Some probabilistic aspects of FCA II

Richard Emilion (University of Orléans, France)

DATA ANALYSIS AND MODELING LAB

Palacky University, Olomouc, Czech Republic

european social fund in the czech republic

EUROPEAN UNION

MINISTRY OF EDUCATION, YOUTH AND SPORTS

OP Education for Competitiveness

INVESTMENTS IN EDUCATION DEVELOPMENT

II.6 Variance of $|\mathcal{L}|$ in the Bernoulli model case 17 / 51

- Computation of Prob(A and B be closed), A, $B \in \mathcal{P}(J)$

Instead of having just 3 cases, namely $O \times A$, $I - O \times A$, $O \times J - A$, it appears 16 cases. Some formulas in (Emilion-Lévy can be simplified).

- Taking expectation yields $\mathbb{E}(|L|^2)$ and therefore $var(|L|) = \mathbb{E}(|L|^2) - (\mathbb{E}(|L|))^2$

II.6 Variance of $|\mathcal{L}|$ in the Bernoulli model case

- Computation of Prob(A and B be closed), A, $B \in \mathcal{P}(J)$

Instead of having just 3 cases, namely $O \times A$, $I - O \times A$, $O \times J - A$, it appears 16 cases. Some formulas in (Emilion-Lévy can be simplified).

- Taking expectation yields $\mathbb{E}(|L|^2)$ and therefore $var(|L|) = \mathbb{E}(|L|^2) - (\mathbb{E}(|L|))^2$

II.6 Variance of $|\mathcal{L}|$ in the Bernoulli model case

- Computation of Prob(A and B be closed), A, $B \in \mathcal{P}(J)$

Instead of having just 3 cases, namely $O \times A$, $I - O \times A$, $O \times J - A$, it appears 16 cases. Some formulas in (Emilion-Lévy can be simplified).

- Taking expectation yields $\mathbb{E}(|L|^2)$ and therefore $var(|L|) = \mathbb{E}(|L|^2) - (\mathbb{E}(|L|))^2$

II.7 μ, σ exact values in the Bernoulli model case

m	n	p	μ	σ	95% CI for L
14	10	0.3	32.48	6.47	$[3, 62]$
15	15	0.9	489.47	373.74	$[1, 2161]$
20	15	0.25	62.78	11.09	$[13, 113]$
20	20	0.65	1945.49	469.16	$[1, 4044]$
25	15	0.85	3758.31	1625.93	$[1, 11030]$
30	12	0.85	1598.66	538.70	$[1, 4008]$

II.8 Bernoulli context, CL size Expectation

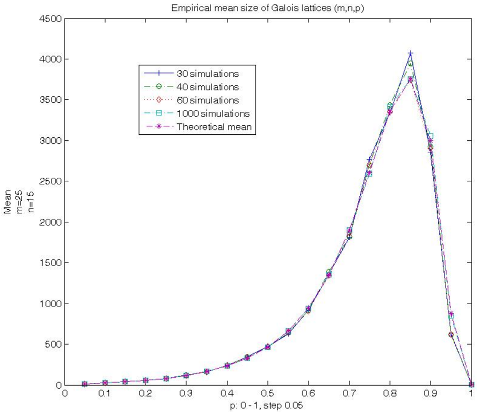

Figure: Estimated and Exact Mean size of Bernoulli Concept Lattices

II.9 Experiments for σ in the Bernoulli model case

m	n	p	σ	S_{300}	95% CI	S_{1000}
14	10	0.3	6.47	5.94	5.03 - 7.94	6.40
15	15	0.9	373.74	321.43	284.39 - 386.57	370.6
20	15	0.25	11.09	11.14	8.92 - 12.96	11.04
20	20	0.65	469.16	469.65	433.60 - 497.42	468.25
25	15	0.85	1625.93	1688.60	1493.90 - 1743.60	1626.20
30	12	0.85	538.70	549.30	503.96 - 566.11	535.79

