Final Year Project

Anthony Gibbons

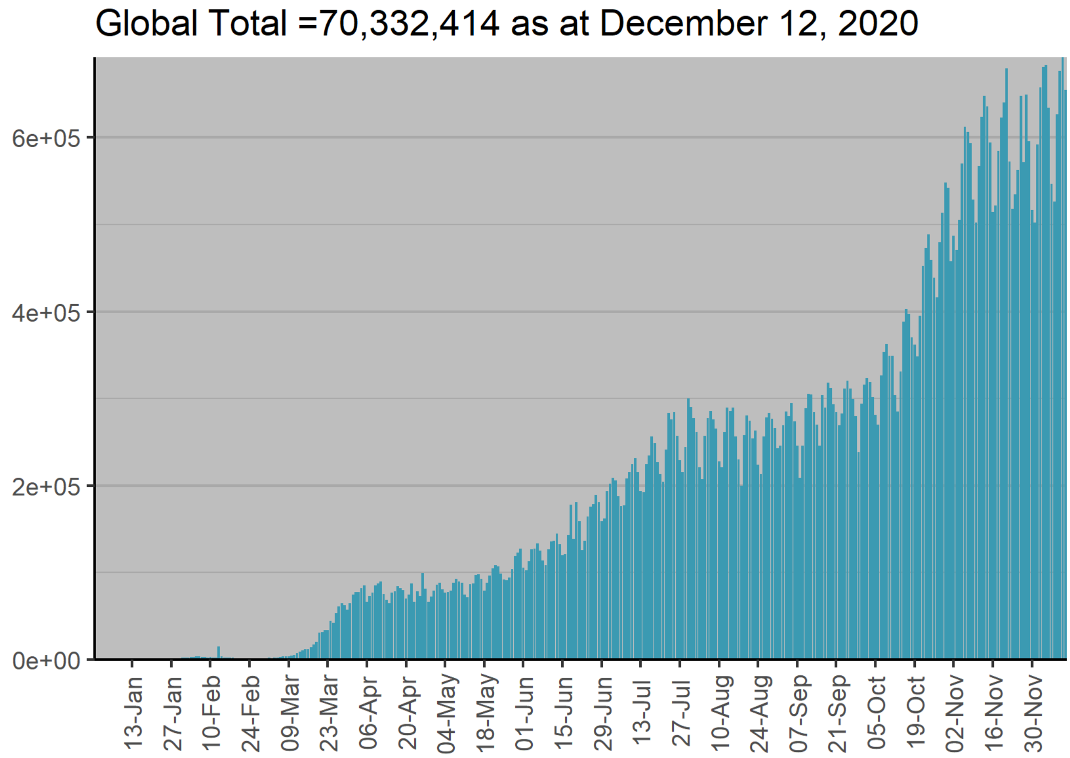
December 13, 2020

Abstract

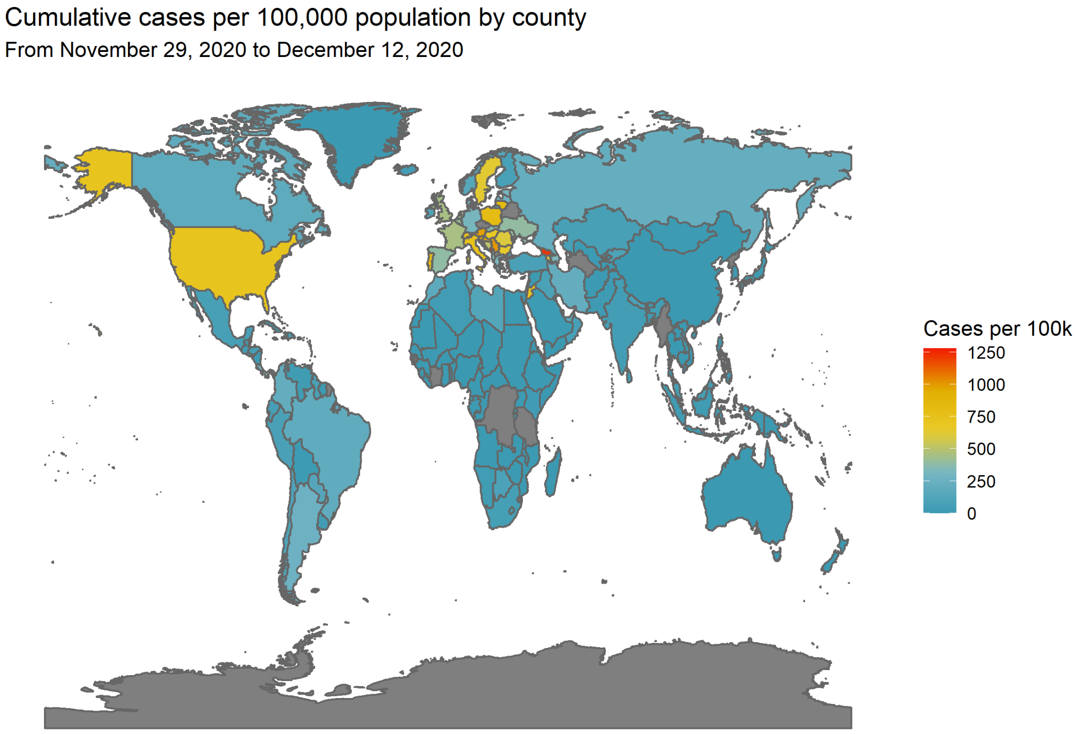
We construct a various models of the Covid-19 pandemic over various periods of 2020. We first construct simple model of a the epidemic by using a recurrence equation. We also add a periodic complexity to these simpler models. We then use more statistical methods, modelling using time series