Introduction to Markov Chains, Queuing Theory, and Network Performance

Marceau Coupechoux
Telecom ParisTech, departement Informatique et Réseaux
marceau.coupechoux@telecom-paristech.fr
IT.2403 Modélisation et Performance des Réseaux

February 8, 2011

Outlines

1. Network Performance
2. Mathematical Background
3. Markov Chains
4. Formalism of Queuing Theory
5. Simple Queues
6. Erlang B and C laws
7. Networks of Queues

1. What is performance evaluation ?
2. Why do we need to evaluate the performance of a system ?
3. How can we evaluate the performance of a system ?
4. Qualitative and quantitative analysis
5. Models
6. Result analysis
7. Stochastic characterization

What is performance evaluation ?

Computation of the performance parameters of a system

- Communication Networks
  - end-to-end delay
  - packet throughput
- Production Chain
  - number of products produced per hour
  - machine load
- Other
  - waiting time of customers
  - number of waiting customers

What is performance evaluation ?

Performance parameters

- mean throughput: $\bar{X}$
- mean response time: $\bar{R}$
- mean number of clients: $\bar{Q}$
- mean utilization ratio: $\bar{U}$
- higher order moments, e.g. variance

Why do we need performance evaluation ?

Design

- the system does not exist
- the system has to be dimensioned according to requirements
- experience is not sufficient or does not even exist
- try to avoid under-dimensioning, over-dimensioning

System in use

- the system exists
- it has to be modified
- it has to be tested in cases of failures or overload

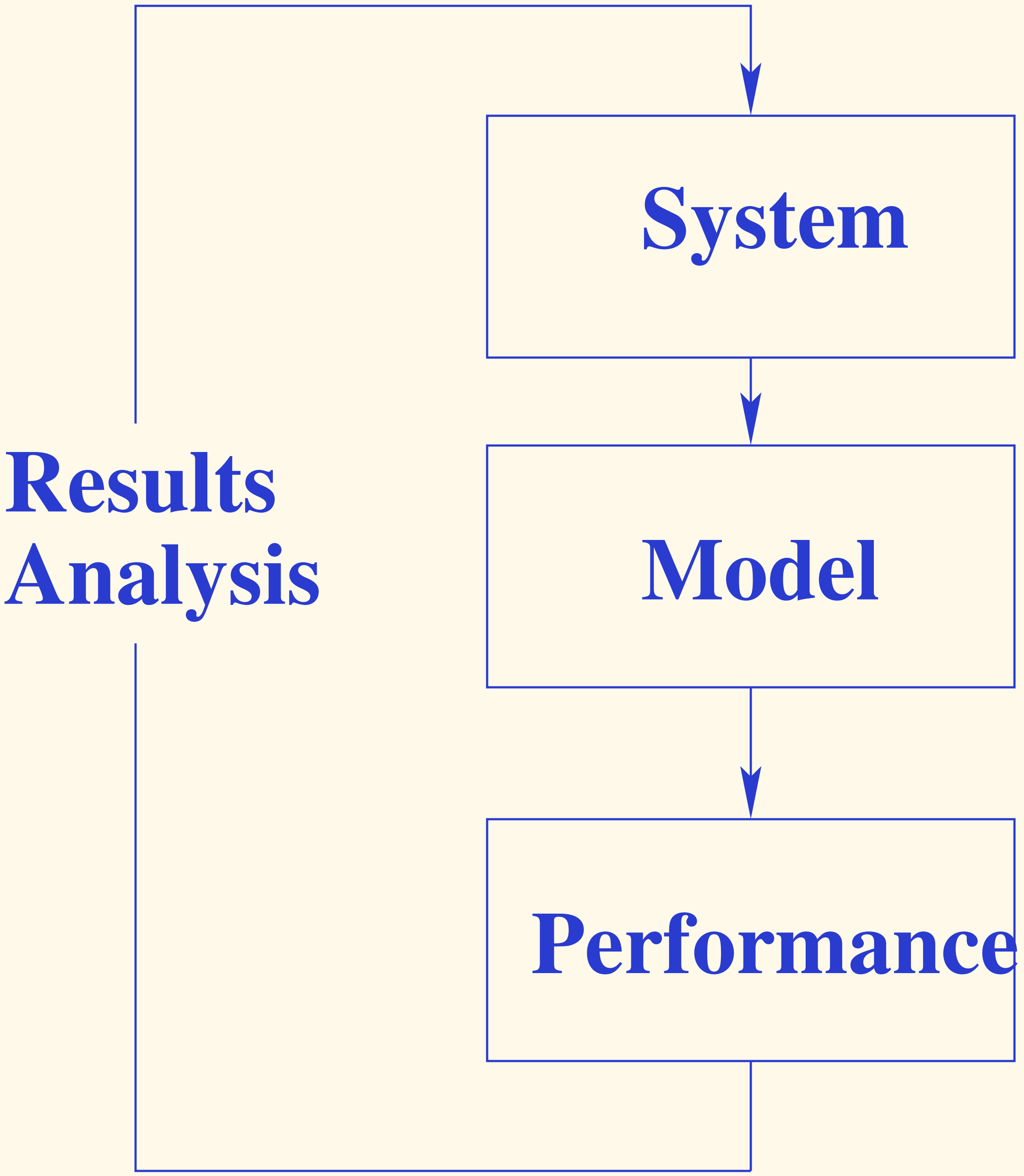
How can we evaluate the performance of a system ?

Propose a mathematical formalism to reproduce the behavior of the system
Model = mathematical abstraction of the system

- characterize a smaller or simpler system
- obtain equations describing the simplified system
- use of a mathematical formalism, e.g.:
  - Markov chains
  - Petri networks
  - queuing theory
  - stochastic automates
  - finite state machines
  - etc...

How can we evaluate the performance of a system ?

