Chapter 2

Sequence Evolution Models

In the previous lectures, we introduced two simple scoring functions for pairwise alignments:

- a similarity function, that assigns a score of M to matches ($M > 0$), m to mismatches ($m < 0$), and g to indels ($g < 0$) and
- an edit distance, which does not reward matches ($M = 0$) and assigns a unit cost to mismatches and gaps ($m, g > 0$).

These scoring functions allow us to compare two alignments by comparing their scores, but are less useful for assessing a pairwise alignment in an absolute sense. Given a pair of aligned sequences with a particular collection of matches, mismatches, and indels, does the alignment reflect enough similarity to suggest that it is of biological interest? One way of assessing an alignment in an absolute sense is to determine whether it reflects more similarity than we would expect by chance. In developing this approach, we must take into account the divergence of related sequences due to mutation. With that in mind, we will explore models of sequence evolution and then discuss how they are used to assess alignments. Sequence evolution models are typically based on Markov chains, so we will begin with a general introduction to Markov models.

2.1 Finite discrete Markov chains

In various computational biology applications, it is useful to track the stochastic variation of a random variable. Here are some examples:

1. **Time-dependent system:** For models of sequence evolving by substitution, the random variable of interest is the nucleotide (or amino acid) observed at a fixed position, or *site*, in the sequence at time t. The goal is to characterize how this random variable changes over time.

2. **Space-dependent system:** It is also useful to consider how the residues in a sequence change as one moves along the sequence from one site to the next. In this case, the random variable is the amino acid (or nucleotide) at site i. We are interested in how the probability of observing a given amino acid at site i depends on the amino acid observed at site $i-1$.

For each of these examples, we can model how the value of the random variable (the nucleotide or amino acid) changes with respect to an independent variable (time or position), using a *Markov* chain with a finite number of *states*, $E_1, E_2, \ldots E_s$. Each state corresponds to one of the possible values of the random variable:

- In a nucleic acid sequence, there are four states, each corresponding to the event of observing one of the four nucleotides at the site of interest, e.g., $E_1 = A, E_2 = G, E_3 = C, E_4 = T$.
- In a protein sequence, there are 20 states, each of which corresponds to the event of observing a given amino acid; for example $E_1 = \text{Ala}, E_2 = \text{Cys}, \ldots E_{20} = \text{Tyr}$.

In our examples above, the states are defined as follows:

1. In a time-dependent system we say the system is in state E_j at time t.
2. In a spatially varying system, we say the system is in state E_j at site i without concerning ourselves with time. This is in contrast to the previous example, where time varies and the position, i, is held fixed.

The probability that a Markov chain is in state E_j at time t is designated $\varphi_j(t)$[1]. Consider the example of modelling the evolution of a given nucleotide site over time. In this example, $\varphi_1(t)$ is the probability of observing an A at site i at time t. The vector $\varphi(t) = (\varphi_1(t), \varphi_2(t), \ldots \varphi_s(t))$ describes the *state probability distribution* over all states at time t. The *initial* state probability distribution is given by $\varphi(0)$. Note that Ewens and Grant[2] use π to denote the initial state distribution: $\pi = (\varphi_1(0), \varphi_2(0), \ldots \varphi_s(0))$.

In order to capture the stochastic variation of the system, we must also define the probability of making a transition from one state to another. The *transition probability*, P_{jk}, is the probability that the chain will be in state E_k at time $t+1$, given that it was in state E_j at the previous time step, t. In the time-dependent, nucleotide sequence example, P_{12} is the probability of an A-to-G substitution at site i.

P is an $s \times s$ matrix specifying the probability of making a transition from any state to any other state. The rows of this matrix sum to one ($\sum_k P_{jk} = 1$) since the chain must be

[1] To simplify the exposition, we will focus on models where time is the independent variable. However, the framework is more general, and can be used to model variation with respect to other independent variables, such as the position in a sequence.

[2] Statistical Methods in Bioinformatics: An Introduction. W. Ewens, G. Grant. Springer 2001.

in some state at every time step. The columns do not have to add up to one, since there is no guarantee that the system will end up in a particular state, k.

The *Markov property* states that Markov chains are memoryless. In other words, the probability that the chain is in state E_j at time $t+1$, depends only on the state at time t and not on the past history of the states visited at times $t-1, t-2$...

