Outline: In this lecture we discussed some properties of discrete-time Markov chains. We discussed several results on Transition Probability Matrices, Memorylessness, and Hitting time.

5.1 Discrete Time Markov Chains

Definition 5.1. Consider a discrete time stochastic process \((X_0, X_1, X_2, \ldots)\) defined on an underlying probability space \((\Omega, \mathcal{F}, \mathbb{P})\). Hence, we have \(X_i : (\Omega, \mathcal{F}, \mathbb{P}) \to (I, \xi)\). The random process is called a discrete-time Markov chain if,

1. \(X_0 \sim \lambda\), where \(\lambda\) is the initial distribution

2. \(\mathbb{P}(X_n = j | X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = i) = \mathbb{P}(X_n = j | X_{n-1} = i) = p_{ij}\), implies, given \(X_{n-1}\), the random number \(X_n\) is independent of the history, \(X_0, X_1, \ldots, X_{n-2}\).

\(P = ((p_{ij}))\) is called the transition probability matrix.

- Discrete time Markov chain is denoted as \((X_i)_{i \geq 0} \sim\) Markov(\(\lambda, P\)).
- If \(p_{ij}\)'s are independent of time (consequently transition probability matrix \(P\) is also independent of time), then, the Markov chain is called time homogeneous Markov chain.

Examples

Lazy Random Walk: Consider an integer line and a Markov process \((X_i)_{i \geq 0}\). The random variables take integer values, i.e. \(I = \mathbb{Z}\). Consider \(X_0 = 1\) (or some distribution \(\lambda\)). The value of random variable in the next step increases by 1 with probability 0.25, decreases by 1 with probability 0.25 and stays the same with probability 0.5. This can be formally written as:

\[
p_{ij} = \begin{cases} 
0.25, & j = i + 1 \\
0.25, & j = i - 1 \\
0.50, & j = i \\
0.00, & \text{otherwise}
\end{cases}
\]

The transition probability matrix, \(P\), has a special structure (tridiagonal matrix) in this example. \((X_0)_{i \geq 0} \sim\) Markov(\(\lambda, P\)).

Lazy Random Walk on Finite Line: Consider an integer line of finite length and a Markov process \((X_i)_{i \geq 0}\). The random variables take integer values, i.e. \(I = \{-N, -(N-1), \ldots, -1, 0, 1, \ldots, N\}\).
Consider $X_0 = 1$ (or some distribution $\lambda$). The value of random variable in the next step increases by 1 with probability 0.25, decreases by 1 with probability 0.25 and stays the same with probability 0.5. Moreover once the random variable reaches the end, $\{-N\}$ or $\{N\}$, the random variable stays there with probability 1. We can write the transition probability matrix as,

$$P = \begin{bmatrix}
1 & 0 & 0 & 0 & \ldots & 0 \\
0.25 & 0.5 & 0.25 & 0 & \ldots & 0 \\
0 & 0.25 & 0.5 & 0.25 & \ldots & 0 \\
0 & 0 & 0.25 & \ddots & \ddots & 0 \\
\vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & 0 & \ldots & 1
\end{bmatrix}.$$

This process, denoted as $(X_t)_{t\geq 0}$ Markov($\lambda, P$), is also called Lazy Random Walk. It is called lazy since the random variable stays the same with probability 1/2.

**Pattern Recognition:** Consider a random variable $Y$ which takes value 1 with probability $p$ and takes value 0 with probability $1 - p$. $Y$ can be considered a Bernoulli random variable with parameter $p$, i.e. $Y \sim \text{Ber}(p)$. Clearly, $Y$ is not a Markov chain, since the random variable at next step is independent of current value of random variable.

Now let us focus at finding the pattern $S = (1, 1)$, i.e. the value of two successive random variables is 1. Now we consider new random variable $X$ defined as $X_0 = (Y_0, Y_1), X_1 = (Y_1, Y_2), \ldots$. The random variable $X$ takes values in set $I = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$. The transition probabilities for $X$ can be written as,

$$P(X_n = (c, d)|X_{n-1} = (a, b)) = \begin{cases}
0, & \text{if } b \neq c \\
p, & \text{if } b = c \text{ and } d = 1 \\
1 - p, & \text{if } b = c \text{ and } d = 0
\end{cases} \rightarrow P = \begin{bmatrix}
1 - p & p & 0 & 0 \\
0 & 0 & 1 - p & p \\
1 - p & p & 0 & 0 \\
0 & 0 & 1 - p & p
\end{bmatrix}.$$
The random process \((X_i)_{i \geq 0}\) is Markov and denoted as \((X_i)_{i \geq 0} \sim \text{Markov}(\lambda, P)\).

