2.2 Uniqueness

Remark 5. Consider a Markov chain $(X_0)_{t\geq 0}$ with finite state set S and transition probabilities K. Initializing the chain with a probability distribution μ on S, the consecutive states have distributions $\mu, \mu K, \mu K^2, \ldots$. Recall from the proof of the existence theorem that any convergent subsequence of the partial averages

$$\frac{1}{n}(\mu + \mu K + \cdots + \mu K^{n-1}) \tag{17}$$

converges to a stationary distribution. If the Markov chain happens to have only one stationary distribution, then it follows that the partial averages (17) converge to this unique stationary distribution, *independent of the initial distribution* μ.

In particular, if $f : S \to \mathbb{R}$ is a function of the state (an *observable*), it follows that

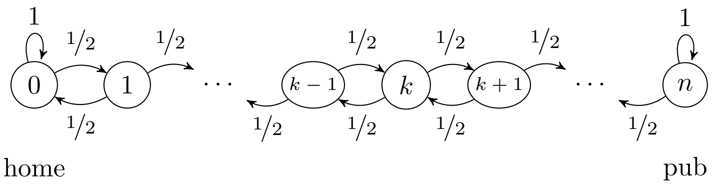
$$\mathbf{E}\left[\frac{1}{n}\left(f(X_0) + f(X_1) + \cdots + f(X_{n-1})\right)\right] \to \pi(f) \ , \tag{18}$$

independent of the distribution of X_0. (Recall the notation: $\pi(f) \triangleq \sum_{a\in S} \pi(a) f(a)$ is the expected value of f with respect to the distribution π.) Later we shall see that a much stronger statement holds: the averages

$$\frac{1}{n}\left(f(X_0) + f(X_1) + \cdots + f(X_{n-1})\right) \tag{19}$$

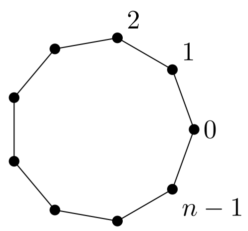
converge to $\pi(f)$ *almost surely* (i.e., with probability 1). In other words, the random variables $f(X_0), f(X_1), \ldots$ follow the law of large numbers. This is the so-called ergodic theorem of Markov chains. ◇

Example 6 (Non-uniqueness)**.** The Markov chain of the drunkard's walk (a.k.a. the gambler's ruin)



has at least two stationary distributions: the distribution δ_{home} concentrated at "home", and the distribution δ_{pub} concentrated at the "pub". In fact, every convex combination $\lambda\delta_{\text{home}} + (1-\lambda)\delta_{\text{pub}}$ (for $0 \leq \lambda \leq 1$) is also a stationary distribution. ○

Example 7 (Uniqueness)**.** The simple random walk on a cycle $\mathbb{Z}_n$ of length n



has an obvious stationary distribution: the uniform distribution $\pi(x) \triangleq {}^1\!/_n$. There is a simple argument showing that this is the only stationary distribution. For, let μ be a stationary distribution that is not uniform. Pick a vertex $a \in \mathbb{Z}_n$ such that

- $\mu(a)$ is maximum, and

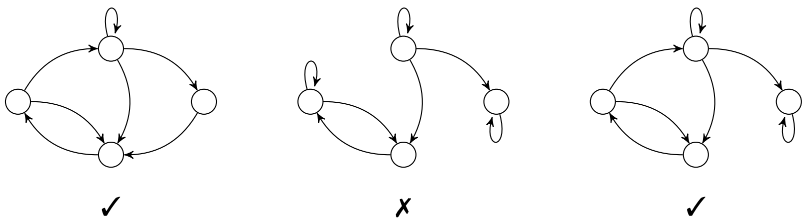
- $\mu(a-1) < \mu(a)$.