R.E., *Selected contributions in Data Analysis and Classification*, 247-259, Springer, 2007

Context: $m \times r$ random binary matrix $\mathcal{C}$
U a latent class variable $\in \{1, \ldots, K\}$ over the individuals

$$\begin{cases} q = (q_1, \ldots, q_K) & \sim & Dirichlet(\gamma_1, \ldots, \gamma_K) \\ U \in \{1, \ldots, K\} : P(U = u|q) & = & q_u \\ \mathcal{C}|_{U=u,q} & \sim & \otimes_{j=1}^{r} B(p_{u,j}) \\ \mathcal{C}|_q & \sim & \sum_{u=1}^{K} q_u \otimes_{j=1}^{r} B(p_{u,j}) \end{cases}$$

Y. W. Teh, D. Gorur, Z. Ghahramani
$Beta - Bernoulli$ Context: $m \times r$ random binary matrix $\mathcal{C}$

$$\begin{cases} p_1, \dots, p_r & \overset{i.i.d.}{\sim} & Beta(\frac{\alpha}{r}, 1) \\ \mathcal{C}_{ij} | p_1, \dots, p_r & \overset{ind}{\sim} & Bernoulli(p_j) \end{cases}$$

Limit:
Step 1: Customer 1 chooses $K^{(1)}$ different items, where $K^{(1)} \sim Poisson(\alpha)$
Step 2: Customer 2 arrives and chooses to enjoy each of the items already chosen with probability $1/2$. In addition, he chooses $K^{(2)}$ new items , where $K^{(2)} \sim Poisson(\alpha/2)$
Steps 3 through N: The ith customer arrives and chooses to enjoy each of the items already chosen with probability m_{ki}/i, where m_{ki} is the number of customers who have chosen the kth item before the ith customer. In addition, the ith customer chooses $K^{(i)} \sim Poisson(\alpha/i)$ new items.

II.12 Latent Block model

G. Govaert, M. Nadif, Bi-clustering.
Context: $m \times r$ random binary matrix $\mathcal{C}$
$\mathcal{Z}$ set of partitions of I into g subsets
$\mathcal{W}$ set of partitions of J into h subsets

$$f(\mathcal{C};\theta) = \sum_{(z,w)\in\mathcal{Z}\times\mathcal{W}} p(z;\theta)p(w;\theta) \prod_{i,j,k,l} Bernoulli(c_{i,j};\alpha_{k,l})^{z_{i,k}w_{j,l}}$$

III - Sampling

Sampling in a large set

Markov Chains in $\mathcal{L}$

Sampling and Counting concepts

III.1 Sampling in a large set

- Selecting an element at random on a large (but finite) set, e.g., $\mathcal{L}$
- At random ? Given a probability measure Q on $\mathcal{L}$, propose an algorithm X which outputs are elements of $\mathcal{L}$ and such that $Prob(X = l) = Q\{l\} = q_l$ for any $l \in \mathcal{L}$
- When Q is uniform, i.e. $q_l = \frac{1}{|\mathcal{L}|}$: sampling at random, in common language .
- Problems : L is very large, listing L is tedious, $|\mathcal{L}|$ is **unknown**
- More general problem : $Prob(X = l) \propto v(l)$ a function of l which no need to sum up to 1.

III.1 Sampling in a large set

- Selecting an element at random on a large (but finite) set, e.g., $\mathcal{L}$
- At random ? Given a probability measure Q on $\mathcal{L}$, propose an algorithm X which outputs are elements of $\mathcal{L}$ and such that $Prob(X = l) = Q\{l\} = q_l$ for any $l \in \mathcal{L}$
- When Q is uniform, i.e. $q_l = \frac{1}{|\mathcal{L}|}$: sampling at random, in common language .
- Problems : L is very large, listing L is tedious, $|\mathcal{L}|$ is **unknown**
- More general problem : $Prob(X = l) \propto v(l)$ a function of l which no need to sum up to 1.