Performance evaluation cycle



Qualitative and quantitative analysis

Qualitative analysis

- Define behavior and properties of the system
- invariants of the system
- stability
- **Petri networks**

Quantitative analysis

- compute the performance of the system
- two different methods :
  - discrete event simulation
  - analytical methods

Qualitative and quantitative analysis

Simulations

- reproduce step by step the behavior of the system
- relies on an event scheduler
- advantage: very generic method for a wide range of applications
- drawbacks: time consuming and difficult interpretation of the results

Analytical methods

- solve equations of the model
- advantages: very fast and good comprehension of the system
- drawbacks: very specific
- approximations can be made to make the resolution of equations simpler and faster

Models

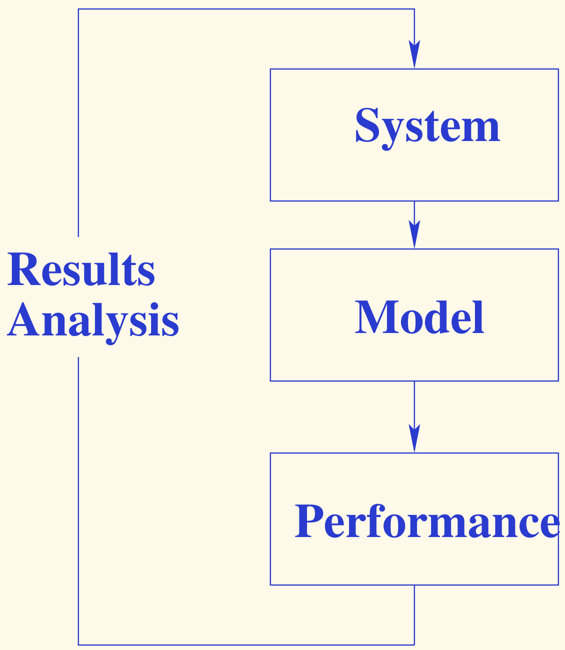
Design of the model is very important because we evaluate the performance of the model and not of the system

- model = trade-off between fidelity to the system and ease of analysis
- complex model implies:
  - model performance close to system performance
  - complex analysis
  - we need a lot of information on the system
- simplified model implies:
  - model performance far from system performance
  - simpler analysis

Results analysis

Does the configuration of the studied system fulfills the initial requirements?

- Yes: the analysis stops
- No: results analysis and new model



Stochastic characterization

In reality, all system are determinist

- in theory, it is thus possible to predict their behavior
- e.g. communication networks

In practice, sufficient information is not available

- behavior of a system is simplified by introducing random variables
- e.g. routing in communication networks
- e.g. location of base stations of a mobile network modeled by a Poisson point process

Mathematical background

1. Probabilities
2. Random variables
3. Examples of random variables
4. Stochastic processes
5. Poisson process
6. Transformations

A probability P is an application from Ω in $[0;1]$ that verifies

- $\forall A$, $0 \leq P[A] \leq 1$
- $P[\emptyset] = 0$ and $P[\Omega] = 1$
- $\forall A, B$, such that $A \cap B = \emptyset$, $P[A \cup B] = P[A] + P[B]$

Conditional probability

- If $P[B] \neq 0$, $P[A|B] = P[A \cap B]/P[B]$
- If A and B are independent, $P[A|B] = P[A]$

Total probability theorem

- $\{A_i\}$ is a partition of Ω iff $\cup A_i = \Omega$ and $\forall i \neq j$, $A_i \cap A_j = \emptyset$
- $\forall B$, $P[B] = \sum_i P[B|A_i] P[A_i]$

Definition

- X application from (Ω, P) in E, state space
- X: $\omega \rightarrow X(\omega)$

Two important types of random variables

- Discrete variable: $E \subset \mathbb{Z}$, X defined by $p(n) = P[X = n], n \in \mathbb{Z}$
- Continuous variable: $E \subset \mathbb{R}$, X defined by the probability density function (pdf) $f_X(x)$
- Discrete variable: $\sum_n p(n) = 1$, continuous variable: $\int_{-\infty}^{\infty} f_X(x)dx = 1$

Cumulative distribution function (cdf)

- Discrete variable: $F_X(n) = \sum_{k=-\infty}^{n} p(k)$
- Continuous variable: $F_X(x) = \int_{-\infty}^{x} f_X(y)dy$

Examples of random variables

Mean and moments of order k

- $E[X] = \sum np(n)$, $E[X^k] = \sum n^k p(n)$
- $E[X] = \int x f_X(x) dx$, $E[X^k] = \int x^k f_X(x) dx$
- Variance: $V[X] = E[X^2] - E[X]^2$, standard deviation: $\sigma[X] = \sqrt{V[X]}$

Geometric law of parameter $0 < a < 1$

- Discrete random variable with positive values
- $P[X = n] = (1 - a)a^n$
- $E[X] = a/(1 - a)$, $V[X] = a/(1 - a)^2$
- memoryless property: $P[X \leq n + n_0 | X \geq n_0] = P[X \leq n]$

Examples of random variables

Mean and moments of order k

- $E[X] = \sum np(n)$, $E[X^k] = \sum n^k p(n)$
- $E[X] = \int x f_X(x) dx$, $E[X^k] = \int x^k f_X(x) dx$
- Variance: $V[X] = E[X^2] - E[X]^2$, standard deviation: $\sigma[X] = \sqrt{V[X]}$

Exponential law of parameter λ

- Continuous random variable with positive values
- $f_T(t) = \lambda e^{-\lambda t}$, $F_T(t) = 1 - e^{-\lambda t}$, for $t \geq 0$
- $E[T] = 1/\lambda$
- memoryless property: $P[T \leq t + t_0 | T \geq t_0] = P[T \leq t]$

Stochastic processes

Definition

- $\{X(t)\}_{t\in\mathbb{T}}$ stochastic process: at each time instant $t \in \mathbb{T}$, $X(t)$ is a random variable
- $\mathbb{T}$ is the time set: discrete or continuous
- $\mathbb{E}$ is the state set: discrete or continuous

Four important types of stochastic processes

	$\mathbb{T} \subset \mathbb{Z}$	$\mathbb{T} \subset \mathbb{R}$
$\mathbb{E} \subset \mathbb{Z}$	discrete time process with discrete state space or discrete time chain	continuous time process with discrete state space or continuous time chain
$\mathbb{E} \subset \mathbb{R}$	discrete time process with continuous state space	continuous time process with continuous state space

Counting process

- stochastic process with discrete state space such that:
- $x_0 = 0$ and $\forall n \in \mathbb{Z}$, $x_{n-1} \leq x_n$ (discrete time chain)
- $x(0) = 0$ and $\forall t < s$, $x(t) \leq x(s)$

Characterization of a stochastic process

- discrete time chain: $\forall n$, $P[X_n = i_n | X_{n-1} = i_{n-1}, ..., X_0 = i_0]$
- continuous time chain: $\forall n$, $P[X(t_n) = i_n | X(t_{n-1}) = i_{n-1}, ..., X(t_0) = i_0]$, $\forall t_n > ... > t_0$

