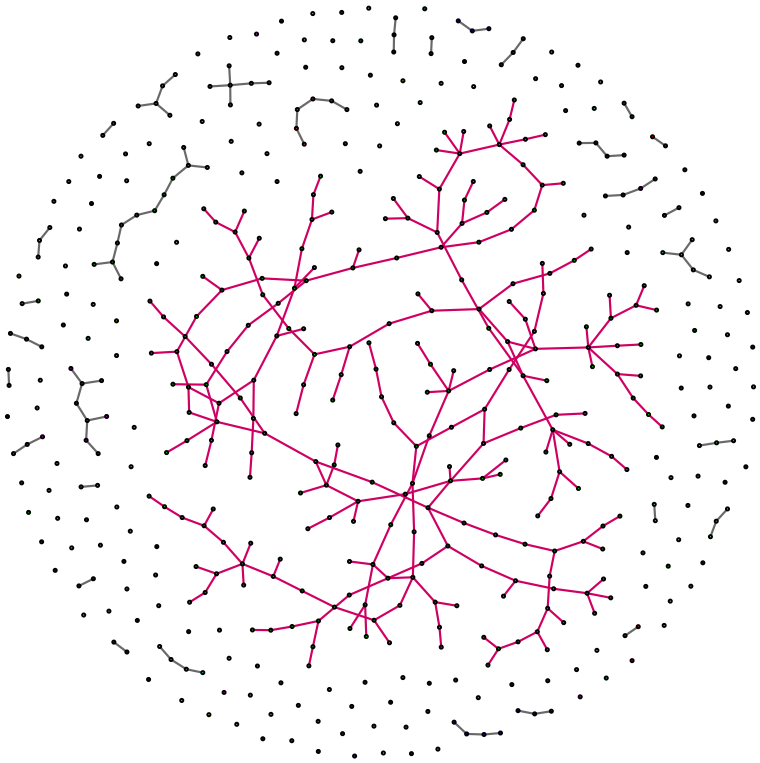
Chapitres Choisis de Théorie des Probabilités

— Yvan Velenik —



UNIVERSITÉ DE GENÈVE

— Version du 20 janvier 2026 —

Dernière version téléchargeable à l'adresse

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

Table des matières

1 La loi forte des grands nombres sous hypothèse de faible dépendance

1.1 La problématique

Dans le cours d'introduction à la théorie des probabilités, vous avez vu une version de la loi forte des grands nombres, valide pour des suites de variables aléatoires intégrables indépendantes. Dans ce chapitre, nous allons nous intéresser à une extension de ce résultat à des suites de variables aléatoires non indépendantes, mais dont les covariances décroissent suffisamment rapidement. Il existe diverses façons d'étendre la loi forte des grands nombres à des situations de dépendance. L'approche suivie dans ce chapitre a le triple avantage d'être directe, de ne reposer que sur des outils élémentaires (inégalité de Bienaymé-Tchebychev et lemme de Borel–Cantelli) et de fournir une vitesse de convergence. L'idée sous-tendant cette approche trouve ses racines dans des méthodes développées en théorie quantique des champs constructive.

Soit $(X_k)_{k\geqslant 1}$ une suite de variables aléatoires réelles centrées sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

Théorème 1.1. *On suppose qu'il existe $\gamma > 0$ et $K > 0$ tels que*

$$\mathbb{E}[X_m X_n] \leqslant \frac{K}{(1+|n-m|)^\gamma} \quad \textit{pour tout } m, n \geqslant 1.$$

Alors,

$$\mathbb{P}\big(\exists C, N_0 \textit{ telle que } \big|\frac{1}{N}\sum_{n=1}^{N} X_n\big| \leqslant C N^{\beta-1}\log N \textit{ pour tout } N \geqslant N_0\big) = 1,$$

pour tout $\beta \in (\frac{1}{2}, 1)$ si $\gamma \geqslant 1$ et tout $\beta \in (1-\frac{\gamma}{2}, 1)$ si $\gamma \in (0,1)$.

Remarque 1.2. *Observez que la condition sur les covariances implique que les variances des variables aléatoires $(X_k)_{k\geqslant 1}$ sont uniformément bornées. De telles conditions allant au-delà de l'intégrabilité sont nécessaires si l'on veut obtenir des taux de convergence explicites.* ◇

1.2 Preuve du Théorème 1.1

Étant donné $k \in \mathbb{N}$ et $\ell \in \mathbb{N}^*$, on définit le bloc $B_{k,\ell}$ par

$$B_{k,\ell} := \big\{ n \in \mathbb{N}^* \,\big|\, (\ell-1)2^k < n \leqslant \ell 2^k \big\}.$$

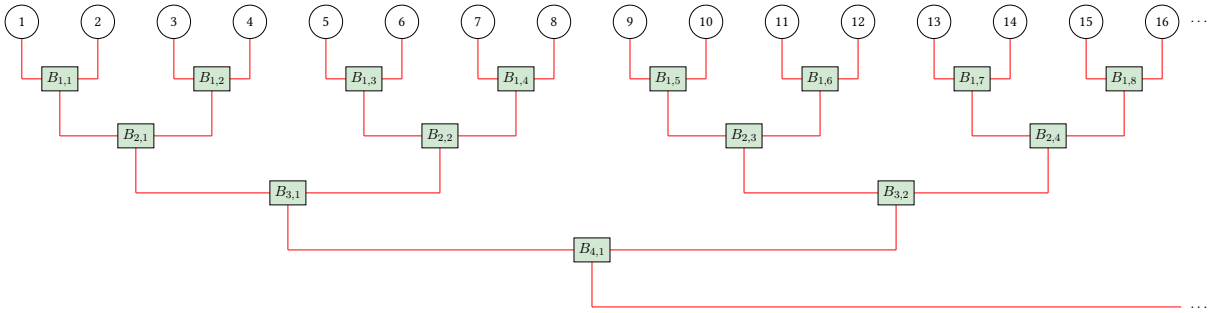


FIGURE 1.1: La décomposition hiérarchique de l'ensemble $\mathbb{N}^*$.

Les blocs $B_{k,\ell}$, $k \geqslant 0, \ell \geqslant 1$, fournissent une décomposition hiérarchique de l'ensemble $\mathbb{N}^*$, *cf.* Fig. 1.1. Fixons $\beta \in (\frac{1}{2}, 1)$ si $\gamma \geqslant 1$ et $\beta \in (1 - \frac{\gamma}{2}, 1)$ si $\gamma \in (0, 1)$. On distinguera deux types de blocs selon la valeur prises par les variables aléatoires X_n, $n \geqslant 1$: $B_{k,\ell}$ est un *mauvais bloc* si

$$\Big| \sum_{n \in B_{k,\ell}} X_n \Big| \geqslant 2^{\beta k} \ell^{\beta},$$

et un *bon bloc* dans le cas contraire.

Pour $\gamma \in \mathbb{R}_+^*$, on définit la fonction croissante $p_\gamma : [1, \infty) \to (0, \infty)$ par

$$p_\gamma(x) := \begin{cases} 1/(\gamma - 1) & \text{si } \gamma > 1, \\ \log x & \text{si } \gamma = 1, \\ x^{1-\gamma}/(1-\gamma) & \text{si } \gamma < 1. \end{cases}$$

Le lemme suivant borne la probabilité qu'un bloc donné soit mauvais.

Lemme 1.3. *Pour tous les $k \geqslant 0$, $\ell \geqslant 1$,*

$$\mathbb{P}(B_{k,\ell} \textit{ est un mauvais bloc}) \leqslant K \ell^{-2\beta} 2^{(1-2\beta)k} \big(1 + 2 p_\gamma(2^k)\big).$$

Démonstration. Par l'inégalité de Bienaymé–Tchebychev,

$$\mathbb{P}(B_{k,\ell} \text{ est un mauvais bloc}) = \mathbb{P}\Big(\Big| \sum_{n \in B_{k,\ell}} X_n \Big| \geqslant 2^{\beta k} \ell^{\beta} \Big) \leqslant \frac{\operatorname{Var}\big[\sum_{n \in B_{k,\ell}} X_n\big]}{2^{2\beta k} \ell^{2\beta}}.$$

Il suit des hypothèses que

$$\operatorname{Var}\Big[\sum_{n \in B_{k,\ell}} X_n \Big] = \sum_{n,m \in B_{k,\ell}} \mathbb{E}[X_m X_n] \leqslant K \sum_{n,m \in B_{k,\ell}} (1 + |n - m|)^{-\gamma}.$$

À présent,

$$\sum_{n,m \in B_{k,\ell}} (1 + |n - m|)^{-\gamma} \leqslant 2^k + 2 \sum_{n \in B_{k,\ell}} \sum_{\substack{m \in B_{k,\ell} \\ m > n}} (1 + |n - m|)^{-\gamma}.$$

Pour tout $n \in B_{k,\ell}$, on a la borne

$$\sum_{\substack{m \in B_{k,\ell} \\ m > n}} (1 + |n - m|)^{-\gamma} \leqslant \sum_{r=1}^{2^k - 1} (1 + r)^{-\gamma} \leqslant \int_0^{2^k - 1} (1 + x)^{-\gamma} \mathrm{d}x \leqslant p_\gamma(2^k),$$

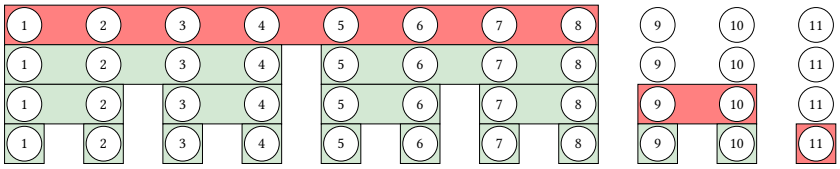


FIGURE 1.2: La partition $\{1,2,3,4,5,6,7,8,9,10,11\} = B_{3,1} \cup B_{1,5} \cup B_{0,11}$ (les blocs retenus sont en rouge).

où la dernière inégalité s'obtient en traitant séparément les 3 cas $\gamma < 1$, $\gamma = 1$ et $\gamma > 1$. Par conséquent,

$$\operatorname{Var}\Big[\sum_{n\in B_{k,\ell}} X_n\Big] \leqslant K2^k\big(1 + 2p_\gamma(2^k)\big),$$

et la conclusion suit. □

Soit $F \coloneqq \{(k,\ell) \in \mathbb{N}\times\mathbb{N}^* \mid B_{k,\ell} \text{ est un mauvais bloc}\}$ l'ensemble des indices de tous les mauvais blocs. Observez que $\beta > \frac{1}{2}$ pour tout γ, et que $2 - 2\beta - \gamma \leqslant -\frac{1}{2}\gamma < 0$ lorsque $\gamma < 1$. La borne fournie par le lemme précédent est donc sommable à la fois en ℓ et en k, et le lemme de Borel–Cantelli implique que l'ensemble F est presque-sûrement fini.

Sans perte de généralité, on suppose donc F fini et on pose $N_F = \max \bigcup_{(k,\ell)\in F} B_{k,\ell} \in \{-\infty\} \cup \mathbb{N}$. Le Théorème 1.1 est alors une conséquence immédiate de l'observation (déterministe) suivante.

Lemme 1.4. *Si $N \geqslant 4N_F - 1$, alors*

$$\Big|\sum_{n=1}^{N} X_n\Big| \leqslant \Big(\frac{\log N}{\log 2} + 1\Big) N^\beta.$$

Démonstration. Observons qu'il existe une unique décomposition $N = 2^{k_1} + 2^{k_2} + \cdots + 2^{k_r}$ avec $k_1 > k_2 > \cdots > k_r \geqslant 0$. Ceci induit une décomposition de $\{1,\dots,N\}$ comme union disjointe de blocs (*cf.* Fig. 1.2) : $\{1,\dots,N\} = B_{k_1,\ell_1} \cup B_{k_2,\ell_2} \cup \cdots \cup B_{k_r,\ell_r}$, où

$$\ell_1 = 1, \quad \ell_2 = 2^{k_1-k_2} + 1, \quad \ell_3 = 2^{k_1-k_3} + 2^{k_2-k_3} + 1, \quad \dots, \quad \ell_r = 2^{k_1-k_r} + \cdots + 2^{k_{r-1}-k_r} + 1.$$

Si tous les blocs de cette décomposition sont bons, il suit de la construction ci-dessus que

$$\Big|\sum_{n=1}^{N} X_n\Big| = \Big|\sum_{i=1}^{r}\sum_{n\in B_{k_i,\ell_i}} X_n\Big| \leqslant \sum_{i=1}^{r}\Big|\sum_{n\in B_{k_i,\ell_i}} X_n\Big| \leqslant \sum_{i=1}^{r} 2^{\beta k_i}\ell_i^\beta = \sum_{i=1}^{r}\big(2^{k_1} + \cdots + 2^{k_i}\big)^\beta \leqslant rN^\beta.$$

De plus, puisque $k_1 > k_2 > \cdots > k_r \geqslant 0$, on a $r \leqslant k_1 + 1$. Mais $2^{k_1} \leqslant N$, ce qui implique que $r \leqslant \log N/\log 2 + 1$. Il nous suffit donc de montrer que, sous l'hypothèse $N \geqslant 4N_F - 1$, tous les blocs de la décomposition sont nécessairement bons.

Il n'y a évidemment rien à montrer si $F = \varnothing$. On suppose donc $F \neq \varnothing$ fini. Dans ce cas, $1 \leqslant N_F < \infty$. Notons $s = \min\{m \in \mathbb{N} \mid 2^m \geqslant N_F\}$. Par définition de N_F, tous les mauvais blocs sont inclus dans $B_{s,1}$. Par définition de s et de k_1, on a $2^{s-1} < N_F \leqslant \frac{1}{4}(N+1)$ et $2^{k_1+1} \geqslant N+1$. Il s'ensuit que $2^{s-1} < \frac{1}{4}2^{k_1+1} = 2^{k_1-1}$ et donc que $s < k_1$. Ceci montre que, sous l'hypothèse $N \geqslant 4N_F - 1$, tous les mauvais blocs sont des sous-ensembles stricts de B_{k_1,ℓ_1} et, par conséquent, qu'aucun des blocs de la partition n'est mauvais. □

Remarques bibliographiques : L'approche suivie dans ce chapitre a été introduite dans [1].

2 Le théorème limite local

2.1 La problématique

Soit $X_1, X_2, \dots$ une suite de variables aléatoires i.i.d. à valeurs dans $\mathbb{Z}$, d'espérance μ et de variance $0 < \sigma^2 < \infty$. Notons $S_n \coloneqq \sum_{i=1}^n X_i$ et $\widetilde{S}_n \coloneqq \frac{1}{\sqrt{\sigma^2 n}}(S_n - n\mu)$. Le Théorème Central Limite affirme que, lorsque $n \to \infty$,

$$\mathbb{P}\big(\widetilde{S}_n \in [a,b]\big) = (1+o(1))\frac{1}{\sqrt{2\pi}}\int_a^b e^{-t^2/2}\mathrm{d}t,$$

pour tout $-\infty \leqslant a < b \leqslant \infty$. Observez que ce résultat ne permet pas, *a priori*, de prendre pour a et b des fonctions de n et ainsi ne permet de décrire que les fluctuations typiques autour de la moyenne. D'autre part, même dans ce régime, il ne fournit aucune information sur la probabilité que S_n prenne précisément une valeur x donnée. Nous allons voir comment remédier à ces deux limitations dans le reste de ce chapitre.

Nous noterons $\phi_X(t) \coloneqq \mathbb{E}[e^{\mathrm{i}tX}]$ la fonction caractéristique associée à la variable aléatoire X. Pour $x \in \mathbb{C}$ et $\delta \in \mathbb{R}_+^*$, nous écrirons $B_\delta(x) \coloneqq \big\{z \in \mathbb{C} \,\big|\, |z-x| < \delta\big\}$, avec la convention que $B_\delta \coloneqq B_\delta(0)$.

Dans la suite de ce chapitre, nous considérerons des variables aléatoires possédant les propriétés suivantes :

- H_1. $\mathbb{E}[X] = \mu$, $\mathrm{Var}[X] = \sigma^2 \in (0,\infty)$;
- H_2. $|\phi_X(t)| < 1, \forall t \in [-\pi,\pi] \setminus \{0\}$;
- H_3. il existe $\delta_1 > 0$ tel que $F(z) \coloneqq \mathbb{E}[e^{zX}]$ est analytique dans B_{δ_1}.

Ces hypothèses, en fait très naturelles, sont discutées ci-dessous, dans la Section 2.3.

2.2 Énoncé du théorème limite local

Les fonctions suivantes jouent un rôle essentiel dans ce chapitre :

$$\mathbb{F}_n(z) \coloneqq \mathbb{E}[e^{zS_n}] = \big(\mathbb{E}[e^{zX_1}]\big)^n = F(z)^n,$$
$$\mathbb{H}_n(z) \coloneqq \log \mathbb{F}_n(z) = n \log F(z) =: nH(z).$$

Les fonctions F et $\mathbb{F}_n$ sont respectivement appelées **fonctions génératrices des moments** de X_1 et de S_n ; les fonctions H et $\mathbb{H}_n$ sont les **fonctions génératrices des cumulants** de X_1 et S_n, respectivement [1].

1. En anglais, ces quantités sont respectivement appelées **moment generating function** et **log-moment generating function**.

L'hypothèse H_3 ci-dessus implique que $\mathbb{F}_n$ est analytique dans B_{δ_1}. Étant donné que $\mathbb{F}(0) = 1$, il suit qu'il existe $\delta_2 \in (0, \delta_1]$ tel que $\mathbb{F}$ ne s'annule pas sur B_{δ_2} ; $\mathbb{H}_n$ est alors analytique dans B_{δ_2}.

Nous pouvons à présent énoncer le résultat principal de ce chapitre.

Théorème 2.1 (Théorème limite local)**.** *Soit $X_1, X_2, \dots$ une suite de variables aléatoires i.i.d. à valeurs dans $\mathbb{Z}$ satisfaisant aux hypothèses H_1, H_2 et H_3 ci-dessus. Soit $\Delta\colon \mathbb{N} \to \mathbb{R}$ telle que $\lim_{n\to\infty} \Delta(n)/n = 0$.*

Alors, uniformément pour toute suite d'entiers $(x_n)_{n\geqslant 1}$ tels que $|x_n - n\mu| \leqslant \Delta(n)$, on a, lorsque $n \to \infty$,

$$\mathbb{P}(S_n = x_n) = \frac{1}{\sqrt{2\pi\sigma^2 n}} \exp\bigl(-nI(x_n/n)\bigr)\,(1 + \mathsf{o}(1)), \tag{2.1}$$

où[2] *$I(t) := \lambda(t)t - H(\lambda(t))$, avec $\lambda(t)$ l'unique solution de l'équation $H'(\lambda) = t$. I est généralement appelée* ***fonction de taux****.*

En particulier, si $\Delta(n) = \mathsf{o}(n^{2/3})$, le comportement asymptotique est gaussien : uniformément dans les suites $(x_n)_{n\geqslant 1}$ d'entiers satisfaisant $|x_n - n\mu| \leqslant \Delta(n)$,

$$\mathbb{P}(S_n = x_n) = \frac{1}{\sqrt{2\pi\sigma^2 n}} \exp\Bigl(-\frac{(x_n - n\mu)^2}{2\sigma^2 n}\Bigr)\,(1 + \mathsf{o}(1)). \tag{2.2}$$

Le corollaire suivant se révèle également utile.

Corollaire 2.2. *Soit $X_1, X_2, \dots$ une suite de variables aléatoires i.i.d. à valeurs dans $\mathbb{Z}$ satisfaisant aux hypothèses H_1, H_2 et H_3. Alors, pour toute suite $(k_n)_{n\geqslant 1}$ de nombres positifs telle que $\lim_{n\to\infty} k_n = \infty$ et $\lim_{n\to\infty} k_n n^{-1/6} = 0$, on a*

$$\mathbb{P}(\widetilde{S}_n \geqslant k_n) = (1 + \mathsf{o}(1))\,\frac{1}{\sqrt{2\pi}k_n} e^{-\frac{1}{2}k_n^2}$$

lorsque $n \to \infty$.

Démonstration. Par l'inégalité de Chernoff[3],

$$\mathbb{P}(\widetilde{S}_n > 2k_n) = \mathbb{P}(S_n > n\mu + 2k_n\sqrt{\sigma^2 n}) \leqslant \exp\Bigl\{-n \sup_{t\geqslant 0}\Bigl(\mu t + \frac{2\sigma k_n}{\sqrt{n}}t - H(t)\Bigr)\Bigr\}.$$

On choisit $t = 2k_n/\sqrt{\sigma^2 n}$; observez que $t \in B_{\delta_2}$ lorsque n est suffisamment grand. La fonction H étant analytique dans ce domaine, on a $H(t) = H(0) + H'(0)t + \frac{1}{2}H''(0)t^2 + \mathsf{O}(t^3) = \mu t + \frac{1}{2}\sigma^2 t^2 + \mathsf{O}(t^3)$. Par conséquent,

$$\mu t + \frac{2\sigma k_n}{\sqrt{n}}t - H(t) = \frac{2k_n^2}{n} + \mathsf{O}\Bigl(\frac{k_n^3}{n^{3/2}}\Bigr) = \frac{2 - \mathsf{o}(1)}{n}k_n^2.$$

Ainsi, $\mathbb{P}(\widetilde{S}_n > 2k_n) = \mathsf{o}\bigl(\frac{1}{k_n}e^{-\frac{1}{2}k_n^2}\bigr)$, lorsque $n \to \infty$, et il suffit donc d'estimer

$$\mathbb{P}(2k_n \geqslant \widetilde{S}_n \geqslant k_n).$$

2. En d'autres termes, I est la transformée de Legendre de H, $I(t) = \sup_{\lambda\in\mathbb{R}}\bigl(\lambda t - H(\lambda)\bigr)$.
3. Rappelons que cette dernière affirme que, pour tout $x \in \mathbb{R}$ et toute variable aléatoire réelle X,

$$\mathbb{P}(X \geqslant x) \leqslant \inf_{t>0} e^{-tx}\mathbb{E}[e^{tX}] = \exp\bigl\{-\sup_{t>0}\bigl(tx - H(t)\bigr)\bigr\}.$$

Notons $J := [n\mu + k_n\sqrt{\sigma^2 n}, n\mu + 2k_n\sqrt{\sigma^2 n}]$. Par hypothèse, $k_n\sqrt{n} = \mathsf{o}(n^{2/3})$. Le Théorème 2.1 implique donc que [4]

$$\begin{aligned}\mathbb{P}(2k_n \geqslant \widetilde{S}_n \geqslant k_n) &= \sum_{x\in\mathbb{Z}\cap J} \mathbb{P}(S_n = x)\\ &= (1+\mathsf{o}(1))\,\frac{1}{\sqrt{2\pi\sigma^2 n}} \sum_{x\in\mathbb{Z}\cap J} e^{-(x-n\mu)^2/2\sigma^2 n}\\ &= (1+\mathsf{o}(1))\,\frac{1}{\sqrt{2\pi}} \int_{k_n}^{2k_n} e^{-t^2/2}\mathrm{d}t + \mathsf{O}(\tfrac{1}{\sqrt{n}}e^{-k_n^2})\\ &= (1+\mathsf{o}(1))\,\frac{1}{\sqrt{2\pi}} \int_{k_n}^{2k_n} e^{-t^2/2}\mathrm{d}t + \mathsf{o}(\tfrac{1}{k_n}e^{-k_n^2}).\end{aligned}$$

La conclusion suit en utilisant les estimées habituelles pour cette dernière intégrale [5]. □

2.3 Brève discussion des hypothèses

Hypothèse H_1. Celle-ci ne nécessite pas de discussion particulière puisqu'elle est déjà nécessaire afin d'avoir un Théorème Central Limite.

Hypothèse H_2. Celle-ci sert à éviter des problèmes de périodicité. Considérons, par exemple, une suite de variables aléatoires i.i.d. $(X_i)_{i\geqslant 1}$ avec $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$. Dans ce cas, on a manifestement $\mathbb{P}(S_{2k+1} = 0) = 0$, pour tout $k \in \mathbb{N}$, et le Théorème 2.1 n'est plus vrai sous la forme énoncée ci-dessus. Le lemme suivant montre que H_2 est précisément la condition requise pour éviter ce type de problèmes.

Lemme 2.3. *Soit X une variable aléatoire à valeurs dans $\mathbb{Z}$. Alors il existe $a \in \mathbb{Z}$ et $2 \leqslant b \in \mathbb{N}$ tels que X soit à valeurs dans $a + b\mathbb{Z}$ si et seulement si H_2 est violée.*

Démonstration. Supposons qu'il existe $a \in \mathbb{Z}$ et $2 \leqslant b \in \mathbb{N}$ tels que $\mathbb{P}(X \notin a + b\mathbb{Z}) = 0$. Alors,

$$\phi_X(t) = \sum_{k=-\infty}^{\infty} \exp\bigl(\mathrm{i}t(a+bk)\bigr)\,\mathbb{P}(X = a+bk).$$

Il suit que $|\phi_X(2\pi/b)| = 1$ et donc que H_2 est violée.

Réciproquement, supposons H_2 violée, c'est-à-dire qu'il existe $0 \neq |\beta| \leqslant 1$ avec $|\phi_X(\beta\pi)| = 1$. Dans ce cas, $\phi_X(\beta\pi) = e^{\mathrm{i}\alpha\pi}$ pour un certain $\alpha \in (-1, 1]$. Par conséquent,

$$1 = e^{-\mathrm{i}\alpha\pi}\phi_X(\beta\pi) = \sum_{k\in\mathbb{Z}} e^{\mathrm{i}\beta\pi k - \mathrm{i}\alpha\pi}\mathbb{P}(X = k) = \sum_{k\in\mathbb{Z}} \cos(\beta\pi k - \alpha\pi)\,\mathbb{P}(X = k).$$

Comme $\sum_{k\in\mathbb{Z}} \mathbb{P}(X = k) = 1$, il suit que $\cos(\beta\pi k - \alpha\pi) = 1$ lorsque $\mathbb{P}(X = k) \neq 0$. Pour ces valeurs de k, on doit donc avoir $\beta k - \alpha \in 2\mathbb{Z}$. Par conséquent, si k_0 et k_1 sont deux points distincts tels que $\mathbb{P}(X = k_0) \neq 0$, $\mathbb{P}(X = k_1) \neq 0$, alors $k_1 - k_0$ est un multiple de $2/\beta \geqslant 2$. □

4. Pour la troisième identité, on utilise l'encadrement (la sommande étant décroissante en $x \in \mathbb{Z}\cap J$)

$$\int_J e^{-(s-n\mu)^2/2\sigma^2 n}\,\mathrm{d}s + \mathsf{O}(e^{-k_n^2/2}) \leqslant \sum_{x\in\mathbb{Z}\cap J} e^{-(x-n\mu)^2/2\sigma^2 n} \leqslant \int_J e^{-(s-n\mu)^2/2\sigma^2 n}\,\mathrm{d}s + \mathsf{O}(e^{-k_n^2/2}).$$

5. On a, pour tout $x > 0$,

$$\frac{e^{-x^2/2}}{x} \geqslant \int_x^\infty e^{-y^2/2}\mathrm{d}y \geqslant (1 - \frac{1}{x^2})\,\frac{e^{-x^2/2}}{x}.$$

En effet, une intégration par parties montre que $\int_x^\infty e^{-y^2/2}\mathrm{d}y = \frac{1}{x}e^{-x^2/2} - \int_x^\infty \frac{1}{y^2}e^{-y^2/2}\mathrm{d}y$. La conclusion suit, puisque $0 \leqslant \int_x^\infty \frac{1}{y^2}e^{-y^2/2}\mathrm{d}y \leqslant \frac{1}{x^3}\int_x^\infty ye^{-y^2/2}\mathrm{d}y = \frac{1}{x^3}e^{-x^2/2}$.

Observons également que l'hypothèse H_2 n'est pas réellement restrictive : si X est une variable aléatoire telle que $\mathbb{P}(X \notin a + b\mathbb{Z}) = 0$, pour certains $a, b \in \mathbb{Z}$ avec $b \geqslant 2$, mais que $\mathbb{P}(X \notin a + b'\mathbb{Z}) > 0$, pour tout $b' > b$, alors la variable aléatoire $Y = (X - a)/b$ satisfait l'hypothèse H_2 et le théorème s'applique.

Hypothèse H_3. Cette hypothèse, appelée **condition de Cramér**, est classique en théorie des grandes déviations, dans laquelle on étudie la probabilité que la moyenne empirique S_n/n fasse une déviation d'ordre 1 loin de son espérance. Comme, ici aussi, nous sommes intéressés à la probabilité que S_n/n dévie loin de son espérance, il est nécessaire d'imposer une condition sur le comportement des queues de la loi des X_i. La condition de Cramér revient à exiger que la loi des X_i décroisse au moins exponentiellement. Sans cette hypothèse, le Théorème 2.1 ne serait plus valide, sous la forme énoncée, dans tout le régime considéré (voir par contre les remarques faites dans la section 2.5). Par exemple, si $\mathbb{P}(X > k) \sim |k|^{-\gamma}$, pour un certain $\gamma > 2$, alors la probabilité de l'événement $\{S_n - \mathbb{E}[S_n] > cn\}$ ne décroît plus que comme une puissance de n, puisqu'il suffit pour réaliser cet événement de faire prendre une grande valeur à l'une des variables X_i. Sous des hypothèses appropriées, on peut montrer que c'est effectivement le mécanisme typique utilisé pour réaliser une telle déviation dans ce contexte.

2.4 Preuve du théorème limite local

Dans cette sous-section, nous démontrons le Théorème 2.1. La preuve se fait en deux étapes.

I. Réduction au cas $\mu = x_n/n$.

L'idée, classique en théorie des grandes déviations, est de déformer la mesure $\mathbb{P}$ en une nouvelle mesure $\mathbb{P}^\lambda$, dépendant d'un paramètre λ choisi de sorte que $\mathbb{E}^\lambda[S_n] = x_n$.

Plus précisément, soit $\mathbb{P}^\lambda$ la mesure de probabilité sous laquelle les variables aléatoires X_k sont indépendantes et identiquement distribuées, avec $\mathbb{P}^\lambda(X_1 = u) = F(\lambda)^{-1} e^{\lambda u} \mathbb{P}(X_1 = u)$. On peut alors écrire

$$\begin{aligned}\mathbb{P}^\lambda(S_n = x_n) &= \sum_{\substack{y_1,\dots,y_n \\ \sum_i y_i = x_n}} \prod_{i=1}^n \mathbb{P}^\lambda(X_i = y_i) \\ &= \sum_{\substack{y_1,\dots,y_n \\ \sum_i y_i = x_n}} \prod_{i=1}^n F(\lambda)^{-1} e^{\lambda y_i} \mathbb{P}(X_i = y_i) = \mathbb{F}_n(\lambda)^{-1} e^{\lambda x_n} \mathbb{P}(S_n = x_n),\end{aligned}$$

c'est-à-dire

$$\mathbb{P}(S_n = x_n) = e^{-\lambda x_n} \mathbb{F}_n(\lambda)\, \mathbb{P}^\lambda(S_n = x_n). \tag{2.3}$$

Nous allons à présent montrer qu'il existe une unique valeur $\lambda_0 = \lambda_0(x_n)$ pour laquelle $\mathbb{E}^{\lambda_0}[X_1] = x_n/n$. On supposera que $x_n/n > \mu$, le cas $x_n/n < \mu$ se traitant de la même manière. Observons tout d'abord que, lorsque $|\lambda| < \delta_2$, $\mathbb{E}^\lambda[X_1] = H'(\lambda)$ et $\mathrm{Var}^\lambda[X_1] = H''(\lambda)$. Posons $\delta_3 := \delta_2/2$ et définissons $\tilde{c} := \min_{t \in [0,\delta_3]} H''(t) > 0$, puisque la variance de $\mathbb{P}^t$ n'est pas dégénérée[6] pour tout $t \in B_{\delta_2}$. Ainsi,

$$H'(\delta_3) = H'(0) + \int_0^{\delta_3} H''(t)\,\mathrm{d}t \geqslant \mu + \tilde{c}\delta_3 \geqslant \frac{x_n}{n} - \frac{\Delta(n)}{n} + \tilde{c}\delta_3 > \frac{x_n}{n} + \tfrac{1}{4}\tilde{c}\delta_2,$$

pour tout n suffisamment grand pour que $\Delta(n)/n < \frac{1}{4}\tilde{c}\delta_2$. Comme $H'(0) = \mu < x_n/n$, il suit du théorème des valeurs intermédiaires qu'il existe $\lambda_0 \in (0, \delta_3)$ tel que $H'(\lambda_0) = x_n/n$. De plus, il y a un unique tel λ_0, puisque H est convexe sur son ensemble de définition et strictement convexe sur $[0, \delta_3]$.

6. En effet, $\mathrm{Var}[X_1] > 0$ implique qu'il existe $x, y \in \mathbb{Z}$, $x \neq y$, tels que $\mathbb{P}(X_1 = x) > 0$ et $\mathbb{P}(X_1 = y) > 0$. Mais cela implique que $\mathbb{P}^t(X_1 = x) > 0$ et $\mathbb{P}^t(X_1 = y) > 0$, et donc $\mathrm{Var}^t[X_1] > 0$.

De plus, de la même façon,

$$\frac{x_n}{n} = H'(\lambda_0) = H'(0) + \int_0^{\lambda_0} H''(t)\,\mathrm{d}t \geqslant \mu + \tilde{c}\lambda_0,$$

ce qui implique que (la valeur absolue est nécessaire pour couvrir le cas où $x_n/n < \mu$)

$$|\lambda_0| \leqslant \frac{1}{\tilde{c}} \left| \frac{x_n}{n} - \mu \right| \leqslant \frac{\Delta(n)}{\tilde{c}n}. \tag{2.4}$$

Finalement,

$$\frac{x_n}{n} = H'(\lambda_0) = H'(0) + H''(0)\lambda_0 + \mathsf{O}(\lambda_0^2) = \mu + \sigma^2\lambda_0 + \mathsf{O}(\Delta(n)^2/n^2),$$

d'où l'on tire

$$\lambda_0 = \frac{1}{\sigma^2}\Big(\frac{x_n}{n} - \mu\Big) + \mathsf{O}(\Delta(n)^2/n^2). \tag{2.5}$$

En choisissant $\lambda = \lambda_0$, (2.3) devient

$$\begin{aligned}
\mathbb{P}(S_n = x_n) &= e^{-\lambda_0 x_n}\mathbb{F}_n(\lambda_0)\,\mathbb{P}^{\lambda_0}(S_n = x_n) \\
&= \exp\big(-(\lambda_0 x_n - \mathbb{H}_n(\lambda_0))\big)\,\mathbb{P}^{\lambda_0}(S_n = x_n) \\
&= \exp\big(-nI(x_n/n)\big)\,\mathbb{P}^{\lambda_0}(S_n = x_n).
\end{aligned}$$

En particulier, si $\Delta(n) = \mathsf{o}(n^{2/3})$, (2.4) montre que $\lambda_0 = \mathsf{o}(n^{-1/3})$. On peut donc utiliser l'analyticité de H et (2.5) afin d'obtenir

$$\begin{aligned}
nI(x_n/n) &= \lambda_0 x_n - \mathbb{H}_n(\lambda_0) \\
&= n\,\lambda_0\frac{x_n}{n} - n\,H'(0)\lambda_0 - n\tfrac{1}{2}H''(0)\lambda_0^2 + \mathsf{O}(n\lambda_0^3) \\
&= n\,\lambda_0\frac{x_n}{n} - n\,\mu\lambda_0 - n\tfrac{1}{2}\sigma^2\lambda_0^2 + \mathsf{O}(n\lambda_0^3) \\
&= n\,\lambda_0(\frac{x_n}{n} - \mu - \tfrac{1}{2}\sigma^2\lambda_0) + \mathsf{O}(n\lambda_0^3) \\
&= \frac{1}{2\sigma^2 n}(x_n - n\mu)^2 + \mathsf{O}(n\lambda_0^3) \\
&= \frac{1}{2\sigma^2 n}(x_n - n\mu)^2 + \mathsf{o}(1),
\end{aligned} \tag{2.6}$$

ce qui montre que (2.1) implique (2.2) dans ce régime.

II. Comportement asymptotique de $\mathbb{P}^{\lambda_0}(S_n = x_n)$.

Notre but est à présent d'établir que

$$\mathbb{P}^{\lambda_0}(S_n = x_n) = \frac{1}{\sqrt{2\pi\sigma^2_{\lambda_0} n}}\,(1 + \mathsf{o}(1)), \tag{2.7}$$

où

$$\sigma^2_{\lambda_0} = \mathrm{Var}^{\lambda_0}[X_1] = H''(\lambda_0) = H''(0) + \mathsf{O}(\lambda_0) = \sigma^2 + \mathsf{o}(1).$$

Les affirmations (2.1) et (2.2) suivront alors immédiatement.

Observez que la fonction caractéristique de S_n sous $\mathbb{P}^{\lambda_0}$ peut s'écrire

$$\begin{aligned}
\phi^{\lambda_0}_{S_n}(t) := \mathbb{E}^{\lambda_0}(e^{\mathrm{i}tS_n}) &= \sum_{y\in\mathbb{Z}} e^{\mathrm{i}ty}\mathbb{P}^{\lambda_0}(S_n = y) \\
&= \sum_{y\in\mathbb{Z}} e^{(\mathrm{i}t+\lambda_0)y}\frac{1}{\mathbb{F}_n(\lambda_0)}\mathbb{P}(S_n = y) = \frac{\mathbb{F}_n(\mathrm{i}t+\lambda_0)}{\mathbb{F}_n(\lambda_0)} = e^{nH_{\lambda_0}(\mathrm{i}t)},
\end{aligned}$$

et l'on a introduit $H_{\lambda_0}(z) \coloneqq H(\lambda_0 + z) - H(\lambda_0)$.

On a donc, en utilisant le théorème d'inversion,

$$\mathbb{P}^{\lambda_0}(S_n = x_n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-\mathrm{i}tx_n} \phi_{S_n}^{\lambda_0}(t)\, \mathrm{d}t. \tag{2.8}$$

Soient $\epsilon \in (0, \frac{1}{6})$ et $\delta > 0$ suffisamment petit (les conditions sur δ sont données plus bas). On partitionne le domaine d'intégration en trois régions :

$$[-\pi, \pi] = A_\epsilon \cup A_{\epsilon,\delta} \cup A_\delta,$$

où

$$\begin{aligned} A_\epsilon &\coloneqq \big\{ t \in [-\pi, \pi] \,\big|\, |t| < n^{-1/2+\epsilon} \big\}, \\ A_{\epsilon,\delta} &\coloneqq \big\{ t \in [-\pi, \pi] \,\big|\, n^{-1/2+\epsilon} \leqslant |t| < \delta \big\}, \\ A_\delta &\coloneqq \big\{ t \in [-\pi, \pi] \,\big|\, |t| \geqslant \delta \big\}. \end{aligned}$$

Considérons tout d'abord la contribution des $t \in A_\epsilon$. Dans ce cas, $\lambda_0 + \mathrm{i}t \in B_{\delta_2}$ lorsque n est assez grand. On en déduit que, uniformément en $t \in A_\epsilon$,

$$\phi_{S_n}^{\lambda_0}(t) = e^{nH_{\lambda_0}(\mathrm{i}t)} = \exp\big(\mathrm{i}x_n t - \tfrac{1}{2}\sigma_{\lambda_0}^2 n t^2 + \mathsf{O}(n^{-1/2+3\epsilon})\big),$$

où l'on a utilisé les relations $H_{\lambda_0}(0) = 0$, $H'_{\lambda_0}(0) = x_n/n$ et $H''_{\lambda_0}(0) = \sigma_{\lambda_0}^2$. Par conséquent, la contribution à (2.8) provenant de l'intégration sur A_ϵ est

$$\begin{aligned} &\frac{1}{2\pi} \int_{A_\epsilon} \exp\big(-\tfrac{n}{2}\sigma_{\lambda_0}^2 t^2 + \mathsf{O}(n^{-1/2+3\epsilon})\big)\, \mathrm{d}t \\ &\qquad = (1 + \mathsf{o}(1)) \frac{1}{2\pi\sqrt{\sigma_{\lambda_0}^2 n}} \int_{|s| < n^\epsilon \sigma_{\lambda_0}} \exp\big(-\tfrac{1}{2}s^2\big)\, \mathrm{d}s = (1 + \mathsf{o}(1)) \frac{1}{\sqrt{2\pi\sigma_{\lambda_0}^2 n}}. \end{aligned}$$

Il reste donc, afin d'établir (2.7), à démontrer que la contribution provenant de l'intégration sur $A_{\epsilon,\delta} \cup A_\delta$ est négligeable (c'est-à-dire, $\mathsf{o}(n^{-1/2})$).

Analysons pour commencer la contribution des $t \in A_{\epsilon,\delta}$. On choisit $\delta < \delta_2$, de sorte que $\lambda_0 + \mathrm{i}t \in B_{\delta_2}$ pour tout $t \in A_{\epsilon,\delta}$ dès que n est assez grand. On en conclut que, pour δ suffisamment petit et n suffisamment grand,

$$|\phi_{S_n}^{\lambda_0}(t)| = \exp\big(-\tfrac{n}{2}(\sigma_{\lambda_0}^2 + \mathsf{O}(\delta))\, t^2\big) \leqslant \exp\big(-\tfrac{1}{3}\sigma^2 t^2 n\big) \leqslant \exp\big(-\tfrac{1}{3}\sigma^2\, n^{2\epsilon}\big),$$

puisque $t^2 n \geqslant n^{2\epsilon}$. Comme $\sigma^2 > 0$, la contribution de $A_{\epsilon,\delta}$ est donc bien négligeable.

Considérons finalement les $t \in A_\delta$. Comme $\mathbb{P}^{\lambda_0}(X_1 = k) > 0 \Leftrightarrow \mathbb{P}(X_1 = k) > 0$, il suit de H_2 et du Lemme 2.3 ci-dessus que

$$|\mathbb{E}^{\lambda_0}[e^{\mathrm{i}tX_1}]| < 1,$$

pour tout $t \in [-\pi, \pi] \setminus \{0\}$. Par continuité de la fonction caractéristique, il existe donc $\alpha > 0$ tel que

$$\sup_{t \in A_\delta} |\mathbb{E}^{\lambda_0}[e^{\mathrm{i}tX_1}]| < 1 - \alpha,$$

et donc

$$|\phi_{S_n}^{\lambda_0}(t)| = |\mathbb{E}^{\lambda_0}[e^{\mathrm{i}tS_n}]| = |\mathbb{E}^{\lambda_0}[e^{\mathrm{i}tX_1}]|^n \leqslant (1 - \alpha)^n,$$

ce qui montre que la contribution de A_δ est elle aussi négligeable. □

2.5 Généralisations, extensions

Dans cette section, nous présentons, sans preuves, diverses extensions du Théorème 2.1.

Déviations d'ordre n. Dans l'énoncé du Théorème 2.1, on se restreint à des points situés à distance $\mathsf{o}(n)$ de l'espérance de S_n. On peut toutefois facilement étendre le théorème limite local aux fluctuations d'ordre n pour peu que l'on renforce la condition de Cramér en exigeant que $\mathbb{E}[e^{tX_1}] < \infty$ pour tous les $t \in \mathbb{R}$ (et pas seulement dans un voisinage de 0). Plus précisément, soient

$$B^- := \inf\big\{b \in \mathbb{Z} \,\big|\, \mathbb{P}(X_1 = b) > 0\big\}, \qquad B^+ := \sup\big\{b \in \mathbb{Z} \,\big|\, \mathbb{P}(X_1 = b) > 0\big\}.$$

Alors, sous la condition discutée ci-dessus, l'asymptotique (2.1) s'étend à toutes les suites $(x_n)_{n\geqslant 1}$ telles que

$$B^- < \liminf_{n\to\infty} \frac{x_n}{n} \leqslant \limsup_{n\to\infty} \frac{x_n}{n} < B^+.$$

Une preuve peut être trouvée dans [16].

TCL local sous les hypothèses H_1 et H_2. Comme mentionné précédemment, la condition de Cramér H_3 est nécessaire pour la validité de (2.1) lorsque x_n se trouve loin de l'espérance de S_n. Si l'on est uniquement intéressé au comportement proche de l'espérance, on peut se contenter des hypothèses H_1 et H_2. On a, par exemple, sous ces deux hypothèses, le résultat suivant [24, 43] :

$$\lim_{n\to\infty} \sup_{x\in\mathbb{Z}} \Big| \sqrt{2\pi\sigma^2 n}\, \mathbb{P}(S_n = x) - \exp\Big\{\frac{-(x-n\mu)^2}{2\sigma^2 n}\Big\}\Big| = 0.$$

De plus, la vitesse de convergence peut être estimée sous l'hypothèse que X_1 possède des moments d'ordres supérieurs. Par exemple, si X_1 possède un moment d'ordre 3, alors le supremum est $\mathsf{O}(n^{-1/2})$.

Remarquons encore qu'il est possible de démontrer des résultats similaires pour des variables à queues lourdes pour lesquelles l'hypothèse H_1 n'est pas satisfaite ; la limite n'est alors plus gaussienne. Nous renvoyons à [24] pour plus de détails.

Comportement de la fonction de taux au-delà du régime gaussien. La dérivation de (2.2) repose sur l'hypothèse que $|x_n - n\mu| = \mathsf{o}(n^{2/3})$. Une telle hypothèse est nécessaire, car l'approximation quadratique de la fonction de taux utilisée dans (2.6) n'est plus valable lorsque cette condition est violée. La preuve donnée ci-dessus permet toutefois obtenir une succession de raffinements, pour des déviations de plus en plus importantes, en conservant un nombre suffisant de termes dans le développement asymptotique de la fonction de taux. On obtient ainsi le résultat suivant [16]. Soit $3 \leqslant k \in \mathbb{N}$; sous les hypothèses du Théorème 2.1, lorsque $|x_n - \mu| = \mathsf{o}(n^{(k-1)/k})$, on a l'asymptotique

$$\mathbb{P}(S_n = x_n) = \frac{1}{\sqrt{2\pi\sigma^2 n}} \exp\Big\{-\frac{(x_n - n\mu)^2}{2\sigma^2 n} - n\sum_{j=3}^{k} \frac{K_j}{j!} \frac{(x_n - \mu n)^j}{(\sigma^2 n)^j}\Big\}(1 + \mathsf{o}(1)),$$

où les coefficients K_j peuvent être déterminés explicitement [16]. On obtient ainsi $K_3 = -H^{(3)}(0)$, par exemple.

Variables à valeurs dans $\mathbb{Z}^d$. Afin de simplifier au maximum les notations, l'énoncé du Théorème 2.1 est restreint aux variables aléatoires à valeurs dans $\mathbb{Z}$. Le résultat et sa preuve s'étendent sans problème au cas de variables aléatoires à valeurs dans $\mathbb{Z}^d$, sous l'analogue approprié des hypothèses $\mathsf{H}_{1,2,3}$. Soient $X_1, X_2, \ldots$ des variables aléatoires i.i.d. à valeurs dans $\mathbb{Z}^d$. On notera $S_n := \sum_{i=1}^n X_i$, et X_n^j la $j^{\text{ème}}$ coordonnée de X_n. On suppose que

- ▷ $\mathbb{E}(X_1) = \mu \in \mathbb{R}^d$, et la matrice $d \times d$ D avec éléments de matrice $D_{ij} := \mathrm{Cov}[X_1^i, X_1^j]$ est définie positive.
- ▷ $|\mathbb{E}(e^{\mathrm{i}\langle \tau, X_1\rangle})| < 1$, pour tout $\tau \in [-\pi, \pi]^d \setminus \{0\}$.
- ▷ $F(z) := \mathbb{E}(e^{\langle z, X_1\rangle})$ est holomorphe dans un voisinage de 0 dans $\mathbb{C}^d$.

Soit $\Delta : \mathbb{N} \to \mathbb{R}$ telle que $\lim_{n\to\infty} \Delta(n)/n = 0$. Alors, uniformément pour toute suite $(x_n)_{n\geqslant 1}$ de points de $\mathbb{Z}^d$ tels que $\|x_n - n\mu\|_2 < \Delta(n)$, on a, lorsque $n \to \infty$,

$$\mathbb{P}(S_n = x_n) = \frac{1}{\sqrt{(2\pi n)^d \det D}} \exp\bigl(-nI(x_n/n)\bigr)(1 + \mathsf{o}(1)),$$

où $I(t) := \langle \lambda(t), t\rangle - H(\lambda(t))$, avec $H(z) := \log F(z)$ et $\lambda(t)$ est l'unique solution de l'équation $\nabla H(\lambda) = t$. En particulier, si $\Delta(n) = \mathsf{o}(n^{2/3})$, on a

$$\mathbb{P}(S_n = x_n) = \frac{1}{\sqrt{(2\pi n)^d \det D}} \exp\bigl(-\langle x_n - n\mu, D^{-1}(x_n - n\mu)\rangle/2n\bigr)(1 + \mathsf{o}(1)).$$

Variables à densité. On a jusqu'à présent considéré uniquement le cas de variables aléatoires discrètes. Dans le cas de variables à densité, le problème correspondant revient à obtenir une description asymptotique de la densité de S_n en un certain point. Des théorèmes similaires peuvent être obtenus, par des méthodes du même type, *cf.*, par exemple, les livres [43, 24].

Autres extensions. Mentionnons pour terminer que des résultats de ce type peuvent également être obtenus dans des situations plus générales : variables aléatoires non identiquement distribuées [43] ou non-indépendantes (*cf.*, par exemple, [12]).

3 La loi du logarithme itéré

3.1 La problématique

Soit $X_1, X_2, \dots$ une suite de variables aléatoires i.i.d. à valeurs dans $\mathbb{Z}$, d'espérance $\mu \coloneqq \mathbb{E}[X_1]$ et de variance $\sigma^2 \coloneqq \mathrm{Var}[X_1] \in (0, \infty)$. On sait du Théorème Central Limite que

$$\widetilde{S}_n \coloneqq \frac{1}{\sqrt{n\sigma^2}} \sum_{i=1}^{n} (X_i - \mu)$$

converge en loi vers une variable aléatoire de loi $\mathcal{N}(0,1)$. En particulier, $\tilde{S}_n$ prend typiquement des valeurs d'ordre 1. Toutefois, il est clair que $\tilde{S}_n$ va occasionnellement prendre des valeurs très grandes et il est naturel d'essayer de déterminer l'ordre de grandeur des fluctuations extrêmes de cette variable aléatoire.

3.2 Énoncé de la loi du logarithme itéré

Théorème 3.1 (Loi du logarithme itéré)**.** *Sous les hypothèses du Théorème 2.1, on a, presque-sûrement,*

$$\limsup_{n\to\infty} \frac{\widetilde{S}_n}{\sqrt{2\log\log n}} = 1.$$

Remarque 3.2. *Le résultat reste vrai sous la seule condition* H_1*. Le résultat devient par contre nettement plus difficile à démontrer. Supposer également la validité de* H_2 *et* H_3 *va nous permettre d'utiliser les résultats du Chapitre 2 (plus précisément, le Corollaire 2.2).* ◇

Remarque 3.3. *Pour des raisons évidentes de symétrie, il suit de ce théorème que*

$$\liminf_{n\to\infty} \frac{\widetilde{S}_n}{\sqrt{2\log\log n}} = -1.$$ ◇

Il est important de bien comprendre ce qu'affirme ce théorème. Soit $\epsilon > 0$ arbitraire ; le processus stochastique $S_n \coloneqq \sum_{i=1}^{n} X_i$ satisfait alors presque sûrement (*cf.* Fig. 3.1)

$$S_n \leqslant n\mu + (1+\epsilon)\sqrt{2n\sigma^2 \log\log n}, \tag{3.1}$$

pour tout n suffisamment grand, mais également

$$S_n > n\mu + (1-\epsilon)\sqrt{2n\sigma^2 \log\log n}, \tag{3.2}$$

pour une infinité de valeurs de n.

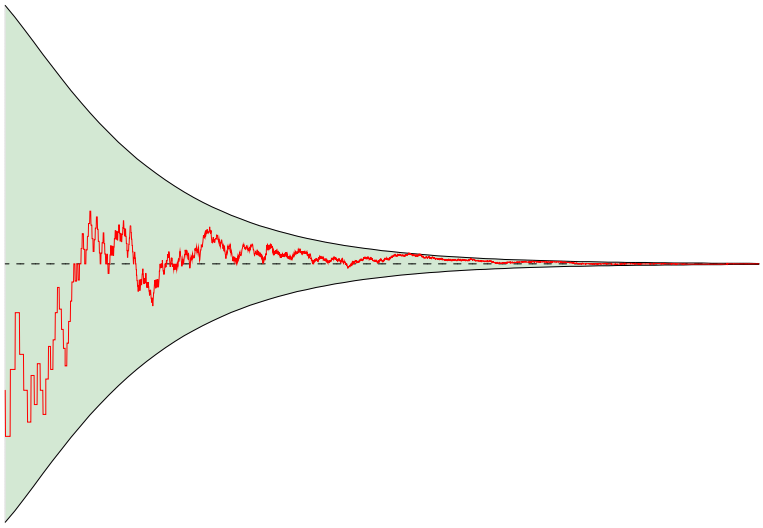


FIGURE 3.1: Une réalisation du processus $\frac{1}{n}S_n$, lorsque X_1 suit une loi de Bernoulli de paramètre $1/2$, pour n allant de 1 à 10 000 000 ; l'échelle horizontale est logarithmique. La région en vert s'étend entre les courbe $\mu \pm \sqrt{2\sigma^2 \log\log n}/\sqrt{n}$.

3.3 Preuve de la loi du logarithme itéré

Clairement, on peut supposer, sans perte de généralité, que $\mathbb{E}[X_1] = \mu = 0$.
La preuve repose sur le lemme suivant.

Lemme 3.4. *Soit $(X_k)_{k\geqslant 1}$ une suite de variables aléatoires i.i.d. telles que $\mathbb{E}(X_1) = 0$ et $\mathrm{Var}[X_1] < \infty$; on pose, pour $n \in \mathbb{N}$, $S_n := \sum_{k=1}^n X_k$. Il existe alors une constante $c < \infty$ telle que, pour tout $x \in \mathbb{R}$ fixé,*

$$\mathbb{P}(\max_{0\leqslant k\leqslant n} S_k > x) \leqslant c\,\mathbb{P}(S_n > x).$$

Preuve du Lemme 3.4. Soient $\sigma_x := \inf\{k \geqslant 0 \,|\, S_k > x\}$ et $\mathcal{B}_k := \{S_n - S_k \geqslant 0\}$. Observons que les événements $\{\sigma_x = k\} \cap \mathcal{B}_k$, $1 \leqslant k \leqslant n$, sont disjoints et que la réalisation de l'un d'entre eux entraîne automatiquement celle de l'événement $\{S_n > x\}$. On peut donc écrire

$$\begin{aligned}\mathbb{P}(S_n > x) \geqslant \sum_{k=0}^n \mathbb{P}(\{\sigma_x = k\} \cap \mathcal{B}_k) &= \sum_{k=0}^n \mathbb{P}(\sigma_x = k)\,\mathbb{P}(\mathcal{B}_k) \\ &\geqslant \min_{0\leqslant k\leqslant n} \mathbb{P}(\mathcal{B}_k) \sum_{k=1}^n \mathbb{P}(\sigma_x = k) \\ &= \min_{0\leqslant k\leqslant n} \mathbb{P}(S_{n-k} \geqslant 0)\,\mathbb{P}(\max_{0\leqslant k\leqslant n} S_k > x),\end{aligned}$$

où l'on a utilisé l'indépendance de $\{\sigma_x = k\}$ (qui ne dépend que de $X_1, \ldots, X_k$) et $\mathcal{B}_k$ (qui ne dépend que de $X_{k+1}, \ldots, X_n$) pour la première égalité.

L'observation suivante est que la condition $\mathbb{E}[X_1] = 0$ garantit que $\mathbb{P}(S_m \geqslant 0) \neq 0$ pour tout $m \geqslant 0$. De plus, par le Théorème Central Limite, $\lim_{m\to\infty} \mathbb{P}(S_m \geqslant 0) = 1/2$. Par conséquent, il existe $c' > 0$ telle que $\mathbb{P}(S_m \geqslant 0) \geqslant c' > 0$ pour tout $m \geqslant 0$. La conclusion suit, avec $c = 1/c'$. □

Nous allons à présent établir (3.1). Soit $\epsilon > 0$, $\gamma \in (1, 1+\epsilon)$, et notons $n_r := \lfloor \gamma^r \rfloor$. On considère l'événement $\mathcal{C}_r$ stipulant que l'inégalité

$$S_n > (1+\epsilon)\sqrt{2\sigma^2 n_r \log\log n_r}$$

est vérifiée pour au moins une valeur de n telle que $n_r \leqslant n < n_{r+1}$. Il est évident qu'il suffit de montrer que, presque-sûrement, seul un nombre fini des événements $\mathcal{C}_r$, $r \geqslant 1$, sont réalisés. Il suit du Lemme 3.4 que

$$\mathbb{P}(\mathcal{C}_r) \leqslant c\,\mathbb{P}\bigl(S_{n_{r+1}} > (1+\epsilon)\sqrt{2n_r\sigma^2 \log\log n_r}\bigr).$$

On peut aisément estimer cette dernière probabilité à l'aide du Corollaire 2.2 : pour r grand, $n_{r+1}/n_r \leqslant \gamma/(1-\gamma^{-r}) < 1+\epsilon$, et donc

$$\begin{aligned}\mathbb{P}\bigl(S_{n_{r+1}} > (1+\epsilon)\sqrt{2n_r\sigma^2 \log\log n_r}\bigr) &= \mathbb{P}\bigl(\widetilde{S}_{n_{r+1}} > (1+\epsilon)\sqrt{2(n_r/n_{r+1}) \log\log n_r}\bigr)\\ &\leqslant e^{-(1+\epsilon)^2(n_r/n_{r+1})\log\log n_r}\\ &\leqslant e^{-(1+\epsilon)\log\log n_r} \leqslant (\tfrac{1}{2}\log\gamma)^{-1-\epsilon} r^{-1-\epsilon}.\end{aligned}$$

Par conséquent, $\sum_r \mathbb{P}(\mathcal{C}_r) < \infty$ et (3.1) suit du premier lemme de Borel–Cantelli.