### 5.1.1 Properties and Remarks

**Theorem 5.2.** Given a distribution \(\lambda\) and a transition probability matrix \(P\), and \((X_n, n \geq 0) \sim \text{Markov}(\lambda, P)\) if and only if, 
\[
P(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) = \lambda(x_0)p_{x_0x_1}p_{x_1x_2}\cdots p_{x_{n-1}x_n}
\]
for all \(x_0, x_1, \ldots, x_n\).

**Proof.** We first “only if” statement. \((\Leftarrow)\)

1) \(X_0 \sim \lambda\)  

(Markov Chain definition)

2) Next we start from the definition of conditional probability of \(X_n\) given history and prove Markov property for \((X_n, n \geq 0)\).

\[
P(X_n = x_n|X_0 = x_0, \ldots, X_{n-1} = x_{n-1}) = \frac{P(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n)}{P(X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1})}
\]

\[
= \frac{\lambda(x_0)p_{x_0x_1}p_{x_1x_2}\cdots p_{x_{n-1}x_n}}{\lambda(x_0)p_{x_0x_1}p_{x_1x_2}\cdots p_{x_{n-2}x_{n-1}}}
\]

\[
= p_{x_{n-1}x_n}
\]

\[
= P(X_n = x_n|X_{n-1} = x_{n-1})
\]

This proves that if \(P(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) = \lambda(x_0)p_{x_0x_1}p_{x_1x_2}\cdots p_{x_{n-1}x_n}\) for all \(x_0, x_1, \ldots, x_n\) then \((X_n, n \geq 0) \sim \text{Markov}(\lambda, P)\).

Now we prove the “if” statement. \((\Rightarrow)\)

We begin with the definition of conditional probability followed by the Markov property.

\[
P(X_0 = x_0, \ldots, X_n = x_n) = P(X_n = x_n|X_0 = x_0, \ldots, X_{n-1})P(X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1})
\]

\[
= p_{x_{n-1}x_n}P(X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1})
\]
We recursively use Markov property. Specifically, in the above expression we try to rewrite \(P(X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1})\) as a conditional probability followed by use of Markov property.

\[
P(X_0 = x_0, \ldots, X_n = x_n) = p_{x_{n-1}x_n}P(X_{n-1} = x_{n-1}|X_0 = x_0, \ldots, X_{n-2} = x_{n-2})
\]

\[
P(X_0 = x_0, X_1 = x_1, \ldots, X_{n-2} = x_{n-2}) = p_{x_{n-1}x_n}P(X_{n-2} = x_{n-2})
\]

And so on. We get,

\[
P(X_0 = x_0, \ldots, X_n = x_n) = \lambda(x_0)p_{x_0x_1}p_{x_1x_2} \cdots p_{x_{n-1}x_n}.
\]

We have proved that if \((X_n, n \geq 0)\) is Markov chain then \(P(X_0 = x_0, \ldots, X_n = x_n) = \lambda(x_0)p_{x_0x_1}p_{x_1x_2} \cdots p_{x_{n-1}x_n}\) is true.

We next consider some consequences of Markov property.

**Definition 5.3.** We call \(\delta_i\) as a degenerate distribution at \(i\) if \(X \sim \delta_i \iff P(X = i) = 1\).

**Remark 5.4.** As a consequence of the Markov property (Theorem 5.2), we can show that if \((X_n, n \geq 0)\) \(\sim\) Markov\((\lambda, P)\), then given \(X_k = i\), we have \((X_n, n \geq k)\) \(\sim\) Markov\((\delta_i, P)\).

This remark reinforces the idea that Markov chains don’t have memory (Memorylessness). If we consider the state of a Markov chain at any arbitrary time instant, the random process starting that instant is also Markov. It has the same transition probability matrix, however, the initial distribution is a degenerate distribution.

