

Chaînes de Markov

Jean Bérard

Avertissement

Ces notes sont en cours d'élaboration. Il se peut donc qu'y subsistent un certain nombre d'erreurs, d'incohérences, et/ou de passages inachevés.

Table des matières

Introduction	7
1 Propriété de Markov	9
1.1 Définition fondamentale	9
1.2 Exemples	12
1.3 Mesure sur l'espace des trajectoires	17
1.4 Propriété de Markov forte	19
1.5 Décomposition d'une trajectoire selon les retours successifs en un point	21
1.6 Action sur les mesures, action sur les fonctions : le noyau comme opérateur linéaire	23
1.7 Extensions de la notion de chaîne de Markov	26
1.7.1 Chaîne de Markov en temps discret et espace quelconque	26
1.7.2 Chaîne de Markov en temps continu et espace discret	27
1.7.3 Processus de Markov en temps continu	35
1.7.4 Champs markoviens	36
1.7.5 Chaînes de Markov avec plusieurs pas de mémoire	37
1.7.6 Semi-chaînes de Markov	38
1.7.7 Chaînes de Markov cachées	38
1.8 Exercices supplémentaires	41
2 Décompositions de l'espace d'états, récurrence/transience	45
2.1 Points essentiels, communication entre points, irréductibilité, période	45
2.1.1 Points essentiels	45
2.1.2 Communication entre points, irréductibilité	46
2.1.3 Période	48
2.2 Récurrence et transience	49
3 Théorie du potentiel, mesures et lois invariantes	63
3.1 L'équation de Poisson avec frontière absorbante	64
3.1.1 Des fonctions de Green	65

3.1.2	Des martingales	68
3.1.3	Questions d'unicité	70
3.1.4	Quelques exemples classiques	73
3.2	Mesures et lois invariantes	75
3.2.1	Renversé dans le temps d'un noyau par rapport à une mesure invariante	77
3.2.2	Mesures invariantes et récurrence/transience	79
3.2.3	Réversibilité	85
3.3	Exercices supplémentaires	87
4	Fonctionnelles additives : loi des grands nombres	93
4.1	Résultat principal	93
4.2	Preuves	94
4.2.1	Approche par renouvellement	95
4.2.2	Approche par la théorie ergodique	98
4.3	Exercices	98
5	Comportement asymptotique de la loi de X_n	101
5.1	Plan de cette partie	104
5.2	Preuve par renouvellement	104
5.3	Preuve par couplage et distance en variation totale	106
5.3.1	Distance en variation totale entre deux mesures de probabilité et couplage	106
5.3.2	Approche par couplage pour la convergence des chaînes de Markov	108
5.4	Temps stationnaires forts et distance en séparation	113
5.4.1	Deux exemples de temps stationnaires forts	117
5.5	Approche spectrale	117
5.5.1	Approche spectrale dans le cas où S est fini	118
5.6	Théorie L^2	120
5.6.1	Généralités	120
5.6.2	Formes de Dirichlet	121
5.6.3	Le cas réversible	123
5.7	Entropie relative	125
6	Une première approche quantitative de l'ergodicité pour la distance en variation totale	129
6.1	Ergodicité de degré 2	129
6.2	Ergodicité géométrique	135
6.2.1	Ergodicité uniforme	140

7	Fonctionnelles additives : théorème de la limite centrale	145
7.1	Approche par renouvellement	146
7.2	Approche par les martingales et l'équation de Poisson	158
7.3	Calculs asymptotiques de variance	163
8	Critères de dérive	167
8.1	Un critère de non-réurrence positive.	167
8.2	Un critère de transience	168
8.3	Un critère de récurrence	169
8.4	Un critère de récurrence positive	170
8.4.1	Un critère d'ergodicité géométrique	172
9	Principe des méthodes de Monte-Carlo par chaînes de Markov	175
9.1	Intérêt des méthodes MCMC	175
9.2	Qualité de l'approximation	176
9.3	Deux exemples	177
9.3.1	Algorithme de Metropolis	177
9.3.2	Echantillonneur de Gibbs	177
10	Appendice : Mémento sur les martingales	179
11	Appendice : Mémento sur la théorie ergodique	181

Introduction

Les suites de variables aléatoires présentant une structure de dépendance markovienne sont probablement, après les suites de variables aléatoires i.i.d., l'une des catégories d'objets les plus étudiées en probabilités. Elles jouent un grand rôle à la fois dans la théorie et dans les applications, et ont fait l'objet de développements considérables depuis leur introduction par A. A. Markov au début du vingtième siècle. Dans ce cours, nous ne pouvons bien entendu présenter qu'une petite sélection de résultats sur le sujet. Ces notes traitent principalement des chaînes à temps et espace discret, et, parmi les questions abordées, mentionnons : la propriété de Markov et ses diverses extensions, les propriétés de récurrence/transience, le problème des mesures invariantes, le comportement asymptotique de la loi d'une chaîne de Markov ainsi que des fonctions additives de ses trajectoires, les méthodes de Monte-Carlo par chaînes de Markov. Un large éventail de techniques (couplage, régénération, étude du semi-groupe, réseaux électriques) est présenté pour l'étude de ces questions. La lecture de ces notes, issues d'un cours de 2ème année de Master, suppose une connaissance préalable de la théorie des probabilités discrètes, et, pour certains points, fait appel à la théorie générale de la mesure, aux propriétés des martingales et à des rudiments de théorie ergodique¹. De nombreux ouvrages sur le sujet existent (nous avons par exemple utilisé l'ouvrage [7] pour la préparation de ces notes), consacrés aux chaînes de Markov en général, ou à certains aspects plus spécialisés de la question. En général, des références sont données au fur et à mesure du texte pour permettre un approfondissement des sujets présentés.

¹Les propriétés et définitions qui nous seront utiles concernant les martingales et la théorie ergodique sont rappelées en appendice.

Chapitre 1

Propriété de Markov

1.1 Définition fondamentale

Avant même de pouvoir donner la définition d'une chaîne de Markov, il nous faut définir la notion fondamentale de noyau de transition. Précisément, étant donné un ensemble fini ou dénombrable S , on appelle **noyau de transition sur S** toute application $p = p(\cdot, \cdot) : S \times S \rightarrow \mathbb{R}_+$ telle que, pour tout $x \in S$, $p(x, \cdot)$ est une loi de probabilité sur S .

Dans tout ce cours, la convention de notation suivante sera utilisée : étant donnée une probabilité ν définie sur un ensemble S fini ou dénombrable, nous nous autoriserons à utiliser indifféremment la notation $\nu(x)$ avec $x \in S$, en voyant ν comme une fonction définie sur S et à valeurs dans \mathbb{R}_+ vérifiant $\sum_{x \in S} \nu(x) = 1$, et la notation $\nu(A)$ avec $A \subset S$, en voyant ν comme une fonction définie sur l'ensemble des parties de S et à valeurs dans \mathbb{R}_+ , les deux notations étant reliées par la relation $\nu(A) = \sum_{x \in A} \nu(x)$.

Par ailleurs, nous utiliserons abondamment la notation vectorielle abrégée $x_{m:n}$ au lieu de $(x_m, x_{m+1}, \dots, x_n)$, étendue en $x_{m:\infty} = (x_m, x_{m+1}, \dots)$.

Nous pouvons maintenant donner la définition de l'objet fondamental dont traite ce cours. Étant donné un ensemble fini ou dénombrable S , une suite de variables aléatoires $(X_n)_{n \geq 0}$ définie sur un espace de probabilité (Ω, \mathcal{F}, P) et à valeurs dans S est appelée une **chaîne de Markov d'espace d'états S** lorsqu'il existe une famille de noyaux de transitions $(p_n(\cdot, \cdot))_{n \geq 0}$ sur S et une loi de probabilité ν sur S tels que : pour tout $n \geq 0$, et toute suite x_0, \dots, x_n d'éléments de S ,

$$P(X_{0:n} = x_{0:n}) = \nu(x_0) \prod_{i=0}^{n-1} p_i(x_i, x_{i+1}). \quad (1.1)$$

On dit alors que $(X_n)_{n \geq 0}$ est une **chaîne de Markov de loi initiale ν et de famille de noyaux de transition $(p_n(\cdot, \cdot))_{n \geq 0}$** . On vérifie sur la définition ci-dessus que la

loi de la suite $(X_n)_{n \geq 0}$ est complètement déterminée par la donnée de la loi initiale et de la famille de noyaux de transition.

On peut donner une définition similaire pour une chaîne de longueur finie : étant donné $m \geq 1$, une suite de variables X_0, \dots, X_m à valeurs dans S est une chaîne de Markov de loi initiale ν et de famille de noyaux de transition $(p_n(\cdot, \cdot))_{0 \leq n \leq m-1}$ si (1.1) est satisfaite pour tout $n \leq m$.

Exercice 1 *Montrer que si la propriété (1.1) est vérifiée pour un entier donné $n \geq 1$, elle est automatiquement vérifiée pour tout $m \leq n$.*

On déduit de cette définition la propriété fondamentale suivante, appelée **propriété de Markov**.

Proposition 1 *Etant donné une chaîne de Markov $(X_n)_{n \geq 0}$ sur un ensemble fini ou dénombrable S , définie sur un espace de probabilité (Ω, \mathcal{F}, P) , pour tous $m \geq 0$ et $n \geq m + 1$, et toute suite x_0, \dots, x_n d'éléments de S telle que $P(X_{0:m} = x_{0:m}) > 0$,*

$$P(X_{m+1:n} = x_{m+1:n} | X_{0:m} = x_{0:m}) = P(X_{m+1:n} = x_{m+1:n} | X_m = x_m).$$

On énonce souvent cette propriété de manière informelle en disant que, pour une chaîne de Markov (où l'indice m est interprété comme un temps), les états futurs de la chaîne ne dépendent de ses états passés que par l'intermédiaire de l'état présent. La preuve résulte de (1.1) par un calcul immédiat conduisant à l'identité suivante :

$$P(X_{m+1:n} = x_{m+1:n} | X_{0:n} = x_{0:n}) = p_m(x_m, x_{m+1}) \cdots p_{n-1}(x_{n-1}, x_n). \quad (1.2)$$

On note que l'expression ci-dessus n'est autre que la probabilité pour que les $n - m$ premiers pas d'une chaîne de Markov de loi initiale δ_{x_m} et de famille de noyaux de transition $(p_{m+k})_{k \geq 0}$ soient donnés par x_{m+1}, \dots, x_n . Dit de manière informelle : «Sachant toutes les valeurs passées jusqu'au temps n (y compris la valeur de X_n), la chaîne se comporte à partir du temps n comme une nouvelle chaîne de Markov, issue de X_n .»

On déduit de ce qui précède le lien suivant entre la loi ν et les noyaux de transition $p_i(\cdot, \cdot)$ intervenant dans la définition que nous avons donnée d'une chaîne de Markov :

$$\nu = \text{loi de } X_0, \quad (1.3)$$

et, pour tout $n \geq 0$, tout $x \in S$ tel que $P(X_m = x) > 0$, et tout $y \in S$,

$$p_n(x, y) = P(X_{n+1} = y | X_n = x). \quad (1.4)$$

Observons que la donnée de $(X_n)_{n \geq 0}$ n'impose aucune contrainte sur la valeur de $p_n(x, \cdot)$ lorsque $P(X_n = x) = 0$. Il est cependant plus commode de supposer *a priori* dans la définition d'une chaîne de Markov que l'on a affaire à un noyau de transition.

Inversement, si une suite de variables aléatoires $(X_n)_{n \geq 0}$ vérifie la propriété énoncée dans la proposition 1, on vérifie que, en définissant ν et $p_n(\cdot, \cdot)$ par les identités (1.3) et (1.4) ci-dessus, et en définissant arbitrairement $p_n(x, \cdot)$ lorsque $P(X_n = x) = 0$, la définition d'une chaîne de Markov est satisfaite. Autrement dit, **la propriété de Markov caractérise les chaînes de Markov.**

Exercice 2 Vérifier que $(X_n)_{n \geq 0}$ est une chaîne de Markov si et seulement si, pour tous $n \geq 0$, il existe une fonction f_n définie sur $S \times S$ telle que, pour tous $x_0, \dots, x_n \in S$ tels que $P(X_0 = x_0, \dots, X_n = x_n) > 0$,

$$P(X_{n+1} = y | X_{0:n} = x_{0:n}) = f_n(x_n, y)$$

Exercice 3 Considérons une famille de variables aléatoires X_0, \dots, X_n à valeurs dans un ensemble fini ou dénombrable S et définies sur (Ω, \mathcal{F}, P) . Supposons qu'il existe des fonctions g_0, \dots, g_{n-1} définies sur $S \times S$ telles que $P(X_{0:n} = x_{0:n}) = \prod_{i=0}^{n-1} g_i(x_i, x_{i+1})$. Montrer que X_0, \dots, X_n est une chaîne de Markov sur S . Préciser la loi initiale et les noyaux de transition.

Une manière plus sophistiquée, mais équivalente, d'énoncer la propriété de Markov, consiste à faire appel à la notion de tribu et de probabilité conditionnelle à une tribu. Etant donnée une variable aléatoire Z définie sur (Ω, \mathcal{F}) , nous utiliserons la notation classique $\sigma(Z)$ pour la sous-tribu de \mathcal{F} engendrée par Z . Plus particulièrement, pour tout $n \geq 0$, nous utiliserons la notation $\mathcal{F}_n := \sigma(X_0, \dots, X_n)$. On peut alors reformuler la propriété de Markov de la manière suivante.

Proposition 2 Etant donné une chaîne de Markov $(X_n)_{n \geq 0}$ sur un ensemble fini ou dénombrable S , définie sur un espace de probabilité (Ω, \mathcal{F}, P) , pour tous $m \geq 0$ et $n \geq m + 1$, et toute suite x_{m+1}, \dots, x_n d'éléments de S , l'égalité suivante a lieu P -presque sûrement :

$$P(X_{m+1:n} = x_{m+1:n} | \mathcal{F}_m) = P(X_{m+1:n} = x_{m+1:n} | \sigma(X_m)).$$

Exercice 4 Vérifier l'équivalence entre les propriétés énoncées dans les propositions 1 et 2.

Dans certains exemples, on peut être amené à étudier la validité de la propriété énoncée dans la proposition 2 en remplaçant la filtration $(\mathcal{F}_n)_{n \geq 0}$ par une plus grosse. Plus précisément, supposons donnée une famille de sous-tribus $(\mathcal{G}_m)_{m \geq 0}$ telle que, pour tous $n \geq m$, $\mathcal{G}_m \subset \mathcal{G}_n$, et vérifiant $\mathcal{F}_m \subset \mathcal{G}_m$ pour tout $m \geq 0$. On dit que $(X_n)_{n \geq 0}$ vérifie la propriété de Markov par rapport à $(\mathcal{G}_m)_{m \geq 0}$ si la propriété énoncée dans la proposition 2 est valable lorsque l'on remplace \mathcal{F}_n par \mathcal{G}_n .

Exercice 5 Vérifier que la propriété énoncée ci-dessus entraîne la propriété de Markov énoncée dans la proposition 2.

Exercice 6 Supposons donnée en plus de $(X_n)_{n \geq 0}$ une famille de variables aléatoires $(Y_n)_{n \geq 0}$ à valeurs dans un ensemble fini ou dénombrable (et définies sur (Ω, \mathcal{F}, P)), et que $(X_n)_{n \geq 0}$ vérifie la propriété de Markov par rapport à la famille de tribus définies par $\mathcal{G}_n := \sigma(\mathcal{F}_n, Y_0, \dots, Y_n)$. Donner une formulation de la propriété de Markov reprenant la forme de l'énoncé de la proposition 1.

Lorsque les noyaux de transition associés aux différents pas de la chaîne sont identiques, c'est-à-dire qu'il existe un noyau p telle que $p_n = p$ pour tout n , on dit que l'on a affaire à une **chaîne de Markov homogène** (ou parfois, mais cette terminologie n'est pas très heureuse, de chaîne de Markov à probabilités de transition stationnaires). Dans ce cours, nous considérerons principalement des chaînes de Markov homogènes, et cette propriété d'homogénéité sera sous-entendue la plupart du temps. Quand nous aurons affaire à des chaînes pour lesquelles cette propriété est en défaut, nous le spécifierons en parlant de **chaîne de Markov inhomogène**.

1.2 Exemples

Exercice 7 Vérifier qu'une suite de variables aléatoires *i.i.d.* forme une chaîne de Markov. Quels sont la loi initiale et le noyau de transition correspondants ?

Exercice 8 Etant donnée une suite de variables aléatoires $(Z_n)_{n \geq 0}$ définies sur (Ω, \mathcal{F}, P) , posons, pour tout n , $X_n := (Z_0, \dots, Z_n)$. Montrer que $(X_n)_{n \geq 0}$ est une chaîne de Markov. (Préciser l'espace d'états et les noyaux de transition.)

L'exercice ci-dessus montre que toute suite de variables aléatoires discrètes peut en fait être transformée en une chaîne de Markov. Cependant, la chaîne ainsi obtenue possède souvent une structure trop complexe pour que l'on puisse utilement analyser son comportement, ce qui limite quelque peu la portée de la remarque.

Exercice 9 (Fonctions itérées)

1) Considérons une suite de *i.i.d.* de fonctions aléatoires $(f_i)_{i \geq 0}$ de S dans S (S^S étant muni de la tribu produit), définies sur un espace de probabilité (Ω, \mathcal{F}, P) . Vérifier que, pour tout $x \in S$, la suite de variables aléatoires définie par $X_0(x) := x$ et, pour $n \geq 1$, $X_n(x) := f_{n-1} \circ \dots \circ f_0(x)$, est une chaîne de Markov homogène dont les probabilités de transition sont données par $p(x, y) = P(f(x) = y)$.

2) Est-il nécessaire que les fonctions aléatoires que l'on compose possèdent soient identiquement distribuées pour que l'on obtienne une chaîne de Markov ? Même question avec l'indépendance de ces fonctions entre elles ?

3) Inversement, montrer que, étant donnée un noyau de transition $p(\cdot, \cdot)$ sur $S \times S$, on peut toujours définir une suite de fonctions telle que $(X_n(x))_{n \geq 0}$ soit une chaîne de Markov initialisée en x et possédant p pour noyau de transition. Indication : on peut par exemple chercher à écrire $f_i(x)$ sous la forme $f(x, U_i)$, la suite $(U_i)_{i \geq 0}$ étant i.i.d. et de loi uniforme sur $[0, 1]$. Discuter de la pertinence de ce résultat pour la simulation.

D'après l'exercice ci-dessus, on voit que la notion de chaîne de Markov apparaît comme la généralisation stochastique la plus simple d'un système dynamique discret : au lieu de composer de manière répétée une fonction donnée avec elle-même, on compose des fonctions aléatoires indépendantes possédant toujours la même loi. Une propriété intéressante de ce type de construction par fonctions itérées est de définir un flot aléatoire, et, plus spécifiquement, d'autoriser la définition simultanée, sur le même espace de probabilité, de plusieurs trajectoires d'une chaîne de Markov en partant d'états initiaux différents. Nous reviendrons sur ce point lorsque nous discuterons de manière plus générale les liens entre chaînes de Markov et couplage.

Exercice 10 (Marche aléatoire sur un graphe orienté) Etant donné un graphe orienté $G = (V, E)$ fini ou dénombrable (V désigne l'ensemble de sommets, E l'ensemble des arêtes orientées, i.e. un sous-ensemble de $V \times V$), et une application $w : E \rightarrow \mathbb{R}_+$ telle que, pour tout $v_1 \in V$, $0 < \sum_{v_2; (v_1, v_2) \in E} w(v_1, v_2) < +\infty$, la marche aléatoire sur le graphe G associée aux poids w est définie par une loi initiale arbitraire, et les probabilités de transition suivantes : pour tous $v_1, v_2 \in V$ tels que $(v_1, v_2) \in E$,

$$p(v_1, v_2) := w(v_1, v_2) \left(\sum_{v_3; (v_1, v_3) \in E} w(v_1, v_3) \right)^{-1}.$$

- 1) Vérifier que l'on définit bien ainsi un noyau de transition.
- 2) Vérifier que toute chaîne de Markov utilisant le même noyau pour chacun de ses pas (ce que nous appellerons une chaîne de Markov homogène), peut se représenter sous cette forme.

Exercice 11 (Marche aléatoire sur un graphe non-orienté) Etant donné un graphe non-orienté $G = (V, E)$ fini ou dénombrable (V désigne l'ensemble de sommets, E l'ensemble des arêtes non-orientées, i.e. un sous-ensemble de $V \times V$ quotienté par la relation d'équivalence identifiant (x, y) et (y, x) ; nous noterons $\{x, y\}$ l'arête non-orientée ayant x et y pour extrémités), et une application $w : E \rightarrow \mathbb{R}_+$ telle que, pour tout $v \in V$, pour tout $v_1 \in V$, $0 < \sum_{v_2; \{v_1, v_2\} \in E} w(\{v_1, v_2\}) < +\infty$, la marche aléatoire sur le graphe G associée aux poids w est définie par une loi initiale arbitraire,

et les probabilités de transition suivantes : pour tous $v_1, v_2 \in V$ tels que $\{v_1, v_2\} \in E$,

$$p(v_1, v_2) := w(\{v_1, v_2\}) \left(\sum_{v_3; \{v_1, v_3\} \in E} w(\{v_1, v_3\}) \right)^{-1}.$$

- 1) Vérifier que l'on définit bien ainsi un noyau de transition.
- 2) Obtient-on une notion différente en considérant une pondération affectée aux sommets plutôt qu'aux arêtes ?
- 3) Préciser en quoi cette notion diffère de celle définie dans l'exercice précédent.

Un exemple particulièrement simple de marche aléatoire est celui où toutes les arêtes du graphe se voient accorder un poids égal (ce qui suppose qu'un sommet ne puisse être relié qu'à un nombre fini d'autres sommets). On parle alors souvent de "la" marche aléatoire sur le graphe sans préciser les poids.

Exercice 12 (Marche aléatoire sur un groupe) *Etant donné un groupe G fini ou dénombrable, et une famille de variables aléatoires $(g_i)_{i \geq 0}$ i.i.d. à valeurs dans G , on définit, pour toute variable aléatoire X_0 à valeur dans G , les variables aléatoires X_i pour $i \geq 1$ par $X_i := g_{i-1} \cdots g_0 X_0$.*

- 1) Montrer que la suite $(X_i)_{i \geq 0}$ est une chaîne de Markov sur G . Quel est le noyau de transition correspondant ? Même question en définissant $X_i := X_0 g_0 \cdots g_{i-1}$.
- 2) Cette chaîne de Markov est-elle nécessairement une marche aléatoire sur un graphe non-orienté au sens de l'exercice précédent ?
- 3) Inversement une marche aléatoire définie sur le graphe de Cayley de G (associé à un jeu donné de générateurs) non-orienté, est-elle toujours une marche aléatoire sur le groupe au sens du présent exercice ?

Un exemple canonique de marche aléatoire sur un groupe est la marche aléatoire simple symétrique sur \mathbb{Z}^d , définie par $P(g_0 = \pm e_i) := 1/(2d)$, où (e_1, \dots, e_d) désigne la base canonique de \mathbb{Z}^d et $B := \{+e_1, \dots, +e_d, -e_1, \dots, -e_d\}$.

