

Mémoire de L3 - Département de mathématiques et
applications (ENS Ulm)

Chaînes de Markov et temps de mélange: l'exemple de la marche aléatoire sur l'hypercube

Linus Bleistein * Dowan Koo † Linus Rösler ‡

Directrices de mémoire:

Anna Ben-Hamou (Sorbonne Université) et Claire Boyer (Sorbonne Université, ENS Ulm)

Résumé

Nous introduisons tout d'abord le cadre d'étude des chaînes de Markov, processus stochastiques à temps discret. Nous développons ensuite les notions de temps de mélange et de *cut-off* en introduisant une distance sur l'espace des distributions de probabilités discrètes (sur des espaces finis). Nous nous concentrons plus spécifiquement sur le cas de la marche aléatoire sur l'hypercube. Ce mémoire comporte également quelques simulations numériques (marches aléatoires, temps de mélange, modèle d'Ising).

*linus.bleistein@ens.fr

†dowan.koo@ens.fr

‡linus.rosler@epfl.ch

Les chaînes de Markov font partie des processus stochastiques à temps discret les plus élémentaires; étudiées pour la première fois en 1906 par le célèbre mathématicien russe Andreï Markov, elles fournissent encore aujourd'hui un cadre théorique souple et extraordinairement riche largement utilisé en physique, en biologie, en économie ou encore en informatique. Dans ce mémoire, nous nous intéressons à une question particulière, étudiée en probabilités depuis les années 1980 - 1990, qui est celle du temps de mélange et du *cut-off* des chaînes de Markov.

Sous de bonnes conditions, une chaîne de Markov converge vers une distribution dite stationnaire. Cette convergence se fait, pour certaines chaînes, de manière très brusque autour d'un temps qu'on appelle le temps de mélange: ce phénomène est qualifié de *cut-off*. Si les chaînes de Markov et les phénomènes de *cut-off* ont bien sûr une importance théorique, elles sont également très utiles à comprendre pour la simulation de lois de probabilité sur de grands ensembles. Cette question est au cœur de la méthode de Metropolis, que nous évoquons brièvement dans ce mémoire. Elle consiste plus ou moins à approcher une loi de probabilité sur un grand ensemble par une chaîne de Markov qui converge vers cette loi de probabilité. Savoir où se situe le *cut-off* est crucial: si notre simulation s'arrête juste avant, nos efforts de simulations sont vains; cependant, dès que le point de *cut-off* est passé, on ne gagne plus en précision.

Nous introduisons d'abord les concepts fondamentaux nécessaires pour l'étude des chaînes de Markov (chapitre 1). Une question qui émerge naturellement de l'étude de ces chaînes est la suivante: combien d'étapes sont nécessaires pour que la chaîne soit "bien mélangée"? Pour prendre l'exemple du mélange d'un jeu de cartes, combien de fois faut-il battre les cartes pour que celles-ci soient "éloignées" de la disposition initiale du paquet? La formulation de ces questions nous indique d'emblée un premier problème, qui est celui de la définition d'une bonne distance permettant de quantifier cet écart entre plusieurs distributions de probabilité. Ce premier problème résolu, on peut ensuite revenir à notre préoccupation initiale - étudier l'évolution du temps de mélange, c'est à dire de la distance entre la chaîne de Markov et une certaine distribution de probabilité. Plusieurs problèmes surgissent alors: comment se comporte cette distance au cours du temps? Connaît-elle des variations brusques? Si oui, à quels moments? Ces dernières questions sont au centre de l'étude du *cut-off* des chaînes de Markov. Nous abordons tous ces concepts dans le chapitre 2. Dans le chapitre 3, nous explorons les principales techniques utilisées pour estimer le temps de mélange et pour étudier le phénomène de *cut-off*. Tout ce mémoire est traversé par un objet théorique très étudié, la marche aléatoire sur l'hypercube de dimension n , $\{0, 1\}^n$: nous en déterminons quelques propriétés dans le chapitre 1, avant de finalement parvenir à la détermination précise de son temps de mélange dans les chapitres 2 et 3.

Le petit chapitre 4 est dédié à une application rapide des objets étudiés à un objet de physique statistique, le modèle d'Ising. Nous effectuons quelques simulations sur ce modèle et sur l'hypercube; le code MATLAB utilisé est rapidement présenté dans l'annexe A.

Une présentation de ce mémoire a été faite le mercredi 20 juin 2018, à l'ENS Ulm.

Nous remercions chaleureusement Anna Ben-Hamou et Claire Boyer pour leurs conseils avisés et l'aide qu'elles nous ont apportée dans la rédaction et la conception de ce mémoire.

Beaucoup de preuves et en général beaucoup d'exemples, explications et remarques s'inspirent très fortement resp. sont en grande partie reprises de l'ouvrage remarquable [2].

Table des matières

1	Une introduction aux chaînes de Markov: quelques définitions et exemples	4
1.1	Chaîne de Markov sur un espace d'états fini	4
1.2	Irréductibilité et apériodicité	5
1.3	Représentation d'une chaîne de Markov par une suite récurrente aléatoire	7
1.4	Existence et unicité de la distribution stationnaire d'une chaîne de Markov	8
1.5	Réversibilité	11
1.6	L'exemple de la marche aléatoire sur l'hypercube	11
2	Temps de mélange et <i>cut-off</i>	12
2.1	Une distance sur l'espace des distributions de probabilité: la distance de variation totale	12
2.2	Le théorème de convergence	15
2.3	Le temps de mélange	16
2.4	Le phénomène de <i>cut-off</i>	16
3	Techniques d'étude du temps de mélange et du <i>cut-off</i>	24
3.1	Couplage	25
3.2	Temps d'arrêt	30
3.2.1	Temps stationnaires forts	30
3.2.2	Raffinement: temps forts	34
3.3	Valeurs propres	39
3.4	Statistiques distinctives	42
4	Le modèle d'Ising et la dynamique de Glauber	44
A	Annexe de code	47
A.1	Calculs numérique pour l'hypercube	47
A.2	Distance de variation totale	48
A.3	Distance \bar{d}	50
A.4	Distance de séparation	51
A.5	Le modèle d'Ising	53

1 Une introduction aux chaînes de Markov: quelques définitions et exemples

1.1 Chaîne de Markov sur un espace d'états fini

Les chaînes de Markov font partie d'une vaste catégorie d'objets mathématiques, les processus stochastiques à temps discret. De manière simplifiée, on peut décrire une chaîne de Markov comme une suite de variables aléatoires indexée sur les entiers dont le comportement aléatoire en un temps t dépend uniquement de l'événement qui s'est réalisé au temps précédent. On peut par exemple imaginer un promeneur amnésique se baladant par pas unitaire sur une grande partie bornée de \mathbb{Z}^2 et qui choisit uniformément un chemin à emprunter parmi les 4 possibles, tout en oubliant le chemin parcouru jusqu'ici. La probabilité qu'il emprunte en t le chemin liant le point $(1, 1)$ au point $(1, 2)$ dépend uniquement du chemin emprunté en $t - 1$: si en l'instant $t - 1$ il est sur le point $(1, 1)$, la probabilité d'emprunter le chemin allant à $(1, 2)$ vaut $\frac{1}{4}$; sinon, elle vaut 0. Formalisons cette définition.

Définition 1.1. Soit $(X_t)_{t \in \mathbb{N}}$ (on va souvent écrire aussi $(X_t)_{t \geq 0}$) une suite de variables aléatoires à valeurs dans un ensemble fini Ω . On dit que (X_t) est une *chaîne de Markov* si elle vérifie la propriété dite de Markov: il existe une matrice $P \in [0, 1]^{\Omega \times \Omega}$, dite *matrice de transition*, tel que pour tout $x, y \in \Omega$, tout $t \geq 0$ et tout $x_0, \dots, x_{t-1} \in \Omega$ avec $\mathbb{P}(X_0 = x_0, \dots, X_{t-1} = x_{t-1}, X_t = x) > 0$ l'on a

$$\mathbb{P}(X_{t+1} = y \mid X_0 = x_0, \dots, X_{t-1} = x_{t-1}, X_t = x) = \mathbb{P}(X_{t+1} = y \mid X_t = x) = P(x, y). \quad (1)$$

Remarque 1.2. Faisons dès à présent une remarque de notation pour ne pas égarer le lecteur dès les premières pages: les coefficients d'une matrice P sont ici notés $P(i, j)$ et non pas p_{ij} , comme pourrait s'y attendre le lecteur français.

Remarque 1.3. On peut aussi considérer le cas plus général, dit *non homogène*, où cette matrice dépend de t . Une chaîne de Markov donnée comme dans la définition ci-dessus est donc plus précisément qualifiée d'*homogène*. C'est le cas de toutes les chaînes de Markov étudiées ici, et on l'a donc incluse dans la définition.

Par la propriété de Markov, on peut ainsi résumer l'ensemble des informations sur la loi du processus étudié en un vecteur de taille $|\Omega|$ correspondant à la distribution, dite initiale, de X_0 et une matrice de taille $|\Omega| \times |\Omega|$! Puisque $P(x, \cdot)$ est une distribution de probabilité (la distribution de X_{t+1} sachant $X_t = x$), on remarque aussi la chose suivante:

Proposition 1.4. La matrice P est stochastique, c'est à dire que $\forall x \in \Omega, \sum_{y \in \Omega} P(x, y) = 1$.

La matrice P permet de calculer certaines probabilités très aisément à travers des calculs matriciels élémentaires. Elle permet par conséquent d'introduire dans l'analyse des chaînes de Markov des arguments et des idées d'algèbre linéaire très utiles (théorème spectral, valeurs propres, etc.) qu'on étudiera en détail dans la partie 3.3. Faisons quelques remarques sur le calcul avec P .

- Le calcul s'effectue généralement de la manière suivante: on part d'une distribution initiale $X_0 \sim \mu_0$, que l'on écrit sous forme d'un vecteur ligne de taille $|\Omega|$ telle que $\mu_0(x) = \mathbb{P}(X_0 = x)$ pour tout $x \in \Omega$. Par exemple, si μ_0 vaut 0 partout sauf en une coordonnée, la chaîne "part" de ce point. Pour passer à l'étape suivante, i.e. pour calculer la distribution μ_1 de X_1 , on peut se convaincre qu'il suffit de multiplier à droite par P : $\mu_1 = \mu_0 P$. Le vecteur μ_1 représente ainsi la probabilité de se trouver en chacune des points de Ω en étant parti de la distribution initiale μ_0 .

- Plus généralement, on remarque que la distribution μ_t de X_t est donnée par $\mu_t = \mu_s P^{t-s}$, où μ_s est la distribution de X_s pour un $s \leq t$. En particulier, $P^t(x, \cdot)$ est la distribution de X_t si la chaîne est partie du point x . Multiplier un vecteur ligne qui représente une distribution de probabilité à droite par la r -ième puissance de P signifie donc mettre à jour cette distribution de probabilité par le passage de r étapes de temps.
- Un vecteur colonne f de taille $|\Omega|$ représente une fonction de Ω à valeurs réelles. Multiplier f à gauche par P signifie ainsi que pour tout $x \in \Omega$, l'on a $Pf(x) = \mathbb{E}_x[f(X_1)]$, i.e. l'espérance de $f(X_1)$ si la chaîne est partie de x .

Remarque 1.5. Pour une chaîne de Markov de distribution initiale $X_0 \sim \mu_0$, on notera \mathbb{P}_{μ_0} et \mathbb{E}_{μ_0} la probabilité et l'espérance de l'espace de probabilité sous-jacent, sur lequel la chaîne est définie. Dans le cas où μ_0 est une masse de Dirac δ_x , i.e. avec l'interprétation ci-dessus un vecteur ligne dont tout les coordonnées valent 0 sauf la x -ième qui vaut 1, on écrit simplement \mathbb{P}_x et \mathbb{E}_x . On peut penser de cette notation du point de vue du théorème d'extension de Kolmogorov, où pour changer la loi resp. les marginales d'un processus, on change plutôt la mesure de probabilité de l'espace sous-jacent au lieu du processus lui même.

1.2 Irréductibilité et apériodicité

Notre but est d'étudier la convergence de la distribution de X_t pour t tendant vers l'infini, vers une distribution de probabilité qui ne dépend pas de la distribution initiale. On verra plus tard dans la partie 2.2 que l'on a une convergence à un niveau très général, mais il est néanmoins nécessaire d'imposer deux conditions de non-dégénérescence, sans lesquelles la convergence et l'unicité pourraient être impossibles: l'irréductibilité et l'apériodicité.

La notion d'irréductibilité est assez simple à énoncer: une chaîne de Markov sur Ω est dite irréductible si pour tout point, on peut passer de ce point à tout autre point en un nombre fini d'étapes. Plus formellement:

Définition 1.6. Une chaîne de Markov est dite irréductible si pour tout $x, y \in \Omega$, il existe $t \geq 1$ tel que $P^t(x, y) > 0$.

L'hypothèse d'irréductibilité va garantir que la distribution limite sera unique, et qu'elle sera positive en chaque point.

La notion d'apériodicité, tout aussi importante, est moins triviale à exposer. Introduisons-la par l'exemple d'une marche aléatoire sur $\mathbb{Z}/4\mathbb{Z}$.

Exemple 1.7. Notre promeneur amnésique décide cette fois-ci de se balader sur $\mathbb{Z}/4\mathbb{Z}$. Il commence sur un point (par exemple 2), et à chaque tour, choisit uniformément l'un des deux éléments à distance 1 (ici 1 ou 3) et s'y place. Alors l'on remarque que s'il commence sur un état pair (0 ou 2) en $t = 0$, il sera toujours sur un état impair en temps impair, et sur un état pair en temps pair. C'est cette propriété qui correspond à la périodicité d'une chaîne de Markov, et que l'on va essayer d'éliminer car elle empêche la convergence de la distribution de X_t pour t tendant vers l'infini.

Définissons ces notions de manière plus rigoureuse.

Définition 1.8. Pour tout $x \in \Omega$, on définit l'ensemble des temps de retour \mathbb{T}_x par $\mathbb{T}_x = \{t \geq 1 \mid P^t(x, x) > 0\}$. On appelle *période de l'état x* le plus grand commun diviseur des éléments de \mathbb{T}_x .

On voit que l'ensemble \mathbb{T}_x correspond aux temps auxquels il est possible de revenir au point de départ x .

Proposition 1.9. *Les états d'une chaîne irréductible ont tous la même période.*

Preuve. Soient $x, y \in \Omega$. Par irréductibilité, il existent $r, l \in \mathbb{N}^*$ tels que $P^r(x, y) > 0$ et $P^l(y, x) > 0$. Posons $m = l + r$. Alors puisque $P^m(y, y) \geq P^l(y, x)P^r(x, y) > 0$ et similairement $P^m(x, x) \geq P^r(x, y)P^l(y, x) > 0$ l'on a $m \in \mathbb{T}_x \cap \mathbb{T}_y$, et de plus $\mathbb{T}_x \subseteq \mathbb{T}_y - m$. En effet, si $s \in \mathbb{T}_x$, alors $P^{s+m}(y, y) \geq P^l(y, x)P^s(x, x)P^r(x, y) > 0$. Donc le pgcd de \mathbb{T}_y divise tous les éléments de \mathbb{T}_x . Ainsi $\text{pgcd}(\mathbb{T}_y) \leq \text{pgcd}(\mathbb{T}_x)$. Par un argument symétrique, on obtient l'inégalité inverse. \square

Définition 1.10. On appelle *période* d'une chaîne irréductible la période de ses états. On dit qu'une chaîne est *apériodique* si elle est de période 1; autrement, on dit qu'elle est *périodique*.

La périodicité d'une chaîne l'empêche de converger vers une distribution de probabilité, puisque la distribution de X_t va dépendre fortement de la classe de t modulo la période. Pour rendre une chaîne apériodique, on peut introduire la version paresseuse ("lazy version" en anglais) de la chaîne. Pour cela, on remplace sa matrice de transition P par la matrice $Q = \frac{P+I}{2}$. Intuitivement, cela correspond à donner à la chaîne, à tout moment, la possibilité de rester sur place. Ainsi, on se débarrasse du problème de périodicité, puisque $1 \in \mathbb{T}_x$ et donc $\text{pgcd}(\mathbb{T}_x) = 1$ pour tout $x \in \Omega$.

Exemple 1.11. Reprenons l'exemple de la marche aléatoire sur $\mathbb{Z}/4\mathbb{Z}$. On peut supposer que notre promeneur amnésique est très paresseux, et qu'à chaque temps, il fait une sieste d'un temps sur son état actuel avec probabilité $\frac{1}{2}$. La probabilité de passer sur l'un des deux états voisins est donc de $\frac{1}{4}$ pour chacun des deux états. Sa marche n'est plus périodique!



Figure 1: La marche aléatoire symétrique simple est périodique sur $\mathbb{Z}/10\mathbb{Z}$, mais apériodique sur $\mathbb{Z}/9\mathbb{Z}$. L'image provient de [2], page 6.

L'irréductibilité et l'apériodicité donnent lieu à la proposition utile suivante.

Proposition 1.12. *Si (X_t) est une chaîne de Markov irréductible et apériodique, alors il existe $s \in \mathbb{N}^*$ tel que $P^t(x, y) > 0$ pour tout $x, y \in \Omega$ et $t \geq s$.*

Preuve. Pour cette preuve, nous allons utiliser le lemme suivant (il s'agit du Lemme 1.27 dans [2]): si $S \subset \mathbb{N}^*$ a pour pgcd 1 et est stable par addition, alors S contient tous les entiers à partir d'un certain rang.

Pour $x \in \Omega$, comme la chaîne est apériodique, le pgcd de \mathbb{T}_x vaut 1. L'ensemble \mathbb{T}_x est stable par addition: si $s, t \in \mathbb{T}_x$, alors $P^{s+t}(x, x) \geq P^s(x, x)P^t(x, x) > 0$ et donc $s + t \in \mathbb{T}_x$. Ainsi, il existe $t(x) \in \mathbb{N}^*$ tel que si $t \geq t(x)$, alors $t \in \mathbb{T}_x$. Par irréductibilité de la chaîne, on sait que pour tout $y \in \Omega$, il existe $r = r(x, y) \in \mathbb{N}^*$ tel que $P^r(x, y) > 0$. Ainsi, pour $t \geq t(x) + r(x, y)$, $P^t(x, y) \geq P^{t-r}(x, x)P^r(x, y) > 0$. Par conséquent, si $t \geq t'(x) := t(x) + \max_{y \in \Omega} r(x, y)$, il vient que $P^t(x, y) > 0$ pour tout $y \in \Omega$. En posant $s := \max_{x \in \Omega} t'(x)$ et $t \geq s$, on a donc que $P^t(x, y) > 0$ pour tout $x, y \in \Omega$. \square

1.3 Représentation d'une chaîne de Markov par une suite récurrente aléatoire

Il existe une deuxième manière de représenter une chaîne de Markov, appelée "random mapping representation" en anglais, ce que nous traduisons assez librement par le titre ci-dessus.

L'idée est la suivante: plutôt que de décrire une chaîne de Markov à travers sa matrice de transition, on peut décrire la variation qui s'opère entre chaque état par une fonction d'une variable aléatoire, dont on peut facilement connaître la loi.

Définition 1.13. On appelle *représentation par une fonction aléatoire* d'une matrice de transition P sur un espace d'états Ω le triplet (Λ, f, Z) , où Λ est un ensemble, f est une fonction $f : \Omega \times \Lambda \rightarrow \Omega$ et Z une variable aléatoire à valeurs dans Λ , qui satisfait

$$\mathbb{P}(f(x, Z) = y) = P(x, y)$$

pour tout $x, y \in \Omega$.

On pourra vérifier que si (Λ, f, Z) est une représentation par une fonction aléatoire d'une matrice de transition P , alors si $(Z_t)_{t \geq 1}$ est une suite de variables aléatoires iid de même loi que Z , que X_0 a pour loi μ_0 et est indépendante de toute la suite (Z_t) , alors la suite de variables aléatoires $(X_t)_{t \geq 0}$ définie par

$$\forall t \geq 1, \quad X_t = f(X_{t-1}, Z_t)$$

est une chaîne de Markov de matrice de transition P et de distribution initiale μ_0 . Une représentation de la chaîne (X_t) sous cette forme sera dite *représentation par une suite récurrente aléatoire*.

Proposition 1.14. *Pour toute chaîne de Markov sur un espace d'états fini, il existe une représentation par une suite récurrente aléatoire.*

Preuve. La construction est technique et peut être vue en [2] (Proposition 1.5). □

Cette représentation n'est pas unique, contrairement à la matrice P . Donnons un exemple qui fera apparaître le caractère très naturel de ce concept.

Exemple 1.15. Rendons encore une fois visite à notre promeneur amnésique. La matrice de transition P de sa marche aléatoire, variante non paresseuse, sur $\mathbb{Z}/4\mathbb{Z}$ est donnée par

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

On peut également se représenter sa promenade de manière suivante: à chaque tour, notre promeneur lance une pièce équilibrée, portant le chiffre 1 sur une face, et -1 sur l'autre; il ajoute le chiffre qu'il obtient et effectue la division modulo 4. Ainsi, une représentation par une fonction aléatoire de P est donnée par le triplet (Λ, f, Z) où $\Lambda = \{-1, 1\}$,

$$f : \mathbb{Z}/4\mathbb{Z} \times \{-1, 1\} \rightarrow \mathbb{Z}/4\mathbb{Z}, \quad (x, z) \mapsto x + z \pmod{4}$$

et Z une variable aléatoire à valeurs dans $\{-1, 1\}$ telle que $\mathbb{P}(Z = 1) = \mathbb{P}(Z = -1) = \frac{1}{2}$.

1.4 Existence et unicité de la distribution stationnaire d'une chaîne de Markov

Définition 1.16. On appelle *distribution stationnaire* d'une chaîne de Markov une distribution de probabilité π sur Ω vérifiant $\pi = \pi P$.

Intuitivement, la distribution stationnaire va correspondre à la distribution vers laquelle se dirige la chaîne de Markov à long terme. En effet, si la suite $(\mu_t)_{t \geq 0}$ de distributions d'une chaîne de Markov $(X_t)_{t \geq 0}$ converge (en chaque point) vers une distribution π , alors vu que $\mu_{t+1} = P\mu_t$, il est nécessaire que $\pi = \pi P$ (par passage à la limite). On remarque aussi que si $\mu_0 = \pi$, alors pour tout t , on a $\mu_t = \pi$, d'où la qualification de *stationnaire*. Venons-en à présent à un théorème extrêmement important, qui assure l'existence et l'unicité de la distribution de probabilité stationnaire. Pour le prouver, nous devons cependant commencer par introduire d'autres notions.

Définition 1.17. On appelle temps positif de première arrivée en x la variable aléatoire réelle à valeurs dans $\mathbb{N}^* \cup \{\infty\}$ défini par

$$\tau_x^+ := \min \{t \geq 1 \mid X_t = x\}.$$

Si $X_0 = x$, on parle aussi de temps de premier retour en x .

Proposition 1.18. Pour une chaîne de Markov irréductible, l'on a $\mathbb{E}_x[\tau_y^+] < +\infty$ pour tout $x, y \in \Omega$.

Preuve. Comme la chaîne est irréductible, il existe un entier r et $\varepsilon > 0$ tels que pour $x, y \in \Omega$, il existe $1 \leq j \leq r$ un entier vérifiant $P^j(x, y) > \varepsilon$. Cela signifie donc que pour toute valeur de X_t , la probabilité de toucher y entre l'instant t et l'instant $t+r$ vaut au moins ε . Ainsi, pour $k > 0$, il vient que

$$\begin{aligned} \mathbb{P}_x(\tau_y^+ > kr) &= \sum_{z \neq y} \mathbb{P}_x(X_{(k-1)r} = z, X_{(k-1)r+1} \neq y, \dots, X_{kr} \neq y, \tau_y^+ > (k-1)r) = \\ &= \sum_{z \neq y} \mathbb{P}_x(X_{(k-1)r+1} \neq y, \dots, X_{kr} \neq y \mid X_{(k-1)r} = z) \mathbb{P}_x(X_{(k-1)r} = z, \tau_y^+ > (k-1)r) \leq \\ &\leq \sum_{z \neq y} (1 - \varepsilon) \mathbb{P}_x(X_{(k-1)r} = z, \tau_y^+ > (k-1)r) = (1 - \varepsilon) \mathbb{P}_x(\tau_y^+ > (k-1)r) \end{aligned}$$

où l'on a utilisé la propriété de Markov en remarquant que l'événement $\{\tau_y^+ > (k-1)r\}$ dépend uniquement des X_t avec $1 \leq t \leq (k-1)r$. Une récurrence immédiate permet ainsi d'affirmer que

$$\mathbb{P}_x(\tau_y^+ > kr) \leq (1 - \varepsilon)^k.$$

En utilisant une expression bien connue de l'espérance pour les variables aléatoires discrètes positives, et comme $\mathbb{P}_x(\tau_y^+ > t)$ est décroissante en t , la dernière inégalité est suffisante pour majorer l'espérance de τ_y^+ ; on a

$$\mathbb{E}_x(\tau_y^+) = \sum_{t \geq 0} \mathbb{P}_x(\tau_y^+ > t) \leq \sum_{k \geq 0} r \mathbb{P}_x(\tau_y^+ > kr) \leq r \sum_{k \geq 0} (1 - \varepsilon)^k = \frac{r}{\varepsilon} < \infty.$$

□

Avant de construire explicitement la distribution stationnaire d'une chaîne de Markov, il va s'avérer utile de prouver qu'une telle distribution sera unique sous l'hypothèse d'irréductibilité. Cela nécessite un résultat intermédiaire sur les fonctions dites harmoniques.