Il reste à établir (3.2). Soit à nouveau $\epsilon > 0$ et $\gamma \in \mathbb{N}$ (ce dernier sera choisi suffisamment grand par la suite). On pose $n_r := \gamma^r$. La preuve va reposer sur le second lemme de Borel–Cantelli, et il faut donc introduire une famille appropriée d'événements *indépendants* : pour $r \geqslant 1$, on introduit l'événement

$$\mathcal{D}_r := \bigl\{S_{n_r} - S_{n_{r-1}} > (1-\tfrac{1}{2}\epsilon)\sqrt{2\sigma^2 n_r \log\log n_r}\bigr\}.$$

Nous allons tout d'abord vérifier que, presque-sûrement, une infinité des événements $\mathcal{D}_r$ sont réalisés, puis nous montrerons que cela implique (3.2). Les événements $\mathcal{D}_r$ étant indépendants, le second lemme de Borel–Cantelli montre que la divergence de la série $\sum_r \mathbb{P}(\mathcal{D}_r)$ suffit à établir la première affirmation. Or,

$$\begin{aligned}\mathbb{P}(\mathcal{D}_r) &= \mathbb{P}\bigl(S_{n_r - n_{r-1}} > (1-\tfrac{1}{2}\epsilon)\sqrt{2\sigma^2 n_r \log\log n_r}\bigr)\\ &= \mathbb{P}\bigl(\widetilde{S}_{n_r - n_{r-1}} > (1-\tfrac{1}{2}\epsilon)\sqrt{2\tfrac{n_r}{n_r - n_{r-1}} \log\log n_r}\bigr)\\ &\geqslant \mathbb{P}\bigl(\widetilde{S}_{n_r - n_{r-1}} > \sqrt{2(1-\tfrac{1}{2}\epsilon)\log\log n_r}\bigr),\end{aligned}$$

puisque $(n_r - n_{r-1})/n_r = (\gamma-1)/\gamma > 1 - \epsilon/2$, pour peu que l'on choisisse $\gamma > 2/\epsilon$. Cette dernière probabilité peut à nouveau être estimée à l'aide du Corollaire 2.2 :

$$\begin{aligned}\mathbb{P}\bigl(\widetilde{S}_{n_r - n_{r-1}} > \sqrt{2(1-\tfrac{1}{2}\epsilon)\log\log n_r}\bigr) &\geqslant \frac{1/2}{\sqrt{4\pi \log\log n_r}} e^{-(1-\epsilon/2)\log\log n_r}\\ &= \frac{1/2}{\sqrt{4\pi \log\log n_r}(\log n_r)^{1-\epsilon/2}}.\end{aligned}$$

Comme $(\log n_r)^{1-\epsilon/2} = (r\log\gamma)^{1-\epsilon/2}$, la série $\sum_r \mathbb{P}(\mathcal{D}_r)$ diverge comme annoncé.

Il ne reste plus qu'à montrer que la réalisation d'une infinité des événements $\mathcal{D}_r$ implique (3.2). Pour cela, il suffit d'observer que

$$\frac{S_{n_r}}{\sqrt{2n_r\sigma^2\log\log n_r}} = \frac{S_{n_r} - S_{n_{r-1}}}{\sqrt{2n_r\sigma^2\log\log n_r}} + \frac{S_{n_{r-1}}}{\sqrt{2n_{r-1}\sigma^2\log\log n_{r-1}}} \frac{\sqrt{n_{r-1}\log\log n_{r-1}}}{\sqrt{n_r \log\log n_r}}.$$

Étant donné que

$$\lim_{r\to\infty} \frac{\sqrt{n_{r-1}\log\log n_{r-1}}}{\sqrt{n_r\log\log n_r}} = \frac{1}{\sqrt{\gamma}},$$

et que

$$\liminf_{r\to\infty} \frac{S_{n_{r-1}}}{\sqrt{2n_{r-1}\sigma^2\log\log n_{r-1}}} \geqslant -(1+\epsilon),$$

par (3.1) (dont la validité a été établie plus haut), on en déduit que

$$\limsup_{r\to\infty} \frac{S_{n_r}}{\sqrt{2n_r\sigma^2\log\log n_r}} \geqslant (1-\tfrac{1}{2}\epsilon) - \frac{1+\epsilon}{\sqrt{\gamma}} \geqslant 1-\epsilon,$$

avec probabilité 1, pourvu que γ soit choisi suffisamment grand. □

3.4 Remarques historiques et extensions

Soit $0 \leqslant x < 1$ un nombre réel, et notons

$$x = 0, \mathsf{x}_1\mathsf{x}_2\mathsf{x}_3 \ldots \coloneqq \sum_{k=1}^{\infty} \mathsf{x}_k 2^{-k}$$

son développement binaire. Ce développement est unique pour tous les x, à l'exception de ceux de la forme $a2^{-n}$ où $a, n \in \mathbb{N}$, ces derniers pouvant être représentés soit par un développement se terminant par une infinité de 0, soit par un développement se terminant par une infinité de 1. Afin de lever l'ambiguïté, nous utiliserons la convention de n'utiliser que la première représentation. Observez que ce « problème » ne concerne qu'un sous-ensemble dénombrable de $[0, 1)$ et ne jouera par conséquent aucun rôle dans la discussion qui suit.

Considérons à présent une variable aléatoire X uniformément distribuée sur $[0, 1)$. Une réalisation de cette variable aléatoire peut être obtenue en tirant au hasard, indépendemment et uniformément, chacun des bits de son développement binaire ; notons $\mathsf{X}_1, \mathsf{X}_2, \ldots$ les variables aléatoires de Bernoulli correspondantes. En effet, fixons $n \in \mathbb{N}^*$ et $0 \leqslant k < 2^n$ et considérons $a = 0, \mathsf{a}_1\mathsf{a}_2 \ldots \mathsf{a}_n \coloneqq k2^{-n}$ et $b \coloneqq (k+1)2^{-n}$. Alors $X \in [a, b)$ si et seulement si $\mathsf{X}_k = \mathsf{a}_k$ pour chaque $k \in \{1, \ldots, n\}$, ce qui a lieu avec probabilité 2^{-n}, ce qui est bien la mesure de Lebesgue de l'intervalle $[a, b)$. L'affirmation suit, car tout borélien de $[0, 1)$ peut être approximé à partir de tels intervalles.

Étant donné un nombre $x \in [0, 1)$, on note

$$S_n(x) \coloneqq \sum_{i=1}^{n} (\mathbf{1}_{\{\mathsf{x}_i = 1\}} - \mathbf{1}_{\{\mathsf{x}_i = 0\}})$$

l'excès de 1 dans les n premiers bits de son développement binaire. Au début du XX$^{\text{e}}$ siècle, de nombreux mathématiciens se sont intéressés à comprendre comment cette quantité se comporte asymptotiquement pour un nombre typique (c'est-à-dire presque partout relativement à la mesure de Lebesgue). Bien entendu, la loi forte des grands nombres implique que, presque partout, $S_n = \mathsf{o}(n)$. La question était alors d'obtenir des informations plus fines sur le comportement de cette différence. Les progrès ont été obtenus par paliers : Lebesgue-presque partout,

- ▷ Hausdorff, 1913 : $S_n = \mathsf{o}(n^{1/2+\epsilon})$, pour tout $\epsilon > 0$;
- ▷ Hardy et Littlewood, 1914 : $S_n = \mathsf{O}((n \log n)^{1/2})$;
- ▷ Steinhaus, 1922 : $\limsup_{n\to\infty} S_n/(2n \log n)^{1/2} \leqslant 1$;
- ▷ Khintchine, 1923 : $S_n = \mathsf{O}((n \log\log n)^{1/2})$;
- ▷ Khintchine, 1924 : $\limsup_{n\to\infty} S_n/(2n \log\log n)^{1/2} = 1$.

Ce dernier résultat est bien entendu un cas particulier du Théorème 3.1 (observez qu'ici $\sigma^2 = 1$). Ce résultat a ensuite encore été amélioré. Pour énoncer ces améliorations, il est utile d'introduire la terminologie suivante (due à Paul Lévy) : on dit qu'une fonction ϕ appartient à la **classe supérieure** $\mathcal{U}$ si, pour presque tout x, $S_n(x) > n^{1/2}\phi(n)$ pour seulement un nombre fini de valeurs de n ; on dit que ϕ appartient à la **classe inférieure** $\mathcal{L}$ si, pour presque tout x, $S_n(x) > n^{1/2}\phi(n)$ pour une infinité de valeurs de n.[1] Dans ce langage, le résultat de Khintchine peut se reformuler comme suit :

$$a\sqrt{2 \log\log t} \in \begin{cases} \mathcal{U} & \text{si } a > 1, \\ \mathcal{L} & \text{si } a < 1. \end{cases}$$

On a alors les améliorations successives suivantes :

1. Il suit de la loi $0/1$ de Hewitt–Savage que toute fonction ϕ appartient à l'une de ces deux classes.

▷ Lévy, 1933 : $(2\log\log t + a\log\log\log t)^{1/2} \in \begin{cases} \mathcal{U} & \text{si } a > 3, \\ \mathcal{L} & \text{si } a \leqslant 1. \end{cases}$

▷ Erdős, 1942 : si ϕ est une fonction positive croissante, alors

$$\phi \in \mathcal{U} \;\Leftrightarrow\; \sum_n \frac{1}{n}\phi(n)e^{-\phi^2(n)/2} < \infty.$$

Ce dernier résultat donne une réponse complète à la question initiale, dans le cas de variables de Bernoulli. Il a été étendu au cadre général traité dans ce chapitre par Feller [20] (en fait, ce dernier ne suppose même pas les variables aléatoires identiquement distribuées). Cette condition est assez implicite ; on peut toutefois montrer qu'elle implique en particulier que, pour tout $k \geqslant 4$,

$$(2\log_2 t + 3\log_3 t + 2\log_4 t + \cdots + 2\log_{k-1} t + (2+\epsilon)\log_k t)^{1/2} \in \begin{cases} \mathcal{U} & \text{si } \epsilon > 0, \\ \mathcal{L} & \text{si } \epsilon < 0, \end{cases}$$

où on a posé $\log_k t := \log(\log_{k-1} t)$ et $\log_1 t := \log t$.

Remarques bibliographiques : L'approche présentée dans ce chapitre est fortement inspirée de [21, Section VIII.5]. La Section 3.4 est tirée de l'introduction de [19]. Le Théorème 3.1 est originellement dû à [33].

4 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 $(\Omega, \mathcal{P}(\Omega))$.

4.1 Distance en variation totale

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

Définition 4.1. *Soit $\mu, \nu \in \mathcal{M}_1(\Omega)$. La **distance en variation totale** entre μ et ν est*

$$\|\mu - \nu\|_{\mathrm{VT}} \coloneqq \sup_{A \subset \Omega} |\mu(A) - \nu(A)|. \tag{4.1}$$

Lemme 4.2. *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 4.3. *Pour tout $\mu, \nu \in \mathcal{M}_1(\Omega)$,*

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

De plus, le supremum dans (4.1) est atteint lorsque $A = E \coloneqq \{i \in \Omega \,|\, \mu(i) \geqslant \nu(i)\}$.

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

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

Passons aux autres affirmations. Soit $E \coloneqq \{i \in \Omega \,|\, \mu(i) \geqslant \nu(i)\}$. Alors, pour tout $A \subset \Omega$,

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

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

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

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

Lemme 4.4. *Pour tout $\mu,\nu \in \mathcal{M}_1(\Omega)$, $\|\mu-\nu\|_{\mathrm{VT}} = 1 - \sum_{i\in\Omega} \mu(i)\wedge\nu(i)$.*

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

$$\sum_{i\in\Omega} \mu(i)\wedge\nu(i) = \tfrac{1}{2}\sum_{i\in\Omega}\big(\mu(i)+\nu(i)-|\mu(i)-\nu(i)|\big) = 1-\|\mu-\nu\|_{\mathrm{VT}}. \qquad \square$$

4.2 Couplage

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

Définition 4.5. *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 :*

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

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

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

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 4.7. *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 $\geqslant$. Le couplage est dit **monotone** si $\mathbb{P}(X\leqslant Y)=1$.*

Exemple 4.8. Considérons à nouveau $X\sim\mathrm{Bern}(p)$ et $Y\sim\mathrm{Bern}(q)$ avec $0\leqslant p\leqslant q\leqslant 1$. Soit U une variable aléatoire uniforme sur $[0,1]$. On considère les variables aléatoires $X'' := \mathbf{1}_{\{U\leqslant p\}}$ et $Y'' := \mathbf{1}_{\{U\leqslant q\}}$. Alors, (X'',Y'') est un couplage monotone de X et Y :

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

Exemple 4.9. Soit $0\leqslant p\leqslant q\leqslant 1$ et $n\in\mathbb{N}^*$. Soit $G_p\sim\mathsf{G}(n,p)$ et $G_q\sim\mathsf{G}(n,q)$ des graphes aléatoires d'Erdős–Rényi (*cf.* Chapitre 8). G_p et G_q sont des sous-graphes aléatoires du graphe complet $K_n=([\![n]\!],A_n)$, où $A_n := \{\{i,j\}\mid 1\leqslant i<j\leqslant n\}$. 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 $U_e \leqslant p$, resp. $U_e \leqslant q$. 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, $G \subset G' \implies f(G) \leqslant f(G')$. On peut, par exemple, considérer la fonction $f(G) \coloneqq \mathbf{1}_{\{G \text{ est connexe}\}}$ ou encore la fonction $f(G) \coloneqq \max\{|C| \,|\, C \text{ composante connexe maximale de } G\}$. Soit $G \sim \mathsf{G}(n,p)$. Alors, pour une telle fonction f, le couplage précédent implique que l'espérance de $f(G)$ est croissante en p :

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

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

Lemme 4.10. *Soit $\mu, \nu \in \mathcal{M}_1(\Omega)$. Pour tout couplage (X,Y) de μ et ν, on a*

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

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

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

La conclusion suit donc du Lemme 4.3. □

Exemple 4.11. Soit $\lambda > \nu > 0$. On souhaite borner la distance en variation totale entre les lois des variables aléatoires $X \sim \mathrm{Poisson}(\lambda)$ et $Y \sim \mathrm{Poisson}(\nu)$. On va le faire en construisant un couplage de X et Y.

Soit $Y' \sim \mathrm{Poisson}(\nu)$ et $Z' \sim \mathrm{Poisson}(\lambda - \nu)$ deux variables aléatoires indépendantes. On pose $X' \coloneqq Y' + Z'$. Il suit du cours d'introduction à la théorie des probabilités que $X' \sim \mathrm{Poisson}(\lambda)$. Par conséquent, (X', Y') est un couplage de X et Y et le Lemme 4.10 implique que

$$\|\mathscr{L}(X) - \mathscr{L}(Y)\|_{\mathrm{VT}} \leqslant \mathbb{P}(X' \neq Y') = \mathbb{P}(Z' \neq 0) = 1 - e^{-(\lambda-\nu)}. \qquad \diamond$$

Le résultat suivant montre que la borne du Lemme 4.10 est optimale.

Lemme 4.12. *Soit $\mu, \nu \in \mathcal{M}_1(\Omega)$. Il existe un couplage (X,Y) de μ et ν tel que*

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

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

Démonstration. Les cas $\|\mu - \nu\|_{\mathrm{VT}} \in \{0,1\}$ sont triviaux et sont laissés en exercice ; on suppose donc $\|\mu - \nu\|_{\mathrm{VT}} \in (0,1)$. Considérons à nouveau $E \coloneqq \{i \in \Omega \,|\, \mu(i) \geqslant \nu(i)\}$. Par le Lemme 4.4,

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

En particulier,

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

Ainsi,

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

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

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 $\mathbb{P}(X \neq Y) = 1 - \tilde{p} = \|\mu - \nu\|_{\mathrm{VT}}$. □

Exemple 4.13. Retournons au cas de deux variables aléatoires indépendantes $X \sim \mathrm{Bern}(p)$ et $Y \sim \mathrm{Bern}(q)$, avec $0 < p \leqslant q < 1$. On cherche à construire un couplage maximal (X''', Y''') des lois de X et Y. En reprenant les notations de la preuve du Lemme 4.12, on a $\Omega \coloneqq \{0,1\}$, $E \coloneqq \{0\}$, $E^c \coloneqq \{1\}$, $\tilde{p} \coloneqq (1-q)+p$, $\rho_E(0) \coloneqq 1$, $\rho_{E^c}(1) \coloneqq 1$ et $\big(\rho_{\min}(i)\big)_{i\in\Omega} \coloneqq \big((1-q)/(1-q+p), p/(1-q+p)\big)$. On en conclut que

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

ce qui n'est rien d'autre que le couplage monotone de l'Exemple 4.8. ◇

4.3 Quelques applications

4.3.1 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 $G \sim \mathsf{G}(n, p_n)$ avec n sommets et probabilité d'occupation

$$p_n \coloneqq \frac{c}{n}.$$

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

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

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

$$\frac{N_d}{n} \xrightarrow{n\to\infty} \frac{c^d}{d!}e^{-c} \qquad \textit{en probabilité.}$$

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

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

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

$$\lim_{n\to\infty}\frac{(n-1)!}{(n-1-d)!}n^{-d} = \lim_{n\to\infty}\prod_{k=1}^{d}\frac{n-k}{n} = 1 \qquad \text{et} \qquad \lim_{n\to\infty}\Big(1-\frac{c}{n}\Big)^{n-1-d} = e^{-c},$$

ce qui implique que

$$\lim_{n\to\infty}\mathbb{E}\Big[\frac{N_d}{n}\Big] = \frac{c^d}{d!}e^{-c}. \tag{4.2}$$

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 :

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

Notons que $\mathbb{P}(D_1=d)^2 = \mathbb{P}(D_1=d, D_2'=d)$, où $D_2' \sim \mathrm{Binom}(n-1, p_n)$ 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 $Y_1 \sim \mathrm{Binom}(n-2, p_n)$, $Y_2 \sim \mathrm{Binom}(n-2, p_n)$, $X_1 \sim \mathrm{Bern}(p_n)$ et $X_2 \sim \mathrm{Bern}(p_n)$. Alors, $(D_1, D_2) \overset{\text{loi}}{=} (Y_1+X_1, Y_2+X_1)$ et $(D_1, D_2') \overset{\text{loi}}{=} (Y_1+X_1, Y_2+X_2)$. On a donc

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

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

$$\forall \epsilon > 0, \qquad \mathbb{P}\Big(\Big|\frac{N_d}{n} - \mathbb{E}\Big[\frac{N_d}{n}\Big]\Big| \geqslant \epsilon\Big) \leqslant \frac{1+2c}{\epsilon^2 n}. \tag{4.3}$$

La conclusion suit de (4.2) et (4.3). □

4.3.2 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** :

$$\exists i_* \in \mathbb{Z}, \quad \mathbb{P}(X_1 = i_*) > 0 \text{ et } \operatorname{pgcd}\{j \in \mathbb{Z} \,|\, \mathbb{P}(X_1 - i_* = j) > 0\} = 1, \tag{4.4}$$

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

Remarque 4.15. *Cette condition exclut, par exemple, la marche aléatoire simple sur $\mathbb{Z}$, puisque l'on a* $\operatorname{pgcd}\{j \in \mathbb{Z} \,|\, \mathbb{P}(X_1 - 1 = j) > 0\} = \{0, -2\} = 2$. ◇

Nous allons montrer le résultat de perte de mémoire suivant.

Théorème 4.16. *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 (4.4). Soit $S_n := \sum_{k=1}^n X_k$ et $S'_n := a + \sum_{k=1}^n X'_k$. Alors,*

$$\lim_{n\to\infty} \|\mathscr{L}(S_n) - \mathscr{L}(S'_n)\|_{\mathrm{VT}} = 0.$$

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 (4.4) et soit c suffisamment grand pour que

$$\operatorname{pgcd}\big\{j \in \mathbb{Z} \,\big|\, \mathbb{P}(X_1 - i_* = j, |X_1 - i_*| \leqslant c) > 0\big\} = 1. \tag{4.5}$$

Définissons

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

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

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

ce qui implique que $(S''_k)_{k\in\mathbb{N}} \overset{\text{loi}}{=} (S'_k)_{k\in\mathbb{N}}$ (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

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

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

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

et symétriques,

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

De plus,

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

Il suit, par conséquent, de (4.4) et de (4.5) que la loi des incréments de $(R_n)_{n\in\mathbb{N}}$ est **apériodique** :

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

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 4.17. *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, $T := \inf\{k \in \mathbb{N} \,|\, S_k = S_k''\} = \inf\{k \in \mathbb{N} \,|\, R_k = 0\}$ est presque sûrement fini. Finalement, définissons la marche aléatoire $(S_k''')_{k\in\mathbb{N}}$ par

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

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 4.10 que

$$\|\mathscr{L}(S_n) - \mathscr{L}(S_n')\|_{\mathrm{VT}} \leqslant \mathbb{P}(S_n \neq S_n''') \leqslant \mathbb{P}(T > n).$$

T étant presque sûrement fini, on a bien

$$\lim_{n\to\infty} \|\mathscr{L}(S_n) - \mathscr{L}(S_n')\|_{\mathrm{VT}} = 0. \qquad \square$$

Preuve du Lemme 4.17. 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 $\mathbb{P}(|\tilde{X}_1| \leqslant c) = 1$ pour une constante $c < \infty$, $\mathbb{P}(\tilde{X}_1 = i) = \mathbb{P}(\tilde{X}_1 = -i)$ pour tout $i \in \mathbb{Z}$ et $\operatorname{pgcd}\{j \in \mathbb{Z} \,|\, \mathbb{P}(\tilde{X}_1 = j) > 0\} = 1$.

Soit $A := \{j \in \mathbb{Z} \,|\, \exists n \in \mathbb{N}, \mathbb{P}(\tilde{R}_n = j) > 0\}$. Clairement, A est stable sous addition et satisfait $-A = A$ (les incréments étant symétriques) et $\operatorname{pgcd} A = 1$. 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 4.18. *Soit $A \subset \mathbb{Z}$ un ensemble stable sous addition et satisfaisant $-A = A$ et $\operatorname{pgcd} A = 1$. 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 $p := \mathbb{P}(|\tilde{R}_n| \leqslant 2r) < 1$. Alors, pour tout $k \in \mathbb{N}^*$,

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

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

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

Les incréments étant symétriques, il suit que

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

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 4.18. Soit $d := \min\{n \geqslant 1 \,|\, n \in A\}$. 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 $0 \leqslant i - kd < d$. 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, $d = \operatorname{pgcd} A = 1$. $\square$

Marche aléatoire sur $\mathbb{Z}^d$

Notre but dans cette section est d'étendre le résultat de perte de mémoire du Théorème 4.16 aux marches aléatoires sur $\mathbb{Z}^d$. Clairement, on ne peut pas procéder de la même façon, car l'équivalent du Lemme 4.17 n'est pas vrai lorsque $d \geqslant 3$, 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 $S_n^a := a + \sum_{i=1}^n X_k$ la **marche aléatoire paresseuse** sur $\mathbb{Z}^d$ partant de a.

Théorème 4.19. *Soit $a, b \in \mathbb{Z}^d$. Alors,*

$$\lim_{n\to\infty} \|\mathscr{L}(S_n^a) - \mathscr{L}(S_n^b)\|_{\mathrm{VT}} = 0.$$

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 $(S'_n)_{n\in\mathbb{N}} \overset{\text{loi}}{=} (S_n^a)_{n\in\mathbb{N}}$ et $(S''_n)_{n\in\mathbb{N}} \overset{\text{loi}}{=} (S_n^b)_{n\in\mathbb{N}}$. 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 4.17 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 $T := \max_{1\leqslant s\leqslant d} \tau^s$ 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 4.10 que

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

tend vers 0 lorsque $n \to \infty$. □

Avant de conclure cette section, mentionnons une conséquence directe du Théorème 4.19. Le **Laplacien discret** sur $\mathbb{Z}^d$ est l'opérateur défini par

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

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 $f : \mathbb{Z}^d \to \mathbb{R}$ 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 :

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

Corollaire 4.20. *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 $\sup_j |f(j)| \leqslant M$. Alors, pour tout $n \in \mathbb{N}^*$,

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

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

$$\begin{aligned}
|f(a) - f(b)| = \big|\mathbb{E}[f(S_n^a)] - \mathbb{E}[f(S_n^b)]\big| &= \Big|\sum_{j\in\mathbb{Z}^d} f(j)\big(\mathbb{P}(S_n^a = j) - \mathbb{P}(S_n^b = j)\big)\Big| \\
&\leqslant M \sum_{j\in\mathbb{Z}^d} \big|\mathbb{P}(S_n^a = j) - \mathbb{P}(S_n^b = j)\big| \\
&= 2M \|\mathscr{L}(S_n^a) - \mathscr{L}(S_n^b)\|_{\mathrm{VT}}.
\end{aligned}$$

Il suit alors du Théorème 4.19, en laissant $n \to \infty$, que $f(a) = f(b)$. □

Remarques bibliographiques : Ce chapitre est basé sur les notes de cours [45] et sur le livre [51]. Une autre référence classique sur les couplages est [39].

5 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 $\lim_{n\to\infty} np_n = \lambda > 0$. Alors, $S_n \coloneqq \sum_{k=1}^n I_k^{(n)}$ converge en loi vers une variable aléatoire $S \sim \mathrm{Poisson}(\lambda)$ lorsque $n \to \infty$.*

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.

5.1 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 $\rho_1, \dots, \rho_n$ respectivement. Posons $S_n \coloneqq I_1 + \cdots + I_n$ et $\lambda \coloneqq \rho_1 + \cdots + \rho_n$; on supposera $\lambda > 0$. En lieu et place de l'indépendance des I_k, nous supposerons 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. $U_k \overset{\text{loi}}{=} S_n$;
> 2. $1 + V_k \overset{\text{loi}}{=} S_n \,|\, \{I_k = 1\}$.

Théorème 5.1. *Soit p_λ la loi de Poisson de paramètre λ. Sous l'hypothèse H,*

$$\|\mathscr{L}(S_n) - \mathsf{p}_\lambda\|_{\mathrm{VT}} \leqslant \frac{1 - e^{-\lambda}}{\lambda} \sum_{k=1}^n \rho_k \mathbb{E}\big[|U_k - V_k|\big].$$

Exemple 5.2. Appliquons le théorème au cas particulier où les I_k sont indépendantes. Dans ce cas, l'hypothèse H est clairement satisfaite pour le choix $U_k \coloneqq S_n$ et $V_k \coloneqq \sum_{i \in [\![n]\!] \setminus \{k\}} I_i = S_n - I_k$. Comme $\mathbb{E}\big[|U_k - V_k|\big] = \mathbb{E}\big[|I_k|\big] = \rho_k$, il suit du Théorème 5.1 que

$$\|\mathscr{L}(S_n) - \mathsf{p}_\lambda\|_{\mathrm{VT}} \leqslant \frac{1 - e^{-\lambda}}{\lambda} \sum_{k=1}^n \rho_k^2 \leqslant \min\{1, \tfrac{1}{\lambda}\} \sum_{k=1}^n \rho_k^2.$$

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 $\mathrm{Bern}(\rho)$, on obtient

$$\|\mathscr{L}(S_n) - \mathsf{p}_\lambda\|_{\mathrm{VT}} \leqslant \min\{n\rho^2, \rho\},$$

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

Nous démontrerons le Théorème 5.1 dans la Section 5.2. Avant cela, énonçons quelques corollaires utiles pour les applications. (D'autres versions sont également discutées dans les séries d'exercices.)

Corollaire 5.3. *Supposons que, pour chaque $k \in [\![n]\!]$, il existe des variables aléatoires $(Z_i^{(k)})_{i\in[\![n]\!]\setminus\{k\}}$ telles que $(Z_i^{(k)})_{i\in[\![n]\!]\setminus\{k\}} \overset{\text{loi}}{=} (I_i)_{i\in[\![n]\!]\setminus\{k\}} \,|\, \{I_k = 1\}$. Alors,*

$$\|\mathscr{L}(S_n) - \mathsf{p}_\lambda\|_{\mathrm{VT}} \leqslant \frac{1-e^{-\lambda}}{\lambda} \sum_{k=1}^{n} \rho_k \Big(\rho_k + \sum_{i\in[\![n]\!]\setminus\{k\}} \mathbb{E}\big[|I_i - Z_i^{(k)}|\big]\Big).$$

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

$$\mathbb{E}\big[|U_k - V_k|\big] \leqslant \mathbb{E}\big[|I_k|\big] + \sum_{i\in[\![n]\!]\setminus\{k\}} \mathbb{E}\big[|I_i - Z_i^{(k)}|\big] = \rho_k + \sum_{i\in[\![n]\!]\setminus\{k\}} \mathbb{E}\big[|I_i - Z_i^{(k)}|\big].$$ □

Corollaire 5.4. *Supposons que, pour chaque $i \in [\![n]\!]$, il existe $\mathscr{V}_i \subset [\![n]\!] \setminus \{i\}$ tel que I_i soit indépendant de $(I_j)_{j\notin\{i\}\cup\mathscr{V}_i}$. Alors,*

$$\|\mathscr{L}(S_n) - \mathsf{p}_\lambda\|_{\mathrm{VT}} \leqslant \frac{1-e^{-\lambda}}{\lambda} \sum_{k=1}^{n} \Big(\rho_k^2 + \sum_{j\in\mathscr{V}_k} \big(\rho_j\rho_k + \mathbb{E}[I_jI_k]\big)\Big).$$

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

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

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

$$\mathbb{P}(\forall j \in \mathscr{V}_k, I_j^{(k)} = b_j \,|\, \forall i \in [\![n]\!], I_i = a_i) = \mathbb{P}(\forall j \in \mathscr{V}_k, I_j = b_j \,|\, I_k = 1, \forall i \notin \overline{\mathscr{V}}_k, I_i = a_i).$$

La conclusion suit alors de

$$\begin{aligned}\mathbb{E}\big[|U_k - V_k|\big] &= \mathbb{E}\Big[\Big|I_k + \sum_{j\in\mathscr{V}_k}(I_j - I_j^{(k)})\Big|\Big] \leqslant \mathbb{E}\big[|I_k|\big] + \sum_{j\in\mathscr{V}_k}\big(\mathbb{E}\big[|I_j|\big] + \mathbb{E}\big[|I_j^{(k)}|\big]\big)\\ &= \rho_k + \sum_{j\in\mathscr{V}_k}\big(\rho_j + \mathbb{E}[I_j \,|\, I_k = 1]\big) = \rho_k + \sum_{j\in\mathscr{V}_k}\Big(\rho_j + \frac{\mathbb{E}[I_jI_k]}{\rho_k}\Big).\end{aligned}$$ □

5.2 Preuve du Théorème de Chen

5.2.1 L'équation de Chen

On cherche à borner

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

où $N \sim \mathrm{Poisson}(\lambda)$. 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 5.5. *Soit Z une variable aléatoire à valeurs dans $\mathbb{N}$ et $\lambda > 0$. Alors, $Z \sim \mathrm{Poisson}(\lambda)$ si et seulement si*

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

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

Au vu du lemme précédent[1], 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 $f : \mathbb{N} \to \mathbb{R}$ bornée. On peut dès lors se demander s'il n'existerait pas une fonction $f_A : \mathbb{N} \to \mathbb{R}$ bornée et satisfaisant l'**équation de Chen** :

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

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

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

Lemme 5.6. *Soit $A \subset \mathbb{N}$ et $\lambda > 0$. La fonction $f_A : \mathbb{N} \to \mathbb{R}$ définie par $f_A(0) := 0$ et*

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

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

On observe à présent que

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

La conclusion désirée suit donc du lemme suivant.

Lemme 5.7. *Soit $A \subset \mathbb{N}$ et $\lambda > 0$. Alors,*

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

□

5.2.2 Preuves des lemmes

Preuve du Lemme 5.5. Si $Z \sim \mathrm{Poisson}(\lambda)$, 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 $k \in \mathbb{N}$,

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

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

Preuve du Lemme 5.6. On pose $f_A(0) \coloneqq 0$, puis on procède via la relation de récurrence fournie par l'équation de Chen :

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

□

Preuve du Lemme 5.7. Par l'inégalité triangulaire, il suffit de montrer que

$$\forall i \in \mathbb{N}^*, \qquad |f_A(i+1) - f_A(i)| \leqslant \frac{1 - e^{-\lambda}}{\lambda}. \tag{5.1}$$

Observons d'une part que le Lemme 5.6 implique que $f_{A\cup B} = f_A + f_B$ lorsque $A \cap B = \varnothing$. En particulier,

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

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

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

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

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

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

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

La borne (5.1) est démontrée. □

5.3 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.

5.3.1 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 $k \leqslant m$ et $n \coloneqq \binom{m}{k}$. 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 $\mathcal{K} \coloneqq \{\alpha \subset [\![m]\!] \,|\, |\alpha| = k\}$; 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, $I_\alpha \sim \mathrm{Bern}(\rho_\alpha)$ avec $\rho_\alpha \equiv \rho \coloneqq 365^{-k+1}$, mais ne sont pas indépendantes. Le nombre de k-tuples dont tous les membres partagent la même date d'anniversaire est alors donné par $S \coloneqq \sum_{\alpha\in\mathcal{K}} I_\alpha$.

On souhaite appliquer le Corollaire 5.4. Pour ce faire, il faut introduire des voisinages de dépendance appropriés. Il est évident qu'avec le choix $\mathscr{V}_\alpha \coloneqq \{\beta \in \mathcal{K}\setminus\{\alpha\} \,|\, \alpha\cap\beta \neq \varnothing\}$, on a bien que $(I_\beta)_{\beta\notin\mathscr{V}_\alpha\cup\{\alpha\}}$ est indépendant de I_α. On a alors, pour tout $\beta \in \mathscr{V}_\alpha$,

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

Le Corollaire 5.4 donne alors

$$\|\mathscr{L}(S) - \mathsf{p}_\lambda\|_{\mathrm{VT}} \leqslant \frac{1-e^{-\lambda}}{\rho}\Big(\rho^2 + \sum_{\beta\in\mathscr{V}_\alpha}\big(\rho^2 + 365^{-|\alpha\cup\beta|+1}\big)\Big), \tag{5.2}$$

où $\lambda \coloneqq n\rho$ 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 $N \sim \mathrm{Poisson}(\lambda)$, $\mathbb{P}(N \geqslant 1) = 1 - \mathbb{P}(N = 0) = 1 - e^{-\lambda} = 1 - e^{-253/365} \cong 0{,}5$. La borne (5.2) donne dans ce cas $|\mathbb{P}(S > 0) - \mathbb{P}(N > 0)| \leqslant 0{,}059$ 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, $\|\mathscr{L}(S) - \mathsf{p}_\lambda\|_{\mathrm{VT}} \leqslant 0{,}0597$ 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 $0{,}1264$).

5.3.2 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 $(X_k)_{k\geqslant 1}$ 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 $k \geqslant 2$ si et seulement si

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

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

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

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

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

Lorsque $n \gg \ell \gg 1$, il semble naturel d'approximer S_n par une variable de Poisson de paramètre $\lambda \coloneqq \mathbb{E}[S_n] = 2^{-\ell}\big(\frac{1}{2}(n-1)+1\big)$. 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 $\ell = \log_2\big(\frac{1}{2}(n-1)+1\big) + c$, on obtient

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

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

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

On souhaite appliquer le Corollaire 5.4 afin d'évaluer la précision de cette approximation. On introduit, pour chaque $i \in [\![n]\!]$, le voisinage $\mathscr{V}_i \coloneqq \{j \in [\![n]\!] \setminus \{i\} \,|\, |i-j| \leqslant \ell\}$. Avec ce choix, $(I_j)_{j \notin \mathscr{V}_i \cup \{i\}}$ est bien indépendant de I_i. De plus,

$$\forall i \in [\![n]\!], \forall j \in \mathscr{V}_i, \qquad \mathbb{E}[I_i I_j] = 0.$$

En effet, comme $I_k = 1$ uniquement lorsque la série *débute* en k, il est impossible d'avoir $I_i = I_j = 1$ lorsque $|j-i| \leqslant \ell$. Le Corollaire 5.4 donne donc

$$\|\mathscr{L}(S_n) - \mathsf{p}_\lambda\|_{\mathrm{VT}} \leqslant \frac{1-e^{-\lambda}}{\lambda} \sum_{k=1}^{n} \Big(\rho_k^2 + \rho_k \sum_{j \in \mathscr{V}_k} \rho_j\Big) = \frac{1-e^{-\lambda}}{\lambda} \sum_{k=1}^{n} \rho_k \sum_{j \in \mathscr{V}_k \cup \{k\}} \rho_j.$$

En utilisant

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

(l'inégalité vient de la contrainte $j \in [\![n]\!]$) et la borne grossière

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

on obtient

$$\|\mathscr{L}(S_n) - \mathsf{p}_\lambda\|_{\mathrm{VT}} \leqslant 0{,}000054.$$

En particulier,

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

5.3.3 Le problème des rencontres

Le problème suivant a été proposé en 1708 par Pierre Rémond de Montmort [14] (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 ſort 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 $k \in [\![n]\!]$ telle que la k^{e} carte retournée soit la carte numéro k.

Reformulé en d'autre termes, Pierre perd si, en tirant au hasard uniformément une permutation de l'ensemble $[\![n]\!]$, 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 $[\![n]\!]$. On considère les variables aléatoires

$$\forall i \in [\![n]\!], \qquad I_i \coloneqq \mathbf{1}_{\{\sigma_n(i)=i\}}.$$

Ces variables aléatoires sont manifestement identiquement distribuées, $I_i \sim \mathrm{Bern}(\rho_i)$ avec $\rho_i \equiv \rho \coloneqq 1/n$, mais ne sont pas indépendantes (observez que si $\sigma_n(i) \neq i$, alors il existe au moins un autre j avec $\sigma_n(j) \neq j$). Notons $S_n \coloneqq \sum_{i=1}^n I_i$ 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 $n \to \infty$, vers la loi de Poisson de paramètre 1 :

$$\lim_{n\to\infty} \|\mathscr{L}(S_n) - \mathsf{p}_1\|_{\mathrm{VT}} = 0.$$

On souhaite appliquer le Théorème 5.1. Pour ce faire, étant donné $k \in [\![n]\!]$ et la permutation σ_n, considérons la permutation $\tilde{\sigma}_n^k \coloneqq \tau_{k,\sigma_n(k)} \circ \sigma_n$ 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 $U_k \coloneqq S_n$ et $V_k \coloneqq \sum_{i\in[\![n]\!]\setminus\{k\}} \mathbf{1}_{\{\tilde{\sigma}_n^k(i)=i\}}$. Alors, $1 + V_k \overset{\text{loi}}{=} S_n \,|\, \{I_k = 1\}$. Le Théorème 5.1 donne donc

$$\|\mathscr{L}(S_n) - \mathsf{p}_1\|_{\mathrm{VT}} \leqslant (1 - e^{-1})\frac{1}{n}\sum_{k=1}^n \mathbb{E}\big[|U_k - V_k|\big],$$

puisque $\rho_k = \rho = 1/n$ pour tout $k \in [\![n]\!]$ 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,

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

On obtient donc finalement

$$\|\mathscr{L}(S_n) - \mathsf{p}_1\|_{\mathrm{VT}} \leqslant \frac{2(1 - e^{-1})}{n}, \tag{5.3}$$

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

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

Notons que la borne (5.3) est loin d'être optimale : on peut montrer que $\|\mathscr{L}(S_n) - \mathsf{p}_1\|_{\mathrm{VT}} \leqslant \frac{2^n}{n!}$. 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.

5.3.4 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 $G \sim \mathsf{G}(n, p_n)$ (*cf.* Chapitre 8) 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 $\lim_{n\to\infty} np_n = \infty$, il suit des résultats des Sections 8.3 ou 4.3.1 que $\lim_{n\to\infty} \mathbb{P}(N_0 > \epsilon n) = 0$ 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. $(X_{ij})_{1\leqslant i<j\leqslant n}$ suivant chacune une loi de Bernoulli de paramètre p_n : deux sommets distincts $i, j \in [\![n]\!]$ sont connectés par une arête dans G si et seulement si $X_{ij} = 1$ (où l'on a posé $X_{ij} \coloneqq X_{ji}$ lorsque $i > j$).

En termes de ces variables aléatoires, le sommet $i \in [\![n]\!]$ est isolé si et seulement si

$$I_i \coloneqq \prod_{j\in[\![n]\!]\setminus\{i\}} (1 - X_{ij}) = 1.$$

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

Nous allons utiliser le Corollaire 5.3. Pour ce faire, introduisons, pour chaque $k \in [\![n]\!]$, les variables aléatoires

$$\forall i \in [\![n]\!] \setminus \{k\}, \qquad Z_i^{(k)} \coloneqq \prod_{j\in[\![n]\!]\setminus\{i,k\}} (1 - X_{ij}).$$

Manifestement, $(Z_i^{(k)})_{i\in[\![n]\!]\setminus\{k\}} \overset{\text{loi}}{=} (I_i)_{i\in[\![n]\!]\setminus\{k\}} \,|\, \{I_k = 1\}$. Une application du Corollaire 5.3 fournit donc

$$\|\mathscr{L}(S_n) - \mathsf{p}_\lambda\|_{\mathrm{VT}} \leqslant \frac{1-e^{-\lambda}}{\lambda} \sum_{k=1}^n \rho_k \Big(\rho_k + \sum_{i\in[\![n]\!]\setminus\{k\}} \mathbb{E}\big[|I_i - Z_i^{(k)}|\big]\Big),$$

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

$$\big|I_i - Z_i^{(k)}\big| = \Big|\prod_{j\in[\![n]\!]\setminus\{i\}} (1 - X_{ij}) - \prod_{j\in[\![n]\!]\setminus\{i,k\}} (1 - X_{ij})\Big| = X_{ik} \prod_{j\in[\![n]\!]\setminus\{i,k\}} (1 - X_{ij}).$$

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

$$\|\mathscr{L}(S_n) - \mathsf{p}_\lambda\|_{\mathrm{VT}} \leqslant \frac{1-e^{-\lambda}}{\lambda} n\rho\Big(\rho + (n-1)\rho\frac{p_n}{1-p_n}\Big) \leqslant \rho\Big(1 + n\frac{p_n}{1-p_n}\Big).$$

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

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

Dans ce cas,

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

On sait de l'inégalité triangulaire que

$$\|\mathscr{L}(S_n) - \mathsf{p}_{e^{-c}}\|_{\mathrm{VT}} \leqslant \|\mathscr{L}(S_n) - \mathsf{p}_\lambda\|_{\mathrm{VT}} + \|\mathsf{p}_\lambda - \mathsf{p}_{e^{-c}}\|_{\mathrm{VT}}.$$

Comme

$$\|\mathscr{L}(S_n) - \mathsf{p}_\lambda\|_{\mathrm{VT}} = \mathsf{O}\Big(\frac{\log n}{n}\Big)$$

et, par l'Exemple 4.11,

$$\|\mathsf{p}_\lambda - \mathsf{p}_{e^{-c}}\|_{\mathrm{VT}} \leqslant 1 - e^{-|\lambda - e^{-c}|} = \mathsf{O}\Big(\frac{(\log n)^2}{n}\Big),$$

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

$$\forall c \in \mathbb{R}, \qquad \lim_{n\to\infty} \|\mathscr{L}(S_n) - \mathsf{p}_{e^{-c}}\|_{\mathrm{VT}} = 0.$$

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

Dans ce cas,

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

On obtient donc

$$\begin{aligned} \|\mathscr{L}(S_n) - \mathsf{p}_{n^{1-c}}\|_{\mathrm{VT}} &\leqslant \|\mathscr{L}(S_n) - \mathsf{p}_\lambda\|_{\mathrm{VT}} + \|\mathsf{p}_\lambda - \mathsf{p}_{n^{1-c}}\|_{\mathrm{VT}} \\ &= \mathsf{O}\Big(\frac{\log n}{n^c}\Big) + \mathsf{O}\Big(\frac{(\log n)^2}{n^c}\Big) = \mathsf{O}\Big(\frac{(\log n)^2}{n^c}\Big). \end{aligned} \tag{5.4}$$

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

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

ce qui implique que $\mathscr{L}(\hat{Y}) \xrightarrow{n\to\infty} \mathcal{N}(0,1)$. Par (5.4)

$$\forall x \in \mathbb{R}, \qquad \mathbb{P}\big((S_n - n^{1-c})/\sqrt{n^{1-c}} \leqslant x\big) = \mathbb{P}\big(\hat{Y} \leqslant x\big) + \mathsf{O}\Big(\frac{(\log n)^2}{n^c}\Big).$$

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

Remarques bibliographiques : Ce chapitre est basé sur le livre [10] et l'article [5].

6 Concentration de la mesure : l'inégalité de Talagrand

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

$$\lim_{n\to\infty} \mathbb{P}(|s_n - \mathbb{E}[X_1]| > a_n) \leqslant \lim_{n\to\infty} \frac{\mathrm{Var}[s_n]}{a_n^2} = \lim_{n\to\infty} \frac{\mathrm{Var}[X_1]}{na_n^2} = 0.$$

En particulier, l'essentiel de la masse associée à la loi de $(X_1 + \cdots + X_n)/n$ 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 $X_1, \ldots, X_n$ 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.

6.1 Distance de Hamming pondérée

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

Définition 6.1. *Soit $\mathbb{S}_+^n \coloneqq \{\alpha \in \mathbb{R}_+^n \mid \|\alpha\|_2 = 1\}$. La **distance de Hamming pondérée par $\alpha \in \mathbb{S}_+^n$** sur Ω est définie par*

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

Étant donné $x \in \Omega$ et[1] $A \subset \Omega$, on notera

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

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

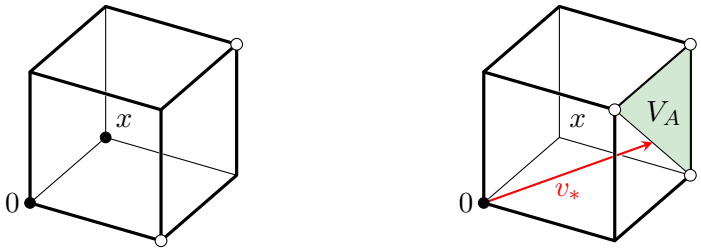


FIGURE 6.1: Illustration du cas où $n = 3$, $x = (0,1,0)$, $A = \{(1,0,0),(1,1,1)\}$. Gauche : représentation de cette situation (en blanc, les sommets de A). Droite : représentation de $U_A = \{(1,1,0),(1,0,1),(1,1,1)\}$ (sommets blancs), V_A (en gris) et du vecteur v_* (en rouge) réalisant le minimum dans le Lemme 6.3. 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 6.2. 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 $\mathcal{D}_A(x) \geqslant \min_{y\in A} \big|\{i \in [\![n]\!] \,|\, x_i \neq y_i\}\big| n^{-1/2}$, puisque cette borne revient à choisir $\alpha_i = n^{-1/2}$ pour chaque $i \in [\![n]\!]$.

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

Considérons l'ensemble

$$U_A(x) := \big\{ s = (s_1, \dots, s_n) \in \{0,1\}^n \,\big|\, \exists y \in A \text{ tel que } x_i \neq y_i \implies s_i = 1 \big\}. \tag{6.1}$$

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. 6.1). Nous pouvons à présent donner une caractérisation alternative de $\mathcal{D}_A(x)$.

Lemme 6.3. *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 $v \in V_A(x)$, $v\cdot v_* \geqslant \|v_*\|_2^2$. On choisit $\alpha_* = v_*/\|v_*\|_2$. Alors, pour tout $s \in U_A(x) \subset V_A(x)$, on a $\alpha_* \cdot s = \frac{1}{\|v_*\|_2} s\cdot v_* \geqslant \|v_*\|_2$. Par conséquent,

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

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 (6.1) va se révéler très utile dans la preuve de l'inégalité de Talagrand.

3. Pour la seconde identité, observons que, pour tout $y \in A$, $\sum_{i=1}^n \alpha_i \mathbf{1}_{\{x_i\neq y_i\}} \geqslant \min_{s\in U_A(x)} \sum_{i=1}^n \alpha_i s_i$, puisque $(\mathbf{1}_{\{x_i\neq y_i\}})_{i=1}^n \in U_A(x)$. On a donc $\inf_{y\in A} \sum_{i=1}^n \alpha_i \mathbf{1}_{\{x_i\neq y_i\}} \geqslant \min_{s\in U_A(x)} \sum_{i=1}^n \alpha_i s_i$. 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 $\sum_{i=1}^n \alpha_i s_i \geqslant \sum_{i=1}^n \alpha_i \mathbf{1}_{\{x_i\neq z_i\}} \geqslant \inf_{y\in A} \sum_{i=1}^n \alpha_i \mathbf{1}_{\{x_i\neq y_i\}}$. On a donc $\min_{s\in U_A(x)} \sum_{i=1}^n \alpha_i s_i \geqslant \inf_{y\in A} \sum_{i=1}^n \alpha_i \mathbf{1}_{\{x_i\neq y_i\}}$.

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

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

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

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

6.2 Inégalité de Talagrand

Pour $t \in \mathbb{R}_+^*$ et $A \subset \Omega$, définissons le t-épaississement de A par $A_t := \{x \in \Omega \,|\, \mathcal{D}_A(x) \leqslant t\}$. 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 6.4 (Inégalité de Talagrand). *Pour tout $A \subset \Omega$ mesurable et tout $t > 0$,*

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

Exemple 6.5. Dans cet exemple, considérons $\Omega := \{0,1\}^n$ muni de la loi uniforme.

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

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

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

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

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 $\mu(A) \geqslant 10^{-6}$. Alors,

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

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

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

puisque, si $\|x\|_1 > \frac{n}{2} + t\sqrt{n}$ et $y \in A$, alors $\sum_{i=1}^n \mathbf{1}_{\{x_i \neq y_i\}} = \|x - y\|_1 \geqslant \|x\|_1 - \|y\|_1 > t\sqrt{n}$. Notons que la borne ci-dessus reproduit qualitativement (les constantes sont moins bonnes) celle du Lemme 8.4. ◇

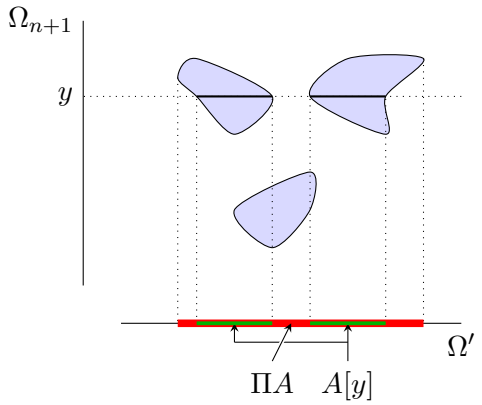


FIGURE 6.2: L'ensemble A (les 3 domaines en bleu) et les projections ΠA (en rouge) et $A[y]$ (en vert) sur Ω'.

6.3 Preuve de l'inégalité de Talagrand

Le Théorème 6.4 est une conséquence du résultat suivant.

Lemme 6.6. *Pour tout $A \subset \Omega$ mesurable et non vide, on a*

$$\int_\Omega \exp\left(\tfrac{1}{4}\mathcal{D}_A(x)^2\right) \mu(\mathrm{d}x) \leqslant \frac{1}{\mu(A)}.$$

Avant de démontrer ce lemme, voyons comment en déduire l'inégalité de Talagrand.

Preuve du Théorème 6.4. Soit X un élément aléatoire de Ω de loi μ. Il suit alors de l'inégalité de Markov et du Lemme 6.6 que, pour tout $t > 0$,

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

Preuve du Lemme 6.6. 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

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

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

Supposons donc que l'affirmation du lemme soit vérifiée pour $\Omega' \coloneqq \times_{i=1}^n \Omega_i$ (et la mesure $\mu' \coloneqq \times_{i=1}^n \mu_i$) et considérons $\Omega \coloneqq \Omega' \times \Omega_{n+1}$. 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. 6.2)

$$\begin{aligned} \forall y \in \Omega_{n+1}, \qquad A[y] &\coloneqq \{x \in \Omega' \,|\, (x, y) \in A\} \\ \Pi A &\coloneqq \{x \in \Omega' \,|\, \exists y \in \Omega_{n+1},\ (x, y) \in A\}. \end{aligned}$$

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 $i \in [\![n]\!]$ 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 $i \in [\![n + 1]\!]$ 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 (6.1) 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 $i \in [\![n]\!]$ 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 $i \in [\![n+1]\!]$ 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,

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

Il suit donc du Lemme 6.3 que

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

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

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

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

$$\begin{aligned}&\int_{\Omega'} \exp\bigl(\tfrac{1}{4}\mathcal{D}_A((x,y))^2\bigr)\, \mu'(\mathrm{d}x) \\ &\qquad \leqslant e^{(1-\lambda)^2/4} \int_{\Omega'} \Bigl(\exp\bigl(\tfrac{1}{4}\mathcal{D}_{\Pi A}(x)^2\bigr)\Bigr)^{1-\lambda} \Bigl(\exp\bigl(\mathcal{D}_{A[y]}(x)^2\bigr)\Bigr)^{\lambda} \mu'(\mathrm{d}x) \\ &\qquad \leqslant e^{(1-\lambda)^2/4} \biggl(\int_{\Omega'} \exp\bigl(\tfrac{1}{4}\mathcal{D}_{\Pi A}(x)^2\bigr)\, \mu'(\mathrm{d}x)\biggr)^{1-\lambda} \biggl(\int_{\Omega'} \exp\bigl(\mathcal{D}_{A[y]}(x)^2\bigr)\, \mu'(\mathrm{d}x)\biggr)^{\lambda},\end{aligned}$$

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

$$\int_{\Omega'} \exp\bigl(\tfrac{1}{4}\mathcal{D}_A((x,y))^2\bigr)\, \mu'(\mathrm{d}x) \leqslant e^{(1-\lambda)^2/4} \frac{1}{\mu'(\Pi A)^{1-\lambda} \mu'(A[y])^{\lambda}} = \frac{1}{\mu'(\Pi A)}\, e^{(1-\lambda)^2/4} r^{-\lambda},$$

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 $r \geqslant e^{-1/2}$ 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

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

En effet, lorsque $r \in [0, e^{-1/2})$, l'affirmation suit de $e^{1/4} \leqslant 2 - e^{-1/2}$. 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

$$\int_{\Omega'} \exp\bigl(\tfrac{1}{4}\mathcal{D}_A((x,y))^2\bigr)\, \mu'(\mathrm{d}x) \leqslant \frac{1}{\mu'(\Pi A)} \biggl(2 - \frac{\mu'(A[y])}{\mu'(\Pi A)}\biggr).$$

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

$$\begin{aligned}\int_{\Omega} \exp\bigl(\tfrac{1}{4}\mathcal{D}_A((x,y))^2\bigr)\, \mu'(\mathrm{d}x)\mu_{n+1}(\mathrm{d}y) &\leqslant \frac{1}{\mu'(\Pi A)} \biggl(2 - \frac{\int_{\Omega_{n+1}} \mu'(A[y])\mu_{n+1}(\mathrm{d}y)}{\mu'(\Pi A)}\biggr) \\ &= \frac{1}{\mu'(\Pi A)} \biggl(2 - \frac{\mu(A)}{\mu'(\Pi A)}\biggr) \leqslant \frac{1}{\mu(A)},\end{aligned}$$

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

6.4 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.

6.4.1 Concentration des fonctions lipschitziennes

Définition 6.7. *Soit $\Omega = \times_{i=1}^n \Omega_i$ et $c > 0$. Une fonction $F : \Omega \to \mathbb{R}$ est **c-lipschitzienne au sens de Talagrand** si, pour tout $x \in \Omega$, il existe $\alpha = \alpha(x) \in \mathbb{S}_+^n$ tel que*

$$\forall y \in \Omega, \qquad F(x) \leqslant F(y) + c d_\alpha(x, y).$$

Remarque 6.8. *Insistons sur le fait que la valeur de α (et donc la "distance" d_α utilisée) n'est pas fixée à 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.* ◇

On a alors le résultat suivant de concentration autour des médianes [5] .

