Markov Chains

Electrical Engineering 126 (UC Berkeley)

Spring 2018

1 Brisk Introduction

This note is not meant to be a comprehensive treatment of Markov chains. Instead, it is intended to provide additional explanations for topics which are not emphasized as much in the course texts.

A sequence of random variables $(X_n)_{n\in\mathbb{N}}$ is a **discrete-time Markov chain (DTMC)** on the **state space** $\mathcal{X}$ if it satisfies the **Markov property**: for all positive integers n and feasible[1] sequences of states $x_0, x_1, \dots, x_{n+1} \in \mathcal{X}$,

$$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n, \dots, X_1 = x_1, X_0 = x_0) = P(x_n, x_{n+1}),$$

where $P(\cdot,\cdot)$ is a set of non-negative numbers such that for all $x \in \mathcal{X}$, $\sum_{y\in\mathcal{X}} P(x,y) = 1$. The Markov property is often summarized by the statement "the past and future are conditionally independent given the present", and it reflects a model in which knowledge of the current state fully determines the distribution of the next state. Be careful however; the random variables in a Markov chain are not in general independent, and in particular, the Markov property does *not* say that X_{n+1} is independent of X_{n-1}.

In this course, we will allow $\mathcal{X}$ to either be finite or countably infinite. When $\mathcal{X}$ is finite, then the numbers $(P(x,y),\ x,y \in \mathcal{X})$ can be organized into a matrix called the **transition probability matrix**; it has the property that its rows sum to 1, and such matrices are called **row stochastic**.

In order to fully specify the joint distribution of a Markov chain, one needs to specify the **initial distribution** π_0, which is a probability distribution on $\mathcal{X}$ representing the distribution of X_0; and the transition probabilities P. The Markov property then gives, for any positive integer n and $x_0, x_1, \dots, x_n \in \mathcal{X}$,

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \pi_0(x_0)P(x_0,x_1)\cdots P(x_{n-1},x_n).$$

[1] The word "feasible" is here because the conditional probability is not well-defined if $\mathbb{P}(X_n = x_n, \dots, X_1 = x_1, X_0 = x_0) = 0$.

Using the rules of probability and the Markov property, the **k-step transition matrix** P_k is

$$\begin{aligned}
P_k(x,y) &:= \mathbb{P}(X_k = y \mid X_0 = x) \\
&= \sum_{x_1,\ldots,x_{k-1}\in\mathcal{X}} \mathbb{P}(X_k = y, X_{k-1} = x_{k-1}, \ldots, X_1 = x_1 \mid X_0 = x) \\
&= \sum_{x_1,\ldots,x_{k-1}\in\mathcal{X}} P(x,x_1)P(x_1,x_2)\cdots P(x_{k-2},x_{k-1})P(x_{k-1},y) \\
&= P^k(x,y),
\end{aligned}$$

where $P^k(x,y)$ is the (x,y) entry of the kth power of P. A consequence is that $P_{k+\ell} = P_k P_\ell$ for all $k,\ell \in \mathbb{N}$; these are known as the **Chapman-Kolmogorov equations**. [2] Usually, we denote the distribution of X_n by the row vector π_n, and then the distribution of π_n is given by matrix-vector multiplication: $\pi_n = \pi_0 P^n$. [3]

As another consequence of the above discussion, if π_0 is a distribution such that $\pi_0 = \pi_0 P$, then the chain will have the same distribution for all time: $\pi_n = \pi_0$ for all $n \in \mathbb{N}$. Such a distribution is called a **stationary distribution** of the chain (also known as an **invariant distribution**). Stationarity plays a central role in the study of Markov chains. The condition $\pi = \pi P$ is more explicitly written as $\pi(x) = \sum_{y\in\mathcal{X}} \pi(y)P(y,x)$ for all $x \in \mathcal{X}$, and these are known as the **balance equations**. In the finite-state case, the balance equations correspond to a set of $|\mathcal{X}|$ linear equations (one of which is redundant), along with the normalization condition $\sum_{x\in\mathcal{X}} \pi(x) = 1$, and the system can be solved via Gaussian elimination (or other methods for calculating eigenvectors).

