MARKOV PROCESSES

Valerio Di Valerio

Stochastic Process

- Definition: a stochastic process is a collection of random variables $\{X(t)\}$ indexed by time $t \in T$

- Each $X(t) \in X$ is a random variable that satisfy some probability law
- X is usually called the *state space* of the process
- A realization of a stochastic process (sample path) is a *specific* sequence $X(t_0) = x_0, X(t_1) = x_1, \ldots$

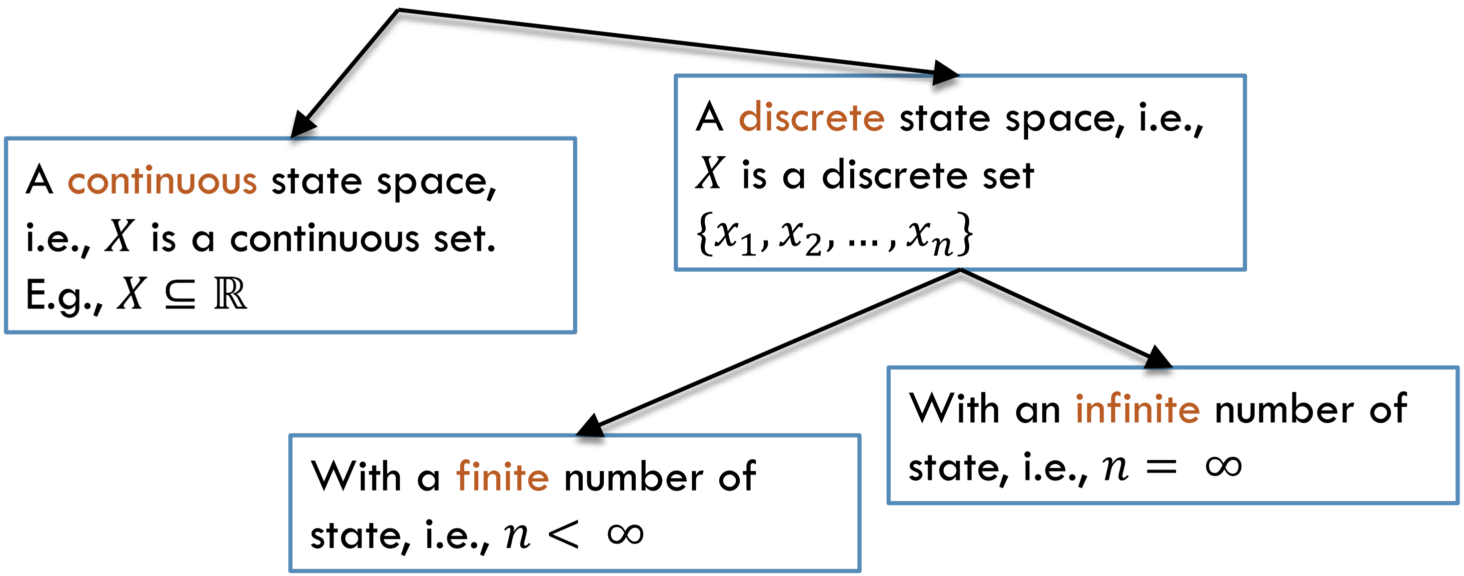
Stochastic Process

- Example: toss a coin an infinite number of times, i.e., $t = 1, 2, 3, \ldots$
- $X = \{\text{Head, Tail}\}$
- Sample path:

t	X(t)
1	Head
2	Head
3	Tail
4	Head

A Simple Classification

- A stochastic process can have:



- The process can be either continuous time, $T = [0, \infty)$, or discrete time ($T = \mathbb{N}$)

Examples

- Example 1: the process represents the number of people queued at the post office
  - $X = \{1, \ldots, \infty\}$ discrete state space
  - $T = \mathbb{R}^+ \cup 0$ continuous time
- Example 2: height of a person on his/her birthday
  - $X = \mathbb{R}$ continuous state space
  - $T = \{1, 2, \ldots\}$ discrete time

Stochastic Process Dynamics

- The process *dynamics* can be defined using the transition probabilities
- They specify the stochastic evolution of the process through its states
- For a discrete time process, transition probabilities can be defined as follows

$$P(X_{k+1} = x_{k+1} | X_k = x_k, X_{k-1} = x_{k-1}, \dots, X_0 = x_0)$$

Stochastic Process Dynamics

- Example: we have a bag with 20 balls.
  - 10 are red and 10 are blue
- At time any $t = 1,2, \ldots, n$, we draw a ball from the bag, without replacements
- Question: what is $P(X_1 = r)$?
- Question: what is $P(X_2 = r \mid X_1 = r)$?
- Question: what is $P(X_3 = b \mid X_2 = r, X_1 = r)$?

Markov Property

- The term **Markov property** refers to the memoryless property of a stochastic process:
- For a discrete time process, the Markov property is defined as:

$$P(X_{k+1} = x_{k+1} | X_k = x_k, X_{k-1} = x_{k-1}, \dots, X_0 = x_0) = \boldsymbol{P(X_{k+1} = x_{k+1} \mid X_k = x_k)}$$

- **Definition:** a stochastic process that satisfies the Markov property is called Markov process
- If the state space is discrete, we refers to these processes as Markov Chains

Time-homogeneous Markov chains

- A Markov chain is time-homogeneous if transition probabilities are time-independent

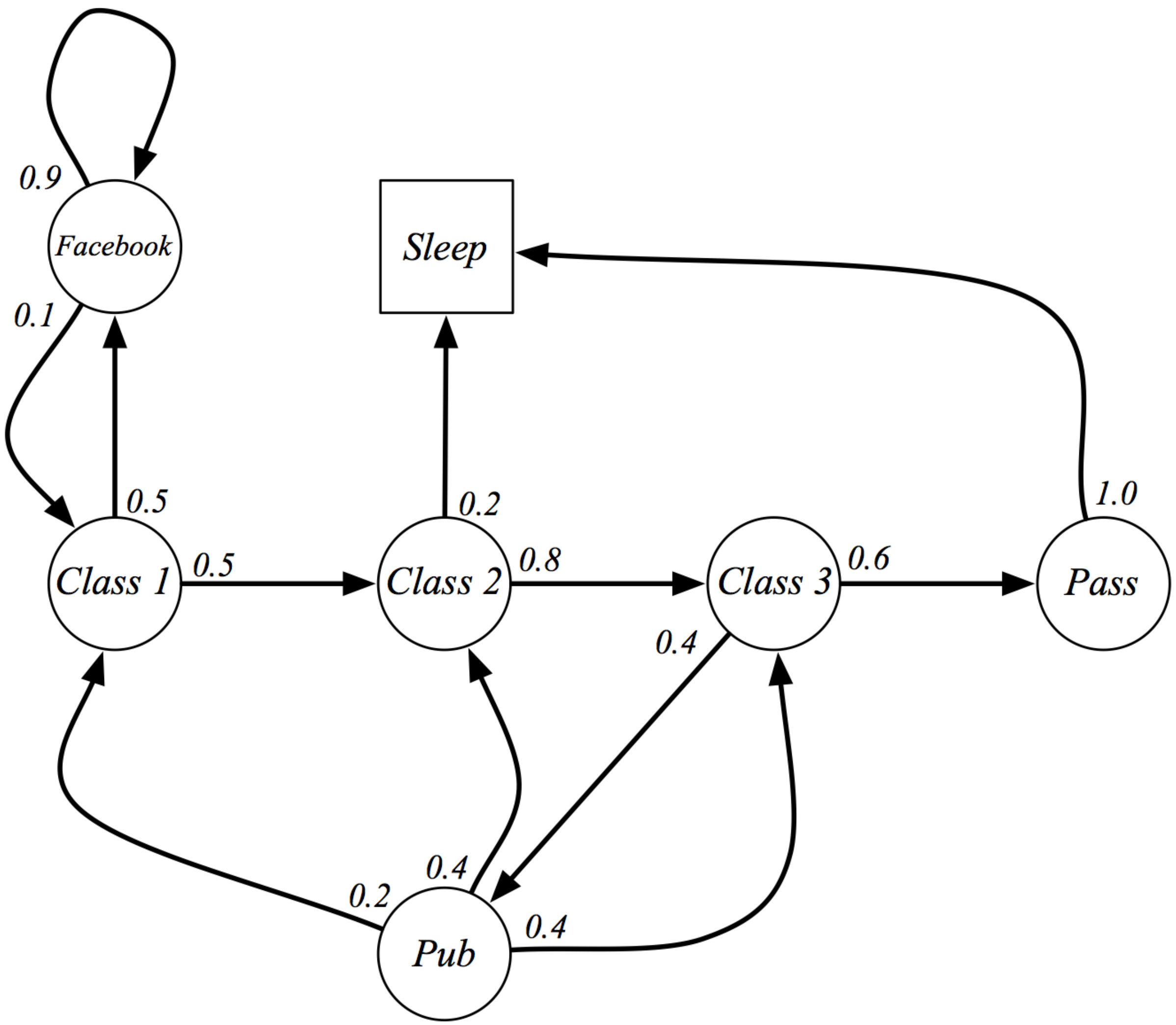
  $P(X_{k+1} = x_{k+1} \mid X_k = x_k)$ is the same for all k

- If the state space is discrete and finite, transition probabilities are usually represented using a matrix…

$$P = \begin{bmatrix} p_{1,1} & \cdots & p_{n,1} \\ \vdots & \ddots & \vdots \\ p_{n,1} & \cdots & p_{n,n} \end{bmatrix}$$

- …and the Markov chain can be easily represented using a graph!

