Singapore International Mathematical and Computational Challenge 2022

Senior Challenge

Continuum Modelling of Infectious Disease Dynamics

There are many ways to model the spread of infectious diseases. Over a small spatial scale, a network-based infection model is often appropriate. However, when considering the spread of diseases over large communities, continuum modelling in the form of ordinary or partial differential equations is an attractive alternative. In this challenge, we will investigate the macroscopic modelling of disease spread using differential equations.

1.1 A Simple ODE Model

Let us start by considering the simplest model variant, where we can fully describe the disease progression by the following 3 numbers as functions of time:

- $S(t)$: The number of *susceptible* individuals in the population at time t.
- $I(t)$: The number of *infected* individuals in the population at time t.
- $R(t)$: The number of *recovered* (or removed) individuals in the population at time t.

For convenience of modelling via differential equations, we will assume that these numbers are real numbers, instead of natural numbers. One can also interpret them as some unnormalized density of individuals. The evolution of these numbers are coupled in the following way:

- $S \to I$: Each susceptible individual has some rate of contracting the disease (i.e. becomes infected) after being in contact with an infected individual. Let us call this rate β.
- $I \to R$: Each infected individual recovers at a certain rate. After recovery, the individual becomes immune to the disease and will no longer be infected even if they come into contact with infected individuals again. Let's call this recovery rate γ.

Now, we make the (unrealistic) assumption that every pair of individuals in the population has the same probability of coming into contact of each other. In other words, the system is *well-mixed* so that we only need to track the total number of each type of individuals. Then,

the above dynamics can be summarized by the following ordinary differential equations (ODE)

$$\begin{aligned} \frac{d}{dt}S(t) &= -\beta I(t)S(t), & S(0) &= S_0 \\ \frac{d}{dt}I(t) &= \beta I(t)S(t) - \gamma I(t) & I(0) &= I_0 \\ \frac{d}{dt}R(t) &= \gamma I(t) & R(0) &= 0 \end{aligned} \tag{1.1}$$

Here, $S_0 > 0$ and $I_0 > 0$ are the initial numbers of susceptible and infected individuals respectively. Let us also denote the initial population size by $N = S_0 + I_0$.

We now investigate various properties of model (1.1).

(a) Show that $S(t), I(t), R(t) > 0$ and $S(t) + I(t) + R(t) = N$ for all $t > 0$.

(b) We say that there is an epidemic if $I(t) > I_0$ for some $t > 0$. That is, the number of infected individuals increases. Derive condition(s) on S_0, I_0, β, γ for which an epidemic occurs. What is the outcome of the disease spread if an epidemic does not occur?

(c) Suppose now that an epidemic occurs. It is useful for policy makers to determine the maximum number of individuals that will become infected. This number is $I_{\max} = \max_{t>0} I(t)$. Derive an expression for $I_{\max}$ in terms of $N, S_0, I_0, \beta, \gamma$.

(d) We say that the disease dies out if $\lim_{t\to\infty} I(t) = 0$. Derive condition(s) on S_0, I_0, β, γ for the disease to die out. If the disease dies out, is it due to a lack of infected individuals, or a lack of susceptible individuals?

(e) Write a program to solve (1.1) for $S(t), I(t), R(t)$ as functions of t. Use your program to validate your analytical results in (a)-(d).

1.2 The Spatial Model

One aspect of disease spread not captured by model (1.1) is geographical effects. We now consider incorporating these effects in the model.

Let us consider M different spatial locations $[M] = \{0, 1, \ldots, M-1\}$. Let $S^i(t), I^i(t), R^i(t)$ denote the number of susceptible, infected and recovered individuals at location i respectively. We also write $\boldsymbol{S}(t) = (S^0(t), S^1(t), \ldots, S^{M-1}(t))^\top$ as the (column) vector of the number of susceptible individuals over all spatial locations. Similar notations will be used for infected ($\boldsymbol{I}$) and recovered ($\boldsymbol{R}$) individuals. As before, the dynamics of $\boldsymbol{R}$ is determined by those of $\boldsymbol{S}$ and $\boldsymbol{I}$, so we only need to model the latter's dynamics.

Let us consider a one dimensional model, where the M locations are arranged in a line, i.e. neighbours of location i with $0 < i < M-1$ are $i+1$ and $i-1$, and the two end points $i = 0, M-1$ only have one neighbour each. The spatial interactions are limited interactions with neighbour(s). The interaction is a simple movement: per unit time, each susceptible or infected individual at one location moves at a rate d to one of its neighbouring location(s).