**Lemma 5.5.** If \(P\) and \(Q\) are transition matrices on \(I\), then \(PQ\) is a transition matrix on \(I\).

**Proof.** The entries of product matrix \(PQ\) can be written as -

\[
(PQ)_{ij} = \sum_k p_{ik} q_{kj} \geq 0 \quad (\forall ij)
\]

The sum of all entries in a row is computed as the sum of \((PQ)_{ij}\) over all \(j\).

\[
\sum_j (PQ)_{ij} = \sum_j \sum_k p_{ik} q_{kj} = \sum_k \left(\sum_j q_{kj}\right) p_{ik} \quad \text{(Fubini’s Theorem)}
\]

\[
= \sum_k p_{ik} \quad \text{(Q is stochastic, i.e. } \sum_j q_{kj} = 1, \forall k)
\]

\[
= 1 \quad \text{(P is stochastic, i.e. } \sum_k p_{ik} = 1, \forall i)
\]

Since, every entry in \(PQ\) matrix is non-negative and the row sum of the matrix is 1, we know that \(PQ\) is a transition probability matrix (row stochastic).

**Corollary 5.6.** If \(P\) is a transition probability matrix then \(P^n\) is also a transition probability matrix.
Proof. We can prove this statement easily by induction.

The base case is when \( n = 1 \). This is true trivially.

We assume that this statement is true for \( n = k - 1 \).

We prove, given that the statement is true for \( n = k - 1 \), it is also true for \( n = k \). Begin by rewriting \( P^k \) as a product of \( P \) and \( P^{k-1} \). From our assumption \( P^{k-1} \) is a transition probability matrix. We know from Lemma 5.5 that since both \( P \) and \( P^{k-1} \) are transition probability matrix, \( P^k \) is also a transition matrix.

\[ \square \]

Note that the converse statement is not true. \( P^n \) being a transition matrix is not sufficient for \( P \) to be a transition matrix. A counter-example can be easily constructed.

**Counter-example (n=2):** Consider a transition matrix
\[
A^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
\]
where \( A = \begin{bmatrix} -1/\sqrt{2} & -1/\sqrt{2} \\ -1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \). This counter-example clearly proves that \( A^2 \) being a transition matrix is not sufficient for \( A \) to be a transition matrix.

**Remark 5.7.** If \((X_n, n \geq 0) \sim \text{Markov}(\lambda, P)\) then,

1. \( \mathbb{P}(X_n = i) = (\lambda P^n)_i \)
2. \( \mathbb{P}(X_n = j | X_0 = i) = (P^n)_{ij} \)

**Proof.**

1) We know from Theorem 5.2, \( \mathbb{P}(X_n = i, \ldots, X_0 = x_0) = \lambda(x_0) p_{x_0 x_1} \ldots p_{x_{n-1} x_n} \). We can write the probability of a state being \( i \) as,

\[
\mathbb{P}(X_n = i) = \sum_{x_{n-1} x_{n-2}} \ldots \sum_{x_0} \lambda(x_0) p_{x_0 x_1} \ldots p_{x_{n-1} x_n} p_{x_n i}
\]

\[
= \sum_{x_{n-2}} \ldots \sum_{x_0} \lambda(x_0) p_{x_0 x_1} \ldots \left( \sum_{x_{n-1}} p_{x_{n-1} x_{n-1}} p_{x_{n-1} i} \right) \quad \text{(Fubini’s Theorem)}
\]

\[
= (\lambda P^n)_i \quad \text{(Recursively using Fubini’s Theorem)}
\]

2) We begin by using the definition of conditional probability and Theorem 5.2. And use Fubini’s theorem recursively.

\[
\mathbb{P}(X_n = j | X_0 = i) = \frac{\mathbb{P}(X_n = j, X_0 = i)}{\mathbb{P}(X_0 = i)} = \frac{\sum_{x_{n-1}} \sum_{x_{n-2}} \ldots \sum_{x_1} \lambda(i) p_{i x_1} \ldots p_{x_{n-1} x_n} p_{x_n j}}{\lambda(i)} \quad \text{(Fubini’s Theorem)}
\]

\[
= (P^n)_{i,j} \quad \text{(Fubini’s Theorem)}
\]

**Corollary 5.8.** Let \((X_n, n \geq 0) \sim \text{Markov}(\lambda, P)\). Then, \((X_{a+nb}, n \geq 0) \sim \text{Markov}(\lambda P^a, P^b)\), for all \( a \geq 0 \) and \( b \geq 1 \).
5.2 Hitting Time

**Definition 5.9.** Consider a random process \( X_n : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (I, \xi) \), which is a discrete-time Markov Chain \((X_n, n \geq 0) \sim \text{Markov}(\lambda, P)\). Now consider a set \( A \subseteq I \). We define the hitting time of \( A \), denoted by \( H_A \), as,

\[
H_A = \inf \{ n \geq 0 \mid X_n \in A \}.
\]

We denote the hitting time of \( A \) given initial state as \( H^A_i = H_A \mid (X_0 = i) \).