Then,

$$\mu K(a) = \underbrace{\mu(a-1)}_{<\mu(a)} \cdot \frac{1}{2} + \underbrace{\mu(a+1)}_{\leq\mu(a)} \cdot \frac{1}{2} < \mu(a) \ , \tag{20}$$

which means μ is not a stationary distribution. ○

The above examples (as well as others we have seen so far) might suggest that whether a finite-state Markov chain has one or more stationary distributions depends not on the actual transition probabilities $K(a, b)$ but only on the combinatorics of the transitions between different states. Indeed, by looking at the underlying graph of probable transitions one could guess which Markov chain has a unique stationary distribution and which has more.



(By the underlying graph of a Markov chain we mean a directed graph with a vertex for each state and a directed edge $a \to b$ for each pair of states a and b with $K(a, b) > 0$.)

A Markov chain $(X_t)_{t\geq 0}$ is said to be *irreducible* if for every two states a, b there is an integer $n > 0$ such that

$$\mathbf{P}(X_{t+n} = b \,|\, X_t = a) = K^n(a, b) > 0 \ . \tag{21}$$

A Markov chain is irreducible if and only if its underlying graph is strongly connected.

Lemma 8. *Any stationary distribution for an irreducible Markov chain is (strictly) positive.*

Proof. Let $(X_t)_{t\geq 0}$ be an irreducible Markov chain initialized according to a stationary distribution π. Let b be an arbitrary state. Choose a state a such that $\mathbf{P}(X_0 = a) = \pi(a) > 0$. By irreducibility, there is a time $n \geq 0$ such that $\mathbf{P}(X_n = b \,|\, X_0 = a) > 0$. It follows that

$$\pi(b) = \mathbf{P}(X_n = b) \geq \mathbf{P}(X_0 = a, X_n = b) \tag{22}$$
$$= \mathbf{P}(X_0 = a) \, \mathbf{P}(X_n = b \,|\, X_0 = a) > 0 \ . \tag{23}$$

□

Theorem 9. *An irreducible Markov chain has at most one stationary distribution.*

Remark 10. The argument used in Example 7 can be adapted to prove the above theorem in case of finite-state irreducible Markov chains (see [LPW08], Exercise 1.13). Various other proofs are possible, each with its own insight. Below, we shall take an alternative approach that has the advantage of providing an explicit probabilistic expression for the unique stationary distribution. Yet another proof is by defining a notion of "entropy" associated to each probability distribution and showing that the entropy increases unless the chain is in equilibrium. ◇

2.3 Equilibrium via Return Times

For each state x, consider the average time m_x it takes for the chain to return to x if started from x. If the chain is irreducible, every state x is visited over and over again, and the gap between every two consecutive visits is on average m_x. The law of large numbers suggests that on the long run, the chain is spending roughly $1/m_x$ of its time in state x. In equilibrium, therefore the chain has roughly chance $1/m_x$ to be in state x.

To be specific, let us denote by T_x^+ the first time $t > 0$ that the chain visits a state x, that is

$$T_x^+ \triangleq \inf\{t > 0 : X_t = x\} \ . \tag{24}$$

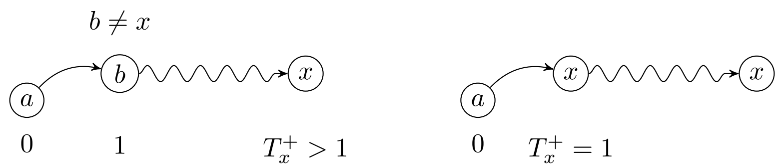
(We understand the infimum to be ∞ if the set is empty.)

Theorem 11 (Kac's recurrence lemma)**.** *Let $(X_t)_{t\geq 0}$ be an irreducible Markov chain and π a stationary distribution. Then,*

$$\pi(x) = \frac{1}{\mathbf{E}_x\, T_x^+} \tag{25}$$

for every $x \in S$.