Théorème 6.9. *Soit $\Omega = \times_{i=1}^n \Omega_i$ et $\mu = \times_{i=1}^n \mu_i$. Soit $F : \Omega \to \mathbb{R}$ c-Lipschitzienne au sens de Talagrand. Soit m_F une médiane pour F. Alors,*

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

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

$$A_t = \{x \in \Omega \,|\, \mathcal{D}_A(x) \leqslant t\} \subset \big\{x \in \Omega \,\big|\, F(x) \leqslant r + ct\big\}.$$

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

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

En particulier, pour tout $t > 0$,

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

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

$$\mu(F \geqslant m_F + u) \leqslant \frac{1}{\mu(F \leqslant m_F)} e^{-u^2/4c^2} \leqslant 2e^{-u^2/4c^2},$$
$$\mu(F \leqslant m_F - u) \leqslant \frac{1}{\mu(F \geqslant m_F)} e^{-u^2/4c^2} \leqslant 2e^{-u^2/4c^2},$$

et la conclusion suit. □

Exemple 6.10. Soit $X_1, \ldots, X_n$ des variables indépendantes, $X_i \sim \text{bernoulli}(p_i)$ pour chaque i. Dans le cadre de ce chapitre, cela revient à considérer $\Omega \coloneqq \{0, 1\}^n$ 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 $\mu(F \geqslant m_F) \geqslant 1/2$ et $\mu(F \leqslant m_F) \geqslant 1/2$.

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

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

où l'on a choisi $\alpha(x)\coloneqq x/\|x\|_2\in\mathbb{S}_+^n$ (et utilisé $\|x\|_2\leqslant\sqrt{n}$). Il suit donc du Théorème 6.9 que

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

Exemple 6.11. 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 $\Omega\coloneqq\{1,\dots,n\}^m$, $\mu\coloneqq\times_{i=1}^m\mu_i$ avec $\mu_i(\{k\})=1/n$ pour chaque $k\in[\![n]\!]$ et chaque $i\in[\![m]\!]$. Ainsi, pour $x\in\Omega$, x_i dénote l'urne dans laquelle est placée la i^{e} boule.

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

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

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

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

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

Il suit donc du Théorème 6.9 que

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

Le Théorème 6.9, 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 6.12. *1. Soit X une variable aléatoire réelle telle que*

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

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

$$\forall t\in\mathbb{R}_+^*,\qquad \mu(|X-\mathbb{E}[X]|\geqslant t)\leqslant e^{\pi a^2/4}\,e^{-t^2/4b}.$$

2. Soit X une variable aléatoire réelle telle que

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

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

Démonstration. Tout d'abord,

$$|m_X - \mathbb{E}[X]| = |\mathbb{E}[m_X - X]| \leqslant \mathbb{E}\big[|m_X - X|\big] = \int_{\mathbb{R}_+} \mathbb{P}(|m_X - X| > t)\,\mathrm{d}t \leqslant a\int_{\mathbb{R}_+} e^{-t^2/b}\,\mathrm{d}t = \frac{\sqrt{\pi}}{2}a\sqrt{b}.$$

Notons $t_0 \coloneqq \frac{\sqrt{\pi}}{2}a\sqrt{b}$. Si $t \geqslant 2t_0$,

$$\mu(|X - \mathbb{E}[X]| \geqslant t) \leqslant \mu(|X - m_X| \geqslant t - t_0) \leqslant \mu(|X - m_X| \geqslant t/2) \leqslant ae^{-t^2/4b}.$$

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

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

On obtient donc

$$\forall t \in \mathbb{R}_+^*, \qquad \mu(|X - \mathbb{E}[X]| \geqslant t) \leqslant (a \vee e^{\pi a^2/4})e^{-t^2/4b},$$

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

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

$$\int_{\mathbb{R}_+} e^{-t^2/b(m_X+t)}\,\mathrm{d}t \leqslant \int_0^{m_X} e^{-t^2/2bm_X}\,\mathrm{d}t + \int_{m_X}^{\infty} e^{-t/2b}\,\mathrm{d}t \leqslant \sqrt{\tfrac{1}{2}\pi b m_X} + 2be^{-m_X/2b}. \qquad \square$$

Il suit ainsi, par exemple, du Théorème 6.9 que $|m_F - \mathbb{E}[F]| \leqslant 4c\sqrt{\pi}$ et

$$\forall t \in \mathbb{R}_+^*, \qquad \mu(|F - \mathbb{E}[F]| \geqslant t) \leqslant e^{4\pi}\,e^{-t^2/16c^2}.$$

Exemple 6.13. Retournons à l'Exemple 6.11. L'espérance du nombre d'urnes non vides est donnée par

$$\mathbb{E}[F] = \big(1 - (1 - \tfrac{1}{n})^m\big)n.$$

L'estimée de concentration obtenue dans l'Exemple 6.11 implique donc

$$\forall t \in \mathbb{R}_+^*, \qquad \mu(|F - \mathbb{E}[F]| \geqslant t) \leqslant e^{4\pi}\,e^{-t^2/16(n\wedge m)}.$$

En particulier, dans la limite où $n \to \infty$ et $m/n \to \rho > 0$, on a $\lim_{n\to\infty} \mathbb{E}[F]/n = (1 - e^{-\rho})$ 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 $t_n \gg \sqrt{n}$. $\diamond$

6.4.2 Suprema de fonctions linéaires

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

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

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

Considérons $\Omega \coloneqq [0, 1]^n$, muni de la mesure produit μ des lois des variables Y_i. On souhaiterait appliquer le Théorème 6.9 à la fonction $F : \Omega \to \mathbb{R}$ définie par

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

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

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

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 6.9 conduit donc au résultat de concentration suivant.

Corollaire 6.14. *Soit m_Z une médiane de Z. Alors,*

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

Exemple 6.15 (**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 donnés $x, y \in S$, 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-évitants 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 6.14 à la variable aléatoire $Z' := -Z = \max_{\pi \in \Pi} \sum_{e \in A} (-\mathbf{1}_{\{e \in \pi\}}) Y_e$ conduit à

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

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 $C'n$ [32]. 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,

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

6.4.3 Fonctions certifiables

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

Définition 6.16. *Soit $\Omega = \times_{i=1}^n \Omega_i$ et $f : \mathbb{R} \to \mathbb{R}$. Une fonction $h : \Omega \to \mathbb{R}$ est **f-certifiable** si, pour tout $x \in \Omega$ tel que $h(x) \geqslant s$, il existe $I \subset [\![n]\!]$ avec $|I| \leqslant f(s)$ tel que*

$$\forall i \in I, y_i = x_i \qquad \Longrightarrow \qquad h(y) \geqslant s.$$

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

Exemple 6.17. Soit $n \in \mathbb{N}$ et $N := \binom{n}{2}$. On peut encoder le graphe d'Erdős–Rényi $\mathsf{G}(n, p)$ 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 $G \sim \mathsf{G}(n,p)$.

On s'intéresse à la fonction $G \mapsto h(G)$ comptant le nombre maximum de triangles distincts[6] dans G. Alors, h est 3-certifiable. En effet, si $h(G) \geqslant s$, alors on peut trouver s triangles dans G, comprenant au plus $3s$ arêtes, et toute autre réalisation G' possédant également ces arêtes satisfait $h(G') \geqslant s$. ◇

Théorème 6.18. *Soit $\Omega = \times_{i=1}^n \Omega_i$, $\mu = \times_{i=1}^n \mu_i$ et $f : \mathbb{N} \to \mathbb{N}$. Soit $h : \Omega \to \mathbb{R}$ une fonction f-certifiable telle que $|h(x) - h(y)| \leqslant c$ lorsque x et y diffèrent en au plus une coordonnée. Alors,*

$$\forall r \in \mathbb{R}, \forall t \in \mathbb{R}_+^*, \qquad \mu\big(h \leqslant r - ct\sqrt{f(r)}\big)\,\mu(h \geqslant r) \leqslant e^{-t^2/4}. \tag{6.2}$$

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

$$\mu\big(h \leqslant m_h - u\big) \leqslant 2e^{-u^2/4c^2 f(m_h)}, \qquad \mu\big(h \geqslant m_h + u\big) \leqslant 2e^{-u^2/4c^2 f(m_h+u)}. \tag{6.3}$$

Démonstration. Soit $t \in \mathbb{R}_+^*, r \in \mathbb{R}$ et $A := \big\{y \in \Omega \,\big|\, h(y) < r - ct\sqrt{f(r)}\big\}$. Soit $x \in \Omega$ tel que $h(x) \geqslant r$ et soit $I \subset [\![n]\!]$ tel que $|I| \leqslant f(r)$ certifiant que $h(x) \geqslant r$. 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 $d_\alpha(x, A) \leqslant t - \epsilon$. Alors, il existe $y \in A$ tel que

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

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

On a donc $\mu(h \geqslant r) \leqslant 1 - \mu(A_{t-\epsilon})$ et le Théorème 6.4 implique que

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

L'inégalité (6.2) suit en laissant $\epsilon \downarrow 0$, puis en exploitant la continuité du membre de droite, comme dans la preuve du Théorème 6.9.

Les deux conséquences énoncées en (6.3) sont obtenues en appliquant (6.2) 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. □

Exemple 6.19. Soit G un graphe d-régulier[7], 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 $|h(x) - h(y)| \leqslant 2$ lorsque x et y diffèrent en au plus une coordonnée. De plus, h est 1-certifiable, puisque l'existence d'au moins s sommets non isolés peut être certifiée en révélant la présence d'une

6. Un **triangle** de G est un triplet de sommets distincts $i, j, k \in [\![n]\!]$ 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.

arête incidente à chacun de s sommets non isolés distincts. On peut donc appliquer le Théorème 6.18, ce qui donne, pour toute médiane m_h de h,

$$\mu\big(h \leqslant m_h - u\big) \leqslant 2e^{-u^2/16m_h}, \qquad \mu\big(h \geqslant m_h + u\big) \leqslant 2e^{-u^2/16(m_h+u)}.$$

En particulier, en combinant ces bornes et le point 2 du Lemme 6.12, 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. ◇

Exemple 6.20 (**Plus longue sous-suite croissante**). Soit σ une permutation aléatoire uniforme de l'ensemble $[\![n]\!]$. Un problème très étudié est celui de déterminer la longueur $L = L(\sigma)$ de la plus grande sous-suite $(\sigma(i_1), \dots, \sigma(i_L))$, $1 \leqslant i_1 < i_2 < \cdots < i_L \leqslant n$, de la suite $(\sigma(1), \dots, \sigma(n))$ telle que $\sigma(i_1) < \sigma(i_2) < \cdots < \sigma(i_L)$.

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 $h : \Omega \to \mathbb{R}$ définie par

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

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 $\ell := \lceil \sqrt{n} \rceil$, $N := \lfloor n/\ell \rfloor$, $I_k := \big\{(k-1)\ell + 1, \dots, k\ell\big\}$ et $J_k := \big[(k-1)n^{-1/2}, kn^{-1/2}\big)$, $k \in \mathbb{N}^*$. 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

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

On utilise l'algorithme suivant pour construire une sous-suite croissante à partir de x : pour chaque $k \in [\![N]\!]$, 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 $Y := Y_1 + \cdots + Y_N$. Par conséquent, il suit du Lemme 8.6 que

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

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

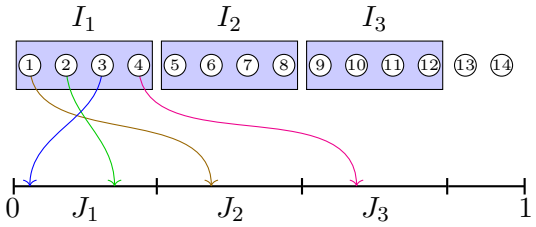


FIGURE 6.3: Extraction d'une sous-suite croissante. Dans chaque bloc I_k, on inclut le premier point $i \in I_k$ tel que $x_i \in J_k$ (s'il en existe au moins un). Pour des raisons de lisibilité, seul les images des points de I_1 sont indiquées. Le premier élément de la sous-suite croissante est donc 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 $i_1 < i_2 < \cdots < i_k$ soit croissante est égale à $1/k!$. Notons M_R le nombre de telles sous-suites de longueur au moins $R \in [\![n]\!]$. On a[8]

$$\mathbb{E}[M_R] = \sum_{k=R}^{n} \binom{n}{k} \frac{1}{k!} \leqslant \sum_{k=R}^{n} \Big(\frac{ne}{k}\Big)^k \Big(\frac{e}{k}\Big)^k = \sum_{k=R}^{n} \frac{n^k e^{2k}}{k^{2k}}.$$

8. On utilise les inégalités établies dans la note en bas de page 66.

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

$$\mathbb{E}[M_R] \leqslant \sum_{k=R}^{n} \frac{n^k e^{2k}}{R^{2k}} \leqslant \sum_{k=R}^{n} 4^{-2k} \leqslant \sum_{k=R}^{\infty} 16^{-k} \leqslant \frac{16}{15} 16^{-4e\sqrt{n}}.$$

Par conséquent, par l'inégalité de Markov,

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

ce qui tend également rapidement vers 0 lorsque $n \to \infty$ 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 $h(x) \geqslant s$, alors les s composantes formant une des sous-suites croissantes de longueur s certifient cette inégalité.

Il suit donc du Théorème 6.18 que

$$\mu\big(h \leqslant m_h - u\big) \leqslant 2e^{-u^2/4m_h}, \qquad \mu\big(h \geqslant m_h + u\big) \leqslant 2e^{-u^2/4(m_h+u)}.$$

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 [6] :

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

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

$$F_2(s) := \exp\Big(- \int_s^\infty (x - s) q^2(x) \, \mathrm{d}x \Big)$$

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

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

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 [46]. ◇

Remarques bibliographiques : Ce chapitre est basé sur les livres [4] et [36], ainsi que sur la source originale, l'article [50].

7 Réduction de la dimensionalité : 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 dimensionalité. 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 $\mathsf{O}(\log m)/\epsilon^2$ 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 7.3.

7.1 Cadre et énoncé du Lemme de Johnson–Lindenstrauss

Théorème 7.1 (Lemme de Johnson–Lindenstrauss)**.** *Soit $N, m \in \mathbb{N}^*$ et $\epsilon \in (0,1)$. Pour tout $n \geqslant 2(\epsilon^2 - \frac{1}{3}\epsilon^3)^{-1} \log m$ et tout ensemble $V \subset \mathbb{R}^N$ composé de m points, il existe une application $f : \mathbb{R}^N \to \mathbb{R}^n$ telle que*

$$\forall \mathbf{x}, \mathbf{y} \in V, \qquad (1-\epsilon)\|\mathbf{x}-\mathbf{y}\|_2 \leqslant \|f(\mathbf{x}) - f(\mathbf{y})\|_2 \leqslant (1+\epsilon)\|\mathbf{x}-\mathbf{y}\|_2. \tag{7.1}$$

Remarque 7.2. *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 (7.1) 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 $n \geqslant (2+\alpha)(\epsilon^2 - \frac{1}{3}\epsilon^3)^{-1} \log m$, 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 $1 - m^{-\alpha}$.* ◇

Remarque 7.3. *On peut montrer [35] qu'il n'est pas possible de faire mieux qu'une borne de la forme $n \geqslant \mathsf{O}(\epsilon^{-2}) \log m$).* ◇

7.2 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 $\mathbf{z} = (z_1, \dots, z_N) \in \mathbb{R}^N$,

$$\mathbb{E}\big[\|A\mathbf{z}\|_2^2\big] = \mathbb{E}\Big[\sum_{i=1}^{n}\Big(\sum_{j=1}^{N} A_{ij} z_j\Big)^2\Big] = \sum_{i=1}^{n}\sum_{j=1}^{N}\sum_{k=1}^{N} \mathbb{E}[A_{ij}A_{ik}] z_j z_k = n\|\mathbf{z}\|_2^2,$$

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 $\|Lz\|_2^2$ est concentrée au voisinage de son espérance.

Lemme 7.4. *Soit $\delta, \epsilon \in (0,1)$. Pour tout $n \geqslant |\log \delta|/(\epsilon^2 - \frac{1}{3}\epsilon^3)$, on a*

$$\forall \mathbf{z} \in \mathbb{R}^N \text{ tel que } \|z\|_2 = 1, \qquad \mathbb{P}\big(\big|\|L\mathbf{z}\|_2 - 1\big| \geqslant \epsilon\big) \leqslant 2\delta.$$

Avant de démontrer ce lemme, vérifions qu'il permet de conclure la preuve du Théorème 7.1.

On prend $\delta := 1/m^2$. Soit $\mathbf{x}, \mathbf{y} \in V$ distincts. Comme $n \geqslant 2(\epsilon^2 - \frac{1}{3}\epsilon^3)^{-1} \log m = |\log\delta|/(\epsilon^2 - \frac{1}{3}\epsilon^3)$, on peut appliquer le Lemme 7.4 à $\mathbf{z} := (\mathbf{x}-\mathbf{y})/\|\mathbf{x}-\mathbf{y}\|_2$, ce qui donne

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

Par conséquent, par l'inégalité de Boole,

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

On en conclut que la probabilité que L possède la propriété (7.1) est strictement positive, ce qui démontre l'existence d'une telle application.

Preuve du Lemme 7.4. Tout d'abord, pour tout $i \in [\![n]\!]$,

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

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

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

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

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

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

Par conséquent,

$$\forall t \in (-\infty, \tfrac{1}{2}), \qquad \mathbb{E}\big[e^{tW}\big] = \mathbb{E}\big[e^{tZ^2}\big]^n = (1-2t)^{-n/2}.$$

Bornons tout d'abord la probabilité que $\|L\mathbf{z}\|_2 \geqslant 1+\epsilon$. Une application de l'inégalité de Chernoff conduit à la borne

$$\mathbb{P}(\|L\mathbf{z}\|_2 \geqslant 1+\epsilon) = \mathbb{P}(\|A\mathbf{z}\|_2^2 \geqslant (1+\epsilon)^2 n) = \mathbb{P}(W \geqslant (1+\epsilon)^2 n) \leqslant \inf_{t>0} e^{-t(1+\epsilon)^2 n}\, \mathbb{E}\big[e^{tW}\big].$$

On en conclut que

$$\mathbb{P}(\|L\mathbf{z}\|_2 \geqslant 1+\epsilon) \leqslant \inf_{t\in(0,1/2)} \exp\big[-n\big(t(1+\epsilon)^2 + \tfrac{1}{2}\log(1-2t)\big)\big].$$

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

$$\mathbb{P}(\|L\mathbf{z}\|_2 \geqslant 1+\epsilon) \leqslant \exp\big[-n\big(\epsilon + \tfrac{1}{2}\epsilon^2 - \log(1+\epsilon)\big)\big] \leqslant \exp\big[-(\epsilon^2 - \tfrac{1}{3}\epsilon^3)n\big],$$

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

Bornons à présent la probabilité que $\|L\mathbf{z}\|_2 \leqslant 1-\epsilon$. On procède de manière similaire :

$$\mathbb{P}(\|L\mathbf{z}\|_2 \leqslant 1-\epsilon) = \mathbb{P}(-W \geqslant -(1-\epsilon)^2 n) \leqslant \inf_{t>0} e^{t(1-\epsilon)^2 n}\, \mathbb{E}\big[e^{-tW}\big] = \inf_{t>0} e^{t(1-\epsilon)^2 n}(1+2t)^{-n/2}.$$

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

$$\mathbb{P}(\|L\mathbf{z}\|_2 \leqslant 1-\epsilon) \leqslant \exp\big[-n\big(-\epsilon + \tfrac{1}{2}\epsilon^2 - \log(1-\epsilon)\big)\big] \leqslant \exp\big[-\epsilon^2 n\big],$$

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

On obtient donc finalement

$$\mathbb{P}(\|L\mathbf{z}\|_2 \notin (1-\epsilon, 1+\epsilon)) \leqslant \exp\big[-(\epsilon^2 - \tfrac{1}{3}\epsilon^3)n\big] + \exp\big[-\epsilon^2 n\big] \leqslant 2\exp\big[-(\epsilon^2 - \tfrac{1}{3}\epsilon^3)n\big] \leqslant 2\delta,$$

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

7.3 Application à l'acquisition comprimée

7.3.1 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.

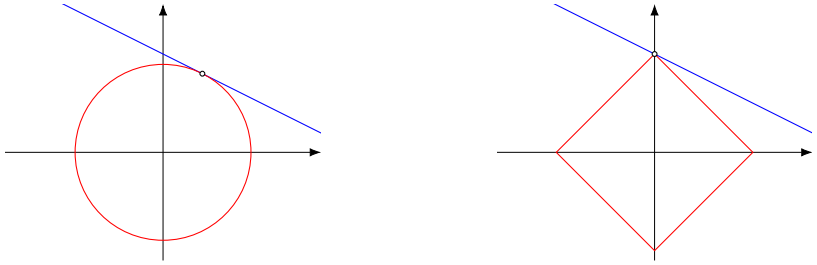


FIGURE 7.1: Minimiser la norme ℓ^1 sur un sous-espace affine tend à produire une solution sparse (droite), contrairement au cas de la norme ℓ^2 (gauche).

Définition 7.5. *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 $\mathscr{S}_k^N$ 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 $\mathbf{x}, \mathbf{x}' \in \mathscr{S}_k^N$, on a $\mathbf{x}-\mathbf{x}' \in \mathscr{S}_{2k}^N$. Par conséquent, il existe deux vecteurs distincts $\mathbf{x}, \mathbf{x}' \in \mathscr{S}_k^N$ 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 $\mathbf{x} \in \mathscr{S}_k^N$ tel que $L\mathbf{x} = \mathbf{y}$. Nous allons à présent décrire une approche plus efficace. Celle-ci repose sur la notion suivante.

Définition 7.6. *Soit $k \in \mathbb{N}^*$ et $\epsilon \in (0,1)$. Une application linéaire $M : \mathbb{R}^N \to \mathbb{R}^n$ possède la **propriété d'isométrie restreinte** d'ordre k et de constante ϵ si*

$$\forall \mathbf{x} \in \mathscr{S}_k^N, \qquad (1-\epsilon)\|\mathbf{x}\|_2 \leqslant \|M\mathbf{x}\|_2 \leqslant (1+\epsilon)\|\mathbf{x}\|_2.$$

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

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.

Lemme 7.7. *Soit $M : \mathbb{R}^N \to \mathbb{R}^n$ une application linéaire vérifiant $\mathsf{RIP}_{10k,1/3}$. Alors, pour chaque $\mathbf{x} \in \mathscr{S}_k^N$, 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 7.8. *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. 7.1.* ◇

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 $L := n^{-1/2}A$.

Lemme 7.9. *Pour tout $n \geqslant 500k \log N$, L vérifie $\mathsf{RIP}_{10k,1/3}$ avec probabilité au moins $1 - 1/N$.*

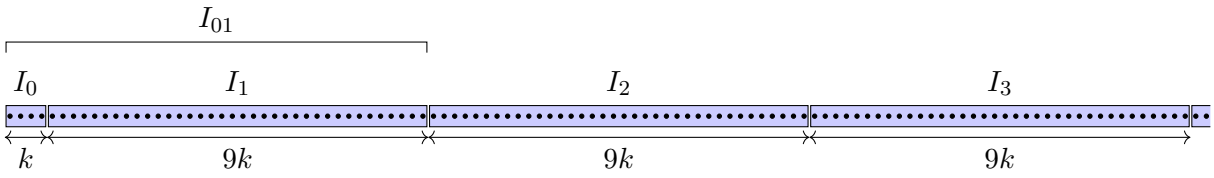


FIGURE 7.2: La décomposition de $\mathbb{N}^*$ employée dans la preuve du Lemme 7.7 (ici, avec $k = 4$).

En combinant les Lemmes 7.7 et 7.9, on obtient le résultat principal de cette section.

Théorème 7.10. *Soit $n \geqslant 500k \log N$. Avec probabilité au moins $1 - 1/N$, on peut reconstruire n'importe quel vecteur $\mathbf{x} \in \mathscr{S}_k^N$ à partir de la connaissance de $L\mathbf{x}$ en résolvant le problème d'optimisation*

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

7.3.2 Preuve des Lemmes 7.7 et 7.9

Preuve du Lemme 7.7. Fixons $\mathbf{x} = (x_1, \dots, x_N) \in \mathscr{S}_k^N$. Sans perte de généralité, nous supposerons que $x_i = 0$ pour tout $i > k$. Soit $\mathbf{z} \in \mathbb{R}^N$ tel que $M\mathbf{z} = M\mathbf{x}$ et $\|\mathbf{z}\|_1 \leqslant \|\mathbf{x}\|_1$. Notons $\mathbf{r} = (r_1, \dots, r_N) \coloneqq \mathbf{z} - \mathbf{x}$. Sans perte de généralité, on supposera que $|r_i| \geqslant |r_{i+1}|$ 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 $a_1 \coloneqq k$ et, pour $i \geqslant 2$, $a_i \coloneqq a_{i-1} + 9k$. On partitionne $\mathbb{N}^* = \bigvee_{i \geqslant 0} I_i$, avec $I_0 \coloneqq [\![k]\!]$ et, pour $i \geqslant 1$, $I_i \coloneqq \{a_i + 1, \dots, a_{i+1}\}$. On notera également $I_{\geqslant j} \coloneqq \bigvee_{j \geqslant i} I_j$ et $I_{01} \coloneqq I_0 \cup I_1$. 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 $(\mathbf{y}_J)_i \coloneqq \mathbf{1}_{\{i \in J\}} y_i$.

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

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

et, d'autre part,

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

En combinant ces deux observations, on conclut que

$$\|\mathbf{r}_{I_{\geqslant 1}}\|_1 \leqslant \|\mathbf{r}_{I_0}\|_1. \tag{7.2}$$

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

$$\tfrac{2}{3}\|\mathbf{r}_{I_{01}}\|_2 \leqslant \|M\mathbf{r}_{I_{01}}\|_2 \leqslant \sum_{i \geqslant 2} \|M\mathbf{r}_{I_i}\|_2 \leqslant \tfrac{4}{3} \sum_{i \geqslant 2} \|\mathbf{r}_{I_i}\|_2, \tag{7.3}$$

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

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

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

Ceci implique que

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

où l'on a utilisé (7.2). Par Cauchy–Schwarz, $\|r_{I_0}\|_1 \leqslant \sqrt{k}\|r_{I_0}\|_2$. On obtient donc finalement

$$\sum_{i\geqslant 2}\|\mathbf{r}_{I_i}\|_2 \leqslant \tfrac{1}{3}\|\mathbf{r}_{I_0}\|_2 \leqslant \tfrac{1}{3}\|\mathbf{r}_{I_{01}}\|_2.$$

Combiné avec (7.3), on en déduit que $\|\mathbf{r}_{I_{01}}\|_2 = 0$ et donc que $\mathbf{r}_{I_{01}} = \mathbf{0}$. Les entrées $|r_\ell|$, $\ell \geqslant k$, étant décroissantes, il suit que $\mathbf{r} = \mathbf{0}$, comme souhaité. □

Preuve du Lemme 7.9. Notons $k' := 10k$ et $\Lambda(\mathbf{z}) := \big|\|L\mathbf{z}\|_2 - 1\big|$.

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

$$\mathscr{S}_I := \big\{\mathbf{x} \in \mathscr{S}^N_{k'} \,\big|\, \mathbf{x}_{I^c} = \mathbf{0},\ \|\mathbf{x}\|_2 = 1\big\},$$

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

Le lemme sera démontrée une fois que l'on aura établi que

$$\mathbb{P}\Big(\sup\Big\{\Lambda(\mathbf{z}) \,\Big|\, \mathbf{z} \in \bigcup_{I\in\mathscr{I}} \mathscr{S}_I\Big\} \leqslant 1/3\Big) \geqslant 1 - \frac{1}{N}. \tag{7.4}$$

Il semble naturel d'utiliser le Lemme 7.4. Notons cependant que, contrairement au lemme de Johnson–Lindenstrauss qui ne s'applique qu'à des collections finies de points, l'événement intervenant dans (7.4) 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 $\mathscr{S}_I$.

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

$$\bigcup_{i=1}^{K} \big\{\mathbf{y} \in \mathscr{S}_I \,\big|\, \|\mathbf{y} - \mathbf{x}_i\|_2 \leqslant \rho\big\} = \mathscr{S}_I$$

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

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

sont disjointes ; leur volume total est donc égal à[2] $K c_{k'}(\rho/2)^{k'}$. 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 $K \leqslant (3/\rho)^{k'}$.

Nous allons à présent appliquer le Lemme 7.4 aux points de $\bigcup_{I\in\mathscr{I}(k',N)} D_I$. On fixe $\epsilon := 1/6$ et $\delta := \big(4N|\bigcup_{I\in\mathscr{I}} D_I|\big)^{-1}$. Notre hypothèse sur n et la borne grossière

$$\Big|\bigcup_{I\in\mathscr{I}} D_I\Big| \leqslant \binom{N}{k'}(3/\rho)^{k'} \leqslant \Big(\frac{3eN}{k'\rho}\Big)^{k'} \leqslant \Big(\frac{N}{\rho}\Big)^{k'}$$

impliquent que $|\log\delta|/(\epsilon^2 - \frac{1}{3}\epsilon^3) \leqslant 38(k'+2)\log N \leqslant 500k\log N \leqslant n$. Il suit donc du Lemme 7.4 que

$$\mathbb{P}\Big(\sup\Big\{\Lambda(\mathbf{z}) \,\Big|\, \mathbf{z} \in \bigcup_{I\in\mathscr{I}} D_I\Big\} \leqslant 1/6\Big) \geqslant 1 - \frac{1}{2N}. \tag{7.5}$$

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 (7.4) suivrait de (7.5), si nous étions en mesure de montrer que

$$\mathbb{P}\big(\forall \mathbf{y}, \mathbf{z} \in \mathbb{R}^N,\ |\Lambda(\mathbf{z}) - \Lambda(\mathbf{y})| \leqslant \sqrt{6N \log N}\, \|\mathbf{z} - \mathbf{y}\|_2\big) \geqslant 1 - \frac{1}{2N}. \tag{7.6}$$

En effet, la construction des ensembles D_I garantit que pour tout $\mathbf{y} \in \mathscr{S}_I$, il existe $\mathbf{x} \in D_I$ tel que $\|\mathbf{x} - \mathbf{y}\|_2 \leqslant \rho$. Les événements dans (7.5) et (7.6) se produisent avec probabilité au moins $1 - 1/N$ et lorsque c'est le cas,

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

Il nous suffit donc d'établir (7.6). Soit $\mathbf{y}, \mathbf{z} \in \mathbb{R}^N$. Tout d'abord, deux applications de l'inégalité du triangle permettent d'écrire

$$|\Lambda(\mathbf{z}) - \Lambda(\mathbf{y})| \leqslant \big|\|L\mathbf{z}\|_2 - \|L\mathbf{y}\|_2\big| \leqslant \|L(\mathbf{z} - \mathbf{y})\|_2.$$

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

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

Il nous reste à contrôler les valeurs typiques de $\|A\|_\infty$. Par une inégalité classique sur les queues de gaussiennes[3] et l'inégalité de Boole,

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

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

Remarques bibliographiques : Ce chapitre est basé sur [45].

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

8 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).

8.1 Le modèle d'Erdős–Rényi

Dans ce chapitre, nous nous intéresserons au modèle de graphe aléatoire $\mathsf{G}(n,p)$, $n \in \mathbb{N}^*$, $p \in [0,1]$. Il s'agit d'une loi de probabilité sur l'ensemble des graphes ayant pour sommets l'ensemble $[\![n]\!]$. Une réalisation $G \sim \mathsf{G}(n,p)$ du graphe aléatoire est alors obtenue de la façon suivante : indépendamment pour chacune des paires $\{i,j\} \subset [\![n]\!]$ 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 [23], la version introduite par Erdős et Rényi [18] étant légèrement différente (ils considèrent la loi uniforme sur tous les graphes avec sommets $[\![n]\!]$ 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 $p = 1/n$.

8.2 La transition de phase

Considérons à présent une réalisation $G \sim \mathsf{G}(n,p)$, où l'on suppose que $p = c/n$ pour un $c \in \mathbb{R}$ fixé. Le résultat principal de ce chapitre est le Théorème 8.1, 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 $\mathsf{O}(\log n)$ sommets avec grande probabilité [1].
- ▷ Lorsque $c > 1$, la plus grande composante connexe de G contient $\mathsf{O}(n)$ 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 chemin [2] de longueur $\mathsf{O}(n)$ lorsque $c > 1$.

1. Lorsqu'une propriété $\mathcal{P}$ du graphe aléatoire $G \sim \mathsf{G}(n,p)$ se produit avec probabilité $1 - \mathsf{o}_{n\to\infty}(1)$, 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 $k \in \{1, \dots, \ell - 1\}$.

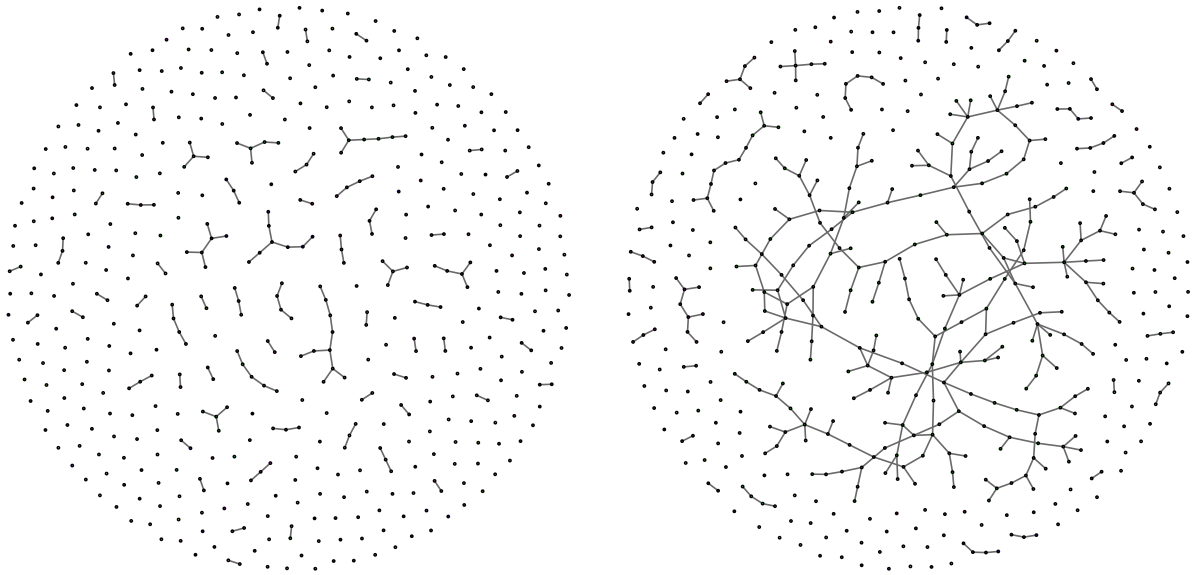


FIGURE 8.1: Deux réalisations du graphe aléatoire d'Erdős–Rényi avec $n = 500$. Gauche : phase sous-critique ($p = 0.5/n$). Droite : phase sur-critique ($p = 1.3/n$).

Ceci est formulé de façon précise dans le théorème suivant, dont la démonstration est donnée dans la Section 8.5.

Théorème 8.1. *Soit $\epsilon > 0$ suffisamment petit. Soit $G \sim \mathsf{G}(n,p)$.*

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 $\frac{1}{5}\epsilon^2 n$.*

8.3 Existence d'un chemin contenant presque tous les sommets

Comme seconde application des méthodes présentées dans ce chapitre, considérons à présent le graphe aléatoire $G \sim \mathsf{G}(n,p)$ 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 $\mathsf{O}(n)$) 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 8.2. *Pour tout $\epsilon > 0$ suffisamment petit, il existe $C = C(\epsilon)$ tel que l'affirmation suivante soit vraie. Si $p = C/n$ et $G \sim \mathsf{G}(n,p)$, alors G contient un chemin de longueur au moins $(1-\epsilon)n$ avec grande probabilité.*

La preuve de ce résultat est donnée en Section 8.6.

8.4 Préliminaires

Cette section contient quelques outils techniques qui se révéleront utiles pour la preuve des Théorèmes 8.1 et 8.2 : 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.

8.4.1 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 8.1 et 8.2 : 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 $S \coloneqq [\![n]\!]$. 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 $P = |a_1, \ldots, a_k\rangle$ pour désigner une pile composée des éléments $a_1, \ldots, a_k$ 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 $S = [\![n]\!]$
Initialisation : $T = \varnothing$, $P = \varnothing$, $I = S$
tant que $P \cup I \neq \varnothing$ **faire**
  si $P = \varnothing$ **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 8.3. Afin d'être certain d'avoir bien compris comment cet algorithme fonctionne, considérons par exemple le graphe représenté sur la Figure 8.2. 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 : $T = \varnothing, P = \varnothing, I = \{1, 2, 3, 4, 5, 6\}$.

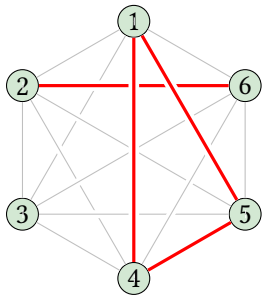


FIGURE 8.2: Le graphe discuté dans l'exemple 8.3.

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

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 : $T = \varnothing, P = |1,4\rangle, I = \{2,3,5,6\}$.

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 : $T = \varnothing, P = |1,4,5\rangle, I = \{2,3,6\}$.

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

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

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

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

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

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 complétera 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 $n \to \infty$, certaines propriétés garantissant que les affirmations des Théorèmes 8.1 et 8.2 sont vraies.

8.4.2 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 8.7 et 8.8 forment la partie probabiliste de la preuve du Théorème 8.1, le reste de la preuve étant alors purement déterministe. Commençons par une borne standard sur la queue supérieure d'une variable binomiale.

Lemme 8.4. *Soit $n \in \mathbb{N}^*$, $p \in [0,1]$, $X \sim \mathrm{Bin}(n,p)$. Alors,*

$$\forall \delta > 0, \qquad \mathbb{P}\big(X \geqslant (1+\delta)\mathbb{E}[X]\big) \leqslant \exp\Big(-\frac{\delta^2 pn}{2+\delta}\Big).$$

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

$$\mathbb{P}\big(X \geqslant (1+\delta)\mathbb{E}[X]\big) = \mathbb{P}(X \geqslant (1+\delta)np) = \mathbb{P}(e^{tX} \geqslant e^{t(1+\delta)np}) \leqslant e^{-t(1+\delta)np}\,\mathbb{E}[e^{tX}].$$

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

$$\mathbb{P}\big(X \geqslant (1+\delta)\mathbb{E}[X]\big) \leqslant \exp\Big(-\big((1+\delta)\log(1+\delta) - \delta\big)pn\Big) \leqslant \exp\Big(-\frac{\delta^2 pn}{2+\delta}\Big),$$

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

Remarque 8.5. *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*

$$\mathbb{P}\big(X \geqslant (1+\delta)\mathbb{E}[X]\big) \leqslant \exp\Big(-2\delta^2 p^2 n\Big),$$

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 6).

Lemme 8.6. *Soit $n \in \mathbb{N}^*$, $p \in [0,1]$, $X \sim \mathrm{Bin}(n,p)$. Alors,*

$$\forall \delta \in (0,1), \qquad \mathbb{P}\big(X \leqslant (1-\delta)\mathbb{E}[X]\big) \leqslant \exp\Big(-\frac{\delta^2 pn}{2}\Big).$$

Démonstration. Pour tout $t > 0$,

$$\mathbb{P}(X \leqslant (1-\delta)pn) = \mathbb{P}(e^{-tX} \geqslant e^{-t(1-\delta)pn}) \leqslant e^{t(1-\delta)pn}\mathbb{E}[e^{-tX}].$$

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

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

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

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

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

Lemme 8.7. *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,*

$$\nexists\, a, b \in [\![N]\!] \text{ tels que } b - a + 1 = kn \text{ et } \sum_{i=a}^{b} X_i \geqslant k. \tag{8.1}$$

Démonstration. Pour tout $a, b \in [\![N]\!]$ tels que $b - a + 1 = kn$, $\sum_{i=a}^{b} X_i \sim \mathrm{Bin}(kn, p)$. Soit donc $X \sim \mathrm{Bin}(kn, p)$. Notons que $\mathbb{E}[X] = kpn = (1-\epsilon)k \leqslant k/(1+\epsilon)$. Il suit donc du Lemme 8.4 que

$$\mathbb{P}(X \geqslant k) \leqslant \mathbb{P}\big(X \geqslant (1+\epsilon)\mathbb{E}[X]\big) \leqslant \exp\big(-\tfrac{1}{3}\epsilon^2(1-\epsilon)k\big),$$

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

$$\begin{aligned}\mathbb{P}\Big(\exists a, b \in [\![N]\!],\, b - a + 1 = kn,\, \sum_{i=a}^{b} X_i \geqslant k\Big) &\leqslant (N - kn + 1)\mathbb{P}(X \geqslant k) \\ &\leqslant \tfrac{1}{2}n^2 \exp\big(-\tfrac{1}{3}\epsilon^2(1-\epsilon)k\big) \leqslant \tfrac{1}{2}n^2 \exp\big(-\tfrac{7}{3}(1-\epsilon)\log n\big) = \mathsf{o}(1),\end{aligned}$$

puisque $\epsilon < 1/7$. □

Lemme 8.8. *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,*

$$\Big|\sum_{i=1}^{N_0} X_i - \frac{\epsilon(1+\epsilon)n}{2}\Big| \leqslant n^{2/3}. \tag{8.2}$$

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

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

8.5 Preuve du Théorème 8.1

Soit $(X_i)_{i=1}^N$ la collection de variables aléatoires déterminant le graphe $G \sim \mathsf{G}(n,p)$ 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 $t \geqslant |T| \cdot |I|$.

(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 $|P| \leqslant 1 + \sum_{i=1}^{t} X_i$ au temps t.

(O3) De plus, tant que $I \neq \varnothing$, 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 $I \neq \varnothing$ après t paires testées, alors on doit nécessairement avoir $|T \cup P| \geqslant \sum_{i=1}^{t} X_i$.

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

8.5.1 Régime sous-critique

Par le Lemme 8.7, nous pouvons supposer que les variables aléatoires $(X_i)_{i=1}^{N}$ satisfont (8.1). 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 $J \subset [\![N]\!]$ de longueur au plus kn et tel que $\sum_{i \in J} X_i \geqslant k$. Mais cela contredit notre hypothèse que la condition (8.1) est satisfaite.

8.5.2 Régime sur-critique

Par le Lemme 8.8, nous pouvons supposer que les variables aléatoires $(X_i)_{i=1}^{N}$ satisfont (8.2). 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 (8.2) satisfaite, il suit de l'observation (O2) que $|P| \leqslant 1 + \sum_{i=1}^{t} X_i \leqslant 1 + \sum_{i=1}^{N_0} X_i < n/3 - 1$, d'où l'on déduit que $|I| = n - |T| - |P| \geqslant n/3$. De plus, par l'observation (O1), on a testé au temps t au moins $|T| \cdot |I| \geqslant n^2/9 > N_0$ paires de sommets, ce qui est impossible puisque $t \leqslant N_0$.

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 $I \neq \varnothing$. 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 (8.2), 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 $|T \cup P| \geqslant \frac{1}{2}\epsilon(1+\epsilon)n - n^{2/3}$ au temps N_0. Comme on a supposé que $|P| < \frac{1}{5}\epsilon^2 n$, on conclut que $|T| \geqslant$

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

$$\frac{\epsilon n^2}{2} \geqslant N_0 \geqslant |T|\Big(n-|T|-\frac{\epsilon^2 n}{5}\Big).$$

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

$$N_0 \geqslant \Big(\frac{\epsilon n}{2}+\frac{3\epsilon^2 n}{10}-n^{2/3}\Big)\Big(n-\frac{\epsilon n}{2}-\frac{\epsilon^2 n}{2}+n^{2/3}\Big) = \frac{\epsilon n^2}{2}+\frac{\epsilon^2 n^2}{20}+\mathsf{O}(\epsilon^3)n^2+\mathsf{o}(n^2) > \frac{\epsilon n^2}{2},$$

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

8.6 Preuve du Théorème 8.2

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

Lemme 8.9. *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 $|T| = |I| \leqslant k - 1$ à cet instant. Comme $|T| + |P| + |I| = n$, on en conclut que $|P| \geqslant n - 2k + 2$. La conclusion suit, puisque les sommets de P forment toujours un chemin. □

Afin de démontrer le Théorème 8.2, il nous suffit donc de prouver que $G \sim \mathsf{G}(n,p)$ 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 $S_1 \cap S_2 = \varnothing$ 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

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

où l'on a utilisé les inégalités élémentaires $1-p \leqslant e^{-p}$ et[3] $\binom{n}{k} < \big(\frac{en}{k}\big)^k$. En choisissant $C := 5|\log\epsilon|/\epsilon$, on a $(en/k)^2 e^{-Ck/n} \leqslant 4e^2\epsilon^{1/2} < 1$ dès que $\epsilon < (2e)^{-4}$. On en conclut que la probabilité tend vers 0 exponentiellement rapidement avec n. □

Remarques bibliographiques : Ce chapitre est basé sur l'article [34].

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, $\binom{n}{k} \leqslant \frac{n^k}{k!} \leqslant \big(\frac{en}{k}\big)^k$.

9 La loi du demi-cercle

9.1 La problématique

L'étude du spectre de grandes matrices aléatoires a débuté dans les années 1930 avec les travaux de statisticiens tels que John Wishart. Ce sujet a ensuite pris une ampleur beaucoup plus importante dans les années 1950, lorsqu'Eugene Wigner, à l'occasion d'une conférence qu'il donnait au sujet du spectre nucléaire des grands atomes (tel l'uranium), suggéra de substituer à l'opérateur hamiltonien associé au noyau une matrice dont les éléments sont aléatoires, simplifiant ainsi considérablement l'analyse, avec l'intuition que les propriétés « grossières » du spectre ne devraient pas dépendre trop fortement des « détails » microscopiques. Cette idée s'est révélée très féconde, plusieurs expériences confirmant rapidement les prédictions faites avec son modèle. Depuis, cette approche a conduit au développement rapide d'un nouveau domaine, très actif, de recherche en physique théorique. Peut-être de façon plus surprenante, la théorie des matrices aléatoires s'est révélée extrêmement importante dans un grand nombre d'autres problèmes, allant des mathématiques pures (connexions avec la fonction zêta de Riemann) jusqu'à l'étude du risque financier, en passant par les systèmes intégrables, le chaos quantique, la gravitation quantique, diverses applications en statistiques, en traitement du signal, aux algèbres d'opérateurs, etc.

9.1.1 Les ensembles gaussiens

Wigner a introduit trois ensembles dans sa théorie des spectres nucléaires :

- ▷ l'**ensemble gaussien orthogonal**, GOE, utilisé pour décrire les systèmes invariants sous renversement du temps ;
- ▷ l'**ensemble gaussien unitaire**, GUE, utilisé pour décrire les systèmes non-invariants sous renversement du temps ;
- ▷ l'**ensemble gaussien symplectique**, GSE, utilisé pour décrire les systèmes avec spin.

L'ensemble GOE est donné par l'ensemble des matrices symétriques $\mathbf{A} = (A_{i,j})_{1 \leqslant i,j \leqslant n}$ réelles $n \times n$, muni de la mesure de probabilité de densité proportionnelle à

$$e^{-\frac{1}{4}\operatorname{Tr}\mathbf{A}^2} d\mathbf{A} = \exp\bigl(-\tfrac{1}{4}\sum_{1\leqslant i,j\leqslant n} A_{i,j}^2\bigr) \prod_{i=1}^{n} \mathrm{d}A_{i,i} \prod_{1\leqslant i<j\leqslant n} \mathrm{d}A_{i,j}.$$

L'adjectif « orthogonal » fait référence à l'invariance de cette loi sous l'action de la conjugaison $\mathbf{A} \mapsto \mathbf{OAO}^{\mathrm{t}}$ par une matrice orthogonale $\mathbf{O}$.

Les ensembles GUE et GSE sont définis similairement, en remplaçant les matrices symétriques par des matrices hermitiennes $n \times n$, resp. $2n \times 2n$, et en introduisant une loi invariante sous la conjugaison par des matrices unitaires, resp. symplectiques.

Un des résultats les plus célèbres dans cette théorie est le théorème dû à Wigner déterminant la loi du spectre de telles matrices dans la limite $n \to \infty$. C'est à une version de ce résultat que le reste de ce chapitre est consacré.

9.2 Énoncé de la loi du demi-cercle

On considère une famille de variables aléatoires (réelles) indépendantes $(X_{i,j})_{1 \leqslant i \leqslant j < \infty}$, d'espérance nulle et de variance 1, et on pose $X_{j,i} = X_{i,j}$ pour $i < j$. Nous supposerons, pour simplifier, que les variables aléatoires $X_{i,j}$ avec $i < j$ sont identiquement distribuées, et que les variables aléatoires $X_{i,i}$ sont également identiquement distribuées (mais éventuellement selon une autre loi). Pour $n \in \mathbb{N}^*$, on considère la matrice aléatoire $\mathbf{A}_n \coloneqq (X_{i,j})_{1 \leqslant i,j \leqslant n}$. Une telle matrice est appelée une **matrice de Wigner réelle**.

La matrice $\mathbf{A}_n$ étant symétrique, ses n valeurs propres sont réelles. Nous les noterons

$$\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n.$$

Le but de ce chapitre est de donner une description de la distribution asymptotique de ces valeurs propres. Comme pour tous les théorèmes limites, il est nécessaire de déterminer tout d'abord la normalisation appropriée afin d'obtenir une limite non triviale. Une façon de le faire ici est d'observer que

$$\frac{1}{n} \sum_{i=1}^{n} \lambda_i^2 = \frac{1}{n} \operatorname{Tr} \mathbf{A}_n^2 = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} X_{i,j} X_{j,i} = \frac{1}{n} \sum_{i,j=1}^{n} X_{i,j}^2.$$

La loi forte des grands nombres implique que cette quantité est d'ordre $\mathsf{O}(n)$ et par conséquent que les valeurs propres sont typiquement d'ordre $\mathsf{O}(\sqrt{n})$. Ceci suggère qu'il sera nécessaire de renormaliser les valeurs propres (ou, de manière équivalente, la matrice $\mathbf{A}_n$) par $1/\sqrt{n}$ si l'on désire obtenir une limite intéressante.

La **mesure spectrale empirique** associée à la matrice $\frac{1}{\sqrt{n}} \mathbf{A}_n$ est la mesure de probabilité aléatoire L_n sur $\mathbb{R}$ attribuant une masse $1/n$ à chaque valeur propre de $\frac{1}{\sqrt{n}} \mathbf{A}_n$,

$$L_n \coloneqq \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i/\sqrt{n}},$$

où δ_x représente la mesure de Dirac[1] avec masse en $x \in \mathbb{R}$. En d'autres termes, la moyenne d'une fonction f sous L_n est simplement la moyenne empirique

$$\int f(x) L_n(\mathrm{d}x) = \frac{1}{n} \sum_{i=1}^{n} f(\lambda_i/\sqrt{n}).$$

Le résultat principal de ce chapitre décrit le comportement de cette mesure de probabilité lorsque n devient grand (*cf.* Fig. 9.1 pour une illustration).

1. C'est-à-dire, $\delta_x(A) = \mathbf{1}_{\{x \in A\}}$, pour tout borélien A.

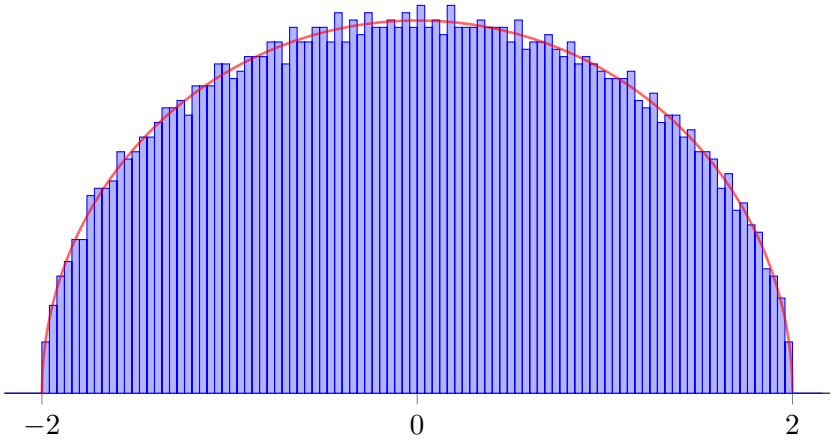


FIGURE 9.1: Un histogramme de la répartition des valeurs propres d'une matrice symétrique aléatoire 4000×4000, composée de variables aléatoires i.i.d. normales standards. En rouge : la loi du demi-cercle.

Théorème 9.1 (Théorème du demi-cercle de Wigner)**.** *Supposons la condition suivante satisfaite : pour tout $k \geqslant 1$,*

$$B_k := \mathbb{E}[|X_{1,1}|^k] \vee \mathbb{E}[|X_{1,2}|^k] < \infty.$$

*Soit μ la **loi du demi-cercle**,*

$$\mu(\mathrm{d}x) := \frac{1}{2\pi}\sqrt{4-x^2}\,\mathbf{1}_{\{|x|\leqslant 2\}}\,\mathrm{d}x.$$

Alors, presque sûrement, la mesure spectrale empirique L_n converge faiblement vers μ, c'est-à-dire

$$\lim_{n\to\infty}\int f(x)L_n(\mathrm{d}x) = \int f(x)\mu(\mathrm{d}x), \quad \textit{presque sûrement,}$$

pour toute fonction f continue et bornée.

9.3 Preuve de la loi du demi-cercle

Dans cette section, nous présentons une preuve du Théorème du demi-cercle de Wigner, le Théorème 9.1. La preuve est décomposée en trois parties.

9.3.1 Moments de la loi du demi-cercle

Lemme 9.2. *Pour $k \in \mathbb{N}$, soit $m_k := \int x^k \mu(\mathrm{d}x)$. Alors, pour tout $k \in \mathbb{N}$,*

$$m_{2k+1} = 0, \quad \textit{et} \quad m_{2k} = C_k,$$

*où les C_k, $k \geqslant 0$, sont les **nombres de Catalan**,*

$$C_k := \frac{1}{k+1}\binom{2k}{k}.$$

Démonstration. Le fait que les moments impairs soient nuls suit immédiatement de la symétrie de μ.

En effectuant le changement de variables $x = 2\sin\theta$, on obtient

$$\begin{aligned} m_{2k} &= \frac{1}{\pi}\int_0^2 x^{2k}\sqrt{4-x^2}\,\mathrm{d}x \\ &= \frac{2^{2k+2}}{\pi}\int_0^{\pi/2}\sin^{2k}\theta\cos^2\theta\,\mathrm{d}\theta \qquad (9.1)\\ &= \frac{2^{2k+2}}{\pi(2k+1)}\int_0^{\pi/2}\sin^{2k+2}\theta\,\mathrm{d}\theta \qquad (9.2)\\ &= \frac{2^{2k+2}}{\pi(2k+1)}\int_0^{\pi/2}\sin^{2k}\theta\,\mathrm{d}\theta - \frac{2^{2k+2}}{\pi(2k+1)}\int_0^{\pi/2}\sin^{2k}\theta\cos^2\theta\,\mathrm{d}\theta \qquad (9.3)\\ &= \frac{4(2k-1)}{2k+1}m_{2k-2} - \frac{1}{2k+1}m_{2k}, \qquad (9.4)\end{aligned}$$

où (9.2) résulte d'une intégration par partie, (9.3) de l'identité $\sin^2\theta = 1-\cos^2\theta$, et (9.4) des représentations (9.1) et (9.2) de m_{2k}. On en déduit que

$$m_{2k} = \frac{2(2k-1)}{k+1}m_{2k-2}.$$

Comme $m_0 = 1$, la conclusion suit en observant que les nombres de Catalan satisfont à la même relation de récurrence :

$$C_k = \frac{2(2k-1)}{k+1}C_{k-1}, \qquad C_0 = 1. \qquad \square$$

9.3.2 Convergence presque sûre des moments de L_n

Nous allons à présent montrer la convergence presque sûre des moments associés à la mesure spectrale empirique vers les moments m_k.

Lemme 9.3. *Sous les hypothèses du Théorème 9.1, on a, pour tout $k \in \mathbb{N}$,*

$$\lim_{n\to\infty}\mathbb{E}\left[\int x^k L_n(\mathrm{d}x)\right] = m_k, \qquad (9.5)$$

et, pour une constante $c(k) < \infty$,

$$\mathrm{Var}\left[\int x^k L_n(\mathrm{d}x)\right] \leqslant \frac{c(k)}{n^2}. \qquad (9.6)$$

En particulier, pour tout $k \in \mathbb{N}$, on a, presque sûrement,

$$\lim_{n\to\infty}\int x^k L_n(\mathrm{d}x) = m_k.$$

Démonstration. On commence par décomposer l'espérance de sorte à faire apparaître les moments des variables aléatoires $X_{i,j}$.

$$\begin{aligned}\mathbb{E}\left[\int x^k L_n(\mathrm{d}x)\right] &= \frac{1}{n}\mathbb{E}\Big[\sum_{i=1}^n\big(\lambda_i/\sqrt{n}\big)^k\Big] = \frac{1}{n}\mathbb{E}\big[\mathrm{Tr}\{(\frac{1}{\sqrt{n}}\mathbf{A}_n)^k\}\big] \\ &= n^{-\frac{k}{2}-1}\sum_{i_1,\dots,i_k=1}^n \mathbb{E}[X_{i_1,i_2}X_{i_2,i_3}\cdots X_{i_k,i_1}]. \qquad (9.7)\end{aligned}$$

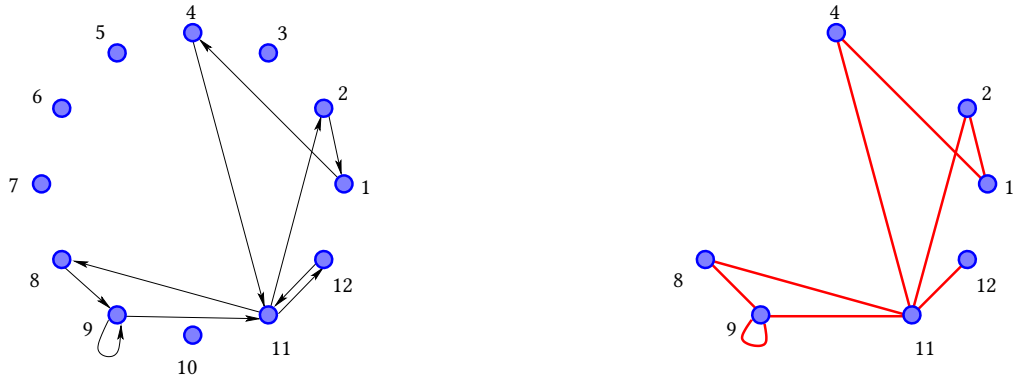


FIGURE 9.2: Illustration pour : $n = 12$, $k = 10$, $\mathbf{i} = (4, 11, 8, 9, 9, 11, 12, 11, 2, 1)$. Gauche : le chemin $\vec{\pi}(\mathbf{i})$ (qui part de 4). Droite : le graphe $G(\mathbf{i})$ correspondant.