Définition 1.19. On dit qu'une fonction $h : \Omega \rightarrow \mathbb{R}$ est *harmonique en x* si

$$h(x) = \sum_{y \in \Omega} P(x, y)h(y).$$

Proposition 1.20. Soit P la matrice de transition d'une chaîne de Markov irréductible. Une fonction harmonique en tout point de Ω est constante.

Preuve. Comme Ω est fini, il existe x_0 tel que h y atteigne son maximum. Ainsi, si pour $z \in \Omega$ tel que $P(x_0, z) > 0$ on avait $h(z) < M$, alors

$$h(x_0) = P(x_0, z)h(z) + \sum_{y \neq z} P(x_0, y)h(y) < M$$

ce qui est absurde. Donc $h(z) = M$ pour tous les états tels que $P(x_0, z) > 0$. En itérant cet argument un nombre fini de fois, l'irréductibilité permet de conclure. \square

On peut à présent prouver l'unicité de la distribution stationnaire lorsqu'elle existe.

Proposition 1.21. Soit P la matrice de transition d'une chaîne de Markov irréductible. S'il existe une distribution stationnaire de P , alors elle est unique.

Preuve. On voit qu'une fonction $h : \Omega \rightarrow \mathbb{R}$, écrit sous forme de vecteur colonne, est harmonique en tout point de Ω si et seulement si $h = Ph$. Comme on a montré que toute telle fonction était constante, et comme toute fonction constante est harmonique en tout point, on en tire que $\dim \ker(P - \text{Id}) = 1$, et donc que $\text{rg}(P - \text{Id}) = |\Omega| - 1$ par le théorème du rang. Comme le rang d'une matrice carrée est égal au rang de sa transposée, on en déduit que l'espace des solutions de $\nu P = \nu$ est de dimension 1. Par conséquent, cet espace contient au plus un vecteur dont la somme des coefficients vaut 1, d'où le résultat. \square

On peut à présent énoncer le théorème suivant (qui donne une construction explicite de la distribution stationnaire!).

Théorème 1.22. Soit (X_t) une chaîne de Markov irréductible. Alors il existe une distribution de probabilité π sur Ω telle que

$$(i) \quad \pi = \pi P.$$

$$(ii) \quad \pi(x) = \frac{1}{\mathbb{E}_x(\tau_x^+)} \text{ et donc en particulier } \pi(x) > 0 \text{ pour tout } x \in \Omega.$$

Preuve. Soit $z \in \Omega$ un état quelconque. Définissons pour $y \in \Omega$ et $t \geq 0$ la variable aléatoire

$$N_z^y(t) := \begin{cases} 1 & \text{si } t < \tau_z^+ \text{ et } X_t = y \\ 0 & \text{sinon} \end{cases},$$

de sorte que $N_z^y := \sum_{t=0}^{\infty} N_z^y(t)$ est égal au nombre de visites en y avant de retourner en z . Posons $\tilde{\pi}(y)$ le nombre moyen de telles visites, i.e.

$$\begin{aligned} \tilde{\pi}(y) &:= \mathbb{E}_z(N_z^y) \\ &= \sum_{t=0}^{\infty} \mathbb{E}_z(N_z^y(t)) = \sum_{t=0}^{\infty} \mathbb{P}_z(X_t = y, \tau_z^+ > t). \end{aligned}$$

Puisque $N_z^y \leq \tau_z^+$, on a que $\tilde{\pi}(y) \leq \mathbb{E}_z(\tau_z^+)$ pour tout $y \in \Omega$. Ainsi, il vient par la Proposition 1.18 que $\tilde{\pi}(y) < \infty$. Vérifions que $\tilde{\pi}$ est stationnaire, i.e. que $\tilde{\pi} = \tilde{\pi}P$. On a

$$\sum_{x \in \Omega} \tilde{\pi}(x)P(x, y) = \sum_{x \in \Omega} \sum_{t=0}^{\infty} \mathbb{P}(X_t = x, \tau_z^+ > t)P(x, y).$$

Comme l'événement $\{\tau_z^+ \geq t+1\} = \{\tau_z^+ > t\}$ est déterminé uniquement par X_0, \dots, X_t , on obtient par la propriété de Markov que

$$\mathbb{P}(X_t = x, X_{t+1} = y, \tau_z^+ \geq t+1) = \mathbb{P}_z(X_t = x, \tau_z^+ \geq t+1)P(x, y)$$

et par conséquent

$$\sum_{x \in \Omega} \tilde{\pi}(x)P(x, y) = \sum_{t=0}^{\infty} \mathbb{P}(X_{t+1} = y, \tau_z^+ \geq t+1) = \sum_{t=1}^{\infty} \mathbb{P}(X_t = y, \tau_z^+ \geq t).$$

On peut ainsi écrire

$$\begin{aligned} \sum_{t=1}^{\infty} \mathbb{P}(X_t = y, \tau_z^+ \geq t) &= \tilde{\pi}(y) - \mathbb{P}_z(X_0 = y, \tau_z^+ > 0) + \sum_{t=1}^{\infty} \mathbb{P}_z(X_t = y, \tau_z^+ = t) \\ &= \tilde{\pi}(y) - \mathbb{P}_z(X_0 = y) + \mathbb{P}_z(X_{\tau_z^+} = y). \end{aligned}$$

Or, si $y = z$, les deux derniers termes valent 1 et s'annulent. Si $y \neq z$, les deux derniers termes sont nuls. On a donc

$$\sum_{x \in \Omega} \tilde{\pi}(x)P(x, y) = \sum_{t=1}^{\infty} \mathbb{P}(X_t = y, \tau_z^+ \geq t) = \tilde{\pi}(y)$$

at ainsi $\tilde{\pi} = \tilde{\pi}P$. Il reste à normaliser $\tilde{\pi}$ par $\sum_{x \in \Omega} \tilde{\pi}(x) = \mathbb{E}_z(\tau_z^+)$ pour bien avoir une distribution de probabilité; il suffit donc de poser

$$\pi(x) := \frac{\tilde{\pi}(x)}{\mathbb{E}_z(\tau_z^+)}$$

pour obtenir une distribution stationnaire. Comme $\tilde{\pi}(z) = 1$ on a en particulier

$$\pi(z) = \frac{1}{\mathbb{E}_z(\tau_z^+)}.$$

Comme par la Proposition 1.21 π est unique, π ne dépend pas du point z choisit au début de la preuve. Par conséquent, on obtient

$$\pi(x) = \frac{1}{\mathbb{E}_x(\tau_x^+)},$$

et en particulier, par la Proposition 1.18, que $\pi(x) > 0$, pour tout $x \in \Omega$. □

On vient donc d'arriver à un résultat important: pour toute chaîne de Markov irréductible, nous sommes en mesure de construire explicitement la distribution stationnaire. La preuve était un peu technique, mais elle a l'avantage de montrer le fait très intuitif que la valeur de la distribution stationnaire en un point x correspond au de temps moyen que passe la chaîne en x .

1.5 Réversibilité

Il nous faut encore présenter un dernier concept utile pour l'étude des chaînes de Markov.

Définition 1.23. On dit qu'une distribution de probabilité π vérifie les *équations d'équilibre détaillées* si

$$\forall x, y \in \Omega : \quad \pi(x)P(x, y) = \pi(y)P(y, x).$$

Proposition 1.24. *Une telle distribution est stationnaire.*

Preuve. Pour tout $y \in \Omega$, en sommant les deux parties de l'équation d'équilibre détaillée, il vient que

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

car P est stochastique, ce qui conclut la démonstration. \square

Une chaîne irréductible dont la distribution stationnaire satisfait les équations d'équilibre détaillées est dit réversible, puisque dans ce cas $\mathbb{P}_\pi(X_0 = x_0, \dots, X_t = x_t) = \mathbb{P}_\pi(X_0 = x_t, \dots, X_t = x_0)$. Si l'on veut vérifier qu'une distribution est stationnaire pour une chaîne de Markov donnée, on peut donc d'abord examiner si les équations d'équilibre détaillées sont vérifiées, et donc le cas échéant éviter des calculs plus compliqués.

1.6 L'exemple de la marche aléatoire sur l'hypercube

Nous illustrons maintenant toutes les notions introduites dans ce chapitre à l'aide de l'exemple de la marche aléatoire paresseuse sur l'hypercube. Définissons tout d'abord celle-ci de manière rigoureuse.

Définition 1.25. La marche aléatoire paresseuse sur l'hypercube de dimension n est une chaîne de Markov $(X_t)_{t \geq 0}$ d'espace d'états $\Omega = \{0, 1\}^n$ dont la matrice de transition P est donnée par

$$P(x, y) = \begin{cases} \frac{1}{2} & \text{si } y = x \\ \frac{1}{2n} & \text{si } y = x + e_i \text{ pour un } i \in \llbracket 1, n \rrbracket \\ 0 & \text{sinon,} \end{cases}$$

où l'on a noté $e_i \in \Omega$ le vecteur dont toutes les coordonnées valent 0 sauf la i -ième qui vaut 1; l'addition est celle de la structure de groupe abélien donnée par l'identification de Ω à $(\mathbb{Z}/2\mathbb{Z})^n$.

Remarque 1.26. La marche du promeneur amnésique paresseux sur $\mathbb{Z}/4\mathbb{Z}$ qu'on a vu dans l'Exemple 1.11 correspond, à une bijection de l'espace d'états près, à $n = 2$ dans la définition ci-dessus.

Cette définition abstraite ne parle peut-être pas au lecteur fatigué par de nombreuses pages traitant de concepts généraux. On peut décrire cette marche aléatoire de la manière suivante.

On se place sur un point de l'hypercube de dimension n . À chaque étape, on lance une pièce équilibrée indépendante. Si le lancer donne pile, on ne bouge pas. Si le lancer donne face, on choisit d'une manière aléatoire et uniforme, indépendante du lancer (et de tous les lancers et choix avant cette étape) une coordonnée; si celle-ci vaut 0, on la remplace par 1, et inversement si elle vaut 1. Bien que cette description soit très riche, il va s'avérer utile de la modifier un petit peu, pour

obtenir la description équivalente suivante.

À chaque étape, on joue à pile ou face avec une pièce équilibrée et on choisit d'une manière aléatoire et uniforme une coordonnée (ces deux actions sont indépendantes entre elles et de tout ce qui s'est passé avant). Si le lancer donne pile, la coordonnée est remplacée par 0. Si le lancer donne face, elle est remplacée par 1.

Exemple 1.27. Cette description correspond exactement à la représentation de la chaîne de Markov par une suite récurrente aléatoire. En effet, soit $\Lambda = \{0, 1\} \times [[1, n]]$ et soit $f : \Omega \times \Lambda \rightarrow \Omega$ définie par

$$f(x, (b, k)) = (x_1, \dots, x_{k-1}, b, x_{k+1}, \dots, x_n).$$

Soit $(Z_t)_{t \geq 1}$ une suite de variables aléatoires iid à valeurs dans Λ et de loi uniforme, et soit X_0 une variable aléatoire à valeur dans Ω , indépendante de toute la suite (Z_t) . Alors la suite de variables aléatoires $(X_t)_{t \geq 0}$ définie récursivement par

$$\forall t \geq 1 : \quad X_t = f(X_{t-1}, Z_t)$$

a comme matrice de transition P :

$$\begin{aligned} P(x, y) &= \mathbb{P}(f(x, Z_1) = y) = \mathbb{P}(Z_1 \in \{(b, k) \in \Lambda \mid (x_1, \dots, x_{k-1}, b, x_{k+1}, \dots, x_n) = y\}) = \\ &= \frac{|\{(b, k) \in \Lambda \mid (x_1, \dots, x_{k-1}, b, x_{k+1}, \dots, x_n) = y\}|}{2n} =: \frac{Q(x, y)}{2n} \end{aligned}$$

Si $y = x$ on a $Q(x, y) = |\{(y_k, k) \in \Lambda \mid k \in [[1, n]]\}| = n$. Si $y = x + e_i$ pour un $i \in [[1, n]]$, alors $Q(x, y) = |\{(y_i, i)\}| = 1$ et si x et y différent de plus de deux coordonnées, on a que $Q(x, y) = 0$. Il s'agit donc effectivement de la marche aléatoire paresseuse sur l'hypercube de dimension n .

Proposition 1.28. *La marche aléatoire paresseuse sur l'hypercube de dimension n est une chaîne irréductible et apériodique. Sa distribution stationnaire $\pi^{(n)}$ est la distribution uniforme sur $\{0, 1\}^n$, i.e. $\pi^{(n)}(x) = \frac{1}{2^n}$ pour tout $x \in \{0, 1\}^n$, et la chaîne est réversible.*

Preuve. Soient $x, y \in \{0, 1\}^n$ et prouvons qu'il existe $r \geq 1$ tel que $P^r(x, y) > 0$. Si $x = y$ il suffit de prendre $r = 1$, supposons donc $x \neq y$. Il existe donc $r \geq 1$ et $1 \leq i_1 < \dots < i_r \leq n$ tel que $y = x + e_{i_1} + \dots + e_{i_r}$; on prend simplement $\{i_1, \dots, i_r\} = \{i \in [[1, n]] \mid x_i \neq y_i\}$. Pour $j \in [[1, r]]$, on pose $x_j = x + e_{i_1} + \dots + e_{i_j}$, ainsi que $x_0 = x$. Comme donc $x_j = x_{j-1} + e_{i_j}$ pour tout $j \in [[1, r]]$, on a par définition de P que $P(x_{j-1}, x_j) = \frac{1}{2^n} > 0$. Par conséquent, $P^r(x, y) = P^r(x_0, x_r) \geq P(x_0, x_1) \cdots P(x_{r-1}, x_r) > 0$, et donc la chaîne est irréductible. Puisque la chaîne est paresseuse, elle est apériodique.

Finalement, puisque pour tout $x, y \in \{0, 1\}^n$ et $i \in [[1, n]]$ l'on a $y = x + e_i \iff x = y + e_i$, P est symétrique, et donc satisfait les équations d'équilibre détaillées pour la distribution uniforme. Par conséquent, celle-ci est stationnaire, et la chaîne est réversible. \square

2 Temps de mélange et *cut-off*

2.1 Une distance sur l'espace des distributions de probabilité: la distance de variation totale

Dans les parties précédentes, nous avons à plusieurs reprises évoqué la notion de convergence vers la distribution stationnaire, i.e. que la distribution de la chaîne à un instant t converge en tout

point vers la distribution stationnaire pour t tendant vers l'infini. Notre but sera d'examiner la vitesse de cette convergence quand elle a lieu, et pour cela il nous faut une distance bien choisie sur l'espace des distributions de probabilités sur Ω . Nous allons nous attaquer à ce problème, qui nous permettra ensuite de préciser les notions de temps de mélange et de *cut-off*.

Définition 2.1. On appelle distance de variation totale entre deux mesures de probabilité μ et ν sur Ω la distance

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

Remarquons que la définition de cette distance est purement probabiliste: on compare deux distributions de probabilité sur la différence maximale de probabilité pour un même événement. Cette formule n'est pas toujours la plus utile en pratique; donnons donc trois formulations alternatives.

Proposition 2.2. Soient μ et ν deux mesures de probabilité sur Ω , alors

$$\|\mu - \nu\|_{\text{TV}} = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|.$$

Preuve. Soit $B = \{x : \mu(x) > \nu(x)\}$. Soit $A \subset \Omega$. Alors

$$\mu(A) - \nu(A) \leq \mu(A \cap B) - \nu(A \cap B) \leq \mu(B) - \nu(B).$$

Un raisonnement parallèle donne

$$\nu(A) - \mu(A) \leq \nu(B^c) - \mu(B^c).$$

Or, il se trouve que ces deux bornes sont égales, car $\mu(B^c) = 1 - \mu(B)$ et de même pour ν . Enfin, on remarque que le cas d'égalité correspond au choix $A = B$, et ainsi

$$\|\mu - \nu\|_{\text{TV}} = \frac{1}{2} (\mu(B) - \mu(B^c) + \nu(B^c) - \nu(B)) = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|.$$

□

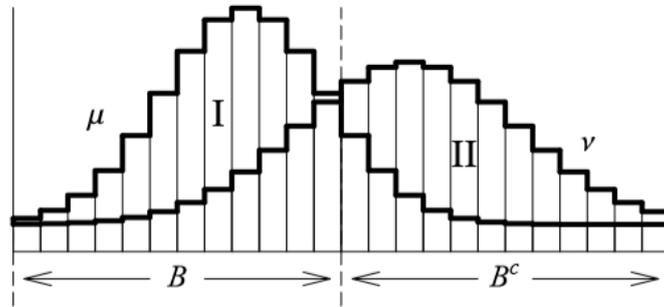


Figure 2: Comme dans la preuve ci-dessus, on pose $B = \{x : \mu(x) > \nu(x)\}$. L'aire I vaut $\mu(B) - \nu(B)$, celle de II vaut $\nu(B^c) - \mu(B^c)$. Par la preuve, les deux sont égales, et valent $\|\mu - \nu\|_{\text{TV}}$. La figure provient de [2], page 48.

Proposition 2.3. Soient μ, ν deux mesures de probabilité sur Ω , alors

$$\|\mu - \nu\|_{\text{TV}} = \sum_{\substack{x \in \Omega \\ \mu(x) \geq \nu(x)}} \mu(x) - \nu(x).$$

Preuve. La preuve découle de la preuve précédente (voire Figure 2). \square

On verra une troisième formule utile pour $\|\cdot\|_{\text{TV}}$ dans le sous-chapitre 3.1 qui fait intervenir la notion de couplage (Lemme 3.2 et Remarque 3.3).

Ayant défini une distance sur l'espace des mesures de probabilité, nous pouvons maintenant l'introduire dans le cadre plus spécifique des chaînes de Markov.

Définition 2.4. Pour une chaîne de Markov irréductible de matrice de transition P et distribution stationnaire π , on appelle, par abus de langage, *distance de variation totale* la fonction $d : \mathbb{N} \rightarrow [0, 1]$ définie par

$$d(t) = \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{\text{TV}}$$

pour tout entier t . On définit de plus $\bar{d} : \mathbb{N} \rightarrow [0, 1]$ par

$$\bar{d}(t) = \max_{x, y \in \Omega} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{\text{TV}}$$

pour tout entier t .

Remarque 2.5. On peut vérifier que pour tout distribution initiale μ_0 d'une chaîne de Markov (X_t) irréductible l'on a $\|\mathbb{P}_{\mu_0}(X_t = \cdot) - \pi\|_{\text{TV}} \leq d(t)$ pour tout $t \geq 0$, et donc d permet de majorer la distance de variation totale de toute chaîne à sa distribution stationnaire.

Il y a deux raisons pour introduire \bar{d} . Premièrement, on a l'encadrement $\bar{d}/2 \leq d \leq \bar{d}$ (la preuve est élémentaire et peut être vue en [2], Lemme 4.11), et parfois il est utile de ne pas avoir à faire référence à la distribution stationnaire. Deuxièmement, \bar{d} est sous-multiplicative dans le sens où $\bar{d}(s+t) \leq \bar{d}(s)\bar{d}(t)$ pour tout $t, s \geq 0$ (la preuve est technique et peut être vue en [2], Lemme 4.12), ce que, combiné avec $d \leq \bar{d}$, peut être utile.

Pour des raisons similaires, il convient enfin d'introduire une deuxième distance, appelée *distance de séparation*.

Définition 2.6. Pour une chaîne de Markov irréductible de matrice de transition P et distribution stationnaire π , on appelle, par abus de langage, *distance de séparation* la fonction $s : \mathbb{N} \rightarrow [0, 1]$ définie par

$$s(t) := \max_{x, y \in \Omega} \left[1 - \frac{P^t(x, y)}{\pi(y)} \right].$$

L'utilité principale ici de s sera la suivante.

Proposition 2.7. On a $d \leq s$.

Preuve. En effet,

$$\begin{aligned} \|P^t(x, \cdot) - \pi\|_{\text{TV}} &= \sum_{\substack{y \in \Omega \\ P^t(x, y) < \pi(y)}} [\pi(y) - P^t(x, y)] = \sum_{\substack{y \in \Omega \\ P^t(x, y) < \pi(y)}} \pi(y) \left[1 - \frac{P^t(x, y)}{\pi(y)} \right] \leq \\ &\leq \max_{y \in \Omega} \left[1 - \frac{P^t(x, y)}{\pi(y)} \right] \end{aligned}$$

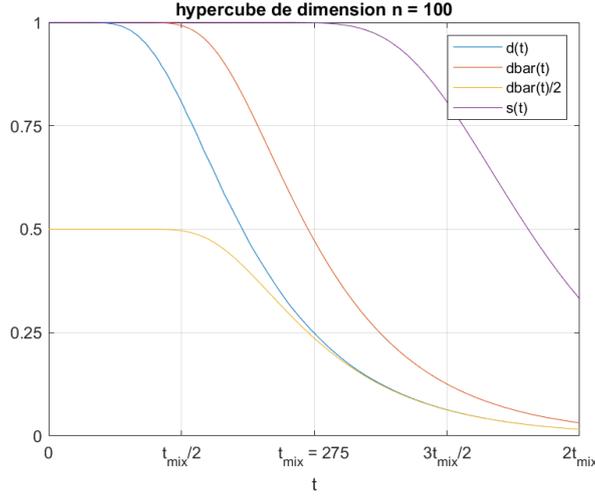


Figure 3: Les différentes distances définies ci-dessus pour l'exemple de l'hypercube en dimension $n = 100$, voire Définition 1.25. Voir les sous-chapitres A.1, A.2, A.3 et A.4 de l'appendice pour les idées et le code utilisé pour créer ces images.

et on conclut en prenant le maximum sur y des deux côtés. □

Proposition 2.8. *Pour tout chaîne irréductible, les fonctions d , \bar{d} et s sont décroissantes.*

Preuve. Pour deux mesures de probabilité μ, ν et une matrice stochastique P l'on a

$$\begin{aligned} \|\mu P - \nu P\|_{\text{TV}} &= \frac{1}{2} \sum_{y \in \Omega} \left| \sum_{x \in \Omega} [\mu(x)P(x, y) - \nu(x)P(x, y)] \right| \leq \\ &\leq \frac{1}{2} \sum_{x \in \Omega} \sum_{y \in \Omega} [|\mu(x) - \nu(x)|P(x, y)] = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)| = \|\mu - \nu\|_{\text{TV}} \end{aligned}$$

d'où la décroissance de d et \bar{d} . Pour la décroissance de s , on remarque que

$$\frac{P^{t+1}(x, y)}{\pi(y)} = \sum_{z \in \Omega} P(x, z) \frac{P^t(z, y)}{\pi(y)} \geq \min_{x', y' \in \Omega} \frac{P^t(x', y')}{\pi(y')} \sum_{z \in \Omega} P(x, z) = \min_{x', y' \in \Omega} \frac{P^t(x', y')}{\pi(y')}$$

et on conclut en prenant le minimum sur x, y . □

2.2 Le théorème de convergence

Nous allons à présent énoncer un fait fondamental: toute chaîne de Markov raisonnable (apériodique et irréductible) converge vers sa distribution stationnaire. Cette convergence est d'ordre exponentiel.

Théorème 2.9. *Soit P la matrice de transition d'une chaîne de Markov irréductible et apériodique, de distribution stationnaire π . Alors il existe deux constantes $C > 0$ et $\alpha \in]0, 1[$ telles que*

$$d(t) \leq C\alpha^t.$$

Preuve. La preuve peut être trouvée dans [2] (Theorem 4.9). □

Un deuxième théorème important est le suivant.

Théorème 2.10. *Il existe $\lambda_* < 1$ tel que*

$$\lim_{t \rightarrow \infty} d(t)^{1/t} = \lambda_*$$

Preuve. On remarque que l'on a pour tout $t, s \geq 0$ que

$$d(t+s) \leq \bar{d}(t+s) \leq \bar{d}(t)\bar{d}(s) \leq 4d(t)d(s)$$

et par conséquent, $(4d(t))_{t \geq 1}$ est sous-multiplicative. Ainsi, par le Lemme de Fekete, il vient que

$$\lim_{t \rightarrow \infty} d(t)^{1/t} = \lim_{t \rightarrow \infty} (4d(t))^{1/t} = \inf\{(4d(t))^{1/t} \mid t \geq 1\} =: \lambda_*.$$

Puisque, par le théorème de convergence, on a $d(t) < \frac{1}{4}$ pour t assez grand, on obtient $\lambda_* < 1$. \square

2.3 Le temps de mélange

Effectuons un bref retour à notre point de départ: nous nous étions initialement demandé combien de temps était nécessaire pour s'approcher d'un état "bien mélangé" de la chaîne. Nous y sommes presque. Nous avons défini les principaux concepts nécessaires à la compréhension des chaînes de Markov et nous avons montré qu'une chaîne de Markov raisonnable (irréductible et apériodique) convergeait exponentiellement vite, pour une distance soigneusement définie, vers sa distribution stationnaire, qui est unique.

