Linear Algebra Applications

Markov Chains

Andrew Berger

MATH 224

11 December 2007

10/10 Very nice.

1. Introduction

Markov chains are named after Russian mathematician Andrei Markov and provide a way of dealing with a sequence of events based on the probabilities dictating the motion of a population among various states (Fraleigh 105). Consider a situation where a population can exist in two or more states. A **Markov chain** is a series of discrete time intervals over which a population distribution at a given time ($t = n;\ n = 0, 1, 2, \ldots$) can be calculated based on the the distribution at an earlier time ($t = n-1$) and the probabilities governing the population changes. More specifically, a future distribution depends only on the most recent previous distribution. In any Markov process there are two necessary conditions (Fraleigh 105):

1. The total population remains fixed
2. The population of a given state can never become negative

If it is known how a population will redistribute itself after a given time interval, the initial and final populations can be related using the tools of linear algebra. A matrix T, called a **transition matrix**, describes the probabilistic motion of a population between various states. The individual elements of the matrix reflect the probability that a population moves to a certain state. The two conditions stated above require that in the transition matrix each column sums to 1 (that is, the total population is unchanging) and there are no negative entries (logically, populations are positive quantities).

2. Theory

The transition matrix T is comprised of elements, denoted t_{ij}, relating the motion of a population from state j to state i.

$$T = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix} \tag{1}$$

For example, the 3×3 matrix above represents transition probabilities between 3 different states. A given element, t_{23} for example, describes the likelihood that a ~~portion~~ member of the population moves from state 3 to state 2.

Having a means to describe the changes in a population distribution, it is also necessary to describe the overall system. The proportions of the population in its various states is given by a column vector

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix} \tag{2}$$

An element p_i of such a vector, known as a **population distribution vector**, provides the probility that a member of a population exists in a the i^{th} state. Similar to the requirements on the transition matrix, the sum of the entries in $\mathbf{p}$ must add to 1 and be nonnegative.

Application of a transition matrix to a population vector provides the population distribution at a later time. If the transition matrix remains valid over n time intervals, the population distribution at time n is given by $T^n\mathbf{p}$. This is simple to demonstrate.

$$\begin{aligned} \mathbf{p_1} &= T\mathbf{p_0} \\ \mathbf{p_2} &= T\mathbf{p_1} = T(T\mathbf{p_0}) = T^2\mathbf{p_o} \\ &\vdots \\ \mathbf{p_n} &= T\mathbf{p_{n-1}} = T^n\mathbf{p_0} \end{aligned}$$

A **regular transition matrix** is one which, when the original matrix T is raised to some power m, the result T^m has no zero entries. A Markov chain governed by such a matrix is called a **regular chain** (Fraleigh 107). For such a matrix, the populations will eventually approach a steady-state. This means that further application of the transition matrix will produce no noticeable population changes. This is not to say that the population is stagnant; rather that it is in a state of dynamic equilibrium such that the net movement into and out of a given state is zero.

$$T\mathbf{s} = \mathbf{s} \tag{3}$$

The vector $\mathbf{s}$ in Equation 3 is known as a **steady-state vector**. Rearranging the equation yields

$$\begin{aligned} T\mathbf{s} - \mathbf{s} &= 0 \\ T\mathbf{s} - I\mathbf{s} &= \mathbf{0} \\ (T - I)\mathbf{s} &= \mathbf{0} \end{aligned}$$

The last line is just a homogeneous linear equation which can be solved easily by a row reduction on the augmented matrix $[T - I|\mathbf{0}]$.

