Lecture 9 - Stochastic Processes & Markov Chains

Stochastic Process (SP): A SP in discrete time $t \in \{0, 1, 2, \ldots\}$ is a sequence of time-indexed RV's $\underbrace{X_0}_{\text{RV at time}=0}, \underbrace{X_1}_{\text{RV at time } t=1 \text{ etc.}}, \ldots$ with $\mathcal{X} = \{X_t, t \geq 0\}$

- If t is continuous then cts SP
- " " " discrete " discrete " .

Markov Chains (MC)
→ Discrete Time MC (DTMC)
→ Cts " MC (CTMC)

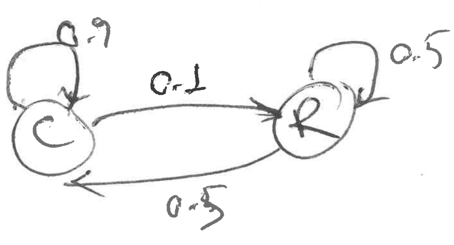
Markovian Property: The transition to the next state depends only on the current state.

Ex: Tracking Daily weather $\mathcal{X} = \{\underbrace{\text{Clear}}_{C}, \underbrace{\text{Rainy}}_{R}\}$

today $C \xrightarrow{0.9} C$ tomorrow $\qquad R \xrightarrow{0.5} R$

$C \xrightarrow{0.1} R \qquad R \xrightarrow{0.5} C$

Assumption: a day's weather depends only on the previous day



Markov Chain

Distribution of escaping a state is Geo(p).?

A Stochastic Process with Markovian property is called a Markov Chain.

Discrete Time Markov Chain (DTMC): A SP

$\mathcal{X} = \{X_t, t \geq 0, t \text{ discrete}\}$ is a DTMC if for all t

$$P(X_{t+1} = j \mid X_t = i, X_{t-1} = a, X_{t-2} = b, \dots) =$$
$$P(X_{t+1} = j \mid X_t = i) = P_{ij} \leftarrow \text{one-step transition probability (from } i \text{ to } j\text{)}$$

Markovian Property: (recall) A transition to state $t+1$, X_{t+1}, only depends on the current state X_t.

$$P_{ij} = P(\underset{\text{e.g. today}}{i} \longrightarrow \underset{\text{tomorrow}}{j})$$

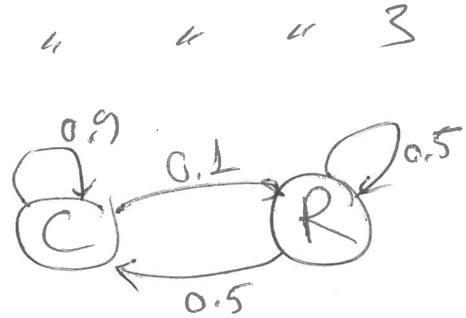
Ex: Recall daily weather $\mathcal{X} = \{\overset{\text{clear}}{C}, \overset{\text{rainy}}{R}\}$

X_1: weather on day 1 $\in \mathcal{X}$

X_2: " " " 2 $\in \mathcal{X}$

X_3: " " " 3 $\in \mathcal{X}$

⋮



e.g. $\underset{\text{clear today}}{C} \xrightarrow{0.9} \underset{\text{clear tomorrow}}{C}$

$P_{CC} = 0.9$ $\quad P_{CR} = 0.1$

$P_{RC} = 0.5$ $\quad P_{RR} = 0.5$ $\quad \leftarrow P_{ij} = P(i \rightarrow j)$

$$P = \begin{array}{c|cc} & C & R \\ \hline C & P_{CC} & P_{CR} \\ R & P_{RC} & P_{RR} \end{array} = \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix} \leftarrow \text{transition matrix}$$

$P_{ij} = P(\text{row } i \to \text{col } j)$
$= P(i \to j)$

rows of P sum to 1:

$$P_{CC} + P_{CR} = 0.9 + 0.1 = 1$$
$$P_{RC} + P_{RR} = 0.5 + 0.5 = 1$$

In general $\sum_j P_{ij} = 1 \Rightarrow \sum_j P(i \to j) = P(i \to \text{any})$

