

Cutoff et pre-cutoff pour les temps de mélange de chaînes de Markov

Alexis Kagan

M1 mathématiques appliquées, Université Paris-Sud/ENSTA ParisTech

Dirigé par Olivier Hénard et Pascal Maillard

Août 2017

Sommaire

Introduction	2
1 Présentations des notions et théorèmes importants	5
1.1 Quelques propriétés intéressantes de la distance en variation totale	5
1.2 Distance de Hellinger : Quelques propriétés et liens avec la distance en variation totale	14
1.3 Quelques propriétés intéressantes de la distance de séparation	21
1.4 Phénomène de cutoff et de pre-cutoff	24
1.4.1 Cas d'une suite de chaînes simples	24
1.4.2 Cas d'une suite de chaînes sur un espace produit	26
2 Un exemple de suite ayant un pre-cutoff sans avoir de cutoff	33
2.1 Construction de l'exemple	33
2.2 Preuve du théorème 2.1.3	37
3 Quelques résultats positifs à propos de la condition produit	51
3.1 Valeurs propres et noyau de transition	51
3.2 Temps d'atteinte de grands ensembles et hit_α -cutoff	56
3.3 Un exemple de suite de chaînes Markov pour laquelle la condition produit est suffisante pour obtenir le cutoff	62
Bibliographie	66

Introduction

L'hypothèse simplificatrice d'absence de mémoire des chaînes de Markov fait d'elles une classe de processus stochastiques très riche, suffisamment riche pour susciter, plus de 100 ans après son apparition, l'intérêt des probabilistes.

L'étude asymptotique d'une chaîne de Markov constitue un sujet de recherche très actif au sein de la communauté probabiliste, notamment la convergence de la chaîne vers sa mesure de probabilité stationnaire, quand elle existe et est unique. Le temps de mélange $t_{\text{mix}}(\varepsilon)$ associé à une chaîne de Markov est le temps mis par la distance à l'équilibre de cette chaîne pour atteindre une certaine valeur $\varepsilon > 0$. $t_{\text{mix}}(\varepsilon)$ est alors le temps à attendre pour que cette chaîne soit "bien mélangée", suivant une certaine tolérance ε . Dans la suite, on notera $t_{\text{mix}} := t_{\text{mix}}(1/4)$.

Nous nous concentrerons sur des chaînes de Markov à espace d'état fini Ω , de noyau de transition P , irréductibles, réversibles, aperiodiques et de mesure stationnaire π . Autrement dit, P sera un opérateur auto-adjoint pour un produit scalaire particulier. Une étude spectrale de cet opérateur semble alors pouvoir nous fournir des informations importantes sur la chaîne de Markov. En effet, une valeur liée au spectre de P apparaît naturellement. Notons $\lambda := \max\{|\text{vp}| : \text{vp} \text{ valeur propre de } P \text{ différente de } 1\}$. $\lambda_* := 1 - \lambda$ est le trou spectral associé à P et $t_{\text{rel}} := \frac{1}{\lambda_*}$ est le temps de relaxation de cette chaîne. On a alors une borne pour la distance à l'équilibre de la chaîne dans un espace de Hilbert particulier : $\sum_{x \in \Omega} |P^t(x, y) - \pi(y)|^2 \pi(x) \leq \pi(y)(1 - \pi(y))e^{-2t/t_{\text{rel}}}$ (obtenue en choisissant $f = \mathbb{1}_y, y \in \Omega$ fixé, dans [5, relation (12.8) p. 163]), ce qui donne tout son sens à t_{rel} . Les notions de temps de mélange et de temps de relaxation sont centrales dans ce projet car elles sont étroitement liées aux phénomènes de cutoff et de pre-cutoff.

Une chaîne de Markov à espace d'état fini présente le phénomène de cutoff lorsque sa distance à l'équilibre d est proche de 1 jusqu'au temps de mélange puis chute brutalement vers 0 en une période négligeable devant le temps de mélange, appelée la fenêtre de cutoff (voir Figure 1). Le terme cutoff a été formulé pour la première fois par David Aldous et Persi Diaconis en 1986 [1].

Le phénomène de pre-cutoff est plus souple. Une chaîne de Markov à espace d'état fini présente le phénomène de pré-cutoff lorsque le temps t_{pc} mis par sa distance à l'équilibre d pour passer de 1 à

0 vérifie $t_{pc}/t_{mix} < \infty$ (voir Figure 2).

La "condition produit" (P) $t_{mix} \times \lambda_* = \infty$ c'est-à-dire $\frac{t_{mix}}{t_{rel}} = \infty$ est une condition nécessaire pour qu'une chaîne de Markov (finie, irréductible, réversible et apériodique) présente un cutoff. (P) est également connue pour être une condition nécessaire pour qu'une chaîne présente un pre-cutoff.

Yuval Peres fait dans un premier temps de (P) une condition suffisante pour qu'une chaîne "suffisamment régulière" présente un cutoff. Une hypothèse aussitôt réfutée, en 2004, par David Aldous grâce à un contre-exemple connu [5, p. 270].

Plus récemment, Hubert Lacoïn [4] prouve que (P) n'est pas une condition suffisante pour qu'une chaîne générale présente un cutoff. Sa démarche est la suivante : Sachant que (P) est nécessaire pour qu'une chaîne de Markov présente un pre-cutoff, si (P) était une condition suffisante pour que cette chaîne présente un cutoff, alors le phénomène de pre-cutoff impliquerait un cutoff. Hubert Lacoïn construit donc une chaîne sans cutoff mais présentant un pre-cutoff.

Le thème de ce projet est donc l'étude des phénomènes de cutoff et de pre-cutoff, avec pour guide la condition produit. Nous montrerons d'abord qu'elle n'est pas forcément suffisante puis nous regarderons quelques hypothèses supplémentaires à faire sur les chaînes de Markov pour que la condition produit soit suffisante. Ainsi la première partie de ce rapport présentera les différents outils nécessaires à la compréhension des notions de temps de mélange et des phénomènes de cutoff et pre-cutoff. Dans la deuxième partie, nous construirons un exemple de chaîne ayant un pre-cutoff mais ne présentant pas de cutoff fortement inspiré de celui de Lacoïn, montrant que la condition produit n'est pas toujours suffisante. La troisième et dernière partie a pour objectif de présenter brièvement quelques résultats positifs sur la condition produit, c'est-à-dire de voir à quelles conditions cette dernière est suffisante pour obtenir le cutoff.

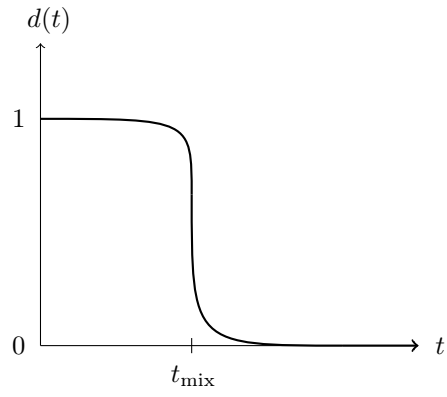


FIGURE 1 – Exemple de cutoff

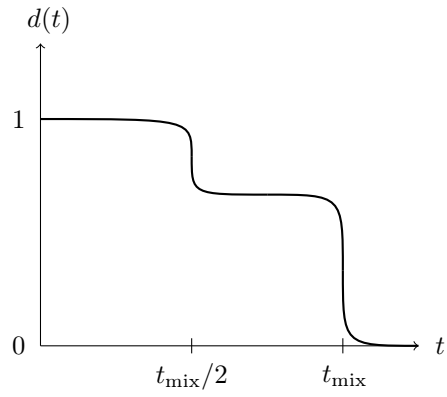


FIGURE 2 – Un exemple de pre-cutoff sans cutoff. On a $(t_{\text{mix}} - t_{\text{mix}}/2) / t_{\text{mix}} = 1/2 < \infty$

Partie 1

Présentations des notions et théorèmes importants

1.1 Quelques propriétés intéressantes de la distance en variation totale

Toutes les variables aléatoires considérées sont définies sur le triplet $(\mathcal{X}, \mathcal{F}, \mathbb{P})$.

La collection $\mathcal{P}(\Omega)$ désigne l'ensemble des parties de l'ensemble Ω fini.

Définition 1.1.1. *Pour tout ensemble Ω fini, on pose $\mathcal{M}_\Omega := \{\lambda : \lambda \text{ mesure de probabilité sur } (\Omega, \mathcal{P}(\Omega))\}$. On munit \mathcal{M}_Ω de la distance d_{TV} , appelée distance en variation totale définie par*

$$d_{TV}(\mu, \nu) := \max_{A \in \mathcal{P}(\Omega)} |\mu(A) - \nu(A)| \quad \forall (\mu, \nu) \in \mathcal{M}_\Omega \times \mathcal{M}_\Omega$$

Remarque 1.1.2. *Il est parfois plus commode de voir que $d_{TV}(\mu, \nu) = \max_{A \in \mathcal{P}(\Omega)} (\mu(A) - \nu(A))$.*

En effet si pour un $A \in \mathcal{P}(\Omega)$ on $d_{TV}(\mu, \nu) = |\mu(A) - \nu(A)|$, alors on a $d_{TV}(\mu, \nu) = |\mu(A^c) - \nu(A^c)|$ et comme $\mu(A) - \nu(A) = -(\mu(A^c) - \nu(A^c))$, le résultat suit.

Remarque 1.1.3. *En considérant l'événement $B := \{x \in \Omega : \mu(x) \geq \nu(x)\}$, on montre aisément que $d_{TV}(\mu, \nu) = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|$.*

Définition 1.1.4. *Soit $X = (X(t))_{t \in \mathbb{R}^+}$ une chaîne de Markov d'espace d'état Ω fini, irréductible, réversible et de mesure invariante π . Notons $(P_t)_{t \in \mathbb{R}^+}$ le semi-groupe de Markov associé à X . On définit la distance d , appelée distance à l'équilibre en variation totale pour la chaîne X par*

$$d(t) := \max_{x \in \Omega} \|P_t(x, \cdot) - \pi\|_{TV} \quad \forall t \in \mathbb{R}^+$$

Définition 1.1.5. Soit $X = (X(t))_{t \in \mathbb{R}^+}$ une chaîne de Markov d'espace d'état Ω fini, irréductible, réversible et de mesure invariante π . Notons $(P_t)_{t \in \mathbb{R}^+}$ le semi-groupe de Markov associé à X . On définit la distance \bar{d} par

$$\bar{d}(t) := \max_{x, y \in \Omega} \|P_t(x, \cdot) - P_t(y, \cdot)\|_{TV} \quad \forall t \in \mathbb{R}^+$$

Proposition 1.1.6. Soit $X = (X(t))_{t \in \mathbb{R}^+}$ une chaîne de Markov d'espace d'état Ω fini, irréductible, réversible et de mesure invariante π . Notons $(P_t)_{t \in \mathbb{R}^+}$ le semi-groupe de Markov associé à X . Notons $l_1^0(\pi) := \{f \in l_1(\pi) : \pi(f) = 0\}$. Soit P_t l'opérateur linéaire sur $l_1^0(\pi)$ tel que

$$P_t f : x \in \mathbb{R}^+ \longrightarrow P_t f(x) := \sum_{y \in \Omega} P_t(x, y) f(y)$$

On a

$$\bar{d}(t) = \|P_t\|_{\mathcal{L}(l_1^0(\pi))} := \max_{f \in l_1^0(\pi)} \frac{\|P_t f\|_{l_1(\pi)}}{\|f\|_{l_1(\pi)}}$$

où $\mathcal{L}(l_1^0(\pi))$ est l'ensemble des applications linéaires de $l_1^0(\pi)$ à valeurs dans $l_1^0(\pi)$.

Rappelons que dans le cas d'une chaîne de Markov irréductible, $\pi(x) > 0$ pour tout $x \in \Omega$. En effet, puisque pour tout $x, y \in \Omega$, $P_t(y, x) > 0$ alors par définition de π , $\pi(x) = \pi P_t(x) = \sum_{y \in \Omega} \pi(y) P_t(y, x) > 0$.

Soient $x, y \in \Omega$ fixés. On définit sur Ω l'application $f_{x,y}$ par

$$f_{x,y} := \frac{\mathbb{1}_x}{\pi(x)} - \frac{\mathbb{1}_y}{\pi(y)}$$

Lemme 1.1.7. Pour toute fonction $f \in l_1^0(\pi)$, il existe $\alpha_{x,y} \geq 0$ tels que

$$f = \sum_{x,y \in \Omega} \alpha_{x,y} f_{x,y}$$

Preuve. On se donne $f \in l_1^0(\pi)$.

Il paraît naturel de découper f en deux parties : d'un côté les valeurs positives, de l'autre les valeurs négatives

$$f = \sum_{x \in \Omega^+} \pi(x) f(x) \frac{\mathbb{1}_x}{\pi(x)} - \sum_{y \in \Omega^-} (-\pi(y) f(y)) \frac{\mathbb{1}_y}{\pi(y)}$$

avec $\sum_{x \in \Omega^+} \pi(x)f(x) = \sum_{y \in \Omega^-} -(\pi(y)f(y))$ et où $\Omega^+ := \{x : f(x) > 0\}$ et $\Omega^- := \{y : f(y) < 0\}$.

Nous allons donc nous intéresser aux deux ensembles de points $(\pi f)^+ := \{\pi(x_i)f(x_i), x_i \in \Omega^+, x_1 \leq x_2 \leq \dots\}$ et $\{\pi(y_i)f(y_i), y_i \in \Omega^-, y_1 \leq y_2 \leq \dots\}$.

Pour obtenir la combinaison linéaire à coefficients positifs attendue, une construction possible est la suivante :

— Si la première valeur de f est strictement positive, alors on prend $\pi(x_1)f(x_1)$.

— Si $\pi(x_1)f(x_1) \leq -\pi(y_1)f(y_1)$, alors on écrit :

$$f = \pi(x_1)f(x_1)\left(\frac{\mathbb{1}_{x_1}}{\pi(x_1)} - \frac{\mathbb{1}_{y_1}}{\pi(y_1)}\right) + (\pi(y_1)f(y_1) + \pi(x_1)f(x_1))\frac{\mathbb{1}_{y_1}}{\pi(y_1)} + \sum_{x \in \Omega^+ \setminus \{x_1\}} \pi(x)f(x)\frac{\mathbb{1}_x}{\pi(x)} - \sum_{y \in \Omega^- \setminus \{y_1\}} (-\pi(y)f(y))\frac{\mathbb{1}_y}{\pi(y)}$$

c'est-à-dire

$$f = \pi(x_1)f(x_1)f_{x_1, y_1} + (\pi(y_1)f(y_1) + \pi(x_1)f(x_1))\frac{\mathbb{1}_{y_1}}{\pi(y_1)} + \sum_{x \in \Omega^+ \setminus \{x_1\}} \pi(x)f(x)\frac{\mathbb{1}_x}{\pi(x)} - \sum_{y \in \Omega^- \setminus \{y_1\}} (-\pi(y)f(y))\frac{\mathbb{1}_y}{\pi(y)}$$

Puis on prend $\pi(y_1)f(y_1) + \pi(x_1)f(x_1)$ et on compare cette valeur à $\pi(x_2)f(x_2)$ et ainsi de suite. Puisque $\sum_{x \in \Omega^+} \pi(x)f(x) = \sum_{y \in \Omega^-} -(\pi(y)f(y))$, on obtiendra bien une combinaison linéaire des $f_{x,y}$ avec des coefficients $\alpha_{x,y}$ qui seront soit strictement positifs, soit nuls par construction.

— Si $\pi(x_1)f(x_1) \geq -\pi(y_1)f(y_1)$, alors on écrit :

$$f = \pi(y_1)f(y_1)\left(\frac{\mathbb{1}_{y_1}}{\pi(y_1)} - \frac{\mathbb{1}_{x_1}}{\pi(x_1)}\right) + (\pi(y_1)f(y_1) + \pi(x_1)f(x_1))\frac{\mathbb{1}_{x_1}}{\pi(x_1)} + \sum_{x \in \Omega^+ \setminus \{x_1\}} \pi(x)f(x)\frac{\mathbb{1}_x}{\pi(x)} - \sum_{y \in \Omega^- \setminus \{y_1\}} (-\pi(y)f(y))\frac{\mathbb{1}_y}{\pi(y)}$$

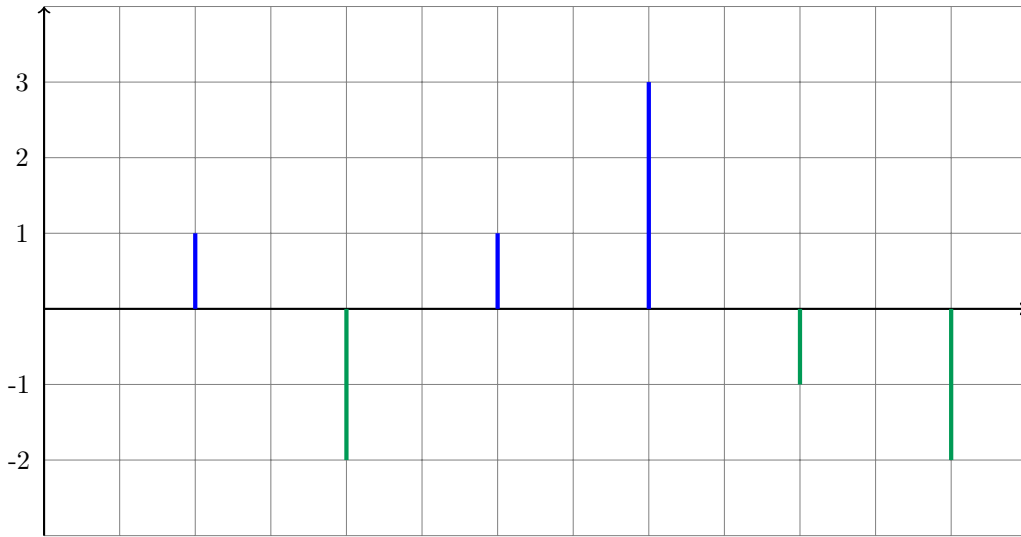
c'est-à-dire

$$f = \pi(y_1)f(y_1)f_{y_1, x_1} + (\pi(y_1)f(y_1) + \pi(x_1)f(x_1))\frac{\mathbb{1}_{x_1}}{\pi(x_1)} + \sum_{x \in \Omega^+ \setminus \{x_1\}} \pi(x)f(x)\frac{\mathbb{1}_x}{\pi(x)} - \sum_{y \in \Omega^- \setminus \{y_1\}} (-\pi(y)f(y))\frac{\mathbb{1}_y}{\pi(y)}$$

Puis on prend $\pi(y_1)f(y_1) + \pi(x_1)f(x_1)$ et on compare cette valeur à $\pi(y_2)f(y_2)$ et ainsi de suite. Puisque $\sum_{x \in \Omega^+} \pi(x)f(x) = \sum_{y \in \Omega^-} -(\pi(y)f(y))$, on obtiendra bien une combinaison linéaire des $f_{x,y}$ avec des coefficients $\alpha_{x,y}$ qui seront soit strictement positifs, soit nuls par construction.

- Si la première valeur de f est strictement négative, alors on prend $\pi(y_1)f(y_1)$ et le raisonnement est analogue.

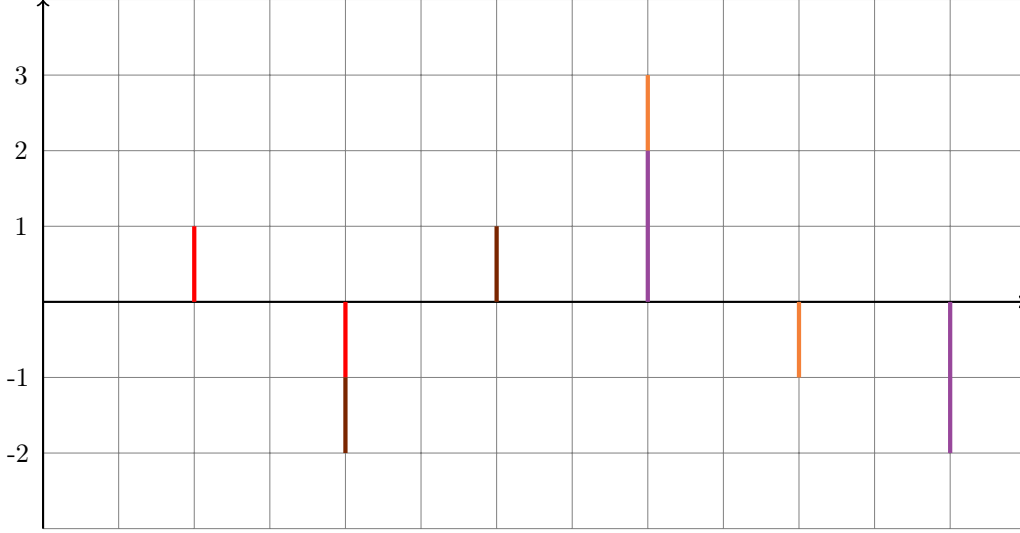
Voici un exemple de cette construction pour $\Omega = \{1, 2, 3, 4, 5, 6\}$. La première valeur de f est strictement positive.



Ci-dessus le graphe de l'application qui à x associe $\pi(x)f(x)$ sur Ω . Ici, $x_1 = 1, x_2 = 3, x_3 = 4$ et $y_1 = 2, y_2 = 5, y_3 = 6$.

On a

$$f = \frac{\mathbb{1}_1}{\pi(1)} + \frac{\mathbb{1}_3}{\pi(3)} + 3 \frac{\mathbb{1}_4}{\pi(4)} - 2 \frac{\mathbb{1}_2}{\pi(2)} - \frac{\mathbb{1}_5}{\pi(5)} - 2 \frac{\mathbb{1}_6}{\pi(6)}$$



Le graphe ci-dessus illustre la construction expliquée précédemment. On regroupe les bâtons ayant la même couleur et on obtient finalement

$$f = f_{1,2} + f_{2,3} + f_{4,5} + 2f_{4,6}$$

Remarque 1.1.8. On voit que la manière de construire la combinaison linéaire n'est pas unique. En effet, puisque la moyenne de f par rapport à la mesure π est nulle, c'est-à-dire que $\sum_{x \in \Omega^+} \pi(x)f(x) = \sum_{y \in \Omega^-} -(\pi(y)f(y))$, on peut découper les bâtons d'une manière quelconque tant qu'on associe une valeur de $(\pi f)^+$ à une valeur de $(\pi f)^-$.

Preuve de la proposition 1.1.6. Dans un premier temps, montrons qu'il existe $f \in l_1^0(\pi)$ telle que

$$\bar{d}(t) = \frac{\|P_t f\|_{l_1(\pi)}}{\|f\|_{l_1(\pi)}}$$

Prenons les $x, y \in \Omega$ tels que

$$\bar{d}(t) = \|P_t(x, \cdot) - P_t(y, \cdot)\|_{TV} \quad \forall t \in \mathbb{R}^+$$

Soit $f_{x,y}$ la fonction définie dans le lemme 1.1.7. Il est évident que $f_{x,y} \in l_1^0(\pi)$. De plus, on a

$$\|f_{x,y}\|_{l_1(\pi)} = \sum_{z \in \Omega} \left| \frac{\mathbf{1}_x(z)}{\pi(x)} - \frac{\mathbf{1}_y(z)}{\pi(y)} \right| \pi(z) = 2$$

Enfin

$$\|P_t f_{x,y}\|_{l_1(\pi)} = \sum_{z \in \Omega} \left| \frac{\pi(z)P_t(z,x)}{\pi(x)} - \frac{\pi(z)P_t(z,y)}{\pi(y)} \right|$$

Mais par réversibilité de X , nous avons

$$\frac{\pi(z)P_t(z,x)}{\pi(x)} = P_t(x,z) \quad \text{et} \quad \frac{\pi(z)P_t(z,y)}{\pi(y)} = P_t(y,z)$$

soit finalement

$$\|P_t f_{x,y}\|_{l_1(\pi)} = 2\|P_t(x,\cdot) - P_t(y,\cdot)\|_{TV}$$

d'où

$$\frac{\|P_t f_{x,y}\|_{l_1(\pi)}}{\|f_{x,y}\|_{l_1(\pi)}} = \|P_t(x,\cdot) - P_t(y,\cdot)\|_{TV} = \bar{d}(t)$$

Pour terminer, il faut montrer que

$$\forall f \in l_1^0(\pi), \quad \bar{d}(t) \geq \frac{\|P_t f\|_{l_1(\pi)}}{\|f\|_{l_1(\pi)}} \quad \forall t \in \mathbb{R}^+$$

Pour cela, on se donne $f \in l_1^0(\pi)$. D'après le lemme 1.1.7, il existe $\alpha_{x,y} \geq 0$ tels que

$$f = \sum_{x,y \in \Omega} \alpha_{x,y} f_{x,y}$$

On a alors

$$\|f\|_{l_1(\pi)} = 2 \sum_{x,y \in \Omega} \alpha_{x,y}$$

On a également par linéarité de l'opérateur P_t et par inégalité triangulaire

$$\|P_t f\|_{l_1(\pi)} \leq \sum_{x,y \in \Omega} \alpha_{x,y} \|P_t f_{x,y}\|_{l_1(\pi)} = 2 \sum_{x,y \in \Omega} \alpha_{x,y} \|P_t(x,\cdot) - P_t(y,\cdot)\|_{TV}$$

ou encore

$$\|P_t f\|_{l_1(\pi)} \leq 2 \sum_{x,y \in \Omega} \alpha_{x,y} \bar{d}(t) \quad \forall t \in \mathbb{R}^+$$

d'où

$$\frac{\|P_t f\|_{l_1(\pi)}}{\|f\|_{l_1(\pi)}} \leq \bar{d}(t) \quad \forall t \in \mathbb{R}^+$$

Ce qui achève la preuve

Remarque sur la preuve de la proposition 1.1.6.

