A Story of COVID-19 Modelling

Robert Schaback, version of April 29, 2020

This is an attempt to deal with the COVID-19 outbreak by comparably simple mathematical and numerical methods. I write this mainly for my own understanding, not aiming at a top-level journal paper. In particular, I observed that the media used certain terms in a rather misleading way, and I wanted to understand properly what is behind them. The final goal is to predict the peak of the epidemic outbreak per country with a reliable technique.

Open things are in red.

1 Classical SIR Modeling

I start with some basic notions that are useful for modelling epidemics, and that were in use in the media during the COVID-19 outbreak. Among other things, I shall give a rigid mathematical underpinning of what media mean by

- *flattening the epidemic outbreak*,
- *basic reproduction number*,
- *Herd Immunity Threshold*, and
- *doubling time*,

pointng out certain abuses of these notions. This will not work without calculus, but we try to keep things simple. Readers should take the opportunity to brush up their calculus knowledge.

1.1 The Model

The simplest standard "SIR" model of epidemics (e.g. [4] and easily retrievable in the wikipedia) deals with three variables

1. Susceptible (S),
2. Infectious (I), and

3. Removed (R).

The Removed cannot infect anybody anymore, being either dead or immune. This model is taking the viewpoint of bacteria of viruses. The difference between death and immunity of subjects is totally irrelevant from their viewpoint: they cannot proliferate anymore. The SIR model cannot say anything about death rates.

The Susceptible are living, not yet infected and not immune, while the Infectious can infect Susceptibles. These classes are disjoint and add up to a fixed total population count $N = S + I + R$. All of these are ideally assumed to be smooth functions of time t, and satisfy the differential equations

$$\begin{aligned} \dot{S} &= -\beta \frac{S}{N} I, \\ \dot{I} &= +\beta \frac{S}{N} I - \gamma I, \\ \dot{R} &= \gamma I. \end{aligned} \tag{1}$$

where the dot stands for the time derivative, and where β and γ are positive parameters. The product $\frac{S}{N} I$ stands for the product probability that an Infectious meets a Susceptible. Note that the Removed of the SIR model are not the Recovered of the Johns Hopkins data that we treat later, and the SIR model does not account for the Confirmed counted there.

Since $\dot{N} = \dot{S} + \dot{I} + \dot{R} = 0$, the equation $N = S + I + R$ is kept valid at all times. The term $\beta \frac{S}{N} I$ moves Susceptibles to Infectious, while γI moves Infectious to Removed. Thus β represents an *infection rate* while the *removal rate* γ accounts for either healing or fatality after infection, i.e. immunity. Political decisions about reducing contact probabilities will affect β, while γ resembles the balance between the medical aggressivity of the infection and the quality of the health care system.

Qualitatively, the system is not really dependent on N, because one can multiply N, R, I, and S by arbitrary factors without changing the system. As an aside, one can also go over to relative quantities with the two differential equations

$$\begin{aligned} \frac{d}{dt} \frac{I}{N} &= +\beta \left(1 - \frac{I}{N} - \frac{R}{N} \right) \frac{I}{N} - \gamma \frac{I}{N}, \\ \frac{d}{dt} \frac{R}{N} &= \gamma \frac{I}{N}. \end{aligned}$$

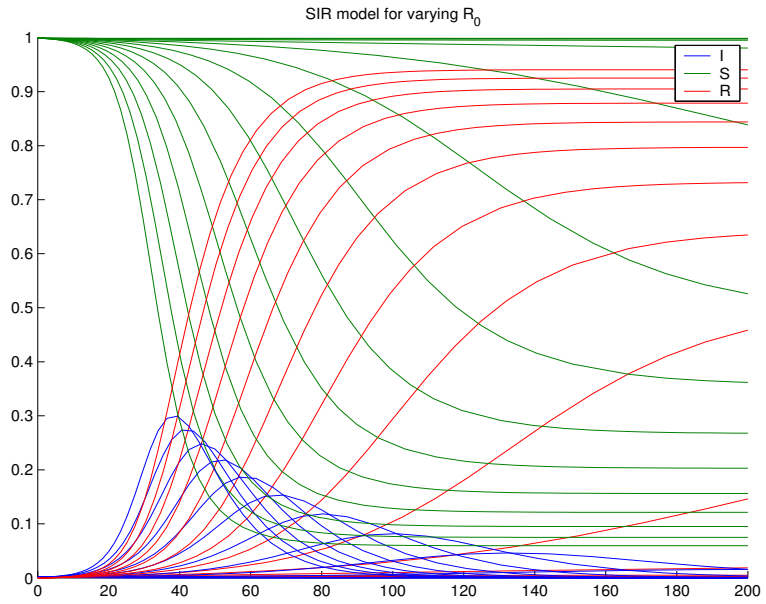

Figure 1: Some typical SIR system solutions

Figure 1 shows some simple examples that will be explained in some detail below.

1.2 Conditions for Outbreaks

We start by looking at the initial conditions. Since everything is invariant under an additive time shift, we can start at time 0 and consider

$$\dot{I}(0) = +\beta \frac{S(0)}{N} I(0) - \gamma I(0)$$

and see that the Infectious decrease right from the start if

$$\frac{S(0)}{N} < \frac{\gamma}{\beta}, \tag{2}$$

and this keeps going on since S must decrease and

$$\frac{\dot{I}}{I}(t) = \beta \frac{S(t)}{N} - \gamma \leq \beta \frac{S(0)}{N} - \gamma < 0. \tag{3}$$

There is no outbreak in this case, because there are not enough Susceptibles at start time. The case $S(0) = N$ means that there is no infection at all, and we ignore it, though it is a solution to the system, with $I = R = 0, S = N$ throughout.

1.3 The Peak

As long as the Infectious I are positive, the Susceptibles S are decreasing, while the Removed R are increasing. Excluding the trivial case of zero Infectious from now on, the Removed and the Susceptible will be strictly monotonic.

