

Le phénomène de cutoff pour les chaînes de Markov

Lucas Rey

June 15, 2022

1 Résumé

Dans ce texte, nous abordons l'étude du temps de mélange des chaînes de Markov et plus précisément le phénomène de transition de phase qu'on appelle cutoff dans ce contexte. Nous donnons d'abord une introduction générale aux concepts et un historique, puis nous décrivons un peu plus en détail l'étude du phénomène de cutoff dans le cadre du zero-range process. Dans une dernière partie, nous évoquons un des problèmes ouverts du domaine: la détermination de conditions nécessaires et suffisantes pour qu'il y ait un cutoff.

2 Introduction: convergence des chaînes de Markov vers leur équilibre et phénomène de cutoff

De manière générale, on notera $(X_t)_{t \in \mathbb{N}}$ une chaîne de Markov à temps discret de matrice de transition P sur un espace d'état \mathcal{X} ou $(X_t)_{t \in \mathbb{R}^+}$ un processus de Markov à temps continu, de générateur \mathcal{L} . On supposera implicitement que les chaînes de Markov sont réversibles, mais le plus souvent ces résultats sont généralisables aux chaînes de Markov non réversibles en remplaçant les méthodes spectrales par des inégalités de Poincaré.

2.1 Mesure du mélange

Un des thèmes centraux de la théorie des chaînes de Markov est la convergence vers une loi stationnaire pour les chaînes irréductibles et apériodiques. Nous allons nous intéresser seulement aux chaînes de Markov irréductibles sur des espaces d'état finis qui sont automatiquement ergodiques : leur loi converge vers une loi stationnaire π . Une première étape de la théorie consiste à étudier la loi stationnaire, sur des graphes finis ou infinis. Elle peut être décrite plus ou moins précisément selon les cas (voir par exemple la discussion sur le zero-range process plus loin dans 3.1). Que l'on comprenne ou non la loi stationnaire, on peut s'intéresser à la vitesse de convergence vers cette loi: ce problème est notamment motivé par l'informatique, où l'on a besoin de bornes quantitatives sur la distance à l'équilibre. Pour échantillonner une loi de probabilité π sur un espace \mathcal{X} qui n'est pas explicite, il est fréquent de passer par une chaîne de Markov dont π est la loi stationnaire. C'est par exemple le cas de la dynamique de Glauber (parfois appelée dynamique heat-bath) que l'on utilise pour simuler le modèle d'Ising ou d'autres modèles de physique statistique. C'est un exemple où l'on connaît explicitement les marginales de la loi qu'on veut échantillonner, mais pas la loi elle-même et on aimerait savoir à quelle vitesse la convergence vers l'équilibre s'effectue. Une première étape pour cette étude est de définir une notion de distance à l'équilibre. Celle que l'on utilisera dans ce texte (et qui est la plus utilisée dans la littérature sur le phénomène de cutoff) est la distance en variation totale entre deux lois μ et ν sur un espace \mathcal{X} :

$$d_{TV}(\mu, \nu) = \sup_{A \subset \mathcal{X}} |\mu(A) - \nu(A)| = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)| \quad (1)$$

C'est une distance L_1 qui a une interprétation probabiliste intuitive (voir la remarque plus loin 2.1). D'autres distances utilisées pour mesurer la distance à l'équilibre sont les distances L_p pour $p > 1$, les distances de transport ou la divergence entropique. La distance en variation totale est bien adaptée à l'étude du mélange car elle vérifie pour toutes lois initiales μ et ν sur \mathcal{X} :

$$d_{TV}(\mu P, \nu P) \leq d_{TV}(\mu, \nu). \quad (2)$$

Si on note $\mu_t = \mu P_t$ la loi de la chaîne au temps t avec condition initiale μ , le théorème ergodique se reformule

$$d_{TV}(\mu_t, \pi) \rightarrow 0 \text{ quand } t \rightarrow \infty \quad (3)$$

donc la distance de la loi d'une chaîne de Markov à sa mesure stationnaire décroît vers 0. Une première notion de vitesse de convergence est celle de vitesse asymptotique. On note $D(\mu, t) = d_{TV}(\mu_t, \pi)$ la distance à l'équilibre depuis la configuration initiale μ et $D_*(t) = \sup_{\mu} D(\mu, t)$ la distance à l'équilibre dans le pire des cas. Cette suite vérifie la propriété de sous-multiplicativité:

$$D_*(t+s) \leq 2D_*(t)D_*(s). \quad (4)$$

Par le lemme de Fekete on montre l'existence d'une limite $\frac{1}{t} \ln D_*(t) \rightarrow -\lambda$, en d'autres termes $D_*(t) \approx \exp\left(-\frac{t}{t_{rel}}\right)$ où $t_{rel} = \frac{1}{\lambda}$ est appelé le temps de relaxation. Dans le cas où la chaîne de Markov est réversible, il est relié simplement au trou spectral de la matrice de transition (ou du générateur pour le cas du temps continu). Ce temps de relaxation est simple à manier, mais en tant que vitesse asymptotique il n'a que peu d'intérêt en pratique. C'est pourquoi on définit le temps de mélange à hauteur ε depuis la loi initiale μ (souvent implicite dans les notations) :

$$t_{mix}(\mu, \varepsilon) = t_{mix}(\varepsilon) = \inf\{t \in \mathbb{N} \text{ tel que } d_{TV}(\mu_t, \pi) \leq \varepsilon\} \quad (5)$$

