Copyright © 2009 by Karl Sigman

1 Limiting distribution for a Markov chain

In these Lecture Notes, we shall study the limiting behavior of Markov chains as time $n \to \infty$. In particular, under suitable easy-to-check conditions, we will see that a Markov chain possesses a limiting probability distribution, $\pi = (\pi_j)_{j \in \mathcal{S}}$, and that the chain, if started off initially with such a distribution will be a stationary stochastic process. We will also see that we can find π by merely solving a set of linear equations.

1.1 Communication classes and irreducibility for Markov chains

For a Markov chain with state space $\mathcal{S}$, consider a pair of states (i, j). We say that j is reachable from i, denoted by $i \to j$, if there exists an integer $n \geq 0$ such that $P^n_{ij} > 0$. This means that starting in state i, there is a positive probability (but not necessarily equal to 1) that the chain will be in state j at time n (that is, n steps later); $P(X_n = j | X_0 = i) > 0$. If j is reachable from i, and i is reachable from j, then the states i and j are said to *communicate*, denoted by $i \longleftrightarrow j$. The relation defined by communication satisfies the following conditions:

1. *All states communicate with themselves:* $P^0_{ii} = 1 > 0$.[1]
2. *Symmetry:* If $i \longleftrightarrow j$, then $j \longleftrightarrow i$.
3. *Transitivity:* If $i \longleftrightarrow k$ and $k \longleftrightarrow j$, then $i \longleftrightarrow j$.

The above conditions imply that communication is an example of an equivalence relation, meaning that it shares the properties with the more familiar equality relation " $=$ ":
$i = i$. If $i = j$, then $j = i$. If $i = k$ and $k = j$, then $i = j$.

Only condition 3 above needs some justification, so we now prove it for completeness: Suppose there exists integers n, m such that $P^n_{ik} > 0$ and $P^m_{kj} > 0$. Letting $l = n + m$, we conclude that $P^l_{ij} \geq P^n_{ik} P^m_{kj} > 0$ where we have formally used the Chapman-Kolmogorov equations. The point is that the chain can (with positive probability) go from i to j by first going from i to k (n steps) and then (independent of the past) going from k to j (an additional m steps).

If we consider the rat in the open maze, we easily see that the set of states $C_1 = \{1, 2, 3, 4\}$ all communicate with one another, but state 0 only communicates with itself (since it is an absorbing state). Whereas state 0 is reachable from the other states, $i \to 0$, no other state can be reached from state 0. We conclude that the state space $\mathcal{S} = \{0, 1, 2, 3, 4\}$ can be broken up into two disjoint subsets, $C_1 = \{1, 2, 3, 4\}$ and $C_2 = \{0\}$ whose union equals $\mathcal{S}$, and such that each of these subsets has the property that all states within it communicate. Disjoint means that their intersection contains no elements: $C_1 \cap C_2 = \emptyset$.

A little thought reveals that this kind of disjoint breaking can be done with any Markov chain:

Proposition 1.1 *For each Markov chain, there exists a unique decomposition of the state space $\mathcal{S}$ into a sequence of disjoint subsets $C_1, C_2, \ldots$,*

$$\mathcal{S} = \cup_{i=1}^{\infty} C_i,$$

in which each subset has the property that all states within it communicate. Each such subset is called a communication class of the Markov chain.

[1] $P^0_{ii} = P(X_0 = i | X_0 = i) = 1$, a trivial fact.

If we now consider the rat in the closed maze, $\mathcal{S} = \{1,2,3,4\}$, then we see that there is only one communication class $C = \{1,2,3,4\} = \mathcal{S}$: all states communicate. This is an example of what is called an *irreducible* Markov chain.

Definition 1.1 *A Markov chain for which there is only one communication class is called an irreducible Markov chain: all states communicate.*

Examples

1. *Simple random walk is irreducible.* Here, $\mathcal{S} = \{\cdots -1, 0, 1, \cdots\}$. But since $0 < p < 1$, we can always reach any state from any other state, doing so step-by-step, using the fact that $P_{i,i+1} = p,\ P_{i,i-1} = 1-p$. For example $-4 \rightarrow 2$ since $P^6_{-4,2} \geq p^6 > 0$, and $2 \rightarrow -4$ since $P^6_{2,-4} \geq (1-p)^6 > 0$; thus $-4 \longleftrightarrow 2$. In general $P^n_{i,j} > 0$ for $n = |i-j|$.