Independent increments

- discrete time chain: $\forall n$, $X_n - X_{n-1}$,...,$X_1 - X_0$ are independent
- continuous time chain: $\forall n$ and $\forall t_n > ... > t_0$, $X(t_n) - X(t_{n-1})$,...,$X(t_1) - X(t_0)$ are independent

Stationary process

- discrete time chain: $\forall n$, $X_n - X_{n-1}$ and $X_1 - X_0$ have the same distribution
- continuous time chain: $\forall s, t$, $X(s+t) - X(s)$ and $X(t) - X(0)$ have the same distribution

A stationary stochastic process with independent increments is characterized by:

- discrete time chain: $X_{n+1} - X_n$
- continuous time chain: $\forall t$, $X(t) - X(0)$

Poisson process

Definition

- a continuous chain N is a Poisson process with parameter λ iff
- N is a counting process
- N is a stationary process with independent increments
- $\forall s, t$ and $k \in \mathbb{N}$, $P[N(s+t) - N(s) = k] = (\lambda t)^k e^{-\lambda t}/k!$

Properties

- $\forall t$, $N(t)$ follows a Poisson law of parameter λt
- for $k \geq k_0$, $P[N(t) = k | N(t_0) = k_0] = [\lambda(t - t_0)]^{k-k_0} e^{-\lambda(t-t_0)}/(k - k_0)!$

Poisson arrivals: the probability that a client arrives in dt is approx. λdt

$$
\begin{aligned}
P[N(t+dt) = k+j | N(t) = k] &= \lambda dt + o(dt) \text{ if } j = 1 \\
&= o(dt) \text{ if } j > 1 \\
&= 1 - \lambda dt + o(dt) \text{ if } j = 0
\end{aligned}
$$

Poisson arrivals: exponential inter-arrivals Properties

- the probabilistic decomposition of a Poisson process gives a set of Poisson processes
- the superposition of independent Poisson processes gives a Poisson process

Transformations

Moment generator function

- $\Psi_X(\theta) = E[e^{\theta X}]$
- $E[X^n] = \frac{d^n \Psi_X(\theta)}{d\theta^n}|_{\theta=0}$
- if X and Y are independent, $\Psi_{X+Y} = \Psi_X \Psi_Y$

z transformation (X with positive values)

- $F(z) = \Psi(\ln z) = \sum_{n=0}^{\infty} p(n) z^n$, for $|z| < 1$
- $E[X] = \frac{dF(z)}{dz}|_{z=1}$

Moment generator function

- $\Psi_X(\theta) = E[e^{\theta X}]$
- $E[X^n] = \frac{d^n \Psi_X(\theta)}{d\theta^n}|_{\theta=0}$
- if X and Y are independent, $\Psi_{X+Y} = \Psi_X \Psi_Y$

Laplace transformation (X with positive values)

- $F^*(s) = \Psi_X(-s) = \int_0^\infty e^{-st} f_X(t)dt$
- $E[X^n] = (-1)^n \frac{d^n F^*(s)}{ds^n}|_{s=0}$
- if X and Y are independent, $F^*_{X+Y} = F^*_X F^*_Y$

1. Discrete time Markov chains
   (a) Definitions
   (b) Graph representation
   (c) Transient regime analysis
   (d) Stationary (permanent) regime analysis
2. Continuous time Markov chains
   (a) Definition
   (b) Characterizations
   (c) Transient regime analysis
   (d) Stationary (permanent) regime analysis

Definition

- $\{X_n\}_{n\in\mathbb{N}}$ stochastic process with discrete state space and discrete time
- state space can be finite or infinite (denumerable)
- $\{X_n\}_{n\in\mathbb{N}}$ is a discrete time Markov chain iff

$$P[X_n = j | X_{n-1} = i_{n-1}, ..., X_0 = i_0] = P[X_n = j | X_{n-1} = i_{n-1}]$$

$$\forall i_0, .., i_n, j \in E$$

Homogeneous DTMC

- these probabilities does not depend on n
- transition probability: $\forall n \in \mathbb{N}$, $p_{ij} = P[X_n = j | X_{n-1} = i]$
- $\sum_{j \in E} p_{ij} = 1$

Transition matrix

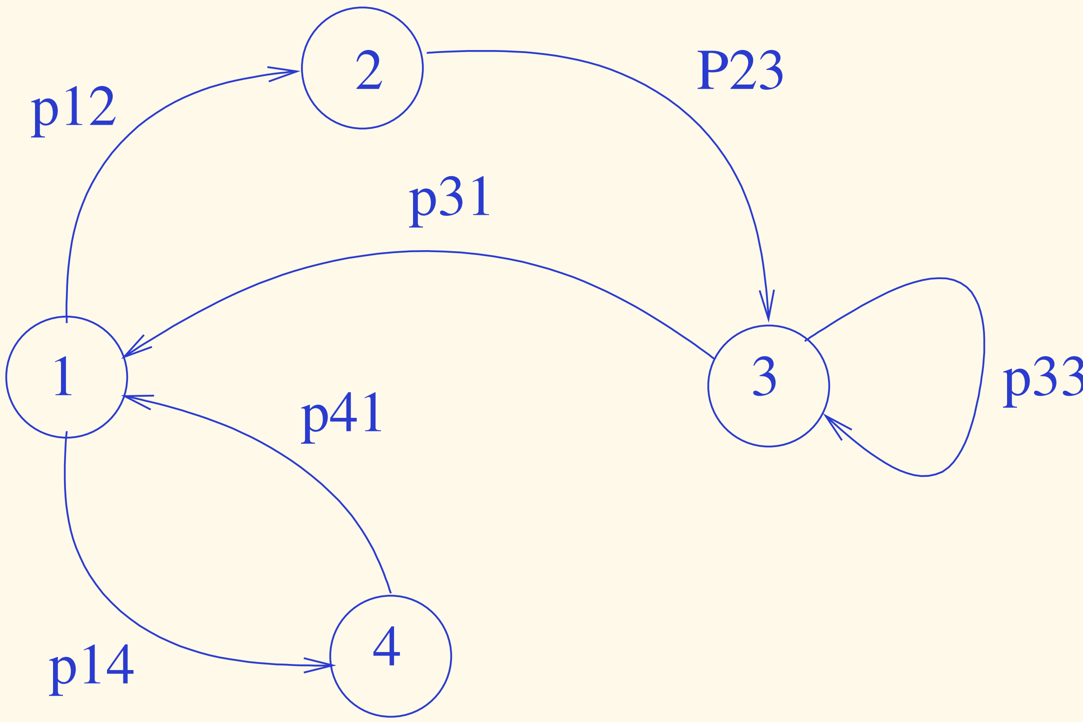
$$P = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1j} & \dots \\ p_{21} & p_{22} & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ p_{i1} & \dots & \dots & p_{ij} & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

DTMC graph representation

- Oriented graph
- state of the chain = graph node
- transitions = vertices associated to the transition probability

DTMC example

- $E = \{1, 2, 3, 4\}$
- $p_{23} = p_{41} = 1$, $p_{12} + p_{14} = 1$, and $p_{31} + p_{33} = 1$
- transition matrix ?



