Math Explorer's Club: The Mouse, the Maze and the Markov Chain

Summer 2008

Contents

1 Introduction

The Math Explorer's Club is an NSF supported project aiming at developing materials and activites to give middle school and high school students an experience of more advanced topics in mathematics.

In this activity, we introduce and develop the notion of Markov chains, consolidating the student's grasp of (simple) *probability theory* and introducing two important *discrete mathematics concepts*: the relationship between *graphs* and *matrices*, and *recurrence relations*.

The material developed here is not difficult and with graphing calculators, the computations are not hard. A basic understanding of probability theory is assumed though. The reader might want to consider having a look at the Probability math explorer's write-up, for example.

We first introduce the example of a mouse in a maze and develop the idea of a *transition probability*. After playing with this toy model, the next section introduces the framework of Markov chains and their matrices showing how it makes it easier to deal with problems like that of the mouse in great generality. We then investigate the model and then provide a few applications to practice the learned skills.

2 The Mouse and the Maze

2.1 Transition Probabilities

Little Ricky is a mouse and she lives in a small maze where she likes to wander from room to room. The layout of her living space is depicted in figure 2.1.

Every thirty minutes or so she likes to change rooms and just moves to one of the rooms adjacent to the one she is in at that moment. For example, from room number 1, she might decide to move on to rooms number 2, 4 or 5, but she wouldn't go all the way to room 3. She is a lazy rodent after all.

A careful study of her habits shows that the room she decides to go to

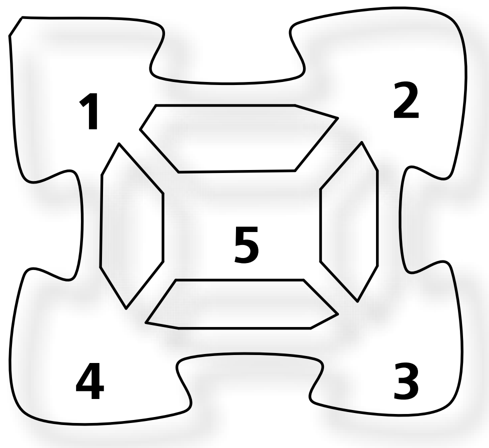


Figure 1: The Maze

next does not depend on any kind of history. She just seems to pick one the rooms she has access to (excluding her current location) randomly, where each accessible room is equally likely to be chosen.

In this setting, given an initial room i for the mouse to be in – i can be equal to 1, 2, 3, 4 or 5 – and for each other room j, there is a well defined number that represents the probability for the mouse to move from room i to room j.

Notice that this number does not depend on the time at which we evaluate that probability nor does it depend on what other rooms the mouse visited before that.

Definition If the mouse is inside room i, we will say it is in the state i. We then define the *transition probability* p_{ji} to be the probability for the mouse to move to room j if she starts in room i. For example, here $p_{21} = \frac{1}{3}$ being the probability for the mouse to go from room 1 to room 2: she has access to 3 rooms (2, 4 and 5) all of which are equally likely, hence the value $\frac{1}{3}$. In our example, she can't stay in the same room more than 30 minutes, so $p_1 1 = 0$ for example, meaning that the probability that she stays in room 1 starting in room 1 is 0.

Question Find all the transition probabilities p_{ji} for and i and j ranging from 1 to 5.

Question What can you say about the sum $p_{21} + p_{41} + p_{51}$? Is this a surprise? How about $p_{12} + p_{15} + p_{14}$? Can you explain these results?

Remember that if $\Omega = \{\omega_1, \cdots \omega_n\}$ is probability model, where the ω_i's are the possible future events or *outcomes*, and if p_i is the probability of the event ω_i occurring, then $p_1 + p_2 + \cdots + p_n =$. This just says that the sum of all probabilities is equal to on, as *something has to happen*.

For example the mouse deciding to move from room 1 to room 2 is one of the outcomes in the probability model describing the mouse in room 1, and it's probability is $\frac{1}{3}$ as calculated before.

2.2 Steps and Probabilities

Let us put Little Ricky back into room number 1 and watch what happens. After about 30 minutes, she moves to another room. This could be room 2, room 4 or room 5.

After yet another 30 minutes, she moves again to one of the adjacent rooms, and so on. For example, she might decide to move on to room 2 and then to room 3. But she might also want to come back to room 1 after she visits room 2.

