Stochastic Pandemic Flu Spread in a Closed Community

Abstract

Mathematical models to chart disease transmission have dated as far back as the late 19th century to better understand and plan for pandemic events [1]. This python simulation takes a simplistic Bernoulli trial approach to model a pandemic flu spread in a classroom of 31 students. One student is initially sick and has the same likelihood to spread the flu to 30 fully susceptible other students each day. The simulation validated the mathematical expected values of infection rates in the classroom; an expected 0.6 students would be infected on day one and 1.55 students would be infected on day two. This scenario also assumes an infectious length of three days so the simulation also found that the duration of the epidemic was typically 10 days and provides scenarios where part of the population is immunized.

Background

Overview & Assumptions

In the given simulation, each of the 30 fully susceptible students have an equal chance of being infected by the flu with a probability of 0.02. Each infected student, starting with primary infected student nicknamed 'Tommy', will have a chance to infect each healthy student in the classroom. When a student is infected, the infectiousness period will last for the next three days, and on the fourth, for the intent of this simulation, the student will have fully recovered and will be considered immune. This simulation will also assume that all students and days are independent, that students attend school every day of the week, and that sick students do not stay home to recover or allow for time to disinfect the room.

Literature Review

Modern models are able to account for many more variables that can predict disease transmission with a higher degree of accuracy. Simpler models like this one can provide a baseline view of how quickly diseases can spread in a closed environment. It also removes a lot of variables like mixing patterns or variable susceptibility that can vary greatly depending on the community.

P.D. En'ko created a discrete-time deterministic model that used similar assumptions: all susceptibles are full susceptible; all immunes are fully immune; as healthy individuals become infected, they are removed from the susceptible population; these newly infected individuals will be infectious for a period of *t* days; and each interaction has the same probability of infection. This model by En'ko goes further by simulating variable contact rates as a percentage of the population. As a whole, this model, while deterministic, was compared with real data of measles outbreaks in boarding school girls from 1875 and 1888 which validated his model as a 'crude' but functioning model [1].

En'ko's model is one of many more simplistic models that can create broad generalizations for disease transmission. This model seeks to provide further insight into the distribution and expected infection rates using similarly broad generalizations while maintaining a limited real world applicability.

More complex models are able to calculate further data that increases the accuracy of modeling. This data can include the generation rate, the elapsed time between infections; the offspring distribution, the amount of secondary infections that are generated by a single individual accounting for dispersion; the infectiousness of a disease; and these value can be used to calculate the reproduction number, the average number of secondary cases produced by a single infected individual in a completely susceptible population [2].

While our model is simplified and limited, it can still be used to calculate the offspring distribution and the secondary infection rate. Our model uses 'Tommy' as the initial primary case, so the model will track all infections stemming from this initial case and create a functional basic transmission model for a closed community.

Approach

The simulation uses Python and NumPy for random number generation combined with NumPy's array features to improve computing time. The simulation starts by creating an array of students designated with the number '0'; students that are designated 0 are healthy while 1 – 3 for infectious with the corresponding number indicating the individual's current day of infection. If the student has an associated number of four or higher, then that student is considered immune for the purposes of this simulation.

An example starting class state would have the initial student 'Tommy' as infected with a number one while his peers are all healthy indicated by a zero as shown below.

[1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

Each day, each infected student from numbers 1 – 3 have will attempt to infect each healthy student designed with a number zero. During each infection attempt, a randomly generated number will be used to simulate the likelihood of that student spreading the flu. In the example below, Tommy has infected one other student after Day 1.

[2. 0. 0. 0. 1. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 1. 1. 1. 0. 0. 0. 0. 0.]

At the start of the second day, Tommy will be on his second day of infection while the newly infected student is on their first day. Both students will now attempt to infect the eight remaining healthy classmates.

Results

The simulation was conducted over 100,000 trials. Starting with Day 1, Tommy should mathematically infect the number of people following the expected value formula:

$$E(X) = \sum X * P(X)$$

$$E(X) = \sum 30 * 0.02 = 0.6$$

The numerical distribution of day 1 are shown in Table 1 and the histogram on Figure 1. With only a 0.02 chance of infecting each student, Tommy does not manage to infect any students in over 50,000 trials. Averaging the number of students he infects over these 100,000 simulations yields an average of 0.60278 students on Day 1. This corresponds to the Law of Large Numbers and validates the average mathematical expected value of 0.6 infected students per day.