Nous allons donc nous intéresser à la vitesse de convergence de la chaîne: comment s'approche-t-elle de sa distribution stationnaire?

Définition 2.11. On appelle ε -temps de mélange la quantité

$$t_{\text{mix}}(\varepsilon) := \inf \{t \geq 0 : d(t) \leq \varepsilon\}.$$

On fixe aussi (de manière un peu arbitraire)

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

Notons que par le théorème de convergence, les ε -temps de mélange sont fini. On définit les ε -temps de séparation $t_{\text{sep}}(\varepsilon)$ de la même façon, mais par rapport à la distance de séparation.

On vient donc de définir le temps de mélange, outil central qui désigne le temps minimal qu'il faut pour que la chaîne de Markov soit "assez proche" de sa distribution stationnaire. Les travaux récents sur le sujet tendent à mettre en avant un comportement remarquable de certaines chaînes de Markov: autour du temps de mélange t_{mix} , leur distance à la distribution stationnaire connaît une chute brutale sur un court laps de temps. C'est ce phénomène qui est qualifié de *cut-off* dans la littérature, et que nous allons à présent étudier en détail.

2.4 Le phénomène de *cut-off*

L'étape conceptuelle finale consiste maintenant à faire évoluer l'espace d'états d'une manière naturelle et ensuite d'examiner le comportement des différentes notions introduites dans les chapitres précédentes. On s'intéressera principalement au comportement des ε -temps de mélange lors de cette évolution. Ainsi, on s'intéressera par exemple au calcul asymptotique du temps de mélange

de la marche aléatoire paresseuse sur l'hypercube de dimension n , dont on a vu la définition au début du sous-chapitre 1.6.

En analysant ensuite la distance de variation totale, il s'avère que, pour cette marche aléatoire, elle décroît d'une manière très brutale autour du temps de mélange, comme on peut le voir dans la Figure 4. C'est ce phénomène que l'on qualifie de *cut-off*, dont on va donner la définition et différentes propriétés abstraites ici.

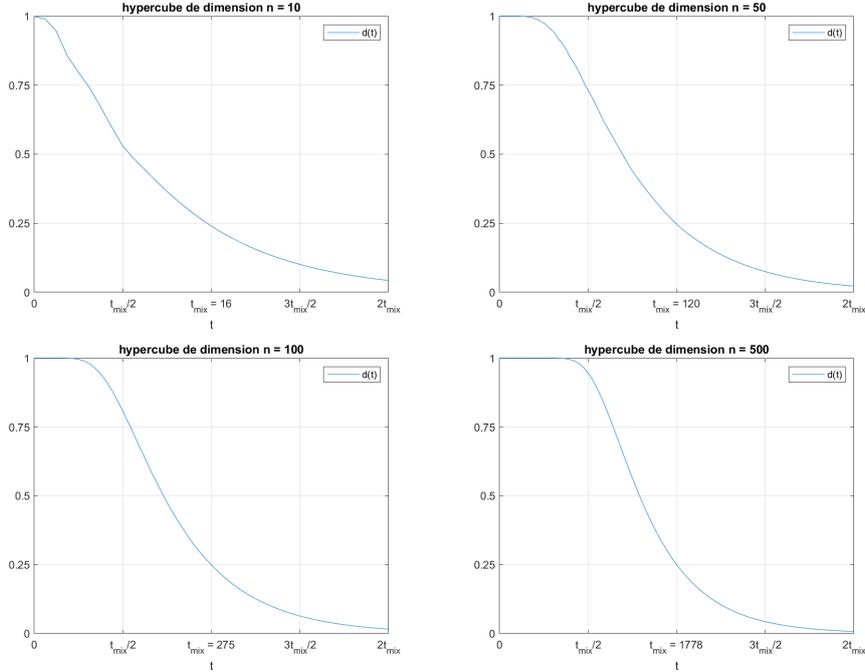


Figure 4: La distance de variation totale pour l'hypercube en différentes dimensions. On voit que la chute autour du temps de mélange devient de plus en plus raide. Voir les sous-chapitres A.1 et A.2 de l'appendice pour les idées et le code utilisé pour créer ces images.

Avant de rentrer dans les détails, faisons plusieurs remarques sur le contenu de ce sous-chapitre. Les Définitions 2.12 et 2.18 en combinaison avec le Corollaire 2.21 suffisent pour comprendre la suite. Le lecteur pressé pourrait donc se limiter à ces trois énoncés. Le reste du sous-chapitre est dédié à donner, avec preuve, diverses caractérisations équivalents du *cut-off*. Les preuves sont composées d'arguments élémentaires d'analyse réelle et n'ont donc pas un grand intérêt en soi, mais vu que ces équivalences sont admis sans démonstration (ou même sans mention) dans tout les ouvrages de notre connaissance, le lecteur sceptique trouvera ici un traitement complet de ces propriétés de base.

Formellement, faire "évoluer l'espace d'états" consiste à considérer une suite d'espaces d'états quelconques $(\Omega_n)_{n \in \mathbb{N}}$ ainsi qu'une suite de chaînes de Markov $((X_t^{(n)})_{t \geq 0})_{n \in \mathbb{N}}$, tels que pour tout $n \in \mathbb{N}$, Ω_n est l'espace d'états de $(X_t^{(n)})_{t \geq 0}$. Dans ce cadre, on notera $t_{\text{mix}}^{(n)}(\varepsilon)$ l' ε -temps de mélange et d_n la distance de variation totale associée à la chaîne de Markov $(X_t^{(n)})_{t \geq 0}$. La discussion du *cut-off* présupposera qu'il y a convergence; on suppose donc dans toute cette section que les chaînes de Markov sont irréductibles et apériodiques.

Définition 2.12. On dit qu'une suite de chaînes de Markov $((X_t^{(n)})_{t \geq 0})_{n \in \mathbb{N}}$ présente un *cut-off*

si et seulement si pour tout $\varepsilon \in]0, 1[$, on a

$$\lim_{n \rightarrow \infty} \frac{t_{\text{mix}}^{(n)}(\varepsilon)}{t_{\text{mix}}^{(n)}(1-\varepsilon)} = 1. \quad (2)$$

On impose implicitement que pour tout $\varepsilon > 0$ l'expression $\frac{t_{\text{mix}}^{(n)}(\varepsilon)}{t_{\text{mix}}^{(n)}(1-\varepsilon)}$ soit définie à partir d'un certain rang, ou en d'autres termes, que à ε fixé, les ε -temps de mélange sont non-nuls à partir d'un certain rang. De plus, pour que la suite de chaînes présente un *cut-off*, il suffit bien sûr que (2) soit vérifiée pour tout $\varepsilon \in]0, 1/2[$.

D'une manière plus explicative, une suite de chaînes de Markov présente un *cut-off* si et seulement si les suites d' ε -temps de mélange sont toutes asymptotiquement équivalentes. C'est le contenu du lemme suivant.

Lemme 2.13. *Une suite de chaînes de Markov présente un cut-off si et seulement si pour tout $\varepsilon, \varepsilon' \in]0, 1[$, on a*

$$\lim_{n \rightarrow \infty} \frac{t_{\text{mix}}^{(n)}(\varepsilon)}{t_{\text{mix}}^{(n)}(\varepsilon')} = 1. \quad (3)$$

Preuve. Si (3) est vérifiée pour tout $\varepsilon, \varepsilon' \in]0, 1[$, alors il suffit de poser $\varepsilon' = 1 - \varepsilon$ pour obtenir la présence du *cut-off*. Il reste donc à prouver l'implication inverse.

Supposons que la suite de chaînes de Markov présente un *cut-off*. Pour $\varepsilon, \varepsilon' \in]0, 1/2]$, on voit que

$$\frac{t_{\text{mix}}^{(n)}(1-\varepsilon)}{t_{\text{mix}}^{(n)}(\varepsilon)} \leq \frac{t_{\text{mix}}^{(n)}(\varepsilon')}{t_{\text{mix}}^{(n)}(\varepsilon)} \leq \frac{t_{\text{mix}}^{(n)}(\varepsilon')}{t_{\text{mix}}^{(n)}(1-\varepsilon')}$$

et donc en faisant tendre n vers l'infini, on obtient par (2) que $\lim_{n \rightarrow \infty} \frac{t_{\text{mix}}^{(n)}(\varepsilon')}{t_{\text{mix}}^{(n)}(\varepsilon)} = 1$. En utilisant à nouveau (2), on obtient donc que $(t_{\text{mix}}^{(n)}(\varepsilon))_{n \in \mathbb{N}}$, $(t_{\text{mix}}^{(n)}(1-\varepsilon))_{n \in \mathbb{N}}$, $(t_{\text{mix}}^{(n)}(\varepsilon'))_{n \in \mathbb{N}}$ et $(t_{\text{mix}}^{(n)}(1-\varepsilon'))_{n \in \mathbb{N}}$ sont tous équivalents, ce qui conclut la preuve. \square

Pour nous familiariser davantage avec cette définition abstraite, nous donnons à présent une autre caractérisation du phénomène de *cut-off*.

Lemme 2.14. *Une suite de chaînes de Markov présente un cut-off si et seulement si pour tout réel $c \neq 1$ positif*

$$\lim_{n \rightarrow \infty} d_n(ct_{\text{mix}}^{(n)}) = \begin{cases} 1 & \text{si } c < 1 \\ 0 & \text{si } c > 1 \end{cases}, \quad (4)$$

où, pour tout $n \in \mathbb{N}$, on étend d_n à $\mathbb{R}_{\geq 0}$ en posant $d_n(x) = d_n(\lfloor x \rfloor)$ pour tout $x \in \mathbb{R}_{\geq 0}$.

Concrètement, un *cut-off* a lieu si et seulement si la distance de variation totale, à une échelle normalisée par le temps de mélange, tend simplement vers la fonction en escalier qui vaut 1 sur $[0, 1[$ et 0 sur $]1, \infty[$.

Remarque 2.15. L'extension de la distance de variation totale d à $\mathbb{R}_{\geq 0}$ par $d(x) = d(\lfloor x \rfloor)$ se comporte bien avec les ε -temps de mélange, dans le sens où on a encore $t_{\text{mix}}(\varepsilon) = \inf\{x \in \mathbb{R}_{\geq 0} \mid d(x) \leq \varepsilon\}$.

Preuve. (\Rightarrow) Supposons qu'il y ait un *cut-off*. Soit $N \in \mathbb{N}$ tel que $t_{\text{mix}}^{(n)} \geq 1$ pour tout $n \geq N$. Fixons $\varepsilon \in]0, 1/4[$ quelconque. Alors par le Lemme 2.13, pour tout $\delta > 0$, il existe un $N_\delta \geq N$ tel

que pour tout $n \geq N_\delta$ on a

$$\frac{t_{\text{mix}}^{(n)}(\varepsilon)}{t_{\text{mix}}^{(n)}} < 1 + \delta, \quad \frac{t_{\text{mix}}^{(n)}(1 - \varepsilon)}{t_{\text{mix}}^{(n)}} > 1 - \delta. \quad (5)$$

Soit maintenant $c > 1$. Soit $\delta > 0$ tel que $c > 1 + \delta$. Par (5), on obtient par décroissance de d_n que

$$d_n(ct_{\text{mix}}^{(n)}) \leq d_n\left(\frac{t_{\text{mix}}^{(n)}(\varepsilon)}{t_{\text{mix}}^{(n)}} t_{\text{mix}}^{(n)}\right) \leq \varepsilon$$

pour tout $n \geq N_\delta$. Comme $\varepsilon \in]0, 1/4[$ était quelconque, on en déduit que

$$\lim_{n \rightarrow \infty} d_n(ct_{\text{mix}}^{(n)}) = 0.$$

Considérons le cas restant $c \in [0, 1[$. Soit $\delta > 0$ tel que $c < 1 - 2\delta$. En modifiant un peu (5), on voit que

$$\frac{t_{\text{mix}}^{(n)}(1 - \varepsilon) - \delta}{t_{\text{mix}}^{(n)}} > 1 - \delta - \frac{\delta}{t_{\text{mix}}^{(n)}} \geq 1 - 2\delta > c$$

pour tout $n \geq N_\delta$, puisque $t_{\text{mix}}^{(n)} \geq 1$ car $n \geq N_\delta \geq N$. Ceci nous donne alors, de nouveau par décroissance de d_n , que

$$d_n(ct_{\text{mix}}^{(n)}) \geq d_n\left(\frac{t_{\text{mix}}^{(n)}(1 - \varepsilon) - \delta}{t_{\text{mix}}^{(n)}} t_{\text{mix}}^{(n)}\right) = d_n(t_{\text{mix}}^{(n)}(1 - \varepsilon) - \delta) > 1 - \varepsilon$$

pour tout $n \geq N_\delta$. Comme $\varepsilon \in]0, 1/4[$ était quelconque, on en déduit

$$\lim_{n \rightarrow \infty} d_n(ct_{\text{mix}}^{(n)}) = 1.$$

(\Leftarrow) Supposons à présent que (4) soit vérifiée. Alors il existe $N \in \mathbb{N}$ tel que $t_{\text{mix}}^{(n)} \geq 1$ pour tout $n \geq N$ (sinon il existerait une sous-suite $t_{\text{mix}}^{(n_k)} = 0 \forall k$ et donc $\lim_{k \rightarrow \infty} d_{n_k}(ct_{\text{mix}}^{(n_k)})$ ne dépendrait pas de c , ce qui est absurde). Fixons $\varepsilon \in]0, 1/2[$ quelconque. Pour $\delta \in]0, 1[$, soit $N_\delta \geq N$ tel que pour tout $n \geq N_\delta$, l'on ait

$$d_n((1 + \delta)t_{\text{mix}}^{(n)}) < \varepsilon, \quad d_n((1 - \delta)t_{\text{mix}}^{(n)}) > 1 - \varepsilon.$$

En particulier, il vient que

$$t_{\text{mix}}^{(n)}(\varepsilon) \leq (1 + \delta)t_{\text{mix}}^{(n)} \quad \text{et} \quad t_{\text{mix}}^{(n)}(1 - \varepsilon) \geq (1 - \delta)t_{\text{mix}}^{(n)}. \quad (6)$$

Puisque $t_{\text{mix}}^{(n)} \geq 1$ pour tout $n \geq N_\delta \geq N$, on obtient donc

$$1 \leq \frac{t_{\text{mix}}^{(n)}(\varepsilon)}{t_{\text{mix}}^{(n)}(1 - \varepsilon)} \leq \frac{1 + \delta}{1 - \delta}.$$

pour tout $n \geq N_\delta$. Comme $\delta \in]0, 1[$ était quelconque, on obtient le résultat souhaité. \square

Ce lemme permet de prouver le corollaire suivant, qui nous permet d'établir à la fois le comportement asymptotique de $t_{\text{mix}}^{(n)}$ et l'existence du *cut-off*.

Corollaire 2.16. *S'il existe $(t_n)_{n \in \mathbb{N}}$ une suite de nombres positifs telle que*

$$\lim_{n \rightarrow \infty} d_n(ct_n) = \begin{cases} 1 & \text{si } c < 1 \\ 0 & \text{si } c > 1 \end{cases} \quad (7)$$

alors la suite de chaînes présente un cut-off, et de plus $\lim_{n \rightarrow \infty} \frac{t_{\text{mix}}^{(n)}}{t_n} = 1$.

Preuve. En imitant la partie (\Leftarrow) de la preuve du Lemme 2.14 précédent, on obtient la présence du cut-off. Pour obtenir $\lim_{n \rightarrow \infty} \frac{t_{\text{mix}}^{(n)}}{t_n} = 1$, on utilise les estimations (6) de cette même partie dans les cas particuliers $\varepsilon = \frac{1}{4}$ et $\varepsilon = \frac{3}{4}$. \square

Remarque 2.17. Toutes les chaînes de Markov ne présentent pas de cut-off. La Figure 5 permet notamment de voir que la marche aléatoire paresseuse sur $\mathbb{Z}/n\mathbb{Z}$ ne présente pas de cut-off.

On sait donc maintenant que la présence d'un cut-off correspond à une décroissance rapide de la distance de variation totale autour du temps de mélange, à une échelle normalisée par celui-ci. Cependant, il est parfois possible de décrire plus précisément l'intervalle sur lequel se produit la chute brutale; c'est l'objectif de la définition suivante.

Définition 2.18. Soit $(w_n)_{n \in \mathbb{N}}$ une suite de nombres réels strictement positifs. On dit qu'une suite de chaînes de Markov présente un cut-off dont la fenêtre est de taille $(w_n)_{n \in \mathbb{N}}$ si et seulement si $\lim_{n \rightarrow +\infty} \frac{t_{\text{mix}}^{(n)}}{w_n} = +\infty$ et

$$\lim_{\alpha \rightarrow -\infty} \liminf_{n \rightarrow +\infty} d_n(t_{\text{mix}}^{(n)} + \alpha w_n) = 1, \quad (8a)$$

$$\lim_{\alpha \rightarrow +\infty} \limsup_{n \rightarrow +\infty} d_n(t_{\text{mix}}^{(n)} + \alpha w_n) = 0. \quad (8b)$$

A première vue, il n'est pas évident pourquoi cette définition donne lieu à un cut-off. Prouvons-le!

Lemme 2.19. *Soit $(w_n)_{n \in \mathbb{N}}$ une suite de nombres réels strictement positifs. Si une suite de chaînes de Markov présente un cut-off dont la fenêtre est de taille $(w_n)_{n \in \mathbb{N}}$, alors elle présente un cut-off.*

Preuve. Pour $n \in \mathbb{N}$, posons $\alpha_n = \frac{t_{\text{mix}}^{(n)}}{w_n}$; par hypothèse on a $\lim_{n \rightarrow \infty} \alpha_n = +\infty$. Fixons $c \neq 1$ un réel positif. Pour $\alpha > 1$, soit $N_\alpha \in \mathbb{N}$ tel que $\alpha_n > \frac{\alpha}{|c-1|}$ (et pour que tout soit bien défini aussi $\alpha_n > \alpha$) pour tout $n \geq N_\alpha$.

Si $c < 1$, alors par décroissance de d_n on a

$$d_n(ct_{\text{mix}}^{(n)}) = d_n(t_{\text{mix}}^{(n)} + (c-1)\alpha_n w_n) \geq d_n(t_{\text{mix}}^{(n)} - \alpha w_n)$$

pour tout $n \geq N_\alpha$. Le côté gauche est bien défini car $\alpha_n > \alpha$. En passant à la lim inf, on obtient alors

$$\liminf_{n \rightarrow +\infty} d_n(ct_{\text{mix}}^{(n)}) \geq \liminf_{n \rightarrow +\infty} d_n(t_{\text{mix}}^{(n)} - \alpha w_n).$$

Comme $\alpha > 1$ était quelconque, on peut le faire tendre vers l'infini; cela nous donne par (8a) que $\liminf_{n \rightarrow +\infty} d_n(ct_{\text{mix}}^{(n)}) \geq 1$, et par conséquent $\lim_{n \rightarrow +\infty} d_n(ct_{\text{mix}}^{(n)}) = 1$.

Si $c > 1$, alors de nouveau par décroissance de d_n on a

$$d_n(ct_{\text{mix}}^{(n)}) = d_n(t_{\text{mix}}^{(n)} + (c-1)\alpha_n w_n) \leq d_n(t_{\text{mix}}^{(n)} + \alpha w_n)$$

pour tout $n \geq N_\alpha$. En passant à la \limsup , on obtient alors

$$\limsup_{n \rightarrow +\infty} d_n(ct_{\text{mix}}^{(n)}) \leq \limsup_{n \rightarrow +\infty} d_n(t_{\text{mix}}^{(n)} + \alpha w_n).$$

Comme $\alpha > 1$ était quelconque, on peut le faire tendre vers l'infini, ce qui nous donne par (8b) que $\limsup_{n \rightarrow +\infty} d_n(ct_{\text{mix}}^{(n)}) \leq 0$, et par conséquent $\lim_{n \rightarrow +\infty} d_n(ct_{\text{mix}}^{(n)}) = 0$. On conclut par le Corollaire 2.16. \square

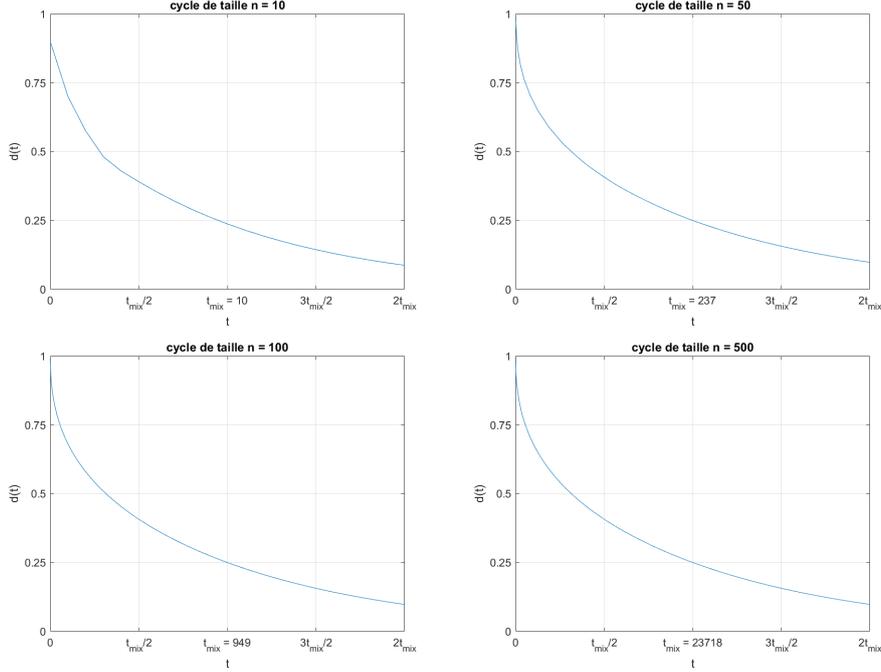


Figure 5: La distance de variation totale pour la marche aléatoire paresseuse sur le cycle $\mathbb{Z}/n\mathbb{Z}$ pour différentes valeurs de n . On voit qu'à l'échelle normalisée par le temps de mélange, la distance ne change pas qualitativement. En effet, on peut montrer que pour cette suite de chaînes il n'y pas de *cut-off*. Voir le sous-chapitre A.2 de l'appendice pour les idées et le code utilisé pour créer ces images.

Cette preuve met en avant la correspondance entre (8a) et le comportement de la distance de variation total avant le temps de mélange, ainsi que la correspondance entre (8b) et le comportement de la distance de variation total après le temps de mélange.

Néanmoins, il n'est pas très clair ce que la suite $(w_n)_{n \in \mathbb{N}}$ dit exactement de l'intervalle de la chute. Pour comprendre ceci, prouvons le lemme explicatif suivant.

Lemme 2.20. *Soit $(w_n)_{n \in \mathbb{N}}$ une suite de nombres réels strictement positifs. Une suite de chaînes de Markov présente un *cut-off* dont la fenêtre est de taille $(w_n)_{n \in \mathbb{N}}$ si et seulement si $\lim_{n \rightarrow \infty} \frac{t_{\text{mix}}^{(n)}}{w_n} = +\infty$ et si pour tout $\varepsilon \in]0, 1[$, il existe $c_\varepsilon > 0$ tel que*

$$t_{\text{mix}}^{(n)}(\varepsilon) - t_{\text{mix}}^{(n)}(1 - \varepsilon) \leq c_\varepsilon w_n \quad (9)$$

pour tout $n \in \mathbb{N}$.

Preuve. (\Rightarrow): Supposons qu'il y ait un *cut-off* dont la fenêtre est de taille $(w_n)_{n \in \mathbb{N}}$; on obtient alors directement $\lim_{n \rightarrow \infty} \frac{t_{\text{mix}}^{(n)}}{w_n} = +\infty$. Fixons $\varepsilon \in]0, 1[$. En combinant (8a) et (8b), il existe $\alpha > 0$

tel que

$$\begin{aligned}\liminf_{n \rightarrow +\infty} d_n(t_{\text{mix}}^{(n)} - \alpha w_n) &> 1 - \varepsilon, \\ \limsup_{n \rightarrow +\infty} d_n(t_{\text{mix}}^{(n)} + \alpha w_n) &< \varepsilon.\end{aligned}$$

Par conséquent, il existe $N \in \mathbb{N}$ tel que

$$\begin{aligned}d_n(t_{\text{mix}}^{(n)} - \alpha w_n) &> 1 - \varepsilon, \\ d_n(t_{\text{mix}}^{(n)} + \alpha w_n) &< \varepsilon\end{aligned}$$

pour tout $n \geq N$. Ceci donne

$$\begin{aligned}t_{\text{mix}}^{(n)} - \alpha w_n &< t_{\text{mix}}^{(n)}(1 - \varepsilon), \\ t_{\text{mix}}^{(n)} + \alpha w_n &> t_{\text{mix}}^{(n)}(\varepsilon)\end{aligned}$$

et donc $t_{\text{mix}}^{(n)}(\varepsilon) - t_{\text{mix}}^{(n)}(1 - \varepsilon) < 2\alpha w_n$ pour tout $n \geq N$, d'où cette implication.

