What this all is about?

Lecture 2

Markov chain

Petr Lukianchenko

9 September 2024

Syllabus

Lecture 1
Recall basic of probability
MGF to start

Lecture 2
Markov chain

Lecture 3
Convergences

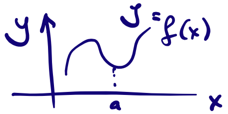
Lecture 4
Conditional Expectations

Lecture 5
Poisson distribution

Lecture 6
G-algebra

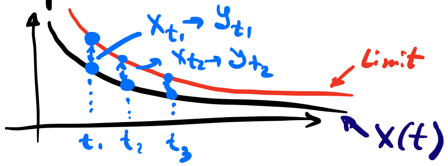
Lecture 7
Filtration

$\lim_{x\to a} f(x) = f(a)$ deterministic

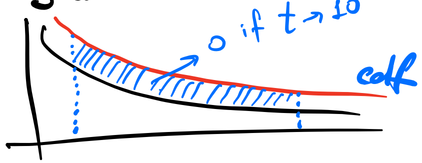


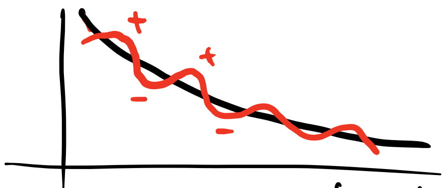
$\boxed{\lim_{t\to\infty} \eta_t^2 = \eta_\infty^2}$

- point-wise



- by distribution





- in probability

$$\lim_{t\to\infty} P(|X(t) - X(\infty)| > \varepsilon) = 0$$

$$\lim_{t\to\infty} E[(X(t) - X(\infty))^q] = 0$$

- almost surely

$$\lim_{t\to\infty} P(|X(t) - X(\infty)| < \varepsilon) = 1$$

Stochastic Processes: Basic Definitions

Stochastic process $\{X_t\}_{t \in T}$

The value of a variable changes in an uncertain way

Discrete vs. continuous time

When can a variable change?

What values can a variable take?

