

Temps de mélange pour les chaînes de Markov

Polycopié pour le cours de M1 Maths Générales, Université
Grenoble Alpes, année 2022/2023

Hugo Vanneuille

Ce polycopié est une version préliminaire des notes de cours, qui comporte donc probablement un certain nombre de (petites ?) erreurs. Il contient toutes les définitions, résultats et preuves du cours, mais pas forcément toutes les motivations, intuitions etc proposées pendant le cours.

Commençons par introduire de façon précise le formalisme que nous allons adopter pendant tout le cours :

- On se donne un ensemble dénombrable V et on note $\text{Pr}(V)$ l'ensemble des mesures de probabilités sur $(V, \mathcal{P}(V))$.¹
- On se donne aussi une matrice de transition **irréductible** $P = (p(x, y))_{x, y \in V}$ sur V .
- Par ailleurs, on se place sur un espace mesurable (Ω, \mathcal{F}) muni de mesures de probabilités \mathbb{P}_μ , pour toute $\mu \in \text{Pr}(V)$.
- Enfin, on considère un processus $X = (X_n)_{n \geq 0}$ sur (Ω, \mathcal{F}) à valeurs dans V ² et on suppose que, pour toute $\mu \in \text{Pr}(V)$:

Sous \mathbb{P}_μ , X est une chaîne de Markov de matrice de transition P et de loi initiale μ .

(Autrement dit, $X : (\Omega, \mathcal{F}, \mathbb{P}_\mu) \rightarrow (V, \mathcal{P}(V))$ est une chaîne de Markov de matrice de transition P et de loi initiale μ .)

On note \mathbb{E}_μ l'espérance associée à \mathbb{P}_μ et on note $\mathbb{P}_x = \mathbb{P}_{\delta_x}$ et $\mathbb{E}_x = \mathbb{E}_{\delta_x}$.

L'existence des objets ci-dessus est admise dans ce cours. Ce formalisme permet de considérer constamment le même processus X sans devoir en introduire un autre à chaque fois qu'on considère une mesure initiale différente, et ainsi pouvoir énoncer de façon assez compacte des propriétés pour n'importe quelle mesure initiale. Mais c'est un

1. Dans le reste du cours, on dira juste "mesures de probabilités sur V " sans spécifier qu'on munit tout ensemble dénombrable de la tribu de l'ensemble des parties.

2. Cela signifie que pour tout $n \geq 0$, X_n est une application mesurable de (Ω, \mathcal{F}) dans V , où on munit toujours V de la tribu $\mathcal{P}(V)$.

choix de formalisme parmi d'autres. On peut avoir à l'esprit que, comme la loi d'une chaîne de Markov est caractérisée par sa matrice de transition et sa loi initiale, on a le fait suivant : Si Y est une chaîne de Markov de matrice de transition P définie sur un espace de probabilité quelconque $(\Omega', \mathcal{F}', \mathbb{P})$, si par ailleurs μ est la loi initiale de Y , et si $F : V^{\mathbb{N}} \rightarrow \mathbb{R}$ est une fonction mesurable positive ou bornée³, alors

$$\mathbb{E}[F(Y)] = \mathbb{E}_{\mu}[F(X)].$$

(Car la loi de Y est égale à la loi de X sous \mathbb{P}_{μ} .)

Table des matières

| | | |
|----------|---|-----------|
| 1 | Section préliminaire sous des hypothèses plus générales : le théorème ergodique | 2 |
| 2 | Le théorème de Perron–Frobenius | 7 |
| 3 | Qu'est-ce que le temps de mélange ? Définitions et premières propriétés | 9 |
| 4 | Un exemple phare pour le cours : des marches aléatoires sur des graphes (et des “méthodes géométriques”) | 18 |
| 5 | Le couplage (ou “méthode probabiliste”) | 23 |
| 6 | Le temps de relaxation (ou “méthode spectrale”) | 32 |
| 7 | Références (et lien vers des “méthodes fonctionnelles”) | 42 |

1 Section préliminaire sous des hypothèses plus générales : le théorème ergodique

Le but de ce cours est d'étudier des théorèmes de convergence pour des chaînes de Markov. Avant d'entrer dans le vif du sujet, rappelons qu'on suppose P irréductible pendant tout le cours, et rappelons un résultat central de la théorie des chaînes de Markov.

Théorème 1.1. *Si P est récurrente positive, alors elle admet une unique mesure de probabilité invariante π . Par ailleurs, pour tout $x \in V$, $\pi(x) > 0$.*

Énonçons maintenant un premier théorème de convergence, le théorème ergodique. D'après ce théorème, moyenne temporelle et moyenne selon π sont asymptotiquement les mêmes, et ce quelle que soit la mesure initiale.

3. Où $V^{\mathbb{N}}$ est équipé de la tribu produit et \mathbb{R} de la tribu borélienne.

Théorème 1.2. *Supposons que P est récurrente positive et notons π l'unique mesure de probabilité invariante. Alors, pour toute $\mu \in \text{Pr}(V)$ et toute fonction mesurable $f : V \rightarrow \mathbb{R}$ bornée,*

$$\mathbb{P}_\mu \left[\frac{1}{n} \sum_{k=0}^n f(X_k) \xrightarrow{n \rightarrow +\infty} \int f(x) \pi(dx) \right] = 1.$$

Avant de démontrer le théorème, rappelons la proposition suivante au sujet des chaînes récurrentes positives.

Proposition 1.3. *Supposons que P est récurrente positive et donnons-nous un point $x \in V$. Notons par ailleurs R_1^x le premier temps de retour en x , i.e.*

$$R_1^x = \inf\{n \geq 1 : X_n = x\} \in \mathbb{N}^* \cup \{+\infty\}.$$

Alors, $\mathbb{E}_x[R_1^x] < +\infty$, et l'unique mesure de probabilité invariante π vérifie

$$\pi(x) = 1/\mathbb{E}_x[R_1^x].$$

De plus, la mesure ν_x définie de la façon suivante est invariante :

$$\nu_x(y) = \mathbb{E} \left[\sum_{k=0}^{R_1^x-1} 1_{\{X_k=y\}} \right].$$

Démonstration du Théorème 1.2. Notons C (pour ‘‘Convergence’’) l'événement de convergence dans l'énoncé du théorème. D'après la propriété de Markov simple (appliquée au temps $n = 0$),

$$\mathbb{P}_\mu[C] = \sum_{x \in V} \mu(x) \mathbb{P}_x[C].$$

Il suffit donc de démontrer le théorème dans le cas où μ est une mesure de Dirac. Fixons donc un point $x \in V$ et plaçons-nous sous la mesure de probabilité \mathbb{P}_x (ainsi, ‘‘p.s.’’ signifiera toujours ‘‘ \mathbb{P}_x -p.s.’’ dans cette preuve). Pour cela, nous allons appliquer la loi des grands nombres à une suite de variables aléatoires i.i.d. et intégrables. Nous allons construire ces variables à l'aide des temps d'arrêt R_n^x qui sont les temps de $n^{\text{ième}}$ retour en x , i.e.

$$R_0^x = 0, \quad R_{n+1}^x = \inf\{k \geq R_n^x + 1 : X_k = x\}.$$

Notons que, comme la chaîne est récurrente, ces temps sont tous finis p.s.

Lemme 1.4. *La suite $(Z_n(f))_{n \geq 0}$ définie par*

$$Z_n(f) = \sum_{k=R_n^x}^{R_{n+1}^x-1} f(X_k)$$

est i.i.d., et $\mathbb{E}_x[|Z_0(f)|] < +\infty$.

Démonstration. Pour montrer que cette suite est i.i.d., il suffit de montrer que pour tout $n \geq 0$ et tout choix de fonctions mesurables bornées $F_0, \dots, F_n : \mathbb{R} \rightarrow \mathbb{R}$ (où on munit \mathbb{R} de la tribu borélienne), on a

$$\mathbb{E}_x \left[\prod_{k=0}^n F_k(Z_k(f)) \right] = \prod_{k=0}^n \mathbb{E}_x [F_k(Z_0(f))]. \quad (1)$$

Afin de démontrer cette égalité, nous allons appliquer la propriété de Markov forte, qui donne en effet des propriétés d'indépendance et d'égalité en loi. Pour cela, on remarque tout d'abord que – comme la chaîne est récurrente – l'événement $\{X_{R_n^x} = x\} := \{R_n^x < +\infty \text{ et } X_{R_n^x} = x\}$ est réalisé presque-sûrement. On peut donc écrire :

$$\mathbb{E}_x \left[\prod_{k=0}^n F_k(Z_k(f)) \right] = \mathbb{E}_x \left[\prod_{k=0}^n F_k(Z_k(f)) \mid X_{R_n^x} = x \right]$$

On remarque par ailleurs que les variables R_k^x sont des temps d'arrêt et que $Z_k(f)$ est $\mathcal{F}_{R_{k+1}^x}$ -mesurable pour tout k .⁴ Cela implique que $Z_k(f)$ est $\mathcal{F}_{R_n^x}$ -mesurable pour tout $k \in \{0, \dots, n-1\}$. En effet, $R_{k+1}^x \leq R_n^x$ pour tout $k \leq n-1$, donc $\mathcal{F}_{R_{k+1}^x} \subset \mathcal{F}_{R_n^x}$ pour ces mêmes entiers k . On en déduit que

$$W_{n-1} := \prod_{k=0}^{n-1} F_k(Z_k(f))$$

est $\mathcal{F}_{R_n^x}$ -mesurable. Notons maintenant

$$G_n : V^{\mathbb{N}} \rightarrow \mathbb{R}$$

la fonction mesurable (où on munit $V^{\mathbb{N}}$ de la tribu produit et \mathbb{R} de la tribu borélienne) définie par

$$G_n(u) = 1_{\{R^x(u) < +\infty\}} F_n \left(\sum_{k=0}^{R^x(u)-1} f(u_k) \right),$$

où

$$R^x : V^{\mathbb{N}} \rightarrow \mathbb{N}^* \cup \{+\infty\}$$

est la fonction mesurable définie par

$$R^x((u_k)_{k \geq 0}) = \inf \{k \geq 1 : u_k = x\}.$$

4. On démontre cette deuxième affirmation en écrivant par exemple que, pour tout ensemble mesurable $A \subset \mathbb{R}$,

$$\{Z_n(f) \in A\} \cap \{R_{n+1}^x = m\} = \bigcup_{\ell=0}^{m-1} \left\{ \sum_{k=\ell}^{m-1} f(X_k) \in A \right\} \cap \{X_\ell = x, X_{\ell+1} \neq x, \dots, X_{m-1} \neq x\} \cap \{R_{n+1}^x = m\} \in \mathcal{F}_m^X.$$

Notons enfin $Y = (Y_k)_{k \geq 0}$ le processus défini par

$$Y_k = X_{R_n^x + k}.$$

On remarque que $F_n(Z_n(f)) = G_n(Y)$ p.s. (notons qu'afin d'affirmer cela, on a utilisé que, sous \mathbb{P}_x , les R_k^x sont tous finis p.s.). Par conséquent,

$$\prod_{k=0}^n F_k(Z_k(f)) = W_{n-1} G_n(Y).$$

Or, d'après la propriété de Markov forte, sous $\mathbb{P}_x[\cdot \mid X_{R_n^x} = x]$, Y est indépendante de $\mathcal{F}_{R_n^x}$, et la loi de Y est égale à la loi de X sous \mathbb{P}_x . Donc

$$\begin{aligned} \mathbb{E}_x \left[\prod_{k=0}^n F_k(Z_k(f)) \mid X_{R_n^x} = x \right] &= \mathbb{E}_x \left[W_{n-1} \mid X_{R_n^x} = x \right] \mathbb{E}_x \left[G_n(X) \right] \\ &= \mathbb{E}_x \left[W_{n-1} \mid X_{R_n^x} = x \right] \mathbb{E}_x \left[F_n(Z_0(f)) \right] = \mathbb{E}_x \left[W_{n-1} \right] \mathbb{E}_x \left[F_n(Z_0(f)) \right]. \end{aligned}$$

Un raisonnement par récurrence implique donc (1).

Il reste à montrer que $\mathbb{E}_x[|Z_0(f)|] < +\infty$. Pour cela, on remarque que

$$\mathbb{E}_x[|Z_0(f)|] \leq \|f\|_\infty \mathbb{E}_x[R_1^x],$$

qui est bien fini d'après la Proposition 1.3. □

D'après le lemme précédent et la loi des grands nombres,

$$\frac{1}{n} \sum_{k=0}^n Z_k(f) \xrightarrow[n \rightarrow +\infty]{} \mathbb{E}_x[Z_0(f)] \quad p.s.$$

Or,

$$\mathbb{E}_x[Z_0(f)] = \mathbb{E}_x \left[\sum_{k=0}^{R_1^x-1} \sum_{y \in V} 1_{\{X_k=y\}} f(y) \right] = \int f(y) \nu_x(dy),$$

où ν_x est la mesure définie dans la Proposition 1.3. Or, la mesure invariante est unique à constante multiplicative près. Comme $\nu_x(x) = 1$, on en déduit que $\pi(y) = \pi(x) \nu_x(y)$ pour tout $y \in V$. Par conséquent, $\int f(y) \nu_x(dy) = \frac{1}{\pi(x)} \int f d\pi$. On a donc montré que

$$\frac{1}{n} \sum_{k=0}^n Z_k(f) \xrightarrow[n \rightarrow +\infty]{} \frac{1}{\pi(x)} \int f d\pi \quad p.s. \quad (2)$$

Concluons maintenant la preuve du théorème ergodique. On remarque tout d'abord que, comme on peut écrire f comme la somme de sa partie positive et sa partie négative, il suffit de montrer le résultat dans le cas où f est positive. On se place dans ce cas et on note

$$N_x(n) = \#\{m \in \{1, \dots, n\} : X_m = x\}$$

et on observe que $R_{N_x(n)}^x \leq n < R_{N_x(n)+1}^x$. On a donc (c'est ici qu'on utilise que f est positive) :

$$\frac{1}{N_x(n)} \sum_{k=0}^{R_{N_x(n)}^x - 1} f(X_k) \leq \frac{1}{N_x(n)} \sum_{k=0}^n f(X_k) \leq \frac{1}{N_x(n)} \sum_{k=0}^{R_{N_x(n)+1}^x - 1} f(X_k),$$

c'est-à-dire :

$$\frac{N_x(n) - 1}{N_x(n)} \frac{1}{N_x(n) - 1} \sum_{k=0}^{N_x(n) - 1} Z_k(f) \leq \frac{1}{N_x(n)} \sum_{k=0}^n f(X_k) \leq \frac{1}{N_x(n)} \sum_{k=0}^{N_x(n)} Z_k(f),$$

Par (2), les termes tout à gauche et tout à droite de la série d'inégalités ci-dessus convergent p.s. vers

$$\frac{1}{\pi(x)} \int f d\pi,$$

et donc

$$\frac{1}{N_x(n)} \sum_{k=0}^n f(X_k) \xrightarrow[n \rightarrow +\infty]{} \frac{1}{\pi(x)} \int f d\pi \quad p.s.$$

Si on applique cela à $f = 1$, on obtient que

$$\frac{n+1}{N_x(n)} \xrightarrow[n \rightarrow +\infty]{} \frac{1}{\pi(x)} \quad p.s.$$

Le théorème ergodique est une conséquence des deux convergences ci-dessus. Cela conclut la démonstration. \square

En appliquant le théorème ergodique à $f(y) = 1_{y=x}$, on obtient le corollaire suivant.

Corollaire 1.5. *Sous les mêmes hypothèses que le théorème ergodique et pour toute $\mu \in \text{Pr}(V)$ et tout $x \in V$,*

$$\mathbb{P}_\mu \left[\frac{1}{n} \sum_{k=0}^n 1_{\{X_k=x\}} \xrightarrow[n \rightarrow +\infty]{} \pi(x) \right] = 1.$$

Par le théorème de convergence dominée, on en déduit que

$$\frac{1}{n} \sum_{k=0}^n \mathbb{P}_\mu[X_k = x] \xrightarrow[n \rightarrow +\infty]{} \pi(x)$$

pour tout $x \in V$ et toute $\mu \in \text{Pr}(V)$.