Table Distribution of Students Infected by Tommy on Day 1

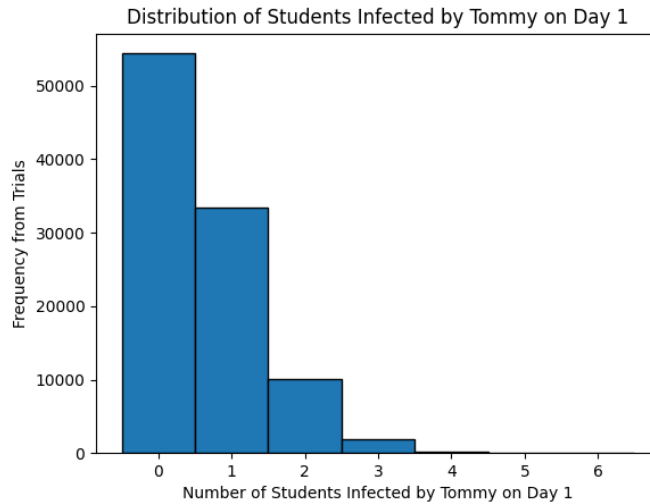| Number of Infections	Frequency
0	5.4680e+04
1	3.3198e+04
2	9.9460e+03
3	1.8960e+03
4	2.4200e+02
5	3.7000e+01
6	1.0000e+00

Table 1

Figure 1

Additional interpretation of Figure 1 shows that the graph doubles as a Day 1 offspring distribution for Tommy. This expected value of 0.6 infected students per day can be interpreted as a secondary infection rate of 0.6.

Analyzing Day 2, the expected number of students that will be infected can be predicted using the expected value formula again. The expected value can be calculated by the number of students likely to get sick on day 1 (30 * 0.02) multiplied by their chance to infect another student on Day 2 (0.02) and multiplied by the number of students who are expected to be still healthy on the second day. This is then added to the number of students that Tommy gets sick on Day 1 and Day 2. This is shown in the equation below:

$$E(X) = \sum 30 * 0.02 * 0.02 * (30 - 0.6) + 0.6 + 0.6 = 1.5528$$

By the end of the second day, the simulated expected value of sick students is 1.53062. This result also closely reflects our mathematically derived expected value due to the Law of Large Numbers. This pandemic is simulated over the course of 15 days and the number of expected infected students are shown in Figure 2.

Table Distribution of Infected
Students Over Each Day

Days	Estimated Infected Students
1	0.60
2	1.52
3	2.88
4	2.38
5	2.56
6	2.09
7	1.72
8	1.18
9	0.70
10	0.26
11	0.00
12	0.00
13	0.00

Table 2

Figure 2

From Figure 2, this epidemic will last no more than 10 days with the largest number of students being infected on Day 2. There are events where the pandemic continue after Day 10, but these are near zero probabilities. There is a steady decline on the number of students getting infected after Day 4 as a result of a smaller susceptible population. Totaling all of the expected sick students each day, it is estimated that approximately 13.33253 students of the 30 will have gotten sick over the course of this classroom pandemic.
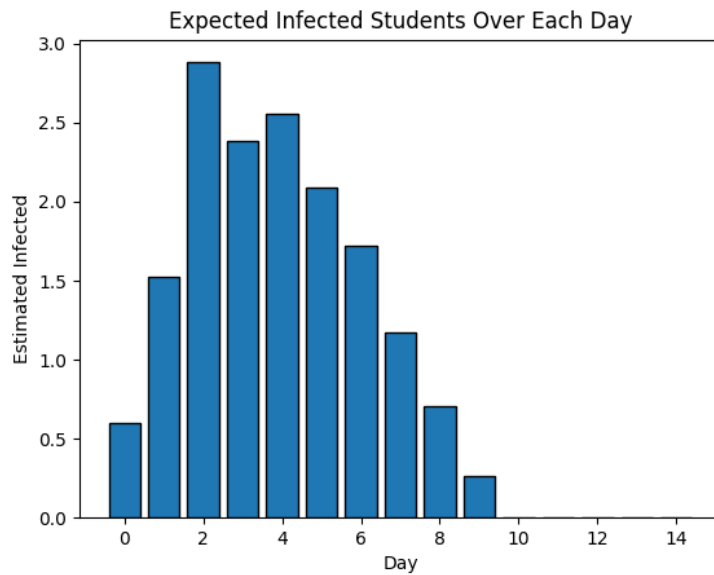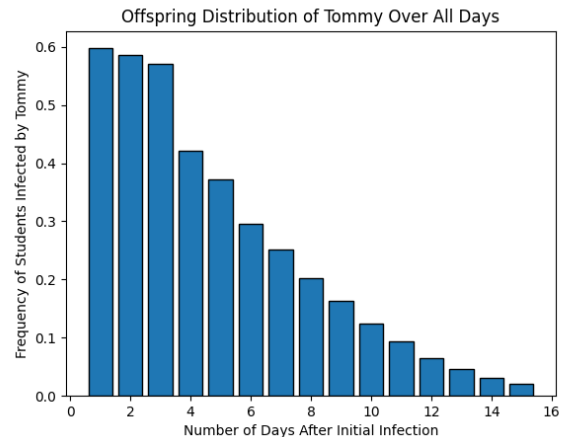

Figure 3

Case studies from real-world spread of SARS showed that initial pandemic situations with low primary cases resulted in an offspring distribution with a skewed binomial distribution [2]. If we plot Tommy's offspring distribution bar chart, we can see from a visual inspection that his secondary infection distribution follows a similar skewed binomial distribution as shown in Figure 3.

