A Convex Optimization Solution for the Effective Reproduction Number R_t

Journal Title
XX(X):1–12
©The Author(s) 2016
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/

SAGE

Joaquín Salas¹

Abstract

COVID-19 is a global infectious disease that has affected millions of people. With new variants emerging with augmented transmission rates, slowing down of vaccine rollouts, and rising new cases threatening sanitary capabilities to the brink of collapse, there is the need to continue studying more effective forms to track its spread. This paper presents a strategy to compute the effective reproduction number R_t . Our method starts with a form of the renewal equation of the birth process specially suitable to compute R_t . After showing that one can express it as a linear system, we proceed to solve it, along with appropriate constraints, using convex optimization. We demonstrate the method's effectiveness using synthetic and real sequences of infections and comparing it with a leading approach.

Keywords

Effective Reproduction Number, Convex Optimization, COVID-19

¹ CICATA Querétaro, Instituto Politécnico Nacional

Corresponding author:

Joaquín Salas, Instituto Politécnico Nacional, Cerro Blanco 141, Colinas del Cimatario, Querétaro, México
Email: salas@ieee.org

Introduction

In December 2019, there was an atypical number of respiratory failure cases in Wuhan, China. By January 2020, laboratories could isolate the SARS-CoV-2, a new type of betacoronavirus¹. Spread around the world, the SARS-CoV-2 has infected more than 100 million people, and more than two million people have died of COVID-19². However, due to sub registration and asymptomatic patients, the actual numbers are likely to be higher. Research has established that the mean incubation time is five days; and that the serial interval, the time between symptoms onset for the infectious and infected, is between 4.1 and 7.0 days with a 95% confidence³. COVID-19 is a highly contagious disease, with a basic reproduction number, the average number of infections caused by an infected person when all population members are susceptible, for SARS-CoV-2 is around 2.24 and 3.58⁴.

Some researchers observe that a small number of cases (10~20%) are responsible for a large number of infections (80~90%)⁵, a significant fraction through asymptomatic or pre-symptomatic transmission⁶. However, super-spreaders act in super-spreader events, *i.e.*, crowded and closed places where people may talk loudly, sing, shout to each other. These places include hospitals, leisure venues, workplaces, schools, family meetings, mass gatherings, prisons, homeless shelters, and transportation vehicles⁷. People with COVID-19 may develop symptoms that include cephalalgia, odynophagia, myalgia, arthralgia, rhinorrhea, dyspnoea, cyanosis, polypnea, anosmia, dysgeusia⁸. However, the definitive diagnosis comes after quantitative reverse transcription Polymerase Chain Reaction (RTq-PCR) tests⁹, which may take days to process. As new variants emerge with augmented transmission rates, vaccine rollouts slow down, and new cases threaten sanitary capabilities to the brink of collapse, there is the need to continue studying more effective forms to track infection spread.

During an epidemic, it is helpful to describe whether to apply lockdowns or permit certain activities. Early in the onset, when the whole population is susceptible, R_0 , the basic reproduction number, is a useful indicator of the epidemic's virulence. Later in the process, it is of more practical value the effective reproduction number R_t . R_t is the average number of infections that an infected person will cause. When R_t is one, conditions will remain the same. While with $R_t > 1$, the number of infected people

will duplicate in a short time. On the contrary, if $R_t < 1$, the infection is receding and is tending to terminate.

This paper presents a strategy to compute R_t . Our method starts with a form of the renewal equation of the birth process specially suitable to compute R_t . After showing that one can express it as a linear system, we proceed to solve it, along with appropriate constraints, using convex optimization. We demonstrate the method's effectiveness using synthetic sequences of infections and real sequences corresponding to cases in Mexico. We assess its performance using Cori *et al.*¹⁰, a leading method of widespread use.

Related Literature

Due to its importance, researchers have explored various strategies to compute the effective reproduction number, R_t ¹¹. One could classify these strategies between compartment-based, time series-based methods, and other methods.

