

Chaînes de Markov ¹

Olivier Hénard

10 Octobre 2018

1. Notes du cours « Chaînes de Markov » (PRB201) du master M1 MA

Ces notes de cours reprennent dans un seul ensemble (supposé cohérent) une toute petite partie des notions présentées dans le livre de Levin Peres (et Wilmer) :

Levin, D. A., & Peres, Y. Markov chains and mixing times Second Edition (Vol. 107). American Mathematical Soc. (2017)

qui est la référence choisie pour ce cours. Notre ambition en proposant ce texte complémentaire n'est pas de nous substituer à cette référence, mais plutôt d'aider le lecteur novice à trouver son chemin dans cet ouvrage dont l'ambition excède très largement le cadre d'un cours de 6 séances d'une heure.

Notre délicat travail d'élagage a été dicté cette année par un principe simple : parvenir à aborder, au terme des six séances, le problème des temps de recouvrement d'une chaîne de Markov ; le choix de ce problème a lui-même répondu à plusieurs critères : confronter le plus rapidement possible les étudiants à une problématique de recherche récente et abordable¹, sans trop de technicité, et qui offre un panorama des méthodes utilisées dans les probabilités modernes.²

Le chemin que nous avons tracé dans l'ouvrage est donc essentiellement dicté par notre contrainte de temps ; il s'agit du plus court chemin qui a pour point de départ la définition d'une chaîne de Markov et pour point d'arrivée le calcul effectif des temps de recouvrement des tores d -dimensionnels. Le contenu de quasiment toutes ces notes découle de ces deux impératifs, comme pourra le constater un lecteur qui pour comprendre le résultat final sur les temps de recouvrement, déroulerait patiemment la bobine des résultats intermédiaires nécessaires à la compréhension du résultat final.

Le problème des temps de recouvrement fournit un exemple particulièrement caractéristique de la théorie dite "moderne" des chaînes de Markov. Alors que la théorie dite classique des chaînes de Markov était centrée sur la vitesse de convergence en temps long d'une chaîne donnée vers sa mesure stationnaire, la théorie moderne, développée depuis les années 80 et motivée par des applications en algorithmique et physique statistique, consiste plutôt à considérer des familles de chaînes de Markov dont l'espace d'état croît, et à estimer pour ces familles des quantités caractéristiques (ici le temps de recouvrement, c'est-à-dire le temps mis par un marcheur aléatoire sur un graphe pour visiter chaque sommet du graphe) lorsque la taille de l'espace d'état tend vers l'infini.

Si aucun passage de cette promenade dans le monde de l'aléa n'est de difficulté très relevée, la promenade se fera à rythme soutenu, et on ne prendra que peu de temps en route pour apprécier le panorama. Mentionnons ici les quelques jolis points de vue qu'un œil averti ne devra pas manquer pas d'apprécier :

- tout d'abord, le théorème de convergence des chaînes de Markov (et sa démonstration en particulier, hautement généralisable),
- la représentation des mesures stationnaires par les fonctions de Green (attribuée à Aldous-Fill), qui justifie à elle seule l'approche probabiliste,
- le calcul des extensions harmoniques au moyen des chaînes de Markov stoppées en un

1. le calcul du temps de recouvrement du tore fait l'objet d'une publication de l'année 2004, Dembo, A., Y. Peres, J. Rosen, and O. Zeitouni. 2004. Cover times for Brownian motion and random walk in two dimensions, *Ann. Math.* 160, 433-464

2. Il s'agit d'un objectif possible, mais certainement pas du seul ; en un volume horaire similaire, on pourrait aussi présenter les temps stationnaires forts et les méthodes de couplage en vue de l'estimation des temps de mélange, ou encore, de façon un peu plus ambitieuse, le phénomène de cutoff et la condition produit.

- temps d'arrêt (problème de Dirichlet),
- l'équivalence entre les chaînes réversibles et les réseaux électriques, qui culmine avec le principe de Thomson (une description énergétique/variationnelle des fonctions harmoniques),
- l'identité du temps de transport (qui motive pour beaucoup l'intérêt de la résistance équivalente) et la borne de Matthews.

Nous le répétons finalement : l'objectif de ces notes est d'aider l'élève à délimiter les notions du livre dont nous ferons usage plutôt que de se substituer à la lecture du dit ouvrage ; la lecture parallèle des chapitres 1, 2, 9, 10 et 11 reste tout à fait recommandée, et même nécessaire à l'assimilation complète des notions.

Mes remerciements aux étudiants de L3 MFA d'Orsay promotion 2017-2018, Damien Girault et Léo Hahn-Leclerc, dont le mémoire "Chaînes de Markov et Arbres couvrants aléatoires" que j'ai encadré a fournit une base solide et stimulé l'écriture de ces notes.

Notations :

$\mathbb{N} = \{0, 1, 2, 3, \dots\}$, $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$, $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$, $\mathbb{R} =] - \infty, +\infty[$ ont leur signification traditionnelle (française), et l'ordre total \leq dont ces ensembles sont munis (ainsi que sa version stricte $<$) également ; en revanche, le symbole d'inclusion \subset fera toujours référence à une inclusion *large* d'ensembles, c'est-à-dire avec égalité possible (et on précisera en toute lettre le cas d'une inclusion stricte). Aussi, par souci de légèreté, la probabilité conditionnelle $\mathbb{P}(B \cap C)$ sera quelquefois abrégée en $\mathbb{P}(B, C)$, de même pour les probabilités conditionnelles où $\mathbb{P}(A|B, C)$ pourra remplacer $\mathbb{P}(A|B \cap C)$. Si S est un ensemble, $\#S$ et $|S|$ sont deux notations pour son cardinal.

Enfin, par souci de concision et de clarté, on ne précisera pas en général l'espace sur lequel sont définies nos variables aléatoires (Ω ayant déjà un rôle autre), ni la tribu dont on munit cet espace : dans le cas d'un espace d'état discret, ces notions n'ont que peu d'intérêt et ne posent pas de difficultés.

Table des matières

1	Matrices stochastiques	7
1.1	Le semi-groupe des matrices stochastiques.	7
1.2	L'ensemble des mesures stationnaires	8
1.2.1	Définition et existence d'une mesure stationnaire	8
1.2.2	Unicité de la mesure stationnaire et irréductibilité	9
1.3	Convergence et périodicité	11
1.4	Matrices stochastiques et réversibilité	14
1.5	Compléments : spectre et périodicité	15
2	Les chaînes associées aux matrices	17
2.1	Construction des chaînes	17
2.1.1	Exemples	19
2.1.2	Loi marginale de la chaîne	21
2.2	La propriété de Markov	22
2.3	Représentation de la mesure stationnaire par des temps d'arrêt	24
2.4	L'exemple de la ruine du joueur	30
2.5	Quelques mots sur la propriété de Markov forte	32
3	Réseaux électriques (a.k.a. chaînes réversibles)	33
3.1	Chaînes de Markov réversibles et réseau.	33
3.2	Principe de Dirichlet pour les fonctions harmoniques	35
3.3	Tension, flot et flot courant	36
3.4	Résistance équivalente	38
3.5	Temps de transport	41
3.6	Énergie	42
4	Temps d'atteinte et temps de couverture	45
4.1	Cas réversible.	45
4.2	Borne de Matthews (de l'aléa pour construire une borne supérieure)	48
5	Application : temps de couverture du tore	51

Chapitre 1

Matrices stochastiques

C'est un jour pluvieux sur le plateau de Saclay. Et on se prend à regretter le bon vieux temps des khôlles quotidiennes. Et si l'on multipliait des matrices ? Mais pas n'importe quelles matrices, des matrices...stochastiques.

1.1 Le semi-groupe des matrices stochastiques.

Soit Ω un ensemble fini, et $P = (P(x, y))_{x, y \in \Omega}$ une matrice à coefficients réels indicée par Ω . On notera que P est une matrice carrée. Noter qu'on pourrait sans perte de généralité prendre $\Omega = \{1, 2, \dots, n\}$, mais il n'est pas plus gênant de travailler avec un ensemble Ω quelconque, sans compter que cela peut éviter des confusions entre les entrées des matrices (des nombres) et les éléments qui indiquent ces entrées (des éléments de Ω). Aussi, dans les applications à venir, nous prendrons souvent pour Ω les sommets d'un graphe, et il n'est pas toujours naturel d'étiquetter par $\{1, \dots, n\}$ les sommets d'un graphe.

Définition 1.1. On appelle matrice stochastique sur Ω une matrice $P = (P(x, y))_{x, y \in \Omega}$ carrée dont les entrées sont positives et dont les lignes somment à 1 c'est-à-dire :

1. Pour tout $x, y \in \Omega$, $P(x, y) \geq 0$.
2. Pour tout $x \in \Omega$, $\sum_{y \in \Omega} P(x, y) = 1$.

Remarque 1.2. Noter qu'une matrice stochastique n'a rien d'aléatoire, contrairement à ce que son nom peut laisser penser. Le lien avec l'aléa sera explicité en Section 2.

Si $Q = (Q(x, y))_{x, y \in \Omega}$ est une autre matrice indicée par Ω , on rappelle que le produit matriciel PQ est défini par

$$PQ(x, z) = \sum_{y \in \Omega} P(x, y)Q(y, z),$$

il possède notamment la propriété d'associativité, et on note, dans ce cadre, le

Lemme 1.3. Si P et Q sont deux matrices stochastiques sur Ω , alors PQ est encore une matrice stochastique.

Démonstration. La positivité est immédiate, et $\sum_z PQ(x, z) = \sum_z (\sum_y P(x, y)Q(y, z)) = \sum_y P(x, y) (\sum_z Q(y, z)) = \sum_y P(x, y) = 1$. \square

En conséquence, la famille obtenue en considérant les produits successifs de P par elle-même, c'est-à-dire la famille $(P^t)_{t \in \mathbb{N}}$, est une famille de matrices stochastiques. (Pour $t = 0$, on obtient P^0 , par convention la matrice identité, qui est aussi stochastique). L'ensemble des matrices stochastiques forme un ensemble compact (en tant que sous-ensemble fermé du compact $[0, 1]^{\Omega \times \Omega}$, puisque Ω est fini), donc la suite $(P^t)_{t \in \mathbb{N}}$ admet des points d'accumulation. (Le choix de la norme n'est pas important ici). On se demande dans la suite de ce chapitre quand l'ensemble des points d'accumulation est réduit à un point, c'est-à-dire quand la suite converge.

1.2 L'ensemble des mesures stationnaires

On rappelle qu'on peut faire le produit d'une matrice par un vecteur (colonne) par la droite ou par un vecteur (ligne) par la gauche. Si $f = (f(y))_{y \in \Omega}$ et $\pi = (\pi(x))_{x \in \Omega}$, on pose pour tout $x, y \in \Omega$,

$$Pf(x) = \sum_{y \in \Omega} P(x, y)f(y) \text{ et } \pi P(y) = \sum_x \pi(x)P(x, y).$$

On peut aussi former des quantités scalaires comme la quantité $\pi Pf = \sum_{x, y \in \Omega} \pi(x)P(x, y)f(y)$. Si l'on munit l'ensemble des fonctions $f : \Omega \rightarrow \mathbb{R}$ de la norme infinie, alors P est une contraction : $\|Pf\|_\infty \leq \|f\|_\infty$; aussi, si π est une mesure de probabilité, alors πP est encore une mesure de probabilité. Les quantités les plus importantes associées à une matrice sont ses valeurs propres et ses vecteurs propres. Dans les cas des matrices stochastiques, on note que, par définition,

$$P\mathbf{1} = \mathbf{1},$$

avec $\mathbf{1} = (1)_{x \in \Omega}$, d'où :

Lemme 1.4. *1 est valeur propre à droite, et pour le sous-espace propre associé, on a $\text{Ker}(P - I) \supset \text{Vect}(\mathbf{1})$.*

Nous verrons plus loin une condition qui assure l'égalité entre ces deux ensembles.

1.2.1 Définition et existence d'une mesure stationnaire

Définition 1.5. Soit P stochastique. Une mesure de probabilité $\pi = (\pi(x))_{x \in \Omega}$ sur Ω est stationnaire pour P si π est un vecteur propre à gauche de P pour la valeur propre 1, c'est-à-dire que,

$$\pi P = \pi. \tag{1.1}$$

Il s'agit d'une égalité entre vecteurs lignes : pour tout $y \in \Omega$, $\pi P(y) = \pi(y)$, avec $(\pi P)(y) = \sum_x \pi(x)P(x, y)$.

Définition 1.6. Soit P stochastique. On note

$$\mathcal{I}_P = \left\{ \pi = (\pi(x))_{x \in \Omega} : \sum_x \pi(x) = 1 \text{ et } \pi P = \pi \right\}$$

l'ensemble des mesures *de probabilité* stationnaires de P .

Il s'agit d'un sous-ensemble convexe de l'ensemble des mesures de probabilité sur Ω . Notons que, pour tout $t \in \mathbb{N}$, $\mathcal{I}_P \subset \mathcal{I}_{P^t}$, c'est-à-dire qu'une mesure de probabilité stationnaire pour P l'est aussi pour P^t , puisque $\pi = \pi P$ implique $\pi P = \pi P^2, \pi P^2 = \pi P^3, \dots, \pi P^{t-1} = \pi P^t$ en multipliant par la droite par P, \dots, P^{t-1} , et partant $\pi = \pi P^t$.

Maintenant, sous quelles conditions \mathcal{I}_P est-il non vide? réduit à un seul élément? Nous commençons par répondre par l'affirmative à la première question, au moyen d'un argument de type Césaro.

Proposition 1.7 (Existence). *Soit P stochastique. Alors $\mathcal{I}_P \neq \emptyset$.*

Remarque 1.8. On rencontre aussi la terminologie "mesure invariante".

Insistons sur le fait que notre espace d'état est fini. Quand ce n'est pas le cas, \mathcal{I}_P peut être vide, il peut ne pas exister de mesure *de probabilité* stationnaire : considérer l'exemple de la matrice stochastique associée au shift vers la droite sur \mathbb{Z} (infini), donné par $P(x, y) = 1_{y=x+1}$ pour tout $x, y \in \mathbb{Z}$, est à cet égard instructif : seule la mesure de comptage et ses multiples sont des mesures stationnaires, mais elles ne sont pas de probabilité.

Démonstration. Soit μ mesure de probabilité sur Ω . Posons, pour $t \in \mathbb{N}$, $\nu_t = \frac{1}{t} \sum_{s=0}^{t-1} \mu P^s$. L'intérêt de cette expression est que la différence fait apparaître une somme télescopique :

$$\nu_t P - \nu_t = \frac{1}{t} \sum_{s=1}^t \mu P^s - \frac{1}{t} \sum_{s=0}^{t-1} \mu P^s = \frac{1}{t} (\mu P^t - \mu).$$

Soit $x \in \Omega$. On a

$$|\nu_t P(x) - \nu_t(x)| = \frac{1}{t} |\mu P^t(x) - \mu(x)| \leq \frac{1}{t} (|\mu P^t(x)| + |\mu(x)|) \leq \frac{2}{t}.$$

De plus, la suite $(\nu_t)_{t \in \mathbb{N}}$ est à valeurs dans $[0, 1]^\Omega$ qui est un compact. On peut donc extraire une sous-suite (ν_{t_k}) qui converge vers une limite ν , et $t_k \geq k$ par définition. On a alors $|\nu_{t_k} P(x) - \nu_{t_k}(x)| \leq \frac{2}{t_k} \leq \frac{2}{k}$, et l'on conclut par continuité que $\nu P(x) = \nu(x)$. \square

1.2.2 Unicité de la mesure stationnaire et irréductibilité

Il n'y a aucune raison pour que la mesure stationnaire soit unique : il suffit pour cela de considérer le cas par exemple de P_1 et P_2 deux matrices stochastiques sur Ω_1 et Ω_2 respectivement, et deux mesures invariantes π_1 et π_2 pour ces matrices stochastiques. Alors on forme sur la réunion disjointe $\Omega = \Omega_1 \cup \Omega_2$ une matrice stochastique

$$P = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix} \quad (1.2)$$

alors les mesures π_1 et π_2 se prolongent sur Ω (en attachant la mesure nulle aux ensembles sur lesquelles elles ne sont pas définies) et ces deux mesures sont deux mesures de probabilité stationnaires distinctes. La question de l'unicité de la mesure invariante réclame donc une nouvelle définition.

Définition 1.9. On dit que P est irréductible lorsque, pour tout $x, y \in \Omega$, il existe $t = t(x, y) \in \mathbb{N}$ tel que $P^t(x, y) > 0$.

Cela équivaut au fait de dire que P n'est pas similaire, via une matrice de *permutation* S , à une matrice triangulaire Q par blocs (avec plus d'un bloc évidemment, c'est-à-dire à une matrice de la forme 1.2), c'est-à-dire que la relation $P = S^{-1}QS$ ne vaut pas quelque soit S matrice de permutation et Q matrice triangulaire par blocs. On n'utilisera pas cette caractérisation en revanche.

La notion de fonction harmonique "précise" de façon plus locale celle de vecteur propre à droite de valeur propre 1 pour P .

Définition 1.10. Soit P stochastique, $h : \Omega \rightarrow \mathbb{R}$, $x \in \Omega$ et $\Omega' \subset \Omega$.

- On dit que h est harmonique en x si $Ph(x) = h(x)$.
- On dit que h est harmonique sur Ω' si h est harmonique en tout point de Ω' .

En particulier, h harmonique sur Ω est donc un vecteur propre à droite pour P . L'irréductibilité a une implication immédiate sur l'espace propre associé à la valeur propre 1, et permet de préciser le lemme 1.4.

Proposition 1.11. Soit P irréductible. Si h est harmonique sur Ω entier, alors h est constante. En d'autres termes,

$$\text{Ker}(P - I) = \text{Vect}(\mathbf{1}).$$

Démonstration. Soit $x \in \Omega$ qui maximise h , et $y \in \Omega$. Il existe $t \in \mathbb{N}$ tel que $P^t(x, y) > 0$. Ainsi : $h(x) = P^t h(x) = \sum_z P^t(x, z)h(z)$ implique, pour tout z tel que $P^t(x, z) > 0$, $h(z) = h(x)$. Ceci vaut en particulier pour y . \square

Lemme 1.12 (Positivité). Soit P irréductible. Si $\pi \in \mathcal{I}_P$, alors pour tout $x \in \Omega$, $\pi(x) > 0$.

Démonstration. Puisque π est une mesure de probabilité, $\sum_{z \in \Omega} \pi(z) = 1$ donc il existe $y \in V$ tel que $\pi(y) > 0$. Soit maintenant $x \in \Omega$. P est irréductible donc il existe $t \in \mathbb{N}$ tel que $P^t(y, x) > 0$. Donc

$$\pi(x) = \sum_{z \in V} \pi(z)P^t(z, x) \geq \pi(y)P^t(y, x) > 0.$$

\square

La question de l'unicité de la mesure stationnaire est dès lors très vite tranchée.

Proposition 1.13 (Unicité). Soit P irréductible. $\#\mathcal{I}_P = 1$

Démonstration. Il suffit au vu de la proposition 1.7 de prouver l'unicité. Nous proposons deux démonstrations, chacune basée sur un des deux résultats précédents.

1. Soient $\pi_1, \pi_2 \in \mathcal{I}_P$. L'application $z \mapsto \pi_1(z)/\pi_2(z)$ est bien définie par le lemme de positivité 1.12, et on note x un élément en lequel cette application atteint un minimum. Si cette application n'est pas constante, alors il existe $y \in V$ tel que $\pi_1(y)/\pi_2(y) < \pi_1(x)/\pi_2(x)$. P est irréductible donc il existe $t \in \mathbb{N}$ tel que $P^t(y, x) > 0$. Puisque

$\pi_1 \in \mathcal{I}_P \subset \mathcal{I}_{P^t}$, on a :

$$\begin{aligned} \pi_1(x) &= \sum_{z \in V} \frac{\pi_1(z)}{\pi_2(z)} \pi_2(z) P^t(z, x) && \text{en forçant l'apparition du terme } \pi_2(z) \\ &> \sum_{z \in V} \frac{\pi_1(x)}{\pi_2(x)} \pi_2(z) P^t(z, x) && \text{car } P^t(y, x) > 0 \\ &= \frac{\pi_1(x)}{\pi_2(x)} \pi_2(x) && \text{car } \pi_2 \in \mathcal{I}_P \subset \mathcal{I}_{P^t} \\ &= \pi_1(x), \end{aligned}$$

ce qui est absurde. On en tire que pour tout $y \in \Omega$, $\pi_1(x)/\pi_2(x) = \pi_1(y)/\pi_2(y)$, c'est-à-dire que π_1 et π_2 sont deux mesures de probabilité proportionnelles, ce qui signifie qu'elles sont égales.

