SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

The SIR Model: Understanding the Spread of Disease

av

Loïs Veen

2020 - No K10

The SIR Model: Understanding the Spread of Disease

Loïs Veen

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Yishao Zhou

2020

Abstract

The SIR (Susceptible-Infected-Recovered) model is an epidemiological model used to estimate the spread of infectious diseases. This paper aims to provide the reader with a description of the model and it's applications, using both mathematical theory and real world data. The first part gives some contextual background and explains the mathematics behind the model. Numerical solutions to the model are then discussed using the methods of Euler and Runge-Kutta. This is followed by a qualitative analysis of the model, where concepts such as epidemic threshold, equilibrium, and epidemic size are investigated. We then study ways of analyzing the effects of vaccination and explain how the World Health Organization (WHO) in 1980 managed to eradicate smallpox. In the last part of this paper we use the newly collected data on COVID-19 to estimate the disease's basic reproduction number R_0, as well as to investigate the effects of public health measures.

Acknowledgements

I wish to express sincere gratitude to my supervisor Yishao Zhou, for sparking my interest in the subjects of numerical analysis and dynamical systems, for providing me with lots of helpful ideas and feedback, and for always being so very quick to respond to my questions.

Contents

1 Introduction **5**
- 1.1 Epidemiology of infectious diseases . . . 5
- 1.2 Disease transmission . . . 6
- 1.3 History of the SIR model . . . 6

2 The SIR Model **8**
- 2.1 Parameters and variables . . . 8
  - 2.1.1 Susceptible - infected - recovered . . . 8
  - 2.1.2 Total population . . . 8
  - 2.1.3 Transmission rate . . . 8
  - 2.1.4 Recovery rate . . . 9
- 2.2 Conditions and assumptions . . . 9
- 2.3 Differential equations . . . 10
  - 2.3.1 The susceptible equation . . . 10
  - 2.3.2 The infected equation . . . 10
  - 2.3.3 The recovered equation . . . 10
- 2.4 Variations on the SIR model . . . 11

3 Numerical solutions **12**
- 3.1 The Euler method . . . 12
- 3.2 Forward Euler and the SIR model . . . 13
- 3.3 Runge-Kutta 4 and the SIR model . . . 14
- 3.4 Experimental error analysis . . . 15
- 3.5 Theoretical error analysis . . . 19

4 Qualitative analysis of the SIR model **21**
- 4.1 Basic- and effective reproduction numbers . . . 21
- 4.2 Epidemic threshold . . . 22
- 4.3 Equilibrium . . . 23

4.4 The size of an epidemic 24

5 Vaccination **27**
5.1 Immunization 27
5.2 Vaccination and the SIR model 28
5.3 Herd immunity and Smallpox 29
5.4 Stability of the equilibrium points 30
5.5 Vaccination and the size of an epidemic 33

6 Model fitting: COVID-19 **36**
6.1 Estimating R_0 for COVID-19 36
6.2 Simulating a COVID-19 epidemic in Sweden 40
6.2.1 Without social distancing 40
6.2.2 With social distancing 40
6.3 The effects of quarantine 43

7 Discussion **45**
7.1 Summary 45
7.2 Topics for future study 46

A Mathematical proofs **49**
A.1 Proof to theorem 5.4 49

Chapter 1

Introduction

The SIR (Susceptible-Infected-Recovered) model is a mathematical model consisting of three ordinary differential equations, used to estimate the spread of infectious diseases. In this paper we will describe the mathematics behind the model as well as use numerical methods find it's solutions. In addition, we will discuss the model's applications in global health strategies for preventing and treating infectious diseases.

1.1 Epidemiology of infectious diseases

Throughout history, infectious diseases such as the plague, cholera and smallpox have killed millions of people and wiped out entire populations. Infamous outbreaks include the Black Death in Europe during the 14th century, killing an estimated 25 million people in just five years, as well as the pandemic influenza of 1918-1919 killing around 40 million people globally. Still today, diseases such as malaria, HIV/AIDS, measles and tuberculosis come at the cost of millions of lives each year. According to UNAIDS, 34 million people were infected with HIV worldwide at the end of 2010. [1] Furthermore, it is estimated that infectious diseases account for up to 70% of all deaths in countries needing humanitarian assistance.[2]

It is clear that forecasting as well as controlling these diseases is of the utmost importance when it comes to improving global health. Ways of doing this include vaccination programs, antibiotics or antiviral medication, behavioural changes, and sanitary measures. However, in order to consider the

right strategies for each specific situation it is essential to understand the contagious dynamics of these diseases. [3]

Epidemiology is defined as *the study of the distribution and determinants of health-related states or events in specified populations, and the applications of this study to the control of health problems* [4]. In other words, it is the branch of science that deals with the quantitative analysis of disease occurrence in populations. When it comes to infectious diseases, epidemiologists use mathematical and statistical models to study rates, risk factors, and the effects of interventions.

1.2 Disease transmission

Infectious disease is defined as illness caused and transmitted by infectious agents such as a viruses, bacteria, fungi or parasites. Depending on the type of agent, transmission from host to host can occur through direct contact or indirect contact. Examples of transmission through direct contact include touch and exchange of bodily fluids, whereas examples of ways for indirect transmission include the air, animals or animal waste, contaminated objects and contaminated food or water. [5]

The rate at which transmission occurs is represented by the reproduction number (R), which stands for the average number of new infections per infected case. Varying across infectious agents, time and space, this number depends on factors such as the rate of contacts in the host population, the mode of transmission, the probability of infection being transmitted during contact and the duration of infectiousness. [6]

Understanding R enables the development of mathematical models aiming to predict and simulate disease outbreaks. An example of such a model is called the SIR model.

1.3 History of the SIR model

The first known mathematician to analyze infectious diseases was Daniel Bernoulli in 1766. [1] Using a model he developed, he estimated the total mortality of smallpox to be 1 in fourteen. He also showed that inoculation would add about three years to life expectancy at birth. In 1911, Ronald

Ross published a paper in which he formulated a mathematical model on the transmission dynamics of malaria. [7] He explored the relationship between the number of mosquitoes and the incidence of malaria, as well as the effect of intervention strategies. Based on his work, between 1927 and 1933 Kermack and McKendrick founded the compartmental epidemic modeling. [1, 8] This is a type of modeling where a population is divided into compartments, assuming those in the same compartment share specific characteristics. Their work suggested that the probability of infection of a susceptible individual was related to the number of contacts with infected individuals.

One of the simplest and most fundamental compartmental models is the so-called SIR model. [9] This model divides individuals of a populations into compartments based on whether they are susceptible, infected or recovered, and estimates rates of transmission between compartments. Using the SIR model, estimates could be made about size and duration of specific outbreaks, as well as the effectiveness of strategies such as vaccination and behavioural changes.

Chapter 2

The SIR Model

2.1 Parameters and variables

2.1.1 Susceptible - infected - recovered

The SIR model defines individuals as susceptible (S) if they are neither infected nor recovered, and therefore at risk of catching the disease. The ones infected (I) are those currently carrying the infection, which makes them potentially contagious. Individuals that are recovered (R) are those that were-, but no longer are, infected. This group is assumed to be immune and noncontagious. [8] Figure 2.1 shows a flow chart of the SIR model.

2.1.2 Total population

The total population size (N) is considered constant and equal to the sum of all three compartments at time t: $N = S(t) + I(t) + R(t)$.

2.1.3 Transmission rate

The transmission rate (β) is the number of individuals that will be infected by one infected individual per time unit, assuming that all contacts are susceptible. This number is the product of the number of contacts κ and the transmissibility τ of the disease, $\beta = \kappa\tau$. For an infectious disease, β will

always be positive. Also, the higher β, the more infectious the disease. [8, 10]

2.1.4 Recovery rate

The recovery rate (γ) is the rate at which infected individuals will recover, which is inversely proportionate to the average time period for infection ($\frac{1}{\gamma}$). For any disease, γ will be equal to- or greater than zero. [8, 10]

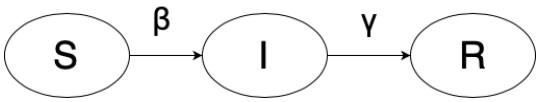


Figure 2.1: A flow chart of the SIR model.

2.2 Conditions and assumptions

The SIR model is based on a number of strong assumptions about the population and disease characteristics [11, 9, 10]:

- The population is large and closed.
- No natural births or deaths occur, as well as no immigration or emigration. The population size (N) is constant.
- The infection has no latency period, which means that infection immediately leads to an individual being infectious.
- Recovery occurs at a constant rate and gives an individual life-long immunity. The only way of requiring immunity is through infection, there is no inherited immunity.
- Individuals from all three compartments are homogeneously distributed across the entire population. This is called mass action mixing and means that each individual is as likely to encounter every other individual.

Comparison with similar but more elaborate models has shown that predictions made by the SIR model are strikingly reliable, despite the fact that many of it's assumptions are highly unrealistic. [10]

2.3 Differential equations

The SIR model consists of three ordinary differential equations (ODE) measuring how the size of each of the three compartments changes over time.

2.3.1 The susceptible equation

The number of susceptible individuals getting infected per time unit (Δt) is equal to the product of the transmission rate (β), the number of infected individuals ($I(t)$) and the fraction of susceptible individuals ($\frac{S(t)}{N}$). Thus, the change in the number of of susceptible individuals per time unit is given by

$$\frac{dS}{dt} = -\beta \cdot I(t) \cdot \frac{S(t)}{N} \tag{2.1}$$

2.3.2 The infected equation

The change in the number of infected individuals is given by the sum of the number of susceptible individuals getting infected minus the number of infected individuals recovering. The latter is given by the product of the recovery rate (γ) and the number of infected individuals ($I(t)$). Thus, the change in size of the compartment of infected individuals per time unit is given by

$$\frac{dI}{dt} = \beta \cdot I(t) \cdot \frac{S(t)}{N} - \gamma \cdot I(t) \tag{2.2}$$

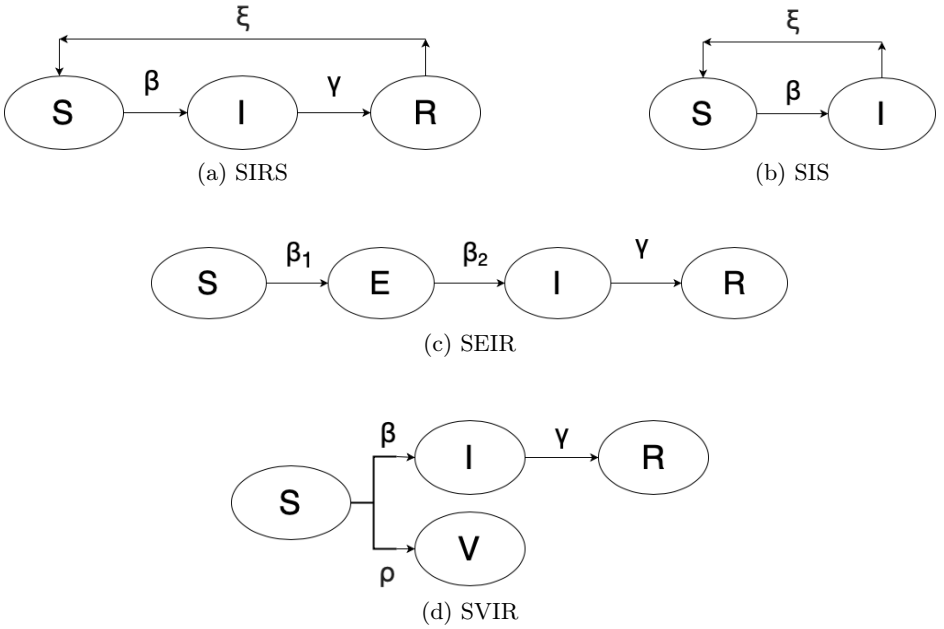
2.3.3 The recovered equation

The number of individuals recovering is given by the product of the recovery rate (γ) and the number of infected individuals ($I(t)$). Thus, the change in the number of recovered individuals per time unit is given by

$$\frac{dR}{dt} = \gamma \cdot I(t) \tag{2.3}$$

2.4 Variations on the SIR model

Figure 2.2 shows four possible variations on the SIR model. The SIRS model allows for the possibility of recovered individuals turning susceptible again at a rate of ξ. That is, infection leads to immunity but the immunity is not necessarily permanent. In the SIS model, infection does not lead to immunity. Therefore, there is no recovered class and all infected individuals return to the susceptible class at a rate of ξ. The SEIR model contains an exposed class, for those that have been exposed to the infection but are not yet infectious themselves. This means the model can take into account a potential incubation period. In the SVIR model, individuals from the susceptible class have the possibility to be vaccinated. This would give them immunity without needing to go through an infection.