2 Le théorème de Perron–Frobenius

Dans cette section, on se demande si la dernière partie du Corollaire 1.5 peut être vraie sans avoir à passer par la moyenne de Cesàro. Plus précisément : a-t-on

$$\mathbb{P}_\mu[X_n = x] \xrightarrow{n \rightarrow +\infty} \pi(x) ?$$

On peut remarquer sur un exemple particulier que ce n'est pas vrai en règle générale : étudions par exemple la marche aléatoire sur un cercle de cardinal pair, c'est-à-dire une chaîne de Markov sur $\mathbb{Z}/(N\mathbb{Z})$ pour un certain N pair (supérieur ou égal à 4, disons), et dont la matrice de transition P est définie par $p(x, y) = 1/2$ si $y = x \pm 1$ et 0 sinon. On peut montrer que la mesure de probabilité stationnaire est la mesure de probabilité uniforme et que $\mathbb{P}_0[X_n = 0] = 0$ pour tout n impair, ce qui rend impossible la convergence désirée (dans le cas $\mu = \delta_0$).

Ajoutons donc une hypothèse pour rendre ce genre de chose impossible. Peut-être que l'utilisation d'un argument de parité ci-dessus rend naturelle que nous allons utiliser la notion d'**apériodicité**. Nous allons par ailleurs nous placer dans le cas V fini (voir la section sur les couplages pour une remarque sur le cas infini). Le **théorème de Perron–Frobenius** est le résultat suivant.

Théorème 2.1. *Supposons que V est fini et que P est apériodique, et notons π l'unique mesure de probabilité invariante.*⁵ *Alors, pour toute $\mu \in \text{Pr}(V)$ et tout $x \in V$,*

$$\mathbb{P}_\mu[X_n = x] \xrightarrow{n \rightarrow +\infty} \pi(x).$$

Ce théorème est le point de départ de cette partie du cours. Ce résultat est particulièrement utile quand on veut simuler une loi π : si cette loi est la loi stationnaire d'une chaîne de Markov, alors on peut la simuler en faisant évoluer cette chaîne pendant un grand nombre d'étapes. Mais que signifie grand ? Autrement dit, quelle est la vitesse de convergence dans le théorème ci-dessus ? Avant de parler de vitesse de convergence, il nous faut définir ce que signifie qu'une loi est proche d'une autre. On utilisera la notion de distance en variation totale, qui fait l'objet de la section suivante.

Mais démontrons tout d'abord le théorème de Perron–Frobenius, et, pour cela, commençons par rappeler le lemme suivant.

Lemme 2.2. *Supposons V fini et donnons-nous une matrice de transition irréductible $Q = (q(x, y))_{x, y \in V}$ sur V . Q est apériodique si et seulement si elle satisfait une des deux propriétés équivalentes suivantes :*

- $\exists n \geq 1, \forall x, y \in V, q_n(x, y) \geq 0$;
- $\exists n_0 \geq 1, \forall x, y \in V, \forall n \geq n_0, q_n(x, y) \geq 0$.

Profitons de cette parenthèse sur les critères d'apériodicité pour énoncer le critère suffisant suivant.

5. Ici, on utilise l'implication (V fini et P irréductible) \Rightarrow P récurrente positive.

Lemme 2.3. *Supposons V fini et donnons-nous une matrice de transition irréductible $Q = (q(x, y))_{x, y \in V}$ sur V . S'il existe $x_0 \in V$ tel que $q(x_0, x_0) > 0$, alors Q est apériodique.*

Démonstration. Le pgcd de $\{n \geq 1 : q_n(x_0, x_0) > 0\}$ est 1 (car 1 appartient à cet ensemble). \square

Donnons une preuve algébrique du théorème de Perron–Frobenius (plus tard dans le cours, nous donnerons une preuve de nature plus probabiliste). Notons $\Pi = (\Pi(x, y))_{x, y \in V}$ la matrice de transition dont toutes les lignes sont égales à π , i.e.

$$\forall x, y \in V, \quad \Pi(x, y) = \pi(y).$$

Démonstration du Théorème 2.1. D'après la propriété de Markov simple (appliquée au temps 0),

$$\mathbb{P}_\mu[X_n = y] = \sum_{x \in V} \mu(x) p_n(x, y).$$

Il est donc suffisant de montrer que, pour tous $x, y \in V$,

$$p_n(x, y) - \pi(y) \xrightarrow{n \rightarrow +\infty} 0. \quad (3)$$

Comme P est apériodique, on peut considérer $m \geq 1$ tel que toutes les entrées de P^m sont strictement positive (cf. le Lemme 2.2). Fixons un tel m . Comme V est fini, ce choix de m implique qu'il existe $\delta > 0$ tel que, pour tous $x, y \in V$,

$$p_m(x, y) > \delta \pi(y).$$

Définissons maintenant une matrice Q par

$$Q = \frac{1}{1 - \delta} (P^m - \delta \Pi).$$

On remarque que Q est une matrice stochastique. En effet, le facteur $1/(1 - \delta)$ est présent afin que la somme sur chaque ligne soit égale à 1, et le choix de δ implique que les coefficients de Q sont tous positifs.

Montrons par récurrence que, pour tout $q \geq 1$,

$$P^{mq} = (1 - \theta^q) \Pi + \theta^q Q^q, \quad (4)$$

où $\theta = 1 - \delta$. Le cas $q = 1$ vient de la définition de Q . Si on suppose que la propriété est vraie à un certain rang q , on obtient que

$$P^{m(q+1)} = (1 - \theta^q) \Pi P^m + \theta^q Q^q P^m.$$

Or, π étant invariante pour P , on a $\pi P^m = \pi P^{m-1} = \dots = \pi$, donc $\Pi P^m = \Pi$. Par ailleurs, en utilisant le cas $q = 1$, on obtient que

$$\theta^q Q^q P^m = \theta^q Q^q ((1 - \theta) \Pi + \theta Q) = \theta^q (1 - \theta) Q^q \Pi + \theta^{q+1} Q^{q+1}.$$

Q^q étant une matrice de transition et les colonnes de Π étant constantes, on a $Q^q\Pi = \Pi$. On obtient donc finalement que

$$P^{m(q+1)} = (1 - \theta^q)\Pi + \theta^q(1 - \theta)\Pi + \theta^{q+1}Q^{q+1} = (1 - \theta^{q+1})\Pi + \theta^{q+1}Q^{q+1},$$

ce qui conclut la preuve de (4) par récurrence.

Terminons maintenant la preuve grâce à (4). Soit $n \geq m$ et soit

$$n = mq + r$$

la division euclidienne de n par m . La quantité $p_n(x, y) - \pi(y)$ est le coefficient d'indice (x, y) de la matrice $P^n - \Pi = P^r(P^{mq} - \Pi)$. Or, les coefficients de P^r sont bornés (car c'est une matrice de transition), et les coefficients de $P^{mq} - \Pi$ convergent vers 0 quand q tend vers $+\infty$ (i.e. quand n tend vers $+\infty$) d'après (4), car les coefficients de Q^q sont bornés (puisque c'est une matrice de transition) et car $\theta < 1$. Cela implique (3) et termine la démonstration. \square

3 Qu'est-ce que le temps de mélange ? Définitions et premières propriétés

3.1 Le contexte à partir de maintenant

À partir de maintenant et jusqu'à la fin du poly, on suppose que V est **fini** et possède au moins deux éléments (i.e. $|V| := \text{Card}(V) \geq 2$). Par conséquent, P est **une matrice de transition irréductible sur un ensemble fini**. P est donc **récurrente positive** et admet **une unique mesure de probabilité invariante** qu'on continue à noter :

$$\pi.$$

La notation π pour cette mesure de probabilité invariante sera utilisée dans tout le cours.

Par ailleurs, on identifie l'ensemble des mesures de probabilité sur V avec

$$\text{Pr}(V) = \{\mu \in [0, 1]^V \subset \mathbb{R}^V : \sum_{x \in V} \mu(x) = 1\}.$$

Ainsi, si on se donne un vecteur $\mu \in [0, 1]^V$ tel que $\sum_{x \in V} \mu(x) = 1$, on note aussi μ la probabilité sur V définie par

$$\forall A \subset V, \quad \mu(A) = \sum_{x \in A} \mu(x).$$

De plus, on note

$$p_n(x, y) = P^n(x, y)$$

pour tout $n \geq 0$.

Souvenons-nous que, sous \mathbb{P}_μ , la loi de X_n est μP^n . En particulier, sous \mathbb{P}_x , la loi de X_n est $\delta_x P^n = p_n(x, \cdot)$. On fera de nombreux aller-retours entre notations algébriques et concepts probabilistes. Il est important de continuer à s'habituer aux écritures algébriques. On peut par exemple méditer sur les deux écritures suivantes :

(i) On peut calculer $\mathbb{P}_\mu[X_n \in A] = \mu P^n(A)$ en écrivant

$$\forall A \subset V, \quad \mu P^n(A) = \sum_{y \in A} \mu P^n(y) = \sum_{y \in A} \sum_{x \in V} \mu(x) p_n(x, y).$$

(ii) On peut aussi écrire :

$$\mu P^n = \sum_{x \in V} \mu(x) p_n(x, \cdot), \quad \text{donc} \quad \forall A \subset V, \quad \mu P^n(A) = \sum_{y \in A} \sum_{x \in V} \mu(x) p_n(x, y).$$

Du fait du théorème de Perron–Frobenius, on va s'intéresser au cas où P est apériodique.

À partir de maintenant et jusqu'à la fin du poly, on suppose aussi que P est apériodique. Il est important d'avoir à l'esprit que de nombreuses propriétés que nous étudierons ne seraient pas vraies sans cette hypothèse.

3.2 La distance en variation totale et le temps de mélange

Définissons la **distance en variation totale** entre deux mesures de probabilité.

Définition 3.1. Soient $\mu, \nu \in \text{Pr}(V)$. La distance en variation totale entre μ et ν est

$$\|\mu - \nu\|_{\text{VT}} := \max_{A \subset V} |\mu(A) - \nu(A)|.$$

Exemple 3.2. (i) Si $x, y \in V$ et $x \neq y$, alors $\|\delta_x - \delta_y\|_{\text{VT}} = 1$.

(ii) Si $V = \{0, 1\}$ et μ_p est la loi de Bernoulli de paramètre p , alors $\|\mu_p - \mu_q\|_{\text{VT}} = |p - q|$.

L'écriture sous forme de norme est justifiée par la définition suivante.

Définition 3.3. Soit $\mu \in \mathbb{R}^V$. La norme en variation totale de μ est

$$\|\mu\|_{\text{VT}} := \max_{A \subset V} \left| \sum_{x \in A} \mu(x) \right|.$$

$\|\cdot\|_{\text{VT}}$ est bien une norme sur \mathbb{R}^V . Par conséquent, la distance en variation totale est bien une distance sur $\text{Pr}(V)$.

On peut maintenant définir le temps de mélange, qui est l'objet central de cette partie du cours. On étudiera – ce qui voudra souvent dire trouver des bornes inf et sup – cet objet à l'aide de techniques diverses : probabilistes, géométriques, spectrales, fonctionnelles...

Définition 3.4. Pour tout $\varepsilon \in (0, 1/2)$, le **temps de mélange** est

$$t_{\text{mél}}(\varepsilon) = \inf \{n \geq 0 : \max_{\mu \in \text{Pr}(V)} \|\mu P^n - \pi\|_{\text{VT}} \leq \varepsilon\}.$$

Pour simplifier les notations, pour tout $n \geq 0$ on va noter

$$\mathcal{D}(n) = \max_{\mu \in \text{Pr}(V)} \|\mu P^n - \pi\|_{\text{VT}},$$

de telle façon que

$$t_{\text{mél}}(\varepsilon) = \inf \{n \geq 0 : \mathcal{D}(n) \leq \varepsilon\}.$$

Remarque 3.5. On va voir dans très peu de temps (cf. le Théorème 3.8) que le Théorème 2.1 implique que $\mathcal{D}(n) \rightarrow 0$, ce qui implique que $t_{\text{mél}}(\varepsilon) < +\infty$ pour tout $\varepsilon \in (0, 1/2)$.

On pourrait bien sûr aussi regarder le cas $\varepsilon \in [1/2, 1)$, mais se restreindre à $(0, 1/2)$ va nous aider à simplifier certains calculs. D'après la remarque ci-dessous, cela implique que $t_{\text{mél}}(\varepsilon)$ est toujours au moins égal à 1. Cela nous permettra de considérer le temps $n = t_{\text{mél}}(\varepsilon) - 1$.

Remarque 3.6. On peut remarquer que $t_{\text{mél}}(\varepsilon) \geq 1$. En effet, comme $|V| \geq 2$, il existe $x \in V$ tel que $\pi(x) \leq 1/2$. Fixons un tel x . On remarque que $1/2 \leq \|\delta_x - \pi\|_{\text{VT}} = \|\delta_x P^0 - \pi\|_{\text{VT}}$, donc $\mathcal{D}(0) \geq 1/2 > \varepsilon$, et $t_{\text{mél}}(\varepsilon) \geq 1$.

On va commencer par remarquer qu'on peut se restreindre au cas où $(X_n)_{n \geq 0}$ part d'un point fixé, i.e. au cas où la loi initiale μ est égale à δ_x pour un certain $x \in V$:

Lemme 3.7. Pour tout $n \geq 0$, $\mathcal{D}(n) = \max_{x \in V} \|p_n(x, \cdot) - \pi\|_{\text{VT}}$.

Démonstration. $p_n(x, \cdot) = \delta_x P^n$ donc $\max_x \|p_n(x, \cdot) - \pi\|_{\text{VT}} \leq \mathcal{D}(n)$. Pour montrer l'inégalité inverse, on montre que $\|\mu P^n - \pi\|_{\text{VT}} \leq \max_x \|p_n(x, \cdot) - \pi\|_{\text{VT}}$ pour toute $\mu \in \text{Pr}(V)$ de la façon suivante :

$$\begin{aligned} \|\mu P^n - \pi\|_{\text{VT}} &= \left\| \sum_x \mu(x) p_n(x, \cdot) - \pi \right\|_{\text{VT}} \\ &= \left\| \sum_x \mu(x) (p_n(x, \cdot) - \pi) \right\|_{\text{VT}} \text{ car } \sum_x \mu(x) = 1 \\ &\leq \sum_x \mu(x) \|p_n(x, \cdot) - \pi\|_{\text{VT}} \text{ car } \|\cdot\|_{\text{VT}} \text{ est une norme} \\ &\leq \max_x \|p_n(x, \cdot) - \pi\|_{\text{VT}} \text{ encore une fois car } \sum_x \mu(x) = 1. \end{aligned}$$

On remarquera que, même si on n'a travaillé qu'avec des mesures de probabilité, on a utilisé dans l'avant-dernière inégalité, le fait que $\|\cdot\|_{\text{VT}}$ est une norme sur \mathbb{R}^V . \square

On a donc

$$t_{\text{mél}}(\varepsilon) = \min\{n \geq 0 : \max_x \|p_n(x, \cdot) - \pi\|_{\text{VT}} \leq \varepsilon\}.$$

Grâce au Lemme 3.7, on déduit le théorème suivant du Théorème 2.1 :

Théorème 3.8. *On a*

$$\mathcal{D}(n) \xrightarrow{n \rightarrow +\infty} 0.$$

En particulier, $t_{\text{mél}}(\varepsilon)$ est fini pour tout $\varepsilon \in (0, 1/2)$.

Démonstration. On a (en utilisant l'inégalité triangulaire)

$$\mathcal{D}(n) = \max_{x \in V} \max_A |p_n(x, A) - \pi(A)| \leq \max_{x, y \in V} |V| |p_n(x, y) - \pi(y)|.$$

V étant fini, le Théorème 2.1 permet de conclure. □

Exemple 3.9. Si $|V| = 2$, disons $V = \{0, 1\}$, alors

$$\mathcal{D}(n) = \max\{|p_n(0, 0) - \pi(0)|, |p_n(1, 1) - \pi(1)|\}. \quad (5)$$

En effet, si $A = \emptyset$ ou $A = V$, alors $p_n(0, A) = p_n(1, A) = \pi(A)$, et donc

$$\mathcal{D}(n) = \max_{x \in V, A \subset V} |p_n(x, A) - \pi(A)| = \max_{x, y \in \{0, 1\}} |p_n(x, y) - \pi(y)|.$$

Enfin, $p_n(0, 1) - \pi(1) = -(p_n(0, 0) - \pi(0))$ et $p_n(1, 0) - \pi(0) = -(p_n(1, 1) - \pi(1))$, ce qui donne (5).

