Probability and statistics, 2022, Computer Science Algorithmics, Undergraduate Course, Part II, lecturer: Mirosław Kutyłowski

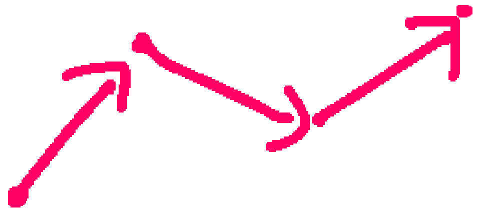
3- Stochastic Processes

Stochastic process

- Time dependent random variables : time+space
  - time: $1, 2, 3, 4, \ldots$
    $t \in (0, +\infty)$
  - space: Ω
  - state: $X(t,\omega)$ where $t \in$ Time, $\omega \in \Omega$

Examples:

- **Trajectory of a particle**



- **Noise**
- **Rain**
- **Messages in a communication bus**

Examples:

- **CPU usage**

- **microcontrollers power consumption**

1) **Discrete time process**

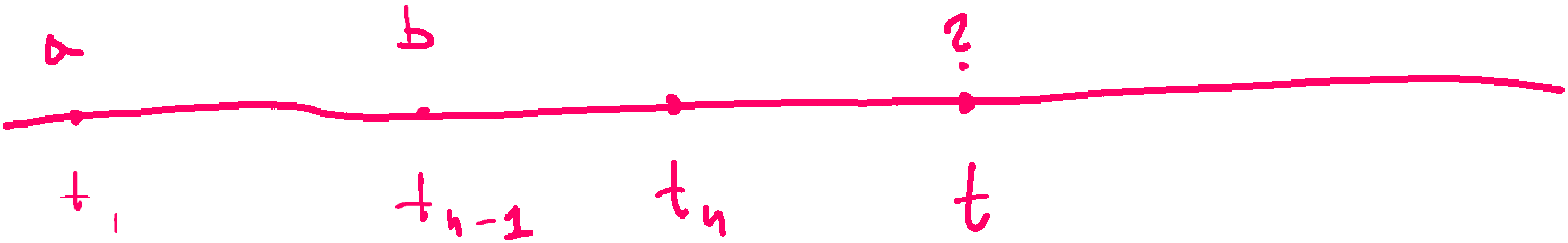
2) **Continuous time process**

Markov process

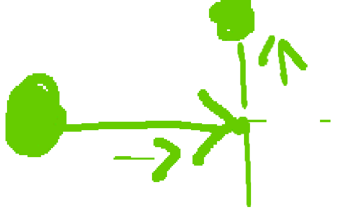
only the most recent state counts

Stochastic process $X(t)$ is Markov if for any $t_1 < \ldots < t_n < t$ and any sets $A; A_1, \ldots, A_n$

$$P\{X(t) \in A \mid X(t_1) \in A_1, \ldots, X(t_n) \in A_n\} = P\{X(t) \in A \mid X(t_n) \in A_n\}. \tag{6.1}$$



Markov chain



- **discrete Markov process**
- **the state at time t+1 depends only on the state at time t**

$$p_{ij}(t) = P\{X(t+1) = j \mid X(t) = i\}$$
$$= \quad P\{X(t+1) = j \mid X(t) = i,\ X(t-1) = h,\ X(t-2) = g, \ldots\}$$

Transition probability:

$$p_{ij}^{(\;)}(t) = P\{X(t+1) = j \mid X(t) = i\}$$

Homogenous Markov chain

- **Transition pbb does not depend on the time**

$p_{ij}(t)$ is constant, notation: p_{ij}

p_{ij} = pbb przejścia ze stanu i do j

- **Transition matrix**

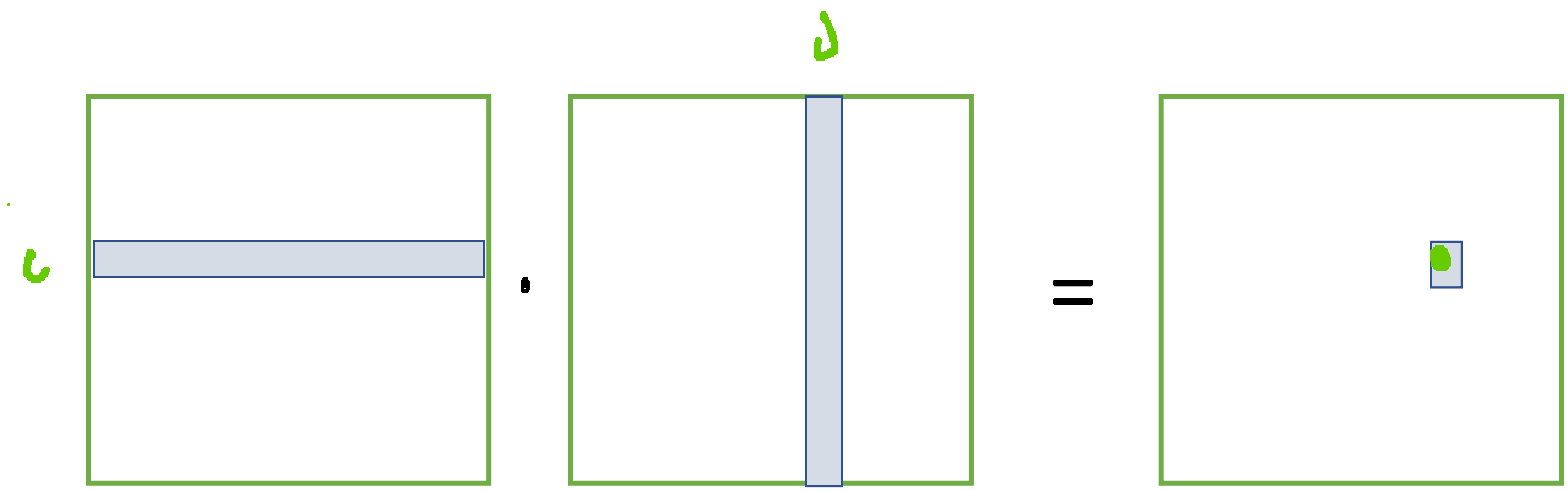
$$\begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix}$$

Transition in 2 steps

$i \to k \to j$

$$\begin{aligned} p_{ij}^{(2)} &= P\{X(2)=j \mid X(0)=i\} \\ &= \sum_{k=1}^{n} P\{X(1)=k \mid X(0)=i\} \cdot P\{X(2)=j \mid X(1)=k\} \\ &= \sum_{k=1}^{n} p_{ik}p_{kj} = (p_{i1},\ldots,p_{in}) \begin{pmatrix} p_{1j} \\ \vdots \\ p_{nj} \end{pmatrix}. \end{aligned}$$