Time instants models

- regular time instants: $t_n = nT$
- just after an event: t_n is the instant of the n^{th} event

Transient regime

- find the vector $\pi^{(n)}$ of the state probabilities
- $\pi_j^{(n)} = P[X_n = j]$, $j \in E$
- $\pi^{(n)} = [\pi_1^{(n)}, \pi_2^{(n)}, ...]$
- this vector depends on P and on $\pi^{(0)}$

Total probabilities theorem

- $\pi_j^{(n)} = P[X_n = j] = \sum_{i \in E} P[X_n = j | X_{n-1} = i] P[X_{n-1} = i] = \sum_{i \in E} \pi_i^{(n-1)} p_{ij}$
- $\pi^{(n)} = \pi^{(n-1)} P$
- $\pi^{(n)} = \pi^{(0)} P^n$

Transitions in m steps

- definition: $p_{ij}^{(m)} = P[X_{n+m} = j | X_n = i]$
- definition of the transition matrix in m steps: $P^{(m)} = [p_{ij}^{(m)}]_{i,j \in E}$
- total probabilities theorem:

$$\begin{aligned} p_{ij}^{(m)} &= P[X_{n+m} = j | X_n = i] \\ &= \sum_{k \in E} P[X_{n+m} = j | X_n = i \text{ and } X_{n+m-1} = k] P[X_{n+m-1} = k | X_n = i] \\ &= \sum_{k \in E} p_{ik}^{(m-1)} p_{kj} \end{aligned}$$

- $P^{(m)} = P^{(m-1)} P$
- $P^{(m)} = P^m$

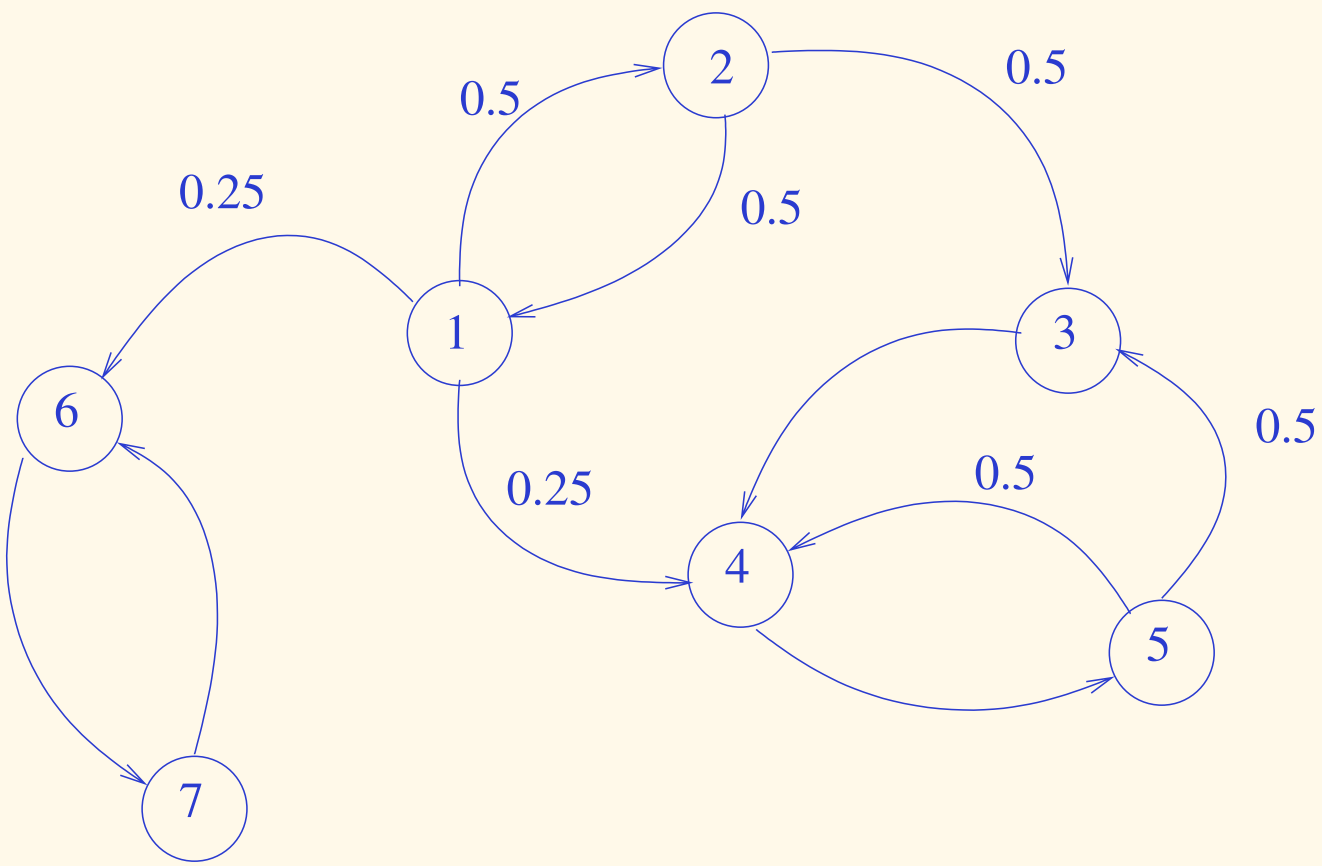
Distribution of the journey time in a state

- the number of steps in a given state j of the DTMC follows a geometrical law ($p_{jj} \neq 0$)