2. *Random walk from the gambler's ruin problem is not irreducible.* Here, the random walk is restricted to the finite state space $\{0, 1, \ldots, N\}$ and $P_{00} = P_{NN} = 1$. $C_1 = \{0\},\ C_2 = \{1, \ldots N-1\},\ C_3 = \{N\}$ are the communication classes.

3. Consider a Markov chain with $\mathcal{S} = \{0,1,2,3\}$ and transition matrix given by

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1/3 & 1/6 & 1/6 & 1/3 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

   Notice how states $0, 1$ keep to themselves in that whereas they communicate with each other, no other state is reachable from them (together they form an absorbing set). Thus $C_1 = \{0, 1\}$. Whereas every state is reachable from state 2, getting to state 2 is not possible from any other state; thus $C_2 = \{2\}$. Finally, state 3 is absorbing and so $C_3 = \{3\}$. This example illustrates the general method of deducing communication classes by analyzing the the transition matrix.

2 Recurrence and Stationary distributions

2.1 Recurrence and transience

Let τ_{ii} denote the *return time* to state i given $X_0 = i$:

$$\tau_{ii} = \min\{n \geq 1 : X_n = i | X_0 = i\},\ \tau_{ii} \stackrel{\text{def}}{=} \infty,\ \text{if } X_n \neq i,\ n \geq 1.$$

It represents the amount of time (number of steps) until the chain returns to state i given that it started in state i. Note how "never returning" is allowed in the definition by defining $\tau_{ii} = \infty$, so a return occurs if and only if $\tau_{ii} < \infty$.

$f_i \stackrel{\text{def}}{=} P(\tau_{ii} < \infty)$ is thus the probability of ever returning to state i given that the chain started in state i. A state i is called *recurrent* if $f_i = 1$; *transient* if $f_i < 1$. By the (strong) Markov property, once the chain revisits state i, the future is independent of the past, and it is as if the chain is starting all over again in state i for the first time: Each time state i is visited, it will be revisited with the same probability f_i independent of the past. In particular, if $f_i = 1$, then the chain will return to state i over and over again, an infinite number of times. That is why the word recurrent is used. If state i is transient ($f_i < 1$), then it will only be visited a finite (random) number of times (after which only the remaining states $j \neq i$ can be visited by the chain). Counting over all time, the total number of visits to state i, given that $X_0 = i$, is given by an infinite sequence of indicator rvs

$$N_i = \sum_{n=0}^{\infty} I\{X_n = i | X_0 = i\} \tag{1}$$

and has a geometric distribution,

$$P(N_i = n) = f_i^{n-1}(1 - f_i), \ n \geq 1.$$

(We count the initial visit $X_0 = i$ as the first visit.)

The expected number of visits is thus given by $E(N_i) = 1/(1 - f_i)$ and we conclude that

A state i is recurrent ($f_i = 1$) if and only if $E(N_i) = \infty$,

or equivalently

A state i is transient ($f_i < 1$) if and only if $E(N_i) < \infty$.

Taking expected value in (1) yields

$$E(N_i) = \sum_{n=0}^{\infty} P_{i,i}^n,$$

because $E(I\{X_n = i | X_0 = i\}) = P(X_n = i | X_0 = i) = P_{i,i}^n$.

We thus obtain

Proposition 2.1 *A state i is recurrent if and only if*

$$\sum_{n=0}^{\infty} P_{i,i}^n = \infty,$$

transient otherwise.

Proposition 2.2 *For any communication class C, all states in C are either recurrent or all states in C are transient. Thus: if i and j communicate and i is recurrent, then so is j. Equivalenly if i and j communicate and i is transient, then so is j. In particular, for an irreducible Markov chain, either all states are recurrent or all states are transient.*

Proof : Suppose two states $i \neq j$ communicate; choose an appropriate n so that $p = P_{i,j}^n > 0$. Now if i is recurrent, then so must be j because every time i is visited there is this same positive probability p ("success" probability) that j will be visited n steps later. But i being recurrent means it will be visited over and over again, an infinite number of times, so viewing this as sequence of Bernoulli trials, we conclude that eventually there will be a success. (Formally, we are using the *Borel-Cantelli theorem*.) ■

Definition 2.1 *For an irreducible Markov chain, if all states are recurrent, then we say that the Markov chain is recurrent; transient otherwise.*

The rat in the closed maze yields a recurrent Markov chain. The rat in the open maze yields a Markov chain that is not irreducible; there are two communication classes $C_1 = \{1, 2, 3, 4\}$, $C_2 = \{0\}$. C_1 is transient, whereas C_2 is recurrent.

