Lecture Notes: Markov chains

Tuesday, September 16

Dannie Durand

In the last lecture, we introduced Markov chains, a mathematical formalism for modeling how a random variable progresses over time. We introduced the following notation for describing the properties of a Markov chain:

- A Markov chain has states \( E_0, E_1, \ldots, E_s \) corresponding to the range of the associated random variable.
- \( \varphi_j(t) \) is the probability that the chain is in state \( E_j \) at time \( t \). The vector \( \varphi(t) = (\varphi_1(t), \ldots, \varphi_s(t)) \) is the state probability distribution at time \( t \).
- \( \pi = \varphi(0) \) is the initial state probability distribution.
- \( P \) is the transition probability matrix. \( P_{jk} \) gives the probability of making a transition to state \( E_k \) at time \( t + 1 \), given that the chain was in state \( E_j \) at time \( t \). The rows of this matrix sum to one (\( \sum_k P_{jk} = 1 \)).
- The state probability distribution at time \( t + 1 \) is given by \( \varphi(t + 1) = \varphi(t) \cdot P \). The probability of being in state \( E_j \) at \( t + 1 \) is

\[
\varphi_j(t + 1) = \sum_j \varphi_j(t) P_{jk}
\]  

(1)

- The Markov property states that Markov chains are memoryless. The probability that the chain is in state \( E_k \) at time \( t + 1 \), depends only on \( \varphi(t) \) and is independent of \( \varphi(t - 1), \varphi(t - 2), \varphi(t - 3) \ldots \)

In this course, we will focus on discrete, finite, time-homogeneous Markov chains. These are models with a finite number of states, in which time (or space) is split into discrete steps. A Markov chain is time-homogeneous if the transition matrix does not change over time.

Absorbing states

Last Thursday, we considered a Markov chain to model the position of a drunk moving back and forth on a railroad track on top of a mesa. When the drunk reaches either end of the railway (either the \( 0^{th} \) or the \( 4^{th} \) tie), he falls off the mesa. In the Markov model, states \( E_0 \) and \( E_4 \) are absorbing states: Once the system enters one of these states, it remains in that state forever, because \( P_{00} = P_{44} = 1 \). Our model of the drunk is an example of a random walk with absorbing boundaries.
The transition matrix of a Markov chain can be represented as a graph, where the nodes represent states and the edges represent transitions with non-zero probability. For example, the random walk with absorbing boundaries can be modeled like this:

**Periodic Markov chains**

In order to save the drunk from an early death, we introduced a random walk with *reflecting* boundaries. At each step, the drunk moves to the left or to the right with equal probability. When the drunk reaches one of the boundary states ($E_0$ or $E_4$), he returns to the adjacent state ($E_1$ or $E_3$) at the next step, with probability one. This yields the following transition probability matrix:

$$
\begin{array}{c|cccc}
 & 0 & 1 & 2 & 3 \\
\hline
0 & 0 & 1 & 0 & 0 \\
1 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\
2 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\
3 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\
4 & 0 & 0 & 0 & 1 & 0 \\
\end{array}
$$

and can be represented graphically like this:
The initial state probability distribution and the state distribution for the first two time steps are the same in both random walk models, namely

\[ \varphi(0) = (0, 0, 1, 0, 0) \]
\[ \varphi(1) = (0, \frac{1}{2}, 0, \frac{1}{2}, 0) \]
\[ \varphi(2) = \left( \frac{1}{4}, 0, \frac{1}{2}, 0, \frac{1}{4} \right). \]

We calculate the state probability distribution at \( t = 3 \) by multiplying the vector \( \varphi(2) \) times the matrix \( P \):

\[ \varphi(3) = \varphi(2) \cdot P \]
\[ = \left( \frac{1}{4}, 0, \frac{1}{2}, 0, \frac{1}{4} \right) \cdot P \]
\[ = \left( 0, \frac{1}{2}, 0, \frac{1}{2}, 0 \right). \]

This demonstrates that \( \varphi(3) = \varphi(1) \). Similarly, \( \varphi(4) = \varphi(2) \) as can be seen from the following calculation:

\[ \varphi(4) = \varphi(3) \cdot P \]
\[ = \left( 0, \frac{1}{2}, 0, \frac{1}{2}, 0 \right) \cdot P \]
\[ = \left( \frac{1}{4}, 0, \frac{1}{2}, 0, \frac{1}{4} \right). \]

From this we can see that the probability state distribution will be \( (0, \frac{1}{2}, 0, \frac{1}{2}, 0) \) at all odd time steps and \( \left( \frac{1}{4}, 0, \frac{1}{2}, 0, \frac{1}{4} \right) \) at all even time steps. Thus, the random walk with reflecting boundaries is a periodic Markov chain. A Markov chain is periodic if there is some state that can only be visited in multiples of \( m \) time steps, where \( m > 1 \).