If there is a time instance t_I (maybe $t_I = 0$ above) where the Infectious are positive and do not change, we have

$$\begin{aligned} 0 = \dot{I}(t_I) &= \beta \frac{S(t_I)}{N} I(t_i) - \gamma I(t_I), \\ \gamma &= \beta \frac{S(t_I)}{N} \leq \beta. \end{aligned} \tag{4}$$

If $\beta < \gamma$ holds, this situation cannot occur, and I must be decreasing all the time, i.e. the infection dies out. This is what everybody wants. There is no outbreak. In case of $\gamma = \beta$ we go back to the initial situation of the previous section and see that there is no outbreak due to $S(0)/N < 1$ if there is an infection at all.

The interesting case is $\beta > \gamma$. Then the first part of (3) shows that as soon as t is larger than the peak time t_I, the Infectious will decrease due to (4). Therefore the zero of $\dot{I}$ must be a maximum, i.e. a *peak*, and it is unique. The Infectious go to zero even in the peak situation.

It is one of the most important practical problems in the beginning of an epidemic to predict

- whether there will be a peak at all,
- when the possible peak will come, and
- how many Infectious will be at the peak.

This can be answered if one has good estimates for β and γ, and we shall deal with this problem later.

It will turn out that is is highly important to avoid the peak situation, and this can only be done by administrative measures that change β and γ to the situation $\beta < \gamma$. This is what management of epidemics is all about, provided that an epidemic follows the SIR model.

1.4 Basic Reproduction Number

The quotient

$$R_0 := \frac{\beta}{\gamma}$$

is called the *basic reproduction number*. If it is not larger than one, there is no outbreak, whatever the initial conditions are. If it is larger than one, there is an outbreak provided that

$$1 > \frac{S(0)}{N} > \frac{\gamma}{\beta} = \frac{1}{R_0} \tag{5}$$

holds. In that case, there is a time t_I where I reaches a maximum, and (4) holds there. When we discuss an outbreak in what follows, we always assume $R_0 > 1$ and (5). If we later let R_0 tend to 1 from above, we also require that $S(0)$ tends to N from below, in order to stay in the outbreak situation.

Both β and γ change under a change of time scale, but the basic reproduction number is invariant. Physically, β and γ have the dimension $time^{-1}$, but $R_0 = \beta/\gamma$ is dimensionless.

1.5 Examples

Figure 1 shows a series of test runs with $S(0) = N \cdot 0.999$ and $R(0) = 0$ with fixed $\gamma = 0.1$ and β varying from 0.02 to 0.3, such that R_0 varies from $1/5$ to 3. Due to the realistically small $I(0)$, one cannot see the decaying cases near startup, but the tails of the blue I curves are decaying examples by starting value, due to $\frac{S(t)}{N} < \frac{\gamma}{\beta} = 1/R_0$ when started at time t. Decreasing R_0 flattens the blue curves for I. One can observe that I always dies out, while S and R tend to fixed positive levels. We shall prove this below. From the system, one can also infer that R has an inflection point where I has its maximum, since $\ddot{R} = \gamma \dot{I}$. If only R would be observable, one could locate the peak of I via the inflection point of R.

Figure 2 shows an artifical case with a large starting value $I(0) = N/2$, fixed $\gamma = 0.1$ and β varying from 0.005 to 0.3, letting R_0 vary from 0.05 to 3. In contrast

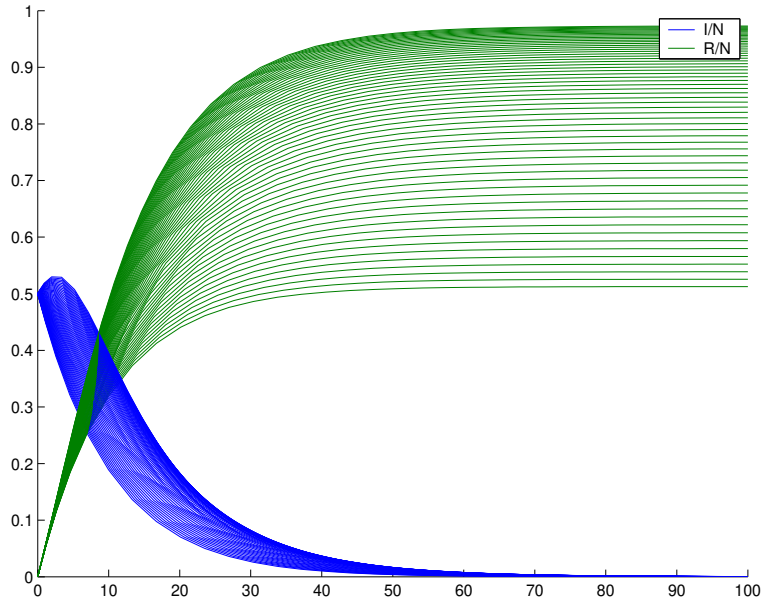

Figure 2: Some other typical SIR system solutions SIR01.m

to Figure 1, this example shows cases with small R_0 properly. The essence is that the Infectious go down, whether they have a peak or not, and there will always be a portion of Susceptibles. Again, we shall prove this below.

1.6 Herd Immunity Threshold

This is a number related to the Basic Reproduction Number R_0 by

$$HIT = 1 - \frac{1}{R_0}$$

following a special scenario. If a population is threatened by an infection with Basic Reproduction Number R_0, what is the number of immune persons needed to prevent an outbreak right from the start? We can read this off equation (2) in the ideal situation that $I(0) = 0$ and $S(0) + R(0) = N$, namely

$$\frac{S(0)}{N} = 1 - \frac{R(0)}{N} = \frac{\gamma}{\beta} = \frac{1}{R_0}$$

implying

$$\frac{R(0)}{N} = 1 - \frac{1}{R_0}$$

as the threshold between outbreak and decay. This does not refer to a whole epidemic scenario, nor to an epidemic outbreak. It is a condition to be checked before anything happens, and useless within a developing epidemic.

