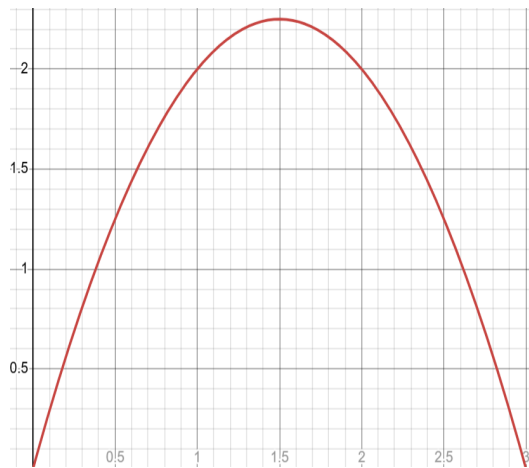


The Mathematics behind Understanding the Spread of Infectious Diseases

Harshit Kumar Raina

Introduction

A mathematical model (according to Wikipedia) is an abstract description of a concrete system using mathematical concepts and language. Simply put, mathematical modelling is where we use mathematical terms to describe real-life problems. For example, a simple mathematical model may use an inverted parabola (an upside down U-shape) to describe the motion of a ball thrown through the air, such as the following graph of $y = -x^2 + 3x$, which I made in 1 minute:



As you may expect, mathematical modelling has a vast number of real world applications, but due to my combined interest in mathematics and biology, I will explain one of its applications: epidemiology.

Along with being difficult to pronounce, epidemiology is also the study of how often different diseases occur in different groups, and why this happens. This is especially useful in strategies to prevent illness. Many of you readers may have even seen graphs modelling this during the Covid-19 pandemic.

In this essay, I am going to outline two models that are used to model the spread of infectious diseases, and evaluate their pros and cons. These are the models I will discuss, in order of increasing complexity:

- The SI Model
- The SIR Model

These 2 models are examples of what are known as compartmental models, as they split the population up into smaller sections which can exchange the things that they contain with each other.

The SI Model

This is the more basic of the two models I will be discussing. This splits the whole population, P , into two groups: Susceptible, S , and Infected, I . The susceptible group are those who do not have the disease yet but are able to get it later and the infected group are those who already have the disease and are infectious to others. A susceptible person can become infected by contact with an infected person.

Let's discuss the modelling assumptions in this model:

- An infectious disease has a Basic Reproduction number, R_0 . This is the average number of people who will be infected by an infected person.
- Initially, the whole population, P , is considered susceptible. The population is considered a constant value, meaning it doesn't change (no births or deaths).
- Since everyone is either susceptible or infected, we can say that $S+I=P$. In order to make the maths simpler, we can consider the proportions of the whole population instead, which we can get by dividing the entire equation by P . We can use the letters s and i to represent the proportion of the population that is susceptible and infected respectively.
- This leads to this equation: $s+i=1$
- An infected person is immediately infectious.
- We can let $\beta\Delta t$ be the probability that a random infective person infects a random susceptible person in the time Δt . This is also known as the transmission rate. This is calculated by dividing R_0 by the infectious period.

In the SI model, we consider how quickly the infective group grows to predict the growth of the disease. In order to study the growth of something, we need to set up a differential equation (an equation that involves derivatives). Before this we must ask ourselves, how many people become infected during the time ΔT ? To do this we can write this equation:

$$i(t + \Delta T) = i(t) + \beta\Delta t s(t)i(t)$$

This equation may look intimidating, but what it's really saying is that the proportion of infectives after time ΔT is equal to the initial proportion of infectives + the new infectives in the time ΔT , which is the transmission rate multiplied by the proportion of susceptibles and the proportion of infectives. This is because $s \times i$ represents how interactive the two groups are with each other.