2. $\text{Ker}(P - I) = \text{Vect}(\mathbf{1})$ implique $\dim(\text{Ker}(P - I)) = 1$ et donc, du théorème du rang, $\text{rg}(P - I) = n - 1$. En passant à la transposée ici notée $^\top$ (à ne pas confondre avec t la puissance t -ième de la matrice), $\text{rg}(P - I)^\top = n - 1$ et finalement $\dim(\text{Ker}((P - I)^\top)) = 1$. Ceci assure que l'équation $\pi P = \pi$ équivalente à $(P - I)^\top \pi^\top = 0$ équivalente à $\pi^\top \in \text{Ker}((P - I)^\top)$ admet une et une seule mesure *de probabilité* solution. □

Une obstruction claire à l'irréductibilité est par exemple la présence d'états absorbants au sens suivant :

Définition 1.14. Soit P matrice stochastique sur Ω . Un état $x \in \Omega$ est dit absorbant pour P si $P(x, x) = 1$.

Lemme 1.15. *S'il existe un état absorbant pour P , alors P n'est pas irréductible.*

Démonstration. En effet, puisque $P(x, x) = 1$, et si $y \neq x$, alors pour tout $t \in \mathbb{N}$, $P^t(x, y) \leq \sum_{z \neq x} P^t(x, z) = 1 - P^t(x, x) = 1 - 1 = 0$. □

1.3 Convergence et périodicité

Nous revenons maintenant à la question de la convergence de la suite $(P^t)_{t \in \mathbb{N}}$ - sans lien avec l'unicité de la mesure stationnaire. Il peut exister des phénomènes de périodicité, illustrés par l'exemple de la matrice stochastique

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

dont les puissances impaires sont égales à P , et les puissances paires à la matrice identité. La description de ces phénomènes nous amène à poser une nouvelle définition.

Définition 1.16. Soit $x \in \Omega$. On pose $\mathcal{T}(x) := \{t \in \mathbb{N}^*, P^t(x, x) > 0\}$ et on appelle *période de x* l'entier $\text{pgcd}(\mathcal{T}(x))$.

Le pgcd d'un sous-ensemble non vide S de \mathbb{N} est le plus grand des diviseurs communs de S (qui existe bien puisque cet ensemble est non vide - il contient 1 - et majoré - par le plus petit élément de S), c'est-à-dire du maximum de l'ensemble $A_S = \{a \in \mathbb{N}^*, S \subset a\mathbb{N}\}$. Par le théorème de Bachet Bézout, si $\mathbb{Z}[S]$ désigne l'ensemble des combinaisons linéaires à coefficients entiers relatifs d'éléments de S , alors

$$\mathbb{Z}[S] = g.\mathbb{Z}, \text{ où } g = \text{pgcd}(S).$$

Proposition 1.17. *Soit P stochastique irréductible. Les éléments de Ω ont tous la même période.*

Démonstration. Soient $x, y \in \Omega$. Il suffit de démontrer que $\text{pgcd}(\mathcal{T}(x)) = \text{pgcd}(\mathcal{T}(y))$. P est irréductible donc il existe $t_1, t_2 \in \mathbb{N}$ tels que $P^{t_1}(y, x) > 0$ et $P^{t_2}(x, y) > 0$. On pose $t_0 = t_1 + t_2$. Alors, si $t \in \mathcal{T}(x)$.

$$P^{t_0+t}(y, y) = \sum_{z \in \Omega} P^{t_1}(y, z)P^t(z, z)P^{t_2}(z, y) \geq P^{t_1}(y, x)P^t(x, x)P^{t_2}(x, y) > 0,$$

d'où $t + t_0 \in \mathcal{T}(y)$, et en particulier, en prenant $t = 0$ dans l'expression ci-dessus $t_0 \in \mathcal{T}(y)$. En terme de pgcd, cela signifie que $\text{pgcd}(\mathcal{T}(y))$ divise t_0 et $t + t_0$ et donc aussi leur différence t . Mais puisque cela vaut pour tout $t \in \mathcal{T}(x)$, ceci implique $\text{pgcd}(\mathcal{T}(y)) \leq \text{pgcd}(\mathcal{T}(x))$. Par symétrie, $\text{pgcd}(\mathcal{T}(x)) \leq \text{pgcd}(\mathcal{T}(y))$ vaut également. Et finalement $\text{pgcd}(\mathcal{T}(x)) = \text{pgcd}(\mathcal{T}(y))$, comme attendu. \square

Cette proposition rend licite la définition suivante :

Définition 1.18. — Soit P irréductible. On appelle période de P la période d'un élément quelconque de Ω .

— On dit que P est *apériodique* si cette période vaut 1.

La proposition clef qu'on montre sous l'hypothèse d'apériodicité est la suivante :

Proposition 1.19. *Si P est irréductible apériodique, alors il existe $\varepsilon > 0$ et $t_0 \in \mathbb{N}$ tel que*

$$\text{pour tout } x, y \in \Omega, \quad P^{t_0}(x, y) \geq \varepsilon. \quad (1.3)$$

La preuve de cette proposition nécessite un lemme d'arithmétique élémentaire, que nous ne démontrerons pas (voir par exemple le livre de Levin Peres) :

Lemme 1.20 (Lemme de Schur). *Un sous-ensemble $S \subset \mathbb{N}$ stable par somme ($t, s \in S \Rightarrow t + s \in S$) tel que $\text{pgcd}(S) = 1$ vérifie :*

$$\#(\mathbb{N} \setminus S) < \infty,$$

c'est-à-dire que S contient tous les entiers sauf un nombre fini.

Démonstration de la Proposition 1.19. Soit $x \in \Omega$. Observons que $\mathcal{T}(x)$ est stable par somme et que $\text{pgcd}(\mathcal{T}(x)) = 1$ par apériodicité. Le lemme de Schur assure alors qu'il existe $t(x)$ tel que pour tout $t \geq t(x)$, $t \in \mathcal{T}(x)$. De plus, pour tout $y \in \Omega$, toujours par irréductibilité, il existe $t(x, y) \in \mathbb{N}$ tel que $P^{t(x, y)}(x, y) > 0$. Donc, pour tout $t \geq t(x) + t(x, y)$, $P^t(x, y) > 0$. Ainsi, pour $t_0 := \max_{x \in \Omega}(t(x) + \max_{y \in \Omega} t(x, y))$, on a pour tout x, y , $P^{t_0}(x, y) > \min_{x, y} P^{t_0}(x, y) > 0$ en utilisant que l'espace d'état est fini. \square

Il reste maintenant un petit pas pour arriver au théorème de convergence, qui est le résultat fondamental de ce chapitre, sinon du cours. On note $\lfloor t \rfloor$ la partie entière de t (définie de façon unique par les deux propriétés : $\lfloor t \rfloor \leq t < \lfloor t \rfloor + 1$ et $\lfloor t \rfloor \in \mathbb{Z}$).

Théorème 1.21. *Soit P matrice stochastique irréductible apériodique, et π son unique mesure stationnaire. Pour ε et t_0 qui satisfont (1.3), on a pour tout $(x, y) \in \Omega^2$,*

$$\sum_y |P^t(x, y) - \pi(y)| \leq 2(1 - \varepsilon)^{\lfloor t/t_0 \rfloor}.$$

Ce résultat implique que $P^t(x, y)$ admet quand $t \rightarrow \infty$ une limite $\pi(y)$ qui ne dépend pas de l'entrée x . En termes de matrice, P^t converge donc vers la matrice de rang 1 (on rappelle que le rang d'une matrice est la dimension de l'image de l'application linéaire associée) dont les lignes sont toutes égales à π . Cette matrice est bien sûr encore une matrice stochastique, puisque cet ensemble est fermé. Si l'on sait que t_0 et ε existent de la propriété d'apériodicité, il importe en pratique de trouver des valeurs numériques de façon à maximiser $(1 - \varepsilon)^{1/t_0}$: c'est ce qui est difficile. Le fait que la vitesse de convergence soit toujours exponentielle est encore une simplification liée à notre espace d'état fini.

Démonstration d'après Aldous–Diaconis. On pose Π la matrice carrée dont toutes les lignes sont égales à π , et on vérifie immédiatement que $P\Pi = \Pi P = \Pi$ (seule la deuxième identité requiert $\pi \in \mathcal{I}_P$). Pour ε et t_0 fournis par la relation 1.19, on définit par la relation :

$$P^{t_0} = \varepsilon\Pi + (1 - \varepsilon)Q$$

une matrice Q à coefficients positifs ou nuls, dont on vérifie sans souci qu'elle est stochastique. De plus, multipliant à gauche par Π , on a $\Pi = \varepsilon\Pi + (1 - \varepsilon)Q\Pi$ d'où $Q\Pi = \Pi$. Et multipliant à droite par Π , on a : $\Pi = \varepsilon\Pi + (1 - \varepsilon)Q\Pi$ d'où $\Pi Q = \Pi$. Soit $k \in \mathbb{N}$. Ces deux relations conduisent aux simplifications suivantes lorsqu'on applique la formule du binôme de Newton :

$$\begin{aligned} P^{t_0 k} &= ((1 - \varepsilon)Q + \varepsilon\Pi)^k \\ &= (1 - \varepsilon)^k Q^k + \sum_{j=0}^{k-1} \binom{k}{j} (1 - \varepsilon)^j \varepsilon^{k-j} \Pi && \text{car } \Pi Q = Q\Pi = \Pi \text{ et } \Pi^2 = \Pi \\ &= (1 - \varepsilon)^k Q^k + \sum_{j=0}^k \binom{k}{j} (1 - \varepsilon)^j \varepsilon^{k-j} \Pi - (1 - \varepsilon)^k \Pi \\ &= (1 - \varepsilon)^k Q^k + \Pi - (1 - \varepsilon)^k \Pi \\ &= (1 - \varepsilon)^k (Q^k - \Pi) + \Pi. \end{aligned}$$

Il faut aussi prendre en compte le reste que peut laisser la division euclidienne d'un entier arbitraire par t_0 . Pour $r \in \{0, \dots, t_0 - 1\}$, on forme donc la différence

$$\begin{aligned} P^{t_0 k + r} - \Pi &= (P^{t_0 k} - \Pi)P^r && \text{car } \Pi = \Pi P^r \\ &= (1 - \varepsilon)^k (Q^k - \Pi) P^r && \text{du calcul précédent} \\ &= (1 - \varepsilon)^k (Q^k P^r - \Pi) \end{aligned}$$

Ainsi $\sum_y |P^{t_0 k + r}(x, y) - \pi(y)| \leq (1 - \varepsilon)^k \sum_y |Q^k P^r(x, y) - \pi(y)| \leq 2(1 - \varepsilon)^k$ en notant que $Q^k P^r$ est stochastique. \square

Une analyse de la démonstration montre que la propriété clef est la minoration "pour tout $x, y \in \Omega$, $P^{t_0}(x, y) \geq \varepsilon \geq \varepsilon \pi(y)$." C'est l'inégalité entre les membres extrêmes que l'on utilise dans la preuve, c'est aussi celle qui s'étend aux espaces d'états infinis (en effet, la première inégalité ne peut avoir lieu lorsque Ω est infini (sommer sur y)).

Nous verrons enfin en TD que les itérées successives $(P^t)_{t \in \mathbb{N}}$ convergent toujours au sens de Césaro, et ceci sans faire aucune hypothèse d'apériodicité en particulier, c'est -à-dire que :

$$\frac{\sum_{0 \leq s < t} P^s}{t} \longrightarrow \Pi \text{ quand } t \rightarrow \infty.$$

De plus, chaque ligne de la matrice limite Π fournit une mesure stationnaire de P (mais si l'on ne suppose pas l'irréductibilité de P , ces mesures ne sont pas forcément égales).

1.4 Matrices stochastiques et réversibilité

Finissons ce chapitre en introduisant une dernière notion : la réversibilité. On peut voir cette notion sous deux aspects : d'une part elle exprime une forme d'invariance en loi par renversement du temps, d'autre part, lorsqu'elle est vérifiée, elle permet de trouver de façon économique une mesure invariante.

Définition 1.22. Soit P stochastique et π une mesure sur Ω . P est dite *réversible* par rapport à π si pour tout $x, y \in \Omega$

$$\pi(x)P(x, y) = \pi(y)P(y, x). \quad (1.4)$$

Par extension, on dira simplement que P est réversible s'il existe une mesure de probabilité π tel que P soit réversible par rapport à π .

Noter que l'équation (1.4) est automatiquement vérifiée sur la diagonale $\{x = y\}$. Noter que la réversibilité de P par rapport à π est équivalente à l'énoncé : pour tout $x_1, \dots, x_n \in \Omega^n$,

$$\pi(x_1)P(x_1, x_2)P(x_2, x_3) \dots P(x_{n-1}, x_n) = \pi(x_n)P(x_n, x_{n-1})P(x_{n-1}, x_{n-2}) \dots P(x_2, x_1)$$

qui implique, si P est irréductible (à l'aide du lemme 1.12) et si $x_1 = x_n$, que

$$P(x_1, x_2)P(x_2, x_3) \dots P(x_{n-1}, x_n) = P(x_n, x_{n-1})P(x_{n-1}, x_{n-2}) \dots P(x_2, x_1) \quad (1.5)$$

Réciproquement, on peut vérifier (il s'agit d'un exercice assez non trivial) que si tout suite $x_1 \dots x_n \in \Omega^n$ avec $x_1 = x_n$ satisfait (2.1), alors P est réversible. C'est le critère dit de Kolmogorov, qui permet de vérifier la réversibilité sans connaître π .

L'intérêt de la réversibilité est exprimée par la proposition ci dessous :

Proposition 1.23. Si P est réversible par rapport à π , alors π est une mesure stationnaire pour P .

Démonstration. Soit $x \in \Omega$. Il suffit d'appliquer la définition :

$$\pi P(x) = \sum_{y \in V} \pi(y)P(y, x) = \sum_{y \in V} \pi(x)P(x, y) = \pi(x)$$

□

Au vu de la simplification que constitue la réversibilité dans la recherche des mesures stationnaires, il importe de comprendre quand une matrice stochastique P a des chances d'être réversible. Nous proposons plusieurs pistes en ce sens.

D'abord, on peut formuler une condition suffisante de réversibilité à l'aide des seules entrées non nulles de la matrice P , c'est-à-dire à l'aide du seul graphe dirigé induit par la matrice stochastique, ainsi défini : on associe à une matrice stochastique un graphe dirigé $\vec{G} = (V, \vec{E})$ d'ensemble de sommets $V = \Omega$ et d'ensemble d'arêtes dirigées \vec{E} donné par : $x\vec{y} \in \vec{E}$ ssi $P(x, y) > 0$. La version non dirigée $G = (V, E)$ du graphe dirigé \vec{G} est définie naturellement en posant $\{x, y\} \in E$ ssi $(x\vec{y} \in \vec{E} \text{ ou } y\vec{x} \in \vec{E})$. On a alors le résultat suivant :

Lemme 1.24. *Soit P stochastique irréductible, et G le graphe associé à P . Si G est un arbre, c'est-à-dire un graphe sans cycle¹ alors P est réversible.*

La démonstration, omise, prend la forme suivante : s'étant donné un sommet $x \in \Omega$, il suffit de déterminer $\pi(y)$ en fonction de $\pi(x)$ pour tout sommet $y \in \Omega$ de proche en proche, ce qui est toujours possible du fait de l'absence de cycles (on peut formaliser ce raisonnement en exercice). Nous verrons plus loin, à la proposition 3.5, une caractérisation (CNS) des matrices stochastiques réversibles, formulée en terme de toute la matrice stochastique cette fois, et s'applique à des graphes plus généraux.

Enfin, il est bon de noter qu'une matrice symétrique est réversible (quelle mesure π choisir alors ?), comme on le verra en exercice.

Notons pour terminer la conséquence suivante facile de la réversibilité sur la période.

Lemme 1.25. — *Soit x tel que $\pi(x) > 0$ et P réversible. Alors la période de x est au plus 2.*

— *Si P est de plus irréductible, alors la période de P est au plus 2.*

Démonstration. Soit x comme dans le premier énoncé. Il existe y tels que $P(x, y) > 0$. Ensuite $\pi(x)P(x, y) = \pi(y)P(y, x)$ implique, puisque $\pi(x) > 0$ que $P(y, x)$ et partant $P(x, y)P(y, x)$ sont des quantités strictement positives. Mais $P^2(x, x) \geq P(x, y)P(y, x)$ donc $2 \in \mathcal{T}(x)$. \square

1.5 Compléments : spectre et périodicité

Notons que la périodicité d'une matrice stochastique a une traduction spectrale simple en terme des valeurs propres de module 1. On rappelle que, pour $a \in \mathbb{N}^*$, une racine a -ième de l'unité est une solution de l'équation $z^a = 1$, d'inconnue complexe $z \in \mathbb{C}$. Cet ensemble est explicite, il s'agit de

$$\{e^{\frac{2ik\pi}{a}}, k \in \{0, \dots, a-1\}\}.$$

Proposition 1.26 (Condition nécessaire d'apériodicité). *Soit P stochastique irréductible et ω une racine a -ième de l'unité. Alors $\mathcal{T}(x) \subset a\mathbb{N}$ si, et seulement si, ω est une valeur propre de P .*

Démonstration. Supposons $\mathcal{T}(x) \subset a\mathbb{N}$ et notons ω une racine a -ième de l'unité. Soit $x \in \Omega$ et, pour $k \in \{0, \dots, a-1\}$,

$$\Omega_k = \{y \in \Omega : \exists t \in a\mathbb{N} + k, P^t(x, y) > 0\}.$$

1. pour un graphe $G = (V, E)$, on appelle cycle une suite d'arêtes $\{x_1, x_2\}, \{x_2, x_3\}, \dots, \{x_{n-1}, x_n\} \in E^{n-1}$ deux à deux distinctes telles que $x_1 = x_n$

Puisque la chaîne est irréductible, on a la décomposition $\Omega = \Omega_0 \cup \dots \cup \Omega_{a-1}$; de plus, la réunion est disjointe. Supposons en effet qu'il existe $y \in \Omega$ avec $y \in \Omega_i \cap \Omega_j$, $0 \leq i \neq j \leq a-1$. Il existe alors $s \in a\mathbb{N} + i$ et $t \in a\mathbb{N} + j$ tels que $P^s(x, y), P^t(x, y) > 0$. De plus, par irréductibilité de P , il existe $r \in \mathbb{N}$ avec $P^r(y, x) > 0$ et donc $s+r, t+r \in \mathcal{T}(x)$, et donc a divise ces deux éléments ainsi que leur différence, ce qui est absurde. Notons aussi que, si $x, y \in \Omega$ vérifient $P(x, y) > 0$ et $x \in \Omega_k$ alors $y \in \Omega_{k+1 \bmod a}$. On pose maintenant $v = \mathbf{1}_{\Omega_0} + \omega \mathbf{1}_{\Omega_1} + \dots + \omega^{a-1} \mathbf{1}_{\Omega_{a-1}}$. Il découle $(Pv)(x) = \omega v(x)$ pour tout $x \in \Omega$: v est bien vecteur propre pour la valeur propre ω .

Supposons réciproquement ω valeur propre de P et montrons que $T(x) \subset a\mathbb{N}$. Soit v vecteur propre associé à la valeur propre ω , et supposons sans perte de généralité que $v(x) = \omega^{-k}$ pour $k \in \{0, \dots, a-1\}$ et pour tout $y \in \Omega$, $|v(y)| \leq 1$. On observe, si $t \in a\mathbb{N} + k$,

$$1 = \omega^k v(x) = \omega^t v(x) = P^t v(x) = \sum_{y \in V} P^t(x, y) v(y)$$

Maintenant une somme pondérée (par une mesure de probabilité) de complexes de module inférieurs ou égal à 1 ne peut valoir 1 que si tous les complexes valent, c'est-à-dire que $P^t(x, y) > 0$ implique $v(y) = 1$. En particulier, $P^t(x, x) = 0$. Comme ceci vaut pour tout $t \in a\mathbb{N} + k$, on en déduit bien $T(x) \subset a\mathbb{N}$. \square

Chapitre 2

Les chaînes associées aux matrices

On se propose dans ce chapitre d'explorer le lien entre les évolutions aléatoires connues sous le nom de chaînes de Markov et les matrices stochastiques étudiées dans le chapitre 1. L'approche classique des chaînes de Markov commence par énoncer une propriété d'indépendance conditionnelle, dite propriété de Markov, puis de montrer que cette propriété définit des suites de variables aléatoires dont l'évolution à un pas est décrite par une matrice stochastique. Nous prenons la chose à revers en définissant les chaînes de Markov par leur loi exprimée au moyen d'une matrice stochastique.

