On Infinite Markov Chains

Steven Labalme

April 8, 2020

Contents

List of Figures

Abstract

This paper was inspired by Problem 9 of the Chapter 6 review problems, which is restated below for convenience.

> **Problem 9**
> Let $A = \begin{bmatrix} 0.9 & 0.4 \\ 0.1 & 0.6 \end{bmatrix}$. Find the limiting value of $A^k \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ as $k \to \infty$.

However, to clarify the findings of this paper[1], Problem 9 will hereafter be considered to be the following.

> **Problem 9**
> Let $A = \begin{bmatrix} 0.9 & 0.4 \\ 0.1 & 0.6 \end{bmatrix}$. Find the limiting value of $A^k \begin{bmatrix} 6 \\ 4 \end{bmatrix}$ as $k \to \infty$.

While Problem 9 presents as a straightforward Markov chain problem, and it certainly can be solved as such, this paper explores a quicker, far simpler method. To begin, Problem 9 will be solved by following the standard Markov chain procedure. Following that, empirical evidence will be assembled to suggest the presence of a simpler method, and this evidence will be fleshed out. After the validity of this new method is shown to extend to Problem 9, the tenets of the method will be rigorously derived.

Having been presented, employed, and justified, the new method will be compared and contrasted with the Markov chain method. Lastly, a number of new areas of research opened up by the derivation of the new method will be explored.

Note that the techniques used in Section 1 are adapted from in-class teachings, summarized in the author's class notes for Chapter 6: [Labalme, 2020]. Note also that this paper will entirely ignore complex arithmetic.

---

[1]So that $c \neq 1$ in Section 2.1.2. If c were to equal 1, the derived-by-inspection null space vector could be confused with the solution vector.

1 Solving Problem 9 With a Markov Chain

Restate Problem 9 for convenience:

> **Problem 9**
> Let $A = \begin{bmatrix} 0.9 & 0.4 \\ 0.1 & 0.6 \end{bmatrix}$. Find the limiting value of $A^k \begin{bmatrix} 6 \\ 4 \end{bmatrix}$ as $k \to \infty$.

The presence of both a **Markov matrix** (A) and of A being raised to a power before mapping an initial condition vector evokes a **Markov chain**. As such, the tools of Markov chains can be employed to evaluate Equation 1, which symbolically represents Problem 9.

$$\lim_{k \to \infty} \begin{bmatrix} 0.9 & 0.4 \\ 0.1 & 0.6 \end{bmatrix}^k \begin{bmatrix} 6 \\ 4 \end{bmatrix} \tag{1}$$

To efficiently raise A to an arbitrarily high power, **diagonalize** it. The reason for the diagonalization is that due to the nature of the $A = S\Lambda S^{-1}$ factorization, $A^k = S\Lambda^k S^{-1}$. Indeed, diagonalization transforms the complex process of raising A to a power into the relatively simple task of raising a diagonal matrix to a power and then recompiling the factorization.

To diagonalize A, first find the eigenvalues, as follows.

$$\begin{aligned}
0 &= \begin{vmatrix} 0.9 - \lambda & 0.4 \\ 0.1 & 0.6 - \lambda \end{vmatrix} \\
&= (0.9 - \lambda)(0.6 - \lambda) - (0.4)(0.1) \\
&= 0.54 - 1.5\lambda + \lambda^2 - 0.04 \\
&= \lambda^2 - 1.5\lambda + 0.5 \\
&= 2\lambda^2 - 3\lambda + 1 \\
&= (2\lambda - 1)(\lambda - 1)
\end{aligned}$$

$$\lambda_1 = 0.5 \qquad\qquad\qquad \lambda_2 = 1$$

Use the eigenvalues to find the eigenvectors, as follows[2].

