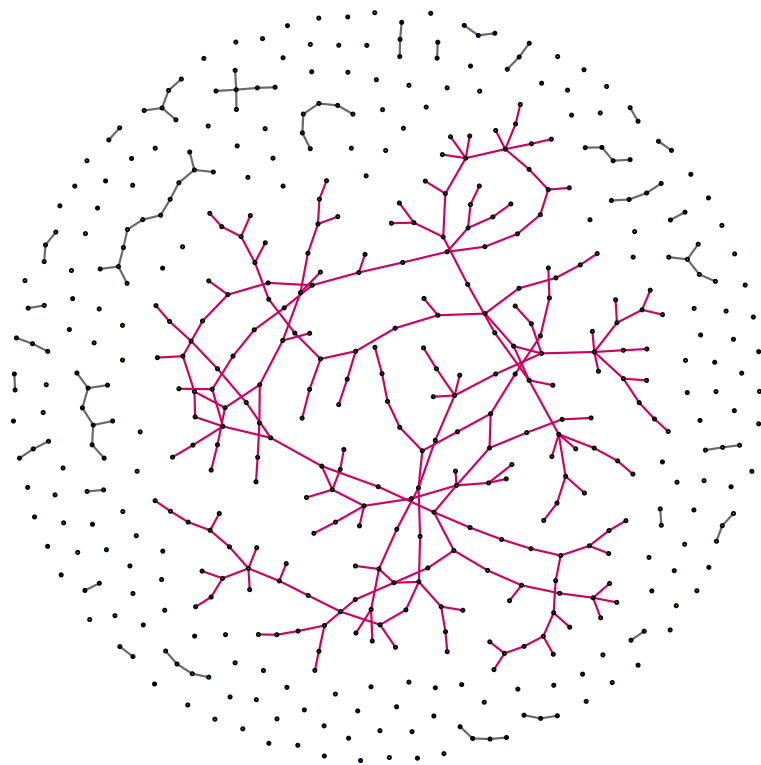


Chapitres Choisis de Théorie des Probabilités

B

– Yvan Velenik –



UNIVERSITÉ
DE GENÈVE

— Version du 16 juin 2024 —

Dernière version téléchargeable à l'adresse

www.unige.ch/math/folks/velenik/cours.html

Table des matières

Table des matières	i
1 Transition de phase dans le graphe d'Erdős–Rényi	1
1.1 Le modèle d'Erdős–Rényi	1
1.2 La transition de phase	1
1.3 Existence d'un chemin contenant presque tous les sommets	2
1.4 Préliminaires	2
1.5 Preuve du Théorème 1.1	6
1.6 Preuve du Théorème 1.2	8
2 Distance en variation totale et couplages	9
2.1 Distance en variation totale	9
2.2 Couplage	10
2.3 Quelques applications	12
3 Approximation de Poisson : la méthode de Chen–Stein	19
3.1 Théorème de Chen	19
3.2 Preuve du Théorème de Chen	20
3.3 Quelques exemples d'application	22
4 Concentration de la mesure : l'inégalité de Talagrand	29
4.1 Distance de Hamming pondérée	29
4.2 Inégalité de Talagrand	31
4.3 Preuve de l'inégalité de Talagrand	32
4.4 Quelques corollaires et applications	34
5 Réduction de la dimensionnalité : le lemme de Johnson–Lindenstrauss	41
5.1 Cadre et énoncé du Lemme de Johnson–Lindenstrauss	41
5.2 Preuve du Lemme de Johnson–Lindenstrauss	42
5.3 Application à l'acquisition comprimée	43
6 Transformée de Doob et marche aléatoire sur \mathbb{Z}	49
6.1 Transformée de Doob d'une chaîne de Markov	49
6.2 Quelques exemples	50
6.3 Marche aléatoire sur \mathbb{Z} conditionnée à rester positive	52
7 Convergence des chaînes de Markov réversibles	57

7.1	Structure algébrique	57
7.2	Vitesse de convergence	59
7.3	Temps de relaxation et temps de mélange	60
7.4	Caractérisation variationnelle du trou spectral	62
7.5	Bornes sur les trous spectraux γ et γ_*	63
8	Méthode de Monte Carlo et simulation parfaite	67
8.1	Méthode de Monte Carlo	67
8.2	Simulation parfaite	69
	Bibliographie	75
	Notations	77
	Index	79

1 Transition de phase dans le graphe d'Erdős–Rényi

Le but de ce chapitre est de présenter une preuve particulièrement simple de l'existence d'une transition de phase dans le modèle de graphe aléatoire d'Erdős–Rényi. La preuve fournit des estimées du bon ordre de grandeur, mais ne cherche pas à obtenir les constantes correctes (ce qui requiert des arguments plus sophistiqués).

1.1 Le modèle d'Erdős–Rényi

Dans ce chapitre, nous nous intéresserons au modèle de graphe aléatoire $G(n, p)$, $n \in \mathbb{N}^*$, $p \in [0, 1]$. Il s'agit d'une loi de probabilité sur l'ensemble des graphes ayant pour sommets l'ensemble $\llbracket n \rrbracket$. Une réalisation $G \sim G(n, p)$ du graphe aléatoire est alors obtenue de la façon suivante : indépendamment pour chacune des paires $\{i, j\} \subset \llbracket n \rrbracket$ de sommets distincts, $\{i, j\}$ est une arête de G avec probabilité p .

Ce modèle est généralement appelé le **graphe aléatoire d'Erdős–Rényi**, bien que cette version ait en réalité été introduite par Gilbert [10], la version introduite par Erdős et Rényi [9] étant légèrement différente (ils considèrent la loi uniforme sur tous les graphes avec sommets $\llbracket n \rrbracket$ possédant m arêtes).

Les propriétés géométriques d'une réalisation typique (connexité, tailles des composantes connexes, etc.) présentent des comportements très variés selon la valeur de la probabilité p de rétention d'arête. Dans ce chapitre, nous nous intéresserons principalement au comportement de la taille de la plus grande composante connexe. Cette dernière change de façon abrupte au voisinage de $p = 1/n$.

1.2 La transition de phase

Considérons à présent une réalisation $G \sim G(n, p)$, où l'on suppose que $p = c/n$ pour un $c \in \mathbb{R}$ fixé. Le résultat principal de ce chapitre est le Théorème 1.1, dans lequel on montre que le comportement de G change abruptement en $c = 1$:

- ▷ Lorsque $c < 1$, la plus grande composante connexe de G contient au plus $O(\log n)$ sommets avec grande probabilité¹.
- ▷ Lorsque $c > 1$, la plus grande composante connexe de G contient $O(n)$ sommets avec grande probabilité. En fait, le théorème ci-dessous montre même davantage : avec grande probabilité, la plus grande composante connexe de G contient un chemin² de longueur $O(n)$ lorsque $c > 1$.

1. Lorsqu'une propriété \mathcal{P} du graphe aléatoire $G \sim G(n, p)$ se produit avec probabilité $1 - o_{n \rightarrow \infty}(1)$, on dit que G possède la propriété \mathcal{P} **avec grande probabilité**. Cette terminologie s'étend à tout autre modèle aléatoire dépendant d'un paramètre $n \in \mathbb{N}$.

2. Dans ce chapitre, un **chemin de longueur** ℓ dans un graphe (S, A) est une collection de sommets tous distincts $i_1, \dots, i_\ell \in S$ tels que $\{i_k, i_{k+1}\} \in A$ pour tout $k \in \{1, \dots, \ell - 1\}$.

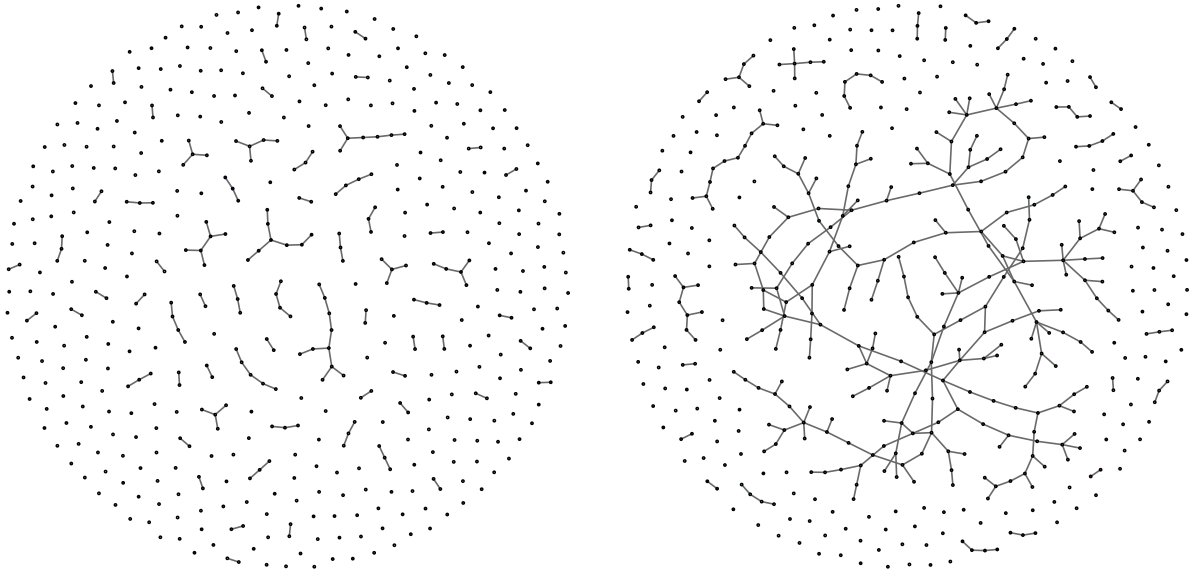


FIGURE 1.1: Deux réalisations du graphe aléatoire d'Erdős-Rényi avec $n = 500$. Gauche : phase sous-critique ($p = 0.5/n$). Droite : phase sur-critique ($p = 1.3/n$).

Ceci est formulé de façon précise dans le théorème suivant, dont la démonstration est donnée dans la Section 1.5.

Théorème 1.1. Soit $\epsilon > 0$ suffisamment petit. Soit $G \sim \mathcal{G}(n, p)$.

1. Soit $p = \frac{1-\epsilon}{n}$. Alors, avec grande probabilité, aucune composante connexe de G ne contient plus de $7\epsilon^{-2} \log n$ sommets.
2. Soit $p = \frac{1+\epsilon}{n}$. Alors, avec grande probabilité, G contient un chemin de longueur au moins $\frac{1}{5}\epsilon^2 n$.

1.3 Existence d'un chemin contenant presque tous les sommets

Comme seconde application des méthodes présentées dans ce chapitre, considérons à présent le graphe aléatoire $G \sim \mathcal{G}(n, p)$ dans le régime où $p = c/n$ avec c tendant vers l'infini. Nous avons vu qu'un chemin de longueur macroscopique (c'est-à-dire, de longueur $O(n)$) apparaît avec grande probabilité dès que $c > 1$. Nous allons voir à présent qu'en choisissant c suffisamment grand, il existe avec grande probabilité un chemin contenant une fraction des sommets aussi proche de 1 que l'on souhaite.

Théorème 1.2. Pour tout $\epsilon > 0$ suffisamment petit, il existe $C = C(\epsilon)$ tel que l'affirmation suivante soit vraie. Si $p = C/n$ et $G \sim \mathcal{G}(n, p)$, alors G contient un chemin de longueur au moins $(1 - \epsilon)n$ avec grande probabilité.

La preuve de ce résultat est donnée en Section 1.6.

1.4 Préliminaires

Cette section contient quelques outils techniques qui se révéleront utiles pour la preuve des Théorèmes 1.1 et 1.2 : tout d'abord, une discussion de l'algorithme de « parcours en profondeur » d'un graphe, qui joue un rôle central dans les preuves de ces théorèmes, puis des estimées sur des sommes de variables de Bernoulli indépendantes.

1.4.1 Exploration d'un graphe par l'algorithme de parcours en profondeur

Dans cette section, nous décrivons un algorithme d'exploration de graphe qui joue un rôle central dans la preuve des Théorèmes 1.1 et 1.2 : l'algorithme du « parcours en profondeur », plus connu sous son nom anglais de *Depth-First Search*, dont nous utiliserons l'acronyme DFS pour désigner l'algorithme dans le reste de ce chapitre.

DFS est un algorithme d'exploration des sommets d'un graphe fini $G = (S, A)$. On supposera les sommets de S ordonnés, disons $S := \llbracket n \rrbracket$. Lors de son application, on fait évoluer une partition de S en 3 ensembles disjoints :

- ▷ l'ensemble T des sommets traités;
- ▷ l'ensemble I des sommets inexplorés;
- ▷ la collection ordonnée P des sommets à traiter, qui est utilisée comme une pile (*stack* en anglais) fonctionnant en mode LIFO (pour *last in, first out*, soit « dernier entré, premier sorti »). On utilisera la notation $P = \langle a_1, \dots, a_k \rangle$ pour désigner une pile composée des éléments a_1, \dots, a_k ordonnés selon leur ordre d'arrivée sur la pile, a_k représentant ainsi l'élément se trouvant au sommet de la pile.

L'algorithme DFS prend la forme suivante :

Algorithme 1 : parcours en profondeur

```

Entrées : un graphe  $G = (S, A)$ , où  $S = \llbracket n \rrbracket$ 
Initialisation :  $T = \emptyset, P = \emptyset, I = S$ 
tant que  $P \cup I \neq \emptyset$  faire
  si  $P = \emptyset$  alors
    // Si la pile est vide, on retire le premier sommet appartenant à  $I$ 
    // et on le dépose sur la pile.
    soit  $i$  le premier sommet de  $I$  // selon l'ordre des sommets
    on retire  $i$  de l'ensemble  $I$ 
    on place  $i$  sur la pile  $P$ 
  fin
  soit  $v$  le sommet se trouvant sur la pile  $P$ 
  // On cherche le premier sommet  $u$  de  $I$  tel que  $\{v, u\} \in A$ .
  répéter
    | soit  $u$  le sommet suivant de  $I$  // selon l'ordre des sommets
  jusqu'à  $u$  est voisin de  $v$  ou  $u$  est le dernier sommet de  $I$ 
  // Si on a trouvé un tel sommet  $u$ ,
  si  $\{v, u\} \in A$  alors
    | on retire  $u$  de l'ensemble  $I$ 
    | on place  $u$  sur la pile  $P$ 
  sinon
    | on retire  $v$  de la pile  $P$ 
    | on ajoute  $v$  à l'ensemble  $T$ 
  fin
fin

```

Exemple 1.3. Afin d'être certain d'avoir bien compris comment cet algorithme fonctionne, considérons par exemple le graphe représenté sur la Figure 1.2. Voici l'évolution de l'algorithme lors de l'exploration de ce dernier (on indique le "temps" à gauche, ce dernier est incrémenté à chaque fois que la présence d'une nouvelle arête est testée) :

0 : Initialisation : $T = \emptyset, P = \emptyset, I = \{1, 2, 3, 4, 5, 6\}$.

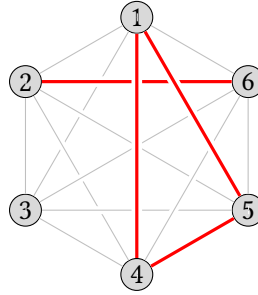


FIGURE 1.2: Le graphe discuté dans l'exemple 1.3.

On déplace 1 sur la pile : $T = \emptyset, P = |1\rangle, I = \{2, 3, 4, 5, 6\}$.

1,2 : On teste, sans succès, les paires suivantes : $\{1, 2\}, \{1, 3\}$.

3 : L'arête $\{1, 4\}$ appartenant au graphe, on déplace 4 sur la pile : $T = \emptyset, P = |1, 4\rangle, I = \{2, 3, 5, 6\}$.

4,5 : On teste, sans succès, les paires suivantes : $\{4, 2\}, \{4, 3\}$.

6 : L'arête $\{4, 5\}$ appartenant au graphe, on déplace 5 sur la pile : $T = \emptyset, P = |1, 4, 5\rangle, I = \{2, 3, 6\}$.

7,8,9 : On teste, sans succès, les paires suivantes : $\{5, 2\}, \{5, 3\}, \{5, 6\}$.

On n'a trouvé aucune arête, on déplace donc 5 dans T : $T = \{5\}, P = |1, 4\rangle, I = \{2, 3, 6\}$.

10 : On sait déjà que les arêtes $\{4, 2\}, \{4, 3\}$ ne sont pas dans le graphe. On teste donc, sans succès, la paire $\{4, 6\}$.

On n'a trouvé aucune arête, on déplace donc 4 dans T : $T = \{4, 5\}, P = |1\rangle, I = \{2, 3, 6\}$.

11 : On sait déjà que les arêtes $\{1, 2\}, \{1, 3\}$ ne sont pas dans le graphe. On teste donc, sans succès, la paire $\{1, 6\}$.

On n'a trouvé aucune arête, on déplace donc 1 dans T : $T = \{1, 4, 5\}, P = \emptyset, I = \{2, 3, 6\}$.

La pile est vide. On déplace donc 2 sur la pile : $T = \{1, 4, 5\}, P = |2\rangle, I = \{3, 6\}$.

12 : On teste, sans succès, la paire $\{2, 3\}$.

13 : L'arête $\{2, 6\}$ appartenant au graphe, on déplace 6 sur la pile : $T = \{1, 4, 5\}, P = |2, 6\rangle, I = \{3\}$.

14 : On teste, sans succès, la paire $\{6, 3\}$.

On n'a trouvé aucune arête, on déplace donc 6 dans T : $T = \{1, 4, 5, 6\}, P = |2\rangle, I = \{3\}$.

On sait déjà que l'arête $\{2, 3\}$ n'est pas dans le graphe.

On n'a trouvé aucune arête, on déplace donc 2 dans T : $T = \{1, 2, 4, 5, 6\}, P = \emptyset, I = \{3\}$.

La pile est vide. On déplace donc 3 sur la pile : $T = \{1, 2, 4, 5, 6\}, P = |3\rangle, I = \emptyset$.

Il n'y a plus de paires à tester. On déplace donc 3 dans T : $T = \{1, 2, 3, 4, 5, 6\}, P = \emptyset, I = \emptyset$.

On a exploré tous les sommets. L'algorithme s'arrête. \diamond

Faisons quelques observations sur l'exemple précédent qui restent manifestement vraies en général.

- ▷ À chaque instant, on sait déjà qu'aucun des sommets placés dans T ne possède de voisin dans I .
- ▷ La pile est vide précisément au moment où l'on a terminé d'explorer une composante connexe du graphe. On appellera **époque** la partie de l'algorithme se trouvant entre deux instants successifs où la pile est vide. Chaque époque correspond donc à l'exploration d'une composante connexe du graphe.
- ▷ Il suit de l'observation précédente que l'ensemble des sommets placés sur la pile appartiennent toujours à la même composante connexe (bien sûr, certains des sommets de cette composante ont déjà pu être déplacés dans T). Mais plus que cela est vrai : par construction ils forment toujours

un chemin. En effet, lorsqu'un nouveau sommet est placé sur la pile, ceci se produit parce que ce sommet est un voisin du sommet se trouvant précédemment au sommet de la pile, incrémentant d'une unité la longueur du chemin déjà présent dans la pile.

Notons que l'algorithme DFS décrit ci-dessus explore tous les sommets du graphe, mais pas nécessairement toutes ses arêtes. Par la suite, on souhaitera déterminer également ces dernières. On complètera donc l'exploration du graphe par une étape finale dans laquelle on teste, dans l'ordre, toutes les paires de sommets dont on n'a pas encore déterminé s'ils sont reliés par une arête. L'algorithme va ainsi demander l'état de chacune des $N := \binom{n}{2}$ arêtes du graphe complet à n sommets K_n .

Soit $(X_k)_{k=1}^N$ des variables aléatoires i.i.d. suivant chacune une loi de Bernoulli de paramètre p . Appliquer l'algorithme DFS à une réalisation du graphe aléatoire d'Erdős-Rényi, est équivalent à lui fournir la valeur prise par la variable X_k lors de sa k ième demande. Ainsi, toute la structure du graphe aléatoire est contenue dans la réalisation des variables aléatoires $(X_k)_{k=1}^N$. Nous allons voir, à présent, qu'une telle réalisation possède, avec probabilité tendant vers 1 lorsque $n \rightarrow \infty$, certaines propriétés garantissant que les affirmations des Théorèmes 1.1 et 1.2 sont vraies.

1.4.2 Quelques estimées sur les sommes de variables de Bernoulli indépendantes

Cette section est consacrée à quelques résultats élémentaires sur les variables aléatoires $(X_k)_{k=1}^N$. Les lemmes 1.7 et 1.8 forment la partie probabiliste de la preuve du Théorème 1.1, le reste de la preuve étant alors purement déterministe. Commençons par une borne standard sur la queue supérieure d'une variable binomiale.

Lemme 1.4. Soit $n \in \mathbb{N}^*$, $p \in [0, 1]$, $X \sim \text{Bin}(n, p)$. Alors,

$$\forall \delta > 0, \quad \mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]) \leq \exp\left(-\frac{\delta^2 pn}{2 + \delta}\right).$$

Démonstration. L'inégalité de Markov implique que, pour tout $t > 0$,

$$\mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]) = \mathbb{P}(X \geq (1 + \delta)np) = \mathbb{P}(e^{tX} \geq e^{t(1+\delta)np}) \leq e^{-t(1+\delta)np} \mathbb{E}(e^{tX}).$$

Évidemment, $X \stackrel{\text{loi}}{=} X_1 + \dots + X_n$, où X_1, \dots, X_n sont i.i.d. et $X_1 \sim \text{Bern}(p)$. Par conséquent, $\mathbb{E}(e^{tX}) = \mathbb{E}(e^{tX_1})^n = (pe^t + 1 - p)^n \leq e^{(e^t - 1)pn}$. On optimise à présent sur t en choisissant $t = \log(1 + \delta)$. On obtient $\mathbb{E}(e^{tX}) \leq e^{\delta pn}$ et donc

$$\mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]) \leq \exp\left(-((1 + \delta) \log(1 + \delta) - \delta)pn\right) \leq \exp\left(-\frac{\delta^2 pn}{2 + \delta}\right),$$

où l'on a utilisé l'inégalité $\log(1 + x) \geq \frac{x}{1 + \frac{1}{2}x}$, valide pour tout $x \geq 0$. □

Remarque 1.5. Observons qu'il sera essentiel dans ce qui suit d'avoir un exposant dépendant linéairement de p dans la borne précédente. En particulier, l'inégalité de Hoeffding donnerait dans le cas présent la borne

$$\mathbb{P}(X \geq (1 + \delta)\mathbb{E}[X]) \leq \exp\left(-2\delta^2 p^2 n\right),$$

ce qui ne serait pas suffisant. ◇

On peut également obtenir une borne similaire sur la queue inférieure (ce résultat n'est pas nécessaire dans ce chapitre, mais sera utilisé dans le Chapitre 4).

Lemme 1.6. Soit $n \in \mathbb{N}^*$, $p \in [0, 1]$, $X \sim \text{Bin}(n, p)$. Alors,

$$\forall \delta \in (0, 1), \quad \mathbb{P}(X \leq (1 - \delta)\mathbb{E}[X]) \leq \exp\left(-\frac{\delta^2 pn}{2}\right).$$

Démonstration. Pour tout $t > 0$,

$$\mathbb{P}(X \leq (1 - \delta)pn) = \mathbb{P}(e^{-tX} \geq e^{-t(1-\delta)pn}) \leq e^{t(1-\delta)pn} \mathbb{E}(e^{-tX}).$$

Or, $\mathbb{E}(e^{-tX}) = (e^{-t}p + 1 - p)^n \leq e^{-(1-e^{-t})pn}$. Par conséquent,

$$\mathbb{P}(X \leq (1 - \delta)pn) \leq e^{-((1-e^{-t})-t(1-\delta))pn}.$$

Optimiser sur t conduit à prendre $t = -\log(1 - \delta)$, ce qui donne

$$\mathbb{P}(X \leq (1 - \delta)pn) \leq e^{-(\delta+(1-\delta)\log(1-\delta))pn} \leq e^{-(1-\sqrt{1-\delta})\delta pn} \leq e^{-\delta^2 pn/2},$$

puisque $\log(1 - \delta) \geq -\delta/\sqrt{1 - \delta}$ et $1 - \sqrt{1 - \delta} \geq \delta/2$ pour tout $\delta \in (0, 1)$. \square

Lemme 1.7. Soit $\epsilon \in (0, \frac{1}{7})$ et $k := \lceil (7/\epsilon^2) \log n \rceil$. Soit $(X_k)_{k=1}^N$ des variables aléatoires i.i.d. suivant chacune une loi de Bernoulli de paramètre $p = (1 - \epsilon)/n$. Alors, avec grande probabilité,

$$\nexists a, b \in \llbracket N \rrbracket \text{ tels que } b - a + 1 = kn \text{ et } \sum_{i=a}^b X_i \geq k. \quad (1.1)$$

Démonstration. Pour tout $a, b \in \llbracket N \rrbracket$ tels que $b - a + 1 = kn$, $\sum_{i=a}^b X_i \sim \text{Bin}(kn, p)$. Soit donc $X \sim \text{Bin}(kn, p)$. Notons que $\mathbb{E}[X] = kpn = (1 - \epsilon)k \leq k/(1 + \epsilon)$. Il suit donc du Lemme 1.4 que

$$\mathbb{P}(X \geq k) \leq \mathbb{P}(X \geq (1 + \epsilon)\mathbb{E}[X]) \leq \exp(-\frac{1}{3}\epsilon^2(1 - \epsilon)k),$$

où l'on a utilisé le fait que $\epsilon < 1$. La conclusion suit facilement de l'inégalité de Boole :

$$\begin{aligned} \mathbb{P}\left(\exists a, b \in \llbracket N \rrbracket, b - a + 1 = kn, \sum_{i=a}^b X_i \geq k\right) &\leq (N - kn + 1)\mathbb{P}(X \geq k) \\ &\leq \frac{1}{2}n^2 \exp(-\frac{1}{3}\epsilon^2(1 - \epsilon)k) \leq \frac{1}{2}n^2 \exp(-\frac{7}{3}(1 - \epsilon)\log n) = o(1), \end{aligned}$$

puisque $\epsilon < 1/7$. \square

Lemme 1.8. Soit $\epsilon > 0$ et $N_0 := \lfloor \epsilon n^2/2 \rfloor$. Soit $(X_k)_{k=1}^N$ des variables aléatoires i.i.d. suivant chacune une loi de Bernoulli de paramètre $p = (1 + \epsilon)/n$. Alors, avec grande probabilité,

$$\left| \sum_{i=1}^{N_0} X_i - \frac{\epsilon(1 + \epsilon)n}{2} \right| \leq n^{2/3}. \quad (1.2)$$

Démonstration. Soit $X := \sum_{i=1}^{N_0} X_i \sim \text{Bin}(N_0, p)$. Alors, $\mathbb{E}[X] = N_0 p = \frac{1}{2}\epsilon(1 + \epsilon)n + O(1/n)$. L'inégalité de Tchebychev implique donc

$$\mathbb{P}\left(\left|X - \frac{\epsilon(1 + \epsilon)n}{2}\right| > n^{2/3}\right) \leq \mathbb{P}(|X - \mathbb{E}[X]| > \frac{1}{2}n^{2/3}) \leq \frac{4N_0 p(1 - p)}{n^{4/3}} \leq \frac{2\epsilon(1 + \epsilon)n}{n^{4/3}} = o(1). \quad \square$$

1.5 Preuve du Théorème 1.1

Soit $(X_i)_{i=1}^N$ la collection de variables aléatoires déterminant le graphe $G \sim \mathcal{G}(n, p)$ lorsqu'on explore ce dernier à l'aide de l'algorithme d'exploration DFS. Commençons par faire trois observations élémentaires :

- (O1) À chaque étape de l'algorithme, toutes les paires $\{i, j\}$ de sommets avec $i \in T$ et $j \in I$ ont été testées (négativement : les sommets ne sont pas reliés par une arête). En particulier, au temps t (c'est-à-dire après t paires testées), on a nécessairement $t \geq |T| \cdot |I|$.
- (O2) Les éléments de P appartenant toujours à une même composante connexe et le placement de n'importe quel sommet d'une composante connexe donnée, sauf le premier, résultant du test d'une paire de sommets ayant obtenu une réponse positive (la variable X_i correspondante prenant la valeur 1), on doit nécessairement avoir $|P| \leq 1 + \sum_{i=1}^t X_i$ au temps t .
- (O3) De plus, tant que $I \neq \emptyset$, chaque test de présence d'une arête recevant une réponse positive conduit à retirer un sommet de I et à le placer sur la pile P . Ce sommet rejoindra ultérieurement T , mais ne peut en aucun cas retourner dans I . Il suit que si l'on a toujours $I \neq \emptyset$ après t paires testées, alors on doit nécessairement avoir $|T \cup P| \geq \sum_{i=1}^t X_i$.

Nous sommes à présent en mesure de démontrer les deux affirmations du théorème. Nous supposons dans la preuve que ϵ est choisi suffisamment petit et n suffisamment grand.

1.5.1 Régime sous-critique

Par le Lemme 1.7, nous pouvons supposer que les variables aléatoires $(X_i)_{i=1}^N$ satisfont (1.1). Nous allons voir que la première affirmation du théorème suit alors de façon déterministe.

Supposons, par l'absurde, que G contienne une composante connexe C constituée de plus de $k := \lceil (7/\epsilon^2) \log n \rceil$ sommets. Concentrons-nous sur l'époque durant laquelle C est explorée et considérons en particulier l'instant où le $(k+1)$ ième sommet de C est découvert, juste avant qu'il soit déplacé sur la pile P .

Notons $T_C := T \cap C$ l'ensemble des sommets de C ayant déjà rejoint l'ensemble T des sommets traités. Alors, $|T_C \cup P| = k$, ce qui signifie que l'algorithme a reçu exactement k réponses positives durant cette époque jusqu'à cet instant (chacune révélant un nouveau sommet de C ; seul le premier sommet de C a été ajouté à P au début de l'époque, sans nécessiter de test).

Observons qu'au cours de cette époque et jusqu'à cet instant, les paires testées contiennent toutes au moins un sommet de $T_C \cup P$. Or, le nombre de telles paires est au plus égal à $\binom{k}{2} + k(n-k) < kn$.

Il suit des observations précédentes qu'on a trouvé un intervalle $J \subset \llbracket N \rrbracket$ de longueur au plus kn et tel que $\sum_{i \in J} X_i \geq k$. Mais cela contredit notre hypothèse que la condition (1.1) est satisfaite.

1.5.2 Régime sur-critique

Par le Lemme 1.8, nous pouvons supposer que les variables aléatoires $(X_i)_{i=1}^N$ satisfont (1.2). La seconde affirmation du théorème en est, à nouveau, une conséquence déterministe.

Nous allons vérifier qu'après les premiers $N_0 := \lfloor \epsilon n^2 / 2 \rfloor$ tests de paires de sommets, la pile P contient au moins $\frac{1}{5} \epsilon^2 n$ sommets (ces sommets formant nécessairement un chemin, comme on l'a vu précédemment).

Observons tout d'abord que $|T| < n/3$ au temps N_0 . En effet, supposons par l'absurde que ce ne soit pas le cas. On peut alors considérer le temps t auquel $|T| = \lceil n/3 \rceil$ (qui est bien défini, puisque T croît d'un sommet à la fois). À cet instant, comme nous avons supposé la propriété (1.2) satisfaite, il suit de l'observation (O2) que $|P| \leq 1 + \sum_{i=1}^t X_i \leq 1 + \sum_{i=1}^{N_0} X_i < n/3 - 1$, d'où l'on déduit que $|I| = n - |T| - |P| \geq n/3$. De plus, par l'observation (O1), on a testé au temps t au moins $|T| \cdot |I| \geq n^2/9 > N_0$ paires de sommets, ce qui est impossible puisque $t \leq N_0$.

Retournons à la situation au temps N_0 . Supposons par l'absurde que $|P| < \frac{1}{5} \epsilon^2 n$. Puisque $|T| < n/3$, ceci implique que $I \neq \emptyset$. En particulier, l'algorithme est encore en train d'explorer les sommets de G (il n'est pas encore entré dans la phase finale lors de laquelle on ne cherche plus qu'à déterminer l'ensemble des arêtes). En utilisant à nouveau notre hypothèse (1.2), on sait que le nombre de réponses positives au temps N_0 est supérieur ou égal à $\frac{1}{2} \epsilon (1 + \epsilon) n - n^{2/3}$. Par l'observation (O3), on a donc $|T \cup P| \geq \frac{1}{2} \epsilon (1 + \epsilon) n - n^{2/3}$ au temps N_0 . Comme on a supposé que $|P| < \frac{1}{5} \epsilon^2 n$, on conclut que $|T| \geq$

$\frac{1}{2}\epsilon n + \frac{3}{10}\epsilon^2 n - n^{2/3}$. En utilisant à nouveau l'observation (O1), on a $N_0 \geq |T| \cdot |I| \geq |T|(n - |T| - \frac{1}{5}\epsilon^2 n)$. On obtient ainsi

$$\frac{\epsilon n^2}{2} \geq N_0 \geq |T| \left(n - |T| - \frac{\epsilon^2 n}{5} \right).$$

Le second facteur du membre de droite est positif, puisque $|T| < n/3$. Le membre de droite restreint aux valeurs $T \in [\frac{1}{2}\epsilon n + \frac{3}{10}\epsilon^2 n - n^{2/3}, \frac{1}{3}n]$ atteint son minimum en $|T| = \frac{1}{2}\epsilon n + \frac{3}{10}\epsilon^2 n - n^{2/3}$. On a donc

$$N_0 \geq \left(\frac{\epsilon n}{2} + \frac{3\epsilon^2 n}{10} - n^{2/3} \right) \left(n - \frac{\epsilon n}{2} - \frac{\epsilon^2 n}{2} + n^{2/3} \right) = \frac{\epsilon n^2}{2} + \frac{\epsilon^2 n^2}{20} + O(\epsilon^3)n^2 + o(n^2) > \frac{\epsilon n^2}{2},$$

ce qui est incompatible avec le fait que $N_0 = \lfloor \epsilon n^2/2 \rfloor$. On en conclut que $|P| \geq \frac{1}{5}\epsilon^2 n$, ce qui termine la preuve. \square

1.6 Preuve du Théorème 1.2

Commençons par établir le résultat élémentaire (et déterministe) suivant.

Lemme 1.9. Soit $n > k > 0$ deux entiers. Soit $G = (S, A)$ un graphe avec $|S| = n$ sommets et tel que pour toute paire de sous-ensembles disjoints $S_1, S_2 \subset S$ de taille k , on peut trouver une arête $\{i, j\} \in A$ avec $i \in S_1$ et $j \in S_2$. Alors, G contient un chemin de longueur $n - 2k + 2$.

Démonstration. On applique l'algorithme d'exploration DFS au graphe G . Considérons l'étape d'exécution de cet algorithme à laquelle $|T| = |I|$. Observons que cela se produit nécessairement, puisque chaque déplacement d'un sommet se fait soit de I vers P , soit de P vers T (dans les deux cas, $|I| - |T|$ décroît d'une unité). Nous savons également que G ne contient jamais d'arête reliant un sommet de T à un sommet de I . Il suit donc de l'hypothèse du lemme que $|T| = |I| \leq k - 1$ à cet instant. Comme $|T| + |P| + |I| = n$, on en conclut que $|P| \geq n - 2k + 2$. La conclusion suit, puisque les sommets de P forment toujours un chemin. \square

Afin de démontrer le Théorème 1.2, il nous suffit donc de prouver que $G \sim \mathbb{G}(n, p)$ contient, avec grande probabilité, une arête entre toute paire de sous-ensembles disjoints $S_1, S_2 \subset S$ de taille $k := \lceil \epsilon n/2 \rceil$. Soit donc $S_1, S_2 \subset S$ tels que $S_1 \cap S_2 = \emptyset$ et $|S_1| = |S_2| = k$. La probabilité que G ne contienne aucune des $|S_1| \cdot |S_2| = k^2$ arêtes reliant ces deux ensembles est égale à $(1 - p)^{k^2}$. Par conséquent, une application de l'inégalité de Boole nous donne que la probabilité de trouver une telle paire de sous-ensembles est bornée supérieurement par

$$\binom{n}{k} \binom{n-k}{k} (1-p)^{k^2} < \binom{n}{k}^2 (1-p)^{k^2} < \left(\frac{\epsilon n}{k} \right)^{2k} e^{-pk^2} < \left(\left(\frac{\epsilon n}{k} \right)^2 e^{-Ck/n} \right)^k,$$

où l'on a utilisé les inégalités élémentaires $1 - p \leq e^{-p}$ et $\binom{n}{k} < \left(\frac{\epsilon n}{k} \right)^k$. En choisissant $C := 5|\log \epsilon|/\epsilon$, on a $(\epsilon n/k)^2 e^{-Ck/n} \leq 4e^2 \epsilon^{1/2} < 1$ dès que $\epsilon < (2e)^{-4}$. On en conclut que la probabilité tend vers 0 exponentiellement rapidement avec n . \square

Remarques bibliographiques : Ce chapitre est basé sur l'article [14].

3. Il suffit d'observer que $\frac{e^k}{k^k} = \frac{1}{k^k} \sum_{\ell=0}^{\infty} \frac{k^\ell}{\ell!} > \frac{1}{k^k} \frac{k^k}{k!} = \frac{1}{k!}$. Par conséquent, $\binom{n}{k} \leq \frac{n^k}{k!} \leq \left(\frac{\epsilon n}{k} \right)^k$.

2 Distance en variation totale et couplages

Dans tout ce chapitre, nous considérons un ensemble Ω dénombrable (fini ou infini). On note $\mathcal{M}_1(\Omega)$ l'ensemble des mesures de probabilité sur $(\Omega, \mathcal{P}(\Omega))$.

2.1 Distance en variation totale

La distance en variation totale fournit une notion de proximité uniforme entre deux mesures de probabilité.

Définition 2.1. Soit $\mu, \nu \in \mathcal{M}_1(\Omega)$. La **distance en variation totale** entre μ et ν est

$$\|\mu - \nu\|_{\text{VT}} := \sup_{A \subset \Omega} |\mu(A) - \nu(A)|. \quad (2.1)$$

Lemme 2.2. La distance en variation totale définit une métrique sur $\mathcal{M}_1(\Omega)$.

Démonstration. Laissée en exercice. □

Le lemme suivant donne quelques expressions équivalentes pour cette quantité.

Lemme 2.3. Pour tout $\mu, \nu \in \mathcal{M}_1(\Omega)$,

$$\|\mu - \nu\|_{\text{VT}} = \sup_{A \subset \Omega} (\mu(A) - \nu(A)) = \frac{1}{2} \sum_{i \in \Omega} |\mu(i) - \nu(i)|.$$

De plus, le supremum dans (2.1) est atteint lorsque $A = E := \{i \in \Omega \mid \mu(i) \geq \nu(i)\}$.

