Mathematical Relationship between Effective Reproduction Number R_t and Epidemic Curve of Daily Cases Demonstration and Details

Bruno CAUDANA -- Adaptive Systems
b.caudana@adaptive.it
http://www.adaptive.it/covid19

24-jan-2021

Abstract

The strict mathematical relationship between R_t and the curve of daily cases $f(t)$ is shown. Up-to-date and statistically robust R_t from the curve of daily cases can be estimated as soon as new cases are added to the curve. That is equivalent to estimating R_t by averaging all detected cases of infection, without any distortion induced by the difficulty of following and weighting trees of secondary cases from original ones, and without needing to wait for secondary cases to manifest infection. With this method, if R_t scaled numbers are of interest, only the average duration of infectivity of subjects has to be estimated directly, but independently of linking secondary cases to primary ones. A new index, instantaneous reproduction number R_{ist} is introduced, which does not depend on the duration of infectivity of subjects. R_{ist}, R_t and the doubling/halving time of the epidemics may be estimated by simple computations at the very detection time of new daily cases. Any smoothed curve of daily cases gives smooth R_t and R_{ist}. No phase lag on R_t estimate is introduced by this method.

KEYWORDS: epidemics, reproduction number, daily cases, cumulative counts, epidemic curve, mathematics

Motivation for the method described here

I am new to epidemiology. I began to think about R_t during the first outbreak of COVID19 epidemics in Italy, while I was tinkering with a diffusion-saturation model trying to fit epidemic data: `http://adaptive.it/covid19/`. So, I do not know if what I found is new, or trivial, or already perfectly know. Excuse me for that. I am submitting my findings to the community in hope they may help.

During the first phase of COVID19 epidemics I encountered estimations of R_t which where incompatible with the doubling time of daily cases and the location in time of the peaks. So, I began to think on the subject.

It seems that R_t was defined from the epidemiological point of view with the assumption in mind that an epidemics can be characterized by a somewhat stable relationship between a pathogen and its infectable host. This in the hope of predicting the evolution of an outbreak. Which is not.

In fact, the initial susceptibility of a population of hosts is always unknown because unknown is the reaction of the immune system spectrum and history of a population. Besides that, both pathogen and host can modify this relationship via several options (decreasing susceptibility of the host population due to the spreading of the epidemics that saturates a population or sub-population of susceptible individuals, reaction of immune systems, reactive behaviors of the host and the pathogen populations, etc).

This writing shows how R_t definition is strictly tied to the curve of daily cases by mathematical equations. The two are essentially the same thing expressed with different words. R_t is a sort of first derivative of the curve of daily cases with respect to time t.

The difficulty of directly estimating R_t in a reliable way is the same as predicting the evolution of an epidemics in a reliable way.

Indeed even much harder, since one has to face the further uncertainty of estimating trees of secondary cases, with all the uncertainty implied by this process. It is very similar to estimating the space traveled by measuring acceleration with very inaccurate accelerometers, but very much harder and error prone.

The excellent articles by Cori, et al. [2] and Dietz [3] clearly show this difficulty.

Epidemiological definition of R_t

The epidemiological definition of R_t states:

> R_t ***is the number of secondary infections caused by a single case of disease during its period of infectivity in a completely susceptible population, on average.***
>
> (see: `https://en.wikipedia.org/wiki/Basic_reproduction_number` [5])

According to this epidemiological definition, R_t is analogous to the multiplier of the initial unit capital after 1 period, in a compound capitalization process.

This analogy allows the estimation of R_t from the epidemic curve of daily cases by introducing the concept of *Instantaneous Reproduction Number* R_{ist}, similar to the instantaneous capitalization rate in actuarial mathematics.

Definition of R_{ist}

The epidemiological definition of R_t (and its cousin R_0, as its limit to the beginning of an epidemics of an uninfected population) in-

dicates an exponential expansion. An infected, after his period of infectious capacity will have infected a new infected plus (or minus) a number of new infected individuals. Let's say, for example, one infected plus another one and a half infected, equal to two and a half infected ($1 + 1.5 = 2.5$). After 2 periods of infectivity, the infected will be those of the previous period (2.5) each of which will have infected new ones ($1 + 1.5 = 2.5$): i.e. the ($1 + 1.5$) of period 1, multiplied by ($1 + 1.5$) of period 2; and so on...

In general:

- after period 1: $1 \cdot (1 + r) = R_t$;
- after period 2: $1 \cdot (1 + r) \cdot (1 + r) = 1 \cdot (1 + r)^2$;
- and so on: $(1 + r)^1, (1 + r)^2 \ldots (1 + r)^p$.

