This lecture first reviewed Markov property, strong Markov property, and probability generating function (from Lecture 07). Then, it covered class structure where the intuition is “... to break a Markov chain into smaller pieces, each of which is relatively easy to understand, and which together give an understanding of the whole.” (see source for the quoted sentence.)

8.1 Review

8.1.1 Markov Property

Markov property says that for each time $m$, conditional on $X_m = i$, the process after time $m$ begins afresh from $i$.

Theorem 8.1 (Markov Property).
Given $X_m = i$, $(X_{m+k}, k \geq 0)$ is independent of $(X_1, X_2, \ldots, X_m)$, and $(X_{m+k}, k \geq 0) \sim \text{Markov}(\delta_i, P)$.

8.1.2 Strong Markov Property

Instead of conditioning on a fixed event $X_m = i$, we can also examine the process after a random stopping time and define strong Markov property.

Definition 8.2 (Stopping Time).
A random variable $T : \Omega \to \{0, 1, 2, \ldots\} \cup \{\infty\}$ is called a stopping time if the event $\{T = n\}$ depends only on $X_0, X_1, \ldots, X_n$ for $n = 0, 1, 2, \ldots$

Theorem 8.3 (Strong Markov Property).
Let $(X_n)_{n \geq 0}$ be Markov($\lambda$, $P$) and let $T$ be a stopping time of $(X_n)_{n \geq 0}$. Then, conditional on $T < \infty$ and $X_T = i$, $(X_n)_{n \geq 0}$ is Markov($\delta_i$, $P$) and independent of $X_0, X_1, \ldots, X_T$.

8.1.3 Probability Generating Function

Probability generating function is a power series representation of the probability mass function for a discrete random variable.

Definition 8.4 (Probability Generating Function).
Let $Y$ be an integer-valued random variable and $Y \geq 0$. Define the probability generating
**function of $Y$ as $\varphi_Y(t) = \mathbb{E}[t^Y]$.**

$$\varphi_Y(t) = \mathbb{E}[t^Y] = \sum_{k=0}^{\infty} t^k \mathbb{P}(Y = k) \text{ for } \forall t \in [0, 1)$$

One shall note that the expectation in the definition is always finite because $t^\infty \mathbb{P}(Y = \infty) = t^\infty \cdot 0 = 0$. It can also be verified that $\varphi_Y(t) : [0, 1) \rightarrow [0, 1]$.

**Example 8.5 (One-Dimensional Random Walk).**

Consider a one-dimensional random walk with $\mathbb{P}(\text{Right}) = p$ and $\mathbb{P}(\text{Left}) = p$ where $p + q = 1$.

![One-Dimensional Random Walk](image)

Figure 8.1: One-Dimensional Random Walk

Define hitting time with initial position $X_0 = k$ as $H^{(0)}_{X_0} = \inf\{n \geq 0 | X_n = 0\}$. We know from previous lectures that the probability of eventually hitting $\{0\}$ with starting position $X_0 = k$ is given by

$$\mathbb{P}_k(H^{(0)} < \infty) = \begin{cases} 1, & \text{if } q \geq p \\ \left(\frac{q}{p}\right)^k, & \text{if } q < p \end{cases}$$

Now, we want to obtain the complete distribution of the time to hit 0 starting form $X_0 = 1$. We will accomplish this task by finding the probability generating function for discrete random variable hitting time $H^{(0)}_{X_0=1}$ with starting position $X_0 = 1$.

