

$$E_{i+1}^r = E_i^r + u_i^r \beta_r I_i^r S_i^r + \sum_{s \in V_r} u_i^{r,s} \beta_{r,s} I_i^s S_i^r - \sigma E_i, \tag{9}$$

$$I_{i+1}^r = I_i^r + \sigma E_i - (\alpha + \gamma + \delta) I_i^r, \tag{10}$$

$$Q_{i+1}^r = Q_i^r + \delta I_i^r - (\alpha + \gamma) Q_i^r, \tag{11}$$

$$R_{i+1}^r = R_i^r + \gamma I_i^r + \gamma Q_i^r, \tag{12}$$

$$D_{i+1}^r = D_i^r + \alpha I_i^r + \alpha Q_i^r, \tag{13}$$

where σ is an indicator of the inverse of the incubation period and δ is the probability for an infected individual to be isolated thus effectively reducing its infection ratio to 0.

Along with these modified equations describing the model's evolution a new loss function also arises replacing the total number of infected individuals with the total number of deceased individuals and now considering also the problem of minimization of the intra-region restriction parameters u_i^r. The new proposed loss function is the following

$$L\left(u_i^{r,s},\ldots,u_i^r\right)=\sum_{r\in\Omega}\left(w_I^r D_N^r+\sum_{i=0}^{N-1}\left(w_r\left(1-u_i^r\right)^2+\sum_{s\in V_r}\frac{w_{r,s}}{2}\left(1-u_i^{r,s}\right)^2\right)\right), \tag{14}$$

where the new parameter w_r is positive and represents the severity weight associated with intra-region restrictions. Notice that, similarly to what happened with equation 7, once appropriate values for w_I^r, w_r and $w_{r,s}$ are set, minimizing equation 14 results in a simultaneous minimization of both the total number of casualties due to the disease and the severity of the imposed restrictions. Moreover the resulting optimal parameters u_i^r and $u_i^{r,s}$ tell us exactly how severe must be the imposed restrictions for every region at each point of time which is exactly what we are interested in.

References

[1] *Optimization of Prognostication Model about the Spread of Ebola Based on SIR Model*, Proceedings of the 2016 6th International Conference on Machinery, Materials, Environment, Biotechnology and Computer, 2016.

[2] Imane Abouelkheir, Fadwa Kihal, Mostafa Rachik, Omar Zakary, and Ilias Elmouki. A multi-regions sirs discrete epidemic model with a travel-blocking vicinity optimal control approach on cells. *British Journal of Mathematics & Computer Science*, 20:1–16, 02 2017.

[3] Fadwa El Kihal, Mostafa Rachik, Omar Zakary, and Ilias Elmouki. A multi-regions seirs discrete epidemic model with a travel-blocking vicinity optimal control approach on cells. *International Journal of Advances in Applied Mathematics and Mechanics*, 2017.

[4] Howard (Howie) Weiss. The sir model and the foundations of public health. Technical report, Universitat Autonoma de Barcelona, 2013.

[5] Omar Zakary, Mostafa Rachik, and Ilias Elmouki. A new epidemic modeling approach: Multi-regions discrete-time model with travel-blocking vicinity optimal control strategy. *Infectious Disease Modelling*, 2(3):304 – 322, 2017.

[6] Omar Zakary, Mostafa Rachik, and Ilias Elmouki. On the analysis of a multi-regions discrete sir epidemic model: an optimal control approach. *International Journal of Dynamics and Control*, 5, 09 2017.