Clearly if the state space is finite for a given Markov chain, then not all the states can be transient (for otherwise after a finite number a steps (time) the chain would leave every state never to return; where would it go?).

Thus we conclude that

Proposition 2.3 *An irreducible Markov chain with a finite state space is always recurrent: all states are recurrent.*

Finally observe (from the argument that if two states communicate and one is recurrent then so is the other) that for an irreducible recurrent chain, even if we start in some other state $X_0 \neq i$, the chain will still visit state i an infinite number of times: *For an irreducible recurrent Markov chain, each state j will be visited over and over again (an infinite number of times) regardless of the initial state $X_0 = i$.*

For example, if the rat in the closed maze starts off in cell 3, it will still return over and over again to cell 1.

2.2 Expected return time to a given state: positive recurrence and null recurrence

A recurrent state j is called *positive recurrent* if the *expected* amount of time to return to state j given that the chain started in state j has finite first moment:

$$E(\tau_{jj}) < \infty.$$

A recurrent state j for which $E(\tau_{jj}) = \infty$ is called *null recurrent.*

In general even for $i \neq j$, we define $\tau_{ij} \stackrel{\text{def}}{=} \min\{n \geq 1 : X_n = j \mid X_0 = i\}$, the time (after time 0) until reaching state j given $X_0 = i$.

Proposition 2.4 *Suppose $i \neq j$ are both recurrent. If i and j communicate and if j is positive recurrent ($E(\tau_{jj}) < \infty$), then i is positive recurrent ($E(\tau_{ii}) < \infty$) and also $E(\tau_{ij}) < \infty$. In particular, all states in a recurrent communication class are either all together positive recurrent or all together null recurrent.*

Proof : Assume that $E(\tau_{jj}) < \infty$ and that i and j communicate. Choose the smallest $n \geq 1$ such that $P^n_{j,i} > 0$. With $X_0 = j$, let $A = \{X_k \neq j,\ 1 \leq k \leq n,\ X_n = i\}$; $P(A) > 0$. Then $E(\tau_{j,j}) \geq E(\tau_{j,j}|A)P(A) = (n + E(\tau_{i,j}))P(A)$; hence $E(\tau_{i,j}) < \infty$ (for otherwise $E(\tau_{j,j}) = \infty$, a contradiction).

With $X_0 = j$, let $\{Y_m : m \geq 1\}$ be iid distributed as $\tau_{j,j}$ denote the interarrival times between visits to state j. Thus the n^{th} revisit of the chain to state j is at time $t_n = Y_1 + \cdots + Y_n$, and $E(Y) = E(\tau_{j,j}) < \infty$. Let

$$p = P(\{X_n\} \text{ visits state } i \text{ before returning to state } j \mid X_0 = j).$$

$p \geq P(A) > 0$, where A is defined above. Every time the chain revisits state j, there is, independent of the past, this probability p that the chain will visit state i before revisiting state j again. Letting N denote the number of revisits the chain makes to state j until first visiting state i, we thus see that N has a geometric distribution with "success" probability p, and so $E(N) < \infty$. N is a stopping time with respect to the $\{Y_m\}$, and

$$\tau_{j,i} \leq \sum_{m=1}^{N} Y_m,$$

and so by Wald's equation $E(\tau_{j,i}) \leq E(N)E(Y) < \infty$.

Finally, $E(\tau_{i,i}) \leq E(\tau_{i,j}) + E(\tau_{j,i}) < \infty$.

■

Proposition 2.2 together with Proposition 2.4 immediately yield

Proposition 2.5 *All states in a communication class C are all together either positive recurrent, null recurrent or transient. In particular, for an irreducible Markov chain, all states together must be positive recurrent, null recurrent or transient.*

Definition 2.2 *If all states in an irreducible Markov chain are positive recurrent, then we say that the Markov chain is positive recurrent. If all states in an irreducible Markov chain are null recurrent, then we say that the Markov chain is null recurrent. If all states in an irreducible Markov chain are transient, then we say that the Markov chain is transient.*

2.3 Limiting stationary distribution

When the limits exist, let π_j denote the *long run proportion of time that the chain spends in state j:*

$$\pi_j = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} I\{X_m = j\} \text{ w.p.1.} \tag{2}$$

Taking into account the initial condition $X_0 = i$, this is more precisely stated as:

$$\pi_j = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} I\{X_m = j \mid X_0 = i\} \text{ w.p.1, for all initial states } i. \tag{3}$$

Taking expected values ($E(I\{X_m = j \mid X_0 = i\}) = P(X_m = j \mid X_0 = i) = P_{ij}^m$) we see that if π_j exists then it can be computed alternatively by (via the bounded convergence theorem)

$$\begin{aligned}
\pi_j &= \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P(X_m = j \mid X_0 = i), \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P_{ij}^m, \text{ for all initial states } i. \qquad (4)
\end{aligned}$$

For simplicity of notation, we assume for now that the state space $\mathcal{S} = \mathbb{N} = \{0, 1, 2, \ldots\}$ or some finite subset of $\mathbb{N}$.

