Markov chains with countable state spaces

S-VALUED RANDOM VARIABLES

Let S be a countable set. An S-valued r.v on a probability space (Ω, P) is a function $X : \Omega \to S$. If $S \subseteq \mathbb{R}$, this is just an ordinary r.v. If $S \subseteq \mathbb{R}^d$, then X is what we called a random vector. One can talk of the distribution or p.m.f of an S-valued r.v (as S is countable, these random variables are discrete) defined as $p : S \to [0,1]$ where $p_x = \mathbf{P}\{\omega : X(\omega) = x\}$. If $X_1, \ldots, X_n$ are S-valued r.v.s on a common probability space, their joint p.m.f may be defined as the function $p : S^n \to [0,1]$ with $p(x_1, \ldots, x_n) = \mathbf{P}\{X_1 = x_1, \ldots, X_n = x_n\}$. Just as for real-valued random variables, one can specify the joint distribution of $X_1, \ldots, X_n$ in two ways.

(1) Specify the joint p.m.f $p_X(x_1, \ldots, x_n) = \mathbf{P}\{X_1 = x_1, \ldots, X_n = x_n\}$.

(2) Specify the successive conditional p.m.f.s $p_{X_1}, p_{X_2|X_1}, \ldots, p_{X_n|X_1,\ldots,X_{n-1}}$ which are defined as

$$p_{X_k|X_1=x_1,\ldots,X_{k-1}=x_{k-1}}(y) = \frac{\mathbf{P}\{X_1 = x_1, \ldots, X_{k-1} = x_{k-1}, X_k = y\}}{\mathbf{P}\{X_1 = x_1, \ldots, X_{k-1} = x_{k-1}\}}.$$

It is clear how to obtain the conditional p.m.f.s from the joint and vice versa.

Example 97. Pick a person at random from a given population. Here S is the given population and the random person picked is an S-valued random variable.

Caution: One cannot take expectations of an S-valued random variable! However, if $f : S \to \mathbb{R}$, then $f(X)$ is a real-valued random variable and we can talk of its expectation. In the above example, it makes no sense to talk of an expected person, but we can talk of the expected height of a random person (height is a function from S to $\mathbb{R}$).

MARKOV CHAINS

Let S be a countable set. Let $X_0, X_1, X_2, \cdots$ be a sequence of random variables on a common probability space. We say that X is a *Markov chain* if the conditional distribution of X_n given $X_0, \ldots, X_{n-1}$ depends only on X_{n-1}. Mathematically, this means that

$$\mathbf{P}\left\{X_n = j \,\middle|\, X_0 = i_0, \ldots, X_{n-2} = i_{n-2}, X_{n-1} = i\right\} = p_n(i,j)$$

for some $p_n(i,j)$ (as the notation says, this depends on n and i and j but not on $i_0, \ldots, i_{n-2}$.

We say that the MC is *time-homogeneous* if $p_n(i,j)$ does not depend on n. This means that

$$\mathbf{P}\left\{X_n = j \,\middle|\, X_0 = i_0, \ldots, X_{n-2} = i_{n-2}, X_{n-1} = i\right\} = p(i,j) = p_{i,j}$$

for all $n \geq 1$, and for all $i, j, i_0, \ldots, i_{n-2}$.

Henceforth, Markov chain will mean a time-homogeneous Markov chain, unless otherwise stated. S is called the *state space* of the chain and $p_{i,j}$ are called the *transition probabilities*. The matrix $P = (p_{i,j})_{i,j \in S}$ is called the *transition matrix*. The distribution of X_0 is called the *initial distribution*.

EXAMPLES OF MARKOV CHAINS

Example 98. (SRW on $\mathbb{Z}$): We defined this as the sequence of r.v X_n with $X_0 = 0$ and $X_n = \xi_1 + \ldots + \xi_n$ where ξ_i are i.i.d. $\text{Ber}_{\pm}(1/2)$ random variables. Clearly, X_n is $\mathbb{Z}$-valued. Further, as $X_k = X_{k-1} + \xi_k$,

$$\begin{aligned}
\mathbf{P}\{X_k = j \,|\, X_0 = i_0, \ldots, X_{k-2} = i_{k-2}, X_{k-1} = i\} &= \mathbf{P}\{\xi_k = j - i \,|\, \xi_1 = i_1 - i_0, \ldots, \xi_{k-1} = i - i_{k-1}\} \\
&= \mathbf{P}\{\xi_k = j - i\} \quad \text{(as } \xi_i \text{ are independent)} \\
&= \begin{cases} 1/2 & \text{if } j = i \pm 1. \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

Thus, X is a MC with state space $\mathbb{Z}$ and transitions $p_{i,i+1} = p_{i,i-1} = 1/2$ and $p_{i,j} = 0$ for all other i, j.

Exercise 99. Write down the state space and transitions for symmetric SRW on $\mathbb{Z}^d$.

Example 100. (Ehrenfest chain): Fix $N \geq 1$. Informally, there is a container with two equal parts separated by a wall in which there is a small aperture. Initially a certain amount of gas is on either side of the partition. By experience we know that the amount of gas in the two sides will equalize eventually. Here is the statistical model of Ehrenfest: At each instant of time (we discretize time into instants $0, 1, 2, \ldots$), one molecule of gas is randomly chosen and transferred to the other side of the partition.

In mathematical language, we are defining a MC with $S = \{0, 1, \dots, N\}$ and transitions

$$p_{i,j} = \begin{cases} \frac{N-i}{N} & \text{if } j = i+1. \\ \frac{i}{N} & \text{if } j = i-1. \\ 0 & \text{otherwise.} \end{cases}$$

Example 101. (SRW on a graph). Let $G = (V, E)$ be a graph with a countable vertex set V and edge set E. For each $v \in V$, its degree d_v is the number of neighbours it has in G. We assume $d_v < \infty$ for all $v \in V$. Then, we may a define a MC with $S = V$ and $p_{v,u} = \frac{1}{d_v}$ if $u \sim v$ and $p_{v,u} = 0$ otherwise. This is called *SRW on G*.

There are variants of this. We can fix a parameter $\tau \in [0, 1]$ and define the transition probabilities $p_{v,v} = 1 - \tau$, $p_{v,u} = \frac{\tau}{d_v}$ if $u \sim v$ and $p_{v,u} = 0$ otherwise. This is called a *lazy SRW on G with laziness parameter* $1 - \tau$.

Many special cases:

(1) SRW on $\mathbb{Z}^d$ is obviously a special case of this. Here $V = \mathbb{Z}^d$ with $\mathbf{x} \sim \mathbf{y}$ if and only if $\sum_{i=1}^{d} |x_i - y_i| = 1$.
(2) Coupon collector problem: Here we have N coupons labelled $1, 2, \dots, N$ from which we pick with replacement. This is lazy SRW on the complete graph K_N with laziness parameter $1/N$.
(3) SRW on the hypercube: Let $V = \{0, 1\}^N$ with $\mathbf{x} \sim \mathbf{y}$ if and only if $\sum_{i=1}^{d} |x_i - y_i| = 1$. This graph is called the hypercube. One can consider SRW on it, which is equivalent to picking a co-ordinate at random and flipping it (if 0 change to 1 and vice versa). If we define $Y_n =$ no. of ones in X_n, then Y_n is a MC which is just the Ehrenfest chain!

Example 102. (Queues). Imagine a queue where one person is served every instant. Also, at each instant a random number of people may join the queue. Let the number who join the queue at the n^{th} instant be ξ_n. We assume ξ_i are i.i.d random variables with $\mathbf{P}(\xi_1 = k) = \alpha_k$ for $k = 0, 1, 2 \dots$.

Let X_0 have some distribution. For $n \geq 1$, define

$$X_n = \begin{cases} X_{n-1} - 1 + \xi_n & \text{if } X_{n-1} \geq 1. \\ X_{n-1} + \xi_n & \text{if } X_{n-1} = 0. \end{cases}$$

X_n has the interpretation as the number of people in the queue at time n.

Show that X is a MC with $S = \mathbb{N} = \{0, 1, 2, \dots\}$ and transition probabilities

$$p_{i,j} = \begin{cases} \alpha_{j-i+1} & \text{if } i \geq 1. \\ \alpha_j & \text{if } i = 0. \end{cases}$$

Example 103. (Card shuffling). There are many algorithms for shuffling a deck of cards, and each of it is a Markov chain with state space $\mathcal{S}_n$ (the set of all permutations of $1, 2, \dots, n$ where n is the number of cards in the deck). The transition probabilities are different for different algorithms. We give just two shuffling algorithms.