(P_{ij}^2) : Prob ($i \to j$ in 2 steps)

$$(P_{ij}^2) = \begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix} \times \begin{bmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{bmatrix} = \begin{bmatrix} (P_{00}^2) & (P_{01}^2) \\ (P_{10}^2) & (P_{11}^2) \end{bmatrix} = P^2$$

e.g. $(P_{00}^2) = \underline{P_{00}} \cdot \underline{P_{00}} + \underline{P_{01}}\, \underline{P_{10}}$

In general (P_{ij}^n) = Prob($i \to j$ in n steps)
(path does not matter

$= P^n \leftarrow$ multiply P with itself n times.

If there are $k+1$ states: $0, 1, \dots, k$

$$P^n = \begin{bmatrix} (P_{00}^n) & (P_{01}^n) & \cdots & (P_{0k}^n) \\ \vdots & & & \\ (P^n)_{k0} & (P^n)_{k1} & \cdots & (P_{kk}^n) \end{bmatrix}$$

Ex: $P^2 = \begin{bmatrix} 0.86 & 0.14 \\ 0.70 & 0.30 \end{bmatrix}$, $P^5 = \begin{bmatrix} 0.83504 & 0.16496 \\ 0.82480 & 0.17520 \end{bmatrix}$

$P^{10} = \begin{bmatrix} 0.8333508 & 0.1666492 \\ 0.8332460 & 0.1667540 \end{bmatrix} \cdots P^{100} = \begin{bmatrix} 0.8\overline{3} & 0.1\overline{6} \\ 0.8\overline{3} & 0.1\overline{6} \end{bmatrix}$

Ex: $\mathcal{X} = \{C, S, R\}$
clear snowy rainy

$C \xrightarrow{0.7} C$ $\quad S \xrightarrow{0.6} C$ $\quad R \xrightarrow{0.4} C$

$C \xrightarrow{0.2} S$ $\quad S \xrightarrow{0.3} S$ $\quad R \xrightarrow{0.4} S$

$C \xrightarrow{0.1} R$ $\quad S \xrightarrow{0.1} R$ $\quad R \xrightarrow{0.2} R$

$$P = \begin{array}{c|c} & \begin{array}{ccc} C & S & R \end{array} \\ \begin{array}{c} C \\ S \\ R \end{array} & \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.6 & 0.3 & 0.1 \\ 0.4 & 0.4 & 0.2 \end{bmatrix} \end{array} \longrightarrow P^5 = \begin{bmatrix} 0.64195 & 0.24690 & 0.11111 \\ 0.64198 & 0.24691 & 0.111111 \\ 0.64188 & 0.24700 & 0.11112 \end{bmatrix}$$

$$\rightarrow P^{10} = \begin{bmatrix} 0.6419753 & 0.2469136 & 0.\overline{1} \\ \text{''} & \text{''} & \text{''} \\ \text{''} & \text{''} & \text{''} \end{bmatrix}$$

In general, for large n,

$$P^n = \begin{bmatrix} \pi_1 & \pi_2 & \cdots & \pi_k \\ \pi_1 & \pi_2 & & \pi_k \\ \vdots & \vdots & & \\ \pi_1 & \pi_2 & & \pi_k \end{bmatrix}$$

sum = 1
sum = 1
sum = 1

Each row is the same

i.e. Prob of ending up at state j after sufficiently large # of steps does not depend on the initial state or first few states

$\lim_{n\to\infty} (P_{ij}^n) = \pi_j \leftarrow$ does not depend on i

$\sum_j \pi_j = 1$, so $\vec{\pi} = (\pi_1, \ldots, \pi_k)$ is a pmf.

$\pi_j = \lim_{n\to\infty} (P_{ij}^n) = \lim_{n\to\infty} P(i \longrightarrow j \text{ in } n \text{ steps})$

= long term prob. of being in state j

= <u>limiting</u> prob of being in state j

$\vec{\pi}$ is called <u>limiting distr.</u> or <u>stationary distr.</u>

<u>Goal:</u> How to find $\vec{\pi}$?