Exercice 13 (Marche aléatoire en environnement aléatoire sur \mathbb{Z}^d) *Pour $d \geq 1$, soit \mathcal{M} l'ensemble des mesures de probabilité sur $B := \{+e_1, \dots, +e_d, -e_1, \dots, -e_d\}$. Pour tout élément $\mathbf{w} = (w_x)_{x \in \mathbb{Z}^d} \in \mathcal{M}^{\mathbb{Z}^d}$ (appelé environnement), on définit un noyau de transition $p(\mathbf{w})(\cdot, \cdot)$ sur \mathbb{Z}^d par $p(\mathbf{w})(x, x \pm e_i) = w_x(\pm e_i)$, pour $1 \leq i \leq d$. A présent, on considère une famille de variables aléatoires i.i.d. $\mathbf{T} = (T_x)_{x \in \mathbb{Z}^d}$ à valeurs dans \mathcal{M} , et une suite de variables aléatoires $(X_n)_{n \geq 0}$ à valeurs dans \mathbb{Z}^d telle que, conditionnellement à \mathbf{T} , $(X_n)_{n \geq 0}$ est une chaîne de Markov admettant $p(\mathbf{T})$ pour famille de noyaux de transition, initialisée en $X_0 = 0$.*

- 1) Montrer que $(X_n)_{n \geq 0}$ n'est pas une chaîne de Markov en général.

2) Montrer que la suite de variables aléatoires définie par $\mathbf{T}_n := (T_{X_n+x})_{x \in \mathbb{Z}^d}$ est une chaîne de Markov. On appelle $(\mathbf{T}_n)_{n \geq 0}$ l'environnement vu de la marche. (Pour simplifier les problèmes de mesurabilité, on pourra supposer que T_x ne peut prendre qu'un nombre fini de valeurs distinctes.)

Exercice 14 (Marche aléatoire renforcée par arêtes) Soit $G = (V, E)$ un graphe (nous donnons ici le cas d'un graphe orienté, il suffit de changer les (v_1, v_2) en $\{v_1, v_2\}$ dans les formules ci-dessous pour obtenir le cas non-orienté), fini ou dénombrable, $\Delta \geq 0$ un paramètre fixé, et $w_0 : E \rightarrow \mathbb{R}_+$ un jeu de poids comme dans l'exercice 11. On spécifie par récurrence la loi d'une suite de variables aléatoires $(X_n)_{n \geq 0}$ à valeurs dans V en supposant la loi de X_0 donnée et en posant, pour toute arête $e \in E$,

$$w_n(e) := w_0 + \Delta \sum_{i=0}^{n-1} \mathbf{1}(e = (X_i, X_{i+1})),$$

et

$$P(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) := w_n(x_n, x_{n+1}) \left(\sum_{e \ni x} w_n(e) \right)^{-1}.$$

- 1) Montrer que $(X_n)_{n \geq 0}$ n'est pas une chaîne de Markov en général.
- 2) Montrer qu'en revanche la suite $(X_n, w_n)_{n \geq 0}$ est une chaîne de Markov.

Exercice 15 (Processus de Galton-Watson) On considère le processus généalogique suivant : la première génération (numérotée 0) est constituée d'un nombre donné $p \geq 1$ d'individus. A chaque génération, chaque individu appartenant à cette génération donne lieu à un nombre aléatoire d'enfants, qui appartiennent à la génération suivante, la règle étant que les nombres d'enfants obtenus par les différents individus aux cours des différentes générations sont des variables aléatoires i.i.d. Montrer que la suite de variables aléatoires $(X_n)_{n \geq 0}$ définie par $X_n :=$ nombre d'individus présents dans la génération numéro n , constitue une chaîne de Markov.

Exercice 16 (Urne d'Ehrenfest) On considère un récipient contenant un nombre total de $N \geq 1$ particules, divisé en deux cavités, numérotées 1 et 2, qui communiquent. A chaque pas de temps, une particule parmi les N est choisie uniformément au hasard, et passe de la cavité où elle se trouve à l'autre. Montrer que, en définissant $X_n :=$ nombre de particules dans la cavité 1, la suite $(X_n)_{n \geq 1}$ forme une chaîne de Markov. Quel est son noyau de transition ?

Exercice 17 (Mutation-Sélection) Etant donné un espace fini ou dénombrable S , un noyau de transition $p(\cdot, \cdot)$ sur S , et une fonction $f : S \rightarrow \mathbb{R}_+^*$, un entier $M \geq 1$ et un entier $m \geq 1$. on définit un noyau de transition q sur S^M de la manière suivante.

Pour tout $1 \leq i \leq M$, on fabrique indépendamment m réalisations indépendantes Y_i^j , $1 \leq j \leq m$ de la loi $p(x_i, \cdot)$. Ensuite, on effectue M tirages indépendants selon la loi de probabilité

$$\frac{\sum_{i,j} f(Y_i^j) \delta_{Y_i^j}}{\sum_{i,j} f(Y_i^j)},$$

et l'on note (Z_1, \dots, Z_M) le M -uplet ainsi obtenu. Le noyau de transition q est alors défini par $q((x_1, \dots, x_M), \cdot) :=$ loi de (Z_1, \dots, Z_M) .

- 1) Donner une expression plus explicite du noyau q .
- 2) On appelle q un noyau de mutation-sélection. Comment expliquer ce terme ? Proposer d'autres types de sélection basés sur la fonction f .

Exercice 18 (Dynamique de Metropolis pour le modèle d'Ising)

Soient M et d des entiers ≥ 1 , et

$$S := \{+1, -1\}^{\{-M, \dots, 0, \dots, +M\}^d}.$$

On définit sur S la fonction H par

$$H(\sigma) = - \sum_{x \sim y \in \{-M, \dots, 0, \dots, +M\}^d} \sigma(x) \sigma(y),$$

où $x \sim y$ signifie que x et y sont voisins au sens du graphe \mathbb{Z}^d , i.e. $\sum_{i=1}^d |x_i - y_i| = 1$.

On définit ensuite un mécanisme de transition sur S de la manière suivante : on choisit un état initial $\sigma_0 \in S$ arbitraire, et la règle pour passer de σ_n à σ_{n+1} est la suivante : on choisit $x \in \{-M, \dots, 0, \dots, +M\}^d$ selon la loi uniforme, et l'on définit σ_n^x par $\sigma_n^x(y) := \sigma_n(y)$ pour $y \neq x$, et $\sigma_n^x(x) := -\sigma_n(x)$. Si $H(\sigma_n^x) \leq H(\sigma_n)$, le nouvel état σ_{n+1} est σ_n^x . Dans le cas contraire, le nouvel état est σ_n avec une probabilité égale à $\exp(-\beta(H(\sigma_n^x) - H(\sigma_n)))$, et reste égal à σ_n sinon, où $\beta > 0$ est un paramètre fixé.

- 1) Vérifier que l'on définit ainsi une chaîne de Markov.
- 2) Comment se comporte la chaîne lorsque $\beta \rightarrow 0$ et $\beta \rightarrow +\infty$?

La chaîne de Markov définie dans l'exercice ci-dessus fournit un modèle très simplifié de la dynamique d'un matériau ferromagnétique, dans lequel des spins qui peuvent valoir $+1$ ou -1 sont disposés aux sommets d'un réseau régulier, l'énergie attachée à une configuration de spins favorisant l'alignement de spins voisins. Elle fournit également un moyen de simulation pour le comportement à l'équilibre de ce type de système, comme nous le verrons ultérieurement, dans un cadre plus général.

Exercice 19 (PageRank) Soit S l'ensemble des pages ouaibe existant à un instant donné. Soit $N := \text{card } S$. Pour une page $x \in S$ donnée, soit k_x le nombre de pages vers lesquelles x contient des liens. Si x contient un lien vers y , on pose $p(x, y) :=$

$\alpha/k_x + (1 - \alpha)/N$, tandis que, si x ne contient pas de lien vers y , on pose $p(x, y) = (1 - \alpha)/N$.

- 1) Montrer que l'on définit bien ainsi un noyau de transition sur S .
- 2) Donner une interprétation probabiliste simple de la forme du noyau.

1.3 Mesure sur l'espace des trajectoires

La propriété (1.1) et les définitions données dans les propositions 1 et 2 ne font intervenir que des tronçons de longueur finie de la trajectoire infinie formée par la suite $(X_n)_{n \geq 0}$. Il est intéressant de pouvoir étudier en tant que telle cette trajectoire aléatoire. Etant donné un ensemble fini ou dénombrable S , munissons donc S de la tribu \mathcal{H} comprenant toutes les parties de S , puis munissons $S^{\mathbb{N}}$ de la tribu produit $\mathcal{H}^{\otimes \mathbb{N}}$ correspondante. Nous obtenons l'espace d'états dans lequel vivent les trajectoires de la chaîne $(X_n)_{n \geq 0}$.

Par un argument d'extension standard, on obtient la généralisation suivante de la proposition 1 (la proposition 2 se généralise de manière identique) :

Proposition 3 *Etant donné une chaîne de Markov $(X_n)_{n \geq 0}$ sur un ensemble fini ou dénombrable S , définie sur un espace de probabilité (Ω, \mathcal{F}, P) , pour tout $m \geq 0$, tout x_0, \dots, x_m tel que $\mathbb{P}(X_{0:m} = x_{0:m}) > 0$, et tout $A \in \mathcal{H}^{\otimes \mathbb{N}}$,*

$$P(X_{m+1:\infty} \in A | X_{0:m} = x_{0:m}) = P(X_{m+1:\infty} \in A | X_m = x_m).$$

Exercice 20 *Ecrire proprement l'argument d'extension nécessaire pour prouver la propriété ci-dessus.*

Nous allons maintenant caractériser la loi d'une chaîne de Markov en tant que probabilité sur $(S^{\mathbb{N}}, \mathcal{H}^{\otimes \mathbb{N}})$ en étendant l'équation (1.1) à des trajectoires infinies. Partons donc d'une loi de probabilité ν sur S , et d'une famille de noyaux de transitions $\mathbf{p} = (p_k(\cdot, \cdot))_{k \geq 0}$ sur S . En vertu, par exemple, du théorème d'extension de Kolmogorov (voir par exemple [18, 49]), il existe alors une unique mesure de probabilité $\mathbb{P}_{\nu, \mathbf{p}}$ sur $(S^{\mathbb{N}}, \mathcal{H}^{\otimes \mathbb{N}})$ vérifiant pour tout $n \geq 0$ l'identité

$$\mathbb{P}_{\nu, \mathbf{p}} \left(\{x_0\} \times \dots \times \{x_n\} \times S^{\{n+1, \dots\}} \right) = \nu(x_0) \prod_{i=0}^{n-1} p_i(x_i, x_{i+1}).$$

Dans ce cadre, le pendant de la définition (1.1) est alors le suivant :

Proposition 4 *Etant donné un ensemble fini ou dénombrable S , une suite de variables aléatoires $(X_n)_{n \geq 0}$ définie sur un espace de probabilité (Ω, \mathcal{F}, P) , $(X_n)_{n \geq 0}$ est une chaîne de Markov de loi initiale ν et de famille de noyaux de transition*

$\mathbf{p} = (p_k(\cdot, \cdot))_{k \geq 0}$ si et seulement si la loi de $(X_n)_{n \geq 0}$, vue comme variable aléatoire définie sur (Ω, \mathcal{F}, P) et à valeurs dans $(S^{\mathbb{N}}, \mathcal{H}^{\otimes \mathbb{N}})$, est égale à $\mathbb{P}_{\nu, \mathbf{p}}$. En d'autres termes, si, pour tout $A \in \mathcal{H}^{\otimes \mathbb{N}}$,

$$P(X_{0:\infty} \in A) = \mathbb{P}_{\nu, \mathbf{p}}(A).$$

Nous utiliserons la notation $\mathbb{E}_{x, \mathbf{p}}$ pour désigner l'espérance relative à $\mathbb{P}_{\nu, \mathbf{p}}$. Dans le cas homogène, nous noterons $\mathbb{P}_{\nu, p}$ plutôt que $\mathbb{P}_{\nu, \mathbf{p}}$, et nous omettrons même souvent de mentionner la dépendance de la loi de la chaîne vis-à-vis de p , pour ne conserver que celle vis-à-vis du point de départ, c'est-à-dire que nous emploierons la notation \mathbb{P}_{ν} en lieu et place de $\mathbb{P}_{\nu, \mathbf{p}}$. La même remarque vaut pour les espérances par rapport à $\mathbb{P}_{\nu, \mathbf{p}}$, que nous noterons plutôt $\mathbb{E}_{\nu, p}$ et \mathbb{E}_{ν} .

La proposition ci-dessus illustre le fait que la propriété d'une famille de variables aléatoires $(X_n)_{n \geq 0}$ d'être une chaîne de Markov (par rapport à sa filtration naturelle) ne dépend que de la loi de probabilité de cette suite de variables aléatoires. Il est parfois très utile, pour étudier les propriétés de la trajectoire $(X_n)_{n \geq 0}$, de se ramener à l'espace-image des trajectoires $(S^{\mathbb{N}}, \mathcal{H}^{\otimes \mathbb{N}}, \mathbb{P}_{\nu, \mathbf{p}})$, que l'on appelle parfois l'espace canonique associé à la chaîne de Markov.

Introduisons donc quelques notations supplémentaires relatives à cet espace. Dans le cas où ν est la mesure de Dirac $\nu = \delta_x$, nous emploierons la notation $\mathbb{P}_{x, \mathbf{p}}$ de préférence à $\mathbb{P}_{\delta_x, \mathbf{p}}$ (et donc $\mathbb{P}_{x, p}$ ou \mathbb{P}_x dans le cas homogène). De même, l'espérance sera notée $\mathbb{E}_{x, \mathbf{p}}$ (et $\mathbb{E}_{x, p}$ ou \mathbb{E}_x dans le cas homogène). On note au passage que la probabilité $\mathbb{P}_{\nu, \mathbf{p}}$ peut s'écrire comme un mélange de telles probabilités associées à des lois initiales concentrées en un point :

$$\mathbb{P}_{\nu, \mathbf{p}} = \sum_{x \in S} \nu(x) \mathbb{P}_{x, \mathbf{p}}.$$

Lorsque nous travaillerons avec l'espace canonique, nous utiliserons encore les notations X_n pour désigner les variables aléatoires associées aux coordonnées successives de la trajectoire, autrement dit, étant donné $(x_0, x_1, \dots) \in S^{\mathbb{N}}$, nous poserons $X_n(x_{0:\infty}) = x_n$ pour tout $n \geq 0$. (Attention, les expressions faisant intervenir simultanément l'espace (Ω, \mathcal{F}) et l'espace $(S^{\mathbb{N}}, \mathcal{H}^{\otimes \mathbb{N}})$ pourront donc se révéler légèrement ambiguës car les notations X_n et \mathcal{F}_n n'auront pas la même signification selon qu'elles portent sur le premier ou le deuxième espace.)

Les opérateurs de décalage θ_n sur $S^{\mathbb{N}}$ sont définis pour tout $n \geq 0$ par

$$\theta_n(x_0, x_1, \dots) = (x_n, x_{n+1}, \dots).$$

Nous utiliserons la même notation pour le décalage correspondant effectué sur les noyaux, autrement dit, $\theta_n(\mathbf{p})$ est la suite de noyaux $(p_{n+k})_{k \geq 0}$. Pour tout $n \geq 0$, et conformément aux notations définies précédemment pour une chaîne de Markov

générale, nous noterons \mathcal{F}_n la sous-tribu de $\mathcal{H}^{\otimes \mathbb{N}}$ engendrée par les variables X_k , $0 \leq k \leq n$.

On vérifie que, sur l'espace probabilisé $(S^{\mathbb{N}}, \mathcal{H}^{\otimes \mathbb{N}}, \mathbb{P}_{\nu, \mathbf{p}})$, la suite de variables aléatoires $(X_n)_{n \geq 0}$, à valeurs dans S , est bien une chaîne de Markov de loi initiale ν et de famille de noyaux de transition $\mathbf{p} = (p_k)_{k \geq 0}$, ce qui prouve en particulier que, pour toute loi ν et toute famille de noyaux de transition \mathbf{p} , il existe une chaîne de Markov qui leur est associée.

La propriété de Markov peut se ré-exprimer dans ce contexte de la manière suivante (en concaténant la proposition 1 et (1.2)) :

Proposition 5 *Pour tout $m \geq 0$, et tout $A \in \mathcal{H}^{\otimes \mathbb{N}}$,*

$$\mathbb{P}_{\nu, \mathbf{p}}(X_{m:\infty} \in A | \mathcal{F}_m) = \mathbb{P}_{X_m, \theta_m(\mathbf{p})}(X_{0:\infty} \in A), \mathbb{P}_{\nu, \mathbf{p}} - p.s. \quad (1.5)$$

L'écriture de l'identité ci-dessus peut paraître un peu sibylline, car elle fait intervenir les variables X_n à la fois pour définir les événements dont on considère la probabilité, et dans l'expression de leurs probabilités. Pour se rassurer, on peut réécrire cette identité sous la forme

$$\mathbb{P}_{\nu, \mathbf{p}}(\theta_m^{-1}(A) | \mathcal{F}_m) = \mathbb{P}_{X_m, \theta_m(\mathbf{p})}(A), \mathbb{P}_{\nu, \mathbf{p}} - p.s.$$

Enfin, si l'on veut repartir d'une expression faisant intervenir l'espace (Ω, \mathcal{F}, P) , on pourra écrire, en faisant attention aux ambiguïtés de notation, que

$$P(X_{m:\infty} \in A | \mathcal{F}_m) = \mathbb{P}_{X_m, \theta_m(\mathbf{p})}(A), P - p.s.$$

1.4 Propriété de Markov forte

Une extension très importante de la propriété de Markov discutée précédemment consiste à considérer un conditionnement par un tronçon de trajectoire dont la longueur n'est pas fixée, mais de longueur aléatoire. Nous devons auparavant définir la notion fondamentale suivante :

Définition 1 *Etant donnée une suite de variables aléatoires $(X_n)_{n \geq 0}$ définies sur un espace de probabilité (Ω, \mathcal{F}, P) et à valeur dans un ensemble fini ou dénombrable S , une variable aléatoire T définie sur (Ω, \mathcal{F}, P) à valeurs dans $\mathbb{N} \cup \{+\infty\}$ est appelée un temps d'arrêt de $(X_n)_{n \geq 0}$ lorsque, pour tout $n \in \mathbb{N}$, l'événement $\{T = n\}$ s'exprime en fonction de X_0, \dots, X_n , ou, en termes plus précis, $\{T = n\} \in \mathcal{F}_n = \sigma(X_0, \dots, X_n)$.*

Un exemple fondamental de temps d'arrêt est fourni par les temps d'atteinte, par exemple définis par $T := \inf\{n \geq 0; X_n \in B\}$, où $B \subset S$.

Exercice 21 *Montrer que les temps d'atteinte définis ci-dessus sont des temps d'arrêt. Donner d'autres exemples de variables aléatoires à valeurs dans $\mathbb{N} \cup \{+\infty\}$ qui ne sont pas des temps d'arrêt.*

Proposition 6 *Etant donnée une chaîne de Markov $(X_n)_{n \geq 0}$, définie sur (Ω, \mathcal{F}, P) , à valeurs dans un ensemble fini ou dénombrable S , de loi initiale ν et de famille de noyaux de transitions $\mathbf{p} = (p_n)_{n \geq 0}$, et un temps d'arrêt T de la suite $(X_n)_{n \geq 0}$, la propriété suivante est vérifiée : pour tout $m \geq 0$ et tout $A \in \mathcal{H}^{\otimes \mathbb{N}}$, et toute suite x_0, \dots, x_m d'éléments de S tels que $P(X_{0:T} = x_{0:T}, T = m) > 0$,*

$$P(X_{T:\infty} \in A | X_{0:T} = x_{0:T}, T = m) = \mathbb{P}_{x_m, \theta_m(\mathbf{p})}(A). \quad (1.6)$$

La propriété énoncée dans la proposition est appelée propriété de Markov forte, par opposition à la propriété discutée jusqu'à présent, que l'on peut rebaptiser propriété de Markov simple. «Sachant toutes les valeurs passées jusqu'au temps T (et en particulier la valeur de X_T), la chaîne se comporte à partir du temps T comme une nouvelle chaîne de Markov, issue de X_T .» On voit bien que, du fait que T est supposé être un temps d'arrêt, l'événement $\{T = m\}$ s'exprime en termes des variables X_0, \dots, X_m , et que (1.6) est donc une conséquence de la propriété de Markov usuelle.

En introduisant la tribu \mathcal{F}_T sur Ω définie comme l'ensemble des événements $C \in \mathcal{F}$ tels que $C \cap \{T = n\} \in \mathcal{F}_n$ pour tout $n \geq 0$, on peut réécrire l'équation (1.6) sous la forme suivante : sur l'événement $\{T < +\infty\}$, on a, pour tout $A \in \mathcal{H}^{\otimes \mathbb{N}}$, l'identité

$$P(X_{T:\infty} \in A | \mathcal{F}_T) = \mathbb{P}_{X_T, \theta_T(\mathbf{p})}(A), P - p.s. \quad (1.7)$$

Exercice 22 *(Propriété de Markov forte)*

- 1) *Prouver la propriété de Markov forte énoncée dans la proposition ci-dessus.*
- 2) *Vérifier l'équivalence des formulations (1.6) et (1.7). En particulier, montrer que T est mesurable par rapport à \mathcal{F}_T .*

Exercice 23 *Vérifier qu'avec notre définition, tout temps d'arrêt T de $(X_n)_{n \geq 0}$ se factorise nécessairement sous la forme $T = \hat{T}((X_n)_{n \geq 0})$, où \hat{T} est une variable aléatoire définie sur $(S^{\mathbb{N}}, \mathcal{H}^{\otimes \mathbb{N}})$ constituant un temps d'arrêt pour la chaîne de Markov formée par la succession des coordonnées sur l'espace canonique (que nous notons également $(X_n)_{n \geq 0}$ en général, mais cela introduirait trop d'ambiguïté d'utiliser cette notation ici).*

Notons que la notion de temps d'arrêt peut-être généralisée de la manière suivante : étant donnée une famille croissante de tribus $(\mathcal{G}_n)_{n \geq 0}$, on dit que T est un temps d'arrêt par rapport à $(\mathcal{G}_n)_{n \geq 0}$ lorsque pour tout $n \geq 0$, $\{T = n\} \in \mathcal{G}_n$. Si, en outre, $\mathcal{F}_n \subset \mathcal{G}_n$ pour tout $n \geq 0$, et si $(X_n)_{n \geq 0}$ vérifie la propriété de Markov

par rapport à $(\mathcal{G}_n)_{n \geq 0}$, la propriété de Markov forte est encore vérifiée. Un exemple simple de cette situation est le cas où, pour tout n , l'événement $T = n$ s'exprime en fonction de X_0, \dots, X_n et d'autres variables aléatoires, indépendantes de la suite $(X_i)_{i \geq 0}$. En revanche, la conclusion de l'exercice ci-dessus n'est plus vérifiée dans ce cadre général.

1.5 Décomposition d'une trajectoire selon les retours successifs en un point

Etant donnée une chaîne de Markov homogène $(X_n)_{n \geq 0}$ définie sur (Ω, \mathcal{F}, P) et à valeurs dans un ensemble fini ou dénombrable S , de noyau de transition p , définissons la variable aléatoire $T_1(a) = \inf\{n \geq 1; X_n = a\}$ (noter que l'on part de $n = 1$ et non pas de $n = 0$), avec la convention $\inf \emptyset = +\infty$. Plus généralement, définissons par récurrence pour tout entier i , $T_{i+1}(a) = \inf\{n \geq T_i(a) + 1; X_n = a\}$.