On s'intéresse aussi au temps de mélange dans le pire des cas:

$$t_{mix}^*(\varepsilon) = \inf\{t \in \mathbb{N} \text{ tel que } D_*(t) \leq \varepsilon\} \quad (6)$$

Remarque 2.1. On peut prendre l'exemple des mélanges de cartes pour montrer pourquoi cet objet est intuitif. On modélise un mélange de cartes par une chaîne de Markov sur l'espace $\mathcal{X} = S_n$ des permutations de n cartes. Les états successifs X_t de la chaîne de Markov représentent l'état du paquet après t mélanges. Si on part d'un paquet trié, i.e. $X_0 = \text{id}$, et que après t étapes $d_{TV}(\mu_t, \pi)$ est encore proche de 1, cela veut dire par définition qu'il existe des événements A ayant une probabilité proche de 1 pour notre paquet mais une probabilité d'ordre 0 pour la distribution uniforme sur S_n . Un parieur habile peut alors exploiter ce mauvais mélange en pariant sur l'événement A . Il lui suffit de connaître la procédure de mélange i.e. la matrice de transition P (même un modèle approximatif du mélange permet de trouver de tels événements, par exemple pour le riffle shuffle [5]). Au contraire si $d_{TV}(\mu_t, \pi)$ est proche de 0 tous les événements ont presque la même probabilité que si le paquet était parfaitement mélangé.

Remarque 2.2. Mesurer la distance à l'équilibre en variation totale permet d'utiliser la technique probabiliste de couplage. En effet, on a le résultat:

$$d_{TV}(\mu, \nu) = \inf\{\mathbb{P}[X \neq Y], (X, Y) \text{ couplage de } \mu \text{ et } \nu\} \quad (7)$$

Dans le cadre des chaînes de Markov, on trouve naturellement des couplages de la loi à tout temps : il suffit de coupler le premier pas depuis chaque paire d'états voisins $(x, y) \in \mathcal{X}$. Si on note $\mu_t(x) = \delta_x P^t$ la loi au temps t partant d'un état x , et si l'on dispose d'un couplage (X_t, Y_t) de $(\mu_t(x), \mu_t(y))$ pour tous $x, y \in \mathcal{X}$, on peut facilement le modifier pour que les chaînes restent identiques une fois qu'elles se sont rejointes, et en notant $\tau(x, y)$ le temps de coalescence:

$$D_*(t) \leq \sup_{x, y \in \mathcal{X}} d_{TV}(\mu_t(x), \mu_t(y)) \leq \sup_{x, y} \mathbb{P}[\tau(x, y) > t]. \quad (8)$$

2.2 Le phénomène de cutoff

Le phénomène de cutoff a été découvert par Diaconis et Shahshahani en cherchant à estimer t_{mix} sur un modèle de mélange de cartes dit random-to-random : à chaque étape, on choisit deux cartes au hasard dans un paquet et on les échange. C'est une marche aléatoire sur le groupe symétrique S_n , la loi des incréments étant donnée par $Q(id) = \frac{1}{n}$, $Q(\tau) = \frac{2}{n^2}$ pour toute transposition τ et $Q(\sigma) = 0$ pour toute autre permutation. Ils montrent dans [7] que le temps de mélange pour ce mélange sur n cartes est d'ordre $\frac{1}{2}n \ln n$, et ce indépendamment de la précision ε :

$$t_{mix}(\varepsilon) \sim \frac{1}{2}n \ln n \quad \forall \varepsilon > 0 \quad (9)$$

Remarque 2.3. La preuve de ce résultat illustre bien l'idée générale des preuves de cutoff : la borne inférieure est relativement simple à montrer, car par le problème du coupon-collector il est facile de voir qu'aux temps $t \approx (\frac{1}{2} - \delta)n \ln n$, un grand nombre $N_t \gg 1$ de cartes n'ont pas changé de places i.e. la permutation X_t a un grand nombre de points fixes (avec grande probabilité) tandis que le nombre de points fixes d'une permutation aléatoire sur n éléments est $O(1)$ (d'espérance 1 par un calcul bien connu). La borne supérieure est quant à elle beaucoup plus difficile à montrer, elle est obtenue en majorant la distance en variation totale par la distance L_2 puis en utilisant l'analyse de Fourier sur les représentations du groupe symétrique pour calculer cette distance L_2 .