(a) SIRS

(b) SIS

(c) SEIR

(d) SVIR

Figure 2.2: Variations on the SIR model.

Chapter 3

Numerical solutions

Since the SIR model consists of nonlinear equations, it is difficult to find analytical solutions. However, using numerical methods we can find approximate these solutions. [12] The numerical methods discussed in this paper are called the Euler method and the Runge-Kutta method. These methods divide time into intervals of length Δt and approximates the solutions at those times using the given differential equations. In order to improve readability, we will refer to the differential equations of the SIR model using $f(S, I) = \frac{dS}{dt}$, $g(S, I) = \frac{dI}{dt}$ and $h(I) = \frac{dR}{dt}$.

3.1 The Euler method

The Euler method is one of the most basic numerical methods and is geometrically very easy to understand. Using a given initial value y_n, the method approximates y_{n+1} by multiplying the length of the time interval $\Delta t = t_{n+1} - t_n$ with an estimated value for the rate of change $\frac{dy}{dt}$. [13] The way different Euler methods estimate this value is what distinguishes them. Forward Euler is an explicit method, since it uses the slope at the current state (t_n, y_n) to calculate the rate of change. Backward Euler is an implicit method which uses the slope at the later state (t_{n+1}, y_{n+1}) and Improved Euler, also known as Heun's method, uses the average of the slopes at the current- and later state. Figure 3.1 displays these three methods in one graph.

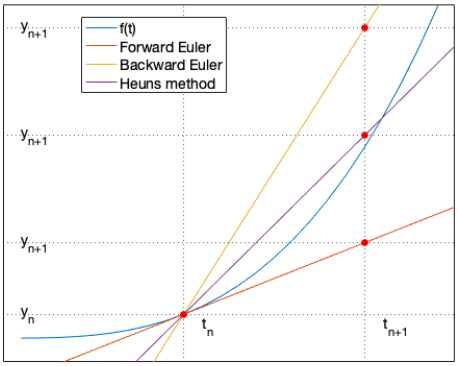


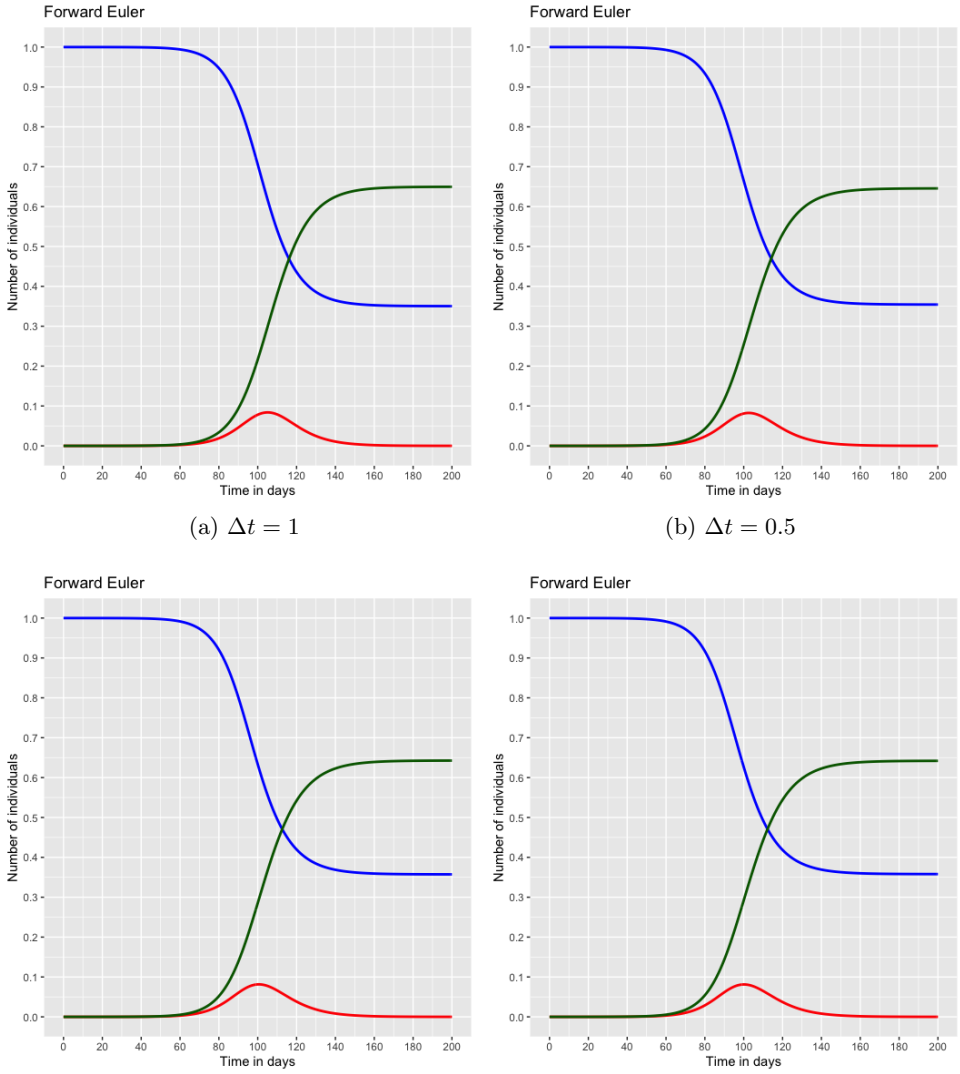
Figure 3.1: Illustration of different Euler methods.

3.2 Forward Euler and the SIR model

Applying Forward Euler to the SIR model gives us the following system of equations, where Δt stands for the step size measured in days:

$$\begin{cases} S_{n+1} = S_n + \Delta t \cdot f(S_n, I_n) = S_n \cdot (1 - \Delta t \cdot \frac{\beta \cdot I_n}{N}) \\ I_{n+1} = I_n + \Delta t \cdot g(S_n, I_n) = I_n \cdot (1 + \Delta t \cdot \frac{\beta \cdot S_n - \gamma \cdot N}{N}) \\ R_{n+1} = R_n + \Delta t \cdot h(I_n) = R_n + \Delta t \cdot \gamma \cdot I_n \end{cases}$$

From this system it is easily verified that tomorrow's number of susceptible individuals (S_{n+1}) always will be equal to-, or lesser than, today's number of susceptible individuals (S_n). In other words, for any t we have $S(0) \geq S(t)$. Likewise can we see that tomorrow's number of recovered individuals (R_{n+1}) always will be equal to-, or greater than, today's number of recovered individuals (R_n). In other words, for any t we have $R(t) \geq R(0)$. More specifically, if and only if there are no infected individuals ($I_n = 0$), we have $S_n = S_{n+1}$ and $R_n = R_{n+1}$. Figure 3.2 shows numerical estimates of the variables of the SIR model for different step sizes (Δt). The figures look almost identical.



(a) $\Delta t = 1$ (b) $\Delta t = 0.5$

(c) $\Delta t = 0.1$ (d) $\Delta t = 0.01$

Figure 3.2: Forward Euler for different step sizes, showing S (blue), I (red), and R (green).

3.3 Runge-Kutta 4 and the SIR model

Just as the different Euler methods, Runge-Kutta 4 (RK4) is an iterative method aiming to approximate the next value using the present value, the

step size, and the rate of change. However, instead of using the rate of change at one point, RK4 uses the weighted average of the rate of change at four points along the step. [13] The rate of change at such a point is called an increment. Applying RK4 to the SIR model gives us the following system of equations and increments:

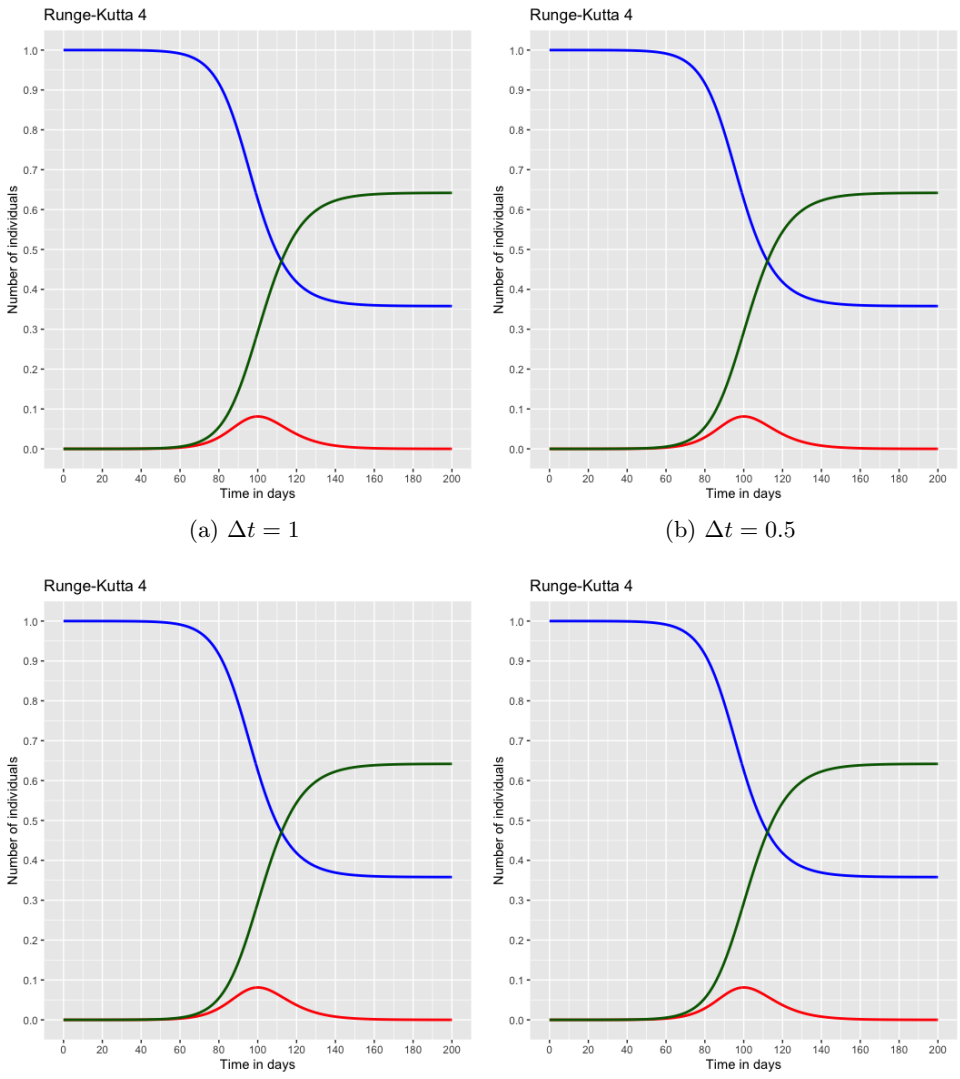
$$\begin{cases} S_{n+1} = S_n + \frac{\Delta t}{6}(k_1 + 2k_2 + 2k_3 + k_4) \\ \\ I_{n+1} = I_n + \frac{\Delta t}{6}(j_1 + 2j_2 + 2j_3 + j_4) \end{cases}$$

$$\begin{aligned} k_1 &= f(t_n, S_n, I_n) \\ k_2 &= f(t_n + \frac{\Delta t}{2}, S_n + \frac{\Delta t}{2}k_1, I_n + \frac{\Delta t}{2}j_1) \\ k_3 &= f(t_n + \frac{\Delta t}{2}, S_n + \frac{\Delta t}{2}k_2, I_n + \frac{\Delta t}{2}j_2) \\ k_4 &= f(t_n + \Delta t, S_n + \Delta t k_3, I_n + \Delta t j_3) \\ j_1 &= g(t_n, S_n, I_n) \\ j_2 &= g(t_n + \frac{\Delta t}{2}, S_n + \frac{\Delta t}{2}k_1, I_n + \frac{\Delta t}{2}j_1) \\ j_3 &= g(t_n + \frac{\Delta t}{2}, S_n + \frac{\Delta t}{2}k_2, I_n + \frac{\Delta t}{2}j_2) \\ j_4 &= g(t_n + \Delta t, S_n + \Delta t k_3, I_n + \Delta t j_3) \end{aligned}$$

Figure 3.3 shows numerical estimates of the variables of the SIR model for different step sizes (Δt). Again, the figures almost look identical.