Exemple 3.10. Soient $a, b \in (0, 1)$. On regarde la chaîne de Markov à deux états définie par $V = \{0, 1\}$ et

$$P = \begin{pmatrix} p(0, 0) & p(0, 1) \\ p(1, 0) & p(1, 1) \end{pmatrix} = \begin{pmatrix} 1 - a & a \\ b & 1 - b \end{pmatrix}.$$

Cherchons π . Pour cela, on résout le système donné par $\pi = \pi P$ et $\pi(0) + \pi(1) = 1$ et on trouve que $\pi(0) = \frac{b}{a+b}$ et $\pi(1) = \frac{a}{a+b}$.

Étudions maintenant $\mathcal{D}(n)$. Rappelons que, d'après (5),

$$\mathcal{D}(n) = \max\{|p_n(0, 0) - \pi(0)|, |p_n(1, 1) - \pi(1)|\}.$$

Étudions tout d'abord $\delta_n := p_n(0, 0) - \pi(0)$. On a $\delta_0 = a/(a+b)$ et

$$\begin{aligned} \delta_{n+1} &= p_n(0, 0)p(0, 0) + p_n(0, 1)p(1, 0) - \frac{b}{a+b} \\ &= p_n(0, 0)(1 - a) + (1 - p_n(0, 0))b - \frac{b}{a+b} \\ &= p_n(0, 0)(1 - a - b) - \frac{b - ab - b^2}{a+b} = (1 - a - b)\delta_n. \end{aligned}$$

Et donc,

$$p_n(0, 0) - \pi(0) = (1 - a - b)^n \frac{a}{a + b}.$$

Par symétrie,

$$p_n(1, 1) - \pi(1) = (1 - a - b)^n \frac{b}{a + b} = p_n(1, 0) - \pi(0).$$

On a donc

$$\mathcal{D}(n) = |1 - a - b|^n \frac{(a \vee b)}{a + b}.$$

On peut maintenant calculer le temps de mélange. Soit $\varepsilon \in (0, 1/2)$. $t_{\text{mél}}(\varepsilon)$ est le plus petit entier positif tel que

$$|1 - a - b|^n \frac{(a \vee b)}{a + b} \leq \varepsilon.$$

On trouve, en notant $\alpha = |1 - a - b| \in [0, 1)$ et $\beta = (a \vee b)/(a + b) \in (0, 1)$, que $t_{\text{mél}}(\varepsilon) = 1$ si $\alpha = 0$ et que si $\alpha \neq 0$,

$$t_{\text{mél}}(\varepsilon) = \left\lceil \frac{\log(1/(\beta\varepsilon))}{\log(1/\alpha)} \right\rceil.$$

3.3 Premières propriétés

Avant d'étudier $\mathcal{D}(n)$ et $t_{\text{mél}}(\varepsilon)$ plus précisément (et dans un cadre général), donnons des formules équivalentes pour la distance en variation totale qui seront utiles tout au long du cours.

Lemme 3.11. *Pour toutes $\mu, \nu \in \text{Pr}(V)$,*

$$\|\mu - \nu\|_{\text{VT}} = \max_{A \subset V} (\mu(A) - \nu(A)) = \sum_{x \in V} (\mu(x) - \nu(x))_+ = \frac{1}{2} \sum_{x \in V} |\mu(x) - \nu(x)|.$$

Démonstration. (i) La première égalité vient du fait que $|\mu(A) - \nu(A)| = \max\{\mu(A) - \nu(A), \mu(A^c) - \nu(A^c)\}$.

(ii) Pour démontrer, la deuxième inégalité, on remarque tout d'abord que

$$\begin{aligned} \mu(A) - \nu(A) &= \sum_{x \in A} \mu(x) - \nu(x) \\ &\leq \sum_{x \in A} (\mu(x) - \nu(x))_+. \end{aligned}$$

Il reste à montrer que $\max_A (\mu(A) - \nu(A)) \geq \sum_x (\mu(x) - \nu(x))_+$. Pour cela, on note $A_0 = \{x \in V : \mu(x) \geq \nu(x)\}$ et on remarque que

$$\begin{aligned} \mu(A_0) - \nu(A_0) &= \sum_{x \in A_0} \mu(x) - \nu(x) \\ &= \sum_{x \in V} (\mu(x) - \nu(x))_+. \end{aligned}$$

(iii) Pour démontrer la troisième égalité, on écrit que

$$\sum_{x \in V} (\mu(x) - \nu(x))_+ + \sum_{x \in V} (\nu(x) - \mu(x))_+ = \sum_{x \in V} |\mu(x) - \nu(x)|$$

et on remarque les deux termes du membre de droite sont égaux (car tous deux égaux à $\|\mu - \nu\|_{\text{VT}}$). \square

Montrons maintenant que $\mathcal{D}(n)$ est décroissante, ce qui implique que π sera bien approximée même si on va au delà de $t_{\text{mél}}(\varepsilon)$.

Lemme 3.12. $\mathcal{D}(n)$ est décroissant en n . Et donc, pour tous $\varepsilon \in (0, 1/2)$ et $n \geq t_{\text{mél}}(\varepsilon)$, on a

$$\mathcal{D}(n) \leq \varepsilon.$$

Pour démontrer ce lemme, on va utiliser le lemme suivant – qui montre que les matrices de transition ont un effet de “contraction” sur la distance en variation totale.

Lemme 3.13. Soit $Q = (q(x, y))_{x, y \in V}$ une matrice de transition sur V et soient $\mu, \nu \in \text{Pr}(V)$. Alors,

$$\|\mu Q - \nu Q\|_{\text{VT}} \leq \|\mu - \nu\|_{\text{VT}}.$$

Utilisons d’abord ce lemme pour montrer le Lemme 3.12.

Démonstration du Lemme 3.12. $\pi = \pi P$, donc

$$\mathcal{D}(n+1) = \max_{\mu} \|\mu P^{n+1} - \pi P\|_{\text{VT}} \leq \max_{\mu} \|\mu P^n - \pi\|_{\text{VT}} = \mathcal{D}(n). \quad \square$$

Démonstration du Lemme 3.13. D’après la première égalité du Lemme 3.11,

$$\begin{aligned} \|\mu Q - \nu Q\|_{\text{VT}} &= \max_A \mu Q(A) - \nu Q(A) \\ &= \max_A \sum_{x \in A} \sum_{y \in E} \mu(y) q(y, x) - \nu(y) q(y, x) \\ &= \max_A \sum_{y \in E} (\mu(y) - \nu(y)) \sum_{x \in A} q(y, x) \\ &\leq \max_A \sum_{y \in E} (\mu(y) - \nu(y))_+ \sum_{x \in A} q(y, x) \\ &\leq \max_A \sum_{y \in E} (\mu(y) - \nu(y))_+ = \|\mu - \nu\|_{\text{VT}}, \end{aligned}$$

où la dernière égalité vient de la deuxième égalité du Lemme 3.11. \square

$\mathcal{D}(n)$ vérifie par ailleurs une propriété de sous-multiplicativité :

Lemme 3.14. Pour tous $m, n \geq 0$,

$$\mathcal{D}(n+m) \leq 2\mathcal{D}(n)\mathcal{D}(m).$$

Afin de démontrer ce lemme, on va donner une nouvelle formule pour $\mathcal{D}(n)$, et voir cette quantité non comme le sup d'une norme mais comme la norme d'une matrice.

Lemme 3.15. *Comme la démonstration du théorème de Perron–Frobenius, on note $\Pi = (\Pi(x, y))_{x, y \in V}$ la matrice stochastique dont toutes les lignes sont égales à π . De plus, si $A = (a(x, y))_{x, y \in V}$ est une matrice réelle, on note*

$$\|A\| = \max_{x \in V} \sum_{y \in V} |a(x, y)|.$$

Alors, pour tout $n \geq 1$,

$$\mathcal{D}(n) = \frac{1}{2} \|(P - \Pi)^n\|.$$

De plus, pour toutes matrices réelles $A = (a(x, y))_{x, y \in V}$ et $B = (b(x, y))_{x, y \in V}$,

$$\|AB\| \leq \|A\| \|B\|.$$

Démonstration du Lemme 3.14. Si $n = 0$ ou $m = 0$ on utilise que $\mathcal{D}(0) \geq 1/2$ (cf. la Remarque 3.6). Supposons que $n, m \geq 1$. D'après le Lemme 3.15,

$$\mathcal{D}(n + m) = \frac{1}{2} \|(P - \Pi)^{n+m}\| \leq 2 \times \frac{1}{2} \|(P - \Pi)^n\| \times \frac{1}{2} \|(P - \Pi)^m\| = 2\mathcal{D}(n)\mathcal{D}(m). \quad \square$$

Démonstration du Lemme 3.15. On remarque que Π^2 , ΠP et $P\Pi$ sont toutes égales à Π (respectivement car $\pi \in \text{Pr}(V)$, car π est une mesure invariante et car P est une matrice de transition). On en déduit par récurrence que $(P - \Pi)^n = P^n - \Pi$ pour tout $n \geq 1$. Or $\mathcal{D}(n) = \|P^n - \Pi\|$, donc on obtient le premier résultat. Pour montrer le deuxième résultat, on écrit que

$$\begin{aligned} \|AB\| &= \max_x \sum_y \left| \sum_z a(x, y)b(y, z) \right| \leq \max_x \sum_y \sum_z |a(x, y)||b(y, z)| \\ &\leq \max_x \sum_z |a(x, y)| \|B\| \leq \|A\| \|B\|. \quad \square \end{aligned}$$

On en déduit une propriété importante : $t_{\text{mél}}(\varepsilon)$ et $t_{\text{mél}}(\varepsilon')$ sont du même ordre de grandeur quels que soient nos choix de ε et ε' . Pour cette raison, on va faire un choix arbitraire et noter

$$t_{\text{mél}} = t_{\text{mél}}(1/4).$$

Corollaire 3.16. *Pour tous $\varepsilon, \varepsilon' \in (0, 1/2)$, il existe une constante $N(\varepsilon, \varepsilon') \geq 1$ qui ne dépend que de ε et ε' telle que*

$$t_{\text{mél}}(\varepsilon) \leq N(\varepsilon, \varepsilon') t_{\text{mél}}(\varepsilon').$$

En particulier, pour tout $\varepsilon \in (0, 1/2)$, il existe une constante $N(\varepsilon) \geq 1$ qui ne dépend que de ε telle que

$$t_{\text{mél}}(\varepsilon) \leq N(\varepsilon) t_{\text{mél}}.$$

Démonstration. On remarque d'abord que, d'après le Lemme 3.14 (et en raisonnant par récurrence sur n), pour tous $m, n \geq 1$, on a

$$\mathcal{D}(nm) \leq (2\mathcal{D}(m))^n.$$

Soient maintenant $\varepsilon, \varepsilon' \in (0, 1/2)$ et soit $N \geq 1$ tel que $(2\varepsilon)^N \leq \varepsilon'$. D'après l'inégalité ci-dessus, appliquée à $m = t_{\text{mél}}(\varepsilon)$ et $n = N$,

$$\mathcal{D}(Nt_{\text{mél}}(\varepsilon)) \leq (2\mathcal{D}(t_{\text{mél}}(\varepsilon)))^N \leq (2\varepsilon)^N \leq \varepsilon',$$

donc $t_{\text{mél}}(\varepsilon') \leq Nt_{\text{mél}}(\varepsilon)$. □

Feuille de TD 1 : distance en variation totale / temps de mélange

On rappelle qu'on se donne un ensemble V fini tel que $|V| \geq 2$, une matrice de transition apériodique $P = (p(x, y))_{x, y \in V}$ sur V , qu'on note $p_n(x, y) = P^n(x, y)$, et qu'on note π la mesure de probabilité invariante pour P . De plus, $X = (X_n)_{n \geq 0}$ est un processus à valeurs dans V et, pour toute $\mu \in \text{Pr}(V)$, \mathbb{P}_μ est une mesure de probabilité sous-laquelle $(X_n)_{n \geq 0}$ est une chaîne de Markov de matrice de transition P et de loi initiale μ . Enfin, on note $\mathbb{P}_x = \mathbb{P}_{\delta_x}$.

A. Un exercice préliminaire

Exercice 3.1. Soit $Q = (q(x, y))_{x, y \in V}$ une matrice irréductible sur V . On suppose qu'il existe $x_0 \in V$ tel que $q(x_0, x_0) > 0$. Pourquoi Q est-elle apériodique ?

B. Trois exercices sur la distance en variation totale

Exercice 3.2. Pour tout $n \geq 1$, on note μ_n la loi binomiale de paramètres n et $1/2$. Par ailleurs, on note ν_n la loi de $Y + 1$ où $Y \sim \mu_n$. Montrer qu'il existe une constante $c > 0$ (qu'on calculera) telle que

$$\|\mu_n - \nu_n\|_{\text{VT}} = \frac{c}{\sqrt{n}}(1 + o_{n \rightarrow +\infty}(1)).$$

Exercice 3.3. Soit $N \in \mathbb{N}$ et soient V, W deux variables aléatoires à valeurs dans $\{0, \dots, N\}$ indépendantes de la chaîne de Markov X . On note μ_V et μ_W leur loi.

1. Montrer que X_V et X_W sont mesurables. On note μ^V et μ^W leur loi.
2. Montrer que

$$\|\mu^V - \mu^W\|_{\text{VT}} \leq \|\mu_V - \mu_W\|_{\text{VT}}.$$

Exercice 3.4. Soient V_1, \dots, V_n des ensembles finis et soient, pour tout $i \in \{1, \dots, n\}$, μ_i et ν_i des mesures de probabilité sur V_i . On note $\mu = \mu_1 \otimes \dots \otimes \mu_n$ et $\nu = \nu_1 \otimes \dots \otimes \nu_n$. Montrer que

$$\|\mu - \nu\|_{\text{VT}} \leq \sum_{i=1}^n \|\mu_i - \nu_i\|_{\text{VT}}.$$

C. La notion de sensibilité à une dynamique markovienne

Exercice 3.5. Dans cet exercice, on suppose que π est réversible.

Soit $A \subset V$ (fixé pour tout l'exercice), tel que $\pi(A) = 1/2$. On note $\pi^A = \pi(\cdot|A)$. On note aussi

$$\mathcal{D}^A(n) = \|\pi^A P^n - \pi\|_{\text{VT}} \quad \text{et} \quad t_{\text{mél}}^A(\varepsilon) = \min\{n \geq 0 : \mathcal{D}^A(n) \leq \varepsilon\}.$$

Soient $\varepsilon, \delta \in (0, 1/4]$, fixés pour tout l'exercice. On dit que A est (ε, δ) -sensible à la dynamique si

$$t_{\text{mél}}^A(\varepsilon) \leq \delta t_{\text{mél}},$$

où on rappelle qu'on utilise la notation $t_{\text{mél}} = t_{\text{mél}}(1/4)$.

(Il est en fait souvent plus pertinent d'étudier les ensembles A qui satisfont $t_{\text{mél}}^A(\varepsilon) \leq \delta t_{\text{rel}}$, où t_{rel} est le temps de relaxation, qu'on introduira plus tard dans le cours.)

1. Montrer que la propriété suivante est vraie sous la mesure \mathbb{P}_π : Pour tout n , $(X_k)_{0 \leq k \leq n}$ a la même loi que $(X_{n-k})_{0 \leq k \leq n}$.
2. (a) Montrer que la loi de X sous \mathbb{P}_{π^A} est égale à la loi de X sous $\mathbb{P}_\pi[\cdot|X_0 \in A]$.
(b) Montrer que, pour tout n ,

$$\text{Var}_\pi(\mathbb{P}_\pi[X_n \in A | X_0]) = \mathbb{P}_\pi[X_0, X_{2n} \in A] - \pi(A)^2,$$

où Var_π est la variance sous la mesure \mathbb{P}_π .