In the peak situation of (4), the fraction

$$\frac{R(t_I) + I(t_I)}{N} = \frac{N - S(t_I)}{N} = 1 - \frac{1}{R_0}$$

of the Non-Susceptible at the peak t_I of I is exactly the Herd Immunity Threshold. Thus it is correct to say that if the Immune of a population are below the Herd Immunity Threshold at startup, and if the Basic Reproduction Number is larger than one, the sum of the Immune and the Infectious will rise up to the Herd Immunity Threshold and then the Infectious will decay. This is often stated imprecisely in the media. The Herd Immunity Threshold has nothing to do with the long-term ratio of Susceptibles to Removed. We shall address this ratio below.

1.7 Locating the Peak

The most interesting questions during an outbreak with $R_0 > 1$ are

- At which time t_I will we reach the maximum of the Infectious, and
- what is $I(t_I)$, i.e. how many people will maximally be infectious at that time?

It will turn out that there are no easy direct answers. From (4) we see that at the maximum of I the Susceptibles S have the value

$$\frac{\gamma}{\beta} = \frac{S(t_I)}{N} = \frac{1}{R_0},$$

i.e. the portion $1/R_0$ of the population is susceptible. From that time on, the Infectious decrease. In terms of R and I, the value

$$\frac{R(t_I) + I(t_I)}{N} = 1 - \frac{1}{R_0}$$

of the Non-Susceptibles marks the peak of the Infectious at the Herd Immunity Threshold. "Flattening the curve", as often mentioned in the media, is intended to mean making the maximum of I smaller, but this is not exactly what happens, since the maximum is described by the penultimate equation concerning the Susceptibles, while for $I(t_I)$ we only know

$$\frac{I(t_I)}{N} \leq \frac{R(t_I)+I(t_I)}{N} = 1 - \frac{1}{R_0} \tag{6}$$

yielding that the left-hand side gets smaller if R_0 gets closer to one. Politically, this requires either making β smaller via reducing contact probabilities or making γ larger by improving the health system, or both. Anyway, "flattening the curve" works by letting R_0 tend to 1 from above, but the basic reproduction number does not directly determine the time t_I of the maximum or the value there. We shall improve the above analysis later.

1.8 Analyzing the Outbreak

In the beginning of the outbreak, S/N is near to one, and therefore

$$\dot{I} \approx +\beta I - \gamma I$$

models an exponential outbreak with exponent $\beta - \gamma > 0$, with a solution

$$I(t) \approx I(t_0)\exp((\beta-\gamma)t).$$

If this is done in discrete time steps Δt, one has

$$\frac{I(t+\Delta t)}{I(t)} \approx \exp((\beta-\gamma)\Delta t).$$

The severity of the outbreak is not controlled by $R_0 = \beta/\gamma$, but rather via $\beta - \gamma$. Publishing single values $I(t)$ does not give any information about $\beta - \gamma$. Better is the ratio of two subsequent values

$$\frac{I(t_2)}{I(t_1)} \approx \exp((\beta-\gamma)(t_2-t_1)),$$

and if this gets smaller over time, the outbreak gets less dramatic because $\beta - \gamma$ gets smaller. Really useful information about an outbreak does not consist of values and not of increments, but of increments of increments, i.e. some second derivative information. This is what the media rarely provided during the outbreak.

1.9 Doubling Time

Another information used by media during the outbreak is the "doubling time", e.g. how many days it takes until daily values double. This is the number n in

$$2 = \frac{I(t+n\Delta t)}{I(t)} \approx \exp((\beta-\gamma)n\Delta t) = (\exp((\beta-\gamma)\Delta t)^n$$

or

$$n = \frac{\log 2}{(\beta-\gamma)\Delta t},$$

i.e. it is inversely proportional to $\beta-\gamma$. If political action doubles the "doubling time", if halves $\beta-\gamma$. If politicians do this repeatedly, they never reach $\beta<\gamma$, and they never escape an exponential outbreak if they do this any finite number of times. When presenting a "doubling time", media should always point out that this makes only sense during an exponential outbreak. And it is not related to the basic reproduction number $R_0 = \beta/\gamma$, but rather to the difference $\beta-\gamma$.

1.10 Spread of Infections

Media often say that the basic reproduction number R_0 gives the number of persons an average Infectious infects while being infectious. This is a rather mystical statement that needs underpinning. The quantity

$$\frac{1}{\gamma} = \frac{I}{\dot{R}}$$

is a time value that describes the ratio between current Infectious and current newly Removed, and thus can be seen as the average time needed for an Infectious to get Removed, i.e. it is the average time that an Infectious can infect others. Correspondingly,

$$\dot{I}+\gamma I = \dot{I}+\dot{R} = \beta\frac{S}{N}I$$

are the newly Infected, and therefore

$$\frac{1}{\beta}\frac{N}{S} = \frac{I}{\dot{I}+\dot{R}}$$

can be seen as the time it needs for an average Infectious to generate a new Infectious. The ratio $\frac{\beta}{\gamma}\frac{S}{N}$ then gives how many new Infectious can be generated by an

Infectious while being infected, but this is only close to R_0 if $S \approx N$, i.e. at the start of an outbreak. Another way to say this is that R_0 is the average number of infections an Infectious generates while being infectious, but within an unlimited supply of Susceptibles.

1.11 Long-term Behavior

If we are at a time t_D behind the possible peak at t_I, or in a decay situation enforced by starting value, like in (2), we know that I must decrease exponentially to zero. This follows from

$$(\log I)^{\cdot} = \frac{\dot{I}}{I} = \beta \frac{S(t)}{N} - \gamma \leq \beta \frac{S(t_D)}{N} - \gamma < 0 \tag{7}$$

showing that $\log I$ must decrease linearly, or I must decrease exponentially. Thus we get rid of the Infectious in the long run, keeping only Susceptibles and Removed. Surprisingly, this happens independent of how large R_0 is. Figures 1 and 2 demonstrate how S and R level out under all circumstances shown, but in which final ratio?