$$\begin{aligned}
0 &= (A - \lambda_1 I)x_1 \\
\begin{bmatrix} 0 \\ 0 \end{bmatrix} &= \left(\begin{bmatrix} 0.9 & 0.4 \\ 0.1 & 0.6 \end{bmatrix} - 0.5 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \begin{bmatrix} x_{1_1} \\ x_{1_2} \end{bmatrix} \\
&= \begin{bmatrix} 0.4 & 0.4 \\ 0.1 & 0.1 \end{bmatrix} \begin{bmatrix} x_{1_1} \\ x_{1_2} \end{bmatrix} \\
&= \begin{bmatrix} 0.4 & 0.4 \\ 0.1 & 0.1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
0 &= (A - \lambda_2 I)x_2 \\
\begin{bmatrix} 0 \\ 0 \end{bmatrix} &= \left(\begin{bmatrix} 0.9 & 0.4 \\ 0.1 & 0.6 \end{bmatrix} - 1 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \begin{bmatrix} x_{1_1} \\ x_{1_2} \end{bmatrix} \\
&= \begin{bmatrix} -0.1 & 0.4 \\ 0.1 & -0.4 \end{bmatrix} \begin{bmatrix} x_{1_1} \\ x_{1_2} \end{bmatrix} \\
&= \begin{bmatrix} -0.1 & 0.4 \\ 0.1 & -0.4 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix}
\end{aligned}$$

$$x_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \qquad\qquad\qquad x_2 = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

From the above information (λ_1, λ_2, x_1, and x_2), A can be diagonalized as follows[3].

$$A = -\frac{1}{5} \begin{bmatrix} -1 & 4 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0.5 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -4 \\ -1 & -1 \end{bmatrix} \tag{2}$$

---

[2]Note that the sudden appearance of the eigenvectors in the following algebra is because inspection (basically guess and check) was used to find them. Inspection is a suitable method for simple instances, such as this 2-square example.

[3]Note that the $-\frac{1}{5}$ pertains to every entry of S^{-1}. It is factored out so that every entry of S^{-1} is not a decimal.

At this point, it is possible to return to the original task of evaluating Equation 1. Begin by substituting Equation 2 into Equation 1, as follows.

$$\lim_{k \to \infty} \left(-\frac{1}{5} \begin{bmatrix} -1 & 4 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0.5 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -4 \\ -1 & -1 \end{bmatrix} \right)^k \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

As was discussed earlier, move the exponent into the factorization so that it only raises Λ to the power k.

$$\lim_{k \to \infty} -\frac{1}{5} \begin{bmatrix} -1 & 4 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0.5 & 0 \\ 0 & 1 \end{bmatrix}^k \begin{bmatrix} 1 & -4 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

Apply the fact that raising a diagonal matrix to a power is the same as raising each of its entries to that power.

$$\lim_{k \to \infty} -\frac{1}{5} \begin{bmatrix} -1 & 4 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0.5^k & 0 \\ 0 & 1^k \end{bmatrix} \begin{bmatrix} 1 & -4 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} 6 \\ 4 \end{bmatrix}$$

It is now possible to evaluate the limit. While it may not be obvious what limit the entirety of the above expression approaches, all that is necessary is to evaluate the limits of 0.5^k and 1^k as $k \to \infty$, respectively, and then recompile the factorization. This can be done as follows.

$$\begin{aligned}
\lim_{k \to \infty} -\frac{1}{5} \begin{bmatrix} -1 & 4 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0.5^k & 0 \\ 0 & 1^k \end{bmatrix} \begin{bmatrix} 1 & -4 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} 6 \\ 4 \end{bmatrix} &= -\frac{1}{5} \begin{bmatrix} -1 & 4 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -4 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} 6 \\ 4 \end{bmatrix} \\
&= -\frac{1}{5} \begin{bmatrix} -1 & 4 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -10 \\ -10 \end{bmatrix} \\
&= -\frac{1}{5} \begin{bmatrix} -1 & 4 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ -10 \end{bmatrix} \\
&= -\frac{1}{5} \begin{bmatrix} -40 \\ -10 \end{bmatrix} \\
&= \begin{bmatrix} 8 \\ 2 \end{bmatrix}
\end{aligned}$$

The vector $\begin{bmatrix} 8 \\ 2 \end{bmatrix}$ is the final simplification of Equation 1 and, therefore, the answer to Problem 9.

