

Estimation of the basic reproduction number of COVID-19 from the incubation period distribution

Lasko Basnarkov^{1,2}, Igor Tomovski², and Florin Avram³

¹SS. Cyril and Methodius University, Faculty of Computer Science and Engineering, P.O. Box 393, 1000 Skopje, Macedonia

²Macedonian Academy of Sciences and Arts, P.O. Box 428, 1000 Skopje, Macedonia

³Laboratoire de Mathématiques Appliquées, Université de Pau, 64000, France

November 4, 2021

Abstract

Background

The estimates of future course of spreading of the SARS-CoV-2 virus are frequently based on Markovian models in which the transitions between the compartments are exponentially distributed. Specifically, the basic reproduction number R_0 is also determined from formulae where it is related to the parameters of such models. The observations show that the start of infectivity of an individual appears nearly at the same time as the onset of symptoms, while the distribution of the incubation period is not an exponential.

Methods

We propose a method for estimation of R_0 for COVID-19 based on the empirical incubation period distribution and assumed very short infectivity period that lasts only few days around the onset of symptoms. It is tested on daily new cases in six major countries in Europe, in the first wave of epidemic in spring, 2020.

Results

The calculations show that even if the infectivity starts two days before the onset of symptoms and stops immediately when they appear, the value of R_0 is larger than that from the classical, Markovian approach. For more realistic cases, when only individuals with mild symptoms spread the virus for few days after onset of symptoms, the respective values are even larger.

Conclusions

The calculations of R_0 and other characteristics of spreading of COVID-19 based on the classical, Markovian approaches should be taken very cautiously. Instead, non-Markovian models with general distribution functions of transition between compartments should be considered as more appropriate.

Key messages

- Although formulae for estimate of the basic reproduction number R_0 , by using general-form functions of infectivity are known since the earliest works in epidemiology, majority of studies are based on exponential distribution function .
- We introduce a new methodology of calculating R_0 with an infectivity function obtained by combining empirical incubation period distribution with infectivity window function that is localized around the onset of symptoms.
- Estimates of R_0 for the first wave of COVID - 19 in the spring 2020, by the proposed methodology are larger than those from the classical SIR model.
- When possible, the estimates of R_0 should be based on empirical distributions of the infectivity functions, while the values obtained with the conventional epidemic spreading models should be taken with caution.

1 Introduction

The mathematical epidemiology is a field of study with increasing public importance by providing valuable insights for the authorities aimed for planning various actions against epidemic diseases. Although its inception dates back to Daniel Bernoulli [1], the major development as separate field of study might be attributed to the works of Ross [2–4] and Kermack and McKendrick [5]. The key quantity that is usually reported from studies about epidemic spreading is the basic reproduction number (or rate, or ratio) R_0 . Its importance is particularly emphasized with the fact that the herd immunity, achieved naturally or through vaccination expressed as fractions of the total population depends on it as $1 - 1/R_0$ [6]. The basic reproduction ratio represents an estimated number of newly infected persons by one infected individual introduced in a completely susceptible population. In its calculations, even in the earliest works it was considered that the infectivity potential of the infector depends on the time passed since she or he become infected - the age of infection, adequately modeled with certain function with general form. However, in majority of works the dominant role has exponential function e^{-t} . This choice provides simple relationships for the basic reproduction ratio, but also makes the epidemic spreading models to have simple form based on ordinary differential equations. This exponential function in the probabilistic approach has a meaning of probability density function of the period in which individuals stay in given compartment (exposed, infected, and so on), before transiting to another one. This is the appropriate function when the Markovian property holds, which means that the probability for transition to another state is independent on the past. However, very often this assumption is not empirically verified, and particularly for the diseases like the COVID-19, which can have particularly long period of incubation [7–9]. This practically means that the spreading potential of an infected individual becomes significant only when the incubation period is near its end. As a consequence, more appropriate, and more accurate estimates would be obtained by non-Markovian approach which allows for using arbitrary distribution functions of the incubation period duration.