Question Suppose Little Ricky starts in room 1 and then moves to room 2. What is the probability of her returning to room 1 next time she moves? Remember that we assumed that she moved from to room independently of any kind of history and always following the transition probabilities that we defined earlier.

Question Suppose she first moves to room 5, instead of 2. What is now the probability of her returning to room 1 at the next step?

Question Starting from room number 1 now. Can you calculate the probability of her returning to that room after 2 steps? How about after 3?

Let us try and solve this problem for 2 steps. To simplify the notation, we will write R_0 to denote the room she is in at the beginning of the experiment,

R_1 to denote the room she first moves to, an so on. In this language we are trying to calculate the probability of $\{R_0 = 1 \text{ and } R_2 = 1\}$. Call this event X

Now this event can be broken apart into 3 events depending on where she first moves to: $A = \{R_0 = 1 \text{ and } R_1 = 2 \text{ and } R_2 = 1\}$, $B = \{R_0 = 1 \text{ and } R_1 = 4 \text{ and } R_2 = 1\}$ and $C = \{R_0 = 1 \text{ and } R_1 = 5 \text{ and } R_2 = 1\}$. By broken apart we mean that in our probability model the event X is equivalent to A or B or C. Mathematically, this means:

$$X = A \cup B \cup C$$

In general, calculating the probability $P(A \cup B)$ of A or B happening is given by $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, where the last term is the probability of both A and B happening. But when A and B are *disjoint*, meaning they can not happen at the same time, this probability simply becomes $P(A) + P(B)$. In our case, the mouse can't be in both rooms 2 and 4 after the first step, for example. The events A, B and C and are all *mutually disjoint*, and thus we get:

$$P(X) = P(A) + P(B) + P(C)$$

While moving from one value to calculate to three may seem like making the problem only harder, it is the very opposite in reality. Let us look at $P(A)$, for example.

The event A occurs when Little Ricky first moves to room number 2, and then to room number 1 again. The probability of A occurring is thus the probability of her moving from room 1 to room 2 at the first move, and from room 2 to room 1 at the second one.

Now remember that we assumed that she moved from room to room independently of any kind of history. This means that her moving from room 1 to room 2 at first, and her moving from room 2 to room 1 at the second step are *two independent events*. Probability theory tells us that in a case like this, where Y and Z are two independent events, the probability of both Y and Z happening is simply the product of their probabilities, that is to say $P(Y \cap Z) = P(Y) \times P(Z)$.

Putting everything together, we get $P(A) = P(\text{moving from room 1 to room 2 at step 1}) \times P(\text{moving from room 2 to room 1 at step 2})$. Remembering that those probabilities do not depend on the time and are the transition probabilities p_{ij} this simplifies to:

$$P(A) = p_{12} \times p_{21}$$

Question Carry the same process and show that $P(B) = p_{14}p_{41}$ and that a similar expression holds for $P(C)$.

We have now shown that:

$$\begin{aligned} P(X) &= p_{12}p_{21} + p_{14}p_{41} + p_{15}p_{51} && (1) \\ &= p_{11}p_{11} + p_{12}p_{21} + p_{13}p_{31} + p_{14}p_{41} + p_{15}p_{51} && (2) \end{aligned}$$

You might now be able to guess the probability of Little Ricky ending up in room 2 after 2 steps, say. Indeed, just break that even apart depending on what room she first moves 2, then calculate the 3 probabilities using the same principles we used in the previous example.

Question Calculate the probability of Little Ricky ending up in room 2 after 2 steps starting from room 1.

Question How about the probability of ending up in room 2 starting from 1 after 3 steps? 4 steps? n steps?

While you should be able to carry on this analysis in theory, the calculations might quickly get out of control in practice. The framework we will introduce in the next section will help solve that problem and make a lot of these calculations much simpler, so that they can even ben carried on by a computer or graphing calculator. This is only a hint of why the Markov chain framework is so popular. Before we do that, we dedicate the next subsection to introducing another important concept.

2.3 States of Mice

Suppose we first flip a coin then put Little Ricky in room 1 or 2 depending on what the outcome is. For example, we put Little Ricky in room 1 if the coin lands on tails and in room 2 otherwise. One might wonder what the probability of Little Ricky ending up in room 5 after 1 step would be, and we will show to find that out. We assume that the coin is fair so that both side are equally likely to come out, with a probability of $\frac{1}{2}$, and that the coin and the mouse are "independent" of one another.

Let then X be the event "she ends up in 5 after 1 step", then following the same kind of analysis we did earlier, we can break this event apart into two events: $A = \{$she ends up in 5 starting from 1$\}$ and $B = \{$she ends up in 5 starting in 2$\}$.