Markov property

Only the current value of a variable is relevant for future predictions

! No information from past prices or path

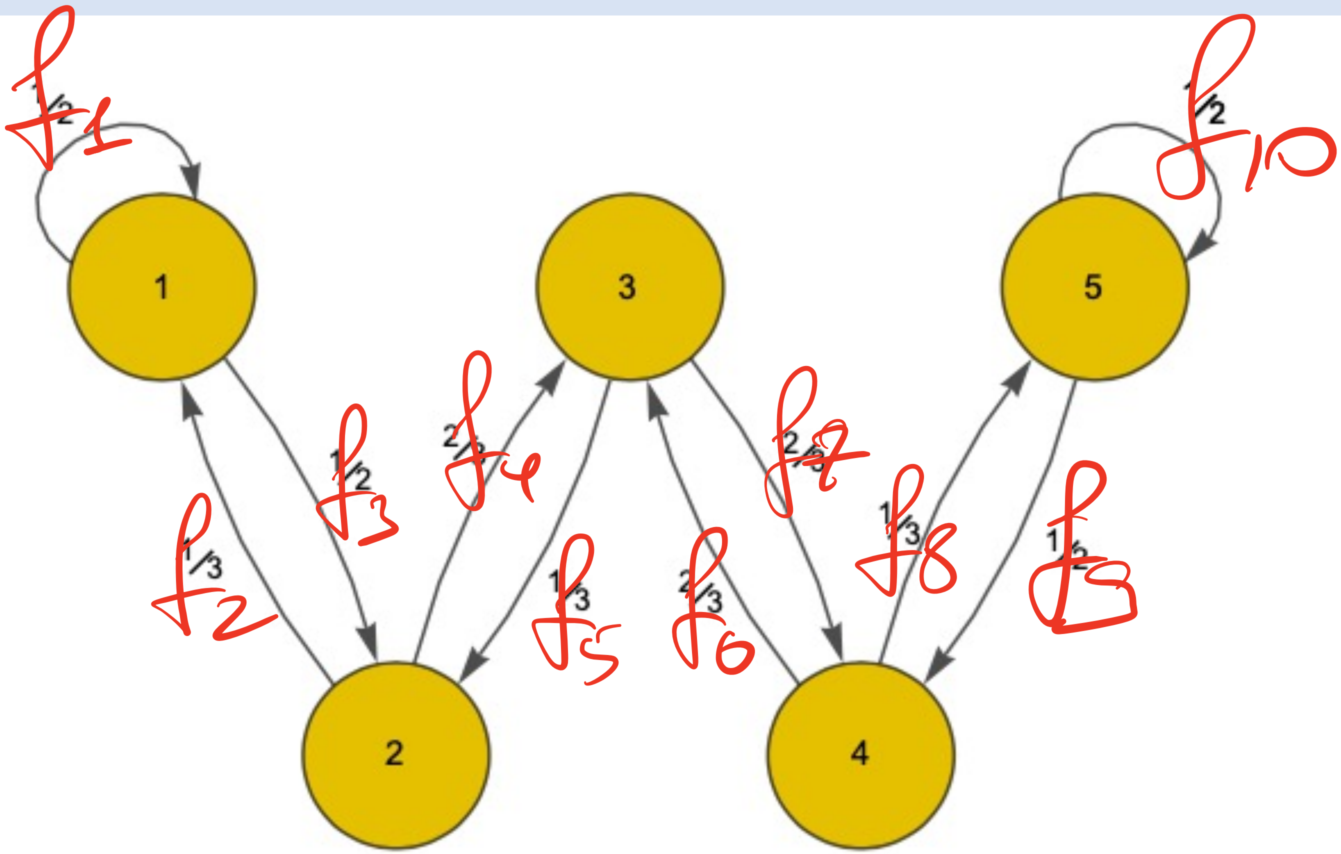
static

$$P_t \quad \begin{array}{c|cc} & 1 & 2 \\ \hline 1 & 1/3 & 2/3 \\ 2 & 1/2 & 1/2 \end{array} \quad P_{t+1}$$

dynamic

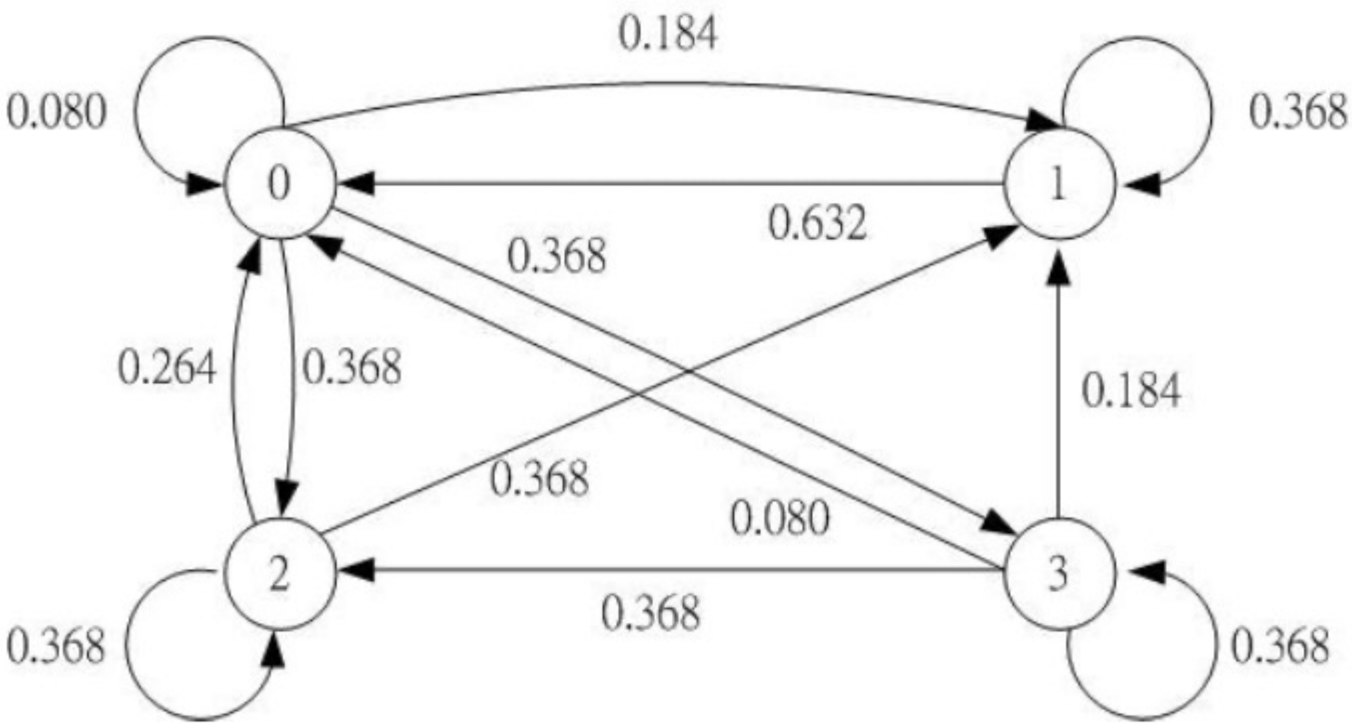
$$P_t \quad \begin{array}{c|cc} & 1 & 2 \\ \hline 1 & \frac{1}{3}+\frac{1}{3t} & \frac{2}{3}-\frac{1}{3t} \\ 2 & \frac{1}{2}+\frac{1}{2t} & \frac{1}{2}-\frac{1}{2t} \end{array} \quad P_{t+1}$$