Notons $\mathcal{M}_s := \{\lambda : \lambda \text{ mesure signée sur } (\Omega, \mathcal{P}(\Omega)) \text{ telle que } \lambda(\Omega) = 0 \text{ et } |\lambda|(\Omega) \leq 1\}$.

Une autre façon, finalement assez liée à la preuve ci-dessus, de démontrer la proposition 1.1.6 est d'utiliser un argument de dualité entre $l_1^0(\pi)$ et \mathcal{M}_s .

En effet, en remarquant que $\bar{d}(t) = \sup_{\mu, \nu \in \mathcal{M}_\Omega} \|\mu P_t - \nu P_t\|_{TV}$ [5, 4.4 Standardizing Distance from Stationarity], on a

$$\bar{d}(t) = \sup_{\mu, \nu \in \mathcal{M}_\Omega} \left\| \frac{(\mu - \nu)}{2} P_t \right\|_{l_1} = \sup_{\lambda \in \mathcal{M}_s} \|\lambda P_t\|_{l_1}$$

La deuxième égalité est vraie car si $(A_n)_{n \in \mathbb{N}}$ est une famille de parties de Ω quelconque, $\sum_{n \in \mathbb{N}} |(\frac{\mu - \nu}{2})(A_n)| \leq$

$1 < \infty$ et $|(\frac{\mu - \nu}{2})|(\Omega) = \sup \left\{ \sum_{n \in \mathbb{N}} |(\frac{\mu - \nu}{2})(A_n)| : (A_n)_{n \in \mathbb{N}} \text{ partition de } \Omega \right\} \leq 1$ (car μ et ν sont des mesures de probabilité).

C'est-à-dire par réversibilité de la chaîne

$$\bar{d}(t) = \sup_{\lambda \in \mathcal{M}_s} \left\| P_t \frac{\lambda}{\pi} \right\|_{l_1(\pi)}$$

d'où en posant $f := \frac{\lambda}{\pi}$ (vue à présent comme une application de Ω à valeurs dans \mathbb{R})

$$\bar{d}(t) = \sup_{f \in l_1^0(\pi); \|f\|_{l_1(\pi)} \leq 1} \|P_t f\|_{l_1(\pi)}$$

et puisque pour toute fonction $f \in l_1^0(\pi)$, $f/\|f\|_{l_1(\pi)}$ est de norme $l_1(\pi)$ égale à 1, par linéarité de P_t

$$\bar{d}(t) = \sup_{f \in l_1^0(\pi); \|f\|_{l_1(\pi)} = 1} \|P_t f\|_{l_1(\pi)} = \sup_{f \in l_1^0(\pi)} \frac{\|P_t f\|_{l_1(\pi)}}{\|f\|_{l_1(\pi)}}$$

Proposition 1.1.9. \bar{d} et d sont liées par la relation suivante :

$$d(t) \leq \bar{d}(t) \leq 2d(t) \tag{1.1}$$

et \bar{d} est sous-multiplicative :

$$\bar{d}(t+u) \leq \bar{d}(t)\bar{d}(u) \tag{1.2}$$

On déduit de (1.2) par une récurrence immédiate que pour tout entier $m \geq 1$ et pour tout $t \geq 0$, $\bar{d}(mt) \leq \bar{d}(t)^m$ d'où (en combinant avec (1.1))

$$d(mt) \leq 2^m d(t)^m \quad (1.3)$$

Preuve. Commençons par (1.1). Pour l'inégalité de gauche, on écrit, pour $A \in \mathcal{P}(\Omega)$ quelconque et pour $x \in \Omega$ fixé

$$|P_t(x, A) - \pi(A)| = \left| \sum_{y \in \Omega} \pi(y) P_t(x, A) - \sum_{y \in \Omega} \pi(y) P_t(y, A) \right| \leq \bar{d}(t) \quad \forall t \in \mathbb{R}^+$$

car π est la mesure invariante de la chaîne et par inégalité triangulaire.

D'où

$$\|P_t(x, \cdot) - \pi\|_{TV} \leq \bar{d}(t) \quad \forall t \in \mathbb{R}^+$$

L'inégalité de droite découle directement de l'inégalité triangulaire. En effet, pour tout $x, y \in \Omega$

$$\|P_t(x, \cdot) - P_t(y, \cdot)\|_{TV} = \|P_t(x, \cdot) - \pi + \pi - P_t(y, \cdot)\|_{TV} \quad \forall t \in \mathbb{R}^+$$

Le résultat est obtenu en prenant dans un premier temps le max sur x puis le max sur y .

Démontrons à présent (1.2). Soit $s \in \mathbb{R}^+$. D'après [5, Proposition 4.7, p. 50], on a

$$\|P_s(x, \cdot) - P_s(y, \cdot)\|_{TV} = \min_{(X, Y) \in \mathcal{C}} \mathbb{P}(X \neq Y) \quad (1.4)$$

où \mathcal{C} est l'ensemble des couplages de $P_s(x, \cdot)$ et $P_s(y, \cdot)$.

Notons (X_s, Y_s) le couplage réalisant le minimum dans (1.4). X_s et Y_s sont donc deux variables aléatoires de \mathcal{X} à valeurs dans Ω . Remarquons que l'application qui à $\omega \in \Omega$ associe $P_t(\omega, A)$ est $\mathcal{P}(\Omega)$ -mesurable, $A \in \mathcal{P}(\Omega)$.

Alors par théorème de transfert et grâce à la propriété de semi-groupe de $(P_t)_{t \in \mathbb{R}^+}$, on a

$$\mathbb{E}(P_t(X_s, A)) = \int_{\Omega} P_s(x, d\omega) P_t(\omega, A) = P_{t+s}(x, A) \quad \text{et} \quad \mathbb{E}(P_t(Y_s, A)) = \int_{\Omega} P_s(y, d\omega) P_t(\omega, A) = P_{t+s}(y, A)$$

On peut alors voir que pour tout $A \in \mathcal{P}(\Omega)$

$$|P_{t+s}(x, A) - P_{t+s}(y, A)| \leq \mathbb{E}[|P_t(X_s, A) - P_t(Y_s, A)|]$$

Mais

$$|P_t(X_s, A) - P_t(Y_s, A)| = |P_t(X_s, A) - P_t(Y_s, A)|\mathbb{1}_{\{X_s \neq Y_s\}} \leq \bar{d}(t)\mathbb{1}_{\{X_s \neq Y_s\}}$$

d'où selon ce qui précède

$$|P_{t+s}(x, A) - P_{t+s}(y, A)| \leq \bar{d}(t)\bar{d}(s)$$

ce qui achève la démonstration.

1.2 Distance de Hellinger : Quelques propriétés et liens avec la distance en variation totale

Définition 1.2.1. Pour tout ensemble Ω fini, on munit \mathcal{M}_Ω de la distance d^H , appelée distance de Hellinger définie par

$$d^H(\mu, \nu) := \sqrt{\sum_{x \in \Omega} \left(\sqrt{\mu(x)} - \sqrt{\nu(x)} \right)^2} \quad \forall (\mu, \nu) \in \mathcal{M}_\Omega \times \mathcal{M}_\Omega$$

On définit également l'affinité de Hellinger I par

$$I(\mu, \nu) := \sum_{x \in \Omega} \sqrt{\mu(x)\nu(x)} \quad \forall (\mu, \nu) \in \mathcal{M}_\Omega \times \mathcal{M}_\Omega$$

I est évidemment liée à d^H et sera très utile pour la suite.

Proposition 1.2.2. $\forall (\mu, \nu) \in \mathcal{M}_\Omega \times \mathcal{M}_\Omega$

$$d_{TV}(\mu, \nu) \leq d^H(\mu, \nu) \leq \sqrt{2d_{TV}(\mu, \nu)}$$

Preuve. Pour la première inégalité, on écrit

$$d_{TV}(\mu, \nu)^2 = \frac{1}{4} \left(\sum_{x \in \Omega} |\mu(x) - \nu(x)| \right)^2 = \frac{1}{4} \left(\sum_{x \in \Omega} |\sqrt{\mu(x)} - \sqrt{\nu(x)}| (\sqrt{\mu(x)} + \sqrt{\nu(x)}) \right)^2$$

donc par inégalité de Cauchy-Schwarz

$$\frac{1}{4} \left(\sum_{x \in \Omega} |\sqrt{\mu(x)} - \sqrt{\nu(x)}| (\sqrt{\mu(x)} + \sqrt{\nu(x)}) \right)^2 \leq \frac{1}{4} d^H(\mu, \nu)^2 \sum_{x \in \Omega} (\sqrt{\mu(x)} + \sqrt{\nu(x)})^2$$

mais on a

$$\sum_{x \in \Omega} (\sqrt{\mu(x)} + \sqrt{\nu(x)})^2 \leq 2 + 2 \sum_{x \in \Omega} \sqrt{\mu(x)\nu(x)} \leq 2 + 2 \sqrt{\sum_{x \in \Omega} \mu(x)} \sqrt{\sum_{x \in \Omega} \nu(x)} = 4$$

où la dernière inégalité provient de l'inégalité de Cauchy-Schwarz.

Pour la deuxième inégalité, on a évidemment $(\mu(x) \wedge \nu(x))^2 \leq \mu(x)\nu(x)$ donc

$$I(\mu, \nu) \geq \sum_{x \in \Omega} \mu(x) \wedge \nu(x)$$

On pose $B := \{x \in \Omega : \mu(x) \geq \nu(x)\}$. Remarquons que

$$d_{TV}(\mu, \nu) = -\frac{1}{2} \sum_{x \in \Omega} ((\nu \mathbf{1}_B)(x) + (\mu \mathbf{1}_{B^c})(x)) + \frac{1}{2} \sum_{x \in \Omega} (1 - (\nu \mathbf{1}_B)(x) + 1 - (\mu \mathbf{1}_{B^c})(x))$$

c'est-à-dire

$$d_{TV}(\mu, \nu) = 1 - \sum_{x \in \Omega} \mu(x) \wedge \nu(x)$$

Finalement

$$I(\mu, \nu) \geq 1 - d_{TV}(\mu, \nu)$$

En remarquant que

$$d^H(\mu, \nu)^2 = 2 - 2I(\mu, \nu)$$

on obtient

$$2d_{TV}(\mu, \nu) \geq d^H(\mu, \nu)^2$$

ce qui est le résultat attendu.

Définition 1.2.3. Soit $X = (X(t))_{t \in \mathbb{R}^+}$ une chaîne de Markov d'espace d'état Ω fini, irréductible, réversible et de mesure invariante π . Notons $(P_t)_{t \in \mathbb{R}^+}$ le semi-groupe de Markov associé à X . On définit la distance d^H par

$$d^H(t) := \sup_{x \in \Omega} d^H(P_t(x, \cdot), \pi) \quad \forall t \in \mathbb{R}^+$$

Proposition 1.2.4. (Hubert Lacoïn, 2015)

Pour tout $t \in \mathbb{R}^+$

$$d^H(2t) \leq 18(d^H(t))^{5/4}$$

(voir la remarque 1.4.19 pour comprendre pourquoi l'exposant $5/4$ est important).

Posons pour tout $t \in \mathbb{R}^+$

$$u := d^H(t) \text{ et } d^H(2t) := d^H(P_{2t}(x, \cdot), \pi) = \sup_{z \in \Omega} d^H(P_t(z, \cdot), \pi)$$

Puisque $P_t(x, \cdot) \ll \pi$ et $P_{2t}(x, \cdot) \ll \pi$, si on pose

$$g := \frac{dP_t(x, \cdot)}{d\pi} \text{ et } g' := \frac{dP_{2t}(x, \cdot)}{d\pi}$$

alors on a le lemme suivant :

Lemme 1.2.5.

$$\int_{\Omega} \left(\sqrt{g'(y)} - 1 \right)^2 \pi(dy) = d^{\mathcal{H}}(2t)^2 \text{ et } \int_{\Omega} \left(\sqrt{g(y)} - 1 \right)^2 \pi(dy) \leq u^2$$

Preuve. Commençons par l'égalité. Comme $\forall y \in \Omega, \pi(y) > 0$, on a

$$d^{\mathcal{H}}(2t)^2 = \sum_{y \in \Omega} \left(\sqrt{\frac{P_{2t}(x, y)}{\pi(y)}} - 1 \right)^2 \pi(y)$$

mais par définition, $P_{2t}(x, y) = g'(y)\pi(y)$, d'où

$$d^{\mathcal{H}}(2t)^2 = \sum_{y \in \Omega} \left(\sqrt{g'(y)} - 1 \right)^2 \pi(y)$$

c'est-à-dire

$$d^{\mathcal{H}}(2t) = \int_{\Omega} \left(\sqrt{g'(y)} - 1 \right)^2 \pi(dy)$$

Pour l'inégalité, on obtient de manière analogue

$$d^H(P_t(x, \cdot), \pi)^2 = \int_{\Omega} \left(\sqrt{g(y)} - 1 \right)^2 \pi(dy)$$

et comme $\forall x \in \Omega, u^2 = d^{\mathcal{H}}(t)^2 \geq d^H(P_t(x, \cdot), \pi)^2$, on obtient le résultat attendu.

Preuve de la proposition 1.2.4 Nous séparons la preuve en deux cas.

Voici d'abord deux relation intéressantes.

En combinant les propositions 1.2.2 et 1.1.9 (1.1), on obtient

$$\frac{\bar{d}(t)}{2} \leq d^{\mathcal{H}}(t) \leq \sqrt{2\bar{d}(t)} \tag{1.5}$$

et donc en utilisant la sous-multiplicativité de \bar{d} (proposition 1.1.9 (1.2))

$$\bar{d}(2t) \leq \bar{d}(t)^2 \leq 4d^{\mathcal{H}}(t)^2 \tag{1.6}$$

Premier cas : $u \geq 1$. C'est évident. Par définition de la distance de Hellinger,

$$d^H(\mu, \nu) \leq \sqrt{2} \quad \forall (\mu, \nu) \in \mathcal{M}_{\Omega} \times \mathcal{M}_{\Omega}$$

Deuxième cas : $u \leq 1$

Soit $\delta \in (0, 1)$. Montrons dans un premier temps que

$$\int_{\Omega} |g(y) - 1| \mathbf{1}_{\{|g(y)-1| \geq u^\delta\}} \pi(dy) \leq 12u^{2-\delta} \quad (1.7)$$

Remarquons que par décroissance sur $[2, +\infty)$ de l'application qui à x associe $\left| \frac{\sqrt{x+1}}{\sqrt{x-1}} \right|$, si $g \geq 2$ alors

$$|g - 1| \leq (3 + 2\sqrt{2}) |\sqrt{g} - 1|^2$$

alors que si $g \in (0, 2)$

$$|g - 1|^2 \leq (\sqrt{2} + 1)^2 |\sqrt{g} - 1|^2$$

et comme $|g - 1| \mathbf{1}_{\{|g(y)-1| \geq u^\delta\}} \leq u^{-\delta} |g - 1|^2 \mathbf{1}_{\{|g(y)-1| \geq u^\delta\}}$, on obtient finalement

$$\int_{\Omega} |g(y) - 1| \mathbf{1}_{\{|g(y)-1| \geq u^\delta\}} \pi(dy) \leq u^{-\delta} (3 + 2\sqrt{2} + (\sqrt{2} + 1)^2) \int_{\Omega} |\sqrt{g(y)} - 1|^2 \mathbf{1}_{\{|g(y)-1| \geq u^\delta\}} \pi(dy)$$

c'est-à-dire

$$\int_{\Omega} |g(y) - 1| \mathbf{1}_{\{|g(y)-1| \geq u^\delta\}} \pi(dy) \leq 12u^{2-\delta}$$

d'après le lemme 1.2.5.

Décomposons maintenant $g - 1$ de la manière suivante

$$h_1 := (g - 1) \mathbf{1}_{\{|g(y)-1| < u^\delta\}} \quad \text{et} \quad h_2 := (g - 1) \mathbf{1}_{\{|g(y)-1| \geq u^\delta\}}$$

puis afin de centrer h_2 , posons

$$\bar{h}_1 := (g - 1) \mathbf{1}_{\{|g(y)-1| < u^\delta\}} + \pi(h_2) \quad \text{et} \quad \bar{h}_2 := (g - 1) \mathbf{1}_{\{|g(y)-1| \geq u^\delta\}} - \pi(h_2)$$

et on a bien

$$g - 1 = \bar{h}_1 + \bar{h}_2$$

Décomposons maintenant $g' - 1$ de la manière suivante

$$\bar{h}'_1 := P_t \bar{h}_1 \quad \text{et} \quad \bar{h}'_2 := P_t \bar{h}_2$$

ce qui nous permet d'écrire que

$$g' - 1 = \bar{h}'_1 + \bar{h}'_2$$

En effet, pour tout $z \in \Omega$

$$P_t(g-1)(z) = \sum_{y \in \Omega} P_t(z, y) \frac{P_t(x, y)}{\pi(y)} - 1$$

mais $\frac{P_t(x, y)}{\pi(y)} = \frac{P_t(y, x)}{\pi(x)}$ par réversibilité de la chaîne, d'où

$$P_t(g-1)(z) = \frac{1}{\pi(x)} \sum_{y \in \Omega} P_t(z, y) P_t(y, x) - 1 = \frac{P_{2t}(z, x)}{\pi(x)} - 1$$

et en utilisant de nouveau la réversibilité, on obtient

$$P_t(g-1)(z) = g' - 1$$

On conclut en utilisant la linéarité de P_t .

On a, d'après l'inégalité triangulaire et (1.7)

$$\|\bar{h}_2\|_{l_1(\pi)} \leq 2\|h_2\|_{l_1(\pi)} \leq 24u^{2-\delta}$$

donc en utilisant la proposition 1.1.6 puis (1.5)

$$\|\bar{h}'_2\|_{l_1(\pi)} \leq \bar{d}(t)\|\bar{h}_2\|_{l_1(\pi)} \leq 48u^{3-\delta}$$

On a évidemment d'après (1.7)

$$\|\bar{h}_1\|_\infty \leq \|h_1\|_\infty + \|h_2\|_{l_1(\pi)} \leq u^\delta + 12u^{2-\delta} \leq 13u^\delta$$

car $u < 1$ et $\delta < 1$.

d'où

$$\|\bar{h}'_1\|_\infty \leq \|\bar{h}_1\|_\infty \leq 13u^\delta \tag{1.8}$$

Enfin, remarquons que

$$\|g' - 1\|_{l_1(\pi)} = 2\|P_{2t}(x, \cdot) - \pi\|_{TV}$$

ce qui permet d'écrire que

$$\|g' - 1\|_{l_1(\pi)} \leq 2d(2t) \leq 2\bar{d}(2t) \leq 8u^2 \tag{1.9}$$

où la première inégalité découle directement de la définition de $d(2t)$, la deuxième vient de la proposition 1.1.9 (1.1) et la dernière est obtenue grâce à (1.6).

Nous sommes maintenant prêt à borner $d^{\mathcal{H}}(2t)^2$. De nouveau, nous découpons les intégrales en deux parties.

Remarquons d'abord que

$$(\sqrt{g'} - 1)^2 \leq |g' - 1|^2 \quad (1.10)$$

et puisque $u \geq 0$, on a

$$|g' - 1|^2 \mathbb{1}_{\{|g(y)-1| < 26u^\delta\}} \leq 26u^\delta |g' - 1| \quad (1.11)$$

d'où

$$\int_{\Omega} (\sqrt{g'} - 1)^2(y) \mathbb{1}_{\{|g(y)-1| < 26u^\delta\}}(y) \pi(dy) \leq 26u^\delta \|g' - 1\|_{l_1(\pi)} \leq 208u^{2+\delta} \quad (1.12)$$

d'après (1.9), (1.10) et (1.11).

Pour borner le deuxième membre, remarquons que

$$(\sqrt{g'} - 1)^2 \leq |g' - 1| \quad (1.13)$$

que

$$\mathbb{1}_{\{|g(y)-1| \geq 26u^\delta\}} \leq \mathbb{1}_{\{|\bar{h}'_2| \geq 13u^\delta\}} \quad (1.14)$$

et enfin que

$$|g' - 1| \mathbb{1}_{\{|\bar{h}'_2| \geq 13u^\delta\}} \leq (|\bar{h}'_2| + \|\bar{h}'_1\|_{\infty}) \mathbb{1}_{\{|\bar{h}'_2| \geq 13u^\delta\}} \leq 2|\bar{h}'_2| \quad (1.15)$$

d'après (1.8).

Ainsi

$$\int_{\Omega} (\sqrt{g'} - 1)^2(y) \mathbb{1}_{\{|g(y)-1| \geq 26u^\delta\}}(y) \pi(dy) \leq \int_{\Omega} 2|\bar{h}'_2(y)| \mathbb{1}_{\{|\bar{h}'_2| \geq 13u^\delta\}}(y) \pi(dy) \leq 2\|\bar{h}'_2\|_{l_1(\pi)}$$

et d'après (1.13), (1.14) et (1.15), on obtient

$$\int_{\Omega} (\sqrt{g'} - 1)^2(y) \mathbb{1}_{\{|g(y)-1| \geq 26u^\delta\}}(y) \pi(dy) \leq 96u^{3-\delta} \quad (1.16)$$

Finalement d'après (1.12) et (1.16)

$$d^{\mathcal{H}}(2t)^2 = \int_{\Omega} (\sqrt{g'} - 1)^2(y) \pi(dy) \leq 96u^{3-\delta} + 208u^{2+\delta}$$

Puisque $u \leq 1$ et $\delta < 1$, si nous voulons maximiser l'exposant, nous devons choisir $\delta = \frac{1}{2}$.
Donc

$$d^{\mathcal{H}}(2t) \leq \sqrt{304}(d^{\mathcal{H}}(t))^{5/4} \leq 18(d^{\mathcal{H}}(t))^{5/4}$$

Ce qui achève la preuve.

1.3 Quelques propriétés intéressantes de la distance de séparation

Définition 1.3.1. Soit $X = (X(t))_{t \in \mathbb{R}^+}$ une chaîne de Markov d'espace d'état Ω fini, irréductible, réversible et de mesure invariante π . Notons $(P_t)_{t \in \mathbb{R}^+}$ le semi-groupe de Markov associé à X . On définit la distance d^s , appelée distance de séparation pour la chaîne X par

$$d^s(t) := \max_{x, y \in \Omega} \left(1 - \frac{P_t(x, y)}{\pi(y)} \right) = 1 - \min_{x, y \in \Omega} \frac{P_t(x, y)}{\pi(y)}$$

Proposition 1.3.2. La distance de séparation est sous-multiplicative, c'est-à-dire que pour tout $t, u \in \mathbb{R}^+$

$$d^s(t + u) \leq d^s(t)d^s(u)$$

Remarque 1.3.3. Pour tout $t \in \mathbb{R}^+$, $d^s(t) \geq 0$. L'égalité a lieu s'il existe $t \in \mathbb{R}^+$ tel que pour tout $x \in \Omega$, $P_t(x, \cdot) = \pi$. Nous allons exclure ce cas.

Lemme 1.3.4. Soit $x \in \Omega$ fixé. Pour tout $t \in \mathbb{R}^+$, il existe une mesure de probabilité $\tilde{P}_t(x, \cdot)$ sur $(\Omega, \mathcal{P}(\Omega))$ telle que

$$P_t(x, \cdot) = (1 - d^s(t))\pi + d^s(t)\tilde{P}_t(x, \cdot)$$

et telle que $\pi = \pi \tilde{P}_t$ ($(\tilde{P}_t)_t$ n'est pas un semi-groupe).

