Math 4MB, winter 2021, assignment 2

This assignment is due in the dropbox on Avenue to Learn by midnight (11:59 PM) on **Friday February 12**. You need to submit two files: (1) a **source file** (Jupyter notebook [`.ipynb`] or Rmarkdown [`.Rmd`] or Sweave [`.Rnw`]) and (2) an **output file** (PDF).

   This is a **group project**. One member of your group should submit a PDF and a source file (notebook or Rmarkdown or Sweave) in the dropbox; **all** members of your group (including the person who submits the main assignment) should separately put in the dropbox a **plain text** file giving a brief description of what each member of the group has contributed.

   Feel free to use web resources, but (1) report on any web resources you use significantly (you don't have to list resources that you look at but don't turn out to be useful); (2) maybe **don't** look at a complete solution to the problem below (i.e. a fully worked analysis of the SIRS equations).

   You can use a computer algebra system (CAS: Mathematica/Maple/sympy/etc.) if you like (again, let me know what you've used), but I usually find it easiest to try the calculations by hand first and then check them with the CAS. Similarly, I strongly recommend that you write all of your equations/derivations out by hand, **then** translate them into LaTeX.

1. The **linear chain trick** consists of breaking up the infectious (or other) compartment in an epidemiological model into multiple, serially linked compartments. By convention, the contact rate β is the same in all compartments and the rate of movement between sub-compartments is n times the original rate. For example, applying the linear chain trick to the SIR model would give

$$\frac{dS}{dt} = -\beta S \sum_{j=1}^{n} I_j$$

$$\frac{dI_1}{dt} = \beta S \sum_{j=1}^{n} I_j - n\gamma I_1$$

$$\frac{dI_2}{dt} = n\gamma I_1 - n\gamma I_2$$

$$\cdots$$

$$\frac{dI_n}{dt} = n\gamma I_{n-1} - n\gamma I_n$$

$$\frac{dR}{dt} = n\gamma I_n$$

   a. Show that the total infectious period does not change when we

    expand an infectious compartment with the linear chain trick (for arbitrary $n \geq 1$.

b. Show that $\mathcal{R}_0$ does not change (either argue verbally or use the integral definition as in Shaw and Kennedy (2021)).

c. "Lazy math": for $n = 4$ and $\gamma = 1$, you can draw one realization of the infectious period by choosing four independent random deviates from an exponential distribution with mean 1 and summing them (in R, `rexp()`; in Python, `scipy.stats.expon.rvs()`). Generate 1000 realizations of the distribution. Draw a histogram **where the area of the bins sums to 1** (in R, `hist(.,freq=FALSE)`; in Python, `matplotlib.pyplot.hist(., density=True)`) and overlay it with the probability density of a Gamma distribution with shape parameter equal to 4 and rate (or scale) parameter equal to 1 (R, `dgamma()`; Python, `scipy.stats.gamma`).

    2a. Reproduce Figure 2 from Shaw and Kennedy (2021) by coding the ODEs in R or Python and integrating them numerically. You don't need to get the colours exactly right.

    2b. Change the model **in some interesting way**. See what changes about the model. Explain whether (and why) $\mathcal{R}_0$ does or doesn't change for your modified model, and derive the new value of $\mathcal{R}_0$ if it does. A few ideas:

- add some rate of testing of non-symptomatic (i.e., pre- and asymptomatic) people to the model.
- add vaccination
- add a variant strain with a different transmission rate
- add behavioural responses, e.g. make one or more of the β values a decreasing function of I
- add an exposed class, i.e. an **non-infectious** (E) category that all infected people go through before going into the P or A compartment
- use the linear chain trick to change the distribution in one or more epidemiological compartments
- add compartments for people in the hospital and/or ICU

3. The file `great_plague.csv` contains a record of weekly deaths in the Great Plague of London in 1665. I **strongly** recommend that you deal with all of your times in units of weeks or weeks^{-1} to minimize confusion.

a. Fit a linear model to the log-death series as a function of time during the exponentially increasing phase of the epidemic to estimate r. Choose a subset of data that seems reasonable to you for this fit. Make sure to include the units of r in your answer! What is the doubling time implied by your estimate?

b. The generation time/infectious period for plague is *approximately* 2.5 weeks. Combine this information with your r estimate to get a rough estimate of contact rate β, removal rate γ, and $\mathcal{R}_0$ for plague in London in 1665.

c. The case fatality rate for plague (let's call it ϕ) was about 40% in the 17th century; the population of London was about 130,000. Using the following crude assumptions:

- ignore the time lag between infections and deaths and assuming therefore that the time series of plague deaths represents 40% of the current **incidence** (rate of new infections) of plague
- parameterize the incidence term in your model as β/NSI (this is likely to match well with the scaling from your linear regression)
- since the number of deaths in the first week is 1, set the initial $S(0) = N$ and the initial $I(0)$ such that $\phi\beta/NS(0)I(0) = 1$. (Yes, this means we will have $S + I$ slightly greater than N; if you want you can subtract $I(0)$ from N, but it will make barely any difference.)
- Numerically solve the ODE for your initial guess. Plot the time series of deaths and superimpose the curve of $\phi \cdot \beta/NSI$ from your solution. **Do not expect them to match exactly**; the point of all the work above was to get a solution that was at least on the right order of magnitude.

d. Adjust the values of β and γ by eye to get a better fit to the data.

- You will need to recalculate $I(0)$ each time to keep the initial number of deaths at 1/week.)
- You should probably ignore the time series after about week 40, when
- Keep in mind that r will control the initial slope, so you will probably want to keep r approximately constant as you modify β and γ.
- It is also allowed to shift the starting time of the epidemic forward (e.g. start the ODE solution at week 2) to skip over some of the initial noisy bits of the epidemics.

Present your final fit along with the values of β, γ and the values of r, $\mathcal{R}_0$, doubling time, and infectious period implied by your parameters.

NOT REQUIRED: if you are good at computers and estimation, you could refine your by-eye estimates by writing a program to adjust β and γ to minimize the sum of squared deviations between your trajectory and the data over some time window (e.g. using `optim()` in R or `scipy.optimize.curve_fit` in Python).

References

Shaw, Clara L., and David A. Kennedy. 2021. "What the Reproductive Number R0 Can and Cannot Tell Us About COVID-19 Dynamics." *Theoretical Population Biology*, January. https://doi.org/10.1016/j.tpb.2020.12.003.