MATH 575: Disease Transmission

1 Introduction

Since early 2020, the world has experienced massive changes in economic and social policies due to the emergence of the novel coronavirus COVID-19. While every country has been affected, poor and less industrialized nations have suffered the most significant and devastating effects of the disease, mostly due to lack of access to reliable healthcare and dependence on just a few key industries [11]. On the other end of the spectrum, highly industrialized first world countries such as the U.S. and the U.K. have also experienced significant impacts. Beyond fatalities caused by COVID-19, business and school closures have had wide ranging consequences for the educations and livelihoods of millions [2]. These closures have been the result of social and economic policies that restrict the way in which individuals interact in hopes of lowering the transmission and spread of COVID-19.

While many have balked and dismissed the science and reasoning behind such actions, here we aim to shed light on the mathematical ideas and tools that allow us to model the transmission of COVID-19 and thereby the justification to enact such measures. Section 2 describes the underlying models in both continuous and discrete cases, while Section 3 gives us a deeper insight into the mathematical underpinnings of the models used.

2 Age Structured Population Dynamics

Age structured population dynamics seek to analyze the rate a population grows, shrinks, or reaches some steady state. Based upon age classes within the population and their corresponding mortality and reproduction rates, these models give insight into the age distribution over time [8]. We will first describe these models intended purposes and then show how they can be adjusted for modelling an epidemic.

2.1 Continuous Method

The continuous model that we first introduce was formulated by Alfred Lotka who unknowingly extended earlier work done by Euler. Their focus was the dynamics of age structured populations and their respective birth and mortality rates in terms of continuous variables with respect to time t. Let's define some of these variables in order to get a handle on the mathematics behind the model. We note that usually only the female sex is considered in this model due to the fact that only females can give birth.

$B(t)dt$ Number of female births during the time interval t to $t + dt$

$n(a, t)da$ Number of females of age a to $a + da$ at a time t

$l(a)$ fraction of newborn females surviving to age a

$m(a)da$ number of females born, on average, to a female during the ages a to $a + da$

We have three quantities to consider here for each age group, the maternity or reproductive rate, the survivor rate, and the number of individuals in that age group. We note that there exists a 'terminal' age group γ where the survivor rate is essentially zero and indicates death by natural

causes like old age. Putting these together we have:

$$B(t) = \int_0^t B(t-a)l(a)m(a)da + G(t) \tag{2.1}$$

We integrate over all possible age groups a to find the overall birth rate of the population at a time t. The extra function $G(t)$ is the birthrate of the women of age a when $t = 0$ and can be given by:

$$G(t) = \int_0^{\gamma - t} n(a,0)\frac{l(a+t)}{l(a)}m(a+t)da$$

Now we know that at a certain age the females in the younger age group will begin to be able to reproduce and conversely females that reach an older age group ϕ will cease to reproduce. Therefore if we let $t > \beta_0$, then we are essentially saying that nobody in the initial population from $t = 0$ can reproduce, and so $G(t) = 0$. So Equation 2.1 will become:

$$B(t) = \int_{\beta_0}^t B(t-a)l(a)m(a)da$$

Therefore we assume that our initial population from $t = 0$ is no longer contributing the reproductive capacity of the population. Being somewhat loose with our notation, we then let $\alpha \geq \beta_0$ such that we can represent our integral as:

$$B(t) = \int_\alpha^\beta B(t-a)l(a)m(a)da \tag{2.2}$$

So what information can we deduce and what operations can we do with such an integral? It is linear, and if we subtract $B(t)$ to the right side then it is a homogeneous equation. We can "guess" an exponential solution, let $B(t) := Qe^{rt}$ and substituting in:

$$Qe^{rt} = \int_\alpha^\beta Qe^{r(t-a)}l(a)m(a)da$$

$$1 = \int_\alpha^\beta e^{-ra}l(a)m(a)da$$

which is the famous Euler-Lotka equation. Let's differentiate this function to see the behavior of the function.

