Markov Chains

Céline Comte and Fabien Mathieu

TELECOM ParisTech

NOKIA Bell Labs

ACN – Propagation in Graphs

November 12, 2018

References

- Strongly inspired by the chapter "Chaînes de Markov" of the first-year class "MDI104: Cours de Probabilités" at Télécom ParisTech

- Sergey Brin and Lawrence Page (1998). "The Anatomy of a Large-scale Hypertextual Web Search Engine". In: Proceedings of the Seventh International Conference on World Wide Web 7

Outline

Definition

Irreducibility

Aperiodicity

Example: PageRank algorithm

Markov Chain

$(X_k)_{k\in\mathbb{N}}$ sequence of random variables with values in $\{1,\dots,n\}$.
Typically, $X_k \sim$ state of a system at time k.

$(X_k)_{k\in\mathbb{N}}$ is a Markov chain if **the future is conditionally independent of the past given the present**, that is

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

for all $k \in \mathbb{N}$ and $x_0, x_1, \dots, x_k, x_{k+1} \in \{1,\dots,n\}$.

$1,\dots,n$ are called the **states** of the Markov chain.

Transition graph

(Time) Homogeneity: We assume that the probability of the next transition is independent of the time k when it occurs.

Transition graph: A directed graph with

- Nodes ~ the states 1,…,n of the Markov chain
- Arrows ~ the transitions, weighted by their probability

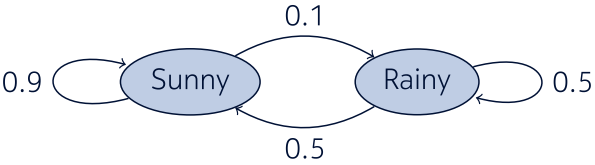
Useful representation for reasoning about the Markov chain.

The weights of the outgoing edges from each node sum to 1.

Example: Weather (Wikipedia)

$X_k \in \{\text{sunny}, \text{rainy}\}$ is the weather on day k

- If it's sunny today, it'll also be sunny tomorrow with probability 0.9, otherwise it'll be rainy;
- If it's rainy today, it'll also be rainy tomorrow with probability 0.5, otherwise it'll be sunny.



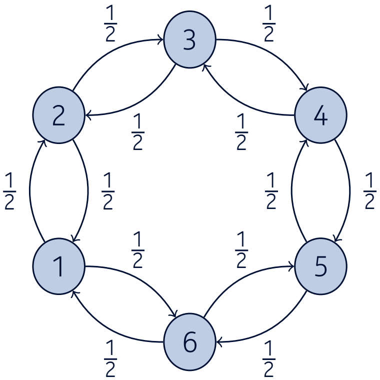
Question: On average, is it more often sunny or rainy?

Example: Simple game

$X_k \in \{1, \ldots, 6\}$ is the position of the player at time k.

At each turn, the player flips an unbiased coin,

- If it's heads, the player moves down;
- It it's tails, the player moves up.



Question: How does the position evolve after a long time?

Transition matrix

Adjacency matrix $A = (a_{i,j})_{i,j=1,\dots,n}$ of the transition graph

$$
\begin{aligned}
a_{i,j} &= \text{weight of the arrow from state i to state j,} \\
&= \text{probability } \mathbb{P}(X_{k+1} = j \,|\, X_k = i) \text{ that the} \\
&\quad\ \text{Markov chain moves from state i to state j,}
\end{aligned}
$$

The matrix A is **right-stochastic**, because

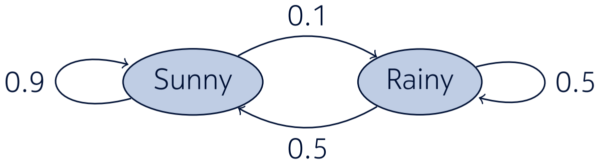
$$
a_{i,j} \geq 0, \quad \forall i,j = 1,\dots,n,
$$

$$
\text{and} \quad \sum_{j=1}^{n} a_{i,j} = 1, \quad \forall i = 1,\dots,n.
$$

Useful representation for doing calculations.

Example: Weather (Wikipedia)

$X_k \in \{sunny, rainy\}$ is the weather on day k

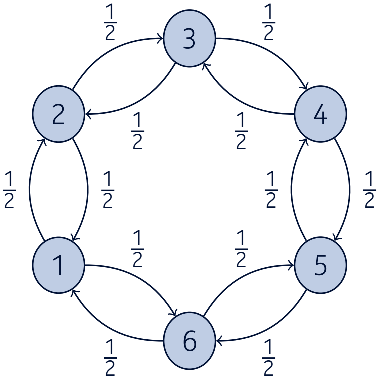


$$A = \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix}$$

Example: Simple game

$X_k \in \{1, \ldots, 6\}$ is the position of the player at time k.

