MathExp 2018: Chaînes de Markov

Chaînes de Markov – définition, propriétés, algorithmes

Ana Busic

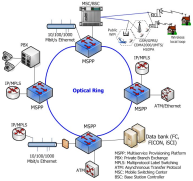
Inria Paris - DI ENS

http://www.di.ens.fr/~busic/

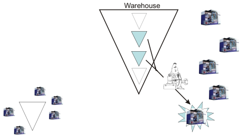
Saint Flour, Mai 2018

Systèmes complexes

Réseaux de communications :

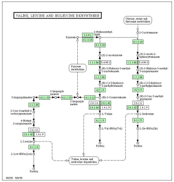


Génie industriel :



Centres de calcul :

Réseaux métaboliques :



Modèles markoviens

- Largement utilisés - simplicité de modélisation des systèmes complexes : décrire les états et les transitions

Modèles markoviens

- Largement utilisés - simplicité de modélisation des systèmes complexes : décrire les états et les transitions
- Différents formalismes de haut niveau (Petri nets, Stochastic Automata Networks, ...) $\rightarrow$ modélisation et stockage encore plus simples : décrire différentes composantes et leurs interactions

Modèles markoviens

- Largement utilisés - simplicité de modélisation des systèmes complexes : décrire les états et les transitions
- Différents formalismes de haut niveau (Petri nets, Stochastic Automata Networks, ...) $\rightarrow$ modélisation et stockage encore plus simples : décrire différentes composantes et leurs interactions
- Problème : l'explosion de l'espace d'états - difficile/impossible à analyser

Modèles markoviens

- Largement utilisés - simplicité de modélisation des systèmes complexes : décrire les états et les transitions
- Différents formalismes de haut niveau (Petri nets, Stochastic Automata Networks, ...) $\rightarrow$ modélisation et stockage encore plus simples : décrire différentes composantes et leurs interactions
- Problème : l'explosion de l'espace d'états - difficile/impossible à analyser
- Les méthodes d'étude :
  - Méthodes analytiques (sous des hypothèses restrictives)

Modèles markoviens

- Largement utilisés - simplicité de modélisation des systèmes complexes : décrire les états et les transitions
- Différents formalismes de haut niveau (Petri nets, Stochastic Automata Networks, ...) $\rightarrow$ modélisation et stockage encore plus simples : décrire différentes composantes et leurs interactions
- Problème : l'explosion de l'espace d'états - difficile/impossible à analyser
- Les méthodes d'étude :
  - Méthodes analytiques (sous des hypothèses restrictives)
  - Méthodes numériques (limitées par la taille de l'espace d'états)

Modèles markoviens

- Largement utilisés - simplicité de modélisation des systèmes complexes : décrire les états et les transitions
- Différents formalismes de haut niveau (Petri nets, Stochastic Automata Networks, ...) $\rightarrow$ modélisation et stockage encore plus simples : décrire différentes composantes et leurs interactions
- Problème : l'explosion de l'espace d'états - difficile/impossible à analyser
- Les méthodes d'étude :
  - Méthodes analytiques (sous des hypothèses restrictives)
  - Méthodes numériques (limitées par la taille de l'espace d'états)
  - Approximations (avec ou sans contrôle d'erreur) et bornes

Modèles markoviens

- Largement utilisés - simplicité de modélisation des systèmes complexes : décrire les états et les transitions
- Différents formalismes de haut niveau (Petri nets, Stochastic Automata Networks, ...) $\rightarrow$ modélisation et stockage encore plus simples : décrire différentes composantes et leurs interactions
- Problème : l'explosion de l'espace d'états - difficile/impossible à analyser
- Les méthodes d'étude :
  - Méthodes analytiques (sous des hypothèses restrictives)
  - Méthodes numériques (limitées par la taille de l'espace d'états)
  - Approximations (avec ou sans contrôle d'erreur) et bornes
  - Simulation (problème de convergence vers la limite stationnaire)

Plan

Plan

Chaînes de Markov

$\{X_n, n \in \mathbb{N}\}$ une suite de variables aléatoires (un processus stochastique) à valeurs dans un espace d'états $\mathcal{E}$ dénombrable est appelé chaîne de Markov à temps discret (CMTD) sur $\mathcal{E}$ si pour tout $n \in \mathbb{N}$, pour tous $i, j, i_0, \ldots, i_{n-1} \in \mathcal{E}$,

$$\mathbf{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \mathbf{P}(X_{n+1} = j \mid X_n = i).$$

Chaînes de Markov

$\{X_n, n \in \mathbb{N}\}$ une suite de variables aléatoires (un processus stochastique) à valeurs dans un espace d'états $\mathcal{E}$ dénombrable est appelé chaîne de Markov à temps discret (CMTD) sur $\mathcal{E}$ si pour tout $n \in \mathbb{N}$, pour tous $i, j, i_0, \ldots, i_{n-1} \in \mathcal{E}$,

$$\mathbf{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \mathbf{P}(X_{n+1} = j \mid X_n = i).$$

Si pour tout $i, j \in \mathcal{E}$, $\mathbf{P}(X_{n+1} = j \mid X_n = i)$ ne dépend pas de n, la CMTD est dite homogène.

Chaînes de Markov

$\{X_n, n \in \mathbb{N}\}$ une suite de variables aléatoires (un processus stochastique) à valeurs dans un espace d'états $\mathcal{E}$ dénombrable est appelé chaîne de Markov à temps discret (CMTD) sur $\mathcal{E}$ si pour tout $n \in \mathbb{N}$, pour tous $i, j, i_0, \ldots, i_{n-1} \in \mathcal{E}$,

$$\mathbf{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \mathbf{P}(X_{n+1} = j \mid X_n = i).$$

Si pour tout $i, j \in \mathcal{E}$, $\mathbf{P}(X_{n+1} = j \mid X_n = i)$ ne dépend pas de n, la CMTD est dite homogène.

Propriété : une CMTD est un processus stochastique pour lequel le futur $(X_{n+1}, \ldots)$ est indépendant du passé $(X_0, \ldots, X_{n-1})$ conditionnellement au présent (X_n).

Représentation matricielle

Les $p_{ij} := \mathbf{P}(X_{n+1} = j \mid X_n = i)$ sont appelés les probabilités de transitions et la matrice (de dimension possiblement infinie) $\mathbf{P} = (p_{ij})_{i,j \in \mathcal{E}}$ est la matrice de transition de la CMTD.

La matrice $\mathbf{P}$ est une matrice stochastique : la somme des coefficients sur chaque ligne est égale à 1 ($\forall i \in \mathcal{E}, \sum_{j \in \mathcal{E}} p_{ij} = 1$).

Représentation matricielle

Les $p_{ij} := \mathbf{P}(X_{n+1} = j \mid X_n = i)$ sont appelés les probabilités de transitions et la matrice (de dimension possiblement infinie) $\mathbf{P} = (p_{ij})_{i,j\in\mathcal{E}}$ est la matrice de transition de la CMTD.

La matrice $\mathbf{P}$ est une matrice stochastique : la somme des coefficients sur chaque ligne est égale à 1 ($\forall i \in \mathcal{E}, \sum_{j\in\mathcal{E}} p_{ij} = 1$).

Soit μ_0 la distribution initiale de la chaîne, $\mu_0(i) = \mathbf{P}(X_0 = i)$. Alors

$$\mathbf{P}(X_0 = i_0, X_1 = i_1) = \mathbf{P}(X_0 = i_0)\mathbf{P}(X_1 = i_1 \mid X_0 = i_0) = \mu_0(i_0)p_{i_0 i_1}.$$

Par récurrence on a aussi

$$\begin{aligned}
&\mathbf{P}(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \\
&\quad \mathbf{P}(X_0 = i_0, \ldots, X_{n-1} = i_{n-1})\mathbf{P}(X_n = i_n \mid X_0 = i_0, \ldots, X_{n-1} = i_{n-1}) = \\
&\quad\quad \mathbf{P}(X_0 = i_0, \ldots, X_{n-1} = i_{n-1})\mathbf{P}(X_n = i_n \mid X_{n-1} = i_{n-1}) = \\
&\quad\quad\quad \mu_0(i_0)p_{i_0 i_1} \ldots p_{i_{n-1} i_n}.
\end{aligned}$$

Donc μ_0 et $\mathbf{P}$ déterminent entièrement la loi de $\{X_n\}$.

Équations de Chapman-Kolmogorov

Notons $\mu_n(i) = \mathbf{P}(X_n = i)$; μ_n est la mesure de probabilité de la chaîne à la n-ième étape.

On a $\mu_{n+1}(i) = \sum_{j \in \mathcal{E}} \mu_n(j) p_{ji}$, donc $\forall n \in \mathbb{N}, \mu_{n+1} = \mu_n \mathbf{P}$ et $\mu_n = \mu_0 \mathbf{P}^n$.