Example: Student Markov Chain



Transitory Analysis of a Markov Chain

- We can define the state probability as
$$\pi_j(k) = P(X_k = j)$$
- Definition: it is the probability of finding the process in state j at time k
- Simple theory allows us to compute "next step" probabilities as

$$\pi_j(k+1) = \sum_{i \in X} P(X_{k+1} = j \mid X_k = i) \cdot \pi_i(k)$$

Transitory Analysis of a Markov Chain

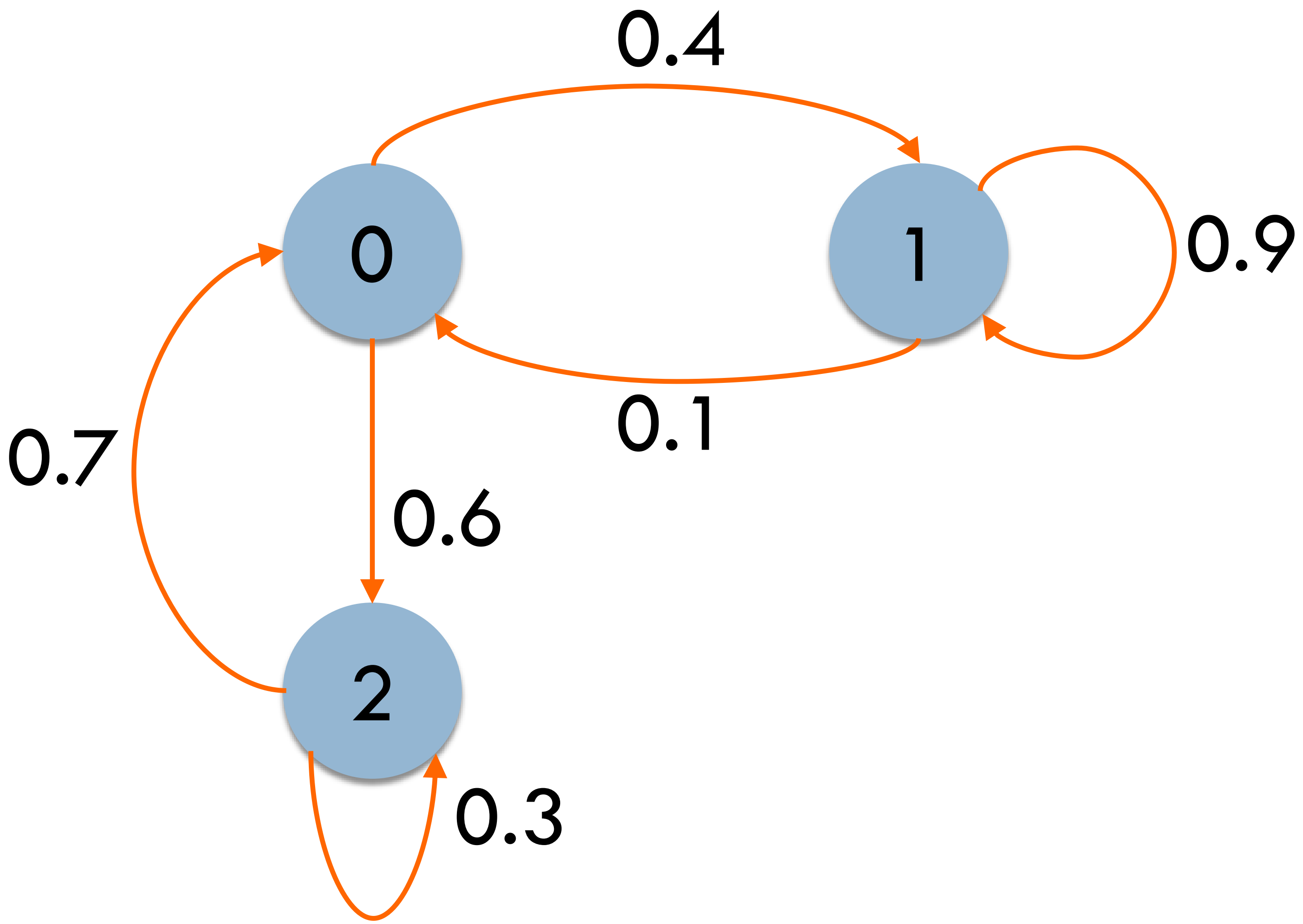
- If we consider all states, we can use the vector
  $$\pi(k) = [\pi_0(k), \pi_1(k), \pi_2(k), \dots]$$
- In matrix notation it becomes

$$\pi(k+1) = \pi(k) \cdot P$$

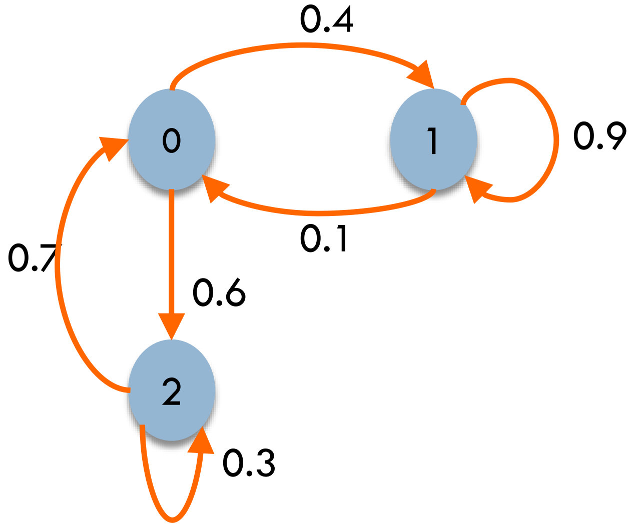
- But, if we know initial probabilities $\pi(0)$, then

$$\boldsymbol{\pi}(\boldsymbol{k}+\mathbf{1}) = \boldsymbol{\pi}(\mathbf{0}) \cdot \boldsymbol{P}^{\boldsymbol{k}}$$