Since the COVID-19 has become serious threat of the health and seriously disrupted the normal life, large number of studies have been made for modeling its spread and particularly about estimates of the basic reproduction ratio. It is calculated on a daily basis by various institutions and individuals. We have an impression that majority of the methods for its estimate are based on relationships that result from compartmental models with ordinary differential equations, that is, with the Markovian setting and the related exponential distribution of incubation period. It is thus our motivation to provide an estimate of R_0 with an approach that is based on the non-exponential distribution of the incubation period, in a non-Markovian framework, which we believe is particularly relevant for the spread of the SARS-CoV-2. A basis of our calculations are the empirically obtained distributions of the time to onset of symptoms. Such distributions are further combined with certain window function that assumes that an individual is infectious only few days around the onset of symptoms. The combination is an infectivity function that determines the basic reproduction number. The obtained results for different European countries in the first wave of the epidemic, when the virus was assumed to spread freely, are larger than those with the classical, Markovian approach. The key outcome of these observations is that the estimates of R_0 with the classical models should be considered very carefully. Finally, it is worth noting that the method used here has also similarities with the early studies in demography by Richard Böckh [10] and in epidemiology by Alfred Lotka [11].

The paper is organized as follows. In the Section 2 we first explain the formula used for estimation of R_0 . In the following Section 3 it is elaborated on the infectivity function that is used for estimates of R_0 . In the next, Section 4 are presented the results and a discussion on them and we finish the paper with the Conclusions.

2 Methods

The derivation of the formula that we use for estimation of the basic reproduction rate R_0 for general infectivity function can be found in various works in the literature (for example in [12–14]). For completeness we present it here and chose to use a discrete-time approach since it is more appropriate for the available data. To start with, assume that the epidemic is in inception phase which means that the number of newly infected individuals grows exponentially. As is the case of COVID-19, when daily new cases are reported, by denoting with $I_d(t)$ the confirmed cases for day t given as fraction of the total population, one has the following exponential form

$$I_d(t) = I_0 e^{\lambda t}, \quad (1)$$

where λ is a parameter for the growth rate, while I_0 is a constant. The growth rate λ is related with the period of doubling with $T_d = \ln 2 / \lambda$. For many diseases the individuals are not able to infect the others immediately, but

after certain period has passed. We consider that the infectivity potential of the individuals that have contracted the spreading agent at the same moment t , is described with certain infectivity function $i(\tau)$. We assume that this function depends on the time elapsed since the individuals have become exposed, but not on the moment when that happened and has finite support T which means that $i(\tau) = 0$, for $\tau > T$. The shape of this function will be elaborated later on. Denote with $I(t, \tau)$ the fraction of infectees that have become infected at moment t , by having a contact with infectors that have become infected τ time units earlier. The function $I(t, \tau)$ depends on the fraction of susceptibles, but also on the fraction of those previously infected $I_d(t - \tau)$ and on their infectivity potential encoded in $i(\tau)$. Thus one has the following form

$$I(t, \tau) = R_0 S(t) I_d(t - \tau) i(\tau), \quad (2)$$

where the constant of proportionality R_0 will be shown to be exactly the basic reproduction ratio. The fraction of the new infectees will be obtained as sum of contributions from all possible infectors

$$I_d(t) = \sum_{\tau=1}^T I(t, \tau), \quad (3)$$

which will further result in the following recurrent relationship

$$I_d(t) = R_0 S(t) \sum_{\tau=1}^T I_d(t - \tau) i(\tau). \quad (4)$$

By applying the exponential form for of the function of newly infected individuals (1), in the last relationship, one will have

$$I_0 e^{\lambda t} = I_0 R_0 S(t) e^{\lambda t} \sum_{\tau=1}^T e^{-\lambda \tau} i(\tau), \quad (5)$$

from where it follows that

$$1 = R_0 S(t) \sum_{\tau=1}^T e^{-\lambda \tau} i(\tau), \quad (6)$$

Since at the inception of the epidemic $S(t) \approx 1$, one has the reproduction number

$$R_0 = \frac{1}{\sum_{\tau=1}^T e^{-\lambda \tau} i(\tau)}. \quad (7)$$

If one has obtained R_0 from the last expression and it is assumed to be the same during the epidemic, than the herd immunity threshold S_{th} , which corresponds to unity reproduction number R , can be obtained from (6) as

