MARKOV CHAINS AND MIXING TIMES

BEAU DABBS

Abstract. This paper introduces the idea of a Markov chain, a random process which is independent of all states but its current one. We analyse some basic properties of such processes, introduce the notion of a stationary distribution, and examine methods of bounding the time it takes to become close to such a distribution.

Contents

1. Markov Chains on Finite State Spaces 1
1.1. Markov Chains 1
1.2. Stationary Distributions 3
2. Total Variation Metric and Mixing Times 6
3. Eigenvalues and Eigenfunctions 9
4. Hitting and Covering Times 13
4.1. Hitting Time 13
4.2. Covering Times 18
Acknowledgments 20
References 20

1. Markov Chains on Finite State Spaces

This paper defines Markov chains, and shows some methods which can be used to analyse them. We begin with some basic definitions and properties related to Markov chains.

1.1. Markov Chains.

Definition 1.1. We say a sequence of random variables $X_0, X_1, X_2, \ldots$ with state space Ω and *transition matrix* P is a *Markov chain* if it satisfies the following *Markov property*: for all $x, y \in \Omega$,

$$\mathbb{P}\left(X_{t+1} = y | \bigcap_{s=0}^{t} X_t\right) = \mathbb{P}(X_{t+1}|X_t) = P(x, y).$$

Provided $\mathbb{P}(\bigcap_{s=1}^{t}) > 0$.

Intuitively a Markov chain is a sequence of random variables, such that the next step in the chain of variables is only dependent on the previous variable. Since the xth row of the transition matrix P for a chain is the distribution of $P(x, \cdot)$,

Date: September 18, 2009.

the probability of moving from x to another state in Ω, the matrix P is stochastic, meaning all entries are non-negative, and each row sums to 1.

Now we define some basic properties of Markov chains, and then prove some basic results which will be useful throughout the paper. First consider the following two definitions.

Definition 1.2. We say a Markov chain with transition matrix P is *irreducible* if for any two states $x, y \in \Omega$ there is some $t > 0$ such that $P^t(x, y) > 0$, meaning that there are no two states for which there is no probability of going from one to the other in any number of steps.

Definition 1.3. Let $T(x) = \{t > 1 : P^t(x, x) > 0\}$ be the set of all times when it is possible to have returned to x starting from x. We then define the *period* of a state x to be the greatest common divisor of this set, $\gcd(T(x))$.

Lemma 1.4. *If P is irreducible, then $gcd(T(x)) = gcd(T(y))$ for all $x, y \in \Omega$.*

Proof. Let $g_x = \gcd(T(x))$ and $g_y = \gcd(T(y))$. Since P is irreducible, for any $x, y \in \Omega$ there exist positive integers r and s such that $P^r(x, y) > 0$ and $P^s(y, x) > 0$. Let $t = r + s$, then for any $a \in T(x)$ one has:

$$P^{a+t}(y, y) \geq P^r(x, y) \cdot P^a(x, x) \cdot P^s(x, y) > 0$$

The first inequality follows since restricting the path from y to y can't increase the probability, and the second from our definition of r, s and a. Since Hence for all $a \in T(x)$ we have $a + t \in T(y)$, so g_y divides $a + t$. But since $t \in T(y)$, g_y also divides t, implying that g_y divides a. Thus g_y is a common divisor of $T(x)$ and divides the greatest common divisor g_x. A similar argument shows that g_x also divides g_y, hence the two must be equal. □

This lemma allows us to define periodicity as a property of a Markov chain when the chain is irreducible. We define the period of the chain to be $\gcd(T(x))$ for any state $x \in \Omega$. We say a chain is *aperiodic* if its period is 1. Aperiodicity along with irreducibility then gives us the following proposition.

Proposition 1.5. *If P is aperiodic and irreducible, then there exists $r > 0$ such that $P^r(x, y) > 0$ for all $x, y \in \Omega$.*

The proof of this fact is left to the reader, but can also be found in [Levin, Peres, Wilmer]. The third property we will frequently consider is reversibility.

Definition 1.6. One says a distribution π on Ω is *reversible* with respect to the transition matrix P if it satisfies the following *detail balanced equations*:

$$\pi(x)P(x, y) = P(y, x)\pi(y) \qquad \forall x, y \in \Omega$$

Now let us give an example. Let G be graph with vertex set V and edge set E. If $x, y \in V$ are connect by an edge we write $x \sim y$. We then take the degree of a vertex x to be $deg(x) = |\{y : x \sim y\}|$. We then define the simple random walk on G by letting:

$$P(x, y) = \begin{cases} \frac{1}{\deg(x)} & \text{if } x \sim y \\ 0 & \text{else} \end{cases}$$

While we can consider a random walk on any graph for now consider to simple examples:

Example 1.7. Random Walk on the n-cycle.

Consider the walk on the n-cycle. We can consider this walk as a random walk on $\mathbb{Z}_n$ by having:

$$P(i,j) = \begin{cases} \frac{1}{2} & \text{if } i+1 \equiv j \pmod{n} \\ \frac{1}{2} & \text{if } i-1 \equiv j \pmod{n} \\ 0 & \text{else} \end{cases}$$

For any integer n, the walk on the n-cycle is irreducible, but not necessarily aperiodic. If n is an odd-integer, then the walk is aperiodic. Consider the walk that starts at 0. $T(0)$ as defined above contains both 2, the event that one takes one step forward and one step back, or vice versa, and n, the event that one walks all the way around the cycle. Then since n is odd, $1 = \gcd(2, n) \mid \gcd(T(0))$. But if n is even note that the only way to get back to a starting point is to go an equal number of steps forwards, and backwards, taking an even number of steps, or to have the difference between the number of steps forwards and backwards be a multiple of n. Hence $T(0) = \{2k + tn : k, t > 0\}$, and $\gcd(T(0)) = 2$.

However one can construct a *lazy random walk* on a graph by using the transition matrix $Q = P/2 + I/2$ instead. Hence at every step there is $\frac{1}{2}$ probability that the walk stays where it is, and a $\frac{1}{2}$ chance that the walk follows the distribution of P. Then one can see that $Q(x, x) = 1/2$, and hence $1 \in T(x)$ for all $x \in \mathbb{Z}_n$, making $\gcd(T(x)) = 1$. As one can see, this argument was independent of the distribution of P, so this method of making a *lazy chain* works more generally to make an aperiodic version of any Markov chain.

