Année universitaire 2003-2004

UNIVERSITÉ D'ORLÉANS

Olivier GARET

Chaînes de Markov et algorithmes stochastiques

Table des matières

Chapitre 1

Chaînes de Markov

1.1 Dynamique markovienne

Définition: Soit S un ensemble fini ou dénombrable, ν une mesure de probabilité sur S et $P = (p_{i,j})_{(i,j)\in S\times S}$ une matrice à coefficients positifs. Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires définies sur un espace $(\Omega, \mathcal{F}, P)$. On dit que la suite $(X_n)_{n\geq 0}$ est une *chaîne de Markov* de loi initiale ν et de matrice de passage M si l'on a, pour tout entier $n \geq 1$ et toute suite $x_0, \dots x_n$ d'éléments de S :

$$P(X_0 = x_0, X_1 = x_1, \dots X_n = x_n) = \nu(x_0) \prod_{i=0}^{n-1} p_{x_i,x_{i+1}}.$$

Exemple : une suite $(X_n)_{n\geq 0}$ de variables aléatoires indépendantes de même loi ν à valeurs dans S dénombrable est une chaîne de Markov. En effet, il suffit de poser pour $(i,j) \in S \times S$ $p_{i,j} = \nu(j)$.

Propriétés :

1. Si $(X_n)_{n\geq 0}$ est une chaîne de Markov de matrice de passage M et que $x_0, \dots, x_{n-1}$ sont tels que $P(X_0 = x_0, X_1 = x_1, \dots X_{n-1} = x_{n-1})$, alors $P(X_n = x_n)|X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}) = p_{X_{n-1},x_n}$. Autrement dit,
$$P(X_n = x_n|X_0, \dots, X_{n-1}) = p_{X_{n-1},x_n}. \tag{1.1}$$
Cela signifie que toute l'information que $X_0, \dots, X_{n-1}$ peuvent nous apporter sur X_n est comprise dans X_n.
2. (1.1) implique que $P(X_n = x_n|X_{n-1}) = p_{X_{n-1},x_n}$

Qu'est ce concrètement, qu'une chaîne de Markov ? On va voir que c'est une suite de réalisations, au cours du temps, des états d'un système soumis à des transformations aléatoires, la suites des transformations est une suite de transformations indépendantes, de même loi. Évidemment, le résultat de la transformation dépend de la transformation choisie et de l'état du système avant la transformation.

Lemme 1. *Soit S un ensemble fini ou dénombrable, ν une loi sur S et θ une mesure sur $S^S = \mathcal{F}(S, S)$.*

Soit $(f_n)_{n\geq 1}$ une suite de v.a.i.i.d. de loi θ et X_0 une variable aléatoire de loi μ indépendante de $(f_n)_{n\geq 1}$. On définit $(X_n)_{n\geq 1}$ par

$$\forall n \geq 0 \quad X_{n+1} = f_{n+1}(X_n)$$

Alors $(X_n)_{n\geq 0}$ est une chaîne de Markov de loi initiale ν et de matrice de transition M, où M est définie par

$$\forall (i,j) \in S \times S \quad m_{i,j} = \theta(\{f \in S^S; f(i) = j\}).$$

Démonstration. Soit $A \subset S^{\{0,\dots,n\}}$.

$$\begin{aligned}
& P(\{(X_0,\dots,X_n) \in A\} \cap \{X_n = i\} \cap \{X_{n+1} = j\}) \\
= \; & P(\{(X_0,\dots,X_n) \in A\} \cap \{X_n = i\} \cap \{f_{n+1}(i) = j\}) \\
= \; & P(\{(X_0,\dots,X_n) \in A\} \cap \{X_n = i\}) P(f_{n+1}(i) = j) \\
= \; & P(\{(X_0,\dots,X_n) \in A\} \cap \{X_n = i\}) P(f_{n+1} \in S \times \dots \{j\} \times \dots S) \\
= \; & P(\{(X_0,\dots,X_n) \in A\} \cap \{X_n = i\}) \theta(S \times \dots \{j\} \times \dots S) \\
= \; & P(\{(X_0,\dots,X_n) \in A\} \cap \{X_n = i\}) m_{i,j}
\end{aligned}$$

□

Exemple : la marche de l'ivrogne (ou marche aléatoire sur $\mathbb{Z}$)
Un ivrogne sort du café passablement éméché. À chaque pas, il prend une décision (enfin, si tant est que cela lui soit possible...) : aller à gauche, ou aller à droite. Si on repère par X_n sa position dans la rue au temps n, on a $S = \mathbb{Z}$, $X_{n+1} = f_{n+1}(X_n)$, où f_n est une suite de translations indépendantes : $P(f_n = (x \mapsto x+1)) = P(f_n = (x \mapsto x-1)) = 1/2$.

Comme on va le voir, ce procédé permet de fabriquer toutes les chaînes de Markov.

1.2 Matrice stochastique

Définition: Soit S un ensemble dénombrable et $P = (p_{i,j})_{(i,j)\in S\times S}$ une matrice à coefficients positifs. On dit que P est une matrice stochastique si on a

$$\forall i \in S \quad \sum_{j\in S} p_{i,j} = 1.$$

1.2.1 Existence des chaînes de Markov

Théorème 1. *Soit S un ensemble dénombrable, $P = (p_{i,j})_{(i,j)\in S\times S}$ une matrice stochastique et ν une mesure de probabilité sur S. Alors, on peut construire une chaîne de Markov de loi initiale ν et de matrice de passage P.*

Démonstration. Définissons une mesure θ_P sur S^S par

$$\theta_P = \bigotimes_{i\in S} \mu_i,$$

où μ_i est la mesure sur S définie par $\mu_i(j) = p_{i,j}$. Alors θ_P vérifie $\theta_P(S \times \dots \{j\} \times \dots S) = p_{i,j}$ et il suffit d'appliquer le lemme précédent. □

Lorsque la matrice P est fixée, on note souvent $\mathbb{P}_\nu$ une probabilité sous laquelle $(X_n)_{n\geq 0}$ est une chaîne de Markov de matrice de transition P telle que la loi de X_0 sous P_ν est ν. De même, on note $\mathbb{E}_\nu$ l'espérance correspondante. Dans le cas où la loi initiale est une masse de Dirac, on écrit simplement $\mathbb{P}_i$ (resp. $\mathbb{E}_i$) au lieu de $\mathbb{P}_{\delta_i}$ (resp. $\mathbb{E}_{\delta_i}$).

1.2.2 Puissances des matrices stochastiques

Théorème 2. *Soit (X_n) une chaîne de Markov de matrice de transition P et de loi initiale $\mathbb{P}_{X_0} = \nu$. Alors, la loi μ_n de la chaîne au temps n s'écrit $\mu_n = \nu P^n$, où on a écrit ν et μ_n comme des vecteurs lignes.*

Démonstration. Il suffit de montrer que $\mu_{n+1} = \mu_n P$, puis procéder par récurrence sur n. D'après le principe de partition, on a

$$\begin{aligned}
\mu_{n+1}(j) &= \mathbb{P}_\nu(X_{n+1} = j) \\
&= \sum_{i\in S} \mathbb{P}_\nu(X_n = i, X_{n+1} = j) \\
&= \sum_{i\in S} \mathbb{P}_\nu(X_n = i,)p_{i,j} \\
&= \sum_{i\in S} \mu_n(i) p_{i,j} \\
&= (\mu_n M)(j)
\end{aligned}$$

□

1.3 Graphe associé à une matrice stochastique

Soit $P = (p_{i,j})_{(i,j)\in S\times S}$ une matrice stochastique. On peut associer à la matrice P (où aux chaînes de Markov correspondantes) un graphe orienté $G = (S, A)$ avec
$$A = \{(x, y) \in S \times S; p_{i,j} > 0\}.$$
Considérons une chaîne de Markov associée à la matrice stochastique P avec la condition initiale déterministe x_0, autrement dit $\nu = \delta_{x_0}$ et notons $\mathbb{P}_{x_0}$ la mesure de probabilité correspondante Alors, comme

$$\mathbb{P}_{x_0}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \prod_{i=0}^{n-1} p_{x_i,x_{i+1}},$$

il est clair que $\mathbb{P}_{x_0}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n)$ est non nul si et seulement si $(x_0, x_1, \dots, x_n)$ constitue un chemin dans le graphe G.

D'après le principe de partition, on a pour une chaîne de Markov avec une loi initiale ν

$$P(\nu)(X_1 = x_1, X_n = x_n) = \sum_{(x_0,\dots x_{n-1})\in S^n} \mathbb{P}_i(X_0 = x_0, X_2 = X_2, \dots X_{n-1} = x_{n-1}, X_n = x_n). \tag{1.2}$$

En particulier, si l'on pose $p^{(n)}_{i,j} = \mathbb{P}_i(X_n = j)$, on a

$$p^{(n)}_{i,j} = \sum_{x\in S^{n-1}} \mathbb{P}_i(X_1 = x_1, X_2 = X_2, \dots X_{n-1} = x_{n-1}, X_n = j).$$

Donc $p^{(n)}_{i,j} > 0$, autrement dit il est possible d'aller en n étapes de l'état i à l'état j si et seulement si on peut trouver dans le graphe G un chemin de longueur n allant de i à j.

On en déduit que

$$\mathbb{P}_i(\exists n > 0 X_n = j) = \mathbb{P}_i(\underset{i\geq 1}{\cup} \{X_n = j\}),$$

qui représente la probabilté que, partant de i, on puisse arriver à j, est non nulle si et seulement si il existe dans le graphe G un chemin allant de i à j.

Si il y a à la fois un chemin de i vers j et un chemin de j vers i, on dit que les états i et j communiquent et on écrit $i \leftrightarrow j$.

Si tous les états communiquent, on dit que la chaîne de Markov est *irréductible*.

On appelle *période* d'un état x d'une chaîne de Markov et on note $d(x)$ le pgcd (plus grand commun diviseur) des longueurs des circuits du graphe G contenant x. Lorsque la période est 1, on dit que l'état x est *apériodique*.

Lemme 2. *Si deux états communiquent, alors ils ont même période.*

Démonstration. Soient i, j avec $i \leftrightarrow j$. Soit γ un chemin de i à j, γ' un chemin de j à i. Soit $\mathcal{C}$ un circuit quelconque (éventuellement vide) contenant j . $\gamma - \gamma'$ et $\gamma - \mathcal{C} - \gamma'$ sont deux circuits contenant i. Donc $d(i)$ divise leurs longueurs ainsi que la différence de leurs longueurs, soit la longueur de $\mathcal{C}$. Ainsi $d(i)$ divise les longueurs de tous les circuits contenant j, donc divise leur pgcd, soit $d(j)$. De la même manière, on montre que $d(j)$ divise $d(i)$, d'où $d(i) = d(j)$. □

Définition: Si une chaîne irréductible a ses états de période 1, on dit qu'elle est apériodique.