On note $\mathbf{i} := (i_1, \dots, i_k)$ et $\rho_{\mathbf{i}} := \mathbb{E}[X_{i_1,i_2} X_{i_2,i_3} \cdots X_{i_k,i_1}]$. À chaque $\mathbf{i}$, on fait correspondre un chemin (orienté) fermé $\vec{\pi}(\mathbf{i})$ partant de i_1, puis visitant successivement les sommets $i_2, \dots, i_k, i_1$. Observez que le chemin peut passer plusieurs fois par le même sommet, qu'il peut effectuer plusieurs fois la même transition, et qu'il est également possible de faire du « sur place », c'est-à-dire d'avoir des transitions du type (i, i).

On associe ensuite au chemin $\vec{\pi}(\mathbf{i})$ le graphe connexe non orienté, $G(\mathbf{i}) = (V(\mathbf{i}), E(\mathbf{i}))$, avec sommets $V(\mathbf{i}) = \{i_1, \dots, i_k\}$ et une arête entre les sommets i et j de $V(\mathbf{i})$ si et seulement si le chemin $\vec{\pi}(\mathbf{i})$ contient (au moins une fois) la transition (i, j) ou la transition (j, i) (*cf.* Fig. 9.2).

Observons tout d'abord que l'indépendance des $X_{i,j}$ implique que

$$\rho_{\mathbf{i}} = \prod_{e \in E(\mathbf{i})} \mathbb{E}[X_e^{\ell_e(\mathbf{i})}],$$

où $\ell_e(\mathbf{i}) \in \{1, \dots, k\}$ est le nombre de passage de $\vec{\pi}(\mathbf{i})$ à travers l'arête $e \in E(\mathbf{i})$ (dans n'importe quelle direction) et où l'on a écrit $X_e \equiv X_{i,j}$ si $e = (i, j)$. Par conséquent,

$$|\rho_{\mathbf{i}}| \leqslant \big(1 \vee \sup_{1 \leqslant m \leqslant k} B_m\big)^k =: \bar{\rho}_k < \infty. \tag{9.8}$$

De plus, si $\rho_{\mathbf{i}} \neq 0$ et $e = (i, j) \in E(\mathbf{i})$, alors $\ell_e(\mathbf{i}) \geqslant 2$. En effet, si l'arête e n'était traversée qu'une seule fois, on aurait

$$\rho_{\mathbf{i}} = \mathbb{E}[X_e]\, \mathbb{E}[\prod_{\substack{e' \in E(\mathbf{i}) \\ e' \neq e}} X_{e'}^{\ell_{e'}(\mathbf{i})}] = 0,$$

puisque $\mathbb{E}[X_{i,j}] = 0$, pour tout $1 \leqslant i \leqslant j \leqslant n$. On en conclut immédiatement que $|E(\mathbf{i})| \leqslant \lfloor k/2 \rfloor$, lorsque $\rho_{\mathbf{i}} \neq 0$. Le lemme suivant, dont la démonstration est donnée plus bas, permet d'en déduire également une borne sur $V(\mathbf{i})$.

Lemme 9.4. *Soit $G = (V, E)$ un graphe fini connexe. Alors, $|V| \leqslant |E| + 1$, et l'égalité a lieu si et seulement si G est un arbre.*

Il suit du lemme que $|V(\mathbf{i})| \leqslant \lfloor k/2 \rfloor + 1$, lorsque $\rho_{\mathbf{i}} \neq 0$. On aura également besoin de la borne (grossière) suivante : comme on peut tirer ℓ éléments de $\{1, \dots, n\}$ d'au plus n^ℓ façons et, une fois tirés, y faire passer un chemin de longueur k d'au plus ℓ^k manières différentes, on a $\#\{\mathbf{i} \,|\, |V(\mathbf{i})| = \ell\} \leqslant n^\ell \ell^k$.

Supposons à présent k impair. Il suit de (9.7), (9.8) et de la borne $|V(\mathbf{i})| \leqslant \lfloor k/2 \rfloor + 1$ que

$$\left| \mathbb{E}\left[\int x^k L_n(\mathrm{d}x) \right] \right| = n^{-\frac{k}{2}-1} \left| \sum_{\mathbf{i}} \rho_{\mathbf{i}} \right| \leqslant n^{-\frac{k}{2}-1}\, \bar{\rho}(k) \sum_{\ell=1}^{\lfloor \frac{k}{2} \rfloor + 1} n^\ell \ell^k.$$

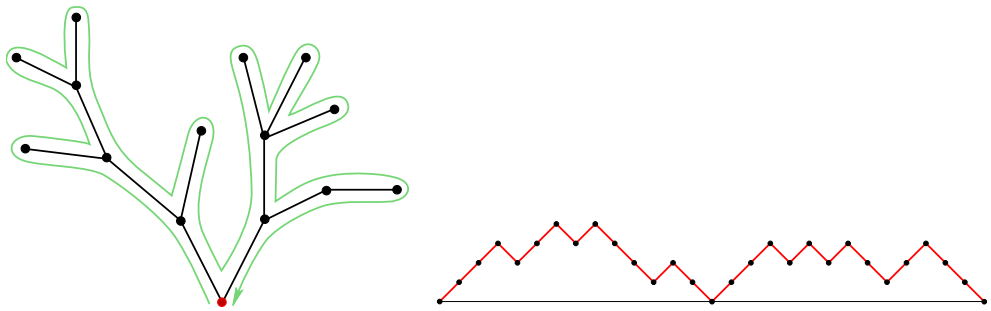


FIGURE 9.3: Gauche : le plongement de l'arbre $G(\mathbf{i})$ dans le plan, induit par le chemin $\vec{\pi}(\mathbf{i})$; la racine est donnée par le point de départ du chemin (i_1). Droite : l'excursion d'une marche aléatoire simple obtenue en appliquant la bijection à l'arbre de gauche.

Manifestement, $\sum_{\ell=1}^{\lfloor \frac{k}{2} \rfloor+1} n^\ell \ell^k \leqslant (\lfloor \frac{k}{2}+1 \rfloor)^k \sum_{\ell=1}^{\lfloor \frac{k}{2} \rfloor+1} n^\ell \leqslant 2(\lfloor \frac{k}{2}+1 \rfloor)^k \, n^{\lfloor \frac{k}{2} \rfloor+1}$, pour $n \geqslant 2$, et il existe donc une constante $M(k)$, indépendante de n, telle que

$$\mathbb{E}\left[\int x^k L_n(\mathrm{d}x)\right] \leqslant M(k)\, n^{\lfloor \frac{k}{2} \rfloor - \frac{k}{2}},$$

d'où l'on conclut que $\lim_{n\to\infty} \mathbb{E}[\int x^k L_n(\mathrm{d}x)] = 0$.

Il reste à considérer les valeurs paires de k. Le même argument montre que la contribution à $\mathbb{E}[\int x^k L_n(\mathrm{d}x)]$ des $\mathbf{i}$ tels que $|V(\mathbf{i})| < \frac{k}{2}+1$ disparaît dans la limite $n \to \infty$. Par conséquent, on peut se restreindre aux $\mathbf{i}$ tels que $|V(\mathbf{i})| = \frac{k}{2}+1$. Dans ce cas, comme $\frac{k}{2}+1 = |V(\mathbf{i})| \leqslant |E(\mathbf{i})|+1 \leqslant \frac{k}{2}+1$, il suit que $|E(\mathbf{i})| = k/2$. Le Lemme 9.4 implique donc que $G(\mathbf{i})$ est un arbre, et donc que $\vec{\pi}(\mathbf{i})$ parcourt précisément chaque arête de $G(\mathbf{i})$ deux fois (une fois dans chaque direction). On en déduit que, pour de tels $\mathbf{i}$,

$$\rho_{\mathbf{i}} = \prod_{(i,j)\in E(\mathbf{i})} \mathbb{E}[X_{i,j}^2] = 1.$$

Ceci implique que, pour k pair,

$$\lim_{n\to\infty} \mathbb{E}\left[\int x^k L_n(\mathrm{d}x)\right] = \lim_{n\to\infty} n^{-\frac{k}{2}-1} \#\big\{\mathbf{i} \,\big|\, |V(\mathbf{i})| = \tfrac{k}{2}+1\big\}.$$

Il ne reste donc plus qu'à déterminer le nombre de $\mathbf{i}$ avec $|V(\mathbf{i})| = \frac{k}{2}+1$. Comme on l'a vu ci-dessus, on peut associer à chaque $\mathbf{i}$ de ce type un arbre composé de $k/2$ arêtes et un chemin fermé $\vec{\pi}(\mathbf{i})$ partant de i_1 et parcourant tous les sommets de l'arbre en passant précisément deux fois par chaque arête (une fois dans chaque direction). L'information fournie par le chemin permet d'enraciner l'arbre (le point de départ du chemin tenant lieu de racine) et de le plonger dans le plan (l'orientation de l'arbre plongé étant donnée par le chemin), *cf.* Fig. 9.3. Étant donné un tel arbre planaire enraciné, on peut reconstruire les $\mathbf{i}$ correspondants simplement en étiquetant ses sommets à l'aide de $\frac{k}{2}+1$ éléments distincts de $\{1,\dots,n\}$.

Lemme 9.5. *Le nombre d'arbres planaires enracinés constitués de N arêtes est égal à C_N.*

La preuve est donnée plus bas. À l'aide de ce lemme, on obtient donc finalement (le facteur $\binom{n}{\frac{k}{2}+1}(\frac{k}{2}+1)!$ prenant en compte le choix des $\frac{k}{2}+1$ étiquettes associées aux différent sommets de l'arbre parmi les n à disposition)

$$\lim_{n\to\infty} \mathbb{E}\left[\int x^k L_n(\mathrm{d}x)\right] = \lim_{n\to\infty} n^{-\frac{k}{2}-1}\, C_{k/2} \binom{n}{\frac{k}{2}+1} (\frac{k}{2}+1)! = C_{k/2},$$

puisque, pour tout $k \in 2\mathbb{N}^*$ fixé, $\lim_{n\to\infty} n^{-\frac{k}{2}-1}\binom{n}{\frac{k}{2}+1}(\frac{k}{2}+1)! = \lim_{n\to\infty}\prod_{i=0}^{k/2}\frac{n-i}{n} = 1$.

Ceci conclut la preuve de (9.5).

Tournons-nous à présent vers l'estimation de la variance (9.6). On observe tout d'abord que

$$\begin{aligned}\operatorname{Var}\Big[\int x^k L_n(\mathrm{d}x)\Big] &= \frac{1}{n^2}\mathbb{E}\Big[\Big\{\operatorname{Tr}\big((\frac{1}{\sqrt{n}}\mathbf{A}_n)^k\big)\Big\}^2\Big] - \frac{1}{n^2}\Big\{\mathbb{E}\Big[\operatorname{Tr}\big((\frac{1}{\sqrt{n}}\mathbf{A}_n)^k\big)\Big]\Big\}^2 \\ &= n^{-k-2}\sum_{\substack{i_1,\dots,i_k=1\\ j_1,\dots,j_k=1}}^{n}(\rho_{\mathbf{i},\mathbf{j}} - \rho_{\mathbf{i}}\rho_{\mathbf{j}}),\end{aligned}$$

où l'on a introduit $\rho_{\mathbf{i},\mathbf{j}} \coloneqq \mathbb{E}[X_{i_1,i_2}X_{i_2,i_3}\cdots X_{i_k,i_1}X_{j_1,j_2}\cdots X_{j_k,j_1}]$. Similairement à ce que l'on a fait pour l'espérance, on associe à $\mathbf{i}$ et $\mathbf{j}$ les chemins fermés $\vec{\pi}(\mathbf{i})$ et $\vec{\pi}(\mathbf{j})$, et les graphes $G(\mathbf{i})$ et $G(\mathbf{j})$. On considère également le graphe $G(\mathbf{i},\mathbf{j}) = (V(\mathbf{i},\mathbf{j}), E(\mathbf{i},\mathbf{j}))$ avec sommets $V(\mathbf{i}) \cup V(\mathbf{j})$ et une arête entre $i, j \in V(\mathbf{i},\mathbf{j})$ si $(i,j) \in E(\mathbf{i})$ ou $(i,j) \in E(\mathbf{j})$.

Comme précédemment, si $\rho_{\mathbf{i},\mathbf{j}} - \rho_{\mathbf{i}}\rho_{\mathbf{j}} \neq 0$, il faut nécessairement que toute arête de $E(\mathbf{i},\mathbf{j})$ soit traversée soit au moins deux fois par $\vec{\pi}(\mathbf{i})$, soit au moins deux fois par $\vec{\pi}(\mathbf{j})$, soit au moins une fois par chacun. On en conclut que $|E(\mathbf{i},\mathbf{j})| \leqslant k$, et donc, par le Lemme 9.4 que $|V(\mathbf{i},\mathbf{j})| \leqslant k+1$.

Comme $|\rho_{\mathbf{i},\mathbf{j}} - \rho_{\mathbf{i}}\rho_{\mathbf{j}}| \leqslant 2\bar{\rho}_{2k}$ (*cf.* (9.8)) et que $\#\big\{\mathbf{i},\mathbf{j}\,\big|\,|V(\mathbf{i},\mathbf{j})| \leqslant k+1\big\} \leqslant 2(k+1)^{2k}n^{k+1}$, on en déduit immédiatement que $\operatorname{Var}[\int x^k L_n(\mathrm{d}x)] = \mathsf{O}(n^{-1})$. Ceci n'est cependant pas suffisant pour établir la convergence presque sûre (mais serait suffisant pour établir la convergence en probabilité), et il nous faut améliorer un peu cette estimée. Nous allons donc montrer que les paires $\mathbf{i},\mathbf{j}$ telles que $|V(\mathbf{i},\mathbf{j})| = k+1$ ne contribuent pas à la variance. On obtiendra alors la conclusion désirée, à savoir que la variance est d'ordre $\mathsf{O}(n^{-2})$.

Lorsque $|V(\mathbf{i},\mathbf{j})| = k+1$, le graphe $G(\mathbf{i},\mathbf{j})$ est un arbre, par le Lemme 9.4, et chacune de ses arêtes est parcourue précisément deux fois, et trois cas se présentent pour chaque arête : soit les deux traversées correspondantes se trouvent dans $\vec{\pi}(\mathbf{i})$, soit elles ont lieu dans $\vec{\pi}(\mathbf{j})$, soit une traversée a lieu dans chacun de ces deux chemins. En outre, pour obtenir une contribution strictement positive à la variance, il faut que cette dernière alternative ait lieu pour au moins une des arêtes, puisque, dans le cas contraire, $\rho_{\mathbf{i},\mathbf{j}} = \rho_{\mathbf{i}}\rho_{\mathbf{j}}$. Mais ceci est impossible, car $\vec{\pi}(\mathbf{i})$ est un chemin fermé sur $G(\mathbf{i},\mathbf{j})$ et doit donc visiter chaque arête précisément deux fois, $G(\mathbf{i},\mathbf{j})$ ne contenant pas de boucles (c'est un arbre !).

Nous sommes maintenant en mesure de démontrer la convergence presque sûre. La variance étant bornée par $c(k)/n^2$, une application de l'inégalité de Bienaymé-Tchebychev implique que

$$\mathbb{P}\Big(\Big|\int x^k L_n(\mathrm{d}x) - \mathbb{E}\Big[\int x^k L_n(\mathrm{d}x)\Big]\Big| > \epsilon\Big) \leqslant \frac{c(k)}{\epsilon^2 n^2},$$

et il suit donc du premier lemme de Borel–Cantelli que, presque sûrement,

$$\lim_{n\to\infty}\Big|\int x^k L_n(\mathrm{d}x) - \mathbb{E}\Big[\int x^k L_n(\mathrm{d}x)\Big]\Big| = 0. \qquad \square$$

Preuve du Lemme 9.4. Supposons tout d'abord que G soit un arbre. On supprime de E un sommet de degré 1 et l'unique arête qui lui est incidente, obtenant à nouveau un arbre. On peut itérer cette opération jusqu'à ce qu'il ne reste plus qu'un unique sommet (et aucune arête). On a donc bien $|V| = |E| + 1$.

Considérons à présent un graphe G qui n'est pas un arbre. Il doit donc exister au moins une arête $e \in E$ dont la suppression laisse G connexe ; on enlève e. On répète cette opération jusqu'à ce que le graphe obtenu soit un arbre (on dit que le graphe obtenu est un **arbre couvrant** de G). Comme l'on a dû enlever au moins une arête à G, il suit que dans ce cas $|V| < |E| + 1$. $\square$

Preuve du Lemme 9.5. Afin de compter les arbres planaires enracinés composés de N arêtes, on utilise le fait qu'ils sont en bijection avec les excursions positives de longueur $2N$ d'une marche aléatoire simple

sur $\mathbb{Z}$, c'est-à-dire les trajectoires $S_0, \dots, S_{2N}$ satisfaisant $S_0 = S_{2N} = 0$, $S_r \geqslant 0$ pour $0 < r < 2N$, et $|S_r - S_{r-1}| = 1$, pour tout $1 \leqslant r \leqslant 2N$. Cette bijection est obtenue en prenant pour S_r la distance dans l'arbre (c'est-à-dire le nombre d'arêtes du chemin reliant les deux sommets) entre i_1 et i_r, *cf.* Fig. 9.3.

Il nous suffit donc de compter ces trajectoires. Ceci peut être fait à l'aide du principe de réflexion. Considérons une trajectoire joignant $(0,0)$ à $(2N,0)$ et intersectant la ligne $y = -1$. Notons $\tau = \min\{1 \leqslant \ell \leqslant 2N-1 \,|\, S_\ell = -1\}$. On associe à cette trajectoire une nouvelle trajectoire en réfléchissant à travers la droite $y = -1$ la partie de la trajectoire initiale comprise entre les temps 0 et τ. On obtient ainsi une trajectoire joignant $(0,-2)$ et $(2N,0)$. Cette transformation étant évidemment inversible, elle définit une bijection entre ces deux types de trajectoires. Le nombre recherché et donc donné par

$$\binom{2N}{N} - \binom{2N}{N+1} = \frac{1}{N+1}\binom{2N}{N} = C_N. \qquad \square$$

9.3.3 Convergence faible de L_n

On fixe $B > 2$ et $\delta > 0$. Par le théorème d'approximation de Weierstrass, on peut trouver un polynôme P_δ tel que

$$\sup_{|x| \leqslant B} \big| f(x) - P_\delta(x) \big| \leqslant \delta.$$

Par conséquent,

$$\Big| \int f(x) L_n(\mathrm{d}x) - \int f(x)\mu(\mathrm{d}x) \Big| \leqslant 2\delta + \Big| \int P_\delta(x) L_n(\mathrm{d}x) - \int P_\delta(x)\mu(\mathrm{d}x) \Big| + \Big| \int_{|x|>B} (f(x) - P_\delta(x)) L_n(\mathrm{d}x) \Big|, \quad (9.9)$$

où l'on a utilisé le fait que μ est supportée sur $[-2,2]$.

Les moments de L_n convergeant presque sûrement vers ceux de μ, on a

$$\lim_{n\to\infty} \Big| \int P_\delta(x) L_n(\mathrm{d}x) - \int P_\delta(x)\mu(\mathrm{d}x) \Big| = 0, \qquad \text{presque sûrement.} \quad (9.10)$$

Notons p le degré du polynôme P_δ. f étant bornée, il existe une constante $K = K(\delta, B) < \infty$ telle que

$$\Big| \int_{|x|>B} (f(x) - P_\delta(x)) L_n(\mathrm{d}x) \Big| \leqslant K \int_{|x|>B} |x|^p L_n(\mathrm{d}x) \leqslant K B^{-p-2q} \int x^{2(p+q)} L_n(\mathrm{d}x),$$

où l'on a introduit un paramètre $q \in \mathbb{N}^*$ et utilisé l'inégalité $|x|^{p+2q} B^{-p-2q} \geqslant 1$ pour $|x| > B$. En utilisant une nouvelle fois la convergence des moments de L_n, on en déduit que

$$\limsup_{n\to\infty} \Big| \int_{|x|>B} (f(x) - P_\delta(x)) L_n(\mathrm{d}x) \Big| \leqslant K B^{-p-2q} \int x^{2(p+q)} \mu(\mathrm{d}x) \leqslant K B^{-p-2q} C_{p+q},$$

presque sûrement. Les nombres de Catalan satisfont l'inégalité $C_k \leqslant 2^{2k}$, pour tout $k \geqslant 0$. En effet, on a vu dans la preuve de Lemme 9.5 que ces derniers comptent le nombre d'excursions de longueur $2k$ d'une marche simple et, par conséquent, sont bornés par le nombre total de trajectoires de longueur $2k$ d'une marche simple, soit 2^{2k}. Comme $B > 2$, on obtient donc, en laissant $q \to \infty$,

$$\lim_{n\to\infty} \Big| \int_{|x|>B} (f(x) - P_\delta(x)) L_n(\mathrm{d}x) \Big| = 0, \qquad \text{presque sûrement.} \quad (9.11)$$

Finalement, en laissant $\delta \to 0$, on obtient de (9.9), (9.10) et (9.11) que

$$\lim_{n\to\infty} \Big| \int f(x) L_n(\mathrm{d}x) - \int f(x)\mu(\mathrm{d}x) \Big| = 0, \qquad \text{presque sûrement.}$$

Remarques bibliographiques : Ce chapitre est fortement inspiré de [25, Ch. 1]. Le Théorème 9.1, sous une forme un peu moins générale, est originellement dû à [53].

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

10.1 Transformée de Doob d'une chaîne de Markov

Soit $(X_n)_{n\geqslant 0}$ 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 $h : S \to \mathbb{R}_+$ une fonction strictement positive sur A et P-harmonique sur A :

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

On définit alors

$$\forall i \in A,\ \forall j \in S, \qquad q(i,j) \coloneqq \frac{h(j)}{h(i)}p(i,j).$$

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 10.1. *Soit $i_0 \in A$. La chaîne de Markov $(Y_n)_{n\geqslant 0}$ 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 $i_0, i_1, \ldots, i_{n-1}, i_n$, avec $i_0, \ldots, i_{n-1} \in A$ et $i_n \in S$, est donnée par

$$\begin{aligned}\mathbb{Q}_{i_0}(Y_0 = i_0, \ldots, Y_n = i_n) &= \prod_{k=1}^{n} q(i_{k-1}, i_k) \\ &= \frac{h(i_n)}{h(i_0)}\prod_{k=1}^{n} p(i_{k-1}, i_k) = \frac{h(i_n)}{h(i_0)}\mathbb{P}_{i_0}(X_0 = i_0, \ldots, X_n = i_n). \quad (10.1)\end{aligned}$$

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.*

10.2 Quelques exemples

10.2.1 Processus conditionné à quitter A par un certain état

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

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

où $a \in S \setminus A$ et $\tau_B \coloneqq \min\{n \geqslant 0 \,|\, X_n \in B\}$. 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 $i \in A$,

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

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).

Théorème 10.2. *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 (10.1) : pour toute trajectoire finie $i_0, i_1, \ldots, i_{n-1}, i_n$, avec $i_0, \ldots, i_{n-1} \in A$ et $i_n = a$,

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

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.

10.2.2 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 $h : S \to \mathbb{R}_+$ définie par

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

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 10.1 et pouvons considérer la h-transformée de Doob (Y_n).

Théorème 10.3. *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 (10.1) : 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 $0 \leqslant k \leqslant n$),

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

puisque, par la propriété de Markov,

$$\mathbb{P}_{i_n}(\tau_{S\setminus A} = \infty) = \mathbb{P}_{i_0}\big(\tau_{S\setminus A} = \infty \;\big|\; (X_0, \dots, X_n) = (i_0, \dots, i_n)\big). \qquad \square$$

10.2.3 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 $h : \{0, \dots, N\} \times S \to \mathbb{R}_+$ donnée par

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

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

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

lorsque $h(n, i) \neq 0$ et $q_n(i, j) \coloneqq 0$ 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

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

démarrée en i_0 et stoppée au temps N.

Théorème 10.4. *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 $0 \leqslant n \leqslant N-1$ (observons que la probabilité de toute trajectoire de (Y_n) ne satisfaisant pas cette dernière condition est nécessairement nulle). Alors, on a bien

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

10.3 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 $(S_n)_{n\geqslant 0}$ 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.

10.3.1 Deux fonctions (super)harmoniques

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

$$h^{\downarrow}(i) \coloneqq \mathbb{P}_0(\tau_{-i}<\infty), \qquad h^{\uparrow}(i)\coloneqq h^{\downarrow}(0)+\cdots+h^{\downarrow}(i).$$

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

Lemme 10.5. *Pour tout $i\in\mathbb{N}$, $h^{\downarrow}(i)=\alpha^i$ et $h^{\uparrow}(i)=\sum_{k=0}^{i}\alpha^k$. De plus,*

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

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

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

On en déduit que $h^{\downarrow}(i)=\alpha^i$, pour tout $i\in\mathbb{N}^*$ et, par conséquent, $h^{\uparrow}(i)=1+\alpha+\cdots+\alpha^i$. 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 $\mu(-1)\geqslant\mu(1)$.

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, $\lim_{n\to\infty}S_n/n=\mu(1)-\mu(-1)=:c>0$ presque sûrement. En particulier,

$$\lim_{N\to\infty}\mathbb{P}_0\big(\forall n\geqslant N,\ S_n>\tfrac{1}{2}cn\big)=\mathbb{P}_0\big(\exists N\in\mathbb{N},\ \forall n\geqslant N,\ S_n/n>c/2\big)=1.$$

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

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

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

Lemme 10.6. *$h^{\downarrow}$ est μ-harmonique sur $\mathbb{N}^*$ et $h^{\uparrow}$ est μ-superharmonique sur $\mathbb{N}$:*

$$\forall i \in \mathbb{N}^*, \qquad h^{\downarrow}(i) = \sum_{j\in\{-1,0,1\}} \mu(j) h^{\downarrow}(i+j),$$
$$\forall i \in \mathbb{N}, \qquad h^{\uparrow}(i) \geqslant \sum_{j\in\{-1,0,1\}} \mu(j) h^{\uparrow}(i+j).$$

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

Démonstration. La μ-harmonicité de $h^{\downarrow}$ suit immédiatement du Lemme 10.5, mais peut être déduite de façon plus transparente de la propriété de Markov : pour tout $i \in \mathbb{N}^*$,

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

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 $\mu(-1) \geqslant \alpha\mu(1)$, 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 $\mu(-1) \leqslant \mu(1)$. L'extension de la μ-(super)harmonicité de $h^{\uparrow}$ à $\mathbb{N}$ est immédiate :

$$\mu(-1)h^{\uparrow}(-1) + \mu(0)h^{\uparrow}(0) + \mu(1)h^{\uparrow}(1) = \mu(0) + (1+\alpha)\mu(1) \leqslant 1 = h^{\uparrow}(0),$$

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

Le Lemme 10.6 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 10.1.

10.3.2 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 10.2.1, avec $X_n := S_n$, $i_0 \in \mathbb{N}^*$, $A := \mathbb{N}^*$ et $a := 0$. Une application du Théorème 10.2 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.

10.3.3 La $h^\uparrow$-transformée de Doob lorsque $\mu(1) \geqslant \mu(-1)$

Sous l'hypothèse que $\mu(1) \geqslant \mu(-1)$, 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.

Lemme 10.7. *On suppose que $\mu(1) \geqslant \mu(-1)$. Alors, la chaîne de Markov $(S_n^\uparrow)$ est transiente.*

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

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

la seconde identité suivant de (10.1).

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 $\mathbb{P}_i^\uparrow(\tau_{i-1} < \infty) < 1$. □

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

Le théorème suivant montre que, lorsque $\mu(1) > \mu(-1)$, on se trouve dans le cadre de la Section 10.2.2 : la fonction $h^\uparrow$ ne diffère de la fonction harmonique de la Section 10.2.2 que par une constante multiplicative strictement positive. Notons que l'hypothèse $\mu(1) > \mu(-1)$ garantit que $\mathbb{P}_0(\tau_{\mathbb{Z}_{<0}} = \infty) = 1 - \alpha > 0$.

Théorème 10.8. *Lorsque $\mu(1) > \mu(-1)$, on a*

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

Démonstration. Pour tout $i \geqslant 0$,

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

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,

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

Une application du Théorème 10.3 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 : $\mu(1) = \mu(-1)$

Le Lemme 10.6 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

$$\Lambda_N := \{S_n \geqslant 0 \text{ pour tout } 0 \leqslant n \leqslant N\},$$

ce dernier ayant probabilité strictement positive. Commençons par montrer un lemme technique, qui est un pendant du Théorème 10.8 (la preuve est d'ailleurs presque identique).

Lemme 10.9. *Lorsque $\mu(1) = \mu(-1)$, on a*

$$\forall i \in \mathbb{N}, \qquad h^\uparrow(i) \leqslant \liminf_{N\to\infty} \frac{\mathbb{P}_i(\Lambda_N)}{\mathbb{P}_0(\Lambda_N)}.$$

Démonstration. Pour tout $i \geqslant 0$,

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

Par conséquent,

$$\liminf_{N\to\infty} \frac{\mathbb{P}_i(\Lambda_N)}{\mathbb{P}_0(\Lambda_N)} \geqslant \lim_{N\to\infty} \sum_{k=0}^{i} \mathbb{P}_0(\tau_{-k} \leqslant N) = \sum_{k=0}^{i} \mathbb{P}_0(\tau_{-k} < \infty) = h^\uparrow(i). \qquad \square$$

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

Théorème 10.10. *On suppose que $\mu(-1) = \mu(1)$. Alors, pour tout $i_0 = 0, i_1, \dots, i_n \geqslant 0$,*

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

Démonstration. Pour tout $i_0 = 0, i_1, \dots, i_n \in \mathbb{N}$ et tout $N \geqslant n$, il suit de la propriété de Markov et de (10.1) que

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

Le Lemme 10.5 implique donc que

$$\liminf_{N\to\infty} \mathbb{P}_0(S_0 = i_0, \dots, S_n = i_n \,|\, \Lambda_N) \geqslant \mathbb{P}_0^{\uparrow}(S_0^{\uparrow} = i_0, \dots, S_n^{\uparrow} = i_n). \tag{10.2}$$

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

$$\limsup_{N\to\infty} \mathbb{P}_0(S_0 = j_0, \dots, S_n = j_n \,|\, \Lambda_N) = \mathbb{P}_0^{\uparrow}(S_0^{\uparrow} = j_0, \dots, S_n^{\uparrow} = j_n) + c. \tag{10.3}$$

Par (10.2), il existe $N_0 \in \mathbb{N}$ tel que $\forall N \geqslant N_0$,

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

Par (10.3), il existe $N \geqslant N_0$ tel que

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

On obtient donc

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

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. □

Remarques bibliographiques : Ce chapitre s'inspire largement du chapitre 5 des notes de cours [13]. Pour le lecteur intéressé, ces dernières traitent de marches aléatoires générales sur $\mathbb{Z}$, pas uniquement de celles dont les incréments prennent valeur dans $\{-1, 0, 1\}$.

11 Chaînes de Markov réversibles et réseaux électriques

Soit S un ensemble fini et $(X_n)_{n\geqslant 0}$ une chaîne de Markov irréductible sur S, de matrice de transition $P = (p_{ij})_{i,j\in S}$. Comme cela a été vu dans le cours d'introduction à la théorie des probabilités, il y a dans ce cas une unique distribution stationnaire $\pi = (\pi_i)_{i\in S}$ et celle-ci satisfait $\pi_i > 0$ pour tout $i \in S$. Dans le reste de ce chapitre, on supposera la chaîne $(X_n)_{n\geqslant 0}$ réversible (par rapport à π), c'est-à-dire telle que

$$\pi_i p_{ij} = \pi_j p_{ji}, \qquad \forall i, j \in S.$$

On notera $i \sim j$ si les états $i, j \in S$ sont voisins, c'est-à-dire tels que $p_{ij} > 0$ (et donc également $p_{ji} > 0$).

Étant donnée une telle chaîne, on peut construire un graphe pondéré connexe (S, A, c), avec sommets S, ensemble d'arêtes (non orientées)

$$A \coloneqq \big\{(i,j) \,\big|\, i,j \in S,\, i \sim j\big\},$$

et une fonction de valuation $c : A \to (0,\infty)$ associant à toute arête $e = (i,j) \in A$ le poids

$$c_e \equiv c_{ij} = \pi_i p_{ij} = \pi_j p_{ji}.$$

Inversement, si l'on se donne un graphe pondéré connexe, avec au plus une arête (non orientée) entre deux sommets donnés (mais avec la possibilité d'avoir une arête connectant un sommet avec lui-même), (S, A, c), avec S fini, on peut construire une chaîne de Markov réversible $(X_n)_{n\geqslant 0}$ sur S en posant, pour $(i,j) \in A$,

$$p_{ij} = \frac{c_{ij}}{c[i]},$$

où $c[i] \coloneqq \sum_{j\sim i} c_{ij}$. La chaîne ainsi définie satisfait

$$c[i]p_{ij} = c_{ij} = c[j]p_{ji},$$

et est donc bien réversible et de distribution stationnaire $\pi_i = c[i]/\sum_{j\in S} c[j]$, $\forall i \in S$.

11.1 Réseaux de conductances

11.1.1 Flots

Définition 11.1. *Un **flot** ϕ sur un graphe (S, A) est une fonction $\phi : S \times S \to \mathbb{R}$, $(i,j) \mapsto \phi_{ij}$, telle que $\phi_{ij} = -\phi_{ji}$ si $(i,j) \in A$, et $\phi_{ij} = 0$ sinon. Le **flot sortant de $i \in S$** est $\phi[i] = \sum_{j\sim i} \phi_{ij}$.*

On a donc, en particulier,

$$\sum_{i\in S}\phi[i]=\sum_{i\in S}\sum_{j\sim i}\phi_{ij}=\sum_{(i,j)\in A}\big(\phi_{ij}+\phi_{ji}\big)=0.$$

Définition 11.2. *Soient $B, C \subset S$ non-vides et disjoints. Un **flot unité de B vers C** est un flot ϕ tel que $\sum_{i\in B}\phi[i]=1$ et $\phi[j]=0, \forall j\notin B\cup C$.*

11.1.2 Réseaux électriques

Définition 11.3. *Soit $i_0 \in S$ et $C \subset S\setminus\{i_0\}$. Un **réseau électrique** est un graphe pondéré connexe (S, A, c), muni d'une fonction $V : S \to \mathbb{R}$, le **potentiel**, et d'un flot I, le **courant**, satisfaisant les relations suivantes :*

$$\textbf{\textit{Loi d'Ohm :}}\qquad I_{ij}=c_{ij}\big(V(i)-V(j)\big),\qquad \forall (i,j)\in A, \tag{11.1}$$

$$\textbf{\textit{Loi de Kirchhoff :}}\qquad I[i]:=\sum_{j\sim i}I_{ij}=0,\qquad \forall i\notin\{i_0\}\cup C, \tag{11.2}$$

ainsi que la normalisation suivante[1] *:*

$$V(i_0)=1,\qquad V_{|_C}\equiv 0. \tag{11.3}$$

*Dans ce contexte, les poids c_{ij} sont appellés **conductances**, et leurs inverses $r_{ij}=1/c_{ij}$ **résistances**.*

Le résultat suivant montre que le potentiel V (et donc également le courant I) est complètement caractérisé par la loi d'Ohm, la loi de Kirchhoff, et la normalisation (11.3).

Lemme 11.4. *Soient V et I satisfaisant (11.1), (11.2) et (11.3). Alors, V est l'unique solution du problème de Dirichlet suivant :*

$$\begin{cases} V(i)=\sum_{j\sim i}p_{ij}V(j) & \forall i\notin\{i_0\}\cup C,\\ V(i_0)=1,\\ V_{|_C}\equiv 0, \end{cases} \tag{11.4}$$

*où $p_{ij}=c_{ij}/c[i]$, pour tout $(i,j)\in A$. On dit que V est **harmonique** hors de i_0 et C.*

Démonstration. Montrons tout d'abord que V est bien solution du problème de Dirchlet (11.4). Les conditions au bord sont automatiquement vérifiées puisque V satisfait (11.3). D'autre part, pour tout $i\notin\{i_0\}\cup C$,

$$0=\sum_{j\sim i}I_{ij}=\sum_{j\sim i}\big(V(i)-V(j)\big)c_{ij}=c[i]\Big\{V(i)-\sum_{j\sim i}V(j)p_{ij}\Big\},$$

la première égalité suivant de (11.2) et la seconde de (11.1). Comme $c[i]>0$ pour tout $i\in S$, l'affirmation est démontrée.

Montrons que le problème de Dirichlet possède une unique solution. Soit donc f et g deux solutions. La fonction $h:=f-g$ est alors solution du problème de Dirichlet

$$\begin{cases} h(i)=\sum_{j\sim i}p_{ij}h(j) & \forall i\notin\{i_0\}\cup C,\\ h(i)=0 & \forall i\in\{i_0\}\cup C. \end{cases}$$

1. Pour $f : S \to \mathbb{R}$ et $C \subset S$, on note $f_{|_C}$ la restriction de f à C.

Soit $k \in S \setminus (\{i_0\} \cup C)$ tel que $|h(k)|$ soit maximal. Sans perte de généralité, on peut supposer que $h(k) \geqslant 0$. On a alors

$$\sum_{j \sim k} p_{kj}(h(j) - h(k)) = 0.$$

Or, $p_{kj} > 0$ pour tout $j \sim k$, et $h(j) - h(k) \leqslant 0$ par choix de k. Il suit que $h(j) = h(k)$ pour tout $j \sim k$. En itérant cet argument, on en déduit que h est constante sur S, et la condition au bord impose que $h \equiv 0$ et donc que $f = g$.

L'existence d'une solution peut être établie à partir de l'unicité et de considérations élémentaires d'algèbre linéaire. Alternativement, elle suit du Théorème 11.6 donné plus loin. □

11.2 Liens avec les chaînes de Markov réversibles

Nous allons voir à présent que le potentiel V et le courant I d'un réseau électrique peuvent être réinterprétés de façon probabiliste, en termes de la chaîne de Markov $(X_n)_{n \geqslant 1}$. Une telle identification permet d'utiliser l'intuition que l'on a de tels réseaux afin d'obtenir des résultats sur les chaînes de Markov (et vice-versa).

Pour $B \subset S$, on note

$$T_B := \inf\{n \geqslant 0 \,|\, X_n \in B\}, \qquad T_B^+ := \inf\{n \geqslant 1 \,|\, X_n \in B\}.$$

Lorsque B est réduit à un point, $B = \{i\}$, on écrit simplement $T_i \equiv T_{\{i\}}$ et $T_i^+ \equiv T_{\{i\}}^+$.

Lemme 11.5. *Soient $i_0 \in S$ et $C \subset S \setminus \{i_0\}$. Alors*

$$\phi_{ij} = \mathbb{E}_{i_0}\Big[\sum_{n=1}^{T_C} \big(\mathbf{1}_{\{X_{n-1}=i, X_n=j\}} - \mathbf{1}_{\{X_{n-1}=j, X_n=i\}}\big)\Big]$$

est un flot unité de i_0 vers C, que l'on notera $\phi^{i_0 \to C}$.

Démonstration. Il est clair que $\phi^{i_0 \to C}$ est un flot. De plus, (X_n) étant récurrente positive, $\mathbb{E}_{i_0}(T_C) < \infty$. Par conséquent,

$$\begin{aligned}
\phi[i] &= \mathbb{E}_{i_0}\Big[\sum_{n=1}^{T_C} \sum_{j \sim i} \big(\mathbf{1}_{\{X_{n-1}=i, X_n=j\}} - \mathbf{1}_{\{X_{n-1}=j, X_n=i\}}\big)\Big] \\
&= \mathbb{E}_{i_0}\Big[\sum_{n=1}^{T_C} \big(\mathbf{1}_{\{X_{n-1}=i\}} - \mathbf{1}_{\{X_n=i\}}\big)\Big] \\
&= \mathbb{E}_{i_0}\big[\mathbf{1}_{\{X_0=i\}} - \mathbf{1}_{\{X_{T_C}=i\}}\big] \\
&= \begin{cases} 0, & \text{si } i \notin \{i_0\} \cup C, \\ 1 & \text{si } i = i_0, \\ -\mathbb{P}_{i_0}(X_{T_C} = i) & \text{si } i \in C. \end{cases}
\end{aligned}$$

□

Théorème 11.6. *Soient $i_0 \in S$ et $C \subset S \setminus \{i_0\}$. Le potentiel V et le courant I possèdent les représentations probabilistes suivantes :*

$$V(i) = \mathbb{P}_i(T_{i_0} < T_C), \tag{11.5}$$

$$V(i) = \frac{1}{r_{\mathrm{eff}}^{i_0 \to C} c[i]} \mathbb{E}_{i_0}\Big[\sum_{n=0}^{T_C - 1} \mathbf{1}_{\{X_n = i\}}\Big], \tag{11.6}$$

$$I_{ij} = \frac{1}{r_{\mathrm{eff}}^{i_0 \to C}} \phi_{ij}^{i_0 \to C}, \tag{11.7}$$

où $r_{\mathrm{eff}}^{i_0 \to C} := \big(c[i_0] \mathbb{P}_{i_0}(T_{i_0}^+ > T_C)\big)^{-1}$.

Définition 11.7. *La fonction* $\mathcal{G}_C(i_0, i) := \mathbb{E}_{i_0}\Big(\sum_{n=0}^{T_C - 1} \mathbf{1}_{\{X_n = i\}}\Big)$ *est appelée* ***fonction de Green****. La probabilité* $p_{\mathrm{esc}} := \mathbb{P}_{i_0}(T_{i_0}^+ > T_C)$ *est appelée* ***probabilité de fuite****. La quantité* $r_{\mathrm{eff}}^{i_0 \to C}$ *est appelée* ***résistance effective***

Démonstration. Pour établir (11.5), il suffit de vérifier que $\mathbb{P}_i(T_{i_0} < T_C)$ est (l'unique) solution du problème de Dirichlet (11.4). Manifestement, les conditions au bord sont vérifiées. Quant à l'harmonicité, elle suit immédiatement de la propriété de Markov : pour tout $i \notin \{i_0\} \cup C$,

$$\mathbb{P}_i(T_{i_0} < T_C) = \sum_{j \sim i} \mathbb{P}_i(T_{i_0} < T_C \,|\, X_1 = j)\, p_{ij} = \sum_{j \sim i} \mathbb{P}_j(T_{i_0} < T_C)\, p_{ij}.$$

Pour dériver les représentations (11.6) et (11.7), on observe tout d'abord que

$$\begin{aligned}
\phi_{ij}^{i_0 \to C} &= \mathbb{E}_{i_0}\Big[\sum_{n=1}^{T_C} \mathbf{1}_{\{X_{n-1} = i, X_n = j\}}\Big] - \mathbb{E}_{i_0}\Big[\sum_{n=1}^{T_C} \mathbf{1}_{\{X_{n-1} = j, X_n = i\}}\Big] \\
&= \sum_{n=1}^{\infty} \mathbb{P}_{i_0}\big(X_{n-1} = i, X_n = j, T_C \geqslant n\big) - \sum_{n=1}^{\infty} \mathbb{P}_{i_0}\big(X_{n-1} = j, X_n = i, T_C \geqslant n\big) \\
&= \sum_{n=1}^{\infty} \mathbb{P}_{i_0}\big(X_{n-1} = i, T_C \geqslant n\big)\, p_{ij} - \sum_{n=1}^{\infty} \mathbb{P}_{i_0}\big(X_{n-1} = j, T_C \geqslant n\big)\, p_{ji} \\
&= \mathbb{E}_{i_0}\Big[\sum_{n=1}^{T_C} \mathbf{1}_{\{X_{n-1} = i\}}\Big]\, p_{ij} - \mathbb{E}_{i_0}\Big[\sum_{n=1}^{T_C} \mathbf{1}_{\{X_{n-1} = j\}}\Big]\, p_{ji} \\
&= \big(\frac{1}{c[i]} \mathcal{G}_C(i_0, i) - \frac{1}{c[j]} \mathcal{G}_C(i_0, j)\big)\, c_{ij},
\end{aligned}$$

où la troisième égalité suit de la propriété de Markov (et du fait que $\{T_C \geqslant n\}$ ne dépend que des $n-1$ premiers pas de la chaîne) et la dernière de $c_{ij} = p_{ij} c[i] = p_{ji} c[j]$. On en déduit que le couple $\big(\frac{1}{c[\cdot]} \mathcal{G}_C(i_0, \cdot), \phi^{i_0 \to C}\big)$ vérifie la loi d'Ohm (11.1). D'autre part, il suit du Lemme 11.5 que $\phi^{i_0 \to C}$ satisfait la loi de Kirchhoff (11.2). De plus, comme, pour tout $i \in C$,

$$\mathcal{G}_C(i_0, i) = \mathbb{E}_{i_0}\Big[\sum_{n=0}^{T_C - 1} \mathbf{1}_{\{X_n = i\}}\Big] = 0,$$

on déduit que

$$\frac{c[i_0]}{\mathcal{G}_C(i_0, i_0)} \frac{\mathcal{G}_C(i_0, i)}{c[i]} \equiv \frac{1}{r} \frac{\mathcal{G}_C(i_0, i)}{c[i]}$$

satisfait la normalisation (11.3). Par conséquent, il suit du lemme 11.4 que

$$V(i) = \frac{1}{r}\,\frac{\mathcal{G}_C(i_0, i)}{c[i]}, \qquad \text{et} \qquad I_{ij} = \frac{1}{r}\,\phi_{ij}^{i_0 \to C}.$$

Pour conclure, il suffit donc de montrer que $r = r_{\text{eff}}^{i_0 \to C}$. On observe tout d'abord qu'il suit du Lemme 11.5 que $\phi^{i_0 \to C}[i_0] = 1$, et par conséquent, $I[i_0] = 1/r$. On peut donc écrire

$$\begin{aligned}
\frac{1}{r} = I[i_0] &= \sum_{j \sim i_0} \big(V(i_0) - V(j)\big) c_{i_0 j} && \text{par (11.1)} \\
&= c[i_0] \Big\{ 1 - \sum_{j \sim i_0} p_{i_0 j} V(j) \Big\} && \text{puisque } V(i_0) = 1 \\
&= c[i_0] \Big\{ 1 - \sum_{j \sim i_0} p_{i_0 j} \mathbb{P}_j(T_{i_0} < T_C) \Big\} && \text{par (11.5)} \\
&= c[i_0] \Big\{ 1 - \mathbb{P}_{i_0}(T_{i_0}^+ < T_C) \Big\} \\
&= c[i_0]\, \mathbb{P}_{i_0}(T_{i_0}^+ > T_C).
\end{aligned}$$

□

La terminologie « résistance effective » vient du raisonnement suivant : on a imposé une différence de potentiel égale à 1 entre i_0 et C. Or, on a vu que

$$r_{\text{eff}}^{i_0 \to C} I[i_0] = 1 = V(i_0) - V(C).$$

On peut alors interpréter la relation ci-dessus comme une version effective de la loi d'Ohm, le reste du réseau électrique jouant le rôle d'une grosse résistance placée entre les « points » i_0 et C. La valeur de cette « résistance effective » est alors bien donnée par $r_{\text{eff}}^{i_0 \to C}$.

Définition 11.8. *On appelle* ***conductance effective*** *entre i_0 et C la grandeur $c_{\text{eff}}^{i_0 \to C} := 1/r_{\text{eff}}^{i_0 \to C}$.*

Observez que la probabilité de fuite p_{esc} prend une forme particulièrement simple en termes de la conductance effective :

$$p_{\text{esc}} = \frac{c_{\text{eff}}^{i_0 \to C}}{c[i_0]}. \tag{11.8}$$

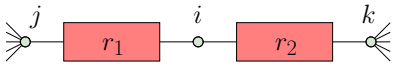
Bien sûr, pour que les représentations probabilistes de cette section soient réellement utiles, on a besoin de techniques permettant de déterminer, ou au moins d'estimer, la résistance effective d'un réseau. La section suivante est consacrée à ce problème.

11.3 Calcul de la résistance effective

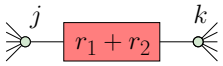
Afin de déterminer la résistance effective d'un réseau électrique, on procède par réduction du réseau à l'aide des 4 règles de transformation suivantes.

Remarque 11.9. *Dans cette section, les transformations appliquées vont en général nous faire momentanément sortir du cadre précédent, le graphe sous-jacent pouvant posséder plusieurs arêtes entre deux sommets donnés. Toutefois, à la fin de la procédure le graphe obtenu sera toujours du type considéré plus haut.* ◇

1. Résistances en série. Supposons que le graphe possède un sommet $i \notin \{i_0\} \cup C$ de degré 2, avec voisins j et k, comme dans le dessin suivant :



Remplaçons les arêtes (i,j) et (i,k), de résistance r_1 et r_2 respectivement, par une unique arête (j,k) de résistance r_1+r_2, comme suit :



Alors, les valeurs prises par le potentiel V sur $S\setminus\{i\}$ et le courant I sur $A\setminus\big\{(i,j),(i,k)\big\}$ sont inchangées. Le courant I_{jk} traversant la nouvelle arête (j,k) de j vers k est égal à I_{ji}.

Pour vérifier cela, il suffit de s'assurer que (11.1) et (11.2) sont encore vérifiées pour le nouveau graphe, avec les mêmes valeurs du potentiel et du courant, ce qui est élémentaire (et laissé en exercice).

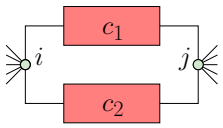
Exemple 11.10. Ceci conduit à une solution particulièrement simple du problème de la ruine du joueur. Soient $p\in(0,1)$ et $q=1-p$; on considère une marche aléatoire simple sur $\{0,\dots,n\}$, absorbée en 0 et n, avec probabilités de transition $p_{i,i+1}=p$, $p_{i,i-1}=q$, pour $1\leqslant i\leqslant n-1$. Que vaut la probabilité $\mathbb{P}_k(T_0<T_n)$ que la marche partant de $k\in\{1,\dots,n-1\}$ atteigne 0 avant d'atteindre n ?

Considérons le réseau électrique dont les conductances sont données par $c_{i,i+1}=(p/q)^i$, pour $0\leqslant i\leqslant n-1$. Si l'on applique un potentiel 1 en 0 et 0 en n, alors la probabilité recherchée, $\mathbb{P}_k(T_0<T_n)$, est donnée par la valeur du potentiel en k, par (11.5). Le résultat ci-dessus nous permet de remplacer les k résistances en série séparant k et 0 par une unique résistance égale à $\sum_{i=0}^{k-1}(q/p)^i$. De même, on peut remplacer les $n-k$ résistances en série placées entre k et n par une résistance unique $\sum_{i=k}^{n-1}(q/p)^i$.

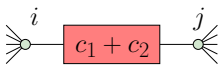
La probabilité recherchée devient alors simplement la probabilité de faire un pas de k à 0, plutôt que de k à n. Or la probabilité de cet événement est donnée par

$$\frac{\sum_{i=k}^{n-1}(q/p)^i}{\sum_{i=0}^{n-1}(q/p)^i}=\begin{cases}(n-k)/n & \text{si } p=q=1/2,\\ \big((p/q)^{n-k}-1\big)\;/\;\big((p/q)^n-1\big) & \text{sinon.}\end{cases}$$ ◇

2. Résistances en parallèle. Supposons que le graphe possède deux arêtes e_1 et e_2, de conductances c_1 et c_2, joignant deux sommets i,j, comme dans le dessin suivant :



Remplaçons les arêtes e_1 et e_2 par une unique arête (i,j) de conductance c_1+c_2, comme suit :



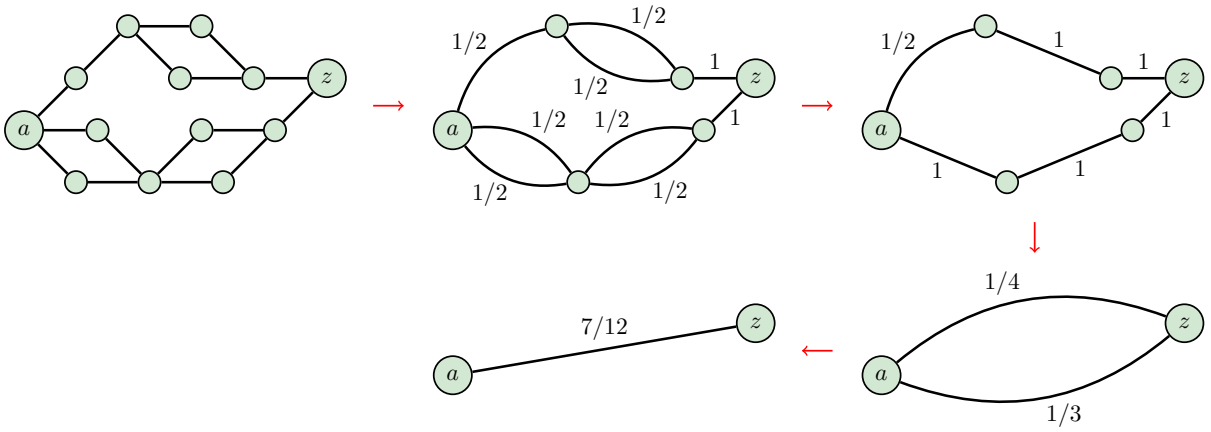
Alors, les valeurs prises par le potentiel V sur S et le courant I sur $A\setminus\{e_1,e_2\}$ sont inchangées. Le courant I_{ij} traversant la nouvelle arête (i,j) de i vers j est égal à la somme des courants traversant e_1 et e_2 de i vers j. La vérification, élémentaire, est à nouveau laissée en exercice.

Exemple 11.11. Nous allons appliquer les deux règles précédentes afin de déterminer la probabilité de fuite de la marche aléatoire simple sur un graphe. On considère la marche aléatoire simple sur le graphe représenté ci-dessous (en haut, à gauche), c'est-à-dire la marche aléatoire se déplaçant à chaque pas du sommet occupé à cet instant vers l'un de ses voisins choisi uniformément. Pour une marche partant de a, on désire déterminer la probabilité qu'elle visite z avant de retourner en a. Cette dernière probabilité est donnée par (*cf.* (11.8))

$$\mathbb{P}_a(T_a^+>T_z)=\frac{c_{\mathrm{eff}}^{a\to z}}{c[a]},$$

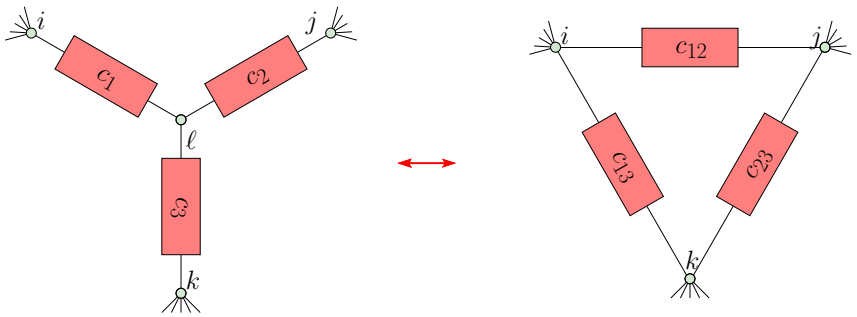
où chaque arête du réseau correspondant a une conductance égale à 1. Comme $c_{\mathrm{eff}}^{a\to z}=I[a]/(V(a)-V(z))$, la conductance effective entre a et z demeure inchangée lors de l'application des deux transformations précédentes. On peut donc les utiliser successivement afin de simplifier le réseau (les nombres

indiqués représentent les conductances associées aux différentes arêtes, toutes étant égales à 1 dans le réseau initial) :



On obtient donc que la conductance effective vaut $7/12$, et donc $\mathbb{P}_a(T_a^+ > T_z) = \frac{7/12}{3} = \frac{7}{36}$. ◇

3. Transformation étoile-triangle. Soit $\ell \in S \setminus (\{i_0\} \cup C)$. Les valeurs prises par le potentiel V sur $S \setminus \{\ell\}$ et le courant I sur les arêtes non modifiées sont inchangées sous les transformations suivantes :

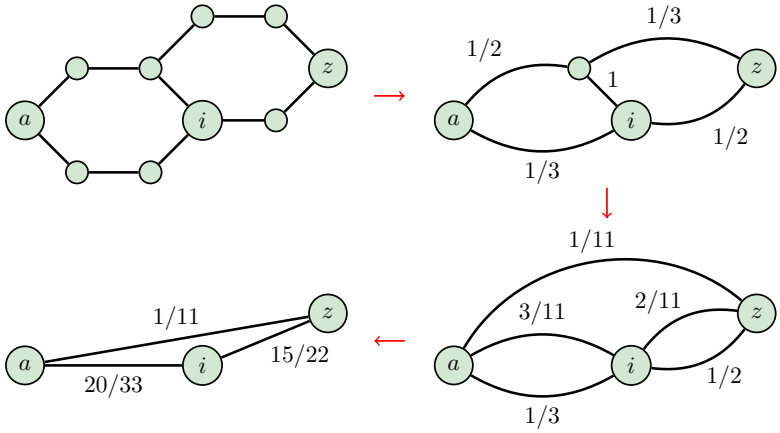


pour peu que les conductances satisfassent

$$c_{12} = \frac{c_1 c_2}{c_1 + c_2 + c_3}, \qquad c_{13} = \frac{c_1 c_3}{c_1 + c_2 + c_3}, \qquad c_{23} = \frac{c_2 c_3}{c_1 + c_2 + c_3}.$$

La vérification est à nouveau laissée en exercice.