1.2. Stationary Distributions.

Definition 1.8. We call a distribution π on the state space Ω of a Markov chain *stationary* with respect to the transition matrix P if:

$$\pi = \pi P$$

Or equivalently

$$\pi(y) = \sum_{x \in \Omega} \pi(x) P(x, y) \qquad \forall x, y \in \Omega$$

As an example, let $G = (V, E)$ be a graph, and consider the distribution:

$$\pi(x) = \frac{\deg(x)}{2 \cdot |E|}$$

This distribution is stationary with respect to the simple random walk on G since:

$$\sum_{x \in \Omega} \pi(x) P(x, y) = \sum_{\{x \in \Omega \,:\, x \sim y\}} \frac{\deg(x)}{2 \cdot |E|} \frac{1}{\deg(x)} = \frac{\deg(y)}{2 \cdot |E|} = \pi(y)$$

Also note that this distribution is also reversible with respect to our transition matrix since:

$$\pi(x) P(x, y) = \frac{\deg(x)}{2|E|} \cdot \frac{1}{\deg(x)} = \frac{1}{2|E|} = \frac{\deg(y)}{2|E|} \cdot \frac{1}{\deg(y)} = \pi(y) P(y, x)$$

if $x \sim y$, and otherwise $P(x, y) = 0 = P(y, x)$ so the equality holds trivially. In fact the distribution being both reversible and stationary is not just a coincidence:

Proposition 1.9. *If a distribution π on Ω is reversible with respect to the transition matrix P on Ω, then π is a stationary with respect to P.*

Proof. For any $x, y \in \Omega$:

$$\sum_{x\in\Omega} \pi(x)P(x,y) = \sum_{x\in\Omega} \pi(y)P(y,x) = \pi(y) \cdot \left(\sum_{x\in\Omega} P(y,x)\right) = \pi(y)$$

The first equality follows from reversibility, and the second by using the fact that P is a stochastic matrix. Hence by definition π is stationary. $\square$

In general showing a distribution is reversible can be the simplest way to show that a distribution is stationary, but it is not always possible, since there exist Markov chains with stationary distributions that are not reversible. Yet whether there exists a reversible distribution with respect to a transition matrix or not, one might want to know if every transition matrix at least has a stationary distribution, and this in fact turns out to be true. But before we show this let us introduce the concept of a *hitting time*, and a *stopping time.*

Definition 1.10. We define the *hitting time* for x to be:

$$\tau_x := \min\{t \geq 0 : X_t = x\}$$

and also define:

$$\tau_x^+ := \min\{t \geq 1 : X_t = x\}$$

to be the first positive time one visits x. If a chain begins at x then τ_x^+ is also called the *first return time*

As one might expect, for finite Markov chains, if the chain is irreducible, then the expected time to hit any given point is finite. We state this as the following lemma, but leave the proof to the reader.

Lemma 1.11. *For any states x and y of an irreducible Markov chain, $E_x(\tau_y^+) < \infty$*

Definition 1.12. More generally we define a *stopping time* for (X_t) to be a $\{0, 1, ...\} \cup \{\infty\}$-valued random variable such that for each time t, the event that $\{\tau = t\}$ is determined by $X_0, ...X_t$.

Hence the hitting time can be seen to be one example of a stopping time. We also define the $Green's function$ for a Markov chain with stopping time τ to be,

$$G_\tau(a,x) := E_a(\text{number of visits to } x \text{ before } \tau) = E_a\left(\sum_{t=0}^{\infty} \mathbf{1}_{X_t=x,\tau>t}\right).$$

Theorem 1.13. *If τ is a stopping time for a finite, irreducible Markov chain satisfying $\mathbb{P}_a\{X_\tau = a\} = 1$ and $G_\tau(a,x)$ is the Green's function, then*

$$\frac{G_\tau(a,x)}{E_a(\tau)} = \pi(x) \qquad \textit{for all } x$$

where $\pi(x)$ is the stationary distribution for P.

Proof. We will show directly that $G_\tau(a,x)$ satisfies the stationary distribution property directly. Consider

$$\begin{aligned}\sum_{x\in\Omega} G_\tau(a,x)P(x,y) &= \sum_{x\in\Omega}\sum_{t=1}^{\infty}\mathbb{P}_a(X_t = x, \tau > t)P(x,y)\\ &= \sum_{x\in\Omega}\sum_{t=0}^{\infty}\mathbb{P}_a(X_t = x, X_{t+1} = y, \tau > t)\end{aligned}$$

Now if we switch the order of summations, and sum over $x \in \Omega$ we obtain

$$\begin{aligned}\sum_{x\in\Omega} G_\tau(a,x)P(x,y) &= \sum_{t=0}^{\infty}\mathbb{P}_a(X_{t+1} = y, \tau \geq t+1)\\ &= \sum_{t=1}^{\infty}\mathbb{P}_a(X_t = y, \tau \geq t)\\ &= \sum_{t=0}^{\infty}\mathbb{P}_a(X_t = y, \tau \geq t) - \mathbb{P}_a(X_0 = y, \tau \geq 0)\\ &= \sum_{t=0}^{\infty}\mathbb{P}_a(X_t = y, \tau > t) + \sum_{t=0}^{\infty}\mathbb{P}_a(X_t = y, \tau = t) - \mathbb{P}_a(X_0 = y)\\ &= G_\tau(a,y) + \mathbb{P}_a(X_\tau = y) - \mathbb{P}_a(X_0 = y)\end{aligned}$$

Now to complete the proof consider two cases.

(i) If $y = a$ then by our assumptions

$$\mathbb{P}_a(X_\tau = y) = \mathbb{P}_a(X_0 = y) = 1$$

(ii) If $y \neq a$ then we have

$$\mathbb{P}_a(X_\tau = y) = \mathbb{P}_a(X_0 = y) = 0$$

In either case our last line is equal to $G_\tau(a,y)$, so we have shown that $G_\tau(a,\cdot) = G_\tau(a,\cdot)P$. To normalize to a probability distribution note that $\sum_{x\in\Omega} G_\tau(a,x) = E_a(\tau)$ so for all x

$$\frac{G_\tau(a,x)}{E_a(\tau)} = \pi(x)$$

Is a stationary distribution on Ω □

Since $G_{\tau_x^+}(x,x) = 1$ for all x we obtain the following corollary.

Corollary 1.14. *If P is the transition matrix for an irreducible Markov chain, then*

$$\pi(x) = \frac{1}{E_x(\tau_x^+)}$$

is a stationary distribution for P.

It can further be shown that for an aperiodic Markov chain this stationary distribution is unique, and that if a chain has period n then there are precisely n stationary distributions on Ω. For a proof of these two facts see [Levin, Peres, Wilmer].

Similar to the idea of a stationary distribution is that of a harmonic function.

Definition 1.15. We say a function $h : \Omega \to \mathbb{R}$ is harmonic at x if:

$$h(x) = \sum_{y \in \Omega} P(x, y)h(y)$$

Another way of writing this fact is by considering h as a column vector we have $h = Ph$. If h is harmonic for all $x \in D \subset \Omega$ then we say h is harmonic on D.

We finish up this section with the following proposition, demonstrating one of the most useful facts about harmonic functions. The proof here is left to the reader.

Lemma 1.16. *Suppose that P is an irreducible transition matrix. A function h which is harmonic on Ω is a constant function.*

2. Total Variation Metric and Mixing Times

Now that we have set up the framework for the study of Markov chains we introduce a distance metric for probability distributions, and consider repeated applications of a transition matrix to a distribution, and its effects. We start out with some definitions, and then prove an important convergence theorem.

Definition 2.1. We define the *total variation distance* between two probability distributions μ and ν on the same state space Ω to be:

$$||\mu - \nu||_{TV} = \max_{A \subset \Omega} |\mu(A) - \nu(A)|$$

That is, our distance metric is the largest difference in probability of any single event under the each probability distribution. The following proposition shows a very useful equivalence to this definition.