Remarque 1.3.5. Cette relation a une interprétation probabiliste. En effet, pour tout $t \in \mathbb{R}^+$, $\epsilon(t)$ est une variable aléatoire sur \mathcal{X} suivant une loi de Bernoulli de paramètre $d^s(t)$. Pour tout $t \in \mathbb{R}^+$, les $\epsilon(t)$ sont indépendantes.

Considérons pour tout $t \in \mathbb{R}^+$ le couplage $(\tilde{X}(t), \tilde{Y}(t))$ obéissant aux règles suivantes :

- $\tilde{X} = (\tilde{X}(t))_{t \in \mathbb{R}^+}$ et $\tilde{Y} = (\tilde{Y}(t))_{t \in \mathbb{R}^+}$ sont deux suites de variables aléatoires. Pour tout t , $\tilde{X}(t)$ et $\tilde{Y}(t)$ sont définies sur le même espace \mathcal{X} et à valeurs dans Ω .
- Partant d'un état $x \in \Omega$, la loi de la variable aléatoire $\tilde{X}(t)$ est la mesure $\mathbb{P}_x(\tilde{X}(t) \in \cdot)$ et la loi de la variable aléatoire $\tilde{Y}(t)$ est la mesure $\mathbb{P}_x(\tilde{Y}(t) \in \cdot)$.
- A un temps t quelconque, si $\epsilon(t) = 0$, \tilde{X} saute vers un état $y \in \Omega$ avec probabilité $\pi(y)$.
- A un temps t quelconque, si $\epsilon(t) = 1$, \tilde{X} saute vers un état $y \in \Omega$ avec probabilité $\mathbb{P}_x(\tilde{Y}(t) = y)$.

Autrement dit, pour tout $y \in \Omega$ et pour tout $t \in \mathbb{R}^+$

$$\mathbb{P}_x(\tilde{X}(t) = y) = (1 - d^s(t))\pi(y) + d^s(t)\mathbb{P}_x(\tilde{Y}(t) = y)$$

c'est-à-dire en posant $\mathbb{P}_x(\tilde{Y}(t) \in \cdot) := \tilde{P}_t(x, \cdot)$

$$\mathbb{P}_x(\tilde{X}(t) = y) = (1 - d^s(t))\pi(y) + d^s(t)\tilde{P}_t(x, y)$$

Le but du lemme 1.3.4 est finalement de trouver, pour tout t , une mesure $\tilde{P}_t(x, \cdot)$ (qui est en réalité la loi de $\tilde{Y}(t)$) de sorte que $(\tilde{X}(t), \tilde{Y}(t))$ soit un couplage de $P_t(x, \cdot)$ et $\tilde{P}_t(x, \cdot)$.

Preuve.

Pour tout $t \in \mathbb{R}^+$, on définit

$$\tilde{P}_t(x, \cdot) := \frac{P_t(x, \cdot) - (1 - d^s(t))\pi}{d^s(t)} \quad (1.17)$$

$\tilde{P}_t(x, \cdot)$ est une mesure de probabilité sur $(\Omega, \mathcal{P}(\Omega))$ et $\pi = \pi \tilde{P}_t$.

Pour tout $A \in \mathcal{P}(\Omega)$, $\tilde{P}_t(x, A) \geq 0$ par définition de $d^s(t)$ et $\tilde{P}_t(x, \cdot)$ est évidemment une mesure sur $(\Omega, \mathcal{P}(\Omega))$.

De plus,

$$\tilde{P}_t(x, \Omega) = \frac{P_t(x, \Omega) - (1 - d^s(t))\pi(\Omega)}{d^s(t)} = 1$$

Enfin remarquons que pour tout $y \in \Omega$, et pour x quelconque fixé dans Ω

$$\pi(x)\tilde{P}_t(x, y) = \frac{\pi(x)P_t(x, y) - \pi(x)(1 - d^s(t))\pi(y)}{d^s(t)}$$

d'où

$$(\pi \tilde{P}_t)(y) = \sum_{x \in \Omega} \pi(x)\tilde{P}_t(x, y) = \frac{\pi(y) - (1 - d^s(t))\pi(y)}{d^s(t)} = \pi(y)$$

Ceci étant vrai pour tout $y \in \Omega$, on obtient le résultat voulu.

Preuve de la proposition 1.3.2. D'après le lemme 1.3.4, on a pour tout $u, t \in \mathbb{R}^+$ et pour tout $y, z \in \Omega$

$$\begin{cases} P_t(x, z) &= (1 - d^s(t))\pi(z) + d^s(t)\tilde{P}_t(x, z) \\ P_u(z, y) &= (1 - d^s(u))\pi(y) + d^s(u)\tilde{P}_u(z, y) \end{cases}$$

d'où en multipliant la première relation avec la deuxième et en sommant sur z

$$\sum_{z \in \Omega} P_t(x, z)P_u(z, y) = \pi(y) - \pi(y)d^s(u) + d^s(u)(\pi \tilde{P}_u)(y) - d^s(t)d^s(u)(\pi \tilde{P}_u)(y) + d^s(t)d^s(u) \sum_{z \in \Omega} \tilde{P}_t(x, z)\tilde{P}_u(z, y)$$

mais puisque $\pi \tilde{P}_u = \pi$, on a

$$P_{t+u}(x, y) = \pi(y)(1 - d^s(t)d^s(u)) + d^s(t)d^s(u) \sum_{z \in \Omega} \tilde{P}_t(x, z) \tilde{P}_u(z, y)$$

c'est-à-dire

$$1 - \frac{P_{t+u}(x, y)}{\pi(y)} = d^s(t)d^s(u) \left(1 - \frac{\sum_{z \in \Omega} \tilde{P}_t(x, z) \tilde{P}_u(z, y)}{\pi(y)} \right) \leq d^s(t)d^s(u) \quad \forall x, y \in \Omega$$

soit encore

$$\max_{y \in \Omega} \max_{x \in \Omega} \left(1 - \frac{P_{t+u}(x, y)}{\pi(y)} \right) \leq d^s(t)d^s(u)$$

ce qui achève la preuve.

1.4 Phénomène de cutoff et de pre-cutoff

1.4.1 Cas d'une suite de chaînes simples

On considère une suite de chaînes de Markov $(X^n)_{n \in \mathbb{N}}$ dont le n -ème terme $X^n = (X^n(t))_{t \in \mathbb{R}^+}$ est une chaîne de Markov d'espace d'état Ω_n fini, continue à droite, réversible, irréductible et de mesure invariante π_n . On note $(P_t^n)_{t \in \mathbb{R}^+}$ le semi-groupe de Markov associé à X^n .

Comme précédemment, voici des distances intéressantes associées à la chaîne de Markov X^n :

Pour tout $n \in \mathbb{N}$

$$d_n(t) := \max_{x \in \Omega(n)} \|P_t^n(x, \cdot) - \pi_n\|_{TV} \quad \forall t \in \mathbb{R}^+$$

$$d_n^H(t) := \sup_{x \in \Omega(n)} d^H(P_t^n(x, \cdot), \pi_n) \quad \forall t \in \mathbb{R}^+$$

$$d_n^s(t) := \max_{x, y \in \Omega(n)} \left(1 - \frac{P_t^n(x, y)}{\pi_n(y)} \right) = 1 - \min_{x, y \in \Omega(n)} \frac{P_t^n(x, y)}{\pi_n(y)} \quad \forall t \in \mathbb{R}^+$$

Définition 1.4.1. Soit $a \in (0, 1)$. On définit le temps de mélange $t_{mix}^n(a)$ de la suite de chaînes de Markov $(X^n)_{n \in \mathbb{N}}$ pour la distance en variation totale par

$$t_{mix}^n(a) := \inf\{t : d_n(t) < a\}$$

On note $t_{mix}^n = t_{mix}^n(1/4)$

Définition 1.4.2. Soit $a \in (0, 1)$. On définit le temps de mélange $t_s^n(a)$ de la suite de chaînes de Markov $(X^n)_{n \in \mathbb{N}}$ pour la distance de séparation par

$$t_s^n(a) := \inf\{t : d_n^s(t) < a\}$$

$t_s^n = t_s^n(1/4)$

Définition 1.4.3. La suite de chaînes de Markov $(X^n)_{n \in \mathbb{N}}$ présente un cutoff pour la distance en variation totale si pour tout $\varepsilon \in (0, 1/2]$

$$\lim_{n \rightarrow +\infty} \frac{t_{mix}^n(\varepsilon)}{t_{mix}^n(1 - \varepsilon)} = 1 \quad (1.18)$$

Définition 1.4.4. La suite de chaînes de Markov $(X^n)_{n \in \mathbb{N}}$ présente un cutoff pour la distance de séparation si pour tout $\varepsilon \in (0, 1/2]$

$$\lim_{n \rightarrow +\infty} \frac{t_s^n(\varepsilon)}{t_s^n(1 - \varepsilon)} = 1 \quad (1.19)$$

Voici un résultat assez intuitif que nous ne démontrons pas :

Proposition 1.4.5. *La suite de chaînes de Markov $(X^n)_{n \in \mathbb{N}}$ présente un cutoff pour la distance en variation totale (respectivement pour la distance de séparation) si et seulement si*

$$\lim_{n \rightarrow \infty} d_n(ct_{mix}^n) = \begin{cases} 1 & \text{si } c < 1 \\ 0 & \text{si } c > 1 \end{cases} \quad (1.20)$$

respectivement si et seulement si

$$\lim_{n \rightarrow \infty} d_n^s(ct_s^n) = \begin{cases} 1 & \text{si } c < 1 \\ 0 & \text{si } c > 1 \end{cases} \quad (1.21)$$

Définition 1.4.6. *La suite de chaînes de Markov $(X^n)_{n \in \mathbb{N}}$ présente un pre-cutoff pour la distance en variation totale si*

$$\limsup_{\varepsilon \rightarrow 0+} \limsup_{n \rightarrow +\infty} \frac{t_{mix}^n(\varepsilon)}{t_{mix}^n(1 - \varepsilon)} < \infty \quad (1.22)$$

Définition 1.4.7. *La suite de chaînes de Markov $(X^n)_{n \in \mathbb{N}}$ présente un pre-cutoff pour la distance de séparation si*

$$\limsup_{\varepsilon \rightarrow 0+} \limsup_{n \rightarrow +\infty} \frac{t_s^n(\varepsilon)}{t_s^n(1 - \varepsilon)} < \infty \quad (1.23)$$

Remarque 1.4.8. *On peut voir que*

$$\limsup_{\varepsilon \rightarrow 0+} \limsup_{n \rightarrow +\infty} \frac{t_{mix}^n(\varepsilon)}{t_{mix}^n(1 - \varepsilon)} = \limsup_{k \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \frac{t_{mix}^n(\varepsilon_k)}{t_{mix}^n(1 - \varepsilon_k)}$$

pour toute suite $(\varepsilon_k)_{k \in \mathbb{N}}$ strictement positive qui converge vers 0.

1.4.2 Cas d'une suite de chaînes sur un espace produit

Soit $(X^n)_{n \in \mathbb{N}}$ la suite chaînes de Markov définie dans la sous-partie précédente.

On considère à présent une suite de chaînes de Markov $(Y^n)_{n \in \mathbb{N}}$ dont le n -ème terme $Y^n = (Y^n(t))_{t \in \mathbb{R}^+}$ est une chaîne de Markov d'espace d'état $\Omega(n) := \underbrace{\Omega_n \times \cdots \times \Omega_n}_n$ définie par

$$Y^n(t) := (X_1^n(t), \dots, X_n^n(t)) \quad \forall t \in \mathbb{R}^+$$

où les chaînes $X_1^n = (X_1^n(t))_t, \dots, X_n^n = (X_n^n(t))_t$ sont n copies indépendantes de la chaîne $X^n = (X^n(t))_{t \in \mathbb{R}^+}$.

Notons $(Q_t^n)_{t \in \mathbb{R}^+}$ le semi-groupe de Markov de Y^n .

Étant donné que les n chaînes X_1^n, \dots, X_n^n sont indépendantes, que chacune de ces chaînes a respectivement $(P_t^n)_{t \in \mathbb{R}^+}$ et π_n pour semi-groupe de Markov et pour mesure invariante, on a pour tout $t \in \mathbb{R}^+$

$$Q_t^n(x, \cdot) = P_t^n(x_1, \cdot) \otimes \cdots \otimes P_t^n(x_n, \cdot) \quad \forall x = (x_1, \dots, x_n) \in \Omega(n) \quad (1.24)$$

et

$$\pi_n^{\otimes n} Q_t^n = \pi_n^{\otimes n} \quad (1.25)$$

On note $\mu_n = \pi_n^{\otimes n}$.

De manière analogue que pour la chaîne simple, nous pouvons définir les distances suivantes pour la chaîne produit :

$$D_n(t) := \max_{x \in \Omega(n)} \|Q_t^n(x, \cdot) - \mu_n\|_{TV} \quad \forall t \in \mathbb{R}^+ \quad (1.26)$$

$$D_n^{\mathcal{H}}(t) := \sup_{x \in \Omega(n)} d^{\mathcal{H}}(Q_t^n(x, \cdot), \mu_n) \quad \forall t \in \mathbb{R}^+ \quad (1.27)$$

$$D_n^s(t) := \max_{x, y \in \Omega(n)} \left(1 - \frac{Q_t^n(x, y)}{\mu_n(y)} \right) = 1 - \min_{x, y \in \Omega(n)} \frac{Q_t^n(x, y)}{\mu_n(y)} \quad \forall t \in \mathbb{R}^+ \quad (1.28)$$

Proposition 1.4.9. *Pour tout $t \in \mathbb{R}^+$, les distances $D_n^{\mathcal{H}}$ et $d_n^{\mathcal{H}}$ sont liées par la relation suivante*

$$1 - \frac{1}{2} (D_n^{\mathcal{H}}(t))^2 = \left(1 - \frac{1}{2} (d_n^{\mathcal{H}}(t))^2 \right)^n \quad (1.29)$$

Preuve. Rappelons la définition de l'affinité d'Hellinger. Pour deux mesures μ et ν de probabilité sur Ω fini

$$I(\mu, \nu) = \sum_{x \in \Omega} \sqrt{\mu(x)\nu(x)}$$

Soit $x = (x_1, \dots, x_n) \in \Omega(n)$. On a pour tout $t \in \mathbb{R}^+$

$$d^H(Q_t^n(x, \cdot), \mu_n)^2 = 2 - 2I(Q_t^n(x, \cdot), \mu_n) \quad (1.30)$$

Cependant d'après (1.24) et (1.25), on a

$$I(Q_t^n(x, \cdot), \mu_n) = \sum_{(y_1, \dots, y_n) \in \Omega(n)} \sqrt{P_t^n(x_1, y_1) \cdots P_t^n(x_n, y_n) \pi_n(y_1) \cdots \pi_n(y_n)}$$

c'est-à-dire

$$I(Q_t^n(x, \cdot), \mu_n) = \prod_{i=1}^n \sum_{y_i \in \Omega_n} \sqrt{P_t^n(x_i, y_i) \pi_n(y_i)}$$

d'où

$$I(Q_t^n(x, \cdot), \mu_n) = \prod_{i=1}^n I(P_t^n(x_i, \cdot), \pi_n) \quad (1.31)$$

De plus, remarquons que

$$(D_n^{\mathcal{H}}(t))^2 = \left(\sup_{x \in \Omega(n)} d^H(Q_t^n(x, \cdot), \mu_n) \right)^2 = \sup_{x \in \Omega(n)} d^H(Q_t^n(x, \cdot), \mu_n)^2$$

ce qui grâce à (1.30) et (1.31) permet d'écrire

$$(D_n^{\mathcal{H}}(t))^2 = 2 - 2 \prod_{i=1}^n \inf_{x_i \in \Omega_n} I(P_t^n(x_i, \cdot), \pi_n)$$

soit encore

$$1 - \frac{1}{2} (D_n^{\mathcal{H}}(t))^2 = \prod_{i=1}^n \left(1 - \frac{1}{2} \inf_{x_i \in \Omega_n} d^H(P_t^n(x_i, \cdot), \pi_n) \right) = \prod_{i=1}^n \left(1 - \frac{1}{2} (d_n^{\mathcal{H}}(t))^2 \right)$$

et finalement

$$1 - \frac{1}{2} (D_n^{\mathcal{H}}(t))^2 = \left(1 - \frac{1}{2} (d_n^{\mathcal{H}}(t))^2 \right)^n$$

ce qui achève la démonstration.

Proposition 1.4.10. *Pour tout $t \in \mathbb{R}^+$, les distances D_n^s et d_n^s sont liées par la relation suivante*

$$D_n^s(t) = 1 - (1 - d_n^s(t))^n \quad (1.32)$$

Preuve. En utilisant, (1.24) et (1.25), on a, pour tout $t \in \mathbb{R}^+$ et pour tout $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in \Omega(n)$

$$1 - \frac{Q_t^n(x, y)}{\mu_n(y)} = 1 - \prod_{i=1}^n \frac{P_t^n(x_i, y_i)}{\mu_n(y_i)}$$

d'où

$$\max_{x \in \Omega(n)} \left(1 - \frac{Q_t^n(x, y)}{\mu_n(y)} \right) = 1 - \prod_{i=1}^n \min_{x_i \in \Omega_n} \frac{P_t^n(x_i, y_i)}{\mu_n(y_i)} = 1 - \prod_{i=1}^n (1 - d_n^s(t))$$

soit

$$D_n^s(t) = 1 - (1 - d_n^s(t))^n$$

ce qui achève la preuve.

Définition 1.4.11. *Soit $a \in (0, 1)$. On définit le temps de mélange $T_{mix}^n(a)$ de la suite de chaînes de Markov $(Y^j)_{j \in \mathbb{N}}$ pour la distance en variation totale par*

$$T_{mix}^n(a) := \inf\{t : D_n(t) < a\}$$

On note $T_{mix}^n = T_{mix}^n(1/4)$

Définition 1.4.12. *Soit $a \in (0, 1)$. On définit le temps de mélange $T_s^n(a)$ de la suite de chaînes de Markov $(Y^j)_{j \in \mathbb{N}}$ pour la distance de séparation par*

$$T_s^n(a) := \inf\{t : D_n^s(t) < a\}$$

$T_s^n = T_s^n(1/4)$

Définition 1.4.13. *La suite de chaînes de Markov $(Y^n)_{n \in \mathbb{N}}$ présente un cutoff pour la distance en variation totale si pour tout $\varepsilon \in (0, 1/2]$*

$$\lim_{n \rightarrow +\infty} \frac{T_{mix}^n(\varepsilon)}{T_{mix}^n(1 - \varepsilon)} = 1 \quad (1.33)$$

Définition 1.4.14. La suite de chaînes de Markov $(Y^n)_{n \in \mathbb{N}}$ présente un cutoff pour la distance de séparation si pour tout $\varepsilon \in (0, 1/2]$

$$\lim_{n \rightarrow +\infty} \frac{T_s^n(\varepsilon)}{T_s^n(1-\varepsilon)} = 1 \quad (1.34)$$

Proposition 1.4.15. La suite de chaînes de Markov $(Y^n)_{n \in \mathbb{N}}$ présente un cutoff pour la distance en variation totale (respectivement pour la distance de séparation) si et seulement si

$$\lim_{n \rightarrow \infty} D_n(cT_{mix}^n) = \begin{cases} 1 & \text{si } c < 1 \\ 0 & \text{si } c > 1 \end{cases} \quad (1.35)$$

respectivement si et seulement si

$$\lim_{n \rightarrow \infty} D_n^s(cT_s^n) = \begin{cases} 1 & \text{si } c < 1 \\ 0 & \text{si } c > 1 \end{cases} \quad (1.36)$$

Définition 1.4.16. La suite de chaînes de Markov $(Y^n)_{n \in \mathbb{N}}$ présente un pre-cutoff pour la distance en variation totale si

$$\limsup_{\varepsilon \rightarrow 0+} \limsup_{n \rightarrow +\infty} \frac{T_{mix}^n(\varepsilon)}{T_{mix}^n(1-\varepsilon)} < \infty \quad (1.37)$$

Définition 1.4.17. La suite de chaînes de Markov $(Y^n)_{n \in \mathbb{N}}$ présente un pre-cutoff pour la distance de séparation si

$$\limsup_{\varepsilon \rightarrow 0+} \limsup_{n \rightarrow +\infty} \frac{T_s^n(\varepsilon)}{T_s^n(1-\varepsilon)} < \infty \quad (1.38)$$

Dans le théorème qui va suivre, nous allons évoquer le terme "triviale" pour qualifier une chaîne de Markov. Par triviale on entend par exemple une chaîne $X = (X(t))_t$ de semi-groupe de Markov $(P_t)_t$ d'espace d'état $\Omega = \{x\}$. Notons π sa mesure invariante. Il est évident que pour tout t , $P_t(x, \cdot) = 1$ et $\pi = 1$. Ainsi pour tout t , $d(t) = 0$. Il vient de même que pour tout t , $Q_t((x, \dots, x), \cdot) = 1$ et $\mu = 1$. Donc $D(t) = 0$ pour tout t .

Théorème 1.4.18. (Hubert Lacoïn, 2015)

Pour toute suite de chaînes de Markov $(X^n)_{n \in \mathbb{N}}$ non triviales vérifiant les propriétés définies dans la sous-partie précédente, on a, pour tout $\varepsilon \in (0, 1/2]$

$$\limsup_{n \rightarrow +\infty} \frac{T_{mix}^n(\varepsilon)}{T_{mix}^n(1-\varepsilon)} \leq 2 \quad (1.39)$$

et

$$\limsup_{n \rightarrow +\infty} \frac{T_s^n(\varepsilon)}{T_s^n(1-\varepsilon)} \leq 2 \quad (1.40)$$

Corollaire 1.4.18.1. *Pour toute suite de chaîne de Markov $(X^n)_{n \in \mathbb{N}}$ non triviale vérifiant les propriétés définies dans la sous-partie précédente, la suite $(Y^n)_{n \in \mathbb{N}}$ admet un pre-cutoff pour la distance en variation totale et pour la distance en séparation.*

Preuve. La preuve est immédiate. En effet, le théorème 1.4.18 dit en particulier que

$$\limsup_{\varepsilon \rightarrow 0^+} \limsup_{n \rightarrow +\infty} \frac{T_{\text{mix}}^n(\varepsilon)}{T_{\text{mix}}^n(1-\varepsilon)} \leq 2 < \infty$$

et

$$\limsup_{\varepsilon \rightarrow 0^+} \limsup_{n \rightarrow +\infty} \frac{T_s^n(\varepsilon)}{T_s^n(1-\varepsilon)} \leq 2 < \infty$$

Preuve du théorème 1.4.18. Commençons par (1.40).

Pour tout $\varepsilon \in (0, 1/2)$ et pour tout n , l'inégalité

$$T_s^n(1-\varepsilon) \leq T_s^n(\varepsilon)$$

est toujours vraie.

Montrons alors qu'il existe $\alpha < 0$ tel que pour tout $\varepsilon \in (0, 1/2)$ et pour n assez grand

$$t_n^s(n^\alpha) \leq T_s^n(1-\varepsilon) \leq T_s^n(\varepsilon) \leq t_n^s(n^{2\alpha}) \quad (1.41)$$

Posons

$$A_n := \left\{ \alpha < 0 : (1-n^\alpha)^n \leq \varepsilon \text{ et } 1 - (1-n^{2\alpha})^n \leq \varepsilon, \forall \varepsilon \in (0, 1/2) \right\}$$

D'après la proposition 1.4.10, pour montrer qu'il existe $\alpha < 0$ satisfaisant (1.41), il suffit de montrer que A_n est non vide pour n assez grand.

Remarquons que comme $\alpha < 0$

$$(1-n^\alpha)^n = e^{n \log(1-n^\alpha)} = e^{-n^{\alpha+1} + o(n^{\alpha+1})}$$

et donc si $\alpha > -1$ alors $\lim_{n \rightarrow +\infty} (1-n^\alpha)^n = 0$.

De même

$$(1-n^{2\alpha})^n = e^{n \log(1-n^{2\alpha})} = e^{-n^{2\alpha+1} + o(n^{2\alpha+1})}$$

et donc si $\alpha < -\frac{1}{2}$ alors $\lim_{n \rightarrow +\infty} 1 - (1-n^{2\alpha})^n = 0$.

Ce qui prouve que pour n assez grand

$$(-1, -1/2) \subseteq A_n$$

et bien évidemment A_n est non vide.