Un cas d'étude important sera l'étude des marches aléatoires sur des graphes finis, qui se trouve relié aux matrices stochastiques réversibles.

2.1 Construction des chaînes

Une probabilité sur un ensemble fini (ou dénombrable) Ω est une collection de nombres qui somment à 1, et sur un tel ensemble, on prend en général la tribu discrète engendrée par les singletons, qui est égale à l'ensemble des parties de Ω : tout ensemble est alors mesurable. En revanche, dès qu'on travaille sur un espace non dénombrable, la précision de la tribu a son importance.

Considérons l'ensemble $\Omega^{\mathbb{N}}$ des suites à valeurs dans Ω . Excepté le cas trivial où Ω est pas réduit à un point, cet ensemble n'est pas dénombrable (on le montre par un argument de diagonalisation à la Cantor, voir la fiche wiki "Cantor's diagonal argument"). Le premier point est de munir cet ensemble d'une tribu (famille de parties stable par passage au complémentaire, par réunion dénombrable, et qui contient l'ensemble entier). La tribu de toutes les parties de $\Omega^{\mathbb{N}}$ est trop grosse en général. Deux projections seront importantes dans ce cadre :

$$\pi_s : \Omega^{\mathbb{N}} \rightarrow \Omega, (x_t)_{t \in \mathbb{N}} \rightarrow x_s \text{ et également } \pi_{\leq s} : \Omega^{\mathbb{N}} \rightarrow \Omega^{s+1}, (x_t)_{t \in \mathbb{N}} \rightarrow (x_t)_{0 \leq t \leq s}.$$

Les tribus associées, notées :

$$\sigma\{\pi_s, s \in \mathbb{N}\} \text{ et } \sigma\{\pi_{\leq s}, s \in \mathbb{N}\},$$

sont par définition les plus petite tribus sur $\Omega^{\mathbb{N}}$ qui rende chacune des projections mesurables. Ces deux tribus sont égales (le montrer), et on appelle la tribu résultante la tribu cylindrique. Celle-ci n'est pas la tribu de toutes les parties de $\Omega^{\mathbb{N}}$. Notons que, pour s fixé, la tribu $\sigma\{\pi_{\leq s}\}$ est explicite :

$$A \in \sigma\{\pi_{\leq s}\} \text{ ssi il existe } B \in \Omega^{s+1} \text{ tel que } A = (\pi_{\leq s})^{-1}(B) = \{x = (x_t)_{t \geq 0}, \pi_{\leq s}(x) \in B\}.$$

En particulier, les cylindres suivants

$$\bigcap_{0 \leq s \leq t} \pi_s^{-1}(\{y_s\}) = \{(x_t)_{t \geq 0}, \forall 0 \leq s \leq t, x_s = y_s\} \in \sigma\{\pi_s, s \in \mathbb{N}\}.$$

sont dans la tribu en question, et engendrent cette tribu. Par exemple, l'événement que $(x_t)_{t \in \mathbb{N}}$ est constant à partir d'un certain rang est dans la tribu cylindrique puisqu'il admet l'expression suivante, en terme de réunions et d'intersections dénombrables :

$$\begin{aligned} \{(x_t)_{t \in \mathbb{N}} \text{ constant à partir d'un certain rang}\} &= \{\exists x \in \Omega, \exists t \in \mathbb{N} : \forall s \in \mathbb{N}, (s \geq t \Rightarrow x_s = x)\} \\ &= \bigcup_{x \in \Omega} \bigcup_{t \in \mathbb{N}} \bigcap_{s \geq t} \{x_s = x\} \end{aligned}$$

Maintenant, le modèle probabiliste depuis Kolmogorov est le suivant. On suppose qu'il existe un espace probabilisé (bien souvent non explicite), $(E, \mathcal{F}, \mathbb{P})$, appelé l'espace des événements, et une application mesurable

$$(E, \mathcal{F}, \mathbb{P}) \rightarrow \Omega^{\mathbb{N}}, \quad \omega \mapsto (X_t(\omega))_{t \in \mathbb{N}} \quad (2.1)$$

Ce qu'on appelle alors loi de X_1 est la mesure image de \mathbb{P} par l'application $\omega \mapsto X_1(\omega)$, c'est une mesure de probabilité sur Ω , notée $\mathbb{P}(X_1 \in \cdot) = \mathbb{P}(\{\omega : X_1(\omega) \in \cdot\})$. En tant que mesure de probabilité sur un espace fini, cette mesure est bien sûr caractérisée simplement par la collection des nombres $(\mathbb{P}(X_1 = k))_{k \in \mathbb{N}}$. Dire que l'application (2.1) est mesurable, c'est dire que pour tout ensemble A de la tribu cylindrique sur $\Omega^{\mathbb{N}}$, $\{\omega : X_t(\omega) \in A\} \in \mathcal{F}$, mais aussi, par définition de la tribu cylindrique, que pour tout $t \in \mathbb{N}, x \in \Omega$, $\{\omega : X_t = x\} = \{X_t = x\} \in \mathcal{F}$. Maintenant, on peut considérer la plus petite sous-tribu de \mathcal{F} qui rend les applications X_0, X_1, \dots, X_t mesurables :

Définition 2.1. Pour tout $t \in \mathbb{N}$, on pose $\mathcal{F}_t = \sigma\{X_0, X_1, \dots, X_t\}$ la plus petite sous tribu qui rend mesurable les applications coordonnées X_0, X_1, \dots, X_t , et \mathcal{F}_∞ la plus petite tribu qui comprend tous les \mathcal{F}_t pour $t \in \mathbb{N}$.

(Point de détail : \mathcal{F}_∞ n'est pas nécessairement égal à $\cup_{t \geq 0} \mathcal{F}_t$; en effet, une réunion de tribus n'est pas en général une tribu). Pour tout, \mathcal{F}_t est une sous tribu de \mathcal{F} , et la suite $(\mathcal{F}_t)_{t \in \mathbb{N}}$ est une sous-suite croissante de sous tribus de \mathcal{F} . On peut alors montrer le résultat fondamental suivante : dire qu'une fonction $F : E \rightarrow \mathbb{R}$ est une fonction mesurable signifie alors qu'il existe une fonction mesurable de $G : \Omega^{t+1} \rightarrow \mathbb{R}$ telle que

$$F(\omega) = G(X_0(\omega), \dots, X_t(\omega)).$$

On définit la loi d'une chaîne de Markov en définissant une mesure de probabilité sur $(\cup_{t \in \mathbb{N}} \mathcal{F}_t)$. Pour cela il suffit de définir la mesure des cylindres. Le théorème suivant est aussi une définition. Nous omettons sa démonstration¹.

Theorème 2.2. Soit μ mesure de probabilité sur Ω et P matrice stochastique sur Ω . La propriété

$$\forall t \in \mathbb{N}, (x_s)_{0 \leq s \leq t} \in \Omega^{t+1}, \quad \mathbb{P} \left(\bigcap_{s=0}^t \{X_s = x_s\} \right) = \mu(x_0) P(x_0, x_1) \cdots P(x_{t-1}, x_t) \quad (2.2)$$

1. le lecteur intéressé pourra googler "théorème de Daniell-Kolmogorov" ou "théorème d'extension de Kolmogorov"

définit de façon unique la loi d'une suite de variables aléatoires $(X_t)_{t \in \mathbb{N}}$. Cette suite est la chaîne de Markov de distribution initiale μ et de matrice de transition P .

Rappelons la notation δ_x pour la masse de Dirac en x , définie par

$$\delta_x(A) = \mathbf{1}_A(x) \text{ pour tout } x \in \Omega, A \subset \Omega.$$

Si la chaîne de Markov est définie sur un espace probabilisé dont la mesure de probabilité est notée \mathbb{P} , on notera \mathbb{P}_μ par un léger abus² la loi de la chaîne de Markov issue de μ , et si $\mu = \delta_x$, on note simplement $\mathbb{P}_x = \mathbb{P}_{\delta_x}$ la loi de la chaîne issue de x ,

Lemme 2.3. *Soit μ mesure de probabilité sur Ω . On a la décomposition :*

$$\mathbb{P}_\mu = \sum_{x \in \Omega} \mu(x) \mathbb{P}_x$$

Démonstration. Il suffit de le vérifier pour un événement $A = \bigcap_{s=0}^t \{X_s = x_s\}$

$$\begin{aligned} \sum_{x \in \Omega} \mu(x) \mathbb{P}_x(A) &= \sum_{x \in \Omega} \mu(x) \mathbb{P}_x\left(\bigcap_{s=0}^t \{X_s = x_s\}\right) \\ &= \sum_{x \in \Omega} \mu(x) \mathbf{1}_{\{x\}}(x_0) P(x_0, x_1) \cdots P(x_{t-1}, x_t) \\ &= \mu(x_0) P(x_0, x_1) \cdots P(x_{t-1}, x_t) \\ &= \mathbb{P}_\mu\left(\bigcap_{s=0}^t \{X_s = x_s\}\right) \\ &= \mathbb{P}_\mu(A). \end{aligned}$$

□

Au sujet du vocabulaire, la chaîne de Markov hérite des propriétés de sa matrice de transition : si celle-ci est irréductible ou apériodique, on dira que la chaîne est irréductible ou apériodique. Une mesure de probabilité stationnaire de la matrice de transition sera encore appelée mesure de probabilité stationnaire de la chaîne de Markov.

2.1.1 Exemples

Variables aléatoires iid Il est utile de comparer la forme de la loi des chaînes de Markov à celle de la loi d'une suite de variables aléatoires indépendantes de loi μ :

$$\forall t \in \mathbb{N}, (x_s)_{0 \leq s \leq t} \in \Omega^{t+1}, \quad \mathbb{P}\left(\bigcap_{s=0}^t \{X_s = x_s\}\right) = \prod_{0 \leq s \leq t} \mu(x_s),$$

on voit ainsi, en comparant avec (2.2), que les chaînes de Markov introduisent une première forme de dépendance entre les différentes coordonnées de la suite. On note aussi que cette suite définit une chaîne de Markov (certes peu intéressante du point de vue de ce cours) de mesure initiale et de matrice de transition respectivement données par :

$$\mathbb{P}(X_0 = x) = \mu(x) \text{ et } P(x, y) = \mu(y).$$

2. plusieurs chaînes pourraient être définies sous \mathbb{P} en effet, il faut donc veiller à ce que le contexte soit clair, et savoir de quelle suite de variables aléatoires μ est la loi initiale

Marche aléatoire simple sur un graphe non dirigé Un autre exemple est fourni par une la marche aléatoire (simple) sur un graphe non dirigé³ $G = (V, E)$. L'ensemble de sommets est V et d'ensemble d'arêtes est E (un sous-ensemble des parties de V à deux éléments). Pour alléger les notations on note $x \sim y$ si $\{x, y\} \in E$. On suppose que le graphe G est sans sommet isolé, c'est à dire que pour tout $x \in V$, il existe $y \in V$ tel que $x \sim y$. Notons qu'on peut avoir $x \sim x$: cela signifie que la boucle $\{x, x\}$ appartient à l'ensemble des arêtes. Alors la matrice de transition de la marche aléatoire sur le graphe est définie par :

$$P(x, y) = \frac{\mathbf{1}_{x \sim y}}{\deg(x)} \text{ avec } \deg(x) = \sum_{y \in V} \mathbf{1}_{x \sim y}$$

Alors

$$\mathbb{P} \left(\bigcap_{s=0}^t \{X_s = x_s\} \right) = \mathbb{P}(X_0 = x_0) \frac{1}{\deg(x_0)} \cdot \dots \cdot \frac{1}{\deg(x_{t-1})} \mathbf{1}_{x_0 \sim x_1 \sim \dots \sim x_{t-1} \sim x_t}$$

Marche aléatoire sur Ω Cela concerne la cas où Ω est un groupe additif, de sorte qu'on puisse faire des additions d'éléments de Ω (pour les cas finis, on peut penser à $\mathbb{Z}/n\mathbb{Z} = \{0, 1, \dots, n-1\}$ ou même $(\mathbb{Z}/n\mathbb{Z})^d$; le cas le plus naturel est bien sûr celui de \mathbb{Z}^d , qui n'est pas un graphe fini, mais poser la définition suivante ne pose cependant aucun problème). On pose alors la matrice de transition :

$$\forall x, y \in \Omega, \quad P(x, y) = \eta(y - x), \quad \eta \text{ mesure de probabilité sur } \Omega,$$

Alors

$$\mathbb{P} \left(\bigcap_{s=0}^t \{X_s = x_s\} \right) = \mathbb{P}(X_0 = x_0) \eta(x_1 - x_0) \cdot \dots \cdot \eta(x_t - x_{t-1}) \quad (2.3)$$

La variable aléatoire X_0 étant donnée, on peut construire comme suit cette chaîne à partir d'une suite de variables aléatoires i.i.d. $(Y_t)_{t \in \mathbb{N}}$ de loi η , indépendante de X_0 . On pose pour tout $t \in \mathbb{N}$

$$X_{t+1} = X_t + Y_t.$$

Alors 2.3 vaut en $t = 0$ par hypothèse, et si elle vaut en t , on la montre comme suit en $t + 1$:

$$\begin{aligned} \mathbb{P} \left(\bigcap_{s=0}^{t+1} \{X_s = x_s\} \right) &= \mathbb{P} \left(\bigcap_{s=0}^t \{X_s = x_s\} \cap \{X_t + Y_t = x_{t+1}\} \right) && \text{par définition de } X_{t+1} \\ &= \mathbb{P} \left(\bigcap_{s=0}^t \{X_s = x_s\} \cap \{x_t + Y_t = x_{t+1}\} \right) \\ &= \mathbb{P} \left(\bigcap_{s=0}^t \{X_s = x_s\} \right) \mathbb{P}(Y_t = x_{t+1} - x_t) && \text{par indépendance} \\ &= \mathbb{P}(X_0 = x_0) \eta(x_1 - x_0) \cdot \dots \cdot \eta(x_t - x_{t-1}) \eta(x_{t+1} - x_t) && \text{par la récurrence} \\ &= \mathbb{P}(X_0 = x_0) \eta(x_1 - x_0) \cdot \dots \cdot \eta(x_{t+1} - x_t) \end{aligned}$$

3. L'adjectif "non dirigé" a son importance, car toute chaîne de Markov peut être vue comme une marche aléatoire sur un graphe dirigé

2.1.2 Loi marginale de la chaîne

Le lien crucial entre loi de la chaîne et produit matriciel est exprimé dans le lemme suivant, qui donne les lois marginales de la chaîne :

Lemme 2.4 (Loi marginale et produit matriciel). *On a*

$$\forall x, y \in \Omega, \quad \mathbb{P}_x(X_t = y) = P^t(x, y).$$

Plus généralement, pour μ une probabilité sur Ω et $f : \Omega \rightarrow \mathbb{R}$, et $x \in \Omega$,

$$\mathbb{E}_\mu[f(X_t)] = \mu P^t f, \quad \text{et en particulier } \mathbb{E}_x[f(X_t)] = P^t f(x).$$

Démonstration. On pose la convention $x_t = x$. On commence par exprimer l'événement $\{X_t = y\}$ comme une réunion sur des chemins dont on évalue ensuite la probabilité en explicitant la loi de la chaîne de Markov :

$$\begin{aligned} \mathbb{P}_x(X_t = y) &= \mathbb{P}_x \left(\bigcup_{(x_s)_{0 \leq s \leq t-1} \in \Omega^t} \bigcap_{s=1}^t \{X_s = x_s\} \right) \\ &= \sum_{(x_s)_{0 \leq s \leq t-1} \in \Omega^t} \mathbb{P}_x \left(\bigcap_{s=1}^t \{X_s = x_s\} \right) && \text{car la réunion est disjointe} \\ &= \sum_{(x_s)_{0 \leq s \leq t-1} \in \Omega^t} \mathbb{P}(X_0 = x_0) \prod_{s=0}^{t-1} P(x_s, x_{s+1}) \\ &= \sum_{(x_s)_{0 \leq s \leq t-1} \in \Omega^t} \delta_x(x_0) \prod_{s=0}^{t-1} P(x_s, x_{s+1}) \\ &= \sum_{(x_s)_{1 \leq s \leq t-1} \in \Omega^t} P(x, x_1) P(x_1, x_2) \dots P(x_{t-1}, y) \\ &= P^t(x, y) && \text{par définition du produit matriciel} \end{aligned}$$

Il s'ensuit que

$$\begin{aligned} \mu P^t f(x) &= \sum_{x, y \in \Omega} \mu(x) P^t(x, y) f(y) && \text{par définition de } \mu P^t f \\ &= \sum_{x, y \in \Omega} \mu(x) \mathbb{P}_x(X_t = y) f(y) && \text{vu le résultat précédent} \\ &= \sum_{x \in \Omega} \mu(x) \mathbb{E}_x[f(X_t)] \\ &= \mathbb{E}_\mu[f(X_t)]. \end{aligned}$$

Le cas particulier où $\mu = \delta_x$ la mesure de Dirac en x donne la dernière identité de l'énoncé. \square

On reformulera en guise d'exercice la version probabiliste des résultats vus pour les matrices stochastiques, en particulier du théorème de convergence (on connaît plusieurs modes de convergence pour les suites de variables aléatoires, convergence p.s., convergence en probabilité, convergence en loi : à quel type de convergence correspond-il ?).

2.2 La propriété de Markov

Définition 2.5. La suite de variables aléatoires $(X_t)_{t \in \mathbb{N}}$ satisfait la propriété de Markov lorsque pour tout $t \in \mathbb{N}$, pour tout $(x_s)_{0 \leq s \leq t-1} \in \Omega^t$ tel que $\mathbb{P}(X_t = x \cap H_{t-1}) > 0$ où $H_{t-1} = \bigcap_{0 \leq s \leq t-1} \{X_s = x_s\}$, on a

$$\mathbb{P}(X_{t+1} = y \mid X_t = x, H_{t-1}) = \mathbb{P}(X_1 = y \mid X_0 = x). \quad (2.4)$$

Notons que la relation 2.4 peut encore s'écrire :

$$\mathbb{P}(X_{t+1} = y, X_t = x, H_{t-1}) = \mathbb{P}(X_1 = y \mid X_0 = x) \mathbb{P}(X_t = x, H_{t-1}). \quad (2.5)$$

ce qui a l'avantage de ne pas demander à ce que $\mathbb{P}(X_t = x \cap H_{t-1}) > 0$, c'est-à-dire que le conditionnement soit bien défini. Intuitivement, la propriété de Markov énonce une propriété d'indépendance conditionnelle parfois ainsi formulée : "le futur est indépendant du passé conditionnellement au présent".

Remarque 2.6. Pour être précis, c'est la propriété de Markov dite homogène que nous venons de présenter. La propriété de Markov inhomogène autorise de plus une dépendance des transitions en t , dans le sens où le membre de droite de (2.4) se trouve remplacé par : $\mathbb{P}(X_{t+1} = y \mid X_t = x)$.

Notons d'ors et déjà que si (X_0, X_1) est un couple de variables aléatoires, alors

$$P(x, y) := \mathbb{P}(X_1 = y \mid X_0 = x)$$

définit une matrice stochastique. On obtient directement le lien entre chaîne de Markov et propriété de Markov, exprimé dans le théorème suivant :

Théorème 2.7. — *Si une suite de variables aléatoires $(X_t)_{t \in \mathbb{N}}$ satisfait la propriété de Markov, alors c'est une chaîne de Markov de matrice de transition $P(x, y) = \mathbb{P}(X_1 = y \mid X_0 = x)$.*

— *Réciproquement, une chaîne de Markov satisfait la propriété de Markov.*

Seule la mesure initiale est laissée indéterminée dans l'énoncé de la propriété de Markov.

