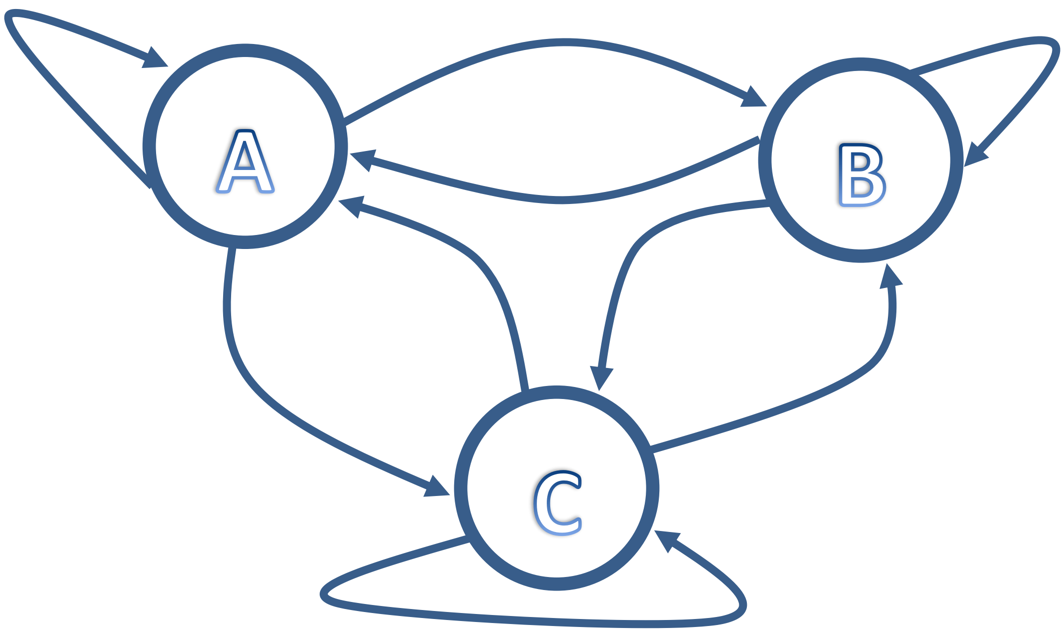
Markov Chains

A **Markov chain** is a model of change in a system with discrete conditions that change according to a probabilistic model in such a way that the future state of the system depends only on the present state and not what may have come before. Consider a graph of a three-state system below.

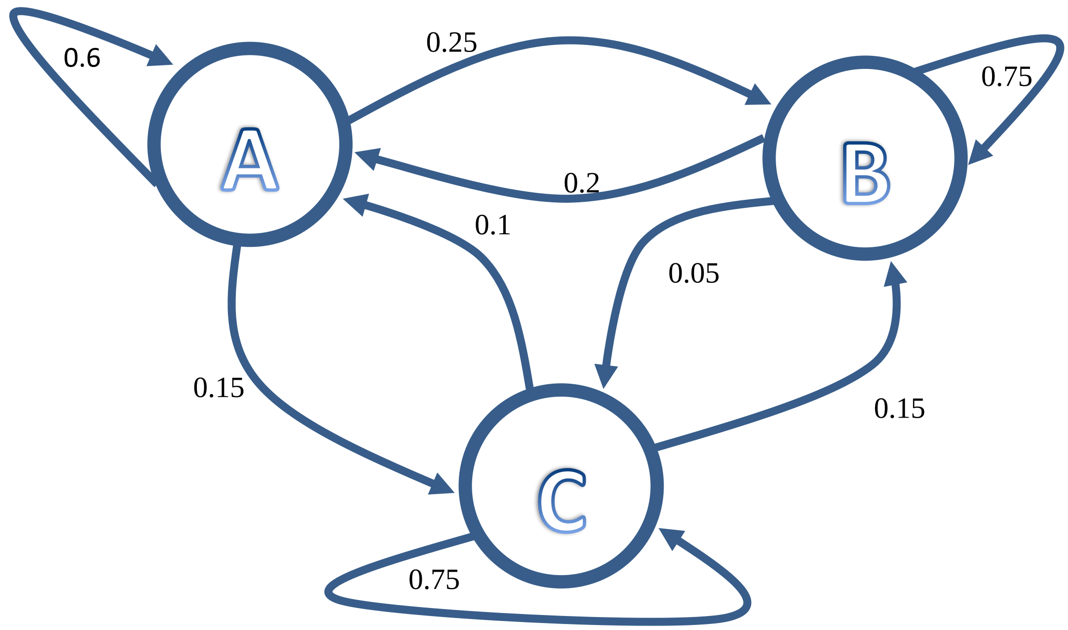


Each of the three states (A, B, C) has three arrows leaving it representing transitions from that state to each of the other states (including the possibility of remaining in the same state). If each of these transitions is associated with a certain probability, we have a Markov chain.

