AN INTRODUCTION TO BASIC PROPERTIES OF MARKOV CHAINS

NEAL PANNALA

ABSTRACT. This paper serves to give an introduction to Markov Chains, focusing on thoroughly building the properties of Markov Chains with a finite state space. We then extend these properties for infinite state space, going over a few examples. Specifically, we look at Gambler's Ruin and other examples of random walks by using the previously defined properties of Markov Chains.

CONTENTS

1. TIME HOMOGENEOUS FINITE MARKOV CHAIN

Markov Chains are based on the stochastic property which assumes that the transition probabilities between states depend only on the current state rather than the past state (Markov property). Each transition is assigned a probability that defines the chance of the system changing from one state to another.

Definition 1.1. A stochastic process $\{X_n\}$ is defined as a *Markov Chain* if for all $n \in \mathbb{N}$ and $s_0, ..., s_n \in S$,

$$\mathbb{P}(X_n = s_n | X_0 = s_0, X_1 = s_1, \dots, X_{n-1} = s_{n-1}) = \mathbb{P}(X_n = s_n | X_{n-1} = s_{n-1})$$

where $\mathbb{P}(X_n = s_n | X_{n-1} = s_{n-1})$ refers to the conditional probability of being in state s_n at the nth time step given that, at the $(n-1)$th time step, the process is in state s_{n-1}. Note that S is our set of all the possible states the Markov Chain can reach.

Definition 1.2. A Markov Chain $\{X_n\}$ is *time homogeneous* if for all $n \in \mathbb{N}$ and $x, y \in S$,

$$\mathbb{P}(X_n = y | X_{n-1} = x) = \mathbb{P}(X_1 = y | X_0 = x).$$

For the rest of this paper, every Markov Chain discussed will be time homogenous. For (time homogenous) Markov Chains, we only need to specify the distribution of X_0 and the transition probabilities. For example, the probability of going from state x to state y can be expressed as follows.

$$p(x,y) := \mathbb{P}(X_1 = y | X_0 = x)$$

This assumption is important as it allows as to us to reduce the Markov Chain into a finite set of time-independent probabilities (this will be further discussed in the next definition). Without time homogeneity, we would have to work with infinite sets of transition probabilities, which would make predicting future events very difficult as future events may happen at arbitrary times, which we will see going forward.

Definition 1.3. The transition matrix P for the Markov chain is an $N \times N$ matrix, where N refers to the number of states, size of S. Denote the entry as P_{ij}, where $P_{ij} = p(i,j) = P(X_n = j | X_{n-1} = i) = P(X_1 = j | X_0 = i)$, due to time homogeneity. The matrix P is stochastic, that is, each entry is between 0 and 1, and

$$\sum_j P_{ij} = 1$$

for each i. This is because, if the Markov Chain starts at position i at time n, it must, with probability 1, be at "some" state at time $n+1$. Thus, the probability of all the possible jumps the Markov Chain can make from state i must sum to exactly 1. The $N \times N$ transition matrix is represented as

$$P = \begin{bmatrix} P_{11} & \cdots & P_{1N} \\ \vdots & \ddots & \vdots \\ P_{N1} & \cdots & P_{NN} \end{bmatrix}$$

Let $\pi = (\pi_1, ..., \pi_N)$ be the vector such that $\pi_j = \mathbb{P}[X_0 = j]$. We call π the initial distribution vector.

A time-homogeneous finite Markov chain (X_n), for $n \geq 0$, is entirely determined by the following:

(1) An initial distribution $\pi = (\pi_i)_{i \in S}$
(2) A stochastic matrix $P = (P_{ij})_{i,j \in S}$, with $p_{ij} \geq 0$ for all $i, j \in S$ and $\sum_j p_{ij} = 1$ for each i

Definition 1.4. We can define *n-step transition properties* as the probability to get from state x to state y in n steps. We can notate it as follows:

$$p^n(x,y) := \mathbb{P}(X_n = y | X_0 = x).$$

Proposition 1.5. *For all $n, m \in \mathbb{N}$ and for all $x, y \in S$*

$$p^{n+m}(x,y) = \sum_{z \in S} p^n(x,z) p^m(z,y).$$

Proof. By time homogeneity, followed by the Markov property,

$$\begin{aligned} p^n(x,z)p^m(z,y) &= \mathbb{P}(X_n = z | X_0 = x)\mathbb{P}(X_{n+m} = y | X_n = z) \\ &= \mathbb{P}(X_n = z | X_0 = x)\mathbb{P}(X_{n+m} = y | X_n = z, X_0 = x) \qquad (1.6) \\ &= \mathbb{P}(X_{n+m} = y, X_n = z | X_0 = x). \end{aligned}$$

Thus,

$$\begin{aligned}\sum_{z\in S} p^n(x,z)p^m(z,y) &= \sum_{z\in S}\mathbb{P}(X_{n+m}=y, X_n=z|X_0=x)\\ &= \mathbb{P}(X_{n+m}=y|X_0=x)\\ &= p^{n+m}(x,y).\end{aligned} \tag{1.7}$$

□

Intuitively speaking, this proposition says the following. Let N be the size of S. The probability of getting from x to y in $n+m$ steps is equal to the sum of the probabilities of getting from x to z in n steps and z to y in m steps for each z. This makes sense because we are guaranteed to be "some" state after n steps. This gives us N total paths we can take to get from x to y, depending on where we are after n steps. We can sum over the probabilities of each of our N paths to determine the total probability of getting from x to y after $n+m$ steps.