Démonstration. La première identité suit immédiatement de l'observation que

$$|\mu(A) - \nu(A)| = \max\{\mu(A) - \nu(A), \mu(A^c) - \nu(A^c)\}.$$

Passons aux autres affirmations. Soit $E := \{i \in \Omega \mid \mu(i) \geq \nu(i)\}$. Alors, pour tout $A \subset \Omega$,

$$\mu(A) - \nu(A) = \sum_{i \in A} (\mu(i) - \nu(i)) \leq \sum_{i \in A \cap E} (\mu(i) - \nu(i)) \leq \sum_{i \in E} (\mu(i) - \nu(i)) = \mu(E) - \nu(E),$$

avec égalité lorsque $A = E$, ce qui montre que $\|\mu - \nu\|_{\text{VT}} = \mu(E) - \nu(E)$. Finalement, la seconde identité suit de

$$\mu(E) - \nu(E) = \frac{1}{2} (\mu(E) - \nu(E) + \nu(E^c) - \mu(E^c)) = \frac{1}{2} \sum_{i \in \Omega} |\mu(i) - \nu(i)|. \quad \square$$

Il est clair de la définition que $\|\mu - \nu\|_{\text{VT}} \in [0, 1]$ et que $\|\mu - \nu\|_{\text{VT}} = 0$ si et seulement si $\mu = \nu$. Le lemme suivant implique, entre autre, que $\|\mu - \nu\|_{\text{VT}} = 1$ si et seulement si les supports de μ et ν sont disjoints.

Lemme 2.4. *Pour tout $\mu, \nu \in \mathcal{M}_1(\Omega)$, $\|\mu - \nu\|_{\text{VT}} = 1 - \sum_{i \in \Omega} \mu(i) \wedge \nu(i)$.*

Démonstration. Puisque $x \wedge y = \frac{1}{2}(x + y - |x - y|)$ pour tout $x, y \in \mathbb{R}$, on a

$$\sum_{i \in \Omega} \mu(i) \wedge \nu(i) = \frac{1}{2} \sum_{i \in \Omega} (\mu(i) + \nu(i) - |\mu(i) - \nu(i)|) = 1 - \|\mu - \nu\|_{\text{VT}}. \quad \square$$

2.2 Couplage

Rappelons la notion de couplage, qui a probablement déjà été présentée dans le cours d'introduction.

Définition 2.5. *Soit $\mu, \nu \in \mathcal{M}_1(\Omega)$. Un **couplage** des mesures μ et ν est une mesure de probabilité $\rho \in \mathcal{M}_1(\Omega \times \Omega)$ dont les marginales coïncident avec μ et ν :*

$$\forall A \subset \Omega, \quad \rho(A \times \Omega) = \mu(A), \quad \rho(\Omega \times A) = \nu(A).$$

*Si X et Y sont deux variables aléatoires (pas nécessairement définies sur un même espace de probabilité) toutes deux à valeurs dans Ω , un **couplage** de X et Y est une paire de variables aléatoires (définies sur un même espace de probabilité) (X', Y') dont la loi conjointe est un couplage des lois de X et de Y . Plus généralement, on dit que (X', Y') est un **couplage** de μ et ν si la loi conjointe de (X', Y') est un couplage de μ et ν .*

Exemple 2.6. Soit $0 \leq p \leq q \leq 1$. On considère les variables aléatoires $X \sim \text{Bern}(p)$ et $Y \sim \text{Bern}(q)$. On considère deux variables aléatoires indépendantes $X' \sim \text{Bern}(p)$ et $Y' \sim \text{Bern}(q)$. Alors, (X', Y') est un couplage de X et Y , appelé **couplage indépendant**. On a

$$(\mathbb{P}(X' = i, Y' = j))_{i, j \in \{0, 1\}} = \begin{pmatrix} (1-p)(1-q) & (1-p)q \\ p(1-q) & pq \end{pmatrix}. \quad \diamond$$

Évidemment, considérer la mesure produit, comme dans l'exemple précédent, fournit toujours un couplage (le couplage indépendant). Il est cependant généralement plus intéressant de considérer des couplages possédant des propriétés supplémentaires. Une forme fréquemment utilisée est celle de couplage monotone.

Définition 2.7. *Soit (X, Y) un couplage des lois μ et ν , où X et Y sont à valeurs dans un ensemble muni d'un ordre partiel \geq . Le couplage est dit **monotone** si $\mathbb{P}(X \leq Y) = 1$.*

Exemple 2.8. Considérons à nouveau $X \sim \text{Bern}(p)$ et $Y \sim \text{Bern}(q)$ avec $0 \leq p \leq q \leq 1$. Soit U une variable aléatoire uniforme sur $[0, 1]$. On considère les variables aléatoires $X'' := \mathbf{1}_{\{U \leq p\}}$ et $Y'' := \mathbf{1}_{\{U \leq q\}}$. Alors, (X'', Y'') est un couplage monotone de X et Y :

$$(\mathbb{P}(X'' = i, Y'' = j))_{i, j \in \{0, 1\}} = \begin{pmatrix} 1-q & q-p \\ 0 & p \end{pmatrix}. \quad \diamond$$

Exemple 2.9. Soit $0 \leq p \leq q \leq 1$ et $n \in \mathbb{N}^*$. Soit $G_p \sim \text{G}(n, p)$ et $G_q \sim \text{G}(n, q)$ des graphes aléatoires d'Erdős-Rényi (cf. Chapitre 1). G_p et G_q sont des sous-graphes aléatoires du graphe complet $K_n = ([n], A_n)$, où $A_n := \{\{i, j\} \mid 1 \leq i < j \leq n\}$. Ces sous-graphes sont naturellement partiellement ordonnés par la relation d'inclusion. Nous allons étendre le couplage monotone de l'exemple précédent aux graphes aléatoires G_p et G_q .

Soit $(U_e)_{e \in A_n}$ une collection de variables aléatoires i.i.d., uniformes sur $[0, 1]$. Soit G'_p , resp. G'_q , le sous-graphe de K_n obtenu en conservant tous les sommets et chaque arête $e \in A_n$ telle que $U_e \leq p$, resp. $U_e \leq q$. Alors (G'_p, G'_q) est un couplage de G_p et G_q tel que

$$\mathbb{P}(G'_p \subset G'_q) = 1.$$

L'existence d'un tel couplage fournit immédiatement des informations intéressantes, pas toujours faciles à établir directement. Soit f une fonction définie sur les sous-graphes de K_n et à valeur dans \mathbb{R} . On suppose f croissante : pour toute paire de sous-graphes G, G' de K_n , $G \subset G' \implies f(G) \leq f(G')$. On peut, par exemple, considérer la fonction $f(G) := \mathbf{1}_{\{G \text{ est connexe}\}}$ ou encore la fonction $f(G) := \max\{|C| \mid C \text{ composante connexe maximale de } G\}$. Soit $G \sim \mathbf{G}(n, p)$. Alors, pour une telle fonction f , le couplage précédent implique que l'espérance de $f(G)$ est croissante en p :

$$\forall p \leq q, \quad \mathbb{E}[f(G_p)] = \mathbb{E}[f(G'_p)] \leq \mathbb{E}[f(G'_q)] = \mathbb{E}[f(G_q)]. \quad \diamond$$

Tout couplage entre deux mesures de probabilité fournit une borne sur leur distance en variation totale.

Lemme 2.10. Soit $\mu, \nu \in \mathcal{M}_1(\Omega)$. Pour tout couplage (X, Y) de μ et ν , on a

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

Démonstration. Pour tout $A \subset \Omega$,

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

La conclusion suit donc du Lemme 2.3. □

Exemple 2.11. Soit $\lambda > \nu > 0$. On souhaite borner la distance en variation totale entre les lois des variables aléatoires $X \sim \text{Poisson}(\lambda)$ et $Y \sim \text{Poisson}(\nu)$. On va le faire en construisant un couplage de X et Y .

Soit $Y' \sim \text{Poisson}(\nu)$ et $Z' \sim \text{Poisson}(\lambda - \nu)$ deux variables aléatoires indépendantes. On pose $X' := Y' + Z'$. Il suit du cours d'introduction à la théorie des probabilités que $X' \sim \text{Poisson}(\lambda)$. Par conséquent, (X', Y') est un couplage de X et Y et le Lemme 2.10 implique que

$$\|\mathcal{L}(X) - \mathcal{L}(Y)\|_{\text{VT}} \leq \mathbb{P}(X' \neq Y') = \mathbb{P}(Z' \neq 0) = 1 - e^{-(\lambda - \nu)}. \quad \diamond$$

Le résultat suivant montre que la borne du Lemme 2.10 est optimale.

Lemme 2.12. Soit $\mu, \nu \in \mathcal{M}_1(\Omega)$. Il existe un couplage (X, Y) de μ et ν tel que

$$\|\mu - \nu\|_{\text{VT}} = \mathbb{P}(X \neq Y).$$

Un tel couplage est appelé un **couplage maximal** de μ et ν .

Démonstration. Les cas $\|\mu - \nu\|_{\text{VT}} \in \{0, 1\}$ sont triviaux et sont laissés en exercice ; on suppose donc $\|\mu - \nu\|_{\text{VT}} \in (0, 1)$. Considérons à nouveau $E := \{i \in \Omega \mid \mu(i) \geq \nu(i)\}$. Par le Lemme 2.4,

$$\tilde{p} := \sum_{i \in \Omega} \mu(i) \wedge \nu(i) = 1 - \|\mu - \nu\|_{\text{VT}}.$$

En particulier,

$$\sum_{i \in E} (\mu(i) - \nu(i)) = \sum_{i \notin E} (\nu(i) - \mu(i)) = \|\mu - \nu\|_{\text{VT}} = 1 - \tilde{p}.$$

Ainsi,

- ▷ $\tilde{p} \in (0, 1)$;
- ▷ $\rho_E(\cdot) := \frac{1}{1-\tilde{p}}(\mu(\cdot) - \nu(\cdot))$ définit une mesure de probabilités sur E ;
- ▷ $\rho_{E^c}(\cdot) := \frac{1}{1-\tilde{p}}(\nu(\cdot) - \mu(\cdot))$ définit une mesure de probabilités sur E^c ;
- ▷ $\rho_{\min}(\cdot) := \frac{1}{\tilde{p}}(\mu(\cdot) \wedge \nu(\cdot))$ définit une mesure de probabilités sur Ω .

On peut à présent définir le couplage de la façon suivante :

- ▷ Avec probabilité \tilde{p} , on tire au hasard la valeur de $X = Y$ dans Ω selon la loi ρ_{\min} .
- ▷ Sinon, on tire X au hasard dans E selon la loi ρ_E et on tire, indépendamment, Y au hasard dans E^c selon la loi ρ_{E^c} .

Vérifions que (X, Y) est un couplage de μ et ν .

$$\forall i \in \Omega, \quad \mathbb{P}(X = i) = \tilde{p}\rho_{\min}(i) + (1 - \tilde{p})\rho_E(i)\mathbf{1}_{\{i \in E\}} = (\mu(i) \wedge \nu(i)) + (\mu(i) - \nu(i))\mathbf{1}_{\{i \in E\}}.$$

En considérant séparément les cas $i \in E$ et $i \notin E$, on en déduit aisément que $\mathbb{P}(X = i) = \mu(i)$ pour tout $i \in \Omega$, c'est-à-dire que $X \sim \mu$. Un calcul similaire montre que $Y \sim \nu$.

Finalement, ce couplage est maximal, puisque $\mathbb{P}(X \neq Y) = 1 - \tilde{p} = \|\mu - \nu\|_{\text{VT}}$. \square

Exemple 2.13. Retournons au cas de deux variables aléatoires indépendantes $X \sim \text{Bern}(p)$ et $Y \sim \text{Bern}(q)$, avec $0 < p \leq q < 1$. On cherche à construire un couplage maximal (X''', Y''') des lois de X et Y . En reprenant les notations de la preuve du Lemme 2.12, on a $\Omega := \{0, 1\}$, $E := \{0\}$, $E^c := \{1\}$, $\tilde{p} := (1 - q) + p$, $\rho_E(0) := 1$, $\rho_{E^c}(1) := 1$ et $(\rho_{\min}(i))_{i \in \Omega} := ((1 - q)/(1 - q + p), p/(1 - q + p))$. On en conclut que

$$(\mathbb{P}(X''' = i, Y''' = j))_{i, j \in \{0, 1\}} = \begin{pmatrix} \tilde{p}\rho_{\min}(0) & (1 - \tilde{p})\rho_E(0)\rho_{E^c}(1) \\ 0 & \tilde{p}\rho_{\min}(1) \end{pmatrix} = \begin{pmatrix} 1 - q & q - p \\ 0 & p \end{pmatrix},$$

ce qui n'est rien d'autre que le couplage monotone de l'Exemple 2.8. \diamond

2.3 Quelques applications

2.3.1 Distribution asymptotique des degrés dans le graphe aléatoire d'Erdős-Rényi

Soit $n \in \mathbb{N}^*$ et $c \in \mathbb{R}_+^*$. On considère le graphe aléatoire d'Erdős-Rényi $G \sim \mathcal{G}(n, p_n)$ avec n sommets et probabilité d'occupation

$$p_n := \frac{c}{n}.$$

Pour chaque $i \in \llbracket n \rrbracket$, notons D_i le degré du sommet i (c'est-à-dire, le nombre de voisins de i dans G) et

$$N_d := \sum_{i=1}^n \mathbf{1}_{\{D_i=d\}}$$

le nombre de sommets de degré d . Dans cette section, on s'intéresse à la fraction asymptotique de sommets de degré d , N_d/n . Observons que les variables aléatoires $\mathbf{1}_{\{D_i=d\}}$ sont identiquement distribuées, mais pas indépendantes. On ne peut donc pas simplement recourir à la loi des grands nombres.

Théorème 2.14. Soit $c \in \mathbb{R}_+^*$ et $d \in \mathbb{N}$. Alors,

$$\frac{N_d}{n} \xrightarrow{n \rightarrow \infty} \frac{c^d}{d!} e^{-c} \quad \text{en probabilité.}$$

Démonstration. Les variables aléatoires D_i étant toutes de loi Binom($n-1, p_n$), on a

$$\mathbb{E}\left[\frac{N_d}{n}\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{P}(D_i = d) = \mathbb{P}(D_1 = d) = \frac{c^d}{d!} \frac{(n-1)!}{(n-1-d)!} n^{-d} \left(1 - \frac{c}{n}\right)^{n-1-d},$$

pour tout $d \in \{0, \dots, n-1\}$. Évidemment, pour tout $d \in \mathbb{N}$ fixé,

$$\lim_{n \rightarrow \infty} \frac{(n-1)!}{(n-1-d)!} n^{-d} = \lim_{n \rightarrow \infty} \prod_{k=1}^d \frac{n-k}{n} = 1 \quad \text{et} \quad \lim_{n \rightarrow \infty} \left(1 - \frac{c}{n}\right)^{n-1-d} = e^{-c},$$

ce qui implique que

$$\lim_{n \rightarrow \infty} \mathbb{E}\left[\frac{N_d}{n}\right] = \frac{c^d}{d!} e^{-c}. \quad (2.2)$$

Il nous suffit donc de montrer que la variable aléatoire N_d/n est concentrée proche de son espérance. Pour ce faire, estimons sa variance :

$$\begin{aligned} \text{Var}\left[\frac{N_d}{n}\right] &= \frac{1}{n^2} \left\{ \mathbb{E}\left[\left(\sum_{i=1}^n \mathbf{1}_{\{D_i=d\}}\right)^2\right] - \mathbb{E}\left[\sum_{i=1}^n \mathbf{1}_{\{D_i=d\}}\right]^2 \right\} \\ &= \frac{1}{n^2} \left\{ n\mathbb{P}(D_1 = d) + n(n-1)\mathbb{P}(D_1 = d, D_2 = d) - n^2\mathbb{P}(D_1 = d)^2 \right\} \\ &\leq \frac{1}{n} + \mathbb{P}(D_1 = d, D_2 = d) - \mathbb{P}(D_1 = d)^2. \end{aligned}$$

Notons que $\mathbb{P}(D_1 = d)^2 = \mathbb{P}(D_1 = d, D'_2 = d)$, où $D'_2 \sim \text{Binom}(n-1, p_n)$ est indépendante de D_1 . La différence entre les lois conjointes de (D_1, D_2) et (D_1, D'_2) est la présence de l'arête $\{1, 2\}$ dont l'état affecte à la fois D_1 et D_2 . Ceci suggère d'utiliser le couplage suivant. Soit Y_1, Y_2, X_1 et X_2 des variables aléatoires indépendantes telles que $Y_1 \sim \text{Binom}(n-2, p_n)$, $Y_2 \sim \text{Binom}(n-2, p_n)$, $X_1 \sim \text{Bern}(p_n)$ et $X_2 \sim \text{Bern}(p_n)$. Alors, $(D_1, D_2) \stackrel{\text{loi}}{=} (Y_1 + X_1, Y_2 + X_1)$ et $(D_1, D'_2) \stackrel{\text{loi}}{=} (Y_1 + X_1, Y_2 + X_2)$. On a donc

$$\begin{aligned} \mathbb{P}(D_1 = d, D_2 = d) - \mathbb{P}(D_1 = d)^2 &= \mathbb{P}((Y_1 + X_1, Y_2 + X_1) = (d, d)) - \mathbb{P}((Y_1 + X_1, Y_2 + X_2) = (d, d)) \\ &\leq \mathbb{P}((Y_1 + X_1, Y_2 + X_1) = (d, d), (Y_1 + X_1, Y_2 + X_2) \neq (d, d)) \\ &= \mathbb{P}(X_1 = 0, Y_1 = Y_2 = d, X_2 = 1) + \mathbb{P}(X_1 = 1, Y_1 = Y_2 = d-1, X_2 = 0) \\ &\leq \mathbb{P}(X_2 = 1) + \mathbb{P}(X_1 = 1) = \frac{2c}{n}. \end{aligned}$$

On obtient ainsi $\text{Var}[N_d/n] \leq (1 + 2c)/n$ et donc, par l'inégalité de Tchebychev,

$$\forall \epsilon > 0, \quad \mathbb{P}\left(\left|\frac{N_d}{n} - \mathbb{E}\left[\frac{N_d}{n}\right]\right| \geq \epsilon\right) \leq \frac{1 + 2c}{\epsilon^2 n}. \quad (2.3)$$

La conclusion suit de (2.2) et (2.3). \square

2.3.2 Le couplage d'Ornstein

Dans cette section, nous présentons un couplage dû à Ornstein, avec application aux marches aléatoires sur \mathbb{Z}^d . Ce couplage peut s'appliquer également dans des contextes substantiellement plus généraux.

Marche aléatoire sur \mathbb{Z}

Soit $(X_k)_{k \in \mathbb{N}^*}$ une collection de variables aléatoires i.i.d. à valeurs dans \mathbb{Z} et soit $a \in \mathbb{Z}$. On note $S_n := a + \sum_{k=1}^n X_k$ la marche aléatoire sur \mathbb{Z} correspondante, partant du point a . On supposera que la loi des incréments X_k est **fortement apériodique** :

$$\exists i_* \in \mathbb{Z}, \quad \mathbb{P}(X_1 = i_*) > 0 \text{ et } \text{pgcd}\{j \in \mathbb{Z} \mid \mathbb{P}(X_1 - i_* = j) > 0\} = 1, \quad (2.4)$$

où l'on a noté, pour tout $A \subset \mathbb{Z}$, $\text{pgcd } A := \max\{n \in \mathbb{N}^* \mid \forall m \in A, m/n \in \mathbb{Z}\}$.

Remarque 2.15. Cette condition exclut, par exemple, la marche aléatoire simple sur \mathbb{Z} , puisque l'on a $\text{pgcd}\{j \in \mathbb{Z} \mid \mathbb{P}(X_1 - 1 = j) > 0\} = \{0, -2\} = 2$. \diamond

Nous allons montrer le résultat de perte de mémoire suivant.

Théorème 2.16. Soit $a \in \mathbb{Z}$. Soit $(X_k)_{k \in \mathbb{N}^*}$ et $(X'_k)_{k \in \mathbb{N}^*}$ des variables aléatoires i.i.d. à valeurs dans \mathbb{Z} . On suppose que (X_k) et (X'_k) satisfaisant chacune (2.4). Soit $S_n := \sum_{k=1}^n X_k$ et $S'_n := a + \sum_{k=1}^n X'_k$. Alors,

$$\lim_{n \rightarrow \infty} \|\mathcal{L}(S_n) - \mathcal{L}(S'_n)\|_{\text{VT}} = 0.$$

Démonstration. Nous allons construire un couplage adéquat de S_n et S'_n en couplant les trajectoires des deux marches aléatoires. Soit i_* comme dans (2.4) et soit c suffisamment grand pour que

$$\text{pgcd}\{j \in \mathbb{Z} \mid \mathbb{P}(X_1 - i_* = j, |X_1 - i_*| \leq c) > 0\} = 1. \quad (2.5)$$

Définissons

$$X''_k := \begin{cases} X'_k & \text{si } |X_k - X'_k| \leq c, \\ X_k & \text{si } |X_k - X'_k| > c, \end{cases}$$

et considérons la marche aléatoire $S''_n := a + \sum_{k=1}^n X''_k$. Faisons quelques observations. Tout d'abord, comme $(X_k, X'_k) \stackrel{\text{loi}}{=} (X'_k, X_k)$, on a

$$\begin{aligned} \mathbb{P}(X''_k = j) &= \mathbb{P}(X'_k = j, |X_k - X'_k| \leq c) + \mathbb{P}(X_k = j, |X_k - X'_k| > c) \\ &= \mathbb{P}(X'_k = j, |X_k - X'_k| \leq c) + \mathbb{P}(X'_k = j, |X_k - X'_k| > c) = \mathbb{P}(X'_k = j), \end{aligned}$$

ce qui implique que $(S''_k)_{k \in \mathbb{N}} \stackrel{\text{loi}}{=} (S'_k)_{k \in \mathbb{N}}$ (rappelons que les deux marches partent de a par définition).

Considérons à présent la marche aléatoire $(R_n)_{n \in \mathbb{N}}$ sur \mathbb{Z} définie par

$$\forall n \in \mathbb{N}, \quad R_n := S_n - S''_n.$$

Manifestement, $R_n = -a + \sum_{k=1}^n (X_k - X''_k)$ et ses incréments sont bornés,

$$\mathbb{P}(|X_1 - X''_1| > c) = \mathbb{P}(|X_1 - X'_1| > c, |X_1 - X'_1| \leq c) + \mathbb{P}(|X_1 - X_1| > c, |X_1 - X'_1| > c) = 0,$$

et symétriques,

$$\begin{aligned} \forall j \neq 0, \quad \mathbb{P}(X_1 - X''_1 = j) &= \mathbb{P}(X_1 - X'_1 = j, |X_1 - X'_1| \leq c) \\ &= \mathbb{P}(X'_1 - X_1 = j, |X_1 - X'_1| \leq c) = \mathbb{P}(X_1 - X''_1 = -j). \end{aligned}$$

De plus,

$$\begin{aligned} \mathbb{P}(X_1 - X''_1 = j) &\geq \mathbb{P}(X_1 - X'_1 = j, |X_1 - X'_1| \leq c) \\ &\geq \mathbb{P}(X_1 - i_* = j, |X_1 - X'_1| \leq c, X'_1 = i_*) \\ &= \mathbb{P}(X_1 - i_* = j, |X_1 - i_*| \leq c) \mathbb{P}(X'_1 = i_*). \end{aligned}$$

Il suit, par conséquent, de (2.4) et de (2.5) que la loi des incréments de $(R_n)_{n \in \mathbb{N}}$ est **apériodique** :

$$\text{pgcd}\{j \in \mathbb{Z} \mid \mathbb{P}(X_1 - X_1'' = j) > 0\} = 1.$$

Le lemme suivant, dont la preuve est donnée plus bas, implique donc que la marche $(R_n)_{n \in \mathbb{N}}$ est irréductible et récurrente.

Lemme 2.17. *Toute marche aléatoire sur \mathbb{Z} dont les incréments sont bornés, symétriques et apériodiques est nécessairement irréductible et récurrente.*

$(R_n)_{n \in \mathbb{N}}$ étant irréductible et récurrente, $T := \inf\{k \in \mathbb{N} \mid S_k = S_k''\} = \inf\{k \in \mathbb{N} \mid R_k = 0\}$ est presque sûrement fini. Finalement, définissons la marche aléatoire $(S_k''')_{k \in \mathbb{N}}$ par

$$\forall n \in \mathbb{N}, \quad S_k''' := \begin{cases} S_k'' & \text{si } k < T, \\ S_k & \text{si } k \geq T. \end{cases}$$

Observons que (S_n, S_n''') est un couplage de S_n et S_n' . De plus, $\{S_n \neq S_n'''\} \subset \{T > n\}$. Il suit donc du Lemme 2.10 que

$$\|\mathcal{L}(S_n) - \mathcal{L}(S_n''')\|_{\text{VT}} \leq \mathbb{P}(S_n \neq S_n''') \leq \mathbb{P}(T > n).$$

T étant presque sûrement fini, on a bien

$$\lim_{n \rightarrow \infty} \|\mathcal{L}(S_n) - \mathcal{L}(S_n''')\|_{\text{VT}} = 0. \quad \square$$

Preuve du Lemme 2.17. Sans perte de généralité, soit $\tilde{R}_n := \sum_{k=1}^n \tilde{X}_k$ une marche aléatoire sur \mathbb{Z} dont les incréments i.i.d. $(\tilde{X}_k)_{k \in \mathbb{N}^*}$ satisfont $\mathbb{P}(|\tilde{X}_1| \leq c) = 1$ pour une constante $c < \infty$, $\mathbb{P}(\tilde{X}_1 = i) = \mathbb{P}(\tilde{X}_1 = -i)$ pour tout $i \in \mathbb{Z}$ et $\text{pgcd}\{j \in \mathbb{Z} \mid \mathbb{P}(\tilde{X}_1 = j) > 0\} = 1$.

Soit $A := \{j \in \mathbb{Z} \mid \exists n \in \mathbb{N}, \mathbb{P}(\tilde{R}_n = j) > 0\}$. Clairement, A est stable sous addition et satisfait $-A = A$ (les incréments étant symétriques) et $\text{pgcd } A = 1$. Le lemme suivant, dont la preuve est donnée plus bas, implique donc que $A = \mathbb{Z}$, ce qui montre que $(\tilde{R}_n)_{n \in \mathbb{N}}$ est irréductible.

Lemme 2.18. *Soit $A \subset \mathbb{Z}$ un ensemble stable sous addition et satisfaisant $-A = A$ et $\text{pgcd } A = 1$. Alors, $A = \mathbb{Z}$.*

Il nous reste donc à établir la récurrence de $(\tilde{R}_n)_{n \in \mathbb{N}}$. Soit $r \in \mathbb{N}^*$ et soit n suffisamment grand pour que $p := \mathbb{P}(|\tilde{R}_n| \leq 2r) < 1$. Alors, pour tout $k \in \mathbb{N}^*$,

$$\mathbb{P}(|\tilde{R}_n| \leq r, |\tilde{R}_{2n}| \leq r, \dots, |\tilde{R}_{kn}| \leq r) \leq p^k.$$

En prenant la limite $k \rightarrow \infty$, on en déduit que $\mathbb{P}(\sup_n |\tilde{R}_n| \leq r) = 0$ et donc, r étant arbitraire,

$$\mathbb{P}(\sup_n |\tilde{R}_n| = \infty) = 1.$$

Les incréments étant symétriques, il suit que

$$\mathbb{P}(\sup_n \tilde{R}_n = \infty) = \mathbb{P}(\inf_n \tilde{R}_n = -\infty) \geq 1/2.$$

Les événements $\{\sup_n \tilde{R}_n = \infty\}$ et $\{\inf_n \tilde{R}_n = -\infty\}$ appartenant à la tribu asymptotique, la loi 0-1 de Kolmogorov implique finalement que

$$\mathbb{P}(\sup_n \tilde{R}_n = \infty) = \mathbb{P}(\inf_n \tilde{R}_n = -\infty) = 1,$$

ce qui montre que $(\tilde{R}_n)_{n \in \mathbb{N}}$ change de signe infiniment souvent, presque sûrement. Les incréments étant bornés, il suit que l'ensemble $\{-c, \dots, c\}$ doit être visité infiniment souvent. Par conséquent, la récurrence de $(\tilde{R}_n)_{n \in \mathbb{N}}$ suit de son irréductibilité. \square

Preuve du Lemme 2.18. Soit $d := \min\{n \geq 1 \mid n \in A\}$. A étant stable sous addition et symétrique, il suit que $d\mathbb{Z} \subset A$. Pour chaque $i \in A$, il existe $k \in \mathbb{Z}$ tel que $0 \leq i - kd < d$. Par définition de d , ceci n'est possible que si $i - kd = 0$, ce qui montre que $A \subset d\mathbb{Z}$, et donc $A = d\mathbb{Z}$. Finalement, $d = \text{pgcd } A = 1$. \square

Marche aléatoire sur \mathbb{Z}^d

Notre but dans cette section est d'étendre le résultat de perte de mémoire du Théorème 2.16 aux marches aléatoires sur \mathbb{Z}^d . Clairement, on ne peut pas procéder de la même façon, car l'équivalent du Lemme 2.17 n'est pas vrai lorsque $d \geq 3$, la marche aléatoire étant alors transiente. Afin de simplifier l'exposition au maximum, nous ne discuterons que d'un cas particulier, mais il est possible de combiner les idées de cette section avec le couplage d'Ornstein afin de parvenir à un même niveau de généralité que dans le cas $d = 1$.

Soit $(X_k)_{k \in \mathbb{N}^*}$ une suite de variables aléatoires i.i.d. à valeurs dans \mathbb{Z}^d telles que

$$\mathbb{P}(X_1 = j) = \frac{1}{2} \mathbf{1}_{\{j=0\}} + \frac{1}{4d} \mathbf{1}_{\{\|j\|_1=1\}}.$$

Étant donné $a \in \mathbb{Z}^d$, on note $S_n^a := a + \sum_{i=1}^n X_i$ la **marche aléatoire paresseuse** sur \mathbb{Z}^d partant de a .

Théorème 2.19. *Soit $a, b \in \mathbb{Z}^d$. Alors,*

$$\lim_{n \rightarrow \infty} \|\mathcal{L}(S_n^a) - \mathcal{L}(S_n^b)\|_{\text{VT}} = 0.$$

Démonstration. On définit, itérativement, deux processus stochastiques $(S'_n)_{n \in \mathbb{N}}$ et $(S''_n)_{n \in \mathbb{N}}$ sur un même espace de probabilité de la façon suivante :

- ▷ On pose $S'_0 := a$ et $S''_0 := b$.
- ▷ Une fois S'_k et S''_k définis, on construit S'_{k+1} et S''_{k+1} de la façon suivante :
 - ▶ On choisit une composante $I \in \{1, \dots, d\}$ au hasard uniformément.
 - ▶ On pose $S'_{k+1}(s) := S'_k(s)$ et $S''_{k+1}(s) := S''_k(s)$ pour tout $s \in \{1, \dots, d\} \setminus \{I\}$.
 - ▶ Si $S'_k(I) = S''_k(I)$, alors
 - avec probabilité 1/2, on pose $S'_{k+1}(I) := S'_k(I)$ et $S''_{k+1}(I) := S''_k(I)$;
 - avec probabilité 1/2, on tire $W \in \{-1, 1\}$ au hasard uniformément et on pose $S'_{k+1}(I) := S'_k(I) + W$, $S''_{k+1}(I) := S''_k(I) + W$.
 - ▶ Si $S'_k(I) \neq S''_k(I)$, alors on tire $W \in \{-1, 1\}$ au hasard uniformément et
 - avec probabilité 1/2, on pose $S'_{k+1}(I) := S'_k(I) + W$ et $S''_{k+1}(I) := S''_k(I)$;
 - avec probabilité 1/2, on pose $S'_{k+1}(I) := S'_k(I)$ et $S''_{k+1}(I) := S''_k(I) + W$.

Il est clair que $(S'_n)_{n \in \mathbb{N}} \stackrel{\text{loi}}{=} (S_n^a)_{n \in \mathbb{N}}$ et $(S''_n)_{n \in \mathbb{N}} \stackrel{\text{loi}}{=} (S_n^b)_{n \in \mathbb{N}}$. En outre, pour chaque composante $s \in \{1, \dots, d\}$, la marche aléatoire définie par $R_n^s := S'_n(s) - S''_n(s)$ est une marche aléatoire sur \mathbb{Z} dont les incréments sont symétriques, bornés (égaux à $-1, 0$ ou 1) et apériodiques, jusqu'à la première visite en 0, après quoi la marche ne quitte plus 0. Il suit donc du Lemme 2.17 que $\tau^s := \inf\{n \in \mathbb{N} \mid R_n^s = 0\}$ est fini presque sûrement, pour chaque $s \in \{1, \dots, d\}$. On en conclut que $T := \max_{1 \leq s \leq d} \tau^s$ est fini presque sûrement.

Finalement, (S'_n, S''_n) fournit un couplage de S_n^a et S_n^b , et $S'_n \neq S''_n$ implique que $T > n$. Il suit donc du Lemme 2.10 que

$$\|\mathcal{L}(S_n^a) - \mathcal{L}(S_n^b)\|_{\text{VT}} \leq \mathbb{P}(S'_n \neq S''_n) \leq \mathbb{P}(T > n)$$

tend vers 0 lorsque $n \rightarrow \infty$. □

Avant de conclure cette section, mentionnons une conséquence directe du Théorème 2.19. Le **Laplacien discret** sur \mathbb{Z}^d est l'opérateur défini par

$$(\Delta f)(i) := \frac{1}{2d} \sum_{j \sim i} (f(j) - f(i)),$$

où la somme porte sur tous les sommets de \mathbb{Z}^d voisins du sommet i (c'est-à-dire tels que $\|j - i\|_1 = 1$). Une **fonction harmonique** sur \mathbb{Z}^d est une fonction $f : \mathbb{Z}^d \rightarrow \mathbb{R}$ telle que $\Delta f \equiv 0$, c'est-à-dire telle que la valeur en chaque sommet $i \in \mathbb{Z}^d$ est égal à la moyenne des valeurs prises sur les voisins de i :

$$\forall i \in \mathbb{Z}^d, \quad f(i) = \frac{1}{2d} \sum_{j \sim i} f(j).$$

Corollaire 2.20. *Toutes les fonctions harmoniques bornées sur \mathbb{Z}^d sont constantes.*

Démonstration. Soit $(S_n^a)_{n \in \mathbb{N}}$ la marche aléatoire paresseuse sur \mathbb{Z}^d partant de $a \in \mathbb{Z}^d$ et f une fonction harmonique telle que $\sup_j |f(j)| \leq M$. Alors, pour tout $n \in \mathbb{N}^*$,

$$\begin{aligned} \mathbb{E}[f(S_n^a)] &= \sum_{i \in \mathbb{Z}^d} f(i) \mathbb{P}(S_n^a = i) \\ &= \sum_{i \in \mathbb{Z}^d} f(i) \sum_{j \in \mathbb{Z}^d} \mathbb{P}(S_n^a = i \mid S_{n-1}^a = j) \mathbb{P}(S_{n-1}^a = j) \\ &= \sum_{i \in \mathbb{Z}^d} f(i) \left(\frac{1}{2} \mathbb{P}(S_{n-1}^a = i) + \sum_{j \in \mathbb{Z}^d} \frac{\mathbf{1}_{\{j \sim i\}}}{4d} \mathbb{P}(S_{n-1}^a = j) \right) \\ &= \frac{1}{2} \mathbb{E}[f(S_{n-1}^a)] + \frac{1}{2} \sum_{j \in \mathbb{Z}^d} \mathbb{P}(S_{n-1}^a = j) \frac{1}{2d} \sum_{i \sim j} f(i) \\ &= \frac{1}{2} \mathbb{E}[f(S_{n-1}^a)] + \frac{1}{2} \sum_{j \in \mathbb{Z}^d} \mathbb{P}(S_{n-1}^a = j) f(j) = \mathbb{E}[f(S_{n-1}^a)]. \end{aligned}$$

En itérant, on obtient $\mathbb{E}[f(S_n^a)] = \mathbb{E}[f(S_0^a)] = f(a)$. Alors, pour tout $a, b \in \mathbb{Z}^d$,

$$\begin{aligned} |f(a) - f(b)| &= |\mathbb{E}[f(S_n^a)] - \mathbb{E}[f(S_n^b)]| = \left| \sum_{j \in \mathbb{Z}^d} f(j) (\mathbb{P}(S_n^a = j) - \mathbb{P}(S_n^b = j)) \right| \\ &\leq M \sum_{j \in \mathbb{Z}^d} |\mathbb{P}(S_n^a = j) - \mathbb{P}(S_n^b = j)| \\ &= 2M \|\mathcal{L}(S_n^a) - \mathcal{L}(S_n^b)\|_{\text{VT}}. \end{aligned}$$

Il suit alors du Théorème 2.19, en laissant $n \rightarrow \infty$, que $f(a) = f(b)$. □

Remarques bibliographiques : Ce chapitre est basé sur les notes de cours [22] et sur le livre [26]. Une autre référence classique sur les couplages est [18].

3 Approximation de Poisson : la méthode de Chen–Stein

Le résultat élémentaire suivant, parfois appelé **la loi des petits nombres**, est souvent établi dans les cours d'introduction à la théorie des probabilités :

Pour chaque $n \in \mathbb{N}^$, soit $I_1^{(n)}, \dots, I_n^{(n)}$ des variables aléatoires i.i.d. suivant chacune une loi de Bernoulli de paramètre p_n . Supposons que $\lim_{n \rightarrow \infty} np_n = \lambda > 0$. Alors, $S_n := \sum_{k=1}^n I_k^{(n)}$ converge en loi vers une variable aléatoire $S \sim \text{Poisson}(\lambda)$ lorsque $n \rightarrow \infty$.*

Le but de ce chapitre est d'introduire une approche permettant d'étendre ce type de résultats au cas de variables aléatoires ni forcément indépendantes, ni forcément identiquement distribuées. De plus, cette approche fournira un contrôle de la distance entre les lois de S_n et de S .

3.1 Théorème de Chen

On considère le cadre suivant. Soit $n \in \mathbb{N}^*$ et soit I_1, \dots, I_n des variables aléatoires de Bernoulli de paramètre ρ_1, \dots, ρ_n respectivement. Posons $S_n := I_1 + \dots + I_n$ et $\lambda := \rho_1 + \dots + \rho_n$; on supposera $\lambda > 0$. En lieu et place de l'indépendance des I_k , nous supposons que la propriété suivante est satisfaite :

Hypothèse H : il existe des variables aléatoires $U_1, \dots, U_n, V_1, \dots, V_n$ définies sur le même espace de probabilité et telles que

1. $U_k \stackrel{\text{loi}}{=} S_n$;
2. $1 + V_k \stackrel{\text{loi}}{=} S_n \mid \{I_k = 1\}$.