[2] This may seem like an awfully long name for a relatively simple result. It is difficult to appreciate at this stage, but the Chapman-Kolmogorov equations are important because characterize the structure of the transition dynamics of the chain. They are also useful for computations.

[3] Notice that our convention of writing π_n as a row vector means that the transition matrix P appears on the right; this is the standard notation in probability theory, and it carries advantages (such as $P(x,y)$ being the probability of transitioning *from x to y*).

2 Long-Run Behavior of a Countable-State Markov Chain

This section aims to give a big picture overview of the results about the long-run behavior of countable-state discrete-time Markov chains. The proofs will not be given here.

2.1 Recurrence & Transience

When discussing Markov chains, the mental picture to have is that of a particle which jumps from state to state in the state space according to the transition probabilities of the chain, and the random variable X_n keeps track of the location of the particle at time n.

The first key insight about Markov chains is that, while some states will be visited by the chain over and over again, other states will only be visited a few times and then never seen again. To formalize this, let us define some standard notation. For each $x \in \mathcal{X}$, the random variable T_x represents the first time that the chain visits state x, i.e., $T_x := \min\{n \in \mathbb{N} : X_n = x\}$ (this is called the **hitting time** of state x). We will also need the random variable $T_x^+ := \min\{n \in \mathbb{Z}_+ : X_n = x\}$, which is the hitting time for state x except that we do not let T_x^+ equal 0 when the chain starts at x. Also, the notations $\mathbb{P}_x$ and $\mathbb{E}_x$ mean that the chain is started at state x, that is,

$$\mathbb{P}_x(\cdot) := \mathbb{P}(\cdot \mid X_0 = x),$$
$$\mathbb{E}_x[\cdot] := \mathbb{E}[\cdot \mid X_0 = x].$$

Then, for $x, y \in \mathcal{X}$, we define $\rho_{x,y} := \mathbb{P}_x(T_y^+ < \infty)$, the probability that starting from state x we eventually reach state y, and $\rho_x := \rho_{x,x}$ for simplicity. Now, we say that a state x is **recurrent** if $\rho_x = 1$ and **transient** if $\rho_x < 1$.

Proposition 1. *Let N_x denote the total number of visits to state x, that is, $N_x := \sum_{n\in\mathbb{N}} \mathbb{1}\{X_n = x\}$. If x is recurrent, then $N_x = \infty$ $\mathbb{P}_x$-a.s., so in particular $\mathbb{E}_x[N_x] = \infty$. If x is transient, then $\mathbb{E}_x[N_x] < \infty$; in fact,*

$$\mathbb{E}_x[N_x] = \frac{\rho_x}{1 - \rho_x} < \infty.$$

In particular, $N_x < \infty$ $\mathbb{P}_x$-a.s. [4]

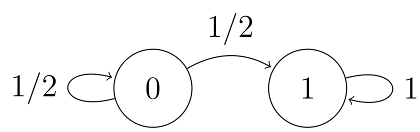
[4] Suppose that X is non-negative. If X takes on the value ∞ with positive probability p, then $\mathbb{E}[X] \geq \infty \cdot p = \infty$. Thus, if $\mathbb{E}[X] < \infty$, then $X < \infty$ with probability 1.

In the above result, the notation $\mathbb{P}_x$-a.s. means that the event occurs almost surely when the chain is started from state x, i.e., $\mathbb{P}_x(\cdot) = 1$.

The above result formalizes the intuition about recurrent and transient states: starting from a recurrent state x, then x will be visited infinitely many times by the chain. Starting from a transient state x, the chain will only visit x finitely many times, and then never return.

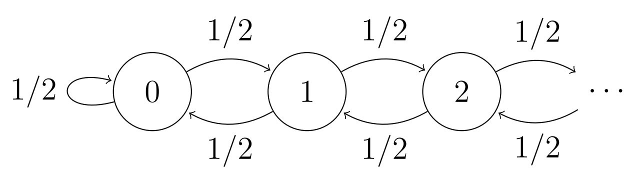
What is an example of a recurrent state or a transient state? We will shortly describe the main classification result, which gives an easy way of figuring out which states are transient and which states are recurrent by looking at the **transition diagram** of the Markov chain. The transition diagram is the directed graph associated with the Markov chain, where the vertices are the states in $\mathcal{X}$, and the edge (x, y) is drawn in the transition diagram if and only if $P(x, y) > 0$. However, we can treat a few examples from definitions alone.