Exemple 11.12. On veut déterminer la probabilité que la marche aléatoire simple sur le graphe représenté ci-dessous (en haut, à gauche) et partant de i visite a avant de visiter z. Celle-ci est égale au potentiel $V(i)$ obtenu en imposant $V(a) = 1$ et $V(z) = 0$. On peut donc la déterminer en réduisant le réseau par applications successives des transformations ci-dessus (comme précédemment, les nombres représentent les conductances ; les conductances initiales sont toutes égales à 1) :



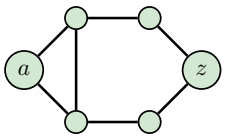
La probabilité recherchée est donc

$$\mathbb{P}_i(T_a < T_z) = \frac{20/33}{20/33 + 15/22} = \frac{8}{17}. \qquad \diamond$$

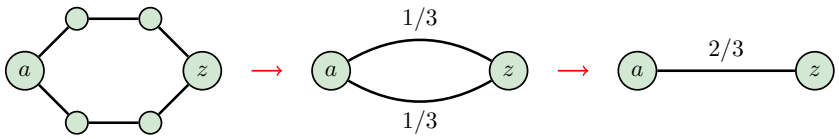
Les trois transformations ci-dessus permettent de réduire n'importe quel réseau planaire (théorème d'Epifanov). Une autre transformation est cependant souvent utile en pratique.

4. Ajout/Retrait d'arêtes. Si deux sommets i et j ont même potentiel, $V(i) = V(j)$, alors on peut modifier c_{ij} sans affecter ni les potentiels, ni les courants (puisque $V(i) = V(j) \implies I_{ij} = 0$). Ceci permet en particulier, l'ajout ou le retrait d'arête entre deux tels sommets (c_{ij} passant respectivement de 0 à une valeur strictement positive, ou d'une valeur strictement positive à 0). Ceci permet également l'identification de deux tels sommets, c'est-à-dire le remplacement de i et j par un sommet unique, avec les mêmes arêtes incidentes, mais sans la boucle correspondant à l'arête (i, j) (moralement, ceci revient à mettre $c_{ij} = \infty$; il est toutefois plus simple de vérifier directement que l'identification de i et j ne modifie pas le potentiel).

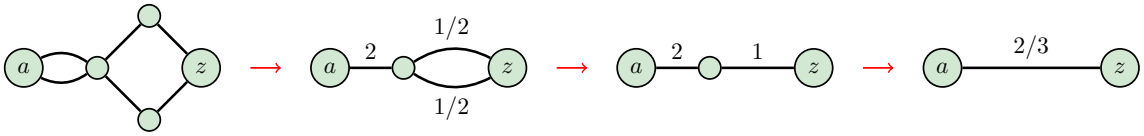
Exemple 11.13. On considère le graphe suivant :



Quelle est la probabilité que la marche aléatoire simple sur ce graphe, partant de a, visite z avant de retourner en a ? Le réseau équivalent a, à nouveau, toutes ses conductances égales à 1. Par symétrie, le potentiel aux deux extrémités de l'arête e sont nécessairement égales. Il suit que l'on peut modifier la conductance associée sans modifier la conductance effective entre a et z. Une première possibilité est de mettre $c_e = 0$, ce qui revient à éliminer l'arête correspondante. On obtient alors :



Une approche alternative revient à identifier les extrémités de l'arête e. On obtient dans ce cas :

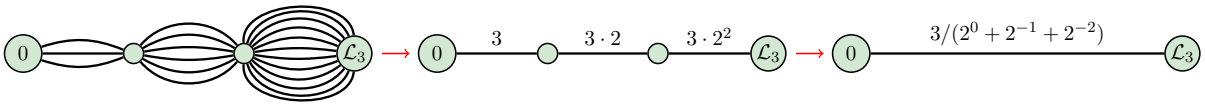


On en déduit dans chaque cas que $\mathbb{P}_a(T_z < T_a^+) = \frac{2/3}{2} = \frac{1}{3}$. ◇

Exemple 11.14. Un exemple plus intéressant est le suivant : on considère la marche aléatoire sur l'arbre binaire, et on désire déterminer la probabilité de ne jamais revisiter le sommet de départ 0. Nous n'avons, dans ce chapitre, considéré que des graphes finis. Ceci ne pose pas de problème dans le cas présent. En effet, si l'on dénote par $\mathcal{L}_n$ l'ensemble des sommets à distance n de 0 (la distance étant donnée par le nombre d'arêtes séparant les deux sommets), la probabilité que 0 ne soit plus jamais revisité est donnée par

$$\mathbb{P}_0(T_0^+ = +\infty) = \lim_{n\to\infty} \mathbb{P}_0(T_0^+ > T_{\mathcal{L}_n}),$$

et l'on peut donc se restreindre à l'arbre constitué de n générations. Par symétrie, la valeur prise par le potentiel en un sommet donné ne dépend que de la distance entre ce sommet et le sommet de départ. On peut donc identifier tous les sommets de $\mathcal{L}_m$ (et ce pour tout $1 \leqslant m \leqslant n$). On obtient ainsi le réseau équivalent, le nombre d'arêtes entre $\mathcal{L}_m$ et $\mathcal{L}_{m+1}$ étant égal à $3 \cdot 2^m$ (ici pour $n = 3$) :



On obtient donc

$$\mathbb{P}_0(T_0^+ = \infty) = \lim_{n\to\infty} \frac{c_{\text{eff}}^{0\to\mathcal{L}_n}}{c[0]} = \lim_{n\to\infty} \frac{3}{3\sum_{m=0}^{n-1} 2^{-m}} = \frac{1}{2}.$$

On conclut en particulier que la marche aléatoire simple sur l'arbre binaire est transiente. ◇

L'exemple précédent suggère l'existence d'un lien entre la résistance effective et les propriétés de récurrence et transience. Nous y reviendrons dans la Section 11.5. Avant cela, nous allons introduire une autre quantité naturelle du point de vue des réseaux électriques dont l'importance est considérable dans l'étude des chaînes de Markov réversibles.

11.4 Énergie et principe variationnel

Définition 11.15. *La **forme de Dirichlet** est la forme quadratique associant à une fonction g sur S le nombre*

$$\mathcal{E}(g,g) := \frac{1}{2}\sum_{i,j\in S} c_{ij}\big(g(j)-g(i)\big)^2 = \sum_{(i,j)\in A} c_{ij}\big(g(j)-g(i)\big)^2.$$

*L'**énergie** associée à un flot ϕ est définie par*

$$\mathcal{E}(\phi) := \sum_{e\in A} r_e \phi_e^2.$$

En particulier, si g est le potentiel imposé aux nœuds du réseau électrique ou si ϕ est le courant traversant le réseau, la forme de Dirichlet et l'énergie correspondent toutes deux à l'énergie dissipée par le réseau.

À l'aide de ces notions, il est possible d'exprimer la résistance effective $r_{\text{eff}}^{i_0\to C}$ comme solution de deux problèmes variationnels. Ce qui est remarquable, c'est que l'un caractérise la résistance effective comme un supremum, et l'autre comme un infimum ; ceci permet donc d'obtenir des bornes supérieures et inférieures sur la résistance effective en évaluant $\mathcal{E}(g,g)$ ou $\mathcal{E}(\phi)$ en une fonction g ou un flot ϕ bien choisis.

Théorème 11.16. *Soit $i_0 \in S$ et $C \subset S \setminus \{i_0\}$. Alors,*

$$c_{\text{eff}}^{i_0 \to C} = \min\{\mathcal{E}(g,g) \,|\, g(i_0) = 1, g|_C \equiv 0\}, \tag{11.9}$$

$$r_{\text{eff}}^{i_0 \to C} = \min\{\mathcal{E}(\phi) \,|\, \phi \text{ flot unité de } i_0 \text{ vers } C\}. \tag{11.10}$$

De plus, ces minima sont uniquement atteints, respectivement, lorsque $g = V$, le potentiel engendré par l'application d'un potentiel 1 en i_0 et 0 sur C, et lorsque $\phi = \phi^{i_0 \to C} = r_{\text{eff}}^{i_0 \to C} I$ avec I le courant correspondant.

Démonstration. On démontre d'abord (11.9). Soit $g_{\min}$ un extremum de $\mathcal{E}(g,g)$ parmi les fonctions sur S satisfaisant $g(i_0) = 1$ et $g|_C \equiv 0$. Alors, on doit avoir

$$0 = \frac{\partial}{\partial g(i)} \mathcal{E}(g,g)|_{g=g_{\min}} = 2c[i] \sum_{j \sim i} p_{ij} \big(g_{\min}(i) - g_{\min}(j)\big),$$

pour tout $i \notin C \cup \{i_0\}$. En particulier, $g_{\min}$ est harmonique hors de $C \cup \{i_0\}$. Il suit donc du Lemme 11.4 et du Théorème 11.6 que $g_{\min}$ satisfait

$$g_{\min}(i) = V(i) = \mathbb{P}_i(T_{i_0} < T_C).$$

Manifestement, $g_{\min}$ est un minimum puisque $\mathcal{E}(g,g)$ est une forme quadratique semi-définie positive. On observe à présent que

$$\begin{aligned}
\mathcal{E}(g_{\min}, g_{\min}) &= \frac{1}{2} \sum_{i,j \in S} c_{ij} \big(g_{\min}(j) - g_{\min}(i)\big)^2 \\
&= \sum_{i \in S} c[i] g_{\min}(i)^2 - \sum_{i \in S} c[i] g_{\min}(i) \sum_{j \sim i} p_{ij} g_{\min}(j) \\
&= \sum_{i \in S} c[i] g_{\min}(i) \Big(g_{\min}(i) - \sum_{j \sim i} p_{ij} g_{\min}(j)\Big) \\
&= c[i_0] \Big(1 - \sum_{j \sim i_0} p_{i_0 j} g_{\min}(j)\Big),
\end{aligned}$$

puisque $g_{\min}$ est harmonique sur $S \setminus \{i_0\} \cup C$ et satisfait $g_{\min}(i_0) = 1$ et $g|_C \equiv 0$. On a donc

$$\begin{aligned}
\mathcal{E}(g_{\min}, g_{\min}) &= c[i_0] \Big(1 - \sum_{j \sim i_0} p_{i_0 j} \mathbb{P}_j(T_{i_0} < T_C)\Big) \\
&= c[i_0] \big(1 - \mathbb{P}_{i_0}(T_{i_0}^+ < T_C)\big) \\
&= c[i_0] \mathbb{P}_{i_0}(T_{i_0}^+ > T_C) \\
&= c_{\text{eff}}^{i_0 \to C}.
\end{aligned}$$

Passons à présent à la preuve de (11.10). Soit V le potentiel résultant de l'application d'un potentiel 1 en i_0 et 0 sur C et I le courant correspondant. Dans ce cas, on a vu que $\phi^{i_0 \to C} = r_{\text{eff}}^{i_0 \to C} I$. De plus, tout flot unité ϕ de i_0 vers C peut s'écrire $\phi = \phi^{i_0 \to C} + \psi$, avec ψ un flot tel que $\psi[i] = 0$ pour tout $i \in S \setminus C$. On a alors

$$\mathcal{E}(\phi) = \mathcal{E}(\phi^{i_0 \to C} + \psi) = \mathcal{E}(\phi^{i_0 \to C}) + \mathcal{E}(\psi) + 2 \sum_{e \in A} r_e \phi_e^{i_0 \to C} \psi_e.$$

On observe à présent que $\mathcal{E}(\psi) > 0$ si $\psi \not\equiv 0$, et que

$$\mathcal{E}(\phi^{i_0 \to C}) = (r_{\text{eff}}^{i_0 \to C})^2 \sum_e r_e I_e^2 = (r_{\text{eff}}^{i_0 \to C})^2 \sum_{(i,j) \in A} c_{ij} \big(V(j) - V(i)\big)^2 = (r_{\text{eff}}^{i_0 \to C})^2 c_{\text{eff}}^{i_0 \to C} = r_{\text{eff}}^{i_0 \to C},$$

la troisième égalité résultant de la première partie de la preuve. Finalement, comme

$$\sum_{\substack{i,j\in S\\ i\sim j}} V(j)\psi_{ij} = -\sum_{\substack{i,j\in S\\ i\sim j}} V(j)\psi_{ji} = -\sum_{\substack{i,j\in S\\ i\sim j}} V(i)\psi_{ij},$$

on obtient

$$\begin{aligned} 2\sum_{e\in A} r_e\phi_e^{i_0\to C}\psi_e = r_{\text{eff}}^{i_0\to C}\sum_{\substack{i,j\in S\\ i\sim j}} \big(V(i)-V(j)\big)\psi_{ij} &= 2r_{\text{eff}}^{i_0\to C}\sum_{\substack{i,j\in S\\ i\sim j}} V(i)\psi_{ij} \\ &= 2r_{\text{eff}}^{i_0\to C}\sum_{i\in S} V(i)\sum_{j\sim i}\psi_{ij} = 2r_{\text{eff}}^{i_0\to C}\sum_{i\in S} V(i)\psi[i] = 0, \end{aligned}$$

puisque $\psi[i] = 0$ si $i \notin C$ et $V(i) = 0$ si $i \in C$. Par conséquent, on a bien

$$\mathcal{E}(\phi) > \mathcal{E}(\phi^{i_0\to C}) = r_{\text{eff}}^{i_0\to C},$$

pour tous les flots unité $\phi \neq \phi^{i_0\to C}$. □

Le théorème précédent possède un corollaire particulièrement utile, lorsque le calcul explicite de la résistance effective est impossible.

Corollaire 11.17 (Principe de monotonicité de Rayleigh)**.** *La résistance effective $r_{\text{eff}}^{i_0\to C}$ est une fonction croissante des résistances du réseau. En particulier,*

- ▷ *identifier des sommets ne peut que réduire $r_{\text{eff}}^{i_0\to C}$;*
- ▷ *supprimer des arêtes ne peut qu'augmenter $r_{\text{eff}}^{i_0\to C}$.*

Démonstration. L'affirmation principale suit directement de la monotonicité de la fonction de Dirichlet par rapport aux r_{ij}. Les deux affirmations secondaires correspondent à des cas particuliers : mettre une résistance à 0 revient à identifier les extrémités de l'arêtes correspondante, et la mettre à $+\infty$ revient à enlever cette arête. □

11.5 Récurrence et transience

On considère dans cette section un graphe pondéré connexe (S, A, c) avec S dénombrable. Pour simplifier, on supposera que $\{j \in S \mid j \sim i\}$ est fini pour tout $i \in S$. On note $(X_n)_{n\geqslant 0}$ la chaîne de Markov irréductible sur S avec probabilités de transition $p_{ij} = c_{ij}/c[i]$ si $(i, j) \in A$ et $p_{ij} = 0$ sinon. On fixe un sommet arbitraire $i_0 \in S$ et, pour chaque $n \in \mathbb{N}$, on introduit $\mathcal{L}_n := \{i \in S \mid d(i_0, i) = n\}$, où la distance $d(i, j)$ entre $i, j \in S$ est la longueur du plus court chemin composé d'arêtes de A joignant i et j. On a alors la caractérisation suivante de la récurrence de la chaîne.

Théorème 11.18. *La chaîne de Markov $(X_n)_{n\geqslant 0}$ est récurrente si et seulement si $\lim_{n\to\infty} r_{\text{eff}}^{i_0\to\mathcal{L}_n} = \infty$.*

Démonstration. Par des résultats standards sur les chaînes de Markov, la chaîne est récurrente si et seulement si $\mathbb{P}_{i_0}(T_{i_0}^+ = \infty) = 0$. On vérifie aisément que $\mathbb{P}_{i_0}(\exists n \geqslant 1,\, T_{\mathcal{L}_n}^+ = \infty) = 0$. Par conséquent,

$$\mathbb{P}_{i_0}(T_{i_0}^+ = \infty) = \lim_{n\to\infty} \mathbb{P}_{i_0}(T_{i_0}^+ > T_{\mathcal{L}_n}) = \frac{1}{c[i_0]}\lim_{n\to\infty} c_{\text{eff}}^{i_0\to\mathcal{L}_n}.$$ □

Exemple 11.19. Nous allons montrer que les marches aléatoires simples sur $\mathbb{Z}$ et sur $\mathbb{Z}^2$ sont récurrentes. Pour cela, il suffit de montrer que $\lim_{n\to\infty} r_{\text{eff}}^{0\to\mathcal{L}_n} = \infty$. On utilise le principe de monotonicité de Rayleigh (Corollaire 11.17) afin d'obtenir une borne inférieure convenable.

Tout d'abord, on observe qu'il suffit de démontrer le résultat pour $\mathbb{Z}^2$, puisque $\mathbb{Z}$ peut être obtenu à partir de ce dernier graphe en retirant des arêtes (c'est-à-dire, en mettant les résistances correspondantes à $+\infty$).

Pour $\mathbb{Z}^2$, on procède comme suit : pour chaque $1 \leqslant m \leqslant n$, on identifie tous les sommets de $\mathcal{L}_m$. Cette opération diminue la résistance effective du réseau entre 0 et $\mathcal{L}_n$. On obtient ainsi une chaîne de $n+1$ sommets, le sommet m étant relié au sommet $m+1$ par $4(2m+1)$ arêtes de conductance 1 ou, ce qui est équivalent, une arête de conductance $4(2m+1)$. On obtient donc

$$\lim_{n\to\infty} r_{\text{eff}}^{0\to\mathcal{L}_n} \geqslant \frac{1}{4} \lim_{n\to\infty} \sum_{m=0}^{n-1} \frac{1}{2m+1} = \infty. \qquad \diamond$$

Exemple 11.20. Nous allons à présent montrer la transience de la marche simple sur $\mathbb{Z}^d$, $d \geqslant 3$. À nouveau, il suffit de considérer le cas de $\mathbb{Z}^3$.

Pour ce faire, nous allons construire un flot unité ϕ_n de 0 vers $\mathcal{L}_n$ dont l'énergie $\mathcal{E}(\phi_n)$ est bornée uniformément en n. La conclusion suivra alors des Théorèmes 11.16 et 11.18.

Commençons par quelques considérations générales. N'importe quel chemin $e_1, \dots, e_m$, composé d'arêtes (orientées) distinctes de $\mathbb{Z}^3$, reliant 0 à $\mathcal{L}_n$ fournit un flot unité ϕ : il suffit de poser

$$\phi_e = \sum_{k=1}^{m} \big(\mathbf{1}_{\{e=e_k\}} - \mathbf{1}_{\{e=-e_k\}}\big).$$

Supposons à présent que $\mathbb{Q}$ soit une mesure de probabilité sur un ensemble de tels chemins. Alors, on peut à nouveau aisément définir un flot unité ϕ, en posant

$$\phi_e = \sum_{k=1}^{\infty} \big(\mathbb{Q}(e = e_k) - \mathbb{Q}(e = -e_k)\big).$$

(Observez qu'on a juste pris l'espérance de l'expression précédente.)

Dans le cas de $\mathbb{Z}^3$, une classe naturelle de tels chemins, et la mesure $\mathbb{Q}$ associée, est obtenue comme suit : on tire au hasard, selon la loi uniforme, un vecteur unité dans $\mathbb{R}^3$. On considère la demi-droite D partant de 0 et de direction donnée par ce vecteur. On associe à celle-ci le chemin $\mathcal{P}_D$ connectant 0 à $\mathcal{L}_n$, ne visitant qu'une fois chaque arête et tel que $\max_{i\in\mathcal{P}_D} d(i, D)$ soit minimal (il y a presque sûrement un unique tel chemin). Notons ϕ^n le flot obtenu à partir de la distribution des chemins $\mathcal{P}_D$.

Nous allons montrer que le flot ϕ^n ainsi défini possède une énergie uniformément bornée en n, ce qui impliquera la transience. Étant donné une arête $e = (i, j)$, on notera $d(0, e) \coloneqq \max\{d(0, i), d(0, j)\}$. On observe tout d'abord qu'il existe une constante A telle que si e est une arête telle que $d(0, e) = m$, alors $\mathbb{Q}(e \in \mathcal{P}_D) \leqslant A/m^2$. D'autre part, il existe une constante B telle que le nombre d'arêtes e satisfaisant $d(0, e) = m$ soit borné par Bm^2. Par conséquent, puisqu'on peut prendre $r_e = 1$ pour tout e,

$$\mathcal{E}(\phi^n) = \sum_e r_e(\phi_e^n)^2 \leqslant \sum_{m=1}^{n} \sum_{e:d(0,e)=m} \mathbb{Q}(e \in \mathcal{P}_D)^2 \leqslant \sum_{m=1}^{n} \frac{A^2}{m^4} Bm^2 \leqslant A^2 B \sum_{m\geqslant 1} m^{-2} < \infty,$$

uniformément en n. $\diamond$

Remarque 11.21. *Un des avantages de cette approche pour démontrer la récurrence ou la transience d'une chaîne de Markov est qu'elle ne repose que sur des propriétés grossières des graphes sous-jacents, et n'exige pas d'avoir des informations détaillées sur la structure du graphe (par exemple, la régularité de $\mathbb{Z}^d$).* $\diamond$

Remarques bibliographiques : Une référence, très facile d'accès, pour ce chapitre est [17]. D'autres références, plus avancées, sont [40, 3, 47].

12 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 : $(X_n)_{n\geqslant 0}$ 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 $p_{ij}(n) \coloneqq (P^n)_{ij} = \mathbb{P}(X_n = j \mid X_0 = i)$.

12.1 Structure algébrique

12.1.1 Une caractérisation de la réversibilité

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

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

La norme associée sera notée $\|f\|_\pi \coloneqq \sqrt{\langle f,\, f\rangle_\pi}$. Finalement, nous écrirons

$$\langle f\rangle_\pi \coloneqq \sum_{i\in S} f(i)\pi(i) = \langle f,\, \mathbf{1}\rangle_\pi \tag{12.1}$$

pour l'espérance de f sous π (nous avons noté $\mathbf{1}$ la fonction $\mathbf{1}(i) \coloneqq 1$ 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 $(Pf)(i) \coloneqq \sum_{j\in S} p_{ij}f(j)$.

Le théorème suivant fournit une caractérisation des chaînes de Markov réversibles.

Théorème 12.1. *Une chaîne de Markov $(X_n)_{n\geqslant 0}$ 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 $\ell^2(\pi)$:*

$$\forall f,\, g \in \ell^2(\pi), \qquad \langle Pf,\, g\rangle_\pi = \langle f,\, Pg\rangle_\pi.$$

Démonstration. ▷ Supposons tout d'abord que, $\pi(i)p_{ij} = \pi(j)p_{ji}$ pour tout $i, j \in S$. Alors, pour tout $f,\, g \in \ell^2(\pi)$,

$$\begin{aligned}\langle Pf,\, g\rangle_\pi &= \sum_{i\in S}\Big[\sum_{j\in S} p_{ij}f(j)\Big]g(i)\pi(i)\\ &= \sum_{i\in S}\sum_{j\in S} p_{ji}f(j)g(i)\pi(j)\\ &= \sum_{j\in S} f(j)\Big[\sum_{i\in S} p_{ji}g(i)\Big]\pi(j) = \langle f,\, Pg\rangle_\pi.\end{aligned}$$

▷ Supposons à présent que P soit autoadjoint dans $\ell^2(\pi)$. Alors, pour tout $i, j \in S$,

$$\pi(j)p_{ji} = \sum_{r\in S}\Big[\sum_{s\in S} p_{rs}\delta_i(s)\Big]\delta_j(r)\pi(r) = \langle P\delta_i,\, \delta_j\rangle_\pi = \langle \delta_i,\, P\delta_j\rangle_\pi = \pi(i)p_{ij},$$

où l'on a noté δ_u, $u \in S$, la fonction définie par $\delta_u(v) \coloneqq \delta_{u,v}$ pour tout $v \in S$. □

12.1.2 Valeurs propres et vecteurs propres de P

P étant autoadjointe, elle possède $|S|$ valeurs propres réelles, que l'on ordonne en ordre décroissant :

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_{|S|}.$$

On vérifie facilement que $\lambda_1 = 1$ et $\lambda_{|S|} \geqslant -1$. 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 $|\lambda||v(i)| = |\lambda v(i)| = |\sum_{j\in S} p_{ij}v(j)| \leqslant |v(i)|$. Deux quantités vont jouer un rôle crucial dans la suite.

Définition 12.2. *Le **trou spectral** est défini par $\gamma \coloneqq 1 - \lambda_2$. Le **trou spectral absolu** est défini par $\gamma_* \coloneqq \min\{\gamma, 1 - |\lambda_{|S|}|\}$.*

Théorème 12.3. *Soit P une matrice stochastique irréductible, apériodique et réversible. Alors, $\gamma \geqslant \gamma_* > 0$.*

Démonstration. L'inégalité $\gamma \geqslant \gamma_*$ 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

$$|\lambda^{2n}w(i)| = |(P^{2n}w)(i)| = \Big|\sum_{k\in S} p_{ik}(2n)w(k)\Big| \leqslant \sum_{k\in S} p_{ik}(2n)|w(k)| < |w(i)|\sum_{k\in S} p_{ik}(2n) = |w(i)|,$$

d'où l'on conclut à nouveau que $|\lambda^{2n}| < 1$, et donc que $|\lambda| < 1$. □

P étant autoadjointe, on sait du cours d'algèbre linéaire que l'on peut trouver une base orthonormale (par rapport au produit scalaire $\langle \cdot\,,\,\cdot \rangle_\pi$) composée de vecteurs propres $(v_k)_{k=1,\dots,|S|}$, avec $v_1 \coloneqq \mathbf{1}$ 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 (12.1), $\langle f,\, v_1 \rangle_\pi = \langle f \rangle_\pi$. En particulier,

$$f - \langle f \rangle_\pi \mathbf{1} = \sum_{k=2}^{|S|} \langle f,\, v_k \rangle_\pi \, v_k, \tag{12.2}$$

et donc

$$\begin{aligned}
\mathrm{Var}_\pi[f] = \sum_{i=1}^{|S|} \big(f(i) - \langle f \rangle_\pi\big)^2 \pi(i) &= \sum_{i=1}^{|S|} \Big(\sum_{k=2}^{|S|} \langle f,\, v_k \rangle_\pi \, v_k(i) \Big)^2 \pi(i) \\
&= \sum_{i=1}^{|S|} \sum_{k=2}^{|S|} \sum_{\ell=2}^{|S|} \langle f,\, v_k \rangle_\pi \langle f,\, v_\ell \rangle_\pi \, v_k(i) v_\ell(i) \pi(i) \\
&= \sum_{k=2}^{|S|} \sum_{\ell=2}^{|S|} \langle f,\, v_k \rangle_\pi \langle f,\, v_\ell \rangle_\pi \, \langle v_k,\, v_\ell \rangle_\pi \\
&= \sum_{k=2}^{|S|} \langle f,\, v_k \rangle_\pi^2.
\end{aligned} \tag{12.3}$$

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 $n \in \mathbb{N}^*$,

$$P^n f = \sum_{k=1}^{|S|} \langle f,\, v_k \rangle_\pi \, \lambda_k^n v_k. \tag{12.4}$$

En particulier, pour tout $i, j \in S$,

$$p_{ij}(n) = (P^n \delta_j)(i) = \sum_{k=1}^{|S|} \langle \delta_j,\, v_k \rangle_\pi \, \lambda_k^n v_k(i) = \sum_{k=1}^{|S|} \pi(j) v_k(j) \lambda_k^n v_k(i). \tag{12.5}$$

12.2 Vitesse de convergence

Le résultat suivant fournit une borne spectrale sur la vitesse de convergence vers l'équilibre.

Théorème 12.4. *Soit P irréductible et réversible sur S. Alors,*

$$\|\mathbb{P}(X_n = \cdot \,|\, X_0 = i) - \pi\|_{\mathrm{VT}} \leqslant \frac{1}{2} \sqrt{\frac{1 - \pi(i)}{\pi(i)}} (1 - \gamma_*)^n.$$

Démonstration. On a

$$\begin{aligned} 4\|\mathbb{P}(X_n = \cdot \,|\, X_0 = i) - \pi\|_{\mathrm{VT}}^2 &= \Big(\sum_{j\in S} |p_{ij}(n) - \pi(j)|\Big)^2 = \Big(\sum_{j\in S} \sqrt{\tfrac{\pi(j)}{\pi(j)}}\,|p_{ij}(n) - \pi(j)|\Big)^2 \\ &\leqslant \Big(\sum_{j\in S} \pi(j)\Big)\Big(\sum_{j\in S} \frac{1}{\pi(j)} |p_{ij}(n) - \pi(j)|^2\Big) \\ &= \sum_{j\in S} \frac{1}{\pi(j)} \big(p_{ij}(n)^2 - 2p_{ij}(n)\pi(j) + \pi(j)^2\big) \\ &= \sum_{j\in S} \frac{p_{ij}(n)^2}{\pi(j)} - 1, \end{aligned} \tag{12.6}$$

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

$$\sum_{j\in S} \frac{p_{ij}(n)^2}{\pi(j)} = \sum_{j\in S} \frac{p_{ij}(n)p_{ji}(n)}{\pi(j)} \frac{\pi(j)}{\pi(i)} = \frac{1}{\pi(i)} \sum_{j\in S} p_{ij}(n)p_{ji}(n) = \frac{1}{\pi(i)} p_{ii}(2n). \tag{12.7}$$

Par (12.5) et la définition de γ_*,

$$\begin{aligned} p_{ii}(2n) = \sum_{k=1}^{|S|} \pi(i)v_k(i)\lambda_k^{2n}v_k(i) &= \pi(i)v_1(i)^2\lambda_1^{2n} + \pi(i)\sum_{k=2}^{|S|} v_k(i)^2\lambda_k^{2n} \\ &\leqslant \pi(i) + \pi(i)(1-\gamma_*)^{2n}\sum_{k=2}^{|S|} v_k(i)^2. \end{aligned} \tag{12.8}$$

En appliquant (12.2) à la fonction $f = \delta_i$, on obtient

$$\sum_{k=2}^{|S|} v_k(i)^2 = \frac{1}{\pi(i)} \sum_{k=2}^{|S|} \langle \delta_i,\, v_k\rangle_\pi v_k(i) = \frac{1}{\pi(i)} \big(\delta_i(i) - \langle \delta_i\rangle_\pi\big) = \frac{1-\pi(i)}{\pi(i)}. \tag{12.9}$$

La conclusion suit en combinant (12.6), (12.7), (12.8) et (12.9). □

Pour être utile quantitativement, le Théorème 12.4 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.

12.3 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 $d(n) \coloneqq \max_{i\in S} \|\mathbb{P}(X_n = \cdot \,|\, X_0 = i) - \pi\|_{\mathrm{VT}}$. La première notion, celle de temps de mélange, est très naturelle.

Définition 12.5. *Soit $\epsilon > 0$. Le **ϵ-temps de mélange** est défini par*

$$t_{\mathrm{mix}}(\epsilon) \coloneqq \min\{n \in \mathbb{N} \,|\, d(n) \leqslant \epsilon\}.$$

*On définit le **temps de mélange** par*

$$t_{\mathrm{mix}} \coloneqq t_{\mathrm{mix}}(1/4).$$

(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 12.6. *Le **temps de relaxation** est défini par*

$$t_{\mathrm{rel}} := \frac{1}{\gamma_*}.$$

Afin de mieux appréhender ce que mesure cette seconde notion, rappelons qu'il suit du théorème de convergence que $\lim_{n\to\infty} P^n f(i) = \langle f\rangle_\pi$ pour tout $i \in S$. En particulier, $\lim_{n\to\infty} \mathrm{Var}_\pi(P^n f) = 0$. La notion de temps de relaxation permet de rendre cette dernière convergence quantitative : par (12.3) et (12.4), pour toute fonction $f \in \ell^2(\pi)$ et tout $n \in \mathbb{N}^*$,

$$\begin{aligned}\mathrm{Var}_\pi[P^n f] = \sum_{k=2}^{|S|} \langle P^n f,\, v_k\rangle_\pi^2 = \sum_{k=2}^{|S|} \langle \sum_{\ell=1}^{|S|} \langle f,\, v_\ell\rangle_\pi \lambda_\ell^n v_\ell,\, v_k\rangle_\pi^2 = \sum_{k=2}^{|S|} \langle f,\, v_k\rangle_\pi^2 \lambda_k^{2n} \\ \leqslant (1-\gamma_*)^{2n}\,\mathrm{Var}_\pi[f] \leqslant e^{-2n/\gamma_*}\,\mathrm{Var}_\pi[f] = e^{-2n/t_{\mathrm{rel}}},\end{aligned}$$

ce qui montre que $\mathrm{Var}_\pi[P^n f]$ 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 12.7. *Soit P irréductible et réversible sur S. Soit $\pi_{\min} := \min_{i\in S} \pi(i)$. Alors, pour tout $\epsilon > 0$,*

$$t_{\mathrm{mix}}(\epsilon) \leqslant \frac{1}{2}\log\Big(\frac{1-\pi_{\min}}{4\pi_{\min}\epsilon^2}\Big)\, t_{\mathrm{rel}}.$$

Démonstration. Cela suit presque immédiatement du Théorème 12.4. Ce dernier implique que

$$d(n) = \max_{i\in S} \|\mathbb{P}(X_n = \cdot \,|\, X_0 = i) - \pi\|_{\mathrm{TV}} \leqslant \frac{1}{2}\max_{i\in S}\sqrt{\frac{1-\pi(i)}{\pi(i)}}\,(1-\gamma_*)^n \leqslant \sqrt{\frac{1-\pi_{\min}}{4\pi_{\min}}}\,(1-\gamma_*)^n.$$

Comme $(1-\gamma_*)^n \leqslant e^{-\gamma_* n} = e^{-n/t_{\mathrm{rel}}}$, la conclusion suit, puisque

$$\forall \epsilon > 0, \qquad n \geqslant \frac{1}{2}\log\Big(\frac{1-\pi_{\min}}{4\pi_{\min}\epsilon^2}\Big) t_{\mathrm{rel}} \implies d(n) \leqslant \epsilon. \qquad \square$$

Théorème 12.8. *Soit P irréductible, apériodique et réversible sur S. Alors, pour tout $\epsilon > 0$,*

$$t_{\mathrm{mix}}(\epsilon) \geqslant |\log(2\epsilon)|(t_{\mathrm{rel}} - 1).$$

Démonstration. Il suit de l'orthogonalité des vecteurs propres que

$$\forall k \in \{2,\dots,|S|\}, \qquad \sum_{j\in S} \pi(j) v_k(j) = \langle \mathbf{1},\, v_k\rangle_\pi = \langle v_1,\, v_k\rangle_\pi = 0.$$

On a donc, pour tout $i \in S$ et tout $n \in \mathbb{N}^*$,

$$|\lambda_k^n v_k(i)| = |P^n v_k(i)| = \Big|\sum_{j\in S} \big(p_{ij}(n) v_k(j) - \pi(j) v_k(j)\big)\Big| \leqslant \|v_k\|_\infty 2d(n).$$

En particulier, en choisissant $i \in S$ tel que $v_k(i) = \|v_k\|_\infty$, on obtient

$$\forall k \in \{2,\dots,|S|\}, \qquad |\lambda_k|^n \leqslant 2d(n).$$

La meilleure borne est obtenue lorsque $|\lambda_k| = 1 - \gamma_*$, ce qui donne

$$d(n) \geqslant \frac{1}{2}(1-\gamma_*)^n \geqslant \frac{1}{2}e^{-n\gamma_*/(1-\gamma_*)},$$

puisque $1 - x \geqslant e^{-x/(1-x)}$ pour tout $x \in [0, 1)$. On en conclut que

$$t_{\text{mix}}(\epsilon) \geqslant |\log(2\epsilon)|\frac{1-\gamma_*}{\gamma_*} = |\log(2\epsilon)|(t_{\text{rel}} - 1). \qquad \square$$

12.4 Caractérisation variationnelle du trou spectral

Définition 12.9. *La **forme de Dirichlet** associée à la chaîne de Markov $(X_n)_{n\geqslant 0}$ 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

$$\begin{aligned}\mathcal{E}_\pi(f) &= \sum_{i\in S} f(i)\Big(f(i) - \sum_{j\in S} p_{ij}f(j)\Big)\pi(i) = \sum_{i,j\in S}\pi(i)p_{ij}f(i)(f(i)-f(j)) \\ &= \sum_{i,j\in S}\pi(j)p_{ji}f(i)(f(i)-f(j)) = \sum_{i,j\in S}\pi(i)p_{ij}f(j)(f(j)-f(i)),\end{aligned} \tag{12.10}$$

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

$$\mathcal{E}_\pi(f) = \tfrac{1}{2}\sum_{i,j\in S}\pi(i)p_{ij}(f(i)-f(j))^2. \tag{12.11}$$

Un calcul similaire donne

$$\mathrm{Var}_\pi[f] = \tfrac{1}{2}\sum_{i,j\in S}\pi(i)\pi(j)(f(i)-f(j))^2. \tag{12.12}$$

En particulier, $\mathrm{Var}_\pi[f] = 0$ si et seulement si f est un multiple de $\mathbf{1}$.

Théorème 12.10. *On suppose P irréductible et réversible. Alors,*

$$\gamma = \inf\left\{\frac{\mathcal{E}_\pi(f)}{\mathrm{Var}_\pi[f]} \;\middle|\; \mathrm{Var}_\pi[f] \neq 0\right\}.$$

Une inégalité de la forme $\mathrm{Var}_\pi[f] \leqslant c\mathcal{E}_\pi(f)$, pour un $c > 0$, est appelée une **inégalité de Poincaré**.

Démonstration. Observons tout d'abord que (12.10) et (12.12) impliquent que, pour tout $c \in \mathbb{R}$,

$$\frac{\mathcal{E}_\pi(f - c\mathbf{1})}{\mathrm{Var}_\pi[f - c\mathbf{1}]} = \frac{\mathcal{E}_\pi(f)}{\mathrm{Var}_\pi[f]}.$$

On peut donc supposer sans perte de généralité que $\langle f\rangle_\pi = 0$. Il suit alors de (12.2) que

$$f = \sum_{k=2}^{|S|}\langle f,\, v_k\rangle_\pi\, v_k.$$

Par conséquent,

$$\mathcal{E}_\pi(f) = \langle (I-P)f,\, f\rangle_\pi = \sum_{k=2}^{|S|}\sum_{\ell=2}^{|S|}\langle f,\, v_k\rangle_\pi\langle f,\, v_\ell\rangle_\pi(1-\lambda_k)\langle v_k,\, v_\ell\rangle_\pi = \sum_{k=2}^{|S|}\langle f,\, v_k\rangle_\pi^2(1-\lambda_k)$$
$$\geqslant (1-\lambda_2)\sum_{k=2}^{|S|}\langle f,\, v_k\rangle_\pi^2 = (1-\lambda_2)\langle f,\, f\rangle_\pi = (1-\lambda_2)\operatorname{Var}_\pi[f].$$

La conclusion suit puisque l'inégalité ci-dessus est saturée lorsque $f = v_2$. □

12.5 Bornes sur les trous spectraux γ et γ_*

12.5.1 Conséquences directes de la caractérisation variationnelle

Commençons par un corollaire immédiat du Théorème 12.10.

Corollaire 12.11. *On suppose P irréductible et réversible. S'il existe $A > 0$ tel que*

$$\forall f \in \ell^2(\pi), \qquad \operatorname{Var}_\pi[f] \leqslant A\mathcal{E}_\pi(f), \tag{12.13}$$

alors $\gamma \geqslant 1/A$.

Remarque 12.12. *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 12.4. 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).* ◇

Plus généralement, des résultats de comparaison entre chaînes de Markov sont souvent utiles. En voici un autre.

Lemme 12.13. *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 $f : S \to \mathbb{R}$ et tout $i \in S$,*

$$\pi(i) \leqslant A\tilde{\pi}(i) \qquad \text{et} \qquad \mathcal{E}_\pi(f) \geqslant B\mathcal{E}_{\tilde{\pi}}(f).$$

Alors, $\gamma \geqslant \frac{B}{A}\tilde{\gamma}$, où γ et $\tilde{\gamma}$ sont les trous spectraux associés à P et $\tilde{P}$ respectivement.

Démonstration. Observons que

$$\min_{c\in\mathbb{R}}\|f - c\mathbf{1}\|_\pi^2 = \|f - \langle f\rangle_\pi\mathbf{1}\|_\pi^2 = \operatorname{Var}_\pi[f].$$

On a donc

$$\operatorname{Var}_\pi[f] \leqslant \|f - \langle f\rangle_{\tilde{\pi}}\mathbf{1}\|_\pi^2 = \sum_{i\in S}(f(i) - \langle f\rangle_{\tilde{\pi}})^2\pi(i) \leqslant A\sum_{i\in S}(f(i) - \langle f\rangle_{\tilde{\pi}})^2\tilde{\pi}(i) = A\operatorname{Var}_{\tilde{\pi}}][f].$$

Par conséquent,

$$\frac{\mathcal{E}_\pi(f)}{\operatorname{Var}_\pi[f]} \geqslant \frac{B}{A}\frac{\mathcal{E}_{\tilde{\pi}}(f)}{\operatorname{Var}_{\tilde{\pi}}[f]}.$$

La conclusion suit du Théorème 12.10. □

12.5.2 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 12.14. *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 12.15. *Soit P irréductible et réversible. Alors, $\gamma \geqslant 1/\kappa$.*

Démonstration. Il suffit de vérifier que la condition (12.13) est vérifiée avec $A = \kappa$. Tout d'abord,

$$\mathrm{Var}_\pi[f] = \tfrac{1}{2}\sum_{i,j\in S} \pi(i)\pi(j)(f(i) - f(j))^2 = \tfrac{1}{2}\sum_{i,j\in S} \pi(i)\pi(j)\Big[\sum_{e\in\gamma_{ij}} (f(e^+) - f(e^-))\Big]^2.$$

Observons que l'inégalité de Cauchy–Schwarz donne

$$\begin{aligned}\Big[\sum_{e\in\gamma_{ij}} (f(e^+) - f(e^-))\Big]^2 &= \Big[\sum_{e\in\gamma_{ij}} \frac{1}{\sqrt{Q(e)}}\sqrt{Q(e)}(f(e^+) - f(e^-))\Big]^2 \\ &\leqslant \sum_{e'\in\gamma_{ij}} \frac{1}{Q(e')} \sum_{e\in\gamma_{ij}} Q(e)(f(e^+) - f(e^-))^2 \\ &= |\gamma_{ij}|_Q \sum_{e\in\gamma_{ij}} Q(e)(f(e^+) - f(e^-))^2.\end{aligned}$$

Par conséquent,

$$\begin{aligned}\mathrm{Var}_\pi[f] &\leqslant \tfrac{1}{2}\sum_{i,j\in S} \pi(i)\pi(j)|\gamma_{ij}|_Q \sum_{e\in\gamma_{ij}} Q(e)(f(e^+) - f(e^-))^2 \\ &= \tfrac{1}{2}\sum_e Q(e)(f(e^+) - f(e^-))^2 \sum_{\gamma_{ij}\ni e} |\gamma_{ij}|_Q \pi(i)\pi(j) \\ &\leqslant \kappa\tfrac{1}{2}\sum_e Q(e)(f(e^+) - f(e^-))^2 = \kappa\mathcal{E}_\pi(f),\end{aligned}$$

où l'on a utilisé (12.11) pour la dernière identité. □

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 $p_{ii_1}\cdots p_{i_mi} > 0$ et m est impair. Afin de pouvoir satisfaire cette dernière condition, on supposera P apériodique[1]. Soit Σ la collection de tous ces chemins.

De façon similaire à ce que l'on a fait précédemment, on définit

$$|\sigma_i|_Q := \sum_{e\in\sigma_i} \frac{1}{Q(e)}, \qquad \text{et} \qquad \alpha = \alpha(\Sigma) := \max_e \sum_{\sigma_i\ni e} \pi(i)|\sigma_i|_Q.$$

1. Si tous les chemins menant de i à i sont de longueur paire, alors i est de période au moins 2.

Théorème 12.16. *Soit P irréductible, apériodique et réversible. Alors, $\lambda_{|S|} \geqslant -1 + \frac{2}{\alpha}$.*

Démonstration. Il suffit de montrer que, pour tout $f \in \ell^2(\pi)$,

$$\|f\|_\pi^2 \leqslant \frac{\alpha}{2}\big(\|f\|_\pi^2 + \langle Pf,\, f\rangle_\pi\big). \tag{12.14}$$

En effet, appliqué à la fonction $f = v_{|S|}$, cette inégalité devient $1 \leqslant \frac{\alpha}{2}(1 + \lambda_{|S|})$.

Observons tout d'abord que, pour tout $\sigma_i = (i, i_1, \dots, i_m, i) \in \Sigma$,

$$\begin{aligned} f(i) &= \tfrac{1}{2}\big((f(i) + f(i_1)) - (f(i_1) + f(i_2)) + (f(i_2) + f(i_3)) - \cdots + (f(i_m) + f(i))\big) \\ &= \tfrac{1}{2}\sum_{e\in\sigma_i} (-1)^{n(e)}(f(e^-) + f(e^+)), \end{aligned}$$

où l'on a noté $n(e) \coloneqq k$ si $e = (i_k, i_k + 1)$ (avec la convention que $i_0 \coloneqq i$) et utilisé l'hypothèse que $|\sigma_i|$ est impair. Ainsi,

$$\begin{aligned} \|f\|_\pi^2 &= \sum_{i\in S} \tfrac{1}{4}\pi(i)\Big[\sum_{e\in\sigma_i} (-1)^{n(e)}(f(e^-) + f(e^+))\Big]^2 \\ &= \sum_{i\in S} \tfrac{1}{4}\pi(i)\Big[\sum_{e\in\sigma_i} \sqrt{\tfrac{Q(e)}{Q(e)}}(-1)^{n(e)}(f(e^-) + f(e^+))\Big]^2 \\ &\leqslant \sum_{i\in S} \tfrac{1}{4}\pi(i)|\sigma_i|_Q \sum_{e\in\sigma_i} Q(e)(f(e^-) + f(e^+))^2, \end{aligned}$$

où l'on a procédé comme dans la preuve du Théorème 12.15. On obtient donc

$$\|f\|_\pi^2 = \frac{1}{4}\sum_e Q(e)(f(e^-) + f(e^+))^2 \sum_{\sigma_i \ni e} \pi(i)|\sigma_i|_Q \leqslant \frac{\alpha}{4}\sum_e Q(e)(f(e^-) + f(e^+))^2.$$

(12.14) 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^+. □

Exemple 12.17. 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 $p_{ij} \coloneqq \mathbf{1}_{\{i,j\}\in E}/d(i)$, où $d(i) \coloneqq |\{j \in S \,|\, \{i, j\} \in E\}|$ est le degré du sommet i. On vérifie aisément que cette chaîne de Markov est réversible pour la mesure de probabilité $\pi(i) \coloneqq d(i)/2|E|$. En particulier, $Q(\{i, j\}) = \pi(i)p_{ij} = 1/2|E|$ pour tout arête $\{i, j\} \in E$.

On souhaite appliquer le Théorème 12.15. 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,

$$\kappa = \max_e \sum_{\gamma_{ij}\ni e} |\gamma_{ij}|_Q \pi(i)\pi(j) = \max_e \sum_{\gamma_{ij}\ni e} 2|E|\,|\gamma_{ij}| \frac{d(i)}{2|E|}\frac{d(j)}{2|E|} \leqslant K\frac{d_{\max}^2}{2|E|},$$

où $K \coloneqq \max_e \sum_{\gamma_{ij} \ni e} |\gamma_{ij}|$ et $d_{\max} \coloneqq \max_i d(i)$. On conclut donc du Théorème 12.15 que

$$\gamma \geqslant \frac{2|E|}{K d_{\max}^2}.$$

À présent, bornons inférieurement $\lambda_{|S|}$ à l'aide du Théorème 12.16. On suppose donc que le graphe G est tel que la marche aléatoire est apériodique. On a

$$\alpha = \max_e \sum_{\sigma_i \ni e} \pi(i) |\sigma_i|_Q = \max_e \sum_{\sigma_i \ni e} \frac{d(i)}{2|E|} 2|E| \, |\sigma_i| \leqslant K' d_{\max},$$

où $K' \coloneqq \max_e \sum_{\sigma_i \ni e} |\sigma_i|$. Par conséquent,

$$\lambda_{|S|} \geqslant -1 + \frac{2}{K' d_{\max}}.$$

En combinant ces deux bornes, on conclut que

$$\gamma_* \geqslant \min\Big\{ \frac{2|E|}{K d_{\max}^2}, \frac{2}{K' d_{\max}} \Big\}. \qquad \diamond$$

Remarques bibliographiques : La source principale pour ce chapitre est le livre [11], auquel on pourra se référer pour des informations supplémentaires. Il existe de nombreux autres excellents livres sur ce sujet, par exemple [37].

13 Le phénomène de *cutoff*

Soit $(X_n)_{n\geqslant 0}$ une chaîne de Markov irréductible et apériodique sur un espace des états S fini, de loi initiale μ_0 et de matrice de transition P. Comme cela a été démontré dans le cours d'introduction à la théorie des probabilités, une telle chaîne possède toujours une unique distribution stationnaire π et la loi $\mu_n = \mu_0 P^n$, décrivant la chaîne au temps n, converge vers la distribution stationnaire π lorsque $n \to \infty$:

$$\|\mu_n - \pi\|_{\mathrm{VT}} \coloneqq \max_{A\in\mathcal{F}} |\mu_n(A) - \pi(A)| = \tfrac{1}{2}\sum_{i\in S} |\mu_n(i) - \pi(i)| \xrightarrow{n\to\infty} 0.$$

Considérons à présent une famille $\{(X_n^N)_{n\geqslant 0}, N \geqslant 1\}$ de chaînes de Markov irréductibles, chacune sur un espace des états fini. On note $\mu_n^N(\cdot) \coloneqq \mathbb{P}(X_n^N = \cdot\,)$ et π^N la distribution stationnaire associée à la chaîne $(X_n^N)_{n\geqslant 0}$. On dit que cette famille possède un *cutoff* s'il existe $t_N \uparrow \infty$ tel que

$$\lim_{N\to\infty} \|\mu_{c\cdot t_N}^N - \pi^N\|_{\mathrm{VT}} = \begin{cases} 1, & \text{si } c < 1, \\ 0, & \text{si } c > 1. \end{cases}$$

Pour une telle famille, lorsque N est grand, la distance en variation totale entre la loi de la chaîne au temps n et la distribution stationnaire passe très rapidement (à l'échelle de t_N) d'une valeur proche de 1 à une valeur proche de 0 ; *cf.* Figure 13.1.

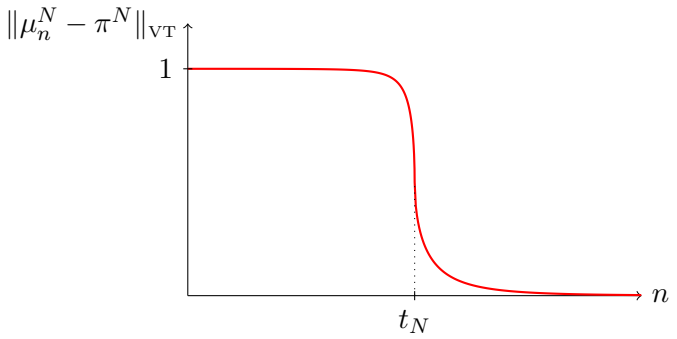


FIGURE 13.1: Illustration du phénomène de *cutoff*.

Dans ce chapitre, nous allons étudier un exemple particulier sur lequel un tel phénomène peut être exhibé de façon relativement simple et explicite.

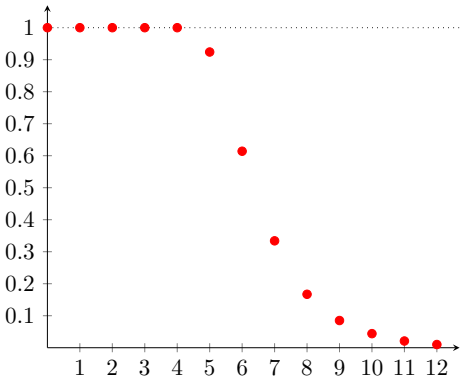


FIGURE 13.2: Gauche : le mélange américain (source : wikipedia). Droite : évolution de la distance en variation totale entre la loi de la distribution des 52 cartes après n répétitions du battage et la loi uniforme.

13.1 Une méthode de battage de cartes

Le résultat le plus célèbre dans ce contexte concerne le battage d'un jeu de cartes à l'aide de la méthode classique du *mélange américain* (plus connue sous ses dénominations en anglais, *riffle shuffle* ou *dovetail shuffle*) : on sépare le paquet en deux tas d'approximativement la même taille, tenus chacun dans une main dans une position légèrement pliées, puis on fait s'entrelacer aléatoirement les cartes des deux tas en les relâchant progressivement. Pour un modèle approprié de cette technique de mélange de cartes, Bayer et Diaconis [8] ont montré qu'il fallait répéter la procédure 7 ou 8 fois pour qu'un jeu de 52 cartes soit bien mélangé : après 6, le paquet présente encore des anomalies statistiques notables, mais en faire plus de 8 n'améliore plus sensiblement le résultat (voir la Fig. 13.2).

Dans ce chapitre, nous allons considérer une méthode de mélange d'un jeu de N cartes dont l'analyse est substantiellement plus simple (quoique peu efficace) : le battage *top-to-random*. Ce dernier revient à itérer la procédure suivante :

- ▷ On prend la carte se trouvant au-dessus.
- ▷ On l'insère dans le paquet au hasard, uniformément (éventuellement à nouveau au sommet de la pile).

Ceci conduit à une chaîne de Markov $(X_n)_{n\geqslant 0}$ sur le groupe S_N des permutations d'un ensemble à N éléments. Comme on le vérifiera plus bas, cette chaîne est irréductible et de loi stationnaire uniforme. En particulier, ce battage conduit bien, asymptotiquement, à un paquet bien mélangé. Une question naturelle est de déterminer le nombre d'itérations nécessaires afin d'obtenir un jeu raisonnablement bien mélangé.

Avant de retourner à notre exemple, considérons le cas général d'une marche aléatoire sur un groupe fini.

13.2 Marche aléatoire sur un groupe fini

Soit G un groupe fini et $\mathbb{Q}$ une mesure de probabilité sur G ; on notera e l'élément neutre de G et g^{-1} l'inverse de $\mathsf{g} \in \mathsf{G}$. Soit $(\mathsf{g}_k)_{k\geqslant 1}$ une suite d'éléments de G, tirés indépendamment selon $\mathbb{Q}$. La **marche aléatoire sur G** est le processus stochastique $(X_n)_{n\geqslant 0}$, à valeurs dans G, défini par $X_0 := \mathsf{e}$ et, pour $n \geqslant 1$, $X_n := \mathsf{g}_n X_{n-1}$. En d'autres termes, les premiers pas de la marche sont $X_0 = \mathsf{e}$, $X_1 = \mathsf{g}_1$, $X_2 = \mathsf{g}_2\mathsf{g}_1$, $X_3 = \mathsf{g}_3\mathsf{g}_2\mathsf{g}_1$, etc.

Clairement, la loi de X_1 est donnée par $\mathbb{Q}$. Déterminons la loi de X_n. Pour $n = 2$, on peut écrire

$$\begin{aligned}\mathbb{P}(X_2 = \mathsf{g}) &= \sum_{\mathsf{h}\in\mathsf{G}} \mathbb{P}(X_2 = \mathsf{g}, X_1 = \mathsf{h}) = \sum_{\mathsf{h}\in\mathsf{G}} \mathbb{P}(\mathsf{g}_2\mathsf{h} = \mathsf{g}, \mathsf{g}_1 = \mathsf{h}) = \sum_{\mathsf{h}\in\mathsf{G}} \mathbb{P}(\mathsf{g}_2 = \mathsf{g}\mathsf{h}^{-1}, \mathsf{g}_1 = \mathsf{h}) \\ &= \sum_{\mathsf{h}\in\mathsf{G}} \mathbb{Q}(\mathsf{g}\mathsf{h}^{-1})\mathbb{Q}(\mathsf{h}) =: \mathbb{Q} * \mathbb{Q}(\mathsf{g}) \equiv \mathbb{Q}^{*2}(\mathsf{g}).\end{aligned}$$

Par induction, on a donc, pour $n \geqslant 3$,

$$\mathbb{P}(X_n = \mathsf{g}) = \sum_{\mathsf{h}\in\mathsf{G}} \underbrace{\mathbb{P}(X_n = \mathsf{g} \,|\, X_{n-1} = \mathsf{h})}_{\mathbb{Q}(\mathsf{g}\mathsf{h}^{-1})} \underbrace{\mathbb{P}(X_{n-1} = \mathsf{h})}_{\mathbb{Q}^{*(n-1)}(\mathsf{h})} = \mathbb{Q}^{*n}(\mathsf{g}),$$

où on a utilisé la notation $\mathbb{Q}^{*n} := \mathbb{Q} * \mathbb{Q}^{*(n-1)}$.

