Analysis and Computational Study of SIR-Type Epidemiological Models for the Covid-19 Pandemic

Evridiki Klagkou

Committee
Michael Plexousakis (Supervisor)
Konstadia Lika
Georgios Makrakis

Department of Mathematics and Applied Mathematics

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
UNIVERSITY OF CRETE

Περίληψη

Οι επιδημίες έχουν υπάρξει μια τεράστια ανησυχία για την παγκόσμια υγεία. Όμως από τον 14ο αιώνα μέχρι τώρα έχουν γίνει πολλές ανακαλύψεις για τον τρόπο που λειτουργούν. Ειδικότερα, όταν εισήχθησαν μαθηματικές μέθοδοι ώστε να υποστηρίξουν τα δεδομένα.
Στα τέλη του 2019, ο ιός SARS-CoV-2 ή αλλιώς Covid-19 ξεκίνησε να εξαπλώνεται σε όλο τον κόσμο. Πολύ σύντομα, ο αριθμός των μολυσμένων που εμφάνισαν συμπτώματα καθώς και οι βαριά άρρωστοι συνέτριψαν το σύστημα υγείας σε πολλές χώρες. Ακόμη αυτός ο ιός οδήγησε σε πάνω από 4 εκατομμύρια θανάτους μέχρι τον Ιούλιο του 2020. Αυτή η πανδημία είχε τρομερές επιπτώσεις στην παγκόσμια οικονομία λόγω της αύξησης της ανεργίας, της μείωσης του ειδοδήματος, τις διαταραχές σε επιχειρήσεις.
Οι υπάλληλοι της δημόσιας υγείας χρησιμοποιούν επιδημιολογικά μοντέλα για την παρακολούθηση ασθενειών και έρευνα για πιθανές εξάρσεις καθώς και για παρατηρητική μελέτη, προκειμένου να εντοπιστούν οι παράγοντες κινδύνου και να εφαρμοστούν μέτρα ελέγχου της νόσου. Παρόλο που τα δεδομένα είναι σχεδόν πάντα διαθέσιμα από εμφανιζόμενες επιδημίες, είναι συχνά ελλιπή λόγω ανεπαρκούς αναφοράς. Συγκεκριμένα, για την επιδημία Covid-19 υπάρχουν ολοένα και περισσότερες ενδείξεις ότι μέρος της ταχείας εξάπλωσης αυτού του ιού οφείλεται σε ασυμπτωματικές λοιμώξεις. Λόγω αυτής της έλλειψης αξιόπιστων δεδομένων, χρησιμοποιήθηκαν μαθηματικά μοντέλα και προσομοιώσεις υπολογιστών για την εκτέλεση θεωρητικών πειραμάτων για την εκτίμηση των παραμέτρων του μηχανισμού μετάδοσης και της εξάπλωσης της νόσου. Επιπλέον, τέτοια πειράματα μπορεί να είναι χρήσιμα στη σύγκριση των αποτελεσμάτων των προληπτικών μέτρων, όπως η κοινωνική αποστασιοποίηση ή η καραντίνα.
Ένα από τα πιο γνωστά και επιτυχή επιδημιολογικά μοντέλα είναι το μοντέλο SIR, τα αποτελέσματα του οποίου είναι παρόμοια με αυτά που παίρνουμε χρησιμοποιώντας πραγματικά δεδομένα. Ο σκοπός αυτής της εργασίας είναι η αναλυτική και αριθμητική μελέτη ενός γενικευμένου SIR μοντέλου το οποίο περιλαμβάνει την κλάση των ασυμπτωματικών, και η σύγκριση των αποτελεσμάτων του με τα πραγματικά δεδομένα για τον Covid-19 από την Ελλάδα και άλλες χώρες.

Abstract

Epidemic outbreaks have been a major concern in public health throughout history. However from the 14th century till now there have been a lot of discoveries about them. Especially when mathematical methods were introduced to statistically support the data.
In late 2019 SARS-CoV-2 virus, or Covid-19 started spreading around the world. Soon after the number of symptomatically infected and severely ill individuals overwhelmed the medical system in many countries. It also lead to more than 4 million deaths by July 2020. This pandemic also had severe consequences in the global economy due to disruption in manufacturing and services, income reductions and rize of unemployment.
Public health officials use epidemiological models for disease surveilance and the investigation of outbreaks, along with observational studies, in order to identify risk factors and implement disease control measures. Although data are almost always available from occuring epidemics, they are often incomplete due to underreporting. In particluar, for the Covid-19 epidemic there is mounting evidence that some of the rapid spread of this virus has been driven by asymptomatic infections. Due to this lack of reliable data mathematical modeling and computer simulations have been used to perform theoretical experiments to estimate the parameters of the transmission mechanism and the spread of the disease. Moreover, such experiments may be useful in comparing the effects of preventive measures, such as social distancing or quarantine.
A well known epidemical model is the SIR model, as it gives results that are similar with the real data.
The aim of this thesis is the analytical and computational study of an extended SIR model which includes the class of asymptomatic individuals and compare its predictions with real Covid-19 data from Greece and elsewhere.

Acknowledgements

First of all, I would like to thank Professor Michael Plexousakis for supervising this thesis. Our collaboration was great, and it made me more interested in the field of Mathematical Biology and Epidemics. I also want to thank all the professors that helped me learn all these years and guided me to the path that I'm following right now. Lastly, I thank my family and friends for the love and support they provided me.

List of Figures

Contents

Chapter 1

Introduction

1.1 Epidemiology

Epidemiology is the study and analysis of the distribution and determinants of health-related events in specified populations, and the application of this study to the control of health problems. The word epidemiology comes from the Greek words ‘epi’,‘demos’ and ‘logos’, meaning ‘upon’, ‘people’ and ‘study’, respectively, in other words, the study of ‘what is upon a population’. The mathematical aspect of Epidemiology is creating a model that can project how an infectious disease progresses with time. This is a very useful tool because it helps to better understand the infection and to make sure that the right interventions are being made.
The Greek physicist Hippocrates was the first epidemiologist,suggesting that a disease has a logic behind it and that there is a relation between the disease, its spread and the environment. In the 16th century an Italian doctor named Girolamo Fracastoro was the first one to propose that the particles which cause a disease are alive. He also promoted that personal and environmental hygiene help to prevent a disease. Not long after that physician Quinto Tiberio Angelerio published a manual with 57 rules that help to prevent a disease such as social distancing and washing produce.
In the 19th century John Snow, known as the father of epidemiology, investigated the causes of cholera epidemics. He used chlorine to clean the water and he managed to end the outbreak. This was a major event in the history of public health and it was the fist time that epidemiology helped to shape public health policies around the world. The early 20th century was the first time that mathematical methods were used in epidemiology, adding statistical support to the field, by Ronald Ross, Janet Lane–Clayton, Anderson, Gray, McKendrick, and others. Also at that time epidemiologists extended their methods to noninfectious diseases such as cancer, proving the suspicion that smoking was linked to lung cancer. The latest epidemic outbreak still going on is the Coronavirus disease.

1.2 Covid-19

The coronavirus disease,or Covid-19, is an infectious disease, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that started infecting people in December 2019. The first confirmed case was in Wuhan, China and soon after it spread all over the world, making it an ongoing pandemic. There are several symptoms but the most common ones are fever, cough,

fatigue, breathing difficulties, and loss of smell and taste. However, there are several cases where the infected person does not show any symptoms but is still able to spread the disease.
According to [12] from the people that do show symptoms about 81% develop mild symptoms, 14% develop severe symptoms and 5% develop critical symptoms. At least a third of the people who are infected don't develop any notable symptoms. The symptoms begin to show between 1 to 14 days after the exposure and they can last for more than 2 weeks in a lot of cases.
As of July 2021 there have been 185 million confirmed cases and 4 million deaths worldwide. There have been several lockdowns worldwide in order to contain the outbreak. Vaccines against the SARS-CoV-2 virus were granted marketing authorisation and use in the EU and USA in mid 2020 with the purpose of ensuring public health and controlling the epidemic.

1.3 Infection and Recovery rate

We will study briefly the main compartmental model called SIR. We divide the population into three groups: Susceptible, Infected and Recovered individuals. In order to understand how it came about we shall make the following, usual among epidemiological models, assumptions:

- After contacting the infection a person either dies or develops immunity to the disease
- The rate of infection is proportional to the number of contacts between Infected and Susceptible individuals
- All individuals are equally vulnerable
- The total population is stable in the sense that we ignore births, but not deaths, and the latter are included in the total
- If the total population is N, then each individual makes βN contacts per unit time

To find the infection and recovery rates, using the above assumptions, we argue as follows: We denote by $S(t)$ the susceptible individuals and by $I(t)$ the infected individuals, with t denoting time.
Firstly for the infection rate, we already know the number of contacts that each individual makes, but we also need the probability of an Infected person to make contact with a Susceptible individual. This probability is $p = S/N$. Therefore, the number of new infections per unit time per infected individual is $\beta N \frac{S}{N} I = \beta SI$.
The calculation of the recovery rate is much simpler because it is not affected by the contacts someone makes, therefore, it is simply αI. The important thing now is to understand what this means. Assume that a group of people got infected at time $t = 0$ and u(s) are still infected at time $t = s$. If a fraction α of them recovers per unit time then the equation for the recovery is $u' = -\alpha u$, or, if we integrate,

$$u(s) = u(0)e^{-\alpha s} \Rightarrow \frac{u(s)}{u(0)} = e^{-\alpha s}.$$

This means that the time that an individual remains infected follows the exponential distribution with average value

$$\int_0^\infty s e^{-\alpha s} ds = \frac{1}{\alpha}.$$

Therefore, the infection rate is the inverse of the recovery period.

1.4 The basic reproduction number

The *basic reproduction number*, R_0, is the expected number of secondary cases produced by a single infection in a completely susceptible population. Therefore

$$R_0 \propto (\frac{\text{infection}}{\text{contact}})(\frac{\text{contact}}{\text{time}})(\frac{\text{time}}{\text{infection}}),$$

which means that R_0 is a dimensionless number and not a rate. Another way of writing the expression for R_0 is

$$R_0 = \tau * c * d,$$

where τ is the transmissibility, c is the average rate of contact between susceptible and infected individuals and d is the duration of infectiousness.
In simpler models, R_0 is usually calculated easily through the equations, but this is not the case for all models. If we have a system of equations

$$x_i' = f_i(x) = \mathcal{F}_i(x) - \mathcal{V}_i(x), \qquad 1 \leq i \leq m,$$

with m the number of components that we consider and $\mathcal{F}_i(x)$ is the rate of new infections in compartment i, while $\mathcal{V}_i(x)$ is the rate of transfer of individuals in i minus the rate of transfer of individuals out of i. We define

$$F_{ij} = [\frac{\partial \mathcal{F}_i}{\partial x_j}(x_0)] \text{ and } V_{ij} = [\frac{\partial \mathcal{V}_i}{\partial x_j}(x_0)] \text{ with } 1 \leq i, j \leq m.$$

Then,

$$R_0 = \rho(FV^{-1}),$$

where $\rho(M)$ denotes the spectral radius of a matrix M.
There is a second method for calculating R_0. If we have a system of equations $x_i' = f_i(x)$, $1 \leq i \leq m$, we focus on the equations for the population groups that have the disease, whether that is Asymptomatics, Symptomatics, Infected, Exposed, etc. Let's assume that in our example indicators 2 to $m-1$ are the ones that we focus on. We calculate the Jacobian matrix for these equations

$$J_{ij} = [\tfrac{\partial f_i}{\partial x_j}], \text{ where } 2 \leq i, j \leq m-1.$$

If we set the determinant of this matrix equal to zero, this gives us the crucial value $R_0 = 1$.

1.5 Ruth-Hurwitz stability criteria

The Ruth-Hurwitz criteria is a helpful tool that will be used in the upcoming chapters to determine the sign of the roots of a polynomial equation. We will employ this lemma for second and third degree polynomials.

Lemma 1.5.1.

1. *The second-degree polynomial* $P(s) = s^2 + \alpha_1 s + \alpha_0$ *has both roots in the open left half plane (and the system with characteristic equation* $P(s) = 0$ *is stable) if and only if both coefficients satisfy* $\alpha_i > 0$

2. *The third-degree polynomial* $P(s) = s^3 + \alpha_2 s^2 + \alpha_1 s + \alpha_0$ *has all roots in the open left half plane if and only if* α_2, α_0 *are positive and* $\alpha_2 \alpha_1 > \alpha_0$

Chapter 2

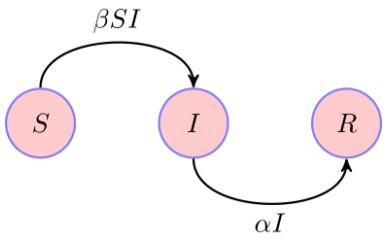
The SIR Model

2.1 The basic model

One of the first successes of Mathematical Biology was the introduction of a model the results of which agreed with the behaviour of epidemics.This model is the so-called SIR model and was proposed by Kermack and McKendrick in 1927. In the SIR model the total population is divided into three groups, Susceptible $S(t)$, those who are capable of catching the disease, Infected $I(t)$, those who have the disease, and Recovered $R(t)$, those who have had the disease and are now immune to it. We shall assume that

- Infected individuals can transfer the disease to Susceptible individuals with a rate proportional to the the contacts between those two groups
- Infected individuals recover with a rate proportional to their number
- We don't have any deaths due to the disease
- The total population is stable, in the sence that births are neglected

Schematically the so called compartments of the SIR model and the transfer of individuals between them are shown below:



Our model is then

$$\frac{dS}{dt} = -\beta SI, \tag{2.1.1}$$

$$\frac{dI}{dt} = \beta SI - \alpha I, \tag{2.1.2}$$

$$\frac{dR}{dt} = \alpha I. \tag{2.1.3}$$

Here, β is the *infection rate*, α is the *recovery rate*, and the ratio $\rho = \frac{\alpha}{\beta}$ is the *relative recovery rate*. The model is supplemented with the initial conditions

$$S(0) = S_0 > 0, I(0) = I_0 > 0 \text{ and } R(0) = 0.$$

The fact that the population is stable, is also clear from the model equations because if we add equations (2.1.1)-(2.1.3) we have

$$(S + I + R)' = 0 \Rightarrow N' = 0 \Rightarrow N = S_0 + I_0.$$

Here and in the sequel, prime will denote the derivative with respect to the variable t. From the first equation it is clear that the number of susceptible individuals decreases with time. From the second equation we see that the infected increase as long as $S > \alpha/\beta$. If the initial number of susceptibles is smaller than α/β then the infection dies out since

$$S_0 < \alpha/\beta \Rightarrow N - I_0 < \alpha/\beta \Rightarrow I_0 > \frac{\alpha}{\beta}(R_0 - 1), \tag{2.1.4}$$

where

$$R_0 = \frac{\beta N}{\alpha} = \frac{1}{\rho}N,$$

is the basic reproduction number. Using the previous definition of R_0 we can see that for the SIR model $\beta = \tau c$ and $d = 1/\alpha$. From the equation (2.1.4) it is obvious that we have a very important phenomenon happening: if $R_0 > 1$ the infection will spread, while if $R_0 < 1$ the infection will die out.

