Markov Chains

Consider a physical or mathematical system undergoing a process of change. We assume that at any given time the system can be in only one of a finite number of states. For example, the weather in Peoria can be in one of three possible states: sunny, cloudy, or rainy. Or a person may be in one of three economic classes: upper class, middle class, or lower class. Suppose that the system changes over time. Although the exact state of the system might not be known with certainty at any given time, it might be possible, based on previous observations, to know the probability with which each state occurs. Such a process is called a **Markov Chain** after the Russian mathematician Andrei Andreevich Markov (1856 – 1922) who first studied them.

Transition matrices

Suppose we consider a Markov chain with n possible states, which we label $1, \ldots, n$. The state of the system is described by specifying the probabilities for each of the states occurring. The **state distribution vector** for an observation is the $n \times 1$ matrix

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

having for its i'th entry x_i, the probability that the system is in state i. A vector with this property is called a **probability vector**.

The tendency of the system to change between states over time determines certain transition parameters. The probability that the system is in state i after it was in state j is denoted p_{ij} and is called the **transition probability** from state j to state i. The matrix $T = (p_{ij})$ is called the **transition matrix** of the Markov chain.

A three state Markov chain transition matrix has the form

$$\begin{array}{cc} & \text{old state} \\ & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \text{new state} \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix} \end{array}$$

EXAMPLE: A car rental agency has three rental locations, L_1, L_2, L_3. A customer may rent a car from any of the rental locations and return it to any of the rental locations. The manager finds that the cars are returned according to the following probabilities.

$$\begin{array}{cc} & \text{Rental Location} \\ & \begin{array}{ccc} L_1 & L_2 & L_3 \end{array} \\ \text{Return Location} \begin{array}{c} L_1 \\ L_2 \\ L_3 \end{array} & \begin{pmatrix} 0.8 & 0.3 & 0.2 \\ 0.1 & 0.2 & 0.6 \\ 0.1 & 0.5 & 0.2 \end{pmatrix} \end{array}$$

You can see from this matrix that the probability is $.6$ that a car rented at location L_3 will be returned to location L_2 .

Observe several things about the transition matrix. The entries of the matrix are all non-negative and the entries of each column sum to 1 . Any matrix satisfying these two properties is called a **stochastic matrix**.

Suppose now that we know the initial state distribution vector, $\mathbf{x}^{(0)}$. We examine how to determine the next state distribution vector, $\mathbf{x}^{(1)}$. From an initial state 1 , the system has a probability p_{11} of staying in state 1 ; similarly, from state 2 the system has a probability of p_{12} of reaching state 1 ; also, from state 3 the system has a probability of p_{13} of reaching state 1 , and so on... Hence

$$\left\{\begin{array}{c}\text{new}\\ \text{probability}\\ \text{of being in}\\ \text{state } 1\end{array}\right\} = \left\{\begin{array}{c}\text{probability}\\ \text{of going from}\\ \text{state } 1\\ \text{to}\\ \text{state } 1\end{array}\right\} \cdot \left\{\begin{array}{c}\text{old}\\ \text{probability}\\ \text{of being in}\\ \text{state } 1\end{array}\right\} + \cdots + \left\{\begin{array}{c}\text{probability}\\ \text{of going from}\\ \text{state } n\\ \text{to}\\ \text{state } 1\end{array}\right\} \cdot \left\{\begin{array}{c}\text{old}\\ \text{probability}\\ \text{of being in}\\ \text{state } n\end{array}\right\} \tag{1}$$

and, in general

$$\left\{\begin{array}{c}\text{new}\\ \text{probability}\\ \text{of being in}\\ \text{state } i\end{array}\right\} = \left\{\begin{array}{c}\text{probability}\\ \text{of going from}\\ \text{state } 1\\ \text{to}\\ \text{state } i\end{array}\right\} \cdot \left\{\begin{array}{c}\text{old}\\ \text{probability}\\ \text{of being in}\\ \text{state } 1\end{array}\right\} + \cdots + \left\{\begin{array}{c}\text{probability}\\ \text{of going from}\\ \text{state } n\\ \text{to}\\ \text{state } i\end{array}\right\} \cdot \left\{\begin{array}{c}\text{old}\\ \text{probability}\\ \text{of being in}\\ \text{state } n\end{array}\right\} \tag{2}$$

In fact, if we put

$$\mathbf{x}^{(0)} = \begin{pmatrix} x_1^{(0)} \\ \vdots \\ x_n^{(0)} \end{pmatrix}, \qquad \mathbf{x}^{(1)} = \begin{pmatrix} x_1^{(1)} \\ \vdots \\ x_n^{(1)} \end{pmatrix},$$

we can write (1) in symbols as

$$\begin{array}{ccccccccc} x_1^{(1)} & = & p_{11}x_1^{(0)} & + & p_{12}x_2^{(0)} & + & \cdots & + & p_{1n}x_n^{(0)} \\ x_2^{(1)} & = & p_{21}x_1^{(0)} & + & p_{22}x_2^{(0)} & + & \cdots & + & p_{2n}x_n^{(0)} \\ & & & & \vdots & & & & \\ x_n^{(1)} & = & p_{n1}x_1^{(0)} & + & p_{n2}x_2^{(0)} & + & \cdots & + & p_{nn}x_n^{(0)} \end{array}$$

We have found the following result.