Ce résultat a amené à définir le phénomène de cutoff de façon plus générale par la suite, pour des chaînes de Markov "de plus en plus grosses":

Définition 2.4. On se donne pour tout n une chaîne de Markov $(X_t^n)_{t \in \mathbb{N}}$ sur un espace d'état \mathcal{X}^n . pour alléger les notations, la dépendance en n est souvent implicite : on notera $t_{mix}(\varepsilon)$ pour $t_{mix}^n(\varepsilon)$. On dit que cette suite présente un cutoff si $\forall \varepsilon, \varepsilon' > 0$,

$$\frac{t_{mix}^*(\varepsilon)}{t_{mix}^*(\varepsilon')} \rightarrow 1 \quad \text{quand } n \rightarrow \infty \quad (10)$$

Ce phénomène de cutoff est un phénomène de transition de phase qui arrive au temps $t_{mix}(0.25)$: avant ce temps, la chaîne de Markov n'est pas du tout mélangée, après ce temps elle l'est complètement.

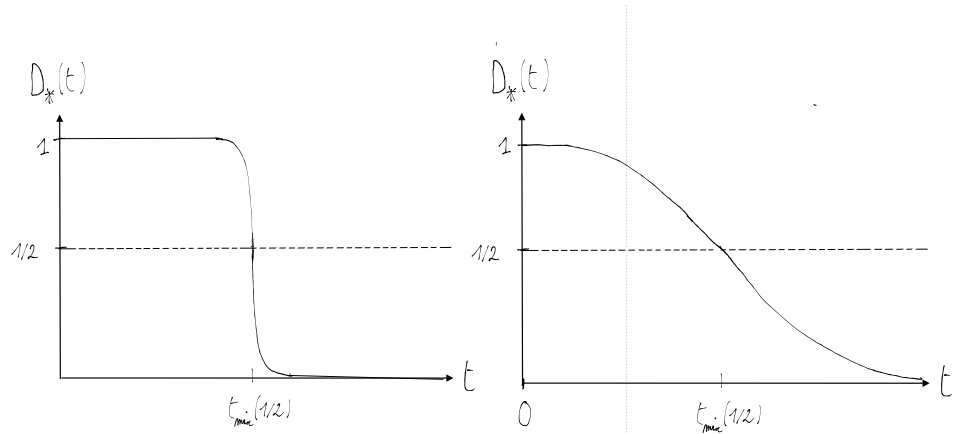


Figure 1: à gauche, distance à l'équilibre en fonction du temps en situation de cutoff, par exemple la marche par transposition sur S_n pour n grand. A droite, même chose sans cutoff, par exemple la marche simple sur le cercle $\mathbb{Z}/n\mathbb{Z}$ pour n grand

Suite à cet article fondateur, ce phénomène de cutoff a été observé et démontré pour de nombreuses chaînes de Markov, voir [4] pour les premiers exemples historiques. Un exemple élémentaire est celui de

la marche aléatoire paresseuse sur l'hypercube $\{0,1\}^n$ dont on peut montrer qu'elle présente un cutoff au temps $\frac{1}{2}n \ln n$. La preuve consiste à donner une borne inférieure simple (en utilisant la méthode de Wilson et une analyse de l'espérance et la variance dans le problème du collecteur de coupon) et à donner une borne supérieure correspondante par couplage (2.2). Parmi les exemples les plus marquants (cette liste est loin d'être exhaustive), des cutoffs ont été démontrés sous des hypothèses parfois restrictives pour

- d'autres mélanges de carte, comme le riffle shuffle par Diaconis dans [5])
- de nombreux modèles de la physique statistique: le modèle d'Ising dans [12], le processus d'exclusion en dimension 1 dans [10], le mean-field zero-range-process dans [9]
- la marche aléatoire simple sur des graphes aléatoires dans [16]

Des phénomènes de cutoff ont aussi été démontrés dans le cadre plus général de processus de Markov sur des espaces d'états infinis (et notamment continus) :

- le mouvement brownien sur des groupes compacts dans [14]
- la marche d'adjacence sur le simplexe et des généralisations dans une suite de travaux de Caputo, Labbé, Lacoïn culminant en [19].

Enfin, le cutoff est conjecturé dans de nombreux autres cas pour lesquels on ne connaît qu'un ordre de grandeur du temps de mélange, par exemple la marche de Kac ([6]) i.e. la marche par rotations aléatoires sur le groupe unitaire $SO(n)$, ou le problème d'apparence plus simple du mélange de cartes où à chaque étape on décide soit d'échanger les deux premières cartes soit de mettre la première carte à la fin du paquet (soit de ne rien faire pour éviter les problèmes de parité de la signature). Dans de nombreux cas où un cutoff a été démontré, les méthodes de preuve ne s'adaptent pas à des modèles plus généraux : le zero-range process hors du champ moyen, le processus d'exclusion en dimension plus grande que 1, la marche simple sur des graphes aléatoires plus généraux.

