Introduction to Bioinformatics for Computer Scientists

Lecture 8

A Detour to Markov Chains

- Before we start looking at likelihood models for phylogenetics
- We will review the concept of Markov Chains
- This will be useful to
  - Better understand likelihood models
  - Better understand Markov Chain Monte Carlo Sampling for Bayesian statistics

Markov Chains - Outline

- We will mostly talk about discrete Markov chains as this is conceptually easier
- Then, we will talk how to get from discrete Markov chains to continuous Markov chains … which are used in phylogenetics

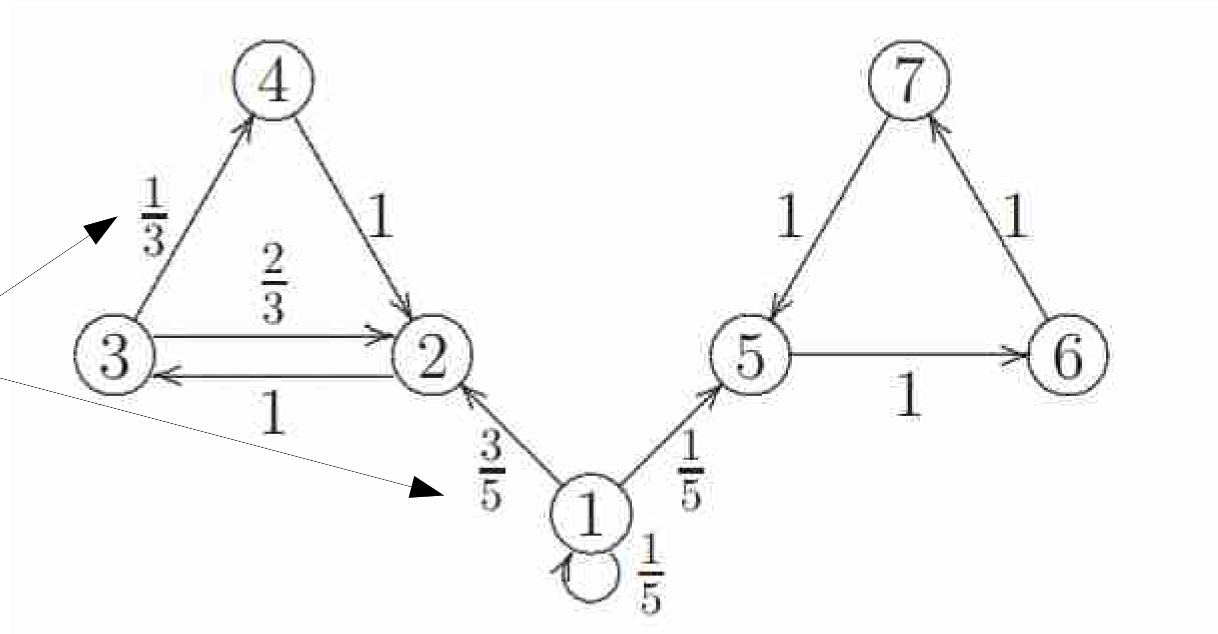
Markov Chains

- Stochastic processes with transition diagrams
- Process, is written as $\{X_0, X_1, X_2, \ldots\}$

  where X_t is the state at discrete time t
- Markov property: X_{t+1} **ONLY** depends on X_t
- Such processes are called **Markov Chains**

An Example

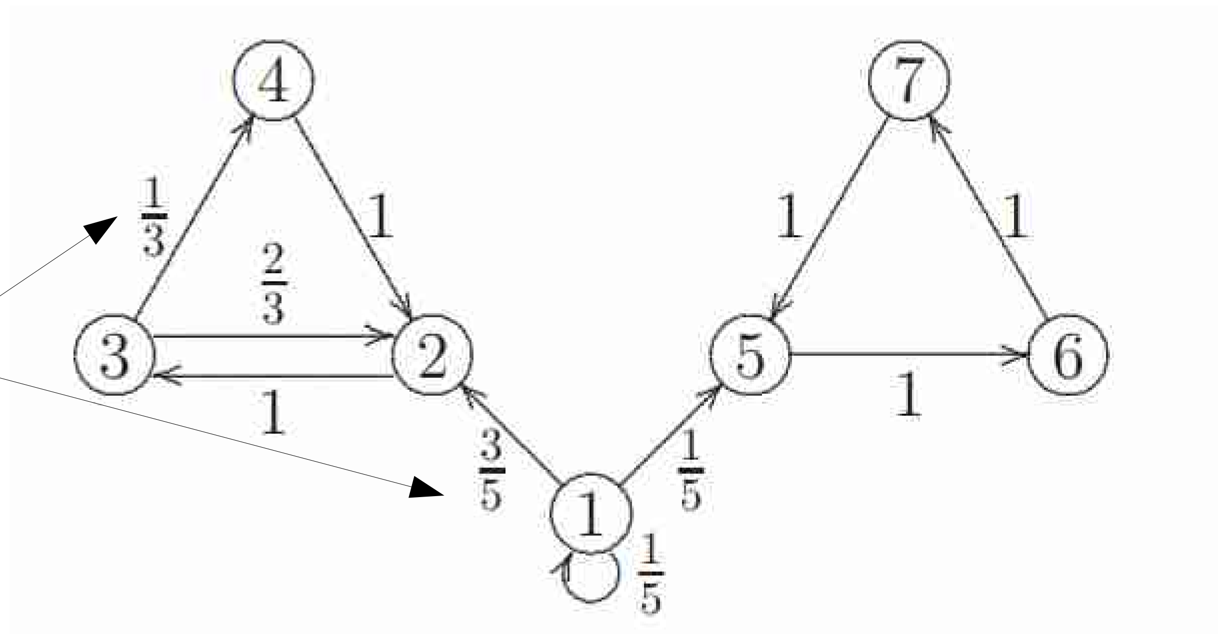
State transition
probabilities



The Markov flea example: flea hopping around **at random** on this diagram **according to the probabilities** shown

An Example

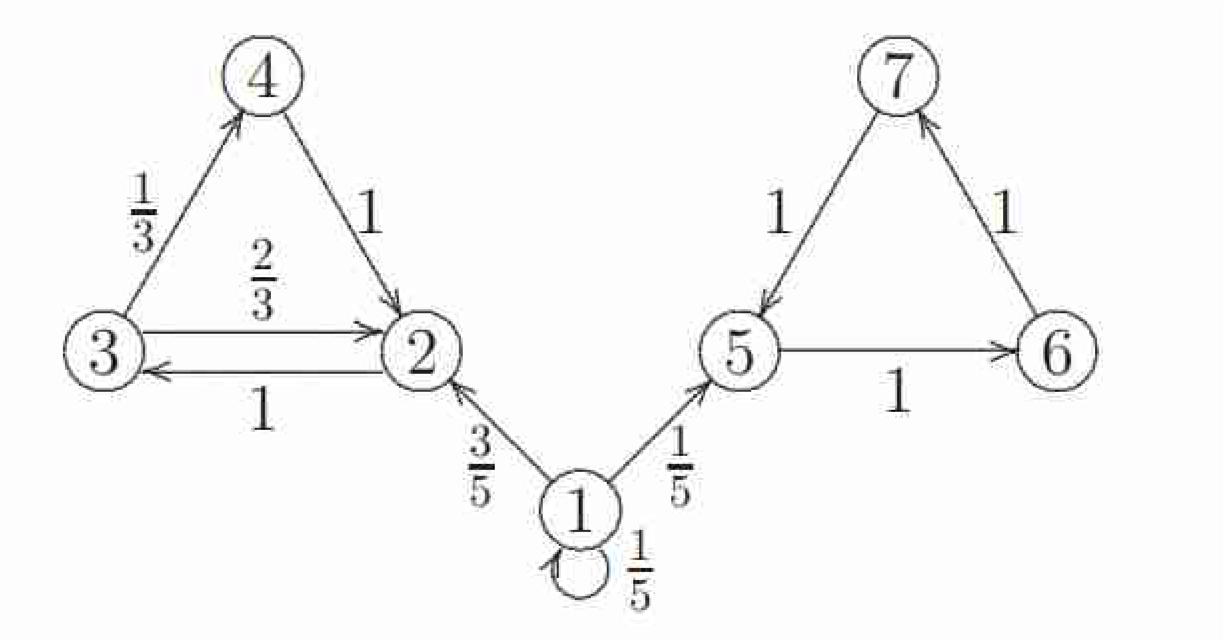
State transition probabilities



The Markov flea example: flea hopping around **at random** on this diagram **according to the probabilities** shown

State space *S = {1,2,3,4,5,6,7}*

An Example



- What is the probability of ever reaching state *7* from state *1*?
- Starting from state *2*, what is the expected time taken to reach state *4*?
- Starting from state *2*, what is the long-run proportion of time spent in state 3?
- Starting from state 1, what is the probability of being in state 2 at time t? Does the probability converge as $t \to \infty$, and if so, to what?

Definitions

- The Markov chain is the process $X_0, X_1, X_2, \ldots$.

- **Definition:** The state of a Markov chain at time t is the value of X_t

  For example, if $X_t = 6$, we say the process is in state 6 at time t.

- **Definition**: The state space of a Markov chain, S, is the set of values that each X_t can take.

  For example, $S = \{1, 2, 3, 4, 5, 6, 7\}$.

  Let S have size N (possibly infinite).

- **Definition**: A trajectory of a Markov chain is a particular set of values for $X_0, X_1, X_2, \ldots$

  For example, if $X_0 = 1$, $X_1 = 5$, and $X_2 = 6$, then the trajectory up to time $t = 2$ is $1, 5, 6$.

  More generally, if we refer to the trajectory $s_0, s_1, s_2, s_3, \ldots$ we mean that

  $X_0 = s_0, X_1 = s_1, X_2 = s_2, X_3 = s_3, \ldots$

  ‘Trajectory’ is just a word meaning ‘path'

Markov Property

- Only the most recent point X_t affects what happens next, that is, X_{t+1} only depends on X_t, but not on X_{t-1}, X_{t-2}, . . .
- More formally:

$$\mathbb{P}(X_{t+1} = s \mid X_t = s_t, X_{t-1} = s_{t-1}, \ldots, X_0 = s_0) \; = \; \mathbb{P}(X_{t+1} = s \mid X_t = s_t),$$

Markov Property

- Only the most recent point X_t affects what happens next, that is, X_{t+1} only depends on X_t, but not on X_{t-1}, X_{t-2}, . . .
- More formally:

$$\mathbb{P}(X_{t+1} = s \mid X_t = s_t, X_{t-1} = s_{t-1}, \ldots, X_0 = s_0) \;=\; \mathbb{P}(X_{t+1} = s \mid X_t = s_t),$$

- Explanation

$$\mathbb{P}(X_{t+1} = s \quad \mid \quad X_t = s_t, \quad \underbrace{\cancel{X_{t-1} = s_{t-1}, X_{t-2} = s_{t-2}, \ldots, X_1 = s_1, X_0 = s_0}}\,)$$

↑ *distribution of X_{t+1}*

↑ *depends on X_t*

↑ *but whatever happened before time t doesn't matter.*

Definition

Definition: Let $\{X_0, X_1, X_2, \ldots\}$ be a sequence of discrete random variables. Then $\{X_0, X_1, X_2, \ldots\}$ is a **Markov chain** if *it satisfies the Markov property:*

$$\mathbb{P}(X_{t+1} = s \mid X_t = s_t, \ldots, X_0 = s_0) = \mathbb{P}(X_{t+1} = s \mid X_t = s_t),$$

for all $t = 1, 2, 3, \ldots$ *and for all states* $s_0, s_1, \ldots, s_t, s$.

Definition

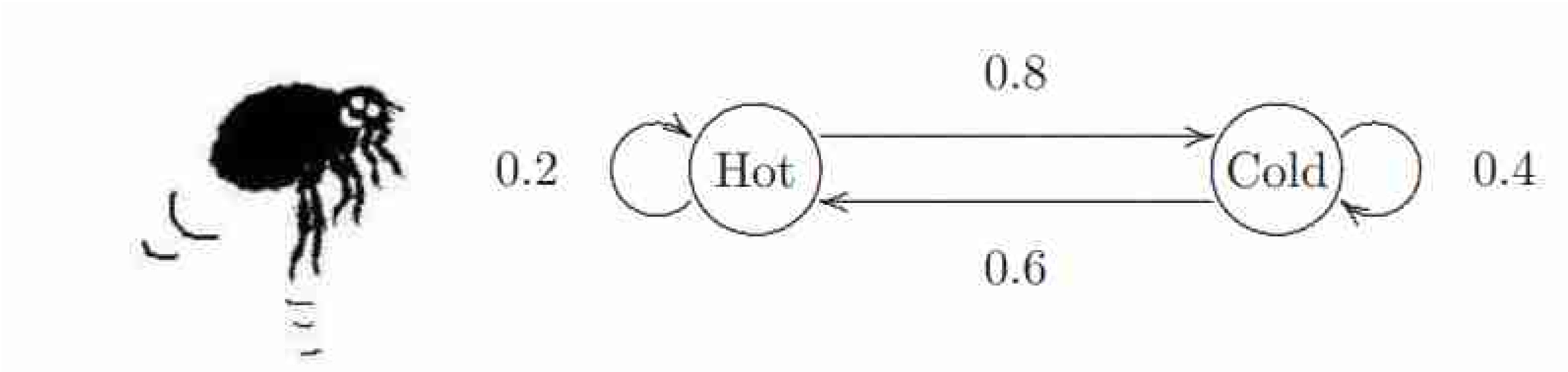
Discrete states, e.g., *A, C, G, T*

Definition: Let $\{X_0, X_1, X_2, \ldots\}$ be a sequence of discrete random variables. Then $\{X_0, X_1, X_2, \ldots\}$ is a **Markov chain** if *it satisfies the Markov property:*

$$\mathbb{P}(X_{t+1} = s \mid X_t = s_t, \ldots, X_0 = s_0) \;=\; \mathbb{P}(X_{t+1} = s \mid X_t = s_t),$$

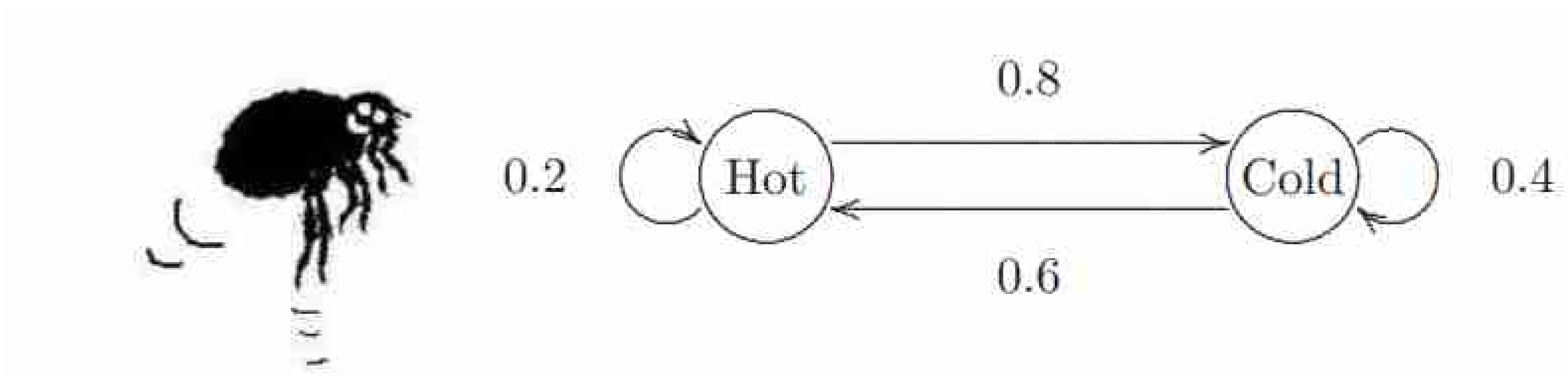
for all $t = 1, 2, 3, \ldots$ *and for all states* $s_0, s_1, \ldots, s_t, s$.