Soit $\alpha \in (-1, -1/2)$. D'après la proposition 1.3.2, on a

$$d(2t_n^s(n^\alpha)) \leq d(t_n^s(n^\alpha))^2 \leq n^{2\alpha}$$

donc

$$t_n^s(n^{2\alpha}) \leq 2t_n^s(n^\alpha)$$

d'où selon (1.41), pour tout $\varepsilon \in (0, 1/2)$ et pour n assez grand

$$t_n^s(n^\alpha) \leq T_s^n(1 - \varepsilon) \leq T_s^n(\varepsilon) \leq 2t_n^s(n^\alpha)$$

c'est-à-dire pour tout $\varepsilon \in (0, 1/2)$,

$$\limsup_{n \rightarrow +\infty} \frac{T_s^n(\varepsilon)}{T_s^n(1 - \varepsilon)} \leq 2$$

Montrons à présent (1.39). L'idée est la même. Il s'agit de trouver $\alpha < 0$ tel que pour tout $\varepsilon \in (0, 1/2)$ et pour n assez grand

$$t_n \leq T_{mix}^n(1 - \varepsilon) \leq T_{mix}^n(\varepsilon) \leq t'_n \tag{1.42}$$

où $t_n := \inf\{t : d_n^{\mathcal{H}}(t) \leq n^\alpha\}$ et $t'_n := \inf\{t : d_n^{\mathcal{H}}(t) \leq 18n^{5\alpha/4}\}$.

Remarquons que pour toute fonction réelle $a \geq 0$, $\sup_x a(x)^2 = \left(\sup_x a(x)\right)^2$ ce qui permet, d'après la proposition 1.2.2, d'écrire

$$D_n(t) \leq D_n^{\mathcal{H}}(t) \leq \sqrt{2D_n(t)} \tag{1.43}$$

Posons

$$A'_n := \left\{ \alpha < 0 : \left(1 - \frac{1}{2}n^{2\alpha}\right)^n \leq \varepsilon \text{ et } 1 - \left(1 - 162n^{5\alpha/2}\right)^n \leq \varepsilon, \forall \varepsilon \in (0, 1/2) \right\}$$

D'après la proposition 1.4.10 et (1.43), pour montrer qu'il existe $\alpha < 0$ satisfaisant (1.42), il suffit de montrer que A'_n est non vide pour n assez grand.

Remarquons que comme $\alpha < 0$

$$\left(1 - \frac{1}{2}n^{2\alpha}\right)^n = e^{n \log(1 - \frac{1}{2}n^{2\alpha})} = e^{-\frac{1}{2}n^{2\alpha+1} + o(n^{2\alpha+1})}$$

et donc si $\alpha > -\frac{1}{2}$ alors $\lim_{n \rightarrow +\infty} \left(1 - \frac{1}{2}n^{2\alpha}\right)^n = 0$.

De même

$$\left(1 - 162n^{5\alpha/2}\right)^n = e^{n \log(1 - 162n^{5\alpha/2})} = e^{-162n^{5\alpha/2} + o(n^{5\alpha/2+1})}$$

et donc si $\alpha < -\frac{2}{5}$ alors $\lim_{n \rightarrow +\infty} 1 - (1 - 162n^{5\alpha/2})^n = 0$.

Ce qui prouve que pour n assez grand

$$(-1/2, -2/5) \subseteq A'_n$$

et bien évidemment A'_n est non vide.

Soit $\alpha \in (-1/2, -2/5)$. D'après la proposition 1.2.4, on a

$$d^{\mathcal{H}}(2t_n) \leq 18(d^{\mathcal{H}}(t_n))^{5/4} \leq 18n^{5\alpha/4}$$

donc

$$t'_n \leq 2t_n$$

d'où d'après (1.42), pour tout $\varepsilon \in (0, 1/2)$ et pour n assez grand

$$t_n \leq T_{\text{mix}}^n(1 - \varepsilon) \leq T_{\text{mix}}^n(\varepsilon) \leq 2t_n$$

c'est-à-dire pour tout $\varepsilon \in (0, 1/2)$,

$$\limsup_{n \rightarrow +\infty} \frac{T_{\text{mix}}^n(\varepsilon)}{T_{\text{mix}}^n(1 - \varepsilon)} \leq 2$$

ce qui prouve finalement le théorème 1.4.18.

Remarque 1.4.19. *L'exposant 5/4 de la proposition 1.2.4 est très important dans la preuve que nous venons de faire. En effet, si nous remplacions 5/4 par $0 < \gamma < 1$, nous n'arriverions pas à trouver un $\alpha > 0$ satisfaisant (1.42) puisque α devrait alors satisfaire $\alpha > -1/2$ et $\alpha < -1/2\gamma$ ce qui est évidemment impossible.*

Partie 2

Un exemple de suite ayant un pre-cutoff sans avoir de cutoff

2.1 Construction de l'exemple

Hubert Lacoïn a montré que la condition produit n'est pas forcément suffisante en exhibant un exemple de suite de chaînes de Markov ayant un pre-cutoff mais ne présentant pas de cutoff [4, 3. An example without cutoff]. En particulier, son exemple montre que la borne 2 dans le théorème 1.4.18 est optimale.

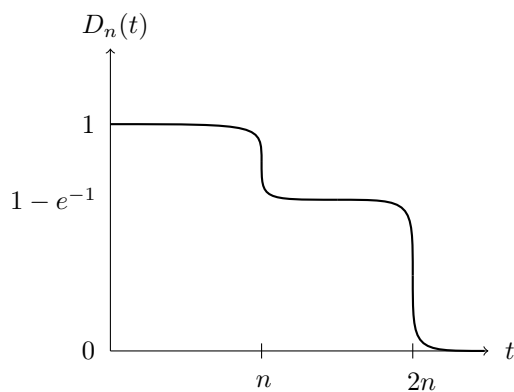


FIGURE 2.1 – Graphe de $\lim_{n \rightarrow \infty} D_n(t)$. La borne 2 est en effet optimale car pour tout $\varepsilon \in (0, e^{-1})$, $T_{\text{mix}}^n(\varepsilon)/T_{\text{mix}}^n(1 - \varepsilon) = 2n/n = 2$.

Nous allons construire une variante très proche de l'exemple proposé par Hubert Lacoïn permettant de répondre à la question suivante : de quelle manière la géométrie de la chaîne influence-t-elle le comportement de sa distance à l'équilibre ?

Soit $G_n := (V_n, E_n)$ un graphe fini tel que :

- $V_n = \{A_0, A_1, \dots, A_{5n}\}$ est l'ensemble des sommets de G_n et $|V_n| = 5n + 1$.
Notons $A := A_0, B := A_{3n}, C := A_{4n}$ et $D := A_{5n}$.
- E_n est l'ensemble des arêtes de G_n et $|E_n| = 5n + 2$.
- Il y a un chemin de $5n$ arêtes reliant les $5n + 1$ sommets.
- Il y a deux arêtes supplémentaires : une arête reliant directement B et C et une autre arête reliant directement C et D .

Soit $(X^n)_{n \in \mathbb{N}}$ une suite de chaînes de Markov dont le n -ème terme $X^n = (X^n(t))_{t \in \mathbb{R}^+}$ est une chaîne de Markov d'espace d'état V_n , de sous-groupe de Markov $(P_t^n)_{t \in \mathbb{R}^+}$ et de mesure stationnaire π_n .

Voici une illustration du graphe G_n où nous avons indiqué les taux de transition de la chaîne X^n :

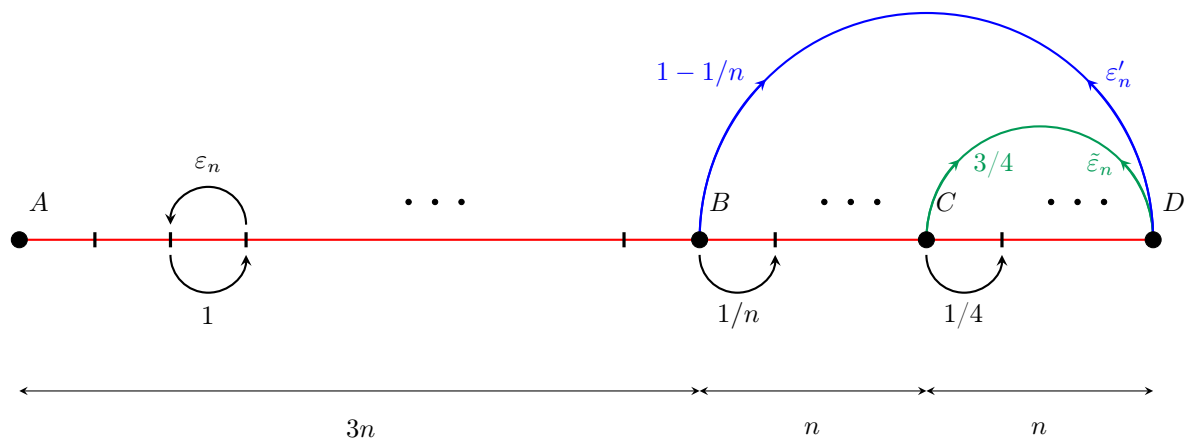


FIGURE 2.2 – Le graphe G_n avec les taux de transition de la chaîne X^n . De A vers D , sur le chemin rouge, X^n saute toujours avec un taux égal à 1 sauf au point B où X^n saute sur A_{3n+1} avec un taux égale à $1/n$ et au point C où X^n saute vers A_{4n+1} avec taux $1/4$. Au point B , X^n peut sauter avec taux $1 - 1/n$ directement vers D en empruntant le chemin bleu. Au point C , X^n peut sauter avec taux $3/4$ directement vers D en empruntant le chemin vert.

De D vers A , sur le chemin rouge, X^n saute toujours avec un taux égal à $\varepsilon_n := 2^{-n^2}$ sauf au point D où X^n peut sauter avec taux ε'_n directement vers B en empruntant le chemin bleu ou sauter avec taux $\tilde{\varepsilon}_n$ directement vers C en empruntant le chemin vert.

Remarque 2.1.1. Prendre ε_n exponentiellement petit rend les calculs plus commodes. De plus, avec cette construction, nous avons bien une suite de chaînes de Markov continues, irréductibles mais pas nécessairement réversibles. Nous choisissons alors ε'_n et $\tilde{\varepsilon}_n$ de sorte que pour tout $n \in \mathbb{N}$, X^n soit réversible. Notons $(Q_t^n)_{t \in \mathbb{R}^+}$ le générateur infinitésimal de X^n . On alors $\pi_n Q_t^n = 0$ d'où

$$\pi_n(B) \cdot \frac{1}{n} \cdot \frac{1}{4} = \pi_n(D) \cdot (\varepsilon_n)^{2n} \quad \text{et} \quad \pi_n(C) \cdot \frac{1}{4} = \pi_n(D) \cdot (\varepsilon_n)^n$$

et pour que X^n soit réversible, il suffit que

$$\pi_n(B)Q_t^n(B, D) = \pi_n(D)Q_t^n(D, B) \quad \text{et} \quad \pi_n(C)Q_t^n(C, D) = \pi_n(D)Q_t^n(D, C)$$

soit finalement

$$\varepsilon'_n = 4(n-1)2^{-2n^3} \quad \text{et} \quad \tilde{\varepsilon}_n = 3 \cdot 2^{-n^3}$$

Notons $(J_i)_{i \in \mathbb{N}}$ la suite des temps de saut de X^n . J_i dépend bien sûr de n . Afin de ne pas charger les notations, nous écrirons $J_i^n = J_i$.

Rigoureusement, on a $J_0 = 0$ et pour tout $i \geq 1$

$$J_i := \inf\{t > J_{i-1} : X(t) \neq X(J_{i-1})\}$$

Posons pour tout $n \geq 1$

$$S_i := J_i - J_{i-1}$$

S_i est alors le temps passé par X^n dans l'état $X^n(J_{i-1})$.

Notons $Q_t^n(k)$ le k -ème élément diagonal de Q_t^n . En utilisant la définition donnée dans [6, p. 87-88], le processus $(X^n(J_i))_{i \in \mathbb{N}}$ est une chaîne de Markov discrète de matrice de transition Π telle que

$$\Pi(A, A_l) = \mathbb{1}_{\{l=1\}}$$

$$\Pi(A_k, A_l) = \frac{\varepsilon_n}{1 + \varepsilon_n} \mathbb{1}_{\{l=k-1\}} + \frac{1}{1 + \varepsilon_n} \mathbb{1}_{\{l=k+1\}} \quad \forall k \notin \{3n, 4n, 5n\}$$

$$\Pi(B, A_l) = \frac{\varepsilon_n}{1 + \varepsilon_n} \mathbb{1}_{\{l=3n-1\}} + \frac{1/n}{1 + \varepsilon_n} \mathbb{1}_{\{l=3n+1\}} + \frac{1 - 1/n}{1 + \varepsilon_n} \mathbb{1}_{\{l=5n\}}$$

$$\Pi(C, A_l) = \frac{\varepsilon_n}{1 + \varepsilon_n} \mathbb{1}_{\{l=4n-1\}} + \frac{1/4}{1 + \varepsilon_n} \mathbb{1}_{\{l=4n+1\}} + \frac{3/4}{1 + \varepsilon_n} \mathbb{1}_{\{l=5n\}}$$

$$\Pi(D, A_l) = \frac{\varepsilon_n}{\varepsilon_n + \varepsilon'_n + \tilde{\varepsilon}_n} \mathbb{1}_{\{l=5n-1\}} + \frac{\tilde{\varepsilon}_n}{\varepsilon_n + \varepsilon'_n + \tilde{\varepsilon}_n} \mathbb{1}_{\{l=4n\}} + \frac{\tilde{\varepsilon}_n}{\varepsilon_n + \varepsilon'_n + \tilde{\varepsilon}_n} \mathbb{1}_{\{l=3n\}}$$

De plus, conditionnellement à $X^n(J_0), \dots, X^n(J_{i-1})$, les variables aléatoires S_1, \dots, S_i sont *iid* et suivent une loi exponentielle respectivement de paramètre $Q(X^n(J_0)), \dots, Q(X^n(J_{i-1}))$. Dans notre cas, le temps passé en A suit une loi exponentielle de paramètre 1, le temps passé en D suit une loi exponentielle de paramètre $\varepsilon_n + \varepsilon'_n + \tilde{\varepsilon}_n$ et enfin le temps passé dans un état x différent de A et D suit une loi exponentielle de paramètre $1 + \varepsilon_n$.

Soit $(Y^n)_{n \in \mathbb{N}}$ la suite de chaînes de Markov définie à partir de $(X^n)_{n \in \mathbb{N}}$ comme dans la section 4.2. Rappelons alors que la chaîne Y^n a pour espace d'état l'espace produit $V(n) := \underbrace{V_n \times \dots \times V_n}_n$.

Le but du prochain théorème est de montrer que $(X^n)_{n \in \mathbb{N}}$ présente un cutoff mais que $(Y^n)_{n \in \mathbb{N}}$ n'est présente pas. $(Y^n)_{n \in \mathbb{N}}$ présente bien sûr un pre-cutoff d'après 1.4.18.1.

La clé de la démonstration, très inspirée une nouvelle fois de celle d'Hubert Lacoïn, est que D_n (respectivement d_n) peut-être exprimée en fonction de la loi du temps d'atteinte de l'état $\mathcal{D} := \underbrace{(D, \dots, D)}_n$ (respectivement de l'état D).

Bien sûr le choix du chemin pour atteindre \mathcal{D} (respectivement D) est très important. La probabilité que la chaîne X^n , une fois arrivée en B , continue sa route le long du chemin rouge est $\frac{1/n}{1+2^{-n^2}}$, ce qui décroît assez rapidement vers 0 quand n tend vers l'infini.

En revanche $\mathbb{P}(\#\{X_i^n \text{ empruntant le chemin rouge}\} = k) = \binom{n}{k} \left(\frac{1/n}{1+2^{-n^2}}\right)^k \left(\frac{1-1/n}{1+2^{-n^2}}\right)^{n-k} \xrightarrow{n \rightarrow \infty} \frac{e^{-1}}{k!}$. Donc, pour n assez grand, la probabilité qu'au moins une des coordonnées X_i^n de la chaîne Y^n , une fois arrivée en B , continue sa route le long du chemin rouge est $e^{-1} \sum_{k \geq 1} \frac{1}{k!} = 1 - e^{-1} \approx 0,63$.

Ainsi, contrairement à la chaîne X^n , dont la probabilité d'emprunter le chemin rouge entre B et D est négligeable pour n assez grand, au moins l'une des coordonnées X_i^n de la chaîne Y^n empruntera le chemin rouge entre B et D avec une assez forte probabilité pour n assez grand. Le comportement de D_n en sera donc fortement affecté ce qui explique l'absence de cutoff pour la suite Y^n .

Théorème 2.1.2. *La construction précédente vérifie les propriétés suivantes :*

- (i) *La suite $(X^n)_{n \in \mathbb{N}}$ présente un cutoff en un temps proche de $3n$.*
- (ii) *La suite $(Y^n)_{n \in \mathbb{N}}$ ne présente pas de cutoff et*

$$T_{mix}^n(a) = \begin{cases} 5n(1 + o(1)) & \text{si } a \in (0, 1 - e^{-1/4}) \\ 4n(1 + o(1)) & \text{si } a \in (1 - e^{-1/4}, 1 - e^{-1}) \\ 3n(1 + o(1)) & \text{si } a \in (1 - e^{-1}, 1) \end{cases}$$

Remarque 2.1.3. *(1.39) est bien vérifiée. En effet, d'après le théorème 2.1.2, on a*

$$\limsup_{n \rightarrow \infty} \frac{T_{mix}^n(a)}{T_{mix}^n(1-a)} \leq \frac{5}{3} \leq 2 \quad \forall a \in (0, 1)$$

2.2 Preuve du théorème 2.1.3

Avant de démontrer le théorème 2.1.2, nous allons énoncer quelques propriétés vérifiées par les suites $(X^n)_{n \in \mathbb{N}}$ et $(Y^n)_{n \in \mathbb{N}}$ puis démontrer un lemme.

La mesure invariante de la chaîne de Markov X^n , π_n , charge presque uniquement D et la mesure invariante de la chaîne de Markov Y^n , μ_n , charge presque uniquement \mathcal{D} .

En effet, puisque $\pi_n Q_t^n = 0$, on a :

$$\begin{cases} \pi_n(A_i) \frac{1}{4n} = \varepsilon_n^{5n-i} \pi_n(D) & \text{si } i \in \{0, \dots, 3n\} \\ \pi_n(A_i) \frac{1}{4} = \varepsilon_n^{5n-i} \pi_n(D) & \text{si } i \in \{3n+1, \dots, 4n\} \\ \pi_n(A_i) = \varepsilon_n^{5n-i} \pi_n(D) & \text{si } i \in \{4n+1, \dots, 5n\} \end{cases}$$

On a alors

$$\pi_n(\{A_i\}_{i=0}^{5n-1}) = \sum_{i=0}^{5n-1} \pi_n(A_i) = A_n \pi_n(D) \quad (2.1)$$

où

$$A_n = 4n \sum_{i=2n}^{5n} \varepsilon_n^i + 4 \sum_{i=n}^{2n-1} \varepsilon_n^i + \sum_{i=1}^{n-1} \varepsilon_n^i \quad (2.2)$$

Mais comme π_n est une mesure de probabilité,

$$\pi_n(\{A_i\}_{i=0}^{5n-1}) + \pi_n(D) = 1$$

donc d'après (2.1)

$$\pi_n(D) = 1 - \frac{A_n}{1 + A_n} \quad (2.3)$$

Il s'agit maintenant de montrer que $\frac{A_n}{1+A_n} = O(2^{-n^2})$.

En effet,

$$\frac{A_n}{1 + A_n} = 2^{-n^2} \frac{4n2^{-5n^3} - 4n2^{-2n^3+n^2} + 4 \times 2^{-3n^3+n^2} - 4 \times 2^{-n^3+n^2} + 2^{-n^3} - 1}{4n2^{-5n^3-n^2} - 4n2^{-2n^3} + 4 \times 2^{-3n^3} - 4 \times 2^{-n^3} + 2^{-n^3-n^2} - 2 \times 2^{-n^2} - 2}$$

d'où, pour tout $\eta > 0$ et pour n assez grand

$$\frac{A_n}{1 + A_n} \leq \left(\eta + \frac{1}{2}\right) 2^{-n^2} \quad (2.4)$$

et finalement, d'après (2.3) et (2.4), on a

$$\pi_n(D) = 1 - O(2^{-n^2}) \quad (2.5)$$

Enfin, puisque $\mu_n = \pi_n^{\otimes n}$

$$\mu_n(\mathcal{D}) = \pi_n(D)^n = \left(1 - \frac{A_n}{1 + A_n}\right)^n = e^{-k_n}$$

où $k_n = n \frac{A_n}{1 + A_n} + o\left(n \frac{A_n}{1 + A_n}\right)$

ou encore

$$\mu_n(\mathcal{D}) = 1 - k_n + o(k_n)$$

donc d'après (2.4),

$$\mu_n(\mathcal{D}) = 1 - O(n2^{-n^2}) \quad (2.6)$$

Enfin, d'après la remarque 1.1.2 et puisque pour tout $E \in \mathcal{P}(V_n)$, $E = (E \cap \{D\} \cup E \cap \{D\}^c)$, on a, pour tout $t \in \mathbb{R}^+$

$$d_n(t) = \max_{x \in V_n} \max_{E \in \mathcal{P}(V_n)} (\pi_n(E \cap \{D\}) - P_t^n(x, E \cap \{D\}) + \pi_n(E \cap \{D\}^c) - P_t^n(x, E \cap \{D\}^c))$$

Si $E \cap \{D\} = \emptyset$ alors $D \notin E$ et donc $\pi_n(E) \leq \pi_n(\{A_i\}_{i=0}^{5n-1})$. D'où d'après (2.4)

$$d_n(t) \leq \pi_n(\{A_i\}_{i=0}^{5n-1}) - \min_{x \in V_n} \min_{E \in \mathcal{P}(V_n)} P_t^n(x, E \cap \{D\}^c) \leq \pi_n(\{A_i\}_{i=0}^{5n-1}) = O(2^{-n^2})$$

Donc nécessairement $E \cap \{D\} \neq \emptyset$ et finalement d'après (2.5)

$$d_n(t) = \pi_n(D) - \min_{x \in V_n} P_t^n(x, D) = 1 - O(2^{-n^2}) - \min_{x \in V_n} P_t^n(x, D) \quad (2.7)$$

De manière analogue, on a

$$D_n(t) = 1 - O(n2^{-n^2}) - \min_{x \in V(n)} Q_t^n(x, \mathcal{D}) \quad (2.8)$$

Définissons $\tau_n := \min\{t \geq 0 : X^n(t) = D\}$ et $\mathcal{T}_n := \min\{t \geq 0 : Y^n(t) = \mathcal{D}\}$.

$\mathbb{P}_x := \mathbb{P}(\cdot | X_0 = x)$ et $\mathbb{Q}_x := \mathbb{P}(\cdot | Y_0 = x)$ sont deux mesures de probabilité sur $(\mathcal{X}, \mathcal{F})$.

Lemme 2.2.1. *On a les deux propriétés suivantes :*

- (i) *La suite de fonctions $(d_n - \mathbb{P}_A(\tau_n > \cdot))_{n \in \mathbb{N}}$ converge uniformément vers 0.*
- (ii) *La suite de fonctions $(D_n - \mathbb{Q}_A(\mathcal{T}_n > \cdot))_{n \in \mathbb{N}}$ converge uniformément vers 0.*

Preuve. Nous allons seulement démontrer (i) car la preuve de (ii) est analogue.

Calculons dans un premier temps la probabilité que la chaîne X^n ne fasse aucun backtrack (déplacement en direction de A) avant d'avoir atteint D .

Par propriété de Markov simple, on a, pour tout $x \in V_n$ (notons $x = A_i, i < 5n$)

$$\mathbb{P}_x(X_{J_1} = A_1, \dots, X_{J_{5n-i}} = D) = \left(\frac{1}{1 + \varepsilon_n} \right)^{5n-i} = \left(1 - \frac{2^{-n^2}}{1 + 2^{-n^2}} \right)^{5n-i} = e^{-(5n-i) \frac{2^{-n^2}}{1 + 2^{-n^2}} + o\left(n \frac{2^{-n^2}}{1 + 2^{-n^2}}\right)} \quad (2.9)$$

Donc pour n assez grand, la chaîne X^n ne fait, $\mathbb{P}_x - p.s.$, pas de backtrack.

