Editor: Dianne P. O'Leary, oleary@cs.umd.edu

MODELS OF INFECTION: PERSON TO PERSON

By Dianne P. O'Leary

WHEN FACED WITH A SPREADING INFECTION, PUBLIC HEALTH WORKERS WANT TO PREDICT ITS PATH AND SEVERITY SO THEY CAN MAKE DECISIONS ABOUT VACCINATION strategies, quarantine policy, and the use of public health resources. This is true whether the pathogen's dispersion is natural (for example, the spread of influenza in 1918) or deliberate (for example, the spread of anthrax via terrorism). Effective mathematical models can help us test a public health policy's potential outcome and initiate an effective response.

In this problem, we focus on a simplified model of the spread of an infection and develop some tools that lend insight into its behavior. To make our problem as easy as possible, we impose some rather artificial assumptions. Suppose we have nm patients in a hospital ward and that their beds are arranged as n rows of m beds. For convenience, we'll let m be an even number. Suppose also that one of the patients, the one in bed $m/2$ in row $\lceil n/2 \rceil$, becomes infected and can infect any patient in a neighboring bed. How will this infection spread through the ward?

Insight through Monte Carlo Simulation

First, we'll need some model parameters. A patient, once infected, stays contagious for k days and then recovers, never to be infected again. For every day of infection, the probability that each susceptible neighbor (north, south, east, or west) becomes infected is τ.

This gives us three parts of the population to track. At day t, $I(t)$ is the infected proportion of the population; $S(t)$ is the proportion of the population that has never been infected. These quantities satisfy $0 \le I(t) \le 1$ and $0 \le S(t) \le 1$ for $t \ge 0$. We derive the third part, $R(t)$—the proportion of the population that was once infected but has now recovered—from the first two: $R(t) = 1 - I(t) - S(t)$.

We can use this model by running a simulation of it. Each patient is in one of $k + 2$ states: the patient has recovered from an infection, is susceptible to infection, or is in the ith day ($i = 1, \ldots, k$) of the k day infection. The integer values -1, 0, and 1, …, k are convenient for representing these different states. Each day, we update each infected patient's status by incrementing that patient's state; for each susceptible neighbor, we generate a random number between 0 and 1. If that number is less than τ, then the neighbor's state changes from 0 to 1, which indicates infection. We continue this process until there are no more infected patients; at that point, our model allows no possibility of any additional infections, so the epidemic ends.

Let's see how this model behaves.

Problem 1. Run the model for $m = n = 10$, $k = 4$, and $\tau = 0.2$ until there are no infected patients. Plot $I(t)$, $S(t)$, and $R(t)$ in a single graph. If possible, display the epidemic as a movie, where each pixel's color represents a patient's state.

The model is stochastic, so if we run it 10 times, we might get 10 different results, possibly ranging from no infections other than the original patient to infection of every patient. (These are both very low-probability events, however.) We must investigate the variation in results, but first let's add two complications.

The patients in our model are immobile and can only contact their four nearest neighbors. In most situations, the

In this issue's problem, we use Monte Carlo simulations and Markov models to gain insight into a simple model of the spread of an infection. In the solution section, we revisit the direction-of-arrival problem presented in the previous issue.

population would move in more arbitrary ways—for example, epidemics jump from continent to continent by air or ship travel. In our hospital ward, let's assume that the nursing staff sometimes moves patients to other beds. For definiteness, we'll assume that each patient initiates a swap with probability δ. Then, at each time and for each patient, we must decide whether that patient initiates a swap. If so, we'll choose the bed indices for the second patient randomly as $\lfloor r_2 n + 1 \rfloor, \lfloor r_2 m + 1 \rfloor$, where r_1 and r_2 are random samples from a uniform distribution on [0, 1].

Problem 2. Modify your model to include mobility and run it for $\delta = 0.01$ until there are no infected patients. Display the results as in Problem 1.