State classification

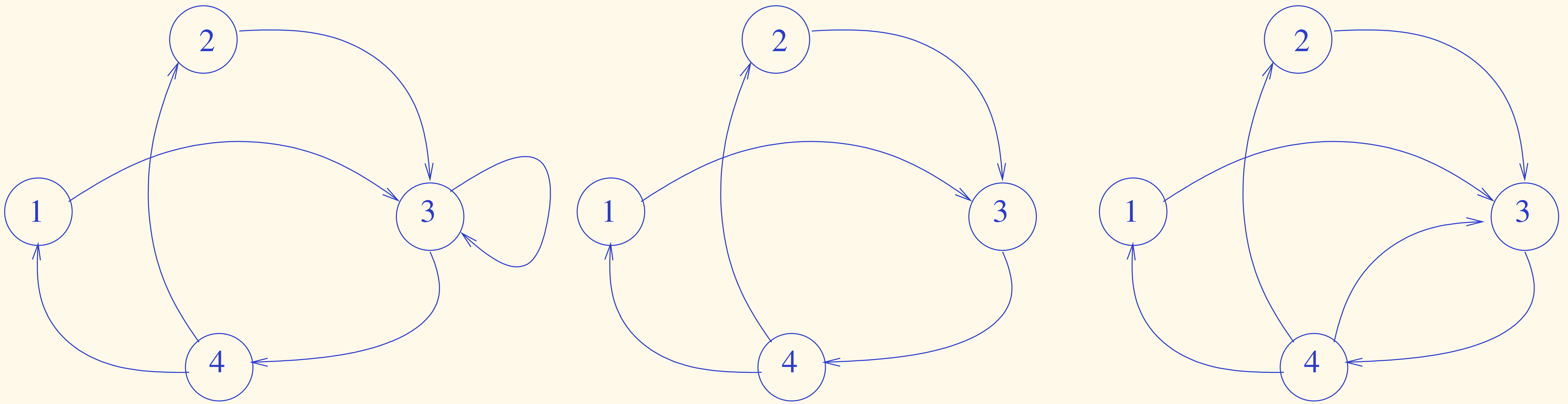
- a DTMC is irreducible iff from any state i, we can reach each state j in a finite number of steps

- $\forall i, j \in E$, there is a m such that $p_{ij}^{(m)} \neq 0$



Periodic states

- there is $k > 1$, such that $p_{jj}^{(m)} = 0$ for m non multiple of k
- the period of state j is the greatest integer k that verifies this definition
- the period of a DTMC is the greatest common divisor of the periods of its states
- the period of a DTMC is the greatest common divisor of the the length of all paths in the chain graph



Definitions

- $f_{jj}^{(n)}$: probability that the first come back to state j occurs after n steps
- f_{jj}: probability to come back in state j: $f_{jj} = \sum_{n=1}^{\infty} f_{jj}^{(n)}$
- M_j: mean return time in state j: $M_j = \sum_{n=1}^{\infty} n f_{jj}^{(n)}$

Recurrent and transient states

- state j is transient iff $f_{jj} < 1$
- state j is recurrent iff $f_{jj} = 1$
- state j is null recurrent if the mean return time is infinite: $M_j = \infty$
- state j is non null recurrent if the mean return time is finite: $M_j < \infty$

Properties

- all states of an irreducible DTMC are of the same nature: transients, null recurrents, or non null recurrents
- if all states are periodic, they have all the same period
- all states of a finite irreducible DTMC are non null recurrent

Performance parameters

- $f_{ij}^{(n)}$: probability to go from i to j in exactly n steps
- $f_{ij}^{(1)} = p_{ij}$ and $f_{ij}^{(n)} = \sum_{k \neq j} p_{ik} f_{kj}^{(n-1)}$ for $n > 2$
- f_{ij}: probability to go from i to j in any number of steps: $f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}$
- $f_{ij} = p_{ij} + \sum_{k \neq j} p_{ik} f_{kj}$
- R_{ij}: mean number of paths through state j given that we started from i
- $R_{ij} = \sum_{n=1}^{\infty} nP[n \text{ steps in } j | \text{ initial state} = i]$
- $R_{ij} = \frac{f_{ij}}{1 - f_{jj}}$

Stationary regime

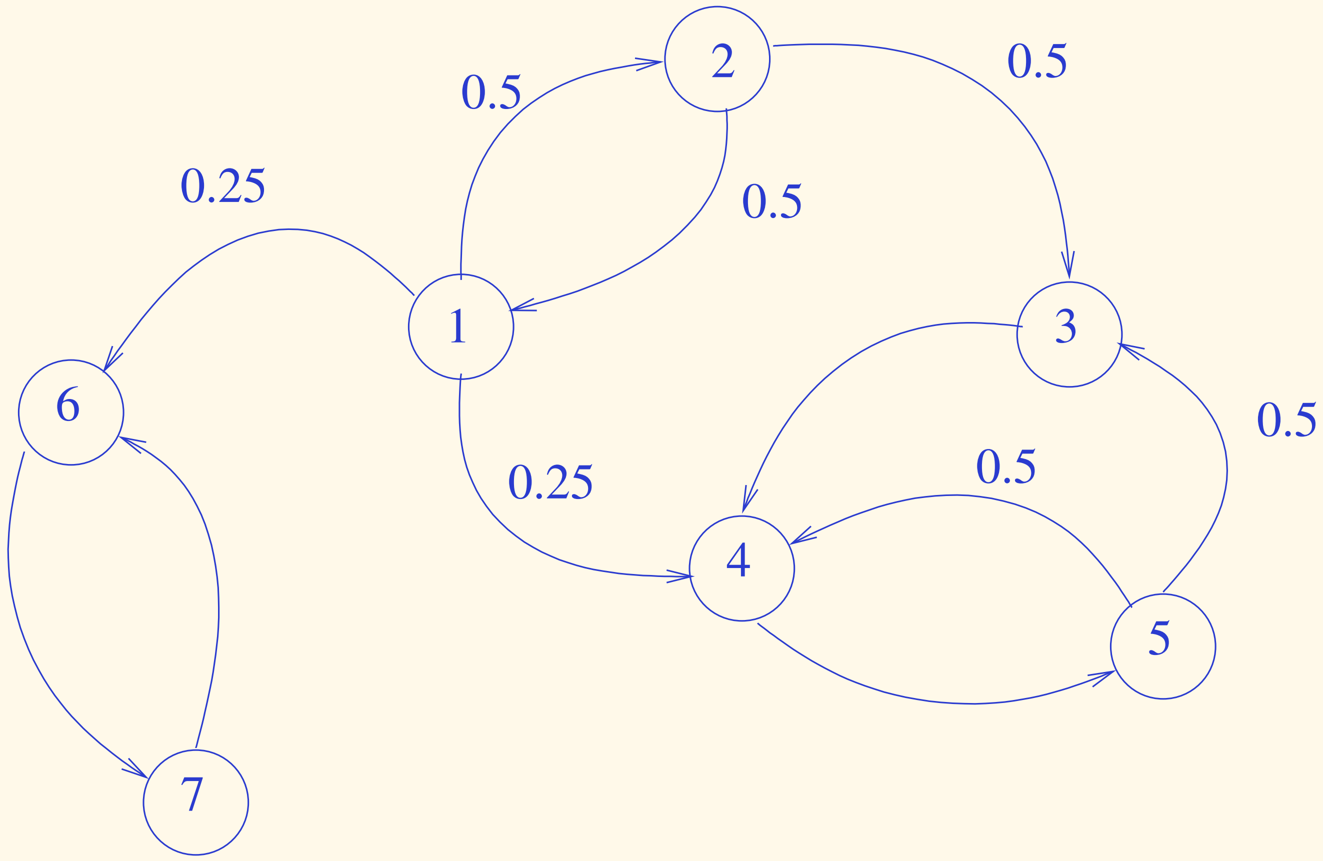
- we are interested in the limit of vector $\pi^{(n)}$ when n tends toward infinity
- in an irreducible and aperiodic DTMC, the limit vector π such that $\pi_j = \lim_{n\to\infty} \pi_j^{(n)}$ always exists and is independent on the initial probability vector $\pi^{(0)}$
- if states are transient or null recurrent, $\forall j \in E$, $\pi_j = 0$
- if states are non null recurrent, π_j are solutions of the following equations:
- $\forall j \in E$, $\pi_j = \sum_{i\in E} \pi_i p_{ij}$, i.e., $\pi = \pi P$
- and $\sum_{i\in E} \pi_i = 1$