A simple example



A simple example

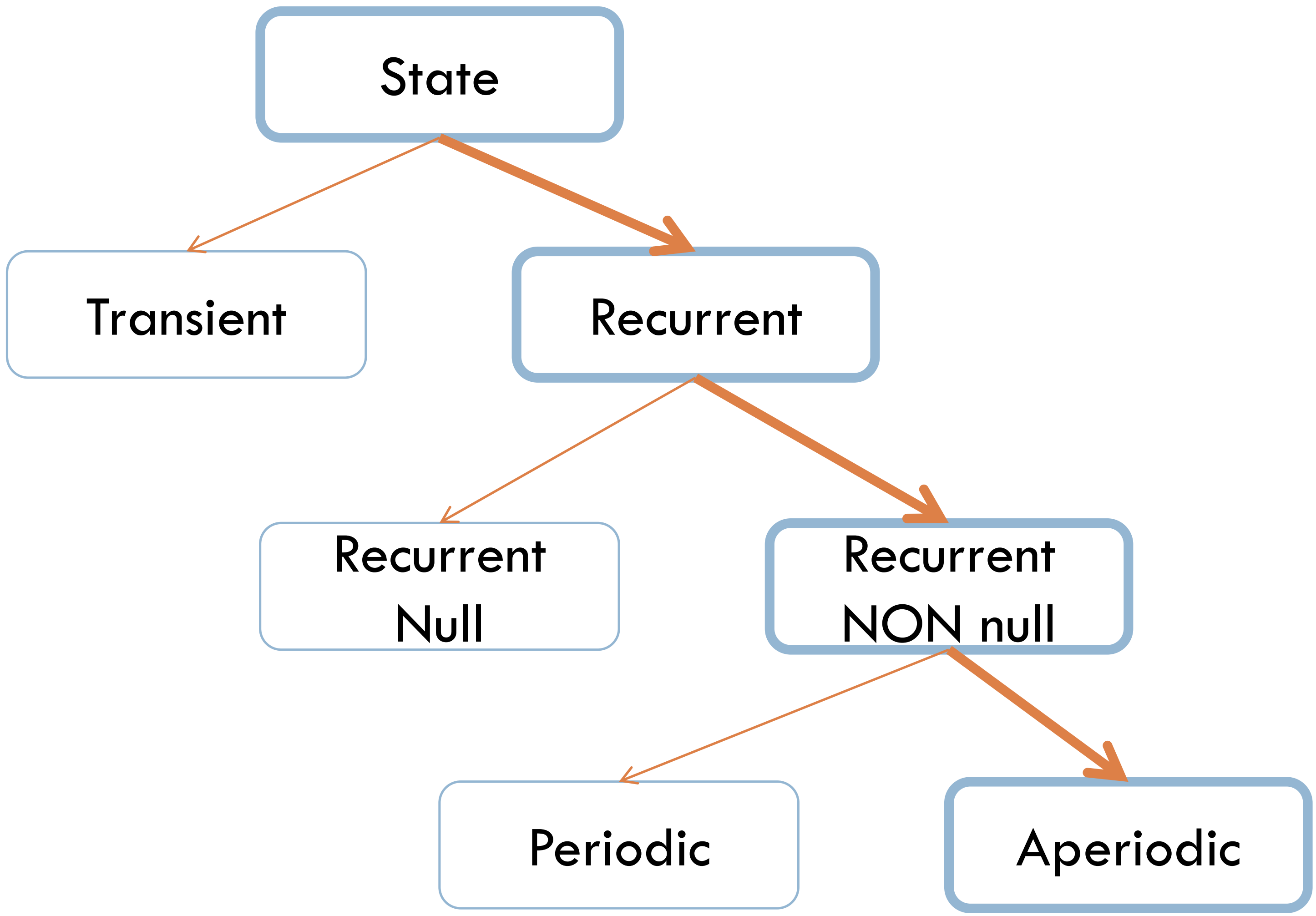


Transition Probabilities

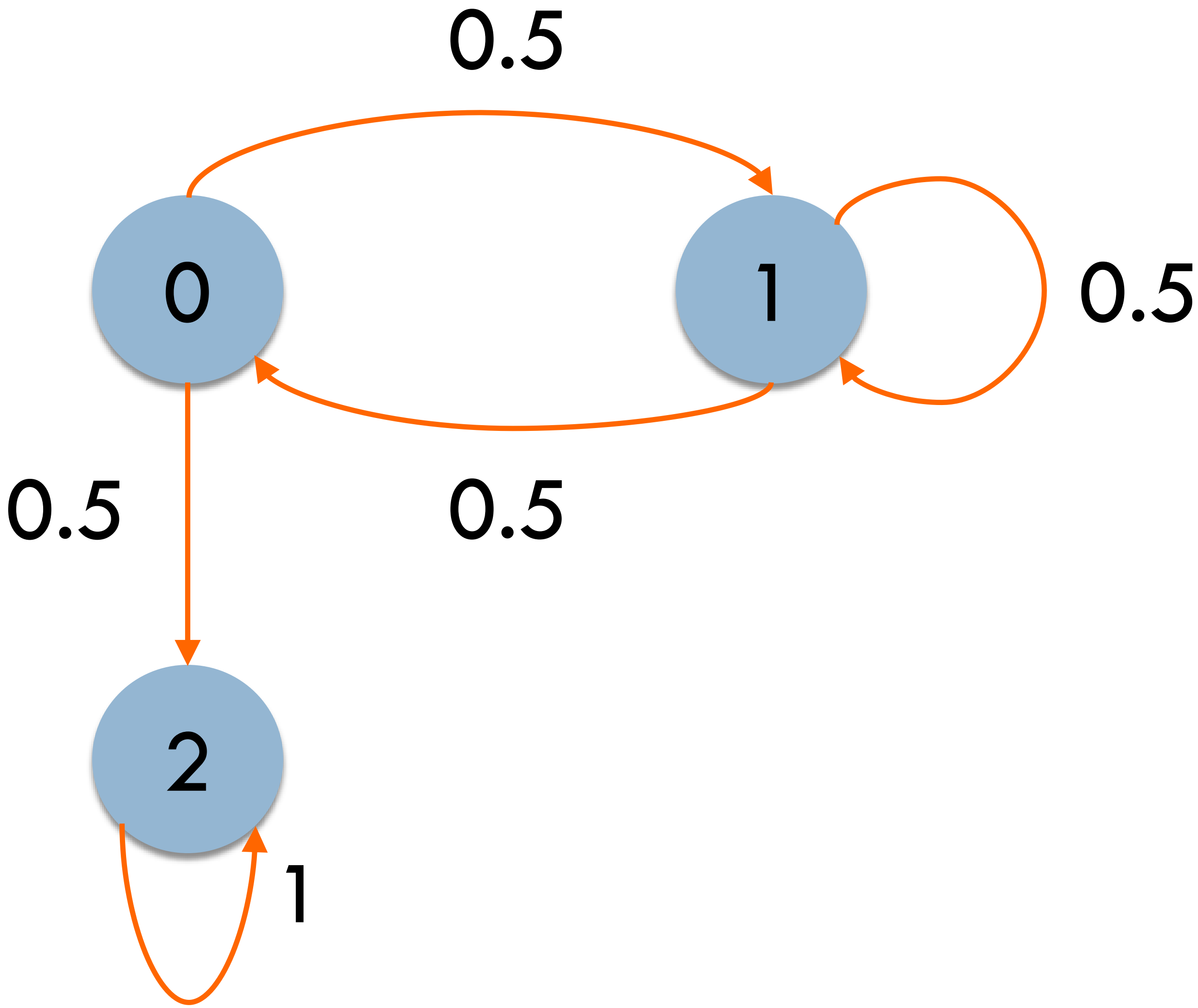
$$P = \begin{bmatrix} 0 & 0.4 & 0.6 \\ 0.1 & 0.9 & 0 \\ 0.7 & 0 & 0.3 \end{bmatrix}$$

$$\begin{cases} \pi_0(k+1) = 0.1\pi_1(k) + 0.7\pi_2(k) \\ \pi_1(k+1) = 0.4\pi_0(k) + 0.9\pi_1(k) \\ \pi_2(k+1) = 0.6\pi_0(k) + 0.3\pi_2(k) \\ \end{cases}$$

