Markov Chains

Definition

A Markov chain or Markov process is a stochastic model describing a sequence of possible events (Also known as state by convention) in which **the probability of each event depends only on the state attained in the previous event.**

In mathematical denotation, it is written as

$$P(X_{n+1} = x | X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = P(X_{n+1} = x | X_n = x_n)$$

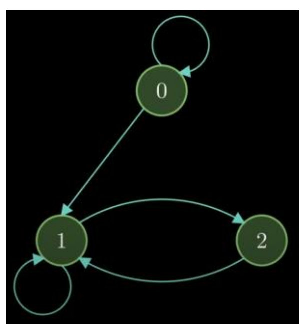
This assumption makes a lot of problems in the real world easier.

Another rule is the sum of arrows from one state must be 1 since it represents probability.

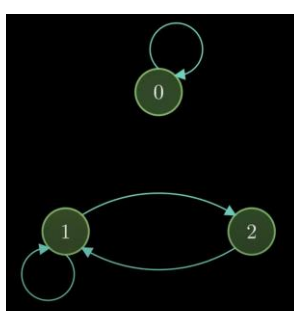
Transient state and recurrent state

Assume we have the following transition relations. Since there is no edge (arrow / probability) from other states to state 0, we called state 0 as **Transient state. Which means the probability of revisiting 0 is less than 1.**

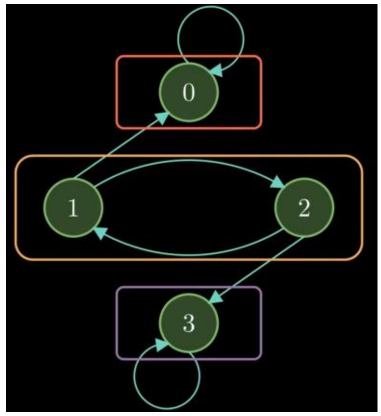
In the contrast, states 1 and 2 are known as **recurrent state, which means the probability of revisiting is 1.**



In this situation, we claimed this Markov chain is **reducible**, it can be divided into two **irreducible** situations (**all states are recurrent**).



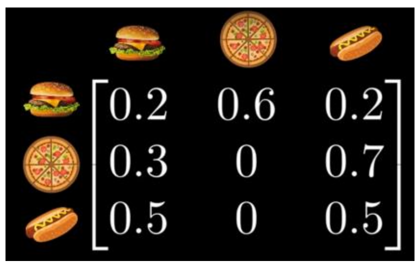
The most famous reducible Markov chain is **Gambler's ruin,** which can be divided into three irreducible one, as follows

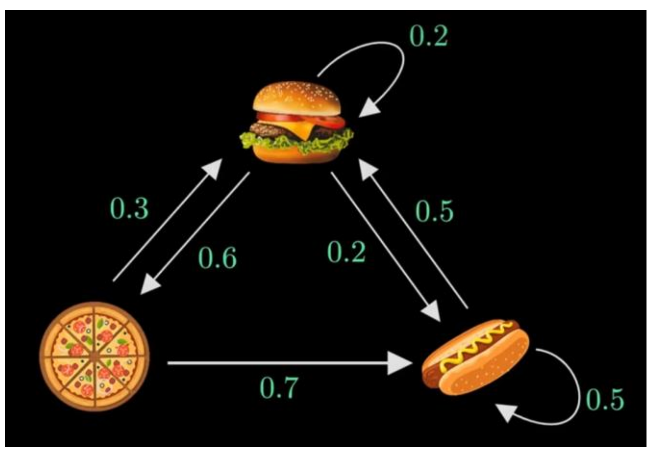


Example

Find the stationary state of a Markov chain.

Imagine we have the following relationships among three states. We could denote those relationships by a **transition matrix A** (0 means no edge between two vertices)





If we have the pizza now, denote as the vector $[0\ 1\ 0]$, use the vector to multiply the transition matrix, we could get the probability of next day

$$\pi_0 A = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.2 & 0.6 & 0.2 \\ 0.3 & 0 & 0.7 \\ 0.5 & 0 & 0.5 \end{bmatrix} = \begin{bmatrix} 0.3 & 0 & 0.7 \end{bmatrix}$$

If π_0 is the eigenvector of the transition matrix A, **we get $\boldsymbol{\pi_0 A = \pi_0 \lambda}$ by the constrain $\boldsymbol{\pi[1] + \pi[2] + \pi[3] = 1}$.** After solving those equations, we get (The number of eigenvectors is usually more than 1)