$$A = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \end{bmatrix}$$



Distribution at time k

- Distribution of the state of the Markov chain at time k:

$$P_k = \left[\mathbb{P}(X_k = 1) \quad \dots \quad \mathbb{P}(X_k = n-1) \quad \mathbb{P}(X_k = n)\right].$$

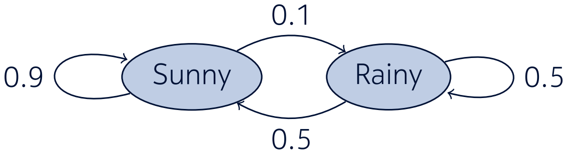
- Recursive formula between times k and $k+1$:

$$\begin{aligned}\mathbb{P}(X_{k+1} = j) &= \sum_{i=1}^{n} \mathbb{P}(X_k = i)\mathbb{P}(X_{k+1} = j \mid X_k = i) \quad \left(\begin{array}{c}\text{law of total}\\ \text{probabilities}\end{array}\right),\\ &= \sum_{i=1}^{n} \mathbb{P}(X_k = i) a_{i,j}.\end{aligned}$$

In other words: $\quad P_{k+1} = P_k A, \quad \forall k \in \mathbb{N}.$

Example: Weather (Wikipedia)

$X_k \in \{\text{sunny}, \text{rainy}\}$ is the weather on day k.



$$A = \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix}$$

Assume that, on the initial day, it's sunny with probability 0.6 and rainy with probability 0.4: $P_0 = [0.6 \quad 0.4]$.

On the next day, it's sunny with probability 0.74 and rainy with probability 0.26:

$$P_1 = P_0 A = [0.6 \quad 0.4] \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix} = [0.74 \quad 0.26].$$

Stationary distribution

- If the distribution has a limit P, then:

$$P = \lim_{k \to +\infty} P_{k+1} = \lim_{k \to +\infty} P_k A = PA.$$

Therefore, the limit P is a **“fixed-point”** of the matrix A (that is, a left eigenvector for the eigenvalue 1):

$$P = PA.$$

- P is called a **stationary distribution** of the Markov chain:

If $P_0 = P$, then we have:

$$P_1 = P_0 A = PA = P,$$
$$P_2 = P_1 A = PA = P,$$
and so on …

Outline

Definition

Irreducibility

Aperiodicity

Example: PageRank algorithm

Irreducible Markov chain

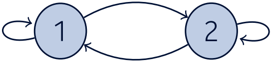
The Markov chain $(X_k)_{k \in \mathbb{N}}$ is **irreducible** if one of the two following equivalent conditions is satisfied:

- Its transition graph is **strongly connected**, that is, there exists a directed path between any two nodes.
- Whichever the current state of the Markov chain, it can reach any other state with a non-zero probability.

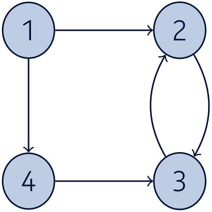
Remark: This condition only depends on the existence of the arrows, but not on their weight.

A Markov chain that is not irreducible is said **reducible**.

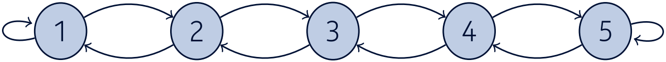
Examples



Irreducible
(path 1 → 2 → 1)



Reducible
(no path from 2 to 1)



Irreducible
(path 1 → 2 → 3 → 4 → 5 → 4 → 3 → 2 → 1)

Perron-Frobenius theorem (Part 1)

Theorem: Consider an **irreducible** Markov chain $(X_k)_{k\in\mathbb{N}}$.

- $(X_k)_{k\in\mathbb{N}}$ has a unique stationary distribution P.
- P is the unique **normalized** solution of $P = PA$.
  (In other words, P is the unique normalized, left eigenvector of A for the eigenvalue 1.)
- All components of P are positive.

Remark: This theorem proves the existence and the unicity of the stationary distribution P but not necessarily the convergence of P_k to P as $k \to +\infty$.

You will see in the practical what happens if the Markov chain is reducible.

Ergodic theorem

Theorem: Let $(X_k)_{k\in\mathbb{N}}$ be an **irreducible** Markov chain and P its stationary distribution. Then, for each state $i = 1,\dots,n$,

$$\frac{1}{K}\sum_{k=1}^{K} 1_{\{X_k=i\}} \to P(i) \quad \text{almost surely as } K \to +\infty.$$

In other words, the i^{th} component $P(i)$ of P is the fraction of time the Markov chain spends in state i during one realization.

Remark: In the rest of the course, we will systematically work with irreducible Markov chains.

Outline

Definition

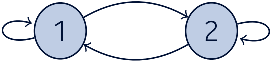
Irreducibility

Aperiodicity

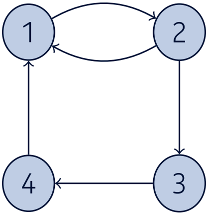
Example: PageRank algorithm

Period

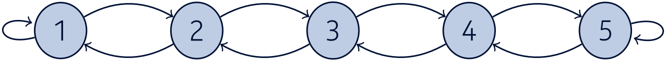
The **period** of a node is the greatest common divisor (GCD) of the lengths of its cycles.



