

Markov Chains and Their Application to Actuarial Science

Ronnie Kolodziej
MAT 2240 Section 102

What material that we learned in this course helps us understand Markov Chains?

The Matrix Equation $A\mathbf{x} = \mathbf{b}$

Fundamentally, the matrix equation $A\mathbf{x} = \mathbf{b}$ helps us understand how Markov chains work. Recollect the definition of this equation: If A is an $m \times n$ matrix, with columns $\mathbf{a}_1, \dots, \mathbf{a}_n$ and if \mathbf{x} is in \mathbb{R}^n , then the product of A and \mathbf{x} , denoted by $A\mathbf{x}$, is the linear combination of the columns of A using the corresponding

entries in \mathbf{x} as weights; that is, $A\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots + x_n\mathbf{a}_n$. In other words, the

product $A\mathbf{x}$ makes sense only when the columns of A equal the number of entries in \mathbf{x} . Furthermore, \mathbf{b} is a $m \times 1$ resultant column vector.

Next, we discussed how a system of linear equations could be represented in three different but equivalent ways: If A is an $m \times n$ matrix, with columns $\mathbf{a}_1, \dots, \mathbf{a}_n$ and if \mathbf{x} is in \mathbb{R}^n , the matrix equation $A\mathbf{x} = \mathbf{b}$ has the same solution set as the vector equation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots + x_n\mathbf{a}_n = \mathbf{b}$ which in turn has the same solution set as the system of linear equations whose augmented matrix is $\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n & \mathbf{b} \end{bmatrix}$. This is a very important theorem for constructing and computing a mathematical model. This allows us to switch to either of these methods when convenient for computational purposes.

A useful fact we learned was that the equation $A\mathbf{x} = \mathbf{b}$ has a solution if and only if \mathbf{b} is a linear combination of the columns of A . Consider the equation when A is I_2 (the 2×2 identity matrix):

$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 + 0x_2 \\ 0x_1 + x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. This example is the most obvious in explaining the

mentioned fact. As we can see, $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ is a scalar multiple of the columns of I_2 .

Remember one of the ways to compute the product $A\mathbf{x}$ is the row-vector rule, commonly known as the *dot product*. Consider the example: $\begin{bmatrix} 2 & 3 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ Let's first compute this problem by using the dot product. Each entry in the resultant \mathbf{b} is the sum of the products of the i th row in A and the entries of \mathbf{x} . Thus,

$\begin{bmatrix} 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 + 3x_2 \end{bmatrix}$ is our first entry of \mathbf{b} by the dot product method. In the same

way, the second entry is computed as: $\begin{bmatrix} 5 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5x_1 + x_2 \end{bmatrix}$. By the definition $A\mathbf{x} = \mathbf{b}$, $\begin{bmatrix} 2 & 3 \\ 5 & 1 \end{bmatrix}$

$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 2 \\ 5 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 5x_1 \end{bmatrix} + \begin{bmatrix} 3x_2 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 + 3x_2 \\ 5x_1 + x_2 \end{bmatrix}$. However, this is only applicable when

we want to know if a specific vector is a resultant of the matrix equation.

The properties of the matrix-vector product $A\mathbf{x}$ are that the product satisfies addition and scalar multiplication. Specifically, If A is an $m \times n$ matrix, \mathbf{u} and \mathbf{v} are vectors in \mathbb{R}^n , and c is a scalar, then: $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$ and $A(c\mathbf{u}) = c(A\mathbf{u})$.

So far, my examples have focused on square $n \times n$ matrices. In general, If A is an $n \times n$ matrix, and \mathbf{x} and \mathbf{b} are $n \times 1$ vectors, then $A\mathbf{x} = \mathbf{b}$ has 0, 1, or infinite solutions. This fact has much to do with the number of pivots in the coefficient matrix A . If the matrix A has n pivots, then the columns span \mathbb{R}^n and we have a unique solution \mathbf{b} . If A The matrix equation can have no solutions (0 solutions) if the augmented matrix $A\mathbf{x} = \mathbf{b}$ has inconsistency (i.e. in Gauss-Jordan form, we have a row such as $[0 \ 0 \ 0 \ \dots \ n]$).

