Markov Matrix Model

Terri Donovan

recorded: January, 2010

Today we are going to use a spreadsheet to model how a pixel can change in state from one time period to the next.

A spreadsheet is really a good way to think about this, because a spreadsheet itself is a bunch of pixels. Every cell in a spreadsheet has an address. That address identifies its column number and its row number. Because of that, we know not only what the address of a particular cell is, but we also know where that cell is in relationship to every single other cell on the spreadsheet.

What Tony talked about was when you are thinking about large ecosystems - it's useful to think about first, are the processes inside one pixel of most importance - or second, are the dynamics that are happening around a pixel more important for you to consider first?

Today our task is to think about the dynamics that occur within one pixel. We are going to use an example that Tony talked about in class and we are going to assume that a single pixel can have one of three states: Jack Pine, White Pine, or Spruce.

Let's go ahead and enter 'Ecosystem State' in cell A1. We are going to let 'J' be equal to Jack Pine, 'W' be equal to White Pine, and 'S' be equal to Spruce.

And what we are going to do is set up a model in which we track the dynamics of one pixel through time. We are going to start that with year zero. So let's enter the number 0 in cell B10. Our time-step for this model will be a 10-year time-step. And we'll go ahead and project this model out through the year 120.

Start with a 'State' for our pixel. And we are going to let that initial state be represented in cell D11. That's a model input, so let's go ahead and shade that green. If we entered a 'J' in there, for example, that would suggest that we start off in a Jack Pine forest. And what we're interested in doing is knowing how that Jack Pine forest, where a state of that one pixel, changed through time. And just to remind ourselves, we are going to go ahead and enter, this is pixel number 1. Knowing that the dynamics that we enter here are for one pixel, but once we know how the dynamics work for one pixel we can apply the same rational and same methodology across thousands of pixels should we choose to do so.

To start our model we're going to enter what is called a **Transition Matrix**. Let's go ahead and enter that in cell E2. We are going to write the word 'FROM'. We have three states and we'll enter those across the top row. So 'FROM' J, W, and S. That means if we are in J, W, or S we're going to enter some transition probabilities that will let us know how we move from J to another system; to a W, or to an S, or remain a J.

We'll enter this as a giant matrix. And we'll enter J, W, and S down column D, and this is going to be 'TO:'. So here we have our entries are going to be in these cells: E4 through G6. Let's go ahead and center those. And we'll shade those green and also give those a border. These represent our transition probabilities.

From J, if we are a pixel in J, this would be the probability of staying in J. Tony said that was .7. Here's the probability transition from J to W. Tony said that was .1. This cell, E6, is the probability of transitioning from J to Spruce. Again, that was .2. Let's just take a second here and clean this up, and provide these with some kind of heading.

Now, the transition probabilities must sum to 1. And so to double check our work we'll just enter a sum equation and sum cells E4 to E6, so that we always know that we're not making a mistake in our inputs.

The next set of values is the transition probabilities from W into J, W, and S. Tony said that those were .1, .8, and .1. If you are in Spruce the transition probability to Jack Pine is .5, to White Pine was 0, and to Spruce was .5. This represents our transition probability matrix. It's worth double checking your numbers to make sure that these numbers all add to 1 down the column. That's not necessary across the rows because we are starting in a state and it needs to transition to a new state and we need to account for all possibilities, so they must sum to 1.

As we have done in other spreadsheets, we're going to take this matrix and create what is called a **Cumulative Probability Matrix**. This is the matrix we are going to use in our model itself.

Let's select cell I1 and we'll type in the word 'Cumulative Probability'. As we had done before, we have J, W, and S across the top. And our actual cumulative probabilities are going to be represented in cells I3 through K6. It has one additional row of values. These are outputs because we are going to be using the values inside this matrix and computing our cumulative probabilities over here. So we will shade those blue and border them as well.

What the cumulative probabilities are going to do is take these numbers, and let's start off with the Jack Pine, and it's going to cumulate them down. And they're going to begin with the number 0, and they're going to end with the number 1. We are going to use these values here for Jack Pine.