In fact, this is a process equivalent to the amount of a compound capitalization of the interest rate r, where R_t is the amount after period 1.

To obtain which interest rate r should be used for a continuous compound capitalization of n fractions of a period that gives the amount R_t after 1 period, we can write as follows:

$$R_t = \left(1 + \frac{r}{n}\right)^n$$

equivalent to

$$R_t = \left[\left(1 + \frac{r}{n}\right)^{\frac{n}{r}}\right]^r$$

Passing to the limit for $n \to \infty$, and noting that $\lim_{n\to\infty}(1 + \frac{r}{n})^{\frac{n}{r}} = e$, we get:

$$R_t = \left[\lim_{n\to\infty}(1 + \frac{r}{n})^{\frac{n}{r}}\right]^r = e^r$$

hence

$$r = \ln(R_t)$$

In other terms, r is the exponent to be given to e to obtain R_t after a period of infectious duration equal to 1. That is:

$$R_{ist} \cdot 1 = r = \ln(R_t)$$

If we want to express R_{ist} in a unit of time g_i other than the dimensionless unit period, for example the days (or hours) with which we measure the duration of the infectivity period of an infectious subject and with which we measure the progress of the epidemic, we can write:

$$R_{ist} \cdot g_i = r = \ln(R_t)$$

from wich:

$$R_{ist} = \frac{\ln(R_t)}{g_i}$$

In this way we have the parameter R_{ist} which characterizes the exponential growth (as per the definition of R_t) at the point in time t that the increase (or decrease) of the daily cases generates.

Connecting R_{ist} to the epidemic curve of daily cases

Whenever an exponential function $y = e^{ax}$ is represented in logarithmic scale $\ln(y) = ax$, it becomes a straight line. Its shape factor a becomes the slope of the straight line (the angular coefficient).

If we represent the curve of the daily cases $f(t)$ in logarithmic scale $h(t) = \ln(f(t))$, the slope of the tangent of $h(t)$ at point t is the slope R_{ist}, corresponding to the exponential growth of the

epidemiological definition of the effective reproduction number R_t, represented in logarithmic scale, at time t, and scaled in time units of the curve of daily cases. But the tangent of $h(t)$ at point t is also the first derivative of $h(t)$

that is:

$$\frac{\ln(R_t)}{g_i} = R_{ist} = \frac{\mathrm{d}}{\mathrm{d}t}h(t)$$

A different reasoning perhaps better illustrates the concept of estimating R_t from epidemic curves.

R_t is basically the ratio between the daily cases at time $t+1$ compared to the cases at time t, where 1 is the infecting period. Given the point a on the curve of daily cases that precedes the point b, then $R_t = b/a$.

Differentiating the curve of daily cases, expressed in logarithmic scale with base e, means making the difference between two values, spaced by a unitary period of time tending to zero, that is: $\ln(b) - \ln(a)$. This expression is equivalent to doing $\ln(b/a)$, as those who have used slide rules easily remember: $\ln(b) - \ln(a) = \ln(b/a)$. (see: `https://en.wikipedia.org/wiki/Slide_rule`)

By doing the inverse operation of extracting a logarithm from a number, i.e. raising the base of the logarithm to a power of the value of the logarithm in question, one obtains the ratio b/a in the scale of daily cases of infection: $e^{\ln(b/a)} = b/a$.

This ratio represents the rate of increase (if > 1), or decrease (if < 1), of the infections averaged over **all** the infections observed, including **all** the information on the overall average resistance to the spread of the infection that may have formed meanwhile, for any known or unknown reason it was formed. It also takes in properly weighted account **all** the overlappings of the infection trees defined by R_t and of the varying susceptibility of the hosts.

Furthermore, the value obtained in this way is a very accurate

value of R_t acting at current time of b, that is, at the very moment in which the current value of the infected cases is known. The passage to the limit of a period that tends to the instant, implicit in the differentiation operation with respect to t, allows to have a curve of R_t trend that is always updated in real-time.

According to the epidemiological definition of R_t, we have the following correspondence of classical outstanding cases, direct consequence of that epidemiological definition:

$R_{ist} > 0$
: when the daily cases increase and the epidemic is expanding: therefore the associated $e^{R_{ist} \cdot g_i} = R_t > 1$.

$R_{ist} = 0$
: when the daily cases remain constant and the epidemic is stationary:
therefore the associated $e^{R_{ist} \cdot g_i} = R_t = 1$.
In this case the curve of daily cases has a minimum or a maximum;
R_{ist} crosses 0; R_t crosses 1.