We do not require periodic Markov chains for modeling sequence evolution and will only consider aperiodic Markov chains going forward.

### Stationary distributions

A state probability distribution, \( \varphi^* \), that satisfies the equation

\[ \varphi^* = \varphi^* P \tag{2} \]

is called a stationary distribution. A key question for a given Markov chain is whether such a stationary distribution exists. Equation 2 is equivalent to a system of \( s \) equations in \( s \) unknowns.
One way to determine the steady state distribution is to solve that system of equations. The stationary distribution can also be obtained using matrix algebra, but that approach is beyond the scope of this course.

The random walk with reflecting boundaries clearly does not have a stationary distribution, since every state with non-zero probability at time $t$ has zero probability at time $t+1$. The random walk with absorbing boundaries does have a stationary distribution, but it is not unique. For example, both $(1,0,0,0,0)$ and $(0,0,0,0,1)$ are stationary distributions of the random walk with absorbing boundaries.

For the rest of this course, we will concern ourselves only with aperiodic Markov chains that do not have absorbing states. In fact, we will make an even stronger assumption and restrict our consideration to Markov chains in which every state is connected to every other state via a series of zero or more states. If a finite Markov chain is aperiodic and connected in this way, it has a unique stationary distribution. We will not attempt to prove this or even to state the theorem in a rigorous way. That is beyond the scope of this class. For those who are interested, a very nice treatment can be found in Chapter 15 of *Probability Theory and its Applications (Volume I)* by William Feller (John Wiley & Sons).

As an example of a Markov chain with a unique stationary distribution, we introduced a random walk that has neither absorbing, nor reflecting boundaries. In this third random walk model, if the drunk is in one of the boundary states ($E_0$ or $E_4$) at time $t$, then at time $t+1$ he either remains in the boundary state or returns to the adjacent state ($E_1$ or $E_3$). In the model we considered in class, we assigned a probability of 0.5 to each of these events, resulting in the following state transition matrix, $P$:

$$
\begin{array}{c|cccc}
  & 0 & 1 & 2 & 3 \\
\hline
0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\
1 & \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2} \\
2 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\
3 & \frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{2} \\
4 & 0 & 0 & 0 & 1 \\
\end{array}
$$

We can determine the stationary state distribution for this random walk model by substitution this
transition matrix into Equation 2. The probability of being in state $E_0$ is

$$\varphi_0 = \sum_{j=0}^{4} \varphi_j P_{j0}$$

$$= \varphi_0 P_{00} + \varphi_1 P_{10} + \varphi_2 P_{20} + \varphi_3 P_{30} + \varphi_4 P_{40}$$

$$= \frac{1}{2} \varphi_0 + \frac{1}{2} \varphi_1,$$

since $P_{20}$, $P_{30}$ and $P_{40}$ are all equal to zero. The other steady state probabilities are derived similarly, yielding

$$\varphi_0 = \frac{1}{2} \varphi_0 + \frac{1}{2} \varphi_1$$  \hspace{1cm} (3)$$

$$\varphi_1 = \frac{1}{2} \varphi_0 + \frac{1}{2} \varphi_2$$  \hspace{1cm} (4)$$

$$\varphi_2 = \frac{1}{2} \varphi_1 + \frac{1}{2} \varphi_3$$  \hspace{1cm} (5)$$

$$\varphi_3 = \frac{1}{2} \varphi_2 + \frac{1}{2} \varphi_4$$  \hspace{1cm} (6)$$

$$\varphi_4 = \frac{1}{2} \varphi_3 + \frac{1}{2} \varphi_4.$$  \hspace{1cm} (7)$$

In addition, the probability that the system is in some state is unity, imposing an additional constraint:

$$\varphi_0 + \varphi_1 + \varphi_2 + \varphi_3 + \varphi_4 = 1.$$  \hspace{1cm} (8)$$

The model has a stationary distribution if the above equations have a solution. In class, we showed that Equations 3 - 7 reduce to $\varphi_0 = \varphi_1 = \varphi_2 = \varphi_3 = \varphi_4$. Applying the constraint in Equation 8, we see that the solution to the above equations must be $\varphi^* = (0.2, 0.2, 0.2, 0.2, 0.2)$.

If we know the stationary state distribution, or have an educated guess, we can verify that it indeed satisfies Equation 2. For example, it is easy to verify that $(0.2, 0.2, 0.2, 0.2, 0.2) \cdot P = (0.2, 0.2, 0.2, 0.2, 0.2)$.  