The Transition Matrix



Let us transform this into an equivalent transition matrix which is just another way of describing this diagram.

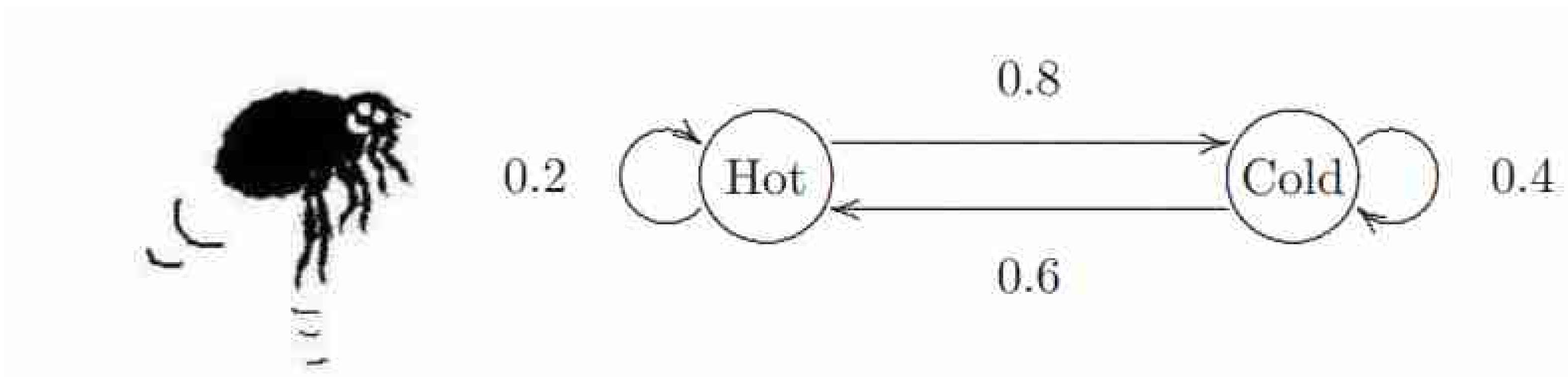
The Transition Matrix



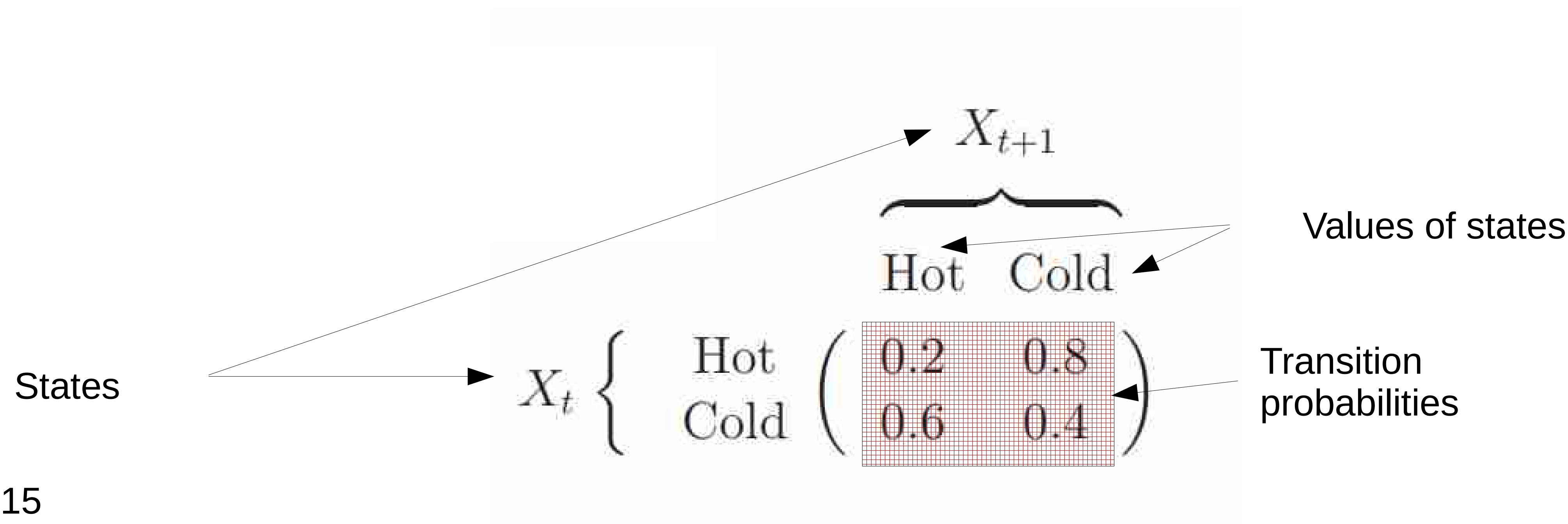
Let us transform this into an equivalent transition matrix which is just another
Equivalent way of describing this diagram.

$$
\begin{array}{c}
 & & \overbrace{\begin{array}{cc}\text{Hot} & \text{Cold}\end{array}}^{X_{t+1}} \\
X_t \left\{ \begin{array}{c} \text{Hot} \\ \text{Cold} \end{array} \right. & & \begin{pmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \end{pmatrix}
\end{array}
$$

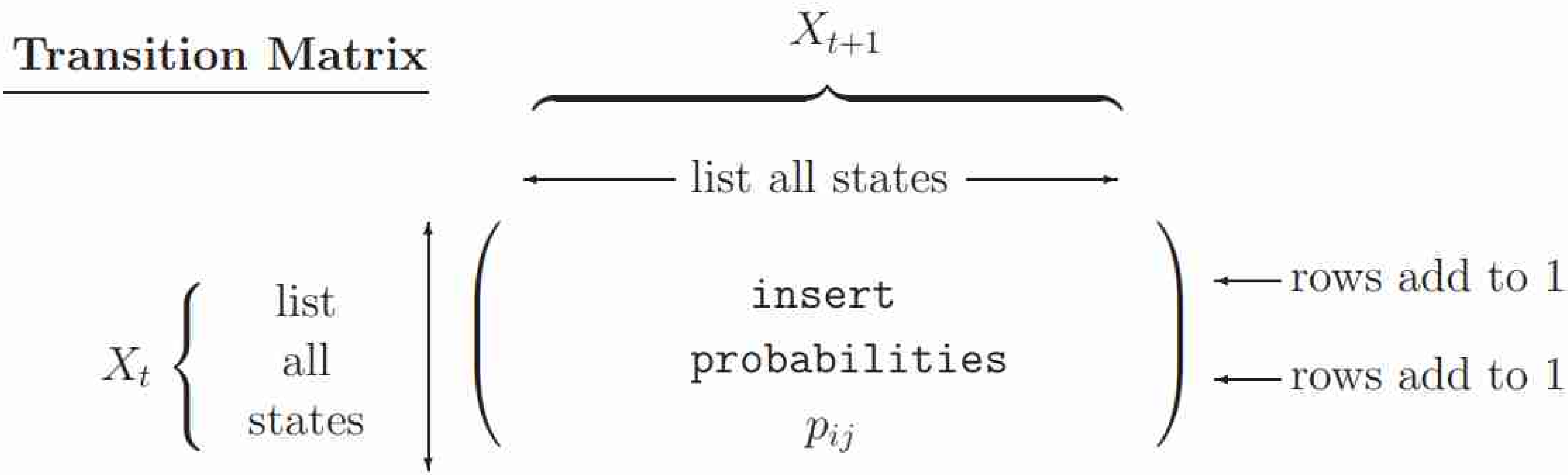
The Transition Matrix



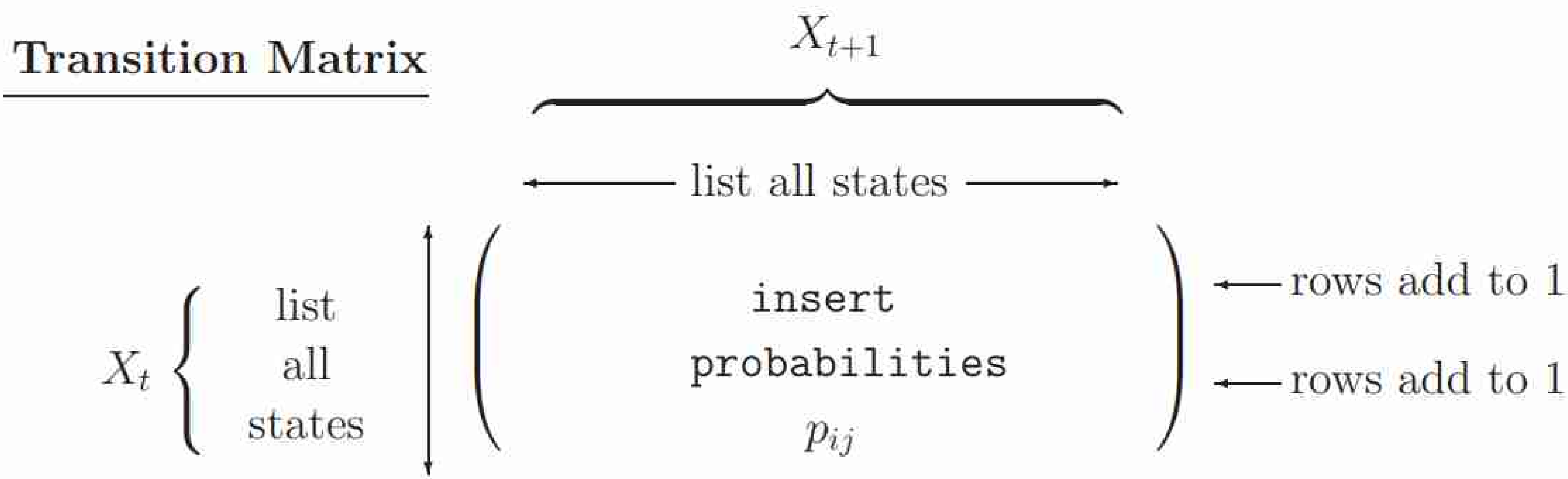
Let us transform this into an equivalent transition matrix which is just another Equivalent way of describing this diagram.



More formally



More formally



The transition matrix is usually given the symbol $P = (p_{ij})$

In the transition matrix P:

the **ROWS** represent **NOW**, or **FROM** X_t

the **COLUMNS** represent **NEXT**, or **TO** X_{t+1}

Matrix entry *i,j* is the **CONDITIONAL** probability that **NEXT =** ***j***, given that **NOW =** ***i***: the probability of going **FROM** state ***i*** **TO** state ***j***.

$p_{ij} = P(X_{t+1} = j \mid X_t = i)$.

A Review of Probabilities

This is not a transition matrix!