<u>Application:</u> mean or average state (in the long run)

$$= \sum_{i=1}^{k} i\pi_i$$

Recall that for large n — each row is vector $\vec{\pi}$

$$P^{n+1} = P^n = \begin{pmatrix} \vec{\pi} \longrightarrow \\ \vec{\pi} \longrightarrow \\ \vdots \\ \vec{\pi} \longrightarrow \end{pmatrix}$$

$$\Rightarrow \quad P^n P = \begin{pmatrix} \vec{\pi} \longrightarrow \\ \vec{\pi} \longrightarrow \\ \vec{\pi} \longrightarrow \end{pmatrix}$$

$$\Rightarrow \quad \begin{pmatrix} \vec{\pi} \longrightarrow \\ \vec{\pi} \longrightarrow \\ \vdots \\ \vec{\pi} \longrightarrow \end{pmatrix} P = \begin{pmatrix} \vec{\pi} \longrightarrow \\ \vec{\pi} \longrightarrow \\ \vdots \\ \vec{\pi} \longrightarrow \end{pmatrix}$$

$$\Rightarrow \quad \boxed{\vec{\pi} P = \vec{\pi}}$$

<u>Ex:</u> $\mathcal{X} = \{C, R\}$ $\quad \vec{\pi} = (\pi_C, \pi_R)$

$$(\pi_C \;\; \pi_R) \begin{pmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{pmatrix} = (\pi_C \;\; \pi_R)$$

both eqns yield this

$$\left.\begin{array}{l} \Rightarrow 0.9\pi_C + 0.5\pi_R = \pi_C \\ 0.1\pi_C + 0.5\pi_R = \pi_R \end{array}\right\} \Rightarrow 0.5\pi_R = 0.1\pi_C$$

$$\Rightarrow 5\pi_R = \pi_C$$

& $\pi_C + \pi_R = 1$ (since $\vec{\pi}$ is a pmf)

$$\Rightarrow 5\pi_R + \pi_R = 1 \Rightarrow \pi_R = 1/6$$

$$\rightarrow \pi_C = 5(1/6) = 5/6$$

$$\Rightarrow \vec{\pi} = \left(5/6 \quad 1/6\right)$$

<u>Recap:</u> $\vec{\pi} P = \vec{\pi}$ The distribution $\vec{\pi}$ that solves this equation is called the <u>stationary distr.</u> of that <u>M.C.</u>

<u>Thm:</u> If limiting distr. exists, it is also the <u>unique</u> stationary distr.

<u>Ex:</u> $\pi_R = 1/6$ $\quad P(\text{rains on day } 400) = 1/6$

$P(\quad " \quad " \quad " \quad 401) = 1/6$

but $P(R \text{ on } 401 \mid R \text{ on } 400) = P_{RR} = 0.5$

check: $P(R \text{ on } 401) = P(R \text{ on } 401 \mid R \text{ on } 400)P(R \text{ on } 400)$
$+ P(R \text{ on } 401 \mid C \text{ on } 400)P(C \text{ on } 400)$

$$= 0.5(1/6) + 0.1(5/6) = 2 \times \frac{0.5}{6}$$

$$= 1/6$$

<u>Ex:</u> Find $\vec{\pi}$ for $\mathcal{X} = \{C, S, R\}$ example.
(left as exercise)

Ex:

$$R \xrightarrow{0.5} RR \xrightarrow{0.5} RRR$$

day: 1   1 2   1 2 3

$$R \xrightarrow{0.5} RC \xrightarrow{0.1} RCR$$

prob only depends on the previous state (ordering)

Ex: <u>Umbrella Problem</u>: Prof X commutes between Home & office.

Suppose Prof X has 2 umbrellas.
If it rains, he takes one with him (if available)
prob (rain) = p & prob (no rain) = $1-p$

Let $i = 0, 1, 2$ be the # of umbrellas he has @ home or @ office in a given day.