2 Solving Problem 9 More Concisely

2.1 A New Method

2.1.1 Assembling the New Method

The limit given by Equation 1 converges on an element of $\mathbb{R}^2$ as $k \to \infty$. But there is another, simpler way to think about this process. Because the limit converges, as k increases, the change that each repeated application of A causes to the product diminishes, i.e., the difference vector $Ax_0 - x_0$ is greater in magnitude than the difference vector $A^2 x_0 - Ax_0$, and so on and so forth.

The consistently diminishing nature of the difference vector implies that at infinity, this difference vector is 0: symbolically, $AA^\infty x_0 - A^\infty x_0 = 0$. If $x = A^\infty x_0$ is the solution vector (the evaluation of Equation 1), then substitution gives $Ax - x = 0$, or

$$Ax = x \tag{3}$$

This result actually makes quite a bit of sense — if x truly is the limit of repeated multiplications by A, it stands to reason that A could no longer do anything to change it. Indeed, A would be an identity map on x, and the above equation reflects that fact, as it can be restated $Ax = Ix$.

Equation 3 provides a computationally far simpler method than Markov chains to find x. However, the solution to $Ax = x$ is the one-dimensional subspace $\operatorname{span}(x)$, only one vector of which will be the solution to Problem 9. Thus, to pin down a single vector, it will be necessary to find a second equation describing this situation.

A good first place to look is at the initial conditions x_0, which have not yet been used. Since the solution most certainly depends on them, perhaps it is possible to use them to find the desired vector of $\operatorname{span}(x)$, killing two birds with one stone. Indeed, finding a single solution depends on the initial conditions, but in a manner best visualized by applying the Markov chain to a more tangible model, as follows.
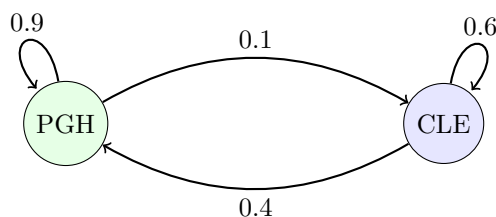
Figure 1: Modeling population changes with a Markov chain.

Consider matrix A from Problem 9, but with the initial conditions adjusted to be 600,000 and 400,000 respectively. Figure 1 visualizes Problem 9 in the following ways.

1. Imagine that 600,000 is the population of Pittsburgh (PGH) in year 0.

2. Imagine that 400,000 is the population of Cleveland (CLE) in year 0.

   a. Together, the populations are represented by the vector $x_0 = \begin{bmatrix} \text{PGH} \\ \text{CLE} \end{bmatrix} = \begin{bmatrix} 600,000 \\ 400,000 \end{bmatrix}$.

3. Every year, 90% (0.9) of Pittsburgh's population remains in Pittsburgh, and 10% (0.1) moves to Cleveland.

4. Every year, 60% (0.6) of Cleveland's population remains in Cleveland, and 40% (0.4) make the right choice (☺).

   a. Together, the movements are represented by the linear map $A = \begin{bmatrix} 0.9 & 0.4 \\ 0.1 & 0.6 \end{bmatrix}$.

5. In year k, the populations of each city are given by $A^k x_0$.

a. For example, in year 1, the populations are $\begin{bmatrix} \text{PGH} \\ \text{CLE} \end{bmatrix} = A^1 x_0 = \begin{bmatrix} 0.9 & 0.4 \\ 0.1 & 0.6 \end{bmatrix}^1 \begin{bmatrix} 600,000 \\ 400,000 \end{bmatrix} = \begin{bmatrix} 700,000 \\ 300,000 \end{bmatrix}$.

6. In the long run, the populations of each city are given by the limit of $A^k x_0$ as $k \to \infty$.

a. Eventually, the populations converge on $\begin{bmatrix} \text{PGH} \\ \text{CLE} \end{bmatrix} = A^\infty x_0 = \begin{bmatrix} 0.9 & 0.4 \\ 0.1 & 0.6 \end{bmatrix}^\infty \begin{bmatrix} 600,000 \\ 400,000 \end{bmatrix} = \begin{bmatrix} 800,000 \\ 200,000 \end{bmatrix}$ [4].