- (c) En déduire que, pour tout n ,

$$\mathcal{D}^A(n) \leq \sqrt{2} \sqrt{\pi^A P^{2n}(A) - \pi(A)}.$$

- (d) On définit donc

$$\widetilde{t}_{\text{mél}}^A(\varepsilon) = \min\{n \geq 0 \text{ pair} : \pi^A P^n(A) - \pi(A) \leq \varepsilon\}.$$

Montrer que

$$t_{\text{mél}}^A(\varepsilon) \leq \widetilde{t}_{\text{mél}}^A(\varepsilon^2/2)/2.$$

3. Le but de cette dernière partie est de montrer que, si A est (ε, δ) -sensible à la dynamique, alors,

$$\mathbb{P}_\pi[\exists n \in \{0, \dots, t_{\text{mél}}\} : X_n \in A] \geq 1 - \left(11\varepsilon + \frac{1}{4^{1/\delta}}\right).$$

Supposons donc que A est (ε, δ) -sensible à la dynamique. Pour ne pas ajouter d'hypothèses techniques, nous supposons par ailleurs dans cette partie que $t_{\text{mél}}^A(\varepsilon)$ est pair.

- (a) Montrer que

$$\mathbb{P}_\pi[\mathbb{P}_\pi[X_{t_{\text{mél}}^A(\varepsilon)/2} \in A | X_0] \geq 1/4] \leq 8\varepsilon.$$

(b) Notons $N_A = \{n \geq 0 : X_n \in A\}$ et

$$W_\varepsilon = \{x \in V : \mathbb{P}_\pi[X_{t_{\text{mél}}^A(\varepsilon)/2} \in A \mid X_0 = x] \geq 1/4\}.$$

Notons aussi $k = \lfloor 2t_{\text{mél}}/t_{\text{mél}}^A(\varepsilon) \rfloor$ et

$$t_i = it_{\text{mél}}^A(\varepsilon)/2$$

pour tout $i \in \{1, \dots, k\}$.

Montrer que, pour tout $i \in \{0, \dots, k\}$,

$$\begin{aligned} \mathbb{P}_\pi[N_A \cap \{0, \dots, t_i\} = \emptyset] &\leq \mathbb{P}_\pi[X_{t_{i-1}} \notin W_\varepsilon] \\ + \mathbb{E}_\pi[1_{\{X_{t_{i-1}} \in W_\varepsilon\}} \mathbb{P}_\pi[N_A \cap \{0, \dots, t_{i-1}\} = \emptyset \mid X_{t_{i-1}}] \times \mathbb{P}_\pi[N_A \cap \{t_{i-1}, \dots, t_i\} = \emptyset \mid X_{t_{i-1}}]]. \end{aligned}$$

(c) En déduire que

$$\mathbb{P}_\pi[N_A \cap \{0, \dots, t_i\} \leq \emptyset] \leq 8\varepsilon + \frac{1}{4} \mathbb{P}_\pi[N_A \cap \{0, \dots, t_{i-1}\} = \emptyset].$$

(d) Conclure.

4 Un exemple phare pour le cours : des marches aléatoires sur des graphes (et des “méthodes géométriques”)

Une grande partie du cours sera consacrée à l’étude de chaînes de Markov qu’on appelle “marches aléatoires simples paresseuses sur des graphes finis connexes”. Commençons par définir ce qu’est un graphe fini connexe.

Définition 4.1. Un **graphe** fini G est un couple (V, E) où V est un ensemble fini et E est un sous-ensemble de $\{\{x, y\} \subset V : x \neq y\}$. V est appelé l’ensemble des **sommets** et E est appelé l’ensemble des **arêtes**. On dit que x et y sont voisins – ce qu’on note $x \sim y$ – si $\{x, y\} \in E$.

On utilisera aussi les notions suivantes :

- (i) Le **degré** d’un sommet x est $\deg(x) = |\{y \in V : y \sim x\}|$ (on rappelle qu’on utilise $|\cdot|$ pour noter le cardinal d’un ensemble fini) ;
- (ii) G dit est **connexe** si, pour tous $x, y \in V$, il existe un entier $n \geq 0$ et un n -uplet (x_0, x_1, \dots, x_n) tel que
 - (a) $x_0 = x, x_n = y$ et
 - (b) $\{x_{k-1}, x_k\} \in E$ pour tout $k \in \{1, \dots, n\}$.

Dans cette sous-section, on fixe un graphe fini connexe $G = (V, E)$, avec V de cardinal au moins égal à 2, comme précédemment. La marche aléatoire simple “paresseuse” sur G est la chaîne de Markov sur V de matrice de transition P définie par

$$p(x, y) = \begin{cases} \frac{1}{2} & \text{si } x = y, \\ \frac{1}{2\deg(x)} & \text{si } x \sim y, \\ 0 & \text{sinon.} \end{cases}$$

Dans cette section, P sera toujours la matrice ci-dessus, et π sa mesure de probabilité invariante pour P . Vérifions qu'on respecte bien nos conventions selon lesquelles P est apériodique.

Lemme 4.2. P est apériodique.

Démonstration. Comme G est connexe, P est irréductible. En effet, si $x, y \in V$ et si n et (x_0, \dots, x_n) sont comme dans la définition de la connexité de G , on a

$$p_n(x, y) = \mathbb{P}_x[X_n = y] \geq \mathbb{P}_x[X_1 = x_1, \dots, X_n = x_n] = p(x_0, x_1) \dots p(x_{n-1}, x_n) > 0.$$

Le lemme est donc une conséquence du Lemme 2.3. \square

Un aspect agréable de la marche aléatoire simple est que π a une expression simple en fonction des degrés.

Lemme 4.3. π est la mesure de probabilité définie par

$$\forall x \in V, \quad \pi(x) = \frac{\deg(x)}{2|E|}.$$

Démonstration. Notons $\mu(x) = \frac{\deg(x)}{2|E|}$. Montrons déjà que μ est une mesure de probabilité :

$$\sum_{x \in V} \deg(x) = \sum_{x \in V, y \in V} 1_{\{x, y\} \in E} = 2|E|.$$

Par unicité de la mesure de probabilité invariante, il suffit donc de montrer que μ est une mesure invariante. Pour cela, il suffit de montrer que μ est réversible, i.e. $\mu(x)p(x, y) = \mu(y)p(y, x)$ pour tous $x \neq y$. Soient $x \neq y$. On a

$$\mu(x)p(x, y) = \frac{\deg(x)}{2|E|} \frac{1}{2\deg(x)} 1_{x \sim y} = \frac{1}{4|E|} 1_{x \sim y} = \frac{\deg(y)}{2|E|} \frac{1}{2\deg(y)} 1_{x \sim y} = \mu(y)p(y, x). \quad \square$$

Remarque 4.4. Par une preuve similaire, on peut aussi montrer que π est aussi la mesure de probabilité invariante pour la marche non paresseuse, i.e. définie par la matrice de transition \tilde{P} définie par $\tilde{p}(x, y) = 1/\deg(x)$ si $x \sim y$, et $\tilde{p}(x, y) = 0$ sinon. Cette matrice n'étant pas forcément apériodique, nous étudions plutôt la marche paresseuse.

4.1 L'hypercube

L'hypercube de dimension $d \geq 1$ est le graphe (V_d, E_d) défini par $V_d = \{0, 1\}^d$ et $x = (x_1, \dots, x_d) \sim y = (y_1, \dots, y_d)$ si et seulement s'il existe un unique $i \in \{1, \dots, d\}$ tel que $x_i \neq y_i$. Nous étudierons l'asymptotique de $t_{\text{mél}}(\varepsilon)$ quand d tend vers $+\infty$.

4.2 Le tore

Soit $N \geq 2$. Le tore de dimension $d \geq 1$ est le graphe $\mathbb{T}_N^d = (V_N^d, E_N^d)$ avec

$$V_N^d = (\mathbb{Z}/N\mathbb{Z})^d$$

et $x = (x_1, \dots, x_d) \sim (y_1, \dots, y_d)$ si et seulement s'il existe $i \in \{1, \dots, d\}$ tel que $x_i = y_i \pm 1$ et $x_j = y_j$ pour tout $j \neq i$. Nous étudierons l'asymptotique de $t_{\text{mél}}(\varepsilon)$ à d fixé et quand N tend vers $+\infty$. On va aussi comparer le temps de mélange sur le tore au temps de mélange sur "deux tores collés à un sommet".

Remarque 4.5. Le tore de dimension 1 est appelé le cercle.

Feuille de TD 2 : Temps de mélange pour des marches aléatoires

On rappelle qu'on se donne un ensemble V fini tel que $|V| \geq 2$, une matrice de transition apériodique $P = (p(x, y))_{x, y \in V}$ sur V , qu'on note $p_n(x, y) = P^n(x, y)$, et qu'on note π la mesure de probabilité invariante pour P . De plus, $X = (X_n)_{n \geq 0}$ est un processus à valeurs dans V et, pour toute $\mu \in \text{Pr}(V)$, \mathbb{P}_μ est une mesure de probabilité sous-laquelle $(X_n)_{n \geq 0}$ est une chaîne de Markov de matrice de transition P et de loi initiale μ . Enfin, on note $\mathbb{P}_x = \mathbb{P}_{\delta_x}$.

Exercice 4.1. Dans cet exercice, on considère la marche aléatoire simple paresseuse sur le tore (V_N^d, E_N^d) , qu'on note $X_n = (X_n^{(1)}, \dots, X_n^{(d)})$. On note $t_{\text{mél}}^{(N, d)}$ le temps de mélange associé (pour $\varepsilon = 1/4$). Pour simplifier, on suppose que $N \geq 4$ et que N est divisible par 4. On note

$$W_N^d = \{x = (x_1, \dots, x_d) \in V_N^d : x_1 \text{ a un représentant dans } \{N/4, \dots, 3N/4 - 1\}\}.$$

1. Soient ξ_1, ξ_2, \dots des variables i.i.d. de loi donnée par $\mathbb{P}[\xi_1 = 1] = \mathbb{P}[\xi_1 = -1] = 1/(4d)$ et $\mathbb{P}[\xi_1 = 0] = 1 - 1/(2d)$. Montrer que, pour tout $k \in \{1, \dots, d\}$, le processus $(X_n^{(k)})_{n \geq 0}$ sous la mesure de probabilité $\mathbb{P}_{(0, \dots, 0)}$ a la même loi que le processus $(\sum_{i=1}^n \xi_i \bmod N)_{n \geq 0}$.
2. En déduire que, pour tout $n \geq 0$, $\mathbb{P}_{(0, \dots, 0)}[X_n \in W_N^d] \leq \mathbb{P}[|\sum_{i=1}^n \xi_i| \geq N/4]$.
3. Montrer qu'il existe $c_0 > 0$ (indépendant de d et N) tel que $\mathbb{P}_{(0, \dots, 0)}[X_{n_N} \in W_N^d] < 1/4$, où $n_N = \lfloor c_0 d N^2 \rfloor$.
4. En déduire qu'il existe $c_1 > 0$ (indépendant de d et N) tels que $t_{\text{mél}}^{(N, d)} \geq c_1 d N^2$.

Exercice 4.2. Dans cet exercice, on se place dans le cadre de la marche aléatoire simple paresseuse sur un graphe fini connexe $G = (V, E)$. Si $x, y \in V$, on appelle chemin entre x et y un $(n+1)$ -uplet (x_0, x_1, \dots, x_n) , pour un certain $n \geq 0$, tel que

- (i) $x_0 = x, x_n = y$ et
- (ii) $\{x_{k-1}, x_k\} \in E$ pour tout $k \in \{1, \dots, n\}$.

On appelle distance de graphe entre x et y , notée $\text{dist}(x, y)$, le minimum de tous les n tels qu'un tel $(n + 1)$ -uplet existe. Le diamètre du graphe est alors

$$\text{diam} := \max_{x, y \in V} \text{dist}(x, y).$$

Le but de cet exercice est de montrer et appliquer la borne suivante :

$$\forall \varepsilon \in (0, 1/2), \quad t_{\text{mél}}(\varepsilon) \geq \text{diam}/2. \quad (6)$$

1. Montrer que dist définit bien une distance sur V .
2. Appliquer (6) à l'hypercube et au tore.
3. Si $x \in V$ et $n \geq 0$, on note $B_x(n) = \{y \in V : \text{dist}(x, y) \leq n\}$. Que pouvez-vous dire de $p_n(x, B_x(n))$ (pour tous $x \in V$ et $n \geq 0$) ? En déduire que $\mathcal{D}(n) \geq 1 - \min_{x \in V} \pi(B_x(n))$ pour tout $n \geq 0$.
4. Soit $\varepsilon \in (0, 1/2)$. Montrer que, si $n = t_{\text{mél}}(\varepsilon)$, alors il ne peut pas exister $x, y \in V$ tels que $B_x(n) \cap B_y(n) = \emptyset$.
5. Démontrer (6).

Exercice 4.3 (Temps de mélange et expansion (ou “méthode géométrique”)). On se place dans le cadre d'une marche aléatoire simple paresseuse sur un graphe fini connexe $G = (V, E)$, avec $|V| \geq 2$ (rappelons que $|\cdot|$ est utilisé pour noter le cardinal).

Le but de cet exercice est de lier le temps de mélange à une quantité géométrique appelée expansion. Pour définir l'expansion, définissons tout d'abord la frontière et le volume d'un ensemble $S \subset V$.

Définition 4.6. Soit $S \subset V$. Le volume de S est la quantité

$$\text{Vol}(S) = \sum_{x \in S} \text{deg}(x).$$

La frontière de S , notée ∂S , est l'ensemble des arêtes dont une extrémité est dans S et l'autre n'est pas dans S . Autrement dit,

$$\partial S = \{\{x, y\} \in E : x \in S, y \notin S\}.$$

- a. Quel est le volume d'un singleton ? De V ?
- b. Quelle est la frontière d'un singleton ? De V ?

Définition 4.7. L'expansion du graphe G est la quantité

$$\Phi_\star = \min_{S \subset V : S \neq \emptyset \text{ et } \text{Vol}(S) \leq |E|} \frac{|\partial S|}{\text{Vol}(S)}.$$

Le minimum est ainsi pris sur tous les $S \subset V$ non vides et de volume plus petit que $|E|$.

L'expansion mesure à quel point la frontière d'un ensemble peut être petite si on la compare au volume de cet ensemble. On se restreint à des ensembles S tels que $\text{Vol}(S) \leq |E|$ pour des raisons qu'on comprendra au fur et à mesure de l'énoncé.

1. Que vaut

$$\min_{S \subset V: S \neq \emptyset} \frac{|\partial S|}{\text{Vol}(S)} ?$$

(Qui est défini comme Φ_* mais sans restriction sur le volume de S .)

2. Montrer que, pour toute $\mu \in \text{Pr}(V)$, tout $n \geq 0$ et tout $S \subset V$, on a

$$\mathbb{P}_\mu[X_0 \in S, X_n \notin S] \leq \sum_{k=0}^{n-1} \mathbb{P}_\mu[X_k \in S, X_{k+1} \notin S].$$

En déduire que, pour tout $n \geq 0$ et tout $S \subset V$, on a

$$\mathbb{P}_\pi[X_0 \in S, X_n \notin S] \leq \frac{n}{4|E|} |\partial S|.$$

3. Montrer que, pour tout $n \geq 0$ et tout $S \subset V$,

$$\mathbb{P}_\pi[X_0 \in S, X_n \notin S] \geq \sum_{x \in S} \pi(x) (\pi(S^c) - \mathcal{D}(n)).$$

4. Montrer que, pour tout $\varepsilon \in]0, 1/2[$,

$$t_{\text{mél}}(\varepsilon) \geq \frac{1 - 2\varepsilon}{\Phi_*}.$$

On a ainsi trouvé une borne inférieure pour le temps de mélange à l'aide d'une quantité très géométrique : l'expansion. On avait déjà fait une telle chose en TD avec le diamètre.