De plus, sachant qu'il n'y a pas de backtrack, S_{5n-i+1} suit une loi exponentielle de paramètre $\varepsilon_n + \varepsilon'_n + \tilde{\varepsilon}_n$, d'où

$$\mathbb{P}_x(S_{5n-i+1} > 6n) = e^{-6n(\varepsilon_n + \varepsilon'_n + \tilde{\varepsilon}_n)} \quad (2.10)$$

Ainsi grâce à (2.9) et (2.10), on peut dire que pour n assez grand, partant d'un sommet $x \in V_n$, la chaîne X^n ne se déplace qu'en direction de D tant que D n'a pas été atteint. Une fois D atteint, X^n y reste un temps exponentiellement long. On a donc $\{X^n \text{ ne backtrack pas avant d'avoir atteint } D\} = \{X^n \text{ ne backtrack pas avant un temps linéaire en } n\}$.

Autrement dit, pour tout $\eta > 0$, il existe $N \in \mathbb{N}$ tel que pour tout $n \geq N$

$$\mathbb{P}_x(\tau_n \leq t) \leq P_t^n(x, D) + \frac{\eta}{2} \quad \forall t < 6n \quad (2.11)$$

et puisqu'on a toujours $P_t^n(x, D) \leq \mathbb{P}_x(\tau_n \leq t)$ alors

$$|P_t^n(x, D) - \mathbb{P}_x(\tau_n \leq t)| \leq \frac{\eta}{2} \quad \forall t < 6n \quad (2.12)$$

Il nous faut maintenant montrer que $\mathbb{P}_A(\tau_n \leq t) = \min_{x \in V_n} \mathbb{P}_x(\tau_n \leq t)$.

Soit $X^{n,A}$ une chaîne de Markov de distribution initiale δ_A sur V_n et $X^{n,x}$ une chaîne de Markov de distribution initiale δ_x sur V_n , $x \in \{A_1, \dots, B\}$ fixé. Les deux chaînes sautent en même et si elle se croisent, alors elles restent ensemble pour toujours. $X^{n,A}$ ne pourra donc jamais dépasser $X^{n,x}$. Pour tout $t < 6n$, la loi de $X^{n,A}(t)$ est évidemment $P_t^n(A, \cdot)$ et celle de $X^{n,x}(t)$ est bien sûr $P_t^n(x, \cdot)$.

Ce couplage permet ainsi de voir facilement que

$$\mathbb{P}_A(\tau_n \leq t) \leq \mathbb{P}_x(\tau_n \leq t) \quad \forall x \in \{A_1, \dots, B\} \quad (2.13)$$

Fixons maintenant $x \in \{A_{3n+1}, \dots, A_{5n-1}\}$.

Rappelons que, sachant que X^n ne backtrack pas avant d'avoir atteint D , $S_1, (1 + \varepsilon_n)S_2, \dots, (1 +$

$\varepsilon_n)S_{5n}$ et $(1 + \varepsilon_n)S'_i$ pour tout i suivent une loi exponentielle de paramètre 1.

Par construction, partant de A et sachant que X^n ne backtrack pas avant un temps inférieur à $6n$, τ_n est minoré par $S_1 + \sum_{k=2}^{3n+1} S_k$ alors que partant de x et sachant que X^n ne backtrack pas avant un temps inférieur à $6n$, τ_n est majoré par $\sum_{k=1}^{2n} S'_k$ et donc

$$\mathbb{P}_A(\tau_n \leq t) \leq \mathbb{P}\left(S_1 + \sum_{k=2}^{3n+1} S_k \leq t\right) \leq \mathbb{P}\left(\sum_{k=1}^{2n} S'_k \leq t\right) \leq \mathbb{P}_x(\tau_n \leq t) \quad (2.14)$$

Les expressions (2.13) et (2.14) permettent de conclure que $\mathbb{P}_A(\tau_n \leq t) = \min_{x \in V_n} \mathbb{P}_x(\tau_n \leq t)$.

On a en particulier, en utilisant (2.12)

$$\left| \min_{x \in V_n} P_t^n(x, D) - \mathbb{P}_A(\tau_n \leq t) \right| \leq \frac{\eta}{2} \quad \forall t < 6n \quad (2.15)$$

Finalement

$$|\mathbb{P}_A(\tau_n > t) - d_n(t)| = \left| \min_{x \in V_n} P_t^n(x, D) - \mathbb{P}_A(\tau_n \leq t) + 1 - \min_{x \in V_n} P_t^n(x, D) - d_n(t) \right|$$

Par inégalité triangulaire puis en utilisant (2.15) et (2.7), on a, pour tout $\eta > 0$, il existe $N' \in \mathbb{N}$ tel que pour tout $n \geq N'$

$$|\mathbb{P}_A(\tau_n > t) - d_n(t)| \leq \left| \min_{x \in V_n} P_t^n(x, D) - \mathbb{P}_A(\tau_n \leq t) \right| + \left| 1 - \min_{x \in V_n} P_t^n(x, D) - d_n(t) \right| \leq \frac{\eta}{2} + \frac{\eta}{2} = \eta \quad \forall t < 6n \quad (2.16)$$

Maintenant, si $t = 6n$, nous savons par construction, que, partant de A et sachant que X^n ne backtrack pas avant un temps inférieur à $6n$, τ_n est majoré par $S_1 + \sum_{k=2}^{5n} S_k$ d'où, pour tout $\lambda \in (0, 1/6)$

$$\mathbb{P}_A(\tau_n > t) \leq \mathbb{P}\left(S_1 + \sum_{k=2}^{5n} S_k > 6n\right) = \mathbb{P}\left(e^{\lambda S_1 + \sum_{k=2}^{5n} \lambda S_k} > e^{\lambda 6n}\right) \leq e^{\psi(\lambda) - \lambda 6n} \quad (2.17)$$

où $\psi(\lambda) := \log \mathbb{E}\left(e^{\lambda S_1 + \sum_{k=2}^{5n} \lambda S_k}\right)$ et où la dernière inégalité est obtenue par inégalité de Markov.

Par indépendance des variables S_1, \dots, S_{5n} , on a

$$\log \mathbb{E}\left(e^{\lambda S_1 + \sum_{k=2}^{5n} \lambda S_k}\right) = -\log(1 - \lambda) - (5n - 1) \log\left(1 - \frac{\lambda}{1 + 2^{-n^2}}\right) \underset{n \rightarrow \infty}{\sim} -5n \log(1 - \lambda)$$

Mais l'application qui à λ associe $5 \log(1 - \lambda) + 6\lambda$ est positive sur $(0, 1/6)$ donc

$$\lim_{n \rightarrow \infty} e^{-n(5 \log(1-\lambda)+6\lambda)} = 0$$

et finalement d'après (2.17)

$$\lim_{n \rightarrow \infty} \mathbb{P}_A(\tau_n > t) = 0 \tag{2.18}$$

D'après (2.7), on a, pour n assez grand

$$d_n(t) = d_n(6n) = \max_{x \in V_n} (1 - P_{6n}^n(x, D)) - O(2^{-n^2})$$

Mais

$$1 - P_{6n}^n(x, D) = \mathbb{P}_x(X_{6n}^n \neq D) = \mathbb{P}_x(\tau_n > 6n)$$

car nous avons dit précédemment que pour n assez grand, la chaîne X^n ne pouvait pas avoir atteint D et avoir quitter D en un temps linéaire en n . Donc si $X_{6n}^n \neq D$, nécessairement D n'a pas été atteint au temps $6n$. Ainsi

$$\max_{x \in V_n} (1 - P_{6n}^n(x, D)) = 1 - \min_{x \in V_n} \mathbb{P}_x(\tau_n \leq 6n) = 1 - \mathbb{P}_A(\tau_n \leq 6n) = \mathbb{P}_A(\tau_n > 6n)$$

car nous avons montré que, si X^n ne backtrack pas avant le temps $6n$, $\mathbb{P}_A(\tau_n \leq t) = \min_{x \in V_n} \mathbb{P}_x(\tau_n \leq t)$ pour tout $t < 6n$ mais le résultat reste évidemment vrai si $t = 6n$.

Finalement, en utilisant (2.18)

$$\lim_{n \rightarrow \infty} d_n(6n) = 0 \tag{2.19}$$

Enfin, grâce à (2.18) et (2.19), on peut conclure que pour tout $\eta > 0$, il existe $\tilde{N} \in \mathbb{N}$ tel que pour tout $n \geq \tilde{N}$

$$\mathbb{P}_A(\tau_n > 6n) \leq \frac{\eta}{4} \quad \text{et} \quad d_n(6n) \leq \frac{\eta}{4}$$

d'où, par inégalité triangulaire et décroissance de d_n et de $\mathbb{P}_A(\tau_n > \cdot)$, pour tout $t > 6n$

$$|d_n(t) - \mathbb{P}_A(\tau_n > t)| \leq d_n(t) + \mathbb{P}_A(\tau_n > t) \leq d_n(6n) + \mathbb{P}_A(\tau_n > 6n) \leq \frac{\eta}{4} + \frac{\eta}{4} = \frac{\eta}{2} \tag{2.20}$$

et donc en combinant (2.16) et (2.20), on a, pour tout $\eta > 0$ et pour tout $n \geq \max(N', \tilde{N})$

$$|d_n(t) - \mathbb{P}_A(\tau_n > t)| \leq \frac{\eta}{2} + \frac{\eta}{2} = \eta \quad \forall t \in \mathbb{R}^+$$

ce qui achève la démonstration.

Preuve du théorème 2.1.2. Nous allons utiliser des notations plus commodes pour certains événements. Ainsi $CR = \{X^n \text{ empreinte le chemin rouge}\}$, $CV = \{\text{Arrivé en } C, X^n \text{ empreinte le chemin vert}\}$ et $CB = \{\text{Arrivé en } B, X^n \text{ empreinte le chemin bleu}\}$

Commençons par (i).

Soit $s > 0$. On a par construction

$$\mathbb{P}_A(\tau_n > 3ns) = \mathbb{P}_A(\tau_n > 3ns|CR) \frac{1}{4n} + \mathbb{P}_A(\tau_n > 3ns|CV) \frac{3}{4n} + \mathbb{P}_A(\tau_n > 3ns|CB) \frac{n-1}{n}$$

Mais puisque $\mathbb{P}_A(\tau_n > 3ns|CR) \leq 1$ et $\mathbb{P}_A(\tau_n > 3ns|CV) \leq 1$, il nous faut uniquement savoir comment $\mathbb{P}_A(\tau_n > 3ns|CB)$ se comporte quand n tend vers l'infini.

Par construction, nous savons que, conditionnellement aux événements CB et $\{X^n \text{ ne backtrack pas avant d'avoir atteint } D\}$ et partant de A , $\tau_n = S_1 + \sum_{k=2}^{3n+1} S_k$.

Si $s < 1$, alors pour tout $\lambda \in (\frac{s-1}{s}, 0)$

$$\mathbb{P}_A(\tau_n > 3ns|CB) = \mathbb{P}\left(S_1 + \sum_{k=2}^{3n+1} S_k > 3ns\right) = 1 - \mathbb{P}\left(e^{\lambda S_1 + \sum_{k=2}^{5n} \lambda S_k} > e^{\lambda 3ns}\right) \quad (2.21)$$

Mais par inégalité de Markov et par indépendance des variables S_1, \dots, S_{3n+1}

$$\mathbb{P}\left(e^{\lambda S_1 + \sum_{k=2}^{5n} \lambda S_k} > e^{\lambda 3ns}\right) \leq e^{-\log(1-\lambda) - 3n \log\left(1 - \frac{\lambda}{1+2^{-n^2}}\right) - \lambda 3ns}$$

et $-\log(1-\lambda) - 3n \log\left(1 - \frac{\lambda}{1+2^{-n^2}}\right) \underset{n \rightarrow \infty}{\sim} -(3n+1) \log(1-\lambda)$.

Mais l'application qui à λ associe $\log(1-\lambda) + \lambda s$ est positive sur $(\frac{s-1}{s}, 0)$ donc

$$\lim_{n \rightarrow \infty} e^{-3n(\log(1-\lambda) + \lambda s) - \log(1-\lambda)} = 0$$

et finalement, en passant à la limite dans (2.21), on obtient

$$\lim_{n \rightarrow \infty} \mathbb{P}_A(\tau_n > 3ns|CB) = 1 \quad (2.22)$$

Si $s > 1$, alors pour tout $\lambda \in (0, \frac{s-1}{s})$

$$\mathbb{P}_A(\tau_n > 3ns|CB) = \mathbb{P}\left(S_1 + \sum_{k=2}^{3n+1} S_k > 3ns\right) = \mathbb{P}\left(e^{\lambda S_1 + \sum_{k=2}^{5n} \lambda S_k} > e^{\lambda 3ns}\right) \quad (2.23)$$

Mais par inégalité de Markov et par indépendance des variables S_1, \dots, S_{3n+1}

$$\mathbb{P} \left(e^{\lambda S_1 + \sum_{k=2}^{5n} \lambda S_k} > e^{\lambda 3ns} \right) \leq e^{-\log(1-\lambda) - 3n \log \left(1 - \frac{\lambda}{1+2^{-n^2}} \right) - \lambda 3ns}$$

et $-\log(1-\lambda) - 3n \log \left(1 - \frac{\lambda}{1+2^{-n^2}} \right) \underset{n \rightarrow \infty}{\sim} -(3n+1) \log(1-\lambda)$.

Mais l'application qui à λ associe $\log(1-\lambda) + \lambda s$ est positive sur $(0, \frac{s-1}{s})$ donc

$$\lim_{n \rightarrow \infty} e^{-3n(\log(1-\lambda) + \lambda s) - \log(1-\lambda)} = 0$$

et finalement, en passant à la limite dans (2.23), on obtient

$$\lim_{n \rightarrow \infty} \mathbb{P}_A(\tau_n > 3ns | CB) = 0 \quad (2.24)$$

Ainsi, en utilisant le (i) du lemme 2.2.1 puis les relations (2.22) et (2.24) on obtient

$$\lim_{n \rightarrow \infty} d_n(3ns) = \begin{cases} 1 & \text{si } s < 1 \\ 0 & \text{si } s > 1 \end{cases}$$

ce qui prouve que la suite $(X_n)_{n \in \mathbb{N}}$ présente un cutoff vers le temps $3n$.

Démontrons maintenant (ii).

Soit $s > 0$. On a par construction

$$n\mathbb{P}_A(\tau_n > 3ns) = \mathbb{P}_A(\tau_n > 3ns | CR) \frac{1}{4} + \mathbb{P}_A(\tau_n > 3ns | CV) \frac{3}{4} + \mathbb{P}_A(\tau_n > 3ns | CB)(n-1) \quad (2.25)$$

En conditionnant par le fait que X^n ne backtrack pas avant d'avoir atteint D , sachant CR et partant de A , $\tau_n = S_1 + \sum_{k=2}^{5n} S_k$, sachant CV et partant de A , $\tau_n = S_1 + \sum_{k=2}^{4n+1} S_k$ et sachant CB et partant de A , $\tau_n = S_1 + \sum_{k=2}^{3n+1} S_k$.

Donc en procédant exactement comme dans la preuve de (i), on a

Si $s < \frac{5}{3}$, en prenant cette fois $\lambda \in (\frac{3s-5}{3s}, 0)$ de sorte que $5 \log(1-\lambda) + 3\lambda s > 0$, alors

$$\lim_{n \rightarrow \infty} \mathbb{P}_A(\tau_n > 3ns | CR) = 1$$

Si $s > \frac{5}{3}$, en prenant cette fois $\lambda \in (0, \frac{3s-5}{3s})$ de sorte que $5 \log(1-\lambda) + 3\lambda s > 0$, alors

$$\lim_{n \rightarrow \infty} \mathbb{P}_A(\tau_n > 3ns | CR) = 0$$

Si $s < \frac{4}{3}$, en prenant cette fois $\lambda \in (\frac{3s-4}{3s}, 0)$ de sorte que $4 \log(1-\lambda) + 3\lambda s > 0$, alors

$$\lim_{n \rightarrow \infty} \mathbb{P}_A(\tau_n > 3ns | CV) = 1$$

Si $s > \frac{4}{3}$, en prenant cette fois $\lambda \in (0, \frac{3s-4}{3s})$ de sorte que $4 \log(1 - \lambda) + 3\lambda s > 0$, alors

$$\lim_{n \rightarrow \infty} \mathbb{P}_A(\tau_n > 3ns | CV) = 0$$

Si $s < 1$, en prenant cette fois $\lambda \in (\frac{s-1}{s}, 0)$ de sorte que $\log(1 - \lambda) + \lambda s > 0$, alors

$$\lim_{n \rightarrow \infty} \mathbb{P}_A(\tau_n > 3ns | CB) = 1$$

Si $s > 1$, en prenant cette fois $\lambda \in (0, \frac{s-1}{s})$ de sorte que $\log(1 - \lambda) + \lambda s > 0$, alors

$$\lim_{n \rightarrow \infty} \mathbb{P}_A(\tau_n > 3ns | CB)(n-1) = 0$$

car une décroissance exponentielle l'emporte sur une croissance linéaire.

Donc en passant à la limite dans (2.25)

$$\lim_{n \rightarrow \infty} n \mathbb{P}_A(\tau_n > 3ns) = \begin{cases} \infty & \text{si } s < 1 \\ \frac{1}{4} + \frac{3}{4} & \text{si } 1 < s < \frac{4}{3} \\ \frac{1}{4} & \text{si } \frac{4}{3} < s < \frac{5}{3} \\ 0 & \text{si } s > \frac{5}{3} \end{cases} \quad (2.26)$$

A présent, si on note $\tau_{n,i} := \min\{t \geq 0 : X_i^n(t) = D\}$, alors $\tau_{n,i}$ et τ_n ont évidemment la même loi et on a

$$\{\mathcal{T}_n \leq 3ns\} = \bigcap_{i=1}^n \{\tau_{n,i} \leq 3ns\}$$

donc par indépendance des variables aléatoires $X_i^n, i = 1, \dots, n$, on a

$$\mathbb{Q}_A(\mathcal{T}_n > 3ns) = 1 - (1 - \mathbb{P}_A(\tau_n > 3ns))^n \quad (2.27)$$

Si $s < 1$, on a, par concavité de la fonction log

$$(1 - \mathbb{P}_A(\tau_n > 3ns))^n = e^{n \log(1 - \mathbb{P}_A(\tau_n > 3ns))} \leq e^{-n \mathbb{P}_A(\tau_n > 3ns)} \xrightarrow[n \rightarrow \infty]{} 0$$

Si $s > 1$ alors d'après le (i) du théorème 2.1.2, $\lim_{n \rightarrow \infty} \mathbb{P}_A(\tau_n > 3ns) = 0$ ce qui nous permet d'écrire

$$(1 - \mathbb{P}_A(\tau_n > 3ns))^n = e^{n \log(1 - \mathbb{P}_A(\tau_n > 3ns))} = e^{-n \mathbb{P}_A(\tau_n > 3ns) + o(1)}$$

Ainsi en utilisant (2.27) puis le (ii) du lemme 2.2.1 on obtient

$$\lim_{n \rightarrow \infty} D_n(3ns) = \begin{cases} 1 & \text{si } s < 1 \\ 1 - e^{-1} & \text{si } 1 < s < \frac{4}{3} \\ 1 - e^{-1/4} & \text{si } \frac{4}{3} < s < \frac{5}{3} \\ 0 & \text{si } s > \frac{5}{3} \end{cases}$$

en donc la suite $(Y^n)_{n \in \mathbb{N}}$ ne présente pas de cutoff et le théorème 2.1.2 est démontré.

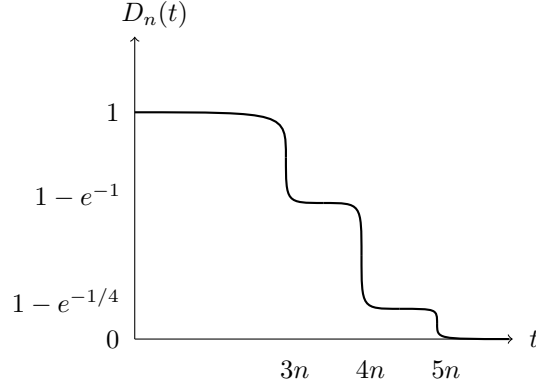


FIGURE 2.3 – Graphe de $\lim_{n \rightarrow \infty} D_n(t)$. On obtient 2 paliers distincts entre 0 et 1.

Remarque 2.2.2. Notons l_n le nombre d'arêtes constituant le chemin rouge et t_n le nombre d'arêtes séparant A du premier point à partir duquel A_{l_n} peut-être atteint sans emprunter le chemin rouge. Dans notre exemple, $l_n = 5n$, $A_{l_n} = D$ et $t_n = 3n$, $A_{t_n} = B$.

Il est important que $l_n = ct_n$, $c > 0$. En effet si ce n'était pas le cas, la suite $(Y^n)_{n \in \mathbb{N}}$ pourrait par "malchance" présenter un cutoff. Si par exemple $l_n = n^2 + n$, c'est-à-dire que le graphe G_n a $n^2 + n + 1$ sommets et $n^2 + n$ arêtes rouges. On ajoute une arête supplémentaire reliant directement les points A_{n^2} et A_{n^2+n} (le dernier point du chemin rouge). On a bien sûr $t_n = n^2$. G_n a donc $n^2 + n + 1$ sommets et $n^2 + n + 1$ arêtes. Alors les suites $(X^n)_{n \in \mathbb{N}}$ et $(Y^n)_{n \in \mathbb{N}}$ présente un cutoff au temps $t_n = n^2$.

Remarque 2.2.3. On voit dans la preuve du théorème 2.1.2 à quel point le lemme 2.2.1 est important. Mais la clé de la preuve de ce lemme est la relation suivante :

$$\min_{x \in V_n} \mathbb{P}_x(\tau_n \leq t) = \mathbb{P}_A(\tau_n \leq t)$$

A est en fait le point du graphe maximisant le temps d'atteinte de D .

Posons $l_n = a_1 n$ et $t_n = a_2 n$ $a_1 > 0$ de sorte que la suite $(Y^n)_{n \in \mathbb{N}}$ ne présente pas de cutoff d'après la remarque 2.2.2. Compte tenu de la structure du graphe, A est le point du graphe maximisant le temps d'atteinte de A_{l_n} (D dans notre exemple) si et seulement si $t_n \geq l_n - t_n$ c'est-à-dire

$$\frac{a_1}{a_2} \leq 2 \tag{2.28}$$

Dans notre exemple, $\frac{a_1}{a_2} = \frac{5}{3}$ et dans celui de Lacoïn, $\frac{a_1}{a_2} = 2$.

Si (2.28) est vérifiée on a en fait pour tout $a \in (0, 1)$

$$\limsup_{n \rightarrow \infty} \frac{T_{mix}^n(a)}{T_{mix}^n(1-a)} \leq \frac{a_1}{a_2}$$

Ainsi pour toute suite de graphe $(G_n)_{n \in \mathbb{N}}$ ayant la même structure que précédemment et telle que (2.28) soit vérifiée, le théorème 1.4.18 est vérifié.

Remarque 2.2.4. Le théorème 2.1.2 reste vrai pour la distance de séparation. L'idée est de montrer que pour n assez grand, la différence entre la distance en variation totale et la distance de séparation est négligeable.

Lemme 2.2.5. Soit \mathbb{M} une mesure de probabilité sur $(\mathcal{X}, \mathcal{F})$, $(E_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires sur \mathcal{X} iid suivant une loi exponentielle de paramètre 1 et \mathcal{P}_a une variable aléatoire sur \mathcal{X} suivant une loi de Poisson de paramètre $a \geq 0$. Alors pour tout entier $j \geq 1$

$$\mathbb{M} \left(\sum_{k=1}^j E_k \leq a \right) = \mathbb{M}(\mathcal{P}_a \geq j)$$

Preuve. On a par définition

$$\frac{d}{da} \mathbb{M} \left(\sum_{k=1}^j E_k \leq a \right) = \frac{a^{j-1} e^{-a}}{\Gamma(j)}$$

De plus,

$$\frac{d}{da} \mathbb{M}(\mathcal{P}_a \geq j) = \frac{d}{da} \sum_{k \geq j} \frac{a^k}{k!} e^{-a} = \sum_{k \geq j-1} \frac{a^k}{k!} e^{-a} - \sum_{k \geq j} \frac{a^k}{k!} e^{-a}$$

d'où

$$\frac{d}{da} \mathbb{M}(\mathcal{P}_a \geq j) = \frac{a^{j-1} e^{-a}}{(j-1)!} = \frac{a^{j-1} e^{-a}}{\Gamma(j)}$$

donc

$$\mathbb{M} \left(\sum_{k=1}^j E_k \leq a \right) = \mathbb{M}(\mathcal{P}_a \geq j) + c$$

Mais par convention, \mathcal{P}_0 est la fonction nulle sur \mathcal{X} , $E_k \geq 0$ \mathbb{M} -p.s et $j \geq 1$ donc

$$c = 0$$

ce qui achève la preuve.