forecasting methods such as HoltWinters and ARIMA methods. All of this is with the aim of predicting the course of the epidemic.

1 Introduction

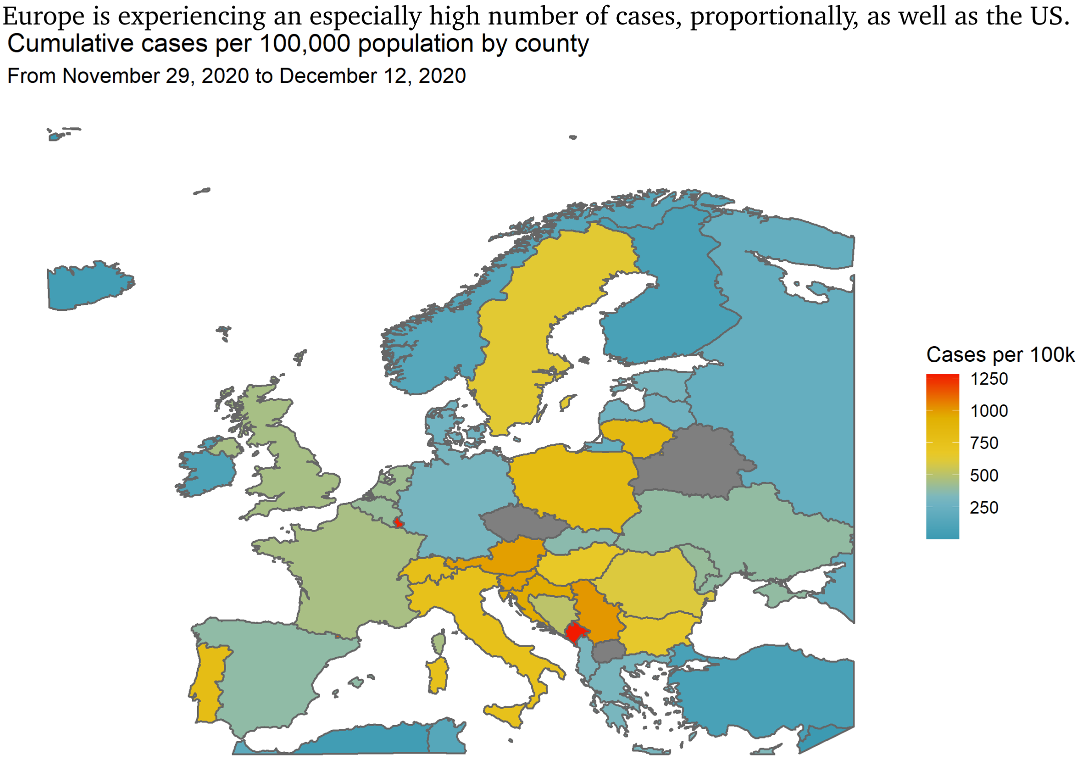
The Coronavirus disease (COVID-19) was first characterized by the World Health Organisation as pandemic on 11th March 2020 [7]. The outbreak has affected almost every aspect of human life throughout 2020, and is expected to continue for much of 2021.