On vérifie que les variables aléatoires $T_i(a)$ sont des temps d'arrêt de $(X_n)_{n \geq 0}$. En appliquant la propriété forte de Markov, on obtient le résultat suivant.

Proposition 7 *Conditionnellement au fait que $T_i(a) < +\infty$ et à $X_0, \dots, X_{T_i(a)}$, la loi de $(X_t)_{T_i(a) \leq t < T_{i+1}(a)}$ est identique à la loi (sur l'espace canonique) de $(X_t)_{0 \leq t < T_1(a)}$ par rapport à la probabilité \mathbb{P}_a .*

Remarque 1 *Le conditionnement utilisé dans l'énoncé de la proposition peut se lire de deux manières équivalentes : comme dans l'énoncé de la propriété (1.6), en conditionnant par la valeur exacte prise par $T_1(a)$ et par toute la trajectoire $X_0, \dots, X_{T_1(a)}$, ou en termes de la tribu $\sigma(X_0, \dots, X_{T_1(a)})$, comme dans l'énoncé de la propriété (1.7).*

Nous obtenons donc la décomposition d'une trajectoire de la chaîne en tronçons de trajectoire suivant les retours successifs en a , que l'on peut résumer de la manière suivante. On commence par le tronçon $(X_t)_{0 \leq t < T_1(a)}$. Si ce tronçon est infini (c'est-à-dire si la chaîne n'atteint jamais a après son point de départ), la décomposition est terminée. Si ce tronçon est fini, on lui ajoute un tronçon de trajectoire possédant la loi de $(X_t)_{0 \leq t < T_1(a)}$ par rapport à la probabilité \mathbb{P}_a , et l'on continue d'ajouter de manière i.i.d. de tels tronçons jusqu'à obtenir éventuellement un tronçon infini, auquel cas la décomposition s'arrête. Si la décomposition ne s'arrête jamais, la trajectoire est constituée, en plus de son tronçon initial, d'une suite infinie i.i.d. de tronçons de trajectoire de longueurs finies.

Définition 2 *Posons $N(a) = \sum_{i=1}^{+\infty} \mathbf{1}(X_i = a)$ (le nombre de visites au point a au cours d'une trajectoire, sans inclure le point de départ).*

En appliquant de manière répétée la proposition ci-dessus, on obtient le résultat suivant, qui explicite la loi de $N(a)$.

Proposition 8 *La loi de $N(a)$ est caractérisée de la manière suivante :*

- Si $\mathbb{P}_a(T_1(a) < +\infty) < 1$, on a, pour tout $k \geq 1$

$$P(N(a) = k) = P(T_1(a) < +\infty) \mathbb{P}_a(T_1(a) < +\infty)^{k-1} (1 - \mathbb{P}_a(T_1(a) < +\infty)).$$

Bien entendu, $P(N(a) = 0) = 1 - P(T_1(a) < +\infty)$. Par conséquent, $N(a)$ est presque sûrement fini, et l'on a en particulier

$$E(N(a)) = P(T_1(a) < +\infty) (1 - \mathbb{P}_a(T_1(a) < +\infty))^{-1}.$$

- Si $\mathbb{P}_a(T_1(a) < +\infty) = 1$, $N(a)$ vaut 0 avec probabilité $1 - P(T_1(a) < +\infty)$, et $+\infty$ avec probabilité $P(T_1(a) < +\infty)$.

On peut préciser un peu la décomposition d'une trajectoire décrite précédemment. Avec probabilité $P(T_1(a) < +\infty)$, la trajectoire revient en a . S'y greffent alors, selon que $\mathbb{P}_a(T_1(a) < +\infty) = 1$ ou que $\mathbb{P}_a(T_1(a) < +\infty) < 1$, une infinité de tronçons i.i.d. possédant la loi de $(X_t)_{0 \leq t < T_1(a)}$ par rapport à la probabilité \mathbb{P}_a , ou un nombre fini $N(a) - 1$ de tronçons i.i.d. possédant la loi de $(X_t)_{0 \leq t < T_1(a)}$ par rapport à la probabilité $\mathbb{P}_a(\cdot | T_1(a) < +\infty)$, suivi d'une trajectoire infinie possédant la loi de $(X_t)_{0 \leq t < T_1(a)}$ par rapport à la probabilité $\mathbb{P}_a(\cdot | T_1(a) = +\infty)$.

Pour formaliser précisément la discussion, le lemme élémentaire suivant est utile :

Lemme 2 *Soit un espace mesurable (A, \mathcal{A}) et une loi μ sur A . Considérons l'espace C formé par les suites finies d'éléments de A de longueur supérieure ou égale à 1, c'est-à-dire la réunion disjointe*

$$C := \bigsqcup_{n \in \mathbb{N} \cup \{+\infty\}} \{n\} \times A^n.$$

On munit C de la tribu engendrée par les événements de la forme $\{n\} \times D$, où $D \in \mathcal{A}^{\otimes n}$, pour n décrivant $\mathbb{N} \cup \{+\infty\}$. Supposons que A est partitionné en deux parties mesurables A_1 et A_2 , et considérons à présent une variable aléatoire $(N, (Z_i)_{1 \leq i < N+1})$ à valeurs dans C vérifiant les deux propriétés suivantes :

- $N = \inf\{i \geq 1; X_i \in A_2\}$ (avec la convention $\inf \emptyset = +\infty$);
- pour tout $k \geq 1$, conditionnellement au fait que $N > k$, la variable aléatoire Z_{k+1} est indépendante de (Z_1, \dots, Z_k) et suit la loi μ .

Si $\mu(A_2) = 0$, on a que $N = +\infty$ p.s. et que, conditionnellement à l'événement $N = +\infty$, la suite $(Z_n)_{n \geq 0}$ est i.i.d. de loi μ . Dans le cas où $0 < \mu(A_2) < 1$, on a les propriétés suivantes :

- la loi de N est géométrique de paramètre $\mu(A_2)$;

- pour tout $k \in \mathbb{N}$, conditionnellement à $N = k$, les variables (Z_1, \dots, Z_k) sont indépendantes;
- pour tout $k \in \mathbb{N}$, conditionnellement à $N = k$, la loi de Z_i est $\mu(\cdot|A_1)$ si $1 \leq i < k$ et $\mu(\cdot|A_2)$ si $i = k$.

Exercice 24 Prouver le lemme ci-dessus.

Exercice 25 Appliquer le lemme ci-dessus pour vérifier la validité de la discussion qui le précède.

1.6 Action sur les mesures, action sur les fonctions : le noyau comme opérateur linéaire

Dans toute cette partie, S désigne un ensemble fini ou dénombrable. Nous commençons par définir diverses opérations associant noyaux, fonctions, et mesures sur S . Tout d'abord, étant donnés deux noyaux de transition p et q sur S , on définit le produit pq par la formule : pour tout $x, y \in S$

$$(pq)(x, y) := \sum_{z \in S} p(x, z)q(z, y),$$

et l'on vérifie que l'on a ainsi défini un nouveau noyau de transition sur S . A présent, étant donnée une mesure positive ν sur S , et un noyau de transition p sur S , le produit νp est défini par la formule : pour tout $x \in S$

$$(\nu p)(x) = \sum_{y \in S} \nu(y)p(y, x),$$

avec la convention $0 \times +\infty = 0$, et l'on vérifie que l'on a ainsi défini une nouvelle mesure positive sur S , possédant la même masse que ν : $(\nu p)(S) = \nu(S)$. Pour une fonction positive (à valeurs dans $[0, +\infty]$) f sur S , et un noyau de transition p sur S , on définit également le produit pf , par la formule

$$(pf)(x) = \sum_{y \in S} f(y)p(x, y),$$

avec la convention $0 \times +\infty = 0$. On vérifie que l'on a ainsi défini une nouvelle fonction positive sur S , et l'on note que $\sup_{x \in S} (pf)(x) \leq \sup_{x \in S} f(x)$. Enfin, étant donnée une mesure positive ν sur S et une fonction positive f , nous noterons

$$\nu f := \sum_{x \in S} f(x)\nu(x).$$

Les définitions ci-dessus s'étendent naturellement au cas d'une mesure signée $\nu = \nu_+ - \nu_-$ (différence de deux mesures positives) et d'une fonction à valeurs réelles $f = f_+ - f_-$ (écrite comme différence de sa partie positive et négative), pourvu que les parties positives et négatives donnent chacune lieu à une valeur finie.

Proposition 9 *Les produits précédents sont associatifs :*

- si p, q, r sont trois noyaux de transition sur S , $(pq)r = p(qr)$;
- si p, q sont deux noyaux de transition sur S et ν une mesure positive sur S , $\nu(pq) = (\nu p)q$;
- si p, q sont deux noyaux de transition sur S et f une fonction positive sur S , $(pq)f = p(qf)$;
- si p est un noyau de transition sur S , ν une mesure positive, et f une fonction positive sur S , $(\nu p)f = \nu(pf)$.

Ces propriétés restent vérifiées en considérant des mesures signées et des fonctions réelles donnant lieu à des parties positives et négatives finies dans toutes les expressions.

Même s'il ne couvre pas la totalité des cas possibles, on peut néanmoins retenir le résultat suivant.

Proposition 10 *L'action d'un noyau de transition à droite (sur les fonctions) définit un opérateur linéaire continu de norme 1 de $\ell^\infty(S)$ dans lui-même. L'action d'un noyau de transition à gauche (sur les mesures) définit un opérateur linéaire continu de norme 1 de $\ell^1(S)$ dans lui-même. Les actions à gauche et à droite sont duales pour la dualité entre mesures signées de masse finie et fonctions bornées.*

Exercice 26 *Prouver les propositions ci-dessus.*

Dans le cas où S est un ensemble fini, on peut interpréter matriciellement les opérations précédentes : les noyaux s'identifient à des matrices indexées par $S \times S$, les mesures comme des vecteurs-lignes indexés par S , et les fonctions comme des vecteurs-colonnes indexés par S . Les produits précédents s'identifient alors exactement aux produits usuels entre matrices.

Dans la suite, nous utiliserons la notation p^n pour désigner le produit itéré n fois du noyau de transition p avec lui-même, avec la convention selon laquelle $p^0(x, y) = 1$ si $x = y$, et $p^0(x, y) = 0$ sinon.

L'interprétation de ces opérations en termes probabilistes est résumée par les propositions suivantes.

Proposition 11 *Considérons une loi initiale ν sur S , et une suite de noyaux de transition $\mathbf{p} = (p_n)_{n \geq 0}$. On a alors, pour tout $n \geq 0$, tout $x \in S$,*

$$\mathbb{P}_{\nu, \mathbf{p}}(X_n = x) = (\nu p_0 \cdots p_{n-1})(x).$$

De plus, pour toute fonction positive f définie sur S ,

$$\mathbb{E}_{x, \mathbf{p}}(f(X_n)) = (p_0 \cdots p_{n-1}f)(x)$$

et, plus généralement,

$$\mathbb{E}_{\nu, \mathbf{p}}(f(X_n)) = \nu p_0 \cdots p_{n-1} f.$$

On note que le résultat de la proposition est encore valable si l'on considère des fonctions réelles pour lesquelles les parties positives et négatives donnent toutes les deux une valeur finie dans les expressions considérées.

Exercice 27 *Prouver la proposition ci-dessus et l'affirmation qui la suit.*

Dans le cas homogène, on déduit en particulier de cette proposition que, pour tout $n \geq 0$, et tout $x \in S$,

$$\mathbb{P}_x(X_n = y) = p^n(x, y).$$

On en déduit la propriété suivante, qui ne fait que traduire la propriété de semi-groupe de la suite des noyaux itérés $(p^n)_{n \geq 0}$, et que l'on peut déduire directement de la propriété de Markov, connue sous le nom d'équation de Chapman-Kolmogorov :

$$\mathbb{P}_x(X_{n+m} = y) = \sum_{z \in S} \mathbb{P}_x(X_n = z) \mathbb{P}_z(X_m = y).$$

On retient notamment de ce qui précède que la loi de probabilité de l'état dans lequel se trouve une chaîne de Markov après n pas s'obtient en composant l'action successive de n opérateurs linéaires sur sa loi initiale ; de la même façon, la fonction indiquant, pour chaque état initial possible dans S , la valeur moyenne de f sur l'état de la chaîne après n pas en partant de cet état initial, s'obtient en composant l'action de n opérateurs linéaires sur f . (Attention au fait que les compositions ne s'effectuent pas dans le même ordre dans les deux cas !) C'est cette remarque fondamentale qui permet d'utiliser, pour étudier les chaînes de Markov, les techniques liées à la composition d'opérateurs linéaires (et notamment les méthodes spectrales), en particulier dans le cas homogène où l'on a affaire à l'action d'un même opérateur linéaire composé avec lui-même.

Exercice 28 *Soit p un noyau de transition, $(X_n)_{n \geq 0}$ une chaîne de Markov homogène de noyau de transition p , et $f : S \rightarrow S$ une fonction bornée. On définit une suite de variables aléatoires $(M_n)_{n \geq 0}$ par*

$$M_n := f(X_n) - f(X_0) - \sum_{k=0}^{n-1} (pf)(X_k) - f(X_k).$$

- 1) Montrer que $(M_n)_{n \geq 0}$ est une martingale par rapport à la filtration $(\mathcal{F}_n)_{n \geq 0}$.
- 2) Montrer que, réciproquement, une suite de variables aléatoires $(X_n)_{n \geq 0}$ telle que propriété ci-dessus est vérifiée, est nécessairement une chaîne de Markov homogène associée au noyau p .

1.7 Extensions de la notion de chaîne de Markov

La notion de chaîne de Markov que nous étudions dans ce cours traite de processus en **temps discret**, car elle porte sur des suites de variables aléatoires indexées par les entiers, et en **espace discret**, car les variables en questions prennent leurs valeurs dans des espaces d'états S finis ou dénombrables. De nombreuses généralisations de cette notion existent, pouvant porter sur le caractère discret du temps, de l'espace, ou la propriété de Markov elle-même. Sans beaucoup entrer dans les détails, nous donnons ci-après une petite liste des plus courantes de ces extensions.

1.7.1 Chaîne de Markov en temps discret et espace quelconque

L'hypothèse que l'espace des états de la chaîne est fini ou dénombrable est une restriction importante, et l'on peut assez facilement définir une notion de chaîne de Markov à valeur dans des espaces généraux, en tenant correctement compte des questions de mesurabilité, qui sont inexistantes dans le cas où l'espace d'états est discret.

On étend d'abord de manière convenable la notion de noyau de transition.

Définition 3 *Etant donné un espace mesurable (S, \mathcal{S}) , on appelle noyau de transition toute application $p(\cdot, \cdot)$ de $S \times \mathcal{S}$ dans \mathbb{R}_+ telle que :*

- *pour tout $A \in \mathcal{S}$, $x \mapsto p(x, A)$ est une application mesurable de (S, \mathcal{S}) dans \mathbb{R} muni de sa tribu Borélienne ;*
- *pour tout $x \in S$, $p(x, \cdot)$ est une mesure de probabilité sur (S, \mathcal{S}) .*

Ensuite, l'action d'un noyau sur les mesures et sur les fonctions se définit de manière analogue à celle décrite précédemment, de même que la mesure $\mathbb{P}_{\nu, \mathbf{p}}$ sur l'espace des trajectoires. On peut alors par exemple caractériser la propriété de Markov par l'identité (1.5), mais les autres versions de la propriété de Markov peuvent également se généraliser sans grande difficulté.

Si la définition des chaînes de Markov dans ce contexte ne pose guère de problèmes, le comportement des objets ainsi définis peut se révéler sensiblement plus complexe que dans le cas où l'espace est discret, et de nombreuses pathologies sont susceptibles d'apparaître. Dans une certaine mesure, il est possible d'adapter les méthodes et les résultats qui prévalent dans le cas des espaces discrets, en général au prix d'hypothèses de régularité supplémentaires sur le noyau de transition de la chaîne. Pour un exemple simple, consulter par exemple la discussion de la notion de chaîne de Harris dans [18].

Pour des développements beaucoup plus poussés, voir par exemple les ouvrages [36, 40, 43].

1.7.2 Chaîne de Markov en temps continu et espace discret

Même en se restreignant à des espaces d'états discret, il peut être intéressant de considérer des familles de variables aléatoires indexées par une variable de temps continue plutôt que discrète. Du point de vue des applications, on obtient ainsi des modèles plus réalistes de phénomènes se déroulant en temps réel. D'un point de vue plus mathématique, les modèles en temps continu se prêtent davantage à l'utilisation des outils du calcul différentiel, et ont également l'avantage de s'intégrer directement dans la théorie plus générale des processus à temps continu (voir la section suivante). Deux chapitres de [7] sont consacrés aux chaînes de Markov en temps continu. On peut également consulter [39], et [4] pour des développements plus poussés.

Une manière naturelle de définir la notion de chaîne de Markov en temps continu consiste à généraliser la définition (1.1) et les caractérisations de la propriété de Markov données à la suite de celle-ci. Nous ne traiterons que de la généralisation de la notion de chaîne de Markov homogène, et nous aurons besoin de la notion suivante. Une famille $(p^t)(\cdot, \cdot)_{t \in \mathbb{R}_+}$ de noyaux de transition est appelée un **semi-groupe** si elle vérifie, pour tous $s, t \in \mathbb{R}_+$, la relation

$$p^{s+t} = p^s p^t,$$

ainsi que l'identité $p^0 = I$, où I est le noyau de transition défini pour tout $x \in S$ par $I(x, x) = 1$ et $I(x, y) = 0$ pour $y \neq x$.

Une famille de variables aléatoires $(X_t)_{t \in \mathbb{R}_+}$ définies sur un espace probabilisé (Ω, \mathcal{F}, P) et à valeurs dans l'ensemble fini ou dénombrable S , sera appelée une **chaîne de Markov (homogène) en temps continu** s'il existe une loi de probabilité ν sur S et un semi-groupe $(p^t(\cdot, \cdot))_{t \in \mathbb{R}_+}$ de noyaux de transition¹ tels que, pour toute famille d'indices $0 =: t_0 < t_1 < \dots < t_n$, et toute suite x_0, \dots, x_n d'éléments de S , on ait l'identité

$$P(X_{t_0:t_n} = x_{0:n}) = \nu(x_0) \times \prod_{i=1}^n p^{t_{i+1}-t_i}(x_{i-1}, x_i).$$

De manière équivalente, on peut demander que, pour tous $m \geq 0$ et $n \geq m+1$, toute famille d'indices $0 =: t_0 < t_1 < \dots < t_n$, et toute suite x_0, \dots, x_n d'éléments de S telle que $P(X_{t_0:t_m} = x_{0:m}) > 0$,

$$P(X_{t_{m+1}:t_n} = x_{t_{m+1}:t_n} | X_{t_0:t_m} = x_{t_0:t_m}) = P(X_{t_{m+1}:t_n} = x_{t_{m+1}:t_n} | X_{t_m} = x_{t_m}),$$

¹Attention : l'indice t dans la notation $p^t(\cdot, \cdot)$ employée ici ne joue pas le même rôle que l'indice n dans la notation $p_n(\cdot, \cdot)$ utilisée dans le cas discret. Dans le cas discret, l'indice n permet de considérer des chaînes de Markov n'ayant pas le même noyau de transition à chaque pas, c'est-à-dire inhomogènes. Dans le cas présent, nous décrivons une chaîne de Markov homogène, et l'indice t doit être employé du fait que, les trajectoires étant indexées par une variable de temps continue, celles-ci ne peuvent pas être découpées en "pas" élémentaires comme dans le cas discret. En fait, il faut plutôt rapprocher la notation p^t de la notation p^n employée pour désigner la puissance n -ème du noyau p .

les autres caractérisations de la propriété de Markov données dans le cas discret pouvant être adaptées de manière similaire.

Comme dans le cas discret, ν apparaît alors comme la loi de X_0 , et $p^t(x, y)$ doit vérifier $p^t(x, y) := P(X_{s+t}(y) | X_s = x)$ lorsque $P(X_s = x) > 0$. On voit facilement d'après la définition que, dans le cas où $P(X_u = x) > 0$, on doit avoir l'équation (dite de Chapman-Kolmogorov) $p^{s+t}(x, y) = P(X_{s+t+u} = y | X_u = x) = \sum_{z \in S} p^s(x, z) p^t(z, y) = (p^s p^t)(x, y)$, ce qui explique le fait que nous avons supposé *a priori* que $(p^t)_{t \in \mathbb{R}_+}$ forme un semi-groupe. Bien entendu, comme dans le cas discret, il n'y a en réalité aucune contrainte portant sur la valeur de $p^s(x, \cdot)$ lorsque $P(X_t = x) = 0$ pour tout $t \in \mathbb{R}_+$ (on peut alors toujours supposer que, dans ce cas, $p^s(x, x) = 1$ et $p^s(x, y) = 0$ pour $y \neq x$, ce qui permet de garantir la propriété de semi-groupe).

On note que la donnée du semi-groupe $(p^t)_{t \in \mathbb{R}_+}$ et de la loi initiale ν caractérise complètement² la loi de $(X_t)_{t \in \mathbb{R}_+}$ en temps que variable aléatoire de $S^{\mathbb{R}_+}$ munie de la tribu engendrée par les coordonnées, cette loi étant uniquement déterminée par ses marginales de dimension finie. La propriété de Markov de $(X_t)_{t \in \mathbb{R}_+}$ apparaît en fait comme une propriété de cette loi, et ne peut donc servir directement à caractériser des propriétés telles que la régularité des trajectoires de la forme $t \mapsto X_t(\omega)$, qu'il est pourtant nécessaire de considérer pour obtenir une théorie satisfaisante. Nous dirons qu'une fonction de \mathbb{R}_+ dans S est à **sauts réguliers** si elle continues à droite avec des limites à gauche en tout point, et n'effectue qu'un nombre fini de sauts dans tout intervalle borné. Nous dirons qu'une famille de variables aléatoires $(X_t)_{t \in \mathbb{R}_+}$ définie sur un espace probabilisé (Ω, \mathcal{F}, P) est à sauts réguliers si, hormis pour ω dans un ensemble de probabilité nulle sous P , la fonction $t \mapsto X_t(\omega)$ est à sauts réguliers. Ensuite, un semi-groupe de noyaux de transition $(p_t)_{t \geq 0}$ sera dit **régulier** si, pour toute loi de probabilité ν sur S (ou, simplement toute loi de probabilité de la forme $\nu = \delta_x$ avec $x \in S$), il existe une chaîne de Markov à sauts réguliers de loi initiale ν et de semi-groupe $(p_t)_{t \geq 0}$.

Il est possible de développer la théorie des chaînes de Markov en temps continu à partir de l'étude des propriétés analytiques de ce semi-groupe (voir par exemple [7, 4]). Cependant, comme nous ne souhaitons donner dans ce cours, qui traite essentiellement du temps discret, qu'un bref aperçu de la notion, nous présentons plutôt une construction, essentiellement équivalente, mais s'appuyant sur les chaînes de Markov à temps discret. Nous expliquons dans la suite le lien précis entre les deux points de vue.

Plutôt que de partir d'un semi-groupe $(p^t)_{t \in \mathbb{R}_+}$ de noyaux de transition, nous supposons donc donnée une **famille de taux de sauts** $\lambda = (\lambda(x, y), x, y \in S)$,

²Inversement, on peut toujours construire une chaîne de Markov en temps continu associée à un semi-groupe et une loi initiale donnés.

les $\lambda(x, y)$ étant des nombres réels vérifiant la condition $\lambda(x, y) \geq 0$ pour tous $x, y \in S$, $x \neq y$, ainsi que la condition $\lambda(x, x) = -\sum_{y \in S} \lambda(x, y)$ (qui impose donc que la somme figurant dans le membre de droite de l'équation soit finie).