13.3 Retour à l'exemple

Dans l'exemple qui nous intéresse, $\mathsf{G} = \mathsf{S}_N$, le groupe des permutations de l'ensemble $\{1, 2, \ldots, N\}$. On note $[1, 2, \ldots, N]$ l'ordre initial du paquet (la carte 1 correspondant à celle se trouvant sur le paquet). Après n itérations, l'état du paquet est donné par une permutation $\mathsf{g} \in \mathsf{S}_N$, l'ordre des cartes étant $[\mathsf{g}(1), \mathsf{g}(2), \ldots, \mathsf{g}(N)]$. À l'itération suivante, la carte se trouvant sur le paquet, $\mathsf{g}(1)$, est insérée à la $k^{\text{ème}}$ position, k étant choisi uniformément dans l'ensemble $\{1, \ldots, N\}$. On a donc :

	X_n	X_{n+1}	
1	$\mathsf{g}(1)$	$\mathsf{g}(2)$	Permutation cyclique de $(1, \ldots, k)$
2	$\mathsf{g}(2)$	$\mathsf{g}(3)$	
3	$\mathsf{g}(3)$	$\mathsf{g}(4)$	
$\vdots$	$\vdots$	$\vdots$	
$k-1$	$\mathsf{g}(k-1)$	$\mathsf{g}(k)$	
k	$\mathsf{g}(k)$	$\mathsf{g}(1)$	
$k+1$	$\mathsf{g}(k+1)$	$\mathsf{g}(k+1)$	Identité
$\vdots$	$\vdots$	$\vdots$	
N	$\mathsf{g}(N)$	$\mathsf{g}(N)$	

On notera $\mathsf{c}_k \in \mathsf{S}_N$ la permutation correspondant à permuter cycliquement les k premiers éléments en laissant les $N - k$ éléments suivants inchangés. On déduit donc de la discussion précédente que la mesure de probabilité $\mathbb{Q}$ correspondant à un pas de la chaîne est donnée par

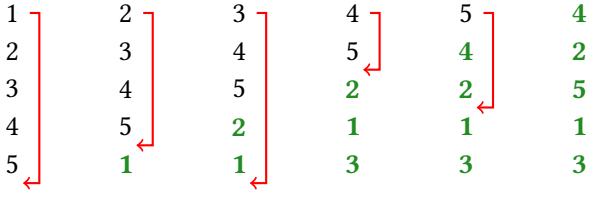
$$\mathbb{Q}(\mathsf{g}) = \begin{cases} 1/N & \text{si } \mathsf{g} \in \{\mathsf{c}_1, \ldots, \mathsf{c}_N\}, \\ 0 & \text{sinon.} \end{cases}$$

Comme on l'a vu précédemment, la loi de la chaîne au temps n est alors donnée par $\mathbb{Q}^{*n}$.

Notons $\mathbb{U}$ la loi uniforme sur S_N : $\mathbb{U}(\mathsf{g}) := 1/|\mathsf{S}_N| = 1/N!$ pour tout $\mathsf{g} \in \mathsf{S}_N$.

Lemme 13.1. *La chaîne de Markov $(X_n)_{n\geqslant 0}$ est irréductible et de loi stationnaire $\mathbb{U}$.*

Démonstration. L'irréductibilité est évidente ; pour passer de $[1, 2, 3, 4, 5]$ à $[4, 2, 5, 1, 3]$, par exemple, on peut procéder ainsi :



Clairement, la même procédure permet d'atteindre n'importe quel ordre.

La stationnarité se vérifie également aisément : pour tout $\mathsf{g} \in \mathsf{S}_N$,

$$\sum_{\mathsf{h}\in\mathsf{S}_N} \mathbb{U}(\mathsf{h})\,\mathbb{Q}(\mathsf{g}\mathsf{h}^{-1}) = \sum_{\tilde{\mathsf{g}}\in\mathsf{S}_N} \underbrace{\mathbb{U}(\tilde{\mathsf{g}}^{-1}\mathsf{g})}_{1/|\mathsf{S}_N|}\,\mathbb{Q}(\tilde{\mathsf{g}}) = \frac{1}{|\mathsf{S}_N|} = \mathbb{U}(\mathsf{g}),$$

la première identité étant obtenue en introduisant $\tilde{\mathsf{g}} = \mathsf{g}\mathsf{h}^{-1}$. □

Remarque 13.2. *Le lecteur devrait vérifier que le résultat précédent (et la preuve) reste vrai pour un groupe fini* G *arbitraire, pour peu que* $\{\mathsf{g} \mid \mathbb{Q}(\mathsf{g}) > 0\}$ *engendre* G. ◇

Nous pouvons à présent énoncer le résultat principal de ce chapitre, qui établit l'existence d'un phénomène de *cutoff* pour le battage de cartes *top-to-random*.

Théorème 13.3. *Soit* $d(n) := \|\mathbb{Q}^{*n} - \mathbb{U}\|_{\mathrm{VT}}$. *Alors,*

1. $d(N\log N + cN) \leqslant e^{-c}$ *pour tout* $c \geqslant 0$ *et tout* $N \geqslant 2$.
2. $d(N\log N - c_N N) \to 1$ *pour toute suite* $(c_N)_{N\geqslant 1}$ *telle que* $c_N \to \infty$.

Afin de démontrer ce résultat, nous aurons besoin de la notion de temps d'arrêt fortement uniforme, que nous introduisons pour une marche aléatoire sur un groupe fini quelconque.

13.4 Temps d'arrêt fortement uniformes

Définition 13.4. *Soit* G *un groupe fini et* $\mathsf{G}^\infty := \{\underline{\mathsf{g}} = (\mathsf{g}_k)_{k\geqslant 1} \mid \mathsf{g}_k \in \mathsf{G}$ *pour tout* $k \geqslant 1\}$. *Une fonction* $\hat{T} : \mathsf{G}^\infty \to \mathbb{N}^\star \cup \{+\infty\}$ *est une* ***règle d'arrêt*** *si*

$$\hat{T}(\underline{\mathsf{g}}) = \ell \implies \hat{T}(\underline{\hat{\mathsf{g}}}) = \ell \text{ pour tout } \underline{\hat{\mathsf{g}}} \in \mathsf{G}^\infty \text{ satisfaisant } \hat{\mathsf{g}}_k = \mathsf{g}_k \text{ pour tout } k \leqslant \ell.$$

Soit $\mathbb{Q}$ une mesure de probabilité sur G et $(X_n)_{n\geqslant 0}$ la marche aléatoire associée. Soit $\hat{T}$ une règle d'arrêt ; on lui associe le temps d'arrêt $T := \hat{T}(X_1, X_2, \ldots)$.

Définition 13.5. *Le temps d'arrêt* T *est* ***fortement uniforme*** *si, pour tout* $k \geqslant 1$,

$$\mathbb{P}(T = k, X_k = \mathsf{g}) \text{ ne dépend pas de } \mathsf{g}.$$

Observez que, si T est fortement uniforme, on a

$$\mathbb{P}(T = k, X_k = \mathsf{g}) = \frac{1}{|\mathsf{G}|}\sum_{\mathsf{h}\in\mathsf{G}} \mathbb{P}(T = k, X_k = \mathsf{h}) = \mathbb{U}(\mathsf{g})\mathbb{P}(T = k).$$

Si $\mathbb{P}(T = k) > 0$, on peut réécrire cette dernière identité sous la forme

$$\mathbb{P}(X_k = \mathsf{g} \mid T = k) = \mathbb{U}(\mathsf{g}), \tag{13.1}$$

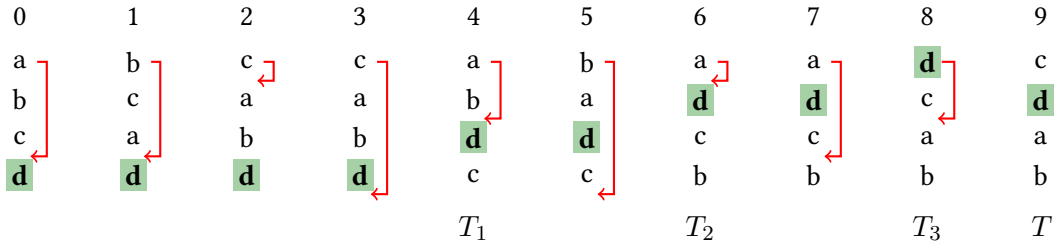
ce qui montre que, lorsque $T < \infty$, la chaîne est distribuée uniformément au temps T. Plus généralement, si $\mathbb{P}(T \leqslant k) > 0$, on a

$$\begin{aligned}
\mathbb{P}(X_k = \mathsf{g} \mid T \leqslant k) &= \sum_{\ell=1}^{k}\sum_{\mathsf{h}\in\mathsf{G}} \mathbb{P}(X_k = \mathsf{g} \mid T = \ell, X_\ell = \mathsf{h})\,\mathbb{P}(X_\ell = \mathsf{h} \mid T = \ell)\,\mathbb{P}(T = \ell \mid T \leqslant k) \\
&= \sum_{\ell=1}^{k}\sum_{\mathsf{h}\in\mathsf{G}} \mathbb{P}(X_k = \mathsf{g} \mid X_\ell = \mathsf{h})\,\underbrace{\mathbb{P}(X_\ell = \mathsf{h} \mid T = \ell)}_{\mathbb{U}(\mathsf{h})}\,\mathbb{P}(T = \ell \mid T \leqslant k) \\
&= \mathbb{U}(\mathsf{g}),
\end{aligned} \tag{13.2}$$

pour tout $\mathsf{g} \in \mathsf{G}$. La deuxième égalité suit de la propriété de Markov et de (13.1), la troisième de la stationnarité de la distribution uniforme $\mathbb{U}$ sur G : $\sum_{\mathsf{h}\in\mathsf{G}} \mathbb{U}(\mathsf{h})\,\mathbb{P}(X_k = \mathsf{g} \mid X_\ell = \mathsf{h}) = \mathbb{U}(\mathsf{g})$.

13.5 Retour à l'exemple

Soit T_k le premier temps auquel k cartes ont été placées sous la carte initialement sous le paquet, et posons $T := T_{N-1} + 1$:



Lemme 13.6. *T est fortement uniforme.*

Démonstration. L'affirmation suit immédiatement des trois observations élémentaires suivantes :

- ▷ Au temps T_k, les k cartes placées sous la carte initialement sous le paquet sont réparties uniformément.
- ▷ Au temps T_{N-1}, la carte initialement sous le paquet se retrouve au-dessus.
- ▷ Au temps $T_{N-1} + 1$, la distribution des N cartes est uniforme. □

13.6 Convergence des marches aléatoires sur les groupes finis

Lemme 13.7. *Soit $\mathbb{Q}$ une mesure de probabilité sur un groupe fini G et T un temps d'arrêt fortement uniforme pour la marche aléatoire associée à $\mathbb{Q}$. Alors,*

$$\|\mathbb{Q}^{*k} - \mathbb{U}\|_{\mathrm{VT}} \leqslant \mathbb{P}(T > k) \text{ pour tout } k \geqslant 0.$$

Démonstration. Soit $\mathsf{A} \subset \mathsf{G}$. On a, en utilisant (13.2),

$$\begin{aligned}\mathbb{Q}^{*k}(\mathsf{A}) = \mathbb{P}(X_k \in \mathsf{A}) &= \mathbb{P}(X_k \in \mathsf{A} \,|\, T \leqslant k)\,\mathbb{P}(T \leqslant k) + \mathbb{P}(X_k \in \mathsf{A} \,|\, T > k)\,\mathbb{P}(T > k) \\ &= \mathbb{U}(\mathsf{A})\,\mathbb{P}(T \leqslant k) + \mathbb{P}(X_k \in \mathsf{A} \,|\, T > k)\,\mathbb{P}(T > k) \\ &= \mathbb{U}(\mathsf{A}) + \big\{\mathbb{P}(X_k \in \mathsf{A} \,|\, T > k) - \mathbb{U}(\mathsf{A})\big\}\mathbb{P}(T > k).\end{aligned}$$

Par conséquent,

$$\big|\mathbb{Q}^{*k}(\mathsf{A}) - \mathbb{U}(\mathsf{A})\big| = \big|\mathbb{P}(X_k \in \mathsf{A} \,|\, T > k) - \mathbb{U}(\mathsf{A})\big| \cdot \mathbb{P}(T > k) \leqslant \mathbb{P}(T > k). \qquad \square$$

13.7 Retour à l'exemple : preuve du Théorème 13.3

Preuve de $d(N \log N + cN) \leqslant e^{-c}$, pour tout $c \geqslant 0$ et tout $N \geqslant 2$.

Soit $M_N = N \log N + cN$. Il suit du Lemme 13.7 que $d(M_N) \leqslant \mathbb{P}(T > M_N)$. Le temps d'arrêt T peut s'écrire comme une somme de variables aléatoires indépendantes :

$$T = T_1 + (T_2 - T_1) + (T_3 - T_2) + \cdots + (T_{N-1} - T_{N-2}) + \underbrace{(T - T_{N-1})}_{=1}. \tag{13.3}$$

Déterminons la loi de $T_{k+1} - T_k$. En T_k, k cartes ont été placées sous la carte initialement sous le paquet. Pour atteindre le temps T_{k+1}, il faut effectuer un mouvement plaçant la carte se trouvant au-dessus du paquet parmi les $k+1$ cartes se trouvant au bas du paquet, ce qui peut se faire de $k+1$ façons

différentes. Ainsi, à chaque itération, la probabilité d'y parvenir est de $(k+1)/N$. On a donc, pour tout $r \geqslant 1$,

$$\mathbb{P}(T_{k+1} - T_k = r) = \Big(1 - \frac{k+1}{N}\Big)^{r-1} \cdot \frac{k+1}{N}. \tag{13.4}$$

Ceci implique que T a la même loi que le nombre V de tirages nécessaires afin de tirer toutes les boules lors d'une série de tirages uniforme avec remise à partir d'une urne contenant N boules numérotées (autrement dit, le problème du collectionneur de coupons). En effet, on peut également décomposer V sous la forme

$$V = (V - V_{N-1}) + (V_{N-1} - V_{N-2}) + \cdots + (V_2 - V_1) + \underbrace{V_1}_{=1},$$

où V_k est le nombre de tirages nécessaire pour obtenir k boules distinctes. Une fois que l'on a tiré k boules distinctes, il reste $N-k$ boules n'ayant pas encore été tirées. La probabilité de tirer l'une d'entre elles est donc de $(N-k)/N$ à chaque tentative. On a donc, pour tout $r \geqslant 1$,

$$\mathbb{P}(V_{k+1} - V_k = r) = \Big(1 - \frac{N-k}{N}\Big)^{r-1} \cdot \frac{N-k}{N}.$$

Il suit que $T_{k+1} - T_k \overset{\text{loi}}{=} V_{N-k} - V_{N-k-1}$ et donc que $T \overset{\text{loi}}{=} V$.

Il nous suffit donc de vérifier que $\mathbb{P}(V > M_N) \leqslant e^{-c}$. Pour $b \in \{1, \dots, N\}$, soit $A_{M_N,b}$ l'événement « la boule b n'a jamais été tirée lors des M_N premiers tirages ». On peut alors écrire

$$\begin{aligned}\mathbb{P}(V > M_N) = \mathbb{P}\Big(\bigcup_{b=1}^{N} A_{M_N,b}\Big) \leqslant \sum_{b=1}^{N} \mathbb{P}(A_{M_N,b}) = N\mathbb{P}(A_{M_N,1}) = N\Big(1 - \frac{1}{N}\Big)^{M_N} \\ \leqslant N\exp(-M_N/N) = e^{-c}.\end{aligned}$$

Preuve de $\lim_{N\to\infty} d(N\log N - c_N N) = 1$, pour toute suite $(c_N)_{N\geqslant 1}$ telle que $c_N \to \infty$.

Fixons une suite $(c_N)_{N\geqslant 1}$ telle que $c_N \to \infty$ et notons $R_N := N\log N - c_N N$.

Soit D_j l'événement « les j cartes initialement sous le paquet sont dans le même ordre relatif qu'au départ ». On a évidemment $\mathbb{U}(D_j) = 1/j!$ pour tout $1 \leqslant j \leqslant N$. Nous allons montrer que

$$\lim_{N\to\infty} \mathbb{Q}^{*R_N}(D_j) = 1 \quad \text{pour tout } j \text{ fixé.} \tag{13.5}$$

Avec ce résultat, on a immédiatement la conclusion désirée : pour tout $j \geqslant 1$,

$$\liminf_{N\to\infty} d(R_N) \geqslant \lim_{N\to\infty} \big(\mathbb{Q}^{*R_N}(D_j) - \mathbb{U}(D_j)\big) = 1 - \frac{1}{j!}.$$

En laissant $j \to \infty$, on conclut que $\lim_{N\to\infty} d(R_N) = 1$ comme désiré.

Il reste à démontrer (13.5). Notons T'_j le temps nécessaire pour que la carte initialement à la $j^{\text{ème}}$ position depuis le bas du paquet se retrouve tout en haut du paquet et soit réintroduite au hasard dans le paquet. Observez que si $T'_j > R_N$, alors les j cartes initialement placées sous le paquet sont encore nécessairement dans le même ordre relatif au temps R_N. On a donc

$$\mathbb{Q}^{*R_N}(D_j) \geqslant \mathbb{P}(T'_j > R_N).$$

Or, un instant de réflexion montre que T'_j et $T - T_{j-1}$ ont la même loi, ce qui implique que

$$\mathbb{Q}^{*R_N}(D_j) \geqslant \mathbb{P}(T - T_{j-1} > R_N).$$

Il suffit donc d'établir que $\lim_{N\to\infty} \mathbb{P}(T - T_{j-1} \leqslant R_N) = 0$. On a vu en (13.4) que $T_{k+1} - T_k$ suit une loi géométrique de paramètre $(k+1)/N$; par conséquent,

$$\mathbb{E}[T_{k+1} - T_k] = \frac{N}{k+1}, \qquad \mathrm{Var}[T_{k+1} - T_k] = \Big(\frac{N}{k+1}\Big)^2\Big(1 - \frac{k+1}{N}\Big).$$

On déduit alors immédiatement de (13.3) et de l'indépendance de ces incréments que, lorsque $N \to \infty$ avec j fixé,

$$\mathbb{E}[T - T_{j-1}] = \sum_{i=j}^{N} \frac{N}{i} = N \log N + \mathsf{O}(N),$$
$$\mathrm{Var}[T - T_{j-1}] = \sum_{i=j}^{N-1} \Big(\frac{N}{i}\Big)^2 \Big(1 - \frac{i}{N}\Big) = \mathsf{O}(N^2).$$

La conclusion suit donc de l'inégalité de Tchebychev : pour j fixé et N suffisamment grand,

$$\mathbb{P}\big(T - T_{j-1} \leqslant R_N\big) \leqslant \mathbb{P}\big(T - T_{j-1} \leqslant \mathbb{E}[T - T_{j-1}] - \tfrac{1}{2} c_N N\big) \leqslant \frac{\mathrm{Var}[T - T_{j-1}]}{(\frac{1}{2} c_N N)^2} = O\big(c_N^{-2}\big),$$

ce qui tend bien vers 0 lorsque N tend vers l'infini.

Remarques bibliographiques : Ce chapitre est basé sur l'article [2]. Le lecteur intéressé peut trouver des informations supplémentaires sur le sujet de ce chapitre dans le livre [38].

14 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.

14.1 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 $(X_n)_{n\geqslant 0}$ 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

$$\forall j \in S, \qquad \lim_{n\to\infty} \mathbb{P}(X_n = i \,|\, X_0 = j) = \pi(i).$$

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.

14.1.1 Construction d'une chaîne de Markov appropriée

Considérons tout d'abord le problème de construire une chaîne de Markov irréductible et apériodique $(X_n)_{n\geqslant 0}$ 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

$$\forall i, j \in S, \qquad \pi(i)p_{ij} = \pi(j)p_{ji}. \tag{14.1}$$

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 (14.1) de la forme

$$\forall i, j \in S \text{ distincts}, \qquad p_{ij} = a_{ij}q_{ij},$$

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 $u \leqslant a_{ij}$, 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 (14.1). 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.

14.1.2 Quelques exemples

Exemple 14.1 (**Algorithme de Metropolis–Hastings**). Observons qu'afin de garantir que $a_{ij} \in [0,1]$, on doit nécessairement avoir $s_{ij} \leqslant 1 + \min\{t_{ij}, t_{ji}\}$. L'algorithme de Metropolis consiste à saturer cette inégalité, ce qui conduit à

$$\forall i, j \in S \text{ distincts}, \qquad a_{ij} = \min\Big\{1, \frac{\pi(j)q_{ji}}{\pi(i)q_{ij}}\Big\}.$$

La première version de cet algorithme a été publiée par Metropolis et Ulam en 1949 [42, 41]. La version générale est due à Hastings [27]. ◇

Exemple 14.2 (**Algorithme de Barker**). L'algorithme proposé par Barker en 1965 [7] consiste à prendre $s_{ij} = 1$ pour tout $i, j \in S$ distincts. On obtient alors

$$\forall i, j \in S \text{ distincts}, \qquad a_{ij} = \frac{\pi(j)q_{ji}}{\pi(j)q_{ji}+\pi(i)q_{ij}}.$$ ◇

Exemple 14.3 (**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 $S \coloneqq \{\sigma = (\sigma_i)_{i\in V} \in \{-1,1\}^V\}$ l'ensemble des **configurations** du modèle (*cf.* Figure 14.1). Le **hamiltonien** du modèle est la fonction $H : S \to \mathbb{R}$ définie par

$$H(\sigma) \coloneqq -\sum_{\{i,j\}\in A} \sigma_i\sigma_j.$$

Pour chaque $\beta \in \mathbb{R}$, on considère la mesure de probabilité sur S définie par

$$\pi_{G,\beta}(\sigma) \coloneqq \frac{e^{-\beta H(\sigma)}}{Z_{G,\beta}}, \qquad Z_{G,\beta} \coloneqq \sum_{\sigma\in S} e^{-\beta H(\sigma)},$$

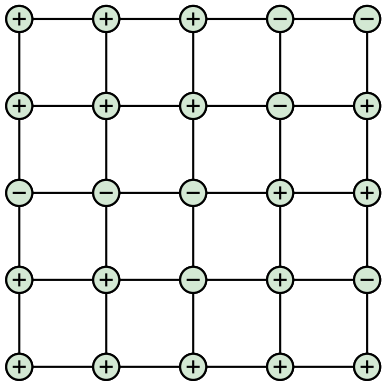


FIGURE 14.1: Une configuration du modèle d'Ising sur le graphe $G = (V, A)$ avec $V = \{1, \dots, 5\}^2$ et $\{i, j\} \in A$ si et seulement si $\|i - j\|_1 = 1$.

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

$$\sigma' := \begin{cases} \sigma^{i+} & \text{si } u \leqslant \dfrac{\pi_{G,\beta}(\sigma^{i+})}{\pi_{G,\beta}(\sigma^{i+}) + \pi_{G,\beta}(\sigma^{i-})} = \Big(1 + \exp\Big[-2\beta \sum_{j \sim i} \sigma_j\Big]\Big)^{-1}, \\ \sigma^{i-} & \text{sinon,} \end{cases} \tag{14.2}$$

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 $i \in V$,

$$\begin{aligned} \pi_{G,\beta}(\sigma^{i+})\, p_{\sigma^{i+}\sigma^{i-}} &= \pi_{G,\beta}(\sigma^{i+}) \frac{1}{|V|} \frac{\pi_{G,\beta}(\sigma^{i-})}{\pi_{G,\beta}(\sigma^{i+}) + \pi_{G,\beta}(\sigma^{i-})} \\ &= \frac{1}{|V|} \frac{\pi_{G,\beta}(\sigma^{i+})\pi_{G,\beta}(\sigma^{i-})}{\pi_{G,\beta}(\sigma^{i+}) + \pi_{G,\beta}(\sigma^{i-})} = \pi_{G,\beta}(\sigma^{i-})\, p_{\sigma^{i-}\sigma^{i+}}. \end{aligned}$$ ◇

14.2 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 12). 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 [44] 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.

14.2.1 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. $f_k : S \to S$ dont la loi $\mathbb{Q}$ est définie par

$$\forall j \in S, \qquad \mathbb{Q}(f_k(i) = j) \coloneqq p_{ij},$$

indépendamment pour chaque $i \in S$. On a alors, pour tout $m, n \in \mathbb{Z}$ tels que $m < n$,

$$\mathbb{P}(X_n = j \,|\, X_m = i) = \mathbb{Q}\big(F_m^n(i) = j\big),$$

où l'on a introduit la notation $F_m^n(i) \coloneqq f_{n-1}\circ f_{n-2}\circ\cdots\circ f_{m+1}\circ f_m(i)$. En effet, on a $\mathbb{Q}(F_m^{m+1}(i) = j) = \mathbb{Q}(f_m(i) = j) = p_{ij} = \mathbb{P}(X_{m+1} = j \,|\, X_m = i)$. Procédons donc par récurrence en supposant que $\mathbb{P}(X_\ell = j \,|\, X_m = i) = \mathbb{Q}\big(F_m^\ell(i) = j\big)$ pour tout $\ell \in [\![m+1, n-1]\!]$ et montrons qu'elle reste vraie pour $\ell = n$:

$$\begin{aligned}
\mathbb{Q}(F_m^n(i) = j) &= \mathbb{Q}(f_{n-1} \circ F_m^{n-1}(i) = j) \\
&= \sum_{k\in S} \mathbb{Q}(f_{n-1}(k) = j,\, F_m^{n-1}(i) = k) \\
&= \sum_{k\in S} \mathbb{Q}(f_{n-1}(k) = j)\mathbb{Q}(F_m^{n-1}(i) = k) \\
&= \sum_{k\in S} \mathbb{P}(X_n = j \,|\, X_{n-1} = k)\mathbb{P}(X_{n-1} = k \,|\, X_m = i) = \mathbb{P}(X_n = j \,|\, X_m = i),
\end{aligned}$$

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 14.4. Considérons la chaîne de Markov $(X_n)_{n\in\mathbb{Z}}$ sur $\{0,1\}$ avec probabilités de transition $p_{01} \coloneqq p$, $p_{00} \coloneqq 1-p$, $p_{10} \coloneqq q$ et $p_{11} \coloneqq 1-q$. Il existe 4 fonctions $g_i : \{0,1\} \to \{0,1\}$, $i = 1,\dots,4$: $g_1(0) = 0$, $g_1(1) = 1$; $g_2(0) = g_2(1) = 0$; $g_3(0) = 1$, $g_3(1) = 0$; $g_4(0) = g_4(1) = 1$. 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

$$\mathbb{Q}(f_k = g_1) = (1-p)(1-q), \quad \mathbb{Q}(f_k = g_2) = (1-p)q, \quad \mathbb{Q}(f_k = g_1) = pq, \quad \mathbb{Q}(f_k = g_1) = p(1-q).$$

La Figure 14.2 illustre cette construction. ◇

14.2.2 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 14.3. 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. $f_k : S \to S$ de loi

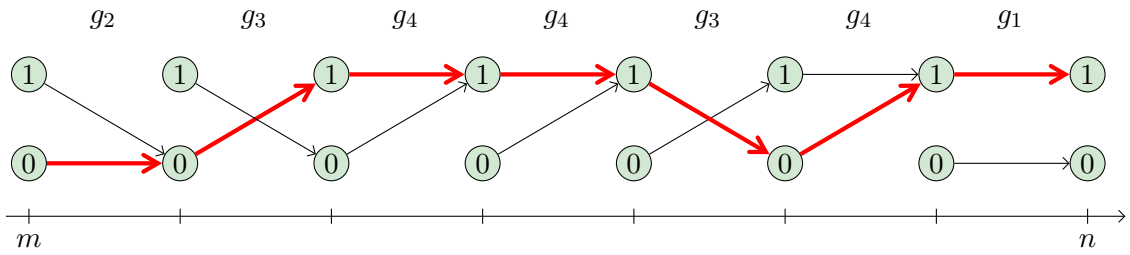


FIGURE 14.2: Une partie d'une réalisation de la trajectoire de la chaîne de Markov $(X_n)_{n\in\mathbb{Z}}$ de l'Exemple 14.4, exprimée en termes des fonctions aléatoires $(f_k)_{k\in\mathbb{Z}}$. Dans cette réalisation, $F_m^n(0) = 1$ (en rouge).

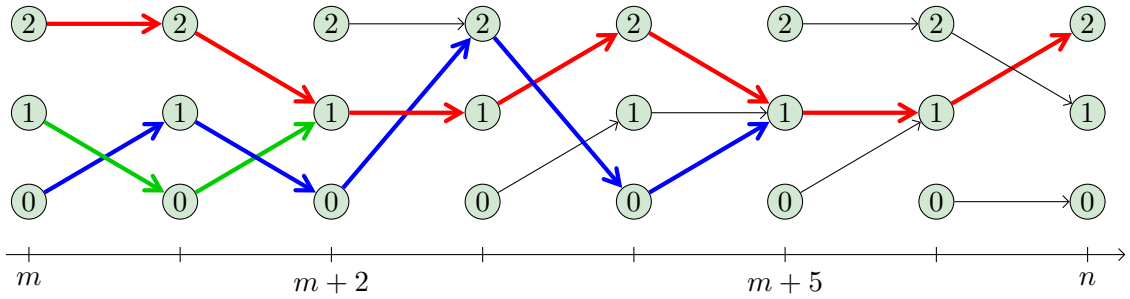


FIGURE 14.3: Trois copies d'une chaîne de Markov sur $S = \{0, 1, 2\}$, partant de trois états différents au temps m. Les trajectoires sont indiquées par les flèches en gras. Les trajectoires issues de 1 et 2 se rejoignent au temps $m + 2$, puis leur trajectoire commune fusionne avec celle issue de 0 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 $F_m^n(i) := f_{n-1} \circ f_{n-2} \circ \cdots \circ f_{m+1} \circ f_m(i)$. 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 14.5. *$\mathbb{Q}$-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 $\{F_{nK}^{(n+1)K} = \text{const}\}$, $n \geqslant 0$, 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 $s > r \geqslant 0$; observant que $F_r^s = \text{const}$ implique $F_0^n = \text{const}$ pour tout $n \geqslant s$, on en déduit que F_0^n sera constante pour tout $n \geqslant (N + 1)K$. □

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 14.4. 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 $F_{-n}^0 = f_{-1} \circ f_{-2} \circ \cdots f_{-n+1} \circ f_{-n}$ est constante pour tout n suffisamment grand, puisque $F_{-n}^0 \overset{\text{loi}}{=} F_0^n$. 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.

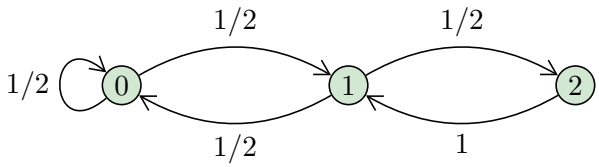


FIGURE 14.4: Un exemple montrant qu'à l'instant où les trajectoires ont toutes fusionné, la chaîne n'est pas nécessairement distribuée selon la loi stationnaire : l'état 2 ne pouvant être atteint qu'en venant de l'état 1, il est impossible que les chaînes soient dans l'état 2 lorsque la fusion de toutes les trajectoires se produit, alors que 2 a évidemment une probabilité strictement positive sous la loi stationnaire.

Théorème 14.6. *Soit M une variable aléatoire $\mathbb{Q}$-presque sûrement finie telle que F^0_{-M} soit constante. Alors, l'unique image de F^0_{-M} est distribuée selon la loi π.*

Démonstration. Fixons $i, j \in S$. Pour tout $n \in \mathbb{N}$,

$$\begin{aligned}\mathbb{Q}(F^0_{-n}(i) = j) &= \mathbb{Q}(F^0_{-n}(i) = j,\, M \leqslant n) + \mathbb{Q}(F^0_{-n}(i) = j,\, M > n)\\ &= \mathbb{Q}(F^0_{-M}(i) = j,\, M \leqslant n) + \mathbb{Q}(F^0_{-n}(i) = j,\, M > n)\\ &= \mathbb{Q}(F^0_{-M}(i) = j) - \mathbb{Q}(F^0_{-M}(i) = j, M > n) + \mathbb{Q}(F^0_{-n}(i) = j,\, M > n).\end{aligned}$$

Donc,

$$\begin{aligned}\big|\mathbb{Q}(F^0_{-M}(i) = j) - \pi(j)\big| &\leqslant \big|\mathbb{Q}(F^0_{-n}(i) = j) - \pi(j)\big| + \big|\mathbb{Q}(F^0_{-M}(i) = j) - \mathbb{Q}(F^0_{-n}(i) = j)\big|\\ &\leqslant \big|\mathbb{P}(X_n = j \,|\, X_0 = i) - \pi(j)\big| + \mathbb{Q}(M > n),\end{aligned}$$

où l'on a utilisé le fait que $\mathbb{Q}(F^0_{-n}(i) = j) = \mathbb{P}(X_n = j \,|\, X_0 = i)$ par construction. La conclusion suit en prenant la limite $n \to \infty$, puisque $\lim_{n\to\infty}\big|\mathbb{P}(X_n = j \,|\, X_0 = i) - \pi(j)\big| = 0$, la chaîne $(X_k)_{k\geqslant 0}$ étant irréductible et apériodique, et $\lim_{n\to\infty} \mathbb{Q}(M > n) = 0$ par hypothèse. □

Remarque 14.7. *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 $F^n_0(i) = F^M_0(i)$ pour tout $n \geqslant M$. Un coup d'œil à la Figure 14.5 (bas) devrait rendre cela évident.* ◇

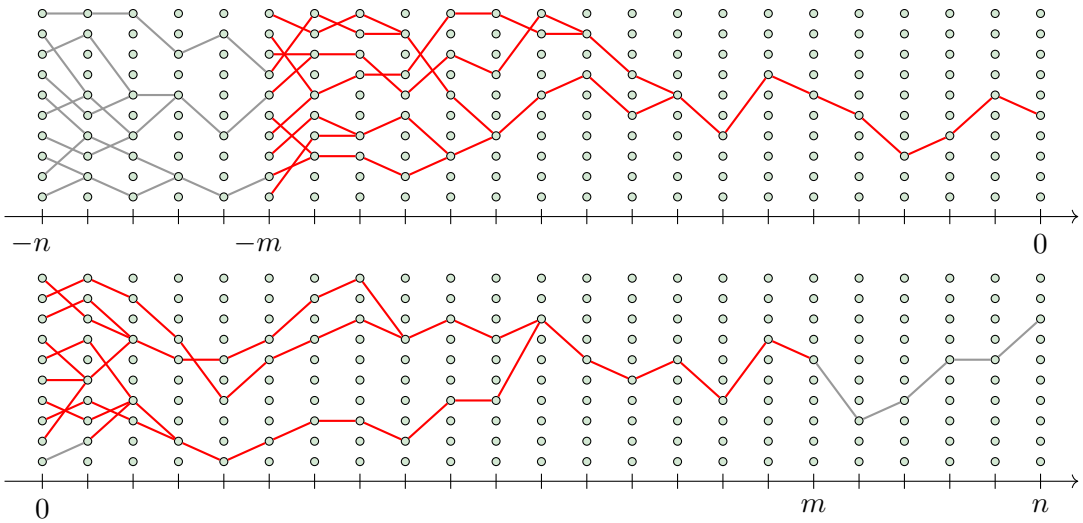


FIGURE 14.5: Haut : couplage depuis le passé ; F^0_{-m} étant constante, c'est également le cas de F^0_{-n}, et leur unique image est identique. Bas : évolution vers le futur ; F^m_0 étant constante, F^n_0 l'est aussi, mais il n'est plus vrai en général que leur unique image est identique.

En résumé, l'algorithme suivant renvoie un élément de S distribué selon π après un temps presque-sûrement fini (voir aussi la Figure 14.6) :

Algorithme 2 : couplage depuis le passé

Initialisation : $n = 0$, $F_0^0 =$ 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

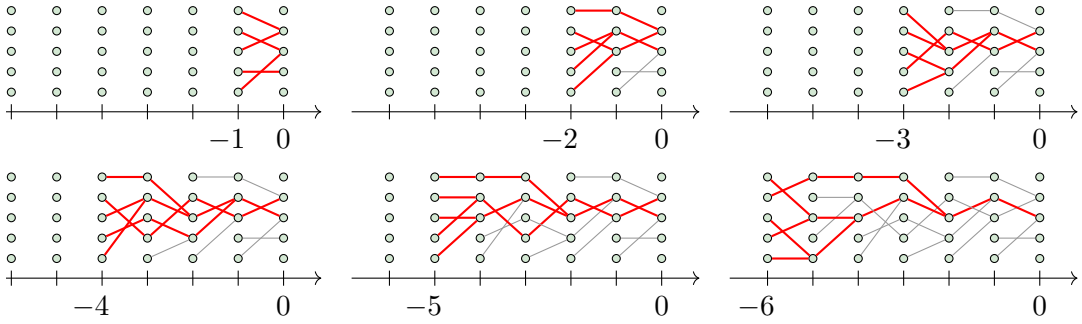


FIGURE 14.6: Une réalisation de l'algorithme de couplage depuis le passé. La fonction F_{-n}^0 ne devient constante qu'à partir de $n = 6$. Observez que le temps auquel les 5 trajectoires fusionnent n'est pas égal à 0, mais à -2.

14.2.3 Quelques commentaire 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 $\lim_{k\to\infty}(-n_k) = -\infty$, et de vérifier successivement, pour $k = 1, 2, \ldots$, 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 [49]).

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 $2^{1\,000\,000} \simeq 10^{301\,030}$. 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 : $\sigma \leqslant \tilde{\sigma}$ si et seulement si $\sigma_i \leqslant \tilde{\sigma}_i$ pour tout $i \in V$. On dit qu'une fonction $f : S \to \mathbb{R}$ est **croissante** si

$$\forall \sigma, \tilde{\sigma} \in S, \qquad \sigma \leqslant \tilde{\sigma} \implies f(\sigma) \leqslant f(\tilde{\sigma}).$$

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 (14.2), c'est-à-dire que l'on pose $f_0(\sigma) = \sigma^{i+}$ avec probabilité

$$\Big(1 + \exp\Big[-2\beta \sum_{j \sim i} \sigma_j\Big]\Big)^{-1}$$

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

$$\sigma \mapsto \Big(1 + \exp\Big[-2\beta \sum_{j \sim i} \sigma_j\Big]\Big)^{-1}$$

est croissante, on constate que ce couplage possède la propriété remarquable suivante :

$$\forall \sigma, \tilde{\sigma} \in S, \qquad \sigma \leqslant \tilde{\sigma} \implies f_0(\sigma) \leqslant f_0(\tilde{\sigma}).$$

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 14.6 reste valide pour cette variante. En effet, la seule chose à vérifier est que le Lemme 14.5 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 [10, 26].

15 Le théorème ergodique

Le but de ce chapitre est de démontrer une version du théorème ergodique dont nous aurons occasionnellement besoin.

15.1 Cadre mathématique et motivation

15.1.1 Transformation préservant une mesure, stationnarité

Le cadre naturel pour énoncer le théorème ergodique est le suivant. On considère un espace probabilisé $(S, \mathcal{S}, \mathbb{Q})$, sur lequel agit une transformation mesurable $\theta : S \to S$ préservant la mesure $\mathbb{Q}$, c'est-à-dire telle que

$$\theta(\mathbb{Q})(A) \coloneqq \mathbb{Q}\big(\theta^{-1}(A)\big) = \mathbb{Q}(A), \qquad \forall A \in \mathcal{S}.$$

On se donne ensuite une variable aléatoire $Y : S \to \mathbb{R}$, et on s'intéresse à l'existence de la limite

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} Y(\theta^k \xi), \tag{15.1}$$

lorsque ξ est un élément aléatoire de S de loi $\mathbb{Q}$.

Du point de vue probabiliste, le problème ci-dessus apparaît de façon naturelle dans le contexte suivant, qui correspond à une extension de la problématique de la loi des grands nombres au-delà du cadre des variables aléatoires indépendantes.

On considère une suite de variables aléatoires $\mathbf{X} = (X_k)_{k\geqslant 0}$ sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. La suite $\mathbf{X}$ est dite stationnaire si

$$(X_n, X_{n+1}, \dots, X_{n+m}) \overset{\text{loi}}{=} (X_0, X_1, \dots, X_m), \qquad \forall n, m \geqslant 0.$$

En particulier, les variables aléatoires X_i sont identiquement distribuées.

Exemple 15.1. Une suite de variables aléatoires i.i.d. est toujours stationnaire. Plus généralement, c'est également le cas de toute chaîne de Markov dont la loi initiale est stationnaire. ◇

La suite $\mathbf{X}$ peut être vue comme une application mesurable de $(\Omega, \mathcal{F})$ dans l'espace $(S, \mathcal{S})$, où $S = \mathbb{R}^{\mathbb{N}}$ est l'ensemble des suites de nombres réels $\mathbf{x} = (x_k)_{k\geqslant 0}$, et $\mathcal{S}$ est la tribu produit correspondante. La mesure $\mathbb{P}$ induit, via $\mathbf{X}$, une mesure de probabilité $\mathbb{Q}$ sur $(S, \mathcal{S})$ par

$$\mathbb{Q}(A) \coloneqq \mathbb{P}(\mathbf{X} \in A), \qquad \forall A \in \mathcal{S}.$$

Sur S, on considère l'endomorphisme mesurable

$$\theta\mathbf{x} = \theta(x_0, x_1, x_2, \ldots) \coloneqq (x_1, x_2, \ldots).$$

Il suit alors de la stationnarité de $\mathbf{X}$ que

$$\mathbb{Q}\big(\theta^{-1}(A)\big) = \mathbb{P}(\theta(X_0, X_1, X_2, \ldots) \in A) = \mathbb{P}((X_1, X_2, \ldots) \in A) = \mathbb{P}((X_0, X_1, \ldots) \in A) = \mathbb{Q}(A),$$

ce qui montre que l'opérateur de translation θ préserve $\mathbb{Q}$.

On s'intéresse à l'existence de la limite des moyennes empiriques

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} X_k.$$

Observez qu'on peut écrire

$$X_k = \pi_0(\theta^k \mathbf{X}),$$

où $\pi_0 : S \to \mathbb{R}$ est défini par $\pi_0(\mathbf{x}) = \pi_0(x_0, x_1, \ldots) = x_0$, et donc que la limite ci-dessus prend la forme

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \pi_0(\theta^k \mathbf{X}),$$

où $\mathbf{X}$ peut être considéré comme un élément aléatoire de S de loi $\mathbb{Q}$. On est donc ramené précisément à une expression du type (15.1).

15.1.2 Événements invariants, ergodicité

Sur $(S, \mathcal{S})$, on introduit la tribu $\mathcal{I} \subset \mathcal{S}$ des événements **invariants sous translation**,

$$\mathcal{I} \coloneqq \big\{A \in \mathcal{S} \,\big|\, \theta^{-1}(A) = A\big\}.$$

La tribu invariante $\mathcal{I}_{\mathbf{X}} \subset \mathcal{F}$ associée à la suite stationnaire $\mathbf{X}$ est définie par

$$\mathcal{I}_{\mathbf{X}} \coloneqq \mathbf{X}^{-1}(\mathcal{I}) = \big\{\mathbf{X}^{-1}(A) \,\big|\, A \in \mathcal{I}\big\}.$$

La suite $\mathbf{X}$ est dite **ergodique** si $\mathcal{I}_{\mathbf{X}}$ est $\mathbb{P}$-triviale, c'est-à-dire si $\mathbb{P}(A) \in \{0, 1\}$, pour tout $A \in \mathcal{I}_{\mathbf{X}}$. Clairement, ceci est équivalent à dire que la tribu $\mathcal{I}$ est $\mathbb{Q}$-triviale. Observez que dans ce cas, une variable aléatoire $Z : S \to \mathbb{R}$ est $\mathcal{I}$-mesurable si et seulement si Z est constante.

Exemple 15.2. Il suit immédiatement de la définition que $\mathcal{I}_{\mathbf{X}} \subset \mathcal{T}$, où $\mathcal{T}$ est la tribu asymptotique. En particulier, si $\mathbf{X}$ est une suite (nécessairement stationnaire) de variables aléatoires i.i.d., la loi $0/1$ de Kolmogorov implique la $\mathbb{P}$-trivialité de $\mathcal{I}_{\mathbf{X}}$ et donc l'ergodicité de $\mathbf{X}$. ◇

15.1.3 Parenthèse sur l'espérance conditionnelle.

Soit X une variable aléatoire intégrable sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, et $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. On appelle **espérance conditionnelle** de X sachant $\mathcal{G}$, notée $\mathbb{E}[X \,|\, \mathcal{G}]$, toute variable aléatoire $Y : \Omega \to \mathbb{R}$ telle que

- ▷ Y est $\mathcal{G}$-mesurable ;
- ▷ $\mathbb{E}\big[Y\mathbf{1}_A\big] = \mathbb{E}\big[X\mathbf{1}_A\big]$, pour tout $A \in \mathcal{G}$.

On peut montrer qu'une telle variable aléatoire Y existe et qu'elle est unique, dans le sens que deux telles variables aléatoires Y_1 et Y_2 (on dit que ce sont deux **versions** de l'espérance conditionnelle) satisfont nécessairement $Y_1 = Y_2$, $\mathbb{P}$-presque partout.

De plus, l'espérance conditionnelle satisfait les propriétés usuelles de l'espérance (*cf.* exercices). En particulier, l'inégalité de Jensen s'applique : si $\psi : \mathbb{R} \to \mathbb{R}$ est convexe, alors

$$\mathbb{E}(\psi(X) \,|\, \mathcal{G}) \geqslant \psi\big(\mathbb{E}(X \,|\, \mathcal{G})\big).$$

Observez finalement que si $\mathcal{G}$ est $\mathbb{P}$-triviale, alors $\mathbb{E}(X \,|\, \mathcal{G}) = \mathbb{E}(X)$ $\mathbb{P}$-presque sûrement.

Remarque 15.3. *Dans le cas particulier où $\mathcal{G} = \sigma(Z)$ est la tribu engendrée par une variable aléatoire $Z : \Omega \to \mathbb{R}$, on écrit simplement $\mathbb{E}[X \mid Z]$ pour l'espérance conditionnelle $\mathbb{E}[X \mid \sigma(Z)]$. On vérifiera en exercice que cette définition se réduit à celle donnée dans le cours d'introduction à la théorie des probabilités.* ◇

15.2 Le théorème ergodique de Birkhoff

Nous sommes à présent en mesure d'énoncer le théorème ergodique de Birkhoff.

Théorème 15.4 (Théorème ergodique de Birkhoff)**.** *Soit θ un endomorphisme mesurable préservant la mesure $\mathbb{Q}$ sur l'espace de probabilité $(S, \mathcal{S}, \mathbb{Q})$. Soit $Y : S \to \mathbb{R}$ une variable aléatoire telle que $Y \in L^p(\mathbb{Q})$ pour un $p \geqslant 1$, et posons $X_k = Y \circ \theta^k$. Alors,*

$$\frac{1}{n}\sum_{k=0}^{n-1} X_k \to \mathbb{E}[X_0 \,|\, \mathcal{I}], \qquad n \to \infty,$$

$\mathbb{Q}$-presque sûrement et dans $L^p(\mathbb{Q})$. En particulier, si la suite X_k est ergodique, alors

$$\frac{1}{n}\sum_{k=0}^{n-1} X_k \to \mathbb{E}[X_0], \qquad n \to \infty,$$

$\mathbb{Q}$-presque sûrement et dans $L^p(\mathbb{Q})$.

Remarque 15.5. *Au vu des Exemples 15.1 et 15.2, ce théorème implique la loi forte des grands nombres pour une suite de variables aléatoires i.i.d..* ◇

15.3 Preuve du théorème ergodique ponctuel

La première observation est que la suite $\mathbf{X} = (X_k)_{k\geqslant 0}$ est stationnaire. La preuve du théorème se fait en trois parties : tout d'abord, on démontre un résultat technique (le lemme ergodique maximal), puis on en déduit la convergence presque sûre, et finalement la convergence dans L^p, à l'aide d'un argument d'intégrabilité uniforme.

15.3.1 Lemme ergodique maximal

Lemme 15.6 (Lemme ergodique maximal)**.** *Sous les hypothèses du théorème, soit $S_n = X_0 + \cdots + X_{n-1}$. Alors*

$$\mathbb{E}(X_0 \mathbf{1}_{\{\sup_{k\geqslant 1} S_k > 0\}}) \geqslant 0.$$

Preuve du Lemme 15.6. On introduit, pour $n \geqslant 1$, $S_n^* \coloneqq X_1 + \cdots + X_n$ et

$$M_n \coloneqq \max(0, S_1, \ldots, S_n), \qquad M_n^* \coloneqq \max(0, S_1^*, \ldots, S_n^*).$$

Lorsque l'événement $M_n > 0$ est réalisé, on peut écrire

$$M_n = \max(0, S_1, \dots, S_n) = \max(S_1, \dots, S_n) = M_{n-1}^* + X_0 \leqslant M_n^* + X_0. \tag{15.2}$$

On peut donc écrire

$$\begin{aligned}\mathbb{E}[X_0 \mathbf{1}_{\{M_n>0\}}] &\geqslant \mathbb{E}[(M_n - M_n^*)\mathbf{1}_{\{M_n>0\}}] \\ &= \mathbb{E}[M_n - M_n^*] + \mathbb{E}[(M_n^* - M_n)\mathbf{1}_{\{M_n=0\}}] \\ &\geqslant \mathbb{E}[M_n - M_n^*] \\ &= 0.\end{aligned}$$

La première inégalité suit de (15.2) ; la seconde du fait que $M_n^* \geqslant 0$; la dernière identité de la stationnarité de $\mathbf{X}$. La conclusion suit du théorème de convergence dominée puisque $\lim_{n\to\infty} \mathbf{1}_{\{M_n>0\}} = \mathbf{1}_{\{\sup_{k\geqslant 1} S_k>0\}}$. □

15.3.2 Convergence presque sûre

Posons $Z_k \coloneqq X_k - \mathbb{E}[X_k \,|\, \mathcal{I}]$. Il suit de la stationnarité de $\mathbf{X}$ que $\mathbb{E}[X_k \,|\, \mathcal{I}] = \mathbb{E}[X_0 \,|\, \mathcal{I}]$. Par conséquent,

$$\frac{1}{n}\sum_{k=0}^{n-1} Z_k = \tfrac{1}{n} S_n - \mathbb{E}[X_0 \,|\, \mathcal{I}].$$

On peut donc, quitte à passer aux variables Z_k, se ramener au cas où $\mathbb{E}[X_0 \,|\, \mathcal{I}] = 0$, $\mathbb{Q}$-presque sûrement.

Soit $\epsilon > 0$. On introduit

$$A \coloneqq \big\{\limsup_{n\to\infty} \tfrac{1}{n} S_n > \epsilon\big\} \quad \text{et} \quad \widehat{X}_k \coloneqq (X_k - \epsilon)\mathbf{1}_A.$$

Observons d'une part que la suite $(\widehat{X}_0, \widehat{X}_1, \dots)$ est stationnaire, puisque $A \in \mathcal{I}$. D'autre part, en posant $\widehat{S}_n \coloneqq \widehat{X}_0 + \cdots + \widehat{X}_{n-1}$, on observe que

$$\big\{\sup_n \widehat{S}_n > 0\big\} = \big\{\sup_n \tfrac{1}{n}\widehat{S}_n > 0\big\} = \big\{\sup_n \tfrac{1}{n} S_n > \epsilon\big\} \cap A = A.$$

Une application du Lemme 15.6 donne donc

$$0 \leqslant \mathbb{E}[\widehat{X}_0 \, \mathbf{1}_{\{\sup_n \widehat{S}_n>0\}}] = \mathbb{E}[(X_0 - \epsilon)\mathbf{1}_A] = \mathbb{E}\big[\mathbb{E}[X_0 \,|\, \mathcal{I}]\mathbf{1}_A\big] - \epsilon\mathbb{Q}(A) = -\epsilon\mathbb{Q}(A),$$

puisque $\mathbb{E}[X_0 \,|\, \mathcal{I}] = 0$, $\mathbb{Q}$-presque sûrement. Comme $\epsilon > 0$, l'inégalité ci-dessus implique donc nécessairement que $\mathbb{Q}(A) = 0$, c'est-à-dire que

$$\limsup_{n\to\infty} \tfrac{1}{n} S_n \leqslant \epsilon, \qquad \mathbb{Q}\text{-presque sûrement,}$$

et donc, ϵ pouvant être pris arbitrairement petit,

$$\limsup_{n\to\infty} \tfrac{1}{n} S_n \leqslant 0, \qquad \mathbb{Q}\text{-presque sûrement.}$$

En appliquant le même raisonnement à la suite stationnaire $-\mathbf{X} = (-X_0, -X_1, \dots)$, on obtient que

$$\liminf_{n\to\infty} \tfrac{1}{n} S_n = -\limsup_{n\to\infty} \tfrac{1}{n}(-S_n) \geqslant 0, \qquad \mathbb{Q}\text{-presque sûrement.}$$

Par conséquent, on a

$$\limsup_{n\to\infty} \tfrac{1}{n} S_n \leqslant 0 \leqslant \liminf_{n\to\infty} \tfrac{1}{n} S_n, \qquad \mathbb{Q}\text{-presque sûrement,}$$

et donc $\lim_{n\to\infty} \frac{1}{n} S_n = 0$, $\mathbb{Q}$-presque sûrement.

15.3.3 Convergence dans L^p

On fixe $r > 0$, et on décompose $X_k = X'_k + X''_k$, où

$$X'_k \coloneqq X_k \mathbf{1}_{\{|X_k| \leqslant r\}} \qquad \text{et} \qquad X''_k \coloneqq X_k \mathbf{1}_{\{|X_k| > r\}}.$$

On note S'_n et S''_n les sommes partielles correspondantes. Évidemment,

$$\|\tfrac{1}{n} S_n - \mathbb{E}[X_0 \,|\, \mathcal{I}]\|_p \leqslant \|\tfrac{1}{n} S'_n - \mathbb{E}[X'_0 \,|\, \mathcal{I}]\|_p + \|\tfrac{1}{n} S''_n - \mathbb{E}[X''_0 \,|\, \mathcal{I}]\|_p$$

Il suffit donc de montrer que le membre de droite tend vers 0 lorsqu'on prend la limite $n \to \infty$, suivie par la limite $r \to \infty$.

Il suit de la première partie de la preuve que $\frac{1}{n} S'_n \to \mathbb{E}[X'_0 \,|\, \mathcal{I}]$, $\mathbb{Q}$-presque sûrement. Comme, par construction, $|\frac{1}{n} S'_n| \leqslant r$, il suit du théorème de convergence dominée que la convergence a également lieu dans L^p.

D'autre part, l'inégalité de Jensen implique que

$$\|\mathbb{E}[X''_0 \,|\, \mathcal{I}]\|_p = \big\{\mathbb{E}\big[|\mathbb{E}[X''_0 \,|\, \mathcal{I}]|^p\big]\big\}^{1/p} \leqslant \big\{\mathbb{E}\big[\mathbb{E}[|X''_0|^p \,|\, \mathcal{I}]\big]\big\}^{1/p} = \|X''_0\|_p.$$

On a donc

$$\|\tfrac{1}{n} S''_n - \mathbb{E}[X''_0 \,|\, \mathcal{I}]\|_p \leqslant \tfrac{1}{n} \sum_{k=0}^{n-1} \|X''_k\|_p + \|\mathbb{E}[X''_0 \,|\, \mathcal{I}]\|_p \leqslant 2\|X''_0\|_p,$$

ce qui tend vers 0 lorsque $r \to \infty$.

Remarques bibliographiques : Ce chapitre est fortement inspiré de [31, chapitre 9]. Le Théorème 9.1 est originellement dû à Birkhoff [9] ; la preuve courte donnée ici est due à Garsia.

16 La marche aléatoire en milieu aléatoire

16.1 Description du processus

Comme cela a été vu dans le cours d'introduction à la théorie des probabilités, les marches aléatoires permettent, parmi de nombreuses autres applications, de décrire le déplacement de « particules » diffusant de façon aléatoire. Dans de nombreuses situations, on se retrouve confronté à des problèmes où le milieu dans lequel la particule se meut est inhomogène et désordonné. Il n'est en général pas raisonnable d'essayer de décrire explicitement le milieu : cette information n'est souvent simplement pas accessible expérimentalement, et deux environnements préparés selon le même processus, bien qu'indistinguables au niveau macroscopique, vont être très différents au niveau microscopique. Une modélisation mathématique possible est alors de représenter le milieu lui-même comme étant issu d'une expérience aléatoire : on commence par tirer un milieu au hasard, selon une loi reproduisant au mieux les propriétés statistiques du milieu que l'on désire modéliser, puis on considère une marche aléatoire se mouvant dans cette réalisation du milieu.