Les lemme suivant se révélera très utile par la suite

Lemme 3. *Soit x un état de période 1. Il existe un entier $N(x)$ tel que pour tout $n \geq N(x)$ le graphe associé à la chaîne de Markov possède un circuit de longueur n contenant x*

Soit A l'ensemble des valeurs de n telles que le graphe associé à la chaîne de Markov possède un circuit de longueur n contenant x. Il est clair que A est stable par addition (concaténation des circuits). Il existe $p \geq 1$ et $n_1, n_2, \dots, n_p$ tels que le pgcd de $n_1, n_2, \dots, n_p$ soit 1. D'apès le lemme de Bezout, il existe des relatifs $a_1, \dots a_p$ tels que $1 = \sum_{k=1}^{p} a_k n_k$. Posons $P = \sum_{p:a_p>0} a_p n_p$ et $N = \sum_{p:a_p<0} (-a_p) n_p$. On a $P \in A, N \in A$ et $1 = P - N$. Soit $n \geq N(N-1)$. On peut écrire $n = bN + r$, avec $b \geq N - 1$ et $r \in \{0, 1, \dots N-1\}$. On a $n = bN + r = bN + r(P - N) = rP + (b - r)N \in A$ car $b - r \in \mathbb{N}$ et A est stable par addition.

Corollaire 1. *Si x est un état de période 1 et qu'il existe un chemin de longueur $d(x, y)$ allant de x à y, alors pour tout $n \geq N(x, y) = N(x) + d(x, y)$, il existe un chemin de longueur n allant de x à y.Ainsi, si P est la matrice associée, $P^n(x, y) > 0$.*

Démonstration. Il suffit de concaténer le chemin allant de x à x avec un chemin allant de x à y. □

Corollaire 2. *Si une chaîne de Markov est irréductible, apériodique, à valeurs dans un ensemble fini S, alors il existe un entier N tel que pour tout $n \geq N$ et tout couple (i, j), il existe un chemin de longueur n allant de i à j. Ainsi, si P est la matrice associée, P^n est à coefficients strictement positifs.*

Démonstration. Il suffit de prendre $N = \max(N(x), x \in S) + \mathrm{diam}(G)$. □

1.4 Exercices

1. *Chaîne à deux états.* Soit $\{X_n : n \geq 0\}$ une chaîne de Markov à valeurs dans $\{0, 1\}$ et de probabilité de transition :

$$P := \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}, \ 0 \leq \alpha, \beta \leq 1.$$

   (a) Montrer que pour $(\alpha, \beta) \neq (0, 0)$:

$$P^n = \frac{1}{\alpha+\beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{(1-\alpha-\beta)^n}{\alpha+\beta} \begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix}.$$

   Que se passe-t-il lorsque $\alpha = 0$ ou $\beta = 0$ ou $\alpha = \beta = 0$? On supposera pour la suite de l'exercice que $(\alpha, \beta) \neq (0, 0)$.

   (b) Vérifier (par récurrence) que, pour toute loi initiale μ :

$$\mathrm{P}_\mu(X_n = 0) = \frac{\beta}{\alpha+\beta} + (1-\alpha-\beta)^n \left(\mu(0) - \frac{\beta}{\alpha+\beta}\right).$$

   (c) Si $(\alpha, \beta) \neq (1, 1)$, montrer que $\{X_n : n \geq 0\}$ converge en loi vers une variable aléatoire de loi ν. Que vaut ν ? On supposera pour la suite de l'exercice que $(\alpha, \beta) \neq (1, 1)$.

   (d) (*Mesure stationnaire*) Prouver que, pour tout $n \in \mathbb{N}$,

$$\mathrm{P}_\nu(X_n \in A) = \nu(A).$$

2. *Représentation canonique et simulation des chaînes de Markov.*

   (a) Soit $(Z_n)_{n\geq 1}$ une suite de vaiid à valeurs dans F, soit $g : E \times F \to E$ et soit X_0 une variable aléatoire à valeurs dans E indépendante de $(Z_n)_{n\geq 1}$. Montrer que la suite $(X_n)_{n\geq 0}$ définie par $X_{n+1} = g(X_n, Z_{n+1})$ est une chaîne de Markov homogène. Donner sa matrice de transition.

(b) On suppose qu'on dispose d'un générateur de nombres aléatoire de loi uniforme sur $[0,1]$, noté 'rand'. Soit μ une mesure de probabilité sur $\mathbb{N}$. Donner un algorithme pour générer des nombres aléatoires suivant la loi μ.

(c) Soit $P = (p_{i,j})$ une matrice de transition sur $\mathbb{N}$. On note $s_{i,k} = \sum_{j=0}^{k} p_{i,j}$. Soit $(Z_n)_{n\geq 1}$ une suite de vaiid de loi uniforme sur $[0,1]$ et X_0 une variable aléatoire à valeurs dans $\mathbb{N}$ indépendante de $(Z_n)_{n\geq 1}$. On construit la suite $(X_n)_{n\geq 0}$ par récurrence de la façon suivante :

$$\text{si } X_n(\omega) = i \text{ et } Z_{n+1}(\omega) \in]s_{i,j}, s_{i,j+1}] \text{ alors } X_{n+1} = j.$$

Montrer que la suite $(X_n)_{n\geq 0}$ ainsi définie est une chaîne de Markov homogène. Donner sa matrice de transition.

(d) Application. Comment simuler une chaîne de Markov homogène de matrice de transition $P = (p_{i,j})$? Ecrire un algorithme explicite si

$$P = \begin{pmatrix} 0.25 & 0.5 & 0.25 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{pmatrix}.$$

3. *La ruine du joueur* Un joueur possédant une fortune de a unités joue à pile ou face jusqu'à ce qu'il ait fait sauter la banque ou qu'il soit ruiné. Les réserves de la banque sont de b unités. Chaque victoire rapporte une unité et chaque défaite en coûte une. On suppose que les lancers sont indépendants et que la probabilité de gain de la banque est $p = 1 - q$. On veut déterminer la probabilité p_g que la banque résiste.
On note $(X_n)_{n\geq 1}$ une suite de v.a.i.i.d. de loi $p\delta_1 + q\delta_{-1}$, puis $S_n = \sum_{k=1}^{n} X_k$ et $T = \inf\{n \geq 0; S_n = -b \text{ ou } S_n = a\}$. Si l'on pose $S'_n = S_{n\wedge T}$, il est aisé de constater que S'_n représente la suite des gains relatifs de la banque.

(a) Montrer que S'_n est une chaîne de Markov homogène à espace d'états $E = \{-b, \dots, a\}$ dont on déterminera la loi initiale et la matrice de transition.

(b) Considérons les chaînes de Markov ayant la même matrice de transition que $(S'_n)_{n\geq 0}$ Montrer que la suite $(u_n)_{-b\leq n\leq a}$ définie par

$$u_n = \mathbb{P}_n(\{\text{la banque résiste}\})$$

vérifie la récurrence linéaire

$$pu_{n+1} - u_n + qu_{n-1} = 0.$$

Que valent u_a et u_{-b} ?

(c) Résoudre l'équation de récurrence et en déduire

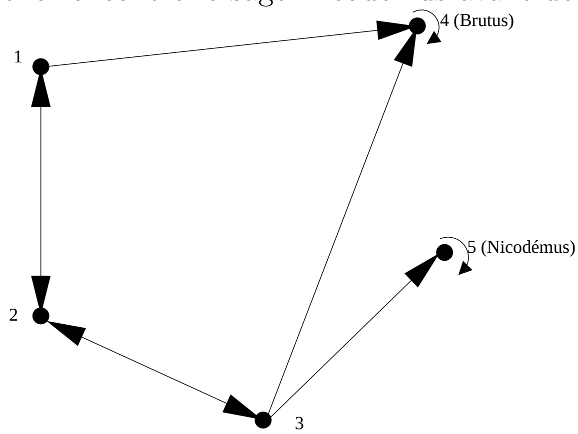
$$p_g = \frac{(\frac{q}{p})^b - 1}{(\frac{q}{p})^{a+b} - 1}. \tag{1.3}$$

4. *Le joueur inruinable* Le problème est le même que le précédent, a ceci près que l'on suppose maintenant que le joueur est infiniment riche. On cherche toujours la probabilité que la banque résiste (ce qui ne signifie pas ici que le joueur est ruiné).
   Intuitivement, il suffit de faire tendre a vers $+\infty$ dans la formule (1.3), le tout étant de le justifier...

   Indications : poser $T' = \inf\{n; S_n \le -b\}$ et, pour tout $a > 0$, $U_a = \inf\{n; S_n \ge a\}$ et $G^a = \{U_a \le T'\}$, puis montrer que

$$\lim_{a \to +\infty} \mathbb{1}_{G^a} = \mathbb{1}_{\{T' = +\infty\}}.$$

5. Madame Brisby dans le labyrinthe Madame Brisby s'est perdue dans le labyrinthe que forment les galeries où vivent les rats de Nim. Quelle est la probabilité qu'elle rencontre le sage Nicodémus avant de croiser



   le belliqueux Rufus ?
6. Soit M la matrice d'une chaîne de Markov. Montrer que si $m_{i,i} > 0$, alors l'état i est apériodique. Qu'en déduire pour une chaîne irréductible ?
7. Soit a un entier supérieur ou égal à 2, $(D_n)_{n \ge 1}$ une suite de variables aléatoires i.i.d. à valeurs dans $\mathbb{Z}/a\mathbb{Z} \times \mathbb{Z}/a\mathbb{Z}$ vérifiant $P(D_1 = (0,1)) = P(D_1 = (1,0)) = \frac{1}{2}$. On pose

$$S_n = S_0 + \sum_{k=1}^{n} D_k.$$

   Montrer que (S_n) est une chaîne de Markov . Est-elle irréductible, apériodique ?