2. Recurrence and Transience

Definition 2.1. Consider a Markov chain $\{X_n\}$ on a finite state space S. Two states $x, y \in S$ *communicate* if there exist natural numbers, n and m such that $p^n(x,y) > 0$ and $p^m(y,x) > 0$. In this case, we write $x \leftrightarrow y$.

Lemma 2.2. *Communication is an equivalence class.*

(1) *Reflexive:* $p^0(x,x) = 1$
(2) *Symmetric: if* $x \leftrightarrow y$ *then* $y \leftrightarrow x$
(3) *Transitive: if* $x \leftrightarrow y$ *and* $y \leftrightarrow z$ *then* $x \leftrightarrow z$

Proof. Choose n, m, l, k such that $p^n(x,y), p^m(y,x), p^l(y,z)$, and $p^k(z,y)$ are all positive. Then

$$p^{n+l}(x,z) = \sum_{w\in S} p^n(x,w)p^l(w,z) \geq p^n(x,y)p^l(y,z) > 0$$

Similarly, $p^{m+k}(z,x) > 0$. So $x \leftrightarrow z$.

□

Definition 2.3. We can use the previous definition to define a *communication class* C, $C \subset S$, such that, for all $s_i, s_j \in C, s_i \leftrightarrow s_j$.

Noticing that every state in S is at least in a communication class with itself, we can use these disjoint communication classes to make a disjoint partition of S.

Definition 2.4. A communication class C is *recurrent* if $p(x,y) = 0$ for each $x \in C$ and each $y \notin C$. Essentially, once the Markov Chain enters C, it can never leave. A communication class is *transient* if it is not recurrent, i.e., $p(x,y) > 0$ for some $x \in C$ and each $y \notin C$. We say that $x \in S$ is recurrent if its communication class is recurrent, and $y \in S$ is transient if its communication class is transient.

As shown in the following examples, we care about recurrence and transience because they give us more detailed information about how a given Markov chain operates. By telling how different states communicate with each other, we can make predictions about how the chain will function over long periods of time.

Example 2.5. The following transition matrix represents a markov chain where $S = \{1, 2, 3\}$

$$P = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0 & 0.5 & 0.5 \\ 0 & 1 & 0 \end{bmatrix}$$

Breaking P into Communication classes, $C_1 = \{1\}$, which is transient, and $C_2 = \{2, 3\}$, which is recurrent.

Definition 2.6. A Markov chain is *irreducible* if there is only one communication class (which is necessarily recurrent). In other words, we can always get from any state to any other state given a sufficient number of steps.

Proposition 2.7. *If C is a recurrent communication class, then if $\{X_n\}$ starts in C, with probability 1, $\{X_n\}$ visits each state in C infinitely many times, that is, for each $x, y \in C$,*

$$\mathbb{P}(X_n = y \textit{ for infinitely many } n \,|X_0 = x) = 1.$$

Proof. Let $x, y \in C$ and assume that $X_0 = x$. We want to show that there are infinitely many values of n such that $X_n = y$. The idea is that in every long enough interval of time, we have a positive probability q to hit y. So, the probability that we don't hit y after k time intervals is at most $(1 - q)^k$, and, as we increase our number of time intervals, this expression tends to 0.
Since C is a communication class, for each $z \in C$ we can choose n_z such that $p^{n_z}(z, y) > 0$. Let

$$n = \max\{n_z : z \in C\}, q = \min\{p^{n_z}(z, y) : z \in C\}.$$

For each $k \in \mathbb{N}$ let

$$E_k := \{\text{y is visited at a time in}\{n(k-1)+1, ..., nk\}\}.$$

Then, E_k depends only on where the process is at that time, i.e. $X_{n(k-1)+1}, ..., X_{nk}$
Now let's look at the conditional probability of E_{k+1} given all the past states.

$$\mathbb{P}[E_{k+1}|X_0 = s_0, ..., X_{nk} = s_{nk}] = \mathbb{P}[E_{k+1}|X_{nk} = s_{nk}] = \mathbb{P}[E_1|X_0 = s_{nk}].$$

Thus our first conditional probability, is the same as the probability that a Markov Chain starting at state s_{nk} visits y before time n. Because $s_{nk} \in C$, $n_{s_{nk}} \leq max\{n_z : z \in C\} = n$. Thus, $q = min\{p^{n_z}(z, y) : z \in C\} \leq p^{n_{s_{nk}}}(s_{nk}, y)$. Thus,

$$\mathbb{P}[E_{k+1}|X_0 = s_0, ..., X_{nk} = s_{nk}] \geq q.$$

Let E_k^c denote the event complement to E_k. Now let $M, K \in \mathbb{N}, M > K$. Then,

$$\begin{aligned} \mathbb{P}[E_k \text{ does not occur for any } k \in \{K, ..., M\}] &= \mathbb{P}[\bigcap_{k=K}^{M} E_k^c] \\ &= \mathbb{P}[E_M^c|\bigcap_{k=K}^{M-1} E_k^c] \cdot \mathbb{P}[\bigcap_{k=K}^{M-1} E_k^c] \qquad (2.8) \\ &\leq (1-q) \cdot \mathbb{P}[\bigcap_{k=K}^{M-1} E_k^c]. \end{aligned}$$

Iterating this process, we see that

$$\mathbb{P}[E_k \text{ does not occur for any } k \in \{K, ..., M\}] \leq (1-q)^{M-K}.$$

As M tends to ∞, this expression tends to 0, so

$$\mathbb{P}[X \text{ hits } y \text{ after time } nK] = 1.$$

□

Note that this proposition implies that, in an irreducible Markov Chain, with probability 1, every state will be reached infinite times.