Since $S(t)$ is a positive, decreasing function this means that the limit $S_\infty = \lim_{t\to\infty} S(t)$ exists. Similarly $R'(t) \geq 0$, therefore $R(t)$ is an increasing function but because $R \leq N$ always, we have that the limit $R_\infty = \lim_{t\to\infty} R(t)$ exists as well.Lastly, because the population is stable at all times $I(t) = N - R(t) - S(t)$, so that the limit $I_\infty = \lim_{t\to\infty} I(t)$ also exists.

If we combine the equations (2.1.1) and (2.1.3) we get that

$$\frac{dS}{dR} = -\frac{\beta}{\alpha}S,$$

and with integration by parts we have

$$S = S_0 \exp(-\frac{\beta}{\alpha}R) \geq S_0 \exp(-\frac{\beta}{\alpha}N),$$

which proves us that $S_\infty > 0$, meaning that the infection will not die out due to the lack of susceptibles.

We will now deal with the equation (2.1.1) and (2.1.2) only since the Recovered individuals can be found through the Susceptible and Infected. Every point of the I=0 axis is an equilibrium point for the system (2.1.1),(2.1.2). Because of that, the only equilibrium that can occur is a disease free equilibrium. The phase plane between I and S is

$$\frac{\frac{dI}{dt}}{\frac{dS}{dt}} = \frac{\beta SI - \alpha I}{-\beta SI} = -1 + \frac{\alpha}{\beta S} \Rightarrow I(S) = (N - S) + \frac{\alpha}{\beta} \log \frac{S}{S_0}$$

The solution curves in the SI phase plane are described by the equation $\phi(S, I) = S + I - \rho \log S =$ constant and are shown in 2.1. Since S(t) is decreasing all the curves are going from right to left.

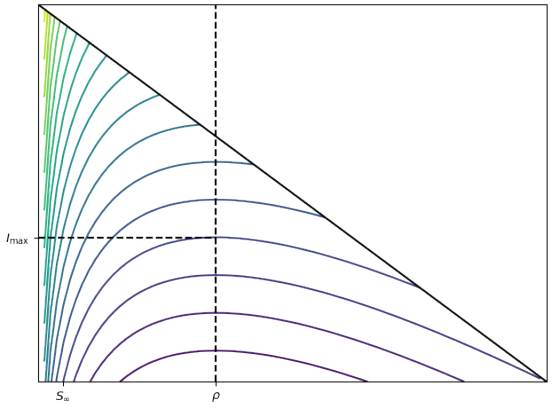


Figure 2.1: SIR phase plane

As we can see from Figure 2.1, the Infected individuals reach a maximum when $S = \rho$. Therefore we have that

$$I_{max} = (N - \rho) + \rho \log \frac{\rho}{S_0}.$$

However, from Figure 2.1 something really important can also be observed. If the initial condition S_0 satisfies $S_0 < \rho$ then the epidemic cannot occur, since $I(t)$ goes to zero. However, if $S_0 > \rho$ the number of infectives increases until S gets its maximum value at $S = \rho$ and then falls to zero. Since $I_\infty = 0$, we have that $S_\infty = N - R_\infty$ and by using the expresion for S and R we get that S_∞ is a root of the equation

$$z = S_0 \exp(-\frac{\beta}{\alpha}(N - z)).$$

If $f(z) = S_0 \exp(-\frac{\beta}{\alpha}(N - z)) - z$, then $f(0) = S_0 \exp(-\frac{\beta}{\alpha}N) > 0$, and $f(N) = S_0 - N < 0$ since $N = S_0 + I_0 > S_0$. From the intermediate value theorem we see that there exist a root z_* of the equation $f(z) = 0$ in the interval $(0, N)$. Moreover, $f'(z_*) = \frac{\beta}{\alpha}z_* - 1$ and $f''(z) = \frac{\beta^2 S_0}{\alpha^2}\exp(-\frac{\beta}{\alpha}(N - z)) > 0$. And since $f(N) < 0$ there is exactly one root and $z_* < \rho = \alpha/\beta$.
If we combine equations (2.1.1) and (2.1.2) we have that $(S + I)' = -\alpha I < 0$ meaning that it is a decreasing function. Therefore, the limit $\lim_{t\to\infty}(S+I)(t)$ exists and $\lim_{t\to\infty}(S+I)'(t) = 0$. This means that $I_\infty = \lim_{t\to\infty} I(t) = 0$, and $\lim_{t\to\infty}(S + I)(t) = \lim_{t\to\infty} S(t) = S_\infty$. If we integrate the sum of (2.1.1) and (2.1.2) we get

$$\alpha \int_0^\infty I(t)dt = -\int_0^\infty (S + I)'(t) = S_0 + I_0 - S_\infty = N - S_\infty.$$

From (2.1.1) we have $\frac{S'}{S} = -\beta I \Rightarrow \ln(S)' = -\beta I$, which gives

$$\ln \frac{S_0}{S_\infty} = \beta \int_0^\infty I(t)dt = \frac{\beta}{\alpha}(N - S_\infty) = R_0(1 - \frac{S_\infty}{N}).$$

Therefore, we have

$$\ln \frac{S_0}{S_\infty} = R_0(1 - \frac{S_\infty}{N}).$$

This is a relationship between the basic reproduction number and the final size of the disease. Knowing this relation helps us find how many individuals will not get sick. This equation defines unambiguously S_∞. Indeed, let us define the function $g(x) = \ln \frac{S_0}{x} - R_0(1 - \frac{x}{N})$. Then $\lim_{x\to 0^+} g(x) > 0$ and $g(N) < 0$.Therefore we have at least one solution between 0 and N. $g'(x) = \frac{R_0}{N} - \frac{1}{x}$ which is zero if $x = \frac{N}{R_0}$. If $R_0 \leq 1$ then $g'(x) < 0$ if $0 < x < \frac{N}{R_0}$ so we have only one solution. If $R_0 > 1$, then g is decreasing in $(0, \frac{N}{R_0})$ and increasing in $(\frac{N}{R_0}, N)$, therefore once again there is only one point where $g(x) = 0$, and this is S_∞. In both cases we get that $S_\infty < \frac{N}{R_0}$ and $g(\frac{S_0}{R_0}) = \ln R_0 - R_0 + \frac{S_0}{N} \leq \ln R_0 - R_0 + 1 < 0$, since $\ln x < x + 1$ if $x > 0$. This gives us a simpler estimate

$$S_\infty < \frac{S_0}{R_0}.$$

The Jacobian matrix along the axis $I = 0$ is

$$J = \begin{pmatrix} 0 & -\beta S \\ 0 & \beta S - \alpha \end{pmatrix}.$$

The eigenvalues of this matrix are the roots of the equation

$$\lambda^2 + (\alpha - \beta S)\lambda = 0.$$

One eigenvalue is zero. If $R_0 < 1$ then $\beta N < \alpha$ so then $\beta S < \alpha$. The other eigenvalue is negative. This means that the disease-free equilibrium is stable, but not asymptomatically stable. We recall that an equilibrium is(locally) stable if initial conditions that start near an equilibrium point stay near that equilibrium point and an equilibrium point is (locally) asymptotically stable if it is stable and, in addition, the state of the system converges to the equilibrium point as time increases.

2.2 The SIR model with deaths

The previous version of the SIR model did not take into account deaths due to the disease. In this section the model will be slightly altered to include deaths. To do that, we assume:

- Infected individuals may die after a period of time with mortality rate δ
- We will not count dead individuals as part of the total population, making in not stable
- Because of the fact that the population is not stable, it makes sence to have a population-dependent infection rate

With these assumptions in mind we consider the model

$$S' = -\beta(N)SI, \tag{2.2.1}$$

$$I' = \beta(N)SI - (\alpha + \delta)I, \tag{2.2.2}$$

$$R' = \alpha I. \tag{2.2.3}$$

If we add those three equations we get

$$N' = -\delta I. \tag{2.2.4}$$

The original population is $N(0) = N_0 = S_0 + I_0$, and $N(t) = S(t) + I(t) + R(t)$ for every t. In this model the basic reproduction number is

$$R_0 = \frac{N_0\beta(N_0)}{\alpha + \delta}.$$

As we can see from equation (2.2.2), in order for the infection to spread we need $\beta(N)S-(\alpha+\delta) > 0$. If we combine equations (2.2.1) and (2.1.2) we see that

$$\int_0^\infty I(s)ds = \frac{1}{\alpha + \delta}N_0 - \frac{1}{\alpha + \delta}S_\infty.$$

Also, from equation (2.2.4) we have

$$N_\infty = N_0 - \mu\int_0^\infty I(s)ds \Rightarrow N_\infty = \frac{\alpha}{\alpha + \delta}N_0 + \frac{\delta}{\alpha + \delta}S_\infty.$$

In analogy with the basic SIR model we can find the relationship between the basic reproduction number and the final size of the disease. From (2.2.1) we have $\frac{S'}{S} = -\beta(N)I$, which by integration gives

$$\ln\frac{S_0}{S_\infty} = -\int_0^\infty \beta N(t)I(t)dt < -\beta N(0)\int_0^\infty I(t)dt = \frac{\beta N(0)}{\alpha + \delta}N_0 - \frac{\beta N(0)}{\alpha + \delta}S_\infty.$$

Therefore we have

$$\ln\frac{S_0}{S_\infty} < R_0 - \frac{\beta N(0)}{\alpha + \delta}S_\infty = R_0(1 - \frac{S_\infty}{N_0}).$$

This is very similar to the result that we got in the previous model, the only differences are that we have an inequality instead of an equality, and we use the initial total population because this model is not stable. As with the SIR model we also have the simpler estimate

$$S_\infty < \frac{S_0}{R_0}.$$

In this model we have only one equilibrium which is the disease free one since every point in the $I = 0$ axis is an equilibrium point for the system (2.2.1)-(2.2.2). Along the axis $I = 0$ the Jacobian matrix is

$$J = \begin{pmatrix} 0 & -\beta(N)S \\ 0 & \beta(N)S - (\alpha + \delta) \end{pmatrix}.$$

The eigenvalues of this matrix are the roots of the equation

$$\lambda^2 + (\alpha + \delta - \beta(N)S)\lambda = 0.$$

One eigenvalue is zero. If $R_0 < 1$ then $N_0\beta(N_0) < \alpha + \delta$ and so $\beta(N)S < \alpha + \delta$. Since the population is decreasing with time, then the second equilibrium is negative. This means that the disease-free equilibrium is stable, but not asymptomatically stable.

2.3 The SIR model with vaccination

In this section, a version of the SIR model with vaccination,births and deaths will be introduced. So we assume

- Susceptible individuals can get vaccinated
- All individuals are susceptible at birth
- Individuals in each population group may die from natural causes
- The birth and death rate are equal

We have the model

$$S' = -\beta SI + \mu(N - S) - \phi S, \tag{2.3.1}$$

$$I' = \beta SI - (\alpha + \mu)I, \tag{2.3.2}$$

$$R' = \alpha I + \phi S - \mu R. \tag{2.3.3}$$

As we can see from the equations $N' = (S + I + R)' = 0$ which means that the population is stable. For the system (2.3.1)-(2.3.2) we have two equilibrium points. The first one is the disease free equilibrium $(S_*, I_*) = (\frac{\mu N}{\mu+\phi}, 0)$ and the second one is the endemic equilibrium point $(S_*, I_*) = (\frac{\alpha+\mu}{\beta}, \frac{\mu N}{\mu+\alpha} - \frac{\mu+\phi}{\beta})$.
The basic reproduction number for this model is

$$R_0 = \frac{\mu\beta N}{(\mu + \phi)(\mu + \alpha)}.$$

The Jacobian matrix at the point $(\frac{\mu N}{\mu+\phi}, 0)$ is

$$J = \begin{pmatrix} -\mu - \phi & -\frac{\beta\mu N}{\mu+\phi} \\ 0 & \frac{\beta\mu N}{\mu+\phi} - (\alpha + \mu) \end{pmatrix}.$$

The eigenvalues of this matrix are the roots of the equation

$$\lambda^2 + \lambda(\alpha + \mu - \frac{\beta\mu N}{\mu + \phi} + \mu + \phi) + (\alpha + \mu)(\mu + \phi) - \beta\mu N = 0.$$

If $R_0 < 1$ then $\beta\mu N < (\mu + \phi)(\mu + \alpha)$ so all the coefficients are positive and therefore all roots have negative real parts. This means that the disease-free equilibrium is stable.
The Jacobian matrix at the point $(\frac{\alpha+\mu}{\beta}, \frac{\mu N}{\mu+\alpha} - \frac{\mu+\phi}{\beta})$ is

$$J = \begin{pmatrix} -\frac{\beta\mu N}{\mu+\alpha} & -(\alpha + \mu) \\ \frac{\beta\mu N}{\mu+\alpha} - (\mu + \phi) & 0 \end{pmatrix}.$$

The eigenvalues of this matrix are the roots of the equation

$$\lambda^2 + R_0(\mu + \phi)\lambda + (\mu + \alpha)(\mu + \phi)(R_0 - 1) = 0.$$

If $R_0 > 1$ then all the coefficients are positive and therefore all roots have negative real parts. This means that the endemic equilibrium is also stable.