In linear algebra, we can deal with a discrete-time Markov chain, where the probabilities are associated with a jump in each time step (say a day, or second, or a year). The "chain" comes from the list of states that a particular element of interest might transition through in a particular sequence of time steps.

Suppose that the probability of moving from the A state to the B state is 25%, and the probability of moving from the A state to the C state is 15% (this leaves a 60% chance of remaining in the A state). Suppose that the probability of moving from the B state to the A state is 20%, and the probability of moving from the B state to the C state is 5% (leaving 75% chance of remaining in the B state). And lastly, suppose that there is a 10% chance of moving from the C state to the A state, and a 15% chance of moving from the C state to the B state (and a 75% chance of remaining in the C state). We can represent this system by a matrix:

$$P = \begin{bmatrix} .6 & .2 & .10 \\ .25 & .75 & .15 \\ .15 & .05 & .75 \end{bmatrix}$$



The columns of the matrix P represent the state the object starts in, and the rows are the states the object ends up in. Because the entries represent probabilities, the entries in each column must add up to 1.

While the Markov chain terminology comes from the behaviour of a specific object run through such a system, what we are interested in studying is the behaviour of the system. If we run hundreds of objects through the system and look at what happens to them in the long run, what happens? If the three states are three types of food eaten by caged rats at each feeding, we aren't so much interested in what one individual rat is doing, but the long-term behaviour of the population, especially if you are concerned about trying to plan for how much of each food type is required for the whole population, or at any point in time, given figures at a previous feeding cycle.

Consider our P matrix from before. Suppose that at a certain feeding time, we'll call this situation $\overrightarrow{x_k}$, there are 50% of the rats in the population eating from Feed C, 20% from Feed B, and 30% from Feed A. Organize these in a vector in the same order they were presented in the matrix. Thus $\overrightarrow{x_k} = \begin{bmatrix} .3 \\ .2 \\ .5 \end{bmatrix}$. These represent all the possible states, and the entries add to 1, accounting for the entire population. We can multiple this vector by the matrix to determine How much of the population will be eating at each feeding station at the next feeding cycle.

$$P\overrightarrow{x_k} = \begin{bmatrix} .6 & .2 & .10 \\ .25 & .75 & .15 \\ .15 & .05 & .75 \end{bmatrix} \begin{bmatrix} .3 \\ .2 \\ .5 \end{bmatrix} = \begin{bmatrix} .27 \\ .3 \\ .43 \end{bmatrix} = \overrightarrow{x_{k+1}}$$

We can continue this procedure to find the sequence of states that the system will pass through over time.

Since $P\overrightarrow{x_k} = \overrightarrow{x_{k+1}}$, and $P\overrightarrow{x_{k+1}} = P(P\overrightarrow{x_k}) = P^2\overrightarrow{x_k} = \overrightarrow{x_{k+2}}$, and so forth so that we can jump to any point in the sequence n steps in time later by multiplying by P^n: $P^n\overrightarrow{x_k} = \overrightarrow{x_{k+n}}$.

In our example:

$$\begin{bmatrix}.3\\.2\\.5\end{bmatrix}, \begin{bmatrix}.27\\.3\\.43\end{bmatrix}, \begin{bmatrix}.265\\.357\\.378\end{bmatrix}, \begin{bmatrix}.2682\\.3907\\.3411\end{bmatrix}, \begin{bmatrix}.27317\\.41124\\.31559\end{bmatrix}, \begin{bmatrix}.277709\\.424061\\.29823\end{bmatrix}, \begin{bmatrix}.2812606\\.4322075\\.2865319\end{bmatrix} \dots. \begin{bmatrix}.28940984\\.4472650252\\.2633251348\end{bmatrix} \dots \begin{bmatrix}.289473627\\.4473683284\\.2631580446\end{bmatrix} \dots$$

With each step, the changes in probabilities for objects in each state get smaller until all the vectors in the sequences start to look alike. Up to the tolerance of the calculator the system limit will converge to what is called the **equilibrium vector**. This is a vector $\vec{q}$ such that when you multiply the Markov chain matrix by the equilibrium vector, the equilibrium vector is the result. In other words:

$$P\vec{q} = \vec{q}$$

There are two methods for calculating the equilibrium vectors. One is technology intensive: raise the matrix P to some very large power. Because the initial condition doesn't matter in the long-run, all the long-run behaviour is contained in the matrix itself. The second can be more easily be done on paper, and that is to solve the $P\vec{q} = \vec{q}$ equation.

