Note 13. Discrete-time Markov chains III

Alex Fu

Fall 2022

6 First-step analysis

We have looked in detail at the distributions of X_n and $(X_k, \ldots, X_{n+k})$, but distributions do not tell the whole story about the random variables $(X_n)_{n \in \mathbb{N}}$. Taking the perspective of an individual particle performing a random walk, or an individual realization $(X_n(\omega))_{n \in \mathbb{N}}$, we can ask the exact **times** we visit certain states $j \in S$ or revisit the initial state x_0.

These times, defined on every outcome, are thus random variables. While their distributions can be quite complicated, we often wish to examine their expectations or probabilities of being finite. This in turn leads us to *classify* states visited infinitely often (i.o.) or finitely often, and states in which a particle spends most of its time, revisiting repeatedly — in turn leading us right back to stationarity and asymptotic behavior.

Definition 1 (Hitting time; positive hitting time; time to revisit).

The **hitting time** T_A of a subset of states $A \subseteq S$ is the $\mathbb{N}$-valued random variable for the time when the chain first enters a state in A:

$$T_A := \min_n \{n \in \mathbb{N} : X_n \in A\}.$$

If no such n exists, then $T_A := \infty$. When A consists of a single state j, we also write T_j. The **positive hitting time** differs from the hitting time only when $X_0 \in A$:

$$T_A^+ := \min_n \left\{n \in \mathbb{Z}^+ : X_n \in A\right\}.$$

Likewise, we write T_j^+ when $A = \{j\}$. Given the initial state $X_0 = i$, the **time to revisit** or *time of first return* for the state i is T_i^+.

For the sake of convenience, we also introduce the following conventional notations:

$$\mathbb{P}_i(\cdot) := \mathbb{P}(\cdot \mid X_0 = i)$$
$$\mathbb{E}_i(\cdot) := \mathbb{E}(\cdot \mid X_0 = i).$$

Definition 2 (Total number of visits; long-term fraction of time).

The **total number of visits** to a state i (or a subset of states $A \subseteq S$ accordingly) is

$$N_i := \sum_{n \in \mathbb{N}} \mathbb{1}_{X_n = i}.$$

The **long-term fraction of time** or *proportion of time* spent in state i (or $A \subseteq S$) is

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N} \mathbb{1}_{X_n = i}.$$

Note that the long-term fraction of time is always defined, so it is distinguished from the probability mass of state i in the *limiting* distribution, which may not always exist.

Definition 3 (Probability of eventually reaching; probability of eventual revisit).

The probability of **eventually reaching** state j from state i is

$$p_{i \to j} = p(i \to j) := \mathbb{P}_i(T_j^+ < \infty).$$

The probability of **eventual revisit** or *eventual return* to state i is $p_{i\circlearrowleft} = p(i \to i)$.

Note that $p(i \to j)$ is not the one-step transition probability or any k-step transition probability. The notation $i \to j$ is meant to remind of **reachability**, the existence of a sequence of transitions from i to j of *some* finite length. Thus, for any particular number of steps $k \in \mathbb{Z}^+$,

$$p(i \to j) = \mathbb{P}\left(\bigcup_{k \in \mathbb{Z}^+} \{X_k = j\} \,\middle|\, X_0 = i \right) \geq \mathbb{P}(X_k = j \mid X_0 = i) = p^{(k)}(i, j).$$

First-step analysis is a class of techniques that set up a system of equations, most often recurrence relations, to find certain probabilities or expectations involving times. The system relies on the law of total probability or expectation to break down the relevant quantity into a sum over all possible one-step transitions out of the initial state. By the Markov property, what follows the "*first step*" is the same as simply taking a new initial state.

Proposition 1 (Hitting probabilities).

The **hitting probability** of state $j \in A \subseteq S$ from state $i \in S$ is

$$h_i(j) := \mathbb{P}_i(X_{T_A} = j) = \sum_{k \in S} p(i, k) \cdot h_k(j).$$

Proof. As stated, we will use the law of total probability and the Markov property (starred):

$$
\begin{aligned}
h_i(j) = \mathbb{P}_i(X_{T_A} = j) &= \sum_{k \in S} \mathbb{P}(X_{T_A} = j \mid X_1 = k, X_0 = i) \cdot \mathbb{P}(X_1 = k \mid X_0 = i) \\
&\overset{*}{=} \sum_{k \in S} \mathbb{P}(X_{T_A} = j \mid X_1 = k) \cdot p(i,k) \\
&= \sum_{k \in S} \mathbb{P}_k(X_{T_A} = j) \cdot p(i,k) \\
&= \sum_{k \in S} p(i,k) \cdot h_k(j).
\end{aligned}
$$

Notice that the equations indeed only involve one-step transition probabilities $p(i,k)$. Of course, recurrence relations also require base cases, in this case for $i \in A$, or $T_A = 0$:

$$
h_i(j) = \mathbb{1}_{i=j} \qquad \forall i \in A.
$$

□

Proposition 2 (Expected hitting times).

The **expected hitting time** or *mean hitting time* of $A \subseteq S$ is given by

$$
\mathbb{E}_i(T_A) = \begin{cases} 1 + \sum_{j \in S} p(i,j) \cdot \mathbb{E}_j(T_A) & \text{if } i \notin A \\ 0 & \text{if } i \in A. \end{cases}
$$

When $A = \{j\}$, this may remind you of breadth-first search over a graph, though with sums of "breadths" weighted by transition probabilities rather than the simpler incrementing. As more foreshadowing for CTMCs, note that the expression above is equal to $\sum_{j \in S} p_{i,j} \cdot (1 + \mathbb{E}_j(T_A))$.