Another property of regular transition matrices is that as $m \to \infty$

$$T^m = \begin{bmatrix} s_1 & s_1 & \cdots & s_1 \\ s_2 & s_2 & & \vdots \\ \vdots & & \ddots & \\ s_n & \cdots & & s_n \end{bmatrix} \tag{4}$$

In words, as T is raised to higher and higher powers, the column vectors of T^m approach the steady-state vector $\mathbf{s}$ (Fraleigh 108). For a matrix Q comprised of the steady-state vector as its column vectors (Anton 614):

$$
\begin{aligned}
Q\mathbf{x} &= \begin{bmatrix} s_1 & s_1 & \cdots & s_1 \\ s_2 & s_2 & & \vdots \\ \vdots & & \ddots & \\ s_n & \cdots & & s_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} s_1x_1 & s_1x_2 & \cdots & s_1x_n \\ s_2x_1 & s_2x_2 & & \vdots \\ \vdots & & \ddots & \\ s_nx_1 & \cdots & & s_nx_n \end{bmatrix} \\
&= (x_1 + x_2 + \cdots + x_n) \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix} = (1)\mathbf{s} = \mathbf{s}
\end{aligned} \tag{5}
$$

This result is precisely what we would expect, since multiplication by Q is equivalent to m applications of T to the original vector $\mathbf{x}$. Note that in the last line of Equation 5, we rely on the requirement that the sum of the entries of a population distribution vector is 1. Therefore, for a steady-state transition matrix ($Q = T^m$ as $m \to \infty$), an arbitrary population distribution vector $\mathbf{x}$ is taken to the steady-state vector $\mathbf{s}$.

Let us re-examine Equation 3. Clearly this is an eigenvalue equation of the form $A\mathbf{x} = \lambda\mathbf{x}$ with $\lambda = 1$. As such, a regular transition matrix T is shown to have eigenvector $\mathbf{s}$ with eigenvalue $\lambda = 1$. Remarkably, it can be shown that *any* transition matrix obeying conditions 1 and 2 must have $\lambda = 1$ as an eigenvalue.

To prove this, we begin with an $n \times n$ transition matrix whose characteristic polynomial $\det(T - \lambda I)$ can be shown in matrix form

$$\begin{vmatrix} t_{11}-\lambda & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22}-\lambda & & \vdots \\ \vdots & & \ddots & \\ t_{n1} & \cdots & & t_{nn}-\lambda \end{vmatrix} \tag{6}$$

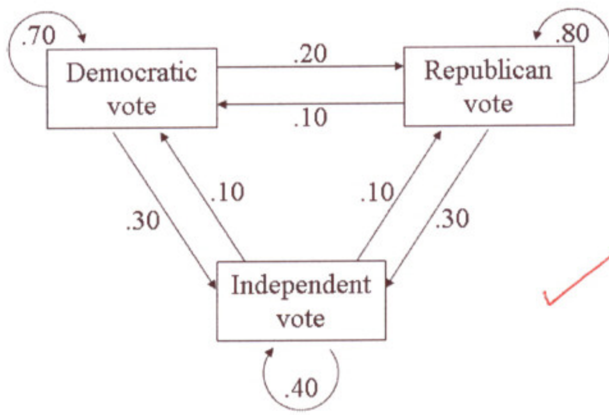
Recall condition 1 for a transition matrix: the sum of the entries of a column vector is 1. Subtraction of the eigenvalue $\lambda = 1$ exactly once from each column (as is done above in Equation 6) results in a new sum of zero for the elements of each column vector. Next, perform row operations by adding each row (2 through n) to the first row (Williams):

$$\begin{vmatrix} \sum t_{i1}-1 & \sum t_{i2}-1 & \cdots & \sum t_{in}-1 \\ t_{21} & t_{22}-1 & & \vdots \\ \vdots & & \ddots & \\ t_{n1} & \cdots & & t_{nn}-1 \end{vmatrix} \tag{7}$$

For clarity, the (-1) is repeated in each of the row 1 elements, but this simply reflects that $T-(1)I$ results in each diagonal element t_{ii} having 1 subtracted from it. The Row-Addition property of determinants guarantees that the above row operations will not alter the determinant of the matrix (Fraleigh 258). Since the summations of elements in each column are required to add to 1, and 1 is also being subtracted from each sum, the first row entry of each column is equal to zero. Furthermore, because the determinant of a matrix can be calculated by expansion by minors on any column or row of the matrix, we are free to choose expansion across the top row, which clearly results in a determinant of zero (Fraleigh 254). As such, $\lambda = 1$ is a solution to the eigenvalue equation and is therefore an eigenvalue of any transition matrix T.