Dividing the first equation in (1) by the third leads to

$$\frac{\frac{d}{dt}S}{\frac{d}{dt}R} = \frac{dS}{dR} = -\frac{\beta}{\gamma}\frac{S}{N},$$

and when setting $\sigma = S/N$ and $\rho = R/N$, we get

$$\frac{d\sigma}{d\rho} = -\frac{\beta}{\gamma}\sigma$$

with the solution

$$\sigma(\rho) = \sigma(0)\exp\left(-\frac{\beta}{\gamma}\rho\right) \tag{8}$$

when assuming $R(0) = 0$ at startup. Since ρ is increasing, it has a limit $0 < \rho_\infty \leq 1$ for $t \to \infty$, and in this limit we get

$$\sigma(\rho_\infty) = \sigma(0)\exp\left(-\frac{\beta}{\gamma}\rho_\infty\right)$$

together with the condition $\rho_\infty + \sigma(\rho_\infty) = 1$. The equation

$$\sigma(0)\exp\left(-\frac{\beta}{\gamma}\rho_\infty\right) = 1 - \rho_\infty$$

has a unique solution in $(0,1)$ dependent on $\sigma(0) < 1$ and $R_0 = \beta/\gamma$. See Figure 3 for illustration. Looking at both sides of the equation as functions of ρ_∞, an increase of $R_0 = \beta/\gamma$ for fixed $S(0) < 1$ lets the intersection point move towards 1.

This has some serious implications, if the model is correct for an epidemic situation. First, the Infectious always go to zero, but Susceptibles always remain. This means that a new infection can always arise whenever some infected person enters the sanitized population. The outbreak risk is dependent on the portion $\sigma_\infty = 1 - \rho_\infty$ of the Susceptibles. This illustrates the importance of vaccination, e.g. against measles or influenza.

The above analysis shows that large values of R_0 lead to large relative values of Removed to Susceptible in the limit. The consequence is that systems with large R_0 have a dramatic outbreak and lead to a large portion of Removed. This is good news in case that the rate of fatalities within the Removed is low, but very bad news otherwise.

When politicians try to "flatten the curve" by bringing R_0 below 1 from some time on, this will automatically decrease the asymptotic rate of Removed and increase the asymptotic rate of Susceptibles in the population. This is particularly important if the rate of fatalities within the Removed is high, but by the first argument the risk of re-infection rises due to the larger portion of Susceptibles.

The decay situation (7) implies that

$$\sigma_\infty = \frac{S(\infty)}{N} \leq \frac{\gamma}{\beta} = \frac{1}{R_0}$$

and consequently

$$\rho_\infty = 1 - \sigma_\infty \geq 1 - \frac{1}{R_0}.$$

Therefore the final rate of the Removed is not smaller than the Herd Immunity Threshold. This is good news for possible re-infections.

1.12 Asymptotic Exponential Decay

In a decay situation like in (7), we get

$$I(t_D)\exp(-(\gamma - \beta\sigma_\infty)(t - t_D)) \leq I(t) \leq I(t_D)\exp(-(\gamma - \beta\sigma(t_D))(t - t_D))$$

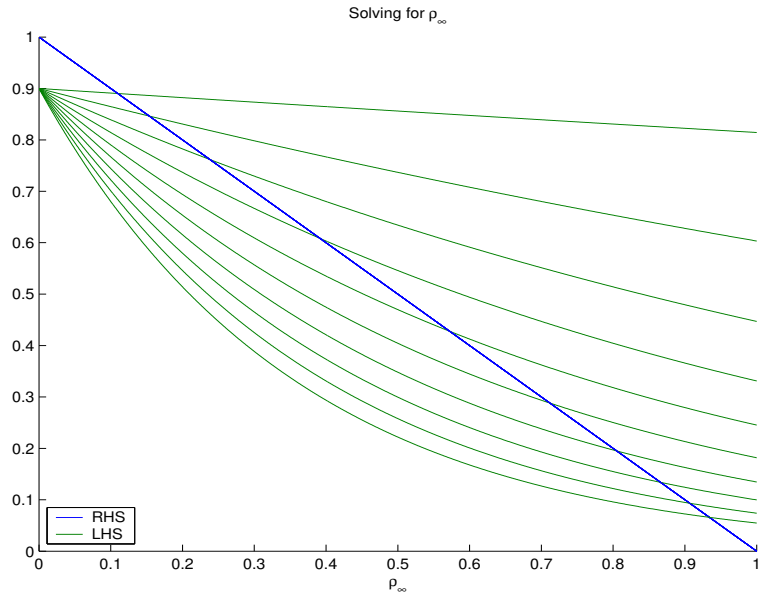

Figure 3: Solving for ρ_∞ for fixed $C(0) = 0.9$ and varying R_0

to see that the exponential decay is not ruled by $\beta - \gamma$ as in the outbreak case with $R_0 > 1$, but rather by $-\gamma + \beta\sigma_\infty$. This also holds for large $R_0 = \beta/\gamma$ because σ_∞ counteracts. The bell shapes of the peaked I curves are not symmetric with respect to the peak.