Proposition 2.2. *Let μ and ν be two probability distributions on Ω. Then:*

$$||\mu - \nu||_{TV} = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|.$$

Proof. Let $B = \{x : \mu(x) \geq \nu(x)\}$ and let $A \subset \Omega$ be any event. Then

$$\mu(A) - \nu(A) \leq \mu(A \cap B) - \nu(A \cap B) \leq \mu(B) - \nu(B)$$

The first inequality follows since $x \in B^C \Rightarrow \mu(x) - \nu(x) < 0$, so removing these terms can only increase the difference in probabilities. The second inequality follows since adding more elements from B also can't decrease the difference.

Similarly one can say that:

$$\nu(A) - \mu(A) \leq \nu(B^C) - \mu(B^C)$$

and in fact these two upper bounds are exactly the same since:

$$\nu(B^C) - \mu(B^C) = (1 - \nu(B)) - (1 - \mu(B)) = \mu(B) - \nu(B)$$

So taking A to be B or B^C in the definition of total variation will give us the maximum difference in probabilities, hence:

$$\begin{aligned}
||\mu-\nu||_{TV} &= \frac{1}{2}\left[\mu(B)-\nu(B)+\nu(B^C)-\mu(B^C)\right] \\
&= \frac{1}{2}\left[\left(\sum_{x\in B}\mu(x)-\nu(x)\right)+\left(\sum_{x\in B^C}\nu(x)-\mu(x)\right)\right] \\
&= \frac{1}{2}\left[\left(\sum_{x\in B}|\mu(x)-\nu(x)|\right)+\left(\sum_{x\in B^C}|\mu(x)-\nu(x)|\right)\right] \\
&= \frac{1}{2}\sum_{x\in\Omega}|\mu(x)-\nu(x)|
\end{aligned}$$

Giving us the desired result. □

Also note that by the proof above one can also say that:

$$||\mu-\nu||_{TV} = \sum_{x\in B}[\mu(x)-\nu(x)] = \sum_{\mu(x)\geq\nu(x)}\mu(x)-\nu(x) \tag{2.3}$$

We already know that all irreducible Markov chains have a stationary distribution, a distribution such that repeated application of the Markov process results in the same distribution, but we don't know if this distribution will ever be reached, unless the chain is started at such a distribution. The following convergence theorem guarantees that repeated application of a transition matrix on any distribution will tend to the stationary distribution. To be more precise:

Theorem 2.4. (**Convergence Theorem**) *Suppose that P is irreducible and aperiodic with stationary distribution π. Then:*

$$||P^t(x,\cdot)-\pi||_{TV}\leq c\alpha^t$$

Proof. First recall that by proposition 1.5 there exists $r>0$ such that P^r contains all strictly positive entries, since P is an irreducible transition matrix. We let Π be the $|\Omega|\times|\Omega|$ matrix with each row being the stationary distribution π. For sufficiently small $\delta>0$ we have

$$P^r(x,y)\geq\delta\pi(y),$$

for all $x,y\in\Omega$. Hence if we let $\theta=1-\delta$, the equation:

$$P^r=(1-\theta)\Pi+\theta Q$$

defines a matrix Q which is guaranteed to be stochastic.

One can further show that for any stochastic matrix M, the matrix $M\Pi=M$ consider the entry in the xth row and yth column of $M\Pi$:

$$(M\Pi)(x,y)=\sum_{z\in\Omega}M(x,z)\Pi(z,y)=\sum_{z\in\Omega}M(x,z)\pi(z)=\pi(z)$$

and for any matrix which satisfies $\pi M=\pi$ we have that $\Pi M=\Pi$, since

$$\begin{pmatrix}\pi\\ \cdots\\ \pi\end{pmatrix}M=\begin{pmatrix}\pi M\\ \cdots\\ \pi M\end{pmatrix}=\begin{pmatrix}\pi\\ \cdots\\ \pi\end{pmatrix}$$

Now we will prove by induction that for every $k > 0$:

$$P^{kr} = (1-\theta^k)\Pi + \theta^k Q^k$$

The case for $n = 1$ is true by definition of our matrix Q. Now suppose that for $n > 0$ we have $P^{nr} = (1-\theta^n)\Pi + \theta^n Q^n$, and consider the case for $n+1$:

$$\begin{aligned} P^{(n+1)r} &= \left((1-\theta^n)\Pi + \theta^n Q^n\right) P^r \\ &= (1-\theta^n)\Pi P^r + \theta^n Q^n P^n \\ &= (1-\theta^n)\Pi + \theta^n Q^n \left[(1-\theta)\Pi + \theta Q)\right] \\ &= (1-\theta^n)\Pi + \theta^n(1-\theta)\Pi + \theta^{n+1}Q^{n+1} = (1-\theta^{n+1})\Pi + \theta^{n+1}Q^{n+1} \end{aligned}$$

This completes the inductive step. Now for any k if we multiply by P^j and rearrange the terms we obtain:

$$P^{rk+j} - \Pi = \theta^k(Q^k P^j - \Pi)$$

Since these matrices are equal, so are each of their rows. Summing the absolute values of the x_0 row on each side of the equation gives us:

$$||P^{rk+j}(x_0,\cdot) - \pi||_{TV} = \theta^k \cdot ||Q^k P^j(x_0,\cdot) - \pi||_{TV} \leq \theta^k$$

The last inequality holding since the greatest value the total variation distance can take is 1. □

Because of this theorem the stationary distribution is also sometimes called the *equilibrium distribution*. But, even though we now know P^t tends to it stationary distribution as t tends to infinity for any transition matrix, we have no indication of how quick this process is, in general, or in specific cases. The rest of this paper develops methods for analyzing the speed at which transition matrices converge to their equilibrium distributions. We define here three measures of distance at a given time t, and show a few relations between them.

Definitions 2.5. Let P be the transition matrix for a Markov chain on state space Ω, with $x, y \in \Omega$. (i)

$$d(t) := \max_{x\in\Omega} ||P^t(x,\cdot) - \pi||_{TV}$$

(ii)

$$\bar{d}(t) := \max_{x,y\in\Omega} ||P^t(x,\cdot) - P^t(y,\cdot)||_{TV}$$

(iii) *separation distance.*

$$s_x(t) := \max_{y\in\Omega}\left[1 - \frac{P^t(x,y)}{\pi(y)}\right]$$

$$s(t) := \max_{x\in\Omega} s_x(t)$$

Proposition 2.6.

$$d(t) \leq \bar{d}(t) \leq 2d(t)$$

Proof. The second inequality simply follows from the triangle inequality. Let $x, y \in \Omega$, then

$$\begin{aligned}||P^t(x,\cdot) - P^t(y,\cdot)||_{TV} &= \sum_{z\in\Omega} |P^t(x,z) - \pi(z) + \pi(z) - P^(y,z)| \\ &\leq \sum_{z\in\Omega} |P^t(x,z) - \pi(z)| + \sum_{z\in\Omega} |P^t(y,z) - \pi(z)| \leq 2d(t)\end{aligned}$$

Since this holds for all x and y it also holds for the max.