Théorème 3.1. Soit \mathfrak{p}_λ la loi de Poisson de paramètre λ . Sous l'hypothèse H,

$$\|\mathcal{L}(S_n) - \mathfrak{p}_\lambda\|_{\text{VT}} \leq \frac{1 - e^{-\lambda}}{\lambda} \sum_{k=1}^n \rho_k \mathbb{E}[|U_k - V_k|].$$

Exemple 3.2. Appliquons le théorème au cas particulier où les I_k sont indépendantes. Dans ce cas, l'hypothèse H est clairement satisfaite pour le choix $U_k := S_n$ et $V_k := \sum_{i \in [n] \setminus \{k\}} I_i = S_n - I_k$. Comme $\mathbb{E}[|U_k - V_k|] = \mathbb{E}[|I_k|] = \rho_k$, il suit du Théorème 3.1 que

$$\|\mathcal{L}(S_n) - \mathfrak{p}_\lambda\|_{\text{VT}} \leq \frac{1 - e^{-\lambda}}{\lambda} \sum_{k=1}^n \rho_k^2 \leq \min\{1, \frac{1}{\lambda}\} \sum_{k=1}^n \rho_k^2.$$

Cette version est connue sous le nom de théorème de Le Cam. En particulier, dans le cas où les (I_k) sont i.i.d., de loi $\text{Bern}(\rho)$, on obtient

$$\|\mathcal{L}(S_n) - \mathbf{p}_\lambda\|_{\text{VT}} \leq \min\{n\rho^2, \rho\},$$

où l'on a utilisé le fait que $\lambda = n\rho$. \diamond

Nous démontrons le Théorème 3.1 dans la Section 3.2. Avant cela, énonçons quelques corollaires utiles pour les applications. (D'autres versions sont également discutées dans les séries d'exercices.)

Corollaire 3.3. *Supposons que, pour chaque $k \in \llbracket n \rrbracket$, il existe des variables aléatoires $(Z_i^{(k)})_{i \in \llbracket n \rrbracket \setminus \{k\}}$ telles que $(Z_i^{(k)})_{i \in \llbracket n \rrbracket \setminus \{k\}} \stackrel{\text{loi}}{=} (I_i)_{i \in \llbracket n \rrbracket \setminus \{k\}} \mid \{I_k = 1\}$. Alors,*

$$\|\mathcal{L}(S_n) - \mathbf{p}_\lambda\|_{\text{VT}} \leq \frac{1 - e^{-\lambda}}{\lambda} \sum_{k=1}^n \rho_k \left(\rho_k + \sum_{i \in \llbracket n \rrbracket \setminus \{k\}} \mathbb{E}[|I_i - Z_i^{(k)}|] \right).$$

Démonstration. L'hypothèse H est satisfaite pour le choix $U_k := S_n$ et $V_k := \sum_{i \in \llbracket n \rrbracket \setminus \{k\}} Z_i^{(k)}$. Ainsi,

$$\mathbb{E}[|U_k - V_k|] \leq \mathbb{E}[|I_k|] + \sum_{i \in \llbracket n \rrbracket \setminus \{k\}} \mathbb{E}[|I_i - Z_i^{(k)}|] = \rho_k + \sum_{i \in \llbracket n \rrbracket \setminus \{k\}} \mathbb{E}[|I_i - Z_i^{(k)}|]. \quad \square$$

Corollaire 3.4. *Supposons que, pour chaque $i \in \llbracket n \rrbracket$, il existe $\mathcal{V}_i \subset \llbracket n \rrbracket \setminus \{i\}$ tel que I_i soit indépendant de $(I_j)_{j \notin \{i\} \cup \mathcal{V}_i}$. Alors,*

$$\|\mathcal{L}(S_n) - \mathbf{p}_\lambda\|_{\text{VT}} \leq \frac{1 - e^{-\lambda}}{\lambda} \sum_{k=1}^n \left(\rho_k^2 + \sum_{j \in \mathcal{V}_k} (\rho_j \rho_k + \mathbb{E}[I_j I_k]) \right).$$

Démonstration. Notons $\overline{\mathcal{V}}_k := \{k\} \cup \mathcal{V}_k$. On vérifie facilement que l'on peut choisir $U_k := S_n$ et

$$V_k := \sum_{j \notin \overline{\mathcal{V}}_k} I_j + \sum_{j \in \mathcal{V}_k} I_j^{(k)},$$

où les variables aléatoires $(I_j^{(k)})_{j \in \mathcal{V}_k}$ sont tirées selon la loi conditionnelle

$$\mathbb{P}(\forall j \in \mathcal{V}_k, I_j^{(k)} = b_j \mid \forall i \in \llbracket n \rrbracket, I_i = a_i) = \mathbb{P}(\forall j \in \mathcal{V}_k, I_j = b_j \mid I_k = 1, \forall i \notin \overline{\mathcal{V}}_k, I_i = a_i).$$

La conclusion suit alors de

$$\begin{aligned} \mathbb{E}[|U_k - V_k|] &= \mathbb{E}\left[\left| I_k + \sum_{j \in \mathcal{V}_k} (I_j - I_j^{(k)}) \right| \right] \leq \mathbb{E}[|I_k|] + \sum_{j \in \mathcal{V}_k} (\mathbb{E}[|I_j|] + \mathbb{E}[|I_j^{(k)}|]) \\ &= \rho_k + \sum_{j \in \mathcal{V}_k} (\rho_j + \mathbb{E}[I_j \mid I_k = 1]) = \rho_k + \sum_{j \in \mathcal{V}_k} \left(\rho_j + \frac{\mathbb{E}[I_j I_k]}{\rho_k} \right). \quad \square \end{aligned}$$

3.2 Preuve du Théorème de Chen

3.2.1 L'équation de Chen

On cherche à borner

$$\|\mathcal{L}(S_n) - \mathbf{p}_\lambda\|_{\text{VT}} = \sup_{A \subset \mathbb{N}} (\mathbb{P}(S_n \in A) - \mathbb{P}(N \in A)) = \sup_{A \subset \mathbb{N}} \mathbb{E}[\mathbf{1}_A(S_n) - \mathbb{P}(N \in A)],$$

où $N \sim \text{Poisson}(\lambda)$. L'idée au cœur de l'approche de Chen, inspirée d'une idée originellement introduite par Stein dans un contexte gaussien, est d'utiliser la caractérisation suivante de la loi de Poisson.

Lemme 3.5. Soit Z une variable aléatoire à valeurs dans \mathbb{N} et $\lambda > 0$. Alors, $Z \sim \text{Poisson}(\lambda)$ si et seulement si

$$\mathbb{E}[\lambda f(Z+1) - Zf(Z)] = 0,$$

pour toute fonction bornée $f : \mathbb{N} \rightarrow \mathbb{R}$.

Au vu du lemme précédent¹, on peut espérer que si la loi de S_n est proche d'une loi de Poisson, alors $\mathbb{E}[\lambda f(S_n+1) - S_n f(S_n)] \approx 0$ pour toute fonction $f : \mathbb{N} \rightarrow \mathbb{R}$ bornée. On peut dès lors se demander s'il n'existerait pas une fonction $f_A : \mathbb{N} \rightarrow \mathbb{R}$ bornée et satisfaisant l'équation de Chen :

$$\forall i \in \mathbb{N}, \quad \mathbf{1}_A(i) - \mathbb{P}(N \in A) = \lambda f_A(i+1) - i f_A(i).$$

En effet, l'existence d'une telle fonction permettrait d'écrire

$$\mathbb{E}[\mathbf{1}_A(S_n) - \mathbb{P}(N \in A)] = \mathbb{E}[\lambda f_A(S_n+1) - S_n f_A(S_n)].$$

Lemme 3.6. Soit $A \subset \mathbb{N}$ et $\lambda > 0$. La fonction $f_A : \mathbb{N} \rightarrow \mathbb{R}$ définie par $f_A(0) := 0$ et

$$\forall i \in \mathbb{N}^*, \quad f_A(i) := \frac{\mathbb{P}(N \in A, N < i) - \mathbb{P}(N \in A)\mathbb{P}(N < i)}{\lambda \mathbb{P}(N = i-1)}$$

est l'unique solution de l'équation de Chen satisfaisant $f_A(0) = 0$.²

On observe à présent que

$$\begin{aligned} \mathbb{E}[\lambda f_A(S_n+1) - S_n f_A(S_n)] &= \sum_{k=1}^n \mathbb{E}[\rho_k f_A(S_n+1) - I_k f_A(S_n)] \\ &= \sum_{k=1}^n \rho_k (\mathbb{E}[f_A(S_n+1)] - \mathbb{E}[f_A(S_n) | I_k = 1]) \\ &= \sum_{k=1}^n \rho_k \mathbb{E}[f_A(U_k+1) - f_A(V_k+1)]. \end{aligned}$$

La conclusion désirée suit donc du lemme suivant.

Lemme 3.7. Soit $A \subset \mathbb{N}$ et $\lambda > 0$. Alors,

$$\forall i, j \in \mathbb{N}^*, \quad |f_A(j) - f_A(i)| \leq \frac{1 - e^{-\lambda}}{\lambda} |j - i|.$$

□

3.2.2 Preuves des lemmes

Preuve du Lemme 3.5. Si $Z \sim \text{Poisson}(\lambda)$, alors

$$\mathbb{E}[\lambda f(Z+1) - Zf(Z)] = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k+1}}{k!} f(k+1) - e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} f(k) = 0.$$

Réciproquement, en appliquant l'identité avec $f(i) = \mathbf{1}_{\{i=k\}}$ avec $k \in \mathbb{N}$, on obtient

$$0 = \mathbb{E}[\lambda \mathbf{1}_{\{Z+1=k\}} - Z \mathbf{1}_{\{Z=k\}}] = \lambda \mathbb{P}(Z = k-1) - k \mathbb{P}(Z = k).$$

1. Ce lemme n'est en fait pas utilisé dans la démonstration. Il n'est énoncé que dans le but de motiver la suite de la preuve.
2. On peut également vérifier que f_A est bornée; nous ne le faisons pas, car cela ne jouera aucun rôle par la suite.

Par conséquent, pour tout $k \in \mathbb{N}$,

$$\mathbb{P}(Z = k) = \frac{\lambda}{k} \mathbb{P}(Z = k - 1) = \dots = \frac{\lambda^k}{k!} \mathbb{P}(Z = 0).$$

La conclusion suit, puisque $1 = \sum_{k \in \mathbb{N}} \mathbb{P}(Z = k) = e^\lambda \mathbb{P}(Z = 0)$ et donc $\mathbb{P}(Z = 0) = e^{-\lambda}$. \square

Preuve du Lemme 3.6. On pose $f_A(0) := 0$, puis on procède via la relation de récurrence fournie par l'équation de Chen :

$$\begin{aligned} f_A(i+1) &= \frac{i}{\lambda} f_A(i) + \frac{1}{\lambda} \mathbf{1}_A(i) - \frac{1}{\lambda} \mathbb{P}(N \in A) \\ &= \frac{i(i-1)}{\lambda^2} f_A(i-1) + \frac{1}{\lambda} \mathbf{1}_A(i) + \frac{i}{\lambda^2} \mathbf{1}_A(i-1) - \frac{1}{\lambda} \mathbb{P}(N \in A) - \frac{i}{\lambda^2} \mathbb{P}(N \in A) \\ &= \dots \\ &= \frac{i!}{\lambda^{i+1}} \sum_{k=0}^i \frac{\lambda^k}{k!} \mathbf{1}_A(k) - \mathbb{P}(N \in A) \frac{i!}{\lambda^{i+1}} \sum_{k=0}^i \frac{\lambda^k}{k!} \\ &= \frac{\mathbb{P}(N \in A, N < i+1) - \mathbb{P}(N \in A) \mathbb{P}(N < i+1)}{\lambda \mathbb{P}(N = i)}. \end{aligned} \quad \square$$

Preuve du Lemme 3.7. Par l'inégalité triangulaire, il suffit de montrer que

$$\forall i \in \mathbb{N}^*, \quad |f_A(i+1) - f_A(i)| \leq \frac{1 - e^{-\lambda}}{\lambda}. \quad (3.1)$$

Observons d'une part que le Lemme 3.6 implique que $f_{A \cup B} = f_A + f_B$ lorsque $A \cap B = \emptyset$. En particulier,

$$f_A(i+1) - f_A(i) = \sum_{j \in A} (f_{\{j\}}(i+1) - f_{\{j\}}(i)).$$

D'autre part, pour tout $i \in \mathbb{N}^*$ et tout $j \in \mathbb{N}$,

$$f_{\{j\}}(i) = \frac{(i-1)!}{\lambda^i} e^\lambda (\mathbb{P}(N = j, N < i) - \mathbb{P}(N = j) \mathbb{P}(N < i)) = \begin{cases} -\frac{(i-1)!}{j!} \lambda^{j-i} \mathbb{P}(N < i) & \text{si } j \geq i, \\ \frac{(i-1)!}{j!} \lambda^{j-i} \mathbb{P}(N \geq i) & \text{si } j < i. \end{cases}$$

Comme $\mathbb{P}(N < i) \leq \frac{i}{\lambda} \mathbb{P}(N < i+1)$ et $\mathbb{P}(N \geq i) \geq \frac{i}{\lambda} \mathbb{P}(N \geq i+1)$, il suit que $f_{\{j\}}$ est décroissante sur $\{1, \dots, j\}$, ainsi que sur $\{j+1, j+2, \dots\}$. En particulier, $f_{\{j\}}(i+1) - f_{\{j\}}(i) \leq 0$ pour tout $i \neq j$, ce qui implique que

$$\begin{aligned} f_A(i+1) - f_A(i) &\leq f_{\{i\}}(i+1) - f_{\{i\}}(i) = \frac{1}{\lambda} (\mathbb{P}(N \geq i+1) + \frac{\lambda}{i} \mathbb{P}(N < i)) \\ &\leq \frac{1}{\lambda} (1 - \mathbb{P}(N \leq i) + \mathbb{P}(0 < N \leq i)) = \frac{1}{\lambda} (1 - \mathbb{P}(N = 0)) = \frac{1 - e^{-\lambda}}{\lambda}. \end{aligned}$$

Comme $f_A + f_{A^c} = f_{\mathbb{N}} \equiv 0$, on obtient, en appliquant l'inégalité précédente à f_{A^c} ,

$$f_A(i+1) - f_A(i) = -(f_{A^c}(i+1) - f_{A^c}(i)) \geq -\frac{1 - e^{-\lambda}}{\lambda}.$$

La borne (3.1) est démontrée. \square

3.3 Quelques exemples d'application

Dans cette section, nous présentons quelques exemples élémentaires d'application des résultats de ce chapitre. D'autres exemples seront considérés durant les séances d'exercices.

3.3.1 Problème des anniversaires

On considère un groupe de m personnes. On suppose que leurs dates d'anniversaire (jour et mois) sont i.i.d., de loi uniforme sur $\{1, \dots, 365\}$. Soit $k \leq m$ et $n := \binom{m}{k}$. Quelle est la probabilité de trouver k personnes partageant toutes la même date d'anniversaire? Le calcul explicite, pour m et k arbitraires, est difficile. Nous nous contenterons donc d'estimations, basée sur l'approximation de Poisson. La qualité de cette approximation dépendra évidemment des valeurs choisies, mais l'approche de ce chapitre permet au moins d'estimer l'erreur commise.

On considère l'ensemble de toutes les k -tuples $\mathcal{K} := \{\alpha \subset \llbracket m \rrbracket \mid |\alpha| = k\}$; clairement, $|\mathcal{K}| = n$. Pour un k -tuple donné $\alpha \in \mathcal{K}$, notons I_α l'indicatrice de l'événement « les k membres du groupe α ont la même date d'anniversaire ». Évidemment, ces variables aléatoires sont identiquement distribuées, $I_\alpha \sim \text{Bern}(\rho_\alpha)$ avec $\rho_\alpha \equiv \rho := 365^{-k+1}$, mais ne sont pas indépendantes. Le nombre de k -tuples dont tous les membres partagent la même date d'anniversaire est alors donné par $S := \sum_{\alpha \in \mathcal{K}} I_\alpha$.

On souhaite appliquer le Corollaire 3.4. Pour ce faire, il faut introduire des voisinages de dépendance appropriés. Il est évident qu'avec le choix $\mathcal{V}_\alpha := \{\beta \in \mathcal{K} \setminus \{\alpha\} \mid \alpha \cap \beta \neq \emptyset\}$, on a bien que $(I_\beta)_{\beta \notin \mathcal{V}_\alpha \cup \{\alpha\}}$ est indépendant de I_α . On a alors, pour tout $\beta \in \mathcal{V}_\alpha$,

$$\mathbb{E}[I_\alpha I_\beta] = 365^{-|\alpha \cup \beta|+1}.$$

Le Corollaire 3.4 donne alors

$$\|\mathcal{L}(S) - \mathfrak{p}_\lambda\|_{\text{VT}} \leq \frac{1 - e^{-\lambda}}{\rho} \left(\rho^2 + \sum_{\beta \in \mathcal{V}_\alpha} (\rho^2 + 365^{-|\alpha \cup \beta|+1}) \right), \quad (3.2)$$

où $\lambda := n\rho$ et on a fixé un $\alpha \in \mathcal{K}$ arbitraire. Considérons pour commencer le problème classique : quelle est la probabilité qu'au moins deux personnes dans un groupe de 23 partagent la même date d'anniversaire? Ce cas correspond à $m = 23$ et $k = 2$, et on est intéressé à la probabilité que $S > 0$. Évidemment, si $N \sim \text{Poisson}(\lambda)$, $\mathbb{P}(N \geq 1) = 1 - \mathbb{P}(N = 0) = 1 - e^{-\lambda} = 1 - e^{-253/365} \cong 0,5$. La borne (3.2) donne dans ce cas $|\mathbb{P}(S > 0) - \mathbb{P}(N > 0)| \leq 0,059$ et donc $\mathbb{P}(S > 0) \in [0,44, 0,56]$, ce qui est bien sûr compatible avec la probabilité exacte qui est approximativement égale à 0,492.

Estimons à présent la probabilité de trouver trois personnes partageant la même date d'anniversaire parmi un groupe de 50 personnes. On prend donc $m = 50$ et $k = 3$. Dans ce cas, $\|\mathcal{L}(S) - \mathfrak{p}_\lambda\|_{\text{VT}} \leq 0,0597$ et on obtient que la probabilité de trouver 3 personnes partageant une même date d'anniversaire appartient à l'intervalle $[0,077, 0,197]$ (ici, le résultat exact est d'environ 0,1264).

3.3.2 Séries de « pile » dans une succession de lancers

On lance une pièce de monnaie équilibrée un grand nombre de fois et on s'intéresse à la longueur de la plus longue série (suite de lancers consécutifs) de « pile ». Formalisons le problème. On considère une suite de variables aléatoires $(X_k)_{k \geq 1}$ i.i.d., suivant chacune une loi de Bernoulli de paramètre $1/2$; on interprétera $X_i = 1$ comme signifiant qu'un « pile » a été obtenu au lancer numéro i . Une série de « pile » de longueur au moins ℓ débute au lancer numéro $k \geq 2$ si et seulement si

$$I_k := (1 - X_{k-1}) \prod_{i=k}^{k+\ell-1} X_i = 1.$$

De même, une série de longueur au moins ℓ débute au premier lancer si

$$I_1 := \prod_{i=1}^{\ell} X_i = 1.$$

Évidemment, $I_k \sim \text{Bern}(\rho_k)$ avec $\rho_1 := 2^{-\ell}$ et $\rho_k := 2^{-\ell-1}$ pour $k \geq 2$. Notons R_n la longueur de la plus longue série débutant au plus tard au temps n . On a évidemment

$$R_n < \ell \quad \Leftrightarrow \quad S_n := \sum_{i=1}^n I_i = 0.$$

Lorsque $n \gg \ell \gg 1$, il semble naturel d'approximer S_n par une variable de Poisson de paramètre $\lambda := \mathbb{E}[S_n] = 2^{-\ell} \left(\frac{1}{2}(n-1) + 1 \right)$. On s'attend donc à ce que

$$\mathbb{P}(R_n < \ell) = \mathbb{P}(S_n = 0) \approx e^{-\lambda} = e^{-2^{-\ell} \left(\frac{1}{2}(n-1) + 1 \right)}.$$

Ainsi, si on prend $\ell = \log_2 \left(\frac{1}{2}(n-1) + 1 \right) + c$, on obtient

$$\mathbb{P}(R_n < \ell) \approx e^{-\lambda} = e^{-2^{-c}}.$$

Considérons à titre d'exemple, $n := 2047$. On a alors $\log_2 \left(\frac{1}{2}(n-1) + 1 \right) = \log_2(1024) = 10$. Estimons la probabilité d'observer une série de longueur 14 ou plus. Fixons $\ell := 14$ et calculons

$$\mathbb{P}(R_{2047} \geq 14) \approx 1 - e^{-2^{-4}} \cong 0,06059.$$

On souhaite appliquer le Corollaire 3.4 afin d'évaluer la précision de cette approximation. On introduit, pour chaque $i \in \llbracket n \rrbracket$, le voisinage $\mathcal{V}_i := \{j \in \llbracket n \rrbracket \setminus \{i\} \mid |i-j| \leq \ell\}$. Avec ce choix, $(I_j)_{j \notin \mathcal{V}_i \cup \{i\}}$ est bien indépendant de I_i . De plus,

$$\forall i \in \llbracket n \rrbracket, \forall j \in \mathcal{V}_i, \quad \mathbb{E}[I_i I_j] = 0.$$

En effet, comme $I_k = 1$ uniquement lorsque la série *début* en k , il est impossible d'avoir $I_i = I_j = 1$ lorsque $|j-i| \leq \ell$. Le Corollaire 3.4 donne donc

$$\|\mathcal{L}(S_n) - \mathfrak{p}_\lambda\|_{\text{VT}} \leq \frac{1 - e^{-\lambda}}{\lambda} \sum_{k=1}^n \left(\rho_k^2 + \rho_k \sum_{j \in \mathcal{V}_k} \rho_j \right) = \frac{1 - e^{-\lambda}}{\lambda} \sum_{k=1}^n \rho_k \sum_{j \in \mathcal{V}_k \cup \{k\}} \rho_j.$$

En utilisant

$$\sum_{k=\ell+2}^n \rho_k \sum_{j \in \mathcal{V}_k \cup \{k\}} \rho_j \leq (n - \ell - 1)(2\ell + 1)2^{-2\ell-2}$$

(l'inégalité vient de la contrainte $j \in \llbracket n \rrbracket$) et la borne grossière

$$\sum_{k=1}^{\ell+1} \rho_k \sum_{j \in \mathcal{V}_k \cup \{k\}} \rho_j \leq (\ell + 1)(2\ell + 1)2^{-2\ell-1},$$

on obtient

$$\|\mathcal{L}(S_n) - \mathfrak{p}_\lambda\|_{\text{VT}} \leq 0,000054.$$

En particulier,

$$\mathbb{P}(R_{2047} \geq 14) \in [0,06053, 0,06065].$$

3.3.3 Le problème des rencontres

Le problème suivant a été proposé en 1708 par Pierre Rémond de Montmort [8] (et est très souvent discuté en cours d'introduction aux probabilités, car il fournit une application naturelle du principe d'inclusion-exclusion)³ :

« Pierre a un certain nombre de cartes différentes qui ne sont point répétées, & qui sont mêlées à discrétion ; il parie contre Paul que s'il les tire de suite, & qu'il les nomme selon l'ordre des cartes, en commençant, ou par la plus haute, ou par la plus basse, il lui arrivera au moins une fois de tirer celle qu'il nommera. [...] On demande quel est le fort ou l'espérance de Pierre, pour quelque nombre de cartes que ce puisse être, depuis deux jusqu'à treize. »

L'énoncé précédent n'étant pas forcément limpide, reformulons-le précisément. On suppose que l'on a un paquet de n cartes numérotées de 1 à n . Le paquet est supposé bien mélangé, dans le sens que l'ordre des cartes est tiré uniformément parmi les $n!$ ordres possibles. Le jeu consiste à retourner les cartes une à une et Pierre gagne s'il existe au moins une valeur de $k \in \llbracket n \rrbracket$ telle que la k^{e} carte retournée soit la carte numéro k .

Reformulé en d'autres termes, Pierre perd si, en tirant au hasard uniformément une permutation de l'ensemble $\llbracket n \rrbracket$, celle-ci ne possède aucun point fixe (une telle permutation est appelée un dérangement). Nous utiliserons ce langage dorénavant. Soit σ_n une permutation aléatoire uniforme de $\llbracket n \rrbracket$. On considère les variables aléatoires

$$\forall i \in \llbracket n \rrbracket, \quad I_i := \mathbf{1}_{\{\sigma_n(i)=i\}}.$$

Ces variables aléatoires sont manifestement identiquement distribuées, $I_i \sim \text{Bern}(\rho_i)$ avec $\rho_i \equiv \rho := 1/n$, mais ne sont pas indépendantes (observez que si $\sigma_n(i) \neq i$, alors il existe au moins un autre j avec $\sigma_n(j) = i$). Notons $S_n := \sum_{i=1}^n I_i$ le nombre total de points fixes de σ_n . Notre but est de montrer que la loi de S_n converge en variation totale, lorsque $n \rightarrow \infty$, vers la loi de Poisson de paramètre 1 :

$$\lim_{n \rightarrow \infty} \|\mathcal{L}(S_n) - \mathbf{p}_1\|_{\text{VT}} = 0.$$

On souhaite appliquer le Théorème 3.1. Pour ce faire, étant donné $k \in \llbracket n \rrbracket$ et la permutation σ_n , considérons la permutation $\tilde{\sigma}_n^k := \tau_{k, \sigma_n(k)} \circ \sigma_n$ où $\tau_{i,j}$ est la transposition échangeant i et j . Manifestement $\tilde{\sigma}_n^k$ est distribuée comme une permutation aléatoire uniforme conditionnée à avoir un point fixe en k . On pose $U_k := S_n$ et $V_k := \sum_{i \in \llbracket n \rrbracket \setminus \{k\}} \mathbf{1}_{\{\tilde{\sigma}_n^k(i)=i\}}$. Alors, $1 + V_k \stackrel{\text{loi}}{=} S_n | \{I_k = 1\}$. Le Théorème 3.1 donne donc

$$\|\mathcal{L}(S_n) - \mathbf{p}_1\|_{\text{VT}} \leq (1 - e^{-1}) \frac{1}{n} \sum_{k=1}^n \mathbb{E}[|U_k - V_k|],$$

puisque $\rho_k = \rho = 1/n$ pour tout $k \in \llbracket n \rrbracket$ et $\lambda = n\rho = 1$. On observe ensuite que

$$|U_k - V_k| = \begin{cases} 1 & \text{si } \sigma_n(k) = k, \\ 1 & \text{si } k \text{ appartient à un 2-cycle de } \sigma_n, \\ 0 & \text{sinon.} \end{cases}$$

Par conséquent,

$$\mathbb{E}[|U_k - V_k|] = \mathbb{P}(\sigma_n(k) = k) + \mathbb{P}(\exists j \in \llbracket n \rrbracket \setminus \{k\}, \sigma_n(k) = j, \sigma_n(j) = k) = \frac{1}{n} + \frac{1}{n} = \frac{2}{n}.$$

On obtient donc finalement

$$\|\mathcal{L}(S_n) - \mathbf{p}_1\|_{\text{VT}} \leq \frac{2(1 - e^{-1})}{n}, \quad (3.3)$$

3. L'intérêt de Montmort pour ce problème était motivé par le problème plus complexe de déterminer l'avantage du banquier dans le jeu de hasard connu sous le nom de « Treize ».

ce qui tend bien vers 0. En particulier, la probabilité que Pierre perde, soit $\mathbb{P}(S_n = 0)$, satisfait

$$e^{-1} - \frac{2(1 - e^{-1})}{n} \leq \mathbb{P}(S_n = 0) \leq e^{-1} + \frac{2(1 - e^{-1})}{n}.$$

Notons que la borne (3.3) est loin d'être optimale : on peut montrer que $\|\mathcal{L}(S_n) - \mathfrak{p}_1\|_{\text{VT}} \leq \frac{2^n}{n!}$. En particulier, la convergence est en réalité super-exponentielle ! L'avantage de l'approche employée ici est toutefois son adaptabilité à de nombreuses variantes plus complexes.

3.3.4 Sommets isolés dans le graphe aléatoire d'Erdős–Rényi

Les exemples précédents illustrent le fait que si la méthode de ce chapitre permet parfois des estimées quantitatives relativement précises, celles-ci sont souvent loin d'être optimales. Son intérêt se trouve plutôt dans la possibilité de démontrer des résultats asymptotiques (ce que l'on aurait également pu faire dans les exemples précédents). Nous allons à présent en voir deux exemples.

Soit $n \in \mathbb{N}^*$. On considère le graphe aléatoire d'Erdős–Rényi $G \sim G(n, p_n)$ (cf. Chapitre 1) avec n sommets et probabilité d'occupation p_n . Notons N_0 le nombre de sommets isolés de G (c'est-à-dire, de sommets de degré 0). Lorsque $\lim_{n \rightarrow \infty} np_n = \infty$, il suit des résultats des Sections 1.3 ou 2.3.1 que $\lim_{n \rightarrow \infty} \mathbb{P}(N_0 > \epsilon n) = 0$ pour tout $\epsilon > 0$. Nous allons à présent établir deux résultats beaucoup plus précis.

Le graphe aléatoire G est encodé à l'aide d'une famille de variables aléatoires i.i.d. $(X_{ij})_{1 \leq i < j \leq n}$ suivant chacune une loi de Bernoulli de paramètre p_n : deux sommets distincts $i, j \in \llbracket n \rrbracket$ sont connectés par une arête dans G si et seulement si $X_{ij} = 1$ (où l'on a posé $X_{ij} := X_{ji}$ lorsque $i > j$).

En termes de ces variables aléatoires, le sommet $i \in \llbracket n \rrbracket$ est isolé si et seulement si

$$I_i := \prod_{j \in \llbracket n \rrbracket \setminus \{i\}} (1 - X_{ij}) = 1.$$

Évidemment, les variables aléatoires I_i ont même loi, $I_i \sim \text{Bern}(\rho_i)$ avec $\rho_i \equiv \rho := (1 - p_n)^{n-1}$, mais elles ne sont pas indépendantes. Le nombre de sommets isolés du graphe G est alors simplement donné par $S_n := \sum_{i=1}^n I_i$.

Nous allons utiliser le Corollaire 3.3. Pour ce faire, introduisons, pour chaque $k \in \llbracket n \rrbracket$, les variables aléatoires

$$\forall i \in \llbracket n \rrbracket \setminus \{k\}, \quad Z_i^{(k)} := \prod_{j \in \llbracket n \rrbracket \setminus \{i, k\}} (1 - X_{ij}).$$

Manifestement, $(Z_i^{(k)})_{i \in \llbracket n \rrbracket \setminus \{k\}} \stackrel{\text{loi}}{=} (I_i)_{i \in \llbracket n \rrbracket \setminus \{k\}} | \{I_k = 1\}$. Une application du Corollaire 3.3 fournit donc

$$\|\mathcal{L}(S_n) - \mathfrak{p}_\lambda\|_{\text{VT}} \leq \frac{1 - e^{-\lambda}}{\lambda} \sum_{k=1}^n \rho_k \left(\rho_k + \sum_{i \in \llbracket n \rrbracket \setminus \{k\}} \mathbb{E}[|I_i - Z_i^{(k)}|] \right),$$

où $\lambda := \sum_{i=1}^n \rho_i = n\rho$. Il nous faut à présent estimer $\mathbb{E}[|I_i - Z_i^{(k)}|]$. Observons que

$$|I_i - Z_i^{(k)}| = \left| \prod_{j \in \llbracket n \rrbracket \setminus \{i\}} (1 - X_{ij}) - \prod_{j \in \llbracket n \rrbracket \setminus \{i, k\}} (1 - X_{ij}) \right| = X_{ik} \prod_{j \in \llbracket n \rrbracket \setminus \{i, k\}} (1 - X_{ij}).$$

En particulier, $\mathbb{E}[|I_i - Z_i^{(k)}|] = p_n(1 - p_n)^{n-2} = \rho p_n / (1 - p_n)$. On obtient ainsi

$$\|\mathcal{L}(S_n) - \mathfrak{p}_\lambda\|_{\text{VT}} \leq \frac{1 - e^{-\lambda}}{\lambda} n\rho \left(\rho + (n-1)\rho \frac{p_n}{1 - p_n} \right) \leq \rho \left(1 + n \frac{p_n}{1 - p_n} \right).$$

Considérons à présent deux formes spécifiques pour la probabilité d'occupation p_n .

Cas 1 : $p_n := \frac{\log n + c}{n}$, avec $c \in \mathbb{R}$ et n suffisamment grand.

Dans ce cas,

$$\rho = \frac{e^{-c}}{n} + O\left(\frac{(\log n)^2}{n^2}\right), \quad \text{et} \quad \lambda = e^{-c} + O\left(\frac{(\log n)^2}{n}\right).$$

On sait de l'inégalité triangulaire que

$$\|\mathcal{L}(S_n) - \mathbf{p}_{e^{-c}}\|_{\text{VT}} \leq \|\mathcal{L}(S_n) - \mathbf{p}_\lambda\|_{\text{VT}} + \|\mathbf{p}_\lambda - \mathbf{p}_{e^{-c}}\|_{\text{VT}}.$$

Comme

$$\|\mathcal{L}(S_n) - \mathbf{p}_\lambda\|_{\text{VT}} = O\left(\frac{\log n}{n}\right)$$

et, par l'Exemple 2.11,

$$\|\mathbf{p}_\lambda - \mathbf{p}_{e^{-c}}\|_{\text{VT}} \leq 1 - e^{-|\lambda - e^{-c}|} = O\left(\frac{(\log n)^2}{n}\right),$$

on en conclut que, dans la limite $n \rightarrow \infty$, la loi du nombre S_n de sommets isolés de G converge vers la loi de Poisson de paramètre e^{-c} :

$$\forall c \in \mathbb{R}, \quad \lim_{n \rightarrow \infty} \|\mathcal{L}(S_n) - \mathbf{p}_{e^{-c}}\|_{\text{VT}} = 0.$$

Cas 2 : $p_n := \frac{c \log n}{n}$, avec $0 < c < 1$.

Dans ce cas,

$$\rho = n^{-c} + O\left(\frac{(\log n)^2}{n^{1+c}}\right), \quad \text{et} \quad \lambda = n^{1-c} + O\left(\frac{(\log n)^2}{n^c}\right).$$

On obtient donc

$$\begin{aligned} \|\mathcal{L}(S_n) - \mathbf{p}_{n^{1-c}}\|_{\text{VT}} &\leq \|\mathcal{L}(S_n) - \mathbf{p}_\lambda\|_{\text{VT}} + \|\mathbf{p}_\lambda - \mathbf{p}_{n^{1-c}}\|_{\text{VT}} \\ &= O\left(\frac{\log n}{n^c}\right) + O\left(\frac{(\log n)^2}{n^c}\right) = O\left(\frac{(\log n)^2}{n^c}\right). \end{aligned} \quad (3.4)$$

Posons $\lambda_n := n^{1-c}$. Soit $Y \sim \text{Poisson}(\lambda_n)$ et $\hat{Y} := (Y - \lambda_n)/\sqrt{\lambda_n}$. La fonction caractéristique de \hat{Y} satisfait, lorsque $n \rightarrow \infty$,

$$\forall t \in \mathbb{R}, \quad \varphi_{\hat{Y}}(t) = e^{-it\sqrt{\lambda_n}} \varphi_Y(t/\sqrt{\lambda_n}) = e^{-it\sqrt{\lambda_n}} e^{\lambda_n(e^{it/\sqrt{\lambda_n}} - 1)} = e^{-\frac{1}{2}t^2 + O(t^3/\sqrt{\lambda_n})},$$

ce qui implique que $\mathcal{L}(\hat{Y}) \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1)$. Par (3.4)

$$\forall x \in \mathbb{R}, \quad \mathbb{P}((S_n - n^{1-c})/\sqrt{n^{1-c}} \leq x) = \mathbb{P}(\hat{Y} \leq x) + O\left(\frac{(\log n)^2}{n^c}\right).$$

On en conclut que $\mathcal{L}((S_n - n^{1-c})/\sqrt{n^{1-c}}) \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1)$.

Remarques bibliographiques : Ce chapitre est basé sur le livre [5] et l'article [2].

4 Concentration de la mesure : l'inégalité de Talagrand

Soit $(X_k)_{k \geq 1}$ des variables aléatoires identiquement distribuées, non corrélées et de variance finie. Soit $s_n := \frac{1}{n}(X_1 + \dots + X_n)$ la moyenne empirique des n premières variables aléatoires. L'inégalité de Bienaymé-Tchebychev montre alors que, pour toute suite $(a_n)_{n \geq 1}$ satisfaisant $\lim_{n \rightarrow \infty} na_n^2 = +\infty$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|s_n - \mathbb{E}(X_1)| > a_n) \leq \lim_{n \rightarrow \infty} \frac{\text{Var}(s_n)}{a_n^2} = \lim_{n \rightarrow \infty} \frac{\text{Var}(X_1)}{na_n^2} = 0.$$

En particulier, l'essentiel de la masse associée à la loi de $(X_1 + \dots + X_n)/n$ est concentrée dans un intervalle de longueur $n^{-1/2+\epsilon}$ ($\epsilon > 0$ arbitraire) autour de $\mathbb{E}(X_1)$ lorsque n devient grand.

Le but de ce chapitre est d'introduire certaines approches utiles à l'étude de ce type de **phénomènes de concentration**. Il s'agit d'un domaine très vaste, que nous ne ferons qu'effleurer. Plus précisément, nous allons démontrer l'inégalité de Talagrand, qui constitue une contribution majeure à cette théorie. Nous travaillerons toujours sous des hypothèses d'indépendance. Par contre, nous ne nous restreindrons pas à des fonctions linéaires comme la moyenne empirique. Un principe général dans cette théorie est qu'une fonction f de variables aléatoires X_1, \dots, X_n indépendantes ne dépendant que faiblement de la valeur prise par chacune de ces variables, devrait être approximativement constante avec une probabilité proche de 1. Observez que c'est le cas de la moyenne empirique : la contribution de chacune des n variables étant typiquement d'ordre $1/n$ lorsque ces variables sont d'espérance finie.