1.13 Back to the Peak

If we go back to analyzing the peak of I at t_I for $R_0 > 1$, we know

$$\sigma(t_I) = \frac{\gamma}{\beta} = \frac{1}{R_0} = \sigma(\rho(t_I)) = \sigma(0)\exp(-R_0\rho(t_I))$$

and get

$$\rho(t_I) = \frac{1}{R_0}\log(\sigma(0)R_0)$$

leading to

$$\frac{I(t_I)}{N} = 1 - \sigma(t_I) - \rho(t_I) = 1 - \frac{1}{R_0} - \frac{1}{R_0}\log(\sigma(0)R_0)$$

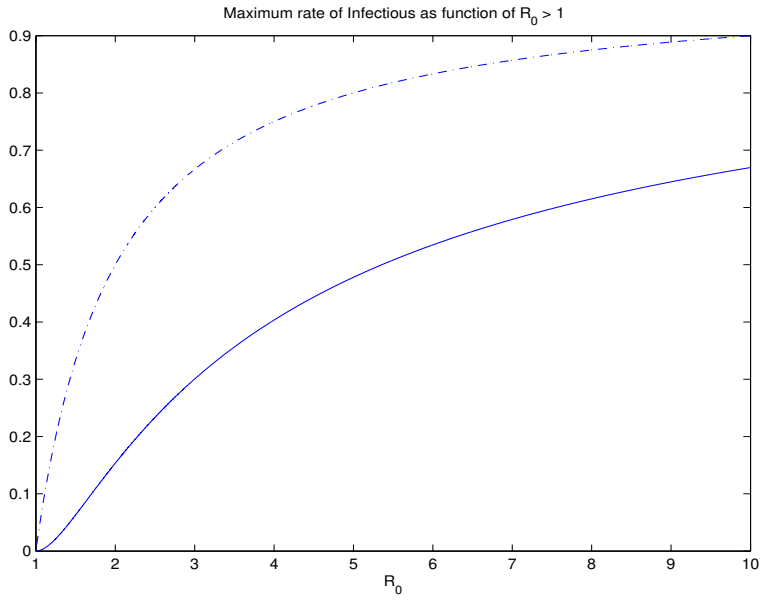

Figure 4: The effect of R_0 on the maximum rate of Infectious within the population

as the exact value at the maximum, improving (6). Note that the final log is positive due to the condition (5) for an outbreak.

For standard infections that have starting values $\sigma(0) = S(0)/N$ very close to one, the maximal ratio of Infectious is

$$\frac{I(t_I)}{N} \approx 1 - \frac{1}{R_0} - \frac{1}{R_0}\log(R_0).$$

Figure 4 shows the behaviour of the function, and this is what "flattening the curve" is all about. A value of $R_0 = 4$ gets a maximum of more than 40% of the population infectious at a single time. If 5% need hospital care, this implies that a country needs hospital beds for 2% of the population. The dotted line leaves the log term out, i.e. is marks the rate of the Susceptibles at the peak, and by (6) the difference is the rate $R(t_I)/N$ of the Recovered at the peak.

To analyze the peak time t_I, we use

$$\frac{\dot{I}}{I} = \beta\frac{S}{N} - \gamma \leq \beta - \gamma$$

to get an upper bound for the exponential outbreak

$$I(t) \leq I(0)\exp((\beta-\gamma)t)$$

that implies a lower bound for t_I of the form

$$1-\frac{1}{R_0}-\frac{1}{R_0}\log(R_0) \leq 1-\frac{1}{R_0}-\frac{1}{R_0}\log(\sigma(0)R_0) = \frac{I(t_I)}{N} \leq \frac{I(0)}{N}\exp((\beta-\gamma)t_I).$$

To be improved ???

1.14 Flattening the Curve

To get a quantitative result about "flattening the curve", we first evaluate the integral

$$\int_0^\infty \frac{I(s)}{N}ds = \frac{1}{\gamma}\int_0^\infty \frac{R'(s)}{N}ds = \frac{1}{\gamma}\rho(\infty)$$

assuming $R(0)=0$, and set it equal to an integral over the constant value at the maximum, i.e. we squeeze the area under the curve into a rectangle of length $b-a$ under the maximal value, i.e.

$$\frac{1}{\gamma}\rho(\infty) = (b-a)\frac{I(t_I)}{N}.$$

This implies

$$b-a = \frac{\frac{1}{\gamma}\rho(\infty)}{\frac{I(t_I)}{N}} \geq \frac{1}{\gamma}\frac{\rho(\infty)}{1-\frac{1}{R_0}},$$

and if we "flatten the curve" by letting R_0 tend to 1 from above, we see that the length $b-a$ of the above rectangle goes to infinity like $R_0/(1-R_0)$, because ρ_∞ tends to 1.

If there is no peak, e.g. if $R_0=\beta/\gamma$ is below 1 either at the beginning or after some political intervention, one can repeat the above argument starting with the Infectious at some time t looking at the area under I from t to infinity:

$$\int_t^\infty \frac{I(s)}{N}ds = \frac{1}{\gamma}\int_t^\infty \frac{R'(s)}{N}ds = \frac{1}{\gamma}(\rho(\infty)-\rho(t)) = (b-a)\frac{I(t)}{N}$$

leading to

$$b-a=\frac{1}{\gamma}\frac{\rho(\infty)-\rho(t)}{\frac{I(t)}{N}}=\frac{1}{\gamma}\frac{\rho(\infty)-\rho(t)}{1-\sigma(t)-\rho(t)}$$

For $t\to\infty$, this is 0/0 and needs some L'Hospital argument. To be done ...

1.15 The Infection Timescale

Here is an aside that is well-known in the SIR literature. The SIR system can be written as

$$\begin{aligned}\frac{dS}{N} &= -\beta\frac{S}{N}\frac{I}{N}dt\\ \frac{dI}{N} &= (+\beta\frac{S}{N}-\gamma)\frac{I}{N}dt\\ \frac{dR}{N} &= \gamma\frac{I}{N}dt\end{aligned}$$

and in a new time variable τ with $d\tau=\frac{I}{N}dt$, one gets the system

$$\begin{aligned}\frac{d\sigma}{d\tau} &= -\beta\sigma(\tau),\\ \frac{d\rho}{d\tau} &= -\gamma\end{aligned}$$

for $\sigma=S/N$ and $\rho=R/N$ as functions of the new *infection timescale* τ that one can fix as

$$\tau(t)=\int_0^t\frac{I(s)}{N}ds$$

to make sure that $\tau(0)=0$. This implies

$$\begin{aligned}\sigma(\tau) &= \sigma(0)\exp(-\beta\tau),\\ \rho(\tau) &= \rho(0)+\gamma\tau.\end{aligned}$$

The beauty of this is that the role of β and γ are perfectly split. In the new timescale, ρ increases linearly and σ decreases exponentially. The Basic Reproduction Number then describes the fixed ratio

$$\frac{\beta}{\gamma}=R_0=\frac{\log S(\tau)-\log S(0)}{\rho(\tau)-\rho(0)},$$

and the result (8) of section 1.11 comes back as

$$\sigma(\tau) = \sigma(0)\exp\left(-\frac{\beta}{\gamma}\rho(\tau)\right)$$

for the case $\rho(0) = 0$. This approach has the disadvantage to conceal the peak within the new timescale.

