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:* $P^0_{ii} = 1 > 0$.
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 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.

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.

> *A Markov chain for which there is only one communication class is called an irreducible Markov chain.*

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 $-4 \to 2$ since $P^6_{-4,2} \geq p^6 > 0$, and $2 \to -4$ 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 "staring" at the transition matrix.

2 Positive Recurrence and Stationarity

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 until the chain returns to state i given that it started in state i. Note how never returning is allowed 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 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 $X_0 = i$, denoted by N_i, therefore has a geometric distribution with "success" probability f_i,

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

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

The expected number of visits is given by $E(N_i) = 1/(1 - f_i)$ and so 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 $E(N_i) < \infty$.

Since N_i has representation

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

taking expectations 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.

States that are recurrent/transient all lie in the same communication class, and thus for any given communication class C, all states $i \in C$ are recurrent or transient together. This is easily seen as follows: Suppose two state i and 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 a kind of Bernoulli trials experiment, we conclude that eventually there wil be a success. 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, an irreducible Markov chain must have all it states together be recurrent or all its states together be transient. If all states are recurrent 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 *an irreducible Markov chain with a finite state space is 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

A 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 positive recurrent state j is always recurrent: If $E(\tau_{jj}) < \infty$, then $f_j = P(\tau_{jj} < \infty) = 1$, but the converse is not true: *a recurrent state need not be positive recurrent.*

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

Positive recurrence is a communication class property: all states in a communication class are all together positive recurrent, null recurrent or transient. In particular, *for an irreducible Markov chain, all states together must be positive recurrent, null recurrent or transient.* If all states in an irreducible Markov chain are positive recurrent, then we say that the Markov chain is positive recurent. If all states in an irreducible Markov chain are null recurrent, then we say that the Markov chain is null recurent.

In general $\tau_{ij} \stackrel{\text{def}}{=} \min\{n \geq 1 : X_n = j \; |X_0 = i\}$, the time (after time 0) until reaching state j given $X_0 = i$, and it is easily seen and intuitive that if a chain is positive recurrent, then $E(\tau_{ij}) < \infty$ for $i \neq j$ also: the expected amount of time to reach state j given that the chain started in some other state i has finite first moment.

2.3 Limiting or stationary distribution

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

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

Including the initial condition $X_0 = i$, this is more precisely stated as:

$$\pi_j = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} I\{X_m = j \mid X_0 = i\} \text{ w.p.1, for all initial states } i. \tag{2}$$

Taking expected values (recall that $E(I\{X_m = j\}) = P(X_m = j)$) yields

$$\begin{aligned} \pi_j &= \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P(X_m = j \mid X_0 = i), \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P^m_{ij}, \text{ for all initial states } i. \end{aligned} \tag{3}$$

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

If for each $j \in \mathcal{S}$, π_j exists with $\pi_j > 0$, then $\pi = (\pi_0, \pi_1, \ldots)$ forms a probability distribution on the state space $\mathcal{S}$, and 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 (P multiplied by itself m times), we conclude that in matrix form (3) is expressed by

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P^m = \begin{pmatrix} \pi \\ \pi \\ \vdots \end{pmatrix} = \begin{pmatrix} \pi_0, \pi_1, \ldots \\ \pi_0, \pi_1, \ldots \\ \vdots \end{pmatrix} \tag{4}$$

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 (3), 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 time) have a uniform distribution over the integers $\{1, 2, \ldots n\}$ where n is large; $P(N = m) = 1/n,\ m \in \{1, 2, \ldots n\}$. N is taken independent of the chain. Now assume that $X_0 = i$. Then by conditioning on $N = m$ we obtain

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

where we used (3) for the last line.

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

The following is intuitive and very useful:

Proposition 2.2 *If $\{X_n\}$ is a positive recurrent Markov chain, then the stationary distribution π exists and is given by*

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

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

Proof :

First assume that $X_0 = j$. If we let Y_n denote the amount of time spent between the $(n-1)^{th}$ and n^{th} visit to state j, then we visit state j for the n^{th} time at time $Y_1 + \cdots + Y_n$, where $Y_1 = \tau_{jj}$. The idea here is to break up the evolution of the Markov chain into *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 i.i.d 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 visits to state j is precisely n visits at time $Y_1 + \cdots + Y_n$, and thus the long run proportion of visits to state j per unit time can be computed as

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

where the last equality follows from the Strong Law of Large Numbers (SLLN). $\pi_j > 0$ because $E(\tau_{jj}) < \infty$ by definition of positive recurrence.

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.

