Introduction to Markov Chains

What is a Markov **Chain**?

$$\boxed{X_1} \quad \boxed{X_2} \quad \boxed{X_3} \quad \boxed{X_n}$$

Random Variables

Could model as i.i.d.
(independent and identically distributed)

Realistic? What if index is time?

What is a Markov **Chain**?

$$X_1 \rightarrow X_2 \rightarrow X_3 \dashrightarrow X_n$$

Let's add a dash of dependence
(but not too much!)

The Markov Property

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

The Markov Property

Everything Before

Next Now Previous First Next Now

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

Memoryless!

State Space

$$\boxed{X_1} \rightarrow \boxed{X_2} \rightarrow \boxed{X_3} \dashrightarrow \boxed{X_n}$$

$$X_i \in \{Rainy, Sunny\}$$

http://setosa.io/ev/markov-chains/

State Spaces

$$\boxed{X_1} \rightarrow \boxed{X_2} \rightarrow \boxed{X_3} \dashrightarrow \boxed{X_n}$$

$$X_i \in \{Rainy, Sunny\}$$

$$X_i \in \{1, 2, 3, 4, 5, 6\}$$

$$X_i \in \{A, C, G, T\}$$

$$X_i \in \{AAA, AAC, AAG, \ldots, TTG, TTT\}$$

State Spaces (Discrete)

$$\boxed{X_1} \rightarrow \boxed{X_2} \rightarrow \boxed{X_3} \dashrightarrow \boxed{X_n}$$

$$X_i \in \{Rainy, Sunny\}$$

$$X_i \in \{1, 2, 3, 4, 5, 6\}$$

$$X_i \in \{A, C, G, T\}$$

$$X_i \in \{AAA, AAC, AAG, \ldots, TTG, TTT\}$$

State Spaces (Continuous)

$$X_1 \rightarrow X_2 \rightarrow X_3 \dashrightarrow X_n$$

$$X_i \in \mathbb{R}$$

$$X_i \in \mathbb{R}_{>0}$$

$$X_i \in [0, 1]$$

What sorts of continuous state spaces might we have in phylogenetics?

Transition Matrix

$$\begin{array}{cc} & \begin{array}{cc} R & S \end{array} \\ \begin{array}{c} R \\ S \end{array} & \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix} \end{array}$$

Transition Matrix

To

$$\begin{array}{cc} & \begin{array}{cc} R & S \end{array} \\ \text{From} \quad \begin{array}{c} R \\ S \end{array} & \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix} \end{array}$$

To

$$\text{From}\quad \begin{array}{cc} & \begin{array}{cc} R & S \end{array} \\ \begin{array}{c} R \\ S \end{array} & \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix} \end{array}$$

$$P(X_{n+1} = R | X_n = R) = 0.7$$

$$P(X_{n+1} = S | X_n = R) = 0.3$$

To

From

$$\begin{array}{c|c} & \begin{array}{cc} R & S \end{array} \\ \begin{array}{c} R \\ S \end{array} & \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix} \end{array}$$

$$P(X_{n+1} = R | X_n = R) = 0.7$$

$$P(X_{n+1} = S | X_n = R) = 0.3$$

$$P(X_{n+1} = R | X_n = S) = 0.3$$

$$P(X_{n+1} = S | X_n = S) = 0.7$$

http://setosa.io/markov/index.html#%7B%22tm%22%3A%5B%5B0.7%2C0.3%5D%2C%5B0.3%2C0.7%5D%5D%7D

Transition Matrix

$$Q = \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix}$$

$$q_{ij} = P(X_{n+1} = j | X_n = i)$$

$$q_{11} = q_{RR} = 0.7$$

Q and ***q*** give us a sense for what will happen in the next step. But what about 2,3,4,…,100 steps in the future?

$$q_{ij}^{(100)} = \,?$$

Transition Matrix

$$Q = \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix}$$

$$q_{ij} = P(X_{n+1} = j | X_n = i)$$

$$q_{11} = q_{RR} = 0.7$$

Q and ***q*** give us a sense for what will happen in the next step. But what about 2,3,4,…,100 steps in the future?