Chapter 3

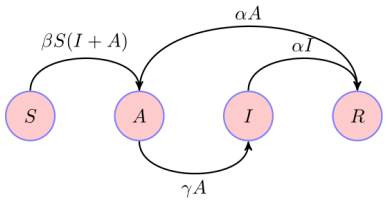
The SAIR model

3.1 Introduction of the model

In Covid-19 and many other diseases there are infected individuals that never show any symptoms, collectively called Asymptomatics. In this section we will introduce a new model, similar to the SIR model that includes this new group of individuals. We assume that:

- We have four groups of individuals, the Susceptibles, Asymptomatics, Infected and Recovered
- Both Infected and Asymptomatics can spread the disease with the same infection rate
- A fraction of Asymptomatics become infected after some time
- Both Asymptomatic and Infected recover with the same rate

These assumptions are shown schematically below: This leads us to the model



$$\frac{dS}{dt} = -\beta S(I + A), \tag{3.1.1}$$

$$\frac{dA}{dt} = \beta S(I + A) - (\gamma + \alpha)A, \tag{3.1.2}$$

$$\frac{dI}{dt} = \gamma A - \alpha I, \tag{3.1.3}$$

$$\frac{dR}{dt} = \alpha(I + A). \tag{3.1.4}$$

The population in this model is stable. The only equilibrium point is the point $(N, 0, 0, 0)$. To find R_0 we need to calculate the determinant of the Jacobian matrix for the Asymptomatics and Infected in the equilibrium point and see where it is zero. This gives us that $R_0 = \frac{\beta N}{\alpha}$. For this model we may only deal with the first three groups of people since the Recovered can be found from the other ones.
It is important to note that in this model, if we set $P = A + I$, then this combined population group has the exact same behaviour as the Infected in the basic SIR model. This means that

$$P(S) = (N - S) + \frac{\alpha}{\beta} \ln \frac{S}{S_0},$$

and similarly we get the largest value for this combined population group when $S = \frac{\alpha}{\beta}$. It is equal to

$$P_{max} = (N - \frac{\alpha}{\beta}) + \frac{\alpha}{\beta} \ln \frac{\alpha}{\beta S_0}.$$

If we combine (3.1.1),(3.1.2) and (3.1.3) we have that

$$S_\infty - N = \int_0^\infty (S + A + I)'(t)dt = -\alpha \int_0^\infty (A + I)(t)dt.$$

Equation (3.1.1) gives us that $\frac{S'}{S} = -\beta(A + I)$, so that

$$\ln \frac{S_0}{S_\infty} = \beta(N - S_\infty) = R_0(1 - \frac{S_\infty}{N}).$$

As with the SIR model the simpler estimate also holds:

$$S_\infty < \frac{S_0}{R_0}.$$

The Jacobian matrix at the point $(N, 0, 0, 0)$ is

$$J = \begin{pmatrix} 0 & -\beta N & -\beta N \\ 0 & \beta N - (\gamma + \alpha) & \beta N \\ 0 & \delta & -\alpha \end{pmatrix}.$$

The eigenvalues of this matrix are the roots of the equation

$$\lambda^3 - \lambda^2(R_0 - 1 - (\gamma + \alpha) - \lambda(R_0 - (\gamma + \alpha) + \gamma R_0) = 0.$$

If $R_0 < 1$ then all the coefficients are positive and since the constant term is zero it satisfies the Ruth–Hurwitz criteria and therefore all roots have negative real parts. This means that the disease-free eigenvalue is stable.

3.2 The SEIR model

In this section we will study a different model, in which we will use Exposed individuals instead of Asymptomatics. We make the following assumptions:

- Newborns are Susceptible and the birth rate is μN
- There are deaths in each group due to natural causes
- Contacts of Suscetibles with Infected individuals lead to exposure at a rate proportional to the number of contacts
- Exposed individuals don't spread the disease
- Infected individuals may die from the disease

With all that, the model is

$$\frac{dS}{dt} = \mu(N - S) - \beta SI, \tag{3.2.1}$$

$$\frac{dE}{dt} = \beta SI - (\mu + \gamma)E, \tag{3.2.2}$$

$$\frac{dI}{dt} = \gamma E - (\mu + \delta + \alpha)I, \tag{3.2.3}$$

$$\frac{dR}{dt} = \alpha I - \mu R, \tag{3.2.4}$$

$$\frac{dD}{dt} = \delta I. \tag{3.2.5}$$

This model has two equilibrium points, the disease-free equilibrium $(S_*, E_*, I_*, R_*) = (N, 0, 0, 0)$ and the endemic equilibrium

$$(S_*, E_*, I_*, R_*) = \left(\frac{(\mu + \gamma)(\mu + \delta + \alpha)}{\beta\gamma}, \frac{\mu N}{\mu + \gamma} + \frac{\mu(\mu + \delta + \alpha)}{\beta\gamma}, \right.$$
$$\left. \frac{\gamma\mu N}{(\mu + \gamma)(\mu + \delta + \alpha)} - \frac{\mu}{\beta}, \frac{\gamma\alpha N}{(\mu + \gamma)(\mu + \delta + \alpha)} - \frac{\alpha}{\beta} \right)$$

We can find the basic reproduction number by taking the Jacobian matrix of the equations (3.2.2)-(3.2.3) in one of the equilibrium points and set its determinant equal to zero. We have

$$R_0 = \frac{\beta\gamma N}{(\mu + \gamma)(\mu + \alpha + \delta)}.$$

The Jacobian matrix for the disease-free equilibrium is

$$J = \begin{pmatrix} -\mu & 0 & -\beta N \\ 0 & -(\mu + \gamma) & \beta N \\ 0 & \gamma & -(\mu + \delta + \alpha) \end{pmatrix}.$$

The eigenvalues of this matrix are the roots of the equation

$$\lambda^3 + \lambda^2(\mu(2\mu + \gamma + \delta + \alpha)) + \lambda(\mu(1 - R_0)) = 0.$$

If $R_0 < 1$ then all the coefficients are positive and since the constant term is zero it satisfies the Routh-Hurwitz criteria and therefore all roots have negative real parts. This means that the disease-free equilibrium is stable.
The Jacobian matrix for the endemic equilibrium is

$$J = \begin{pmatrix} -\mu - \beta(\frac{\gamma\mu N}{(\mu+\gamma)(\mu+\delta+\alpha)} - \frac{\mu}{\beta}) & 0 & -\frac{(\mu+\gamma)(\mu+\delta+\alpha)}{\gamma} \\ \beta(\frac{\gamma\mu N}{(\mu+\gamma)(\mu+\delta+\alpha)} - \frac{\mu}{\beta}) - (\mu+\gamma) & \frac{(\mu+\gamma)(\mu+\delta+\alpha)}{\gamma} & \\ 0 & \gamma & -(\mu+\delta+\alpha) \end{pmatrix}.$$

The eigenvalues of this matrix are the roots of the equation

$$\lambda^3 + \lambda^2(2\mu + \gamma + \delta + \alpha) + \lambda(2\mu + \gamma + \delta + \alpha)(2\mu + R_0) + \mu(R_0 - 1) = 0.$$

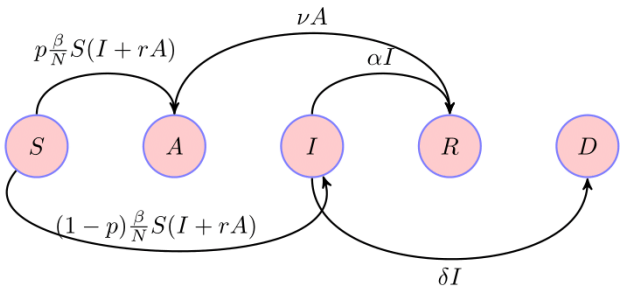
If $R_0 > 1$ then all the coefficients are positive and also the product of the coefficient of λ and λ^2 is greater than the constant term. Therefore all roots have negative real parts. This means that the endemic equilibrium is stable.

3.3 A general model

In this section we will introduce a very general model that in some cases can be simplified to be SIR or the SAIR model that we saw earlier.

- Firstly, we assume that both Infective and Asymptomatic individuals can infect Susceptible individuals, but this time with *different* infection rates
- We have included a probability as to whether a person who gets the disease will show symptoms or not
- There are different recovery rates for Asymptomatic and Infected individuals
- Lastly we assume that only Infected people may die from the disease

The model is shown schematically below: With these assumptions, we now analyse the model



$$\frac{dS}{dt} = -\frac{\beta}{N}S(I + rA), \tag{3.3.1}$$

$$\frac{dA}{dt} = p\frac{\beta}{N}S(I + rA) - \nu A, \tag{3.3.2}$$

$$\frac{dI}{dt} = (1-p)\frac{\beta}{N}S(I + rA) - (\alpha + \delta)I, \tag{3.3.3}$$

$$\frac{dR}{dt} = \nu A + \alpha I, \tag{3.3.4}$$

$$\frac{dD}{dt} = \delta I. \tag{3.3.5}$$

As we can see if $r = 0$, $p = 0$ and $\nu = 0$ then we have the SIR model.
This model has only one equilibrium point and it is $(N, 0, 0, 0, 0)$, and at this point the Jacobian matrix for the Asymptomatic and Infeted is

$$J = \begin{pmatrix} p\beta r - \nu & p\beta \\ (1-p)\beta r & (1-p)\beta - (\alpha + \delta) \end{pmatrix}.$$

If we set the determinant of this matrix to be zero we get that

$$R_0 = \frac{(\alpha + \delta)p\beta r + \nu(1-p)\beta}{\nu(\alpha + \delta)}.$$

If we combine (3.3.1)-(3.3.3) we have that $(S + A + I)'(t) = -\nu A - (\alpha + \delta)I$ and by integration we get

$$S_\infty - N = -\nu \int_0^\infty A(t)dt - (\alpha + \delta)\int_0^\infty I(t)dt. \tag{3.3.6}$$

From (3.3.1) we have that $\frac{S'}{S} = -\frac{\beta}{N}(I + rA)$ which gives us

$$\ln \frac{S_0}{S_\infty} = \frac{\beta}{N}(\int_0^\infty I(t)dt + r\int_0^\infty A(t)dt). \tag{3.3.7}$$

From (3.3.2) we get that

$$-\nu \int_0^\infty A(t)dt = -A_0 - \frac{p\beta}{N}\int_0^\infty S(I + rA)dt. \tag{3.3.8}$$

and similarly from (3.3.3)

$$-(\alpha + \delta)\int_0^\infty I(t)dt = -I_0 - \frac{(1-p)\beta}{N}\int_0^\infty S(I + rA)dt. \tag{3.3.9}$$

If we combine the equations (3.3.6),(3.3.8) and (3.3.9) we get

$$\int_0^\infty S(I + rA)dt = \frac{N}{\beta}(S_0 - S_\infty).$$

Now we can go back to the equation (3.3.7) and with the above equation it becomes

$$\ln\frac{S_0}{S_\infty} = \frac{R_0}{N}(S_0 - S_\infty) + \frac{\beta}{N}(\frac{I_0}{\alpha+\delta} + \frac{rA_0}{\nu}).$$

But $R_0 S_0 + \beta(\frac{I_0}{\alpha+\delta} + r\frac{A_0}{\nu}) \simeq R_0 N$, which means that in this model

$$\ln\frac{S_0}{S_\infty} \simeq R_0(1 - \frac{S_\infty}{N}).$$

The simpler estimate $S_\infty < \frac{S_0}{R_0}$ also holds in this case as well.
The Jacobian matrix at the first three groups is

$$J = \begin{pmatrix} 0 & -\beta r & \beta \\ 0 & p\beta r - \nu & p\beta \\ 0 & (1-p)r\beta & (1-p)\beta - (\alpha+\delta) \end{pmatrix}.$$

The eigenvalues of this matrix are the roots of the equation

$$\lambda^3 - \lambda^2[(\alpha+\delta+\nu)(R_0-1) - \frac{(\alpha+\delta)p\beta r}{\nu} - \frac{\nu(1-p)\beta}{\alpha+\delta}] - \lambda(\nu(\alpha+\delta)(R_0-1)) = 0$$

If $R_0 < 1$ then all the coefficients are positive and because there isn't a fixed term it satisfies the Routh-Hurwitz criteria and therefore all roots have negative real parts. This means that the disease free equilibrium is stable.

3.4 The dimensionless form of the model

To simplify the numerical calculations of this new model we derive its dimensionless form. The variables S, A, I, R, D and N denote population, the variable β denotes $(time \cdot population)^{-1}$ and the variables ν, α, δ have the dimension of $time^{-1}$. The variables r and p are dimensionless.
Now that we have determined the dimensions of each variable we may transform them to non-dimensional ones using $S = S_{new} = \frac{S}{N}$ and similarly for the sizes of the other compartments.
With these changes the model becomes

$$\frac{dS}{dt} = -\beta S(I + rA)$$

$$\frac{dA}{dt} = p\beta S(I + rA) - \nu A$$

$$\frac{dI}{dt} = (1-p)\beta S(I + rA) - (\alpha+\delta)I$$

$$\frac{dR}{dt} = \nu A + \alpha I$$

$$\frac{dD}{dt} = \delta I$$

Chapter 4

Data Fitting

4.1 Estimating the model parameters

There are several studies concerning the values of the various parameters of the epidemiological models discussed in the previous chapter. Here and for the remaining of this thesis we shall consider two particular values for the infection rate β, namely, $\beta = 0.52$ and $\beta = 1.12$, corresponding to time periods with or without social distancing or lockdown measures in effect. The recovery rates both for Symptomatic and Asymptomatic are $\alpha = \nu = 0.143 \simeq 1/7$ for simplicity and the death rate is $\delta = 0.056 \simeq 1/18$. Lastly, for the variables p and r we will see later on some analysis on which values are correct, but p is in $[0, 1]$ and r in $[0, 2]$.

4.2 The critical value $R_0 = 1$

We have already seen that $R_0 = 1$ is a critical value of the basic reproductive rate and that the stability of the compartmental models discussed depend on R_0. In this section we will use the value for the parameters that we indicated above and we will see how R_0 changes with r and p, for each infection rate.

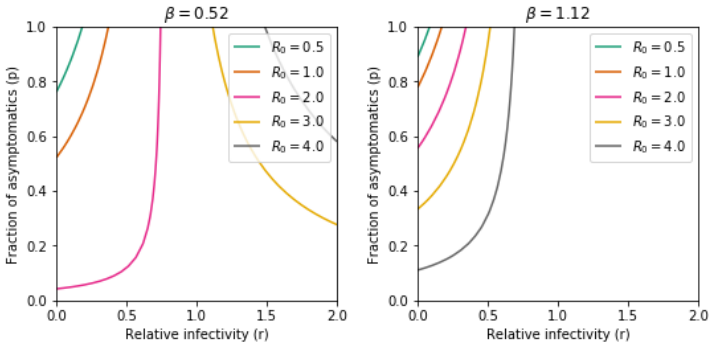


Figure 4.1: R_0 for each infection rate

As we can see from Figure 4.1 in order for $R_0 < 1$, p needs to be in $[0.5, 1]$ if $\beta = 0.52$, or in $[0.8, 1]$ if $\beta = 1.12$, and r needs to be in $[0, 0.3]$ in both cases.
We will now look at the fraction of Infectives and the peak of the infection depending on the values for r and p.

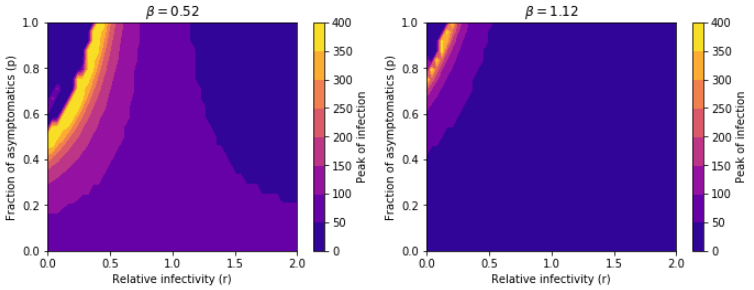


Figure 4.2: Peak of the infection

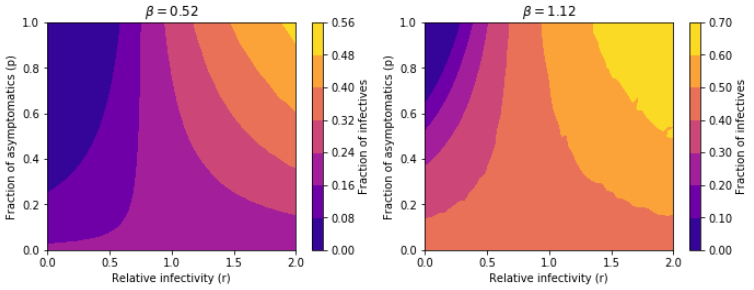


Figure 4.3: Fraction of Infectives

A careful observation of Figures 4.2 and 4.3 reveals that the positions of the contour lines match, approximately, meaning that the value of R_0 is a critical value both for the peak of the infection and for the fraction of Infectives. Therefore, every time that R_0 changes there is a drastic change in the outbreak.
Also, from the colours of the above graphs we can see that if $R_0 < 1$, both the peak of the infection and the fraction of Infectives is much smaller. This is an expected result since reducing the value for R_0 to be less than one, leads to the end of the infection.