Proof. We proceed in an identical fashion as the previous proof. For $i \notin A$,

$$
\begin{aligned}
\mathbb{E}_i(T_A) &= \sum_{j \in S} \mathbb{E}(T_A \mid X_1 = j, X_0 = i) \cdot \mathbb{P}(X_1 = j \mid X_0 = i) \\
&\overset{*}{=} \sum_{j \in S} \mathbb{E}(T_A \mid X_1 = j) \cdot \mathbb{P}(X_1 = j \mid X_0 = i) \\
&= \sum_{j \in S} p(i,j) \cdot (1 + \mathbb{E}_j(T_A)).
\end{aligned}
$$

□

Recall that a state i is **absorbing** if no edges lead out of it in the transition diagram, or $p(i,i) = 1$: it has a self-loop with probability 1. If a particle on the chain enters an absorbing state, it will remain there almost surely, so absorbing states, and more generally sinks or *closed* classes, can model *persistent* conditions such as wins, losses, system exits, etc.

Definition 4 (Absorbing chain).

A *chain* is called **absorbing** if from any state $j \in S$, some absorbing state i is reachable.

Any convex combination of the indicators for absorbing states, if one exists, gives a stationary distribution for the chain. For absorbing chains in particular, the absorbing states describe a "complete" set of conditions in which any particle must eventually end up. We may wish to find the time it takes to reach absorption, which is a slightly special hitting time:

Proposition 3 (Expected absorption time).

Let $A, B \subseteq S$ be classes of absorbing states so that every absorbing state belongs to A or B. Then the **expected absorption time** or *mean absorption time* to reach A (before reaching B) from state i is given by

$$\mathbb{E}_i(T_A) = \begin{cases} 1 + \sum_{j \in S} p(i,j) \cdot \mathbb{E}_j(T_A) & \text{if } i \in S \setminus (A \cup B) \\ 0 & \text{if } i \in A \\ \infty & \text{if } i \in B. \end{cases}$$

A and B most often describe "good" persistent conditions to seek and "bad" conditions to avoid, such as win conditions and loss conditions, and more situation-specific modifications can certainly be made. For absorbing chains, Proposition 3 also gives an "expected time to convergence" when A contains every absorbing state.

Proposition 4 (Expected time to return).

The **expected time to return** or *mean return time* of state i is given by

$$\mathbb{E}_i(T_i^+) = 1 + \sum_{j \neq i} p(i,j) \cdot \mathbb{E}_j(T_i^+).$$

Note that there is no base case; it is possible for the mean return time to be infinite. The proof proceeds in the same manner that we have already seen, so it is omitted and left to the reader. We also invite you to make similar generalizations from individual states $i \in S$ to subsets $A \subseteq S$ for the results above; the analogous counterparts to absorbing states in Proposition 3 are closed classes, which "once you enter, you will never leave" (almost surely).

Beyond the *absorption probability* for a particular absorbing state $j \in A$, which we leave to you, one final quantity of interest for the moment, the *expected number of visits* $\mathbb{E}_i(N_j)$, will require an important definition in the following section which turns out to be a class property.

7 Classification

7.1 Recurrence and transience

Definition 5 (Recurrence; transience).

A state $i \in S$ is **recurrent** if the chain almost surely revisits i given the starting state $X_0 = i$, i.e. the probability of eventually revisiting i is 1:

$$p_{i\circlearrowleft} = 1.$$

A state is **transient** if it is not recurrent, i.e. $p_{i\circlearrowleft} < 1$.

Equivalently, i is recurrent if

$$1 - p_{i\circlearrowleft} = \mathbb{P}_i(T_i^+ = \infty) = \lim_{n\to\infty} \mathbb{P}_i(T_i^+ \geq n) = 0.$$

From here, we see that transience implies the expected time to revisit $\mathbb{E}_i(T_i^+)$ must be infinite. This explains why the definition of recurrence is not simply $p_{i\circlearrowleft} > 0$: a recurrent state should be revisited in finite time *almost surely*, not simply with some nonzero probability.

Proposition 5 (Revisited in finite time a.s. iff revisited infinitely often a.s.).

If state $i \in S$ is recurrent, then i is revisited **infinitely often** almost surely:

$$\mathbb{P}_i(N_i = \infty) = 1.$$

This implies that $\mathbb{P}_i(N_i < \infty) = 0$ and $\mathbb{E}_i(N_i) = \infty$. Conversely, a transient state is revisited only finitely often almost surely, which is far stronger than $\mathbb{P}_i(N_i < \infty) > 0$:

$$\mathbb{P}_i(N_i < \infty) = 1.$$

Moreover, the expected number of *re*visits to state i can be found explicitly as

$$\mathbb{E}_i(N_i) = \frac{p_{i\circlearrowleft}}{1 - p_{i\circlearrowleft}} < \infty.$$

Proof. For every time the chain visits state i, it eventually revisits i with probability $p_{i\circlearrowleft}$, or never revisits i with probability $1 - p_{i\circlearrowleft}$. By the conditional independence given by the Markov property, the probability that the chain revisits i exactly k times is

$$\mathbb{P}_i(N_i = k) = (p_{i\circlearrowleft})^k \cdot (1 - p_{i\circlearrowleft}).$$