III.1 Sampling in a large set

• Selecting an element at random on a large (but finite) set, e.g., $\mathcal{L}$

• At random ? Given a probability measure Q on $\mathcal{L}$, propose an algorithm X which outputs are elements of $\mathcal{L}$ and such that $Prob(X = l) = Q\{l\} = q_l$ for any $l \in \mathcal{L}$

• When Q is uniform, i.e. $q_l = \frac{1}{|\mathcal{L}|}$: sampling at random, in common language .

• Problems : L is very large, listing L is tedious, $|\mathcal{L}|$ is **unknown**

• More general problem : $Prob(X = l) \propto v(l)$ a function of l which no need to sum up to 1.

III.1 Sampling in a large set

• Selecting an element at random on a large (but finite) set, e.g., $\mathcal{L}$

• At random ? Given a probability measure Q on $\mathcal{L}$, propose an algorithm X which outputs are elements of $\mathcal{L}$ and such that $Prob(X = l) = Q\{l\} = q_l$ for any $l \in \mathcal{L}$

• When Q is uniform, i.e. $q_l = \frac{1}{|\mathcal{L}|}$: sampling at random, in common language .

• Problems : L is very large, listing L is tedious, $|\mathcal{L}|$ is **unknown**

• More general problem : $Prob(X = l) \propto v(l)$ a function of l which no need to sum up to 1.

III.1 Sampling in a large set

- Selecting an element at random on a large (but finite) set, e.g., $\mathcal{L}$

- At random ? Given a probability measure Q on $\mathcal{L}$, propose an algorithm X which outputs are elements of $\mathcal{L}$ and such that $Prob(X = l) = Q\{l\} = q_l$ for any $l \in \mathcal{L}$

- When Q is uniform, i.e. $q_l = \frac{1}{|\mathcal{L}|}$: sampling at random, in common language .

- Problems : L is very large, listing L is tedious, $|\mathcal{L}|$ is **unknown**

- More general problem : $Prob(X = l) \propto v(l)$ a function of l which no need to sum up to 1.

III.1 Sampling in a large set

• Selecting an element at random on a large (but finite) set, e.g., $\mathcal{L}$

• At random ? Given a probability measure Q on $\mathcal{L}$, propose an algorithm X which outputs are elements of $\mathcal{L}$ and such that $Prob(X = l) = Q\{l\} = q_l$ for any $l \in \mathcal{L}$

• When Q is uniform, i.e. $q_l = \frac{1}{|\mathcal{L}|}$: sampling at random, in common language .

• Problems : L is very large, listing L is tedious, $|\mathcal{L}|$ is **unknown**

• More general problem : $Prob(X = l) \; \propto \; v(l)$ a function of l which no need to sum up to 1.

- Classical Central limit Theorem (CLT) holds for i.i.d. r.v.s
- Andrey Markov, triying to generalizing CLT for non i.i.d. r.v.s, found his famous definition:
- $X_0, \ldots, X_n, \ldots : \Omega \to \mathcal{L}$ (the finite or countable *state space*)
- Let p be a Markov matrix $p(i,j)_{i,j \in \mathcal{L}}$: $p(i,j) \geq 0$ and $\sum_j p(i,j) = 1$
- The chain 'forgets' its past:

$$P(X_{n+1} = x_{n+1} | X_0 = x_0, \ldots, X_n = x_n) = P(X_{n+1} = x_{n+1} | X_n = x_n) = p(x_n, x_{n+1})$$

- $P(X_0 = x_0) = \nu(x_0)$ ν: initial distribution.
- We have $P(X_n = j | X_0 = i) = p^{(n)}(i,j)$
- Simulation Algorithm: draw x_0 from ν. Once x_n is drawn, draw $x_{n+1} =$ from $p(x_n, .)$
- Notations: $(\Omega, \mathbb{P}_\nu), (\Omega, \mathbb{P}_i)$