Definition 2.3 *If for each $j \in \mathcal{S}$, π_j exists as defined in (3) and is independent of the initial state i, and $\sum_{j\in\mathcal{S}} \pi_j = 1$, then the probability distribution $\pi = (\pi_0, \pi_1, \ldots)$ on the state space $\mathcal{S}$ is called the limiting or stationary or steady-state distribution of the Markov chain.*

Recalling that P^m_{ij} is precisely the $(ij)^{th}$ component of the matrix P^m, we conclude that (4) can be expressed in matrix form by

$$\lim_{n\to\infty} \frac{1}{n}\sum_{m=1}^{n} P^m = \begin{pmatrix} \pi \\ \pi \\ \vdots \end{pmatrix} = \begin{pmatrix} \pi_0, \pi_1, \ldots \\ \pi_0, \pi_1, \ldots \\ \vdots \end{pmatrix} \tag{5}$$

That is, when we average the m-step transition matrices, each row converges to the vector of stationary probabilities $\pi = (\pi_0, \pi_1, \ldots)$. The i^{th} row refers to the intial condition $X_0 = i$ in (4), and for each such fixed row i, the j^{th} element of the averages converges to π_j.

A nice way of interpreting π: If you observe the state of the Markov chain at some random time way out in the future, then π_j is the probability that the state is j.

To see this: Let N (our random observation time) have a uniform distribution over the integers $\{1, 2, \ldots n\}$, and be independent of the chain; $P(N = m) = 1/n,\ m \in \{1, 2, \ldots n\}$. Now assume that $X_0 = i$ and that n is very large. Then by conditioning on $N = m$ we obtain

$$\begin{aligned} P(X_N = j) &= \sum_{m=1}^{n} P(X_m = j | X_0 = i) P(N = m) \\ &= \frac{1}{n}\sum_{m=1}^{n} P^m_{i,j} \\ &\approx \pi_j, \end{aligned}$$

where we used (4) for the last line.

2.4 Connection between $E(\tau_{jj})$ and π_j

The following is intuitive and very useful:

Proposition 2.6 *If $\{X_n\}$ is a positive recurrent Markov chain, then a unique stationary distribution π exists and is given by*

$$\pi_j = \frac{1}{E(\tau_{jj})} > 0, \quad \textit{for all states } j \in \mathcal{S}.$$

If the chain is null recurrent or transient then the limits in (2) are all 0 wp1; no stationary distribution exists.

The intuition: On average, the chain visits state j once every $E(\tau_{jj})$ amount of time; thus $\pi_j = \frac{1}{E(\tau_{jj})}$.

Proof : First, we immediately obtain the transient case result since by definition, each fixed state i is then only visited a finite number of times; hence the limit in (2) must be 0 wp1.

Thus we need only consider now the two recurrent cases. (We will use the fact that for any fixed state j, returns to state j constitute recurrent regeneration points for the chain; thus this result is a consequence of standard theory about regenerative processes; but even if the reader has not yet learned such theory, the proof will be understandable.)

First assume that $X_0 = j$. Let $t_0 = 0,\ t_1 = \tau_{jj},\ t_2 = \min\{k > t_1 : X_k = j\}$ and in general $t_{n+1} = \min\{k > t_n : X_k = j\},\ n \geq 1$. These t_n are the consecutive times at which the chain visits state j. If we let $Y_n = t_n - t_{n-1}$ (the interevent times) then we revisit state j for the n^{th} time at time $t_n = Y_1 + \cdots + Y_n$. The idea here is to break up the evolution of the Markov chain into iid *cycles* where a cycle begins every time the chain visits state j. Y_n is the n^{th} *cycle-length*. By the Markov property, the chain starts over again and is independent of the past everytime it enters state j (formally this follows by the *Strong Markov Property*). This means that the cycle lengths $\{Y_n : n \geq 1\}$ form an iid sequence with common distribution the same as the first cycle length τ_{jj}. In particular, $E(Y_n) = E(\tau_{jj})$ for all $n \geq 1$.