Proof. Let us initialize the chain with the stationary distribution π and write the expected hitting time of x by conditioning on the state after one step. For clarity, we first assume that $\mathbf{E}\, T_x^+ < \infty$. Note that this also implies $\mathbf{E}_a\, T_x^+ < \infty$ for every $a \in S$ (why?). There are two cases, based on whether the chain hits x in its first step or not:



$$\mathbf{E}[T_x^+ \mid X_1 = b] = \begin{cases} 1 + \mathbf{E}_b\, T_x^+ & \text{if } b \neq x, \\ 1 & \text{if } b = x. \end{cases} \tag{26}$$

Therefore,

$$\mathbf{E}\, T_x^+ = \mathbf{E}[\mathbf{E}[T_x^+ \mid X_1]] = \sum_b \overbrace{\mathbf{P}(X_1 = b)}^{\pi(b)}\, \mathbf{E}[T_x^+ \mid X_1 = b] \tag{27}$$

$$= \sum_b \pi(b)\left(1 + \mathbf{E}_b\, T_x^+\right) - \pi(x)\, \mathbf{E}_x\, T_x^+ \ . \tag{28}$$

It follows that

$$\cancel{\mathbf{E}\, T_x^+} = 1 + \cancel{\sum_b \pi(b)\, \mathbf{E}_b\, T_x^+} - \pi(x)\, \mathbf{E}_x\, T_x^+ \ , \tag{29}$$

hence proving $\pi(x)\, \mathbf{E}_x\, T_x^+ = 1$.

To remove the assumption $\mathbf{E}\, T_x^+ < \infty$, we use a common truncation trick. Let us write $r \wedge s$ for the minimum of two numbers r and s. Using a similar conditioning as above, for any $m \geq 0$ we can write

$$\mathbf{E}[T_x^+ \wedge (m+1)] = \sum_b \pi(b)\left(1 + \mathbf{E}_b[T_x^+ \wedge m]\right) - \pi(x)\, \mathbf{E}_x[T_x^+ \wedge m] \tag{30}$$

$$= 1 + \mathbf{E}[T_x^+ \wedge m] - \pi(x)\, \mathbf{E}_x[T_x^+ \wedge m] \ . \tag{31}$$

Therefore,

$$\pi(x)\,\mathbf{E}_x[T_x^+ \wedge m] = 1 - \mathbf{E}[T_x^+ \wedge (m+1)] + \mathbf{E}[T_x^+ \wedge m] \;, \tag{32}$$

which reduces to

$$\pi(x)\,\mathbf{E}_x[T_x^+ \wedge m] = 1 - \mathbf{P}(T_x^+ > m) \;. \tag{33}$$

The rest follows by analyzing what happens if we let $m \to \infty$. As $m \to \infty$,

- $\mathbf{E}_x[T_x^+ \wedge m] \to \mathbf{E}_x\,T_x^+$ (independently of whether $\mathbf{E}_x\,T_x^+$ is finite or not), and
- $\mathbf{P}(T_x^+ > m) \to \mathbf{P}(T_x^+ = \infty)$.

These are by the monotone continuity of probabilities and expectation. Recall from Lemma 8 that $\pi(x) > 0$. Since the righthand side of (33) is bounded from above by 1, it follows that $\mathbf{E}_x\,T_x^+$ is finite, and in particular, $\mathbf{P}(T_x^+ = \infty) = 0$. Therefore, $\pi(x)\,\mathbf{E}_x\,T_x^+ = 1$. □

For irreducible chains, the following theorem gives another dynamical interpretation of the stationary distribution (when it exists), although the construction is valid even if the chain is not irreducible.

Theorem 12. *Let $(X_t)_{t\geq 0}$ be a Markov chain and x a state such that $\mathbf{E}_x\,T_x^+$ is finite. Then,*

$$\pi_x(z) \triangleq \frac{1}{\mathbf{E}_x\,T_x^+} \underbrace{\sum_{n=0}^{\infty} \mathbf{P}_x(X_n = z, T_x^+ > n)}_{\substack{\text{expected number of visits to } z \\ \text{before returning to } x}} \tag{34}$$

is a stationary distribution.

