The material from this lesson is drawn from "Chapter 11: Markov Chains" and Dr. Robert Talbert's Spring 2010 MAT 233 Linear Algebra course at Franklin College.

I. Terminology

1. States, Time, and Steps

   Consider a set of outcomes $\mathcal{S} = \{s_1, \ldots, s_r\}$. We call $\mathcal{S}$ the state space. (It is possible to have a continuous state space, but we'll assume that the state space is discrete.)

   Consider random variables $X_1, \ldots, X_n$. We call $1, \ldots, n$ the time index. (It is possible to have a continuous time index, but we'll assume that our time index is always discrete.)

   Our random variables $X_1, \ldots, X_n$ can only take on the values of our state space $\mathcal{S}$. Each move is called a "step."

   Let's go through an example. Suppose that the weather on any given day can be characterized as clear, rainy, or snowy.

   - Our state space $\mathcal{S}$ can be written $\{C, R, S\}$, where 'C', 'R', and 'S' are clear, rainy, and snowy, respectively.
   - If "day one" is clear, we write $X_1 = C$.
   - If "day two" is rainy, we write $X_2 = R$.
   - We call the move from X_1 to X_2 a "step."

   Two questions of interest:

   i. Given a particular state, can we predict what the random variables will look like k steps from now?
   ii. Can we estimate what the long-run behavior of our sequence of random variables will be?

2. Markov Property

   $$\mathbb{P}(X_{n+1} = x_i | X_n = x_j) = \mathbb{P}(X_{n+1} = x_i | X_n = x_j, X_{n-1} = x_k, \ldots, X_1 = x_z)$$

   In English, knowing X_n gives us as much information about X_{n+1} as we would get from knowing $X_1, \ldots, X_n$. In other words, we might say that knowing $X_1, \ldots, X_{n-1}$ gives us no additional information about X_{n+1}.

   Some applications of this property:

   - Genetics. Knowing someone's genotype (AA,Aa,aa) about a particular trait gives you all of the information you can get about their potential child's genotype; knowing the parents' ancestors' genotypes provides no additional information.
   - Monopoly. Knowing someone's current location on the Monopoly game board gives you all of the information about where they can go on their next turn; knowing where they have previously been on the board provides no additional information.
   - Baseball. Knowing the current positions of runners and number of outs gives you all of the information about what outcomes can occur next; knowing the previous positions of the runners and number of outs provides no additional information.
   - PageRank. Google's PageRank algorithm uses the current position of an individual on the Web to estimate where that individual will travel next; knowing the previous positions of that individual provides no additional information.

3. Transition Matrix

   Consider our example above, where the weather on any given day can be characterized as clear, rainy, or snowy. Recall that the state space $\mathcal{S} = \{C, R, S\}$. Let's assume the following:

   - If $X_i = C$, then $\mathbb{P}(X_{i+1} = C) = 0.6$, $\mathbb{P}(X_{i+1} = R) = 0.3$, $\mathbb{P}(X_{i+1} = S) = 0.1$.
   - If $X_i = R$, then $\mathbb{P}(X_{i+1} = C) = 0.4$, $\mathbb{P}(X_{i+1} = R) = 0.4$, $\mathbb{P}(X_{i+1} = S) = 0.2$.
   - If $X_i = S$, then $\mathbb{P}(X_{i+1} = C) = 0.3$, $\mathbb{P}(X_{i+1} = R) = 0.3$, $\mathbb{P}(X_{i+1} = S) = 0.4$.

Let's further assume that the Markov property holds in this situation. Then the matrix

$$\mathbf{A} = \begin{array}{cc} \begin{matrix} C & \quad R & \quad S \end{matrix} & \\ \begin{bmatrix} 0.6 & 0.4 & 0.3 \\ 0.3 & 0.4 & 0.3 \\ 0.1 & 0.2 & 0.4 \end{bmatrix} & \begin{matrix} \mathrm{C} \\ \mathrm{R} \\ \mathrm{S} \end{matrix} \end{array}$$

is called the transition matrix, which represents how we move from one state to another.

Note that each column has non-negative entries that sum to one. When this is the case, we refer to each column as a probability vector and call the entire matrix a stochastic matrix. (This will always be the case for a valid transition matrix.)

Suppose it is snowy today. While we can just look at the S column to see what is likely to happen tomorrow, let's express this as the vector $\mathbf{y}_1 = [0 \;\; 0 \;\; 1]^T$. Then we can find $\mathbf{y}_2$ by evaluating $\mathbf{Ay}_1$:

$$\begin{aligned}
\mathbf{Ay}_1 &= \begin{bmatrix} 0.6 & 0.4 & 0.3 \\ 0.3 & 0.4 & 0.3 \\ 0.1 & 0.2 & 0.4 \end{bmatrix} \times \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} 0 \times 0.6 + 0 \times 0.4 + 1 \times 0.3 \\ 0 \times 0.3 + 0 \times 0.4 + 1 \times 0.3 \\ 0 \times 0.1 + 0 \times 0.2 + 1 \times 0.4 \end{bmatrix} \\
&= \begin{bmatrix} 0.3 \\ 0.3 \\ 0.4 \end{bmatrix} \\
&= \mathbf{y}_2
\end{aligned}$$

This makes sense - if it's snowy today, there's a 30% chance of clear skies tomorrow, a 30% chance of rain, and a 40% chance of snow. This was easy, but let's find $\mathbf{y}_3$.

$$\begin{aligned}
\mathbf{Ay}_2 &= \begin{bmatrix} 0.6 & 0.4 & 0.3 \\ 0.3 & 0.4 & 0.3 \\ 0.1 & 0.2 & 0.4 \end{bmatrix} \times \begin{bmatrix} 0.3 \\ 0.3 \\ 0.4 \end{bmatrix} \\
&= \begin{bmatrix} 0.3 \times 0.6 + 0.3 \times 0.4 + 0.4 \times 0.3 \\ 0.3 \times 0.3 + 0.3 \times 0.4 + 0.4 \times 0.3 \\ 0.3 \times 0.1 + 0.3 \times 0.2 + 0.4 \times 0.4 \end{bmatrix} \\
&= \begin{bmatrix} 0.18 + 0.12 + 0.12 \\ 0.09 + 0.12 + 0.12 \\ 0.03 + 0.06 + 0.16 \end{bmatrix} \\
&= \begin{bmatrix} 0.42 \\ 0.33 \\ 0.25 \end{bmatrix} \\
&= \mathbf{y}_3
\end{aligned}$$

If it's snowing today, there's a 42% chance of clear skies in two days, a 33% chance of rain, and a 25% chance of snow.

What will happen if we look at $\mathbf{y}_n$ for large n? What happens if we change $\mathbf{y}_1$?