3.4 Experimental error analysis

In this section we will compare Forward Euler and RK4 to each other for different step sizes. We choose our parameters and initial values to approximate those at the start of the Ebola outbreak in Liberia 2014. That is, $\beta = 0.32$, $\gamma = 0.2$, $N = 4000000$, $S(0) = (N-1)/N$, and $I(0) = 1/N$. Figure 3.4 shows the numerical simulation using RK4 with $\Delta t = 1$. Note that this is not a realistic simulation, since in real life the sizes of β and γ change as time progresses and society adapts to the epidemic. However, if that would



(a) $\Delta t = 1$ (b) $\Delta t = 0.5$

(c) $\Delta t = 0.1$ (d) $\Delta t = 0.01$

Figure 3.3: Runge-Kutta 4 for different step sizes, showing S (blue), I (red), and R (green).

not have been the case we see that by the end of the epidemic more than 60% of the total Liberian population would have been infected. Since Ebola is a very deadly virus with an approximated fatality rate of 50% [14], this

would have equaled a death toll of $0.6 \cdot 0.5 \cdot N = 1200000$ individuals. Again, this is by far not a correct representation of reality. The actual number of cases and fatalities in Libera were 10678 and 4810 respectively [15].

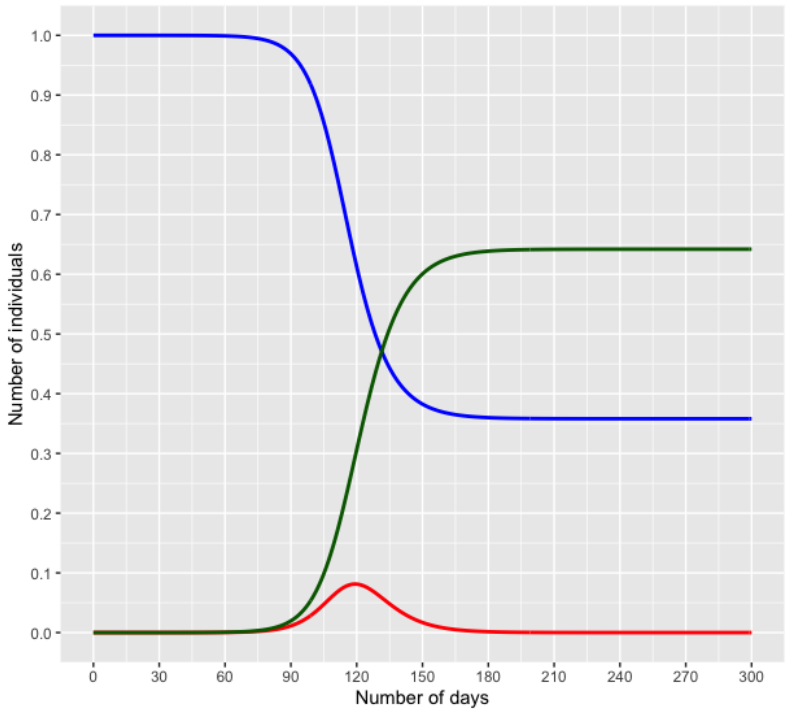
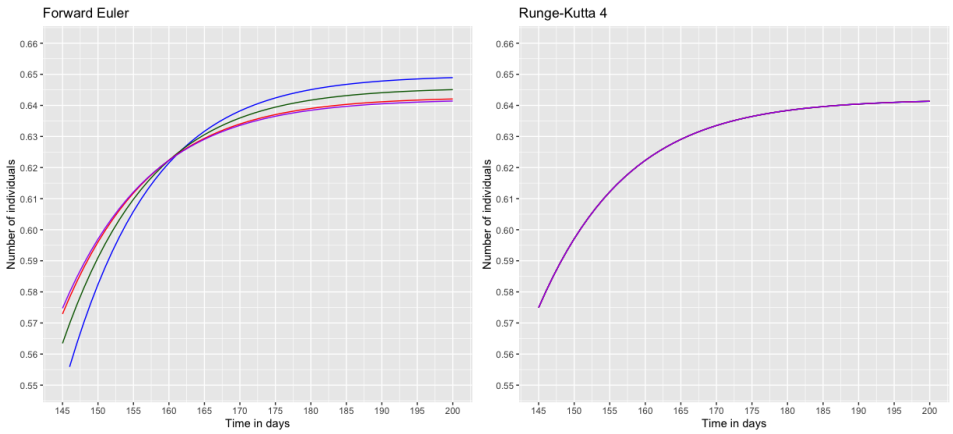


Figure 3.4: Numerical simulation of an (unrealistic) Ebola outbreak with $\beta = 0.32$ and $\gamma = 0.2$. The figure shows S (blue), I (red), and R (green).

Figure 3.5 shows the proportion of recovered individuals ($R(t)$) for different step sizes (Δt), where $\Delta t = 1$ corresponds to the discrete SIR model. It is clear that the results obtained by the Forward Euler method vary more depending on step size than those obtained by RK4. The same data is presented in table 3.1. The error values in this table show how much the results deviate from the result for $\Delta t = 0.001$. For RK4, the results were not affected by the change in step size. In the next section we will discuss this difference in accuracy from a theoretical point of view.



(a) (b)

Figure 3.5: The proportion of recovered individuals after 400 days for different step sizes (Δt) and different numerical methods (a=Euler, b=RK4), with $\beta = 0.32$ and $\gamma = 0.2$.

Δt	RK4	RK4 error (%)	Euler	Euler error (%)
0.001	0.6419834	0.0	0.6419909	0.0
0.01	0.6419834	0.0	0.6420578	0.01
0.1	0.6419834	0.0	0.6427298	0.12
0.5	0.6419834	0.0	0.6457640	0.59
1	0.6419834	0.0	0.6496703	1.20

Table 3.1: The proportion of recovered individuals after 400 days for different step sizes (Δt), with $\beta = 0.32$ and $\gamma = 0.2$.

3.5 Theoretical error analysis

One of the main types of errors in numerical analysis are so called round-off errors, which occur when a finite number of digits is used to represent real numbers. Another common type of errors is called truncation errors, which are caused by simple methods to approximate more complex functions. Error propagation is the term used to describe the combined error of a function, as a result of each variable's individual error.

The error propagation in numerical methods can be estimated using the Taylor series. The exact expression for the point y_{n+1} can be derived by doing a Taylor series expansion around the point:

$$y_{n+1} = y_n + y'_n h + y''_n \frac{h^2}{2} + y'''_n \frac{h^3}{3!} + \ldots + y_n^{(i)} \frac{h^i}{i!} + \ldots$$

Now we want to compare this exact expression to the expressions given by Euler's- and Runge-Kutta's methods. This will give us their estimated errors.

Using Forward Euler we get

$$\tilde{y}_{n+1} = y_n + y'_n h,$$

so the difference between the exact solution and Euler's solution is

$$y_{n+1} - \tilde{y}_{n+1} = y''_n \frac{h^2}{2} + y'''_n \frac{h^3}{3!} + \ldots + y_n^{(i)} \frac{h^i}{i!} + \ldots$$

Since for really small h this error is mostly going to be determined by the value of $y''_n \frac{h^2}{2}$, we can say that the local error e_l for the Forward Euler method has order $\mathcal{O}(h^2)$. We then get the global error e_g by multiplying the local error with the total number of steps. Because we know that the number of steps is inversely proportional to the step size h, we get a global error of $e_g = e_l \cdot \frac{1}{h}$. Hence, the global error for the Forward Euler method has order $\mathcal{O}(h^2) \cdot \frac{1}{h} = \mathcal{O}(h)$. In other words, the global error is proportional to the step size.

This conclusion agrees with table 3.1 from the previous section.

For Runge-Kutta's method we say $\frac{dy}{dx} = f(x, y)$. In a neighbourhood of $(x, y) = (a, b)$, the Taylor series expansion becomes

$$f(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b) +$$

$$+\frac{1}{2}(f_{xx}(a,b)(x-a)^2 + f_{xy}(a,b)(x-a)(y-b) + f_{yy}(a,b)(y-b)^2) + \dots$$

Using RK4 we get the following expression for $\tilde{y}_{n+1}$:

$$\tilde{y}_{n+1} = y_n + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

where

$$k_1 = hf(x,y)$$
$$k_2 = hf(x + \frac{h}{2}, y + \frac{k_1}{2})$$
$$k_3 = hf(x + \frac{h}{2}, y + \frac{k_2}{2})$$
$$k_4 = hf(x + h, y + k_3)$$

Using the Taylor series to rewrite the expression for $\tilde{y}_{n+1}$ (not shown here) reveals that the local error e_l for RK4 has order $\mathcal{O}(h^5)$. Hence, the global error e_g has order $\mathcal{O}(h^4)$.

This conclusion also agrees with table 3.1 from the previous section.

Chapter 4

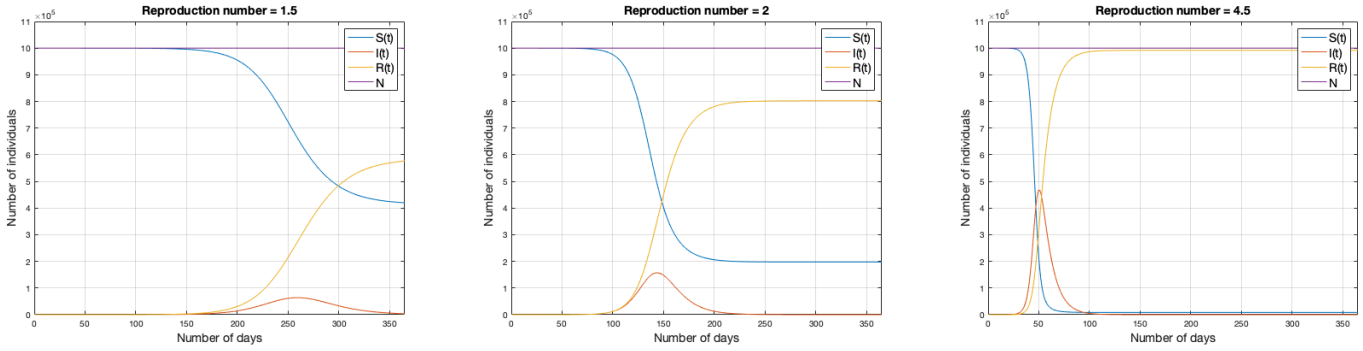
Qualitative analysis of the SIR model

4.1 Basic- and effective reproduction numbers

The **basic reproduction number** (R_0) is the average number of new infections per infectious individual surrounded by a completely susceptible population. In other words, it is the product of the transmission rate (β) and the time period of infection ($\frac{1}{\gamma}$). This means we can write $R_0 = \frac{\beta}{\gamma}$. [8]

Since it is rarely the case that a population entirely consists of susceptible individuals, the actual number of new infections per infected individual will be lower than R_0. To account for this, the **effective reproduction number** (R_e) multiplies R_0 with the fraction of the population consisting of susceptible individuals (S). Hence we have $R_e = \frac{S}{N} \cdot \frac{\beta}{\gamma}$. Note that since the fraction of susceptible individuals (S) changes over time, R_e will also change over time. [11]

Figure 4.1 shows how the variables of the SIR model change over time for different reproduction numbers (R_0). Chosen parameters were $N = 10^6$, $I(0) = 1$, $S(0) = N - 1$, $\gamma = \frac{1}{10}$ and $\beta = R_0 \cdot \gamma = \frac{R_0}{10}$. It can be seen that increasing R_0 resulted in an earlier and increased peak for $I(t)$ as well as the system returning to a state of equilibrium sooner. It can also be seen that the size of R_0 affected the final size of $S(t)$. That is, the larger R_0, the larger the epidemic.