4.1 Distance de Hamming pondérée

Soit $n \in \mathbb{N}^*$ et soit $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i \in \llbracket n \rrbracket$, des espaces probabilisés. Nous nous intéresserons à la mesure produit $\mu := \mu_1 \times \dots \times \mu_n$ sur $\Omega := \Omega_1 \times \dots \times \Omega_n$.

Définition 4.1. Soit $\mathbb{S}_+^n := \{\alpha \in \mathbb{R}_+^n \mid \|\alpha\|_2 = 1\}$. La **distance de Hamming pondérée par $\alpha \in \mathbb{S}_+^n$** sur Ω est définie par

$$\forall x, y \in \Omega, \quad d_\alpha(x, y) := \sum_{i=1}^n \alpha_i \mathbf{1}_{\{x_i \neq y_i\}}.$$

Étant donné $x \in \Omega$ et $A \subset \Omega$, on notera

$$\mathcal{D}_A(x) := \sup_{\alpha \in \mathbb{S}_+^n} d_\alpha(x, A) = \sup_{\alpha \in \mathbb{S}_+^n} \inf_{y \in A} d_\alpha(x, y).$$

1. Dans tout ce chapitre, les sous-ensembles de Ω considérés seront toujours implicitement supposés mesurables.

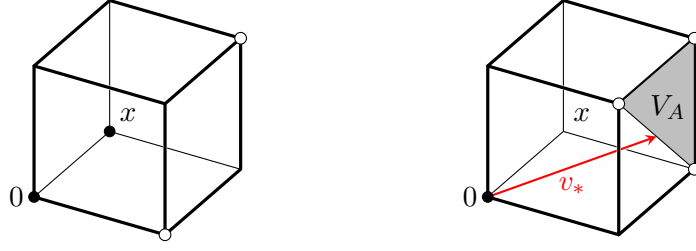


FIGURE 4.1: Illustration du cas où $n = 3$, $x = (0, 1, 0)$, $A = \{(1, 0, 0), (1, 1, 1)\}$. Gauche : représentation de cette situation (en blanc, les sommets de A). Droite : représentation de $U_A = \{(1, 1, 0), (1, 0, 1), (1, 1, 1)\}$ (sommets blancs), V_A (en gris) et du vecteur v_* (en rouge) réalisant le minimum dans le Lemme 4.3. On obtient ainsi $\mathcal{D}_A(x) = \sqrt{3}/2$, ce qui correspond aux poids $\alpha = (\sqrt{2}/3, \sqrt{1}/6, \sqrt{1}/6)$.

Exemple 4.2. Considérons le cas où $\Omega := \{0, 1\}^n$ muni de la loi uniforme. Soit $A \subsetneq \Omega$ non vide et $x \in \Omega \setminus A$. On a alors $\mathcal{D}_A(x) \geq \min_{y \in A} |\{i \in \llbracket n \rrbracket \mid x_i \neq y_i\}| n^{-1/2}$, puisque cette borne revient à choisir $\alpha_i = n^{-1/2}$ pour chaque $i \in \llbracket n \rrbracket$.

D'un autre côté, s'il existe un ensemble de coordonnées $I \subset \llbracket n \rrbracket$ tel que, pour se retrouver dans A , il est nécessaire que toutes les composantes x_i , $i \in I$, changent de valeur, alors $\mathcal{D}_A(x) \geq |I|^{1/2}$, la borne correspondant au choix $\alpha = |I|^{-1/2}$ pour $i \in I$ et $\alpha_i = 0$ pour $i \in \llbracket n \rrbracket \setminus I$. \diamond

Considérons l'ensemble

$$U_A(x) := \{s = (s_1, \dots, s_n) \in \{0, 1\}^n \mid \exists y \in A \text{ tel que } x_i \neq y_i \implies s_i = 1\}. \quad (4.1)$$

En d'autres termes, $s \in U_A(x)$ s'il est possible de passer de x à un élément de A en ne changeant que des composantes x_i avec $s_i = 1$.²

Avec cette notation,³

$$\mathcal{D}_A(x) = \sup_{\alpha \in \mathbb{S}_+^n} \inf_{y \in A} \sum_{i=1}^n \alpha_i \mathbf{1}_{\{x_i \neq y_i\}} = \sup_{\alpha \in \mathbb{S}_+^n} \min_{s \in U_A(x)} \sum_{i=1}^n \alpha_i s_i = \sup_{\alpha \in \mathbb{S}_+^n} \min_{s \in U_A(x)} \alpha \cdot s.$$

Soit $V_A(x)$ l'enveloppe convexe de $U_A(x)$ (considéré comme un sous-ensemble de \mathbb{R}^n) (cf. Fig. 4.1). Nous pouvons à présent donner une caractérisation alternative de $\mathcal{D}_A(x)$.

Lemme 4.3. Pour tout $A \subset \Omega$ et tout $x \in \Omega \setminus A$,

$$\mathcal{D}_A(x) = \min_{v \in V_A(x)} \|v\|_2.$$

Démonstration. Soit $v_* \in V_A(x)$ tel que $\|v_*\|_2 = \min_{v \in V_A(x)} \|v\|_2$. Par convexité de $V_A(x)$, l'hyperplan normal à v_* et passant par v_* sépare l'origine et $V_A(x)$. En particulier, pour tout $v \in V_A(x)$, $v \cdot v_* \geq \|v_*\|_2^2$. On choisit $\alpha_* = v_*/\|v_*\|_2$. Alors, pour tout $s \in U_A(x) \subset V_A(x)$, on a $\alpha_* \cdot s = \frac{1}{\|v_*\|_2} s \cdot v_* \geq \|v_*\|_2$. Par conséquent,

$$\mathcal{D}_A(x) = \sup_{\alpha \in \mathbb{S}_+^n} \min_{s \in U_A(x)} \alpha \cdot s \geq \min_{s \in U_A(x)} \alpha_* \cdot s \geq \|v_*\|_2 = \min_{v \in V_A(x)} \|v\|_2.$$

2. On pourrait trouver plus naturel de définir $U_A(x)$ de sorte à ce que $x_i \neq y_i \iff s_i = 1$. Il se trouve cependant que la flexibilité additionnelle donnée par (4.1) va se révéler très utile dans la preuve de l'inégalité de Talagrand.

3. Pour la seconde identité, observons que, pour tout $y \in A$, $\sum_{i=1}^n \alpha_i \mathbf{1}_{\{x_i \neq y_i\}} \geq \min_{s \in U_A(x)} \sum_{i=1}^n \alpha_i s_i$, puisque $(\mathbf{1}_{\{x_i \neq y_i\}})_{i=1}^n \in U_A(x)$. On a donc $\inf_{y \in A} \sum_{i=1}^n \alpha_i \mathbf{1}_{\{x_i \neq y_i\}} \geq \min_{s \in U_A(x)} \sum_{i=1}^n \alpha_i s_i$. Pour l'autre direction, fixons $s \in U_A(x)$ et observons que pour tout $z \in A$ tel que $x_i \neq z_i \implies s_i = 1$, on a $\sum_{i=1}^n \alpha_i s_i \geq \sum_{i=1}^n \alpha_i \mathbf{1}_{\{x_i \neq z_i\}} \geq \inf_{y \in A} \sum_{i=1}^n \alpha_i \mathbf{1}_{\{x_i \neq y_i\}}$. On a donc $\min_{s \in U_A(x)} \sum_{i=1}^n \alpha_i s_i \geq \inf_{y \in A} \sum_{i=1}^n \alpha_i \mathbf{1}_{\{x_i \neq y_i\}}$.

Pour l'autre borne, considérons $\alpha \in \mathbb{S}_+^n$ arbitraire. Alors, $\alpha \cdot v_* \leq \|v_*\|_2$. Par définition de $V_A(x)$, $v_* = \sum_{i=1}^n \lambda_i s_i$ avec $s_1, \dots, s_n \in U_A(x)$, $\lambda_1, \dots, \lambda_n \geq 0$ et $\lambda_1 + \dots + \lambda_n = 1$. On a donc

$$\|v_*\|_2 \geq \alpha \cdot v_* = \sum_{i=1}^n \lambda_i (\alpha \cdot s_i),$$

ce qui implique qu'il existe $i \in \llbracket n \rrbracket$ tel que $\alpha \cdot s_i \leq \|v_*\|_2$. Par conséquent, quel que soit $\alpha \in \mathbb{S}_+^n$, on a $\min_{s \in U_A(x)} \alpha \cdot s \leq \|v_*\|_2$ et donc

$$\mathcal{D}_A(x) = \sup_{\alpha \in \mathbb{S}_+^n} \min_{s \in U_A(x)} \alpha \cdot s \leq \|v_*\|_2 = \min_{v \in V_A(x)} \|v\|_2. \quad \square$$

4.2 Inégalité de Talagrand

Pour $t \in \mathbb{R}_+^*$ et $A \subset \Omega$, définissons le t -épaississement de A par $A_t := \{x \in \Omega \mid \mathcal{D}_A(x) \leq t\}$. Le résultat central de ce chapitre est l'inégalité suivante qui montre que le t -épaississement d'un ensemble A de mesure d'ordre 1 (disons, $\mu(A) = \frac{1}{2}$) a une mesure qui tend vers 1 extrêmement rapidement lorsque t augmente.

Théorème 4.4 (Inégalité de Talagrand). *Pour tout $A \subset \Omega$ mesurable et tout $t > 0$,*

$$\mu(A)(1 - \mu(A_t)) \leq e^{-t^2/4}.$$

Exemple 4.5. Dans cet exemple, considérons $\Omega := \{0, 1\}^n$ muni de la loi uniforme.

▷ Soit $A \subset \Omega$ tel que $\mu(A) \geq p > 0$. L'inégalité de Talagrand implique donc

$$\mu(A_t) \geq 1 - \frac{1}{\mu(A)} e^{-t^2/4} \geq 1 - \frac{1}{p} e^{-t^2/4}.$$

Pour comprendre ce qu'implique cette inégalité, il peut être utile de considérer les poids $\alpha_i := 1/\sqrt{n}$ pour tout $i \in \llbracket n \rrbracket$. On a alors

$$x \in A_t \implies d_\alpha(x, A) \leq \mathcal{D}_A(x) \leq t \implies \exists y \in A \text{ tel que } \sum_{i=1}^n \mathbf{1}_{\{x_i \neq y_i\}} \leq t\sqrt{n}.$$

Ainsi, si l'on tire au hasard un élément de Ω , alors, avec probabilité au moins $1 - \frac{1}{p} e^{-t^2/4}$, il existe un point de A ne différant de celui-ci qu'en au plus $t\sqrt{n}$ coordonnées.

▷ Soit $A \subset \Omega$ avec $\mu(A) \geq 10^{-6}$. Alors,

$$\mu(A_t) \geq 1 - 10^6 e^{-t^2/4}.$$

En particulier, cela signifie qu'avec une probabilité d'au moins 0,999986, un point tiré au hasard uniformément dans Ω différera en au plus $10\sqrt{n}$ coordonnées d'un point de l'ensemble A .

▷ Considérons à présent $A := \{x \in \{0, 1\}^n \mid \|x\|_1 \leq n/2\}$. Par symétrie, $\mu(A) \geq 1/2$. Il suit donc des considérations ci-dessus que, si X est un élément aléatoire de Ω tiré uniformément,

$$\mu(\|X\|_1 > \frac{n}{2} + t\sqrt{n}) \leq \mu(A_t^c) \leq \frac{1}{\mu(A)} e^{-t^2/4} = 2e^{-t^2/4},$$

puisque, si $\|x\|_1 > \frac{n}{2} + t\sqrt{n}$ et $y \in A$, alors $\sum_{i=1}^n \mathbf{1}_{\{x_i \neq y_i\}} = \|x - y\|_1 \geq \|x\|_1 - \|y\|_1 > t\sqrt{n}$. Notons que la borne ci-dessus reproduit qualitativement (les constantes sont moins bonnes) celle du Lemme 1.4. \diamond

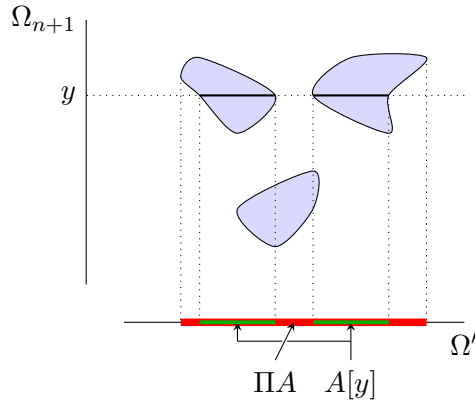


FIGURE 4.2: L'ensemble A (les 3 domaines en bleu) et les projections ΠA (en rouge) et $A[y]$ (en vert) sur Ω' .

4.3 Preuve de l'inégalité de Talagrand

Le Théorème 4.4 est une conséquence du résultat suivant.

Proposition 4.6. *Pour tout $A \subset \Omega$ mesurable et non vide, on a*

$$\int_{\Omega} \exp\left(\frac{1}{4}\mathcal{D}_A(x)^2\right) \mu(dx) \leq \frac{1}{\mu(A)}.$$

Avant de démontrer cette proposition, voyons comment en déduire l'inégalité de Talagrand.

Preuve du Théorème 4.4. Soit X un élément aléatoire de Ω de loi μ . Il suit alors de l'inégalité de Markov et de la Proposition 4.6 que, pour tout $t > 0$,

$$1 - \mu(A_t) = \mu(\mathcal{D}_A(X) > t) = \mu(e^{\mathcal{D}_A(X)^2/4} > e^{t^2/4}) \leq \mathbb{E}_{\mu}(e^{\mathcal{D}_A(X)^2/4})e^{-t^2/4} \leq \frac{1}{\mu(A)}e^{-t^2/4}. \quad \square$$

Preuve de la Proposition 4.6. La preuve se fait par récurrence sur la dimension n . Le cas $n = 1$ est facile. En effet, dans ce cas, $\mathcal{D}_A(x) = \inf_{y \in A} \mathbf{1}_{\{x \neq y\}} = \mathbf{1}_{\{x \notin A\}}$, ce qui implique

$$\int_{\Omega} \exp\left(\frac{1}{4}\mathcal{D}_A(x)^2\right) \mu(dx) = (1 - e^{1/4})\mu(A) + e^{1/4} \leq \frac{1}{\mu(A)},$$

puisqu'on a $(1 - a)x + a \leq 1/x$ pour tout $x \in (0, 1]$, lorsque $a \in [1, 2]$.

Supposons donc que l'affirmation de la proposition soit vérifiée pour $\Omega' := \times_{i=1}^n \Omega_i$ (et la mesure $\mu' := \times_{i=1}^n \mu_i$) et considérons $\Omega := \Omega' \times \Omega_{n+1}$. Nous noterons les éléments de Ω sous la forme (x, y) avec $x \in \Omega'$ et $y \in \Omega_{n+1}$. Introduisons les sous-ensembles (cf. Fig. 4.2)

$$\begin{aligned} \forall y \in \Omega_{n+1}, \quad A[y] &:= \{x \in \Omega' \mid (x, y) \in A\} \\ \Pi A &:= \{x \in \Omega' \mid \exists y \in \Omega_{n+1}, (x, y) \in A\}. \end{aligned}$$

Observons que si $(x, y) \in \Omega \setminus A$, il existe deux façons de modifier (x, y) pour atteindre A : on peut laisser y varier, ce qui réduit le problème à modifier x pour atteindre ΠA , ou on peut conserver y , ce qui réduit le problème à modifier x pour atteindre $A[y]$. Ceci conduit à deux observations cruciales.

La première est que si $s \in U_{\Pi A}(x)$, alors $(s, 1) \in U_A((x, y))$. En effet, si $s \in U_{\Pi A}(x)$, alors il existe $\tilde{x} \in \Pi A$ tel que $s_i = 1$ pour tout $i \in \llbracket n \rrbracket$ tel que $x_i \neq \tilde{x}_i$. Or, \tilde{x} appartenant à ΠA , on sait qu'il existe $\tilde{y} \in \Omega_{n+1}$ tel que $(\tilde{x}, \tilde{y}) \in A$. L'affirmation suit, puisque $(s, 1)_i = 1$ pour tout $i \in \llbracket n + 1 \rrbracket$ tel que $(\tilde{x}, \tilde{y})_i \neq (x, y)_i$.⁴

4. C'est ici que le fait d'avoir utilisé une implication plutôt qu'une équivalence dans la définition (4.1) se révèle utile, car on ne sait pas si $y = \tilde{y}$.

La seconde observation est que si $t \in U_{A[y]}(x)$, alors $(t, 0) \in U_A((x, y))$. En effet, si $t \in U_{A[y]}(x)$, alors il existe $\tilde{x} \in A[y]$ tel que $t_i = 1$ pour tout $i \in \llbracket n \rrbracket$ tel que $x_i \neq \tilde{x}_i$. Or, \tilde{x} appartenant à $A[y]$, on sait que $(\tilde{x}, y) \in A$. L'affirmation suit, puisque $(t, 0)_i = 1$ pour tout $i \in \llbracket n+1 \rrbracket$ tel que $(\tilde{x}, y)_i \neq (x, y)_i$.

Il suit immédiatement des deux observations précédentes que si $s \in V_{\Pi A}(x)$ et $t \in V_{A[y]}(x)$, alors $(s, 1) \in V_A((x, y))$ et $(t, 0) \in V_A((x, y))$. Par conséquent,

$$\forall \lambda \in [0, 1], \quad ((1 - \lambda)s + \lambda t, 1 - \lambda) \in V_A((x, y)).$$

Il suit donc du Lemme 4.3 que

$$\begin{aligned} \mathcal{D}_A((x, y))^2 &= \min_{v \in V_A(x)} \|v\|_2^2 \leq \|((1 - \lambda)s + \lambda t, 1 - \lambda)\|_2^2 = \|(1 - \lambda)s + \lambda t\|_2^2 + (1 - \lambda)^2 \\ &\leq (1 - \lambda)\|s\|_2^2 + \lambda\|t\|_2^2 + (1 - \lambda)^2, \end{aligned}$$

où l'on a utilisé la convexité de la fonction $u \mapsto \|u\|_2^2$. On obtient donc, en optimisant sur s et t ,

$$\begin{aligned} \mathcal{D}_A((x, y))^2 &\leq (1 - \lambda) \min_{s \in V_{\Pi A}(x)} \|s\|_2^2 + \lambda \min_{t \in V_{A[y]}(x)} \|t\|_2^2 + (1 - \lambda)^2 \\ &= (1 - \lambda)\mathcal{D}_{\Pi A}(x)^2 + \lambda\mathcal{D}_{A[y]}(x)^2 + (1 - \lambda)^2. \end{aligned}$$

En particulier, en fixant $y \in \Omega_{n+1}$ et en intégrant sur $x \in \Omega'$,

$$\begin{aligned} \int_{\Omega'} \exp\left(\frac{1}{4}\mathcal{D}_A((x, y))^2\right) \mu'(dx) &\leq e^{(1-\lambda)^2/4} \int_{\Omega'} \left(\exp\left(\frac{1}{4}\mathcal{D}_{\Pi A}(x)^2\right)\right)^{1-\lambda} \left(\exp\left(\mathcal{D}_{A[y]}(x)^2\right)\right)^\lambda \mu'(dx) \\ &\leq e^{(1-\lambda)^2/4} \left(\int_{\Omega'} \exp\left(\frac{1}{4}\mathcal{D}_{\Pi A}(x)^2\right) \mu'(dx)\right)^{1-\lambda} \left(\int_{\Omega'} \exp\left(\mathcal{D}_{A[y]}(x)^2\right) \mu'(dx)\right)^\lambda, \end{aligned}$$

où la dernière ligne suit d'une application de l'inégalité de Hölder. Nous pouvons à présent utiliser l'hypothèse de récurrence afin d'obtenir

$$\int_{\Omega'} \exp\left(\frac{1}{4}\mathcal{D}_A((x, y))^2\right) \mu'(dx) \leq e^{(1-\lambda)^2/4} \frac{1}{\mu'(\Pi A)^{1-\lambda} \mu'(A[y])^\lambda} = \frac{1}{\mu'(\Pi A)} e^{(1-\lambda)^2/4} r^{-\lambda},$$

où l'on a posé $r := \mu'(A[y])/\mu'(\Pi A)$. Observons que $r \in [0, 1]$, puisque $A[y] \subset \Pi A$. Un petit calcul montre que la fonction $\lambda \mapsto e^{(1-\lambda)^2/4} r^{-\lambda}$ est minimisée en $\lambda = 1 + 2 \log r$ si $r \geq e^{-1/2}$ et $\lambda = 0$ sinon. En choisissant ainsi λ , on obtient

$$e^{(1-\lambda)^2/4} r^{-\lambda} = \begin{cases} e^{-\log r - (\log r)^2} & \text{si } r \in [e^{-1/2}, 1], \\ e^{1/4} & \text{si } r \in [0, e^{-1/2}). \end{cases}$$

Il s'ensuit que

$$\forall r \in [0, 1], \quad e^{(1-\lambda)^2/4} r^{-\lambda} \leq 2 - r.$$

En effet, lorsque $r \in [0, e^{-1/2})$, l'affirmation suit de $e^{1/4} \leq 2 - e^{-1/2}$. Lorsque $r \in [e^{-1/2}, 1]$, il suffit d'observer que la droite $r \mapsto 2 - r$ est tangente au graphe de la fonction $g(r) := e^{-\log r - (\log r)^2}$ en $r = 1$ et que g est concave sur tout cet intervalle.

On a donc finalement

$$\int_{\Omega'} \exp\left(\frac{1}{4}\mathcal{D}_A((x, y))^2\right) \mu'(dx) \leq \frac{1}{\mu'(\Pi A)} \left(2 - \frac{\mu'(A[y])}{\mu'(\Pi A)}\right).$$

La conclusion suit en intégrant sur y (et en utilisant Fubini)

$$\begin{aligned} \int_{\Omega} \exp\left(\frac{1}{4}\mathcal{D}_A((x, y))^2\right) \mu'(dx) \mu_{n+1}(dy) &\leq \frac{1}{\mu'(\Pi A)} \left(2 - \frac{\int_{\Omega_{n+1}} \mu'(A[y]) \mu_{n+1}(dy)}{\mu'(\Pi A)}\right) \\ &= \frac{1}{\mu'(\Pi A)} \left(2 - \frac{\mu(A)}{\mu'(\Pi A)}\right) \leq \frac{1}{\mu(A)}, \end{aligned}$$

puisque $u(2 - u) \leq 1$ pour tout $u \in \mathbb{R}$. □

4.4 Quelques corollaires et applications

Dans cette section sont présentés quelques corollaires intéressants de l'inégalité de Talagrand. Ceux-ci facilitent son utilisation dans de nombreuses situations. Nous en profiterons également pour donner quelques exemples simples de telles applications.

4.4.1 Concentration des fonctions lipschitziennes

Définition 4.7. Soit $\Omega = \times_{i=1}^n \Omega_i$ et $c > 0$. Une fonction $F : \Omega \rightarrow \mathbb{R}$ est *c-lipschitzienne au sens de Talagrand* si, pour tout $x \in \Omega$, il existe $\alpha = \alpha(x) \in \mathbb{S}_+^n$ tel que

$$\forall y \in \Omega, \quad F(x) \leq F(y) + cd_\alpha(x, y).$$

Remarque 4.8. Insistons sur le fait que la valeur de α (et donc la "distance" d_α utilisée) n'est pas fixée a priori, mais peut dépendre du point x choisi. Cette flexibilité est une des grandes forces de cette approche et se révèle essentielle dans de nombreuses applications. \diamond

On a alors le résultat suivant de concentration autour des médianes⁵.

Théorème 4.9. Soit $\Omega = \times_{i=1}^n \Omega_i$ et $\mu = \times_{i=1}^n \mu_i$. Soit $F : \Omega \rightarrow \mathbb{R}$ c-Lipschitzienne au sens de Talagrand. Soit m_F une médiane pour F . Alors,

$$\forall t \in \mathbb{R}_+^*, \quad \mu(|F - m_F| \geq t) \leq 4e^{-t^2/4c^2}.$$

Démonstration. Soit $r \in \mathbb{R}$, $A := \{F \leq r\} \subset \Omega$ et $x \in \Omega$. Par hypothèse, il existe $\alpha = \alpha(x) \in \mathbb{S}_+^n$ tel que $F(x) \leq F(y) + cd_\alpha(x, y)$ pour tout $y \in \Omega$. En particulier, si $y \in A$, on a $F(x) \leq r + cd_\alpha(x, y)$, ce qui implique que $\mathcal{D}_A(x) \geq \inf_{y \in A} d_\alpha(x, y) \geq (F(x) - r)/c$. Ainsi, pour tout $t \in \mathbb{R}_+^*$,

$$A_t = \{x \in \Omega \mid \mathcal{D}_A(x) \leq t\} \subset \{x \in \Omega \mid F(x) \leq r + ct\}.$$

Il suit donc de l'inégalité de Talagrand que, pour tout $t > 0$,

$$\mu(F \leq r)\mu(F > r + ct) \leq \mu(A)(1 - \mu(A_t)) \leq e^{-t^2/4}.$$

En particulier, pour tout $t > 0$,

$$\mu(F \leq r)\mu(F \geq r + ct) = \lim_{\epsilon \downarrow 0} \mu(F \leq r)\mu(F > r + c(t - \epsilon)) \leq \lim_{\epsilon \downarrow 0} e^{-(t-\epsilon)^2/4} = e^{-t^2/4}.$$

Fixons $u \in \mathbb{R}_+^*$. En appliquant cette inégalité avec $t := u/c$, et en choisissant successivement $r := m_F$ et $r := m_F - u$, on obtient

$$\begin{aligned} \mu(F \geq m_F + u) &\leq \frac{1}{\mu(F \leq m_F)} e^{-u^2/4c^2} \leq 2e^{-u^2/4c^2}, \\ \mu(F \leq m_F - u) &\leq \frac{1}{\mu(F \geq m_F)} e^{-u^2/4c^2} \leq 2e^{-u^2/4c^2}, \end{aligned}$$

et la conclusion suit. \square

Exemple 4.10. Soit X_1, \dots, X_n des variables indépendantes, $X_i \sim \text{bernoulli}(p_i)$ pour chaque i . Dans le cadre de ce chapitre, cela revient à considérer $\Omega := \{0, 1\}^n$ et $\mu = \times_{i=1}^n \mu_i$ avec $\mu_i(\{1\}) = p_i$ pour

5. Rappelons que m_F est une médiane pour F si $\mu(F \geq m_F) \geq 1/2$ et $\mu(F \leq m_F) \geq 1/2$.

chaque i . Soit $F(x_1, \dots, x_n) := x_1 + \dots + x_n$. Évidemment, pour tout $y \in \Omega$, $F(0) \leq F(y)$. Supposons donc $x \neq 0$. Alors, pour tout $y \in \Omega$,

$$F(x) = \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \mathbf{1}_{\{x_i=y_i\}} + \sum_{i=1}^n x_i \mathbf{1}_{\{x_i \neq y_i\}} \leq \sum_{i=1}^n y_i + \sum_{i=1}^n x_i \mathbf{1}_{\{x_i \neq y_i\}} \leq F(y) + \sqrt{n} d_\alpha(x, y),$$

où l'on a choisi $\alpha(x) := x/\|x\|_2 \in \mathbb{S}_+^n$ (et utilisé $\|x\|_2 \leq \sqrt{n}$). Il suit donc du Théorème 4.9 que

$$\forall t \in \mathbb{R}_+^*, \quad \mu(|F - m_F| \geq t) \leq 4e^{-t^2/4n}. \quad \diamond$$

Exemple 4.11. On considère n urnes et m boules. Chaque boule est placée au hasard, uniformément et indépendamment, dans une urne. Soit F le nombre d'urnes non vides et m_F une médiane pour F . Cela correspond à prendre $\Omega := \{1, \dots, n\}^m$, $\mu := \times_{i=1}^m \mu_i$ avec $\mu_i(\{k\}) = 1/n$ pour chaque $k \in \llbracket n \rrbracket$ et chaque $i \in \llbracket m \rrbracket$. Ainsi, pour $x \in \Omega$, x_i dénote l'urne dans laquelle est placée la i^e boule.

Soit $x, y \in \Omega$, $I := \{i \in \llbracket m \rrbracket \mid x_i = y_i\}$ et soit $B := \{x_i \mid i \in I\}$ l'ensemble des urnes occupées par les boules dont le numéro appartient à I . Par définition de I , $|B| = |\{y_i \mid i \in I\}| \leq F(y)$. On a donc

$$F(x) = |B| + |\{x_i \mid x_i \notin B\}| \leq F(y) + \sum_{i=1}^m a_i(x) \mathbf{1}_{\{x_i \neq y_i\}},$$

où l'on a posé $a_i(x) := 1$ si la i^e boule est la boule de plus petit numéro dans son urne (relativement à la répartition associée à x), et $a_i(x) := 0$ sinon. L'inégalité suit de l'observation que, pour chaque $u \in \{x_i \mid x_i \notin B\}$, si $j := \min\{i \in \llbracket m \rrbracket \mid x_i = u\}$, alors $x_j \neq y_j$ et $a_j(x) = 1$.

En choisissant $\alpha(x) := a(x)/\|a(x)\|_2$ (et en utilisant $\|a(x)\|_2 \leq \sqrt{n \wedge m}$), on obtient ainsi, pour chaque $x \in \Omega$,

$$\forall y \in \Omega, \quad F(x) \leq F(y) + \sqrt{n \wedge m} \sum_{i=1}^m \alpha_i(x) \mathbf{1}_{\{x_i \neq y_i\}}.$$

Il suit donc du Théorème 4.9 que

$$\forall t \in \mathbb{R}_+^*, \quad \mu(|F - m_F| \geq t) \leq 4e^{-t^2/4(n \wedge m)}. \quad \diamond$$

Le Théorème 4.9, énonçant un résultat de concentration autour d'une médiane de F , est typique des applications de l'inégalité de Talagrand. Le fait qu'une médiane apparaisse, plutôt que l'espérance, peut sembler ennuyeux, les médianes étant généralement beaucoup plus difficiles à déterminer. Cependant, lorsqu'un résultat de concentration est disponible, on peut généralement remplacer la médiane par l'espérance, le prix à payer étant que les constantes se détériorent quelque peu.

Lemme 4.12. 1. Soit X une variable aléatoire réelle telle que

$$\forall t \in \mathbb{R}_+^*, \quad \mathbb{P}(|X - m_X| \geq t) \leq ae^{-t^2/b},$$

où m_X est une médiane de X et $a, b \in \mathbb{R}_+^*$. Alors, $|m_X - \mathbb{E}(X)| \leq \frac{\sqrt{\pi}}{2} a \sqrt{b}$ et

$$\forall t \in \mathbb{R}_+^*, \quad \mu(|X - \mathbb{E}(X)| \geq t) \leq e^{\pi a^2/4} e^{-t^2/4b}.$$

2. Soit X une variable aléatoire réelle telle que

$$\forall t \in \mathbb{R}_+^*, \quad \mathbb{P}(|X - m_X| \geq t) \leq ae^{-t^2/b(m_X+t)},$$

où m_X est une médiane de X et $a, b \in \mathbb{R}_+^*$. Alors, $|\mathbb{E}(X) - m_X| \leq a \sqrt{\frac{1}{2} \pi b m_X} + 2abe^{-m_X/2b}$.

Démonstration. Tout d'abord,

$$\begin{aligned} |m_X - \mathbb{E}(X)| &= |\mathbb{E}(m_X - X)| \leq \mathbb{E}(|m_X - X|) \\ &= \int_{\mathbb{R}_+} \mathbb{P}(|m_X - X| > t) dt \leq a \int_{\mathbb{R}_+} e^{-t^2/b} dt = \frac{\sqrt{\pi}}{2} a \sqrt{b}. \end{aligned}$$

Notons $t_0 := \frac{\sqrt{\pi}}{2} a \sqrt{b}$. Si $t \geq 2t_0$,

$$\mu(|X - \mathbb{E}(X)| \geq t) \leq \mu(|X - m_X| \geq t - t_0) \leq \mu(|X - m_X| \geq t/2) \leq a e^{-t^2/4b}.$$

Si $t < 2t_0$, on peut utiliser la borne triviale

$$\mu(|X - \mathbb{E}(X)| \geq t) \leq 1 \leq e^{(2t_0)^2/4b} e^{-t^2/4b} = e^{\pi a^2/4} e^{-t^2/4b}.$$

On obtient donc

$$\forall t \in \mathbb{R}_+^*, \quad \mu(|X - \mathbb{E}(X)| \geq t) \leq (a \vee e^{\pi a^2/4}) e^{-t^2/4b},$$

et l'on vérifie aisément que $a \leq e^{\pi a^2/4}$ pour tout $a \in \mathbb{R}$.

La seconde affirmation se démontre de la même façon en observant que

$$\int_{\mathbb{R}_+} e^{-t^2/b(m_X+t)} dt \leq \int_0^{m_X} e^{-t^2/2bm_X} dt + \int_{m_X}^{\infty} e^{-t/2b} dt \leq \sqrt{\frac{1}{2}\pi b m_X} + 2be^{-m_X/2b}. \quad \square$$

Il suit ainsi, par exemple, du Théorème 4.9 que $|m_F - \mathbb{E}(F)| \leq 4c\sqrt{\pi}$ et

$$\forall t \in \mathbb{R}_+^*, \quad \mu(|F - \mathbb{E}(F)| \geq t) \leq e^{4\pi} e^{-t^2/16c^2}.$$

Exemple 4.13. Retournons à l'Exemple 4.11. L'espérance du nombre d'urnes non vides est donnée par

$$\mathbb{E}[F] = \left(1 - \left(1 - \frac{1}{n}\right)^m\right)n.$$

L'estimée de concentration obtenue dans l'Exemple 4.11 implique donc

$$\forall t \in \mathbb{R}_+^*, \quad \mu(|F - \mathbb{E}(F)| \geq t) \leq e^{4\pi} e^{-t^2/16(n \wedge m)}.$$

En particulier, dans la limite où $n \rightarrow \infty$ et $m/n \rightarrow \rho > 0$, on a $\lim_{n \rightarrow \infty} \mathbb{E}[F]/n = (1 - e^{-\rho})$ et la borne ci-dessus nous permet d'affirmer que $F \in [\mathbb{E}[F] - t_n, \mathbb{E}[F] + t_n]$ avec grande probabilité dès que $t_n \gg \sqrt{n}$. \diamond

4.4.2 Suprema de fonctions linéaires

Soit Y_1, \dots, Y_n des variables aléatoires indépendantes à valeurs dans $[0, 1]$ et considérons la variable aléatoire

$$Z := \max_{\xi \in \Xi} \sum_{i=1}^n \xi_i Y_i,$$

où $\Xi \subset \mathbb{R}^n$ est un ensemble fini. Notons $\rho := \max_{\xi \in \Xi} \|\xi\|_2$.

Considérons $\Omega := [0, 1]^n$, muni de la mesure produit μ des lois des variables Y_i . On souhaiterait appliquer le Théorème 4.9 à la fonction $F : \Omega \rightarrow \mathbb{R}$ définie par

$$F(x) := \max_{\xi \in \Xi} \sum_{i=1}^n \xi_i x_i.$$

Étant donné $x = (x_1, \dots, x_n) \in \Omega$, soit $\hat{\xi} = \hat{\xi}(x) \in \Xi$ réalisant le maximum. Alors, pour tout $y \in \Omega$,

$$F(x) = \sum_{i=1}^n \hat{\xi}_i x_i \leq \sum_{i=1}^n \hat{\xi}_i y_i + \sum_{i=1}^n |\hat{\xi}_i| |x_i - y_i| \leq F(y) + \rho \sum_{i=1}^n \frac{|\hat{\xi}_i|}{\|\hat{\xi}\|_2} \mathbf{1}_{\{x_i \neq y_i\}}.$$

Ainsi, la fonction F est ρ -Lipschitzienne au sens de Talagrand, pour les poids $\alpha(x) := |\hat{\xi}(x)| / \|\hat{\xi}(x)\|_2$. Une application du Théorème 4.9 conduit donc au résultat de concentration suivant.

Corollaire 4.14. Soit m_Z une médiane de Z . Alors,

$$\forall t \in \mathbb{R}_+^*, \quad \mu(|Z - m_Z| \geq t) \leq 4e^{-t^2/4\rho^2}.$$

Exemple 4.15 (Percolation de premier passage). Soit $G = (S, A)$ un graphe fini et $(Y_e)_{e \in A}$ des variables aléatoires i.i.d. à valeurs dans $[0, 1]$, appelées **temps de passage**. On interprète Y_e comme le temps nécessaire pour traverser l'arête e . Étant donné $x, y \in S$, une quantité centrale dans l'étude de la percolation de premier passage est le temps minimum nécessaire pour aller de x à y . Soit Π une collection finie de chemins auto-évitant de G reliant x à y . Le temps nécessaire pour parcourir le chemin $\pi \in \Pi$ est $Y_\pi := \sum_{e \in \pi} Y_e$. Le temps minimum pour aller de x à y en parcourant l'un des chemins de Π est alors donné par la variable aléatoire

$$Z := \min_{\pi \in \Pi} Y_\pi = \min_{\pi \in \Pi} \sum_{e \in A} \mathbf{1}_{\{e \in \pi\}} Y_e.$$

Soit m_Z une médiane de Z . Alors, une application du Corollaire 4.14 à la variable aléatoire $Z' := -Z = \max_{\pi \in \Pi} \sum_{e \in A} (-\mathbf{1}_{\{e \in \pi\}}) Y_e$ conduit à

$$\forall t \in \mathbb{R}_+^*, \quad \mu(|Z - m_Z| \geq t) \leq 4e^{-t^2/4\ell},$$

où $\ell := \max_{\pi \in \Pi} |\pi|$.

Sur le graphe \mathbb{Z}^2 (avec une arête entre $i, j \in \mathbb{Z}^2$ si $\|i - j\|_1 = 1$), lorsque les (Y_e) sont i.i.d. et telles que $\mu(Y_e = 0) < 1/2$, il existe C et C' indépendantes de n telles qu'avec probabilité au moins $1 - Ce^{-n/C}$, le temps minimal pour aller de $(0, 0)$ à $(n, 0)$ ne change pas si on se restreint à la famille Π des chemins de longueur au plus $C'n$ [13]. Combiné avec le résultat de concentration ci-dessus, cela montre qu'il existe $\epsilon > 0$, indépendant de n , telle que, pour tout n suffisamment grand,

$$\forall t \in [0, n], \quad \mu(|Z - m_Z| \geq t) \leq Ce^{-n/C} + 4e^{-t^2/4C'n} \leq e^{-\epsilon t^2/n}. \quad \diamond$$

4.4.3 Fonctions certifiables

Une classe de fonctions pour lesquelles l'inégalité de Talagrand est particulièrement efficace est celle des fonctions $h : \Omega \rightarrow \mathbb{R}$ ayant la propriété que lorsque $h(x) \geq s$, on peut trouver un « petit » nombre de coordonnées *certifiant* que $h(x) \geq s$.

