Monitoring the spread of COVID-19 by estimating reproduction numbers over time

Thomas Hotz[1], Matthias Glock[1], Stefan Heyder[1], Sebastian Semper[1], Alexander Krämer[2], Anne Böhle[2]

[1] Institut für Mathematik, Technische Universität Ilmenau
{thomas.hotz,matthias.glock,stefan.heyder,sebastian.semper}@tu-ilmenau.de

[2] School of Public Health, Bielefeld University
{alexander.kraemer,anne.boehle}@uni-bielefeld.de

06/04/20 – 10:24 GMT

Contents

1 Introduction

As the Coronavirus Disease 2019 (COVID-19) threatens humanity, unprecedented measures to stop its spread have been adopted around the globe. In many countries, schools have closed and curfews have been imposed. Given the enormous burden these measures place on the economy, sooner or later they have to be relaxed. This raises important questions for policymakers and public health specialists. How large is the effect of these measures? Do they effectively stop the spread of COVID-19? What will happen if restrictions get relaxed? And in the future, how can we see whether the epidemic is getting out of hands again?

To answer these questions, one needs to know how fast the epidemic is growing. In epidemiology, this is measured by the **reproduction number**, i.e. the mean number of people someone who got infected will infect in the course of time. Its **critical value** clearly is 11: for larger values the epidemic will grow, for smaller values it will diminish.

Since conditions may change in the future, e.g. when countermeasures are introduced or lifted, this may change. We therefore follow Fraser (2007) and consider what he calls the *instantaneous* reproduction number $R(t)$ at time t, and for which he suggests the estimator

$$\hat{R}(t) = \frac{I(t)}{\sum_{\tau=1}^{\infty} w(\tau) I(t-\tau)} \tag{1}$$

where $I(t)$ is the number of incident cases at time t and w specifies the so-called **infectivty profile** which is assumed to be known. To the best of our knowledge, this estimator has first been published by Fraser and others in (Grassly et al.; 2006).

We explain the probabilistic model behind this estimator following (Cori et al.; 2013, Web Appendix 1) in Section 2. In addition, we analytically derive asymptotic confidence intervals (with details given in Appendix A) which are simple to compute. Here, we differ from Grassly et al. (2006) who use computationally more elaborate resampling techniques, namely the bootstrap, to obtain confidence intervals; Cori et al. (2013) on the other hand take a Bayesian approach, assuming a certain gamma prior distribution for $R(t)$.

The estimator and corresponding confidence intervals are validated on simulated data in Section 4. Then, we apply this methodology to real data from Germany in Section 5, followed by a sensitivity analysis in Section 6. Finally, the results are summarised in Section 7, also discussing difficulties with this approach.

In order to continuously monitor the spread of COVID-19, a designated website has been created where the results of our analysis are shown and updated daily. It is available at `https://stochastik-tu-ilmenau.github.io/COVID-19/` in English for all affected countries based on the data from (Johns Hopkins University Center for Systems Science and Engineering; 2020) as well as in German for Germany and its federal states based on the data from (Robert Koch-Institut; 2020) at `https://stochastik-tu-ilmenau.github.io/COVID-19/germany.html`. We note that a similar analysis using the Bayesian approach of (Cori et al.; 2013) was presented by Abbott et al. (2020). However, as of 06/04/2020, the analysis appears not to have been updated since 19/03/2020.

2 Derivation of the estimator

The following is an adaptation of the modelling in (Fraser; 2007) and (Cori et al.; 2013, Web Appendix 1).

Time is taken to be discrete, i.e. we consider days $t \in \mathbb{Z}$, since the spread of the epidemic shows a strong intraday variability (e.g., there are fewer infections during the night when people are at sleep), and the time scales of incubation and infectious period are on the order of days. Also, cases are reported on a daily basis.

The number of number of **incidences**, i.e. newly infected cases, at day t will be given as $I(t)$. The **infection age** of an infected person in days, i.e. the number of day elapsed since the infection, is denoted by $\tau \in \mathbb{N}_0$.