We'll illustrate both methods in the examples below.

Example 1. Find the equilibrium vector by raising the matrix P to a large power for the matrix P given above: $P = \begin{bmatrix} .6 & .2 & .10 \\ .25 & .75 & .15 \\ .15 & .05 & .75 \end{bmatrix}$. Put another way, we are finding $\lim_{k\to\infty} P^k$ numerically.

How quickly the matrix will converge to the equilibrium will depend on properties of the matrix, but I usually start with P^{60} and then square it (once or more after that) if need be. You'll know when you've reached the correct form of the matrix (subject to the tolerance of your calculator) when all the columns of the matrix are identical.

In the case of this matrix $P^{60} = \begin{bmatrix} .2894736842 & .2894736842 & .2894736842 \\ .4473684211 & .4473684211 & .447368421 \\ .2631578947 & .2631578947 & .2631578947 \end{bmatrix}$.

The matrix isn't quite equivalent to P^{∞}. There is a discrepancy in the second row, and if you try to convert these to fractions, they don't all convert. That's a sign that there are still differences in memory that are not displayed.

If I square this matrix and try to convert to fraction, it does work:

$$P^{120} \approx P^{\infty} = \begin{bmatrix} \frac{11}{38} & \frac{11}{38} & \frac{11}{38} \\ \frac{17}{38} & \frac{17}{38} & \frac{17}{38} \\ \frac{5}{19} & \frac{5}{19} & \frac{5}{19} \end{bmatrix}$$

Notice that all the column vectors are identical (and their decimal equivalents are what we saw above).

Let's check to see that the vector $\vec{q} = \begin{bmatrix} \frac{11}{38} \\ \frac{17}{38} \\ \frac{5}{19} \end{bmatrix}$ does satisfy the conditions of the equilibrium vector. If we multiply it by P, do we get the same vector back again?

$$P\vec{q} = \begin{bmatrix} .6 & .2 & .10 \\ .25 & .75 & .15 \\ .15 & .05 & .75 \end{bmatrix} \begin{bmatrix} 11/38 \\ 17/38 \\ 5/19 \end{bmatrix} = \begin{bmatrix} .6\left(\frac{11}{38}\right) + .2\left(\frac{17}{38}\right) + .1\left(\frac{5}{19}\right) \\ .25\left(\frac{11}{38}\right) + .75\left(\frac{17}{38}\right) + .15\left(\frac{5}{19}\right) \\ .15\left(\frac{11}{38}\right) + .05\left(\frac{17}{38}\right) + .75\left(\frac{5}{19}\right) \end{bmatrix} = \begin{bmatrix} 11/38 \\ 17/38 \\ 5/19 \end{bmatrix} = \vec{q}$$

So it does check out. One important thing to recall here is that although this high power of P is a method of calculating the equilibrium vector, it is not itself the vector that is sought. You'll need to pull the vector out of the matrix and label it appropriately or the method is not complete.

Example 2. Find the equilibrium vector by algebraically solving the equation $P\vec{q} = \vec{q}$ for the same matrix P.

We want to solve for $\vec{q}$ so we want to collect all the $\vec{q}$ terms on the left side of the equation.

$$P\vec{q} - \vec{q} = 0$$

Now we'd like to factor out the $\vec{q}$, but that would leave us with P-1 but this is poorly defined. We cannot subtract a scalar from a matrix, so we will rewrite the equation in an equivalent form we can proceed with.

$$P\vec{q} - I\vec{q} = 0$$

$$(P - I)\vec{q} = 0$$

The I matrix is the identity matrix, but now we can subtract one matrix from another, so we can proceed. Since the system is now a homogeneous equation, we are interested in the null space of the (P-I) matrix.

$$P-I=\begin{bmatrix} .6 & .2 & .10 \\ .25 & .75 & .15 \\ .15 & .05 & .75 \end{bmatrix}-\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}=\begin{bmatrix} -.4 & .2 & .1 \\ .25 & -.25 & .15 \\ .15 & .05 & -.25 \end{bmatrix}$$