In this model, it should be clear that no person is ever introduced or removed from the system — they only move from one city to the other, or stay in place. This is reflected by the fact that the net number of people in every "year" given above is always one million — $600,000 + 400,000 = 1,000,000$, $700,000 + 300,000 = 1,000,000$, and $800,000 + 200,000 = 1,000,000$.

Herein lies the key: Multiplying the initial conditions by a Markov matrix A appears to preserve the sum of the values in x_0. Therefore, it is possible to formalize this second rule as follows, where each sum iterates through the n entries in its respective vector.

$$\sum_{i=1}^{n} x_{0_i} = \sum_{i=1}^{n} x_i \tag{4}$$

With the results of Equations 3 and 4, solving Problem 9 is a trivial matter.

2.1.2 Applying the New Method

Restate Problem 9 for convenience:

Problem 9

Let $A = \begin{bmatrix} 0.9 & 0.4 \\ 0.1 & 0.6 \end{bmatrix}$. Find the limiting value of $A^k \begin{bmatrix} 6 \\ 4 \end{bmatrix}$ as $k \to \infty$.

In parallel with Section 2.1.1, A is a 2-square Markov matrix and x_0 will be the variable name for the initial conditions of the Markov chain in Problem 9. Also, once again let x be the sought-after solution vector.

Use Equation 3 to solve for the subspace in which x lies (see Footnote 2).

$$x = Ax$$
$$0 = Ax - x$$
$$= Ax - Ix$$
$$= (A - I)x$$
$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \left(\begin{bmatrix} 0.9 & 0.4 \\ 0.1 & 0.6 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$
$$= \begin{bmatrix} -0.1 & 0.4 \\ 0.1 & -0.4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$
$$= \begin{bmatrix} -0.1 & 0.4 \\ 0.1 & -0.4 \end{bmatrix} \begin{bmatrix} 4c \\ 1c \end{bmatrix}, \qquad c \in \mathbb{R}$$
$$x = \text{span}\left(\begin{bmatrix} 4 \\ 1 \end{bmatrix} \right)$$

Use Equation 4 to solve for a single x.

$$\sum_{i=1}^{n} x_{0_i} = \sum_{i=1}^{n} x_i$$
$$6 + 4 = 4c + 1c$$
$$c = 2$$

---

[4]Note that the final vector can be found by scaling the solution vector from Section 1 by the same factor (100,000) that scaled x_0. This is further evidence that the solution depends on the initial conditions.

$$x = \begin{bmatrix} 4(2) \\ 1(2) \end{bmatrix}$$
$$= \begin{bmatrix} 8 \\ 2 \end{bmatrix}$$

The vector $\begin{bmatrix} 8 \\ 2 \end{bmatrix}$ is the specific element of span $\left(\begin{bmatrix} 4 \\ 1 \end{bmatrix} \right)$ that is the solution to Problem 9.

2.2 Rigorously Deriving the New Method

Let A be an n-square Markov matrix, and let $x_0 \in \mathbb{R}^n$ be a vector of initial conditions for the Markov chain.

Lemma 1 $A - I$ is a **singular** matrix.

Proof By the definition of a markov matrix, the entries in each column of A sum to 1. However, when the difference $A - I$ is taken, one entry of every column of A decreases in value by 1. Thus, the entries in each column of $A - I$ sum to 0.

If the sum of the entries in each column of $A - I$ is 0, every column j (and every vector $v \in C(A - I)$) can be expressed in the following form.

$$v = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{(n-1)j} \\ -a_{1j} - a_{2j} - \cdots - a_{(n-1)j} \end{bmatrix}$$

Notably, no vector *not* in this form can be an element of $C(A - I)$, and, similarly, any vector not in this form would be an element of $N(A - I)$. Thus, proving that $A - I$ is singular becomes a task of finding a vector $u \in \mathbb{R}^n$ that is not in the above form. The following is an example of such a vector.

$$u = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

It is obvious that the above vector is not in the form necessary for it to be an element of $C(A - I)$.