3 Le cas du Zero-Range Process

3.1 Le modèle

Pour rentrer dans le détail des idées et des arguments utilisés dans l'étude du phénomène de cutoff, on va prendre l'exemple du zero-range process (ZRP) tel qu'il est étudié dans [9]. Ce modèle est plus naturel à décrire en temps continu, mais tous les concepts définis précédemment pour les chaînes de Markov s'adaptent immédiatement aux processus de Markov à temps continu. Le ZRP est un modèle de physique statistique qui décrit des particules indistinguables en interaction, dont l'interaction est paramétrée par une fonction $r : \mathbb{N} \rightarrow \mathbb{R}^+$ décrivant à quel point les particules se repoussent ou s'attirent. On modélise le milieu environnant par un graphe G . A un instant t donné chaque particule est sur un des sommets du graphe, et il peut y avoir plusieurs particules sur le même sommet. On représente cette configuration par un vecteur $X(t)$ où $X_i(t)$ indique le nombre de particules présentes sur le site $i \in G$ au temps t . Chaque site expulse indépendamment une particule vers un des sommets voisins (choisi uniformément au hasard) selon une loi de Poisson dont le paramètre $r(X_i(t))$ ne dépend que du nombre de particules présentes sur le site. C'est pourquoi on parle de zero-range process: les particules ne "voient" que ce qu'il se passe là où elles sont. Dans ce modèle, le nombre m de particules est conservé au cours du temps. L'espace d'état de ce processus de Markov est alors $\Omega = \{x \in \mathbb{N}^G \text{ tels que } \sum_{i \in G} x_i = m\}$, et on peut décrire formellement ce modèle en donnant l'action de son générateur L sur les observables $\phi : \Omega \rightarrow \mathbb{R}$: pour $x \in \Omega$,

$$(L\phi)(x) = \frac{1}{d_n} \sum_{i \in G} \sum_{j \sim i} r(x_i) (\phi(x + \delta_j - \delta_i) - \phi(x)). \tag{11}$$

Ce modèle a été introduit par Spitzer en 1970 dans [22], et a été très étudié depuis notamment car sa loi stationnaire prend une forme produit simple quels que soient le graphe et la fonction de taux : pour $x \in \Omega$, elle est proportionnelle à

$$\pi(x) \propto \prod_{i \in G} \prod_{k=1}^{x_i} \frac{1}{r(k)} \quad (12)$$

Pour étudier le cutoff on considère une famille de graphes G_n "de plus en plus gros" et on se place dans un régime de densités bornées et convergentes $\rho_n = \frac{m}{|G_n|} \rightarrow \rho \in]0, \infty[$. Si G_n est le graphe complet sur n sommets, on parle de mean-field zero-range process. Dans ce cadre, on peut montrer (voir [9]) que

Théorème 3.1. *Il y a cutoff pour le mean field-zero ZRP sur n sommets avec fonction de taux r bornée croissante au temps $t_{mix}^*(\epsilon) = \gamma n$ où γ est une constante ne dépendant que de r et ρ .*

Ce théorème est en fait un corollaire d'un résultat plus précis qui décrit le mélange depuis les états $x \in \Omega$ où au moins un des sites contient une fraction importante de l'ensemble des particules dans la limite où $n \rightarrow \infty$. Dans le cas du graphe complet, notons (x_1^n, \dots, x_n^n) une telle configuration initiale et supposons que $\frac{x_i^n}{n} \rightarrow u_i$ quand $n \rightarrow \infty$ pour tout i fixé. On suppose de plus que $u_1 > 0$ et que les u_i sont décroissants en i à n fixé (cela ne change rien, par symétrie du ZRP sur le graphe complet). Dans le cas du ZRP avec fonction de taux $r(k) = \mathbf{1}_{k>0}$ (voir [13]), l'équation différentielle limite se résout explicitement et donne

$$\frac{t_{mix}(\delta_x, \epsilon)}{n} \rightarrow (1 + \rho)u_1 - \frac{1}{2} \sum_{i=1}^{\infty} u_i^2. \quad (13)$$

Remarque 3.2. Il est facile de se convaincre que c'est le bon ordre de grandeur, au moins pour la borne inférieure: la mesure d'équilibre π étant explicite, on peut l'étudier et observer qu'à l'équilibre le nombre maximum de particules en un site est $O(\ln n)$. On voit donc que le mélange va être long depuis les configurations initiales où les particules sont concentrées, par exemple si $x^n = (\rho n, 0, \dots, 0)$, comme les particules sont expulsées du premier site avec un taux 1, on voit qu'il faut un temps d'ordre au moins ρn pour que le nombre de particules dans le premier site devienne $o(n)$.

La preuve de ce résultat est faite en passant à la limite hydrodynamique pour le ZRP i.e. en renormalisant et en étudiant plutôt $\tilde{X}^n(t) = \frac{1}{n}X(nt)$. Cette procédure est une généralisation de la loi de grands nombres, où l'objet limite est obtenu comme solution d'une certaine équation différentielle, voir par exemple [3]. Un objet important dans l'étude de cette limite est la "température":