Transition pbb in two steps



macierz przejścia T

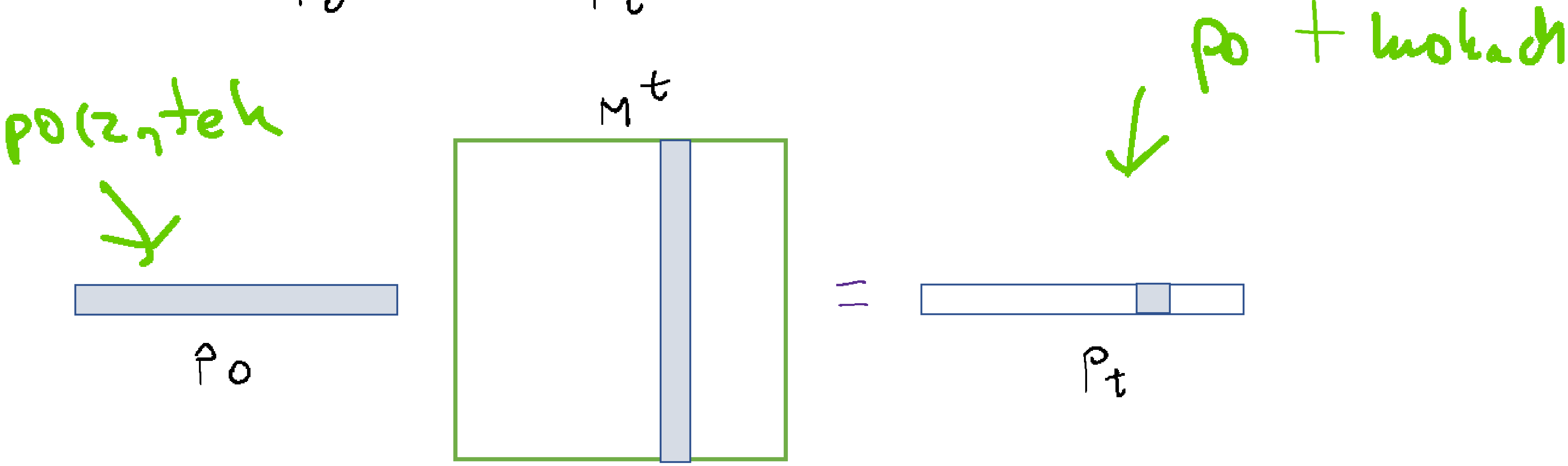
dla 2 kroków T^2

n kroków T^n

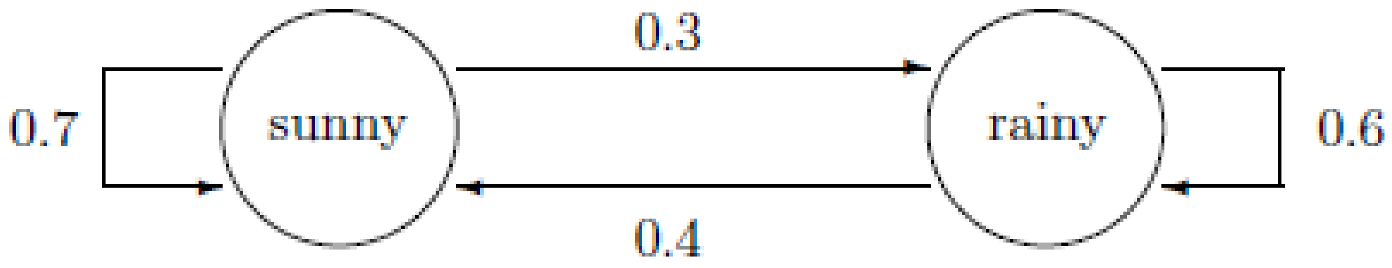
Probabilities at time t

- **Transition matrix M of a homogenous chain**

$$p_0 \cdot M^t = p_t$$



Description via a Transition diagram

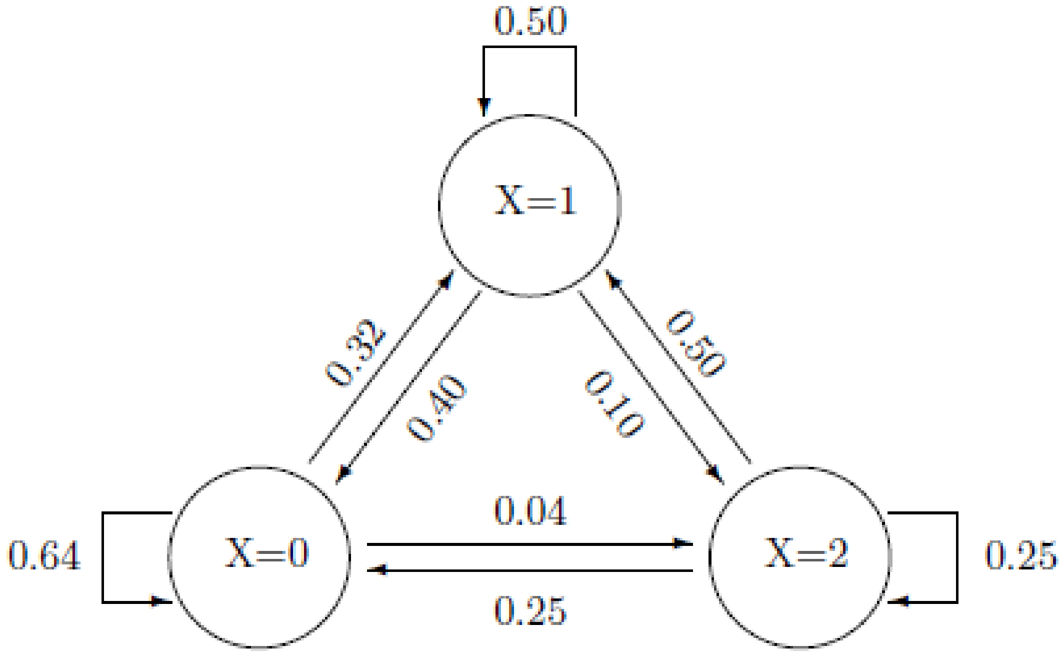


Transition diagram

2 users: active user disconnects with pbb 0.5

inactive user connects with ppb 0.2

X= number of active users



Steady state distribution

„eventually it does not depend on the initial state"

A collection of limiting probabilities

$$\pi_x = \lim_{h \to \infty} P_h(x)$$

is called a steady-state distribution of a Markov chain $X(t)$.