Hair color

Eye color

	brown	blonde	Σ
light	5/40	15/40	20/40
dark	15/40	5/40	20/40
Σ	20/40	20/40	**40/40**

A Review of Probabilities

Hair color

Eye color

	brown	blonde	Σ
light	5/40	15/40	20/40
dark	15/40	5/40	20/40
Σ	20/40	20/40	**40/40**

Joint probability: probability of observing both A and B: *Pr(A,B)*
For instance, *Pr(brown, light) = 5/40 = 0.125*

A Review of Probabilities

Hair color

Eye color

	brown	blonde	Σ
light	5/40	15/40	20/40
dark	15/40	5/40	20/40
Σ	20/40	20/40	**40/40**

Marginalize over hair color

Marginal Probability: *unconditional* probability of an observation *Pr(A)*
For instance, *Pr(dark) = Pr(dark,brown) + Pr(dark,blonde) = 15/40 + 5/40 = 20/40 = 0.5*

A Review of Probabilities

Hair color

Eye color

	brown	blonde	Σ
light	5/40	15/40	20/40
dark	15/40	5/40	20/40
Σ	20/40	20/40	**40/40**

Conditional Probability: The probability of observing A given that B has occurred:
Pr(A|B) is the fraction of cases *Pr(B)* in which *B* occurs where *A* also occurs with *Pr(AB)*
$Pr(A|B) = Pr(AB) / Pr(B)$

For instance, $Pr(blonde|light) = Pr(blonde,light) / Pr(light) = (15/40) / (20/40) = 0.75$

A Review of Probabilities

Hair color

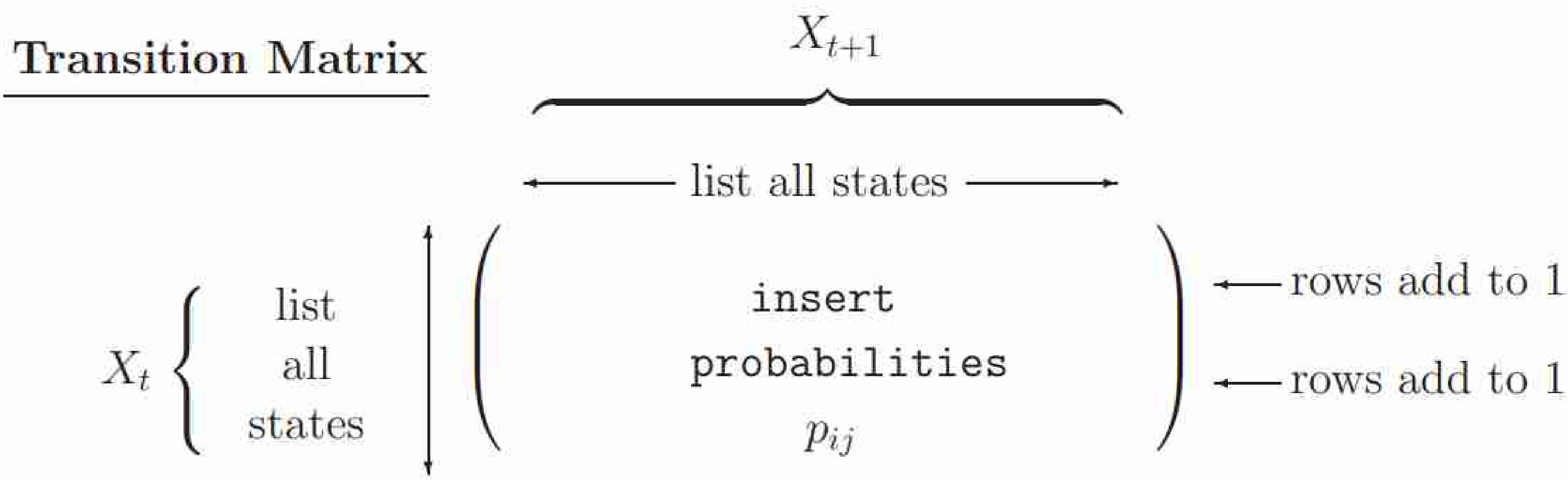
Eye color

	brown	blonde	Σ
light	5/40	15/40	20/40
dark	15/40	5/40	20/40
Σ	20/40	20/40	**40/40**

Statistical Independence: Two events A and B are independent
If their joint probability *Pr(A,B)* equals the product of their marginal probability *Pr(A) Pr(B)*

For instance, *Pr(light,brown) ≠ Pr(light) Pr(brown)*, that is, the events are not independent!

More formally



The transition matrix is usually given the symbol *P* = (p_{ij})

In the transition matrix *P*:

the **ROWS** represent **NOW**, or **FROM** X_t

the **COLUMNS** represent **NEXT**, or **TO** X_{t+1}

Matrix entry *i,j* is the **CONDITIONAL** probability that **NEXT = *j***, given that **NOW = *i***: the probability of going **FROM** state ***i*** **TO** state ***j***.

$p_{ij} = P(X_{t+1} = j \mid X_t = i)$.

Notes

1. The transition matrix *P* must list all possible states in the state space *S*.

2. *P* is a square $N \times N$ matrix, because X_{t+1} and X_t both take values in the same state space *S* of size *N*.

3. The **rows** of *P* should each sum to *1*:

$$\sum_{j=1}^{N} p_{ij} = \sum_{j=1}^{N} \mathbb{P}(X_{t+1} = j \mid X_t = i) = \sum_{j=1}^{N} \mathbb{P}_{\{X_t = i\}}(X_{t+1} = j) = 1.$$

   The above simply states that X_{t+1} must take one of the listed values.

4. The columns of *P* do in general NOT sum to 1.

Notes

This is just another way of writing this conditional probability.

1. The transition matrix *P* must list all possible states in the state space *S*.
2. *P* is a square $N \times N$ matrix, because X_{t+1} and X_t both take values in the same state space *S* of size *N*.
3. The **rows** of *P* should each sum to *1*:

$$\sum_{j=1}^{N} p_{ij} = \sum_{j=1}^{N} \mathbb{P}(X_{t+1} = j \mid X_t = i) = \boxed{\sum_{j=1}^{N} \mathbb{P}_{\{X_t = i\}}(X_{t+1} = j)} = 1.$$

   The above simply states that X_{t+1} must take one of the listed values.

4. The columns of *P* do in general NOT sum to 1.

t-step Transition Probabilites

- Let $\{X_0, X_1, X_2, ...\}$ be a Markov chain with state space $S = \{1, 2, ..., N\}$
- Recall that the elements of the transition matrix P are defined as

  $(P)_{ij} = p_{ij} = P(X_1 = j \mid X_0 = i) = P(X_{n+1} = j \mid X_n = i)$ for any n.

- p_{ij} is the probability of making a transition **FROM** state i **TO** state **j** in a **SINGLE** step
- **Question:** what is the probability of making a transition from state i to state j over two steps? i.e. what is

  $P(X_2 = j \mid X_0 = i)$?

t-step transition probs

$$\mathbb{P}(X_2 = j \mid X_0 = i) \ =$$

Any ideas?

t-step transition probs

$$\begin{aligned}
\mathbb{P}(X_2 = j \mid X_0 = i) &= \sum_{k=1}^{N} \mathbb{P}(X_2 = j \mid X_1 = k)\mathbb{P}(X_1 = k \mid X_0 = i) \\
& \qquad \textit{(Markov Property)} \\
&= \sum_{k=1}^{N} p_{kj} p_{ik} \qquad \textit{(by definitions)} \\
&= \sum_{k=1}^{N} p_{ik} p_{kj} \qquad \textit{(rearranging)} \\
&= (P^2)_{ij}.
\end{aligned}$$

t-step transition probs

$$
\begin{aligned}
\mathbb{P}(X_2 = j \mid X_0 = i) &= \sum_{k=1}^{N} \mathbb{P}(X_2 = j \mid X_1 = k)\mathbb{P}(X_1 = k \mid X_0 = i) \quad \textit{(Markov Property)} \\
&= \sum_{k=1}^{N} p_{kj} p_{ik} \\
&= \sum_{k=1}^{N} p_{ik} p_{kj} \\
&= (P^2)_{ij}.
\end{aligned}
$$

Sum of probabilities (OR!!!) over all possible paths with 1 intermediate state *k* that will take us from *i* to *j*

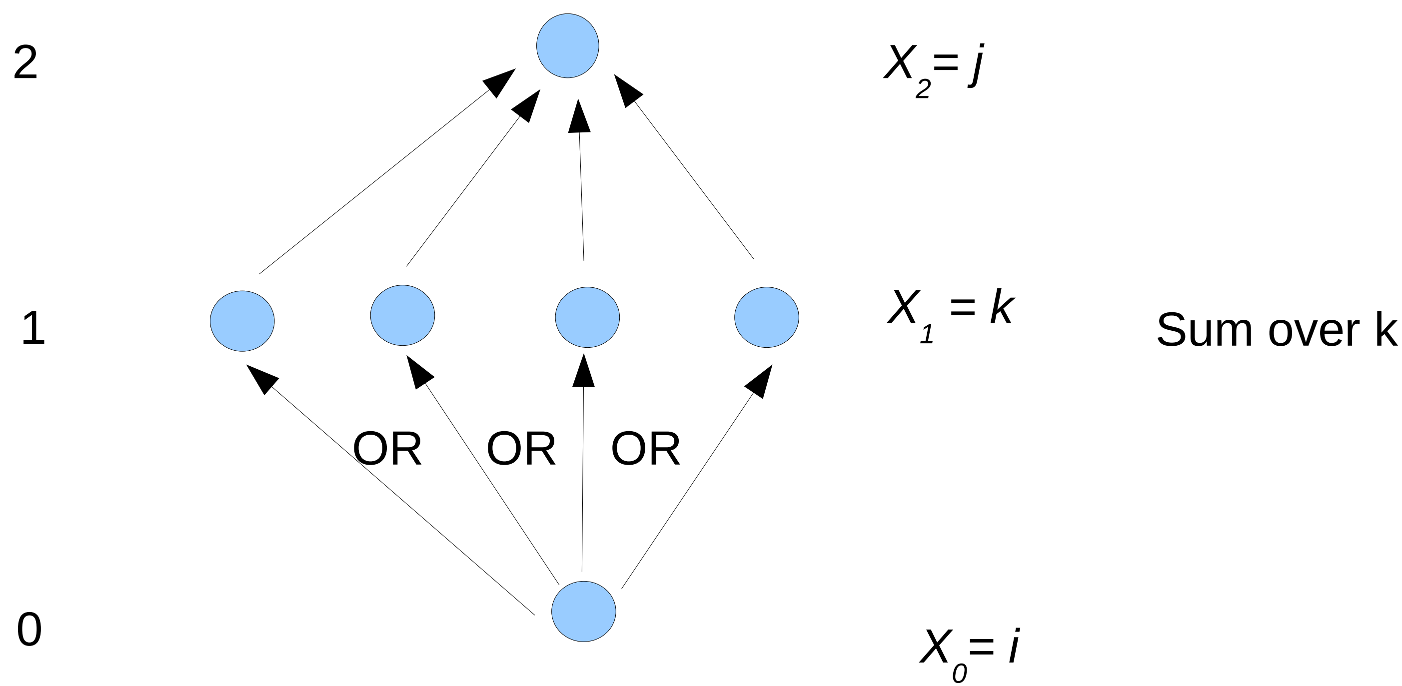
t-step transition probs

$$\begin{aligned}
\mathbb{P}(X_2 = j \mid X_0 = i) &= \sum_{k=1}^{N} \mathbb{P}(X_2 = j \mid X_1 = k)\mathbb{P}(X_1 = k \mid X_0 = i) \\
&\qquad \text{(Markov Property)} \\
&= \sum_{k=1}^{N} p_{kj} p_{ik} \qquad \text{(by definitions)} \\
&= \sum_{k=1}^{N} p_{ik} p_{kj} \qquad \text{(rearranging)} \\
&= (P^2)_{ij}.
\end{aligned}$$

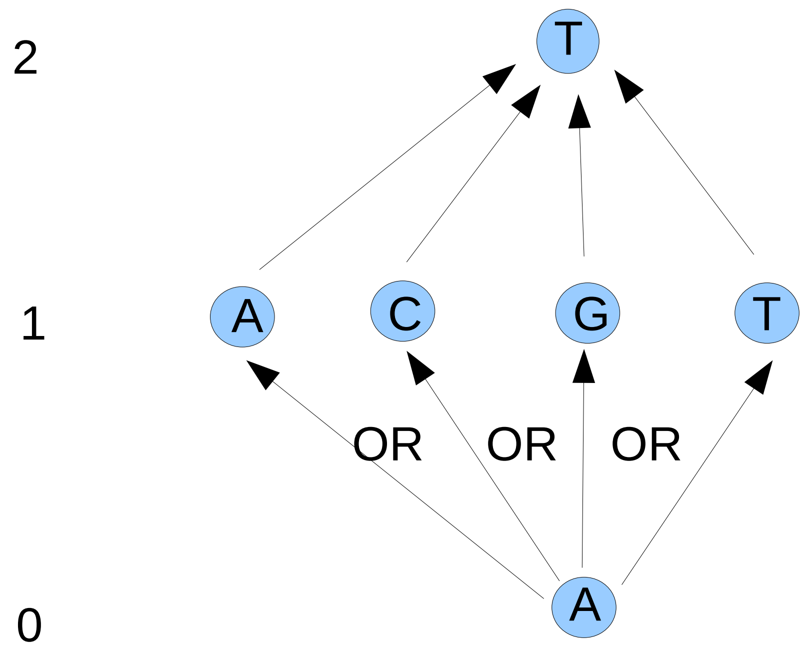
The two step-transition probabilities, in fact, for any n are thus:

$$\mathbb{P}(X_2 = j \mid X_0 = i) = \mathbb{P}(X_{n+2} = j \mid X_n = i) = \left(P^2\right)_{ij}$$

All possible paths



All possible paths



$X_2 = j$

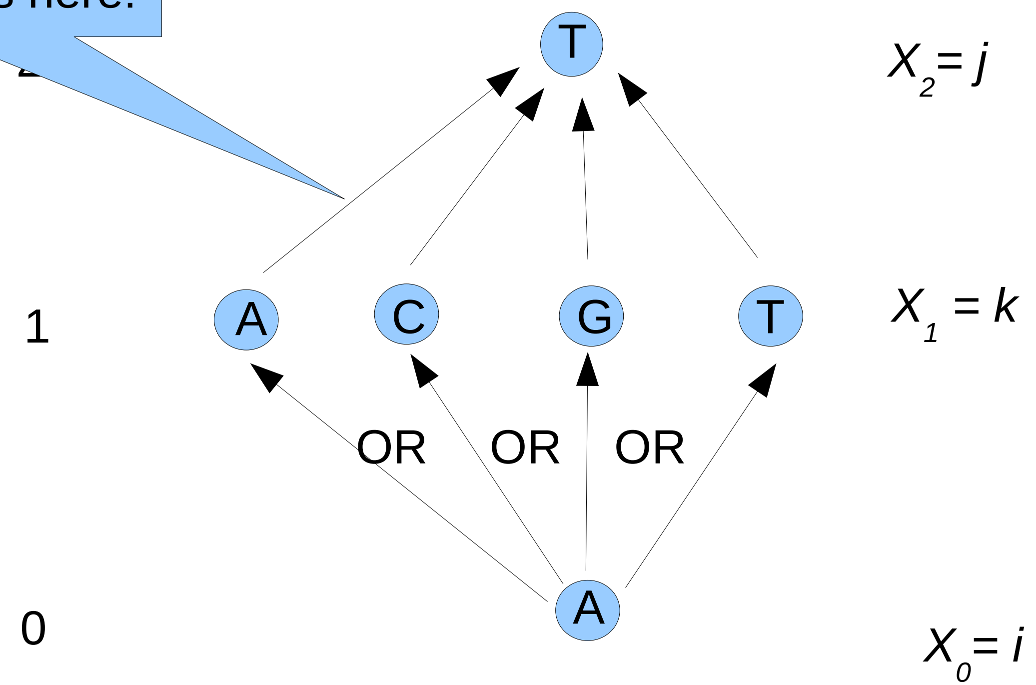
$X_1 = k$

Sum over k

$X_0 = i$

All possible paths

We are still thinking
In discrete steps here!