Lemme 2.2.6. *Pour n assez grand, pour tout $x, y \in V_n \setminus \{D\}$ et pour tout $\alpha n \leq t \leq \beta n$ avec $0 < \alpha < 3 < \beta$, on a*

$$P_t^n(x, y) \geq \pi_n(y)$$

Preuve. Commençons par le cas où $x = A_1$. Soit $y \in V_n \setminus \{A, D\}$ fixé. Alors il existe $d \in \{1, \dots, 5n - 2\}$ tel que $y = A_{d+1}$ (c'est-à-dire qu'il y a d arêtes rouges séparant x et y).

Nous avons

$$P_t^n(x, y) \geq \mathbb{P}_x(A_{d,[0,t]}) \tag{2.29}$$

où $\{A_{d,[0,t]}\} := \{X^n \text{ fait exactement } d \text{ sauts dans l'intervalle de temps } [0, t] \text{ entre } x \text{ et } y\}$

$$\mathbb{P}_x(A_{d,[0,t]}) = \mathbb{P}_x(X_{J_1} = A_2, \dots, X_{J_d} = y \text{ et } J_d \leq t < J_{d+1})$$

Posons $\mathbb{M} := \mathbb{P}_x(\cdot | X_{J_1} = A_2, \dots, X_{J_d} = y)$.

Si $d \leq 3n - 1$ alors en utilisant la propriété de Markov simple, on a

$$\mathbb{P}_x(A_{d,[0,t]}) = \mathbb{M}(J_d \leq t < J_{d+1}) \left(\frac{1}{1 + 2^{-n^2}} \right)^d$$

Mais rappelons que $\sum_{k=1}^N S_k = J_N$ et que sous \mathbb{M} , les variables aléatoires $(1 + 2^{-n^2})S_1, \dots, (1 + 2^{-n^2})S_N$ sont *iid* et suivent une loi exponentielle de paramètre 1. Ainsi d'après le lemme 2.2.5, on a

$$\mathbb{M}(J_d \leq t < J_{d+1}) = \mathbb{M}\left(\sum_{k=1}^d S_k \leq t < \sum_{k=1}^{d+1} S_k\right) = \frac{\left(\left(1 + 2^{-n^2}\right)t\right)^d}{d!} e^{-t(1+2^{-n^2})}$$

et

$$\frac{\left(\left(1 + 2^{-n^2}\right)t\right)^d}{d!} e^{-t(1+2^{-n^2})} \underset{n \rightarrow \infty}{\sim} \frac{t^d}{d!} e^{-t}$$

car $\alpha n \leq t \leq \beta n$ et au pire, d dépend linéairement de n et est donc battu par 2^{-n^2} en passant à la limite.

De plus pour n assez grand

$$\frac{t^d}{d!}e^{-t} \geq \frac{(\alpha n)^d}{d!}e^{-\beta n} \geq e^{-\beta n}$$

Si $3n - 1 < d \leq 4n - 1$ alors de la même façon, on a

$$\mathbb{P}_x(A_{d,[0,t]}) = n^{-1}\mathbb{M}(J_d \leq t < J_{d+1}) \left(\frac{1}{1+2^{-n^2}} \right)^{d-1}$$

et

$$n^{-1}\mathbb{M}(J_d \leq t < J_{d+1}) \underset{n \rightarrow \infty}{\sim} n^{-1} \frac{t^d}{d!} e^{-t}$$

De plus pour n assez grand

$$n^{-1} \frac{t^d}{d!} e^{-t} \geq n^{-1} \frac{(\alpha n)^d}{d!} e^{-\beta n} \geq e^{-\beta n - \log n} \geq e^{-n(\beta+1)}$$

Si $4n - 1 < d \leq 5n - 2$ alors de la même façon, on a

$$\mathbb{P}_x(A_{d,[0,t]}) = \frac{1}{4n}\mathbb{M}(J_d \leq t < J_{d+1}) \left(\frac{1}{1+2^{-n^2}} \right)^{d-1}$$

et

$$\frac{1}{4n}\mathbb{M}(J_d \leq t < J_{d+1}) \underset{n \rightarrow \infty}{\sim} \frac{1}{4n} \frac{t^d}{d!} e^{-t}$$

De plus pour n assez grand

$$\frac{1}{4n} \frac{t^d}{d!} e^{-t} \geq \frac{1}{4n} \frac{(\alpha n)^d}{d!} e^{-\beta n} \geq e^{-\beta n - \log 4n} \geq e^{-n(\beta+4)}$$

donc d'après (2.29), on a, pour n assez grand

$$P_t^n(x, y) \geq \max(e^{-\beta n}, e^{-n(\beta+1)}, e^{-n(\beta+4)}) = e^{-\beta n}$$

Mais nous savons que pour tout $y \in V_n \setminus \{A, D\}$, il existe $K > 0$ tel que

$$\pi_n(y) \leq K2^{-n^2}$$

Finalement, nous avons montré que pour tout $y \in V_n \setminus \{A, D\}$

$$P_t^n(A_1, y) \geq \pi_n(y)$$

Il semble évident que A_1 peut-être remplacé par x quelconque différent de A et de D tant que x appartient au segment $[A_1, y]$. Si ce n'est pas la cas, alors pour $x \in V_n \setminus \{A, D\}$ fixé nous montrons de manière analogue que tout $y \in V_n \setminus \{A, D\}$

$$P_t^n(y, x) \geq \pi_n(x)$$

et par réversibilité de la chaîne X^n , on obtient le résultat voulu.

Le cas où x et y peuvent valoir A est à traiter à part car le temps passé en A suit une loi exponentielle de paramètre 1 et non pas de paramètre $(1 + 2^{-n^2})$.

L'idée est la suivante :

Si d est le nombre d'arêtes rouges entre A et y , alors $S_1, (1 + 2^{-n^2})S_2, \dots, (1 + 2^{-n^2})S_{d+1}$ suivent une loi exponentielle de paramètre A et par inégalité de Markov

$$\mathbb{M}(J_d \leq t < J_{d+1}) \geq 1 - e^{\varphi_d(\lambda)} - e^{\varphi_{d+1}(\lambda')}$$

où

$$\varphi_N(\gamma) = \log \mathbb{E}_{\mathbb{M}} \left(e^{\gamma S_1 + \sum_{k=1}^N \gamma S_k} \right) - \gamma t$$

et $\lambda < 0$ et $\lambda' > 0$ sont à ajuster de sorte que $\varphi_d(\lambda)$ et $\varphi_{d+1}(\lambda')$ puissent s'écrire respectivement sous la forme $-nf(\lambda)$ et $-ng(\lambda')$ avec $f(\lambda), g(\lambda') > 0$, ce qui est possible comme nous l'avons déjà vu à plusieurs reprises (voir la preuve du théorème 2.1.2 pour un exemple).

Concluons maintenant l'explication de la remarque 2.2.4.

D'après le lemme 2.2.6, puisque pour tout x , $P_t^n(x, \cdot)$ et π_n sont des mesures de probabilité, on a nécessairement pour n assez grand et pour tout $\alpha n \leq t \leq \beta n$

$$P_t^n(x, D) < \pi_n(D)$$

Autrement dit

$$\min_{y \in V_n} \frac{P_t^n(x, y)}{\pi_n(y)} = \frac{P_t^n(x, D)}{\pi_n(D)}$$

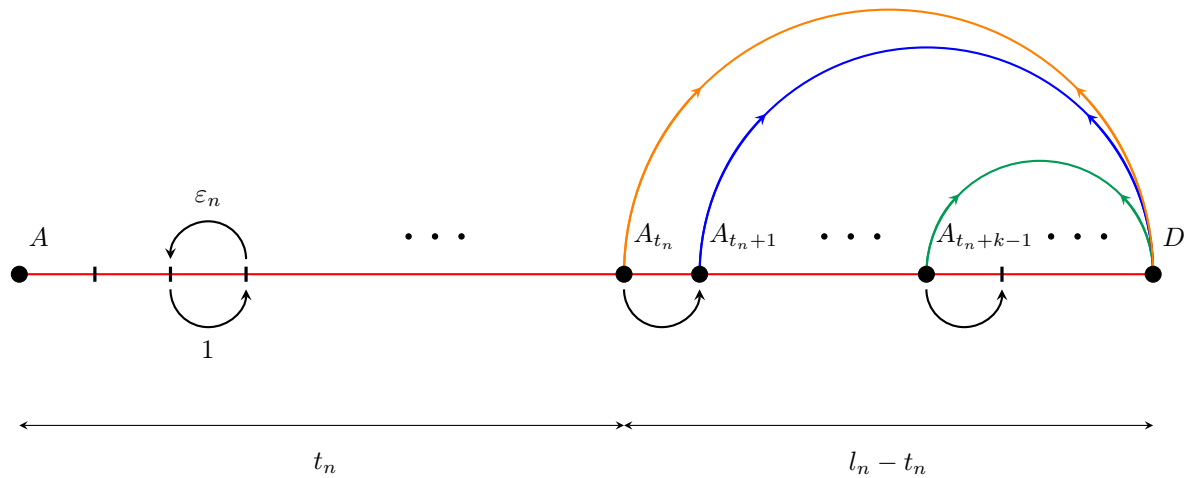
d'où

$$d_n^s(t) = 1 - \min_{x \in V_n} \frac{P_t^n(x, D)}{\pi_n(D)}$$

et comme $\pi_n(D) = 1 - O(2^{-n^2})$ alors selon (2.7), la différence entre la distance en variation totale et la distance de séparation est négligeable.

Finalement, nous venons de montrer qu'en ajoutant une arête allant directement vers D au graphe d'origine de Lacoïn et en choisissant judicieusement le nombre d'arêtes du graphe (voir la remarque 2.2.3) et les taux de saut, nous ajoutons un palier supplémentaire entre 0 et 1 au graphe

de la distance à l'équilibre. On a ainsi exhibé un exemple pour lequel la condition produit n'est pas suffisante et pour lequel la géométrie du graphe influence beaucoup le comportement de la distance à l'équilibre. On peut même se permettre de choisir le nombre de paliers entre 0 et 1 :



En considérant le graphe G'_n ci-dessus, en choisissant bien les taux de saut et en prenant en compte la remarque 2.2.3, la suite $(X_n)_{n \in \mathbb{N}}$ présentera un cutoff au temps t_n tandis que la suite $(Y_n)_{n \in \mathbb{N}}$, elle, n'en présentera pas. Enfin, D_n aura exactement k paliers entre 0 et 1.

Partie 3

Quelques résultats positifs à propos de la condition produit

3.1 Valeurs propres et noyau de transition

Comme nous l'avons dit dans l'introduction, la condition produit fait intervenir les valeurs propres de la matrice de transition d'une chaîne de Markov sur un espace fini. Nous allons donc à présent nous intéresser aux propriétés spectrales d'un tel opérateur.

Voici d'abord quelques définitions très importantes :

Définition 3.1.1. Soit $X = (X(t))_{t \in \mathbb{N}}$ une chaîne de Markov réversible, irréductible d'espace d'état Ω fini, de matrice de transition P et de mesure invariante π . Soit $f \in \mathbb{R}^\Omega$. On définit

$$\mathbb{E}_\pi(f) := \int_\Omega f(x)\pi(dx) \quad \text{et} \quad \text{Var}_\pi(f) := \mathbb{E}_\pi[(f - \mathbb{E}_\pi(f))^2]$$

Soit $1 \leq p < \infty$. Pour tout $f, g \in \mathbb{R}^\Omega$, on définit

$$\langle f, g \rangle_\pi := \mathbb{E}_\pi(fg) \quad \text{et} \quad \|f\|_p := (\mathbb{E}_\pi[|f|^p])^{1/p}$$

$\langle \cdot, \cdot \rangle_\pi$ est un produit scalaire sur \mathbb{R}^Ω et $\|\cdot\|_p$ est une norme sur \mathbb{R}^Ω .

Posons $\mathcal{L}^p := \{f \in \mathbb{R}^\Omega : \|f\|_p < \infty\}$ et $L^p := \{[f] : f \in \mathcal{L}^p\}$ où $[f] := \{g \in \mathbb{R}^\Omega : g = f \pi\text{-p.s}\}$

Ainsi défini, $(L^p, \|\cdot\|_p)$ est un espace vectoriel normé et $(L^2, \langle \cdot, \cdot \rangle_\pi)$ est un espace de Hilbert.

Rappelons que P_t est l'opérateur linéaire de L^p à valeurs dans L^p tel que $P_t f(x) = \sum_{y \in \Omega} P_t(x, y) f(y)$.

On a $P_1 = P$.

Voyons dans un premier temps comment, grâce aux valeurs propres de P_t , le temps de relaxation t_{rel} apparaît naturellement.

Par réversibilité de X , on a immédiatement $\langle P_t f, g \rangle_\pi = \langle f, P_t g \rangle_\pi$ pour tout $f, g \in L^2$, pour tout entier $t \geq 1$. Autrement dit, P_t est auto-adjoint et est donc diagonalisable dans une base orthonormée $(f_1, \dots, f_{|\Omega|})$ de vecteurs propres, $f_i \in L^2$. Notons $P f_i := \lambda_i f_i$ (c'est-à-dire $P_t f_i := \lambda_i^t f_i$) et ordonnons ses valeurs propres (qui sont toutes réelles) ainsi : $\lambda_1 > \lambda_2 \geq \dots \geq \lambda_{|\Omega|} > -1$. Alors $f_1 \equiv 1$, $\lambda_1 = 1$ et son sous-espace propre est de dimension 1, d'où l'inégalité stricte.

Soit $f \in \mathbb{R}^\Omega$. Alors on a bien sûr $f = \sum_{i=1}^{|\Omega|} \langle f, f_i \rangle_\pi f_i = \mathbb{E}_\pi(f) f_1 + \sum_{i=2}^{|\Omega|} \langle f, f_i \rangle_\pi f_i$. Donc par linéarité de P_t , on a

$$P_t f = \mathbb{E}_\pi(f) + \sum_{i=2}^{|\Omega|} \langle f, f_i \rangle_\pi \lambda_i^t f_i$$

Il vient immédiatement que $\|P_t f - \mathbb{E}_\pi(f)\|_2^2 \leq \lambda^{2t} \|f - \mathbb{E}_\pi(f)\|_2^2 = (1 - \lambda_\star)^{2t} \text{Var}_\pi(f)$ où $\lambda := \max_{i \geq 2} |\lambda_i|$. $\lambda_\star := 1 - \lambda$ est appelé le trou spectral de X . Finalement en posant $t_{\text{rel}} := \frac{1}{\lambda_\star}$ (le fameux temps de relaxation), on obtient

$$\sqrt{\text{Var}_\pi(P_t f)} = \|P_t f - \mathbb{E}_\pi(f)\|_2 \leq e^{-t/t_{\text{rel}}} \sqrt{\text{Var}_\pi(f)}$$

On savait déjà que $P_t f \rightarrow \mathbb{E}_\pi(f)$ [5]. Mais t_{rel} apparaît comme le temps à partir duquel la variance de $P_t f$ est majorée par e^{-1} fois la variance de f .

Définition 3.1.2. Soit $X = (X(t))_{t \in \mathbb{N}}$ une chaîne de Markov irréductible, réversible d'espace d'état Ω et de matrice de transition P . On dira de X qu'elle est lazy, littéralement feignante, si elle peut rester sur un état x avec une probabilité supérieure à $1/2$, pour tout $x \in \Omega$. Autrement dit, pour tout $x \in \Omega$, $P(x, x) \geq 1/2$.

Remarque 3.1.3. Supposons que X est lazy et notons P pour matrice de transition et λ_i les valeurs propres de P , $i = 1, \dots, |\Omega|$. Alors pour tout $i = 1, \dots, |\Omega|$, $\lambda_i > 0$. En effet, puisque la chaîne est lazy, il existe une matrice stochastique \tilde{P} irréductible telle que $P = (\tilde{P} + I)/2$. Donc si \tilde{f}_i est un vecteur propre de \tilde{P} associé à $\tilde{\lambda}_i$, alors \tilde{f}_i est un vecteur propre de P associé à la valeur propre $\lambda_i = (1 + \tilde{\lambda}_i)/2$. Puisque pour tout i , $|\tilde{\lambda}_i| < 1$ alors $\lambda_i > 0$. Ainsi, $\lambda := \max_{i \geq 2} |\lambda_i| = \lambda_2$ et le trou spectral devient $\lambda_\star := 1 - \lambda_2$. Le temps de relaxation devient alors $t_{\text{rel}} := \frac{1}{\lambda_\star} = \frac{1}{1 - \lambda_2}$.

Posons pour tout $f \in \mathbb{R}^\Omega$, $\mathcal{E}(f) := \langle (I - P)f, f \rangle_\pi$.

Proposition 3.1.4. Soit $X = (X(t))_{t \in \mathbb{N}}$ une chaîne de Markov réversible, irréductible d'espace d'état Ω fini, de matrice de transition P et de mesure invariante π . Notons λ_2 la plus grande valeur propre de P différente de 1. Alors

$$1 - \lambda_2 = \min_{\substack{f \in \mathbb{R}^\Omega \\ f \perp_\pi f_1; \|f\|_2=1}} \mathcal{E}(f) = \min_{\substack{f \in \mathbb{R}^\Omega \\ f \perp_\pi f_1; f \neq 0}} \frac{\mathcal{E}(f)}{\|f\|_2^2} \quad (3.1)$$

De plus

$$1 - \lambda_2 = \min_{\substack{f \in \mathbb{R}^\Omega \\ \text{Var}_\pi(f) \neq 0}} \frac{\mathcal{E}(f)}{\text{Var}_\pi(f)} \quad (3.2)$$

Preuve. Commençons par (3.1). Remarquons que la deuxième égalité découle directement de la première. En effet, pour tout $f \in \mathbb{R}^\Omega$, $\tilde{f} := f/\|f\|_2$ est de norme 1 et $\mathcal{E}(\tilde{f}) = \mathcal{E}(f)/\|f\|_2^2$.

En gardant les mêmes notations que précédemment, on a $(I-P)f_2 = (1-\lambda_2)f_2$ donc $\mathcal{E}(f_2) = 1-\lambda_2$. Toujours par définition, $f_2 \perp f_1$ et $\|f_2\|_2 = 1$. Ainsi, le minimum est bien atteint.

Soit $f \in \mathbb{R}^\Omega$ telle que $f \perp f_1$ et $\|f\|_2 = 1$. Rappelons que f_1 est la fonction constante égale à 1 sur Ω donc $f \perp_\pi f_1$ signifie que $\mathbb{E}_\pi(f) = 0$, d'où $f = \sum_{i=2}^{|\Omega|} \langle f, f_i \rangle_\pi f_i$. D'après le théorème de Pythagore,

$\|f\|_2 = 1$ c'est-à-dire $\sum_{i=2}^{|\Omega|} \langle f, f_i \rangle_\pi^2 = 1$. Nous avons alors

$$\mathcal{E}(f) = \sum_{i=2}^{|\Omega|} \langle f, f_i \rangle_\pi^2 \mathcal{E}(f_i) = \sum_{i=2}^{|\Omega|} \langle f, f_i \rangle_\pi^2 (1 - \lambda_i) \geq 1 - \lambda_2$$

Terminons par (3.2). On a pour toute constance c et pour tout $f \in \mathbb{R}^\Omega$, $\mathcal{E}(f+c) = \mathcal{E}(f)$. En particulier, $\mathcal{E}(f - \mathbb{E}_\pi(f)) = \mathcal{E}(f)$. Donc d'après (3.1)

$$1 - \lambda_2 = \min_{\substack{f \in \mathbb{R}^\Omega \\ f \perp_\pi f_1; f \neq 0}} \frac{\mathcal{E}(f)}{\|f\|_2^2} = \min_{\substack{f \in \mathbb{R}^\Omega \\ f \neq \mathbb{E}_\pi(f)}} \frac{\mathcal{E}(f - \mathbb{E}_\pi(f))}{\|f - \mathbb{E}_\pi(f)\|_2^2} = \min_{\substack{f \in \mathbb{R}^\Omega \\ \text{Var}_\pi(f) \neq 0}} \frac{\mathcal{E}(f)}{\text{Var}_\pi(f)}$$

Ce qui termine la preuve.

Remarque 3.1.5. Si la chaîne est lazy, alors selon la remarque 3.1.3, on peut remplacer $1 - \lambda_2$ par $1/t_{\text{rel}}$ dans la proposition 3.1.4. Si la chaîne n'est pas lazy, on pourra néanmoins se contenter de l'inégalité suivante : pour toute fonction $f \in \mathbb{R}^\Omega$ telle que $\text{Var}_\pi(f) \neq 0$,

$$\frac{1}{t_{\text{rel}}} \leq \frac{\mathcal{E}(f)}{\text{Var}_\pi(f)} \quad (3.3)$$

Proposition 3.1.6. Soit $X = (X(t))_{t \in \mathbb{N}}$ une chaîne de Markov réversible, irréductible d'espace d'état Ω fini, de matrice de transition P et de mesure invariante π . Notons λ_2 la plus grande valeur propre de P différente de 1. Alors

$$\frac{1}{1 - \lambda_2} \geq \frac{|\Omega| - 1}{|\Omega|}$$

En particulier, si $\tilde{X} = (\tilde{X}(t))_{t \in \mathbb{N}}$ une chaîne de Markov réversible, irréductible, lazy d'espace d'état Ω fini, de matrice de transition \tilde{P} et de mesure invariante $\tilde{\pi}$ alors en notant $\tilde{\lambda}_2$ la plus grande valeur propre de \tilde{P} différente de 1

$$\frac{1}{1 - (2\tilde{\lambda}_2 - 1)} \geq \frac{|\Omega| - 1}{|\Omega|}$$

c'est-à-dire

$$\tilde{t}_{rel} \geq 2 \frac{|\Omega| - 1}{|\Omega|} \quad (3.4)$$

On pourra consulter [2, Proposition 3.8 p. 80] pour une preuve.

Nous pouvons définir à partir de la mesure de probabilité π une mesure "conditionnelle" : Soit $A \subset \Omega$ fixé. Pour tout $E \subset \Omega$, $\pi_A(E) := \frac{\pi(A \cap E)}{\pi(A)}$.

Proposition 3.1.7. (Basu, Hermon et Peres, 2017)

Soit $X = (X(t))_{t \in \mathbb{N}}$ une chaîne de Markov réversible, irréductible d'espace d'état Ω fini, de matrice de transition P et de mesure invariante π . Soit $A \subsetneq \Omega$ non vide et $B := \Omega \setminus A$. Notons $k = |B|$. On définit P_B comme la matrice de transition de la sous-chaîne de X restreinte à l'ensemble B . Supposons de plus que cette sous-chaîne est irréductible.

(i) P_B a k valeurs propres réelles telles que $1 - \frac{\pi(A)}{t_{rel}} \geq \gamma_1 > \gamma_2 \geq \dots \geq \gamma_k \geq -\gamma_1$

(ii) Il existe $a_1, \dots, a_k \geq 0$ vérifiant $\sum_{i=1}^k a_i = 1$ tels que pour tout $t \geq 0$

$$\mathbb{P}_{\pi_B}(T_A > t) = \sum_{i=1}^k a_i \gamma_i^t \quad (3.5)$$

(iii) Pour tout $t \geq 0$

$$\mathbb{P}_{\pi_B}(T_A > t) \leq \left(1 - \frac{\pi(A)}{t_{rel}}\right)^t \leq e^{-t \frac{\pi(A)}{t_{rel}}} \quad (3.6)$$

Preuve. Commençons par (i). On peut définir sur \mathbb{R}^B le produit scalaire $\langle f, g \rangle_{\pi_B} := \sum_{x \in B} f(x)g(x)\pi_B(x)$.

P_B est auto-adjoint pour ce produit scalaire par réversibilité de la chaîne X donc diagonalisable dans une base orthonormée (g_1, \dots, g_k) de vecteurs propres, $g_i \in L^2$. Notons $P_B g_i := \gamma_i g_i$ et ordonnons ses valeurs propres (qui sont toutes réelles) ainsi : $\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_k$. Comme nous l'avons dit plus haut, P_B est irréductible et a bien-sûr ses coefficients positifs. Dans ces conditions, le théorème de Perron-Frobenius s'applique et nous dit que le rayon spectral de P_B , $\rho_B := \max_{1 \leq i \leq k} |\gamma_i|$, est une valeur propre de P_B dont le sous-espace propre est de dimension 1. Ainsi $\rho = \gamma_1 > 0$ et $\gamma_1 \geq |\gamma_k| \geq -\gamma_k$. Finalement, $\gamma_1 > \gamma_2 \geq \dots \geq \gamma_k \geq -\gamma_1$.

