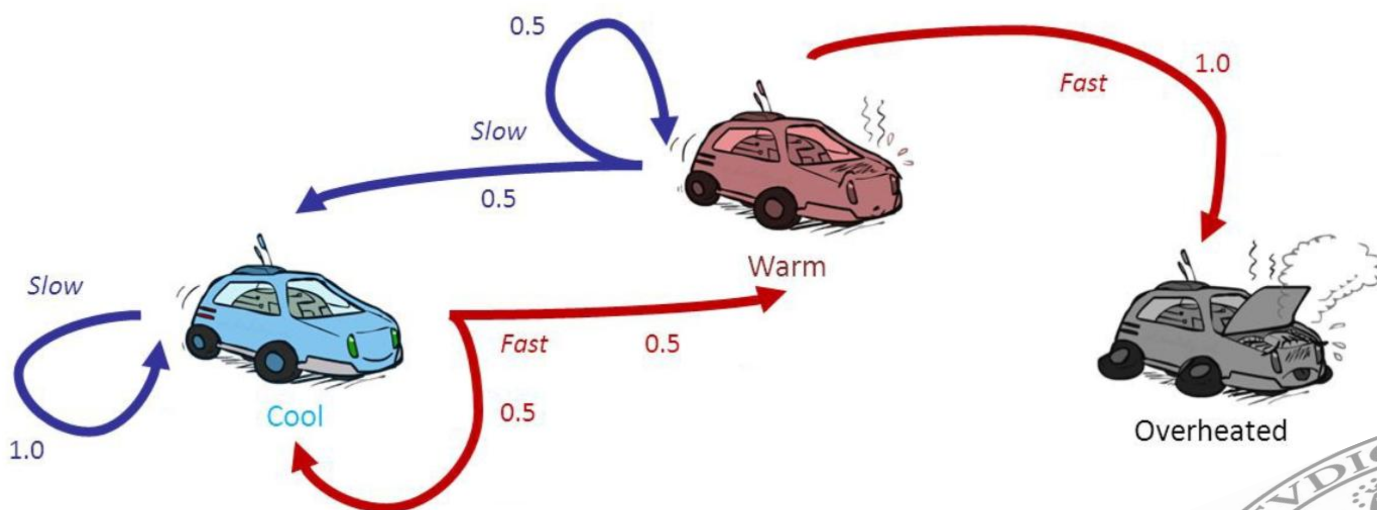
UNIVERSITY OF CAGLIARI
DIEE - Department of Electrical and Electronic Engineering

STOCHASTIC MODELS
-
Discrete-Time Markov Chain

А. А. Марков (1886).



alessandro.pilloni@unica.it

Markov Chain

- A Markov chain $\{x(t), t \in T\}$ is a Markov Process which state-space X is discrete

$$x(t) \in X = \{x_0, x_1, x_2, \ldots\} \quad \forall\, t \in T$$

  either **finite**, or **countably infinite**.

- A Markov chain can evolve in:
  - discrete-time (DT), i.e. $t = k\Delta T$ where $dt > 0$ is the sampling rate and $k \in \mathbb{N}$
  - continuous-time (CT), i.e. $t \in \mathbb{R}$ state transitions may occur at any time
- depending on the considered model and/or application.
- **Note:** a CT-MC can always be approximated by a DT-MC for sufficiently small dt.

- Example of CT-MC: Every queue system where arrivals can occur at every point in the time line.
- Example of DT-MC: Every queue system where arrivals are recorded at a given sampling rate, e.g., every minute.

Discrete-time Markov Chains (DT-CM)

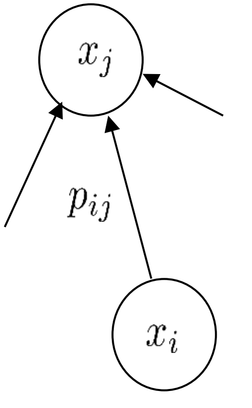
- In DT-MCs $\{x(k),\ k \in \mathbb{N}\}$ transitions can occur only at discrete time instants

$$t_k = k \cdot dt > 0,\ \ dt = cost.\ \ \forall\, k = 1,2, \ldots$$

- In a DT-MC the Markov property can be rewritten as follows

$$\Pr(\ x(k+1) = x_j \mid x(k) = x_i,\ x(k-1) = \ldots, \ldots,\ x(0) = x_0\ ) =$$
$$= \Pr(\ x(k+1) = x_j \mid x(k) = x_i\ )\ =\ p_{ij}(k)$$

- where $x(k)$ denote the random variable of $\{x(k)\}$ associated to the k-th sampling-interval, i.e., $x(k) \equiv x(t_k)$, and $t_k = k \cdot dt,\ \ dt > 0$
- Moreover by exploiting the **Law of Total Probability** it results:



$$\Pr(x(k+1) = x_j)\ =\ \sum_{\forall\, i} p_{ij} \cdot \Pr(x(k) = x_j)$$
$$\equiv\ \pi_j(k+1)\ =\ \sum_{\forall\, i} p_{ij} \cdot \pi_i(k)$$

Discrete-time Markov Chains (DT-CM)

- **Definition:** A DT-MC is a Markov SP defined by the triplet

$$C = (X, \boldsymbol{P}(k), \Pi(0))$$

1. X : is the sample space of the MC, namely its collection of states;
2. $\boldsymbol{P}(k)$: is the transition probability matrix at time $k \in \mathbb{N}$

$$\boldsymbol{P}(k) = \begin{pmatrix} p_{11}(k) & p_{12}(k) & \cdots \\ p_{21}(k) & \ddots & \ddots \\ \vdots & \ddots & p_{ij}(t) \end{pmatrix} \qquad p_{ij}(k) = \Pr(\ x(k+1) = x_j \mid x(k) = x_i\ )$$

$$\forall\, x_i,\ x_j \in X,\ \forall\, k \in \mathbb{N}$$

3. $\Pi(0)$: is the initial marginal probability distribution of being in x_i at $t = 0$

$$\Pi(0) = [\pi_0(0), \pi_1(0), \ldots] \qquad \pi_i(0) = \Pr(x(0) = x_i) \qquad \sum_i \pi_i(t) = 1\ \forall\, t \geq 0$$

Remark: In the MC literature letter 'p' is used for the transition probabilities, whereas the pmf distribution at time t is $\Pi(\mathrm{t}) = [\pi_0(t), \pi_1(t) \ldots]$, where $\pi_i(t) = \Pr(x(t) = x_i)$

Time Homogenous DT-MCs

- **Definition:** Consider a DT-MC defined by the triplet

$$C = \big(X, \boldsymbol{P}(k), \Pi(0)\big) \text{ with } k \in \mathbb{N}.$$

- If $\boldsymbol{P}(k)$ is constant ($\boldsymbol{P}(k) = \boldsymbol{P} \quad \forall\, k$), the DT-MC is said to be time-homogenous.

- **NOTE:** A time-homogenous **finite-state** DT-MC is **always WSS**, since the independence by k implies the independence to time-shifts $m > 0$, moreover note that since the number of states is finite, then $E[x(k)^2] < +\infty$

- Example: Consider the SP associated to the weather day $\{X_t\}$ with $dt = 1$ day, and which sample space is $X = \{x_1, x_2\}$

$$x_1 = \text{cloudy day} \quad , \quad x_2 = \text{sunny day}$$

- with

$$p_{11} = Pr(\text{Tomorrow} = x_1 | \text{Today} = x_1) = a$$
$$p_{22} = Pr(\text{Tomorrow} = x_2 | \text{Today} = x_2) = b$$

then $\boldsymbol{P} = \begin{pmatrix} a & 1-a \\ 1-b & b \end{pmatrix}$

Transition probabilities of a DT-MC

- Each transition probability $p_{ij}(k)$ denotes the probability to jump in state x_j from state x_i at the generic k-th time instant

$$p_{ij}(k) = \Pr(\ x(k+1) = x_j \mid x(k) = x_i\ ) \qquad \forall\, x_i,\, x_j \in X,\ \forall\, k \in \mathbb{N}$$

- It follows that each $p_{ij}(k)$ satisfies

$$p_{ij}(k) \in [0,1] \qquad \forall\, x_i,\, x_j \in X \qquad\qquad p_{ij}(k)\ :\ \sum_{j=0}^{\infty} p_{ij} = 1 \quad \forall\, x_i \in X$$

- By construction $\boldsymbol{P}(k)$ is right-stochastic (row sum equal to 1) , thus $\boldsymbol{P}(k)\mathbf{1} = \mathbf{1}$

1. Follows that **1** is an eigenvector of $\boldsymbol{P}(k)$ and “$\lambda_1 = 1$” is the corresponding eigenvalue, i.e

$$\mathbf{1} = [1,1,1,\ldots]^\top \qquad \boldsymbol{P}(k)\mathbf{1} = \mathbf{1} \implies \boldsymbol{P}(k)\mathbf{1} = \lambda_1 \cdot \mathbf{1} \quad \text{with} \quad \lambda_1 = 1$$

2. By the Gershgorin Theorem (given with no proof), the eigenvalues of $\boldsymbol{P}(k)$ satisfies

$$|\lambda_i(k)| \leq 1 \qquad \forall\, \lambda_i(k) \in \mathrm{eig}\,\{\boldsymbol{P}(k)\} \quad \forall\, k$$

DT-MC can denoted by means of Graphs

- Time-homogeneous DT-MC are often described by a means of directed graphs

$$G = (V, E)$$

V is the vertex set
$E \subseteq V \times V$ is the edge set of G

- Each possible outcome (state) of the DT-MC corresponds to a vertex on G

$$V = X = \{x_1, x_2, \dots\}$$

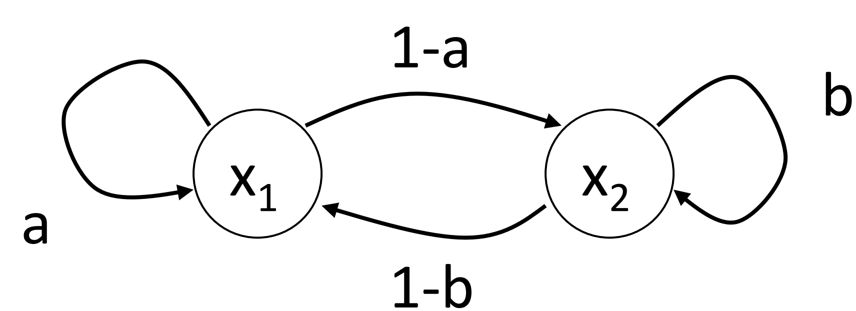
- whereas an edge $(x_i, x_j) \in E$ from the vertex x_i to x_j exists if and only if

$$p_{ij} = \Pr\big(\, x(k+1) = x_j \,\big|\, x(k) = x_i \,\big) \neq 0$$

- The **weight** associated with edge (x_i, x_j) **corresponds** to p_{ij} .

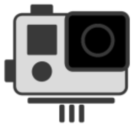
- Example 1: Weather prediction model.

$$\boldsymbol{P} = \begin{pmatrix} a & 1-a \\ 1-b & b \end{pmatrix}$$



Example 2: Inventory Example

- A camera store stocks a new action-camera



- Assume the demand of cameras $\{D_t, t = T, 2T, \ldots\}$ per-week T follows a DT-Poisson process

$$\Pr(D_t = i) = \frac{(\lambda t)^i}{i!} e^{-\lambda t} \qquad \begin{cases} \lambda = 1 \left[\dfrac{\text{demand}}{\text{week}}\right] \\ T = 1[\text{week}] \end{cases}$$

- **Assume:** the stocks are restored using an (s, S) policy with s=1 and S=3:

  ✓ If the inventory is less than s: **place an order** to **replenish** the supply up to S
  ✓ If the inventory is greater or equal to s: **do nothing**

- **Assume:** if placed, orders are made the Saturday night, and stocks arrive the early Monday

Model the weekly behavior of stocks each Saturday night
as a DT-MC $\{x(t), t = T, 2T, \ldots\}$ counting the #cameras in the INVENTORY

- **Assume:** At the startup the stock contains 3 cameras, i.e. $x(0) = 3 \;\Rightarrow\; \Pi(0) = [0\ 0\ 0\ 1]$

Example 2: Inventory Example (cont'd)

- **Sample Space:** No more then 3 cameras can be stored, thus, $X_t \in \{0,1,2,3\}$
- **Transition probability matrix:**

$$\boldsymbol{P} = \begin{bmatrix} \Pr(X_{t+1}=0|X_t=0) & \Pr(X_{t+1}=1|X_t=0) & \Pr(X_{t+1}=2|X_t=0) & \Pr(X_{t+1}=3|X_t=0) \\ \Pr(X_{t+1}=0|X_t=1) & \Pr(X_{t+1}=1|X_t=1) & \Pr(X_{t+1}=2|X_t=1) & \Pr(X_{t+1}=3|X_t=1) \\ \Pr(X_{t+1}=0|X_t=2) & \Pr(X_{t+1}=1|X_t=2) & \Pr(X_{t+1}=2|X_t=2) & \Pr(X_{t+1}=3|X_t=2) \\ \Pr(X_{t+1}=0|X_t=3) & \Pr(X_{t+1}=1|X_t=3) & \Pr(X_{t+1}=2|X_t=3) & \Pr(X_{t+1}=3|X_t=3) \end{bmatrix}$$

- it clearly depends by the demand process for $t = 1\ Week$ as next

$$\boldsymbol{P} = \begin{bmatrix} \Pr(D_t \geq 3) & \Pr(D_t = 2) & \Pr(D_t = 1) & \Pr(D_t = 0) \\ \Pr(D_t \geq 1) & \Pr(D_t = 0) & 0 & 0 \\ \Pr(D_t \geq 2) & \Pr(D_t = 1) & \Pr(D_t = 0) & 0 \\ \Pr(D_t \geq 3) & \Pr(D_t = 2) & \Pr(D_t = 1) & \Pr(D_t = 0) \end{bmatrix}$$

$$\Pr(D_t = 0) = \frac{1^0}{0!}e^{-1} = 0.367$$

$$\Pr(D_t = 1) = \frac{1}{1!}e^{-1} = 0.367$$

$$\Pr(D_t = 2) = \frac{1^2}{2!}e^{-2} = 0.184$$

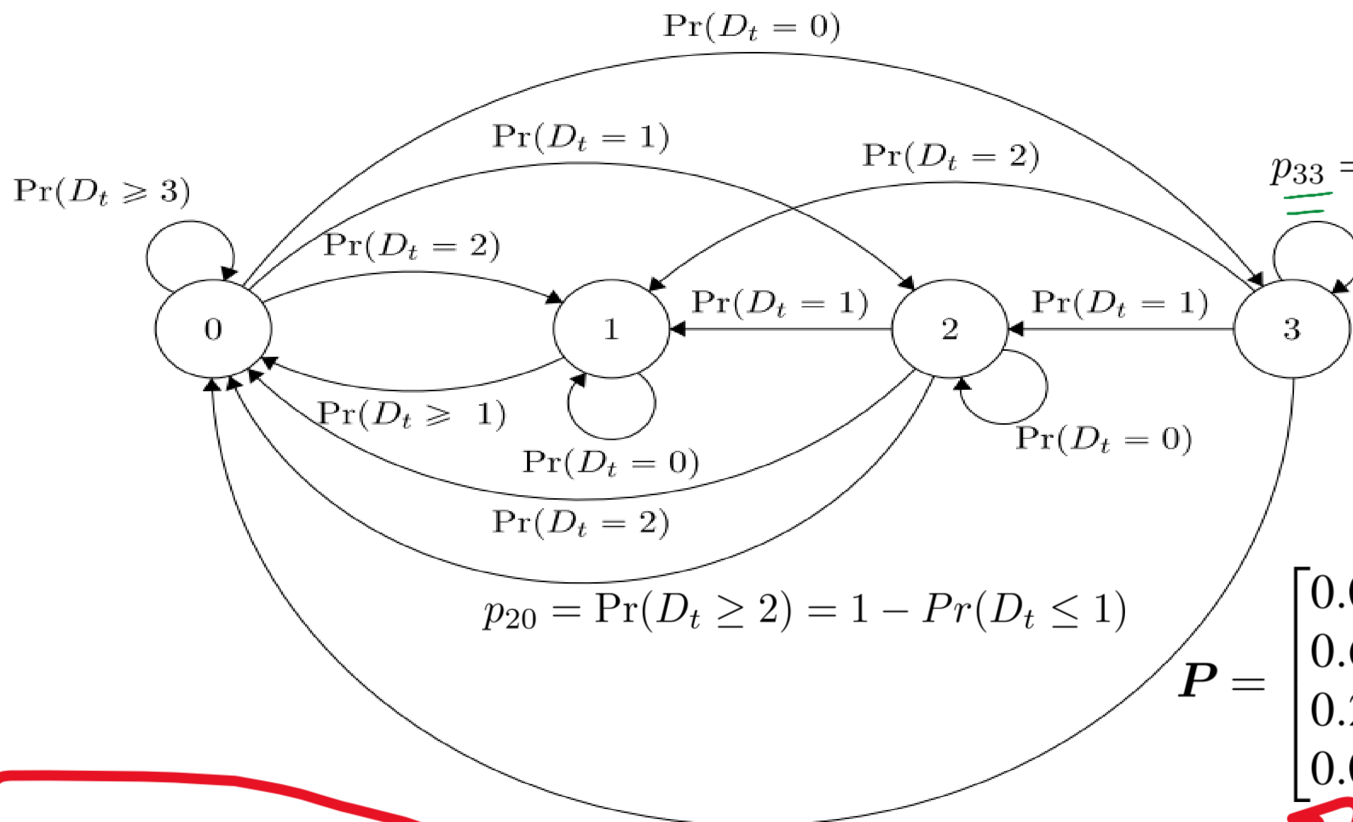
$$\Pr(D_t \geq 1) = 1 - \Pr(D_t = 0) = 0.632$$

$$\Pr(D_t \geq 2) = 1 - \Pr(D_t = 0) - \Pr(D_t = 1) = 0.264$$

$$\Pr(D_t \geq 3) = 1 - \Pr(D_t = 0) + \\ - \Pr(D_t = 1) - \Pr(D_t = 2) = 0.080$$