Définition 3.3. A un instant t , on définit la température en un sommet $i \in G$ par

$$\xi_i^n(t) = \sum_{j \sim i} \frac{1}{\deg(j)} r(X_j^n(t)) \quad (14)$$

et la température renormalisée par

$$\tilde{\xi}_i^n(t) = \xi_i^n(nt) \quad (15)$$

Cette température décrit le taux auquel les particules arrivent en i , donc le taux net de particules éjectées par i est donné par $r(X_i^n(t)) - \xi_i^n(t)$. Dans le modèle en champ moyen, cette température est la même pour tous les sites, on la note $\tilde{\xi}^n(t)$. Dans la limite hydrodynamique, le nombre de sauts des processus de Poisson est concentré et l'évolution devient déterministe: on montre que ces suites de variables aléatoires sont tendues et donc quitte à extraire (par argument diagonal) on a la convergence $\tilde{\xi}^n(t) \rightarrow \tilde{\xi}(t)$ et $\tilde{X}_i^n(t) \rightarrow U_i(t)$ pour tout $i \in \mathbb{N}$ quand $n \rightarrow \infty$ (au sens de la convergence en loi). On obtient une équation différentielle reliant $\tilde{\xi}$ et les $\tilde{U}_i(t)$: pour tout site i

$$\frac{dU_i}{dt} = \tilde{\xi}(t) - 1 \text{ tant que } U_i > 0 \quad (16)$$

On peut aussi exprimer la température au temps t en utilisant que la majorité des sites ne contiennent qu'un nombre $o(n)$ de particules, que ce sont eux qui contribuent majoritairement à la température, qu'ils

contiennent une densité de particules $\rho - \sum_i U_i(t)$ au temps t et que ces sites sont à l'équilibre (le mélange se fait à des échelles de temps $\ll n$ pour ces sites contenant peu de particules, c'est en fait le point le plus délicat de la preuve). On peut écrire

$$\tilde{\xi}(t) = \phi \left(\rho - \sum_i U_i(t) \right) \quad (17)$$

où ϕ est une fonction explicite ne dépendant que de r et décrivant la température d'un ZRP à l'équilibre. En résolvant le système composé des (16) et de (17), on peut expliciter une fonction $f : \mathbb{R}^+ \rightarrow \infty$ telle que

$$U_i(t) = [u_i - f(t)]_+ \text{ où } [x]_+ = \max(0, x) \quad (18)$$

Ces équations décrivent précisément l'évolution de la densité de particules des sites contenant un nombre initial de particules $\Theta(n)$. L'étape clé du reste de la preuve est de montrer que lorsque $\|X(t)\|_\infty = o(n)$, le mélange se fait en un temps $o(n)$, ce qui est obtenu dans [9] par un couplage délicat suivant la méthode introduite par Bubley et Dyer dans [1].

3.2 Des questions ouvertes

Le mean-field ZRP dont on vient de parler est en fait un modèle simpliste, et les modèles vraiment intéressants de ZRP pour la physique statistique sont ceux où le graphe sous-jacent G est plus complexe, typiquement le cas du mélange sur la grille \mathbb{Z}^d , avec $G_n = \{1, \dots, n\}^d$ ou $G_n = (\mathbb{Z}/n\mathbb{Z})^d$ le tore d -dimensionnel. Des résultats plus faibles sur le mélange sont déjà connus dans ce cadre, par exemple Morris montre dans [15] que le temps de relaxation est d'ordre $\Theta((\rho+1)^2 n^2)$, en analysant finement un couplage pour obtenir la borne supérieure. Ce résultat est généralisé par une technique différente dans [8], où l'étude du temps de relaxation du ZRP sur un graphe quelconque G est ramenée à l'étude du temps de relaxation du ZRP pour une seule particule (i.e. la marche aléatoire sur le graphe sous-jacent G) en prouvant une relation algébrique entre la forme de Dirichlet du ZRP avec m particules et celle du ZRP avec 1 particule. C'est une variante de la spectral-gap conjecture d'Aldous, qui concerne initialement le processus d'échanges de particules sur un graphe G . Tous ces résultats sont beaucoup moins fins que ceux que l'on a sur le graphe complet, et soulèvent les questions suivantes:

Question 3.4. Quel est le temps de mélange du ZRP avec fonction de taux monotone bornée et densité de particules ρ sur un graphe G quelconque ? Pour quels types de graphe y a-t-il cutoff ? En particulier, que se passe-t-il dans le cas où $G_n = (\mathbb{Z}/n\mathbb{Z})^d$?

4 Une question ouverte: CNS pour qu'il y ait cutoff

4.1 Motivation: la condition produit

Le phénomène de cutoff survient pour de nombreuses chaînes de Markov, mais dans presque tous les exemples où l'on arrive à démontrer ce cutoff la preuve passe par une étude détaillée du temps de mélange de la chaîne, et le plus souvent elle consiste à trouver une borne inférieure relativement facile par la méthode de Wilson et à montrer une borne supérieure correspondante par des couplages ou une analyse très détaillée du mélange. Une des questions ouvertes les plus intéressantes du domaine est de trouver des critères qui permettraient d'affirmer l'existence d'un cutoff sans avoir à calculer précisément le temps de mélange. Cette idée a été lancée par Peres en 2004 ([18]), où il proposait une condition simple pour que le cutoff ait lieu : la condition produit.

Question 4.1. Est-il vrai que si $t_{rel} = o(t_{mix}(0.25))$ quand $n \rightarrow \infty$, le phénomène de cutoff a lieu ? Et sinon, peut-on trouver des conditions supplémentaires pour que cette condition produit implique le cutoff ?