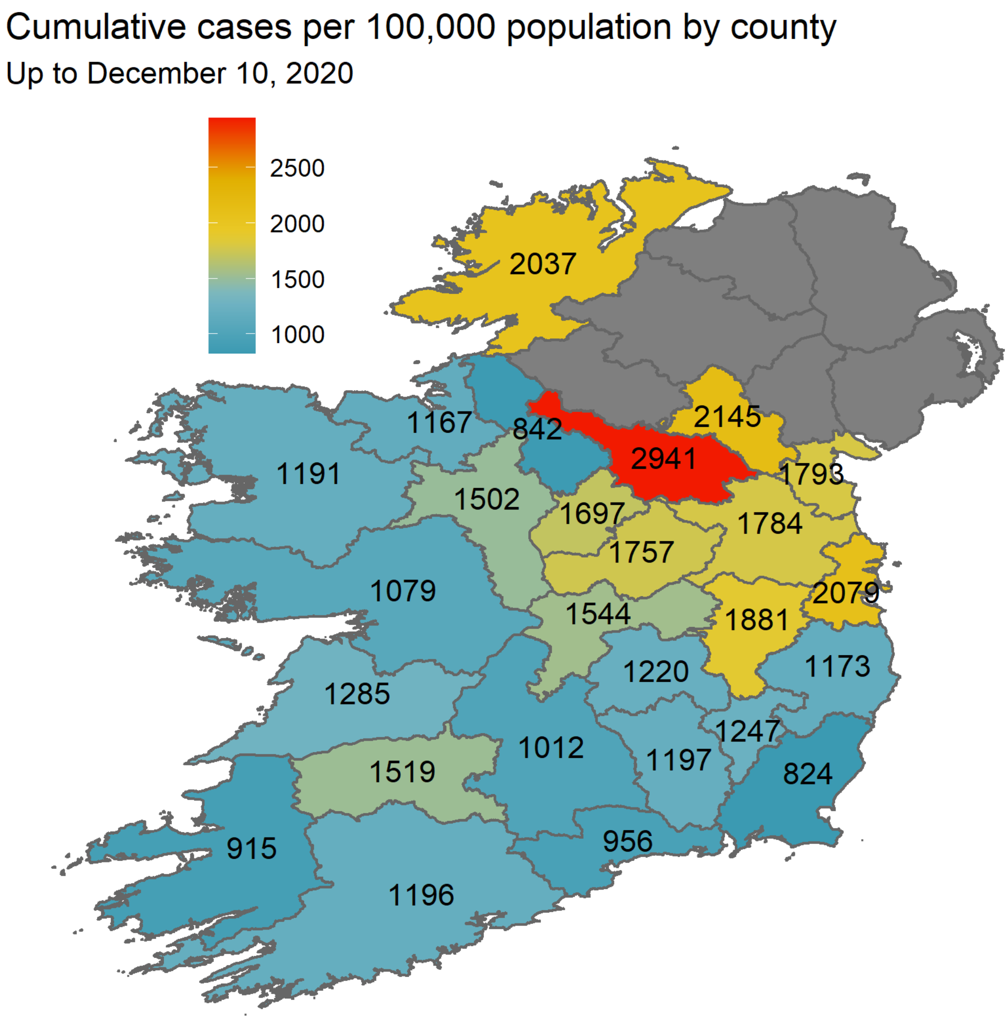
We can map the cumulative number of cases per 100,000 population for each country to see the varying severity of disease spread.