$$q_{ij}^{(100)} \neq \ (q_{ij})^{100}$$

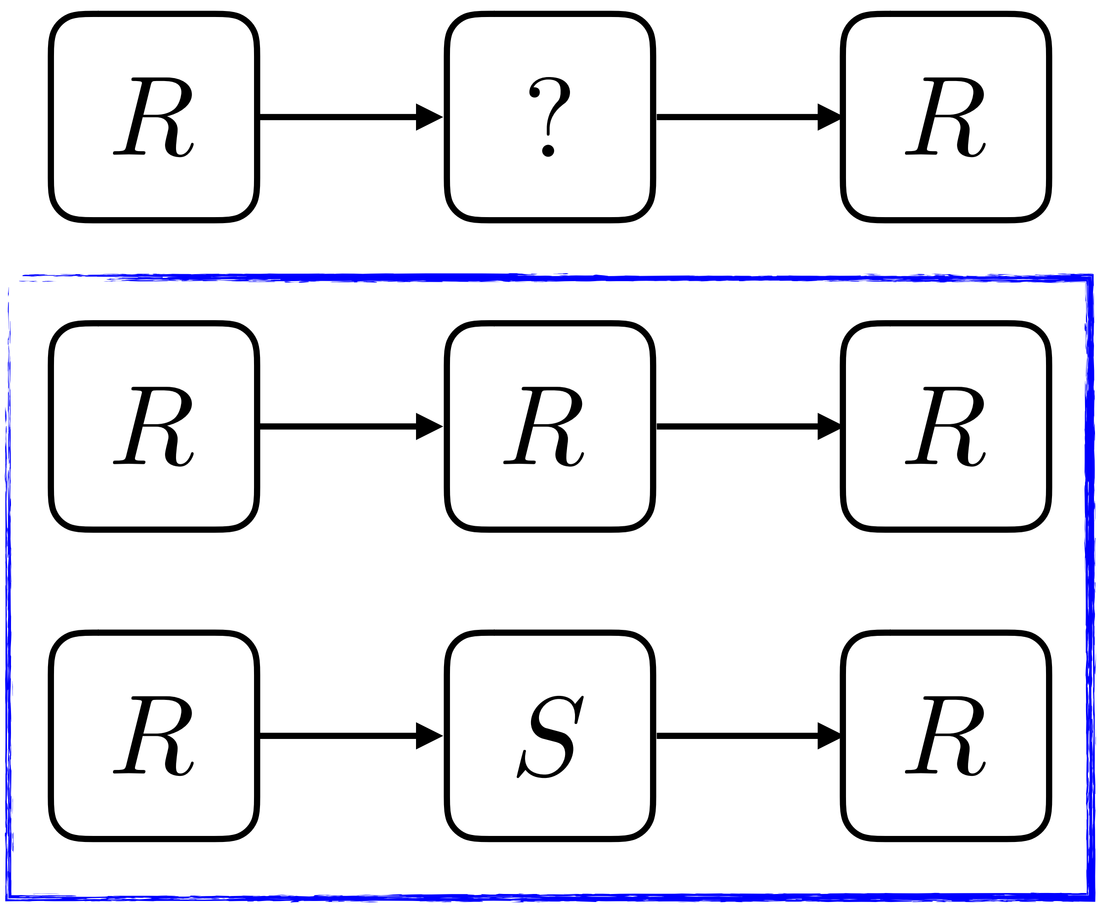
Transition Matrix

$$P(X_{n+2} = R | X_n = R)$$

$$\boxed{R} \longrightarrow \boxed{?} \longrightarrow \boxed{R}$$

Transition Matrix

$$P(X_{n+2} = R | X_n = R)$$



Transition Matrix

$$P(X_{n+2} = R | X_n = R)$$

$$\boxed{R} \longrightarrow \boxed{?} \longrightarrow \boxed{R}$$

$$(q_{RR})^2 = \boxed{\begin{array}{c} \boxed{R} \longrightarrow \boxed{R} \longrightarrow \boxed{R} \\ \\ \boxed{R} \longrightarrow \boxed{S} \longrightarrow \boxed{R} \end{array}}$$

Transition Matrix

$$P(X_{n+3} = R | X_n = R)$$

$$R \to ? \to ? \to R$$

$$(q_{RR})^3 = \quad R \to R \to R \to R$$

$$R \to R \to S \to R$$

$$R \to S \to R \to R$$

$$R \to S \to S \to R$$

Transition Matrix

$$P(X_{n+3} = R | X_n = R)$$

$$R \to ? \to ? \to R$$

$$q_{RR}^{(3)} = \boxed{\begin{array}{c} R \to R \to R \to R \\ R \to R \to S \to R \\ R \to S \to R \to R \\ R \to S \to S \to R \end{array}}$$