$$\varphi_{H^{(0)}_{H^{(0)}_{X_0=1}}}^1(t) = \mathbb{E}_1[t^{H^{(0)}_{X_0=1}}]$$

$$= q \cdot t + p \cdot \mathbb{E}_1[t^{H^{(0)}_{X_0=1}} | X_1 = 2]$$

We need to find a recursion relation to proceed. By Markov property at time 1, we can condition on $X_1 = 2$ and observe that

$$H^{(0)}_{X_0=1, X_1=2} = 1 + H^{(1)}_{X_0=2} + \tilde{H}^{(0)}_{X_0=1}$$

This equation can be interpreted as the time needed to hit $\{0\}$ with the first two steps being $X_0 = 1$ and $X_1 = 2$ is equal to first moving from $X_0 = 1$ and $X_1 = 2$ (cost 1 step), then wait to move back from position 2 to position 1 ($H^{(1)}_{X_0=2}$), and finally from position 1 to position 0 ($\tilde{H}^{(0)}_{X_0=1}$). Note that $\tilde{H}^{(0)}_{X_0=1}$ is an independent copy of $H^{(0)}_{X_0=1}$. Now, we can compute the expectation in the previous equation.

$$\mathbb{E}_1[t^{H^{(0)}_{H^{(0)}_{X_0=1}}} | X_1 = 2] = \mathbb{E}[t^{H^{(0)}_{X_0=1}} | X_0 = 1, X_1 = 2]$$

$$= \mathbb{E}[t^{1+H^{(1)}_{X_0=2}} + \tilde{H}^{(0)}_{X_0=1} | H^{(1)}_{X_0=2} < \infty]$$

$$= \mathbb{E}[t \cdot t^{H^{(1)}_{X_0=2}} \cdot \tilde{H}^{(0)}_{H^{(0)}_{X_0=1}} | H^{(1)}_{X_0=2} < \infty]$$
We can condition on \( H_2^{(1)} < \infty \) because if \( H_2^{(1)} = H_{X_0=1,X_1=2} = \infty \), then \( H_{X_0=1,X_1=2} = \infty \), hence \( t^{H_{X_0=1,X_1=2}} \to t^{\infty} = 0 \), therefore the expectation is zero as well. We will apply strong Markov property to proceed with the calculation.

\[
\begin{align*}
E[t \cdot t^{H_2^{(1)}} \cdot t^{\tilde{H}_1^{(0)}} | H_2^{(1)} < \infty] &= t \cdot (E_2[t^{H_2^{(1)}} | H_2^{(1)} < \infty] \cdot P_2(H_2^{(1)} < \infty)) \cdot E_1[t^{\tilde{H}_1^{(0)}}] \\
&= t \cdot (E_2[t^{H_2^{(1)}} | H_2^{(1)} < \infty] \cdot E_1[t^{\tilde{H}_1^{(0)}}] \\
&= t \cdot \phi(t) \cdot \phi(t)
\end{align*}
\]

Plug this result back in \( \phi_{H_1^{(0)}}(t) \), we get \( \phi(t) = qt + pt(\phi(t))^2 \), which results in \( \phi(t) = \frac{1 \pm \sqrt{1 - 4pqt^2}}{2pt} \).

Since \( \phi(t) \leq 1 \) for all \( t \in [0, 1] \), then \( \phi(t) = \frac{1 - \sqrt{1 - 4pqt^2}}{2pt} \). To recover the distribution of \( H_1^{(0)} \), we will expand \( \phi(t) \) in power series form,

\[
\phi(t) = \frac{1}{2pt} (1 - \sqrt{1 - 4pqt^2}) \\
\approx \frac{1}{2pt} \{1 - (1 + \frac{1}{2}(-4pqt^2) + \frac{1}{2}(-\frac{1}{2})(-4pqt^2)^2 + \cdots)\} \\
= qt + pq^2t^3 + \cdots \\
= \sum_{k=1}^{\infty} P_1(H^{(0)} = k) \cdot t^k
\]

Thus, by comparing the terms, \( P_1(H^{(0)} = 2l + 1) = \binom{2l}{l} \frac{p^{l+1}}{t+1} \) for all \( l \geq 0 \).

We can also inspect the probability that the hitting time will be finite

\[
\lim_{t \to 1^-} \phi(t) = \lim_{t \to 1^-} E_1[t^{H^{(0)}}] = P_1(H^{(0)} < \infty) = \frac{1 - \sqrt{1 - 4pq}}{2p} = \begin{cases} 
1, & \text{if } q \geq p \iff p \leq \frac{1}{2} \\
\frac{2}{p}, & \text{if } q < p \iff p > \frac{1}{2} \end{cases}
\]

This agrees with what we already have from previous lectures.