Example 1. Consider the two-state chain:



Here it is clear that once we are in state 1, we will never leave 1, whereas if we are in state 0, then eventually we will leave state 0 and move to state 1. Therefore state 0 is transient and state 1 is recurrent.

Example 2. Consider the simple random walk which is “reflected” at 0:



We claim that state 0 is recurrent. If we start in state 0, then with probability 1/2 we will return to state 0 immediately; otherwise, we will move to state 1, which yields the equation

$$\rho_0 = \frac{1}{2} + \frac{1}{2}\rho_{1,0}.$$

Now, once we are in state 1, then with probability 1/2 we will reach state 0; otherwise, with probability 1/2 we will reach state 2, so

$$\rho_{1,0} = \frac{1}{2} + \frac{1}{2}\rho_{2,0}.$$

However, in order to reach state 0 from 2, we must reach 1 from 2 and 0 from 1, so $\rho_{2,0} = \rho_{2,1}\rho_{1,0}$. By symmetry, $\rho_{2,1} = \rho_{1,0}$, so $\rho_{2,0} = \rho_{1,0}^2$. Substituting this into the above equations yields

$$\rho_{1,0} = \frac{1}{2} + \frac{1}{2}\rho_{1,0}^2$$

and this is seen to imply $\rho_{1,0} = 1$; thus, $\rho_0 = 1$ as well. Already we have arrived at a result that is not immediately obvious: the reflected random walk will visit the origin infinitely often (instead of "drifting off to ∞").

Proposition 2. *A finite-state DTMC has at least one recurrent state.*

After all, the chain has to spend its time *somewhere*, and if it visits each of its finitely many state finitely many times, then where else could it go? However, this is not true for infinite-state Markov chains.

Example 3. Consider the following chain:

$$0 \xrightarrow{1} 1 \xrightarrow{1} 2 \xrightarrow{1} \cdots$$

Clearly the chain drifts off to ∞ and every state is transient.

2.2 Classification of States

We say that state x **communicates** with state y if $\rho_{x,y} > 0$ and $\rho_{y,x} > 0$. In words, it is possible (through a sequence of transitions with non-zero probability) to reach state y from state x, and it is also possible to reach state x from state y. A **communicating class** is a maximal set of states which communicate with each other. In graph terminology, a communicating class is a *strongly connected component (SCC)* in the transition diagram of the chain. The set of communicating classes of the chain partition the state space, and this concept will allow us to take a Markov chain with a complicated structure and decompose it into smaller chains.

We say that a Markov chain is **irreducible** if it consists of only a single communicating class. An alternate way to describe irreducibility is that for any pair of states x and y, it is possible to reach x from y and vice versa.

We say that a property of a state is a **class property** if the property is necessarily shared by all members of a communicating class. In this case, then the property is not really a property of the *state*, but rather a property of the entire communicating class. The key classification result is:

Theorem 1 (Classification of States)**.** *Recurrence and transience are class properties.*

We can now speak of recurrent and transient *classes*, rather than restricting ourselves to recurrent and transient *states*. In the case when the Markov chain is irreducible, then there is only one communicating class, so we can speak of the entire Markov chain as being recurrent or transient. The following observations are helpful for classifying the states of a Markov chain.

- From Proposition 2 it follows that every finite-state irreducible chain is recurrent. More generally, any finite communicating class which has no edges leaving the communicating class (the class is **closed**) is recurrent.
- If a state has an edge which leads outside of the communicating class in which it belongs, then the state is transient.
- If a state is recurrent, then any state it can reach is also recurrent ("recurrence is contagious").

Given the above observations, see if you can formulate an algorithm for classifying all of the states of a finite-state Markov chain given only the transition diagram.

