Introduction to Markov Chains

What is a Markov **Chain**?

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

Random Variables

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

Realistic? What if index is time?

What is a Markov **Chain**?

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

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

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

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

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

State Spaces

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

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

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

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

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

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)

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

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

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

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

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

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

$$\begin{array}{ccc} & & \text{To} \\ & R & S \\ \text{From} \quad \begin{array}{c} R \\ S \end{array} & \multicolumn{2}{c}{\begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix}} \end{array}$$

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

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

To

$$
\begin{array}{ccc}
 & R & S \\
R & \left(\begin{array}{c} 0.7 \\ 0.3 \end{array} \right. & \left. \begin{array}{c} 0.3 \\ 0.7 \end{array} \right) \\
S & &
\end{array}
$$

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

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

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

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

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

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

In-Class Exercise (pairs)

(1) Create a Markov chain class with these values:

- Number of steps (positive integer)
- State space (list)
- Transition matrix (list of lists of floats - or own class!)
- Sampled states (list)

and these methods:

- run (sample states for each step)
- clear (remove any sampled states)

(2) Create a list (or lists) to hold frequencies of states for different runs. For the {Rainy,Sunny} example, start each run in *S*. Now run 100 chains for 1 step. Record state frequencies across chains. Then run 100 chains for 2 steps. Record state frequencies. Then 5, then 10.