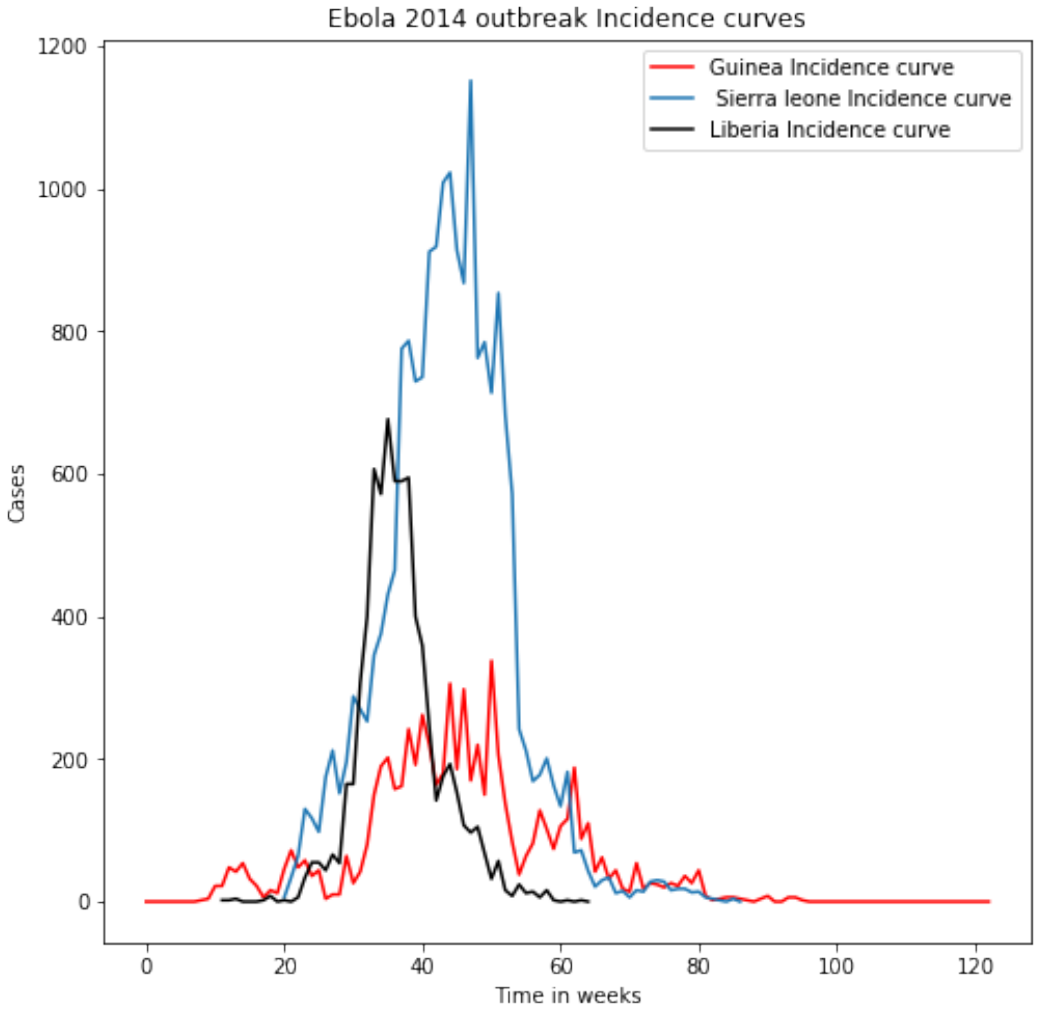
Ebola project Report

Estimating Epidemics models parameters



1 Introduction

The 2014 outbreak of Ebola virus in West Africa was the "largest, most severe and most complex Ebola epidemic" in history, according to the World Health Organization (WHO). More than 28,000 people were infected, and 11,000 people died before the international public health emergency ended in 2016. Most of the cases occurred in three countries: Guinea, Sierra Leone, and Liberia, In order to measure the severity of an outbreak a number called Basic Reproductive number is defined to determine if an epidemic breaks if it's greater than 1 or it dies out if it's less than 1. And our Role in this project will be using the data rendered from the (WHO) website to estimate this number by fitting the data.