Une première remarque est que cette condition est très pratique, car elle ne demande pas de déterminer les constantes multiplicatives précises pour le temps de mélange et le temps de relaxation. Souvent, on arrive à obtenir des bornes inférieures et supérieures grossières sur le temps de relaxation ou le temps de mélange

par couplage, mais il est très difficile d’obtenir des bornes inférieure et supérieure qui coïncident. Une deuxième remarque est qu’il est facile de voir que cette condition est nécessaire: si un cutoff a lieu, $\forall \varepsilon > 0$,

$$\frac{t_{mix}(\varepsilon)}{t_{mix}(0.25)} \rightarrow 1. \quad (19)$$

Mais on sait facilement montrer pour toute chaîne de Markov réversible que

$$t_{mix}(\varepsilon) \geq t_{rel} \ln \left(\frac{1}{2\varepsilon} \right). \quad (20)$$

Donc $\forall \varepsilon > 0$,

$$t_{rel} \leq \frac{t_{mix}(\varepsilon)}{\ln \left(\frac{1}{2\varepsilon} \right)} \sim \frac{t_{mix}(0.25)}{\ln \left(\frac{1}{2\varepsilon} \right)} \quad (21)$$

Et comme cela est vrai pour tout ε il découle que $t_{rel} = o(t_{mix}(0.25))$. Ainsi cette condition est nécessaire, et dans les cas les plus simples où le cutoff n’a pas lieu cette condition est mise en défaut

Exemple 4.2. Pour la marche aléatoire sur le cercle $\mathbb{Z}/n\mathbb{Z}$, on peut diagonaliser la matrice de transition ce qui permet de calculer explicitement $t_{rel} \sim \frac{n^2}{\pi^2}$, et d’autre part par couplage on montre facilement que $t_{mix}(\varepsilon) \leq c_\varepsilon n^2$ pour une certaine constante c_ε ne dépendant que de ε : il n’y a donc pas cutoff.

Il a été montré par Chen et Saloff-Coste dans [2] que pour le mélange en distance L^p avec $p > 1$, la condition produit est une condition nécessaire et suffisante pour que le phénomène de cutoff se produise. Il a aussi été montré que pour les chaînes birth-and-death et plus généralement les chaînes de Markov fainéantes réversibles sur les arbres (voir le Théorème 1 de [20]), la condition produit est nécessaire et suffisante. En toute généralité cependant, cette condition produit n’implique pas le cutoff: un contre-exemple a été décrit par Aldous, que l’on peut trouver dans [11]. Ce contre-exemple est général dans le sens où en modifiant un peu une suite de chaînes de Markov qui présente un cutoff on produit une autre suite de chaînes vérifiant aussi la condition produit mais qui ne présente plus de cutoff. On le décrit dans la remarque suivante.

Remarque 4.3. Précisément, si une suite de chaînes de Markov réversibles est donnée par la suite des matrices de transition P_n et qu’elle présente un cutoff, on sait que $t_{rel} = o(t_{mix}(0.25))$. Si on note $L_n = \sqrt{t_{rel} t_{mix}(0.25)}$, on peut définir une nouvelle suite de matrices de transition

$$\tilde{P}_n = \left(1 - \frac{1}{L_n} \right) P_n + \frac{1}{L_n} \Pi_n \quad (22)$$

(où Π est la matrice de mélange instantané i.e. la matrice dont les colonnes sont identiques et égales à la mesure stationnaire π). A chaque étape, la chaîne fait un saut selon la loi Π_n avec probabilité $\frac{1}{L_n}$ et selon la loi P_n avec probabilité $\left(1 - \frac{1}{L_n} \right)$. Cette suite de chaînes ne présente pas de cutoff: le mélange a lieu progressivement autour du temps L_n , quand la chaîne fait son premier saut selon Π_n . Mais on montre que $\tilde{t}_{rel} \sim t_{rel}$ donc la condition produit est encore vérifiée pour cette suite de chaînes.

4.2 Le critère entropique

Une autre piste est de trouver des conditions supplémentaires à rajouter pour que la condition de Peres soit suffisante. Cette piste est poursuivie dans le travail récent de Salez [21] sur lequel nous allons nous attarder. Pour développer ce critère, il introduit la notion de varentropie d’une loi par rapport à la loi d’équilibre (cette notion existait déjà dans d’autres contextes mais n’était pas utilisée dans l’étude des temps de mélange). Si une mesure μ admet une densité f par rapport à une mesure de référence π , son entropie relative à la mesure de référence (aussi appelée divergence de Kullback-Leibler) est définie comme

$$d_{KL}(\mu|\pi) = \mathbb{E}_\pi[f \log f]. \quad (23)$$

Cette divergence est utilisée dans la littérature sur les cutoffs, notamment quand la majoration de la distance en variation totale par la distance L_2 est mauvaise. Par analogie avec la définition de la variance par rapport à la moyenne, il est naturel de considérer la quantité

$$\mathcal{V}_{KL}(\mu|\pi) = \mathbb{E}_\pi \left[(f \log f - d_{KL}(\mu|\pi))^2 \right]. \quad (24)$$

Le résultat général de [21] est que cette varentropie permet de contrôler le temps de mélange:

Théorème 4.4. *Soit (X_t) une chaîne de Markov irréductible ergodique sur un espace d'état fini \mathcal{X} de loi μ_t au temps t . Soit $\varepsilon \in]0, 0.5[$ quelconque. Alors*

$$t_{mix}(\mu, \varepsilon) - t_{mix}(\mu, 1 - \varepsilon) \leq \frac{2t_{rel}}{\varepsilon^2} \left(1 + \sqrt{\mathcal{V}_{KL}(\mu_{t_{mix}(1-\varepsilon)}|\pi)} \right) \quad (25)$$

Remarque 4.5. Dans [21], Salez étudie le temps de mélange depuis la pire configuration possible, donc la varentropie dans le membre de droite est remplacée par un maximum sur toutes les configurations initiales.