As further proof that $u \notin C(A - I)$ and $u \in N(A - I)$, the fundamental theorem of linear algebra asserts that all elements of $C(A - I)$ are orthogonal to all elements of $N(A - I)$, or, alternatively, that the dot product of any element of the column space with any element of the null space is 0. Since v encapsulates every possible element of $C(A - I)$, the following definitively proves that $u \notin C(A - I)$.

$$0 = v \cdot u$$
$$= \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{(n-1)j} \\ -a_{1j} - a_{2j} - \cdots - a_{(n-1)j} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$
$$= (1)(a_{1j}) + (1)(a_{2j}) + \cdots + (1)(a_{(n-1)j}) + (1)(-a_{1j} - a_{2j} - \cdots - a_{(n-1)j})$$
$$= 0$$

Since there always exist elements of $\mathbb{R}^n$ (at least span (u)) in $N(A - I)$, $A - I$ is singular.

$\blacksquare$

Lemma 2   For all A, there exists a vector $x \in \mathbb{R}^n$ satisfying $Ax = x$.

Proof   By Lemma 1, the matrix $A - I$ has a null space. In other words, the following equation has at least one family of solutions.
$$(A - I)x = 0$$
The above equation can be algebraically rearranged as follows, proving the lemma.
$$\begin{aligned} (A - I)x &= 0 \\ Ax - Ix &= 0 \\ Ax &= Ix \\ Ax &= x \end{aligned}$$

■

Lemmas 1 and 2 hold true for any Markov matrix A. However, before tackling convergence, it is necessary to point out that not all Markov chains converge. For example, if A is the following **permutation matrix**, the Markov chain diverges periodically for infinitely many (but not all) x_0.

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

Theorem 1   The solution to $\lim_{k \to \infty} A^k x_0$ lies in span (x), where x satisfies $Ax = x$, for any x_0 if and only if the limit converges.

Proof   If $\lim_{k \to \infty} A^k x_0$ converges, the difference vector $A^{k+1} x_0 - A^k x_0$ between an instance $A^k x_0$ and the next instance $A^{k+1} x_0$ converges to 0. The convergence of the distance vector implies the validity of the following algebra.

$$\begin{aligned} 0 &= \lim_{k \to \infty} \left(A^{k+1} x_0 - A^k x_0 \right) \\ &= \lim_{k \to \infty} \left(AA^k x_0 - IA^k x_0 \right) \\ &= \lim_{k \to \infty} (A - I) A^k x_0 \\ &= (A - I) \lim_{k \to \infty} A^k x_0 \end{aligned}$$

By the above algebra, $\lim_{k \to \infty} A^k x_0$ is an element of $N(A - I)$. If $\lim_{k \to \infty} A^k x_0 \in N(A - I)$, Lemma 2 guarantees that it lies in span (x), where x satisfies $Ax = x$.

■

Theorem 1 concludes the proof of Equation 3. All that is left at this point is to prove Equation 4.

Lemma 3   The sum of the entries in x_0 equals the sum of the entries in Ax_0. Symbolically, the following equation is true for any A and x_0: $\sum_{i=1}^n x_{0_i} = \sum_{i=1}^n (Ax_0)_i$.

Proof   Let A and x_0 be visualized as follows.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \qquad x_0 = \begin{bmatrix} x_{0_1} \\ x_{0_2} \\ \vdots \\ x_{0_n} \end{bmatrix}$$

Then Ax_0 is given by the following.

$$Ax_0 = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_{0_1} \\ x_{0_2} \\ \vdots \\ x_{0_n} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}x_{0_1} + a_{12}x_{0_2} + \cdots + a_{1n}x_{0_n} \\ a_{21}x_{0_1} + a_{22}x_{0_2} + \cdots + a_{2n}x_{0_n} \\ \vdots \\ a_{n1}x_{0_1} + a_{n2}x_{0_2} + \cdots + a_{nn}x_{0_n} \end{bmatrix}$$

With both x_0 and Ax_0 defined, algebraically transform one of the expressions in the hypothesis into the other, as follows[5].

$$\sum_{i=1}^{n}(Ax_0)_i = (a_{11}x_{0_1} + \cdots + a_{1n}x_{0_n}) + (a_{21}x_{0_1} + \cdots + a_{2n}x_{0_n}) + \cdots + (a_{n1}x_{0_1} + \cdots + a_{nn}x_{0_n})$$

$$= (a_{11} + \cdots + a_{n1})x_{0_1} + (a_{12} + \cdots + a_{n2})x_{0_2} + \cdots + (a_{1b} + \cdots + a_{nb})x_{0_n}$$

$$= 1x_{0_1} + 1x_{0_2} + \cdots + 1a_{0_n}$$

$$= \sum_{i=1}^{n} x_{0_i}$$

∎

Theorem 2   The sum of the entries in x_0 equals the sum of the entries in $A^k x_0$ for any $k \in \mathbb{N}$.