So $N_i \mid X_0 = i$ is distributed as Geometric$(1 - p_{i\circlearrowleft}) - 1$, which explains the condition of $p_{i\circlearrowleft} = 1$: if i is recurrent, then $\mathbb{P}_i(N_i = k) = 0$ for all finite $k \in \mathbb{N}$, so we must have $\mathbb{P}_i(N_i = \infty) = 1$. Or,

$$\mathbb{P}_i(N_i = \infty) = \lim_{k\to\infty} \mathbb{P}_i(N_i > k) = \lim_{k\to\infty} (p_{i\circlearrowleft})^{k+1} = 1$$

We can also thus see why transience, $p_{i\circlearrowleft} < 1$, implies that $\mathbb{P}_i(N_i = \infty) = 0$, or $\mathbb{P}_i(N_i < \infty) = 1$. The expected number of *re*visits can also be found as

$$\mathbb{E}_i(N_i) = \mathbb{E}(\text{Geometric}(1 - p_{i\circlearrowleft}) - 1) = \frac{1}{1 - p_{i\circlearrowleft}} - 1 = \frac{p_{i\circlearrowleft}}{1 - p_{i\circlearrowleft}}.$$

□

We will highlight the result found in the proof above as its own proposition.

Proposition 6 (Distribution of number of visits).

Given $X_0 = i$, the number of revisits N_i is distributed as Geometric$(1 - p_{i\circlearrowleft}) - 1$. In general, given that $X_0 = i$, the number of visits to state j has distribution

$$\mathbb{P}_i(N_j = k) = \begin{cases} 1 - p(i \to j) & \text{if } k = 0 \\ p(i \to j) \cdot (p_{i\circlearrowleft})^k (1 - p_{i\circlearrowleft}) & \text{if } k \geq 1. \end{cases}$$

The more general result can be shown by "reachability analysis," which is very similar to first-step analysis. We also note that $N_j = 0$ iff $T_j^+ = \infty$.

Proposition 5 is quite beautiful in its connection between the memorylessness of the Geometric distribution (the only such discrete distribution) and that of the Markov property: every time the particle lands on state i, it independently "decides" to revisit i with probability $p_{i\circlearrowleft}$.

Additionally, it draws a connection between N_i and T_i^+ in $\mathbb{P}_i(T_i^+ < \infty) = 1$ iff $\mathbb{P}_i(N_i = \infty) = 1$, as well as $\mathbb{P}_i(T_i^+ = \infty) > 0$ iff $\mathbb{P}_i(N_i < \infty) = 1$. We will find a similar connection between *number* (of visits) and *time* (between visits) important to the later Poisson processes.

We also remark that Proposition 5 justifies that the definitions of recurrence and transience in terms of $p_{i\circlearrowleft}$ are "correct": *recurrent* means "occurring often or repeatedly," while *transient* means "impermanent; lasting only for a brief amount of time," which are really $N_i = \infty$ or $N_i < \infty$. From the perspective of the wandering particle, transient states "disappear" after a finite amount of time, leaving only recurrent states in sight, so we might rightfully guess some deep connections to long-term behavior.

We must also warn at this point that while the expected time to revisit a transient state is infinite, the hitting times from other states may still be finite: it may be visited from elsewhere in finite time, but not *re*visit itself in finite time.

Before we move onto an optional technical result, we also observe that the long-term fraction of time, which seem very similar to the total number of visits, may be nonetheless "unrelated": even if $N_i = \infty$, the state does not have to take over a positive proportion of time. This is the boundary case of *null* recurrence we will soon see.

Proposition 7 (*).

State i is recurrent iff the sum of its k-step revisit probabilities is infinite:

$$\sum_{n=1}^{\infty} p^{(n)}(i,i) = \infty.$$

Proof. We define the *strict* k-step revisit probabilities to be

$$f^{(n)}(i,i) := \mathbb{P}_i(T_i^+ = n) = \mathbb{P}(X_n = i, X_k \neq i \text{ for } k = 1, \dots, n-1 \mid X_0 = i).$$

We note that $f^{(n)}(i,i) \leq p^{(n)}(i,i)$, and by countable additivity for a union of disjoint events,

$$p_{i\circlearrowleft} = \sum_{n=1}^{\infty} f^{(n)}(i,i).$$

To relate $p^{(n)}(i,i)$ and $f^{(n)}(i,i)$, we observe that the events $\{T_i^+ = k\}$ and $\{X_n = i\}$ together require the chain to revisit state i in $n-k$ time steps. That is,

$$\begin{aligned} p^{(n)}(i,i) = \mathbb{P}_i(X_n = i) &= \sum_{k=1}^{n} \mathbb{P}_i(X_n = i \mid T_i^+ = k) \cdot \mathbb{P}_i(T_i^+ = k) \\ &= \sum_{k=1}^{n} p^{(n-k)}(i,i) \cdot f^{(k)}(i,i). \end{aligned}$$