- Classical Central limit Theorem (CLT) holds for i.i.d. r.v.s
- Andrey Markov, triying to generalizing CLT for non i.i.d. r.v.s, found his famous definition:
- $X_0, \ldots, X_n, \ldots : \Omega \to \mathcal{L}$ (the finite or countable *state space*)
- Let p be a Markov matrix $p(i,j)_{i,j \in \mathcal{L}}$: $p(i,j) \geq 0$ and $\sum_j p(i,j) = 1$
- The chain 'forgets' its past:

$$P(X_{n+1} = x_{n+1} | X_0 = x_0, \ldots, X_n = x_n) = P(X_{n+1} = x_{n+1} | X_n = x_n) = p(x_n, x_{n+1})$$

- $P(X_0 = x_0) = \nu(x_0)$ ν: initial distribution.
- We have $P(X_n = j | X_0 = i) = p^{(n)}(i,j)$
- Simulation Algorithm: draw x_0 from ν. Once x_n is drawn, draw $x_{n+1} =$ from $p(x_n, .)$
- Notations: $(\Omega, \mathbb{P}_\nu), (\Omega, \mathbb{P}_i)$

• Classical Central limit Theorem (CLT) holds for i.i.d. r.v.s
• Andrey Markov, triying to generalizing CLT for non i.i.d. r.v.s, found his famous definition:
• $X_0, \ldots, X_n, \ldots : \Omega \to \mathcal{L}$ (the finite or countable *state space*)
• Let p be a Markov matrix $p(i,j)_{i,j \in \mathcal{L}}$: $p(i,j) \geq 0$ and $\sum_j p(i,j) = 1$
• The chain 'forgets' its past:

$$P(X_{n+1} = x_{n+1} | X_0 = x_0, \ldots, X_n = x_n) = P(X_{n+1} = x_{n+1} | X_n = x_n) = p(x_n, x_{n+1})$$

• $P(X_0 = x_0) = \nu(x_0)$ ν: initial distribution.
• We have $P(X_n = j | X_0 = i) = p^{(n)}(i,j)$
• Simulation Algorithm: draw x_0 from ν. Once x_n is drawn, draw $x_{n+1} =$ from $p(x_n, .)$
• Notations: $(\Omega, \mathbb{P}_\nu), (\Omega, \mathbb{P}_i)$

- Classical Central limit Theorem (CLT) holds for i.i.d. r.v.s
- Andrey Markov, triying to generalizing CLT for non i.i.d. r.v.s, found his famous definition:
- $X_0, \ldots, X_n, \ldots : \Omega \to \mathcal{L}$ (the finite or countable *state space*)
- Let p be a Markov matrix $p(i,j)_{i,j \in \mathcal{L}}$: $p(i,j) \geq 0$ and $\sum_j p(i,j) = 1$
- The chain 'forgets' its past:

$$P(X_{n+1} = x_{n+1} | X_0 = x_0, \ldots, X_n = x_n) = P(X_{n+1} = x_{n+1} | X_n = x_n) = p(x_n, x_{n+1})$$

- $P(X_0 = x_0) = \nu(x_0)$ ν: initial distribution.
- We have $P(X_n = j | X_0 = i) = p^{(n)}(i,j)$
- Simulation Algorithm: draw x_0 from ν. Once x_n is drawn, draw $x_{n+1} =$ from $p(x_n, .)$
- Notations: $(\Omega, \mathbb{P}_\nu), (\Omega, \mathbb{P}_i)$

