Simulating Epidemics with a Probabilistic Toy Model

Bernd Kärcher

Abstract

During the coronavirus pandemic, a simplistic model has been developed as an attempt of a non-specialist scientist in lockdown to better understand the evolution of infectious diseases. The model is based on Monte Carlo simulation and applied to a closed, homogeneous population. Basic characteristics of epidemics dynamics are analyzed and compared to results from a deterministic model. Sensitivity studies corroborate key recommendations of public health officials.

Contents

Chapter 1

Situation

This report was conceived and written as the coronavirus (Covid-19) pandemic unfolded in the year 2020, sending some to intensive care units (that includes UK's prime minister) and trapping others in their homes (including HM The Queen, for that matter, who self-isolated at Windsor Castle). Most members of the Institute of Atmospheric Physics (IPA) at DLR telework from March 9 to April 19 and as I write these lines, it is unclear when we will be able to return to our offices. The public media announced the most important Dos and Dont's required to 'flatten the curve', meaning to keep the number of infected people as low as possible to prevent putting too much stress on the healthcare system.

This spurred my interest to learn more about the mathematics of epidemics. My goal was to explore how viruses spread with simple means, but based on scientific principles. Out of curiosity and without the slightest idea about the modeling machinery available in epidemiology, I developed a toy model that describes how a given population affected by a virus outbreak evolves over time. I picked a probabilistic framework using the Monte Carlo method that is well-suited to tackle problems with statistical behavior. Variability in spatial movements, infection risk, and recovery time of individuals are captured by random processes with prescribed probability distributions.

My simulations are based on population data corresponding to Wessling in Upper Bavaria, Germany, for a number of reasons. Firstly, DLR is located in Oberpfaffenhofen, a village that is part of Wessling. Secondly, Wessling, where many who work at DLR live, has a population density that is practically identical to the national average. Thirdly, as I had no ambition to write the most computationally efficient computer code and deliberately relied on rather modest computer resources, simulations based on the number of residents in Wessling could be carried out quickly.

I am neither a physician nor an epidemiologist and have no background in biology. I am discussing a simplistic model playing with a limited population.

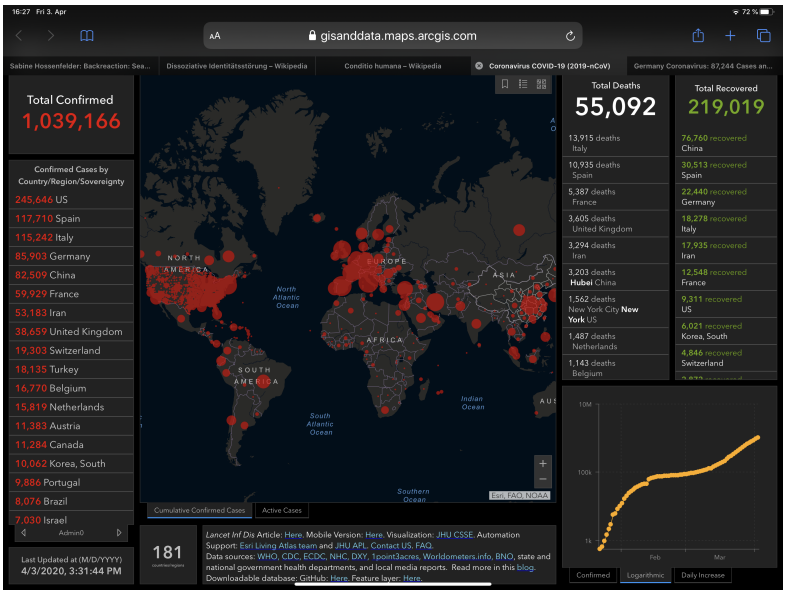


Figure 1.1: Statistics about Covid-19 provided online by the Johns-Hopkins-University on April 3, 2020.

Of course, picking Wessling as the playground of my model is largely a tongue-in-cheek decision. *The simulations discussed here do not represent the actual conditions and behavior of residents of Wessling.* Professionals employ much more sophisticated models that fall far less short from simulating real epidemics and can be used to make sensible predictions.

Looking at actual data is safe only in hindsight, after careful analyses. For example, the numbers shown in Fig. 1.1 for Germany differed from those officially reported by Germany's public health institute, presumably because the Baltimore-based Johns-Hopkins-University based their output mainly on non-official, online media sources.

On the one hand, interpretation of data sampled during an epidemic can easily be misleading due mainly to the lack of representative testing to identify the infectious, and, as in the case of Covid-19, when there are many asymptomatic cases that escape detection. The data themselves are not perfect. For example, the true number of infected people is difficult to estimate. For this reason, it is so important to have guidance from scientists and public health professionals to make sense out of real-time statistical data. On the other hand, I was surprised to see how well my toy model reproduced key features of the dynamics of an epidemic. I thought it would be interesting to find out whether this is also true concerning the outcome of mitigation options.