Now observe that the number of revisits to state j is precisely n visits at time $t_n = Y_1 + \cdots + Y_n$, and thus the long-run proportion of visits to state j per unit time can be computed as

$$\begin{aligned}\pi_j &= \lim_{m\to\infty} \frac{1}{m}\sum_{k=1}^{m} I\{X_k = j\} \\ &= \lim_{n\to\infty} \frac{n}{\sum_{i=1}^{n} Y_i} \\ &= \frac{1}{E(\tau_{jj})}, \text{ w.p.1,}\end{aligned}$$

where the last equality follows from the Strong Law of Large Numbers (SLLN). Thus in the positive recurrent case, $\pi_j > 0$ for all $j \in \mathcal{S}$, where as in the null recurrent case, $\pi_j = 0$ for all $j \in \mathcal{S}$. Finally, if $X_0 = i \neq j$, then we can first wait until the chain enters state j (which it will eventually, by recurrence), and then proceed with the above proof. (Uniqueness follows by the unique representation $\pi_j = \frac{1}{E(\tau_{jj})}$.)

■

The above result is useful for computing $E(\tau_{jj})$ if π has already been found: For example, consider the rat in the closed off maze problem from HMWK 2. Given that the rat starts off in cell 1, what is the expected number of steps until the rat returns to cell 1? The answer is simply $1/\pi_1$. But how do we compute π? We consider that problem next.

2.5 Computing π algebraically

Theorem 2.1 *Suppose $\{X_n\}$ is an irreducible Markov chain with transition matrix P. Then $\{X_n\}$ is positive recurrent if and only if there exists a (non-negative, summing to 1) solution, $\pi = (\pi_0, \pi_1, \ldots)$, to the set of linear equations $\pi = \pi P$, in which case π is precisely the unique stationary distribution for the Markov chain.*

For example consider state space $\mathcal{S} = \{0, 1\}$ and the matrix

$$P = \begin{pmatrix} 0.5 & 0.5 \\ 0.4 & 0.6 \end{pmatrix},$$

which is clearly irreducible. For $\pi = (\pi_0, \pi_1)$, $\pi = \pi P$ yields the two equations

$$\begin{aligned}\pi_0 &= 0.5\pi_0 + 0.4\pi_1, \\ \pi_1 &= 0.5\pi_0 + 0.6\pi_1.\end{aligned}$$

We can also utilize the "probability" condition that $\pi_0 + \pi_1 = 1$. Solving yields $\pi_0 = 4/9$, $\pi_1 = 5/9$. We conclude that this chain is positive recurrent with stationary distribution $(4/9, 5/9)$. The long run proportion of time the chain visits state 0 is equal to 4/9 and the long run proportion of time the chain visits state 1 is equal to 5/9. Furthermore, since $\pi_j = 1/E(\tau_{jj})$, we conclude that the expected number of steps (time) until the chain revisits state 1 given that it is in state 1 now is equal to 9/5.

How to use Theorem 2.1

Theorem 2.1 can be used in practice as follows: if you have an irreducible MC, then you can try to solve the set of equations: $\pi = \pi P$, $\sum_{j \in \mathcal{S}} \pi_j = 1$. If you do solve them for some π, then this solution π is unique and is the stationary distribution, and the chain is positive recurrent. It might not even be necessary to "solve" the equations to obtain π: suppose you have a candidate π for the stationary distribution (perhaps by guessing), then you need only plug in the guess to verify that it satisfies $\pi = \pi P$. If it does, then your guess π is the stationary distribution, and the chain is positive recurrent.[2]

Proof of Theorem 2.1

Proof : Assume the chain is irreducible and positive recurrent. Then we know from Proposition 2.6 that π exists (as defined in Equations (3), (4)), has representation $\pi_j = 1/E(\tau_{jj}) > 0$, $j \in \mathcal{S}$, and is unique.

On the one hand, if we multiply (on the right) each side of Equation (5) by P, then we obtain

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P^{m+1} = \begin{pmatrix} \pi \\ \pi \\ \vdots \end{pmatrix} P.$$

But on the other hand,

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P^{m+1} &= \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P^{m} + \lim_{n\to\infty} \frac{1}{n}(P^{n+1} - P) \\ &= \begin{pmatrix} \pi \\ \pi \\ \vdots \end{pmatrix} + \lim_{n\to\infty} \frac{1}{n}(P^{n+1} - P) \quad \text{(from (5))} \\ &= \begin{pmatrix} \pi \\ \pi \\ \vdots \end{pmatrix}, \end{aligned}$$

because for any $k \geq 1$, $\lim_{n\to\infty} \frac{1}{n} P^k = 0$ (since $p_{ij}^k \leq 1$ for all i, j).