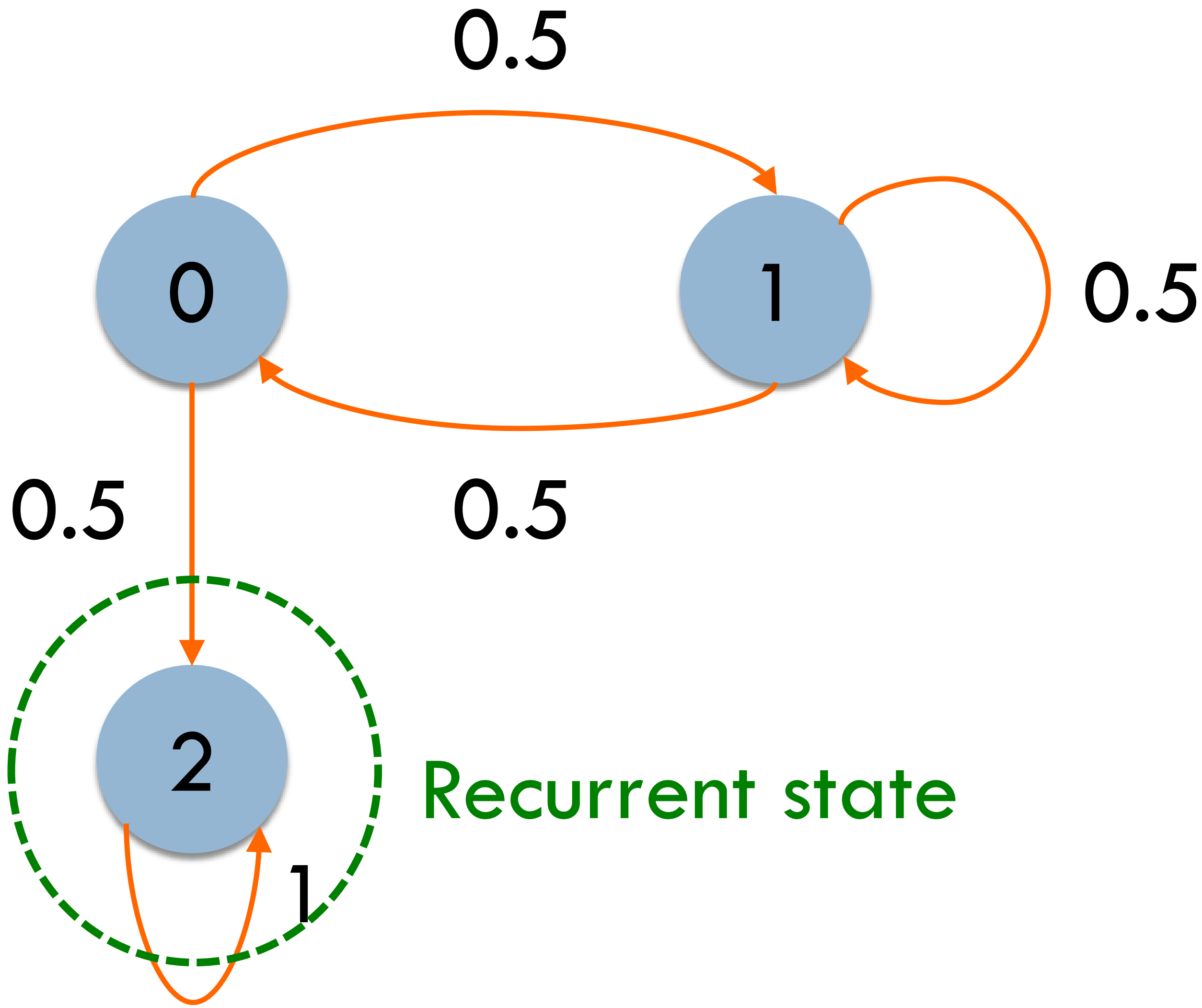
State Classification



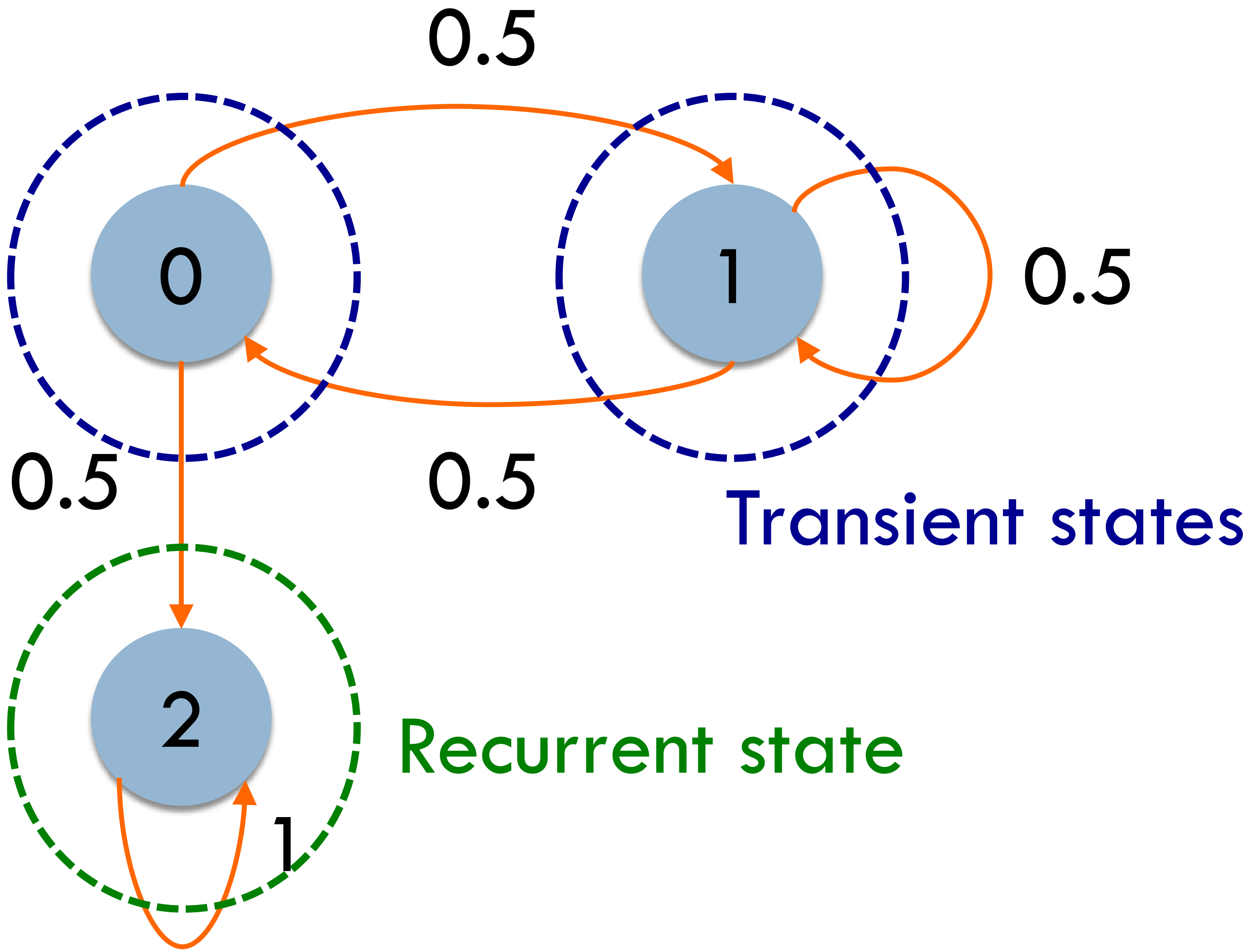
State Classification



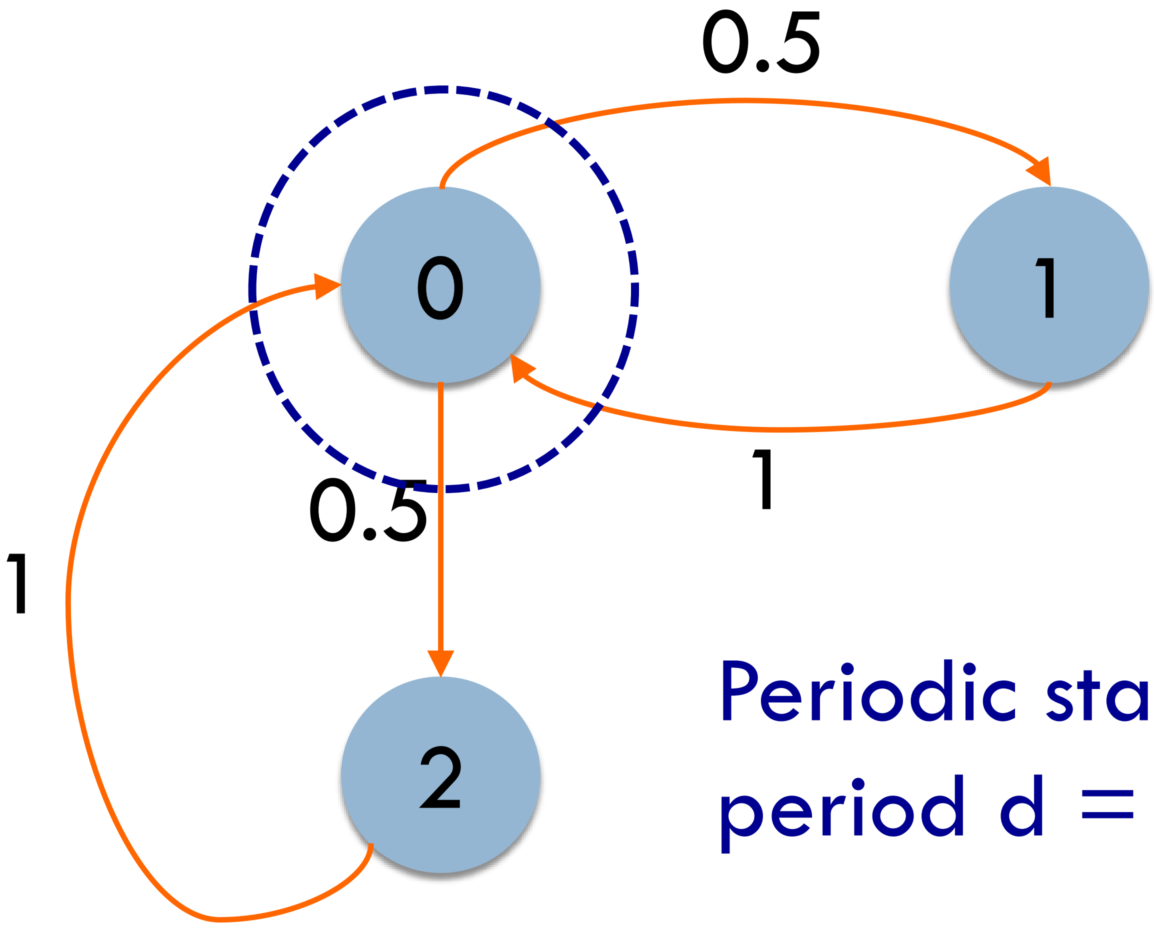
State Classification



State Classification

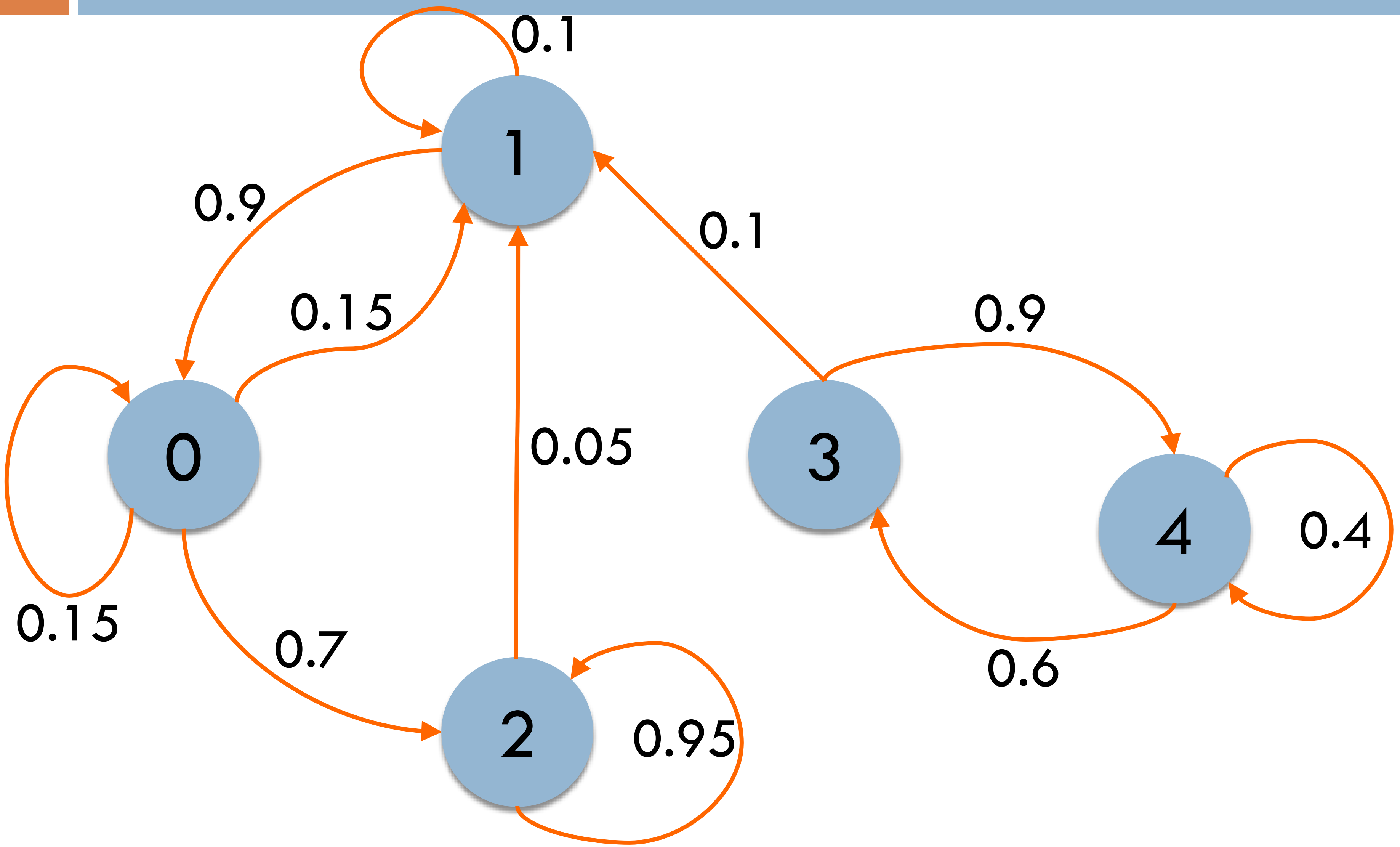


State Classification

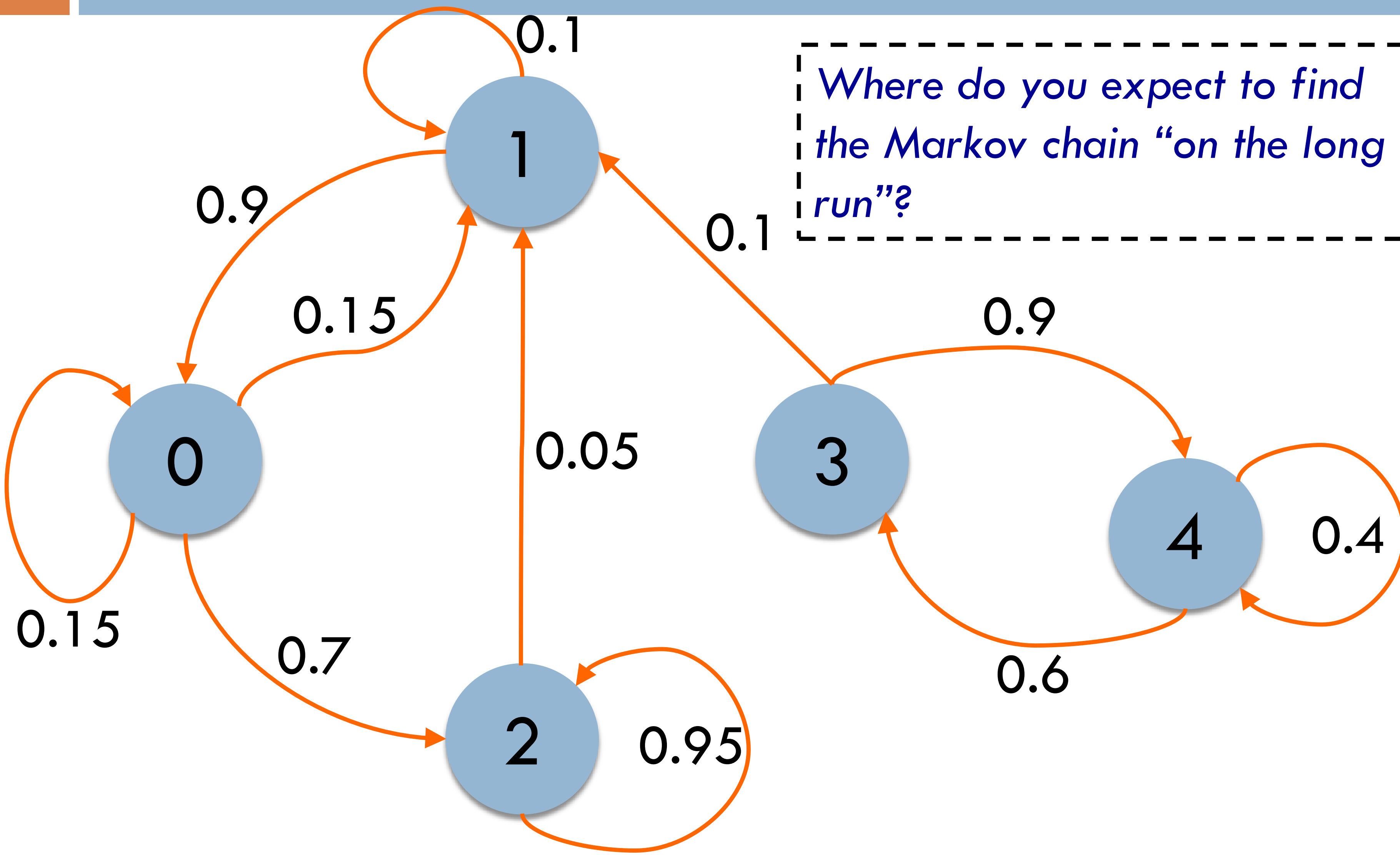


Periodic state, with period d = 2

Simple Exercise



Simple Exercise



Where do you expect to find the Markov chain "on the long run"?

Analysis of a DTMC

- Let us define the stationary probability of a DTMC as

$$\pi_j = \lim_{k \to \infty} \pi_j(k)$$

- It is the probability to find, on the long run, the DTMC in a certain state j

- Question 1: there exists this steady-state probability?
- Question 2: if any, what is the stationary probability that the DTCM is in state j, i.e., how can I compute it?

Some Definitions…

- A state j is said to be **accessible** from a state i (written $i \rightarrow j$) if a system started in state i has a non-zero probability of transitioning into state j

- A state i is said to **communicate** with state j (written $i \leftrightarrow j$) if both $i \rightarrow j$ and $j \rightarrow i$

- A set of states C is a **communicating class** if every pair of states in C communicates with each other, and no state in C is communicating with any state not in C

- A Markov chain is said to be **irreducible** if its state space is a single communicating class

…and some useful results

- Result 1: if a DTMC has a <u>finite</u> number of states, then at least one state is **recurrent**