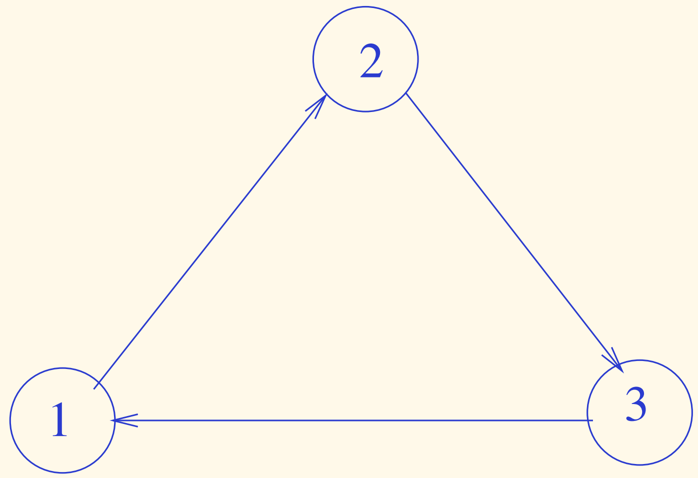
Discrete time Markov chains - Stationary regime analysis

Remarks

- in a finite non irreducible DTMC, we are interested in the probability to fall in an absorbent irreducible sub-chain and in the stationary probability given that we are in a sub-chain

- when a finite and irreducible DTMC is periodic, there is no limit vector. However, the system $p = pP$, $\sum_i p_i = 1$ has a solution. p_i is the proportion of time spent in state i





Definition

- let $\{X(t)\}_{t\geq 0}$ be a stochastic process with discrete state space and continuous time
- the state space E can be finite or infinite (denumerable)
- $\{X(t)\}_{t\geq 0}$ is a CTMC iff $\forall n$, $\forall t_0 < ... < t_n$

$$P[X(t_n) = j | X(t_{n-1}) = i_{n-1}, ..., X(t_0) = i_0] = P[X(t_n) = j | X(t_{n-1}) = i_{n-1}]$$

Homogeneous CTMC

- these probabilities does not depend on the observation points t_n and t_{n-1} but only on the duration $t_n - t_{n-1}$ between two instants

Transition probability from i to j in t

$$\forall s \geq 0 \; p_{ij}(t) = P[X(s+t) = j | X(s) = i]$$

Property

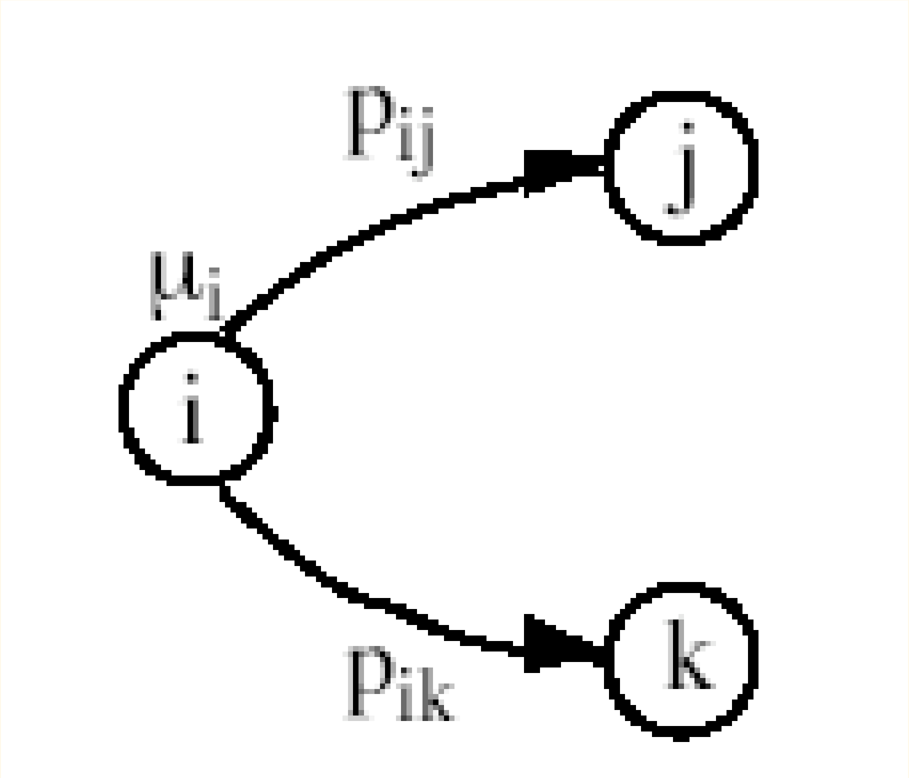
- the time spent in a state of a CTMC has an exponential distribution
- we associate to each state an exponential law T_i of rate μ_i