Chapter 2

Probabilistic Model

The purpose of this chapter is to introduce the model without resorting to mathematical detail. I rely on the conceptual framework of the well-established SIR-model. The basic version of this classical non-probabilistic (deterministic) model developed almost a century ago is briefly described in Appendix A. Results of both approaches are compared in Appendix B.

People are categorized into three compartments: S stands for members of a population, the susceptibles, capable of contracting an infection. The symbol I denotes infectious cases, and R indicates those recovered from the infection. Probabilistic approach including basic model assumptions and movements of and contact processes between humans are described next.

In the model, populations are homogeneous, i.e. all persons in a given compartment are equally susceptible. There is no delay time between becoming infected and being infectious in the model.[1] Moreover, recovered members stay immune, as typical for viral infections.[2] Immunity is achieved instantly once an I-member is transferred to the post-exposure R-compartment, which are not allowed to affect S-members. At this point, it is not known if Covid-19 confers long lasting immunity. Instead of separately specifying those who do not recover, but, sadly, die after getting sick, this group is subsumed in R, in which case the abbreviation can be thought of as short for removed.

Many variants of compartmental epidemiological models exist. My model ia basically a probabilistic SIR-model variant. In addition, I introduce the quarantined, Q. Members of this compartment are former I-members that have been identified after infection and then isolated after a time delay that can be set arbitrarily. By contrast, the R-compartment is filled by removing I-members after a time delay that is not arbitrary, but matches the time people need to

[1] Effects of a nonzero incubation time can be simulated by assigning S-members to an exposed compartment who are transferred to I when the latent period terminates.

[2] For bacterial infections, R-members can transform back to S due to the lack of immunity.

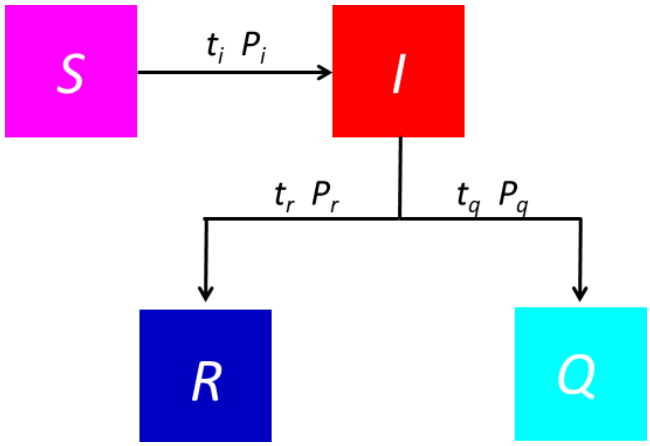


Figure 2.1: Compartments and parameters of the probabilistic model developed in this study. It includes submodels for movement of individuals and contact between them potentially leading to infection. For the meaning of symbols, see text or list of symbols.

recover (or die). Figure 2.1 schematically depicts the four model compartments and the flow between them along with key model parameters.

The number S+I+R+Q stays constant over time, t, and equals the total population, N_t. This implicitly assumes that natural births and deaths are neglected over the course of a simulation (3 months). All members of the population are allowed to roam freely within the computational domain, but are not allowed to leave it.[3] A possible net inflow of people into the computational domain is also neglected. The population initially contains only few infected cases in I, N_i, hence, $N_t - N_i$ susceptibles. Once moved to R, humans are assumed to be permanently immune, hence there is no $R \to S$ flow. The Q-compartment is initially always empty.

The rate of getting sick depends on movements of S-I-pairs. Here, I apply the following contact process submodel: when a susceptible person is closer than a certain infection distance, δ, to someone already infected, there is some probability, P_i, that S contracts the virus due to close proximity to I. I resort on the idealization that P_i is temporally and spatially constant. Clearly, a more socially engaged population is represented by a smaller δ-value and P_i depends on hand-washing, face-touching, and the like.

The age since infection, τ, is recorded for each I-member, as the time elapsed after an infection occurred. At the infection age $\tau = t_r$, infected persons will recover or die and will thus be moved into the R-category. As a sort of more realistic vital dynamics, I prescribe t_r as an expectation value of an exponential

[3] see Eagles, Hotel California, Asylum Records [1977].