4.3 The Nelder-Mead Method

In order to apply our model we use real data from OurWorldInData in several locations in the United States and Greece in early Spring of 2020. We fit the number of infected and the number of dead to our model by minimizing the sum of square residuals using the Nelder—Mead algorithm.
Nelder-Mead is a simplex method that begins with a randomly-generated simplex. With every iteration it proceeds to reshape it, one vertex at a time, in order to achieve its optimal shape. In order to see how this method works we will assume that we are in the n-dimensional space. This simplex consists the points $x_1, x_2, ..., x_{n+1}$, and the function we will try to minimize is $f(x)$. The algorithm will take the following steps with every iteration.

- **Step 1**
  All the points need to be order such that the value of f in the first point is highest and the value of f in the last point the lowest. We will denote the first(worst), second(second-worst) and last(best) points as x_h, x_s, x_l respectively.

- **Step 2**
  We will compute the mean of all the points except the worst x_h as $c = \frac{1}{n}\sum_{i\neq h} x_i$.

- **Step 3**
  We will begin the transformation by computing the reflected point as $x_r = c + \alpha(c - x_h)$, where α is the reflection parameter and is usually equal to 1. If $f(x_s) > f(x_r) > f(x_l)$, which means that x_r is better than the second-worst point but not better than the best point, we replace x_h with x_r in the simplex and we move to the next iteration.

- **Step 4**
  If the reflected point x_r happens to be better than the best point x_l ($f(x_r) < f(x_l)$) we will move a little bit more in the direction of x_r from c in order to see if there is an even better solution. The expanded point is $x_e = c+\gamma(x_r-c)$, where γ is called the expansion parameter and is usually 2.

- **Step 5**
  We will then replace x_h with the better of the two points: x_e and x_r in the simplex.

- **Step 6**
  If the reflection point was worst than x_s maybe the direction defined by x_r is not the one we should move. So we will need to contract our simplex using the contraction point $x_c = c + \beta(x_h - c)$, where β is the contraction parameter and is usually 0.5.If $f(x_c) < f(x_h)$ this means that the contraction point is better than the worst point and we replace x_h with x_c in the simplex.

- **Step 7**
  If however $f(x_c) > f(x_h)$ we will need to redefine the entire simplex. We will keep the best point x_l and we will define the other points using that. The j-th new point will be $x_j = x_l + \delta(x_j - x_l)$, where δ is the shrinkage parameter and is usually 0.5. What we will essentially be doing with the above definition, is moving each point in the simplex towards the current best point, in the hope of converging onto the best neighbourhood.

It is important to say that this last step is the most expensive because we have to replace multiple points in the simplex. However it has been found that this transformation rarely needs to happen in practice. Now, we will see a simple example of this method using the Python code down below. As an example we compute the minimum of the function

$$f(x, y) = (a - x)^2 + b(y - x^2)^2.$$

This is the Rosenbrock function or Rosenbrock's banana function. This function is proved to have a minimum in the point (a, a^2) and is often computed for $a = 1$ and $b = 100$.

```
#!/usr/bin/python
# -*- coding: utf-8 -*-
class Vector(object):
    def __init__(self, x, y):
        """ Create a vector, example: v = Vector(1,2) """
        self.x = x
        self.y = y
    def __repr__(self):
        return "({0}, {1})".format(self.x, self.y)
    def __add__(self, other):
        x = self.x + other.x
        y = self.y + other.y
        return Vector(x, y)
    def __sub__(self, other):
        x = self.x - other.x
        y = self.y - other.y
        return Vector(x, y)
    def __rmul__(self, other):
        x = self.x * other
        y = self.y * other
        return Vector(x, y)
    def __truediv__(self, other):
        x = self.x / other
        y = self.y / other
        return Vector(x, y)
    def c(self):
        return (self.x, self.y)
# objective function
def f(point):
    x, y = point
    a=1; b=100
    return (a-x)**2 +b*(y-x**2)**2
def nelder_mead(alpha=1, beta=0.5, gamma=2, maxiter=10):
    # initialization
    v1 = Vector(1.0, 1.0)
    v2 = Vector(1.0, 0)
    v3 = Vector(0, 1)
    for i in range(maxiter):
        adict = {v1:f(v1.c()), v2:f(v2.c()), v3:f(v3.c())}
        points = sorted(adict.items(), key=lambda x: x[1])
        b = points[0][0]
        g = points[1][0]
        w = points[2][0]
        mid = (g + b)/2
```

```
        # reflection
        xr = mid + alpha * (mid - w)
        if f(xr.c()) < f(g.c()):
            w = xr
        else:
            if f(xr.c()) < f(w.c()):
                w = xr
            c = (w + mid)/2
            if f(c.c()) < f(w.c()):
                w = c
        if f(xr.c()) < f(b.c()):
            # expansion
            xe = mid + gamma * (xr - mid)
            if f(xe.c()) < f(xr.c()):
                w = xe
            else:
                w = xr
        if f(xr.c()) > f(g.c()):
            # contraction
            xc = mid + beta * (w - mid)
            if f(xc.c()) < f(w.c()):
                w = xc
        # update points
        v1 = w
        v2 = g
        v3 = b
    return b
print("Result of Nelder-Mead algorithm: ")
xk = nelder_mead()
print("Best poits is: %s"%(xk))
```

Running the Nelder–Mead code above produces

```
Result of Nelder-Mead algorithm:
Best poits is: (1.0, 1.0)
```

Chapter 5

Comparison with real data

In this section we will compare numerically the SIR and the SAIRD model with real data from Greece, California, Florida, New York and Texas, that track the cumulative number of cases and deaths. The Python code that we used, is shown in the Appendix.
We have used different time periods in those location since the virus didn't spread at the same time everywhere. We assumed that the epidemic started with a single individual and no deaths, that is we took $I_0 = 1$. To determine the start date of the infection we included a free parameter T_s to shift the data in time. There will obviously be differences on how well the model works in each location because the real number of new cases is different from the ones that actually get announced, meaning in some places it might be closer to the real number than others.
The data that we have used come from OurWorldInData and CDC. We use the cumulative number of cases and deaths in each location and modify equation 2.1.2 and 3.3.3 slightly to reflect this fact. We also plot the total Symptomatics and Asymptomatics for the SAIRD model, that determine the peak of the infection.

5.1 The SIR model

5.1.1 Greece

For Greece, we used data in the time period February 26, 2020 to April 5, 2020. In the first graph we computed the SIR model with the initial parameter estimates being $\beta = 1.12$, $\alpha = 1.143$,$\delta = 0.056$, $T_s = 32.7$.

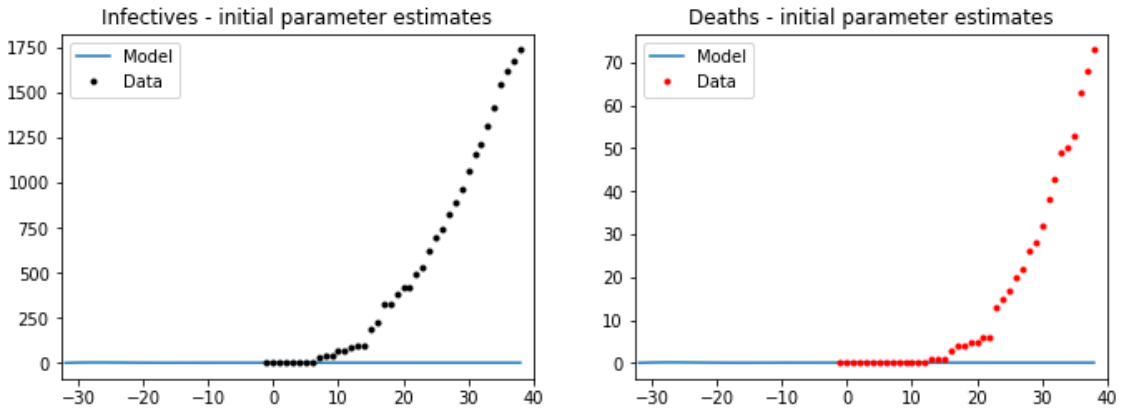


Figure 5.1: SIR initial Greece

It is obvious that the difference between the real data and the result from the simulation is quite big, meaning we need to use different parameters. In order to optimize that difference and find the right parameters we will use the Nelder-Mead algorithm. This gives us the improved parameter estimates $\beta = 1.59382275$, $\alpha = 1.43799552$ and $T_s = 10.04436537$.

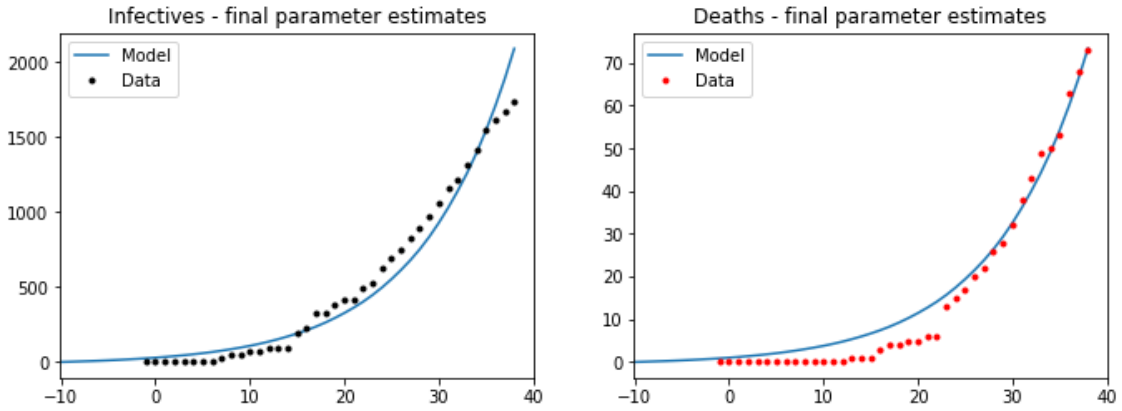


Figure 5.2: SIR final Greece

As we can see, the graph is now much closer to the real data now, and that means that we can use this model with these parameters to predict the evolution of the disease.
In the next paragraph we use data from CDC for the total number og cases and deaths in four populous states in the United States of America, specifically California, Florida, New York and Texas for a time interval of 40 days after approximately March 1, 2020.

5.1.2 United States of America

California

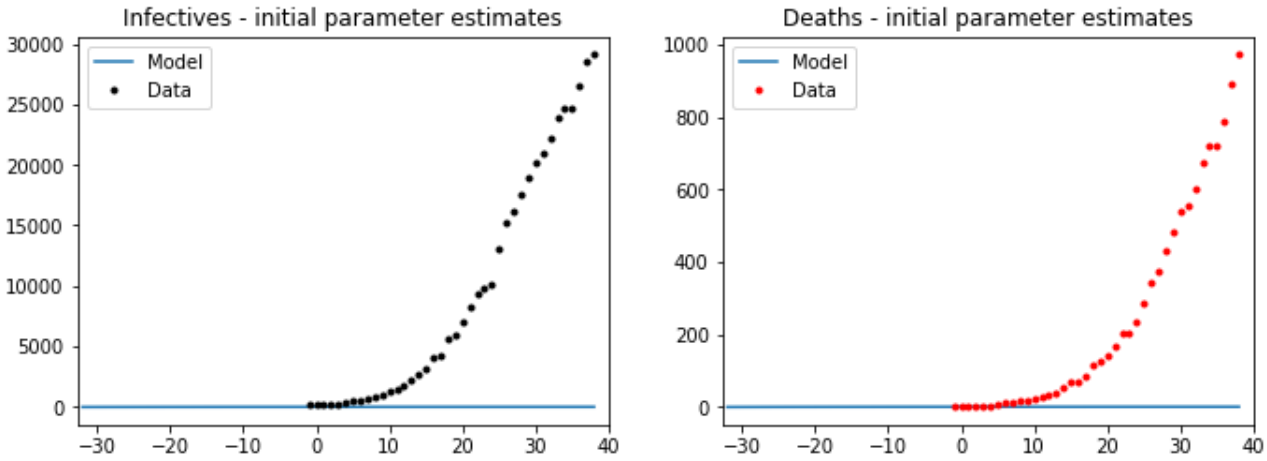


Figure 5.3: SIR initial California

Similarly with the data from Greece, we used the same original estimates and they are not the best, and after we optimize them we have $\beta = 1.99989353$, $\alpha = 1.85249568$ and $T_s = 42.08582669$.

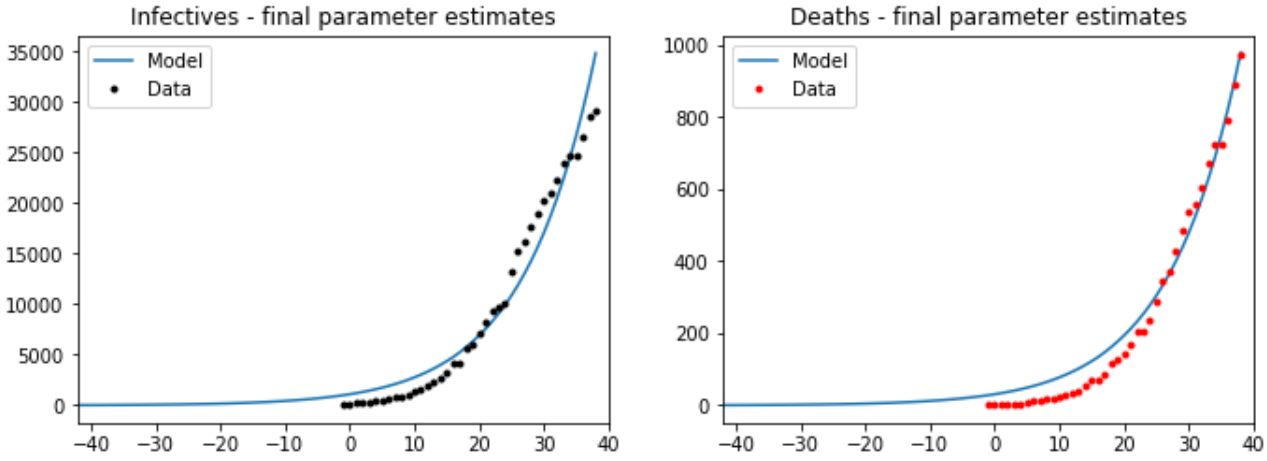


Figure 5.4: SIR final California

These results are very close to the real data, so once again we can use them to study the disease in the area.