(a) $R_0 = 1.5$ (b) $R_0 = 2$ (c) $R_0 = 4.5$

Figure 4.1: SIR for different sizes of R_0

4.2 Epidemic threshold

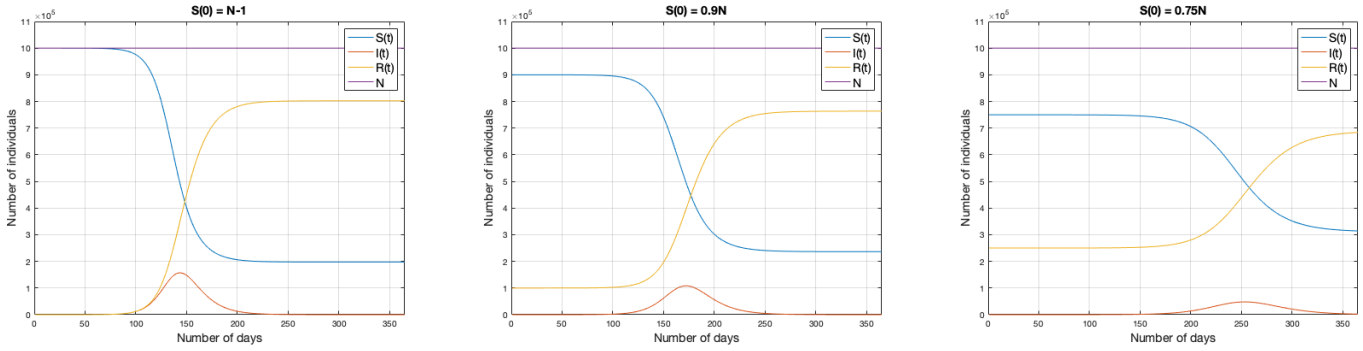
In this section we will investigate the factors that determine whether a disease fades out or develops into an epidemic. We will start by examining how R_e affects the change in the number of infected individuals (I). Combining equation (2.2) from section 2.3.2 with the previously discussed fact that $S(0) \geq S(t)$, we derive

$$\frac{\Delta I}{\Delta t} = \beta \cdot I(t) \cdot \frac{S(t)}{N} - \gamma \cdot I(t) \leq \beta \cdot I(t) \cdot \frac{S(0)}{N} - \gamma \cdot I(t) = \gamma \cdot (R_e - 1) \cdot I$$

This gives us that $\frac{\Delta I}{\Delta t} < 0$ for $R_e < 1$, as well as $\frac{\Delta I}{\Delta t} > 0$ for $R_e > 1$. In other words, the number of infected individuals (I) will decrease only when $R_e < 1$ as well as increase only when $R_e > 1$.

Figure 4.2 shows how the variables of the SIR model change over time for different values of $S(0)$. Chosen parameters were $N = 10^6$, $I(0) = 1$, $\gamma = \frac{1}{10}$ and $\beta = \frac{1}{5}$. It can be seen that decreasing $S(0)$ resulted in a delayed and decreased peak for $I(t)$ as well as the system returning to a state of equilibrium later. It can also be seen that the size of $S(0)$ affected the final size of $S(t)$. That is, the smaller $S(0)$, the smaller the epidemic.

Figure 4.3 displays the **epidemic threshold**. This is the critical number of susceptible individuals (S) required for an epidemic to occur. Using our previously derived understanding of R_e, we see that an epidemic will occur only when $R_e = \frac{S}{N} \cdot \frac{\beta}{\gamma} > 1$. Rewriting this equation gives us $S > N \cdot \frac{\gamma}{\beta} = \frac{N}{R_0}$.



(a) $S(0) = N - 1$ (b) $S(0) = 0.9N$ (c) $S(0) = 0.75N$

Figure 4.2: SIR for different sizes of $S(0)$

Hence, the epidemic threshold can be expressed as $S = \frac{N}{R_0}$. For an entirely susceptible population ($S = N$), an epidemic will occur only when $1 > \frac{1}{R_0}$. In other words, a disease will cause an epidemic only when it's transmission rate is greater than it's recovery rate ($\beta > \gamma$).

4.3 Equilibrium

A population is at equilibrium when the number of infected individuals is constant. That is, when we have $\frac{dI}{dt} = 0$. Using the formula for $\frac{dI}{dt}$ derived in section 2.3.2, we can find out for which situations a population is at equilibrium by solving the following equation:

$$\beta \cdot I \cdot \frac{S}{N} - \gamma \cdot I = 0$$

$$I = 0 \quad \vee \quad S = \frac{\gamma N}{\beta}$$

The population is at equilibrium when there are no infected individuals ($I = 0$). This is called the **Disease Free Equilibrium** (DFE), denoted by $DFE = (N, 0, 0)$. For $I > 0$, the population is at equilibrium when $S = \frac{\gamma N}{\beta}$. This is called the **Endemic Equilibrium** (EE), denoted by $EE = (\frac{\gamma N}{\beta}, I*, R*)$.

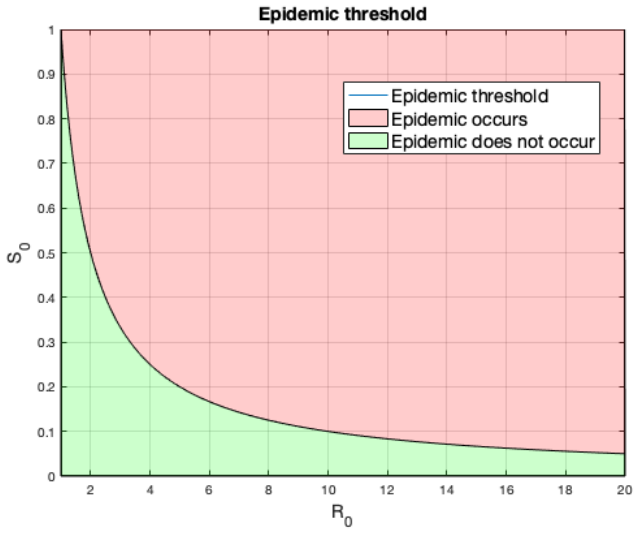


Figure 4.3: The epidemic threshold for different sizes of S_0 and R_0.

Theorem 4.3.1. *For all values of R_0, there is a unique disease free equilibrium (DFE).*

Proof. All populations are at equilibrium when there are no infected individuals, no matter the value of R_0. Therefore, both when $R_0 \leq 1$ and when $R_0 > 1$ there is a unique DFE with $I = 0$.

□

4.4 The size of an epidemic

In this section we are interested in the size of an epidemic. That is, we want to know the value of $S(t)$ as t goes to infinity. We can find this value by dividing the infected equation (2.3.2) by the susceptible equation (2.3.1), and then integrating the resulting equation:

$$\frac{\frac{dI}{dt}}{\frac{dS}{dt}} = \frac{dI}{dS} = \frac{\beta IS - \gamma I}{-\beta IS} = \frac{\gamma}{\beta S} - 1$$

$$I(t) = \frac{\gamma}{\beta} \cdot \ln S(t) - S(t) + C$$

Through substituting $S(0) = S_0$ and $I(0) = I_0$, we can find an expression for C:

$$I_0 = \frac{\gamma}{\beta} \cdot \ln S_0 - S_0 + C$$

$$C = S_0 + I_0 - \frac{\gamma}{\beta} \cdot \ln S_0$$

$$I(t) = -S(t) + S_0 + I_0 + \frac{\gamma}{\beta} \cdot \ln \frac{S(t)}{S_0}$$

In section 4.3 we showed that I reaches it's peak for $S = \frac{1}{R_0} = \frac{\gamma}{\beta}$. Hence, by substituting this into our equation we can find an expression for I_{max}:

$$I_{max} = \frac{\gamma}{\beta} \cdot \ln \frac{\gamma}{\beta} - \frac{\gamma}{\beta} + S_0 + I_0 - \frac{\gamma}{\beta} \cdot \ln S_0$$

$$I_{max} = S_0 + I_0 + \frac{\gamma}{\beta} \cdot \left(\ln \frac{\gamma}{\beta} - \ln S_0 - 1 \right)$$

Figure 4.4 shows that as R_0 increases, I_{max} also increases. It also shows that as S_0 decreases, I_{max} decreases as well.

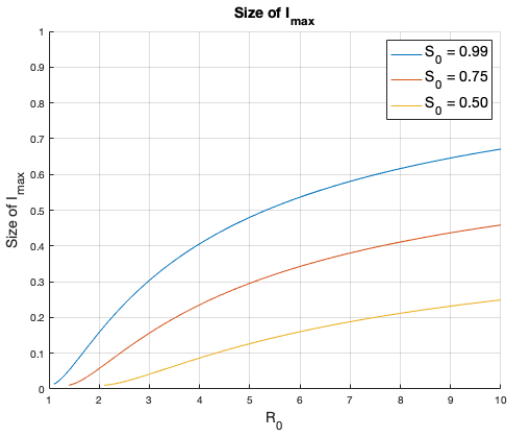


Figure 4.4: I_{max} for different sizes of S_0 and R_0.

The expression for $S(\infty)$ can be found by substituting $I(\infty) = 0$ into our

equation for $I(t)$:

$$0 = -S(\infty) + S_0 + 0 + \frac{\gamma}{\beta} \cdot \ln \frac{S(\infty)}{S_0}$$

$$S(\infty) = S_0 + \frac{\gamma}{\beta} \cdot \ln \frac{S(\infty)}{S_0}$$

We call the proportion of all individuals getting infected during an epidemic for I_{total}:

$$I_{total} = S_0 - S(\infty) = -\frac{\gamma}{\beta} \cdot \ln \frac{S(\infty)}{S_0}$$

From these expressions for $S(\infty)$ and I_{total}, it can be seen that both depend on β, γ and S_0. In other words, the size of an epidemic depends on β, γ and S_0. Figure 4.5 shows that as R_0 increases, I_{total} increases and $S(\infty)$ decreases. As S_0 decreases, I_{total} decreases and $S(\infty)$ increases.

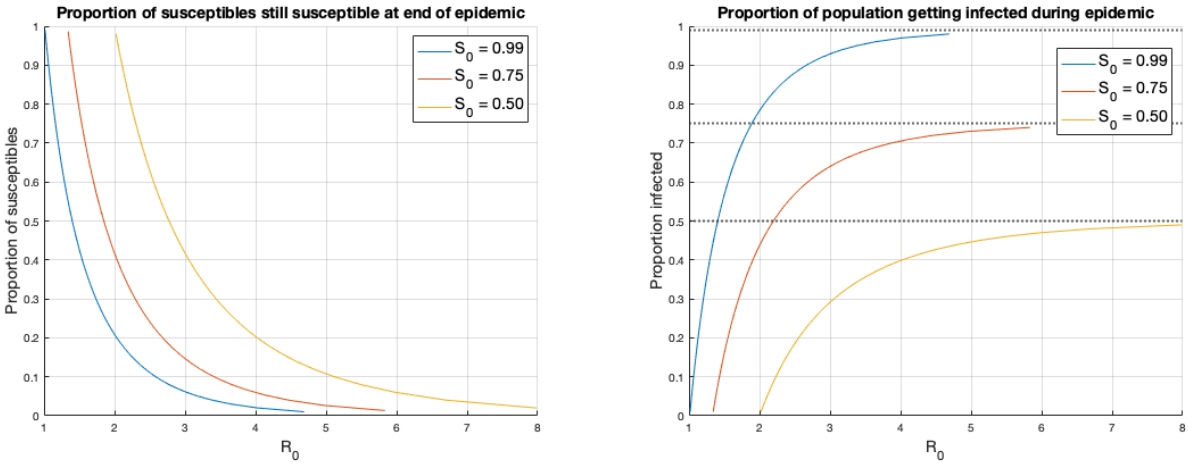


Figure 4.5: $S(\infty)$ (left) and I_{total} (right) for different sizes of S_0 and R_0.

Chapter 5

Vaccination

In previous chapters we investigated ways of determining the epidemic threshold as well as the size of an epidemic. We saw that an epidemic will occur only when $\frac{S(0)}{N} > \frac{1}{R_0}$ and that the size of an epidemic depends on β, γ and S_0. In this chapter we will see how this knowledge is used in the development of public health interventions aiming to prevent and control infectious disease.

5.1 Immunization

The immune system consists of many different components and mechanisms [16]. There are the skin and mucous membranes, body temperature, gastric acidity, proteins and antibodies able to kill pathogenic microorganisms, and inflammation. All these general mechanisms are part of the innate immune response, which is the response that is not specific to particular pathogenic agents and has no memory.

In contrast to this part of our immune system, there is the adaptive immune response which does have memory and is able to recognize specific pathogenic agents [16]. This memory enables a more rapid and effective response the next time the body exposed to the same pathogen. This process is called immunization, and can occur either naturally or artificially. Exposure to influenza is an example of natural immunization, whereas artificial immunization is achieved through vaccination.