We learned that we can solve for a general \mathbf{b} for $A\mathbf{x} = \mathbf{b}$. This was achieved by augmenting a matrix

A with a general \mathbf{b} , and subsequently using Gaussian elimination. Consider the example: $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$

with *(LinearAlgebra)* : $A := \text{Matrix}(\left[\left[1, 2, 3, b_1 \right], \left[4, 5, 6, b_2 \right], \left[7, 8, 9, b_3 \right] \right]);$

$$\begin{bmatrix} 1 & 2 & 3 & b_1 \\ 4 & 5 & 6 & b_2 \\ 7 & 8 & 9 & b_3 \end{bmatrix} \tag{1}$$

GaussianElimination(A);

$$\begin{bmatrix} 1 & 2 & 3 & b_1 \\ 0 & -3 & -6 & b_2 - 4b_1 \\ 0 & 0 & 0 & b_3 + b_1 - 2b_2 \end{bmatrix} \tag{2}$$

The fourth entry in the fourth column (the equal column) must equal zero in order for the system to be consistent. In other words, only some solutions will satisfy the system because $b_3 + b_1 - 2b_2$ must equal 0. This is a graphical description of \mathbf{b} , namely a plane in \mathbb{R}^3 through the origin. This description is equal to the Span of A where each \mathbf{b} is a linear combination of the columns of A .

Tenuous Relationship to Linear Transformations

We learned that the matrix equation $A\mathbf{x} = \mathbf{b}$ is not necessarily connected with linear combinations of vectors. Linear transformations consider a matrix A to act on a vector \mathbf{x} by multiplication to produce a new vector $A\mathbf{x}$. This is notated as: $\mathbf{x} \Rightarrow A\mathbf{x}$. Consider this example:

$$\begin{bmatrix} 1 & 5 & 6 \\ 2 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 + 5 + 6 \\ 2 + 3 + 1 \end{bmatrix} = \begin{bmatrix} 12 \\ 6 \end{bmatrix}. \text{ In this perspective, we say that } A \text{ acted on the vector } \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \text{ and}$$

transformed it into \mathbf{b} , a vector in \mathbb{R}^2 . Thus, solving the matrix equation $A\mathbf{x} = \mathbf{b}$ consists of finding all the vectors in \mathbb{R}^n that are transformed into the vector \mathbf{b} in \mathbb{R}^m by the multiplication of A , an $m \times n$ matrix.

Finding Eigenvalues

We learned that eigenvectors are nontrivial solution vectors of the equation $A\mathbf{x} = \lambda\mathbf{x}$, where A is an $n \times n$ matrix. The scalar λ corresponds to what is known as an eigenvalue of the matrix A . We can determine if a specified number (other than the trivial solution) is an eigenvalue of A by concluding if that number satisfies $\det(A - \lambda I)\mathbf{x} = \mathbf{0}$. One example that we saw presented us with a matrix A , and asked us to determine if 7 was an eigenvalue of A . We set up the equation $A\mathbf{x} = 7\mathbf{x}$. Then we subtracted $7\mathbf{x}$ from both sides to arrive at $A\mathbf{x} - 7\mathbf{x} = \mathbf{0}$. We then applied the multiplicative identity property: $A\mathbf{x} - 7I\mathbf{x} = \mathbf{0}$. We then distribute \mathbf{x}

out so that $(A - 7I)\mathbf{x} = \mathbf{0}$. Thus, $\left(\begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} - \begin{bmatrix} 7 & 0 \\ 0 & 7 \end{bmatrix} \right) = \begin{bmatrix} -6 & 6 \\ 5 & -5 \end{bmatrix}$. Finally, we compute the determinant,

hoping that it equals zero. $\det(A) = (-6)(-5) - (6)(5) = 30 - 30 = 0$. Thus, 7 must be an eigenvalue of the matrix A .