In this course, we will focus on discrete, finite, time-homogeneous Markov chains. These are models with a *finite* number of states, in which time (or space) is split into *discrete* steps. The assumption of discrete steps is quite natural for a spatially varying system, because sequences of symbols are inherently discrete, but somewhat artificial for the sequence evolution over time model, since time is continuous. Our models are *time-homogeneous*, because the transition matrix does not change over time.

Summary of Markov chain notation

A Markov chain has *states* $E_0, E_1, \ldots E_s$ corresponding to the range of the associated random variable.

$\varphi_j(t)$ is the probability that the chain is in state E_j at time t. The vector $\varphi(t) = (\varphi_1(t), \ldots \varphi_s(t))$ is the *state probability distribution* at time t.

$\pi = \varphi(0)$ is the *initial state probability distribution.*

P is the *transition probability matrix.* P_{jk} gives the probability of making a transition to state E_k at time $t+1$, given that the chain was in state E_j at time t. The rows of this matrix sum to one: $\sum_k P_{jk} = 1$.

The state probability distribution at time $t+1$ is given by $\varphi(t+1) = \varphi(t) \cdot P$. The probability of being in state E_k at $t+1$ is

$$\varphi_k(t+1) = \sum_j \varphi_j(t) P_{jk}$$

The *Markov property* states that Markov chains are memoryless. The probability that the chain is in state E_k at time $t+1$, depends only on $\varphi(t)$ and is independent of $\varphi(t-1), \varphi(t-2), \varphi(t-3) \ldots$

2.1.1 Higher Order Markov Chains

The memoryless requirement that the probability of occupying state E_k at time $t+1$ depends only on $\varphi(t)$ can sometimes be relaxed to allow for a more general Markov process. Markov chains that uphold the memoryless property are often called *first order* Markov chains. Higher-order dependencies can also be modelled. For example, consider a case the nucleotide at site i depends not only on the nucleotide observed at site $i-1$ but also the nucleotide observed at site $i-2$. This case can be modeled with a *second order* Markov chain because the nucleotide at site i depends on the previous two sites. The transition matrix P would be of size 16×4; columns would represent the nucleotide at the site under consideration and rows would represent all combinations of the two nucleotides preceeding. More generally, an *n-th order* Markov chain is a system where the probability of a given state at time t depends on the previous n sites. The transition matrix P would be of size $s^n \times s$.

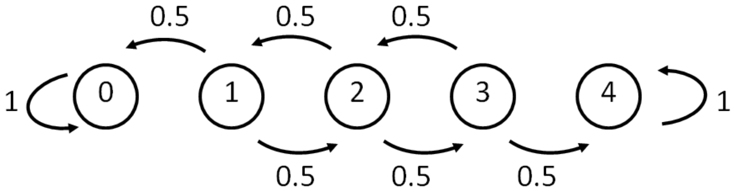
2.2 Random walks

To illustrate these concepts, let us consider a simple example: A drunk is staggering about on a very short railway track with five ties on top of a mesa (a high hill with a flat top and steep sides.) Here state E_j corresponds to the event that the drunk is standing on the j^{th} tie, where $0 \leq j \leq 4$. At each time step, the drunk staggers either to the left or to the right with equal probability. If the drunk reaches either end of the track (either the 0^{th} or the 4^{th} tie), he falls off the mesa. This model is called a *random walk with absorbing boundaries*, because once the drunk falls off the mesa, he can never get back on the railroad track. States E_0 and E_4 are *absorbing* states. Once the system enters one of these states, it remains in that state forever, since $P_{00} = P_{44} = 1$. This results in the following transition probability matrix:

$$P = \begin{bmatrix} & E_0 & E_1 & E_2 & E_3 & E_4 \\ E_0 & 1 & 0 & 0 & 0 & 0 \\ E_1 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ E_2 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ E_3 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ E_4 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.1}$$

Note that each row sums to one, consistent with the definition of a Markov chain.

The transition matrix of a Markov chain can be represented as a graph, where the nodes represent states and the edges represent transitions with non-zero probability. For example, the random walk with absorbing boundaries can be modeled like this:



Note that the sum of the weights on all outgoing edges from any given state sum to 1, just as row sums are equal to 1 in the transition matrix.

How does the state probability distribution change over time? If we know the state probability distribution at time t, the distribution at the next time step is given by:

$$\varphi_k(t+1) = \sum_j \varphi_j(t) P_{jk} \tag{2.2}$$

or

$$\varphi(t+1) = \varphi(t) P \tag{2.3}$$

in matrix notation.