$$S_{th} = \frac{1}{R_0}. \quad (8)$$

2.1 Self-consistency relationship for R_0

The basic reproduction number R_0 represents an estimated multiplicative factor determining the number of newly infected individuals that will contract the spreading agent in a contact with certain infected individual. Here we use it to denote the growth factor that corresponds to the group of individuals that have been exposed in the same period. Thus, from one side, the fraction of all infectees that have contracted the pathogen from infectors that become infected at certain moment t should be $R_0 I_d(t)$. From another side, this fraction can be represented as forward-time sum as follows

$$R_0 I_d(t) = \sum_{\tau=1}^T I(t + \tau, \tau), \quad (9)$$

which accounts for all admissible combinations $I(t + \tau, \tau)$. By using (2), one has

$$R_0 I_d(t) = R_0 \sum_{\tau=1}^T S(t + \tau) I_d(t) i(\tau). \quad (10)$$

For slowly growing epidemic one might assume that the fraction of susceptibles does not change significantly in the considered period $S(t) \approx 1$. Then, one can see that the constant R_0 in the relationship (2) will be the basic reproduction number, if the infectivity function is normalized $\sum_{\tau=1}^T i(\tau) = 1$.

2.2 Continuous-time formula

Although analysis of the data in this work is based on discrete time, for completeness we provide the continuous-time version, that can be found in the literature in similar forms. Denote with $I(t)$ the fraction of the population that has become infected within infinitesimal interval $(t, t + dt)$, and assume that it grows exponentially at the onset of epidemics $I(t) = I_0 e^{\lambda t}$. These infectees have appeared from contacts with others that have been infected in the past $I(t - \tau)$. Again, let the function $R_0 i(\tau)$ denote the infectivity potential at some later moment τ of the group of individuals infected nearly at the same time. As for the discrete-time case assume that $i(\tau)$ has some finite support $(0, T)$. Now, the formula for newly infected population will read

$$I(t) = R_0 S(t) \int_0^T I(t - \tau) i(\tau) d\tau, \quad t \geq T. \quad (11)$$

One should note that similar relationship has appeared earlier in the works of Ross and Hudson [15]. By plugging in the exponential form of the newly infected individuals one will obtain

$$I_0 e^{\lambda t} = R_0 S(t) \int_0^T I_0 i^{\lambda(t-\tau)} i(\tau) d\tau. \quad (12)$$

which will reduce to similar relationship as the discrete-time one

$$1 = R_0 S(t) \int_0^T e^{-\lambda \tau} i(\tau) d\tau. \quad (13)$$

At the onset of an epidemic $S(t) \approx 1$, the reproduction number will be

$$R_0 = \frac{1}{\int_0^T e^{-\lambda \tau} i(\tau) d\tau}. \quad (14)$$

In reality, one does not have observations of the short-interval infections $I(t)$, but for certain unit interval, like a day. So, one has to choose a function $I(t)$ such that its daily integral $\int_t^{t+1} I(t) dt$ would fit the observed data.

3 Shape of the infectivity function

The function $i(\tau)$ should model the infectivity potential of the fraction of the population, becoming infected within the same unit interval, at some later moment τ after contracting the spreading agent. Its shape can be directly deduced from epidemiological tracing of infector-infectee pairs. Our approach is to relate it with other characteristics of the respective disease. First, its shape depends on the fractions of the individuals that have not been recovered yet and thus is related to the function that describes the healing process. However, since the COVID-19 is a disease which has long period of recovery we consider that the function modeling it can be considered as constant for the period under study. Second, the infectivity function depends on the incubation period distribution, for which one can find many studies in the literature. Third, the viral load in reality is not constant over time, which means that during infectiousness, a person cannot spread the pathogen with equal intensity [16]. For simplicity in this work it is also considered to be constant. As a side note we mention here that the product of the healing and incubation period functions was applied in non-Markovian epidemic spreading models recently [17, 18], while the same relationship with other functions was used in the pioneering work of demography [10]. Thus, in this work we are focusing on using the distribution of the time to onset of symptoms as basis. To reach the final shape of the infectivity function we combine it with an infectivity window function $w(t)$. It is non-zero in the period $t \in [t_{\text{init}}; t_{\text{end}}]$, between certain initial moment t_{init} before the onset of symptoms, and ends at t_{end} after. From practical perspective, the shape of the window in the segment $[t \in t_{\text{init}}, 0]$, with $t = 0$ being the time of onset of symptoms, may be obtained from sensitivity analysis conducted on known infector-infectee pairs, as suggested in [16]. The shape of $w(t)$ for $t \in [0; t_{\text{end}}]$ represents the distribution of