2 The problem

The problem of this project:
is to figure out The epidemiological model best fitting the data and estimating the parameters of each model and the basic reproductive number R_0
Proposed Solution:
Many publications suggested using an SEIR model but since they're quite similar we are using both SIR and SEIR models to fit the data corresponding to the following system of Equations:

$$\begin{aligned} \frac{dS}{dt} &= -\beta(t)\frac{SI}{N} \\ \frac{dE}{dt} &= \beta(t)\frac{SI}{N} - \sigma E \\ \frac{dI}{dt} &= \sigma E - \mu I \\ \frac{dR}{dt} &= \mu I \end{aligned} \tag{1}$$

$$\begin{aligned} \frac{dS}{dt} &= -\beta(t)\frac{SI}{N} \\ \frac{dI}{dt} &= \beta(t)\frac{SI}{N} - \mu I \\ \frac{dR}{dt} &= \mu I \end{aligned} \tag{2}$$

Assumptions:
Since these kinds of analyses is conducted numerously we will be following in part the analysis in the research paper by Christian L. Althaus [1] and we will some of the parameters he used in his paper as default parameters, such as

- μ which is $\frac{1}{5.61 days}$ which is the reciprocal of the duration of the infection thus it's customary to use it from literature
- also σ in the SEIR model which is the transition rate from latency to infection which is specified by physicians rather than epidemiologists, that's why we're also assuming it to be $\sigma = \frac{1}{5.3 days}$

[1] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4169395/

- and based on that we are fitting for the β parameter and it's decaying parameter in the case of β being a function of t not a constant following this function: $\beta(t) = e^{-kt}$, we will be considering both cases ,constant and a decaying function of t, to fit the data.
- The initial number of susceptibles $S(0)$ is the entire population of each country. Should be noted that this assumption limits the model strongly if beta is considered constant, whilst introducing an exponential decay in β mitigates the dependence on $S(0)$ of the transition rates[2].

3 Methods

The fitting procedure we're following in our analysis is the Maximum Likelihood estimation, in other words we will be trying different parameters in our models and infer which of these parameters will maximize the likelihood function of our data.
We assume that each real case in the epidemic will be observed with equal probability d, and therefore the epidemic data approximately follows a Poisson distribution of parameter $\lambda(t) = d \cdot \Delta I_{model}(t)$ [3]. The likelihood and log-likelihood functions for a set of data are:

$$\begin{aligned} L(\lambda_1, ...\lambda_n; x_1, ...x_n) &= \prod_{i=1}^{n} \frac{e^{-\lambda_i} \lambda_i^{x_i}}{x_i!} \\ l(\lambda_1, ...\lambda_n; x_1, ...x_n) &= \sum_{i=1}^{n} [-\lambda_i - log(x_i!) + x_i log(\lambda_i)] \end{aligned} \tag{3}$$

4 Parameters estimation porcedure

To have a rough idea of the parameter range we used linear regression on the initial stage of the epidemic (invasion), fitting the following equation: $I(t) = I_0 e^{(R_0 - 1)(\gamma + \mu)}$ and considering γ the death rate to be much less than μ, R_0 can be roughly estimated as $\frac{slope}{\mu} + 1$. This allow us to get an idea of the availability of an epidemic but it cannot be reliable estimate of parameters.

The next step of our analysis is to solve the differential equations of each model given parameters β, μ and σ and the assumed initial values of each compartment in the model,S(0),E(0), X(0) and R(0), this model is a key part of our likelihood function as it's the input of the Poisson distribution.
This allows us to compute the likelihood of a set of parameters following equation 3 according to Each model. This is crucial in the analysis, since we're trying to estimate β (and k) via the grid search method, and trying the fit for various values of d.
Furthermore, we allowed for delay to the data in order to relax the assumption X(0) = 1, so the real epidemic could start sometime before the first cases are observed. This allowed for more flexibility in the fitting procedure.
Last but not least we distinguished between the models with β decaying and β constant in time. For a constant β we fit the models only on the increasing part of the incidence curve

[2] the parameters $S(0)$ and k in the first order approximation for $\beta(t)\frac{SI}{N}$, are multiplied together and therefore coupled.