3. Applications

Markov chains can be used to model situations in many fields, including biology, chemistry, economics, and physics (Lay 288). As an example of Markov chain application, consider voting behavior. A population of voters are distributed between the Democratic (D), Republican (R), and Independent (I) parties. Each election, the voting population $\mathbf{p} = [D, R, I]$ obeys the redistribution shown in Figure 1.



Nice example.

Figure 1: Voter shift between two elections (Lay 290).

For example, in an upcoming election, of those that voted Republican in the previous election, 80% will remain Republican, 10% will vote Democrat, and the remaining 10% will vote Independent.

The transition matrix describing this tendency of voter shift is given as

$$T = \begin{array}{c} \text{Party} \\ D \\ R \\ I \end{array} \begin{array}{c} \begin{array}{ccc} D & R & I \end{array} \\ \begin{bmatrix} .70 & .10 & .30 \\ .20 & .80 & .30 \\ .10 & .10 & .40 \end{bmatrix} \end{array} \tag{8}$$

The column labels indicate the initial party of a portion of the voting population, while the row labels indicate the final party. For example in column R, row I we see that 10% of those who previously voted Republican will vote Independent (in agreement with Figure 1).

In the 2004 presidential election, the voters were distributed according to the distribution vector (*CNN.com*)

$$\mathbf{p_1} = \begin{bmatrix} .48 \\ .51 \\ .01 \end{bmatrix} \tag{9}$$

If the transition matrix in Equation 12 dictates the changes between two primary elections, we can expect the outcome of the 2008 election as follows:

$$\mathbf{p_2} = T\mathbf{p_1} = \begin{bmatrix} .70 & .10 & .30 \\ .20 & .80 & .30 \\ .10 & .10 & .40 \end{bmatrix} \begin{bmatrix} .48 \\ .51 \\ .1 \end{bmatrix} = \begin{bmatrix} .390 \\ .507 \\ .103 \end{bmatrix} \tag{10}$$

More explicitly, 70% of the original Democrats (48%) remain Democrat, 10% of the 51% Republican population will vote Democrat, and 30% of the 1% Libertarian population also vote Democrat:

$$.70(.48) + .10(.51) + .30(.01) = .336 + .051 + .003 = .390 \tag{11}$$

The same goes for the shift of votes to the Republican and Libertarian parties (Lay 291).

If this voter behavior is valid for a long time, we can predict the outcome of a given election by applying the transition matrix the appropriate number of times to the initial distribution vector. Particularly, it can be shown that since the transition matrix here is obviously regular (T^1 has no zero entries), a steady-state solution is expected. Using the row-reduction method mentioned above to solve the system $[T - I|\mathbf{0}]$, we can solve for the steady-state vector.

$$[T - I|\mathbf{0}] = \begin{bmatrix} -.30 & .10 & .30 & 0 \\ .20 & -.20 & .30 & 0 \\ .10 & .10 & -.60 & 0 \end{bmatrix} \sim \begin{bmatrix} -3 & 1 & 3 & 0 \\ 2 & -2 & 3 & 0 \\ 1 & 1 & -6 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -9/4 & 0 \\ 0 & 1 & -15/4 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The eigenvector of this system corresponding to eigenvalue $\lambda = 1$ therefore, has components $x_1 = 9/4,\ x_2 = 15/4,\ x_3 =$ free. In order to transform this into a valid population vector (with the sum of the entries equal to 1), the entries are summed and then each is divided by the result.

$$\mathbf{s} = \begin{bmatrix} 9/28 \\ 15/28 \\ 1/7 \end{bmatrix} \approx \begin{bmatrix} .32 \\ .54 \\ .14 \end{bmatrix}$$

The vector $\mathbf{s}$ therefore demonstrates that the voting population will eventually settle into a state in which 54% of the votes will be cast for the Republican candidate (Lay 295).