Example 4. Let us classify all of the states in the examples above.

- Example 1. The communicating classes are $\{0\}$ and $\{1\}$. Since $\{0\}$ has an edge to $\{1\}$, it is transient, and since $\{1\}$ is closed it is recurrent.
- Example 2. The chain is irreducible, and since we have shown that 0 is recurrent, it follows that the entire chain is recurrent.
- Example 3. The communicating classes are $\{i\}$ for $i \in \mathbb{N}$ and each class is transient.

2.3 Positive Recurrence & Null Recurrence

In this section we address the existence and uniqueness of the stationary distribution.

First of all, to understand the terminology, a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ is called **stationary** if for all positive integers k, n, and all events $A_1, \dots, A_n$, then

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{P}(X_{k+1} \in A_1, \dots, X_{k+n} \in A_n).$$

In other words, the distribution of $(X_1, \dots, X_n)$ is the same as the joint distribution after we shift the time index by k to $(X_{k+1}, \dots, X_{k+n})$. Stationarity is important because many stochastic processes that have a stationary regime will, under suitable conditions, *converge* to stationarity in some sense. Consequently, stationarity is a powerful simplifying assumption that is justified for systems that have been running for a long period of time. [5] For a Markov chain $(X_n)_{n\in\mathbb{N}}$, convince yourself that $(X_n)_{n\in\mathbb{N}}$ is stationary if and only if the chain is started from its stationary distribution (so the terminology is consistent).

We will start with a crucial interpretation of the stationary distribution. We will focus on the irreducible case.

Theorem 2. *Suppose that the Markov chain is irreducible with a stationary distribution π. Then, for each $x \in \mathcal{X}$,*

$$\pi(x) = \frac{1}{\mathbb{E}_x[T_x^+]}.$$

Understanding this theorem carefully sheds light on much of the convergence theory, so let us take the time to sketch the ideas involved. Markov chains are a generalization of i.i.d. random variables because they introduce a dependence structure, so they are more difficult to study than i.i.d. random variables. The key to analyzing Markov chains is to look for the underlying i.i.d. structure hidden within the Markov chain.

The Markov property says that once the current state is known, then the past history of the chain is irrelevant from the point of view of predicting the future. Therefore, if at two different points in time we are in state x, then

[5] For example, it turns out that the AEP still holds if we replace the i.i.d. assumption with the assumption of stationarity (along with another property called *ergodicity* which we will not discuss).

at both times the future of the Markov chain has the same distribution. We can formalize the idea by defining random variables $T_x(k)$ to be the time at which we visit state x for the kth time after 0, where $T_x(0) = 0$. Start the chain at state x. Then, the *blocks*

$$(X_0, \dots, X_{T_x(1)-1}), (X_{T_x(1)}, \dots, X_{T_x(2)-1}), (X_{T_x(2)}, \dots, X_{T_x(3)-1}), \dots$$

are i.i.d. To interpret this, note that each block above starts at state x, so each block can be viewed as an *excursion* which leaves x and eventually makes its way back to x; when the chain reaches x again, a new block begins. Notice that in order to make sense of this idea, the state x must be recurrent, or else there is a change that the excursion outside of x will never end.

Since the blocks are i.i.d., then any function applied to the blocks will produce i.i.d. random variables. So, let $\tau_1, \tau_2, \tau_3, \dots$ denote the *length* of the blocks. They are i.i.d., with mean $\mathbb{E}[\tau_1] = \mathbb{E}_x[T_x^+]$, so by the SLLN, $n^{-1} \sum_{i=1}^n \tau_i \to \mathbb{E}_x[T_x^+]$. In the time $\tau_1 + \cdots + \tau_n$ we have visited x exactly n times, so $n^{-1} \sum_{i=1}^n \tau_i$ is the total time divided by the number of visits to x. Alternatively, if we fix a time t, then the total time divided by the number of visits to x is $t/(\sum_{i=0}^{t-1} \mathbb{1}\{X_i = x\})$, so we can expect that

$$\frac{t}{\sum_{i=0}^{t-1} \mathbb{1}\{X_i = x\}} \to \mathbb{E}_x[T_x^+]$$

or

$$\frac{1}{t} \sum_{i=0}^{t-1} \mathbb{1}\{X_i = x\} \to \frac{1}{\mathbb{E}_x[T_x^+]}.$$