(1) Random transposition: *Pick two distinct positions at random and interchange the cards in those locations.* In this case,

$$p_{\pi,\sigma} = \begin{cases} \frac{1}{\binom{n}{2}} & \text{if } \sigma = (i, j) \circ \pi \text{ for some } i < j. \\ 0 & \text{otherwise.} \end{cases}$$

Here (i, j) is the transposition, which is the permutation that interchanges i and j.
(2) Top to random shuffle: *Pick the top card and place it in a random position.* The transitions are

$$p_{\pi,\sigma} = \begin{cases} \frac{1}{n} & \text{if } \sigma = (1, i) \circ \pi \text{ for some } i. \\ 0 & \text{otherwise.} \end{cases}$$

Example 104. (Branching process). Let $L_{n,i}$, $n \geq 0$, $i \geq 1$, be i.i.d non-negative integer valued random variables. Let $Z_0 = 1$ and for $n \geq 1$ let

$$Z_n = \begin{cases} L_{n,1} + \ldots + L_{n,Z_{n-1}} & \text{if } Z_{n-1} \geq 1. \\ 0 & \text{if } Z_{n-1} = 0. \end{cases}$$

Check that $Z_0, Z_1, \dots$ is a MC on $\{0, 1, 2 \dots\}$.

Example 105. (Two state MC). Let $S = \{0, 1\}$ and let $p_{0,1} = \alpha = 1 - p_{0,0}$ and $p_{1,0} = \beta = 1 - p_{1,1}$. here $\alpha, \beta \in [0, 1]$ are fixed. This is a useful example because we can make explicit calculations very easily.

Multi-step transition probabilities

By the considerations of the first section, if we know the transition probabilties of a Markov chain, together with the distribution of X_0, we can calculate the joint distribution of $X_0, X_1, X_2, \ldots$. Indeed,

$$\begin{aligned}\mathbf{P}\{X_0 = i_0, \ldots, X_n = i_n\} &= \mathbf{P}\{X_0 = i_0\}\prod_{k=1}^{n}\mathbf{P}\{X_k = i_k \,\big|\, X_0 = i_0, \ldots, X_{k-1} = i_{k-1}\} \\ &= \mathbf{P}\{X_0 = i_0\} p_{i_0,i_1} p_{i_1,i_2} \cdots p_{i_{n-1},i_n}.\end{aligned}$$

Thus the initial distribution and the transition probabilities determine the distribution of the Markov chain.

Let $p^{(k)}_{i,j} = \mathbf{P}(X_k = j \,\big|\, X_0 = i)$. Then $p^{(1)}_{i,j} = p_{i,j}$. Can we calculate $p^{(2)}$? Observe that

$$\begin{aligned}\mathbf{P}\{X_{n+2} = j \,\big|\, X_n = i\} &= \sum_{\ell\in S}\mathbf{P}\{X_{n+1} = \ell, X_{n+2} = j \,\big|\, X_n = i\} \\ &= \sum_{\ell\in S}\mathbf{P}\{X_{n+1} = \ell, X_{n+2} = j \,\big|\, X_n = i\} \\ &= \sum_{\ell\in S} p_{i,\ell} p_{\ell,j}.\end{aligned}$$

In particular, taking $n = 0$ we get $p^{(2)}_{i,j} = \sum_{\ell\in S} p_{i,\ell} p_{\ell,j}$. Similarly, check that

$$p^{(k)}_{i,j} = \sum_{\ell_1,\ldots,\ell_{k-1}\in S} p_{i,\ell_1} p_{\ell_1,\ell_2} \cdots p_{\ell_{k-2},\ell_{k-1}} p_{\ell_{k-1},j}.$$

Thus, in terms of matrices, we can simply write $P^{(k)} = P^k$ (as product of matrices!).

An additional useful observation which follows from the above calculation is that $X_0, X_k, X_{2k}, \ldots$ is a MC with the same state space S and transition matrix P^k.

An illustration of how to work with Markov chains

Let us illustrate the useful and important technique of working with Markov chains - "condition on the first step"!

Gambler's ruin problem: Let a, b be positive integers and let $S = \{-a, -a+1, \ldots, b-1, b\}$. Let $X_0, X_1, \ldots$ be a Markov chain with state space S, and transition probabilities

$$p_{-a,-a} = 1, \qquad p_{b,b} = 1, \qquad p_{i,i+1} = p_{i,i-1} = \frac{1}{2} \text{ if } -a < i < b.$$

One can regard this as the fortunes of a gambler who starts with a units of money and his opponent has b units of money. In each game she wins 1 or loses 1 and his fortune goes up or down accordingly. Of course, the game stops when X_n hits $-a$ or b (either our gambler or his opponent has become bankrupt) and then the fortunes do not change anymore - hence the transitions are different at $-a$ and b.

Question 1: The main question is, what is the chance that our gambler goes bankrupt rather than her opponent?

Let $\tau = \min\{n : X_n = -a \text{ or } b\}$ be the time at which one of the two gamblers becomes bankrupt. Firstly, we claim that $\tau < \infty$ with probability 1 (why?). We want $\mathbf{P}\{X_\tau = -a\}$.

An equivalent problem is to fix $L = a+b$ and take $S = \{0, 1, \ldots, L\}$ and the transitions

$$p_{0,0} = 1, \qquad p_{L,L} = 1, \qquad p_{i,i+1} = p_{i,i-1} = \frac{1}{2} \text{ if } 0 < i < L.$$

More generally, let $f(i) = \mathbf{P}_i(X_\tau = 0) := \mathbf{P}(X_\tau = 0 \,\big|\, X_0 = i)$. We want $f(a)$.

Observe that $f(0) = 1$ and $f(L) = 0$. For $0 < i < L$, by conditioning on X_1, we get

$$\begin{aligned}f(i) &= \mathbf{P}(X_\tau = 0 \,\big|\, X_0 = i) \\ &= \sum_j \mathbf{P}(X_\tau = 0 \,\big|\, X_0 = i, X_1 = j)\mathbf{P}(X_1 = j \,\big|\, X_0 = i) \\ &= \frac{1}{2}\mathbf{P}(X_\tau = 0 \,\big|\, X_1 = i-1) + \frac{1}{2}\mathbf{P}(X_\tau = 0 \,\big|\, X_1 = i+1) \\ &= \frac{f(i+1) + f(i-1)}{2}.\end{aligned}$$

Using these equations for $i = 1, 2, \ldots, L-1$, we successively get $f(2) = 2f(1)$, $f(3) = 2f(2) - f(1) = 3f(1)$, etc., i.e., $f(i) = if(1)$. For $i = L$ we must get 1 and hence $f(1) = 1/L$. Thus, $f(i) = i/L$. For the original problem, this means $\mathbf{P}\{X_\tau = -a\} = \frac{b}{a+b}$.

Another question of interest here is

Question 2: What is the distribution of τ or at least its expected value? This is the time for one of the players to go bankrupt.

Again, one may consider the chain on the state space $S = \{0, 1, \ldots, L\}$ and consider the starting point as a variable. Let $g(i) = \mathbf{E}_i[\tau] := \mathbf{E}[\tau \,|\, X_0 = i]$. Then, $g(0) = 0 = g(L)$, while for $0 < i < L$ we condition on the fist step to get

$$
\begin{aligned}
g(i) &= \frac{1}{2}\mathbf{E}\left[\tau \,\middle|\, X_0 = i, X_1 = i-1\right] + \frac{1}{2}\mathbf{E}\left[\tau \,\middle|\, X_0 = i, X_1 = i+1\right] \\
&= \frac{1}{2}\left(1 + \mathbf{E}\left[\tau - 1 \,\middle|\, X_1 = i-1\right]\right) + \frac{1}{2}\left(1 + \mathbf{E}\left[\tau - 1 \,\middle|\, X_1 = i+1\right]\right) \\
&= 1 + \frac{1}{2}g(i-1) + \frac{1}{2}g(i+1).
\end{aligned}
$$

Thus, $g_{i+1} - g_i = g_i - g_{i-1} - 2$ for $i \geq 1$. Summing this we get $g_i - g_0 = ig_1 - i(i-1)$. Recall that $g_0 = 0$ and hence $g_{i+1} = ig_1 - i(i-1)$. Now, observe that $g_{L-i} = g_i$ by symmetry. Hence, we must have $g_1 = g_{L-1} = (L-1)g_1 - (L-1)(L-2)$ which implies $g_1 = L-1$. Plugging back we get $g_i = i(L-i)$ for all i.