Partant de λ , on définit ensuite un noyau de transition $q(\cdot, \cdot)$ sur S de la manière suivante. Si $|\lambda(x, x)| > 0$, on pose, pour tout $y \neq x$,

$$q(x, y) := \frac{\lambda(x, y)}{\sum_{z \neq x} \lambda(x, z)},$$

et par conséquent $q(x, x) := 0$. Si $\lambda(x, x) = 0$, $q(x, x) := 1$, et, par conséquent, $q(x, y) := 0$ pour tout $y \neq x$.

On considère ensuite une chaîne de Markov $(Z_n)_{n \geq 0}$ de loi initiale ν et de noyau de transition $q(\cdot, \cdot)$, et une suite de variables aléatoires $(\tau_n)_{n \geq 0}$ vérifiant la propriété suivante : conditionnellement à $(Z_k)_{k \geq 0}$ les variables aléatoires $(\tau_n)_{n \geq 0}$ sont indépendantes, la loi (conditionnelle) de τ_n étant exponentielle de paramètre $-\lambda(Z_n, Z_n) = \sum_{y \neq X_n} \lambda(X_n, y)$, lorsque cette quantité est strictement positive, et τ_n étant pris égal à $+\infty$ dans le cas contraire³. La variable aléatoire $T_{expl.} := \sum_{n=0}^{+\infty} \tau_n$ est appelée le **temps d'explosion**. On définit ensuite, pour tout $t \in [0, T_{expl.}[$, la variable aléatoire X_t , de la manière suivante : pour tout $n \geq 0$, $X_t := Z_n$ pour tout $t \in [\tau_0 + \dots + \tau_{n-1}, \tau_0 + \dots + \tau_n[$. Lorsque $P(T_{expl.} < +\infty) = 0$, on dit qu'il n'y a **pas d'explosion en temps fini**, et l'on peut définir X_t pour tout $t \in \mathbb{R}_+$. On observe alors que, presque sûrement, la trajectoire $(X_t)_{t \in \mathbb{R}_+}$ ainsi obtenue est à sauts réguliers.

On a en fait le résultat suivant, dont la preuve utilise essentiellement la propriété d'absence de mémoire de la loi exponentielle.

Théorème 1 *Avec les notations précédentes, dans le cas où il n'y a pas d'explosion en temps fini, la famille de variables aléatoires $(X_t)_{t \in \mathbb{R}_+}$ est une chaîne de Markov en temps continu à sauts réguliers, au sens défini précédemment.*

Exercice 29 *Prouver le théorème 1.*

Remarquablement, le théorème 1 admet une réciproque, ce qui signifie que la construction ci-dessus permet en fait d'obtenir toutes les chaînes de Markov en temps continu à sauts réguliers (voir [7] pour une preuve).

³Il serait souhaitable de vérifier que l'on peut effectivement construire une telle famille de variables aléatoires $(Z_n)_{n \geq 0}$ et $(\tau_n)_{n \geq 0}$. Une construction explicite à partir de processus de Poisson sera donnée par la suite. Les anxieux peuvent tenter de donner eux-mêmes une construction élémentaire utilisant par exemple des variables aléatoires uniformes pour générer les variables τ_n à partir des variables X_n .

Théorème 2 Si $(X_t)_{t \in \mathbb{R}_+}$ est une chaîne de Markov en temps continu associée à un semi-groupe régulier de noyaux de transition $p^t(\cdot, \cdot)$, il est possible d'obtenir une chaîne de Markov à sauts réguliers de même loi à partir de la construction basée sur les taux de sauts présentée ci-dessus. On peut employer pour cette construction la famille de taux de sauts λ définie par (l'existence des limites ci-dessous, et le fait qu'elles définissent effectivement une famille de taux de sauts font partie du théorème) :

$$\lambda(x, y) = \lim_{h \rightarrow 0} \frac{p^h(x, y)}{h}, \quad x \neq y,$$

$$\lambda(x, x) = \lim_{h \rightarrow 0} \frac{(p^h(x, x) - 1)}{h}.$$

Nous allons vérifier directement, pour la construction basée sur les taux de sauts, que les limites figurant dans le théorème 2 sont bien satisfaites. Partons donc d'une chaîne de Markov ainsi construite, initialisée selon la loi δ_x . Posons $\alpha(x) := -\lambda(x, x)$, et supposons que $\alpha(x) > 0$, sans quoi le résultat annoncé est trivial. Pour $y \neq x$, on a

$$P(X_h = y) = P(Z_1 = y, \tau_1 < h, \tau_1 + \tau_2 > h) + P(X_h = y, \tau_1 < h, \tau_1 + \tau_2 < h). \quad (1.8)$$

Constatons déjà que, par définition, $P(Z_1 = y, \tau_1 < h) = q(x, y)(1 - e^{-\alpha(x)h})$, d'où le fait que $\lim_{h \rightarrow 0} h^{-1}P(Z_1 = y, \tau_1 < h) = q(x, y)\alpha(x) = \lambda(x, y)$. Au vu de (1.8), on en déduit qu'il suffit de montrer que $P(\tau_1 + \tau_2 < h) = o(h)$ lorsque h tend vers zéro pour prouver le résultat concernant $p_h(x, y)$. On peut par exemple écrire que $P(\tau_1 + \tau_2 < h) \leq P(\tau_1 < h, \tau_2 < h)$, puis, en découpant selon la valeur de Z_1 , écrire que

$$P(\tau_1 < h, \tau_2 < h) = \sum_{z \in S} P(\tau_1 < h, Z_1 = z, \tau_2 < h),$$

d'où

$$P(\tau_1 < h, \tau_2 < h) = \sum_{z \in S} q(x, z)(1 - e^{-\alpha(x)h})(1 - e^{-\alpha(z)h}).$$

Par convergence dominée, on vérifie que $\lim_{h \rightarrow 0} \sum_{z \in S} q(x, z)(1 - e^{-\alpha(z)h}) = 0$, d'où le résultat lorsque $y \neq x$. Le cas $y = x$ peut être traité de la même manière.

Dans ce contexte, la famille de taux $\lambda(x, y)$ constitue ce que l'on appelle le **générateur infinitésimal** du semi-groupe $p^t(\cdot, \cdot)$. Comme pour les noyaux de transition, on peut définir l'action du générateur à droite sur les fonctions par $(\lambda f)(x) = \sum_{y \in S} \lambda(x, y)f(y)$, et à gauche sur les mesures par $(\mu \lambda)(x) = \sum_{y \in S} \mu(y)\lambda(y, x)$. Pour une fonction ou une mesure positive ne prenant que des valeurs finies, la définition a toujours un sens car $\lambda(x, y) \geq 0$ sauf si $x = y$. On étend la définition aux fonctions et aux mesures non-nécessairement positives par le procédé habituel.

Au vu des théorèmes précédents, on déduit la proposition suivante.

Proposition 12 *Un semi-groupe régulier de noyaux de transition $(p^t)_{t \in \mathbb{R}_+}$ est entièrement déterminé par la donnée de son générateur infinitésimal.*

Exercice 30 *Prouver la proposition ci-dessus à partir des théorèmes précédents.*

En revanche, compte-tenu de la possibilité d'explosion en temps fini, une famille de taux de sauts λ telle que définie ci-dessus ne donne pas nécessairement lieu à un semi-groupe régulier dont λ serait le générateur infinitésimal. Telle que présentée ici, la construction n'est possible qu'en l'absence d'explosion (mais diverses extensions sont possibles, voir par exemple [4]).

L'exercice 34 ci-dessous donne une caractérisation des familles de taux de sauts pour lesquels il n'y a pas d'explosion en temps fini.

Exercice 31 *Donner un exemple de famille de taux de sauts pour laquelle il y a explosion en temps fini avec probabilité positive.*

Exercice 32 *Montrer que, dans le cas où S est de cardinal fini, on a toujours $P(T_{\text{expl.}} < +\infty) = 0$, et plus généralement lorsque $\sup |\lambda(x, x)| < +\infty$.*

Exercice 33 *Montrer que, sur l'événement $\{T_{\text{expl.}} < +\infty\}$, avec probabilité 1, chaque élément de S ne doit être visité qu'un nombre fini de fois par la chaîne $(Z_n)_{n \geq 0}$.*

Exercice 34 *Le but de cet exercice est d'établir la caractérisation suivante, dite critère de Reuter : étant donné une famille de taux de sauts λ , il n'y a pas d'explosion en temps fini si et seulement si, pour tout nombre réel positif $\gamma > 0$, le système d'équations*

$$\gamma + \lambda(x, x)u_x = \sum_{y \in S, y \neq x} \lambda(x, y)u_y; \quad x \in S,$$

n'admet aucune solution $u = (u_x)_{x \in S}$ positive, bornée, et distincte de $u_x \equiv 0$.

1) *On définit, pour tout $x \in S$ et tout $\gamma > 0$, $u_x(\gamma) := E(\exp(-\gamma T_{\text{expl.}}))$, où l'espérance est prise par rapport à la construction ayant pour loi initiale δ_x . Montrer que, dans le cas où $P(T_{\text{expl.}} < +\infty) > 0$, $u_x(\gamma)$ fournit la solution recherchée au système.*

2) *On suppose à présent que nous disposons d'une solution $u = (u_x)_{x \in S}$ négative, bornée, et distincte de $u_x \equiv 0$ au système, et que, par ailleurs, $P(T_{\text{expl.}} < +\infty) = 0$. Montrer alors que la famille de variables aléatoires $(\exp(-\gamma t)u_{X_t})_{t \geq 0}$ constitue une martingale (on peut se contenter de le vérifier pour les t de la forme $t = \tau_0 + \dots + \tau_n$). Conclure à une contradiction en analysant le comportement lorsque $t \rightarrow +\infty$.*

On peut réécrire le résultat final du théorème 2 sous la forme

$$\lim_{h \rightarrow 0} \frac{p^h - I}{h} = \lambda,$$

où la limite a lieu coordonnée par coordonnée pour tout $(x, y) \in S$.

En écrivant, pour $t \in \mathbb{R}_+$, que $p^{t+h} - p^t = (p^h - I)p^t = p^t(p^h - I)$, on s'attend à ce que les deux équations suivantes soient vérifiées (bien entendu, ce n'est qu'un calcul purement formel pour l'instant, car il y a de sérieux problèmes de convergence à régler) :

$$\frac{dp^t}{dt} = \lambda p^t, \quad \frac{dp^t}{dt} = p^t \lambda,$$

soit, de manière plus explicite :

$$\frac{dp^t(x, y)}{dt} = \sum_{z \in S} \lambda(x, z) p^t(z, y), \quad (1.9)$$

$$\frac{dp^t(x, y)}{dt} = \sum_{z \in S} p^t(x, z) \lambda(z, y). \quad (1.10)$$

L'équation (1.9) est appelée **équation de Kolmogorov rétrograde**, tandis que (1.10) est appelée **équation de Kolmogorov progressive**. Pour que cette équation est un sens, on doit s'assurer et de la différentiabilité des fonctions $t \mapsto p^t(x, y)$, et de la convergence des séries qui apparaissent dans les équations. Nous n'en donnerons pas la preuve (voir par exemple [4]), mais il est en fait possible de vérifier rigoureusement ces propriétés.

Théorème 3 *Etant donné un semi-groupe régulier de noyaux de transition $(p^t)_{t \in \mathbb{R}_+}$, les équations de Kolmogorov progressive et rétrograde sont satisfaites.*

Pour résumer, le générateur infinitésimal joue en quelque sorte dans le cas continu le rôle dévolu au noyau p (ou plus exactement $p - I$) dans le cas discret, et les équations de Kolmogorov, progressives ou rétrogrades, correspondent simplement au fait que p^n s'obtient par produit itéré de p avec lui-même. Nous n'approfondirons pas davantage ici le lien entre semi-groupe et générateur infinitésimal. Mentionnons simplement le fait important suivant, qui constitue en quelque sorte une forme intégrée de la définition du générateur à partir du semi-groupe : si $(X_t)_{t \geq 0}$ est une chaîne de Markov à sauts réguliers associée à un semi-groupe régulier $(p^t)_{t \in \mathbb{R}_+}$ dont λ est le générateur infinitésimal, et si f est une fonction définie sur S et à valeurs réelles, alors, sous certaines hypothèses de régularité sur f , la famille de variables aléatoires $(M_t)_{t \geq 0}$ définie par

$$M_t := f(X_t) - f(X_0) - \int_0^t (\lambda f)(X_s) ds,$$

est une martingale par rapport à la filtration $(\mathcal{F}_t)_{t \geq 0}$.

Exercice 35 *Prouver le fait ci-dessus lorsque la fonction f est bornée. Comparer avec l'exercice 28.*

Dans le cas où l'on souhaite construire la dynamique à partir d'une famille de taux de sauts, il est possible de donner une réalisation explicite de celle-ci au moyen d'une famille de processus de Poisson. Plus précisément, on supposera donnée une famille $\mathcal{P}(x, y)$, $x, y \in S$, $x \neq y$, de processus de Poisson sur \mathbb{R}_+ (vus comme des sous-ensembles aléatoires), mutuellement indépendants, chaque $\mathcal{P}(x, y)$ ayant un taux constant égal à $\lambda(x, y)$. Considérons $n \geq 0$, et supposons donnés les valeurs $Z_n = x$ et $\tau_0 + \dots + \tau_{n-1} = t$ utilisés pour la construction. On définit alors t_* comme le plus petit élément de la réunion des ensembles $\mathcal{P}(x, y) \cap]t, +\infty[$ pour $y \in S$, et x_* comme l'élément de S vérifiant $t_* \in \mathcal{P}(x, x_*)$. Compte-tenu de nos hypothèses sur λ , avec probabilité 1, t_* existe et vérifie $t_* > t$, tandis que x_* est unique. On définit alors $Z_{n+1} := x_*$ et $\tau_n := t_* - t$.

Une interprétation intuitive utile de la construction donnée ci-dessus peut se résumer ainsi : lorsque la chaîne est dans l'état x , au cours d'un intervalle de temps infinitésimal dt , la probabilité que la chaîne effectue un saut vers $y \neq x$ est donnée par $\lambda(x, y)dt$.

Exercice 36 *Prouver que la construction définie ci-dessus satisfait bien les propriétés attendues de la construction par les taux de sauts.*

Dans le cas où la famille de taux de sauts est bornée, il est possible de donner une représentation plus simple de la dynamique, ne faisant appel qu'à un seul processus de Poisson. Supposons donc donné un nombre réel $\rho > 0$ tel que $\sup_{x \in S} |\lambda(x, x)| \leq \rho$. Définissons alors le noyau de transition $r(\cdot, \cdot)$ sur S de la manière suivante : si $x \neq y$, on pose $r(x, y) := \frac{\lambda(x, y)}{\rho}$, et $r(x, x) = 1 - \sum_{y \neq x} r(x, y)$. En d'autres termes,

$$r = I + \lambda/\rho.$$

Notre hypothèse sur ρ assure que l'on définit bien ainsi un noyau de transition. La construction reprend alors les étapes de la construction générale à partir des taux de sauts, mais en modifiant les définitions de Z_n et τ_n . Plus précisément, on considère donc une chaîne de Markov $(Z_n)_{n \geq 0}$ de loi initiale ν et de noyau de transition $r(\cdot, \cdot)$, et une suite de variables aléatoires $(\tau_n)_{n \geq 0}$ i.i.d. de loi exponentielle de paramètre ρ . En posant, pour tout $t \in \mathbb{R}_+$, $N_t = \text{card}\{i; \tau_0 + \dots + \tau_{i-1} < t\}$ on obtient la représentation $X_t = Z_{N_t}$, $(N_t)_{t \geq 0}$ étant le processus de comptage associé au processus de Poisson dont les accroissements sont donnés par les variables aléatoires $(\tau_n)_{n \geq 0}$. Insistons sur le fait que N_t est indépendant de la chaîne $(Z_n)_{n \geq 0}$.

Exercice 37 *Prouver que la construction définie ci-dessus satisfait bien les propriétés attendues de la construction par les taux de sauts.*

Par rapport à la construction générale, on constate que les lois des dates auxquelles ont lieu les sauts ne dépendent pas des états dans lesquels la chaîne se trouve,

ce que l'on corrige en introduisant des sauts fictifs d'un état vers lui-même. On vérifie facilement que l'on a alors, pour tout $t \in \mathbb{R}_+$, la relation

$$p^t = \sum_{n=0}^{+\infty} r^k \frac{t \rho^k}{k} e^{-\rho t} = e^{\rho t(r-I)} = e^{t\lambda}.$$

(Cette identité a lieu coordonnée par coordonnée.)

Exercice 38 Vérifier que, dans le cas particulier ci-dessus, les équations de Kolmogorov progressive et rétrograde sont vérifiées.

Mentionnons une famille importante de chaînes de Markov en temps continu : les processus de naissance et de mort. L'espace d'états de ces processus est l'ensemble \mathbb{N} des entiers naturels, et les seuls sauts autorisés consistent à ajouter ou retrancher 1 à l'état courant. Plus précisément, on suppose que l'on a deux suites $(\alpha_i)_{i \geq 0}$ et $(\beta_i)_{i \geq 1}$, et que, pour tout $x \geq 0$, $\lambda(x, x+1) := \alpha_x$, tandis que pour tout $x \geq 1$, $\lambda(x, x-1) := \beta_x$. Les taux $\lambda(x, y)$ avec $y \in \{x-1, x, x+1\}$ sont égaux à 0.

Exercice 39 On considère une population d'individus dans laquelle, au cours de chaque intervalle de temps infinitésimal dt , chaque individu de la population peut mourir avec probabilité μdt , ou, au contraire, donner naissance à un autre individu, avec probabilité νdt , indépendamment des autres individus présents. Montrer que la population totale est décrite par un processus de naissance et de mort. Quels sont les paramètres correspondants ?

Exercice 40 On considère un processus de naissance $(X_t)_{t \geq 0}$ construit à partir des taux de sauts $\alpha_x := x$ (et pas de mort, donc $\beta_x := 0$), et vérifiant $X_0 := 1$.

- 1) Montrer qu'il n'y a pas d'explosion en temps fini.
- 2) On pose, pour tout $\gamma > 0$, $f_\gamma(t) := E[\exp(-\gamma X_t)]$. Grâce aux équations de Kolmogorov, trouver une équation différentielle satisfaite par f_γ , et la résoudre.
- 3) En déduire que, pour tout $t \geq 0$ et $n \geq 1$, $P(X_t = n) = a(1-a)^n$, où $a := \exp(-t)$.
- 4) Retrouver la loi de X_t à l'aide d'un argument purement probabiliste. (Indication : représenter la généalogie du processus, et la considérer à rebrousse-temps).
- 5) Que dire de la loi de $\exp(-t)X_t$ lorsque $t \rightarrow +\infty$?
- 6) Montrer que $(\exp(-t)X_t)_{t \geq 0}$ est une martingale. En déduire que $\lim_{t \rightarrow +\infty} \exp(-t)X_t$ existe presque sûrement.
- 7) On considère à présent non plus un, mais deux processus de naissance indépendants $(X_t)_{t \geq 0}$ et $(Y_t)_{t \geq 0}$ avec $X_0 := A$ et $Y_0 := B$. Soit $(\tau_n)_{n \geq 0}$ la suite croissante des instants des sauts de X_t et Y_t (avec $\tau_0 := 0$). Quelle est la loi de la suite $(X_{\tau_n}, Y_{\tau_n})_{n \geq 0}$?
- 8) Retrouver à partir des questions précédentes le fait que la proportion de boules rouges dans une urne de Pólya contenant initialement A boules rouges et B boules

blanches (avec $A, B \geq 1$), et dans laquelle on ajoute à chaque tirage une boule supplémentaire de la même couleur que celle qui vient d'être tirée, converge presque sûrement vers une valeur limite distribuée selon une loi Beta de paramètres (A, B) .

Exercice 41 (*Urne d'Ehrenfest en temps continu*) On reprend le contexte du modèle de l'urne d'Ehrenfest, mais en temps continu cette fois. Les N particules évoluent de manière indépendante, chaque particule sautant d'une cavité à l'autre lors d'instantants de sauts donnés par un processus de Poisson d'intensité constante associé à cette particule.

- 1) Montrer que la suite des configurations lors des instants de saut correspond au modèle de l'urne d'Ehrenfest en temps discret.
- 2) Expliciter la loi du nombre de particules se trouvant dans la cavité 1 au temps t .
- 3) En déduire une estimation sur le temps nécessaire pour passer d'une configuration équilibrée (autant de particules dans chaque cavité) à une configuration déséquilibrée, et, inversement, une estimation sur le temps nécessaire pour passer d'une configuration déséquilibrée à une configuration équilibrée.

Exercice 42 Montrer qu'un processus de naissance et de mort n'explose pas en temps fini si et seulement si

$$\sum_{n=1}^{+\infty} \left(\frac{1}{\alpha_n} + \frac{\beta_n}{\alpha_n \alpha_{n-1}} + \cdots + \frac{\beta_n \cdots \beta_1}{\alpha_n \cdots \alpha_1 \alpha_0} \right) = +\infty.$$

1.7.3 Processus de Markov en temps continu

Un saut technique très important est nécessaire pour pouvoir traiter des processus évoluant en temps continu sur des espaces qui ne sont pas discrets, et nous nous contenterons de mentionner la définition correspondante, assortie de quelques commentaires et références.

Plutôt que de définir d'abord la propriété de Markov dans un cadre général, puis de restreindre l'étude à des processus pouvant se réaliser au moyen de trajectoires suffisamment régulières, comme nous l'avons fait dans le cas temps continu/espace discret, nous supposons d'emblée que les processus étudiés possèdent des trajectoires satisfaisant les propriétés de régularité souhaitées.

Ainsi, pour définir un processus Markovien (homogène) en temps continu à valeur dans un espace S général, on suppose que S est muni d'une topologie, et de la tribu Borélienne associée. On prend pour espace des trajectoires possibles pour le processus, l'ensemble $\mathcal{D}([0, +\infty), S)$ des fonctions continues à droite avec des limites à gauche en tout point, de $[0, +\infty)$ dans S , que l'on munit de la tribu \mathcal{F} engendrée par les projections X_t , où $X_t((x_s)_{s \geq 0}) := x_t$. Un processus de Markov est alors caractérisé par la famille des lois des trajectoires obtenues en considérant tous les

point de départ possibles pour le processus, c'est-à-dire ici une famille de mesures de probabilité $(\mathbb{P}_x)_{x \in S}$ sur $\mathcal{D}([0, +\infty), S)$, vérifiant les conditions suivantes :

- $\mathbb{P}_x(X_0 = x) = 1$;
- pour tout $A \in \mathcal{F}$, $x \mapsto \mathbb{P}_x(A)$ est mesurable ;
- pour tout $x \in S$ et tout $t \geq 0$,

$$\mathbb{P}_x(\theta_t^{-1}(A) | \mathcal{F}_t) = \mathbb{P}_{X_t}(A), \mathbb{P}_x - p.s.,$$

où \mathcal{F}_t et θ_t sont définis de manière analogue au cas discret. On voit que la propriété de Markov est ici traduite par la troisième condition.

Comme dans le cas temps continu/espace discret, on peut développer une théorie décrivant les liens entre semi-groupe et générateur infinitésimal, mais celle-ci est beaucoup plus difficile et sophistiquée.