Sum over k

3-step transitions

- What is: $P(X_3 = j \mid X_0 = i)$?

3-step and *t*-step transitions

- What is: $P(X_3 = j \mid X_0 = i)$?

  $\rightarrow (P^3)_{ij}$

- General case with *t* steps for any *t* **and** any *n*

$$\mathbb{P}(X_t = j \mid X_0 = i) = \mathbb{P}(X_{n+t} = j \mid X_n = i) = \left(P^t\right)_{ij}$$

Distribution of X_t

- Let $\{X_0, X_1, X_2, ...\}$ be a Markov chain with state space $S = \{1, 2, ..., N\}$.
- Now each X_t is a random variable → it has a **probability distribution**.
- We can write down the probability distribution of X_t as vector with N elements.
- For example, consider X_0. Let π be a vector with N elements denoting the probability distribution of X_0.

The π vector

$$\boldsymbol{\pi} = \begin{pmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_N \end{pmatrix} = \begin{pmatrix} \mathbb{P}(X_0 = 1) \\ \mathbb{P}(X_0 = 2) \\ \vdots \\ \mathbb{P}(X_0 = N) \end{pmatrix}$$

This means that our Markov process choses at random in which state (e.g., A, C, G, or T) it **starts** with probability: P(start in state A) = π_A

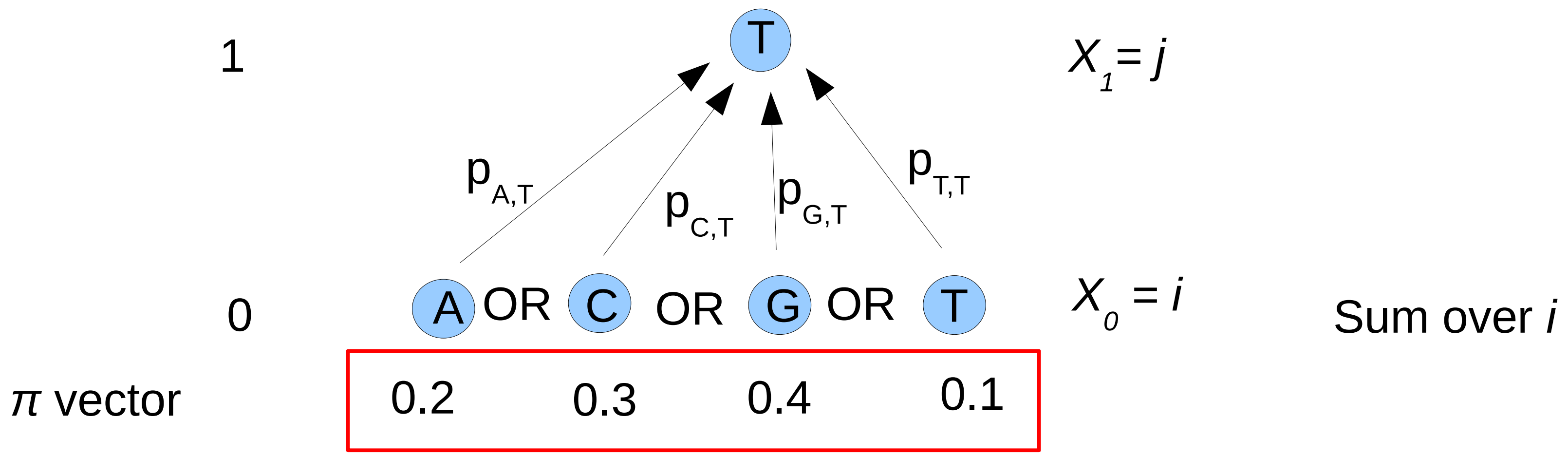
This is why those vectors are also called prior probabilities.

Probability of X_1

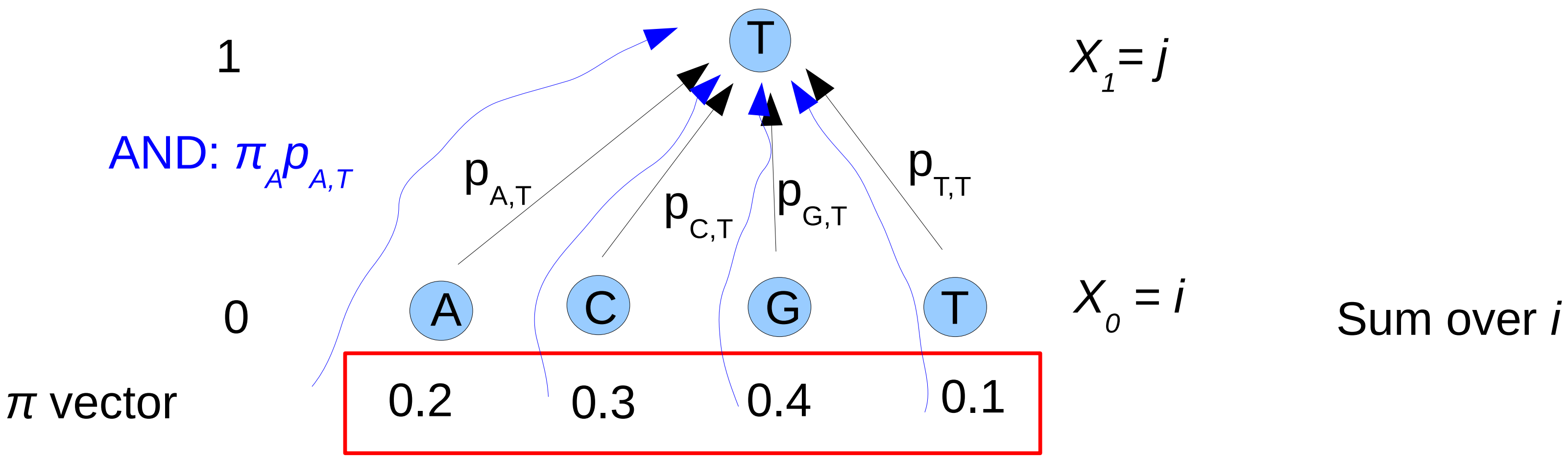
$$\begin{aligned}
\mathbb{P}(X_1 = j) &= \sum_{i=1}^{N} \mathbb{P}(X_1 = j \mid X_0 = i)\mathbb{P}(X_0 = i) \\
&= \sum_{i=1}^{N} p_{ij}\pi_i \quad \textit{by definitions} \\
&= \sum_{i=1}^{N} \pi_i p_{ij} \\
&= \left(\boldsymbol{\pi}^T P\right)_j .
\end{aligned}$$

So, here we are asking what the probability of ending up in state *j* at X_1 is, for starting in all possible states *N* at X_0

All possible paths



All possible paths



Probability Distribution of X_1

$$
\begin{aligned}
\mathbb{P}(X_1 = j) &= \sum_{i=1}^{N} \mathbb{P}(X_1 = j \mid X_0 = i)\mathbb{P}(X_0 = i) \\
&= \sum_{i=1}^{N} p_{ij}\pi_i \quad \textit{by definitions} \\
&= \sum_{i=1}^{N} \pi_i p_{ij} \\
&= \left(\boldsymbol{\pi}^T P\right)_j .
\end{aligned}
$$

This shows that $P(X_1 = j) = \pi^T P_j$ for all j .

The row vector $\pi^T P$ is therefore the probability distribution over all possible states for X_1, more formally:

$X_0 \sim \pi^T$

$X_1 \sim \pi^T P$

Distribution of X_2

- What do you think?

Distribution of X_2

- What do you think?

$$\mathbb{P}(X_2 = j) = \sum_{i=1}^{N} \mathbb{P}(X_2 = j \mid X_0 = i)\mathbb{P}(X_0 = i) = \sum_{i=1}^{N} \left(P^2\right)_{ij} \pi_i = \left(\boldsymbol{\pi}^T P^2\right)_j .$$

and in general:

$$\boxed{\begin{array}{rcl} X_0 & \sim & \boldsymbol{\pi}^T \\ X_1 & \sim & \boldsymbol{\pi}^T P \\ X_2 & \sim & \boldsymbol{\pi}^T P^2 \\ & \vdots & \\ X_t & \sim & \boldsymbol{\pi}^T P^t. \end{array}}$$

Theorem

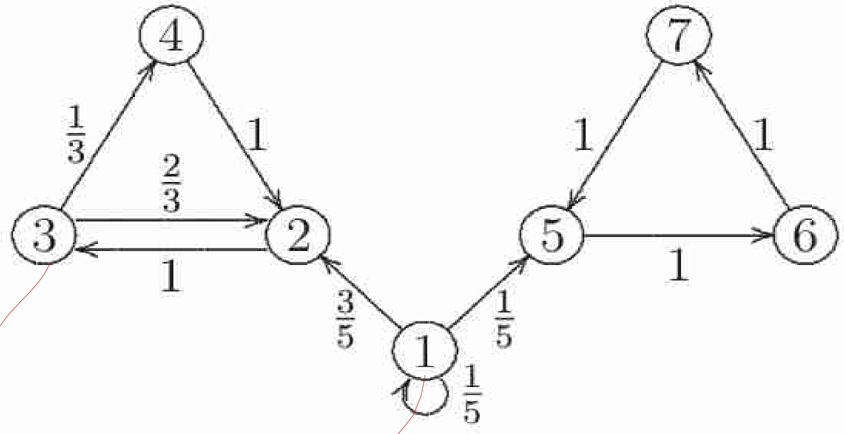
- Let $\{X_0, X_1, X_2, . . .\}$ be a Markov chain with a $N \times N$ transition matrix P.
- If the probability distribution of X_0 is given by the $1 \times N$ row vector π^T, then the probability distribution of X_t is given by the $1 \times N$ row vector $\pi^T P^t$. That is,

  $X_0 \sim \pi^T \Rightarrow X_t \sim \pi^T P^t$.

Example – Trajectory probability

Recall that a trajectory is a sequence of values for $X_0, X_1, \ldots, X_t$.

Because of the Markov Property, we can find the probability of any trajectory by multiplying together the starting probability and all subsequent single-step probabilities.



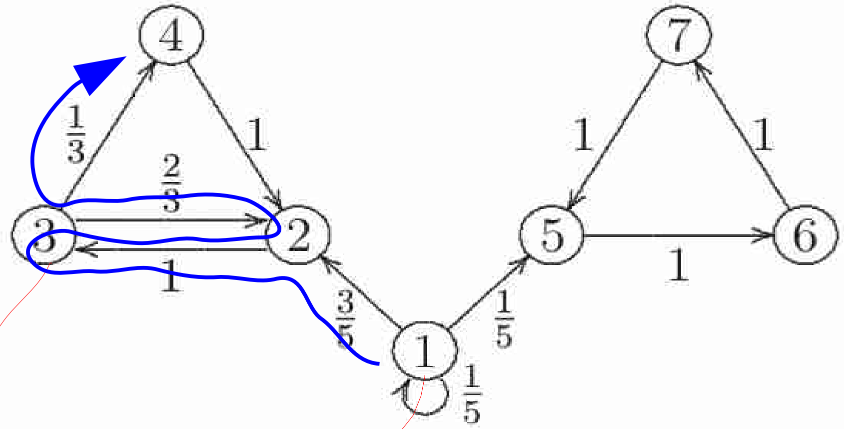
Example: Let $X_0 \sim (\frac{3}{4}, 0, \frac{1}{4}, 0, 0, 0, 0)$. What is the probability of the trajectory 1, 2, 3, 2, 3, 4?

$$
\begin{aligned}
\mathbb{P}(1,2,3,2,3,4) &= \mathbb{P}(X_0 = 1) \times p_{12} \times p_{23} \times p_{32} \times p_{23} \times p_{34} \\
&= \frac{3}{4} \times \frac{3}{5} \times 1 \times \frac{2}{3} \times 1 \times \frac{1}{3} \\
&= \frac{1}{10}.
\end{aligned}
$$

Example – Trajectory probability

Recall that a trajectory is a sequence of values for $X_0, X_1, \ldots, X_t$.

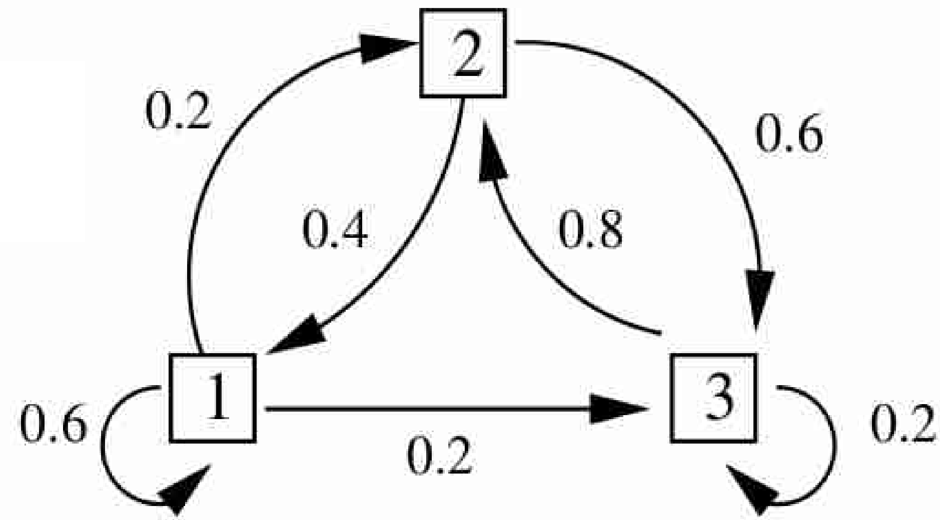
Because of the Markov Property, we can find the probability of any trajectory by multiplying together the starting probability and all subsequent single-step probabilities.