Vaccination reduces or removes the risk for an individual to get infected.

An analysis on a hypothetical birth cohort of 4,2 million infants showed that the routine US childhood immunization schedule will prevent around 42 thousand early deaths and 20 million cases of disease, as well as save $13.5 billion in direct costs and $68.8 billion in total societal costs [17, 18].

5.2 Vaccination and the SIR model

Since vaccination reduces or removes the risk for an individual to get infected, we can say it removes the individual from the susceptible class. In other words, it causes $S(0)$ to decrease. In section 4.4 we saw that both $S(\infty)$ and I_{total} depend on $S(0)$, and in section 4.2 on the epidemic threshold we saw that an epidemic will occur only when $\frac{S(0)}{N} > \frac{1}{R_0}$ (or when $R_e > 1$). Therefore, using vaccination we can decrease the sizes of $S(\infty)$ and I_{total}, as well as raise the epidemic threshold.

Say we call the fraction of vaccinated individuals p, and we introduce the variables b, d and ω representing birth rate, natural death rate, and infection death rate respectively. Assuming all newborns are born into the susceptible class, we get the following system of ODE's:

$$\begin{cases} \frac{dS}{dt} = b - (1-p)\beta SI - dS \\ \frac{dI}{dt} = ((1-p)\beta S - (\gamma + d + \omega)) \cdot I \\ \frac{dR}{dt} = \gamma I - dR \end{cases}$$

β	transmission rate
γ	recovery rate
d	natural death rate
ω	infection death rate
b	birth rate
p	fraction of vaccinated individuals

Table 5.1: The variables of the SIR model.

In section 4.1 we defined R_0 as the average number of new infections per infectious individual surrounded by a completely susceptible population. Including our newly defined variables d and ω, we get $R_0 = \frac{\beta}{\gamma + d + \omega}$. In the same

section, we defined R_e as the actual number of new infections per infected individual. Therefore, we get $R_e = (1-p) \cdot \frac{S}{N} \cdot R_0$.

5.3 Herd immunity and Smallpox

As defined in section 4.2, the epidemic threshold is the critical number of susceptible individuals (S) required for an epidemic to occur. We saw that $\frac{dI}{dt} < 0$ for $R_e < 1$, as well as $\frac{dI}{dt} > 0$ for $R_e > 1$. In other words, the number of infected individuals (I) will increase only when $(1-p) \cdot \frac{S}{N} \cdot R_0 > 1$.

Assuming that before vaccination all individuals are susceptible, we have $S = N$ and hence $(1-p) \cdot R_0 > 1$. That is, I will increase only when $p < 1 - \frac{1}{R_0}$. So, in order to prevent an infection from spreading, the fraction of individuals needed to be vaccinated is $p > 1 - \frac{1}{R_0}$.

This leads to the crucial insight that not all individuals of a population need to be vaccinated, in order to still be able to protect all individuals from getting infected. This phenomenon is called **herd immunity** and p is called the herd immunity threshold. Table 5.2 shows the herd immunity threshold for several different infectious diseases.

Disease	R_0	p
Smallpox	5	.80
Polio	5	.80
Rubella	7	.86
Chicken pox	11	.91
Mumps	12	.92
Measles	16	.94

Table 5.2: Several diseases and their values for R_0 and p. [19]

Smallpox, caused by the variola virus, was one of the worlds deadliest diseases. It killed approximately 30% of those infected, and left others blind, sterile, and with deep scars. Following the European colonization of America, 90 percent of all indigenous casualties are thought to have been caused by smallpox as well as the deaths of around 3 million Aztecs. During the 18th century in Europe, an estimated 60 million people died as the result of smallpox. In the 20th century, it is estimated to have killed 300 million

people globally. Figure 5.1 shows the number of smallpox cases globally per million individuals for the period 1920 - 2016.

For smallpox, R_0 is approximated to be around 5 and therefore p has to be around .80. This means that at least 80 percent of the susceptible population needs to be vaccinated in order to achieve herd immunity. In 1967, using the understanding of the herd immunity threshold, the World Health Organization launched a program aiming to do exactly this [19]. Over the following decade, they vaccinated millions of people and because of this, in 1980 smallpox was officially declared eradicated. As of today, smallpox is the only disease that has been eradicated.

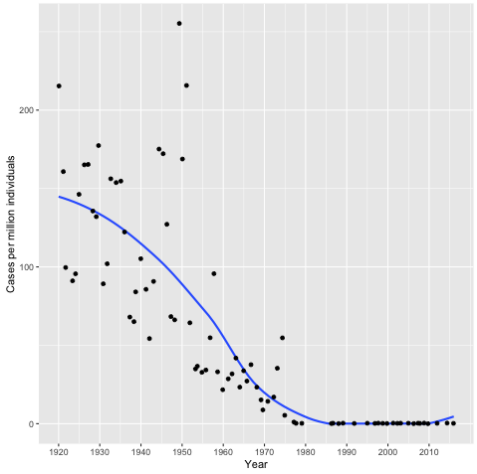


Figure 5.1: The number of smallpox cases per million individuals. [20]

5.4 Stability of the equilibrium points

As showed in section 4.3, we calculate the system's equilibrium points by solving $\frac{dS}{dt} = \frac{dI}{dt} = \frac{dR}{dt} = 0$. This gives us $I = 0$ and $S = \frac{b}{d}$ for the disease free equilibrium (DFE), as well as $S = \frac{\gamma+d+\omega}{\beta(1-p)}$ for the endemic equilibrium (EE). Substituting R_0 into this last equation gives $EE = (S^*, I^*, R^*)$, with

$$S^* = \frac{1}{(1-p)R_0} \quad \text{and} \quad I^* = \frac{d}{\beta}(R_0 - \frac{1}{1-p})$$

An equilibrium point is stable if the system returns to it after a small disturbance, whereas an equilibrium point is unstable if the system moves away from it after a small disturbance. Theorem 5.4.1 states the mathematical definition of stability. [21]

Theorem 5.4.1. *An equilibrium point $\hat{x}$ is*

1. *stable, if for any $\epsilon > 0$, there exists $\delta > 0$ such that $|x(0) - \hat{x}| < \delta \rightarrow |x(t) - \hat{x}| < \epsilon$ for $t \geq 0$*
2. *asymptotically stable, if it is stable and if there exists $\delta > 0$ such that $|x(0) - \hat{x}| < \delta \rightarrow \lim_{t\to\infty} x(t) = \hat{x}$*
3. *unstable, if it is not stable.*

In order to analyze the stability of the equilibrium points, we study the eigenvalues of the Jacobian matrix of the system of equations from section 5.2. We write $\frac{dS}{dt} = f(S, I)$ and $\frac{dI}{dt} = g(S, I)$, and get:

$$J(S,I) = \begin{pmatrix} \frac{\delta f}{\delta S} & \frac{\delta f}{\delta I} \\ \frac{\delta g}{\delta S} & \frac{\delta g}{\delta I} \end{pmatrix} = \begin{pmatrix} -(1-p)\beta I - d & -(1-p)\beta S \\ (1-p)\beta I & (1-p)\beta S - (\gamma + d + \omega) \end{pmatrix}$$

Theorem 5.4.2. *An equilibrium point $\hat{x}$ of $\dot{x} = f(x)$, with λ being the eigenvalues of $J = Df(\hat{x})$, is*

1. *asymptotically stable, if Re $\lambda < 0$ for all λ*
2. *unstable, if Re $\lambda > 0$ for at least one λ*

Theorem 5.4.3. *If $R_0 < \frac{1}{1-p}$, the DFE is asymptotically stable. Otherwise it is unstable.*

Proof. When it comes to the DFE, we have $S = \frac{b}{d}$ and $I = 0$. For the sake of simplicity we say $b \approx d$. This transforms our Jacobian matrix into

$$J(S,I) = \begin{pmatrix} -d & -(1-p)\beta \\ 0 & (1-p)\beta - (\gamma + d + \omega) \end{pmatrix} = \begin{pmatrix} -d & -(1-p)\beta \\ 0 & (\gamma + d + \omega)((1-p)R_0 - 1) \end{pmatrix}$$

Obviously, the eigenvalues of this matrix are $-d$ and $(\gamma+d+\omega)((1-p)R_0-1)$. As stated by theorem 5.4.2, if the real parts of all eigenvalues are negative, then the equilibrium is asymptotically stable. If one or more of the eigenvalues have positive real parts, then the equilibrium is unstable. For $R_0 < \frac{1}{1-p}$ both eigenvalues are negative, which means that the DFE is asymptotically stable. However, for $R_0 > \frac{1}{1-p}$ the second eigenvalue is positive which means that in that case the DFE is unstable. □

Theorem 5.4.4. *If J is a 2×2 matrix, then the sum of the eigenvalues of J is equal to the trace of J and the product of the eigenvalues is equal to the determinant of J.*

Proof. See appendix. □

Theorem 5.4.5. *If $R_0 > \frac{1}{1-p}$, the endemic equilibrium (EE) is asymptotically stable.*

Proof. When it comes to the EE, we have $S^* = \frac{1}{(1-p)R_0}$ and $I^* = \frac{d}{\beta}(R_0 - \frac{1}{1-p})$. Again, we say $b \approx d$. This transforms our Jacobian matrix into

$$J(S^*, I^*) = \begin{pmatrix} -(1-p)\beta I^* - d & -(1-p)\beta S^* \\ (1-p)\beta I^* & (1-p)\beta S^* - (\gamma + d + \omega) \end{pmatrix} =$$
$$= \begin{pmatrix} -(1-p)dR_0 & -(\gamma + d + \omega) \\ d((1-p)R_0 - 1) & 0 \end{pmatrix}$$

We have $\text{trace}(J) = -(1-p)dR_0$ and $\det(J) = d((1-p)R_0 - 1)(\gamma + d + \omega)$. From theorem 5.4.2 we know that an equilibrium is asymptotically stable if the real parts of all eigenvalues are negative. From theorem 5.4.4 we know that the sum of the eigenvalues of J is equal to the trace of J, and that the product of the eigenvalues of J is equal to the determinant of J.

Since our Jacobian has two eigenvalues, a positive determinant either means that both eigenvalues are negative or that both eigenvalues are positive. A positive determinant combined with a negative trace therefore means that

both eigenvalues are negative, and that the equilibrium is asymptotically stable.

Hence our EE is asymptotically stable for trace$(J) < 0$ and det$(J) > 0$, which is true for $R_0 > \frac{1}{1-p}$. □

In reality, this means that a disease free population will return to being disease free as long as $R_0 < \frac{1}{1-p}$. This confirms our findings from section 5.3, where we saw that the occurrence of an epidemic can be prevented by making sure $p > 1 - \frac{1}{R_0}$.

5.5 Vaccination and the size of an epidemic

Just as in section 4.4, we will derive an expression for the size of an epidemic. Only this time we will be focusing the relationship between the vaccination coverage (p) and the epidemic size (I_{total}). Again, we start by dividing the infected equation by the susceptible equation and then integrating the resulting equation:

$$\frac{\frac{dI}{dt}}{\frac{dS}{dt}} = \frac{dI}{dS} = \frac{(1-p)\beta IS - \gamma I}{-(1-p)\beta IS} = \frac{\gamma}{(1-p)\beta S} - 1$$

$$I(t) = \frac{\gamma}{(1-p)\beta} \cdot \ln S(t) - S(t) + C$$

Through substituting $S(0) = (1-p)S_0$ and $I(0) = I_0$, we find $C = (1-p)S_0 + I_0 - \frac{\gamma}{(1-p)\beta} \cdot \ln (1-p)S_0$. Hence we get:

$$I(t) = -S(t) + (1-p)S_0 + I_0 + \frac{\gamma}{(1-p)\beta} \cdot \ln \frac{S(t)}{(1-p)S_0}$$

We derive the expression for $S(\infty)$ by substituting $I(\infty) = 0$, $S_0 + I_0 = 1 - p$ and $R_0 = \frac{\beta}{\gamma}$ into our equation for $I(t)$:

$$S(\infty) = 1 - pS_0 + \frac{1}{(1-p)R_0} \cdot \ln \frac{S(\infty)}{(1-p)S_0} \tag{5.1}$$

The last step is to find the total number of individuals infected during the epidemic:

$$I_{total} = 1 - pS_0 - S(\infty) = -\frac{1}{(1-p)R_0} \cdot \ln \frac{S(\infty)}{S_0} \tag{5.2}$$

At the beginning of this section we mentioned that we would be investigating the relationship between vaccination coverage and epidemic size. We will do this using the example of measles. Measles is an extremely contagious virus with $R_0 = 16$ approximately. Despite the existence of an effective vaccine, during 2018 alone it caused the death of more than 140 000 individuals globally. During the period 2000-2018, vaccination against measles prevented an estimated 23.2 million deaths. (WHO, 2020)

Figure 5.2 shows a numerical estimate of the size of a measles epidemic (left) for different vaccination rates, as well as actual WHO data on reported cases of measles for different vaccination rates (right). For the numerical estimate, we used Matlab to numerically solve and plot equation 5.1. The data from WHO contained an annual report on global vaccination coverage as well as the total number of reported cases globally (1980-2018). We used R to convert the total number of cases to the number of cases per million individuals, and and then plotted this number against the reported vaccination coverage. Even though both plots have very different y-values, their curves are strikingly similar.