A Chain



Markov Chain

- The state transition diagram:



Markov Chain

- Consider time index $n = 0, 1, 2, \ldots$ & time dependent random state X_n
- State X_n takes values on a countable number of states
  - In general denotes states as $i = 0, 1, 2, \ldots$
  - Might change with problem
- Denote the history of the process $\mathbf{X}_n = [X_n, X_{n-1}, \ldots, X_0]^T$
- Denote stochastic process as $\mathbf{X}_{\mathbb{N}}$
- The stochastic process $X_{\mathbb{N}}$ is a Markov chain (MC) if

$$\mathrm{P}\left[X_{n+1} = j \,\middle|\, X_n = i, \mathbf{X}_{n-1}\right] = \mathrm{P}\left[X_{n+1} = j \,\middle|\, X_n = i\right] = P_{ij}$$

- Future depends only on current state X_n

Observations

- Process's history $\mathbf{X}_{n-1}$ irrelevant for future evolution of the process
- Probabilities P_{ij} are constant for all times (time invariant)
- From the definition we have that for arbitrary m

$$\mathrm{P}\left[X_{n+m} \,\middle|\, X_n, \mathbf{X}_{n-1}\right] = \mathrm{P}\left[X_{n+m} \,\middle|\, X_n\right]$$

- X_{n+m} depends only on X_{n+m-1}, which depends only onX_{n+m-2}, $\ldots$ which depends only on X_n
- Since P_{ij}'s are probabilities they're positive and sum up to 1