Evolution of a CTMC

- we stay an exponential time in a state, then we choose another state as next step
- destination doesn't depend neither on the time spent in the state, nor on the path to arrive in this state

First characterization of a CTMC

- a CTMC is a stochastic process with discrete state space and continuous time such that:
- the time spent in a state i has an exponential distribution of rate μ_i and
- the transitions from state i to other states are probabilistic



Transition time between two states

- probability to go from i to j in dt

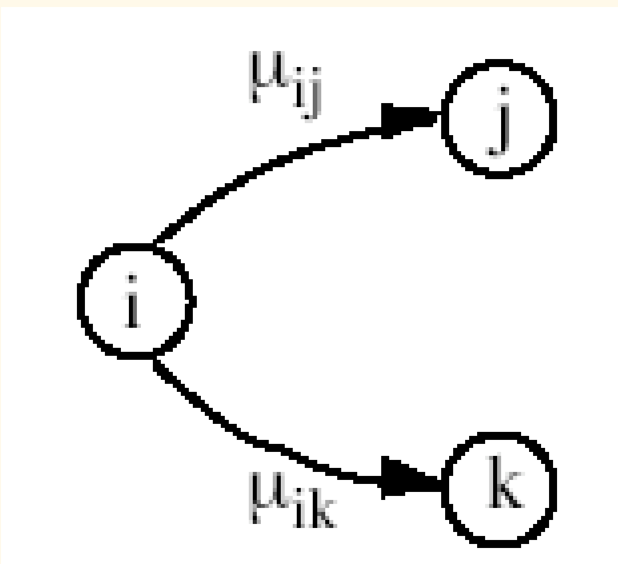
$$\begin{aligned} p_{ij}(dt) &= P[X(t+dt)=j|X(t)=i] \\ &= p_{ij}P_i[T_i \leq t+dt|T_i > t] \\ &= p_{ij}P_i[T_i \leq dt] \\ &= p_{ij}[1-e^{-\mu_i dt}] \\ &= p_{ij}(1-(1-\mu_i dt + o(dt))) \\ &= p_{ij}(\mu_i dt + o(dt)) \end{aligned}$$

- transition time between i and j has an exponential distribution of rate

$$\mu_{ij} = \mu_i p_{ij}$$

Second characterization

- starting from a state i, the random variable measuring the transition time to state j has an exponential law of rate μ_{ij}

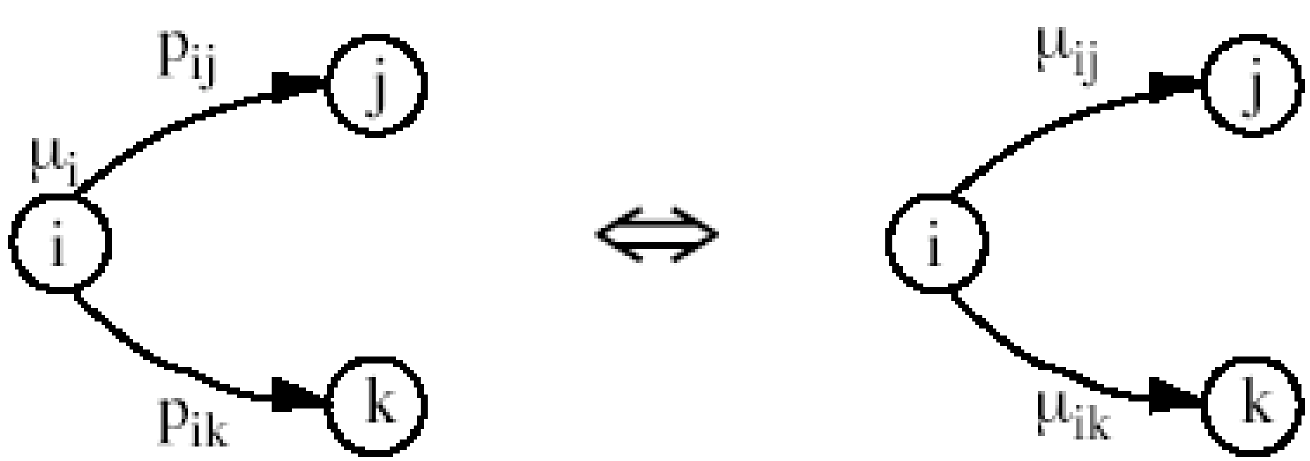


Equivalence between the two characterizations

$$\mu_i = \mu_{ij} + \mu_{ik}$$

$$p_{ij} = \frac{\mu_{ij}}{\mu_{ij} + \mu_{ik}}$$

$$p_{ik} = \frac{\mu_{ik}}{\mu_{ij} + \mu_{ik}}$$



$$\mu_{ij} = \mu_i \, p_{ij}$$

$$\mu_{ik} = \mu_i \, p_{ik}$$

DTMC vs. CTMC

- DTMC : transition matrix
- CTMC : infinitesimal generator

$$\forall i \neq j \; q_{ij} = \mu_{ij}$$

$$q_{ii} = -\sum_{i \neq j} \mu_{ij} = -\mu_i$$

$$Q = \begin{pmatrix} -\sum_{1 \neq j} \mu_{1j} & \mu_{12} & \dots & \mu_{1j} & \dots \\ \mu_{21} & -\sum_{2 \neq j} \mu_{2j} & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \mu_{i1} & \dots & \dots & \mu_{ij} & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