Europe is experiencing an especially high number of cases, proportionally, as well as the US.



More locally, we see that Ireland also has a clear variation in concentration of cases to date, with Donegal and much of Leinster experiencing sometimes twice as many cases per 100,000 population as the rest of the country.



This project is based on the work in [8], where I attempt to reconstruct the recurrence relation to model the pandemic. This largely mathematical model (based on practical assumptions), but of course does not fit well in the long run. It is efficient at explaining singular phases of the pandemic, and calculating the infamous R_0 number, defined below, from [2].

Definition. The number R_0 is called the *basic reproduction number* and is unquestionably the most important quantity to consider when analyzing any epidemic model for an infectious disease. Each infective individual can be expected to infect R_0 individuals.

1. Intro
2. Basic Model
3. Theorems and results
4. Periodic model
5. Multi-phase periodic model
6. Statistical models

2 Theorems

2.1 Model Assumptions

(I) Any infected person becomes ill and infectious on the q-th day after infection.[1]

(A) During each day, each ill person unconfined infects on average a other persons.

(B) During each day, a fraction b of ill people loose gets isolated (hospitalized or otherwise) and withdrawn from a further spread of the epidemic.

Many models use a set of differential equations for to describe the movement of people between *groups* or *compartments*[9, 4, 6]. The SIR (Susceptible–Infectious–Recovered) model, the most frequently used model in epidemiology, uses a set of 3 such differential equations [3, 5].

Our main mathematical model (and even some of the statistical models) make use recurrence equations, which have some correspondence to differential equations [1].

2.2 Notation

- x_n - the number of infected people that are detected and isolated during the day n;
- y_n – the cumulative number of detected cases from the beginning of epidemic by the beginning of the day n;
- z_n – the number of ill people at large by the beginning of the day n (that is, those who were infected at least q days ago and stay unisolated);
- u_n – the number of people newly infected during the day n.

We will obtain the following relation between the leading root r and the basic reproductive rate R_0 that is a main characteristic of an epidemic in epidemiology:

$$r \approx R_0^{\frac{1}{2q}}. \tag{1}$$

Recurrence relation for z_n:

$$z_{n+1} = z_n - x_n + u_{n-q}. \tag{2}$$

Using $x_n = bz_n$ we obtain the following equation for x_n:

$$x_{n+1} = (1-b)x_n + ax_{n-q}. \tag{3}$$

We let the model equal the actual data for the first $q+1$ days

$$x_n = x_n^* \text{ for } n = 0, 1, \dots, q, \tag{4}$$

To fit our model we optimize against the normalized 1-norm:

$$||x - x^*|| := \frac{1}{N+1} \sum_{n=0}^{N} |x_n - x_n^*|, \tag{5}$$

Similarly we define $||y - y^*||$

In order to determine values a, b, q, we want to minimize both

$$||x - x^*|| \text{ and } ||y - y^*|| \tag{6}$$

2.2.1 Why minimize both $||x - x^*||$ and $||y - y^*||$?

Do 3 pairs of plots: - xn/yn for just x-norm - xn/yn for just y-norm - xn/yn for both x-norm and y-norm

[1] The number of days before an infected person becomes infectious is called the latent period, and before he/she becomes symptomatically ill – the incubation period. Here we assume for simplicity that these two periods are equal.