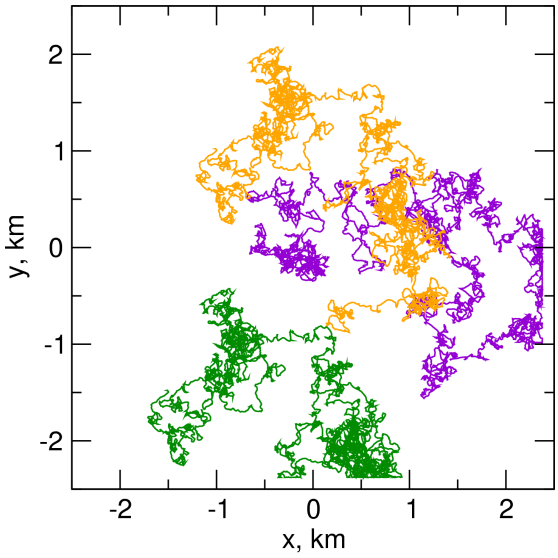


Figure 2.2: Three autocorrelated random walks over 90 days within the computational domain about 2.5 km wide in each direction.

distribution. An I-member is accepted as being recovered if a random number $0 \leq r \leq 1$ exceeds the probability $\exp(-\Delta t/t_r)$ within the time step, Δt. Hence, the transition $I \rightarrow R$ occurs after t_r only on average, it may happen earlier or later.

I-members might also be transferred to the Q-compartment.[4] Persons are quarantined with a probability P_q exactly t_q days after contracting the virus; this should represent a strategy to identify (test) and isolate infectious cases after the time lag t_q. Here, P_q takes into account that some fraction of infectious cases escapes identification. Clearly, less stringent quarantine ($P_q < 1$) and long waiting times for (or insufficient) testing degrade mitigation.

To track individual movements in the movement submodel, each member of the population is described by a pair of spatial coordinates, $\{x, y\}$, within a quadratic domain with area A and side length $L = \sqrt{A}$. Here, $A = N_t/\mathcal{P}$ is estimated from a prescribed population density, $\mathcal{P}$, and the total population, N_t. Starting locations of S- and I-members are randomly distributed. Roaming patterns are assumed to be random walks in both directions.[5]

As illustrated in Fig. 2.2, each member moves with random speeds normally distributed around zero for both directions. The random walks are characterized by the speed spread (standard deviation), σ. When applying a normal, or Gaus-

[4] While known as an efficient strategy to eradicate epidemics, I do not consider additional tracking and isolation of S-members that have been in close contact with I.

[5] This approach cannot simulate the effect of locations that people regularly and frequently visit. Such hubs play an important role in accelerating the spread of diseases.

sian, distribution, values much smaller than σ are picked much more often than those greater than σ. Values $> 3\sigma$ happen to be exceedingly rare. As most real people do not walk completely random (without preferred direction), I further assume that individual steps are autocorrelated over a time t_c, corresponding to correlated distances of on average $d_c = \sigma/t_c$.

Walks are advanced with a time step, $\Delta t = \delta/\sigma$, chosen such that one walk path length does not fall below δ, but only for the average speed (σ). This means that for encounters that occur with speeds $< \sigma$, subsequent steps may still find susceptibles within the infection region, so that these persons are exposed to an infection risk multiple times, while those walking with speeds $\geq \sigma$ enter a given infection region only once per encounter. While I do vary P_i, I keep δ, hence Δt, constant in all simulations. Otherwise, the number of susceptibles within and outside of an infection distance would vary, preventing a direct comparison of results from various sensitivity studies.

While not entirely unrealistic, modeling infections upon contact in this way is crude. A more physically-based submodel for the contact process needs to take into account contact time and virus transport via respiratory droplets in a cough or sneeze. These particles can linger in the air for several minutes and travel up to 8 m. Clearly, this added complexity quickly complicates matters and there is much room for improvement in this simplistic submodel.

Travel of S, I, and R is restricted to take place only inside the computational domain. This is achieved by applying reflective boundary conditions such that speeds reverse their signs once one of the domain boundaries would be crossed. As soon as persons are assigned to the Q-compartment, they no longer walk inside the domain.

Although the simulations are probabilistic in nature, I discuss only one statistical representation for each of them, for brevity. This is an important limitation. In a serious scientific study, it would be important to carry out repeated simulations with variations of random variables (ensemble simulations) to generate a distribution of possible outcomes. One could then analyze either ensemble statistics (at a given time across all simulations of an ensemble) or ergodic statistics (over time for each simulation of an ensemble) and decide with the help of appropriate hypotheses tests whether two scenarios are really different due to altered input parameters and not just different by chance.

During the Covid-19 epidemic, reproduction numbers were at times mentioned in media reports. The basic reproduction number, $\mathcal{R}_0 > 1$, is defined as the average number of secondary (new) infectious cases caused by a single primary case (while the latter is infectious) in a fully susceptible population. It describes the exponential growth of the infection in the initial phase of an epidemic and is used to estimate how infectious diseases are transmitted and to define appropriate intervention or mitigation strategies. Appendix A provides a more quantitative interpretation of $\mathcal{R}_0$.