Transient regime

- find the vector $\pi(t)$ at each instant of the processus

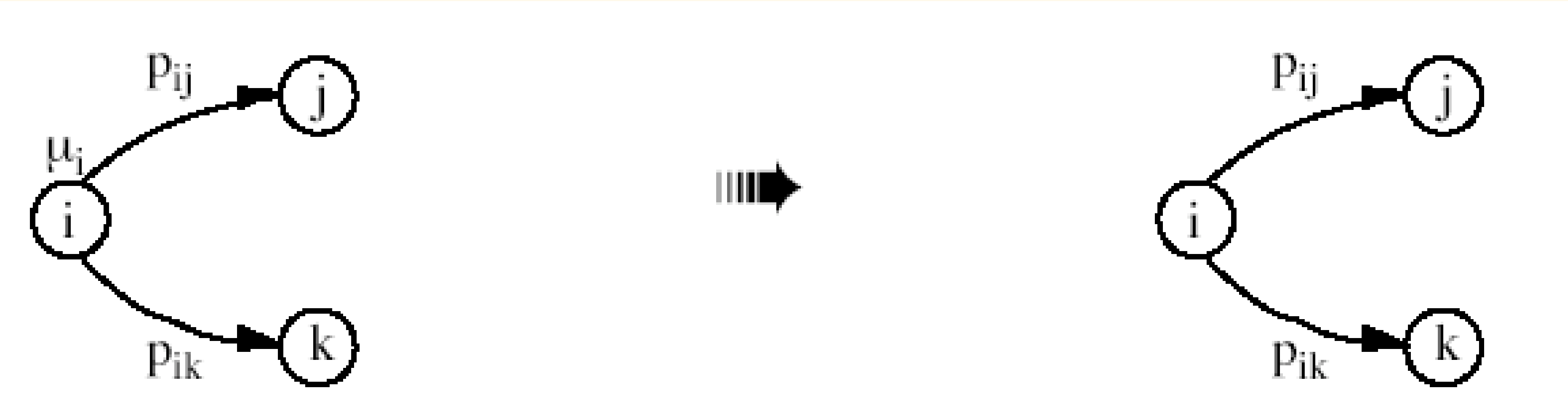
$$\pi(t) = [\pi_1(t), \pi_2(t), ...]$$

- this vector depends on Q and $\pi(0)$

$$\begin{aligned}
\pi_j(t+dt) &= P[X(t+dt)=j] \\
&= \sum_{i \in E} P[X(t+dt)=j | X(t)=i] P[X(t)=i] \\
&= \pi_j(t) P[X(t+dt)=j|X(t)=j] + \sum_{i \neq j} \pi_i(t) P[X(t+dt)=j|X(t)=i] \\
&= \pi_j(t) \left[1 - \sum_{i \neq j} \mu_{ji} dt\right] + \sum_{i \neq j} \pi_i(t) \mu_{ij} dt + o(dt) \\
\frac{d\pi_j(t)}{dt} &= \sum_{i \in E} \pi_i(t) q_{ij} \\
\frac{d\pi(t)}{dt} &= \pi(t) Q \\
\pi(t) &= \pi(0) e^{Qt}
\end{aligned}$$

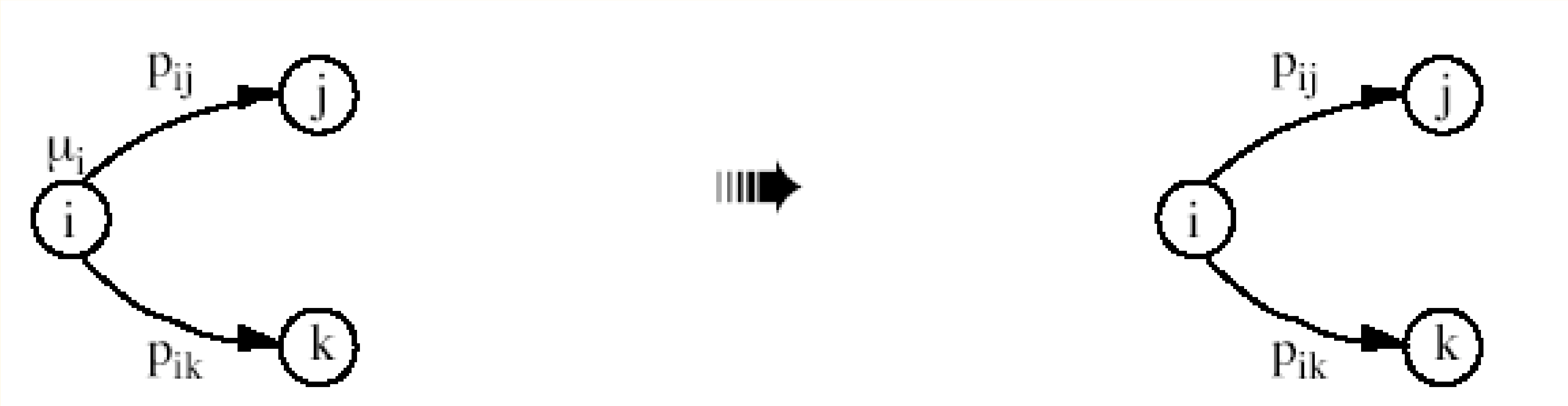
Stationary regime

- stationary solution w/ a DTMC: irreducible and aperiodic chain
- CTMC: no periodicity concept ! states are either transient or recurrent
- DTMC included in a CTMC:



Property

- a CTMC is irreducible iff the included DTMC is irreducible



- state i is transient in a CTMC iff i is transient in the included DTMC
- state i is recurrent in a CTMC iff i is recurrent in the included DTMC
- all states of an irreducible CTMC have the same nature (transient or recurrent)
- if moreover, E is finite, all states are non null recurrent

Condition of existence of the stationary regime

- in an irreducible CTMC, the limit vector $\pi = \lim_{t \to \infty} \pi(t)$ always exists
- it is independent on the initial probabilities $\pi(0)$
- either all states are transient or null recurrent and $\pi_j = 0$ for all $j \in E$
- or all states are non null recurrent and π is solution of

$$\begin{aligned} \pi Q &= 0 \\ \sum_{i \in E} \pi_i &= 1 \end{aligned}$$

State equations :

$$\begin{aligned}\sum_{i\in E} \pi_i q_{ij} &= 0\\ \sum_{i\neq j} \pi_i q_{ij} &= -\pi_j q_{jj}\\ \sum_{i\neq j} \pi_i q_{ij} &= \sum_{i\neq j} \pi_j q_{ji}\end{aligned}$$

Interpretation: outgoing flow from j = arriving flow in j
Border equation :

$$\sum_{i\in E_1}\sum_{j\in E_2} \pi_i \mu_{ij} = \sum_{j\in E_2}\sum_{i\in E_1} \pi_j \mu_{ji}$$

Interpretation: flow from E_1 to E_2 = flow from E_2 to E_1