The spread of the epidemic depends strongly on the time-dependent **transmissibility** $\beta(t, \tau) \geq 0$ specifying the expected number of susceptibles an infectious person at infection age τ, a so-called **primary case**, will infect at time t. The transmissibility is in particular affected by the **contact rate**, i.e. the mean number of people an infected person meets per day, and the **infectiousness** of the primary case. The former is addressed by **non-pharmaceutical interventions** such as school closures and curfews, the latter is a virological feature of the disease. Therefore we make a crucial **structural assumption**, namely that they separate:

$$\beta(t, \tau) = R(t)\, w(\tau) \tag{2}$$

where $R(t) \geq 0$ denotes the (instantaneous) **reproduction number** at time t of transmission, i.e. when the **secondary case** gets infected by the primary case, and $w(\tau) \in [0, 1]$ specifies the **infectivity profile** at infection age τ. This models the belief that contact rates change over time but the infectiousness of the primary case depends only on τ. This is debatable: when rules for isolation or quarantine change are loosend, e.g. because hospital capacities are exhausted, β will change differently for different values of τ; we will reiterate this point in Section 7. It also shows

in the fact that any constant factor may be alternatively incorporated into R or w. The latter is therefore standardised such that

$$\sum_{\tau=0}^{\infty} w(\tau) = 1\,, \tag{3}$$

i.e. w is a **probability distribution** which can be interpreted as follows: for a fixed time t randomly pick a pair of individuals where the first one is a primary case that got infected at time t, in turn infecting the second one later; $w(\tau)$ is the probability that the second case got infected at time $t+\tau$, i.e. at infection age τ of the primary case. w is assumed to be **known**; see Section 3 on how we model it for COVID-19.

In a **stochastic model** for the dynamics of the epidemic, $I(t)$ is given as the number of successful transmissions from an infectious person to someone who is susceptible to the disease. Assuming that for each possible transmission succeeds independently (thus ignoring the possibility of multiple infections) with a probability corresponding to β, and if there are many possible transmissions, $I(t)$ will be (approximately) **Poisson distributed** conditional on the past with intensity equal to

$$\mathbf{E}(I(t)\,|\,I(t-1),\dots) = \sum_{\tau=1}^{\infty} \beta(t,\tau) I(t-\tau) = R(t) \sum_{\tau=1}^{\infty} w(\tau) I(t-\tau)\,. \tag{4}$$

Here, transmissions **on the same day** are ruled out, i.e. $w(0) = 0$, which is a realistic assumption since the incubation period will be at least one day.

The last equation suggests the **estimator** $\hat{R}(t)$ for $R(t)$ given in (Fraser; 2007, Equation (9)),

$$\hat{R}(t) = \frac{I(t)}{\sum_{\tau=1}^{\infty} w(\tau) I(t-\tau)}\,. \tag{5}$$

Note that the **case reproduction number** $R_c(t)$, i.e. the expected number of people a primary case infected at time t will infect, is given by, cf. (Fraser; 2007, Equations (2) and (8)),

$$R_c(t) = \sum_{\tau=1}^{\infty} \beta(t+\tau,\tau) = \sum_{\tau=1}^{\infty} R(t+\tau)\, w(\tau)\,. \tag{6}$$

This is of course difficult or even impossible to estimate as it depends on future contact rates, i.e. on countermeasures that will be imposed. However, assuming that conditions remain the same in the future, i.e. $R(s) = R(t)$ for $s > t$, we obtain $R(t)$ again, cf. (Fraser; 2007, Equation (3)),

$$\sum_{\tau=1}^{\infty} R(t+\tau) w(\tau) = R(t) \sum_{\tau=1}^{\infty} w(\tau) = R(t)\,. \tag{7}$$

This explains why $R(t)$ is called *reproduction number*.

For large intensities, i.e. if the conditional expectation in Equation (4) is large, the distribution of $\hat{R}(t)$ can be well approximated by a Gaussian distribution, with small standard errors. Thus using **asymptotic confidence intervals** can be derived, see Section A. If q denotes the $(1-\frac{\alpha}{2})$-quantile of the standard normal distribution then

$$\left[\hat{R}(t) - q\sqrt{\frac{\hat{R}(t)}{\sum_{\tau=1}^{\infty} w(\tau) I(t-\tau)}}, \hat{R}(t) + q\sqrt{\frac{\hat{R}(t)}{\sum_{\tau=1}^{\infty} w(\tau) I(t-\tau)}}\right] \tag{8}$$

is an (asymptotic) $(1-\alpha)$-confidence interval for $R(t)$. Note that in practice 10 or more incident cases should suffice for the asymptotics to be reliable.

