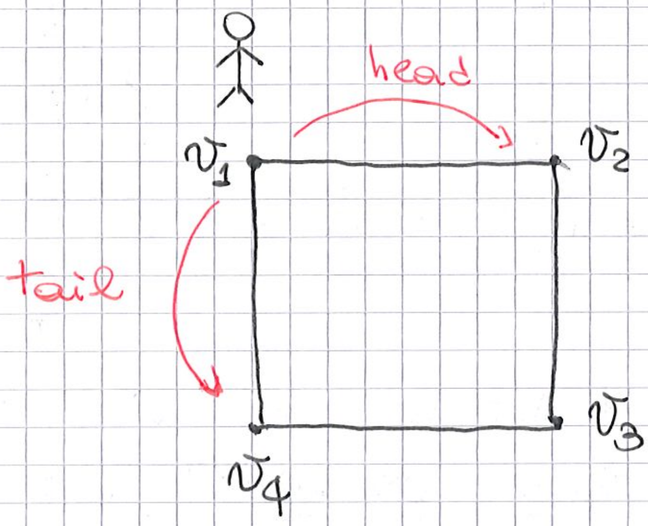
MARKOV CHAINS (O. Häggström, Finite Markov Chains and Alg.)

Example

Consider a "random walker" in a very small town consisting of 4 streets.



v_i = corners among streets

At time 0 the random walker stands in v_1.
At time 1 he tosses a fair coin and moves one step:

clockwise if comes up head
counterclockwise " " " tail

The procedure is iterated.

X_n := index of the corner at which the walker stands at time n

$(X_0, X_1, X_2, \dots)$ is a random process taking values in $\{1, 2, 3, 4\}$.

the walker starts from $v_1 \Rightarrow$

$$\mathbb{P}(X_0 = 1) = 1$$

then he moves in the 2 directions with equal probability $\Rightarrow$

$$\left.\begin{aligned} &\mathbb{P}(X_1 = 2) = \tfrac{1}{2} \\ &\mathbb{P}(X_1 = 4) = \tfrac{1}{2} \\ \text{and} \quad &\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = 3) = 0 \end{aligned}\right\} \text{distribution at time } 1$$

Let's compute the distribution at time $n+1$:
let's consider <u>conditional probabilities</u>:
Suppose that at time n the walker stands at, say, v_2. Then we have the conditional probabilities

$$(*) \quad \begin{cases} \mathbb{P}(X_{n+1} = v_1 \mid X_n = v_2) = \frac{1}{2} \\ \mathbb{P}(X_{n+1} = v_3 \mid X_n = v_2) = \frac{1}{2} \end{cases}$$

and $\mathbb{P}(\text{other vertices}) = 0$

We obtain again $(*)$ also if we condition on the full history of the process up to n:

$$\mathbb{P}(X_{n+1} = v_1 \mid X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = v_2) = \frac{1}{2}$$

$$\mathbb{P}(X_{n+1} = v_3 \mid X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = v_2) = \frac{1}{2}$$

for any choice of $i_0, \dots, i_{n-1}$ (because of independence of coin flips!)

This is the memoryless property or Markov Property:

The conditional distribution of X_{n+1} given $(X_0, \dots, X_n)$ depends only on X_n.

In other words to predict at best what will happen "tomorrow" we just need to know what happened "today", and we don't need to know the past.

This property is shared by many real processes, but not all $\Rightarrow$ it is a modelling assumption

other property of our process: if at any time t we know that $X_t = v_2$ (say), the distribution of X_{t+1} is always the same. This is due to the fact that the choice at each step is done with the same procedure. This property is called stationarity or time homogeneity.

Def: Let P be a $k \times k$ matrix with elements $\{P_{ij};\ i=1,\dots,k\ \ j=1,\dots,k\}$. A random process $(X_0, X_1, \dots)$ with finite state space $S = \{s_1, \dots, s_k\}$ is said to be a (homogeneous) Markov chain with transition matrix P, if for all n, all $i,j \in \{1 \dots k\}$ and all $i_0, \dots, i_{n-1} \in \{1, \dots, k\}$ we have

$$\mathbb{P}(X_{n+1} = s_j \mid X_0 = s_{i_0}, X_1 = s_{i_1}, \dots, X_{n-1} = s_{i_{n-1}}, X_n = s_i) =$$

$$= \mathbb{P}(X_{n+1} = s_j \mid X_n = s_i)$$

$$= P_{ij}$$

The elements of P are called transition probabilities:

$$P_{ij} = \mathbb{P}(\text{"tomorrow" } X \text{ is in state } s_j \mid \text{"today" } X \text{ is in state } s_i)$$

$$= \mathbb{P}(\text{passing from state } s_i \text{ to } s_j \text{ in one step})$$

In our example of the random walker:

$S = \{v_1, v_2, v_3, v_4\}$

$$\begin{array}{cc} & \begin{array}{cccc} v_1 & v_2 & v_3 & v_4 \end{array} \leftarrow \text{going to} \\ P = \begin{array}{c} v_1 \\ v_2 \\ v_3 \\ v_4 \\ \uparrow \\ \text{starting from} \end{array} & \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix} \end{array}$$

- Since P_{ij} are probabilities, $P_{ij} \in [0, 1]$ $\forall i, j$
- Each row is a distribution, thus

$$\sum_{j=1}^{K} P_{ij} = 1 \quad \text{(rows sum to 1)}$$

Another important characteristics of a Markov Chain is the

initial distribution = how the Markov Chain starts: it is a probability distribution

$$\underline{\mu}^{(0)} = (\mu_1^{(0)}, \dots, \mu_k^{(0)})$$
$$= (\mathbb{P}(X_0 = s_1), \dots, \mathbb{P}(X_0 = s_k))$$

with $\sum_{i=1}^{k} \mu_i^{(0)} = 1$

In our example: $\underline{\mu}^{(0)} = (1, 0, 0, 0)$

We call $\underline{\mu}^{(n)} = (\mu_1^{(n)}, \mu_2^{(n)}, \dots, \mu_k^{(n)})$

$$= (\mathbb{P}(X_n = s_1), \mathbb{P}(X_n = s_2), \dots, \mathbb{P}(X_n = s_k))$$

the distribution of the MC at time n.

In our example:

$$\underline{\mu}^{(1)} = (0, \tfrac{1}{2}, 0, \tfrac{1}{2})$$

Since our MC is homogeneous, we can compute $\underline{\mu}^{(n)}$ from $\underline{\mu}^{(0)}$ and P:

Theorem: for a homogeneous Markov Chain $(X_0, X_1, \dots)$ with state space $\{s_1, \dots, s_k\}$, initial distribution $\underline{\mu}^{(0)}$ and transition matrix P, we have that the distribution $\underline{\mu}^{(n)}$ at time n satisfies

$$\boxed{\mu^{(n)} = \underline{\mu}^{(0)} P^n}$$

Example 2 - The Gothenburg weather

Let's assume that we use the following method to predict the weather of tomorrow: we guess that it will be the same of today.
Assume that we are in Gothenburg (Sweden) and we observe that we are right 75% of times. We use only 2 states to describe the weather: S_1 = sunshine
S_2 = rain
Then the weather is a MC with transition matrix:

$$P = \begin{bmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{bmatrix} \qquad (*)$$

Example 3 : the Los Angeles weather

If we are in Los Angeles, where sunshine is more frequent than rain, we probably can not assume the perfect symmetry of matrix P in (*) between the two states. A more realistic transition matrix could be:
(S_1 = sunshine, S_2 = rain)

$$P = \begin{array}{c} \\ \text{sun} \\ \text{rain} \end{array} \begin{array}{c} \begin{array}{cc} \text{sun} & \text{rain} \end{array} \\ \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix} \end{array}$$

← what are we assuming here?

Example 4 - Visiting web pages

Imagine that surfing on the internet users visiting a page click on its hyperlink uniformly at random.

Let X_n = page where you are after n click

State space: $S = \{\text{all web pages}\}$

Transition matrix: $P = [p_{ij}]$

$$p_{ij} = \begin{cases} \frac{1}{d_i} & \text{if page } s_i \text{ has a link to page } s_j \\ 0 & \text{otherwise} \end{cases}$$

d_i = n. of links from page s_i

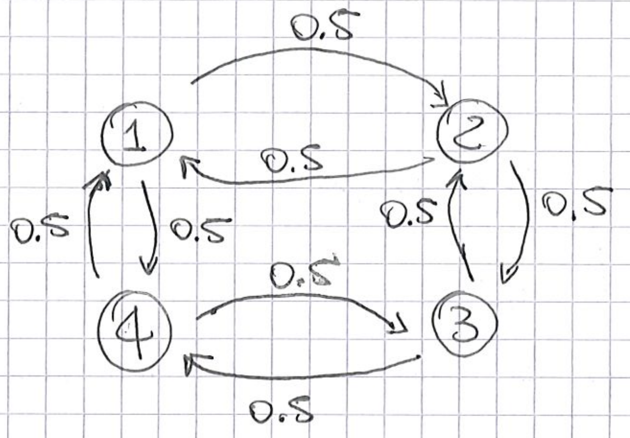
If $\underset{\substack{\uparrow \\ \text{page with no links}}}{d_i = 0}$ we define $\begin{cases} p_{ii} = 1 & \leftarrow \text{stop at this page} \\ p_{ij} = 0 & \forall i \neq j \end{cases}$

Note that since S is not infinite, but is very big, this process is very complicated
(or just seems?)