Discrete Dynamical Systems

Our study of predator-prey systems have an application to Markov chains because we used these systems to determine long-term behavior. As we learned, a discrete dynamical system is a sequence of vectors $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots$ related to one another by a square matrix A such that $\mathbf{x}_{k+1} = A\mathbf{x}_k$ for $k = 0, 1, 2, \dots$, where k refers to a specific unit of time, namely years. One example we covered dealt with spotted owls and wood rats in the redwood forests of California. The difference equations were:

$O_{k+1} = (0.4)O_k + (0.3)S_k$ and $R_{k+1} = (-p)O_k + (1.2)S_k$. We learned that $(0.4)O_k$ represented the survival rate of owl per month when no squirrels were available for food. In contrast, with no owls as predators, the $(1.2)S_k$ represents the 20% monthly growth of the squirrel population. The $(0.3)S_k$ represented the growth of the owl population when the squirrel population was plentiful. Lastly, $(-p)O_k$ measures the mortality rate of wood rats when they are hunted down by owls. The parameter p is the predation parameter. In Problem Set #4, we set this parameter to .5 to observe the long-term behavior of the spotted owls and squirrels. We entered these difference equations into matrices and solved for their eigenvalues and eigenvectors. Our results showed that the populations both declined over time, with the squirrels dying off first and then the owls.

Let $p = .5$.

```
> Ex5a := Matrix([[ (4/10) , (3/10) ], [ (-1/2) , (12/10) ]]);
```

$$Ex5a := \begin{bmatrix} \frac{2}{5} & \frac{3}{10} \\ -\frac{1}{2} & \frac{6}{5} \end{bmatrix} \quad (3)$$

```
> Eigenvectors (Ex5a) ;
```

$$\begin{bmatrix} \frac{9}{10} \\ \frac{7}{10} \end{bmatrix}, \begin{bmatrix} \frac{3}{5} & 1 \\ 1 & 1 \end{bmatrix} \quad (4)$$

The eigenvector decomposition is

$$\mathbf{x}_k = c_1(0.9)^k \begin{pmatrix} .6 \\ 1 \end{pmatrix} + c_2(.7)^k \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$
 \mathbf{x}_k approaches $\mathbf{0}$ as $k \rightarrow \infty$ because both $(0.9)^k$ and $(0.7)^k$ approach 0, though the squirrels will perish more quickly than the owls (Lay, 303). To stabilize the populations, we must choose the eigenvalue $\lambda = 1$.

What are Markov Chains?

Markov chains are trials which test whether vectors have either remained in the same state or have changed states. A Markov chain uses a square matrix called a **stochastic matrix** comprised of **probability vectors**. A **probability vector** is a vector with positive coefficients that add up to 1. A Markov chain, by notation, is a sequence of probability vectors $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots$ and a stochastic matrix P , such that $\mathbf{x}_1 = P\mathbf{x}_0, \mathbf{x}_2 = P\mathbf{x}_1$, etc. Using a difference equation, we can represent the Markov chain as $\mathbf{x}_{k+1} = P\mathbf{x}_k$ for $k = 0, 1, 2, \dots$. Because the stochastic matrix acts on the probability vectors, it is sometimes referred to as a *transition matrix*, or *Markov matrix*.

Consider the stochastic matrix $P = \begin{bmatrix} .3 & .7 \\ .7 & .3 \end{bmatrix}$. This matrix is what is known as a **regular** stochastic

matrix. What this means is that the matrix has strictly positive entries. Now, consider how the entries in both of the column probability vectors equal to 1. Because of this property, this matrix is considered a **left stochastic matrix**. However, consider how the entries in each of the row vectors equal 1 as well. Thus, the matrix is also a **right stochastic matrix**. When a stochastic matrix has both of these characteristics, it is what is known as a **doubly stochastic matrix**.

Sometimes, when a stochastic matrix acts upon a vector \mathbf{w} , the vector remains in the same state (\mathbf{w}). Notation-wise, this is shown as $P\mathbf{w} = \mathbf{w}$. What this means is that when the system is in state \mathbf{w} , there is no change in the system from one calculation to the next. This vector \mathbf{w} is what is known as the **steady-state** or **equilibrium vector**. Every stochastic matrix has a steady-state vector.