Both nodes have period 1
(there is a loop)



All nodes have period 2



All nodes have period 1 (again, there is a loop)

  Public

NOKIA Bell Labs TELECOM ParisTech

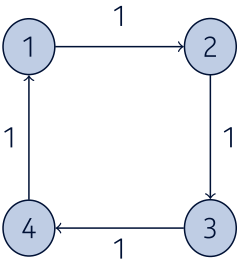
Period

The **period** of a node is the greatest common divisor (GCD) of the lengths of its cycles.

Remarks:

- Again, this definition only depends on the existence of the arrows and not on their weights.
- In an irreducible Markov chain, all states have the same period. Therefore, we can define the **period** of the Markov chain as the shared period of its states.
- The Markov chain is said **aperiodic** it it has period 1.

Example: Period 4



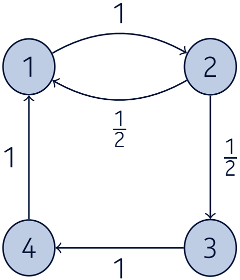
$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

The Markov chain is initially in state 1

$P_0 = [1 \quad 0 \quad 0 \quad 0]$

- Time 1: State 2
  $P_1 = [0 \quad 1 \quad 0 \quad 0]$
- Time 2: State 3
  $P_2 = [0 \quad 0 \quad 1 \quad 0]$
- Time 3: State 4
  $P_3 = [0 \quad 0 \quad 0 \quad 1]$
- Time 4: State 1
  $P_4 = [1 \quad 0 \quad 0 \quad 0]$

Example: Period 2



$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

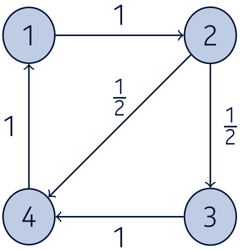
The Markov chain is initially in state 1

$P_0 = [1 \quad 0 \quad 0 \quad 0]$

- Time 1: State 2
  $P_1 = [0 \quad 1 \quad 0 \quad 0]$
- Time 2: State 1 or 3
  $P_2 = [\frac{1}{2} \quad 0 \quad \frac{1}{2} \quad 0]$
- Time 3: State 2 or 4
  $P_3 = [0 \quad \frac{3}{4} \quad 0 \quad \frac{1}{4}]$
- Time 4: State 1 or 3
  $P_4 = [\frac{5}{8} \quad 0 \quad \frac{3}{8} \quad 0]$

→ The Markov chain keeps alternating between $\{1, 3\}$ and $\{2, 4\}$.

Example: Period 1



$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

The Markov chain is initially in state 1

- Time 1: State 2
- Time 2: State 3 or 4
- Time 3: State 1 or 4
- Time 4: State 1 or 2
- Time 5: State 2, 3, or 4
- Time 6: State 1, 3, or 4
- Time 7: State 1, 2, or 4
- Time 8: State 1, 2, 3, or 4

→ If we wait long enough, the Markov chain can be in any state at each instant

Perron-Frobenius theorem (Part 2)

Theorem: Consider an **irreducible** Markov chain $(X_k)_{k \in \mathbb{N}}$.

- $(X_k)_{k \in \mathbb{N}}$ has a unique stationary distribution P.
- P is the unique normalized solution of $P = PA$.
  (In other words, P is the unique normalized, left eigenvector of A for the eigenvalue 1.)
- All components of P are positive.

If $(X_k)_{k \in \mathbb{N}}$ is also **aperiodic**, then $P_k \to P$ as $k \to +\infty$.

Outline

Definition

Irreducibility

Aperiodicity

Example: PageRank algorithm

Introduction

- Proposed by Larry Page and Sergey Brin in 1996.
- **Idea**: Associate to each web page a rank that is proportional to its popularity.
- **PageRank algorithm**: Compute the rank of each page. Imagine a random surfer who wanders through the web graph, following edges at random. The rank of a page is the proportion of time the surfer spends on this page.
- **Web graph**: Directed graph
  - Nodes ~ Web pages
  - Edges ~ Hyperlinks

Markov chain formulation (1st version)

At each step, the surfer chooses an outgoing edge uniformly at random and jumps to the page pointed by this edge.

$X_k \in \{1,\ldots,n\}$ is the position of the random surfer at time k.

If this Markov chain is irreducible and aperiodic, then we can apply Perron-Frobenius theorem:

- The (Page)rank is defined as its stationary distribution P.
- We can approximate P by starting from an arbitrary distribution P_0 and iterating over $P_{k+1} = P_k A$.

Otherwise, see the second exercice of the practical ...