Example: Let $X_0 \sim (\frac{3}{4}, 0, \frac{1}{4}, 0, 0, 0, 0)$. What is the probability of the trajectory 1, 2, 3, 2, 3, 4?

$$\begin{aligned}\mathbb{P}(1,2,3,2,3,4) &= \mathbb{P}(X_0 = 1) \times p_{12} \times p_{23} \times p_{32} \times p_{23} \times p_{34} \\ &= \frac{3}{4} \times \frac{3}{5} \times 1 \times \frac{2}{3} \times 1 \times \frac{1}{3} \\ &= \frac{1}{10}.\end{aligned}$$

Exercise



- Find the transition matrix *P*
- Find $P(X_2=3 \mid X_0=1)$
- Suppose that the process is equally likely to start in any state at time 0
  → Find the probability distribution of X_1
- *Suppose that the process begins in state 1 at time 0*
  → *Find the probability distribution of* X_2
- *Suppose that the process is equally likely to start in any state at time 0*
  → *Find the probability of obtaining the trajectory (3, 2, 1, 1, 3).*

Class Structure

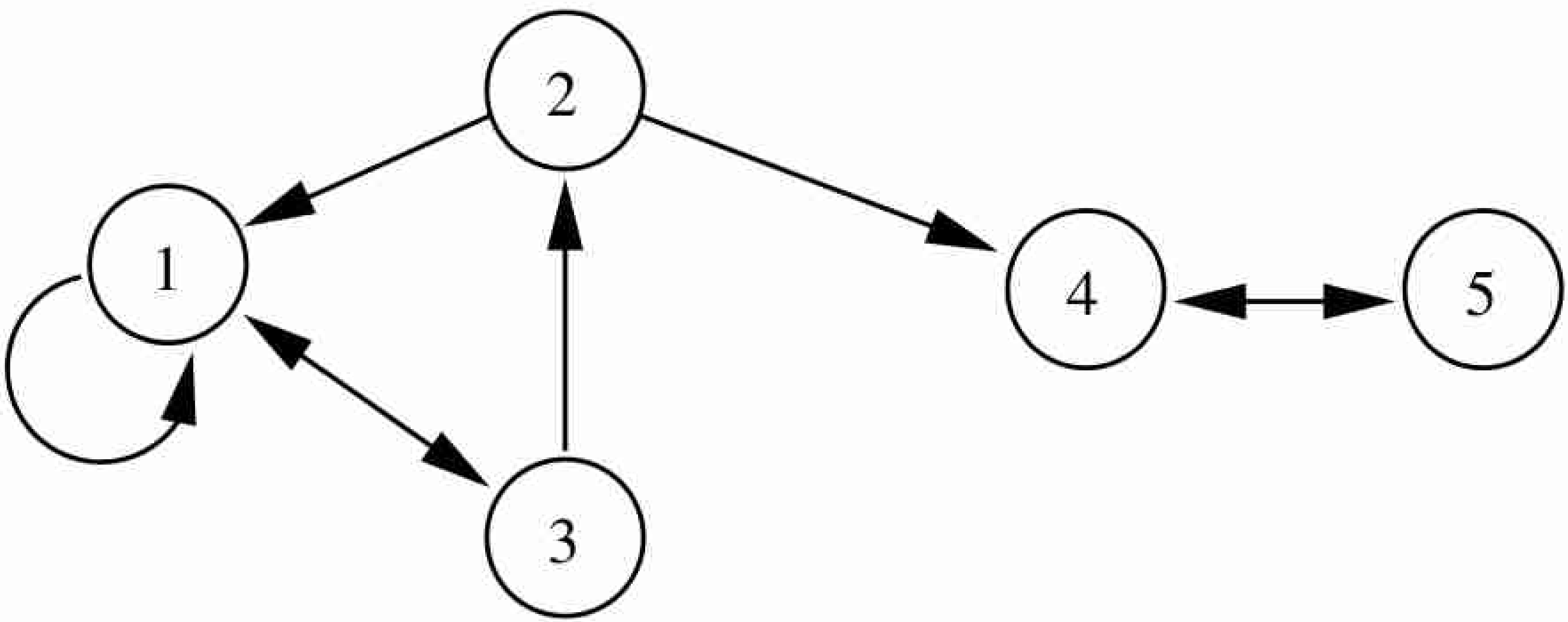
- The state space of a Markov chain can be partitioned into a set of non-overlapping *communicating classes*.
- States i and j are in the same communicating class if there is some way of getting from state $i \rightarrow j$, **AND** there is some way of getting from state $j \rightarrow i$.
- It needn't be possible to get from $i \rightarrow j$ in a single step, but it must be possible over some number of steps to travel between them both ways.
- We write: $i \leftrightarrow j$

Definition

- Consider a Markov chain with state space *S* and transition matrix *P*, and consider states *i, j in S.* Then state *i* communicates with state *j* if:
  - there exists some *t* such that $(P^t)_{ij} > 0$, **AND**
  - there exists some *u* such that $(P^u)_{ji} > 0$.
- Mathematically, it is easy to show that the communicating relation $\leftrightarrow$ is an equivalence relation, which means that it *partitions* the state space *S* into *non-overlapping* equivalence classes.
- **Definition:** States *i* and *j* are in the same communicating class if $i \leftrightarrow j$: i.e., if each state is accessible from the other.
- Every state is a member of *exactly one* communicating class.

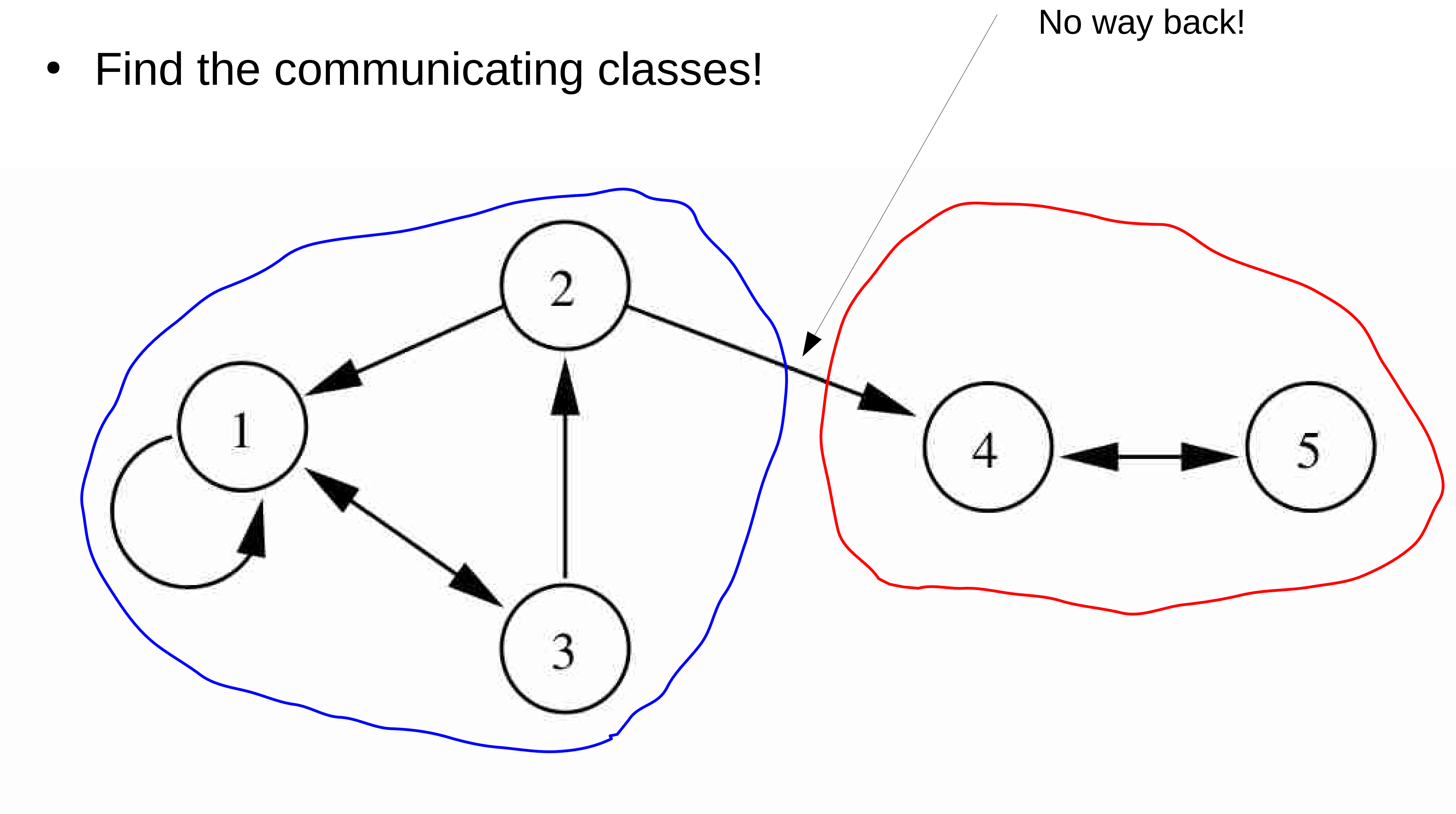
Example

- Find the communicating classes!



Example

- Find the communicating classes!



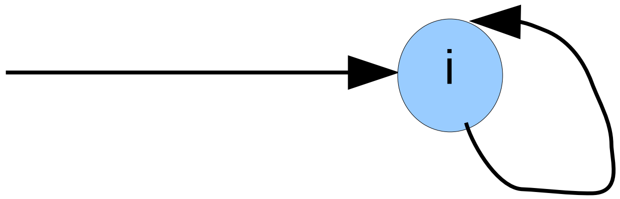
{1, 2, 3} and *{4, 5}*

Properties of Communicating Classes

- **Definition:** A communicating class of states is closed if it is not possible to leave that class.

  That is, the communicating class *C* is closed if $p_{ij} = 0$ whenever *i in C* and *j not in C*

- **Example:** In the transition diagram from the last slide:
  - Class *{1, 2, 3}* is not closed: it is possible to escape to class *{4, 5}*
  - Class *{4, 5}* is closed: it is not possible to escape.
- **Definition:** A state *i* is said to be absorbing if the set *{i}* is a closed class.

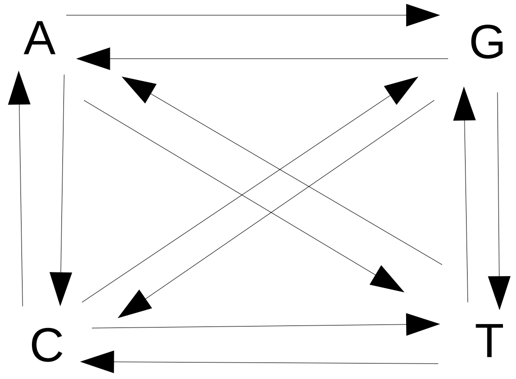


Irreducibility

- **Definition:** A Markov chain or transition matrix P is said to be **irreducible** if $i \leftrightarrow j$ (i communicates with j) for all $i, j \in S$. That is, the chain is irreducible if the state space S is a single communicating class.
- Do you know an example for an irreducible transition matrix P?

Irreducibility

- **Definition:** A Markov chain or transition matrix P is said to be **irreducible** if $i \leftrightarrow j$ for all $i, j \in S$. That is, the chain is irreducible if the state space S is a single communicating class.
- Do you know an example for an irreducible transition matrix P?



Equilibrium

- We saw that if $\{X_0, X_1, X_2, \ldots\}$ is a Markov chain with transition matrix P, then $X_t \sim \pi^T \Rightarrow X_{t+1} \sim \pi^T P$
- **Question:** is there any distribution π at some time t such that $\pi^T P = \pi^T$?
- If $\pi^T P = \pi^T$, *then*

$$X_t \sim \pi^T \quad \Rightarrow X_{t+1} \sim \pi^T P = \pi^T$$

$$\Rightarrow X_{t+2} \sim \pi^T P = \pi^T$$

$$\Rightarrow X_{t+3} \sim \pi^T P = \pi^T$$

$$\Rightarrow \ldots$$

Equilibrium

- We saw that if $\{X_0, X_1, X_2, \ldots\}$ is a Markov chain with transition matrix P, then $X_t \sim \pi^T \Rightarrow X_{t+1} \sim \pi^T P$
- **Question:** is there any distribution π at some time t such that $\pi^T P = \pi^T$?
- If $\pi^T P = \pi^T$, *then*

$$X_t \sim \pi^T \quad \Rightarrow X_{t+1} \sim \pi^T P = \pi^T$$

$$\Rightarrow X_{t+2} \sim \pi^T P = \pi^T$$

$$\Rightarrow X_{t+3} \sim \pi^T P = \pi^T$$

$$\Rightarrow \ldots$$

- In other words, if $\pi^T P = \pi^T$ AND $X_t \sim \pi^T$, then

$$X_t \sim X_{t+1} \sim X_{t+2} \sim X_{t+3} \sim \ldots$$

- Thus, once a Markov chain has reached a distribution π^T such that $\pi^T P = \pi^T$,

it will stay there

Equilibrium

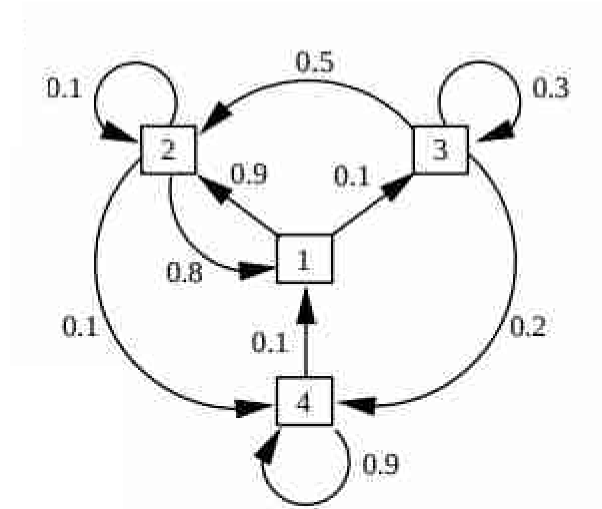
- If $\pi^T P = \pi^T$, we say that the distribution π^T is an equilibrium distribution.
- Equilibrium means there will be no further change in the distribution of X_t as we wander through the Markov chain.
- **Note:** Equilibrium **does not mean** that the actual **value** of X_{t+1} equals the value of X_t
- It means that the distribution of X_{t+1} is the same as the distribution of X_t, e.g.