Soit $p_{ij}(n) = (P^n)_{ij}$. Pour tout $m \in \mathbb{N}$, on a

$$\mathbf{P}(X_{m+n} = j \mid X_m = i) = \sum_{i_1, \ldots, i_{n-1} \in \mathcal{E}} p_{ii_1} p_{i_1 i_2} \ldots p_{i_{n-1} j} = p_{ij}(n).$$

$\mathbf{P}^n$ est la matrice de transition à n étapes et $\mathbf{P}^{m+n} = \mathbf{P}^m \mathbf{P}^n$ s'écrit sous la forme des équations de Chapman-Kolmogorov

$$p_{ij}(m+n) = \sum_{k \in \mathcal{E}} p_{ik}(m) p_{kj}(n) \quad \forall m, n \in \mathbb{N}.$$

Représentation graphique

On peut représenter une chaîne de Markov homogène par un graphe orienté pondéré dont

- les sommets sont les états de la chaîne et
- les arcs représentent les probabilités de transitions : il y a un arc de i à j étiqueté par $p_{i,j}$ si et seulement si $p_{i,j} \neq 0$.

Exemple

$\mathcal{E} = \{1, 2, 3\}$ et

$$\mathbf{P} = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \\ 3/4 & 1/4 & 0 \end{pmatrix}.$$

Donner la représentation graphique d'une chaîne de Markov avec la matrice de transition $\mathbf{P}$.

Classes de communication

On appelle classe de communication de la chaîne une composante fortement connexe de son graphe de transition.

Une chaîne de Markov est dite irréductible si elle est composée d'une seule classe de communication (son graphe de transition est fortement connexe).

La période d'un état i d'une CMTD est le pgcd des entiers $\{n \in \mathbb{N} \mid p_{ii}(n) > 0\}$, que l'on note d_i. Si $d_i = 1$, alors l'état est dit apériodique.

Classes de communication

Lemme

Tous les états d'une même classe de communication ont la même période.

Démonstration.

Soient i et j deux états de la même classe de communication. On a (équations de Chapman-Kolmogorov + positivité des matrices)

$$p_{ii}(m+k+n) \geq p_{ij}(m)p_{jj}(k)p_{ji}(n).$$

Prenons m et n tels que $p_{ij}(m)p_{ji}(n) > 0$. Alors $m+n$ est un multiple de d_i. Si k n'est pas un multiple de d_i, alors $p_{ii}(m+k+n)$ est égal à 0 et donc $p_{jj}(k)$ aussi. Donc $d_i | d_j$. Pas symétrie, on a aussi $d_j | d_i$ et donc $d_i = d_j$. $\square$

Représentation fonctionnelle

Théorème

La relation de récurrence

$$X_{n+1} = f(X_n, Z_{n+1}), n \geq 0$$

où $f : \mathcal{E} \times F \to \mathcal{E}$ *et où* $\{Z_n\}_{n\geq 1}$ *est une suite de v.a. dans* F *dénombrable, définit une chaîne de Markov si* Z_{n+1} *est indépendant de* $X_0, \ldots, X_{n-1}, Z_1, \ldots, Z_n$ *conditionnellement à* X_n*.*
Si $\mathbf{P}(Z_{n+1} = k \mid X_n = i)$ *est indépendant de* n*, alors la chaîne de Markov est homogène.*

Cas particulier : $\{Z_n\}$ i.i.d. et indépendants de $\{X_n\}$.

Exemple : marche aléatoire sur $\mathbb{Z}$. $X_n = X_{n-1} + Z_n$, où $\{Z_n\}$ est i.i.d. avec $\mathbf{P}(Z_n = 1) = p, \mathbf{P}(Z_n = -1) = 1 - p$.

Exemple : marche aléatoire sur $\mathbb{N}$. $X_n = (X_{n-1} + Z_n)^+$, où $\{Z_n\}$ est i.i.d. avec $\mathbf{P}(Z_n = 1) = p, \mathbf{P}(Z_n = -1) = 1 - p$.

Preuve

Démonstration. On utilise des égalités usuelles des probabilités conditionnelles :

- $\mathbf{P}(A \mid B) = \mathbf{P}(A \cap B)/\mathbf{P}(B)$;
- Si $A = \sqcup A_k$, alors $\mathbf{P}(A \cap B) = \sum_k \mathbf{P}(A_k \cap B) = \sum_k \mathbf{P}(A_k \mid B) P(B) = P(A \mid B) P(B)$.

donc

$$\mathbf{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \\ \frac{\mathbf{P}(X_{n+1} = j, X_n = i, X_{n-1} = i_{n-1} \ldots, X_0 = i_0)}{\mathbf{P}(X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0)} = \\ \frac{\mathbf{P}(f(i, Z_{n+1}) = j, X_n = i, \ldots, X_0 = i_0)}{\mathbf{P}(X_n = i, \ldots, X_0 = i_0)}$$

Preuve

Les événements $\{Z_{n+1} = k\}$ sont disjoints, donc

$$\mathbf{P}(f(i, Z_{n+1}) = j, X_n = i, \ldots, X_0 = i_0) = \\ \sum_{k: f(i,k)=j} \mathbf{P}(Z_{n+1} = k \mid X_n = i, \ldots, X_0 = i_0)\mathbf{P}(X_n = i, \ldots, X_0 = i_0).$$

Par hypothèse, on a
$\mathbf{P}(Z_{n+1} = k \mid X_n = i, \ldots, X_0 = i_0) = \mathbf{P}(Z_{n+1} = k \mid X_n = i)$ et donc

$$\mathbf{P}(f(i, Z_{n+1}) = j, X_n = i, \ldots, X_0 = i_0) = \\ \sum_{k: f(i,k)=j} \mathbf{P}(Z_{n+1} = k \mid X_n = i)\mathbf{P}(X_n = i, \ldots, X_0 = i_0)$$

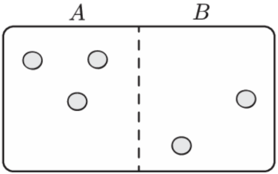
On obtient donc

$$\mathbf{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1} \ldots, X_0 = i_0) = \\ \sum_{k: f(i,k)=j} \mathbf{P}(Z_{n+1} = k | X_n = i) = \mathbf{P}(X_{n+1} = j | X_n = i).$$

$\square$

Exemple : urne d'Ehrenfest

Modèle idéalisé de diffusion à travers d'une membrane poreuse (Tatiana et Paul Ehrenfest 1907)



N particules divisées en deux compartiments, à chaque instant une particule est choisi de manière uniforme et déplacée dans l'autre compartiment.

L'état : X_n nombre de particules dans le compartiment A à la date n. Pour $n \geq 0$, on a

$$X_{n+1} = X_n + Z_{n+1}$$

avec $Z_n \in \{-1, 1\}$ et $\mathbf{P}(Z_{n+1} = -1 | X_n = i) = i/N$.

Distribution stationnaire

Soit $\mu = (\mu(i), i \in \mathcal{E})$ une mesure non nulle sur $\mathcal{E}$. La mesure μ est dite invariante si $\mu = \mu\mathbf{P}$, c'est-à-dire si $\forall i \in \mathcal{E}$,

$$\mu(i) \geq 0 \quad \text{et} \quad \mu(i) = \sum_{j \in \mathcal{E}} \mu(j) p_{ji}.$$

Soit $\pi = (\pi(i), i \in \mathcal{E})$ une distribution de probabilités sur $\mathcal{E}$ telle que $\pi = \pi\mathbf{P}$, c'est-à-dire telle que $\forall i \in \mathcal{E}$,

$$\pi(i) \geq 0, \quad \sum_{j \in \mathcal{E}} \pi_j = 1 \quad \text{et} \quad \pi(i) = \sum_{j \in \mathcal{E}} \pi(j) p_{ji}.$$

La distribution π est appelée distribution (de probabilités) stationnaire de $\mathbf{P}$ ou de $\{X_n\}$.

Si $\mu_0 = \pi$, alors, pour tout $n \in \mathbb{N}, \mu_n = \pi$. On dit alors que la chaîne est stationnaire.

Chaînes de Markov

Exercice

Trouver une CMTD

1. qui n'a pas de mesure invariante ;
2. qui a une mesure invariante, mais pas de distribution stationnaire ;
3. qui a plusieurs distributions stationnaires.

Chaînes de Markov

Exercice

Trouver une CMTD

1. qui n'a pas de mesure invariante ;
2. qui a une mesure invariante, mais pas de distribution stationnaire ;
3. qui a plusieurs distributions stationnaires.

Exercice

La marche aléatoire non biaisée $X_n = X_0 + Z_1 + Z_2 + \ldots + Z_n$, où $\{Z_n\}$ est i.i.d. avec $\mathbf{P}(Z_n = 1) = \mathbf{P}(Z_n = -1) = 1/2$.

1. Quel est l'espace des états ? Montrer que c'est une chaîne de Markov.
2. Donner le graphe de transition et sa matrice de transition
3. La chaîne est-elle irréductible ? quelle est sa période ?
4. Existe-t-il une distribution stationnaire ? une mesure stationnaire ?