Now, we may take expectations[6] to get $t^{-1} \sum_{i=0}^{t-1} \mathbb{P}(X_i = x) \to 1/\mathbb{E}_x[T_x^+]$. If we start the chain at the stationary distribution π, then $\mathbb{P}(X_i = x) = \pi(x)$ for all $i \in \mathbb{N}$, so we get $\pi(x) = 1/\mathbb{E}_x[T_x^+]$.

We already observed that the argument does not make sense if x is transient, and also the argument does not make sense if $\mathbb{E}_x[T_x^+] = \infty$. If x is a transient state, then intuitively it cannot support any stationary probability mass because all of the probability mass in x must eventually flow out of x and never return. If $\mathbb{E}_x[T_x^+] = \infty$, then yet another phenomenon occurs: although the state x may be recurrent, so x is visited infinitely many times,

[6] In order to do this, we must use the Dominated Convergence Theorem, which is valid because $t^{-1} \sum_{i=0}^{t-1} \mathbb{1}\{X_i = x\}$ is bounded above by 1 for all $t \in \mathbb{N}$.

the times between successive visits to x are so long that x cannot support any stationary probability mass either!

We define x to be **positive recurrent** if x is recurrent and $\mathbb{E}_x[T_x^+] < \infty$; otherwise, we say x is **null recurrent** if x is recurrent and $\mathbb{E}_x[T_x^+] = \infty$.

Theorem 3. *Positive recurrence and null recurrence are class properties.*

Our picture of the classification of states is now more subtle. We have transient states and recurrent states, and then recurrent states further separate into positive recurrent and null recurrent states. Null recurrence is a *new* phenomenon that only comes up in the infinite-state case.

Proposition 3. *Any finite-state irreducible chain is positive recurrent.*

Intuitively, null recurrence is a subtle phenomenon which lies at the boundary between positive recurrence and transience.

We are now ready to state the formal theorems.

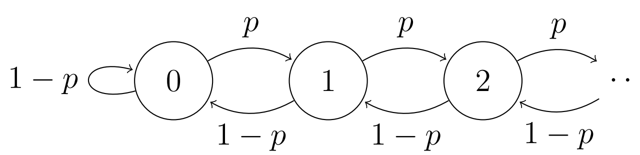
Theorem 4. *An irreducible positive recurrent Markov chain has a unique stationary distribution. In fact, the converse is true too: an irreducible Markov chain is positive recurrent if and only if a stationary distribution exists.*

The above result says that "positive recurrent" and "stationary distribution" are practically synonymous. What does it mean when a stationary distribution does *not* exist? It means that if μ is a non-negative solution to the balance equations, $\mu = \mu P$, then it is impossible to normalize it: $\sum_{x\in\mathcal{X}} \mu(x) = 0$ or $\sum_{x\in\mathcal{X}} \mu(x) = \infty$ (this is one way to show that the chain is *not* positive recurrent). Otherwise, if $\sum_{x\in\mathcal{X}} \mu(x)$ is a positive finite constant c, then $\pi := c^{-1}\mu$ is the stationary distribution.

Theorem 5. *For an irreducible positive recurrent Markov chain with stationary distribution π, for each state x, the fraction of time spent in state x, $n^{-1}\sum_{i=0}^{n-1} \mathbb{1}\{X_i = x\}$, converges a.s. as $n \to \infty$ to $\pi(x)$.*

Finally, we will say what happens when the chain is not irreducible. Using the classification of states, we can separate the chain into communicating classes, and the transient and null recurrent classes can support no stationary mass (so if there are no positive recurrent classes, there are no stationary distributions). If there is more than one positive recurrent class, then each one can be viewed as an irreducible positive recurrent Markov chain with its own stationary distribution. Then, any convex combination of these individual stationary distributions will yield a stationary distribution for the overall chain, so in this case the stationary distribution is not unique.