- Result 2: if i is recurrent and $i \rightarrow j$, then even state j is **recurrent**
- Results 3: if X' is an irreducible set of states, then states are **all** positive recurrent, recurrent null or transient
- Results 4: if X' is a <u>finite</u> irreducible subset of the state space X, then every state in X' is **positive recurrent**

Analysis of a DTMC

- Theorem 1: in a DTMC irreducible and aperiodic there exists the limits

$$\pi_j = \lim_{k \to \infty} \pi_j(k)\,, \forall j \in X$$

and they are independent from the initial distribution π_0

- Theorem 2: in a DTMC irreducible and aperiodic in which all states are transient or recurrent null

$$\pi_j = \lim_{k \to \infty} \pi_j(k) = 0, \forall j \in X$$

Existence of steady-state distribution

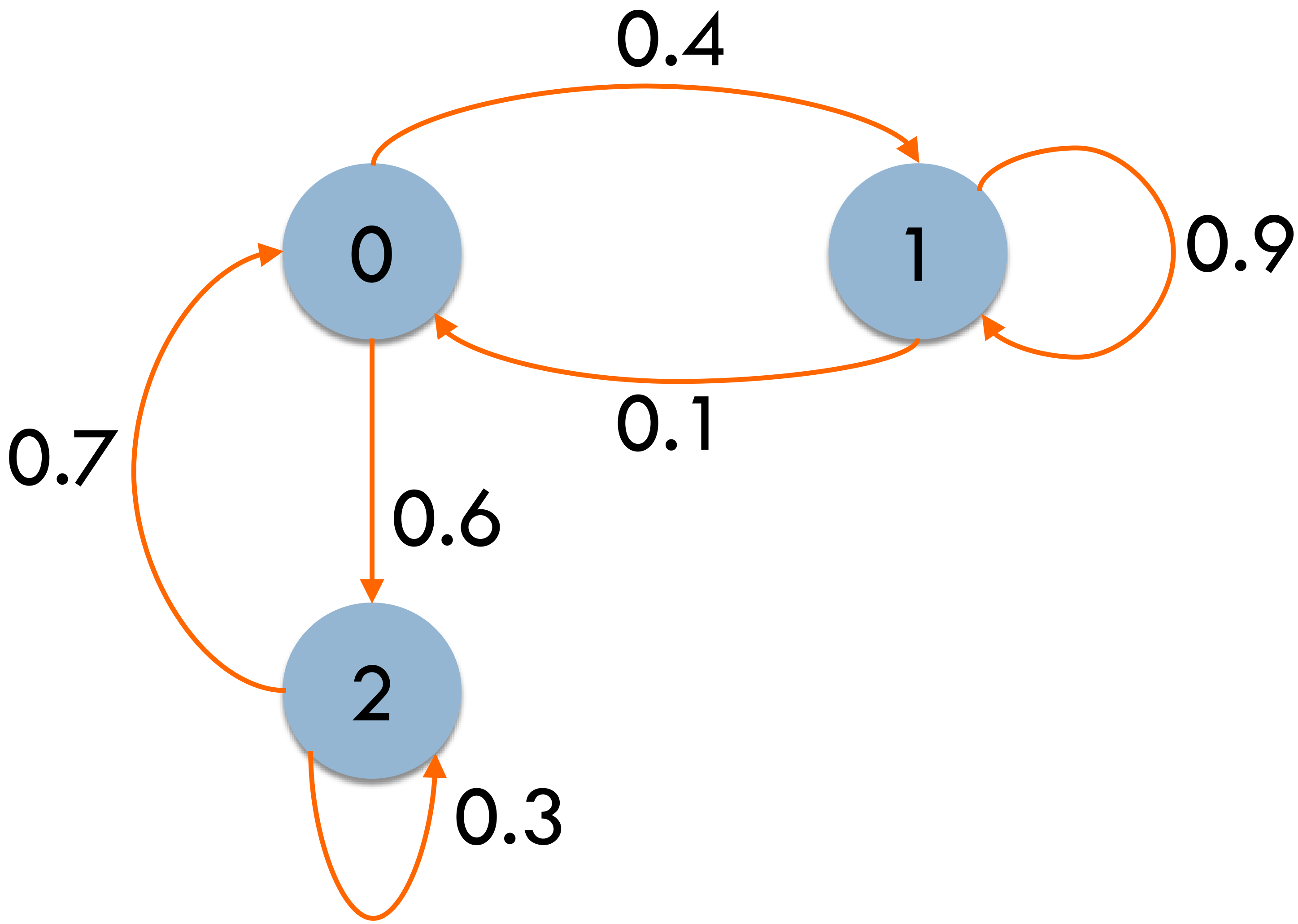
- Consider a time-homogeneous Markov chain is irreducible and aperiodic. Then, the following results hold:
  - If the Markov chain is positive recurrent, then there exists a unique π so that $\pi_j = \lim_{k \to \infty} \pi_j(k), \forall j$, and $\pi = \pi \cdot P$
  - If there exists a positive vector π such $\pi = \pi \cdot P$ and $\sum_{j \in X} \pi_j = 1$, then it must be the stationary distribution and the Markov chain is positive recurrent
  - If there exists a positive vector π such that $\pi = \pi \cdot P$ and $\sum_{j \in X} \pi_j = \infty$ is infinite, then a stationary distribution does not exist and $\lim_{k \to \infty} \pi_j(k) = 0$ for all j