Transition Matrix

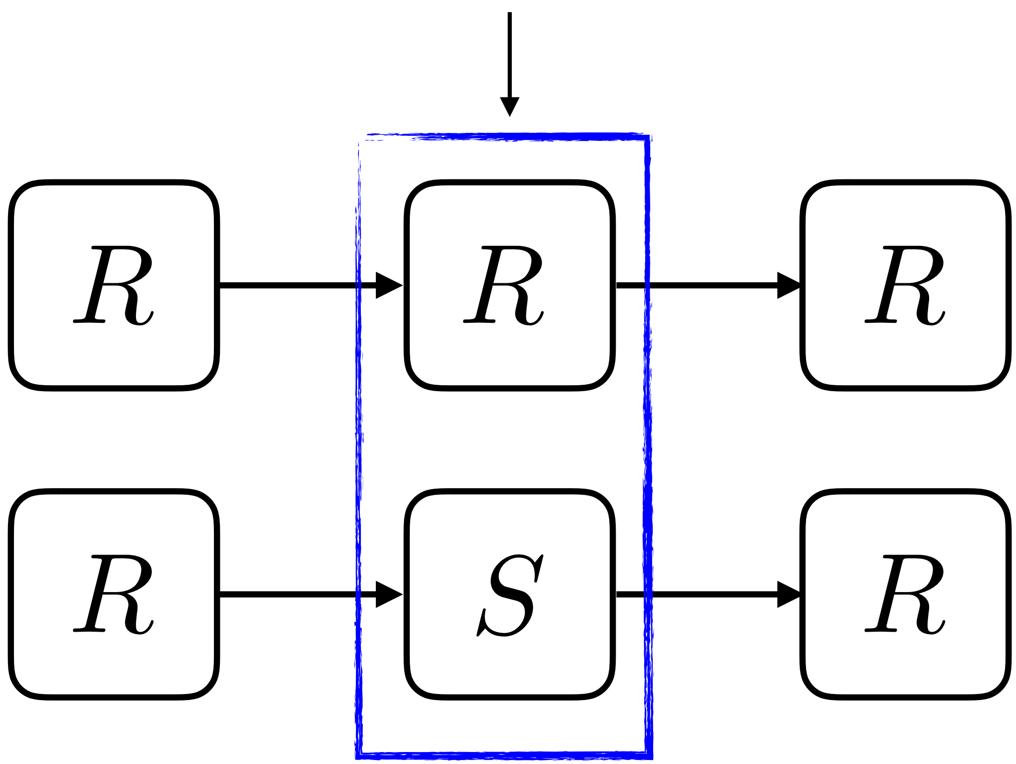
$q_{ij}^{(n)}$ is the (i, j) entry of Q^n

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

Transition Matrix

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

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



Stationary Distribution

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

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

Stationary Distribution

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

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

What does this mean??

Stationary Distribution

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

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

Simply put, if the states have certain probabilities of appearing in one iteration, they will have those same probabilities in the next iteration. The **probabilities do not change**!

Reversibility

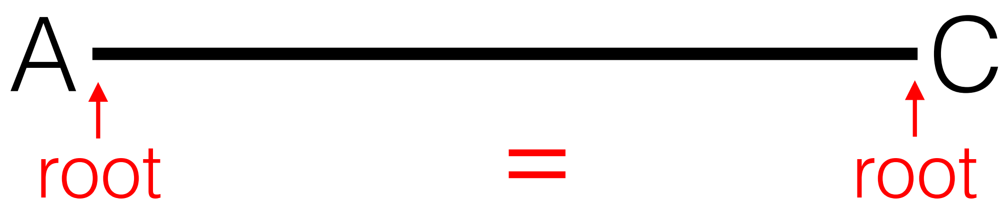
$$s_i q_{ij} = s_j q_{ji}$$

Detailed Balance Equation

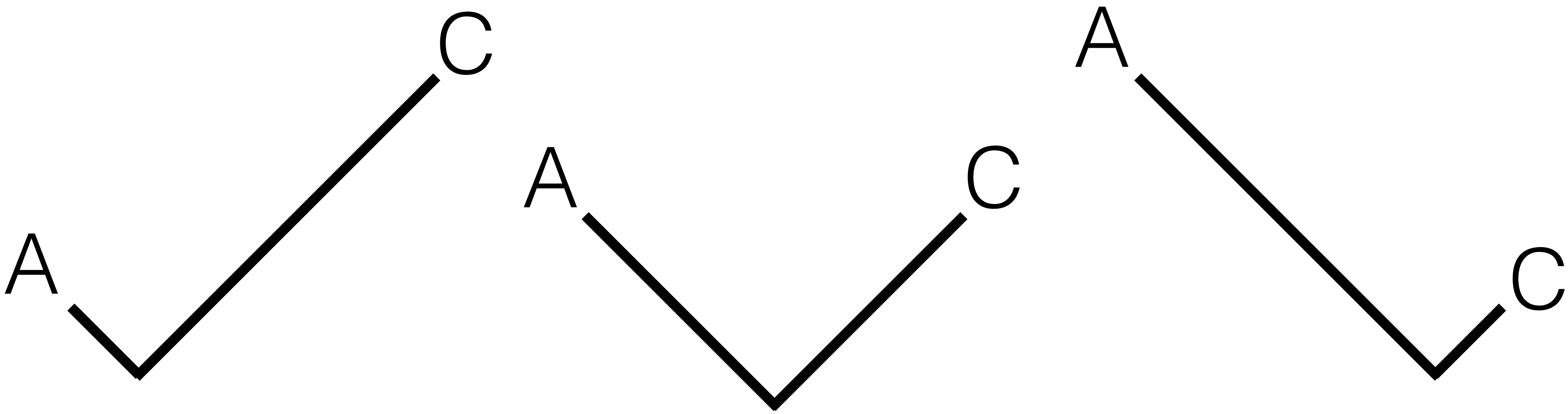
The probability of a series of states in the chain is the same forward, as it is in reverse? Why might this be important for phylogenetics?

The Pulley Principle

The Pulley Principle



The Pulley Principle



If the chain is reversible, the likelihood will be the same no matter where we put our "root". This is known as the pulley principle (Felsenstein 1981).