Proof. Let K denote the transition matrix of the chain. We verify that $(\pi_x K)(z) = \pi_x(z)$ for each state z. We consider two cases, based on whether $z \neq x$ or $z = x$.

For any two states y and z, we can write

$$\begin{aligned}
\pi_x(y)K(y,z) &= \frac{1}{\mathbf{E}_x\,T_x^+} \sum_{n=0}^{\infty} \mathbf{P}_x(X_n = y, T_x^+ > n) K(y,z) && (35)\\
&= \frac{1}{\mathbf{E}_x\,T_x^+} \sum_{n=0}^{\infty} \mathbf{P}_x(X_n = y, X_{n+1} = z, T_x^+ > n) \;. && (36)
\end{aligned}$$

First, suppose that $z \neq x$. Then,

$$\mathbf{P}_x(X_n = y, X_{n+1} = z, T_x^+ > n) = \mathbf{P}_x(X_n = y, X_{n+1} = z, T_x^+ > n+1) \tag{37}$$

and summing (36) over all y and exchanging the order of the sums we get

$$\begin{aligned}
\sum_y \pi_x(y)K(y,z) &= \frac{1}{\mathbf{E}_x\,T_x^+} \sum_{n=0}^{\infty} \sum_y \mathbf{P}_x(X_n = y, X_{n+1} = z, T_x^+ > n+1) && (38)\\
&= \frac{1}{\mathbf{E}_x\,T_x^+} \sum_{n=0}^{\infty} \mathbf{P}_x(X_{n+1} = z, T_x^+ > n+1) && (39)\\
&= \pi_x(z) \;. && (40)
\end{aligned}$$

The last step follows from the fact that $\mathbf{P}_x(X_0 = z, T_x^+ > 0) = 0$.

If on the other hand $z = x$, we have

$$\mathbf{P}_x(X_n = y, X_{n+1} = z, T_x^+ > n) = \mathbf{P}_x(X_n = y, T_x^+ = n+1) \;. \tag{41}$$

Summing (36) over all y and exchanging the order of the sums, this time we get

$$\sum_y \pi_x(y)K(y,x) = \frac{1}{\mathbf{E}_x T_x^+} \sum_{n=0}^{\infty} \sum_y \mathbf{P}_x(X_n = y, T_x^+ = n+1) \tag{42}$$

$$= \frac{1}{\mathbf{E}_x T_x^+} \sum_{n=0}^{\infty} \mathbf{P}_x(T_x^+ = n+1) \tag{43}$$

$$= \frac{1}{\mathbf{E}_x T_x^+} = \pi_x(x)\ . \tag{44}$$

Therefore, in either case, $(\pi_x K)(z) = \pi_x(z)$, and hence π_x is stationary. □

Remark 13. Kac's recurrence lemma implies the uniqueness theorem of irreducible Markov chains (Theorem 9): an irreducible Markov chain (with finite or countable state space) has at most one stationary distribution. If a stationary distribution exists, it must be strictly positive (Lemma 8) and therefore the expected return time $\mathbf{E}_x T_x^+$ of each state x is finite. A state x with finite expected return time is said to be *positively recurrent*.

On the other hand, Theorem 12 implies that any Markov chain that has a positively recurrent state has at least one stationary distribution. This condition is also necessary for the existence: if a Markov chain has a stationary distribution, it also has a stationary distribution concentrated on one of its irreducible components. Kac's lemma then implies that every state in that component is positively recurrent. (This will be more clear once we learn more about reducible Markov chains.) ◇

References

[Dur10] Rick Durrett. *Probability: Theory and Examples.* Cambridge University Press, 4th edition, 2010.

[LPW08] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times.* American Mathematical Society, 2008.