2 Using Available Data

Now we want to confront the modelling of the previous section with available data.

2.1 Johns Hopkins Data

In this text, we work with the COVID-19 data from the CSSEGISandData repository of the Johns Hopkins University [5]. They are the only source that provides comparable data on a worldwide scale.

The numbers there are

1. Confirmed (C) or *cumulative infected*
2. Dead (D), and
3. Recovered (R)

as cumulative integer valued time series in days from Jan. 22nd, 2020. All these values are absolute numbers, not relative to a total population. Note that the unconfirmed cases are not accessible at all, while the Confirmed contain the Dead and the Recovered of earlier days.

At this point, we do not question the integrity of the data, but there are many well-known flaws. In particular, the values for specific days are partly belonging to previous days, due to delays in the chains of data transmission in different countries. This is why, at some points, we shall apply some conservative smoothing of the data. Finally, there are inconsistencies that possibly need data changes. For an example, consider that usually COVID-19 cases lead to recovery or death within a rather fixed period, e.g. $k \approx 15-20$ days. But some Johns Hopkins data have

less new Infectioned at day n than the sum of Recovered and Dead at day $n+k$. And, there are countries like Germany who deliver data of Recovered in a very questionable way. The law in Germany does not enforce authorities to collect data of Recovered, and the United Kingdom does not report numbers of Dead and Recovered from places outside the National Health System, e.g. from Senior's retirement homes.

We might assume that the Dead plus the Recovered of the Johns Hopkins data are the Removed of the SIR model, and that the Infectious $I = C - R - D$ of the Johns Hopkins data are the Infectious of the SIR model. But this is not strictly valid, because registration or confirmation come in the way.

On the other hand, one can take the radical viewpoint that facts are not interesting if they do not show up in the Johns Hopkins data. Except for the United Kingdom, the important figures concern COVID-19 casualties that are actually registered as such, others do not count, and serious cases needing hospitalization or leading to death should not go unregistered. If they do in certain countries, using such data will not be of any help, unless other data sources are available. If SIR modelling does not work for the Johns Hopkins data, it is time to modify the SIR technique appropriately, and this will be tried here, partially.

An important point for what follows is that the data come as daily values. To make this compatible with differential equations, we shall replace derivatives by differences.

2.2 Examples

To get a first impression about the Johns Hopkins data, Figure 5 shows raw data up to day 97 (April 28th, as of this writing). The presentation is logarithmic, because then linear increasing or decreasing parts correspond to exponential increasing or decreasing numbers in the real data. Many presentations in the media are non-logarithmic, and then all exponential outbreaks look similar. The real interesting data are the Infectious $I = C - R - D$ in black that show a peak or not. The other curves are cumulative. The data for other countries tell similar stories and are suppressed.

The media, in particular German TV, present COVID-19 data in a rather debatable way. When mentioning Johns Hopkins data, they provide C, D, and R separately without stating the most important figures, namely $I = C - D - R$, their change,

and the change of their change. When mentioning data of the Infectious from the Robert Koch institute alongside, they do not say precisely that these are non-cumulative and should be compared to the $I = C - R - D$ data of the Johns Hopkins University. And, in most cases during the outbreak, they did not mention the change of the change. Quite like all other media.

One can see in Figure 5 that Germany and South Korea have passed the peak of the Infectious, while France is roughly at the peak and the United States are still in an exponential outbreak. The early figures, below day 40, are rather useless, but then an exponential outbreak is visible in all cases. This outbreak changes its slope due to political actions, and we shall analyze this later. See [3] for a detailed early analysis of slope changes.

There are strange anomalies in the Recovered (green). France seems not to have delivered any data between days 40 and 58, Germany changed the data delivery policy between days 62 and 63, and the UK data for the Recovered are a mess.

It should be noted that the available medical results on the COVID-19 disease often state that Confirmed will die or survive after a more or less fixed number of days, roughly 14 to 18. This would imply that the red curves for the Dead and the green curves for the Recovered should roughly follow the blue curves for the Confirmed with a fixed but measurable delay. This is partially observable, but much less accurately for the Recovered.

2.3 Estimating R_0

To get going, we use the correspondences

$$\begin{aligned} I_{SIR} &\Leftrightarrow C_{JH} - D_{JH} - R_{JH}, \\ R_{SIR} &\Leftrightarrow D_{JH} + R_{JH}, \\ (I+R)_{SIR} &\Leftrightarrow C_{JH} \end{aligned}$$

without being able to do something about S_{SIR} at this point. From now on, we shall omit the subscript JH when we use the Johns Hopkins data, but we shall use SIR when we go back to the SIR model.

