Lecture Notes: Markov chains
Dannie Durand

At the beginning of the semester, we introduced two simple scoring functions for pairwise alignments: edit distance, which does not reward matches and assigns a unit cost to mismatches and gaps, and a simple similarity function, that assigns a score of \( M \) to matches \( (M > 0) \), \( m \) to mismatches \( (m < 0) \), and \( g \) to indels \( (g < 0) \). These scoring functions allowed us to compare two alignments by comparing their scores. We also used these functions in pairwise alignment algorithms to find the alignment that is optimal with respect to the scoring function.

These simple scoring functions 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 a certain amount of divergence between 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 introduce this framework first.

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. For models of sequence evolving by point mutation, the random variable of interest is the nucleotide observed at a given position, or site, in the sequence at time \( t \). The goal is to characterize how this random variable changes over time.
2. It is also useful consider spatial variation; for example, how residues in a sequence change as one moves along the sequence. 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 \( i - 1 \).

Each of these scenarios can be modeled by a finite, discrete Markov chain with a finite number of states, \( E_1, E_2, \ldots, E_s \). States are defined in our three examples as follows:

1. There are four states, each corresponding to the event of observing one of the four nucleotides at the site of interest, e.g., \( E_1 = A, E_2 = C, E_3 = G, E_4 = T \). In a time dependent system, such as this one, we say the system is in state \( E_j \) at time \( t \).
2. Each of the 20 states 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 a spatially varying system, we say the system is in
state $E_j$ at site $i$. This is in contrast to the previous example, where time varies and the position, $i$, is held fixed.

The probability that the chain is in state $j$ at time $t$ is designated $\varphi_j(t)$. 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 $\varphi(0)$. Ewens and Grant use $\pi$ 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 $k$ at time $t + 1$ given that it was in state $j$ at the previous time step, $t$.

$P$ is a $s \times s$ matrix describing 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 always be in some state at every time step. There are no constraints on the column sums, since there is no guarantee that you will arrive at 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 somewhat artificial for the sequence evolution model in example 1, but quite natural for Example 1 and 2, because sequences of symbols are inherently discrete. Our models are time-homogeneous, because the transition matrix, $P_{jk}$, does not change over time.

In class, we considered a simple example: A drunk is staggering about on a very short railway track with seven ties on top of a mesa (a high hill with a flat top and steep sides.) At each time step, the drunk staggers either to the left or to the right. Here $E_j$ is the event that the drunk is standing on the $j$th tie. At each step, the drunk moves to the left or to the right with equal probability, resulting in the following transition probability matrix, $P$:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>$\frac{1}{2}$</td>
<td>0</td>
<td>$\frac{1}{2}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>$\frac{1}{2}$</td>
<td>0</td>
<td>$\frac{1}{2}$</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>$\frac{1}{2}$</td>
<td>0</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
If the drunk reaches either end of the railway (either the 0th or the 4th 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 $P_{00} = P_{44} = 1$.

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:

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

or

$$\varphi(t+1) = \varphi(t) P \quad (2)$$

in matrix notation.

To obtain the state probability distribution after one time step, we apply Equation 1:

$$\varphi_1(1) = \sum_{j=0}^4 \varphi_j(0) P_{jk}$$

$$= 0 \cdot 0 + 0 \cdot 0 + 1 \cdot \frac{1}{2} + 0 \cdot 0 \cdot 0$$

$$= \frac{1}{2}.$$

Since the Markov chain is symmetrical, $\varphi_3(1)$ is also equal to $1/2$. 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, 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, following the same procedure, we calculate the probability distribution at time $t = 2$ from $\varphi(1)$:

$$\varphi_0(2) = \sum_{j=0}^4 \varphi_j(1) P_{j0}$$

$$= 0 \cdot 0 + \frac{1}{2} \cdot \frac{1}{2} + 0 \cdot 0 + \frac{1}{2} \cdot 0 + 0 \cdot 0$$

$$= \frac{1}{4}.$$
and

$$\varphi_2(2) = \sum_{j=0}^{4} \varphi_j(1) P_{j2}$$

$$= 0 \cdot 0 + \frac{1}{2} \cdot \frac{1}{2} + 0 \cdot 0 + \frac{1}{2} \cdot \frac{1}{2} + 0 \cdot 0$$

$$= \frac{1}{2}.$$ 

As above, $\varphi_4(1) = \varphi_0(1)$, by symmetry, yielding

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

### Higher order Markov chains

Suppose we wish to know the state of the system after two time steps. From Equation 1, we obtain

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

and

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

Substituting the right hand side of Equation 4 for $\varphi_l(t + 1)$ in Equation 5 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, yielding

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

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. Let $P^{(2)} = P^2 = P \times P$. Then

$$P_{jk}^{(2)} = \sum_{l=0}^{s} P_{jl} P_{lk}.$$
so that Equation 6 may be rewritten as

\[ \varphi_k(t + 2) = \sum_{j=0}^{s} \varphi_j(t) P_{jk}^{(2)}. \]

Matrix \( P^{(2)} \) is the transition matrix of a 2\textsuperscript{nd} order Markov chain that has the same states as the 1\textsuperscript{st} order Markov chain described by \( P \). However, a single time step in \( P^{(2)} \) is equivalent to two time steps in \( P \). Similarly, an \( n \textsuperscript{th} \) Markov chain models change after \( n \) time steps with a transition probability matrix

\[ P^{(n)} = P^n = P \times P \ldots P. \]

As an example, we use this approach to investigate the periodicity of our 5-state random walk with absorbing boundaries. The transition matrix, \( P \), for the 1\textsuperscript{st} order random walk is

\[
\begin{array}{c|ccccc}
 & 0 & 1 & 2 & 3 & 4 \\
\hline
0 & 1 & 0 & 0 & 0 & 0 \\
1 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\
2 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\
3 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\
4 & 0 & 0 & 0 & 0 & 1 \\
\end{array}
\]

Multiplying \( P \) times itself yields the 2\textsuperscript{nd} order transition matrix, \( P^{(2)} \):

\[
\begin{array}{c|ccccc}
 & 0 & 1 & 2 & 3 & 4 \\
\hline
0 & 1 & 0 & 0 & 0 & 0 \\
1 & \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} & 0 \\
2 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & \frac{1}{4} \\
3 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{2} \\
4 & 0 & 0 & 0 & 0 & 1 \\
\end{array}
\]

In class, we calculated the state probability distribution at \( t = 2 \) by applying \( P^{(2)} \) to \( \varphi(0) \):

\[
\varphi(2) = \varphi(0) \cdot P^{(2)} = (0, 0, 1, 0, 0) \cdot P^{(2)} = (\frac{1}{4}, 0, \frac{1}{2}, 0, \frac{1}{4})
\]
and obtained the same result as Equation 3, which we got by applying the first order Markov chain twice.