Note 12. Discrete-time Markov chains II

Alex Fu

Fall 2022

4 Stationarity I

Now that we understand general distributions of a Markov chain given the initial distribution π_0 and one-step transition probabilities $p(i, j)$, let us consider special distributions — ones invariant over time, which are deeply connected to the long-term asymptotic behavior of Markov chains.

Definition 1 (Stationary random process).

Let $(X_n)_{n \in \mathbb{N}}$ be a S-valued discrete-time random process. The process is **stationary**, or *at stationarity*, if for every time step $n \in \mathbb{N}$, time shift $k \in \mathbb{N}$, and states $x_0, \ldots, x_n \in S$,

$$\mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) = \mathbb{P}(X_k = x_0, \ldots, X_{n+k} = x_k).$$

A stationary random process does not imply that X_n are invariant as random variables: the values of $X_0, \ldots, X_n$ may change, but their *distribution* is fixed under translation in time.

In particular to Markov chains, stationarity has several seemingly weaker but equivalent conditions, which allow us to more easily check for stationarity. Specifically, Markov chains at stationarity are characterized by their stationary distributions.

Definition 2 (Stationary distribution).

A **stationary distribution**, or *invariant*, *steady-state*, or *equilibrium distribution* of a Markov chain $(X_n)_{n \in \mathbb{N}}$ with transition probability matrix P is a distribution π such that

$$\pi = \pi P.$$

We will also find it useful to introduce the following related definition here, as it describes possible candidates to check for the existence of stationary distributions.

Definition 3 (Invariant measure).

An **invariant measure** or *stationary measure* of a Markov chain is a nonnegative measure $\mu \colon S \to [0, \infty)$, represented as a row vector, such that $\mu = \mu P$.

Every stationary distribution is an invariant measure, but not the converse: the entries of μ do not have to sum to 1, i.e. μ may not be a *probability* distribution.

Proposition 1.

Stationarity implies time-homogeneity.

Proof. Recall that a chain is time-homogeneous if $\mathbb{P}(X_{k+2} = j \mid X_{k+1} = i) = \mathbb{P}(X_{k+1} = j \mid X_k = i)$. Now, for a chain at stationarity, $\pi_0 = \pi_k$ for every $k \in \mathbb{N}$ by Definition 1 for $n = 0$. Then, by choosing $n = 1$,

$$\begin{aligned}
\mathbb{P}(X_{k+2} = j \mid X_{k+1} = i) &= \frac{1}{\pi_{k+1}(i)} \cdot \mathbb{P}(X_{k+2} = j, X_{k+1} = i) \\
&= \frac{1}{\pi_k(i)} \cdot \mathbb{P}(X_{k+1} = j, X_k = i) \\
&= \mathbb{P}(X_{k+1} = j \mid X_k = i).
\end{aligned}$$

□

While Definition 1 requires the invariance of every finite-dimensional joint distribution $p_{X_0,\ldots,X_n}$, for processes with the Markov property, invariance of the marginal distributions π_n is sufficient.

Proposition 2.

A Markov chain is at stationarity iff $\pi_0 = \pi_n$ for all $n \in \mathbb{N}$.

Proof. The forward direction follows from Definition 1, so we will show the converse. If $\pi_0 = \pi_k$ for all $k \in \mathbb{N}$, by the Markov property with time-homogeneity,

$$\begin{aligned}
\mathbb{P}\big(X^{(0:n)} = x^{(0:n)}\big) &= \pi_0(x_0) \prod_{i=0}^{n-1} p(x_i, x_{i+1}) \\
&= \pi_k(x_0) \prod_{i=0}^{n-1} p(x_i, x_{i+1}) \\
&= \mathbb{P}\big(X^{(k:n+k)} = x^{(k:n+k)}\big)
\end{aligned}$$

for any $n \in \mathbb{N}$ and any sequence of states $x_0, \ldots, x_k$, even those with zero probability. □

Proposition 3 (Chain is stationary iff initial distribution is stationary).