Qu: (i) Prob of he gets wet in the long run? (ii) long term prob of ending up with zero umbrella

# of Umbrellas @ office

$$P = \begin{array}{c|ccc} & 0 & 1 & 2 \\ \hline 0 & 0 & 0 & 1 \\ 1 & 0 & 1-p & p \\ 2 & 1-p & p & 0 \end{array}$$

# of Umbrellas @ home

$$\vec{\pi} = (\pi_0, \pi_1, \pi_2)$$

$$\vec{\pi} P = \vec{\pi}$$

$$\Rightarrow (\pi_0 \ \ \pi_1 \ \ \pi_2) \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1-p & p \\ 1-p & p & 0 \end{pmatrix} = (\pi_0 \ \ \pi_1 \ \ \pi_2)$$

$$\Rightarrow \pi_2(1-p) = \pi_0$$

$$(1-p)\pi_1 + p\pi_2 = \pi_1 \rightarrow \pi_1 = \pi_2$$

$$\pi_0 + p\pi_1 = \pi_2$$

$$\rightarrow \pi_2(1-p) + p\pi_2 = \pi_2$$

$$\Rightarrow \pi_2 = \pi_1 = \frac{1}{3-p}$$

(ii) $\pi_0 = \dfrac{1-p}{3-p}$

(i) $P(\text{wet}) = \pi_0 \cdot p = \dfrac{(1-p)p}{3-p}$.

Random Walk on integers starting at 0

$$\cdots \quad -2 \quad -1 \quad 0 \quad 1 \quad 2 \quad \cdots$$

$$\text{Prob(left)} = \text{Prob(right)} = \frac{1}{2}$$

$$\begin{tikzcd}
\cdots & -2 & -1 \arrow[l, bend left, "1/2"] \arrow[r, bend left, "1/2"] & 0 \arrow[l, bend left, "1/2"] \arrow[r, bend right, "1/2"'] & 1 \arrow[l, bend right, "1/2"'] \arrow[r, bend right, "1/2"'] & 2 \arrow[l, bend right, "1/2"'] \arrow[r, bend left, "1/2"] & \cdots
\end{tikzcd}$$

Markov Chains Handout for Stat 110

Prof. Joe Blitzstein (Harvard Statistics Department)

1 Introduction

Markov chains were first introduced in 1906 by Andrey Markov, with the goal of showing that the Law of Large Numbers does not necessarily require the random variables to be independent. Since then, they have become extremely important in a huge number of fields such as biology, game theory, finance, machine learning, and statistical physics. They are also very widely used for simulations of complex distributions, via algorithms known as MCMC (Markov Chain Monte Carlo).

To see where the Markov model comes from, consider first an i.i.d. sequence of random variables $X_0, X_1, \dots, X_n, \dots$ where we think of n as time. Independence is a very strong assumption: it means that the X_j's provide no information about each other. At the other extreme, allowing general interactions between the X_j's makes it very difficult to compute even basic things. Markov chains are a happy medium between complete independence and complete dependence.

The space on which a Markov process "lives" can be either discrete or continuous, and time can be either discrete or continuous. In Stat 110, we will focus on Markov chains $X_0, X_1, X_2, \dots$ in discrete space and time (continuous time would be a process X_t defined for all real $t \geq 0$). Most of the ideas can be extended to the other cases. Specifically, we will assume that X_n takes values in a finite set (the *state space*), which we usually take to be $\{1, 2, \dots, M\}$ (or $\{0, 1, \dots, M\}$ if it is more convenient). In Stat 110, we will always assume that our Markov chains are on finite state spaces.

Definition 1. A sequence of random variables $X_0, X_1, X_2, \dots$ taking values in the *state space* $\{1, \dots, M\}$ is called a *Markov chain* if there is an M by M matrix $Q = (q_{ij})$ such that for any $n \geq 0$,

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = j | X_n = i) = q_{ij}.$$

The matrix Q is called the *transition matrix* of the chain, and q_{ij} is the *transition probability* from i to j.