For example, suppose that at time $t = 0$, the drunk is standing on the middle tie (state E_2); that is, $\varphi(0) = (0, 0, 1, 0, 0)$. To obtain the state probability distribution after one time step, we apply Equation 2.2:

$$\varphi_k(1) = \sum_{j=0}^{4} \varphi_j(0) P_{jk}.$$

Thus, the probability of being in state E_1 when $t = 1$ is given by

$$\begin{aligned}\varphi_1(1) &= \sum_{j=0}^{4} \varphi_j(0) P_{j1}.\\ &= 0 \cdot 0 + 0 \cdot 0 + 1 \cdot \frac{1}{2} + 0 \cdot 0 \cdot 0 \cdot 0\\ &= \frac{1}{2}.\end{aligned}$$

Note this is equivalent to multiplying the vector $(0, 0, 1, 0, 0)$ by the second column of the transition matrix (Equation 2.1).

Since the Markov chain is symmetrical, it is easy to show that $\varphi_3(1)$ is also equal to 1/2. (Try it.) It is not possible to reach state E_0 or state E_4 in a single step from state E_2, so $\varphi_0(1) = \varphi_4(1) = 0$. Nor is it possible to remain in state E_2 for two consecutive time steps since $P_{22} = 0$, so $\varphi_2(1) = 0$. Since state E_2 is the only state with non-zero probability at time $t = 0$, we obtain,

$$\varphi(1) = (0, \frac{1}{2}, 0, \frac{1}{2}, 0).$$

Now that we have the probability distribution at time $t = 1$, we can calculate the probability distribution at time $t = 2$ using the same procedure

$$\varphi_k(2) = \sum_{j=0}^{4} \varphi_j(1) P_{jk}.$$

The probability of being in state E_0 at $t = 2$ is given by

$$\begin{aligned}\varphi_0(2) &= \sum_{j=0}^{4} \varphi_j(1) P_{j0} \\ &= 0 \cdot 1 + \frac{1}{2} \cdot \frac{1}{2} + 0 \cdot 0 + \frac{1}{2} \cdot 0 + 0 \cdot 0 \\ &= \frac{1}{4}.\end{aligned}$$

Again, because the matrix is symmetrical, $\varphi_4(2) = \varphi_0(2)$. The probability of being in state E_2 is

$$\begin{aligned}\varphi_2(2) &= \sum_{j=0}^{4} \varphi_j(1) P_{j2} \\ &= 0 \cdot 0 + \frac{1}{2} \cdot \frac{1}{2} + 0 \cdot 0 + \frac{1}{2} \cdot \frac{1}{2} + 0 \cdot 0 \\ &= \frac{1}{2}.\end{aligned}$$

The probabilities of being in state E_1 or E_3 at time $t = 2$ are zero, because $P_{11} = 0$ and $P_{33} = 0$. The probability distribution vector at time $t = 1$ is, therefore,

$$\varphi(2) = (\frac{1}{4}, 0, \frac{1}{2}, 0, \frac{1}{4}). \tag{2.4}$$

2.2.1 Calculating "n-step" Transition Probabilities

Suppose that we wish to know the state of the system after two time steps. In the previous section, we used Equation 2.2 to calculate $\varphi(1)$, given $\varphi(0)$, and then we applied Equation 2.2 again to calculate $\varphi(2)$, from $\varphi(1)$. We can approach this from another linear algebra perspective by constructing a *two-step* transition probability matrix in which each time step corresponds to two time steps in the original Markov chain.

Here, we derive a general expression for $\varphi(t+2)$ in terms of $\varphi(t)$ and P^2. From Equation 2.2, we obtain

$$\varphi_l(t+1) = \sum_{j=0}^{s} \varphi_j(t) P_{jl} \tag{2.5}$$

and

$$\varphi_k(t+2) = \sum_{l=0}^{s} \varphi_l(t+1) P_{lk}. \tag{2.6}$$

Substituting the right hand side of Equation 2.5 for $\varphi_l(t+1)$ in Equation 2.6 yields

$$\varphi_k(t+2) = \sum_{l=0}^{s} \left(\sum_{j=0}^{s} \varphi_j(t) P_{jl} \right) P_{lk}.$$