8. *Madame Brisby II*
   Soit $f : E \to \mathbb{C}$ telle que $\forall x \in A \quad f(x) = 0$.
   On pose $F = \sum_{k=1}^{+\infty} f(X_k)$.
   Montrer que $\mathbb{E}|F| \le (\mathbb{E}T - 1)\|f\|_\infty$.
   Montrer l'identité
   $$(I-N)\begin{pmatrix} \mathbb{E}_1 F \\ \mathbb{E}_2 F \\ \mathbb{E}_3 F \end{pmatrix} = \begin{pmatrix} \mathbb{E}_1 f(X_1) \\ \mathbb{E}_2 f(X_1) \\ \mathbb{E}_3 f(X_1) \end{pmatrix},$$
   En déduire $\mathbb{E}_1 T, \mathbb{E}_2 T, \mathbb{E}_3 T$.
9. *Évolution d'un genotype avec fixation*
   Nous travaillons ici sur une population de taille fixe formée de $2N$ gènes. Il y a deux types de gènes possibles : le type "a" et le type "A". Chacun des gènes au temps $n+1$ est engendré par deux des $2N$ gènes présents au temps N. Son type est celui d'un de ses deux parents (choisi au hasard).
   On considère la variable aléatoire X_n égale au nombre d'individus de type "A" dans la population à l'étape n.
   (a) Montrer que X_n est une chaîne de Markov à valeurs dans $E = \{0, \dots, 2N\}$.
   (b) Montrer que la loi de X_{n+1} sachant $X_n = k$ est une loi binomiale de paramètre $2N$ et $(k/2N)$.
   (c) Montrer que $(X_n)_{n\ge 0}$ converge presque sûrement vers une variable aléatoire X_∞.
   (d) Déterminer la loi de X_∞ en fonction de la loi de X_0. (On pourra remarquer que la suite $(\mathbb{E}X_n)_{n\ge 1}$ est constante.)
10. *L'image d'une chaîne de Markov n'est pas (toujours) une chaîne de Markov.* On considère la chaîne de Markov (X_n) sur $E = \{0,1,2\}$ de transition $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$ et de loi initiale $\pi_0 = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Soit $f : E \to \{0,1\}$ telle que $f(0) = f(1) = 0$, $f(2) = 1$. Montrer que $(f(X_n))$ n'est pas une chaîne de Markov.
11. *L'image d'une chaîne de Markov peut être une chaîne de Markov.* Soit (X_n) une chaîne de Markov sur un ensemble dénombrable E de matrice de transition P. Soit ψ une application surjective de E dans un ensemble F telle que
   $$\forall z \in F\ ,\ \forall x, y \in E \quad \psi(x) = \psi(y) \Rightarrow P(x, \psi^{-1}(z)) = P(y, \psi^{-1}(z))\ .$$

Montrer que la suite (Y_n) définie par $Y_n = \psi(X_n)$ est une chaîne de Markov et déterminer sa matrice de transition. Montrer que si π est une probabilité stationnaire pour la chaîne (X_n) alors l'image de π par ψ est stationnaire pour (Y_n).

Chapitre 2

Récurrence et mesures invariantes

2.1 Temps d'arrêt et propriété de Markov forte

On dit que variable aléatoire T à valeurs dans $\mathbb{N} \cup \{+\infty\}$ est un *temps d'arrêt* adapté à la suite $X_0, X_1, \dots, X_n, \dots$ si pour tout $n \in \mathbb{N}$, l'événement $T \leq n$ est $\sigma(X_0, X_1, \dots X_n)$-mesurable.

En d'autres termes pour tout n , il existe un borélien B de $\mathbb{R}^{n+1}$ tel que $\{T = n\} = \{(X_0, X_1, \dots, X_n) \in B\}$.

Comme $\{T = n\} = \{T \leq n\} \backslash \{T \leq n-1\}$, il s'ensuit que $T = n$ est également $\sigma(X_0, X_1, \dots X_n)$-mesurable.

Exemple : si $X_0, \dots, X_n$ est une suite de variables aléatoires à valeurs dans S et A un borélien de S, alors

$$T_A = \inf\{n \geq 1; X_n \in A\}$$

est un temps d'arrêt.

Preuve : $\{T_A \leq n\} = \cup_{k=1}^{n} \{X_k \in A\}$.

Définition: On dit qu'un événement A se produit avant T si pour tout n, l'événement $A \cap \{T \leq n\}$ est $\sigma(X_0, X_1, \dots X_n)$-mesurable.

Théorème 3. *Soit $(X_k)_{k\geq 0}$ une chaîne de Markov de matrice de passage P et T un temps d'arrêt adapté à cette suite. Soit A un événement se produisant avant le temps T. Conditionnellement à l'événement $\{T < +\infty\} \cap A$, la suite $(X_{T+k})_{k\geq 0}$ est une chaîne de Markov de matrice de passage P.*

Démonstration. Comme être une chaîne de Markov est une propriété de la loi, on peut supposer que $(X_n)_{n\geq 0}$ est obtenue par le procédé décrit plus haut : $X_{n+1} = f_{n+1}(X_n)$ où $(f_n)_{n\geq 1}$ est une suite de variables aléatoires indépendantes de loi θ_M indépendante de X_0.

Posons $Y_k = X_{T+k}$ si $T < +\infty$ et $Y_k = X_k$ sinon. De même posons $g_n = f_{n+k}$ si $T < +\infty$ et $g_k = f_k$ sinon. Il est facile de voir que $(Y_k)_{k\geq 0}$ vérifie la récurrence $Y_{n+1} = g_{n+1}(Y_n)$

Soit p un entier et B un borélien de $\mathcal{F}(S,S)^{\mathbb{N}_*}$. On a

$$\begin{aligned} P(T=p,A,g\in B) &= P(T=p,A,f_{p+.}\in B) \\ &= P(T=p,A)P(f_{p+.}\in B) \\ &= P(T=p,A)P(f_{p+.}\in B) \end{aligned}$$

car comme l'événement $\{T=p\}\cap A$ est $\sigma(X_0,f_1,\dots,f_p)$-mesurable, il est indépendant de l'événement $\{f_{p+.}\in B\}$ qui est $\sigma(f_{p+1},f_{p+2},)$-mesurable. Maintenant la loi de $f_{p+.}$ est la même loi que celle de f : c'est $\Theta_M^{\otimes\mathbb{N}^*}$. On a donc $P(T=p,A,g\in B) = P(T=p,A)P(f\in B)$. En faisant la somme pour p variant de 1 à $+\infty$, on obtient

$$P(T<+\infty,A,g\in B) = P(T<+\infty,A)P(f\in B)$$

Autrement dit , sachant $\{T<+\infty\}\cap A$ $(g_n)_{n\geq 1}$ $(g_n)_{n\geq 1}$ est une suite de variables aléatoires indépendantes de loi θ_M indépendante de A, donc $(X_{T+k})_{k\geq 0}$ est une chaîne de Markov de matrice de passage P. □

Remarque : une constante étant un temps d'arrêt, le théorème précédent s'applique lorsque T est une constante. Dans ce cas, le résultat porte simplement le nom de *Propriété de Markov*.

2.2 Classification des états

Définition: Soit $P=(p_{i,j})_{(i,j)\in S\times S}$ une matrice stochastique. Pour $i\in S$, on considère une chaîne de Markov $(X_n)_{n\geq 0}$ partant de l'état i et de matrice de passage P. On pose $T_i = \inf\{n\geq 1; X_n = i\}$. Si $\mathbb{P}_i(T_i<+\infty)=1$, on dit que l'état i est *récurrent*. Inversement, si $\mathbb{P}_i(T_i<+\infty)<1$, on dit que l'état i est *transient*.

Théorème 4. *Soit $P=(p_{i,j})_{(i,j)\in S\times S}$ une matrice stochastique. Pour $i\in S$, on considère une chaîne de Markov $(X_n)_{n\geq 0}$ partant de l'état i et de matrice de passage P. On pose $T_i=\inf\{n\geq 1; X_n=i\}$ et $N_i=\sum_{k=1}^{+\infty}\mathbb{1}_{\{i\}}(X_i)$. N_i représente le nombre de passage de la chaîne en i à partir de l'instant 1.*

- *Si i est transient, alors $1+N_i$ suit la loi géométrique de paramètre $1-\mathbb{P}_i(T_i<+\infty)$. En particulier N_i est presque sûrement fini et intégrable.*

– *Si i est récurrent, alors N_i est presque sûrement infinie. En particulier $\mathbb{E}N_i = +\infty$.*

Démonstration. Si $T_i < +\infty$ (ou de manière équivalente si $N_i > 0$, on a

$$N_i = 1 + \sum_{k=1}^{+\infty} 1\!\!1_{\{i\}}(X_{T+i}).$$

Soit k un entier positif ou nul. On a

$$\mathbb{P}_i(N_i \geq k{+}1) = \mathbb{P}_i(N_i \geq k{+}1, T_i < +\infty) = \mathbb{P}_i(T_i < +\infty)\mathbb{P}_i(\sum_{k=1}^{+\infty} 1\!\!1_{\{i\}}(X_{T+i} \geq k | T_i < +\infty).$$

Or T_i est un temps d'arrêt. Donc, d'après la propriété de Markov forte, sachant $T_i < +\infty$, $(X_{T+i})_{i\geq 0}$ a la loi d'une chaîne de Markov commençant en i et de matrice de transition P, c'est à dire la même loi que $(X_i)_{i\geq 0}$. On en déduit

$$\mathbb{P}_i(N_i \geq k+1) = \mathbb{P}_i(T_i < +\infty)\mathbb{P}_i(N_i \geq k).$$

Par récurrence, on en déduit

$$\mathbb{P}_i(N_i \geq k) = \mathbb{P}_i(T_i < +\infty)^k$$

D'après le théorème de continuité séquencielle décroissante, on a $P(N_i = +\infty) = \lim_{k\to+\infty} \mathbb{P}_i(N_i \geq k)$. Cette limite vaut donc 0 si i est transient, 1 si i est récurrent. Pour $k \geq 1$, on a

$$\begin{aligned} \mathbb{P}_i(1 + N_i = k) &= \mathbb{P}_i(1 + N_i \geq k) - P(N_i \geq k) \\ &= \mathbb{P}_i(T_i < +\infty)^{k-1} - \mathbb{P}_i(T_i < +\infty)^k \\ &= \mathbb{P}_i(T_i < +\infty)^{k-1}(1 - \mathbb{P}_i(T_i < +\infty)), \end{aligned}$$

ce qui montre que $1 + N_i$ suit bien une loi géométrique de paramètre $1 - \mathbb{P}_i(T_i < +\infty)$. De plus

$$\mathbb{E}_i N_i = \sum_{k=1}^{+\infty} \mathbb{P}_i(N_i \geq k) = \sum_{k=1}^{+\infty} = \mathbb{P}_i(T_i < +\infty)^k = \frac{\mathbb{P}_i(T_i < +\infty)}{1 - \mathbb{P}_i(T_i < +\infty)} < +\infty.$$

□

Corollaire 3. *Un état i est transient si et seulement si*

$$\sum_{k=1}^{+\infty} \mathbb{P}_i(X_k = i) < +\infty.$$

Démonstration. D'après le théorème précédent, i est transient si et seulement si N_i est intégrable sous $\mathbb{P}_i$. Or

$$\begin{aligned}\mathbb{E}_i N_i &= \mathbb{E}_i \sum_{k=1}^{+\infty} 1\!\!1_{\{i\}}(X_k) \\ &= \sum_{k=1}^{+\infty} \mathbb{E}_i 1\!\!1_{\{i\}}(X_k) \\ &= \sum_{k=1}^{+\infty} \mathbb{P}_i(X_k = i)\end{aligned}$$

(On utilise le théorème de Tonelli pour échanger la somme et l'espérance) Ceci achève la preuve. □

Corollaire 4. *Soient i et j deux états d'une chaîne de Markov de matrice de transition $P = (p_{i,j})_{(i,j)\in S\times S}$. On suppose que i et j communiquent. Alors i et j sont tous les deux transients ou tous les deux récurrents*