**Remark 5.10.** Let us look at hitting time of \( A \) for some special cases:

1. \( H^A = 0 \) if \( X_0 \in A \)
2. \( H^A = 1 \) if \( X_0 \notin A \) and \( X_1 \in A \)
3. \( H^A = n \) if \( X_0, X_1, \ldots, X_{n-1} \notin A \) and \( X_n \in A \)

Note that the hitting time of a set is a random variable that takes values in \( \{0, 1, 2, \ldots\} \cup \{\infty\} \).

**Remark 5.11.** We denote the probability that hitting time of set \( A \) is finite as \( h_A \). That is

\[
h_A = \mathbb{P}(H_A < \infty).
\]

We can also consider probability of finite hitting time conditioned on the initial state,

\[
h^A_i = \mathbb{P}(H_A < \infty \mid X_0 = i).
\]

Clearly, using law of total probability, \( h_A = \sum_i \lambda(i) h^A_i \).

**Corollary 5.12.** If \( h^A_i < 1 \), then, \( \mathbb{E}[H^A_i] = \infty \)

The proof for this corollary is simple. One can write down the definition of expectation and note that since \( h^A_i < 1 \) for some \( i \), hitting time when starting at \( i \) is infinity \( H^A_i = \infty \) with non-zero probability. Hence the expectation of hitting time is infinity.

**Theorem 5.13.** The finite hitting time probabilities \((h^A_i)_{i \in I}\) satisfy the following equation -

\[
h^A_i = \begin{cases} 1, & \text{if } i \in A \\ \sum_j p_{ij} h^A_j & \text{if } i \notin A. \end{cases}
\]

Moreover, if \((x_i)_{i \in I}\) is another non-negative solution of the above equation, then \((h^A_i)_{i \in I}\) is the smallest non-negative solution of the above equation.

**Proof.** We first prove the first statement, i.e. the finite time hitting time probabilities satisfy the given equations.

Note if \( i \in A \) then the equations are trivially satisfied. Since \( i \in A \) implies \( H^A_i = 0 \implies h^A_i = 1 \).

Consider \( i \notin A \).

\[
h^A_i = \mathbb{P}(H^A < \infty \mid X_0 = i)
= \sum_j \mathbb{P}(H^A < \infty \mid X_1 = j, X_0 = i) \mathbb{P}(X_1 = j \mid X_0 = i)
= \sum_j \mathbb{P}(H^A < \infty \mid X_1 = j, X_0 = i) p_{ij}
= \sum_j p_{ij} \mathbb{P}(H^A < \infty \mid X_1 = j) = \sum_j p_{ij} h^A_j
\quad \text{ (Markov Property)}
\]
Next, we prove the second statement, i.e. \((h_i^A)_{i \in I}\) is the smallest non-negative solution to the equation above.

Note if \(i \in A\), then \(h_i^A = 1 = x_i\). All solutions \((x_i)'s\) are same.

Consider \(i \notin A\).

\[
x_i = \sum_{j \in I} p_{ij} x_j = \sum_{j \in A} p_{ij} 1 + \sum_{j \notin A} p_{ij} \sum_{k \in I} p_{jk} x_k
\]

\[
= \sum_{j \in A} p_{ij} 1 + \sum_{j \notin A} p_{ij} \left( \sum_{k \in A} p_{jk} 1 + \sum_{k \notin A} p_{jk} \sum_{l \in I} p_{kl} x_l \right)
\]

\[
= \mathbb{P}_i(X_1 \in A) + \mathbb{P}_i(X_2 \in A, X_1 \notin A) + \ldots + \mathbb{P}_i(X_n \in A, X_{n-1} \notin A, \ldots, X_0 \notin A)
\]

\[
+ \sum_{t \in I} p_{ij} p_{jk} \ldots p_{wt} x_t \geq \mathbb{P}_i(H_i^A \leq n)
\]

We have shown that any non-negative solution \(x_i\) is lower bounded by \(\mathbb{P}_i(H_i^A \leq n)\), for all \(n\).

Therefore, \(x_i \geq \mathbb{P}_i(H_i^A < \infty) = h_i^A\). \(\blacksquare\)

---

\(^1\)Notation: We use \(\mathbb{P}_i(X_1 \in A)\) instead of \(\mathbb{P}(X_1 \in A | X_0 = i)\). Both the expression mean the same quantity, it is the probability that \(X_1 \in A\) conditioned on the initial state to be \(i\).