To prove the first inequality first note that for any $A \in \Omega$ $\pi(A) = \sum_{y\in\Omega} P(y,A)\pi(y)$. This gives us:

$$\begin{aligned}||P^t(x,\cdot) - \pi(x)||_{TV} = \max_{A\subset\Omega} |P^t(x,A) - \pi(A)| &= \max_{A\subset\Omega} |\sum_{y\in\Omega} \pi(y)(P^t(x,A) - P^t(y,A)| \\ &\leq \max_{A\subset\Omega} \sum_{y\in\Omega} \pi(y)|P^t(x,A) - P^t(y,A)| \\ &\leq \sum_{y\in\Omega} \pi(y) \max_{A\in\Omega} |P^t(x,A) - P^t(y,A)| \\ &= \sum_{y\in\Omega} \pi(y)||P^t(x,\cdot) - P^t(y,\cdot)||_{TV} \\ &\leq \max_{y\in\Omega} ||P^t(x,\cdot) - P^t(y,\cdot)||_{TV}\end{aligned}$$

□

Proposition 2.7. *The separation distance $s_x(t)$ satisfies:*

$$||P^t(x,\cdot) - \pi||_{TV} \leq s_x(t)$$

And thus $d(t) \leq s(t)$

Proof.

$$\begin{aligned}||P^t(x,\cdot) - \pi||_{TV} = \sum_{P^t(x,y)<\pi(y)} \pi(y) - P^t(x,y) &= \sum_{P^t(x,y)<\pi(y)} \pi(y)\left(1 - \frac{P^t(x,y)}{\pi(y)}\right) \\ &\leq \max_{y\in\Omega}\left(1 - \frac{P^t(x,y)}{\pi(y)}\right) = s_x(t)\end{aligned}$$

□

Definition 2.8. We define the mixing time for a distance $0 < \epsilon < 1$to be:

$$t_{mix}(\epsilon) = \min\{t : d(t) < \epsilon\}$$

And we also let $t_{mix} = t_{mix}(\frac{1}{4})$

3. Eigenvalues and Eigenfunctions

In this section we will define a new inner product which is useful when considering transition matrices of Markov. We begin by stating some basic results about transition matrices, and our new inner product. The first result, is left to the reader.

Lemma 3.1. *Let P be the transition matrix of a finite Markov chain.*

(i) If λ is an eigenvalue of P, then $|\lambda| \leq 1$.

(ii) If P is irreducible, the vector space of eigenfunctions corresponding to the eigenvalue 1 is the one-dimensional space generated by the column vecter $\mathbf{1} := (1, 1, ..., 1)^T$.
(iii) If P is irreducible and aperiodic, then -1 is not an eigenvalue of P.

Proposition 3.2. *For a lazy Markov chain with state space Ω, all the eigenvalues of the transition matrix $\hat{P}$ are non-negative.*

Proof. Since $\bar{P}$ is the transition matrix for a lazy Markov chain we can write $\bar{P} = P/2 + I/2$ with P the transition matrix for some other Markov chain. λ is an eigenvalue of $\bar{P}$, with corresponding eigenfunction f iff

$$\bar{P}f - \lambda If = 0$$

But this is equivalent to requiring

$$(P/2 + I/2)f - \lambda If = P/2 - (\lambda - 1/2)If = 0$$

Hence we must have that $P - (2\lambda - 1)If = 0$, and $2\lambda - 1$ be an eigenvalue for some Markov chain. By lemma 3.1 this implies that

$$2\lambda - 1 \geq -1 \qquad \Leftrightarrow \qquad \lambda \geq 0$$

□

In addition to these basic properties, we further define a convenient inner product when dealing with transition matrices. Let $\langle \cdot, \cdot \rangle$ denote the usual inner product on $\mathbb{R}^\Omega$, $\langle f, g \rangle = \sum_{x \in \Omega} f(x)g(x)$. We also define our own inner product denoted

$$\langle f, g \rangle_\pi = \sum_{x \in \Omega} \pi(x)f(x)g(x).$$

The following lemma exemplifies why this inner product useful.

Lemma 3.3. *Let P be a reversible transition matrix w.r.t. π.*
(i) The inner product space $(\mathbb{R}^\Omega, \langle \cdot, \cdot \rangle_\pi)$ has an orthonormal basis of real-valued functions $\{f_i\}_{i=1}^{|\Omega|}$ corresponding to real eigenvalues $\{\lambda_i\}$.
(ii) The matrix P can be decomposed as

$$\frac{P^t(x, y)}{\pi(y)} = \sum_{j=1}^{|\Omega|} f_j(x)f_j(y)\lambda_j^t$$

(iii) The eigenfunction f_1 corresponding to the eigenvalue 1 can be taken to be the constant vector $\mathbf{1}$, in which case

$$\frac{P^t(x, y)}{\pi(y)} = 1 + \sum_{j=1}^{|\Omega|} f_j(x)f_j(y)\lambda_j^t$$

Proof. First we define a new matrix A by letting $A(x, y) = \pi^{1/2}(x)P(x, y)\pi^{-1/2}$, for all $x, y \in \Omega$. One can check that A is a symmetric matrix since P is reversible:

$$\begin{aligned} A(x, y) &= \pi^{-1/2}(x)[\pi(x)P(x, y)]\pi^{-1/2}(y) \\ &= \pi^{-1/2}(x)[P(y, x)\pi(y)]\pi^{-1/2}(y) = A(y, x) \end{aligned}$$

Then the spectral theorem for symmetric matrices guarantees an orthonormal eigenbasis $\{\varphi_j\}_{j=1}^{|\Omega|}$, with each φ_j an eigenfunction with corresponding eigenvalue λ_j. Now consider the eigenfunction $\sqrt{\pi}$.

$$\begin{aligned}(A\sqrt{\pi})(y) &= \sum_{x\in\Omega} A(x,y)\pi^{1/2}(x) \\ &= \sum_{x\in\Omega} \pi^{1/2}(x)\pi^{1/2}(x)P(x,y)\pi^{-1/2}(y) \\ &= \pi^{-1/2}(y)\sum_{x\in\Omega}\pi(x)P(x,y) = \pi^{-1/2}(y)\pi(y) = \sqrt{\pi}(y)\end{aligned}$$

Hence $\sqrt{\pi}$ is an eigenfunction of A with corresponding eigenvalue 1. We now define D_π to be the diagonal matrix with $D_\pi(x,x) = \pi(x)$ for each $x \in \Omega$. Then we can write our new matrix as $A = D_\pi^{1/2}PD_\pi^{-1/2}$. For each j we also let $f_j = D_\pi^{-1/2}\varphi_j$. Then we can show that each f_j is an eigenfunction for the matrix P with corresponding λ_j.

$$Pf_j = PD_\pi^{-1/2}\varphi_j = D_\pi^{-1/2}\left(D_\pi^{1/2}PD_\pi^{-1/2}\right)\varphi_j = D_\pi^{-1/2}A\varphi_j = \lambda_j D^{-1/2}\varphi_j = \lambda_j f_j$$

Further since the φ_j form an orthonormal basis with respect to the normal inner product, for each i, j we have

$$\langle\varphi_i, \varphi_j\rangle = \langle D_\pi^{1/2}f_i, D_\pi^{1/2}f_j\rangle = \langle f_i, f_j\rangle_\pi$$

Hence the f_j form an orthonormal basis with respect to $\langle\cdot,\cdot\rangle_\pi$.