III.2 Markov Chains

- Classical Central limit Theorem (CLT) holds for i.i.d. r.v.s
- Andrey Markov, triying to generalizing CLT for non i.i.d. r.v.s, found his famous definition:
- $X_0, \ldots, X_n, \ldots : \Omega \to \mathcal{L}$ (the finite or countable *state space*)
- Let p be a Markov matrix $p(i,j)_{i,j \in \mathcal{L}}$: $p(i,j) \geq 0$ and $\sum_j p(i,j) = 1$
- The chain 'forgets' its past:

$$P(X_{n+1} = x_{n+1} | X_0 = x_0, \ldots, X_n = x_n) = P(X_{n+1} = x_{n+1} | X_n = x_n) = p(x_n, x_{n+1})$$

- $P(X_0 = x_0) = \nu(x_0)$ ν: initial distribution.
- We have $P(X_n = j | X_0 = i) = p^{(n)}(i,j)$
- Simulation Algorithm: draw x_0 from ν. Once x_n is drawn, draw $x_{n+1} =$ from $p(x_n, .)$
- Notations: $(\Omega, \mathbb{P}_\nu), (\Omega, \mathbb{P}_i)$

- Classical Central limit Theorem (CLT) holds for i.i.d. r.v.s
- Andrey Markov, triying to generalizing CLT for non i.i.d. r.v.s, found his famous definition:
- $X_0, \dots, X_n, \dots : \Omega \to \mathcal{L}$ (the finite or countable *state space*)
- Let p be a Markov matrix $p(i,j)_{i,j\in\mathcal{L}}$: $p(i,j) \geq 0$ and $\sum_j p(i,j) = 1$
- The chain 'forgets' its past:

$$P(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) = P(X_{n+1} = x_{n+1} | X_n = x_n) = p(x_n, x_{n+1})$$

- $P(X_0 = x_0) = \nu(x_0)$ ν: initial distribution.
- We have $P(X_n = j | X_0 = i) = p^{(n)}(i,j)$
- Simulation Algorithm: draw x_0 from ν. Once x_n is drawn, draw $x_{n+1} =$ from $p(x_n, .)$
- Notations: $(\Omega, \mathbb{P}_\nu), (\Omega, \mathbb{P}_i)$

• Sampling with MC. Main idea: find a MC such that

$$lim_{n \to +\infty} P(X_n = l) = q_l$$

(if the limit exists : ergodicity, steady state)

• Problems:

- Theoretical proof of ergodicity
- From which n can we consider that the steady state is reached
- This n should not be too large (time consuming)
- Precision: Perfect sampling

III.3 Markov Chains and sampling

• Sampling with MC. Main idea: find a MC such that

$$lim_{n \to +\infty} P(X_n = l) = q_l$$

(if the limit exists : ergodicity, steady state)

• Problems:

- Theoretical proof of ergodicity
- From which n can we consider that the steady state is reached
- This n should not be too large (time consuming)
- Precision: Perfect sampling

• Sampling with MC. Main idea: find a MC such that

$$lim_{n \to +\infty} P(X_n = l) = q_l$$

(if the limit exists : ergodicity, steady state)

• Problems:

- Theoretical proof of ergodicity
- From which n can we consider that the steady state is reached
- This n should not be too large (time consuming)
- Precision: Perfect sampling

• Sampling with MC. Main idea: find a MC such that

$$lim_{n \to +\infty} P(X_n = l) = q_l$$

(if the limit exists : ergodicity, steady state)

• Problems:

- Theoretical proof of ergodicity
- From which n can we consider that the steady state is reached
- This n should not be too large (time consuming)
- Precision: Perfect sampling

III.3 Markov Chains and sampling

• Sampling with MC. Main idea: find a MC such that

$$lim_{n \to +\infty} P(X_n = l) = q_l$$

(if the limit exists : ergodicity, steady state)

• Problems:

- Theoretical proof of ergodicity
- From which n can we consider that the steady state is reached
- This n should not be too large (time consuming)
- Precision: Perfect sampling

• Sampling with MC. Main idea: find a MC such that

$$lim_{n \to +\infty} P(X_n = l) = q_l$$

(if the limit exists : ergodicity, steady state)

• Problems:

- Theoretical proof of ergodicity
- From which n can we consider that the steady state is reached
- This n should not be too large (time consuming)
- Precision: Perfect sampling

- *Stopping Time* $T : \Omega \longrightarrow \mathbb{N} \cup \{+\infty\}$ such that $\{T = k\} \in \sigma(X_0, \dots X_k)$ for any k
- $\sigma(X_0, \dots X_k)$: all the information before time k
- Ex: $T_i = \inf\{n \geq 1 : X_n = i\} \quad (\inf \emptyset = +\infty)$
- Strong Markov property: Let $Y_n(\omega) := X_{T(\omega)+n}(\omega)$ if $T(\omega) < +\infty$ and $Y_n(\omega) := X_n(\omega)$ otherwise, then conditionnally to $\{T < +\infty\}$, Y_n is a Markov chain with transition matrix p

III.4 Stopping Time

- *Stopping Time* $T : \Omega \longrightarrow \mathbb{N} \cup \{+\infty\}$ such that $\{T = k\} \in \sigma(X_0, \dots X_k)$ for any k
- $\sigma(X_0, \dots X_k)$: all the information before time k
- Ex: $T_i = \inf\{n \geq 1 : X_n = i\} \quad (\inf \emptyset = +\infty)$
- Strong Markov property: Let $Y_n(\omega) := X_{T(\omega)+n}(\omega)$ if $T(\omega) < +\infty$ and $Y_n(\omega) := X_n(\omega)$ otherwise, then conditionnally to $\{T < +\infty\}$, Y_n is a Markov chain with transition matrix p

• *Stopping Time* $T : \Omega \longrightarrow \mathbb{N} \cup \{+\infty\}$ such that $\{T = k\} \in \sigma(X_0, \dots X_k)$ for any k
• $\sigma(X_0, \dots X_k)$: all the information before time k
• Ex: $T_i = \inf\{n \geq 1 : X_n = i\} \quad (\inf \emptyset = +\infty)$
• Strong Markov property: Let $Y_n(\omega) := X_{T(\omega)+n}(\omega)$ if $T(\omega) < +\infty$ and $Y_n(\omega) := X_n(\omega)$ otherwise, then conditionnally to $\{T < +\infty\}$, Y_n is a Markov chain with transition matrix p

- Oriented Graph G: $(i,j) \in G$ if $p(i,j) > 0$
- $i \rightarrow j$: j is reachable from i : $\mathbb{P}_i(T_j < +\infty) = 1$, i.e. $\exists n \geq 1 : p^{(n)}(i,j) > 0$
- $i \sim j$: i communicates with j iff both $i \rightarrow j$ and $j \rightarrow i$: equivalence relation, comunicating classes
- *Irreducible chain* : single communicating class

- *Period* of state $i = gcd\{n \geq 1 : p^{(n)}(i,i) > 0\}$, $gcd(\emptyset) = +\infty$. gcd(lengths of circuits)
- i is said *aperiodic* if its period is 1. *Proposition*: $\exists r : \forall n \geq r \; p^{(n)}(i,i) > 0$.
- The chain is said *aperiodic* if all its states are aperiodic
- *Proposition*: All the states of a communicating class have same period

III.5 Communicating classes, Periodicity

- Oriented Graph G: $(i,j) \in G$ if $p(i,j) > 0$
- $i \to j$: j is reachable from $i : \mathbb{P}_i(T_j < +\infty) = 1$, i.e. $\exists n \geq 1 : p^{(n)}(i,j) > 0$
- $i \sim j : i$ communicates with j iff both $i \to j$ and $j \to i$: equivalence relation, comunicating classes
- *Irreducible chain* : single communicating class
- *Period* of state $i = gcd\{n \geq 1 : p^{(n)}(i,i) > 0\}$, $gcd(\emptyset) = +\infty$. gcd(lengths of circuits)
- i is said *aperiodic* if its period is 1. *Proposition*: $\exists r : \forall n \geq r \; p^{(n)}(i,i) > 0$.
- The chain is said *aperiodic* if all its states are aperiodic
- *Proposition*: All the states of a communicating class have same period