2.2.2 Recurrence equation

This is our general linear recurrence equation with constant coefficients:

$$x_{n+1} = a_0 x_n + a_1 x_{n-1} + a_2 x_{n-2} + \cdots + a_q x_{n-q} \tag{7}$$

The characteristic polynomial of 7

$$f(\lambda) = \lambda^{q+1} - a_0\lambda^q - a_1\lambda^{q-1} - a_2\lambda^{q-2} - \cdots - a_{q-1}\lambda - a_q. \tag{8}$$

Definition 1. A root λ of f with the maximal absolute value $|\lambda|$ will be referred to as a leading root of the general linear recurrence relation 7.

2.3 Theorems

Theorem 1. Let $a_k \geq 0$ for all $k \in \{0, \dots, q\}$ and $a_{k_0} > 0$ for some $k_0 \in \{0, \dots, q\}$.

(a) (Cauchy, 1829) The polynomial $f(\lambda)$ from 8 has exactly one positive real root r. Besides, the root r is simple and, for any other root $\lambda \in \mathbb{C}$, we have $|\lambda| < r$. Consequently, r is the leading root of 7.

(b) For any positive solution x_n of 7, there exists $C > 0$ such that

$$x_n \sim Cr^n \text{ as } n \to \infty. \tag{9}$$

It follows from 9 that if $r < 1$ then the epidemic fades away, whereas if $r > 1$ then it spreads unlimited.

Proof:

(a) Although this statement is not new, we give here the proof as it is quite simple and a part of the argument will be used below. The equation $f(\lambda) = 0$ is equivalent to

$$0 = \lambda^{q+1} - a_0\lambda^q - a_1\lambda^{q-1} - a_2\lambda^{q-2} - \cdots - a_{q-1}\lambda - a_q$$

dividing across by λ^{q+1}

$$= 1 - \frac{a_0}{\lambda} - \frac{a_1}{\lambda^2} - \frac{a_2}{\lambda^3} - \cdots - \frac{a_{q-1}}{\lambda^q} - \frac{a_q}{\lambda^{q+1}}$$

And so

$$1 = \underbrace{\frac{a_0}{\lambda} + \frac{a_1}{\lambda^2} + \frac{a_2}{\lambda^3} + \cdots + \frac{a_{q-1}}{\lambda^q} + \frac{a_q}{\lambda^{q+1}}}_{g(\lambda)} \tag{10}$$

Since $a_{k_0} > 0$ for some k_0, and the remaining a_k are non-negative, $g(\lambda)$ is strictly monotone decreasing in $\lambda > 0$ (if $c\lambda$ is increasing,then $\frac{c}{\lambda}$ is decreasing), and we have the limits

- $\lim_{\lambda \to 0^+} g(\lambda) = +\infty$
- $\lim_{\lambda \to +\infty} g(\lambda) = 0^+$

Hence, there is exactly one positive value $\lambda = r$ that satisfies this $g(r) = 1$, that is,

$$1 = \frac{a_0}{r} + \frac{a_1}{r^2} + \frac{a_2}{r^3} + \cdots + \frac{a_{q-1}}{r^q} + \frac{a_q}{r^{q+1}}.$$

Now, let $\lambda \in \mathbb{C}\backslash\{0\}$ be another root of f. We obtain from 10 (using the triangle inequality) that

$$1 \leq \frac{a_0}{|\lambda|} + \frac{a_1}{|\lambda|^2} + \frac{a_2}{|\lambda|^3} + \cdots + \frac{a_{q-1}}{|\lambda|^q} + \frac{a_q}{|\lambda|^{q+1}}$$

And so $g(r) \leq g(|\lambda|)$ which implies $|\lambda| \leq r$ by the definition of decreasing functions.

We next need to show that the root r is simple. Denote by r' the largest non-negative root of the derivative $f'(\lambda)$ that exists for the following reason. If $a_k > 0$ for some $k < q$ then the polynomial

$\frac{1}{q+1}f'(\lambda)$ satisfies the hypotheses of the present theorem and, by the above argument, $f'(\lambda)$ has exactly one positive root, that is r'. If $a_k = 0$ for all $k < q$ then $f'(\lambda) = (q+1)\lambda^q$ has the only root 0, and, hence, $r' = 0$.