Définition 4.16. Soit $\Omega = \times_{i=1}^n \Omega_i$ et $f : \mathbb{R} \rightarrow \mathbb{R}$. Une fonction $h : \Omega \rightarrow \mathbb{R}$ est **f -certifiable** si, pour tout $x \in \Omega$ tel que $h(x) \geq s$, il existe $I \subset \llbracket n \rrbracket$ avec $|I| \leq f(s)$ tel que

$$\forall i \in I, y_i = x_i \quad \implies \quad h(y) \geq s.$$

On dira que h est **r -certifiable** ($r \in \mathbb{R}$) si h est f -certifiable avec $f(s) = rs$.

Exemple 4.17. Soit $n \in \mathbb{N}$ et $N := \binom{n}{2}$. On peut encoder le graphe d'Erdős-Rényi $G(n, p)$ comme l'ensemble $\Omega := \{0, 1\}^N$ muni de la mesure produit $\mu = \nu^{\times N}$, où $\nu(\{1\}) = 1 - \nu(\{0\}) = p$. En effet, numérotons les arêtes de K_n , le graphe complet à n sommets, de façon arbitraire. On peut associer à

chaque $x \in \Omega$ un sous-graphe $G = G(x)$ de K_n en déclarant que l'arête numéro k appartient à G si et seulement si $x_k = 1$. On vérifie aisément que $G \sim G(n, p)$.

On s'intéresse à la fonction $G \mapsto h(G)$ comptant le nombre maximum de triangles distincts⁶ dans G . Alors, h est 3-certifiable. En effet, si $h(G) \geq s$, alors on peut trouver s triangles dans G , comprenant au plus $3s$ arêtes, et toute autre réalisation G' possédant également ces arêtes satisfait $h(G') \geq s$. \diamond

Théorème 4.18. Soit $\Omega = \times_{i=1}^n \Omega_i$, $\mu = \times_{i=1}^n \mu_i$ et $f : \mathbb{N} \rightarrow \mathbb{N}$. Soit $h : \Omega \rightarrow \mathbb{R}$ une fonction f -certifiable telle que $|h(x) - h(y)| \leq c$ lorsque x et y diffèrent en au plus une coordonnée. Alors,

$$\forall r \in \mathbb{R}, \forall t \in \mathbb{R}_+^*, \quad \mu(h \leq r - ct\sqrt{f(r)}) \mu(h \geq r) \leq e^{-t^2/4}. \quad (4.2)$$

En particulier, si m_h est une médiane de h , on a, pour tout $t \in \mathbb{R}_+^*$,

$$\mu(h \leq m_h - u) \leq 2e^{-u^2/4c^2f(m_h)}, \quad \mu(h \geq m_h + u) \leq 2e^{-u^2/4c^2f(m_h+u)}. \quad (4.3)$$

Démonstration. Soit $t \in \mathbb{R}_+^*$, $r \in \mathbb{R}$ et $A := \{y \in \Omega \mid h(y) < r - ct\sqrt{f(r)}\}$. Soit $x \in \Omega$ tel que $h(x) \geq r$ et soit $I \subset \llbracket n \rrbracket$ tel que $|I| \leq f(r)$ certifiant que $h(x) \geq r$. On définit $\alpha = \alpha(x) \in \mathbb{S}_+^n$ par

$$\alpha_i := \begin{cases} |I|^{-1/2} & \text{si } i \in I, \\ 0 & \text{sinon.} \end{cases}$$

Soit $\epsilon > 0$ arbitraire. Montrons que $d_\alpha(x, A) > t - \epsilon$ et donc, en particulier, que $x \notin A_{t-\epsilon}$. Supposons, par l'absurde que $d_\alpha(x, A) \leq t - \epsilon$. Alors, il existe $y \in A$ tel que

$$t \geq d_\alpha(x, y) = |I|^{-1/2} \sum_{i \in I} \mathbf{1}_{\{x_i \neq y_i\}},$$

et donc $\sum_{i \in I} \mathbf{1}_{\{x_i \neq y_i\}} \leq t|I|^{1/2} \leq t\sqrt{f(r)}$. Définissons $z \in \Omega$ par $z_i := x_i$ pour tout $i \in I$ et $z_i := y_i$ pour tout $i \in \llbracket n \rrbracket \setminus I$. Comme z et x coïncident sur I , on a $h(z) \geq r$. De plus, y et z différant en au plus $t\sqrt{f(r)}$ coordonnées, on a $|h(y) - h(z)| \leq ct\sqrt{f(r)}$. Il s'ensuit que $h(y) \geq h(z) - |h(y) - h(z)| \geq r - ct\sqrt{f(r)}$, ce qui contredit le fait que $y \in A$.

On a donc $\mu(h \geq r) \leq 1 - \mu(A_{t-\epsilon})$ et le Théorème 4.4 implique que

$$\mu(h < r - ct\sqrt{f(r)}) \mu(h \geq r) \leq \mu(A)(1 - \mu(A_{t-\epsilon})) \leq e^{-(t-\epsilon)^2/4}.$$

L'inégalité (4.2) suit en laissant $\epsilon \downarrow 0$, puis en exploitant la continuité du membre de droite, comme dans la preuve du Théorème 4.9.

Les deux conséquences énoncées en (4.3) sont obtenues en appliquant (4.2) avec $r := m_h$ et $t := u/c\sqrt{f(m_h)}$ pour la première, et avec $r := m_h + u$ et $t := u/c\sqrt{f(m_h + u)}$ pour la seconde. \square

Exemple 4.19. Soit G un graphe d -régulier⁷, sans boucle, avec n sommets. Soit H le sous-graphe aléatoire de G obtenu en conservant chaque arête de G indépendamment avec probabilité p . On peut encoder H par un élément de $\Omega := \{0, 1\}^{nd/2}$, en numérotant les arêtes de G et en associant à l'arête k la valeur 1 si l'arête est présente dans H . Étant donné $x \in \Omega$, on dénotera $H(x)$ le sous-graphe correspondant.

Notons $h(x)$ le nombre de sommets non isolés de $H(x)$ (c'est-à-dire, de degré au moins 1). Clairement, $\mathbb{E}[h] = n(1 - (1-p)^d)$. Modifier une coordonnée de $x \in \Omega$ correspond à retirer ou ajouter une arête à $H(x)$ et ne peut donc changer le nombre de sommets isolés que d'au plus 2; on a donc $|h(x) - h(y)| \leq 2$ lorsque x et y diffèrent en au plus une coordonnée. De plus, h est 1-certifiable,

6. Un **triangle** de G est un triplet de sommets distincts $i, j, k \in \llbracket n \rrbracket$ tels que $\{i, j\}, \{i, k\}, \{j, k\}$ soient des arêtes de G .

7. Un graphe est **d -régulier** si tous ses sommets sont de degré d .

puisque l'existence d'au moins s sommets non isolés peut être certifiée en révélant la présence d'une arête incidente à chacun de s sommets non isolés distincts. On peut donc appliquer le Théorème 4.18, ce qui donne, pour toute médiane m_h de h ,

$$\mu(h \leq m_h - u) \leq 2e^{-u^2/16m_h}, \quad \mu(h \geq m_h + u) \leq 2e^{-u^2/16(m_h+u)}.$$

En particulier, en combinant ces bornes et le point 2 du Lemme 4.12, on conclut que lorsque $u_n \gg \sqrt{n}$, on a $h \in [n(1 - (1-p)^d) - u_n, n(1 - (1-p)^d) + u_n]$ avec grande probabilité. \diamond

Exemple 4.20 (Plus longue sous-suite croissante). Soit σ une permutation aléatoire uniforme de l'ensemble $\llbracket n \rrbracket$. Un problème très étudié est celui de déterminer la longueur $L = L(\sigma)$ de la plus grande sous-suite $(\sigma(i_1), \dots, \sigma(i_L))$, $1 \leq i_1 < i_2 < \dots < i_L \leq n$, de la suite $(\sigma(1), \dots, \sigma(n))$ telle que $\sigma(i_1) < \sigma(i_2) < \dots < \sigma(i_L)$.

Commençons par reformuler ce problème dans le langage de ce chapitre. Soit $\Omega := [0, 1]^n$ et $\mu := \nu^{\times n}$ le produit des mesures uniformes sur $[0, 1]$. On note $x = (x_1, \dots, x_n)$ les éléments de Ω . On considère la fonction $h : \Omega \rightarrow \mathbb{R}$ définie par

$$h(x) := \max\{k \in \mathbb{N} \mid \exists 1 \leq i_1 < i_2 < \dots < i_k \leq n, x_{i_1} < x_{i_2} < \dots < x_{i_k}\}.$$

En d'autres termes, $h(x)$ est la longueur de la plus grande sous-suite croissante de x . μ n'ayant aucun atome, on vérifie facilement que h possède la même loi que $L(\sigma)$. Commençons par montrer que h est d'ordre \sqrt{n} avec grande probabilité.

Soit $\ell := \lceil \sqrt{n} \rceil$, $N := \lfloor n/\ell \rfloor$, $I_k := \{(k-1)\ell + 1, \dots, k\ell\}$ et $J_k := [(k-1)n^{-1/2}, kn^{-1/2}]$, $k \in \mathbb{N}^*$. Notons $Y_k(x) := \mathbf{1}_{\{\exists j \in I_k \text{ tel que } x_j \in J_k\}}$. Les variables aléatoires Y_1, \dots, Y_N sont i.i.d., chacune de loi de Bernoulli de paramètre

$$p := \mu(Y_i = 1) = 1 - (1 - n^{-1/2})^\ell \geq 1 - e^{-n^{-1/2}\ell} \geq 1 - e^{-1}.$$

On utilise l'algorithme suivant pour construire une sous-suite croissante à partir de x : pour chaque $k \in \llbracket N \rrbracket$, on inclut i dans la sous-suite si i est le premier nombre de l'intervalle I_k tel que $x_i \in J_k$. Ceci produit une sous-suite de longueur $Y := Y_1 + \dots + Y_N$. Par conséquent, il suit du Lemme 1.6 que

$$\mu(h \leq \frac{1}{2}pN) \leq \mu(Y \leq \frac{1}{2}\mathbb{E}[Y]) \leq e^{-pN/8} \leq e^{-(1-e^{-1})(\sqrt{n}-2)/8},$$

ce qui tend rapidement vers 0 lorsque $n \rightarrow \infty$ et établit la borne inférieure.

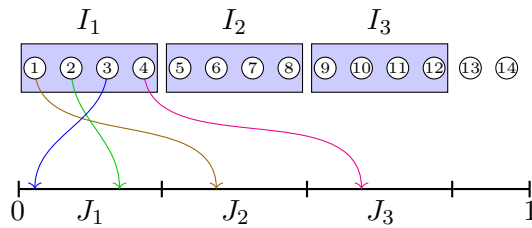


FIGURE 4.3: Extraction d'une sous-suite croissante. Dans chaque bloc I_k , on inclut le premier point $i \in I_k$ tel que $x_i \in J_k$ (s'il en existe au moins un). Pour des raisons de lisibilité, seul les images des points de I_1 sont indiquées. Le premier élément de la sous-suite croissante est donc ici égal à 2. Dans cet exemple, $n = 14$, $\ell = 4$ et $N = 3$.

Passons à la borne supérieure. La probabilité qu'une sous-suite donnée $i_1 < i_2 < \dots < i_k$ soit croissante est égale à $1/k!$. Notons M_R le nombre de telles sous-suites de longueur au moins $R \in \llbracket n \rrbracket$. On a⁸

$$\mathbb{E}(M_R) = \sum_{k=R}^n \binom{n}{k} \frac{1}{k!} \leq \sum_{k=R}^n \left(\frac{ne}{k}\right)^k \left(\frac{e}{k}\right)^k = \sum_{k=R}^n \frac{n^k e^{2k}}{k^{2k}}.$$

8. On utilise les inégalités établies dans la note en bas de page 8.

Avec le choix $R := \lceil 4e\sqrt{n} \rceil$, on obtient

$$\mathbb{E}(M_R) \leq \sum_{k=R}^n \frac{n^k e^{2k}}{R^{2k}} \leq \sum_{k=R}^n 4^{-2k} \leq \sum_{k=R}^{\infty} 16^{-k} \leq \frac{16}{15} 16^{-4e\sqrt{n}}.$$

Par conséquent, par l'inégalité de Markov,

$$\mu(h > 4e\sqrt{n}) \leq \mu(M_R \geq 1) \leq \mathbb{E}[M_R] \leq \frac{16}{15} 16^{-4e\sqrt{n}},$$

ce qui tend également rapidement vers 0 lorsque $n \rightarrow \infty$ et établit la borne supérieure.

Soit m_h une médiane de h . Il suit des estimées précédentes que m_h est d'ordre \sqrt{n} . Nous allons à présent établir la concentration de h autour de m_h .

La première observation est que changer une unique composante de x ne peut faire croître ou décroître la longueur de la plus longue sous-suite croissante que d'au plus 1. La seconde observation est que h est 1-certifiable. En effet, si $h(x) \geq s$, alors les s composantes formant une des sous-suites croissantes de longueur s certifient cette inégalité.

Il suit donc du Théorème 4.18 que

$$\mu(h \leq m_h - u) \leq 2e^{-u^2/4m_h}, \quad \mu(h \geq m_h + u) \leq 2e^{-u^2/4(m_h+u)}.$$

En particulier, ces bornes montrent que si $u_n \gg n^{1/4}$, alors $h \in [m_h - u_n, m_h + u_n]$ avec grande probabilité.

On dispose aujourd'hui d'informations considérablement plus détaillées sur ce problème. En particulier, $\mathbb{E}[L(\sigma)] = 2\sqrt{n}$ et le remarquable résultat suivant détermine le comportement asymptotique des fluctuations autour de l'espérance [3] :

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{L(\sigma) - 2\sqrt{n}}{n^{1/6}} \leq s\right) = F_2(s),$$

où F_2 est la fonction de répartition de la loi de Tracy–Widom,

$$F_2(s) := \exp\left(-\int_s^{\infty} (x-s)q^2(x) dx\right)$$

où q est l'unique solution de l'équation différentielle de Painlevé de type II,

$$y''(s) = sy(s) + 2y(s)^3,$$

avec des conditions au bord appropriées. En particulier, ce résultat montre que l'approche que l'on a utilisée ci-dessus ne fournit pas l'ordre de grandeur correct des fluctuations. Le lecteur intéressé par ces questions trouvera une discussion approfondie dans le livre [23]. \diamond

Remarques bibliographiques : Ce chapitre est basé sur les livres [1] et [16], ainsi que sur la source originale, l'article [25].

5 Réduction de la dimensionnalité : le lemme de Johnson–Lindenstrauss

À l'ère du big data, on est fréquemment confronté à des jeux de données consistant en un très grand nombre de points dans un espace de très grande dimension. Malheureusement, les algorithmes usuels permettant l'extraction d'information à partir de ces données souffrent alors du « fléau de la dimension » (*curse of dimensionality* en anglais) : ils deviennent extrêmement inefficaces sur un plan pratique, lorsque la dimension croît. Il est alors désirable de réduire la dimension des données tout en préservant les propriétés essentielles. C'est ce que l'on appelle le problème de la réduction de la dimensionnalité. Il en existe de nombreuses approches. Le but de ce chapitre est de présenter un résultat classique de ce type : le lemme de Johnson–Lindenstrauss. Pour une importante classe d'algorithmes (par exemple, l'algorithme de recherche des plus proches voisins et ses nombreuses variantes), la propriété à préserver est la distance entre chaque paire de points. Le lemme de Johnson–Lindenstrauss affirme qu'un ensemble de m points dans un espace de dimension arbitraire peut être plongé dans un espace de dimension $O(\log m)/\epsilon^2$ tout en ne modifiant la distance entre chaque paire de points qu'au plus d'un facteur compris entre $1 - \epsilon$ et $1 + \epsilon$, avec $\epsilon \in (0, 1)$ arbitraire. Ce lemme (et ses variantes) trouve des applications dans de nombreux domaines. Un exemple est donné dans la section 5.3.

5.1 Cadre et énoncé du Lemme de Johnson–Lindenstrauss

Théorème 5.1 (Lemme de Johnson–Lindenstrauss). *Soit $N, m \in \mathbb{N}^*$ et $\epsilon \in (0, 1)$. Pour tout $n \geq 2(\epsilon^2 - \frac{1}{3}\epsilon^3)^{-1} \log m$ et tout ensemble $V \subset \mathbb{R}^N$ composé de m points, il existe une application $f : \mathbb{R}^N \rightarrow \mathbb{R}^n$ telle que*

$$\forall \mathbf{x}, \mathbf{y} \in V, \quad (1 - \epsilon)\|\mathbf{x} - \mathbf{y}\|_2 \leq \|f(\mathbf{x}) - f(\mathbf{y})\|_2 \leq (1 + \epsilon)\|\mathbf{x} - \mathbf{y}\|_2. \quad (5.1)$$

Remarque 5.2. *La preuve fait recours à la méthode probabiliste : nous allons construire une application linéaire aléatoire dont nous montrerons qu'elle satisfait la condition désirée avec probabilité positive, ce qui implique l'existence d'une fonction f comme dans l'énoncé. En particulier, cet argument n'est pas constructif. Ce n'est toutefois pas un problème en pratique, car la probabilité d'obtenir une application satisfaisant (5.1) est supérieure ou égale à $1/m$; il suffit donc de faire de l'ordre de m tirages pour obtenir l'application désirée. En fait, quitte à empirer très légèrement la borne sur n , on peut obtenir une probabilité de succès proche de 1 : si $n \geq (2 + \alpha)(\epsilon^2 - \frac{1}{3}\epsilon^3)^{-1} \log m$, pour un $\alpha > 0$, alors la même preuve (il suffit de choisir $\delta = m^{-2-\alpha}$) montre que la probabilité de succès est d'au moins $1 - m^{-\alpha}$. \diamond*

Remarque 5.3. *On peut montrer [15] qu'il n'est pas possible de faire mieux qu'une borne de la forme $n \geq O(\epsilon^{-2}) \log m$. \diamond*

5.2 Preuve du Lemme de Johnson–Lindenstrauss

Soit $A = (A_{ij})$ une matrice $n \times N$ dont les éléments sont des variables aléatoires i.i.d. de loi $\mathcal{N}(0, 1)$. Alors, pour tout $\mathbf{z} = (z_1, \dots, z_N) \in \mathbb{R}^N$,

$$\mathbb{E}[\|\mathbf{Az}\|_2^2] = \mathbb{E}\left[\sum_{i=1}^n \left(\sum_{j=1}^N A_{ij}z_j\right)^2\right] = \sum_{i=1}^n \sum_{j=1}^N \sum_{k=1}^N \mathbb{E}[A_{ij}A_{ik}]z_jz_k = n\|\mathbf{z}\|_2^2,$$

puisque $\mathbb{E}[A_{ij}A_{ik}] = \delta_{jk}$ par hypothèse. Il suit que l'application linéaire $L := n^{-1/2}A$ préserve, en moyenne, le carré de la norme euclidienne : $\mathbb{E}[\|L\mathbf{z}\|_2^2] = \|\mathbf{z}\|_2^2$. L est donc un bon candidat pour l'application recherchée, pour peu que l'on parvienne à montrer que la distribution de $\|L\mathbf{z}\|_2^2$ est concentrée au voisinage de son espérance.

Lemme 5.4. Soit $\delta, \epsilon \in (0, 1)$. Pour tout $n \geq |\log \delta|/(\epsilon^2 - \frac{1}{3}\epsilon^3)$, on a

$$\forall \mathbf{z} \in \mathbb{R}^N \text{ tel que } \|\mathbf{z}\|_2 = 1, \quad \mathbb{P}(|\|\mathbf{Lz}\|_2 - 1| \geq \epsilon) \leq 2\delta.$$

Avant de démontrer ce lemme, vérifions qu'il permet de conclure la preuve du Théorème 5.1.

On prend $\delta := 1/m^2$. Soit $\mathbf{x}, \mathbf{y} \in V$ distincts. Comme $n \geq 2(\epsilon^2 - \frac{1}{3}\epsilon^3)^{-1} \log m = |\log \delta|/(\epsilon^2 - \frac{1}{3}\epsilon^3)$, on peut appliquer le Lemme 5.4 à $\mathbf{z} := (\mathbf{x} - \mathbf{y})/\|\mathbf{x} - \mathbf{y}\|_2$, ce qui donne

$$\mathbb{P}\left(\frac{\|\mathbf{Lx} - \mathbf{Ly}\|_2}{\|\mathbf{x} - \mathbf{y}\|_2} \notin (1 - \epsilon, 1 + \epsilon)\right) \leq 2\delta.$$

Par conséquent, par l'inégalité de Boole,

$$\begin{aligned} \mathbb{P}\left(\exists \{\mathbf{x}, \mathbf{y}\} \subset V, \frac{\|\mathbf{Lx} - \mathbf{Ly}\|_2}{\|\mathbf{x} - \mathbf{y}\|_2} \notin (1 - \epsilon, 1 + \epsilon)\right) &\leq \sum_{\{\mathbf{x}, \mathbf{y}\} \subset V} \mathbb{P}\left(\frac{\|\mathbf{Lx} - \mathbf{Ly}\|_2}{\|\mathbf{x} - \mathbf{y}\|_2} \notin (1 - \epsilon, 1 + \epsilon)\right) \\ &\leq \binom{m}{2} 2\delta = 1 - \frac{1}{m}. \end{aligned}$$

On en conclut que la probabilité que L possède la propriété (5.1) est strictement positive, ce qui démontre l'existence d'une telle application.

Preuve du Lemme 5.4. Tout d'abord, pour tout $i \in \llbracket m \rrbracket$,

$$(\mathbf{Az})_i = \sum_{j=1}^N A_{ij}z_j \sim \mathcal{N}(0, 1),$$

puisque les variables aléatoires $A_{ij}z_j \sim \mathcal{N}(0, z_j^2)$ sont indépendantes et $z_1^2 + \dots + z_N^2 = \|\mathbf{z}\|_2^2 = 1$. En particulier, la variable aléatoire

$$W := \|\mathbf{Az}\|_2^2 = \sum_{i=1}^n (\mathbf{Az})_i^2$$

est la somme des carrés de n variables aléatoires indépendantes de loi $\mathcal{N}(0, 1)$ ¹. Pour $Z \sim \mathcal{N}(0, 1)$, on a

$$\forall t \in (-\infty, \frac{1}{2}), \quad \mathbb{E}[e^{tZ^2}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} e^{tx^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(1-2t)x^2} dx = \frac{1}{\sqrt{1-2t}}.$$

1. En d'autres termes, W suit une loi du χ^2 à n degrés de liberté.

Par conséquent,

$$\mathbb{E}[e^{tW}] = \mathbb{E}[e^{tZ^2}]^n = (1 - 2t)^{-n/2}.$$

Bornons tout d'abord la probabilité que $\|Lz\|_2 \geq 1 + \epsilon$. Une application de l'inégalité de Chernoff conduit à la borne

$$\mathbb{P}(\|Lz\|_2 \geq 1 + \epsilon) = \mathbb{P}(\|Az\|_2^2 \geq (1 + \epsilon)^2 n) = \mathbb{P}(W \geq (1 + \epsilon)^2 n) \leq \inf_{t \geq 0} e^{-t(1+\epsilon)^2 n} \mathbb{E}[e^{tW}].$$

On en conclut que

$$\mathbb{P}(\|Lz\|_2 \geq 1 + \epsilon) \leq \inf_{t \geq 0} \exp[-n(t(1 + \epsilon)^2 + \frac{1}{2} \log(1 - 2t))].$$

L'infimum est atteint en $t = \frac{1}{2}(1 - (1 + \epsilon)^{-2})$, ce qui conduit à

$$\mathbb{P}(\|Lz\|_2 \geq 1 + \epsilon) \leq \exp[-n(\epsilon + \frac{1}{2}\epsilon^2 - \log(1 + \epsilon))] \leq \exp[-(\epsilon^2 - \frac{1}{3}\epsilon^3)n],$$

la dernière inégalité suivant de la borne $\log(1 + \epsilon) \leq \epsilon - \frac{1}{2}\epsilon^2 + \frac{1}{3}\epsilon^3$ valide pour tout $\epsilon \geq 0$.

Bornons à présent la probabilité que $\|Lz\|_2 \leq 1 - \epsilon$. On procède de manière similaire :

$$\mathbb{P}(\|Lz\|_2 \leq 1 - \epsilon) = \mathbb{P}(-W \geq -(1 - \epsilon)^2 n) \leq \inf_{t \geq 0} e^{t(1-\epsilon)^2 n} \mathbb{E}[e^{-tW}] = \inf_{t \geq 0} e^{t(1-\epsilon)^2 n} (1 + 2t)^{-n/2}.$$

L'infimum étant atteint en $t = \frac{1}{2}((1 - \epsilon)^{-2} - 1)$, on obtient

$$\mathbb{P}(\|Lz\|_2 \leq 1 - \epsilon) \leq \exp[-n(-\epsilon + \frac{1}{2}\epsilon^2 - \log(1 - \epsilon))] \leq \exp[-\epsilon^2 n],$$

puisque $\log(1 - \epsilon) \leq -\epsilon - \frac{1}{2}\epsilon^2$ pour tout $\epsilon \in [0, 1)$.

On obtient donc finalement

$$\mathbb{P}(\|Lz\|_2 \notin (1 - \epsilon, 1 + \epsilon)) \leq \exp[-(\epsilon^2 - \frac{1}{3}\epsilon^3)n] + \exp[-\epsilon^2 n] \leq 2 \exp[-(\epsilon^2 - \frac{1}{3}\epsilon^3)n] \leq 2\delta,$$

pour tout $n \geq |\log \delta| / (\epsilon^2 - \frac{1}{3}\epsilon^3)$. □

5.3 Application à l'acquisition comprimée

5.3.1 Sparsité et propriété d'isométrie restreinte

De nombreux types de signaux (par exemple, le son, les images ou la vidéo) peuvent être encodés sous la forme d'un vecteur $\mathbf{x} \in \mathbb{R}^N$ avec N (très) grand. Plutôt que de mesurer directement le signal \mathbf{x} , on mesure $\mathbf{y} = M\mathbf{x} \in \mathbb{R}^n$ avec $n \ll N$, où M est une matrice $n \times N$ (par exemple en ne conservant qu'une partie des coefficients de Fourier). Le problème est de reconstruire le signal \mathbf{x} à partir de la mesure \mathbf{y} (et de la connaissance de la matrice de mesure M). Comme on a supposé $n \ll N$, ce problème semble mal posé : après tout, le problème est sévèrement sous-déterminé ! Néanmoins, dans de très nombreuses situations pratiques, on a des raisons de s'attendre à ce que le signal \mathbf{x} soit sparse (épars en français, mais presque tout le monde utilise le terme en anglais). Par exemple, la norme de compression d'images JPEG 2000 repose sur le fait qu'une image peut être encodée dans une base d'ondelettes, de sorte à ce que la qualité visuelle soit peu impactée même lorsque l'on ne conserve qu'un petit nombre de coefficients. La question devient alors : dans l'hypothèse où le signal \mathbf{x} est sparse, sous quelles conditions sur n est-il possible de reconstruire \mathbf{x} à partir de \mathbf{y} ? Nous allons voir dans cette section qu'une réponse à cette question peut être obtenue à l'aide du lemme de Johnson-Lindenstrauss.

Commençons par préciser la notion de vecteur sparse.

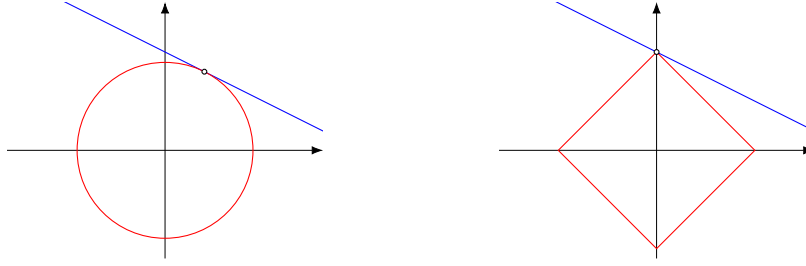


FIGURE 5.1: Minimiser la norme ℓ^1 sur un sous-espace affine tend à produire une solution sparse (droite), contrairement au cas de la norme ℓ^2 (gauche).

Définition 5.5. Soit $k \in \mathbb{N}^*$. Un vecteur $\mathbf{x} \in \mathbb{R}^N$ est k -sparse s'il possède au plus k composantes non nulles. On notera \mathcal{S}_k^N l'ensemble de tous les vecteurs k -sparse.

Des considérations élémentaires d'algèbre linéaire montrent qu'une condition nécessaire et suffisante pour que l'application $\mathbf{x} \mapsto M\mathbf{x}$ soit injective (de sorte à ce que le problème soit bien posé) est qu'aucune famille de $2k$ colonnes de M ne soit linéairement indépendante. En effet, observons que pour toute paire $\mathbf{x}, \mathbf{x}' \in \mathcal{S}_k^N$, on a $\mathbf{x} - \mathbf{x}' \in \mathcal{S}_{2k}^N$. Par conséquent, il existe deux vecteurs distincts $\mathbf{x}, \mathbf{x}' \in \mathcal{S}_k^N$ tels que $M(\mathbf{x} - \mathbf{x}') = \mathbf{0}$ si et seulement si les colonnes de M correspondant aux composantes non nulles de $\mathbf{x} - \mathbf{x}'$ sont linéairement dépendantes. Observons que cette condition est satisfaite (presque sûrement) par la matrice L introduite dans la preuve du Lemme de Johnson–Lindenstrauss. Le problème est qu'étant donné $\mathbf{y} \in \mathbb{R}^n$, il n'est pas aisé de trouver le vecteur $\mathbf{x} \in \mathcal{S}_k^N$ tel que $L\mathbf{x} = \mathbf{y}$. Nous allons à présent décrire une approche plus efficace. Celle-ci repose sur la notion suivante.

Définition 5.6. Soit $k \in \mathbb{N}^*$ et $\epsilon \in (0, 1)$. Une application linéaire $M : \mathbb{R}^N \rightarrow \mathbb{R}^n$ possède la **propriété d'isométrie restreinte** d'ordre k et de constante ϵ si

$$\forall \mathbf{x} \in \mathcal{S}_k^N, \quad (1 - \epsilon)\|\mathbf{x}\|_2 \leq \|M\mathbf{x}\|_2 \leq (1 + \epsilon)\|\mathbf{x}\|_2.$$

Dans ce cas, on dira que M vérifie $\text{RIP}_{k,\epsilon}$.

L'intérêt de cette notion est démontré par le résultat suivant, qui réduit le problème de reconstruction d'un signal sparse à un problème d'optimisation pouvant être implémenté de façon numériquement efficace.

Proposition 5.7. Soit $M : \mathbb{R}^N \rightarrow \mathbb{R}^n$ une application linéaire vérifiant $\text{RIP}_{10k,1/3}$. Alors, pour chaque $\mathbf{x} \in \mathcal{S}_k^N$, le problème d'optimisation

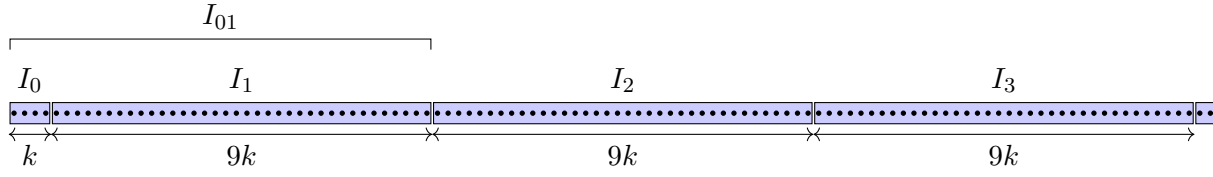
$$\text{minimiser } \|\mathbf{z}\|_1 \text{ sous la contrainte que } M\mathbf{z} = M\mathbf{x}$$

possède une unique solution (évidemment donnée par \mathbf{x}).

Remarque 5.8. Il est important d'observer que minimiser la norme ℓ^1 sur un sous-espace affine tend à générer une solution sparse, contrairement à ce qui se produit avec la norme ℓ^2 . Ceci est illustré (dans le cas de \mathbb{R}^2) dans la Fig. 5.1. \diamond

Évidemment, le problème à présent est d'obtenir des transformations linéaires vérifiant cette condition. Il se trouve que la même approche que pour le lemme de Johnson–Lindenstrauss fonctionne ici. Soit A une matrice $n \times N$ dont les éléments sont des variables aléatoires i.i.d. de loi $\mathcal{N}(0, 1)$ et soit $L := n^{-1/2}A$.

Proposition 5.9. Pour tout $n \geq 500k \log N$, L vérifie $\text{RIP}_{10k,1/3}$ avec probabilité au moins $1 - 1/N$.

FIGURE 5.2: La décomposition de \mathbb{N}^* employée dans la preuve de la Proposition 5.7 (ici, avec $k = 4$).

En combinant les Propositions 5.7 et 5.9, on obtient le résultat principal de cette section.

Théorème 5.10. *Soit $n \geq 500k \log N$. Avec probabilité au moins $1 - 1/N$, on peut reconstruire n'importe quel vecteur $\mathbf{x} \in \mathcal{S}_k^N$ à partir de la connaissance de $L\mathbf{x}$ en résolvant le problème d'optimisation*

$$\text{minimiser } \|\mathbf{z}\|_1 \text{ sous la contrainte que } L\mathbf{z} = L\mathbf{x}.$$

5.3.2 Preuve des Propositions 5.7 et 5.9

Preuve de la Proposition 5.7. Fixons $\mathbf{x} = (x_1, \dots, x_N) \in \mathcal{S}_k^N$. Sans perte de généralité, nous supposons que $x_i = 0$ pour tout $i > k$. Soit $\mathbf{z} \in \mathbb{R}^N$ tel que $M\mathbf{z} = M\mathbf{x}$ et $\|\mathbf{z}\|_1 \leq \|\mathbf{x}\|_1$. Notons $\mathbf{r} = (r_1, \dots, r_N) := \mathbf{z} - \mathbf{x}$. Sans perte de généralité, on supposera que $|r_i| \geq |r_{i+1}|$ pour tout $i > k$. Nous allons montrer que l'on a alors nécessairement $\mathbf{r} = \mathbf{0}$, c'est-à-dire $\mathbf{z} = \mathbf{x}$, ce qui conclura la preuve.

Soit $a_1 := k$ et, pour $i \geq 2$, $a_i := a_{i-1} + 9k$. On partitionne $\mathbb{N}^* = \bigcup_{i \geq 0} I_i$, avec $I_0 := [1, k]$ et, pour $i \geq 1$, $I_i := \{a_i + 1, \dots, a_{i+1}\}$. On notera également $I_{\geq j} := \bigcup_{i \geq j} I_i$ et $I_{01} := I_0 \cup I_1$. Si $\mathbf{y} = (y_1, \dots, y_N) \in \mathbb{R}^N$ et $J \subset \mathbb{N}^*$, on notera \mathbf{y}_J le vecteur de \mathbb{R}^N avec composantes $(\mathbf{y}_J)_i := \mathbf{1}_{\{i \in J\}} y_i$.

On a supposé $\mathbf{x}_{I_{\geq 1}} = \mathbf{0}$. Par conséquent, on a d'une part

$$\|\mathbf{z}\|_1 = \|\mathbf{z}_{I_0}\|_1 + \|\mathbf{z}_{I_{\geq 1}}\|_1 = \|\mathbf{z}_{I_0}\|_1 + \|\mathbf{r}_{I_{\geq 1}}\|_1,$$

et, d'autre part,

$$\|\mathbf{z}\|_1 \leq \|\mathbf{x}\|_1 = \|\mathbf{x}_{I_0}\|_1 \leq \|\mathbf{z}_{I_0}\|_1 + \|\mathbf{r}_{I_0}\|_1.$$

En combinant ces deux observations, on conclut que

$$\|\mathbf{r}_{I_{\geq 1}}\|_1 \leq \|\mathbf{r}_{I_0}\|_1. \quad (5.2)$$

Observons à présent que $M\mathbf{r} = M\mathbf{z} - M\mathbf{x} = \mathbf{0}$, ce que l'on peut réécrire $M\mathbf{r}_{I_{01}} = -\sum_{i \geq 2} M\mathbf{r}_{I_i}$. Il suit donc de l'hypothèse que M satisfait $\text{RIP}_{10k, 1/3}$ que

$$\frac{2}{3} \|\mathbf{r}_{I_{01}}\|_2 \leq \|M\mathbf{r}_{I_{01}}\|_2 \leq \sum_{i \geq 2} \|M\mathbf{r}_{I_i}\|_2 \leq \frac{4}{3} \sum_{i \geq 2} \|\mathbf{r}_{I_i}\|_2, \quad (5.3)$$

puisque $\mathbf{r}_{I_{01}}$ est $10k$ -sparse et chaque \mathbf{r}_{I_i} , $i \geq 2$, est $9k$ -sparse.

On a supposé que $|r_\ell| \geq |r_{\ell+1}|$ pour tout $\ell > k$. En particulier, pour tout $i \geq 1$ et tout $j \in I_{i+1}$, $|r_j| \leq \min\{|r_k| \mid k \in I_i\} \leq \|\mathbf{r}_{I_i}\|_1 / (9k)$. On a donc

$$\forall i \geq 1, \quad \|\mathbf{r}_{I_{i+1}}\|_2^2 \leq 9k \left(\frac{\|\mathbf{r}_{I_i}\|_1}{9k} \right)^2 = \frac{\|\mathbf{r}_{I_i}\|_1^2}{9k}.$$

Ceci implique que

$$\sum_{i \geq 2} \|\mathbf{r}_{I_i}\|_2 \leq \sum_{i \geq 1} \frac{\|\mathbf{r}_{I_i}\|_1}{3\sqrt{k}} = \frac{\|\mathbf{r}_{I_{\geq 1}}\|_1}{3\sqrt{k}} \leq \frac{\|\mathbf{r}_{I_0}\|_1}{3\sqrt{k}},$$

où l'on a utilisé (5.2). Par Cauchy–Schwarz, $\|r_{I_0}\|_1 \leq \sqrt{k}\|r_{I_0}\|_2$. On obtient donc finalement

$$\sum_{i \geq 2} \|r_{I_i}\|_2 \leq \frac{1}{3} \|r_{I_0}\|_2 \leq \frac{1}{3} \|r_{I_{01}}\|_2.$$

Combiné avec (5.3), on en déduit que $\|r_{I_{01}}\|_2 = 0$ et donc que $r_{I_{01}} = \mathbf{0}$. Les entrées $|r_\ell|$, $\ell \geq k$, étant décroissantes, il suit que $\mathbf{r} = \mathbf{0}$, comme souhaité. \square

Preuve de la Proposition 5.9. Notons $k' := 10k$ et $\Lambda(\mathbf{z}) := \|\|L\mathbf{z}\|_2 - 1\|$.