On peut montrer, de manière analogue à la preuve de la proposition 3.1.4, que

$$1 - \gamma_1 = \min_{\substack{f=f|_B \mathbf{1}_B; f \in \mathbb{R}^\Omega \setminus \{0\} \\ \|f\|_2=1}} \mathcal{E}(f) = \min_{f=f|_B \mathbf{1}_B; f \in \mathbb{R}^\Omega \setminus \{0\}} \frac{\mathcal{E}(f)}{\|f\|_2^2} \quad (3.7)$$

Nous pouvons aussi démontrer (3.7) en remarquant que $1 - \gamma_1$ est la plus petite valeur propre de la matrice $I - P_B$ et en appliquant le théorème min-max de Courant-Fischer :

$$1 - \gamma_1 = \min_{G \in F} \max_{g \in G; g \neq 0} \frac{\langle (I - P_B)g, g \rangle_{\pi_B}}{\langle g, g \rangle_{\pi_B}}$$

où $F := \{G \text{ sev de } \mathbb{R}^B : G = \text{Vect}(f), f \in \mathbb{R}^B\}$.

Mais comme pour tout $\alpha \in \mathbb{R}$ et pour tout $f \in \mathbb{R}^B$, $\frac{\langle (I - P_B)\alpha f, \alpha f \rangle_{\pi_B}}{\langle \alpha f, \alpha f \rangle_{\pi_B}} = \frac{\langle (I - P)f \mathbf{1}_B, f \rangle_{\pi}}{\langle f \mathbf{1}_B, f \mathbf{1}_B \rangle_{\pi}}$, on a

$$1 - \gamma_1 = \min_{f=f|_B \mathbf{1}_B; f \in \mathbb{R}^\Omega \setminus \{0\}} \frac{\mathcal{E}(f)}{\|f\|_2^2}$$

Soit $f \in \mathbb{R}^\Omega \setminus \{0\}$ telle que $f = f|_B \mathbf{1}_B$. En notant $\mathbb{E}_{\pi_B}(f) := \sum_{x \in \Omega} f(x) \pi_B(x)$, on obtient par inégalité de Cauchy-Schwarz :

$$\mathbb{E}_{\pi_B}(f)^2 = \langle f|_B, \mathbf{1} \rangle_{\pi_B}^2 \leq \langle f|_B, f|_B \rangle_{\pi_B} = \mathbb{E}_{\pi_B}(f^2)$$

D'où par bilinéarité de $\langle \cdot, \cdot \rangle_{\pi}$, on a $\text{Var}_{\pi}(f) = \|f\|_2^2 - \mathbb{E}_{\pi}(f)^2 = \|f\|_2^2 - \pi(B)^2 \mathbb{E}_{\pi_B}(f)^2 \geq \|f\|_2^2 - \pi(B)^2 \mathbb{E}_{\pi_B}(f^2)$. Ainsi

$$\text{Var}_{\pi}(f) \geq \|f\|_2^2 - \frac{\pi(B)^2}{\pi(B)} \|f\|_2^2 = \pi(A) \|f\|_2^2 \quad (3.8)$$

Remarquons que $\{f \in \mathbb{R}^\Omega : f = f|_B \mathbf{1}_B, f \in \mathbb{R}^\Omega \setminus \{0\}\} \subset \{f \in \mathbb{R}^\Omega : \text{Var}_{\pi}(f) \neq 0\}$ donc d'après la relation (3.3) on a

$$1/t_{\text{rel}} \leq \min_{f=f|_B \mathbf{1}_B; f \in \mathbb{R}^\Omega \setminus \{0\}} \frac{\mathcal{E}(f)}{\text{Var}_{\pi}(f)}$$

Mais d'après (3.8) pour tout $f \in \{f \in \mathbb{R}^\Omega : f = f|_B \mathbf{1}_B, f \in \mathbb{R}^\Omega \setminus \{0\}\}$, $\frac{\pi(A)}{\text{Var}_{\pi}(f)} \leq \frac{1}{\|f\|_2^2}$. De plus, on sait que $\mathcal{E}(f) = \frac{1}{2} \sum_{x, y \in \Omega} (f(x) - f(y))^2 \pi(x) P(x, y) \geq 0$ [5, lemme 13.6]. Donc d'après (3.7)

$$\frac{\pi(A)}{t_{\text{rel}}} \leq 1 - \gamma_1$$

ce qu'on voulait démontrer.

Montrons (ii). Remarquons que si on trouve une telle décomposition, alors nécessairement, $\sum_{i=1}^k a_i =$

1. En effet, $\mathbb{P}_{\pi_B}(T_A > 0) = \sum_{i=1}^k a_i$ et $\mathbb{P}_{\pi_B}(T_A > 0) = 1$ car π_B est la mesure initiale.

Pour tout entier t , on a $\mathbb{P}_{\pi_B}(T_A > t) = \sum_{x \in \Omega} \pi_B(x) \mathbb{P}_x(T_A > t) = \sum_{x \in B} \pi_B(x) \mathbb{P}_x(T_A > t)$ car pour tout $x \in A$, $\mathbb{P}_x(T_A > t) = 0$. Donc

$$\mathbb{P}_{\pi_B}(T_A > t) = \sum_{x \in B} \pi_B(x) \mathbb{P}_x(\forall i \leq t, X_i \in B) = \sum_{x, y \in B} \pi_B(x) (P_B)_t(x, y) = \sum_{x, y \in B} \pi_B(x) (P_B)_t \mathbf{1}_y(x)$$

Décomposons $(P_B)_t \mathbf{1}_y$ dans la base orthonormée (g_1, \dots, g_k) .

$$(P_B)_t \mathbf{1}_y = \sum_{i=1}^k \langle (P_B)_t \mathbf{1}_y, g_i \rangle_{\pi_B} g_i = \sum_{i=1}^k \langle \mathbf{1}_y, (P_B)_t g_i \rangle_{\pi_B} g_i = \sum_{i=1}^k \gamma_i^t g_i(y) \pi_B(y) g_i$$

La deuxième égalité est vraie car P_B est auto-adjoint pour ce produit scalaire et la troisième est vraie car $(P_B)_t g_i = \gamma_i^t g_i$. Ainsi

$$\mathbb{P}_{\pi_B}(T_A > t) = \sum_{x, y \in B} \pi_B(x) \sum_{i=1}^k \gamma_i^t g_i(y) g_i(x) \pi_B(y) = \sum_{i=1}^k a_i \gamma_i^t$$

où $a_i := \sum_{x, y \in B} \pi_B(x) \pi_B(y) g_i(x) g_i(y)$.

La preuve de (iii) est immédiate. En effet en utilisant (ii) puis (i) pour majorer les γ_i , on a

$$\mathbb{P}_{\pi_B}(T_A > t) \leq \sum_{i=1}^k a_i \left(1 - \frac{\pi(A)}{t_{\text{rel}}}\right)^t \leq e^{-t \frac{\pi(A)}{t_{\text{rel}}}} \quad \forall t \geq 0$$

où la dernière inégalité provient de la concavité du log, ce qui termine la preuve.

Grâce à la relation (3.6), on peut montrer (voir [3, Lemme 3.5]) que pour tout $t \geq 0$ et pour tout $A \subsetneq \Omega$ non vide

$$\mathbb{P}_{\pi}(T_A > t) \leq \pi(A^c) e^{-t \frac{\pi(A)}{t_{\text{rel}}}} \quad (3.9)$$

Définition 3.1.8. (Condition produit)

Soit $(X^n)_{n \in \mathbb{N}}$ une suite de chaînes de Markov telle que pour tout n le n -ème terme $X^n = (X^n(t))_{t \in \mathbb{N}}$ est une chaîne irréductible, réversible, d'espace d'état Ω_n fini, de matrice de transition P^n et de mesure invariante π_n . On note t_{mix}^n et t_{rel}^n les temps de mélange et de relaxation associés à la chaîne X^n . On dira que $(X^n)_{n \in \mathbb{N}}$ vérifie la condition produit si $\lim_{n \rightarrow \infty} \frac{t_{\text{mix}}^n}{t_{\text{rel}}^n} = \infty$, ou de manière équivalente, si $t_{\text{rel}}^n = o(t_{\text{mix}}^n)$.

3.2 Temps d'atteinte de grands ensembles et hit_{α} -cutoff

Définition 3.2.1. Soit $X = (X(t))_{t \in \mathbb{N}}$ une chaîne de Markov irréductible d'espace d'état Ω , de semi-groupe de Markov $(P_t)_{t \in \mathbb{N}}$ et de mesure invariante π .

Soit $A \subset \Omega$. On note $T_A := \inf\{t > 0 : X_t \in A\}$ le temps d'atteinte de l'ensemble A .

Soit $\alpha, \varepsilon \in (0, 1)$. Posons $p_x(\alpha, t) := \max_{A \subset \Omega, \pi(A) \geq \alpha} \mathbb{P}_x(T_A > t)$ et $p(\alpha, t) := \max_{x \in \Omega} p_x(\alpha, t)$.

Enfin, on définit

$$\text{hit}_{\alpha, x}(\varepsilon) := \min\{t > 0 : p_x(\alpha, t) \leq \varepsilon\} \quad \text{et} \quad \text{hit}_\alpha(\varepsilon) := \min\{t > 0 : p(\alpha, t) \leq \varepsilon\}$$

Remarque 3.2.2. Grâce à la propriété de Markov simple, on a la sous-multiplicativité de hit_α c'est-à-dire que pour tout $\alpha, \varepsilon, \delta \in (0, 1)$ $\text{hit}_\alpha(\varepsilon\delta) \leq \text{hit}_\alpha(\varepsilon) + \text{hit}_\alpha(\delta)$.

Voici une inégalité pratique reliant la notions ci-dessus avec le temps de mélange.

Proposition 3.2.3. Soit $x \in \Omega, \alpha \in (0, 1)$ fixés. Alors pour tout $\varepsilon \in (0, 1)$, on a

$$\text{hit}_{1-\alpha}((\alpha + \varepsilon) \wedge 1) \leq t_{\text{mix}}(\varepsilon)$$

Preuve. Soit $\varepsilon \in (0, 1)$. Posons $t := t_{\text{mix}}(\varepsilon)$. Soit $A \subset \Omega$ tel que $\pi(A) \geq 1 - \alpha$. On a évidemment $p(1 - \alpha, t) \leq 1$. De plus, $\mathbb{P}_x(T_A \leq t) = \mathbb{P}_x(\exists i \leq t : X_i \in A) \geq \mathbb{P}_x(X_t \in A)$. Mais par définition de t et de la distance en variation totale, $\mathbb{P}_x(X_t \in A) - \pi(A) \geq -\varepsilon$ d'où $\mathbb{P}_x(T_A \leq t) \geq 1 - \alpha - \varepsilon$. En prenant le max sur l'ensemble des A tel que $\pi(A) \geq 1 - \alpha$ puis le max sur le x dans Ω , on a $p(1 - \alpha, t) \leq \alpha + \varepsilon$. Finalement

$$p(1 - \alpha, t) \leq (\alpha + \varepsilon) \wedge 1$$

ce qui implique par définition de $\text{hit}_{1-\alpha}((\alpha + \varepsilon) \wedge 1)$ que $\text{hit}_{1-\alpha}((\alpha + \varepsilon) \wedge 1) \leq t$.

Corollaire 3.2.3.1. Soit $m > 1$ fixé. Pour tout $\alpha \in (0, 1)$, il existe $k \in \mathbb{N}$ tel que

$$\text{hit}_\alpha\left(1 - \frac{\alpha}{m}\right) \leq kt_{\text{mix}}$$

Preuve. En remplaçant α par $1 - \alpha$ et en prenant $\varepsilon = \frac{m-1}{m}\alpha$ dans la proposition 3.2.3, on a $\text{hit}_\alpha\left(1 - \frac{\alpha}{m}\right) \leq t_{\text{mix}}\left(\frac{m-1}{m}\alpha\right)$. Pour obtenir le résultat voulu, il suffit, par définition du temps de mélange, de trouver $k \in \mathbb{N}$ tel que $d(kt_{\text{mix}}) \leq \frac{m-1}{m}\alpha$ (où d est la distance définie la première partie). Mais puisque $d(kt_{\text{mix}}) \leq 2^k d(t_{\text{mix}})^k \leq 2^{-k}$ (d'abord d'après (1.3) puis par définition de t_{mix}), il suffit de trouver $k \in \mathbb{N}$ tel que $2^{-k} \leq \frac{m-1}{m}\alpha$. Par croissance du \log_2 , on voit qu'il suffit de prendre

$$k = \left\lceil \left\lceil \log_2 \left(\frac{m-1}{m} \alpha \right) \right\rceil \right\rceil$$

Définition 3.2.4. Soit $(X^n)_{n \in \mathbb{N}}$ une suite de chaînes de Markov telle que pour tout n le n -ème terme $X^n = (X^n(t))_{t \in \mathbb{N}}$ est une chaîne irréductible, d'espace d'état Ω_n , de semi-groupe de Markov

$(P_t^n)_{t \in \mathbb{N}}$ et de mesure invariante π_n . On dit que $(X^n)_{n \in \mathbb{N}}$ présente un hit_α -cutoff si pour tout $\varepsilon \in (0, 1/4)$

$$\text{hit}_\alpha^n(\varepsilon) - \text{hit}_\alpha^n(1 - \varepsilon) = o(\text{hit}_\alpha^n(1/4))$$

Nous allons maintenant énoncer une condition nécessaire et suffisante pour qu'une suite présente un hit_α -cutoff.

Proposition 3.2.5. (Basu, Hermon et Peres, 2017)

Soit $(X^n)_{n \in \mathbb{N}}$ une suite de chaînes de Markov telle que pour tout n le n -ème terme $X^n = (X^n(t))_{t \in \mathbb{N}}$ est une chaîne irréductible, réversible, lazy d'espace d'état Ω_n fini, de matrice de transition P^n et de mesure invariante π_n vérifiant la condition produit. Alors les deux assertions suivantes sont équivalentes :

- (i) Il existe $\alpha \in (0, 1)$ tel que $(X^n)_{n \in \mathbb{N}}$ présente un hit_α -cutoff.
- (ii) Pour tout $\alpha \in (0, 1)$, $(X^n)_{n \in \mathbb{N}}$ présente un hit_α -cutoff.

De plus il existe $k_1, k_2 > 0$ tel que pour n assez grand

$$k_1 t_{mix}^n \leq \text{hit}_\alpha^n(1/4) \leq k_2 t_{mix}^n \quad \forall \alpha \in (0, 1) \quad (3.10)$$

et si (i) ou (ii) est vérifiée, alors

$$\text{hit}_\alpha^n(1/4) \underset{n \rightarrow \infty}{\sim} \text{hit}_{1/2}^n(1/4) \quad \forall \alpha \in (0, 1) \quad (3.11)$$

voir [3, proposition 3.6 p. 1463-1464].

Continuons avec une comparaison asymptotique :

Proposition 3.2.6. (Basu, Hermon et Peres, 2017)

Soit $(X^n)_{n \in \mathbb{N}}$ une suite de chaînes de Markov telle que pour tout n le n -ème terme $X^n = (X^n(t))_{t \in \mathbb{N}}$ est une chaîne irréductible, réversible, lazy d'espace d'état Ω_n fini, de matrice de transition P^n et de mesure invariante π_n vérifiant la condition produit. Alors pour tout $\varepsilon \in (0, 1/4]$

$$\text{hit}_{1/2}^n(3\varepsilon/2) \leq t_{mix}^n(\varepsilon) \leq \text{hit}_{1/2}^n(\varepsilon/2) \quad (3.12)$$

et

$$\text{hit}_{1/2}^n(1 - \varepsilon/2) \leq t_{mix}^n(1 - \varepsilon) \leq \text{hit}_{1/2}^n(1 - 2\varepsilon) \quad (3.13)$$

voir [3, proposition 1.8 p. 1452] et pour la preuve, voir [3, p. 1462-1463].

Voici une proposition très importante que nous allons démontrer :

Proposition 3.2.7. (Basu, Hermon et Peres, 2017)

Soit $(X^n)_{n \in \mathbb{N}}$ une suite de chaînes de Markov telle que pour tout n le n -ème terme $X^n = (X^n(t))_{t \in \mathbb{N}}$ est une chaîne irréductible, réversible, lazy d'espace d'état Ω_n fini, de matrice de transition P^n et de mesure invariante π_n . Supposons que la condition produit ne soit pas vérifiée. Alors pour tout $\alpha \leq 1/2$, $(X^n)_{n \in \mathbb{N}}$ ne présente pas de hit_α -cutoff.

Preuve. Soit $m \in \mathbb{R}^+$ tel que $m < 4$ et $\alpha \in (0, m/4]$. Commençons par montrer qu'il existe $k_\varepsilon \in \mathbb{N}$ tel que $\text{hit}_\alpha^n(\varepsilon/2) \geq k_\varepsilon$ pour tout $\varepsilon \in (0, 1)$. En effet, supposons que la loi initiale de la chaîne X^n est δ_{x_n} avec $\pi_n(x_n) \leq 1 - \alpha$. Posons $A := \Omega \setminus x_n$. Alors $\pi(A) \geq \alpha$ et $\mathbb{P}_{x_n}(T_A^n > k_\varepsilon) = \mathbb{P}_{x_n}(\forall i \leq k_\varepsilon, X_i^n = x_n) = (P^n(x_n, x_n))^{k_\varepsilon} \geq 2^{-k_\varepsilon}$ où la deuxième égalité est obtenue par propriété de Markov simple. Pour obtenir le résultat voulu, il suffit que $2^{-k_\varepsilon} \geq \varepsilon/2$. Par croissance du \log_2 , on peut choisir $k_\varepsilon = \lceil \log_2(\varepsilon) \rceil$. Ainsi, $p(\alpha, k_\varepsilon) = \max_{A \subset \Omega_n; \pi(A) \geq \alpha} \max_{x_n \in \Omega_n} \mathbb{P}_{x_n}(T_A^n > k_\varepsilon) \geq 2^{-k_\varepsilon} \geq \varepsilon/2$, exactement ce que nous voulions.

Enfin, d'après le corollaire 3.2.3.1, $\text{hit}_\alpha^n(1 - \frac{\alpha}{m}) \leq \lceil \log_2(\frac{m-1}{m}\alpha) \rceil t_{\text{mix}}^n$.

Finalement, pour tout $0 < \varepsilon \leq \alpha/m$

$$\frac{\text{hit}_\alpha^n(\varepsilon)}{\text{hit}_\alpha^n(1 - \varepsilon)} \geq \frac{\lceil \log_2(\varepsilon) \rceil}{\lceil \log_2(\frac{m-1}{m}\alpha) \rceil t_{\text{mix}}^n} \quad (3.14)$$

Nous pouvons supposer ici que t_{mix}^n ne tend pas vers l'infini quand n tend vers l'infini. En effet, si t_{mix}^n tendait vers l'infini, (3.14) n'aurait aucun intérêt. En faisant tendre ε vers 0 dans (3.14), la quantité de droite tend vers l'infini et ainsi $(X^n)_{n \in \mathbb{N}}$ ne présente pas de hit_α -cutoff pour tout $\alpha \leq m/4$.

Supposons maintenant que $\lim_{n \rightarrow \infty} t_{\text{mix}}^n = \infty$. La condition produit n'étant pas vérifiée, on a $\lim_{n \rightarrow \infty} \frac{t_{\text{rel}}^n}{t_{\text{mix}}^n} \neq \infty$ donc il existe une sous-suite $\left(\frac{t_{\text{rel}}^{\varphi(n)}}{t_{\text{mix}}^{\varphi(n)}}\right)_{n \in \mathbb{N}}$ extraite de $\left(\frac{t_{\text{rel}}^n}{t_{\text{mix}}^n}\right)_{n \in \mathbb{N}}$ telle que $\frac{t_{\text{rel}}^{\varphi(n)}}{t_{\text{mix}}^{\varphi(n)}} < C, C > 0$.

On a donc $\lim_{n \rightarrow \infty} t_{\text{rel}}^{\varphi(n)} = \infty$. La chaîne X^n étant réversible, P^n est diagonalisable dans une base orthonormée de vecteurs propres. Notons λ_2^n la plus valeur propre de P^n différente de 1 et $P_n f_2^n = \lambda_2^n f_2^n$, $f_2^n \in \mathbb{R} \setminus \{0\}$. On alors que $\left(\lambda_2^{\varphi(n)}\right)_{t_{\text{mix}}^{\varphi(n)}} = \left(1 - 1/t_{\text{rel}}^{\varphi(n)}\right)_{t_{\text{mix}}^{\varphi(n)}} \geq e^{-C}$ pour n assez grand. Pour éviter de trop charger les notations, nous n'écrivons plus la dépendance en n .

Posons $A := \{y \in \Omega : f_2(y) \leq 0\}$. Supposons dans un premier temps que $\pi(A) \geq \alpha$. Soit $x \in \Omega$ tel que $f_2(x) = \max_{y \in \Omega} f_2(y) =: L$. Puisque $\lambda_2 < 1$ et que $\mathbb{E}_\pi(f_2) = \mathbb{E}_\pi(Pf_2) = \lambda_2 \mathbb{E}_\pi(f_2)$ alors $\mathbb{E}_\pi(f_2) = 0$. De plus $L > \mathbb{E}_\pi(f_2)$ donc $L > 0$.

Supposons que $X_0 = x$. $\mathcal{F}_k := \sigma(X_j, j \leq k)$. On définit $N_k := \lambda_2^{-k} f(X_k)$ et $M_k = N_{k \wedge T_A}$. Alors $(N_k)_{k \in \mathbb{N}}$ et $(M_k)_{k \in \mathbb{N}}$ sont des $(\mathcal{F}_k)_{k \in \mathbb{N}}$ -martingales. En effet :

N_k est bien sûr \mathcal{F}_k -mesurable et intégrable. Notons que $Pf_2(X_k) := \mathbb{E}(f(X_1)|X_k) = \mathbb{E}(f(X_{k+1})|\mathcal{F}_k)$ par propriété de Markov simple.

On a $\mathbb{E}(N_{k+1}|\mathcal{F}_k) = \lambda_2^{-k-1}\mathbb{E}(f(X_{k+1})|\mathcal{F}_k) = \lambda_2^{-k-1}Pf_2(X_k) = \lambda_2^{-k}f_2(X_k) = N_k$. Enfin, $(M_k)_{k \in \mathbb{N}}$ est une martingale arrêtée définie à partir de $(N_k)_{k \in \mathbb{N}}$, donc est une $(\mathcal{F}_k)_{k \in \mathbb{N}}$ -martingale.

Maintenant, puisque $M_k \mathbb{1}_{\{T_A \leq k\}} = N_{T_A} \mathbb{1}_{\{T_A \leq k\}} \leq 0$ car $\lambda_2 > 0$ et par définition de A et puisque $M_k \mathbb{1}_{\{T_A > k\}} = N_k \mathbb{1}_{\{T_A > k\}} \leq \lambda_2^{-k}L$ car $\lambda_2 > 0$ par définition de L , on a $M_k \leq \lambda_2^{-k}L \mathbb{1}_{\{T_A > k\}}$ (bien que non nécessaire, nous conservons $\mathbb{1}_{\{T_A > k\}}$ car nous cherchons à minorer $\mathbb{P}_x(T_A > k)$). De plus, comme $(M_k)_{k \in \mathbb{N}}$ est une martingale, $\mathbb{E}_x(M_k) = \mathbb{E}_x(M_0) = \mathbb{E}_x(N_0) = f_2(x) = L$. Donc $L \leq \lambda_2^{-k}L \mathbb{P}_x(T_A > k)$ pour tout k . Puisque $L > 0$ alors pour tout $a > 0$, on obtient :

$$\mathbb{P}_x(T_A > at_{\text{mix}}) \geq \lambda_2^{at_{\text{mix}}} \geq e^{-aC} \quad (3.15)$$

On cherche maintenant $a_\varepsilon > 0$ tel que pour tout $\varepsilon \in (0, 1)$, $\text{hit}_\alpha(\varepsilon/2) \geq a_\varepsilon t_{\text{mix}}$. Pour cela, il suffit que $p(\alpha, a_\varepsilon t_{\text{mix}}) \geq \varepsilon/2$. Mais par définition, $p(\alpha, a_\varepsilon t_{\text{mix}}) \geq \mathbb{P}_x(T_A > a_\varepsilon t_{\text{mix}})$. Donc d'après (3.15), il suffit que $e^{-a_\varepsilon C} \leq \varepsilon$. $a_\varepsilon = C^{-1}|\log \varepsilon|$ convient.

