10/10

Markov Chains

Melanie Holden

Although linear algebra is considered a specialized field, many people fail to realize its applications to other areas of life, such as physics, biology, economy and more. One particular use of linear algebra that is useful to many fields is the Markov Chain. Used to describe movement of populations over a period of time, the Markov Chain is simple to interpret, and carries a lot of information. A unique matrix, called a transition matrix, can represent each chain. Both Markov Chains and transition matrices have interesting properties that will be explored within this paper.

Nice intro :-)

Consider a situation where you are splitting a population into groups. For example, you might want to describe the student body of a college as science majors, humanity majors, and art majors. Or, you could split the cat population of a county into those with homes, and those without homes. A situation where you are describing a population in terms of different states, especially if there is movement between the states, can be described by a transition matrix, defined as $T=[t_{ij}]$. For k number of states, there are k^2 transition probabilities, which form the $k \cdot k$ transition matrix. Note that a transition matrix must be a square matrix. Each entry $[t_{ij}]$ represents the portion of the population moving from state j to state i over a certain time period. Therefore, all entries in the transition matrix T must be either zero, or a positive number. Since each column of the matrix represents the movement of the entire population, the sum of the column entries, $t_{1j}+t_{2j}+. . . +t_{kj}$ always adds up to one. In our example with college majors, let's say the following:

- Of those who start as science majors, 35% become humanities majors, and 5% art majors
- Of humanities majors, 10% go to science, and 10% to art
- Of art majors, 1% changes to science, and 4% to humanities

We could then make the transition matrix

Good example

$$T=\begin{bmatrix} .60 & .10 & .01 \\ .35 & .80 & .04 \\ .05 & .10 & .95 \end{bmatrix}$$

In addition to the transition matrix, we can also describe the population in the population distribution vector **p**, where **p** is the column vector

$$\mathbf{p}=\begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{bmatrix}$$

Each entry in **p** is a nonnegative number, and the sum of the entries is one. For example, if a population was equally distributed over three states, the vector **p** would be

$$\mathbf{p}=\begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

In order to find the proportion of the whole population after a time period, one must take into consideration the contribution to the state of all the states. If there are three states, then the proportion of the population in state one is given by the proportion in state one that stays in state one, given by $t_{11}p_1$. The contributions from states two and three are found in the same way, resulting in the final equation $t_{11}p_1+t_{12}p_2+t_{13}p_3$. For those familiar with matrix multiplication, it is easy to see that this equation is the result of multiplying the transition matrix T with **p**.

$$\mathbf{Tp}=\begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix} \cdot \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}$$

Using the properties of the transition matrix and **p**, we can show that the sum of the entries in T**p** is always one (Fraleigh & Beauregard, 106). In our three state example, the sum of the entries is:

$$= t_{11}p_1 + t_{21}p_2 + t_{31}p_3 + t_{12}p_1 + t_{22}p_2 + t_{32}p_3 + t_{13}p_1 + t_{23}p_2 + t_{33}p_3$$

$$= p_1(t_{11} + t_{12} + t_{13}) + p_2(t_{21} + t_{22} + t_{23}) + p_3(t_{31} + t_{32} + t_{33})$$

$$= p_1(1) + p_2(1) + p_3(1) \qquad \text{[sum of columns in T is 1]}$$

$$= 1 \qquad \text{[sum of } \mathbf{p} \text{ entries is 1]}$$

Since each column of a transition matrix is also a population distribution vector, it is easy to find the probability that a system is in a certain state at any observation *m*, by raising **p** to the m^{th} power (Rorres & Anton, 83). This can be extended to the entire transition matrix as well. For example, T after 2 observation periods can be described by T^2. This is because the distribution vector of the j^{th} column in an $n*n$ matrix A is denoted by $A\boldsymbol{e}_j$. So after two periods, the population vector $\mathbf{e}_j$ in T is written $T(T\mathbf{e}_j) = T^2\mathbf{e}_j$. This is applicable for any number of observations *m*, where T^m describes the population. (Fraleigh & Beauregard, 106)

Once these properties of a transition matrix are understood, we can properly define a Markov Chain. When the transition between two states is not predetermined, but can only be denoted in terms of the probability determined by past observations, it is called a stochastic process. In addition to being a stochastic process, if the transition is dependent only on the state directly before the current observation, then it is a Markov Chain (Rorres & Anton, 82). In simpler terms, a situation in which the same transition matrix can be applied for multiple observations, each occurring over a time period, is a Markov Chain.