Compartment-based Models

Bettencourt and Ribeiro¹² introduce a Bayesian scheme for the real-time estimation of the probability distribution of the effective reproduction number. Their approach develops a probabilistic formulation of the standard Suspected, Infected, Recovered (SIR) disease transmission model where the number of new cases is considered a stochastic discrete variable. At their end, Arenas *et al.*¹³ develop a method to estimate R_t based on an age-stratified compartment model. They require the transmission among individuals, specificities of a certain demographic group, and the human mobility patterns inside and among regions. Under assumptions of the infection probability, several contacts at each time step, and the fraction of susceptible individuals, they propose a Microscopic Markov Chain Approach (MMCA). Zhao *et al.*¹⁴ present a SIR vector-host compartmental model to compute R_t that is the product between R_0 and the geometric average of the susceptibles of the host and vector populations. They found their approach consistent with past epidemics. Medina-Ortiz *et al.*¹⁵ use an autoregressive integrated moving average (ARIMA) model to forecast the number of infected I , recovered R , and dead individuals. Using these predictions, they estimate R_t , based on a SIR model. Arroyo-Marioli *et al.*¹⁶ introduce a method to estimate R_t

employing the growth rate of the number of infected individuals derived from the SIR model. However, the model seems to work well with the SEIR (SIR extended with a compartment for exposed) model. Their model combines epidemiology theory and time-series techniques in the form of the Kalman filter.

De Oliveira¹⁷ introduce a model where they constraint the model variables and parameters during the estimation process. Starting with a SIRDC (susceptible-infected-resolving-deceased-recovered) model, they construct a convex quadratic optimization formulation with linear and convex quadratic constraints. Their model permits them to trade-off accuracy versus smoothness of the estimates. Cao *et al.*¹⁸ compute R_t from the daily reported cases using a combination of the average latent period, average latent infectious period, and the logarithmic growth rate of the case counts. To predict the future epidemic profile, they employ a deterministic Susceptible-Exposed-Infectious-Recovered-Death-Cumulative (SEIRDC) model. Chong¹⁹ uses the same model as Cao *et al.* by dynamically estimating the value of K . Cislighi's method²⁰ construct their formulation based on the observation that the number of cases increases by a factor of R_t every d days. Annunziato and Asikainen¹¹ extend this method solving for R_t the difference of the exponential grow equation for two extremes in an interval. The Robert Koch Institute²¹ estimates R_t as the ratio between the number of infections for two non-adjacent intervals of four days.

Time Series-based Models

Wallinga and Lipsitch²² show that different equations to estimate the observed epidemic growth rate and the reproductive number differ on their respective generation interval distribution. Furthermore, they demonstrate that using the mean of the generation intervals results in an upper bound to R_t for a given growth rate. Alvarez *et al.*²³ introduce a variational model from the incident cases and the serial interval. Using a formulation of the renewal equation, linking the incidence with R_t , they detect the initial exponential growth date. They then compute R_t by minimizing an energy function with considers the solution's ability to reproduce the incident curve, the regularity of R_t , and the adjustment of the initial value to an initial estimate of R_0 . Das²⁴ propose a method that combines time series of recent cases at the population level with the time interval between

infection of primary and secondary patients in family clusters. They feed the data to Walling and Lipsitch²². Kight and Mishra²⁵ use the incidence time series and the generation time distribution. Although researchers often use the serial interval distribution as a proxy to the generation time distribution, in an illness such as COVID-19, the infections start before the onset of symptoms. They infer the generation time distribution from parametric distributions of the serial interval, a Normal function, and the incubation period, a Gamma function, distributions.