$$\frac{d}{dr}\left(1 = \int_\alpha^\beta e^{-ra}l(a)m(a)da\right)$$

$$\int_\alpha^\beta -ae^{-ra}l(a)m(a)da$$

which is clearly less than 0 for all a, so our function is decreasing. Differentiating once more:

$$\frac{d^2}{dr^2}\left(\int_\alpha^\beta e^{-ra}l(a)m(a)da\right) \tag{2.3}$$

$$\frac{d}{dr}\left(\int_\alpha^\beta -ae^{-ra}l(a)m(a)da\right) \tag{2.4}$$

$$\int_\alpha^\beta a^2e^{-ra}l(a)m(a)da \tag{2.5}$$

The second derivative is greater than 0, and so the function is concave up. This means that there is exactly one point R_0 at which the solution satisfies $1 = \int_\alpha^\beta e^{-ra} l(a)m(a)da$. It should be noted that this is the only real root of Equation 2.2 and that all other roots are complex conjugates [6], of which a deep analysis is far beyond the scope of this paper. Finding this r will give us the R_0 which is the "lifetime reproductive potential of a female, corrected for mortality" [6].

Let's note a few observations before moving on. First we immediately notice that R_0 behaves exactly as we expect, for $|R_0| < 1$, a population will decrease and eventually die out (more succintly as $\lim_{t\to\infty} P \to 0$). Similarly, for $R_0 > 1$ a population will increase and if $R_0 = 1$ then the population will eventually reach a steady state, meaning some constant age distribution.

Next, observe that Equation 2.2 is an integral with an exponential term and some other function, all involving some temporal parameter t. This allows us to use a LaPlace transform $\mathcal{L}$ [14] to conduct a further analysis of Equation 2.2.

Finally, we note that a criticism of the Euler-Lotka integral is that the choice of age classes and age ranges and how they are defined is determined by the researcher. Though it will not yield wildly different results for two carefully chosen constructs, the interpretations may lead to differing conclusions.

Now, R_0 is arguably the most important piece of information that we can garner from the Euler-Lotka integral and though the problem was phrased in terms of the reproductive rate of a population, it can also be viewed as the reproductive rate of a virus. Before we further explore this scenario, let's differentiate between a reproductive rate and a transmission rate. The transmission rate will describe how easily an infected individual will transmit the virus, where as the reproductive rate describes the lifetime average number of people that an infected person will transmit the virus to, which is dependent on the transmission rate. We also note that the spread of a virus or disease changes due to societal, economic, and public health changes and policies. However due to the nature of viral transmission and how we will not necessarily reach any steady state (unless $R_0 = 1$), we cannot hope that R_0 will remain our lifetime reproductive rate. Therefore we define R_0 to be a baseline effective reproductive rate and calculate some $R(t)$ on a rolling basis. $R(t)$ may also be compromised of functions similar to the maternity rate $m(a)$ and morbidity rates $l(a)$ but slightly different, such as biological infectiousness and contact rates [5]. Now that we have clarified the terminology, let's examine an application of R_0 in disease modeling.

2.2 Reproductive Rates

2.2.1 2020 COVID Modeling in India

Written in late 2020, [12] builds a model for estimating the COVID-19 daily positive case count and the effective reproductive rate. Though now outdated, it provides a key glimpse into the application of the Euler-Lotka integral to find this effective reproductive rate. R_0 is defined literally, meaning the initial period of ~ 5 to ~ 14 days of disease spread. They use a product of the reproductive rate $m(a)$ and mortality rate $l(a)$ of Equation 2.2 such that $l(a) \cdot m(a) = g(a)$, their setup is:

$$R(t) = \frac{1}{\int_{a=0}^{\infty} e^{-ra} g(a)da}$$

This function $g(a)$ represents "serial interval distribution" [12] where serial interval is defined as the time between diagnosis and transmission [10]. Because these intervals need to be generalized, different probability distributions were tested in [1] using Chinese data from January to February

of 2020. While the accuracy of Chinese reported data is dubious, they determined a Gamma probability distribution most accurately modelled the serial intervals for the data given. We also note that the results of the paper indicate a predicted spike of COVID-19 in India in September 2020 of $\sim 105,000$ positive cases. This seems accurate, since reported cases around this time peaked at $\sim 75,000$ [3], indicating that total possible cases (including unconfirmed) were in line with the article. However, the highly contagious delta variant of COVID-19 clearly shows that future modeling needed to be adjusted for the higher transmission rate.