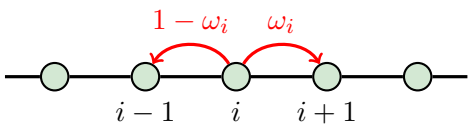
La marche aléatoire en milieu aléatoire fait donc intervenir deux formes d'aléas.

Le milieu. Dans le cas le plus simple, que l'on considère dans ce chapitre, le **milieu aléatoire** est donné par un élément $\omega = (\omega_i)_{i\in\mathbb{Z}} \in \Omega := [0,1]^{\mathbb{Z}}$ (muni de sa tribu borélienne $\mathcal{F}$), choisi selon une mesure de probabilité produit $\mathbb{P}$ sous laquelle les composantes ω_i sont indépendantes et identiquement distribuées[1]. Nous supposerons également que $\mathbb{P}$ satisfait la condition d'ellipticité suivante : il existe $\epsilon > 0$ tel que

$$\mathbb{P}(\omega_0 \in [\epsilon, 1-\epsilon]) = 1. \tag{16.1}$$

La marche aléatoire. Étant donnée une réalisation ω du milieu, la **marche aléatoire en milieu aléatoire** est décrite par la chaîne de Markov $(S_n)_{n\geqslant 0}$ sur $\mathbb{Z}$, avec probabilités de transition données par

$$p(i,i+1) = \omega_i, \qquad p(i,i-1) = 1-\omega_i,$$



pour tout $i \in \mathbb{Z}$. On notera P_x^ω la loi induite sur $(\mathbb{Z}^{\mathbb{N}}, \mathcal{G})$, où $\mathcal{G}$ est la tribu engendrée par les cylindres, telle que $\mathsf{P}_x^\omega(S_0 = x) = 1$.

1. La preuve donnée dans ce chapitre s'étend au cas d'un milieu stationnaire et ergodique [54].

Nous aurons également besoin de la mesure jointe $\mathbf{P}_x \coloneqq \mathbb{P} \otimes \mathsf{P}_x^\omega$ sur $\big(\Omega \times \mathbb{Z}^{\mathbb{N}}, \mathcal{F} \times \mathcal{G}\big)$, définie par[2]

$$\mathbf{P}_x(F \times G) \coloneqq \int_F \mathsf{P}_x^\omega(G)\, \mathbb{P}(\mathrm{d}\omega), \qquad F \in \mathcal{F},\, G \in \mathcal{G}.$$

Observez qu'une propriété de (S_n) est vraie $\mathbf{P}_x$-presque sûrement si et seulement si elle est vraie P_x^ω-presque sûrement pour $\mathbb{P}$-presque tout milieu ω. Dans la suite de ce chapitre, nous serons donc intéressés à obtenir des propriétés $\mathbf{P}_0$-presque certaines du processus (S_n).

Remarque 16.1. *Il est important d'observer que la loi de (S_n) sous $\mathbf{P}_x$ n'est pas markovienne ! En effet, on a, par exemple,*

$$\mathbf{P}_0(S_1 = 1) = \mathbb{E}[\omega_0],$$

alors que

$$\mathbf{P}_0(S_3 = 1 \,|\, S_1 = 1, S_2 = 0) = \frac{\mathbf{P}_0(S_1 = 1, S_2 = 0, S_3 = 1)}{\mathbf{P}_0(S_1 = 1, S_2 = 0)} = \frac{\mathbb{E}[\omega_0(1-\omega_1)\omega_0]}{\mathbb{E}[\omega_0(1-\omega_1)]} = \frac{\mathbb{E}[\omega_0^2]}{\mathbb{E}[\omega_0]}.$$

Clairement, ces deux quantités ne sont égales que si $\mathbb{E}[\omega_0^2] = \mathbb{E}[\omega_0]^2$, c'est-à-dire si ω_0 est déterministe. Intuitivement, la non-markovianité est due au fait que l'observation de la trajectoire de la marche nous renseigne sur le milieu (qui est inconnu sous $\mathbf{P}_0$, sauf dans le cas déterministe). ◇

16.2 Récurrence et transience

Afin de formuler les résultats de façon aussi simple que possible, il est utile d'introduire la notation $\rho_i \coloneqq (1-\omega_i)/\omega_i$.

Théorème 16.2. *On suppose que $\mathbb{E}[\log \rho_0]$ est bien définie (éventuellement égale à $\pm\infty$). Alors, la classification suivante est vraie $\mathbf{P}_0$-presque sûrement :*

1. *si $\mathbb{E}[\log \rho_0] < 0$, alors $\lim_{n\to\infty} S_n = +\infty$;*
2. *si $\mathbb{E}[\log \rho_0] > 0$, alors $\lim_{n\to\infty} S_n = -\infty$;*
3. *si $\mathbb{E}[\log \rho_0] = 0$, alors $\liminf_{n\to\infty} S_n = -\infty$ et $\limsup_{n\to\infty} S_n = +\infty$.*

Démonstration. Fixons tout d'abord une réalisation ω du milieu. Conditionnellement à cette dernière, S_n est une chaîne de Markov. Soient $m_-, m_+ \in \mathbb{N}^*$. Notons $T_i = \inf\{n \geqslant 0 \,|\, S_n = i\}$[3]. Nous allons à présent déterminer la probabilité $\mathsf{P}_0^\omega(T_{-m_-} < T_{m_+})$ que la chaîne partant de 0 atteigne $-m_-$ avant m_+. Cet événement est bien défini, puisque la marche restreinte à l'intervalle $\{-m_-, \dots, m_+\}$ atteint $\{-m_-, m_+\}$ avec probabilité 1 (c'est une chaîne de Markov irréductible sur un espace d'états fini).

Afin d'évaluer cette probabilité, on peut utiliser les méthodes introduites dans le chapitre 11. On vérifie aisément que les probabilités de transition de la marche (S_n) correspondent au choix suivant de conductances :

$$c_{j,j+1} = \begin{cases} \prod_{k=j+1}^{0} \rho_k & \text{si } j < 0, \\ 1 & \text{si } j = 0, \\ \prod_{k=1}^{j} \rho_k^{-1} & \text{si } j > 0. \end{cases}$$

On a un problème de résistances en série. Notons $A_{m_+}^\omega$ la résistance entre 0 et m_+, et $B_{m_-}^\omega$ la résistance entre 0 et $-m_-$. On obtient

$$A_{m_+}^\omega = 1 + \sum_{j=1}^{m_+-1} \prod_{k=1}^{j} \rho_k, \qquad B_{m_-}^\omega \coloneqq \sum_{j=-m_-}^{-1} \prod_{k=j+1}^{0} \frac{1}{\rho_k}.$$

2. On peut montrer que l'application $\omega \mapsto \mathsf{P}_x^\omega(G)$, $G \in \mathcal{G}$, est mesurable.
3. Avec la convention usuelle que $\inf \varnothing \coloneqq \infty$.

La probabilité recherchée est donc donnée par

$$\mathsf{P}_0^\omega\big(T_{-m_-} < T_{m_+}\big) = \frac{A^\omega_{m_+}}{A^\omega_{m_+} + B^\omega_{m_-}}. \tag{16.2}$$

Les propriétés de récurrence et transience de la marche s'expriment facilement en termes du comportement asymptotique des fonctions $A^\omega_{m_+}$ et $B^\omega_{m_-}$. On introduit à cet effet les événements

$$\mathcal{A} \coloneqq \big\{\lim_{m_+\to\infty} A^\omega_{m_+} < \infty\big\}, \qquad \mathcal{B} \coloneqq \big\{\lim_{m_-\to\infty} B^\omega_{m_-} < \infty\big\}.$$

On déduit immédiatement les conclusions suivantes de (16.2).

▷ Si $\omega \in \mathcal{R} \coloneqq \mathcal{A}^{\mathrm{c}} \cap \mathcal{B}^{\mathrm{c}}$, alors

$$\forall m_+ \in \mathbb{N}^*, \qquad \lim_{m_-\to\infty} \mathsf{P}_0^\omega\big(T_{-m_-} < T_{m_+}\big) = 0,$$
$$\forall m_- \in \mathbb{N}^*, \qquad \lim_{m_+\to\infty} \mathsf{P}_0^\omega\big(T_{-m_-} < T_{m_+}\big) = 1.$$

En particulier, pour tout $k > 0$, $\mathsf{P}_0^\omega\big(T_k = \infty\big) = \lim_{m_-\to\infty} \mathsf{P}_0^\omega\big(T_{-m_-} < T_k\big) = 0$ et, similairement, $\mathsf{P}_0^\omega\big(T_{-k} = \infty\big) = 0$. On en conclut aisément que, P_0^ω-presque sûrement, chaque $k \in \mathbb{Z}$ est visité infiniment souvent, et donc que

$$\mathsf{P}_0^\omega\big(\liminf_{n\to\infty} S_n = -\infty,\ \limsup_{n\to\infty} S_n = +\infty\big) = 1.$$

▷ Si $\omega \in \mathcal{T}_+ \coloneqq \mathcal{A} \cap \mathcal{B}^{\mathrm{c}}$, alors

$$\lim_{m_-\to\infty} \lim_{m_+\to\infty} \mathsf{P}_0^\omega\big(T_{-m_-} < T_{m_+}\big) = 0.$$

Il suit que, pour tout $\delta > 0$, on peut trouver un nombre $m^\omega_-(\delta)$ suffisamment grand pour que $\mathsf{P}_0^\omega(T_{-m^\omega_-(\delta)} = \infty) = \lim_{m_+\to\infty} \mathsf{P}_0^\omega\big(T_{-m^\omega_-(\delta)} \geqslant T_{m_+}\big) \geqslant 1 - \delta$. On a donc

$$\mathsf{P}_0^\omega(\exists m_-\ :\ T_{-m_-} = \infty) \geqslant \mathsf{P}_0^\omega(T_{-m^\omega_-(\delta)} = \infty) \geqslant 1 - \delta.$$

$\delta > 0$ étant arbitraire, on en conclut que la trajectoire est P_0^ω-presque sûrement bornée inférieurement : $\liminf_{n\to\infty} S_n > -\infty$. Supposons que $\liminf_{n\to\infty} S_n = m < \infty$. Dans ce cas, m doit être visité infiniment souvent. Or, la condition d'ellipticité (16.1) et la propriété de Markov forte impliquent qu'alors $m - 1$ devrait également être visité infiniment souvent, ce qui est impossible puisque cela contredirait la définition de m. On en conclut que $\liminf_{n\to\infty} S_n = +\infty$, et donc $\lim_{n\to\infty} S_n = +\infty$, P_0^ω-presque sûrement.

▷ Si $\omega \in \mathcal{T}_- \coloneqq \mathcal{A}^{\mathrm{c}} \cap \mathcal{B}$, alors

$$\lim_{m_+\to\infty} \lim_{m_-\to\infty} \mathsf{P}_0^\omega\big(T_{-m_-} < T_{m_+}\big) = 1,$$

d'où l'on déduit, en raisonnant comme précédemment, que

$$\mathsf{P}_0^\omega(S_n \to -\infty) = 1.$$

Afin de conclure la démonstration, il ne nous reste donc plus qu'à établir les relations

$$\mathbb{E}[\log \rho_0] < 0 \implies \mathbb{P}(\mathcal{T}_+) = 1,$$
$$\mathbb{E}[\log \rho_0] > 0 \implies \mathbb{P}(\mathcal{T}_-) = 1,$$
$$\mathbb{E}[\log \rho_0] = 0 \implies \mathbb{P}(\mathcal{R}) = 1.$$

Observons tout d'abord que $\prod_{k=1}^{j} \rho_{k+1} = \frac{\rho_{j+1}}{\rho_1} \prod_{k=1}^{j} \rho_k$, avec $\epsilon^2 < \rho_{j+1}/\rho_1 < \epsilon^{-2}$, par la condition d'ellipticité (16.1). Cela implique que $\mathcal{A}$ appartient à la tribu asymptotique (en fait, à celle des événements invariants sous translation) et est donc de probabilité 0 ou 1. Un argument similaire montre que cela est également le cas pour $\mathcal{B}$.

Supposons que $\mathbb{P}(\mathcal{A}) = 1$. Dans ce cas, la série $\sum_{j\geqslant 1} \prod_{k=1}^{j} \rho_k$ est $\mathbb{P}$-presque-sûrement convergente et, par conséquent, $\lim_{n\to\infty} \rho_1 \cdots \rho_n = 0$. Il suit que $\lim_{n\to\infty} (\rho_1 \cdots \rho_n)^{-1} = \infty$, $\mathbb{P}$-presque sûrement. Or, $\rho_{-n+1} \cdots \rho_0 \overset{\text{loi}}{=} \rho_1 \cdots \rho_n$, et donc la série $\sum_{j\leqslant -1} \prod_{k=j+1}^{0} (1/\rho_k)$ diverge, $\mathbb{P}$-presque sûrement. On en déduit que $\mathbb{P}(\mathcal{B}) = 0$.

Le même argument montrant que $\mathbb{P}(\mathcal{B}) = 1 \implies \mathbb{P}(\mathcal{A}) = 0$, il nous suffit donc d'établir que

$$\mathbb{E}[\log \rho_0] < 0 \iff \mathbb{P}(\mathcal{A}) = 1, \qquad \mathbb{E}[\log \rho_0] > 0 \iff \mathbb{P}(\mathcal{B}) = 1.$$

Ces deux affirmations étant symétriques, il suffit de démontrer la première.

Supposons donc, pour commencer, que $\mathbb{E}[\log \rho_0] < 0$. La loi forte des grands nombres implique[4] alors que, $\mathbb{P}$-presque sûrement, $\frac{1}{n}\sum_{k=1}^{n} \log \rho_k \to \mathbb{E}[\log \rho_0] < 0$, et donc, en particulier, qu'il existe $n_0 = n_0(\omega)$, $\mathbb{P}$-presque sûrement fini, et $c > 0$ tel que $\sum_{k=1}^{n} \log \rho_k < -cn$, pour tout $n > n_0(\omega)$. Par conséquent,

$$\sum_{j\geqslant 1} \prod_{k=1}^{j} \rho_k \leqslant \sum_{j=1}^{n_0(\omega)} \prod_{k=1}^{j} \rho_k + \sum_{j>n_0(\omega)} e^{-cj} < \infty,$$

d'où l'on tire $\mathbb{P}(\mathcal{A}) = 1$.

Supposons à présent que $\mathbb{E}[\log \rho_0] > 0$. Par symétrie, on déduit du cas précédent que $\mathbb{P}(\mathcal{B}) = 1$ et, donc, que $\mathbb{P}(\mathcal{A}) = 0$.

Il reste à considérer le cas $\mathbb{E}[\log \rho_0] = 0$. Observons tout d'abord que $\mathbb{P}(\mathcal{A}) = 0$, dans ce cas, si $\mathbb{P}(\rho_0 = 1) = 1$. On peut donc supposer que la loi de $\log \rho_0$ possède une variance strictement positive (et finie, par la condition d'ellipticité), ce qui permet d'appliquer le théorème central limite. Évidemment,

$$\mathbb{P}(\mathcal{A}^{\mathrm{c}}) = \lim_{M\to\infty} \mathbb{P}\Big(\sum_{j\geqslant 1} \prod_{k=1}^{j} \rho_k > M\Big).$$

Or, les ρ_k étant tous positifs, on a, pour tout $n \geqslant 1$,

$$\mathbb{P}\Big(\sum_{j\geqslant 1} \prod_{k=1}^{j} \rho_k > M\Big) \geqslant \mathbb{P}\Big(\prod_{k=1}^{n} \rho_k > M\Big) = \mathbb{P}\Big(\sum_{k=1}^{n} \log \rho_k > \log M\Big),$$

et cette dernière probabilité tend vers $1/2$, lorsque $n \to \infty$, par le théorème central limite. On en déduit donc que $\mathbb{P}(\mathcal{A}^{\mathrm{c}}) \geqslant \frac{1}{2}$, et par conséquent, $\mathcal{A}$ étant un événement asymptotique, que $\mathbb{P}(\mathcal{A}) = 0$. □

16.3 Loi des grands nombres

Maintenant que nous sommes en possession de critères caractérisant les régimes de transience et récurrence, une question naturelle est de déterminer la vitesse à laquelle le processus part vers l'infini. En d'autres termes, on désire étudier la convergence $\mathbf{P}_0$-presque-sûre de $\frac{1}{n}S_n$.

4. Observez que la loi forte des grands nombres reste valide lorsque l'espérance est infinie. Soit $(X_k)_{k\geqslant 1}$ une collection de variables aléatoires i.i.d. telles que $\mathbb{E}[(X_1)_+] = +\infty$ et $\mathbb{E}[(X_1)_-] < \infty$. Soit $X_k^M := \min\{X_k, M\}$ et $S_n^M := X_1^M + \cdots + X_n^M$. La loi forte des grands nombres donne $\liminf S_n/n \geqslant \lim S_n^M/n = \mathbb{E}[X_1^M]$ presque-sûrement. Par le théorème de convergence monotone, $\lim_{M\to\infty} \mathbb{E}[(X_1^M)_+] = \mathbb{E}[(X_1)_+] = +\infty$ et donc $\mathbb{E}[X_1^M] = \mathbb{E}[(X_1^M)_+] - \mathbb{E}[(X_1^M)_-] \to +\infty$ lorsque $M \to \infty$.

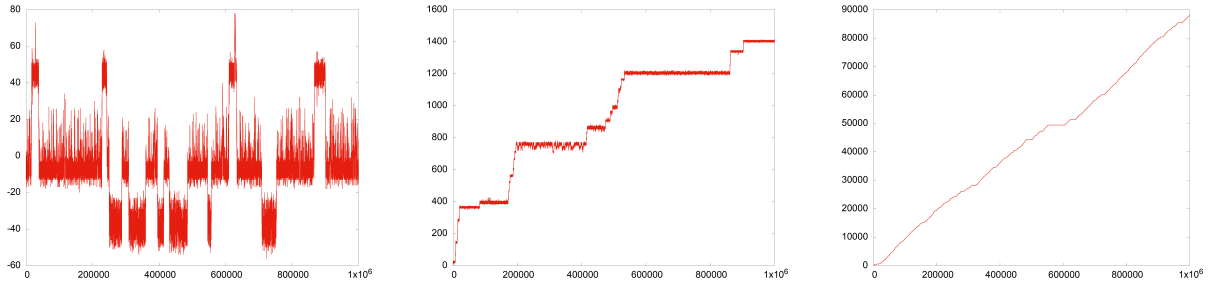


FIGURE 16.1: Trois trajectoires de la marche aléatoire en milieu aléatoire pour une réalisation de l'environnement échantillonné à partir de trois distributions différentes. Gauche : $\mathbb{P}(\omega_0 = \frac{3}{4}) = \mathbb{P}(\omega_0 = \frac{1}{4}) = \frac{1}{2}$, ce qui correspond à un cas récurrent. Milieu : $\mathbb{P}(\omega_0 = \frac{3}{4}) = \frac{5}{8}$, $\mathbb{P}(\omega_0 = \frac{1}{4}) = \frac{3}{8}$, ce qui correspond à un cas transient avec vitesse nulle. Droite : $\mathbb{P}(\omega_0 = \frac{3}{4}) = \frac{4}{5}$, $\mathbb{P}(\omega_0 = \frac{1}{4}) = \frac{1}{5}$, ce qui correspond à un cas transient avec vitesse positive. Notez la grande différence dans les distances parcourues (graduation verticale).

Théorème 16.3. *1. Si $\mathbb{E}[\rho_0] < 1$, alors, $\mathbf{P}_0$-presque sûrement,*

$$\lim_{n\to\infty} \tfrac{1}{n} S_n = \frac{1 - \mathbb{E}[\rho_0]}{1 + \mathbb{E}[\rho_0]}.$$

2. Si $\mathbb{E}[1/\rho_0] < 1$, alors, $\mathbf{P}_0$-presque sûrement,

$$\lim_{n\to\infty} \tfrac{1}{n} S_n = -\frac{1 - \mathbb{E}[1/\rho_0]}{1 + \mathbb{E}[1/\rho_0]}.$$

3. Si $\mathbb{E}[\rho_0] \wedge \mathbb{E}[1/\rho_0] \geqslant 1$, alors, $\mathbf{P}_0$-presque sûrement,

$$\lim_{n\to\infty} \tfrac{1}{n} S_n = 0.$$

Remarque 16.4. *Il suit de l'inégalité de Jensen que $\mathbb{E}[\log \rho_0] < \log \mathbb{E}[\rho_0]$ lorsque $\mathbb{P}$ n'est pas dégénérée. Par conséquent, le régime pour lequel la vitesse est nulle ne coïncide pas avec le régime où la marche est récurrente ! Un exemple pour lequel la marche est transiente (vers $+\infty$), mais possède une vitesse nulle est le suivant : $\mathbb{P}(\omega_0 = \frac{3}{4}) = \frac{5}{8}$ et $\mathbb{P}(\omega_0 = \frac{1}{4}) = \frac{3}{8}$ (dans ce cas, on trouve $\mathbb{E}[\rho_0] \wedge \mathbb{E}[1/\rho_0] = \frac{4}{3} \wedge 2 > 1$, mais $\mathbb{E}[\log \rho_0] = -\frac{1}{4}\log 3 < 0$). Des trajectoires typiques pour chacun des régimes (marche récurrente, transiente avec vitesse nulle et transiente avec vitesse non-nulle) sont représentées sur la Figure 16.1.* ◇

16.3.1 Preuve du Théorème 16.3

Sans perte de généralité, nous supposerons que $\limsup_{n\to\infty} S_n = +\infty$, $\mathbf{P}_0$-presque sûrement.

Nous allons déterminer la limite de $\frac{1}{n} S_n$ en faisant usage du théorème ergodique. On ne peut pas travailler directement avec les incréments $S_{n+1} - S_n$, car ceux-ci ne sont même pas stationnaires (sauf si $\mathbb{P}$ est dégénérée) : on vérifie, par exemple, facilement qu'imposer $\mathbf{P}_0(S_1 - S_0 = 1, S_2 - S_1 = 1) = \mathbf{P}_0(S_2 - S_1 = 1, S_3 - S_2 = 1)$ revient à exiger que $\mathbb{E}[\omega_0^2] = \mathbb{E}[\omega_0]^2$. On doit donc construire une autre famille de variables aléatoires plus appropriées.

Rappelons que $T_k = \inf\{n \geqslant 0 \,|\, S_n = k\}$ est le temps de première visite en $k \in \mathbb{Z}$. On introduit, pour $k \geqslant 1$,

$$\tau_k := T_k - T_{k-1},$$

avec la convention que $\tau_k = \infty$ si $T_k = \infty$. Les propriétés suivantes, démontrées plus bas, sont cruciales pour l'analyse qui va suivre : elles vont nous permettre d'avoir recours au théorème ergodique.

Lemme 16.5. *Si* $\limsup_{n\to\infty} S_n = +\infty$, $\mathbf{P}_0$*-presque sûrement, alors la suite* $(\tau_k)_{k\geqslant 1}$ *est stationnaire et ergodique sous* $\mathbf{P}_0$.

Lemme 16.6. *Notons* $\mathbf{E}_0$ *l'espérance sous la mesure* $\mathbf{P}_0$. *Alors,*

$$\mathbf{E}_0[\tau_1] = \begin{cases} (1+\mathbb{E}[\rho_0])/(1-\mathbb{E}[\rho_0]), & \text{si } \mathbb{E}[\rho_0] < 1, \\ +\infty, & \text{si } \mathbb{E}[\rho_0] \geqslant 1. \end{cases}$$

Lorsque $\mathbf{E}_0[\tau_1] < \infty$, il suit de ces deux lemmes et du théorème ergodique que

$$\lim_{n\to\infty} \tfrac{1}{n} T_n = \lim_{n\to\infty} \tfrac{1}{n} \sum_{i=1}^{n} \tau_i = \mathbf{E}_0[\tau_1], \qquad \mathbf{P}_0\text{-presque sûrement.}$$

En fait ce résultat reste vrai si $\mathbf{E}_0[\tau_1] = \infty$. En effet, on peut appliquer le théorème ergodique à la suite $(\tau_i \wedge M)$, $M \in \mathbb{N}$, ce qui donne

$$\liminf_{n\to\infty} \tfrac{1}{n} T_n = \liminf_{n\to\infty} \tfrac{1}{n} \sum_{i=1}^{n} \tau_i \geqslant \lim_{n\to\infty} \tfrac{1}{n} \sum_{i=1}^{n} (\tau_i \wedge M) = \mathbf{E}_0[\tau_1 \wedge M],$$

et le membre de droite tend vers l'infini lorsque $M \to \infty$ par convergence monotone.

On doit à présent considérer deux cas.

Cas 1 : $\mathbf{E}_0[\tau_1] < \infty$. Notons $M_n := \max\{S_k \,|\, 0 \leqslant k \leqslant n\}$, $n \geqslant 1$. Par définition, $T_{M_n} \leqslant n < T_{M_n+1}$. On en déduit que

$$\frac{T_{M_n}}{M_n} \leqslant \frac{n}{M_n} < \frac{T_{M_n+1}}{M_n+1} \frac{M_n+1}{M_n}.$$

À présent, $M_n \to \infty$, par hypothèse, et $T_n/n \to \mathbf{E}_0[\tau_1]$, lorsque $n \to \infty$. Par conséquent, on obtient

$$\lim_{n\to\infty} \frac{M_n}{n} = \frac{1}{\mathbf{E}_0[\tau_1]}, \qquad \mathbf{P}_0\text{-presque sûrement.}$$

On a donc également

$$\lim_{n\to\infty} \frac{T_{M_n}}{n} = \lim_{n\to\infty} \frac{T_{M_n}}{M_n} \frac{M_n}{n} = 1, \qquad \mathbf{P}_0\text{-presque sûrement.}$$

D'autre part, les incréments de la marche étant de ± 1, on a l'encadrement $M_n \geqslant S_n \geqslant M_n - (n - T_{M_n})$, c'est-à-dire

$$\frac{M_n}{n} \geqslant \frac{S_n}{n} \geqslant \frac{M_n}{n} - (1 - \frac{T_{M_n}}{n}),$$

ce qui montre que

$$\lim_{n\to\infty} \frac{S_n}{n} = \frac{1}{\mathbf{E}_0[\tau_1]}, \qquad \mathbf{P}_0\text{-presque sûrement.}$$

Cas 2 : $\mathbf{E}_0[\tau_1] = +\infty$. En procédant comme ci-dessus, on obtient

$$0 \leqslant \limsup_{n\to\infty} \frac{S_n}{n} \leqslant \limsup_{n\to\infty} \frac{M_n}{n} = \limsup_{n\to\infty} \frac{M_n}{T_{M_n}} \frac{T_{M_n}}{n} \leqslant \lim_{n\to\infty} \frac{M_n}{T_{M_n}} = 0,$$

$\mathbf{P}_0$-presque sûrement. Si la marche est transiente, $\liminf_{n\to\infty} S_n/n \geqslant 0$, $\mathbf{P}_0$-p.s., d'où l'on déduit que $\lim_{n\to\infty} \frac{1}{n} S_n = 0$, $\mathbf{P}_0$-p.s.. Si la marche est récurrente, $\liminf_{n\to\infty} S_n = -\infty$, $\mathbf{P}_0$-p.s., et on peut donc conclure comme ci-dessus (en répétant l'argument avec le processus $(-S_n)$) que $\liminf_{n\to\infty} \frac{1}{n} S_n = 0$, $\mathbf{P}_0$-p.s., ce qui conduit à nouveau au résultat désiré.

Il ne reste plus qu'à démontrer les Lemmes 16.5 et 16.6.

Preuve du Lemme 16.5. Soient $A_1, \dots, A_n \subset \mathbb{N}$. On a, par simple changement d'origine,

$$\mathsf{P}_0^\omega(\tau_1 \in A_1, \dots, \tau_n \in A_n) = \mathsf{P}_1^{\theta^{-1}\omega}(\tau_2 \in A_1, \dots, \tau_{n+1} \in A_n),$$

où θ est l'opérateur de translation défini par $(\theta(\omega))_k \coloneqq \omega_{k+1}$. D'un autre côté, il suit de l'hypothèse que $\mathbf{P}_0(T_1 < \infty) = 1$, et donc

$$\begin{aligned}\mathsf{P}_0^{\theta^{-1}\omega}(\tau_2 \in A_1, \dots, \tau_{n+1} \in A_n) &= \sum_{k=1}^{\infty} \mathsf{P}_0^{\theta^{-1}\omega}(\tau_2 \in A_1, \dots, \tau_{n+1} \in A_n \,|\, T_1 = k)\, \mathsf{P}_0^{\theta^{-1}\omega}(T_1 = k) \\ &= \mathsf{P}_1^{\theta^{-1}\omega}(\tau_2 \in A_1, \dots, \tau_{n+1} \in A_n),\end{aligned}$$

la dernière identité suivant de la propriété de Markov forte.

On obtient ainsi

$$\mathsf{P}_0^\omega(\tau_1 \in A_1, \dots, \tau_n \in A_n) = \mathsf{P}_0^{\theta^{-1}\omega}(\tau_2 \in A_1, \dots, \tau_{n+1} \in A_n).$$

Étant donné que $\mathbb{P} \circ \theta^{-1} = \mathbb{P}$ (les ω_k sont i.i.d.. sous $\mathbb{P}$), la stationnarité suit facilement :

$$\begin{aligned}\mathbf{P}_0(\tau_1 \in A_1, \dots, \tau_n \in A_n) &= \mathbb{E}\big[\mathsf{P}_0^\omega(\tau_1 \in A_1, \dots, \tau_n \in A_n)\big] \\ &= \mathbb{E}\big[\mathsf{P}_0^{\theta^{-1}\omega}(\tau_2 \in A_1, \dots, \tau_{n+1} \in A_n)\big] \\ &= \mathbf{P}_0(\tau_2 \in A_1, \dots, \tau_{n+1} \in A_n).\end{aligned}$$

Afin de démontrer l'ergodicité, il va se révéler utile de considérer un nouvel espace de probabilité $\big(\Omega \times \Xi, \mathcal{F} \times \mathcal{H}, \mathbb{Q}\big)$, où $\Xi = [0,1]^{\mathbb{N}^*}$, $\mathcal{H}$ est la tribu borélienne sur Ξ, et $\mathbb{Q}$ la mesure-produit définie par $\mathbb{Q} \coloneqq \mathbb{P} \times \mathbb{U}$, avec $\mathbb{U} \coloneqq \mathsf{U}^{\times \mathbb{N}^*}$ et U la mesure uniforme sur $[0,1]$.

Étant donné $(\omega, \xi) \in \Omega \times \Xi$, on définit un processus $\widetilde{S}_n$ en posant $\widetilde{S}_0 = 0$ et

$$\widetilde{S}_{n+1} = \widetilde{S}_n - \mathbf{1}_{\{\xi_{n+1} > \omega_{\widetilde{S}_n}\}} + \mathbf{1}_{\{\xi_{n+1} \leqslant \omega_{\widetilde{S}_n}\}}.$$

Notons $\widetilde{\mathsf{P}}_0^\omega$ la loi de $\widetilde{S}_n$ conditionnellement à une réalisation $\omega \in \Omega$ du milieu. On a alors $\widetilde{\mathsf{P}}_0^\omega = \mathsf{P}_0^\omega$. En effet,

$$\begin{aligned}\widetilde{\mathsf{P}}_0^\omega(\widetilde{S}_{n+1} = i+1 \,|\, \widetilde{S}_n = i, \widetilde{S}_{n-1} = x_{n-1}, \dots, \widetilde{S}_1 = x_1) &= \mathsf{U}(\xi_{n+1} \leqslant \omega_i) \\ &= \omega_i = \mathsf{P}_0^\omega(S_{n+1} = i+1 \,|\, S_n = i).\end{aligned}$$

On identifiera donc dans le reste de cet argument les processus S_n (partant de 0) et $\widetilde{S}_n$[5].

Soit $A \in \sigma(\tau_k, k \geqslant 1)$, c'est-à-dire de la forme $A = \{(\omega, \xi) \,|\, (\tau_k)_{k \geqslant 1} \in B\}$ avec $B \subset \mathbb{N}^{\mathbb{N}}$. On définit $\hat{\theta}A \coloneqq \{(\omega, \xi) \,|\, (\tau_k)_{k \geqslant 2} \in B\}$. Nous allons montrer que si A est invariant, c'est-à-dire $A = \hat{\theta}A$, alors $\mathbb{Q}(A) \in \{0, 1\}$, ce qui établira l'ergodicité de la suite (τ_k).

Il suit du théorème de Fubini que

$$\mathbb{Q}(A) = \int \mathbf{1}_A(\omega, \xi)\, \mathrm{d}\mathbb{Q}(\omega, \xi) = \int \Big(\int \mathbf{1}_A(\omega, \xi)\, \mathrm{d}\mathbb{U}(\xi) \Big)\, \mathrm{d}\mathbb{P}(\omega).$$

On fixe $\omega \in \Omega$. Clairement, $\hat{\theta}^k A$, qui ne dépend que des τ_i avec $i \geqslant k+1$, est indépendante des k premiers pas de la marche (puisque $T_k \geqslant k$). Comme $A = \hat{\theta}^k A$, il s'ensuit que la fonction $\mathbf{1}_A(\omega, \cdot)$ est mesurable par rapport à la tribu $\sigma(\xi_i, i \geqslant k)$, et donc, k étant arbitraire, cette fonction est mesurable par rapport à la tribu asymptotique $\bigcap_{k \geqslant 1} \sigma(\xi_i, i \geqslant k)$. Cette dernière tribu étant triviale (les ξ_k sont indépendants), on en déduit que $\mathbf{1}_A(\omega, \cdot)$ est $\mathbb{U}$-presque sûrement constante : il existe $\tilde{A} \in \mathcal{F}$ telle que

$$\mathbf{1}_A(\omega, \xi) = \mathbf{1}_{\tilde{A}}(\omega), \qquad \mathbb{U}\text{-presque sûrement.}$$

5. Observez que si cette construction peut sembler un peu abstraite au premier abord, elle correspond probablement à l'approche que vous utiliseriez afin de simuler S_n sur un ordinateur !

On a donc

$$\mathbb{Q}(A) = \int \mathbf{1}_{\tilde{A}}(\omega)\mathrm{d}\mathbb{P}(\omega) = \mathbb{P}(\tilde{A}).$$

De plus,

$$\mathbf{1}_{\tilde{A}}(\theta\omega) = \mathbf{1}_A(\theta\omega,\cdot) = \mathbf{1}_{\hat{\theta}A}(\omega,\cdot) = \mathbf{1}_A(\omega,\cdot) = \mathbf{1}_{\tilde{A}}(\omega),$$

ce qui montre que $\tilde{A}$ appartient à la sous-tribu $\mathcal{I} \subset \mathcal{F}$ des événements invariants sous les translations. Par définition, la tribu $\mathcal{F}$ est engendrée par les cylindres. On peut donc trouver une suite d'événements $(C_n)_{n\geqslant 1}$ et une suite d'entiers $(N_n)_{n\geqslant 1}$ telles que $N_n \uparrow \infty$, $C_n \in \sigma(\omega_k, |k| \leqslant N_n)$ et $\mathbb{P}(\tilde{A} \vartriangle C_n) \leqslant 2^{-n}$ pour tout $n \geqslant 1$. Pour chaque $n \geqslant 1$, notons $C'_n \coloneqq \theta^{2N_n} C_n$. On a alors,

$$\mathbb{P}(\tilde{A} \vartriangle C'_n) = \mathbb{P}(\theta^{2N_n}(\theta^{-2N_n}\tilde{A} \vartriangle C_n)) = \mathbb{P}(\theta^{2N_n}(\tilde{A} \vartriangle C_n)) = \mathbb{P}(\tilde{A} \vartriangle C_n) \leqslant 2^{-n},$$

où l'on a utilisé le fait que $\tilde{A} \in \mathcal{I}$ pour la deuxième égalité et l'invariance sous translation de $\mathbb{P}$ pour la troisième. Observez que, par construction, $B = \bigcap_n \bigcup_{m\geqslant n} C'_m \in \mathcal{T}$, où $\mathcal{T}$ dénote la tribu asymptotique $\mathcal{T} \coloneqq \bigcap_{n\geqslant 1} \sigma(\omega_k, |k| \geqslant n)$. On a donc

$$\mathbb{P}(\tilde{A} \vartriangle B) \leqslant \mathbb{P}\Big(\bigcap_n \cup_{m\geqslant n}(\tilde{A} \vartriangle C'_m)\Big) \leqslant \lim_{n\to\infty} \sum_{m\geqslant n} \mathbb{P}(\tilde{A} \vartriangle C'_m) = 0.$$

La tribu $\mathcal{T}$ étant triviale (par indépendance des ω_i), $\mathbb{P}(B) \in \{0,1\}$, ce qui implique que $\mathbb{P}(\tilde{A}) \in \{0,1\}$ et donc que $\mathbb{Q}(A) \in \{0,1\}$. □

Preuve du Lemme 16.6. La preuve repose sur la décomposition suivante :

$$\tau_1 = \mathbf{1}_{\{S_1=1\}} + \mathbf{1}_{\{S_1=-1\}}(1 + \tau'_0 + \tau'_1) = 1 + \mathbf{1}_{\{S_1=-1\}}(\tau'_0 + \tau'_1),$$

où l'on a introduit

$$\tau'_0 \coloneqq \inf\big\{n > 0 \,\big|\, S_{T_{-1}+n} = 0\big\}, \qquad \tau'_1 \coloneqq \inf\big\{k > 0 \,\big|\, S_{\tau'_0+k} = 1\big\},$$

le temps de premier retour en 0 après avoir visité -1, et le temps écoulé entre le premier retour en 0 et la première visite ultérieure en 1. Fixons à présent un milieu $\omega \in \Omega$. On a alors, en observant que $\mathsf{P}_0^\omega(S_1 = -1) = (1 - \omega_0)$,

$$\mathsf{E}_0^\omega(\tau_1) = 1 + (1-\omega_0)\big\{\mathsf{E}_0^\omega(\tau'_0 \,|\, S_1 = -1) + \mathsf{E}_0^\omega(\tau'_1 \,|\, S_1 = -1)\big\}.$$

À présent, il suit de la propriété de Markov forte que, pour tout $k \in \mathbb{N}$,

$$\begin{aligned}\mathsf{P}_0^\omega(\tau'_0 = k \,|\, S_1 = -1) &= \mathsf{P}_0^{\theta^{-1}\omega}(\tau_1 = k),\\ \mathsf{P}_0^\omega(\tau'_1 = k \,|\, S_1 = -1, \tau'_0 < \infty) &= \mathsf{P}_0^\omega(\tau_1 = k).\end{aligned}$$

Supposons tout d'abord que $\mathbf{E}_0[\tau_1] < \infty$. On a alors, $\mathbb{P}$-presque sûrement, $\mathsf{E}_0^\omega(\tau_1) < \infty$, et donc $\mathsf{P}_0^\omega(\tau'_0 < \infty \,|\, S_1 = -1) = \mathsf{P}_0^{\theta^{-1}\omega}(\tau_1 < \infty) = 1$, $\mathbb{P}$-presque sûrement. On obtient donc

$$\mathsf{E}_0^\omega(\tau_1) = 1 + (1-\omega_0)\big\{\mathsf{E}_0^{\theta^{-1}\omega}(\tau_1) + \mathsf{E}_0^\omega(\tau_1)\big\},$$

et, après réarrangement,

$$\mathsf{E}_0^\omega(\tau_1) = \frac{1}{\omega_0} + \rho_0 \mathsf{E}_0^{\theta^{-1}\omega}(\tau_1).$$

L'observation suivante est que $\mathsf{E}_0^{\theta^{-1}\omega}(\tau_1)$ est mesurable par rapport à $\sigma(\omega_k, k < 0)$ et est donc indépendante (sous $\mathbb{P}$) de ρ_0. On a donc, en prenant l'espérance par rapport à $\mathbb{P}$,

$$\mathbf{E}_0[\tau_1] = \mathbb{E}[\frac{1}{\omega_0}] + \mathbb{E}[\rho_0]\mathbf{E}_0[\tau_1],$$

ce qui donne bien, après réarrangement,

$$\mathbf{E}_0[\tau_1] = \frac{\mathbb{E}[1/\omega_0]}{1-\mathbb{E}[\rho_0]} = \frac{1+\mathbb{E}[\rho_0]}{1-\mathbb{E}[\rho_0]}. \tag{16.3}$$

En particulier, l'implication $\mathbf{E}_0[\tau_1] < \infty \implies \mathbb{E}[\rho_0] < 1$ suit de (16.3). La preuve sera donc achevée si l'on parvient à montrer que $\mathbb{E}[\rho_0] < 1 \implies \mathbf{E}_0[\tau_1] < \infty$.

Supposons donc que $\mathbb{E}[\rho_0] < 1$. Pour $M > 0$, on obtient, en procédant comme ci-dessus,

$$\mathsf{E}_0^\omega(\tau_1 \mathbf{1}_{\{\tau_1<M\}}) = \mathsf{P}_0^\omega(\tau_1 < M) + (1-\omega_0)\mathsf{E}_0^\omega\big((\tau_0'+\tau_1')\mathbf{1}_{\{\tau_1<M\}} \;\big|\; S_1 = -1\big)$$

Lorsque $S_1 = -1$, on a $\tau_1 = 1 + \tau_0' + \tau_1'$; par conséquent,

$$\mathsf{E}_0^\omega(\tau_0' \mathbf{1}_{\{\tau_1<M\}} \,|\, S_1 = -1) \leqslant \mathsf{E}_0^\omega(\tau_0' \mathbf{1}_{\{\tau_0'<M\}} \,|\, S_1 = -1) = \mathsf{E}_0^{\theta^{-1}\omega}(\tau_1 \mathbf{1}_{\{\tau_1<M\}}),$$

et

$$\begin{aligned}\mathsf{E}_0^\omega(\tau_1' \mathbf{1}_{\{\tau_1<M\}} \,|\, S_1 = -1) &\leqslant \mathsf{E}_0^\omega(\tau_1' \mathbf{1}_{\{\tau_0'<M, \tau_1'<M\}} \,|\, S_1 = -1) \\ &\leqslant \mathsf{E}_0^\omega(\tau_1' \mathbf{1}_{\{\tau_1'<M\}} \,|\, S_1 = -1, \tau_0' < \infty) = \mathsf{E}_0^\omega(\tau_1 \mathbf{1}_{\{\tau_1<M\}}).\end{aligned}$$

On obtient donc

$$\mathsf{E}_0^\omega(\tau_1 \mathbf{1}_{\{\tau_1<M\}}) \leqslant 1 + (1-\omega_0)\big\{\mathsf{E}_0^{\theta^{-1}\omega}(\tau_1 \mathbf{1}_{\{\tau_1<M\}}) + \mathsf{E}_0^\omega(\tau_1 \mathbf{1}_{\{\tau_1<M\}})\big\},$$

que l'on peut réécrire

$$\mathsf{E}_0^\omega(\tau_1 \mathbf{1}_{\{\tau_1<M\}}) \leqslant \frac{1}{\omega_0} + \rho_0 \mathsf{E}_0^{\theta^{-1}\omega}(\tau_1 \mathbf{1}_{\{\tau_1<M\}}).$$

On peut à présent en déduire comme précédemment que

$$\mathbf{E}_0[\tau_1 \mathbf{1}_{\{\tau_1<M\}}] \leqslant \frac{1+\mathbb{E}[\rho_0]}{1-\mathbb{E}[\rho_0]} < \infty.$$

Il suffit alors de laisser $M \uparrow \infty$ pour conclure, par convergence monotone, que

$$\mathbf{E}_0[\tau_1] \leqslant \frac{1+\mathbb{E}[\rho_0]}{1-\mathbb{E}[\rho_0]} < \infty. \qquad \square$$

Remarques bibliographiques : Ce chapitre est inspiré de la présentation dans [54, Section 2]. Les Théorèmes 16.2 et 16.3 sont originellement dûs à Solomon [48]

17 Adsorption d'un polymère

17.1 Le modèle

On notera $\mathbb{P}$ la loi de la marche aléatoire simple symétrique $(S_k)_{k\geqslant 0}$ sur $\mathbb{Z}^d$ partant de 0.[1] Soit $\lambda \in \mathbb{R}$ et $\mathcal{L}_N := \sum_{i=1}^N \mathbf{1}_{\{S_i=0\}}$ le nombre de retours en 0 jusqu'au temps N. Nous allons nous intéresser à la mesure de probabilité suivante :

$$\mathbb{P}_N^\lambda := \frac{e^{\lambda \mathcal{L}_N}}{\mathbf{Z}_N^\lambda}\,\mathbb{P}.$$

La constante de normalisation

$$\mathbf{Z}_N^\lambda := \mathbb{E}[\exp(\lambda \mathcal{L}_N)]$$

est appelée **fonction de partition**.

Dans ce chapitre, on interprétera la trajectoire finie $(k, S_k)_{0\leqslant k\leqslant N}$ comme décrivant un polymère, chaque sommet visité par la marche correspondant à un monomère. Lorsque $\lambda > 0$, la mesure de probabilité $\mathbb{P}_N^\lambda$ favorise les trajectoires $(k, S_k)_{0\leqslant k\leqslant N}$ visitant souvent 0 ; elle les défavorise si $\lambda < 0$. Ceci fournit l'un des modèles les plus simples décrivant l'interaction entre un polymère et un potentiel spatialement localisé, ici placé le long de la ligne $\{(n, x) \in \mathbb{N} \times \mathbb{Z}^d \mid x = 0\}$. Les monomères se trouvant sur la ligne sont dits **adsorbés**.

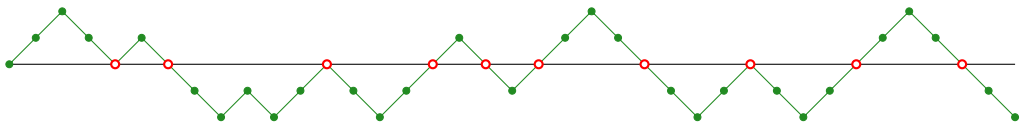


FIGURE 17.1 : Les N premiers pas d'une trajectoire sous $\mathbb{P}_N^\lambda$, lorsque $d = 1$. Les monomères affectés par le potentiel sont marqués (intérieur blanc).

Remarque 17.1. *Observez que le monomère en* $(0, 0)$*, bien qu'adsorbé, n'entre pas dans le calcul de l'effet du potentiel. Cela est fait pour des raisons de commodité. Notez cependant que ceci n'a aucun effet sur la mesure* $\mathbb{P}_N^\lambda$ *puisque remplacer* $\sum_{i=1}^N \mathbf{1}_{\{S_i=0\}}$ *par* $\sum_{i=0}^N \mathbf{1}_{\{S_i=0\}}$ *affecterait de la même façon le numérateur et le dénominateur (la fonction de partition) dans la définition de cette mesure.* ◇

Le problème qui nous intéresse dans ce chapitre est de déterminer l'effet du potentiel sur le comportement à grande échelle du polymère. En particulier, nous serons intéressés par l'effet du potentiel sur la fraction moyenne $\rho_N(\lambda) := \mathbb{E}_N^\lambda[\mathcal{L}_N/N]$ de monomères adsorbés.

1. Les méthodes décrites dans ce chapitre s'étendent sans difficulté à toute marche aléatoire sur $\mathbb{Z}^d$ à laquelle le théorème limite local s'applique.

Notons $\mathsf{p}(k) = \mathbb{P}(S_k = 0)$. Il suit (par exemple) du théorème limite local (voir Section 2.5) qu'il existe deux constantes $c(d), c'(d) > 0$ telles que

$$\frac{c(d)}{k^{d/2}} \leqslant \mathsf{p}(2k) \leqslant \frac{c'(d)}{k^{d/2}}, \qquad \text{pour tout } k \geqslant 1. \tag{17.1}$$

Lorsque $\lambda = 0$, on a donc

$$\rho_N(0) = \mathbb{E}_N^0[\mathcal{L}_N/N] = \frac{1}{N}\sum_{k=1}^{\lfloor N/2 \rfloor} \mathsf{p}(2k),$$

ce qui tend vers 0 lorsque $N \to \infty$, pour toute dimension d.

Afin d'étudier le comportement de $\rho_N(\lambda)$ lorsque $\lambda \neq 0$, il va être utile d'introduire un nouveau concept issu de la physique statistique : l'énergie libre.

17.2 L'énergie libre

Introduisons $f_N(\lambda) := \frac{1}{N}\log \mathbf{Z}_N^\lambda$. Observez que $\log \mathbf{Z}_N^\lambda = \log \mathbb{E}[e^{\lambda \mathcal{L}_N}]$ est essentiellement[2] la fonction génératrice des cumulants de $\mathcal{L}_N$ (sous $\mathbb{P}_N^\lambda$). En particulier,

$$\rho_N(\lambda) = \mathbb{E}_N^\lambda[\mathcal{L}_N/N] = \frac{\mathrm{d}}{\mathrm{d}\lambda} f_N(\lambda). \tag{17.2}$$

En outre, puisque

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\rho_N(\lambda) = \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2} f_N(\lambda) = \frac{1}{N}\mathrm{Var}_N^\lambda[\mathcal{L}_N] \geqslant 0, \tag{17.3}$$

ρ_N est une fonction croissante de λ et f_N est convexe.

Un rôle crucial sera joué par la limite des fonctions f_N lorsque $N \to \infty$.

Théorème 17.2. *L'**énergie libre** $f(\lambda) := \lim_{N\to\infty} f_N(\lambda)$ existe, pour tout $\lambda \in \mathbb{R}$, et est croissante et convexe. De plus, $f(\lambda) = 0$ pour tout $\lambda \leqslant 0$, et $\frac{1}{2}\lambda \geqslant f(\lambda) \geqslant \frac{1}{2}(\lambda - \log(2d))$ pour tout $\lambda > 0$.*

Démonstration. D'une part, observons que f_N est croissante par (17.2) et que $f_N(0) = 0$ pour tout N. On a donc $f_N(\lambda) \leqslant 0$ pour tout $\lambda \leqslant 0$, ce qui implique que

$$\limsup_{N\to\infty} f_N(\lambda) \leqslant 0, \qquad \text{pour tout } \lambda \leqslant 0.$$

D'autre part, il suit de l'inégalité de Jensen que

$$\mathbf{Z}_N^\lambda = \mathbb{E}[e^{\lambda \mathcal{L}_N}] \geqslant e^{\lambda \mathbb{E}[\mathcal{L}_N]}.$$

Or, $\mathbb{E}[\mathcal{L}_N] = \sum_{k=1}^{\lfloor N/2 \rfloor} \mathsf{p}(2k) = \mathsf{o}(N)$, par (17.1). Par conséquent,

$$\liminf_{N\to\infty} f_N(\lambda) \geqslant 0, \qquad \text{pour tout } \lambda \in \mathbb{R}.$$

En particulier, cela démontre que $\lim_{N\to\infty} f_N(\lambda) = 0$ pour tout $\lambda \leqslant 0$. Nous supposerons donc, dans le reste de la preuve, que $\lambda > 0$.

Remarquons tout d'abord que $\mathbf{Z}_{2N+1}^\lambda = \mathbf{Z}_{2N}^\lambda$, puisque S_{2N+1} est toujours différent de 0. Il suffit donc de montrer la convergence de la suite $(\frac{1}{2N}\log \mathbf{Z}_{2N}^\lambda)_{N\geqslant 1}$.

2. La fonction génératrice des cumulants de $\mathcal{L}_N$ sous $\mathbb{P}_N^\lambda$ est donnée par la fonction $t \mapsto \log \mathbb{E}_N^\lambda[e^{t\mathcal{L}_N}] = \log(\mathbf{Z}_N^{\lambda+t}/\mathbf{Z}_N^\lambda)$. En particulier, le cumulant d'ordre k de $\mathcal{L}_N$ sous $\mathbb{P}_N^\lambda$ est donné par $\frac{\mathrm{d}^k}{\mathrm{d}t^k}\log \mathbb{E}_N^\lambda[e^{t\mathcal{L}_N}]\big|_{t=0} = \frac{\mathrm{d}^k}{\mathrm{d}\lambda^k}\log \mathbf{Z}_N^\lambda$.

Observons à présent que $e^{\lambda \mathbf{1}_{\{S_{2i}=0\}}} = (e^\lambda - 1)\mathbf{1}_{\{S_{2i}=0\}} + 1$. On peut donc écrire

$$e^{\lambda \mathcal{L}_{2N}} = \prod_{i=1}^{N} \{(e^\lambda - 1)\mathbf{1}_{\{S_{2i}=0\}} + 1\} = \sum_{A \subset \{1,\dots,N\}} (e^\lambda - 1)^{|A|} \mathbf{1}_{\{S_{2i}=0 \ \forall i \in A\}}$$
$$= 1 + \sum_{L=1}^{N} \sum_{\substack{A \subset \{1,\dots,L\} \\ A \ni L}} (e^\lambda - 1)^{|A|} \mathbf{1}_{\{S_{2i}=0 \ \forall i \in A\}}. \tag{17.4}$$

Par conséquent,

$$\mathbf{Z}_{2N}^\lambda = 1 + \sum_{L=1}^{N} \sum_{\substack{A \subset \{1,\dots,L\} \\ A \ni L}} (e^\lambda - 1)^{|A|} \, \mathbb{P}(S_{2i} = 0 \ \forall i \in A). \tag{17.5}$$

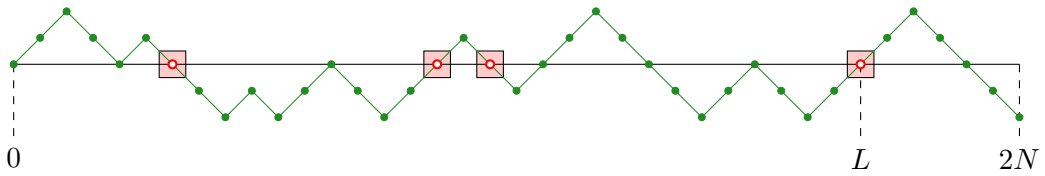


FIGURE 17.2: La décomposition de la fonction de partition $\mathbf{Z}_{2N}^\lambda$ en termes de l'ensemble A. Les points de ce dernier sont marqués.

De même, en utilisant à nouveau (17.4),

$$\hat{\mathbf{Z}}_{2N}^\lambda := \mathbb{E}(e^{\lambda \mathcal{L}_{2N}} \mathbf{1}_{\{S_{2N}=0\}}) = e^\lambda \mathbb{E}(e^{\lambda \mathcal{L}_{2N-2}} \mathbf{1}_{\{S_{2N}=0\}})$$
$$= e^\lambda \Big\{ \mathsf{p}(2N) + \sum_{L=1}^{N-1} \sum_{\substack{A \subset \{1,\dots,L\} \\ A \ni L}} (e^\lambda - 1)^{|A|} \, \mathbb{P}(S_{2i} = 0 \ \forall i \in A) \, \mathsf{p}(2N - 2L) \Big\}$$
$$\geqslant e^\lambda \, c(d) \, N^{-d/2} \, \mathbf{Z}_{2N-2}^\lambda \geqslant c(d) \, N^{-d/2} \, \mathbf{Z}_{2N}^\lambda,$$

où la première inégalité suit de la borne $\mathsf{p}(2N - 2L) \geqslant c(d) N^{-d/2}$ pour tout $0 \leqslant L \leqslant N$ (*cf.* (17.1)) et de (17.5), alors que la seconde suit de $\mathbf{Z}_{2N}^\lambda = \mathbb{E}(e^{\lambda \mathcal{L}_{2N-2}} e^{\lambda \mathbf{1}_{\{S_{2N}=0\}}}) \leqslant e^\lambda \mathbb{E}(e^{\lambda \mathcal{L}_{2N-2}}) = e^\lambda \mathbf{Z}_{2N-2}^\lambda$.

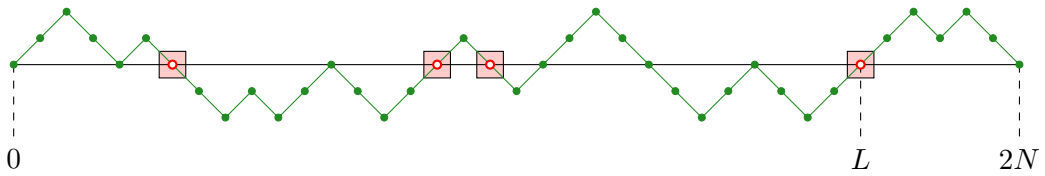


FIGURE 17.3: La décomposition de la fonction de partition $\hat{\mathbf{Z}}_{2N}^\lambda$ en termes de l'ensemble A. Les points de ce dernier sont marqués.