[3] $\Delta I_{model}(t)$ is the incoming flow into compartment I, that is $\beta\frac{SI}{N}$ for SIR and σE for SEIR

"The beginning of the epidemic" not the whole data set. When trying to fit the whole data, the results are unsatisfying: either the shape is completely off, or the delay is too large, and the estimates for R_0 are too low to be reasonable [4]. This is further confirmation that the assumption of a decaying β is reasonable.

5 Resulting plots

5.1 Guinea fit with β decay model

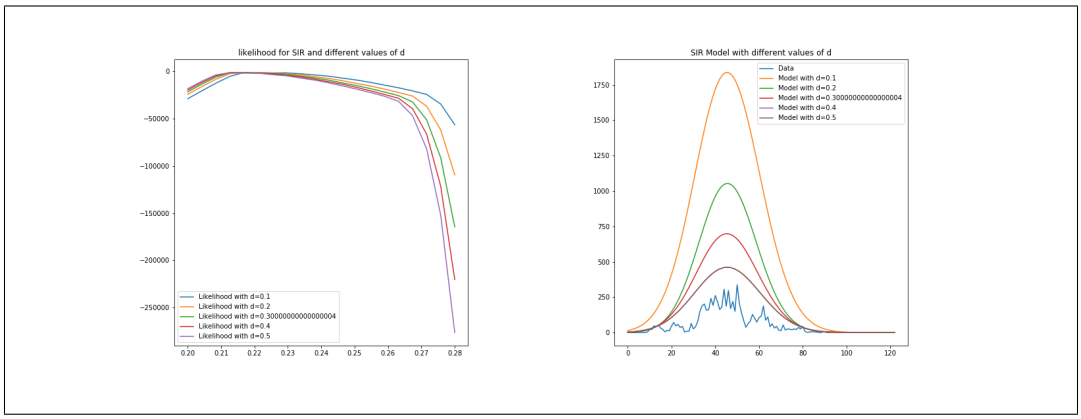


Figure 1: Likelihood function and fitting of Guinea's Data with β decay SIR model

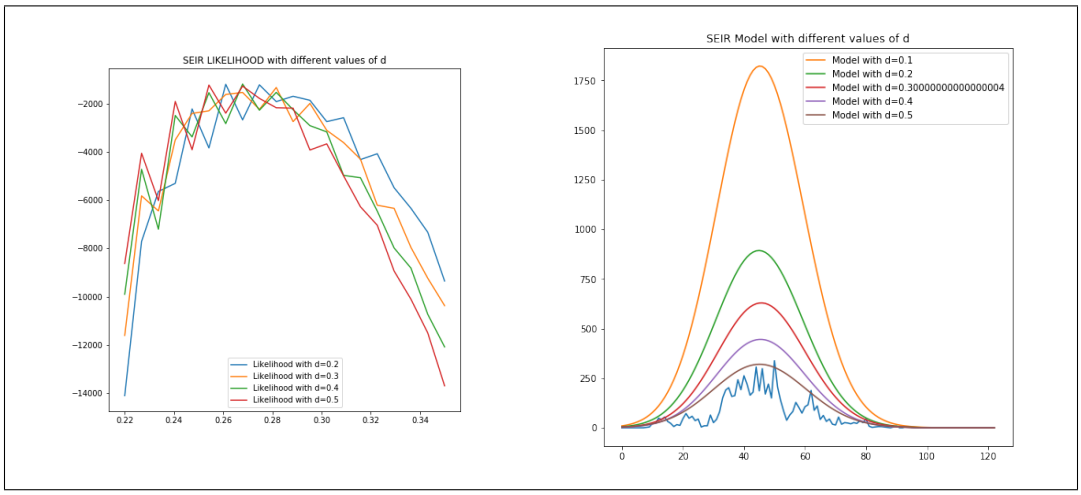


Figure 2: Likelihood function and fitting of Guinea's Data with β decay SEIR model