Analysis of a DTMC

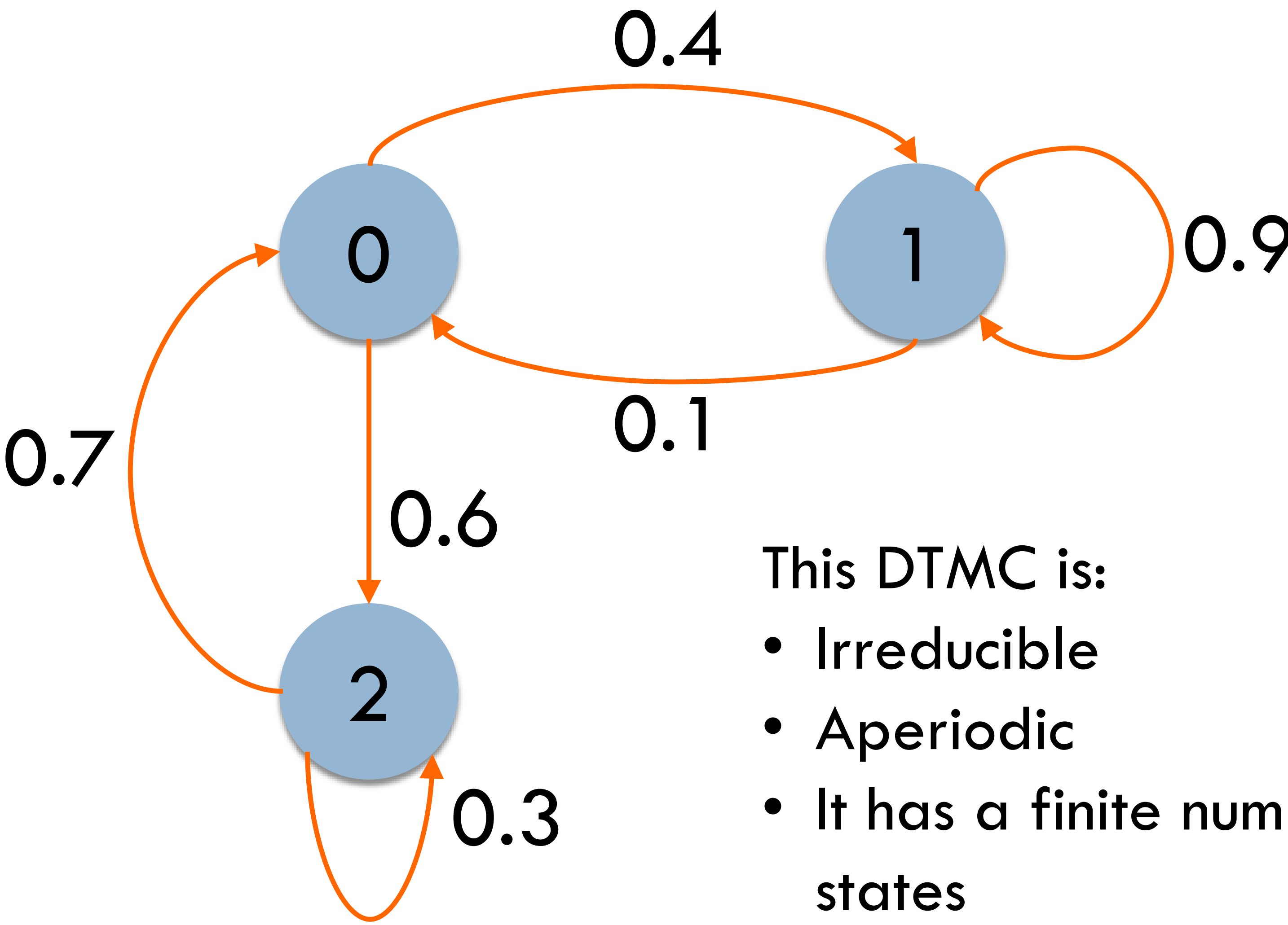
- **To sum up:** In order to compute the steady-state probabilities, we have to solve the following linear system:

$$\begin{cases} \boldsymbol{\pi} = \boldsymbol{\pi} \cdot \boldsymbol{P} \\ \sum_j \boldsymbol{\pi}_j = \mathbf{1} \end{cases}$$

A simple example



A simple example



This DTMC is:

- Irreducible
- Aperiodic
- It has a finite number of states

A simple example

Linear System

$$\begin{cases} \pi_0 = 0.1\pi_1 + 0.7\pi_2 \\ \pi_1 = 0.4\pi_0 + 0.9\pi_1 \\ \pi_2 = 0.6\pi_0 + 0.3\pi_2 \\ \pi_0 + \pi_1 + \pi_2 = 1 \end{cases}$$

Transition Probabilities

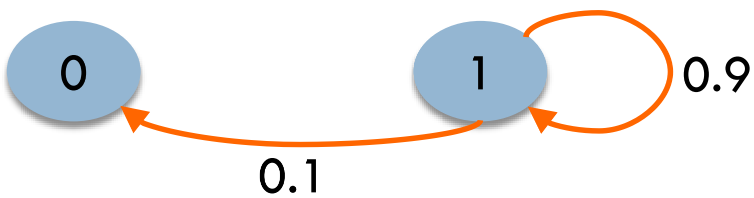
$$P = \begin{bmatrix} 0 & 0.4 & 0.6 \\ 0.1 & 0.9 & 0 \\ 0.7 & 0 & 0.3 \end{bmatrix}$$

Solution

$$\begin{cases} \pi_0 = 0.17 \\ \pi_1 = 0.68 \\ \pi_2 = 0.15 \end{cases}$$

Time spent in a state

- Can we characterize the time spent in each state by the DTMC?
- Let's focus on state *1 of* the previous example



- With $p = 0.1$ the DTMC will "jump" to state *0*, while with *1-p* = 0.1 will remain in state *1*
- Question: do you remind something similar??

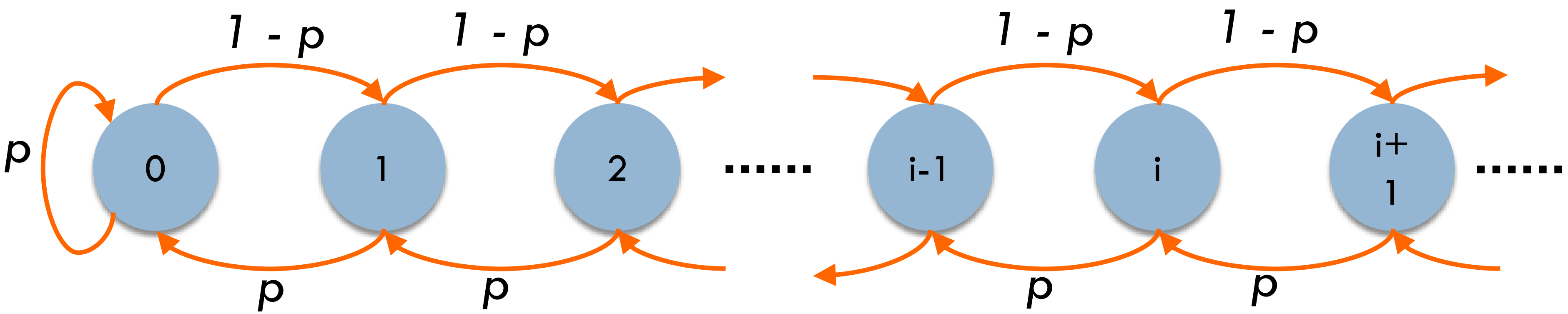
Time spent in a state

- The time spent in a state follows a *geometric distribution!*
- The geometric distribution is used for modelling the number of trials up to and including the first success
  - p = success
  - $1 - p$ = failure
  - P(Success in *K* trials) = $p \cdot (1 - p)^K$
- *Key feature of this distribution:* the geometric distribution is <u>memoryless!!</u>