Example 2: Inventory Example (cont'd)

- No more then 3 cameras can be stored, thus, $X_t \in \{0,1,2,3\}$, whereas the MC transition probability matrix is



$$p_{33} = \Pr(D_t = 0) = \left.\frac{(\lambda T)^k}{k!} e^{\lambda T}\right|_{k=0,\lambda T=1} = e^{-1} = 0.368$$

$$p_{20} = \Pr(D_t \geq 2) = 1 - Pr(D_t \leq 1)$$

$$p_{30} = \Pr(D_t \geq 3) = 1 - Pr(D_t \leq 2)$$

$$\boldsymbol{P} = \begin{bmatrix} 0.080 & 0.184 & 0.368 & 0.368 \\ 0.632 & 0.368 & 0 & 0 \\ 0.264 & 0.368 & 0.368 & 0 \\ 0.080 & 0.184 & 0.368 & 0.368 \end{bmatrix}$$

```
% Matlab code:
P=[1-poisscdf(2,1), poisspdf(2,1), poisspdf(1,1), poisspdf(0,1);
    1-poisscdf(0,1), poisspdf(0,1), 0, 0;
    1-poisscdf(1,1), poisspdf(1,1), poisspdf(0,1), 0;
    1-poisscdf(2,1), poisspdf(2,1), poisspdf(1,1), poisspdf(0,1)]
```

DT-MCs describe how Marginal probabilities change w.r.t the Time

- Let

$$\Pi(k) = [\pi_1(k), \pi_2(k), \ldots]$$

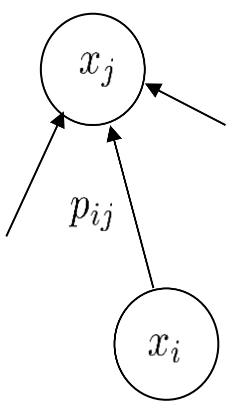
- be the marginal pmf of the DT-MC at time k, where

$$\pi_i(k) = \Pr(x(k) = x_i), \forall\, x_i \in X$$

- is the probability to be in state x_i at time k.

- For all $k > 0$, $\forall\ j$, following the **Law of Total Probability** the next relation is in force

$$\pi_j(k) = \sum_{x_i \in X} p_{ij}(k-1)\pi_i(k-1)$$



DT Chapman-Kolmogorov Equation

$$\Pi(k) = \Pi(k-1)\boldsymbol{P}(k-1)$$

Time evolution of marginal probabilities of a DT-MC (cont'd)

- For a time-homogeneous DT-MC the Chapman-Kolmogorov equation can be rewritten as follows:

$$
\begin{aligned}
\Pi(1) &= \Pi(0)\boldsymbol{P} \\
\Pi(2) &= \Pi(1)\boldsymbol{P} = \Pi(0)\boldsymbol{P}^2 \\
\Pi(3) &= \Pi(2)\boldsymbol{P} = \Pi(1)\boldsymbol{P}^2 = \Pi(0)\boldsymbol{P}^3 \\
\vdots &= \vdots \\
\Pi(k) &= \Pi(k-1)\boldsymbol{P} = \Pi(0)\boldsymbol{P}^k
\end{aligned}
$$

Chapman-Kolmogorov Equation for time-homogeneous DT-MC

This is the explicit solution of the DT-MC dynamic $\Pi(k) = \Pi(0)\boldsymbol{P}^k$

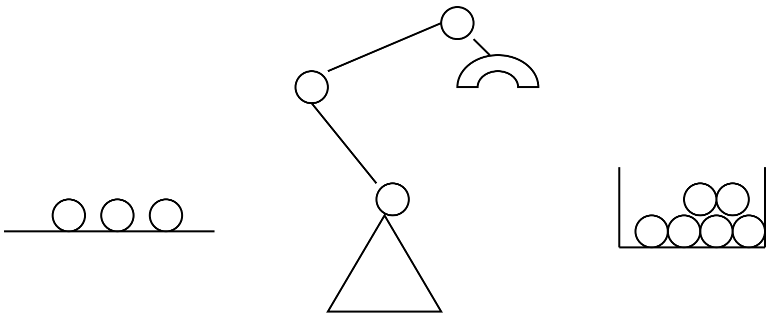
Examples (cont'd)

- Example 3: Consider a robot-arm which moves pieces from a conveyor belt to a depot of infinite capacity.
- The robot has 3 possible states

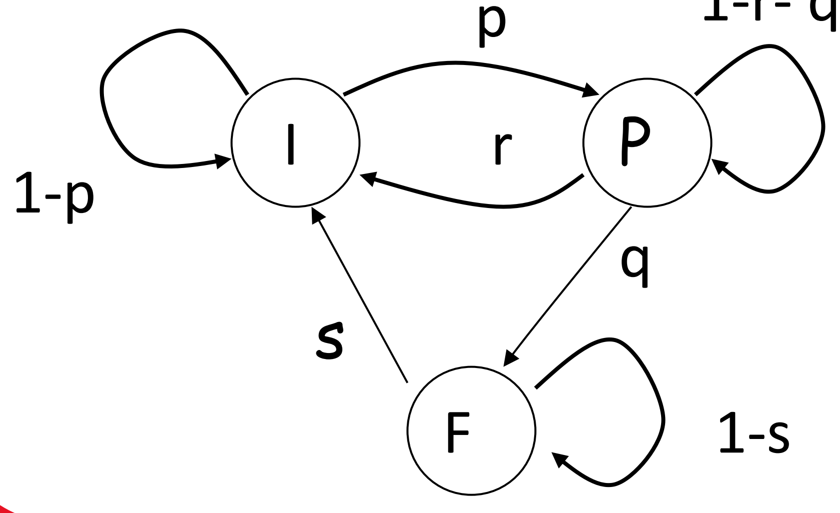
  I = Idle mode

  P = Process of mooving a piece

  F = Faulty mode.



$$X = \{I, P, F\}$$



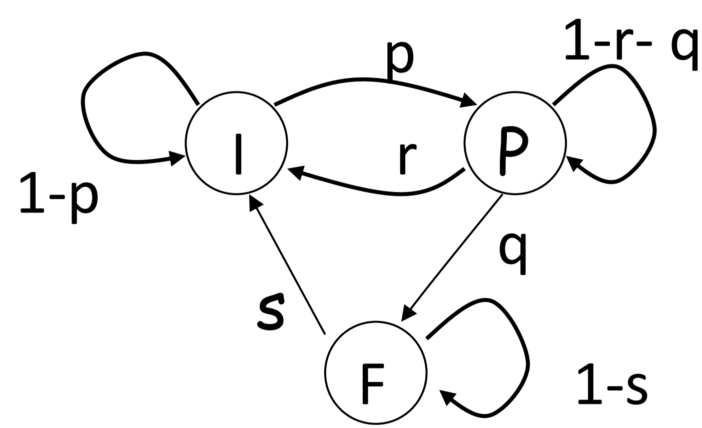
$$p = \Pr(x(k+1) = P \mid x(k) = I)$$

$$r = \Pr(x(k+1) = I \mid x(k) = P)$$

$$q = \Pr(x(k+1) = F \mid x(k) = P)$$

$$s = \Pr(x(k+1) = I \mid x(k) = F)$$

Examples (cont'd)



- The transition probability matrix is

$$\boldsymbol{P} = \begin{pmatrix} 1-p & p & 0 \\ r & 1-r-q & q \\ s & 0 & 1-s \end{pmatrix}$$

- By the Chapman-Kolmogorov equation, an estimation of the probability of being in a given state, k time-steps later can be easily achieved.
- Assume the robot in the IDLE state at time $k = 0$

$$\Pi(0) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

Prob. to be in 2 steps to P because $\Pi(0)$ throght the path $I \to P \to P$ or $I \to I \to P$

- Then the pmf of this process after k time units, is as follows

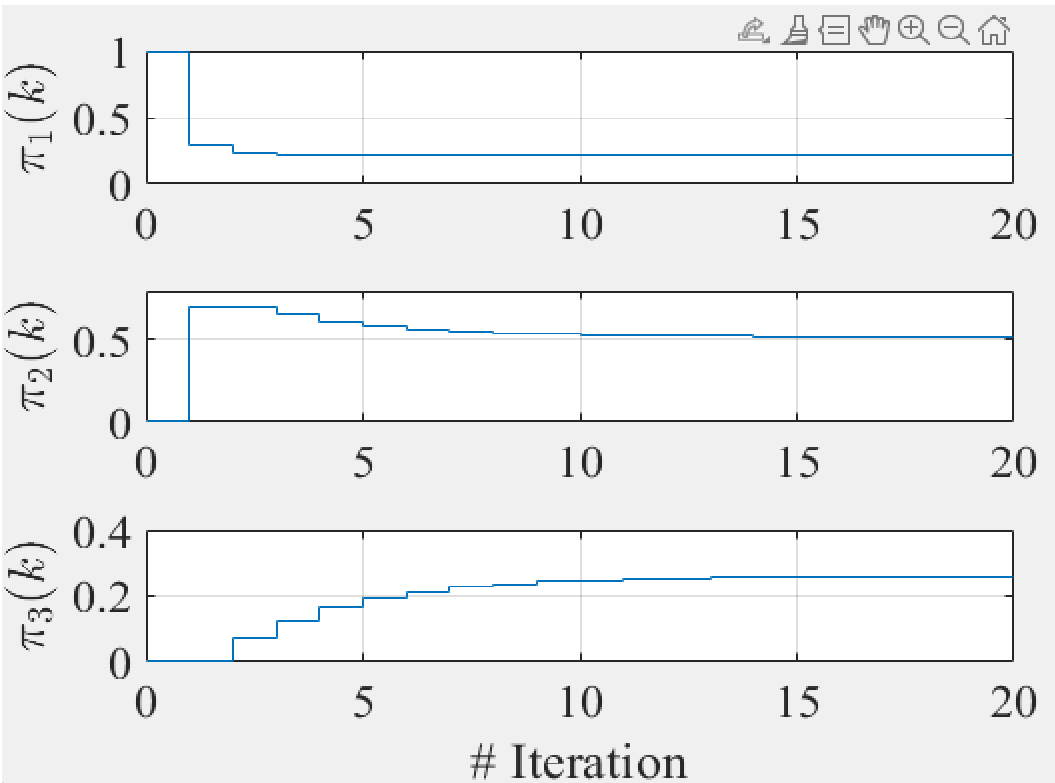
$$\begin{aligned} \Pi(1) &= \Pi(0)\boldsymbol{P} = \begin{bmatrix} 1-p, & p, & 0 \end{bmatrix} \\ \Pi(2) &= \Pi(1)\boldsymbol{P} = \begin{bmatrix} (1-p)^2 + rp, & p(1-p) + p(1-r-q), & pq \end{bmatrix} \\ &\vdots \\ \Pi(n) &= \Pi(0) \cdot \boldsymbol{P}^n \end{aligned}$$

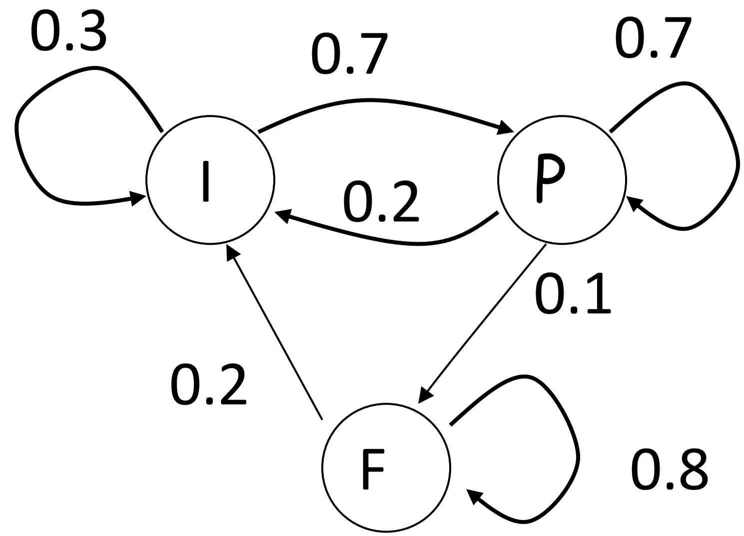
Prob. to be in 2 steps to Q, i.e. only the path $I \to P \to Q$

*Each entry of $\Pi(n)$ is an application of the Law of Total Probability and denotes the **n-step forward marginal probability** to be in x_i from the prior info $\Pi(0)$*

Evolution of marginal probabilities in Example 3

$$\Pi(k) = \Pi(0)\boldsymbol{P}^k$$





$$\boldsymbol{P} = \begin{pmatrix} 1-p & p & 0 \\ r & 1-r-q & q \\ s & 0 & 1-s \end{pmatrix}$$

Interpretation of the simulation:

- The **unconditional probability** to be in the **faulty state from** the **Idle state**, increases monotonically, and after **5 time-units** becomes larger then the 20%
- Thus, to keep $\pi_3(k) \leq 0.2$ the **system needs a restart** every $k = 5$ time-units

Matlab Simulation of Example 3 (cont’d)

```
k_end=60; % Simulation lenght

pi_0=[1 0 0]; % Initial marginal probability

p=0.7; q=0.1; r=0.2; s=0.2;

P=[1-p p 0; r 1-q-r q; s 0 1-s]; % Transition Prob. Matrix

vec_pi=zeros(k_end+1,3); % DT-MC evolution vector
vec_pi(1,:)=pi_0; % Initialization of vec_pi @ k=0

for k=2:1:k_end+1
    vec_pi(k,:)=vec_pi(k-1,:)*P % DT-MC state update
End

% Verification of the Chapman-Kolgomorov Equation
round( vec_pi(k_end,:),5) == round( pi_0*P^(k_end), 5)
```

Matlab Simulation of Example 3 (cont'd)

```
figure(1)
subplot(311)
        stairs(0:k_end,vec_pi(:,1))
        ylim([0,1])
        ylabel('$\pi_1(k)$','fontsize',14,'interpreter','latex')
        set(gca,'FontSize',18,'FontName','times')
        grid
subplot(312)
        stairs(0:k_end,vec_pi(:,2))
        ylabel('$\pi_2(k)$','fontsize',14,'interpreter','latex')
        set(gca,'FontSize',18,'FontName','times')
        grid
        ylim([0,0.8])
subplot(313)
stairs(0:k_end,vec_pi(:,3))
xlabel('# Iteration')
ylabel('$\pi_3(k)$','fontsize',14,'interpreter','latex')
set(gca,'FontSize',18,'FontName','times')
grid
ylim([0,0.4])
```

N-steps forward Unconditional (i.e *marginal*) Probability

- The N-steps forward Transition Probability $p_{ij}^{(n)}$ is a conditional probability,

$$p_{ij}^{(n)} = \Pr(\ x(n) = x_j \mid x(0) = x_i\ )$$

- From that, it can thus be possible determine the marginal (thus **unconditional**) probability to reach x_j at time n, from a given initial probability distribution.