Hence, the disease spread model becomes

$$\begin{aligned}\frac{d}{dt}S^i(t) &= -\beta I^i(t)S^i(t) + dD(\boldsymbol{S})^i \\ \frac{d}{dt}I^i(t) &= \beta I^i(t)S^i(t) - \gamma I^i(t) + dD(\boldsymbol{I})^i\end{aligned} \tag{1.2}$$

where

$$D(\boldsymbol{S}(t))^i = \begin{cases} -S^i(t) + S^{i+1}(t) & i = 0 \\ S^{i-1}(t) - 2S^i(t) + S^{i+1}(t) & 0 < i < M-1 \\ S^{i-1}(t) - S^i(t) & i = M-1 \end{cases} \tag{1.3}$$

and similarly for $D(\boldsymbol{I})$. We study the case where the initial susceptible population is uniformly distributed in space, i.e. $S^i(0) = S_0$ for all i. On the other hand, there is an initial population of infected individuals concentrated at $i = 0$ (a cluster). That is, we set $I^0(0) = I_0$ and $I^i(0) = 0$ for all $i \neq 0$.

(a) Extend your program from 1.1(e) to simulate the outcome of the disease in the spatial model with the following parameters: $\beta = 1/5, \gamma = 5, M = 100, d = 1, S_0 = 100, I_0 = 100$. Plot the spatial profiles (i.e, x-axis being i and y-axis being the value of $S^i(t)$) of $\boldsymbol{S}(t)$ and $\boldsymbol{I}(t)$ for t in the range $[0, 10]$.

A further continuum modelling technique is to treat space also as a continuum variable. Instead of considering the spatial locations $\{0, 1, \ldots, M-1\}$, we consider a continuum of spatial locations indicated by a real vector x. In the one-dimensional case, x is a real number. Then, the 2-variable functions $S(t, x)$ and $I(t, x)$ represent the number of susceptible and infected individuals at time t and spatial location x respectively. In this case, we have the following *partial differential equation* (PDE) model for the disease progression

$$\begin{aligned}\frac{\partial}{\partial t}S(t,x) &= -\beta I(t,x)S(t,x) + \delta\frac{\partial^2}{\partial x^2}S(t,x), \\ \frac{\partial}{\partial t}I(t,x) &= \beta I(t,x)S(t,x) - \gamma I(t,x) + \delta\frac{\partial^2}{\partial x^2}I(t,x),\end{aligned} \tag{1.4}$$

where δ is a positive real number.

(b) What is the relationship between d in (1.2) and δ in (1.4)?

(c) We say that a pandemic occurs if an initial disease cluster propagates towards uninfected areas as a waveform. See Figure 1.1. Derive conditions on $S_0, I_0, \beta, \gamma, \delta$ for which a pandemic *cannot* occur.

(d) Verify your solution with your computer program.

Now, suppose that instead of wandering aimlessly, the susceptible individuals become aware of an incoming pandemic wave from the negative x side, and hence begin moving forwards (positive x direction) in an attempt to escape this wave.

(e) Design a PDE model for this modified disease spreading problem.

(f) At what speed should the susceptible individuals travel for it to be possible to prevent a pandemic?

(g) Suppose a pandemic still occurs, and let us define the fraction of individuals who survive the pandemic uninfected as $\rho = S_-/S_0 = \lim_{t\to\infty} S(t,x)/S_0$ (Figure 1.1). Derive a relationship between ρ and the other parameters in your PDE model.