Figure 5.3 also shows a numerical estimate of the size of a measles epidemic (left) for different vaccination rates, as well as actual WHO data on reported cases of measles for different vaccination rates (right). The aim of this plot is to have a closer look at higher vaccination rates. The numerical estimate is the same one as displayed in figure 5.2, except now only showing vaccination rates of 75% and above. The plot on the right displays annual data from all countries in the WHO regions of America and Europe (1980-2018). Numbers of cases per million individuals and year are plotted against vaccination rates. Again even though both plots have very different y-values, their curves are strikingly similar.

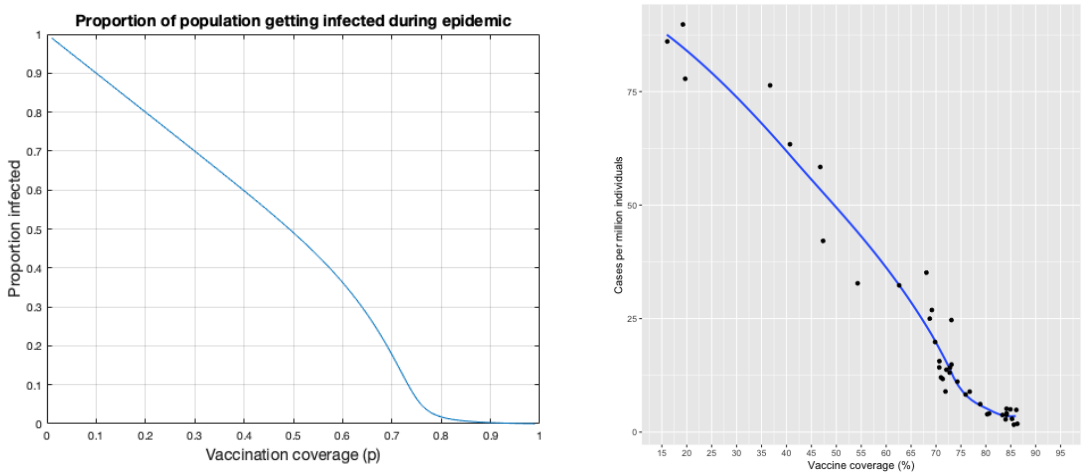


Figure 5.2: The size of a measles epidemic for different vaccination rates. Numerical estimate (left) compared to actual WHO data (right).

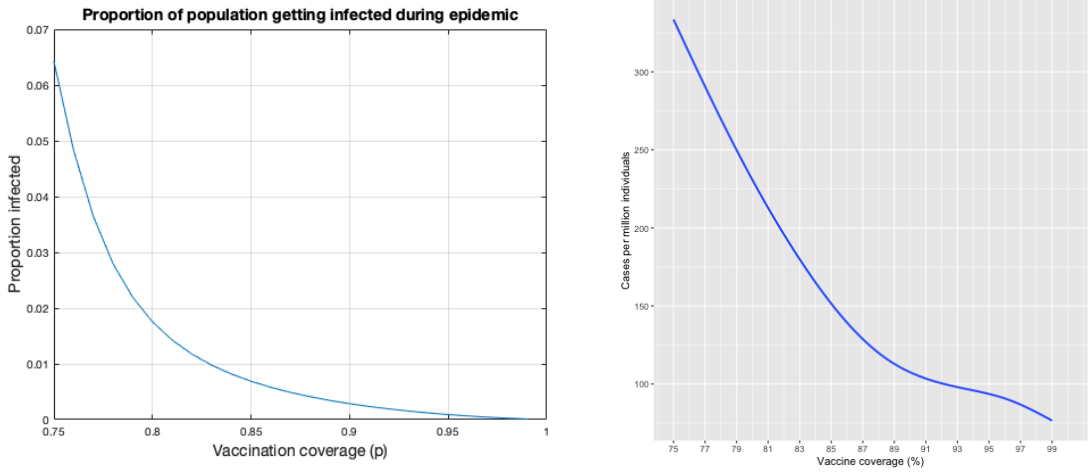


Figure 5.3: The size of a measles epidemic for different vaccination rates. Numerical estimate (left) compared to actual WHO data (right).

Chapter 6

Model fitting: COVID-19

In this chapter we will attempt to fit the SIR-model to data on the current COVID-19 pandemic. COVID-19 is a disease caused by a novel Coronavirus, discovered in December of 2019. The Coronaviruses are a group of viruses, of which seven are known to cause disease in humans. Four of these result in the relatively harmless symptoms of the common cold, whereas the other three cause potentially severe respiratory diseases known as the *Middle East Respiratory Syndrom* (MERS), the *Severe Acute Respiratory Syndrom*the *(SARS), and Corona Virus Disease 19* (COVID-19). For MERS and SARS R_0 has been estimated to be 0.3-0.8 and 2-5 respectively. Since COVID-19 is a new virus, R_0 is not yet known. In the first section of this chapter we will use COVID-19 data to estimate a range for R_0. In the other sections we simulate the COVID-19 outbreak in Sweden for various scenarious.

6.1 Estimating R_0 for COVID-19

Table 6.1 summarizes the parameters we need to estimate in order to model the COVID-19 outbreak. The time period for infection has been approximated to be around three to eight days [22]. Therefore, we investigate the scenarios where $\gamma = \frac{1}{3}$, $\gamma = \frac{1}{4}$, $\gamma = \frac{1}{5}$, $\gamma = \frac{1}{6}$, $\gamma = \frac{1}{7}$, and $\gamma = \frac{1}{8}$. The basic reproduction number (R_0) is the transmission rate (β) divided by the recovery rate (γ), so we have $R_0 = \frac{\beta}{\gamma}$.

Theoretically, R_0 and β are estimated using $S(t)$ and $I(t)$ for different values of t. In the case of COVID-19 this is tricky since these exact numbers

R_0	basic reproduction number
β	transmission rate
κ	contacts
τ	transmissibility
γ	probability of recovery or death
$\frac{1}{\gamma}$	time period for infection

Table 6.1: Parameters to be estimated.

are not yet known. Only certain patients have been tested for the virus, and we can assume that the confirmed cases are only the tip of the iceberg. However, we might also be able to assume that the rate at which this *tip of the iceberg* (the number of confirmed cases) changes is roughly equal to the rate at which the *entire iceberg* (the total number of infected individuals) changes. So, despite the lack of data, we will try to roughly estimate values for R_0 and β.

Early in an outbreak, we still have $S \approx N$, which enables us to simplify the expression for the infected equation as follows:

$$\frac{dI}{dt} = \beta \cdot I(t) \cdot \frac{S(t)}{N} - \gamma \cdot I(t) \approx \beta \cdot I(t) - \gamma \cdot I(t) = I(t) \cdot (\beta - \gamma)$$

$$\beta \approx \frac{dI}{dt} \cdot \frac{1}{I(t)} + \gamma$$

Here $\frac{dI}{dt}$ is the new number of confirmed cases each day and $I(t)$ can be approximated by $I(t) \approx \frac{dI}{dt} + (1-\gamma)I(t-1) \approx \frac{dI}{dt} + \frac{4}{5}I(t-1)$.

For our analysis we selected fourteen countries with a substantial number of reported cases. Data on new cases per day as well as the total number of cases was retrieved from the *European Center for Disease Prevention and Control* ECDC, spanning the period between December 31st (2019) and March 21st (2020). A simple script in *R* was used to estimate values for β and R_0 for each country and each day. Hence, a total of 1148 different values for β and R_0 were estimated. However, since the epidemic started at different dates for each country, many of the values were completely off. This was solved by removing all values where either $R_0 < 1$ or $R_0 > 5$. This resulted in 348 different values. Since we assumed $S \approx N$, we decided to limit our

analysis to the first 20 days of each country's outbreak. Note that for some countries the total number of days since the outbreak was less than 20. The final number of values for β and R_0 was 208. Figures 6.1 and 6.2 show the average values of R_0 for each value of γ and for the number of days since the start of the outbreak. Table 6.3 shows the average R_0 for each value of γ. As the time period of infection increases with one day, R_0 appears to increase with 0.1. Overall, we approximate the R_0 of COVID-19 to lie within the range 2.8-3.3.

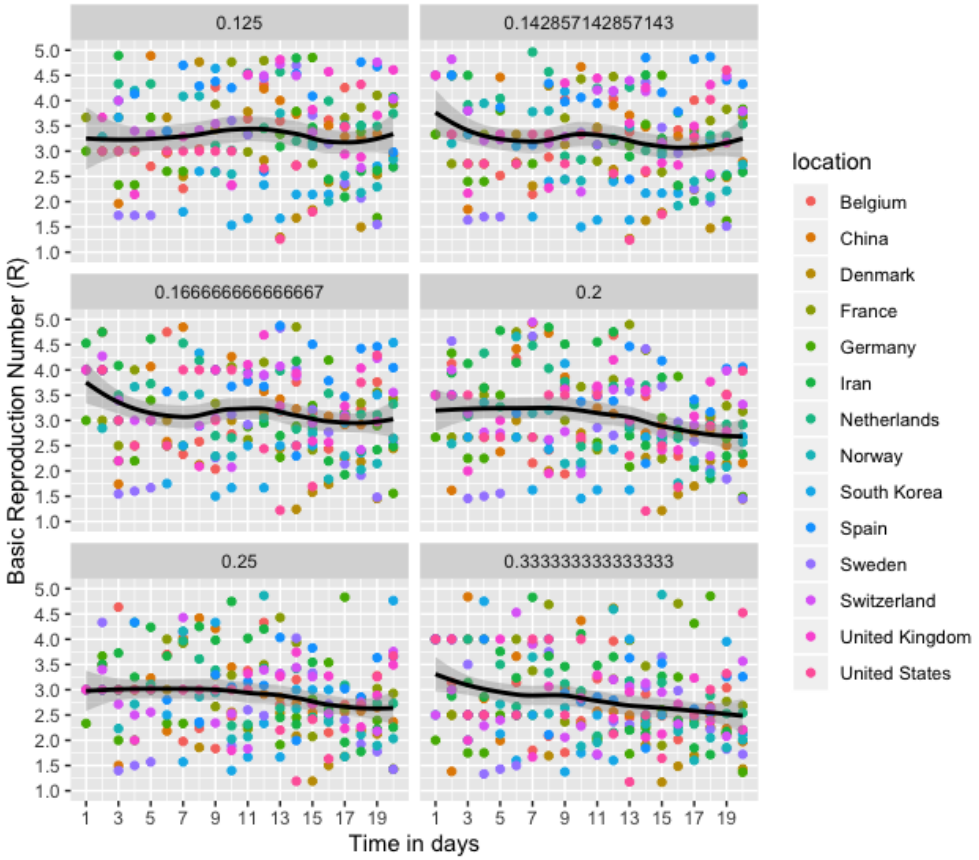


Figure 6.1: The average reproductive number R_0 for different γ.