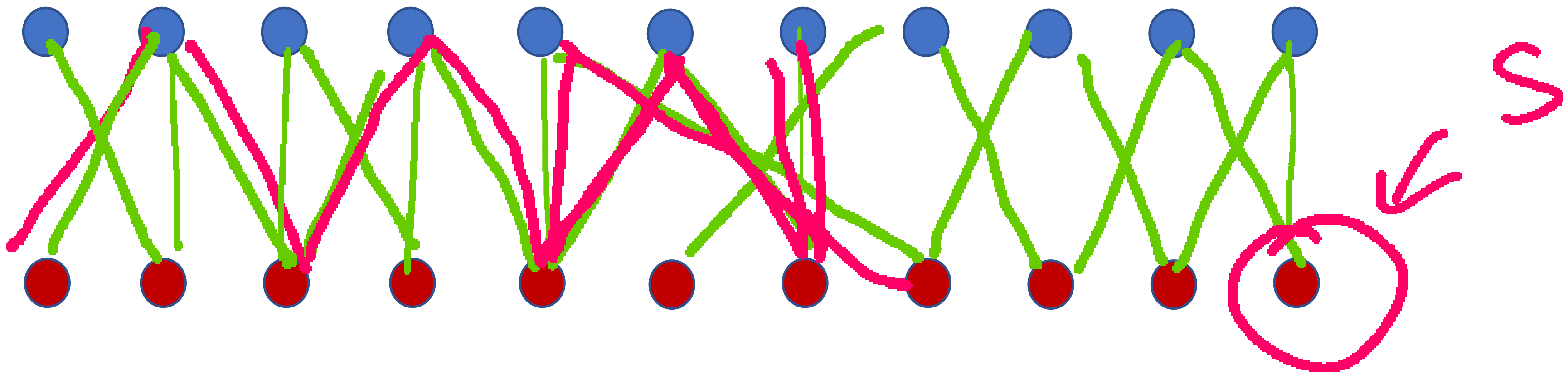
It is not clear in advance that a steady-state distribution exists

Another name used: *stationary distribution*

Example: no steady state distribution

proces periodyczny

random walk in a bipartite graph



parzyste kroki: na dole

nieparzyste na górze

$P_{17}(s)=0$

$P_{18}(s) \geq ?$

$P_{19}(s)=0$

Computing steady state distribution

$$P_h P = P_0 P^h P = P_0 P^{h+1} = P_{h+1}.$$

$$\pi P = \pi.$$

$$(\pi_0, \pi_1)\begin{pmatrix} a & \cdot \\ b & \cdot \end{pmatrix} = (\pi_0, \pi_1)$$

this is a system of linear equations. Moreover:

$$\begin{cases} \pi_0 \cdot a + \pi_1 b = \pi_0 \\ \end{cases}$$

$$\sum \pi_i = 1$$

(the probabilities must sum up to 1)