On suppose à présent que $\pi(A) \leq \alpha$. Alors $\pi(A^c) \geq 1 - \alpha$ et $A^c = \{y \in \Omega : -f_2(y) < 0\}$. Mais puisque $P(-f_2) = \lambda_2(-f_2)$ alors la démonstration précédente est toujours valable, il suffit de remplacer A par A^c et f_2 par $-f_2$. Seule la conclusion change : on obtient pour tout $\varepsilon \in (0, 1)$, $\text{hit}_{1-\alpha}(\varepsilon/2) \geq a_\varepsilon t_{\text{mix}}$. Puisque nous voulions plutôt minorer $\text{hit}_\alpha(\varepsilon/2)$, il suffit de choisir α tel que $\text{hit}_\alpha(\varepsilon/2) \geq \text{hit}_{1-\alpha}(\varepsilon/2)$ c'est-à-dire tel que $1 - \alpha \geq \alpha$ donc $\alpha \geq 1/2$ (et donc $m = 2$).

On obtient ainsi quelque chose de plus satisfaisant : Pour tout $0 < \varepsilon \leq \alpha/2$, en utilisant de nouveau le corollaire 3.2.3.1,

$$\frac{\text{hit}_\alpha(\varepsilon)}{\text{hit}_\alpha(1 - \varepsilon)} \geq \frac{|\log(\varepsilon)|t_{\text{mix}}}{C^{-1} \lceil |\log_2(\frac{\alpha}{2})| \rceil t_{\text{mix}}} = \frac{|\log(\varepsilon)|}{C^{-1} \lceil |\log_2(\frac{\alpha}{2})| \rceil} \quad (3.16)$$

En faisant tendre ε vers 0 dans (3.16), la quantité de droite tend vers l'infini et ainsi, pour tout $\alpha \leq 1/2$, la suite $(X^n)_{n \in \mathbb{N}}$ ne présente pas de hit_α -cutoff.

Nous sommes maintenant prêts à démontrer un théorème affirmant que pour certaines suites de chaînes de Markov, la condition produit est "presque" suffisante pour que ces suites présentent un cutoff.

Théorème 3.2.8. (Basu, Hermon et Peres, 2017)

Soit $(X^n)_{n \in \mathbb{N}}$ une suite de chaînes de Markov telle que pour tout n le n -ème terme $X^n = (X^n(t))_{t \in \mathbb{N}}$ est une chaîne irréductible, réversible, lazy d'espace d'état Ω_n fini, de matrice de transition P^n et de mesure invariante π_n . Alors les assertions suivantes sont équivalentes :

- (i) $(X^n)_{n \in \mathbb{N}}$ présente un cutoff.
- (ii) Il existe $\alpha \in (0, 1/2]$ tel que $(X^n)_{n \in \mathbb{N}}$ présente un hit_α -cutoff.
- (iii) Pour tout $\alpha \in (1/2, 1)$, $(X^n)_{n \in \mathbb{N}}$ présente un hit_α -cutoff et la condition produit est vérifiée.

La condition "presque" suffisante dont nous parlions est l'implication $(iii) \Rightarrow (i)$.

Preuve. L'équivalence entre (ii) et (iii) est quasiment immédiate.

Supposons (ii) . Il existe donc $\alpha \in (0, 1/2]$ tel que $(X^n)_{n \in \mathbb{N}}$ présente un hit_α -cutoff. En prenant la contraposée de la proposition 3.2.7, on obtient que la condition produit est vérifiée. Cela implique donc, d'après la proposition 3.2.5, que pour tout $\alpha \in (0, 1)$, $(X^n)_{n \in \mathbb{N}}$ présente un hit_α -cutoff, ce qui est vrai en particulier pour tout $\alpha \in (1/2, 1)$, c'est-à-dire (iii) .

Supposons (iii) . La condition produit est donc vérifiée et pour tout $\alpha \in (1/2, 1)$, $(X^n)_{n \in \mathbb{N}}$ présente un hit_α -cutoff. Donc d'après la proposition 3.2.5, ceci est vrai pour tout $\alpha \in (0, 1)$. En particulier, il existe $\alpha \in (0, 1/2]$ tel que $(X^n)_{n \in \mathbb{N}}$ présente un hit_α -cutoff, c'est-à-dire (ii) .

Montrons maintenant que (i) et (ii) sont équivalentes.

Supposons (i) . Alors $(X^n)_{n \in \mathbb{N}}$ présente un cutoff et donc la condition produit est vérifiée (puisque nous avons dit dans l'introduction que la condition produit est nécessaire pour avoir un cutoff).

Montrons que $(X^n)_{n \in \mathbb{N}}$ présente un $\text{hit}_{1/2}$ -cutoff. Pour tout $\varepsilon \in (0, 1/4)$, on a

$$\text{hit}_{1/2}^n(2\varepsilon) - \text{hit}_{1/2}^n(1 - 2\varepsilon) \leq \text{hit}_{1/2}^n(3\varepsilon/2) - \text{hit}_{1/2}^n(1 - 2\varepsilon) \leq t_{\text{mix}}^n(\varepsilon) - t_{\text{mix}}^n(1 - \varepsilon)$$

où la dernière inégalité est obtenue en soustrayant (3.13) à (3.12). Mais puisque $(X^n)_{n \in \mathbb{N}}$ présente un cutoff et que $t_{\text{mix}}^n(1 - \varepsilon) \leq t_{\text{mix}}^n$, on obtient que $t_{\text{mix}}^n(\varepsilon) - t_{\text{mix}}^n(1 - \varepsilon) = o(t_{\text{mix}}^n)$. Enfin d'après (3.10), il existe $k_1 > 0$ tel que $k_1 t_{\text{mix}}^n \leq \text{hit}_\alpha^n(1/4)$. Finalement, $\text{hit}_{1/2}^n(2\varepsilon) - \text{hit}_{1/2}^n(1 - 2\varepsilon) = o(\text{hit}_\alpha^n(1/4))$ pour tout $\varepsilon \in (0, 1/4)$ d'où le $\text{hit}_{1/2}$ -cutoff. Enfin, la proposition 3.2.5 permet de conclure qu'il y a hit_α -cutoff pour tout $\alpha \in (0, 1)$, donc pour un certain $\alpha \in (0, 1/2]$ c'est-à-dire (ii) .

Supposons (ii) . Alors il existe $\alpha \in (0, 1/2]$ tel que la suite $(X_n)_{n \in \mathbb{N}}$ présente un hit_α -cutoff. En prenant la contraposée de la proposition 3.2.7, on obtient que la condition produit est vérifiée. Donc $(X_n)_{n \in \mathbb{N}}$ présente un $\text{hit}_{1/2}$ -cutoff d'après la proposition 3.2.5. En divisant (3.13) par (3.12), on a, pour tout $\varepsilon \in (0, 1/4]$

$$1 \leq \frac{t_{\text{mix}}^n(\varepsilon)}{t_{\text{mix}}^n(1 - \varepsilon)} \leq \frac{\text{hit}_{1/2}^n(\varepsilon/2)}{\text{hit}_{1/2}^n(1 - \varepsilon/2)} \xrightarrow{n \rightarrow \infty} 1$$

car $(X_n)_{n \in \mathbb{N}}$ présente un $\text{hit}_{1/2}$ -cutoff et que $\text{hit}_{1/2}^n(1/4) \leq \text{hit}_{1/2}^n(\varepsilon/2)$.

Donc pour tout $\varepsilon \in (0, 1/4]$,

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

Pour $\varepsilon \in (1/4, 1)$, on a

$$1 \leq \frac{t_{\text{mix}}^n(\varepsilon)}{t_{\text{mix}}^n(1 - \varepsilon)} \leq \frac{t_{\text{mix}}^n(1/4)}{t_{\text{mix}}^n(1 - 1/4)}$$

On conclut ainsi que la suite $(X_n)_{n \in \mathbb{N}}$ présente un cutoff, c'est-à-dire (i) .

3.3 Un exemple de suite de chaînes Markov pour laquelle la condition produit est suffisante pour obtenir le cutoff

Pour conclure ce rapport, nous allons énoncer puis démontrer un théorème affirmant que la condition produit est parfois suffisante pour qu'une suite de chaînes de Markov présente le phénomène de cutoff.

Théorème 3.3.1. *Soit $(X^n)_{n \in \mathbb{N}}$ une suite de chaînes de Markov telle que pour tout n le n -ème terme $X^n = (X^n(t))_{t \in \mathbb{N}}$ est une chaîne réversible, lazy, d'espace d'état V_n sur un arbre fini $T_n = (V_n, E_n)$ avec $|V_n| \geq 3$. Notons P^n la matrice de transition de X^n et π_n sa mesure invariante. Si la condition produit est vérifiée pour $(X^n)_{n \in \mathbb{N}}$ alors cette suite présente le phénomène de cutoff.*

On peut affiner le théorème précédent :

Théorème 3.3.2. (Basu, Hermon et Peres, 2017)

Soit $X = (X(t))_{t \in \mathbb{N}}$ une chaîne de Markov réversible, lazy d'espace d'état V sur un arbre fini $T = (V, E)$ avec $|V| \geq 3$. Notons P sa matrice de transition et π sa mesure invariante. Alors pour tout $\varepsilon \in (0, 1/4]$

$$t_{mix}(\varepsilon) - t_{mix}(1 - \varepsilon) \leq 35\sqrt{\varepsilon^{-1}t_{relmix}}$$

En particulier, si $(X^n)_{n \in \mathbb{N}}$ est une suite de chaînes de Markov telle que pour tout n le n -ème terme $X^n = (X^n(t))_{t \in \mathbb{N}}$ est une chaîne réversible, lazy, d'espace d'état V_n sur un arbre fini $T_n = (V_n, E_n)$ et que la condition produit est vérifiée, alors cette suite présente un cutoff avec une fenêtre de cutoff égale à $\sqrt{t_{relmix}^n}$.

Ainsi le théorème 3.3.2 implique le théorème 3.3.1 et nous donne une précision intéressante : la fenêtre de cutoff. Nous ne démontrons néanmoins pas ce théorème dont la preuve se trouve dans [3, p. 1473-1474].

Commençons par quelques définitions : Soit $T := (V, E)$ un arbre fini. V est l'ensemble des sommets de cette arbre et E est l'ensemble des arêtes de T . Soit $X = (X(t))_{t \in \mathbb{N}}$ une chaîne de Markov lazy d'espace d'état V sur T . Notons P sa matrice de transition. On a alors $P(x, y) > 0$ si et seulement si $\{x, y\} \in E$ (c'est-à-dire x et y reliés par une arête) ou $x = y$ (et dans ce cas $P(x, y) \geq 1/2$). X est irréductible et notons π sa mesure invariante. Grâce à la structure d'arbre, X est réversible. En effet, puisqu'un arbre est un graphe connexe n'ayant aucun cycle, le critère de Kolmogorov s'applique.

Définition 3.3.3. *Soit $v \in V$. Notons $\mathcal{C}(T, v)$ l'ensemble des composantes connexes de $T \setminus \{v\}$. v est un sommet central (littéralement central-vertex) de T si pour tout $C \in \mathcal{C}(T, v)$, $\pi(C) \leq 1/2$.*

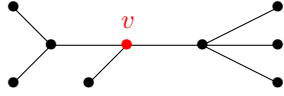


FIGURE 3.1 – Exemple d'un arbre $T = (V, E), v \in V$.

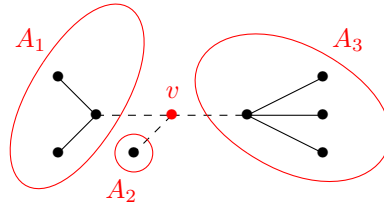


FIGURE 3.2 – Composantes connexes de $T \setminus \{v\}$.

Les figures 3.1 et 3.2 illustre la définition ci-dessus : les arbres $A_i = (V_i, E_i), i = 1, 2, 3$ sont les composantes connexes de l'arbre $T \setminus \{v\}$. On suppose que v est un sommet central de T donc $\bigcup_{i=1}^3 V_i = V \setminus \{v\}$ et $V_i \cap V_j = \emptyset, i \neq j, i, j = 1, 2, 3$. De plus pour tout $i \in \{1, 2, 3\}, \pi(V_i) \leq 1/2$.

Remarque 3.3.4. *Il existe toujours un sommet central mais il y en a au plus 2.*

Nous allons énoncer un lemme avant de démontrer le théorème. On fixe un sommet central c de T . Notons T_c le temps d'atteinte de c . Posons enfin pour tout $\varepsilon \in (0, 1)$

$$\tau_c(\varepsilon) := \min\{t : \mathbb{P}_x(T_c > t) \leq \varepsilon \forall x \in V\}$$

Lemme 3.3.5. *Pour tout $0 < \delta < \varepsilon < 1$, il existe un entier $s > 0$ tel que*

$$\tau_c(\varepsilon) \leq \text{hit}_{1/2}(\varepsilon) \leq \tau_c(\varepsilon - \delta) + s$$

Preuve. Commençons par l'inégalité de gauche. Fixons $\varepsilon \in (0, 1)$. Soit $x \in V$ quelconque. Notons C_x la composante connexe de $T \setminus \{c\}$. Par définition de v , $\pi(C_x) \leq 1/2$. Ainsi $A_x := V \setminus C_x$ est tel que $\pi(A_x) \geq 1/2$. Soit t tel que $p(1/2, t) \leq \varepsilon$ c'est-à-dire tel que $\mathbb{P}_x(T_A > t) \leq \varepsilon$ pour tout $A \subset V$ vérifiant $\pi(A) \geq 1/2$. En particulier, $\mathbb{P}_x(T_{A_x} > t) \leq \varepsilon$. Cependant, par définition du sommet central c , en partant de x , nous sommes forcés de passer par c pour atteindre A_x et ce qui donne $\mathbb{P}_x(T_c > t) \leq \mathbb{P}_x(T_{A_x} > t) \leq \varepsilon$. Finalement nous venons de montrer que pour tout $\varepsilon \in (0, 1)$

$$\{t : p(1/2, t) \leq \varepsilon\} \subset \{t : \mathbb{P}_x(T_c > t) \leq \varepsilon \forall x \in V\}$$

ce qui donne le résultat voulu par définition de $\tau_c(\varepsilon)$ et $\text{hit}_{1/2}$.

Démontrons à présent l'inégalité de droite. Il suffit de trouver $s > 0$ tel que $p(1/2, \tau_c(\varepsilon) + s) \leq \varepsilon$. Soit $A \subset V$ tel que $\pi(A) \geq 1/2$ fixé et $x \in V$. Remarquons que pour tout $t, u > 0, \mathbb{P}_x(T_A > t + u) \leq$

$\mathbb{P}_x(T_c > t) + \mathbb{P}_c(T_A > u)$. En effet, $\mathbb{P}_x(T_A > t + u) = \mathbb{P}_x(T_A > t + u, T_c > t) + \mathbb{P}_x(T_A > t + u, T_c \leq t) \leq \mathbb{P}_x(T_c > t) + \mathbb{P}_x(T_A > t + u, T_c \leq t)$. Si $x \in A$, le résultat est immédiat. Si $x \notin A$ alors puisque $\pi(A) \geq 1/2$ et $\pi(C_x) \leq 1/2$ ce qui implique que A ne peut-être contenu dans $C_x \setminus \{x\}$ ainsi partant de x , par définition du sommet central c , il est indispensable de passer par c avant d'atteindre A . Donc par propriété de Markov forte appliquée au temps d'arrêt T_c , $\mathbb{P}_x(T_A > t + u, T_c \leq t) \leq \mathbb{P}_c(T_A > u)$ ce qui prouve la remarque. En particulier, par définition de $\tau_c(\varepsilon - \delta)$, on a pour tout $u > 0$

$$\mathbb{P}_x(T_A > \tau_c(\varepsilon - \delta) + u) \leq \mathbb{P}_x(T_c > \tau_c(\varepsilon - \delta)) + \mathbb{P}_c(T_A > u) \leq \varepsilon - \delta + \mathbb{P}_c(T_A > u)$$

On voit donc qu'il suffit de trouver $s > 0$ tel que $\mathbb{P}_c(T_A > u) \leq \delta$. si $c \in A$ alors le résultat est immédiat. Supposons donc $c \notin A$. Par définition du sommet central c , il facile de voir que l'on peut trouver T_1 et T_2 tels que $T \setminus \{c\} = T_1 \cup T_2$ et T_1 et T_2 sont chacun l'union de certaines composantes connexes de $T \setminus \{c\}$ avec $\pi(T_1), \pi(T_2) \leq 2/3$ et $\pi(T_1 \cap A) > 0$. Ainsi en posant $A_1 := A \cap T_1$, on peut supposer que $\pi(A_1) \geq 1/4$ (quitte à partitionner d'une autre façon). Posons également $B := T_2 \cup \{c\}$. Puisque $A_1 \subset A$ et $c \notin A$, on a bien-sûr $\mathbb{P}_c(T_A > u) \leq \mathbb{P}_c(T_{A_1} > u)$ pour tout $u > 0$. Toujours par définition du sommet central c , partant de n'importe quel point $b \in B$, nous sommes contraint de passer par c pour atteindre A_1 . Donc pour tout $u > 0$

$$\mathbb{P}_c(T_{A_1} > u) \leq \mathbb{P}_{\pi_B}(T_{A_1} > u) = \pi(B)^{-1} \mathbb{P}_\pi(T_{A_1} > u) \leq \frac{1 - \pi(A_1)}{\pi(B)} e^{-u \frac{\pi(A_1)}{t_{\text{rel}}}}$$

où la dernière inégalité est obtenue grâce à (3.9).

On a donc

$$\mathbb{P}_c(T_{A_1} > u) \leq \frac{3/4}{\pi(B)} e^{-u \frac{1}{4t_{\text{rel}}}}$$

Puisque $1 - \pi(c) \leq \pi(T_2) + 2/3$ et $\pi(B) = \pi(T_2) + \pi(c)$, on a $\pi(B) \geq 1/3$. Ainsi en posant $s := \lceil 4t_{\text{rel}} \lceil \log(\delta \frac{4}{9}) \rceil \rceil$, on a $\mathbb{P}_c(T_A > s) \leq \mathbb{P}_c(T_{A_1} > s) \leq \delta$. Finalement, nous avons trouvé $s > 0$ tel que pour tout $x \in V$ et pour tout $A \subset V$ tel que $\pi(A) \geq 1/2$, $\mathbb{P}_x(T_A > \tau_c(\varepsilon - \delta) + s) \leq \delta$ ce qui signifie, en maximisant sur les ensembles $A \subset V$ tel que $\pi(A) \geq 1/2$ puis sur les $x \in V$ que pour tout $0 < \delta < \varepsilon < 1$, $p(1/2, \tau_c(\varepsilon - \delta) + s) \leq \varepsilon$ ce qui donne le résultat voulu par définition de $\text{hit}_{1/2}(\varepsilon)$.

Voici un dernier lemme avant de démontrer le théorème :

Lemme 3.3.6. *Soit $x \in V$ et $\varepsilon \in (0, 1/4]$. Posons $t_x := \mathbb{E}_x(T_c)$. Si on $\rho := \max_{x \in V} t_x$ et $\kappa_\varepsilon := \sqrt{4\varepsilon^{-1} \rho t_{\text{rel}}}$ alors $\tau_c(\varepsilon) - \tau_c(1 - \varepsilon) \leq 2\kappa_\varepsilon \leq 32\sqrt{\varepsilon^{-1} t_{\text{mix}} t_{\text{rel}}}$.*

Pour la preuve, voir [3, p. 1473].

Preuve du théorème 3.3.1. Notons c_n le sommet central de l'arbre T_n pour tout entier n . Montrons que la suite $(X^n)_{n \in \mathbb{N}}$ présente un $\text{hit}_{1/2}$ -cutoff. Pour cela, fixons $\varepsilon \in (0, 1/4]$. Puisque $0 < \varepsilon/2 < \varepsilon < 1 - \varepsilon < 1$ alors grâce au lemme 3.3.5 (avec $\delta = \varepsilon/2$) on peut écrire

$$\text{hit}_{1/2}^n(\varepsilon) \leq \tau_{c_n}(\varepsilon/2) + s_n \quad (3.17)$$

et

$$\text{hit}_{1/2}^n(1 - \varepsilon) \geq \tau_{c_n}(1 - \varepsilon) \geq \tau_{c_n}(1 - \varepsilon/2) \quad (3.18)$$

En soustrayant (3.18) à (3.17) puis en appliquant le lemme 3.3.6, on obtient

$$\text{hit}_{1/2}^n(\varepsilon) - \text{hit}_{1/2}^n(1 - \varepsilon) \leq \tau_{c_n}(\varepsilon/2) - \tau_{c_n}(1 - \varepsilon/2) + s_n \leq 32\sqrt{(2/\varepsilon)t_{\text{mix}}t_{\text{rel}}} + s_n$$

c'est-à-dire

$$\text{hit}_{1/2}^n(\varepsilon) - \text{hit}_{1/2}^n(1 - \varepsilon) \leq 32\sqrt{(2/\varepsilon)t_{\text{mix}}^n t_{\text{rel}}^n} + \left\lceil 4t_{\text{rel}}^n \left| \log \left(\frac{2\varepsilon}{9} \right) \right| \right\rceil \quad (3.19)$$

Mais puisque $\lceil 4t_{\text{rel}}^n \left| \log \left(\frac{2\varepsilon}{9} \right) \right| \rceil \leq 4t_{\text{rel}}^n \left| \log \left(\frac{2\varepsilon}{9} \right) \right| + 1$ et selon (3.4), $t_{\text{rel}}^n \geq 2 \frac{|V|-1}{|V|} \geq 4/3 > 1$ car $|V| \geq 3$, alors en divisant par t_{mix}^n dans (3.19)

$$0 \leq \frac{\text{hit}_{1/2}^n(\varepsilon) - \text{hit}_{1/2}^n(1 - \varepsilon)}{t_{\text{mix}}^n} \leq +32\sqrt{(2/\varepsilon)\frac{t_{\text{rel}}^n}{t_{\text{mix}}^n}} + \frac{t_{\text{rel}}^n}{t_{\text{mix}}^n} \left(4 \left| \log \left(\frac{2\varepsilon}{9} \right) \right| + 1 \right) \xrightarrow{n \rightarrow \infty} 0$$

car $(X_n)_{n \in \mathbb{N}}$ vérifie la condition produit.

Nous avons finalement montré que pour tout $\varepsilon \in (0, 1/4]$, $\text{hit}_{1/2}^n(\varepsilon) - \text{hit}_{1/2}^n(1 - \varepsilon) = o(t_{\text{mix}}^n)$. Mais d'après la proposition 3.2.5, il existe $k_1 > 0$ tel que pour n assez grand, $k_1 t_{\text{mix}}^n \leq \text{hit}_{\alpha}^n(1/4)$. Donc $\varepsilon \in (0, 1/4]$, $\text{hit}_{1/2}^n(\varepsilon) - \text{hit}_{1/2}^n(1 - \varepsilon) = o(\text{hit}_{\alpha}^n(1/4))$. Autrement dit, $(X_n)_{n \in \mathbb{N}}$ présente un $\text{hit}_{1/2}^n$ -cutoff, ce qui implique que la suite $(X_n)_{n \in \mathbb{N}}$ présente le phénomène de cutoff d'après le théorème 3.2.8 (en utilisant l'implication (ii) \Rightarrow (i)).

Remarque 3.3.7. *La structure d'arbre sur laquelle est définie la suite $(X_n)_{n \in \mathbb{N}}$ joue un rôle essentiel dans la preuve de ce théorème, en particulier pour obtenir la relation (3.19). En effet cette dernière n'a aucune raison d'être vérifiée pour des suites de chaînes de Markov quelconques.*

Bibliographie

- [1] D. ALDOUS and P. DIACONIS. Shuffling Cards And Stopping times. *Amer. Math. Monthly*, 93(5) :333–348, 05 1986. doi : 10.2307/2323590.
- [2] D. ALDOUS and A. FILL, James. Reversible Markov Chains and Random Walks on Graphs. URL <https://www.stat.berkeley.edu/~aldous/RWG/book.html>.
- [3] R. BASU, J. HERMON, and Y. PERES. Characterization of cutoff for reversible markov chains. *Ann. Probab.*, 45(3) :1448–1487, 05 2017. doi : 10.1214/16-AOP1090.
- [4] H. LACOIN. A product chain without cutoff. *Electron. Commun. Probab.*, 20 :9 pp., 2015. doi : 10.1214/ECP.v20-3765.
- [5] D. A. LEVIN, Y. PERES, and E. L. WILMER. Markov Chains and Mixing Times, second edition. URL <http://pages.uoregon.edu/dlevin/MARKOV/>.
- [6] J. R. NORRIS. *Markov Chains*. Cambridge University Press, 1997.