If we differentiate the equation $i(t + \Delta T) = i(t) + \beta\Delta t s(t)i(t)$ of i with respect to t , we get this:

$$\frac{di}{dt} = \beta si$$

The rate of change of susceptibles would be the opposite of this and so would just be $(-\beta si)$. By solving this equation, we will have a relationship between the proportion of infected people and time. First of all, we can eliminate s from the equation as we know that $s + i = 1$. So now we get this equation:

$$\frac{di}{dt} = \beta(1 - i)i$$

In order to solve this we need to separate the variables:

$$\int \frac{1}{(1-i)i} di = \int \beta dt$$

From here, we can separate the integral on the left using partial fractions and then solve the integrals which leads to this:

$$\ln \left| \frac{i}{1-i} \right| = \beta t + c$$

In order to work out the constant c we can consider the point when $t=0$ (the start of the outbreak). At this point, i is equal to the initial number of infectives. We can call this i_0 . Now if we substitute $t=0$ and $i=i_0$ into the equation above, we can find c :

$$c = \ln \left| \frac{i_0}{1-i_0} \right|$$

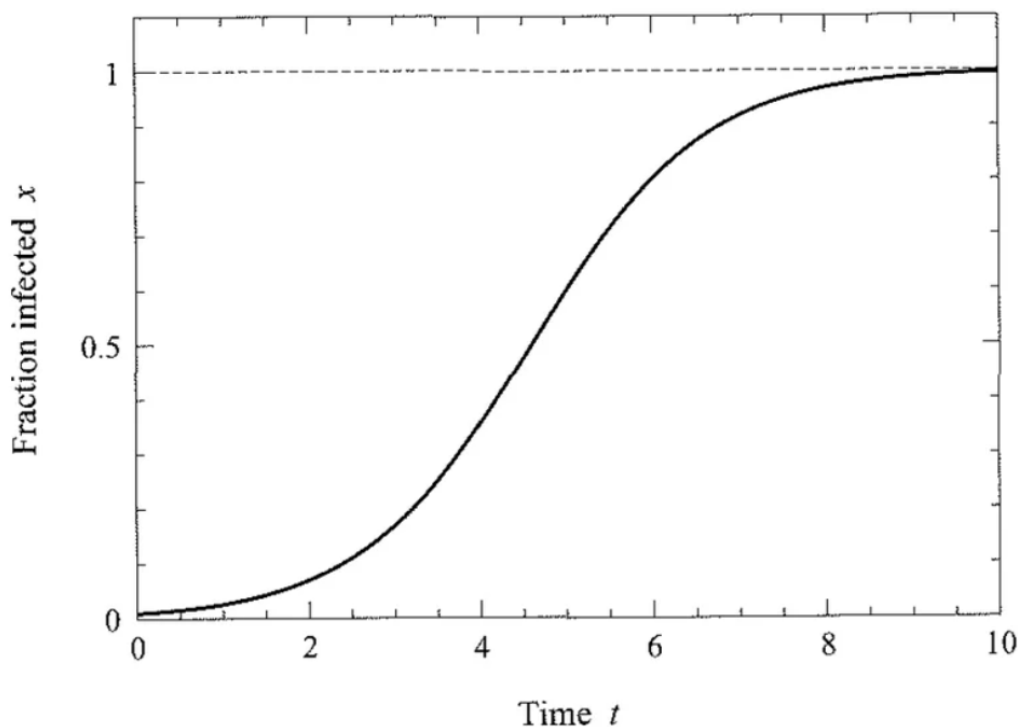
This means that our equation to solve is this:

$$\ln \left| \frac{i}{1-i} \right| = \beta t + \ln \left| \frac{i_0}{1-i_0} \right|$$

Now we can rearrange this to get i on its own which results in this equation linking the proportion of infective people over time:

$$i = \frac{e^{\beta t} i_0}{1 - i_0 + e^{\beta t} i_0}$$

We can now interpret the results of this equation: as $\text{time} \rightarrow \infty$ (as time tends to infinity), $i \rightarrow 1$. What this means is that given enough time, the entire population becomes infected, as shown in this graph:



Although this model is simple, it has its limitations as it predicts that over time the entire population would be infected, which is not accurate as in real life people would recover or die, and the curve would slope back down. This brings us on to our next model: the SIR model.

The SIR Model

This model is an advancement from the SI model as it accounts for recovery (and death) by adding a third compartment, the Removed group, R . In the first model, people only moved from the susceptible group to the infected group. Building on this, in the SIR model, people move from the susceptible group to the infected group, and then from the infected group to the removed group. Those in the removed group are people who have recovered or died, and in either case they are now considered immune to the disease (you can learn more about the biology behind this, but for now, let's consider the maths).