Chapter 3

Applications

3.1 Scenario overview

The purpose of sensitivity studies that will be compared to a baseline scenario is to explore how changes in model parameters related to health intervention or mitigation options alter the time evolution of an epidemic. Let me once again emphasize that in no case do the simulations represent real world Wessling conditions at any time.

I designed a baseline scenario (B) and two sensitivity studies with higher risk of infection (R) and quick containment of infectious persons (Q).

The complete set of parameter values for all scenarios is given in Tab. 3.1. Population density, $\mathcal{P} = 240\,000/\mathrm{km}^2$, and total population, $N_t = 5\,476$, have been representative for Wessling in 2018, corresponding to a domain area of $A = 22.6\,\mathrm{km}^2$ ($L = 4.76\,\mathrm{km}$). For comparison, the mean free path per resident is about 6 m, estimated from the average area per person, $A/N_t = 41.3\,\mathrm{m}^2$. Residents walk randomly with an autocorrelation time of 30 min, at least during the virtual epidemics, giving a correlated path of length $d_c = 21\,\mathrm{m}$ for the typical speed of 1 km/d.

	t_c, min	σ, km/d	δ, m	t_r, d	N_i	P_i	P_q	t_q, d
B	30	1	2	10	5	0.2	0	∞
R						0.4		
Q							0.2	5

Table 3.1: Parameter values used in the simulations. For the sensitivity studies R and Q, only parameter changes relative to the baseline scenario B are listed.

Each scenario works with 5 initial infected residents placed randomly within the domain. Values reported in the media for the recovery time (up to about 17 days) varied. The value 10 d used here is guided by the rule established during Covid-19 that persons who had contact with infected people were told to self-isolate for 2 weeks.[1] Recall that the model uses a probability distribution for the recovery time with t_r as the mean value. An infection probability of 0.2 is a plausible guess; I have not found reliable information to better constrain this parameter for Covid-19. AN interaction distance of $1.5 - 2$ m has been recommended by epidemiologists for this virus.

Although it would be instructive, I did not carry out simulations with changed N_i (more infectious cases at the beginning accelerate the epidemic) or σ (faster travel spreads the virus more aggressively). In the same vein, longer infection periods (longer t_r) lead in total to more infected persons. All simulations run up to 90 days with a time step of 2 min.

3.2 Baseline Scenario

I analyze scenario B with the help of Fig. 3.1 and deduce first implications for Wessling. This scenario does not include quarantine measures.

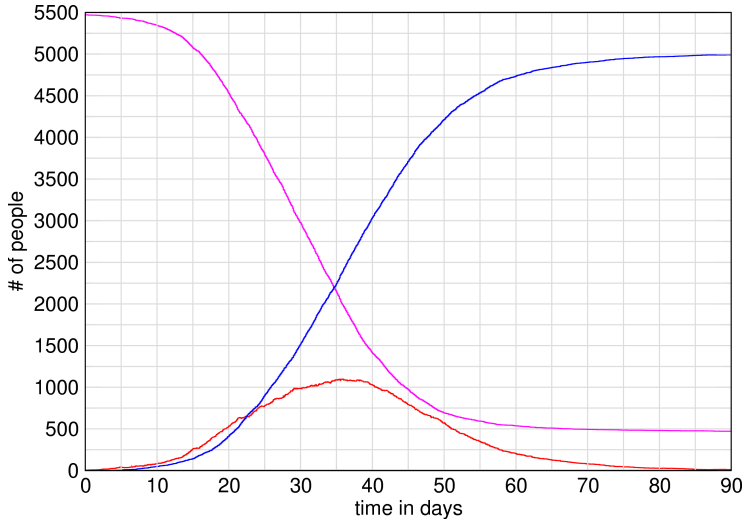


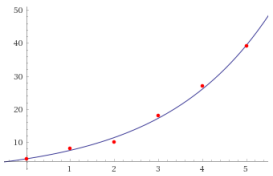
Figure 3.1: Temporal evolution of occupation of model compartments S, I, and R for the baseline scenario.

First focus on the solid curves. The number of susceptibles decreases over the

[1] On April 6, the country with the lowest fraction of infected people was Vatican City: 0.88% based on 7 confirmed cases in a total population of 799. Clearly, the isolation strategy worked well there.

course of the epidemic and approaches a constant value (472) at the end of the simulation. Therefore, about 8% of the residents would need vaccination to immunize to be prepared for the next outbreak. The number of infectious cases rises quickly initially, levels off taking a maximum[2] of 1 101 at day 36, and then approaches zero. Hence, at the peak of the epidemic, about 20% of the residents were infected. The removed compartment is filled over time reaching 4 992 after 90 days.