Plan

Chaînes de Markov reversibles

Soit $\{X_t\}_{t\geq 0}$ une chaîne de Markov avec l'espace d'états $S = \{s_1, \ldots, s_n\}$ et la matrice de transition P. Une distribution de probabilités π sur S est dite reversible si pour tout $i, j \in \{1, \ldots, n\}$,

$$\pi_i P_{i,j} = \pi_j P_{j,i}.$$

Chaînes de Markov reversibles

Soit $\{X_t\}_{t\geq 0}$ une chaîne de Markov avec l'espace d'états $S = \{s_1, \ldots, s_n\}$ et la matrice de transition P. Une distribution de probabilités π sur S est dite reversible si pour tout $i, j \in \{1, \ldots, n\}$,

$$\pi_i P_{i,j} = \pi_j P_{j,i}.$$

Theorem

Si π est une distribution reversible, alors elle est aussi une distribution stationnaire.

Chaînes de Markov reversibles

Soit $\{X_t\}_{t\geq 0}$ une chaîne de Markov avec l'espace d'états $S = \{s_1, \ldots, s_n\}$ et la matrice de transition P. Une distribution de probabilités π sur S est dite reversible si pour tout $i, j \in \{1, \ldots, n\}$,

$$\pi_i P_{i,j} = \pi_j P_{j,i}.$$

Theorem

Si π est une distribution reversible, alors elle est aussi une distribution stationnaire.

Démonstration.

Pour tout $j \in \{1, \ldots, n\}$, on a

$$\pi_j = \pi_j \sum_{i=1}^{n} P_{j,i} = \sum_{i=1}^{n} \pi_j P_{j,i} = \sum_{i=1}^{n} \pi_i P_{i,j}.$$

$\square$

Temps inverse

Theorem

Soit $\{X_t\}_{t\geq 0}$ *une chaîne de Markov avec l'espace d'états* S *et la matrice de transition* P *et la distribution stationnaire* π *reversible.*

Si $X_0 \sim \pi$*, alors pour tout* $k \in \mathbb{N}$ *et tous* $s_{i_0}, s_{i_1}, \ldots, s_{i_k} \in S$*,*

$$P(X_0 = s_{i_0}, X_1 = \ldots, X_k = s_{i_k}) = P(X_0 = s_{i_k}, X_1 = s_{i_{k-1}}, \ldots, X_k = s_{i_0}).$$

La probabilité d'un chemin dans un sens est égale à la probabilité du même chemin dans le sens inverse.

Exemple : Processus de naissance et de mort

Espace d'états : $\mathcal{E} = \{0, 1, \dots, N\}$; matrice de transition :

$$\mathbf{P} = \begin{pmatrix} r_0 & p_0 & & & & & & \\ q_1 & r_1 & p_1 & & & & & \\ & q_2 & r_2 & p_2 & & & & \\ & & \ddots & & & & & \\ & & & q_i & r_i & p_i & & \\ & & & & \ddots & \ddots & \ddots & \\ & & & & & q_{n-1} & r_{N-1} & p_{N-1} \\ & & & & & & q_N & r_N \end{pmatrix}$$

avec $p_i > 0, q_i > 0, r_i \geq 0$ et $p_i + r_i + q_i = 1, \forall i$.

Exemple : Processus de naissance et de mort

Espace d'états : $\mathcal{E} = \{0, 1, \ldots, N\}$; matrice de transition :

$$\mathbf{P} = \begin{pmatrix} r_0 & p_0 & & & & & & \\ q_1 & r_1 & p_1 & & & & & \\ & q_2 & r_2 & p_2 & & & & \\ & & \ddots & & & & & \\ & & & q_i & r_i & p_i & & \\ & & & & \ddots & \ddots & \ddots & \\ & & & & & q_{n-1} & r_{N-1} & p_{N-1} \\ & & & & & & q_N & r_N \end{pmatrix}$$

avec $p_i > 0, q_i > 0, r_i \geq 0$ et $p_i + r_i + q_i = 1, \forall i$.

C'est une chaîne reversible : $\pi_{i-1} p_{i-1} = \pi_i q_i$.

Exemple : Processus de naissance et de mort

Espace d'états : $\mathcal{E} = \{0, 1, \dots, N\}$; matrice de transition :

$$\mathbf{P} = \begin{pmatrix} r_0 & p_0 & & & & & & \\ q_1 & r_1 & p_1 & & & & & \\ & q_2 & r_2 & p_2 & & & & \\ & & \ddots & & & & & \\ & & & q_i & r_i & p_i & & \\ & & & & \ddots & \ddots & \ddots & \\ & & & & & q_{n-1} & r_{N-1} & p_{N-1} \\ & & & & & & q_N & r_N \end{pmatrix}$$

avec $p_i > 0, q_i > 0, r_i \geq 0$ et $p_i + r_i + q_i = 1, \forall i$.

C'est une chaîne reversible : $\pi_{i-1} p_{i-1} = \pi_i q_i$.

On pose $w_0 = 1$ et $w_i = \prod_{k=1}^{i} \frac{p_{k-1}}{q_k}, 1 \leq i \leq N$.

Alors, $\pi_i = \frac{w_i}{\sum_{i=0}^{N} w_i}$ est l'unique probabilité stationnaire.

Exemple : Processus de naissance et de mort

Espace d'états : $\mathcal{E} = \mathbb{N}$.

Le vecteur w, avec $w_0 = 1$ et $w_i = \prod_{k=1}^{i} \frac{p_{k-1}}{q_k}, 1 \leq i < \infty$ est toujours une mesure invariante, mais la probabilité stationnaire existe ssi

$$\sum_{i=0}^{\infty} w_i < \infty.$$

Dans ce cas, la probabilité stationnaire est unique et $\pi_i = \frac{w_i}{\sum_{i=0}^{\infty} w_i}$.

Exemple : Processus de naissance et de mort

Espace d'états : $\mathcal{E} = \mathbb{N}$.

Le vecteur w, avec $w_0 = 1$ et $w_i = \prod_{k=1}^{i} \frac{p_{k-1}}{q_k}, 1 \leq i < \infty$ est toujours une mesure invariante, mais la probabilité stationnaire existe ssi

$$\sum_{i=0}^{\infty} w_i < \infty.$$

Dans ce cas, la probabilité stationnaire est unique et $\pi_i = \frac{w_i}{\sum_{i=0}^{\infty} w_i}$.

La chaîne est

- apériodique ssi $\exists r_i > 0$.
- irréductible (car $p_i > 0$ et $q_i > 0, \forall i$).
- récurrente positive ssi $\sum_{i=0}^{\infty} w_i < \infty$.

Exemple : Processus de naissance et de mort

Espace d'états : $\mathcal{E} = \mathbb{N}$.

Le vecteur w, avec $w_0 = 1$ et $w_i = \prod_{k=1}^{i} \frac{p_{k-1}}{q_k}, 1 \leq i < \infty$ est toujours une mesure invariante, mais la probabilité stationnaire existe ssi

$$\sum_{i=0}^{\infty} w_i < \infty.$$

Dans ce cas, la probabilité stationnaire est unique et $\pi_i = \frac{w_i}{\sum_{i=0}^{\infty} w_i}$.

La chaîne est

- apériodique ssi $\exists r_i > 0$.
- irréductible (car $p_i > 0$ et $q_i > 0, \forall i$).
- récurrente positive ssi $\sum_{i=0}^{\infty} w_i < \infty$.

Exemple : $p_i = p, i \geq 0,\ q_i = q = 1 - p, \imath > 0,\ r_0 = 1 - p_0,\ r_i = 0, i > 0$.
Nous avons $w_i = (\frac{p}{q})^i, i > 0$. Alors $\sum_{i=0}^{\infty} w_i < \infty$ ssi $p < q$. Dans ce cas,

$$\pi_i = \left(1 - \frac{p}{q}\right)\left(\frac{p}{q}\right)^i, \forall i.$$

Exemple : marche aléatoire sur un graphe

Un graphe $G = (V, E)$ avec les sommets $V = \{v_1, \dots, v_n\}$ et les arêtes $E = \{e_1, \dots, e_m\}$. Soit d_i le degré du sommet v_i.

Exemple : marche aléatoire sur un graphe

Un graphe $G = (V, E)$ avec les sommets $V = \{v_1, \dots, v_n\}$ et les arêtes $E = \{e_1, \dots, e_m\}$. Soit d_i le degré du sommet v_i.

Un marcheur qui est au sommet v_i à l'instant t change sa position pour un sommet voisin de v_i avec la même probabilité pour tous les voisins.

Exemple : marche aléatoire sur un graphe

Un graphe $G = (V, E)$ avec les sommets $V = \{v_1, \ldots, v_n\}$ et les arêtes $E = \{e_1, \ldots, e_m\}$. Soit d_i le degré du sommet v_i.

Un marcheur qui est au sommet v_i à l'instant t change sa position pour un sommet voisin de v_i avec la même probabilité pour tous les voisins.

Matrice de transition P :

$$P_{i,j} = 1/d_i, \text{ si } (v_i, v_j) \in E$$

et sinon $P_{i,j} = 0$.

Exemple : marche aléatoire sur un graphe

Un graphe $G = (V, E)$ avec les sommets $V = \{v_1, \ldots, v_n\}$ et les arêtes $E = \{e_1, \ldots, e_m\}$. Soit d_i le degré du sommet v_i.

Un marcheur qui est au sommet v_i à l'instant t change sa position pour un sommet voisin de v_i avec la même probabilité pour tous les voisins.

Matrice de transition P :

$$P_{i,j} = 1/d_i, \text{ si } (v_i, v_j) \in E$$

et sinon $P_{i,j} = 0$.