We can reverse the order of the summations since the terms may be added in any order:

$$\varphi_k(t+2) = \sum_{j=0}^{s} \left(\sum_{l=0}^{s} \varphi_j(t) P_{jl} \right) P_{lk}.$$

Since $\varphi_j(t)$ does not depend on l, it can be moved out of the summation over l, yielding:

$$\varphi_k(t+2) = \sum_{j=0}^{s} \varphi_j(t) \left(\sum_{l=0}^{s} P_{jl} P_{lk} \right). \tag{2.7}$$

The term in the inner summation is simply the element in row j and column k of the matrix obtained by multiplying matrix P by itself. In other words,

$$P^{(2)}_{jk} = \sum_{l=0}^{s} P_{jl} P_{lk},$$

where $P^{(2)} = P \times P$, so that Equation 2.7 may be rewritten as

$$\varphi_k(t+2) = \sum_{j=0}^{s} \varphi_j(t) P^{(2)}_{jk}.$$

Matrix $P^{(2)}$ is the transition matrix for moving two time steps over the Markov chain described by P. In other words, a single time step in $P^{(2)}$ is equivalent to two time steps in

P. Similarly, the *n-step* transition probability matrix, $P^{(n)}$, models change after n time steps such that:

$$P^{(n)} = \underbrace{P \times P .. \times P}_{n \text{ times}} = P^n.$$

The n-step equivalent of Equation 2.3 is

$$\varphi(t+n) = \varphi(t) \cdot P^{(n)}.$$

As an example, let's apply this approach to our 5-state random walk with absorbing boundaries. Recall that the transition matrix for the random walk, given in Equation 2.1, is

$$P = \left[\begin{array}{cccccc} & E_0 & E_1 & E_2 & E_3 & E_4 \\ E_0 & 1 & 0 & 0 & 0 & 0 \\ E_1 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ E_2 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ E_3 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ E_4 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \tag{2.8}$$

Multiplying P times itself yields the two-step transition matrix, $P^{(2)}$:

$$P^{(2)} = \left[\begin{array}{cccccc} & E_0 & E_1 & E_2 & E_3 & E_4 \\ E_0 & 1 & 0 & 0 & 0 & 0 \\ E_1 & \frac{1}{2} & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ E_2 & \frac{1}{4} & 0 & \frac{1}{2} & 0 & \frac{1}{4} \\ E_3 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{2} \\ E_4 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \tag{2.9}$$

Try this matrix multiplication to convince yourself that this is correct. The state probability distribution at $t = 2$ can be calculated by applying $P^{(2)}$ to $\varphi(0)$:

$$\begin{aligned} \varphi(2) &= \varphi(0) \cdot P^{(2)} \\ &= (0, 0, 1, 0, 0) \cdot P^{(2)} \\ &= (\frac{1}{4}, 0, \frac{1}{2}, 0, \frac{1}{4}). \end{aligned}$$

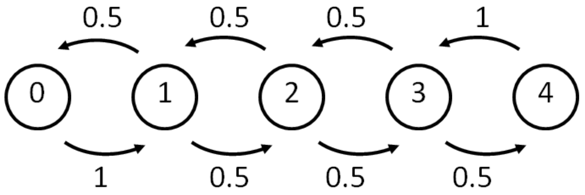
Note that this gives the same result as Equation 2.4, which we got by applying the original Markov chain twice.

2.2.2 Periodic Markov chains

Let us consider a second example. In order to save the drunk from an early death, we introduce a random walk with *reflecting* boundaries. At each step, the drunk moves to the left or to the right with equal probability. When the drunk reaches one of the boundary states (E_0 or E_4), he returns to the adjacent state (E_1 or E_3) at the next step, with probability one. This yields the following transition probability matrix:

$$P = \left[\begin{array}{cccccc} & E_0 & E_1 & E_2 & E_3 & E_4 \\ E_0 & 0 & 1 & 0 & 0 & 0 \\ E_1 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ E_2 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ E_3 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ E_4 & 0 & 0 & 0 & 1 & 0 \end{array}\right] \qquad (2.10)$$

The random walk with reflecting boundaries can be represented graphically like this:



Again, for any given state, the outgoing edges sum to 1.