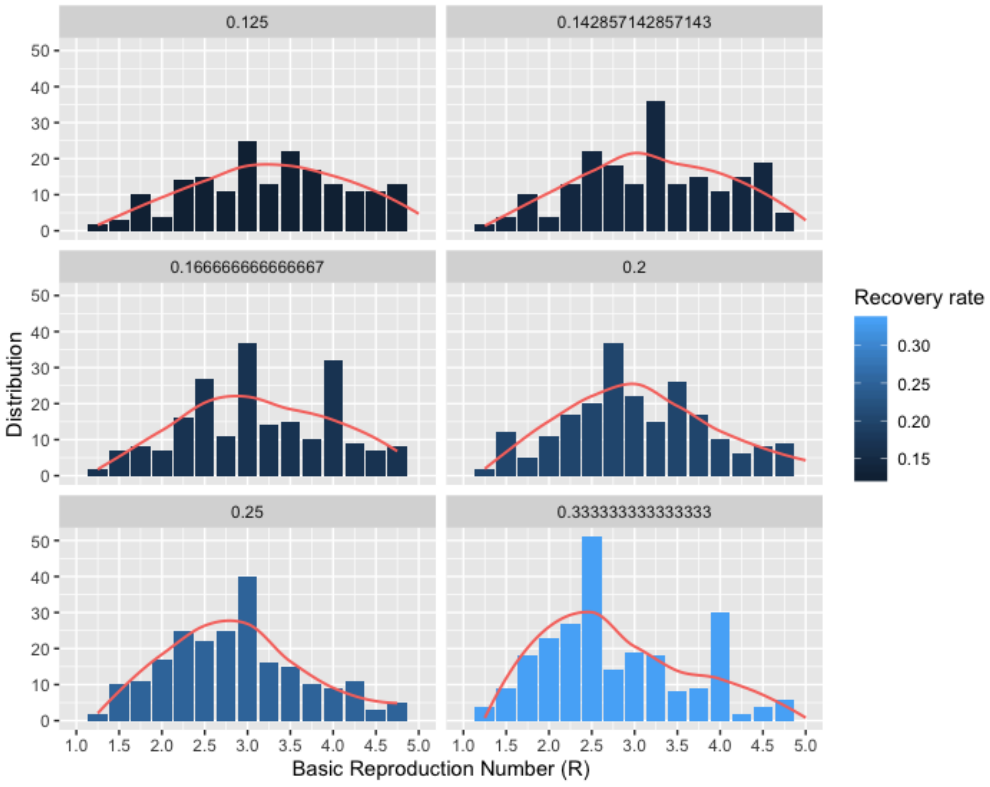


Figure 6.2: The distribution of the estimates of R_0 for different values of γ.

Infection period	Recovery rate	Count	R
3	0.3333333	243	2.8
4	0.2500000	221	2.9
5	0.2000000	220	3.0
6	0.1666667	210	3.1
7	0.1428571	201	3.2
8	0.1250000	186	3.3

Figure 6.3: The average reproductive number R_0 for different γ.

6.2 Simulating a COVID-19 epidemic in Sweden

In this section we will use the previously derived values of R_0 to simulate an outbreak of COVID-19 in Sweden. Sweden has a population of approximately 10 million people and we say the outbreak starts with one infected individual. Therefore, we have $S(0) = 10000000$, $I(0) = 1$, and $R(0) = 0$.

6.2.1 Without social distancing

Figure 6.4 shows the number of infected and recovered (including deceased) individuals respectively, for the scenario where R_0 is constant throughout the entire outbreak. That is, without social distancing or quarantine. The figure shows that, depending on R_0 and γ, the outbreak would reach it's peak after approximately 30-60 days. Around this peak, roughly 2.7-3.3 million individuals would be infected simultaneously. We also see that an estimated 90-95% of the total population would become infected at some point during the outbreak.

6.2.2 With social distancing

Figure 6.5 shows the number of infected and recovered (including deceased) individuals respectively, for the scenario where R_0 varies throughout the outbreak. As explained in chapter 2.1.3, the transmission rate (β) is equal to the product of the number of contacts (κ) a person has per time unit and the transmissibility (τ) of the disease. Hence, R_0 can be reduced both by reducing κ as well as by reducing τ. In practice, this can be done through personal hygiene, social distancing and/or (self-)quarantine. For this scenario, we assumed individuals would start doing this from day 20 and onward, and that on average they would reduce their social exposure/risk of transmission by 50%. Figure 6.5 shows that the curve of infected individuals has been flattened substantially. Depending on R_0 and γ, the outbreak would reach it's peak after approximately 60-160 days. Around this peak, roughly 45 000 - 90 000 individuals would be infected simultaneously. We also see that an estimated 50-70% of the total population would become infected at some point during the outbreak. Thus, both the size of the peak as well as the total size of the epidemic will be much smaller when the individuals of the population practice social distance.

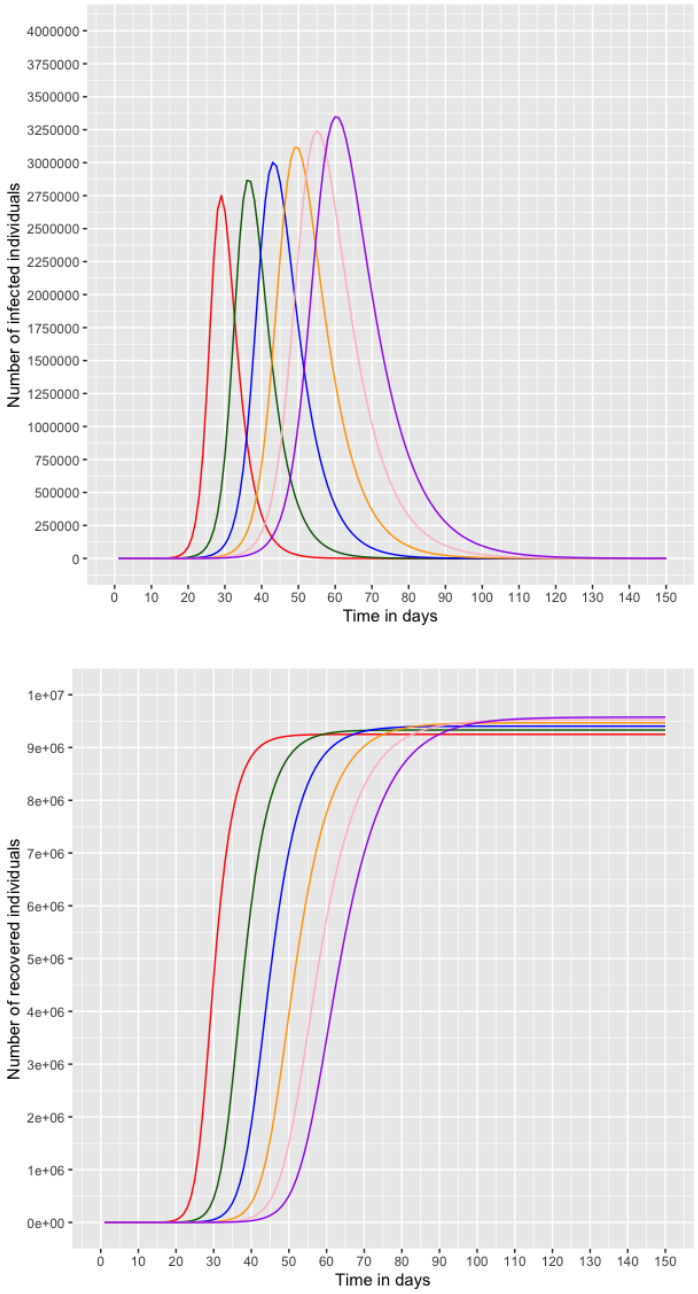


Figure 6.4: The number of infected individuals (top) and recovered individuals (bottom) for different values of R_0 and γ. See figure 6.5 for a legend specifying the different colors.

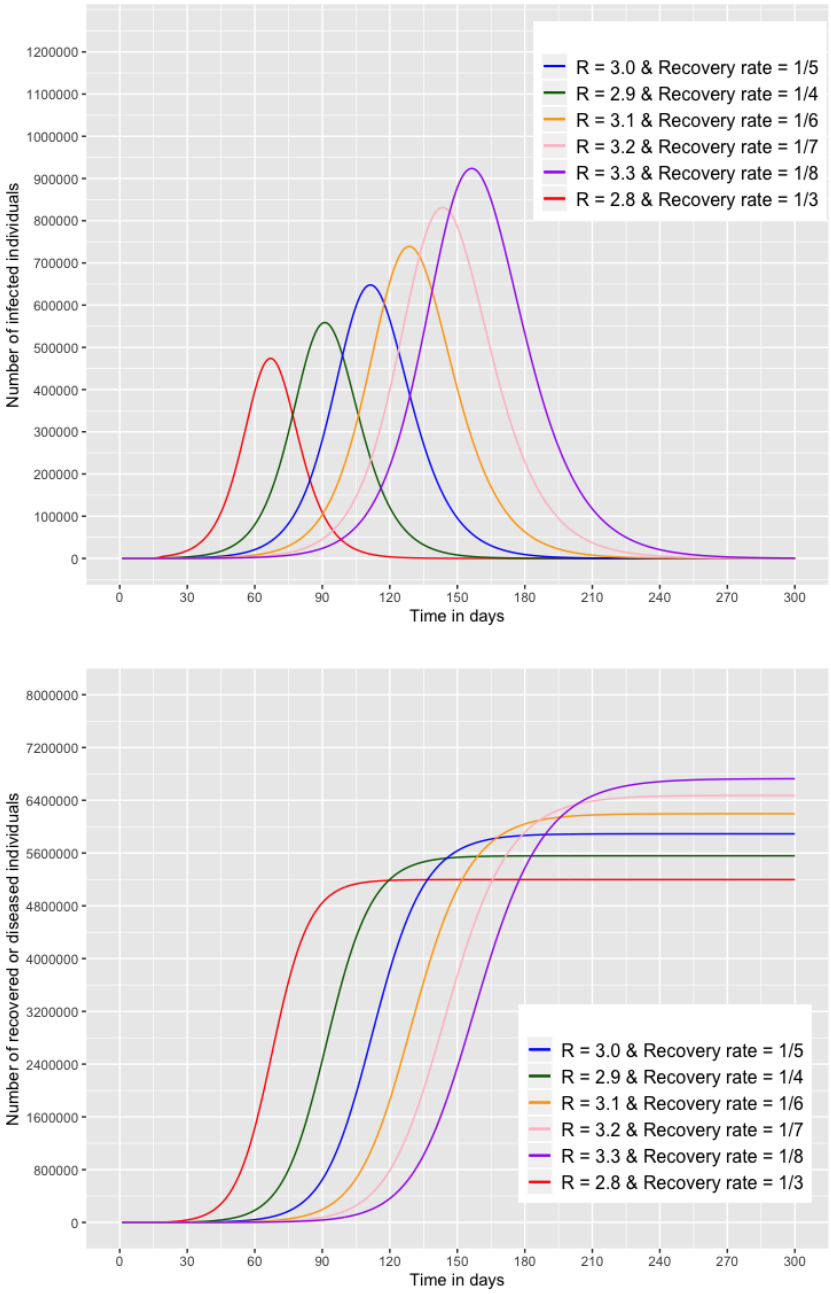


Figure 6.5: The number of infected individuals (top) and recovered individuals (bottom) for different values of R_0 and γ, where after 20 days measures such as personal hygiene and social distance reduce β by 50%.

6.3 The effects of quarantine

In this final section on COVID-19 we investigate the effects of quarantine as well as the effects of the timing of quarantine. For each of these simulations we assume that from day ten of the outbreak the transmission rate will be reduced by 25% due to increased personal hygiene and social distancing. While under quarantine, we chose the transmission rate to be 20% of the original transmission rate ($\beta/5$). The aim of these simulations is to determine the effects of quarantine and the timing of quarantine on the height of the peak of infected individuals as well as on the size of the epidemic.

Figure 6.6 shows that for our specific scenario, a month long quarantine after four or six weeks merely delays the epidemic. It decreases neither the height of the peak of infected individuals nor the size of the epidemic.

