EE 520: Random Processes	**Fall 2020**

Lecture 16
Markov Chains

Instructor Name: John Lipor

Recommended Reading: Pishro-Nik: 11.1.1 - 11.2.5; Gubner: 12.1 - 12.5

1 Markov Chains

Continuing our study of classes of RPs that are generally useful, we will spend the last few lectures discussing processes with the *Markov property*.

Definition 1. Markov property: Conditioned on the present, the future is independent of the past. Formally, for a sequence of integer-valued RVs $X_0, X_1, \ldots,$ we have

$$P\left(X_{n+1} = i_{n+1} \mid X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\right) = P\left(X_{n+1} = i_{n+1} \mid X_n = i_n\right).$$

Definition 2. A sequence of integer-valued RVs that satisfy the Markov property is called a **Markov chain** (MC).

Example 1. Problem 5 of Homework 1 and Problem 1 of Homework 2 both describe Markov chains. In each case, the probability of passing to a given state depends only on the current state and not how the system arrived at that state.

Example 2. This process is known as a *random walk*. Let $X_0 \in \mathbb{Z}$ be a RV and let Z_i be a sequence of i.i.d. RVs with

$$P_Z(z) = \begin{cases} 1, & \text{with probability } \alpha \\ 0, & \text{with probability } 1 - \alpha - \beta \\ -1, & \text{with probability } \beta. \end{cases}$$

Then $X_n = X_{n-1} + Z_n$ is a Markov chain. To see this, note that

$$\begin{aligned} X_1 &= X_0 + Z_1 \\ X_2 &= X_1 + Z_2 = X_0 + Z_1 + Z_2 \\ &\vdots \\ X_n &= X_{n-1} + Z_n = X_0 + Z_1 + Z_2 + \cdots + Z_n, \end{aligned}$$

so Z_{n+1} and $X_0, \ldots, X_n$ are independent. Therefore

$$\begin{aligned} P\left(X_{n+1} = i_{n+1} \mid X_n = i_n, \ldots, X_0 = i_0\right) &= P\left(X_n + Z_{n+1} = i_{n+1} \mid X_n = i_n, \ldots, X_0 = i_0\right) \\ &= P\left(Z_{n+1} = i_{n+1} - i_n \mid X_n = i_n, \ldots, X_0 = i_0\right) \\ &= P\left(Z_{n+1} = i_{n+1} - i_n\right), \end{aligned}$$

where the last line follows by independence of Z_{n+1} and $X_0, \ldots, X_n$. The above depends only on i_n and i_{n+1}. To show the random walk is Markov, we need one more tool.

Lemma 1. Given any event A and two integer-valued RVs X and Y, if $P(A|X = i, Y = j) = h(i)$, then

$$P\left(A \mid X = i, Y = j\right) = P(A \mid X = i).$$

Proof. We want to show that $h(i) = P(A|X = i)$. Write

$$
\begin{aligned}
P\left(A \cap \{X = i\}\right) &= \sum_j P\left(A \cap \{X = i\} \mid Y = j\right) P(Y = j) \\
&= \sum_j P\left(A \mid X = i, Y = j\right) P(X = i \mid Y = j) P(Y = j) \\
&= \sum_j h(i) P(X = i \mid Y = j) P(Y = j) \\
&= h(i) P(X = i).
\end{aligned}
$$

The above implies that

$$
h(i) = \frac{P(A \cap \{X = i\}}{P(X = i)} = P(A \mid X = i).
$$

□

Now let $A = \{X_{n+1} = i_{n+1}\}$, $X = \{X_n = i_n\}$, and $Y = \{X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\}$ and apply the above lemma to see that

$$
P\left(X_{n+1} = i_{n+1} \mid X_n = i_n, \ldots, X_0 = i_0\right) = h(i_n)
$$

and therefore

$$
P\left(X_{n+1} = i_{n+1} \mid X_n = i_n, \ldots, X_0 = i_0\right) = P\left(X_{n+1} = i_{n+1} \mid X_n = i_n\right),
$$

which shows that the random walk is Markov as desired.

Fact 1. The Markov property for a joint PDF is given below

$$
P\left(X_{n+m} = j_m, \ldots, X_{n+1} = j_1 \mid X_n = i_n, \ldots, X_0 = i_0\right) = P\left(X_{n+m} = j_m, \ldots, X_{n+1} = j_1 \mid X_n = i_n\right).
$$

2 State Space Models

Below we define some terminology that is used when analyzing MCs.

Definition 3. The set of possible values the RVs X_n can take is called the **state space** of the MC.