Weather example cnt

$$(\pi_1,\ \pi_2) = (\pi_1,\ \pi_2)\begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix} = (0.7\pi_1 + 0.4\pi_2,\ 0.3\pi_1 + 0.6\pi_2)\,.$$

$$\begin{cases} 0.7\pi_1 + 0.4\pi_2 &= \pi_1 \\ 0.3\pi_1 + 0.6\pi_2 &= \pi_2 \end{cases} \Leftrightarrow \begin{cases} 0.4\pi_2 &= 0.3\pi_1 \\ 0.3\pi_1 &= 0.4\pi_2 \end{cases} \Leftrightarrow \pi_2 = \frac{3}{4}\pi_1.$$

$$\pi_1 + \pi_2 = \pi_1 + \frac{3}{4}\pi_1 = \frac{7}{4}\pi_1 = 1,$$

$$\underline{\pi_1 = 4/7} \quad \text{and} \quad \underline{\pi_2 = 3/7}.$$

15.12.2022

Existence of stationary distribution

A Markov chain is **regular** if

$$p_{ij}^{(h)} > 0$$

for some h and all i, j. That is, for some h, matrix $P^{(h)}$ has only non-zero entries, and h-step transitions from any state to any state are possible.

Any regular Markov chain has a steady-state distribution.

Example: random walk on a bipartite graph --- this property does not hold!

h is even:
H is odd:

Breaking periodicity

The trick is to modify the transition function T:

- **with pbb 0.5 keep the old state**
- **with pbb 0.5 change state according to T**

Algorithms based on Markov chains

Example: choose a maximal independent set in a graph at random (with uniform probability)

Difficult:

we can create a maximal independent set (e.g. via a greedy algorithm) , enumerating all MIS might be extremely hard

Approach: via a Markov chain

- states = independent sets
- transitions: simple modifications (removing or adding nodes)
- … so that the steady distribution is uniform

Uniform steady distribution

ds

The **case of double stochastic matrix:**

sum of each row is 1 (must be)

sum of each column is 1 (not for all transition matrices)

Theorem:

stationary distribution is uniform for a double stochastic transition matrix

Proof: to check that:

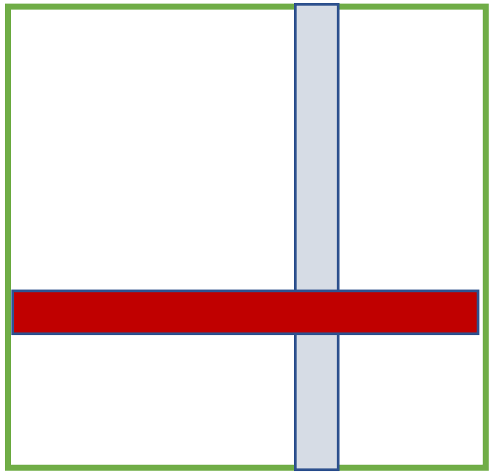
$$\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right) \cdot P \overset{?}{=} \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$$

Checking one column:

$$\frac{1}{n} \cdot p_{1j} + \frac{1}{n} \cdot p_{2j} + \dots + \frac{1}{n} \cdot p_{nj} = \frac{1}{n}\left(p_{1j} + \dots + p_{nj}\right) = \frac{1}{n} \cdot 1 = \frac{1}{n}$$