Ce résultat est prometteur car pour démontrer un cutoff, il suffit de montrer qu'à ε fixé, le membre de droite est négligeable devant $t_{mix}(\varepsilon)$ (ou $t_{mix}(0.25)$). Pour cela le terme délicat à estimer est la varentropie, mais une borne inférieure sur le temps de mélange (qui est souvent facile à obtenir par la méthode de Wilson) peut permettre de contrôler \mathcal{V}_{KL} au temps $t_{mix}(1 - \varepsilon)$ depuis n'importe quelle configuration.

Exemple 4.6. On peut illustrer cette idée dans le cas simple d'une chaîne produit, par exemple la marche aléatoire simple sur l'hypercube $\{0, 1\}^n$ partant de la condition initiale $(0, \dots, 0)$. On peut montrer en étudiant le nombre de coordonnées inchangées (on analyse l'espérance et la variance dans le problème du collectionneur de coupons puis on applique la méthode de Wilson) que

$$t_{mix}(\varepsilon) \geq \frac{1}{2}n \ln n - \phi(\varepsilon) \quad (26)$$

pour une certaine fonction ϕ indépendante de n . Comme l'entropie, la varentropie d'une chaîne produit est facile à calculer: en écrivant $\mu_t = \mu_t^1 \times \dots \times \mu_t^n$ (les μ_t^i étant identiques) et $\pi = \pi^1 \times \dots \times \pi^n$ (où $\pi_i = 0.5(\delta_0 + \delta_1)$ pour tout i est la loi stationnaire pour une des coordonnées),

$$\mathcal{V}_{KL}(\mu_{t_{mix}(1-\varepsilon)}|\pi) = \sum_{i=1}^n \mathcal{V}_{KL}(\mu_{t_{mix}(1-\varepsilon)}^i|\pi^i) = n \mathcal{V}_{KL}(\mu_{t_{mix}(1-\varepsilon)}^1|\pi^1) \quad (27)$$

On se ramène donc à l'étude de la varentropie en dimension 1 quand le temps tend vers l'infini. Or, il n'est pas difficile de montrer que $\mathcal{V}_{KL}(\mu_t^1|\pi^1) \approx \exp\left(-\frac{2t}{t_{rel}}\right)$ (au sens où le ln converge), car en temps grand les méthodes spectrales donnent les bons ordres de grandeur. La borne inférieure sur le temps de mélange permet donc de montrer que pour t d'ordre $\frac{t_{rel}}{2} \ln n = \frac{1}{2}n \ln n$, le terme varentropique est borné et (25) se ramène à la condition produit, qui est bien vérifiée ici.

Dans la suite de l'article, Salez montre que pour certaines chaînes la varentropie peut être majorée sans faire de calculs précis. Ce travail s'applique aux chaînes de Markov à courbure positive, une notion inspirée de la géométrie riemannienne qui est étudié par exemple dans [17]. Nous n'allons pas présenter cette condition en détail dans ce texte, mais nous l'introduisons dans la remarque suivante.

Remarque 4.7. Pour prouver qu'une chaîne de Markov sur un espace fini \mathcal{X} a une courbure positive, il suffit de trouver pour tous états voisins $x \sim y \in \mathcal{X}$ un couplage (X, Y) des lois $\mu_1^x = \delta_x P$ et $\mu_1^y = \delta_y P$ (la loi de la chaîne après un pas) tel que $\mathbb{E}[\text{dist}(X, Y)] \leq \text{dist}(x, y)$. Un tel couplage est souvent facile à trouver, par exemple pour la marche par transposition sur le groupe symétrique (et plus généralement les marches invariantes par conjugaison sur le groupe symétrique), pour le ZRP ou pour les marches aléatoires simples sur les groupes abéliens.

Sous cette condition de courbure positive, Salez développe un critère ne faisant intervenir que des ordres de grandeur qui assure qu'une suite de chaînes de Markov présente un cutoff.