Thus, we obtain

$$\begin{pmatrix} \pi \\ \pi \\ \vdots \end{pmatrix} = \begin{pmatrix} \pi \\ \pi \\ \vdots \end{pmatrix} P,$$

yielding (from each row) $\pi = \pi P$.

Summarizing: If a Markov chain is positive recurrent, then the stationary distribution π exists as defined in Equations (3), (4), is given by $\pi_j = 1/E(\tau_{jj})$, $j \in \mathcal{S}$, and must satisfy $\pi = \pi P$. Now we prove the converse : For an irreducible Markov chain, suppose that $\pi = \pi P$ has a non-negative, summing to 1 solution π; that is, a probability solution. We must show that the chain is thus positive recurrent, and that this solution π is the stationary distribution as defined by (4). We first will prove that the chain cannot be either transient or null recurrent,

[2] $0 = (0, 0, \ldots, 0)$ is always a solution to $\pi = \pi P$, but is not a probability. Moreover, for any solution π, and any constant c, $c\pi = c\pi P$ by linearity, so $c\pi$ is also a solution.

hence must be positive recurrent from Proposition 2.5. To this end assume the chain is either transient or null recurrent. From Proposition 2.6, we know that then the limits in (4) are identically 0, that is, as $n \to \infty$,

$$\frac{1}{n}\sum_{m=1}^{n} P^m \to 0. \tag{6}$$

But if $\pi = \pi P$, then (by multiplying both right sides by P) $\pi = \pi P^2$ and more generally $\pi = \pi P^m,\ m \geq 1$, and so

$$\pi\Big[\frac{1}{n}\sum_{m=1}^{n} P^m\Big] = \frac{1}{n}\sum_{m=1}^{n} \pi P^m = \pi,\ n \geq 1,$$

yielding from (6) that $\pi 0 = \pi$, or $\pi = (0, \ldots, 0)$, contradicting that π is a probability distribution. Having ruled out the transient and null recurrent cases, we conclude from Proposition 2.5 that the chain must be positive recurrent. For notation, suppose $\pi' = \pi' P$ denotes the non-negative, summing to 1 solution, and that π denotes the stationary distribution in (4), given by $\pi_j = 1/E(\tau_{jj}),\ j \in \mathcal{S}$. We now show that $\pi' = \pi$. To this end, multiplying both sides of (5) (on the left) by π', we conclude that

$$\pi' = \pi' \begin{pmatrix} \pi \\ \pi \\ \vdots \end{pmatrix} = \pi' \begin{pmatrix} \pi_0, \pi_1, \cdots \\ \pi_0, \pi_1, \cdots \\ \vdots \end{pmatrix}.$$

Since $\sum_{i \in \mathcal{S}} \pi'_i = 1$, we see that the above yields $\pi'_j = \pi_j,\ j \in \mathcal{S}$; $\pi' = \pi$ as was to be shown. ■

2.6 Finite state space case

When the state space of a Markov chain is finite, then the theory is even simpler:

Theorem 2.2 *Every irreducible Markov chain with a finite state space is positive recurrent and thus has a stationary distribution (unique probability solution to $\pi = \pi P$).*

Proof : From Prop 2.3, we know that the chain is recurrent. We will now show that it can't be null recurrent, hence must be positive recurrent by Proposition 2.5. To this end, note that for any fixed $m \geq 1$, the rows of $P^m = P^{(m)}$ (m-step transition matrix) must sum to 1, that is,

$$\sum_{j \in \mathcal{S}} P^m_{i,j} = 1,\ i \in \mathcal{S}. \tag{7}$$

Moreover, if null recurrent, we know from Proposition 2.6, that for all $i \in \mathcal{S}$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{m=1}^{n} P^m_{i,j} = 0,\ j \in \mathcal{S}. \tag{8}$$

Summing up (8) over j then yields

$$\sum_{j \in \mathcal{S}} \lim_{n\to\infty} \frac{1}{n}\sum_{m=1}^{n} P^m_{i,j} = 0,\ i \in \mathcal{S}.$$

But since the state space is finite, we can interchange the outer finite sum with the limit,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j\in\mathcal{S}} \sum_{m=1}^{n} P_{i,j}^{m} = 0,\ i \in \mathcal{S}.$$

But then we can interchange the order of summation and use (7), to obtain a contradiction:

$$0 = \lim_{n\to\infty} \frac{1}{n} \sum_{j\in\mathcal{S}} \sum_{m=1}^{n} P_{i,j}^{m} = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \sum_{j\in\mathcal{S}} P_{i,j}^{m} = 1,\ i \in \mathcal{S}.$$

Thus the chain must be positive recurrent. ■

This is a very useful result. For example, it tells us that the rat in the maze Markov chain, when closed off from the outside, is positive recurrent, and we need only solve the equations $\pi = \pi P$ to compute the stationary distribution.