Soit $\mathcal{I} = \mathcal{I}(k', N) := \{I \subset [N] \mid |I| = k'\}$. Étant donné $I \in \mathcal{I}$, notons

$$\mathcal{S}_I := \{\mathbf{x} \in \mathcal{S}_{k'}^N \mid \mathbf{x}_{I^c} = \mathbf{0}, \|\mathbf{x}\|_2 = 1\},$$

où l'on a noté $I^c := [N] \setminus I$.

La proposition sera démontrée une fois que l'on aura établi que

$$\mathbb{P}\left(\sup\left\{\Lambda(\mathbf{z}) \mid \mathbf{z} \in \bigcup_{I \in \mathcal{I}} \mathcal{S}_I\right\} \leq 1/3\right) \geq 1 - \frac{1}{N}. \quad (5.4)$$

Il semble naturel d'utiliser le Lemme 5.4. Notons cependant que, contrairement au lemme de Johnson–Lindenstrauss qui ne s'applique qu'à des collections finies de points, l'événement intervenant dans (5.4) porte sur une infinité non dénombrable de vecteurs \mathbf{z} , ce qui rend impossible un simple argument basé sur l'inégalité de Boole. Afin de contourner cette difficulté, nous allons tout d'abord « discrétiser » les ensembles \mathcal{S}_I .

On pose $\rho := (6^3 N \log N)^{-1/2}$. Soit $D_I := \{\mathbf{x}_1, \dots, \mathbf{x}_K\} \subset \mathcal{S}_I$ tel que

$$\bigcup_{i=1}^K \{\mathbf{y} \in \mathcal{S}_I \mid \|\mathbf{y} - \mathbf{x}_i\|_2 \leq \rho\} = \mathcal{S}_I$$

et K soit minimal. En particulier, $K \leq (3/\rho)^{k'}$. En effet, on peut générer une telle famille $\mathbf{x}_1, \dots, \mathbf{x}_K$ en choisissant successivement les points $\mathbf{x}_i \in \mathcal{S}_I$ de sorte à ce que $\|\mathbf{x}_i - \mathbf{x}_j\|_2 > \rho$ pour tout $1 \leq j < i$; la sélection s'achève lorsqu'il n'est plus possible de trouver de tels points. Notons $\Pi_I : \mathbb{R}^N \rightarrow \mathbb{R}^{k'}$ la projection sur les coordonnées appartenant à I et $\tilde{\mathbf{x}}_i := \Pi_I \mathbf{x}_i \in \mathbb{R}^{k'}$. Par construction, les boules

$$\{\mathbf{y} \in \mathbb{R}^{k'} \mid \|\tilde{\mathbf{x}}_i - \mathbf{y}\|_2 \leq \rho/2\}$$

sont disjointes; leur volume total est donc égal à $2^k K c_{k'} (\rho/2)^{k'}$. De plus, toutes ces boules sont incluses dans la boule de $\mathbb{R}^{k'}$ de rayon $3/2$ et centrée en 0. Le volume de cette dernière étant égal à $c_{k'} (3/2)^{k'}$, on doit bien avoir $K \leq (3/\rho)^{k'}$.

Nous allons à présent appliquer le Lemme 5.4 aux points de $\bigcup_{I \in \mathcal{I}(k', N)} D_I$. On fixe $\epsilon := 1/6$ et $\delta := (4N |\bigcup_{I \in \mathcal{I}} D_I|)^{-1}$. Notre hypothèse sur n et la borne grossière

$$\left| \bigcup_{I \in \mathcal{I}} D_I \right| \leq \binom{N}{k'} (3/\rho)^{k'} \leq \left(\frac{3eN}{k'\rho} \right)^{k'} \leq \left(\frac{N}{\rho} \right)^{k'}$$

impliquent que $|\log \delta| / (\epsilon^2 - \frac{1}{3}\epsilon^3) \leq 38(k' + 2) \log N \leq 500k \log N \leq n$. Il suit donc du Lemme 5.4 que

$$\mathbb{P}\left(\sup\left\{\Lambda(\mathbf{z}) \mid \mathbf{z} \in \bigcup_{I \in \mathcal{I}} D_I\right\} \leq 1/6\right) \geq 1 - \frac{1}{2N}. \quad (5.5)$$

2. Le volume de la boule de rayon R dans \mathbb{R}^d est $c_d R^d$ avec $c_d := \pi^{d/2} R^d / \Gamma(\frac{d}{2} + 1)$. Notons toutefois que la valeur de la constante c_d ne joue aucun rôle dans l'argument.

Observons à présent que (5.4) suivrait de (5.5), si nous étions en mesure de montrer que

$$\mathbb{P}(\forall \mathbf{y}, \mathbf{z} \in \mathbb{R}^N, |\Lambda(\mathbf{z}) - \Lambda(\mathbf{y})| \leq \sqrt{6N \log N} \|\mathbf{z} - \mathbf{y}\|_2) \geq 1 - \frac{1}{2N}. \quad (5.6)$$

En effet, la construction des ensembles D_I garantit que pour tout $\mathbf{y} \in \mathcal{S}_I$, il existe $\mathbf{x} \in D_I$ tel que $\|\mathbf{x} - \mathbf{y}\|_2 \leq \rho$. Les événements dans (5.5) et (5.6) se produisent avec probabilité au moins $1 - 1/N$ et lorsque c'est le cas,

$$\Lambda(\mathbf{y}) \leq \Lambda(\mathbf{x}) + |\Lambda(\mathbf{y}) - \Lambda(\mathbf{x})| \leq \frac{1}{6} + \sqrt{6N \log N} \rho = \frac{1}{3}.$$

Il nous suffit donc d'établir (5.6). Soit $\mathbf{y}, \mathbf{z} \in \mathbb{R}^N$. Tout d'abord, deux applications de l'inégalité du triangle permettent d'écrire

$$|\Lambda(\mathbf{z}) - \Lambda(\mathbf{y})| \leq \left| \|L\mathbf{z}\|_2 - \|L\mathbf{y}\|_2 \right| \leq \|L(\mathbf{z} - \mathbf{y})\|_2.$$

Notons $\|A\|_\infty := \max_{i,j} |A_{ij}|$. Par l'inégalité de Cauchy-Schwarz,

$$\|L(\mathbf{z} - \mathbf{y})\|_2^2 = \sum_{i=1}^n \left(\sum_{j=1}^N L_{ij}(\mathbf{z} - \mathbf{y})_j \right)^2 \leq \sum_{i=1}^n \left(\sum_{j=1}^N L_{ij}^2 \right) \left(\sum_{j=1}^N (\mathbf{z} - \mathbf{y})_j^2 \right) \leq N \|A\|_\infty^2 \|\mathbf{z} - \mathbf{y}\|_2^2.$$

Il nous reste à contrôler les valeurs typiques de $\|A\|_\infty$. Par une inégalité classique sur les queues de gaussiennes³ et l'inégalité de Boole,

$$\mathbb{P}(\|A\|_\infty \geq \sqrt{6 \log N}) = \mathbb{P}(\exists i, j \text{ t.q. } |A_{ij}| \geq \sqrt{6 \log N}) \leq Nn \sqrt{\frac{2}{6\pi \log N}} N^{-3} \leq \frac{1}{2N}.$$

Ainsi, avec probabilité au moins $1 - \frac{1}{2N}$, on a $\|A\|_\infty < \sqrt{6 \log N}$, ce qui garantit que $|\Lambda(\mathbf{z}) - \Lambda(\mathbf{y})| \leq \sqrt{6N \log N} \|\mathbf{z} - \mathbf{y}\|_2$. \square

Remarques bibliographiques : Ce chapitre est basé sur [22].

3. Si $Z \sim \mathcal{N}(0, 1)$ et $a > 0$, alors $\mathbb{P}(|Z| \geq a) = \frac{2}{\sqrt{2\pi}} \int_a^\infty e^{-x^2/2} dx \leq \sqrt{\frac{2}{\pi}} \int_a^\infty \frac{x}{a} e^{-x^2/2} dx = \sqrt{\frac{2}{\pi}} \frac{1}{a} e^{-a^2/2}$.

6 Transformée de Doob et marche aléatoire sur \mathbb{Z}

6.1 Transformée de Doob d'une chaîne de Markov

Soit $(X_n)_{n \geq 0}$ une chaîne de Markov sur un espace d'états S dénombrable ou fini, avec matrice de transition $P = (p(i, j))_{i, j \in S}$. On notera \mathbb{P}_i la loi de la chaîne démarrée en $i \in S$ au temps 0. Soit $A \subset S$ non vide et soit $h : S \rightarrow \mathbb{R}_+$ une fonction strictement positive sur A et P -harmonique sur A :

$$\forall i \in A, \quad h(i) > 0 \quad \text{et} \quad h(i) = \sum_{j \in S} p(i, j)h(j).$$

On définit alors

$$\forall i \in A, \forall j \in S, \quad q(i, j) := \frac{h(j)}{h(i)}p(i, j).$$

Il suit de la P -harmonicité de h que, pour tout $i \in A$,

$$\sum_{j \in S} q(i, j) = \frac{1}{h(i)} \sum_{j \in S} h(j)p(i, j) = \frac{1}{h(i)}h(i) = 1.$$

Les $q(i, j)$ peuvent donc être interprétés comme les probabilités de transition d'une chaîne de Markov.

Définition 6.1. Soit $i_0 \in A$. La chaîne de Markov $(Y_n)_{n \geq 0}$ avec probabilités de transition $q(i, j)$, partant de $Y_0 = i_0$ et stoppée lorsqu'elle quitte A est appelée la **h -transformée de Doob**¹ de la chaîne (X_n) . On notera \mathbb{Q}_{i_0} la loi de la chaîne (Y_n) partant de i_0 .

Observons que la probabilité d'une trajectoire finie $i_0, i_1, \dots, i_{n-1}, i_n$, avec $i_0, \dots, i_{n-1} \in A$ et $i_n \in S$, est donnée par

$$\begin{aligned} \mathbb{Q}_{i_0}(Y_0 = i_0, \dots, Y_n = i_n) &= \prod_{k=1}^n q(i_{k-1}, i_k) \\ &= \frac{h(i_n)}{h(i_0)} \prod_{k=1}^n p(i_{k-1}, i_k) = \frac{h(i_n)}{h(i_0)} \mathbb{P}_{i_0}(X_0 = i_0, \dots, X_n = i_n). \end{aligned} \quad (6.1)$$

La construction ci-dessus peut sembler artificielle. Pourtant, comme nous allons le voir à présent sur quelques exemples, le processus (Y_n) possède souvent une interprétation très naturelle.

1. En anglais : *Doob h -transform*.

6.2 Quelques exemples

6.2.1 Processus conditionné à quitter A par un certain état

Soit $A \subset S$ non vide. Considérons la fonction $h : S \rightarrow \mathbb{R}_+$ définie par

$$\forall i \in S, \quad h(i) := \mathbb{P}_i(\tau_{S \setminus A} < \infty, X_{\tau_{S \setminus A}} = a),$$

où $a \in S \setminus A$ et $\tau_B := \min\{n \geq 0 \mid X_n \in B\}$. On supposera que, pour tout $i \in A$, la chaîne (X_n) peut quitter A via a avec probabilité strictement positive; ceci garantit que $h(i) > 0$ pour tout $i \in A$. La fonction h est également P -harmonique sur A : pour tout $i \in A$,

$$\begin{aligned} h(i) &= \mathbb{P}_i(\tau_{S \setminus A} < \infty, X_{\tau_{S \setminus A}} = a) \\ &= \sum_{j \in S} \mathbb{P}_i(X_1 = j) \mathbb{P}_i(\tau_{S \setminus A} < \infty, X_{\tau_{S \setminus A}} = a \mid X_1 = j) \\ &= \sum_{j \in S} p(i, j) \mathbb{P}_j(\tau_{S \setminus A} < \infty, X_{\tau_{S \setminus A}} = a) = \sum_{j \in S} p(i, j) h(j), \end{aligned}$$

la troisième identité suivant de la propriété de Markov. Nous sommes donc dans le cadre de la section précédente et pouvons considérer la h -transformée de Doob (Y_n) .

Proposition 6.2. *Soit $i_0 \in A$ et (Y_n) la h -transformée de Doob partant de i_0 . Alors, (Y_n) a la même loi que la chaîne (X_n) démarrée en i_0 , stoppée à la première visite en a et conditionnée à quitter A par l'état a en un temps fini.*

Démonstration. C'est une conséquence de (6.1) : pour toute trajectoire finie $i_0, i_1, \dots, i_{n-1}, i_n$, avec $i_0, \dots, i_{n-1} \in A$ et $i_n = a$,

$$\begin{aligned} \mathbb{Q}_{i_0}((Y_0, \dots, Y_n) = (i_0, \dots, i_n)) &= \frac{h(a)}{h(i_0)} \prod_{k=1}^n p(i_{k-1}, i_k) \\ &= \frac{\mathbb{P}_{i_0}((X_0, \dots, X_n) = (i_0, \dots, i_n))}{\mathbb{P}_{i_0}(\tau_{S \setminus A} < \infty, X_{\tau_{S \setminus A}} = a)} \\ &= \mathbb{P}_{i_0}((X_0, \dots, X_n) = (i_0, \dots, i_n) \mid \tau_{S \setminus A} < \infty, X_{\tau_{S \setminus A}} = a). \quad \square \end{aligned}$$

Insistons sur le caractère remarquable d'avoir une description *markovienne* d'un processus (X_n) conditionné sur un événement faisant intervenir le comportement futur de (X_n) . La même chose se produit dans les exemples suivants.

6.2.2 Processus conditionné à ne pas quitter A (cas transient)

Soit $A \subset S$ non vide. On suppose cette fois que, pour tout $i_0 \in A$, il y a une probabilité strictement positive que la chaîne ne quitte jamais A . On considère alors la fonction $h : S \rightarrow \mathbb{R}_+$ définie par

$$\forall i \in S, \quad h(i) := \mathbb{P}_i(\tau_{S \setminus A} = \infty).$$

Par notre hypothèse ci-dessus, $h(i) > 0$ pour tout $i \in A$. De plus, $h(i) = 0$ pour tout $i \notin A$. On vérifie aisément que la propriété de Markov implique à nouveau que h est P -harmonique sur A . Nous sommes donc dans le cadre de la Section 6.1 et pouvons considérer la h -transformée de Doob (Y_n) .

Proposition 6.3. *Soit $i_0 \in A$ et (Y_n) la h -transformée de Doob partant de i_0 . Alors, (Y_n) a la même loi que la chaîne (X_n) démarrée en i_0 et conditionnée à ne jamais quitter A .*

Démonstration. C'est à nouveau une conséquence immédiate de (6.1) : pour toute trajectoire finie i_0, i_1, \dots, i_n de probabilité strictement positive sous \mathbb{Q}_{i_0} (observez que cela implique en particulier que $i_k \in A$ pour tout $0 \leq k \leq n$),

$$\begin{aligned} \mathbb{Q}_{i_0}((Y_0, \dots, Y_n) = (i_0, \dots, i_n)) &= \frac{h(i_n)}{h(i_0)} \prod_{k=1}^n p(i_{k-1}, i_k) \\ &= \frac{\mathbb{P}_{i_0}((X_0, \dots, X_n) = (i_0, \dots, i_n)) \mathbb{P}_{i_n}(\tau_{S \setminus A} = \infty)}{\mathbb{P}_{i_0}(\tau_{S \setminus A} = \infty)} \\ &= \mathbb{P}_{i_0}((X_0, \dots, X_n) = (i_0, \dots, i_n) \mid \tau_{S \setminus A} = \infty), \end{aligned}$$

puisque, par la propriété de Markov,

$$\mathbb{P}_{i_n}(\tau_{S \setminus A} = \infty) = \mathbb{P}_{i_0}(\tau_{S \setminus A} = \infty \mid (X_0, \dots, X_n) = (i_0, \dots, i_n)). \quad \square$$

6.2.3 Processus conditionné à appartenir à B au temps N

Dans ce dernier exemple, nous allons voir que cette approche peut être étendue à des situations dans lesquelles l'horizon temporel est spécifié. Cela nécessite d'adapter quelque peu l'approche précédente. Soit $B \subset S$ non vide et $N \in \mathbb{N}^*$. On considère la fonction $h : \{0, \dots, N\} \times S \rightarrow \mathbb{R}_+$ donnée par

$$h(n, i) := \mathbb{P}_i(X_{N-n} \in B).$$

Observons que cette fonction satisfait, pour tout $n \in \{0, \dots, N-1\}$ et tout $i \in S$,

$$h(n, i) = \mathbb{P}_i(X_{N-n} \in B) = \sum_{j \in S} p(i, j) \mathbb{P}_j(X_{N-n-1} \in B) = \sum_{j \in S} p(i, j) h(n+1, j).$$

Définissons

$$q_n(i, j) := \frac{h(n+1, j)}{h(n, i)} p(i, j)$$

lorsque $h(n, i) \neq 0$ et $q_n(i, j) := 0$ sinon. On a alors, pour tout $i \in S$ tel que $h(n, i) > 0$,

$$\sum_{j \in S} q_n(i, j) = \frac{1}{h(n, i)} \sum_{j \in S} h(n+1, j) p(i, j) = \frac{1}{h(n, i)} h(n, i) = 1.$$

On peut donc considérer la chaîne de Markov *inhomogène* $(Y_n)_{n \in \mathbb{N}}$ avec probabilités de transition

$$\mathbb{Q}_{i_0}(Y_{n+1} = j \mid Y_n = i) = q_n(i, j)$$

démarrée en i_0 et stoppée au temps N .

Proposition 6.4. *Soit $N \in \mathbb{N}^*$, $B \subset S$ non vide et $i_0 \in S$ tel que $\mathbb{P}_{i_0}(X_N \in B) > 0$. Soit (Y_n) la chaîne de Markov définie ci-dessus, partant de i_0 . Alors, (Y_n) a la même loi que la chaîne (X_n) démarrée en i_0 et conditionnée à visiter B au temps N .*

Démonstration. Soit $i_0, i_1, \dots, i_N \in S$ avec $i_N \in B$ et $h(n, i_n) > 0$ pour tout $0 \leq n \leq N-1$ (observons que la probabilité de toute trajectoire de (Y_n) ne satisfaisant pas cette dernière condition est nécessairement nulle). Alors, on a bien

$$\begin{aligned} \mathbb{Q}_{i_0}(Y_0 = i_0, \dots, Y_N = i_N) &= \prod_{n=0}^{N-1} \frac{h(n+1, i_{n+1})}{h(n, i_n)} p(i_n, i_{n+1}) \\ &= \frac{h(N, i_N)}{h(0, i_0)} \prod_{n=0}^{N-1} p(i_n, i_{n+1}) \\ &= \frac{\mathbb{P}_{i_N}(X_0 \in B)}{\mathbb{P}_{i_0}(X_N \in B)} \mathbb{P}_{i_0}(X_0 = i_0, \dots, X_N = i_N) \\ &= \mathbb{P}_{i_0}(X_0 = i_0, \dots, X_N = i_N \mid X_N \in B). \quad \square \end{aligned}$$

6.3 Marche aléatoire sur \mathbb{Z} conditionnée à rester positive

Dans les exemples précédents, le conditionnement était toujours par rapport à un événement de probabilité strictement positive. Nous allons à présent nous intéresser au cas d'un conditionnement par un événement de probabilité nulle. Un exemple typique serait celui d'une chaîne de Markov récurrente que l'on conditionne à éviter (éternellement) un sous-ensemble non vide de S . Il s'agit d'un problème nettement plus subtil et, plutôt que de discuter de situations générales, nous ne considérerons qu'une classe de marches aléatoires sur \mathbb{Z} .

Soit μ une mesure de probabilité sur $\{-1, 0, 1\}$ telle que $\mu(0) < 1$. Soit $(S_n)_{n \geq 0}$ la marche aléatoire sur \mathbb{Z} dont les accroissements sont i.i.d. et de loi μ . On note \mathbb{P}_i la loi de (S_n) partant de i au temps 0.

6.3.1 Deux fonctions (super)harmoniques

On considère les fonctions suivantes : pour tout $i \in \mathbb{N}$,

$$h^\downarrow(i) := \mathbb{P}_0(\tau_{-i} < \infty), \quad h^\uparrow(i) := h^\downarrow(0) + \dots + h^\downarrow(i).$$

Notons $\alpha := \mathbb{P}_0(\tau_{-1} < \infty)$.

Lemme 6.5. Pour tout $i \in \mathbb{N}$, $h^\downarrow(i) = \alpha^i$ et $h^\uparrow(i) = \sum_{k=0}^i \alpha^k$. De plus,

$$\alpha = \begin{cases} 1 & \text{si } \mu(-1) \geq \mu(1), \\ \mu(-1)/\mu(1) & \text{si } \mu(-1) < \mu(1). \end{cases}$$

Démonstration. Clairement, $h^\downarrow(0) = 1$. Il suit de la propriété de Markov forte que, pour tout $i \geq 1$,

$$\begin{aligned} h^\downarrow(i) &= \mathbb{P}_0(\tau_{-i} < \infty) = \mathbb{P}_0(\tau_{-i} < \infty \mid \tau_{-i+1} < \infty) \mathbb{P}_0(\tau_{-i+1} < \infty) \\ &= \mathbb{P}_0(\tau_{-1} < \infty) \mathbb{P}_0(\tau_{-i+1} < \infty) = \alpha h^\downarrow(i-1). \end{aligned}$$

On en déduit que $h^\downarrow(i) = \alpha^i$, pour tout $i \in \mathbb{N}^*$ et, par conséquent, $h^\uparrow(i) = 1 + \alpha + \dots + \alpha^i$. Déterminons α . En conditionnant sur le premier pas de la marche et en utilisant la propriété de Markov, on obtient

$$\begin{aligned} \alpha &= \mathbb{P}_0(\tau_{-1} < \infty) = \mu(-1) \mathbb{P}_{-1}(\tau_{-1} < \infty) + \mu(0) \mathbb{P}_0(\tau_{-1} < \infty) + \mu(1) \mathbb{P}_1(\tau_{-1} < \infty) \\ &= \mu(-1) + \mu(0)\alpha + \mu(1)\alpha^2. \end{aligned}$$

Manifestement, $\alpha = 1$ est la seule solution dans $[0, 1]$ lorsque $\mu(-1) \geq \mu(1)$.

Supposons donc que $\mu(-1) < \mu(1)$. Dans ce cas, l'équation pour α possède deux solutions dans $[0, 1]$: celle donnée dans l'énoncé et $\alpha = 1$. Nous allons exclure cette dernière. Par la loi forte des grands nombres, $\lim_{n \rightarrow \infty} S_n/n = \mu(1) - \mu(-1) =: c > 0$ presque sûrement. En particulier,

$$\lim_{N \rightarrow \infty} \mathbb{P}_0(\forall n \geq N, S_n > \frac{1}{2}cn) = \mathbb{P}_0(\exists N \in \mathbb{N}, \forall n \geq N, S_n/n > c/2) = 1.$$

Il existe donc $N \in \mathbb{N}^*$ tel que $\mathbb{P}_0(\forall n \geq N, S_n \geq 0) \geq 1/2$. Étant donné que l'on a également $\mathbb{P}_0(\forall n \in \llbracket N-1 \rrbracket, S_n \geq 0) > 0$, on a bien $\alpha < 1$. \square

Il suit immédiatement de la loi forte des grands nombres que le cas $\mu(1) > \mu(-1)$ correspond au cas où $\lim_{n \rightarrow \infty} S_n = +\infty$ presque sûrement ; on dit que la marche (S_n) **dérive vers** $+\infty$. Symétriquement, le cas $\mu(1) < \mu(-1)$ correspond au cas où $\lim_{n \rightarrow \infty} S_n = -\infty$ presque sûrement ; on dit que la marche (S_n) **dérive vers** $-\infty$. Finalement, le cas $\mu(-1) = \mu(1)$ correspond au cas récurrent, dans lequel la marche **oscille** : $\liminf_{n \rightarrow \infty} S_n = -\infty$ et $\limsup_{n \rightarrow \infty} S_n = +\infty$ presque sûrement.

Pour le prochain résultat, on étend h^\uparrow de façon naturelle en posant $h^\uparrow(-1) := 0$.

Proposition 6.6. h^\downarrow est μ -harmonique sur \mathbb{N}^* et h^\uparrow est μ -superharmonique sur \mathbb{N} :

$$\begin{aligned}\forall i \in \mathbb{N}^*, \quad h^\downarrow(i) &= \sum_{j \in \{-1,0,1\}} \mu(j) h^\downarrow(i+j), \\ \forall i \in \mathbb{N}, \quad h^\uparrow(i) &\geq \sum_{j \in \{-1,0,1\}} \mu(j) h^\uparrow(i+j).\end{aligned}$$

De plus, h^\uparrow est μ -harmonique sur \mathbb{N} si et seulement si $\mu(-1) \leq \mu(1)$.

Démonstration. La μ -harmonicité de h^\downarrow suit immédiatement du Lemme 6.5, mais peut être déduite de façon plus transparente de la propriété de Markov : pour tout $i \in \mathbb{N}^*$,

$$\begin{aligned}h^\downarrow(i) &= \mathbb{P}_0(\tau_{-i} < \infty) = \sum_{k \in \{-1,0,1\}} \mu(k) \mathbb{P}_0(\tau_{-i} < \infty \mid S_1 = k) \\ &= \sum_{k \in \{-1,0,1\}} \mu(k) \mathbb{P}_0(\tau_{-i-k} < \infty) = \sum_{k \in \{-1,0,1\}} \mu(k) h^\downarrow(i+k).\end{aligned}$$

Passons à h^\uparrow . Pour tout $i \in \mathbb{N}^*$,

$$\begin{aligned}h^\uparrow(i) &= 1 + \sum_{j=1}^i h^\downarrow(j) = 1 + \sum_{j=1}^i \sum_{k \in \{-1,0,1\}} \mu(k) h^\downarrow(j+k) \\ &= 1 + \sum_{k \in \{-1,0,1\}} \mu(k) \sum_{j=1+k}^{i+k} h^\downarrow(j) \\ &= \mu(-1) - \alpha\mu(1) + \sum_{k \in \{-1,0,1\}} \mu(k) h^\uparrow(i+k).\end{aligned}$$

Étant donné que $\mu(-1) \geq \alpha\mu(1)$, il suit que h^\uparrow est μ -superharmonique sur \mathbb{N}^* et qu'elle est μ -harmonique sur \mathbb{N}^* précisément lorsque $\mu(-1) = \alpha\mu(1)$, ce qui se produit si et seulement si $\mu(-1) \leq \mu(1)$. L'extension de la μ -(super)harmonicité de h^\uparrow à \mathbb{N} est immédiate :

$$\mu(-1)h^\uparrow(-1) + \mu(0)h^\uparrow(0) + \mu(1)h^\uparrow(1) = \mu(0) + (1 + \alpha)\mu(1) \leq 1 = h^\uparrow(0),$$

puisque $\alpha\mu(1) \leq \mu(-1)$, ce qui établit la μ -superharmonicité. À nouveau, le cas d'égalité correspond à $\alpha\mu(1) = \mu(-1)$. \square

La Proposition 6.6 nous fournit deux fonctions harmoniques sur \mathbb{N} ou \mathbb{N}^* (sous des conditions adéquates). Nous pouvons donc essayer d'implémenter l'approche de la Section 6.1.

6.3.2 La h^\downarrow -transformée de Doob lorsque $\mu(-1) > 0$

h^\downarrow étant μ -harmonique sur \mathbb{N}^* , on souhaite considérer la h^\downarrow -transformée de Doob de (S_n) , partant de $i_0 \in \mathbb{N}^*$. Observons que, pour tout $i \in \mathbb{N}$, $h^\downarrow(i) > 0$ lorsque $\mu(-1) > 0$ et

$$h^\downarrow(i) = \mathbb{P}_0(\tau_{-i} < \infty) = \mathbb{P}_i(\tau_{\mathbb{Z} \setminus \mathbb{N}^*} < \infty, S_{\tau_{\mathbb{Z} \setminus \mathbb{N}^*}} = 0).$$

Par conséquent, on se trouve précisément dans le cadre de la Section 6.2.1, avec $X_n := S_n$, $i_0 \in \mathbb{N}^*$, $A := \mathbb{N}^*$ et $a := 0$. Une application de la Proposition 6.2 montre ainsi que la h^\downarrow -transformée de Doob de (S_n) coïncide avec la marche aléatoire (S_n) partant de $i_0 \in \mathbb{N}^*$ et conditionnée à visiter 0.

Évidemment, il n'y a pas de difficultés particulières ici, le conditionnement étant par rapport à un événement de probabilité positive.

6.3.3 La h^\uparrow -transformée de Doob lorsque $\mu(1) \geq \mu(-1)$

Sous l'hypothèse que $\mu(1) \geq \mu(-1)$, c'est-à-dire lorsque la marche aléatoire (S_n) ne dérive pas vers $-\infty$, la fonction h^\uparrow est μ -harmonique et strictement positive sur \mathbb{N} . On peut donc considérer la h^\uparrow -transformée de Doob de (S_n) démarrée en $i_0 \in \mathbb{N}$. Notons (S_n^\uparrow) la h^\uparrow -transformée de Doob et \mathbb{P}_i^\uparrow la loi de cette chaîne de Markov partant de $i \in \mathbb{N}$. Le premier résultat est que (S_n^\uparrow) est transiente.

Proposition 6.7. *On suppose que $\mu(1) \geq \mu(-1)$. Alors, la chaîne de Markov (S_n^\uparrow) est transiente.*

Démonstration. Pour tout $i \in \mathbb{N}$,

$$\mathbb{P}_i^\uparrow(\tau_{i-1} < \infty) = \sum_{n \geq 0} \mathbb{P}_i^\uparrow(\tau_{i-1} = n) = \frac{h^\uparrow(i-1)}{h^\uparrow(i)} \sum_{n \geq 0} \mathbb{P}_i(\tau_{i-1} = n) = \frac{h^\uparrow(i-1)}{h^\uparrow(i)} \mathbb{P}_i(\tau_{i-1} < \infty),$$

la seconde identité suivant de (6.1).

Lorsque $\mu(-1) = 0$, on a $\mathbb{P}_i(\tau_{i-1} < \infty) = 0$ et donc $\mathbb{P}_i^\uparrow(\tau_{i-1} < \infty) = 0$.

Lorsque $\mu(-1) > 0$, on a $h^\uparrow(i) = h^\uparrow(i-1) + h^\downarrow(i) > h^\uparrow(i-1)$ et donc $\mathbb{P}_i^\uparrow(\tau_{i-1} < \infty) < 1$. \square

Cas transient : $\mu(1) > \mu(-1)$

La proposition suivante montre que, lorsque $\mu(1) > \mu(-1)$, on se trouve dans le cadre de la Section 6.2.2 : la fonction h^\uparrow ne diffère de la fonction harmonique de la Section 6.2.2 que par une constante multiplicative strictement positive. Notons que l'hypothèse $\mu(1) > \mu(-1)$ garantit que $\mathbb{P}_0(\tau_{\mathbb{Z}_{<0}} = \infty) = 1 - \alpha > 0$.

Proposition 6.8. *Lorsque $\mu(1) > \mu(-1)$, on a*

$$\forall i \in \mathbb{N}, \quad h^\uparrow(i) = \frac{\mathbb{P}_i(\tau_{\mathbb{Z}_{<0}} = \infty)}{\mathbb{P}_0(\tau_{\mathbb{Z}_{<0}} = \infty)}.$$

Démonstration. Pour tout $i \geq 0$,

$$\begin{aligned} \mathbb{P}_i(\tau_{\mathbb{Z}_{<0}} = \infty) &= \mathbb{P}_i(\min\{S_n \mid n \geq 0\} \geq 0) \\ &= \sum_{k=0}^i \mathbb{P}_i(\min\{S_n \mid n \geq 0\} = i - k) \\ &= \sum_{k=0}^i \mathbb{P}_i(\tau_{i-k} < \infty, \tau_{i-k-1} = \infty) \\ &= \sum_{k=0}^i \mathbb{P}_i(\tau_{i-k} < \infty) \mathbb{P}_{i-k}(\tau_{i-k-1} = \infty), \end{aligned}$$

par la propriété de Markov forte. On a évidemment $\mathbb{P}_i(\tau_{i-k} < \infty) = \mathbb{P}_0(\tau_{-k} < \infty) = h^\downarrow(k)$ et $\mathbb{P}_{i-k}(\tau_{i-k-1} = \infty) = \mathbb{P}_0(\tau_{\mathbb{Z}_{<0}} = \infty)$. Par conséquent,

$$\mathbb{P}_i(\tau_{\mathbb{Z}_{<0}} = \infty) = \mathbb{P}_0(\tau_{\mathbb{Z}_{<0}} = \infty) \sum_{k=0}^i h^\downarrow(k) = \mathbb{P}_0(\tau_{\mathbb{Z}_{<0}} = \infty) h^\uparrow(i). \quad \square$$

Une application de la Proposition 6.3 montre que la loi de (S_n^\uparrow) démarrée en $i_0 \in \mathbb{N}$, coïncide, lorsque $\mu(1) > \mu(-1)$, avec celle de (S_n) démarrée en i_0 et conditionnée à ne jamais visiter $\mathbb{Z}_{<0}$.

Cas récurrent : $\mu(1) = \mu(-1)$

La Proposition 6.6 montre que la fonction h^\uparrow demeure μ -harmonique et strictement positive sur \mathbb{N} lorsque $\mu(1) = \mu(-1)$. Dans ce cas, la marche aléatoire (S_n) est récurrente et la probabilité de ne jamais visiter $\mathbb{Z}_{<0}$ est donc nulle. Conditionner (S_n) par cet événement est donc mal défini. Ce que l'on peut faire, par contre, c'est considérer le processus (S_n) conditionné par l'événement

$$\Lambda_N := \{S_n \geq 0 \text{ pour tout } 0 \leq n \leq N\},$$

ce dernier ayant probabilité strictement positive. Commençons par montrer un lemme technique, qui est un pendant de la Proposition 6.8 (la preuve est d'ailleurs presque identique).

Lemme 6.9. Lorsque $\mu(1) = \mu(-1)$, on a

$$\forall i \in \mathbb{N}, \quad h^\uparrow(i) \leq \liminf_{N \rightarrow \infty} \frac{\mathbb{P}_i(\Lambda_N)}{\mathbb{P}_0(\Lambda_N)}.$$

Démonstration. Pour tout $i \geq 0$,

$$\begin{aligned} \mathbb{P}_i(\Lambda_N) &= \mathbb{P}_i(\min\{S_n \mid 0 \leq n \leq N\} \geq 0) \\ &= \sum_{k=0}^i \mathbb{P}_i(\min\{S_n \mid 0 \leq n \leq N\} = i - k) \\ &= \sum_{k=0}^i \mathbb{P}_i(\tau_{i-k} \leq N \text{ et } \forall n \in \{\tau_{i-k}, \dots, N\}, S_n \geq i - k) \\ &\geq \sum_{k=0}^i \mathbb{P}_i(\tau_{i-k} \leq N) \mathbb{P}_{i-k}(\forall n \in \{0, \dots, N\}, S_n \geq i - k) \\ &= \mathbb{P}_0(\forall n \in \{0, \dots, N\}, S_n \geq 0) \sum_{k=0}^i \mathbb{P}_0(\tau_{-k} \leq N). \end{aligned}$$

Par conséquent,

$$\liminf_{N \rightarrow \infty} \frac{\mathbb{P}_i(\Lambda_N)}{\mathbb{P}_0(\Lambda_N)} \geq \lim_{N \rightarrow \infty} \sum_{k=0}^i \mathbb{P}_0(\tau_{-k} \leq N) = \sum_{k=0}^i \mathbb{P}_0(\tau_{-k} < \infty) = h^\uparrow(i). \quad \square$$

Le résultat suivant montre que la limite en loi des processus conditionnés lorsque $N \rightarrow \infty$ est bien donnée par le processus (S_n^\uparrow) .

Proposition 6.10. On suppose que $\mu(-1) = \mu(1)$. Alors, pour tout $i_0 = 0, i_1, \dots, i_n \geq 0$,

$$\lim_{N \rightarrow \infty} \mathbb{P}_0(S_0 = i_0, \dots, S_n = i_n \mid \Lambda_N) = \mathbb{P}_0^\uparrow(S_0^\uparrow = i_0, \dots, S_n^\uparrow = i_n).$$

Démonstration. Pour tout $i_0 = 0, i_1, \dots, i_n \in \mathbb{N}$ et tout $N \geq n$, il suit de la propriété de Markov et de (6.1) que

$$\begin{aligned} \mathbb{P}_0(S_0 = i_0, \dots, S_n = i_n \mid \Lambda_N) &= \frac{\mathbb{P}_0(S_0 = i_0, \dots, S_n = i_n) \mathbb{P}_{i_n}(\Lambda_{N-n})}{\mathbb{P}_0(\Lambda_N)} \\ &\geq \frac{\mathbb{P}_{i_n}(\Lambda_N)}{\mathbb{P}_0(\Lambda_N)} \frac{h^\uparrow(0)}{h^\uparrow(i_n)} \mathbb{P}_0^\uparrow(S_0^\uparrow = i_0, \dots, S_n^\uparrow = i_n). \end{aligned}$$

Le Lemme 6.5 implique donc que

$$\liminf_{N \rightarrow \infty} \mathbb{P}_0(S_0 = i_0, \dots, S_n = i_n | \Lambda_N) \geq \mathbb{P}_0^\uparrow(S_0^\uparrow = i_0, \dots, S_n^\uparrow = i_n). \quad (6.2)$$

Supposons, par l'absurde, qu'il existe $j_0, \dots, j_n \in \mathbb{N}$ et $c > 0$ tel que

$$\limsup_{N \rightarrow \infty} \mathbb{P}_0(S_0 = j_0, \dots, S_n = j_n | \Lambda_N) = \mathbb{P}_0^\uparrow(S_0^\uparrow = j_0, \dots, S_n^\uparrow = j_n) + c. \quad (6.3)$$

Par (6.2), il existe $N_0 \in \mathbb{N}$ tel que $\forall N \geq N_0$,

$$\forall i_1, \dots, i_n \in \mathbb{N}, \quad \mathbb{P}_0(S_0 = i_0, \dots, S_n = i_n | \Lambda_N) \geq \mathbb{P}_0^\uparrow(S_0^\uparrow = i_0, \dots, S_n^\uparrow = i_n) - 3^{-n-1}c.$$

Par (6.3), il existe $N \geq N_0$ tel que

$$\mathbb{P}_0(S_0 = j_0, \dots, S_n = j_n | \Lambda_N) \geq \mathbb{P}_0^\uparrow(S_0^\uparrow = j_0, \dots, S_n^\uparrow = j_n) + \frac{c}{2}.$$

On obtient donc

$$\begin{aligned} 1 &= \sum_{i_1, \dots, i_n \in \mathbb{N}} \mathbb{P}_0(S_0 = 0, S_1 = i_1, \dots, S_n = i_n | \Lambda_N) \\ &\geq \frac{c}{2} - \frac{c}{3} + \sum_{i_1, \dots, i_n \in \mathbb{N}} \mathbb{P}_0^\uparrow(S_0^\uparrow = 0, S_1^\uparrow = i_1, \dots, S_n^\uparrow = i_n) = 1 + \frac{c}{6}, \end{aligned}$$

la dernière identité suivant du fait que $h^\uparrow(-1) = 0$, puisqu'il suit que toute trajectoire visitant -1 a probabilité nulle sous \mathbb{P}_0^\uparrow . La contradiction conclut la preuve. \square

Remarques bibliographiques : Ce chapitre s'inspire largement du chapitre 5 des notes de cours [7]. Pour le lecteur intéressé, ces dernières traitent de marches aléatoires générales sur \mathbb{Z} , pas uniquement de celles dont les incréments prennent valeur dans $\{-1, 0, 1\}$.