Question Explain why $P(X) = P(A) + P(B)$.

Now the event A is equivalent to $\{$ She moves from 1 to 5$\}$ and $\{$The coin landed on tail$\}$.

Question Show that $P(A) = \frac{1}{2}p_{15}$. Remember that the mouse and the coin are independent.

Question Conclude that $P(X) = \frac{1}{2}p_{51} + \frac{1}{2}p_{52}$

Let us now assume that we roll an unfair 5-sided die such that the side i comes out the probability q_i. We do not assume the die to be fair, so we do not require $q_1 = q_2 = \cdots = q_5$. On the other hand, the q_i are probability values so they are all non-negative and satisfy $q_1 + q_2 + \cdots + q_5 = 1$.

Suppose we roll the die and drop Little Ricky in the room whose number comes out. In other words, she might be dropped in room 1 with probability q_1, or in room 2 with probability q_2, and so on.

Question What is the probability of her ending up in room 5 after 0 steps?

Question What is the probability of her ending up in room 5 after 1 step? To answer this question, break this event apart again depending on where she is dropped first. You should get

$$p_{51}q_1 + p_{52}q_2 + p_{53}q_3 + p_{54}q_4 + p_{55}q_5 \tag{3}$$

Question Can you do the same calculation for 2 steps? Use what we've calculated earlier.

3 Processes, Matrices and Markov Chains

The kind of process we described in the previous section is an example of a very general kind of processes called *Markov chains*. To describe such a Markov chain, we need the following data:

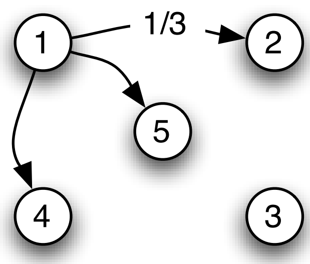
First, we have a set of *states*, $S = \{s_1, s_2, ..., s_r\}$. The process starts in one of these states and moves successively from one state to another, in the very same way our mouse moves from room to room in the maze. Each move is called a *step*.

If the chain is currently in the state s_i, then it moves to the state s_j at the next step with a probability denoted by p_{ji}, and this probability does not depend upon which states the chain was in before the current state. The probabilities p_{ji} are called, once again, *transition probabilities*. While we did not allow it in our first example, in general, the process can remain in the state it is in, and this occurs with probability p_{ii}.

An initial *probability distribution*, defined on S, specifies the starting state. Usually this is done by specifying a particular state as the starting state. More generally, this is given by the distribution of probabilities of starting in each of the states. For example, in the 5-state process describing our mouse, the distribution $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ would mean that the mouse is in the first room at the beginning of the experiment, while the state $\begin{pmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \\ q_5 \end{pmatrix}$ would represent the starting state of the mouse in the unfair 5-sided die experiment.

We can represent the process with a *graph* whose *vertices* (nodes) represent the different states the system can be in, while the arrows represent the possible steps. Over each arrow we write the transition probability associated with that step.

Question Going back to the mouse example, the following is part of the graph representing the mouse process. Can you complete it?



Such a graph makes it easy to visualize the process, but what really makes Markov chains a computationally efficient framework is the matrix representation.

We write the transition probabilities in an array as follows:

$$\begin{array}{c} \\ s_1 \\ \vdots \\ s_r \end{array} \begin{array}{c} \begin{array}{ccc} s_1 & \cdots & s_r \end{array} \\ \begin{pmatrix} p_{11} & \cdots & p_{1r} \\ \vdots & & \vdots \\ p_{r1} & \cdots & p_{rr} \end{pmatrix} \end{array}$$

This means that to find the transition probability from state i to state j, we look up the initial state on the *top line* and the new state on the *left*. The number at the intersection of that row and that column and the transition probability we're looking for.

Question Write the transition matrix for Little Ricky in the maze.

The power of this framework comes to light when one realizes that equations (3) and (1) are nothing but a rewriting of the *multiplication rule for vectors and matrices*. In summary if we rewrite what we showed in the previous section using this matrix language, we can prove the following theorem:

Theorem *Fundamental Theorem of Markov Chains*

If $P = (p_{ij})$ is the transition matrix for a Markov chain and $q = \begin{pmatrix} q_1 & \cdots & q_n \end{pmatrix}^t$ is a distribution of probabilities on the states of that chain then:

(i) The distribution of probability values on the states after 1 step starting on s is given by the matrix-vector product Pq where the i^{th} component of the new vector is given by:

$$(Pq)_i = \sum_{j=1}^{r} p_{ij} q_j = p_{i1} q_1 + \cdots p_{ir} q_r$$

.