Country	Start date	End date	λ
Germany	February 25	March 12	0.298
France	February 25	March 11	0.322
Italy	February 21	March 9	0.235
Russia	March 12	March 30	0.171
Spain	February 25	March 11	0.343
United Kingdom	February 23	March 22	0.237

Table 1: Countries under study and the estimated exponential growth factor

time from onset of symptoms to self isolation, quarantining or viral clearance, obtained from epidemiological data. In our approach we assume that after the onset of symptoms, a person with more severe symptoms would immediately reduce the contacts with the others and thus would not be a significant infector any more. Those with mild symptoms, that constitute about 80% of the cases [19], would continue with their normal daily life and would spread the virus until t_{end} , when they receive positive test results. Thus the infectivity window function has a shape of two steps

$$w(t) = \begin{cases} 1, & t_{\text{init}} \leq t < 0 \\ 0.8, & 0 \leq t \leq t_{\text{end}} \end{cases}. \quad (15)$$

The infectivity function was considered to be the convolution of the incubation period distribution $\beta(\tau)$ and the infectivity window $w(\tau)$

$$i(\tau) = C(\beta * w)(\tau) = C \sum_{\nu} \beta(\tau - \nu)w(\nu), \quad (16)$$

where C is a constant determined from the normalization condition $\sum_{\tau=1}^T i(\tau) = 1$.

We compare our calculations with those obtained from the classical (Markovian) SIR model where the basic reproduction ratio with the growth rate λ is related with [13]

$$R_0 = 1 + \lambda/\beta = 1 + \lambda T_c, \quad (17)$$

where β is the infectivity rate, and T_c is the mean infectivity period – the average value of the duration of the exponentially distributed infective state.

4 Results and discussion

We have used the data from the Our World in Data database for the numbers of daily cases in the first wave in the spring 2020 for six major European countries. As a data window for study we have taken the period from the first day when every day new cases were reported until the moment when lockdown measures were introduced. This choice was made under the assumption that in that period the virus SARS-CoV-2 was spreading nearly freely in the population, with only positive cases being isolated. From those numbers we have fit an exponential curve and estimated the exponential growth rate factor λ . The results are summarized in the table 1.

We have used the estimates of the growth factor λ in the expression for the calculation of the basic reproduction number (7). The infectivity function $i(\tau)$ was obtained from the convolution (16) of windows with variable width, and incubation period functions from two sources in the literature [7, 8].

The onset of infectiousness was considered to be two [16] or one day before the appearance of symptoms. We have also assumed that after the onset of symptoms only those with mild symptoms, that are about 80% [19] can be considered as further infectors. As the end of the infectiousness period was considered either the same day or two days after the onset of the symptoms. The first case is very conservative and is related to the assumption that all individuals would be very cautious in their contacts with the others and are not infectors at all. The second one is less conservative and likely a more realistic scenario, where those with mild symptoms do not change their daily routines and spread the virus freely before receiving positive test. Finally, the results from the classical SIR model are also provided for comparison. Details about the considered scenarios is given in table 2.

Notation	Description	Function	Parameters
S1	Start two days before onset, stop immediately	Log-normal ^a [7]	$P_1 = 1.62; P_2 = 0.42$
S2	Start two days before onset, stop two days after	Log-normal [7]	$P_1 = 1.62; P_2 = 0.42$
S3	Start one day before onset, stop two days after	Log-normal [7]	$P_1 = 1.62; P_2 = 0.42$
S4	Classical relationship	Log-normal [7]	$P_1 = 1.62; P_2 = 0.42$
S5	Start two days before onset, stop immediately	Weibull ^b [8]	$P_1 = 2.04; P_2 = 0.103$
S6	Start two days before onset, stop two days after	Weibull [8]	$P_1 = 2.04; P_2 = 0.103$
S7	Start one day before onset, stop two days after	Weibull [8]	$P_1 = 2.04; P_2 = 0.103$
S8	Classical relationship	Weibull [8]	$P_1 = 2.04; P_2 = 0.103$

Table 2: Description of different scenarios for calculation of basic reproduction ratio

^aFor the log-normal distribution, P_1 is the mean and P_2 is the standard deviation of the logarithm of the distribution.