Just to give you an idea of where we're heading here, we want to know, what's the state of this pixel in year 10. And in lecture Tony said one way you can do that is to use a random number. And you can look up random numbers, and you'll let 70% of the random numbers result in a J. You'll let 10% of those random numbers provide a result to W, and you'll let 20% of those random numbers represent an S. You choose a random number within that pot then, and you return the result.

One way to make this simple is to just cumulate these random numbers in order by using a SUM function. We'll do that in cell I4. Enter the word 'sum'. We're going to start by summing cell E4 through E4. That looks a little funny. We're going to anchor the first E4 reference, then press the Enter key and drag that formula down.

Let's just think about what we have done here. We've made it easy for us to know that if we draw a random number between 0 and .7, it will stay a Jack Pine. If we draw a random number between .7 and .8 it will convert to White Pine. If we draw a random number between .8 and 1, it will convert to Spruce. So it's the same idea, same concept. All we have done though, is to express these numbers on a scale of values that begin with the number 0 and end with the number 1, and knowing that these are ordered values. That is what this matrix is going to do. So if we change any of our values here, your cumulative probabilities will be automatically updated.

Let's just go ahead and try that. If our first number is .1 and our last number is .6, our second number must be .3. Then we know that random numbers between 0 and .1 stay as Jack Pine. We know that numbers between .1 and .4 represent a change to White Pine. Numbers between .4 and 1 represent a change to Spruce. So 60% of these values, random numbers, would result in a change to Spruce. Let's set these numbers back to the way they were originally: .7, .1, and .2.

<u>PAUSE the video</u>: Go ahead and pause the video, and fill out the equations for these other transitions, creating cumulative probabilities for the White Pine and the Spruce.

Your numbers should look just like this. Cell I4 we have a sum of E4. We are referencing the J's. Now we are referencing the W's in column J. And that would be referencing cell F4. And cell K4 has the equation related to G4. The key here is that the cumulative probabilities all begin with a 0, and they all end with a 1. What that lets us do is to use a random number now, and we can look up the random number in these ordered values, and then return what the state actually would be.

Let's go ahead in cell C10 and we'll enter our random numbers. We can start that in cell C12. Remember that the random number function is just r-a-n-d, open parenthesis, closed parenthesis, because there are no arguments in this function. And we'll drag that down so that every single time-step has a unique random number.

Let's just talk our way through this. We're in State J. How are we going to use the random number to the left, in Year 10, and determine where we transition to? We're going to use something called a 'Lookup' function. On this portion of my spreadsheet we'll demonstrate the Lookup function. But it's going to go as part of a nested function when we actually get down to cell D12, which is our main equation for our whole model.

But let's just refresh ourselves with the Lookup function itself. We're in Year 10. We have a random number. We know we're in State J. We want to look up this random number in this series of ordered numbers. So we'll find the number .45 here. We'll see that it falls between a 0 and a .7. So we know where that random number falls in this order. And when we find that, we're going to return the state associated with that position. Let's try it.

The Lookup function is just LOOKUP. And when you click OK, you'll see that it has two kinds of syntax. And we're going to use the one associated with the first syntax. This syntax has three arguments.

The first we want to look up the value, and here we'll reference the random number. We are going to identify the lookup vector. The Lookup vector is where do we want to look up this random number? A key thing here is that the Lookup vector is a range that contains one row or column of text, numbers, or logical values placed in ascending order. So they must be ordered. Once we find where the random number falls, we want to return the value that it transitions to, and that is located over here in cells D4 to D6. We press OK. And what we see is our random number, our new state is provided by that Lookup function.

In my spreadsheet my random number was .9. I looked up the value .95 in this series of data. I find that it falls between these two values. And I return the value associated with the bottom of those two numbers.

By using this system we are going to return a J, if the random number is between 0 and .7. It will return a W, if the random number is between .7 and .8. And it will return an S, if it is between .8 and 1. So, that's the Lookup function.

Now we are going to delete that. And we are going to think about what kind of an equation needs to go in cell D12, So that we can simulate and project the dynamics of this pixel, pixel number 1, through time.

Again it's useful to just think about it out loud in words. First of all, what state were we in, in the previous time-step? If we were in State J, then what we would do is look up this random number in this set of values, and then return the result in cells D4 to D6.