(\Leftarrow): Supposons (9) pour tout $\varepsilon \in]0, 1[$ ainsi que $\lim_{n \rightarrow \infty} \frac{t_{\text{mix}}^{(n)}}{w_n} = +\infty$. Fixons $\varepsilon \in]0, 1/4[$. Par (9) on a $t_{\text{mix}}^{(n)}(1 - \varepsilon) > t_{\text{mix}}^{(n)}(\varepsilon) - c_\varepsilon w_n \geq t_{\text{mix}}^{(n)} - c_\varepsilon w_n$. Donc en particulier, il existe $N \in \mathbb{N}$ tel que $t_{\text{mix}}^{(n)}(1 - \varepsilon) \geq 1$ pour $n \geq N$. Posons $\delta_n = \min\{w_n, 1\}$. Il vient que

$$\frac{t_{\text{mix}}^{(n)} - t_{\text{mix}}^{(n)}(1 - \varepsilon) + \delta_n}{w_n} \stackrel{(9)}{\leq} \frac{t_{\text{mix}}^{(n)} - t_{\text{mix}}^{(n)}(\varepsilon) + c_\varepsilon w_n + \delta_n}{w_n} \leq c_\varepsilon + 1.$$

Donc, si $n \geq N$ et $\alpha \geq c_\varepsilon + 1$, l'on a

$$d_n(t_{\text{mix}}^{(n)} - \alpha w_n) \geq d_n\left(t_{\text{mix}}^{(n)} - \frac{t_{\text{mix}}^{(n)} - t_{\text{mix}}^{(n)}(1 - \varepsilon) + \delta_n}{w_n} w_n\right) = d_n(t_{\text{mix}}^{(n)}(1 - \varepsilon) - \delta_n) > 1 - \varepsilon$$

ce qui prouve (8a).

D'une manière similaire, mais plus simple, on voit que par (9), on a

$$\frac{t_{\text{mix}}^{(n)}(\varepsilon) - t_{\text{mix}}^{(n)}}{w_n} \leq \frac{c_\varepsilon w_n + t_{\text{mix}}^{(n)}(1 - \varepsilon) - t_{\text{mix}}^{(n)}}{w_n} \leq c_\varepsilon$$

et donc pour $\alpha \geq c_\varepsilon$

$$d_n(t_{\text{mix}}^{(n)} + \alpha w_n) \leq d_n\left(t_{\text{mix}}^{(n)} + \frac{t_{\text{mix}}^{(n)}(\varepsilon) - t_{\text{mix}}^{(n)}}{w_n} w_n\right) = d_n(t_{\text{mix}}^{(n)}(\varepsilon)) \leq \varepsilon$$

ce qui prouve (8b). □

Ceci montre donc que le deuxième terme dans le développement asymptotique des ε -temps de mélange, qui contrairement au premier terme peut dépendre de ε , est d'un ordre majoré par (w_n) . Voir la Figure 6 pour une illustration.

Un corollaire utile de ces deux derniers lemmes sera une version analogue du Corollaire 2.16 adaptée à cette définition.

Corollaire 2.21. *Soient $(w_n)_{n \in \mathbb{N}}$ une suite de nombres réels strictement positifs et $(t_n)_{n \in \mathbb{N}}$ une suite de nombres réels positifs telles que $\lim_{n \rightarrow +\infty} \frac{t_n}{w_n} = +\infty$. Si une suite de chaînes de Markov*

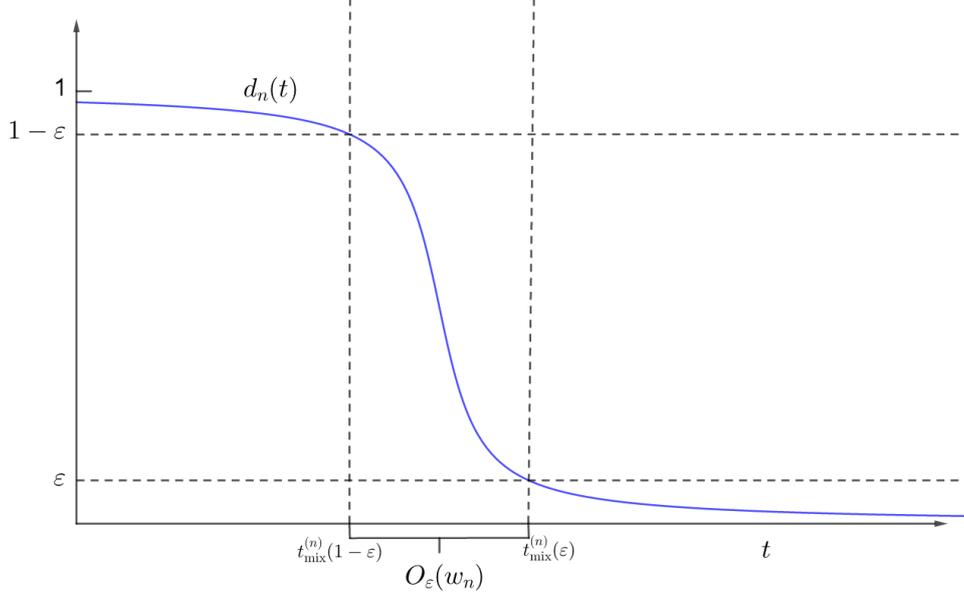


Figure 6: Le Lemme 2.19 montre que l'intervalle sur lequel se produit la chute de $1 - \varepsilon$ à ε est d'une longueur $O_\varepsilon(w_n)$.

satisfait

$$\begin{aligned} \lim_{\alpha \rightarrow -\infty} \liminf_{n \rightarrow +\infty} d_n(t_n + \alpha w_n) &= 1, \\ \lim_{\alpha \rightarrow +\infty} \limsup_{n \rightarrow +\infty} d_n(t_n + \alpha w_n) &= 0, \end{aligned}$$

alors elle présente un *cut-off* de taille de fenêtre $(w_n)_{n \in \mathbb{N}}$, et de plus $\lim_{n \rightarrow \infty} \frac{t_{\text{mix}}^{(n)}}{t_n} = 1$.

Preuve. En imitant exactement la preuve du Lemme 2.19 avec (t_n) au lieu de $(t_{\text{mix}}^{(n)})$ et en appliquant le Corollaire 2.16, on obtient la présence du *cut-off* ainsi que $\lim_{n \rightarrow \infty} \frac{t_{\text{mix}}^{(n)}}{t_n} = 1$. En imitant maintenant la partie (\Rightarrow) de la preuve du Lemme 2.20 de nouveau avec (t_n) au lieu de $(t_{\text{mix}}^{(n)})$, on obtient par ce même Lemme 2.20 que le *cut-off* a une fenêtre de taille (w_n) . \square

La dernière chose importante qu'il faut noter à ce niveau de très grande généralité est que tout *cut-off* se produit dans une fenêtre d'une certaine taille. Dit d'une autre manière, pour toute suite de chaînes de Markov qui présente un *cut-off*, il est possible de trouver une suite (w_n) telle que le *cut-off* se passe dans une fenêtre de taille (w_n) .

Lemme 2.22. *Supposons qu'une suite de chaînes de Markov présente un cut-off. Alors il existe une suite $(w_n)_{n \in \mathbb{N}}$ de nombres réels strictement positifs telle que le cut-off admette une fenêtre de taille (w_n) .*

Preuve. Soit $N \in \mathbb{N}$ tel que $t_{\text{mix}}^{(n)} \geq 1$ pour $n \geq N$ et posons pour $n \geq N$ et $\nu \geq 1$

$$\alpha_{n,\nu} := \frac{t_{\text{mix}}^{(n)}(2^{-\nu}) - t_{\text{mix}}^{(n)}(1 - 2^{-\nu})}{t_{\text{mix}}^{(n)}}.$$

Comme par le Lemme 2.13 on a $\lim_{n \rightarrow \infty} \alpha_{n,\nu} = 0$ pour tout $\nu \geq 1$, on construit récursivement une suite $(N_\nu)_{\nu \geq 1}$ de nombres naturels strictement plus grand que N et strictement croissante telle

que pour tout n, ν, ν'

$$n \geq N_\nu, \nu \geq \nu' \implies \alpha_{n, \nu'} \leq 2^{-\nu}. \quad (14)$$

On pose de plus $N_0 = N$. Alors pour tout $n \geq N$, il existe un unique nombre naturel ν_n tel que $N_{\nu_n} \leq n < N_{\nu_n+1}$; on observe que (ν_n) est monotone croissante et tend vers l'infini. Posons pour $\nu \geq 1$

$$\gamma_\nu := 1 + \max_{\substack{N \leq n < N_\nu \\ 1 \leq \nu' \leq \nu}} 2^{\nu_n} \alpha_{n, \nu'}$$

et pour $n \geq N$

$$\delta_n := 2^{-\nu_n}.$$

Pour $n \geq N$ et $\nu \geq 1$ on a donc, si $\nu_n \geq \nu$

$$n \geq N_{\nu_n}, \nu_n \geq \nu \xrightarrow{(14)} \alpha_{n, \nu} \leq 2^{-\nu_n} \leq \gamma_\nu \delta_n,$$

et si $\nu_n < \nu$ alors $n < N_\nu$, et par conséquent

$$\alpha_{n, \nu} \leq 2^{-\nu_n} \max_{\substack{N \leq n' < N_\nu \\ 1 \leq \nu' \leq \nu}} 2^{\nu_{n'}} \alpha_{n', \nu'} \leq \gamma_\nu \delta_n.$$

Posons enfin $c_\nu = \max\{\gamma_\nu, \max_{n < N} t_{\text{mix}}^{(n)}(2^{-\nu})\}$, $w_n = 1$ si $n < N$ et $w_n = t_{\text{mix}}^{(n)} \delta_n$ si $n \geq N$. En rassemblant tous ces morceaux, on obtient que

$$t_{\text{mix}}^{(n)}(2^{-\nu}) - t_{\text{mix}}^{(n)}(1 - 2^{-\nu}) \leq c_\nu w_n$$

pour tout $n \in \mathbb{N}$ et $\nu \geq 1$, ainsi que $\lim_{n \rightarrow \infty} \frac{t_{\text{mix}}^{(n)}}{w_n} = \lim_{n \rightarrow \infty} 2^{\nu_n} = \infty$. Par conséquent, si $\varepsilon \in]0, 1[$, on prend $\nu \geq 1$ tel que $\varepsilon > 2^{-\nu}$ et donc

$$t_{\text{mix}}^{(n)}(\varepsilon) - t_{\text{mix}}^{(n)}(1 - \varepsilon) \leq t_{\text{mix}}^{(n)}(2^{-\nu}) - t_{\text{mix}}^{(n)}(1 - 2^{-\nu}) \leq c_\nu w_n.$$

On conclut par le Lemme 2.20. □

Remarque 2.23. On peut généraliser la notion de *cut-off* aux différentes distances utilisées pour quantifier l'écart à la distribution stationnaire, le *cut-off* pour la distance de séparation étant la variante la plus importante. Il est à noter que tous les résultats prouvés dans cette section pour la distance de variation total sont vrais pour toute distance décroissante et convergente vers 0; nous avons uniquement utilisé ces propriétés de d . Vu que la distance de séparation est décroissante, on peut appliquer toute cette section, pourvu qu'il y ait convergence.

3 Techniques d'étude du temps de mélange et du *cut-off*

Nous allons à présent explorer différentes techniques qui permettent de prouver l'existence d'un *cut-off* et d'en estimer l'emplacement. Nous illustrons ces techniques à l'aide de l'exemple de la marche aléatoire paresseuse sur l'hypercube, qui a été défini dans le sous-chapitre 1.6, où l'on voyait que celle-ci possède de bonnes propriétés. Nous allons montrer que la marche aléatoire

sur l'hypercube présente en effet un *cut-off*! Pour la première démonstration de ceci (en temps continu), voire [1].

3.1 Couplage

Commençons par définir la notion de couplage de mesures de probabilité.

Définition 3.1. Soient μ, ν deux mesures de probabilité sur un même ensemble fini Ω . Un *couplage* de μ et ν est la donnée d'une paire de variables aléatoires (X, Y) telles que X et Y sont définies sur le même espace de probabilité et à valeurs dans Ω , et telles que la loi de X soit μ et la loi de Y soit ν .

La notion de couplage est utile dans un grand nombre de contextes différents, mais pour notre objectif (estimer la distance de variation totale), l'intuition suivante est particulièrement importante. Soient μ, ν deux mesures de probabilités sur un ensemble fini Ω . Supposons qu'il existe $x \in \Omega$ tel que $\mu(x) \gg \nu(x)$. Si (X, Y) est un couplage de (μ, ν) , alors vu que l'événement $\{X = x\}$ a une forte probabilité comparé à $\{Y = x\}$, on s'attend à ce que la contribution de $\mathbb{P}(X = x, Y \neq x)$ à $\mathbb{P}(X \neq Y)$ soit relativement grande, de sorte que $\mathbb{P}(X \neq Y) \gg 0$. En résumé: si μ et ν diffèrent très fortement quelque part dans l'espace d'états, alors la quantité $\mathbb{P}(X \neq Y)$ va être grande, peu importe le couplage (X, Y) choisi. La quantité $\mathbb{P}(X \neq Y)$ semble donc bien choisie pour quantifier l'écart entre μ et ν .

Lemme 3.2. Soient μ, ν deux distributions de probabilité sur un ensemble fini Ω et soit (X, Y) un couplage de (μ, ν) . Alors

$$\|\mu - \nu\|_{\text{TV}} \leq \mathbb{P}(X \neq Y)$$

Preuve. Pour $A \subset \Omega$, on a

$$\mu(A) - \nu(A) = \mathbb{P}(X \in A) - \mathbb{P}(Y \in A) \leq \mathbb{P}(X \in A, Y \notin A) \leq \mathbb{P}(X \neq Y).$$

En prenant le maximum sur tout les $A \subset \Omega$, on obtient le résultat. \square

Remarque 3.3. On a même

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

avec l'inf atteint. La construction du couplage optimal est technique et peut être vue en [2], (Proposition 4.7).

Spécifions maintenant la notion de couplage dans le cadre des chaînes de Markov, afin d'utiliser le Lemme 3.2 pour estimer la distance de variation totale.

Définition 3.4. Un *couplage de chaînes de Markov* sur un espace d'états Ω et de matrice de transition P est un processus $((X_t, Y_t))_{t \geq 0}$ à valeurs dans $\Omega \times \Omega$ tel que $(X_t)_{t \geq 0}$ et $(Y_t)_{t \geq 0}$ soient des chaînes de Markov d'espace d'états Ω et de matrice de transition P .

Le couplage est dit *coalescent* si la propriété

$$\forall s \geq 0 : X_s = Y_s \implies (X_t = Y_t \forall t \geq s) \quad (\text{coal})$$

est vérifiée.

Un couplage de chaînes de Markov *Markovien* de matrice de transition P est un processus $((X_t, Y_t))_{t \geq 0}$ qui est lui-même une chaîne de Markov dont la matrice de transition Q satisfait

- (i) pour tout $x, y, x' \in \Omega$, $\sum_{y' \in \Omega} Q((x, y), (x', y')) = P(x, x')$
- (ii) pour tout $x, y, y' \in \Omega$, $\sum_{x' \in \Omega} Q((x, y), (x', y')) = P(y, y')$.

On peut vérifier que (X_t) et (Y_t) sont donc toutes deux des chaînes de Markov de matrice de transition P , i.e. qu'en effet, $((X_t, Y_t))_{t \geq 0}$ est un couplage de chaînes de Markov de matrice de transition P .

Remarque 3.5. Si $((X_t, Y_t))_{t \geq 0}$ est un couplage Markovien, alors le processus $((\tilde{X}_t, Y_t))_{t \geq 0}$ défini par

$$\tilde{X}_t = \begin{cases} X_t & \text{si } X_0 \neq Y_0, \dots, X_t \neq Y_t \\ Y_t & \text{sinon} \end{cases}$$

est un couplage Markovien coalescent. Ainsi, on peut généralement supposer qu'un couplage Markovien est coalescent.

Remarque 3.6. Si $((X_t, Y_t))_{t \geq 0}$ est un couplage de chaînes de Markov de matrice de transition P , et si de plus $X_0 = x$ et $Y_0 = y$, alors (X_t, Y_t) est un couplage de $(P^t(x, \cdot), P^t(y, \cdot))$.

Exemple 3.7. Construisons maintenant un couplage de chaînes de Markov $((X_t, Y_t))_{t \geq 0}$ de la marche aléatoire paresseuse sur l'hypercube de dimension n . On procède comme dans l'Exemple 1.27. Étant donné (X_t, Y_t) pour un instant $t \geq 0$, on choisit de manière aléatoire et uniforme (et indépendante de tout ce qui s'est passé avant) une coordonnée $k \in [[1, n]]$. Ensuite, on lance une pièce équilibrée (indépendante de tout ce qui s'est passé avant). Si on obtient pile, on met les k -ièmes coordonnées de X_t et de Y_t à 0. Sinon, on les met toutes deux à 1.

Il est clair que si l'on considère (X_t) resp. (Y_t) séparément, alors ce sont des marches aléatoires paresseuses sur l'hypercube de dimension n . Ce couplage est markovien.

En effet, la procédure décrite correspond à une description en terme de suite récurrente aléatoire. Soient Ω , Λ , f et (Z_t) comme dans l'Exemple 1.27. Définissons $g : (\Omega \times \Omega) \times \Lambda \rightarrow \Omega \times \Omega$ par

$$g((x, y), (b, k)) = (f(x, (b, k)), f(y, (b, k)))$$

et soit (X_0, Y_0) un couple de variables aléatoires défini sur le même espace de probabilité que (Z_t) , à valeur dans Ω , et indépendant de toute la suite (Z_t) . Alors la suite de variables aléatoires $((X_t, Y_t))_{t \geq 0}$ définie récursivement par

$$\forall t \geq 1 : \quad (X_t, Y_t) = g((X_{t-1}, Y_{t-1}), Z_t)$$

est une chaîne de Markov de matrice de transition

$$Q((x, y), (x', y')) = \mathbb{P}\{g((x, y), Z_1) = (x', y')\}.$$

On vérifie immédiatement que

$$\begin{aligned} \sum_{y' \in \Omega} Q((x, y), (x', y')) &= \sum_{y' \in \Omega} \mathbb{P}\{g((x, y), Z_1) = (x', y')\} = \mathbb{P}\{f(x, Z_1) = x'\} = P(x, x'), \\ \sum_{x' \in \Omega} Q((x, y), (x', y')) &= \sum_{x' \in \Omega} \mathbb{P}\{g((x, y), Z_1) = (x', y')\} = \mathbb{P}\{f(y, Z_1) = y'\} = P(y, y'). \end{aligned}$$

De plus, on voit facilement que le couplage est coalescent.

Évoquons à présent le Lemme central du couplage de chaînes de Markov.

Lemme 3.8. *Soit $((X_t, Y_t))_{t \geq 0}$ un couplage coalescent de chaînes de Markov de matrice de transition P tel que $X_0 = x$ et $Y_0 = y$. Notons $\mathbb{P}_{x,y}$ la probabilité de l'espace de probabilité sur lequel le couplage est défini et posons*

$$\tau_{\text{couple}} := \inf\{t \geq 0 \mid X_t = Y_t\} \in \mathbb{N} \cup \{\infty\}.$$

Alors

$$\|P^t(x, \cdot) - P^t(y, \cdot)\|_{\text{TV}} \leq \mathbb{P}_{x,y}(\tau_{\text{couple}} > t).$$

Preuve. Comme déjà mentionné dans la Remarque 3.6, (X_t, Y_t) est un couplage de $(P^t(x, \cdot), P^t(y, \cdot))$. Par le Lemme 3.2 on a donc

$$\|P^t(x, \cdot) - P^t(y, \cdot)\|_{\text{TV}} \leq \mathbb{P}_{x,y}(X \neq Y).$$

Comme le couplage est coalescent, on a $\{X_t \neq Y_t\} \subseteq \{\tau_{\text{couple}} > t\}$, d'où le résultat. \square

Corollaire 3.9. *Supposons que pour tout $x, y \in \Omega$ il existe $((X_t, Y_t))_{t \geq 0}$ un couplage de chaînes de Markov de matrice de transition P avec $X_0 = x$ et $Y_0 = y$, et définissons τ_{couple} de la même manière que dans le Lemme 3.8. Alors*

$$d(t) \leq \max_{x,y} \mathbb{P}_{x,y}(\tau_{\text{couple}} > t).$$

On peut en déduire un premier résultat très intéressant qui concerne l'Exemple 3.7.

Exemple 3.10. Reprenons les notations de l'Exemple 3.7. Écrivons pour $t \geq 1$ que $Z_t = (B_t, K_t)$ (i.e. $B_t \in \{0, 1\}$ et $K_t \in [[1, n]]$) et considérons

$$C_t = |\{K_1, \dots, K_t\}|, \quad C_0 = 0$$

i.e. le nombre de coordonnées qui ont été actualisées par un bit aléatoire au moins une fois jusqu'à l'instant t . Il s'avère que la suite $(C_t)_{t \geq 0}$ est elle-même une chaîne de Markov (on parle du modèle du "collectionneur de coupons") sur l'espace d'états $[[0, n]]$, de matrice de transition

$$P_{\text{coupon}}(i, j) = \begin{cases} \frac{i}{n} & \text{si } j = i \\ \frac{n-i}{n} & \text{si } j = i + 1 \\ 0 & \text{sinon.} \end{cases}$$

Il faut s'imaginer la chaîne de la manière suivante, en tentant de retrouver pour un court instant l'âme de collectionneur féroce que nous avons tous lorsqu'enfants, nous collectionnions les vignettes Panini ou les cartes Pokémon. Supposons qu'il y a n types de vignettes différents, que l'on achète dans des paquets unitaires. On suppose que les vignettes sont réparties de manière uniforme entre les différents paquets. Si à un moment donné, on possède déjà i types de vignettes différents et qu'on achète un paquet, alors la probabilité que l'on obtienne une vignette que l'on a déjà est $\frac{i}{n}$, tandis que la probabilité d'en obtenir une nouvelle est $\frac{n-i}{n}$.

Par construction du couplage dans l'Exemple 3.7, on voit que si pour un $t \geq 0$ on a $C_t = n$, alors

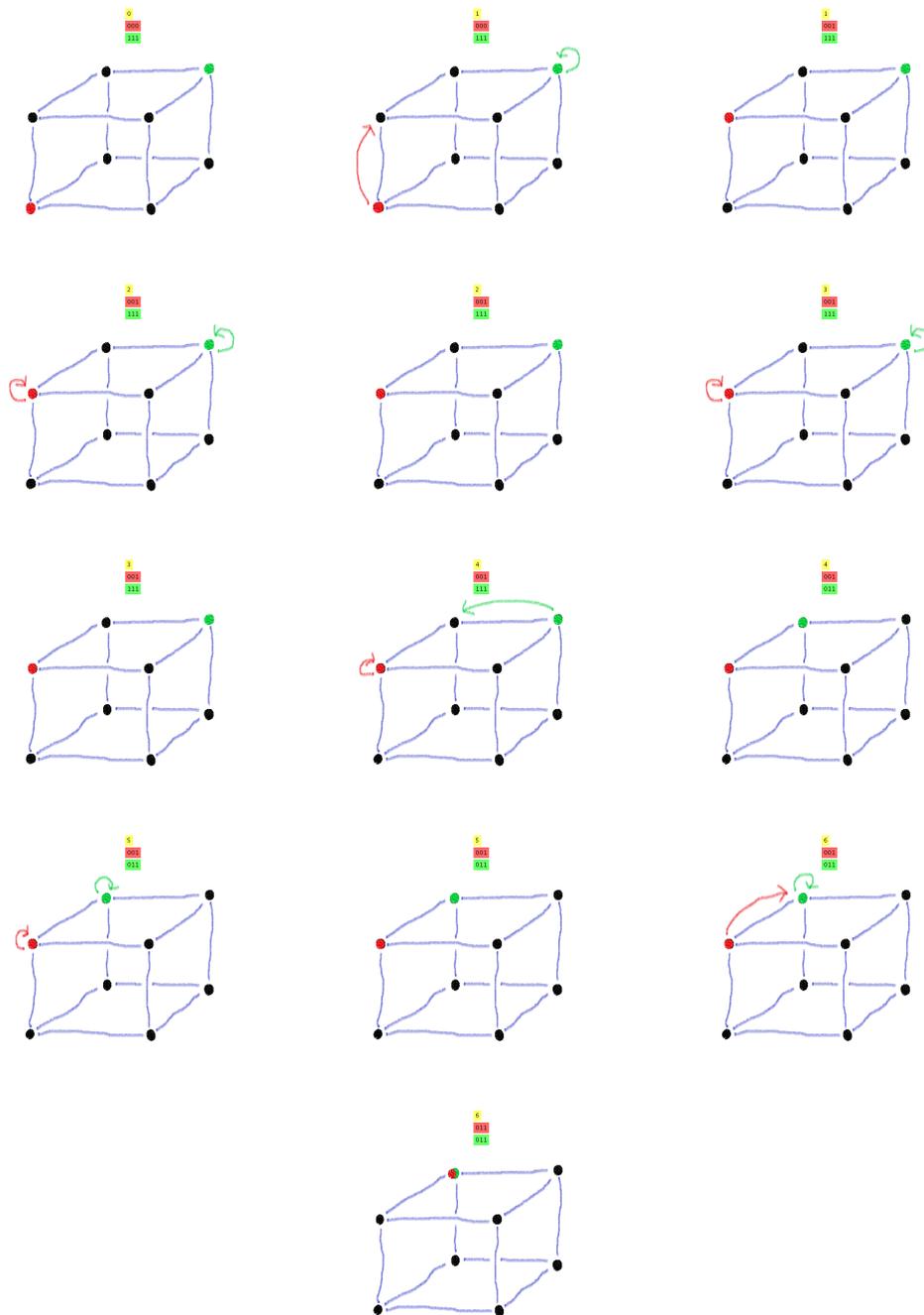


Figure 7: Simulation du couplage de l'hypercube en dimension 3 jusqu'au moment de rencontre.

toutes les coordonnées des points de départs ont été actualisées au moins une fois par un même bit, et donc $X_t = Y_t$. Ainsi, si on pose

$$\tau_{\text{act}}^{(n)} := \inf\{t \geq 1 \mid C_t = n\}$$

alors on a $\tau_{\text{couple}} \leq \tau_{\text{act}}^{(n)}$ et donc en particulier $\mathbb{P}(\tau_{\text{couple}} > t) \leq \mathbb{P}(\tau_{\text{act}}^{(n)} > t)$. Essayons d'estimer cette dernière quantité.