2.7 Stationarity of positive recurrent chains

The use of the word "stationary" here means "does not change with time", and we now proceed to show why that describes π.

For a given probability distribution $\nu = (\nu_j)$ on the state space $\mathcal{S}$, we use the notation $X \sim \nu$ to mean that X is a random variable with distribution ν: $P(X = j) = \nu_j,\ j \in \mathcal{S}$.

Given any such distribution $\nu = (\nu_j)$, and Markov chain with transition matrix P, note that if $X_0 \sim \nu$ then the vector νP is the distribution of X_1, that is, the j^{th} coordinate of the vector νP is $\sum_{i\in\mathcal{S}} P_{i,j}\nu_i$, which is precisely $P(X_1 = j)$: For each $j \in \mathcal{S}$,

$$\begin{aligned} P(X_1 = j) &= \sum_{i\in\mathcal{S}} P(X_1 = j | X_0 = i) P(X_0 = i) \\ &= \sum_{i\in\mathcal{S}} P_{i,j}\nu_i. \end{aligned}$$

In other words if the initial distribution of the chain is ν, then the distribution one time unit later is given by νP and so on:

Lemma 2.1 *If $X_0 \sim \nu$, then $X_1 \sim \nu P$, $X_2 \sim \nu P^2, \ldots, X_n \sim \nu P^n,\ n \geq 1$.*

Proposition 2.7 *For a positive recurrent Markov chain with stationary distribution π, if $X_0 \sim \pi$, then $X_n \sim \pi$ for all $n \geq 0$. In words: By starting off the chain initially with its stationary distribution, the chain remains having that same distribution for ever after. X_n has the same distribution for each n. This is what is meant by stationary, and why π is called the stationary distribution for the chain.*

Proof : From Theorem 2.1, we know that π satisfies $\pi = \pi P$, and hence by multiplying both sides (on the right) by P yields $\pi = \pi P^2$ and so on yielding $\pi = \pi P^n,\ n \geq 1$. Thus from Lemma 2.1, we conclude that if $X_0 \sim \pi$, then $X_n \sim \pi,\ n \geq 1$. ■

Definition 2.4 *A stochastic process $\{X_n\}$ is called a stationary process if for every $k \geq 0$, the process $\{X_{k+n} : n \geq 0\}$ has the same distribution, namely the same distribution as $\{X_n\}$. "same distribution" means in particular that all the finite dimensional distributions are the same: for any integer $l \geq 1$ and any integers $0 \leq n_1 < n_2 < \ldots < n_l$, the joint distribution of the $l-$ vector $(X_{k+n_1}, X_{k+n_2}, \ldots, X_{k+n_l})$ has the same distribution for each k, namely the same distribution as $(X_{n_1}, X_{n_2}, \ldots, X_{n_l})$.*

It is important to realize (via setting $l = 1$ in the above definition) that for a stationary process, it of course holds that X_n has the same distribution for each fixed n, namely that of X_0; but this is only the marginal distribution of the process. Stationarity means much more and in general is a much stronger condition. It means that if we change the time origin 0 to be any time k in the future, then the entire process still has the same distribution (e.g., same finite dimensional distributions) as it did with 0 as the origin. For Markov chains, however, stationarity is the same as for the marginals only:

Proposition 2.8 *For a positive recurrent Markov chain with stationary distribution π, if $X_0 \sim \pi$, then the chain is a stationary process.*

Proof : This result follows immediately from Proposition 2.7 because of the Markov property: Given X_k, the future is independent of the past; its entire distribution depends only on the distribution of X_k and the transition matrix P. Thus if X_k has the same distribution for each k, then it's future $\{X_{k+n} : n \geq 0\}$ has the same distribution for each k. ■

The above result is quite intuitive: We obtain π by choosing to observe the chain way out in the infinite future. After doing that, by moving k units of time further into the future does not change anything (we are already out in the infinite future, going any further is still infinite).