Cori *et al.*¹⁰ introduce a method that employs a Bayesian estimation for R_t . Assuming a Gamma *prior distribution*, they employ a Poisson time of service model to generate the posterior, from where they employ Maximum Likelihood to compute R_t . Lytras *et al.*²⁶ extend Cori *et al.*¹⁰ adding imputations for missing data and adjustment for reporting delay. Menendez²⁷ shows that Cori *et al.*¹⁰ method can be reproduced effectively using the ratio between the incidence at time t and the value at half the serial interval. Na *et al.*²⁸ propose a probabilistic methodology where they predict the number of daily new infections from the daily death count. They base their quadratic programming approach on the probability distribution of delays associated with symptom onset and death reporting. They model this delay with a Gamma distribution. Bonifazi *et al.*²⁹ relate R_t to the estimation of the doubling time, *i.e.*, the time it takes to double the number of infected persons, assuming exponential growth. They obtain a closed-form solution and compute uncertainty by changing their assumption about the generation period length.

Other Approaches

Rudiger *et al.*³⁰ propose a method to estimate R_t based on the contact data extracted from mobile phones. They track de-identified cell phones' position and declare a contact when two devices are within a short distance. In their method, nodes are devices, and edges are contacts. Interestingly, they found a better correlation between the contact index, a metric for number and heterogeneity of contacts, than their quantity. They represent heterogeneity using the moments of the number of contacts distribution. Lai *et al.*³¹ use phylogeny, the change of genomes in 52 samples of COVID-19, to estimate the evolutionary rate, times of the most recent common ancestor (tMRCA), and demographic growth. They inferred the effective reproduction number employing the birth-rate

skyline model, a method that extracts the presence of rate shifts based on phylogenetic trees, and Markov chain Monte Carlo analysis.

Computing R_t

Assuming that individuals infected present a common generative interval distribution $\mathbf{w}^\top = (w_1, \dots, w_T)$, *i.e.*, the relative frequency of infections one day, two days, or s days before day t remains constant, one may express the expected number of infections $\mathbb{E}[I_t]$ observed at time t as

$$\mathbb{E}[I_t] = R_t \sum_{s=1}^T I_{t-s} w_s, \quad (1)$$

where one considers that the effective reproduction number R_t remains as if the conditions are as they were at time t , and I_t is the number of people infected at time t . The term $\mathbf{w}$ express a probability mass function and thus $\sum_{s=1}^T w_s = 1$ and $w_s \geq 0$ for $s = 1, \dots, T$. Once we know I_t , we may express (1) as

$$\sum_{s=1}^T I_{t-s} w_s - \frac{I_t}{R_t} = \begin{bmatrix} I_{t-1} & \dots & I_{t-T} & -I_t \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_T \\ \frac{1}{R_t} \end{bmatrix} = \mathbf{a}^\top \mathbf{x} = 0, \quad (2)$$

where $\mathbf{a}^\top = [I_{t-1}, \dots, I_{t-T}, -I_t]$ and $\mathbf{x}^\top = [w_1, \dots, w_T, 1/R_t]$. To find the optimal $\mathbf{x}$, one may express the dot product in (2) as a quadratic loss term such as

$$\| \mathbf{a}^\top \mathbf{x} \|^2 = (\mathbf{a}^\top \mathbf{x})^\top (\mathbf{a}^\top \mathbf{x}) = \mathbf{x}^\top \mathbf{a} \mathbf{a}^\top \mathbf{x} = \mathbf{x}^\top \mathbf{A} \mathbf{x} = 0, \quad (3)$$

where $\mathbf{A}_{(T+1) \times (T+1)}$ is a positive semi-definite symmetric matrix that has the form

$$\mathbf{A} = \begin{bmatrix} I_{t-1}^2 & I_{t-1}I_{t-2} & \dots & I_{t-1}I_{t-T} & -I_{t-1}I_t \\ I_{t-1}I_{t-2} & I_{t-2}^2 & \dots & I_{t-2}I_{t-T} & -I_{t-2}I_t \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -I_{t-1}I_t & -I_{t-2}I_t & \dots & I_tI_{t-T} & I_t^2 \end{bmatrix}. \quad (4)$$

By construction, $\mathbf{A}$ is rank one, with its largest singular value equal to $\sum_{\tau=0}^T I_{t-\tau}^2$; and nullspace with rank T , *i.e.*, T of its singular values are

equal to zero. We can now formulate the problem as a convex optimization using the standard formulation as

