ISYE - 6644 - Pandemic Flu Simulation

Team Members - Rajarshi Ghoshal, Sujoy Sadhan Bose, Suneet Taparia
Project Group No: 139

Abstract 1

Background 1

Problem Statement 1

Introducing Simulation Models 1

Results and Observations 2

(a) What is the distribution of the number of kids that Tommy infects on Day 1? 2

(b) What is the expected number of kids that Tommy infects on Day 1? 3

(c) What is the expected number of kids that are infected by Day 2? 3

(d) Simulate the number of kids that are infected on Days 1,2, ... Do this many times. What are the (estimated) expected numbers of kids that are infected by Day i, i = 1, 2, . . .? Produce a histogram detailing how long the "epidemic" will last. 3

Conclusion 5

References 5

Abstract

For understanding the Pandemic Flu Spread and longevity of the pandemic, it is important to understand the contact patterns of the population, the probability with which the infection spreads, the time for a body to recover from infection and the immunization achieved after the infection has been cured. Traditional theoretical epidemiology assumes that any individual can meet any other individual with equal probability. We use this assumption to develop our SIR (Susceptible-Infected-Recovered) model and check the spread of the infection. In this report, we estimate the number of days the pandemic/infection will last. The expected number of kids getting infected day on day.

We have used Excel and Python to simulate the Pandemic spread and answer the questions discussed in the report below. The assumptions considered in simulation are listed along with the explanation in respective sections.

To avoid the transmission of a contagious illness like the flu, it has been found that it is ideal for the affected person to isolate themselves from the rest of the cohort or social circle for a brief period. When we expanded or created holidays to lessen student interaction, this was clearly reflected.

Background

Infectious disease epidemics are complex phenomena that involve processes at various scales. The smallest and most rapid processes take place in the bodies of infected individuals as the pathogen enters and colonizes the host. The development of new treatments and the evolution of the pathogen and hosts are the biggest and slowest processes, respectively. The movements of infected hosts and the way government and healthcare systems react to an outbreak are intermediate steps between these. In this research, we consider the challenge of predicting an outbreak using (limited) understanding of the small- and intermediate-scale processes while treating evolutionary processes as constant.
An Alternative way to look at it is, with limited knowledge of the pathogen, the outbreak, and the ways in which human beings move and interact, the disease can spread. When a new pathogen or one that hasn't previously affected a population spreads, modelers are faced with challenges like these [1,2,3].

Problem Statement

In the problem statement we have a classroom of 21 school kids. 1 kid (Tommy) is infected on Day 0, and the other 20 kids are healthy. The infected kid walks in the class on Day 1 and starts interacting with the kids as a result of which infection can potentially spread. The probability at which the infection spreads is given by $p = 0.02$. A person remains sick for 3 days. Thus infected kids can infect the other kids on— Days 1, 2, and 3.
We need to find:
(a) What is the distribution of the number of kids that Tommy infects on Day 1?
(b) What is the expected number of kids that Tommy infects on Day 1?
(c) What is the expected number of kids that are infected by Day 2?
(d) Simulate the number of kids that are infected on Days 1,2,. . . . Do this many times. What are the (estimated) expected numbers of kids that are infected by Day i, i = 1, 2, . . .? Produce a histogram detailing how long the “epidemic” will last.

Introducing Simulation Models

We have built 2 models:

Model 1 - To answer the questions “*a and b”* we have leveraged MS Excel to answer the questions and try to back the answers with mathematical explanations. We have tested the spread with different infection spread probability values, which are captured in the report and attachments.

Model 2 - The “c and *d”* part of the problem statement is simulated using Python, wherein we simulate the infection spread for n^{th} day. The Model and simulation is based on the S I R[4] - Susceptible, Infected and Recovered compartmental model in epidemiology. The whole population would be in the three classes S , I and R. In this model we have assumed that once a person is recovered she / he will not contract the disease again.

Results and Observations

(a) What is the distribution of the number of kids that Tommy infects on Day 1?

With the information given in the problem, we understand that probability of infection spread is given by probability p = 0.02.

The number of kids infected in Bern(p) trials is given by Binomial Distribution. Hence, to get the number of children infected on Day1, we need to simulate Binomial Distribution.

The formula of Binomial Distribution is given by:

$$P_x = \binom{n}{x} p^x q^{n-x}$$

P = binomial probability

x = number of times for a specific outcome within n trials

$\binom{n}{x}$ = number of combinations

p = probability of success on a single trial

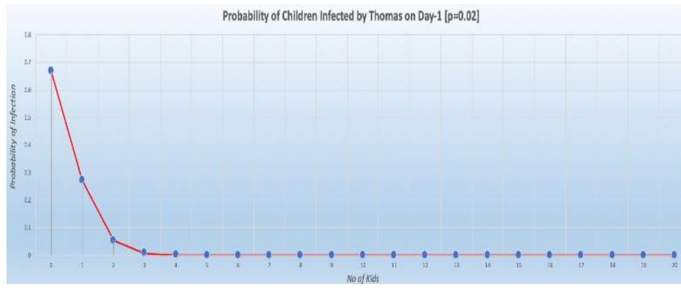
q = probability of failure on a single trial

n = number of trials

On simulating binomial distribution, we observe that the probability keeps decreasing for a higher number of kids getting infected on Day 1.