Alors,

$$\pi = (d_1, \ldots, d_n)/(\sum_{i=1}^{n} d_i)$$

est une distribution de probabilités reversible.

Processus en temps retourné

Processus en temps retourné

Soit $\{X_t\}$ un processus stationnaire et irréductible. On construit $\{X_t^*\}$ en inversant le temps :

$$\boxed{X_t^* = X_{\tau-t}}$$

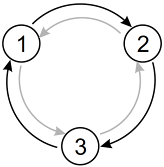
Remarque : τ n'est pas important, il determine uniquement l'origine pour le processus retourné.

Applications

- Permet de mieux comprendre les propriétés d'un processus.
- Les preuves plus élégantes. Exemple : thm de Burke pour les files d'attente (le processus de départs des files plus facile à analyser).
- Parfois plus facile de "deviner" la forme de la loi stationnaire.

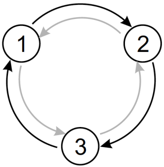
Processus en temps retourné

En général, $\{X_t^*\}$ est différent de $\{X_t\}$.
Exemple : processus cyclique.



Processus en temps retourné

En général, $\{X_t^*\}$ est différent de $\{X_t\}$.
Exemple : processus cyclique.

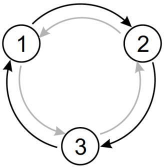


Loi stationnaire
Thm. Suppose que $\{X_t\}$ a la distribution stationnaire π, $\pi_i = \mathbf{P}(X_t = i)$.
Alors $\{X_t^*\}$ a aussi une loi stationnaire π^* et

$$\boxed{\pi^* = \pi}$$

Processus en temps retourné

En général, $\{X_t^*\}$ est différent de $\{X_t\}$.
Exemple : processus cyclique.



Loi stationnaire

Thm. Suppose que $\{X_t\}$ a la distribution stationnaire π, $\pi_i = \mathbf{P}(X_t = i)$.
Alors $\{X_t^*\}$ a aussi une loi stationnaire π^* et

$$\boxed{\pi^* = \pi}$$

Preuve. π_i et π_i^* représentent les proportions de temps que $\{X_t\}$ et $\{X_t^*\}$ passent en état i. Cette proportion ne dépend pas de la direction du temps.

CMH stationnaire

Prop. Le processus retourné $\ldots, X_{n+1}, X_n, X_{n-1}, \ldots$ d'une chaîne de Markov stationnaire en temps discret $\ldots, X_{n-1}, X_n, X_{n+1}, \ldots$ est aussi une chaîne de Markov stationnaire.

Les probabilités de transitions sont

$$P^*_{i,j} = \mathbf{P}(X_n = j | X_{n+1} = i) = \frac{\pi_j P_{j,i}}{\pi_i}.$$

Processus réversible

Déf. Si les processus $\{X_n\}$ et $\{X_n^*\}$ sont statistiquement non-distinguables, on dit que $\{X_n\}$ est réversible (en temps).

$$(X_{t_1}, X_{t_2}, \dots, X_{t_n}) \sim (X^*_{\tau-t_1}, X^*_{\tau-t_2}, \dots, X^*_{\tau-t_n})$$

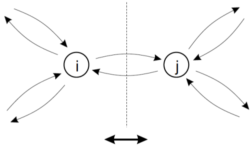
pour tout n, τ et $t_1, \dots, t_n$.

Intuitivement : un spectateur ne peut pas dire si le "film" est projeté en avant ou en arrière.

CMH réversible

$p^*_{i,j} = p_{i,j}, \forall i,j$, i.e. $\boxed{\pi_i p_{i,j} = \pi_j p_{j,i}, \forall i,j}$

Équations de balance détaillée : les flots de probabilité entre deux états sont en équilibre.



- Balance détaillée $\Rightarrow$ balance globale
- Si les conditions de balance détaillée sont vérifiées pour un vecteur π positif et tel que $\sum_i \pi_i < \infty$, alors π normalisé tel que $\sum_i \pi = 1$ est la loi stationnaire.
- Mais balance globale $\not\Rightarrow$ balance détaillée (tous les processus de Markov ne sont pas réversibles).

Exemple : les arbres sont réversibles

Prop. Si une CMH est reversible, alors son graphe de transition est symétrique.

Exemple : les arbres sont réversibles

Prop. Si une CMH est reversible, alors son graphe de transition est symétrique.

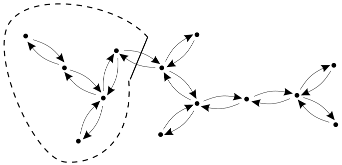
Prop. Si le graphe de transition non-orienté d'une CMH est un arbre, alors elle est réversible.

Exemple : les arbres sont réversibles

Prop. Si une CMH est reversible, alors son graphe de transition est symétrique.

Prop. Si le graphe de transition non-orienté d'une CMH est un arbre, alors elle est réversible.

Preuve. En utilisant la balance détaillée.

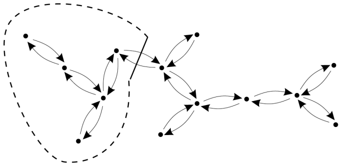


Exemple : les arbres sont réversibles

Prop. Si une CMH est reversible, alors son graphe de transition est symétrique.

Prop. Si le graphe de transition non-orienté d'une CMH est un arbre, alors elle est réversible.

Preuve. En utilisant la balance détaillée.

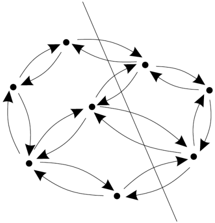


Cor. Tous les CMH de naissance et de mort sont réversibles.

Troncation d'un processus réversible

Soit $\{X_n\}$ une CMH avec l'espace d'états $\mathcal{E}$ et la loi stationnaire π. Soit $\mathcal{E}' \subset \mathcal{E}$. CMH tronqué $\{X'_n\}$ défini par :

$$p'_{i,j} = p_{i,j}, \quad i,j \in \mathcal{E}'; \quad p'_{i,i} = 1 - \sum_{j \neq i} p_{ij}$$



Prop. Si $\{X'_n\}$ est irréductible, alors $\{X'_n\}$ est réversible et sa distribution stationnaire est

$$\boxed{\pi'_i = \frac{\pi_i}{\sum_{j \in \mathcal{E}'} \pi_j}}$$

Plan

Temps d'arrêt

Définition

Soit τ une v. a. sur $N \cup \{\infty\}$. τ est un temps d'arrêt du processus stochastique $\{X_n, n \in \mathbb{N}\}$ si pour tout $k \geq 0$, l'événement $\{\tau = k\}$ s'exprime en fonction de $X_0, \ldots, X_k$:

$$\{\tau = k\} = \sqcup_{(i_0,\ldots,i_k)\in\mathcal{B}_k}\{X_0 = i_0, \ldots, X_k = i_k\}, \qquad \text{avec } \mathcal{B}_k \subset \mathcal{E}^{k+1},$$

où $\sqcup$ denote l'union disjointe.

Exemples

1. $\tau = n_0$, où n_0 est une constante non aléatoire.
2. Temps de retour en $F \subset \mathcal{E}$:

$$T_F = \inf\{n \geq 1 \mid X_n \in F\} \qquad (+\infty \text{ si } \forall n \in \mathbb{N}^*, X_n \notin F).$$

   $\{T_F = k\}$ s'exprime comme $\{X_1 \notin F, \ldots, X_{k-1} \notin F, X_k \in F\}$.

3. Temps d'atteinte de F :

$$T'_F = \inf\{n \geq 0 \mid X_n \in F\} \qquad (+\infty \text{ si } \forall n \in \mathbb{N}, X_n \notin F).$$

Propriété de Markov forte

τ un temps d'arrêt pour $\{X_n, n \in \mathbb{N}\}$, deux nouveaux processus :

- le processus avant τ : $\{X_{n\wedge\tau}\}$;
- le processus après τ : $\{X_{n+\tau}\}$ (si $\tau = \infty, X_\infty = i^* \notin \mathcal{E}$).

Théorème (propriété de Markov forte)

Soit $\{X_n, n \in \mathbb{N}\}$ une chaîne de Markov sur $\mathcal{E}$ et τ un temps d'arrêt presque sûrement fini pour X. Alors,

1. $\forall i \in \mathcal{E}, \{X_{\tau+n}, n \geq 0\}$ *et* $\{X_{\tau\wedge n}, n \geq 0\}$ *sont indépendants cond. à* $\{X_\tau = i\}$ *:* $\forall n, k \in \mathbb{N}, \forall i_0, \ldots, i_n, j_1, \ldots, j_k \in \mathcal{E}$, *on a*

$$\mathbf{P}(X_{\tau+1} = j_1, \ldots, X_{\tau+k} = j_k \,|\, X_\tau = i, X_{0\wedge\tau} = i_0, \ldots, X_{n\wedge\tau} = i_n) = \mathbf{P}(X_{\tau+1} = j_1, \ldots, X_{\tau+k} = j_k \,|\, X_\tau = i).$$

2. *Étant donné* $\{X_\tau = i\}$, $i \in \mathcal{E}$, $\{X_{\tau+n}, n \geq 0\}$ *est une CMH à valeurs dans* $\mathcal{E}$ *de même matrice de transition que* X *et d'état initial* i *:*

$$\mathbf{P}(X_{\tau+1} = j_1, \ldots, X_{\tau+k} = j_k \,|\, X_\tau = i) = p_{ij_1} p_{j_1 j_2} \ldots p_{j_{k-1} j_k}.$$

Méthode du pas en avant

Théorème

Soit $\{X_n\}$ *une CMH de matrice de transition* $\mathbf{P}$, $F \subset \mathcal{E}$ *et*
$T_F = \inf\{n \geq 0 \mid X_n \in F\}$.