Florida

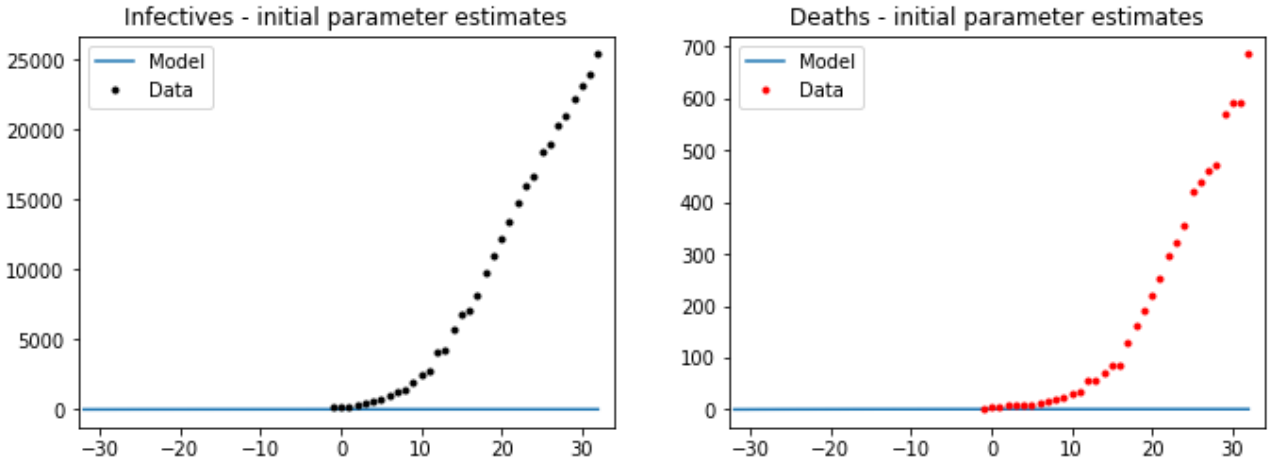


Figure 5.5: SIR initial Florida

In this data set we also used the same initial parameters, and the final ones are $\beta = 2.43894006$,$\alpha = 2.28659767$ and $T_s = 41.06620599$

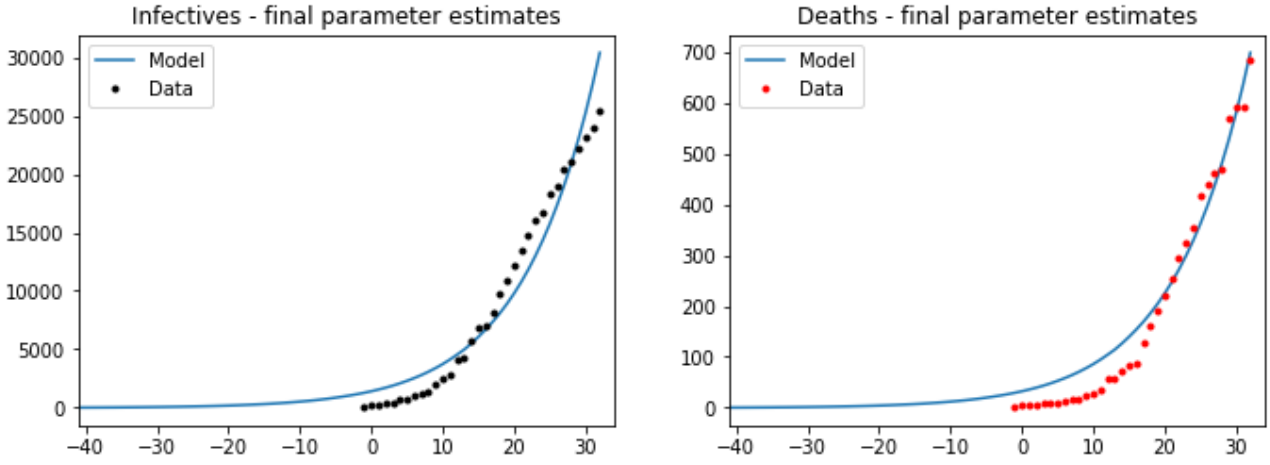


Figure 5.6: SIR final Florida

New York

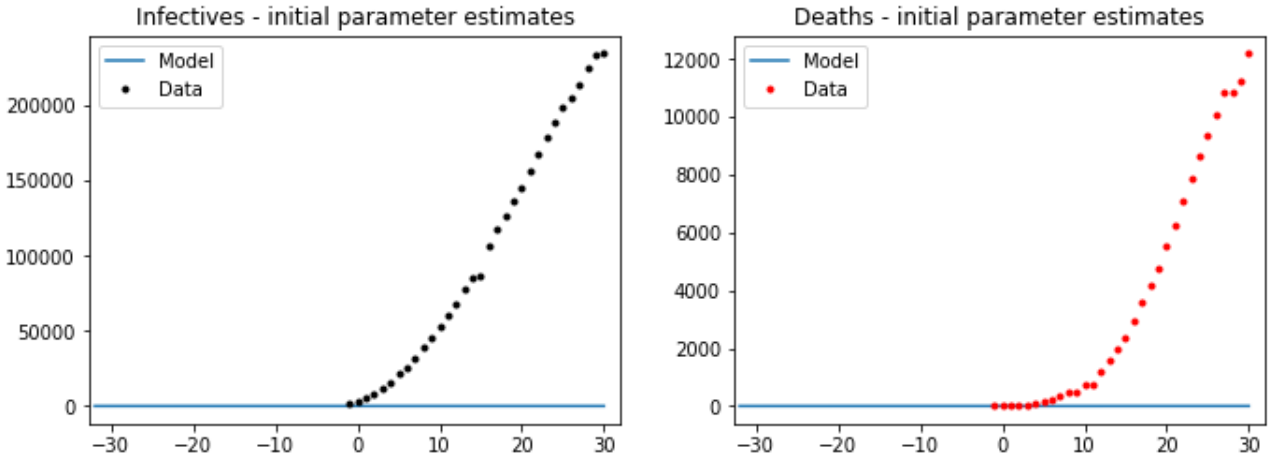


Figure 5.7: SIR initial New York

Similar, with the other data sets we used the same initial parameters and we need to optimize them. This gives us $\beta = 1.2861537$,$\alpha = 1.13089925$ and $T_s = 71.69888294$.

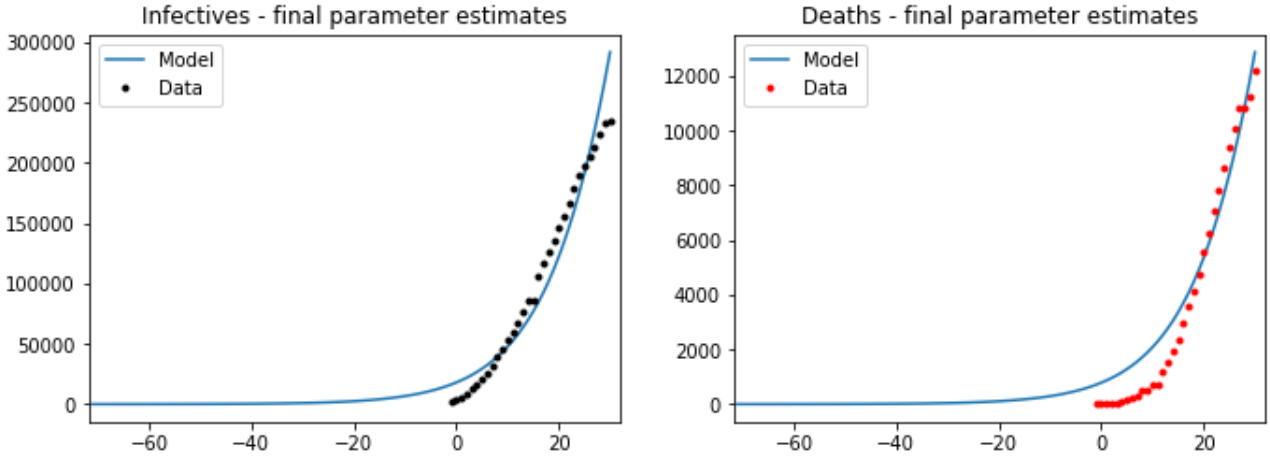


Figure 5.8: SIR final New York

Texas

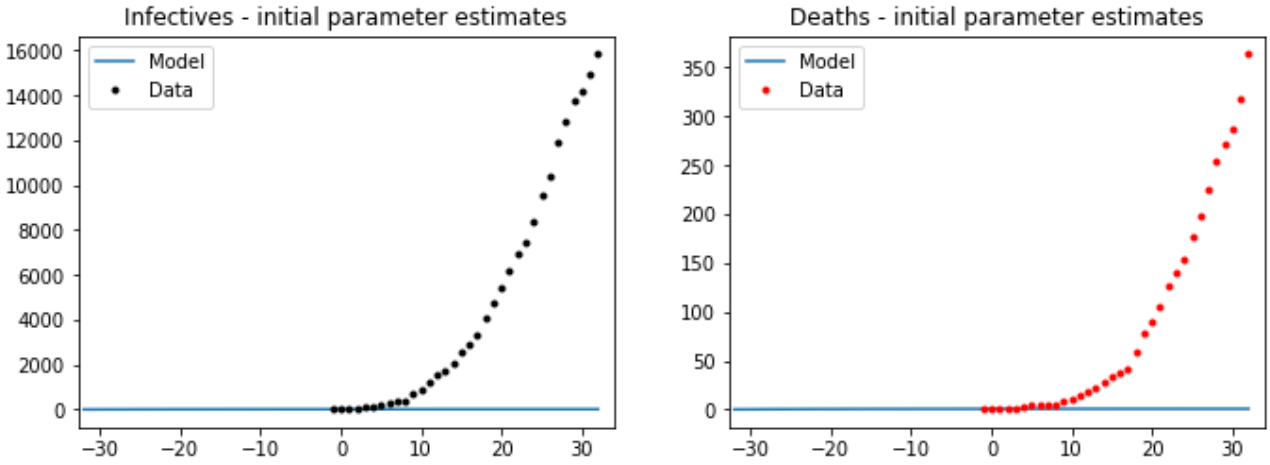


Figure 5.9: SIR initial Texas

Lastly, we will optimize the initial parameters in this data set too and we have $\beta = 2.81556726$,$\alpha = 2.64593284$ and $T_s = 25.63889835$.

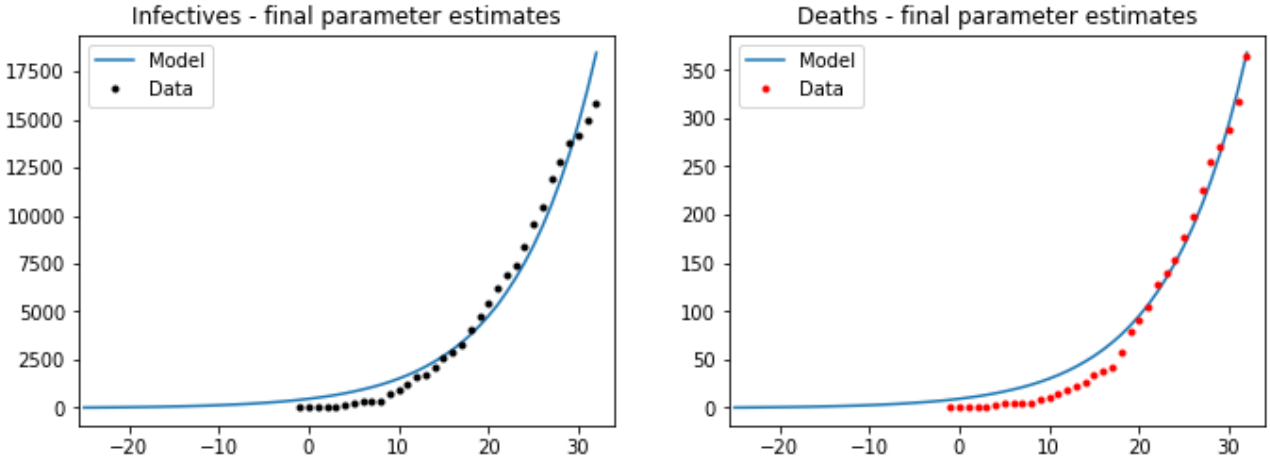


Figure 5.10: SIR final Texas

5.2 The SAIRD model

We will now work on the SAIRD model for the same data sets and see the results that we get this time. We will use some fit parameters $\delta = 0.056$ and an initial estimation of $p = 0.99$ in all data sets.

5.2.1 Greece

As we did for the SIR model we will compute the SAIRD model with data in the time period February 26, 2020 to April 5, 2020. In the first graph we computed the SIR model with the initial parameter estimates being $\beta = 2.3000$, $\alpha = 1.75$, $\nu = 0.143$, $r = 0.153$ and $T_s = 25$. Similarly,

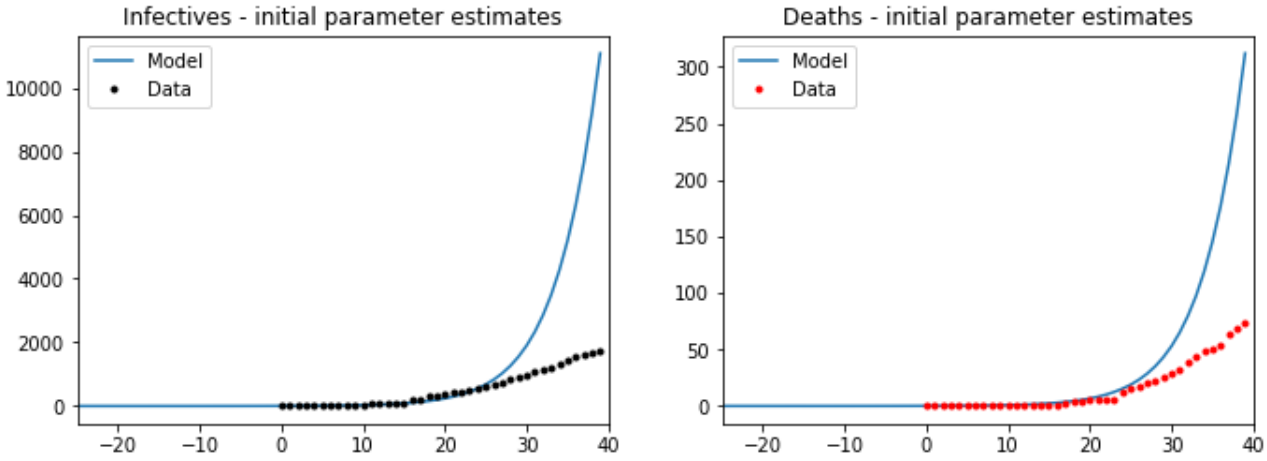


Figure 5.11: SAIRD initial Greece

the results are far different from the real data, so we need to change them. We will use the Nelder-Mead algorithm again to find the best parameters, which are $\beta = 3.3052$, $\alpha = 1.5505$, $\nu = 0.1269$, $r = 0.0721$, $p = 0.9791$ and $T_s = 48.5611$.