One specific kind of Markov Chain and transition matrix is a regular chain and corresponding matrix. If there is some fixed number m, where the transition matrix T^m has no zero entries, meaning that it is possible to move to or from any state, then the transition matrix T is regular. Any Markov Chain with a regular transition matrix is called a regular chain. For example, the matrix

$$T = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

is not a regular matrix. We compute T^2 and T^3 to find

$$T^2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \text{ and } T^3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I.$$

Therefore, $T^4=IT$, $T^5=IT^2$, and so on, proving that T is not a regular matrix (Fraleigh & Beauregard, 107).

As m approaches infinity, the transition matrix approaches what is called a steady state. That is to say, for a regular transition matrix, there exists a unique column vector **s** such that each of its entries are positive and add up to one so that, according to Fraleigh and Beauregard:

1. As m approaches infinity, all columns of T^m approach the column vector **s**

2. **T****s**=**s**, and **s** is the unique column vector with this property and whose components add up to one (108).

To show that this is true, we consider a three state system again:

$$\mathbf{p}=\begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}, \text{ and } \mathbf{s}=\begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix}$$

$$\mathrm{T}^m\mathbf{p}=\begin{bmatrix} s_1 & s_1 & s_1 \\ s_2 & s_2 & s_2 \\ s_3 & s_3 & s_3 \end{bmatrix}\cdot\begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}=\begin{bmatrix} s_1p_1+s_1p_2+s_1p_3 \\ s_2p_1+s_2p_2+s_2p_3 \\ s_3p_1+s_3p_2+s_3p_3 \end{bmatrix}$$

but $p_1 + p_2 + p_3 = 1$, so the vector becomes

$$\begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix}.$$

For our matrix describing college majors, we can experiment (using Maple) to find the value of *m* needed to reach its steady state:

$$T=\begin{bmatrix} .60 & .10 & .01 \\ .35 & .80 & .04 \\ .05 & .10 & .95 \end{bmatrix}, \mathrm{T}^{75}=\begin{bmatrix} .085109 & .085109 & .085105 \\ .27660 & .27660 & .27659 \\ .63829 & .63829 & .63830 \end{bmatrix}, \text{ and}$$

$$\mathrm{T}^{100}=\begin{bmatrix} .085106 & .085106 & .085106 \\ .27660 & .27660 & .27660 \\ .63830 & .63830 & .63830 \end{bmatrix}$$

So our value of *m* is 100, and our **s** vector is

$$\mathbf{s}=\begin{bmatrix} .085106 \\ .27660 \\ .63830 \end{bmatrix}.$$

Another way to solve for **s** is solving the equation T**s**=**s**. From our studies of eigenvalues, we know that we can rearrange the equation as follows:

T**s**=λI**s** How do we know that such an $\vec{s}$ always exists?

T**s** – λI**s** = 0

(T-λI)**s** = 0

Since T**s**=**s**, we know that the value of lambda must be one for every transition matrix. (Fraleigh & Beauregard, 109) We can then solve for **s** by forming the augmented matrix:

$$\left[\begin{array}{ccc|c} -.40 & .10 & .01 & 0 \\ .35 & -.20 & .04 & 0 \\ .05 & .10 & -.05 & 0 \end{array}\right] rref\text{->} \left[\begin{array}{ccc|c} 1 & 0 & -2/15 & 0 \\ 0 & 1 & -13/30 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

so s_1=-2/15r, s_2=-13/30r, and s_3=r, where r is some non-zero scalar. We also know that $s_1+s_2+s_3=1$, so by substitution we find r=30/13. Plugging that into our values for **s**, we should get our original value for **s**.

Beyond our math classes, Markov Chains can be used for many things. Searching for "Markov Chain applications" on a popular internet search engine yields hundreds of thousands of results. One website uses Markov Chains for a completely pointless reason: To create new words. The website analyzes pairs of letters in roughly 45,000 different words to create new words with reasonable syntax. They are also used in Monte Carlo theory, a theory which also has applications in many fields other than math. In Geology, Markov chains are used to describe the probability that a large area of rock with change from one lithology to another, such as from sand to shale. Another commonly talked about application is to predict what happens in baseball. A matrix can be made where the states depend on how many outs there are and which bases have runners on them.

Works Cited

Fraleigh, John B., and Raymond A. Beauregard. Linear Algebra. Addison-Wesley Publishing Company, 1995.

"Markov Chains." 9 Dec 2007. <http://www.ucsb.edu/~g.legrady/academic/courses/042102/markov.html>.

Rorres, Chris, and Howard Anton. Applications of Linear Algebra. 2nd ed. John Wile & Sons, Inc., 1979.