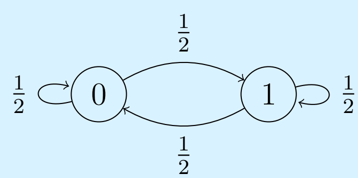
A Markov chain is stationary iff $\pi_n = \pi_{n+1}$ for all $n \in \mathbb{N}$. Equivalently, a chain is stationary iff it is started at a stationary distribution $\pi_0 = \pi_0 P = \pi_1$.

Proof. We observe that $\pi_0 = \pi_n$ and $\pi_n = \pi_{n+1}$ for all $n \in \mathbb{N}$ are equivalent conditions. By the equations $\pi_{n+1} = \pi_n P$ or $\pi_n = \pi_0 P^n$, we find that $\pi_0 = \pi_1$ is equivalent as well. □

Note that $\pi_n = \pi_{n+1}$ for *some* $n \in \mathbb{N}$ does not imply the chain is stationary, though it does imply the truncated chain $(X_{n+k})_{k\in\mathbb{N}}$ is stationary, as the *distribution* π_n is stationary. A chain is stationary iff its distribution is stationary at all times. Henceforth, we will only consider stationary distributions, as this distinction is negligible.

Example 1 (Reaching a stationary distribution from a different initial distribution).

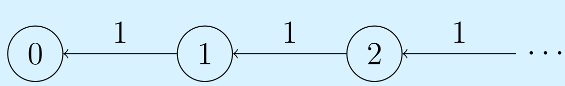
Consider the following chain. If $\pi_0 = [1, 0]$, then $\pi_1 = [\frac{1}{2}, \frac{1}{2}]$ is the stationary distribution.



In fact, we can show that any π_0 will converge to the stationary distribution $[\frac{1}{2}, \frac{1}{2}]$ in one step, which is not true in general. Using intuition about sequences of real numbers, we might guess that sequences of distributions $(\pi_n)_{n\in\mathbb{N}}$ can also converge in finitely many steps, converge eventually, or oscillate forever, though not diverge to infinity, as distributions have total mass 1.

Example 2 (Eventual convergence to stationary distribution).

In the following chain, if $\pi_0(i) \neq 0$ for only finitely many states i, then the chain will reach the stationary distribution $[1, 0, 0, \ldots]$ in finite time. If $\pi_0(i) = 2^{-i}$, then π_n converges to it eventually, though not in any finite number of time steps.



Definition 4 (Limiting distribution).

The **limiting distribution** of a Markov chain is the probability distribution

$$\pi_\infty = \lim_{n\to\infty} \pi_n$$

if it exists. Equivalently, if $X_n \xrightarrow{d} X_\infty$ describes convergence in distribution, then $X_\infty \sim \pi_\infty$. The term *steady-state distribution* is also sometimes used, though we will reserve this term for stationary distributions.

Proposition 4 (Limiting distributions are stationary).

The limiting distribution may not exist, but when it does, it is a stationary distribution. The converse is not true: not every stationary distribution is limiting.

Proof. If π_∞ exists, then its entries are bounded in $[0, 1]$, so we can write

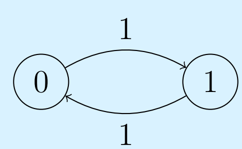
$$\pi_\infty = \lim_{n\to\infty} \pi_{n+1} = \lim_{n\to\infty} \pi_0 P^{n+1} = \left(\lim_{n\to\infty} \pi_0 P^n\right) P = \pi_\infty P.$$

The limiting distribution is unique as a limit, but the stationary distribution may not be unique as we will soon see, so not every stationary distribution is limiting. □

We will return to the question of when the limiting distribution exists, or when the Markov chain converges, in part III. For now, we will look at two examples of divergence.