This says that given the history $X_0, X_1, X_2, \dots, X_n$, only the *most recent* term, X_n, matters for predicting X_{n+1}. If we think of time n as the present, times before n as the past, and times after n as the future, the Markov property says that given the present, the past and future are conditionally independent.

The Markov assumption greatly simplifies computations of conditional probability: instead of having to condition on the entire past, we only need to condition on the *most recent* value.

2 Transition Matrix

The transition probability q_{ij} specifies the probability of going from state i to state j in one step. The *transition matrix* of the chain is the $M \times M$ matrix $Q = (q_{ij})$. Note that Q is a nonnegative matrix in which each row sums to 1.

Definition 2. Let $q_{ij}^{(n)}$ be the n-step transition probability, i.e., the probability of being at j exactly n steps after being at i:

$$q_{ij}^{(n)} = P(X_n = j | X_0 = i).$$

Note that

$$q_{ij}^{(2)} = \sum_k q_{ik} q_{kj}$$

since to get from i to j in two steps, the chain must go from i to some intermediary state k, and then from k to j (these transitions are independent because of the Markov property). So the matrix Q^2 gives the 2-step transition probabilities. Similarly (by induction), powers of the transition matrix give the n-step transition probabilities:

$$q_{ij}^{(n)} \text{ is the } (i, j) \text{ entry of } Q^n.$$

Example. Figure 1 shows an example of a Markov chain with 4 states. The chain

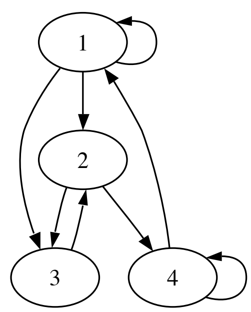


Figure 1: A Markov Chain with 4 Recurrent States

can be visualized by thinking of a particle wandering around from state to state, randomly choosing which arrow to follow. Here we assume that if there are a arrows originating at state i, then each is chosen with probability $1/a$, but in general each arrow could be given any probability, such that the sum of the probabilities on the arrows leaving i is 1. The transition matrix of the chain shown above is

$$Q = \begin{pmatrix} 1/3 & 1/3 & 1/3 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 1 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 \end{pmatrix}.$$

To compute, say, the probability that the chain is in state 3 after 5 steps, starting at state 1, we would look at the (3,1) entry of Q^5. Here (using a computer to find Q^5),

$$Q^5 = \begin{pmatrix} 853/3888 & 509/1944 & 52/243 & 395/1296 \\ 173/864 & 85/432 & 31/108 & 91/288 \\ 37/144 & 29/72 & 1/9 & 11/48 \\ 499/2592 & 395/1296 & 71/324 & 245/864 \end{pmatrix},$$

so $q_{13}^{(5)} = 52/243$.

To fully specify the behavior of the chain $X_0, X_1, \ldots$, we also need to give initial conditions. This can be done by setting the initial state X_0 to be a particular state x_0, or by randomly choosing X_0 according to some distribution. Let $(s_1, s_2, \ldots, s_M)$ be a vector with $s_i = P(X_0 = i)$ (think of this as the PMF of X_0, displayed as a vector). Then the distribution of the chain at any time can be computed using the transition matrix.

Proposition 3. *Define* $\mathbf{s} = (s_1, s_2, \ldots, s_M)$ *by* $s_i = P(X_0 = i)$*, and view* $\mathbf{s}$ *as a row vector. Then* $\mathbf{s}Q^n$ *is the vector which gives the distribution of* X_n*, i.e., the* j*th component of* $\mathbf{s}Q^n$ *is* $P(X_n = j)$*.*

Proof. Conditioning on X_0, the probability that the chain is in state j after n steps is $\sum_i s_i q_{ij}^n$, which is the jth component of $\mathbf{s}Q^n$. □

3 Recurrence and Transience

In the Markov chain shown in Figure 1, a particle moving around between states will continue to spend time in all 4 states in the long run. In contrast, consider the chain shown in Figure 2, and let the particle start at state 1. For a while, the chain may linger in the triangle formed by states 1, 2, and 3, but eventually it will reach state 4, and from there it can never return to states 1, 2, or 3. It will then wander around between states 4, 5, and 6 forever. In this example, states 1, 2, and 3 are *transient* and states 4, 5, and 6 are *recurrent*. In general, these concepts are defined as follows.