3. Strong Markov Property

Definition 3.1. A random time $\tau \in \mathbb{N} \cup \infty$ is called a *stopping time* if $\forall n \in N_0$, the event $\{\tau = n\}$ is determined by $X_0, ..., X_n$

Intuitively put, a stopping time is a time in a Markov Chain which can be chosen based on all the previous states the Markov Chain has been in.

Example 3.2. Here are some examples of times based on the Markov Chain, $\{X_n\}$, defined in (2.4)

(1) $\tau = 5$ (Stopping Time)
(2) $\tau =$ kth time for which $X_n = 2$ (Stopping Time)
(3) $\tau =$ last time for which $X_n = 1$ (Not a Stopping Time)

For the last example, as soon as we see that $X_n = 1$ and $X_{n+1} = 2$, we can determine with certainty that $\tau = n$. We, however, required $n+1$ states to determine that $\tau = n$, so it is not a stopping time.

Loosely put, something qualifies as a "stopping time" if we can "stop" the Markov Chain on the state which that event occurs.

Theorem 3.3. *(Strong Markov Property) Let τ be a stopping time for our Markov process. Let $n, m \geq 1$, let $x_0, ..., x_n \in S$ such that $\mathbb{P}[X_0 = x_0, ..., X_\tau = x_n] > 0$, and let $y_1, ..., y_m \in S$. Then,*

$$\mathbb{P}[X_{\tau+1} = y_1, ..., X_{\tau+m} = y_m | X_0 = x_0, ..., X_\tau = x_n]$$
$$= \mathbb{P}[X_1 = y_1, ..., X_m = y_m | X_0 = x_n].$$

For intution, this theorem basically says that, if τ is a stopping time, what happens after τ, depends solely on X_τ. This theorem extends the Markov Property (which only applied to a fixed time, n) to any stopping time τ.

Proof. We can reformulate the event $\{X_0 = x_0, ..., X_\tau = x_n\}$ into the event where $\tau = n$ and $\{X_0 = x_0, ..., X_n = x_n\}$. Since τ is a stopping time, the event that $\tau = n$ depends only on the values of $\{X_0, ..., X_n\}$. So, the fact that

$$\mathbb{P}[X_0 = x_0, ..., X_\tau = x_n] > 0$$

(i. e. that it is possible to stop at x_n if we follow this path) implies that if $X_0 = x_0, ..., X_n = x_n$, then in fact $\tau = n$. Or, in other words, because our "rule" for the stopping time is fixed, if we follow the same path, we will stop at the same point along the path, x_n. Hence,

$$\{X_0 = x_0, ..., X_\tau = x_n\} = \{X_0 = x_0, ..., X_n = x_n\},$$

and on this event we have $\tau = n$. Thus,

$$\mathbb{P}[X_{\tau+1} = y_1, ..., X_{\tau+m} = y_m | X_0 = x_0, ..., X_\tau = x_n]$$

$$(3.4) \qquad = \mathbb{P}[X_{n+1} = y_1, ..., X_{n+m} = y_m | X_0 = x_0, ..., X_n = x_n]$$
$$= \mathbb{P}[X_1 = y_1, ..., X_m = y_m | X_0 = x_n].$$

□

So why do we care about the Strong Markov Property? Basically, it allows us to treat the time range between stopping times as their own time steps. In other words, instead of analyzing the jump from X_1 to X_2, we can analyze the jump from X_{τ_1} to X_{τ_2}. Essentially, this allows us to apply useful benefits of the normal Markov Property to more complicated problems.

Proposition 3.5. *Suppose that $X_0 = x \in S$ and that with probability 1, $X_n = x$ for infinitely many n. Let $\tau_0, \tau_1, ...$ be the successive times n for which $X_n = x$. The increments $(X_{\tau_k}, ..., X_{\tau_{k+1}})$ for $k \in \mathbb{N}$ are independent and identically distributed.*

Proof. This follows almost immediately from the strong Markov Property. Notice that every $\tau_{k+1} - \tau_k$ is the first time after τ_k such that $X_{\tau_k+n} = x$. Thus, we can say the following for any X_{τ_k}:

$$\mathbb{P}[X_{\tau_k+1} = y_1, ..., X_{\tau_k+n} = x | X_0 = x_0, ..., X_{\tau_k} = x]$$

$$= \mathbb{P}[X_1 = y_1, ..., X_n = x | X_0 = x].$$

Thus, the probability distribution does not depend on k, so each distribution is i.i.d. □

4. Periodicity

Let $\{X_n\}$ be a Markov Chain on countable (not necessarily finite) state space S.

Definition 4.1. For $x \in S$, the *period* of x is the greatest common divisor of

$$J_x := \{n \geq 1 : p^n(x,x) > 0\}.$$

Note that, if it is possible to get from x to x in one time-step, then the period is 1.

Example 4.2. Consider the following Markov Chain P, defined by the following Transition Matrix on the state space S = $\{1, 2, 3\}$.

$$P = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{bmatrix}$$

Here, even though no state can return to itself in one time-step, each state can return to itself in 2 or 3 time-steps. Thus, the period at each state is the gcd(2, 3) = 1.

Example 4.3. Consider the following Markov Chain P, defined by the following Transition Matrix on the state space S = $\{1, 2, 3\}$.

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

Now we can only return to a state after cycling through the other 2 states so the period of P at every state is 3.

Let $d(x)$ denote the period of state x.