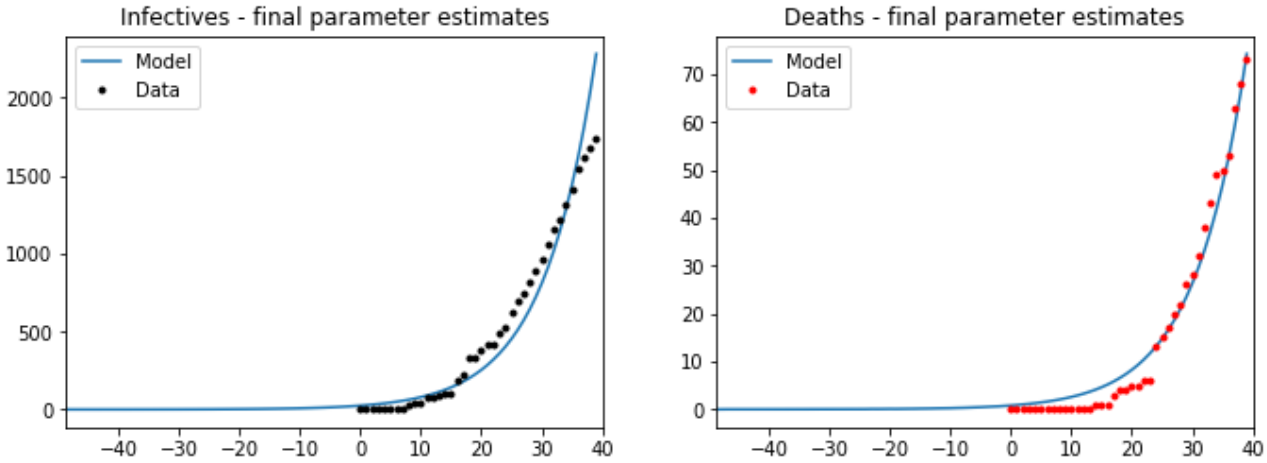


Figure 5.12: SAIRD final Greece

As we can see now the results are really close to the real data, and we can use this model to further analyze this pandemic.

5.2.2 United States of America

California

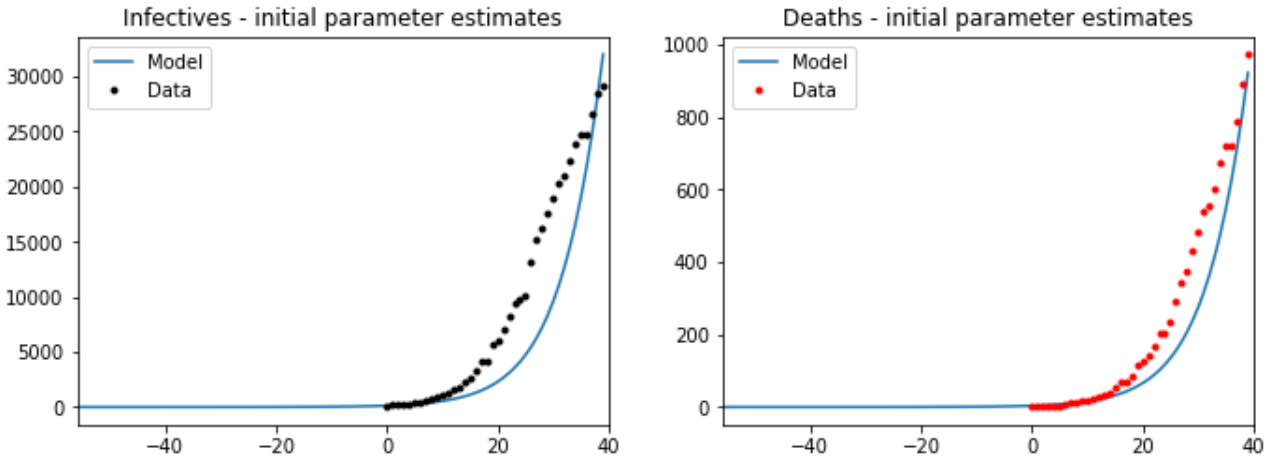


Figure 5.13: SAIRD initial California

The initial parameters are $\beta = 3.3052$, $\alpha = 1.5505$, $\nu = 0.1269$, $r = 0.0721$ and $T_s = 48.5611$. And after the optimization we have $\beta = 3.6107$, $\alpha = 1.8683$, $\nu = 0.2274$, $r = 0.1111$, $p = 0.9986$ and $T_s = 59.8449$.

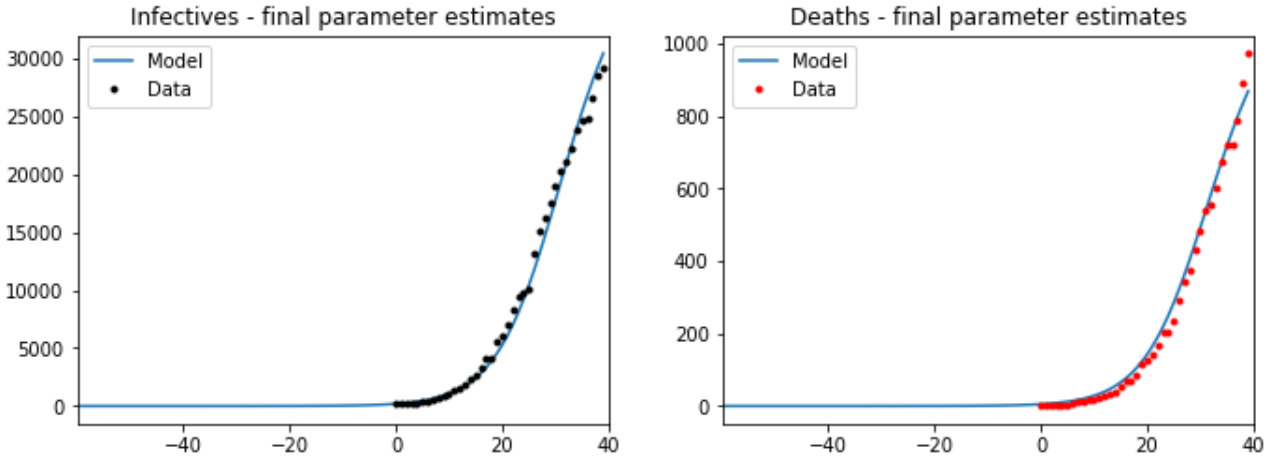


Figure 5.14: SAIRD final California

As we can see the graph and the real data are almost identical.

Florida

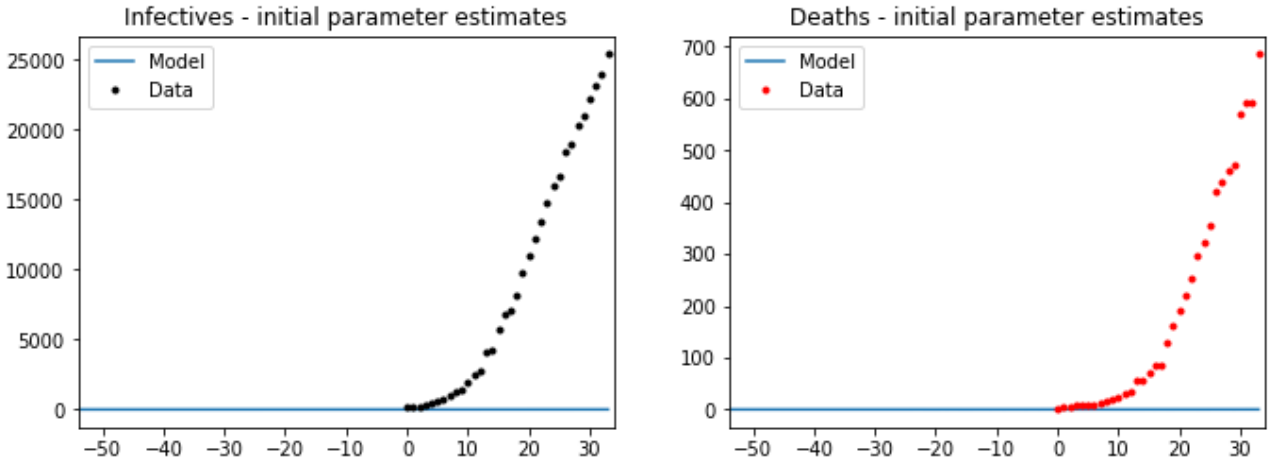


Figure 5.15: SAIRD initial Florida

The initial parameters are $\beta = 4.63$, $\alpha = 2.26$, $\nu = 0.212$, $r = 0.053$ and $T_s = 54$. And the optimization gives us $\beta = 5.0615$, $\alpha = 2.2796$, $\nu = 0.209$, $r = 0.0447$, $p = 0.8072$ and $T_s = 67.6453$.

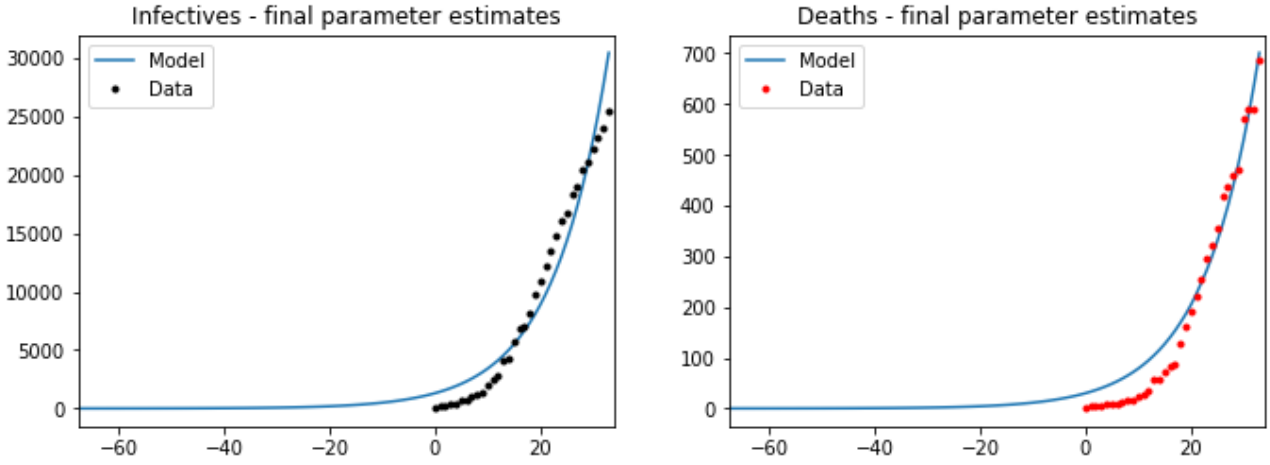


Figure 5.16: SAIRD final Florida

As we can see, the graph, after some days, is a lot closer to the real data.

New York

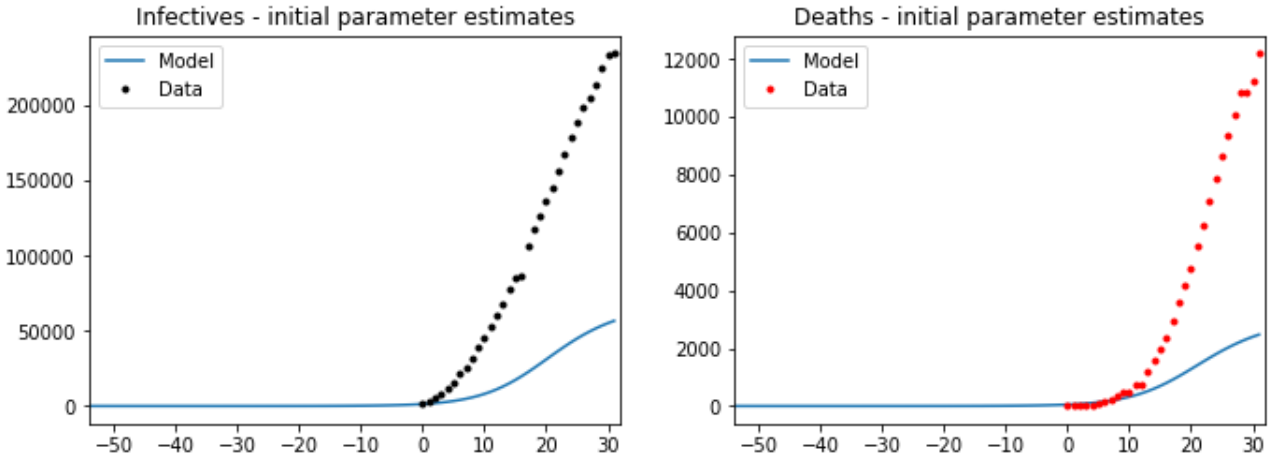


Figure 5.17: SAIRD initial New York

The initial parameters that we used are $\beta = 2.63$, $\alpha = 1.2$, $\nu = 0.212$, $r = 0.153$ and $T_s = 54$. Similarly, the optimization gives us $\beta = 2.3434$, $\alpha = 1.1356$, $\nu = 0.2412$, $r = 0.1858$, $p = 0.954$ and $T_s = 47.8019$.