(ii) The probability of ending up in state j after 2 steps starting from state i is given by $(P^2)_{ji}$ where P^2 denotes the product of the matrix P with itself. Explicetely:

$$(P^2)_{ji} = \sum_{k=1}^{r} p_{jk} p_{ki}$$

.

(ii') More generally, the transition probabilities after n steps are given by the transition matrix P^n, the n-fold product of the matrix P with itself.

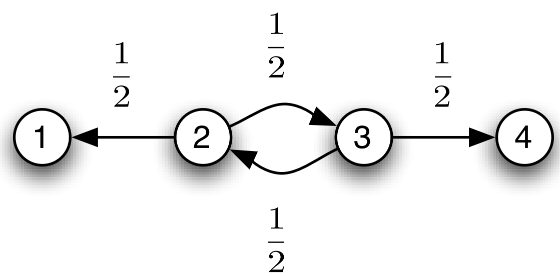
4 Will the Mouse ever find the exit? How Fast?

So far, given a process modeled as a Markov chain, we are able to calculate the various probabilities of jumping from one state to another in a certain given number of steps. This was given by taking successive powers of the transition matrix and reading a coefficient in the result matrix.

For example, we are to able to calculate the probability for Little Ricky to reach state 5 in 2 steps starting in state 1, or in 3 steps starting in 1, or even in 4 steps starting in 1.

What we are still lacking though, is the ability to answer questions like this: forgetting about how many steps it takes her to do it, what is the probability for Little Ricky to ever reach state 5 starting in state 1? While it might be possible to guess what the answer should be in this case, more complicated examples might require a new kind of approach and this is what we will focus in this section. Rather than giving a general statement we develop the idea using a simple example.

Let us create a new maze by describing it using the following graph:



The missing arrows indicate a zero transition probability. For example, state 1 and 4 don't have any arrows coming out them. We call such states *absorbing* states. This simply means once the process reaches one of those states, it stays in them forever. In this example, rooms 1 and 4 are two exits. Once Little Ricky leaves the maze, she stays "out" forever – or till the end of the experiment, that is. Non-absorping states are also called *transient*.

In the following, we show how to calculate the average time it takes Little Ricky to get to one of the exits (rooms 1 and 4) and the probability that she exists through room 4. More precisely:

Definition Define π_i to be the probability of ever reaching state 4 starting for state i, and τ_i to be the average number of steps required to get from state i to either state 4 or state 1. Finally, we will write $\{i \to j\}$ to denote the event {Little Ricky reaches state j starting from state i}.

4.1 Absorption Probabilities

Let us first focus on π_i's. Clearly, $\pi_1 = 0$ as she can't reach state 4 starting from state 1. Also $\tau_1 = 0$ as 1 is one of the target states: it takes Little Ricky no steps to get from 1 to either 1 or 4.

Question Find π_4 and τ_4.

Now suppose she starts in state 2. We break the event $\{2 \to 4\}$ apart depending on whether she first jumps to state 1 or 3. These two events being disjoints, we get that:

$$\begin{aligned} P(\{2 \to 4\}) &= P(\{2 \to 4\} \cap \{\text{She first moves to } 1\}) \\ &+ P(\{2 \to 4\} \cap \{\text{She first moves to } 3\}) \end{aligned}$$

We first simplify this into:

$$\begin{aligned} P(\{2 \to 4\}) &= P(\{1 \to 4\} \cap \{\text{She first moves to } 1\}) \\ &+ P(\{3 \to 4\} \cap \{\text{She first moves to } 3\}) \end{aligned}$$

Finally, using the transition probabilities and the independence of the transitions in a Markov chain, this becomes:

$$\begin{aligned} \pi_2 = p_{12}\pi_1 + p_{32}\pi_3 &= \frac{1}{2}\pi_1 + \frac{1}{2}\pi_3 \\ &= \frac{1}{2}\pi_3 \end{aligned}$$

Question Show that $\pi_3 = \frac{1}{2}\pi_2 + \frac{1}{2}\pi_4 = \frac{1}{2}\pi_2 + \frac{1}{2}$

Question Using the two previous identities, find π_2 and π_3.

Question Compare π_3 and p_{43}. Can you explain this result?

