Markov Chains: An Overview

**Definition.** Given a finite or countable set $S = \{1, 2, \cdots, K\}$, $(K$ is a finite integer or $+\infty)$ a sequence of random variables $X_0, X_1, \cdots$ taking values in $S$ is said to define a Markov chain if

$$P[X_{n+1} = j | X_0 = i_0, \cdots, X_{n-1} = i_{n-1}, X_n = i] = P[X_{n+1} = j | X_n = i]$$

for all $i, j, i_0, \cdots, i_{n-1} \in S, n = 1, 2, \cdots$.

The last equation says that given the history of the states visited up to time $n$ the probability of being in a given state at time $n + 1$ only depends on the current state. Loosely speaking, a Markov chain *forgets* about its past.

**Definition.** A Markov chain is said to have a stationary transition probabilities if $P[X_{n+1} = j | X_n = i]$ does not depend on $n$. If this is the case we refer to the matrix

$$P_{ij} \equiv P[X_{n+1} = j | X_n = i]$$

as the transition matrix for the Markov chain.

From the definition it follows that the transition matrix $P$ is stochastic, meaning that each row sums to 1.

**Definition.** We refer to the row vector $\pi = (\pi_1, \pi_2, \cdots)$ where

$$P[X_0 = i] = \pi_i, \text{ for all } i \in S$$

as the initial distribution for the Markov chain.

For the remainder of this handout, all Markov chains will have state space $S$ and stationary transition matrix $P$.

It is an elementary exercise to check that

$$P[X_n = j | X_0 = i] = P^n_{ij}, \text{ for all } i, j \in S, n = 0, 1, \cdots,$$

and if $\pi$ denotes the initial distribution for the chain, then

$$P[X_n = i] = (\pi P^n)_i, \text{ for all } i \in S, n = 0, 1, \cdots.$$
Definition. State $j$ is said to be accessible from state $i$ if there is a positive probability of getting from state $i$ to state $j$ in a finite number of steps, i.e., $P^n_{ij} > 0$ for some $n \geq 0$. We denote this by $i \rightarrow j$. Note that by definition $i \rightarrow i$ for all states $i$.

Definition. States $i$ and $j$ are said to communicate if $i \rightarrow j$ and $j \rightarrow i$. If $i$ and $j$ communicate we denote this by $i \sim j$.

Exercise. Show $\sim$ defines an equivalence relation on $S$.

Definition. The Markov chain is said to be irreducible if there is a single $\sim$-equivalence class, i.e. $i \sim j$ for all $i, j \in S$.

Definition. The period of a state $i$ is defined to be the greatest common divisor of the set of positive integers $n$ for which $P^n_{ii} > 0$. This number is denoted by $d(i)$. If the set is empty we define $d(i) = 0$.

Exercise. Show $d(i) = d(j)$ if $i \sim j$. Also, show for every $i \in S$ for which $d(i) > 0$ there exists $N(i)$ for which $P^n_{ii} > 0$ for all $n \geq N(i)$.

Definition. A Markov chain for which $d(i) = 1$ for every state is said to be aperiodic.

Given a state $i$ define

$$H_{ii}^{(n)} = P[X_n = i, X_{n-1} \neq i, X_{n-2} \neq i, \ldots, X_2 \neq i, X_1 \neq i, X_0 = i]$$

the probability that starting from state $i$ the first return to state $i$ occurs at time $n$. Then

$$H_{ii}^{(0)} = 1$$
$$H_{ii}^{(1)} = P_{ii}.$$ 

$H_{ii}^{(n)}$ satisfy the recursion formulae

$$P^n_{ii} = \sum_{k=1}^{n} H_{ii}^{(k)} P_{ii}^{n-k}.$$ 

To see this, let $E$ denote the event that the chain is in state $i$ at time $n$, and let $F_k$ denote the event that the chain visits state $i$ for the first positive time at time $k$. Then the $F_k$ are disjoint events and

$$E = \bigcup_{k=1}^{n} F_k \cap \{X_n = i\}.$$ 

Thus

$$P(E|X_0 = i) = \sum_{k=1}^{n} P(F_k \cap \{X_n = i\}|X_0 = i) = \sum_{k=1}^{n} H_{ii}^{(k)} P(X_n = i|X_k = i) = \sum_{k=1}^{n} H_{ii}^{(k)} P_{ii}^{n-k}.$$ 

2
**Definition.** A state \( i \) is said to be **recurrent** if the probability of returning to state \( i \) in finitely many steps is 1, that is:

\[
H_{ii} = \sum_{n=1}^{\infty} H_{ii}^n = 1.
\]

If \( H_{ii} < 1 \) we call the state **transient.**

Equivalently, a state \( i \) is recurrent if, conditionally given that \( X_0 = i \), with probability 1 there are infinitely many returns to that state. A standard theorem for classifying states is the following:

**Theorem.** A state \( i \) is recurrent if

\[
\sum_{n=1}^{\infty} P_{ii}^n = +\infty
\]

and is transient if

\[
\sum_{n=1}^{\infty} P_{ii}^n < +\infty.
\]