^bFor the Weibull distribution P_1 is the scale, while P_2 is the shape parameter.

Country	S1	S2	S3	S4	S5	S6	S7	S8
France	3.22	4.06	5.02	2.87	5.37	6.72	8.01	3.50
Germany	3.00	3.73	4.53	2.73	4.93	6.10	7.17	3.31
Italy	2.45	2.95	3.42	2.37	3.85	4.60	5.22	2.83
Russia	1.97	2.27	2.51	1.99	2.87	3.29	3.60	2.33
Spain	3.43	4.38	5.49	2.99	5.78	7.31	8.81	3.66
UK	2.47	2.97	3.45	2.38	3.89	4.65	5.28	2.84

Table 3: Estimated basic reproduction ratio R_0 for six European countries for the eight different scenarios.

The estimates of R_0 for the different scenarios for all six European countries are presented in table 3. One can easily note that in all scenarios the values of R_0 from the non-Markovian approach considered here are larger than those from the classical, Markovian one. In the overly restricted scenarios S1 and S5 when onset of symptoms, even for those with mild symptoms, is a reason for immediate self-quarantine leads to larger estimates than the reference cases S4, and S8, respectively. The more realistic assumptions in the other cases produce even larger estimates. In the reality there would have been cases for even longer period of infectivity than two days after the onset of symptoms, and this could be the true for significant fraction of infected population. Although we have not tried to model this, the logic suggests that the longer the infectivity period, after the onset of symptoms, the larger R_0 is. This implies that the values of R_0 based on the classical Markovian SIR model for the COVID-19 are likely significant underestimates. From this observation one has that the more accurate estimate of R_0 needs knowledge of the infectivity window function $w(\tau)$, that is the period when an infected person is infectious, in relation to the onset of symptoms. The situation can be even more complicated if the infectivity window function is dependent on the incubation period duration. It could be possible for those that develop symptoms earlier to correspond one infectivity window function, while for those that develop them after longer period another one. Thus, besides being very cautious with using the estimated R_0 from the classical approach, it is needed to have good estimate of the infectivity function in order to have more correct value of the basic reproduction ratio.

5 Conclusions

We have estimated the basic reproduction ratio R_0 by using the more general non-Markovian framework, that besides being known from the emergence of mathematical epidemiology, has not been widely applied. The approach was used to determine the value of R_0 in six major countries in Europe during the first wave of the COVID-19 epidemic. The onset of infectiousness, instead of starting immediately after contraction of the pathogen, was taken to be related to the onset of symptoms, for which the empirical evidence suggests that is not distributed exponentially as the Markovian assumption implies. The incubation period distribution was further combined with an infectiousness window function

which was considered to have short period - one or two days before and finish up to two days after the onset of symptoms. From both functions we have constructed an infectivity function that uniquely determines R_0 . In all scenarios we have considered, the calculated value for R_0 was obtained to exceed the one from the classical relationship. This suggests that the calculations with the classical, Markovian approach, should be taken rather cautiously.

Better estimates of R_0 would be obtained with empirical function of the infectiousness window, or direct estimation of the shape of the infectivity function. This needs more involving epidemiological tracing, that is not an easy task. However, we hope that the observation that the classically obtained value of R_0 is an underestimate would lead to more intensive work for gathering epidemiological data and increase the awareness that non-Markovian setting in the epidemic models should be given more attention.

6 Competing interests

There is NO Competing Interest.

7 Author contributions statement

L.B. and I.T. initiated the study of non-Markovian approach for estimation of the basic reproduction number, formulated the basic model and made the numerical calculations. L.B., I.T. and F.A. contributed to the final model and the overall form of the manuscript.