Un exemple classique de processus de Markov en temps continu et espace non-discret est le mouvement Brownien (voir par exemple [44, 28, 46]), pour lequel l'espace d'états est typiquement \mathbb{R} ou \mathbb{R}^d . Un autre exemple est constitué par les systèmes de particules en interaction (voir par exemple [32, 31]), pour lesquels l'espace est typiquement de la forme $S^{\mathbb{Z}^d}$, où S est discret.

1.7.4 Champs markoviens

La notion de champ markovien constitue une généralisation des chaînes de Markov, dans laquelle la structure uni-dimensionnelle de l'indexation des variables par \mathbb{N} est remplacée par une structure de graphe générale. Pour simplifier la discussion, nous ne considérerons que des graphes finis. Considérons donc un graphe fini $G = (V, E)$ non-orienté, et une famille de variables aléatoires à valeurs dans un ensemble fini ou dénombrable S , indexée par les sommets de G , c'est-à-dire de la forme $(X_v)_{v \in V}$. Etant donné un sous-ensemble $A \subset V$, nous noterons \mathcal{F}_A la tribu engendrée par les variables aléatoires $(X_v; v \in A)$, et ∂A la frontière de A au sens de la structure de graphe de G , c'est-à-dire l'ensemble des éléments de $V \setminus A$ reliés par une arête à un élément de A .

On dira que $(X_v)_{v \in V}$ est un champ markovien si la propriété suivante, qui généralise la propriété de Markov, est vérifiée : pour tout sous-ensemble $A \subset V$, et tout $\Gamma \subset S^A$,

$$P((X_v)_{v \in A} \in \Gamma | \mathcal{F}_{V \setminus A}) = P((X_v)_{v \in A} \in \Gamma | \mathcal{F}_{\partial A}).$$

En d'autres termes, les valeurs du champ sur l'ensemble A ne dépendent des valeurs du champ à l'extérieur de A que par l'intermédiaire des valeurs prises sur la frontière séparant A et l'extérieur de A .

Exercice 43 *Montrer que, si $(X_m)_{0 \leq m \leq n}$ est une chaîne de Markov, elle constitue également un champ Markovien associé au graphe G dont les sommets sont formés par l'ensemble $\llbracket 1, n \rrbracket$, deux entiers étant reliés par une arête s'ils diffèrent de ± 1 .*

Exercice 44 *Etant donné un graphe fini $G = (V, E)$, on dit qu'un sous-ensemble C de V constitue une clique de G si toute paire de sommets distincts de C est reliée par une arête de G . Un potentiel de Gibbs sur S^V associé à la structure de graphe de G est la donnée, pour toute clique C , d'une fonction W_C définie sur S^C et à valeurs dans $]-\infty, +\infty]$. On construit alors une fonction d'énergie associée définie sur S^V et à valeurs dans $]-\infty, +\infty]$ en posant*

$$H((x_v)_{v \in V}) := \sum_C W_C((x_v)_{v \in C}).$$

Une famille de variables aléatoires à valeurs dans S indexée par V constitue un champ de Gibbs lorsque sa loi jointe est associée à un potentiel de Gibbs par l'intermédiaire de la relation :

$$P((X_v)_{v \in V} = (x_v)_{v \in V}) = \frac{1}{Z} \exp(-\beta H((x_v)_{v \in V})),$$

où $\beta > 0$, et où Z est une constante de normalisation. Montrer que tout champ de Gibbs est un champ markovien.

En fait, la réciproque au résultat de l'exercice ci-dessus est vraie : tout champ markovien (sous réserve d'une condition de non-dégenescence) est un champ de Gibbs pour un certain potentiel, au moins dans le cas d'un graphe fini (la situation est bien plus complexe dans le cas d'un graphe infini).

Exercice 45 *Vérifier que, dans l'exercice 43, le champ markovien que l'on obtient en considérant une chaîne de Markov est bien un champ de Gibbs. Préciser le potentiel correspondant.*

Exercice 46 *Vérifier que la fonction H définie dans l'exercice 18 est un potentiel de Gibbs.*

L'ouvrage [7] contient une discussion des champs markoviens. Pour un traitement plus détaillé, voir par exemple [23] et les références qu'il contient.

1.7.5 Chaînes de Markov avec plusieurs pas de mémoire

Du point de vue de la modélisation, la propriété de Markov peut se révéler trop restrictive : au sein d'une séquence, la dépendance d'une valeur vis-à-vis de l'ensemble des valeurs passées peut s'étendre au-delà de la simple valeur précédente. Par conséquent, il est naturel de s'intéresser à des modèles incorporant des dépendances à plus longue portée. Dans le cas où l'on se restreint à une dépendance de portée bornée, on obtient la notion de chaîne de Markov d'ordre M , ou encore de chaîne de Markov avec mémoire de longueur M (la notion standard correspondant à $M = 1$).

Pour caractériser une telle chaîne de Markov, on supposera donnée une loi initiale ν sur S^M , et une famille de noyaux de transition $(p_n)_{n \geq 0}$ de S^M vers S (c'est-à-dire que, pour tout $n \geq 0$ et $x \in S^M$, $p_n(x, \cdot)$ est une probabilité sur S). L'analogie de la propriété de Markov est alors le suivant : pour toute suite x_0, \dots, x_n (avec $n + 1 \geq M$)

$$P(X_{0:n} = x_{0:n}) = \nu(x_0, \dots, x_{M-1}) \prod_{i=0}^{n-M} p_i(x_{i:i+M-1}, x_{i+M}).$$

D'un point de vue formel, cette notion se ramène en fait exactement à la notion usuelle de chaîne de Markov, comme le montre l'exercice suivant.

Exercice 47 *Etant donnée une suite de variables aléatoires $(X_n)_{n \geq 0}$, montrer qu'il y a équivalence entre les deux propriétés suivantes :*

- la suite $(X_n)_{n \geq 0}$ est une chaîne de Markov d'ordre M ;
- la suite $(X_{i:i+M-1})_{i \geq 0}$ est une chaîne de Markov sur S^{M-1} .

Préciser le lien entre les noyaux de transition dans les deux cas.

1.7.6 Semi-chaînes de Markov

Une chaîne de Markov homogène possède la propriété suivante : le nombre de pas passés par la chaîne dans un état donné suit toujours une loi géométrique. Pour se donner la possibilité de modéliser des séquences qui ne satisfont pas cette propriété, on introduit la notion de semi-chaîne de Markov, dans laquelle, partant d'une chaîne de Markov au sens usuel et vérifiant $p(x, x) = 0$ pour tout $x \in S$, on fait passer au processus, dans chaque état visité par la chaîne de Markov, un nombre de pas dont la loi est prescrite par le modèle, et dépend de cet état.

Exercice 48 *Ecrire en détail un tel modèle.*

Pour en apprendre davantage sur le sujet, voir par exemple [5].

1.7.7 Chaînes de Markov cachées

Pour définir une chaîne de Markov cachée, on se donne un espace d'états cachés S et un espace d'états observés V (que nous supposons tous les deux finis ou dénombrables), une loi initiale ν sur S , une famille de noyaux de transition $(p_k)_{k \geq 0}$ sur S , et une famille $(q_k)_{k \geq 0}$ de noyaux de transition de S vers V , c'est-à-dire que, pour tout $k \geq 0$ et $x \in S$, $q_k(x, \cdot)$ est une probabilité sur V .

Une chaîne de Markov cachée associée à ces éléments est une suite de couples de variables aléatoires $(X_n, Y_n)_{n \geq 0}$ telle que $(X_n)_{n \geq 0}$ est une chaîne de Markov associée

à la loi initiale ν et à la famille de noyaux de transition $(p_k)_{k \geq 0}$, tandis que, conditionnellement à $(X_n)_{n \geq 0}$, les variables aléatoires $(Y_k)_{k \geq 0}$ sont indépendantes, la loi de Y_k sur V étant donnée par $q_k(X_k, \cdot)$.

Comme la terminologie le suggère, ce type de modèles est employé pour décrire des suites de variables observées, dont la dépendance est en fait contrôlée par une suite markovienne de variables non-observées. Ce type de modèle est employé dans de nombreuses applications, par exemple pour le traitement du signal, la reconnaissance de la parole, ou la modélisation de séquences biologiques telles que l'ADN.

D'un point de vue théorique, la notion de chaîne de Markov cachée s'identifie en fait avec celle de chaîne de Markov partiellement observée. Les exercices suivants développent un peu ce point de vue.

Exercice 49 Dans cet exercice, on montre que les notions de chaîne de Markov cachée et de chaîne de Markov partiellement observée sont en fait équivalentes.

- 1) Etant donnée une chaîne de Markov $(X_n)_{n \geq 0}$, et une fonction $f : S \rightarrow V$, montrer que la suite $(X_n, f(X_n))$ est une chaîne de Markov cachée.
- 2) Réciproquement, montrer que, si $(X_n, Y_n)_{n \geq 0}$ est une chaîne de Markov cachée, on peut représenter Y_n sous la forme $Y_n = f(Z_n)$, où $(Z_n)_{n \geq 0}$ est une chaîne de Markov.

Exercice 50 Considérons une suite de variables aléatoires i.i.d. de Bernoulli de paramètre $1/2$, $(\varepsilon_n)_{n \geq 0}$. Définissons $X_n := (\varepsilon_n, \varepsilon_{n+1})$.

- 1) Montrer que $(X_n)_{n \geq 0}$ est une chaîne de Markov. Que peut-on dire de la suite $(X_{2n})_{n \geq 0}$?
- 2) Sur $\{0, 1\}^2$, définissons la fonction f par $f(0, 0) := 0$, $f(0, 1) = f(1, 0) := 1$, $f(1, 1) := 2$, et considérons la suite $(Y_n)_{n \geq 0}$ définie par $Y_n := f(X_n)$. Montrer que $(Y_n)_{n \geq 0}$ n'est pas une chaîne de Markov, même si l'on s'autorise à considérer des chaînes à plusieurs pas de mémoire.

Exercice 51 On considère deux ensembles finis ou dénombrables S et V , une application $f : S \rightarrow V$, et une chaîne de Markov $(X_n)_{n \geq 0}$ sur S de noyau de transition p .

- 1) Supposons que p vérifie la propriété (*) suivante : pour tous $x, y \in S$ tels que $f(x) = f(y)$, on a l'égalité $p(x, f^{-1}(v)) = p(y, f^{-1}(v))$ pour tout $v \in V$. (Nous employons la notation $p(x, A) := \sum_{z \in A} p(x, z)$.) Montrer qu'alors $(f(X_n))_{n \geq 0}$ est une chaîne de Markov. Quel est le noyau de transition correspondant ?
- 2) Montrer que, pour un noyau irréductible, il y a en fait équivalence entre la propriété (*) définie à la question précédente, et la propriété suivante : pour toute loi initiale ν sur S , $(f(X_n))_{n \geq 0}$ est une chaîne de Markov pour la probabilité $\mathbb{P}_{\nu, p}$.

Tout un ensemble de méthodes et d'algorithmes ont été développés pour l'inférence statistique dans le contexte chaînes de Markov cachées. Pour motiver l'exercice

qui suit, posons-nous la question de savoir comment calculer la probabilité d'une séquence de valeurs observées. Dans le cas d'une chaîne de Markov, la réponse est immédiate : la probabilité d'observer la suite de valeurs x_0, \dots, x_n est donnée par le produit $\nu(x_0)p_0(x_0, x_1) \cdots p_{n-1}(x_{n-1}, x_n)$. Dans le cas d'une chaîne de Markov cachée, la probabilité d'une suite de valeurs observée y_0, \dots, y_n est donnée par

$$\sum_{x_{0:n} \in S^{n+1}} \nu(x_0)p_0(x_0, x_1) \cdots p_{n-1}(x_{n-1}, x_n)q_0(x_0, y_0) \cdots q_n(x_n, y_n).$$

Clairement, calculer une telle somme est très rapidement impossible, même si S ne comporte qu'un petit nombre d'états. L'algorithme décrit dans l'exercice suivant montre comment des algorithmes de calcul itératifs peuvent être employés pour résoudre ce type de problème.

Exercice 52 (*Algorithme progressif-rétrograde de Viterbi*) Dans cet exercice, on suppose que les ensembles d'états S et V sont finis. On considère une chaîne de Markov cachée $(X_n, Y_n)_{n \geq 0}$ avec les notations ν , p_i , q_i définies précédemment, et une suite de valeurs observées y_0, \dots, y_n .

1) On définit une suite $(\alpha_i(x))_{0 \leq i \leq n}$, $x \in S$, de la manière suivante. Pour initialiser, on pose $\alpha_0(x) := \nu(x)q_0(x, y_0)$. Puis par récurrence :

$$\alpha_{i+1}(x) := \sum_{z \in S} \alpha_i(z)p_i(z, x)q_{i+1}(x, y_{i+1}).$$

Montrer que, pour tout $x \in S$ et $0 \leq i \leq n$, on a

$$\alpha_i(x) := P(Y_{0:i} = y_{0:i}, X_i = x).$$

Expliquer comment on peut en déduire la probabilité $P(Y_{0:n} = y_{0:n})$

2) On cherche à présent une suite d'états cachés x_0, \dots, x_n qui maximise la probabilité $P(Y_{0:n} = y_{0:n}, X_{0:n} = x_{0:n})$.

On définit deux suites $(\delta_i(x))_{0 \leq i \leq n}$ et $(\psi_i(x))_{1 \leq i \leq n}$, $x \in S$, de la manière suivante. Pour initialiser, on pose $\delta_0(x) := \nu(x)q_0(x, y_0)$. Ensuite,

$$\delta_{i+1}(x) := \max_{z \in S} \delta_i(z)p_i(z, x)q_{i+1}(x, y_{i+1}),$$

et $\psi_{i+1}(x)$ est défini comme l'un quelconque des z réalisant le maximum dans l'expression ci-dessus. On définit ensuite x_n^* comme l'un quelconque des x réalisant le maximum de $\delta_n(x)$, puis, par récurrence, $x_i^* := \psi_{i+1}(x_{i+1}^*)$. Montrer que la suite x_0^*, \dots, x_n^* réalise le maximum recherché.

Pour approfondir le sujet, voir par exemple l'ouvrage [8].

1.8 Exercices supplémentaires

Exercice 53 *Considérons un ensemble fini ou dénombrable S , un entier $m \geq 1$, et un sous-ensemble A de S^{m+1} , vu comme un ensemble de trajectoires de longueur m à valeurs dans S .*

1) *Pour $n \in \llbracket 0, m \rrbracket$, posons $\Pi_n := \{x_{0:n}; x_{0:m} \in A\}$, et, pour tout $x_{0:n} \in \Pi_n$, soit $\phi_n(x_{0:n}) := \{x_{n+1:m}; x_{0:m} \in A\}$. On dit que A est compatible-Markov⁴ s'il vérifie la propriété suivante :*

$$\forall n \in \llbracket 0, m \rrbracket, \forall x_{0:n}, y_{0:n} \in \Pi_n, x_n = y_n \Rightarrow \phi_n(x_{0:n}) = \phi_n(y_{0:n}) \quad (1.11)$$

Si A est compatible-Markov, on peut donc définir sans ambiguïté, pour tout x tel qu'il existe $x_{0:m} \in A$ vérifiant $x_n = x$, la notation $\phi_n(x) := \phi_n(x_{0:n})$. Montrer qu'un ensemble de la forme $C_0 \times \dots \times C_m$ est toujours compatible-Markov. Donner un exemple d'ensemble compatible-Markov ne pouvant pas se mettre sous cette forme.

2) *A présent, considérons une chaîne de Markov $(X_n)_{0 \leq n \leq m}$ d'espace d'états S , définie sur (Ω, \mathcal{F}, P) , et de famille de noyaux de transition $(p_n)_{0 \leq n \leq m-1}$. Montrer que, si A est compatible-Markov, X_0, \dots, X_m forme une chaîne de Markov par rapport à la probabilité conditionnelle $P(\cdot | X_{0:m} \in A)$. Montrer que les noyaux de transition correspondants sont donnés, pour tout x_n tel que $P(X_n = x_n | X_{0:m} \in A) > 0$ par la formule :*

$$q_n(x_n, x_{n+1}) := p_n(x_n, x_{n+1}) \times \frac{g_{n+1}(x_{n+1})}{g_n(x_n)},$$

où

$$g_n(x_n) := P(X_{n:m} \in \phi_n(x_n) | X_n = x_n).$$

3) *Donner un exemple d'ensemble A tel que X_0, \dots, X_n n'est pas une chaîne de Markov par rapport à la probabilité $P(\cdot | X_{0:m} \in A)$.*

Exercice 54 *Soit p un noyau de transition sur un ensemble fini ou dénombrable S , et $A \subset S$. On utilise la notation $0 =: T_0(A) < T_1(A) < T_2(A) < \dots$ pour désigner les instant successifs d'atteinte de A . Sur l'espace canonique, on définit la suite $(Z_n)_{n \geq 0}$ par $Z_n := X_{T_n(A)}$. Montrer que, si $x \in A$, la suite $(Z_n)_{n \geq 0}$ est une chaîne de Markov sous la probabilité \mathbb{P}_x .*

Exercice 55 *Soit un ensemble fini ou dénombrable S , et q un noyau de transition de S^3 vers S , c'est-à-dire une fonction associant à tout triplet d'éléments de S une mesure de probabilité sur S . Soit $N \geq 1$, et $a, b \in S$. On définit alors un noyau de transition sur S^N de la manière suivante. Etant donné $(x^1, \dots, x^N) \in S^N$, soit (Z^1, \dots, Z^N) une famille de variables aléatoires indépendantes telles que, pour tout*

⁴Cette terminologie n'est pas standard.

$1 \leq i \leq N$, la loi de Z_i soit donnée par $p((x^{i-1}, x^i, x^{i+1}), \cdot)$, avec la convention $x^{-1} := a$ et $x^{N+1} := b$. On définit alors $p((x^1, \dots, x^N), \cdot)$ comme la loi de (Z^1, \dots, Z^N) .

On se donne ensuite un point de départ fixé $x_0 := (x_0^1, \dots, x_0^N)$, et l'on considère une chaîne de Markov $(X_n)_{n \geq 0}$ initialisée en $X_0 := x_0$ et de noyau de transition p . On note $X_n := (X_n^1, \dots, X_n^N)$.

1) Montrer que, pour tout $n \geq 1$ fixé, la suite $(X_0^i, \dots, X_n^i)_{1 \leq i \leq N}$ est une chaîne de Markov d'ordre 2 sur S^{n+1} .

2) Montrer que, pour tout $n \geq 1$ et $1 \leq i \leq N$ fixé, conditionnellement à la donnée de (X_0^i, \dots, X_n^i) pour tout $j \neq i$, la suite (X_0^j, \dots, X_n^j) est une chaîne de Markov. Préciser le noyau de transition.

Exercice 56 Dans cet exercice, on considère un ensemble fini S , muni d'une relation d'ordre total notée \leq . On rappelle (ou non ?) la définition suivante : étant données deux mesures de probabilité μ_1 et μ_2 sur S , on dit que μ_1 est dominée par μ_2 (ou que μ_2 domine μ_1) si $\mu_1(f) \leq \mu_2(f)$ pour toute fonction croissante $f : S \rightarrow \mathbb{R}$.

1) Étant données deux variables aléatoires Y_1 et Y_2 définies sur un même espace de probabilité (Ω, \mathcal{F}, P) , on suppose que $P(Y_1 \leq Y_2) = 1$. Montrer que la loi de X_2 domine la loi de X_1 .

2) Montrer que la réciproque de la question précédente est vraie : si μ_1 et μ_2 sont deux mesures de probabilité sur S , et si μ_1 est dominée par μ_2 , alors il existe un espace de probabilité (Ω, \mathcal{F}, P) et deux variables aléatoires Y_1 et Y_2 de lois respectives μ_1 et μ_2 , et vérifiant $P(Y_1 \leq Y_2) = 1$.

Indication : comparer $\mu_1(\{x; a \leq x\})$ et $\mu_2(\{x; a \leq x\})$ pour $a \in S$, et utiliser un découpage d'intervalles.

Remarque : ce résultat est encore vrai pour des ensembles partiellement ordonnés, mais la preuve est plus délicate (c'est le théorème de Strassen, voir par exemple [33] pour une preuve).

3) Appliquer les résultats précédents afin de montrer la propriété suivante. Étant donné un noyau de transition p sur \mathbb{Z} tel que, pour tout x , $p(x, x-1) + p(x, x+1) = 1$, et étant donnés $x < y$ tels que $x - y \in 2\mathbb{Z}$, et $n \geq 1$, la loi de X_n sous la probabilité $\mathbb{P}_{x,p}$ est dominée par la loi de X_n sous la probabilité $\mathbb{P}_{y,p}$. Le résultat est-il encore vrai si l'on ne suppose pas que $x - y \in 2\mathbb{Z}$?

Indication : fabriquer un couplage approprié.

4) Appliquer les résultats précédents afin de montrer la propriété suivante. Étant donné un noyau de transition p sur \mathbb{Z} tel que, pour tout x , $p(x, x-1) + p(x, x+1) = 1$, et étant donnés $a \in \mathbb{Z}$, et $x < y \leq a$ tels que $x - y \in 2\mathbb{Z}$, et $n \geq 1$ tel que $\mathbb{P}_{y,p}(X_i \leq a, 0 \leq i \leq n) > 0$, montrer que la loi de X_n sous la probabilité $\mathbb{P}_{x,p}(\cdot | X_i \leq a, 0 \leq i \leq n)$ est dominée par la loi de X_n sous la probabilité $\mathbb{P}_{y,p}(\cdot | X_i \leq a, 0 \leq i \leq n)$. Le résultat est-il encore vrai si l'on ne suppose pas que $x - y \in 2\mathbb{Z}$?

Indication : fabriquer un couplage approprié.

Exercice 57 (Théorème de Diaconis-Freedman) *Le but de cet exercice est d'établir le théorème de Diaconis-Freedman (voir [13]) sur la caractérisation des mélanges de chaînes de Markov. Plus précisément, nous dirons qu'une suite de variables aléatoires $(X_n)_{n \geq 0}$ à valeurs dans un ensemble fini ou dénombrable S est un **mélange de chaînes de Markov** s'il existe une loi initiale ν et une loi Q sur l'ensemble des noyaux de transition sur S (muni de la tribu engendrée par les applications $(x, y) \mapsto p(x, y)$) tels que, pour tout $n \geq 0$ et toute suite $x_0, \dots, x_n \in S$,*

$$P(X_{0:n} = x_{0:n}) = \int \nu(x_0) \prod_{i=0}^{n-1} p_i(x_i, x_{i+1}) dQ(p).$$

*On définit ensuite une relation d'équivalence sur l'ensemble des suites finies d'éléments de S en considérant deux suites comme équivalentes lorsqu'elles ont le même état initial et que, pour tous $x, y \in S$, elles comportent exactement le même nombre de transitions de l'état x vers l'état y . (En particulier, deux telles suites ont obligatoirement la même longueur). Nous dirons enfin que $(X_n)_{n \geq 0}$ est **partiellement échangeable** lorsque, pour tout $n \geq 0$ et tout couple $x_{0:n}$ et $y_{0:n}$ de suites équivalentes, on a l'égalité*

$$P(X_{0:n} = x_{0:n}) = P(X_{0:n} = y_{0:n}).$$

Le théorème de Diaconis-Freedman est alors le suivant : une suite $(X_n)_{n \geq 0}$ pour laquelle $P(X_i = X_0 \text{ pour une infinité d'indices } i) = 1$ est un mélange de chaînes de Markov si et seulement si elle est partiellement échangeable.