Proof   The following will be an induction proof.
Basis step: $\sum_{i=1}^{n}(Ax_0)_i \overset{\checkmark}{=} \sum_{i=1}^{n} x_{0_i}$ (guaranteed by Lemma 3).
Induction hypothesis: $\sum_{i=1}^{n}(A^k x_0)_i = \sum_{i=1}^{n} x_{0_i}$.
Induction step:

$$\sum_{i=1}^{n}(A^{k+1}x_0)_i \overset{?}{=} \sum_{i=1}^{n}(Ax_0)_i$$

$$\sum_{i=1}^{n}(A^{k+1}x_0)_i \overset{\checkmark}{=} \sum_{i=1}^{n} x_{0_i}$$

∎

Although Theorem 2 does not directly treat the infinite case, stating that the sum of the entries never changes under repeated multiplications by A implies that the sum of the entries never changes as $k \to \infty$, proving Equation 4.

Together, Theorems 1 and 2 guarantee that the method of solving Problem 9 used in Section 2.1.2 stands on solid logical footing.

---

[5]Note that the transition from the second to the third line of the following derivation is legal by the definition of a Markov matrix — every sum $a_{1j} + \cdots + a_{nj}$ is the sum of the values in a column j of A, which is, by definition, equal to 1.

3   Comparing and Contrasting the Two Methods

Markov chains are useful not only for their prediction of long-run behavior, but for their prediction of short-term behavior, too. For example, in the population model discussed in association with Figure 1, the model's 1, 2, or 5-year predictions are likely far more useful than its long-run behavior; after all, those rates of population movement are unlikely to remain static for infinitely many years.

However, when determining the long-run behavior of a convergent Markov chain, the new method is more than useful. It eliminates the need for the time-consuming process of diagonalization, removes the need for limits entirely, and avoids the lengthy process of recompiling the factorization.

Most importantly, the new method is a gateway to discovering and proving fascinating results on the nature of infinite Markov chains, some of which are discussed in the next section.

4   Diving Deeper

The method of evaluating a convergent, infinite Markov chain discussed in Section 2 opens up several new, fascinating ways to learn about the properties of Markov matrices. Several of these will be explored below as corollaries to the lemmas and theorems of Section 2.2.

Let A be an n-square Markov matrix. Let $x \in \mathbb{R}^n$ satisfy $Ax = x$. Let $x_0 \in \mathbb{R}^n$ be a vector of initial conditions for the Markov chain such that $\lim_{k \to \infty} A^k x_0$ converges.

Corollary 1   For all A, there exists an eigenvalue λ_i, where $\lambda_i = 1$.

Proof   By Lemma 2, there always exists a vector x. By definition, an eigenvalue (of an arbitrary m-square matrix B) is a scalar $\lambda \in \mathbb{R}$ satisfying $By = \lambda y$, where $y \in \mathbb{R}^m$.

The parallel between $Ax = x$, or $Ax = 1x$, and $By = \lambda y$ is unmistakable. Since $Ax = 1x$ is always solvable, $\lambda_i = 1$ must be an eigenvalue of A.

∎

It is worth noting that since 1 is an eigenvalue of A, x is an eigenvector. In other words, $\lim_{k \to \infty} A^k x_0$ is an eigenvector of A. Eigenvectors are also known as "characteristic vectors," and the appearance of one as the solution to $\lim_{k \to \infty} A^k x_0$ is yet another surprising expression of their characterization of A.

For the next corollary, let $P_A(\lambda)$ denote the characteristic polynomial of A.

Corollary 2 $P_A(\lambda)$ is evenly divisible by $(\lambda - 1)$.