[4] We found the best fit to be an epidemic starting decades in advance and barely contagious

5.2 Sierra leone fit with β decay model

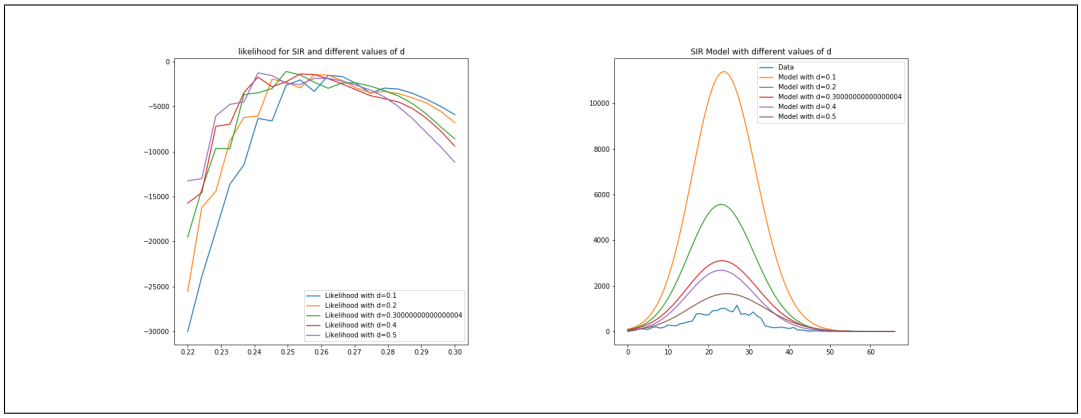


Figure 3: Likelihood function and fitting of Sierra's Data with β decay SIR model

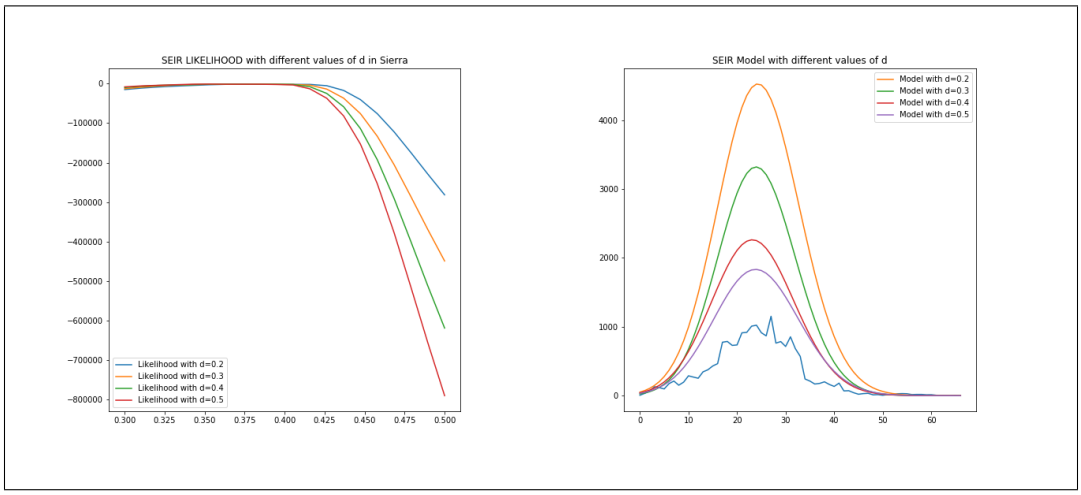


Figure 4: Likelihood function and fitting of sierra's Data with β decay SEIR model