- Following the Law of Total Probability and the Chapman-Kolgomorov equation, and by considering the initial probability distribution $\pi_i(0) = \Pr(x(0) = x_i\ )$, one derives

$$\begin{aligned} \pi_j(n) \ = \ \Pr(x(n) = x_j) \quad &= \quad \sum_{x_i \in X} \Pr(x(n) = x_j \mid x(0) = x_i) \cdot \Pr(x(0) = x_i) \\ &= \quad \sum_{x_i \in X} p_{ij}^{(n)} \cdot \pi_i(0) \ \equiv \ [\Pi(0) P^n]_{j\text{-th entry}} \end{aligned}$$

Examples (cont'd)

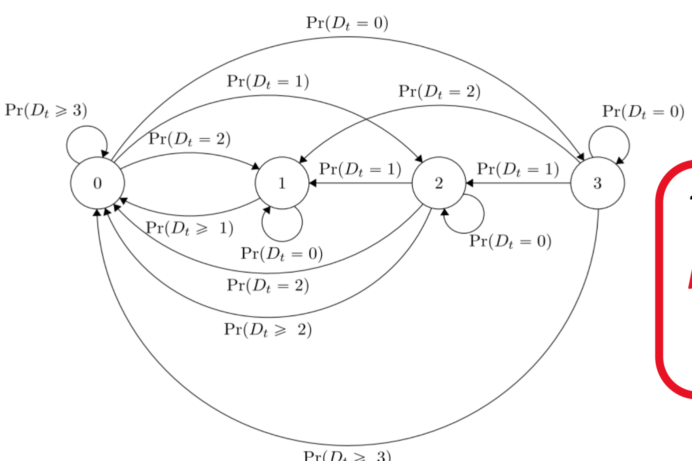
- Example 2: Inventory Example.
- If initial conditions were unknown we can assume uniform initial distribution

$$\Pi(0) = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

- then, what is the **unconditional probability** an order is placed in the 2[nd] week?

$$\Pr(\text{order is placed the 2-nd week}) = \Pr(x(2) = 0) = \pi_2(0)$$

$$\begin{bmatrix} \pi_0(2) & \pi_1(2) & \pi_2(2) & \pi_3(2) \end{bmatrix} = \Pi(0)\boldsymbol{P}^2 = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \cdot \begin{bmatrix} 0.249 & 0.285 & 0.300 & 0.164 \\ 0.283 & 0.251 & 0.232 & 0.232 \\ 0.351 & 0.319 & 0.232 & 0.097 \\ 0.249 & 0.285 & 0.300 & 0.164 \end{bmatrix}$$



$$\pi_0(2) = \frac{1}{4}(0.249) + \frac{1}{4}(0.283) + \frac{1}{4}(0.351) + \frac{1}{4}(0.249) = 0.283$$

This example allows to introduce another notion, the
*N-step forward **transition** probability:*

$$p_{ij}^{(n)} = \Pr(x(n) = x_j | x(0) = x_i)$$

$$\Longrightarrow \quad \pi_0(2) = \pi_0(0) \cdot p_{00}^{(2)} + \pi_1(0) \cdot p_{10}^{(2)} + \pi_2(0) \cdot p_{20}^{(2)} + \pi_3(0) \cdot p_{30}^{(2)}$$

N-steps forward Transition Probabilities

- It denotes the **conditional probability** to be in x_j after n time-units (or steps) given that at step k the SP was in x_i, i.e.,

$$p_{ij}(k,k+n) = \Pr(\, x(k+n) = x_j \mid x(k) = x_i \,)$$

- If the DT-MC is time-homogeneous $p_{ij}(k, k+n)$ the **SP** is **WSS** each p_{ij} does not depend by k (*actual time*) anymore, but only on n (i.e. *prediction horizon length*)

Important property:

$$p_{ij}(k,k+n) = p_{ij}^{(n)} = \sum_{\forall\, z} p_{iz}^{(m)} \cdot p_{zj}^{(n-m)}$$

$p_{ij}^{(n)}$ is the (i,j) entry of the n-th power of $\boldsymbol{P}$, i.e. $\boldsymbol{P}^{\mathrm{n}}$

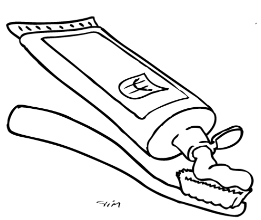
- **Note:** It corresponds to the product between the i-th row of $\boldsymbol{P}^m$ and the j-th column of $\boldsymbol{P}^{n-m}$, $m < n$.
- **Interpretation:** Due to time-homogeneity the $p_{ij}^{(n)}$ equals that of being after m steps in any state x_z from x_i and then reach x_j from each possible x_z, in the remaining $n - m$ steps.

How to use Chapman-Kolgomorov Equations

- To answer the following question:

> *"What is the probability that, starting from x_i the Markov chain will reach ("hit") state x_j after n steps?"*

- Example 4: Suppose 2 toothpaste brands, $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$, are on the market.
- Let $\{x(k), k = 0,1,2, \ldots\}$ be the SP associated with the k-th toothpaste's purchase by an average person between the 2 brands
- Suppose that



$$\Pr(\, x(k) = \boldsymbol{B}_1 \mid x(k-1) = \boldsymbol{B}_1 \,) = 0.9$$
$$\Pr(\, x(k) = \boldsymbol{B}_2 \mid x(k-1) = \boldsymbol{B}_2 \,) = 0.8$$

$$\boldsymbol{P} = \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix}$$

- Question 1: If a person is currently a $\boldsymbol{B}_2$ purchaser, what is the probability s/he will purchase $\boldsymbol{B}_1$ two purchases from now?

Examples (cont'd)

- Question 1: If a person is currently a $\boldsymbol{B}_2$ purchaser, what is the probability s/he will purchase $\boldsymbol{B}_1$ two purchases from now?

- Solution: We are looking for $\Pr(x(2) = \boldsymbol{B}_1 \mid x(0) = \boldsymbol{B}_2)$
- That is equivalent to the evaluation of $p_{21}^{(2)}$, namely,

$$\boldsymbol{P}^2 = \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix} \cdot \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix} = \begin{pmatrix} 0.83 & 0.17 \\ 0.34 & 0.66 \end{pmatrix} \rightarrow p_{21}^{(2)} = 0.34$$

- Question 2: What is instead the $\Pr(x(3) = \boldsymbol{B}_1 \mid x(0) = \boldsymbol{B}_1)$?
- Solution: From the Total probability Law it results that:

$$\boldsymbol{P}^3 = \boldsymbol{P}^2 \cdot \boldsymbol{P} = \begin{pmatrix} 0.83 & 0.17 \\ 0.34 & 0.66 \end{pmatrix} \cdot \begin{pmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix} = \begin{pmatrix} 0.781 & 0.219 \\ 0.438 & 0.562 \end{pmatrix} \rightarrow p_{11}^{(3)} = 0.781$$

States Classification of a DT-MC

- Definition: State x_j is reachable/ accessible from state x_i iff there is a **directed path** from x_i to x_j in the transition diagram of the DT-MC.

- Namely, x_j is reachable from x_i if there are chances to go to x_j from x_i after some steps

$$\exists\, n \,:\, p_{ij}^{(n)} > 0, \qquad \text{for some } n \geq 0$$

- This is written:

$$\boxed{x_j \leftarrow x_i}$$

- It further follows that

$$p_{ii}^{(0)} = \Pr(\, x(0) = x_i \mid x(0) = x_i \,) = 1$$

- Definition: States x_i and x_j communicates if x_j is accessible from x_i, and vice-versa.

- This is written:

$$\boxed{x_j \leftrightarrow x_i}$$

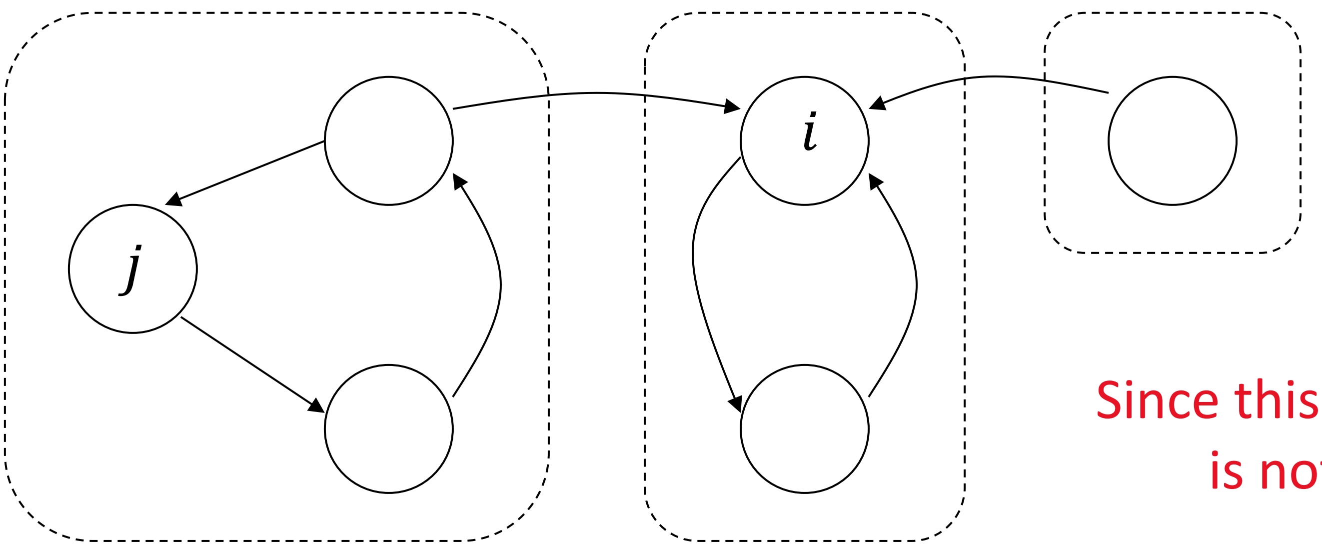
Connectedness notions on Graphs

Definition: Consider a directed graph $G(V,E)$. A vertex "i" is said reachable form vertex "j" iff, on $G(V,E)$, if there exists an oriented path form i to j.

Definition: A strongly connected component **(SSC)** denotes a collection of mutually reachable vertices

Definition: A directed graph $G(V,E)$ is said to be strongly connected (or irriducible) if every vertex is reachable from every other vertex.

- Have more then one SCCs allows reducing $G(V,E)$ into many subgraphs that are themselves strongly connected.

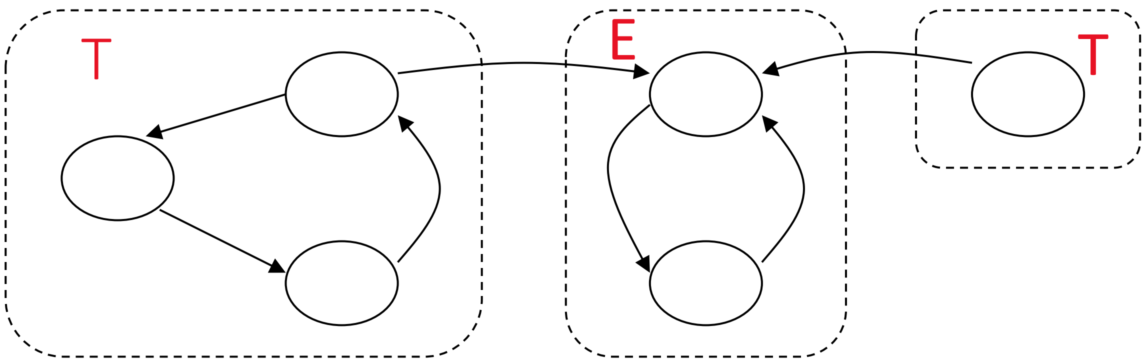


Vertex "i" is reachable from "j" , not the vice versa

Since this graph has 3 SCCs then, it is not strongly connected

States Classification of a DT-MC (cont'd)

- If two states communicate with each other, they are said to be in the same class



- Definition: A DT-MC is irreducible if all states belong to one class (all states communicate with each other, i.e. ($i \leftrightarrow j \; \forall \; i, j \in V$).

i.e. The **DT-MC** is **irreducible IFF** its transition graph is **strongly connected**

- It follows that:

If a **DT-MC** has **more than one class**, it is **reducible**.

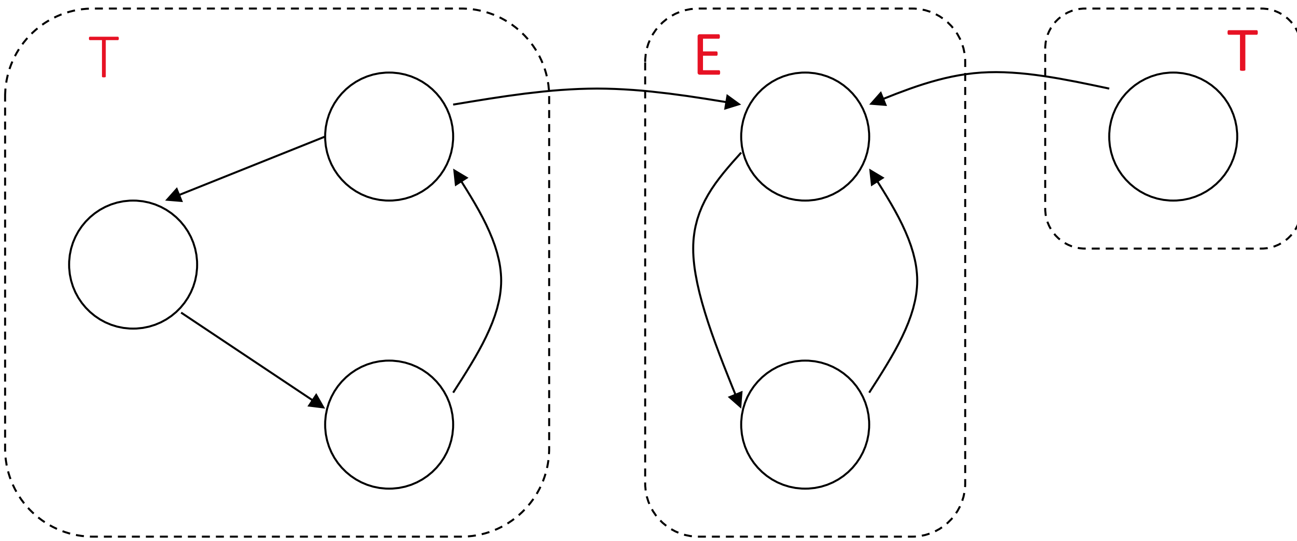
Strongly connected component of a graph

- Let us now remember the following definitions.