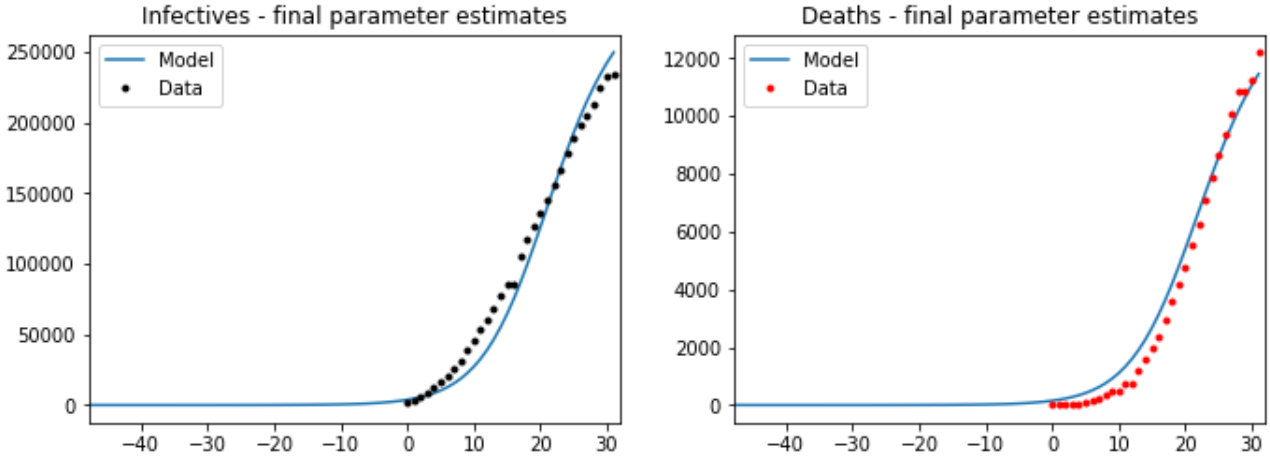


Figure 5.18: SAIRD final New York

Texas

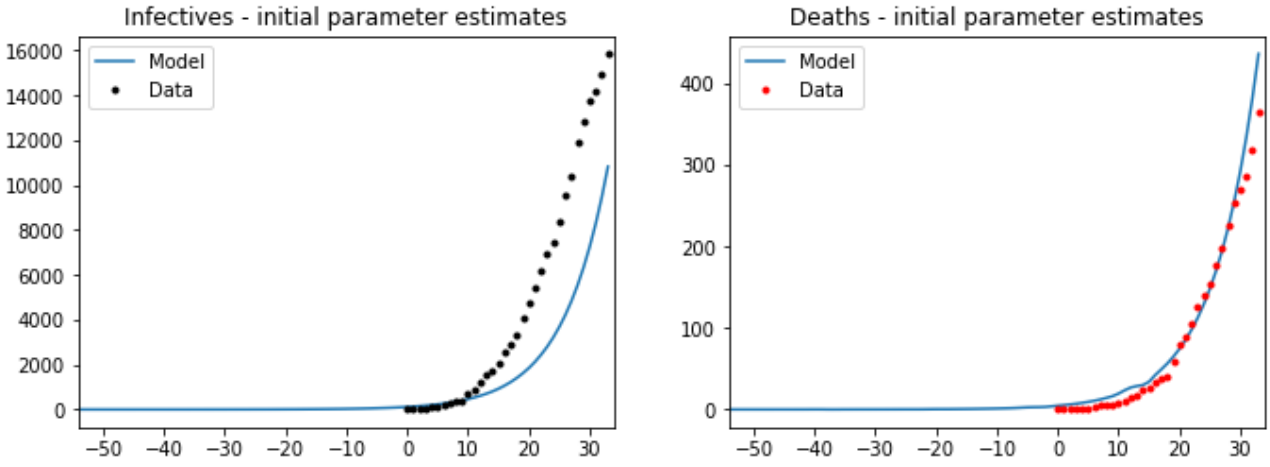


Figure 5.19: SAIRD initial Texas

And lastly, the initial parameters are $\beta = 4.63$, $\alpha = 1.2$, $\nu = 0.112$, $r = 0.053$, $p = 0.99$ and $T_s = 54.1$. After we optimize the parameters we have $\beta = 6.0734$, $\alpha = 2.6427$, $\nu = 0.1754$, $r = 0.0466$, $p = 0.9754$ and $T_s = 68.9820$.

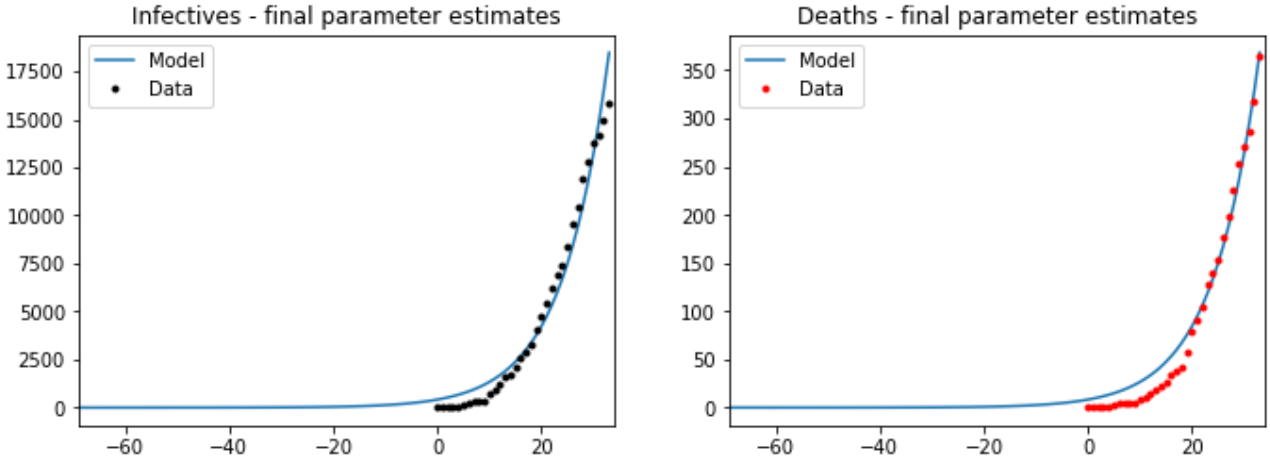


Figure 5.20: SAIRD final Texas

5.3 Symptomatics and Asymptomatics

In this section we plot Symptomatics and Asymptomatics in each location, and find when they have the greatest value.

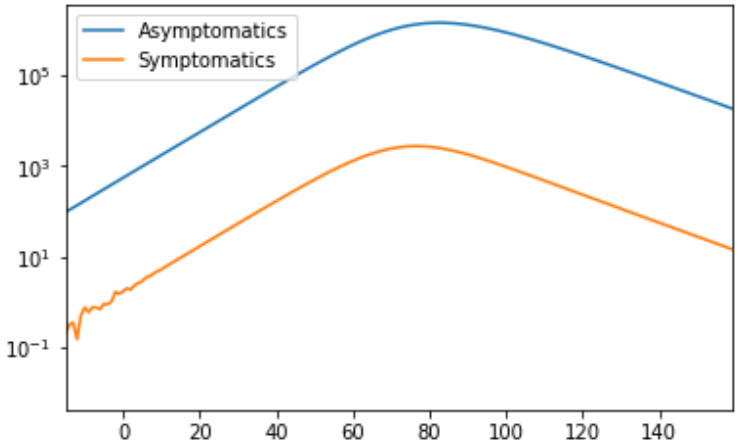


Figure 5.21: Symptomatics and Asymptomatics, Greece.

As we can see from (5.21), the peak for the Asymptomatics is about 85 days after the start of the epidemic, while the peak for Symptomatics is about 80 days.

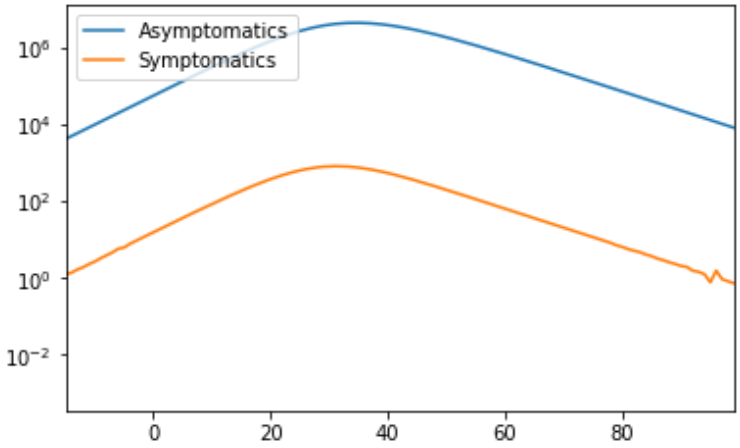


Figure 5.22: Symptomatics and Asymptomatics, California.

As we can see from (5.22), the peak for the Asymptomatics is about 40 days after the epidemic started, while for the Symptomatics is about 30 days.

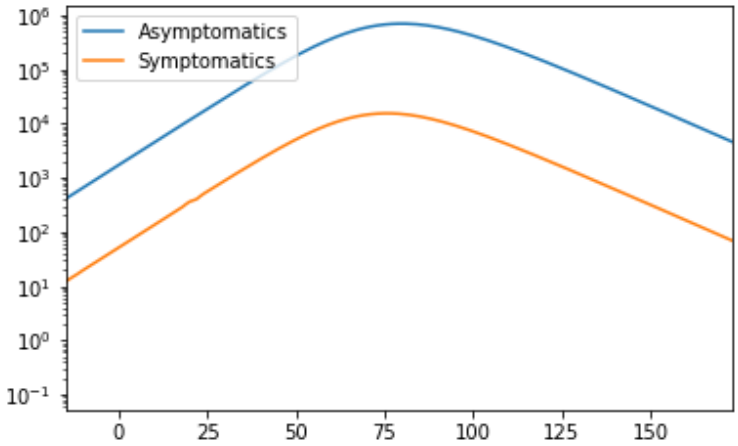


Figure 5.23: Symptomatics and Asymptomatics, Florida.

As we can see from (5.23) the peak for the Asymptomatics is about 80 days after the epidemic started, while for the Symptomatics is about 75 days.

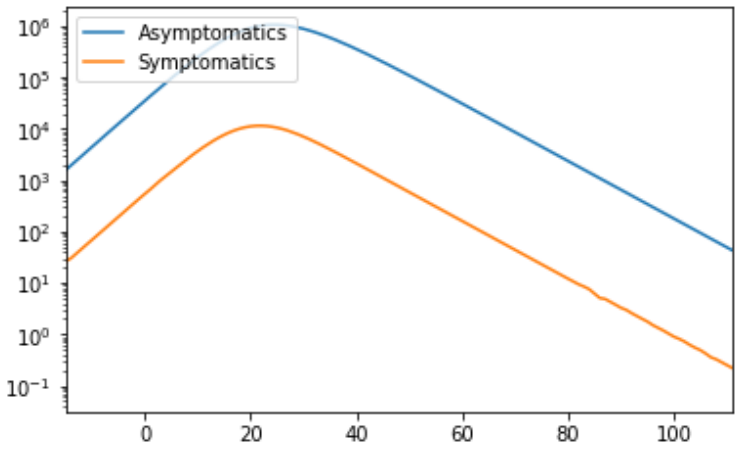


Figure 5.24: Symptomatics and Asymptomatics, New York

As we can see from (5.24) the peak for the Asymptomatics is about 25 days after the epidemic started, while for the Symptomatics is about 20 days.

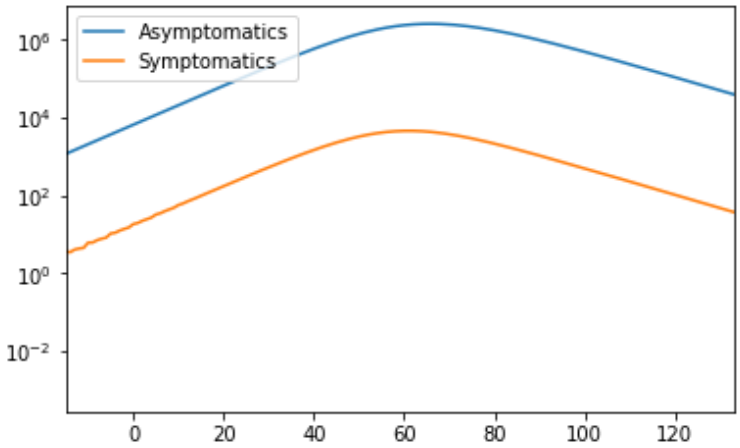


Figure 5.25: Symptomatics and Asymptomatics, Texas

As we can see from (5.25), the peak for the Asymptomatics is about 65 days after the epidemic started, while for the Symptomatics is about 60 days.

We can easily observe that in every location the peak of the Symptomatics is some days earlier than the peak of the Symptomatics. This means that they have a very similar behaviour. This can also be seen from the graph since the one is almost a parallel transport of the other. However the main difference that they have, is their values, since Asymptomatics are in order $10^2 - 10^4$, while the Symptomatics are in order $1 - 10^2$, meaning it's almost 100 times smaller than the size of the Symptomatics. This does make sense because in our code we used a higher value for p, so a bigger fraction of Susceptibles becomes Asymptomatics, rather than Infected.
However, this suggests something really important. Asymptomatics play a huge role in the development of the outbreak due to the fact that they are a lot more than the Infected.

Conclusions

In this thesis we studied some representative SIR-type models with practical emphasis on the classic SIR model and a more general compartmental model SAIRD . There is an infinite number of mathematical models one can create. But despite their differences, here are similarities in all of them.
First, if births are included this leads to a system having two equilibrium points, a disease-free equilibrium and an endemic one. In the models that we saw, the disease-free equilibrium is stable when $R_0 < 1$, while the endemic equilibrium is stable if $R_0 > 1$, with R_0 the basic reproduction number, as defined in Section 1.4. This doesn't come as a surprise since the pandemic in all model in evolving as long as $R_0 > 1$ and it's the only way to stay at a value different than zero.
We also saw that the dependence of R_0 and the evolution of the pandemic is true, since in all the plots, when $R_0 < 1$ we had the smallest amount of Infected.
From the numerical experiments we also saw that both the SIR and the SAIRD model, with the right fitting of the parameters can simulate an epidemic. This was proven from the comparison with the real data for several locations, including Greece.
Lastly, we proved that the Asymptomatics are much greater than the Infected, almost by a factor of $\simeq 10^2$. This is an important conclusion of this thesis because it suppproves the assumption that Asymptomatics play a significant role in the evolution of a disease. One of the most notable reasons why we need to study epidemiological models with more compartments and population transfer among them, is because they are more likely to show the force of the epidemic better. A simpler model such as the SIR cannot show things like this. To sum things, we can create a more complex model that includes more assumptions which will be more difficult to study analytically and numeriically, but it can help view the epidemic more globally and perhaps more accurately.