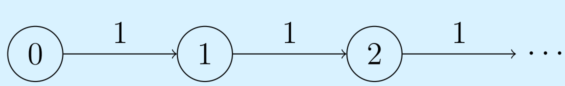
Example 3 (Infinite oscillation between distributions).

For any $\pi_0 \neq [\frac{1}{2}, \frac{1}{2}]$, the following chain will oscillate between $[p, 1-p]$ and $[1-p, p]$.



Example 4 (Divergence to infinity, or the "arrow of time").

The distributions π_n cannot "diverge to infinity," but the random variables X_n can when the state space S is infinite, as in the following chain for any π_0.



We might now ask how to find stationary distributions explicitly. We can look at the equivalent condition of $\pi_0 = \pi_1$ from two familiar perspectives: as the balancing or conservation of flow, or as eigenvectors of the linear stochastic system $\pi = \pi P$.

The perspective of graph properties of the transition diagrams will prove useful later in conditions for the existence of, uniqueness of, and convergence to a stationary distribution.

Proposition 5 (Conservation of flow).

π is a stationary distribution iff *flow-in equals flow-out* for every state j, or iff π satisfies the **(global) balance equations**:

$$\sum_{i \in S} \pi(i) \cdot p(i,j) \overset{*}{=} \pi(j) = \sum_{k \in S} \pi(j) \cdot p(j,k) \qquad \forall j \in S.$$

Recall that mass always equals flow-out, so the GBEs are the starred equations for *flow-in equals mass*. To show Proposition 5, we can simply expand $\pi = \pi P$:

$$(\pi P)(j) = \pi \cdot \mathsf{col}_j(P) = \sum_{i \in S} \pi(i) \cdot p(i,j) = \pi(j).$$

Stationary distributions thus have a very intuitive interpretation: if the amount of flow in always equals the amount of flow out, then the mass is conserved at each state, which is precisely the condition that $\mathbb{P}(X_n = i) = \mathbb{P}(X_{n+1} = i)$.

When the state space S is finite, the global balance equations give a collection of $|S| + 1$ many linear equations along with the normalization condition $\sum_{i \in S} \pi(i) = 1$, making one equation redundant. The system can be solved for small $|S|$ by Gaussian elimination for instance, but it is computationally intractable to solve in general.

As a bit of foreshadowing for CTMCs, by removing self-loops, the GBEs are the same as

$$\pi(j) \sum_{i \neq j} p(i,j) = \sum_{k \neq j} \pi(j) \cdot p(j,k) \qquad \forall j \in S.$$

The global balance equations amount to solving the full linear system given by $\pi = \pi P$, but with the observation that π must be an eigenvector of P, we have access to more efficient methods of finding stationary distributions, as well as a partial guarantee of existence:

Proposition 6 (Existence of invariant measure*).

Every row-stochastic matrix P has a row eigenvector μ of eigenvalue 1.

Proof. The left and right eigenvalues of any square matrix are the same:

$$\det(A - \lambda I) = \det((A - \lambda I)^{\mathsf{T}}) = \det(A^{\mathsf{T}} - \lambda I).$$

Then, every stochastic matrix has a right- or column- eigenvector with right eigenvalue 1:

$$P\mathbf{1} = \mathbf{1},$$

the same as every row of P summing to 1, where $\mathbf{1}$ is the vector with all entries 1. ☐

The sum of entries of an invariant measure μ, represented as a row vector with $|S|$ many nonnegative entries, is also called its ℓ^1-*norm* $\|\mu\|_1$. Eigenvectors are closed under linear combination: if μ is a row eigenvector of eigenvalue 1, then so is any $c\mu$, including $0\mu = 0$. In particular, any $0 < \|\mu\|_1 < \infty$ can be normalized to obtain a stationary distribution $\pi = \frac{\mu}{\|\mu\|_1}$. The converse gives us another method of checking that a stationary distribution does not exist:

Proposition 7 (No stationary distribution without invariant measure of finite nonzero norm).