- *If T is the transition matrix for a Markov chain and $\mathbf{x}^{(0)}$ is the initial distribution vector, then*

$$\mathbf{x}^{(1)} = T\mathbf{x}^{(0)}. \tag{3}$$

If we now define $\mathbf{x}^{(k)}$ to be the state distribution vector for the k'th observation, then equation (3) gives us

$$\begin{aligned}
\mathbf{x}^{(1)} &= T\mathbf{x}^{(0)} \\
\mathbf{x}^{(2)} &= T\mathbf{x}^{(1)} = T(T\mathbf{x}^{(0)}) = T^2\mathbf{x}^{(0)} \\
\mathbf{x}^{(3)} &= T\mathbf{x}^{(2)} = T(T^2\mathbf{x}^{(0)}) = T^3\mathbf{x}^{(0)} \\
&\vdots
\end{aligned}$$

We have the following important relationship.

$$\boxed{\mathbf{x}^{(k)} = T^k\mathbf{x}^{(0)}} \tag{4}$$

Consider the situation described in Example 1 once again. If a car is initially rented at location 2, then the initial state distribution vector is

$$\mathbf{x}^{(0)} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

Using the result of (4) we get the later state vectors.

n	0	1	2	3	4	5	6	7	8	9	10	11
	0	.300	.400	.477	.511	.533	.544	.550	.553	.555	.556	.557
$\mathbf{x}^{(n)}$	1	.200	.370	.252	.261	.240	.238	.233	.232	.231	.230	.230
	0	.500	.230	.271	.228	.227	.219	.217	.215	.214	.214	.213

All distributions after this have the state distribution vector equal to $\mathbf{x}^{(11)}$ to three decimal places.

Two important things need to be observed here. First, it is not necessary for the time periods between observations to be regular; after all, we do not always know when a customer is going to return the car. But more important is to observe that *the state distribution vectors approach a fixed vector as time proceeds*. This is the key to further study of Markov chains.

Steady States

It is not always the case that the state distribution vectors of a Markov chain always approach a fixed vector. However, if we impose a very mild condition on the transition matrix, we will see that such a limiting vector exists.

We will call a transition matrix **regular** if some power of it has strictly positive entries. A Markov chain governed by a regular transition matrix will be called a **regular Markov chain**. (What else?) This is the case in the example we have been considering; in fact, the first power of the transition matrix already has all entries strictly positive. The following major result is proven in advanced courses in matrix theory.

• *Consider a regular Markov chain with transition matrix* T . *Then*

(a) As $k \to \infty$

$$T^k \to \begin{pmatrix} s_1 & s_1 & \cdots & s_1 \\ s_2 & s_2 & \cdots & s_2 \\ \vdots & \vdots & & \vdots \\ s_n & s_n & \cdots & s_n \end{pmatrix}$$

where the vector $\mathbf{s} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix}$ *is a probability vector.*

(b) As $k \to \infty$

$$T^k \mathbf{x}^{(0)} \to \mathbf{s}$$

for **any** *initial state distribution vector* $\mathbf{x}^{(0)}$.

The vector $\mathbf{s}$ is called the **steady state vector**. It is unique and represents the long–term probability of the distribution of the states of the system.

Thus one way to determine the steady state vector is simply to compute $T^k \mathbf{x}^{(0)}$ for some large integer k . This can be somewhat inefficient. In the example above, however, you can easily check that the steady state vector, $\mathbf{s}$, has the property that $T\mathbf{s} = \mathbf{s}$. We have the following additional important property.

• *The steady state vector* $\mathbf{s}$ *of a regular Markov chain is the unique probability vector satisfying the equation* $T\mathbf{s} = \mathbf{s}$.

Note that the following equations are equivalent.

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

Hence the steady state vector is the unique probability vector satisfying the following.

$$\boxed{(T - I)\mathbf{s} = 0}$$

We will later see that this means that $\mathbf{s}$ is an eigenvector for the eigenvalue 1 of the transition matrix, T .