Proof   By definition, the roots of $P_A(\lambda)$ are equal to the eigenvalues of A. By Corollary 1, 1 is an eigenvalue of any A and, therefore, is a root of $P_A(\lambda)$. For any root r of a polynomial $P(x)$, $(x - r)$ is a factor of $P(x)$. Since 1 is a root of $P_A(\lambda)$, $(\lambda - 1)$ is a factor of $P_A(\lambda)$, which implies that $P_A(\lambda)$ is evenly divisible by $(\lambda - 1)$.

∎

Let A^∞ denote[6] $\lim_{k \to \infty} A^k$. If u_1, u_2 are vectors in the same dimension, let $u_1 \cup u_2$ denote the matrix created by horizontally concatenating u_1 and u_2.

Corollary 3   Every vector $v_1, v_2, \ldots, v_n$ composing A^∞ equals the element of span (x) whose components sum to 1 if and only if $v_1, v_2, \ldots, v_n$ satisfy the conditions necessary to be an x_0.

Proof   By Theorem 1, $\lim_{k \to \infty} A^k x_0$ is an element of span (x). By Theorem 2, the sum of the entries in x_0 equals the sum of the entries in $\lim_{k \to \infty} A^k x_0$. Using an arbitrary matrix B to map an arbitrary matrix C (taking the product BC) has the same effect as using B to map every vector c in C individually and then concatenating the products Bc_i.

From these three assertions, the following algebra can be constructed.

$$
\begin{aligned}
A^\infty &= \lim_{k \to \infty} A^k \\
&= \lim_{k \to \infty} A^k A \\
&= \left(\lim_{k \to \infty} A^k \right) (v_1 \cup v_2 \cup \cdots \cup v_n) \\
&= \lim_{k \to \infty} A^k v_1 \cup \lim_{k \to \infty} A^k v_2 \cup \cdots \cup \lim_{k \to \infty} A^k v_n \\
&= x \cup x \cup \cdots \cup x
\end{aligned}
$$

---

[6]For convenience only.

Although the symbol x is used in the conclusion of the above algebra, note that it is not just any element of span (x), but specifically the one whose entries sum to 1 that is represented. It is this element of span (x) because the sum of the entries of v_i is 1, implying by Theorem 2 that the sum of the entries of $x = \lim_{k \to \infty} A^k v_i$ is 1.

∎

After Corollary 3, it is clear that A^∞ has some interesting properties. As an interesting direct follow-up, because every vector in $C(A^\infty)$ is the same, dim $C(A^\infty) = 1$. More importantly, however, is that despite the fact that A is raised to the infinite power to generate A^∞, every entry of A^∞ is finite (because every entry of x is finite). This finding will be instrumental in proving the next corollary.

New corollaries coming soon!

Glossary

Diagonalize To factor an n-square matrix A into the form $S\Lambda S^{-1}$, where S is an n-square matrix containing the eigenvectors $x_1, x_2, \ldots, x_n$ (from left to right) of A and Λ is an n-square diagonal matrix containing the corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ (from top-left to bottom-right) of A. 2

Markov chain "A[n] [algebraic] system that experiences transitions from one state to another according to certain probabilistic rules" [Maltby et al., nd]. 1, 2, 4–7, 9, 10

Markov matrix An n-square matrix A such that every entry $a_{ij} \in [0, 1]$ and the sum of all entries a_{ij} in each column j is 1. Symbolically, $\sum_{i=1}^{n} a_{ij} = 1$ for all $j = 1, 2, \ldots, n$. 2, 5–8, 10

Permutation matrix A matrix with exactly one 1 in every row and column and zeroes everywhere else. 7

Singular Not invertible. Other characteristics of singular matrices include not being full column rank and having a nontrivial null space. 6

References

[Labalme, 2020] Labalme, S. (2020). Chapter 6: Eigenvalues. `https://github.com/shadypuck/LinAlgNotes/blob/master/Chapter%206/Chapter6.pdf`. Accessed: 2020-04-04.

[Maltby et al., nd] Maltby, H., Pakornrat, W., and Jackson, J. (n.d.). Markov chains. `https://brilliant.org/wiki/markov-chains/`. Accessed: 2020-04-04.