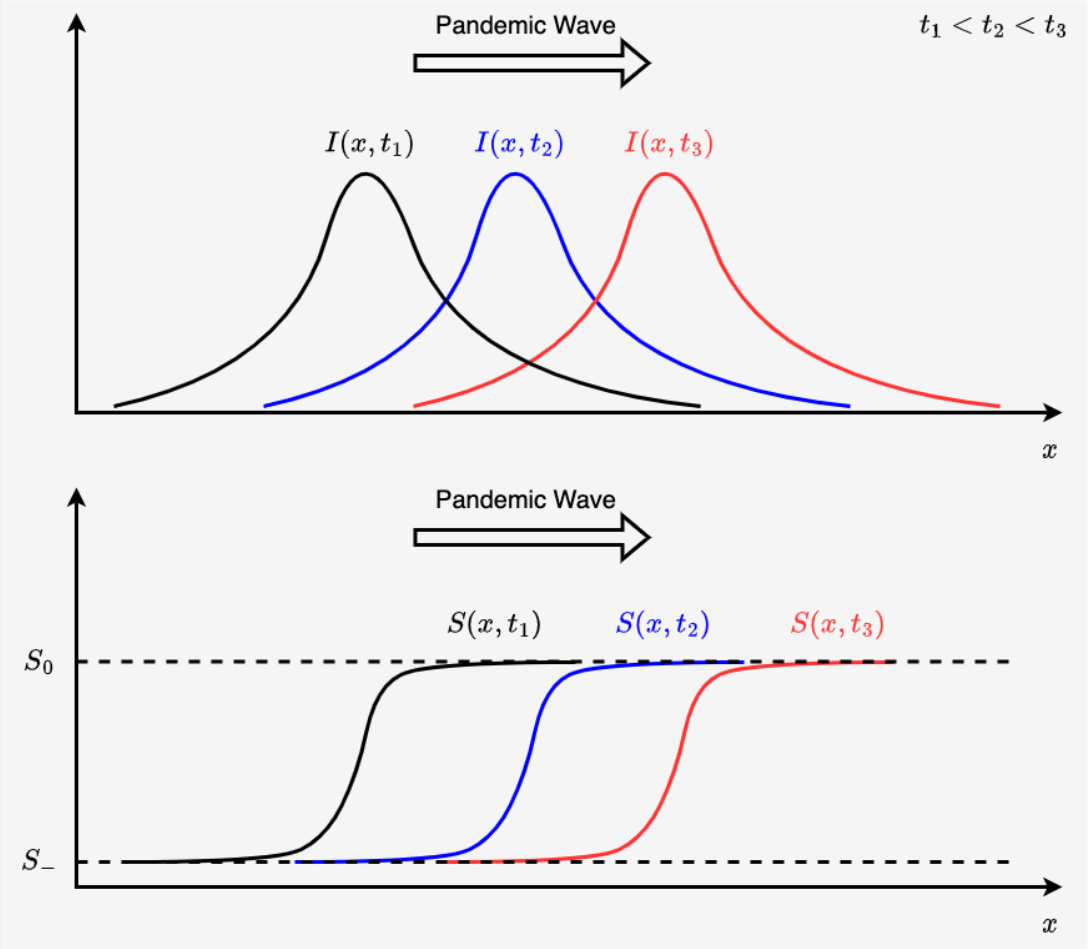

Figure 1.1: Pandemic wave. A pandemic occurs if there exists solutions to (1.4) where a waveform as shown above travels in the positive x direction as t increasees. Top plot: the waveform for infected individuals. Note that $I \to 0$ for $|x| \to \infty$. Bottom plot: the waveform for susceptible individuals. S_0 (the right limit) is the initial number of susceptible individuals and S_- is the number of remaining susceptible individuals after the pandemic wave passes.

1.3 Prediction of Disease Spreading Dynamics from Data

You are now given a dataset of the progress of the spread of a particular disease in a certain geographical area. The file structure is as follows:

```
data
├── locations.txt
├── data_year_1.txt
├── data_year_2.txt
└── data_year_3.txt
```

The file `locations.txt` contains a set of spatial locations of disease data collection points (in terms of scaled coordinates in the unit square $[0, 1] \times [0, 1]$).

The files `data_year_1.txt` and `data_year_2.txt` contains two years of historical data of the number of susceptible, infected and recovered individuals recorded in each data collection location. The last file `data_year_3.txt` contains partial data in year 3.

Your task is to predict the number of susceptible, infected and recovered individuals at each spatial location for the remaining days of year 3. You may use any combination of purely data-driven approaches, differential equation models, or any other methods, as long as you clearly explain your methodology. Credit will be equally distributed according to the following two criteria:

- the accuracy of your prediction measured against the hidden ground truth
- the soundness and creativeness of your methodology

Please submit your prediction as a file named `predicted_year_3.txt`, in exactly the same format as `data_year_1.txt`, and `data_year_2.txt`, i.e. you should copy the given part of `data_year_3.txt` into your prediction file. In addition, your report/presentation should detail your methodology and thought process.