1) *Montrer que si $(X_n)_{n \geq 0}$ est un mélange de chaînes de Markov, $(X_n)_{n \geq 0}$ doit nécessairement être partiellement échangeable.*

2) *Donner un exemple d'une suite $(X_n)_{n \geq 0}$ partiellement échangeable qui n'est pas un mélange de chaînes de Markov. Indication : un exemple déterministe peut suffire.*

Dans la suite de l'exercice, on suppose donnée une suite $(X_n)_{n \geq 0}$ partiellement échangeable vérifiant la condition $P(X_i = X_0 \text{ pour une infinité d'indices } i) = 1$, et l'on cherche à prouver que celle-ci est un mélange de chaînes de Markov.

3) *Montrer que l'on peut, sans perte de généralité, supposer que $P(X_0 = x) = 1$ pour un certain $x \in S$. On supposera donc cette condition vérifiée dans la suite.*

4) *On définit la suite $(L_i)_{i \geq 1}$ de suites finies de S de la manière suivante. Appelons S_0, S_1, \dots les valeurs successives des indices n pour lesquels $X_n = x$ (on a donc $S_0 = 0$). Pour tout $i \geq 1$, on pose alors $L_i := X_{S_i:S_{i+1}-1}$. Montrer que la suite $(L_i)_{i \geq 1}$ est échangeable, au sens suivant : pour tout entier $n \geq 1$, et toute permutation σ des entiers $\{1, \dots, n\}$,*

$$(L_1, \dots, L_n) \stackrel{\text{loi}}{=} (L_{\sigma(1)}, \dots, L_{\sigma(n)}).$$

5) *Définissons la tribu asymptotique de la suite $(L_i)_{i \geq 1}$ par $\mathcal{G} := \bigcap_{i \geq 1} \sigma(L_j; j \geq i)$. Vérifier que la suite $(X_n)_{n \geq 0}$ conserve la propriété d'échangeabilité partielle conditionnellement à \mathcal{G} .*

6) Montrer que, si les $(L_i)_{i \geq 0}$ sont i.i.d., la suite $(X_n)_{n \geq 0}$ est une chaîne de Markov.
 7) Conclure en utilisant le théorème de De Finetti, qui affirme que, du fait de leur propriété d'échangeabilité établie en 4), les variables $(L_i)_{i \geq 1}$ sont i.i.d. conditionnellement à \mathcal{G} .

Exercice 58 Montrer que, si (X_0, \dots, X_n) est une chaîne de Markov (a priori inhomogène), c'est également le cas de (X_n, \dots, X_0) . Dans le cas où l'on part d'une chaîne homogène, obtient-on toujours une chaîne homogène en effectuant ce retournement du temps ?

Exercice 59 Soient S, V , deux ensembles finis ou dénombrables, et $f : S \rightarrow V$ une application. Soient par ailleurs un noyau de transition p sur S et un noyau de transition q sur V . Si μ désigne une probabilité sur S , on note $f_*\mu$ la probabilité image de μ par f , c'est-à-dire la probabilité définie sur V par $(f_*\mu)(A) = \mu(f^{-1}(A))$.
 1) On suppose que, pour toute probabilité μ sur S ,

$$(f_*\mu)q = f_*(\mu p). \quad (1.12)$$

Montrer que, si $(X_n)_{n \geq 0}$ est une chaîne de Markov sur S de loi initiale ν , $(f(X_n))_{n \geq 0}$ est une chaîne de Markov sur V de loi initiale $f_*\nu$ et de noyau de transition q .

2) Donner une condition plus explicite portant sur p, q, f équivalente à la condition (1.12) ci-dessus.

3) Lorsque la propriété (1.12) ci-dessus est vérifiée avec $V = S$ et $q = p$, on dit que p est invariante sous l'action de f . Montrer qu'une marche aléatoire (à gauche) sur un groupe est invariante par la multiplication (à gauche) par un élément du groupe.

Chapitre 2

Décompositions de l'espace d'états, récurrence/transience

Dans ce chapitre, nous introduisons diverses décompositions importantes de l'espace d'états basées sur les propriétés de communication entre les points vis-à-vis du noyau.

2.1 Points essentiels, communication entre points, irréductibilité, période

2.1.1 Points essentiels

Etant donné un noyau markovien p sur un ensemble fini ou dénombrable S , un point $x \in S$ est dit **inessentiel** (pour $p(\cdot, \cdot)$) s'il existe $m \geq 1$ et $y \in S$ tels que $p^m(x, y) > 0$ et $p^k(y, x) = 0$ pour tout k . Un point est dit **essentiel** lorsqu'il n'est pas inessentiel. Autrement dit, un point est inessentiel lorsqu'il existe une probabilité positive pour que, partant de x , la chaîne atteigne un point d'où elle ne peut jamais revenir en x .

Une première décomposition de l'espace d'états S s'écrit alors

$$S = S_{iness.} \cup S_{ess.},$$

où $S_{iness.}$ et $S_{ess.}$ désignent les ensembles de points inessentiels et essentiels respectivement.

Exercice 60 *Donner un exemple dans lequel il n'y a que des points inessentiels, un exemple dans lequel il n'y a que des points essentiels, et un exemple où les deux types de points coexistent.*

Proposition 13 *Si x est un point essentiel et y un point inessentiel (par rapport à $p(\cdot, \cdot)$), alors $p^m(x, y) = 0$ pour tout m . En d'autres termes, l'ensemble des points essentiels est stable par $p(\cdot, \cdot)$: si $x \in S_{ess.}$, $p(x, S_{ess.}) = 1$.*

Preuve :

Par l'absurde. On sait qu'il existe z et k tels que $p^k(y, z) > 0$ et $p^n(z, y) = 0$ pour tout n . S'il existe m tel que $p^m(x, y) > 0$, on en déduit que $p^{m+k}(x, z) > 0$. Comme x est essentiel, il existe ℓ tel que $p^\ell(z, x) > 0$. Par conséquent, $p^{\ell+m}(z, y) > 0$, ce qui contredit le fait que la probabilité d'atteindre y en partant de z soit nulle. \square

Exercice 61 *Vérifier que $S_{iness.}$ n'est, quant à lui, pas nécessairement stable.*

Lorsque x est un point inessentiel, la proposition 8 peut alors s'appliquer à x , car $\mathbb{P}_x(T_1(x) < +\infty) < 1$ par définition, et l'on voit donc que, quelle que soit la loi initiale ν choisie, le nombre de visites en x par une chaîne de noyau de transition p est fini presque sûrement, et possède même une queue à décroissance géométrique. Cette remarque justifie en quelque sorte la terminologie employée.

2.1.2 Communication entre points, irréductibilité

On dit que x et y **communiquent** (par rapport à $p(\cdot, \cdot)$) s'il existe m_1 et m_2 tels que $p^{m_1}(x, y) > 0$ et $p^{m_2}(y, x) > 0$. Cette définition entraîne que x communique toujours avec lui-même, car par convention $p^0(x, x) = 1$.

Proposition 14 *La relation de communication entre points définit une relation d'équivalence sur S .*

Exercice 62 *Prouver la proposition.*

Exercice 63 *Montrer que la classe de communication de x n'est autre que la réunion des points se trouvant sur les cycles issus de x possédant une probabilité positive d'être parcourus.*

Proposition 15 *Les ensembles $S_{ess.}$ et $S_{iness.}$ sont stables par cette relation d'équivalence : si x communique avec y et si $x \in S_{ess.}$, $y \in S_{ess.}$.*

Exercice 64 *Prouver la proposition.*

Les ensembles $S_{ess.}$ et $S_{iness.}$ peuvent donc eux-mêmes être partitionnés en classes d'équivalence pour la relation de communication.

Proposition 16 *Si C est une classe d'équivalence pour la communication incluse dans $S_{ess.}$, C est stable par $p(\cdot, \cdot)$: si $x \in C$, $p(x, C) = 1$.*

Preuve :

Par l'absurde. Si $x \in C$, $p(x, y) > 0$ et y ne communique pas avec x , on a nécessairement que $p^k(y, x) = 0$ pour tout k . Ainsi, x est inessentiel, ce qui contredit le fait que $x \in S_{ess}$. \square

On tire de ce qui précède la représentation suivante d'une trajectoire de la chaîne. Initialisée en un point essentiel, la chaîne reste pour toujours confinée à la classe de communication du point dont elle est issue. Initialisée en un point inessentiel, au contraire, elle finit toujours par sortir de la classe de communication de son point initial, et cette sortie est obligatoirement définitive. Si le point atteint lors de cette sortie est inessentiel, le même scénario se répète. Inversement, si le point atteint lors de la sortie est essentiel, la chaîne reste ensuite pour toujours confinée dans la classe de communication du point qu'elle vient d'atteindre.

Exercice 65 *Montrer que, si S est de cardinal fini, toute trajectoire finit, avec une probabilité égale à 1, par atteindre un point essentiel.*

On dit que $p(\cdot, \cdot)$ est **irréductible** si x communique avec y pour tous x, y dans S . D'après les propositions précédentes, la restriction de $p(\cdot, \cdot)$ restreint à une classe de communication de S_{ess} est irréductible. Quitte à enlever de S les points inessentiels par rapport au noyau de transition considéré (qui ne sont visités qu'un nombre fini de fois sur toute la durée d'une trajectoire), et à étudier séparément les chaînes obtenues par restriction sur chaque classe de S_{ess} , (rappelons que celles-ci sont stables), on peut donc se ramener à étudier des noyaux de transition irréductibles. Ceci ne signifie pas que le comportement de la chaîne à partir de points inessentiels soit inintéressant !

Remarque 3 *On constate que, si une chaîne de Markov possède un noyau de transition irréductible, tous les points de S ont une probabilité positive d'être visités, quelle que soit la loi initiale, et, par conséquent, ce noyau est le seul compatible avec cette chaîne. On peut donc parler d'une chaîne de Markov irréductible sur un ensemble S . Attention cependant, la notion d'irréductibilité dépend de l'ensemble S dans lequel on considère que la chaîne prend ses valeurs : une chaîne, vue comme famille de variables aléatoires dans un ensemble S_1 pourra se révéler irréductible sur S_1 , mais pas sur un ensemble S_2 contenant S_1 .*

Remarque 4 *La définition de l'irréductibilité entraîne le fait que, pour tout $x \in S$, il existe $m \geq 1$ tel que $p^m(x, x) > 0$.*

Exercice 66 *Dans le cas où S est fini, et où p peut donc se représenter comme une matrice, à quelles propriétés de cette matrice l'existence de sous-ensembles de S stables par p correspond-elle ?*

Exercice 67 Reprendre tous les exemples de noyaux de transition vus précédemment et étudier leurs ensembles de points essentiels, inessentiels, et les classes d'équivalence pour la relation de communication de l'ensemble des points essentiels.

2.1.3 Période

On définit la **période** (par rapport à $p(\cdot, \cdot)$) d'un point x par

$$d(x) = \text{p.g.c.d.}\{n \geq 1; p^n(x, x) > 0\}.$$

(Avec la convention que $\text{p.g.c.d.}\emptyset = +\infty$).

Pour un noyau irréductible, on constate que la période est nécessairement finie pour tout point. Par exemple, on voit facilement que, pour (le noyau de transition de) la marche aléatoire simple sur \mathbb{Z} , la période de tout point est égale à 2.

Exercice 68 Montrer par un exemple que l'on n'a pas nécessairement $p^{d(x)}(x, x) > 0$.

Proposition 17 Si $p(\cdot, \cdot)$ est irréductible, $d(x) = d(y)$ pour tous $x, y \in S$.

Preuve :

Considérons x et y dans S . Par irréductibilité, il existe des entiers a, b, c non-nuls tels que $p^a(x, y) > 0$, $p^b(y, y) > 0$ et $p^c(y, x) > 0$. On voit ainsi que $p^{a+nb+c}(x, x) > 0$ pour tout $n \geq 0$, donc, par définition, que $d(x)$ divise $a + nb + c$ pour tout n . En faisant $n = 0$, on voit que $d(x)$ doit diviser $a + c$. En faisant $n = 1$, on en déduit que $d(x)$ doit diviser b . Par conséquent, tout $b \geq 1$ tel que $p^b(y, y) > 0$ vérifie que $d(x)$ divise b . Par conséquent, $d(x)$ divise $d(y)$. Symétriquement, $d(y)$ divise $d(x)$, donc $d(x) = d(y)$. \square

Lorsque $p(\cdot, \cdot)$ est irréductible, on parle donc de la période d du noyau ou de la chaîne, sans préciser le point. Lorsque $d = 1$, on dit que $p(\cdot, \cdot)$ est apériodique.

Remarque 5 Une condition suffisante très simple pour avoir apériodicité dans le cas d'une chaîne irréductible est qu'il existe au moins un point x tel que $p(x, x) > 0$.

Proposition 18 Supposons $p(\cdot, \cdot)$ irréductible et de période d . Considérons x et y dans S . Si $p^m(x, y) > 0$ et $p^n(x, y) > 0$, alors d divise $m - n$.

Preuve :

Soit a tel que $p^a(y, x) > 0$. On a donc $p^{m+a}(x, x) > 0$ et $p^{n+a}(x, x) > 0$. Donc d divise $m + a$ et $n + a$. \square

Proposition 19 *Supposons $p(\cdot, \cdot)$ irréductible et de période d . La relation définie sur S par $x \sim_{per.} y$ s'il existe $n \geq 0$ tel que $p^{nd}(x, y) > 0$ est une relation d'équivalence sur S .*

Sous les hypothèses de la proposition ci-dessus, définissons, pour tout $x \in S$ et tout $h \in \mathbb{Z}/d\mathbb{Z}$, l'ensemble

$$C_h(x) = \{y \in S; \exists m \geq 0, p^m(x, y) > 0, \bar{m} = h\},$$

où \bar{m} désigne la classe d'équivalence de m dans $\mathbb{Z}/n\mathbb{Z}$.

Proposition 20 *Supposons $p(\cdot, \cdot)$ irréductible et de période d . L'ensemble S est partitionné en d classes d'équivalence pour la relation $\sim_{per.}$, et, pour tout $x \in S$ les classes sont exactement l'ensemble des $C_h(x)$, $h \in \mathbb{Z}/d\mathbb{Z}$. Qui plus est, pour tout $x \in S$, $p(x, C_{h+1}) = 1$.*

Exercice 69 *Prouver les deux propositions ci-dessus.*

Proposition 21 *Supposons $p(\cdot, \cdot)$ irréductible et de période d . Pour tout $h \in \mathbb{Z}/d\mathbb{Z}$, les classes d'équivalence de S pour $\sim_{per.}$ sont stables par $p^d(\cdot, \cdot)$, et la restriction de $p^d(\cdot, \cdot)$ à l'une quelconque de ces classes est irréductible et apériodique.*

Preuve :

L'irréductibilité est une conséquence immédiate de l'irréductibilité de $p(\cdot, \cdot)$ et de la discussion ci-dessus. Quant à l'apériodicité, c'est également évident vu la définition de la période. \square

Ainsi, quitte à décomposer une chaîne irréductible en d sous-chaînes correspondant chacune à un état sur d de la chaîne originale, on peut se ramener à étudier des chaînes de Markov irréductibles et apériodiques.

Exercice 70 *Reprendre tous les exemples de noyaux de transition vus précédemment et étudier leur période.*

2.2 Récurrence et transience

Définition 4 *On dit que $x \in S$ est récurrent (pour $p(\cdot, \cdot)$) si $\mathbb{P}_x(T_1(x) < +\infty) = 1$. Si $\mathbb{P}_x(T_1(x) < +\infty) < 1$, x est dit transient.*

Remarque 6 *Noter la différence de définition entre un point inessentiel et un point transient!*

Proposition 22 *Si x est récurrent, $\mathbb{P}_x(T_i(x) < +\infty) = 1$ pour tout i , autrement dit, x est visité une infinité de fois, ou encore $\mathbb{P}_x(N(x) = +\infty) = 1$.*

Preuve :

Voir la proposition 8. □

Proposition 23 *Si x est transient, $\mathbb{P}_x(N(x) < +\infty) = 1$ et $\mathbb{E}_x(N(x)) < +\infty$. On a même que la loi de $N(x)$ sous \mathbb{P}_x est exactement une loi géométrique de paramètre $\mathbb{P}_x(T_1(x) = +\infty) > 0$.*

Preuve :

Idem. □

Corollaire 7 *Le point x est récurrent si et seulement si*

$$\sum_{i=1}^{+\infty} \mathbb{P}_x(X_i = x) = +\infty.$$

Preuve :

On écrit $N(x) = \sum_{i=1}^{+\infty} \mathbf{1}(X_i = x)$. Par conséquent (Fubini ou convergence monotone),

$$\mathbb{E}_x(\tilde{N}(x)) = \sum_{i=1}^{+\infty} \mathbb{E}_x(\mathbf{1}(X_i = x)) = \sum_{i=1}^{+\infty} \mathbb{P}_x(X_i = x).$$

Si x est transient, $\mathbb{E}_x(N(x))$ est fini d'après la proposition précédente. Si x est récurrent, $N(x)$ est infini avec probabilité 1 sous \mathbb{P}_x , donc évidemment $\mathbb{E}_x(N(x))$ est infini. □

On note que, dans le cas où $(X_n)_{n \geq 0}$ est une suite de variables aléatoires i.i.d., le critère ci-dessus correspond simplement à l'application du lemme de Borel-Cantelli.

Exercice 71 *On utilise vraiment la propriété de Markov pour prouver le résultat ci-dessus. Pour une suite de variables aléatoires quelconque telle que $X_0 = x$, le fait que $\sum_{i=1}^{+\infty} P(X_i = x) = +\infty$ (ou, de manière équivalente, le fait que $E(N(x)) = +\infty$), n'entraîne pas en général que $P(N(x) = +\infty) = 1$. Donner un exemple d'une telle situation.*

Proposition 24 *Supposons $p(\cdot, \cdot)$ irréductible. La récurrence d'un point de S implique celle de tous les autres points.*

On parlera donc de chaînes de Markov irréductibles récurrentes ou transientes, ou de noyau récurrent ou transient.

Preuve :

Donnons-nous un point $x \in S$ récurrent, et un point quelconque $y \in S$ différent de x . Par irréductibilité, la probabilité

$$\mathbb{P}_x(\text{visiter } y \text{ avant de revenir en } x \text{ pour la première fois}),$$

autrement dit $\mathbb{P}_x(T_1(y) < T_1(x))$ est strictement positive. A chaque retour en x (et chaque passage en x est toujours suivi d'un nouveau passage en x), on a donc une probabilité $\mathbb{P}_x(T_1(y) < T_1(x))$ de visiter y avant de revenir en x , indépendamment de tout ce qui précède. Par conséquent, $\mathbb{P}_x(T_1(y) < +\infty) = 1$. A présent, $\mathbb{P}_y(T_1(x) < +\infty)$ doit nécessairement être égale à 1, ce sans quoi on aurait, lors du premier passage en y , une probabilité non-nulle de ne jamais revenir en x , ce qui contredirait la récurrence de x . Enfin, il est clair que $\mathbb{P}_y(T_1(x) < +\infty) = 1$ et $\mathbb{P}_x(T_1(y) < +\infty) = 1$ entraînent que $\mathbb{P}_y(T_1(y) < +\infty) = 1$, donc y est récurrent. \square

Lorsque $p(\cdot, \cdot)$ est irréductible, on parle donc du fait que $p(\cdot, \cdot)$ est récurrent ou transient (sans spécifier de point), ou que la chaîne est récurrente ou transiente.

Proposition 25 *Supposons $p(\cdot, \cdot)$ irréductible. Si $p(\cdot, \cdot)$ est récurrent, $\mathbb{P}_x(N(y) = +\infty) = 1$ pour tous x et y .*

Preuve :

D'après la preuve précédente, $\mathbb{P}_x(T_1(y) < +\infty) = 1$. On utilise ensuite le fait que y est récurrent. \square

Proposition 26 *Supposons $p(\cdot, \cdot)$ irréductible. Si $p(\cdot, \cdot)$ est transient, on a nécessairement que $\mathbb{E}_x(N(y)) < +\infty$ pour tous x et y , et l'on a même que, pour tout $k \geq 1$,*

$$\mathbb{P}_x(N(y) = k) = \mathbb{P}_x(T_1(y) < +\infty)(\mathbb{P}_y(T_1(y) < +\infty))^{k-1}\mathbb{P}_x(T_1(y) = +\infty),$$

et

$$\mathbb{P}_x(N(y) = 0) = \mathbb{P}_x(T_1(y) = +\infty).$$

Preuve :

Proposition 8. \square

Exercice 72 *Montrer qu'une marche aléatoire sur le groupe \mathbb{Z}^d est irréductible si et seulement si le sous-groupe de \mathbb{Z}^d engendré par le support de la loi des pas est égal à \mathbb{Z}^d .*

Exercice 73 *(Récurrence/transience des marches aléatoires simples symétriques sur \mathbb{Z}^d) L'objectif de cet exercice est de prouver que la marche aléatoire simple symétrique sur \mathbb{Z}^d est récurrente lorsque $d \leq 2$ et transiente lorsque $d \geq 3$. Pour cela, on se propose d'appliquer le critère du corollaire 7 en prouvant que*

$$p^{2n}(0, 0) \sim_{n \rightarrow +\infty} C_d n^{-d/2}. \quad (2.1)$$

1) Pour $d = 1$, utiliser l'expression explicite de $p^n(0, 0)$ et la formule de Stirling.

2) Noter que l'on pourrait déduire le résultat pour $d \geq 1$ du résultat pour $d = 1$ si les coordonnées de la marche aléatoire étaient des marches aléatoires simples unidimensionnelles indépendantes, ce qui n'est pas le cas. En revanche, il y a effectivement indépendance entre les coordonnées conditionnellement aux nombres de pas effectués sur chacune des coordonnées. Prouver d'abord ce résultat, puis l'utiliser pour déduire (2.1) à partir du cas $d = 1$. Dans le cas $d = 2$, il est possible d'éviter cet argument en considérant les projections de la marche aléatoires sur les deux bissectrices, ce qui permet d'obtenir deux marches aléatoires exactement indépendantes.

3) Reprouver le résultat avec l'approche suivante, dont le potentiel de généralisation est plus important : calculer explicitement la transformée de Fourier de $x \mapsto p^n(0, x)$ et l'inverser.

On notera au passage que (2.1) s'obtient en extrapolant (non-rigoureusement) le théorème de la limite centrale pour en faire un théorème limite local. Par ailleurs, malgré la récurrence en dimension 1 et 2, le théorème de la limite centrale entraîne que, après n pas, la marche aléatoire est typiquement à une distance de l'ordre de $n^{1/2}$ de son point de départ. Le même résultat est valable en dimension $d \geq 3$, pour lequel la chaîne est transiente. Par ailleurs, on note qu'en dimension 3, bien qu'il y ait transience, aucune coordonnée de la marche ne tend vers $\pm\infty$.

Exercice 74 (Récurrence/transience des marches aléatoires sur \mathbb{Z} : cas d'une loi symétrique à support borné)

On considère une marche aléatoire irréductible sur le groupe \mathbb{Z} que l'on note $(X_n)_{n \geq 0}$, initialisée en 0. On suppose que la loi des incréments est symétrique et à support borné. Nous allons montrer que la chaîne est récurrente. Notons $(\varepsilon_n)_{n \geq 1}$ la suite des incréments de la chaîne.