Démonstration. Soit $t \in \mathbb{N}$, $(x_s)_{0 \leq s \leq t} \in \Omega^{t+1}$. Supposons la propriété de Markov vérifiée dans un premier temps.

$$\begin{aligned} \mathbb{P}\left(\bigcap_{s=0}^t \{X_s = x_s\}\right) &= \mathbb{P}\left(X_t = x_t \mid \bigcap_{s=0}^{t-1} \{X_s = x_s\}\right) \mathbb{P}\left(\bigcap_{s=0}^{t-1} \{X_s = x_s\}\right) \\ &= \mathbb{P}(X_1 = x_t \mid X_0 = x_{t-1}) \mathbb{P}\left(\bigcap_{s=0}^{t-1} \{X_s = x_s\}\right) \\ &= \prod_{s=1}^t \mathbb{P}(X_1 = x_s \mid X_0 = x_{s-1}) \cdot \mathbb{P}(X_0 = x_0) \quad \text{par récurrence} \end{aligned}$$

Réciproquement si l'on dispose d'une chaîne de Markov, alors si l'on pose $H_{t-1} = \bigcap_{s=0}^{t-1} \{X_s = x_s\}$ et $x_t = x$ et $x_{t+1} = y$, on a :

$$\mathbb{P}(\{X_{t+1} = y\} \cap \{X_t = x\} \cap H_{t-1}) = \mathbb{P}\left(\bigcap_{s=0}^{t+1} \{X_s = x_s\}\right) = \mu(x_0)P(x_0, x_1) \dots P(x_t, x_{t+1})$$

tandis que

$$\mathbb{P}(\{X_t = x\} \cap H_{t-1}) = \mathbb{P}\left(\bigcap_{s=0}^{t+1} \{X_s = x_s\}\right) = \mu(x_0)P(x_0, x_1) \dots P(x_t, x_{t+1})$$

de sorte que la probabilité conditionnelle vaut, comme attendu,

$$\mathbb{P}(X_{t+1} = y | \{X_t = x\} \cap H_{t-1}) = P(x_t, x_{t+1}) = P(x, y)$$

□

On peut facilement obtenir des énoncés plus généraux de la propriété de Markov, d'abord en étendant le futur de la trajectoire après l'instant $t + 1$: la suite de variables aléatoires $(X_t)_{t \in \mathbb{N}}$ satisfait la propriété de Markov ssi pour tout $t, r \in \mathbb{N}$, et pour tout $(x_s)_{0 \leq s \leq t+r} \in \Omega^{t+r}$

$$\mathbb{P}\left(\bigcap_{s=1}^r \{X_{t+s} = x_{t+s}\} \mid \bigcap_{s=0}^t \{X_s = x_s\}\right) = \mathbb{P}\left(\bigcap_{s=1}^r \{X_s = x_{t+s}\} \mid X_0 = x_t\right) \quad (2.6)$$

Il peut être commode d'utiliser l'opérateur de translation, "shift" en anglais.

Définition 2.8. Pour $s \in \mathbb{N}$, l'opérateur de translation en temps de s unités est défini par :

$$\theta_s : V^{\mathbb{N}} \rightarrow V^{\mathbb{N}}, (X_t)_{t \geq 0} \mapsto (X_{t+s})_{t \geq 0}.$$

On note que $\theta_s \circ \theta_t = \theta_{t+s}$.

Proposition 2.9. On a pour tout $A \in \Omega^{t+1}$, et $B \in \mathcal{F}_\infty$,

$$\mathbb{P}(\theta_t(X) \in B | (X_s)_{0 \leq s \leq t} \in A, X_t = x) = \mathbb{P}(X \in B | X_0 = x)$$

De façon équivalente, on a, pour toutes applications $F : \Omega^{\mathbb{N}} \rightarrow \mathbb{R}$, $G : \Omega^{t+1} \rightarrow \mathbb{R}$ mesurables bornées :

$$\mathbb{E}[F(\theta_t(X))G((X_s)_{0 \leq s \leq t})\mathbf{1}_{\{X_t=x\}}] = \mathbb{E}_x[F(X)] \mathbb{E}[G((X_s)_{0 \leq s \leq t})\mathbf{1}_{\{X_t=x\}}] \quad (2.7)$$

En toutes lettres, la loi de $\theta_t(X)$ conditionnellement à $\{X_t = x\}$ et à un autre événement quelconque de \mathcal{F}_t , ici l'événement $(X_s)_{0 \leq s \leq t} \in A$, coïncide avec la loi de X conditionnellement à $\{X_0 = x\}$, c'est-à-dire avec la loi de X sous \mathbb{P}_x . Le cas suivant nous sera en particulier utile : si $(X_t)_{t \in \mathbb{N}}$ est une chaîne de Markov de matrice de transition P , et si $A \in \Omega^{t+1}$, on a :

$$\mathbb{P}(X_{t+1} = y | (X_s)_{0 \leq s \leq t} \in A, X_t = x) = P(x, y) \quad (2.8)$$

Exemple 2.10. Présentons à titre d'exemple, le calcul, pour $t_1 < t_2$ et $f_1, f_2 : \Omega \rightarrow \mathbb{R}$, de la valeur de $\mathbb{E}_x[f_1(X_{t_1})f_2(X_{t_2})]$ en fonction des éléments caractéristiques de la chaîne, à savoir P et μ :

$$\begin{aligned}
\mathbb{E}_\mu[f_1(X_{t_1})f_2(X_{t_2})] &= \sum_{x_1, x_2} \mathbb{P}_\mu(X_{t_1} = x_1, X_{t_2} = x_2) f_1(x_1) f_2(x_2) \\
&= \sum_{x_1, x_2} \mathbb{P}_\mu(X_{t_1} = x_1, (\theta_{t_1} X)_{t_2-t_1} = x_2) f_1(x_1) f_2(x_2) \\
&= \sum_{x_1, x_2} \mathbb{P}_\mu(X_{t_1} = x_1) \mathbb{P}_{x_1}(X_{t_2-t_1} = x_2) f_1(x_1) f_2(x_2) \\
&= \sum_{x, x_1, x_2} \mu(x) \cdot P^{t_1}(x, x_1) P^{t_2-t_1}(x_2) f_1(x_1) f_2(x_2) \\
&= \sum_{x, x_1} \mu(x) P^{t_1}(x, x_1) f_1(x_1) \sum_{x_2} P^{t_2-t_1}(x_2) f_2(x_2) \\
&= \sum_{x, x_1} \mu(x) P^{t_1}(x, x_1) f_1(x_1) P^{t_2-t_1} f_2(x_1) \\
&= \sum_{x, x_1} \mu(x) P^{t_1}(x, x_1) f_1(x_1) P^{t_2-t_1} f_2(x_1) \\
&= \mu P^{t_1} (f_1 \cdot P^{t_2-t_1} f_2)
\end{aligned}$$

Le produit \cdot désigne le produit usuel des fonctions, c'est-à-dire ici le produit terme à terme des vecteurs colonnes. Le lecteur pourra s'amuser à titre d'exercice à calculer, pour $t_1 < t_2 < t_3$ et $f_1, f_2, f_3 : \Omega \rightarrow \mathbb{R}$, la valeur de $\mathbb{E}_\mu[f_1(X_{t_1})f_2(X_{t_2})f_3(X_{t_3})]$.

2.3 Représentation de la mesure stationnaire par des temps d'arrêt

Pour l'instant nous avons donné un cadre qui permet de reformuler/d'interpréter les résultats sur les matrices stochastiques en terme de suite de variables aléatoires appelées chaînes de Markov. Si l'on s'arrêtait là cependant, l'intérêt serait minime. On va maintenant voir des énoncés de saveur probabiliste sans contrepartie immédiate dans le monde des matrices. On va notamment obtenir un lien entre des temps d'arrêt particuliers, les temps de retour, et les mesures stationnaires.

Définition 2.11. On appelle temps d'arrêt une variable aléatoire τ à valeurs dans \mathbb{N} telle que

$$\forall t \in \mathbb{N}, \{\tau \leq t\} \in \mathcal{F}_t \quad (2.9)$$

Noter que notre cadre où le temps est discret, il est équivalent de demander que $\forall t \in \mathbb{N}, \{\tau = t\} \in \mathcal{F}_t$, ou encore $\{\tau > t\} \in \mathcal{F}_t$. Pratiquement, $\{\tau \leq t\} \in \mathcal{F}_t$ signifie que, pour tout $t \in \mathbb{N}$, $\mathbf{1}_{\{\tau \leq t\}}$ est une fonction mesurable de $(X_s, 0 \leq s \leq t)$. On peut exprimer cela sans parler de tribu engendrée : cela signifie encore que, pour tout $t \in \mathbb{N}$, il existe $A \in \Omega^{t+1}$ tel que :

$$\{\tau \leq t\} = \{(X_s)_{0 \leq s \leq t} \in A\}$$

Définition 2.12. On considère une chaîne de Markov $(X_t)_{t \in \mathbb{N}}$, un temps d'arrêt τ et deux sommets $a, x \in V$. On appelle *fonction de Green* la fonction définie de la manière suivante

$$G_\tau(a, x) = \sum_{t=0}^{+\infty} \mathbb{P}_a(X_t = x, \tau > t)$$

On notera l'inégalité stricte dans $\{\tau > t\}$. Par Fubini positif, on a l'identité

$$G_\tau(a, x) = \mathbb{E}_a \left[\sum_{t=0}^{+\infty} \mathbf{1}_{\{X_t = x, \tau > t\}} \right],$$

c'est-à-dire que la fonction de Green mesure l'espérance du temps passé en x *strictement* avant τ partant de a .

Théorème 2.13 (Représentation de Aldous-Fill). *Soit $(X_t)_{t \in \mathbb{N}}$ une chaîne de Markov irréductible, et τ un temps d'arrêt qui vérifie*

$$\mathbb{P}_a(X_\tau = a, 0 < \tau < \infty) = 1, \tag{2.10}$$

et soit π l'unique distribution stationnaire de la chaîne de Markov. On a alors

$$\forall x \in \Omega, \frac{G_\tau(a, x)}{\mathbb{E}_a(\tau)} = \pi(x). \tag{2.11}$$

Nous verrons qu'un exemple de tel temps d'arrêt τ est le temps de retour en a dont la définition est donnée en 2.14. Cependant, comme la preuve n'est pas plus difficile dans le cadre un peu plus général du théorème ci-dessus, aussi nous préférons démontrer cette version.

Démonstration du théorème. Soit $x, y \in \Omega$. On note tout d'abord que

$$\begin{aligned} \mathbb{P}_a(X_t = x, X_{t+1} = y, \tau > t) &= \mathbb{E}_a(\mathbf{1}_{X_t = x, \tau > t} \mathbf{1}_{X_{t+1} = y}) \\ &= \mathbb{P}_a(X_t = x, \tau > t) P(x, y) \end{aligned}$$

découle de (2.8) avec $A = \{\tau > t\} \in \sigma\{(X_s)_{0 \leq s \leq t}\}$ par définition d'un temps d'arrêt (c'est ici qu'on utilise cette propriété de façon cruciale). Soit maintenant $y \in \Omega$. C'est un calcul, long

mais transparent, qui donne le résultat.

$$\begin{aligned}
\sum_{x \in \Omega} G_\tau(a, x)P(x, y) &= \sum_{x \in \Omega} \sum_{t=0}^{+\infty} \mathbb{P}_a(X_t = x, \tau > t)P(x, y) \\
&= \sum_{t=0}^{+\infty} \sum_{x \in \Omega} \mathbb{P}_a(X_t = x, \tau > t)P(x, y) \text{ par Fubini positif} \\
&= \sum_{t=0}^{+\infty} \sum_{x \in \Omega} \mathbb{P}_a(X_t = x, X_{t+1} = y, \tau > t) \text{ de la propriété de Markov} \\
&= \sum_{t=0}^{+\infty} \mathbb{P}_a(X_{t+1} = y, \tau > t) \\
&= \sum_{t=1}^{+\infty} \mathbb{P}_a(X_t = y, \tau \geq t) \\
&= \sum_{t=1}^{+\infty} \mathbb{P}_a(X_t = y, \tau > t) + \sum_{t=1}^{+\infty} \mathbb{P}_a(X_t = y, \tau = t) \\
&= \sum_{t=0}^{+\infty} \mathbb{P}_a(X_t = y, \tau > t) - \mathbb{P}_a(X_0 = y, \tau > 0) + \sum_{t=1}^{+\infty} \mathbb{P}_a(X_\tau = y) \\
&= \sum_{t=0}^{+\infty} \mathbb{P}_a(X_t = y, \tau > t) - \mathbf{1}_{\{y=a\}} + \mathbf{1}_{\{y=a\}} \text{ de l'hypothèse sur } \tau \\
&= G_\tau(a, y)
\end{aligned}$$

On calcule la normalisation nécessaire pour obtenir une mesure de probabilité.

$$\sum_{y \in \Omega} G_\tau(a, y) = \sum_{y \in \Omega} \sum_{t=0}^{+\infty} \mathbb{P}_a(X_t = y, \tau > t) = \sum_{t=0}^{+\infty} \mathbb{P}_a(\tau > t) = \mathbb{E}_a[\sum_{t=0}^{+\infty} \mathbf{1}_{\tau > t}] = \mathbb{E}_a[\tau]$$

Ainsi

$$\sum_{x \in \Omega} \frac{G_\tau(a, x)}{\mathbb{E}_a[\tau]} P(x, y) = \frac{G_\tau(a, y)}{\mathbb{E}_a[\tau]},$$

et l'unicité de la mesure de probabilité stationnaire, qui découle d'après la proposition 1.13 de l'hypothèse d'irréductibilité, permet de déduire (2.11). \square

Voici les deux exemples canoniques de temps d'arrêt (vérifier qu'il s'agit effectivement de tels temps) :

Définition 2.14. Soit $A \subset \Omega$, et $x \in \Omega$. Le temps d'atteinte de A et le temps de retour en A sont respectivement définis par

$$\tau_A = \min\{t \geq 0, X_t \in A\} \text{ et } \tau_A^+ = \min\{t \geq 1, X_t \in A\},$$

et on écrira simplement $\tau_x := \tau_{\{x\}}$ et $\tau_x^+ := \tau_{\{x\}}^+$ dans le cas de singletons.

Nous énonçons dès maintenant le corollaire du théorème précédent.

2.3. REPRÉSENTATION DE LA MESURE STATIONNAIRE PAR DES TEMPS D'ARRÊT 27

Corollaire 2.15. Soit $(X_t)_{t \in \mathbb{N}}$ une chaîne de Markov irréductible, soit $a \in \Omega$, et τ_a^+ le premier temps de retour en a , $\tau_a^+ = \min\{t \geq 1, X_t = a\}$. Alors :

$$\pi(a) = \frac{1}{\mathbb{E}_a[\tau_a^+]}.$$

Ce résultat est d'autant plus remarquable qu'il n'existe pas de façon simple d'obtenir $\mathbb{E}_a[\tau_b]$ pour $a \neq b$ en général (on verra plus tard une approche dans le cas des graphes réversibles et transitifs). On notera aussi que cette représentation est purement probabiliste (essayer d'exprimer la quantité $\mathbb{E}_a[\tau_a^+]$ à l'aide du semigroupe P^t pour s'en persuader).

Démonstration du corollaire. On applique le théorème en observant que la condition de finitude des temps d'atteinte (2.10) découle de la Proposition 2.17 à venir. Ensuite, on fait le calcul :

$$G_{\tau_a^+}(a, x) = \mathbb{E}_a \left[\sum_{t=0}^{+\infty} \mathbf{1}_{\{X_t = x, \tau_a^+ > t\}} \right] = \mathbb{E}_a[\mathbf{1}_{\tau_a^+ < \infty}] = \mathbb{P}(\tau_a^+ < \infty) = 1.$$

□

Cette représentation de π au moyen de τ^+ est des plus satisfaisantes. Elle permet notamment d'avoir accès au calcul de $\mathbb{E}_a[\tau_a^+]$ lorsque la mesure stationnaire est connue ; cependant, pour calculer $\mathbb{E}_a[\tau_b^+]$ dans le cas où $a \neq b$, il va nous falloir développer une autre stratégie. De façon peut-être surprenante, c'est en mobilisant nos connaissances en électricité que nous allons pouvoir répondre à cette question dans le chapitre suivant.

Pour étudier les temps de retour et d'atteinte, on note le lemme suivant au sujet de l'espérance de variables aléatoires à valeurs entières.

Lemme 2.16. Si N est une variable aléatoire à valeurs dans $\mathbb{N} \cup \{\infty\}$ p.s. (i.e. $\mathbb{P}(X \in \mathbb{N} \cup \{\infty\}) = 1$), alors :

$$\mathbb{E}(N) = \sum_{t=1}^{\infty} \mathbb{P}(N \geq t) = \sum_{t=0}^{\infty} \mathbb{P}(N > t)$$

Démonstration. Pour tout $n \in \mathbb{N}$, on peut écrire $n = \sum_{t=1}^{\infty} \mathbf{1}_{n \geq t} = \sum_{t=0}^{\infty} \mathbf{1}_{n > t}$. Il suffit alors de prendre l'espérance des deux membres. □

On note aussi que les temps d'atteinte sont presque sûrement finis.

Proposition 2.17. Si la matrice de transition P de la chaîne de Markov $(X_t)_{t \geq 0}$ est irréductible sur Ω , alors quelque soit $x, y \in \Omega$

$$\mathbb{E}_x[\tau_y] \leq \mathbb{E}_x[\tau_y^+] < \infty$$

Démonstration. On commence par montrer la propriété sur les temps d'atteinte. La propriété avec les temps de retour en découlera ensuite. Notons $t(x, y)$ l'entier tel que $P^t(x, y) > 0$, dont l'existence nous est assurée par la définition de l'irréductibilité. On fixe alors y et on choisit $t = \max_x t(x, y)$ de sorte que

$$\mathbb{P}_x(\tau_y \leq t) = \mathbb{P}_x\left(\bigcup_{s=0}^t \{X_s = y\}\right) \geq \mathbb{P}_x(X_{t(x,y)} = y)$$

Prenant le minimum sur x , on obtient :

$$\mathbb{P}_x(\tau_y \leq t) \geq \min_x \mathbb{P}_x(X_{t(x,y)} = y) =: \delta > 0$$

Toujours à y fixé, on montre par récurrence sur l'entier $k \in \mathbb{N}$ la propriété :

$$\forall x \in \Omega, \quad \mathbb{P}_x(\tau_y > kt) \leq (1 - \delta)^k.$$

Pour $k = 0$ et $k = 1$, on l'a déjà vérifiée ; puis si la propriété vaut en k , alors on peut écrire, en notant $\tau_y \circ \theta_t = (\tau_y \circ \theta_t)(X) := \tau_y(\theta_t(X))$ le temps d'atteinte de la chaîne shiftée :

$$\begin{aligned} \mathbb{P}_x(\tau_y > (k+1)t) &= \mathbb{P}_x(\{\tau_y \circ \theta_t > kt\} \cap \{\tau_y > t\}) \\ &= \mathbb{E}_x(\mathbf{1}_{\{\tau_y \circ \theta_t > kt\}} \mathbf{1}_{\{\tau_y > t\}}) \\ &= \mathbb{E}_x[\mathbb{E}_{X_t}[\mathbf{1}_{\tau_y > kt}] \mathbf{1}_{\tau_y > t}] \\ &\leq (1 - \delta)^k \mathbb{E}_x[\mathbf{1}_{\tau_y > t}] \\ &\leq (1 - \delta)^{k+1} \end{aligned}$$

Mais alors, puisque $\tau_y \in \mathbb{N}$, appliquant le lemme 2.16 on obtient

$$\mathbb{E}_x[\tau_y] = \sum_{k \in \mathbb{N}} \mathbb{P}_x(\tau_y > k) \leq \sum_{k \in \mathbb{N}} t \mathbb{P}_x(\tau_y > kt) \leq t \sum_{k \in \mathbb{N}} (1 - \delta)^k = \frac{t}{\delta} < \infty$$

conclut la preuve. Pour ce qui est des temps de retour on note que si $x \neq y$, alors $\mathbb{P}_x(\tau_y^+ = \tau_y) = 1$ d'où $\mathbb{E}_x[\tau_y^+] = \mathbb{E}_x[\tau_y] < \infty$ dans ce cas. Ensuite, pour le cas restant, on a, à l'aide de la proposition 2.18 ci-dessous, que :

$$\mathbb{E}_y[\tau_y^+] = 1 + \sum_x P(y, x) \mathbb{E}_x[\tau_y] \leq 1 + \max_x \mathbb{E}_x[\tau_y] < \infty$$

□

Proposition 2.18. *Soit $A \subset \Omega$. Le temps de retour τ_A^+ en A d'une chaîne de Markov $(X_t)_{t \in \mathbb{N}}$ de matrice de transition $P = (P(x, y))_{x, y \in \Omega}$ satisfait :*

$$\mathbb{E}_x[\tau_A^+] = 1 + \sum_y P(x, y) \mathbb{E}_y[\tau_A].$$

Notons que la proposition est vraie sans supposer l'irréductibilité, mais il est alors possible que les deux termes soient simultanément infinis.

