MATH245 Project

Instructions

The aim of this project is to produce a written report on the topic outlined below.

Assessment weighting. This project is out of 100 marks and is worth 50% of the total mark for this module.

Submission. You must complete this assignment using R Markdown. Knit the final report to a pdf file, and submit the pdf file using the assessment portal on Moodle. **You should submit only the pdf file in which you will need to show your code and findings in a report as detailed below.**

Deadline for submission is **14:00 on Friday 28 April 2023.**

Late submissions. Reports submitted after the deadline, and without an agreed extension, will be penalised in line with the university regulations.

Length. The final document should be **approximately 11 pages long**. However, the bulk of the report should consist of R code, graphics and/or tables. As a rough guide, if you remove all the code, plots and tables, there should be around two pages of text remaining. Reports will be penalised for being overly lengthy.

Packages. You should *not* use R packages.

Plagiarism. This assignment is to be completed individually and must be your own work. It is possible to answer all questions using only the course materials provided for this module. However, if you wish to include content from any external sources, for example textbooks, websites, or computer packages, you may do so but these must be acknowledged and referenced. **To *copy* from a textbook, website or another student's work, and present it as if it were one's own is plagiarism and all forms of plagiarism are considered to be serious academic offences.** All submitted reports will be run through software to check for plagiarism.

Mark allocation

Presentation [10 marks]: For producing a well written, logically structured and clearly laid out document using R Markdown. Mathematics should be accurately typeset, R code, graphics and tables must be well integrated, and the report should not be long-winded. *A mark of zero for presentation will be awarded to those submissions which have not been knitted to pdf.*

Good programming practice [10 marks]: For including R code which demonstrates the principles of good programming practice that have been discussed during the module.

Initiative [10 marks]: For including in your report *relevant* material that goes beyond just providing answers to the questions below. Examples include providing additional background material, performing numerical investigations that go beyond those requested, or including mathematical analysis that goes beyond that requested. When including additional material, be mindful of the length constraints on the report.

Content [70 marks]: For answering the specific questions listed (the mark breakdown is specified below). You should indicate clearly in your report where you are answering each question. Content which answers a specific question, but has not been explicitly labelled as such, will not necessarily gain the credit. Where a question asks you to write a program, you must display the code. If a question asks you to plot a graph, you only need to give the graph (not the code used to generate the graph).

Introduction

This project is to develop epidemic simulations which evolve over time using a very general modelling technique known as compartmental models. A literature search will uncover the origin of such models dating back to the early 20th century, significant works include that of Ross in 1916, Ross and Hudson in 1917, Kermack and McKendrick in 1927 and Kendall in 1956.

For a given infectious disease e.g. COVID-19 these models may be used to estimate epidemiological parameters such as the reproductive number and also used to make predictions including: how a disease spreads; the duration of an epidemic; and the number infected individuals at any point in time. Furthermore simulations from these models can be used to show how various public health interventions may affect the outcome of the epidemic, such as the effects of vaccination programs or lockdowns.

Consider a population of N individuals where at time t each individual is assigned to a compartment. For instance take an SIR model with three compartments: susceptible (S), infectious (I) and recovered (R). The order of the model labels indicates the flow pattern between compartments; for example in SIR model individuals may progress through the compartments from susceptible, to infectious, then to recovered. In other words for a given disease an individual may be: susceptible to the disease (but not yet infected); infected with the disease; or recovered. Note that in this particular model the recovered compartment includes any fatalities from the disease. Over time individuals may progress from one compartment to the next until there are no more infected individuals in the population.

Compartmental models may have two or many compartments. For example in the SIS model infected individuals return to the susceptible compartment after being infected (as indicated by final S in the model name), this means that individuals can be reinfected. Another example is the SEIR model where the E compartment refers to an individual who has been exposed to the disease but is not yet infectious.

In the first instance consider an SIR model where at time $0 \leq t \leq T$ the triple $(x(t), y(t), z(t))$ denotes the proportion of a given population in the susceptible, infectious, and recovered compartments respectively. For example if at a given time there were 10 infected individuals and the population size was N then $y(t) = 10/N$; note that the population size N is constant over time. The model depends on two parameters β and γ and also initial values $(x(0), y(0), z(0)) = (x_0, y_0, 1 - (x_0 + y_0))$. The values of $(x(t), y(t), z(t))$ evolve according to the following ordinary differential equations (ODEs):

$$\frac{dx}{dt} = -\beta xy \tag{1}$$

$$\frac{dy}{dt} = \beta xy - \gamma y \tag{2}$$

$$\frac{dz}{dt} = \gamma y. \tag{3}$$

Numerical methods are required to solve this system of ODEs because in general it cannot be solved analytically.