Évidemment, on a également la borne $\mathbf{Z}_{2N}^\lambda \geqslant \hat{\mathbf{Z}}_{2N}^\lambda$. On conclut donc que

$$\lim_{N \to \infty} \frac{1}{2N} \log(\mathbf{Z}_{2N}^\lambda / \hat{\mathbf{Z}}_{2N}^\lambda) = 0. \tag{17.6}$$

Or, pour tout $M, N \in \mathbb{N}^\star$, la propriété de Markov a pour conséquence que

$$\hat{\mathbf{Z}}_{2(N+M)}^\lambda = \mathbb{E}[e^{\lambda \mathcal{L}_{2(N+M)}} \mathbf{1}_{\{S_{2(N+M)}=0\}}] \geqslant \mathbb{E}[e^{\lambda \mathcal{L}_{2(N+M)}} \mathbf{1}_{\{S_{2N}=0\}} \mathbf{1}_{\{S_{2(N+M)}=0\}}]$$
$$= \mathbb{E}[e^{\lambda \mathcal{L}_{2N}} \mathbf{1}_{\{S_{2N}=0\}}] \, \mathbb{E}[e^{\lambda \mathcal{L}_{2M}} \mathbf{1}_{\{S_{2M}=0\}}] = \hat{\mathbf{Z}}_{2N}^\lambda \hat{\mathbf{Z}}_{2M}^\lambda.$$

Par conséquent, la suite $(-\log \hat{\mathbf{Z}}_{2N}^{\lambda})_{N\geqslant 1}$ est sous-additive et le Lemme 17.5 s'applique :

$$\lim_{N\to\infty} \frac{1}{2N} \log \hat{\mathbf{Z}}_{2N}^{\lambda} = \sup_{N\geqslant 1} \frac{1}{2N} \log \hat{\mathbf{Z}}_{2N}^{\lambda}.$$

Combiné avec (17.6), ceci établit l'existence de l'énergie libre f.

La croissance et la convexité de f découlent des propriétés correspondantes de f_N établies en (17.2) et (17.3). Le fait que $f(\lambda) = 0$ lorsque $\lambda \leqslant 0$ a été établi en début de preuve. Finalement, lorsque $\lambda > 0$, on a d'une part

$$\mathbf{Z}_{2N}^{\lambda} = \mathbb{E}[e^{\lambda \mathcal{L}_{2N}}] \leqslant e^{\lambda N},$$

et donc $f(\lambda) \leqslant \frac{1}{2}\lambda$, et d'autre part,

$$\mathbf{Z}_{2N}^{\lambda} \geqslant e^{\lambda N} \mathbb{P}(S_{2k} = 0 \ \forall k \in \{1, \dots, N\}) = e^{\lambda N} (2d)^{-N},$$

et donc $f(\lambda) \geqslant \frac{1}{2}\lambda - \frac{1}{2}\log(2d)$. □

17.3 La transition de phase

Nous allons à présent nous intéresser au comportement de l'énergie libre lorsque λ varie. On a déjà vu que $f(\lambda) = 0$ pour tout $\lambda \leqslant 0$. Ceci implique que f ne sera pas analytique au point $\lambda_{\mathrm{c}}(d) := \inf\{\lambda \geqslant 0 \,|\, f(\lambda) > 0\}$. Le Théorème 17.2 implique que $\lambda_{\mathrm{c}}(d) \leqslant \log(2d)$. Le théorème suivant détermine $\lambda_{\mathrm{c}}(d)$ explicitement et montre que λ_{c} est le seul point auquel l'énergie libre n'est pas analytique.

Théorème 17.3. *On a*

$$\lambda_{\mathrm{c}}(d) = |\log \mathbb{P}(\tau < \infty)|,$$

où $\tau = \inf\{n \geqslant 1 \,|\, S_n = 0\}$. *En particulier,*

$$\lambda_{\mathrm{c}}(1) = \lambda_{\mathrm{c}}(2) = 0 \quad \textit{et} \quad \lambda_{\mathrm{c}}(d) > 0 \textit{ pour tout } d \geqslant 3.$$

De plus, λ_{c} *est la seule valeur de* λ *à laquelle* f *n'est pas réelle-analytique.*

Démonstration. Au vu du Théorème 17.2, on peut supposer que $\lambda > 0$. De plus, comme on l'a vu dans la preuve de ce théorème, il suffit de travailler avec

$$\hat{\mathbf{Z}}_{2N}^{\lambda} = \mathbb{E}[e^{\lambda \mathcal{L}_{2N}} \mathbf{1}_{\{S_{2N}=0\}}].$$

En procédant comme dans la preuve de ce théorème, on obtient

$$\begin{aligned}
\hat{\mathbf{Z}}_{2N}^{\lambda} &= \frac{e^{\lambda}}{e^{\lambda}-1} \sum_{\substack{A\subset\{1,\dots,N\}\\ A\ni N}} (e^{\lambda}-1)^{|A|}\, \mathbb{P}(S_{2i} = 0 \ \forall i \in A) \\
&= \frac{e^{\lambda}}{e^{\lambda}-1} \sum_{n=1}^{N} \sum_{\substack{\ell_1+\cdots+\ell_n=N\\ \ell_i\geqslant 1\ \forall i}} \prod_{i=1}^{n} (e^{\lambda}-1)\mathsf{p}(2\ell_i) \\
&= e^{\lambda}\mathsf{p}(2N) + \sum_{\ell=1}^{N-1} (e^{\lambda}-1)\mathsf{p}(2\ell) \cdot \frac{e^{\lambda}}{e^{\lambda}-1} \sum_{n=1}^{N-\ell} \sum_{\substack{\ell_1+\cdots+\ell_n=N-\ell\\ \ell_i\geqslant 1\ \forall i}} \prod_{i=1}^{n} (e^{\lambda}-1)\mathsf{p}(2\ell_i) \\
&= \sum_{\ell=1}^{N} (e^{\lambda}-1)\mathsf{p}(2\ell)\, \hat{\mathbf{Z}}_{2N-2\ell}^{\lambda}, \qquad (17.7)
\end{aligned}$$

où l'on a pris pour convention $\hat{\mathbf{Z}}_0^\lambda := e^\lambda/(e^\lambda - 1)$.

Introduisons les fonctions génératrices

$$\mathbb{A}_\lambda(s) := \sum_{N\geqslant 0} \hat{\mathbf{Z}}_{2N}^\lambda s^{2N} \quad \text{et} \quad \mathbb{B}_\lambda(s) := \sum_{\ell\geqslant 1} (e^\lambda - 1)\mathsf{p}(2\ell)s^{2\ell}$$

et notons $r_{\mathbb{A}_\lambda}$ et $r_{\mathbb{B}_\lambda}$ leurs rayons de convergence. Observez que (17.1) implique que $r_{\mathbb{B}_\lambda} = 1$, alors que la preuve du Théorème 17.2 montre que $\hat{\mathbf{Z}}_{2N}^\lambda = e^{f(\lambda)2N + o(N)}$ et donc que $r_{\mathbb{A}_\lambda} = e^{-f(\lambda)} \in (0, 1]$. En particulier, $r_{\mathbb{A}_\lambda} \leqslant r_{\mathbb{B}_\lambda}$.

La relation (17.7) peut alors se reformuler en termes des fonctions génératrices :

$$\mathbb{A}_\lambda(s)(1 - \mathbb{B}_\lambda(s)) = \frac{e^\lambda}{e^\lambda - 1}, \tag{17.8}$$

pour tout $s \in \big\{z \in \mathbb{C} \,\big|\, |z| < r_{\mathbb{A}_\lambda}\big\}$.

Notons $s_* := \sup\big\{s > 0 \,\big|\, \mathbb{B}_\lambda(s) < 1\big\}$. On a évidemment $0 < s_* \leqslant r_{\mathbb{B}_\lambda}$. De plus, comme $|\mathbb{B}_\lambda(s)| \leqslant \mathbb{B}_\lambda(|s|)$, on a nécessairement $\mathbb{B}_\lambda(s) \neq 1$ pour tout $s \in \big\{z \in \mathbb{C} \,\big|\, |z| < s_*\big\}$. Observons également que $\mathbb{B}_\lambda(s_*) = 1$ lorsque $s_* < r_{\mathbb{B}_\lambda}$.

Le membre de droite de (17.8) ne s'annulant jamais, on a nécessairement $s_* \geqslant r_{\mathbb{A}_\lambda}$. On peut donc réécrire (17.8) sous la forme

$$\mathbb{A}_\lambda(s) = \frac{e^\lambda}{e^\lambda - 1} \cdot \frac{1}{1 - \mathbb{B}_\lambda(s)} \tag{17.9}$$

pour tout $s \in \big\{z \in \mathbb{C} \,\big|\, |z| < r_{\mathbb{A}_\lambda}\big\}$. Le membre de gauche de (17.9) définit une fonction analytique dans le disque de rayon $r_{\mathbb{A}_\lambda}$ divergeant hors de ce disque. Le membre de droite définit une fonction analytique à l'intérieur du disque de rayon s_*. On en conclut que $s_* = r_{\mathbb{A}_\lambda}$.

Nous allons à présent déterminer les valeurs de λ pour lesquelles $f(\lambda) > 0$. Observez que ceci est équivalent à $r_{\mathbb{A}_\lambda} < 1$. La discussion précédente montre que $r_{\mathbb{A}_\lambda} < 1$ si et seulement si il existe $s_* \in (0, 1)$ tel que $\mathbb{B}_\lambda(s_*) = 1$. La fonction $s \mapsto \mathbb{B}_\lambda(s)$ étant strictement croissante sur $[0, 1)$ et satisfaisant $\mathbb{B}_\lambda(0) = 0$, l'existence d'un tel s_* est équivalente[3] à $\mathbb{B}_\lambda(1) > 1$. Or, $\mathbb{B}_\lambda(1) = (e^\lambda - 1)\sum_{\ell=1}^\infty \mathsf{p}(2\ell) = (e^\lambda - 1)\mathbb{E}(\mathcal{L}_\infty)$, où $\mathcal{L}_\infty := \sum_{n=1}^\infty \mathbf{1}_{\{S_{2n}=0\}}$ est le nombre total de retours en 0. On vérifie aisément (voir le cours d'introduction à la théorie des probabilités) que $\mathbb{E}(\mathcal{L}_\infty) = r/(1 - r)$ avec $r = \mathbb{P}(\tau < \infty)$. On a donc finalement les équivalences suivantes :

$$f(\lambda) > 0 \quad \Leftrightarrow \quad r_{\mathbb{A}_\lambda} < 1 \quad \Leftrightarrow \quad \mathbb{B}_\lambda(1) > 1 \quad \Leftrightarrow \quad \lambda > |\log r|,$$

ce qui montre que $\lambda_c = |\log \mathbb{P}(\tau < \infty)|$.

Supposons à présent que $\lambda > \lambda_c$. Dans ce cas, $r_{\mathbb{A}_\lambda} < 1$ et $\mathbb{B}_\lambda(r_{\mathbb{A}_\lambda}) = 1$. En d'autres termes, la fonction

$$\mathbb{W}(\lambda, r) := (e^\lambda - 1)\sum_{\ell\geqslant 1} \mathsf{p}(2\ell)r^{2\ell}$$

satisfait

$$\mathbb{W}(\lambda, r_{\mathbb{A}_\lambda}) = 1 \quad \text{pour tout } \lambda > \lambda_c. \tag{17.10}$$

Observez que $\mathbb{W}$ est analytique sur $\mathbb{C} \times \big\{r \in \mathbb{C} \,\big|\, |r| < 1\big\}$. De plus, $\partial\mathbb{W}/\partial r$ étant strictement positif en $(\lambda, r_{\mathbb{A}_\lambda})$, la version analytique du théorème des fonctions implicites implique l'existence d'une fonction *analytique* $\lambda' \mapsto r(\lambda')$, définie sur un voisinage V de λ, telle que

$$\mathbb{W}(\lambda', r(\lambda')) = 1 \quad \text{pour tout } \lambda' \in V,$$

c'est-à-dire $\mathbb{B}_{\lambda'}(r(\lambda')) = 1$. Cela montre que $r_{\mathbb{A}_\lambda} = r(\lambda)$ dépend analytiquement de λ lorsque $\lambda > \lambda_c$. La même chose est donc vraie de $f(\lambda) = -\log r_{\mathbb{A}_\lambda}$. □

3. Notons que, par le théorème de convergence monotone, $\mathbb{B}_\lambda(1) = \lim_{s\uparrow 1} \mathbb{B}_\lambda(s)$ (dans $\mathbb{R} \cup \{+\infty\}$).

17.4 Conséquences pour la fraction de monomères adsorbés

On peut à présent extraire les conséquences sur la densité asymptotique de monomères adsorbés.

Corollaire 17.4. *La limite* $\rho(\lambda) \coloneqq \lim_{N\to\infty} \rho_N(\lambda)$ *existe pour tout* $\lambda \neq \lambda_c$. *De plus,*

$$\rho(\lambda) = 0 \quad \forall \lambda < \lambda_c \qquad \textit{et} \qquad \rho(\lambda) > 0 \quad \forall \lambda > \lambda_c.$$

Démonstration. Par le Théorème 17.3, l'énergie libre est différentiable en tout point $\lambda \neq \lambda_c$. Il suit donc de (17.2) et du Lemme 17.6 que

$$\lim_{N\to\infty} \rho_N(\lambda) = \lim_{N\to\infty} \frac{\mathrm{d}}{\mathrm{d}\lambda} f_N(\lambda) = \frac{\mathrm{d}}{\mathrm{d}\lambda} f(\lambda).$$

Pour $\lambda < \lambda_c$, l'affirmation est triviale puisque $f \equiv 0$ sur cet intervalle. Pour $\lambda > \lambda_c$, on peut utiliser la stricte positivité et la convexité de f, puisque celles-ci impliquent que $\frac{\mathrm{d}}{\mathrm{d}\lambda} f(\lambda) \geqslant f(\lambda)/(\lambda - \lambda_c) > 0$. □

On peut en fait, avec un peu plus de travail, déterminer le comportement de f et ρ au voisinage de λ_c à l'aide de la relation (17.10). Nous nous contenterons d'illustrer ceci lorsque $d = 1$. Dans ce cas, il suit du théorème limite local (ou de la formule de Stirling) que $|\mathsf{p}(2\ell) - (\pi\ell)^{-1/2}| \leqslant c/\ell$ pour tout $\ell \geqslant 1$. On a donc, lorsque $\epsilon \downarrow 0$,

$$\begin{aligned}
\sum_{\ell\geqslant 1} \mathsf{p}(2\ell) e^{-\epsilon\ell} &= \frac{1}{\sqrt{\pi}} \sum_{\ell\geqslant 1} \ell^{-1/2} e^{-\epsilon\ell} + \mathsf{O}(|\log\epsilon|) \\
&= \frac{1}{\sqrt{\pi}} \epsilon^{-1/2} \sum_{\ell\geqslant 1} (\epsilon\ell)^{-1/2} e^{-\epsilon\ell} \epsilon + \mathsf{O}(|\log\epsilon|) \\
&= \frac{1+\mathsf{o}(1)}{\sqrt{\pi}} \epsilon^{-1/2} \int_0^\infty x^{-1/2} e^{-x} \mathrm{d}x + \mathsf{O}(|\log\epsilon|) \\
&= (1+\mathsf{o}(1))\, \epsilon^{-1/2}.
\end{aligned}$$

Par l'identité (17.10),

$$\sum_{\ell\geqslant 1} \mathsf{p}(2\ell) e^{-f(\lambda)2\ell} = (e^\lambda - 1)^{-1}.$$

Par conséquent, lorsque $\lambda \downarrow 0$ (et donc $f(\lambda) \downarrow 0$),

$$(1+\mathsf{o}(1))\,(2f(\lambda))^{-1/2} = (e^\lambda - 1)^{-1}$$

et donc

$$f(\lambda) = \frac{\lambda^2}{2}(1+\mathsf{o}(1)). \tag{17.11}$$

En particulier, dans le cas $d = 1$, on conclut que f est $\mathscr{C}^1$, mais pas $\mathscr{C}^2$ en $\lambda_c = 0$. On dit que la transition de phase en λ_c est *du deuxième ordre*.

En effet, d'une part, si f était $\mathscr{C}^2$ en 0, alors il suivrait du théorème de Taylor–Young que $f(\lambda) = \mathsf{o}(\lambda^2)$ lorsque $\lambda \downarrow 0$, ce qui contredirait (17.11). D'autre part, f étant convexe et f' étant bien définie partout, sauf éventuellement en 0, $\rho_+(0) \coloneqq \lim_{\lambda\downarrow 0} f'(\lambda) > 0$ impliquerait $f'(\lambda) \geqslant \rho_+(0)\lambda$, ce qui contredirait à nouveau (17.11).

À titre purement informatif, voici les résultats auxquels conduit une analyse similaire en dimensions supérieures (en omettant les constantes explicites pour simplifier) : lorsque $\lambda \downarrow \lambda_c$,

$$\begin{aligned}
d = 2 &: \quad f(\lambda) = \exp\bigl[-\mathsf{O}\bigl((\lambda-\lambda_c)^{-1}\bigr)\bigr], \\
d = 3 &: \quad f(\lambda) = \mathsf{O}\bigl((\lambda-\lambda_c)^2\bigr), \\
d = 4 &: \quad f(\lambda) = \mathsf{O}\bigl((\lambda-\lambda_c)/|\log(\lambda-\lambda_c)|\bigr), \\
d \geqslant 5 &: \quad f(\lambda) = \mathsf{O}(\lambda-\lambda_c).
\end{aligned}$$

17.5 Compléments techniques.

Lemme sous-additif. Une suite $(a_n)_{n\geqslant 1} \subset \mathbb{R}$ est **sous-additive** si

$$a_{n+m} \leqslant a_n + a_m \quad \forall m, n\,.$$

Lemme 17.5 (Lemme sous-additif, Lemme de Fekete)**.** *Soit a_n une suite sous-additive. Alors,*

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf_n \frac{a_n}{n}\,.$$

Démonstration. Soit $\alpha := \inf_n a_n/n$ et $\alpha' > \alpha$. Soit ℓ tel que $a_\ell/\ell \leqslant \alpha'$.

On peut écrire tout $n \geqslant 1$ sous la forme $n = k\ell + j$ avec $k \in \mathbb{N}$ et $0 \leqslant j < \ell$. Il suit donc de la définition de α et de la sous-additivité de $(a_n)_{n\geqslant 1}$ que

$$\alpha n \leqslant a_n = a_{k\ell+j} \leqslant k a_\ell + a_j\,.$$

On en déduit que

$$\alpha \leqslant \liminf_{n\to\infty} \frac{a_n}{n} \leqslant \limsup_{n\to\infty} \frac{a_n}{n} \leqslant \frac{a_\ell}{\ell} \leqslant \alpha'\,.$$

Le résultat suit en prenant la limite $\alpha' \downarrow \alpha$. □

Limite de fonctions convexes.

Lemme 17.6. *Soit $(g_n)_{n\geqslant 1}$ une suite de fonctions convexes différentiables sur $\mathbb{R}$ convergeant ponctuellement vers une fonction g. Alors, si g est différentiable en x_0,*

$$\frac{\mathrm{d}}{\mathrm{d}x} g(x) = \lim_{n\to\infty} \frac{\mathrm{d}}{\mathrm{d}x} g_n(x).$$

Démonstration. Voir, par exemple, [22, Appendix B.2]. □

18 Le modèle de monomères-dimères

Dans ce chapitre, nous allons introduire le modèle de monomères-dimères et étudier certaines de ses propriétés. La motivation principale est de montrer comment des techniques de couplage (ici, la percolation de désaccord[1]) peuvent être utilisées pour étudier de tels champs aléatoires.

Le modèle de monomères-dimères est un modèle classique de physique statistique, avec de nombreuses applications dans d'autres domaines, que ce soit en recherche opérationnelle, en optimisation combinatoire ou encore en informatique théorique.

18.1 Définition du modèle

On considère le graphe $(\mathbb{Z}^d, \mathbb{A}^d)$, où $\mathbb{A}^d \coloneqq \{\{i,j\} \subset \mathbb{Z}^d \,|\, \|j-i\|_1 = 1\}$ est l'ensemble des arêtes joignant les sommets voisins de $\mathbb{Z}^d$. On écrira $e \sim e'$ pour signifier que deux arêtes $e \neq e' \in \mathbb{A}^d$ partagent une même extrémité. Pour tout[2] $\Lambda \Subset \mathbb{A}^d$, on note $\partial\Lambda \coloneqq \{e' \in \mathbb{A}^d \setminus \Lambda \,|\, \exists e \in \Lambda,\, e \sim e'\}$.

On notera $\Omega_\Lambda \coloneqq \{0,1\}^\Lambda$ l'ensemble des **configurations** $\omega = (\omega_e)_{e\in\Lambda}$ du modèle de monomères-dimères dans Λ. On écrira $|\omega| \coloneqq \#\{e \in \Lambda \,|\, \omega_e = 1\}$. Si Λ et Λ' sont deux sous-ensembles disjoints de $\mathbb{A}^d$, et $\omega \in \Omega_\Lambda$, $\omega' \in \Omega_{\Lambda'}$, on notera $\omega\omega'$ la configuration dans $\Lambda \cup \Lambda'$ coïncidant avec ω sur Λ et avec ω' sur Λ'. Étant donné $\Delta \subset \Lambda$, on notera $\omega_\Delta = (\omega_e)_{e\in\Delta}$ la restriction de la configuration ω à Δ. Finalement, on notera $\sigma_\Delta : \Omega_\Lambda \to \Omega_\Delta$ la variable aléatoire définie par $\sigma_\Delta(\omega) \coloneqq \omega_\Delta$, avec la notation abrégée $\sigma_e \equiv \sigma_{\{e\}}$ lorsque Δ se réduit à l'arête e.

Une configuration $\omega \in \Omega_\Lambda$ est **admissible** si $\omega_e = \omega_{e'} = 1$ implique que $e \not\sim e'$. L'interprétation est la suivante : les arêtes de Λ pour lesquelles $\omega_e = 1$ sont supposées occupées par un dimère (un polymère composé de deux monomères) ; chaque dimère occupe deux sommets voisins de $\mathbb{Z}^d$ et il est impossible pour deux dimères d'occuper un même sommet (*cf.* Fig. 18.1). En particulier, $|\omega|$ représente donc le nombre de dimères dans la configuration ω.

Soit $\alpha \in \Omega_{\partial\Lambda}$ admissible. Le modèle de monomères-dimères dans $\Lambda \Subset \mathbb{A}^d$ avec **condition au bord** α et **fugacité** $\lambda \in \mathbb{R}_+^*$ est la mesure de probabilité sur Ω_Λ associant à toute configuration $\omega \in \Omega_\Lambda$ la probabilité

$$\mu^\alpha_{\Lambda;\lambda}(\omega) \coloneqq \frac{\lambda^{|\omega|}\, \mathbf{1}_{\{\omega\alpha \text{ est admissible}\}}}{Z^\alpha_{\Lambda;\lambda}}.$$

La constante de normalisation $Z^\alpha_{\Lambda;\lambda} \coloneqq \sum_{\omega\in\Omega_\Lambda} \lambda^{|\omega|}\, \mathbf{1}_{\{\omega\alpha \text{ est admissible}\}}$ est appelée **fonction de partition**. Pour alléger les notations, on omettra généralement l'indice λ.

1. Traduction littérale du terme anglais *disagreement percolation* habituellement utilisé.
2. La notation $A \Subset B$ signifie que A est un sous-ensemble *fini* de B.

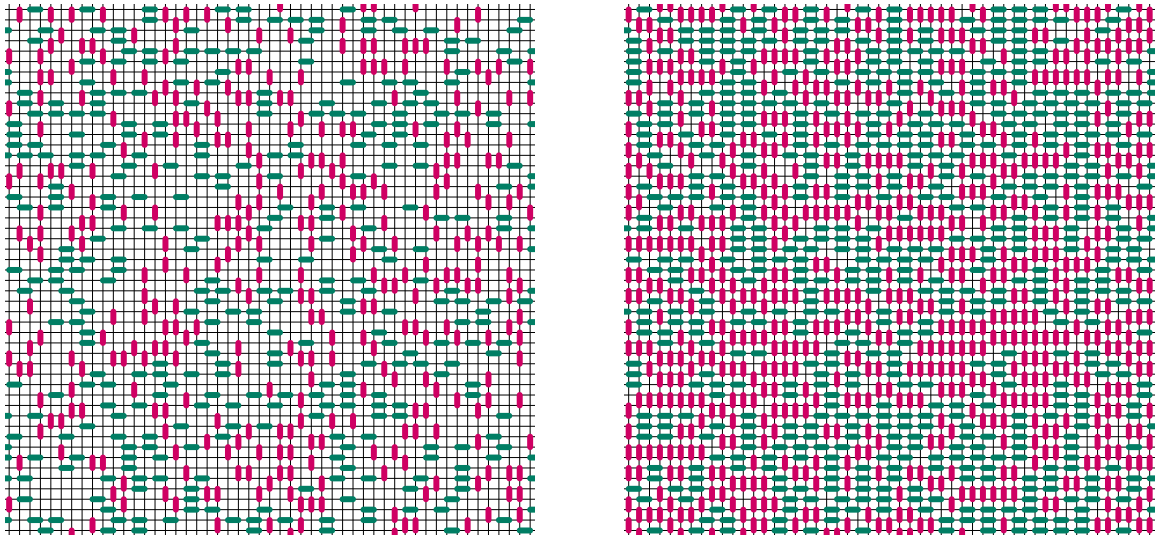


FIGURE 18.1: Deux configurations typiques du modèle de monomères-dimères ($\lambda = 0{,}5$ et $\lambda = 120$).

18.2 Propriété de Markov spatiale

Une propriété élémentaire, mais essentielle, de la mesure μ_Λ^α est la **propriété de Markov spatiale** suivante. Soit $\Delta \subset \Lambda \Subset \mathbb{A}^d$ et $\alpha \in \Omega_{\partial\Lambda}$ admissible. Pour toute configuration $\omega' \in \Omega_{\Lambda\setminus\Delta}$ telle que $\omega'\alpha$ soit admissible, et pour tout événement $A \subset \Omega_\Delta$,

$$\begin{aligned}\mu_\Lambda^\alpha(\sigma_\Delta \in A \,|\, \sigma_{\Lambda\setminus\Delta} = \omega') &= \frac{(Z_\Lambda^\alpha)^{-1} \sum_{\omega\in A} \mathbf{1}_{\{\omega\omega'\alpha \text{ est admissible}\}} \lambda^{|\omega|+|\omega'|}}{(Z_\Lambda^\alpha)^{-1} \sum_{\omega\in\Omega_\Delta} \mathbf{1}_{\{\omega\omega'\alpha \text{ est admissible}\}} \lambda^{|\omega|+|\omega'|}} \\ &= \frac{\sum_{\omega\in A} \mathbf{1}_{\{\omega\beta \text{ est admissible}\}} \lambda^{|\omega|}}{\sum_{\omega\in\Omega_\Delta} \mathbf{1}_{\{\omega\beta \text{ est admissible}\}} \lambda^{|\omega|}} \\ &= (Z_\Delta^\beta)^{-1} \sum_{\omega\in A} \mathbf{1}_{\{\omega\beta \text{ est admissible}\}} \lambda^{|\omega|} = \mu_\Delta^\beta(A),\end{aligned}$$

où l'on a noté $\beta = (\omega'\alpha)_{\partial\Delta}$. Un cas particulier qui nous sera utile plus tard est lorsque Δ est réduit à une unique arête $e \in \Lambda$. Dans ce cas, pour toute configuration $\omega' \in \Omega_{\Lambda\setminus\{e\}}$ telle que $\alpha\omega'$ soit admissible,

$$\mu_\Lambda^\alpha(\sigma_e = 0 \,|\, \sigma_{\Lambda\setminus\{e\}} = \omega') = \mu_{\{e\}}^\beta(\sigma_e = 0) = \begin{cases} (1+\lambda)^{-1} & \text{si } \beta_f = 0 \; \forall f \sim e, \\ 1 & \text{sinon.} \end{cases}$$

En particulier,

$$\min_{\omega'\in\Omega_{\Lambda\setminus\{e\}}} \mu_\Lambda^\alpha(\sigma_e = 0 \,|\, \sigma_{\Lambda\setminus\{e\}} = \omega') \geqslant \frac{1}{1+\lambda}. \tag{18.1}$$

Nous aurons besoin d'une autre conséquence élémentaire de la propriété de Markov spatiale. Étant donné $\Delta \subset \Lambda \Subset \mathbb{A}^d$, on notera $\mu_{\Lambda,\Delta;\lambda}^\alpha$ la marginale de $\mu_{\Lambda;\lambda}^\alpha$ dans Δ, c'est-à-dire la mesure de probabilité sur Ω_Δ associant à toute configuration $\omega \in \Omega_\Delta$ la probabilité

$$\mu_{\Lambda,\Delta;\lambda}^\alpha(\omega) \coloneqq \mu_{\Lambda;\lambda}^\alpha(\sigma_\Delta = \omega) = \sum_{\omega'\in\Omega_{\Lambda\setminus\Delta}} \mu_{\Lambda;\lambda}^\alpha(\omega\omega').$$

Soit $\Delta_1 \subset \Delta_2 \subset \Lambda \Subset \mathbb{A}^d$ et $\bar{\Delta}_2 = \Delta_2 \cup \partial\Delta_2$. Soit $A \subset \Omega_{\Delta_1}$ et $A' = \{\omega \in \Omega_{\Delta_2} \,|\, \omega_{\Delta_1} \in A\}$. Alors,[3]

$$\begin{aligned}
\mu^{\alpha}_{\Lambda,\Delta_1;\lambda}(A) &= \mu^{\alpha}_{\Lambda,\Delta_2;\lambda}(A') \\
&= \sum_{\beta\in\Omega_{\partial\Delta_2}} \sum_{\omega\in\Omega_{\Lambda\setminus\bar{\Delta}_2}} \mu^{\alpha}_{\Lambda}(\sigma_{\Delta_2} \in A' \,|\, \sigma_{\partial\Delta_2} = \beta, \sigma_{\Lambda\setminus\bar{\Delta}_2} = \omega)\, \mu^{\alpha}_{\Lambda}(\sigma_{\partial\Delta_2} = \beta, \sigma_{\Lambda\setminus\bar{\Delta}_2} = \omega) \\
&= \sum_{\beta\in\Omega_{\partial\Delta_2}} \mu^{\beta}_{\Delta_2}(A') \sum_{\omega\in\Omega_{\Lambda\setminus\bar{\Delta}_2}} \mu^{\alpha}_{\Lambda}(\sigma_{\partial\Delta_2} = \beta, \sigma_{\Lambda\setminus\bar{\Delta}_2} = \omega) \\
&= \sum_{\beta\in\Omega_{\partial\Delta_2}} \mu^{\beta}_{\Delta_2,\Delta_1}(A)\, \mu^{\alpha}_{\Lambda}(\sigma_{\partial\Delta_2} = \beta), \qquad (18.2)
\end{aligned}$$

où l'on a utilisé la propriété de Markov spatiale pour obtenir la troisième ligne.

18.3 Relaxation exponentielle

Le principal résultat de ce chapitre est que, pour tout $\lambda > 0$, le modèle de monomères-dimères est très peu sensible à la condition au bord utilisée. Afin de quantifier cela, nous utiliserons la notion de **distance en variation totale**. Étant donné deux mesures de probabilités μ et ν sur un même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, la distance en variation totale $\|\mu - \nu\|_{\mathrm{VT}}$ entre μ et ν est définie par

$$\|\mu - \nu\|_{\mathrm{VT}} \coloneqq \sup_{A\in\mathcal{F}} |\mu(A) - \nu(A)|.$$

Pour $n \geqslant 1$, on notera $\Delta_n \coloneqq \{\{i,j\} \in \mathbb{A}^d \,|\, i,j \in \{-n,\dots,n\}^d\}$.

Théorème 18.1. *Pour tout $\lambda > 0$, il existe K et $\kappa > 0$ tels que, pour tout $n \geqslant 1$, pour tous $\Lambda, \Lambda' \Subset \mathbb{A}^d$ tels que $\Lambda \cap \Lambda' \supset \Delta_{2n}$, et pour toutes conditions au bord $\alpha \in \Omega_{\partial\Lambda}$ et $\alpha' \in \Omega_{\partial\Lambda'}$ admissibles, on ait*

$$\|\mu^{\alpha}_{\Lambda,\Delta_n;\lambda} - \mu^{\alpha'}_{\Lambda',\Delta_n;\lambda}\|_{\mathrm{VT}} \leqslant K e^{-\kappa n}.$$

Intuitivement, ce résultat affirme que le comportement du système loin des parois du domaine Λ dépend très faiblement de la géométrie du domaine et de la condition au bord. En particulier, on peut montrer que ce résultat implique que les dimères ne vont pas s'orienter majoritairement dans une même direction (sinon on pourrait sélectionner cette direction en choisissant Λ et α de façon appropriée). On dit que ce modèle reste dans la **phase isotrope** quel que soit $\lambda > 0$ (*cf.* Fig. 18.1).

Notons que ceci n'est plus vrai lorsque l'on remplace les dimères par des k-mères (c'est-à-dire des « bâtonnets » de longueur k) avec k suffisamment grand [15], comme dans la simulation de la Fig. 18.2, ou que l'on autorise des k-mères de longueurs variables [30]. C'est également le cas pour le modèle de monomères-dimères lorsque l'on ajoute une contribution énergétique récompensant l'alignement des dimères [29]. Dans tous ces derniers cas, la majorité des bâtonnets s'alignent dans une même direction pour certains régimes de valeurs de la fugacité ; on dit alors que le système se trouve dans la **phase nématique**.

Le premier résultat démontrant l'absence d'ordre nématique pour le modèle de monomères-dimères a été obtenu dans le célèbre article [28], mais sous une forme assez différente. L'approche suivie ici, dont la preuve est considérablement plus simple, est due à [52].

18.4 Preuve du Théorème 18.1

La preuve du Théorème 18.1 sera une conséquence du Lemme suivant.

3. Comme d'habitude, si $\mathbb{P}$ est une mesure de probabilité et A, B deux événements, nous utilisons la convention que $\mathbb{P}(A \,|\, B)\mathbb{P}(B) = 0$ si $\mathbb{P}(B) = 0$.

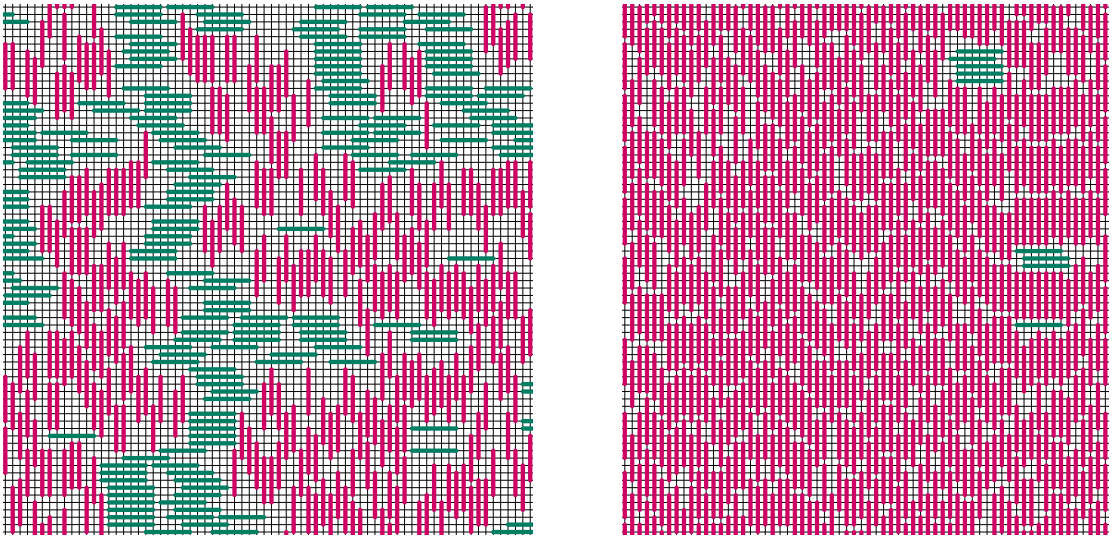


FIGURE 18.2: Lorsque les dimères sont remplacés par des k-mères avec k assez grand (conjecturalement, $k \geqslant 7$), on peut observer une phase isotrope à faible fugacité (gauche) et une phase nématique à fugacité intermédiaire (droite). À très grande fugacité, il est conjecturé qu'une nouvelle phase isotrope apparaît.

Lemme 18.2. *Pour tout $\lambda > 0$, il existe $\kappa' > 0$ tel que, pour tout $n \geqslant 1$, $e \in \partial\Delta_{2n}$ et $\alpha, \alpha' \in \Omega_{\partial\Delta_{2n}}$ admissibles satisfaisant $\alpha_f = \alpha'_f$ pour tout $f \neq e$, on ait*

$$\|\mu^{\alpha}_{\Delta_{2n},\Delta_n;\lambda} - \mu^{\alpha'}_{\Delta_{2n},\Delta_n;\lambda}\|_{\mathrm{VT}} \leqslant e^{-\kappa' n}.$$

Commençons par vérifier que ce lemme implique effectivement la validité du Théorème 18.1. Soit $A \subset \Omega_{\Delta_n}$ un événement ne dépendant que de la configuration à l'intérieur de Δ_n. On peut alors écrire, en utilisant (18.2),

$$\begin{aligned}\mu^{\alpha}_{\Lambda,\Delta_n}(A) &= \sum_{\beta\in\Omega_{\partial\Delta_{2n}}} \mu^{\alpha}_{\Lambda}(\sigma_{\partial\Delta_{2n}} = \beta)\, \mu^{\beta}_{\Delta_{2n},\Delta_n}(A) \\ &= \sum_{\beta,\beta'\in\Omega_{\partial\Delta_{2n}}} \mu^{\alpha}_{\Lambda}(\sigma_{\partial\Delta_{2n}} = \beta)\, \mu^{\alpha'}_{\Lambda'}(\sigma_{\partial\Delta_{2n}} = \beta')\, \mu^{\beta}_{\Delta_{2n},\Delta_n}(A),\end{aligned}$$

puisque $\sum_{\beta'\in\Omega_{\partial\Delta_{2n}}} \mu^{\alpha'}_{\Lambda'}(\sigma_{\partial\Delta_{2n}} = \beta') = 1$. Par conséquent,

$$\left|\mu^{\alpha}_{\Lambda,\Delta_n}(A) - \mu^{\alpha'}_{\Lambda',\Delta_n}(A)\right| \leqslant \sum_{\beta,\beta'\in\Omega_{\partial\Delta_{2n}}} \mu^{\alpha}_{\Lambda}(\sigma_{\partial\Delta_{2n}} = \beta)\, \mu^{\alpha'}_{\Lambda'}(\sigma_{\partial\Delta_{2n}} = \beta') \left|\mu^{\beta}_{\Delta_{2n},\Delta_n}(A) - \mu^{\beta'}_{\Delta_{2n},\Delta_n}(A)\right|.$$

Afin de borner $\left|\mu^{\beta}_{\Delta_{2n},\Delta_n}(A) - \mu^{\beta'}_{\Delta_{2n},\Delta_n}(A)\right|$, on peut utiliser le Lemme 18.2. Considérons une suite de conditions au bord $\beta_1 = \beta, \beta_2, \dots, \beta_{N-1}, \beta_N = \beta'$ telle que β_k et β_{k+1} ne diffèrent qu'en une unique arête. Évidemment, on peut supposer que $N \leqslant |\partial\Delta_{2n}|$. Le lemme implique alors

$$\left|\mu^{\beta}_{\Delta_{2n},\Delta_n}(A) - \mu^{\beta'}_{\Delta_{2n},\Delta_n}(A)\right| \leqslant \sum_{k=2}^{N} \left|\mu^{\beta_k}_{\Delta_{2n},\Delta_n}(A) - \mu^{\beta_{k-1}}_{\Delta_{2n},\Delta_n}(A)\right| \leqslant |\partial\Delta_{2n}|\, e^{-\kappa' n} \leqslant K e^{-\kappa n},$$

avec $\kappa = \kappa'/2$ et K une constante ne dépendant que de κ' et d. On a donc

$$\left|\mu^{\alpha}_{\Lambda,\Delta_n}(A) - \mu^{\alpha'}_{\Lambda',\Delta_n}(A)\right| \leqslant K e^{-\kappa n},$$

uniformément en $A \subset \Omega_{\Delta_n}$, ce qui démontre le Théorème 18.1.

Preuve du Lemme 18.2. Étant donné deux configurations $\omega, \omega' \in \Omega_{\Delta_{2n}}$ et une arête $e' \in \Delta_{2n} \cup \partial\Delta_{2n}$, on écrira $e' \overset{\neq}{\longleftrightarrow} \Delta_n$ si $(\omega\alpha)_{e'} \neq (\omega'\alpha')_{e'}$ et s'il existe une suite $e_1 \sim e_2 \sim \cdots \sim e_m$ d'arêtes distinctes de $\Delta_{2n} \setminus \{e'\}$ avec $e_1 \sim e'$, $e_m \in \Delta_n$ et telle que $\omega_{e_j} \neq \omega'_{e_j}$ pour tout $1 \leqslant j \leqslant m$. Soit $B = \big\{(\omega, \omega') \in \Omega_{\Delta_{2n}} \times \Omega_{\Delta_{2n}} \,\big|\, e \overset{\neq}{\longleftrightarrow} \Delta_n\big\}$, où l'on rappelle que e est l'unique arête de $\partial\Delta_{2n}$ sur laquelle les conditions au bord α et α' diffèrent. Finalement, on notera

$$C_{\neq} = C_{\neq}(\omega, \omega') \coloneqq \Delta_n \cup \big\{e' \in \Delta_{2n} \setminus \Delta_n \,\big|\, e' \overset{\neq}{\longleftrightarrow} \Delta_n\big\}$$

l'« amas de désaccord » de Δ_n. On peut alors définir une application $T : \Omega_{\Delta_{2n}} \times \Omega_{\Delta_{2n}} \to \Omega_{\Delta_{2n}} \times \Omega_{\Delta_{2n}}$ interchangeant ω et ω' sur $C_{\neq}$:

$$T(\omega, \omega') \coloneqq (\omega'_{C_{\neq}} \omega_{\Delta_{2n} \setminus C_{\neq}}, \omega_{C_{\neq}} \omega'_{\Delta_{2n} \setminus C_{\neq}}).$$

Évidemment, T est bijective et $(\omega, \omega') \in B \Leftrightarrow T(\omega, \omega') \in B$. De plus, lorsque B a lieu, $(\omega\alpha)_{\partial C_{\neq}} = (\omega'\alpha')_{\partial C_{\neq}}$. Par la propriété de Markov spatiale, T préserve donc la mesure produit $\mu^{\alpha}_{\Delta_{2n}} \times \mu^{\alpha'}_{\Delta_{2n}}$ sur B.

Soit $A \subset \Omega_{\Delta_n}$. On peut alors écrire

$$\begin{aligned}
\mu^{\alpha}_{\Delta_{2n},\Delta_n}(A) &= \mu^{\alpha}_{\Delta_{2n}} \times \mu^{\alpha'}_{\Delta_{2n}}\big(\{(\omega, \omega') \in \Omega_{\Delta_{2n}} \times \Omega_{\Delta_{2n}} \,|\, \omega_{\Delta_n} \in A\}\big) \\
&= \mu^{\alpha}_{\Delta_{2n}} \times \mu^{\alpha'}_{\Delta_{2n}}\big(\{(\omega, \omega') \in \Omega_{\Delta_{2n}} \times \Omega_{\Delta_{2n}} \,|\, \omega_{\Delta_n} \in A, (\omega, \omega') \in B\}\big) \\
&\quad + \mu^{\alpha}_{\Delta_{2n}} \times \mu^{\alpha'}_{\Delta_{2n}}\big(\{(\omega, \omega') \in \Omega_{\Delta_{2n}} \times \Omega_{\Delta_{2n}} \,|\, \omega_{\Delta_n} \in A, (\omega, \omega') \in B^{\mathrm{c}}\}\big).
\end{aligned}$$

De façon similaire,

$$\begin{aligned}
\mu^{\alpha'}_{\Delta_{2n},\Delta_n}(A) &= \mu^{\alpha}_{\Delta_{2n}} \times \mu^{\alpha'}_{\Delta_{2n}}\big(\{(\omega, \omega') \in \Omega_{\Delta_{2n}} \times \Omega_{\Delta_{2n}} \,|\, \omega'_{\Delta_n} \in A, (\omega, \omega') \in B\}\big) \\
&\quad + \mu^{\alpha}_{\Delta_{2n}} \times \mu^{\alpha'}_{\Delta_{2n}}\big(\{(\omega, \omega') \in \Omega_{\Delta_{2n}} \times \Omega_{\Delta_{2n}} \,|\, \omega'_{\Delta_n} \in A, (\omega, \omega') \in B^{\mathrm{c}}\}\big).
\end{aligned}$$

Or, puisque T préserve la mesure sur B, on a l'identité

$$\begin{aligned}
&\mu^{\alpha}_{\Delta_{2n}} \times \mu^{\alpha'}_{\Delta_{2n}}\big(\{(\omega, \omega') \in \Omega_{\Delta_{2n}} \times \Omega_{\Delta_{2n}} \,|\, \omega_{\Delta_n} \in A, (\omega, \omega') \in B\}\big) \\
&\qquad = \mu^{\alpha}_{\Delta_{2n}} \times \mu^{\alpha'}_{\Delta_{2n}}\big(\{(\omega, \omega') \in \Omega_{\Delta_{2n}} \times \Omega_{\Delta_{2n}} \,|\, \omega'_{\Delta_n} \in A, (\omega, \omega') \in B\}\big).
\end{aligned}$$

On conclut donc que

$$\big|\mu^{\alpha}_{\Delta_{2n},\Delta_n}(A) - \mu^{\alpha'}_{\Delta_{2n},\Delta_n}(A)\big| \leqslant \mu^{\alpha}_{\Delta_{2n}} \times \mu^{\alpha'}_{\Delta_{2n}}(B^{\mathrm{c}}).$$

Il ne reste donc plus qu'à borner la probabilité dans le membre de droite. Par définition, B^{c} se produit si et seulement si il existe une suite $e_1 \sim e_2 \sim \cdots \sim e_m$ d'arêtes distinctes de Δ_{2n} telle que $e_1 \sim e$, $e_m \in \Delta_n$ et $\omega_{e_i} \neq \omega'_{e_i}$ pour tout $1 \leqslant i \leqslant m$.

Soit $k \in \mathbb{N}^*$. L'observation clé est que, presque-sûrement, il ne peut exister qu'au plus un chemin $e \sim e_1 \sim \cdots \sim e_k$ avec $\omega_{e_i} \neq \omega'_{e_i}$ pour tout $1 \leqslant i \leqslant k$ (un chemin de désaccord de longueur k partant de e). En effet, supposons qu'on puisse en trouver deux : $e_0 = e \sim e_1 \sim \cdots \sim e_k$ et $e'_0 = e \sim e'_1 \sim \cdots \sim e'_k$. Soit $r = \min\{\ell \geqslant 1 \,|\, e_\ell \neq e'_\ell\}$. Alors les trois arêtes $e_{r-1} = e'_{r-1}$, e_r et e'_r partagent toutes une même extrémité. Mais comme ω et ω' diffèrent sur chacune de ces arêtes, on conclut qu'il y a soit 2 arêtes sur lesquelles ω prend la valeur 1, soit 2 arêtes sur lesquelles ω' prend la valeur 1. Les configurations $\alpha\omega$ et $\alpha'\omega'$ ne peuvent ainsi pas être toutes deux admissibles.

Soit donc $e \sim e_1 \sim \cdots \sim e_k$ un chemin de désaccord de longueur k. Sans perte de généralité, on supposera que $\alpha_e = 1$. L'observation précédente implique alors que tout chemin de désaccord de longueur $k+1$ partant de e doit être de la forme $e \sim e_1 \sim \cdots \sim e_k \sim \tilde{e}$. Supposons que k est impair (sinon le même argument s'applique en remplaçant ω par ω'). Dans ce cas, $\omega_{\tilde{e}} = 1$. Mais il suit de (18.1) que la probabilité que $\omega_f = 0$ pour toutes les arêtes $f \sim e_k$ non encore utilisées est au moins égale

à $(1+\lambda)^{-(2d-1)}$ indépendamment des valeurs prises par ω sur les autres arêtes. Par conséquent, la probabilité qu'il existe un chemin de désaccord de longueur $k+1$ partant de e sachant qu'il en existe un de longueur k est au plus égale à $1-(1+\lambda)^{-(2d-1)}$. On conclut que la probabilité qu'il y ait un chemin de désaccord joignant e à Δ_n (et donc de longueur au moins n) est au plus égale à $\left(1-(1+\lambda)^{-(2d-1)}\right)^n$, et le lemme est démontré. □

Remarques bibliographiques : Ce chapitre s'inspire fortement de l'article [52].

Bibliographie

[1] A. Abdesselam. The weakly dependent strong law of large numbers revisited. *Grad. J. Math.*, 3(2) :94–97, 2018.

[2] D. Aldous and P. Diaconis. Shuffling cards and stopping times. *Amer. Math. Monthly*, 93(5) :333–348, 1986.

[3] D. Aldous and J. Fill. Reversible markov chains and random walks on graphs. Disponible à l'adresse http://www.stat.berkeley.edu/~aldous/RWG/book.html.

[4] 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.

[5] R. Arratia, L. Goldstein, and L. Gordon. Two moments suffice for Poisson approximations : the Chen-Stein method. *Ann. Probab.*, 17(1) :9–25, 1989.

[6] 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.

[7] A. A. Barker. Monte Carlo calculations of the radial distribution functions for a proton–electron plasma. *Aust. J. Phys.*, 18(2) :119–134, 1965.

[8] D. Bayer and P. Diaconis. Trailing the dovetail shuffle to its lair. *Ann. Appl. Probab.*, 2(2) :294–313, 1992.

[9] G. D. Birkhoff. Proof of the ergodic theorem. *Proc. Nat. Acad. Sci. U.S.A.*, 17 :656–660, 1931.

[10] P. Brémaud. *Discrete probability models and methods*, volume 78 of *Probability Theory and Stochastic Modelling*. Springer, Cham, 2017.

[11] P. Brémaud. *Markov chains—Gibbs fields, Monte Carlo simulation and queues*, volume 31 of *Texts in Applied Mathematics*. Springer, Cham, 2020.

[12] M. Campanino, D. Ioffe, and Y. Velenik. Ornstein-Zernike theory for finite range Ising models above T_c. *Probab. Theory Related Fields*, 125(3) :305–349, 2003.

[13] 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.

[14] P. R. de Montmort. *Essay d'analyse sur les jeux de hazard*. J. Quillau, 1713.

[15] M. Disertori and A. Giuliani. The nematic phase of a system of long hard rods. *Comm. Math. Phys.*, 323(1) :143–175, 2013.

[16] R. L. Dobrushin and S. B. Shlosman. Large and moderate deviations in the Ising model. In *Probability contributions to statistical mechanics*, volume 20 of *Adv. Soviet Math.*, pages 91–219. Amer. Math. Soc., Providence, RI, 1994.

[17] P. G. Doyle and J. L. Snell. *Random walks and electric networks*, volume 22 of *Carus Mathematical Monographs*. Mathematical Association of America, Washington, DC, 1984. Également disponible à l'adresse http://arxiv.org/abs/math.PR/0001057.

[18] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6 :290–297, 1959.

[19] W. Feller. The general form of the so-called law of the iterated logarithm. *Trans. Amer. Math. Soc.*, 54 :373–402, 1943.

[20] W. Feller. The law of the iterated logarithm for identically distributed random variables. *Ann. of Math. (2)*, 47 :631–638, 1946.

[21] W. Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons, Inc., New York-London-Sydney, 1968.

[22] S. Friedli and Y. Velenik. *Statistical mechanics of lattice systems : a concrete mathematical introduction*. Cambridge University Press, Cambridge, 2018. Également disponible à l'adresse http://www.unige.ch/math/folks/velenik/smbook.

[23] E. N. Gilbert. Random graphs. *Ann. Math. Statist.*, 30 :1141–1144, 1959.

[24] B. V. Gnedenko and A. N. Kolmogorov. *Limit distributions for sums of independent random variables*. Addison-Wesley Publishing Company, Inc., Cambridge, Mass., 1954. Translated and annotated by K. L. Chung. With an Appendix by J. L. Doob.

[25] A. Guionnet. *Large random matrices : lectures on macroscopic asymptotics*, volume 1957 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2009.

[26] O. Häggström. *Finite Markov chains and algorithmic applications*, volume 52 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 2002.

[27] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1) :97–109, 1970.

[28] O. J. Heilmann and E. H. Lieb. Theory of monomer-dimer systems. *Comm. Math. Phys.*, 25 :190–232, 1972.

[29] O. J. Heilmann and E. H. Lieb. Lattice models for liquid crystals. *J. Statist. Phys.*, 20(6) :679–693, 1979.

[30] D. Ioffe, Y. Velenik, and M. Zahradník. Entropy-driven phase transition in a polydisperse hard-rods lattice system. *J. Stat. Phys.*, 122(4) :761–786, 2006.

[31] O. Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[32] 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.

[33] A. Khintchine. Über einen neuen Grenzwertsatz der Wahrscheinlichkeitsrechnung. *Math. Ann.*, 101(1) :745–752, 1929.

[34] M. Krivelevich and B. Sudakov. The phase transition in random graphs : a simple proof. *Random Structures Algorithms*, 43(2) :131–138, 2013.

[35] 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.

[36] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[37] D. A. Levin and Y. Peres. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2017.

[38] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.

[39] T. Lindvall. *Lectures on the coupling method*. Dover Publications, Inc., Mineola, NY, 2002. Corrected reprint of the 1992 original.

[40] R. Lyons and Y. Peres. *Probability on Trees and Networks*. Cambridge University Press, New York, 2016. Disponible à l'adresse http://pages.iu.edu/~rdlyons/.

[41] 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.

[42] N. Metropolis and S. Ulam. The Monte Carlo method. *J. Amer. Statist. Assoc.*, 44 :335–341, 1949.

[43] V. V. Petrov. *Sums of independent random variables*. Springer-Verlag, New York, 1975. Translated from the Russian by A. A. Brown, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82.

[44] 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.

[45] S. Roch. Modern discrete probability : An essential toolkit, 2023. Notes de cours, disponibles à l'adresse https://people.math.wisc.edu/~roch/mdp/.

[46] D. Romik. *The surprising mathematics of longest increasing subsequences*, volume 4 of *Institute of Mathematical Statistics Textbooks*. Cambridge University Press, New York, 2015.

[47] P. M. Soardi. *Potential theory on infinite networks*, volume 1590 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1994.

[48] F. Solomon. Random walks in a random environment. *Ann. Probability*, 3 :1–31, 1975.

[49] R. H. Swendsen and J.-S. Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.*, 58 :86–88, Jan 1987.

[50] M. Talagrand. A new look at independence. *Ann. Probab.*, 24(1) :1–34, 1996.

[51] H. Thorisson. *Coupling, stationarity, and regeneration*. Probability and its Applications (New York). Springer-Verlag, New York, 2000.

[52] J. van den Berg. On the absence of phase transition in the monomer-dimer model. In *Perplexing problems in probability*, volume 44 of *Progr. Probab.*, pages 185–195. Birkhäuser Boston, Boston, MA, 1999.

[53] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)*, 67 :325–327, 1958.

[54] O. Zeitouni. *Lecture notes on random walks in random environment*, volume 1837 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2004.

Notations

Ensembles, objets

K_n	le graphe complet à n sommets
$[\![n]\!]$	l'ensemble $\{1, \dots, n\}$
$\mathbb{Z}_{\geqslant \ell}$	l'ensemble $\{k \in \mathbb{Z} \mid k \geqslant \ell\}$
$\mathbb{Z}_{> \ell}$	l'ensemble $\{k \in \mathbb{Z} \mid k > \ell\}$
$\mathbb{Z}_{\leqslant \ell}$	l'ensemble $\{k \in \mathbb{Z} \mid k \leqslant \ell\}$
$\mathbb{Z}_{< \ell}$	l'ensemble $\{k \in \mathbb{Z} \mid k < \ell\}$
$\mathbb{N}$	l'ensemble $\mathbb{Z}_{\geqslant 0}$ des entiers positifs
$\mathbb{N}^*$	l'ensemble $\mathbb{Z}_{> 0}$ des entiers strictement positifs
$\mathbb{R}_+$	l'ensemble $\{x \in \mathbb{R} \mid x \geqslant 0\}$ 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

$\mathsf{G}(n, p)$	graphe d'Erdős–Rényi avec n sommets et probabilité de rétention p
$\mathrm{Bern}(p)$	loi de Bernoulli de paramètre p
$\mathrm{Binom}(n, p)$	loi binomiale de paramètres n et p
$\mathcal{N}(\mu, \sigma^2)$	loi normale de paramètres μ et σ^2
$\mathrm{Poisson}(\lambda)$	loi de Poisson de paramètre λ
p_λ	mesure de probabilité associée à la loi de Poisson de paramètre λ

Divers

$a \coloneqq b$	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
$\lVert \mu - \nu \rVert_{\mathrm{VT}}$	distance en variation totale entre les mesures μ et ν

i	$\sqrt{-1}$
$\mathscr{L}(X)$	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 : $F_X(x) \coloneqq \mathbb{P}(X \leqslant x)$
φ_X	fonction caractéristique de X : $\varphi_X(t) \coloneqq \mathbb{E}[e^{\mathrm{i}tX}]$
$X \overset{\text{loi}}{=} Y$	X a la même loi que Y
$X \overset{\text{loi}}{=} Y \mid A$	X a la même loi que Y conditionnée par l'événement A
$a \vee b$	$\max\{a, b\}$
$a \wedge b$	$\min\{a, b\}$
$\lVert\cdot\rVert_1$	la norme ℓ^1 ou L^1
$\lVert\cdot\rVert_2$	la norme ℓ^2 ou L^2
$\lVert\cdot\rVert_\infty$	la norme ℓ^∞ ou L^∞

Index