Hitting times and probabilities

The ideas illustrated in case of the Gambler's ruin problem can be used in any MC, except that we may not be able to find the solution explicitly at the end. We illustrate this by studying the two important concepts illustrated above - *hitting probabilities* and *hitting times*.

Let $X_0, X_1, X_2, \ldots$ be a Markov chain with state space S and transition matrix $P = (p_{i,j})_{i,j \in S}$. For $A \subseteq S$, let $\tau_A := \min\{n \geq 0 : X_n \in A\}$ and $\tau_A^+ := \min\{n \geq 1 : X_n \in A\}$. If $X_0 \in A$, then $\tau_A = 0$ while τ^+ may not be. If $X_0 \notin A$, then $\tau = \tau^+$. Note that τ, τ^+ can take the value $+\infty$ with positive probability, depending on the chain and the subset A. τ_A is called the *hitting time of the set* A and τ^+ is called the *return time to the set* A.

Hitting probabilities: Let A and B be two disjoint subsets of S. We investigate the probability that A is hit before B, i.e., $\mathbf{P}_i(\tau_A < \tau_B)$. The idea of conditioning on the first step can be used to get a "difference equation" for this quantity.

Exercise 106. Let $A, B \subseteq S$ be disjoint. Assume that $\mathbf{P}_i(\tau_{A \cup B} < \infty) = 1$ for all i (this is trivially true for $i \in A \cup B$ but in general, may not be so for $i \notin A \cup B$). Let $f(i) = \mathbf{P}_i(\tau_A < \tau_B)$ be the probability that the MC started at i hits A before B. Show that $f(i) = 1$ for $i \in T_1$, $f(i) = 0$ for $i \in T_2$ and

$$f(i) = \sum_j p_{i,j} f(j), \text{ if } i \notin A \cup B.$$

If S is finite (or even when $S \backslash (A \cup B)$ is finite), can you show that f is the unique function that satisfies these equations and has the value 1 on A and 0 on B? [**Hint:** If g is another such function, consider the point where $f - g$ attains its maximum].

Remark: When the MC is a SRW on a graph, the equations for f just say that $f(u)$ is equal to the average value of $f(v)$ over all neighbours v of u. We say that f has the *mean-value property* or that f is a *discrete harmonic function*.

Hitting times: Let $A \subseteq S$. τ_A takes values in $\{0, 1, 2 \ldots\} \cup \{\infty\}$. For $i \in A$, clearly $\mathbf{P}_i(\tau_A = 0) = 1$. What is the distribution of τ_A? For $i \notin A$, we can write,

$$\mathbf{P}_i(\tau_A = n+1) = \sum_{i_1, \ldots, i_n \notin A} \sum_{j \in A} p_{i,i_1} p_{i_1,i_2} \cdots p_{i_{n-1},i_n} p_{i_n,j}.$$

An expression not very helpful in general. We can also condition on the fist step and write recursive formulas such as

$$\mathbf{P}_i(\tau_A = n+1) = \sum_j \mathbf{P}_i(\tau_A = n+1 \,|\, X_1 = j) \mathbf{P}_i(X_1 = j) = \sum_j p_{i,j} \mathbf{P}_j(\tau_A = n).$$

Note that this is valid only for $i \notin A$.

Let τ_j denote the hitting time to the singleton $\{j\}$. We then have the following identity.

$$p^{(n)}(i,j) = \sum_{k=1}^{n} \mathbf{P}_i(\tau_j = k) p^{(n-k)}(j,j).$$

Intutively this says that if $X_0 = i$ and $X_n = j$, then there must be some $k \le n$ when the state j was visited for the first time, and then starting from the state j it must return to the same state after $n-k$ steps. To give a rigorous proof, we do the following.

$$p^{(n)}(i,j) = \mathbf{P}_i(X_n = j) = \sum_{k=1}^{n} \mathbf{P}_i(\tau_j = k, X_n = j)$$

All we need to show is that $\mathbf{P}_i(\tau_j = k, X_n = j) = p_{j,j}^{(n-k)} \mathbf{P}_i(\tau_j = k)$ or equivalently $\mathbf{P}_i(X_n = j \,\big|\, tau_j = k) = p_{j,j}^{(n-k)}$. What is clear is that $\mathbf{P}_i(X_n = j \,\big|\, X_k = j) = p_{j,j}^{(n-k)}$. Intuitively, $\tau_j = k$ appears to give no more information than $X_k = j$, but this needs justification (if you do not feel so, observe that if $n < k$, then $\mathbf{P}_i(X_n = j \,\big|\, \tau_j = k) = 0$ but $\mathbf{P}_i(X_n = j \,\big|\, X_k = j)$ need not be). Now for the justification when $n \ge k$.

$$\begin{aligned}
\mathbf{P}_i(\tau_j = k, X_n = j) &= \mathbf{P}_i(X_n = j, X_k = j, X_m \neq k \text{ for } m < k) \\
&= \sum_{i_1,\dots,i_{k-1} \neq j} \mathbf{P}_i(X_n = j \,\big|\, X_k = j, X_m = i_m \text{ for } m < k)\mathbf{P}_i(X_k = j, X_m = i_m \text{ for } m < k) \\
&= \sum_{i_1,\dots,i_{k-1} \neq j} p_{j,j}^{(n-k)} \mathbf{P}_i(X_k = j, X_m = i_m \text{ for } m < k) \\
&= p_{j,j}^{(n-k)} \mathbf{P}_i(\tau_j = k).
\end{aligned}$$

This completes the proof.

RECURRENCE AND TRANSIENCE

Recall that $\tau_i^+ = \inf\{n \ge 1 : X_n = i\}$. Let $\alpha_{i,j} = \mathbf{P}_i(\tau_j^+ < \infty)$. We say that j is a *recurrent state* if $\alpha_{j,j} = 1$ and that j is a *transient state* if $\alpha_{j,j} < 1$. Let $N_j = \sum_{n=1}^{\infty} 1_{X_n = j}$ be the number of returns to the state j. Define $G(i,j) = \mathbf{E}_i[N_j]$, called the *Green's function* of the Markov chain. These quantities may be infinite.

We want to get a criterion for recurrence and transience in terms of transition probabilities. Recall the proof we gave for Pólya's theorem about simple random walks on $\mathbb{Z}^d$. The key to it was that observation that N_j has a Geometric distribution, since every time the SRW returns to the origin, it is like starting the whole process again. We did not give a proof at that time. We now state the analogous claim for general Markov chains and prove it.

Lemma: $\mathbf{P}_i(N_j \ge m) = \alpha_{i,j}\alpha_{j,j}^{m-1}$ for $m \ge 1$.

Proof: For $m = 1$ this is just the definition. For $m = 2$,

$$\begin{aligned}
\mathbf{P}_i(N_j = 2) &= \sum_{1 \le k < \ell} \mathbf{P}(X_k = X_\ell = j, X_n \neq j \text{ for other } n \le \ell) \\
&= \sum_{1 \le k < \ell} \mathbf{P}_i(\tau_j = k)\mathbf{P}_i(X_\ell = j, X_n \neq j \text{ for } k < n < \ell \,\big|\, \tau_j = k) \\
&= \sum_{1 \le k < \ell} \mathbf{P}_i(\tau_j = k)\mathbf{P}_j(\tau_j = \ell - k) \\
&= \sum_{k,m \ge 1} \mathbf{P}_i(\tau_j = k)\mathbf{P}_j(\tau_j = m) \\
&= \alpha_{i,j}\alpha_{j,j}.
\end{aligned}$$

Similarly one can continue and get the result for general m.