$$P_{ij} \geq 0 \qquad \sum_{j=1}^{\infty} P_{ij} = 1$$

Matrix Representation

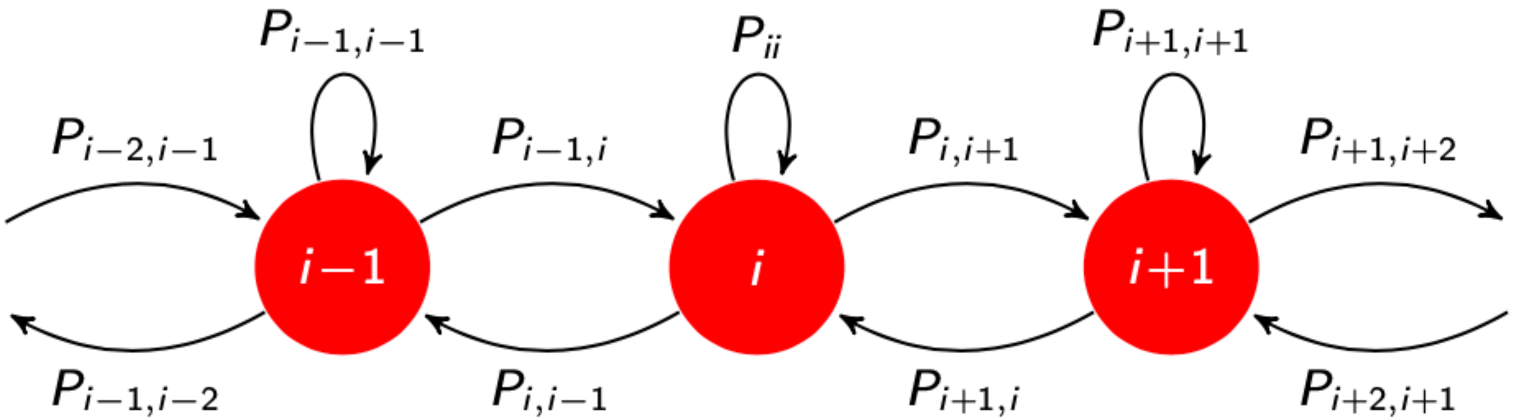
- Group transition probabilities P_{ij} in a "matrix" $\mathbf{P}$

$$\mathbf{P} := \begin{pmatrix} P_{00} & P_{01} & P_{02} & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ P_{i0} & P_{i1} & P_{i2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

- Not really a matrix if number of states is infinite

Graph Representation

- A graph representation is also used



- Useful when number of states is infinite

Homogeneous Chain

- The evolution of a markov chain is defined by its transition probability, defined by $\mathbb{P}(X_{n+1} = j | X_n = i)$ (where without loss of generality we may assume that S is an integer set.

Definition 45

- The chain $\{X_n\}$ is called *homogeneous* if its transition probabilities do not depend on the time, i.e.,

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(X_1 = j | X_0 = i)$$

for all n, i, j. The *transition probability matrix* $\boldsymbol{P} = [p_{i,j}]$ *is the* $|S|$ x $|S|$ matrix of the transition probabilities, such that $p_{i,j} = \mathbb{P}(X_{n+1} = j | X_n = i)$

Persistent and Transient States

- A state i ∈ S is called *persistent* (or recurrent) if

$$\mathbb{P}(X_n = i \text{ } for \text{ } some \text{ } n \geq 1 | X_0 = i) = 1$$

- Otherwise, if the above probability is strictly less than 1, the state is called *transient*.

- We are interested in the *first passage* probability

$$f_{i,j}(n) = \mathbb{P}(X_1 \neq j, X_2 \neq j, \ldots, X_{n-1} \neq j, X_n = j | X_0 = i)$$

- We define $f_{i,j} = \sum_{n=1}^{\infty} f_{i,j}(n)$. **Note**: state j is persistent if and only if $f_{j,j} = 1$.

Ex 1

$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}$

eigenvalues $(\frac{1}{6} \quad 1)$

eigenvectors $\times \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$A \cdot \vec{v}_1 = \lambda_1 \cdot \vec{v}_1$

$A^{\infty} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

Stationary Distribution

Definition 50

The vector π is called a *stationary distribution* of the chain if it has entries $\{\pi_j : j \in S\}$ such that:

a) $\pi_j \geq 0$ for all j, and $\sum_{j \in S} \pi_j = 1$.

b) it satisfies $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$, that is, $\pi_j = \sum_i \pi_i p_{i,j}$ for all $j \in S$.

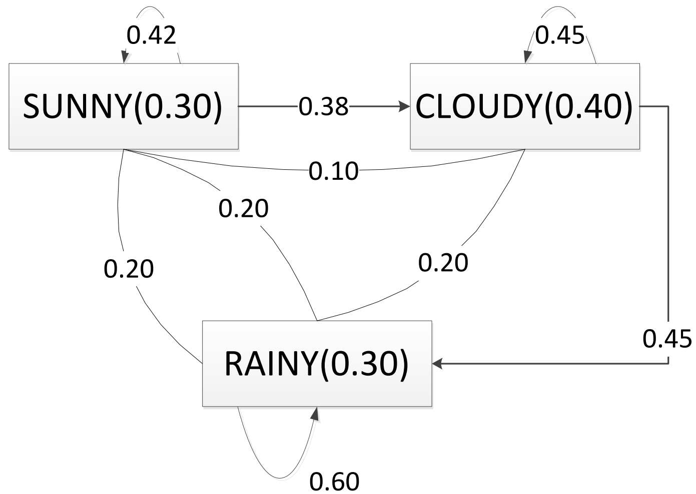
- This is called "stationary distribution" since if X_0 is distributed with $\boldsymbol{u}(0) = \boldsymbol{\pi}$, then all X_n will have the same distribution, in fact