7 Convergence des chaînes de Markov réversibles

Dans ce chapitre, nous allons étudier la vitesse de convergence vers l'équilibre des chaînes de Markov réversibles sur un espace d'états fini à l'aide de méthodes spectrales.

Cadre mathématique dans ce chapitre. Dans ce chapitre, nous considérerons le cadre suivant : $(X_n)_{n \geq 0}$ est une chaîne de Markov irréductible et réversible sur un espace d'états S fini. Nous noterons $P = (p_{ij})_{i,j \in S}$ sa matrice de transition et $\pi = (\pi(i))_{i \in S}$ son unique loi stationnaire. Nous utiliserons également la notation $p_{ij}(n) := (P^n)_{ij} = \mathbb{P}(X_n = j \mid X_0 = i)$.

7.1 Structure algébrique

7.1.1 Une caractérisation de la réversibilité

Soit $\ell^2(\pi)$ l'espace vectoriel \mathbb{R}^S muni du produit scalaire

$$\langle f, g \rangle_\pi := \sum_{i \in S} f(i)g(i)\pi(i).$$

La norme associée sera notée $\|f\|_\pi := \sqrt{\langle f, f \rangle_\pi}$. Finalement, nous écrivons

$$\langle f \rangle_\pi := \sum_{i \in S} f(i)\pi(i) = \langle f, \mathbf{1} \rangle_\pi \tag{7.1}$$

pour l'espérance de f sous π (nous avons noté $\mathbf{1}$ la fonction $\mathbf{1}(i) := 1$ pour tout $i \in S$).

Naturellement, la matrice de transition P peut être interprétée comme un opérateur dans $\ell^2(\pi)$ agissant sur $f \in \ell^2(\pi)$ par $(Pf)(i) := \sum_{j \in S} p_{ij}f(j)$.

Le théorème suivant fournit une caractérisation des chaînes de Markov réversibles.

Théorème 7.1. Une chaîne de Markov $(X_n)_{n \geq 0}$ sur S est réversible par rapport à sa loi stationnaire π si et seulement si sa matrice de transition P est un opérateur autoadjoint dans $\ell^2(\pi)$:

$$\forall f, g \in \ell^2(\pi), \quad \langle Pf, g \rangle_\pi = \langle f, Pg \rangle_\pi.$$

Démonstration. \triangleright Supposons tout d'abord que, $\pi(i)p_{ij} = \pi(j)p_{ji}$ pour tout $i, j \in S$. Alors, pour tout $f, g \in \ell^2(\pi)$,

$$\begin{aligned} \langle Pf, g \rangle_\pi &= \sum_{i \in S} \left[\sum_{j \in S} p_{ij} f(j) \right] g(i) \pi(i) \\ &= \sum_{i \in S} \sum_{j \in S} p_{ji} f(j) g(i) \pi(j) \\ &= \sum_{j \in S} f(j) \left[\sum_{i \in S} p_{ji} g(i) \right] \pi(j) = \langle f, Pg \rangle_\pi. \end{aligned}$$

\triangleright Supposons à présent que P soit autoadjoint dans $\ell^2(\pi)$. Alors, pour tout $i, j \in S$,

$$\pi(j)p_{ji} = \sum_{r \in S} \left[\sum_{s \in S} p_{rs} \delta_i(s) \right] \delta_j(r) \pi(r) = \langle P\delta_i, \delta_j \rangle_\pi = \langle \delta_i, P\delta_j \rangle_\pi = \pi(i)p_{ij},$$

où l'on a noté $\delta_u, u \in S$, la fonction définie par $\delta_u(v) := \delta_{u,v}$ pour tout $v \in S$. \square

7.1.2 Valeurs propres et vecteurs propres de P

P étant autoadjointe, elle possède $|S|$ valeurs propres réelles, que l'on ordonne en ordre décroissant :

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{|S|}.$$

On vérifie facilement que $\lambda_1 = 1$ et $\lambda_{|S|} \geq -1$. En effet, $(P\mathbf{1})(i) = \sum_{j \in S} p_{ij} \mathbf{1}(j) = \sum_{j \in S} p_{ij} = 1$ et donc 1 est une valeur propre de P . Le fait que ce soit la plus grande en valeur absolue suit de l'observation que si $Pv = \lambda v$ et $i \in S$ est tel que $|v(i)| = \max_{j \in S} |v(j)|$, alors $|\lambda||v(i)| = |\lambda v(i)| = |\sum_{j \in S} p_{ij} v(j)| \leq |v(i)|$. Deux quantités vont jouer un rôle crucial dans la suite.

Définition 7.2. Le **trou spectral** est défini par $\gamma := 1 - \lambda_2$. Le **trou spectral absolu** est défini par $\gamma_* := \min\{\gamma, 1 - |\lambda_{|S|}|\}$.

Théorème 7.3. Soit P une matrice stochastique irréductible, apériodique et réversible. Alors, $\gamma \geq \gamma_* > 0$.

Démonstration. L'inégalité $\gamma \geq \gamma_*$ est triviale. Montrons que $\gamma_* > 0$.

Soit w un vecteur propre de P non colinéaire à $\mathbf{1}$ et de valeur propre λ . Comme cela a été vu dans le cours d'introduction à la théorie des probabilités, l'irréductibilité et l'apériodicité de P garantissent l'existence de $n \in \mathbb{N}$ tel que $p_{ij}(2n) > 0$ pour tout $i, j \in S$. Il y a deux cas possibles :

Cas 1 : $|w(i)| = c$ pour tout $i \in S$. On peut alors supposer, sans perte de généralité, que $w(1) = c$. On a donc

$$0 > \sum_{i \in S} p_{1i}(2n)(w(i) - c) = \lambda^{2n} w(1) - c = (\lambda^{2n} - 1)c,$$

ce qui montre que $\lambda^{2n} < 1$ et donc $|\lambda| < 1$.

Cas 2 : il existe $i, j \in S$ tel que $\max_{k \in S} |w(k)| = |w(i)| > |w(j)|$. On a alors

$$|\lambda^{2n} w(i)| = |(P^{2n} w)(i)| = \left| \sum_{k \in S} p_{ik}(2n) w(k) \right| \leq \sum_{k \in S} p_{ik}(2n) |w(k)| < |w(i)| \sum_{k \in S} p_{ik}(2n) = |w(i)|,$$

d'où l'on conclut à nouveau que $|\lambda^{2n}| < 1$, et donc que $|\lambda| < 1$. \square

P étant autoadjointe, on sait du cours d'algèbre linéaire que l'on peut trouver une base ortho-normale (par rapport au produit scalaire $\langle \cdot, \cdot \rangle_\pi$) composée de vecteurs propres $(v_k)_{k=1, \dots, |S|}$, avec $v_1 := \mathbf{1}$ et, pour chaque $k \in \{1, \dots, |S|\}$, $Pv_k = \lambda_k v_k$. (Notons que v_1 est bien normalisé puisque $\|v_1\|_\pi = \sum_{i=1}^{|S|} \pi(i) = 1$.)

On peut alors décomposer toute fonction $f \in \ell^2(\pi)$ dans cette base :

$$f = \sum_{k=1}^{|S|} \langle f, v_k \rangle_\pi v_k.$$

Observons que, comme on l'a vu dans (7.1), $\langle f, v_1 \rangle_\pi = \langle f \rangle_\pi$. En particulier,

$$f - \langle f \rangle_\pi \mathbf{1} = \sum_{k=2}^{|S|} \langle f, v_k \rangle_\pi v_k, \quad (7.2)$$

et donc

$$\begin{aligned} \text{Var}_\pi(f) &= \sum_{i=1}^{|S|} (f(i) - \langle f \rangle_\pi)^2 \pi(i) = \sum_{i=1}^{|S|} \left(\sum_{k=2}^{|S|} \langle f, v_k \rangle_\pi v_k(i) \right)^2 \pi(i) \\ &= \sum_{i=1}^{|S|} \sum_{k=2}^{|S|} \sum_{\ell=2}^{|S|} \langle f, v_k \rangle_\pi \langle f, v_\ell \rangle_\pi v_k(i) v_\ell(i) \pi(i) \\ &= \sum_{k=2}^{|S|} \sum_{\ell=2}^{|S|} \langle f, v_k \rangle_\pi \langle f, v_\ell \rangle_\pi \langle v_k, v_\ell \rangle_\pi \\ &= \sum_{k=2}^{|S|} \langle f, v_k \rangle_\pi^2. \end{aligned} \quad (7.3)$$

Notons également que

$$Pf = \sum_{k=1}^{|S|} \langle Pf, v_k \rangle_\pi v_k = \sum_{k=1}^{|S|} \langle f, Pv_k \rangle_\pi v_k = \sum_{k=1}^{|S|} \langle f, v_k \rangle_\pi \lambda_k v_k,$$

et plus généralement, pour tout $n \in \mathbb{N}^*$,

$$P^n f = \sum_{k=1}^{|S|} \langle f, v_k \rangle_\pi \lambda_k^n v_k. \quad (7.4)$$

En particulier, pour tout $i, j \in S$,

$$p_{ij}(n) = (P^n \delta_j)(i) = \sum_{k=1}^{|S|} \langle \delta_j, v_k \rangle_\pi \lambda_k^n v_k(i) = \sum_{k=1}^{|S|} \pi(j) v_k(j) \lambda_k^n v_k(i). \quad (7.5)$$

7.2 Vitesse de convergence

Le résultat suivant fournit une borne spectrale sur la vitesse de convergence vers l'équilibre.

Théorème 7.4. *Soit P irréductible et réversible sur S . Alors,*

$$\|\mathbb{P}(X_n = \cdot | X_0 = i) - \pi\|_{\text{VT}} \leq \frac{1}{2} \sqrt{\frac{1 - \pi(i)}{\pi(i)}} (1 - \gamma_*)^n.$$

Démonstration. On a

$$\begin{aligned}
4\|\mathbb{P}(X_n = \cdot | X_0 = i) - \pi\|_{\text{VT}}^2 &= \left(\sum_{j \in S} |p_{ij}(n) - \pi(j)| \right)^2 = \left(\sum_{j \in S} \sqrt{\frac{\pi(j)}{\pi(j)}} |p_{ij}(n) - \pi(j)| \right)^2 \\
&\leq \left(\sum_{j \in S} \pi(j) \right) \left(\sum_{j \in S} \frac{1}{\pi(j)} |p_{ij}(n) - \pi(j)|^2 \right) \\
&= \sum_{j \in S} \frac{1}{\pi(j)} (p_{ij}(n)^2 - 2p_{ij}(n)\pi(j) + \pi(j)^2) \\
&= \sum_{j \in S} \frac{p_{ij}(n)^2}{\pi(j)} - 1, \tag{7.6}
\end{aligned}$$

où l'on a utilisé l'inégalité de Cauchy–Schwarz pour obtenir la seconde ligne. Observons à présent que la réversibilité de P permet de réécrire

$$\sum_{j \in S} \frac{p_{ij}(n)^2}{\pi(j)} = \sum_{j \in S} \frac{p_{ij}(n)p_{ji}(n)}{\pi(j)} \frac{\pi(j)}{\pi(i)} = \frac{1}{\pi(i)} \sum_{j \in S} p_{ij}(n)p_{ji}(n) = \frac{1}{\pi(i)} p_{ii}(2n). \tag{7.7}$$

Par (7.5) et la définition de γ_* ,

$$\begin{aligned}
p_{ii}(2n) &= \sum_{k=1}^{|S|} \pi(i) v_k(i) \lambda_k^{2n} v_k(i) = \pi(i) v_1(i)^2 \lambda_1^{2n} + \pi(i) \sum_{k=2}^{|S|} v_k(i)^2 \lambda_k^{2n} \\
&\leq \pi(i) + \pi(i) (1 - \gamma_*)^{2n} \sum_{k=2}^{|S|} v_k(i)^2. \tag{7.8}
\end{aligned}$$

En appliquant (7.2) à la fonction $f = \delta_i$, on obtient

$$\sum_{k=2}^{|S|} v_k(i)^2 = \frac{1}{\pi(i)} \sum_{k=2}^{|S|} \langle \delta_i, v_k \rangle_{\pi} v_k(i) = \frac{1}{\pi(i)} (\delta_i(i) - \langle \delta_i, \pi \rangle) = \frac{1 - \pi(i)}{\pi(i)}. \tag{7.9}$$

La conclusion suit en combinant (7.6), (7.7), (7.8) et (7.9). \square

Pour être utile quantitativement, le Théorème 7.4 requière d'être capable de calculer (ce qui est en général difficile, voire impossible) ou au moins de borner γ_* . Il existe diverses manières de procéder, dont nous décrirons certaines plus tard. Introduisons avant cela un peu de terminologie classique.

7.3 Temps de relaxation et temps de mélange

Il existe plusieurs façons de quantifier le temps nécessaire pour qu'une chaîne de Markov s'approche de l'équilibre. Nous en introduisons deux dans cette section et démontrons un résultat permettant de les comparer sous des hypothèses appropriées.

Notons $d(n) := \max_{i \in S} \|\mathbb{P}(X_n = \cdot | X_0 = i) - \pi\|_{\text{VT}}$. La première notion, celle de temps de mélange, est très naturelle.

Définition 7.5. Soit $\epsilon > 0$. Le ϵ -temps de mélange est défini par

$$t_{\text{mix}}(\epsilon) := \min\{n \in \mathbb{N} \mid d(n) \leq \epsilon\}.$$

On définit le temps de mélange par

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

(Évidemment, la valeur 1/4 dans cette définition n'a rien de spécial et est purement conventionnelle.)
La seconde notion est celle de temps de relaxation.

Définition 7.6. Le *temps de relaxation* est défini par

$$t_{\text{rel}} := \frac{1}{\gamma_*}.$$

Afin de mieux appréhender ce que mesure cette seconde notion, rappelons qu'il suit du théorème de convergence que $\lim_{n \rightarrow \infty} P^n f(i) = \langle f \rangle_\pi$ pour tout $i \in S$. En particulier, $\lim_{n \rightarrow \infty} \text{Var}_\pi(P^n f) = 0$. La notion de temps de relaxation permet de rendre cette dernière convergence quantitative : par (7.3) et (7.4), pour toute fonction $f \in \ell^2(\pi)$ et tout $n \in \mathbb{N}^*$,

$$\begin{aligned} \text{Var}_\pi(P^n f) &= \sum_{k=2}^{|S|} \langle P^n f, v_k \rangle_\pi^2 = \sum_{k=2}^{|S|} \left\langle \sum_{\ell=1}^{|S|} \langle f, v_\ell \rangle_\pi \lambda_\ell^n v_\ell, v_k \right\rangle_\pi^2 = \sum_{k=2}^{|S|} \langle f, v_k \rangle_\pi^2 \lambda_k^{2n} \\ &\leq (1 - \gamma_*)^{2n} \text{Var}_\pi(f) \leq e^{-2n/\gamma_*} \text{Var}_\pi(f) = e^{-2n/t_{\text{rel}}}, \end{aligned}$$

ce qui montre que $\text{Var}_\pi(P^n f)$ converge vers 0 exponentiellement rapidement à une échelle de temps de l'ordre du temps de relaxation.

Les deux théorèmes suivants fournissent une comparaison entre les deux notions précédentes.

Théorème 7.7. Soit P irréductible et réversible sur S . Soit $\pi_{\min} := \min_{i \in S} \pi(i)$. Alors, pour tout $\epsilon > 0$,

$$t_{\text{mix}}(\epsilon) \leq \frac{1}{2} \log\left(\frac{1 - \pi_{\min}}{4\pi_{\min}\epsilon^2}\right) t_{\text{rel}}.$$

Démonstration. Cela suit presque immédiatement du Théorème 7.4. Ce dernier implique que

$$d(n) = \max_{i \in S} \|\mathbb{P}(X_n = \cdot | X_0 = i) - \pi\|_{\text{VT}} \leq \frac{1}{2} \max_{i \in S} \sqrt{\frac{1 - \pi(i)}{\pi(i)}} (1 - \gamma_*)^n \leq \sqrt{\frac{1 - \pi_{\min}}{4\pi_{\min}}} (1 - \gamma_*)^n.$$

Comme $(1 - \gamma_*)^n \leq e^{-\gamma_* n} = e^{-n/t_{\text{rel}}}$, la conclusion suit, puisque

$$\forall \epsilon > 0, \quad n \geq \frac{1}{2} \log\left(\frac{1 - \pi_{\min}}{4\pi_{\min}\epsilon^2}\right) t_{\text{rel}} \implies d(n) \leq \epsilon. \quad \square$$

Théorème 7.8. Soit P irréductible, apériodique et réversible sur S . Alors, pour tout $\epsilon > 0$,

$$t_{\text{mix}}(\epsilon) \geq |\log(2\epsilon)| (t_{\text{rel}} - 1).$$

Démonstration. Il suit de l'orthogonalité des vecteurs propres que

$$\forall k \in \{2, \dots, |S|\}, \quad \sum_{j \in S} \pi(j) v_k(j) = \langle \mathbf{1}, v_k \rangle_\pi = \langle v_1, v_k \rangle_\pi = 0.$$

On a donc, pour tout $i \in S$ et tout $n \in \mathbb{N}^*$,

$$|\lambda_k^n v_k(i)| = |P^n v_k(i)| = \left| \sum_{j \in S} (p_{ij}(n) v_k(j) - \pi(j) v_k(j)) \right| \leq \|v_k\|_\infty 2d(n).$$

En particulier, en choisissant $i \in S$ tel que $v_k(i) = \|v_k\|_\infty$, on obtient

$$\forall k \in \{2, \dots, |S|\}, \quad |\lambda_k|^n \leq 2d(n).$$

La meilleure borne est obtenue lorsque $|\lambda_k| = 1 - \gamma_*$, ce qui donne

$$d(n) \geq \frac{1}{2}(1 - \gamma_*)^n \geq \frac{1}{2}e^{-n\gamma_*/(1-\gamma_*)},$$

puisque $1 - x \geq e^{-x/(1-x)}$ pour tout $x \in [0, 1)$. On en conclut que

$$t_{\text{mix}}(\epsilon) \geq |\log(2\epsilon)| \frac{1 - \gamma_*}{\gamma_*} = |\log(2\epsilon)|(t_{\text{rel}} - 1). \quad \square$$

7.4 Caractérisation variationnelle du trou spectral

Définition 7.9. La *forme de Dirichlet* associée à la chaîne de Markov $(X_n)_{n \geq 0}$ est définie par

$$\mathcal{E}_\pi(f) := \langle (I - P)f, f \rangle_\pi.$$

L'opérateur $I - P$ est le **laplacien** associé à la chaîne de Markov.

Observons que

$$\begin{aligned} \mathcal{E}_\pi(f) &= \sum_{i \in S} f(i) \left(f(i) - \sum_{j \in S} p_{ij} f(j) \right) \pi(i) = \sum_{i, j \in S} \pi(i) p_{ij} f(i) (f(i) - f(j)) \\ &= \sum_{i, j \in S} \pi(j) p_{ji} f(i) (f(i) - f(j)) = \sum_{i, j \in S} \pi(i) p_{ij} f(j) (f(j) - f(i)), \end{aligned} \quad (7.10)$$

où l'on a utilisé la réversibilité pour la troisième identité et interchangé i et j pour la dernière. En particulier, en sommant la troisième et la dernière expressions, on obtient

$$\mathcal{E}_\pi(f) = \frac{1}{2} \sum_{i, j \in S} \pi(i) p_{ij} (f(i) - f(j))^2. \quad (7.11)$$

Un calcul similaire donne

$$\text{Var}_\pi(f) = \frac{1}{2} \sum_{i, j \in S} \pi(i) \pi(j) (f(i) - f(j))^2. \quad (7.12)$$

En particulier, $\text{Var}_\pi(f) = 0$ si et seulement si f est un multiple de $\mathbf{1}$.

Théorème 7.10. On suppose P irréductible et réversible. Alors,

$$\gamma = \inf \left\{ \frac{\mathcal{E}_\pi(f)}{\text{Var}_\pi(f)} \mid \text{Var}_\pi(f) \neq 0 \right\}.$$

Une inégalité de la forme $\text{Var}_\pi(f) \leq c \mathcal{E}_\pi(f)$, pour un $c > 0$, est appelée une **inégalité de Poincaré**.

Démonstration. Observons tout d'abord que (7.10) et (7.12) impliquent que, pour tout $c \in \mathbb{R}$,

$$\frac{\mathcal{E}_\pi(f - c\mathbf{1})}{\text{Var}_\pi(f - c\mathbf{1})} = \frac{\mathcal{E}_\pi(f)}{\text{Var}_\pi(f)}.$$

On peut donc supposer sans perte de généralité que $\langle f \rangle_\pi = 0$. Il suit alors de (7.2) que

$$f = \sum_{k=2}^{|S|} \langle f, v_k \rangle_\pi v_k.$$

Par conséquent,

$$\begin{aligned} \mathcal{E}_\pi(f) &= \langle (I - P)f, f \rangle_\pi = \sum_{k=2}^{|S|} \sum_{\ell=2}^{|S|} \langle f, v_k \rangle_\pi \langle f, v_\ell \rangle_\pi (1 - \lambda_k) \langle v_k, v_\ell \rangle_\pi = \sum_{k=2}^{|S|} \langle f, v_k \rangle_\pi^2 (1 - \lambda_k) \\ &\geq (1 - \lambda_2) \sum_{k=2}^{|S|} \langle f, v_k \rangle_\pi^2 = (1 - \lambda_2) \langle f, f \rangle_\pi = (1 - \lambda_2) \text{Var}_\pi(f). \end{aligned}$$

La conclusion suit puisque l'inégalité ci-dessus est saturée lorsque $f = v_2$. \square

7.5 Bornes sur les trous spectraux γ et γ_*

7.5.1 Conséquences directes de la caractérisation variationnelle

Commençons par un corollaire immédiat du Théorème 7.10.

Corollaire 7.11. *On suppose P irréductible et réversible. S'il existe $A > 0$ tel que*

$$\forall f \in \ell^2(\pi), \quad \text{Var}_\pi(f) \leq A \mathcal{E}_\pi(f), \quad (7.13)$$

alors $\gamma \geq 1/A$.

Remarque 7.12. *Évidemment, le résultat précédent ne nous renseigne pas sur la valeur du trou spectral absolu, qui est celui pertinent pour l'application, par exemple, du Théorème 7.4. L'observation suivante est donc intéressante. Si l'on remplace la matrice de transition P par la matrice de transition $\frac{1}{2}(I + P)$, c'est-à-dire que l'on considère la version **paresseuse** de la chaîne, en ajoutant à chaque pas une probabilité $1/2$ de rester sur place, alors on peut vérifier (cela sera fait lors des séances d'exercices) que les valeurs propres de cette nouvelle matrice de transition sont toutes positives; en particulier, $\gamma_* = \gamma$ pour la variante paresseuse de la chaîne de Markov. Notons également que cette modification ne modifie pas l'ordre de grandeur du temps de mélange (en gros, la chaîne évolue deux fois plus lentement, puisqu'elle reste sur place la moitié du temps). \diamond*

Plus généralement, des résultats de comparaison entre chaînes de Markov sont souvent utiles. En voici un autre.

Lemme 7.13. *Soit P et \tilde{P} deux matrices stochastiques irréductibles et réversibles, sur le même espace d'états fini S . Soit π et $\tilde{\pi}$ les lois stationnaires associées. Supposons qu'il existe $A, B \in \mathbb{R}_+^*$ tels que, pour tout $f : S \rightarrow \mathbb{R}$ et tout $i \in S$,*

$$\pi(i) \leq A \tilde{\pi}(i) \quad \text{et} \quad \mathcal{E}_\pi(f) \geq B \mathcal{E}_{\tilde{\pi}}(f).$$

Alors, $\gamma \geq \frac{B}{A} \tilde{\gamma}$, où γ et $\tilde{\gamma}$ sont les trous spectraux associés à P et \tilde{P} respectivement.

Démonstration. Observons que

$$\min_{c \in \mathbb{R}} \|f - c\mathbf{1}\|_\pi^2 = \|f - \langle f \rangle_\pi \mathbf{1}\|_\pi^2 = \text{Var}_\pi(f).$$

On a donc

$$\text{Var}_\pi(f) \leq \|f - \langle f \rangle_{\tilde{\pi}} \mathbf{1}\|_\pi^2 = \sum_{i \in S} (f(i) - \langle f \rangle_{\tilde{\pi}})^2 \pi(i) \leq A \sum_{i \in S} (f(i) - \langle f \rangle_{\tilde{\pi}})^2 \tilde{\pi}(i) = A \text{Var}_{\tilde{\pi}}(f).$$

Par conséquent,

$$\frac{\mathcal{E}_\pi(f)}{\text{Var}_\pi(f)} \geq \frac{B}{A} \frac{\mathcal{E}_{\tilde{\pi}}(f)}{\text{Var}_{\tilde{\pi}}(f)}.$$

La conclusion suit du Théorème 7.10. \square

7.5.2 Chemins pondérés

Décrivons à présent une autre approche possible pour obtenir des bornes sur γ et γ_* , basée sur la géométrie du graphe de transition associé à P .

Nous utiliserons les conventions suivantes : si $e = (i, j)$ est une arête orientée, notons $e^- := i$ et $e^+ := j$, et définissons $Q(e) := \pi(i)p_{ij} = \pi(j)p_{ji}$.

À chaque paire de sommets distincts $i, j \in S$, associons (de façon arbitraire) un unique chemin $\gamma_{ij} = (i, i_1, \dots, i_m, j)$ tel que $p_{ii_1}p_{i_1i_2} \cdots p_{i_mj} > 0$ et ne visitant aucune arête plus d'une fois. Soit Γ l'ensemble de tous ces chemins. Pour chaque $\gamma_{ij} \in \Gamma$, on pose

$$|\gamma_{ij}|_Q := \sum_{e \in \gamma_{ij}} \frac{1}{Q(e)}.$$

Définition 7.14. *Le coefficient de Poincaré de P est défini par*

$$\kappa = \kappa(\Gamma) := \max_e \sum_{\gamma_{ij} \ni e} |\gamma_{ij}|_Q \pi(i) \pi(j).$$

Théorème 7.15. *Soit P irréductible et réversible. Alors, $\gamma \geq 1/\kappa$.*

Démonstration. Il suffit de vérifier que la condition (7.13) est vérifiée avec $A = \kappa$. Tout d'abord,

$$\text{Var}_\pi(f) = \frac{1}{2} \sum_{i,j \in S} \pi(i) \pi(j) (f(i) - f(j))^2 = \frac{1}{2} \sum_{i,j \in S} \pi(i) \pi(j) \left[\sum_{e \in \gamma_{ij}} (f(e^+) - f(e^-)) \right]^2.$$

Observons que l'inégalité de Cauchy–Schwarz donne

$$\begin{aligned} \left[\sum_{e \in \gamma_{ij}} (f(e^+) - f(e^-)) \right]^2 &= \left[\sum_{e \in \gamma_{ij}} \frac{1}{\sqrt{Q(e)}} \sqrt{Q(e)} (f(e^+) - f(e^-)) \right]^2 \\ &\leq \sum_{e' \in \gamma_{ij}} \frac{1}{Q(e')} \sum_{e \in \gamma_{ij}} Q(e) (f(e^+) - f(e^-))^2 \\ &= |\gamma_{ij}|_Q \sum_{e \in \gamma_{ij}} Q(e) (f(e^+) - f(e^-))^2. \end{aligned}$$

Par conséquent,

$$\begin{aligned} \text{Var}_\pi(f) &\leq \frac{1}{2} \sum_{i,j \in S} \pi(i) \pi(j) |\gamma_{ij}|_Q \sum_{e \in \gamma_{ij}} Q(e) (f(e^+) - f(e^-))^2 \\ &= \frac{1}{2} \sum_e Q(e) (f(e^+) - f(e^-))^2 \sum_{\gamma_{ij} \ni e} |\gamma_{ij}|_Q \pi(i) \pi(j) \\ &\leq \kappa \frac{1}{2} \sum_e Q(e) (f(e^+) - f(e^-))^2 = \kappa \mathcal{E}_\pi(f), \end{aligned}$$

où l'on a utilisé (7.11) pour la dernière identité. \square

Tournons-nous à présent vers la dérivation d'une borne inférieure sur $\lambda_{|S|}$. À chaque $i \in S$, on associe un unique chemin fermé $\sigma_i = (i, i_1, \dots, i_m, i)$ traversant chaque arête au plus une fois et tel que $p_{ii_1} \cdots p_{i_m i} > 0$ et m est impair. Afin de pouvoir satisfaire cette dernière condition, on supposera P a périodique¹. Soit Σ la collection de tous ces chemins.

De façon similaire à ce que l'on a fait précédemment, on définit

$$|\sigma_i|_Q := \sum_{e \in \sigma_i} \frac{1}{Q(e)}, \quad \text{et} \quad \alpha = \alpha(\Sigma) := \max_e \sum_{\sigma_i \ni e} \pi(i) |\sigma_i|_Q.$$

1. Si tous les chemins menant de i à i sont de longueur paire, alors i est de période au moins 2.

Théorème 7.16. *Soit P irréductible, apériodique et réversible. Alors, $\lambda_{|S|} \geq -1 + \frac{2}{\alpha}$.*

Démonstration. Il suffit de montrer que, pour tout $f \in \ell^2(\pi)$,

$$\|f\|_\pi^2 \leq \frac{\alpha}{2} (\|f\|_\pi^2 + \langle Pf, f \rangle_\pi). \quad (7.14)$$

En effet, appliqué à la fonction $f = v_{|S|}$, cette inégalité devient $1 \leq \frac{\alpha}{2}(1 + \lambda_{|S|})$.

Observons tout d'abord que, pour tout $\sigma_i = (i, i_1, \dots, i_m, i) \in \Sigma$,

$$\begin{aligned} f(i) &= \frac{1}{2}((f(i) + f(i_1)) - (f(i_1) + f(i_2)) + (f(i_2) + f(i_3)) - \dots + (f(i_m) + f(i))) \\ &= \frac{1}{2} \sum_{e \in \sigma_i} (-1)^{n(e)} (f(e^-) + f(e^+)), \end{aligned}$$

où l'on a noté $n(e) := k$ si $e = (i_k, i_k + 1)$ (avec la convention que $i_0 := i$) et utilisé l'hypothèse que $|\sigma_i|$ est impair. Ainsi,

$$\begin{aligned} \|f\|_\pi^2 &= \sum_{i \in S} \frac{1}{4} \pi(i) \left[\sum_{e \in \sigma_i} (-1)^{n(e)} (f(e^-) + f(e^+)) \right]^2 \\ &= \sum_{i \in S} \frac{1}{4} \pi(i) \left[\sum_{e \in \sigma_i} \sqrt{\frac{Q(e)}{Q(e)}} (-1)^{n(e)} (f(e^-) + f(e^+)) \right]^2 \\ &\leq \sum_{i \in S} \frac{1}{4} \pi(i) |\sigma_i|_Q \sum_{e \in \sigma_i} Q(e) (f(e^-) + f(e^+))^2, \end{aligned}$$

où l'on a procédé comme dans la preuve du Théorème 7.15. On obtient donc

$$\|f\|_\pi^2 = \frac{1}{4} \sum_e Q(e) (f(e^-) + f(e^+))^2 \sum_{\sigma_i \ni e} \pi(i) |\sigma_i|_Q \leq \frac{\alpha}{4} \sum_e Q(e) (f(e^-) + f(e^+))^2.$$

(7.14) suit immédiatement, puisque

$$\begin{aligned} \frac{1}{2} \sum_e Q(e) (f(e^-) + f(e^+))^2 &= \sum_e Q(e) f(e^-)^2 + \sum_e Q(e) f(e^-) f(e^+) \\ &= \sum_i \pi(i) f(i)^2 \sum_j p_{ij} + \sum_i \pi(i) f(i) \sum_j p_{ij} f(j) = \|f\|_\pi^2 + \langle Pf, f \rangle_\pi, \end{aligned}$$

où la première identité suit du fait que l'expression est symétrique en e^- , e^+ . \square

Exemple 7.17. Soit $G = (S, E)$ un graphe fini simple connexe. Considérons la marche aléatoire simple sur G , c'est-à-dire la chaîne de Markov sur S dont les probabilités de transition sont données par $p_{ij} := \mathbf{1}_{\{i,j\} \in E} / d(i)$, où $d(i) := |\{j \in S \mid \{i,j\} \in E\}|$ est le degré du sommet i . On vérifie aisément que cette chaîne de Markov est réversible pour la mesure de probabilité $\pi(i) := d(i) / 2|E|$. En particulier, $Q(\{i,j\}) = \pi(i) p_{ij} = 1/2|E|$ pour tout arête $\{i,j\} \in E$.

On souhaite appliquer le Théorème 7.15. Observons que, pour tout $\gamma_{ij} \in \Gamma$,

$$|\gamma_{ij}|_Q = \sum_{e \in \gamma_{ij}} \frac{1}{Q(e)} = 2|E| |\gamma_{ij}|,$$

où l'on a dénoté $|\gamma|$ la longueur du chemin γ . Par conséquent,

$$\kappa = \max_e \sum_{\gamma_{ij} \ni e} |\gamma_{ij}|_Q \pi(i) \pi(j) = \max_e \sum_{\gamma_{ij} \ni e} 2|E| |\gamma_{ij}| \frac{d(i)}{2|E|} \frac{d(j)}{2|E|} \leq K \frac{d_{\max}^2}{2|E|},$$

où $K := \max_e \sum_{\gamma_{ij} \ni e} |\gamma_{ij}|$ et $d_{\max} := \max_i d(i)$. On conclut donc du Théorème 7.15 que

$$\gamma \geq \frac{2|E|}{Kd_{\max}^2}.$$

À présent, bornons inférieurement $\lambda_{|S|}$ à l'aide du Théorème 7.16. On suppose donc que le graphe G est tel que la marche aléatoire est apériodique. On a

$$\alpha = \max_e \sum_{\sigma_i \ni e} \pi(i) |\sigma_i|_Q = \max_e \sum_{\sigma_i \ni e} \frac{d(i)}{2|E|} 2|E| |\sigma_i| \leq K' d_{\max},$$

où $K' := \max_e \sum_{\sigma_i \ni e} |\sigma_i|$. Par conséquent,

$$\lambda_{|S|} \geq -1 + \frac{2}{K' d_{\max}}.$$

En combinant ces deux bornes, on conclut que

$$\gamma_* \geq \min \left\{ \frac{2|E|}{Kd_{\max}^2}, \frac{2}{K'd_{\max}} \right\}. \quad \diamond$$

Remarques bibliographiques : La source principale pour ce chapitre est le livre [6], auquel on pourra se référer pour des informations supplémentaires. Il existe de nombreux autres excellents livres sur ce sujet, par exemple [17].

8 Méthode de Monte Carlo et simulation parfaite

Dans ce chapitre, nous allons décrire un algorithme permettant d'obtenir un échantillon distribué exactement selon la loi stationnaire d'une chaîne de Markov. Couplé avec un algorithme de Monte Carlo, ceci fournit une approche pour obtenir sans approximation des échantillons distribués selon une mesure de probabilité quelconque sur un ensemble fini.

8.1 Méthode de Monte Carlo

Soit S un ensemble fini et π une mesure de probabilité sur S associant à chaque $i \in S$ une probabilité $\pi(i) > 0$. Un problème d'importance centrale dans de nombreux domaines est d'effectuer numériquement des tirages d'éléments de S selon π . Une approche classique à ce problème consiste à construire une chaîne de Markov irréductible et apériodique $(X_n)_{n \geq 0}$ sur S dont π est l'unique loi stationnaire. En effet, en notant \mathbb{P} la loi de la chaîne, le théorème de convergence garantit que

$$\forall j \in S, \quad \lim_{n \rightarrow \infty} \mathbb{P}(X_n = i \mid X_0 = j) = \pi(i).$$

Ainsi, pour n suffisamment grand, X_n sera approximativement distribué selon π . Évidemment, quantifier ce que l'on entend par « suffisamment grand », ainsi qu'estimer l'erreur commise n'est pas chose triviale ; le but de ce chapitre sera de décrire une méthode permettant d'éliminer ces problèmes.

8.1.1 Construction d'une chaîne de Markov appropriée

Considérons tout d'abord le problème de construire une chaîne de Markov irréductible et apériodique $(X_n)_{n \geq 0}$ sur S ayant π comme loi stationnaire. Il y a évidemment une infinité de telles chaînes. Il suffit par exemple de choisir les probabilités de transition p_{ij} de telle sorte que

$$\forall i, j \in S, \quad \pi(i)p_{ij} = \pi(j)p_{ji}. \quad (8.1)$$

En effet, il suit du cours d'introduction à la théorie des probabilités que π est alors automatiquement l'unique loi stationnaire de la chaîne (et que celle-ci est réversible par rapport à π).

Soit $Q = (q_{ij})_{i,j \in S}$ une matrice stochastique irréductible arbitraire sur S , que l'on utilisera comme référence lors de la construction. On cherche une solution à (8.1) de la forme

$$\forall i, j \in S \text{ distincts}, \quad p_{ij} = a_{ij}q_{ij},$$

avec $a_{ij} \in [0, 1]$ pour tout $i, j \in S$ distincts. Ceci peut être interprété comme suit : si l'état de la chaîne au temps n est $i \in S$, on tire au hasard son état au temps $n + 1$ de la façon suivante :

- ▷ On tire au hasard un candidat $j \in S$ avec probabilité q_{ij} .
- ▷ On tire au hasard uniformément un nombre $u \in [0, 1]$.
- ▷ Si $j \neq i$ et $u \leq a_{ij}$, on déclare que l'état de la chaîne au temps $n + 1$ est j .
- ▷ Si $j = i$ ou $u > a_{ij}$, on déclare que l'état de la chaîne au temps $n + 1$ est i .