The above is in good general agreement with the predictions of the deterministic model outlined in Appendix A. For this reason, I fitted the simulated $I(t)$-curve to an exponential function during the first 5 days and obtained a net growth rate of 0.410166 per day. (The inset shows the fit as a blue curve and the data as red circles.) This yields a basic reproduction number $\mathcal{R}_0 \simeq 5$, somewhat above the value currently estimated for Covid-19. The rate of infection per I-member is $\alpha t_r \simeq 5/\text{d}$. The time to infect one susceptible individual by one infected person is initially $1/(\alpha S_0) \simeq 2$ days; for comparison, I recall that an I-member recovers on average in $t_r = 10$ days.



3.3 Sensitivity Studies

Health risk changes are modeled by doubling the probability of infection in scenario R. Results are shown in Fig. 3.2.

Compared to B, scenario R assumes a probability of infection that is twice as large than in the baseline case. As a result, the number of susceptibles falls off more quickly than in case B. At the same time, the peak number of infected individuals occurs about 10 days earlier and is larger by about a factor of 2. However, roughly 500 more people have been infected than in B. People recover earlier and after 90 days, there are practically no susceptible and infected persons left.

I have also carried out a simulation halving P_i (not shown) and found that this stopped the epidemic early in its tracks. In the context of my model, this sensitivity calls for an improvement of the manner contacts between susceptible and infectious are modeled. Regardless, being more conscious of better hygiene has a strong effect on the spreading of the virus.

[2]Obviously, this point indicates the time where available health care facilities are most likely to be overloaded.

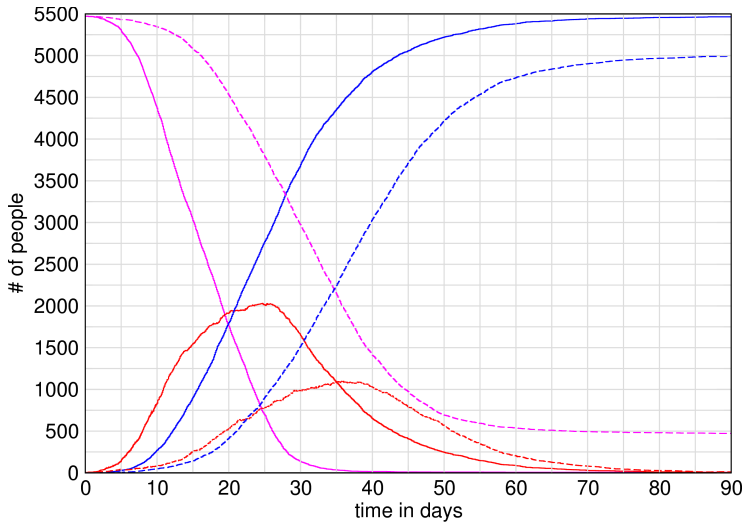


Figure 3.2: Temporal evolution of S, I, and R for the increased risk scenario (solid curves). The baseline scenario is repeated from Fig. 3.1 (dashed).

This comparison clearly highlights the sensitivity of epidemic dynamics on the infection probability. By the same token, I expect to see a similarly strong sensitivity on the interaction distance. I have now convinced myself about the great importance of physical distancing relative to other precautions such as wearing gloves.

Scenario Q differs from B by removing infected individuals, with a probability of 0.2, 5 days after they have been infected. This may be judged as a rather mild containment measure, but in view of practical difficulties to safely identify and isolate infectious persons, perhaps not unrealistic. In this simulation, quarantine happens from day 0. In reality, such a step would be most likely taken only after some period of time into the epidemic, e.g. at the instant when the number of infected persons reached a peak. Results are shown in Fig. 3.3.

The simulation ran out of infected people and stopped on day 83. In this case, quarantine did bring their peak number down dramatically, from 1 101 to 305. A total of 2 327 susceptibles escaped infection, thanks to 1 938 quarantined and 1 211 removed residents. Clearly, the identify-and-isolate strategy can be very powerful in attempts to 'flatten the curve'.

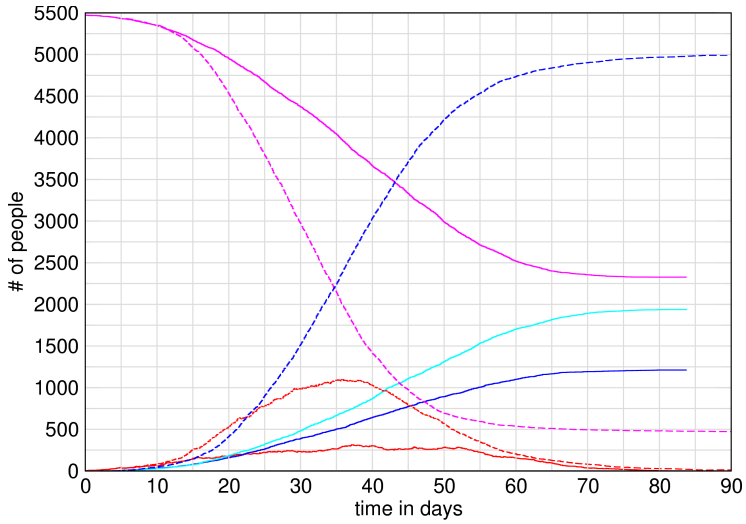


Figure 3.3: Temporal evolution of S, I, R, and Q for the scenario with quarantine (solid curves). The baseline scenario (dashed, with $Q = 0$) is repeated from Fig. 3.1.

