

Analysis of COVID-19 data with PRISM: preliminary thoughts

Paolo Milazzo^[0000-0002-7309-6424]

Department of Computer Science, University of Pisa
Largo B. Pontecorvo, 3, 56127, Pisa, Italy – paolo.milazzo@unipi.it

Extended Abstract

Epidemic phenomena are often studied by means of a *SIR model* [6]. This happened also for COVID-19 (or better, SARS-COV-2) pandemic, with several extensions of the model proposed to take into account its peculiarities [2,3,4,5,6]. The SIR acronym summarizes the classes of individual into which the population is partitioned. They are: *Susceptible*, individuals who can be infected, *Infected*, individuals who have been infected and that can infect susceptible ones, and *Recovered*, individuals who passed the infection phase and cannot infect other individuals any longer (by assuming they do not die).

The dynamics of an epidemic phenomenon is described by means of a system of Ordinary Differential Equations (ODEs). In its simplest formulation, the model includes one equation for each class of individuals. The population size is assumed constant over time (and normalized to 1). Hence, variables $S, I, R \in [0, 1]$ with $S + I + R = 1$ are used to describe the *ratios* of each class of individual in the population. Model parameters are the *infection coefficient* β , describing the probability of infection after a contact between a healthy individual and an infected one, and the *recovery coefficient* γ , describing the rate of recovery of each infected individual (in other words, $1/\gamma$ is the recovery time after infection).

In order to apply such a model for an investigation of the COVID-19 epidemic and, in particular, in order to analyze data collected during the first few months of the epidemic, it is necessary to take into account restriction and prevention measures (e.g. lockdown) that have been enforced by the national governments. Hence, we propose a variant of the SIR model which includes a time dependent coefficient $p(t)$ expressing the effect of such measures on the infection rate.

Once the model is defined, we estimate its parameters by using standard ODE solving and optimization methods provided by Python libraries such as SciPy and Numpy. We perform this estimation by fitting data from the ten Tuscany provinces¹, thus obtaining ten estimations for β, γ , and $p(t)$. By working at the province level, we will mitigate the assumption of the SIR model that the population is uniformly distributed on the territory, and that all individuals can freely meet with each other. Moreover, this will allow us to evaluate and compare differences in the disease spread in different provinces.

¹ freely available at <http://dati.toscana.it/dataset/open-data-covid19>.

The next step we perform is to translate our extended SIR model into a stochastic model, by discretizing variables and by considering infection and recovery rates as parameters of a Continuous Time Markov Chain (CTMC). This allows us to obtain a model that is, in principle, more accurate in capturing the epidemic dynamics, by taking into account random fluctuations that may have a significant role in the case of small numbers of infected individuals.

Dynamical properties of the obtained CTMC could then be analyzed using the stochastic model checker PRISM [7,1]. Stochastic model checking, compared for instance to analysis by stochastic simulation, allows computing in a systematic way the probability of occurrence of emerging behaviors with specific properties of interest. The main problem of model checking is, however, related to the scalability of the approach to models with a very large state space. A stochastic SIR model representing a population of hundreds of thousand of individuals (like in a Tuscan province) can be very likely affected from this kind of scalability problems.

A way to solve scalability issues can be to resort to *statistical* model checking methods: a variant of stochastic model checking which provides approximate results by exploiting stochastic simulation result. PRISM itself has built-in statistical model checking algorithms that can be exploited. However, before considering this solution, there are a few modelling tricks that can be considered to significantly reduce the state space.

In this presentation, after introducing the (extended) SIR model and the results of parameter estimation, we will describe a preliminary attempt to the analysis with PRISM, by discussing some ideas for state space reduction and by considering some dynamical properties of interest.

Acknowledgements This work is supported by the Università di Pisa under the “PRA – Progetti di Ricerca di Ateneo” (Institutional Research Grants) - Project no. PRA_2020-2021_26 “Metodi Informatici Integrati per la Biomedica”

References

1. PRISM probabilistic model checker. <https://www.prismmodelchecker.org/>
2. Acemoglu, D., et al.: A multi-risk SIR model with optimally targeted lockdown. Tech. rep., National Bureau of Economic Research (2020)
3. Calafiore, G.C., Novara, C., Possieri, C.: A modified SIR model for the COVID-19 contagion in italy. arXiv preprint arXiv:2003.14391 (2020)
4. Chen, Y.C., Lu, P.E., Chang, C.S.: A time-dependent SIR model for COVID-19. arXiv preprint arXiv:2003.00122 (2020)
5. D’Arienzo, M., Coniglio, A.: Assessment of the SARS-CoV-2 basic reproduction number, R_0 , based on the early phase of COVID-19 outbreak in italy. Biosafety and Health (2020)
6. Kermack, W.O., McKendrick, A.G.: A contribution to the mathematical theory of epidemics. Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character **115**(772), 700–721 (1927)
7. Kwiatkowska, M., Norman, G., Parker, D.: PRISM 4.0: Verification of probabilistic real-time systems. In: Proc. 23rd International Conference on Computer Aided Verification (CAV’11). LNCS, vol. 6806, pp. 585–591. Springer (2011)