If every eigenvector of P with eigenvalue 1 has zero or infinite ℓ^1-norm, then the chain with transition probability matrix P does not have a stationary distribution.

For example, we can show that Example 4 fails to converge to a limiting distribution because no *stationary* distribution exists at all. Any solution to the GBEs μ must be a uniform distribution by symmetry, but either $\mu(0) = 0$ or $\mu(0) = c > 0$, so

$$\sum_{i=0}^{\infty} \mu(0) = 0 \text{ or } \infty,$$

so no stationary distribution exists. Interestingly, this is precisely the same proof that there is no uniform distribution over any countable space!

Lastly, we can explicitly find the stationary distribution-eigenvector in a special case.

Proposition 8 (k-step transition probabilities converge to stationary mass).

If the transition probability matrix P has strictly positive entries $P_{i,j} > 0$, then P has a unique stationary distribution π such that for *any* (row or initial state) $i \in S$,

$$\pi(j) = \lim_{k\to\infty} (P^k)_{i,j} = \lim_{k\to\infty} p^{(k)}(i,j).$$

Proof. Let $Q = \lim_{k\to\infty} P^k$. Then $Q = QP$, so for every $i \in S$,

$$\text{row}_i(Q) = \text{row}_i(Q)P.$$

The entries of P^k are k-step transition probabilities $p^{(k)}(i,j) \in [0,1]$, so the entries of Q belong in $[0,1]$ as well, and $\text{row}_i(Q)$ is a stationary distribution. The condition that every $p(i,j) > 0$ ensures that no row of Q sums to 0. □

One final remark: we have been "choosing different distributions for the same chain" throughout this section, but the distributions $(\pi_n)_{n\in\mathbb{N}}$ are determined by the chain $(X_n)_{n\in\mathbb{N}}$, so we have been informal with our terminology in referring to the *transition probabilities* as the chain. (The Markov property says that the chain is determined by the initial distribution and transition probabilities.)

5 Reversibility

Let us now recall the reversed chain. Stationarity gives the invariance of every finite-dimensional joint distribution under translation in time, but what about invariance under reflection in time? That is, $(X_0, \ldots, X_n) \overset{d}{=} (X_n, \ldots, X_0)$. We find that this defines *reversibility*, a stronger condition than even stationarity, in which a chain is indistinguishable run forwards or backwards.

Proposition 9 (Reversed stationary chain is stationary).

Let $(X_n)_{n\in\mathbb{N}}$ be a stationary chain. Then the reversed chain $(Y_n)_{n=0}^N$ is time-homogeneous, and moreover itself a stationary chain with the same stationary distribution.

Proof. Recall the reverse transition probabilities $\tilde{p}(i,j)$, which we want to show are well-defined. By the stationarity and time-homogeneity of the forward chain $\{X_n\}$,

$$\begin{aligned}\mathbb{P}(Y_{n+1} = j \mid Y_n = i) &= \mathbb{P}(X_{N-n-1} = j \mid X_{N-n} = i) \\ &= \frac{\pi(j) \cdot p(j,i)}{\pi(i)} \\ &= \mathbb{P}(Y_n = j \mid Y_{n-1} = i).\end{aligned}$$

We also find a formula for $\tilde{p}$ in terms of π and p. By rearranging,

$$\sum_{i\in S} \pi(i) \cdot \tilde{p}(i,j) = \sum_{i\in S} \pi(j) \cdot p(j,i) = \pi(j),$$

so π also satisfies the GBEs for the transition probabilities $\tilde{p}(i,j)$. Alternatively, $\{Y_n\}$ is stationary as $\mathbb{P}(Y_n = i) = \mathbb{P}(X_{N-n} = i) = \pi(i)$ for every $i \in S$ and $0 \le n \le N$. □

Definition 5 (Reversibility).