Let δ_y be the function

$$\delta_y(x) = \begin{cases} 1 & \text{if } y = x \\ 0 & \text{if } y \neq x \end{cases}$$

Then using basis decomposition we can rewrite.

$$\delta_y = \sum_{j=1}^{|\Omega|}\langle\delta_y, f_j\rangle f_j = \sum_{j=1}^{|\Omega|} f_j(y)\pi(y)f_j \tag{3.4}$$

Then using the fact that $P^t f_j = \lambda_j^t f_j$ for each j we can write

$$P^t(x,y) = (P^t\delta_y)(x) = \sum_{j=1}^{|\Omega|} f_j(y)\pi(y)P^t f_j(x) = \sum_{j=1}^{|\Omega|} f_j(y)\pi(y)\lambda_j^t f_j(x),$$

completing the proof of (ii). (iii) follows from our earlier consideration of the eigenfunction $\sqrt{\pi}$ of A. Since $D^{-1/2}\sqrt{\pi}(x) = \mathbf{1}$ is an eigenfunction for P with corresponding eigenvalue 1. □

When talking about the eigenvalues for transition matrices, we'll label them from largest to smallest

$$1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_{|\Omega|}$$

Also define the largest eigenvalue in absolute value to be:

$$\lambda_* = \max\{|\lambda| : \lambda \text{ an eigenvalue of } P, \lambda \neq 1\}$$

then $\gamma_* = 1 - \lambda_*$ is called the absolute spectral gap, and if P is aperiodic and irreducible, then lemma 3.1 implies that $\gamma_* > 0$.

We then define the *relaxation time*, t_{rel} to be

$$t_{rel} := \frac{1}{\gamma_*}$$

With this value it is possible to get upper, and lower, bounds on the mixing time for a Markov chain. We first consider an upper bound on the mixing time.

Theorem 3.5. *Let P be the transition matrix of a reversible, irreducible chain with state space Ω, and let $\pi_{min} := \min_{x\in\Omega} \pi(x)$. Then*

$$t_{mix}(\epsilon) \leq \log\left(\frac{1}{\epsilon\pi_{min}}\right) t_{rel}.$$

Proof. Using lemma 3.3 (iii) and the Cauchy-Schwarz inequality we can write

$$\left|\frac{P^t(x,y)}{\pi(y)} - 1\right| \leq \sum_{j=2}^{|\Omega|} |f_j(x)f_j(y)\lambda_*^t| \leq \lambda_*^t \left(\sum_{j=2}^{|\Omega|} f_j^2(x) \sum_{j=2}^{|\Omega|} f_j^2(y)\right)^{\frac{1}{2}} \tag{3.6}$$

Then using equation 3.4 and the orthonormality of $\{f_j\}$ one obtains

$$\begin{aligned}\pi(x) = \langle\delta_x, \delta_x\rangle &= \left\langle \sum_{j=1}^{|\Omega|} f_j(x)\pi(x)f_j, \sum_{i=1}^{|\Omega|} f_i(x)\pi(x)f_i \right\rangle \\ &= \sum_{j=1}^{|\Omega|}\sum_{i=1}^{|\Omega|} f_j(x)f_i(x)\pi^2(x)\langle f_j, f_i\rangle = \pi^2(x)\sum_{j=1}^{|\Omega|} f_j^2(x)\end{aligned}$$

Hence we can say that $\sum_{j=2}^{|\Omega|} f_j^2(x) \leq \pi(x)^{-1}$. Using this fact and equation 3.6 we obtain

$$\left|\frac{P^t(x,y)}{\pi(y)} - 1\right| \leq \frac{\lambda_*^t}{\sqrt{\pi(x)\pi(y)}} \leq \frac{\lambda_*^t}{\pi_{min}} = \frac{(1-\gamma_*)^t}{\pi_{min}} \leq \frac{e^{-\gamma_* t}}{\pi_{min}}$$

Then by proposition 2.7 $d(t) \leq \pi_{min}^{-1} e^{-\lambda_* t}$. Hence by our definition of $t_{mix}(\epsilon)$ we obtain the desired result. □

Theorem 3.7. *For a reversible, irreducible, and aperiodic Markov chain,*

$$t_{mix}(\epsilon) \geq (t_{rel} - 1)\log\left(\frac{1}{2\epsilon}\right).$$

Proof. Begin by considering any eigenfunction f of P with eigenvalue $\lambda \neq 1$, then by orthonormality we have that $\langle f, \mathbf{1}\rangle_\pi = \sum_{x\in\Omega} f(x)\pi(x) = 0$. So letting $||f||_\infty = \max_{x\in\Omega} |f(x)|$, we can write

$$\begin{aligned}|\lambda^t f(x)| = |P^t f(x)| &= \left|\sum_{y\in\Omega} P^t(x,y)f(y)\right| \\ &= \left|\sum_{y\in\Omega} P^t(x,y)f(y) - f(y)\pi(y)\right| \\ &\leq ||f||_\infty \sum_{y\in\Omega} \left|P^t(x,y) - \pi(y)\right| = ||f||_\infty 2d(t)\end{aligned}$$

Then if we take x to be such that $f(x) = ||f||_\infty$ we can obtain the bound $|\lambda|^t \leq 2d(t)$. If we use the definition of t_{mix} we can say that

$$|\lambda|^{t_{mix}(\epsilon)} \leq 2\epsilon \qquad \Leftrightarrow \frac{1}{2\epsilon} \leq \left(\frac{1}{|\lambda|^t}\right)^{t_{mix}(\epsilon)}$$

Considering the logarithm of each side of this inequality we obtain

$$\log\left(\frac{1}{\epsilon}\right) \leq t_{mix}(\epsilon)\log\left(\frac{1}{|\lambda|}\right) \leq t_{mix}(\epsilon)\left(1 - \frac{1}{|\lambda|}\right)$$

Now if we minimize the right hand side over all choices of λ we obtain

$$t_{mix}(\epsilon) \geq \frac{\lambda_*}{1-\lambda_*}\log\left(\frac{1}{2\epsilon}\right) = (t_{rel}-1)\log\left(\frac{1}{2\epsilon}\right).$$

□

Again we can use these methods to consider the random walk on the n-cycle. We'll consider the walk as a walk on the nth roots of unity as a multiplicative group.