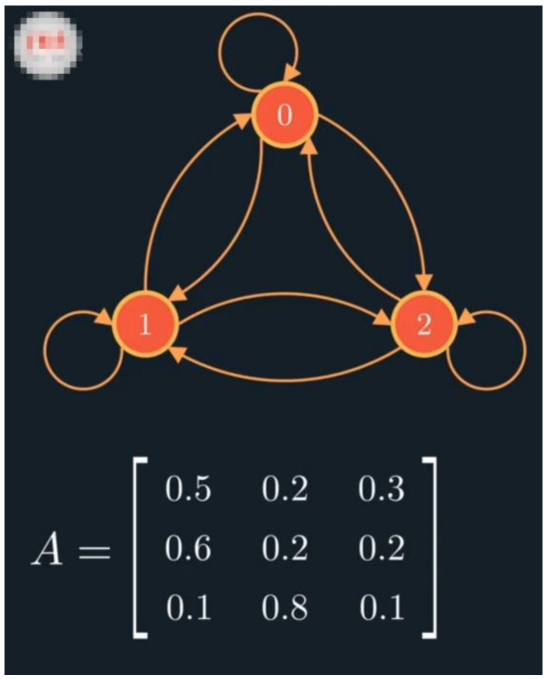
$$\boldsymbol{\pi} = [\mathbf{0.35211} \quad \mathbf{0.21127} \quad \mathbf{0.43662}]$$

So, this is the stationary probability of the given situation.

Chapman-Kolmogorov Theorem

Design purpose

Assume we have the following Markov chain, **what is the probability of reaching the state $\boldsymbol{i}$ from state $\boldsymbol{j}$ after exactly $\boldsymbol{n}$ steps.**



Before dig into the general situation, we restrict the situation to "**probability of reaching state 2 from state 0 after 2 steps**", it equals to $p(1|0)p(2|1) + p(2|0)p(2|2) + p(0|0)p(2|0) =$

$$A_{00}A_{02} + A_{01}A_{12} + A_{02}A_{22} = [A_{00} \quad A_{01} \quad A_{02}] \begin{bmatrix} A_{02} \\ A_{12} \\ A_{22} \end{bmatrix} = 0.22$$

It equals to $\pi_0 = [1 \quad 0 \quad 0], \pi_0 * (A * A) = [1 \quad 0 \quad 0] \begin{bmatrix} 0.4 & 0.38 & 0.22 \\ 0.44 & 0.32 & 0.24 \\ 0.54 & 0.26 & 0.20 \end{bmatrix} = [0.4\ 0.38\ 0.22]$

So, **the chapman-Kolmogorov theorem** says, the probability from state i to j after n steps, denote as $\boldsymbol{P_{ij}(n) = \sum_k P_{ik}(r) * P_{kj}(n-r)}$

Proof

$$\begin{aligned} P_{ij}(n) &= P(X_n = j \mid X_0 = i) \\ &= \sum_k P(X_n = j, X_r = k \mid X_0 = i) \\ &= \sum_k \frac{P(X_n = j, X_r = k,\ X_0 = i)}{P(X_0 = i)} \end{aligned}$$

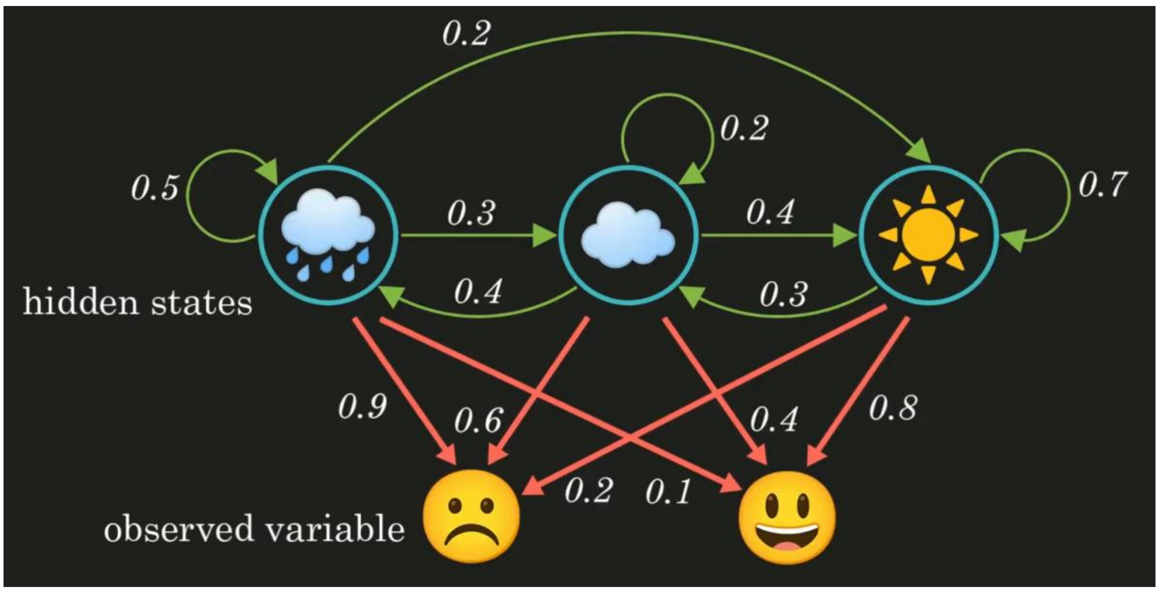
$$= \sum_k \frac{P(X_n = j \mid X_r = k, X_0 = i) * P(X_r = k, X_0 = i)}{P(X_0 = i)}$$

