Why Do Markov Chains Converge?

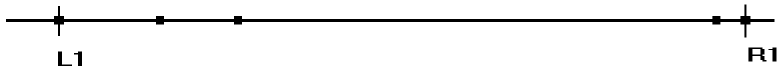
First, recall that a Markov Chain is represented by a square (*k* x *k*) matrix *T*, each row of which sums to 1. We know that for any initial probability distribution X_0, the sequence $X_1, X_2, X_3, \ldots$ generated by $X_{n+1} = X_n \cdot T$ converges to a steady state $S = \left[s_1, s_2, s_{3,} \cdots s_k\right]$. We further recognize that the matrix X_{n+1} is also given by $X_{n+1} = X_0 \cdot T^{n+1}$ and that successive powers of *T* also converge to a matrix in which each row is *S*. Why?

Perhaps the simplest way to see why the sequence converges is to look at the product of a simple transition matrix $T \cdot T = T^2$. For our argument let $T = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}$.

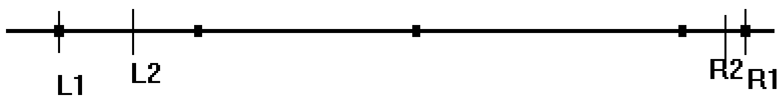
$$\text{So } T \cdot T = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \cdot \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}. =$$

$$\begin{bmatrix} a_1a_1 + a_2b_1 + a_3c_1 & a_1a_2 + a_2b_2 + a_3c_2 & a_1a_3 + a_2b_3 + a_3c_3 \\ b_1a_1 + b_2b_1 + b_3c_1 & b_1a_2 + b_2b_2 + b_3c_2 & b_1a_3 + b_2b_3 + b_3c_3 \\ c_1a_1 + c_2b_1 + c_3c_1 & c_1a_2 + c_2b_2 + c_3c_2 & c_1a_3 + c_2b_3 + c_3c_3 \end{bmatrix}$$

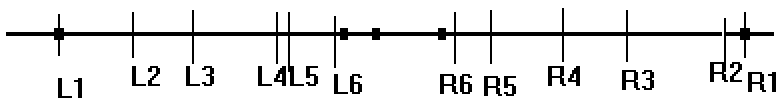
The entry in the first row first column of this product has a values $a_1a_1 + a_2b_1 + a_3c_1$. If each of the *a*'s is $\frac{1}{3}$, then this value is simply the average of the numbers in the first column, $\frac{1}{3}\left(a_1 + b_1 + c_1\right)$. If $a_1 = \frac{1}{3}$, $a_2 = \frac{1}{6}$, and $a_3 = \frac{1}{2}$, then the value of $a_1a_1 + a_2b_1 + a_3c_1$ is $\frac{1}{3}a_1 + \frac{1}{6}b_1 + \frac{1}{2}c_1 = \frac{2a_1 + b_1 + 3c_1}{6}$, the average of the set $a_1, a_1, b_1, c_1, c_1, c_1$. The key to understanding the convergence of Markov Chains is to think about averages. Regardless of the values of the *a*'s, the value of the first entry in the product matrix, $a_1a_1 + a_2b_1 + a_3c_1$, can be thought of as a weighted average of the values in the first column, a_1, b_1, and c_1, with the weights determined by the *a*'s. What is the important characteristic of averages that makes this work? Regardless of the weights, **the average of a set of numbers is always less than or equal to the largest and greater than or equal to the smallest** (equal only if all weights but one are zero). So the value in the first row and first column of the product must be between the largest and smallest of a_1, b_1, and c_1. The value in the second row and first column is another weighted average of a_1, b_1, and c_1 and so it too, lies in the same interval. Likewise for the term in the third row and first column. They are different because the weights are likely different, but all lie in the same interval, between L1 the smallest value and R1 the largest



If we multiply by T again, we have the same result. Now, the new values in the first column all must be less than or equal to the new largest value and greater than or equal to the new smallest value. That is, the interval in which the three values fall is smaller. Now they lie between L2 and R2.