Proposition 4.4. *If $x \leftrightarrow y$, then $d(x) = d(y)$*

Proof. Choose n, m s.t. $p^n(x, y) > 0, p^m(y, x) > 0$. Then $p^{n+m}(x, x) > 0$ and $p^{n+m}(y, y) > 0$. Hence $n + m \in J_x \cap J_y$ so both $d(x)$ and $d(y)$ divide $n + m$.
Assume for contradiction that (without loss of generality) $d(x) < d(y)$. Then $\exists k \in J_x$ not divisible by $d(y)$. We have $n + m + k \in J_y$ as we can go from y to y in $n + m + k$ steps. So, $d(y)$ divides $n + m + k$ and $n + m$ but not k, which is a contradiction. □

Definition 4.5. A Markov chain is called *aperiodic* if every state has period 1.

5. Recurrence and transience for countably infinite state space

Suppose that the state space S is countably infinite.

Definition 5.1. $\{X_n\}$ is *irreducible* if for each $x, y \in S$, there exists n such that $p^n(x, y) > 0$.

Definition 5.2. A state $x \in S$ is *recurrent* if

$$\mathbb{P}[\exists \text{ infinitely many } n \geq 1 \text{ s.t. } X_n = x | X_0 = x] = 1$$

and transient if the state, x, is not recurrent.

Proposition 5.3. *If $\{X_n\}$ is irreducible, then either every state is recurrent or every state is transient.*

Proof. It is sufficient to prove the following statement. If there exists a recurrent state, $x \in S$, then every state must be recurrent. Assume that $X_0 = x$. Let $\tau_1, \tau_2, ...$ be the times of the successive visits to x (these are all finite since x is recurrent). By the strong Markov property, the increments $X_{\tau_k}, ..., X_{\tau_{k+1}}$ are i.i.d.
Consider some $y \in S$. Because $\{X_n\}$ is irreducible, $p^n(x, y) > 0$ for some n, and there must exist a k such that $\tau_k < n < \tau_{k+1}$. Thus, $\mathbb{P}[y \in \{X_{\tau_k}, ..., X_{\tau_{k+1}}\}] > 0$. Since the increments (between each τ) are i.i.d., we can define the following probability, q,

$$q := \mathbb{P}[y \in \{X_{\tau_k}, ..., X_{\tau_{k+1}}\}]$$

which is positive and does not depend on k. Since the events $\{y \in \{X_{\tau_k}, ..., X_{\tau_{k+1}}\}\}$ are independent and each has the same probability q, with probability 1 infinitely many of these events occur. Hence, if we start at $X_0 = x$, then with probability 1 we visit y infinitely many times.

Let $\sigma = min\{n \geq 1 : X_n = y\}$. Then $\mathbb{P}[\sigma < \infty | X_0 = x] = 1$. By the strong Markov property, $\{X_{\sigma+j}\}_{j \geq 0}$ (no matter where we start) has the same distribution as the original Markov chain started at y. We know that $\{X_{\sigma+j}\}_{j \geq 0}$ visits y infinitely many times if we start at $X_0 = x$, so also $\{X_j\}_{j \geq 0}$ visits y infinitely many times if we start at $X_0 = y$. Thus if x is a recurrent state, then any other arbitrary state y is also recurrent. □

Proposition 5.4. *A state x is transient if and only if*

$$\sum_{n=0}^{\infty} p^n(x,x) < \infty.$$

Proof. First, we will do the backward direction. Let the total number of visits to x be represented as follows (imagine $\mathbb{1}$ as an indicator variable).

$$R_x := \sum_{n=0}^{\infty} \mathbb{1}_{X_n = x}$$

Then, we can make the following statement:

$$\begin{aligned} \mathbb{E}[R_x | X_0 = x] &= \sum_{n=0}^{\infty} \mathbb{E}[\mathbb{1}_{X_n=x} | X_0 = x] \\ &= \sum_{n=0}^{\infty} p^n(x,x). \end{aligned} \tag{5.5}$$

Hence, if $\sum_{n=0}^{\infty} p^n(x,x) < \infty$ (which we assume for the backward direction) then $\mathbb{E}[R_x] < \infty$, and hence $R_x < \infty$ with probability 1, i.e., x is only visited finitely many times and is transient.

For the forward direction, we will assume that x is transient. Let $\tau_1, \tau_2, ...$ be the times of the successive visits to x. Since x is transient, we know that with positive probability there is some k such that $\tau_k = \infty$. By the strong Markov property, the increments $\tau_{k+1} - \tau_k$ are i.i.d., so each has the same probability, let's call it q, to be infinite.

Note that R_x is equal to the smallest k such that $\tau_{k+1} = \infty$. Therefore, R_x has a geometric distribution with success probability $q > 0$. Hence,

$$\sum_{n=0}^{\infty} p^n(x,x) = \mathbb{E}[R_x | X_0 = x] = \frac{1}{q} < \infty.$$

The above statement is what we wanted to prove. The result implies that $\mathbb{P}[R_x < \infty] = 1$, and that, with probability 1, x is visited finitely many times. □

Example 5.6. Consider the Markov chain with state space $\mathbb{N}_0$ and transition probabilities:

$$p(x,0) = \frac{1}{x+2}, \quad p(x,x+1) = 1 - \frac{1}{x+2}.$$

This Markov chain is irreducible as the chain can always reach 0 and, from 0, can reach every other state. Assume $X_0 = 0$. We have $X_n \leq n$, so at every X_n, we are anywhere from 0 to n inclusive, so the probability to go to 0 on the next step is at least $\frac{1}{n+2}$ in every case. Thus,

$$p^n(0,0) \geq \frac{1}{n+2}.$$

Hence,

$$\sum_{n=0}^{\infty} p^n(0,0) \geq \sum_{n=0}^{\infty} \frac{1}{n+2} = \infty.$$

Therefore, 0 is not a transient state and must be recurrent. Then, by Proposition 5.3, every state is recurrent.