It's not immediately obvious at this stage, but if an equilibrium vector exists, this matrix must have dependent vectors (in a 2x2 matrix, you will be able to see this straight away, but larger matrices will generally need to be row-reduced to see it).

$$\begin{bmatrix} -.4 & .2 & .1 \\ .25 & -.25 & .15 \\ .15 & .05 & -.25 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 & -1.1 \\ 0 & 1 & -1.7 \\ 0 & 0 & 0 \end{bmatrix}$$

So, we find from solving the equations $x_1 - 1.1x_3 = 0, x_2 - 1.7x_3 = 0$ we obtain the vector (by setting $x_3 = 1$) $\vec{x} = \begin{bmatrix} 1.1 \\ 1.7 \\ 1 \end{bmatrix}$. We still have one more step to obtain the vector $\vec{q}$: all the entries in the equilibrium vector must add up to 1 and these clearly don't. So, we'll have to scale the vector.

$$1.1 + 1.7 + 1 = 3.8$$

$$\vec{q} = \begin{bmatrix} \frac{1.1}{3.8} \\ \frac{1.7}{3.8} \\ \frac{1}{3.8} \end{bmatrix} = \begin{bmatrix} \frac{11}{38} \\ \frac{17}{38} \\ \frac{10}{38} \end{bmatrix} = \begin{bmatrix} 11/38 \\ 17/38 \\ 5/19 \end{bmatrix}$$

We can see that this is the same vector we found by the other method. Once we learn about eigenvectors, you'll see that there is a close relationship between this solution method and finding eigenvectors (in the case of the Markov chain, the equilibrium vectors also corresponds to the eigenvalue λ=1; the matrix may have other eigenvector/eigenvalues, but we are not interested in them in this instance).

How do we know that there is an equilibrium vector at all? Couldn't the system just keep changing forever?

There is theory to show to that under certain conditions one and only one equilibrium vector must exist. One requirement is that all states must be accessible from all other states. If some states are isolated, more than one equilibrium vector will exist, and which one you end up in will depend on where you start. Consider the system with 4 states:

$$P=\begin{bmatrix} .9 & .3 & 0 & 0 \\ .1 & .7 & 0 & 0 \\ 0 & 0 & .5 & .4 \\ 0 & 0 & .5 & .6 \end{bmatrix}$$

Notice that in this set of states, if you are in either state A or B, you will remain in A or B, you will never be in C or D since the probability of moving to those states is zero. Similarly, if you are in states C or D, there is no possibility of moving to states A or B. This system, if we solve it, will have two vectors as the basis for the null space, and any linear combination of those vectors will be an equilibrium state of the system (with the appropriate scaling factor). They need not be accessible in one step; it can be in multiple steps. Compare with the matrix below.

$$P = \begin{bmatrix} .9 & .2 & 0 & .09 \\ .1 & .7 & .05 & 0 \\ 0 & .1 & .5 & .31 \\ 0 & 0 & .45 & .6 \end{bmatrix}$$

In this matrix, if you start in state A, you can only go to state A or B, but if you end up in B, you can also end up in A, B or C; and if you end up in C, you can move to states B, C, or D; and if you end up in state D, you can move to A, C or D. So by starting in any state, you can pass through the whole sequence of states. It doesn't matter how unlikely it is, just that it is possible. This matrix has enough **communication** between states that there will be only one equilibrium vector.

One feature of Markov chain matrices that sometimes occurs is the possibility that one state is **absorbing**. An absorbing state is one where an object that moves into that state permanently remains in that state. Consider the Markov chain matrix below.

$$P = \begin{bmatrix} .8 & .2 & 0 \\ .19 & .78 & 0 \\ .01 & .02 & 1 \end{bmatrix}$$

In this matrix, state C is an absorbing state. The probability of getting to state C from either A or B is low, but once an object gets into state C, the probability of leaving it is zero. It may take a very long time computationally for this state to settle down to an equilibrium vector, but algebraically, it will be easy to show that it must be $\vec{q} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. (Before my TI-84 gave up on the decimal places, I needed to square the matrix about 13 times or about P^{8000}.)