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \mathbf{x}^T \mathbf{A} \mathbf{x}, \text{ subject to } -\mathbf{I} \mathbf{x} \leq 0, \text{ and } \mathbf{b}^T \mathbf{x} = 1, \quad (5)$$

where $\mathbf{I}_{(T+1) \times (T+1)}$ is the identity matrix, $\mathbf{b}^T = [1, 1, \dots, 1, 0]_{1 \times (T+1)}$ is a vector of T ones followed by one zero, $t \geq T$, and $t \in \mathbb{N}$, the set of natural numbers. Note that in (5) the objective function and the inequality constraints are convex, that is, if $f_0 = \mathbf{x}^T \mathbf{A} \mathbf{x}$ and $f_1 = \mathbf{I} \mathbf{x}$, they satisfy

$$f_i(\theta \mathbf{x} + (1 - \theta) \mathbf{y}) \leq \theta f_i(\mathbf{x}) + (1 - \theta) f_i(\mathbf{y}), \text{ for } i \in \{0, 1\}, \quad (6)$$

while the equality is affine; thus it has a global minimum³².

A possible strategy to solve (5) is through convex optimization. Fu *et al.*³³ propose a novel strategy where first, they verify the cost function sign through disciplined convex programming (DCP). Then, they produce a standard form using graph implementation before finally applying a quadratic solver, such as OSQP³⁴.

Assessing Performance

To test our approach's performance to compute R_t , we relied on synthetic sequences made available by Gostic *et al.*³⁵. They constructed their series based on an SEIR model, simulating an increasing level of contagion, followed by lockdowns to terminate in some release of activities (see Fig. 1(a)). We applied our model to this sequence using and compared with Cori *et al.*¹⁰. In Fig. 1(b), we show in red the ground-truth value, in green the result of Cori *et al.*'s algorithm, and in blue the outcome of our method. We notice that after a slightly off start, with large uncertainty, Cori *et al.*'s method converges and faithfully follows the ground-truth value. Our approach follows the ground-truth value but, at inflection points, loses track and overshoots. However, it is worth observing that the optimization error remains low, with a largest value of 4×10^{-11} , signaling that the method faithfully follows the infections curve. A by-product of our approach, the generative period distribution spreads without a significant mode or bias throughout the domain (see Fig. 1(c)), *i.e.*, the contagious spreads equally over the generative interval.

We then proceed to test our method using a real sequence of confirmed positives. To that effect, we took the COVID-19 dataset made public by the

health ministry. In México, the authorities confirmed the first COVID-19 infection on February 28, 2020. Our sequence takes records until January 22, 2021 (see Fig. 1(d)). Before processing the information, we organize the sequence of infections by the onset of symptoms date and not by the date the authorities release the information. Thus, more recent dates will observe a reduced number of cases. We compare the output of our algorithm with Cori *et al.*¹⁰, in green in Fig. 1(e). We observe that our estimation tracks the spiking behavior observed in the input sequence, Cori *et al.* show a smooth development. However, as with the synthetic sequence, our method estimates an objective function error of at most 10^{-10} and a corresponding difference in estimating the next value I_t below $|2.5 \times 10^{-5}|$. Finally, we observe in Fig. 1(f) that the generative interval's data-driven estimation shows an even spread of the infection through time.

Discussion and Conclusion

As a recent study justifies³⁵, Cori *et al.*¹⁰ is regarded as a sound method to compute the effective reproduction number, R_t . Cori *et al.*¹⁰ base their method in a Bayesian scheme that permits a variety of combinations between the priors and likelihood probability distributions. Our approach is data-driven and derives directly from a widespread definition for R_t , requiring the determination of the generative period distribution length T . The optimization of our objective function results in a neglectable error, and since the resulting optimization function is convex, it has a single solution³². Other methods rely on assumptions on the nature of the priors and posteriors of R_t , the transmission process dynamics. That may result in a more pleasant and smooth solution that deviates from the reported number of infections and possibly adding its own bias to the formulation.