Continuing on, we find the three values in the first column of successive powers of T necessarily lying in a region that is always smaller than the previous region. In fact, although it is difficult to prove, the length of the region containing the values shrinks to zero, carrying with it all three values. Hence the convergence.



The other columns work similarly.

Of course, this doesn't prove the values converge, only that the interval containing them decreases in size. To prove that the columns converge, we need to show that the difference between the largest and smallest in the column goes to zero.

Let q be the smallest element in matrix T, and let M_0 and m_0 be the largest and smallest terms in a given column of T. Consider the product

$$\begin{bmatrix} t_{11} & t_{12} & t_{13} & \cdots & t_{1k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a & b & q & \cdots & c \\ \vdots & & \vdots & \cdots & \vdots \\ t_{k1} & t_{k2} & t_{k3} & \cdots & t_{kk} \end{bmatrix} \cdot \begin{bmatrix} \vdots \\ M_0 \\ \vdots \\ m_0 \\ \vdots \end{bmatrix}.$$

Let M_1 and m_1 be the largest and smallest elements in this product. The largest possible value of M_1 would happen if the number q were multiplied by m_0 and all the other terms in the vector were M_1. So we know that

$$M_1 \le qm_0 + (1-q)M_0 \le qM_0 + (1-q)M_0 = M_0. \tag{1}$$

In a similar manner, we can argue that the minimum value in the product will be a result of multiplying q by M_0 with all other terms m_0. So,

$$m_1 \ge qM_0 + (1-q)m_0 \le qm_0 + (1-q)m_0 = m_0. \tag{2}$$

Thus, we know that

$$M_0 \geq M_1 \geq M_2 \geq \cdots \geq M_n$$

and

$$m_0 \leq m_1 \leq m_2 \leq \cdots \leq m_n .$$

This is the result that our "weighted averages" argument above gave us.

To prove convergence, subtract equation (2) from equation (1) to find

$$M_1 - m_1 \leq \left(qm_0 + (1-q)M_0\right) - \left(qM_0 + (1-q)m_0\right) = (1-2q)(M_0 - m_0).$$

If we repeat the process, we find that

$$M_2 - m_2 \leq (1-2q)(M_1 - m_1) \leq (1-2q)^2 (M_0 - m_0)$$

$$M_3 - m_3 \leq (1-2q)(M_2 - m_2) \leq (1-2q)^3 (M_0 - m_0)$$

and

$$M_n - m_n \leq (1-2q)(M_{n-1} - m_{n-1}) \leq (1-2q)^n (M_0 - m_0).$$

Finally, we notice that since q is the smallest element of T, $q \leq \frac{1}{2}$, and so $\lim_{n\to\infty}(1-2q)^n = 0$.

Additionally, since we know that $T^n \to \begin{bmatrix} s_1 & s_2 & s_3 \\ s_1 & s_2 & s_3 \\ s_1 & s_2 & s_3 \end{bmatrix}$, we know that for any initial vector $X_0 = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$,

$X_0 \cdot T^n = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} s_1 & s_2 & s_3 \\ s_1 & s_2 & s_3 \\ s_1 & s_2 & s_3 \end{bmatrix}$. This product is the row vector

$$\begin{bmatrix} x_1 s_1 + x_2 s_1 + x_3 s_1 & x_1 s_2 + x_2 s_2 + x_3 s_2 & x_1 s_3 + x_2 s_3 + x_3 s_3 \end{bmatrix}.$$

Notice that each entry in this vector has a common term and can be rewritten as

$$\begin{bmatrix} s_1(x_1 + x_2 + x_3) & s_2(x_1 + x_2 + x_3) & s_3(x_1 + x_2 + x_3) \end{bmatrix}$$

and that $x_1 + x_2 + x_3 = 1$. This just leaves the steady state vector $\begin{bmatrix} s_1 & s_2 & s_3 \end{bmatrix}$.

Reference: Olinick, Michael, *An Introduction to Mathematical Models in the Social and Life Sciences*, Addison-Wesley Publishing Company, Reading, Massachusetts, 1978.