$R_{ist} < 0$
: when the daily cases decrease and the epidemic is contracting: therefore the associated $e^{R_{ist} \cdot g_i} = R_t < 1$.

Since these outstanding cases derive from the epidemiological definition of R_t, they also are criterion for evaluating the correct estimate of R_t. A contrasting value of R_t respect to the epidemic curve is also an indication that R_t or the epidemic curve are wrong.

Summary of conversion formulas

The curve of daily cases $f(t)$ expressed in logarithmic scale with base e is obviously given by:

$$h(t) = \ln(f(t))$$

R_{ist} is given by the first derivative (numerically or analytically determined) of **any** smoothed curve of daily cases, given in logarithmic scale with base e:

$$R_{ist} = \frac{\mathrm{d}}{\mathrm{d}t} h(t)$$

Please notice that if we have a smoothing procedure of the curve of daily cases that introduces any phase lag, as we have using mobile averages or FIR/IIR filters, we will have the same phase lag in the estimation of R_{ist} and R_t. Otherwise if we have some form of static averaging, as using some least squares fitting procedure, **no phase lag is introduced**.

R_t is given by:

$$R_t = e^{(R_{ist} \cdot g_i)}$$

R_{ist} is also equivalen to:

$$R_{ist} = \frac{\ln(R_t)}{g_i}$$

The doubling or halving time of infection $g_{d \vee h}$ is given by imposing 2.0 as R_t and computing the number of resulting days (negative numbers represent halving time):

$$g_{d \vee h} = \frac{\ln(2.0)}{R_{ist}}$$

Some charting outcome

The following charts show how R_t may be estimated starting from a fitting of the curve of cumulative cases, with a sort of derivative of order 2. The curve of daily cases obviously is the first derivative of cumulative cases.

The fitting is primarily done on cumulative cases because they automatically compensate some kind of errors (for example: a missed case one day may be detected in the following days, etc.). Model and fitting techniques used for the following figures are outside the scope of this writing. Here the model is simply used as source of a smoothed daily data set. The other formulas used to generate the following charts are summarized in the section above. The data-source used for this fitting is the COVID-19 official one for Italy: `https://github.com/pcm-dpc/COVID-19/` [1]

Just a glance at the dispersion of a ample set of daily data around a good fitting of these data let easily imagine how difficult and unreliable could be any attempt to estimate a trend of the epidemics from small samples of their derivatives and relying on considerations of the spread of these samples over overlapping trees of secondary cases, which is what the epidemiological definition of R_t asks to do.

Moreover, the dynamics of an epidemic seems to follow unpredictable and chaotic behavior. We are used to think of populations involved in an epidemic as an isotropic material, like steel, which has equal behavior in all directions respect to stress and strain.

Perhaps, an epidemics may be better depicted as acting on many different relationship fabrics entangled together. A burst of infections occurs when two or more entangled fabrics – which may be in a stable infective condition that eventually saturate – mix and new connections merge in a new more extended fabric.

If this is a plausible landscape of an infection of a population, not every link in this entaglement of networks has the same infection capacity and not all nodes of these networks are isotropically connected.

In other words there may be several networks that may have poor connections with each other, while having strong connection among the members of each network. For example, the network of families with children that go to the same school may have strong link between families of teachers and classmates, but may have weak connections with other unrelated networks of parents-children-teachers. Some of these networks may saturate eventually, while others may not have even been infected. The same thing happens with other types of relational networks. This is a very anisotropic environment.

This landscape shows a very challenging non linear object to investigate. Maybe it has some emerging regularities at the macroscopic level, like sequences of overlapping sigmoidal shapes in the curve of cumulative cases.

References

[1] Protezione Civile. Dati COVID-19 Italia, 2020-2021. URL: https://github.com/pcm-dpc/COVID-19/.

[2] Anne Cori, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. *American Journal of Epidemiology*, 178(9):1505–1512, 09 2013. URL: https://doi.org/10.1093/aje/kwt133, doi:10.1093/aje/kwt133.

[3] K. Dietz. The estimation of the basic reproduction number for infectious diseases . *Statistical Methods in Medical Research*, 2(1):23–41, March 1993. URL: https://doi.org/10.1177/096228029300200103, doi:10.1177/096228029300200103.

[4] Modello a diffusione-saturazione per andamento COVID-19, 2020-2021. URL: http://adaptive.it/COVID-19/.