2.2.2 Alternate Method: SEIR

While Section 2.2.1 implemented the Euler-Lotka equation to predict the reproductive rate of COVID-19, there are other methods available to quantify and discover the reproductive rates of the disease among a population. The most notable of these is SEIR model which was implemented in [13] to predict the spread of COVID-19 in Hubei province China.

Since we have studied a discrete SIR model in [4], let's review our method so we can compare it to this continuous and slightly more complex SEIR model. Our variables are:

- S denotes individuals not currently infected but that become infected.
- I denotes a current infection.
- R denotes recovered or removed, this category of individuals was deemed to 'no longer become infected'.
- k denotes the average length of infection.
- a denotes some constant transmission rate.

We established a system of difference equations:

$$\begin{aligned} R_{n+1} &= R(n) + kI_n \\ I_{n+1} &= I_n - kI_n + aI_nS_n \\ S_{n+1} &= S_n + aI_nS_n \end{aligned}$$

Let's note that we made a number of assumptions:

- Recovery from infection implies complete immunity.
- Constant population size with no outside contact.
- No precautionary measures such as lock-downs or social policies to curb transmission.
- Every individual falls into exactly one of the three categories S, I, or R.

Now, let's take a look at the system of differential equations. Notice the immediate similarities in which they define each letter in SEIR almost exactly to our discrete SIR model:

- N denotes the total population size.
- S denotes the population that is susceptible to the catching and/or carrying the disease, with a probability of β to enter the class E.

- Exposed individuals E have a probability α to display symptoms, and therefore enter the next class I. We also denote β_2 as the probability that they will infect a member of S.
- The infected population I eventually join the recovered class R, either through recovery or death. The probability of recovery is denoted γ and the probability of infecting a member of S is β_1.

Therefore with these quantities defined we can set up a system of linear differential equations where r_1 is the average number of people from S who are in close contact with a member of I and similarly r_2 denotes the average number of people from S in close contact with a member of E. We have:

$$\begin{aligned} \frac{dS}{dt} &= \frac{-r_1\beta_1 IS}{N} - \frac{-r_2\beta_2 IS}{N} \\ \frac{dE}{dt} &= \frac{-r_1\beta_1 IS}{N} - \alpha E + \frac{-r_2\beta_2 IS}{N} \\ \frac{dI}{dt} &= \alpha E - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned}$$

Within this system, we still need to calculate the reproductive rate, which they term the 'basic reproductive number' R_0. Their method is for R_0 to be first estimated by $R_0 = \frac{\beta}{\gamma}$, but just as above we cannot hope to ever reach. Additionally preventative measures such as lock-downs will undoubtedly affect $R(t)$ and so it must be calculated on a rolling basis. Therefore their final calculation is:

$$R_e(t) = \frac{\beta(t)}{\gamma(t)}$$

Criticisms of this method are based on the fact that infection rates β_i are constants, and may not be accurate over a given time interval. In the case of Euler-Lotka, we have multiple reproductive (or transmission) rates corresponding to how the population is structured. This allows more quantitative information to be factored into how we ultimately determine the effective reproduction rate $R(t)$, though care needs to be given in determining the population structure.

