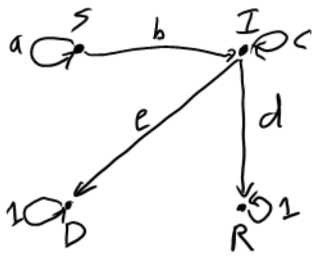
MAT 220—Project 5, Part II

Goal: Use the data for your country from April 1-30 to find the entries of a matrix to model the outbreak of COVID-19 there. Apply the model to estimate the behavior for May 1-10. Compare your results with the actual data for May 1-10.

Model: Use a Markov model. Let S represent the susceptible state, I the infected state, R the recovered state, and D the deceased state.

- Assume that recovered people stay recovered and that deceased people stay deceased.
- Assume that, each day, infected people either stay infected, recover, or die. (There still isn't data showing that surviving infection gives immunity, but it is probably a reasonable assumption when we are only modeling about a week and a half of the pandemic.)

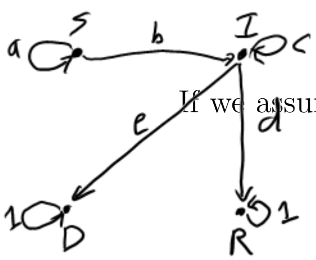
These assumptions give the following transition diagram and transition matrix:



$$A = \begin{bmatrix} a & 0 & 0 & 0 \\ b & c & 0 & 0 \\ 0 & d & 1 & 0 \\ 0 & e & 0 & 1 \end{bmatrix}$$

Let x_1 be the data on April 1 and x_2 be the data on April 2, with

$$x_1 = \begin{bmatrix} s_1 \\ i_1 \\ r_1 \\ d_1 \end{bmatrix} \text{ and } x_2 = \begin{bmatrix} s_2 \\ i_2 \\ r_2 \\ d_2 \end{bmatrix}.$$



If we assume that $Ax_1 = x_2$, then we get the following linear equations:

$$\begin{aligned} as_1 &= s_2 \\ bs_1 + ci_1 &= i_2 \\ di_1 + r_1 &= r_2 \\ ei_1 + d_1 &= d_2. \end{aligned}$$

In this situation, the components of x_1 and x_2 are numbers from our data. It is the a, b, c, d, and e that are the variables for which we need to find a solution. Generally, four equations is not enough to solve for five variables. However, the fact that A is a transition matrix also tells us that

$$\begin{aligned} a + b &= 1 \\ c + d + e &= 1. \end{aligned}$$

We could solve these equations with just two data points. However, this would make the model very susceptible to any noise in those two data points. As a first attempt, we will average our solutions for many data points to try to get rid of this noise in our model.

Assumptions about the data:

- Assume the population is constant, other than from deaths from COVID-19. In other words, $s_k + i_k + r_k + d_k$ should equal the same total population for all vectors x_k.
- Assume that the number of confirmed cases is the total number of cases so far. In terms of this model, that means that the number of confirmed cases is $i_k + r_k + d_k$ and s_k is equal to the total population minus the number of confirmed cases.

Turn in:

1. Solve for a, b, c, d, and e by hand in terms of x_1 and x_2. Plug your solution into the six equations to show that your solution is correct. (Hint: You will need to use your assumption that the population is constant, i.e., $s_1 + i_1 + r_1 + d_1 = s_2 + i_2 + r_2 + d_2$.) Show this work separately from your Jupyter notebook.

2. Import your data into your Jupyter notebook.

3. Use the data from April 1-30 to create a transition matrix for each step. In other words, let
$$A_k = \begin{bmatrix} a_k & 0 & 0 & 0 \\ b_k & c_k & 0 & 0 \\ 0 & d_k & 1 & 0 \\ 0 & e_k & 0 & 1 \end{bmatrix},$$
let x_k be the data for April k, and solve $A_k x_k = x_{k+1}$ for $k = 1, \ldots 29$.

4. For your model matrix A, take a to be the average of the a_k's, b to be the average of the b_k's, etc.

5. Use A and the data from April 30 to model the output for May 1-10. That is, take y_0 to be the data for April 30. Then calculate $y_{k+1} = Ay_k$ for $k = 0, \ldots, 9$, so that y_k is an estimate for the data for May k.

6. Plot the model and the actual data on the same set of axes.
   - Plot i_k, r_k, d_k, and $i_k + r_k + d_k$ (total number of confirmed cases) for both the model and the actual data. Don't plot s_k, since those values will be so large, it will be hard to see the other plots on the same scale.
   - Use coloring or shading or some other method to show the difference between the model and the actual data.

7. Discuss how well your model fits the data. If there is a discrepancy, try to explain why it might be occurring. Discuss whether it indicates that there might be errors in your model or errors in your data. How might you tell the difference? What adjustments might be helpful?

8. Go back to the Part I discussion area. Update your post (you can reply to your original one) to give your matrix A and your graph.

9. **Extra credit:** Repeat with fewer data points (maybe two, maybe fifteen) to get different models. Compare the results for the different models.