5. Fixons une dimension $d \geq 1$ et notons $t_{\text{mél}}^{(N)}$ le temps de mélange (pour $\varepsilon = 1/4$) sur le tore $\mathbb{T}_N^d = (V_N^d, E_N^d)$. On a vu dans l'Exercice 1 la borne inférieure suivante et on verra plus tard la borne supérieure suivante : il existe des constantes $c, C > 0$ (qui peuvent dépendre de d mais de rien d'autre) telles que

$$cN^2 \leq t_{\text{mél}}^{(N)} \leq CN^2.$$

Regardons maintenant le graphe formé par deux tores collés, qui est défini en considérant deux tores identiques, en choisissant un sommet dans chacun des tores, et en reliant ces deux sommets par une arête (voir le dessin ci-dessous qui est fait dans le cas $d = 1$). Plus formellement, le graphe des deux tores collés est le graphe dont l'ensemble des sommets est

$$V_N^d \times \{0, 1\}$$

et l'ensemble des arêtes est défini par

- (i) $(x, 0) \sim (y, 0)$ si et seulement si $x \sim y$ dans \mathbb{T}_N^d ;
- (ii) $(x, 1) \sim (y, 1)$ si et seulement si $x \sim y$ dans \mathbb{T}_N^d ;
- (iii) $(x, 0) \sim (y, 1)$ si et seulement si $x = y = (0, \dots, 0)$.

Fixons une dimension $d \geq 3$. Montrer que, quand $N \rightarrow +\infty$, le temps de mélange pour deux tores collés est bien plus grand que le temps de mélange pour un tore.

6. Comparer ce que vous avez répondu à la question suivante à ce que vous auriez pu faire en utilisant une autre quantité géométrique : le diamètre (i.e. en utilisant l'Exercice 2 plutôt que l'Exercice 3). (La réponse à cette question peut être informelle.)

5 Le couplage (ou “méthode probabiliste”)

On rappelle qu'on se donne un ensemble V fini tel que $|V| \geq 2$, une matrice de transition apériodique $P = (p(x, y))_{x, y \in V}$ sur V , qu'on note $p_n(x, y) = P^n(x, y)$, et qu'on note π la mesure de probabilité invariante pour P . De plus, $X = (X_n)_{n \geq 0}$ est un processus à valeurs dans V et, pour toute $\mu \in \text{Pr}(V)$, \mathbb{P}_μ est une mesure de probabilité sous-laquelle $(X_n)_{n \geq 0}$ est une chaîne de Markov de matrice de transition P et de loi initiale μ . Enfin, on note $\mathbb{P}_x = \mathbb{P}_{\delta_x}$.

5.1 Couplage et distance en variation totale

Commençons par définir la notion centrale de cette section.

Définition 5.1. Soient $\mu, \nu \in \text{Pr}(V)$. Un **couplage** de (μ, ν) est un couple de variables aléatoires (X, Y) défini sur un même espace de probabilité tel que la loi de X est μ et la loi de Y est ν .

Pour déterminer la distance entre deux mesures de probabilités, on pourrait décider qu'elles sont proches s'il est possible de définir un couplage (X, Y) tel qu'avec grande probabilité, $X = Y$. La proposition ci-dessous nous dit que cela donne la même notion que la distance en variation totale.

Proposition 5.2. Pour toutes $\mu, \nu \in \text{Pr}(V)$,

$$\|\mu - \nu\|_{\text{VT}} = \inf_{(X, Y) \text{ couplage de } (\mu, \nu)} \mathbb{P}[X \neq Y],$$

où l'infimum est sur tous les couplages (définis sur n'importe quel espace de probabilité). De plus, il existe un couplage tel que l'infimum est atteint.

Exemple 5.3. On note μ_p la loi de Bernoulli de paramètre p . On a vu que, si $p, q \in [0, 1]$, alors $\|\mu_p - \mu_q\|_{\text{VT}} = |p - q|$. Par ailleurs, on peut définir un couplage (X_p, X_q) de (μ_p, μ_q) de la façon suivante : on se donne une variable uniforme U sur $[0, 1]$, et on définit X_p et X_q par $X_p = 1_{\{U \leq p\}}$ et $X_q = 1_{\{U \leq q\}}$, de telle façon que (si par exemple $p \leq q$), $\mathbb{P}[X_p \neq X_q] = \mathbb{P}[p < U \leq q] = |p - q|$.

Démonstration de la Proposition 5.2. Soit (X, Y) un couplage de (μ, ν) . D'après le Lemme 3.11,

$$\begin{aligned} \|\mu - \nu\|_{\text{VT}} &= \max_A (\mu(A) - \nu(A)) \\ &= \max_A \mathbb{P}[X \in A] - \mathbb{P}[Y \in A] \\ &\leq \max_A \mathbb{P}[X \in A] - \mathbb{P}[X \in A, Y \in A] \\ &= \max_A \mathbb{P}[X \in A, Y \notin A] \\ &\leq \mathbb{P}[X \neq Y]. \end{aligned}$$

C'est cette borne qu'on utilisera en pratique. Mais d'un point de vue théorique, il est aussi intéressant de savoir qu'il existe un couplage tel que l'autre inégalité est aussi vraie. Construisons donc un couplage (X, Y) de (μ, ν) tel que $\mathbb{P}[X \neq Y] = \|\mu - \nu\|_{\text{VT}}$.

On va procéder de la façon suivante : on note $V = \{x_1, \dots, x_N\}$ où $N = |V|$, on se donne une variable U uniforme dans $[0, 1]$ et on définit des points $i_0, j_0, \ell_0, \dots, i_N, j_N, \ell_N \in [0, 1]$ par

$$\begin{aligned} i_n &= \sum_{k=1}^n \mu(x_k) \wedge \nu(x_k); \\ j_n &= i_N + \sum_{k=1}^n (\mu(x_k) - \nu(x_k))_+; \\ \ell_n &= i_N + \sum_{k=1}^n (\nu(x_k) - \mu(x_k))_+. \end{aligned}$$

(Avec la convention qu'une somme sur un ensemble vide vaut 0 donc que $i_0 = 0$ et $j_0 = \ell_0 = i_N$.) On remarque que, pour tous $a, b \in \mathbb{R}$, $a \wedge b + (a - b)_+ = a$, ce qui implique que $j_N = \sum_{x \in V} \mu(x) = \ell_N = \sum_{x \in V} \nu(x) = 1$.

On définit (X, Y) par

$$\begin{aligned} \forall k \in \{1, \dots, N\}, \quad X &= x_k \text{ si et seulement si } U \in [i_{k-1}, i_k) \cup [j_{k-1}, j_k); \\ \forall k \in \{1, \dots, N\}, \quad Y &= x_k \text{ si et seulement si } U \in [i_{k-1}, i_k) \cup [\ell_{k-1}, \ell_k). \end{aligned}$$

(Le seul cas où cela ne définit pas les variables X et Y est quand $U = 1$; dans ce cas, qui arrive avec probabilité 0, on donne une valeur arbitraire à X et Y .) On a alors

$$\forall x \in V, \quad \mathbb{P}[X = x] = \mu(x) \wedge \nu(x) + (\mu(x) - \nu(x))_+ = \mu(x),$$

donc $X \sim \mu$. De même, $Y \sim \nu$. Enfin,

$$\mathbb{P}[X \neq Y] = \mathbb{P}[U \geq i_N] = \sum_{x \in V} (\mu(x) - \nu(x))_+ = \|\mu - \nu\|_{\text{VT}},$$

où la dernière égalité vient du Lemme 3.11. □

5.2 Estimer $\mathcal{D}(n)$ sans connaître π

En pratique, on utilisera souvent la Proposition 5.2 avec le Lemme 5.4 ci-dessous. Dans de nombreux cas, on sait que π existe mais elle n'est pas explicite. Il est donc très utile de trouver une façon d'estimer $\mathcal{D}(n)$ sans avoir besoin de parler de π . On note

$$\overline{\mathcal{D}}(n) = \max_{x,y \in V} \|p_n(x, \cdot) - p_n(y, \cdot)\|_{\text{VT}}.$$

Avec une preuve très proche de celle du Lemme 3.7, on montre le lemme suivant.

Lemme 5.4. *Pour tout $n \geq 0$,*

$$\mathcal{D}(n) \leq \overline{\mathcal{D}}(n) \leq 2\mathcal{D}(n).$$

Démonstration. La deuxième inégalité vient de l'inégalité triangulaire. Démontrons la première inégalité. Pour cela, on remarque tout d'abord que, pour tout $n \geq 0$,

$$\pi = \pi P^n = \sum_{y \in V} \pi(y) p_n(y, \cdot).$$

En utilisant cela et le Lemme 3.7, on obtient que

$$\begin{aligned} \mathcal{D}(n) &= \max_{x \in V} \left\| p_n(x, \cdot) - \sum_{y \in V} \pi(y) p_n(y, \cdot) \right\|_{\text{VT}} \\ &= \max_{x \in V} \left\| \sum_{y \in E} \pi(y) (p_n(x, \cdot) - p_n(y, \cdot)) \right\|_{\text{VT}} \quad \text{car } \sum_y \pi(y) = 1 \\ &\leq \max_{x \in V} \sum_{y \in V} \pi(y) \|p_n(x, \cdot) - p_n(y, \cdot)\|_{\text{VT}} \quad \text{car } \|\cdot\|_{\text{VT}} \text{ est une norme} \\ &\leq \max_{x,y \in V} \|p_n(x, \cdot) - p_n(y, \cdot)\|_{\text{VT}} \quad \text{encore une fois car } \sum_y \pi(y) = 1. \quad \square \end{aligned}$$

5.3 Noyau de couplage

Afin d'utiliser le Lemme 5.4 et la Proposition 5.2, on va étudier des chaînes de Markov sur V^2 dont les deux marginales sont des chaînes de Markov de matrice de transition P . En fait, pour que la théorie des couplages nous soit utile, on va se restreindre à une classe particulière de telles chaînes sur V^2 . On a besoin de la notion de noyau de couplage.

Définition 5.5. On appelle **noyau de couplage** (sous-entendu : pour P) une matrice de transition

$$Q = q((x, y), (x', y'))_{(x,y),(x',y') \in V^2}$$

sur V^2 qui satisfait la propriété suivante : Si $(X_n, Y_n)_{n \geq 0}$ est un processus à valeurs dans V^2 et si pour tout $(x, y) \in V^2$, $\mathbb{P}_{(x,y)}$ est une mesure de probabilité sous-laquelle $(X_n, Y_n)_{n \geq 0}$ est une chaîne de Markov de matrice de transition Q issue de (x, y) , alors, pour tous $x, x', y, y' \in V$,

$$\mathbb{P}_{(x,y)}[X_1 = x'] = p(x, x') \quad \text{et} \quad \mathbb{P}_{(x,y)}[Y_1 = y'] = p(y, y'). \quad (7)$$

Exemple 5.6. La matrice de transition Q sur V^2 définie par

$$\forall (x, y), (x', y') \in V^2, \quad q((x, y), (x', y')) = p(x, x')p(y, y')$$

est un noyau de couplage. En effet, si $x, x', y \in V$, alors

$$\begin{aligned} \mathbb{P}_{(x,y)}[X_1 = x'] &= \sum_{y' \in V} \mathbb{P}_{(x,y)}[X_1 = x', Y_1 = y'] \\ &= \sum_{y' \in V} q((x, y), (x', y')) = \sum_{y' \in V} p(x, x')p(y, y') = p(x, x'). \end{aligned}$$

De même, si $x, y, y' \in V$, alors $\mathbb{P}_{(x,y)}[Y_1 = y'] = p(y, y')$.

On se donne un noyau de couplage Q pour tout le reste de la section. De plus dans toute cette section, $(X_n, Y_n)_{n \geq 0}$ est un processus à valeurs dans V^2 et, pour toute $\mu \in \text{Pr}(V^2)$, \mathbb{P}_μ est une mesure de probabilité sous-laquelle $(X_n, Y_n)_{n \geq 0}$ est une chaîne de Markov de matrice de transition Q et de loi initiale μ . (Et comme précédemment, on note $\mathbb{P}_{(x,y)} = \mathbb{P}_{\delta_{(x,y)}}$ pour tous $x, y \in V$.)

Montrons que les marginales d'une chaîne de matrice de transition Q sont bien des chaînes de matrice de transition P (ce qui justifie l'utilisation de la notation X_n). On va aussi trouver une autre chaîne de matrice de transition P , en réunissant les deux chaînes lorsqu'elles se rencontrent.

Lemme 5.7. 1. X et Y sont des chaînes de Markov de matrice de transition P .

2. Définissons

$$T = \inf\{n \geq 0 : X_n = Y_n\} \in \mathbb{N} \cup \{+\infty\}.$$

Alors, T est un temps d'arrêt pour (la filtration canonique de) la chaîne de Markov $(X_n, Y_n)_{n \geq 0}$. Définissons un processus $(Z_n)_{n \geq 0}$ par

$$Z_n = \begin{cases} Y_n & \text{si } n \leq T, \\ X_n & \text{si } n \geq T. \end{cases}$$

Alors, $(Z_n)_{n \geq 0}$ est aussi une chaîne de Markov de matrice de transition P .

On va écrire la démonstration pour X et pour Z (la démonstration pour Y est la même que pour X). On a vu deux façons d'appliquer les propriétés de Markov. Pour illustrer chacune d'elles, on va appliquer l'une à X et l'autre à Z .

Démonstration. 1. On démontre le résultat pour X , la preuve étant la même pour Y . Soit $\mu \in \text{Pr}(V^2)$, soient $x_0, \dots, x_n \in V$ tels que $\mathbb{P}_\mu[X_0 = x_0, \dots, X_n = x_n] > 0$ et soit $x_{n+1} \in V$. On a (en utilisant la propriété de Markov simple à la deuxième

égalité)

$$\begin{aligned}
& \mathbb{P}_\mu[X_0 = x_0, \dots, X_n = x_n, X_{n+1} = x_{n+1}] \\
&= \sum_{y_n \in V} \mathbb{P}_\mu[X_0 = x_0, \dots, X_n = x_n, Y_n = y_n, X_{n+1} = x_{n+1}] \\
&= \sum_{y_n \in V} \mathbb{P}_\mu[X_0 = x_0, \dots, X_n = x_n, Y_n = y_n] \mathbb{P}_{(x_n, y_n)}[X_1 = x_{n+1}] \\
&= \sum_{y_n \in V} \mathbb{P}_\mu[X_0 = x_0, \dots, X_n = x_n, Y_n = y_n] p(x_n, x_{n+1}) \\
&= \mathbb{P}_\mu[X_0 = x_0, \dots, X_n = x_n] p(x_n, x_{n+1}),
\end{aligned}$$

où l'avant-dernière égalité est due au fait que Q est un noyau de couplage. On a donc bien

$$\mathbb{P}_\mu[X_{n+1} = x_{n+1} \mid X_0 = x_0, \dots, X_n = x_n] = p(x_n, x_{n+1}).$$

2. Tout d'abord, on observe que, pour tout $n \geq 0$, $\{T \leq n\} = \cup_{k=0}^n \{X_k = Y_k\}$ est mesurable par rapport à $\sigma(X_0, Y_0, \dots, X_n, Y_n)$. Ainsi, T est bien un temps d'arrêt pour la filtration canonique de la chaîne de Markov $(X_n, Y_n)_{n \geq 0}$.

Montrons maintenant que Z est une chaîne de Markov de matrice de transition P . Soit $\mu \in \text{Pr}(V^2)$, soient $z_0, \dots, z_n \in V$ tels que $\mathbb{P}_\mu[Z_0 = z_0, \dots, Z_n = z_n] > 0$ et soit $z_{n+1} \in V$.