2.3 Discrete Method

Next we move onto a discrete analogy of stable population theory, but with a greater emphasis placed on the age distribution and not the reproductive rate. We employ almost identical variables and again consider only females:

$$\begin{aligned} P_{k,t} &\text{ Number of females in the population in age class } k \text{ at time } t \\ f_k &\text{ the number of daughters born per female from age } t \text{ and } t+1 \\ l(k) &\text{ the probability of a female surviving until age } k+1 \end{aligned}$$

We let K be the maximum age class and assume that no member of the population exceeds the range defined for K. Similarly we know that at some threshold age class i then $f_i = 0$ and so

females are no longer able to reproduce for classes $i = i+1, i+2, \ldots, K$. This gives us a linear system of equations:

$$\begin{aligned} P_{0,t+1} = f_0 \cdot P_{0,t} + f_1 \cdot P_{1,t} + \ldots f_K \cdot P_{K,n} \\ P_{1,t+1} = l_0 \cdot P_{0,t} \\ P_{2,t+1} = l_1 \cdot P_{1,t} \\ \vdots \\ P_{K,t+1} = l_{K-1} \cdot P_{K-1,t} \end{aligned}$$

but this can be split and written into a matrix and a vector, where the population $P_{i,t}$ is a vector of size $K-1$ but for ease of notation we can set this to $P_{i,t} = P_t$ and A is a square matrix of the same dimensions $(K+1)$.

$$A = \begin{pmatrix} f_0 & f_1 & \cdots & f_K \\ l_0 & 0 & \ldots & 0 \\ 0 & l_1 & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & l_{K-1} & 0 \end{pmatrix}; P = \begin{pmatrix} P_0 \\ P_1 \\ P_2 \\ \vdots \\ P_K \end{pmatrix}$$

We are then solving:

$$P_{t+1} = AP_t$$

But we are interested in finding the stable age distribution so we aim to find the dominant eigenvalue and corresponding eigenvector such that:

$$P_t = A^t P_0$$

It should come as no surprise that the characteristic equation of a Leslie matrix is also the Lotka-Euler equation. We know that there exists only one real root of Equation 2.2 and this real root R_0 is exactly the dominant eigenvalue that corresponds to the eigenvector representing the steady state age distribution[6]. In this sense we can also garner the same information that we expected from $R_0 = \lambda_0$:

$$\lambda_0 = \begin{cases} \to \infty & \text{iff } \lambda_0 > 1 \\ \to 0 & \text{iff } \lambda_0 < 1 \\ \to c & \text{iff } \lambda_0 = 1 \text{ for some constant } c \end{cases}$$

We notice immediately that this is much of the same mathematics that we encountered in Project 1 while studying the machinery of Google's Page Rank algorithm. In fact, Leslie matrices and Markov matrices differ only in the fact that the sum of all entries in a column vector of a Leslie matrix may exceed 1. We conclude both this section and paper by examining a theorem that seemed of little importance when we first encountered it but now explains much of the recurring themes within steady state and asymptotic growth problems.

3 Perron-Frobenius Theorem

Theorem 1 ([7]). *The eigenvalue of largest absolute value of a positive square matrix A is both simple and positive and belongs to a positive eigenvector. All other eigenvectors are smaller in absolute value.*

Perhaps a little anti-climatic, this theorem has profound implications. First we must observe that in all the models we described previously in this paper all consisted of non-negative quantities. A population can certainly be zero, but never negative. Similarly, the mortality and reproductive rates of each class a cannot be negative. And in the case of the Leslie matrices, we were determining the steady state value by iteratively raising our matrix A to a power. Additionally, models of this form "appear widely in the literature on social sciences (occupational mobility, group interaction, human resource management, socio-economic planning), engineering (communications, control, production processes, information processing, power systems, water-resource systems), biology (population dynamics), etc."[9]. Camille Jordan stated:

Theorem 2 ([7]). *An iteration scheme $x^k = A^k x_0$ converges for every initial state x_0 iff the eigenvalues λ of A distinct from 1 have modulus $|\lambda| < 1$, and if $\lambda = 1$ its eigenspace is of full rank, i.e., its rank equals the multiplicity of the root $\lambda = 1$ in the characteristic equation $\phi(\lambda) = det(\lambda I - A)$.*

This further enforces what we know about the behavior of the principle eigenvalue $\lambda_0 \equiv R_0$ and a systems behavior. In a sense we are describing the foundation of many currently used mathematical models by simply setting up systems of equations with non-negative values.