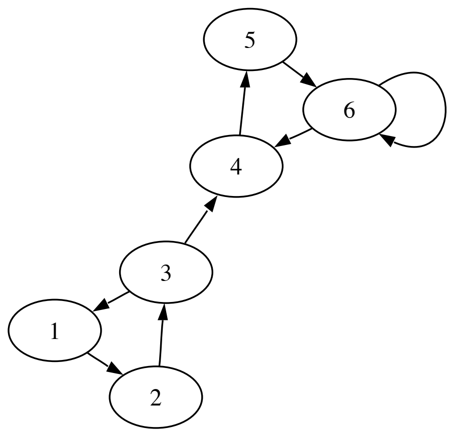


Figure 2: A Markov Chain with States 1, 2, and 3 Transient

Definition 4. State i is *recurrent* if the probability is 1 that the chain will return to i (eventually) if it starts at i. Otherwise, the state is *transient*, which means that if the chain starts out at i, there is a positive probability of never returning to i.

Classifying the states as recurrent or transient is important in understanding the long-run behavior of the chain. Early on in the history, the chain may spend time in transient states. Eventually though, the chain will spend all its time in recurrent states. But what fraction of the time will it spend in each of the recurrent state? To answer this question, we need the concept of a *stationary distribution*.

4 Stationary Distribution

The *stationary distribution* of a Markov chain, also known as the *steady state distribution*, describes the long-run behavior of the chain. Such a distribution is defined as follows.

Definition 5. A row vector $\mathbf{s} = (s_1, \ldots, s_M)$ such that $s_i \geq 0$ and $\sum_i s_i = 1$ is a *stationary distribution* for a Markov chain with transition matrix Q if

$$\sum_i s_i q_{ij} = s_j$$

or equivalently,

$$\mathbf{s}Q = \mathbf{s}.$$

Often π is used to denote a stationary distribution, but some prefer to reserve the letter π for some other obscure purpose. The equation $\mathbf{s}Q = \mathbf{s}$ means that if X_0 has distribution given by $\mathbf{s}$, then X_1 also has distribution $\mathbf{s}$. But then X_2 also has distribution $\mathbf{s}$, etc. That is, a Markov chain which starts out with a stationary distribution will stay in the stationary distribution forever.

In terms of linear algebra, the equation $\mathbf{s}Q = \mathbf{s}$ says that $\mathbf{s}$ is a left eigenvector of Q with eigenvalue 1. To get the usual kind of eigenvector (a right eigenvector), take transposes. Writing A' for the transpose of any matrix A, we have $Q'\mathbf{s}' = \mathbf{s}'$.

For example, if $Q = \begin{pmatrix} 1/3 & 2/3 \\ 1/2 & 1/2 \end{pmatrix}$, then $(3/7, 4/7)$ is a stationary distribution since

$$\begin{pmatrix} 3/7 & 4/7 \end{pmatrix} \begin{pmatrix} 1/3 & 2/3 \\ 1/2 & 1/2 \end{pmatrix} = \begin{pmatrix} 3/7 & 4/7 \end{pmatrix}.$$

This was obtained by finding the eigenvector of Q' with eigenvalue 1, and then normalizing (by dividing by the sum of the components).

But does a stationary distribution always exist? Is it unique? It turns out that a stationary distribution always exists. (Recall that we are assuming a finite state space; there may be no stationary distribution if the state space is infinite. For example, it can be shown that if X_n is a random walk on the integers, with $X_n = Y_1 + Y_2 + \cdots + Y_n$ where Y_j are i.i.d. with $P(Y_j = -1) = P(Y_j = 1) = 1/2$, then the random walk $X_1, X_2, \ldots$ does not have a stationary distribution.)

There are in fact examples where there is not a unique stationary distribution, e.g., in the Gambler's Ruin problem. But under the reasonable assumption that it is possible to get from any state to any other state, then there will be a unique stationary distribution.