Consider the following random walk on $\mathbb{Z}$:

$$p(x, x+1) = p, \quad p(x, x-1) = 1-p, \quad \forall x \in \mathbb{Z}.$$

This random walk models a game where a player bets a dollar each round and doubles it with probability p. We can make the following proposition about this specific walk, known as "Gambler's Ruin".

Proposition 5.7. *Let $N \geq 1$. For each $x \in \{1, ..., N-1\}$,*

$$\mathbb{P}[X \textit{ hits } N \textit{ before } 0 | X_0 = x] = \frac{(\frac{1-p}{p})^x - 1}{(\frac{1-p}{p})^N - 1}, p \neq \frac{1}{2}.$$

$$\mathbb{P}[X \textit{ hits } N \textit{ before } 0 | X_0 = x] = \frac{x}{N}, p = \frac{1}{2}.$$

Proof. For any $x \in \{1, ..., N-1\}$, let

$$\alpha(x) := \mathbb{P}[\text{X hits N before } 0 | X_0 = x].$$

The boundary conditions that $\alpha(0) = 0$, $\alpha(N) = 1$. Also, we can use conditional probability to say that

$$\alpha(x) = p\alpha(x+1) + (1-p)\alpha(x-1), \quad \forall x \in \{1, ..., N-1\}.$$

This is a system of N + 1 equations with N + 1 unknowns, so there should be a unique solution. Let us first consider p = 1/2. Then

$$\alpha(x) = \frac{\alpha(x+1) + \alpha(x-1)}{2}.$$

This is satisfied for a linear function, $\alpha(x)$ = ax + b for some a, b. The boundary conditions give $b = 0, a = \frac{1}{N}$.

For $p \neq \frac{1}{2}$, we try to guess a solution of the form $\alpha(x) = b^x$ for some b. This gives

$$b^x = pb^{x+1} + (1-p)b^{x-1}.$$

Thus,

$$b = pb^2 + (1-p).$$

Using the quadratic formula to solve for b, we get that

$$b = \frac{1 \pm (1-2p)}{2p}.$$

Hence, b = 1 or b = $\frac{1-p}{p}$. This gives the general solution:

$$\alpha(x) = c_1 + c_2 \left(\frac{1-p}{p}\right)^x.$$

To satisfy out boundary conditions of $\alpha(0) = 0$, $\alpha(N) = 1$, we need $c_1 = -c_2$ and

$$c_1 = \left(1 - \left(\frac{1-p}{p}\right)^N\right)^{-1}.$$

□

6. Null Recurrence vs. Positive Recurrence

We previously discussed how, for an irreducible Markov Chain $\{X_n\}$ with an infinite state space S, every state is recurrent or transient. Thus, we can categorize every chain as entirely transient or recurrent. We, however, can break down recurrent Markov Chains even further based if they contain a stationary distribution, which we will define next.

Definition 6.1. Let $\pi : S \to [0,1]$ be a probability distribution on S, so that $\sum_{x\in S} \pi_x = 1$. We say that π is *stationary* (a.k.a. invariant) for $\{X_n\}$ if

$$\pi_y = \sum_{x\in S} \pi_x p(x,y), \forall y \in S.$$

A stationary distribution describes the long term states of the Markov chain. We can express it as a row vector π where the i-th entry represents that given arbitrary long amount of time that X_n will be at state i with some probability π_i(i-th component of π). This implies that all the entries of π must sum to one. Then, if we multiply our infinite transition matrix by π, the result is π.

Definition 6.2. Assume that $\{X_n\}$ is irreducible and recurrent. We say that $\{X_n\}$ is *null recurrent* if

$$\lim_{n\to\infty} p^n(x,y) = 0, \forall x,y \in S.$$

We say that $\{X_n\}$ is *positive recurrent* otherwise.

Theorem 6.3. *A null recurrent Markov chain cannot have a stationary distribution.*

Proof. Assume for the sake of contradiction that a null recurrent Markov chain, $\{X_n\}$ has a transition matrix, P, and a stationary distribution, π. Then,

$$\begin{aligned}\pi &= \pi P \\ &= (\pi P)P = \pi P^2 \qquad (6.4)\\ &= ((\pi P)P)P = \pi P^3. \\ &\cdots \\ \pi &= \pi P^n, \forall n \in \mathbb{N}.\end{aligned}$$

However,

$$\begin{aligned}&\lim_{n\to\infty} p^n(x,y) = 0, \forall x,y \in S \\ &\Rightarrow \lim_{n\to\infty} P^n = 0 \\ &\Rightarrow \lim_{n\to\infty} \pi P^n = 0.\end{aligned}$$

Thus, π can only be the **0** vector, which violates the condition that the entries of π sum to 1, Thus, a stationary distribution does not exist for $\{X_n\}$. □

Theorem 6.5. *An aperiodic, positive recurrent Markov Chain will have a stationary distribution (existence)*

Proof. Let our positive recurrent Markov Chain, $\{X_n\}$ have an infinite state space S. Fix $z \in S$ and suppose that we start at $X_0 = z$. Let $T := \min\{n \geq 1 : X_n = z\}$ be the first time we return to z. Note that this is finite since $\{X_n\}$ is irreducible. For $x \in S$,

$$\pi_x := \mathbb{E}[\text{ number of } \{n \in \{0,...,T-1\} : X_n = x\}].$$

More intuitively put, in the infinite row vector π, the value for x is the expected number of times that we reach x, given that we start at z and stop as soon as we return to z.