Example 3.8. Let $\zeta_n = e^{2\pi i/n}$, and let $W_n = \{\zeta_n, \zeta_n^2, ..., \zeta_n^{n-1}, 1\}$ be the set of the nth roots of unity. These points can be seen as a regular n-gon inscribed in the unit circle, and since $\zeta_n^n = 1$, for all j, k one can say:

$$\zeta_n^j \zeta_n^k = \zeta_n^{j+k \mod n}.$$

So if we consider the random walk as a walk on this group, with transition matrix P, then for any j, and any eigenfunction f of P we have

$$\lambda f(\zeta_n^j) = Pf(\zeta_n^k) = \frac{f(\zeta_n^{k-1}) + f(\zeta_n^{k+1})}{2}.$$

So for $0 \leq k \leq n-1$ define $\varphi_k(\zeta_n^j) := \zeta_n^{jk}$. Then we have

$$P\varphi_k(\zeta_n^j) = \frac{\varphi_k(\zeta_n^{j-1}) + \varphi_k(\zeta_n^{j+1})}{2} = \frac{\zeta_n^{jk-k} + \zeta_n^{jk+k}}{2} = \zeta_n^{jk}\left(\frac{\zeta_n^k + \zeta_n^{-k}}{2}\right).$$

Hence for each k, φ_k is an eigenfunction for our transition matrix, with corresponding eigenvalue $\frac{\zeta_n^k + \zeta_n^{-k}}{2} = \cos(2\pi k/n)$. Hence the largest eigenvalue, that isn't 1 is

$$\lambda_2 = \cos(2\pi/n) = 1 - \frac{4\pi^2}{n^2} + O(n^{-4}).$$

Hence the spectral gap γ has order n^{-2} and t_{rel} has order n^2, and by the previous theorems, the mixing time is also of order n^2.

4. Hitting and Covering Times

4.1. **Hitting Time.** We now return to the idea of a hitting time. Before we defined the hitting time as the expected time to visit a point y from a starting point x. When considering the hitting time for an entire Markov chain, we define the hitting time as the maximum hitting time value between any two points in the chain's state space. More rigorously, for a chain with state space Ω we define:

$$t_{hit} = \max_{x,y\in\Omega} E_x\tau_y$$

Since this value is in a sense the worst case scenario of the time it takes the chain to reach one state from another, it isn't surprising that we can use this value to bound the time is takes the entire chain to settle down to its stationary distribution. We study some properties of hitting times, and give an upper bound on the mixing time in terms of the hitting time. But first we study the target time, whose definition we first justify with the following lemma.

Lemma 4.1. (**Random Target Lemma**) *For an irreducible Markov chain with state space Ω, transition matrix P, and stationary distribution π. the quantity:*

$$\sum_{x\in\Omega} E_a(\tau_x)\pi(x)$$

does not depend on $a \in \Omega$

Proof. For simplicity let $h_x(a) = E_a(\tau_x)$. Then for all $x \neq a$ in Ω we have

$$\begin{aligned} h_x(a) &= \sum_{y\in\Omega} E_a(\tau_x | X_1 = y)P(a,y) \\ &= \sum_{y\in\Omega}(1 + E_y(\tau_x))P(a,y) = \sum_{y\in\Omega}(1 + h_x(y))\, P(a,y) = (Ph_x)(a) + 1 \end{aligned}$$

Hence we can say that

$$(Ph_x)(a) = h_x(a) - 1$$

Now if $x = a$, then:

$$E_a(\tau_a^+) = \sum_{y\in\Omega} E_y(\tau_a^+ | X_1 = y)P(a,y) = \sum_{y\in\Omega}(1 + h_a(y))\, P(a,y) = 1 + (Ph_a)(a)$$

Now by equation corollary 1.14 we can say that

$$Ph_a(a) = \frac{1}{\pi(a)} - 1$$

now let $h(a) = \sum_{x\in\Omega} h_x(a)\pi(x)$. Then

$$\begin{aligned} Ph(a) &= \sum_{x\in\Omega} Ph_x(a)\pi(x) \\ &= \sum_{x\neq a} Ph_x(a)\pi(x) + Ph_a(a)\pi(a) \\ &= \sum_{x\in\Omega}(h_x(a) - 1)\pi(a) - (h_a(a) - 1)\pi(a) + \left(\frac{1}{\pi(a)} - 1\right)\pi(a) \\ &= h(a) - 1 + \pi(a) + (1 - \pi(a)) = h(a) \end{aligned}$$

Note that the last line follows since $h_a(a) = 0$. Hence we have shown that $Ph(a) = h(a)$ for any $a \in \Omega$ so h is a harmonic function and by lemma 1.16 this implies that h is a constant function, thus independent of the choice of a. □

So now we define the *target time* to be

$$t_{\odot} := \sum_{x\in\Omega} E_a(\tau_x)\pi(x) = E_\pi(\tau_\pi)$$

By the random target lemma this definition is equivalent to considering a random draw from Ω for the starting point, giving us:

$$\sum_{x,y\in\Omega} E_x(\tau_y)\pi(x)\pi(y) = E_\pi(\tau_\pi)$$

Lemma 4.2. *For an irreducible Markov chain with state space Ω and stationary distribution π,*

$$t_{hit} \leq 2 \max_{w} E_{\pi}(\tau_w).$$

Proof. for any $a, y \in \Omega$ we have

$$E_a(\tau_y) \leq E_a(\tau_\pi) + E_\pi(\tau_y)$$

Since requiring the path to first pass through some point x selected according to π can only increase the expected time for the path to be completed. Then by the random target lemma we have:

$$E_a(\tau_\pi) = E_\pi(\tau_\pi) \leq \max_{w} E_\pi(\tau_w)$$

Combining these two equations we obtain the desired result. $\square$

Now we find a relationship between hitting time and the mixing time.

Theorem 4.3. *Consider a finite reversible chain with transition matrix P and stationary distribution π on Ω.*

(i) For all $k \geq 0$ and $x \in \Omega$ we have

$$||P^k(x,\cdot) - \pi||_{TV}^2 \leq \frac{1}{4}\left[\frac{P^{2k}(x,x)}{\pi(x)} - 1\right].$$

(ii) If the chain is lazy, that is it satisfies $P(x,x) \geq 1/2$ for all x, then

$$t_{mix}(1/4) \leq 2 \max_{x \in \Omega} E_\pi(\tau_\pi) + 1$$

The first inequality above basically tells us that, for reversible chains, in order to bound the total variation distance from stationarity, one only need make the probability of returning to x close to its stationary probability. Also note that our second inequality is able to relate t_{mix} and t_{hit} since

$$E_\pi(\tau_x) = \sum_{y \in \Omega} E_y(\tau_x)\pi(y) \leq \max_{y \in \Omega} E_y(\tau_x) \leq t_{hit}$$

Hence the theorem implies that

$$t_{mix}(1/4) \leq 2t_{hit} + 1$$

But before we are able to prove this theorem we need a few more results.

Proposition 4.4. *Let P be the transition matrix for a reversible Markov chain on a finite transition space Ω with stationary distribution π.*

(i) For all $t \geq 0$ and $x \in \Omega$ we have $P^{2t+2}(x,x) \leq P^{2t}(x,x)$.