Pour simplifier les notations, notons $E_{n-1} = \{Z_0 = z_0, \dots, Z_{n-1} = z_{n-1}\}$. On a

$$\begin{aligned}
& \mathbb{P}_\mu[Z_0 = z_0, \dots, Z_n = z_n, Z_{n+1} = z_{n+1}] \\
&= \mathbb{P}_\mu[T \leq n, E_{n-1}, X_n = z_n, X_{n+1} = z_{n+1}] \\
&\quad + \mathbb{P}_\mu[T \geq n+1, E_{n-1}, Y_n = z_n, Y_{n+1} = z_{n+1}].
\end{aligned}$$

Notons A et B les deux termes ci-dessus. Le but est de montrer que $A + B = p(z_n, z_{n+1}) \mathbb{P}_\mu[Z_0 = z_0, \dots, Z_n = z_n]$. Or,

$$A = \mathbb{E}_\mu \left[\mathbb{1}_{\{T \leq n, E_{n-1}, X_n = z_n\}} \mathbb{P}_\mu[X_{n+1} = z_{n+1} \mid \mathcal{F}_n^{(X, Y)}] \right]$$

car $\{T \leq n, E_{n-1}, X_n = z_n\} \in \mathcal{F}_n^{(X, Y)}$. Par la propriété de Markov simple appliquée à la chaîne de Markov (X, Y) et au temps n , on a

$$\mathbb{P}_\mu[X_{n+1} = z_{n+1} \mid \mathcal{F}_n^{(X, Y)}] = \mathbb{P}_{(X_n, Y_n)}[X_1 = z_{n+1}].$$

Comme Q est un noyau de couplage, cette quantité est égale à $p(X_n, z_{n+1})$. Par conséquent,

$$\begin{aligned}
A &= \mathbb{E}_\mu \left[\mathbb{1}_{\{T \leq n, E_{n-1}, X_n = z_n\}} p(X_n, z_{n+1}) \right] \\
&= p(z_n, z_{n+1}) \mathbb{P}_\mu[T \leq n, E_{n-1}, X_n = z_n] \\
&= p(z_n, z_{n+1}) \mathbb{P}_\mu[T \leq n, Z_0 = z_0, \dots, Z_n = z_n].
\end{aligned}$$

En suivant les mêmes lignes, on montre que

$$B = p(z_n, z_{n+1}) \mathbb{P}_\mu [T > n, Z_0 = z_0, \dots, Z_n = z_n],$$

ce qui donne bien

$$A + B = p(z_n, z_{n+1}) \mathbb{P}_\mu [Z_0 = z_0, \dots, Z_n = z_n]. \quad \square$$

Le prochain théorème est un des résultats centraux de la théorie des temps de mélange.

Théorème 5.8. *Notons $T = \min\{n \geq 0 : X_n = Y_n\}$ comme précédemment. Alors, pour tout $n \geq 0$,*

$$\mathcal{D}(n) \leq \max_{x,y \in E} \mathbb{P}_{(x,y)} [T > n].$$

Démonstration. On utilise les mêmes notations que dans le Lemme 5.7. D'après le Lemme 5.4 puis la Proposition 5.2,

$$\mathcal{D}(n) \leq \max_{x,y \in V} \|p_n(x, \cdot) - p_n(y, \cdot)\|_{\text{VT}} \leq \max_{x,y \in V} \mathbb{P}_{(x,y)} [X_n \neq Z_n] = \max_{x,y \in V} \mathbb{P}_{(x,y)} [T > n]. \quad \square$$

Donnons une preuve alternative du théorème de Perron–Frobenius, de nature plus probabiliste que la première preuve qu'on a proposée. Notons que cette deuxième preuve s'étend bien mieux au cas infini dénombrable ; on pourrait suivre ces idées pour démontrer le Théorème 2.1 dans le cas plus général d'une matrice de transition irréductible récurrente positive sur un espace dénombrable éventuellement infini.

Preuve du Théorème 2.1. Choisissons Q comme dans l'Exemple 5.6. Montrons que, comme P est apériodique, Q l'est aussi. Pour cela, on démontre par récurrence que, pour tout $n \geq 0$,

$$q_n((x, y), (x', y')) = p_n(x, x') p_n(y, y').$$

Si $n = 0$, cela vient de la définition de la matrice identité. Supposons donc le résultat vrai pour un certain $n \geq 0$. On a alors

$$\begin{aligned} q_{n+1}((x, y), (x', y')) &= \sum_{(x'', y'')} q((x, y), (x'', y'')) \times q_n((x'', y''), (x', y')) \\ &= \sum_{(x'', y'')} p(x, x'') p(y, y'') p_n(x'', x') p_n(y'', y') \\ &= p_{n+1}(x, x') p_{n+1}(y, y'). \end{aligned}$$

Comme Q est apériodique, elle est irréductible. Comme de plus V est fini, tout point de V^2 est récurrent et donc, pour tout $(x, y) \in V^2$, on a

$$\mathbb{P}_{(x,y)} [T = +\infty] \leq P_{(x,y)} [T_{(x,x)} = +\infty] = 0,$$

où T est comme dans le Théorème 5.8 et $T_{(x,x)} = \min\{n \geq 0 : (X_n, Y_n) = (x, x)\}$. Par conséquent (et comme V est fini), le Théorème 5.8 implique que

$$\mathcal{D}(n) \xrightarrow{n \rightarrow +\infty} 0,$$

ce qui implique le Théorème 2.1. □

5.4 Application à la marche sur le cercle (ceci donne une correction pour le prochain TD)

Appliquons le Théorème 5.8 à la marche aléatoire sur le cercle. Soit $N \geq 3$. La marche sur le cercle est la chaîne de Markov sur $\mathbb{Z}/N\mathbb{Z}$ dont la matrice de transition est

$$p(x, y) = \begin{cases} 1/2 & \text{si } x = y, \\ 1/4 & \text{si } x = y + 1, \\ 1/4 & \text{si } x = y - 1. \end{cases}$$

Autrement dit, c'est la marche aléatoire simple paresseuse sur le tore de dimension $d = 1$. Afin de démontrer une borne supérieure pour $\mathcal{D}(n)$ – et donc aussi pour $t_{\text{mél}}(\varepsilon)$ – on va utiliser le Théorème 5.8. La première idée serait d'utiliser le noyau de couplage donné par l'Exemple 5.6. Cependant, avec un tel noyau, il pourrait arriver que X_n et Y_n soient à distance 1 l'une de l'autre mais décident toutes les deux de prendre la place l'une de l'autre alors qu'on souhaiterait les faire se rencontrer. Il va être plus efficace de choisir un couplage dans lequel, à chaque temps, une seule composante change. On choisit le noyau de couplage Q de la façon suivante :

$$q((x, y), (x', y')) = \begin{cases} 1/4 & \text{si } x = x' \text{ et } y = y' \pm 1, \\ 1/4 & \text{si } y = y' \text{ et } x = x' \pm 1, \\ 0 & \text{sinon.} \end{cases}$$

On peut construire une chaîne de Markov $(X, Y) = (X_n, Y_n)_{n \geq 0}$ de matrice de transition Q et issue de $(x \bmod N, y \bmod N)$ de la façon suivante : Soient $\xi_1, \tilde{\xi}_1, \xi_2, \tilde{\xi}_2, \dots$ des variables indépendantes telles que, pour tout $i \geq 1$,

$$\mathbb{P}[\xi_i = 1] = \mathbb{P}[\xi_i = -1] = 1/2 \quad \text{et} \quad \mathbb{P}[\tilde{\xi}_i = 1] = \mathbb{P}[\tilde{\xi}_i = 0] = 1/2.$$

Étant donnés $x, y \in \{0, \dots, N-1\}$, on définit (X, Y) par

$$X_n = V_n \bmod N \quad \text{et} \quad Y_n = W_n \bmod N$$

où

$$V_n = x + \sum_{i=1}^n \xi_i \tilde{\xi}_i \quad \text{et} \quad W_n = y + \sum_{i=1}^n \xi_i (1 - \tilde{\xi}_i).$$

Définissons T comme dans le Théorème 5.8 et notons

$$U_n = V_n - W_n = x - y + \sum_{i=1}^n \xi_i (2\tilde{\xi}_i - 1).$$

Notons que les $\xi_i (2\tilde{\xi}_i - 1)$ sont des variables indépendantes de même loi que ξ_i .

Supposons que $x \geq y$ (le cas $x < y$ est traité de la même façon). On a alors

$$T = \min\{n \geq 0 : U_n \in \{0, N\}\}.$$

Notons que, comme $U_n \bmod N$ est une chaîne de Markov irréductible sur un ensemble fini – donc est récurrente – on a $T < +\infty$ p.s.

Soit \mathcal{F}_n la tribu engendrée par $\xi_1, \tilde{\xi}_1, \dots, \xi_n, \tilde{\xi}_n$. Alors, $(U_n)_{n \geq 0}$ est une $(\mathcal{F}_n)_{n \geq 0}$ -martingale. Montrons que $(U_n^2 - n)_{n \geq 0}$ est aussi une $(\mathcal{F}_n)_{n \geq 0}$ -martingale :

- (i) $U_n^2 - n$ est bien \mathcal{F}_n -mesurable ;
- (ii) U_n est bien intégrable pour tout n fixé (car borné pour tout n fixé) ;
- (iii) De plus, pour tout $n \geq 0$,

$$\begin{aligned} & \mathbb{E}[U_{n+1}^2 - (n+1) \mid \mathcal{F}_n] \\ &= \mathbb{E}[U_n^2 + 2U_n(\xi_{n+1} - \tilde{\xi}_{n+1}) + (\xi_{n+1} - \tilde{\xi}_{n+1})^2 - n - 1 \mid \mathcal{F}_n] \\ &= U_n^2 - n + 2U_n\mathbb{E}[\xi_{n+1} - \tilde{\xi}_{n+1}] + \mathbb{E}[(\xi_{n+1} - \tilde{\xi}_{n+1})^2] - 1 \\ &= U_n^2 - n + 0 + 2\mathbb{E}[\xi_1^2] - 1 = U_n^2 - n. \end{aligned}$$

On en déduit que

$$U_{T \wedge n} \quad \text{et} \quad U_{T \wedge n}^2 - (T \wedge n)$$

sont deux $(\mathcal{F}_n)_{n \geq 0}$ -martingales. Par convergence monotone,

$$\mathbb{E}[T] = \lim_{n \rightarrow +\infty} \mathbb{E}[T \wedge n].$$

Or, $U_{T \wedge n}^2 - (T \wedge n)$ est une martingale, donc

$$\mathbb{E}[U_{T \wedge n}^2] - \mathbb{E}[T \wedge n] = (x - y)^2,$$

et donc

$$\mathbb{E}[T] = \lim_{n \rightarrow +\infty} \mathbb{E}[U_{T \wedge n}^2] - (x - y)^2.$$

Par convergence dominée (et comme $T < +\infty$ p.s.),

$$\lim_{n \rightarrow +\infty} \mathbb{E}[U_{T \wedge n}^2] = \mathbb{E}[U_T^2].$$

De plus, $U_T^2 = NU_T$ donc de nouveau par convergence dominée, puis en utilisant que $U_{T \wedge n}$ est une martingale, on a

$$\mathbb{E}[U_T^2] = N\mathbb{E}[U_T] = \lim_{n \rightarrow +\infty} N\mathbb{E}[U_{T \wedge n}] = N(x - y).$$

On obtient donc que

$$\mathbb{E}[T] = (N - (x - y))(x - y) \leq \min_{x \in [0, N]} Nx - x^2 = \frac{N^2}{4}.$$

Finalement (par l'inégalité de Markov), le Théorème 5.8 implique que, pour tout $n \geq 0$,

$$\mathcal{D}(n) \leq \frac{N^2}{4(n+1)}.$$

et donc, pour tout $\varepsilon \in (0, 1/2)$,

$$t_{\text{mél}}(\varepsilon) \leq \frac{N^2}{4\varepsilon}.$$

C'est le bon ordre de grandeur : nous avons déjà vu que $t_{\text{mél}} \geq cN^2$ pour un certain $c > 0$.

Feuille de TD 3 : Couplage (1/2)

On rappelle qu'on se donne un ensemble V fini tel que $|V| \geq 2$, une matrice de transition apériodique $P = (p(x, y))_{x, y \in V}$ sur V , qu'on note $p_n(x, y) = P^n(x, y)$, et qu'on note π la mesure de probabilité invariante pour P . De plus, $X = (X_n)_{n \geq 0}$ est un processus à valeurs dans V et, pour toute $\mu \in \text{Pr}(V)$, \mathbb{P}_μ est une mesure de probabilité sous-laquelle $(X_n)_{n \geq 0}$ est une chaîne de Markov de matrice de transition P et de loi initiale μ . Enfin, on note $\mathbb{P}_x = \mathbb{P}_{\delta_x}$. Enfin, on se donne un noyau de couplage $Q = (q(x, y), q(x', y'))_{(x, x'), (y, y') \in V^2}$ pour P .

Exercice 5.1. Dans cet exercice, $V = \{1, \dots, N\}$ pour un certain entier $N \geq 2$. Soient $\mu, \nu \in \text{Pr}(V)$. On dit que μ est stochastiquement plus grande que ν si, pour tout $k \in \{1, \dots, N\}$,

$$\mu[[k, N]] \geq \nu[[k, N]].$$

Montrer que μ est stochastiquement plus grande que ν si et seulement s'il existe un couplage (X, Y) de μ et ν tel que $X \geq Y$ p.s.

Exercice 5.2. Soit $N \geq 3$. On considère la marche aléatoire (paresseuse) sur le cercle – i.e. le tore de dimension 1 – qui est la chaîne de Markov sur $\mathbb{Z}/N\mathbb{Z}$ dont la matrice de transition P est donnée par

$$p(x, y) = \begin{cases} 1/2 & \text{si } x = y, \\ 1/4 & \text{si } x = y + 1, \\ 1/4 & \text{si } x = y - 1. \end{cases}$$

Le but de cet exercice est de démontrer que

$$\forall \varepsilon \in (0, 1/2), \quad t_{\text{mél}}(\varepsilon) \leq \frac{N^2}{4\varepsilon}. \quad (8)$$

C'est le bon ordre de grandeur (pour l'asymptotique $N \rightarrow +\infty$) : on avait en effet montré dans le TD 1 qu'il existait $c > 0$ tel que pour tout N , $t_{\text{mél}} \geq cN^2$ (en fait, on l'avait montré pour N divisible par 4, mais la preuve se généralise à tout N).

1. Montrer que Q définie de la façon suivante est bien un noyau de couplage :

$$q((x, y), (x', y')) = \begin{cases} 1/4 & \text{si } x = x' \text{ et } y = y' \pm 1, \\ 1/4 & \text{si } y = y' \text{ et } x = x' \pm 1, \\ 0 & \text{sinon.} \end{cases}$$

2. Soient $x, y \in \{0, \dots, N-1\}$. On peut construire une chaîne de Markov $(X, Y) = (X_n, Y_n)_{n \geq 0}$ de matrice de transition Q avec pour loi initiale $\mathbb{P}_{(x, y)}$ de la façon suivante⁶ : Soient $\xi_1, \tilde{\xi}_1, \xi_2, \tilde{\xi}_2, \dots$ des variables indépendantes telles que, pour tout $i \geq 1$,

$$\mathbb{P}[\xi_i = 1] = \mathbb{P}[\xi_i = -1] = 1/2 \quad \text{et} \quad \mathbb{P}[\tilde{\xi}_i = 1] = \mathbb{P}[\tilde{\xi}_i = 0] = 1/2.$$

6. On remarquera qu'on fait ici un abus de notation : on devrait écrire $\mathbb{P}_{(x \bmod N, y \bmod N)}$.

On définit (X, Y) par

$$X_n = V_n \pmod N \quad \text{et} \quad Y_n = W_n \pmod N$$

où

$$V_n = x + \sum_{i=1}^n \xi_i \tilde{\xi}_i \quad \text{et} \quad W_n = y + \sum_{i=1}^n \xi_i (1 - \tilde{\xi}_i).$$

Quelle est la loi de $\xi_i(2\tilde{\xi}_i - 1)$?

3. On note

$$U_n = V_n - W_n.$$

Soit \mathcal{F}_n la tribu engendrée par $\xi_1, \tilde{\xi}_1, \dots, \xi_n, \tilde{\xi}_n$. Montrer que $(U_n)_{n \geq 0}$ et $(U_n^2 - n)_{n \geq 0}$ sont des $(\mathcal{F}_n)_{n \geq 0}$ -martingales.

4. On note

$$T = \inf\{n \geq 0 : X_n = Y_n\}.$$

Montrer que T est un temps d'arrêt pour la filtration $(\mathcal{F}_n)_{n \geq 0}$.

Ainsi, $(U_{T \wedge n})_{n \geq 0}$ et $(U_{T \wedge n}^2 - T \wedge n)_{n \geq 0}$ sont aussi des $(\mathcal{F}_n)_{n \geq 0}$ -martingales.