Consequently,

$$G(i,j) = \mathbf{E}_i[N_j] = \sum_{k \ge 1} \mathbf{P}_i(N_j \ge k) = \frac{\alpha_{i,j}}{1 - \alpha_{j,j}}.$$

Of course, this is finite only if $\alpha_{j,j} < 1$, but the equation holds even when $\alpha_{j,j} = 1$. Indeed, if $\alpha_{j,j} = 1$, then by the lemma, $\mathbf{P}_i(N_j \ge m) = \alpha_{i,j}$ for all m, which shows that $\mathbf{P}_i(N_j = \infty) = \alpha_{i,j}$. Thus, if $\alpha_{i,j} > 0$ and $\alpha_{j,j} = 1$ (in particular, if $\alpha_{j,j} = 1$ and $i = j$), then N_j is infinite with positive probability, hence its expected value is also infinite. Thus, we have shown that j is recurrent (transient) if and only if $G(j,j) = \infty$ (respectively, $G(j,j) < \infty$).

To write a criterion for recurrence in terms of the transition probabilities, we observe that

$$G(i,j) = \mathbf{E}_i\left[\sum_{n=1}^{\infty} \mathbf{1}_{X_n = j}\right] = \sum_{n=1}^{\infty} p_{i,j}^{(n)}.$$

Thus we have proved the following theorem.

Theorem: Let $X_0, X_1, \ldots$ be a Markov chain with state space S and transition matrix $P = (p_{i,j})_{i,j \in S}$. Then, for any $i, j \in S$with $\alpha_{i,j} > 0$ (in particular, if $i = j$), we have

$$j \text{ is } \textit{recurrent} \iff G(i,j) = \infty \iff \sum_{n=1}^{\infty} p_{i,j}^{(n)} = \infty$$

$$j \text{ is } \textit{transient} \iff G(i,j) < \infty \iff \sum_{n=1}^{\infty} p_{i,j}^{(n)} < \infty$$

If $\alpha_{i,j} = 0$, then $\mathbf{P}_i(N_j = 0) = 1$ and $G(i,j) = 0$ and $\sum_{n=1}^{\infty} p_{i,j}^{(n)} = 0$.

From this theorem, to check for recurrence/transience, we only need to estimate the transition probabilities. This may or may not be easy depending on the example at hand.

Example 107. For SRW on $\mathbb{Z}^d$, we have

$$\begin{aligned}
p^{(2n)}(0,0) &= \left(\frac{1}{2d}\right)^{2n} \sum_{\substack{k_1,\ldots,k_d \\ k_1+\ldots+k_d=n}} \frac{(2n)!}{(k_1!)^2 \ldots (k_d!)^2} \\
&\leq \frac{(2n)!}{2^{2n}(n!)^2} \left(\max_{\substack{k_1,\ldots,k_d \\ k_1+\ldots+k_d=n}} \frac{n!}{k_1! \ldots k_d!} \left(\frac{1}{d}\right)^n \right) \sum_{\substack{k_1,\ldots,k_d \\ k_1+\ldots+k_d=n}} \frac{n!}{k_1! \ldots k_d!} \left(\frac{1}{d}\right)^n \\
&= \frac{(2n)!}{2^{2n}(n!)^2} \frac{n!}{[(n/d)!]^d}
\end{aligned}$$

where by n/d we mean really integers closest to n/d (but they have to sum up to n, so they may not be all exactly equal). Using Stirling's formula, this tells us that

$$p^{(2n)}(0,0) \sim 2\left(\frac{d}{2\pi}\right)^{d/2} \frac{1}{n^{d/2}}$$

which is summable for $d \geq 3$ but not summable for $d = 1$ or 2. In this chain, observe that $p_{j,j}^{(n)} = p_{0,0}^{(n)}$ by symmetry, and hence all states are recurrent for $d = 1, 2$ and all states are transient for $d \geq 3$.

Example 108. Consider SRW on the triangular lattice in $\mathbb{R}^2$. First convince yourself that this is the Markov chain whose state space is $S = \{a + b\omega + c\omega^2 : a, b, c \in \mathbb{Z}\}$ and ω is a cube-root of 1. The transition probabilities are given by

$$p(i,j) = \begin{cases} \frac{1}{6} & \text{if } j - i \in \{\pm 1, \pm\omega, \pm\omega^2\}. \\ 0 & \text{otherwise.} \end{cases}$$

Now show that $p_{0,0}^{(n)} \sim \frac{c}{n}$ for some constant c and hence conclude that all states are recurrent.

In this example, every state has six neighbours, just as in $\mathbb{Z}^3$. But the latter MC is transient while the former is recurrent, not surprising because the graph structures are completely different (in the triangular lattice one can return in three steps, but in $\mathbb{Z}^3$ it is not possible). This is to emphasize that recurrence and transience do not depend on the degrees of vertices or such local properties. The triangular lattice is more like $\mathbb{Z}^2$ than like $\mathbb{Z}^3$.

Example 109. On the state space $\mathbb{Z}$, consider the transition probabilities $p_{i,i+1} = p$ and $p_{i,i-1} = 1 - p$ where $p \neq \frac{1}{2}$. Show that all states are transient. This is called *biased random walk* on $\mathbb{Z}$.

Show that same (transience of all states) if the state space is $\{0, 1, 2, \ldots\}$ with $p_{i,i+1} = \frac{2}{3}$, $p_{i,i-1} = \frac{1}{3}$ for $i \geq 1$ and $p_{0,1} = 1$.

Example 110. Consider SRW on the binary tree. The state space here is $S = \{(x_0, \ldots, x_n) : n \geq 0,\ x_i = 0 \text{ or } 1 \text{ for } i \geq 1,\ x_0 = 0\}$ (all finite strings of 0s and 1s). The vertex $(x_0, \ldots, x_n)$ is joined by an edge to $(x_0, \ldots, x_{n-1})$ (its parent) and $(x_0, \ldots, x_n, 1), (x_0, \ldots, x_n, -1)$ (its children). Let o denote the vertex (0).

Let $Y_n = h(X_n)$ denote the distance from o of the Markov chain. Show that Y_n is a MC with state space $\{0, 1, 2, \ldots\}$ with $p_{i,i+1} = \frac{2}{3}$, $p_{i,i-1} = \frac{1}{3}$ for $i \geq 1$ and $p_{0,1} = 1$. Use the previous example to conclude that the state o in the original chain X on the binary tree is transient.

STATES IN A MC

All the interesting examples of MCs that we have given have some nice properties which are not implied by the definition. For example, in all these chains, it is possible to go from any state to any other state (if allowed sufficiently many steps). However, this is by no means a requirement in the definition and not having such properties brings some annoying issues.

Let $X_0, X_1, \dots$ be a MC with state space S and transition probabilities $p_{i,j}$. Let $p_{i,j}^{(n)}$ denote the n-step transition probabilities. We define $p_{i,i}^{(0)} = 1$ and $p_{i,j}^{(0)} = 0$ if $j \neq i$.

We say that i leads to j and write $i \rightsquigarrow j$ if there exists some $n \geq 0$ such that $p_{i,j}^{(n)} > 0$. We say that i and j communicate with each other and write $i \leftrightsquigarrow j$ if $i \rightsquigarrow j$ and $j \rightsquigarrow i$. Observe that $i \rightsquigarrow j$ if and only if $\alpha_{i,j} > 0$ and $i \leftrightsquigarrow j$ if and only if $\alpha_{i,j} > 0$ and $\alpha_{j,i} > 0$.

Observation: The relationship "communicates with" is an equivalence relationship on S.

proof: Clearly, $i \leftrightsquigarrow i$ because $p_{i,i}^{(0)} = 1$. It is equally obvious that $i \leftrightsquigarrow j$ if and only if $j \leftrightsquigarrow i$. Lastly, if $i \rightsquigarrow j$ and $j \rightsquigarrow k$, then there exist n, m such that $p_{i,j}^{(n)} > 0$ and $p_{j,k}^{(m)} > 0$ and hence $p_{i,k}^{(n+m)} \geq p_{i,j}^{(n)} p_{j,k}^{(m)} > 0$ and hence $i \rightsquigarrow k$. Thus, if $i \leftrightsquigarrow j$ if and only if $j \leftrightsquigarrow i$ then $i \leftrightsquigarrow k$.