Démonstration. On note que par définition des temps de retour

$$\{\tau_A^+ \geq t\} = \{\tau_A \circ \theta_1 \geq t - 1\}$$

avec θ_1 le shift en temps de 1 unité de temps, puis on écrit, pour $t \geq 2$:

$$\begin{aligned} \mathbb{P}_x(\tau_A^+ \geq t) &= \sum_{y \in \Omega} \mathbb{P}_x(\tau_A \circ \theta_1 \geq t - 1, X_1 = y) \\ &= \sum_y P(x, y) \mathbb{P}_y(\tau_A \geq t - 1). \end{aligned}$$

et on note que l'identité vaut encore pour $t = 1$. Ainsi,

$$\begin{aligned}
 \mathbb{E}_x[\tau_z^+] &= \sum_{t=1}^{\infty} \mathbb{P}_x(\tau_A^+ \geq t) \\
 &= \sum_{t=1}^{\infty} \sum_y P(x, y) \mathbb{P}_y(\tau_A \geq t-1) \\
 &= \sum_y P(x, y) \sum_{t=1}^{\infty} \mathbb{P}_y(\tau_A \geq t-1) \\
 &= \sum_y P(x, y) \left(1 + \sum_{t=1}^{\infty} \mathbb{P}_y(\tau_A \geq t)\right) \\
 &= 1 + \sum_y P(x, y) \mathbb{E}_y[\tau_A].
 \end{aligned}$$

□

Notons aussi la proposition suivante, dans la même veine :

Proposition 2.19. *Soit $A \subset \Omega$. Le temps de retour τ_A^+ en A d'une chaîne de Markov $(X_t)_{t \in \mathbb{N}}$ de matrice de transition $P = (P(x, y))_{x, y \in \Omega}$ satisfait, pour tout $x, z \in \Omega$,*

$$\mathbb{P}_x(X_{\tau_A^+} = z) = \sum_y P(x, y) \mathbb{P}_y(X_{\tau_A} = z).$$

En particulier, pour toute fonction $f : \Omega \rightarrow \mathbb{R}$, on a :

$$\mathbb{E}_x[f(X_{\tau_A^+})] = \sum_y P(x, y) \mathbb{E}_y[f(X_{\tau_A})].$$

Démonstration. En effet, de la définition de τ_A^+ et τ_A , on a l'égalité des deux événements :

$$\{X_{\tau_A^+} = z\} = \{(X \circ \theta)_{\tau_A(X \circ \theta)} = z\}$$

puis il en découle que, si l'on pose $F(X) = \mathbf{1}_{\{X_{\tau_A} = z\}}$, alors

$$\begin{aligned}
 \mathbb{P}_x(X_{\tau_A^+} = z) &= \mathbb{P}_x((X \circ \theta)_{\tau_A(X \circ \theta)} = z) \\
 &= \sum_y \mathbb{E}_x(F(\theta \circ X) = z, X_1 = y) \\
 &= \sum_y P(x, y) \mathbb{E}_y(F(X) = z) \\
 &= \sum_y P(x, y) \mathbb{P}_y(X_{\tau_A(X)} = z),
 \end{aligned}$$

et l'égalité avec les espérances découle de l'expression générale $\mathbb{E}[f(Y)] = \sum_{y \in \Omega} \mathbb{P}(Y = y) f(y)$.

□

2.4 L'exemple de la ruine du joueur

Il s'agit de la chaîne de Markov $(X_t)_{t \in \mathbb{N}}$ sur $\Omega = \{0, \dots, n\}$ de matrice de transition

$$P(i, j) = \begin{cases} \frac{1}{2} \mathbf{1}_{\{|j-i|=1\}} & \text{si } i \in \{1, \dots, n-1\} \\ \mathbf{1}_{i=j} & \text{si } i \in \{0, n\}. \end{cases}$$

C'est la marche aléatoire des gains d'un joueur qui joue à un jeu équilibré, gagne ou perd 1 à chaque tour de jeu et s'arrête lorsqu'il atteint un gain de n ou lorsqu'il atteint 0 et n'a plus d'argent à parier. On s'intéresse au temps aléatoire $\tau = \min\{t \geq 0, X_t \in \{0, n\}\} \in \mathbb{N} \cup \{\infty\}$ qui est le premier temps où le joueur atteint la fortune 0 (il a perdu) ou n (il a gagné); est-il fini p.s., ou le joueur peut-il jouer indéfiniment? S'il ce temps aléatoire est fini, admet-il dans ce cas une espérance? La variable aléatoire X_τ (qui est définie sur l'événement $\{\tau < \infty\}$) rend compte de l'événement de gain ou de perte finale. Quelle est sa loi?

La méthode présentée ci dessous, dite *méthode à un pas*, est une méthode récursive sur la position initiale de la chaîne. L'opérateur de shift $\theta = \theta_1$ défini par $(\theta \circ X)_t = X_{t+1}$, peut être composé avec $\tau = \tau(X)$ pour donner

$$\tau \circ \theta = \tau(\theta(X)) = \inf\{t \geq 0, X_{t+1} \in \{0, n\}\} \in \mathbb{N} \cup \{\infty\}.$$

On commence par appliquer le lemme 2.18, avec $\tau = \tau_A$ pour $A = \{0, n\}$, cela donne :

$$\begin{aligned} h(k) &= \mathbb{E}_k[\tau] \\ &= 1 + P(k, k+1)\mathbb{E}_{k+1}[\tau] + P(k, k-1)\mathbb{E}_{k-1}[\tau] \\ &= 1 + \frac{1}{2}(\mathbb{E}_{k+1}[\tau] + \mathbb{E}_{k-1}[\tau]) \\ &= 1 + \frac{h(k+1) + h(k-1)}{2} \end{aligned}$$

de conditions "initiales" $h(0) = h(n) = 0$ puisque, par définition, $\tau = 0$ p.s. sous \mathbb{P}_0 et \mathbb{P}_n . (Noter que l'on a pas, comme d'habitude $h(0)$ et $h(1)$). On pose, pour $k \in \{1, \dots, n-1\}$, $\delta(k) = h(k+1) - h(k)$. Alors la récurrence double précédente devient une simple récurrence simple :

$$\delta(k) = \delta(k-1) - 2.$$

qui se résout comme suit : $\delta(k) = \delta(0) - 2k$. Maintenant la somme des $\delta(k)$ est nulle, donc

$$0 = \sum_{k=0}^{n-1} \delta(k) = n\delta(0) - n(n-1),$$

donne $\delta(0) = n-1$ et $\delta(k) = n-1-2k$. On en déduit que

$$h(k) = h(k) - h(0) = \sum_{j=0}^{k-1} \delta(j) = (n-1)k - (k-1)k = (n-k)k.$$

Notons que les quantités k et $n-k$ jouent bien un rôle symétrique dans cette expression comme attendu. Si $k = \alpha n$ est un entier, on obtient donc $h(\alpha n) = \alpha(1-\alpha)n^2$ dont la maximum est atteint en $\alpha = 1/2$. Il est intéressant de noter la dépendance quadratique en n .

Pour calculer $\mathbb{P}_k(X_\tau = n | X_1 = k + 1)$, on applique cette fois le lemme 2.19, on en tire à nouveau une équation de récurrence double, plus simple que la précédente cette fois. Pour $k \in \Omega \setminus \{0, n\}$

$$\begin{aligned} g(k) &= \mathbb{P}_k(X_\tau = n) \\ &= \mathbb{P}_k(X_\tau^+ = n) && \text{car } k \in \Omega \setminus \{0, n\} \\ &= P(k, k + 1)\mathbb{P}_{k+1}(X_\tau = n) + P(k, k - 1)\mathbb{P}_{k-1}(X_\tau = n) \\ &= \frac{g(k + 1) + g(k - 1)}{2} \end{aligned}$$

avec les conditions "initiales" $g(0) = 0$ et $g(n) = 1$. On peut résoudre cette équation par la méthode classique (le polynôme caractéristique $X^2 - 2 + 1 = (X - 1)^2$ admet une racine double égale à 1 d'où la base de solution $(1)_{k \in \mathbb{N}}$ et $(k)_{k \in \mathbb{N}}$...). On peut aussi dans ce cas précis très simple noter que les différences satisfont simplement :

$$g(k + 1) - g(k) = g(k) - g(k - 1),$$

Ces différences étant constantes, la fonction g est la fonction affine $g(k) = \frac{k}{n}$.

Quelques commentaires sur les spécificités de cette chaîne de Markov sont nécessaires : Noter que la présence d'états absorbants (au sens de la définition 1.14) empêche la chaîne d'être irréductible; ici, la chaîne compte deux états absorbants. On n'a pas unicité de la mesure de probabilité stationnaire, et toute combinaison linéaire des masses de Dirac en les deux états absorbants $p\delta_0 + (1 - p)\delta_n$ est une mesure de probabilité stationnaire (soit encore tout vecteur ligne π avec $\pi(0) = p$, et $\pi(n) = 1 - p$). On a aussi $\lim_{t \rightarrow \infty} P^t(k, n) = \lim_{t \rightarrow \infty} \mathbb{P}(X_t = n) = \lim_{t \rightarrow \infty} \mathbb{P}(\bigcup_{t \in \mathbb{N}} \{X_t = n\}) = \mathbb{P}(X_\tau = n)$ on a :

$$\lim_{t \rightarrow \infty} P^t(k, n) = \frac{k}{n}, \quad \text{et} \quad \lim_{t \rightarrow \infty} P^t(k, 0) = \frac{n - k}{n},$$

ce qui implique aussi, pour tout $k' \in \Omega \setminus \{0, n\}$, $\lim_{t \rightarrow \infty} P^t(k, k') = 0$. En particulier,

$$\lim_{t \rightarrow \infty} P^t(k, n) + P^t(k, 0) = 1.$$

Le résultat sur l'espérance de τ permet en fait de préciser la vitesse de convergence dans la limite précédente, puisque

$$\begin{aligned} \mathbb{E}_k[\tau] &= \sum_{t=1}^{\infty} \mathbb{P}_k(\tau \geq t) \\ &= \sum_{t=1}^{\infty} (1 - \mathbb{P}_k(\tau < t)) \\ &= \sum_{t=1}^{\infty} (1 - \mathbb{P}_k(X_{t-1} \in \{0, n\})) \\ &= \sum_{t=0}^{\infty} (1 - \mathbb{P}_k(X_t \in \{0, n\})) \\ &= \sum_{t=0}^{\infty} (1 - P^t(k, n) - P^t(0, n)) \end{aligned}$$

2.5 Quelques mots sur la propriété de Markov forte

Cette section est un supplément au cours, et n'est pas exigible : en pratique, lorsqu'on voudra appliquer la propriété de Markov en un temps aléatoire τ qui est un temps d'arrêt, on décomposera simplement suivant la valeur de ce temps d'arrêt, c'est à dire selon la réunion d'événements disjoints $\bigcup_{t \in \mathbb{N}} \{\tau = t\}$, et on se ramènera à la propriété de Markov simple.

On peut décréter que l'espace de probabilité (E, \mathcal{F}) est l'espace $\Omega^{\mathbb{N}}$ lui-même (ce qu'on appelle le choix canonique); alors $\omega \rightarrow X_t(\omega)$ dans (2.1) est simplement l'application coordonnée : $X_t(\omega) = \omega_t$. L'avantage de cette opération est qu'on peut alors faire agir l'opérateur de translation θ directement sur E , et changer légèrement de point de vue sur la propriété de Markov simple : pour $F : E \rightarrow \mathbb{R}$ \mathcal{F}_∞ -mesurable bornée, et $G : E \rightarrow \mathbb{R}$ \mathcal{F}_t -mesurable bornée :

$$\mathbb{E} [F \circ \theta_t \cdot G \mathbf{1}_{\{X_t=x\}}] = \mathbb{E}_x [F] \mathbb{E}[G \mathbf{1}_{\{X_t=x\}}]$$

Pour le moment il n'y a pas vraiment de gain par rapport à (2.7). Si l'on essaie d'écrire une relation similaire à (2.7) pour un temps d'arrêt, on voit que le domaine de définition de G (disons le nombre d'arguments que cette fonction doit prendre) n'est pas bien défini. Pour contourner cet obstacle on pose pour τ un temps d'arrêt, la tribu \mathcal{F}_τ :

$$A \in \mathcal{F}_\tau \quad \text{ssi} \quad \forall t \in \mathbb{N}, \quad A \cap \{\tau \leq t\} \in \mathcal{F}_t$$

On considère ensuite directement des fonctions F et G définies directement sur (E, \mathcal{F}) (et non plus des fonctions F et G définies sur $\Omega^{\mathbb{N}}$ et Ω^{t+1}), ce qui permet d'écrire la propriété suivante, appelée propriété de Markov forte par opposition à la propriété de Markov simple :

Proposition 2.20 (Propriété de Markov forte). *Soit τ un temps d'arrêt presque sûrement fini, et $(X_t)_{t \in \mathbb{N}}$ une chaîne de Markov. Alors pour tout F \mathcal{F}_∞ -mesurable bornée, et G \mathcal{F}_τ -mesurable bornée,*

$$\mathbb{E}[F \circ \theta_\tau \cdot G \mathbf{1}_{X_\tau=x}] = \mathbb{E}_x[F] \mathbb{E}[G \mathbf{1}_{X_\tau=x}]$$

Démonstration. L'idée est de décomposer suivant les valeurs prises par le temps d'arrêt, puis d'utiliser la propriété 2.6, en notant que par définition de la filtration \mathcal{F}_τ , $G \mathbf{1}_{X_t=x} \mathbf{1}_{\tau=t}$ est \mathcal{F}_t mesurable,

$$\begin{aligned} \mathbb{E}[F \circ \theta_\tau \cdot G \mathbf{1}_{X_\tau=x}] &= \sum_{t \in \mathbb{N}} \mathbb{E}[F \circ \theta_\tau \cdot G \mathbf{1}_{X_\tau=x} \mathbf{1}_{\tau=t}] \\ &= \sum_{t \in \mathbb{N}} \mathbb{E}[F \circ \theta_t \cdot G \mathbf{1}_{X_t=x} \mathbf{1}_{\tau=t}] \\ &= \sum_{t \in \mathbb{N}} \mathbb{E}_x[F] \mathbb{E}[G \mathbf{1}_{X_t=x} \mathbf{1}_{\tau=t}] \\ &= \mathbb{E}_x[F] \mathbb{E}[G \mathbf{1}_{X_\tau=x}] \end{aligned}$$

□

Dans les faits cependant, nous n'aurons pas grand besoin de ce point de vue dans le reste du cours.

Chapitre 3

Réseaux électriques (a.k.a. chaînes réversibles)

On étudie dans ce chapitre les temps d'atteinte de chaînes de Markov réversibles au moyen de calculs de résistance dans des réseaux électriques, les mêmes réseaux que ceux qu'on a pu rencontrer en cours d'électricité. La tension aux sommets du graphe lorsqu'on branche un générateur entre deux sommets quelconque d'une part, et la probabilité d'atteindre un de ces deux sommets avant l'autre sont des quantités étroitement liées : cela découle de la propriété d'unicité du prolongement harmonique d'une fonction. De cette observation fondamentale découlent un certain nombre de propriétés surprenantes, entre autre le calcul de fonctions harmoniques au moyen de la réduction de réseaux électriques.

3.1 Chaînes de Markov réversibles et réseau.

Définition 3.1. Un graphe $G = (V, E)$ est la donnée d'un ensemble V , appelé ensemble de sommets et d'un sous-ensemble E des paires non ordonnées de sommets de V , appelé ensemble d'arêtes¹. On appelle boucle une arête de type $\{x, x\}$. Un graphe sans boucles est un graphe où $E \subset \{\{x, y\} : x, y \in V^2, x \neq y\}$

Si $x, y \in V$, les deux paires $\{x, y\}$ et $\{y, x\}$ étant égales, représentent la même arête. À ensemble de sommets V fixé de cardinal n , le plus gros graphe possible (au sens de l'inclusion des ensembles d'arêtes) est celui où les $\binom{n}{2}$ arêtes sont retenues dans E , on l'appelle le graphe complet.

Définition 3.2. Le graphe induit par G sur $V' \subset V$ est le graphe $G' = (V', E')$ avec $E' = \{e \in E : e = \{x, y\}, (x, y) \in V'^2\}$

Ainsi on ne conserve dans G' que les arêtes dont les deux extrémités sont dans V' . Il peut être utile d'attacher aux arêtes une grandeur scalaire.

Définition 3.3. Un *réseau* $\{G, c\}$ est la donnée un graphe fini $G = (V, E)$ non orienté et connexe et d'une collection de conductances $c = (c(e))_{e \in E} \in (\mathbb{R}^+ \setminus \{0\})^E$, d'inverses $r(e) = 1/c(e)$ appelés résistances.

1. On appelle multigraphe un graphe où les répétitions de mêmes arêtes sont autorisées dans E ; techniquement E est alors un "multiset".

Si $e = \{x, y\} \in E$, on écrira parfois $x \sim y$ pour alléger les notations et on notera indifféremment

$$c(x, y) = c(y, x) := c(e)$$

Définition 3.4. Soit $\{G, c\}$ un réseau. La chaîne de Markov sur V associée à ce réseau de matrice de transition $P = (P(x, y))_{x, y \in V^2}$ donnée par

$$P(x, y) = \frac{c(x, y)}{c(x)} \mathbf{1}_{\{x, y\} \in E} \text{ avec } c(x) = \sum_{y, \{x, y\} \in E} c(x, y)$$

sera simplement appelée marche aléatoire sur le réseau $\{G, c\}$.

Posons

$$c_G := \sum_{x \in V} c(x) = 2 \sum_{e: e=\{x, y\}, x \neq y} c(e) + \sum_{e: e=\{x, x\}} c(e) \quad (3.1)$$

où l'on distingue à la deuxième égalité les arêtes qui forment des boucles ou non : si une arête n'est pas une boucle, elle est comptée deux fois, s'il s'agit d'une boucle, c'est-à-dire d'une arête du type $\{x, x\}$, elle est comptée une seule fois. L'intérêt qu'on porte aux réseaux est justifié par la proposition suivante.

Proposition 3.5. *La marche aléatoire sur le réseau $\{G, c\}$ est réversible par rapport à la mesure de probabilité π donnée par*

$$\forall x \in V, \pi(x) = \frac{c(x)}{c_G}.$$

Réciproquement, à toute chaîne de Markov irréductible et réversible $(X_t)_{t \in \mathbb{N}}$ sur Ω , on peut associer un réseau (unique à isomorphisme près, où si l'on impose $V = \Omega$) tel que que la chaîne $(X_t)_{t \in \mathbb{N}}$ soit la marche aléatoire sur le réseau.

En conséquence, puisque π est alors l'unique mesure stationnaire (Proposition 1.23), le calcul de la mesure stationnaire sur un réseau ne pose aucune difficulté.

Démonstration. Le sens direct est aisé, il suffit de montrer que P est réversible par rapport à π .

$$\pi(x)P(x, y) = \frac{c(x)}{c_G} \frac{c(x, y)}{c(x)} = \frac{c(x, y)}{c_G} = \frac{c(y, x)}{c_G} = \frac{c(y)}{c_G} \frac{c(y, x)}{c(y)} = \pi(y)P(y, x).$$

Quant au sens opposé, il suffit de poser, si π est la mesure par rapport à laquelle X est réversible, $c(\{x, y\}) := \pi(x)P(x, y)$ dès lors que cette quantité est non nulle (formellement, on pose pour définir le graphe $V = \Omega$ et $E = \{\{x, y\} \in V^2, \pi(x)P(x, y) \neq 0\}$). La définition de $c(\{x, y\})$ n'est pas ambiguë du fait de la propriété de réversibilité. Le graphe est bien connexe par irréductibilité. En outre, le calcul $c(x) = \sum_{y: \{x, y\} \in E} c(\{x, y\}) = \sum_{y: \{x, y\} \in E} \pi(x)P(x, y) = \pi(x)$ assure que, pour tout $x, y \in \Omega$

$$\frac{c(x, y)}{c(x)} = \frac{\pi(x)P(x, y)}{c(x)} = P(x, y).$$

□

On va maintenant approfondir notre étude des fonctions harmoniques, introduites à la définition 1.10. La proposition suivante généralise la proposition 1.11, selon laquelle une fonction harmonique sur l'espace entier pour P stochastique irréductible est constante.

3.2 Principe de Dirichlet pour les fonctions harmoniques

Proposition 3.6 (Principe du maximum). *Soit P stochastique irréductible, et $B \subset V$. Si h est harmonique sur $V \setminus B$ alors h atteint son maximum en un point de B .*

Démonstration. Soit x_0 tel que $h(x_0) = \max_{y \in V} h(y)$. Si $x_0 \in B$ il n'y a rien à prouver. Sinon, si $x_0 \notin B$, soit $b \in B$. On peut trouver une suite finie $(x_i)_{1 \leq i \leq r}$ tel que $P(x_i, x_{i+1}) > 0$ pour tout $0 \leq i \leq r-1$ et $x_r = b$. On note s le plus petit entier tel que $x_s \in B$. On montre alors que $h(x_i) = h(x_0)$ pour tout $i \leq s$ par récurrence (finie) sur l'entier i . C'est vrai en $i = 0$. Si $h(x_i) = h(x_0)$ et $i < s$ alors, puisque $x_i \notin B$ par définition, h est encore harmonique en x_i , et $h(x_i) = \sum P(x_i, y)h(y)$ implique $h(y) = h(x_i)$ pour tout y tel que $P(x_i, y) > 0$, en particulier pour $y = x_{i+1}$. Ainsi $h(x_s) = h(x_0)$ et $x_s \in B$ est un élément en lequel h atteint son maximum. \square

Une fonction définie sur un sous-ensemble arbitraire de sommets s'étend de façon unique en une fonction harmonique sur le complémentaire de cet ensemble. On a une représentation probabiliste simple de cette extension à l'aide des temps d'atteinte.