5.3 Liberia fit with β decay model

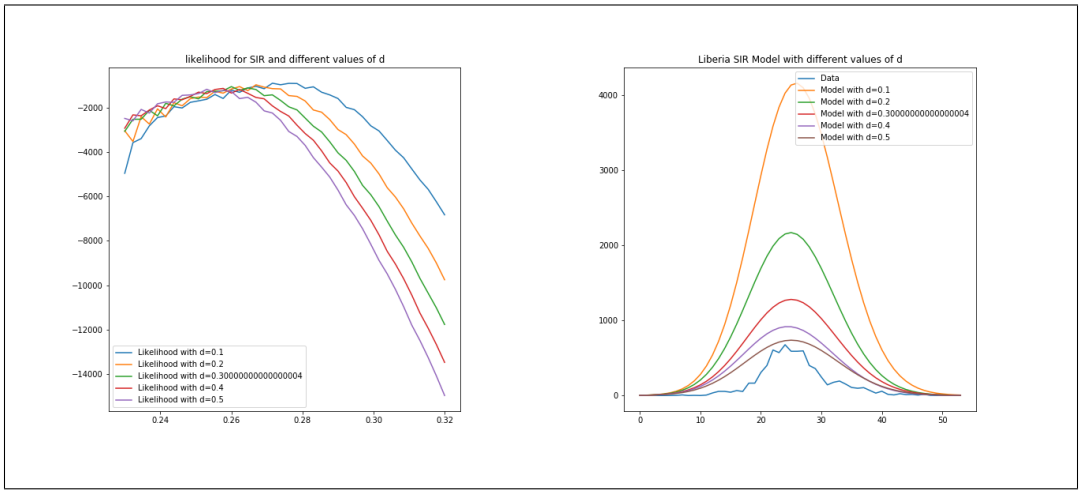


Figure 5: Likelihood function and fitting of Liberia's Data with β decay SIR model

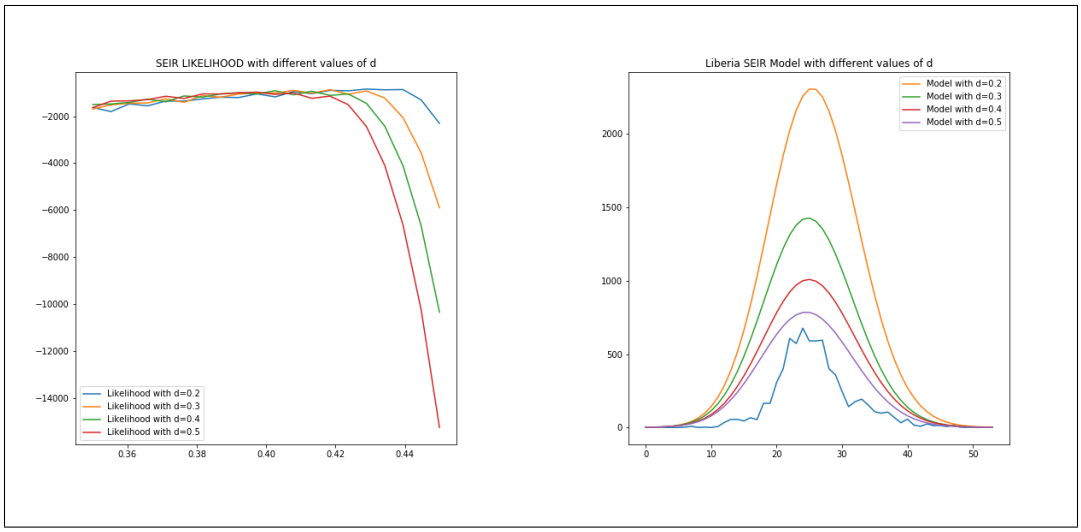


Figure 6: Likelihood function and fitting of Liberia's Data with β decay SEIR model

