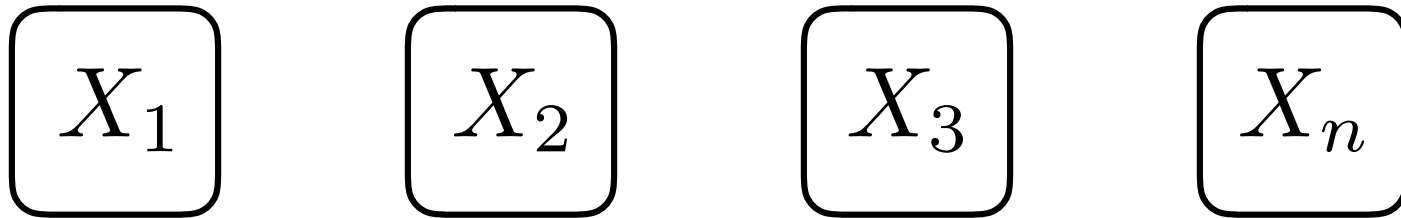


Introduction to Markov Chains

What is a Markov **Chain**?

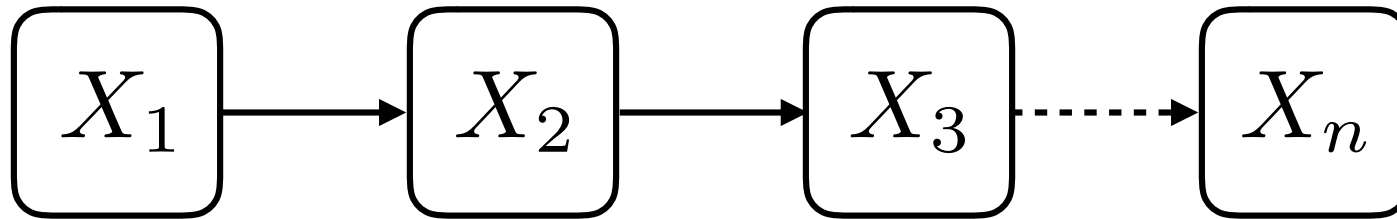


Random Variables

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

Realistic? What if index is time?

What is a Markov **Chain**?



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

The Markov Property

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

The Markov Property

Everything Before

Next

Now

Previous

First

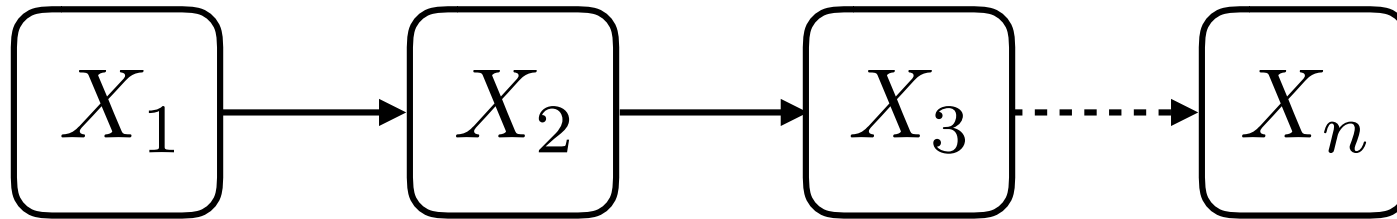
Next

Now

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

Memoryless!

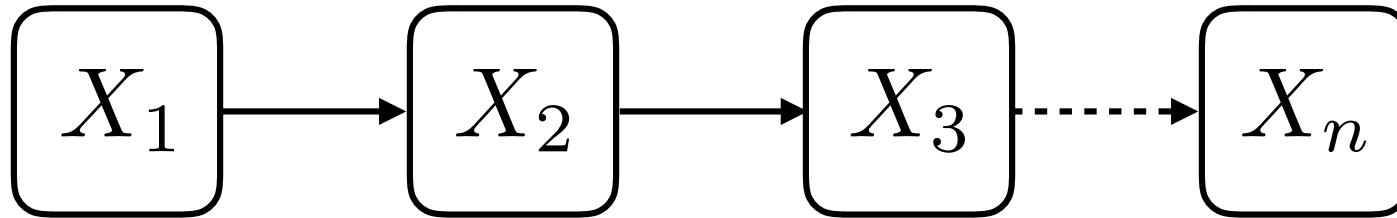
State Space



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

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

State Spaces



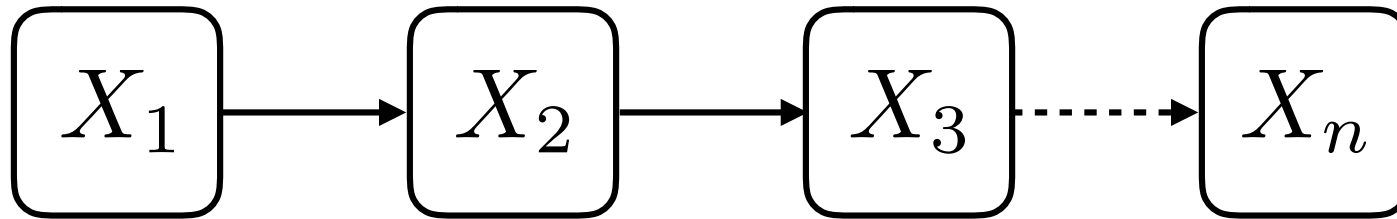
$$X_i \in \{Rainy, Sunny\}$$

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

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

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

State Spaces (Discrete)



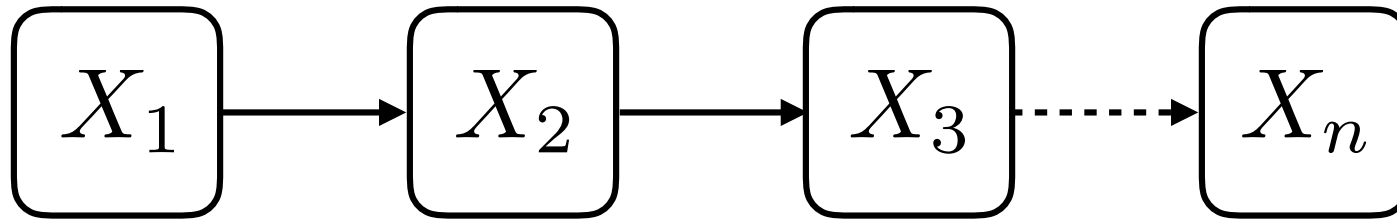
$$X_i \in \{Rainy, Sunny\}$$

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

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

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

State Spaces (Continuous)



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

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

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

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

Transition Matrix

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

Transition Matrix

From

To

$$\begin{matrix} & R & S \\ \begin{matrix} R \\ S \end{matrix} & \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix} \end{matrix}$$

		To	
		R	S
From	R	0.7	0.3
	S	0.3	0.7

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

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

		To	
		R	S
From	R	0.7	0.3
	S	0.3	0.7

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

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

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

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

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.