- A SCC is said:

1. Transitory (T) if, once leaved, it is not possible reach it again by following any oriented path.
2. Ergodic (E) or absorbing if, once reached, there are no path enabling the possibility to leave it in the future.

- It thus results that



N-steps return probability & Return probability

Definition: The n-step return probability $\rho_i(n)$ is the probability that the first return to x_i left at time 0 occurs **exactly** at time n is defined as next:

$$\rho_i(n) = \Pr(\, x(1) \neq x_i,\ x(2) \neq x_i, \ldots,\ x(n-1) \neq x_i\ ,\ x(n) = x_i \mid x(0) = x_i\ )$$

Definition: The return probability to state x_i, (i.e *to return sooner or later, since it accounts all the possible path*), is

$$\rho_i(\infty) = \sum_{n=1}^{\infty} \rho_i(n)$$

From that it is said that x_i is:

1) transient, if $\rho_i(\infty) < 1$ ($\exists\ 1 - \rho_i(\infty)$ of chances x_i will not be hit again)

2) recurrent, if $\rho_i(\infty) = 1$ (we will return on x_i sooner or later);

2.1) recurrent with period d if $\exists\, d > 1\ :\ d$ is the g.c.d of $\left\{ n > 0 \mid p_{ii}^{(n)} > 0 \right\}$

2.2.1) recurrent and aperiodic if $\exists\, d = 1\ :\ d$ is the g.c.d of $\left\{ n > 0 \mid p_{ii}^{(n)} > 0 \right\}$

2.2.2) absorbing if it is recurrent aperiodic and $p_{ii} = 1 \Rightarrow p_{ii}^{(n)} = 1$

see the next slide for the **meaning** of "d"

Period of a recurrent state

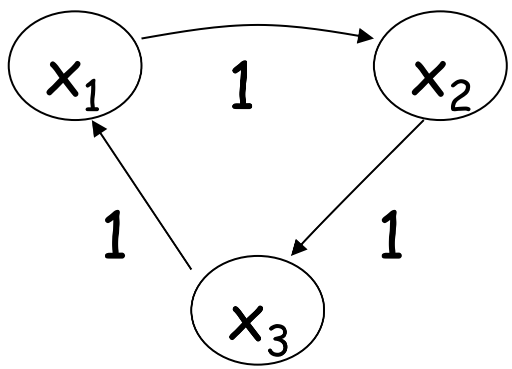
Interpretation: d is the greatest common divisor of the lengths of all the paths from state $x_i \rightarrow x_i$

- A state x_i is recurrent with period d if

$$\exists\, d \in \mathbb{N} \;:\; d \text{ is the g.c.d of } \left\{ n > 0 \mid p_{ii}^{(n)} > 0 \right\}$$

Let us study the recurrence properties of "x_1"

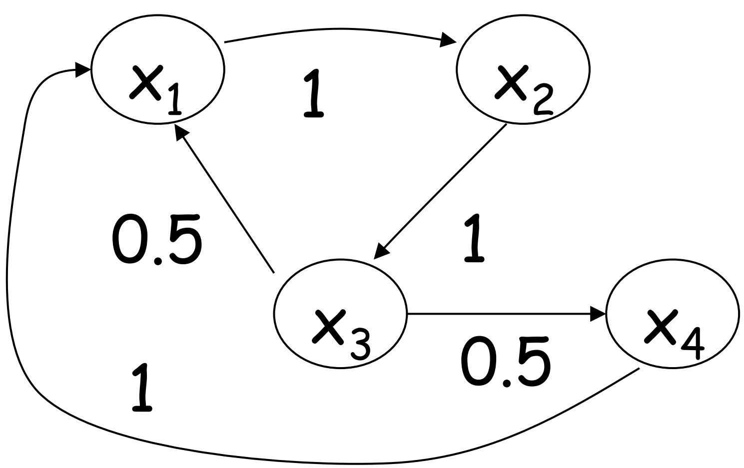
- Example 5:



$$\left\{ n > 0 \mid p_{11}^{(n)} > 0 \right\} = \{3, 6, 9, \ldots\}$$

g.c.d. is equal to 3
x_1 is recurrent with period 3

- Example 6:

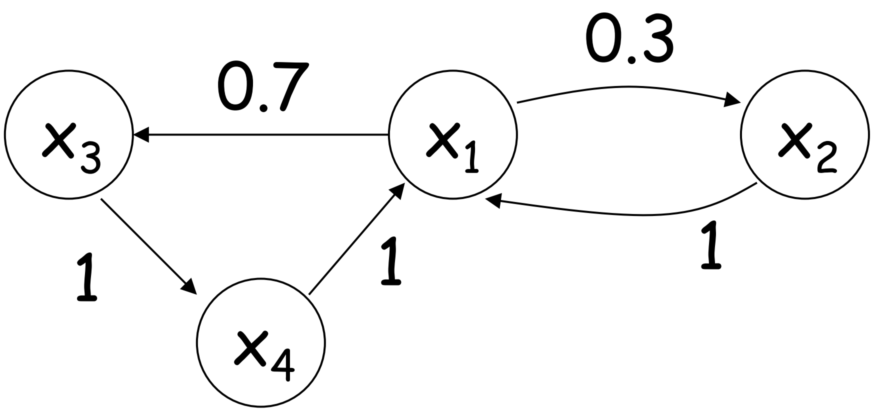


$$\left\{ n > 0 \mid p_{11}^{(n)} > 0 \right\} = \{3, 6, 9, \ldots, 4, 8, 12, \ldots, 7, 10, 11, 13, 14, \ldots\}$$

x_1 is recurrent aperiod, i.e. d=1
(the g.c.d between 2 prime numbers is always 1)

Period of a recurrent state

- Example 7:

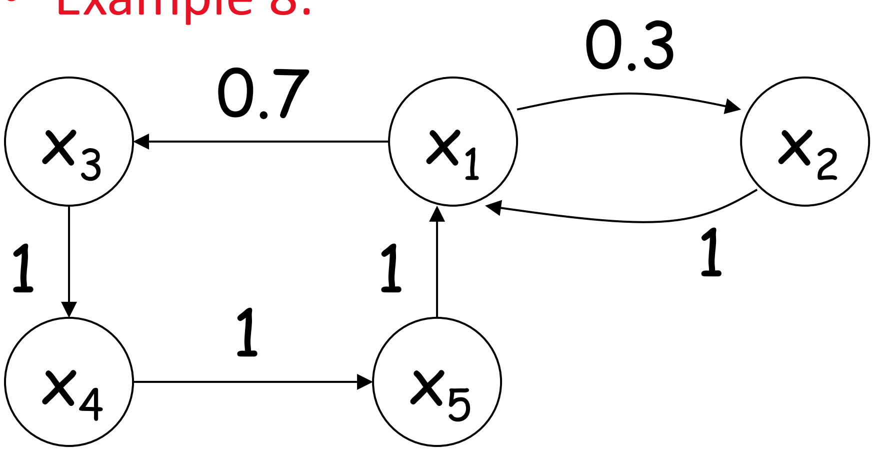


$$\left\{n > 0 \mid p_{11}^{(n)} > 0\right\} = \{3, 6, 9, \ldots \\ 2, 4, 8, \ldots \\ 5, 7, \ldots, \}$$

g.c.d. is 1

x_1 is recurrent aperiodic

- Example 8:



$$\left\{n > 0 \mid p_{11}^{(n)} > 0\right\} = \{2, 4, 6, 8, \ldots\}$$

g.c.d. is 2

x_1 is recurrent with period $d = 2$

Irreducible DT-MC

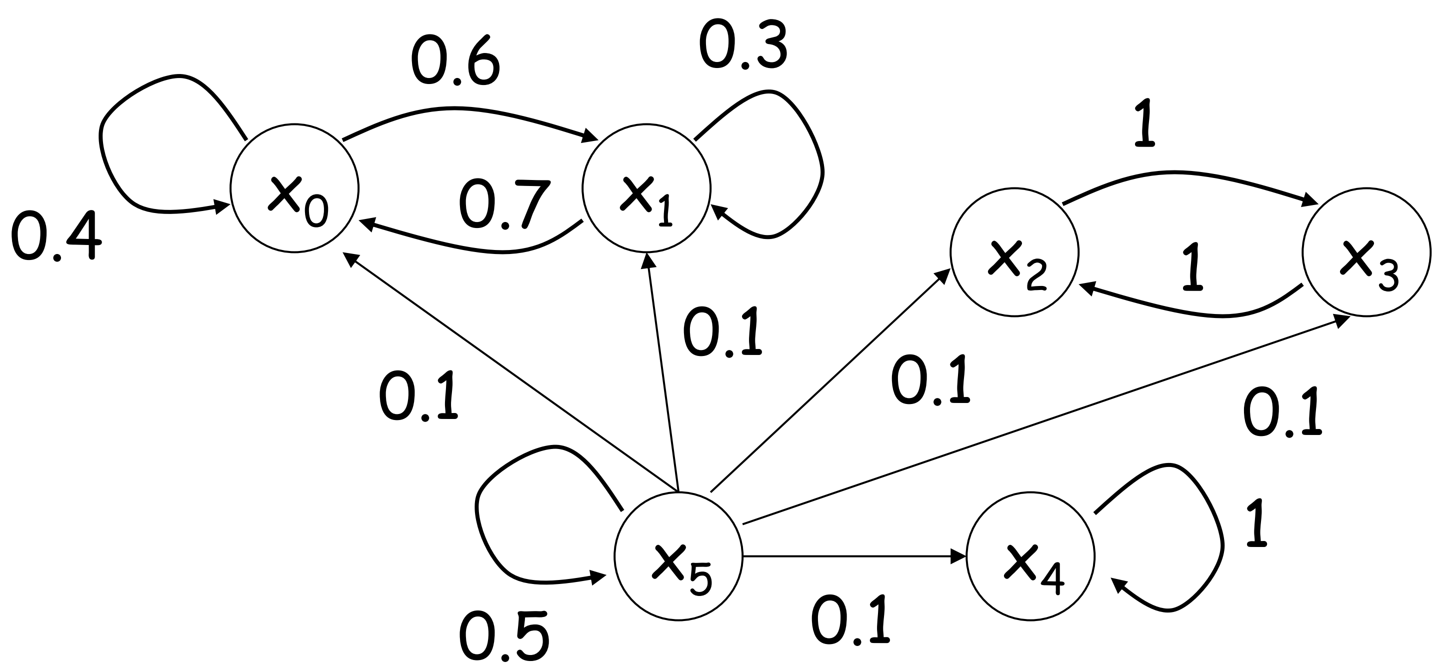
- If a DT-CM is irriducible then:

a) **Or** all the **states** are **aperiodic**;

b) **Or** all the **states** are **periodic** with the **same period**.

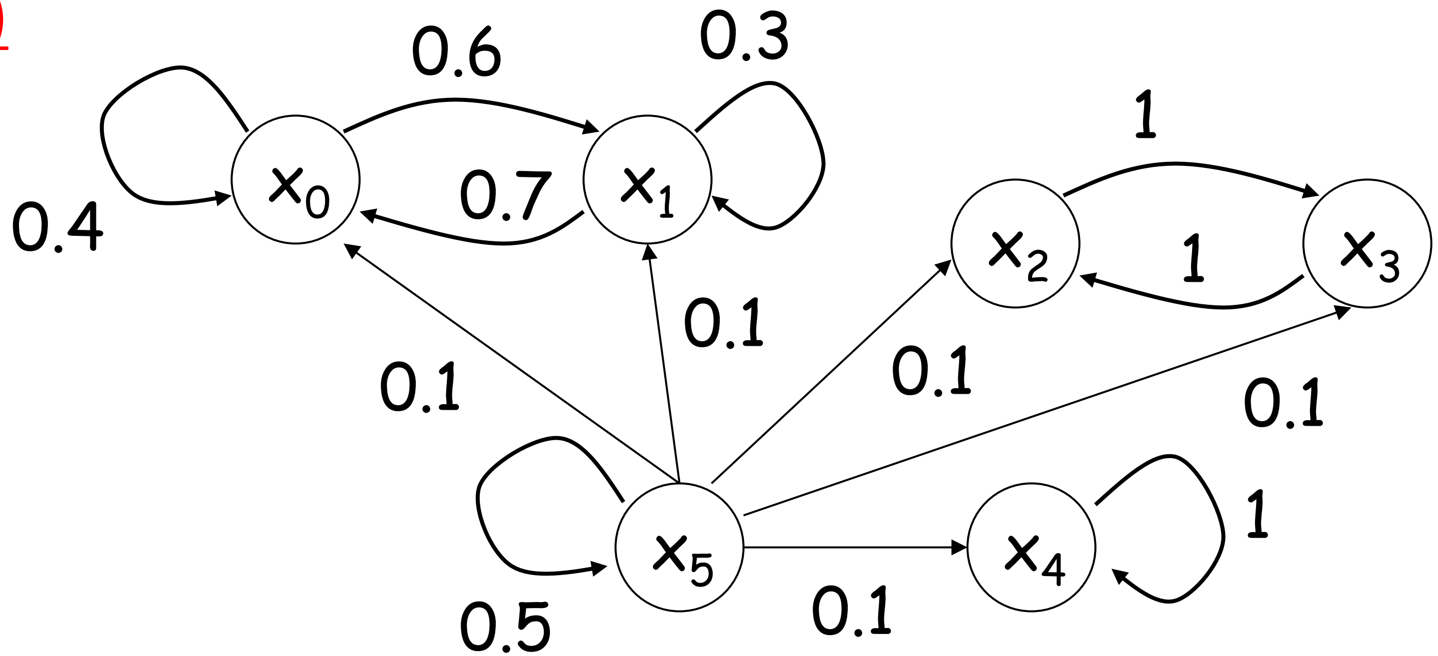
- If so, this implies **all** its **states** are **mutually reachable** and **recurrent**

- Example 11: Example of reducible DT-MC.



Examples (cont'd)

- Example 11:



- There are 4 strongly connected components:

a) 3 classes ergodic: $E_1 = \{x_0, x_1\}, E_2 = \{x_2, x_3\}, E_3 = \{x_4\}$

b) 1 class is transient: $T = \{x_5\}$

c) All states are recurrent except x_5, which is transient

d) States x_0, x_1 and x_4 are recurrent and aperiodic.

e) States x_4 is also absorbing

f) States x_2 and x_3 are periodic with period $d = 2$.

Summarizing….

All the previously properties can be verified by inspecting the graph associated with the DT-MC and they do not need counts!!!

All these sentences are equivalent:

a) State x_i is accessible from x_j (i.e. $x_i \rightarrow x_j$) in "n steps" if there exist at least an oriented path of length n from x_j to x_i.

b) Two states communicate ($x_i \leftrightarrow x_j$) if they belong to the same SCC;

c) A DT-MC is irreducible if its graph is **strongly connected**.

*It follows that if there is at least 1 transient state, then the DT-MC is **reducible***

d) A state x_i is transient iff it belongs to a transient SCC;

e) A state x_i that is not transient is called **recurrent** (upon entering on a recurrent state, the process will hit x_i to again sooner or later);

f) A state x_i is called absorbing it has a self-loop with probability 1

Summarizing….