Consequence: The above equivalence relationship partitions the state space into pairwise disjoint equivalence classes $S = S_1 \sqcup S_2 \sqcup S_3 \dots$ such that $i \leftrightsquigarrow j$ if and only if i and j belong to the same S_p. Note that it may happen that $i \in S_1$ and $j \in S_2$ but $i \rightsquigarrow j$. However, then there can be no other $i' \in S_1, j' \in S_2$ such that $j' \rightsquigarrow i'$ (because since $j \leftrightsquigarrow j'$ and $i \leftrightsquigarrow i'$, we would get $j \rightsquigarrow i$ leading to the conclusion that $j \leftrightsquigarrow i$ which cannot be). We say that a state i is *absorbing* if $p_{i,i} = 1$. In this case $\{i\}$ is an equivalence class by itself.

Example 111. (Gambler's ruin). Let $S = \{0, 1, \dots, L\}$ where $L \geq 1$ is an integer. Let $p_{i,i+1} = p_{i,i-1} = 1/2$ for $0 < i < L$ and $p_{0,0} = 1$ and $p_{L,L} = 1$. This is SRW on $\mathbb{Z}$ absorbed at 0 and L.

In this case, if $0 < i < L$ and j is any state, then $i \rightsquigarrow j$. But $0 \rightsquigarrow j$ only for $j = 0$ and $L \rightsquigarrow j$ only for $j = L$. Thus, we get the equivalence classes $S_1 = \{1, \dots, L-1\}$, $S_2 = \{0\}$ and $S_3 = \{L\}$. Here 0 and L are (the only) absorbing states.

Example 112. Let $S = \mathbb{Z}^2 = \{(m,n) : m, n \in \mathbb{Z}\}$. Let $p_{(u_1,v_1),(u_2,v_2)} = \frac{1}{2}$ if $u_1 - u_2 = \pm 1,\ v_1 = v_2$. Then, it is easy to see that each horizontal line $S_k := \{(u,k) : u \in \mathbb{Z}\}$ is an irreducible class. In this chain, a MC will always stay inside the irreducible class. In otherwords, if $i \in S_k$ and $j \in S_\ell$, then $i \rightsquigarrow j$ if and only if $k = \ell$.

Let us keep the same state space and change the transition probabilities to

$$p_{(u_1,v_1),(u_2,v_2)} = \begin{cases} \frac{1}{4} & \text{if } u_1 - u_2 = \pm 1,\ v_1 = v_2. \\ \frac{1}{2} & \text{if } v_1 - v_2 = 1,\ u_1 = u_2. \\ 0 & \text{otherwise.} \end{cases}$$

In this case, again the irreducible classes are $S_k := \{(u,k) : u \in \mathbb{Z}\}$. Further, if $i \in S_k$ and $j \in S_\ell$, then $i \rightsquigarrow j$ if and only if $\ell \geq k$. Transitions can happen from one equivalence class to another (but never both ways!).

The MC is said to be *irreducible* if there is only one class. Then all states communicate with each other. Irreducible chains are nice, and most of the examples we see are irreducible. However, occasionally we will consider chains with absorbing states and that invariably leads to more classes.

Many properties of states are *class properties* in the sense that if two states are in the same irreducible class, then either both have that property or both do not.

Lemma: Recurrence and transience are class properties.

Proof: Suppose i is recurrent and $i \leftrightsquigarrow j$. Then, there exist $r, s \geq 1$ such that $p_{i,j}^{(r)} > 0$ and $p_{j,i}^{s} > 0$. Now, it is clear that $p_{j,j}^{(n+r+s)} \geq p_{j,i}^{(s)} p_{i,i}^{(n)} p_{i,j}^{(r)}$. Therefore, $\sum_n p_{j,j}^{(n+r+s)} \geq p_{j,i}^{(s)} p_{i,j}^{(r)} \sum_n p_{i,i}^{(n)} = \infty$ as i is recurrent. Thus, $\sum_n p_{j,j}^{(n)} = \infty$ and we see that j is recurrent. Thus states in one irreducible class are all recurrent or all transient.

We can now prove the following fact.

Theorem: All states in an irreducible, finite state space Markov chain are recurrent.

Proof: By irreducibility, all states must be transient or all states recurrent. Observe that $\sum_{j \in S} p_{i,j}^{(n)} = 1$ for each n and hence $\sum_{n=1}^{\infty} \sum_{i \in S} p_{i,j}^{(n)} = \infty$. Rerrange the sum to write $\sum_{i \in S} \sum_{n=1}^{\infty} p_{i,j}^{(n)} = \infty$. As the outer sum is over finitely many states, there must be at leasy one j such that $\sum_{n=1}^{\infty} p_{i,j}^{(n)} = \infty$. Find r such that $p_{j,i}^{(r)} > 0$ and use $p_{i,i}^{(n+r)} \geq p_{i,j}^{(n)} p_{j,i}^{(r)}$ to conclude that i is recurrent. Thus all states are recurrent.

Observe that this is false in infinite state space. For example, for biased random walk on $\mathbb{Z}$ with $p_{i,i+1} = 0.8$ and $p_{i,i-1} = 0.2$, all states are transient, but the chain is irreducible.

STATIONARY MEASURE AND DISTRIBUTION

Let $X_0, X_1, \ldots$ be a Markov chain with stat space S and transition matrix $P = (p_{i,j})_{i,j \in S}$.

Let $\pi = (\pi_i)_{i \in S}$ be a vector of non-negative numbers indexed by S. We say that π is a *stationary measure* for the MC above if $\sum_{i \in S} \pi_i p_{i,j} = \pi_j$ for every $j \in S$. If π is a probability vector, i.e., if $\sum_{i \in S} \pi_i = 1$, we say that π is a *stationary distribution*. If π is a stationary measure, then $c\pi$ is also a stationary measure for any $c > 0$.

If we write $\pi = (\pi_i)_{i \in S}$ as a row vector, the condition for π to be a stationary measure may be written in matrix form as $\pi \cdot P = \pi$. Thus, π is a left eigenvector of P with eigenvalue 1. However, á priori, not every such eigenvector may qualify as a stationary measure as we need the non-negativity condition $\pi_i \geq 0$. It is important to note that we are talking about *left* eigenvector here. The vector $u_i = 1$ for al i is a right eigenvector with eigenvalue 1 but not necessarily a stationary measure.

What is the interpretation of stationary measure? If π is a stationary distribution and the initial state $X_0 \sim \pi$, then

$$\mathbf{P}(X_1 = j) = \sum_i \mathbf{P}(X_1 = j, X_0 = i) = \sum_i \pi_i p_{i,j} = \pi_j.$$

Thus, X_1 also has the same distribution. Similarly, it is easy to see that the marginal distribution of X_n is π for each n. Thus, X_n are identically distributed, of course, they are not independent. If π is just a stationary measure (i.e., not a probability vector), this probabilistic interpretation makes no sense.

Exercise 113. If π is a stationary measure, show that $\sum_{i \in S} \pi_i p_{i,j}^{(n)} = \pi_j$ for all $j \in S$ and all $n \geq 0$.

Example 114. Consider the Ehrenfest chain. Let $\pi_i = \binom{N}{i}$. Then, for any $j \in S$,

$$\sum_{i \in S} \pi_i p_{i,j} = \pi_{j-1} p_{j-1,j} + \pi_{j+1} p_{j+1,j} = \binom{N}{j-1} \frac{N-j+1}{N} + \binom{N}{j+1} \frac{j+1}{N} = \binom{N}{j}.$$

This calculation also holds for $j = 0$ and $j = N$. Thus, π is a stationary measure. In this case, we can get a stationary distribution by normalizing π to get $\sigma_i = \binom{N}{i} 2^{-N}$.

Example 115. Consider SRW on a graph $G = (V, E)$. Recall that this means $S = V$ and $p_{u,v} = \frac{1_{v \sim u}}{d_u}$ where d_u is the degree of the vertex u. Let $\pi_u = d_u$ for $u \in V$. Then,

$$\sum_u \pi(u) p(u,v) = \sum_u d_u \frac{1_{u \sim v}}{d_u} = \sum_u 1_{u \sim v} = d_v = \pi(v).$$

Thus, π is a stationary measure. If the graph is finite, then we can normalize it by $\sum_{u \in V} \pi_u$ to get a stationary distribution.

For example, for Random walk on the hypercube, the stationary distribution is $\pi_x = 2^{-N}$ for each $x \in \{0,1\}^N$. For SRW on $\mathbb{Z}$, a stationary measure is $\pi_i = 1$ for all i. For random transposition shuffling of cards, the uniform distribution on all permutations is again stationary.