[5] Wikipedia. Basic reproduction number. URL: https://en.wikipedia.org/wiki/Basic_reproduction_number.

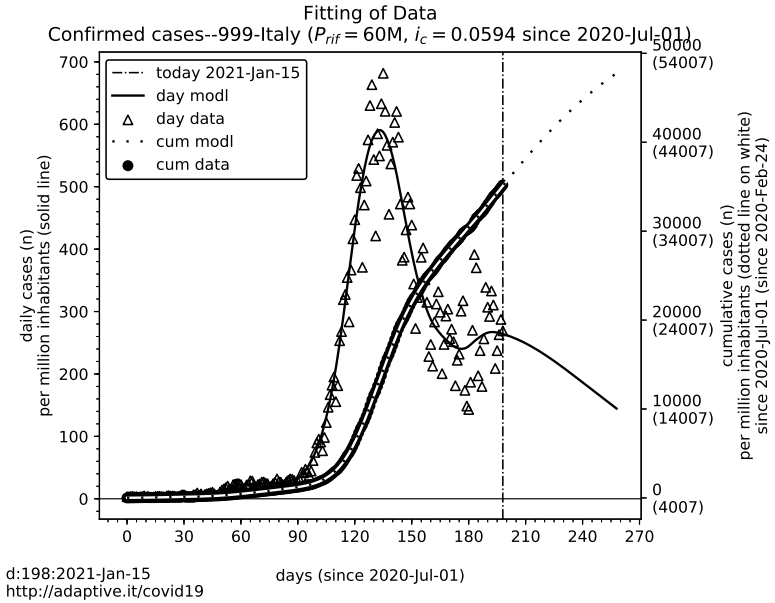

Figure 1: Data fit with an adaptive diffusion/saturation model on cumulative cases. Model and fitting techniques used for these charts are outside the scope of this writing. Fitting model and details at `http://adaptive.it/covid19/` [4]
data-source:`https://github.com/pcm-dpc/COVID-19/` [1]

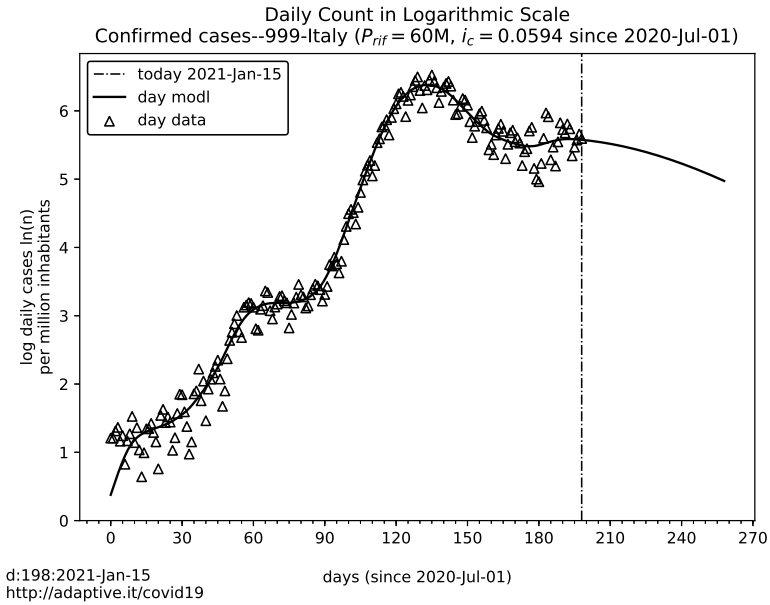

Figure 2: Daily cases smoothed by fitting, in logarithmic scale with base e. Fitting model and details at`http://adaptive.it/covid19/` [4] data-source:`https://github.com/pcm-dpc/COVID-19/` [1]