4.2 Average Time to Absorption

To calculate the average time to absorption we need to remember that in probability theory, the average of a *random variable* is defined as the *expectation* of that variable:

Definition The *expectation* of random variable X that can take values $1, 2, \cdots, k$ (k can be infinity, but mathematical care is needed to is defined the sum) as

$$\sum_{i=1}^{k} k \times P(X = k)$$

Getting back to our original problem, we want to calculate the average time to get to rooms 1 or 4.

Let us write $P(i \to E \text{ in } k)$ for the probability of reaching state 1 or 4 in k steps starting from state i and let us try to find an expression for the average time τ_2 it takes Little Ricky to find one of the exits starting in room 2. By the definition, we have:

$$\tau_2 = \sum_{k \geq 1} k \times P(2 \to E \text{ in } k)$$

Here again, we split those probabilities depending on whether the first step is a step towards state 1 or 3, and use the same kind of trick we used many times before:

$$\begin{aligned}
\tau_2 &= \sum_{k\geq 1} k p_{12} P(1 \to E \text{ in } k-1) \\
&+ \sum_{k\geq 1} k p_{32} P(3 \to E \text{ in } k-1) \\
&= p_{12} \sum_{k\geq 1} (k-1) P(1 \to E \text{ in } k-1) \\
&+ p_{32} \sum_{k\geq 1} (k-1) P(3 \to E \text{ in } k-1) \\
&+ p_{12} \sum_{k\geq 1} P(1 \to E \text{ in } k-1) \\
&+ p_{32} \sum_{k\geq 1} P(3 \to E \text{ in } k-1) \\
&= p_{12}\tau_1 + p_{32}\tau_3 + (p_{12} + p_{32}) \\
&= 1 + p_{12}\tau_1 + p_{32}\tau_3
\end{aligned}$$

Question Conclude that the following identity holds in our example:

$$\begin{aligned}\tau_2 &= 1 + \frac{1}{2}\tau_1 + \frac{1}{2}\tau_3 \\ &= 1 + \frac{1}{2}\tau_3\end{aligned}$$

Question Using a similar derivation or working by analogy, show that the following holds:

$$\begin{aligned}\tau_3 &= 1 + \frac{1}{2}\tau_2 + \frac{1}{2}\tau_4 \\ &= 1 + \frac{1}{2}\tau_2\end{aligned}$$

Question Putting everything together, calculate the values of τ_1, τ_2, τ_3 and τ_4.

In this section we tried to show how one can solve absorption problems explicitly when the graph of the Markov chain is small enough to make the calculations feasible by hand.

A more general treatment of this kind of problems requires tools from *linear algebra* beyond the scope of this introduction.

5 Other Examples and Applications

Following are examples and applications that have been collected from the internet, clipped, adjusted and/or rewritten to suit our needs and notations. Nothing is original in here, but the author believes that the situations described were interesting enough to be included here. All of these exercises are probably better carried on with the aid of a computer or a graphing calculator.

5.1 Amusement Park

An amusement park has five adult attractions: Roller coaster, Flume, Wheel, Smasher and Crunch.

Two of these, the exciting Roller coaster and Flume, are usually the last attractions visited before people leave the park.

There is a path between the Roller coaster and Flume, Roller coaster and Wheel, Roller coaster and Smasher, Flume and Wheel, Flume and Crunch, Smasher and Wheel, and Wheel and Crunch.

Question Assuming that people are equally likely to take any path leaving an attraction and that 25% of the people who ride the Roller coaster or the Flume then leave the park, draw the graph that shows the five attractions and the exit and shows the probability that someone at one vertex goes to any other vertex. Use E for exit, R for roller coaster and so on. Assume that someone who has exited the park does not return.

Question From this graph, write the transition probability matrix P for going from one attraction to another. The entries in column E would be 0 everywhere except at row E to indicate that someone who exits the park does not return ("stays out").

Question Use this matrix to determine the probability that a person who rides the Roller Coaster rides it once again after four changes of rides. To do this, you need to create an initial state vector q with zeroes everywhere expect at row R indicating you started on the Roller Coaster. Multiply P and q to determine the probability distribution for the next ride. Keep multiplying by P to get subsequent observations.

Question What is the probability that someone who is now riding the Roller Coaster rides the Wheel after changing rides four times?

Question Find the expected number of times that someone who is now riding the Roller Coaster rides the Wheel before leaving the park.

Question Find the expected number of rides that a person who is now riding the roller coaster rides on any ride whatsoever before leaving the park.