We want to show that the row vector $\pi/\mathbb{E}[T]$ is a stationary distribution for $\{X_n\}$. By partitioning the set $\{0, ..., T-1\}$ based on the value of $\{X_n\}$, we see that $\sum_{x\in S}\pi_x = \mathbb{E}[T]$. Therefore, $\sum_{x\in S}\pi_x/\mathbb{E}[T] = 1$. Next, we need to show that $\forall y \in S$, $\pi_y = \sum_{x\in S}\pi_x p(x,y)$. The idea of the proof is that π_x is the expected number of visits to x in $\{0, ..., T-1\}$, so $\sum_{x\in S}\pi_x p(x,y)$ is the expected number of visits to y in $\{1, ..., T\}$ (this step relies on positive recurrence). This equals the expected number of visits to y in $\{0, ..., T-1\}$ because T is a stopping time, we can treat time T and time 0 as analogous. The expected number of visits to y in $\{0, ..., T-1\}$, finally, just equals π_y.

Here is the more formal proof. We can use an indicator variable, $\mathbb{1}$, like we did before to help us rewrite π_x.

$$
\begin{aligned}
\pi_x &= \mathbb{E}[\sum_{n=0}^{T-1}\mathbb{1}_{X_n=x}] \\
&= \mathbb{E}[\sum_{n=0}^{\infty}\mathbb{1}_{X_n=x,n<T}] \\
&= \sum_{n=0}^{\infty}\mathbb{E}[\mathbb{1}_{X_n=x,n<T}] \\
&= \sum_{n=0}^{\infty}\mathbb{P}[X_n=x, n<T]
\end{aligned} \tag{6.6}
$$

Because T is a stopping time, the event $\{n < T\}$ is only determined by the prior states of $\{X_n\}$. By the Markov Property, for $y \neq z$, and by Fubini's Theorem for absolutely convergent sums,

$$
\begin{aligned}
\sum_{x\in S}\pi_x p(x,y) &= \sum_{x\in S}\sum_{n=0}^{\infty}\mathbb{P}[X_n=x, n<T]p(x,y) \\
&= \sum_{n=0}^{\infty}\sum_{x\in S}\mathbb{P}[X_n=x, n<T, X_{n+1}=y] \\
&= \sum_{n=0}^{\infty}\mathbb{P}[n<T, X_{n+1}=y] \\
&= \sum_{n=0}^{\infty}\mathbb{P}[n+1<T, X_{n+1}=y] (\text{because } \mathbb{P}[X_T=y]=0) \\
&= \sum_{n=1}^{T-1}\mathbb{E}[\mathbb{1}_{X_n=y}] \\
&= \pi_y - \mathbb{P}[X_0=y] \\
&= \pi_y.
\end{aligned} \tag{6.7}
$$

Note that we used the fact that $y \neq z$, so $\mathbb{P}[X_T = y] = 0$). What if, however, $y = z$?

$$\begin{aligned}\sum_{x \in S} \pi_x p(x,z) &= \sum_{n=0}^{\infty} \mathbb{P}[n < T, X_{n+1} = z] \\ &= \sum_{n=0}^{\infty} \mathbb{P}[n+1 = T] \\ &= 1 = \pi_z\end{aligned} \tag{6.8}$$

□

Next, we need to show that the stationary distribution we found is unique.

Theorem 6.9. *If π is a stationary distribution for an irreducible, aperiodic, and positive recurrent Markov Chain, $\{X_n\}$, for any $x, y \in S$,*

$$\lim_{n \to \infty} \mathbb{P}[X_n = y | X_0 = x] = \pi_y.$$

Proof. We use a technique called coupling: define Markov chains started from x and started from π which interact with each other. Coupling is a probabilistic technique that allows you to consolidate multiple sources of randomness into one single source. For example, we have 2 Markov chains (2 sources of randomness) that we combine into one source of randomness.

Consider the Markov chain (X_n, Y_n) with state space $S \times S$, with transition probabilities

$$\overline{p}((x,y)(x',y')) = \begin{cases} p(x,x')p(y,y'), \text{ if } x \neq y \\ p(x,x'), \text{ if } x = y, x' = y' \\ 0, \text{ if } x = y, x' \neq y' \end{cases}.$$

Basically, the coupled Markov chains work as independent Markov Chains on the X and Y axis until they meet (when x = y), then they move as one along the diagonal. If $x \neq y$, then

$$\begin{aligned}\mathbb{P}[X_1 = x' | X_0 = x, Y_0 = y] &= \sum_{y' \in S} p(x,x')p(y,y') \\ &= p(x,x').\end{aligned} \tag{6.10}$$

If $x = y$, then

$$\mathbb{P}[X_1 = x' | X_0 = x, Y_0 = y] = p(x,x').$$

Thus, $\{X_n\}$ has the same distribution has the same distribution as our original Markov Chain.

Let

$$\tau = \min\{n \geq 0 | X_n = Y_n\}.$$

Note that $X_n = Y_n$ for all $n \geq \tau$. Let us assume the following claim for any $x, y \in S$:

$$\mathbb{P}[\tau < \infty | X_0 = x, Y_0 = y] = 1.$$

Assuming this claim is correct (we will prove it later), consider (X_n, Y_n) where $X_0 = x$ and Y_0 starts from a stationary distribution, π. Then Y_n has distribution π for all n, and $X_n = Y_n$ for all large enough n. Hence for any y,

$$\lim_{n \to \infty} (\mathbb{P}[X_n = y | X_0 = x] - \pi_y) = \lim_{n \to \infty} (\mathbb{P}[X_n = y] - \mathbb{P}[Y_n = y]) = 0.$$