  $P(X_{t+1} = 1) = P(X_t = 1) = \pi_1$;

  $P(X_{t+1} = 2) = P(X_t = 2) = \pi_2$, etc.

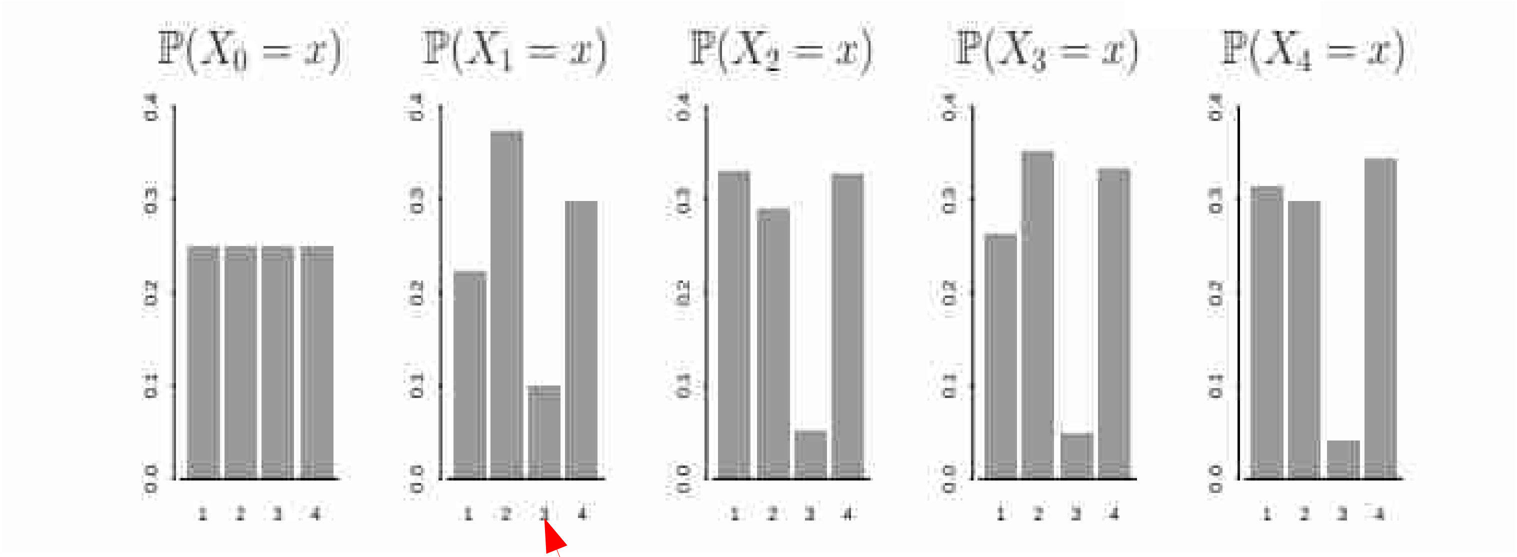
Example

$$P = \begin{pmatrix} 0.0 & 0.9 & 0.1 & 0.0 \\ 0.8 & 0.1 & 0.0 & 0.1 \\ 0.0 & 0.5 & 0.3 & 0.2 \\ 0.1 & 0.0 & 0.0 & 0.9 \end{pmatrix}$$



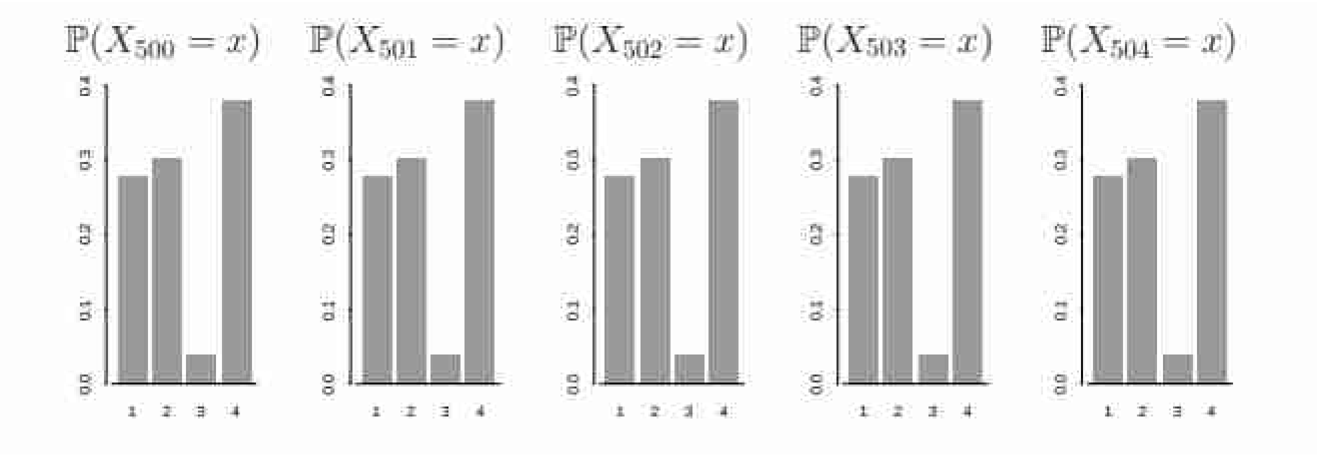
Suppose we start at time *t=0* with $X_0 \sim (¼, ¼, ¼, ¼)$: so the chain is equally likely to start in any of the four states.

First Steps



Probability of being in state 1, 2, 3, or 4

Later Steps



We have reached equilibrium, the chain has forgotten about the initial Probability distribution of *(¼, ¼, ¼, ¼).*

Note: There are several other names for an equilibrium distribution. If π^T is an equilibrium distribution, it is also called:

- **invariant:** it doesn't change π^T
- **stationary:** the chain 'stops' here

Calculating the Equilibrium Distribution

- For the example, we can explicitly calculate the equilibrium distribution by solving $\pi^T P = \pi^T$, under the restriction that:

1. The sum over all entries π_i in vector π^T is *1*
2. All π_i are larger or equal to *0*

- I will spare you the details, the equilibrium frequencies for our example are: *(0.28, 0.30, 0.04, 0.38)*

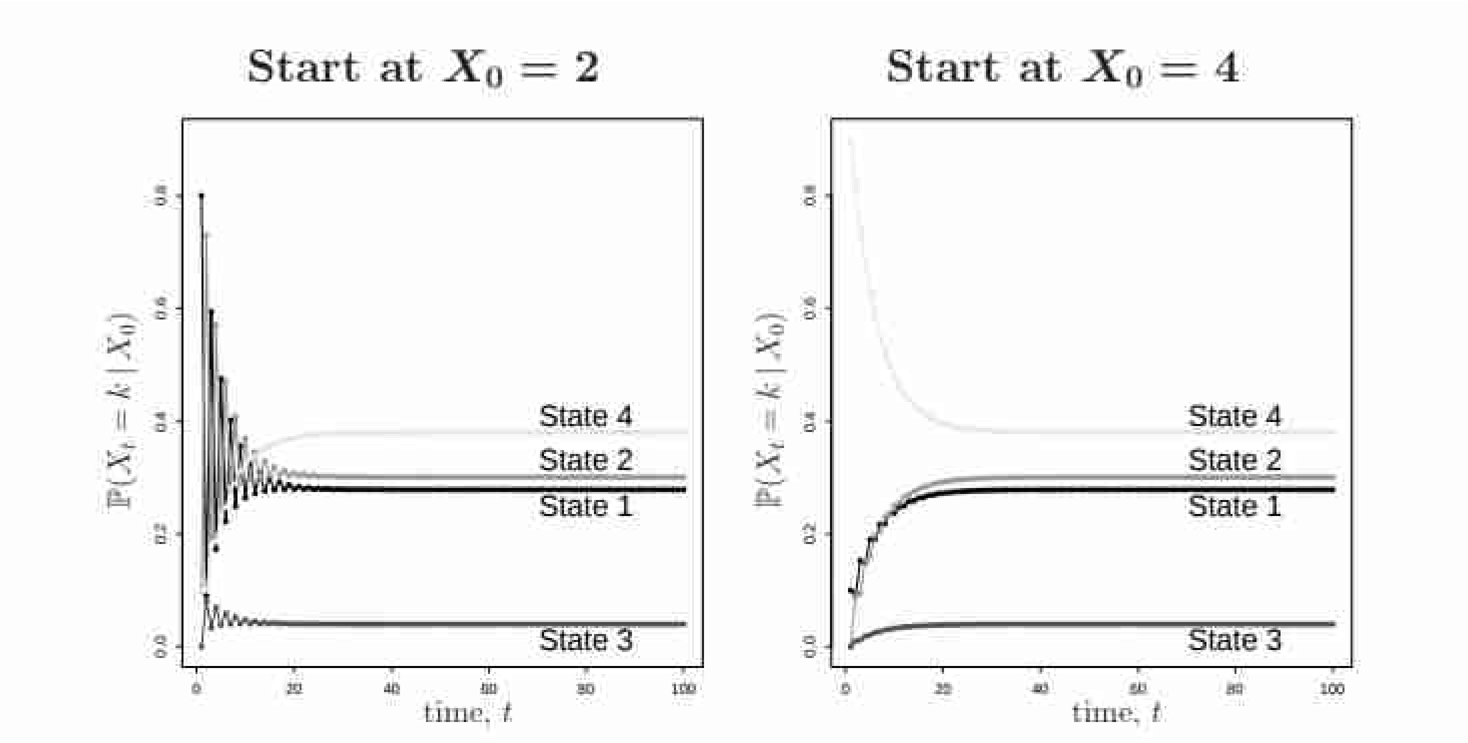
Convergence to Equilibrium

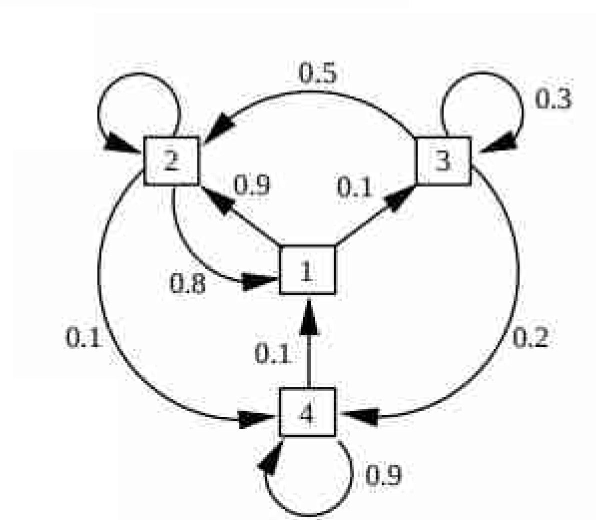
- What is happening here is that each row of the transition matrix Pt converges to the equilibrium distribution (0.28, 0.30, 0.04, 0.38) as t → ∞

$$P = \begin{pmatrix} 0.0 & 0.9 & 0.1 & 0.0 \\ 0.8 & 0.1 & 0.0 & 0.1 \\ 0.0 & 0.5 & 0.3 & 0.2 \\ 0.1 & 0.0 & 0.0 & 0.9 \end{pmatrix} \quad \Rightarrow \quad P^t \to \begin{pmatrix} 0.28 & 0.30 & 0.04 & 0.38 \\ 0.28 & 0.30 & 0.04 & 0.38 \\ 0.28 & 0.30 & 0.04 & 0.38 \\ 0.28 & 0.30 & 0.04 & 0.38 \end{pmatrix} \text{ as } t \to \infty.$$