Démonstration. Soit n, p tels que $p_{i,j}^{(n)} > 0$ et $p_{j,i}^{(p)} > 0$. Pour tout $k \geq 0$, on a

$$p_{j,j}^{(n+p+k)} \geq p_{j,i}^{(p)} p_{i,i}^{(k)} p_{i,j}^{(n)}.$$

Ainsi, si la série de terme général $p_{i,i}^{(k)}$ diverge, la série de terme général $p_{j,j}^{(k)}$ aussi. Comme les rôles de i et j sont symétriques, les deux séries sont de même nature. Comme $p_{i,i}^{(k)} = \mathbb{P}_i(X_k = i)$, le résultat découle du corollaire précédent. □

Corollaire 5. *Considérons une chaîne de Markov irréductible de matrice de transition $P = (p_{i,j})_{(i,j)\in S\times S}$ et pour tous les $i \in S$, notons $\mathbb{P}_i$ les lois markoviennes correspondantes. Les propriétés suivantes sont équivalentes*

1. $\exists i, j \in S\ \mathbb{P}_j(N_i = +\infty) > 0.$
2. $\exists i \in S$, *i est récurrent*
3. $\forall i \in S$, *i est récurrent*
4. $\forall i, j \in S\ \mathbb{P}_j(N_i = +\infty) = 1.$

Démonstration.
- (1) $\Longrightarrow$ (2). Soit l tel que $\mathbb{P}_i(X_l = j) > 0$. On a $\mathbb{P}_i(N = i = +\infty) \geq \mathbb{P}_i(X_l = j, \sum_{k=1}^{+\infty} X_{k+l} = i) \geq \mathbb{P}_i(X_l = j)\mathbb{P}_j(N_i = +\infty) > 0$, donc i est récurrent.
- (2) $\Longrightarrow$ (3). C'est une conséquence du corollaire précédent.

- (3) $\Longrightarrow$ (4). Considérons $\mathbb{P}_i(T_j < +\infty, \forall k > T_y X_k \neq x)$. Comme i et j communiquent $\mathbb{P}_i(T_j < +\infty) > 0$). D'après la propriété de Markov forte, on a

  $$\mathbb{P}_i(T_j < +\infty, \forall k > T_j X_k \neq i) = \mathbb{P}_i(T_j < +\infty)\mathbb{P}_j(\forall k > 0 X_k \neq i) = \mathbb{P}_i(T_j < +\infty)\mathbb{P}_j(T_i = +\infty)$$

  Mais $\{T_j < +\infty, \forall k > T_j X_k \neq i\} \subset \{N_j < +\infty\}$, donc comme i est récurrent, $\mathbb{P}_i(N_i < +\infty) = 0$, donc $\mathbb{P}_j(T_i = +\infty) = 0$. Mais

  $$N_i = 1 + \sum_{k=1}^{+\infty} \mathbb{1}_{\{i\}}(X_{k+T_i}),$$

  Donc d'après la propriété de Markov forte

  $$\mathbb{P}_j(N_i = +\infty) = \mathbb{P}_i(\sum_{k=1}^{+\infty} \mathbb{1}_{\{i\}}(X_k) = +\infty) = \mathbb{P}_i(N_i = +\infty) = 1.$$

- (4) $\Longrightarrow$ (1). Évident.

$\square$

Définition: Si une chaîne de Markov vérifie une des 4 propriétés équivalentes ci-dessus, on dit que c'est une *chaîne récurrente*.

2.3 Mesures invariantes

Définition: On dit qu'une mesure μ est *invariante* sous l'action de la matrice de transition markovienne M si $\mu M = \mu$, c'est à dire.

$$\forall j \in S \quad \sum_{i \in S} \mu(i) m_{i,j} = \mu(j).$$

Si μ est *invariante* sous l'action de M, une récurrence immédiate donne $\forall n \geq 0 \mu M^n = \mu$. Ainsi, si $(X_n)_{n\geq 0}$ est une chaîne de Markov de matrice de transition M et de mesure initiale $\mathbb{P}_{X_0} = \mu$, alors pour tout n, la loi de X_n est $\mathbb{P}_{X_n} = \mu$.

Définition: On dit qu'une mesure μ est réversible sous l'action de la matrice de transition markovienne M si

$$\forall i, j \in S \quad \mu(i) m_{i,j} = \mu(j) m_{j,i}.$$

Théorème 5. *Soit $(X_n)_{n\geq 0}$ une chaîne de Markov de loi initiale ν réversible sous l'action de M. Alors*

$$\forall n \geq 1 (X_0, X_1, \dots X_n) \text{ et } (X_n, X_{n-1}, \dots, X_0) \text{ ont même loi sous} \mathbb{P}_\nu.$$

Démonstration. Il suffit de procéder par récurrence sur n que

$\forall(X_0,\dots X_n)\in S^{n+1}\mathbb{P}_\nu(X_0=x_0,X_1=x_1,\dots,X_n=x_n)=\mathbb{P}_\nu(X_0=x_n,X_1=x_{n-1},\dots,X_n=$

Pour $n=1$, il suffit de voir que

$\mathbb{P}_\nu(X_0=x_0,X_1=x_1)=\nu(x_0)m_{x_0,x_1}=\nu(x_1)m_{x_1,x_0}=\mathbb{P}_\nu(X_0=x_1,X_1=x_0).$

Ensuite

$$\begin{aligned}
\mathbb{P}_\nu(X_0=x_0,X_1=x_1,\dots,X_n=x_n) &= \mathbb{P}_\nu(X_0=x_0,X_1=x_1,\dots,X_{n-1}=x_{n-1})m_{x_{n-1},x_n}\\
&= m_{x_{n-1},x_n}\mathbb{P}_\nu(X_0=x_{n-1},X_1=x_{n-2},\dots,X_{n-1}=x_0)\\
&= m_{x_{n-1},x_n}\nu(x_{n-1})\prod_{i=1}^{n-1}m_{x_{n-i},x_{x-i-1}}\\
&= \nu(x_{n-1})m_{x_{n-1},x_n}\prod_{i=1}^{n-1}m_{x_{n-i},x_{x-i-1}}\\
&= \nu(x_n)m_{x_n,x_{n-1}}\prod_{i=1}^{n-1}m_{x_{n-i},x_{x-i-1}}\\
&= \nu(x_n)\prod_{i=0}^{n-1}m_{x_{n-i},x_{x-i-1}}\\
&= \mathbb{P}_\nu(X_0=x_n,X_1=x_{n-1},\dots,X_n=x_0).
\end{aligned}$$

□

Il est facile de voir que toute mesure réversible est invariante.

Théorème 6. *Si la matrice de transition M est irréductible et admet une probabilité invariante, alors les chaînes de Markov associées à M sont récurrentes.*

Démonstration. Soit μ une probabilité invariante Pour tout $n\ge 0$, on a $\mu M^n=\mu$, soit

$$\forall j\in S\quad \forall n\ge 0\quad \sum_{i\in S}\mu(i)m_{i,j}^{(n)}=\mu(j)$$

Si une chaîne de Markov irréductible n'est pas récurrente, les états sont tous transitoires et $\lim_{n\to+\infty}\mu(i)m_{i,j}^{(n)}=0$ quels que soient i et j. D'après le théorème de convergence dominée, on a alors

$$\forall j\in S\quad 0=\mu(j),$$

ce qui est impossible. □

Théorème 7. *Toute chaîne de Markov sur un espace d'états S fini admet une probabilité invariante.*

Démonstration. L'ensemble $\mathcal{M}(S)$ des mesures de probabilité sur S s'identifie au compact $K = \{(x_1, \dots, x_n) \in \mathbb{R}_+^n; \sum_{k=1}^n x_k = 1\}$, avec $n = |S|$. $\mathcal{M}(S)$ est un convexe stable par $\mu \mapsto \mu M$. Ainsi, si μ est une mesure quelconque sur S, la suite $(\mu_n)_{n\geq 0}$ définie par

$$\mu_n = \frac{1}{n} \sum_{k=0}^{n-1} \mu M^k$$

est à valeurs dans $\mathcal{M}(S)$. On a $\mu_n(I - M) = \frac{\mu(I-M^n)}{n}$. Comme la suite $(\mu I - M^n))_{n\geq 0}$, est bornée, il s'ensuit que toute valeur d'adhérence de $(\mu_n)_{n\geq 0}$ est laissée fixe par M. Comme $\mathcal{M}(S)$ est compacte, $(\mu_n)_{n\geq 0}$ a au moins une valeur d'adhérence donc M au moins une mesure invariante. $\square$

Corollaire 6. *Une chaîne de Markov irréductible dont l'espace d'états est fini est récurrente.*

2.4 Théorème de la probabilité stationnaire

Théorème 8. *Soit M la matrice de transition d'une chaîne de Markov irréductible apériodique admettant μ comme loi stationnaire. Alors pour toute loi ν sur S, la chaîne de Markov de matrice de transition M et de loi initiale ν converge vers μ.*

Démonstration. Soit X_0, X'_0 deux variables aléatoires indépendantes, X_0 suivant la loi μ, X'_0 la loi ν. On note également $Y_0 = X'_0$. Soit également $(f_n)_{n\geq 1}$ et $(f'_n)_{n\geq 1}$ deux suites de variables aléatoires i.i.i.d. de loi θ_M définie au lemme 1, ces deux suites étant indépendantes de X_0 et X'_0. On définit par récurrence les suites $(g_n)_{n\geq 1}, (X_n)_{n\geq 1}$ et $(X'_n)_{n\geq 1}, (Y_n)_{n\geq 1}$ par

$$\begin{cases} X_{n+1} &= f_{n+1}(X_n) \\ Y_{n+1} &= f'_{n+1}(Y_n) \\ g_{n+1} &= \begin{cases} f_{n+1} & \text{si } X_n = X'_n \\ f'_{n+1} & \text{sinon} \end{cases} \\ X'_{n+1} &= g_{n+1}(X'_n) \end{cases}$$

Il n'est pas difficile de voir qu'en tout point ω on a

$$(X_n(\omega) = X'_n(\omega)) \Longrightarrow (f_{n+1}(\omega) = g_{n+1}(\omega)) \Longrightarrow (X_{n+1}(\omega) = X'_{n+1}(\omega))$$

Ainsi, les processus X_n et X'_n évoluent de manière indépendante jusqu'au moment où ils se rencontrent. À partir de là, X'_n demeure scotché à X_n.