This proves the convergence part of the proposition. The uniqueness follows since if $\pi, \tilde{\pi}$ are two stationary distributions, then

$$\pi = \lim_{n\to\infty} \mathbb{P}[X_n = y | X_0 = x] = \tilde{\pi}.$$

Now, we just need to prove the assumed claim. Consider $(\tilde{X_n}, \tilde{Y_n})$ which move independently for all time (even after they meet). It suffices to show that this Markov chain is irreducible. Let $x, x', y, y' \in S$. Since our original Markov chain is aperiodic, there exists k_0 such that $p^k(x,x) > 0$ for any $k \geq k_0$. Since our original Markov chain is irreducible, there exists n such that $p^n(x,x') > 0$ and there exists $m \geq n + k_0$ such that $p^m(y,y') > 0$. Since $\tilde{X_M}$ and $\tilde{Y_m}$ are independent,

(6.11)

$$\begin{aligned}\mathbb{P}[(\tilde{X_m}, \tilde{Y_m}) = (x', y') | (\tilde{X_0}, \tilde{Y_0}) = (x,y)] &= p^m(x,x')p^m(y,y') \\ &\geq p^{m-n}(x,x)p^n(x,x')p^m(y,y') > 0.\end{aligned}$$

Thus, if we set $x' = y'$, we can see that there is a positive chance the coupled chains will meet; therefore, they will meet with probability 1 in a finite time. □

Thus, positive recurrence means that a stationary distribution exists, and null recurrence means that a stationary distribution does not exist. This is useful as a unique stationary distribution can help us make useful predictions about the evolution of a Markov Chain.

As seen in the next example, often times the easiest way to determine if an infinite Markov Chain is positive recurrent or null recurrent is to see if a has a stationary distribution.

Example 6.12. Consider the biased random walk, $\{X_n\}$ on $\mathbb{Z}$, which has transition probabilities:

$$p(x, x-1) = 1 - p \quad p(x, x+1) = p$$
$$X_0 = 0$$

This is clearly irreducible. Is it positive recurrent, null recurrent, or transient?

Proof. We can look to our Gambler's Ruin result for inspiration. For $x \geq 1$,

$$\mathbb{P}[\text{X hits N before } 0 | X_0 = x] = \frac{(\frac{1-p}{p})^x - 1}{(\frac{1-p}{p})^N - 1}, \quad p \neq \frac{1}{2}.$$

$$\mathbb{P}[\text{X hits N before } 0 | X_0 = x] = \frac{x}{N}, \quad p = \frac{1}{2}.$$

Sending N to ∞ from our Gambler's Ruin example, we see that

$$\mathbb{P}[\text{X does not hit } 0 | X_0 = x] = 0, \quad p < \frac{1}{2}.$$

$$\mathbb{P}[\text{X does not hit } 0 | X_0 = x] = 1 - (\frac{1-p}{p})^x, \quad p > \frac{1}{2}.$$

$$\mathbb{P}[\text{X does not hit } 0 | X_0 = x] = 0, \quad p = \frac{1}{2}.$$

So then what happens when $X_0 = 0$? First, let us consider what happens when $p > \frac{1}{2}$. Then, X_1 will be 1 with probability p, and, from there, the chain has a positive chance to never return to 0, making the chain transient. Next, let us consider what happens when $p < \frac{1}{2}$. Then, X_1 will be -1 with probability $1 - p$. Using symmetry, $1 - p$ is greater than $\frac{1}{2}$. Thus, if the chain gets to -1 (which has a positive chance), it has a positive chance of not getting back to 0. Therefore, the whole chain, considering Proposition 5.3, when $p < \frac{1}{2}$ has a positive chance of not returning to 0. So, when $p \neq \frac{1}{2}$, the chain is transient.

So, what happens when $p = \frac{1}{2}$? Then, X_1 will equal 1 or -1. If it equals 1, we can use the formulas above to see that will return to 0 with probability 1. Furthermore, by symmetry, if $X_1 = -1$ it will also return to 0 with probability 1. Thus, when $p = \frac{1}{2}$, the chain is recurrent.

We need to know, however, if this chain is null recurrent or positive recurrent. As previously explained, the easiest way to do this would be to check if a stationary distribution exists. However, the chain is NOT aperiodic and has a period of 2, so, even if the chain was positive recurrent, a stationary distribution would not exist. To get around this problem, however, we can check if $\{X_{2n}\}$, the version of this random walk where we take two steps at a time (or just look at $\{X_n\}$ every 2 steps), has a stationary distribution. Notice that our state space, $S = 2\mathbb{Z}$. Then, the stationary distribution, π, would have to satisfy the following property:

$$\pi_x = \frac{1}{4}\pi_{x-2} + \frac{1}{2}\pi_x + \frac{1}{4}\pi_{x+2}$$

$$\Rightarrow \pi_x = \frac{1}{2}\pi_{x-2} + \frac{1}{2}\pi_{x+2}.$$

From here, let's consider 2 cases regarding the values of π_{-2} and π_2:
Case 1: $\pi_{-2} = \pi_2$. Then, clearly, $\pi_0 = \pi_{-2} = \pi_2$ which implies using the formula above, that $\pi_x = \pi_y, \forall x, y \in S$. Then, $\sum_{x \in S} \pi_x \neq 1$, so a stationary distribution does not exist.
Case 2: $\pi_{-2} \neq \pi_2$. Without loss of generality, let $\pi_{-2} < \pi_2$. Then, $\pi_{-2} < \pi_0 < \pi_2$. Let $\pi_2 - \pi_0 = d$. Then, $\pi_{-4} = \pi_0 - 2d$, and $\pi_{-2n} = \pi_0 - nd$. Thus, by Archimedes Principle, there exists an $n \in \mathbb{N}$ such that $\pi_{-2n} = \pi_0 - nd < 0$, which contradicts the notion that all entries of π are positive, so a stationary distribution does not exist.