Example 3. Suppose that we have a two-state system. In one state, people are living in the city, and the second state is people living in the suburbs. Suppose that the probability that in any given year someone living in the city moves to a suburb is 20%, and someone living in a suburb moves to the city is 8%. a) Give the Markov chain matrix that models this system. b) Suppose that in 2013, the percent of residents living in the city is 64% and the percent of residents living in the suburbs is 36%. What percentage of people will be living in the city in 2015? c) In the long term, what is the percentage of people who will be living in the city versus living in the suburbs?

a) The matrix for P is given by $P = \begin{bmatrix} .8 & .08 \\ .2 & .92 \end{bmatrix}$. We obtain these values by remembering that the values in each columns must add up to 1. Also, recall that the order left-to-right across the top, must match the order top-to-bottom down the side, and it must be the same in both columns. Changing the order will confuse the system and make it more difficult (or impossible) to solve.

$$\begin{array}{cc} & \begin{array}{cc} city & suburb \end{array} \\ \begin{array}{c} city \\ suburb \end{array} & \begin{bmatrix} .8 & .08 \\ .2 & .92 \end{bmatrix} \end{array}$$

b) The current state vector is $\overrightarrow{x_0} = \begin{bmatrix} .64 \\ .36 \end{bmatrix}$. This represents 2013, so 2015 is two states away. So we multiply by P twice to obtain the requested information.

$$P^2\overrightarrow{x_0} = \begin{bmatrix} .8 & .08 \\ .2 & .92 \end{bmatrix}^2 \begin{bmatrix} .64 \\ .36 \end{bmatrix} = \begin{bmatrix} .469376 \\ .530624 \end{bmatrix}$$

After two years, in 2015, 53% of the population will be living in the suburbs, and 47% will be living in the city.

c) We need the equilibrium vector to find the long-term percentage.

$$P - I = \begin{bmatrix} -.2 & .08 \\ .2 & -.08 \end{bmatrix}$$

We can see clearly these equations are linearly dependent. Reducing we obtain $x_1 = .4x_2$, and $\vec{x} = \begin{bmatrix} .4 \\ 1 \end{bmatrix}$, and scaling gives us $\vec{q} = \begin{bmatrix} \frac{.4}{1.4} \\ \frac{1}{1.4} \end{bmatrix} = \begin{bmatrix} \frac{2}{7} \\ \frac{5}{7} \end{bmatrix}$. So in the long-run, we can say that approximately 29% will live in the city, and approximately 71% will live in the suburbs.

Practice Problems.

1. Solve the equation $P\vec{q} = \vec{q}$ for the following Markov chain matrices. How many equilibrium vectors are there? Can you tell by looking at the matrix? (Why or why not?)
   a. $P = \begin{bmatrix} .7 & .4 \\ .3 & .6 \end{bmatrix}$
   b. $P = \begin{bmatrix} .5 & .04 \\ .5 & .96 \end{bmatrix}$
   c. $P = \begin{bmatrix} .99 & .08 \\ .01 & .92 \end{bmatrix}$
   d. $P = \begin{bmatrix} .4 & .3 & .1 \\ .3 & .5 & .2 \\ .3 & .2 & .7 \end{bmatrix}$
   e. $P = \begin{bmatrix} .88 & .03 & .15 \\ .08 & .75 & .21 \\ .04 & .22 & .64 \end{bmatrix}$
   f. $P = \begin{bmatrix} .9 & .002 & .12 \\ .099 & .95 & .004 \\ .001 & .048 & .876 \end{bmatrix}$
   g. $P = \begin{bmatrix} .8 & 0 & .5 \\ .1 & 1 & .1 \\ .1 & 0 & .4 \end{bmatrix}$

h. $P = \begin{bmatrix} .7 & 0 & 0 & .4 \\ 0 & .5 & .2 & 0 \\ 0 & .5 & .8 & 0 \\ .3 & 0 & 0 & .6 \end{bmatrix}$

i. $P = \begin{bmatrix} .7 & 0 & 0 & .4 \\ .2 & .5 & .1 & 0 \\ 0 & .5 & .1 & 0 \\ .1 & 0 & .8 & .6 \end{bmatrix}$

j. $P = \begin{bmatrix} .36 & .09 & 0 & .4 \\ 2 & .5 & .2 & .13 \\ .25 & .4 & .8 & 0 \\ .19 & .01 & 0 & .47 \end{bmatrix}$