Once again, there are some modelling assumptions we need to consider first:

- We still keep most of our assumptions from the SI model. This means we still consider the Basic Reproduction number, R_0 , and the transmission rate, $\beta\Delta t$. We also still consider the whole Population, P , to be constant (even though this model does account for deaths).
- We let γ be the recovery rate, which is how quickly a person recovers from the disease. This is equal to 1 divided by the number of infectious days.
- Since the population is constant, we have the equation $S+I+R=P$. Once again, we can consider proportions, with s being the proportion of susceptibles, i being the proportion of infectives, and r being the proportion of removed people. This gives the following equation: $s+i+r=1$. It is also worth stating that, r_0 , which is the proportion of recovered people at the beginning is equal to 0, so at the beginning, $s_0+i_0=1$, where s_0 and i_0 are the initial proportions of susceptibles and infectives respectively.

We still initially consider the whole population to be susceptible, and we once again consider the rates of change in the number of people in the different compartments. The overall rate of change should be 0, so when we differentiate each term with respect to time, we get this equation:

$$\frac{ds}{dt} + \frac{di}{dt} + \frac{dr}{dt} = 0$$

The rate of change of susceptibles is still $\frac{ds}{dt} = -\beta si$. The rate of change of i is the rate of people entering the infective group and the rate of people leaving the infective group. The rate of people leaving the infective group is proportional to γi . This gives the following equation:

$$\frac{di}{dt} = \beta si - \gamma i$$

Following on, we can tell that the rate of change of the removed group is simply:

$$\frac{dr}{dt} = \gamma i$$

In practical use, these equations are often solved using iterative techniques (meaning they repeatedly calculate values to solve the equations) such as Euler's Method or the Runge-Kutta Method. However, we can still partially solve this analytically by dividing $\frac{di}{dt}$ by $\frac{ds}{dt}$ to get an invariant relationship between i and s . This gives us the following:

$$\frac{di}{ds} = -1 + \frac{\gamma}{\beta s}$$

Now we integrate and calculate the constant c :

$$i + s - \frac{\gamma}{\beta} \ln s = i_0 + s_0 - \frac{\gamma}{\beta} \ln s_0$$

This can help us with a number of things, such as calculating the maximum number of infectives, which is crucial for distributing health resources in real life. To do this we can substitute $s = \frac{\gamma}{\beta}$ into the equation above, as this value of s is when the peak of infectives would occur (as $\frac{di}{ds} = 0$). Doing this leads to the equation:

$$i_{max} = i_0 + s_0 - \frac{\gamma}{\beta} (1 + \ln s_0)$$

Where i_{max} is the maximum proportion of infectives at a given time.

Furthermore, we can use this information to predict the total number of people who will catch the disease, which is also helpful in understanding the impact of an outbreak. This will be the number of people in the recovered group at the end of the outbreak. For this, we need two formulas. First:

$$r_{end} = -s_{end} + i_0 + s_0$$

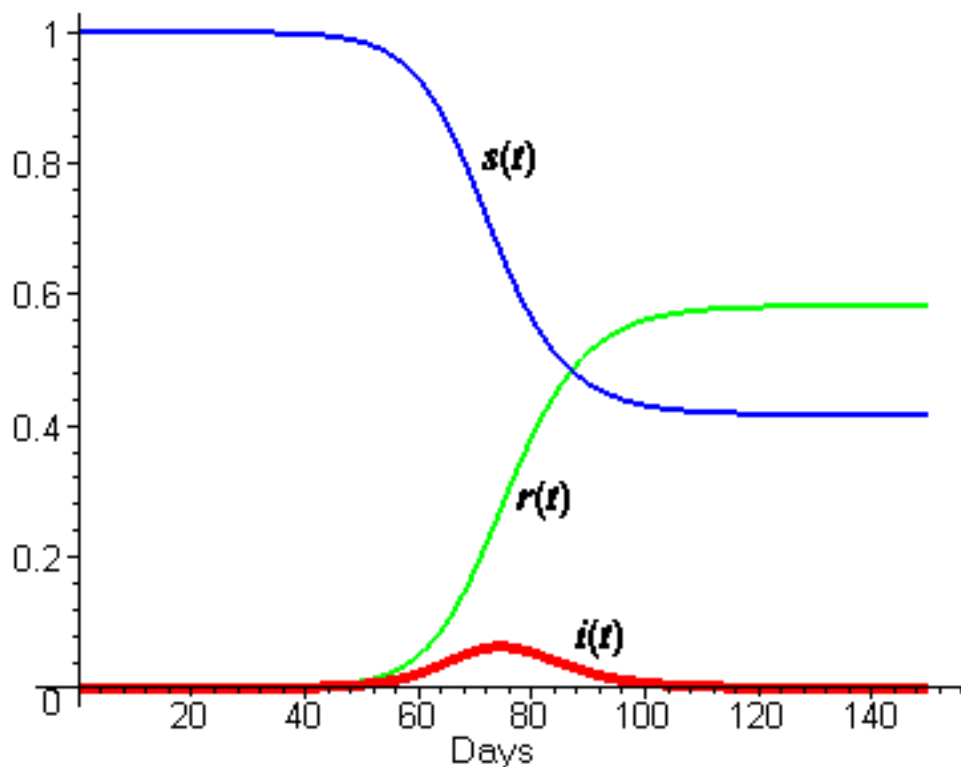
Where s_{end} and r_{end} are the proportions of susceptibles and recovered people at the end of the outbreak. We can also find s_{end} by using this formula from before:

$i + s_0 - \frac{\gamma}{\beta} \ln s = i_0 + s_0 - \frac{\gamma}{\beta} \ln s_0$. This gives us this formula:

$$s_{end} - \frac{\gamma}{\beta} \ln s_{end} = i_0 + s_0 - \frac{\gamma}{\beta} \ln s_0$$

From this, the value of s_{end} can be substituted into the formula for r_{end} to find its value.

As we can see, $\frac{\gamma}{\beta}$ repeatedly comes into these equations and is a key factor. Indeed, the reciprocal of this, $\frac{\beta}{\gamma}$ can be known as q , the contact ratio, which is the fraction of the population that an infective individual comes into contact with during the time that they are infective. For large values of q the max value of i increases and so does the number of people who will catch the disease. This shows us how useful measures such as social distancing are during an outbreak! This is because these methods “flatten the curve” as you will see in the graph below:



Although an improvement from the SI model, this model still leaves a lot to be desired, which I will discuss in the conclusion.

Conclusion

In this essay I have discussed two simple, compartmental models of disease spread (hopefully in a way that was easy to understand)!

Of course, as with all models, these two aren't perfect: they don't account for many things, such as those who are exposed to a disease, herd immunity and natural immunity, the introduction of vaccines or other precautions, or waning resistance over time. Furthermore, human behaviour is too erratic to predict with high accuracy.

However, that is not to say that they aren't very useful. These simple yet effective models allow us to plan ahead by predicting the impacts of disease outbreaks and plan necessary precautions, like the ones we had during the Covid-19 pandemic.

Initially, I had hoped to explain a third model: the SEIR model as well, however, I will refrain from this in order to stay within the word count, although I hope that I have interested you enough to research it yourself.

Hopefully I have managed to share my enthusiasm for this topic with you, and I thank you for reading!

References

https://en.wikipedia.org/wiki/Mathematical_model

<https://www.bmj.com/about-bmj/resources-readers/publications/epidemiology-uninitiated/1-with-epidemiology>

<https://nrich.maths.org/sites/default/files/thumbnails/content-id-14880-Using%20A%20Level%20Maths%20to%20Model%20the%20Spread%20of%20an%20Infectious%20Disease%20%282%29.pdf>

[https://math.libretexts.org/Bookshelves/Applied_Mathematics/Mathematical_Biology_\(Chasnov\)/04%3A_Infectious_Disease_Modeling/4.01%3A_The_SI_Model](https://math.libretexts.org/Bookshelves/Applied_Mathematics/Mathematical_Biology_(Chasnov)/04%3A_Infectious_Disease_Modeling/4.01%3A_The_SI_Model)

https://www.researchgate.net/figure/This-graph-shows-classic-logistic-growth-of-SI-epidemic-model-with-small-initial-number_fig8_320508832

<https://medium.com/@shaliniharkar/sir-model-for-spread-of-disease-the-differential-equation-model-7e441e8636ab>

<https://sites.math.duke.edu/education/ccp/materials/diffcalc/sir/sir2.html>

<https://www.youtube.com/watch?v=NKMHhm2Zbkw&t=1287s>