Chapter 4

Epilogue

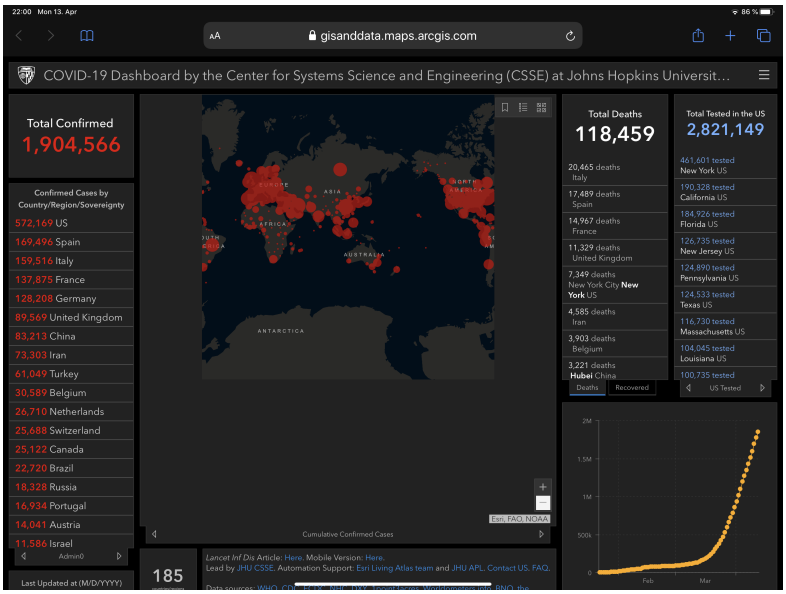


Figure 4.1: Statistics about Covid-19 provided online by the Johns-Hopkins-University on April 13, 2020.

Despite of the limitations of my simplistic approach, I felt that learning in this way is better than solely relying on news headlines. On April 13, the pandemic is still in an exponential growth stage (Fig. 4.1). Few active cases are seen in China and East Asia; however, uncertainty remains whether this region is on knife edge between recovery and recurrence. In Germany, public health officials noted with caution that they see signs of relief, but lockdown rules will remain in place in one way or another for months to come. These are trying times.

On a positive note, one outcome of the pandemic might be that we will have learned to better cope with 'not knowing' and nonetheless live our lives fully. (After all, in physics we know that uncertainty is built into the very fabric of nature.) And it is my hope that when all this passed, we question some of the habitual patterns and aspects of our 'way of life' that seemed to be set in stone.

Although it goes against the grain of scientifically-oriented minds, I refrain from tackling model issues, notably conducting ensemble simulations to add a statistical analysis or developing less idealized human movement and contact submodels to prevent this study from becoming a time sink.

Regardless, this transient episode kicked off by a microbe in one single infected person might sharpen our senses of how delicately all things work together to create the world we inhabit. Right now, we have first-hand experience that real exponential growth does occur in nature, not in economics. At the end of the day, we all may want to reflect on how we organize our societies, and on the true values of well-knit social security nets, well-kept public health care systems, and thriving scientific research communities.

Bernd Kärcher
Munich, April 2020

List of symbols used in the main text

d_c	correlated distance for random walk
t_c	autocorrelation time for random walk
t_i	average infection time
t_q	time lag for quarantine
x, y	horizontal coordinates
t	time
A	domain area
I	infectious
L	domain length
N_i	initial number of infected persons
N_t	total population
N_v	initial number of immune persons
Q	quarantined
R	removed (recovered, dead)
P_i	probability of infection
P_q	probability of quarantine
S	susceptibles
δ	pair interaction distance
Δt	time step
$\mathcal{P}$	population density
$\mathcal{R}_0$	basic reproduction number
σ	speed spread for random walk
τ	infection age

Appendix A

Deterministic Model

The model discussed in the main text is a version of a stochastic compartmental model. For large populations one quickly needs to resort to high performance computing and specialized programming techniques. The model outlined next is an example of a deterministic compartmental model. Since it treats individuals as continuous functions, it is suitable for sufficiently large populations.