5.2 Genetics

A given plant species has *red*, *pink*, or *white* flowers according to the genotypes RR, RW, and WW respectively. If each genotype is crossed with a pink flowering plant (genotype RW), the transition matrix is as follows:

$$\begin{array}{c} \\ \text{Red} \\ \text{Pink} \\ \text{White} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Red} & \text{Pink} & \text{White} \end{array} \\ \begin{pmatrix} 0.5 & 0.25 & 0 \\ 0.5 & 0.5 & 0.5 \\ 0 & 0.25 & 0.5 \end{pmatrix} \end{array}$$

These numbers are determined as follows:

1. RR x RW yields RR, RW, RR, and RW.
2. RW x RW yields RR, RW, WR, WW.
3. WW x RW yields WR, WW, WR, and WW.

Assume the plants of each generation are crossed only with pink plants to produce the next generation.

Question This is the transition matrix is P. Now create a state vector q indicating an initial distribution of genotypes of 70% Red, 10% Pink, and 20% White.

Question Determine the distribution of phenotypes for the next generation starting from q. Multiply the answer with P and keep multiplying. Eventually the matrix product will *stabilize*. What does this show? What will be the eventual distribution of genotypes after many generations?

Question Suppose the initial distribution is different. Modify the values in vector q to indicate a different distribution (remember the three numbers must add up to 1) and repeat the multiplications as in the previous question. Explain your results.

Question Suppose we decide to cross plants of each generation with only white plants. Create a transition matrix for this situation. (To figure the probabilities, recall that RR x WW yields RW all the time. Calculate the other probabilities.) Use the same initial distribution of genotypes and determine the eventual distribution after many generations.

5.3 Predicting the Weather

Suppose a weather forecaster has collected data to predict whether tomorrow will be *sunny* (S), *cloudy* (C), or *rainy.* (R), given today's weather conditions. Here is a summary of his analysis:

- If today is sunny, then tomorrow's probabilities are 70% for S, 20% for C, and 10% for R.
- If today is cloudy, then the probabilities for tomorrow are 30% for S, 60% for C, and 10% for R.
- If today is cloudy, then the probabilities for tomorrow are 25% for S, 20% for C, and 55% for R.

Question With this information form a Markov chain describing the weather changes. Take as states the kinds of weather S, C, and R. From the above information determine the transition probabilities:

$$\begin{array}{c c} & \begin{array}{ccc} \text{S} & \text{C} & \text{R} \end{array} \\ \begin{array}{c} \text{S} \\ \text{C} \\ \text{R} \end{array} & \begin{pmatrix} ? & ? & ? \\ ? & ? & ? \\ ? & ? & ? \end{pmatrix} \end{array}$$

Question Start with today's weather and determine the probabilities for tomorrow's weather. Multiply by the transition matrix P again to determine the weather for 2 days from now. Repeat this so that you have a set of probabilities for the next 5 days.

Question Keep multiplying by P. Eventually the matrix product will stabilize. What does this show?

Question Change today's weather to indicate a rainy day. Repeat the multiplications as before to formulate a 5-day forecast.

Question Now forget about state vectors and just take P and raise it to some power, such as 20. What do you see? What does that tell you about the significance the state vector here? For what value of n does P^n

stabilize? Once it stabilizes it no longer forecasts the weather, but rather tells you about the *climate*.

Question Create a matrix that would be appropriate for Ithaca, NY. You may want to include snow as well as rain. Decide on today's weather and illustrate a 5-day forecast.

5.4 Predicting the Weather: The Land of Oz

According to Kemeny, Snell, and Thompson, the Land of Oz is blessed by many things, but not by good weather. They never have two nice days in a row. If they have a nice day, they are just as likely to have snow as rain the next day. If they have snow or rain, they have an even chance of having the same the next day. If there is change from snow or rain, only half of the time is this a change to a nice day.

Question With this information form a Markov chain describing the weather changes. Take as states the kinds of weather R, N, and S. From the above information determine the transition probabilities:

$$\begin{array}{c} \\ \text{R} \\ \text{N} \\ \text{S} \end{array} \begin{array}{c} \begin{array}{ccc} \text{R} & \text{N} & \text{S} \end{array} \\ \begin{pmatrix} ? & ? & ? \\ ? & ? & ? \\ ? & ? & ? \end{pmatrix} \end{array}$$

Question Study the climate of the Land of Oz. If you don't understand this question, make sure you read the previous example.