Proposition 3.11. *Pour tout $\alpha > 0$*

$$\mathbb{P}(\tau_{\text{act}}^{(n)} > \lceil n \log(n) + \alpha n \rceil) \leq e^{-\alpha}.$$

Preuve. Fixons $\alpha > 0$ et posons $t = \lceil n \log(n) + \alpha n \rceil$. Pour $i \in \llbracket 1, n \rrbracket$, soit A_i l'événement $\{i \notin \{K_1, \dots, K_t\}\}$, i.e. qu'au temps t , on n'a pas encore la i -ième vignette. On a

$$\mathbb{P}(\tau_{\text{act}}^{(n)} > t) = \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \mathbb{P}(A_i) \quad (15)$$

Comme $\mathbb{P}(A_i) = \frac{(n-1)^t}{n^t} = \left(1 - \frac{1}{n}\right)^t$ on obtient donc

$$\mathbb{P}(\tau_{\text{act}}^{(n)} > t) \leq n \left(1 - \frac{1}{n}\right)^t \leq n \exp\left(-\frac{t}{n}\right) \leq n \exp\left(-\frac{n \log(n) + \alpha n}{n}\right) = e^{-\alpha}.$$

□

En utilisant ce Corollaire 3.9, il vient que

$$d_n(\lceil n \log(n) + \alpha n \rceil) \leq e^{-\alpha}$$

ce que nous donne déjà

$$\lim_{\alpha \rightarrow \infty} \limsup_{n \rightarrow \infty} d_n(n \log(n) + \alpha n) = 0$$

i.e. une identité du type (8b), et aussi

$$t_{\text{mix}}^{(n)} \leq \lceil n \log(n) + \log(4)n \rceil.$$

Malheureusement, ceci n'aidera pas à prouver l'existence du *cut-off*. On verra avec des méthodes plus avancées que $t_{\text{mix}}^{(n)} \sim \frac{1}{2}n \log(n)$. Mais au moins, on a déjà le bon ordre! Il reste à améliorer la constante.

Remarque 3.12. La raison pour laquelle on n'obtient pas la bonne constante vient du couplage, et non pas des estimations $\mathbb{P}(\tau_{\text{couple}} > t) \leq \mathbb{P}(\tau_{\text{act}}^{(n)} > t)$ et (15). En utilisant le principe d'inclusion-exclusion, on peut calculer $\mathbb{P}(\tau_{\text{couple}} > t)$ explicitement et s'en convaincre. Même si l'on raffine le couplage en imposant $Y_0 \sim \pi$ (ce que fait que (X_t, Y_t) est un couplage de $(P^t(x, \cdot), \pi)$), ce qui améliore l'estimation donné par Lemme 3.2, on n'arrive pas à améliorer la constante.

3.2 Temps d'arrêt

3.2.1 Temps stationnaires forts

Changeons maintenant un peu de perspective en considérant la chaîne de Markov suivante. Le cadre des chaînes de Markov est très convenable pour modéliser le mélange de cartes, afin qu'on puisse étudier par exemple le nombre d'itérations d'une certaine technique nécessaires pour que les cartes soient bien mélangées. La technique pour mélanger qu'on va considérer ici s'appelle *top-to-random shuffle*.

Exemple 3.13. Le *top-to-random shuffle* pour n cartes est une chaîne de Markov $(X_t)_{t \geq 0}$ d'espace d'états $\Omega = S_n$ et de matrice de transition

$$P(x, y) = \begin{cases} \frac{1}{n} & \text{si } y = (i \ i - 1 \ \dots \ 1) \circ x \text{ pour un } i \in [[1, n]] \\ 0 & \text{sinon.} \end{cases}$$

On impose de plus que $X_0 = \text{id}$. Cela modélise le mélange de cartes comme suit.

Pour pouvoir distinguer les cartes on les numérote avec les éléments de $[[1, n]]$. Le nombre $X_t(i)$ correspond à la position de la carte d'indice i au temps t , où la position 1 correspond à la carte au sommet de la pile. Au début, i.e. à l'instant $t = 0$, le jeu est en ordre parfait. La méthode de mélange que modélise P correspond à prendre à chaque étape la carte au sommet de la pile (i.e. la carte d'indice $X_t^{-1}(1)$) et de la placer quelque part dans le jeu, d'une manière aléatoire et uniforme (et indépendante de tout ce qui se passait avant). Bien évidemment personne mélange les cartes comme ceci (du moins à notre connaissance), et donc cet exemple est d'un intérêt purement mathématique.

On observe maintenant qu'avec cette technique, la seule possibilité pour une carte d'être au-dessous de la carte qui était tout en bas de la pile au temps $t = 0$ (i.e. celle d'indice $X_0^{-1}(n) = n$) au temps t , est d'avoir été mise en dessous de celle-ci en un temps antérieur; on s'attend donc à ce que les positions relatives des cartes au dessous de la carte initialement au fond soient toutes équiprobables. Par conséquent, si τ_{top} est le premier temps où la carte initialement au fond est au sommet de la pile, les $n - 1$ cartes au-dessous vont être arrangées d'une manière aléatoire et uniforme. Il y a plus étonnant encore: au temps $\tau_{\text{top}} + 1$, on a placé la carte initialement au fond du tas dans l'arrangement uniformément aléatoire des $n - 1$ cartes au dessous d'elle, et ceci d'une manière aléatoire, uniforme, et indépendante de l'arrangement au-dessous; ce que signifie que tout les arrangements des n cartes sont équiprobables. Le jeu est parfaitement mélangé!

D'un point de vue conceptuel, pour notre objectif qui reste d'estimer la distance de variation totale, cet exemple a l'intérêt suivant: si pour une chaîne de Markov donnée $(X_t)_{t \geq 0}$ on trouve une variable aléatoire τ tel que la distribution X_τ est la distribution stationnaire (ou peut-être même que la distribution de X_t sur l'événement $\{\tau = t\}$ est stationnaire, ce qui est plus fort!), et si ensuite on observe que τ est fortement concentrée autour d'un certain valeur, cela nous donnera des informations sur le temps de mélange. Formalisons donc, dans l'esprit de cette méthode d'approche, les notions qu'on vient de discuter.

Définition 3.14. Soit $(X_t)_{t \geq 0}$ une suite de variables aléatoires à valeurs dans Ω (pas forcément une chaîne de Markov). Une variable aléatoire τ à valeurs dans $\mathbb{N} \cup \{\infty\}$ est un *temps d'arrêt* pour (X_t) si et seulement si pour tout $t \in \mathbb{N}$, il existe un sous-ensemble $B_t \subseteq \Omega^{t+1}$ tel que

$$\{\tau = t\} = \{(X_0, \dots, X_t) \in B_t\}$$

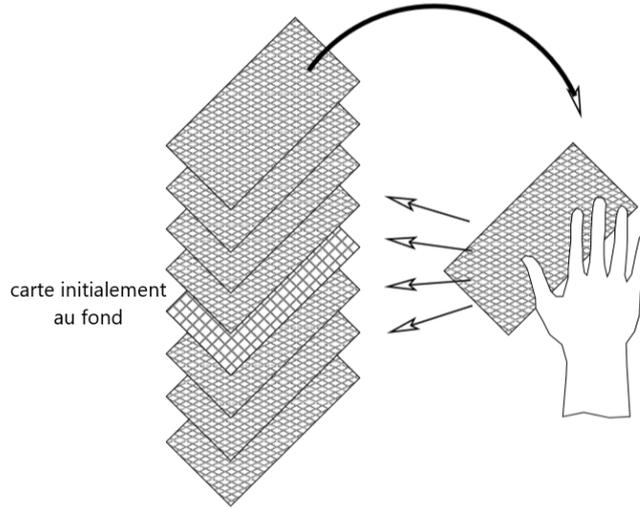


Figure 8: Le *top-to-random-shuffle*. L'image est prise (et légèrement altérée) de [2], page 75.

Si $(X_t)_{t \geq 0}$ est une chaîne de Markov avec $X_0 = x$ pour un $x \in \Omega$ et donnée sous forme d'une représentation comme suite récurrente aléatoire avec $(Z_t)_{t \geq 1}$ la suite de variables aléatoires i.i.d. y associée, et si τ est un temps d'arrêt pour (Z_t) (dans le sens où $\{\tau = t\} = \{(Z_1, \dots, Z_t) \in B_t\}$ pour un $B_t \subseteq \Lambda^t$), alors τ est dit *temps d'arrêt randomisé* pour (X_t) .

Remarque 3.15. Le terme de "temps d'arrêt" vient du fait que si l'on simule la suite (X_t) un pas de temps à la fois, alors à tout instant t , l'événement $\{\tau = t\}$ dépend seulement des observations déjà réalisées, et on peut donc décider si l'événement $\{\tau = t\}$ est réalisé ou non. On pourrait donc utiliser $\{\tau = t\}$ comme critère pour arrêter la simulation.

Remarque 3.16. La variable aléatoire τ_{top} qu'on a considérée dans l'Exemple 3.13 est un temps d'arrêt pour la chaîne de Markov (et par conséquent aussi $\tau_{\text{top}} + 1$) puisque

$$\{\tau_{\text{top}} = t\} = \{X_0(n) \neq 1, X_1(n) \neq 1, \dots, X_{t-1}(n) \neq 1, X_t(n) = 1\}.$$

En général, le temps de première visite d'un sous-ensemble fixé de l'espace d'états est un temps d'arrêt pour la chaîne de Markov considérée.

Remarque 3.17. Contrairement à ce qu'on pourrait penser à première vue, un temps d'arrêt d'une chaîne de Markov avec un point de départ fixe est toujours un temps d'arrêt randomisé, mais un temps d'arrêt randomisé par rapport à une représentation comme suite récurrente aléatoire n'est pas forcément un temps d'arrêt pour la chaîne de Markov. La spécification "temps d'arrêt randomisé" n'indique donc pas que c'est une sous-classe de temps d'arrêts pour la chaîne de Markov en question, mais qu'on la considère dans le cadre d'une représentation comme suite récurrente aléatoire donnée (ceci pour des raisons techniques).

Conceptualisons maintenant la propriété plus remarquable de τ_{top} dans l'Exemple 3.13.

Définition 3.18. Soit $(X_t)_{t \geq 0}$ une chaîne de Markov à position initiale fixé, d'espace d'états Ω et de distribution stationnaire π . Un *temps stationnaire fort* pour (X_t) est un temps d'arrêt randomisé τ , qui peut éventuellement dépendre du point de départ de la chaîne, tel que pour

tout $x, y \in \Omega$ et $t \geq 0$

$$\mathbb{P}_x(X_\tau = y, \tau = t) = \pi(y)\mathbb{P}_x(\tau = t)$$

i.e. tel que la distribution de X_t sur l'événement $\{\tau = t\}$ soit π .

Remarque 3.19. Il existe aussi une notion de *temps stationnaire*. On demande seulement que la distribution de X_τ soit π , ce qui est plus faible, puisque pour un temps stationnaire fort, on demande qu'en plus X_τ soit indépendant de τ . Pour des raisons techniques, plus précisément pour le lemme suivant, on aura besoin de cette propriété plus forte.

Le lemme technique suivant montre le fait intuitif qu'après être arrivée au temps stationnaire fort, la distribution reste stationnaire. La preuve peut être vue en [2], (Lemma 6.9), ou bien en s'inspirant de la preuve du Lemme 3.27.

Lemme 3.20. Soit $(X_t)_{t \geq 0}$ une chaîne de Markov à position initiale fixé $X_0 = x$, d'espace d'états Ω et de distribution stationnaire π . Si τ est un temps stationnaire fort pour (X_t) , alors pour tout $y \in \Omega$ et $t \geq 0$

$$\mathbb{P}_x(X_t = y, t \geq \tau) = \pi(y)\mathbb{P}_x(t \geq \tau)$$

i.e. la distribution de X_t sur l'événement $\{t \geq \tau\}$ est π .

On peut maintenant prouver le lemme central de cette partie.

Lemme 3.21. Si τ est un temps stationnaire fort pour une chaîne de Markov avec distance de séparation s , alors

$$s(t) \leq \max_{x \in \Omega} \mathbb{P}_x(\tau > t).$$

Preuve. Pour $x, y \in \Omega$ on a

$$1 - \frac{P^t(x, y)}{\pi(y)} = 1 - \frac{\mathbb{P}_x(X_t = y)}{\pi(y)} \leq 1 - \frac{\mathbb{P}_x(X_t = y, t \geq \tau)}{\pi(y)} \stackrel{\text{Lemme 3.20}}{=} 1 - \mathbb{P}_x(t \geq \tau) = \mathbb{P}_x(\tau > t)$$

et en prenant le maximum sur tout les $x, y \in \Omega$, on obtient le résultat. \square

Puisque $d \leq s$, on obtient donc en particulier que pour un temps stationnaire fort, $d(t) \leq \max_{x \in \Omega} \mathbb{P}_x(\tau > t)$. Voyons ce que cela nous donne pour l'hypercube.

Exemple 3.22. Premièrement, essayons de trouver un temps stationnaire fort pour la marche aléatoire paresseuse sur l'hypercube. Comme un tel temps est toujours défini par rapport à une représentation comme fonction récurrente aléatoire, reprenons celle des Exemples 1.27 et 3.7. Intuitivement, si à un instant t tout les coordonnées ont été remplacées au moins une fois par un bit aléatoire, i.e. si $t \geq \tau_{\text{act}}^{(n)}$, alors le point sur l'hypercube devrait être uniformément aléatoire. De plus, on voit directement que $\tau_{\text{act}}^{(n)}$ est un temps d'arrêt randommisé. Notre candidat pour un temps stationnaire fort est alors $\tau_{\text{act}}^{(n)}$; prouvons donc que c'en est effectivement un.

Lemme 3.23. Soit $(X_t)_{t \geq 0}$ une marche aléatoire paresseuse sur l'hypercube de dimension n . Alors si pour $t \geq 0$ on note $A_t = \{K_1, \dots, K_t\}$ et si $A \subset [[1, n]]$, on a pour tout $y \in \{0, 1\}^n$

$$\mathbb{P}(X_t|_A = y|_A, A_t = A) = \frac{1}{2^{|A|}} \mathbb{P}(A_t = A)$$

où pour $x \in \{0,1\}^n$, on note $x|_A$ le vecteur de $\{0,1\}^{|A|}$ contenant les coordonnées x_i de x pour lesquels $i \in A$.

Preuve. Prouvons le résultat par récurrence sur $|A|$. Si $|A| = 0$ le résultat est trivialement vrai; supposons le donc vrai pour tout B avec $|B| < |A|$. Faisons de plus une récurrence sur t ; pour $t = 0$ le résultat est aussi trivialement vrai, supposons le donc vrai pour tout temps inférieur à t . On a

$$\begin{aligned} \mathbb{P}(X_t|_A = y|_A, A_t = A) &= \sum_{a \in A} \mathbb{P}(X_t|_A = y|_A, K_t = a, A_t = A) = \\ &= \sum_{a \in A} [\mathbb{P}(X_t|_A = y|_A, K_t = a, A_{t-1} = A) + \mathbb{P}(X_t|_A = y|_A, K_t = a, A_{t-1} = A \setminus \{a\})] \end{aligned}$$

Maintenant, en posant $y(a) = y + e_a$

$$\begin{aligned} \mathbb{P}(X_t|_A = y|_A, K_t = a, A_{t-1} = A) &= \\ &= \mathbb{P}(X_{t-1}|_A = y|_A, K_t = a, B_t = y_a, A_{t-1} = A) \\ &\quad + \mathbb{P}(X_{t-1}|_A = y(a)|_A, K_t = a, B_t = y_a, A_{t-1} = A) = \\ &= \frac{1}{2n} \mathbb{P}(X_{t-1}|_A = y|_A, A_{t-1} = A) + \frac{1}{2n} \mathbb{P}(X_{t-1}|_A = y(a)|_A, A_{t-1} = A) = \\ &= \frac{1}{2n} \frac{1}{2^{|A|}} \mathbb{P}(A_{t-1} = A) + \frac{1}{2n} \frac{1}{2^{|A|}} \mathbb{P}(A_{t-1} = A) = \\ &= \frac{1}{n} \frac{1}{2^{|A|}} \mathbb{P}(A_{t-1} = A) = \frac{1}{2^{|A|}} \mathbb{P}(A_{t-1} = A, K_t = a) \end{aligned}$$

ainsi que

$$\begin{aligned} \mathbb{P}(X_t|_A = y|_A, K_t = a, A_t = A \setminus \{a\}) &= \\ &= \mathbb{P}(X_{t-1}|_{A \setminus \{a\}} = y|_{A \setminus \{a\}}, K_t = a, B_t = y_a, A_t = A \setminus \{a\}) = \\ &= \frac{1}{2n} \mathbb{P}(X_{t-1}|_{A \setminus \{a\}} = y|_{A \setminus \{a\}}, A_{t-1} = A \setminus \{a\}) = \frac{1}{2n} \frac{1}{2^{|A|-1}} \mathbb{P}(A_{t-1} = A \setminus \{a\}) = \\ &= \frac{1}{2^{|A|}} \mathbb{P}(A_{t-1} = A \setminus \{a\}, K_t = a). \end{aligned}$$

On a donc en résumé

$$\begin{aligned} \mathbb{P}(X_t|_A = y|_A, A_t = A) &= \sum_{a \in A} \frac{1}{2^{|A|}} \mathbb{P}(A_{t-1} = A, K_t = a) + \frac{1}{2^{|A|}} \mathbb{P}(A_{t-1} = A \setminus \{a\}, K_t = a) = \\ &= \frac{1}{2^{|A|}} \sum_{a \in A} \mathbb{P}(A_{t-1} \in \{A, A \setminus \{a\}\}, K_t = a) = \frac{1}{2^{|A|}} \mathbb{P}(A_t = A). \end{aligned}$$

□

Lemme 3.24. Avec les mêmes notations que dans le Lemme 3.23 on a pour tout $y \in \{0,1\}^n$

$$\mathbb{P}(X_t|_A = y|_A, A_t = A \supsetneq A_{t-1}) = \frac{1}{2^{|A|}} \mathbb{P}(A_t = A \supsetneq A_{t-1}).$$

Preuve.

$$\begin{aligned}
\mathbb{P}(X_t|_A = y|_A, A_t = A \supseteq A_{t-1}) &= \sum_{a \in A} \mathbb{P}(X_t|_A = y|_A, K_t = a, A_t = A \supseteq A_{t-1}) = \\
&= \sum_{a \in A} \mathbb{P}(X_t|_A = y|_A, K_t = a, A_{t-1} = A \setminus \{a\}) = \sum_{a \in A} \frac{1}{2} \frac{1}{2^{|A|-1}} \mathbb{P}(K_t = a, A_{t-1} = A \setminus \{a\}) = \\
&= \frac{1}{2^{|A|}} \mathbb{P}(A_t = A \supseteq A_{t-1}).
\end{aligned}$$

□

En appliquant ce lemme à $A = [[1, n]]$ on obtient donc que

Corollaire 3.25. $\tau_{\text{act}}^{(n)}$ est un temps stationnaire fort de la marche aléatoire paresseuse sur l'hypercube.

Par le Lemme 3.21 on obtient donc que $d_n(t) \leq s_n(t) \leq \mathbb{P}(\tau_{\text{act}}^{(n)} > t)$. Pour d_n cela donne donc malheureusement la même estimation qu'on a déjà obtenue par le couplage et qui n'aidera pas à prouver l'existence du *cut-off*. Ce qu'on vient de démontrer est quand même un tout petit peu mieux, puisque on a pu borner s . En particulier, on a obtenu

$$\lim_{\alpha \rightarrow \infty} \limsup_{n \rightarrow \infty} s_n(n \log(n) + \alpha n) = 0$$

et $t_{\text{sép}}^{(n)} \leq \lceil n \log(n) + \log(4)n \rceil$. On peut montrer qu'on a effectivement $t_{\text{sép}}^{(n)} \sim n \log(n)$ et même un *cut-off* pour la distance de séparation; pour la borne inférieure, voire [2], partie 18.4.

3.2.2 Raffinement: temps forts

On peut néanmoins raffiner les arguments ci-dessus pour obtenir la bonne estimation $t_{\text{mix}}^{(n)} \leq \frac{1}{2}n \log n + cn$, l'idée clé étant la suivante: si l'on trouve un temps d'arrêt randomnisé τ tel que la distribution de X_t sur l'événement $\{\tau = t\}$ n'est pas égal à π , mais au moins suffisamment proche en distance de variation totale, alors on pourra quand même espérer en tirer une bonne estimation de d . En effet, si cette relaxation fait que τ prend des valeurs suffisamment plus petites qu'un temps stationnaire fort donné, cela pourrait compenser la perte de ne plus imposer la stationnarité, et on obtiendra peut-être même une meilleure estimation.

Essayons donc d'appliquer cette idée raffinée à la marche aléatoire paresseuse sur l'hypercube. On a construit $\tau_{\text{act}}^{(n)}$ afin que la distribution de X_t sur $\{\tau_{\text{act}}^{(n)} = t\}$ soit $\pi^{(n)}$, i.e. soit uniforme, ou, dit d'une autre manière, que la distribution du nombre de 1 dans X_t soit binomiale. On relâche maintenant cette condition forte en introduisant

$$\tau_{\text{act}}^{(n, \varepsilon)} := \inf\{t \geq 1 \mid C_t > n - \varepsilon\sqrt{n}\}.$$

Sur $\{\tau_{\text{act}}^{(n, \varepsilon)} = t\}$, $n - \varepsilon\sqrt{n}$ bits ont été touchés, et donc si l'on est parti de 0, la distribution du nombre de 1 dans X_t va généralement être plus faible que selon $\pi^{(n)}$, et ceci avec une différence d'ordre $\varepsilon\sqrt{n}$. Mais par le théorème central limite, les fluctuations d'une loi binomiale autour de son espérance sont de toute façon de l'ordre \sqrt{n} , et donc on s'attend à ce que, en variation totale, la distribution de X_t sur $\{\tau_{\text{act}}^{(n, \varepsilon)} = t\}$ soit suffisamment proche de π pour pouvoir améliorer l'estimation, parce qu'en même temps, $\tau_{\text{act}}^{(n, \varepsilon)}$ est concentré autour de valeurs plus faibles que τ_{act} . Attendre que les dernières coordonnées soient touchées peut prendre très longtemps, et il va s'avérer que le gain lié au fait de ne plus attendre les dernières $\varepsilon\sqrt{n}$ coordonnées va être substantiel.

Définition 3.26. Soit $(X_t)_{t \geq 0}$ une chaîne de Markov à position initiale fixé, d'espace d'états Ω . Un *temps fort* pour (X_t) est un temps d'arrêt randomnisé τ , qui peut éventuellement dépendre

du point de départ de la chaîne, tel que pour tout $x \in \Omega$, il existe une mesure de probabilité $\tilde{\pi}_x$ sur Ω satisfaisant

$$\mathbb{P}_x(X_\tau = y, \tau = t) = \tilde{\pi}_x(y)\mathbb{P}_x(\tau = t) \quad (16)$$

pour tout $y \in \Omega$ et $t \geq 0$.