To get a grip on what γ is in the SIR model, one can use the SIR equation $\dot{R}_{SIR} = \gamma I_{SIR}$ in the Johns Hopkins form

$$\gamma_n = \frac{R_{n+1} + D_{n+1} - R_n - D_n}{C_n - D_n - R_n}$$

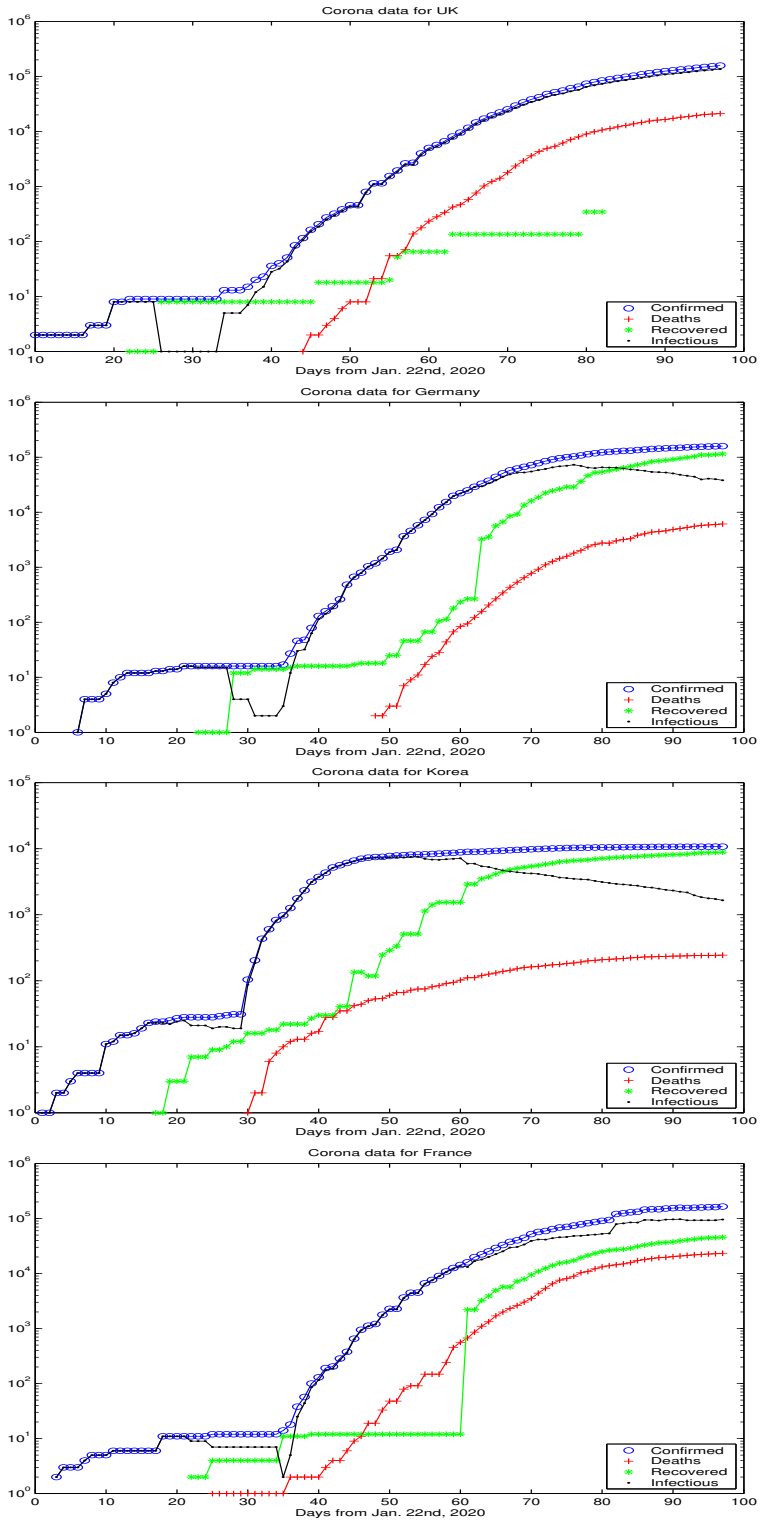

Figure 5: Raw Johns Hopkins data in logarithmic presentation up to day 97, from top: UK, Germany, South Korea, and France

to get a time series from the Johns Hopkins data that models γ. It gives the fraction of the newly Removed against the Infectious of the previous day, in the subpopulation of the Confirmed.

Similarly, the SIR equation $\dot{I}_{SIR} + \dot{R}_{SIR} = I_{SIR} \cdot \beta \frac{S_{SIR}}{N}$ leads to a time series

$$b_n = \frac{C_{n+1} - C_n}{C_n - D_n - R_n}$$

that models $\beta \frac{S_{SIR}}{N}$ without knowing N or S. By brute force, one can consider

$$r_n = \frac{b_n}{\gamma_n} = \frac{C_{n+1} - C_n}{R_{n+1} + D_{n+1} - R_n - D_n}$$

as a data-driven substitute for

$$\frac{\beta}{\gamma} \frac{S_{SIR}}{N} = R_0 \frac{S_{SIR}}{N}.$$

Then, using $C = I + R + D$,

$$\begin{aligned} C_{n+1} - C_n &= r_n(R_{n+1} + D_{n+1} - R_n - D_n) \\ I_{n+1} - I_n + R_{n+1} - R_n + D_{n+1} - D_n &= r_n(R_{n+1} + D_{n+1} - R_n - D_n) \\ I_{n+1} - I_n &= -(1 - r_n)(R_{n+1} - R_n + D_{n+1} - D_n) \end{aligned}$$

shows that if $r_n < 1$ holds, there is no increase in the Confirmed Infectious, very much like for the Infectious under the condition $R_0 < 1$ in the SIR model.

As long as S/N is very close to 1, this technique can be used to estimate R_0. At this point, it is not intended to model the epidemics. The focus is on extracting relevant parameters from the data. The good news is that $r_n < 1$ will lead to a decrease of I_n, but this is visible in the data anyway and not of much help.

Figure 6 shows R_0 estimates via r_n for the last four weeks before day 93, i.e. April 25. Except for US, UK, and Sweden who still have values about 5, the other countries fight for pressing R_0 below one, with varying success. In all cases, S/N is too close to one to have any influence. The variation in r_n is not due to the decrease in S/N, but should rather be attributed to political action.

For the figure, the raw Johns Hopkins data were smoothed by a double action of a $1/4, 1/2, 1/4$ filter on the logarithms of the data. This smoother keeps constants and linear sections of the logarithm invariant, i.e. it does not change local exponential behavior. This smoothing was not applied to Figure 5.