Example 5. Consider the reflected random walk where the probability of moving forwards is $p \in (0, 1)$.



The chain is irreducible, and it can be classified as:

- positive recurrent, when $p < 1/2$;
- null recurrent, when $p = 1/2$;
- transient, when $p > 1/2$.

In the first case, $p < 1/2$, check that $\pi(k) := (1 - \rho)\rho^k$ for $k \in \mathbb{N}$, where $\rho := p/(1 - p)$, is the stationary distribution of the chain. [7] The existence of the stationary distribution confirms positive recurrence.

In the second case, $p = 1/2$, we already showed in Example 2 that the chain is recurrent. We can check by hand that $\mathbb{E}_0[T_0^+] = \infty$. Indeed, using the first-step equations,

$$\mathbb{E}_1[T_0^+] = 1 + \frac{1}{2}\,\mathbb{E}_0[T_0^+] + \frac{1}{2}\,\mathbb{E}_2[T_0^+].$$

Observe that $\mathbb{E}_2[T_0^+]$ is the expected time to travel two steps to the left, which by symmetry is $\mathbb{E}_2[T_1^+] + \mathbb{E}_1[T_0^+] = 2\,\mathbb{E}_1[T_0^+]$, so

$$\mathbb{E}_1[T_0^+] = 1 + \frac{1}{2}\,\mathbb{E}_0[T_0^+] + \mathbb{E}_1[T_0^+].$$

This equation can only be satisfied if $\mathbb{E}_1[T_0^+] = \infty$. Then, since

$$\mathbb{E}_0[T_0^+] = 1 + \frac{1}{2}\,\mathbb{E}_1[T_0^+]$$

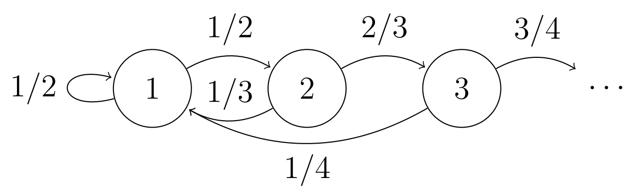
it follows that $\mathbb{E}_0[T_0^+] = \infty$ as well, so the chain is null recurrent.

In the third case, $p > 1/2$, intuitively the chain drifts more towards the right than to the left and it is therefore transient. Here is an argument using

[7] To check this, it is easiest to verify the detailed balance conditions.

the SLLN, which is trickier than the previous two arguments. For each $n \in \mathbb{N}$, let $Y_n := -1$ if the chain takes a transition to the "left" on the nth transition, and let $Y_n := +1$ if the chain takes a transition to the "right" on the nth transition (so $Y_1, Y_2, Y_3, \dots$ are i.i.d. with $\mathbb{P}(Y_1 = 1) = p = 1 - \mathbb{P}(Y_1 = -1)$). Then, $X_n = 0$ only if $\sum_{i=1}^n Y_i \le 0$, which occurs if and only if $n^{-1} \sum_{i=1}^n Y_i \le 0$. However, by the SLLN, $n^{-1} \sum_{i=1}^n Y_i \xrightarrow[n\to\infty]{\text{a.s.}} \mathbb{E}[Y_1] > 0$, which means that $n^{-1} \sum_{i=1}^n Y_i \le 0$ can only happen finitely many times, a.s. Therefore, $X_n = 0$ can only happen finitely many times a.s., but we know that any recurrent state is visited infinitely many times, so 0 is transient (and thus the entire chain is transient).