Proposition 3.7 (Principe de Dirichlet). *Soit P stochastique irréductible. Soient $B \subset V$ et $h_B : B \rightarrow \mathbb{R}$ une fonction définie sur B . Alors, la fonction $h : V \rightarrow \mathbb{R}$ définie par*

$$h(x) = \mathbb{E}_x[h_B(X_{\tau_B})]$$

est l'unique extension de h_B telle que $h(x) = h_B(x)$ pour tout $x \in B$ et h est harmonique pour P sur $V \setminus B$.

On appelle la fonction h de l'énoncé précédent l'extension harmonique de h_B .

Démonstration. Vérifions d'abord que h donnée dans l'intitulé est bien une extension. Si $x \in B$, alors $\tau_B = 0$ et donc $h(x) = \mathbb{E}_x[h_B(X_{\tau_B})] = h_B(x)$. Cette extension est de plus harmonique. Soit $x \notin B$. Notons que si

$$f(X) = X_{\tau_B^+} \quad \text{alors} \quad f(X \circ \theta) = (X \circ \theta)_{\tau_B(X \circ \theta)}$$

Et

$$\{X_{\tau_B^+} = y\} = \{(X \circ \theta)_{\tau_B(X \circ \theta)} = y\}$$

Alors, de la propriété de Markov,

$$h(x) = \mathbb{E}_x[h_B(X_{\tau_B})] = \mathbb{E}_x[h_B(X_{\tau_B^+})] = \mathbb{E}_x[h_B \circ f(X)] = \mathbb{E}_x[h_B \circ f(X \circ \theta)] = \sum_{y \in V} P(x, y) \mathbb{E}_y[h(X_{\tau_B}) | X_1 = y].$$

Comme $x \notin B$, $\mathbb{P}_x(\tau_B \geq 1) = 1$, et on peut écrire :

$$\begin{aligned}
\mathbb{E}_x[h(X_{\tau_B}) \mid X_1 = y] &= \sum_{t \geq 1} \mathbb{E}_x[h(X_t) \mathbf{1}_{\{\tau_B=t\}} \mid X_1 = y] \\
&= \sum_{t \geq 1} \mathbb{E}_x[h((\theta \circ X)_{t-1}) \mathbf{1}_{\{\tau_B \circ \theta = t-1\}} \mid X_1 = y] \\
&= \sum_{t \geq 1} \mathbb{E}_y[h(X_{t-1}) \mathbf{1}_{\{\tau_B=t-1\}}] \\
&= \sum_{t \geq 0} \mathbb{E}_y[h(X_t) \mathbf{1}_{\{\tau_B=t\}}] \\
&= \mathbb{E}_y[h(X_{\tau_B})] \\
&= h(y)
\end{aligned}$$

Ce calcul entraîne que h est harmonique en x puisque :

$$h(x) = \sum_{y \in V} P(x, y) \mathbb{E}_y[h(X_{\tau_B}) \mid X_1 = y] = \sum_{y \in V} P(x, y) h(y).$$

Montrons maintenant l'unicité. Pour cela considérons $g : V \rightarrow \mathbb{R}$ harmonique sur $V \setminus B$ et nulle sur B . Le principe du maximum appliqué à g assure que $g \leq 0$. Appliqué à $-g$, il assure que $g \geq 0$. Ainsi $g = 0$. Si maintenant h et \tilde{h} sont deux extensions harmoniques de h_B , alors $g = h - \tilde{h}$ est harmonique sur $V \setminus B$, nulle sur B , et donc nulle partout. \square

3.3 Tension, flot et flot courant

La tension est le nom des fonctions harmonique en électricité.

Définition 3.8. Soient $a, z \in V$ deux sommets distingués de V , appelés respectivement *source* et *puits* du réseau. On appelle *tension* sur le réseau $\{G, c\}$ (relativement à a et z) une fonction harmonique sur $V \setminus \{a, z\}$.

D'après le principe de Dirichlet, une tension W est entièrement déterminée par ses valeurs $W(a)$ et $W(z)$ en les points source et puits. Il existe une autre façon, duale, de représenter les fonctions harmoniques à constante additive près : il s'agit de définir une fonction sur les arêtes orientées et non les sommets ; cette fonction proportionnelle aux différences de tension est appelée le courant en électricité. On introduit maintenant cette fonction et on énonce ses deux propriétés caractéristiques.

Définition 3.9. On appelle arête orientée $\vec{e} = \vec{x}y$ est un couple ordonné de sommets. Dans le cas où l'on dispose au préalable d'un ensemble d'arêtes non orientées E , on note $\vec{E} = \{\vec{x}y, \{x, y\} \in E\}$ l'ensemble des arêtes orientées formées à partir des arêtes de E .

Noter qu'à une arête $\{x, y\}$, sont associées deux arêtes orientées $\vec{x}y$ et $\vec{y}x$.

Définition 3.10. Soit $f : \vec{E} \rightarrow \mathbb{R}$ et $x \in V$. On note $\text{div} f(x)$ et on appelle divergence de f en x la quantité :

$$\text{div} f(x) := \sum_{y: \{x, y\} \in E} f(\vec{x}y).$$

Définition 3.11 (Flot). Soit $\{G, c\}$ un réseau et $a, z \in V$. On appelle flot de a à z toute fonction $\theta : \vec{E} \rightarrow \mathbb{R}$ qui vérifie les propriétés suivantes :

- $\forall \vec{xy} \in \vec{E}, \theta(\vec{xy}) + \theta(\vec{yx}) = 0$ (antisymétrie).
- $\forall x \notin \{a, z\},$

$$\operatorname{div} \theta(x) = 0 \quad (\text{loi des noeuds})$$

- $\operatorname{div} \theta(a) \geq 0.$

Définition 3.12. L'intensité du flot θ de a à z est définie par

$$\|\theta\| = \operatorname{div} \theta(a) = \sum_{x: \{a, x\} \in E} \theta(\vec{ax}).$$

Définition 3.13. Un flot unitaire est un flot d'intensité 1.

Le rôle des points source a et puits z n'est pas symétrique du fait de la seule inégalité $\operatorname{div} \theta(a) \geq 0$ dans la définition du flot. Notons que pour un flot θ , l'antisymétrie assure que :

$$\operatorname{div} \theta(a) + \operatorname{div} \theta(z) = \sum_{x \in V} \operatorname{div} \theta(x) = \sum_{x \in V} \sum_{y: \{x, y\} \in E} \theta(\vec{xy}) = \sum_{\{x, y\} \in E} (\theta(\vec{xy}) + \theta(\vec{yx})) = 0.$$

On étend la définition des conductances et résistances aux arêtes orientées par les relations

$$c(x, y) = c(\vec{xy}) = c(\vec{yx}) \text{ et } r(x, y) = r(\vec{xy}) = r(\vec{yx}).$$

Définition 3.14. Soit $\{G, c\}$ un réseau et W une tension sur $\{G, c\}$. Le courant I associé à W est défini par

$$\forall \vec{xy} \in \vec{E}, I(\vec{xy}) = \frac{W(x) - W(y)}{r(\vec{xy})} = c(\vec{xy})(W(x) - W(y)).$$

Noter que cette définition correspond à ce qu'on appelle en physique la loi d'Ohm :

$$r(\vec{xy})I(\vec{xy}) = W(x) - W(y).$$

Pour éviter tout problème de signe, on retiendra que le courant va des sommets de potentiel maximal à ceux de potentiel minimal. I ainsi définie est antisymétrique, et puisque W est une tension, elle satisfait également la loi des noeuds. Enfin, $\operatorname{div} I(a) \geq 0$ dès lors que $W(a) \geq W(z)$ (utiliser le principe du maximum, Proposition 3.6). On a donc établi le :

Lemme 3.15. Si $W(a) \geq W(z)$, alors le courant I est un flot de a à z .

Définition 3.16. On dit qu'un flot $\theta : \vec{E} \rightarrow \mathbb{R}$ vérifie la loi des cycles si, pour toute suite de sommets $\vec{e}_1, \vec{e}_2, \dots, \vec{e}_m$ qui forme un cycle orienté, on a

$$\sum_{i=1}^m r(\vec{e}_i)\theta(\vec{e}_i) = 0.$$

On note que le flot courant permet de retrouver la fonction de tension à constante additive près. On a enfin le résultat suivant dont la démonstration est laissée au lecteur.

Proposition 3.17. Le flot courant I vérifie la loi des cycles

La proposition suivante assure que les deux propriétés précédentes caractérisent le flot courant.

Proposition 3.18. *Si θ est un flot de a à z qui vérifie :*

- *la loi des cycles pour tout cycle $\vec{e}_1, \vec{e}_2, \dots, \vec{e}_m$*
- *la normalisation $\|\theta\| = \|I\|$,*

alors $\theta = I$.

Démonstration. Posons $f = \theta - I$. Alors f satisfait la loi des nœuds et la loi des cycles. Supposons par l'absurde que $f \neq 0$, et par exemple $f(\vec{e}_1) > 0$ pour \vec{e}_1 une arête du réseau. Alors, par la loi des nœuds, il existe \vec{e}_2 tel que $f(\vec{e}_2) > 0$. On construit ainsi une suite d'arêtes sur lesquelles f est strictement positive. Or, V est fini donc cette suite va passer 2 fois sur un même nœud, et en sommant, on contredit la loi des cycles (rappelons que la loi des cycles est vérifiée même en les cycles qui comprennent a ou z). Donc $f = 0$ et $\theta = I$. \square

3.4 Résistance équivalente

Définition 3.19 (Résistance équivalente). Soit un réseau $\{G, c\}$ irréductible, et $a, z \in V$. Il existe une unique extension harmonique W telle que $W(a) = 1$ et $W(z) = 0$. On appelle résistance équivalente entre a et z le rapport :

$$\mathcal{R}(a \leftrightarrow z) = \frac{1}{\|I\|},$$

avec I le flot courant associé à la tension W . La conductance équivalente est définie comme l'inverse de la résistance équivalente, par $\mathcal{C}(a \leftrightarrow z)\mathcal{R}(a \leftrightarrow z) = 1$.

Il est bien sûr possible de définir la résistance équivalente entre a et z sans préciser la valeur de la tension en a et z .

Proposition 3.20. *Soit un réseau $\{G, c\}$ irréductible, et $a, z \in V$. Si W est harmonique sur $V \setminus \{a, z\}$, et $W(a) \neq W(z)$, alors*

$$\mathcal{R}(a \leftrightarrow z) = \frac{W(a) - W(z)}{\|I\|},$$

avec I le flot courant associé à la tension W .

Démonstration. On pose

$$x \mapsto \bar{W}(x) := \frac{W(x) - W(z)}{W(a) - W(z)}$$

et on note que la fonction \bar{W} ainsi définie est encore harmonique sur $V \setminus \{a, z\}$, avec les conditions au bord $\bar{W}(a) = 1$ et $\bar{W}(z) = 0$ (c'est donc encore une tension, mais les valeurs

aux bornes sont différentes) donc, si I' désigne le flot courant associé à W' :

$$\begin{aligned}
\mathcal{R}(a \leftrightarrow z) &= \frac{1}{\|I'\|} \\
&= \frac{1}{\sum_{x:\{a,x\} \in E} \bar{I}(a\vec{x})} \\
&= \frac{1}{\sum_{x:\{a,x\} \in E} c(ax) \bar{W}(a) - \bar{W}(x)} \\
&= \frac{1}{\sum_{x:\{a,x\} \in E} c(ax) \frac{W(a) - W(x)}{W(a) - W(z)}} \\
&= \frac{W(a) - W(z)}{\sum_{x:\{a,x\} \in E} c(ax) (W(a) - W(x))} \\
&= \frac{W(a) - W(z)}{\|I\|}
\end{aligned}$$

□

L'interprétation est la suivante : si l'on remplace le réseau $\{G, c\}$ entre a et z par *une seule arête*, quelle résistance/conductance lui attribuer pour que, à tension fixée, l'intensité du courant soit identique ou, à courant d'intensité fixée, la différence de tension entre a et z soit identique : la réponse est dans les deux cas donnée par la résistance/conductance équivalente.

On note une première représentation probabiliste des tensions :

Lemme 3.21 (Probabilités d'atteinte comme tension). *Pour tout $a, z \in V$, l'application*

$$x \mapsto \mathbb{P}_x(\tau_z < \tau_a)$$

est une tension, il s'agit de l'unique tension de valeur 0 en a et 1 en z .

Noter qu'on aurait aussi bien pu considérer la fonction $x \mapsto \mathbb{P}_x(\tau_a < \tau_z)$, c'est encore une tension, avec seulement des valeurs aux bords sont modifiées.

Démonstration. Il s'agit d'une simple application du principe de Dirichlet puisque : $\mathbb{P}_x(\tau_z < \tau_a) = \mathbb{E}_x[h(X_{\tau_{a,z}})]$ où $h : \{a, z\} \rightarrow \{0, 1\}$, $x \mapsto \mathbf{1}_{\{z\}}(x)$. □

On en tire une représentation intéressante de la conductance équivalente qui fait apparaître $\mathbb{P}_a(\tau_z < \tau_a^+)$ comme le facteur d'amortissement entre $c(a)$ et $\mathcal{C}(a \leftrightarrow z)$. Notons qu'on pourrait aussi bien prendre cette propriété comme définition.

Proposition 3.22. *Pour tout $a, z \in V$,*

$$\mathcal{C}(a \leftrightarrow z) = c(a) \mathbb{P}_a(\tau_z < \tau_a^+)$$

On tire de cette proposition la linéarité des conductances équivalentes en les conductances : si l'on multiplie toutes les conductances des arêtes $c(e)$ par un même facteur, alors pour tout $a \in V$, $c(a)$ est multiplié par ce même facteur, tandis les probabilités de transition restent inchangées.

Démonstration. Il suit du lemme 3.21 que

$$\begin{aligned}
\mathbb{P}_a\{\tau_z < \tau_a^+\} &= \sum_{x, \{a, x\} \in E} P(a, x) \mathbb{P}_x\{\tau_z < \tau_a\} \\
&= \sum_{x, \{a, x\} \in E} \frac{c(a, x)}{c(a)} \frac{W(a) - W(x)}{W(a) - W(z)} \\
&= \sum_{x, \{a, x\} \in E} \frac{I(a\vec{x})}{c(a)(W(a) - W(z))} \\
&= \frac{\|I\|}{c(a)(W(a) - W(z))} \\
&= \frac{\mathcal{C}(a \leftrightarrow z)}{c(a)} \qquad \text{par la définition 3.19}
\end{aligned}$$

□

Proposition 3.23 (Fonction de Green et résistance équivalente). *Pour tout $a \neq z \in V$, la résistance équivalente admet l'expression suivante en fonction de la fonction de Green :*

$$G_{\tau_z}(a, a) = c(a)\mathcal{R}(a \leftrightarrow z)$$

Démonstration. On commence par noter que pour tout $n \in \mathbb{N}$, $n = \sum_{t=0}^{+\infty} \mathbf{1}_{\{n > t\}}$; partant, toute variable aléatoire N à valeurs dans \mathbb{N} satisfait $\mathbb{E}[N] = \sum_{t=0}^{\infty} \mathbb{P}(N > t)$. On a donc

$$\begin{aligned}
G_{\tau_z}(a, a) &= \mathbb{E}_a \left[\sum_{s \geq 0} \mathbf{1}_{\{X_s = a, s < \tau_z\}} \right] \\
&= \sum_{t=0}^{+\infty} \mathbb{P}_a \left(\sum_{s \geq 0} \mathbf{1}_{\{X_s = a, s < \tau_z\}} > t \right) \\
&= \sum_{t=0}^{+\infty} \mathbb{P}_a(\tau_a^+ < \tau_z)^t \qquad \text{par la propriété de Markov forte} \\
&= \frac{1}{1 - \mathbb{P}_a(\tau_z > \tau_a^+)} \\
&= \frac{1}{\mathbb{P}_a(\tau_z \leq \tau_a^+)} \\
&= \frac{1}{\mathbb{P}_a(\tau_z < \tau_a^+)} \\
&= c(a)\mathcal{R}(a \leftrightarrow z) \qquad \text{par la Proposition 3.22}
\end{aligned}$$

□

On a alors une deuxième interprétation probabiliste des tensions.

Proposition 3.24 (Fonction de Green comme tension). *Pour tout $a, z \in V$, l'application*

$$x \mapsto \frac{G_{\tau_z}(a, x)}{c(x)}$$

est une tension, et c'est l'unique tension associée à un courant unitaire, nul en z .

Démonstration. On note que : $\frac{G_{\tau_z}(a,x)}{c(x)} = \frac{G_{\tau_z}(x,a)}{c(a)}$ par réversibilité, puis $G_{\tau_z}(x, a) = \sum_{t \geq 0} \mathbb{P}_x(X_t = a, t < \tau_z)$ est harmonique en tout point x distinct de a , puisque :

$$\begin{aligned} G_{\tau_z}(x, a) &= \sum_{t \geq 0} \mathbb{P}_x(X_t = a, t < \tau_z) \\ &= \sum_{t \geq 0} \sum_y P(x, y) \mathbb{P}_y(X_t = a, t < \tau_z) \text{ car } x \neq a \\ &= \sum_y P(x, y) \sum_{t \geq 0} \mathbb{P}_y(X_t = a, t < \tau_z) \\ &= \sum_y P(x, y) G_{\tau_z}(y, a) \end{aligned}$$

Maintenant, on sait de plus que la tension en a vaut $\mathcal{R}(a \leftrightarrow z)$ du lemme 3.23. Il en découle que $I = \mathcal{R}(a \leftrightarrow z)$. L'unicité est bien sûr conséquence du principe de Dirichlet. \square

3.5 Temps de transport

Il est temps de tirer les fruits de l'introduction de la résistance équivalente. On rappelle que θ désigne l'opérateur de translation en temps, défini en 2.8.

Définition 3.25. Soit $a, b \in V$ et $(X_t)_{t \in \mathbb{N}} \in V^{\mathbb{N}}$ une trajectoire issue de a . On note $\tau_{b,a}$ la variable aléatoire :

$$\tau_{b,a} := \min\{t \geq \tau_b, X_t = a\} = \min\{t \in \mathbb{N} | \exists s < t, X_s = b, X_t = a\} = \tau_a \circ \theta_{\tau_b}.$$

où l'on rappelle la convention que le minimum d'un ensemble vide est égal à $+\infty$. On note $t_{a \leftrightarrow b}$ et on appelle temps de transport entre a et b :

$$t_{a \leftrightarrow b} := \mathbb{E}_a[\tau_{b,a}].$$

Ainsi le temps de transport est le temps espéré, partant de a , pour revenir en a après avoir visité b (ce qui correspond bien au temps de transport quotidien si a ="maison" et b ="travail"). Notons que la propriété de Markov forte implique :

$$\mathbb{E}_a[\tau_{b,a}] = \mathbb{E}_a[\tau_b] + \mathbb{E}_b[\tau_a],$$

et donc $t_{a \leftrightarrow b}$ est bien une expression symétrique en a et b , comme le laissait présager son écriture. L'identité suivante sera notre outil principal pour évaluer des espérances de temps d'atteinte. Elle nécessite le calcul préalable de la résistance équivalente et justifie au passage l'intérêt de cette notion.

Proposition 3.26 (Identité du temps de transport). *Soit P irréductible réversible. Pour tout $a, b \in V$,*

$$t_{a \leftrightarrow b} = c_G \mathcal{R}(a \leftrightarrow b).$$

On se rappelle que c_G défini en (3.1) comptabilise la somme des conductances des sommets (la conductance d'une arête qui n'est pas une boucle est comptée deux fois dans cette somme). Nous verrons en TD que cette identité fondamentale combinée permet bien souvent de se passer de la méthode à un pas (et de la résolution de la récurrence qui lui est souvent associée).