All these sentences are equivalent:

a) If x_i is recurrent and periodic with period to $d > 1$,

- ✓ All the sequences originating from x_i to x_i, are such that $p_{ii}^{(n)} > 0$ only for those of length $n = d > 0$ and multiple of d.
- ✓ A sequence originating from x_i with length n not multiple of d do not terminate on x_i, this implies $p_{ii}^{(n)} = 0$

b) If x_i is recurrent and aperiodic, every path with length $n \geq \bar{n}$, where $\bar{n}$ is the minimum path length from $x_i \to x_i$, may terminate in x_i ,

i.e. $\exists\ \bar{n} : p_{ii}^{(n)} > 0\ \forall\ n \geq \bar{n}$

- ✓ This implies, after a while, the probability to hit again x_i is always nonzero

c) If x_i has a self-loop with probability 1, then x_i is for sure recurrent and aperidic. Indeed $p_{ii} = 1 \Rightarrow p_{ii}^{(n)} = 1 \Rightarrow d = 1$

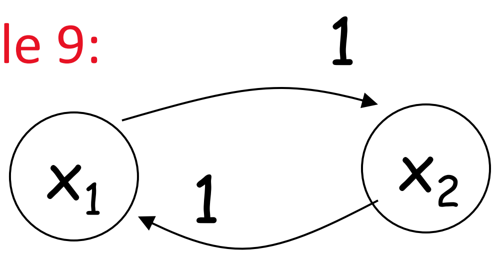
IMPORTANT: The periodicity of a state **depends only from the graph topology** and thus **it is <u>independent</u> from the edge's weight** (i.e. ***probabilities***).

Period of an ergodic component

- Let $\bar{\boldsymbol{P}} = [\bar{p}_{ij}]$ be the portion of the transition probability matrix $\boldsymbol{P}$ of a DT-MC, associated with **an ergodic SCC**, with entries $\bar{p}_{ij}$
- Let Λ be the set of all the periods' length "n" associated with the all the states of the considered ergodic component, i.e.,

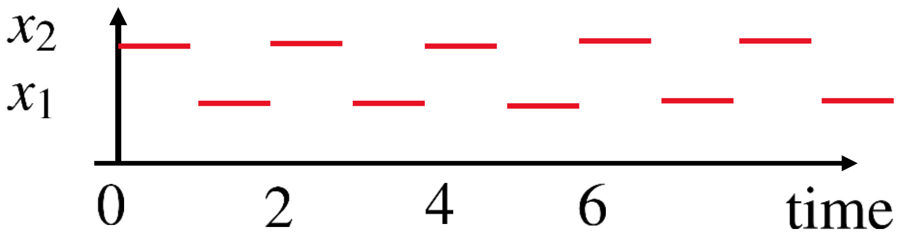
$$\Lambda = \left\{ n > 0 \mid \bar{p}_{ii}^{(n)} > 0,\ \forall\, i \right\}$$

- If $D = GCD\{\Lambda\}$, the ergodic SCC has period D;
- If $D = 1$ the ergodic SCC is aperiodic.
- Example 9:



$$\Lambda = \left\{ n > 0 \mid p_{ii}^{(n)} > 0,\ \forall\, i \right\} = \{2, 4, 6, \ldots\}$$

$$D = \mathrm{GCD}\{\Lambda\} = 2$$



Here it is easy see that all its realizations will have period $D = 2$

Period of an ergodic component (cont'd)

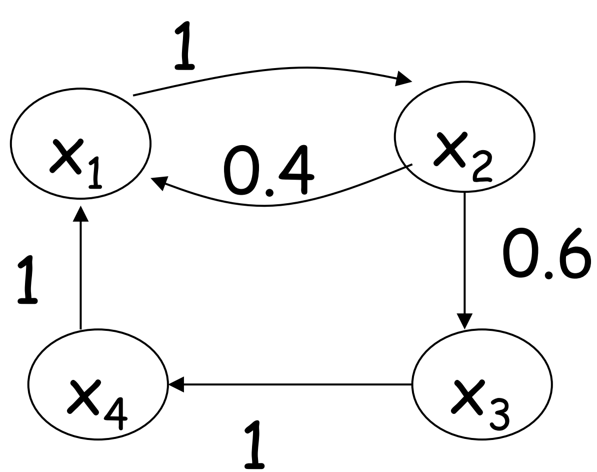
- Proposition 1: An ergodic component is periodic with period D if only if all its states are periodic with the same period d.
- This implies that the states of an ergodic component can be either

a) All periodic

b) All aperiodic

- Proposition 2: If an ergodic SCC hass at least **1 self-loop** then it is aperiodic

- Example 10:



All states have periodicity equal to

$$d = 2$$

Thus , this ergodic SCC has period

$$D = 2.$$

Limiting and stationary distribution

- Let us consider the evolution of a homogeneous DT-MC, for which it holds

$$\Pi(k+1) = \Pi(k)\boldsymbol{P}$$

Definition: A DT-MC admits a limiting distribution Π_ℓ if, for a given initial distribution $\Pi(0) = [\pi_1(0), \pi_2(0), \dots]$ it holds that

$$\exists \quad \Pi_\ell = \lim_{k\to\infty} \Pi(k+1) = \lim_{k\to\infty} \Pi(0)\boldsymbol{P}^k \quad \Longrightarrow \quad \Pi_\ell \equiv \Pi(\infty)$$

- **Definition:** A probability distribution Π_s is called stationary (or invariant) if and only if it satisfies the so-called "balanced equations", namely

$$1)\ \ \Pi_s = \Pi_s \boldsymbol{P} \qquad\qquad 2)\ \ \Pi_s \cdot \mathbf{1} = 1 \quad \equiv \quad \sum_i \pi_{i,s} = 1$$

- For a DT-MC a limiting distribution is also stationary, **not true** the vice-versa.
- Indeed, there are not guarantees that every Π_s would be reached from $\Pi(0)$

Limiting and stationary distribution

- Remark 1: Not necessarily Π_ℓ may exists

$$\nexists \quad \Pi_\ell = \lim_{k\to\infty} \Pi(k+1) = \lim_{k\to\infty} \Pi(0)\boldsymbol{P}^k$$

To exists the limit must converges to a constant vector Π_ℓ

- Remark 2: The above limit it is not necessarily unique, and it may be influenced by the considered initial probability distribution

$$\Pi(0) = [\pi_1(0), \pi_2(0), \ldots]$$

If the limit changes with $\Pi(0)$ the process is not erogdic

- Remark 4: Not all stationary distributions Π_s are limiting distributions Π_ℓ.

- Remark 5: If a Π_s is reached at some step k probabilities will remain unchanged as the time progresses

Ergodicity and Stationary Distribution

- Definition: A DT-MC is said to be ergodic if and only if:

1) If it admits an unique limiting probability distribution, namely,

$$\exists\, \Pi_\ell = \lim_{k\to\infty} \Pi(k+1) = \lim_{k\to\infty} \Pi(0)\boldsymbol{P}^k \quad \forall \quad \Pi(0) \qquad \Longrightarrow \qquad \Pi_s \equiv \Pi_\ell$$

If these conditions are verified, to understand the behavior of the DT-MC we can simply study one their possible realizations.

- Remark: If a DT-MC is ergodic, the calculus of its Π_ℓ reduces to the calculus of its stationary component Π_s, called:

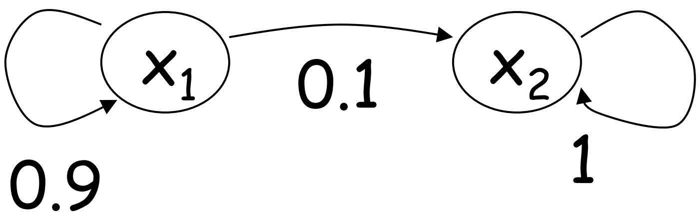
Balance equations

$$\Pi_s = \Pi_s \boldsymbol{P}$$

$$\Pi_s \cdot \mathbf{1} = 1 \quad \equiv \quad \sum_i \pi_{i,s} = 1 \qquad \Longrightarrow \qquad \Pi_s \equiv \Pi_\ell$$

Calculus of a limiting distribution from its definition

- Example 12: Calculus of the limiting distribution



$$P = \begin{pmatrix} 0.9 & 0.1 \\ 0 & 1 \end{pmatrix}$$

2 SCC $\begin{cases} \text{1 Transient} \\ \text{1 Ergodic} \end{cases}$

- Due to its definition $\lim_{k\to\infty} \Pi(k+1) = \lim_{k\to\infty} \Pi(0)\boldsymbol{P}^k$ we need to find $\boldsymbol{P}^k$

$$\boldsymbol{P}^2 = \begin{pmatrix} 0.9 & 0.1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0.9 & 0.1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0.9^2 & 0.9 \cdot 0.1 + 0.1 \\ 0 & 1 \end{pmatrix}$$

$$\boldsymbol{P}^3 = \begin{pmatrix} 0.9^2 & 0.9 \cdot 0.1 + 0.1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0.9 & 0.1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0.9^3 & 0.9^2 \cdot 0.1 + 0.9 \cdot 0.1 + 0.1 \\ 0 & 1 \end{pmatrix}$$

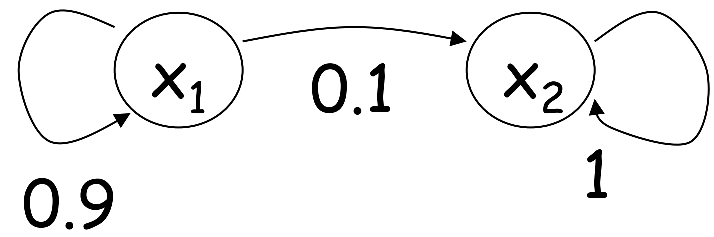
$$\boldsymbol{P}^k = \begin{pmatrix} 0.9^k & 0.1 \cdot \sum_{i=0}^{k-1} 0.9^i \\ 0 & 1 \end{pmatrix}$$

$$(1-\alpha) \cdot \sum_{i=0}^{k-1} \alpha^i = (1-\alpha^k)$$

$$\boldsymbol{P}^k = \begin{pmatrix} 0.9^k & 1-0.9^k \\ 0 & 1 \end{pmatrix}$$

Proof : $(1-\alpha)(1+\alpha+\alpha^2) = (1+\alpha+\alpha^2) - (\alpha+\alpha^2+\alpha^3) = (1-\alpha^3)$

Examples (cont'd)



- Let $\Pi(0) = [\pi_1(0),\ \pi_2(0)]$

$$\Pi(k) = \Pi(0)\boldsymbol{P}^k = [\pi_1(0),\ \pi_2(0)] \cdot \begin{pmatrix} 0.9^k & 1-0.9^k \\ 0 & 1 \end{pmatrix} =$$

$$= \left[0.9^k \pi_1(0) \quad , \quad \pi_1(0) + \pi_2(0) - 0.9^k \pi_1(0)\right]$$

- Then, the limiting distribution is

$$\Pi_\ell = \lim_{k\to\infty} \Pi(k) = \lim_{k\to\infty} \left[0.9^k \pi_1(0) \quad , \quad \pi_1(0) + \pi_2(0) - 0.9^k \pi_1(0)\right] =$$

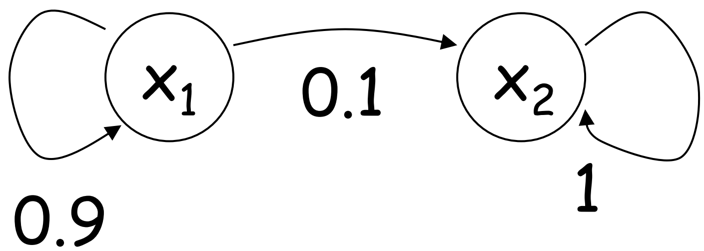
$$= \left[0 \quad , \quad \pi_1(0) + \pi_2(0)\right] = \left[0 \quad , \quad 1\right]$$

Since Π_ℓ does not depend on $\Pi(0)$, then this DT-MC is ergodic.

- Remark: The same result can be obtained by solving its balanced equation

Graphical consideration on the ergodicity to finite-state DT-MCs

- Example 12:



2 SCC $\begin{cases} \text{1 Transient} \\ \text{1 Ergodic} \end{cases}$

- Remark: The ergodicity **depends only** by the graph **not** by the **weights**.
- It can be evaluated by analyzing the graph, in place of computing analytically Π_ℓ.

- By observing the above transition graph, we noticed that:

a) The state x_1 consists of a transient SCC, thus sooner or later it will leaved

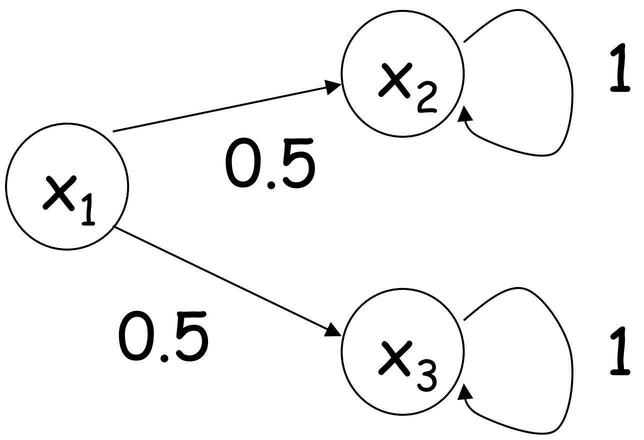
b) The state x_2 consists of an ergodic SCC, thus sooner or later it will be reached.

- Once the ergodic SCC is reached, then the DT-MC will never leave it.
- This occur independently on the given $\Pi(0)$

Graphical consideration on the ergodicity to finite-state DT-MCs

- Example 13:

$$\boldsymbol{P} = \begin{pmatrix} 0 & 0.5 & 0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$



$$\boldsymbol{P}^k = \boldsymbol{P} \text{ (idempotent)}, \quad \forall\, k \in \mathbb{N}^+$$

$$\Pi(k) = \Pi(0)\boldsymbol{P}^k = [\, 0 \;,\; 0.5\pi_1(0) + \pi_2(0\,) \;,\; 0.5\pi_1(0) + \pi_3(0) \,]$$

if $\quad \Pi(0) = [\,1\,,\,0\,,\,0\,] \quad \Longrightarrow \quad \Pi(k) = \Pi_\ell = [\,0\,,\,0.5\,,\,0.5\,]$

if $\quad \Pi(0) = [\,0\,,\,1\,,\,0\,] \quad \Longrightarrow \quad \Pi(k) = \Pi_\ell = [\,0\,,\,1\,,\,0\,]$

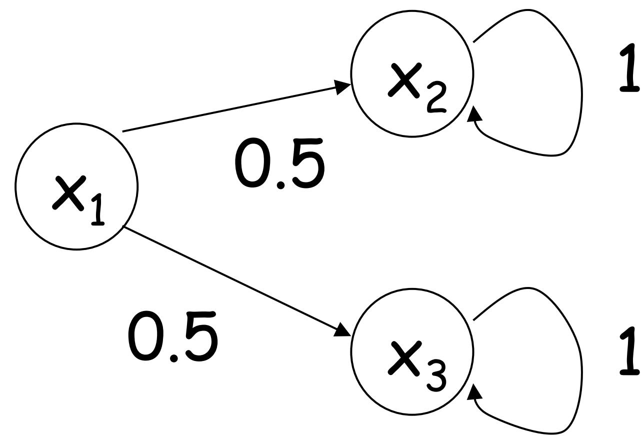
Etc…

Since Π_ℓ is not unique, this DT-MC is not ergodic.

Graphical consideration on the ergodicity to finite-state DT-MCs

- Example 13:

$$P = \begin{pmatrix} 0 & 0.5 & 0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$



- Remark: The previous result was predictable from the transition graph, since it has 2 absorbing SCC.