Example 6. Consider the following chain:



In other words, $P(i, 1) = 1 - P(i, i+1) = 1/(i+1)$ for all positive integers i. The chain is irreducible and the probability of never returning to state 1 starting from state 1 is

$$\lim_{n\to\infty} \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} \cdots \frac{n}{n+1} = \lim_{n\to\infty} \frac{1}{n+1} = 0,$$

so the chain is recurrent. It is null recurrent because the expected time to reach state 1 starting from state 1 is

$$\begin{aligned} \mathbb{E}_1[T_1^+] &= 1 + \frac{1}{2}\,\mathbb{E}_2[T_1^+] = 1 + \frac{1}{2}\Big(1 + \frac{2}{3}\,\mathbb{E}_3[T_1^+]\Big) = 1 + \frac{1}{2} + \frac{1}{3}\,\mathbb{E}_3[T_1^+] = \cdots \\ &= 1 + \frac{1}{2} + \frac{1}{3} + \cdots = \infty. \end{aligned}$$

2.4 Aperiodicity & Convergence

Suppose that we have an irreducible positive recurrent chain with stationary distribution π. In this section we are interested in the question of when $\pi_n \to \pi$, in the sense that $\pi_n(x) \to \pi(x)$ for each $x \in \mathcal{X}$ as $n \to \infty$, *regardless* of the initial distribution π_0. Clearly, if we start with $\pi_0 = \pi$, then $\pi_n \to \pi$, but we want to find assumptions on the chain under which the initial condition π_0 of the chain does not matter in the long run.

The assumption that we need here is aperiodicity. To build up to it, the **period** of a state x is $d(x) := \gcd\{n \in \mathbb{Z}_+ : P^n(x,x) > 0\}$. To parse this definition, look at the set $\{n \in \mathbb{Z}_+ : P^n(x,x) > 0\}$: this is the set of times at which it is possible to return to state x starting from state x. Then, we look at the greatest common divisor of this set. In particular, if a state has a self-loop, then its period is 1.

Theorem 6. *The period of a state is a class property.* [8]

So, for an irreducible positive recurrent chain, the period of every state is the same. We may now define the chain to be **aperiodic** if its period is 1, otherwise it is **periodic**.

The following theorem summarizes what the period of the Markov chain has to say about its behavior, and it provides us with valuable intuition about periodic Markov chains.

Theorem 7. *Suppose that a irreducible recurrent Markov chain has period d.*

1. *The state space $\mathcal{X}$ can be partitioned into d classes $\mathcal{X}_1, \dots, \mathcal{X}_d$ which flow in a cyclic fashion into each other: states in $\mathcal{X}_i$ only transition to states in $\mathcal{X}_{i+1 \bmod d}$.*

2. *The dth power of the transition matrix, P^d, has d closed communicating classes, and P^d is an irreducible transition probability matrix on each of the classes $\mathcal{X}_1, \dots, \mathcal{X}_d$.*

3. *If the chain is aperiodic, then for sufficiently large n, P^n has only positive entries (the matrix is called **regular**).*

For the aperiodic case we have:

Theorem 8 (Convergence Theorem)**.** *For an irreducible positive recurrent aperiodic chain with stationary distribution π, then $\pi_n \to \pi$.*

The proof method used to prove the convergence result is known as **coupling**, and using the idea of coupling can lead to quantitative results about the *rate* of convergence, which in turn leads to the analysis of the performance of algorithms. One famous result in this area is the result that seven riffle shuffles suffices to produce a well-shuffled deck of cards.

[8] It can be awkward to define the period of a transient state, but the period is mainly a useful concept for irreducible recurrent chains anyway.

2.5 A Word on Linear Algebra

The theory of finite-state Markov chains can be viewed either probabilistically, or through the eyes of linear algebra. We will not explore the latter viewpoint in much detail, but here we will briefly explain the connection.

The stationary distribution is a non-negative left eigenvector of eigenvalue 1 for the transition probability matrix. The main theorem from the linear algebra perspective is called the **Perron-Frobenius Theorem**. One of its conclusions is that all of the eigenvalues are bounded in magnitude by 1. Therefore, we would like to argue that the stationary distribution corresponds to the dominant eigenvalue, and as we take $n \to \infty$, the other eigenvalues will go to 0. When all other eigenvalues are strictly less than 1 in magnitude and the eigenspace corresponding to the stationary distribution has dimension one, which is true for the irreducible aperiodic chain, then this argument provides a proof of the convergence result.

In the aperiodic case, then there are multiple eigenvalues of the transition matrix with magnitude 1 (they are the dth roots of unity for period d). This is the essential barrier to convergence for periodic Markov chains.