$$P(T = m + n \mid T > m) = P(T = n)$$

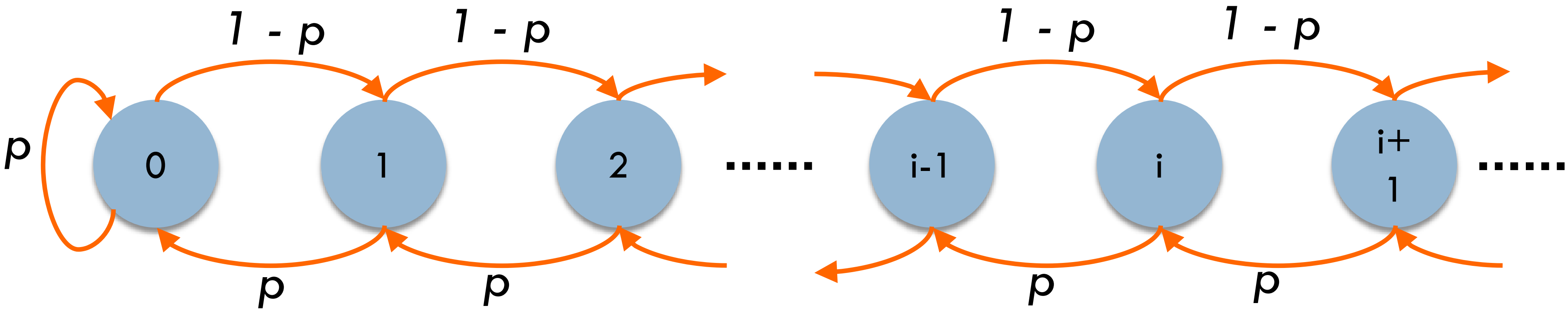
A more complex example

- A discrete time birth-death process



- The DTMC is irreducible and aperiodic

Birth-death process

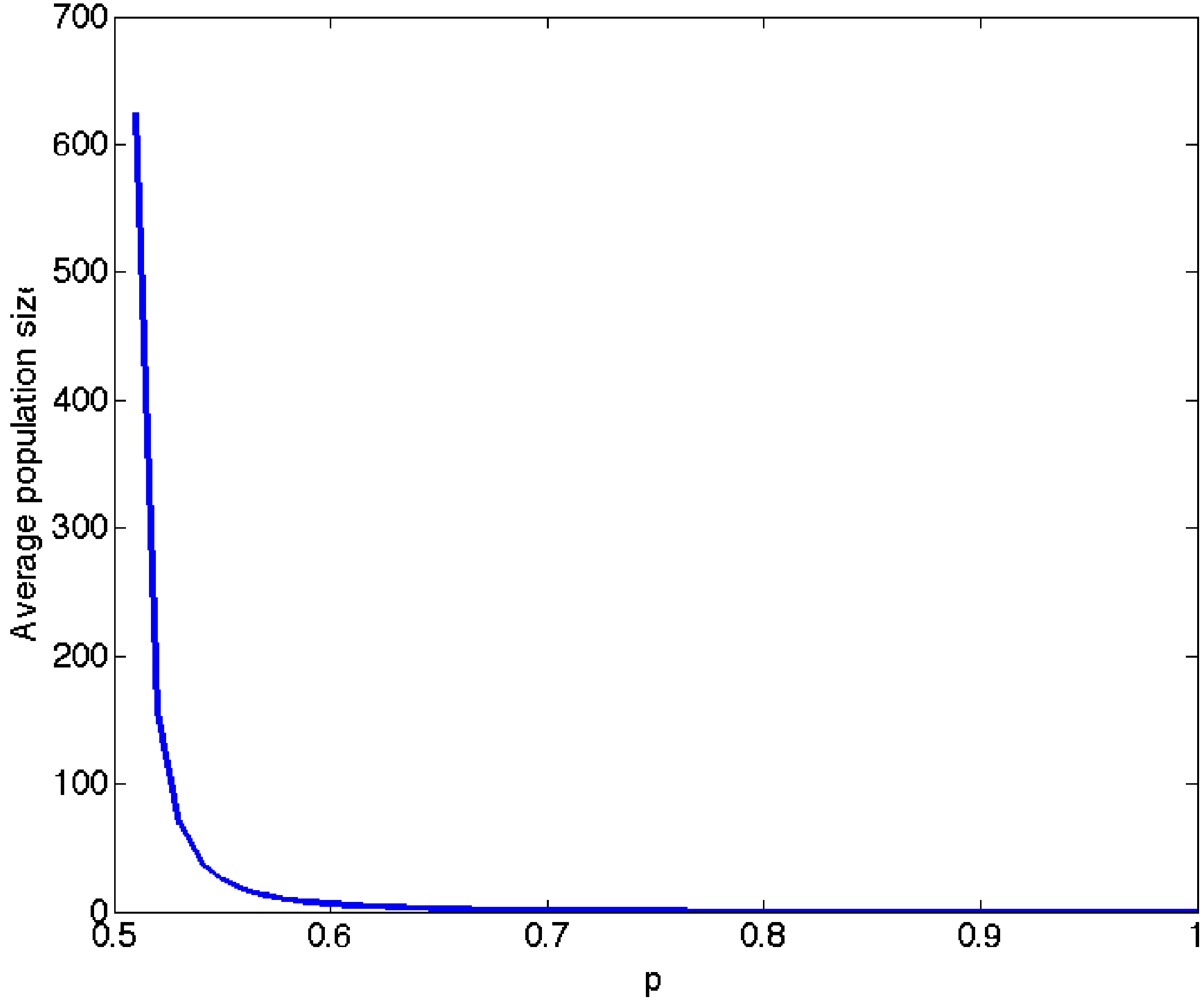


- There exists the steady-state probabilities?
- Intuitively
  - if $p < ½$ the DTMC will probably diverge, so maybe states are ***transient***
  - if $p > ½$ the DTMC will probably remain "near" 0, so state 0 could be *positive recurrent*, and since the DTMC is irreducible, all states would be *positive* ***recurrent***
  - if $p = ½$ the DTMC will probably neither diverge or converge, so maybe states are ***recurrent null***

Birth-death process solution

CHECK OUT THE DASHBOARD!

Birth-death process solution



Continuous Time Markov Chain (CTMC)

- Let s be the current time instant and $\boldsymbol{\tau}$ an arbitrary time interval
- Markov property for continuous time MC

$$\boldsymbol{P}(\boldsymbol{X}(\boldsymbol{s}+\boldsymbol{\tau})=\boldsymbol{j}\mid \boldsymbol{X}(\boldsymbol{s})=\boldsymbol{i})$$

- No state memory: next state depends only on the current state, and not on all history
- No age memory: the time already spent in the current state is irrelevant to determining the remaining time and the next state

DTMC versus CTMC

- The core of Discrete Time MC is the probability matrix P
  - Remember: It defines the probability to "jump" to another state in the next slot
- The core of a Continuous Time MC is the rate matrix Q
- It defines the rate at which the process transits from one state to another
- E.g., the MC transits from state *0* to state *1* with a rate of *5* times per seconds