Lemme 4. *Soit $(X_n)_{n\geq 0}$ une chaîne de Markov de matrice de transition M et de loi initiale μ , $(Y_n)_{n\geq 0}$ une chaîne de Markov de matrice de transition N et de loi initiale ν. On suppose en outre que les suites $(X_n)_{n\geq 0}$ et $(Y_n)_{n\geq 0}$ sont indépendantes sous P. Alors la suite $(Z_n)_{n\geq 0}$ définie par $Z_n = (X_n, Y_n)$ est une chaîne de markov de matrice de transition $M \otimes N$.*

Démonstration. Soient $(x_0, \ldots, x_n) \in S^{n+1}$ et $(y_0, \ldots, y_n) \in S^{n+1}$.

$$\begin{aligned}
P(\forall i \in \{0,n\}(X_i, Y_i) = (x_i, y_i)) &= P(\{\forall i \in \{0,n\} X_i = x_i\} \cap \{\forall i \in \{0,n\} Y_i = y_i\}) \\
&= P(\forall i \in \{0,n\} X_i = x_i) P(\forall i \in \{0,n\} Y_i = y_i) \\
&= \mu(\{x_0\}) \prod_{i=0}^{n-1} m_{x_i,x_{i+1}} \times \nu(\{y_0\}) \prod_{i=0}^{n-1} n_{y_i,y_{i+1}} \\
&= \mu(\{x_0\}) \nu(\{y_0\}) \prod_{i=0}^{n-1} m_{x_i,x_{i+1}} n_{y_i,y_{i+1}} \\
&= (\mu \otimes \nu)(\{x_0, y_0\}) \prod_{i=0}^{n-1} (M \otimes N)((x_i, y_i), (x_{i+1}, y_{i+1}))
\end{aligned}$$

□

Lemme 5. *Soit U, V deux variables aléatoires de loi θ. On suppose que sous P, U et V sont indépendantes de la tribu $\mathcal{A}$. Soit A un événement $\mathcal{A}-$ mesurable. On définit W par*

$$W(\omega) = \begin{cases} U(\omega) \text{ si } \omega \in A \\ V(\omega) \text{ si } \omega \notin A \end{cases}$$

Alors, sous P, W suit la loi θ et W est indépendante de $\mathcal{A}$.

Démonstration. Soit A' un événement $\mathcal{A} - mesurable$ et B un borélien

$$\begin{aligned}
P(A' \cap \{W \in B\}) &= P(A \cap A' \cap \{W \in B\}) + P(A^c \cap A' \cap \{W \in B\}) \\
&= P(A \cap A' \cap \{U \in B\}) + P(A^c \cap A' \cap \{V \in B\}) \\
&= P(A \cap A') P(U \in B) + P(A^c \cap A') P(V \in B) \\
&= P(A \cap A') \theta(B) + P(A^c \cap A') \theta(B) \\
&= (P(A \cap A') + P(A^c \cap A')) \theta(B) \\
&= P(A') \theta(B)
\end{aligned}$$

En prenant $A' = \Omega$, on en déduit d'abord que $P(W \in B) = \theta(B)$ pour tout borélien B. θ est donc la loi de W sous P. En réinsérant dans la formule précédente, on a pour tout événement $\mathcal{A}-mesurable$ A' et pour tout borélien B :

$$P(A' \cap \{W \in B\}) = P(A')P(W \in B),$$

ce qui veut dire que W est indépendante de $\mathcal{A}$. □

En appliquant le lemme précédent à $\mathcal{A} = \sigma(X_0, X'_0, f_1, \dots, f_n, f'_1, \dots, f'_n)$, $A = \{X_n = X'_n\}$, $U = f_{n+1}$, $V = f'_{n+1}$ et $W = g_{n+1}$ on voit que g_{n+1} suit la loi θ_M et que g_{n+1} est indépendante de $\sigma(X_0, X'_0, f_1, \dots, f_n, f'_1, \dots, f'_n)$. Comme $(g_1, \dots, g_n)$ est $\sigma(X_0, X'_0, f_1, \dots, f_n, f'_1, \dots, f'_n)$-mesurable, il s'ensuit que $(g_n)_{n\geq 1}$ est une suite de v.a.i.i.d de loi θ_M.

D'après le lemme 1, (X_n) est une chaîne de Markov de matrice de transition M et de loi initiale μ tandis que (X'_n) est une chaîne de Markov de matrice de transition M et de loi initiale ν.

On va maintenant montrer que $\tau = \inf\{n; X_n = X'_n\}$ est presque sûrement fini. Il est facile de voir que $\tau = \inf\{n; X_n = Y_n\}$. Ce qui est intéressant, c'est que $(X_n)_{n\geq 0}$ et $(Y_n)_{n\geq 0}$ sont indépendants.

Ainsi, d'après le lemme 4, (X_n, Y_n) est une chaîne de Markov de loi initiale $\nu \otimes \mu$ et de matrice de transition $N = M \otimes M$. Soient $(x, y, z, t) \in S^4$. Comme M est la matrice d'une chaîne de Markov irréductible et apériodique, on peut, d'après le lemme 1, trouver un entier n_0 tel que M^{n_0} soit à coefficients strictement positifs. Or $N^{n_0} = (M \otimes M)^{n_0} = M^{n_0} \otimes M^{n_0}$: on a

$$N^{n_0}((x, y), (z, t)) = M^{n_0}(x, z)M^{n_0}(y, t) > 0.$$

Ainsi $((Z_n)_{n\geq 0} = ((X_n, Y_n))_{n\geq 0}$ est une chaîne de Markov irréductible. Comme $M \otimes M$ admet $\mu \otimes \mu$ comme mesure invariante, la dynamique est donc récurrente : $(Z_n)_{n\geq 0}$ passe donc presque sûrement en tout point de $S \times S$. En particulier, elle passe presque sûrement sur sa diagonale, ce qui implique que $P(\tau < +\infty) = 1$.

Soit f une fonction bornée de S dans $\mathbb{R}$. Pour $n \geq \tau$, on a $f(X_n) = f(X'_n)$. Donc $f(X_n) - f(X'_n)$ converge presque sûrement vers 0. D'après le théorème de convergence dominée, on en déduit que $\mathbb{E}(f(X_n) - f(X'_n))$ converge vers 0. Comme μ est invariante $\mathbb{E}(f(X_n) - f(X'_n)) = \int f \, d\mu - \mathbb{E}f(X'_n)$. Ainsi pour toute fonction f, $\mathbb{E}f(X'_n)$ converge vers $\int f \, d\mu$, ce qui veut dire que X'_n converge en loi vers μ.

□

Corollaire 7. *Une chaîne de Markov irréductible apériodique a au plus une loi stationnaire.*

Remarque-exercice : l'hypothèse d'apériodicité est importante. En effet, on peut construire deux chaînes de Markov indépendantes $(X_n)_{n\geq 0}$ et

$(Y_n)_{n\geq 0}$ ayant la même matrice de transition irréductibles, telles que $(X_n, Y_n)_{n\geq 0}$ ne soit pas irréductible et que $(X_n, Y_n)_{n\geq 0}$ ne coupe jamais la diagonale. Donner deux exemples d'un tel phénomène, l'un avec S fini, l'autre avec S infini.

2.5 Théorème ergodique des chaînes de Markov

Théorème 9. *Soit $(X_n)_{n\geq 0}$ une chaîne de Markov irréductible et récurrente. Pour tout $x \in S$, on a*

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{k=0}^{n-1} 1\!\!1_{\{x\}}(X_k) \to \frac{1}{\mathbb{E}_x T_x}.$$

Démonstration. Pour tout $k \geq 1$, posons $T^k = \inf\{n > 0, \sum_{k=1}^{n} 1\!\!1_{\{x\}}(X_k) \geq k$. Comme la chaîne est récurrente, les T^k sont presque sûrement finis. Il est aisé de constater que la suite $(T^k)_{k\geq 1}$ est croissante. Soit $k \geq 1$ et $A \in \sigma(T^1, \dots, T^k)$: il est clair que A se produit avant T^k. Donc

$$\begin{aligned}
\mathbb{P}_x(A, T^{k+1} - T_k > n) &= \mathbb{P}_x(A, \cap_{j=T^k+1}^{T^k+n} 1\!\!1_{\{x\}}(X_j) = 0) \\
&= \mathbb{P}_x(A)\mathbb{P}_x(\cap_{j=1}^{n} 1\!\!1_{\{x\}}(X_j) = 0) \\
&= \mathbb{P}_x(A)\mathbb{P}_x(T^1 > n)
\end{aligned}$$

On en déduit que, sous la loi $\mathbb{P}_x$, les variables aléatoires $T^1, T^2 - T^1, T^3 - T^2, \dots$ forment une suite de variables aléatoires positives indépendantes ayant même loi que $T^1 = T_x$. D'après la loi forte des grands nombres on en déduit que $\frac{T^n}{n} = \frac{1}{n}\big(T^1 + (T^2 - T^1) + (T^3 - T^2) + \dots (T^n - T^{n-1}\big)$ converge presque sûrement vers $\mathbb{E}_x T^x$. Lorsque la loi initiale n'est pas la masse de Dirac en x, il suffit de remarquer que la suite $\frac{T^n}{n}$ a le même comportement asymtotique que la suite $\frac{T^n - T^1}{n-1}$ et appliquer à nouveau la propriété de Markov forte : la loi sous P de $(\frac{T^n - T^1}{n-1})_{n\geq 2}$ est la loi sous $\mathbb{P}_x$ de $(\frac{T^n}{n})_{n\geq 1}$, d'où le résultat. Posons $S_n = \sum_{k=0}^{n-1} 1\!\!1_{\{x\}}(X_k)$. Un instant de réflexion montre que $T^{S_n} \leq n < T^{S_n+1}$. On en déduit

$$\frac{T^{S_n}}{S_n} \leq \frac{n}{S_n} < \frac{T^{S_n+1}}{S_n+1} \frac{S_n+1}{S_n}$$

Comme x est récurrent $\lim_{n\to+\infty} S_n = +\infty$, donc $\lim_{n\to+\infty} \frac{n}{S_n} = \mathbb{E}_x T_x$, d'où le résultat. □

Théorème 10. *Soit $(X_n)_{n\geq 0}$ une chaîne de Markov irréductible admettant une probabilité invariante μ. Pour tout $x \in S$, on a*

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{\{x\}}(X_k) \to \frac{1}{\mathbb{E}_x T_x} = \mu(x).$$

Démonstration. Une chaîne de Markov irréductible admettant une probabilité invariante est toujours récurrente. Le théorème précédent s'applique donc. Comme $|\frac{1}{n}\sum_{k=0}^{n-1}\mathbb{1}_{\{x\}}(X_k)| \leq 1$, le théorème de convergence dominée s'applique et on a

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{P}_\nu(X_k = x) \to \frac{1}{\mathbb{E}_x T_x}.$$

Si l'on prend pour *nu* la mesure invariante μ, on a pour tout $k \geq 0$ $\mathbb{P}_\nu(X_k = x) = \mu(x)$. On en déduit que $\frac{1}{\mathbb{E}_x T_x} = \mu(x)$, ce qui achève la preuve. □

2.6 Exercices

1. *Temps d'atteinte d'un état absorbant.*
   Soit (X_n) une chaîne de Markov sur un ensemble dénombrable E de noyau de transition Q, et $a \in E$ un état absorbant (*ie* : $Q(a,a) = 1$). Montrer que $P(X_n = a) = P(T_a \leq n)$.
2. *Temps d'entrée : une propriété d'invariance.*
   Soit (X_n) une chaîne de Markov sur un ensemble dénombrable E de noyau de transition Q. Pour $f : E \to \mathbb{R}_+$, soit Qf la fonction définie par $Qf(x) = \sum_{y\in E} Q(x,y)f(y)$. Pour $B \subset E$ on note $D_B = \inf\{n \geq 0; X_n \in B\}$ le temps d'entrée dans B. Montrer que la fonction f définie par $f(x) = \mathbb{P}_x(T_B < +\infty)$ vérifie $(I - Q)f = 0$ sur $E \setminus B$, et $f = 1$ sur B.
3. *Chaîne de Markov arrêtée.*
   Soit (X_n) une chaîne de Markov sur un ensemble dénombrable E de noyau de transition Q. Etant donné un ensemble $B \subset E$, on pose $Y_n = X_{n\wedge D_B}$. Montrer que (Y_n) est une chaîne de Markov sur E de noyau de transition Q' défini par $Q'(x,y) = \delta_x(\{y\})$ si $x \in B$ et $Q'(x,y) = Q(x,y)$ si $x \notin B$.
4. *Chaîne observée quand elle bouge.* Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov homogène sur l'espace d'états E, de matrice de transition P. On définit la suite $(T_k)_{k\in\mathbb{N}}$ de la façon suivante : $T_0 = 0$ et

$$T_{k+1} = \min\{n \geq T_k, X_n \neq X_{T_k}\}.$$

On suppose que la chaîne n'admet aucun état absorbant.