(ii) If the chain P_L is lazy, that is $P_L(x,x) \geq 1/2$ for all x, then for all $t \geq 0$ and $x \in \Omega$ we have $P_L^{t+1}(x,x) \leq P_L^t(x,x)$

Proof. By proposition 3.3 we have the following representation:

$$\frac{P^{2t}(x,x)}{\pi(y)} = \sum_{j=1}^{|\Omega|} f_j(x) f_j(x) \lambda_j^{2t}$$

but since $|\lambda_j| \leq 1$ for all j, we know that $0 \leq \lambda_j^2 \leq 1$. So for each term in the above sum we have $f_j^2(x)\lambda_j^{2t} \geq f_j(x)f_j(y)\lambda_j^{2t+2}$ hence we can say:

$$\frac{P^{2t}(x,x)}{\pi(x)} \geq \sum_{j=1}^{|\Omega|} f_j^2(x)\lambda_j^{2t+2} = \frac{P^{2t+2}(x,x)}{\pi(x)}$$

The second inequality follows simply because the transition matrices for lazy Markov chains have all non-negative eigenvalues.

$$\frac{P_L^t(x,x)}{\pi(x)} = \sum_{j=1}^{|\Omega|} f_j^2(x)\lambda_j^{t} \geq \sum_{j=1}^{|\Omega|} f_j^2(x)\lambda_j^{t+1} = \frac{P_L^{t+1}(x,x)}{\pi(x)}$$

□

We also require the following proposition which allows us to relate the expected hitting time to the return probability of a state x.

Proposition 4.5. *Consider a finite irreducible aperiodic chain with transition matrix P, and stationary distribution π on Ω. Then for any $x \in \Omega$,*

$$\pi(x)E_\pi(\tau_x) = \sum_{t=0}^{\infty}[P^t(x,x) - \pi(x)].$$

Proof. Begin by defining $\tau_x^{(m)} := \min\{t \geq m \,:\, X_t = x\}$ the first time x is visited after step m. Also define $\mu_m := P^m(x,\cdot)$. By theorem 2.4 we know that $\mu_m \to \pi$ as $m \to \infty$. First note that we can simplify the Green's function for this stopping time by writing:

$$G_{\tau_x^{(m)}}(x,x) = \sum_{t=0}^{\infty} \mathbb{P}(X_t = x, \tau_x^{(m)} > t) = \sum_{t=0}^{m-1} P^k(x,x)$$

The summation can be stopped at $m-1$ because for all $t \geq m$, $X_t = x \Rightarrow t \geq \tau_x^{(m)}$. Then by theorem 1.13 we can write:

$$\sum_{t=0}^{m-1} P^k(x,x) = \pi(x)E_x\left(\tau_x^{(m)}\right) = \pi(x)[m + E_{\mu_m}(\tau_x)].$$

Manipulating the equation we can write

$$\sum_{t=0}^{m-1} \left[P^k(x,x) - \pi(x)\right] = \pi(x)E_{\mu_m}(\tau_x^).$$

By our earlier statement, if we let m go to infinity on both sides of the equality, we obtain our result. □

With these two propositions, we can now proof the theorem.

Proof. (i) By equation 2.3 we can write:

$$||P^k(x,\cdot)-\pi||_{TV}^2 = \left(\frac{1}{2}\sum_{y\in\Omega} P^k(x,\cdot)-\pi(y)\right)^2$$
$$= \frac{1}{4}\left(\sum_{y\in\Omega}\left[\pi^{1/2}(y)\right]\left[\pi^{1/2}(y)\left(\frac{P^k(x,y)}{\pi(y)}-1\right)\right]\right)^2$$

Now using the Cauchy-Schwarz inequality we can write

$$||P^k(x,\cdot)-\pi||_{TV}^2 \leq \frac{1}{4}\left(\sum_{y\in\Omega}\pi(y)\right)\left(\sum_{y\in\Omega}\pi(y)\left(\frac{P^k(x,y)}{\pi(y)}-1\right)^2\right)$$
$$= \frac{1}{4}\sum_{y\in\Omega}\pi(y)\left(\frac{P^k(x,y)}{\pi(y)}-1\right)^2$$

Expanding and using the fact that P is a reversible transition matrix, we can write:

$$\frac{1}{4}\sum_{y\in\Omega}\frac{P^k(x,y)\cdot P^k(x,y)}{\pi(y)} - 2P^k(x,y) + pi(y)$$
$$= \frac{1}{4}\left(\sum_{y\in\Omega}\frac{P^k(x,y)\cdot P^k(y,x)}{\pi(x)} - 2\sum_{y\in\Omega}P^k(x,y)+1\right)$$
$$= \frac{1}{4}\left[\frac{P^{2k}(x,x)}{\pi(x)}-2+1\right] = \frac{1}{4}\left[\frac{P^{2k}(x,x)}{\pi(x)}-1\right]$$

(ii) By proposition 4.5 and the monotonicity property we proved in proposition 4.4 (ii), we can write

$$\pi(x)E_\pi(\tau_x) = \sum_{t=0}^{\infty}[P^t(x,x)-\pi(x)] \geq \sum_{t=0}^{2} m[P^t(x,x)-\pi(x)] \leq 2m[P^2m(x,x)-\pi(x)]$$

Now if we divide both sides by $8m\pi(x)$ we obtain

$$\frac{E_\pi(\tau_x)}{8m} \geq \frac{1}{4}[\frac{P^2m(x,x)}{\pi(x)}-1] \geq ||P^m(x,\cdot)-\pi||_{TV}^2$$

Now if $m \geq 2E_\pi(\tau_x)$ then

$$\frac{1}{4} \geq ||P^m(x,\cdot)-\pi||_{TV}$$

So we must have $t_{mix} \leq t_{hit}+1$. □

Now let us return to our example of the random walk on the n-cycle.

Example 4.6. Consider the lazy random walk on the n-cycle. Let P be the transition matrix for the lazy random walk on the n-cycle. We consider the walk as one on Z_n as described in example 1.7. Using theorem 4.3 we can bound the mixing time for the lazy walk by $t_hit = \max_{x,y} E_x(\tau_y)$. Without loss of generality we can consider end point to be 0 since for any two points x and y we can shift the points so that $y = 0$ without changing the underlying distribution.First note that for each $j \in \{1, ...n-1\}$:

$$E_j(\tau_0) = \frac{1}{2}(1 + E_j(\tau_0) \ + frac14(1 + E_{j-1}(\tau_0)) + \frac{1}{4}(1 + E_{j+1})$$

and for $j = 0$, $E_0(\tau_0) = 0$. By identifying $n \equiv 0 \mod n$ we have a system of equations with $f_j = E_j(\tau_0$ for $j = 0, ..., n$ such that $f_0 = f_n = 0$ and

$$f_j = 2 + \frac{1}{2}(f_{j-1} + f_{j+1}) \qquad\qquad j = 1, ..., n-1$$

let $\Delta_j = f_j - f_{j-1}$ for each $j = 1, ..., n-1$. Then one can see that

$$\begin{aligned} &2f_j = 4 + f_{j-1} + f_{j+1} \\ \Leftrightarrow &\Delta_j = f_j - f_{j-1} = 4 + f_{j+1} - f_j = 4 + \Delta_{j+1} \end{aligned}$$

Hence we have an arithmetic progression with

$$\sum_{j=1}^{n} \Delta_j = \sum_{j=1}^{n} (f_j - f_{j-1}) = f_n - f_0 = 0$$

Giving us,

$$0 = \frac{n}{2} \cdot (2\Delta_1 - 4(n-1))$$

Thus

$$\Delta_1 = 2(n-1)$$

and

$$f_j = f_0 + \sum_{k=1}^{j} \Delta_k = \frac{j}{2} \cdot (2\Delta_1 - 4(j-1)) = 2j(n-j)$$

for all $1 \le j \le n-1$. So for the lazy random walk we have $E_j(\tau_0) = 2j(n-j)$. Hence we have

$$t_{hit} = \max_j E_j(\tau_0) = \left\lfloor \frac{n^2}{2} \right\rfloor$$

So theorem 4.3 gives us

$$t_{mix} \le n^2 + 1.$$

Again we find the mixing time for the random walk on the n-cycle is of order n^2, as we did in the previous section with our relaxation time bound.