Les nombres $(a_{ij})_{i,j \in S}$ sont appelés **probabilités d'acceptation**. Il reste à déterminer ces derniers afin de satisfaire (8.1). Une famille classique de solutions est donnée par des probabilités d'acceptation de la forme

$$a_{ij} = \frac{s_{ij}}{1 + t_{ij}},$$

où $(s_{ij})_{i,j \in S}$ est une matrice *symétrique* et

$$t_{ij} = \frac{\pi(i)q_{ij}}{\pi(j)q_{ji}}.$$

Il faut évidemment choisir la matrice (s_{ij}) de sorte à ce que $a_{ij} \in [0, 1]$ pour tout $i, j \in S$ distincts. Avec un tel choix, on a bien

$$\pi(i)p_{ij} = \frac{\pi(i)q_{ij}s_{ij}\pi(j)q_{ji}}{\pi(i)q_{ij} + \pi(j)q_{ji}} = \pi(j)p_{ji},$$

puisque l'expression intermédiaire est symétrique en i et j .

8.1.2 Quelques exemples

Exemple 8.1 (Algorithme de Metropolis–Hastings). Observons qu'afin de garantir que $a_{ij} \in [0, 1]$, on doit nécessairement avoir $s_{ij} \leq 1 + \min\{t_{ij}, t_{ji}\}$. L'algorithme de Metropolis consiste à saturer cette inégalité, ce qui conduit à

$$\forall i, j \in S \text{ distincts}, \quad a_{ij} = \min\left\{1, \frac{\pi(j)q_{ji}}{\pi(i)q_{ij}}\right\}.$$

La première version de cet algorithme a été publiée par Metropolis et Ulam en 1949 [20, 19]. La version générale est due à Hastings [12]. \diamond

Exemple 8.2 (Algorithme de Barker). L'algorithme proposé par Barker en 1965 [4] consiste à prendre $s_{ij} = 1$ pour tout $i, j \in S$ distincts. On obtient alors

$$\forall i, j \in S \text{ distincts}, \quad a_{ij} = \frac{\pi(j)q_{ji}}{\pi(j)q_{ji} + \pi(i)q_{ij}}. \quad \diamond$$

Exemple 8.3 (Échantillonneur de Gibbs). On décrit à présent un choix de probabilité de transition de nature différente, très souvent utilisé dans la simulation de **champs markoviens**. Afin de rester aussi concrets que possible, nous nous contenterons de discuter un cas particulier célèbre : le **modèle d'Ising**.

Soit $G = (V, A)$ un graphe fini simple et $S := \{\sigma = (\sigma_i)_{i \in V} \in \{-1, 1\}^V\}$ l'ensemble des **configurations** du modèle (cf. Figure 8.1). Le **hamiltonien** du modèle est la fonction $H : S \rightarrow \mathbb{R}$ définie par

$$H(\sigma) := - \sum_{\{i,j\} \in A} \sigma_i \sigma_j.$$

Pour chaque $\beta \in \mathbb{R}$, on considère la mesure de probabilité sur S définie par

$$\pi_{G,\beta}(\sigma) := \frac{e^{-\beta H(\sigma)}}{Z_{G,\beta}}, \quad Z_{G,\beta} := \sum_{\sigma \in S} e^{-\beta H(\sigma)},$$

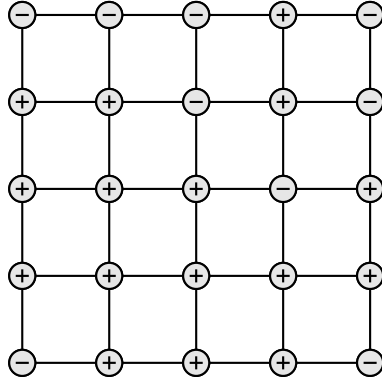


FIGURE 8.1: Une configuration du modèle d'Ising sur le graphe $G = (V, A)$ avec $V = \{1, \dots, 5\}^2$ et $\{i, j\} \in A$ si et seulement si $\|i - j\|_1 = 1$.

appelée la **mesure de Gibbs** associée au modèle d'Ising. Le paramètre $\beta > 0$ est appelé la **température inverse** et joue un rôle essentiel dans l'étude de ce modèle. La constante de normalisation $Z_{G,\beta}$ est appelée la **fonction de partition**.

Étant donné $\sigma \in S$ et $i \in V$, on notera $\sigma^{i+}, \sigma^{i-} \in S$ les configurations définies par $\sigma_j^{i\pm} := \sigma_j$ pour tout $j \neq i$ et $\sigma_i^{i\pm} := \pm 1$.

L'échantillonneur de Gibbs est une implémentation de la méthode de Monte Carlo afin d'échantillonner des configurations de ce modèle (ou d'autres champs markoviens). Il propose de passer d'une configuration σ à une autre configuration σ' de la façon suivante : on tire, indépendamment, un nombre $u \in [0, 1]$ selon la loi uniforme et un sommet $i \in V$ également de façon uniforme. On pose alors

$$\sigma' := \begin{cases} \sigma^{i+} & \text{si } u \leq \frac{\pi_{G,\beta}(\sigma^{i+})}{\pi_{G,\beta}(\sigma^{i+}) + \pi_{G,\beta}(\sigma^{i-})} = \left(1 + \exp\left[-2\beta \sum_{j \sim i} \sigma_j\right]\right)^{-1}, \\ \sigma^{i-} & \text{sinon,} \end{cases} \quad (8.2)$$

où la notation $j \sim i$ signifie que j est un voisin de i , c'est-à-dire que $\{i, j\} \in A$. En d'autres termes, on choisit un sommet $i \in V$ au hasard, et on met à jour σ_i en tirant sa valeur au hasard selon la loi $\pi_{G,\beta}$ conditionnellement au reste de la configuration, $(\sigma_j)_{j \in V \setminus \{i\}}$.

Notons $p_{\sigma\sigma'}$ les probabilités de transition correspondantes. Clairement, $p_{\sigma\sigma'} \neq 0$ si et seulement si σ et σ' diffèrent en au plus un sommet. De plus, il est immédiat que la chaîne de Markov ainsi définie est irréductible et apériodique. Vérifions qu'elle est bien réversible par rapport à $\pi_{G,\beta}$: pour toute configuration $\sigma \in S$ et tout $i \in V$,

$$\begin{aligned} \pi_{G,\beta}(\sigma^{i+}) p_{\sigma^{i+}\sigma^{i-}} &= \pi_{G,\beta}(\sigma^{i+}) \frac{1}{|V|} \frac{\pi_{G,\beta}(\sigma^{i-})}{\pi_{G,\beta}(\sigma^{i+}) + \pi_{G,\beta}(\sigma^{i-})} \\ &= \frac{1}{|V|} \frac{\pi_{G,\beta}(\sigma^{i+}) \pi_{G,\beta}(\sigma^{i-})}{\pi_{G,\beta}(\sigma^{i+}) + \pi_{G,\beta}(\sigma^{i-})} = \pi_{G,\beta}(\sigma^{i-}) p_{\sigma^{i-}\sigma^{i+}}. \end{aligned} \quad \diamond$$

8.2 Simulation parfaite

Nous avons vu qu'il est aisé de construire des chaînes de Markov dont la loi stationnaire est donnée par une loi que l'on désire simuler. Nous sommes à présent confrontés à un problème d'ordre pratique : on ne peut évidemment pas laisser cet algorithme tourner infiniment longtemps, ce qui nous force à l'interrompre après un nombre fini M d'itérations. Comment doit-on choisir M si l'on veut être assuré d'être proche de la distribution stationnaire ? Une approche possible est d'étudier la vitesse de convergence de la chaîne (par exemple, avec les méthodes décrites dans le Chapitre 7). Malheureusement, cela ne fournit en général que des bornes trop peu quantitatives pour être réellement utiles en pratique, même si elles peuvent donner des indications.

Nous allons à présent présenter une approche, due à Propp et Wilson [21] et appelée « couplage depuis le passé » (*coupling from the past* en anglais), qui permet de générer des configurations distribuées *exactement* selon la loi stationnaire d'une chaîne de Markov, tout en ne nécessitant qu'un nombre fini d'itérations. Afin de faciliter sa description, nous allons faire une brève parenthèse pour expliquer comment on peut exprimer une chaîne de Markov en termes de fonctions aléatoires itérées.

8.2.1 Fonctions aléatoires itérées

Soit S un ensemble fini et $(X_n)_{n \in \mathbb{Z}}$ une chaîne de Markov sur S avec probabilités de transition $(p_{ij})_{i,j \in S}$. Une façon équivalente de représenter cette chaîne de Markov consiste à considérer une famille $(f_k)_{k \in \mathbb{Z}}$ de fonctions aléatoires i.i.d. $f_k : S \rightarrow S$ dont la loi \mathbb{Q} est définie par

$$\forall j \in S, \quad \mathbb{Q}(f_k(i) = j) := p_{ij},$$

indépendamment pour chaque $i \in S$. On a alors, pour tout $m, n \in \mathbb{Z}$ tels que $m < n$,

$$\mathbb{P}(X_n = j \mid X_m = i) = \mathbb{Q}(F_m^n(i) = j),$$

où l'on a introduit la notation $F_m^n(i) := f_{n-1} \circ f_{n-2} \circ \dots \circ f_{m+1} \circ f_m(i)$. En effet, on a $\mathbb{Q}(F_m^{m+1}(i) = j) = \mathbb{Q}(f_m(i) = j) = p_{ij} = \mathbb{P}(X_{m+1} = j \mid X_m = i)$. Procédons donc par récurrence en supposant que $\mathbb{P}(X_\ell = j \mid X_m = i) = \mathbb{Q}(F_m^\ell(i) = j)$ pour tout $\ell \in \llbracket m+1, n-1 \rrbracket$ et montrons qu'elle reste vraie pour $\ell = n$:

$$\begin{aligned} \mathbb{Q}(F_m^n(i) = j) &= \mathbb{Q}(f_{n-1} \circ F_m^{n-1}(i) = j) \\ &= \sum_{k \in S} \mathbb{Q}(f_{n-1}(k) = j, F_m^{n-1}(i) = k) \\ &= \sum_{k \in S} \mathbb{Q}(f_{n-1}(k) = j) \mathbb{Q}(F_m^{n-1}(i) = k) \\ &= \sum_{k \in S} \mathbb{P}(X_n = j \mid X_{n-1} = k) \mathbb{P}(X_{n-1} = k \mid X_m = i) = \mathbb{P}(X_n = j \mid X_m = i), \end{aligned}$$

où la troisième identité suit de l'indépendance des $(f_r)_{r \in \mathbb{Z}}$, la quatrième de l'hypothèse de récurrence et la dernière de la propriété de Markov.

Exemple 8.4. Considérons la chaîne de Markov $(X_n)_{n \in \mathbb{Z}}$ sur $\{0, 1\}$ avec probabilités de transition $p_{01} := p, p_{00} := 1-p, p_{10} := q$ et $p_{11} := 1-q$. Il existe 4 fonctions $g_i : \{0, 1\} \rightarrow \{0, 1\}, i = 1, \dots, 4 : g_1(0) = 0, g_1(1) = 1; g_2(0) = g_2(1) = 0; g_3(0) = 1, g_3(1) = 0; g_4(0) = g_4(1) = 1$. La chaîne $(X_n)_{n \in \mathbb{Z}}$ est alors équivalente à l'itération de fonctions aléatoires i.i.d. $(f_k)_{k \in \mathbb{Z}}$ de loi \mathbb{Q} telle que

$$\mathbb{Q}(f_k = g_1) = (1-p)(1-q), \quad \mathbb{Q}(f_k = g_2) = (1-p)q, \quad \mathbb{Q}(f_k = g_3) = pq, \quad \mathbb{Q}(f_k = g_4) = p(1-q).$$

La Figure 8.2 illustre cette construction. ◇

8.2.2 Algorithme de couplage depuis le passé

Soit $(p_{ij})_{i,j \in S}$ les probabilités de transition d'une chaîne de Markov $X = (X_k)_{k \in \mathbb{Z}}$ irréductible, apériodique et de loi stationnaire π . Afin d'analyser la dépendance en l'état initial de la chaîne, on démarre, au temps m , $|S|$ copies de celle-ci, chacune partant d'un état différent. L'évolution se fait de façon indépendante tant que les trajectoires ne coïncident pas. Si deux (ou plus) trajectoires coïncident à un temps $n > m$, alors on les fait évoluer ensemble pour tous les temps ultérieurs, comme représenté sur la Figure 8.3. On réalise ainsi un couplage de ces $|S|$ chaînes. Une façon équivalente de présenter ce couplage est de l'encoder en termes d'une famille $(f_k)_{k \in \mathbb{Z}}$ de fonctions aléatoires i.i.d. $f_k : S \rightarrow S$ de loi

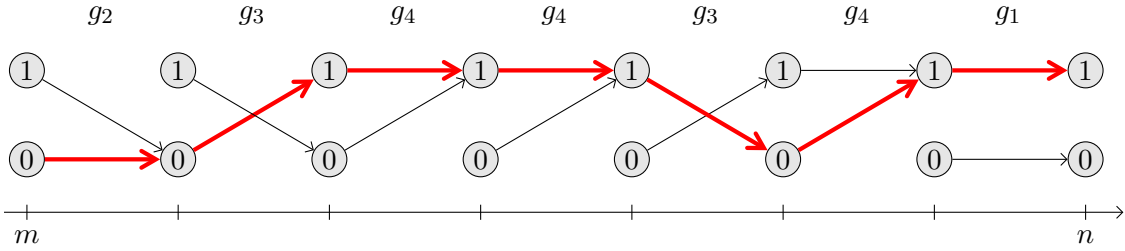


FIGURE 8.2: Une partie d’une réalisation de la trajectoire de la chaîne de Markov $(X_n)_{n \in \mathbb{Z}}$ de l’Exemple 8.4, exprimée en termes des fonctions aléatoires $(f_k)_{k \in \mathbb{Z}}$. Dans cette réalisation, $F_m^n(0) = 1$ (en rouge).

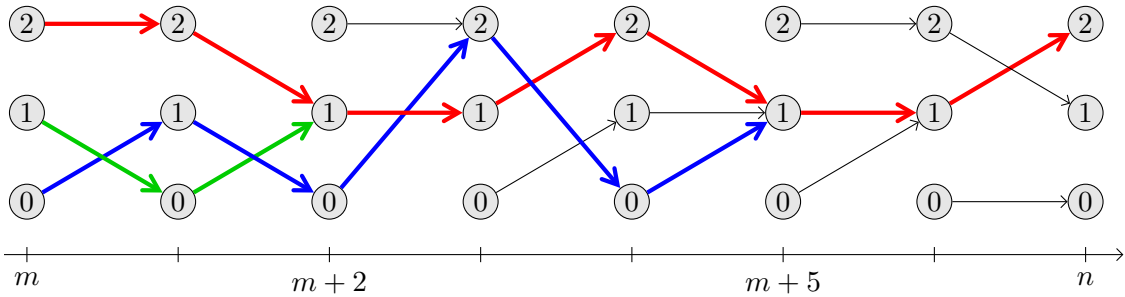


FIGURE 8.3: Trois copies d’une chaîne de Markov sur $S = \{0, 1, 2\}$, partant de trois états différents au temps m . Les trajectoires sont indiquées par les flèches en gras. Les trajectoires issues de 1 et 2 se rejoignent au temps $m + 2$, puis leur trajectoire commune fusionne avec celle issue de 0 au temps $m + 5$.

\mathbb{Q} comme expliqué dans la section précédente. En particulier, la distribution au temps $n \in \mathbb{Z}$ de la chaîne partant de l’état i au temps $m < n$ est identique à celle de $F_m^n(i) := f_{n-1} \circ f_{n-2} \circ \dots \circ f_{m+1} \circ f_m(i)$. Les $|S|$ trajectoires auront donc fusionné si et seulement si la fonction F_m^n est constante.

Le résultat suivant garantit que toutes les trajectoires auront fusionné après un temps presque-sûrement fini.

Lemme 8.5. \mathbb{Q} -presque-sûrement, $F_0^n = \text{const}$ pour tout n suffisamment grand.

Démonstration. Par irréductibilité et apériodicité, il existe $K < \infty$ tel que la probabilité d’aller de i à j en K pas est strictement positive pour toute paire $i, j \in S$. Par conséquent, $\mathbb{Q}(F_0^K = \text{const})$ est également strictement positive. Les indicatrices des événements $\{F_{nK}^{(n+1)K} = \text{const}\}$, $n \geq 0$, étant i.i.d., on peut, \mathbb{Q} -presque sûrement, trouver $N \in \mathbb{N}$ tel que $F_{NK}^{(N+1)K}$ soit constante.

Soit $s > r \geq 0$; observant que $F_r^s = \text{const}$ implique $F_0^n = \text{const}$ pour tout $n \geq s$, on en déduit que F_0^n sera constante pour tout $n \geq (N + 1)K$. \square

Manifestement, une fois que les $|S|$ trajectoires ont fusionné, toute information sur l’état de départ est perdue. On pourrait donc penser que l’on obtient à cet instant un échantillon distribué selon la loi stationnaire. Mais c’est faux, comme on peut le voir facilement sur l’exemple de la Figure 8.4. Le problème est que l’observation n’est pas faite en un temps déterministe, ce qui conduit à un biais. Cette idée n’est cependant pas à rejeter complètement, une modification très simple permettant de la faire fonctionner.

L’idée est de ne pas chercher à coupler « dans le futur » comme on vient de le faire, mais « depuis le passé » : on va démarrer les $|S|$ chaînes à un temps suffisamment reculé dans le passé, puis observer le résultat au temps 0. Plus précisément, on sait que $F_{-n}^0 = f_{-1} \circ f_{-2} \circ \dots \circ f_{-n+1} \circ f_{-n}$ est constante pour tout n suffisamment grand, puisque $F_{-n}^0 \stackrel{\text{loi}}{=} F_0^n$. Observez que l’on n’exige pas que le premier temps auquel toutes les copies ont fusionné soit égal à 0, mais seulement que cela ait eu lieu *avant* 0; cette nuance est cruciale.

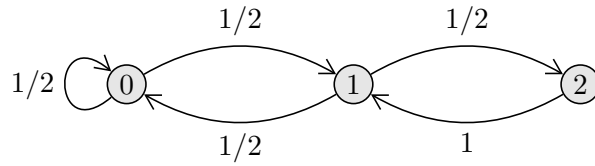


FIGURE 8.4: Un exemple montrant qu'à l'instant où les trajectoires ont toutes fusionné, la chaîne n'est pas nécessairement distribuée selon la loi stationnaire : l'état 2 ne pouvant être atteint qu'en venant de l'état 1, il est impossible que les chaînes soient dans l'état 2 lorsque la fusion de toutes les trajectoires se produit, alors que 2 a évidemment une probabilité strictement positive sous la loi stationnaire.

Théorème 8.6. Soit M une variable aléatoire \mathbb{Q} -presque sûrement finie telle que F_{-M}^0 soit constante. Alors, l'unique image de F_{-M}^0 est distribuée selon la loi π .

Démonstration. Fixons $i, j \in S$. Pour tout $n \in \mathbb{N}$,

$$\begin{aligned} \mathbb{Q}(F_{-n}^0(i) = j) &= \mathbb{Q}(F_{-n}^0(i) = j, M \leq n) + \mathbb{Q}(F_{-n}^0(i) = j, M > n) \\ &= \mathbb{Q}(F_{-M}^0(i) = j, M \leq n) + \mathbb{Q}(F_{-n}^0(i) = j, M > n) \\ &= \mathbb{Q}(F_{-M}^0(i) = j) - \mathbb{Q}(F_{-M}^0(i) = j, M > n) + \mathbb{Q}(F_{-n}^0(i) = j, M > n). \end{aligned}$$

Donc,

$$\begin{aligned} |\mathbb{Q}(F_{-M}^0(i) = j) - \pi(j)| &\leq |\mathbb{Q}(F_{-n}^0(i) = j) - \pi(j)| + |\mathbb{Q}(F_{-M}^0(i) = j) - \mathbb{Q}(F_{-n}^0(i) = j)| \\ &\leq |\mathbb{P}(X_n = j | X_0 = i) - \pi(j)| + \mathbb{Q}(M > n), \end{aligned}$$

où l'on a utilisé le fait que $\mathbb{Q}(F_{-n}^0(i) = j) = \mathbb{P}(X_n = j | X_0 = i)$ par construction. La conclusion suit en prenant la limite $n \rightarrow \infty$, puisque $\lim_{n \rightarrow \infty} |\mathbb{P}(X_n = j | X_0 = i) - \pi(j)| = 0$, la chaîne $(X_k)_{k \geq 0}$ étant irréductible et apériodique, et $\lim_{n \rightarrow \infty} \mathbb{Q}(M > n) = 0$ par hypothèse. \square

Remarque 8.7. Il est instructif d'observer ce qui ne marcherait pas si l'on avait procédé « vers le futur » : dans ce cas, il n'est plus vrai que $F_0^n(i) = F_0^M(i)$ pour tout $n \geq M$. Un coup d'œil à la Figure 8.5 (bas) devrait rendre cela évident. \diamond

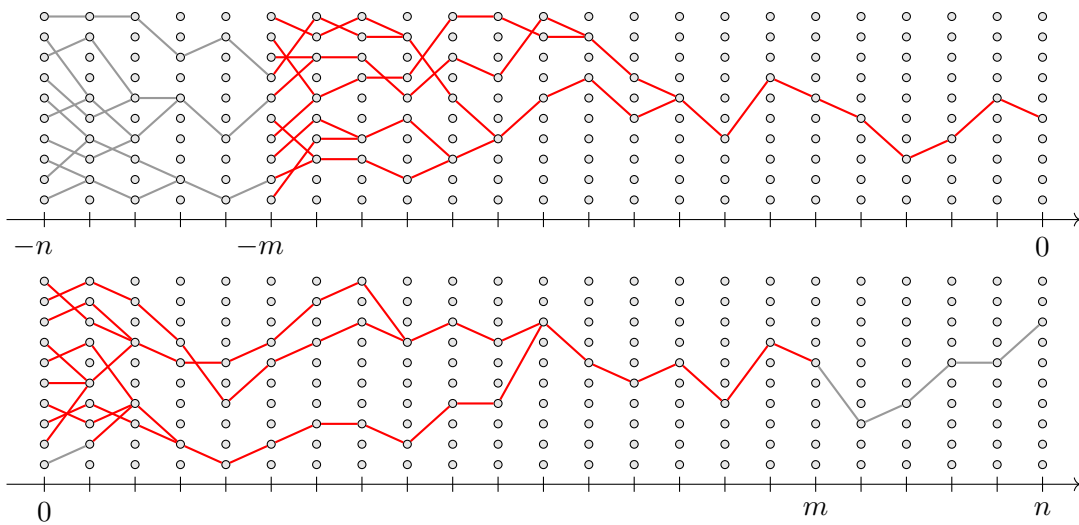


FIGURE 8.5: Haut : couplage depuis le passé ; F_{-m}^0 étant constante, c'est également le cas de F_{-n}^0 , et leur unique image est identique. Bas : évolution vers le futur ; F_0^m étant constante, F_0^n l'est aussi, mais il n'est plus vrai en général que leur unique image est identique.

En résumé, l'algorithme suivant renvoie un élément de S distribué selon π après un temps presque-sûrement fini (voir aussi la Figure 8.6) :

Algorithme 2 : couplage depuis le passé

Initialisation : $n = 0$, $F_0^0 = \text{identité}$

répéter

 On remplace n par $n + 1$

 On tire f selon \mathbb{Q}

 On pose $F_{-n}^0 = F_{-n+1}^0 \circ f$

jusqu'à F_{-n}^0 est constante

retourner l'unique valeur dans l'image de F_{-n}^0

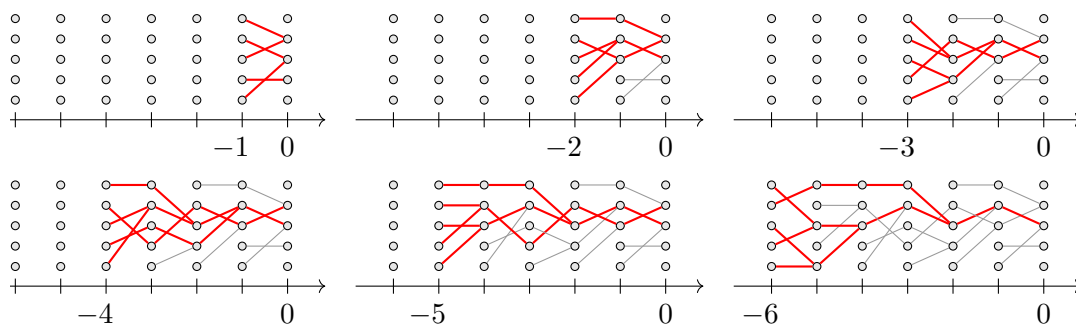


FIGURE 8.6: Une réalisation de l'algorithme de couplage depuis le passé. La fonction F_{-n}^0 ne devient constante qu'à partir de $n = 6$. Observez que le temps auquel les 5 trajectoires fusionnent n'est pas égal à 0, mais à -2 .

8.2.3 Quelques commentaires à propos de cet algorithme

Temps d'échantillonnage. Il n'est pas nécessaire (et pas désirable du tout !) d'appliquer l'algorithme ci-dessus pour chaque $n \in \mathbb{N}$. Il suffit bien sûr de choisir une suite décroissante de temps $-n_k < 0$ telle que $\lim_{k \rightarrow \infty} (-n_k) = -\infty$, et de vérifier successivement, pour $k = 1, 2, \dots$, si $F_{-n_k}^0$ est constante. On peut montrer que le choix $-n_k = -2^k$ est proche du choix optimal.

La dynamique sous-jacente. Bien entendu, un pas de temps dans l'algorithme ci-dessus ne correspond pas nécessairement à l'application d'un pas de la chaîne de Markov sous-jacente. Par exemple, dans le cas de l'échantillonneur de Gibbs, un pas de la chaîne de Markov ne modifie la configuration qu'en au plus un sommet, ce qui ne favorise guère la fusion des trajectoires. Il est évidemment beaucoup plus judicieux d'effectuer un nombre suffisant de pas de la chaîne de Markov pour chaque pas de l'algorithme ci-dessus.

Il est également important de bien choisir cette dynamique sous-jacente. Plus elle converge rapidement, plus l'algorithme de couplage depuis le passé s'arrêtera rapidement. Par exemple, appliquée au modèle d'Ising, l'échantillonneur de Gibbs converge rapidement pour de petites valeurs du paramètre β , mais la vitesse de convergence devient catastrophique lorsque β est grand. Heureusement, il existe des chaînes de Markov simulant ce modèle beaucoup plus efficacement (par exemple, l'algorithme de Swendsen–Wang [24]).

Couplage indépendant. Le couplage utilisé ci-dessus est tel que $f_0(i)$ est choisie comme étant égale à j avec probabilité p_{ij} indépendamment pour chaque $i \in S$. Le choix d'un tel couplage indépendant entre les différentes trajectoires jusqu'à leur rencontre n'est bien entendu pas le seul possible. En fait, on peut en général faire beaucoup mieux, en choisissant un couplage qui favorise une fusion plus rapide du processus. On en verra un exemple ci-dessous.

Couplage monotone. L'approche esquissée ci-dessus présente une faiblesse évidente : la nécessité de considérer des chaînes partant de chaque état de S peut sembler rendre cette approche inutilisable dans les situations les plus intéressantes pratiquement. Après tout, pour le modèle d'Ising sur le graphe avec sommets $V = \{1, \dots, 1000\}^2$, le nombre de configurations est déjà de $2^{1000000} \simeq 10^{301030}$. Ceci rend évidemment totalement impossible la simulation de $|S|$ chaînes de Markov simultanément ! Heureusement, dans de nombreuses situations importantes il est possible d'utiliser des propriétés d'ordre afin d'éliminer cette difficulté. Décrivons le principe dans le cas de l'échantillonneur de Gibbs appliqué au modèle d'Ising.

L'ensemble S des configurations du modèle d'Ising sur un graphe fini $G = (V, A)$ possède un ordre partiel naturel : $\sigma \leq \tilde{\sigma}$ si et seulement si $\sigma_i \leq \tilde{\sigma}_i$ pour tout $i \in V$. On dit qu'une fonction $f : S \rightarrow \mathbb{R}$ est **croissante** si

$$\forall \sigma, \tilde{\sigma} \in S, \quad \sigma \leq \tilde{\sigma} \implies f(\sigma) \leq f(\tilde{\sigma}).$$

On introduit le couplage suivant : on tire un nombre u uniformément dans $[0, 1]$ et un site i uniformément dans V , indépendamment, et on définit $f_0(\sigma)$ comme dans (8.2), c'est-à-dire que l'on pose $f_0(\sigma) = \sigma^{i+}$ avec probabilité

$$\left(1 + \exp\left[-2\beta \sum_{j \sim i} \sigma_j\right]\right)^{-1}$$

et $f_0(\sigma) = \sigma^{i-}$ sinon.

Remarquez que ce couplage n'est pas du tout indépendant, puisque l'on utilise les mêmes i et u pour toute configuration initiale σ . En observant que la fonction

$$\sigma \mapsto \left(1 + \exp\left[-2\beta \sum_{j \sim i} \sigma_j\right]\right)^{-1}$$

est croissante, on constate que ce couplage possède la propriété remarquable suivante :

$$\forall \sigma, \tilde{\sigma} \in S, \quad \sigma \leq \tilde{\sigma} \implies f_0(\sigma) \leq f_0(\tilde{\sigma}).$$

Un tel couplage est dit **monotone**.

L'intérêt d'un tel couplage est qu'il suffit de considérer l'évolution de deux trajectoires seulement : celles partant des configurations constantes $\sigma_+ \equiv 1$ et $\sigma_- \equiv -1$. En effet, comme toute trajectoire partant d'une autre configuration va toujours être prise en sandwich entre ces deux trajectoires-ci, ce seront toujours les dernières à fusionner.

Observons que la preuve du Théorème 8.6 reste valide pour cette variante. En effet, la seule chose à vérifier est que le Lemme 8.5 est encore vrai, mais ceci est encore plus simple à vérifier dans ce contexte : il suffit de réaliser que la fusion de toutes les trajectoires a forcément lieu au plus tard la première fois que la copie de la chaîne partant de σ_+ atteint la configuration σ_- , ce qui se produit presque-sûrement en temps fini, par irréductibilité de la chaîne de Markov.

Remarques bibliographiques : Il y a de très nombreuses sources d'information sur ce sujet, dont les livres [5, 11].

Bibliographie

- [1] N. Alon and J. H. Spencer. *The probabilistic method*. Wiley Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, fourth edition, 2016.
- [2] R. Arratia, L. Goldstein, and L. Gordon. Two moments suffice for Poisson approximations : the Chen-Stein method. *Ann. Probab.*, 17(1) :9–25, 1989.
- [3] J. Baik, P. Deift, and K. Johansson. On the distribution of the length of the longest increasing subsequence of random permutations. *Journal of the American Mathematical Society*, 12(4) :1119–1178, 1999.
- [4] A. A. Barker. Monte Carlo calculations of the radial distribution functions for a proton–electron plasma. *Aust. J. Phys.*, 18(2) :119–134, 1965.
- [5] P. Brémaud. *Discrete probability models and methods*, volume 78 of *Probability Theory and Stochastic Modelling*. Springer, Cham, 2017.
- [6] P. Brémaud. *Markov chains—Gibbs fields, Monte Carlo simulation and queues*, volume 31 of *Texts in Applied Mathematics*. Springer, Cham, 2020.
- [7] N. Curien. Random walks and graphs, 2020. Notes de cours, disponibles à l’adresse <https://www.imo.universite-paris-saclay.fr/~nicolas.curien/enseignement.html>.
- [8] P. R. de Montmort. *Essay d’analyse sur les jeux de hazard*. J. Quillau, 1713.
- [9] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6 :290–297, 1959.
- [10] E. N. Gilbert. Random graphs. *Ann. Math. Statist.*, 30 :1141–1144, 1959.
- [11] O. Häggström. *Finite Markov chains and algorithmic applications*, volume 52 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 2002.
- [12] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1) :97–109, 1970.
- [13] H. Kesten. Aspects of first passage percolation. In *École d’été de probabilités de Saint-Flour, XIV—1984*, volume 1180 of *Lecture Notes in Math.*, pages 125–264. Springer, Berlin, 1986.
- [14] M. Krivelevich and B. Sudakov. The phase transition in random graphs : a simple proof. *Random Structures Algorithms*, 43(2) :131–138, 2013.
- [15] K. G. Larsen and J. Nelson. Optimality of the Johnson-Lindenstrauss lemma. In *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2017*, pages 633–638. IEEE Computer Soc., Los Alamitos, CA, 2017.

- [16] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.
- [17] D. A. Levin and Y. Peres. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2017.
- [18] T. Lindvall. *Lectures on the coupling method*. Dover Publications, Inc., Mineola, NY, 2002. Corrected reprint of the 1992 original.
- [19] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.*, 21(6) :1087–1092, 12 2004.
- [20] N. Metropolis and S. Ulam. The Monte Carlo method. *J. Amer. Statist. Assoc.*, 44 :335–341, 1949.
- [21] J. G. Propp and D. B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. In *Proceedings of the Seventh International Conference on Random Structures and Algorithms (Atlanta, GA, 1995)*, volume 9, pages 223–252, 1996.
- [22] S. Roch. Modern discrete probability : An essential toolkit, 2023. Notes de cours, disponibles à l'adresse <https://people.math.wisc.edu/~roch/mdp/>.
- [23] D. Romik. *The surprising mathematics of longest increasing subsequences*, volume 4 of *Institute of Mathematical Statistics Textbooks*. Cambridge University Press, New York, 2015.
- [24] R. H. Swendsen and J.-S. Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.*, 58 :86–88, Jan 1987.
- [25] M. Talagrand. A new look at independence. *Ann. Probab.*, 24(1) :1–34, 1996.
- [26] H. Thorisson. *Coupling, stationarity, and regeneration*. Probability and its Applications (New York). Springer-Verlag, New York, 2000.

Notations

Ensembles, objets

K_n	le graphe complet à n sommets
$\llbracket n \rrbracket$	l'ensemble $\{1, \dots, n\}$
$\mathbb{Z}_{\geq \ell}$	l'ensemble $\{k \in \mathbb{Z} \mid k \geq \ell\}$
$\mathbb{Z}_{> \ell}$	l'ensemble $\{k \in \mathbb{Z} \mid k > \ell\}$
$\mathbb{Z}_{\leq \ell}$	l'ensemble $\{k \in \mathbb{Z} \mid k \leq \ell\}$
$\mathbb{Z}_{< \ell}$	l'ensemble $\{k \in \mathbb{Z} \mid k < \ell\}$
\mathbb{N}	l'ensemble $\mathbb{Z}_{\geq 0}$ des entiers positifs
\mathbb{N}^*	l'ensemble $\mathbb{Z}_{> 0}$ des entiers strictement positifs
\mathbb{R}_+	l'ensemble $\{x \in \mathbb{R} \mid x \geq 0\}$ des réels positifs
\mathbb{R}_+^*	l'ensemble $\{x \in \mathbb{R} \mid x > 0\}$ des réels strictement positifs
$\mathcal{M}_1(\Omega)$	l'ensemble des mesures de probabilité sur Ω

Lois de probabilité

$G(n, p)$	graphe d'Erdős-Rényi avec n sommets et probabilité de rétention p
$\text{Bern}(p)$	loi de Bernoulli de paramètre p
$\text{Binom}(n, p)$	loi binomiale de paramètres n et p
$\mathcal{N}(\mu, \sigma^2)$	loi normale de paramètres μ et σ^2
$\text{Poisson}(\lambda)$	loi de Poisson de paramètre λ
p_λ	mesure de probabilité associée à la loi de Poisson de paramètre λ

Divers

$a := b$	a est défini comme étant égal à b
$A \subset B$	A est un sous-ensemble de B (le cas $A = B$ est possible)
$A \subsetneq B$	A est un sous-ensemble strict de B
$\ \mu - \nu\ _{\text{VT}}$	distance en variation totale entre les mesures μ et ν

i	$\sqrt{-1}$
$\mathcal{L}(X)$	loi de la variable aléatoire X
$\mu(i)$	notation alternative pour $\mu(\{i\})$ lorsque $\mu \in \mathcal{M}_1(\Omega)$, Ω dénombrable et $i \in \Omega$
F_X	fonction de répartition de la variable aléatoire $X : F_X(x) := \mathbb{P}(X \leq x)$
φ_X	fonction caractéristique de $X : \varphi_X(t) := \mathbb{E}[e^{itX}]$
$X \stackrel{\text{loi}}{=} Y$	X a la même loi que Y
$X \stackrel{\text{loi}}{=} Y \mid A$	X a la même loi que Y conditionnée par l'événement A
$a \vee b$	$\max\{a, b\}$
$a \wedge b$	$\min\{a, b\}$
$\ \cdot\ _1$	la norme ℓ^1 ou L^1
$\ \cdot\ _2$	la norme ℓ^2 ou L^2
$\ \cdot\ _\infty$	la norme ℓ^∞ ou L^∞

Index

A

algorithmes

- couplage depuis le passé 70
- de Barker 68
- de Metropolis–Hastings 68
- échantillonneur de Gibbs 68
- méthode de Monte Carlo 67
- parcours en profondeur 3

apériodique

- loi apériodique 15
- loi fortement apériodique 14

C

configuration 68

couplage 10

- couplage maximal 11
- couplage
 - indépendant 10
 - monotone 10

couplage monotone 74

D

dérangement 25

distance

- de Hamming 29
- en variation totale 59

distance en variation totale 9

F

fonction certifiable 37

fonction de partition 68

fonction lipschitzienne 34

fonctions aléatoires itérées 70

forme de Dirichlet 62

G

graphe aléatoire d’Erdős–Rényi .. 1, 10, 26

H

hamiltonien 68

harmonique 49

I

inégalité de Talagrand 31

isométrie restreinte 44

L

laplacien (discret) 62

lemme de Johnson–Lindenstrauss 42

loi des petits nombres 19

M

marche aléatoire

- paresseuse 16
- sur un graphe 65

médiane et espérance 35

mesure de Gibbs 69

modèle d’Ising 68, 74

P

percolation de premier passage 37

plus longue sous-suite croissante 39

problème des anniversaires 23

problème des rencontres 25

S

- sparse** 44
superharmonique 49

T

- temps**
 de mélange 60
 de relaxation 61
température inverse 69
transformée de Doob 49
transition de phase
 graphe aléatoire d'Erdős-Rényi 1
trou spectral 58