Uniform steady distribution – special case:
Symmetric transition matrix

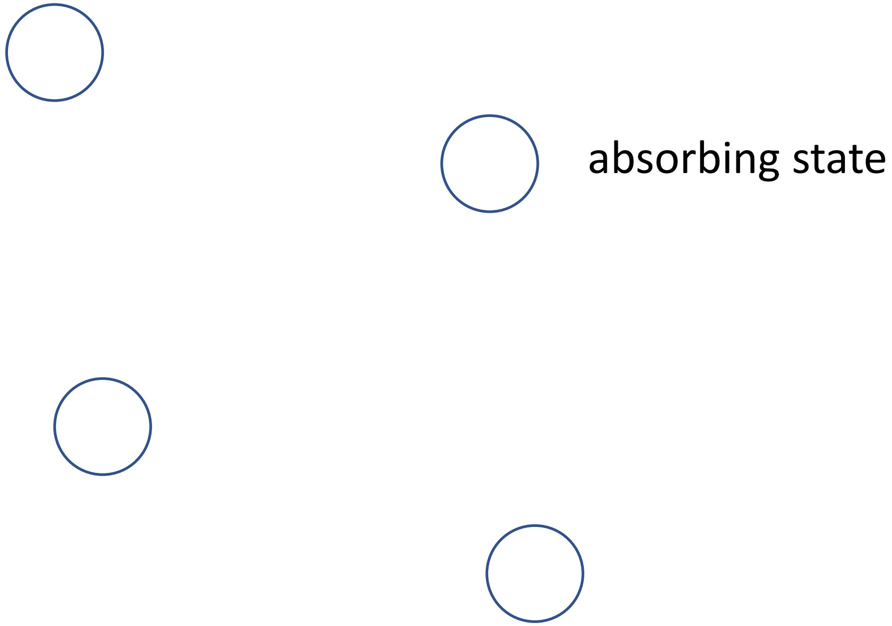
$$p_{ij} = p_{ji}$$



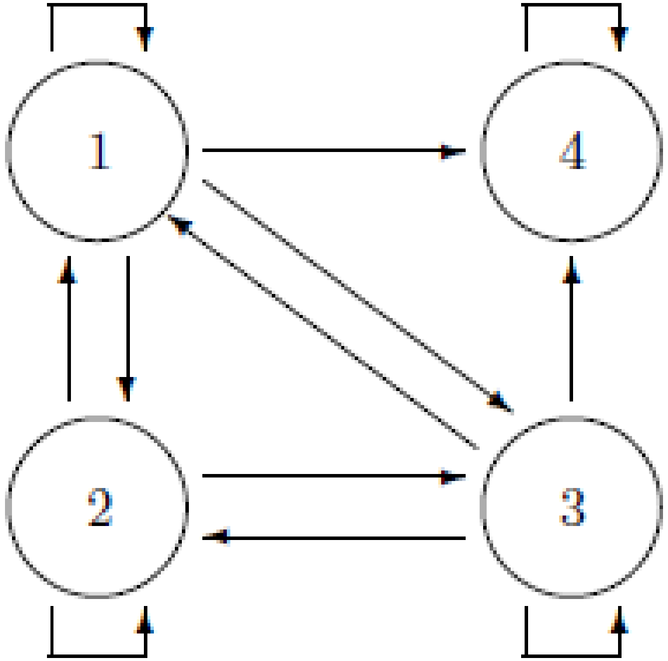
Obviously: the matrix is double stochastic
→ the steady distribution is uniform!

Absorbing states

No exit from an absorbing state



Absorbing state example



Algorithms based on absorbing state

Random walk based on a Markov chain

Run simulation, eventually you are trapped in an absorbing state = a good state, where there is nothing to improve

Counting processes

e.g. Bernoulli trials
State in time t = number of succeses in steps 1 through t

Expected number: $p \cdot t$

Counting processes

state is a counter
counter is nondecreasing

Examples:

- **the number of incoming cars on a bridge**
- **the number of emails arrived**

Binomial proces

independent Bernoulli trials

counter= number of successes
time frame Δ: one Bernoulli trial per Δ seconds

Expected number of successes:

$$\mathbf{E}\left\{X\left(\frac{t}{\Delta}\right)\right\} = \frac{t}{\Delta}p$$

Expected number of successes per second (arrival rate):

$$\lambda = \frac{p}{\Delta}$$

Bernoulli counting process - interarrival time

$$T = Y\Delta$$

$$\lambda = \frac{p}{\Delta}$$

where Y has geometric distribution

$$\begin{aligned} \mathrm{E}(T) &= \mathrm{E}(Y)\Delta = \frac{1}{p}\Delta = \frac{1}{\lambda}; \\ \mathrm{Var}(T) &= \mathrm{Var}(Y)\Delta^2 = (1-p)\left(\frac{\Delta}{p}\right)^2 \text{ or } \frac{1-p}{\lambda^2}. \end{aligned}$$

Continuous counting process

a limit of Bernoulli counting process with time frame Δ → 0

The number of frames during time t increases to infinity,

$$n = \frac{t}{\Delta} \uparrow \infty \text{ as } \Delta \downarrow 0.$$

The probability of an arrival during each frame is proportional to Δ, so it also decreases to 0,

$$p = \lambda\Delta \downarrow 0 \text{ as } \Delta \downarrow 0.$$

Continuous counting process:

Poisson as a limit of Binomial

Then, *the number of arrivals* during time t is a Binomial(n, p) variable with expectation

$$\mathbf{E}\, X(t) = np = \frac{tp}{\Delta} = \lambda t.$$

$$X(t) = \text{Binomial}(n, p) \to \text{Poisson}(\lambda)$$

Based on the following theorem:

$$\lim_{\substack{n \to \infty \\ p \to 0 \\ np \to \lambda}} \binom{n}{x} p^x (1-p)^{n-x} = e^{-\lambda} \frac{\lambda^x}{x!}$$

Continuous counting process

The *interarrival time* T becomes a random variable with the c.d.f.

$$\begin{aligned} F_T(t) &= P\{T \le t\} = P\{Y \le n\} && \text{because } T = Y\Delta \text{ and } t = n\Delta \\ &= 1 - (1-p)^n && \text{Geometric distribution of } Y \\ &= 1 - \left(1 - \frac{\lambda t}{n}\right)^n && \text{because } p = \lambda\Delta = \lambda t/n \\ &\to 1 - e^{-\lambda t}. \end{aligned}$$

This is the "Euler limit":
$(1 + x/n)^n \to e^x$ as $n \to \infty$

The interarrival time converges to exponential distribution

Continuous counting process

let us inspect the time for k arrivals

$$\boldsymbol{P}\{T_k \le t\} = \boldsymbol{P}\{\ k\text{-th arrival before time } t\ \} = \boldsymbol{P}\{X(t) \ge k\}$$

where T_k is Gamma(k, λ) and $X(t)$ is Poisson(λt).

Applications

What is the probability that in time T more than k requests arrive for a webpage P?

We assume that λ is known (λ requests per minute)

Continuous counting process:

Poisson as a limit of Binomial

However: if we take a time interval that is not very small with respect to λ, then some differences occur:

- ❑ **Binomial: at most one arrival in the interval**

- ❑ **Poisson: more than one possible**

… even if the arrival rate is the same λ, there is a greater variance for Poisson

Solution for λ=7 hits per minute , assumed Poisson process

Pbb for 10000 hits within 24 hours?

Solution. The time of the 10,000-th hit T_k has Gamma distribution with parameters $k = 10,000$ and $\lambda = 7\ \text{min}^{-1}$. Then, the expected time of the k-th hit is

$$\mu = \mathrm{E}(T_k) = \frac{k}{\lambda} = \underline{1{,}428.6 \text{ min or } 23.81 \text{ hrs.}}$$

$$\sigma = \mathrm{Std}(T_k) = \frac{\sqrt{k}}{\lambda} = 14.3 \text{ min.}$$

Pbb of more than 10000 hits?

A shortcut: CLT (we do not have to care about Gamma distribution!)

$$P\{T_k < 1440\} = P\left\{\frac{T_k - \mu}{\sigma} < \frac{1440 - 1428.6}{14.3}\right\} = P\{Z < 0.80\} = \underline{0.7881}.$$

Conclusion

Easy way to solve many problems regarding required capacity …