Immunization Scenario

This model can be used to simulate the spread of disease in the classroom if a certain percentage of students are immune. In this scenario, the odds of each student being immunized is 50-50. When

initializing the simulation, each of the 30 students will have a 50% chance of being assigned "4", thus rendering them immune. The results of the simulation are shown in Figure 4 and Figure 5.
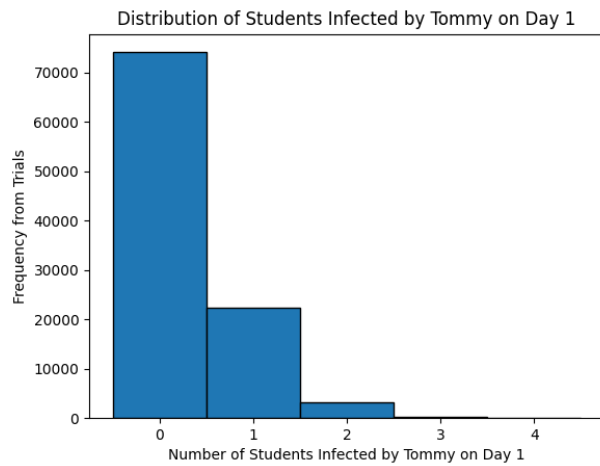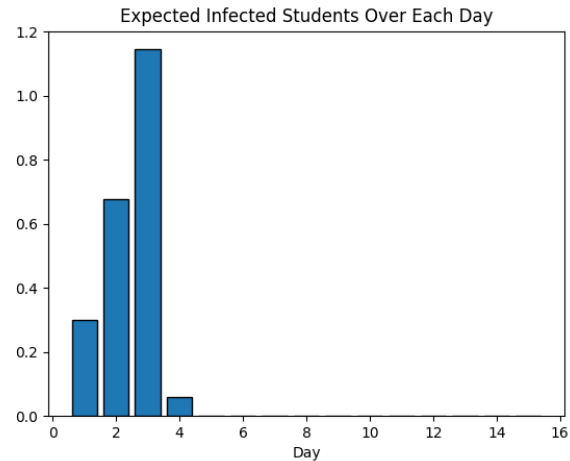

Figure 4

Figure 5

The simulated number of infected students on Day 1 with a 50% immunization rate is 0.2957. This is roughly half of the original expected value as the susceptible population has decreased by half. This simulated result is validated by the mathematical expected value of 0.3 as shown below.

$$E(X) = \sum 15 * 0.02 = 0.3$$

With the already decreased susceptible population, the number of expected infected student's peaks at 1.14 students on Day 3 and the pandemic already ends after Day 4. This could be representative of real world scenarios of herd immunity, where enough individuals are immunized and gives the community indirect protection against a disease as there aren't enough susceptible individuals to perpetuate the chain of infections needed to grow into a pandemic.

Conclusion

This simulation models disease transmission in a closed community. It does so with generalizations that attempt to keep the model relatively applicable in the real world but still does sacrifice a degree of accuracy. The number of estimated students that will be infected on a given day can be calculated mathematically and this model validates that with a simulation based approach.

There are a few areas that this model can be expanded upon. The secondary infections distribution from Tommy can be fit with actual distribution curves rather than through a visual inspection. This distribution can then be matched with real world diseases to get a more accurate representation of real world transmissibility. Separately, the scenarios can also be expanded to account for delays in infectiousness after the initial infection, weekend breaks and recoveries, and even variable infectiousness rates.

In the current state, the model is able to create a functional general model that predicts the number of students that become infected, and the length of the pandemic. It is also able to obtain the secondary infection rates and create an offspring distribution to better quantify the capabilities of this given flu.

Bibliography

[1] Foppa, Ivo M. *A Historical Introduction to Mathematical Modeling of Infectious Diseases : Seminal Papers in Epidemiology*. London, Academic Press, 2016. P. 22-30.

[2] Grassly, Nicholas C., and Christophe Fraser. "Mathematical Models of Infectious Disease Transmission." *Nature Reviews Microbiology*, vol. 6, no. 6, 13 May 2008, pp. 477–487, https://doi.org/10.1038/nrmicro1845.