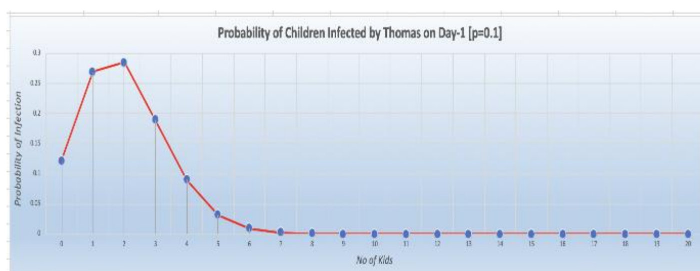
The number of kids Tommy Infects on Day 1 follows something like an exponential or a geometric distribution

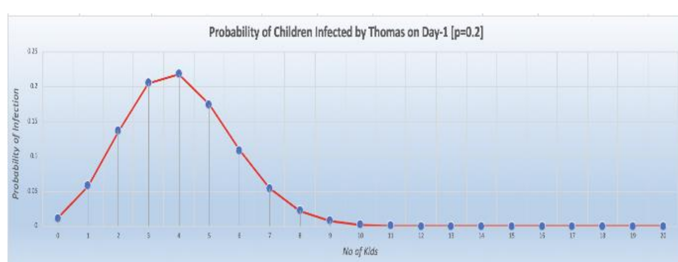
But since number of children infected is discrete, we can say that it follows **Geometric Distribution** (x axis – no. of kids, y axis – probability of infection)

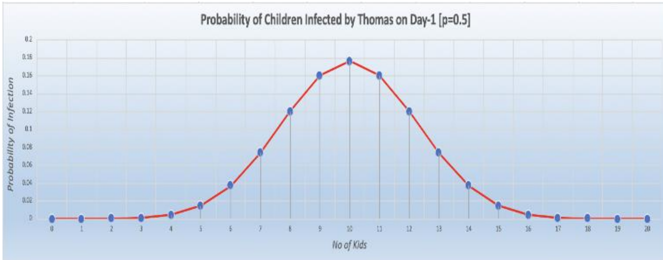


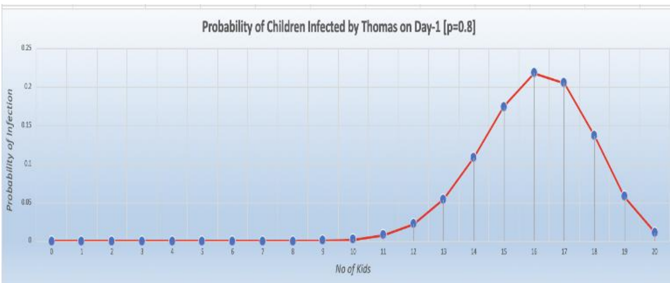
To gain better understanding and from an exploration point of view we also checked the same distribution for different p values.

We observe that, as the probability of infection spread increases, the graph looks more like a normal distribution, and higher the value of p, the more to the right, the graph shifts. This can also be interpreted as, the higher the value of p more kids would be infected on Day 1.









(b) What is the expected number of kids that Tommy infects on Day 1?

Total kids including Tommy = 21

In the attached excel file we observe that the number of infected kids by end of Day-1 (other than Tommy) is **0.4**

Total infected kids on Day 1 (out of 21) = 1 + 0.4 = **1.4**

This result can be verified mathematically using Expected Value of Binomial Distribution which is given by n*p. In this case we are dealing with n = 20 (Non infected kids at the start of Day-1) and p = 0.02.

So, E[X] = 20 * 0.02 = **0.4**

(c) What is the expected number of kids that are infected by Day 2?

On Day 2 out of 21 kids remaining Susceptible = 21 - 1.4 = 19.6

Independent Probability of number of kids infected on Day 2 = 19.6 * 0.02 = **0.392**

Expected number of kids infected on Day 2 including Tommy = 1 + 0.4 + 0.392 = **1.792**

The value for the same in the Python simulations (10,000 times) is coming to be ~**1.93**

(d) Simulate the number of kids that are infected on Days 1,2, ... Do this many times. What are the (estimated) expected numbers of kids that are infected by Day i, i = 1, 2, . . .? Produce a histogram detailing how long the "epidemic" will last.