Appendix A

Code for the SIR model

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from scipy.integrate import solve_ivp
from math import ceil
from scipy.optimize import fmin


# ---------------------------------------------------------------
# SIR model equations for Y = (S, I, R) and initial condition

def sir(t, Y, beta, alpha, delta):
    dS = -beta*Y[0]*Y[1]
    dI =  beta*Y[0]*Y[1] - (alpha + delta)*Y[1]
    dR =                    alpha         *Y[1]
    dC = beta*Y[0]*Y[1]
    return [dS, dI, dR, dC]

def sirIC(i0, N):
    return [1-i0/N, i0/N, 0, i0/N]

# ---------------------------------------------------------------
# Compute the least squares error

def fitscore(c, d, Tf, infd, dead):
    c  = c[-Tf-1:]
    wc = 1 / (np.mean(c)**2 + np.mean(infd)**2)
    d  = d[-Tf-1:]
    wd = 1 / (np.mean(d)**2 + np.mean(dead)**2)
    return wc*np.sum((c - infd)**2) + wd*np.sum((d - dead)**2)

# ---------------------------------------------------------------
# Objective function for the minimization

def cost(x, delta, Tf, N, infd, dead):
    beta, alpha, Ts = x

    days = [i for i in range(ceil(-Ts), Tf)]
```

```
39
40      Y0 = sirIC(1, N)
41      sol = solve_ivp(sir, (-Ts, Tf), Y0, method='LSODA', t_eval=days,\
42                      args=(beta, alpha, delta))
43
44      c =  N*sol.y[3]
45      d = N*(1 - (sol.y[0] + sol.y[1] + sol.y[2]))
46
47      return fitscore(c, d, Tf, infd, dead)
48
49  # ------------------------------------------------------------------
50  # Read total cases and total deaths
51
52  data = pd.read_csv('ds1.csv', usecols=['total_cases','total_deaths'])
53  #data = pd.read_csv('ds2.csv', usecols=['total_cases','total_deaths'])
54  #data = pd.read_csv('ds3.csv', usecols=['total_cases','total_deaths'])
55  #data = pd.read_csv('ds4.csv', usecols=['total_cases','total_deaths'])
56  #data = pd.read_csv('greece-data-small.csv', usecols=['total_cases', '
        total_deaths'])
57  infd = data['total_cases'].to_numpy()
58  dead = data['total_deaths'].to_numpy()
59  dead = np.nan_to_num(dead,nan=0)
60
61  dasz = min(infd.size, dead.size)
62
63  # ------------------------------------------------------------------
64  # Initial estimates of the model parameters. The size of the population, N,
        and the
65  # mortality rate, delta, will be kept fixed.
66
67  beta = 1.12
68  alpha = 1.143
69  Ts = 32.7
70
71  delta = 0.056
72  N = 11000000
73
74  # ------------------------------------------------------------------
75  # Integrate model on [-Ts, Tf]. Ts will be estimated to fix the start of the
        epidemic
76
77  Tf = dasz-1
78  days = [i for i in range(ceil(-Ts), Tf)]
79
80  Y0 = sirIC(1, N)
81  sol = solve_ivp(sir, (-Ts, Tf), Y0, method='LSODA', t_eval=days,\
82                  args=(beta, alpha, delta))
83
84  c = N*sol.y[3]
85  d = N*(1 - (sol.y[0] + sol.y[1] + sol.y[2]))
86
87  # ------------------------------------------------------------------
88  # Plot total number of infectives and number of deaths
89
```

```
fig, (ax1, ax2) = plt.subplots(nrows=1, ncols=2, figsize=(10, 4))
fig.tight_layout(pad=3.0)

ax1.plot(days, N*sol.y[3], days[-Tf-1:], infd, 'ko', markersize=3)
ax1.set_xlim(-Ts, Tf+1)
ax1.set_title('Infectives - initial parameter estimates')
ax1.legend(labels=['Model', 'Data'], loc='upper left')

ax2.plot(days, d, days[-Tf-1:], dead, 'ro', markersize=3)
ax2.set_xlim(-Ts, Tf+1)
ax2.set_title('Deaths - initial parameter estimates')
ax2.legend(labels=['Model', 'Data'], loc='upper left')
plt.show()
plt.show()

# ------------------------------------------------------------------
# Optimize the parameters of the model

x0 = (beta, alpha, Ts)
xnew = fmin(cost, x0, args=(delta, Tf, N, infd, dead))
beta, alpha, Ts = xnew

# ------------------------------------------------------------------
# Integrate model on [-Ts, Tf]. Ts will be estimated to fix the start of the
    epidemic

days = [i for i in range(ceil(-Ts), Tf)]

Y0 = sirIC(1, N)
sol = solve_ivp(sir, (-Ts, Tf), Y0, method='LSODA', t_eval=days,\
                args=(beta, alpha, delta))

c= N*sol.y[3]
d = N*(1 - (sol.y[0] + sol.y[1] + sol.y[2]))

# ------------------------------------------------------------------
# Plot total number of infectives and number of deaths

fig, (ax1, ax2) = plt.subplots(nrows=1, ncols=2, figsize=(10, 4))
fig.tight_layout(pad=3.0)

ax1.plot(days, N*sol.y[3], days[-Tf-1:], infd, 'ko', markersize=3)
ax1.set_xlim(-Ts, Tf+1)
ax1.set_title('Infectives - final parameter estimates')
ax1.legend(labels=['Model', 'Data'], loc='upper left')

ax2.plot(days, d, days[-Tf-1:], dead, 'ro', markersize=3)
ax2.set_xlim(-Ts, Tf+1)
ax2.set_title('Deaths - final parameter estimates')
ax2.legend(labels=['Model', 'Data'], loc='upper left')
plt.show()
plt.show()
```

Appendix B

Code for the SAIRD model

```
from math import ceil
import numpy as np
from scipy.integrate import solve_ivp

# The model equations

def model(t, Y, beta, alpha, nu, delta, r, p):
    S, A, I, R, C = Y

    z = beta*S*(I + r*A)
    q = 1 - p

    dS =  -z
    dA = p*z - nu*A
    dI = q*z -       (alpha + delta)*I
    dR =       nu*A + alpha         *I
    dC = q*z

    return [dS, dA, dI, dR, dC]

def modelIC(i0, N):
    return [1-i0/N, 0, i0/N, 0, i0/N]

# Objective function for the minimization

def cost(x, delta, Tf, N, infd, dead):
    beta, alpha, nu, r, p, Ts = x

    if ceil(-Ts) != -Ts:
        days = [-Ts] + [i for i in range(int(ceil(-Ts)), Tf+1)]
    else:
        days = [i for i in range(int(ceil(-Ts)), Tf+1)]

    Y0 = modelIC(1, N)
    sol = solve_ivp(model, (-Ts, Tf), Y0, method='DOP853', t_eval=days,\
                    args=(beta, alpha, nu, delta, r, p), rtol=1.0e-6, atol=1.0
   e-9)

```

```
    c = N*sol.y[4]
    d = np.maximum(N*(1 - (sol.y[0] + sol.y[1] + sol.y[2] + sol.y[3])), 0)

    return fitscore(c, d, Tf, infd, dead)

# Compute the least squares error

def fitscore(c, d, Tf, infd, dead):
    c  = c[-Tf-1:]
    wc = 1 / (np.mean(c)**2 + np.mean(infd)**2)
    d  = d[-Tf-1:]
    wd = 1 / (np.mean(d)**2 + np.mean(dead)**2)
    return wc*np.sum((c - infd)**2) + wd*np.sum((d - dead)**2)

# Read covid data

def readData(fname):
    import pandas as pd
    data = pd.read_csv(fname, usecols=['total_cases', 'total_deaths'])
    infd = data['total_cases'].to_numpy()
    dead = data['total_deaths'].to_numpy()
    dead = np.nan_to_num(dead,nan=0)
    return infd, dead

# Fit infected and dead to the model

if __name__ == '__main__':

    dataset = 'greece-data-small.csv'
#Total population
    N = 11000000
#Estimates of the model parameters
    beta = 2.3
    alpha = 1.75
    nu = 0.143
    r = 0.153
    Ts = 25
    '''
    dataset = 'ds1.csv' #California data-set
#Total population
    N = 40000000
#Estimates of the model parameters
    beta = 3.05
    alpha = 1.76
    nu = 0.198
    r = 0.112
    Ts = 55.9
    '''
    '''
    dataset = 'ds2.csv' #Florida data-set
#Total population
    N = 22000000
#Estimates of the model parameters
    beta = 4.63
```

```
    alpha = 2.26
    nu = 0.212
    r = 0.053
    Ts = 54
    '''
    '''
    dataset = 'ds3.csv' #New-York data-set
#Total population
    N = 9000000
#Estimates of the model parameters
    beta = 2.63
    alpha = 1.2
    nu = 0.212
    r = 0.153
    Ts = 54
    '''
    '''
    dataset = 'ds4.csv' #Texas data-set
#Total population
    N = 30000000
#Estimates of the model parameters
    beta = 4.63
    alpha = 1.2
    nu = 0.112
    r = 0.053
    Ts = 54.1
    '''
    infd, dead = readData(dataset)
    Tf = infd.size-1

    delta = 0.056
    p = 0.99

    print('Dataset: {:}'.format(dataset))
    print('Initial parameter estimates:')
    print(r ' = {:.4f}    = {:.4f}    = {:.4f}'.format(beta, alpha, nu))
    print(r ' = {:.4f}  r = {:.4f}  p = {:.4f}  Ts = {:.4f}'.format(delta, r,
   p, Ts))

    R0 = p*r*beta/nu + (1-p)*beta/(alpha+delta)
    print(r'R_0 = {:.3f}'.format(R0), end = '\n\n')

#   Run the model

    if ceil(-Ts) != -Ts:
        days = [-Ts] + [i for i in range(int(ceil(-Ts)), Tf+1)]
    else:
        days = [i for i in range(int(ceil(-Ts)), Tf+1)]

    Y0 = modelIC(1, N)
    sol = solve_ivp(model, (-Ts, Tf), Y0, method='DOP853', t_eval=days,\
                    args=(beta, alpha, nu, delta, r, p), rtol=1.0e-6, atol=1.0
   e-9)

```

```
144 #   Plot total number of infectives and number of deaths
145
146     import matplotlib.pyplot as plt
147
148     fig, (ax1, ax2) = plt.subplots(nrows=1, ncols=2, figsize=(10, 4))
149     fig.tight_layout(pad=3.0)
150
151     ax1.plot(days, N*sol.y[4], days[-Tf-1:], infd, 'ko', markersize=3)
152     ax1.set_xlim(-Ts, Tf+1)
153     ax1.set_title('Infectives - initial parameter estimates')
154     ax1.legend(labels=['Model', 'Data'], loc='upper left')
155
156     d = np.maximum(N*(1 - (sol.y[0] + sol.y[1] + sol.y[2] + sol.y[3])), 0)
157
158     ax2.plot(days, d, days[-Tf-1:], dead, 'ro', markersize=3)
159     ax2.set_xlim(-Ts, Tf+1)
160     ax2.set_title('Deaths - initial parameter estimates')
161     ax2.legend(labels=['Model', 'Data'], loc='upper left')
162     plt.show()
163
164 #   Optimize model parameters
165
166     from scipy.optimize import fmin
167
168     x0 = (beta, alpha, nu, r, p, Ts)
169     xnew = fmin(cost, x0, args=(delta, Tf, N, infd, dead))
170
171 #   Run model with the optimized parameters
172
173     beta, alpha, nu, r, p, Ts = xnew
174
175     print('\nOptimized parameter estimates:')
176     print(r ' = {:.4f}    = {:.4f}    = {:.4f}'.format(beta, alpha, nu))
177     print(r ' = {:.4f}  r = {:.4f}  p = {:.4f}  Ts = {:.4f}'.format(delta, r,
    p, Ts))
178     R0 = p*r*beta/nu + (1-p)*beta/(alpha+delta)
179     print(r'R_0 = {:.3f}'.format(R0), end = '\n\n')
180
181     if ceil(-Ts) != -Ts:
182         days = [-Ts] + [i for i in range(int(ceil(-Ts)), Tf+1)]
183     else:
184         days = [i for i in range(int(ceil(-Ts)), Tf+1)]
185
186     Y0 = modelIC(1, N)
187     sol = solve_ivp(model, (-Ts, Tf), Y0, method='DOP853', t_eval=days,\
188                     args=(beta, alpha, nu, delta, r, p), rtol=1.0e-6, atol=1.0
    e-9)
189
190 #   Plot total number of infectives and number of deaths
191
192     fig, (ax1, ax2) = plt.subplots(nrows=1, ncols=2, figsize=(10, 4))
193     fig.tight_layout(pad=3.0)
194
195     ax1.plot(days, N*sol.y[4], days[-Tf-1:], infd, 'ko', markersize=3)
```

```
    ax1.set_xlim(-Ts, Tf+1)
    ax1.set_title('Infectives - final parameter estimates')
    ax1.legend(labels=['Model', 'Data'], loc='upper left')

    d = np.maximum(N*(1 - (sol.y[0] + sol.y[1] + sol.y[2] + sol.y[3])), 0)

    ax2.plot(days, d, days[-Tf-1:], dead, 'ro', markersize=3)
    ax2.set_xlim(-Ts, Tf+1)
    ax2.set_title('Deaths - final parameter estimates')
    ax2.legend(labels=['Model', 'Data'], loc='upper left')
    plt.show()

#   Plot symptomatics and asymptomatics

    Tf = Tf + 165

    if ceil(-Ts) != -Ts:
        days = [-Ts] + [i for i in range(int(ceil(-Ts)), Tf+1)]
    else:
        days = [i for i in range(int(ceil(-Ts)), Tf+1)]

    Y0 = modelIC(1, N)
    sol = solve_ivp(model, (-Ts, Tf), Y0, method='DOP853', t_eval=days,\
                    args=(beta, alpha, nu, delta, r, p), rtol=1.0e-6, atol=1.0
    e-9)

    fig, ax1 = plt.subplots(figsize=(6, 4))
    fig.tight_layout(pad=3.0)

    ax1.plot(days, N*sol.y[1], days, N*sol.y[2])
    ax1.set_yscale('log')
    ax1.set_xlim(-15, Tf)
    ax1.legend(labels=['Asymptomatics', 'Symptomatics'], loc='upper left')
    plt.show()
```

Bibliography

[1] Michael Plexousakis. *Course notes for the class Mathematical Biology.* University of Crete, Spring Semester 2021

[2] James D. Murray. *Mathematical Biology I:An Introduction.* Springer, 2002

[3] Tailei Zhang, Ruini Kang, Kai Wang, Junli Liu. *Global dynamics of an SEIR epidemic model with discontinuous treatment,* 2020

[4] Yan Deng, Wei Liu, Kui Liu, Yuan-Yuan Fang, Jin Shang, Ling Zhou, Ke Wang, Fan Leng, Shuang Wei, Lei Chen, Hui-Guo Liu. *Clinical characteristics of fatal and recovered cases of coronavirus disease 2019 in Wuhan, China: a retrospective study,* 2020

[5] P. van den Driessche, James Watmough. *Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission*, 2002

[6] J. David Logan. *n A first course in Differential Equations. Undergraduate Text in Mathematics*, Springer 2015

[7] Xq. Liu, S. Xue, Jb. Xu. *Clinical characteristics and related risk factors of disease severity in 101 COVID-19 patients hospitalized in Wuhan, China*, 2021

[8] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, J. Shaman. *Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2)*, 2020

[9] Sachin Joglekar. *Nelder-Mead Optimization,* https://codesachin.wordpress.com/2016/01/16/nelder-mead-optimization/

[10] *Epidemiology,* https://en.wikipedia.org/wiki/Epidemiology

[11] *Lesson 1: Introduction to Epidemiology* https://www.cdc.gov/csels/dsepd/ss1978/lesson1/section2.html

[12] Wikipedia, *Covid-19 Pandemic,* https://en.wikipedia.org/wiki/COVID-19_pandemic