3 Specifics of COVID-19

$w(0)$

infectivity profile vs amount of infectious material

4 Validation on simulated data

5 Application to real data

6 Sensitivity analysis

7 Discussion and Outlook

incubation period

structural assumption, isolation, medical treatment

infectivity profile vs amount of infectious material

weekday effects

reporting chain

infected vs reported

mean, individual, stratification, regional effects get averaged

imported cases

Acknowledgements. The authors thank Dr. med. Luise Prüfer-Krämer, Steering Committee Member, German Society of Tropical Medicine and Global Health and practising physician, for many fruitful discussions and insights into the care of COVID-19 patients.

A Derivation of confidence intervals

Starting from Equation (4), the conditional expectation of $\hat{R}(t)$ given the past is

$$\mathbf{E}(\hat{R}(t) \,|\, I(t-1), \dots) = \frac{\mathbb{E}(I(t)|I(t-1), \dots)}{\sum_{\tau=1}^{\infty} w(\tau) I(t-\tau)} = R(t)\,. \tag{9}$$

Therefore, $\hat{R}(t)$ is **unbiased**,

$$\mathbf{E}\,\hat{R}(t) = R(t)\,, \tag{10}$$

and the **conditional variance** of $\hat{R}(t)$ is given by

$$\mathbf{Var}(\hat{R}(t) \,|\, I(t-1), \dots) = \frac{R(t)}{\sum_{\tau=1}^{\infty} w(\tau) I(t-\tau)}\,. \tag{11}$$

An application of Slutsky's lemma gives an asymptotic $(1-\alpha)$-confidence interval for $R(t)$: if q denotes the $(1-\frac{\alpha}{2})$-quantile of the standard normal distribution it is given by

$$\left[\hat{R}(t) - q\sqrt{\frac{\hat{R}(t)}{\sum_{\tau=1}^{\infty} w(\tau) I(t-\tau)}}, \hat{R}(t) + q\sqrt{\frac{\hat{R}(t)}{\sum_{\tau=1}^{\infty} w(\tau) I(t-\tau)}}\right]. \tag{12}$$

Note that (approximate) coverage is always guaranteed conditionally on the past, and hence also without conditioning.

B Derivation of the infectivity profile for the SEIR-model

References

Abbott, S., Hellewell, J., Munday, J. D., Chun, J. Y., Thompson, R. N., Bosse, N. I., Chan, Y.-W. D., Russell, T. W., Jarvis, C. I., CMMID nCov working group, Flasche, S., Kucharski,

A. J., Eggo, R. and Funk, S. (2020). Temporal variation in transmission during the covid-19 outbreak. https://cmmid.github.io/topics/covid19/current-patterns-transmission/global-time-varying-transmission.html.

Cori, A., Ferguson, N. M., Fraser, C. and Cauchemez, S. (2013). A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics, *American Journal of Epidemiology* **178**(9): 1505–1512. https://doi.org/10.1093/aje/kwt133.

Fraser, C. (2007). Estimating individual and household reproduction numbers in an emerging epidemic, *PLOS ONE* **2**(8): 1–12. https://doi.org/10.1371/journal.pone.0000758.

Grassly, N. C., Fraser, C., Wenger, J., Deshpande, J. M., Sutter, R. W., Heymann, D. L. and Aylward, R. B. (2006). New strategies for the elimination of polio from india, *Science* **314**(5802): 1150–1153. https://science.sciencemag.org/content/314/5802/1150.

Johns Hopkins University Center for Systems Science and Engineering (2020). 2019 novel coronavirus covid-19 (2019-ncov) data repository by johns hopkins csse. https://github.com/CSSEGISandData/COVID-19.

Robert Koch-Institut (2020). Rki covid19. dl-de/by-2-0, https://npgeo-corona-npgeo-de.hub.arcgis.com/datasets/dd4580c810204019a7b8eb3e0b329dd6_0.