Example 116. Reflected, biased RW on $\mathbb{Z}_+$. Let $p_{i,i+1} = p = 1 - p_{i,i-1}$ for $i \geq 1$ and $p_{0,1} = 1$. If π is a stationary measure, it must satisfy $\pi_i =$

REVERSIBILITY

Let $X_0, X_1, X_2, \ldots$ be a MC with state space S and transitions $P = (p_{i,j})_{i,j \in S}$. Let $\pi = (\pi_i)_{i \in S}$ where $\pi_i \geq 0$. We say that the MC is reversible with respect to π if $\pi_i p_{i,j} = \pi_j p_{j,i}$ for all $i, j \in S$. We may also say that P is reversible w.r.t. π.

In this case, observe that $\sum_{i \in S} \pi_i p_{i,j} = \sum_i \pi_j p_{j,i} = \pi_j$ since $\sum_i p_{j,i} = 1$ for each j. Thus, π is a stationary measure for P.

The point is that the equations for stationary measure are difficult to solve in general. But it is easier to check if our chain is reversible w.r.t. some π and actually find that π if it is reversible. There are further special properties of reversible measures, but we shall not get there.

Example 117. SRW on $G = (V, E)$. Then, $p_{u,v} = \frac{1_{v \sim u}}{d_u}$ and $p_{v,u} = \frac{1_{u \sim v}}{d_v}$. It is easy to see that $d_u p_{u,v} = d_v p_{v,u}$ for all $u, v \in V$. Thus the chain is reversible and $\pi_u = d_u$ is a stationary measure.

Check that the Ehrenfest chain, random transposition shuffle are reversible. Consequently get stationary measures for them. The top to random shuffle is clearly not reversible (w.r.t any π) because there exist π, σ such that $p_{\pi,\sigma} > 0$ but $p_{\sigma,\pi} = 0$. The following exercise is useful in finding stationary measures in some cases where we do not have reversibility.

Exercise 118. Let P and Q be two transition matrices on the same state space S. Suppose $\pi = (\pi_i)_{i \in S}$ is a vector of non-negative numbers such that $\pi_i p_{i,j} = \pi_j q_{j,i}$ for all i, j. Then show that π is a stationary measure for P as well as for Q.

Exercise 119. Let P be the transition matrix for the top to random shuffle. Define an appropriate "random to top" shuffle and let Q denote its transition matrix and use the previous exercise to show that the uniform distribution on $\mathcal{S}_n$ is stationary for both P and Q.

EXISTENCE OF STATIONARY MEASURE

The following theorem shows that a stationary measure exists.

Theorem: Let $X_0, X_1, X_2, \ldots$ be a MC with state space S and transitions $P = (p_{i,j})_{i,j \in S}$. Fix a *recurrent state* $x \in S$. For any $j \in S$, define $\pi_j := \mathbf{E}_x\left[\sum_{n=0}^{\tau_x^+ - 1} \mathbf{1}_{X_n = j}\right]$ as the expected number of visits to j before returning to x. Then π is a stationary measure for the MC.

Proof: We can rewrite $\pi_j := \mathbf{E}_x\left[\sum_{n=0}^{\infty} \mathbf{1}_{X_n = j, n < \tau_x^+}\right] = \sum_{n=0}^{\infty} \mathbf{P}_x(X_n = j,\ n < \tau_x^+)$.

Now fix $j \in S$ and consider $\sum_{i \in S} \pi_i p_{i,j}$. Before proceeding, observe that $\mathbf{P}_x(X_{n+1} = j \,\big|\, X_n = i, n < \tau_x^+) = p_{i,j}$ for any $n \geq 0$. Hence,

$$\begin{aligned}
\sum_{i \in S} \pi_i p_{i,j} &= \sum_{i \in S} \sum_{n=0}^{\infty} \mathbf{P}_x(X_n = i,\ n < \tau_x^+) p_{i,j} \\
&= \sum_{i \in S} \sum_{n=0}^{\infty} \mathbf{P}_x(X_n = i,\ n < \tau_x^+) \mathbf{P}_x(X_{n+1} = j \,\big|\, X_n = i, n < \tau_x^+) \\
&= \sum_{i \in S} \sum_{n=0}^{\infty} \mathbf{P}_x(X_{n+1} = j,\ X_n = i,\ n < \tau_x^+). \qquad (7)
\end{aligned}$$

We split into two cases.

Case 1: $j \neq x$. Then, $\mathbf{P}_x(X_{n+1} = j,\ X_n = i,\ n < \tau_x^+) = \mathbf{P}_x(X_{n+1} = j,\ X_n = i,\ n+1 < \tau_x^+)$. Thus,

$$\begin{aligned}
\sum_{i \in S} \pi_i p_{i,j} &= \sum_{i \in S} \sum_{n=0}^{\infty} \mathbf{P}_x(X_{n+1} = j,\ X_n = i,\ n+1 < \tau_x^+) \\
&= \sum_{n=0}^{\infty} \sum_{i \in S} \mathbf{P}_x(X_{n+1} = j,\ X_n = i,\ n+1 < \tau_x^+) \\
&= \sum_{n=0}^{\infty} \mathbf{P}_x(X_{n+1} = j,\ n+1 < \tau_x^+) \\
&= \pi_j
\end{aligned}$$

since the $n = 0$ term is irrelevant when $j \neq x$.

Case 2: $j = x$. In this case, $\pi_x = 1$ by the definition. On the other hand, from (7), we see that

$$\begin{aligned}
\sum_{i \in S} \pi_i p_{i,x} &= \sum_{i \in S} \sum_{n=0}^{\infty} \mathbf{P}_x(X_{n+1} = x,\ X_n = i,\ n < \tau_x^+) \\
&= \sum_{n=0}^{\infty} \sum_{i \in S} \mathbf{P}_x(X_{n+1} = j,\ X_n = i,\ n < \tau_x^+) \\
&= \sum_{n=0}^{\infty} \mathbf{P}_x(X_{n+1} = j,\ n < \tau_x^+) \\
&= \sum_{n=0}^{\infty} \mathbf{P}_x(\tau_x^+ = n+1) \\
&= 1
\end{aligned}$$

because we assumed that x is a recurrent state. This finishes the proof.

Remark: If x is a transient state, then the definition of π can still be made and we see that in fact $\pi_i = G(x,i)$ for all $i \in S$. Further, observe that we used the recurrence of x only in the last step. In other words, when x is transient, we still have the equation

$$\sum_{i \in S} G(x,i) p_{i,j} = G(x,j), \text{ for } j \neq x.$$

We say that the function $i \to G(x,i)$ is *discrete harmonic* except at $i = x$.

Proposition: If a MC has a stationary distribution π, then any state $i \in S$ with $\pi_i > 0$ is necessarily recurrent.
Proof: Fix $j \in S$ such that $\pi_j > 0$. Then, $\pi_j = \sum_{i \in S} \pi_i p_{i,j}^{(n)}$ and hence

$$\infty = \sum_{n=1}^{\infty} \pi_j = \sum_{n=1}^{\infty} \sum_{i \in S} \pi_i p_{i,j}^{(n)} = \sum_{i \in S} \pi_i \sum_{n=1}^{\infty} p_{i,j}^{(n)} = \sum_{i \in S} \pi_i \frac{\alpha_{i,j}}{1 - \alpha_{j,j}} \leq \frac{1}{1 - \alpha_{j,j}}$$

since $\alpha_{i,j} \leq 1$ and $\sum_i \pi_i = 1$.