$$\mathbf{u}(n) = \mathbf{u}(0)\mathbf{P}^n = \boldsymbol{\pi}\mathbf{P}^n = \boldsymbol{\pi}\mathbf{P}\mathbf{P}^{n-1} = \boldsymbol{\pi}\mathbf{P}^{n-1} = \cdots = \boldsymbol{\pi}$$

- Given the classification of chains and the decomposition theorem, we shall assume that the chain is *irreducible*, that is, its state space is formed by a single equivalence class of intercommunicating (persistent) states C or by the class of transient states T.

Markov Chain

Stochastic FSM



The transition matrix:

$$A = \begin{pmatrix} 0.42 & 0.38 & 0.20 \\ 0.10 & 0.45 & 0.45 \\ 0.20 & 0.20 & 0.60 \end{pmatrix}$$

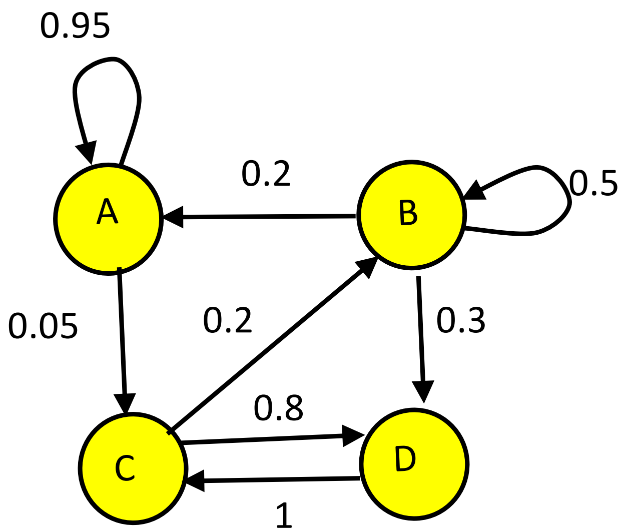
- Stochastic matrix:
  Rows sum up to 1
- Double stochastic matrix:
  Rows and columns sum up to 1

Markov Chain

	System state is fully observable	*System state is partially observable*
System is autonomous	Markov chain	Hidden Markov model
System is controlled	Markov decision process	Partially observable Markov decision process

Markov Chain

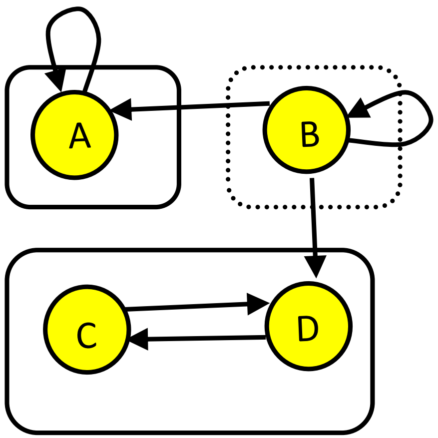
Each directed edge A→B is associated with the **positive** transition probability from A to B.



	A	B	C	D
A	0.95	0	0.05	0
B	0.2	0.5	0	0.3
C	0	0.2	0	0.8
D	0	0	1	0

Markov Chain

- States of Markov chains are classified by the digraph representation (omitting the actual probability values)
- A, C and D are ***recurrent*** states: they are in strongly connected components which are **sinks** in the graph.



