

MATH229 Project: Modelling the Spread of a Virus

Using the Stochastic Reed-Frost Model - A Simplistic Model

Instructions:

- Submit a report of 2-3 pages that addresses the 5 points below. You can work in pairs on this portion of the final.
- You have learned everything you need to answer the probability questions. So, I suggest not asking Google because it probably will confuse you with more complicated things than you need.
- You can create the report in RMarkdown and knit as Word or html before saving as a PDF or you can use another software. However note that you are to **SUBMIT a PDF version of the REPORT** along with the **RCode used for the simulation**.
- This assignment is based on the video: <https://www.youtube.com/watch?v=gSqIwXl6IjQ> . Watch this video before attempting the assignment for context.

Points to address in report:

1. Introduction to the Problem. What is COVID-19? Provide a very brief background into its history and its developments.
2. Describe the binomial distribution, how is the RV defined, what are the parameters, the pdf, the mean, variance?
3. Discuss the way the virus is spread. How can we use the binomial distribution to model this? What assumptions would you have to make for the simplest model as in the video?
4. Generate the following results from the simplest model in the video.
 - a. What is R_0 ? How is R_0 related to the parameters of the binomial distribution?
 - b. time 12:22 of video - How many people will get infected?
 - To answer this question, Use the values of R_0 from the video and simulate barplots to show the number of people who are infected at the different time points (use at least 100 time points). Compare your results for the different values of R_0 .
 - You should start with a susceptible population of 2000 and 2 infected persons.
 - **see the notes below for more hints to help with the simulation of the barplots.**
 - c. In the above simulation, we assume that the probability that a person gets infected at a time point T is $1 - q^{I_{T-1}}$. Why does this make sense?
 - **HINT:** p is the probability that a susceptible person gets the infection from an infected. So that, $q = 1 - p$ is the probability that the susceptible will not get the disease. I_{T-1} is the number of infected people at the time before.
 - **see notes below for more hints to help with this question**
 - d. Why do the curves for the barplots you simulated in question 4b eventually drop off (i.e. it does not keep increasing)?
 - e. Choose a bigger population to start. Notice the curve drops off really fast, even more so if you increase the infected population at the start. One reason is because of an assumption that was made in calculating the probability of being infected (see part 4c above). Can you say what this assumption is and how you would suggest fixing this problem? *This is a more challenging question and there are many ways to think of it, so don't be afraid to think out of the box and be creative.*

5. Provide a discussion of what you have learnt. INCLUDE:
 - a. Discuss why this model is too simple.
 - b. How can this model be modified to be more realistic? Discuss those talked about in the video and suggest 1 other modification you would make. Suggest how you would incorporate your modification into this model.

Some notes for Question 4 parts b and c - How many will get infected?

We have to set up this problem. Consider the following scenarios:

- a. Let us first start with a population of 2 people. One person is sick, the other is susceptible. what is the probability that the other person gets sick at the first time point?
- b. Stepping up: How about a population of 3 people with one sick, notice there are multiple scenarios here.
 - i. What is the probability that no one else gets sick at the first time point?
 - ii. What is the probability that one susceptible person gets sick at the first time point?
 - iii. What is the probability that the other 2 susceptible people get sick at the first time point? So, this probability of getting sick/infected can be calculated using the binomial distribution.

In other words, using part 4c above:

- (i) At time 0: $I_0 = 1$, $S_0 = 2$ and $A_0 = 0$. This means 1 Infected (I), 2 Susceptible (S) and 0 Recovered (A) at the initial time $T = 0$.
- (ii) At time 1: $I_1 \sim \text{binom}(S_0, 1 - q^{I_0})$, $S_1 = S_0 - I_1$ and $A_1 = A_0 + I_1$. This means that at time 1, we can simulate the number of people that will get infected from the total number of Susceptible, S_0 using a binomial distribution. All we need is the number of susceptible people, S_0 and the probability of getting infected $1 - q^{I_0}$.

Question:The probability that a person gets infected at time point 1 is $1 - q^{I_0}$. WHY? (HINT: p is the probability that a susceptible person gets the infection from an infected. So that, $q = 1 - p$ is the probability that the susceptible will not get the disease. I_0 is the initial number of infected individuals.)

Example of simulation: If there are 1000 susceptible people and the probability that a susceptible person will get the disease is 0.0015, then the following code simulated the number of people that would now be infected

```
## 1 means to simulate 1 time
## assuming 1000 susceptible people
## assuming probability of transmission is 0.0015
rbinom(1,1000,0.0015)
```

```
## [1] 1
```

- (iii). Then, similarly, at time $T + 1$, $I_{T+1} \sim \text{binom}(S_T, 1 - q^{I_T})$, $S_{T+1} = S_T - I_{T+1}$ and $A_{T+1} = A_T + I_T$

- a. So, The number of infected people at time $T + 1$ follows a binomial distribution depending on the number of susceptible the time before, S_T , and the updated probability of transmission which depends on the number of infected at time T .
- b. The number of susceptible people is now the number that was susceptible at the previous time T minus the number now infected I_{T+1}
- c. The new number people who had/will recover and be immune is A_{T+1} which is the previous infected plus the newly infected

Example of a recursive sequence simulation in R

Assume I have the following recursive sequence. You can edit this to write your own simulation code for the project.

Initial Values: $A_0 = 1, P_0 = 12, G_0 = 5$

Recursive Sequence:

$$P_{T+1} \sim \text{binom}(100, 0.01)$$

$$A_{T+1} = G_T + A_T + P_T$$

$$G_{T+1} = A_{T+1} + P_{T+1}$$

Basically, to get the Values of A, G, P at a certain time point, we need to know the result of the time point before. To do this in R, we will need to use your previous values of A, G and P to update the new values for the following time period. Here is an example of how I would do this:

R Code + Results:

```
##Creating vectors to store the results for each time point
A=c()
P=c()
G=c()

##Setting Initial Values
A[1]=1 ; P[1]=12 ; G[1]=5

##Doing Recursive Sequence
for (T in 1:50){
  P[T+1]=rbinom(1, 100,0.1)
  A[T+1]=G[T]+P[T]+A[T]
  G[T+1]=A[T+1]+P[T+1]
}

T=1:51 # I assumed 51 time points (inclusive of the initial time point 0).
barplot(G, names.arg=T,xlab="Time Points",ylab="G",main="Example!") #Edit the axis labels and title too
```