On obtient une version du Lemme 3.20 dans ce cadre plus général.

Lemme 3.27. *Soit $(X_t)_{t \geq 0}$ une chaîne de Markov à position initiale fixé, d'espace d'états Ω et de distribution stationnaire π . Si τ est un temps fort pour (X_t) , alors pour tout $x, y \in \Omega$ et $t \geq 0$ avec $\mathbb{P}_x(\tau \leq t) > 0$*

$$\|\mathbb{P}_x(X_t = \cdot \mid \tau \leq t) - \pi\|_{\text{TV}} \leq \|\tilde{\pi}_x - \pi\|_{\text{TV}},$$

i.e. la distribution de X_t sur $\{\tau \leq t\}$ est plus proche de π que sur $\{\tau = t\}$.

Preuve. Soit $(Z_t)_{t \geq 1}$ la suite de variables aléatoires de la représentation par une suite récurrente aléatoire dont τ est un temps d'arrêt. On voit par récurrence que pour tout $r \geq 0$, il existe une fonction $f_r : \Omega \times \Lambda^r \rightarrow \Omega$ tel que

$$\forall t \geq 0 : \quad X_{t+r} = f_r(X_t, Z_{t+1}, \dots, Z_{t+r}).$$

En effet, il suffit de poser $f_0 = \text{id}_\Omega$ et $f_{r+1} = f \circ (f_r \times \text{id}_\Lambda)$. Soient maintenant $x, y \in \Omega$, $t \geq 0$ tel que $\mathbb{P}_x(\tau \leq t) > 0$ et $0 \leq s \leq t$. On a

$$\mathbb{P}_x(X_t = y, \tau = s) = \sum_{\substack{z \in \Omega \\ \mathbb{P}_x(\tau = s, X_s = z) > 0}} \mathbb{P}_x(X_t = y \mid X_s = z, \tau = s) \mathbb{P}_x(X_s = z, \tau = s).$$

Soit $B_s \subseteq \Lambda^s$ tel que $\{\tau = s\} = \{(Z_1, \dots, Z_s) \in B_s\}$. Pour un z apparaissant dans la somme, il vient que

$$\begin{aligned} \mathbb{P}_x(X_t = y \mid X_s = z, \tau = s) &= \\ &= \mathbb{P}_x(f_{t-s}(z, Z_{s+1}, \dots, Z_t) = y \mid f_s(x, Z_1, \dots, Z_s) = z, (Z_1, \dots, Z_s) \in B_s) = \\ &= \mathbb{P}_x(f_{t-s}(z, Z_{s+1}, \dots, Z_t) = y) = P^{t-s}(z, y), \end{aligned}$$

où l'on a utilisé l'indépendance mutuelle de (Z_t) . Par conséquent,

$$\begin{aligned} \mathbb{P}_x(X_t = y, \tau = s) &= \sum_{z \in \Omega} P^{t-s}(z, y) \mathbb{P}_x(X_s = z, \tau = s) = \sum_{z \in \Omega} P^{t-s}(z, y) \tilde{\pi}_x(z) \mathbb{P}_x(\tau = s) = \\ &= (\tilde{\pi}_x P^{t-s})(y) \cdot \mathbb{P}_x(\tau = s) \end{aligned}$$

et donc

$$\mathbb{P}_x(X_t = y \mid \tau \leq t) = \sum_{s \leq t} (\tilde{\pi}_x P^{t-s})(y) \cdot \mathbb{P}_x(\tau = s \mid \tau \leq t).$$

Par la preuve de la Proposition 2.8, on a $\|\tilde{\pi}_x P^{t-s} - \pi\|_{\text{TV}} = \|\tilde{\pi}_x P^{t-s} - \pi P^{t-s}\|_{\text{TV}} \leq \|\tilde{\pi}_x - \pi\|_{\text{TV}}$

et par l'inégalité triangulaire il vient que

$$\begin{aligned} \|\mathbb{P}_x(X_t = \cdot \mid \tau \leq t) - \pi\|_{\text{TV}} &\leq \sum_{s \leq t} \|\tilde{\pi}_x P^{t-s} - \pi\|_{\text{TV}} \mathbb{P}_x(\tau = s \mid \tau \leq t) \leq \\ &\leq \|\tilde{\pi}_x - \pi\|_{\text{TV}} \sum_{s \leq t} \mathbb{P}_x(\tau = s \mid \tau \leq t) = \|\tilde{\pi}_x - \pi\|_{\text{TV}}. \end{aligned}$$

□

Ceci donne ensuite une version généralisé du Lemme 3.21.

Lemme 3.28. *Si τ est un temps fort pour une chaîne de Markov (X_t) avec distance de variation totale d , alors*

$$d(t) \leq \max_{x \in \Omega} [\mathbb{P}_x(\tau > t) + \|\tilde{\pi}_x - \pi\|_{\text{TV}}].$$

Preuve. Soient $t \geq 0$ et $x \in \Omega$. Alors si $\mathbb{P}_x(\tau \leq t) = 0$, on a directement $\|P^t(x, \cdot) - \pi\|_{\text{TV}} \leq 1 \leq \mathbb{P}_x(\tau > t) + \|\tilde{\pi}_x - \pi\|_{\text{TV}}$. Sinon, on obtient en posant $\Omega(x, t) = \{y \in \Omega \mid \pi(y) > P^t(x, y)\}$, que

$$\begin{aligned} \|P^t(x, \cdot) - \pi\|_{\text{TV}} &= \pi(\Omega(x, t)) - \mathbb{P}_x(X_t \in \Omega(x, t)) \leq \pi(\Omega(x, t)) - \mathbb{P}_x(X_t \in \Omega(x, t) \mid \tau \leq t) \mathbb{P}_x(\tau \leq t) = \\ &= \pi(\Omega(x, t)) \mathbb{P}_x(\tau > t) + (\pi(\Omega(x, t)) - \mathbb{P}_x(X_t \in \Omega(x, t) \mid \tau \leq t)) \mathbb{P}_x(\tau \leq t) \leq \mathbb{P}_x(\tau > t) + \|\tilde{\pi}_x - \pi\|_{\text{TV}} \end{aligned}$$

par le Lemme 3.27. On conclut en prenant le maximum sur x des deux côtés. □

Appliquons ceci à $\tau_{\text{act}}^{(n, \varepsilon)}$. Pour commencer, prouvons que $\tau_{\text{act}}^{(n, \varepsilon)}$ est un temps fort. Il est claire que par symétrie, il suffit de vérifier (16) pour $x = 0$. Déterminons de plus la distribution $\tilde{\pi}_0^{(n, \varepsilon)}$ associée.

Lemme 3.29. *Avec les notations introduites ci-dessus, $\tau_{\text{act}}^{(n, \varepsilon)}$ est un temps fort, et on a pour tout $y \in \{0, 1\}^n$*

$$\tilde{\pi}_0^{(n, \varepsilon)}(y) = \binom{n - |y|}{c_n(\varepsilon) - |y|} 2^{-c_n(\varepsilon)} \binom{n}{c_n(\varepsilon)}^{-1},$$

où $|y| := \sum_{i=1}^n y_i$ dénote le nombre de 1 dans y , et $c_n(\varepsilon) := n - \lceil \varepsilon \sqrt{n} \rceil + 1$.

Preuve. Soit $y \in \{0, 1\}^n$ et notons $U_y = \{i \in [[1, n]] \mid y_i = 1\}$. Alors

$$\mathbb{P}_0(X_t = y, \tau_{\text{act}}^{(n, \varepsilon)} = t) = \sum_{\substack{A \subseteq [[1, n]] \\ |A| = c_n(\varepsilon)}} \mathbb{P}_0(X_t = y, A_t = A \supseteq A_{t-1}).$$

Vu que les coordonnées de X_t hors de A n'ont pas été touchées et donc valent 0, il faut $A \supseteq U_y$ pour que $\mathbb{P}_0(X_t = y, A_t = A \supseteq A_{t-1})$ soit non nul. Par conséquent, on a $\mathbb{P}_0(X_t = y, A_t = A \supseteq A_{t-1}) = \mathbb{1}_{\{A \supseteq U_y\}} \mathbb{P}_0(X_t^{(A)} = y^{(A)}, A_t = A \supseteq A_{t-1})$. Ainsi, par le Lemme 3.24, il vient que

$$\mathbb{P}_0(X_t = y, \tau_{\text{act}}^{(n, \varepsilon)} = t) = \sum_{\substack{U_y \subseteq A \subseteq [[1, n]] \\ |A| = c_n(\varepsilon)}} \frac{1}{2^{|A|}} \mathbb{P}_0(A_t = A \supseteq A_{t-1}).$$

Par symétrie, il est clair que $\mathbb{P}_0(A_t = A \supseteq A_{t-1})$ ne dépend pas de $A \subseteq [[1, n]]$ avec $|A| = c_n(\varepsilon)$.

Notons donc $p_t = \mathbb{P}_0(A_t = A \supseteq A_{t-1})$. On obtient alors

$$\mathbb{P}_0(X_t = y, \tau_{\text{act}}^{(n,\varepsilon)} = t) = \sum_{\substack{U_y \subseteq A \subseteq [[1, n]] \\ |A| = c_n(\varepsilon)}} \frac{1}{2^{c_n(\varepsilon)}} p_t = \binom{n - |y|}{c_n(\varepsilon) - |y|} 2^{-c_n(\varepsilon)} p_t.$$

D'autre part, on a

$$\mathbb{P}_0(\tau_{\text{act}}^{(n,\varepsilon)} = t) = \sum_{\substack{A \subseteq [[1, n]] \\ |A| = c_n(\varepsilon)}} \mathbb{P}_0(A_t = A \supseteq A_{t-1}) = \binom{n}{c_n(\varepsilon)} p_t,$$

d'où le résultat. □

Il reste donc à estimer $\mathbb{P}_0(\tau_{\text{act}}^{(n,\varepsilon)} > t)$ et $\|\tilde{\pi}_0^{(n,\varepsilon)} - \pi^{(n)}\|_{\text{TV}}$.

Proposition 3.30. *Pour tout $\alpha > 0$ on a*

$$\mathbb{P}_0\left(\tau_{\text{act}}^{(n,\varepsilon)} > \lceil \frac{1}{2} n \log n + \alpha n \rceil\right) \leq \exp(\lceil \varepsilon \sqrt{n} \rceil (1 - \log(\varepsilon) - \alpha)).$$

Preuve. Pour $A \subseteq [[1, n]]$, notons $E_A = \{A_t \cap A = \emptyset\}$. Il vient que

$$\mathbb{P}_0(\tau_{\text{act}}^{(n,\varepsilon)} > t) \leq \sum_{\substack{A \subseteq [[1, n]] \\ |A| = \lceil \varepsilon \sqrt{n} \rceil}} \mathbb{P}_0(E_A) = \binom{n}{\lceil \varepsilon \sqrt{n} \rceil} \left(1 - \frac{\lceil \varepsilon \sqrt{n} \rceil}{n}\right)^t.$$

En utilisant l'estimation connue $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$, le fait que $\frac{n}{\lceil \varepsilon \sqrt{n} \rceil} \leq \frac{\sqrt{n}}{\varepsilon}$ ainsi que $1 - x \leq e^{-x}$, on obtient

$$\mathbb{P}_0(\tau_{\text{act}}^{(n,\varepsilon)} > t) \leq \left(\frac{e\sqrt{n}}{\varepsilon}\right)^{\lceil \varepsilon \sqrt{n} \rceil} \exp\left(-t \frac{\lceil \varepsilon \sqrt{n} \rceil}{n}\right).$$

En posant maintenant $t = \lceil \frac{1}{2} n \log n + \alpha n \rceil$, on voit que $\exp\left(-t \frac{\lceil \varepsilon \sqrt{n} \rceil}{n}\right) \leq (\sqrt{n})^{-\lceil \varepsilon \sqrt{n} \rceil} \exp(-\alpha \lceil \varepsilon \sqrt{n} \rceil)$ et donc

$$\mathbb{P}_0\left(\tau_{\text{act}}^{(n,\varepsilon)} > \lceil \frac{1}{2} n \log n + \alpha n \rceil\right) \leq \left(\frac{e}{\varepsilon}\right)^{\lceil \varepsilon \sqrt{n} \rceil} \exp(-\alpha \lceil \varepsilon \sqrt{n} \rceil) = \exp(\lceil \varepsilon \sqrt{n} \rceil (1 - \log(\varepsilon) - \alpha)).$$

□

Proposition 3.31.

$$\lim_{\varepsilon \rightarrow 0^+} \limsup_{n \rightarrow \infty} \|\tilde{\pi}_0^{(n,\varepsilon)} - \pi^{(n)}\|_{\text{TV}} = 0$$

Preuve. Pour faciliter les notations, posons $m = c_n(\varepsilon)$. D'après Proposition 2.2 il vient que

$$\|\tilde{\pi}_0^{(n,\varepsilon)} - \pi^{(n)}\|_{\text{TV}} = \frac{1}{2} \sum_{y \in \{0,1\}^n} |\tilde{\pi}_0^{(n,\varepsilon)}(y) - \pi^{(n)}(y)| = \frac{1}{2} \sum_{k=0}^n \binom{n}{k} \left| \binom{n-k}{m-k} 2^{-m} \binom{n}{m}^{-1} - 2^{-n} \right|.$$

Comme $\binom{n}{k} \binom{n-k}{m-k} \binom{n}{m}^{-1} = \binom{m}{k}$ on a donc

$$\|\tilde{\pi}_0^{(n,\varepsilon)} - \pi^{(n)}\|_{\text{TV}} = \frac{1}{2} \sum_{k=0}^n \left| \binom{m}{k} 2^{-m} - \binom{n}{k} 2^{-n} \right| = \|\beta_m - \beta_n\|_{\text{TV}}$$

où l'on a noté β_m resp. β_n la loi binomiale de paramètres $1/2$ et m resp. n . En observant que $\binom{l+1}{k}2^{-l-1} = \frac{l+1}{2(l+1-k)}\binom{l}{k}2^{-l}$, on voit que la suite $\left(\binom{l}{k}2^{-l}\right)_{l \in \mathbb{N}}$ est croissante sur $[[0, 2k]]$, atteint son maximum $\binom{2k}{k}2^{-2k} < \frac{1}{\sqrt{\pi k}}$ sur $\{2k-1, 2k\}$, et est décroissante sur $[[2k-1, \infty[$. Par conséquent,

$$\begin{aligned} & \sum_{k=0}^n \left| \binom{m}{k}2^{-m} - \binom{n}{k}2^{-n} \right| = \\ & = \sum_{0 \leq 2k \leq m} \left(\binom{m}{k}2^{-m} - \binom{n}{k}2^{-n} \right) + \sum_{m < 2k < n} \left| \binom{m}{k}2^{-m} - \binom{n}{k}2^{-n} \right| + \sum_{n \leq 2k \leq 2n} \left(\binom{n}{k}2^{-n} - \binom{m}{k}2^{-m} \right). \end{aligned}$$

En utilisant la formule de symétrie des coefficients binomiaux, ceci donne

$$\sum_{k=0}^n \left| \binom{m}{k}2^{-m} - \binom{n}{k}2^{-n} \right| = \sum_{2m-n < 2k \leq m} \binom{m}{k}2^{-m} + \sum_{m < 2k \leq n} \binom{n}{k}2^{-n} + \sum_{m < 2k < n} \left| \binom{m}{k}2^{-m} - \binom{n}{k}2^{-n} \right|.$$

En notant F_m resp. F_n la fonction de répartition de β_m resp. β_n et en utilisant le fait que $n-m < \varepsilon\sqrt{n}$ et donc aussi $n-m < 2\varepsilon\sqrt{m}$ pour n assez grand, ainsi que $\left| \binom{m}{k}2^{-m} - \binom{n}{k}2^{-n} \right| \leq \frac{1}{\sqrt{\pi m}} + \frac{1}{\sqrt{\pi n}} \leq \frac{2}{\sqrt{\pi m}}$, ceci donne

$$\sum_{k=0}^n \left| \binom{m}{k}2^{-m} - \binom{n}{k}2^{-n} \right| \leq \left[F_m\left(\frac{m}{2}\right) - F_m\left(\frac{m}{2} - 2\varepsilon\frac{\sqrt{m}}{2}\right) \right] + \left[F_n\left(\frac{n}{2}\right) - F_n\left(\frac{n}{2} - \varepsilon\frac{\sqrt{n}}{2}\right) \right] + \frac{2\varepsilon\sqrt{n}}{\sqrt{\pi m}}.$$

Finalement, on obtient par le théorème central limite que

$$\limsup_{n \rightarrow \infty} \sum_{k=0}^n \left| \binom{c_n(\varepsilon)}{k}2^{-c_n(\varepsilon)} - \binom{n}{k}2^{-n} \right| \leq \Phi(0) - \Phi(-2\varepsilon) + \Phi(0) - \Phi(-\varepsilon) + \frac{2\varepsilon}{\sqrt{\pi}},$$

où Φ dénote la fonction de répartition de la loi normale de moyenne 0 et variance 1, d'où le résultat. \square

En rassemblant tout ceci, on obtient donc (en remarquant que par symétrie $\mathbb{P}_x(\tau_{\text{act}}^{(n,\varepsilon)} > t) + \|\tilde{\pi}_x^{(n,\varepsilon)} - \pi^{(n)}\|_{\text{TV}}$ ne dépend pas de $x \in \{0, 1\}^n$)

$$d_n\left(\left\lceil \frac{1}{2}n \log n + \alpha n \right\rceil\right) \leq \exp(\lceil \varepsilon\sqrt{n} \rceil (1 - \log(\varepsilon) - \alpha)) + \|\tilde{\pi}_0^{(n,\varepsilon)} - \pi^{(n)}\|_{\text{TV}}$$

et donc pour $\alpha > 1 - \log(\varepsilon)$

$$\limsup_{n \rightarrow \infty} d_n\left(\frac{1}{2}n \log n + \alpha n\right) \leq \limsup_{n \rightarrow \infty} \|\tilde{\pi}_0^{(n,\varepsilon)} - \pi^{(n)}\|_{\text{TV}},$$

d'où

$$\lim_{\alpha \rightarrow \infty} \limsup_{n \rightarrow \infty} d_n\left(\frac{1}{2}n \log n + \alpha n\right) \leq \limsup_{n \rightarrow \infty} \|\tilde{\pi}_0^{(n,\varepsilon)} - \pi^{(n)}\|_{\text{TV}},$$

et donc en laissant tendre ε vers 0, ceci donne finalement

$$\lim_{\alpha \rightarrow \infty} \limsup_{n \rightarrow \infty} d_n\left(\frac{1}{2}n \log n + \alpha n\right) = 0.$$

On a obtenu la première moitié du *cut-off* (dont la taille de fenêtre semble d'être $w_n = n$) pour la marche aléatoire paresseuse sur l'hypercube!

3.3 Valeurs propres

Considérons maintenant les idées d'algèbre linéaire, que l'on a déjà évoqué plusieurs fois avant.

Lemme 3.32. *Soit P la matrice de transition d'une chaîne de Markov. Alors*

1. *Si λ est une valeur propre de P , alors on a $|\lambda| \leq 1$.*
2. *Si P est irréductible, alors $\dim \ker(P - \text{Id}) = 1$.*

Preuve. Soit $v \in \mathbb{C}^\Omega$ un vecteur propre non nul de P , de valeur propre $\lambda \in \mathbb{C}$. Soit $x \in \Omega$ tel que $|v(x)| = \max_{y \in \Omega} |v(y)| \neq 0$. Alors

$$|\lambda| |v(x)| = \left| \sum_{y \in \Omega} P(x, y) v(y) \right| \leq \sum_{y \in \Omega} P(x, y) |v(y)| \leq \sum_{y \in \Omega} P(x, y) |v(x)| = |v(x)|,$$

et par conséquent $|\lambda| \leq 1$.

Enfin, si P est en plus irréductible, nous avons déjà vu dans la preuve de la Proposition 1.21 que $\dim \ker(P - \text{Id}) = 1$. \square

Maintenant, on va s'intéresser au cas où P est réversible. Rappelons que P est réversible (par rapport à sa distribution stationnaire π) si et seulement si les équations d'équilibre détaillées sont vérifiées, i.e.

$$\pi(x)P(x, y) = \pi(y)P(y, x)$$

pour tout $x, y \in \Omega$ (voire le sous-chapitre 1.5). Ce n'est pas forcément le cas que P est donc symétrique; on peut prendre par exemple

$$P = \begin{pmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{pmatrix}$$

et $\pi = (1/3 \ 2/3)$. Néanmoins, après un changement de variable, on peut s'y ramener.

Avant de continuer, introduisons la notation suivante: pour π une mesure de probabilité quelconque sur Ω , notons $\langle \cdot, \cdot \rangle_\pi$ le produit scalaire sur \mathbb{R}^Ω défini par $\langle f, g \rangle_\pi := \sum_{x \in \Omega} f(x)g(x)\pi(x)$.

Lemme 3.33. *Soit P irréductible et réversible (par rapport à sa distribution stationnaire π). Alors il existe $\lambda_1, \dots, \lambda_{|\Omega|} \in \mathbb{R}$ et $f_1, \dots, f_{|\Omega|} \in \mathbb{R}^\Omega$ tel que $Pf_i = \lambda_i f_i$ pour tout i , la famille $\{f_i\}_{i=1}^{|\Omega|}$ forme une base orthonormal par rapport à $\langle \cdot, \cdot \rangle_\pi$ et*

$$P^t(x, y) = \sum_{i=1}^{|\Omega|} \lambda_i^t f_i(x) f_i(y) \pi(y) \tag{17}$$

pour tout $x, y \in \Omega$ et $t \geq 0$.

Preuve. Définissons $A \in \mathbb{R}^{\Omega \times \Omega}$ par $A(x, y) := \pi(x)^{1/2} \pi(y)^{-1/2} P(x, y)$ pour tout $x, y \in \Omega$. Comme P est réversible, A est symétrique. Par le théorème spectral, il existe $\lambda_1, \dots, \lambda_{|\Omega|} \in \mathbb{R}$ et $\varphi_1, \dots, \varphi_{|\Omega|} \in \mathbb{R}^\Omega$ tel que $A\varphi_i = \lambda_i \varphi_i$ pour tout $1 \leq i \leq |\Omega|$. Soit D_π la matrice diagonal définie par $D_\pi(x, x) := \pi(x)$ pour tout $x \in \Omega$. Remarquons que $A = D_\pi^{1/2} P D_\pi^{-1/2}$ (notons que par le Théorème 1.22, π est strictement positive, donc $D_\pi^{-1/2}$ est bien défini). Posons enfin $f_j := D_\pi^{-1/2} \varphi_j$. Alors,

$$Pf_j = (D_\pi^{-1/2} A D_\pi^{1/2})(D_\pi^{-1/2} \varphi_j) = D_\pi^{-1/2} A \varphi_j = \lambda_j D_\pi^{-1/2} \varphi_j = \lambda_j f_j$$

et de plus

$$\langle f_i, f_j \rangle_\pi = \sum_{x \in \Omega} f_i(x) f_j(x) \pi(x) = \sum_{x \in \Omega} \varphi_i(x) \varphi_j(x) = \delta_{ij}.$$

Par conséquent, la famille $\{f_i\}$ forme une base orthonormal par rapport $\langle \cdot, \cdot \rangle_\pi$. Finalement, on obtient

$$\begin{aligned} P^t(x, y) &= [P^t \delta_y](x) = \left[P^t \left(\sum_{i=1}^{|\Omega|} \langle \delta_y, f_i \rangle_\pi f_i \right) \right] (x) = \\ &= \left[P^t \left(\sum_{i=1}^{|\Omega|} f_i(y) \pi(y) f_i \right) \right] (x) = \left[\sum_{i=1}^{|\Omega|} f_i(y) \pi(y) P^t f_i \right] (x) = \sum_{i=1}^{|\Omega|} \lambda_i^t f_i(x) f_i(y) \pi(y). \end{aligned}$$

□

Remarque 3.34. En posant sans perte de généralité $\lambda_1 = 1$ (et donc f_1 est constante égale à 1), on peut écrire (17) comme

$$\frac{P^t(x, y)}{\pi(y)} - 1 = \sum_{i=2}^{|\Omega|} \lambda_i^t f_i(x) f_i(y). \quad (18)$$

Exemple 3.35. Considérons une marche aléatoire symétrique simple sur $\mathbb{Z}/n\mathbb{Z}$, i.e. la chaîne de Markov de matrice de transition

$$P(k, l) = \begin{cases} \frac{1}{2} & \text{si } k - l \equiv \pm 1 \pmod{n} \\ 0 & \text{sinon,} \end{cases}$$

(voire l'Exemple 1.15 pour le cas $n = 4$). Soit $f_j : \mathbb{Z}/n\mathbb{Z} \rightarrow \mathbb{C}$, $x \mapsto \exp(2\pi i j x)/n$ (où i est l'unité imaginaire). Prouvons que f_j est un vecteur propre de P et calculons-en la valeur propre. On a pour tout $x \in \mathbb{Z}/n\mathbb{Z}$ que

$$P f_j(x) = \frac{f_j(x-1) + f_j(x+1)}{2} = \left(\frac{f_j(-1) + f_j(1)}{2} \right) f_j(x) = \cos(2\pi j/n) f_j(x).$$

Par conséquent, f_j est un vecteur propre de valeur propre $\cos(2\pi j/n)$. Comme on a pu le voir dans l'Exemple 1.7 et la Figure 1, cette chaîne est apériodique si et seulement si n est impair. On remarque de plus, que par le calcul ci-dessus, -1 est une valeur propre si et seulement si n est pair.