The recurrence relation now allows us to more directly relate $p_{i\circlearrowleft}$, the sum of $f^{(n)}(i,i)$, to $p^{(n)}(i,i)$:

$$\begin{aligned} \sum_{n=1}^{\infty} p^{(n)}(i,i) &= \lim_{N\to\infty} \sum_{n=1}^{N} \sum_{k=1}^{n} p^{(n-k)}(i,i) \cdot f^{(k)}(i,i) \\ &= \lim_{N\to\infty} \sum_{k=1}^{N} \left(p^{(0)}(i,i) + \sum_{n=k+1}^{N} p^{(n-k)}(i,i) \right) f^{(k)}(i,i) \\ &= \left(\sum_{k=1}^{\infty} f^{(k)}(i,i) \right) \left(1 + \sum_{n=1}^{\infty} p^{(n)}(i,i) \right). \end{aligned}$$

Therefore $p_{i\circlearrowleft} = 1$ iff the given summation is infinite, as claimed. □

In the converse direction, we could have also simply cited the Borel–Cantelli lemma: if the sum of every $p^{(n)}(i,i)$ is finite, then $\mathbb{P}_i(X_n = i \text{ i.o.}) = 0$, which is precisely transience.

Proposition 7 is a somewhat technical result needed to prove the main result below, but we should not lose sight of the important intuition: recurrent states are those the particle returns to almost surely, or those the particle visits infinitely often. If a state is reachable from a recurrent state, it must receive a visit for some nonzero proportion out of the infinitely many revisits, so we expect that state to be recurrent as well — recurrence is contagious.

Theorem 1 (Classification I).

Recurrence and transience are class properties.

Proof. Let $i \in S$ be recurrent, and let $j \in [i]$. As i and j communicate, there exist $k, \ell \in \mathbb{Z}^+$ so that the probabilities $p^{(k)}(i,j)$ and $p^{(\ell)}(j,i)$ are nonzero. Then

$$p^{(\ell+n+k)}(j,j) \geq p^{(\ell)}(j,i) \cdot p^{(n)}(i,i) \cdot p^{(k)}(i,j)$$

as one possible sequence of transitions for revisiting j consists of $j \to i$, then i revisiting itself, and finally $i \to j$. We take the summation over all possible times to revisit, $n \in \mathbb{Z}^+$:

$$\sum_{n=1}^{\infty} p^{(n)}(j,j) \geq \sum_{n=1}^{\infty} p^{(\ell+n+k)}(j,j) = p^{(\ell)}(j,i) \left(\sum_{n=1}^{\infty} p^{(n)}(i,i) \right) p^{(k)}(i,j) = \infty.$$

Therefore if one state in a communicating class is recurrent, every state is recurrent as well, and we say that the class is recurrent. By the contrapositive, the same holds for transience. □

We can thus speak of recurrent and transient *classes* and *chains*. In general, if every state in a chain has a given class property, then we say the chain has the same property as well.

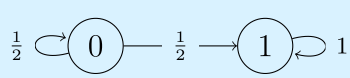
7.2 Graphical results and counterexamples

We just have defined recurrence and transience in terms of the time to revisit T_i^+, found equivalent conditions involving the number of revisits N_i and the k-step revisit probabilities $p^{(k)}(i,i)$, and now characterized them as class properties. With class properties, we can effectively leverage our graphical intuition!

Recall that every transition diagram can be decomposed into a DAG of classes, which are **closed**, isolated or sinks, or **open**, sources or otherwise intermediate nodes. We can imagine that closed classes might be recurrent: the particle stays within the class and cannot leave, and open classes might be transient: any particle, after some finite amount of time, almost surely must make its way to an exit and leave, without any path to return. Even intermediate classes, as in $* \to * \to * \circlearrowleft$, can only be passed through temporarily, and eventually left entirely.

It turns out that this intuition is perfectly correct in the finite-state case! Per usual, the infinite case brings a few minor complications. An infinite closed class may not be recurrent: a particle always has someplace new to visit within the class, without revisiting any past states.

Example 1 (Simple reducible chain).



For instance, find and classify the communicating classes of the chain above. What does a typical realization look like, and what is its stationary distribution?

We are also not limited to the perspective of a single particle: as an ensemble of water particles has mass and flow, the collection of realizations has probability mass supported at each state and flow circulating within classes. If any particle visits any transient state only finitely often, then its proportion of particles must eventually dwindle to zero — transient states cannot support any stationary mass, which indeed turns out to be true.

And, the infinity complication again rears its head, just as it broke the perfect correspondence of recurrence $\leftrightarrow$ closedness, transience $\leftrightarrow$ openness: transient states must have zero mass in a stationary or limiting distribution, but not the converse — recurrent states need not have nonzero mass. We could simply start with zero mass in its class, but infinite recurrent classes can also be *unable* to support a stationary distribution — the boundary case of *null recurrence*. For now,

Proposition 8 (Graphical classification).

a. Any irreducible chain is either recurrent or transient.
b. Any finite-state chain has at least one recurrent state.
c. Any finite-state irreducible chain is recurrent.
d. Every recurrent class is closed, but not the converse.
e. Every finite-state closed class is recurrent.
f. Every finite-state open class is transient.
g. Every transient class is infinite or open.
h. Every state reachable from a recurrent state is recurrent.