$\blacksquare$

The above result is useful for computing $E(\tau_{jj})$ when 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: Main Theorem

Given a $n-$vector $y = (y_1, y_2, \ldots, y_n)$ and a $n \times n$ matrix A, yA denotes mutiplication of y with A yielding a $n-$ vector. (y is dotted with each column of A, and $n = \infty$ is allowed.) Example: $n = 4$, $y = (4, 6, 8, 2)$

$$A = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 \end{pmatrix}$$

$$yA = (6(1/2) + 8(1/2), 4(1/2) + 2(1/2), 4(1/2) + 2(1/2), 6(1/2) + 8(1/2)) = (7, 3, 3, 7).$$

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 this case, π is precisely the stationary distribution for the Markov chain and is unique.*

For example consider the matrix

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

which is clearly irreducible. View it as corresponding to the Markov chain for weather on each day (0 for rain, 1 for no rain). For $\pi = (\pi_0, \pi_1)$, the equations $\pi = \pi P$ yields

$$\pi_0 = 0.5\pi_0 + 0.4\pi_1, \ \pi_1 = 0.5\pi_0 + 0.6\pi_1.$$

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 days that it rains is $4/9$; the long run proportion of days that it does not rain is $5/9$. Furthermore, since $\pi_j = 1/E(\tau_{jj})$, we conclude that the expected number of days until it rains again given that it rained today is $9/4$.

The above theorem is important because if you have an irreducible MC, then: on the one hand you can try to solve the set of equations: $\pi = \pi P$ and $\sum_n \pi_n = 1$. 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.

On the other hand, if before solving, you have a candidate π for the stationary distribution (perhaps by guessing), then you need only plug in the guess and verify that it satisfies $\pi = \pi P$. If it does, then your guess π IS the the stationary distribution, and the chain IS positive recurrent.

Proof of Theorem 2.1

Here we prove one side of Theorem 2.1: Assume the chain is irreducible and positive recurrent. Then we know that π exists (as defined in Equations (1), (3)) and can even be given by $\pi_j = 1/E(\tau_{jj})$.

On the one hand, if we multiply (on the right) each side of Equation (4) 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,

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P^{m+1} &= \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} P^{m} + \lim_{n\to\infty} \frac{1}{n} P^{n+1} \\ &= \begin{pmatrix} \pi \\ \pi \\ \vdots \end{pmatrix} + \lim_{n\to\infty} \frac{1}{n} P^{n+1} \quad \text{(from (4))} \\ &= \begin{pmatrix} \pi \\ \pi \\ \vdots \end{pmatrix}, \end{aligned}$$

because $\lim_{n\to\infty} \frac{1}{n} P^{n+1} = 0$ (since $p_{ij}^{n+1} \leq 1$ 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 (non-negative, summing to 1) exists as defined in Equations (1), (3), is given by $\pi_j = 1/E(\tau_{jj})$, $j \in \mathcal{S}$, and must satisfy $\pi = \pi P$. The converse also turns out to be true for an irreducible Markov chain (proof omitted): For an irreducible Markov chain, if $\pi = \pi P$ has a non-negative, summing to 1 solution π, then the chain is positive recurrent with $\pi_j = 1/E(\tau_{jj})$, $j \in \mathcal{S}$. (Uniqness of π follows since $\pi_j = 1/E(\tau_{jj})$.)

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 in fact positive recurrent and thus has a stationary distribution (unique probability solution to $\pi = \pi P$).*

Finite state space means, for example, that $\mathcal{S} = \{0, 1, 2 \dots b\}$ for some number b. 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 word "stationary" means "does not change with time", and we now proceed to show why that word is used to describe π.

For a given discrete probability distribution $\nu = (\nu_0, \nu_1, \ldots,)$, we use the notation $X \sim \nu$ to mean that X is a random variable with distribution ν: $P(X = j) = \nu_j,\ j \geq 0$.

Proposition 2.3 *For a positive recurrent Markov chain with stationary distribution π, if $X_0 \sim \pi$, that is, if $P(X_0 = j) = \pi_j,\ j \geq 0$, then in fact $X_n \sim \pi$ for all $n \geq 0$, that is, $P(X_n = j) = \pi_j,\ j \geq 0$.*

In other words: By starting off the chain initially with its stationary distribution, the chain remains having that distribution for ever after. This is what is meant by stationary, and why π is called the stationary distribution for the chain.

Proof : Suppose that $X_0 \sim \pi$. Then by conditioning on $\{X_0 = i\}$

$$\begin{aligned} P(X_1 = j) &= \sum_{i=0}^{\infty} P(X_1 = j | X_0 = i) P(X_0 = i) \\ &= \sum_{i=0}^{\infty} P_{i,j} \pi_i \\ &= \pi_j, \end{aligned}$$

where the last equality follows from the fact that π must satisfy $\pi = \pi P$ via Theorem 2.1. Thus $X_1 \sim \pi$, so that $P(X_2 = j | X_1 = i) P(X_1 = i) = P_{i,j} \pi_i$, and we see that the above conditioning argument now yields $X_2 \sim \pi$, and in general (by induction) $X_n \sim \pi,\ n \geq 0$. ∎