5. Montrer que $T < +\infty$ p.s.

6. Calculer $\mathbb{E}[T]$.

7. Conclure.

8. Comment adapteriez-vous cette preuve à un tore de plus grande dimension ?

6 Le temps de relaxation (ou “méthode spectrale”)

On rappelle qu'on se donne un ensemble V fini tel que $|V| \geq 2$, une matrice de transition apériodique $P = (p(x, y))_{x, y \in V}$ sur V , qu'on note $p_n(x, y) = P^n(x, y)$, et qu'on note π la mesure de probabilité invariante pour P . De plus, $X = (X_n)_{n \geq 0}$ est un processus à valeurs dans V et, pour toute $\mu \in \Pr(V)$, \mathbb{P}_μ est une mesure de probabilité sous-laquelle $(X_n)_{n \geq 0}$ est une chaîne de Markov de matrice de transition P et de loi initiale μ . Enfin, on note $\mathbb{P}_x = \mathbb{P}_{\delta_x}$.

Rappelons aussi qu'on voit les fonctions $f : V \rightarrow \mathbb{R}$ (ou \mathbb{C}) comme les matrices colonnes formées par le vecteur $(f(x))_{x \in V}$. Ainsi, si $M = (m(x, y))_{x, y \in V}$ est une matrice, $Mf : V \rightarrow \mathbb{R}$ (ou \mathbb{C}) est la fonction définie par

$$Mf(x) = \sum_{y \in V} m(x, y) f(y).$$

On a donc

$$Pf(x) = \mathbb{E}_x[f(X_1)].$$

Dans toute cette section, on note $N = |V|$. Le but de cette section est d'effectuer un lien entre la décomposition spectrale de P et le temps de mélange. Dans le Lemme 3.11,

on a vu que $\mathcal{D}(n)$ est une quantité “de nature ℓ_1 ”. Dans cette section, on va étudier une base de diagonalisation pour P et on va voir que cette base est orthonormée pour le produit scalaire qui à (f, g) associe $\int fgd\pi$. Ceci est une propriété “de nature ℓ_2 ”, qui possède souvent des propriétés plus agréables que les propriétés de nature ℓ_1 (pensons par exemple à l’inégalité de Cauchy–Schwarz). On trouvera un moyen d’établir des inégalités entre des quantités ℓ_1 et des quantités ℓ_2 afin d’utiliser la décomposition spectrale pour étudier le temps de mélange.

Avant de faire cela, étudions la décomposition spectrale de matrices de transitions.

6.1 Décomposition spectrale des matrices de transition

On remarque déjà que, si Q est une matrice de transition (quelconque) sur V et si $\mathbf{1} : V \rightarrow \mathbb{R}$ est la fonction constante égale à 1, alors $Q\mathbf{1} = \mathbf{1}$. 1 est donc valeur propre de toute matrice de transition. Commençons par le lemme suivant.

Lemme 6.1. *Soit Q une matrice de transition sur V et soit $\lambda \in \mathbb{C}$ une valeur propre de Q . Alors, $|\lambda| \leq 1$. Si de plus $\lambda \neq 1$ et $\mu \in \text{Pr}(V)$ est invariante pour Q , alors $\mu f = \int fd\mu = 0$.*

Démonstration. Soit $\lambda \in \mathbb{C}$. Si $f : V \rightarrow \mathbb{C}$ est une fonction propre de valeur propre λ , alors

$$|\lambda| \|f\|_\infty = \max_x |Qf(x)| \leq \sum_y q(x, y) \|f\|_\infty = \|f\|_\infty,$$

donc $|\lambda| \leq 1$. Supposons maintenant que $\lambda \neq 1$ et que $\mu \in \text{Pr}(V)$ est invariante pour Q . Alors,

$$\mu f = \mu Qf = \lambda \mu f,$$

donc $\mu f = 0$. □

Plaçons-nous maintenant dans le cas des matrices irréductibles.

Lemme 6.2. *Soit Q une matrice de transition irréductible sur V . Si $f : V \rightarrow \mathbb{C}$ satisfait $Qf = f$, alors f est une fonction constante.*

Démonstration. Supposons par l’absurde qu’il existe $f : V \rightarrow \mathbb{C}$ non constante telle que $Qf = f$. On remarque que $Q\text{Re}(f) = \text{Re}(f)$ et $Q\text{Im}(f) = \text{Im}(f)$ et que $\text{Re}(f)$ ou $\text{Im}(f)$ n’est pas constante; il suffit donc de considérer le cas où f prend des valeurs réelles. Comme Q est irréductible et que f n’est pas constante, il existe $x, y \in V$ tels que $f(x) = \max_{z \in V} f(z) > f(y)$ et $q(x, y) > 0$ (sinon, $W := \{u \in V : f(u) = \max_{z \in V} f(z)\}$ vérifierait $\mathbb{P}_x[X_n \in W] = 0$ pour tous $x \in W$ et $n \geq 0$). On a alors

$$0 = Qf(x) - f(x) = \sum_{z \in V} q(x, z)(f(z) - f(x)) \leq q(x, y)(f(y) - f(x)) < 0,$$

ce qui est contradictoire. □

Continuons à ajouter des hypothèses : supposons que Q est apériodique.

Lemme 6.3. Soit Q une matrice de transition apériodique sur V . Alors, 1 est la seule valeur propre de module 1.

Démonstration. Soit n tel que toutes les entrées de Q^n sont strictement positives. On remarque tout d'abord que les entrées de Q^m sont strictement positives pour tout $m \geq n$ (dans le cas contraire, il existerait une colonne de Q qui est nulle, ce qui est contradictoire avec le fait que Q est irréductible). (En fait, on utilisera cela juste pour $m = n$ et $m = n + 1$.)

Soit $\lambda \in \mathbb{C}$ une valeur propre de module 1 et soit $m \geq n$. On note $\varepsilon = \min_{x \in V} q_m(x, x) > 0$. Alors,

$$R = \frac{Q^m - \varepsilon Id}{1 - \varepsilon}$$

est une matrice de transition et $\frac{\lambda^m - \varepsilon}{1 - \varepsilon}$ est une valeur propre de R . Le Lemme 6.1 implique donc que

$$\left| \frac{\lambda^m - \varepsilon}{1 - \varepsilon} \right| \leq 1,$$

ce qui implique (en écrivant le module comme $\sqrt{\Re(\cdot)^2 + \Im(\cdot)^2}$) que $\lambda^m = 1$. On en déduit que $1 = \lambda^{n+1} = \lambda \lambda^n = \lambda$. \square

En fait, on va travailler sous des hypothèses plus fortes encore : **Dans toute cette section, on suppose que notre matrice apériodique P est réversible.** Rappelons la définition de la réversibilité.

Définition 6.4. On dit que P est réversible si π l'est, i.e. si pour tous $x, y \in V$, $\pi(x)p(x, y) = \pi(y)p(y, x)$.

Exemple 6.5. C'est le cas de la marche aléatoire simple (paresseuse) sur tout graphe fini connexe d'après le Lemme 4.3.

On peut maintenant entrer dans le vif du sujet : on va essayer de comprendre la décomposition spectrale de P . Dans ce qui suit, on va démontrer que P est diagonalisable, avec une base de fonctions propres orthonormées pour le produit scalaire $\langle \cdot, \cdot \rangle_\pi$ défini plus bas. On pourrait démontrer cela en remarquant que P est auto-adjointe pour ce produit scalaire (ce qui signifie que $\langle Pf, g \rangle_\pi = \langle f, Pg \rangle_\pi$ pour toutes fonctions f, g), et en utilisant un théorème général de diagonalisation des opérateurs auto-adjoints. Mais on va plutôt choisir de démontrer la propriété de diagonalisation "à la main". On essaiera ensuite d'effectuer un lien entre la décomposition spectrale et de P et le temps de mélange.

Commençons par chercher une matrice symétrique fortement liée à P . On définit la matrice $A = (a(x, y))_{x, y \in V}$ par

$$a(x, y) = p(x, y) \sqrt{\frac{\pi(x)}{\pi(y)}}.$$

On remarque que A est bien définie car $\pi(x) > 0$ pour tout $x \in V$ (cf. le Théorème 1.1). Par ailleurs, on remarque que, comme P est réversible, A est une matrice symétrique. Il existe donc une base orthonormale $(\psi_k)_{k=1}^N$ (pour le produit scalaire $\langle f, g \rangle = \sum_x f(x)g(x)$) et des réels $\lambda_1, \dots, \lambda_N$ tels que, pour tout k , $\psi_k : V \rightarrow \mathbb{R}$ est une fonction propre de A de valeur propre λ_k i.e.

$$A\psi_k = \lambda_k\psi_k.$$

On fixe une telle base $(\psi_k)_{k=1}^N$ avec ses valeurs propres $\lambda_N \leq \dots \leq \lambda_1$.

Une autre façon d'écrire A est la suivante. On note D_π la matrice diagonale définie par

$$D_\pi(x, y) = \pi(x)\delta_{x,y}.$$

Alors,

$$A = \sqrt{D_\pi}P\sqrt{D_\pi}^{-1}.$$

Et donc, si ψ est une fonction propre pour A , alors $\sqrt{D_\pi}^{-1}\psi$ est une fonction propre pour P , de même valeur propre. On note

$$\varphi_k = \sqrt{D_\pi}^{-1}\psi_k, \quad \text{i.e. } \varphi_k : x \in V \mapsto \frac{\psi_k(x)}{\sqrt{\pi(x)}}.$$

On a donc trouvé une base de fonctions propres pour P . A priori, il n'y a pas de raison que la base formée par les fonctions propres φ_k soit elle aussi orthonormée. En fait, elle l'est pour un autre produit scalaire :

Définition 6.6. On définit $\langle \cdot, \cdot \rangle_\pi$, défini sur les fonctions $f, g : V \rightarrow \mathbb{R}$ par

$$\langle f, g \rangle_\pi = \int fgd\pi = \sum_{x \in V} \pi(x)f(x)g(x) = \mathbb{E}_\pi[f(X_n)g(X_n)]$$

pour tout $n \geq 0$.

Ce produit scalaire donne une "bonne structure" (pour l'étude des chaînes de Markov de matrice de transition P) à l'espace vectoriel des fonctions $f : V \rightarrow \mathbb{R}$.

Lemme 6.7. $\langle \cdot, \cdot \rangle_\pi$ est bien un produit scalaire. De plus, $(\varphi_k)_{k=1}^N$ est une base de fonctions propres pour P qui est orthonormée pour ce produit scalaire.

Démonstration. Grâce à la discussion ci-dessus, il ne reste plus qu'à montrer que $\langle \cdot, \cdot \rangle_\pi$ est bien un produit scalaire et que les $\varphi_k = \sqrt{D_\pi}^{-1}\psi_k$ sont orthonormées pour celui-ci.

Montrons que c'est un produit scalaire : c'est bien une forme bilinéaire symétrique. De plus, comme $\pi(x) > 0$ pour tout $x \in V$ (cf. le Théorème 1.1), $\sum_{x \in V} \pi(x)f(x)^2$ est supérieur ou égal à 0, avec égalité si et seulement si $f = 0$.

Montrons maintenant que les $\sqrt{D_\pi}^{-1}\psi_k$ sont orthonormés pour ce produit scalaire. On a

$$\begin{aligned} \langle \sqrt{D_\pi}^{-1}\psi_j, \sqrt{D_\pi}^{-1}\psi_k \rangle_\pi &= \sum_{x \in V} \pi(x) \frac{1}{\sqrt{\pi(x)}}\psi_j(x) \frac{1}{\sqrt{\pi(x)}}\psi_k(x) \\ &= \sum_{x \in V} \psi_j(x)\psi_k(x) = \delta_{j,k}. \quad \square \end{aligned}$$

On déduit des lemmes sur les valeurs propres des matrices de transition que :

Lemme 6.8. *On a*

$$-1 < \lambda_N \leq \dots \leq \lambda_2 < \lambda_1 = 1,$$

et φ_1 est la fonction constante égale à ± 1 . De plus, si $k \geq 2$, on $\int \varphi_k d\pi = 0$.

6.2 Application au temps de mélange

Les deux objets principaux des méthodes spectrales sont probablement les suivants.

Définition 6.9. Le **trou spectral** est

$$\lambda_\star = \max\{|\lambda_k| : k \in \{2, \dots, n\}\} \in [0, 1).$$

Le **temps de relaxation** est

$$t_{\text{rel}} = \frac{1}{\log(1/\lambda_\star)} \in [0, +\infty).$$

Pour toute $f : V \rightarrow \mathbb{R}$, on a $f = \sum_{k=1}^n \langle f, \varphi_k \rangle_\pi \varphi_k$ et donc

$$P^n f = \sum_k \lambda_k^n \langle f, \varphi_k \rangle_\pi \varphi_k = \int f d\pi + \sum_{k=2}^n \langle f, \varphi_k \rangle_\pi \lambda_k^n \varphi_k.$$

Un choix intéressant est $f_y : x \mapsto \delta_{x,y}$ de telle façon que $P^n f_y(x) = p_n(x, y)$, $\int f_y d\pi = \pi(y)$ et $\langle f_y, \varphi_k \rangle_\pi = \varphi_k(y)\pi(y)$. On obtient que

$$p_n(\cdot, y) - \pi(y) = \sum_{k=2}^n \pi(y) \lambda_k^n \varphi_k(y) \varphi_k.$$

Cette expression semble intéressante, car notre but est d'estimer à quel point $p_n(x, y)$ est proche de $\pi(y)$ pour tout $x \in V$. Lorsqu'on l'appliquera, on préférera voir y comme la variable. On a ainsi démontré le lemme suivant :

Lemme 6.10. *Pour tout $x \in V$, la fonction*

$$\frac{p_n(x, \cdot)}{\pi(\cdot)} - 1 : V \rightarrow \mathbb{R}$$

est égale à

$$\sum_{k=2}^n \lambda_k^n \varphi_k(x) \varphi_k.$$

On a enfin les outils nécessaires pour établir un lien fort (du moins quand on a un contrôle suffisant sur $\min \pi$) entre le temps de relaxation et le temps de mélange.

Théorème 6.11. *Pour tout $\varepsilon \in (0, 1/2)$,*

$$t_{\text{rel}} \log \left(\frac{1}{2\varepsilon} \right) \leq t_{\text{mél}}(\varepsilon) \leq t_{\text{rel}} \log \left(\frac{1}{2\varepsilon \min_{x \in V} \sqrt{\pi(x)}} \right) + 1.$$

Démonstration. Démontrons tout d'abord l'inégalité de gauche. Soit φ une fonction propre pour P dont la valeur propre est de valeur absolue λ_* . Soit $x \in V$ tel que $|\varphi(x)| = \|\varphi\|_\infty$. En utilisant que $\pi\varphi = 0$, on obtient que, pour tout $n \geq 0$,

$$\begin{aligned} \lambda_*^n \|\varphi\|_\infty &= |P^n \varphi(x)| = |P^n \varphi(x) - \pi\varphi| \leq \sum_y |p_n(x, y)\varphi(y) - \pi(y)\varphi(y)| \\ &\leq \sum_y |p_n(x, y) - \pi(y)| \|\varphi\|_\infty \leq 2\mathcal{D}(n) \|\varphi\|_\infty, \end{aligned}$$

d'après le Lemme 3.11. En évaluant en $n = t_{\text{mél}}(\varepsilon)$, on obtient que

$$\lambda_*^{t_{\text{mél}}(\varepsilon)} \leq 2\varepsilon,$$

ce qui est le résultat cherché.

Démontrons maintenant l'inégalité de droite. Pour cela, on va tout d'abord lier $\mathcal{D}(n)$ au produit scalaire $\langle \cdot, \cdot \rangle_\pi$, auquel on va alors pouvoir appliquer l'inégalité de Cauchy–Schwarz. D'après le Lemme 3.11,

$$\begin{aligned} \mathcal{D}(n) &= \frac{1}{2} \max_x \sum_y \pi(y) \left| \frac{p_n(x, y)}{\pi(y)} - 1 \right| = \frac{1}{2} \max_x \left\langle \left| \frac{p_n(x, \cdot)}{\pi(\cdot)} - 1 \right|, 1 \right\rangle_\pi \\ &\leq \frac{1}{2} \max_x \left\| \frac{p_n(x, \cdot)}{\pi(\cdot)} - 1 \right\|_\pi \|1\|_\pi = \frac{1}{2} \max_x \left\| \frac{p_n(x, \cdot)}{\pi(\cdot)} - 1 \right\|_\pi, \end{aligned}$$

où l'inégalité est une conséquence de l'inégalité de Cauchy–Schwarz.