Two major tools slow the spread of epidemics: quarantine (to isolate infected individuals) and vaccination (to protect susceptible individuals). To reduce the number of infections in our hospital model, we should move the infected individuals to a corner of the ward, with recovered individuals separating them from susceptible ones whenever possible. You could experiment with this quarantine strategy, but in the next problem we turn our attention to vaccinations. For convenience, you could use the value –2 to indicate a vaccinated patient.

Problem 3. Suppose that each day, each susceptible individual has a probability ν of being vaccinated. Rerun your model with $\nu = 0.1$ and $\Delta = 0$ until no infected patients remain. Display the results as in Problem 1, and then compare the three models' results.

Now we need to see how much variation is possible in the results if we run the model multiple times.

Problem 4. Run Problem 3's model 1,000 times, recording the proportion of individuals who become infected in each run. (This is equal to the number of recovered individuals when the run is terminated.) Plot this data as a histogram, and then compute the mean number of recovered individuals and the variance in this number. Try several different values of ν to see whether the variance changes.

Tools

The 1918 influenza epidemic killed more than 20 million people. Investigators believe that it might have started on a US Army base, but the disaster had a worldwide impact, with millions killed in India alone. Travel of soldiers in World War I aided the spread of the infection. Gina Kolata chronicles these events in a recent book.[1]

When doing Monte Carlo experiments, it's wise to use a high-quality (pseudo-) random number generator to get valid results. Donald Knuth's book is the classic reference for understanding such programs.[2]

You can use Matlab's `fzero` function in Problem 5.

To go beyond the simple-minded models investigated in this article, read, for example, the books by Nick Britton[3] and Frank Hoppensteadt and Charles Peskin.[4]

You can learn more about Markov chain models and computing in William Stewart's book.[5]

Many approaches to aggregation of Markov chains exist. One starting point is an article by Ivo Marek.[6]

References

1. G. Bari Kolata, *Flu: The Story of the Great Influenza Pandemic of 1918 and the Search for the Virus that Caused It*, Farrar, Straus, and Giroux, 1999.
2. D. Knuth, *The Art of Computer Programming Volume 2: Seminumerical Algorithms*, Addison-Wesley, 1981.
3. N. Britton, *Essential Mathematical Biology*, Springer, 2003.
4. F.C. Hoppensteadt and C.S. Peskin, *Mathematics in Medicine and the Life Sciences*, Springer-Verlag, 1992.
5. W.J. Stewart, *Introduction to the Numerical Solution of Markov Chains*, Princeton Univ. Press, 1994.
6. I. Marek, "Iterative Aggregation/Disaggregation Methods for Computing Some Characteristics of Markov Chains. II. Fast Convergence," *Applied Numerical Mathematics*, vol. 45, no. 1, 2003, pp. 11–28.

From Problem 3's results, we see that vaccinations can contain an epidemic's spread. In Problem 5, let's take the role of a public health official trying to limit the spread.

Problem 5. Develop a vaccination strategy that will, on average, limit the epidemic to 20 percent of the population. Do this by using a nonlinear equation solver to solve the problem $R(\nu) - .2 = 0$, where $R(\nu)$ is the mean proportion of recovered individuals when we use a vaccination rate of ν. For each value of ν the solver presents, you will need to get a reliable estimate of R by running the model multiple times. Use Problem 4's variance estimates to determine how many runs to use, and then justify your choice.