In this document, we introduce a novel method to compute the effective reproduction number, R_t , based on the problem's formulation as a convex optimization. Our experiments with synthetic and real data show that our method computes reliably R_t based only on the sequence of infected and the length of the generative period distribution.

Most infection reporting systems contain information delays affecting the number of cases input to the estimation of R_t . In the future, we are planning to incorporate into our method a nowcasting estimation of the number of infections with its corresponding uncertainty assessment.

Figure 1. Assessment of our method in a real sequence. Using an SEIR model, Gostic *et al.* constructed a synthetic series of infections (a). We compare the result of our method (blue) with their ground truth (red) and the output of Cori *et al.*¹⁰ (green) (b). Using the data provided by the Mexican Health Ministry, we processed the sequence of confirmed positives running from February 28, 2020, to January 22, 2021 (d). We compare the result of our method (blue) with the output of Cori *et al.*¹⁰ (green) (e). Along with the result, we obtain the generative period distribution (c) and (f) for the synthetic and real sequence, respectively. The red line highlights the mean value position through the elements of w .

Acknowledgements

Thanks to the Mexican Health Ministry for the data provided and to the CDMX SECTEI discussion group.

Author biography

Joaquín Salas is a professor in the field of Computer Vision at Instituto Politécnico Nacional. Member of the Mexican National Research System, his research interests include the monitoring of natural systems using visual perception and aerial platforms. Salas received a Ph.D. in computer science from ITESM, México. He has been a visiting scholar at Stanford University, Duke University, Oregon State University, Xerox PARC, the Computer Vision Center, and the École Nationale Supérieure des Télécommunications de Bretagne. He has served as co-chairperson of

the Mexican Conference for Patter Recognition three times. Salas was Fulbright scholar for the US State Department. He has been invited editor for Elsevier Pattern Recognition and Pattern Recognition Letters. For his services at the Instituto Politécnico Nacional, he received the *Lázaro Cárdenas* medal from the President of Mexico.

Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This work was partially funded by SIP-IPN 20210219.

Supplemental material

To foster further research, allowing other researchers to verify our results and serve as a stepping stone, we make our code publicly available at <https://github.com/joaquinsalas/computingRt>.

References

1. Zhou P, Yang XL, Wang XG et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *nature* 2020; 579(7798): 270–273.
2. Dong E, Du H and Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet infectious diseases* 2020; 20(5): 533–534.
3. Li Q, Guan X, Wu P et al. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England journal of medicine* 2020; .
4. Zhao S, Lin Q, Ran J et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International journal of infectious diseases* 2020; 92: 214–217.
5. Adam D, Wu P, Wong J et al. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nature Medicine* 2020; 26(11): 1714–1719.
6. Mizumoto K and Chowell G. Estimating risk for death from coronavirus disease, China, january–february 2020. *Emerging infectious diseases* 2020; 26(6): 1251.
7. Kamps BS and Hoffmann C. COVID reference. *Steinhäuser Verlag* 2020; .
8. Salas J, Pulido D, Montoya O et al. Data-Driven Inference of COVID-19 Clinical Prognosis. *medRxiv* 2020; .
9. Lan L, Xu D, Ye G et al. Positive RT-PCR test results in patients recovered from COVID-19. *Jama* 2020; 323(15): 1502–1503.
10. Cori A, Ferguson N, Fraser C et al. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology* 2013; 178(9): 1505–1512.