Suppose that the drunk starts out on the middle tie at $t = 0$, as before. That is, the initial state probability distribution is $\varphi(0) = (0, 0, 1, 0, 0)$. The state distributions for the first two time steps are the same in this random walk and in the random walk with absorbing boundaries specified by Equation 2.1. These are

$$\varphi(1) = (0, \frac{1}{2}, 0, \frac{1}{2}, 0) \qquad (2.11)$$

$$\varphi(2) = (\frac{1}{4}, 0, \frac{1}{2}, 0, \frac{1}{4}). \qquad (2.12)$$

This makes sense because the two random walk models differ only in the boundary states, E_0 and E_4, and $\varphi_0(t) = \varphi_4(t) = 0$ when $t = 0$ or $t = 1$. We calculate the state probability distribution at $t = 3$ by multiplying the vector $\varphi(2)$ with the matrix P:

$$\begin{aligned}\varphi(3) &= \varphi(2) \cdot P \\ &= (\frac{1}{4}, 0, \frac{1}{2}, 0, \frac{1}{4}) \cdot P \\ &= (0, \frac{1}{2}, 0, \frac{1}{2}, 0). \end{aligned} \tag{2.13}$$

Comparing Equations 2.11 and 2.13 demonstrates that the state probability distribution at time $t = 3$ is the same as the distribution at time $t = 1$. In other words, $\varphi(3) = \varphi(1)$. Similarly, $\varphi(4) = \varphi(2)$, as can be seen from the following calculation:

$$\begin{aligned}\varphi(4) &= \varphi(3) \cdot P \\ &= (0, \frac{1}{2}, 0, \frac{1}{2}, 0) \cdot P \\ &= (\frac{1}{4}, 0, \frac{1}{2}, 0, \frac{1}{4}). \end{aligned} \tag{2.14}$$

From this we can see that the probability state distribution will be $(0, \frac{1}{2}, 0, \frac{1}{2}, 0)$ at all odd time steps and $(\frac{1}{4}, 0, \frac{1}{2}, 0, \frac{1}{4})$ at all even time steps. Thus, the random walk with reflecting boundaries is a periodic Markov chain.

A Markov chain is *periodic* if there is some state that can only be visited, with any probability greater than 0, in multiples of m time steps, where $m > 1$. Formally, state j has period

$$m = \gcd\{n > 0 : P_{jj}^{(n)} > 0\},$$

where "gcd" is the greatest common divisor. In our example, $P_{jj}^{(n)} > 0$ for $n = 2, 4, 6, 8, ...$ for all j; therefore, each state has a period of 2, which is the gcd of $\{2, 4, 6, 8, ...\}$. If $m = 1$, the state is aperiodic. To show that an irreducible Markov chain is aperiodic, it is sufficient to show one of the following:

1. Any state in the Markov chain is aperiodic.
2. Any state has a self-loop; i.e., $P_{jj} > 0$ for some state j.
3. All elements of the n-step transition matrix $P^{(n)}$ are greater than 0 for some positive integer n.
4. If $P_{jj}^{(k)} > 0$ and $P_{jj}^{(l)} > 0$, then the $\gcd(k, l) = 1$.

We do not require periodic Markov chains for modeling sequence evolution and will only consider aperiodic Markov chains going forward.

2.2.3 Stationary distributions

A state probability distribution, φ^*, that satisfies the equation

$$\varphi^* = \varphi^* P \tag{2.15}$$

is called a *stationary* distribution. A key question for a given Markov chain is whether such a stationary distribution exists. Equation 2.15 is equivalent to a system of s equations with s unknowns. One way to determine the steady state distribution is to solve that system of equations. The stationary distribution can also be obtained using matrix algebra, but that approach is beyond the scope of this course.

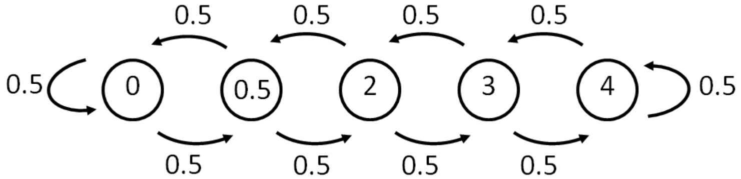
The random walk with reflecting boundaries clearly does not have a stationary distribution, because it is periodic. Every state with non-zero probability at time t has zero probability at time $t+1$ and vice versa. The random walk with absorbing boundaries does not have a *unique* stationary distribution; both $(1,0,0,0,0)$ and $(0,0,0,0,1)$ are stationary distributions of the random walk with absorbing boundaries.