Lemme 3.36. Soit P est irréductible et apériodique. Alors -1 n'est pas une valeur propre de P .

Preuve. Supposons par l'absurde que -1 est une valeur propre. Par le Lemme 3.32, P^2 n'est pas irréductible car $\dim \ker(P^2 - \text{Id}) \geq 2$. Montrons que néanmoins, P^2 est irréductible. Soit $x, y \in \Omega$. Par l'irréductibilité de P , il existe $n \in \mathbb{N}^*$ tel que $P^n(x, y) > 0$. Si n est pair, $(P^2)^{n/2}(x, y) = P^n(x, y) > 0$. Sinon, alors par l'apériodicité il existe un $m \in \mathbb{N}$ impair tel que $P^m(x, x) > 0$ et donc $P^{n+m}(x, y) \geq P^m(x, x) P^n(x, y) > 0$. De plus, $n+m$ est impair, et donc $(P^2)^{(n+m)/2}(x, y) = P^{n+m}(x, y) > 0$. On conclut donc que P^2 est irréductible, ce qui est absurde. □

Les valeurs propre de la matrice de transition ont une fort influence sur de nombreuses propriétés d'une chaîne de Markov.

Lemme 3.37. Soit P irréductible et apériodique avec valeurs propres $1 = \lambda_1 \geq \dots \geq \lambda_{|\Omega|} \geq -1$ et $f_1, \dots, f_{|\Omega|}$ les vecteurs propres orthonormaux (par rapport à $\langle \cdot, \cdot \rangle_\pi$) associés. Alors

$$4\|P^t(x, \cdot) - \pi(\cdot)\|_{\text{TV}}^2 \leq \sum_{j=2}^{|\Omega|} \lambda_j^{2t} (f_j(x))^2. \quad (19)$$

Preuve. Fixons $x \in \Omega$. Par Cauchy-Schwarz et (18), il vient que

$$4\|P^t(x, \cdot) - \pi(\cdot)\|_{\text{TV}}^2 \stackrel{(18)}{=} \left\| \frac{P^t(x, \cdot)}{\pi(\cdot)} - 1 \right\|_{L^1(\pi)}^2 \stackrel{\text{C-S}}{\leq} \left\| \frac{P^t(x, \cdot)}{\pi(\cdot)} - 1 \right\|_{L^2(\pi)}^2 = \sum_{j=2}^{|\Omega|} \lambda_j^{2t} (f_j(x))^2.$$

□

Remarque 3.38. Si une chaîne de Markov possède une grande symétrie, de sorte qu'elle a la même "apparence" partout dans l'espace d'états (on dit que la chaîne est *transitive*; voir partie 2.6.2 de [2] pour un traitement rigoureux de cette notion), ce qui est le cas pour la marche aléatoire paresseuse sur l'hypercube, alors on peut montrer que la distribution stationnaire π est forcément uniforme, i.e. $\pi(x) = \frac{1}{|\Omega|}$ pour tout $x \in \Omega$, et que $\|P^t(x, \cdot) - \pi\|_{\text{TV}}$ ne dépend pas de $x \in \Omega$. Par conséquent,

$$4d(t)^2 = \frac{1}{|\Omega|} \sum_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{\text{TV}}^2 \stackrel{(19)}{\leq} \sum_{j=2}^{|\Omega|} \lambda_j^{2t} \sum_{x \in \Omega} \frac{(f_j(x))^2}{|\Omega|} = \sum_{j=2}^{|\Omega|} \lambda_j^{2t} \sum_{x \in \Omega} (f_j(x))^2 \pi(x) = \sum_{j=2}^{|\Omega|} \lambda_j^{2t} \quad (20)$$

puisque $\{f_j\}$ est orthonormale.

Appliquons maintenant ce que l'on vient de démontrer à notre exemple de la marche aléatoire paresseuse sur l'hypercube.

Théorème 3.39. Soit d_n la distance variation totale de la marche aléatoire paresseuse sur l'hypercube de dimension n . Alors il existe une constante absolue C tel que pour tout $\alpha > 0$

$$d_n \left(\frac{1}{2} n \log n + \alpha n \right) = C e^{-\alpha}. \quad (21)$$

Preuve. Notons que cette marche est irréductible, réversible, apériodique, et transitive. Calculons ses valeurs propres. Soient $\chi_0, \chi_1 : \{0, 1\} \rightarrow \mathbb{R}$ définies par $\chi_0 := 1$ et $\chi_1(1) := -1, \chi_1(0) = 1$, i.e. $\chi_j(x) = (-1)^{jx}$ pour tout $j, x \in \{0, 1\}$. Pour $\xi \in \{0, 1\}^n$, définissons $\chi_\xi : \{0, 1\}^n \rightarrow \mathbb{R}$ par $\chi_\xi(x_1, \dots, x_n) := \chi_{\xi_1}(x_1) \dots \chi_{\xi_n}(x_n)$ pour tout $(x_1, \dots, x_n) \in \{0, 1\}^n$, et définissons de plus $k = k(\xi)$ par $k(\xi) := |\{i \in [1, n] : \xi_i = 1\}|$. En utilisant que $\chi_\xi(x + e_i) = (-1)^{\xi_i} \chi_\xi(x)$, il vient ensuite que

$$P\chi_\xi = \left(\frac{1}{2} + \frac{n-k}{2n} - \frac{k}{2n} \right) \chi_\xi = \left(1 - \frac{k}{n} \right) \chi_\xi,$$

i.e. χ_ξ est un vecteur propre, de valeur propre $(1 - \frac{k}{n})$. Comme les $\{\chi_\xi\}$ sont linéairement indépendants, la valeur propre $1 - \frac{k}{n}$ a une multiplicité (géométrique) de $\binom{n}{k}$. On a donc par (20) que

$$4d_n^2(t) = \sum_{k=1}^n \binom{n}{k} \left(1 - \frac{k}{n} \right)^{2t} \leq \sum_{k=1}^n \binom{n}{k} e^{-2tk/n} = (1 + e^{-2t/n})^n - 1 \leq e^{ne^{-2t/n}} - 1.$$

En prenant $t = \frac{1}{2}n \log n + \alpha n$ avec $\alpha > 1$, on obtient

$$4d_n^2 \left(\frac{1}{2}n \log n + \alpha n \right) \leq e^{e^{-2\alpha}} - 1 \leq 2e^{-2\alpha}$$

d'où le résultat. \square

On a donc obtenu à nouveau, avec un peu moins de travail, la bonne borne supérieure de d_n , qui prouve la moitié 'supérieure' du *cut-off* de la marche aléatoire paresseuse sur l'hypercube. La preuve du Lemme 3.37 montre que l'on a en fait prouvé la moitié 'supérieure' du *cut-off* au sens L^2 , ce qui, en toute généralité, est plus fort qu'au sens L^1 , montré déjà à la fin de la partie 3.2.2.

3.4 Statistiques distinctives

Puisque toutes les techniques que l'on a vu jusqu'ici ont donné des bornes supérieures de la distance de variation totale et du temps de mélange, examinons maintenant une technique pour estimer ces quantités inférieurement, et utilisons la pour finalement montrer qu'il y a *cut-off* pour la marche aléatoire paresseuse sur l'hypercube!

Pour une distribution de probabilité μ sur Ω fini, on notera E_μ l'espérance associée, i.e. $E_\mu(f) := \sum_{x \in \Omega} f(x)\mu(x)$ pour toute fonction à valeurs réelles f . Cette notation n'est pas à confondre avec \mathbb{E}_μ , l'espérance de l'espace de probabilité sous-jacent d'une chaîne de Markov de distribution initiale μ . On note aussi Var_μ la variance sur (Ω, μ) .

Lemme 3.40. *Soit μ, ν deux distributions de probabilité sur Ω et soit $f : \Omega \rightarrow \mathbb{R}$ une fonction. Supposons qu'il existe $r, \sigma \geq 0$ tels que $E_\mu(f) - E_\nu(f) \geq r\sigma$ et $\sigma^2 \geq \max\{\text{Var}_\mu(f), \text{Var}_\nu(f)\}$, alors on a*

$$\|\mu - \nu\|_{\text{TV}} \geq 1 - \frac{8}{r^2}.$$

Preuve. Posons $m := (E_\mu(f) + E_\nu(f))/2$ et $A := f^{-1}((m, \infty))$. Alors, $|E_\nu(f) - m| \geq r\sigma/2$. Par l'inégalité de Tchebychev, $\nu(A) \leq \frac{4}{r^2\sigma^2} \text{Var}_\nu(f) \leq \frac{4}{r^2}$. Par un raisonnement analogue, on obtient $\mu(A^c) \leq \frac{4}{r^2}$. Par conséquent,

$$\|\mu - \nu\|_{\text{TV}} \geq \mu(A) - \nu(A) \geq 1 - \frac{8}{r^2}.$$

\square

Appliquons ce lemme à la marche aléatoire paresseuse sur l'hypercube, afin d'obtenir une bonne borne inférieure de sa distance de variation totale. Pour cela, il nous faut trouver une bonne fonction f qui va nous permettre de 'distinguer' les deux distributions de probabilité dont on veut estimer la distance de variation totale. L'idée est de prendre f comme la distance (i.e. le nombre minimal de pas unitaires) entre la position initiale fixe et celle actuelle. Intuitivement, pour des temps petits, la position de la marche n'est pas loin du point de départ (donc f va être petit) et la variation est relativement petite aussi. Utilisons cette idée pour prouver le *cut-off* pour l'hypercube.

Théorème 3.41. *Pour la marche aléatoire paresseuse sur l'hypercube de dimension n , on a*

$$d_n \left(\frac{1}{2}n \log(n) - \alpha n \right) \geq 1 - 8e^{2\alpha + o_\alpha(1)}$$

où pour tout $\alpha > 0$, $o_\alpha(1) \rightarrow 0$ quand n tend vers l'infini.

Preuve. Soit (X_t) la marche de point de départ $0 := (0, \dots, 0)$ et soit $f : \{0, 1\}^n \rightarrow \mathbb{R}$ définie par $f(x) := \text{dist}(0, x) = |x| = x_1 + x_2 + \dots + x_n$, i.e. le nombre de 1 dans x , pour tout $x \in \{0, 1\}^n$. Notons que $E_\pi(f) = \frac{n}{2}$ et $\text{Var}_\pi(f) = \frac{n}{4}$, puisque la distribution du nombre de 1 selon $\pi = \pi^{(n)}$ est binomiale de paramètres n et $1/2$. Définissons $g : \{0, 1\}^n \rightarrow \mathbb{R}$ par $g := \frac{n}{2} - f$. On a donc $E_\pi(g) = 0$, et $\text{Var}_\pi(g) = \frac{n}{4}$. Soit $Y_t := g(X_t)$. Pour appliquer le Lemme 3.40, essayons d'estimer l'espérance et la variance de $f(X_t)$.

D'abord, notons que $(f(X_t))$ est une marche aléatoire sur $[[0, n]]$ de point de départ 0, et que sa matrice de transition est donnée par

$$\mathbb{P}(f(X_{t+1}) = j \mid f(X_t) = i) = \begin{cases} \frac{1}{2} & \text{si } j = i \\ \frac{n-i}{2n} & \text{si } j = i + 1 \\ \frac{i}{2n} & \text{si } j = i - 1 \\ 0 & \text{sinon.} \end{cases}$$

Il vient que

$$\begin{aligned} \mathbb{E}_0(f(X_{t+1}) \mid f(X_t)) &= \frac{1}{2}f(X_t) + \left(\frac{n-f(X_t)}{2n}\right)(f(X_t) + 1) + \left(\frac{f(X_t)}{2n}\right)(f(X_t) - 1) \\ &= \left(1 - \frac{1}{n}\right)f(X_t) + \frac{1}{2}. \end{aligned}$$

et par conséquent aussi $\mathbb{E}_0(Y_{t+1} \mid Y_t) = \left(1 - \frac{1}{n}\right)Y_t$. Ensuite, on voit que $E_\pi(f) - E_{P^t(0, \cdot)}(f) = \frac{n}{2} - \mathbb{E}_0(f(X_t)) = \mathbb{E}_0(Y_t) = \frac{n}{2} \left(1 - \frac{1}{n}\right)^t$ car $Y_0 = \frac{n}{2}$.

Rappelons que $\text{Var}_0(Y) = \mathbb{E}_0(\text{Var}_0(Y \mid X)) + \text{Var}_0(\mathbb{E}_0(Y \mid X))$, où Var_0 est la variance de l'espace de probabilité sur laquelle la chaîne est définie, et donc

$$\begin{aligned} \text{Var}_0(Y_{t+1}) &= \mathbb{E}_0(\text{Var}_0(Y_{t+1} \mid Y_t)) + \text{Var}_0(\mathbb{E}_0(Y_{t+1} \mid Y_t)) \\ &= \mathbb{E}_0 \left\{ \mathbb{E}_0 \left(\left(Y_{t+1} - \left(1 - \frac{1}{n}\right) Y_t \right)^2 \mid Y_t \right) \right\} + \left(1 - \frac{1}{n}\right)^2 \text{Var}_0(Y_t) \\ &= \mathbb{E}_0 \left\{ \mathbb{E}_0(Y_{t+1}^2 \mid Y_t) - \left(1 - \frac{1}{n}\right)^2 Y_t^2 \right\} + \left(1 - \frac{1}{n}\right)^2 \text{Var}_0(Y_t). \end{aligned}$$

De la même manière que précédemment, on peut vérifier que

$$\mathbb{E}_0(Y_{t+1}^2 \mid Y_t) = \left(1 - \frac{2}{n}\right) Y_t^2 + \frac{1}{2}$$

et par conséquent

$$\begin{aligned} \text{Var}_0(Y_{t+1}) &= \frac{1}{2} - \frac{1}{n^2} \mathbb{E}_0(Y_t^2) + \left(1 - \frac{1}{n}\right)^2 \text{Var}_0(Y_t) \\ &= \frac{1}{2} - \frac{1}{4} \left(1 - \frac{1}{n}\right)^{2t} + \left(1 - \frac{2}{n}\right) \text{Var}_0(Y_t). \end{aligned}$$

L'identité ci-dessus permet de conclure que $\text{Var}_{P^t(0, \cdot)}(f) = \text{Var}_0(f(X_t)) = \text{Var}_0(Y_t) \leq \frac{n}{4}$ pour tout temps t , en effectuant une récurrence sur t .

En posant $r := \sqrt{n} \left(1 - \frac{1}{n}\right)^t$, on obtient par le Lemme 3.40 que

$$\begin{aligned} d_n(t) = \|P^t(0, \cdot) - \pi\|_{\text{TV}} &\geq 1 - \frac{8}{r^2} = 1 - 8 \exp\left(-2t \log\left(1 - \frac{1}{n}\right) - \log(n)\right) \\ &\geq 1 - 8 \exp\left(-2t \left(\frac{1}{n} - \frac{1}{n^2}\right) - \log(n)\right) \end{aligned}$$

En prenant enfin $t := \frac{1}{2}n \log(n) - \alpha n$, il vient que

$$d_n\left(\frac{1}{2}n \log(n) - \alpha n\right) \geq 1 - 8 \exp\left(2\alpha + \frac{\log(n) - 2\alpha}{n}\right)$$

d'où le résultat. □

Un corollaire immédiat du résultat ci-dessus est

$$\lim_{\alpha \rightarrow \infty} \liminf_{n \rightarrow \infty} d_n\left(\frac{1}{2}n \log(n) - \alpha n\right) = 1$$

et donc on a finalement obtenu qu'il y a *cut-off* pour la marche aléatoire paresseuse sur l'hypercube. La fenêtre est de taille n , et le temps de mélange vérifie $t_{\text{mix}}^{(n)} \sim \frac{1}{2}n \log(n)$.

4 Le modèle d'Ising et la dynamique de Glauber

Dans tout ce qui précède, notre point de départ a toujours été la donnée d'une chaîne de Markov; nous en avons étudié la convergence vers sa distribution stationnaire. Dans cette section, nous allons rapidement évoquer le problème inverse: étant donné un ensemble fini Ω et une distribution de probabilité π sur Ω , est-ce que l'on peut construire une chaîne de Markov $(X_t)_{t \geq 0}$ d'espace d'états Ω et dont la distribution stationnaire est π ? Dans l'introduction, nous avons souligné le principal intérêt de cette question: pour simuler une loi de probabilité sur un ensemble de très grande taille, recourir aux chaînes de Markov peut s'avérer judicieux. Le *cut-off* donne, quant à lui, une estimation du temps de simulation nécessaire pour être proche de la loi simulée (au sens de la distance de variation totale). Considérons un exemple célèbre et très profondément étudié pour illustrer ce problème: le modèle d'Ising.

Définition 4.1. Soit $G = (V, E)$ un graphe fini. On appelle *configuration de spins* toute fonction $\sigma : V \rightarrow \{-1, 1\}$; pour $v \in V$, la valeur $\sigma(v)$ est dite le *spin en v* . Notons $\Omega = \{-1, 1\}^V$ l'ensemble des configurations de spins. On appelle *système de spins* toute distribution de probabilité sur l'espace Ω . Le *modèle d'Ising* est le système de spins particulier suivant.

Pour une configuration de spins $\sigma \in \Omega$ on définit son énergie $H(\sigma)$ par

$$H(\sigma) := - \sum_{\{u, v\} \in E} \sigma(u)\sigma(v).$$

La *mesure de Gibbs* de paramètre $\beta \in \mathbb{R}_{\geq 0}$ est la distribution de probabilité

$$\mu_\beta(\sigma) := \frac{1}{Z(\beta)} e^{-\beta H(\sigma)},$$

ou $Z(\beta) = \sum_{\sigma \in \Omega} e^{-\beta H(\sigma)}$ et la constante de normalisation, dite *fonction de partition*.

Le modèle d'Ising modélise très efficacement certains phénomènes physiques complexes, tout en restant suffisamment simple pour être étudié mathématiquement. L'interprétation la plus courante,

inspirant la terminologie ci-dessus, est qu'une configuration de spins $\sigma \in \Omega$ représente un arrangement de petits aimants, et le fait que deux sommets de V soient reliés par une arête modélise leur proximité spatiale.

Pour mieux comprendre la mesure μ_β , on aimerait générer des configurations de spins selon μ_β . Mais si $|V|$ est grand, $|\Omega| = 2^{|V|}$ est beaucoup trop grand pour calculer les probabilités de toutes les configurations, même pour des machines très puissantes. L'idée est donc la suivante.

Si l'on peut construire une chaîne de Markov irréductible et apériodique (X_t) d'espace d'états Ω dont les transitions sont suffisamment simples (d'un point de vue abstrait et numérique) et dont la distribution stationnaire est μ_β , alors pour t suffisamment grand, la distribution de X_t va être très proche de μ_β . Pour générer une configuration selon μ_β , on peut donc faire un nombre suffisamment grand d'itérations de la chaîne, et choisir X_t . Le nombre de transitions nécessaire pour une approximation suffisamment bonne de μ_β peut être étudié avec la théorie développée dans les sections précédentes. En particulier, si l'on peut prouver qu'on a *cut-off*, on aura un temps d'arrêt rigoureux pour un tel algorithme. Considérons maintenant la dynamique de Glauber, qui se formule dans un cadre plus général que le modèle d'Ising (voire la partie 3.2 de [2] sur les chaînes de Metropolis), et consiste en ce que l'on vient de discuter: elle construit une chaîne de Markov d'une distribution stationnaire donnée.

Définition 4.2. Soit un graphe $G = (V, E)$, un ensemble fini S et $\Omega \subset S^V$ un ensemble de fonctions de V vers S . Soit de plus π une distribution de probabilités sur Ω qui ne s'annule pas. Pour $\sigma \in \Omega$ et $w \in V$, posons $\Omega(\sigma, w) := \{\sigma' \in \Omega \mid \sigma'(v) = \sigma(v) \forall v \neq w\}$ l'ensemble des éléments de Ω qui sont égaux à σ sauf éventuellement en un point w . Soit finalement

$$\pi^{\sigma, w}(\sigma') = \pi(\sigma' \mid \Omega(\sigma, w)) = \begin{cases} \frac{\pi(\sigma')}{\pi(\Omega(\sigma, w))} & \text{si } \sigma' \in \Omega(\sigma, w) \\ 0 & \text{sinon,} \end{cases}$$

i.e. la probabilité π conditionnée par l'événement $\Omega(\sigma, w)$. La dynamique de Glauber est une chaîne de Markov (X_t) dont les transitions sont faites de la manière suivante.

Étant donné X_t à un instant t , on choisit d'une manière aléatoire, uniforme et indépendante du passé un sommet $w \in \Omega$. Ensuite on génère X_{t+1} selon $\pi^{X_t, w}$, indépendant du passé.

Ceci correspond donc à la matrice de transition

$$P(\sigma, \sigma') = \frac{1}{|V|} \sum_{w \in \Omega} \pi^{\sigma, w}(\sigma').$$

On peut vérifier que (X_t) est toujours réversible par rapport à π , et donc en particulier π est la distribution stationnaire de (X_t) . Voyons maintenant ce que cela nous donne pour le modèle d'Ising.

Exemple 4.3. Reprenons les notations de la Définition 4.1. Soit $\sigma \in \Omega$ une configuration de spins et soit $w \in V$ un sommet. Alors $\Omega(\sigma, w) = \{\sigma_w^+, \sigma_w^-\}$, où l'on a noté σ_w^+ (resp. σ_w^-) l'élément de $\Omega(\sigma, w)$ tel que $\sigma_w^+(w) = 1$ (resp. $\sigma_w^-(w) = -1$). Calculons les valeurs de $\mu_\beta^{\sigma, w}$.

$$\mu_\beta^{\sigma, w}(\sigma_w^+) = \frac{\frac{1}{Z(\beta)} e^{-\beta H(\sigma_w^+)}}{\frac{1}{Z(\beta)} e^{-\beta H(\sigma_w^+)} + \frac{1}{Z(\beta)} e^{-\beta H(\sigma_w^-)}} = \frac{e^{-\beta H(\sigma_w^+)}}{e^{-\beta H(\sigma_w^+)} + e^{-\beta H(\sigma_w^-)}}$$

La première chose à remarquer est que la constante de normalisation $Z(\beta)$, qui est une somme sur $2^{|V|}$ termes et donc très coûteuse à calculer, s'est simplifiée à cause du conditionnement! Pour simplifier encore plus l'expression ci-dessus, calculons $H(\sigma_w^\pm)$; on s'attend à ce que sa valeur soit

proche de $H(\sigma)$. En effet, on a

$$\begin{aligned} H(\sigma_w^\pm) - H(\sigma) &= \sum_{\{u,v\} \in E} [\sigma(u)\sigma(v) - \sigma_w^\pm(u)\sigma_w^\pm(v)] = \sum_{\substack{\{u,v\} \in E \\ w \in \{u,v\}}} [\sigma(u)\sigma(v) - \sigma_w^\pm(u)\sigma_w^\pm(v)] = \\ &= \sum_{\substack{v \in V \\ \{v,w\} \in E}} [\sigma(v)\sigma(w) - \sigma_w^\pm(v)\sigma_w^\pm(w)] = \sum_{\substack{v \in V \\ \{v,w\} \in E}} \sigma(v) [\sigma(w) - (\pm 1)] = S(\sigma, w)(\sigma(w) \mp 1) \end{aligned}$$

où l'on a noté $S(\sigma, w) := \sum_{\substack{v \in V \\ \{v,w\} \in E}} \sigma(v)$ la somme des spins des voisins de w . Ceci nous donne donc

$$\begin{aligned} \mu_\beta^{\sigma,w}(\sigma_w^+) &= \frac{e^{-\beta H(\sigma_w^+)}}{e^{-\beta H(\sigma_w^+)} + e^{-\beta H(\sigma_w^-)}} = \frac{e^{-\beta(H(\sigma_w^+) - H(\sigma))}}{e^{-\beta(H(\sigma_w^+) - H(\sigma))} + e^{-\beta(H(\sigma_w^-) - H(\sigma))}} = \\ &= \frac{e^{-\beta S(\sigma,w)(\sigma(w)-1)}}{e^{-\beta S(\sigma,w)(\sigma(w)-1)} + e^{-\beta S(\sigma,w)(\sigma(w)+1)}} = \frac{e^{\beta S(\sigma,w)}}{e^{\beta S(\sigma,w)} + e^{-\beta S(\sigma,w)}} = \frac{1 + \tanh(\beta S(\sigma,w))}{2} \end{aligned} \quad (22)$$

et donc, puisque le support de $\mu_\beta^{\sigma,w}$ est réduit à $\{\sigma_w^+, \sigma_w^-\}$, on a aussi

$$\mu_\beta^{\sigma,w}(\sigma_w^-) = \frac{1 - \tanh(\beta S(\sigma,w))}{2}. \quad (23)$$

Pour itérer les transitions de la dynamique de Glauber, il suffit donc de calculer une somme simple d'au plus Δ termes, où Δ est le degré maximal de G . Les transitions étant presque gratuites d'un point de vue numérique, on peut donc à présent se poser la question du nombre de transitions nécessaire pour que l'on soit proche de μ_β , i.e. pour que la dynamique de Glauber ait mélangé. Pour simplifier les choses, prenons Maintenant $G = (\mathbb{Z}/n\mathbb{Z})^2$ le tore en 2 dimensions. Pour ce graphe, le modèle d'Ising présente un comportement étonnant de transitions de phases (voire l'introduction dans [3]): si l'on note $\beta_c = \frac{1}{2} \log(1 + \sqrt{2})$, dit *température inverse critique dans \mathbb{Z}^2* , alors

- si $\beta < \beta_c$, le temps de relaxation est $O(1)$.
- si $\beta = \beta_c$, le temps de relaxation est polynomial en n .
- si $\beta > \beta_c$, le temps de relaxation est exponentiel en n .