- If a process admits many Π_ℓ, then it will admit infinite many Π_s, each satisfying their balanced equations (not true the vice versa):

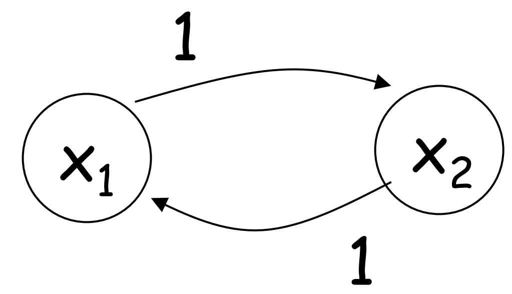
$$\begin{cases} \Pi_s = \Pi_s \cdot \boldsymbol{P} \\ \sum_i \pi_{i,s} = 1 \end{cases} \Rightarrow \begin{cases} \pi_{s,1} = 0 \\ \pi_{s,2} = 0.5\pi_{s,1} + \pi_{s,2} \\ \pi_{s,3} = 0.5\pi_{s,1} + \pi_{s,3} \\ \pi_{s,1} + \pi_{s,2} + \pi_{s,3} = 1 \end{cases} \Rightarrow \begin{cases} \pi_{s,1} = 0 \\ \pi_{s,2} = \pi_{s,2} \\ \pi_{s,3} = 1 - \pi_{s,2} \end{cases}$$

$$\Rightarrow \quad \Pi_s = [\, 0 \,,\ a \,,\ 1-a \,] \qquad \forall\, a \in [0,1]$$

Graphical consideration on the ergodicity to finite-state DT-MCs

- Example 14:

$$\boldsymbol{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \Pi(0) = [\ \pi_1(0)\ ,\ \pi_2(0)\ ]$$



- It results that

$$\boldsymbol{P}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \boldsymbol{P}^3 = \boldsymbol{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\boldsymbol{P}^{2k} = \boldsymbol{P}^2$$
$$\boldsymbol{P}^{2k+1} = \boldsymbol{P}$$

$$\Pi_\ell = \lim_{k \to \infty} \Pi(0)\boldsymbol{P}^k = \begin{cases} [\ \pi_2(0)\ ,\ \pi_1(0)] & \text{if} \quad k \text{ is odd} \\ [\ \pi_1(0)\ ,\ \pi_2(0)] & \text{if} \quad k \text{ is even} \end{cases}$$

The limit exists only for

$$\Pi(0) = [\ 0.5\ ,\ 0.5\ ]$$

while for different initial conditions does not exists

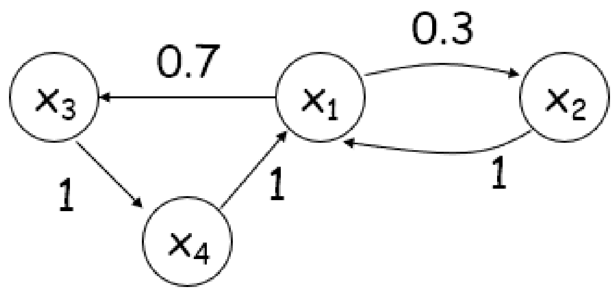
This process is not ergodic

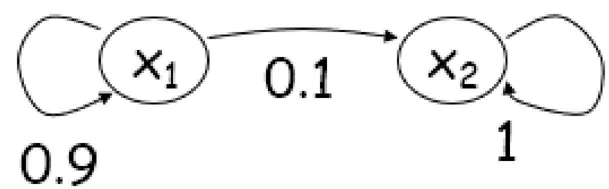
Graphical consideration on the ergodicity to finite-state DT-MCs

- By the previous examples show that:

a) A DT-MC is ergodic if Π_ℓ is **unique** and **independent** from $\Pi(0)$

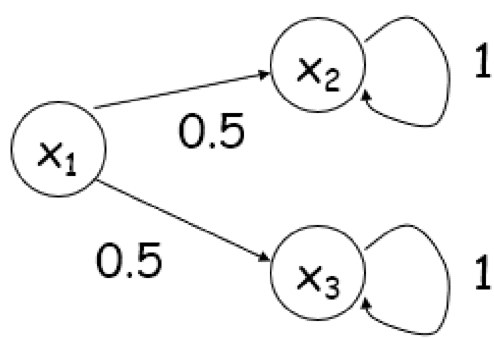
See Examples 7 or 12

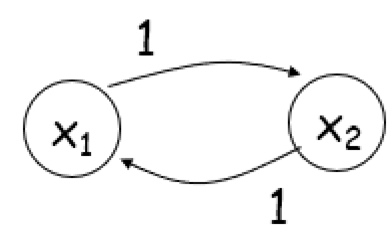




b) If Π_ℓ **not exists for some** $\Pi(0)$, or Π_ℓ is **not unique** $\forall\ \Pi(0)$
the DT−MC is not ergodic

See Example 13 or 14



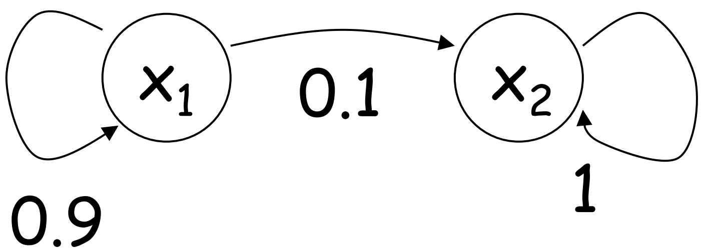


Criteria for ergodicity for finite-state DT-MCs: Graph's Criteria

Graph's Criteria (necessary condition)

- Theorem: Necessary condition (but **not sufficient**) for a finite-state time-homogeneous DT-MC to be **ergodic** is that its transition graph has a single ergodic component.

- Example 12:



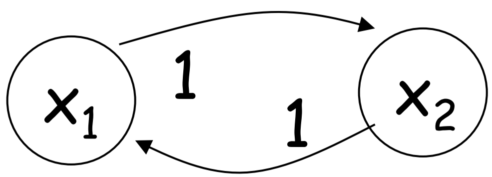
- Since $\{x_2\}$ is the only ergodic SCC, the necessary condition is verified.

This is not enough to conclude that is ergodic

since the limiting distribution could not be unique

Criteria for ergodicity for finite-state DT-MCs: Graph's Criteria

- Example 14:



The **necessary condition is verified**, since

$$\{x_1, x_2\}$$

is the **unique ergodic** (or **recurrent**) **component**.

- However, $\{x_1, x_2\}$ has period $D = 2$

$$\text{if } \pi_1(0) \neq \pi_2(0) \nexists\, \Pi_\ell = \lim_{k\to\infty} \Pi(0)\boldsymbol{P}^k \quad i.e. \quad \begin{matrix} [\,\pi_2(0)\;,\;\pi_1(0)] & \text{if} & k \text{ is odd} \\ [\,\pi_1(0)\;,\;\pi_2(0)] & \text{if} & k \text{ is even} \end{matrix}$$

The limit exists only for

$$\Pi(0) = [\,0.5\;,\;0.5\,]$$

while for different initial conditions does not exists

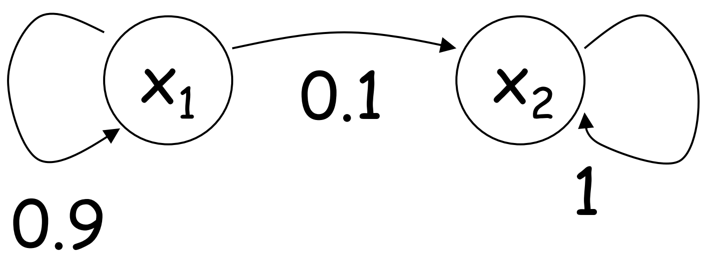
Since the limit is not unique because Π_ℓ depends by $\Pi(0)$ and sometimes id does not exist, this MC is NOT ergodic *(informaly, we can say it depends by the day we test it, i.e. by $\Pi(0)$)*

Criteria for ergodicity for finite-state DT-MCs: Graph's Criteria

Graph's Criteria (necessary and sufficient condition)

- Theorem: Necessary and sufficient condition for a finite-state time-homogeneous DT-MC to be ergodic is that its transition graph has a single ergodic SCC that is also aperiodic.

- Example 12:



- Since $\{x_2\}$ is the only ergodic component, and it is also aperiodic.

This DT-MC is ergodic

Criteria for ergodicity for finite-state DT-MCs: Eigenvalue Criteria

Eigenvalue Criteria

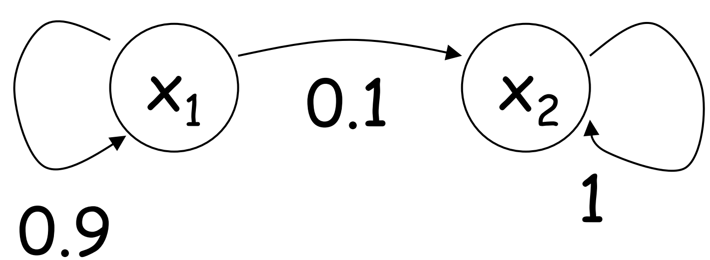
- Theorem: Let λ_i with $i = 1, \ldots, n$, be the eigenvalues of the transition probability matrix $\boldsymbol{P}$ of a finite-state time-homogeneous DT-MC .
- Since $\boldsymbol{P}$ is row stochastic then $1 = \lambda_1 \geq \lambda_2 \geq \cdots$.
- Then, the DT-MC is ergodic **if and only if (IFF)**:

$$|\lambda_i| < 1 \qquad \forall \quad i = 2, 3, \ldots$$

INTERPRETATION: A DT-MC is ergodic if $\boldsymbol{P}$ has at most a single eigenvalue in 1, while the other are **inside the unit circle** $|\lambda_i| < 1$

- Example 12:

$$\boldsymbol{P} = \begin{pmatrix} 0.9 & 0.1 \\ 0 & 1 \end{pmatrix}$$

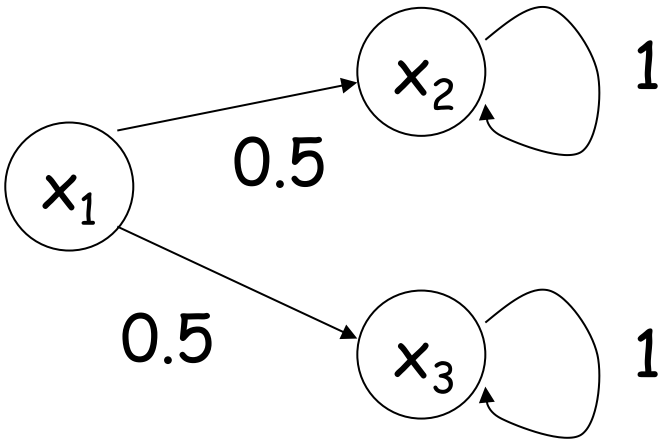


$$\Lambda = \text{eig}\{\boldsymbol{P}\} = \{\, 1\,, 0.9\}$$

This DT-MC is ergodic

Criteria for ergodicity for finite-state DT-MCs: Eigenvalue Criteria

- Example 13:

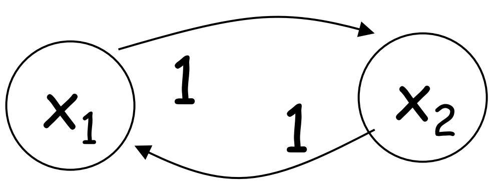


$$P = \begin{pmatrix} 0 & 0.5 & 0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\Lambda = \text{eig}\{P\} = \{\ 1\ ,\ 1\ ,\ 0\ \}$$

This DT-MC is NOT ergodic

- Example 14:



$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\Lambda = \text{eig}\{P\} = \{\ 1\ ,\ -1\ \}$$

This DT-MC is NOT ergodic

Reversible DT-MC

Theorem: An **irreducible DT-MC** is **time-reversible** if there

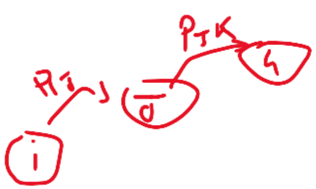
$$\exists\, \Pi = [\pi_1,\ \pi_2, \cdots, \pi_N] \ : \ \text{satisfying} \quad \begin{aligned} &\textstyle\sum_{i=1}^{n} \pi_i = 1 \\ &\pi_i \cdot p_{ij} = \pi_j \cdot p_{ji} \quad \forall\, i, j \quad \text{(detailed balance equations)} \end{aligned}$$

If so, **vector** Π is a **stationary probability distribution**, i.e. $\Pi = \Pi_s$

- **Proof:** Suppose the DT-MC is reversible, then note that

$$\sum_{i=1}^{n} \pi_i \cdot p_{ij} = \sum_{i=1}^{n} \pi_j \cdot p_{ji} = \pi_j \sum_{i=1}^{n} p_{ji} = \pi_j \quad \Longrightarrow \quad \Pi \ : \ \Pi = \Pi \cdot P$$

- The term time-reversible comes from the fact, at the steady-state, a MC *run forward, e.g.* $\{x_i \to x_j \to x_k\}$ and the same *run backword* $\{x_k \to x_j \to x_i\}$, are indistinguable



$$\overbrace{(\pi_i \cdot p_{ij}) \cdot p_{jk}}^{\Pr(x_i \to x_j \to x_k)} = (\pi_j \cdot p_{ji}) \cdot p_{jk} = (\pi_j \cdot p_{jk}) \cdot p_{ji} = \overbrace{(\pi_k \cdot p_{kj}) \cdot p_{ji}}^{\Pr(x_k \to x_j \to x_i)}$$

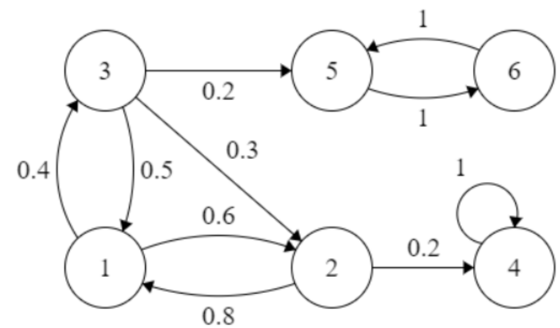
- **Remark:** An irreducible DT-MC with symmetric $\boldsymbol{P} = \boldsymbol{P}^T$ is reversible. For this special DT-MC its stationary distribution is $\Pi_s = Uni_d(0, N) = \left[\frac{1}{N}, \dots, \frac{1}{N}\right]$

- **Remark:** Reversible does not implies the ergodicity (cf. Example 12 with Example 14)

- **Reversible MCs** have important applications in **computer science** (see *randomized decision making, e.g. consensus and distr. opt.*), **queueing theory** (see *PASTA property*), biology, etc…

Review on steady-state behaviour

Consider the DT-MC with 1 transient and 2 absorbing SCCs



$T_1 = \{1,2,3\}$
$E_1 = \{5,6\}$
$E_2 = \{4\}$

Question: *What happen in the long run if the DT-MC has multiple ergodic SCCs?*

- If at $t = 0$, we start in state $j \in E_m \ \forall\ m$, we can compute $\Pi(\infty)$ as next:
  - If E_m is aperiodic, we can consider the reduced DT-MC $C'(E_m, P', \Pi'(0))$, then solve

$$\Pi'_s = \Pi'_s P' \qquad \Pi_\ell :\ \pi_{\ell,j} = \pi'_{s,j}\ \forall\ j \in E_m$$

while

$$\pi_{\ell,j} = 0 \quad \forall\ \ j \in X - E_m$$

E.g. if $x(0) = 4 \in E_1$

$$\Pi(\infty) = [\overbrace{0,0,0,}^{T_1}\ \overbrace{1}^{E_1},\ \overbrace{0,0,0,}^{E_2}]$$

  - If E_m is periodic, we have to verify if there $\exists\ \Pi'_\ell$ for the reduced $C'(E_m, P', \Pi'(0))$

E.g. if $\pi_5(0) = \pi_6(0) = 0.5$

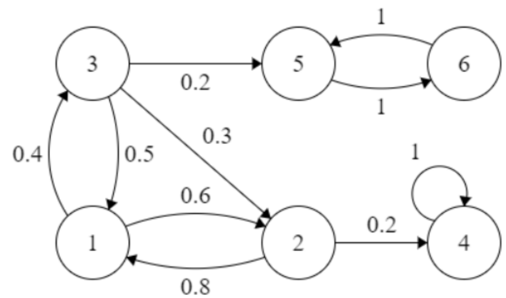
$\exists\ \Pi(\infty)$, otherwise it does not exists (see Example 14)

- **Question:** What if $i \in T_1$? How do we know if whether I'm going to E_1 or E_2?
- I can't! It is a random variable, thus:

It may be of interest evaluate $\Pr(\exists\ k \geq 0 : x(k) \in E_m | x(0) = i) \quad \forall\ m$

Absorption probability: $a_i = \Pr(\exists\, k \geq 0 : x(k) \in E_m | x(0) = i)$

- Example: Which is the probability to reach E_2 given that the process starts in state i?