### 8.2 Class Structure

"It is sometimes possible to break a Markov chain into smaller pieces, each of which is relatively easy to understand, and which together give an understanding of the whole. This is done by identifying the communicating classes of the chain." [see source for the quoted sentence.]

**Definition 8.6 (i leads to j).**

We say "i leads to j" and write "i \( \to j \)" if the probability of hitting j eventually when starting from i is positive, i.e.

\[
P_i(H^{(j)} < \infty) > 0
\]

Negatively, \( i \not\to j \equiv P_i(H^{(j)} < \infty) = 0 \), or \( i \not\to j \equiv P_i(H^{(j)} = \infty) = 1 \)
Lemma 8.7.  
\[ i \rightarrow j \Leftrightarrow \exists m \geq 0 \text{ such that } (P^m)_{ij} = p^{(m)}_{i,j} > 0 \]

Proof.
(\(\Rightarrow\)) Suppose \(i \rightarrow j\), then by definition of \(\rightarrow\), \(P_i(H(j) < \infty) > 0\). Hence, \(\exists m \geq 0\) such that \(P_i(H(j) = m) > 0\). Therefore, \(P^{(m)}_{i,j} = P_i(X_m = j) \geq P_i(H(j) = m) > 0\).

(\(\Leftarrow\)) Suppose \(\exists m \geq 0\) such that \(P^{(m)}_{i,j} > 0\). To show \(P_i(H(j) < \infty) > 0\), it is enough to show \(\exists m \geq 0\) such that \(P_i(H(j) \leq m) > 0\). We will prove the later by contradiction. Assume the opposite, i.e. \(P_i(H(j) \leq m) \leq 0\). Since probabilities are always non-negative, then \(P_i(H(j) \leq m) = 0\). Take the complement, \(P_i(H(j) > m) = 1\). Since the event \(\{H(j) > m\} = \{X_0, X_1, \cdots, X_m \neq j\} \subseteq \{X_m \neq j\}\), then \(P_i(X_m \neq j) \geq P_i(H(j) > m) = 1\). Since all probabilities are less than or equal to 1, then \(P_i(X_m \neq j) = 1\). Take the complement, \(P_i(X_m = j) = 0\), i.e. \(P^{(m)}_{i,j} = 0\), which contradicts with our assumption that \(P^{(m)}_{i,j} > 0\).

\[ \square \]

Definition 8.8 (Communicating Classes).
We say “\(i\) communicates with \(j\)” and write “\(i \leftrightarrow j\)” if both \(i\) leads to \(j\) and \(j\) leads to \(i\).

\(i \leftrightarrow j \equiv (i \rightarrow j) \land (j \rightarrow i)\)

Note that \((i \leftrightarrow j)\) can mean either \((i \rightarrow j)\) or \((j \rightarrow i)\) or both.

Claim 8.9. “\(\leftrightarrow\)” is an equivalence relation

Proof. We will prove the claim by checking the definition of equivalence relation.
1. (Reflexivity)
It is trivial to see that \(i \leftrightarrow i\) for \(\forall i\) in the state space by definitions of \(\rightarrow\)” and “\(\leftrightarrow\)”.
2. (Symmetry)
By definition of “\(\leftrightarrow\)” \((i \leftrightarrow j) \Leftrightarrow (i \rightarrow j) \land (j \rightarrow i) \Leftrightarrow (j \rightarrow i) \land (i \rightarrow j) \Leftrightarrow (j \leftrightarrow i)\).
3. (Transitivity)
First, we show transitivity of “\(\rightarrow\)”.
Suppose \((i \rightarrow j) \land (j \rightarrow k)\), then by the previous lemma, \(\exists m, n \geq 0\) such that \((P^m)_{ij} > 0\) and \((P^n)_{jk} > 0\). Consider the \((i,k)\)-th element of the transition matrix \((P^{m+n})_{ik} = \sum_l (P^m)_i (P^n)_k \geq (P^m)_{ij} (P^n)_{jk} > 0\) because all elements of any transition matrix are non-negative. Therefore, \(\exists (m+n) \geq 0\) such that \((P^{m+n})_{ik} > 0\), which means \(i \rightarrow k\) by the previous lemma.