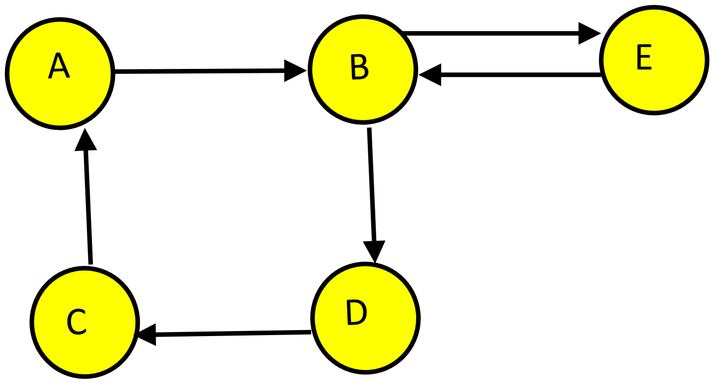
B is not recurrent – it is a ***transient*** state

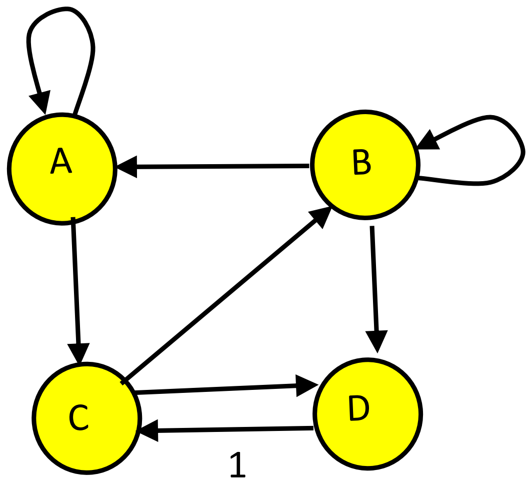
Alternative definitions:

A state ***s*** is ***recurrent*** if it can be reached from any state reachable from ***s;*** otherwise it is ***transient***.

Markov Chain

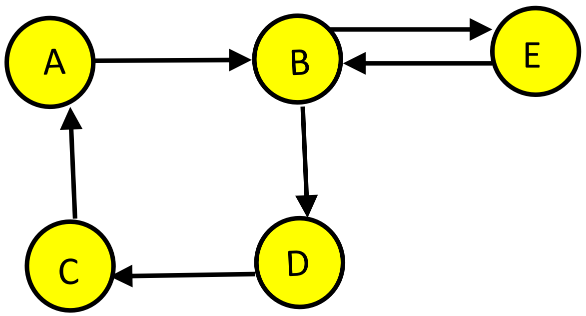
A Markov Chain is ***irreducible*** if the corresponding graph is strongly connected (and thus all its states are recurrent).



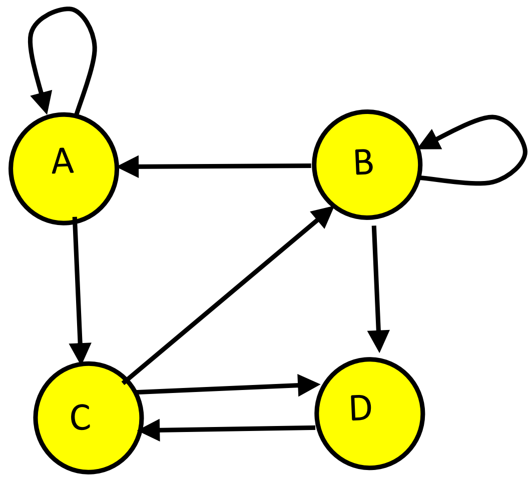


Markov Chain

- A state *s* has a period *k* if *k* is the *GCD* of the lengths of all the cycles that pass via *s*. (in the shown graph the period of A is 2).
- A Markov Chain is *periodic* if all the states in it have a period *k* >1. It is *aperiodic* otherwise.



Markov Chain



A Markov chain is ***ergodic*** if :

1. ***the corresponding graph is strongly connected.***
2. ***It is not peridoic***

Ergodic Markov Chains are important since they guarantee the corresponding Markovian process converges to a unique distribution, in which all states have strictly positive probability.

Thank you for your attention!

See next week!