**Exercise.** Use this theorem to show if \( i \sim j \) then \( i \) is recurrent if and only if \( j \) is recurrent.

For a transient state, we can talk about the expected number of returns to the state, given that we started there at time 0. (For recurrent states this expectation is infinite.) In fact, let \( M_i \) denote the number of times the state \( i \) is visited after starting in state \( i \) at time 0. As above, let \( H_{ii} \) denote the conditional probability that we return to state \( i \) given that we start at state \( i \) at time 0, and assume \( H_{ii} > 0 \). Then

\[
P[M_i = 0] = 1 - H_{ii}.
\]

and

\[
P[M_i \geq 1] = H_{ii}.
\]

Furthermore, conditionally, given that we return to state \( i \) at least once, it is as if time starts all over again when we return to state \( i \) for the first time. Thus, the conditional probability that never return to state \( i \) for a second time given that we return once is

\[
P[M_i = 1] = H_{ii}(1 - H_{ii}).
\]

By repeating this same argument we obtain the following:

**Theorem.** For a transient state \( i \), suppose there is a positive probability \( H_{ii} \) of returning at least once to state \( i \) given that we start in state \( i \). Then the number of times \( M_i \) we return to state \( i \) starting from \( i \) has a geometric distribution. In fact,

\[
P[M_i = k] = H_{ii}^k(1 - H_{ii}), \text{ for } k = 0, 1, 2, \ldots.
\]

**Corollary.** For a transient state \( i \) the expected number of times we return to state \( i \) starting from \( i \) is \( \frac{1}{H_{ii}} \) if \( H_{ii} > 0 \) and is 0 if \( H_{ii} = 0 \).
**Corollary.** For an irreducible Markov chain with a finite state space, all states are recurrent.

Proof. The irreducibility of the chain means that all states are recurrent or all states are transient. In the latter case, fix any state $i$ and let $N_j$ denote the number of times the state $j$ is visited after the chain is started in state $i$. Since there are only finitely many states we see that $EN_j = +\infty$ for some state $j$. It follows that the expected number of visits to state $j$ starting from state $j$ is also infinite. This contradicts the previous Corollary.

**Corollary.** A Markov chain with a finite state space and transition matrix $P$ having all of its entries positive is recurrent, irreducible and aperiodic.

Since a recurrent state $i$ is visited infinitely often with probability 1, it is natural to consider the amount of time between visits. Let $R_i$ denote time of first return to state $i$ when the chain is started in state $i$ so that

$$P[R_i = m|X_0 = i] = H^m_{ii}.$$  
Conditionally, given $X_0 = i$ the random variable $R_i$ is finite with probability 1 and the amounts of time between successive visits are iid replicates of $R_i$.

Suppose we want to know the average amount of time the Markov chain spends in state $i$, starting from state $i$. We can observe the chain for $N$ returns to $i$. By the above comments this involves generating $N$ replicates of $R_i$. For each $k$ let us denote the number of these replicates which are equal to $k$ by $M_k$. By the strong law of large numbers we should expect $M_k$ to be about $N \times P[R_i = k|X_0 = i] = N \times H^k_{ii}$.

If $D$ denotes the total duration of time before the $N$-th return to state $i$ then the contribution towards this duration is 1 time unit for each $R_i$ equal to 1, 2 time units for each $R_i$ equal to 2, and so on. It follows that

$$D = \sum_{k=1}^{\infty} kM_k.$$
Combining with the above statement we see that the average amount of time spent in state $i$ equals

$$\frac{N}{D} = \frac{1}{\sum_{k=1}^{\infty} kM_k/N} \approx \frac{1}{E(R_i)}.$$

On the other hand, under appropriate conditions we can think of the average amount of time spent in a given state as the long run probability of being in that state at time $n$ as the following theorem shows.

**Theorem.** Given a recurrent, irreducible and aperiodic Markov chain

$$\lim_{n \to \infty} P^n_{ii} = \lim_{n \to \infty} P^n_{ji} = \frac{1}{E[R_i|X_0 = i]}.$$

**Definition.** A state in a Markov chain is said to be positive recurrent if $E[R_i|X_0 = i] < +\infty$ for some state $i \in S$. Otherwise we call the state null recurrent.
If a Markov chain is irreducible, recurrent and aperiodic, then positive recurrence for a single state implies positive recurrence for all states. Also, under the same conditions if the state space is finite then the chain is positive recurrent.

**Theorem (Existence and uniqueness of a stationary distribution).** Given a positive recurrent, irreducible and aperiodic Markov chain there exists a unique nonnegative row-vector \( \pi \) satisfying

\[
\sum_i \pi_i = 1
\]

and

\[
\pi P = \pi.
\]

Furthermore, this vector has the additional property

\[
\lim_{n \to \infty} P^n_{ii} = \lim_{n \to \infty} P^n_{ji} = \pi_i, \text{ for all states } i, j \in S.
\]

More general results are available. For example, see Kemeny, Snell and Knapp (1976), which is a classic textbook for Markov chains.

**References**