Proof. Here we go.

a. Any irreducible chain consists of one communicating class.

b. We will outsource an optional result: if j is transient, then $\lim_{n\to\infty} p^{(n)}(i,j) = 0$ for any i. Now suppose that every state is transient. Then we have contradiction:

$$0 = \lim_{n\to\infty} \sum_{j\in S} p^{(n)}(i,j) = 1.$$

c. Follows from parts a) and b).

d. Suppose that $[i]$ is recurrent but not closed. Then there is some $j \notin [i]$ so that $p(j \to i) = 0$ but $p^{(n)}(i,j) > 0$ for some n. We thus have contradiction — i is transient:

$$1 - p_{i\circlearrowleft} = \mathbb{P}_i(T_i^+ = \infty) \geq \mathbb{P}_i(X_n = j) = p^{(n)}(i,j) > 0.$$

e. We may consider any finite-state closed class C as a standalone finite-state irreducible chain, which does not change any probabilities of eventual revisit. Then this follows from part c).

f. A finite-state class is closed iff recurrent by parts d) and e), i.e. it is open iff transient.

g. Follows from part f).

h. We claim that every state j reachable from a recurrent state i is in its communicating class. Suppose that $i \to j$ but $j \not\to i$. But then $[i]$ is open, so i is transient, a contradiction.

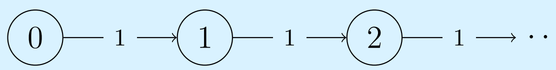
$\square$

See if you can find an algorithm to classify *any* finite-state Markov chain. The adjacency matrix can be obtained by applying $\mathbb{1}_{x \geq 0}$ to the entries of P; then we would only need to decompose the transition diagram and use recurrence $\leftrightarrow$ closedness.

Now, we encourage you to come up with your own counterexamples to the following false claims which we find useful to mention. Again, the following are **counterexamples** to **false** claims.

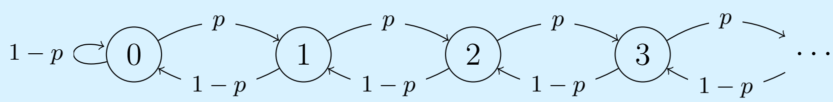
Example 2 (Every chain has a recurrent state?).

Any counterexample must have infinitely many states, so the mass doesn't have to be "somewhere": it can go off to infinity, as in the *arrow of time*, or be dispersed across everywhere.



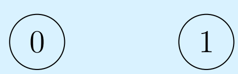
Example 3 (Every closed class is recurrent?).

Consider the irreducible **reflected random walk**, which is transient when $p > 1 - p$:



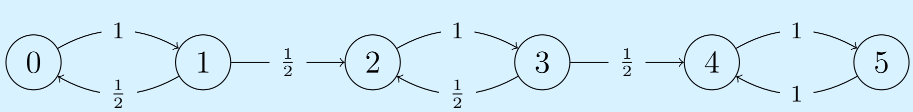
Example 4 (Every finite-state reducible chain is transient?).

The converse of part c in Proposition 8 fails with a very simple counterexample:



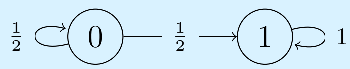
Example 5 (Every state or class with an edge leading into it is recurrent?).

Recurrence implies that no edges lead *out*. Mass can still flow *in*to a transient state, only it cannot stay for long: see state 2 or the intermediate class $\{2, 3\}$ below.



In general, any source class will be transient.

Example 6 (Transience is contagious: every state reachable from a transient state is transient?).



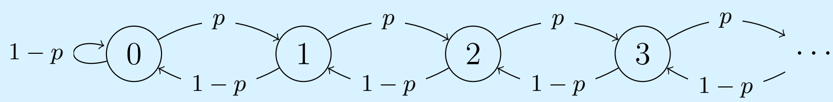
In general, valid reachability relations are recurrent $\to$ recurrent, transient $\to$ recurrent, transient $\to$ transient, but *not* recurrent $\to$ transient.

Take a moment to digest Proposition 8. We have really reconciled quite a few perspectives, both temporal and spatial, both individual and collective: the times between revisits; the total number of visits; the proportion of time visiting; the proportion of particles visiting; the presence of paths or transitions. We now turn to find what happens in infinite time and infinite space.

7.3 Positive recurrence, null recurrence, and transience

Example 7 (Reflected random walk).

Let us reconsider the random walk on $\mathbb{Z}$ which is reflected at the origin 0.



When $p = 0$, this is the "inevitable countdown timer," in which only state 0 is recurrent; when $p = 1$, this is the transient *arrow of time*. For $p \in (0, 1)$,

- When $p > \frac{1}{2}$, the chain is transient: there is a greater tendency to drift right, to infinity.
- When $p < \frac{1}{2}$, the chain is recurrent: the chain will eventually settle at the leftmost 0.
- When $p = \frac{1}{2}$, which defines a **symmetric** random walk, the chain is recurrent: there is no bias towards moving left or right, but there is also no stationary distribution!

A general principle: null recurrence mostly occurs as a *boundary* case.

Definition 6 (Positive recurrence; null recurrence).

Let $\mathbb{E}_i(T_i^+)$ be the expected time to revisit state i. Then

a. i is **positive recurrent** if i is recurrent — $\mathbb{P}_i(T_i^+ = \infty) = 0$, and $\mathbb{E}_i(T_i^+) < \infty$.

b. i is **null recurrent** if i is recurrent — $\mathbb{P}_i(T_i^+ = \infty) = 0$, and $\mathbb{E}_i(T_i^+) = \infty$.

c. i is **transient** if i is not recurrent — $\mathbb{P}_i(T_i^+ = \infty) > 0$.