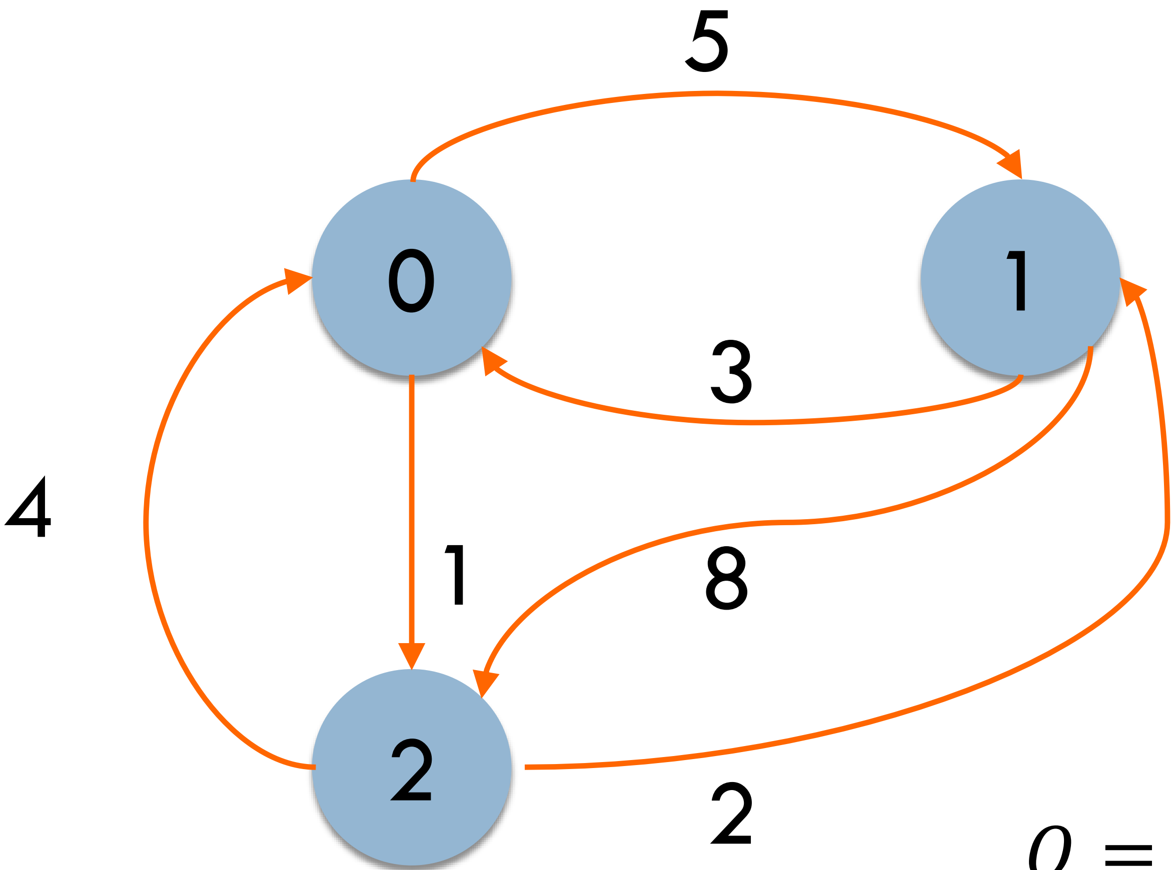
Homogeneous CTMC

A CTMC is said to be homogeneous if

$$\boldsymbol{P(X(s+\tau) = j \mid X(s) = i)}$$

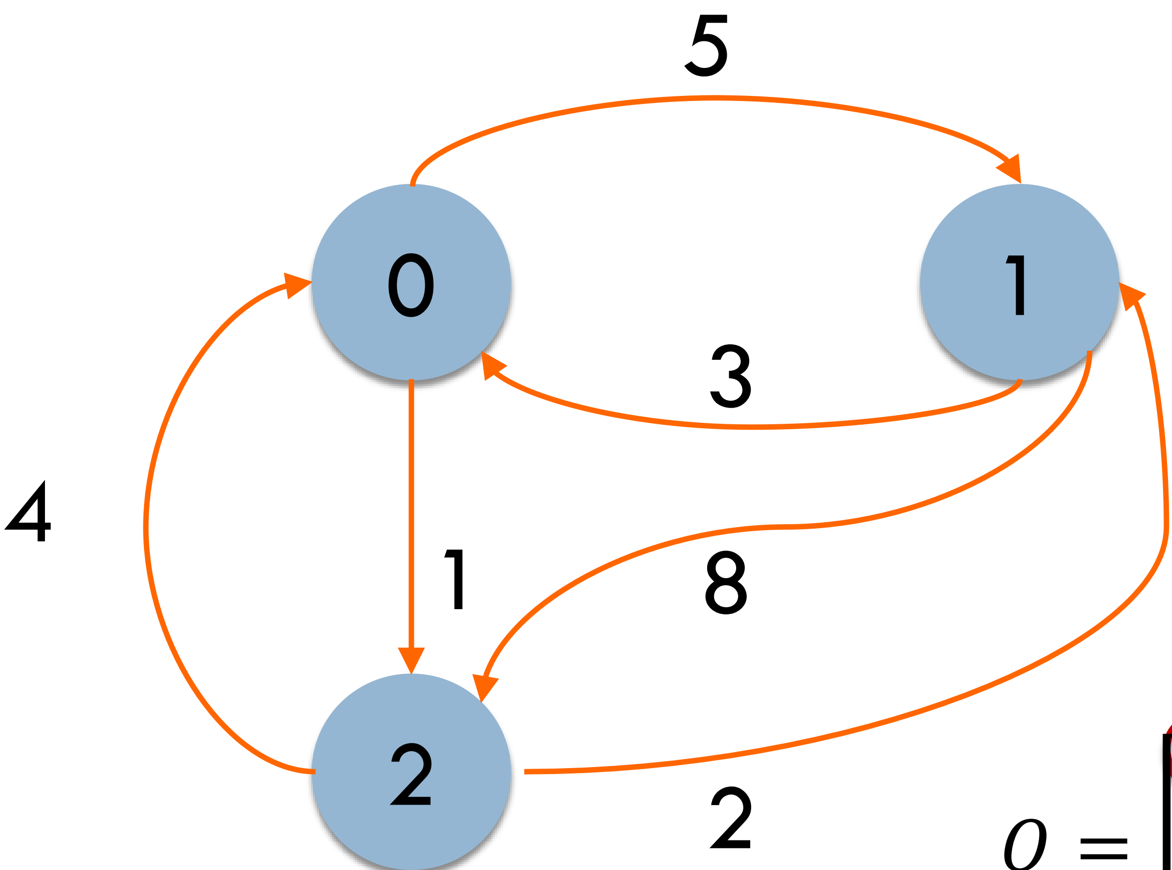
is independent from $\boldsymbol{s}$, i.e., only the "time interval" $\boldsymbol{\tau}$ matters

Design a CTMC



$$Q = \begin{bmatrix} -6 & 5 & 1 \\ 3 & -11 & 8 \\ 4 & 2 & -6 \end{bmatrix}$$

Design a CTMC



$$Q = \begin{bmatrix} -6 & 5 & 1 \\ 3 & -11 & 8 \\ 4 & 2 & -6 \end{bmatrix}$$

Existence of steady state distribution

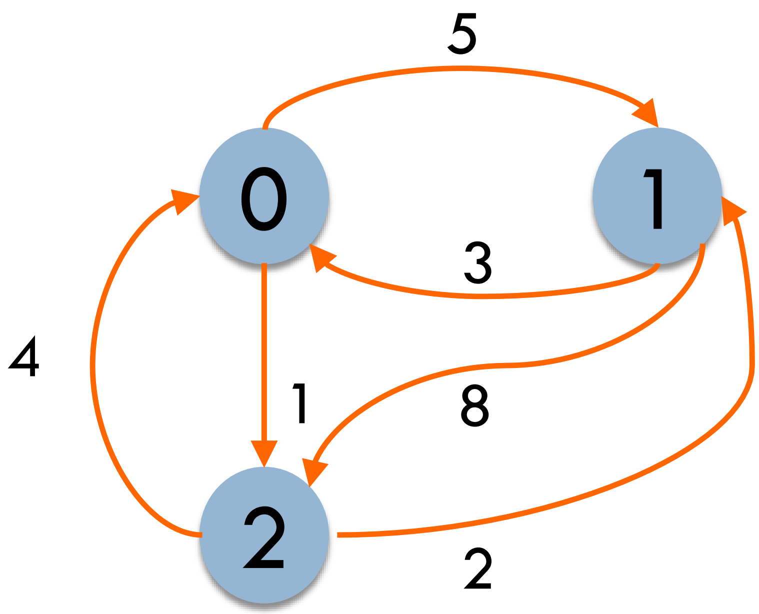
- Consider a time-homogeneous Markov chain is irreducible and aperiodic. Then, the following results hold:
  - If the Markov chain is positive recurrent, then there exists a unique π so that $\pi Q = 0$ and $\pi_j = \lim_{k \to \infty} \pi_j(k), \forall j$
  - If there exists a positive vector π such $\pi Q = 0$ and $\sum_{j \in X} \pi_j = 1$, then it must be the stationary distribution and the Markov chain is positive recurrent

Analysis of a CTMC

- **To sum up:** In order to compute the steady-state probabilities, we have to solve the following linear system:

$$\begin{cases} \boldsymbol{\pi} \cdot \boldsymbol{Q} = \boldsymbol{0} \\ \sum_{j} \boldsymbol{\pi}_{j} = \boldsymbol{1} \end{cases}$$

Example



$$Q = \begin{bmatrix} -6 & 5 & 1 \\ 3 & -11 & 8 \\ 4 & 2 & -6 \end{bmatrix}$$

- *Flow equilibrium equations!*

$$\begin{cases} -6\pi_0 + 3\pi_1 + 4\pi_2 = 0 \\ 5\pi_0 - 11\pi_1 + 2\pi_2 = 0 \\ \pi_0 + 8\pi_1 - 6\pi_2 = 0 \end{cases}$$

Time spent in a state

- If $v(i)$ is the time spent in state i, for CTMC it follows an exponential distribution:

$$\boldsymbol{P(v(j) < t) = 1 - e^{-\Lambda(j)t}}$$

where $\boldsymbol{\Lambda(j)}$ is the exit rate from state j

- Memoryless property: the exponential distribution is memoryless!

Exercise: model a wireless link using a DTMC

Consider a simple model of a wireless link where, due to channel conditions, either one packet or no packet can be served in each time slot. Let $s[k]$ denote the number of packets served in time slot k and suppose that $s[k]$ are i.i.d. Bernoulli random variables with mean μ. Further, suppose that packets arrive to this wireless link according to a Bernoulli process with mean λ, i.e., $a[k]$ is Bernoulli with mean λ where $a[k]$ is the number of arrivals in time slot k and $a[k]$ are i.i.d. across time slots. Assume that $a[k]$ and $s[k]$ are independent processes.

We specify the following order in which events occur in each time slot:

- We assume that any packet arrival occurs first in the time slot, followed by any packet departure, i.e., a packet that arrives in a time slot can be served in the **same** time slot
- Packets that are not served in a time slot are queued in a buffer for service in a future time slot.

Compute, if exists, the steady-state distribution.