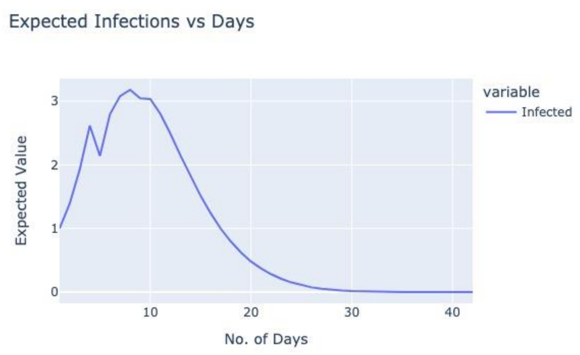
To achieve the results, we have created a model using the SIR concept of compartmental model in epidemiology. There are 2 approaches for the same – 1^{st} using the differential equation method[5] and 2^{nd} using simulation with Binomial Random Variables[6]. We have used the 2^{nd} approach to get the results. We have been able to run simulation for N = 10,000 for 3 different scenarios of school days.

Assumptions:

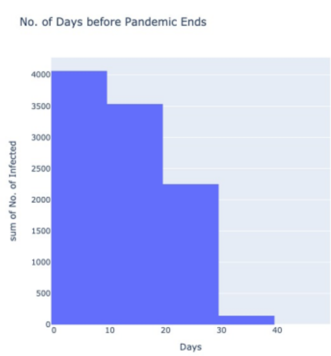
- Once a person has recovered from the infection, she / he will not contract the infection again and develops permanent immunity.
- Everyone interacts with everyone on daily basis
- Everyone interacts with everyone on daily basis
- Everyone interacts with everyone on daily basis

Scenario 1 – School has no off days 🙁

Expected number of kids who would get infected on days 1,2,3…

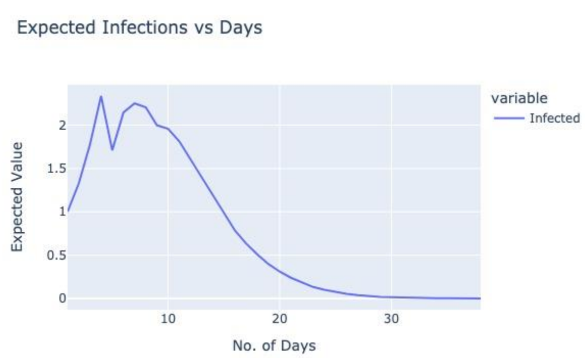


75 % of the simulations indicate that pandemic would last for less than 20 days.

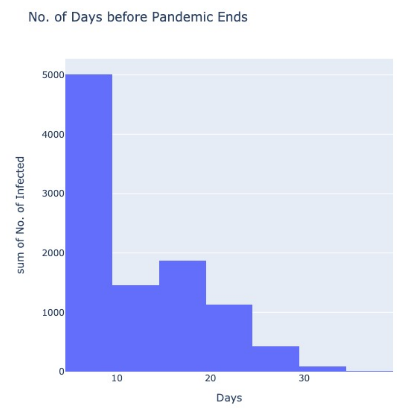


Scenario 2 – School has 1 day off 😐

Expected number of kids who would get infected on days 1,2,3…



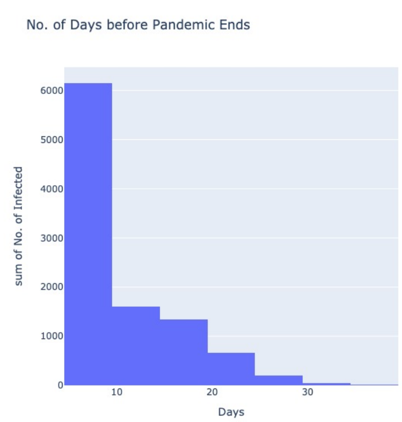
85 % of the simulations indicate that pandemic would last for less than 20 days.



Scenario 3 – School has 2 days off 😊

Expected number of kids who would get infected on days 1,2,3…

93 % of the simulations indicate that pandemic would last for less than 20 days.



Conclusion

We observed that within a given set constraints like limited population, social interactions, and proximity Flu spread can be contained.

However, this can be made more holistic by adding more random factors like any pre-existing medical conditions and strength of immunity of the cohort members and most of the interaction might happen in smaller sub-groups. This would imbibe the complexity of contact pattern[1].

From simulations above, it has been observed that if someone gets infected by a communicable disease like Flu, it's best that infected person stays away or isolated for a short time, from the rest of the cohort / social circle to prevent spread when the person can infect the other people. This is clearly captured when we introduced / increased the holidays to reduce the interaction of students.

References

1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4585889/

2. Anderson R. M. & May R. M. Infectious diseases of humans. (Oxford University Press, 1991). [Google Scholar]

3. Hethcote H. W. The mathematics of infectious diseases. SIAM Rev. 42, 599–653 (2000). [Google Scholar]

4. Compartmental models in epidemiology - https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology

5. The SIR Model for Spread of Disease - The Differential Equation Model https://www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-the-differential-equation-model

6. Almetwally, E.M., Dey, S. & Nadarajah, S. An Overview of Discrete Distributions in Modelling COVID-19 Data Sets. Sankhya A (2022). https://doi.org/10.1007/s13171-022-00291-6