Theorem 2 (Classification II).

Positive recurrence, null recurrence, and transience are class properties. Moreover, they form a trichotomous classification: every state in S satisfies exactly one out of the three properties.

Proof. We need only show that $\mathbb{E}_i(T_i^+) = \infty$ is a class property. Suppose that it holds for i, and consider any $j \in [i]$. There exist $k, \ell \in \mathbb{Z}^+$ so that $p^{(k)}(i,j) > 0$ and $p^{(\ell)}(j,i) > 0$, so

$$\mathbb{P}_j(T_j^+ \geq t) \geq p^{(\ell)}(j,i) \cdot \mathbb{P}_i(T_i^+ \geq t - \ell - k) \cdot p^{(k)}(i,j).$$

Taking the sum over all possible $t \in \mathbb{Z}^+$, we find that $\mathbb{E}_j(T_j^+) \geq p^{(\ell)}(j,i) \cdot \mathbb{E}_i(T_i^+) \cdot p^{(k)}(i,j) = \infty$ by the tail sum formula. $\square$

Revisiting Example 7, let us now see the exact reasoning behind each case.

Proof for transience for $p > \frac{1}{2}$. Let $Y_n = X_n - X_{n-1}$ for $n \in \mathbb{Z}^+$, so $Y_n = +1$ with probability p and -1 w.p. $1-p$. Then $X_n = 0$ only if $\sum_{i=1}^n Y_i \leq 0$. However, by the SLLN,

$$\frac{1}{n}\sum_{i=1}^n Y_i \xrightarrow{\text{a.s.}} \mathbb{E}(Y_1) = 2p - 1 > 0,$$

so $\sum_{i=1}^n Y_i \leq 0$ only happens finitely often almost surely. Thus state 0 is transient, and as the chain is irreducible, every state is transient. $\square$

Proof of positive recurrence for $p < \frac{1}{2}$. By the later Proposition 11, we can show that a stationary distribution exists. Noticing the graph structure of the chain, we can solve the DBEs to find that $\pi =$ Geometric$(\frac{p}{1-p}) - 1$ is the stationary distribution. We leave the details to you; for simplicity, we often write $q := 1 - p$ and $r := \frac{p}{1-p} < 1$. $\square$

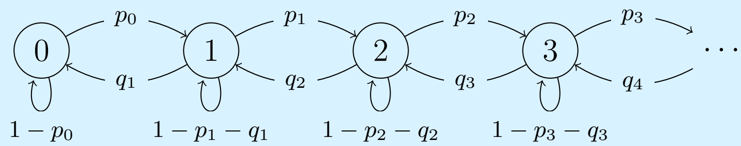
Proof of null recurrence for $p = \frac{1}{2}$. We will first show that state 0 is recurrent:

$$\begin{aligned} p_{0\to 0} &= \frac{1}{2} + \frac{1}{2} \cdot p_{1\to 0} \\ &= \frac{1}{2} + \frac{1}{2} \cdot \left(\frac{1}{2} + \frac{1}{2} \cdot p_{2\to 1} p_{1\to 0}\right) \\ &= \frac{3}{4} + \frac{1}{4} \cdot (p_{1\to 0})^2 \end{aligned}$$

implies that $p_{1\to 0} = 1$, and thus $p_{0\to 0} = 1$. Note that $p_{2\to 0} = p_{2\to 1} p_{1\to 0}$, and by symmetry under translation $p_{2\to 1} = p_{1\to 0}$. $\square$

We will leave two generalizations of the reflected random walk, which find applications in queueing theory and many common models, for you to explore.

Example 8 (Birth-death chain).



Optionally, the birth-death chain above is positive recurrent iff

$$\sum_{n=1}^{\infty}\prod_{i=1}^{n}\frac{p_i}{q_i} = \infty \text{ and } \sum_{n=1}^{\infty}\prod_{i=1}^{n}\frac{q_{i-1}}{p_i} < \infty.$$

It may help to consider the recurrence relations from the detailed balance equations.

Example 9 (d-dimensional symmetric random walk*).

Let $S_n = \sum_{i=1}^{n} X_i$ be a symmetric random walk on $\mathbb{Z}^d$, $d \geq 1$. We jump to every neighbor with equal probability 2^{-d}, so the X_i are i.i.d. Uniform($\{\pm e_1, \ldots, \pm e_d\}$), where $\{e_j\}$ is the standard basis. Then S_n is (null) recurrent when $d = 1$ and $d = 2$, but transient for all $d \geq 3$.

"A drunk man will eventually find his way home, but a drunk bird may get lost forever." — Shizuo Kakutani

Now having refined the notion of recurrence by addressing the somewhat counterintuitive case of null recurrence in infinite space, in which T_i^+ is finite almost surely yet has infinite expectation, we can revisit the correspondence we wished for: positive recurrence $\leftrightarrow$ nonzero stationary mass.

Another general principle: *stationarity is deeply connected to positive recurrence*.

To relate these conditions about T_i^+ and $\pi(i)$, we will intuitively guess that $\pi(i)$ is related to the long-term fraction of time $\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{X_n=i}$, which should depend on the expected *frequency of revisiting* state i: if it takes infinite time between each revisit, then it should have zero probability at stationarity, and the more frequently i is revisited, the more probable it should be. We confirm our intuition below.