Further insight into steady-state solutions can be gathered by considering Markov chains from a dynamical systems perspective. Consider an example of the population distribution of residents between a city and its suburbs. The transition matrix T for this system describes the movement of citizens between the city and the suburbs.

$$T = \begin{array}{r} \\ \textit{City} \\ \textit{Suburb} \end{array} \begin{array}{c} \begin{array}{cc} \textit{City} & \textit{Suburb} \end{array} \\ \begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix} \end{array} \tag{12}$$

In words, the transition matrix demonstrates that each year, 97% of the city-dwellers remain in the city while 5% migrate to the suburbs. Likewise, 97% of the suburbanites remain while 3% move into the city.

Next, we perform eigenvalue and eigenvector analysis by solving the characteristic polynomial $\det(T - \lambda I) = 0$. This yields an equation of the form $\lambda^2 - 1.92\lambda + .92 = 0$. Solving with the quadratic formula, $\lambda_1 = 1$ and $\lambda_2 = 0.92$. By substituting each of these eigenvalues into the homogenous linear system $(T - \lambda I)\mathbf{x} = \mathbf{0}$, the eigenvectors are calculated.

$$\mathbf{v_1} = \begin{bmatrix} 3 \\ 5 \end{bmatrix} \text{ and } \mathbf{v_2} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \tag{13}$$

Any arbitrary population distribution vector $\mathbf{p}$ can be written in terms of $\mathbf{v_1}$ and $\mathbf{v_2}$.

$$\mathbf{p_0} = r_1\mathbf{v_1} + r_2\mathbf{v_2} \tag{14}$$

We use a Markov chain to solve for later population distributions, and write the results in terms of the eigenvectors:

$$\begin{aligned}
\mathbf{p_1} = T\mathbf{p_0} &= c_1 T\mathbf{v_1} + c_2 A\mathbf{v_2} \\
&= c_1(1)\mathbf{v_1} + c_2(.92)\mathbf{v_2} \\
\mathbf{p_2} = T\mathbf{p_1} &= c_1 T\mathbf{v_1} + c_2(.92)A\mathbf{v_2} \\
&= c_1\mathbf{v_1} + c_2(.92)^2\mathbf{v_2}
\end{aligned}$$

Observing the pattern, we see that in general,

$$\mathbf{p_n} = c_1\mathbf{v_1} + c_2(.92)^n\mathbf{v_2}$$

As $n \to \infty$, the second term disappears, and $\mathbf{p_n}$ approaches a steady-state vector $\mathbf{s} = c_1\mathbf{v_1}$ (Lay 316). Of course, this vector corresponds to the eigenvalue $\lambda = 1$, which is indicative of a steady-state solution.

4. Conclusion

Application of linear algebra and matrix methods to Markov chains provides an efficient means of monitoring the progress of a dynamical system over discrete time intervals. Such systems exist in many fields. One main assumption of Markov chains, that only the immediate history affects the next outcome, does not account for all variables observed in the real world. Voting behavior is a case in point – it is unlikely that the voting population will approach a steady state. Nevertheless, the model affords us with good insight and perhaps serves as a helpful starting point from which more complicated and inclusive models can be developed.

References

1. Anton, Howard and Chris Rorres. *Elementary Linear Algebra Applications Version, Ninth Edition.* John Wiley & Sons, Inc. 1973.

2. *CNN.com.* Election 2004.

   http://www.cnn.com/ELECTION/2004/pages/results/president/

3. Fraleigh, John B. and Raymond A. Beauregrad. *Linear Algebra, Third Edition.* Addison-Wesley Publishing Company, Inc. 1995.

4. Lay, David C. *Linear Algebra and Its Applications, Third Edition Update.* Pearson Education, Inc. 2006.

5. Strang, Gilbert. *Linear Algebra and its Applications.* New York, NY: Academic Press, Inc. 1976.

6. Williams, Gareth. *Computational Linear Algebra With Models, Second Edition.* Boston, MA: Allyn and Bacon, Inc. 1978.