(a) Montrer que les $(T_k)_{k\in\mathbb{N}}$ sont des temps d'arrêt pour $(X_n)_{n\in\mathbb{N}}$, finis presque sûrement.

(b) On définit $Y_k = X_{T_k}$. Montrer que $(Y_k)_{K\in\mathbb{N}}$ est une chaîne de Markov homogène, donner son espace d'états et sa matrice de transition.

5. *Chaîne restreinte.* Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov homogène d'espace d'états E dénombrable et de matrice de transition $P = (p_{i,j})_{i,j\in E}$. Soit J une partie de E. On observe cette chaîne de Markov seulement lors de ses passages par J, et on note Y_m la $m^{\text{ième}}$ observation. Plus formellement, pour $m \geq 1$, on note

$$T_m = \inf\left\{n \geq 1 + T_{m-1} \;\middle|\; X_n \in J\right\} \text{ et } T_0 = \inf\left\{n \geq 0 \;\middle|\; X_n \in J\right\}.$$

On suppose que $\forall m \geq 0$, $T_m < +\infty$, et on pose $Y_m = X_{T_m}$.

(a) Montrer que T_m est un temps d'arrêt, que X_{T_m} est mesurable pour $\mathcal{F}_{T_m}$ et que pour $k \leq m$, T_k et X_{T_k} sont $\mathcal{F}_{T_m}$-mesurables.

(b) Montrer que $(Y_n)_{n\in\mathbb{N}}$ est une chaîne de Markov homogène.
Indication. On pourra montrer, en décomposant sur les valeurs possibles de $T_m - T_{m-1}$, que, si $(k_0, \ldots, k_m) \in J^{m+1}$ alors $P(Y_0 = k_0, \ldots Y_m = k_m) = P(Y_0 = k_0, \ldots Y_{m-1} = k_{m-1})\mathbb{P}_{k_{m-1}}(X_S = k_m)$, où $S = \inf\left\{n \geq 1 \;\middle|\; X_n \in J\right\}$.

(c) On note $Q = (q_{i,j})_{i,j\in J}$ la matrice de transition de $(Y_n)_{n\in\mathbb{N}}$. Montrer que

$$q_{i,j} = p_{i,j} + \sum_{l\notin J} p_{i,l}\mathbb{P}_l(X_{T_0} = j)$$

et donner une caractérisation des $(\mathbb{P}_l(X_{T_0} = j))_{l\notin J, j\in J}$.

(d) Exprimer, en fonction de la loi initiale μ_0 de $(X_n)_{n\in\mathbb{N}}$, la loi initiale ν_0 de $(Y_n)_{n\in\mathbb{N}}$.

(e) Exemple. On considère la marche aléatoire symétrique aux plus proches voisins sur l'hypercube $\{0,1\}^d$, et on choisit $J = \{(x_i)_{1\leq i\leq d}, \sum x_i \text{ impaire}\}$. Déterminer les caractéristiques de $(Y_n)_{n\in\mathbb{N}}$ dans ce cas.

6. *Marche aléatoire sur $\mathbb{Z}/d\mathbb{Z}$.* On considère $X = \{X_n : n \geq 0\}$ la marche aléatoire sur $\mathbb{Z}/d\mathbb{Z}$ dont les pas sont indépendants de même loi $p\delta_1 + (1-p)\delta_{-1}$.

(a) Quelle est sa matrice de transition P ? Dessiner son graphe.

(b) Calculer P^n. (Indication : écrire $P = pA + (1-p)B$ et trouver une relation entre A et B.)

(c) Calculer la probabilité stationnaire π.

(d) (calculatoire) Calculer la loi μ_n de X_n. Discuter l'existence de $\lim_n \mu_n$, $\lim_n P^n$.

7. *Modèle d'Ehrenfest.* Etant donné deux enceintes séparées par une paroi poreuse et contenant ensemble N particules diffusant à travers cette paroi, on décrit le nombre aléatoire X_n de particules se trouvant dans la première enceinte ($0 \leq X_n \leq N$) aux instants successifs $n \in \mathbb{N}$ de transition des particules par une chaîne Markov de probabilité de transition :

$$p(x, x-1) := \frac{x}{N},\ p(x, x+1) := \frac{N-x}{N},\ 0 \leq x \leq N.$$

À chaque instant n, les probabilités que la transition se fasse de la première enceinte vers la seconde ou de la seconde enceinte vers la première sont donc proportionnelles aux nombres de particules en présence dans la première et la deuxième enceinte :

$$\frac{p(x, x-1)}{x} = \frac{p(x, x+1)}{N-x}.$$

Montrer que cette chaîne est récurrente irréductible et trouver sa probabilité invariante.

8. *Chaîne de Markov avec décision.*
Le $n^{\text{ème}}$ Lundi de l'année, une petite entreprise reçoit A_n propositions de travail de type A, et B_n propositions de travail de type B. Un travail de type A mobilise toute la capacité de travail de l'entreprise durant une semaine et lui rapporte 200 euros, alors qu'un travail de type B l'occupe deux semaines pour un rapport de 360 euros. Une semaine d'inactivité coûte 100 euros, un travail non traité pendant la semaine où il arrive est perdu. On suppose A_n,B_n indépendants, les couples $(A_n, B_n)_{n\geq 1}$ indépendants, et

$$\mathbb{P}(A_n = 1) = 1 - \mathbb{P}(A_n = 0) = 0,5, \qquad \mathbb{P}(B_n = 1) = 1 - \mathbb{P}(B_n = 0) = 0,6.$$

Modéliser la situation par une chaîne de Markov , avec si possible un nombre d'états minimal. Quelle est la meilleure stratégie, quand on reçoit simultanément une offre de chaque type : donner la préférence à celle de type A ou à celle de type B ? On pourra faire appel au Théorème ergodique pour départager les deux politiques.

9. *Un modèle de prédiction météo (!)* On suppose que le temps qu'il fera demain depend des deux jours précédents. On suppose que :

   $\mathbb{P}($ il pleut demain $|$ il a plu hier et aujourd'hui$) = 0,7$
   $\mathbb{P}($ il pleut demain $|$ il a plu aujourd'hui mais pas hier$) = 0,5$
   $\mathbb{P}($ il pleut demain $|$ il a plu hier mais pas aujourd'hui$) = 0,4$
   $\mathbb{P}($ il pleut demain $|$ il n'a pas plu ni hier, ni aujourd'hui$) = 0,2$

   Montrer qu'on peut modéliser ceci par une chaîne de Markov. Quelle est la probabilité, sachant qu'il a plu lundi et mardi qu'il pleuve jeudi ? Sur le long terme, quelle proportion de jours de pluie observe-t-on ?

10. *Chaîne de Markov réversible*

    (a) Soit P une matrice de transition sur un espace d'états E dénombrable. On suppose qu'il existe une probabilité π telle que
    $$\pi_i p_{i,j} = \pi_j p_{j,i}.$$
    Montrer que π est stationnaire pour la P.

    (b) Trouver rapidement la probabilité stationnaire de la marche aléatoire symétrique sur les sommets de l'hypercube de dimension d.

    (c) Marche aléatoire symétrique sur un damier. Calculer les temps de retours moyens des différents points de l'échiquier.

11. *Modèle de Laplace-Bernoulli.* N boules noires et N boules blanches sont placées dans deux urnes de sorte que chacune contienne N boules. Après chaque unité de temps on choisit au hasard une boule de chaque urne ; les deux boules ainsi choisies changent d'urne. On note par X_n le nombre de boules noires dans la première urne. Montrer que $\{X_n : n \geq 0\}$ est une chaîne de Markov irréductible réversible et trouver sa mesure stationnaire.

Chapitre 3

Quelques algorithmes stochastiques

3.1 Un algorithme de Métropolis

Le but de cette partie est de montrer comment on fabrique une dynamique markovienne admettant une mesure donnée comme mesure d'équilibre.

On suppose que S est un ensemble fini ou dénombrable et μ une mesure sur S chargeant tous les points de S. On peut alors supposer sans restriction que μ s'écrit

$$\mu(x) = K \exp(-V(x)).$$

On suppose qu'on a su construire une matrice de transition $\tilde{P}$ **irréductible et symétrique** sur S correspondant à un processus que l'on sait simuler.

On cherche une matrice stochastique P pour laquelle μ est réversible : on veut

$$\forall i, j \in S \quad \mu(x)P(x, y) = \mu(y)P(y, x) \tag{3.1}$$

Bien sûr, si μ est constante (c'est à dire si μ la mesure invariante sur un ensemble fini), nous n'avons rien à faire car $\tilde{P}$ convient.

Il est évident que l'identité (3.1) est vide si $x = y$. On va chercher, pour x et y distincts, $P(x, y)$ sous la forme $P(x, y) = \tilde{P}(x, y)a(x, y)$, avec $0 < a(x, y) \leq 1$. La matrice A est appelée matrice d'acceptation. Il suffira alors de poser $P(x, x) = 1 - \sum_{y \neq x} P(x, y)$ pour compléter la diagonale de P de manière à avoir une matrice stochastique.

L'équation traduisant la réversibilité se traduit sur A par

$$\forall (x, y) \in S \quad x \neq y \Longrightarrow \exp(-[V(y) - V(x)]) = \frac{a(x, y)}{a(y, x)} \tag{3.2}$$

Nous allons maintenant chercher a sous la forme $a(x,y) = F(V(y) - V(x))$. Ainsi (3.2) devient

$$\forall (x,y) \in S \quad x \neq y \Longrightarrow \exp(-[V(y) - V(x)]) = \frac{F(V(y) - V(x)}{F(V(x) - V(y))} \tag{3.3}$$

il suffit alors d'avoir

$$\forall u \in \mathbb{R} \quad \exp(-u) = \frac{F(u)}{F(-u)} \tag{3.4}$$

Il est facile de voir que la fonction F définie par

$$F(u) = \begin{cases} e^{-u} & \text{si } u \geq 0 \\ 1 & \text{si } u < 0 \end{cases}$$

satisfait cette équation.