Definition 6. A Markov chain with transition matrix Q is *irreducible* if for any two states i and j, it is possible to go from i to j with positive probability (in some number of steps). That is, the (i, j) entry of Q^n is positive for some n. The chain is *reducible* if it is not irreducible.

The chain in Figure 1 is irreducible (in terms of the picture, check that it's possible to go from anywhere to anywhere following the arrows; in terms of the transition matrix, note that all the entries of Q^5 are positive). The chain in Figure 2 is reducible since it is never possible to go from state 4 back to state 3 in any number of steps.

Theorem 7. *Any irreducible Markov chain has a unique stationary distribution. In this distribution, every state has positive probability.*

Definition 8. The *period* of a state i in a Markov chain is the greatest common divisor of the possible numbers of steps it can take to return to i when starting at i. That is, it is the greatest common divisor of numbers n such that the (i, i) entry of Q^n is positive. A state is called *aperiodic* if its period is 1, and the chain itself is called *aperiodic* if all its states are aperiodic, and *periodic* otherwise.

For example, the "clockwork" behavior of states $1, 2, 3$ in Figure 2 makes them periodic with period 3. In contrast, all the states in Figure 1 are aperiodic, so that chain is aperiodic.

Theorem 9. *Let $X_0, X_1, \ldots$ be an irreducible, aperiodic Markov chain with stationary distribution $\mathbf{s}$ and transition matrix Q. Then $P(X_n = i)$ converges to s_i as $n \to \infty$. In other words, Q^n converges to a matrix in which each row is $\mathbf{s}$.*

Therefore, after a large number of steps, the probability that the chain is in state i is close to the stationary probability s_i. We can also use the stationary distribution to find the average time between visits to a state i, and vice versa.

Theorem 10. *Let $X_0, X_1, \ldots$ be an irreducible Markov chain with stationary distribution $\mathbf{s}$. Let r_i be the expected time it takes the chain to return to i, given that it starts at i. Then $s_i = 1/r_i$.*

In the 2×2 example on the previous page, this says that in the long run, the chain will spend 3/7 of its time in state 1 and 4/7 of its time in state 2. Starting at state 1, it will take an average of 7/3 steps to return to state 1. The powers of the transition matrix converge to a matrix where each row is the stationary distribution:

$$\begin{pmatrix} 1/3 & 2/3 \\ 1/2 & 1/2 \end{pmatrix}^n \to \begin{pmatrix} 3/7 & 4/7 \\ 3/7 & 4/7 \end{pmatrix} \text{ as } n \to \infty.$$

In general, it may be computationally difficult to find the stationary distribution when the state space is large. Here are two important cases where working with eigenvalue equations for large matrices can be avoided.

Proposition 11. *If each column of the transition matrix Q sums to 1, then the uniform distribution over all states, $(1/M, 1/M, \dots, 1/M)$, is a stationary distribution.*

Proof. Assuming each column sums to 1, the row vector $\mathbf{v} = (1, 1, \dots, 1)$ satisfies $\mathbf{v}Q = \mathbf{v}$. It then follows that $(1/M, 1/M, \dots, 1/M)$ is stationary. $\square$

For example, if Q is a symmetric matrix (i.e., $q_{ij} = q_{ji}$), then $(1/M, 1/M, \dots, 1/M)$ is stationary. We now define *reversible* Markov chains, which intuitively have the property that they "look the same" if time is reversed (if the chain starts at the stationary distribution), and which are much nicer to work with than general chains.

Definition 12. Let $Q = (q_{ij})$ be the transition matrix of a Markov chain, and that there is $\mathbf{s} = (s_1, \dots, s_M)$ with $s_i \geq 0, \sum_i s_i = 1$, such that

$$s_i q_{ij} = s_j q_{ji}$$

for all states i, j. This equation is called the *reversibility* or *detailed balance* condition, and we say that the chain is *reversible* with respect to $\mathbf{s}$ if it holds.