Cette manipulation est intéressante car on connaît la décomposition de $\frac{p_n(x, \cdot)}{\pi(\cdot)} - 1$ grâce au Lemme 6.10, ce qui nous permet de calculer sa norme. Finalement, on obtient que

$$\mathcal{D}(n) \leq \frac{1}{2} \max_x \sqrt{\sum_{k=2}^n \lambda_k^{2n} \varphi_k(x)^2} \leq \frac{\lambda_*^n}{2} \sqrt{\sum_{k=2}^n \varphi_k(x)^2}. \quad (9)$$

Pour calculer $\sum_{k=2}^n \varphi_k(x)^2$, on utilise encore le Lemme 6.10, mais pour $n = 0$. Cela donne que $\sum_{k=2}^n \varphi_k(x)^2 = 1/\pi(x) - 1 \leq 1/\pi(x)$. Et donc,

$$\mathcal{D}(n) \leq \frac{\lambda_*^n}{2 \min_x \sqrt{\pi(x)}}.$$

Rappelons que $t_{\text{mél}}(\varepsilon) \geq 1$. On évalue l'expression ci-dessus en $n = t_{\text{mél}}(\varepsilon) - 1$, ce qui donne

$$2\varepsilon \min_x \sqrt{\pi(x)} < \lambda_*^{t_{\text{mél}}(\varepsilon)-1},$$

qui implique l'inégalité de droite du théorème. \square

Feuille de TD 4 : Couplage (2/2) et méthodes spectrales

A. La marche aléatoire paresseuse sur l'hypercube Dans cette section, on se place toujours dans le cadre de la marche aléatoire paresseuse sur l'hypercube (V_d, E_d) . On note

$$x = (x(1), \dots, x(d)) \in V_d = \{0, 1\}^d$$

les éléments de l'hypercube et on rappelle que $x \sim y$ si et seulement si x et y diffèrent sur une coordonnée et une seule).

Exercice 6.1 (Un couplage). Soient U_1, U_2, \dots et ξ_1, ξ_2, \dots des variables aléatoires toutes indépendantes, avec les U_i uniformes sur $\{1, \dots, d\}$ et les ξ_i Bernoulli de paramètre $1/2$. Pour tout $i \in \{1, \dots, d\}$ et tout $\alpha \in \{0, 1\}$, on définit l'application

$$\Phi_i^\alpha : \{0, 1\}^d \rightarrow \{0, 1\}^d$$

de l'hypercube vers lui-même de la façon suivante : $\Phi_i^\alpha(x)$ est égal à α sur la $i^{\text{ème}}$ coordonnée et coïncide avec x ailleurs. Autrement dit,

$$\Phi_i^\alpha(x) = (x(1), \dots, x(i-1), \alpha, x(i+1), \dots, x(d)).$$

Soient $x, y \in \{0, 1\}^d$. On définit les processus $(X_n)_{n \geq 0}$ et $(Y_n)_{n \geq 0}$ de la façon suivante :

$$X_0 = x \quad \text{et} \quad \forall n \in \mathbb{N}, X_{n+1} = \Phi_{U_{n+1}}^{\xi_{n+1}}(X_n)$$

et

$$Y_0 = y \quad \text{et} \quad \forall n \in \mathbb{N}, Y_{n+1} = \Phi_{U_{n+1}}^{\xi_{n+1}}(Y_n).$$

Montrer que X et Y sont des marches aléatoires paresseuses sur l'hypercube et que $((X_n, Y_n))_{n \geq 0}$ est une chaîne de Markov sur $\{0, 1\}^d \times \{0, 1\}^d$.

Exercice 6.2 (Couplage et collection de coupons). Avant d'étudier plus en détails la marche sur l'hypercube, commençons par une partie préliminaire sur le "problème de collection de coupons". Ce problème a pour sujet une chaîne de Markov $(A_n)_{n \geq 0}$ sur $\{0, \dots, d\}$ pour un certain $d \geq 1$. Soient U_1, U_2, \dots des variables i.i.d. uniformes sur $\{1, \dots, d\}$. Pour tout $n \geq 0$, on note

$$A_n = |\{U_1, \dots, U_n\}|.$$

On note aussi, pour tout $m \in \{0, \dots, d\}$,

$$T_{d,m} = \inf\{n \geq 0 : A_n = m\}.$$

On voit cette quantité comme le premier temps qu'on a collecté m coupons.

1. Montrer que $(A_n)_{n \geq 0}$ est bien une chaîne de Markov.
2. Montrer que les temps $\tau_m = T_{d,m} - T_{d,m-1}$, $m = 1, \dots, d$, sont indépendants. Quelle est leur loi ?

3. Quelle est l'espérance de $T_{d,m}$? Quelle est sa variance ?
4. En déduire que, pour toute suite $(m_d)_{d \geq 1}$ telle que $0 \leq m_d \leq \log d$ et pour tout $\delta > 0$,

$$\mathbb{P}[T_{d,d-m_d} \notin [(1-\delta)d \log d, (1+\delta)d \log d]] \xrightarrow{d \rightarrow +\infty} 0.$$

Étudions maintenant la marche sur l'hypercube. À l'aide de la question précédente et de l'Exercice 1, trouver une borne supérieure sur le temps de mélange pour la marche aléatoire paresseuse sur l'hypercube (dans l'asymptotique $d \rightarrow +\infty$).

Exercice 6.3 (Couplages et géométrie : un exercice inspiré du CC de l'an dernier). Dans cet exercice, on utilise les mêmes notations que dans l'Exercice 1.

1. Si $x, y \in \{0, 1\}^d$, on note $\text{dist}(x, y)$ le nombre de coordonnées auxquelles ces deux points de l'hypercube diffèrent. Montrer que

$$\mathbb{P}(x(U_1) \neq y(U_1)) = \frac{\text{dist}(x, y)}{d}.$$

En déduire la propriété suivante, appelée propriété de $(1 - \frac{1}{d})$ -contraction :

$$\mathbb{E}(\text{dist}(X_1, Y_1)) = \text{dist}(x, y) \left(1 - \frac{1}{d}\right).$$

2. Montrer que, pour tout $n \in \mathbb{N}$,

$$\mathbb{E}(\text{dist}(X_n, Y_n)) \leq d \left(1 - \frac{1}{d}\right)^n.$$

3. Montrer que

$$\|p_n(x, \cdot) - p_n(y, \cdot)\|_{\text{VT}} \leq \mathbb{E}(\text{dist}(X_n, Y_n)).$$

4. En déduire une majoration du temps de mélange de cette marche (pour $\varepsilon = 1/4$).

Exercice 6.4 (Méthodes spectrales). On étudie toujours la marche aléatoire paresseuse sur l'hypercube. Pour tout $i \in \{1, \dots, d\}$ on note $\sigma_i : \{0, 1\}^d \rightarrow \{0, 1\}^d$ l'application qui change la $i^{\text{ème}}$ coordonnée (et seulement celle-ci). Comme d'habitude, on note P la matrice de transition.

1. Quelle est la mesure de probabilité invariante ?
2. Montrer que, pour tout $f : \{0, 1\}^d \rightarrow \mathbb{R}$,

$$Pf = \frac{1}{2d} \sum_{i=1}^d (f + f \circ \sigma_i).$$

3. Montrer que les fonctions $\chi_S : \{0, 1\}^d \rightarrow \mathbb{R}$ suivantes, définies pour tout $S \subset \{1, \dots, d\}$, forment une base de fonctions propres pour P :

$$\chi_S : x \in \{0, 1\}^d \mapsto (-1)^{\sum_{i \in S} x_i}.$$

Quelles sont les valeurs propres associées ?

Montrer aussi que cette base est orthonormée pour le produit scalaire $\langle \cdot, \cdot \rangle_\pi$ défini par $\langle f, g \rangle_\pi = \sum_x f(x)g(x)\pi(x)$.

4. En utilisant le théorème établissant une connexion entre le temps de mélange et le temps de relaxation, en déduire les bornes inférieure et supérieure suivantes pour le temps de mélange : pour tout $\varepsilon \in (0, 1/2)$, il existe $c, c', d_0 > 0$ tels que, si $d \geq d_0$, alors

$$cd \leq t_{\text{mél}}^{(d)}(\varepsilon) \leq cd^2,$$

où on a ajouté la dépendance en d dans la notation du temps de mélange.

5. Estimer $\mathcal{D}(n)$ plus précisément en utilisant le fait qu'on connaît toute une base orthonormée et toutes les valeurs propres. (On pourra utiliser que $(1-x)^a \leq e^{-ax}$ pour tous $x \in [0, 1]$ et $a \geq 0$ et que $(1+x)^a \leq e^{xa}$ pour tous $x \geq 0$ et $a \geq 0$.)
6. En déduire que, pour tout $\delta > 0$ et tout $\varepsilon \in (0, 1/2)$, il existe $d_0 > 0$ tel que, si $d \geq d_0$, alors

$$t_{\text{mél}}^{(d)}(\varepsilon) \leq (0.5 + \delta)d \log d.$$

En fait, la borne ci-dessus est la bonne d'une façon très précise : on peut aussi montrer que pour tout $\delta > 0$ et tout $\varepsilon \in (0, 1/2)$, il existe $d_0 > 0$ tel que, si $d \geq d_0$, alors

$$t_{\text{mél}}^{(d)}(\varepsilon) \geq (0.5 - \delta)d \log d.$$

Ainsi, pour tous $\varepsilon, \varepsilon' \in (0, 1/2)$,

$$t_{\text{mél}}^{(d)}(\varepsilon) = t_{\text{mél}}^{(d)}(\varepsilon')(1 + o(1)),$$

où $o(1) \rightarrow 0$ quand $d \rightarrow +\infty$.

On avait vu en cours que le temps de mélange ne dépend pas de ε si on ne s'intéresse qu'à son ordre grandeur. Ainsi

$$0 < \inf_N \frac{t_{\text{mél}}^{(N)}(\varepsilon)}{t_{\text{mél}}^{(N)}(\varepsilon')} \leq \sup_N \frac{t_{\text{mél}}^{(N)}(\varepsilon)}{t_{\text{mél}}^{(N)}(\varepsilon')} < +\infty.$$

Ici, la dépendance en ε est encore moins importante. On appelle ce phénomène le phénomène de cutoff, et savoir si une suite de chaînes de Markov satisfait un tel phénomène est une des questions majeures de l'étude des temps de mélange. Par exemple, on peut montrer que la marche sur le tore ne satisfait pas un tel phénomène.

B. Mélange de cartes

Exercice 6.5. Dans cet exercice, on étudie une façon, assez naïve, de mélanger un grand paquet de cartes. On pourra utiliser les résultats sur le problème de collection de coupons vus plus haut.

Plus formellement, on va étudier une marche aléatoire sur le groupe symétrique \mathfrak{S}_N . Pour cela, on définit tout d'abord N permutations $\gamma_i, i = 1, \dots, N$, de la façon suivante :

$$\gamma_i(k) = \begin{cases} i & \text{si } k = 1 \\ k - 1 & \text{si } k = i, \\ k & \text{si } k > i. \end{cases}$$

On définit une matrice de transition P sur \mathfrak{S}_N de la façon suivante :

$$p(\sigma, \sigma') = \begin{cases} 1/N & \text{s'il existe } i \text{ tel que } \sigma' = \sigma \circ \gamma_i, \\ 0 & \text{sinon.} \end{cases}$$

Pourquoi parlions-nous au début de l'énoncé de mélange de paquet de cartes ? Considérons un paquet de N cartes, numérotées de 1 à N . Chaque élément $\sigma \in \mathfrak{S}_N$ est vu comme un état possible du paquet de cartes : celui où la $i^{\text{ème}}$ carte a pour valeur $\sigma(i)$. Avec ce point de vue, la marche aléatoire définie ci-dessus peut être décrite de la façon suivante : à chaque étape, on choisit une position $i \in \{1, \dots, N\}$ uniformément, on prend la carte placée à la $i^{\text{ème}}$ position, et on la place en haut du paquet.

1. Montrer que P est bien apériodique irréductible. Quelle est la mesure stationnaire π ?
2. Une façon de construire une marche aléatoire X de matrice P est la suivante. Soit $\mu \in \text{Pr}(\mathfrak{S}_N)$, et soient X_0 et U_1, U_2, \dots , des variables indépendantes telles que $X_0 \sim \mu$ et les U_i sont des variables uniformes sur $\{1, \dots, N\}$. On définit récursivement X_n en notant

$$X_{n+1} = X_n \circ \gamma_{U_{n+1}}$$

pour tout $n \geq 0$. Construisons un autre processus sur \mathfrak{S}_N , de la façon suivante. On note $Y_0 = X_0$ et on définit récursivement Y_n en notant

$$V_{n+1} = Y_n^{-1}(U_{n+1}) \quad \text{et} \quad Y_{n+1} = Y_n \circ \gamma_{V_{n+1}} \quad \text{pour tout } n \geq 0.$$

V_{n+1} est ainsi la position de la carte dont la valeur est U_{n+1} dans Y_n . Montrer que $(X_n)_{n \geq 0}$ et $(Y_n)_{n \geq 0}$ sont de même loi. Le mélange étudié dans cet exercice est donc, en loi, la même chose que le mélange étrange qui consisterait, à chaque étape, à choisir une valeur de carte uniformément, chercher cette carte dans le paquet, et la placer en haut du paquet. On va voir que, particulièrement pour une méthode de couplage, ce point de vue est intéressant.

3. Soient $\sigma, \sigma' \in \mathfrak{S}_N$. On définit $(Y_n)_{n \geq 0}$ comme ci-dessus en choisissant $\mu = \delta_\sigma$, i.e. $Y_0 = \sigma$ p.s. On définit un processus $(Z_n)_{n \geq 0}$ de la même façon mais en partant de σ' (il est important de noter qu'on utilise les mêmes variables U_i) :

$$Z_0 = \sigma' \quad ; \quad W_{n+1} = Z_n^{-1}(U_{n+1}) \quad \text{et} \quad W_{n+1} = Z_n \circ \gamma_{W_{n+1}} \quad \text{pour tout } n \geq 0.$$

Montrer que, pour tout $n \geq 0$,

$$\mathcal{D}(n) \leq \mathbb{P}[T_{N,N} > n],$$

où on utilise les notations de l'Exercice 2.

4. En déduire que, pour tout $\varepsilon \in (0, 1/2)$,

$$\liminf_{N \rightarrow +\infty} \frac{t_{\text{mél}}^{(N)}(\varepsilon)}{N \log N} \leq 1,$$

où on a ajouté la dépendance en N dans la notation du temps de mélange.

5. Soit $m \in \{1, \dots, N\}$. Notons E_m l'événement que les m cartes du bas du paquet sont rangées par ordre croissant, i.e.

$$E_m = \{\sigma \in \mathfrak{S}_N : \sigma(N) > \sigma(N-1) > \dots > \sigma(N-m-1)\}.$$

Que vaut $\pi(E_N)$?

6. Montrer que

$$\mathbb{P}_{id}[X_n \in A_m] \geq \mathbb{P}[T_{N, N-m+1} > n].$$

7. En déduire que, pour tout $\varepsilon \in (0, 1/2)$,

$$\frac{t_{\text{mél}}^{(N)}(\varepsilon)}{N \log N} \xrightarrow{N \rightarrow +\infty} 1.$$

Ainsi, ce mélange de cartes satisfait le phénomène dit de cutoff défini à la fin de la section précédente.

7 Références (et lien vers des “méthodes fonctionnelles”)

Les deux références principales dans l'écriture de ce poly sont :

- Mixing times of Markov chains (J. Salez)
- Markov Chains and Mixing Times (D. Levin, Y. Peres, E. Wilmer)
- Intégration, Probabilités et Processus Aléatoires (J.-F. Le Gall)

Vous pouvez par exemple consulter la Section 5 des notes de cours de J. Salez pour découvrir les méthodes variationnelles (ou fonctionnelles).