2.8 Convergence to π in the stronger sense $\pi_j = \lim_{n\to\infty} P(X_n = j)$; aperiodicity.

For a positive recurrent chain, the sense in which it was shown to converge to its stationary distribution π was in a *time average* or *Cesàro* sense (recall Equations (3), (4)). If one wishes to have the convergence without averaging, $\pi_j = \lim_{n\to\infty} P(X_n = j),\ j \in \mathcal{S}$ (regardless of initial conditions $X_0 = i$), then a further condition known as aperiodicity is needed on the chain; we will explore this in this section. Since $P(X_n = j | X_0 = i) = P^n_{i,j}$, we need to explore when it holds for each $j \in \mathcal{S}$ that $P^n_{i,j} \to \pi_j$ for all $i \in \mathcal{S}$.

For simplicity of notation in what follows we assume that $\mathcal{S} = \{0, 1, 2, \ldots\}$ or some subset. It is easily seen that for any probability distribution ν, the matrix

$$M = \begin{pmatrix} \pi \\ \pi \\ \vdots \end{pmatrix} = \begin{pmatrix} \pi_0, \pi_1, \ldots \\ \pi_0, \pi_1, \ldots \\ \vdots \end{pmatrix},$$

satisfies $\nu M = \pi$. Thus from Lemma 2.1, we see that in order to obtain $\pi_j = \lim_{n\to\infty} P(X_n = j),\ j \in \mathcal{S}$, regardless of initial conditions, we equivalently need to have $P^n \to M$ as $n \to \infty$. We already know that the averages converge, $\frac{1}{n}\sum_{m=1}^{n} P^n \to M$, and it is easily seen that in general the stronger convergence $P^n \to M$ will not hold. For example take $\mathcal{S} = \{0, 1\}$ with

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

yielding an alternating sequence: If $X_0 = 0$, then $\{X_n : n \geq 0\} = \{0, 1, 0, 1, \ldots\}$; if $X_0 = 1$, then $\{X_n : n \geq 0\} = \{1, 0, 1, 0, \ldots\}$. Clearly, if $X_0 = 0$, then $P(X_n = 0) = 1$, if n is even, but $P(X_n = 0) = 0$, if n is odd; $P(X_n = 0)$ does not converge as $n \to \infty$. In terms of P, this is seen by noting that for any $n \geq 1$,

$$P^{2n} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

whereas

$$P^{2n+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

So, although $\pi = (1/2, 1/2)$ and indeed $\frac{1}{n}\sum_{m=1}^{n} P^n \to M = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$, it does not hold that $P^n \to M$.

The extra condition needed is explained next.

For a state $j \in \mathcal{S}$, consider the set $\mathcal{Q} = \{n \geq 1 : P^n_{j,j} > 0\}$. If there exists an integer $d \geq 2$ such that $\mathcal{Q} = \{kd : k \geq 1\}$, then state j is said to be *periodic of period d*. This implies that $P^n_{j,j} = 0$ whenever n is not a multiple of d. If no such d exists, then the state j is said to be *aperiodic*. It can be shown that periodicity is a class property: all states in a communication class are either aperiodic, or all are periodic with the same period d. Thus an irreducible Markov chain is either *periodic* or *aperiodic*. In the above two-state counterexample, $d = 2$; the chain is periodic with period $d = 2$. Clearly, if a chain is periodic, then by choosing any subsequence of integers $n_m \to \infty$, as $m \to \infty$, such that $n_m \notin \mathcal{Q}$, we obtain for any state j that $P(X_{n_m} = j \mid X_0 = j) = P^{n_m}_{j,j} = 0$ and so convergence to $\pi_j > 0$ is not possible along this subsequence; hence convergence of $P(X_n = j)$ to π_j does not hold. The converse is also true, the following is stated without proof:

Proposition 2.9 *A positive recurrent Markov chain converges to π via $\pi_j = \lim_{n\to\infty} P(X_n = j)$, $j \in \mathcal{S}$, if and only if the chain is aperiodic.*

Remark 1 For historical reasons, in the literature a positive recurrent and aperiodic Markov chain is sometimes called an *ergodic* chain. The word ergodic, however, has a precise meaning in mathematics (ergodic theory) and this meaning has nothing to do with aperiodicity! In fact any positive recurrent Markov chain is ergodic in this precise mathematical sense. (So the historical use of ergodic in the context of aperiodic Markov chains is misleading and unfortunate.)

Remark 2 The type of convergence $\pi_j = \lim_{n\to\infty} P(X_n = j)$, $j \in \mathcal{S}$, is formally called *weak convergence*, and we then say that X_n converges weakly to π.