Pour une illustration de ceci, voire les dessins 9.

De plus, l'article récent et célèbre [4] a montré qu'on a *cut-off* pour $\beta < \beta_c$. Cet exemple illustre en particulier un phénomène encore très mal compris: un *cut-off* semble souvent s'accompagner d'un mélange de la chaîne assez rapide. De plus, il est à noter que la température critique β_c émerge déjà dans l'étude du modèle d'Ising statique, i.e. de la mesure de Gibbs, et ceci n'est que le début d'une interaction apparemment très riche entre la situation statique et la situation dynamique, i.e. la dynamique de Glauber appliqué au modèle d'Ising.

Le lecteur avide de problèmes supplémentaires pourra consulter [2], qui présente une grande partie des problèmes ouverts sur les sujets évoqués ici (et il y en a beaucoup!).

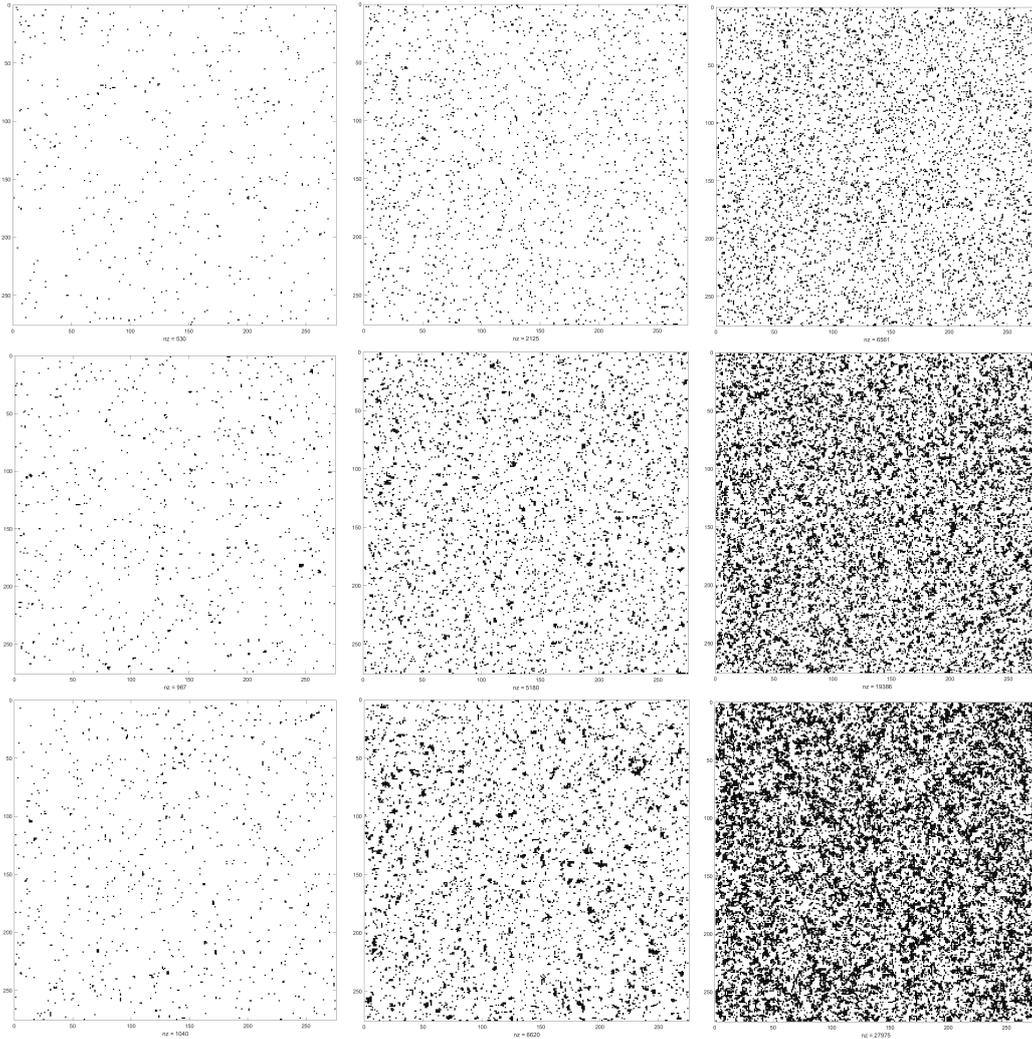


Figure 9: Dynamique de Glauber pour le modèle d’Ising sur le tore $(\mathbb{Z}/275\mathbb{Z})^2$, de configuration initiale la configuration de spins constante égale à -1 . À gauche pour basse température $\beta = 0.6$, au milieu pour la température critique $\beta = \beta_c$, à droite pour haute température $\beta = 0.3$, aux temps 100’000, 500’000 et 1’000’000 de haut en bas. On voit qu’à basse température, la configuration dépend encore très fortement de la configuration initiale, ce qui n’est moins le cas pour la température critique, et pas le cas à haute température.

A Annexe de code

Nous présentons succinctement le code MATLAB et les idées sous-jacentes utilisées pour les simulations.

A.1 Calculs numérique pour l’hypercube

Vu que l’espace d’états de la marche aléatoire paresseuse sur l’hypercube de dimension n est de cardinal 2^n , la matrice de transition devient très vite beaucoup trop grande pour un traitement numérique en temps raisonnable. L’idée principale pour le rendre accessible d’un point de vue numérique est la suivante: si (X_t) est marche aléatoire paresseuse sur l’hypercube de dimension n , alors on peut vérifier que $(|X_t|)$, i.e. la suite de nombres de 1 associée, est une chaîne de Markov

d'espace d'états $[[0, n]]$ et de matrice de transition

$$P(i, j) = \begin{cases} \frac{1}{2} & \text{si } j = i \\ \frac{n-i}{2n} & \text{si } j = i + 1 \\ \frac{i}{2n} & \text{si } j = i - 1 \\ 0 & \text{sinon.} \end{cases}$$

C'est la version paresseuse d'une chaîne de Markov qu'on appelle *urne d'Ehrenfest*. On dit de cette chaîne que c'est une *projection* de la marche aléatoire sur l'hypercube, dans un sens plus largement discuté par [2]. La chose remarquable est que cette projection conserve de nombreuses propriétés probabilistes de la marche sur l'hypercube, dont les distances d, \bar{d}, s . On a donc réduit la taille de la matrice de transition de $2^n \times 2^n$ à $(n + 1) \times (n + 1)$, ce que permet un traitement numérique d'une manière explicite, sans avoir à trouver un algorithme sophistiqué.

A.2 Distance de variation totale

La fonction *totvardis* calcule la distance de variation totale d'une matrice de transition, et éventuellement le temps de mélange. On utilise le fait que $d(t) = \frac{1}{2} \|P^t - \Pi\|_\infty$, où Π est la matrice carrée ayant le nombre qu'il faut de lignes égales à π , et où $\|\cdot\|_\infty$ est la norme matricielle associée à la norme $\|\cdot\|_\infty$ sur \mathbb{R}^n , c'est à dire la norme de l'application linéaire associée à la matrice. À part cela, on n'utilise pas d'algorithme particulier; on calcule tout de manière explicite.

```
function [ d , varargout ] = totvardis( P , pi , varargin )
%TOTVARDIS Calculate total variation distance and, if needed, also tmix.
%   expected input:
%       P: a stochastic matrix, irreducible and aperiodic
%       pi: the stationary distribution of P
%       varargin: empty or equal to T, a nonnegative integer
%   expected number of outputs:
%       1 or 2, any further variables to store outputs are left untouched
%
%   d = totvardis(P,pi) returns total variation distance on [0,2*tmix].
%
%   d = totvardis(P,pi,T) returns total variation distance on [0,T].
%
%   [d, tmix] = totvardis(P,pi) returns total variation distance on [0,2*tmix]
%   and tmix.
%
%   [d, tmix] = totvardis(P,pi,T) returns total variation distance on [0,T] and
%   tmix.
%
n = size(P);
n = n(1,1); % size of P and hopefully pi

Pt = eye(n); % P^t at current t = 0
Pi = ones(n,1)*pi; % square matrix with rows equal to pi

tmix = 0; % mixing time, to be calculated if nargout == 2 (or used if nargin ==
2)

if (nargout > 1) % i.e. if tmix is required to be returned
    d = norm(Pt-Pi, 'inf')/2; % value of d at t=0

    while (d(tmix+1)>0.25) % iterates until d drops below 0.25
```

```

    tmix = tmix + 1; % increases tmix by one as d hasn't dropped below 0.25
    Pt = Pt*P; % updates P^t to the current time t=tmix (conceptually, tmix
        has two roles in this portion of the code; in this while loop, it is
        the current time which is updated until d drops below 0.25, and
        then tmix will be equal to the mixing time)

    d(tmix+1) = norm(Pt-Pi,'inf')/2; % calculates d(t) at current time t=
        tmix
end % calculates tmix and d on [1,tmix]

varargout{1} = tmix; % to make it compatible with more outputs than expected
, all output arguments after the second one are simply left untouched

if (nargin == 2) % i.e. if the interval on which d is to be calculated isn't
    given by the input of the user, and thus is [[0,2*tmix]] by default
    d = [d zeros(1,tmix)]; % allcoates already enough space

    for t = tmix+1 : 2*tmix
        Pt = Pt * P; % updates P^t to current time t
        d(t+1) = norm(Pt-Pi,'inf')/2; % calculates d(t) at current time t
    end % calculates d on remaining [tmix+1,2*tmix]
else % i.e. if the interval [[0,T]] on which d is to be calculated is given
    by the input of the user
    T = varargin{1}; % to make it compatible with more inputs than expected,
        all inputs after the third one are simply ignored
    dif_t = T - tmix;

    if (dif_t < 0) % i.e. if d is already calculated on [0,T]
        d = d(1,1:T+1);
    elseif (dif_t > 0)
        d = [d zeros(1,dif_t)]; % allcoates already enough space

        for t = tmix+1 : T
            Pt = Pt * P; % updates P^t to current time t
            d(1,t+1) = norm(Pt-Pi,'inf')/2; % calculates d(t) at current
                time t
        end % calculates d on the remaining [tmix+1,T]
    end
end
elseif (nargout == 1) % i.e. if tsep isn't to be returned
    if (nargin == 2) % i.e. if the interval on which d is to be calculated isn't
        given by the input of the user, and thus is [[0,2*tmix]] by default
        d = norm(Pt-Pi,'inf')/2; % value of d(t) at t=0

        while (d(tmix+1)>0.25)
            tmix = tmix + 1; % increases tmix by one as d hasn't dropped below
                0.25
            Pt = Pt*P; % updates P^t to the current time t=tmix (conceptually,
                tmix has two roles in this portion of the code; in this while
                loop, it is the current time which is updated until d drops
                below 0.25, and then tmix will be equal to the mixing time)

            d(tmix+1) = norm(Pt-Pi,'inf')/2; % calculates d(t) at current time t
                =tmix
        end % calculates d on [1,tmix]

        d = [d zeros(1,tmix)]; % allcoates already enough space

        for t = tmix+1 : 2*tmix

```

```

        Pt = Pt * P; % updates P^t to current time t
        d(t+1) = norm(Pt-Pi,'inf')/2; % calculates d(t) for current time t
    end % calculates d on remaining [tmix+1,2*tmix]
else % i.e. if the interval [[0,T]] on which d is to be calculated is given
    by the input of the user
    T = varargin{1}; % to make it compatible with more inputs than expected,
        all inputs after the third one are simply ignored
    d = zeros(1,T+1); % allcoates already enough space
    d(1,1) = norm(Pt-Pi,'inf')/2; % value of d at t=0

    for t = 1 : T
        Pt = Pt * P; % updates P^t to current time t
        d(1,t+1) = norm(Pt-Pi,'inf')/2; % calculates d(t) for current time t
    end % calculates d on [1,T]
end
end
end
end

```

A.3 Distance \bar{d}

Les mêmes idée que pour la distance de variation totale sont appliquées ici.

```

function [ dbar ] = totvardisbar( P , T )
%TOTVARDISBAR Calculate "barred" total variation distance on a given interval.
%   expected input:
%       P: a stochastic matrix
%       T: a nonnegative integer
%
%   dbar = totvardisbar(P,T) returns the "barred" total variation distance of P
%   on the interval [0,T]
%
n = size(P);
n = n(1,1); % get the dimension

Pt = eye(n); % initialize P^t for current time t=0

dbar = zeros(1,T+1); % already allocating enough space to store the T+1 values
    of dbar on [[0,T]]
dbar(1,1) = 1; % initalizing dbar(0), which is always equal to 1

for t = 1 : T
    Pt = Pt * P; % update P^t to current time
    for i = 1 : n
        TV = norm(Pt-ones(n,1)*Pt(i,:), 'inf')/2; % calculates the maximum of ||P
            ^t(i,.)-P^t(j,.)||_TV over all j.
        if (TV > dbar(1,t+1))
            dbar(1,t+1) = TV;
        end % if current maximum, where one element of the state space (that is
            i) is fixed, is greater than the current value stored in dbar(t+1) (
            which is 0 befor the current for-loop), then dbar(t+1) is updated to
            this greater value.
    end % iterating the above procedure over all of the state space will store
        in dbar(t+1) the maximum of ||P^t(i,.)-P^t(j,.)||_TV over all i,j, which
        is the "barred" total variation distance at time t.
end % for-loop corresponding to time passing by

end
end

```

A.4 Distance de séparation

À nouveau, on calcule tout de manière explicite.

```
function [ s , varargout ] = sepdis( P , pi , varargin )
%SEPDIS Calculate separation distance and, if needed, also tsep, the mixing time
% with respect to the separation distance.
% expected input:
%   P: a stochastic matrix, irreducible and aperiodic
%   pi: the stationary distribution of P
%   varargin: empty or equal to T, a nonnegative integer. If more inputs are
%   given, they are ignored.
% expected number of outputs:
%   1 or 2, any further variables to store outputs are left untouched
%
% s = sepdis(P,pi) returns separation distance on [0,2*tsep], without
% returning tsep.
%
% d = sepdis(P,pi,T) returns separation distance on [0,T], without calculating
% tsep.
%
% [d, tsep] = sepdis(P,pi) returns separation distance on [0,2*tsep] and tsep.
%
% [d, tsep] = sepdis(P,pi,T) returns separation distance on [0,T] and tsep.
%
n = size(P);
n = n(1,1); % size of P and hopefully pi

Pt = eye(n); % P^t at current time t = 0
Pi = ones(n,1)*pi; % square matrix with rows equal to pi

tsep = 0; % mixing time for separation distance, to be calculated if nargin ==
% 2 (or used if nargin == 2)

if (nargout > 1) % i.e. if tsep is required to be returned
    St = ones(n,n) - Pt ./ Pi; % square matrix where the value at position x,y
    % is 1-P^t(x,y)/pi(y), here at current time t=0
    s(1,1) = max(St(:)); % maximum value of St, which is the separation distance
    % at current time t=0

    while (s(tsep+1)>0.25) % iterates until s drops below 0.25
        tsep = tsep + 1; % increases tsep by one as s hasn't dropped below 0.25
        Pt = Pt*P; % updates P^t to the current time t=tsep (conceptually, tsep
        % has to roles in the code; in this while loop, it is the current time
        % which is updated until s drops below 0.25, and then tsep will be
        % equal to the mixing time for the separation distance)

        St = ones(n,n) - Pt ./ Pi; % calculates St at current time t=tsep
        s(1,tsep+1) = max(St(:)); % calculates s at current time t=tsep
    end % calculates tsep and s on [1,tsep]

    varargout{1} = tsep; % to make it compatible with more outputs than expected
    % , all output arguments after the second one are simply left untouched

    if (nargin == 2) % i.e. if the interval on which s is to be calculated isn't
    % given by the input of the user, and thus is [[0,2*tsep]] by default
        s = [s zeros(1,tsep)]; % allcoates already enough space

        for t = tsep+1 : 2*tsep
```

```

        Pt = Pt * P; % updates P^t to current time t

        St = ones(n,n) - Pt ./ Pi; % calculates St at current time t
        s(1,t+1) = max(St(:)); % calculates s at current time t
    end % calculates s on remaining [tsep+1,2*tsep]
else % i.e. if the interval [[0,T]] on which s is to be calculated is given
    by the input of the user
    T = varargin{1}; % to make it compatible with more inputs than expected,
        all inputs after the third one are simply ignored
    dif_t = T - tsep;

    if (dif_t < 0) % i.e. if s is already calculated on [0,T]
        s = s(1,1:T+1);
    elseif (dif_t > 0)
        s = [s zeros(1,dif_t)]; % allcoates already enough space

        for t = tsep+1 : T
            Pt = Pt * P; % updates P^t to current time t

            St = ones(n,n) - Pt ./ Pi; % calculates St at current time t
            s(1,t+1) = max(St(:)); % calculates s at current time t
        end % calculates s on the remaining [tsep+1,T]
    end
end
elseif (nargout == 1) % i.e. if tsep isn't to be returned
    if (nargin == 2) % i.e. if the interval on which s is to be calculated isn't
        given by the input of the user, and thus is [[0,2*tsep]] by default
        St = ones(n,n) - Pt ./ Pi; % square matrix where the value at position x
            ,y is 1-P^t(x,y)/pi(y), here at current time t=0
        s(1,1) = max(St(:)); % maximum value of St, which is the seperation
            distance at current time t=0

        while (s(tsep+1)>0.25) % iterates until s drops below 0.25
            tsep = tsep + 1; % increases tsep by one as s hasn't dropped below
                0.25
            Pt = Pt*P; % updates P^t to the current time t=tsep (conceptually,
                tsep has to roles in the code; in this while loop, it is the
                current time which is updated until s drops below 0.25, and then
                tsep will be equal to the mixing time for the separation
                distance)

            St = ones(n,n) - Pt ./ Pi; % calculates St at current time t=tsep
            s(1,tsep+1) = max(St(:)); % calculates s at current time t=tsep
        end % calculates tsep and s on [1,tsep]

        s = [s zeros(1,tsep)]; % allcoates already enough space

        for t = tsep+1 : 2*tsep
            Pt = Pt * P; % updates P^t to current time t

            St = ones(n,n) - Pt ./ Pi; % calculates St at current time t
            s(1,t+1) = max(St(:)); % calculates s at current time t
        end % calculates s on remaining [tsep+1,2*tsep]
    else % i.e. if the interval [[0,T]] on which s is to be calculated is given
        by the input of the user
        T = varargin{1}; % to make it compatible with more inputs than expected,
            all inputs after the third one are simply ignored
        s = zeros(1,T+1); % allcoates already enough space
    end
end

```

```

St = ones(n,n) - Pt ./ Pi; % square matrix where the value at position x
    ,y is 1-P^t(x,y)/pi(y), here at current time t=0
s(1,1) = max(St(:)); % maximum value of St, which is the separation
    distance at current time t=0

for t = 1 : T
    Pt = Pt * P; % updates P^t to current time t

    St = ones(n,n) - Pt ./ Pi; % square matrix where the value at
        position x,y is 1-P^t(x,y)/pi(y), here at current time t
    s(1,t+1) = max(St(:)); % maximum value of St, which is the
        separation distance at current time t
end % calculates s on [1,T]
end
end
end

```

A.5 Le modèle d'Ising

On suit en substance la procédure décrite par la dynamique de Glauber dans la Définition 4.2 en la combinaison avec les équations (22) et (23).

```

function [ sigmaT , act ] = ising_simulation( sigma , beta , T , varargin)
%ISING_SIMULATION Simulates the Glauber dynamics for the Ising model on the
    torus for a given number of steps.
% expected input:
%     sigma: a matrix with all entries in {-1,1}, the starting configuration
%     beta: real number, with physical interpretation of inverse temperature
%     T: non negative integer, the number of transitions made according to the
    Glauber dynamics
%     varargin: empty or not, if empty the simulation is on the rectangle, if
    not it is on the torus i.e. with periodic boundary conditions
%
% sigmaT = ising_simulation(sigma,beta,T) retruns a spin configuration of the
    same size as sigma, corresponding to a T-step simulation of the Glauber
    dynamics for the Ising model on the rectangle.
%
% [sigmaT,act] = ising_simulation(sigma,beta,T) additionally retruns the matrix
    'act' of same size as S, with entries in [0,1], where act(i,j) describes
    the point in time where the spin at position (i,j) was last updated, 1
    meaning it was the most recent entry to be updated, 0 meaning it was never
    updated.
%
% sigmaT = ising_simulation(sigma,beta,T,x) retruns a spin configuration of
    the same size as sigma, corresponding to a T-step simulation of the Glauber
    dynamics for the Ising model on the torus, i.e. with periodic boundary
    conditions. x can be anything, only its presence in the input matters.
%
% [sigmaT,act] = ising_simulation(sigma,beta,T,'p') same as above for [sigmaT,
    support] = ising_torus_simulation(sigma,beta,T)
%
sigmaT = sigma;

[M, N] = size(sigma); % gets size of S
act = zeros(M,N);

```

```

periodic = 0;
if (nargin > 3)
    periodic = 1;
end % if no additional inputs are given, 'periodic' contains 0, corresponding to
    the rectangular case, and if additional inputs are given, 'periodic'
    contains 1, corresponding to the periodic case.

t = 0;

while (t < T) % time iteration
    i = ceil(M*rand); % generates random row index
    j = ceil(N*rand); % generates random column index

    Su = 0; Sd = 0; Sl = 0; Sr = 0; % will be used to store the spins of the
        neighbors of the selected entry (i,j); the small letter indicates the
        direction of the neighbor relative to (i,j)

    if (i == 1 && periodic == 1)
        Su = sigmaT(M,j);
    elseif (i > 1)
        Su = sigmaT(i-1,j);
    end

    if (i == M && periodic == 1)
        Sd = sigmaT(1,j);
    elseif (i < M)
        Sd = sigmaT(i+1,j);
    end

    if (j == 1 && periodic == 1)
        Sl = sigmaT(i,N);
    elseif (j > 1)
        Sl = sigmaT(i,j-1);
    end

    if (j == N && periodic == 1)
        Sr = sigmaT(i,1);
    elseif (j < N)
        Sr = sigmaT(i,j+1);
    end

    S = Su + Sd + Sl + Sr; % sum of spins of the neighbors of (i,j)
    p = (1 + tanh(beta*S))/2; % probability that the spin at (i,j) is replaced
        by +1

    if (p > rand)
        sigmaT(i,j) = 1;
    else
        sigmaT(i,j) = -1;
    end % updates spin at (i,j) according to p

    act = (t-1)/t * act;
    act(i,j) = 1; % updates actualisation matrix

    t = t + 1;
end
end

```

Références

- [1] D. Aldous, *Random walks on finite groups and rapidly mixing Markov chains*, Seminar on probability, XVII, Lecture Notes in Math., Springer, Berlin, **986** (1983), 243–297
- [2] D.A. Levin, Y. Peres, E.L. Widmer, *Markov Chains and Mixing Times*, Amer. Math. Soc., USA, 2009, disponible à <http://pages.uoregon.edu/dlevin/MARKOV/markovmixing.pdf>
- [3] E. Lubetzky, A. Sly, *Critical Ising on the square lattice mixes in polynomial time*, ArXiv e-prints, 2010, disponible à <https://arxiv.org/pdf/1001.1613.pdf>
- [4] E. Lubetzky, A. Sly, *Cutoff for the Ising model on the lattice*, Invent. math., **191** (2013), 719–755, disponible à <https://arxiv.org/pdf/0909.4320.pdf>