For the rest of this course, we will concern ourselves only with aperiodic Markov chains that do not have absorbing states. In fact, we will make an even stronger assumption and restrict our consideration to Markov chains in which every state is connected to every other state via a series of zero or more states. If a finite Markov chain is aperiodic and connected in this way, it has a unique stationary distribution. We will not attempt to prove this or even to state the theorem in a rigorous way. For those who are interested, a very nice treatment can be found in Chapter 15 of *Probability Theory and its Applications (Volume I)* by William Feller (John Wiley & Sons).

As an example of a Markov chain with a unique stationary distribution, we introduce a third random walk model that has neither absorbing, nor reflecting boundaries. In this model, if the drunk is in one of the boundary states (E_0 or E_4) at time t, then at time $t+1$ he remains in the boundary state with a probability of 0.5 or returns to the adjacent state (E_1 or E_3) with a probability of 0.5. This results in the following state transition matrix:

$$P = \left[\begin{array}{cccccc} & E_0 & E_1 & E_2 & E_3 & E_4 \\ E_0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ E_1 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ E_2 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ E_3 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ E_4 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{array}\right] \tag{2.16}$$

which can be represented graphically like this:



Yet again, the weights on outgoing edges sum to 1 for every state.

We can determine the stationary state distribution for this random walk model by substituting this transition matrix into Equation 2.15. The probability of being in state E_0 is

$$\begin{aligned}\varphi_0^* &= \sum_{j=0}^{4} \varphi_j^* P_{j0} \\ &= \varphi_0^* P_{00} + \varphi_1^* P_{10} + \varphi_2^* P_{20} + \varphi_3^* P_{30} + \varphi_4^* P_{40}.\end{aligned}$$

This reduces to

$$\varphi_0^* = \frac{1}{2}\varphi_0^* + \frac{1}{2}\varphi_1^*, \tag{2.17}$$

since P_{20}, P_{30} and P_{40} are all equal to zero. The other steady state probabilities are derived similarly, yielding

$$\varphi_1^* = \frac{1}{2}\varphi_0^* + \frac{1}{2}\varphi_2^* \tag{2.18}$$

$$\varphi_2^* = \frac{1}{2}\varphi_1^* + \frac{1}{2}\varphi_3^* \tag{2.19}$$

$$\varphi_3^* = \frac{1}{2}\varphi_2^* + \frac{1}{2}\varphi_4^* \tag{2.20}$$

$$\varphi_4^* = \frac{1}{2}\varphi_3^* + \frac{1}{2}\varphi_4^*. \tag{2.21}$$

In addition, the steady state probabilities must sum to 1, since at any point in time, the drunk must be *somewhere*. This imposes an additional constraint:

$$\varphi_0^* + \varphi_1^* + \varphi_2^* + \varphi_3^* + \varphi_4^* = 1. \tag{2.22}$$

The Markov model specified by Equation 2.16 has a stationary distribution if the above equations have a solution. By repeated substitution, it is possible to show that Equations 2.17 - 2.21 reduce to $\varphi_0^* = \varphi_1^* = \varphi_2^* = \varphi_3^* = \varphi_4^*$. (Do the algebra to convince yourself that this is true.) Applying the constraint in Equation 2.22, we see that $\varphi^* = (0.2, 0.2, 0.2, 0.2, 0.2)$ is a unique solution to the above equations.

In this example, we found a unique solution to Equation 2.22, demonstrating that our third random walk has a unique stationary state. Solving Equation 2.15 is a general approach to finding the stationary distribution. Alternatively, if we know the stationary state distribution, or have an educated guess, it is sufficient to verify that it indeed satisfies Equation 2.15. For example, it is easy to verify that $(0.2, 0.2, 0.2, 0.2, 0.2) \cdot P = (0.2, 0.2, 0.2, 0.2, 0.2)$.

A stationary distribution of $\varphi^* = (0.2, 0.2, 0.2, 0.2, 0.2)$ does not mean that we expect to find 20% of a drunk standing on each railroad tie. Imagine instead that there are an infinite number of co-existing universes and that in each universe, we have a mesa with a railroad track with five ties and a drunk. These drunks are lurching back and forth according to the same Markov model, but they are not synchronized; at any given time point, some of the drunks will be on the 2nd tie, other drunks will be on the 4th tie, and so on. At steady state, for every j, $0 \leq j \leq 4$, 20% of the parallel universes will have a drunk on the j^{th} railroad tie.