Further mathematical details of the SIR epidemic model are given in Chapter 5 of your lecture notes on Moodle. Additionally compartmental models are described in many textbooks and online resources.

Part 1

Question 1 [3 marks]
State three assumptions underpinning the SIR model which make it unrealistic in terms of modelling infectious disease transmission in a human population. To gain full credit you need three valid assumptions.

Question 2 [2 marks] In terms of the SIR model given above it can be seen that $\frac{dx}{dt} + \frac{dy}{dt} + \frac{dz}{dt} = 0$ therefore what does $x(t) + y(t) + z(t)$ equal in this context? Following on explain why the three-dimensional system of

ODEs above is equivalent to the two-dimensional system:

$$\frac{dx}{dt} = -\beta xy$$
$$\frac{dy}{dt} = \beta xy - \gamma y$$

Question 3 [15 marks]
In Chapter 5 Section 3 of the lecture notes on Moodle, four methods are described for solving ODEs numerically. Choose the most appropriate numerical method (of the four) and use it to write a program to solve the SIR system of ODEs on the time interval $[0, T]$. You may re-use or modify any code from the lecture notes or previous assessments, provided you reference the source.

Your SIR model function should:

- have as inputs:
  - the model parameters β and γ;
  - the end time T;
  - the initial values $x_0, y_0 \in [0, 1]$ with $x_0 + y_0 \leq 1$;
  - the step size h that your integration method uses, note that $0 < h < T$ and also that h must have the same units of time as t and T.
- return a matrix with 4 columns in the following format:
  - column one $x(0), \ldots, x(T)$;
  - column two $y(0), \ldots, y(T)$;
  - column three $z(0), \ldots, z(T)$;
  - column four $0, h, 2h, \ldots, T$ i.e. a list of times.
  - Note that each row relates to the approximation of $x(t)$, $y(t)$ and $z(t)$ at a given time

a. State which numerical method for solving the ODEs your program uses and briefly explain why you chose this method.
b. With a population size of $N = 100$ use your code to solve the system of ODEs when $x(0) = 90/N$, $y(0) = 10/N$, $\beta = 0.6$, $\gamma = 0.2$, $h = 0.1$ and $T = 40$ days. Note that $h = 0.1$ represents one tenth of a day.
   - Display the output corresponding to $t = 0, 5, 10, \ldots, 40$ in a table with columns in the following order t, x, y, z (do not display all time steps from your output). Hint this is for visual inspection - for instance if your code is working correctly there should not be any NAs in this table, or columns of all zeros, etc.
   - Plot $x(t)$, $y(t)$ and $z(t)$ against t using all the time steps you computed.
   - In a couple of sentences interpret and comment on each of the quantities in your plots.

Question 4. [10 marks]

a. Plot $y(t)$ against t, for the parameters specified in Question 3b. Then on the same figure plot the solution to your model using all the same parameters except let $\beta = 0.1$ (instead of 0.6). Given these two solutions briefly state the key differences and an explanation for why you might expect to these differences.

b. During an infectious disease outbreak counts of new cases are often aggregated by day or week. Use the parameters from Question 3b and re-run your SIR model with $\beta = 5$ and a value of h equal to 1 week. Keeping everything in units of days plot your results, then in one or two sentences give an explanation of your findings.

c. When solving differential equations, such as the SIR model, finding general solutions analytically can be difficult and sometimes impossible. However it may be possible to find solutions for special cases, these can then be used to benchmark the numerical code to help ensure it is functioning correctly. In terms of the SIR model for all $t \geq 0$ consider the special case where $x(t) = 0$, then:

   - Solve the SIR ODE model by hand and in the context of the SIR model state appropriate constants of integration;

- Create a suitable plot for comparing your numerical solutions with the solutions you derived by hand, let $x(0) = 0$, $y(0) = 20/N$, $z(0) = 80/N$, $\beta = 0.6$, $\gamma = 0.2$, $h = 0.1$, population size $N = 100$ and $0 \leq t \leq 40$ (days);
- Comment on whether or not your solutions obtained by hand agree with your numerical solutions. If they do not agree explain why. One or two sentences will suffice.