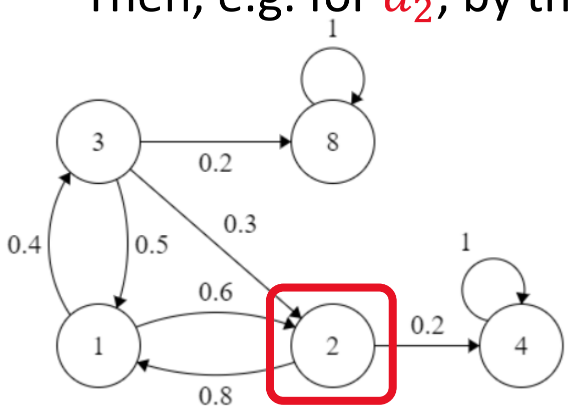


$$T_1 = \{1,2,3\}$$
$$E_1 = \{5,6\}$$
$$E_2 = \{4\}$$

- Let a_i be the absorption probability to hit E_2 from state i, how can we calculate it?

Firstly, we can set: $a_5 = a_6 = 0, a_4 = 1$ (if we hit 4 we couldn't hit neither 5 nor 6)

- To evaluate a_1, a_2, a_3 we need to "**reduce**" the MC by aggregating all the recurrent states belonging to the **same absorbing class** into an absorbing state (e.g. $8 \equiv \{5,6\}$)
- Then, e.g. for a_2, by the Total probability Law, and the Markov property, one has

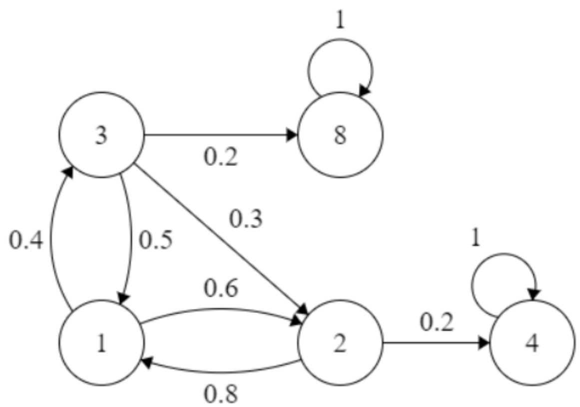


$$\begin{aligned} a_2 &= a_4 \cdot \Pr(x(1) = 4 | x(0) = 2) \; + \; a_1 \cdot \Pr(x(1) = 1 | x(0) = 2) \\ &= a_4 \cdot p_{24} \; + \; a_1 \cdot p_{21} \; = \; a_4 \cdot 0.2 + a_1 \cdot 0.8 \\ &= 1 \cdot 0.2 \; + \; a_1 \cdot 0.8 \end{aligned}$$

What about a_1?

Absorption probability: $a_i = \Pr(\exists\, k \geq 0 : \; x(k) \in E_m | x(0) = i)$

- Due to the Markov property each a_i depends only from the current state
- Thus, by doing the same for all states, a well-posed linear system takes place



$$\begin{cases} a_2 = & 0.2a_4 + 0.8a_1 \\ a_1 = & 0.4a_3 + 0.6a_2 \\ a_3 = & 0.2a_8 + 0.3a_2 \\ a_4 = & 1 \\ a_8 = & 0 \end{cases} \Rightarrow \begin{cases} a_1 = & \frac{18}{28} \\ a_2 = & \frac{20}{28} \\ a_3 = & \frac{15}{28} \\ a_4 = & 1 \\ a_8 = & 0 \end{cases}$$

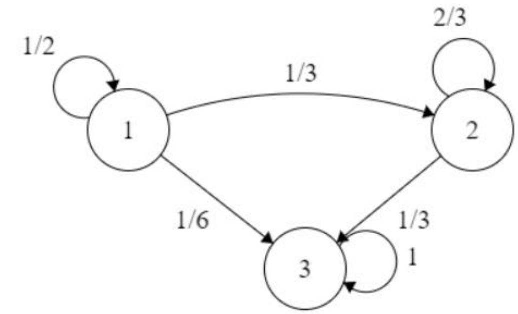
- As a consequence, and let $b_i = Pr(\exists\, k \geq 0 : \; x(k) = 8 | x(0) = i)$ $\Rightarrow$ $\begin{cases} b_1 & = 1 - a_1 \\ b_2 & = 1 - a_2 \\ b_3 & = 1 - a_3 \\ b_4 & = 0 \\ b_8 & = 1 \end{cases}$

Theorem: Consider a finite DT-MC with only **transients** and **absorbing states**.

- Let $a_i = \Pr(\exists k \geq 0 : x(k) = s | x(0) = i)$, where s is an absorbing state.
- Then we have that $a_s = 1$, $a_j = 0$ if j is any other absorbing state.
- Finally to find the unknown a_i we need to solve the well posed linear system:

$$a_i = \sum_{\forall j} p_{ij} \cdot a_j \quad \forall\, i \neq s \neq j$$

Mean time to absorption: $\mu_i = E[T|X_0 = i]$



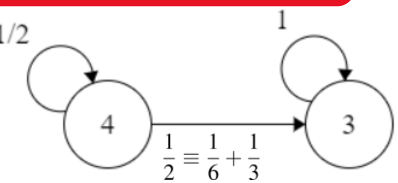
Example A: Consider the ergodic DT-MC on the right

- For that it results $\forall\, \Pi(0) \quad \exists\, \Pi_\ell = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$

Question: How many time steps on average would we need to reach its absorbing state?

Answer: We can't know it, it is a r.v. T function of many random variables… one for each path enabling the MC to hit state 3, given any possible state i reached in the middle

Since T is unknown it is of interest its (conditional) mean, namely $\mu_i = E[T|X_0 = i]$



Prelim. Example A': To warm up consider the "*reduced MC*" where $4 \equiv \{1,2\}$

The r.v. modeling the time to leave $x_0 = 4$ is $t_4 \sim Geo(\,1/2\,)$ ➡ $\mu_4 = \mathrm{E}[t_4] = \dfrac{1}{0.5} = 2$

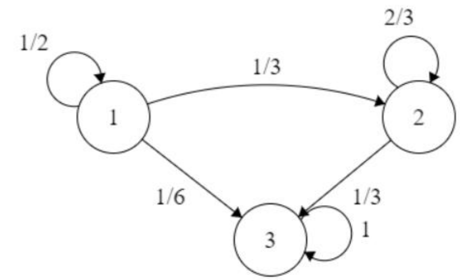
Example A (cont'd): Compared with **Example A'**, from $\{1\}$ we may move to $\{2\}$ in **1 step with a higher probability**, then from $\{2\}$ it results $\mu_2 = E[Geo(1/3)] = 3$, thus it is reasonable that $\mu_1 > 2$

How to account the presence of all the possible paths from $\{1\}$ to $\{3\}$?

Mean time to absorption (**difficult way…!!!**)

Method 1: Let T be the r.v. of the #transition to reach 3 from 1, then we want $\mu_1 = E[T]$,

$$T = \min(k \geq 0 \;:\; X_k = 3 \mid X_0 = 1)$$



We can rewrite it as $T = T^* + dT$ where $T^* \leq T$

$$T^* = \min(k \geq 0 \;:\; X_k \neq 1 \mid X_0 = 1) \sim Geo\left(\frac{1}{3} + \frac{1}{6}\right) \implies \mathrm{E}[T^*] = 2$$

mean time to leave {1}

$$dT = \min(k \geq 0 \;:\; X_k = 3 \mid X_{T^*} \in \{2,3\})$$

min #transition from $X_{T^*} \in \{2,3\}$ to 3

From the Law of Total Expectation one derives

$$\sum_{\forall k} k \cdot \Pr(dT = k | X_{T^*} = i)$$

$$\mathrm{E}[dT] = \sum_{\forall k} k \cdot \overbrace{\sum_{\forall i} \Pr(dT = k | X_{T^*} = i) \cdot \Pr(X_{T^*} = i)}^{\Pr(dT=k)\,:\,\text{Law of Total Probability}} = \overbrace{\sum_{i} \mathrm{E}[dT | X_T^* = i] \cdot \Pr(X_T^* = i)}^{\text{Law of Total Expectation}}$$

$$= \mathrm{E}[dT | X_{T^*} = 3] \cdot \overbrace{\Pr(X_{T^*} = 3)}^{\Pr(X_{T^*}=3|X_{T^*})} + \mathrm{E}[dT | X_{T^*} = 2] \cdot \overbrace{\Pr(X_{T^*} = 2)}^{\Pr(X_{T^*}=2|X_{T^*})}$$

$$= \overbrace{0}^{\mathrm{E}[dT|X_{T^*}=3]} \underbrace{\frac{\frac{1}{6}}{\frac{1}{6}+\frac{1}{3}}}_{1/3} + \overbrace{3}^{\mathrm{E}[Geo(1/3)]} \underbrace{\frac{\frac{1}{3}}{\frac{1}{6}+\frac{1}{3}}}_{2/3} = 2 \quad \Rightarrow \quad \mu_1 = \mathrm{E}[T^*] + \mathrm{E}[dT] = 4$$

From the Bayes Theorem

$$\Pr(X_{T^*} = j | X_{T^*}) = \frac{1 \cdot p_{1j}}{p_{12} + p_{13}} \qquad j = 2, 3$$

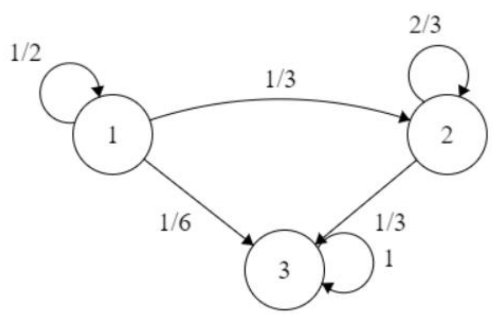
Mean time to absorption (**smart and easy way**)

Method 2: Study the first (i.e. *at 1 step*) hitting time to $\{3\}$ from all $x(0) = j \in X$ by the r.v.

$$T = \overbrace{1}^{\text{1 step}} + \overbrace{dt'}^{\text{random time}} \qquad dt' = \min\{k \geq 0 \,:\, X_k = 3 | X_0 = j, \forall\, j \in X\}$$

Clearly, if $j = 3$ we have $\mu_3 = \mathrm{E}[T|X_0 = 3] = 0$

While, from the others states it results:



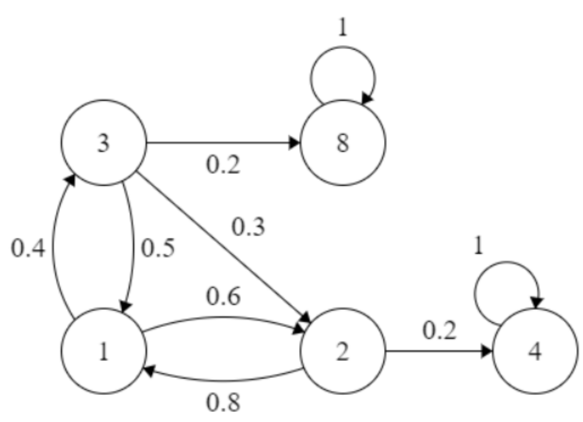
$$\mu_1 = \overbrace{1}^{\text{1step}} + \frac{1}{2}\mu_1 + \frac{1}{3}\mu_2 + \frac{1}{6}\mu_3$$

$$\mu_2 = \overbrace{1}^{\text{1step}} + \frac{2}{3}\mu_2 + \frac{1}{3}\mu_3 = 1 + \frac{2}{3}\mu_2$$

$\Rightarrow$

$$\mu_3 = 0$$
$$\mu_2 = 3$$
$$\mu_1 = 4$$

- Example C: What if the chain have 2 absorbing states?



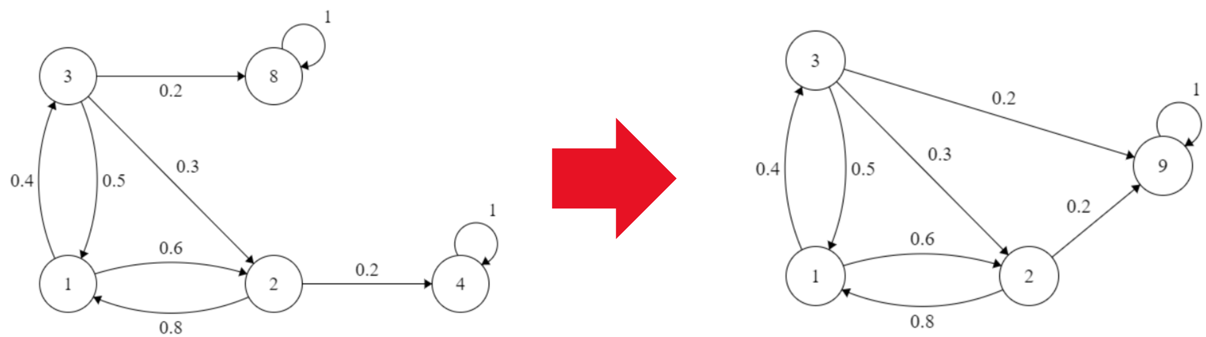
- Which is the mean time $\mu_i = E[T|X_0 = i]$?
- Clearly $\mu_4 = 0$, whereas $\mu_8 = E[T|X_0 = 8] = \infty$

  $\Rightarrow \mu_1 = \mu_2 = \mu_3 = \infty$ Although correct the result is useless ☹

- These strategies works only for ergodic DT-MCs

Mean time to absorption (easy way)

- Example C (cont'd): If we rearrenge the chain in a such a way there is only one absorbing state, everything makes sense



$9 \equiv \{4\} \cup \{8\}$

If the DT-MC has multiple ergodic SCCs, they can be reduced into a unique absorbing state ($p_{kk} = 1$)

Mean Hitting times's Theorem: Let $A \subset X$ be a subset of states of a finite DT-MC. Let T be the r.v. of first hitting time the MC visits a state in A.