$m(i) := \mathbf{E}[T_F \mid X_0 = i], i \in \mathcal{E}$ *(temps moyen d'atteinte de* F*). Alors*

$$m(i) = \begin{cases} 1 + \sum_{j \in \mathcal{E}} p_{ij} m(j), & i \notin F \\ 0, & i \in F. \end{cases}$$

Méthode du pas en avant

Théorème

Soit $\{X_n\}$ *une CMH de matrice de transition* $\mathbf{P}$, $F \subset \mathcal{E}$ *et* $T_F = \inf\{n \geq 0 \mid X_n \in F\}$.

$m(i) := \mathbf{E}[T_F \mid X_0 = i], i \in \mathcal{E}$ *(temps moyen d'atteinte de* F*). Alors*

$$m(i) = \begin{cases} 1 + \sum_{j \in \mathcal{E}} p_{ij} m(j), & i \notin F \\ 0, & i \in F. \end{cases}$$

Démonstration. Soit $Y_n = X_{n+1}$. Si $X_0 \notin F$, alors $T_F(\{X_n\}) = 1 + T_F(\{Y_n\})$ et donc, si $i \notin F$,

$$\begin{aligned} \mathbf{E}(T_F(\{X_n\}) \mid X_0 = i) &= \mathbf{E}(T_F(\{Y_n\}) + 1 \mid X_0 = i) \\ &= 1 + \sum_{j \in \mathcal{E}} \mathbf{E}(T_F(\{Y_n\}) 1_{X_1 = j} \mid X_0 = i) \\ &= 1 + \sum_{j \in \mathcal{E}} \mathbf{E}(T_F(\{Y_n\}) \mid Y_0 = j) \mathbf{P}(X_1 = j \mid X_0 = i) \\ &= 1 + \sum_{j \in \mathcal{E}} p_{ij} m(j). \end{aligned}$$

$\square$

Classification des états

Notation

- $\mathbf{P}_i(A) = \mathbf{P}(A \mid X_0 = i)$.
- $N_i = \sum_{n=1}^{\infty} 1_{X_n = i}$ le nombre de visites de X dans l'état i
- f_{ij} la probabilité d'accéder en j en partant de i ($f_{ij} = \mathbf{P}(T_j < \infty \mid X_0 = i) = \mathbf{P}_i(T_j < \infty)$, avec T_j le temps d'atteinte de j).

Classification des états

Notation

- $\mathbf{P}_i(A) = \mathbf{P}(A \mid X_0 = i)$.
- $N_i = \sum_{n=1}^{\infty} 1_{X_n = i}$ le nombre de visites de X dans l'état i
- f_{ij} la probabilité d'accéder en j en partant de i ($f_{ij} = \mathbf{P}(T_j < \infty \mid X_0 = i) = \mathbf{P}_i(T_j < \infty)$, avec T_j le temps d'atteinte de j).

Le temps moyen de retour en i : $m(i) = \mathbf{E}_i(T_i)$.

$1/m(i)$ la fréquence moyenne de retour en i (nulle si $m(i) = \infty$).

Classification des états

Notation

- $\mathbf{P}_i(A) = \mathbf{P}(A \,|\, X_0 = i)$.
- $N_i = \sum_{n=1}^{\infty} 1_{X_n = i}$ le nombre de visites de X dans l'état i
- f_{ij} la probabilité d'accéder en j en partant de i ($f_{ij} = \mathbf{P}(T_j < \infty \,|\, X_0 = i) = \mathbf{P}_i(T_j < \infty)$, avec T_j le temps d'atteinte de j).

Le temps moyen de retour en i : $m(i) = \mathbf{E}_i(T_i)$.

$1/m(i)$ la fréquence moyenne de retour en i (nulle si $m(i) = \infty$).

Définition

Un état est dit

- transitoire si $\mathbf{P}_i(T_i = \infty) > 0$;
- récurrent si $\mathbf{P}_i(T_i = \infty) = 0$.

Un état récurrent est dit :

- récurrent positif si $\mathbf{E}_i(T_i) < \infty$;
- récurrent nul si $\mathbf{E}_i(T_i) = \infty$.

Classification des états

On a les équivalences suivantes :

Théorème

$$f_{ii} = 1 \ \textit{i.e.} \ i \ \textit{est récurrent} \ \Leftrightarrow \mathbf{P}_i(N_i = \infty) = 1 \Leftrightarrow \mathbf{E}_i(N_i) = \infty.$$

$$f_{ii} < 1 \ \textit{i.e.} \ i \ \textit{est transitoire} \ \Leftrightarrow \mathbf{P}_i(N_i < \infty) = 1 \Leftrightarrow \mathbf{E}_i(N_i) < \infty.$$

Le critère i récurrent $\Leftrightarrow \mathbf{E}_i(N_i) = \infty$ est le critère de la matrice potentiel.

$$\delta_{ij} + \mathbf{E}_i(N_j) = \mathbf{E}(\sum_{n\geq 0} 1_{X_n=j} \,|\, X_0 = i) = \sum_{n\geq 0} \mathbf{P}(X_n = j \,|\, X_0 = i) = \sum_{n\geq 0} p_{ij}(n).$$

La matrice potentiel est $G = \sum_{n\geq 0} \mathbf{P}^n$. Un état i est donc récurrent si et seulement si $G_{ii} = \infty$.

Propriété de classe

Théorème

Si i et j sont deux états de la même classe de communication, alors i et j sont tous les deux soit récurrents, soit transitoires.

Propriété de classe

Théorème

Si i et j sont deux états de la même classe de communication, alors i et j sont tous les deux soit récurrents, soit transitoires.

Propriété de classe : permet de parler de classe récurrente/transitoire et de chaîne récurrente/transitoire pour les chaînes irréductibles.

Propriété de classe

Théorème

Si i et j sont deux états de la même classe de communication, alors i et j sont tous les deux soit récurrents, soit transitoires.

Propriété de classe : permet de parler de classe récurrente/transitoire et de chaîne récurrente/transitoire pour les chaînes irréductibles.

Démonstration. Soient i et j deux états de la même classe de communication. Alors il existe $n, m \in \mathbb{N}$ tels que $p_{ij}(m)p_{ji}(n) = \alpha > 0$. Alors

$$\begin{aligned} p_{ii}(n+\ell+m) &\geq p_{ij}(m)p_{jj}(\ell)p_{ji}(n) = \alpha p_{jj}(\ell) \\ p_{jj}(n+\ell+m) &\geq p_{ji}(n)p_{ii}(\ell)p_{ij}(m) = \alpha p_{jj}(\ell). \end{aligned}$$

Donc $\sum_{\ell\in\mathbb{N}} p_{ii}(\ell)$ et $\sum_{\ell\in\mathbb{N}} p_{jj}(\ell)$ convergent ou divergent tous les deux (en utilisant le critère de la matrice potentiel). $\square$

Classes transitoires

Lemme

Si i et j sont deux états d'une même classe transitoire, alors $\lim_{n\to\infty} p_{ij}(n) = 0$.

Démonstration. $\exists k$ t.q. $p_{ij}(k) = \alpha > 0$ et $p_{ii}(n) \geq p_{ij}(k)p_{ji}(n-k) = \alpha p_{ji}(n-k)$. D'après le critère de la matrice de potentiel, $p_{ii}(n) \to 0$ (la série converge). Donc $p_{ji}(n-k) \to 0$. □

Classes transitoires

Lemme

Si i et j sont deux états d'une même classe transitoire, alors $\lim_{n\to\infty} p_{ij}(n) = 0$.

Démonstration. $\exists k$ t.q. $p_{ij}(k) = \alpha > 0$ et $p_{ii}(n) \geq p_{ij}(k)p_{ji}(n-k) = \alpha p_{ji}(n-k)$. D'après le critère de la matrice de potentiel, $p_{ii}(n) \to 0$ (la série converge). Donc $p_{ji}(n-k) \to 0$. □

Cas des chaînes non irréductibles

Les classes non terminales sont nécessairement transitoires : une fois qu'on soit d'une telle classe, on ne peut plus y revenir.

Si i est un état d'une classe non terminale tel que $p_{ij} > 0$ et j n'est pas dans cette classe, alors $f_{ii} \leq 1 - p_{ij} < 1$.

Mesure invariante canonique

$\{Xn\}$ une CMH irréductible et récurrente ;
$0 \in \mathcal{E}$ et T_0 le temps de retour en 0.