Démonstration. On note π la mesure par rapport à laquelle P est réversible. Par le lemme de représentation de Aldous-Fill et l'unicité de la mesure stationnaire, on a

$$\frac{G_{\tau_{b,a}}(a, a)}{G_{\tau_{b,a}}(a, V)} = \pi(a) = \frac{c(a)}{c_G}.$$

Ensuite, par définition, $G_{\tau_{b,a}}(a, V) = \mathbb{E}_a \left[\sum_{t=0}^{+\infty} \mathbf{1}_{\{X_t=x, \tau_{b,a}>t\}} \right] = \mathbb{E}_a(\tau_{b,a}) = t_{a \leftrightarrow b}$. De plus, d'après la définition des temps d'arrêt, et une application de la proposition 3.23,

$$G_{\tau_{b,a}}(a, a) = G_{\tau_b}(a, a) = c(a)\mathcal{R}(a \leftrightarrow b)$$

On obtient en remplaçant $G_{\tau_{b,a}}(a, a)$ et $G_{\tau_{a,b}}(a, V)$ par leurs deux expressions l'identité du temps de transport. \square

3.6 Énergie

Définition 3.27. Soit θ un flot sur $\{G, c\}$. On définit l'énergie du flot θ par

$$\mathcal{E}(\theta) = \sum_{e \in E} \theta(e)^2 r(e).$$

Le principe variationnel justifie l'introduction de l'énergie.

Theorème 3.28 (Principe de Thomson). *Pour tout graphe fini connexe,*

$$\mathcal{R}(a \leftrightarrow z) = \inf \{ \mathcal{E}(\theta) : \theta \text{ flot unitaire de } a \text{ à } z \},$$

et la borne inférieure est atteinte en le flot courant unitaire.

Ainsi donc, en principe, il suffit de calculer l'extension harmonique de la fonction de valeur 1 en a et 0 en z (c'est-à-dire la tension), de la normaliser, puis de calculer l'énergie du flot associé. En pratique cependant, on ne calcule jamais une résistance équivalente de la sorte (!), et l'intérêt de ce résultat réside dans les conséquences théoriques qu'on peut en tirer (à commencer par le principe de monotonie de Rayleigh discuté en corollaire). Aussi, le principe de Thomson variationnel fournit des bornes supérieures sur la résistance équivalente (moyennant la construction d'un "bon" flot, ce qui, en pratique, nécessite déjà une compréhension fine du problème).

Il existe une version duale de ce principe (elle sera étudiée en TD) où la conductance équivalente est exprimée comme la solution d'un problème de minimisation (ce qui permet donc de calculer des bornes sup pour la conductance équivalente, c'est-à-dire minorer la résistance équivalente.)

Démonstration. Comme l'ensemble des flots unitaires est un fermé borné de $\mathbb{R}^{|E|}$, c'est un compact et il existe donc un flot θ qui minimise \mathcal{E} . Il suffit de montrer que θ vérifie la loi des mailles pour l'identifier avec le flot courant unitaire.

Soient $\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n$ un cycle orienté. Soit γ défini par $\gamma = \sum_i (\mathbf{1}_{\vec{e}_i} - \mathbf{1}_{\vec{e}_i})$. Notons que γ définit un flot.

Soit $\varepsilon \in \mathbb{R}$, comme θ minimise \mathcal{E} , on a

$$\begin{aligned} 0 \leq \mathcal{E}(\theta + \varepsilon\gamma) - \mathcal{E}(\theta) &= \sum_{i=1}^n ((\theta(\vec{e}_i) + \varepsilon)^2 - \theta(\vec{e}_i)^2) r(\vec{e}_i) \\ &= 2\varepsilon \sum_{i=1}^n r(\vec{e}_i)\theta(\vec{e}_i) + \mathcal{O}(\varepsilon^2). \end{aligned}$$

En divisant par $\varepsilon > 0$ et en faisant tendre ε vers 0, on obtient $0 \leq \sum_{i=1}^n r(\vec{e}_i)\theta(\vec{e}_i)$. De même, en divisant par $\varepsilon < 0$ et en faisant tendre ε vers 0, on obtient $0 \geq \sum_{i=1}^n r(\vec{e}_i)\theta(\vec{e}_i)$. Donc, $\sum_{i=1}^n r(\vec{e}_i)\theta(\vec{e}_i) = 0$, θ vérifie la loi des mailles. Il reste à montrer que $\mathcal{E}(I) = \mathcal{R}(a \leftrightarrow z)$.

$$\begin{aligned} \mathcal{E}(I) &= \sum_e r(e)I(e)^2 = \frac{1}{2} \sum_x \sum_y r(x, y) \left(\frac{W(x) - W(y)}{r(x, y)} \right)^2 \\ &= \frac{1}{2} \sum_x \sum_y c(x, y) (W(x) - W(y))^2 \\ &= \frac{1}{2} \sum_x \sum_y I(x\vec{y}) (W(x) - W(y)) \\ &= \sum_x W(x) \sum_y I(x\vec{y}) && \text{car } I \text{ est antisymétrique} \\ &= W(a) \sum_y I(a\vec{y}) + W(z) \sum_y I(z\vec{y}) && \text{par la loi des nœuds} \\ &= \|I\|(W(a) - W(z)) \\ &= \mathcal{R}(a \leftrightarrow z) \end{aligned}$$

□

Proposition 3.29 (Principe de monotonie de Rayleigh). *Soient $\{r(e)\}$ et $\{r'(e)\}$ deux ensembles de résistances sur les arêtes du même graphe G , et $\mathcal{R}(a \leftrightarrow z; r)$ et $\mathcal{R}(a \leftrightarrow z; r')$ les résistances équivalentes associées. Montrer que si, pour tout e , $r(e) \leq r'(e)$, alors*

$$\mathcal{R}(a \leftrightarrow z; r) \leq \mathcal{R}(a \leftrightarrow z; r').$$

Démonstration. On a, pour tout flot θ ,

$$\sum_e r(e)\theta(e)^2 \leq \sum_e r'(e)\theta(e)^2.$$

On minimise alors sur les flots θ unitaire pour conclure à l'aide du principe de Thomson. □

Définition 3.30. On appelle *ensemble d'arêtes séparateur entre a et z* un ensemble d'arêtes tel que tout chemin de a à z emprunte au moins une des arêtes de cet ensemble.

Proposition 3.31 (Borne inférieure de Nash-Williams). *Soit I un ensemble (nécessairement fini puisque E l'est) et $\{\Pi_k\}_{k \in I}$ une famille d'ensembles séparateurs entre a et z deux à deux disjoints, alors*

$$\mathcal{R}(x \leftrightarrow y) \geq \sum_{k \in I} \left(\sum_{e \in \Pi_k} c(e) \right)^{-1}.$$

Démonstration. Soit θ un flot unitaire de a à z . Alors, pour tout $k \in I$, l'inégalité de Cauchy-Schwarz donne

$$\sum_{e \in \Pi_k} c(e) \cdot \sum_{e \in \Pi_k} r(e)\theta(e)^2 \geq \left(\sum_{e \in \Pi_k} \sqrt{c(e)}\sqrt{r(e)}|\theta(e)| \right)^2 = \left(\sum_{e \in \Pi_k} |\theta(e)| \right)^2 \geq 1,$$

(on pourra chercher comment justifier la dernière inégalité). Ainsi,

$$\sum_e r(e)\theta(e)^2 \geq \sum_{k \in I} \sum_{e \in \Pi_k} r(e)\theta(e)^2 \geq \sum_{k \in I} \left(\sum_{e \in \Pi_k} c(e) \right)^{-1}.$$

□

Chapitre 4

Temps d'atteinte et temps de couverture

Pour une chaîne irréductible à espace d'état fini, on sait que les temps d'atteinte sont finis p.s., et même d'espérance finie. Après ces temps d'atteinte, une des quantités les plus naturelles à considérer est le plus grand de ces temps d'atteinte, qui correspond aussi au premier instant où l'espace entier a été visité. Il porte le nom de temps de couverture.

On prendra soin de distinguer, parmi les quantités définies dans ce chapitre, les quantités aléatoires des quantités déterministes. Dans la mesure du possible, on utilisera τ pour une quantité aléatoire et t pour une quantité déterministe.

On travaille dans cette section avec une chaîne $(X_t)_{t \in \mathbb{N}}$ définie sur Ω . On consultera la définition 2.14 si besoin pour la définition du temps d'atteinte d'un sommet.

Définition 4.1. On note t_{hit} et on appelle temps d'atteinte de la chaîne $(X_t)_{t \in \mathbb{N}}$ le temps déterministe :

$$t_{\text{hit}} = \max_{x,y \in V} \mathbb{E}_x[\tau_y].$$

Définition 4.2. On note τ_{cov} et on appelle temps de couverture de la trajectoire $(X_t)_{t \in \mathbb{N}}$ la variable aléatoire

$$\tau_{\text{cov}} = \min\{t \geq 0, \{X_s\}_{0 \leq s \leq t} = V\} = \max_{x \in V} \tau_x.$$

On note t_{cov} et on appelle temps de couverture le temps déterministe :

$$t_{\text{cov}} = \max_{x \in V} \mathbb{E}_x[\tau_{\text{cov}}]$$

Bien entendu une compréhension complète des variables aléatoires τ_{cov} (jusqu'aux fluctuations) est plus informative que celle de t_{cov} ; néanmoins, le calcul de t_{cov} est un premier pas important. Un premier lien simple entre t_{hit} et t_{cov} est :

$$t_{\text{hit}} = \max_{x,y \in V} \mathbb{E}_x[\tau_y] \leq \max_{x \in V} \mathbb{E}_x[\tau_{\text{cov}}] \leq t_{\text{cov}}. \quad (4.1)$$

4.1 Cas réversible.

Commençons par glâner quelques informations supplémentaires sur les temps d'atteinte. Notre résultat principal à ce sujet, l'identité du temps de transport, met en jeu la somme de

deux termes, les temps d'atteinte espérés ; pour connaître chacun de ces deux termes, on a besoin d'une information supplémentaire. Un cas particulièrement pratique est celui où les deux temps sont égaux. On aimerait formuler une condition sous laquelle ceci vaut.

Notons tout d'abord que même dans le cas de marches réversibles, on ne peut espérer avoir l'égalité $\mathbb{E}_a[\tau_b] = \mathbb{E}_b[\tau_a]$: il suffit en effet de considérer un graphe G connexe avec au moins 3 sommets, dont deux sommets a et b tels que a ait pour seul voisin b et des conductances unitaires et sans boucle (par exemple). Alors $1 = \mathbb{E}_a[\tau_b] < \mathbb{E}_b[\tau_a]$.

En revanche, si l'on démarre la chaîne sous sa mesure stationnaire (par rapport à laquelle elle est réversible), des propriétés intéressantes peuvent être énoncées en toute généralité.

On généralise la notion de temps d'atteinte définie en 2.14 à une suite finie de sommets :

Définition 4.3. Soit $x_1, \dots, x_\ell \in V^\ell$. On note $\tau_{x_1, \dots, x_\ell}$ et on appelle temps d'atteinte de x_1, \dots, x_ℓ le temps aléatoire :

$$\tau_{x_1, \dots, x_\ell} = \min\{t_\ell \in \mathbb{N} : \exists 0 \leq t_1 < t_2 < \dots < t_\ell, (i \in \{1, \dots, \ell\} \Rightarrow X_{t_i} = x_i)\}$$

Ayant défini le temps d'atteinte d'un sommet, on aurait aussi pu adopter la définition récursive suivante, à l'aide de l'opérateur θ de translation en temps défini en 2.8 :

$$\tau_{x_1, \dots, x_\ell} = \tau_{x_1} + \theta_{\tau_{x_1}} \circ \tau_{x_2, \dots, x_\ell} \text{ et } \tau_x = \min\{t \in \mathbb{N}, X_t = x\}$$

Il s'agit du premier instant où x_1, x_2, \dots, x_ℓ ont été visités dans cet ordre par la trajectoire. (Noter qu'on a droit aux répétitions parmi les x_i .) Par exemple, $\tau_{abc}(cbbaccb\bar{a}cb\dots) = 8$ (le c souligné apparaît en neuvième position, mais on initialise le compteur à 0). Une propriété importante des marches aléatoires sur réseau est la suivante

Proposition 4.4. Soit un réseau transitif $\{G, c\}$, soit $x_1, \dots, x_\ell \in V^\ell$, et soit $(X_t)_{t \in \mathbb{N}}$ la marche aléatoire associée. On a l'identité

$$\mathbb{E}_\pi[\tau_{x_1, \dots, x_\ell}] = \mathbb{E}_\pi[\tau_{x_\ell, \dots, x_1}].$$

On en déduit immédiatement l'identité suivante, pour les cycles.

Corollaire 4.5 (Lemme cyclique). Sous les hypothèses précédentes, et si de plus $x_1 = x_\ell$, c'est-à-dire si la suite de sommets forme un cycle alors :

$$\mathbb{E}_{x_1}[\tau_{x_2, \dots, x_\ell}] = \mathbb{E}_{x_\ell}[\tau_{x_{\ell-1}, \dots, x_1}].$$

Le Corollaire se déduit directement de la Proposition.

Démonstration du Corollaire. Il suffit d'observer que la Proposition 4.4 s'écrit aussi de la propriété de Markov forte :

$$\mathbb{E}_\pi[\tau_{x_1}] + \mathbb{E}_{x_1}[\tau_{x_2}] + \dots + \mathbb{E}_{x_{\ell-1}}[\tau_{x_\ell}] = \mathbb{E}_\pi[\tau_{x_\ell}] + \mathbb{E}_{x_\ell}[\tau_{x_{\ell-1}}] + \dots + \mathbb{E}_{x_2}[\tau_{x_1}]$$

ce qui implique, en soustrayant le premier terme de chaque membre (puisque $x_1 = x_\ell$),

$$\mathbb{E}_{x_1}[\tau_{x_2, \dots, x_\ell}] = \mathbb{E}_{x_1}[\tau_{x_2}] + \dots + \mathbb{E}_{x_{\ell-1}}[\tau_{x_\ell}] = \mathbb{E}_\pi[\tau_{x_\ell}] + \mathbb{E}_{x_\ell}[\tau_{x_{\ell-1}}] + \dots + \mathbb{E}_{x_2}[\tau_{x_1}] = \mathbb{E}_{x_\ell}[\tau_{x_{\ell-1}, \dots, x_1}]$$

□

Démonstration. On définit un ordre partiel sur l'ensemble des mots finis à valeurs dans V : $m \preceq m'$ si m est sous-mot de m' , c'est-à-dire que $m = (x_1 \dots x_k)$ et $m' = (x'_1 \dots x'_{k'})$ et il existe $x'_i = x_{\pi(i)}$ pour $1 \leq i \leq k$ et π strictement croissante. Notons maintenant que, pour $k \geq \ell - 1$,

$$\begin{aligned} \mathbb{P}_\pi(\tau_{x_1, \dots, x_\ell} \leq k) &= \mathbb{P}_\pi((x_1, \dots, x_\ell) \preceq (X_0, \dots, X_k)) \\ &= \mathbb{P}_\pi((x_1, \dots, x_\ell) \preceq (X_k, \dots, X_0)) && \text{par réversibilité} \\ &= \mathbb{P}_\pi((x_k, \dots, x_1) \preceq (X_0, \dots, X_k)) && \text{par définition} \\ &= \mathbb{P}_\pi(\tau_{x_\ell, \dots, x_1} \leq k) \end{aligned}$$

Il s'ensuit

$$\mathbb{E}_\pi[\tau_{x_1, \dots, x_\ell}] = \sum_{k \geq 1} \mathbb{P}_\pi(\tau_{x_1, \dots, x_\ell} \geq k) = \sum_{k \geq 1} \mathbb{P}_\pi(\tau_{x_\ell, \dots, x_1} \geq k) = \mathbb{E}_\pi[\tau_{x_\ell, \dots, x_1}].$$

□

Pour se débarrasser la mesure initiale stationnaire dans la Proposition 4.4, et pouvoir considérer une mesure initiale arbitraire, une hypothèse supplémentaire est nécessaire. l'hypothèse de transitivité est une hypothèse de symétrie qui énonce que le graphe vu depuis n'importe lequel de ses sommets est "semblable".

Définition 4.6. Un réseau est transitif si pour tout sommets $(a, b) \in V^2$, il existe une bijection $\phi : V \rightarrow V$ telle que $\phi(a) = b$ et ϕ préserve les conductances :

$$c(\phi(x), \phi(y)) = c(x, y)$$

ayant étendu la conductance c à une application sur l'ensemble des paires de sommets (et non les seules arêtes), en posant $c(x, y) = \infty$ si $\{x, y\} \notin E$.

Corollaire 4.7. *Sous les hypothèses de la Proposition 4.4, et si de plus le réseau $\{G, c\}$ est transitif, alors on a l'identité :*

$$\mathbb{E}_{x_1}[\tau_{x_2, \dots, x_\ell}] = \mathbb{E}_{x_\ell}[\tau_{x_{\ell-1}, \dots, x_1}].$$

Démonstration. La proposition 4.4 s'écrit aussi :

$$\mathbb{E}_\pi[\tau_{x_1}] + \mathbb{E}_{x_1}[\tau_{x_2}] + \dots + \mathbb{E}_{x_{\ell-1}}[\tau_{x_\ell}] = \mathbb{E}_\pi[\tau_{x_\ell}] + \mathbb{E}_{x_\ell}[\tau_{x_{\ell-1}}] + \dots + \mathbb{E}_{x_2}[\tau_{x_1}]$$

Mais puisque le réseau est transitif, on sait que

$$\mathbb{E}_\pi[\tau_{x_1}] = \mathbb{E}_\pi[\tau_{x_\ell}].$$

(décomposer selon π si besoin). La différence des deux égalités précédentes fournit le résultat cherché. □

Un corollaire immédiat est le suivant :

Corollaire 4.8. *Soit un réseau transitif irréductible $\{G, c\}$, et $a, b \in V$, et $(X_t)_{t \in \mathbb{N}}$ la marche aléatoire associée. On a l'identité*

$$\mathbb{E}_a[\tau_b] = \frac{\mathbb{E}_a[\tau_{b,a}]}{2} = \frac{c_G \mathcal{R}(a \leftrightarrow b)}{2}.$$

4.2 Borne de Matthews (de l'aléa pour construire une borne supérieure)

On peut compléter la borne inférieure sur t_{cov} par une borne supérieure.

Theorème 4.9 (Borne supérieure de Matthews). *Soit $(X_t)_{t \in \mathbb{N}}$ une chaîne de Markov sur Ω de cardinal n . Alors*

$$t_{\text{cov}} \leq t_{\text{hit}} \left(1 + \frac{1}{2} + \cdots + \frac{1}{n} \right).$$

Démonstration. Soit σ une permutation uniforme de Ω , indépendante de la chaîne de Markov $(X_t)_{t \in \mathbb{N}}$. On note T_k le premier instant où les états $\sigma(1), \sigma(2), \dots, \sigma(k)$ ont été visités et $L_k = X_{T_k}$ le dernier de ces états à avoir été visité. On remarque

$$\mathbb{E}_x[\tau_{\text{cov}}] = \mathbb{E}_x[T_n] = \mathbb{E}_x[T_1] + \sum_{i=2}^n \mathbb{E}_x[T_i - T_{i-1}].$$

Il est clair que

$$\mathbb{E}_x[T_1] = \sum_y \mathbb{E}_x[T_1 \mid \sigma(1) = y] \mathbb{P}(\sigma(1) = y) = \sum_x \mathbb{E}_x[\tau_y] \frac{1}{n} \leq t_{\text{hit}} \sum_y \frac{1}{n} = t_{\text{hit}}$$

Puis, notant que $T_i - T_{i-1}$ est non nul ssi $\sigma(i)$ est le dernier des sommets visités parmi $\sigma(1), \sigma(2), \dots, \sigma(i)$, événement dont la probabilité s'évalue à $\frac{1}{i}$ par échangeabilité, on a :

$$\begin{aligned} \mathbb{E}_x[T_i] - \mathbb{E}_x[T_{i-1}] &= \mathbb{E}_x[T_i - T_{i-1}] \\ &= \mathbb{E}_x[\tau_{\sigma(i)} \circ \theta_{T_{i-1}} \mathbf{1}_{\{\tau_{\sigma(i)} > \max_{j \leq i-1} \tau_{\sigma(j)}\}}] \quad \text{de la propriété de Markov forte} \\ &= \mathbb{E}_x[\mathbb{E}_{X_{T_{i-1}}}[\tau_{\sigma(i)}] \mathbf{1}_{\{\tau_{\sigma(i)} > \max_{j \leq i-1} \tau_{\sigma(j)}\}}] \\ &\leq \mathbb{E}_x[t_{\text{hit}} \mathbf{1}_{\{\tau_{\sigma(i)} > \max_{j \leq i-1} \tau_{\sigma(j)}\}}] \\ &\leq t_{\text{hit}} \mathbb{P}_x(\tau_{\sigma(i)} > \max_{j \leq i-1} \tau_{\sigma(j)}) \\ &= \frac{1}{i} t_{\text{hit}} \end{aligned}$$

On conclut en sommant les inégalités obtenues. □

Il est utile de se demander dans quels cas la borne a des chances d'être précise. On note ensuite qu'une simple adaptation de la démonstration permet de donner un minorant de t_{cov} en fonction des quantités

$$t_{\min}^A := \min_{a, b \in A, a \neq b} \mathbb{E}_a[\tau_b] \text{ où } A \subset \Omega.$$

Theorème 4.10 (Borne inférieure de Matthews). *Soit $(X_t)_{t \in \mathbb{N}}$ une chaîne de Markov sur Ω de cardinal n . Alors*

$$t_{\text{cov}} \geq \max_{A \subset \Omega} t_{\min}^A \cdot \left(1 + \frac{1}{2} + \cdots + \frac{1}{|A| - 1} \right).$$

Par rapport à la borne supérieure de Matthews, on notera que dans le membre de droite, le dénominateur est $A - 1$ et non pas A . Surtout t_{\min}^A remplace t_{hit} . Si tout choix de A donne une borne inférieure, mais la qualité de la borne obtenue réside dans le choix du A .