Proposition 9 (Stationary mass is inverse of expected time to return).

If π is a stationary distribution, then for all $i \in S$,

$$\pi(i) = \frac{1}{\mathbb{E}_i(T_i^+)}.$$

We let $\frac{1}{\infty} := 0$, as $\mathbb{E}_i(T_i^+) = \infty$ when i is null recurrent or transient.

Proof. We will borrow an idea from *renewal theory*. If $\tau_k = T_i^+(k)$ is the time of the kth revisit to state i, given $X_0 = i$, then the *excursions*

$$(X_0, \dots, X_{\tau_1 - 1}), (X_{\tau_1}, \dots, X_{\tau_2 - 1}), \dots$$

are i.i.d. by the Markov property. If i is also recurrent,then there are infinitely many blocks, each with finite length $\Delta\tau_k$, where $\mathbb{E}(\Delta\tau_k) = \mathbb{E}_i(T_i^+)$. By the SLLN,

$$\frac{t}{\sum_{n=0}^{t-1} \mathbb{1}_{X_n = i}} = \frac{1}{n} \sum_{k=1}^{n} \Delta\tau_k \xrightarrow{\text{a.s.}} \mathbb{E}(\Delta\tau_k) = \mathbb{E}_i(T_i^+).$$

The elapsed time t over the total number of revisits is simply the average length of an excursion, which converges almost surely to the expected time to revisit. By the dominated convergence theorem, when $\mathbb{P}(X_n = i) = \pi(i)$ is stationary, we take the inverse to find

$$\mathbb{E}\left(\frac{1}{t} \sum_{n=0}^{t-1} \mathbb{1}_{X_n = i}\right) = \frac{1}{t} \sum_{n=0}^{t-1} \pi(i) \xrightarrow{\text{a.s.}} \frac{1}{\mathbb{E}_i(T_i^+)}.$$

By the SLLN again, we are done. We have just found an inverse relation between the proportion of particles at $\{X_n = i\}$ when stationary and the expected frequency of revisiting state i. □

Proposition 9 is quite a powerful result: we list a few of its consequences below.

Proposition 10 (Uniqueness of stationary distribution for irreducible chains).

For any irreducible chain, exactly one of the following is true.

a. Every state is positive recurrent, and a unique stationary distribution exists as

$$\pi(i) = \mathbb{E}_i(T_i^+)^{-1}.$$

b. Every state is null recurrent or every state is transient, and no stationary distribution for the chain exists.

Recall above that irreducible chains consist of only one communicating class.

Proposition 11 (Existence of stationary distribution for finite-state chains).

Every finite-state irreducible chain is positive recurrent. Also, an irreducible Markov chain is positive recurrent iff a stationary distribution exists.

It may help to recall Proposition 8. It is impossible for $T_i^+ < \infty$ a.s. yet $\mathbb{E}_i(T_i^+)$ to still be infinite with only finitely many states, so we then return to the simpler recurrence-transience dichotomy: recurrence is positive recurrence in finite chains.

7.4 Period and ergodicity

We have resolved quite a few loose threads of questions raised in the first section on stationarity, and as promised, we will now examine conditions for convergence to the limiting distribution (which must be stationary). We gave two examples of "divergence" and "oscillation": the former did not converge because a stationary distribution did not exist, which we now know is due to null recurrence or transience, and the latter constantly alternated between several distributions. Below, we find that these were in fact the only two possible types of examples!

Definition 7 (Period).

The **period** of a state $i \in S$ is the positive integer

$$d_i := \gcd \left\{ n \in \mathbb{Z}^+ : p^{(n)}(i,i) > 0 \right\}.$$

If there is no length of time $n \in \mathbb{Z}^+$ in which i can be revisited with some nonzero probability, then $d_i := \infty$. A state is **aperiodic** if its period is 1, and *periodic* otherwise.

Definition 8 (Ergodicity*).

A state is **ergodic** if it is positive recurrent and aperiodic.

Theorem 3 (Classification III).

Period and ergodicity are class properties.

Proof. Suppose that i has period $d = d_i$. For $j \in [i]$, there exist nonzero $p^{(k)}(i,j)$ and $p^{(\ell)}(j,i)$ just as in the proof of Theorem 1, and

$$p^{(\ell+n+k)}(j,j) \geq p^{(\ell)}(j,i) \cdot p^{(n)}(i,i) \cdot p^{(k)}(i,j) > 0$$

for any n such that $p^{(n)}(i,i) > 0$. We also have

$$p^{(\ell+k)}(j,j) \geq p^{(\ell)}(j,i) \cdot p^{(k)}(i,j) > 0,$$

so the period of j divides $\ell + n + k$ and $\ell + k$. Then n is also a multiple of d_j, and as the choice of $n : p^{(n)}(i,i) > 0$ is arbitrary, we find that $d_j \mid d_i$. By symmetry, $d_i \mid d_j$, so $d_i = d_j$. □

Proposition 12 (Convergence to stationary distribution).

Suppose that a chain has limiting distribution π, which is the unique stationary distribution. Then the chain converges to π from *all* possible initial distributions iff the chain is aperiodic. In particular, positive recurrent irreducible chains are convergent iff aperiodic.

Convergence in distribution $\pi_n \to \pi$ should hold from all possible π_0, not for *some*, because every chain with π can be trivially started from stationarity $\pi_0 = \pi$, like in the example of oscillation.