Théorème 4.8. *Il existe une constante C telle que pour toute chaîne de Markov $(X_t)_{t \in \mathbb{N}}$ à courbure positive sur un espace \mathcal{X} de matrice de transition P à support symétrique, pour tout $t \geq \frac{\text{diam}(\mathcal{X})}{4}$ et $\varepsilon > 0$, on a*

$$\mathcal{V}_{KL}(\mu_t) \leq Ct(1 + \ln(\Delta))^2 \quad (28)$$

où Δ est un paramètre dépendant implicitement de n :

$$\Delta = \max \left\{ \frac{1}{P(x, y)}, x \sim y \in \mathcal{X}_n \right\} \quad (29)$$

Ainsi, si on sait que $t_{mix}(\varepsilon) \geq \frac{\text{diam}(\mathcal{X}_n)}{4}$ (ce qu'on peut vérifier en pratique si on a une borne inférieure sur le temps de mélange), en combinant ces deux résultats on obtient un critère pour le cutoff : il y a cutoff dès que

$$t_{mix}(0.25) \gg t_{rel}^2(1 + \ln \Delta)^2 \quad (30)$$

Comme annoncé, ce critère ressemble à la condition produit, et il ne met en jeu que des ordres de grandeur: des bornes inférieures et supérieures à constante multiplicative près sur le temps de relaxation et le temps de mélange permettent de démontrer un cutoff. Salez applique ce critère aux marches aléatoires sur des graphes aléatoires et redémontre les résultats de [16] dans ce cadre général. Insistons sur le fait que cela n'enlève rien au travail de [16] qui décrit de façon très précise la convergence vers l'équilibre de ces marches aléatoires, apportant beaucoup plus d'information que la simple connaissance du cutoff.

Ce principe de concentration entropique ouvre une nouvelle voie pour prouver des cutoffs. On a vu dans la remarque 4.6 qu'une borne inférieure précise sur le temps de mélange peut permettre de prouver un cutoff si on est capable d'estimer la varentropie. L'utilisation d'arguments de courbure est un moyen d'estimer cette varentropie. On a vu dans la remarque que pour des chaînes produit, il était facile d'estimer cette varentropie. On peut donc essayer de généraliser cette estimation au cas de chaînes "quasi indépendantes". Beaucoup d'autres applications de ce critère sont envisageables.

References

- [1] Russ Bubley and Martin Dyer. Path coupling: a technique for proving rapid mixing in markov chains. *Proceedings 38th Annual Symposium on Foundations of Computer Science*, 1997.
- [2] Guan-Yu Chen and Laurent Saloff-Coste. The cutoff phenomenon for ergodic markov processes. *EJP*, 13:26–78, 2008.
- [3] Richard W. R. Darling. Fluid limits of pure jumps markov processes: a practical guide. 2002.
- [4] Persi Diaconis. The cutoff phenomenon in finite markov chains. *Proc. Natl. Acad. Sci. USA*, 93:1659–1664, 1996.
- [5] Persi Diaconis and Dave Bayer. Trailing the dovetail shuffle to its lair. *Ann. App. Probab.*, 2(2):294–313.
- [6] Persi Diaconis and Laurent Saloff-Coste. Bounds for kac's master equation. *Comm. Math. Phys.*, 209(3):729–755, 2000.
- [7] Persi Diaconis and Mehrdad Shahshahani. Generating a random transposition with random permutations. *Z. Wahrscheinlichkeitstheorie verw Gebiete*, 57:159–179, 1981.
- [8] Jonathan Hermon and Justin Salez. A version of aldous' spectral gap conjecture for the zero range process. *The Annals of Applied Probability*, 2018.
- [9] Jonathan Hermon and Justin Salez. Cutoff for the mean-field zero-range process with bounded monotone rates. *Annals of Probability*, 48(2):742–759, 2020.

- [10] Hubert Lacoin. Mixing time and cutoff for the adjacent transposition shuffle and the simple exclusion. *Ann. Probab.*, 44:1426–1487, 2015.
- [11] David A. Levin and Yuval Peres. *Markov chains and mixing times, second edition*. AMS, 2008.
- [12] Eyal Lubetsky and Allan Sly. Universality of cutoff for the ising model. *Ann. Probab.*, 45:3664–3696, 2014.
- [13] Mathieu Merle and Justin Salez. Cutoff for the mean-field zero-range process. *Ann. of Probab.*, 47(5):3170–3201, 2019.
- [14] Pierre-Loïc Méliot. The cutoff phenomenon for brownian motion on compact symmetric spaces. *Potential anal.*, 40:427–509, 2014.
- [15] Ben Morris. Spectral gap for the zero range process with constant rate. *Ann. Probab.*, 34:1645–1664, 2006.
- [16] Sam Olesker-Taylor and Jonathan Hermon. Cutoff for almost all random walks on abelian groups. *preprint*, 2021.
- [17] Yann Ollivier. Ricci curvature of markov chains on metric spaces. *J. Funct. Anal.*, 256(3):810–864, 2009.
- [18] Yuval Peres. American institute of mathematics (aim) research workshop "sharp thresholds for mixing times", 2004.
- [19] Cyril Labbé Pietro Caputo and Hubert Lacoin. Spectral gap and cutoff phenomenon for the gibbs sampler of $\nabla\phi$ interfaces with convex potential. *to appear in Ann. Institut Henri Poincaré*, 2020.
- [20] Jonathan Hermon Riddhipratim Basu and Yuval Peres. Characterization of cutoff for reversible markov chains. *Annals of Probability*, 45(3):1448–1487, 2017.
- [21] Justin Salez. Cutoff for non-negatively curved markov chains. 2021.
- [22] Frank Spitzer. *Interaction of Markov processes*. 1970.