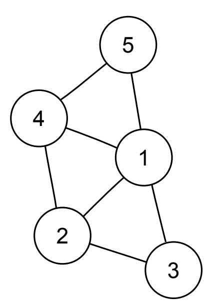
Proposition 13. *Suppose that $Q = (q_{ij})$ is a transition matrix which is reversible with respect to $\mathbf{s} = (s_1, \dots, s_M)$. Then $\mathbf{s}$ is a stationary distribution for the chain with transition matrix Q.*

Proof. We have

$$\sum_i s_i q_{ij} = \sum_i s_j q_{ji} = s_j \sum_i q_{ji} = s_j,$$

where the last equality is because each row sum of Q is 1. Thus, $\mathbf{s}$ is stationary. $\square$

Example 14. Consider a *random walk on an undirected network*, where a wanderer randomly traverses edges. From a node i, the wanderer randomly picks any of the edges at i, with equal probabilities, and then traverses the chosen edge. For example,



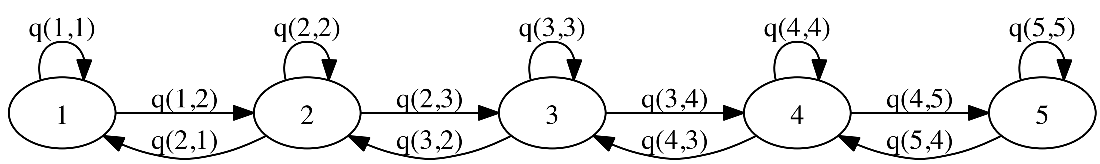
in the network shown above, from node 3 the wanderer goes to node 1 or node 2, with probability 1/2 each.

The *degree* of a node is the number of edges attached to it, and the *degree sequence* of a network with nodes $1, 2, \ldots, n$ is the vector $(d_1, \ldots, d_n)$ listing all the degrees, where d_j is the degree of node j. For example, the network above has degree sequence $\mathbf{d} = (4, 3, 2, 3, 2)$. Note that

$$d_i q_{ij} = d_j q_{ji}$$

for all i, j, since q_{ij} is $1/d_i$ if $\{i, j\}$ is an edge and 0 otherwise, for $i \neq j$. By Proposition 13, we have that the stationary distribution is proportional to the degree sequence. In the example above, this says that $\mathbf{s} = (\frac{4}{14}, \frac{3}{14}, \frac{2}{14}, \frac{3}{14}, \frac{2}{14})$ is the stationary distribution for the random walk.

Example 15. A *birth-death chain* on the state space $\{1, 2, \ldots, M\}$ is a Markov chain with transition matrix $Q = (q_{ij})$ such that $q_{ij} > 0$ if $|i-j| = 1$ and $q_{ij} = 0$ if $|i-j| \geq 2$. Intuitively, this says it's possible to go "one step to the left" and possible to go "one step to the right" (except at boundaries) but it's impossible to jump further in one step. For example, the chain below is a birth-death chain if the labeled transitions have positive probabilities (except for the "loops" from a state to itself, which are allowed to have 0 probability). We will now show that any birth-death chain is



reversible, and construct the stationary distribution. Let s_1 be a positive number (to be specified later). Since we want $s_1 q_{12} = s_2 q_{21}$, let $s_2 = s_1 q_{12}/q_{21}$. Then since we want $s_2 q_{23} = s_3 q_{32}$, let $s_3 = s_2 q_{23}/q_{32} = s_1 q_{12} q_{23}/(q_{32} q_{21})$. Continuing in this way, let

$$s_j = \frac{s_1 q_{12} q_{23} \cdots q_{j-1,j}}{q_{j,j-1} q_{j-1,j-2} \cdots q_{21}},$$

for all states j with $2 \leq j \leq M$. Choose s_1 so that the s_j's sum to 1. Then the chain is reversible with respect to $\mathbf{s}$, since $q_{ij} = q_{ji} = 0$ if $|i-j| \geq 2$ and by construction $s_i q_{ij} = s_j q_{ji}$ if $|i-j| = 1$. Thus, $\mathbf{s}$ is the stationary distribution.