Next, we use this result to show transitivity of “\(\leftrightarrow\)”.
Suppose \((i \leftrightarrow j) \land (j \leftrightarrow k)\), then by definition of “\(\leftrightarrow\)” \(((i \rightarrow j) \land (j \rightarrow i)) \land ((j \rightarrow k) \land (k \rightarrow j))\). By re-grouping the terms, \(((i \rightarrow j) \land (j \rightarrow k)) \land ((k \rightarrow j) \land (j \rightarrow i))\). Hence, \((i \rightarrow k) \land (k \rightarrow i)\), which means \(i \leftrightarrow k\) by definition of “\(\leftrightarrow\)”.

We just showed that “\(\leftrightarrow\)” is an equivalence relation. Thus, it partitions the state space \(E\) into equivalence classes. We call those classes communicating classes.

Definition 8.10 (Communicating Classes).
The equivalence classes corresponding to the communicating relation (“\(\leftrightarrow\)” are called communicating classes.
Definition 8.11 (Closed Classes).
We say a class (set of states) $C$ of state space $E$ is closed if
\[ \forall i \in C, \ i \rightarrow j \implies j \in C \]
A closed class is a “black hole” where there is no escape.

Definition 8.12 (Absorbing States).
We say a state $i \in E$ is an absorbing state if $\{i\}$ is a closed class.
If state $i$ is an absorbing state, then once the Markov chain hits $i$, it will stay at $i$ forever.

Definition 8.13 (Irreducibility).
A Markov chain is irreducible if $\exists$ unique closed class $C$ which is the entire state space $E$.

An irreducible Markov chain cannot be break into multiple closed classes.

Definition 8.14 (Recurrence and Transience).
\begin{itemize}
  \item A state $i \in E$ is recurrent if $P_i(X_n = i \text{ for infinitely many } n) = 1$.
  \item A state $i \in E$ is transient if $P_i(X_n = i \text{ for infinitely many } n) = 0$.
\end{itemize}
If we define return time $T^+_i = \inf\{n \geq 1 | X_n = i\}$, then we will see later that event $\{X_n = i \text{ for infinitely many } n\}$ is equivalent to the event $\{T^+_i < \infty\}$. One may also wonder why the probabilities in the definition only takes values 0 and 1. This will be shown in the next theorem.

Example 8.15 (Illustration of Definitions). First, we note that all “$\rightarrow$” relationships can be summarized into $1 \rightarrow 2 \rightarrow 3 \rightarrow 4$ and $3 \rightarrow 1$. Hence, the communicating classes are $\{1, 2, 3\}$ and $\{4\}$. Class $\{1, 2, 3\}$ is not closed because $3 \in \{1, 2, 3\}$ and $3 \rightarrow 4$, but $4 \notin \{1, 2, 3\}$. On the other hand, $\{4\}$ is closed because $4 \nleftrightarrow i$ for $i = 1, 2, 3$. Since $\{4\}$ is closed and it only contains one element, then 4 is an absorbing state. It is also the unique absorbing state in this example. Since there exists a closed class $\{4\}$ which is not $E = \{1, 2, 3, 4\}$, then the Markov chain is not irreducible, i.e. it is reducible. Since 4 is the unique absorbing state and class $\{1, 2, 3\}$ is not closed, then states 1, 2, and 3 are transient while state 4 is recurrent.

Theorem 8.16. Any state $i \in E$ is either transient or recurrent but never both.
Proof will be covered in the next lecture note.