Si μ n'est pas la mesure uniforme, alors la matrice de transition P est irréductible et apériodique.

Démonstration. L'irréductibilité de P provient de celle de $\tilde{P}$ et du fait que la probabilité d'acceptation $a(x,y)$ est toujours strictement positive.

Passons à l'apériodicité. Remarquons tout d'abord que comme S est fini ou dénombrable et que $\sum_{x \in S} \exp(-V(x)) < +\infty$, V atteint nécessairement son minimum sur S. Soit maintenant M l'ensemble des éléments de S où V est minimale. Comme V n'est pas constante, M n'est pas S tout entier, donc il existe $(x,y) \in M \times (S \backslash M)$ tel que $\tilde{P}(x,y) > 0$, sinon cela contredirait l'irréductibilité de $\tilde{P}$. On a alors

$$\begin{aligned} 1 - P(x,x) &= \sum_{z \neq x} \tilde{P}(x,z) a(x,z) \\ &= \sum_{z \neq x,y} \tilde{P}(x,z) a(x,z) + \tilde{P}(x,y) a(x,y) \\ &\leq \sum_{z \neq x,y} \tilde{P}(x,z) + \tilde{P}(x,y) a(x,y) \\ &\leq \sum_{z \neq x} \tilde{P}(x,z) + \tilde{P}(x,y)(a(x,y) - 1) \\ &\leq 1 + \tilde{P}(x,y)(a(x,y) - 1) \end{aligned}$$

Ainsi $P(x,x) \geq (1 - a(x,y))\tilde{P}(x,y) = (1 - e^{-[V(y) - V(x)]})\tilde{P}(x,y) > 0$.

Ainsi x est de période 1 : comme la chaîne est irréductible, la chaîne est alors apériodique. $\square$

Mise en œuvre concrète : pour passer de $X_n = x$ à X_{n+1}, on propose suivant la méthode utilisée pour simuler $\tilde{P}$ de se déplacer au point y. Si $x = y$, le "changement" est accepté. Sinon, le changement est alors accepté avec probabilité $a(x, y)$: si $V(y) \leq V(x)$, le changement est toujours accepté, tandis que si $V(y) > V(x)$, le changement est accepté avec probabilité $\exp(-(V(y) - V(x))$ – par exemple si une variable aléatoire uniforme U sur $[0, 1]$ vérifie $U \leq \exp(-(V(y) - V(x))$. Si le changement est accepté, on a alors $X_{n+1} = y$, sinon on a $X_{n+1} = x$.

Ainsi, quelle que soit la situation de départ, le processus simulé converge en loi vers μ, puisque $(X_n)_{n\geq 0}$ est une chaîne de Markov irréductible et apériodique dont μ est une mesure invariante.

Notons que la connaissance de K n'est pas nécessaire pour la mise en œuvre de la simulation.

3.1.1 Exemple : mesures de Gibbs associés à une interaction

Soit E un ensemble fini. Pour toute partie finie V de $\mathbb{Z}^d$, soit Φ_V une variable $\sigma(V)$ mesurable. La famille des Φ_V, quand V décrit l'ensemble des parties finies de S, est appelée potentiel d'interaction du système. Soit Λ une partie finie de $\mathbb{Z}^d$.
La quantité

$$H_\Lambda = \sum_{B|B\cap\Lambda\neq\varnothing} \Phi_B$$

est appellée (lorsquelle existe) le hamiltonien du système sur le volume fini Λ. On appelle portée de l'interaction la borne supérieure des diamètres des parties V pour lesquelles Φ_V n'est pas l'application identiquement nulle.

Considérons maintenant $\beta > 0$. On écrit parfois $\beta = \frac{1}{T}$, où T représente la température du système.

Si on se fixe "condition extérieure" $\eta \in E^{\Lambda^c}$, on s'intéresse alors à la mesure de Probabilité μ_Λ définie par E^Λ

$$\mu_\Lambda^\eta(x_\Lambda) = \frac{1}{Z_\Lambda(\eta)} \exp(-\beta H_\Lambda(x_\Lambda \eta_{\Lambda^c})),$$

où

$$Z_\Lambda(\eta) = \sum_{y\in E^\Lambda} \exp(-\beta H_\Lambda(y)\eta_{\Lambda^c})).$$

Modèle d'Ising en dimension d

Ici $E = \{-1, +1\}$. On définit le potentiel d'interaction par :

$$\Phi_{\{i,j\}}(\omega) = -\omega_i\omega_j \quad \text{si } \|i-j\| = 1 \quad \text{les autres interactions étant nulles.}$$

On parle ici d'interaction aux plus proches voisins.
Physiquement, ce modèle représente le phénomène du magnétisme : chaque site correspond à un sommet d'un réseau cristallin occupé par un spin ne pouvant prendre que deux orientations ou deux valeurs $-1, +1$.

Décrivons maintenant un algorithme de Métropolis adapté au modèle Ising dans un volume Λ à température inverse β

Décrivons d'abord la dynamique de $\tilde{P}$.

Voilà comment on passe de l'état au temps n à l'étape au temps $n+1$.

On note x la configuration au temps n.

1. Choisir uniformément $k \in \Lambda$.
2. Définir la configuration y au temps $n+1$ par

$$y_l = \begin{cases} x_l & \text{si } l \in \Lambda \backslash k \\ -x_l & \text{si } l = k \end{cases}$$

Il est alors très facile de voir que la chaîne $\tilde{P}$ est irréductible : on peut toujours passer d'une configuration à une autre en inversant un nombre fini de sites.

Ainsi deux états (ou configurations) x_Λ et y_Λ communiquent (ou de manière équivalente $\tilde{P}(x,y) \neq 0$) si et seulement elles diffèrent en exactement un site. Si x_Λ et y_Λ diffèrent en un unique site k, on a

$$\tilde{P}(x,y) = \frac{1}{|\Lambda|}.$$

Il est alors aisé de constater que la matrice $\tilde{P}$ est bien symétrique.

Si x_Λ et y_Λ diffèrent exclusivement au site k, alors il en est de même pour $x_\Lambda\eta_{\Lambda^c}$ et $y_\Lambda\eta_{\Lambda^c}$ on a alors

$$\begin{aligned} H_\Lambda(y_\Lambda\eta_{\Lambda^c}) - H_\Lambda(x_\Lambda\eta_{\Lambda^c}) &= \sum_{B:k\subset B} \varphi_B(y_\Lambda\eta_{\Lambda^c}) - \sum_{B:k\subset B} \varphi_B(x_\Lambda\eta_{\Lambda^c}) \\ &= \sum_{l:\|k-l\|=1} (-y_k y_l) - \sum_{l:\|k-l\|=1} (-x_k x_l) \\ &= 2 \sum_{l:\|k-l\|=1} x_k x_l \end{aligned}$$

On peut maintenant décrire la dynamique de P.

Voilà comment on passe de l'état au temps n à l'étape au temps $n+1$.

On note x la configuration au temps n.

1. Choisir uniformément $k \in \Lambda$
2. Choisir U suivant la loi uniforme sur $[0,1]$.
3. Définir la configuration y au temps $n+1$ par

$$y_l = \begin{cases} -x_l & \text{si } k = l \text{ et } U \leq \exp(-2\beta x_k \sum\limits_{i:\|k-i\|=1} x_i) \\ x_l & \text{sinon.} \end{cases}$$

3.2 Échantillonneur de Gibbs

Cette méthode est particulièrement utile lorsque l'espace d'états S est inclus dans E^Λ, où E et Λ sont des ensembles finis. On note μ la mesure cible et S son support

Voilà comment on passe de l'état au temps n à l'étape au temps $n+1$.

On note x la configuration au temps n.

1. Choisir uniformément $k \in \Lambda$.
2. Choisir $e \in E$ suivant la loi μ conditionnée par le fait que la configuration coincide avec x en tout point de $\Lambda \backslash k$.
3. Définir la configuration y au temps $n+1$ par

$$y_l = \begin{cases} x_l & \text{si } l \in \Lambda \backslash k \\ e & \text{si } l = k \end{cases}$$

Ainsi, si deux configurations x et y diffèrent en plus d'un site, la probabilité de passage $p(x,y)$ de x à y vaut zéro. En revanche, si x et y diffèrent en un unique site k, on a

$$p(x,y) = \frac{1}{|\Lambda|} \frac{\mu(y)}{\sum\limits_{z \in S; z_{k^c} = x_{k^c}} \mu(z)}.$$

Il est alors aisé de constater que

$$\forall x, y \in S \times S \quad \mu(x)p(x,y) = \mu(y)p(y,x),$$

c'est à dire que μ est réversible sous cette dynamique, donc invariante. Comme $p(x,x) > 0$ pour tout x, la chaîne est apériodique.

Si la chaîne est irréductible, alors c'est une chaîne irréductible apériodique qui converge vers son unique mesure invariante μ.

3.2.1 Application au modèle d'Ising

Soit $x \in \{-1,1\}^\Lambda$ et $k \in \Lambda$. On a

$$\begin{aligned}\frac{\mu_\Lambda(x_{\Lambda\setminus\{k\}}1_k)}{\mu_\Lambda(x_{\Lambda\setminus\{k\}}(-1)_k)} &= \exp(-\beta(H_\Lambda(x_{\Lambda\setminus\{k\}}1_k) - H_\Lambda(x_{\Lambda\setminus\{k\}}(-1)_k))) \\ &= \exp(2\beta \sum_{i:\|i-k\|=1} x_i)\end{aligned}$$

On peut maintenant décrire la dynamique de l'échantilloneur de Gibbs du modèle d'Ising

Voilà comment on passe de l'état au temps n à l'étape au temps $n+1$.

On note x la configuration au temps n.

1. Choisir uniformément $k \in \Lambda$
2. Choisir U suivant la loi uniforme sur $[0,1]$.
3. Définir la configuration y au temps $n+1$ par

$$y_l = \begin{cases} x_l & \text{si } k \neq l \\ 1 & \text{si } k = l \text{ et } U < 1/(1+\exp(-2\beta \sum_{i:\|k-i\|=1} x_i)) \\ -1 & \text{sinon.} \end{cases}$$

Pour tout $n \geq 0$ $x \in \{-1,+1\}^\Lambda$, et tout $k \in \Lambda$, il est possible d'inverser l'état du site k avec une probabilité non nulle. On en déduit que tous les états communiquent ; la chaîne est irréductible apériodique, donc l'algorithme converge bien vers la mesure d'équilibre μ_Λ.