Notation :

- Pour tout $i \in \mathcal{E}$, $Z_i = \sum_{n>0} 1_{X_n=i} 1_{n \leq T_0}$, *le nombre de visites en i entre le temps* 0 *(exclus) et T_0 inclus*.
  En particulier, $Z_0 = 1$.
- $x_i := \mathbf{E}_0(Z_i)$ (on a $x_0 = 1$).

Mesure invariante canonique

$\{Xn\}$ une CMH irréductible et récurrente ;
$0 \in \mathcal{E}$ et T_0 le temps de retour en 0.

Notation :

- Pour tout $i \in \mathcal{E}$, $Z_i = \sum_{n>0} 1_{X_n=i} 1_{n \leq T_0}$, *le nombre de visites en i entre le temps* 0 *(exclus) et T_0 inclus*.
  En particulier, $Z_0 = 1$.
- $x_i := \mathbf{E}_0(Z_i)$ (on a $x_0 = 1$).

Théorème

Pour toute chaîne de Markov irréductible et récurrente, $x = (x_i, i \in \mathcal{E})$ est une mesure invariante telle que $0 < x_i < \infty, i \in \mathcal{E}$.

Mesure invariante canonique

$\{Xn\}$ une CMH irréductible et récurrente ;
$0 \in \mathcal{E}$ et T_0 le temps de retour en 0.

Notation :

- Pour tout $i \in \mathcal{E}$, $Z_i = \sum_{n>0} 1_{X_n=i} 1_{n \leq T_0}$, *le nombre de visites en i entre le temps* 0 *(exclus) et* T_0 *inclus*.
  En particulier, $Z_0 = 1$.
- $x_i := \mathbf{E}_0(Z_i)$ (on a $x_0 = 1$).

Théorème

Pour toute chaîne de Markov irréductible et récurrente, $x = (x_i, i \in \mathcal{E})$ *est une mesure invariante telle que* $0 < x_i < \infty, i \in \mathcal{E}$.

S'il existe une mesure invariante, la chaîne n'est pas nécessairement récurrente ! (e.g. marche aléatoire non symétrique)

Distribution stationnaire et récurrence positive

Théorème

Une chaîne irréductible est récurrente positive si et seulement si elle admet une probabilité stationnaire.

Distribution stationnaire et récurrence positive

Théorème

Une chaîne irréductible est récurrente positive si et seulement si elle admet une probabilité stationnaire.

Démonstration.

$\Rightarrow$ Si une chaîne est irréductible et récurrente, alors elle admet une mesure invariante $x = (x_i),\ x_i = \mathbf{E}_0(Z_i)$. De plus,

$$\sum_{i \in \mathcal{E}} x_i = \sum_{i \in \mathcal{E}} \mathbf{E}_0(Z_i) = \mathbf{E}_0(\sum_{i \in \mathcal{E}} Z_i) = \mathbf{E}_0(T_0).$$

Par définition de la récurrence positive, cette quantité est finie.

Distribution stationnaire et récurrence positive

$\Leftarrow$ Soit une chaîne admettant une probabilité stationnaire π.

On a $\pi = \pi \mathbf{P}^n$ et $\pi_i = \sum_{j \in \mathcal{E}} \pi_j p_{ji}(n)$ pour tout $n \in \mathbb{N}$.

Si la chaîne était transitoire, on aurait $p_{ij}(n) \to 0$ par le critère de la matrice potentiel pour tout i, j, et donc on aurait $\pi_i = 0$. La chaîne est donc récurrente.

Distribution stationnaire et récurrence positive

$\Leftarrow$ Soit une chaîne admettant une probabilité stationnaire π.

On a $\pi = \pi \mathbf{P}^n$ et $\pi_i = \sum_{j \in \mathcal{E}} \pi_j p_{ji}(n)$ pour tout $n \in \mathbb{N}$.

Si la chaîne était transitoire, on aurait $p_{ij}(n) \to 0$ par le critère de la matrice potentiel pour tout i, j, et donc on aurait $\pi_i = 0$. La chaîne est donc récurrente.

Comme la mesure invariante est unique à un multiple près, on a $\sum x_i < \infty$, et donc $\mathbf{E}_0(T_0) < \infty$.

Distribution stationnaire et récurrence positive

$\Leftarrow$ Soit une chaîne admettant une probabilité stationnaire π.

On a $\pi = \pi \mathbf{P}^n$ et $\pi_i = \sum_{j \in \mathcal{E}} \pi_j p_{ji}(n)$ pour tout $n \in \mathbb{N}$.

Si la chaîne était transitoire, on aurait $p_{ij}(n) \to 0$ par le critère de la matrice potentiel pour tout i, j, et donc on aurait $\pi_i = 0$. La chaîne est donc récurrente.

Comme la mesure invariante est unique à un multiple près, on a $\sum x_i < \infty$, et donc $\mathbf{E}_0(T_0) < \infty$.

Ceci est vrai pour tous les états par symétrie. La chaîne est donc récurrente positive.

On a montré en même temps que la récurrence positive est une propriété de classe. $\square$

Temps moyen de récurrence

Soit $\{X_n\}$ une CMH irréductible et récurrente positive de probabilité stationnaire π. On a, $\forall i \in \mathcal{E}$,

$$\pi_i = \frac{1}{\mathbf{E}_i(T_i)}.$$

En effet, d'après les théorèmes et preuves précédents, $\sum_{i \in \mathcal{E}} x_i = \mathbf{E}_0(T_0)$ et $x_0 = 1$. Donc $\pi_0 = 1/\mathbf{E}_0(T_0)$.

Espace d'états fini

Théorème

Une CMH irréductible ayant un nombre fini d'états est récurrente positive.

Espace d'états fini

Théorème

Une CMH irréductible ayant un nombre fini d'états est récurrente positive.

Démonstration.

On montre d'abord la récurrence. Nous avons

$$\sum_{j \in \mathcal{E}} p_{ij}(n) = 1,$$

et donc $\lim_{n\to\infty} \sum_{j\in\mathcal{E}} p_{ij}(n) = 1$.

Si la chaîne était transitoire, on aurait $p_{ij}(n) \to 0$, pour tout i, j, par le critère de la matrice potentiel.

Puisque $|\mathcal{E}| < \infty$, on aurait $\lim_{n\to\infty} \sum_{j\in\mathcal{E}} p_{ij}(n) = 0$, ce qui est en contradiction. Donc la chaîne est récurrente.

La chaîne admet une mesure invariante x telle que $0 < x_i < \infty$, $\forall i$. Puisque $|\mathcal{E}| < \infty$, nous avons $\sum_{i\in\mathcal{E}} x_i < \infty$ et donc la chaîne admet une probabilité stationnaire. Donc la chaîne est récurrente positive. $\square$

Le théorème ergodique

Hypothèse : CMH irréductible et récurrente.

- Mesure invariante μ (unique à un facteur multiplicatif près), $\mu(i) = \mathbf{E}_0(\sum_{n=1}^{T_0} 1_{X_n=i})$.
- $\nu(n) := \sum_{k=1}^{n} 1_{X_k=0}$, le nombre de passages en 0 entre les temps 1 et n.

Théorème

Soit $\{X_n\}$ une CMH, irréductible et récurrente. Pour toute fonction $f : \mathcal{E} \to \mathbb{R}$ vérifiant la condition d'intégrabilité ($\sum_{i\in\mathcal{E}} |f(i)|\mu(i) < \infty$), on a presque sûrement, et pour toute loi initiale :

$$\lim_{n\to\infty} \frac{1}{\nu(n)} \sum_{k=1}^{n} f(X_k) = \sum_{i\in\mathcal{E}} f(i)\mu(i).$$

Le théorème ergodique

Corollaire

Si la chaîne $\{X_n\}$ est irréductible et récurrente positive, alors, presque sûrement et pour toute loi initiale,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f(X_k) = \sum_{i\in\mathcal{E}} f(i)\pi(i).$$

où π est la distribution stationnaire.

Démonstration.

On rappelle qu'on a $\forall i \in \mathcal{E}, \pi(i) = (\sum_{j\in\mathcal{E}} \mu(j))^{-1}\mu(i)$. On a

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f(X_k) = \lim_{n\to\infty} \frac{1}{\nu(n)} \frac{\nu(n)}{n} \sum_{k=1}^{n} f(X_k)$$

Si on applique le théorème à la fonction égale à 1 partout (intégrable, car positive récurrence), on obtient : $\lim_{n\to\infty} \frac{n}{\nu(n)} = \sum_{j\in\mathcal{E}} \mu(j)$. En appliquant le théorème à la fonction f, on obtient le résultat souhaité. $\square$

Le théorème Kolmogorov

Définition

Une CMH irréductible, récurrente positive et apériodique est dite ergodique.

Le théorème Kolmogorov

Définition

Une CMH irréductible, récurrente positive et apériodique est dite ergodique.

Théorème

Si $\{X_n\}$ est une CM ergodique, alors

$$\forall i, j \in \mathcal{E}, \qquad \lim_{n\to\infty} p_{ij}(n) = \pi(j),$$

où π est l'unique distribution stationnaire de la chaîne.

Démonstration : En exercice.