A Markov chain is **reversible** if the forward transition probabilities are the same as the reverse transition probabilities: for every pair of states $i, j \in S$,

$$\tilde{p}(i,j) = p(i,j).$$

Proposition 10 (Detailed balance equations).

An equivalent condition for reversibility is the **detailed balance equations** (DBEs):

$$\pi(i) \cdot p(i,j) = \pi(j) \cdot p(j,i) \qquad \forall i,j \in S.$$

The flow from i to j is equal to the flow from j to i; the flows across every edge are balanced. This is also known as *local balance*, in contrast to the *global balance* of stationarity.

This pairwise equilibrium removes the directionality of time, as if bringing time to a standstill.

Proposition 11 (Reversibility implies stationarity).

Reversibility implies stationarity, but not the converse. Local balance implies global balance.

Proof. Suppose that reversibility holds for the distribution π. Then for any $j \in S$,

$$\pi(j) = \sum_{i \in S} \pi(j) \cdot p(j,i) \overset{*}{=} \sum_{i \in S} \pi(i) \cdot p(i,j).$$

□

Solving the detailed balance equations to find a stationary distribution turns out to be easier than solving the global balance equations usually, even though the DBEs may not have a solution at all! A clock, or 12-cycle, is a counterexample to the converse: its uniform stationary distribution is not reversible. Intuitively, any particle or realization can "tell the direction of time."

Proposition 12 (Forward chain and reverse chain are equal in distribution when reversible).

Let $\{X_n\}$ be a reversible Markov chain started at stationarity $\pi_0 = \pi$. Then

$$(X_0, \ldots, X_N) \overset{\mathsf{d}}{=} (X_N, \ldots, X_0).$$

Proof. We use both the forwards and backwards Markov property below.

$$\begin{aligned} \mathbb{P}\big(X^{(0:N)} = x^{(0:N)}\big) &= \pi(x_N) \prod_{i=0}^{N-1} p(x_{i+1}, x_i) \\ &= \mathbb{P}(X_0 = x_N) \prod_{i=0}^{N-1} \mathbb{P}(X_{N-i} = x_i \mid X_{N-i-1} = x_{i+1}) \\ &= \mathbb{P}\big(X^{(N:0)} = x^{(0:N)}\big). \end{aligned}$$

□

We may also suspect a graphical condition for reversibility, one involving *undirected* edges.

Definition 6 (Graph structure).

The **graph structure** of a transition diagram (S, E) is the undirected graph

$$(S, (E \cup E^{\mathsf{op}}) \setminus \{(i,i) \mid i \in S\}),$$

in which the directions of edges and self-loops are removed from the edge set E.

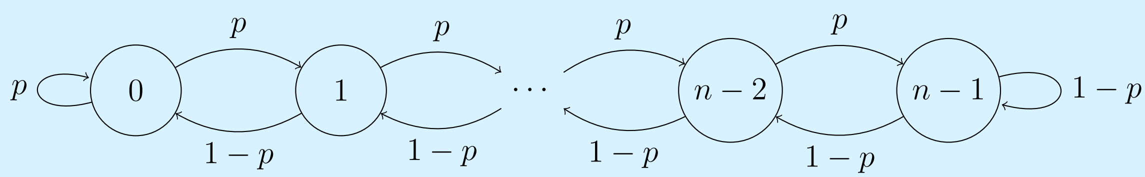
Note that self-loops have no effects on flow-in versus flow-out relations between pairs of states, so they trivially satisfy detailed balance.

Proposition 13 (Tree structure implies reversibility).

If the graph structure of a finite-state irreducible Markov chain is a *tree*, then the stationary distribution of the Markov chain satisfies detailed balance.

The proof is left as an exercise — the property that any pair of vertices in a tree is connected by at most one edge, along with Proposition 11, may help. In particular, chains that resemble lines also satisfy detailed balance, such as the example below.

Example 5 (Finite birth-death chain).



Example 6 (Bernoulli process).