The proof of Proposition 12 involves an out-of-scope technique called **coupling**, which also gives quantitative results about the *rate* of convergence, which we will not cover here. Coupling is also used in the analysis of the performance of algorithms, with one famous result being that seven riffle shuffles suffice to produce a well-shuffled deck of cards.

Proposition 13 (Properties of periodicity*).

For an irreducible recurrent chain with period d,

a. There exists a partition of the state space S into d subsets $S_1, \dots, S_d$ that flow into each other cyclically: the only transitions are from states in S_i to states in $S_{(i \bmod d)+1}$.

b. The dth power of the transition matrix, P^d, has d closed communicating classes, and is an irreducible transition probability matrix on each of the classes $S_1, \dots, S_d$.

c. If $d = 1$, for some sufficiently large n, the matrix P^n is **regular**: its entries are strictly positive. If we recall Proposition 8 from note II, we find a proof of aperiodic convergence! $\pi(j) = \lim_{k\to\infty}(P^k)_{i,j}$, so

$$\pi_\infty = \lim_{k\to\infty} \pi_0 P^{k+1} = \left(\lim_{k\to\infty} \pi_0 P^k\right) P = \pi_\infty P$$

must be the limit from any initial distribution π_0.

d. By the **Perron-Frobenius theorem**, all eigenvalues of P are bounded in absolute value by 1. For irreducible aperiodic chains, the eigenspace corresponding to the stationary distribution is 1-dimensional, and all other eigenvalues are less than 1 in magnitude. As $n \to \infty$, the other eigenvalues tend to 0, which shows convergence.

   In the periodic case, there are multiple eigenvalues of P with magnitude 1: the dth roots of unity. This is the essential barrier to convergence for periodic chains.

We will also briefly mention *ergodicity*, the property of a random process or dynamical system to "cover all of space over all of time." A familiar special case of a result from ergodic theory:

Proposition 14 (Ergodic theorem*: long-term fraction of time converges to stationary mass).

Let (S, Σ, π) be a probability space, let P be a *measure-preserving transformation* such that $\pi(P^{-1}(\{i\})) = \pi(\{i\})$, and let $\mathbb{1}_i \colon S \to \mathbb{R}$ be a random variable. Then the **time average** of $\mathbb{1}_i$ is equal to the **space average** of $\mathbb{1}_i$:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_{X_k = i} = \hat{\mathbb{1}}_i(j) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_i(P^k(j)) \overset{*}{=} \frac{1}{\pi(S)} \int_S \mathbb{1}_i \, d\pi = \bar{\mathbb{1}}_i(j) = \pi(i)$$

Before the final section, we remark that the period is simple to find graphically for small chains: we can simply trace any loops and count their lengths. Self-loops automatically imply aperiodicity, as do multiple loops with coprime lengths.

8 Stationarity II

Theorem 4 (Big theorem).

Let $(X_n)_{n\in\mathbb{N}}$ be a S-valued discrete-time Markov chain with countable state space S. Then

a. A stationary distribution π **exists** iff the chain has at least one positive recurrent communicating class, in which case it is given by the inverse of the expected time to revisit. Positive recurrence implies existence, but not the converse.

b. The stationary distribution π is **unique** (if it exists) iff the chain has at most one positive recurrent communicating class. Irreducibility implies uniqueness, but not the converse.

c. In general, any stationary distribution π will be a *convex combination* of the stationary distributions of the positive recurrent communicating classes of the chain (if any).

d. If a stationary distribution π exists and is unique, the chain **converges** to the limiting distribution π from *any* initial distribution π_0 iff the chain is aperiodic. Moreover, the long-term fraction of time in any state converges almost surely to the stationary mass.

Proof. We encourage you to work out these proofs before reading on. All of the necessary results, and hopefully intuition, have been given to you above.

a. If some π exists, then it must be supported on a positive recurrent class by Proposition 11. Conversely, if a positive recurrent class exists, then we have a stationary distribution for the whole chain by setting zero mass elsewhere, as the class must be closed by Proposition 8.

   By Proposition 9, positive recurrence implies existence. A counterexample to the converse is given by a chain with a positive recurrent class and a transient class: π exists with total mass 0 in the transient class, but the whole chain is not positive recurrent.

b. π does not exist if the chain has no positive recurrent classes, and π is uniquely determined by Proposition 9 on the unique positive recurrent class, π being necessarily zero elsewhere. The converse can be shown from part c). We have seen Proposition 10 already, and for a counterexample to its converse, take any reducible chain with one positive recurrent class.

   Recall a previously stated principle. The existence and uniqueness of a stationary distribution are precisely given by the existence and uniqueness of a positive recurrent class!

c. If there are no positive recurrent classes, then no stationary distributions exist. If the chain is irreducible, then a unique stationary distribution exists iff the chain is positive recurrent. For any reducible chain, by Proposition 8 of note I, stationary distributions are closed under convex combination, and for any given π, its convex coefficients are simply the masses it has in each positive recurrent class.

d. We have already seen Proposition 12 and the proof of Proposition 9.

□

There are many, many more generalizations of Markov processes: Harris chains, Markov kernels, hidden Markov models, semi-Markov decision processes, variable-order Markov models, etc. For now, we will keep the state space countable, but turn to Markov chains over *continuous* time.

■