Let us verify that $r' < r$, which will also imply that r is simple. If $r' = 0$ then it is clear. If $r' > 0$ then it follows from $f'(r') = 0$ that

$$f'(\lambda) = (q+1)\lambda^q - qa_0\lambda^{q-1} - (q-1)a_1\lambda^{q-2} - (q-2)a_2\lambda^{q-3} - \cdots - a_{q-1} - 0$$

$$\frac{1}{q+1}f'(\lambda) = \lambda^q - \frac{q}{q+1}a_0\lambda^{q-1} - \frac{q-1}{q+1}a_1\lambda^{q-2} - \frac{q-2}{q+1}a_2\lambda^{q-3} - \cdots - \frac{1}{q+1}a_{q-1}$$

$$\frac{1}{q+1}f'(r') = (r')^q - \frac{q}{q+1}a_0(r')^{q-1} - \frac{q-1}{q+1}a_1(r')^{q-2} - \frac{q-2}{q+1}a_2(r')^{q-3} - \cdots - \frac{1}{q+1}a_{q-1}$$

$$0 = (r')^q - \frac{q}{q+1}a_0(r')^{q-1} - \frac{q-1}{q+1}a_1(r')^{q-2} - \frac{q-2}{q+1}a_2(r')^{q-3} - \cdots - \frac{1}{q+1}a_{q-1}$$

$$(r')^q = \frac{q}{q+1}a_0(r')^{q-1} + \frac{q-1}{q+1}a_1(r')^{q-2} + \frac{q-2}{q+1}a_2(r')^{q-3} + \cdots + \frac{1}{q+1}a_{q-1}$$

dividing both sides by $(r')^q > 0$

$$\begin{aligned}1 &= \frac{qa_0}{(q+1)r'} + \frac{(q-1)a_1}{(q+1)(r')^2} + \cdots + \frac{a_{q-1}}{(q+1)(r')^q} \\ &= \left(\frac{q+1-1}{q+1}\right)\frac{a_0}{r'} + \left(\frac{q+1-2}{q+1}\right)\frac{a_1}{(r')^2} + \cdots + \left(\frac{q+1-q}{q+1}\right)\frac{a_{q-1}}{(r')^q} \\ &= \left(1 - \frac{1}{q+1}\right)\frac{a_0}{r'} + \left(1 - \frac{2}{q+1}\right)\frac{a_1}{(r')^2} + \cdots + \left(1 - \frac{q}{q+1}\right)\frac{a_{q-1}}{(r')^q} \\ &< \frac{a_0}{r'} + \frac{a_1}{(r')^2} + \cdots + \frac{a_{q-1}}{(r')^q}\end{aligned}$$

So $g(r') > 1$, but $g(r) = 1$

$\implies g(r') > g(r) \implies r' < r$ by the definition of decreasing functions.

(b) Let $\lambda_1, \lambda_2, \ldots$ be all other distinct roots of f apart from r (so that λ_k are negative or imaginary). Any solution x_n of 7 has the form

$$x_n + Cr^n + \tilde{x}_n \tag{11}$$

where $\tilde{x}_n$ is a linear combination of the functions $n^j\lambda_k^n$. Since by (a) we have $|\lambda_k| < r$, it follows that

$$|\tilde{x}_n| = o(r^n) \text{ as } n \to \infty \tag{12}$$

Since $x_n > 0$, it follows from 11 and 12 that $C \geq 0$. Let us verify that $C > 0$, which will finish the proof. It is tempting to say that if $C = 0$ then $x_n = \tilde{x}_n$ is a linear combination of terms of the form $n^j\rho^n\sin(\phi n)$ and $n^j\rho^n\cos(\phi n)$ and, therefore, cannot stay positive. However, it is not easy to make this argument rigorous because different roots of f may have the same absolute value ρ and an uncontrollable cancellation of the terms can occur. We employ here a different, simpler approach that takes advantage of nonnegative coefficients a_k. To that end, consider a new sequence