The mathematical groundwork of the deterministic SIR-model was laid out in a series of 3 articles authored by Scottish epidemiologist Anderson G. McKendrick and biochemist William O. Kermack that appeared in the *Proceedings of the Royal Society* in 1927, 1932, and 1933. In these landmark studies, infectious cases evolve as a function of time, t, and infection age τ, using infection age-dependent transmission ($S \to I$) rates.

In its basic form without age structure, the model consists of three non-linear ordinary differential equations:

$$\frac{dS}{dt} = -\alpha SI \tag{A.1}$$

$$\frac{dI}{dt} = \alpha SI - \gamma I \tag{A.2}$$

$$\frac{dR}{dt} = \gamma I\,; \tag{A.3}$$

here, α quantifies the fraction of S that gets infected per I-person per unit time. As shown below, it may be estimated from data taken during initial (linear) phase of an epidemic; γ is the removal rate. The binary S-I-term makes the equations nonlinear; analytical solutions $\{S(t), I(t), R(t)\}$ do not exist in closed form. However, numerical solutions for $S(t)$ and $I(t)$ can be obtained using standard integration methods; for a closed population, R follows from the conservation law $R(t) = N_t - [S(t) + I(t)]$.

This set of equations with constant rate coefficients α and γ for a closed, homogeneous population can only describe the evolution of simple epidemics, whereby $S(t)$ decreases monotonically. Its relevance stems from the fact that it provides basic insight into the system dynamics and reveals a threshold condition for an epidemic to start and an estimate of the number of people ultimately affected by an epidemic in terms of the basic reproduction number.

Multiplying α with the time a person needs to recover from the infection yields the infection rate per I-member, $z = \alpha/\gamma$. Therefore, multiplying z with the initial number of susceptibles, $S_0 \equiv S(t = 0)$, gives the dimensionless **basic reproduction number, the number of individuals getting infected by one infectious person at the point when the epidemic has not yet spread** (mathematically in the limit $t \to 0$):

$$\mathcal{R}_0 = zS_0 = \frac{\alpha S_0}{\gamma} \,. \tag{A.4}$$

Shortly after an epidemic starts, $R(t)$ can be ignored. Additionally, if $S_0 \gg I_0$, which is typically the case, $S \simeq S_0$, so that $dI/dt \simeq \gamma(\mathcal{R}_0 - 1)I$. This means that $I(t)$ starts growing exponentially at the rate $\gamma(\mathcal{R}_0 - 1)$, but only if $\mathcal{R}_0 > 1$; otherwise, $I(t)$ declines and an epidemic does not occur. Since $S_0 \geq S(t)$, this is the fastest growth rate of $I(t)$ during the epidemic simulated by this model.

Since neither α nor $\mathcal{R}_0$ are known, the exponential onset of $I(t)$ can be used to fit $I(t)$-data (or the result of one of my simulations for that matter) at the onset of an epidemic to the exponent of the exponential solution of the form $I(t) = I_0 \exp(rt)$. Such a fit for r yields reasonable estimates, $\alpha \simeq (r + \gamma)/S_0$ and $\mathcal{R}_0 \simeq 1 + r/\gamma$.

$\mathcal{R}_0$ is also important for the **herd immunity, the threshold fraction of immune members in a population that needs to be surpassed to prevent an outbreak**. The associated necessary condition, $\mathcal{R}_0 < 1$, is written in terms of the initial number of immune persons, R_0:

$$R_0 > N_t - I_0 - \frac{1}{z} \,, \tag{A.5}$$

using the conservation property $S_0 = N_t - (I_0 + R_0)$. Again with the approximation $I_0 \ll N_t$ (consistent with $S_0 \simeq N_t$) and written in terms of a fraction of the total population, $\mathcal{H} = R_0/N_t$, the threshold for herd immunity is given by

$$\mathcal{H} \simeq 1 - \frac{1}{\mathcal{R}_0} \,. \tag{A.6}$$

For Covid-19, initial estimates have been around $\mathcal{R}_0 = 3$, hence, $\mathcal{H} = 2/3$. This simple relationship is intuitively clear: if one infected person transmits Covid-19 to 3 others, more than 2 out of 3 people ($> 66\%$) need to be immune to stop spreading the virus.

In the case of measles, $\mathcal{R}_0 \simeq 12 - 18$, hence, $\mathcal{H} \approx 0.93$. This means that more than 93% of the population must be immune to measles to eradicate

this disease. In such cases, herd immunity can arguably only be achieved by widespread vaccination. In case a virus already spreads, those who are not immune benefit from vaccination of others, but can potentially carry an infection in them, causing an outbreak to recur or unfold elsewhere. At the time I am writing this report, it is expected that a fully tested vaccine for Covid-19 will not be available for another year.