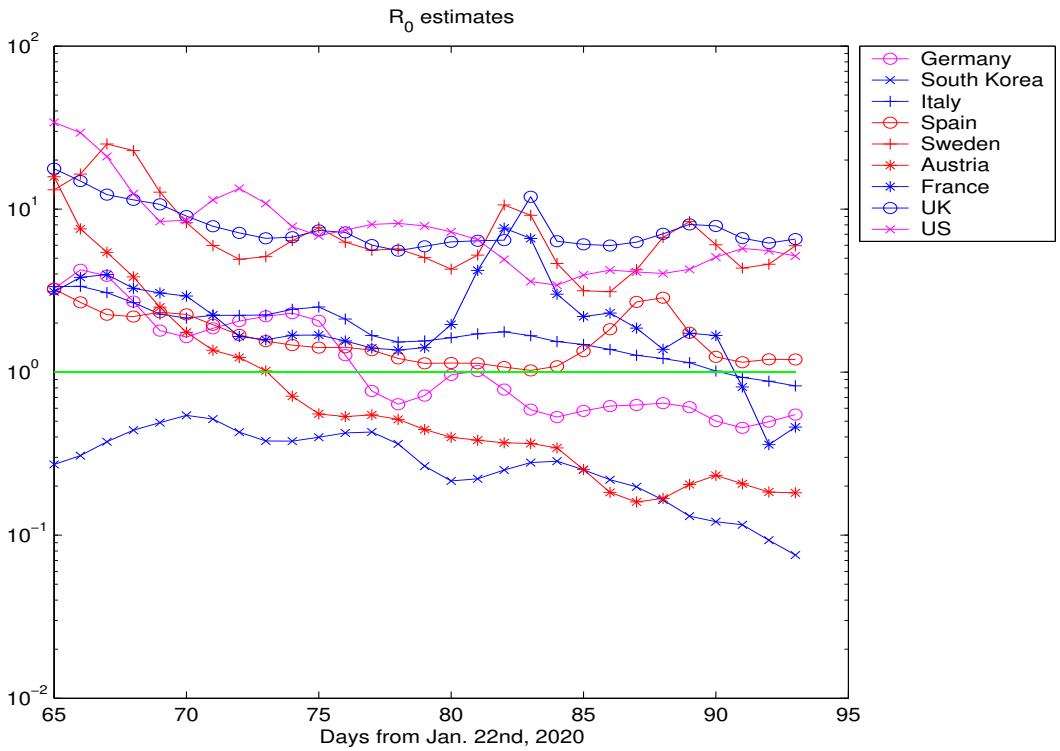

Figure 6: Estimates of R_0 via the time series r_n

As long as r_n is roughly constant, the above approach will always model an exponential outbreak or decay, but never a peak, because the difference equations are linear. It can only help the user to tell if there is a peak ahead or behind, depending on $r_n \approx R_0$ being larger or smaller than 1. If r_n is kept below one, the Confirmed Infectious will not increase, causing no new threats to the health system. Then the S/N factor will not decrease substantially, and a full SIR model is not necessary. The decay can be modeled by

$$\begin{aligned} C_{n+1} - C_n &= b_n I_n \\ R_{n+1} + D_{n+1} - R_n - D_n &= \gamma_n I_n \\ I_{n+1} &= C_{n+1} - R_{n+1} - D_{n+1} = I_n + b_n I_n - \gamma_n I_n \end{aligned}$$

using estimates of b_n and γ_n. This will for constant b and γ always be trivial, because the Infectious decay exponentially like

$$I_{n+j} = I_n (1 + b - \gamma)^j$$

and change the Confirmed C and the Removed $D + R$ accordingly. It takes

$$j = \frac{\log(I_n)}{-\log(1 + b - \gamma)}$$

steps to bring the Confirmed Infectious down to 1. Making $R_0 \approx \frac{b}{\gamma} < 1$ small is not the best strategy. Instead, one should maximize $\gamma - b$.

But, so far, the above argument cannot replace a SIR model. It only interprets the available data. However, monitoring the Johns Hopkins data in the above way will be very useful when it comes to evaluate the effectivity of certain measures taken by politicians. It will be highly interesting to see how the data of Figure 6 continue.

2.4 Extension Towards a SIR Model

From here on: open ended, and in progress, not yet ready for publication.

References

[1] An der Heiden M, Buchholz U:
Modellierung von Beispielszenarien der SARS-CoV-2-Epidemie 2020 in

Deutschland,
DOI 10.25646/6571.2
https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus
/Modellierung_Deutschland.pdf

[2] Christian Bommer & Sebastian Vollmer,
Average detection rate of SARS-CoV-2 infections is estimated around six percent,
www.uni-goettingen.de/de/document/download/
ff656163edb6e674fdbf1642416a3fa1.pdf
/Bommer%20&%20Vollmer%20(2020)%20COVID-19%20detection%20April%202nd.pdf

[3] Jonas Dehning, Johannes Zierenberg, Paul Spitzner, Michael Wibral, Joao Pinheiro Neto, Michael Wilczek, and Viola Priesemann,
Inferring COVID-19 spreading rates and potential change points for case number forecasts
arXiv:2004.01105v2

[4] Hethcote HW: The mathematics of infectious diseases.
SIAM review 2000;42(4):599-653

[5] COVID-19 repository at GitHub,
https://github.com/CSSEGISandData/COVID-19
/tree/master/csse_covid_19_data/csse_covid_19_time_series

[6] Robert-Koch-Institut, SARS-CoV-2 Steckbrief zur Coronavirus-Krankheit-2019 (COVID-19),
dated 24.4.2020
https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Steckbrief.html
#doc13776792bodyText2

[7] Estimates of the severity of coronavirus disease 2019: a model-based analysis,
www.thelancet.com/infection
Published online March 30, 2020
https://doi.org/10.1016/S1473-3099(20)30243-7