The length of the queue in a discrete-time **Bernoulli process** can be modeled as a reversible chain: a customer arrives with probability p at every time step $n \in \mathbb{N}$, and a customer in the queue is served with probability q independently. When the queue is run backwards in time, customer departures become arrivals — the departure process is also a Bernoulli process!

Another way of framing the conservation of flow is as flow-in equalling flow-out *across* collections of edges: for global balance, it is the edges between a single state and all other states. A stronger result in the same vein gives us another perspective on Proposition 13:

Proposition 14 (Cut property).

For any irreducible Markov chain at stationarity, flow-in equals flow-out holds across any *cut* $(T, S \setminus T)$ of the transition diagram:

$$\sum_{i \in T} \sum_{j \in S \setminus T} \pi(i) \cdot p(i,j) = \sum_{i \in T} \sum_{j \in S \setminus T} \pi(j) \cdot p(j,i).$$

Proof. Global balance gives the equality of the summations over $j \in S$:

$$\sum_{i \in T} \sum_{j \in S} \pi(i) \cdot p(i,j) - \sum_{i,j \in T} \pi(i) \cdot p(i,j) = \sum_{i \in T} \sum_{j \in S} \pi(j) \cdot p(j,i) - \sum_{j,i \in T} \pi(j) \cdot p(j,i).$$

□

We conclude with an optional result: a necessary and sufficient graphical condition for reversibility, which almost vacuously implies Proposition 13.

Proposition 15 (Kolmogorov cycle criterion*).

An irreducible, positive recurrent, aperiodic Markov chain with transition probability matrix P is reversible iff for every $i_1, \ldots, i_n \in S$,

$$p(i_1, i_2) \cdot p(i_2, i_3) \cdots p(i_n, i_1) = p(i_1, i_n) \cdots p(i_3, i_2) \cdot p(i_2, i_1).$$

The probability of traversing any cycle in the transition diagram is the same in both directions.

Proof. We will show the converse. If the cycle condition holds, the probability of traversing any particular sequence $i = x_0, x_1, \ldots, x_n, x_{n+1} = j$ given x_0 is

$$\begin{aligned}
\mathbb{P}\big(X^{(n+1:1)} = x^{(n+1:1)} \mid X_0 = x_0\big) &= \frac{1}{p(x_{n+1}, x_0)} \left[p(x_{n+1}, x_0) \prod_{i=0}^{n} p(x_i, x_{i+1}) \right] \\
&\overset{*}{=} \frac{1}{p(x_{n+1}, x_0)} \left[p(x_0, x_{n+1}) \prod_{i=0}^{n} p(x_{i+1}, x_i) \right] \\
&= \frac{p(x_0, x_{n+1})}{p(x_{n+1}, x_0)} \cdot \mathbb{P}\big(X^{(n+1:1)} = x^{(0:n)} \mid X_0 = x_{n+1}\big).
\end{aligned}$$

We take the sum over all possible sequences $x_1, \ldots, x_n \in S$, then take the limit as $n \to \infty$ as Proposition 8 applies to P by assumption:

$$\mathbb{P}(X_{n+1} = j \mid X_0 = i) = p^{(n+1)}(i, j) \to \pi(j)$$

$$\frac{p(i, j)}{p(j, i)} \cdot \mathbb{P}(X_{n+1} = i \mid X_0 = j) = \frac{p(i, j)}{p(j, i)} \cdot p^{(n+1)}(j, i) \to \frac{p(i, j)}{p(j, i)} \pi(i).$$

□

Example 7 (Metropolis–Hastings algorithm).

Monte Carlo Markov chain (MCMC) algorithms allow us to sample even computationally intractable probability distributions π by designing a chain with stationary distribution π, from which samples at stationarity approximate π.

The **Metropolis–Hastings algorithm** is an important MCMC algorithm that only requires information about π up to a constant factor, though the factor is important for tractability, but crucially always produces a *reversible* chain. Continued in part III.

■