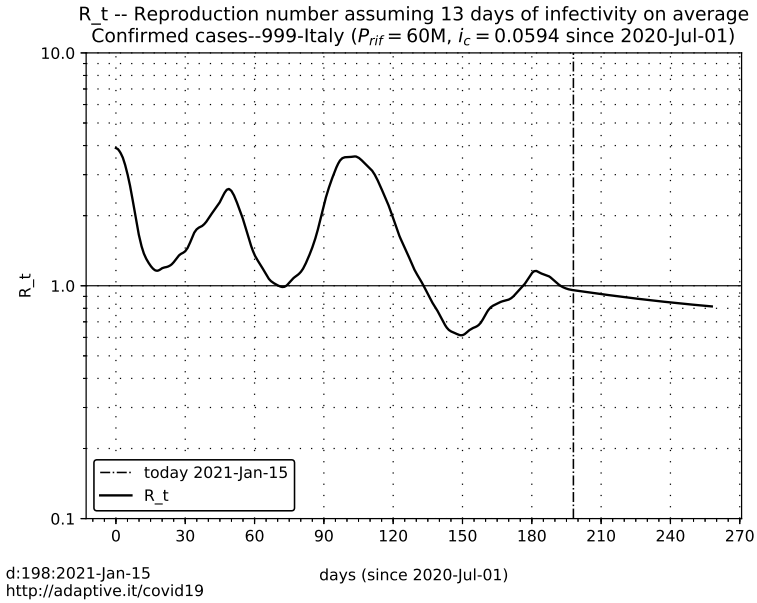

Figure 3: Computed R_t from smoothed daily cases in logarithmic scale with base e. Fitting model and details at `http://adaptive.it/covid19/http://adaptive.it/covid19/` [4] data-source:`https://github.com/pcm-dpc/COVID-19/` [1]

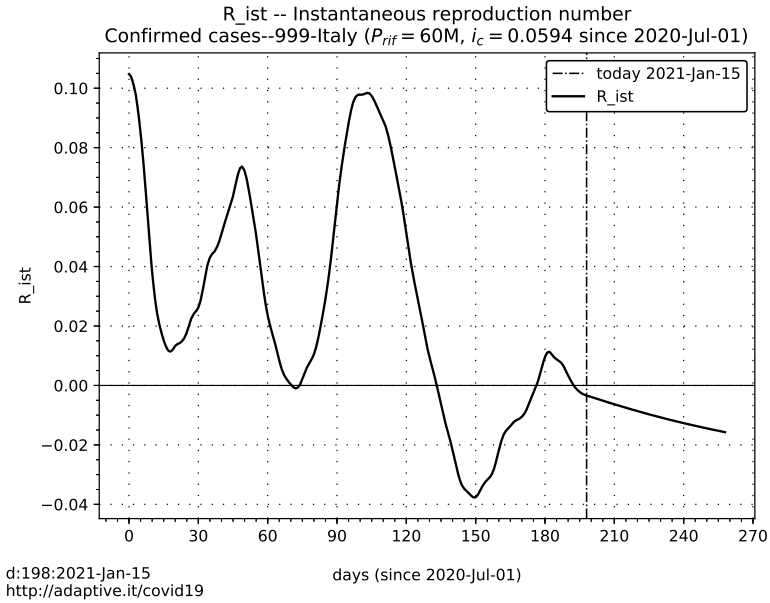

Figure 4: Computed R_{ist}.
Fitting model and details at `http://adaptive.it/covid19/` [4]
data-source:`https://github.com/pcm-dpc/COVID-19/` [1]

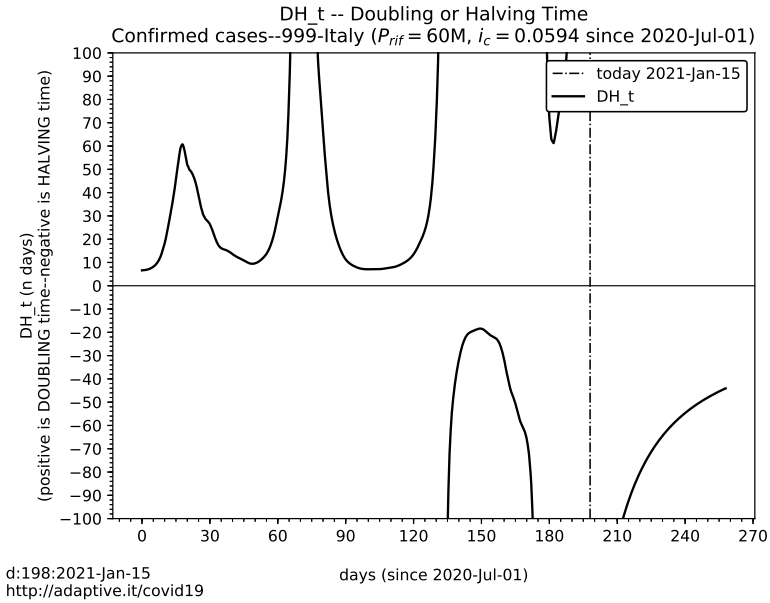

Figure 5: Computed doubling/halving time (days). Negative values mean halving time. Fitting model and details at `http://adaptive.it/covid19/` [4]
data-source:`https://github.com/pcm-dpc/COVID-19/` [1]