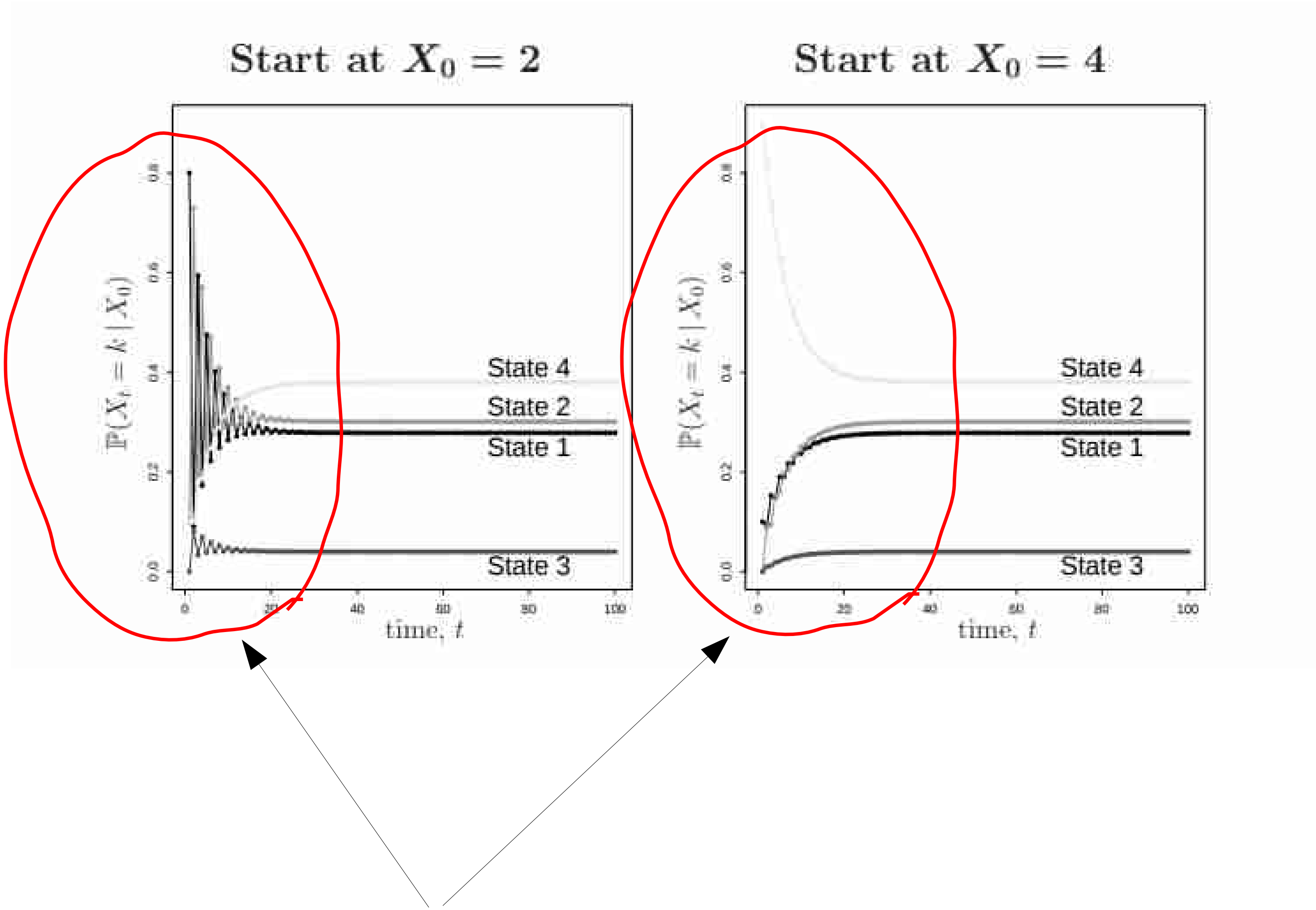
All rows become identical.

Impact of Starting Points



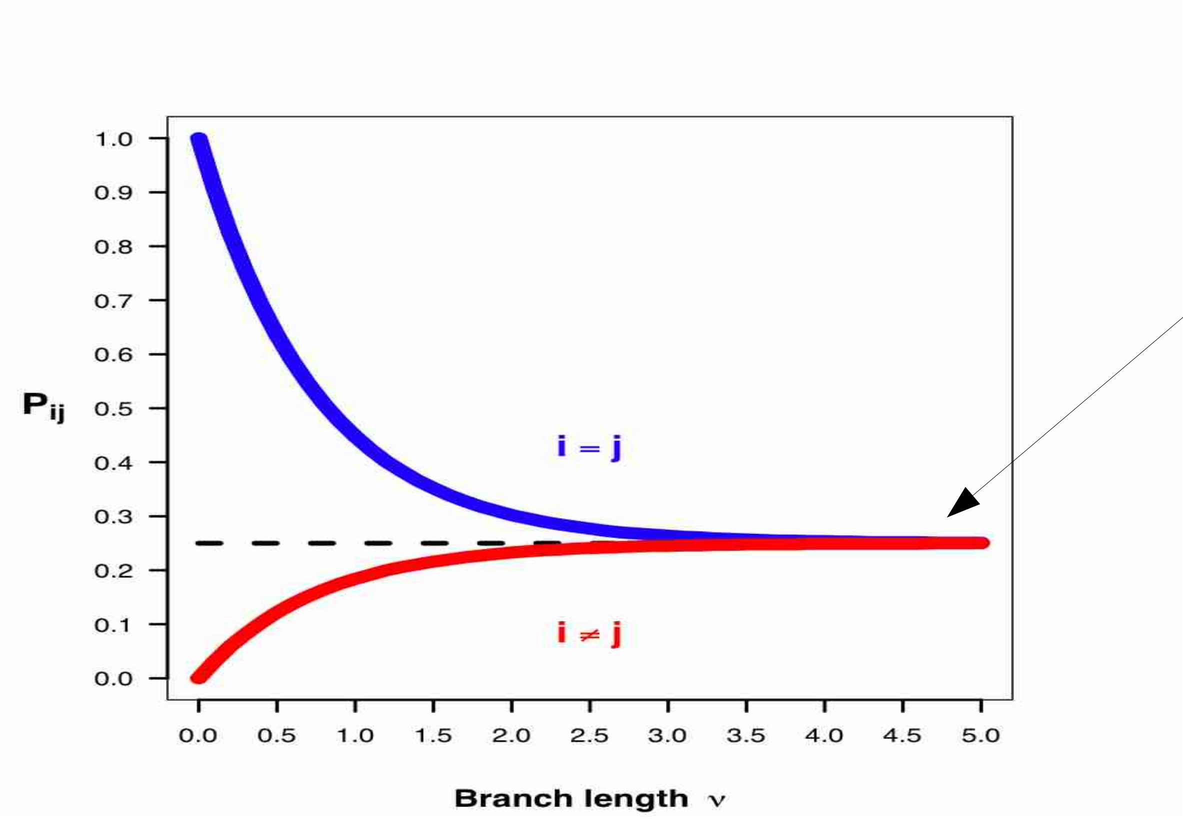


Impact of Starting Points



Initial behavior is different!

Continuous Time Models

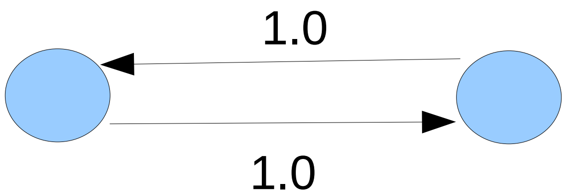


Convergence to stationary distribution of the *Jukes Cantor* Model: *(0.25,0.25,0.25, 0.25)*

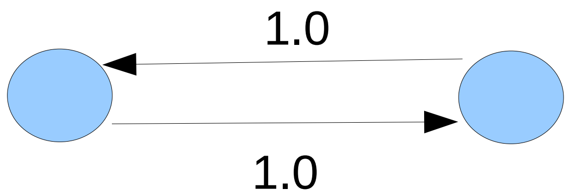
Time steps *t*

Probability of ending in state *j* when starting in state *i* over time (branch length) *v* where *i = j* for the blue curve and i ≠ j for the red one.

Is there always convergence to an equilibrium distribution?



Is there always convergence to an equilibrium distribution?

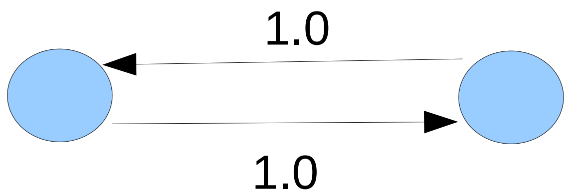


In this example, P^t never converges to a matrix with both rows identical as t becomes large. The chain never 'forgets' its starting conditions as $t \to \infty$.

$$P^t = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \textit{if } t \textit{ is even,}$$

$$P^t = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \textit{if } t \textit{ is odd,}$$

Is there always convergence to an equilibrium distribution?



In this example, P^t never converges to a matrix with both rows identical as t becomes large. The chain never ‘forgets’ its starting conditions as $t \rightarrow \infty$.

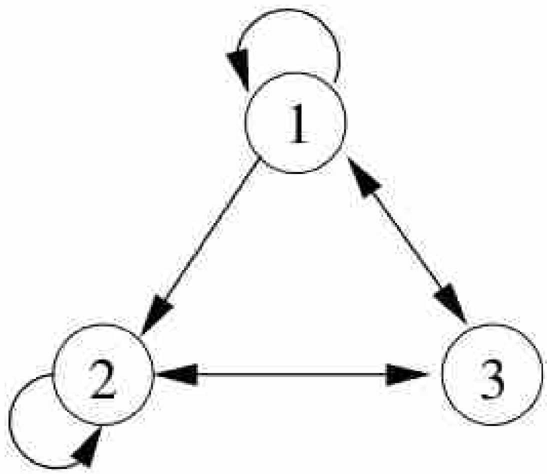
$$P^t = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \textit{if } t \textit{ is even,}$$

$$P^t = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \textit{if } t \textit{ is odd,}$$

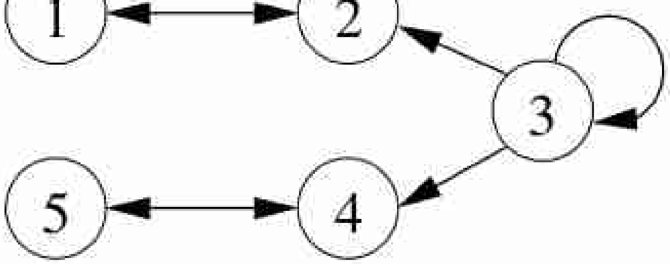
The chain does have an equilibrium distribution $\pi^T = (½, ½)$. However, the chain does not converge to this distribution as $t \rightarrow \infty$.

Convergence

- If a Markov chain is irreducible and aperiodic, and if an equilibrium distribution π^T exists, then the chain converges to this distribution as $t \to \infty$, regardless of the initial starting states.
- Remember: irreducible means that the state space is a single communicating class!



irreducible



non-irreducible

Periodicity

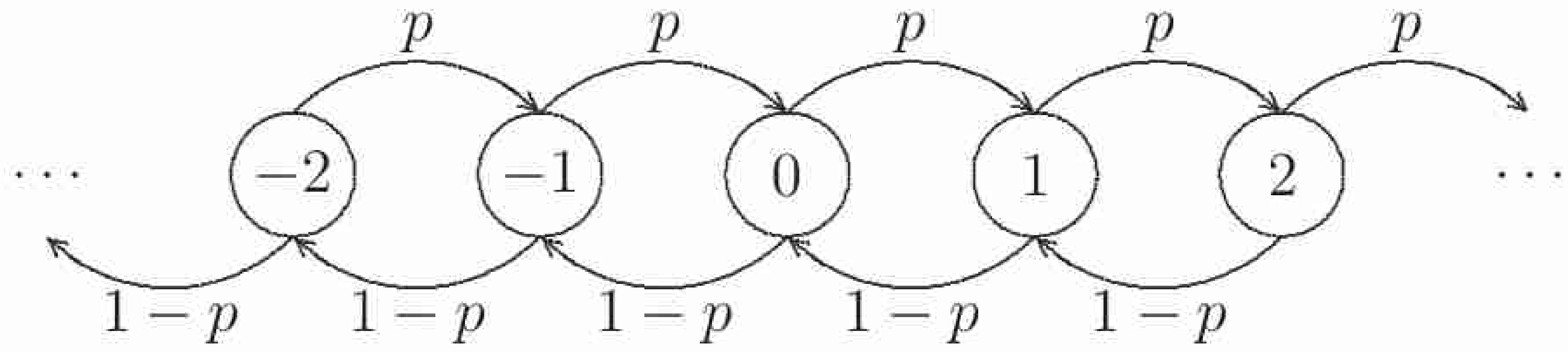
- In general, the chain can return from state i back to state i again in t steps if $(P^t)_{ii} > 0$. This leads to the following definition:
- **Definition:** The period $d(i)$ of a state i is

  $d(i) = gcd\{t : (P^t)_{ii} > 0\}$,

  the greatest common divisor of the times at which return is possible.
- **Definition:** The state i is said to be periodic if $d(i) > 1$

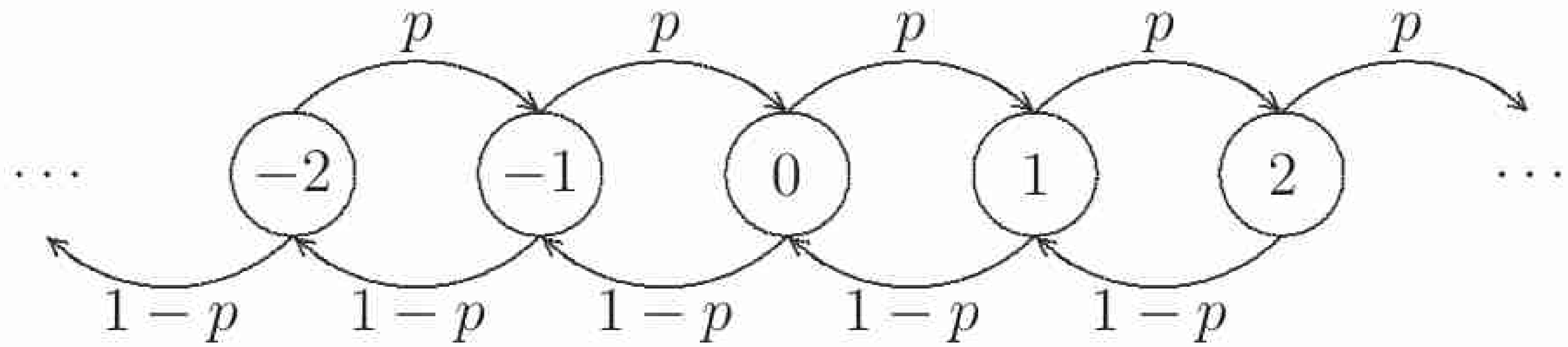
  For a periodic state i, $(P^t)_{ii} = 0$ if t is **not** a multiple of $d(i)$
- **Definition:** The state i is said to be aperiodic if $d(i) = 1$

Example



d(0) = ?

Example



d(0) = gcd{2, 4, 6, …} = 2

The chain is irreducible!

Result

- If a Markov chain is **irreducible** and has one **aperiodic** state, then all states are aperiodic.
- Theorem: Let $\{X_0, X_1, \ldots\}$ be an **irreducible** and **aperiodic** Markov chain with transition matrix P . Suppose that there exists an equilibrium distribution π^T . Then, from any starting state i, and for any end state j,

  $P(X_t = j \mid X_0 = i) \to \pi_j$ as $t \to \infty$.