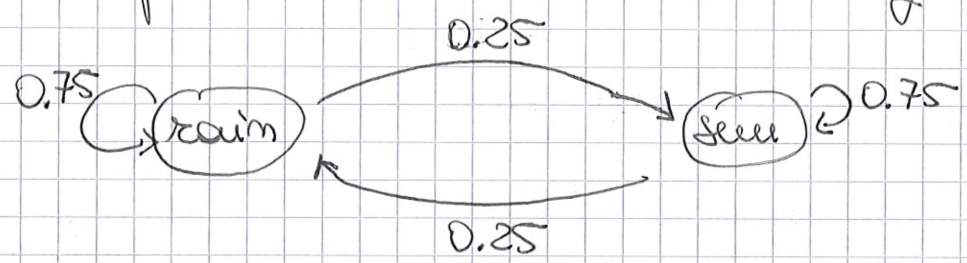
Representation of Markov Chains :

transition graphs

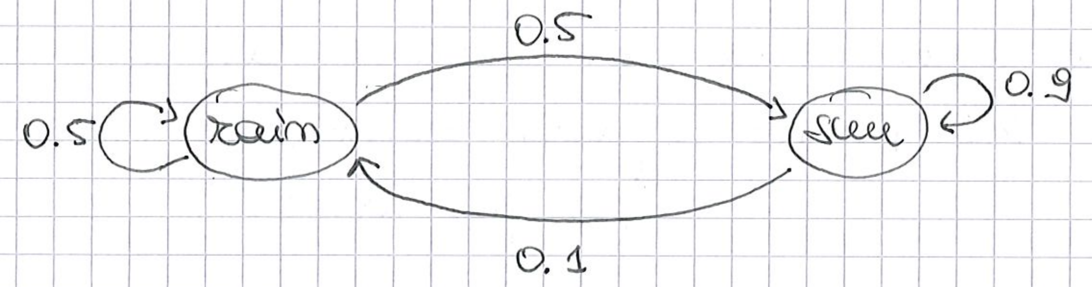
Example 1 : random walker



Example 2 : weather in Gothenburg



Example 3 : weather in Los Angeles



Example 4 : too big S to use transition graphs

INHOMOGENEOUS MARKOV CHAINS

Are MC where the transition probabilities change in time.

Definition: Let $P^{(1)}, P^{(2)}, \dots$ be a sequence of $k \times k$ matrices, each of which has elements $P_{ij}^{(l)}$ such that

$$P_{ij}^{(l)} \in [0,1] \quad \forall i,j \qquad \text{(are probabilities)}$$

$$\sum_{j=1}^{k} P_{ij}^{(l)} = 1 \qquad \text{(rows sum to one)}.$$

A random process $(X_0, X_1, \dots)$ with finite state space $S = \{s_1, \dots, s_k\}$ is said to be an inhomogeneous Markov Chain with transition matrices $P^{(1)}, P^{(2)}, \dots$ if $\forall n \in \mathbb{N}$, $\forall i,j \in \{1, \dots, k\}$, $\forall i_0, \dots, i_{n-1} \in \{1, \dots, k\}$ we have

$$\mathbb{P}(X_{n+1} = s_j \mid X_0 = s_{i_0}, X_1 = s_{i_1}, \dots, X_{n-1} = s_{i_{n-1}}, X_n = s_i) =$$
$$= \mathbb{P}(X_{n+1} = s_j \mid X_n = s_i) = P_{ij}^{(n+1)}$$

Example 5: refined model for Gothenburg weather

Assume now that we take into account seasons, and we want to use different models for summer and winter. Let us extend the state space to $\{s_1, s_2, s_3\}$ with

s_1 = rain

s_2 = sunshine

s_3 = snow

We may assume that the weather X_t evolves in S according to

$$P_{summer} = \begin{bmatrix} 0.75 & 0.25 & 0 \\ 0.25 & 0.75 & 0 \\ 0.5 & 0.5 & 0 \end{bmatrix} \quad \text{in May - September}$$

and

$$P_{winter} = \begin{bmatrix} 0.5 & 0.3 & 0.2 \\ 0.15 & 0.7 & 0.15 \\ 0.2 & 0.3 & 0.5 \end{bmatrix} \quad \text{in October - April}$$

This is an inhomogeneous MC !

How do we compute now the distribution of the process at step n ?

<u>theorem</u>: Let $(X_0, X_1, \dots)$ be an inhomogeneous Markov chain with state space $S = \{s_1, \dots, s_k\}$, initial distribution $\mu^{(0)}$ and transition matrices $P^{(1)}, P^{(2)}, \dots$. Then, for any n, we have

$$\mu^{(n)} = \mu^{(0)} P^{(1)} P^{(2)} \dots P^{(n)}$$