5.4 No β decay models on Guinea's Data

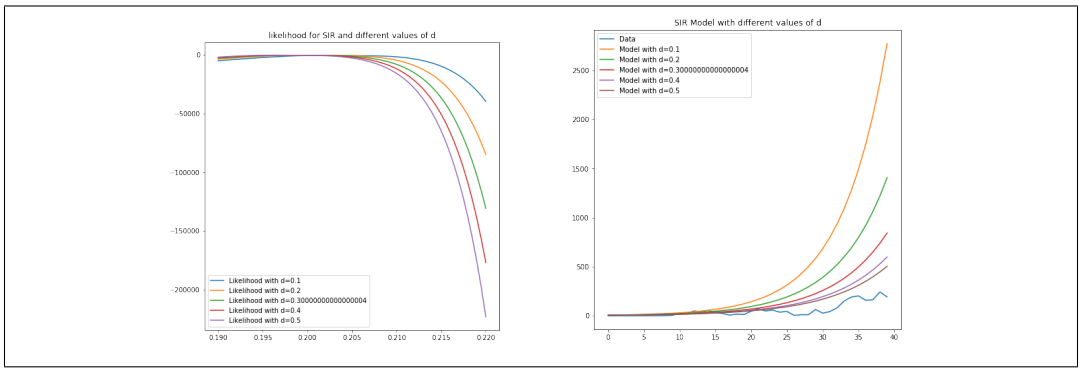


Figure 7: Likelihood function and fitting of Guinea's Data with SIR model without β decay

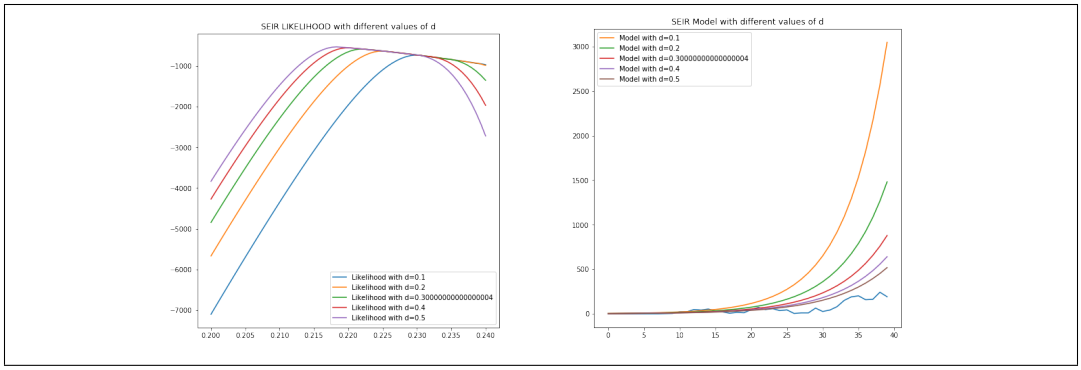


Figure 8: Likelihood function and fitting of Guinea's Data SEIR model without β decay

6 Results

6.1 SIR with β decay

Country	β	R_0	k
Guinea	0.1972	1.1	0.00057
Sierra	0.255	1.43	0.00196
Liberia	0.265	1.49	0.0.0020

6.2 SEIR with β decay

Country	β	R_0	k
Guinea	0.27	1.52	0.0011
Sierra	0.37	1.97	0.0.0029
Liberia	0.41	2.24	0.0046

7 Interpretation of Results and conclusion.

Comparing the results in Christian Althaus' paper we have a very close result on Guinea's dataset with the SEIR fitting model which is the same as the one he use as it's mentioned in his results that he found $R_0 = 1.51$ on Guinea's data and we found it to be 1.52, while our results on the two other country's are a bit off. This is most probably caused by the fact that we are fitting the models on the whole evolution of the epidemics while the original paper was published in september 2014, when the epidemic was still ongoing. Also, us not using deaths data (which gives a better image on the rate of change of infections and therefore a more accurate β estimation) could also partly explain this discrepancy. Overall, while the results on the Giunea's dataset correctly match the data, on the other country this is not the case. But, neither the parameters of the original paper do. This indicates that there is something else going on, and the SEIR model is not able to capture the whole evolution of the epidemic.

8 Further Work

- Deaths Data could make a strong contribution to the fit, as knowing the death rate could contribute to the information of rate of change of infected which would give a much better estimate of the parameters. Basically, knowing the death incidence (and therefore mortality of the disease) would provide with a second observable flow directly proportional to I_{model}. Also, since in reality many infections were caused by funeral rites, death data could make it possible to fit a more detailed model where contagions can be caused by the recently deceased.
- To relax the assumption of homogeneous mixing over the whole country, a meta-population framework could be introduced since the data provided follows the epidemic evolution separately for different parts of the counties. To make this feasible some form of mobility data will be needed since the epidemic data is too scarce to determine mobility parameters.