Uniqueness of stationary measures

In general stationary measures are not unique. For example, suppose $S = S_1 \sqcup S_2$ where S_1, S_2 are the irreducible classes of the Markov chain and suppose both
Theorem: Let $X_0, X_1, X_2, \ldots$ be an irreducible MC with state space S and transitions $P = (p_{i,j})_{i,j \in S}$. Then, the stationary measure is unique up to multiplication by constants. That is, if π and σ are two stationary measures, then there is some $c > 0$ such that $\pi_i = c\sigma_i$ for all $i \in S$.
Proof: Fix $x \in S$ and write for any j

$$\pi_j \quad = \quad \pi_x p_{x,j} + \sum_i^* \pi_i p_{i,j}$$

Using the first equation for each π_i in the second summand we get

$$\pi_j \quad = \quad \pi_x p_{x,j} + \pi_x \sum_i^* p_{x,i} p_{i,j} + \sum_{i,\ell}^* \pi_\ell p_{\ell,i} p_{i,j}$$

Iterating the same gives us for any $m \geq 1$,

$$\pi_j = \pi_x \sum_{k=1}^{m} \mathbf{P}_x(X_k = j, X_{k-1} \neq x, \ldots, X_1 \neq x) \; + \; \sum_i^* \pi_i \mathbf{P}_i(X_m = j, X_{m-1} \neq x, \ldots, X_1 \neq x, X_0 \neq x)$$

As the MC is irreducible and recurrent, for each $i \in S$ we have

$$\mathbf{P}_i(X_m = j, X_{m-1} \neq x, \ldots, X_1 \neq x, X_0 \neq x) \leq \mathbf{P}_i(\tau_x \geq m) \to 0$$

as $m \to \infty$.

As x is a recurrent state,

$$\mathbf{P}_\pi(X_m = j, X_{m-1} \neq x, \ldots, X_1 \neq x, X_0 \neq x) = \sum_{i \in S} \pi_i \mathbf{P}_i(X_m = j, X_{m-1} \neq x, \ldots, X_1 \neq x, X_0 \neq x) \to 0$$

as $m \to \infty$. From this we would like to claim that

$$\lim_{m \to \infty} \sum_i^* \pi_i \mathbf{P}_i(X_m = j, X_{m-1} \neq x, \ldots, X_1 \neq x, X_0 \neq x) = 0.$$

This is obvious if S is finite, because each summand is going to zero and the number of summands is fixed. If S is infinite, one can still argue this using what is called *Monotone convergence theorem*. We omit those details. Thus, we get

$$\pi_j = \pi_x \sum_{k=1}^{\infty} \mathbf{P}_x(X_k = j, X_{k-1} \neq x, \ldots, X_1 \neq x) = \pi_x \mathbf{E}_x \left[\sum_{k=0}^{\tau_x^+ - 1} \mathbf{1}_{X_k = j} \right] = \pi_x \mu_j.$$

where μ is the stationary measure that we constructed, starting from x and using the excursion back to x. Thus, uniqueness follows.

39. Various consequences

The existence and uniqueness of stationary measures and the proof we gave for it has various consequences. Let us say that an state i in the state space is *positive recurrent* if $\mathbf{E}_i[\tau_i^+] < \infty$. Of course, a positive recurrent state has to be recurrent.

Let $X_0, X_1, X_2, \ldots$ be an irreducible, recurrent MC with state space S and transitions $P = (p_{i,j})_{i,j \in S}$.

Suppose $x \in S$. Then, we can define the stationary measure

$$\pi_i^x := \mathbf{E}_x\left[\sum_{n=0}^{\tau_x^+-1} \mathbf{1}_{X_n=i}\right].$$

Clearly,

$$\sum_{i\in S}\pi_i = \sum_{i\in S}\mathbf{E}_x\left[\sum_{n=0}^{\tau_x^+-1} \mathbf{1}_{X_n=i}\right] = \mathbf{E}_x\left[\sum_{n=0}^{\tau_x^+-1}\sum_{i\in S} \mathbf{1}_{X_n=i}\right] = \mathbf{E}_x[\tau_x^+].$$

Thus, x is positive recurrent if and only if $\sum_i \pi_i < \infty$. Thus, the stationary measure π^x can be normalized to a stationary distribution if and only if x is positive recurrent.

By the uniqueness of stationary measure (up to constant multiples), either all stationary measures are normalizable or none is. And the stationary distribution, if it exists, is unique.

From two states x and y we get two stationary measures π^x and π^y. Either both or normalizable or neither. Hence, either both x and y are positive recurrent or neither. More generally you can prove the following (directly).

Exercise 120. In any MC, positive recurrence is a class property.

Note that $\sum_i \pi_i^x = \mathbf{E}_x[\tau_x^+]$ and $\pi_x(x) = 1$. If an irreducible chain is positive recurrent, the unique stationary distribution is given by $\theta_i = \frac{\pi_i^x}{\mathbf{E}_x[\tau_x^+]}$ (for any x). In particular,

$$\theta_x = \frac{1}{\mathbf{E}_x[\tau_x^+]}, \qquad \theta_i = \theta_x \pi_i^x.$$

Now, whether or not the chain is positive recurrent, by the uniqueness of stationary measures up to constant multiples, for any x, y, there is a constant $c > 0$ such that $\pi_i^x = c\pi_i^y$ for all i. In particular, take $i = y$ to get $c = \pi_y^x$. Thus, $\pi_i^x = \pi_i^y \pi_y^x$. In particular, $\pi_x^y \pi_y^x = 1$, a non-trivial identity.

The usefulness of the identity $\theta_i = \frac{1}{\mathbf{E}_i[\tau_i^+]}$ for irreducible positive recurrent chains goes in reverse. Often we can guess the stationary distribution θ (for example, if the chain is reversible) and consequently compute $\mathbf{E}_i[\tau_i^+] < \infty$.

Example 121. Consider the Ehrenfest chain on $\{0,1,2,\ldots,N\}$. We have seen that $\theta_i = \binom{N}{i}2^{-N}$ is the stationary distribution. Consequently,

$$\mathbf{E}_0[\tau_0^+] = \frac{1}{\theta_0} = 2^N, \qquad \mathbf{E}_{N/2}[\tau_{N/2}^+] = \frac{1}{\theta_{N/2}} \sim \sqrt{\pi}\sqrt{N}.$$

With $n = 10^{23}$, these two numbers are about $2^{10^{23}}$ and 10^{12}, vastly differing in magnitude. For example, if we assume that 10^{10} steps of the chain take place in one second (I don't know if this is a reasonable approximation to reality), then to return from $N/2$ to $N/2$ takes 100 seconds (on average) while return from 0 to 0 takes more than $2^{10^{22}}$ years!

Periodicity of Markov chains

Consider an irreducible recurrent Markov chain. Assume that it has a unique stationary distribution π. Ideally, we would like to say that for large n, the p.m.f of X_n is approximately π, regardless of the initial distribution. There is one hitch in this as the following example shows.

Example 122. Let G be the cyclic graph with vertex set $\{1,2,3,4\}$ with edges $\{i,i+1\}$, $1 \le i \le 3$ and $\{4,1\}$. The unique stationary distribution is given by $\pi_i = 1/4$ for $i \in \{1,2,3,4\}$.

Let $X_0 = 1$. Then, $X_n \in \{1,3\}$ if n is even and $X_n \in \{2,4\}$ if n is odd. Thus, we cannot say that $p_{1,i}^{(n)} = \mathbf{P}_1(X_n = i)$ has a limit. Still, it seems that only the parity of n matters. In other words, while $p_{1,i}^{(2n)} = 0$ for $i = 2,4$ and $p_{1,i}^{(2n+1)} = 0$ for $i = 1,3$, it seems reasonable to expect that $p_{1,i}^{(2n+1)} \to \frac{1}{2}$ for $i = 2,4$ and $p_{1,i}^{(2n)} \to \frac{1}{2}$ for $i = 1,3$.

The issue here is called periodicity and is the only obstacle to what we initially wanted to claim.

Definition 123. Let $X_0, X_1, \ldots$ be a Markov chain with state space S and transition matrix $P = (p_{i,j})_{i,j\in S}$. Fix $i \in S$ and consider the set $T_i = \{n \geq 1 : p_{i,i}^{(n)} > 0\}$. Then we say that $d_i = \text{g.c.d.}(T_i)$ is the period of i. If $d_i = 1$ for all $i \in S$, then we say that the Markov chain is *aperiodic*.

Example 124. For SRW on $\mathbb{Z}^d$, clearly $d_i = 2$ for all i. It is periodic (we shall say with period 2).

Consider SRW on the cycle graph with N vertices. If N is even, then again $d_i = 2$ for all i. However, if N is odd, $d_i = 1$ for all i (because $p_{i,i}^{(2)} > 0$ and $p_{i,i}^{(N)} > 0$).

Consider the problem of gambler's ruin with $S = \{0, 1, \ldots, L\}$. Check that $d_0 = d_L = 1$, while $d_i = 2$ for $1 \leq i \leq L-1$.