8 Acknowledgement

This research was partially supported by the Faculty of Computer Science and Engineering, at the Ss. Cyril and Methodius University in Skopje, Macedonia.

References

- [1] K. Dietz and J. Heesterbeek, “Daniel bernoulli’s epidemiological model revisited,” *Mathematical biosciences*, vol. 180, no. 1-2, pp. 1–21, 2002.
- [2] R. Ross, “An application of the theory of probabilities to the study of a priori pathometry.—part i,” *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, vol. 92, no. 638, pp. 204–230, 1916.
- [3] R. Ross and H. P. Hudson, “An application of the theory of probabilities to the study of a priori pathometry.—part ii,” *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, vol. 93, no. 650, pp. 212–225, 1917.
- [4] R. Ross and H. P. Hudson, “An application of the theory of probabilities to the study of a priori pathometry.—part iii,” *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, vol. 93, no. 650, pp. 225–240, 1917.
- [5] W. O. Kermack and A. G. McKendrick, “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*, vol. 115, no. 772, pp. 700–721, 1927.
- [6] P. Fine, K. Eames, and D. L. Heymann, ““herd immunity”: a rough guide,” *Clinical infectious diseases*, vol. 52, no. 7, pp. 911–916, 2011.
- [7] S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, and J. Lessler, “The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application,” *Annals of internal medicine*, vol. 172, no. 9, pp. 577–582, 2020.

- [8] J. Qin, C. You, Q. Lin, T. Hu, S. Yu, and X.-H. Zhou, “Estimation of incubation period distribution of covid-19 using disease onset forward time: a novel cross-sectional and forward follow-up study,” *Science advances*, vol. 6, no. 33, p. eabc1202, 2020.
- [9] C. McAloon, Á. Collins, K. Hunt, A. Barber, A. W. Byrne, F. Butler, M. Casey, J. Griffin, E. Lane, D. McEvoy, *et al.*, “Incubation period of covid-19: a rapid systematic review and meta-analysis of observational research,” *BMJ open*, vol. 10, no. 8, p. e039652, 2020.
- [10] R. Böckh, *Statistisches Jahrbuch der Stadt Berlin*. Leonh. Simion, 1877.
- [11] A. J. Lotka, “A contribution to quantitative epidemiology,” *Journal of the Washington Academy of Sciences*, vol. 9, no. 3, pp. 73–77, 1919.
- [12] O. Diekmann, J. A. P. Heesterbeek, and J. A. Metz, “The legacy of kermack and mckendrick,” *Epidemic models: their structure and relation to data*, vol. 5, p. 95, 1995.
- [13] J. Wallinga and M. Lipsitch, “How generation intervals shape the relationship between growth rates and reproductive numbers,” *Proceedings of the Royal Society B: Biological Sciences*, vol. 274, no. 1609, pp. 599–604, 2007.
- [14] F. Brauer, “Mathematical epidemiology: Past, present, and future,” *Infectious Disease Modelling*, vol. 2, no. 2, pp. 113–127, 2017.
- [15] J. A. P. Heesterbeek, “A brief history of r_0 and a recipe for its calculation,” *Acta biotheoretica*, vol. 50, no. 3, pp. 189–204, 2002.
- [16] X. He, E. H. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, *et al.*, “Temporal dynamics in viral shedding and transmissibility of covid-19,” *Nature medicine*, vol. 26, no. 5, pp. 672–675, 2020.
- [17] I. Tomovski, L. Basnarkov, and A. Abazi, “Discrete-time non-markovian seis model on complex networks,” *IEEE Trans. Netw. Sci. Eng. (accepted)*, 2021.
- [18] L. Basnarkov, I. Tomovski, T. Sandev, and L. Kocarev, “Non-markovian sir epidemic spreading model,” *arXiv preprint arXiv:2107.07427*, 2021.
- [19] Z. Wu and J. M. McGoogan, “Characteristics of and important lessons from the coronavirus disease 2019 (covid-19) outbreak in china: summary of a report of 72 314 cases from the chinese center for disease control and prevention,” *Jama*, vol. 323, no. 13, pp. 1239–1242, 2020.