Part 2

Introduction

In England during November 2021 the COVID-19 variant Omicron emerged, it is generally accepted that it contributed towards a large increase in the number of individuals with COVID-19 during December 2021: for the purposes of this project you should assume this is true. This data is on Moodle in the `omicron.csv` file, it was downloaded from the coronavirus.data.gov.uk website. In `omicron.csv` the `cases` column is the number of new infected individuals of COVID-19 recorded each day. Furthermore it is also useful to note that the Office for National Statistics population estimate for England from the 2021 census 2021 is 56,490,048.

In the SIR model $x(t)$, $y(t)$ and $z(t)$ are the total proportion of individuals in each compartment at any given time. In this modal infected individuals remain in the $y(t)$ compartment for several days. After being infected for several days individuals recover and move into the recovered compartment $z(t)$. It is important to realise that $y(t)$ is *not* the proportion of new infected individuals per time increment (h). If $h = 1$ day then the proportion of new infected individuals per day is $y(kh) - y((k-1)h)$ where $k = 1, 2, 3, ...$

Question 5 [20 marks]

In this question you should make use of the code you wrote in Question 3 which numerically integrates the SIR model.

a. Load the English COVID-19 case data from `omicron.csv` into your Rmarkdown document. Use the `cases` column to compute, for each day, the cases as a proportion of population size (in England). Plot the proportion and then in one or two sentences comment on your observations.
b. In this part the objective is to use your SIR model to estimate a suitable value of β given the COVID-19 data (from part a.). Let the first row in `omicron.csv` represent the initial point in time $t = 0$. At the initial point in time assume there are no individuals who have recovered and that the proportion of infected individuals is 0.004. Let $\gamma = 0.1$ and $h = 1$ (day). Consider the following values of $\beta = [0.1, 0.11, 0.12, ..., 0.2]$. Using the output from the SIR model, for each value of β compute the proportion of new infected individuals per day at every time point. Compare this proportion with the proportion of recorded cases per day (use the proportions from part a.). You may do this comparison by plotting your modelled outputs for different β values on the same figure as the recorded proportions (that were plotted in part a.). Hint: refer to the Part 2. Introduction for the definition of the "proportion of new infected individuals per day".
c. Using your answer from part b. state which value of β best represents the data. In one or two sentences briefly justify your choice of β.
d. Predict how the epidemic will evolve. To this end plot $x(t)$, $y(t)$ and $z(t)$ for $0 \leq t \leq 200$ (days) using the parameters from (Question 5) part b. along with your chosen value of β from part c.
e. Write a short paragraph on your findings from part d.

Part 3

Introduction

The objective here is to extend your SIR model to an SIRV model where 'V' refers to a vaccination compartment. As the name implies this model accounts for vaccination of the susceptible population. The SIRV model has the following system of differential equations:

$$\frac{dx}{dt} = -\beta xy - \mu x \tag{4}$$

$$\frac{dy}{dt} = \beta xy - \gamma y \tag{5}$$

$$\frac{dz}{dt} = \gamma y \tag{6}$$

$$\frac{dv}{dt} = \mu x \tag{7}$$

Here $x(t), y(t), z(z), v(t)$ each represent a proportion of the population. In this model vaccinated individuals are removed from the susceptible compartment $x(t)$ into the vaccinated compartment $v(t)$ at a rate defined by μ. Once vaccinated, individuals are considered immune so do not get infected, as a result they never enter the infected compartment $y(t)$. For the purposes of this project you should assume that the COVID-19 vaccine is effective and once an individual is in the vaccinated compartment they stay there indefinitely. As with the SIR model, once infected individuals recover they move into the recovered compartment $z(t)$. The parameters β and γ have the same meaning in this model as they do in the SIR model, likewise the population size N is constant.

Question 6 [20 marks]

a. Write code to numerically solve the SIRV model by extending your code from Question 3.
b. Plot $x(t)$, $y(t)$, $z(t)$ and $v(t)$ by using the parameters from Question 5. Use the value of β you chose in Question 5c. Let $T = 200$ and $\mu = 0.002$. Let the initial number of vaccinated individuals equal zero.
c. Repeat part b. except with a larger value of μ (you should choose a sensible value). In a couple of sentences state the effect of increasing the value of μ?
d. Using the results you have already computed, write a short paragraph comparing your SIR model from Question 5 with your SIRV model results from this question. You should focus your answer on the effects of vaccination.