1) Montrer que les événements

$$A_+ := \{\limsup_{n \rightarrow +\infty} n^{-1/2} X_n = +\infty\}, \quad A_- := \{\limsup_{n \rightarrow +\infty} n^{-1/2} X_n = -\infty\},$$

sont mesurables par rapport à la tribu asymptotique $\bigcap_{n \geq 1} \sigma(\varepsilon_0, \dots, \varepsilon_n)$. On rappelle que, d'après la loi du 0-1 de Kolmogorov, les v.a. $(\varepsilon_n)_{n \geq 1}$ étant i.i.d., tout événement de cette tribu est de probabilité 0 ou 1.

2) Montrer que $P(A_+) = P(A_-)$.

3) Montrer que, pour tout $K \geq 0$,

$$P(\limsup_{n \rightarrow +\infty} n^{-1/2} X_n > K) \geq \limsup_{n \rightarrow +\infty} P(n^{-1/2} X_n > K).$$

4) Utiliser le théorème de la limite centrale pour déduire de ce qui précède que $P(A_+) > 0$, d'où $P(A_+) = 1$.

5) Conclure.

Exercice 75 (Récurrence/transience des marches aléatoires sur \mathbb{Z}^d : cas biaisé) On considère une marche aléatoire irréductible sur le groupe \mathbb{Z}^d dont la distribution des pas présente un biais, c'est-à-dire possède une espérance non-nulle. Montrer qu'une telle marche est toujours transiente.

Exercice 76 (Récurrence/transience des marches aléatoires sur \mathbb{Z} : cas centré) Le but de cet exercice est de prouver qu'une marche aléatoire irréductible sur le groupe \mathbb{Z} dont la distribution des pas est d'espérance nulle, est toujours récurrente.

A cette fin, on introduit, pour tout $x \in \mathbb{Z}$ et $n \geq 0$,

$$g_n(x) := \sum_{k=0}^n p^k(0, x),$$

où p est le noyau de transition associé la marche.

1) Prouver que $g_n(x)$ est l'espérance du nombre de visites en x après n pas de la marche initialisée en 0 (en incluant le point de départ dans le décompte des sites).

2) En utilisant la propriété de Markov, prouver que, pour tous $n \geq 0$, et $x \in \mathbb{Z}$, $g_n(0) \geq \sum_{k=0}^n p^k(x, 0)$. En déduire que, pour tout $n \geq 0$ et $x \in \mathbb{Z}$, on a $g_n(0) \geq g_n(x)$.

3) On note $A_\ell := \{-\ell, \dots, +\ell\}$. Déduire de ce qui précède que, pour tout $n \geq 0$, et tout $\ell \geq 1$, on a

$$g_n(0) \geq \frac{1}{|A_\ell|} \sum_{x \in A_\ell} g_n(x).$$

4) Prouver l'identité

$$\sum_{x \in A_\ell} g_n(x) = \sum_{k=0}^n \sum_{x \in A_\ell} p^k(0, x).$$

5) Pour un réel $a > 0$ fixé, on pose $\ell_n := \lfloor an \rfloor$. Prouver (en utilisant la loi des grands nombres) que, pour tout $a > 0$, on a

$$\liminf_{n \rightarrow +\infty} \sum_{x \in A_{\ell_n}} p^n(0, x) \geq 1.$$

6) Déduire de ce qui précède que $\sum_{k=0}^{+\infty} p^k(0, 0) = +\infty$, et conclure que 0 est un point récurrent.

7) Qu'est-ce qui empêche d'étendre l'argument en dimension plus grande ?

Exercice 77 (Récurrence/transience des marches aléatoires sur \mathbb{Z}^d : critère général) On considère une marche aléatoire irréductible sur \mathbb{Z}^d . On appelle μ la loi des pas, p le noyau correspondant, et l'on définit, pour tout $t \in \mathbb{R}$, la transformée de Fourier

$$\phi(\theta) := \sum_{x \in \mathbb{Z}^d} \exp(i\theta x) \mu(x).$$

L'objectif de cet exercice est de montrer que la marche est récurrente si et seulement si

$$\lim_{t \rightarrow 1} \int_{[0, 2\pi]^d} \Re \left(\frac{1}{1 - t\phi(\theta)} \right) d\theta = +\infty.$$

1) Montrer que

$$\sum_{n=0}^{+\infty} p^n(0, 0) = \lim_{t \rightarrow 1} \sum_{n=0}^{+\infty} t^n p^n(0, 0).$$

2) Montrer que, pour tout $0 \leq t < 1$,

$$\sum_{n=0}^{+\infty} t^n p^n(0, 0) = \frac{1}{(2\pi)^d} \sum_{n=0}^{+\infty} \int_{[0, 2\pi]^d} t^n \phi(\theta)^n d\theta.$$

3) Conclure.

4) Montrer que la récurrence implique que

$$\int_{[0, 2\pi]^d} \Re \left(\frac{1}{1 - \phi(\theta)} \right) d\theta = +\infty.$$

Il est en fait possible de montrer que la réciproque est vraie, ce qui fournit un critère plus joli que celui prouvé dans l'exercice, mais la preuve est sensiblement plus délicate.

Exercice 78 (Récurrence/transience des marches aléatoires sur \mathbb{Z}^d : encore !)

En utilisant le critère de l'exercice précédent, prouver¹ les résultats suivants :

- 1) Une marche aléatoire irréductible sur \mathbb{Z} avec des pas centrés est récurrente.
- 2) Une marche aléatoire irréductible sur \mathbb{Z}^2 avec des pas centrés et possédant un moment d'ordre 2 fini est récurrente.
- 3) Une marche aléatoire irréductible sur \mathbb{Z}^d avec $d \geq 3$ est toujours transiente.

Pour approfondir les questions de récurrence/transience des marches aléatoires sur le groupe \mathbb{Z}^d , consulter par exemple consulter l'ouvrage [50].

Exercice 79 Considérons le noyau de transition défini sur \mathbb{Z} par $p(x, x+1) = 1 - p(x, x-1) = \alpha$ pour tout $x > 0$, $p(x, x-1) = 1 - p(x, x+1) = \alpha$ pour tout $x < 0$, et $p(0, 1) = p(0, -1) = 1/2$.

- 1) Montrer que, lorsque $\alpha > 1/2$, ce noyau est transient.
- 2) Montrer qu'une trajectoire initialisée en 0 tend vers $+\infty$ avec probabilité $1/2$, et vers $-\infty$ avec probabilité $1/2$.

Exercice 80 Montrer que, si S est fini, toute chaîne de Markov irréductible est nécessairement récurrente.

¹Cela nécessite pas mal de travail!

Définition 5 On dit que x est positivement récurrent (pour $p(\cdot, \cdot)$) si $\mathbb{E}_x(T_1(x)) < +\infty$. Dans le cas contraire, on dit que x est récurrent nul.

Proposition 27 Supposons $p(\cdot, \cdot)$ irréductible et récurrent. La récurrence positive d'un point entraîne celle de tous les autres. Il en va donc de même de la récurrence nulle.

On parlera donc de chaînes de Markov irréductibles positivement récurrentes ou récurrentes nulles, ou de noyau récurrent nul ou transient.

Preuve :

Donnons-nous un point $x \in S$ récurrent positif, et un point quelconque $y \in S$ différent de x . Montrons d'abord que $\mathbb{E}_x(T_1(y)) < +\infty$. Notons d'abord que, du fait de la récurrence de la chaîne, on doit nécessairement avoir que $\mathbb{P}_x(T_1(y) < T_1(x)) > 0$. Introduisons le nombre N de retours en x effectués avant de toucher y pour la première fois, soit

$$N := \sum_{n=1}^{T_1(y)} \mathbf{1}(X_n = x),$$

et le temps nécessaire pour, partant de X_N (qui vaut alors x) à l'instant N , atteindre y pour la première fois, soit

$$U := T_1(y) - T_N(x).$$

On peut alors écrire que

$$T_1(y) = \left(\sum_{i=1}^N T_i(x) - T_{i-1}(x) \right) + U. \quad (2.2)$$

En faisant appel à la propriété de Markov forte et au lemme 2, on déduit les faits suivants. Premièrement, sous \mathbb{P}_x , conditionnellement à la valeur de N , les variables aléatoires

$$T_1(x), T_2(x) - T_1(x), \dots, (T_N(x) - T_{N-1}(x)), U$$

sont indépendantes. Toujours sous \mathbb{P}_x et conditionnellement à la valeur de N , les variables aléatoires $T_1(x), (T_2(x) - T_1(x)), \dots, (T_N(x) - T_{N-1}(x))$ possèdent toutes la même loi, à savoir celle de $T_1(x)$ sous la probabilité $\mathbb{P}_x(\cdot | T_1(y) > T_1(x))$, tandis que U possède la loi de $T_1(y)$ sous la probabilité $\mathbb{P}_x(\cdot | T_1(y) < T_1(x))$. De plus, la loi de N est géométrique de paramètre $\mathbb{P}_x(T_1(y) < T_1(x))$. A présent,

$$\begin{aligned} \mathbb{E}_x(T_1(y) | T_1(y) < T_1(x)) &\leq \mathbb{E}_x(T_1(x) | T_1(y) < T_1(x)) \\ &\leq \frac{\mathbb{E}_x(T_1(x))}{\mathbb{P}_x(T_1(y) < T_1(x))} < +\infty, \end{aligned}$$

car x est récurrent positif. Si $\mathbb{P}_x(T_1(y) < T_1(x)) = 1$, on a $N = 0$ $\mathbb{P}_x - p.s.$, et l'on peut déduire le résultat recherché de (2.2). Dans le cas contraire, on montre comme ci-dessus que $\mathbb{E}_x(T_1(x)|T_1(x) < T_1(y)) < +\infty$. En reprenant (2.2), et en utilisant les propriétés d'indépendance des $(T_i(x) - T_{i-1}(x))_{1 \leq i \leq n}$ conditionnellement à N , on en déduit alors que

$$\mathbb{E}_x(T_1(y)) \leq \mathbb{E}_x(N) \times \mathbb{E}_x(T_1(x)|T_1(x) < T_1(y)) + \mathbb{E}_x(T_1(y)|T_1(y) < T_1(x)) < +\infty.$$

Montrons à présent que $\mathbb{E}_y(T_1(x)) < +\infty$. D'après ce qui précède, $\mathbb{E}_x(T_1(x)|T_1(y) < T_1(x)) < +\infty$. Clairement, la loi de $T_1(x) - T_1(y)$ sous la probabilité $\mathbb{P}_x(\cdot|T_1(y) < T_1(x))$ est identique à la loi de $T_1(x)$ sous \mathbb{P}_y , grâce à la propriété de Markov. Donc $\mathbb{E}_y(T_1(x)) = \mathbb{E}_x(T_1(x) - T_1(y)|T_1(y) < T_1(x)) \leq \mathbb{E}_x(T_1(x)|T_1(y) < T_1(x)) < +\infty$.

À présent, on écrit que, sur l'espace canonique, $T_1(y) \leq T_1(x) + T_1(y) \circ \theta_{T_1(x)}$. Au vu du fait que $\mathbb{E}_y(T_1(x)) < +\infty$ et $\mathbb{E}_x(T_1(y)) < +\infty$, ceci entraîne la conclusion $\mathbb{E}_y(T_1(y)) < +\infty$. □

Corollaire 8 *On note comme corollaire de la preuve du résultat précédent que, pour un noyau irréductible positivement récurrent, on a, pour tous x, y , le fait que $\mathbb{E}_x(T_1(y)) < +\infty$.*

Exercice 81 *Montrer que, même pour un noyau irréductible positivement récurrent, il est possible que $\mathbb{E}_\nu(T_1(y)) = +\infty$ pour certaines lois initiales ν .*

Exercice 82 *Montrer que les marches aléatoires simples symétriques sur \mathbb{Z} et \mathbb{Z}^2 sont récurrentes nulles. Indication : en dimension 1, la loi du temps de premier retour en zéro se calcule explicitement², et l'on obtient que $\mathbb{P}_x(T_1(x) \geq n) \sim Cn^{-1/2}$.*

Exercice 83 *Montrer que, si S est fini, toute chaîne de Markov irréductible est nécessairement positivement récurrente. Indication : pour $x \in S$ fixé, il existe, pour tout $y \in S$, un entier $m(y, x) \geq 1$ tel que $p^{m(y, x)}(y, x) > 0$. En posant $m^*(x) := \max_{y \in S} m(y, x)$, considérer des intervalles de temps successifs de longueur $m^*(x)$.*

Exercice 84 *On se donne un arbre d -régulier enraciné, avec $d \geq 1$, et l'on considère le noyau de transition défini de la manière suivante : seules sont autorisées les transitions d'un sommet vers son père ou vers l'un de ses fils, la probabilité de transition d'un sommet différent de la racine à l'un quelconque de ses fils est un nombre fixé p , la probabilité de transition d'un sommet différent de la racine à son père est un nombre fixé q , les probabilités de transition de la racine vers ses fils étant chacune égale à $1/d$.*

²Par exemple, à l'aide du principe de réflexion, ou encore du calcul des nombres de Catalan basé sur leur relation de récurrence au moyen des séries génératrices, voir également l'exercice 106

1) En fonction de p, q, d , discuter le caractère récurrent positif/récurrent nul/transient, du noyau ainsi défini. Indication : pour la récurrence positive, on peut utiliser le fait que, si les $(\varepsilon_i)_{i \geq 0}$ sont des variables aléatoires i.i.d. valant ± 1 , on a une inégalité, valable pour tout $n \geq 0$, de la forme

$$P \left(\frac{\varepsilon_1 + \dots + \varepsilon_n}{n} \notin [E(\varepsilon_1) - h, E(\varepsilon_1) + h] \right) \leq \exp(-c(h)n),$$

avec $c(h) > 0$ si $h > 0$.

2) Quels sont les noyaux pour lesquels on peut facilement utiliser un couplage avec le cas traité à la question 1) pour conclure quant au caractère récurrent positif/récurrent nul/transient ?

Exercice 85 (Un résultat élémentaire de stabilité)

On se donne un noyau de transition p irréductible sur un ensemble fini ou dénombrable S . On considère ensuite un noyau q sur S tel que, pour tout couple x, y d'éléments de S tels que $(x, y) \notin A \times A$, $q(x, y) = p(x, y)$. En ce sens, q est une modification locale de p . L'objectif de cet exercice est de prouver que, si q est irréductible, q possède le même comportement que p en ce qui concerne la récurrence/transience, et la récurrence positive/récurrence nulle.

1) On utilise la notation $T_1(A)$ pour désigner le premier instant d'atteinte de A strictement postérieur à l'instant 0. Montrer que, pour tout $x \in A$, conditionnellement à l'événement $T_1(A) > 1$, la loi de $(X_n)_{0 \leq n \leq T_1(A)}$ est la même sous $\mathbb{P}_{x,p}$ et $\mathbb{P}_{x,q}$.

2) Conclure.

Exercice 86 (Décomposition de renouvellement pour une marche simple biaisée sur \mathbb{Z}^d)

On considère l'espace A des trajectoires infinies (x_0, x_1, x_2, \dots) à valeurs dans \mathbb{Z}^d , initialisées en 0, et effectuant des pas au plus proche voisin, A étant muni de la tribu engendrée par les coordonnées. La notation (e_1, \dots, e_d) désigne la base canonique de \mathbb{Z}^d , et on pose $B := \{+e_1, \dots, +e_d, -e_1, \dots, -e_d\}$. Pour $(x_0, x_1, x_2, \dots) \in A$, on a donc que $x_0 = 0$ et $x_{n+1} - x_n \in B$ pour tout $n \geq 0$. Pour tout $h \geq 1$, on définit sur A les suites de variables $(R_k)_{k \geq 0}$, $(S_k)_{k \geq 0}$ et $(D_k)_{k \geq 0}$ par récurrence de la manière suivante. Pour initialiser, on pose $R_0 := 0$, $D_0 := 0$. Ensuite, pour tout $k \geq 0$, on pose

$$S_{k+1} := \inf\{n \geq D_k; X_n \cdot e_1 = R_k + 1\},$$

$$D_{k+1} := \inf\{n \geq S_{k+1}; X_n \cdot e_1 = R_k\},$$

$$R_{k+1} := \sup\{X_n \cdot e_1; S_{k+1} \leq n \leq D_{k+1}\}.$$

avec les conventions $\inf \emptyset = +\infty$ et $\sup \emptyset = 0$. On définit ensuite

$$\kappa := \inf\{S_k; k \geq 0, D_k = +\infty\}.$$

A présent, considérons une marche aléatoire $(X_n)_{n \geq 0}$ sur le groupe \mathbb{Z}^d , initialisée en 0, et effectuant des pas au plus proche voisin. Plus précisément, on se donne une famille de variables aléatoires i.i.d. $(\varepsilon_i)_{i \geq 1}$ à valeurs dans B , définies sur un espace de probabilité (Ω, \mathcal{F}, P) , et X_n est défini par $X_0 := 0$ et $X_n := \sum_{i=1}^n \varepsilon_i$ pour $n \geq 1$. On suppose en outre que la marche est biaisée positivement dans la direction $+e_1$, autrement dit, que $E(\varepsilon_1 \cdot e_1) > 0$.

1) Montrer que $\kappa((X_n)_{n \geq 0})$ est fini avec probabilité 1.

2) La variable aléatoire $\kappa((X_n)_{n \geq 0})$ est-elle en général un temps d'arrêt pour la filtration définie par $\mathcal{F}_n := \sigma(X_0, \dots, X_n)$?

On définit à présent par récurrence

$$\kappa_1 := \kappa((X_n)_{n \geq 0}),$$

puis

$$\kappa_{i+1} := \kappa_i + \kappa((X_{\kappa_i+n} - X_{\kappa_i})_{n \geq 0}).$$

On définit également, pour tout $i \geq 0$,

$$\mathcal{F}_{\kappa_i} := \{A \in \mathcal{F}; \forall n \geq 0, A \cap \{\kappa_i = n\} \in \mathcal{F}_n\}.$$

4) Montrer que, pour tout $i \geq 1$, conditionnellement à \mathcal{F}_{κ_i} , la loi de $(X_{\kappa_i+n} - X_{\kappa_i})_{n \geq 0}$ n'est autre que la loi de $(X_n)_{n \geq 0}$ conditionnelle à l'événement $B := \{X_n \cdot e_1 \geq 0 \text{ pour tout } n \geq 0\}$. (Commencer par le cas $i = 1$.)

5) En déduire que $\kappa_1, \kappa_2 - \kappa_1, \kappa_3 - \kappa_2, \dots$ sont des variables indépendantes, possédant toutes la même distribution à l'exception de κ_1 . Même question avec $r_{\kappa_1}, r_{\kappa_2} - r_{\kappa_1}, r_{\kappa_3} - r_{\kappa_2}$.

L'exercice précédent montre comment on peut découper la trajectoire d'une marche aléatoire biaisée en tronçons indépendants et (sauf le premier) identiquement distribués. Comme nous le verrons dans un chapitre ultérieur, une telle décomposition est un premier pas pour prouver des théorèmes limites tels que la loi des grands nombres ou le théorème de la limite centrale, par exemple. Bien entendu, pour la marche aléatoire simple, ce résultat n'est pas très intéressant car on dispose de beaucoup d'autres techniques pour l'étudier. En revanche, l'approche développée dans cet exercice peut encore être utilisée dans certains modèles de marches aléatoires en milieu aléatoire, et de marches aléatoires en auto-interaction, pour lesquels les autres approches ne sont pas utilisables.

Exercice 87 (Marche aléatoire et branchement I) On considère un processus de branchement de Galton-Watson, dans lequel une population initialement constituée d'un individu évolue, chaque individu donnant lieu à un nombre aléatoire de descendants dans la génération suivante, les nombres de descendants des différents individus

étant choisis de manière *i.i.d.* selon une loi de reproduction fixée μ sur \mathbb{N} . On dit qu'il y a extinction si le nombre d'individus présents dans la population s'annule après un certain nombre de générations. Dans le cas contraire, on dit qu'il y a non-extinction. La théorie classique des processus de Galton-Watson, basée sur le calcul des fonctions génératrices, permet (entre autres) d'établir le résultat suivant. En notant m l'espérance (éventuellement infinie) du nombre de descendants d'un individu dans la génération suivante, c'est-à-dire l'espérance de μ , il y a extinction presque sûre lorsque $m \leq 1$, et non-extinction avec probabilité strictement positive si $m > 1$. Nous montrons ici comment relier cette propriété à la récurrence/transience d'une marche aléatoire.

1) Aller lire (ou relire) d'urgence un exposé de la théorie classique des processus de Galton-Watson, par exemple celui de [53].

On définit une suite $(Z_n)_{n \geq 0}$ de parties finies de \mathbb{N} , et une suite $(K_n)_{n \geq 0}$ de variables aléatoires à valeurs entières, de la manière suivante. On suppose donnée une suite $(M_n)_{n \geq 0}$ de variables *i.i.d.* de loi μ . Initialement, $Z_0 := \{1\}$ et $K_0 := 1$. Ensuite, étant donné Z_n et K_n , on distingue deux cas. Si $Z_n = \emptyset$, $Z_{n+1} := Z_n$ et $K_{n+1} := K_n$. Si $Z_n \neq \emptyset$, on pose

$$Z_{n+1} := Z_n \cup \{K_n + 1, K_n + 2, \dots, K_n + M_n\} \setminus \{\min Z_n\},$$

$$K_{n+1} := K_n + M_n.$$

2) Prouver que la probabilité d'extinction d'un processus de Galton-Watson de loi de reproduction μ et contenant initialement un individu, est la probabilité pour qu'il existe un n à partir duquel $Z_n = \emptyset$.

3) Montrer que la suite $(\text{card } Z_n)_{n \geq 0}$ est en fait une marche aléatoire sur le groupe \mathbb{Z} . Quelle est la loi de ses incréments ?

4) A partir des propriétés de récurrence/transience des marches aléatoires sur \mathbb{Z} , en déduire le critère $m \leq 1$ pour avoir une probabilité d'extinction strictement positive.

5) On définit $T := \inf\{n \geq 1; Z_n = \emptyset\}$, avec la convention $\inf \emptyset = +\infty$. Montrer que la variable aléatoire $1 + \sum_{n=0}^{T-1} M_n$ est égale à T et possède la même loi que le nombre total d'individus dans la généalogie (la somme des populations totales de chaque génération) d'un processus de Galton-Watson de loi de reproduction μ et contenant initialement un individu. En déduire que ce nombre est d'espérance finie dans le cas $\mu < 1$, et d'espérance infinie dans le cas $\mu = 1$.

6) L'approche de cet exercice permet de retrouver partiellement le critère de récurrence/transience des marches aléatoires sur le groupe \mathbb{Z} dont l'espérance des pas est définie, à partir de la théorie classique des processus de Galton-Watson. Quel type de loi exactement peut-on traiter au moyen de cette approche ?

Exercice 88 (Marche aléatoire et branchement II)

On considère une marche aléatoire simple $(X_n)_{n \geq 0}$ sur \mathbb{Z} (éventuellement biaisée), initialisée en 1. On note $T_1(0)$ le premier temps de retour en 0 de la chaîne, et, pour tout $j \geq 1$, on définit la variable aléatoire

$$N_j := \text{card} \{n \geq 0; n \leq T_1(0), (X_n, X_{n+1}) = (j, j+1)\}.$$

1) Montrer que la loi de la suite $(N_j)_{0 \leq j < T_1(0)}$ est la même que la loi des nombres d'individus dans les générations successives d'un processus de Galton-Watson comportant initialement un individu, considérées jusqu'à l'extinction du processus (si celle-ci n'a pas lieu, on prend en compte la totalité de la trajectoire).