Idée de démonstration :

- Soient $\{X_n^1\}$ et $\{X_n^2\}$ deux CMH, définies sur le même espace des états, et de même matrice de transition $\mathbf{P}$, de période 1, irréductibles et récurrentes positives. Montrer que la chaîne produit $Z_n = (X_n^1, X_n^2)$ est une CMH irréductible et récurrente positive.

Le théorème Kolmogorov

- En déduire que τ, la première date n à laquelle $X_n^1 = X_n^2$ est presque sûrement fini.
- Montrer que le processus couplée $\{X_n\}$:

$$X_n = \begin{cases} X_n^1, n \leq \tau \\ X_n^2, n \geq \tau \end{cases}$$

  est une CMH, et qu'elle a même matrice de distribution $\mathbf{P}$.
- Montrer que $\mathbf{P}(X_n^2 = i) \leq \mathbf{P}(X_n = i) + \mathbf{P}(n < \tau)$.
- Soit μ la distribution initiale de X^1 et ν celle de X^2. Montrer que $|(\nu\mathbf{P}^n - \mu\mathbf{P}^n)_i| \leq \mathbf{P}(n < \tau)$.
- En déduire le théorème de Kolmogorov.

$\square$

Cas périodique

Si la chaîne est de période d, alors on a $p_{ij}(nd) \neq 0$ si i et j sont dans la même classe (et à partir d'un certain rang, c'est vrai pour tout n).

La chaîne $Y_n = X_{nd}$ est irréductible de période 1 et récurrente positive (matrice de transition $\mathbf{P}^d$ restreinte à la classe de communication.

Le théorème ergodique donne : $\sum_{i \in \mathcal{C}} \pi(i) = 1/d$.

Soit $\tilde{\pi}$ la probabilité stationnaire sur la classe de communication $\mathcal{C}$. On a $\tilde{\pi}(i) = d\pi(i)$ et Kolmogorov donne $\lim_{n\to\infty} p_{ij}(n) = d\pi(j)$.

Plan

Matrice fondamentale

La matrice fondamentale d'une CMH ergodique avec un espace d'états fini $\mathcal{E} = \{1, \dots, r\}$ et distribution stationnaire π est la matrice

$$Z = (I - (\mathbf{P} - \Pi))^{-1},$$

où

$$\Pi = \mathbf{1}^t \pi = \begin{pmatrix} \pi(1) & \cdots & \pi(r) \\ \pi(1) & \cdots & \pi(r) \\ \vdots & & \vdots \\ \pi(1) & \cdots & \pi(r) \end{pmatrix}$$

Théorème

$Z = I + \sum_{n \geq 1} (\mathbf{P}^n - \Pi)$. *En particulier,* $\sum_j Z_{ij} = 1$.

Démonstration

On note que

$$\Pi \mathbf{P} = \Pi \quad (\text{car } \pi \mathbf{P} = \pi \text{ et } \Pi = \mathbf{1}^t \pi),$$
$$\mathbf{P} \Pi = \Pi \quad (\text{car } \mathbf{P1}^t = \mathbf{1}^t \text{ et } \Pi = \mathbf{1}^t \pi),$$
$$\Pi^2 = \Pi \quad (\text{car } \Pi = \mathbf{1}^t \pi \text{ et } \pi \mathbf{1}^t = 1).$$

En particulier, pour $k \geq 1$, $\mathbf{P} \Pi^k = \Pi = \Pi^k \mathbf{P}$ et

$$\begin{aligned} (\mathbf{P} - \Pi)^n &= \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} \mathbf{P}^k \Pi^{n-k} \\ &= \mathbf{P}^n + (\sum_{k=0}^{n-1} \binom{n}{k} (-1)^{n-k}) \Pi = \mathbf{P}^n - \Pi \end{aligned}$$

Alors, pour $A = \mathbf{P} - \Pi$,

$$(I - A)(I + A + A^2 + \cdots + A^{n-1}) = I - A^n = I - \mathbf{P}^n + \Pi.$$

Quand $n \to \infty$, $(I - A)(I + \sum_{n \geq 1} A^n) = I$, ce qui montre que $I - (\mathbf{P} - \Pi)$ est inversible, avec inverse

$$I + \sum_{n \geq 1} (\mathbf{P} - \Pi)^n = I + \sum_{n \geq 1} (\mathbf{P}^n - \Pi). \qquad \square$$

Une extension

Théorème

P *la matrice de CMH ergodique avec un espace d'états fini* $\mathcal{E}$*. Soit* b *un vecteur tel que*

$$b\mathbf{1}^t \neq 0,$$

Alors $Z = (I - \mathbf{P} + \mathbf{1}^t b)^{-1}$ *existe et* $\pi = bZ$*.*

Démonstration. Puisque $\pi\mathbf{1}^t = 1$ et $\pi(I - \mathbf{P}) = 0$,

$$\pi(I - \mathbf{P} + \mathbf{1}^t b) = \pi\mathbf{1}^t b = b, \tag{1}$$

et donc pour tout vecteur x tel que

$$(I - \mathbf{P} + \mathbf{1}^t b)x^t = 0,$$

on a $bx^t = 0$ et $(I - \mathbf{P})x^t = 0$. Donc x doit être un vecteur propre à droite, associé à la valeur propre $\lambda_1 = 1$, et donc x est un multiple de $\mathbf{1}$ (par le théorème de Perron-Frobenius). Ceci est compatible avec $bx^t = 0$ et $b\mathbf{1}^t \neq 0$ seulement si $x = 0$. Donc $(I - \mathbf{P} + \mathbf{1}^t b)$ est inversible et (1) implique le résultat.

Théorème de Perron-Frobenius

Matrice $T = (T_{ij})$ est

- positive si $T_{ij} \geq 0, \ \forall i, j$. Notation : $T \geq 0$.
- strictement positive si $T_{ij} > 0, \ \forall i, j$. Notation : $T > 0$.

Def. Une matrice carrée $T \geq 0$ est dite primitive si $\exists k > 0$ tq. $T^k > 0$.

Thm. (Perron-Frobenius) Soit T une matrice positive et primitive. Alors il existe une valeur propre r tq.

1. $r \in \mathbb{R}_+^*$, *i.e.* $r > 0$;
2. à r sont associés des vecteurs propres à gauche et à droite strictement positifs ;
3. $r > |\lambda|$ pour toute valeur propre $\lambda \neq r$;
4. les vecteurs propres associés à r sont uniques à une constante multiplicative près.

Les temps d'atteinte

Notation

- Pour une matrice carrée B on noté $d(B)$ la matrice diagonale qui a la même diagonale que la matrice B.

  En particulier, $(d(\Pi)^{-1})(i,i) = \pi(i)^{-1}$.
- $M = (m_{ij})_{1 \leq i,j \leq r}$, avec $m_{ij} = \mathbf{E}_i(T_j)$.
- $\mathbf{1}^t\mathbf{1}$ est une matrice avec tous les éléments 1.

Théorème

$M = (I - Z + \mathbf{1}^t\mathbf{1}d(Z))d(\Pi)^{-1}$.

Théorème

Soit Z la matrice fondamentale défini à partir d'un vecteur b. Alors pour tout $i \neq j$,

$$\mathbf{E}_i(T_j) = \frac{z_{jj} - z_{ij}}{\pi(j)}.$$

Variance asymptotique

Soit $\{X_n\}$ une CHM avec l'espace d'états fini $\mathcal{E} = \{1, \ldots, r\}$ et $f : \mathcal{E} \to \mathbb{R}$. On va representer f par un vecteur colonne $f = (f(1), \ldots, f(r))$.

Le théorème ergodique implique que $\frac{1}{n}\sum_{k=1}^{n} f(X_k)$ est un estimateur asymptotiquement sans biais de $< f >_\pi := \mathbf{E}_\pi[f(X_0)]$.

Théorème

Pour une distribution initiale quelconque μ,

$$\lim_{n\to\infty} \frac{1}{n} V_\mu(\sum_{k=1}^{n} f(X_k)) = 2 < f, Zf >_\pi - < f, (I + \Pi)f >_\pi,$$

où V_μ dénote que la variance est calculée par rapport à $\mathbf{P}_\mu$.

Intermède

Matrice stochastique P irréductible apériodique

Méthode de puissances :
$z_k = z_{k-1}P$ avec z_0 un vecteur stochastique initial arbitraire

Convergence vers π ; taux de convergence $|\lambda_2|$ (très lent si proche de 1).

Question : utiliser la multiplication vecteur-matrice ou puissances de matrices ?

$$z_k = z_{k-1}P \qquad \text{vs.} \qquad z_k = z_0 P^k$$

Intermède

Matrice stochastique P irréductible apériodique

Méthode de puissances :
$z_k = z_{k-1}P$ avec z_0 un vecteur stochastique initial arbitraire

Convergence vers π ; taux de convergence $|\lambda_2|$ (très lent si proche de 1).

Question : utiliser la multiplication vecteur-matrice ou puissances de matrices ?

$$z_k = z_{k-1}P \qquad \text{vs.} \qquad z_k = z_0 P^k$$

m produits de matrices pour P^{2^m} vs. 2^m produits vecteur-matrice ?