Therefore, no matter what, a stationary distribution cannot exist for $\{X_{2n}\}$, so $\{X_{2n}\}$ is null recurrent which implies that $\{X_n\}$ is null recurrent. □

Example 6.13. Now, let's consider an unbiased random walk on $\mathbb{Z}^d$, with the following probability distribution:

$$p(x, y) = \begin{cases} \frac{1}{2d}, & \text{if y is adjacent to x} \\ 0, & \text{otherwise} \end{cases}.$$

Is the random walk on $\mathbb{Z}^d$ positive recurrent, null recurrent, or transient?

Proof. Notice that we answered this question in the previous example for the $d = 1$ case, but it seems difficult to expand our previous argument to higher dimensions, so let's try another approach which will hopefully be generalizable.

Assume $X_0 = 0$. Let's try calculating $\sum_{n=0}^{\infty} p^n(0, 0)$ directly to determine whether the sum is infinite or not. Recalling that we can only to 0 on even steps, we can try to calculate $P[X_{2n} = 0]$. Notice that, to return to 0, there must be exactly n positive steps and n negative steps. Furthermore, out of the $2n$ total steps, we must choose n of them to be in the positive direction. Thus, there are $\binom{2n}{n}$ possible paths, each with a $(\frac{1}{2})^n(\frac{1}{2})^n$ probability of happening. Therefore,

$$P[X_{2n} = 0] = \binom{2n}{n}(\frac{1}{2})^n(\frac{1}{2})^n = \frac{(2n)!}{(n!)^2}(\frac{1}{2})^{2n}.$$

Unfortunately, it is rather difficult to calculate this infinite sum as written, so we will use, without proof, *Stirling's Approximation* (displayed below) to help us out.

$$n! \sim \sqrt{2\pi} n^{\frac{n+1}{2}} e^{-n}$$

$$\begin{aligned} [P[X_{2n} = 0] &\sim \frac{\sqrt{2\pi}(2n)^{\frac{2n+1}{2}} e^{-2n}}{(\sqrt{2\pi} n^{\frac{n+1}{2}} e^{-n})^2} 2^{-2n} \\ &= \frac{1}{\sqrt{2\pi}} 2^{\frac{2n+1}{2}} n^{\frac{-1}{2}} 2^{-2n} \\ &= \frac{1}{\sqrt{n\pi}} \end{aligned} \tag{6.14}$$

Note, that we can use Stirling's Approximation, because both the approximation and the factorial behave similarly as we tend to infinity. Thus, they would either converge or diverge together. Thus, because we can only return to 0 on even steps,

$$\sum_{n=0}^{\infty} p^n(0,0) = \sum_{n=0}^{\infty} p^{2n}(0,0) = \sum_{n=0}^{\infty} \frac{1}{\sqrt{n\pi}} = \infty.$$

Additionally, $\lim_{n\to\infty} p^n(0,0) = \lim_{n\to\infty} \frac{1}{\sqrt{n\pi}} = 0$ meaning that, when $d = 1$, the chain is null recurrent, which is consistent with conclusion from the previous example.

Now, let's consider $d \geq 2$. There are d components. By the law of large numbers (which we can use in our infinite sum), in $2n$ steps, about $\frac{2n}{d}$ are in each component. Thus, we can redo the approximation for $P[X_{i_{2n}} = 0]$ for the i-th component specifically using $\frac{2n}{d}$ steps. Then, we can raise the result to the d-th power to get the probability that all d components are at 0. So then, what is $P[X_{i_{2n}} = 0]$ for one component specifically? Because we expect to take only $\frac{n}{d}$ steps in the positive and negative direction instead of n, we can replace n with $\frac{n}{d}$ in the formula. Thus,

$$P[X_{i_{2n}} = 0] \sim \frac{1}{\sqrt{\frac{n\pi}{d}}}.$$

$$P[X_{2n} = 0] \sim (\frac{1}{\sqrt{\frac{n\pi}{d}}})^d = \text{constant} \cdot n^{\frac{-d}{2}}.$$

Using our power series convergence rules, we see that $\sum_{n=0}^{\infty} n^{\frac{-d}{2}}$ diverges when $d = 2$, but converges when $d \geq 3$. Thus, when $d \geq 3$, $\sum_{n=0}^{\infty} p^n(0,0) < \infty$, meaning that the random walk is <u>transient</u>.

On the other hand, $\sum_{n=0}^{\infty} p^n(0,0) = \infty$ when $d = 2$, so that random walk is recurrent. But is it null recurrent or positive recurrent?
Note that, $\lim_{n\to\infty} p^n(0,0) = \lim_{n\to\infty} n^{\frac{-d}{2}} = 0$.
Additionally, for all $n \in \mathbb{N}, s \in S \setminus \{0\}$, $p^{2n}(0,s) < p^{2n}(0,0)$. Thus, for all $s \in S$, $\lim_{n\to\infty} p^n(0,s) = 0$, so the chain is <u>null recurrent</u>. □

Acknowledgments

I want to thank my mentor, Aaron Ho, for consistent support throughout the REU, and I want to thank Peter May for creating this program where I got to learn about many areas of math.

References

[1] Gregory F. Lawler. *Introduction to Stochastic Processes*. Taylor & Francis Group. 2006.
[2] Ewain Gwynne, Lecture notes for Math 23500: Markov Chains, Martingales, and Brownian Motion. The University of Chicago, Autumn 2023.