If it was in State W then we'd look up that random number in this vector and return the result. And iIf it was an S, Spruce, we would look up our random number in this vector of cumulative probabilities, and then we would return the values associated in D4 and D6.

So if you think about it we are going to have to do several IF functions. And we could split this out across multiple columns on our spreadsheet. But since you've done that before, today, we're going to do these as nested functions just to give you a sense for what those look like.

Let's start off by clicking the Insert Function. We're going to use an IF function. Now our logical test is simply identifying if the time-step before is equal to a J. Notice that it's not case sensitive, but just to keep things clean I'm going to enter a capital "J".

The logical test is the State before equal to a J? If it's true, then we need to use our Lookup function to look up the value in these cells here. If it's' false, then what do we need to do? It's not J. So we need another IF function that would say, well, if it's not J let's see if it's W. And then if it's not W, then it must be S, And so let's return the result there.

We happen to have J here so let's walk here. If cell D11 is J, then that's true. We need to go up and we can use this portion of our spreadsheet and find a Lookup function. Press OK. We're going to look up the random number, in the Lookup vector, I3 to I6. And the Result vector will be D4 to D6. Press OK.

Now, let's head back up to our IF function. What you can see is that what we've done is for the second argument of our IF function we've embedded the Lookup function. If D11 is J, look up C12 in cells I3 to I6, and return the value in cell D4 to D6.

If it's not J, then we need to look up whether it's W. And that would be, again, another IF function. So let's embed that IF function here. Now we have nested IF functions. We can head over to the left portion of the spreadsheet, find IF. Now let's ask if cell D11 is equal to W. If it's W, then we can look up the random number in this series and return the result in cells D4 to D6.

When you get good at, and use to using these formulas, you can just enter them straight by hand as we'll do here. So let's look up the random number, comma, in this vector; cells J3 to J6, comma. And then our last argument is cells D4 to D6. We close that Lookup by adding another parenthesis.

If it is W we do this. If it's not W, then what must it be? It must be S, and so now we would enter another Lookup function. Lookup the random number in cells K3 through K6, comma, and return the value in D4 to D6. We close that parenthesis.

Now, if you look up on your formula bar you can see that you have a very complicated looking formula. But when you break it down piece by piece it's really not that bad. This is called a nested function. In some cases in your spreadsheet modeling you'll want to use some nested functions. In other cases, you'll want to break everything apart and then add the pieces back together again. Both ways are fine. When you are starting you probably would be wise to break things apart. As you get more and more comfortable with a spreadsheet and the equations, and you feel comfortable using the nested arguments within them, go ahead and do so.

One thing that's handy, and I'm going to go ahead and push OK now. One thing that is handy is when you click on a complex formula, sometimes it's hard to know where you open and close parenthesis. And if you are ever confused by that, or wondering about that, just click next to a parenthesis inside the equation itself. And then use your arrow keys, and you can see things close and open. Let's choose a simpler one here. This is the open parenthesis for my Lookup function. Here's the closed parenthesis. Watch what happens if I toggle back and forth. You should see that the parenthesis light up and become bolded and that helps you to identify where things are.

Let's find where this IF function closes. You should see that it is popping up over here. And our first IF function, our big function, is closed at the end over here. And that's a really useful way to keep track of your equation.

When you are finished press Enter. Now we have our workhorse equation. We can drag it down. But there is one important thing we need to do first. And that is we need to anchor every reference to these cumulative probabilities, because otherwise we'll be dragging our equation down. And we want to, though we want to reference the random number associated with each time-step, we always want to be referencing the values up here without copying those down. So let's go ahead and put anchors in to those specific places. I'm doing that with the F4 button on a PC.

Now we can drag our formula down through time. As we press random numbers we can see the dynamics of our pixel are changing random. This is called a **Markov Process**, because the state of a particular pixel is dependent on the state that it was previously. Probabilities establish the rules by which things transition. And what you apply depends on the state you were in, in the previous time-step.

Just to finish this up let's just go ahead and give some headings to our actual model. And if you want to clean this up a bit. And make sure that you save your spreadsheet.

And now we'll go back to Tony who will tell us what are some of the challenges and problems with this particular kind of model.

< 00:25:00 END >