III.5 Communicating classes, Periodicity

- Oriented Graph G: $(i,j) \in G$ if $p(i,j) > 0$
- $i \to j$: j is reachable from $i : \mathbb{P}_i(T_j < +\infty) = 1$, i.e. $\exists n \geq 1 : p^{(n)}(i,j) > 0$
- $i \sim j : i$ communicates with j iff both $i \to j$ and $j \to i$: equivalence relation, comunicating classes
- *Irreducible chain* : single communicating class
- *Period* of state $i = gcd\{n \geq 1 : p^{(n)}(i,i) > 0\}$, $gcd(\emptyset) = +\infty$. gcd(lengths of circuits)
- i is said *aperiodic* if its period is 1. *Proposition*: $\exists r : \forall n \geq r \; p^{(n)}(i,i) > 0$.
- The chain is said *aperiodic* if all its states are aperiodic
- *Proposition*: All the states of a communicating class have same period

III.5 Communicating classes, Periodicity 29 / 51

- Oriented Graph G: $(i,j) \in G$ if $p(i,j) > 0$
- $i \to j$: j is reachable from $i : \mathbb{P}_i(T_j < +\infty) = 1$, i.e. $\exists n \geq 1 : p^{(n)}(i,j) > 0$
- $i \sim j : i$ communicates with j iff both $i \to j$ and $j \to i$: equivalence relation, comunicating classes
- *Irreducible chain* : single communicating class

- *Period* of state $i = gcd\{n \geq 1 : p^{(n)}(i,i) > 0\}$, $gcd(\emptyset) = +\infty$. gcd(lengths of circuits)
- i is said *aperiodic* if its period is 1. *Proposition*: $\exists r : \forall n \geq r \; p^{(n)}(i,i) > 0$.
- The chain is said *aperiodic* if all its states are aperiodic
- *Proposition*: All the states of a communicating class have same period

III.5 Communicating classes, Periodicity

- Oriented Graph G: $(i,j) \in G$ if $p(i,j) > 0$
- $i \to j$: j is reachable from i : $\mathbb{P}_i(T_j < +\infty) = 1$, i.e. $\exists n \geq 1 : p^{(n)}(i,j) > 0$
- $i \sim j$: i communicates with j iff both $i \to j$ and $j \to i$: equivalence relation, comunicating classes
- *Irreducible chain* : single communicating class

- *Period* of state $i = gcd\{n \geq 1 : p^{(n)}(i,i) > 0\}$, $gcd(\emptyset) = +\infty$. gcd(lengths of circuits)
- i is said *aperiodic* if its period is 1. *Proposition*: $\exists r : \forall n \geq r\ p^{(n)}(i,i) > 0$.
- The chain is said *aperiodic* if all its states are aperiodic
- *Proposition*: All the states of a communicating class have same period

- Oriented Graph G: $(i,j) \in G$ if $p(i,j) > 0$
- $i \to j$: j is reachable from $i : \mathbb{P}_i(T_j < +\infty) = 1$, i.e. $\exists n \geq 1 : p^{(n)}(i,j) > 0$
- $i \sim j : i$ communicates with j iff both $i \to j$ and $j \to i$: equivalence relation, comunicating classes
- *Irreducible chain* : single communicating class

- *Period* of state $i = gcd\{n \geq 1 : p^{(n)}(i,i) > 0\}$, $gcd(\emptyset) = +\infty$. gcd(lengths of circuits)
- i is said *aperiodic* if its period is 1. *Proposition*: $\exists r : \forall n \geq r \; p^{(n)}(i,i) > 0$.
- The chain is said *aperiodic* if all its states are aperiodic
- *Proposition*: All the states of a communicating class have same period