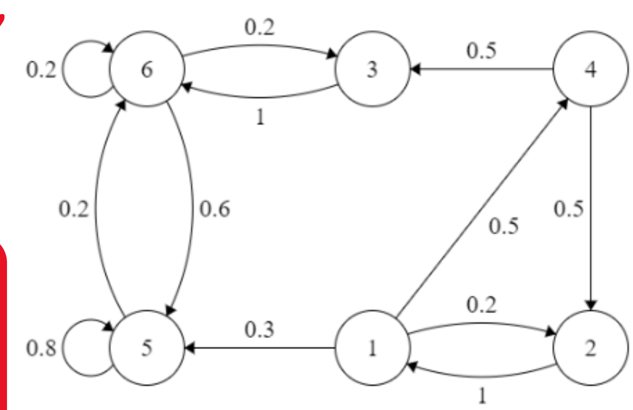
Then, for all $i \in X$, we define $\mu_i = \mathrm{E}[T|x(0) = i]$

By construction, it results that $\mu_j = 0 \;\; \forall\, j \in A$.

whereas the other μ_j are the solution of the linear system of equations:

$$\mu_i = 1 + \sum_j p_{ij} \cdot \mu_j, \quad \forall \quad i \in X \setminus A$$

Mean first passage "t_i" and Recurrence time "t_s^*"

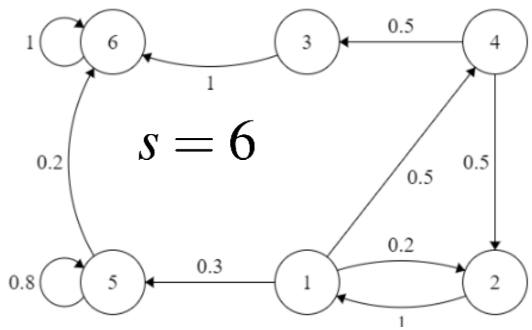


Consider a DT-MC with 1 ergodic class, e.g., $E_1 = \{3,5,6\}$

It is of interest the **mean time of the first passage** from all i to $s \in E_1$

Definition: $t_i = \mathrm{E}[\min\{k \geq 0 \; : \; X_k = s \mid X_0 = i\}]$

Example: Let $s = 6$, since we don't care where the MC goes after reaching s, then estimate «t_i» is like compute the mean time of absorbtion of the **rearrenged** MC where s is **now** an ***absorbing state***



$$t_s = t_6 = 0$$

$$t_i = 1 + \sum_{\forall\; j} p'_{ij} \cdot t_j, \quad \forall \quad i \neq s$$

$$t_1 = 6.36$$
$$t_2 = 7.36$$
$$t_3 = 1$$
$$t_4 = 5.18$$
$$t_5 = 5$$

Then, it is also of interest the **mean recurrence time** of a state $s \in E_1$

Definition: $t_s^* = \mathrm{E}[\min\{k \geq 1 \; : \; X_k = s \mid X_0 = s\}]$

Interpretation: It account the event to be again in s the 1 step later, **OR** leave it and then return to s. Thus $t_s^* \geq 1$, cf. $t_6^* \geq 1$

Example (cont'd): Let $s = 6$, t_6^* can be found as next

$$t_s^* = 1 + \sum_{\forall\; j} p_{sj} \cdot t_j$$

$$t_6^* = 1 + 0.2 \cdot \overbrace{t_6}^{0} + 0.2 \cdot \overbrace{t_3}^{1} + 0.6 \cdot \overbrace{t_5}^{5} = 4.2$$

Discrete-time Birth-Death Processes

- Definition: A discrete-time Birth-Death Process (DT-BDP) is a special class of DT-MC which are characterized by:

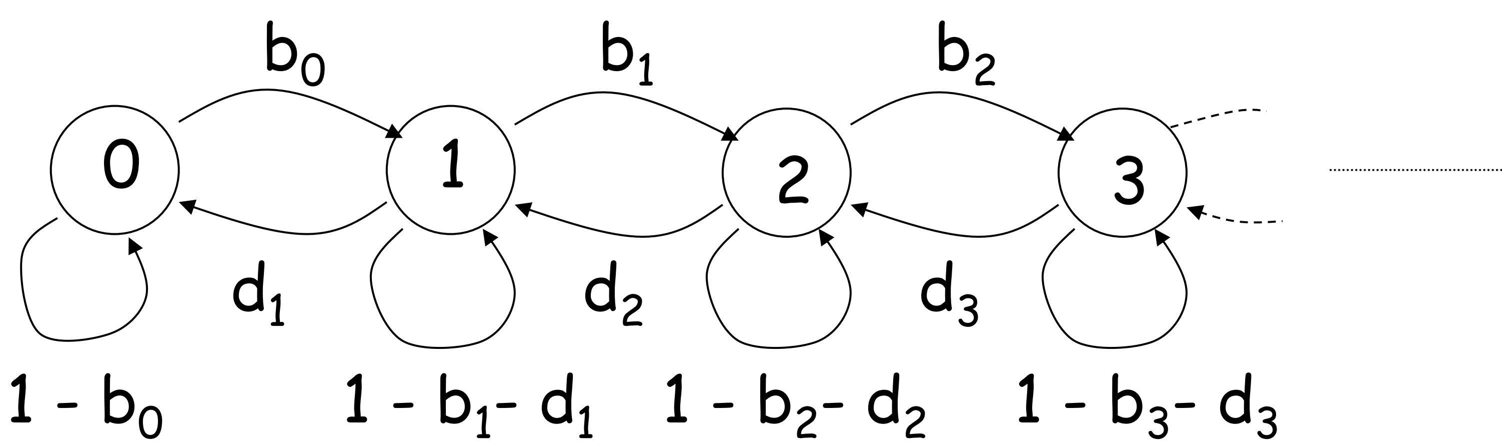
a) A sample space X which can only take non-negative (possibly infinite) integer values

$$X = \{ 0 , 1 , 2 , \ldots \}$$

b) Only transitions between adjacent states are allowed.

c) The terms b_i and d_i denotes the probability of the i-th birth/death event

- The transition probability graph of a DT-BDP process is



Discrete-time Birth-Death Processes (cont'd)

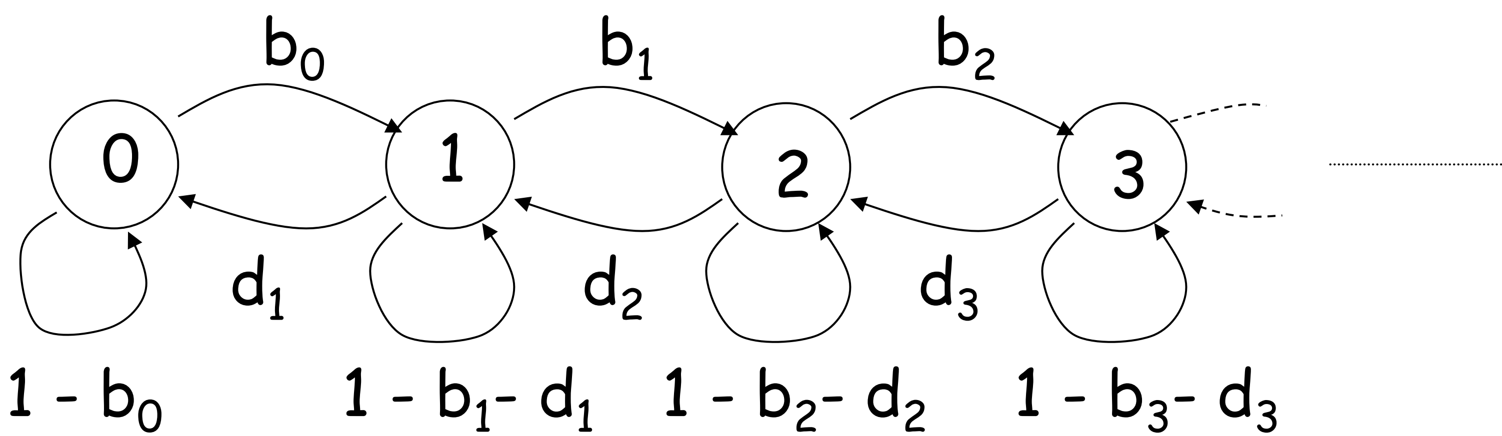
- The sample space X denotes the population of the modeled application.

  1) #messages in a TCL system
  2) #costumers in a queue
  3) #vehicles in a transportation system
  4) #people for analyzing demography/epidemiology applications
  5) Etc…

- They find applications in many fields:

  1) Performance Engineering & Management (e.g. see *resource allocation problems*)
  2) demography control
  3) queueing theory
  4) epidemiology
  5) biology and other areas

Discrete-time Birth-Death Processes (cont'd)



- The transition probabilities matrix takes the following **tri-diagonal structure**

$$
\boldsymbol{P} = \begin{bmatrix} 1-b_0 & b_0 & 0 & 0 & \cdots \\ d_1 & 1-b_1-d_1 & b_1 & 0 & \cdots \\ 0 & d_2 & 1-b_2-d_2 & b_2 & \\ 0 & 0 & d_3 & 1-b_3-d_3 & \\ & 0 & 0 & d_4 & \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}
$$

- If the DT-BDP has infinite states, then $\boldsymbol{P}$ has clearly an infinite dimension.

Discrete-time Birth-Death Processes (cont'd)

- If b_i and d_i are time-invariant $\forall\ k, i$:

The DT-BDP is called time-homogenous → $\boldsymbol{P}$ is constant

- If $b_i(k) > 0$ and $d_i(k) > 0\ \forall\ k, i$

The DT-BDP is irriducible (all states are mutually reachable)

- If $b_i = b \geq 0$ and $d_i = d \geq 0\ \forall\ i$

The DT-BDP is called uniform.

- If additionally $b \in (0,1)$ OR $d \in (0,1)$

The DT-BDP is not only irreducible, uniform
but it is also aperiodic.
Indeed, it exists at least a self-loop in a state

Remark: Unfortunately this is not enough to say the BDP is ergodic… ☹

Stationary distribution of a DT-BDP

- Consider an irreducible non-uniform but aperiodic DT-BDP with infinite many states

$$\boldsymbol{P} = \begin{bmatrix} 1-b_0 & b_0 & 0 & \dots & \dots \\ d_1 & (1-b_1-d_1) & d_1 & 0 & \\ 0 & d_2 & (1-b_2-d_2) & b_2 & \dots \\ \vdots & 0 & d_3 & & \\ \vdots & \vdots & 0 & & \end{bmatrix}$$

- Then, by solving its balance equation one has

$$\begin{cases} \Pi_s = \Pi_s \cdot \boldsymbol{P} \\ \sum_i \pi_{i,s} = 1 \end{cases}$$

$$\begin{cases} \pi_{s,0} & = & (1-b_0)\pi_{s,0} + d_1\pi_{s,1} \\ \pi_{s,1} & = & b_0\pi_{s,0} + (1-b_1-d_1)\pi_{s,1} + d_2\pi_{s,2} \\ \vdots & = & \vdots \\ \sum_i \pi_{s,i} & = & 1 \end{cases}$$

$\Rightarrow$

$$\begin{cases} b_0\pi_{s,0} & = & d_1\pi_{s,1} \\ 0 & = & d_1\pi_{s,1} + (-b_1-d_1)\pi_{s,1} + d_2\pi_{s,2} \\ \vdots & = & \vdots \\ \sum_i \pi_{s,i} & = & 1 \end{cases}$$

(cf. This is a Reversible DT-MC)

Stationary distribution of DT-BDP

- Then, if $b_i = b > 0$ and $d_i = d > 0$, and by letting $\rho = \dfrac{b}{d}$

$$\begin{cases} \pi_{s,1} &= \frac{b_0}{d_1}\cdot\pi_{s,0} \\ \pi_{s,2} &= \frac{b_1}{d_2}\cdot\pi_{s,1} \\ \vdots &= \vdots \\ \pi_{s,i} &= \frac{b_{i-1}}{d_i}\cdot\pi_{s,i-1} \\ \vdots &= \vdots \\ \sum_i \pi_{s,i} &= 1 \end{cases} \Rightarrow \begin{cases} \pi_{s,1} &= \rho\cdot\pi_{s,0} \\ \pi_{s,2} &= \rho^2\cdot\pi_{s,0} \\ \vdots &= \vdots \\ \pi_{s,i} &= \rho^i\cdot\pi_{s,0} \\ \vdots &= \vdots \\ \sum_i \pi_{s,i} &= 1 \end{cases}$$

$$\Rightarrow \sum_{i=0}^{\infty}\pi_{s,i} = \pi_{s,0}+\rho\pi_{s,0}+\rho^2\pi_{s,0}+\cdots = 1 \quad\Longrightarrow\quad \pi_{s,0}\sum_{i=0}^{\infty}\rho^i = 1 \qquad (1)$$

Remark: From that it results a DT-BDP is always a **reversible DT-MC** because it satisfies $\pi_i p_{ij} = \pi_j p_{ji}$, where $\forall\ i < j$ it yields $p_{ij} = b_i$ and $p_{ji} = d_j$

Very important: This is not enought to say it is neither stationary nor ergodic because (1) may not be verified due to the given ρ

Stationary distribution of DT-BDP (cont'd)

$$\rho = \frac{b}{d} < 1$$

If this geometric series is convergent then Π_s exists unique

$$\sum_{i=0}^{k-1} \rho^i = \frac{(1-\rho^k)}{(1-\rho)}\Big|_{k\to\infty} = \frac{1}{1-\rho}$$

Proof : $(1-\rho)(1+\rho+\rho^2) = (1+\rho+\rho^2) - (\rho+\rho^2+\rho^3) = (1-\rho^3)$

$$\begin{matrix} \pi_{s,i} = \rho^i \pi_{s,0} \\ \pi_{s,0} \sum_{i=0}^{\infty} \rho^i = 1 \end{matrix} \quad \Longrightarrow \quad \pi_{s,0} \cdot \frac{1}{1-\rho} = 1 \quad \Longrightarrow \quad \pi_{s,0} = 1-\rho$$

- The irreducibility and aperiodicity of the DT-BDP along with the above constraint $\rho < 1$ implies the ergodicity of the DT-DBP. Thus one has

$$\Pi_\ell \equiv \Pi_s = \begin{cases} \pi_{s,0} & = & 1-\rho \\ \pi_{s,i} & = & \rho^i \cdot \pi_{s,0} = \rho^i \cdot (1-\rho), \qquad \forall\, i > 0 \end{cases}$$

- Remark: If instead the number of states is finite, the process ergodicity can be determined by means of the graphical criteria regardless of the value of ρ

Metrics of interest of DT-BDPs

- It is of interest evaluate the long term averaged population of the DT-BDP (or average number of costumers at steady state

$$\mu_X(\infty) = \lim_{k\to\infty} \mathsf{E}[X] = \sum_{i\geq 0} i\cdot\pi_i(\infty) = \sum_{i\geq 0} i\cdot\pi_{\ell,i}$$

- If the process is ergodic then

$$\pi_{\ell,i} = \pi_{s,i} = \rho^i\cdot(1-\rho) \quad \forall\, i\in X$$

- Since the **population is non-negative** it can be evaluated by means of the moment generating function at the steady state, denoted here by M_ℓ

$$\mu_X(\infty) \quad = \quad -\left.\frac{dM_\ell(z)}{dz}\right|_{z=1}$$

- where

$$M_\ell(z) \quad = \quad \sum_{i=0}^{\infty}\pi_{\ell,i}\cdot z^{-i} = \sum_{i=0}^{\infty}\left(\rho^i\cdot(1-\rho)\right)\cdot z^{-i}$$

$$= \quad (1-\rho)\cdot\sum_{i=0}^{\infty}\left(\frac{\rho}{z}\right)^i = (1-\rho)\cdot\frac{1}{1-\frac{\rho}{z}} = \frac{(1-\rho)\cdot z}{z-\rho}$$

Metrics of interest of DT-BDPs

- Thus the mean #costumer in the DT-BDP is

$$\mu_X(\infty) = -\frac{dM_\ell(z)}{dz}\bigg|_{z=1} = -\frac{d}{dz}\left(\frac{(1-\rho)\cdot z}{z-\rho}\right)\bigg|_{z=1} = -\frac{\rho(1-\rho)}{(z-\rho)^2}\bigg|_{z=1} = \frac{\rho}{1-\rho}$$

- Further note

$$\mathsf{Var}[X_\infty] = \frac{d^2M_\ell(z)}{dz^2}\bigg|_{z=1} - \mu_X(\infty) - (\mu_X(\infty))^2 = \frac{2\rho}{(1-\rho)^2}$$

Remark: If $\rho \geq 1$ the birth probability is larger then the death one. Thus, in the long terms more births than deaths will occur, and thus the process size will diverge

$\mu_X(\infty) \rightarrow \infty$ $\Rightarrow$ the DT-BDP is **not stationary** $\Rightarrow$ **not ergodic**

Remark: For a DT-BDP all the techniques for DT-MC can be also applied by properly reducing/rearranging the DT-BDP transition probability graph

$$\lim_{k\to\infty} \Pr(x(k) \geq s) = 1 - \sum_{i=0}^{s-1} \pi_i(\infty) = 1 - \frac{(1-\rho^s)}{(1-\rho)}$$ (steady state complementary cdf)

$$t_i = \mathsf{E}[\min\{k \geq 0 \,:\, X_k = s \mid X_0 = i\}]$$ (mean first passage)

$$t_s^* = \mathsf{E}[\min\{k \geq 1 \,:\, X_k = s \mid X_0 = s\}]$$ (recurrence time)