To find the steady-state vector of a given stochastic matrix A , we solve the equation $P\mathbf{w} = \mathbf{w}$. We subtract \mathbf{w} from both sides to arrive at $P\mathbf{w} - \mathbf{w} = \mathbf{0}$. Next, we apply the multiplicative identity property: $P\mathbf{w} - I\mathbf{w} = \mathbf{0}$. Finally, we distribute out the \mathbf{w} : $(P - I)\mathbf{x} = \mathbf{0}$. Let's apply this to the matrix P in the

aforementioned example. $(P - I)\mathbf{x} = \left(\begin{bmatrix} .3 & .7 \\ .7 & .3 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \mathbf{x} = \begin{bmatrix} -.7 & .7 \\ .7 & -.7 \end{bmatrix} \mathbf{x} = \mathbf{0}$. Now, we row reduce the

augmented matrix $[A \ \mathbf{0}]$. $B := \text{Matrix} \left(\left[\left[\left(-\frac{7}{10} \right), \left(\frac{7}{10} \right), 0 \right], \left[\left(\frac{7}{10} \right), \left(-\frac{7}{10} \right), 0 \right] \right] \right)$;

$$\begin{bmatrix} -\frac{7}{10} & \frac{7}{10} & 0 \\ \frac{7}{10} & -\frac{7}{10} & 0 \end{bmatrix} \tag{5}$$

$\text{ReducedRowEchelonForm}(B)$;

$$\begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{6}$$

We find that $\mathbf{x}_1 = \mathbf{x}_2$ and \mathbf{x}_2 is free. Thus, the general solution is $\mathbf{x}_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. It follows that one of our

steady-state vectors is then $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ (when $\mathbf{x}_2 = 1$).

A very important application, if not the most important, is that Markov chains can be used to predict the future in the long-term. A very interesting feature is that for each part of the sequence, the subsequent state relies only on the immediately previous vector. In other words, it does not rely on *all* of the previous state vectors, only the state vector that anteceded it. This is reflected in the following theorem: "If P is an $n \times n$ regular stochastic matrix, then P has a unique steady-state vector \mathbf{q} . Further, if \mathbf{x}_0 is any initial state and $\mathbf{x}_{k+1} = P\mathbf{x}_k$ for $k = 0, 1, 2, \dots$, then the Markov chain $\{\mathbf{x}_k\}$ converges to \mathbf{q} as $k \rightarrow \infty$ " (Lay, 259).

Markov Chains in Actuarial Science

Bonus-Malus Systems (BMS) are used throughout the insurance market, though most conspicuously in the auto insurance sector. Bonus-Malus in Latin means "good-bad." In the context of insurance, it is a system that is used to reward or penalize driving classes based on an annual claims history. The entries in the matrix are designed to put policyholders into specific homogeneous classes based on driving performance (i.e. claim frequency of the driver).

The entries in the matrix represent the number of claims required to increase or decrease the discount paid by the classes of drivers. The vertical axis represents the new discounts and maluses (surcharges, penalties, etc.), where the middle point on the axis is the class where neither discounts nor surcharges are assessed to premiums. The horizontal axis represents the current discounts and maluses assessed on a class of drivers. This matrix is then applied to certain vectors that represent driver classes and their claim history for the year. This then adjusts their premiums based on their performance (i.e. if they had claims filed against them).

As you can see, most of my presentation consists of my understanding of what a Markov chain is and how it applies to linear algebra. If I had more time, I would like to find sources that are more readable and more approachable to this application of linear algebra in actuarial science. Most of the literature that I consulted used series and various statistical computations/methods to explain the Markov chain process.

References

Conversation with Dr. Kevin Shirley on April 30. This conversation helped me understand the basic application of Markov chains and state vectors to Bonus-Malus systems.

Lay, David. Section 4.9: "Applications to Markov Chains." *Linear Algebra and Its Applications*. Pgs. 253-259. This section helped with defining and understanding how Markov chains work. From this source, I learned what stochastic matrices and probability vectors are, the notation of each subsequent vector in relation to the stochastic matrix and the previous vector (254). Also, I learned from this source what a steady-state vector is and how to solve for one (258). Furthermore, the theorem regarding convergence was used from this source (259).

"Doubly Stochastic Matrix." Wolfram Mathworld. This helped with defining the characteristics of these rare forms of stochastic matrices.

Neuhaus, Walther (1988). "A bonus-malus system in auto insurance." This source was foundational to understanding the fundamental applications of transition (probability) matrices to Bonus-Malus systems. This enumerated what each side of the matrix represent and how the entries act to move or sustain the state vectors.