$$X_n = \frac{x_n}{r^n}.$$

This satisfies the equation

$$X_{n+1} = A_0X_0 + A_1X_{n-1} + \cdots + A_qX_{n-q} \tag{13}$$

with $A_k = \frac{a_k}{r^{k+1}}$. Since r is a root of f, we have

$$A_0 + A_1 + \cdots + A_q = \frac{a_0}{r^1} + \frac{a_1}{r^2} + \cdots + \frac{a_q}{r^{q+1}}$$
$$= g(r)$$

This implies, by 10, and $g(r) = 1$ that

$$A_0 + A_1 + \cdots + A_q = 1 \tag{14}$$

Set $c := \min(X_1, \ldots, X_{q+1}) > 0$ since x_n have positive initial values. Then we obtain from 13 and 14 by induction that $X_n \geq c$ for all $n \in \mathbb{N}$, which implies

$x_n \geq cr^n$

as required.

□

Theorem 2. Let $a_k \geq 0$ for all $k = 0, \ldots, q$. Denote $a = a_1 + \cdots + a_q, b = 1 - a_0$ and assume that $a > 0, b > 0$.

(a) We have the equivalences: $r < 1 \iff a < b$ and $r > 1 \iff a > b$.

(b) Let $m \geq 1$ be such that $a_1 = \cdots = a_{m-1} = 0$ and $a_m > 0$. Then

$$\min\left(1, \left(\frac{a}{b}\right)^{1/m}\right) \leq r \leq \max\left(1, \left(\frac{a}{b}\right)^{1/m}\right) \tag{15}$$

Remark 1. Although there are in the literature plenty of estimates of the leading roots of polynomial (see, for example, [2]), none of them seems to imply 15. The latter is very useful for a basic model as we will see below in an example.

Proof:

(a) We have

$f(1) = 1 - a_0 - a1 - \cdots - a_q$
$= \underbrace{(1 - a_0)}_{b} - \underbrace{(a_1 + \cdots + a_q)}_{a}$
$= b - a$

We know f is increasing.

So if $r < 1$, we have $f(1) > 0$ and then $b - a > 0 \implies a < b$.

And if $r > 1$, we have $f(1) < 0$ and then $b - a < 0 \implies a > b$

(b) $f(r) = 0$ is equivalent to

$r^{q+1} - a_0 r^q - a_1 r^{q-1} - a_2 r^{q-2} - \cdots - a_{q-1} r - a_q = 0$

But any $a_1, \ldots, a_{m-1}$ are all zero

$\implies r^{q+1} - a_0 r^q - a_m r^{q-m} - a_{m+1} r^{q-m-1} - \cdots - a_{q-1} r - a_q = 0$
$\implies r^{q+1} - (1 - b) r^q - a_m r^{q-m} - \cdots - a_q = 0$
$\implies r^{q+1} - r^q + br^q - a_1 r^{q-m} - \cdots - a_q = 0$
$\implies r^{q+1} - r^q = -br^q + a_m r^{q-m} + \cdots + a_q$

If $r > 1$ then $r^{q+1} > r^q$ and so $r^{q+1} - r^q > 0$

and so

$0 < -br^q + a_m r^{q-m} + \cdots + a_q$

$$
\begin{aligned}
\implies br^q &< a_m r^{q-m} + \cdots + a_q \\
&\leq a_m r^{q-m} + \cdots + a_q r^{q-m} \\
&= (a_m + \cdots + a_q) r^{q-m} \\
&= a r^{q-m}
\end{aligned}
$$

So $br^q < ar^{q-m} \iff r^m = \frac{a}{b} \iff r < \left(\frac{a}{b}\right)^{1/m}$

And if $r < 1$ we get $r < \left(\frac{a}{b}\right)^{1/m}$.

We can combine both cases with $a \leq \max\left(()\, 1, a\right)$ and $a \geq \min\left(1, a\right)$ to get 15, as required.