Using the tendency equations for S and I, a phase space solution $I(S)$ is obtained by solving

$$\frac{dI}{dS} = \frac{1}{zS} - 1 \; : \tag{A.7}$$

$$I(S) = (N_t - R_0) + \frac{1}{z} \ln\Big(\frac{S}{S_0}\Big) - S \, . \tag{A.8}$$

The two latter equations allow to estimate the peak number of infectious persons. In mathematical terms, one searches for the value of S, $\hat{S}$, where I takes its maximum, $\hat{I}$. This value is $\hat{S} = 1/z$, giving $\hat{I}$ as:

$$\hat{I} = (N_t - R_0) - \frac{1}{z}(1 + \ln \mathcal{R}_0) \, . \tag{A.9}$$

$I(S)$ can also be used to estimate the fraction of the population that does not get infected when the epidemic is over (for $t \to \infty$). This fraction is given by $1 - S_\infty/N_t$, with $S_\infty \equiv S(t_\infty)$. It follows immediately from $I(S)$ with $R_0 = 0$ applied at t_∞ and noting that in this limit, no infectious are left ($I_\infty \to 0$):

$$1 - \frac{S_\infty}{N_t} = \frac{1}{\mathcal{R}_0} \ln\Big(\frac{S_0}{S_\infty}\Big) \, . \tag{A.10}$$

A nontrivial solution $S_\infty > 0$ may be computed iteratively from this transcendental relationship known as final size equation (approaching S_∞ as the final size of an epidemic is a characteristic of models with closed populations). It follows that S_∞ must be non-zero, i.e. some susceptibles never get infected, instead the epidemic runs out of infections, a finding that is not immediately obvious.

I examine the difference between the above deterministic and my probabilistic model in Appendix B. Beyond these, other methodologies are employed by public health officials and research scientists, such as stochastic approaches that construct discrete time Markov chains related to equations of a deterministic epidemic model. Epidemiologists also work with data-heavy statistical and diagnostic models.

Appendix B

Probabilistic vs deterministic approaches

Unlike the deterministic SIR model, the stochastic processes (i) spatial encounters between susceptibles and infectious and (ii) recovery or death after exposure are modeled via Monte Carlo simulation in my probabilistic model.

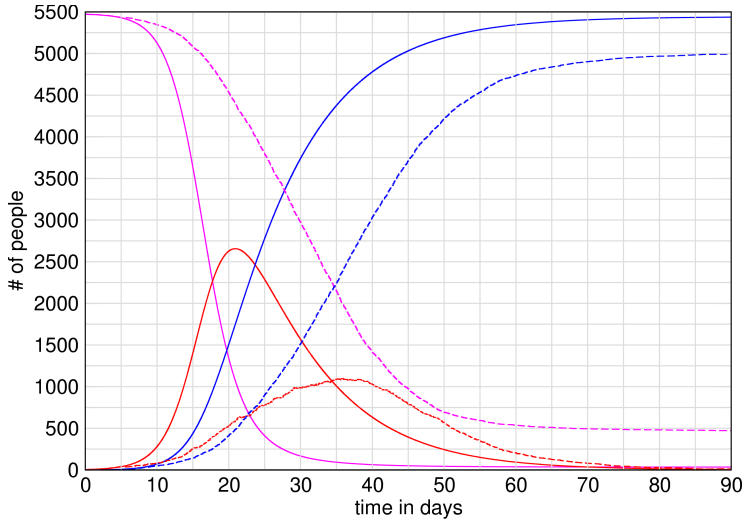


Figure B.1: SIR-solutions for the deterministic (solid curves) and probabilistic baseline scenario (dashed). The latter is repeated from Fig. 3.1.

I compare in Fig. B.1 the deterministic solution of the epidemic model for the baseline scenario with the probabilistic solution discussed in Ch. 3.2, whereby the initial conditions used to compute the former are taken from the latter as de-

scribed in Appendix A: $S_0 = 5471, I_0 = 5; dS/dt(t{=}0) = -\alpha S_0 I_0$, $dI/dt(t{=}0) = (\alpha S_0 - \gamma) I_0$, with $\alpha \simeq (r + \gamma)/S_0$ and $\gamma = 1/t_r$.

The general evolution of both approaches is similar. However, the solutions diverge significantly after five days, although the initial conditions are identical. This should not come as a surprise, for two reasons:

- in the deterministic model, α and γ are constant, while in the probabilistic model, α is computed from randomized contact processes and γ is sampled from an exponential distribution;
- the deterministic solution is uniquely determined by the initial conditions, while each solution of the probabilistic model with randomized process representation will be different.

The preferred way would be to compare the deterministic solution with an ensemble mean solution of my model, whereby the ensemble consists of many tens (> 30) of individual statistical representations of scenario B solutions.