Limiting distributions and stationary distributions

If a Markov chain is finite, irreducible, and aperiodic, then it has a *limiting distribution* and the chain will converge to the stationary distribution φ^*, independent of the starting distribution π. Formally

$$\lim_{n \to \infty} P_{jk}^{(n)} = \varphi_k^*.$$

In other words, as $n \to \infty$ the n-step transition matrix will be

$$P^{(n)} = \left[\begin{array}{cccccc} & E_1 & \cdots & E_j & \cdots & E_s \\ E_1 & \varphi_1^* & \cdots & \varphi_j^* & \cdots & \varphi_s^* \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ E_i & \varphi_1^* & \cdots & \varphi_j^* & \cdots & \varphi_s^* \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ E_s & \varphi_1^* & \cdots & \varphi_j^* & \cdots & \varphi_s^* \end{array}\right]. \tag{2.23}$$

2.2.4 Time reversibility

Most sequence substitution models are *time reversible.* A time reversible model exhibits the same steady state behavior if we run it backward, instead of forward. This is a convenient property that makes many calculations simpler. Most, if not all, of the Markov models we encounter in this course are time reversible. However, when analyzing a data set involving genomes with very different G+C content, a time reversible model of sequence evolution may not provide accurate results. It is therefore helpful to understand the concept of time reversibility and be aware of whether or not the models you are using have this property.

Formally, a Markov chain is *time reversible* if

$$\varphi_j^* P_{jk} = \varphi_k^* P_{kj}$$

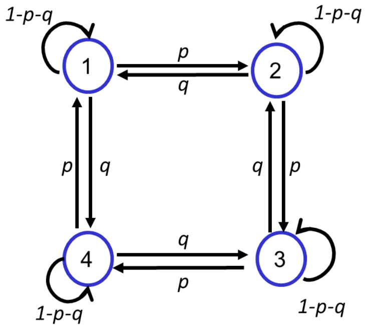
for all states j and k. This criterion is called the detailed balance equation. Earlier, we introduced parallel universes as a metaphor for the stationary state probability distribution,

φ. This metaphor is also helpful in understanding time reversibility: If a Markov chain satisfies the detailed balance equation, then the number of universes that are moving from E_j to E_k is equivalent to the number of universes that are moving from E_k to E_j.

Kolmogorov proposed an alternate criterion for time reversibility that depends only on the transition probability matrix. Let $\mathcal{M}$ be a Markov chain with a unique stationary distribution and let $j_1, \ldots, j_n$ be a sequence of states (a path of length n) through $\mathcal{M}$. Then, $\mathcal{M}$ is time reversible if and only if

$$P_{j_1,j_2}\, P_{j_2,j_3}\, \ldots P_{j_{(n-1)},j_n}\, P_{j_n,j_1} = P_{j_1,j_n}\, P_{j_n,j_{(n-1)}}\, \ldots P_{j_3,j_2}\, P_{j_2,j_1}. \qquad (2.24)$$

Time reversibility and the use of Kolmogorov's criteria are illustrated by the Markov model in the figure below. The four transitions associated with a clockwise circuit have probability p^4, while the probability of a counterclockwise circuit is q^4. When $p = q$, the transition probabilities satisfy Kolmogorov's criterion; the model is time reversible. When $p \neq q$, the probabilities of the clockwise and counterclockwise circuits are not the same ($p^4 \neq q^4$). Kolmogorov's criterion is violated, indicating that the model is not time reversible.



What is the best way to test time reversibility in practise? Kolmogorov's criterion is useful if you have a sequence of states that violates Equation 2.24, providing a direct demonstration that the model at hand is not time reversible. However, it is less useful as a general, systematic test of time reversibility. Given a Markov model with a $k \times k$ transition probability matrix, P, you could use the detailed balance equations to test for time reversibility by checking that $\varphi^*_j P_{jk} = \varphi^*_j P_{kj}$ for all combinations of j and k. However, it is more efficient to do all of these tests using a single matrix product. First, determine the stationary distribution φ^* by solving the system of equations in Equation 2.15. Then, construct an $s \times s$ diagonal matrix, D, where the entries on the main diagonal are $\varphi^*_1, \varphi^*_2, \ldots, \varphi^*_s$ and the off-diagonal elements are zero. The detailed balance equations hold if and only if $D \times P$ is a symmetric matrix. Convince yourself that this is the case.