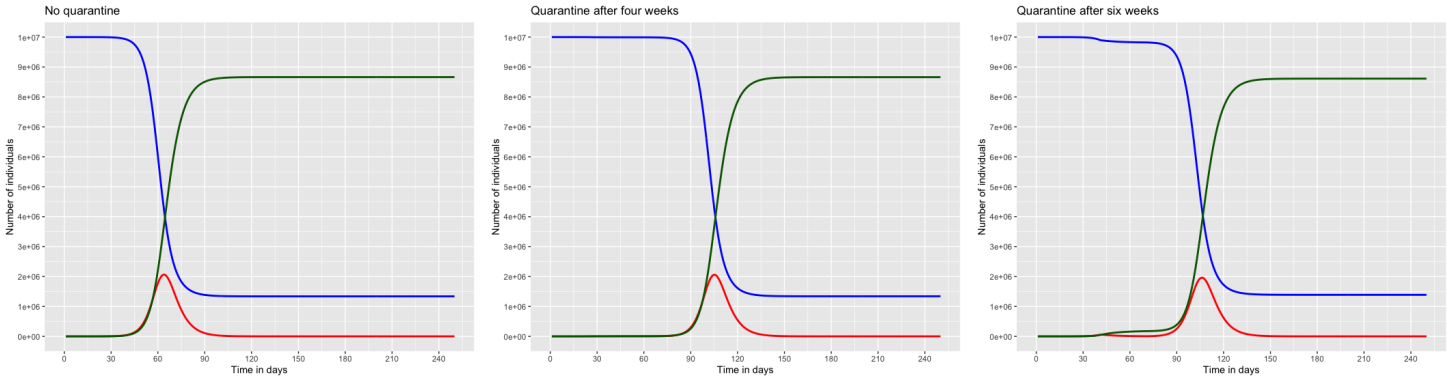


Figure 6.6: A COVID-19 epidemic without quarantine measures (left), and with a month long quarantine starting after four weeks (middle) and after six weeks (right).

Figure 6.7 shows that in this specific case, a month long quarantine after seven, eight, nine, or ten weeks could decrease the height of the peak of infected individuals as well as decrease the size of the epidemic. If the most important goal is to reduce the number of individuals that are infected at the same time, quarantine after eight weeks seems to be most effective. However, if the main goal is to reduce the size of the epidemic, quarantine after nine weeks seems to be the best strategy.

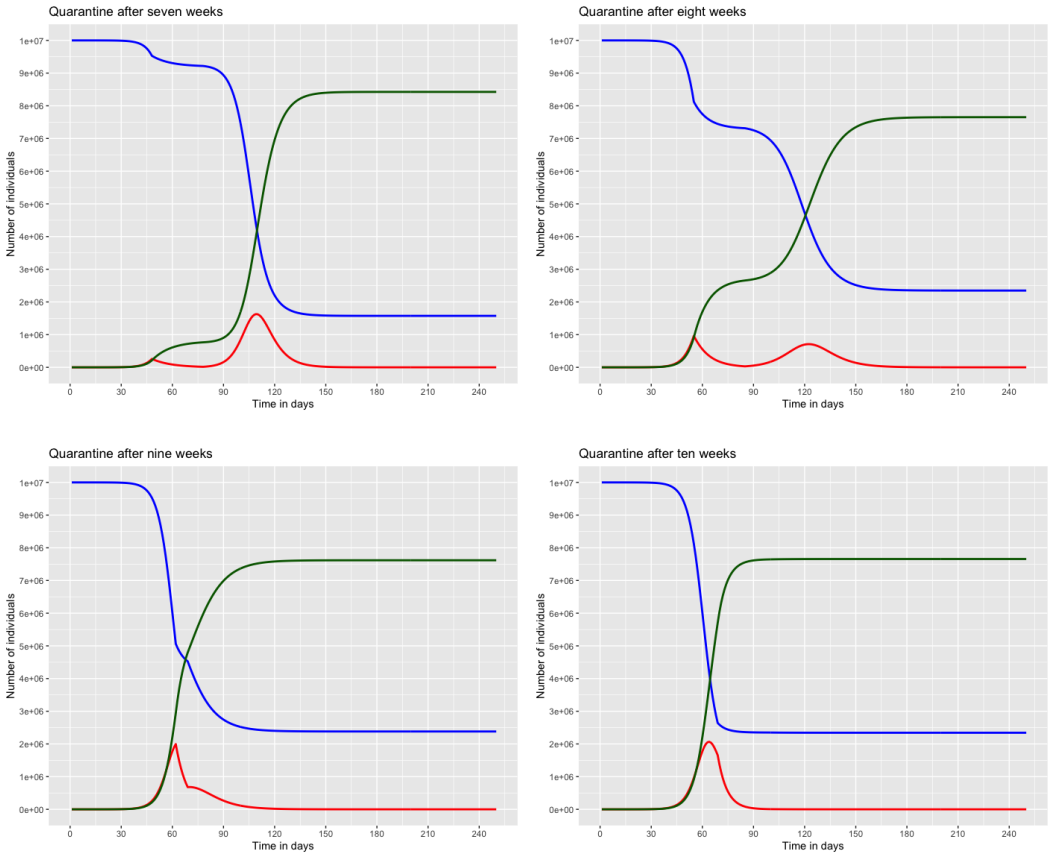


Figure 6.7: A COVID-19 epidemic with a month long quarantine starting after seven, eight, nine, and ten weeks.

Chapter 7

Discussion

7.1 Summary

In this paper we investigated the SIR model and found that it is a compartmental model consisting of three ordinary differential equations. Even though the model is based on a number of strong assumptions about population size and distribution, it's predictions are strikingly reliable.

We performed numerical analyses using the methods of Euler and Runge-Kutta, and we saw that Runge-Kutta 4 is significantly better at making accurate estimates for larger step sizes than Forward Euler. Qualitative analysis of the SIR model revealed the epidemic threshold to be $\frac{S}{N} = \frac{1}{R_0}$. That is, an epidemic will occur only when the proportion of susceptible individuals is greater than the multiplicative inverse of the basic reproduction number. The size of an epidemic, expressed as I_{total} or $S(\infty)$, was found to depend on both R_0 and $S(0)$.

In chapter 5 we derived the herd immunity threshold, the fraction of individuals needed to be vaccinated in order to prevent an epidemic, to be equal to $1 - \frac{1}{R_0}$. Following this, we analyzed the stability of the system's equilibrium points and saw that the disease free equilibrium is stable as long as $R_0 > \frac{1}{1-p}$. Finally, we studied the relationship between the vaccination rate and the epidemic size.

In the last chapter of this thesis we investigated the current COVID-19 pandemic. We approximated R_0 to be in the range of 2.8-3.3 and we illustrated how crucial the timing of quarantine measures is to it's effects.

7.2 Topics for future study

When searching for literature on the SIR model and it's variations, one comes across an overwhelmingly large amount of published articles, many of which are relatively similar. To illustrate, a PubMed search using the MeSH terms *((((((sir) OR svir) OR seir) OR sirs) AND model) AND susceptible)) AND infected*, resulted in 4203 hits, of which 3244 were published during this decade. Frankly, this made it quite difficult to determine how much the model actually has progressed since it was first developed. Therefore, in my opinion the field would benefit from a systematic literature review that summarizes current knowledge and identifies topics for future study.

When it comes to COVID-19, a lot of research is needed. Experiments such as the ones in this thesis provide a sneak peek and are a first step. But in order to really understand the disease, obviously many more studies are needed.

Bibliography

[1] Constantinos I. Siettos and Lucia Russo. "Mathematical modeling of infectious disease dynamics". In: *Virulence* 4:4 (2013), pp. 295–306.

[2] David L Heymann Máire A Connolly. "Deadly comrades: war and infectious diseases". In: *The Lancet* 360 (2002), pp. 23–24.

[3] Pauline van den Driessche. "Reproduction numbers of infectious disease models". In: *Infectious Disease Modelling* 2 (2017), pp. 288–303.

[4] editor. Last JM. *Dictionary of epidemiology.* New York: Oxford University Press, 2001.

[5] Valencia Higuera and Ann Pietrangelo. *How Are Diseases Transmitted?* `https://www.healthline.com/health/disease-transmission`. Accessed on 2019-09-11. Oct. 2016.

[6] Maria Kirwan Helen Barratt and Saran Shantikumar. *Epidemic theory techniques for analysis of infectious disease data.* `https://www.healthknowledge.org.uk/public-health-textbook/research-methods/1a-epidemiology/epidemic-theory`. Accessed on 2019-09-11. 2018.

[7] Md. Samsuzzoha. "A Study on Numerical Solutions of Epidemic Models". In: *Swinburne University of Technology* (2012).

[8] Helena Sofia Rodrigues. "Application of SIR epidemiological model: new trends". In: *International Journal of applied mathematics and informatics* 10 (2016), pp. 92–96.

[9] M. J. Keeling and L. Danon. "Mathematical modelling of infectious diseases". In: *British Medical Bulletin* 92 (2009), pp. 33–42.

[10] MA. Huppert and G. Katriel. "Mathematical modelling and prediction in infectious disease epidemiology". In: *Clin Microbiol Infect* 19 (2013), pp. 999–1005.

[11] Howard (Howie) Weiss. "The SIR model and the Foundations of Public Health". In: *MATerials MATemàtics* 2013.3 (2013), pp. 0001–17.

[12] Muhammad Ozair Ahmad Muhammad Asghar Ali Muhammad Rafiq. "Numerical Analysis of a Modified SIR Epidemic Model with the Effect of Time Delay". In: *Journal of Mathematics* 51 (2019), pp. 79–90.

[13] Germund Dahlquist and Åke Björck. *Numerical methods.* Dover, 2017.

[14] Badejo Oduyomi Micheal Bagbe Atinuke and Ayodeji Samson Bagbe. "Statistical Analysis of Ebola Virus Disease outbreak in some West Africa Countries using S-I-R Model". In: *Annals of Biostatistics Biometric Applications* 2(3) (2019).

[15] Centers for Disease Control and Prevention. *2014-2016 Ebola Outbreak in West Africa, Case Counts.* `https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/case-counts.html`, Last accessed on 2020-01-11. 2020.

[16] Angela S Clem. "Fundamentals of Vaccine Immunology". In: *J Glob Infect Dis.* 3(1) (2011), pp. 73–78.

[17] Zhou et al. "Economic evaluation of the routine childhood immunization program in the United States, 2009." In: *Pediatrics* 133(4) (2014), pp. 577–85.

[18] Walter A. Orenstein and Rafi Ahmed. "Simply put: Vaccination saves lives". In: *PNAS* 114 (2017), pp. 4031–4033.

[19] Ohanian E Glomski M. "Eradicating a Disease: Lessons from Mathematical Epidemiology". In: *The college mathematics journal* 43(2) (2012), pp. 123–132.

[20] World Health Organization. *Global Health Observatory.* `https://www.who.int/data/gho`, Last accessed on 2020-01-11. 2020.

[21] Lennart Råde and Bertil Westergren. *Mathematics Handbook for Science and Engineering.* Studentlitteratur, 2014.

[22] Yang et al. "Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions". In: *Journal of Thoracic Disease* (2020).

Appendix A

Mathematical proofs

A.1 Proof to theorem 5.4

Theorem 5.4. If J is a 2×2 matrix, then the sum of the eigenvalues of J is equal to the trace of J and the product of the eigenvalues is equal to the determinant of J.

Proof. Say $J = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ and $|\lambda I - J| = \begin{vmatrix} \lambda_1 - a_{11} & -a_{12} \\ -a_{21} & \lambda_2 - a_{22} \end{vmatrix} = 0.$

$$(\lambda_1 - a_{11})(\lambda_2 - a_{22}) - a_{12}a_{21} = 0$$

$$\lambda_1\lambda_2 - a_{22}\lambda_1 - a_{11}\lambda_2 + a_{11}a_{22} - a_{12}a_{21} = 0$$

$$\lambda^2 - (a_{22} + a_{11})\lambda + a_{11}a_{22} - a_{12}a_{21} = 0 \quad \lambda = \lambda_1 \vee \lambda = \lambda_2$$

$$\lambda_1 = \frac{(a_{22} + a_{11}) + \sqrt{(a_{22} + a_{11})^2 - 4(a_{11}a_{22} - a_{12}a_{21})}}{2}$$

$$\lambda_2 = \frac{(a_{22} + a_{11}) - \sqrt{(a_{22} + a_{11})^2 - 4(a_{11}a_{22} - a_{12}a_{21})}}{2}$$

From this we find the sum of the eigenvalues to be $\lambda_1 + \lambda_2 = a_{11} + a_{22}$, which indeed equals the trace of J. The product of the eigenvalues becomes

$$\lambda_1\lambda_2 = \frac{(a_{22} + a_{11})^2}{2^2} - \frac{(a_{22} + a_{11})^2 - 4(a_{11}a_{22} - a_{12}a_{21})}{2^2} = a_{11}a_{22} - a_{12}a_{21}$$

which indeed equals the determinant of J.

□