Lemma: In any MC, if $i \leftrightsquigarrow j$, then $d_i = d_j$. Thus, all states in the same irreducible class have the same period. In particular, all states in an irreducible Markov chain have the same period.

Proof: Find r,s such that $p_{i,j}^{(r)} > 0$ and $p_{j,i}^{(s)} > 0$. Then, $p_{i,i}^{(n+r+s)} \geq p_{j,j}^{(n)} p_{i,j}^{(r)} p_{j,i}^{(s)}$. With $n = 0$, this shows that d_i divides $r+s$. Secondly, for any $n \in T_j$, we see that $n+r+s \in T_i$ and hence d_i divided $n+r+s$. In particular, d_i divided n for all $n \in T_j$. Thus, d_i divided d_j. Reversing the roles of i and j we see that d_j must also divide d_i, and hence $d_i = d_j$.

Remark 125. If a MC is irreducible, and $p_{i,i} > 0$ for some $i \in S$, then the chain is aperiodic. This is obvious, since $d_i = 1$ and hence $d_j = 1$ for all j. In particular, any lazy random walk on a graph (with laziness parameter $\tau > 0$) is aperiodic. More generally, starting with any MC (S,P) and a number $0 < \tau < 1$, we can make a lazy version of the MC as follows.

Define $q_{i,j} = (1-\tau)p_{i,j}$ for $j \neq i$ and $q_{i,i} = p_{i,i} + \tau(1-p_{i,i})$. Then, the MC with transitions q is aperiodic because $q_{i,i} > 0$ for all i. This is called a lazy version of the original MC.

The point is that periodicity is a pain. The statements of the theorems are cleaner for aperiodic chains. There are two ways to deal with periodicity. One is to make a lazy version of the chain - this does not really change the nature of the chain, but makes it go slower. As lazy chains are aperiodic, the statements are cleaner for it.

An alternate is to consider the chain at times $0, d, 2d, \ldots$ only. It is easy to check that $X_0, X_d, X_{2d}, \ldots$ is an aperiodic MC (possibly on a smaller state space).

We shall use the following fact about aperiodic chains.

Lemma: Let (S,P) be an irreducible, aperiodic MC. Fix any $i, j \in S$. Then, there exists $N \geq 1$ such that $p_{i,j}^{(n)} > 0$ for all $n \geq N$.

Proof: First consider the case $i = j$. The set $T_i := \{n \geq 1 : p_{i,i}^{(n)} > 0\}$ has g.c.d. 1. Observe that $n, m \in T_i$ implies $an + bm \in T_i$ for all $a, b \geq 0$ since $p_{i,i}^{(an+bm)} \geq (p_{i,i}^{(n)})^a (p_{i,i}^{(m)})^b$. We claim that there is some N such that $N, N+1 \in T_i$. Indeed, take any $n, n+k \in T_i$. If $k = 1$ we are done, else, because the g.c.d. of T_i is 1, we must have $m \in T_i$ that is not divisible by k. Write $m = nk + r$ with $0 \leq r < k$. Then, $nk, nk+r \in T_i$ and their difference is $r < k$. Continue to decrease the difference till you get 1.

Now we have N such that $N, N+1 \in T_i$. Pick any number $m \geq N^2$ and write it as $m = N(N+k) + r$ where $k \geq 0$ and $0 \leq r \leq N-1$. Rewrite this as $m = (N+1)r + (N+k-r)N$. Since $N, N+1 \in T_i$, taking $a = N+k-r$ and $b = r$, we see that $m \in T_i$ for all $m \geq N^2$. This completes the proof for $j = i$.

If $j \neq i$, find r such that $p_{i,j}^{(r)} > 0$ and N such that $p_{i,i}^{(n)} > 0$ for all $n \geq N$. Then, $p_{i,j}^{(n)} \geq p_{i,i}^{(n-r)} p_{i,j}^{(r)} > 0$ for all $n \geq N + r$. This completes the proof.

CONVERGENCE THEOREM FOR IRREDUCIBLE, APERIODIC CHAINS

Theorem: Let $X_0, X_1, X_2, \ldots$ be an irreducible, aperiodic Markov chain on a state space S with transition matrix $P = (p_{i,j})_{i,j\in S}$ and having stationary distribution π. Then, for any $i, j \in S$, we have $p_{i,j}^{(n)} \to \pi_j$ as $n \to \infty$.

Proof: Let $X_0, X_1, \ldots$ be a MC with $X_0 = i$ and transitions given by P. Let $Y_0, Y_1, \ldots$ be an independent MC with the same state space and transitions but with $Y_0 \sim \pi$. Define $\tau := \min\{n \geq 0 : X_n = Y_n\}$ as the first time when the two chains meet.

The key point is that $\mathbf{P}(X_n = \ell, n \geq \tau) = \mathbf{P}(Y_n = \ell, n \geq \tau)$ which can be seen as follows.

$$\begin{aligned}
\mathbf{P}(X_n = \ell, n \geq \tau) &= \sum_{m=1}^{n} \sum_{i \in S} \mathbf{P}(X_n = \ell, \tau = m, X_m = i) \\
&= \sum_{m=1}^{n} \sum_{i \in S} \mathbf{P}(X_n = \ell \,\big|\, X_m = i) \mathbf{P}(\tau = m, X_m = i) \\
&= \sum_{m=1}^{n} \sum_{i \in S} \mathbf{P}(Y_n = \ell \,\big|\, Y_m = i) \mathbf{P}(\tau = m, Y_m = i) \\
&= \mathbf{P}(Y_n = \ell, n \geq \tau).
\end{aligned}$$

Now write

$$\mathbf{P}(X_n = \ell) = \mathbf{P}(X_n = \ell, n \geq \tau) + \mathbf{P}(X_n = \ell, n < \tau).$$
$$\mathbf{P}(Y_n = \ell) = \mathbf{P}(Y_n = \ell, n \geq \tau) + \mathbf{P}(Y_n = \ell, n < \tau).$$

Subtract to get

$$\big|\, \mathbf{P}(X_n = \ell) - \mathbf{P}(Y_n = \ell) \,\big| \leq \mathbf{P}(X_n = \ell, n < \tau) + \mathbf{P}(Y_n = \ell, n < \tau) \leq 2\mathbf{P}(\tau > n).$$

If we show that $\mathbf{P}(\tau < \infty) = 1$, then $\mathbf{P}(\tau > n) \to 0$ as $n \to \infty$ and hence we get

$$\sup_{\ell} \big|\, p_{x,\ell}^{(n)} - \pi_\ell \,\big| = \sup_{\ell} \big|\, \mathbf{P}(X_n = \ell) - \mathbf{P}(Y_n = \ell) \,\big| \leq 2\mathbf{P}(\tau > n) \to 0.$$

In particular, it follows that $p_{x,\ell}^{(n)} \to \pi_\ell$ for any x, ℓ.

It only remains to prove that τ is finite with probability one. To see this, observe that $Z_n := (X_n, Y_n)$ is a Markov chain on $S \times S$ with transition probabilities $q((i,j),(i',j')) = p_{i,i'} p_{j,j'}$. We claim that Z_n is an irreducible chain. To see this, fix $i, j, i', j' \in S$ and invoke aperiodicity of the MC on S to get an N such that $p_{i,i'}^{(n)} > 0$ and $p_{j,j'}^{(n)} > 0$ for all $n \geq N$. Then $q^{(n)}((i,j),(i',j')) = p_{i,i'}^{(n)} p_{j,j'}^{(n)} > 0$ for $n \geq N$. Hence the Q-chain is irreducible. Further, if we define $\theta(i,j) = \pi_i \pi_j$, then $\sum_{(i,j)} \theta(i,j) = 1$ and

$$\sum_{(i,j) \in S} \theta(i,j) q((i,j),(i',j')) = \left(\sum_i \pi_i p_{i,i'} \right) \left(\sum_j \pi_j p_{j,j'} \right) = \pi_{i'} \pi_{j'} = \theta(i',j').$$

Thus, θ is a stationary distribution for the Q-MC and hence the chain must be recurrent. Therefore, for any fixed $i \in S$, the chain Z_n hits (i,i) with probability one. Therefore, $\tau < \infty$ with probability one.