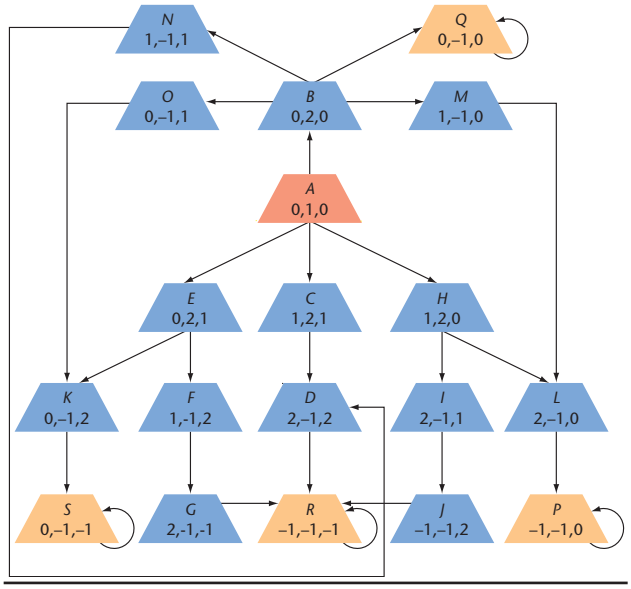

Figure 1. A Markov chain that models three patients in a row of beds, with the middle patient infected and able to infect two neighbors. The red state (*A*) is the state in which we start; the yellow states (*P, Q, R, S*) are the possible outcomes when the infection runs its course, corresponding to one, two, or three patients eventually infected.

A Markov Model

The model we just developed has the *Markov property*: each individual's status depends only on the population's status on the previous day, not on any older history. In fact, the system is a *Markov chain*. The states in the chain correspond to the population's possible statuses; we can label each state $(d_1, \ldots, d_p)$, where there are p beds and d_i ranges from –2 to k, indicating that individual i ($i = 1, \ldots, p$) is vaccinated, recovered, susceptible, or in day j ($1 \le j \le k$) of the infection. There is an edge from one state to a second state if it is possible for the population to move from the first state to the second on the next day; the weight on the edge is the probability of this happening.

Figure 1 illustrates a Markov chain corresponding to three individuals in a single row of beds, with the middle patient initially infected, a disease duration of two days, and no vaccination. (You will determine the edge weights in Problem 6.) For this model, we are interested in the probabilities that we terminate in state *Q*, corresponding to 33 percent of the population becoming infected; state *R*, corresponding to 100 percent; or states *P* or *S*, with the infection contained to 67 percent of the population.

Problem 6.

a. Construct the *transition matrix* A corresponding to this Markov chain: element $a_{i,j}$ is the probability of transitioning to state i from state j.

b. Let $\mathbf{e}_1$ be the column vector with 1 in position 1 and zeroes elsewhere. If we begin in day one in the first state, then vector $A\mathbf{e}_1$ tells us the probabilities of being in each of the states on day two. Prove this.

c. Similarly, $A^2\mathbf{e}_1$ gives the probabilities for day three. For efficiency, this should be computed as $A(A\mathbf{e}_1)$ rather than as $(A^2)\mathbf{e}_1$. Explain why, by doing the operations counts.

d. If we compute $\mathbf{z} = A^j\mathbf{e}_1$ for a large enough j, we will have the (exact) probabilities of being in each state after the epidemic passes. Use this fact to compute the probabilities of having one, two, or three infected individuals, and compare these probabilities with the results of a Monte Carlo experiment as performed in Problem 4 but using three individuals. How many Monte Carlo simulations does it take to get two digits of accuracy in the probabilities?

e. In this simple problem, you can determine the three probabilities directly from Figure 1 by determining the probability of a transition from state *A* to states *P*, *Q*, *R*, and *S*. Show how to derive these probabilities, giving the same answer as you obtained via the Markov chain computation in part d.

Our Markov chain has an enormous number of states (19, just to model three patients), but many of these states provide more detail than we might need. In Figure 1, for example, if we're in state *C*, we always make the transition to state *D*, so these two states can be combined or *aggregated* without a loss of information about infection totals. More importantly, states *P* and *S* represent different outcomes, but they are equivalent to us: in each case, 67 percent of the population becomes infected.

By aggregating states, we can reduce the problem's size. Sometimes we can do this analytically, but when the model is too complicated (for instance, once we add mobility), we can do it by simulation, gathering data to determine the probability of transitions between aggregated states.

These simple Markov models can yield some insight into epidemics, but we've seen that the work in doing a Monte Carlo experiment—or the number of states in the Markov chain—quickly grows with the population's size. In the next issue, we will investigate an alternative set of models.

Meanwhile, you might want to modify the models to explore more realistic variations. You also might consider how to model related systems, such as spreads of fungus on a tree farm, contamination in a set of chicken coops, or disease in a dormitory when the students also interact at school.