11. Annunziato A and Asikainen T. Effective Reproduction Number Estimation from Data Series. Technical report, "European Commission", 2020.
12. Bettencourt LM and Ribeiro RM. Real time bayesian estimation of the epidemic potential of emerging infectious diseases. *PloS one* 2008; 3(5): e2185.
13. Arenas A, Cota W, Gomez-Gardenes J et al. Derivation of the effective reproduction number R for COVID-19 in relation to mobility restrictions and confinement. *medRxiv* 2020; .
14. Zhao S, Musa SS, Hebert JT et al. Modelling the effective reproduction number of vector-borne diseases: the yellow fever outbreak in Luanda, Angola 2015–2016 as an example. *PeerJ* 2020; 8: e8601.
15. Medina-Ortiz D, Contreras S, Barrera-Saavedra Y et al. Country-wise forecast model for the effective reproduction number R_t of coronavirus disease. *Frontiers in Physics* 2020; 8: 304.
16. Arroyo-Marioli F, Bullano F, Kucinskas S et al. Tracking R_t of COVID-19: A new real-time estimation using the Kalman filter. *PloS one* ; 16(1): e0244474.
17. de Oliveira M. Estimating the Effective Reproduction Number and Variables of Disease Models for the COVID-19 Epidemic. *arXiv preprint arXiv:200808201* 2020; .
18. Cao Z, Zhang Q, Lu X et al. Estimating the effective reproduction number of the 2019-nCoV in China. *MedRxiv* 2020; .
19. Chong YC. A Novel Method for the Estimation of a Dynamic Effective Reproduction Number (Dynamic- R) in the CoViD-19 Outbreak. *medRxiv* 2020; .
20. Giraudo MT, Falcone M, Cadum E et al. R_t or rdt , that is the question! *Epidemiologia e Prevenzione* 2020; 44(5-6 Suppl 2): 42–50.
21. Günther F, Bender A, Katz K et al. Nowcasting the COVID-19 pandemic in Bavaria. *Biometrical Journal* 2020; .
22. Wallinga J and Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences* 2007; 274(1609): 599–604.
23. Alvarez L, Colom M and Morel JM. A variational model for computing the effective reproduction number of SARS-CoV-2. *medRxiv* 2020; .
24. Das A. An approximation-based approach for periodic estimation of effective reproduction number: a tool for decision-making in the context of coronavirus disease 2019 (COVID-19) outbreak. *Public Health* 2020; 185: 199–201.
25. Knight J and Mishra S. Estimating effective reproduction number using generation time versus serial interval, with application to COVID-19 in the Greater Toronto Area, Canada. *Infectious Disease Modelling* 2020; 5: 889–896.
26. Lytras T, Sypsa V, Panagiotakos D et al. An improved method to estimate the effective reproduction number of the COVID-19 pandemic: lessons from its application in Greece. *medRxiv* 2020; .
27. Menendez J. A poor-man's approach to the effective reproduction number: the COVID-19 case. *medRxiv* 2020; .
28. Na J, Tibebe H, De Silva V et al. Probabilistic approximation of effective reproduction number of COVID-19 using daily death statistics. *Chaos, Solitons & Fractals* 2020; 140: 110181.
29. Bonifazi G, Lista L, Menasce D et al. A simplified estimate of the Effective Reproduction Number R_t using its relation with the doubling time and application to Italian COVID-19 data. *arXiv preprint arXiv:201205194* 2020; .

30. Rüdiger S, Konigorski S, Edelman J et al. Forecasting the SARS-CoV-2 effective reproduction number using bulk contact data from mobile phones. *medRxiv* 2020; .
31. Lai A, Bergna A, Acciarri C et al. Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2. *Journal of medical virology* 2020; 92(6): 675–679.
32. Boyd S and Vandenberghe L. *Convex optimization*. Cambridge university press, 2004.
33. Fu A, Narasimhan B and Boyd S. CVXR: An R package for disciplined convex optimization. *Journal of Statistical Software* 2020; 94(14): 1–34. DOI:10.18637/jss.v094.i14.
34. Stellato B, Banjac G, Goulart P et al. OSQP: An operator splitting solver for quadratic programs. *Mathematical Programming Computation* 2020; : 1–36.
35. Gostic K, McGough L, Baskerville E et al. Practical considerations for measuring the effective reproductive number, R_t . *PLoS Computational Biology* 2020; 16(12): e1008409.