Démonstration. Soit $x \in A$, et σ une permutation aléatoire uniforme de $A \setminus \{x\}$, indépendante de la chaîne de Markov. On note à nouveau T_k le premier instant où les états $\sigma(1), \sigma(2), \dots, \sigma(k)$ ont été visités et $L_k = X_{T_k}$ le dernier des ces états à avoir été visité. On décompose

$$\mathbb{E}_x[\tau_{\text{cov}}] \geq \mathbb{E}_x[T_A] = \mathbb{E}_x[T_1] + \sum_{i=2}^{|A|-1} \mathbb{E}_x[T_i - T_{i-1}].$$

On a que :

$$\begin{aligned} \mathbb{E}_x[T_i] - \mathbb{E}_x[T_{i-1}] &= \mathbb{E}_x[T_i - T_{i-1}] \\ &= \mathbb{E}_x[\tau_{\sigma(i)} \circ \theta_{T_{i-1}} \mathbf{1}_{\{\tau_{\sigma(i)} > \max_{j \leq i-1} \tau_{\sigma(j)}\}}] \quad \text{de la propriété de Markov forte} \\ &= \mathbb{E}_x[\mathbb{E}_{X_{T_{i-1}}}[\tau_{\sigma(i)}] \mathbf{1}_{\{\tau_{\sigma(i)} > \max_{j \leq i-1} \tau_{\sigma(j)}\}}] \\ &\geq \mathbb{E}_x[t_{\min}^A \mathbf{1}_{\{\tau_{\sigma(i)} > \max_{j \leq i-1} \tau_{\sigma(j)}\}}] \\ &\geq t_{\min}^A \mathbb{P}_x(\tau_{\sigma(i)} > \max_{j \leq i-1} \tau_{\sigma(j)}) \\ &= \frac{1}{i} t_{\text{hit}} \end{aligned}$$

□

Regardons ce que donnent les bornes de Matthews sur l'exemple du tore de dimension 1.

Theorème 4.11 (Bornes de Matthews pour le temps de couverture du n -cycle). *Le temps de couverture du n -cycle vérifie*

$$\frac{n^2}{4} \leq t_{\text{cov}} \leq \frac{n^2}{4} \log(n)(1 + o(1)).$$

Démonstration. On prend pour A un ensemble composé de deux sommets opposés à distance maximale. Les temps d'atteinte sont calculés en utilisant la formule du temps de transport et le lemme concernant les graphes transitifs. □

Les bornes précédentes sont données à titre d'exemple seulement, puisque dans ce cas précis il est possible de faire un calcul exact : on montre qu'on se ramène à un problème de ruine du joueur.

Theorème 4.12 (Temps de couverture exact du n -cycle). *Le temps de couverture du tore du n -cycle vérifie*

$$t_{\text{cov}} = \frac{n(n-1)}{2}.$$

On voit donc que la borne inférieure était ici plus proche de la vérité.

Démonstration. On appelle range à l'instant t l'ensemble $\{X_s, s \leq t\}$. Notons que le range forme un processus croissant pour l'inclusion, le cardinal du range a des incréments égaux à 0 ou 1 à chaque instant. Lorsque la taille du range vaut k pour la première fois, alors le range correspond à un intervalle de longueur k et la marche se trouve à l'une des extrémités de ce range. Le temps d'attente du moment où le range vaudra $k+1$ est alors le temps d'atteinte

de $\{0, k + 1\}$ par la marche simple issue de 1, ce temps peut être évalué à l'aide du temps de transport, il vaut

$$\mathbb{E}[\tau_{\{0, k+1\}}] = \frac{1}{2} c_G \mathcal{R}(1 \leftrightarrow \{0, k + 1\}) = \frac{1}{2} \cdot 2(k + 1) \cdot \frac{1}{1 + \frac{1}{k}} = k$$

Ainsi

$$t_{\text{cov}} = \sum_{1 \leq k \leq n-1} k = \frac{n(n-1)}{2}.$$

□

Un autre graphe très simple où le calcul du temps de recouvrement est possible est le graphe complet (on a déjà étudié cette quantité en TD, sans mentionner alors qu'on calculait le temps de recouvrement du graphe complet, saurez-vous retrouver cet exercice?).

Le chapitre suivant, le dernier, est dévolu aux calculs nécessaires pour passer au tore de dimension supérieure, nettement plus délicats que dans le cas 1 dimensionnel.

Chapitre 5

Application : temps de couverture du tore

Nous avons maintenant tous les éléments en place d'un point de vue théorique pour comprendre les temps de couverture des tores d -dimensionnels de côté $n - 1$, qui sont la généralisation du n -cycle aux dimensions supérieures. Un obstacle de taille demeure néanmoins : des bornes quantitatives sur les résistances entre deux points arbitraires de ces graphes. Nous calculons ici de telles bornes, puis les estimées sur les temps d'atteinte et temps de couverture suivront. On notera que la précision des estimées concernant les résistances conditionne la précision des résultats suivants.

Définition 5.1 (tore et cube d -dimensionnel). On appelle tore d -dimensionnel (de côté $n - 1$) le graphe d'ensemble de sommets $V = \{1, \dots, n\}^d$ où deux sommets $x = (x_i)_{1 \leq i \leq d}$ et $y = (y_i)_{1 \leq i \leq d}$ sont adjacents ssi

$$\sum_{1 \leq i \leq d} |x_i - y_i| = 1 \quad (5.1)$$

avec la différence calculée dans $\mathbb{Z}/n\mathbb{Z}$. On appelle cube d -dimensionnel (de côté $n - 1$) le graphe induit par \mathbb{Z}^d sur l'ensemble de sommets $\{1, \dots, n\}^d$.

Dans le cas du cube, la définition de l'ensemble des arêtes est donc analogue à 5.1 mais la différence n'est pas prise dans $\mathbb{Z}/n\mathbb{Z}$. Si le tore d -dimensionnel est un graphe transitif, ce n'est pas le cas du cube d -dimensionnel. Notre objectif est d'étudier le tore mais nous aurons aussi besoin en chemin du cube.

Théorème 5.2 (Tores : temps d'atteinte). Soient x, y deux sommets du tore d -dimensionnel à distance $k \geq 1$ dans le tore de dimension d à n^d points. Le temps d'atteinte τ_y satisfait la propriété suivante : il existe des constantes $0 < c_d \leq C_d < +\infty$ telles que

$$\begin{aligned} c_2 n^2 \log(k) \leq \mathbb{E}_x[\tau_y] \leq C_2 n^2 \log(k) & \quad \text{si } d = 2 \\ c_d n^d \leq \mathbb{E}_x[\tau_y] \leq C_d n^d & \quad \text{si } d \geq 3. \end{aligned}$$

On notera que l'estimée ne dépend pas de d dans le cas de la dimension $d \geq 3$.

Démonstration. L'identité du temps de transport dans le cas transitif assure que

$$2\mathbb{E}_x[\tau_y] = \mathbb{E}_x[\tau_y] + \mathbb{E}_y[\tau_x] = c_G \mathcal{R}(x \leftrightarrow y) = 2dn^d \mathcal{R}(x \leftrightarrow y).$$

Pour une borne inférieure sur le résistance équivalente, on construit des cutsets d'arêtes deux à deux disjoints pour isoler x de y . On rappelle que $\|x\|_\infty = \max\{|x_i|, 1 \leq i \leq d\}$ définit la norme infinie.

$$\Pi_j = \{\{v, w\} \in V^2, \|v - x\|_\infty = j, \|w - x\|_\infty = j + 1\}.$$

Alors, pour $0 \leq j \leq \|y - x\|_\infty - 1 \leq k - 1$, Π_j est un cutset d'arêtes qui sépare x de y , et qui est de cardinal $2d(2j + 1)^{d-1}$. De plus, ces cutsets sont deux à deux disjoints, donc, par Nash-Williams,

$$\begin{aligned} \mathcal{R}(x \leftrightarrow y) &\geq \sum_{j=0}^{k-1} \left(\sum_{e \in \Pi_j} c(e) \right)^{-1} \\ &\geq \sum_{j=0}^{k-1} \frac{1}{2d(2j+1)^{d-1}} \\ &\geq \begin{cases} c_2 \log(k) & \text{si } d = 2 \\ c_d & \text{si } d \geq 3. \end{cases} \end{aligned}$$

Ainsi, on a finalement, pour des constantes c_1 et c_d différentes :

$$\mathbb{E}_x[\tau_y] \geq \begin{cases} c_2 n^2 \log(k) & \text{si } d = 2 \\ c_d n^d & \text{si } d \geq 3. \end{cases}$$

□

La borne supérieure est plus délicate, elle nécessite la construction de flots, qui eux mêmes nécessitent de comprendre l'urne de Polya tout d'abord ¹

Proposition 5.3 (Urne de Polya à d couleurs). *Soit une urne composée à l'instant $t = 0$ de d boules, dont 1 boule de chacune des d couleurs possibles. À chaque instant $t \geq 1$, on tire une boule choisie uniformément au hasard dans l'urne à l'instant $t - 1$, qu'on replace dans l'urne avec une boule de même couleur. La composition de l'urne forme alors une chaîne de Markov $(X_t)_{t \in \mathbb{N}} = (X_t(i), 1 \leq i \leq d)_{t \in \mathbb{N}}$ à valeurs dans $(\mathbb{N}^*)^d$, et pour chaque $t \in \mathbb{N}$, la loi de X_t est uniforme dans l'ensemble :*

$$\left\{ (y_i)_{1 \leq i \leq d} \in (\mathbb{N}^*)^d : \sum_{1 \leq i \leq d} y_i = d + t \right\}$$

et le cardinal de cet ensemble vaut $\binom{t+d-1}{d-1}$.

Démonstration. La démonstration se fait par exemple par récurrence. La propriété vaut en $t = 0$. On notera pour $x \in (\mathbb{N}^*)^d$ $x^i = (x_j^i)_{1 \leq j \leq d}$ le vecteur $x_j^i = x_j - \mathbf{1}_{\{i\}}(j)$. On suppose que

1. Cette méthode de preuve du théorème de Polya (qui concerne la récurrence/transience des graphes \mathbb{Z}^d) au moyen d'urnes de Polya est relativement récente : voir David A. Levin and Yuval Peres. "Pólya's theorem on random walks via Pólya's urn." *The American Mathematical Monthly* 117.3 (2010) : 220-231

la propriété vaut à l'instant $t - 1$ et on calcule comme suit :

$$\begin{aligned}
\mathbb{P}(X_t = x) &= \sum_{i=1}^d \mathbb{P}(X_{t-1} = x^i, X_t = x) \\
&= \sum_{i=1}^d \mathbb{P}(X_{t-1} = x^i) \frac{x^i(i)}{\sum_j x^i(j)} \\
&= \frac{1}{\binom{t-1+d-1}{d-1}} \sum_{i=1}^d \frac{x^i(i)}{\sum_j x^i(j)} \\
&= \frac{(d-1)!(t-2)!}{(t-1+d-1)!} \frac{t-1}{t+d-1} \\
&= \frac{(d-1)!(t-1)!}{(t+d-1)!} \\
&= \binom{t-1+d}{d-1}
\end{aligned}$$

Donc la propriété est vraie pour tout $t \in \mathbb{N}$. □

Noter que la démonstration par récurrence donne aussi la valeur du cardinal (on peut néanmoins obtenir ceci de manière directe, en notant qu'à chaque d -uplet y est associé de façon unique un chemin dit nord-est de $(0, 0)$ à $(d-1, t)$ dans \mathbb{Z}^2 , c'est-à-dire un chemin de longueur minimale $d-1+t$).

On propose dans le lemme suivant une borne supérieure sur la résistance entre des coins opposés d'un cube d -dimensionnel (et non d'un tore). Ce graphe n'est plus transitif en particulier.

Lemme 5.4. *Soit le cube d -dimensionnel (de côté $n-1$), et notons $\mathbf{1}$ le sommet de coordonnées $(1, \dots, 1)$, et $k \cdot \mathbf{1}$ celui de coordonnées (k, \dots, k) . Soit k tel que $kd < n-1$. Alors*

$$\mathcal{R}(\mathbf{1} \leftrightarrow k \cdot \mathbf{1}) \leq \begin{cases} 2 \log(k) & \text{si } d = 2 \\ \tilde{C}_d & \text{si } d \geq 3. \end{cases}$$

On notera l'hypothèse sur k dont a besoin dans la construction du flot.

Démonstration. On construit un flot unitaire θ de $\mathbf{1}$ à $k \cdot \mathbf{1}$ sur ce graphe. Pour cela, on va utiliser une urne de Polya (et le flot qui lui est naturellement associé plus précisément). La distance de $\mathbf{1}$ à $n \cdot \mathbf{1}$ vaut $(k-1)d$, et on considère l'hyperplan $\{x : \sum x_i = (k-1)d/2\}$.

Posons : $\vec{E}_\ell^+ = \{\vec{e} \in \vec{E}, \|\vec{e}_-\|_1 - 1 = \|\vec{e}_+\|_1 \leq \ell\}$. On définit le flot sur $\vec{E}_{(k-1)d/2}$ en posant

$$\theta(\vec{e}) = \mathbb{P} \left(\vec{e} \in \bigcup_{t \geq 0} \{X_t, X_{t+1}\} \right).$$

Ensuite, la règle d'antisymétrie définit le flot sur les arêtes dirigées opposées. Enfin,

$$\theta(k\mathbf{1} - \vec{e}) = -\theta(\vec{e})$$

plus la règle d'antisymétrie achèvent de définir le flot. Maintenant,

$$\begin{aligned}
\mathcal{E}(\theta) &= \sum_e r(e)\theta(e)^2 \\
&= 2 \sum_{k=0}^{n-2} \sum_{\substack{e=\{x,y\} \in E \\ x \text{ à distance } k \text{ de } a \\ y \text{ à distance } k+1 \text{ de } a}} \theta(e)^2 \\
&\leq 2 \sum_{k=0}^{n-2} \binom{k+d-1}{d-1} \cdot \left(\binom{k+d-1}{d-1}^{-1} \right)^2 \\
&\leq 2 \sum_{k=0}^{n-2} \binom{k+d-1}{d-1}^{-1} \\
&\leq \begin{cases} 2 \log(n) & \text{si } d = 2 \\ \tilde{C}_d & \text{si } d \geq 3. \end{cases}
\end{aligned}$$

□

$x = (x_1, x_2, \dots, x_d)$ et $y = (y_1, y_2, \dots, y_d)$ forment un pavé de dimension d tel que, pour tout $1 \geq i \geq d$, $|y_i - x_i| \geq k$.

Pour tout $1 \geq i \geq d$, il existe s_i tel que $|y_i - x_i| = 2s_i$ ou $2s_i + 1$ selon la parité de $|y_i - x_i|$. Par l'inégalité triangulaire, on a

$$\begin{aligned}
\mathcal{R}(x \leftrightarrow y) &\leq \mathcal{R}((x_1, x_2, \dots, x_d) \leftrightarrow (y_1, x_2, \dots, x_d)) \\
&\quad + \mathcal{R}((y_1, x_2, \dots, x_d) \leftrightarrow (y_1, y_2, \dots, x_d)) \\
&\quad + \dots \\
&\quad + \mathcal{R}((y_1, y_2, \dots, x_d) \leftrightarrow (y_1, y_2, \dots, y_d))
\end{aligned}$$

Or, la résistance équivalente de deux points, qui sont dans le même hyperplan, à distance paire valant $2s$, est majorée, par l'inégalité triangulaire, par 2 fois la résistance équivalente entre les deux extrémités d'un G_s^d .

Donc,

$$\begin{aligned}
\mathcal{R}(x \leftrightarrow y) &\leq \sum_{i=1}^d 1 + 2\mathcal{R}(a \leftrightarrow z, \text{ où } a \text{ et } z \text{ sont les extrémités d'un } G_{s_i}^d) \\
&\leq \begin{cases} \sum_{i=1}^2 1 + 4 \log(s_i) & \text{si } d = 2 \\ \sum_{i=1}^d 1 + 2\tilde{C}_d & \text{si } d \geq 3. \end{cases} \\
&\leq \begin{cases} 2(1 + 4 \log(k+1)) & \text{si } d = 2 \\ d(1 + 2\tilde{C}_d) & \text{si } d \geq 3. \end{cases} \\
&\leq \begin{cases} \frac{C_2}{2} \log(k+1) & \text{si } d = 2 \\ \frac{C_d}{d} & \text{si } d \geq 3. \end{cases}
\end{aligned}$$

Donc, si $d = 2$,

$$\mathbb{E}_x[\tau_y] \leq C_2 n^2 \log(k+1)$$

et si

$$\mathbb{E}_x[\tau_y] \leq C_d n^d.$$

■

Nous en venons finalement au théorème sur les temps de couverture du tore d -dimensionnel, qui conclut ce cours.

Theorème 5.5 (Temps de couverture du tore d -dimensionnel). *Le temps de couverture du tore d -dimensionnel satisfait : il existe des constantes $0 < c_d \leq C_d < +\infty$ telles que*

$$\begin{aligned} \frac{c_2}{2} n^2 (\log(n))^2 (1 + o(1)) &\leq t_{\text{cov}} \leq 2C_2 n^2 (\log(n))^2 (1 + o(1)) && \text{si } d = 2 \\ c_d n^d \log(n) (1 + o(1)) &\leq t_{\text{cov}} \leq C_d n^d \log(n) (1 + o(1)) && \text{si } d \geq 3. \end{aligned}$$

Démonstration. Les bornes supérieures découlent directement de la borne supérieure de Matthews, et des estimées sur les temps d'atteinte. Des précisions doivent en revanche être apportées sur la borne inférieure : il s'agit de choisir un bon ensemble de sommets A d'après 4.10 et pour $d \geq 3$, on prend simplement pour A l'ensemble des sommets, $A = V$, et il suit :

$$\begin{aligned} t_{\text{cov}} &\geq t_{\min}^A \left(1 + \frac{1}{2} + \dots + \frac{1}{|A| - 1} \right) \\ &\geq c_d n^d \left(1 + \frac{1}{2} + \dots + \frac{1}{n^d - 1} \right) \\ &\geq c_d n^d \log(n) \end{aligned}$$

Pour $d = 2$ maintenant, il existe des constantes $0 < c_2 \leq C_2 < +\infty$ telles que pour $x, y \in V$, si l'on note $k = d(x, y)$, on a

$$c_2 n^2 \log(k) \leq \mathbb{E}_x[\tau_y] \leq C_2 n^2 \log(k).$$

On doit être cette fois plus fin dans notre choix de A . Considérons d'abord le cas où n est un carré parfait, alors prenant pour A l'ensemble des sommets dont les coordonnées sont multiples de \sqrt{n} , on trouve :

$$\begin{aligned} t_{\text{cov}} &\geq t_{\min}^A \left(1 + \frac{1}{2} + \dots + \frac{1}{|A| - 1} \right) \\ &\geq c_2 n^2 \log(\sqrt{n}) \left(1 + \frac{1}{2} + \dots + \frac{1}{n - 1} \right) \\ &\geq \frac{c_2}{2} n^2 (\log(n))^2 (1 + o(1)). \end{aligned}$$

Si n n'est pas un carré parfait, alors on observe que le plus grand carré parfait inférieur à n est minoré par $n - 2\sqrt{n}$: en effet $m \leq \sqrt{n} < m + 1$ implique $\sqrt{n} - 1 < m \leq \sqrt{n}$ et $n - 2\sqrt{n} + 1 < m^2 \leq n$ et le même résultat vaut donc. □