□

Lemma 3. For the model described by equation 7 we have

$$R_0 = \frac{a}{b}$$

Proof: Let u be the number of people infected on some day, say 0. On the day $k = 1, \ldots, q$ the number $c_k u$ of them become ill and can infect other people. On the day $k+1$ they infect $ac_k u$ people while $bc_k u$ of them get isolated. On the day $k+1$, the remaining $(1-b)c_k u$ people infect further $a(1-b)c_k u$ people. Continuing this way, we obtain that this group of $c_k u$ people infects in total

$$ac_k u + a(1-b)c_k u + a(1-b)^2 c_k u + \cdots = ac_k u \sum_{n=0}^{\infty} (1-b)^n = \frac{ac_k u}{1-(1-b)} = \frac{a}{b} c_k u$$

since $0 < 1 - b < 1$.

other people.

Hence, the initial group of u people infects in total

$$\sum_{k=0}^{q} \frac{a}{b} c_k u = \frac{a}{b} u \sum_{k=0}^{q} c_k = \frac{a}{b} u$$

So we know R_0 is the unit reprodiction number per infected person ($u = 1$).

And so we get the result $R_0 = \frac{a}{b}$ as required. □

3 R Code and Data Sources

Much of the code was written from scratch for this peoject, or is a close to direct translation of the formulas described in papers such as

3.1 Plotting and colour

wesanderson

3.2 Shapefiles

3.3 Datasets

References

[1] Ravi Agarwal et al. "Dynamic equations on time scales: a survey". In: *Journal of Computational and Applied Mathematics* 141.1 (2002). Dynamic Equations on Time Scales, pp. 1 –26. ISSN: 0377-0427. DOI: https://doi.org/10.1016/S0377-0427(01)00432-0. URL: http://www.sciencedirect.com/science/article/pii/S0377042701004320.

[2] L.J.S. Allen et al. *Mathematical Epidemiology*. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2008. ISBN: 9783540789109. URL: https://books.google.ie/books?id=gcP5l1a22rQC.

[3] Roy M. Anderson. "Discussion: The Kermack-McKendrick epidemic threshold theorem". In: *Bulletin of Mathematical Biology* 53 (). ISSN: 522-9602. DOI: 10.1007/BF02464422. URL: https://doi.org/10.1007/BF02464422.

[4] Derdei Bichara, Abderrahman Iggidr, and Gauthier Sallet. "Global analysis of multi-strains SIS, SIR and MSIR epidemic models". In: *Journal of Applied Mathematics and Computing* 44 (Feb. 2014), pp. 273–292. DOI: 10.1007/s12190-013-0693-x.

[5] Alexander Bird. *A simple introduction to epidemiological modelling—the SIR model*. https://philosophyandmedicine.org/wp-content/uploads/2020/04/Introduction-to-epidemiological-modelling.pdf. 2020.

[6] Jesús Fernández-Villaverde and Charles I Jones. *Estimating and Simulating a SIRD Model of COVID-19 for Many Countries, States, and Cities*. Working Paper 27128. National Bureau of Economic Research, May 2020. DOI: 10.3386/w27128. URL: http://www.nber.org/papers/w27128.

[7] Dr Tedros Adhanom Ghebreyesus. *WHO Director-General's opening remarks at the media briefing on COVID-19*. World Health Organization. Mar. 11, 2020. URL: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

[8] Alexander Grigorian. *Mathematical riddles of COVID-19*. June 2020. URL: https://www.math.uni-bielefeld.de/~grigor/corv.pdf.

[9] Roni Parshani, Shai Carmi, and Shlomo Havlin. "Epidemic Threshold for the Susceptible-Infectious-Susceptible Model on Random Networks". In: *Physical Review Letters* 104.25 (June 2010). ISSN: 1079-7114. DOI: 10.1103/physrevlett.104.258701. URL: http://dx.doi.org/10.1103/PhysRevLett.104.258701.