4.2. **Covering Times.** For any state x in a Markov chain we define the *cover time*τ_{cov} to be the first time at which all states have been visited. If the state space is Ω we say:

$$\tau_{cov} = \min\{t \ : \ \forall y \in \Omega, \exists s \in [0, t] \text{ s.t. } X_s = y\}$$

We also define more generally a covering time for an entire Markov chain.

Definition 4.7. Consider a Markov chain with state space Ω, we define the *covering time* to be

$$t_{cov} = \max_{x \in \Omega} E_x \tau_{cov}$$

Obviously we have a simple lower bound for the covering time in terms of t_{hit}, since the time to go from any one state to another is clearly less than the time it takes to visit all points in the state space. We prove first an upper bound in terms of t_{hit}, and then a tighter lower bound as well.

Theorem 4.8. *Let (X_t) be an irreducible finite Markov chain with n states. Then*

$$t_{cov} \le t_{hit} \cdot \left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right).$$

Proof. Without loss of generality we can consider the n states of our state space to be $1, 2, ..., n$. There are many possible ways to visit all of the states $1, ..., n$, and we let $\sigma \in S_n$ be a permutation of $1, 2, ..., n$ selected in a randomly from S_n independent of the Markov chain. For the ordering $\sigma \in S_n$ we let $T_k = \min\{t : \exists s_1, ...s_k \le t, \text{ s.t. } T_{s_i} = \sigma(i)\}$, and we also let $L_k = X_{T_k}$. It is clear that $E_x[T_1] = E_x[\tau_{\sigma(1)}] \le t_{hit}$. Now we consider the value $E_x[T_k - T_{k-1}]$. This expectation is positive only if $\sigma(k)$ is visited after $\sigma(1), ...\sigma(k-1)$. Otherwise the value will be zero. Define $A_k(r,s)$ to be the event $\{\sigma(k-1) = r, \sigma(k) = s = L_k\}$ for each r, and s we have:

$$E_x[T_k - T_{k-1} | A_k(r,s)] = E_r[\tau_s] \le t_{hit} \tag{4.9}$$

Then if we let

$$A_k = \bigcup_{r \ne s} A_k(r,s)$$

Then A_k is precisely the event that $L_k = \sigma(k)$. Since each of $\sigma(1), ..., \sigma(k)$ are equally likely to be the last state visited $\mathbb{P}(A_k) = 1/k$. Also using equation 4.9 we can say that

$$\begin{aligned} E_x[T_k - T_{k-1} | A_k] &= \sum_{r \ne s} E_x[T_k - T_{k-1} | A_k(r,s)] \mathbb{P}(A_k(r,s) | A_k) \\ &\le \sum_{r \ne s} t_{hit} \mathbb{P}(A_k(r,s) | A_k) = t_{hit} \end{aligned}$$

Hence we can write:

$$E_x[T_k - T_{k-1}] = E_x[T_k - T_{k-1} | A_k] \mathbb{P}(A_k) + E_x[T_k - T_{k-1} | A_k^C] \mathbb{P}(A_k^C) // \quad = \frac{1}{k} E_x[T_k - T_{k-1} | A_k] = \frac{1}{k} t_{hit}$$

Now we consider T_n.

$$\begin{aligned} E_x[T_n] &= E_x[(T_n - T_{n-1}) + (T_{n-1} - T_{n-2}) + \cdots + (T_2 - T_1) + T_1] \\ &= E_x[T_n - T_{n-1}] + \cdots + E_x[T_2 - T_1] + E_x[T_1] \\ &\le t_{hit} \left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right) \end{aligned}$$

□

Now we use a similar method to construct a tighter lower bound on our covering time. We also define τ_{cov}^A to be the first time at which all states of $A \subset \Omega$ have been visited.

Theorem 4.10. *Let $A \subset \Omega$. Set $t_{min}^A = min_{a,b \in A, a \ne b} E_a(\tau_b)$. Then*

$$t_{cov} \ge \max_{A \subset \Omega} t_{min}^A \left(1 + \frac{1}{2} + \cdots + \frac{1}{|A| - 1}\right).$$

Proof. Fix a state $x \in A$, and let σ be permutation chosen randomly from the elements of A, and independent of the chain itself. Let T_k and L_k be defined as above. Then $\mathbb{P}_x(\sigma(1) = x, T_1 = 0) = 1/|A|$, giving us

$$E_x(T_1) \ge \frac{1}{|A|} \cdot 0 + \frac{|A| - 1}{|A|} t_{min}^A = \left(1 - \frac{1}{|A|}\right) t_{min}^A \tag{4.11}$$

Now for $2 \leq k \leq |A|$ and $r, s \in A$, we let

$$B_k(r,s) = \{\sigma(k-1) = r, \sigma(k) = s = L_k\}$$

Then using an argument similar to the above theorem we obtain

$$E_x(T_k - T_{k-1}|B_k(r,s)) = E_r(\tau_s) \geq t^A_{min}. \tag{4.12}$$

Then letting

$$B_k = \bigcup_{r \neq s} B_k(r,s),$$

we find that B_k is the event that $L_k = \sigma(k)$, which has a probability of $1/k$ since σ was chosen uniformly from S_n. Putting this together with equation 4.12 gives us

$$E_x(T_k - T_{k-1} = E_x(T_k - T_{k-1}|B_k)\frac{1}{k} + E_x(T_k - T_{k-1}|B_k^C)\frac{k-1}{k} \geq t^A_{min}\frac{1}{k}$$

Then putting these equations together with 4.11 we obtain

$$E_x(\tau^A_{cov}) = E_x(T_{|A|} - T_{|A|-1}) + \cdots + E_x(T_2 - T_1) + E_x(T_1) \geq t^{|A|}_{min}\left(1 + \frac{1}{2} + \cdots + \frac{1}{|A|-1}\right)$$

Then since for any $A \subset \Omega$, and for every $x \in A$ we have

$$\tau_{cov} \leq E_x(\tau_{cov}) \leq E_x(\tau^A_{cov}),$$

completing the proof. □

Acknowledgments. I'd like to thank both of my mentors, Blair Davey and Shawn Drenning. Blair for help with ideas, and Shawn for helping with the math, and proof reading.

References

[Levin, Peres, Wilmer] Levin, Peres, Wilmer. Markov Chains and Mixing Times. http://www.uoregon.edu/ dlevin/MARKOV/markovmixing.pdf