  In particular,

  $(P^t)_{ij} \to \pi_j$ as $t \to \infty$, for all i and j,

  so P^t converges to a matrix with all rows identical and equal to π^T

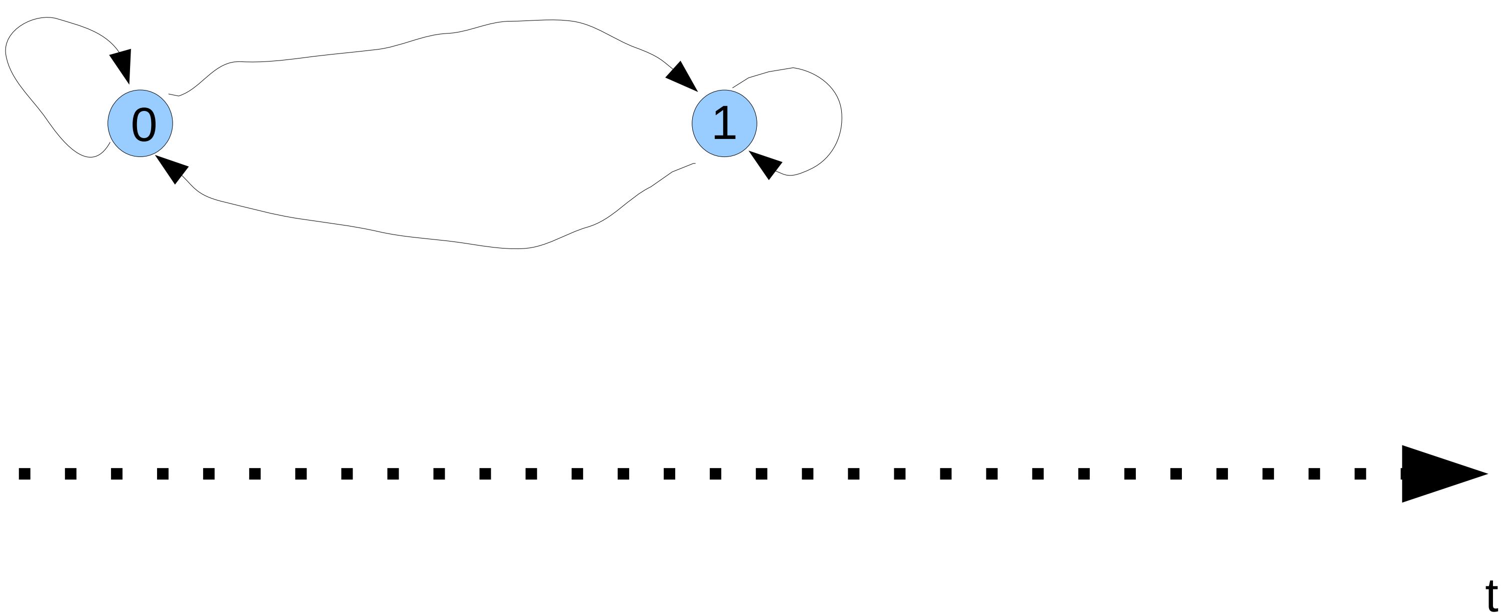
Why?

- The stationary distribution gives information about the stability of a random process.

Continuous Time Markov Chains (CTMC)

- Tranistions/switching between states at random times and not at clock ticks like in a CPU, for example!

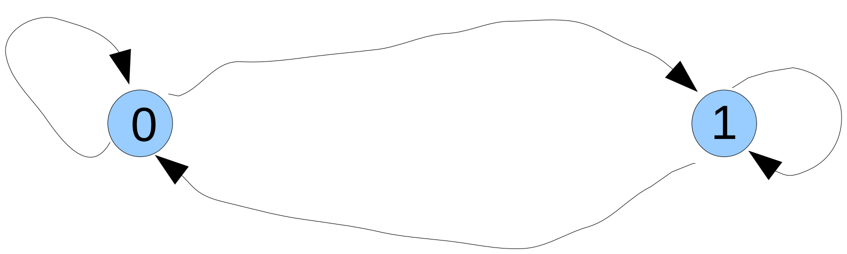
→ no periodic oscillation, concept of waiting times!

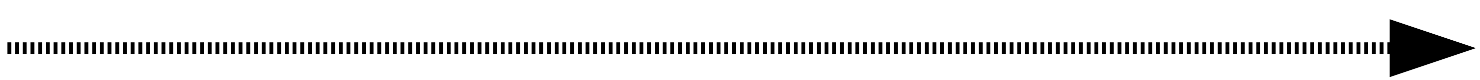


Continuous Time Markov Chains

- Tranistions/switching between states at random times and not at clock ticks like in a CPU, for example!

  → no periodic oscillation, concept of waiting times!



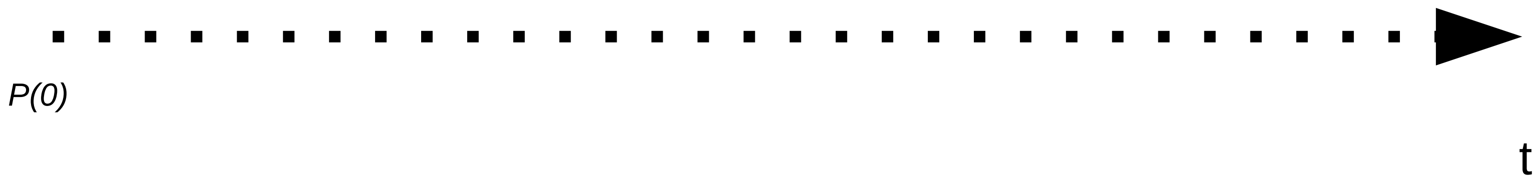


t

Understand what happens as we go toward *dt*

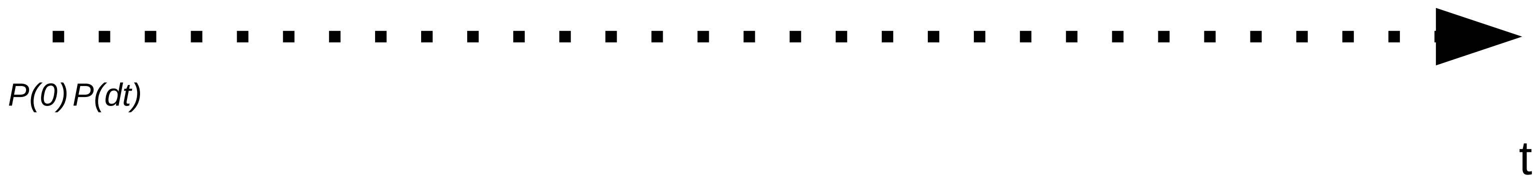
Use Calculus

- Now write the transition probability matrix *P* as a function of time *P(t)*



Use Calculus

- Now write the transition probability matrix *P* as a function of time *P(t)*



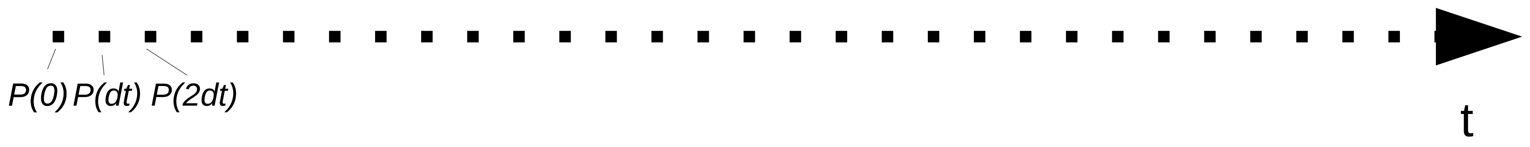
Use Calculus

- Now write the transition probability matrix *P* as a function of time *P(t)*

P(t) is a function that returns a matrix! However, most standard maths on scalar functions can be applied.

Use Calculus

- Now write the transition probability matrix *P* as a function of time *P(t)*



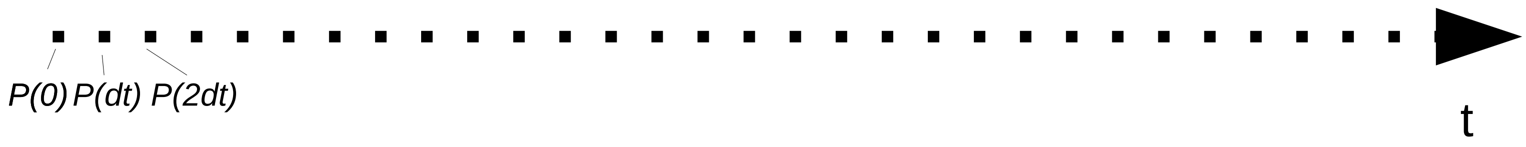
P(t) is a function that returns a matrix! However, most standard maths on scalar functions can be applied.

Derivative: $dP(t) / dt = \lim_{\delta t \to 0} [P(t + \delta t) - P(t)] / \delta t$

Here only *dt* is a scalar value, everything else is a matrix!

Use Calculus

- Now write the transition probability matrix *P* as a function of time *P(t)*



P(t) is a function that returns a matrix! However, most standard maths on scalar Functions can be applied.

Derivative: $dP(t) / dt = \lim_{\delta t \to 0} [P(t + \delta t) - P(t)] / \delta t$

Here only *dt* is a scalar value, everything else is a matrix!

The derivative of a matrix is obtained by individually differentiating all of its entries, the same holds for the limit.

Calculating the limit

- Calculating $lim_{\delta t \to 0}\, [P(t + \delta t) - P(t)] / \delta t$ requires solving a differential equation.
- If we can solve this, then we can calculate *P(t)*
- *Remember, for discrete chains:*

$$\mathbb{P}(X_2 = j \,|\, X_0 = i) \;=\; \sum_{k=1}^{N} \mathbb{P}(X_2 = j \,|\, X_1 = k)\mathbb{P}(X_1 = k \,|\, X_0 = i)$$

This is also known as the **Chapman-Kolmogorov relationship** and can be written differently as

$$P^{n+m} = P^n P^m$$

for any discrete number of steps *n* and *m.*

Calculating the limit

- *Calculating* $\lim_{\delta t \to 0} [P(t + \delta t) - P(t)] / \delta t$ requires solving a differential equation.
- If we can solve this, then we can calculate $P(t)$
- *Remember, for discrete chains:*

$$\mathbb{P}(X_2 = j \mid X_0 = i) = \sum_{k=1}^{N} \mathbb{P}(X_2 = j \mid X_1 = k)\mathbb{P}(X_1 = k \mid X_0 = i)$$

This is also known as the **Chapman-Kolmogorov relationship** and can be written differently as

$P^{n+m} = P^n P^m$

for any discrete number of steps n and m. Thus for continuous time we want: $P(t+h) = P(t)P(h)$

Calculating the limit

$$lim_{\delta t \to 0} \, [P(t + \delta t) - P(t)] / \delta t$$

$$lim_{\delta t \to 0} \, [P(t)P(\delta t) - P(t)] / \delta t$$

$$lim_{\delta t \to 0} \, [P(t)(P(\delta t) - I)] / \delta t$$

Identity matrix, analogous to *1* in the scalar case

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

2 x 2

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

3 x 3

Calculating the limit

$$\lim_{\delta t \to 0} [P(t + \delta t) - P(t)] / \delta t$$

↓

$$\lim_{\delta t \to 0} [P(t)P(\delta t) - P(t)] / \delta t$$

↓

$$\lim_{\delta t \to 0} [P(t)(P(\delta t) - I)] / \delta t$$

↓ The limit doesn't depend on *P(t)*!

$$P(t) \lim_{\delta t \to 0} (P(\delta t) - I) / \delta t$$

Calculating the limit

$$\lim_{\delta t \to 0} [P(t + \delta t) - P(t)] / \delta t$$

$$\lim_{\delta t \to 0} [P(t)P(\delta t) - P(t)] / \delta t$$

$$\lim_{\delta t \to 0} [P(t)(P(\delta t) - I)] / \delta t$$

The limit doesn't depend on *P(t)*!

$$P(t) \lim_{\delta t \to 0} (P(\delta t) - I) / \delta t$$

This is the famous *Q* matrix

Calculating the limit

$$\lim_{\delta t \to 0} [P(t + \delta t) - P(t)] / \delta t$$

$$\lim_{\delta t \to 0} [P(t)P(\delta t) - P(t)] / \delta t$$

$$\lim_{\delta t \to 0} [P(t)(P(\delta t) - I)] / \delta t$$

The limit doesn't depend on *P(t)*!

$$P(t) \lim_{\delta t \to 0} (P(\delta t) - I) / \delta t$$

This is the famous *Q* matrix

The values of *Q* can be anything, but rows must sum to *0*. Remember that rows of *P* must sum to *1*.

What we have so far

$dP(t)/dt = P(t)Q$

Q is also called the **jump rate matrix**, or **instantaneous transition matrix**

Now, imagine that $P(t)$ is a scalar function and Q just some scalar constant:

$P(t) = exp(Qt)$

the same holds for matrices.

What we have so far

$dP(t)/dt = P(t)Q$

Q is also called the **jump rate matrix**, or **instantaneous transition matrix**

Now, imagine that $P(t)$ is a scalar function and Q just some scalar constant:

$P(t) = exp(Qt)$

the same holds for matrices.

However calculating a matrix exponential is not trivial, it's not just taking the exponential of each of its elements!

$exp(Qt) = I + Qt + 1/2!\ Q^2t^2 + 1/3!\ Q^3t^3 + \ldots$

$$P(t)=e^{Qt}$$

- There is no general solution to analytically calculate this matrix exponential, it depends on Q.
- In some cases we can come up with an analytical equation, like for the *Jukes Cantor* model
- For the GTR model we already need to use creepy numerical methods (Eigenvector/Eigenvalue) decomposition, we might see that later
- For non-reversible models it gets even more nasty

Equilibrium Distribution

- Assume there exists a row vector π^T such that $\pi^T Q = 0$

  → *π^T is the equilibrium distribution*