3.3 Algorithme de Propp et Wilson

Théorème 11. *Soit M la matrice d'une chaîne de Markov irréductible admettant μ comme mesure invariante et $(f_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant une loi θ telle que $\forall i,j \in S \quad \theta(f \in F, f(i) = j) = m_{i,j}$.*

On pose $g_0 = Id_S$, puis pour $n \geq 0$ $g_{n+1} = g_n \circ f_{n+1}$. Soit A une partie infinie de $\mathbb{N}$. On note

$$T = \inf\{n \in A, g_n \text{ est une fonction constante}\}.$$

Si $P(T < +\infty) = 1$, alors $g_T(x_0)$ suit la loi μ, où $x_0 \in S$ est quelconque.

Démonstration. Soit X_0 une variable aléatoire suivant la loi ν indépendante de la suite $(f_n)_{n\geq 1}$. Pour $n \geq T$, comme $g_n = g_T \circ (f_{T+1} \circ f_{T+2} \circ \cdots \circ f_n)$, on a $g_n(X_0) = g_T(X_0)$. On en déduit que pour $y \in S$ et $n \geq T$, on a $\mathbb{1}_{\{g_n(X_0)=y\}} = \mathbb{1}_{\{g_T(X_0)=y\}}$. On a donc presque sûrement $\lim_{n\to+\infty} \mathbb{1}_{\{g_n(X_0)=y\}} = \mathbb{1}_{\{g_T(X_0)=y\}}$. On appliquant le théorème de convergence dominée, on obtient

$$\lim_{n\to+\infty} P(g_n(X_0) = y) = P(g_T(X_0) = y).$$

Or $g_n(X_0) = f_1 \circ f_2 \cdots \circ f_n(X_0)$ a la même loi que $f_n \circ f_{n-1} \cdots \circ f_1(X_0)$, c'est à dire la loi que suit au temps n une chaîne de Markov de mesure initiale μ et de matrice de transition M, c'est à dire μ. (Attention : $(g_n(X_0))_{n\geq 0}$ n'est pas une chaîne de Markov !) On a donc pour tout $n \geq 0$ $(g_n(X_0) = y) = \mu(y)$, d'où $P(g_T(X_0) = y) = \mu(y)$. $\square$

Dans la pratique, la suite f_n est souvent générée de la manière suivante : on construit un ensemble X, une suite i.i.d. Z_n de variables aléatoires aisément simulables et une fonction déterministe $f : X \times S \to S$ telle que la suite f_n définie par $f_n(x, y) = f(Z_n, y)$ vérifie les conditions mentionnées ci-dessus.

On peut légitimement se demander pourquoi on ne prend pas tout simplement $A = \mathbb{N}$. Théoriquement, en effet, rien ne l'empêche. Cependant il faut voir que passer de n à $n + 1$ n'augmente que très peu la probabilité que la fonction soit constante et tester si g_n est constante et augmente de manière non négligeable le temps de calcul.

Explication : si $A = \{a_1, a_2, \ldots, \}$ et que $\inf n \geq 0; n$ est constante $= n_0$. l'on a $a_{k-1} < t \leq n_0 \leq a_k$, le temps de calcul est proportionnel à $\sum_{i=1}^{k} a_i$. Ainsi, le choix $a_k = k$ (c'est à dire $A = \mathbb{N}$) conduit à un temps

$$\sum_{i=1}^{n_0} i = \frac{n_0(n_0+1)}{2} \sim \frac{n_0^2}{2},$$

tandis que le choix $a_k = 2^{k-1}$, conduit pour à $\sum_{k:2^{k-2}<n_0} 2^{k-1}$. Cette somme vaut $2n_0 - 1$ si n_0 est une puissance de 2, et au pire $4(n_0 - 1) - 1$.

Ainsi, on considère généralement que le choix $A = \{2^k; k \geq 0\}$ est un bon choix.

3.3.1 Loi 0-1 pour l'algorithme de Propp et Wilson

Théorème 12. *On reprend les hypothèses du théorème précédent*

– *Si il existe $n \geq 0$ tel que $P(g_n$ est une fonction constante$) > 0$, alors $P(T < +\infty) = 1$.*

– *Sinon* $P(T < +\infty) = 0$.

En particulier, l'issue ne dépend pas du choix de A.

Démonstration. Il est clair que si

$$\forall n \geq 0 \quad P(g_n) \text{ est une fonction constante}) = 0,$$

alors par réunion dénombrable on a $P(T < +\infty) = 0$. Supposons donc qu'il existe n tel que $P(g_n$ est une fonction constante$) > 0$. Comme $A \cap (n, +\infty)$ est une partie infinie de $\mathbb{N}$, il est possible d'en extraire un ensemble infini B tel que la différence entre deux éléments distincts quelconques de B dépasse n. Pour $k \in B$, notons

$$A_k = \{f_{k-n+1} \circ f_{k-n+2} \circ \cdots \circ f_{k-1} \circ f_k \text{ est une fonction constante}\}.$$

Les événements A_k sont des éléments indépendants, de même probabilité $P(g_n$ est une fonction constante.$) > 0$. On en déduit que

$$P(\cup_{k \in B} A_k) = 1.$$

Comme $g_k = f_1 \circ f_2 \cdots \circ f_k$, il est facile de voir que $A_k \subset \{g_k$ est une fonction constante$\}$, d'où

$$\begin{aligned} P(T < +\infty) &= P(\cup_{\{k \in A\}} g_k \text{ est une fonction constante}) \\ &\geq P(\cup_{\{k \in B\}} g_k \text{ est une fonction constante}) = 1. \end{aligned}$$

□

3.3.2 Algorithme de Propp et Wilson pour des dynamiques monotones

On suppose ici que S est un ensemble fini muni d'un ordre partiel, possédant un plus grand élément et un plus petit élément. Un exemple classique de tel ensemble est $S = E^L$, où L est un ensemble fini et E une partie finie de $\mathbb{R}$.

Définition: On dit qu'une dynamique associée à une matrice de Markov M est monotone si on peut construire un ensemble $F \subset \mathcal{F}(S, S)$ et une mesure θ sur $\mathcal{F}(S, S)$ tel que

– $\forall i, j \in S \quad \theta(f \in F, f(i) = j) = m_{i,j}$.

– F ne comprend que des fonctions croissantes.

3.3.3 mise en oeuvre

Dans la pratique, la mesure θ est souvent construite comme l'image d'une mesure aisément simulable par l'application

$$x \mapsto (y \mapsto f(x,y)),$$

où f est une fonction de deux variables qui est croissante par rapport à la deuxième variable.

Théorème 13. *Soit M la matrice d'une chaîne de Markov irréductible admettant μ comme mesure invariante. On suppose qu'on a construit un ensemble $F \subset \mathcal{F}(S,S)$ et une loi θ à support dans F telle que*

- $\forall i,j \in S \quad \theta(f \in F, f(i)=j) = m_{i,j}$.
- *F ne contient que des fonctions croissantes.*

(Ceci signifie que la dynamique est monotone).

Soit maintenant $(f_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes suivant la loi θ.

On pose $g_0 = Id_S$, puis pour $n \geq 0$ $g_{n+1} = g_n \circ f_{n+1}$.

Soit A une partie infinie de $\mathbb{N}$. On note

$$T = \inf\{n \in A, g_n \text{ est une fonction constante}\}.$$

Alors $P(T < +\infty) = 1$ et $g_T(x_0)$ suit la loi μ, où $x_0 \in S$ est quelconque.

Démonstration. Notons min le plus petit élément de S et Max le plus grand élément de S. Commençons par une remarque simple : si une fonction croissante h de S dans S vérifie $h(\min) = \text{Max}$, alors elle est constante, car

$$\forall x \in S; \quad \text{Max} = h(\min) \leq h(x) \leq \text{Max},$$

ce qui montre que h est une fonction constante. Soit $n \in A$ fixé. Pour $k \in \{1,\dots,n\}$, notons $h_k = f_n \circ f_{n-1} \circ \dots f_{n-k+1}$. Il est facile de voir que pour tout $k \in \{1,\dots,n\}$, (h_k constante) implique h_n constante. Comme h_k est croissante (comme composée de fonctions croissantes), on a

$$(h_k(m) = \text{Max}) \Longrightarrow (h_k \text{ constante }) \Longrightarrow (h_n \text{ constante }).$$

Ainsi

$$\begin{aligned} P(T > n) &= P(g_n \text{ non constante}) \\ &= P(h_n \text{ non constante}) \\ &\leq P(\forall k \in \{1,\dots n\}, h_k(\min) \neq M). \end{aligned}$$

Mais la suite $(h_k(\min))_{k\in\{1,\dots,n\}}$ est (un morceau d') une chaîne de Markov partant de m est de Matrice de passage M. Ainsi $P(T > n) \leq \mathbb{P}_{\min}(T_{\mathrm{Max}} > n)$, d'où $P(T = +\infty) \leq \mathbb{P}_{\min}(T_{\mathrm{Max}} = +\infty)$ Comme M est irréductible et admet une mesure invariante, elle est récurrente, on a donc $\mathbb{P}_{\min}(T_{\mathrm{Max}} = +\infty) = 0$, d'où $P(T = +\infty) = 0$. □

3.4 Exercices

1. Programmer une dynamique de Metropolis pour le modèle d'Ising.
2. Programmer la dynamique de l'échantillonneur de Gibbs pour le modèle d'Ising.
3. Montrer que la dynamique de l'échantillonneur de Gibbs pour le modèle d'Ising est monotone.
4. *Évolution d'un génotype avec mutation*
   Nous travaillons ici sur une population de taille fixe formée de $2N$ gènes. Il y a deux types de gènes possibles : le type "a" et le type "A". Chacun des gènes au temps $n+1$ est engendré par deux des $2N$ gènes présents au temps N. Son type est normalement celui d'un de ses deux parents (choisi au hasard). Cependant, au cours du processus de reproduction, des erreurs de codage peuvent se produire : un gène qui aurait du être de type "a" peut se transformer en gène de type "A" avec probabilité $u_1 > 0$, tandis qu'un gène qui aurait du être de type "A" peut se transformer en gène de type "a" avec probabilité $u_2 > 0$.

   On considère la variable aléatoire X_n égale au nombre d'individus de type "A" dans la population à l'étape n.

   (a) Montrer que X_n est une chaîne de Markov à valeurs dans $E = \{0, \dots, 2N\}$.

   (b) Montrer que la loi de X_{n+1} sachant $X_n = k$ est une loi binomiale dont on déterminera les paramètres.

   (c) Montrer que cette chaîne admet une unique probabilité invariante μ.

   (d) Simuler une trajectoire de cette chaîne. A l'aide du théorème ergodique, tracer un histogramme approchant celui de la loi invariante μ.

   (e) On suppose maintenant que $u_1 + u_2 < 1$. Montrer que la dynamique markovienne associée à cette chaîne est monotone.

(f) En utilisant l'algorithme de Propp et Wilson, simuler de manière exacte des échantillons indépendant suivant la loi μ. Tracer l'histogramme correspondant et comparer avec le résultat de la simulation précédente.

Index