2. For each of the following problems, draw the diagram for the system (as on page 1 and 2) and use that to set up the Markov chain matrix used to model the system, and the initial state vector. Then answer the question specific to each problem, before solving for the equilibrium vector. In each case, interpret the meaning of the equilibrium vector in the context of the problem.
   a. Suppose that in a certain class, a teacher wants to model the passing and failing rates as a Markov chain to see if additional office hours are required for the struggling students. Suppose that on the first quiz of the semester she finds that 75% of the class passed the first quiz, and 25% failed it. From past experience teaching the course, she knows that if a student passes a quiz, they have a 90% chance of passing the next quiz, but if they failed the quiz, they have a 60% chance of failing the next quiz again. What does she expect the passing rate to be on the third quiz? (Be careful with the subscripts here: how many time steps have passed?) What does she expect the long-term passing rate to be toward the end of the semester?
   b. Suppose that a researcher knows that for a group of test subjects in a driving simulation, 60% of the passengers put on their seat belt, but 40% did not. Suppose that the researcher observes that in the next round of trials, 99% of the people who used a seat belt on the last trial will do so again, and 50% of the people who did not use a seatbelt, will do so on the next trial. How many people will be using their seatbelts on the 15th trial? How many people does the researcher expect will be using their seatbelts by the end of a year-long study?
   c. Suppose that a laboratory has 100 mice in a maze divided up with three feeding stations labeled A, B and C. Suppose that the mice are all initially started at feeding station A. The probability that a mouse will move from feeding station A to feeding station B before the next feeding is 22%, and the probability that a mouse starting at feeding station A will move to station C is 12%. If a mouse starts at feeding station B to feeding station C in the next time step is 34% and the probability it will move from feeding station B to feeding station A is 25%. If a mouse starts at feeding station C, the probability it will move to feeding station B in the next time step is 35%, and the probability it will move to feeding station A is 11%. How many mice will be at each feeding station at the next feeding time? In the long run, about how many mice will appear at each feeding station after many weeks of letting them wander around the maze? (Round your answers to the nearest whole mouse. If your numbers are less than 100, where do you think the extra mouse is? If your number is bigger than 100, which feeding station is most likely a mouse short?)
   d. Suppose that there are three weather conditions available to us in the winter in the snow belt on the east side of Cleveland, Ohio: cold and sunny, cold and cloudy, and snow.

Suppose that a given day the weather is cold and sunny. From past data, we know that the probability that a cold and sunny day will stay cold and sunny is 35%, and the probability that it will become cold and cloudy is 45%. Suppose that the probability a cold and cloudy day will stay cold and cloudy is 70%, and the probability that it will lead to snow is 20%. Suppose that a snowy day will become a cold and sunny day with 25% probability, and that it will snow again with 40% probability. If Christmas is 5 days away, what is the probability of a white Christmas (that it will snow)? What is the long-term probability for each type of day over the whole winter?

e. Suppose that in a college-level linear algebra course, four grades are possible: A, B, C, and fail. On the first quiz of the semester, students get 30% A's, 35% B's, 25% C's and 10% a D or less. From previous classes, the instructor observes that students who receive an A on a quiz have a 70% chance of continuing to get an A, 15% will get a B, 10% will get a C and 5% will fail the next quiz. Students who get a B have a 25% chance of improving their grade, 50% of replicating it, and 15% of getting a C. Students who get a C on the quiz have a 10% chance of getting an A on the next quiz, 20% chance of getting a B, and a 20% of doing worse. Students who fail the first quiz have a 4% chance of getting an A on the next quiz, a 16% chance of getting a B, and 18% chance of getting a C. If the midterm grade will include the first 8 quizzes, what kind of grade distribution on the quiz right before the midterm does the instructor expect? What would she expect to see in the long-run?

f. Suppose that a human being has 4 states: healthy, ill, very ill, and dead. Suppose that we start out with all the population being studied is well. The probability of staying healthy if you are already healthy is 75%. The probability of becoming ill if you are healthy is 22%. The probability of becoming very ill 2.7%. The probability of becoming healthy if you are ill is 70%. The probability of remaining ill is 25%. The probability of becoming very ill is 4.5%. The probability of becoming healthy if you are very ill is 10%. The probability of becoming merely ill is 40%, and the probability of remaining very ill is 45%. The probability of remaining dead once you are dead is 100%. (Dead is an excellent example of an absorbing state.) Suppose that the next time the population is looked at is one month later. If 1000 people were in the original study, how many people are in each state at that time? Approximately how many time steps does it take for everyone to die? (When the sum of probabilities in the living states is less than 5×10^{-4}, you can expect everyone to be dead, or alternatively, when the percentage of the dead exceeds 0.9995.)