Definition 4. The conditional probabilities $P(X_{n+1} = j|X_n = i)$ are called the **transition probabilities**.

Definition 5. When the values of the transition probabilities do not depend on the time n, we say they are **stationary** or that the MC is **time homogeneous**. In this case, we write

$$
p_{ij} = P\left(X_{n+1} = j \mid X_n = i\right).
$$

Stationary MCs are accompanied by a **state transition diagram** and a **transition matrix** P whose (i, j)th entry is p_{ij}.

Example 3. Consider the MC with four states $0, 1, 2, 3$. Draw the state diagram for the transition probability matrix

$$
P = \begin{bmatrix} 4/6 & 1/6 & 1/6 & 0 \\ 0 & 1 & 0 & 0 \\ 4/6 & 1/6 & 0 & 1/6 \\ 0 & 0 & 0 & 1 \end{bmatrix}.
$$

Note that each row of a transition matrix must sum to 1, since it denotes the total probability of transitioning to *any* other state. When we write P as a matrix, we are implicitly assuming there are a finite number of states, but this does not have to be the case.

Definition 6. If $p_{ii} = 0$, we call state i **reflecting**. If $p_{ii} = 1$, we call state i **absorbing**.

The transition probabilities above describe the behavior of a single step of the process. We can also generalize this to m steps.

Definition 7. The m-step transition probability $p_{ij}^{(m)}$ is the probability of reaching state j from state i after m steps

$$p_{ij}^{(m)} = P\left(X_m = j \mid X_0 = i\right).$$

In particular, we can write $p_{ij} = p_{ij}^{(1)}$. Suppose we wish to find $p_{ij}^{(n+m)}$. It turns out the m-step transition probabilities satisfy the following equation.

Definition 8. The m-step transition probabilities are characterized by the **Chapman-Kolmogorov equation**

$$p_{ij}^{(n+m)} = \sum_k p_{ik}^{(n)} p_{kj}^{(m)},$$

or in its matrix form

$$P^{(n+m)} = P^{n+m}.$$

The above allows us to calculate the distribution over the states at a given time. Note that

$$\underbrace{P(X_n = j)}_{\rho_j^{(n)}} = \sum_i \underbrace{P\left(X_n = j \mid X_0 = i\right)}_{p_{ij}^{(n)}} \underbrace{P(X_0 = i)}_{\nu_i}.$$

Breaking our usual assumption that all vectors are columns, let ν be the *row* vector of probabilities over the states at time 0, and let $\rho^{(n)}$ be the distribution over all states at time n. Then

$$\rho^{(n)} = \nu P^n.$$

3 Stationary Distributions

A natural question is to ask whether things settle if we run the experiment long enough, i.e., does the distribution over states converge?

Definition 9. Let $\pi \in \mathbb{R}^N$ be a valid distribution over the N states in a stationary MC, and suppose that

$$\pi = \pi P, \quad \sum_{j=1}^{N} \pi_j = 1, \quad \pi_j \geq 0, \ \forall j.$$

Then π is called the **stationary** or **equilibrium** distribution of the MC.

Note that the equation $\pi = \pi P$ implies that π is a left eigenvector of P with corresponding eigenvalue 1. When dealing with an infinite number of states, we cannot write down the transition matrix and instead use the form

$$\pi_j = \sum_k p_{kj}.$$

Example 4. Consider the MC defined by the transition matrix

$$P = \begin{bmatrix} 0 & 1/4 & 3/4 \\ 0 & 1/2 & 1/2 \\ 2/5 & 2/5 & 1/5 \end{bmatrix}.$$

The stationary distribution then satisfies

$$\pi_0 = \pi_0 p_{00} + \pi_1 p_{10} + \pi_2 p_{20} = \frac{2}{5}\pi_2$$

and

$$\pi_1 = \frac{1}{4}\pi_0 + \frac{1}{2}\pi_1 + \frac{2}{5}\pi_2.$$

Combining these two equations shows that

$$\pi_1 = \pi_2.$$

Since π is a distribution, we also have that

$$\pi_0 + \pi_1 + \pi_2 = 1.$$

Solving this system of equations, we find the stationary distribution to be

$$\pi = \begin{bmatrix} 1/6 & 5/12 & 5/12 \end{bmatrix}.$$

Another question we may ask is when a stationary distribution exists. We typically only consider chains where it does exist, but even if it does, the stationary distribution may not be unique.

Definition 10. If for every pair of states $i \neq j$ there is a path from i to j and from j to i, we say the MC is **irreducible**.

Fact 2. If a chain is irreducible, then it has a unique stationary distribution.