Calcul naïf : utiliser puissances de matrices si

$$nm < 2^m \quad (mn^3 < 2^m n^2)$$

Intermède

Matrice stochastique P irréductible apériodique

Méthode de puissances :
$z_k = z_{k-1}P$ avec z_0 un vecteur stochastique initial arbitraire

Convergence vers π ; taux de convergence $|\lambda_2|$ (très lent si proche de 1).

Question : utiliser la multiplication vecteur-matrice ou puissances de matrices ?

$$z_k = z_{k-1}P \qquad \text{vs.} \qquad z_k = z_0 P^k$$

m produits de matrices pour P^{2^m} vs. 2^m produits vecteur-matrice ?

Calcul naïf : utiliser puissances de matrices si

$$nm < 2^m \quad (mn^3 < 2^m n^2)$$

Attention : si P très large et creux, produit matrice-vecteur : n_z, nombre des éléments non-nuls de P.

Problèmes mémoire pour les produits de matrices.

Théorème de Perron-Frobenius

Matrice $T = (T_{ij})$ est

- positive si $T_{ij} \geq 0$, $\forall i, j$. Notation : $T \geq 0$.
- strictement positive si $T_{ij} > 0$, $\forall i, j$. Notation : $T > 0$.

Def. Une matrice carrée $T \geq 0$ est dite primitive si $\exists k > 0$ tq. $T^k > 0$.

Thm. (Perron-Frobenius) Soit T une matrice positive et primitive. Alors il existe une valeur propre r tq.

1. $r \in \mathbb{R}_+^*$, *i.e.* $r > 0$;
2. à r sont associés des vecteurs propres à gauche et à droite strictement positifs ;
3. $r > |\lambda|$ pour toute valeur propre $\lambda \neq r$;
4. les vecteurs propres associés à r sont uniques à une constante multiplicative près.

Théorème de Perron-Frobenius

Démonstration

(Notation : les vecteurs sont des vecteurs colonnes.)

Pour $\mathbf{x} \geq 0$ avec $\mathbf{x} \neq 0$, on définit

$$r(\mathbf{x}) = \min_j \frac{\sum_i x_i T_{ij}}{x_j},$$

avec $\frac{\sum_i x_i T_{ij}}{x_j} := \infty$ si $x_j = 0$.

Lemme 1. $0 \leq r(\mathbf{x}) < \infty$

Preuve. Consequence directe de l'hypothèse que $\mathbf{x} \geq 0$ et $\mathbf{x} \neq 0$. $\square$

Lemme 2. $r(\mathbf{x})$ est uniformément borné en $\mathbf{x}$, *i.e.* $\exists K \geq 0$ tq. $r(\mathbf{x}) \leq K, \forall x$.

Preuve.

$$\begin{aligned} r(\mathbf{x})x_j &\leq \sum_i x_i T_{ij} \text{ pour tout } j \\ r(\mathbf{x})\mathbf{x}^t &\leq \mathbf{x}^t T \\ r(\mathbf{x})\mathbf{x}^t\mathbf{1} &\leq \mathbf{x}^t T\mathbf{1} \end{aligned}$$

Pour $K = \max_i \sum_j T_{ij}$, on a $T\mathbf{1} \leq K\mathbf{1}$ et

$$r(\mathbf{x}) \leq \frac{\mathbf{x}^t K\mathbf{1}}{\mathbf{x}^t\mathbf{1}} = K.$$

$\square$

Soit
$$r := \sup_{\mathbf{x}\geq 0,\ \mathbf{x}\neq 0} r(\mathbf{x}) = \sup_{\mathbf{x}\geq 0,\ \mathbf{x}\neq 0} \min_j \frac{\sum_i x_i T_{ij}}{x_j}.$$

Lemme 3. $0 < r \leq K$.

Preuve. T est primitive $\Rightarrow$ T ne contient pas de colonne nulle $\Rightarrow r(\mathbf{1}) > 0$. Donc,
$$0 < r(\mathbf{1}) \leq r \leq K.$$
□

On peut normaliser $\mathbf{x}$ sans affecter r, donc
$$0 < r = \sup_{\mathbf{x}\geq 0,\ \mathbf{x}^t\mathbf{x}=\mathbf{1}} r(\mathbf{x}) = \sup_{\mathbf{x}\geq 0,\ \mathbf{x}^t\mathbf{x}=\mathbf{1}} \min_j \frac{\sum_i x_i T_{ij}}{x_j} \leq K.$$

Théorème de Perron-Frobenius

Démonstration

L'ensemble $\{\mathbf{x} \geq \mathbf{0},\ \mathbf{x}^t\mathbf{x} = 1\}$ est un compact de $\mathbb{R}^n$.

La fonction r est semi-continue supérieurement (*i.e.* $\limsup_{x \to x_0} r(x) \leq r(x_0)$).

$\Rightarrow$ r atteint sa borne supérieure.

$\Rightarrow \exists\ \hat{\mathbf{x}} \geq \mathbf{0}, \neq \mathbf{0}$ tq.

$$\sum_i \hat{x}_i T_{ij} \geq r\hat{x}_j, \quad \text{pour tout } j, \tag{2}$$

et avec égalité pour au moins un j.

Lemme 4. $\hat{\mathbf{x}}$ est un vecteur propre à gauche pour T associé à la valeur propre r.

Preuve. Soit $\mathbf{z}^t := \hat{\mathbf{x}}^t T - r\hat{\mathbf{x}}^t \geq \mathbf{0}^t$. Supposons par absurde que $\mathbf{z} \neq \mathbf{0}$. Par hypothèse, T est primitive, et donc $\exists k$ tq. $T^k > 0$. Donc

$$\mathbf{z}^t T^k = \hat{\mathbf{x}}^t T T^k - r\hat{\mathbf{x}}^t T^k = \hat{\mathbf{x}}^t T^k T - r\hat{\mathbf{x}}^t T^k > \mathbf{0}^t.$$

Posons $\mathbf{y}^t = \hat{\mathbf{x}}^t T^k$. Alors

$$\sum_i y_i T_{ij} > r y_j \quad \text{pour tout } j,$$

ce qui est une contradiction avec la définition de r. Donc, $\mathbf{z} = \mathbf{0}$ et $\hat{\mathbf{x}}^t T = r\hat{\mathbf{x}}^t$ □

En itérant $\hat{\mathbf{x}}^t T = r\hat{\mathbf{x}}^t$, on a $\hat{\mathbf{x}}^t T^k = r^k \hat{\mathbf{x}}^t$
En choisissant $T^k > 0$, et comme $\hat{\mathbf{x}} \geq \mathbf{0}, \neq \mathbf{0}$, on a $\hat{\mathbf{x}}^t T^k > \mathbf{0}^t$ et donc $\hat{\mathbf{x}} > \mathbf{0}$.

Théorème de Perron-Frobenius

Démonstration

$r > |\lambda|$ pour toute valeur propre $\lambda \neq r$:

Soit λ une valeur propre de T, c.à.d. pour un $\mathbf{x} \neq \mathbf{0}$ (ayant des valeurs possiblement complexes), $\mathbf{x}^t T = \lambda \mathbf{x}^t$. En prenant le module on a pour $|\mathbf{x}| := (|x_i|)_i$,

$$\sum_i |x_i| T_{ij} \geq |\lambda||x_j|,$$

donc $|\lambda| \leq r$ (par définition de r).

Supposons par l'absurde que $|\lambda| = r$, on a alors $|\mathbf{x}|^t T \geq r|\mathbf{x}|^t$. Par le raisonnement similaire que pour (2), $|\mathbf{x}|^t T = r|\mathbf{x}|^t > \mathbf{0}$.
On a

$$|\sum_i x_i T_{ij}^k| = |\lambda^k||x_j| = \sum_i |x_i| T_{ij}^k$$

et $|x_j| > 0$, $\forall j$.
Pour deux complexes $\mathbf{z}, \mathbf{z}' \neq 0$, si $|\mathbf{z} + \mathbf{z}'| = |\mathbf{z}| + |\mathbf{z}'|$ alors $\mathbf{z}$ et $\mathbf{z}'$ sont colinéaires. Donc tous les x_j sont colinéaires et

$$\sum_i |x_i| T_{ij} = \lambda |x_j|.$$

Comme $|x_j| > 0$ pour tout j, λ est réel positif et donc $\lambda = r$.

Les vect. propres associés à r sont uniques à une constante multip. près.

Soit $\mathbf{y} \neq \mathbf{0}$ un vecteur propre à gauche associé à r. Pour tout c, le vecteur $\eta = \hat{\mathbf{x}} - c\mathbf{y}$ (si pas nul) est un vect. propre à gauche associé à r

$\Rightarrow$ Si $\mathbf{y}$ n'est pas un multiple de $\hat{\mathbf{x}}$, on peut choisir c tq. $\eta \neq \mathbf{0}$ mais que certaines composantes de η soient nulles.

Or par argument précédent, $|\eta|$ est un vecteur propre associé à r et $|\eta| > 0$.

En contradiction avec le fait que $\mathbf{y}$ et $\hat{\mathbf{x}}$ ne soient pas multiples.

Vecteurs propres à droite

Tous les arguments peuvent être répétés pour les vecteurs propres à droite, pour une valeur propre r'. Le point 3 montre que $r' = r$, en observant que le spectre d'une matrice et sa transposé sont égaux.