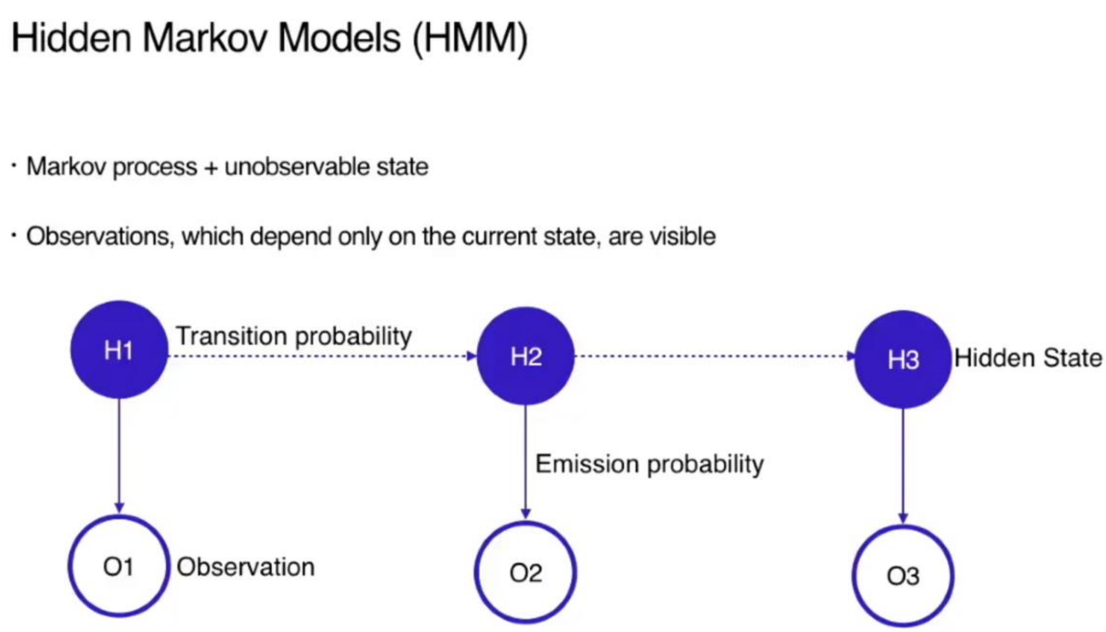
$$= \sum_k \frac{P_{kj}(n-r) * P(X_r = k, X_0 = i)}{P(X_0 = i)}$$

$$= \sum_k P_{ik}(r) * P_{kj}(n-r)$$

Hidden Markov Models (HMM)

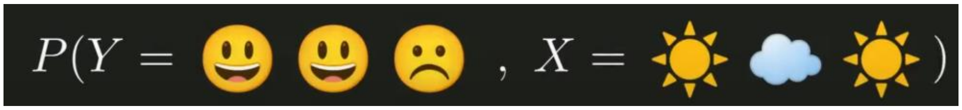
Hidden Markov Chain = hidden states + observed variable



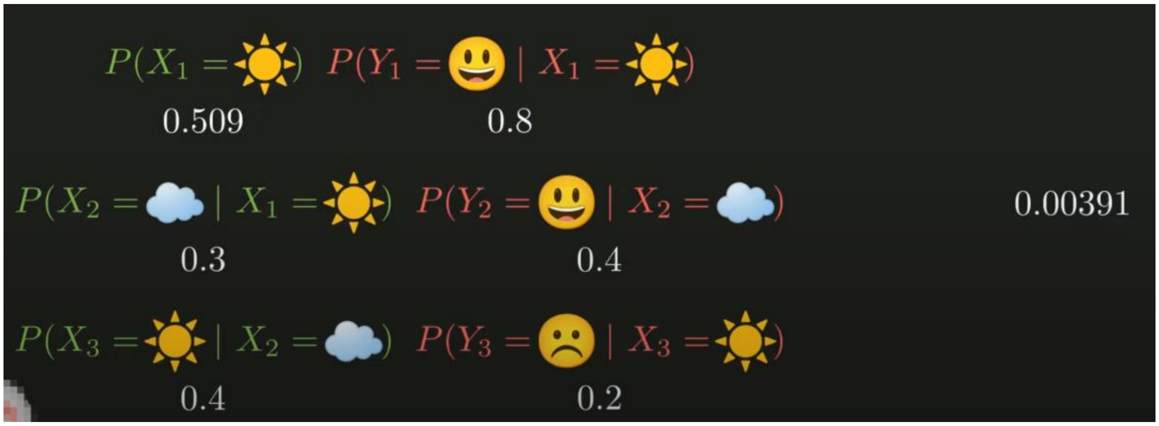


Example

1. Get the probability of



Use the normalized left eigen vector to find the probability of sunny day



2. What is the most likely weather sequence for the observed mood sequence?

$$\underset{X=X_1,X_2,\ldots X_n}{\arg\max} \prod P(Y_i \mid X_i)\ P(X_i \mid X_{i-1})$$

Forward Algorithm