4 Conclusion

In this brief survey paper, we have shown a few key methodologies, derivations, and histories behind mathematically modelling disease transmission. While it may be somewhat of a surprise that they are re-purposed from classifying aged structured population dynamics, we are confident that these models are accurate in their predictions and analysis. In addition we have briefly highlighted how the mechanics of the Perron-Frobenius theorem gives us enormous power and leverage for classifying and understanding any linear system of equations and its behavior.

We are reminded that a model is only as good as the formulation of the problem it seeks to answer and the quality and quantity of the data it utilizes. However, we can be confident that the methods available to us are perfectly valid for influencing social and economic policies as we continue to live through the COVID-19 epidemic.

References

[1] S. T. Ali et al. "Serial intervals and case isolation delays for COVID-19: a systematic review and meta-analysis". In: *Clin Infect Dis* (2021).

[2] Kimberly Chriscaden. *Impact of Covid-19 on People's Livelihoods, Their Health and Our Food Systems*. URL: https://www.who.int/news/item/13-10-2020-impact-of-covid-19-on-people\%27s-livelihoods-their-health-and-our-food-systems.

[3] Dong, Du, and Gardner. *An interactive web-based dashboard to track COVID-19 in real time*. 2020. URL: https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30120-1/fulltext.

[4] Steven B. Horton Frank R. Giordano William P. Fox. *A First Course in Mathematical Modeling*. 5th ed. Cengage Learning, 2013. ISBN: 1-285-05090-8,978-1-285-05090-4.

[5] Nicholas C. Grassly and Christophe Fraser. *Mathematical Models of Infectious Disease Transmission*. 2008. URL: https://www.nature.com/articles/nrmicro1845/.

[6] Mark Kot. *Elements of mathematical ecology*. Cambridge University Press, 2001.

[7] C. R. MacCluer. "The many proofs and applications of Perron's theorem". In: *SIAM Review* 42.3 (2000), 487–498. DOI: 10.1137/s0036144599359449.

[8] Fabio Milner (auth.) Mimmo Iannelli. *The Basic Approach to Age-Structured Population Dynamics: Models, Methods and Numerics*. 1st ed. Lecture Notes on Mathematical Modelling in the Life Sciences. Springer Netherlands, 2017. ISBN: 9789402411461; 9402411461; 9789402411454; 9402411453.

[9] V.G. Rumchev† and D.J.G. James. "The role of non-negative matrices in discrete-time mathematical modelling". In: *International Journal of Mathematical Education in Science and Technology* 21.2 (1990), pp. 169–182. DOI: 10.1080/0020739900210201. eprint: https://doi.org/10.1080/0020739900210201. URL: https://doi.org/10.1080/0020739900210201.

[10] *Serial interval*. 2021. URL: https://en.wikipedia.org/wiki/Serial_interval.

[11] *The COVID-19 Effects on Societies and Economies*. 2021. URL: https://wellcome.org/news/equality-global-poverty-how-covid-19-affecting-societies-and-economies.

[12] Suryakant Yadav and Pawan Kumar Yadav. "The Peak of COVID-19 in India". In: *medRxiv* (2020). DOI: 10.1101/2020.09.17.20197087. URL: https://www.medrxiv.org/content/early/2020/09/25/2020.09.17.20197087.

[13] Yiping Zeng et al. *Forecasting of COVID-19: Spread with dynamic transmission rate*. 2020. URL: https://www.sciencedirect.com/science/article/pii/S2666449620300232.

[14] Dennis G. Zill. *A First Course in Differential Equations: The Classic Fifth Edition*. 5th ed. Brooks Cole, 2000. ISBN: 0534373887,9780534373887.