2) A partir du critère d'extinction pour les processus de Galton-Watson, retrouver la fait que la marche aléatoire retourne presque sûrement en 0 si et seulement si le biais est négatif ou nul.

3) Inversement, retrouver dans ce cas particulier le critère d'extinction pour les processus de Galton-Watson à partir des résultats sur les temps de retour en 0 pour la marche.

L'exercice ci-dessus ne fait que reprouver un résultat que l'on peut obtenir par des moyens plus simples. Cependant, l'approche développée dans cette exercice, basée sur un lien entre les trajectoires d'une marche aléatoire uni-dimensionnelle, et un processus de branchement, peut être appliquée pour traiter certaines marches aléatoires en milieu aléatoire ou en auto-interaction, là où les autres approches ne peuvent pas nécessairement se généraliser.

Exercice 89 (*Méthode de Matthews pour les temps de recouvrement*) Considérons un ensemble fini S , un noyau de transition p irréductible sur S , et $(X_n)_{n \geq 0}$ une chaîne de Markov sur S de noyau de transition p . On appelle temps de recouvrement de S la variable aléatoire

$$T_{\text{rec.}} := \inf\{n; \{X_0, \dots, X_n\} = S\}.$$

Pour $x, y \in S$, posons $t(x, y) := \mathbb{E}_x(T(y))$, où $T(y)$ est le temps d'atteinte de y , et $H_n := \sum_{k=1}^n 1/k$. Le but de cet exercice est de montrer l'encadrement suivant :

$$H_{|S|} \times \min_{x \neq y} t(x, y) \leq E(T_{\text{rec.}}) \leq H_{|S|} \times \max_{x \neq y} t(x, y),$$

où $|S|$ désigne le nombre d'éléments de S .

1) Rappeler pourquoi $t(x, y) < +\infty$ pour tous $x, y \in S$. Montrer ensuite que l'on a $E(T_{\text{rec.}}) < +\infty$.

On considère à présent une permutation aléatoire uniforme $J_1, \dots, J_{|S|}$ des éléments de S indépendante de la chaîne de Markov étudiée, et l'on pose, pour tout $1 \leq m \leq |S|$,

$$C_m := \max_{1 \leq i \leq m} T(J_i),$$

et

$L_m :=$ le dernier élément de $\{J_1, \dots, J_m\}$ à être visité par la chaîne.

2) Montrer que, pour tout $2 \leq m \leq |S|$,

$$E(C_m - C_{m-1} | J_1, \dots, J_m, X_1, \dots, X_{C_{m-1}}) = t(L_{m-1}, J_m) \mathbf{1}(L_m = J_m).$$

3) Montrer à l'aide d'un argument de symétrie que, pour tout $2 \leq m \leq |S|$,

$$P(L_m = J_m) = 1/m.$$

4) En déduire que, pour tout $2 \leq m \leq |S|$,

$$\min_{x \neq y} t(x, y) \leq E(C_m - C_{m-1}) \leq \max_{x \neq y} t(x, y).$$

5) Montrer que $(1 - 1/n) \min_{x \neq y} t(x, y) \leq E(C_1) \leq (1 - 1/n) \max_{x \neq y} t(x, y)$.

6) Conclure.

Chapitre 3

Théorie du potentiel, mesures et lois invariantes

La théorie classique du potentiel s'est développée à partir de la notion de potentiel telle qu'elle apparaît en physique (par exemple le potentiel électrique ou gravitationnel), et étudie notamment l'opérateur Laplacien et les fonctions harmoniques. Il est possible de développer une théorie du potentiel (en fait, plusieurs théories) pour les chaînes de Markov, dans laquelle le rôle du Laplacien est dévolu à l'opérateur $p - I$. Dans le cas du mouvement Brownien, l'analogue de $p - I$ est le générateur infinitésimal du processus, qui s'identifie au Laplacien usuel, et l'on retrouve donc exactement la théorie classique du potentiel. Comme nous le verrons, les analogues probabilistes des objets et des résultats de la théorie du potentiel classique jouent un rôle très important dans l'étude des chaînes de Markov. Pour une introduction à la théorie du potentiel pour les chaînes de Markov, consulter par exemple [39]. Pour approfondir, voir par exemple [43], et [50] pour le cas des marches aléatoires sur le groupe \mathbb{Z}^d .

Un rôle central dans la théorie du potentiel est joué par l'équation de Poisson

$$\Delta V = f,$$

où V est le potentiel et f est une fonction donnée. L'analogue dans notre cadre, est l'équation de Poisson discrète

$$pu = u + c, \tag{3.1}$$

où p est un noyau de transition sur S , et où u et c sont des fonctions définies sur S et à valeurs réelles, c étant supposée à valeurs positives, et u pouvant éventuellement prendre les valeurs $\pm\infty$.

Définissons la **fonction de Green** $G(\cdot, \cdot)$ sur $S \times S$ par

$$G(x, y) := \sum_{n=0}^{+\infty} p^n(x, y) = \sum_{n=0}^{+\infty} \mathbb{P}_x(X_n = y) = \mathbb{E}_x \left(\sum_{n=0}^{+\infty} \mathbf{1}(X_n = y) \right). \tag{3.2}$$

En utilisant simplement l'identité $pp^n(x, y) = p^{n+1}(x, y)$, on voit directement à partir de la définition que, pour tout $y \in S$,

$$pG(\cdot, y) + I(\cdot, y) = G(\cdot, y), \quad (3.3)$$

et on obtient donc une solution positive de l'équation de Poisson discrète (3.1) ci-dessus en posant, pour tout $x \in S$,

$$u(x) := \sum_{y \in S} c(y)G(x, y).$$

De manière plus probabiliste, u se réécrit

$$u(x) = \mathbb{E}_x \left(\sum_{n \geq 0} c(X_n) \right).$$

Exercice 90 *Prouver (3.3) à partir de l'expression probabiliste de $G(x, y)$. Faire de même pour u .*

On note que, pour une chaîne irréductible, $G(x, y) = +\infty$ pour tous $x, y \in S$ dans le cas récurrent, tandis que $G(x, y) < +\infty$ pour tous $x, y \in S$ dans le cas transient. Par conséquent, la formule ci-dessus montre que, dans le cas récurrent, $u \equiv +\infty$, sauf si $c \equiv 0$. En revanche, dans le cas transient, on peut espérer (tout dépend de c) que u fournit une solution non-triviale pour l'équation de Poisson.

Il s'avère que l'on obtient une théorie plus intéressante en introduisant dans l'équation de Poisson une notion de frontière. C'est la théorie que nous présentons brièvement dans ce qui suit, dans le cas où la frontière est un sous-ensemble de S . On peut en fait étendre cette théorie dans le cas où la frontière se trouve en un certain sens "à l'infini" et (notamment) étudier ainsi le comportement asymptotique des chaînes transientes. (Voir par exemple la théorie de la frontière de Martin présentée dans [43] ou le chapitre de [46] sur le sujet.)

3.1 L'équation de Poisson avec frontière absorbante

Considérons un noyau de transition p sur un ensemble fini ou dénombrable S , et un sous-ensemble D de S , non-vide et strictement inclus dans S . Notons ∂D le complémentaire de D dans S , et supposons données deux fonctions $c : D \rightarrow \mathbb{R}_+$ et $\phi : \partial D \rightarrow \mathbb{R}_+$. Nous appellerons **équation de Poisson sur le domaine D avec frontière absorbante** l'équation suivante, dans laquelle l'inconnue est une fonction $u : S \rightarrow [-\infty, +\infty]$.

$$\begin{cases} u(x) = pu(x) + c(x), & x \in D; \\ u(x) = \phi(x), & x \in \partial D. \end{cases} \quad (3.4)$$

Lorsque $c \equiv 0$, l'équation (3.4) se transforme en

$$\begin{cases} u(x) = pu(x), & x \in D; \\ u(x) = \phi(x), & x \in \partial D. \end{cases} \quad (3.5)$$

et l'équation obtenue est **l'équation de Laplace**. Telle qu'écrite ci-dessus, avec une condition au bord fixée, l'équation (3.5) est appelée **problème de Dirichlet**, et les solutions de l'équation sont les **fonctions harmoniques**. Lorsque l'on suppose seulement que l'on a l'inégalité $u(x) \geq pu(x)$, on dit que l'on a affaire à une **fonction sur-harmonique**, et les solutions de l'équation de Poisson générale (3.4) sont donc des fonctions sur-harmoniques, puisque nous supposons toujours que $c \geq 0$.

3.1.1 Des fonctions de Green

Nous allons maintenant décrire l'analogie pour l'équation de Poisson avec frontière absorbante de la fonction de Green $G(\cdot, \cdot)$ décrite plus haut dans le cas sans frontière. Définissons $p_D(\cdot, \cdot)$ sur $S \times S$ par $p_D(x, y) := p(x, y)$ si $x, y \in D$, et $p_D(x, y) = 0$ sinon. Par analogie avec (3.3), posons alors

$$G_D(x, y) := \sum_{n=0}^{+\infty} p_D^n(x, y).$$

Noter que, bien que p_D ne soit pas un noyau de transition, on peut définir toutes les opérations – ici, l'itération de p_D avec lui-même – de la même façon que pour les noyaux de transition. Par convention, $p_D^0(x, y)$ vaut 1 si $x = y$ et $x, y \in D$, et 0 sinon.

Nous utiliserons dans la suite la notation suivante : étant donnée une chaîne de Markov $(X_n)_{n \geq 0}$, la notation T désigne le premier temps d'atteinte de ∂D , soit

$$T := \inf\{n \geq 0; X_n \in \partial D\},$$

avec la convention usuelle $\inf \emptyset = +\infty$. Cette notation permet de réécrire G_D d'une manière plus probabiliste :

$$G_D(x, y) = \sum_{n=0}^{+\infty} \mathbb{P}_x(X_n = y, T > n) = \mathbb{E}_x \left(\sum_{n=0}^{T-1} \mathbf{1}(X_n = y) \right).$$

Considérons $y \in D$ fixé. Comme dans le cas sans frontière, en utilisant simplement l'identité $p_D p_D^n = p_D^{n+1}$, on voit directement à partir de la définition que,

$$[pG_D(\cdot, y)](x) + I(x, y) = G_D(x, y) \text{ pour tout } x \in D. \quad (3.6)$$

Par ailleurs, du fait même de la définition, on a nécessairement que

$$G_D(x, y) = 0 \text{ pour tout } x \in \partial D.$$

Ainsi, on peut construire une solution positive de l'équation de Poisson (3.4) associée à c , mais avec $\phi \equiv 0$, en posant

$$u_D(x) := \sum_{y \in D} c(y) G_D(x, y).$$

Une expression plus probabiliste de u_D est donnée par

$$u_D(x) = \mathbb{E}_x \left(\sum_{n=0}^{T-1} c(X_n) \right).$$

Exercice 91 *Prouver (3.6) à partir de l'expression probabiliste de $G_D(x, y)$. Faire de même pour u_D .*

D'autre part, en utilisant cette fois l'identité $p_D^n p = p_D^{n+1}$, on montre que l'on a l'équation duale pour les mesures $G_D(x, \cdot)$ sur S , à savoir que, pour $x \in D$ fixé,

$$[G_D(x, \cdot)p](y) + I(x, y) = G_D(x, y) \text{ pour tout } y \in D, \quad (3.7)$$

tandis que

$$G_D(x, y) = 0 \text{ pour tout } y \in \partial D.$$

Exercice 92 *Prouver (3.7) à partir de l'expression probabiliste de $G_D(x, y)$.*

Compte-tenu du fait que u_D satisfait l'équation de Poisson (3.4) avec une condition au bord identiquement nulle, nous devons, pour obtenir une solution de l'équation (3.4) initiale, ajouter à u_D une solution du problème de Dirichlet (3.5) associé à la condition au bord ϕ . (C'est une conséquence immédiate de la linéarité de l'équation.) Celui-ci peut se résoudre en introduisant les fonctions suivantes. Définissons $G_{\rightarrow \partial D}(\cdot, \cdot)$ sur $S \times \partial D$ par

$$G_{\rightarrow \partial D}(x, y) := I(x, y) + (G_D p)(x, y),$$

soit, plus explicitement,

$$G_{\rightarrow \partial D}(x, y) = I(x, y) + \sum_{z \in D} G_D(x, z) p(z, y).$$

On note que $G_{\rightarrow \partial D}$ peut se réécrire, pour tout $x \in S$ et $y \in \partial D$, sous la forme

$$G_{\rightarrow \partial D}(x, y) := \sum_{n=0}^{+\infty} \mathbb{P}_x(X_n = y, T = n) = \mathbb{P}_x(X_T = y, T < +\infty). \quad (3.8)$$

Exercice 93 *Vérifier l'identité ci-dessus.*

Considérons $y \in \partial D$ fixé. On vérifie ici encore à partir de la définition que, pour tout $x \in D$, $(pI)(x, y) + (pG_D p)(x, y) = p(x, y) + (pG_D p)(x, y) = G_D(x, y)$. On en déduit donc que

$$[pG_{\rightarrow \partial D}(\cdot, y)](x) = G_{\rightarrow \partial D}(x, y) \text{ pour tout } x \in D,$$

tandis que la définition entraîne immédiatement que

$$G_{\rightarrow \partial D}(x, y) = \mathbf{1}(x = y) \text{ pour tout } x \in \partial D.$$

On en conclut finalement que, si l'on définit

$$u_{\partial D}(X) := \sum_{y \in D} \phi(y) G_{\rightarrow \partial D}(x, y),$$

on obtient une solution positive de l'équation de Poisson 3.4 associée à ϕ , mais avec $c \equiv 0$, c'est-à-dire une solution au problème de Dirichlet. Une expression probabiliste de $u_{\partial D}$ est alors donnée par

$$u_{\partial D}(x) = \mathbb{E}_x(c(X_T)\mathbf{1}(T < +\infty)).$$

Exercice 94 Vérifier (3.8) à partir de l'expression probabiliste de $G_{\rightarrow \partial D}(x, y)$. Faire de même avec $u_{\partial D}$.

Nous pouvons à présent énoncer le résultat suivant.

Théorème 4 On obtient une solution positive de l'équation de Poisson générale (3.4) en posant

$$u(x) := u_D(x) + u_{\partial D}(x).$$

Nous verrons plus tard qu'il est également intéressant de considérer la définition suivante, duale de celle de $G_{\rightarrow \partial D}$. Définissons $G_{\partial D \rightarrow}(\cdot, \cdot)$ sur $\partial D \times S$ par

$$G_{\partial D \rightarrow}(x, y) := I(x, y) + pG_D(x, y) = I(x, y) + \sum_{z \in D} p(x, z)G_D(z, y).$$

On note que $G_{\rightarrow \partial D}$ peut se réécrire, pour tout $x \in S$ et $y \in \partial D$,

$$G_{\rightarrow \partial D}(x, y) := \sum_{n=0}^{+\infty} \mathbb{P}_x(X_n = y, T_1 > n) = \mathbb{E}_x \left(\sum_{n=0}^{T_1-1} \mathbf{1}(X_n = y) \right),$$

où T_1 est le premier temps d'atteinte de D strictement après l'instant 0, soit

$$T_1 := \inf\{n \geq 1; X_n \in \partial D\},$$

avec la convention usuelle $\inf \emptyset = +\infty$. On peut alors effectuer le calcul dual de celui mené pour $G_{\rightarrow \partial D}$, à savoir que, $x \in \partial D$ étant fixé, on a, pour tout $y \in D$, l'identité $(Ip)(x, y) + (pG_{Dp})(x, y) = p(x, y) + (pG_{Dp})(x, y) = G_D(x, y)$, d'où

$$[G_{\partial D \rightarrow p(x, \cdot)}](y) = G_{\partial D \rightarrow}(x, y) \text{ pour tout } y \in D, \quad (3.9)$$

tandis que

$$G_{\partial D \rightarrow}(x, y) = \mathbf{1}(x = y) \text{ pour tout } y \in \partial D.$$

Exercice 95 *Prouver (3.9) à partir de l'expression probabiliste de $G_{\partial D \rightarrow}(x, y)$.*

Notons que, si $x, y \in \partial D$, on a l'identité

$$[pG_{\rightarrow \partial D}(\cdot, y)](x) = [G_{\partial D \rightarrow p(x, \cdot)}](y) = p(x, y) + (pG_{Dp})(x, y), \quad (3.10)$$

mais que cette dernière expression n'est pas en générale égale à $I(x, y)$.

3.1.2 Des martingales

La notion de fonction harmonique est intimement liée à celle de martingale, et ce lien permet notamment d'appliquer les méthodes de martingales à l'étude des fonctions harmoniques, ce que nous ferons à plusieurs reprises. Plus précisément, on a les théorèmes suivants (énoncés pour les fonctions positives ou bornées afin d'éviter d'avoir à discuter les questions d'intégrabilité). Commençons par le cas sans frontière :

Théorème 5 *Considérons une fonction f positive ou bornée, solution de l'équation de Poisson sans frontière (3.1). Alors, pour toute loi initiale ν sur S , et tout $n \geq 0$,*

$$\mathbb{E}_\nu(f(X_{n+1})|\mathcal{F}_n) = f(X_n) - c(X_n), \quad \mathbb{P}_\nu - p.s.$$

Preuve :

La preuve est essentiellement tautologique. Grâce à la propriété de Markov, $\mathbb{E}_\nu(f(X_{n+1})|\mathcal{F}_n) = pf(X_n) \mathbb{P}_\nu - p.s.$ \square

On note que le calcul effectué dans la preuve ci-dessus est exactement celui qui conduit à la martingale de l'exercice 28.

Corollaire 9 *Supposons que f est positive et satisfait $\mathbb{E}_\nu(f(X_0)) < +\infty$, ou que f est bornée, et est solution de l'équation de Poisson sans frontière (3.1). Si $c \equiv 0$, c'est-à-dire dans le cas d'une fonction harmonique, la suite $[f(X_n)]_{n \geq 0}$ est une martingale par rapport à la filtration $(\mathcal{F}_n)_{n \geq 0}$. Dans le cas général de l'équation de Poisson, on a seulement le fait que f est sur-harmonique, et que $[f(X_n)]_{n \geq 0}$ est une sur-martingale.*

Preuve :

Il suffit d'appliquer le résultat du théorème ci-dessus, qui permet de vérifier facilement par récurrence, dans le cas où f est supposée positive mais pas nécessairement bornée, que $\mathbb{E}_\nu(f(X_n)) < +\infty$ pour tout n et que l'on a la propriété de martingale (ou de sur-martingale). \square

Dans le cas avec frontière absorbante, on a le résultat suivant :

Théorème 6 *Considérons une fonction f positive ou bornée, solution de l'équation de Poisson avec frontière absorbante (3.4). Alors, pour toute loi initiale ν sur S , et tout $n \geq 0$,*

$$\mathbb{E}_\nu(f(X_{T \wedge (n+1)}) | \mathcal{F}_n) = f(X_{T \wedge n}) - c(X_{T \wedge n}) \mathbf{1}(T > n), \quad \mathbb{P}_\nu - p.s.$$

Preuve :

Grâce à la propriété de Markov, on a toujours que

$$\mathbb{E}_\nu(f(X_{n+1}) | \mathcal{F}_n) = pf(X_n) \quad \mathbb{P}_\nu - p.s.$$

Si $T > n$, on a nécessairement $X_n \in D$, d'où le fait que

$$pf(X_n) = f(X_n) - c(X_n).$$

Par ailleurs, la condition $T > n$ entraîne que $T \wedge (n+1) = n+1$ et $T \wedge n = n$. En utilisant le fait que l'événement $\{T = n\}$ est mesurable par rapport à \mathcal{F}_n du fait que T est un temps d'arrêt, on constate ainsi que $\mathbb{P}_\nu - p.s.$,

$$\mathbb{E}_\nu(f(X_{T \wedge (n+1)}) | \mathcal{F}_n) \mathbf{1}(T > n) = \mathbb{E}_\nu(f(X_{n+1}) | \mathcal{F}_n) \mathbf{1}(T > n) = pf(X_n) \mathbf{1}(T > n)$$

tandis que

$$pf(X_n) \mathbf{1}(T > n) = [f(X_{T \wedge n}) - c(X_{T \wedge n})] \mathbf{1}(T > n).$$

Si $T \leq n$, on a en revanche l'égalité $f(X_{T \wedge (n+1)}) = f(X_{T \wedge n}) = f(X_T)$, et $\mathbf{1}(T \leq n) f(X_T)$ est mesurable par rapport à \mathcal{F}_n , d'où le fait que, $\mathbb{P}_\nu - p.s.$,

$$\mathbb{E}_\nu(f(X_{T \wedge (n+1)}) | \mathcal{F}_n) \mathbf{1}(T \leq n) = f(X_{T \wedge n}) \mathbf{1}(T \leq n).$$

\square

On note que le calcul ci-dessus revient à considérer la martingale $(M_n)_{n \geq 0}$ de l'exercice 28 et à exploiter le fait que $(M_{T \wedge n})_{n \geq 0}$ est nécessairement une martingale.

Corollaire 10 *Supposons que f est positive et satisfait $\mathbb{E}_\nu(f(X_0)) < +\infty$, ou que f est bornée, et est solution de l'équation de Poisson avec frontière absorbante (3.4). Si $c \equiv 0$, c'est-à-dire dans le cas d'une fonction harmonique, la suite $[f(X_{T \wedge n})]_{n \geq 0}$ est une martingale par rapport à la filtration $(\mathcal{F}_n)_{n \geq 0}$. Dans le cas général de l'équation de Poisson, on a seulement le fait que f est sur-harmonique, et que $[f(X_{T \wedge n})]_{n \geq 0}$ est une sur-martingale.*

Preuve :

Il suffit d'appliquer le résultat du théorème ci-dessus, qui permet de vérifier facilement par récurrence, dans le cas où f est supposée positive mais pas nécessairement bornée, que $\mathbb{E}_\nu(f(X_{T \wedge n})) < +\infty$ pour tout n et que l'on a la propriété de martingale (ou de sur-martingale). \square

3.1.3 Questions d'unicité

Les fonctions de Green introduites ci-dessus nous ont permis de fabriquer une solution $u = u_D + u_{\partial D}$ à l'équation de Poisson avec frontière absorbante. Une question naturelle est alors de demander si cette fonction est l'unique solution de l'équation de Poisson, ou s'il peut en exister d'autres.

Avant de discuter le problème de manière générale, analysons le cas de la marche aléatoire simple sur \mathbb{Z} , pour lequel l'étude des fonctions harmoniques se ramène à celle de suites récurrentes linéaires d'ordre 2.

Exercice 96 Soit le noyau sur \mathbb{Z} défini par $p(x, x+1) := \alpha$ et $p(x, x-1) := 1 - \alpha$, où $\alpha \in]0, 1[$.

- 1) Montrer que, lorsque $\alpha \neq 1/2$, il existe une fonction harmonique positive non-constante.
- 2) Montrer que, lorsque $\alpha = 1/2$, les seules fonctions harmoniques positives sont les constantes, mais qu'il existe néanmoins des fonctions harmoniques non-constantes.

Sur cet exemple, on constate donc qu'il n'y a pas nécessairement unicité pour le problème de Dirichlet (et donc pour l'équation de Poisson en général), même si l'on impose à la solution de ne prendre que des valeurs positives. En revanche, la solution $u_D + u_{\partial D}$ construite dans la section précédente possède la propriété très importante d'être une solution positive minimale de l'équation, ce qui est précisé par le théorème suivant.

Théorème 7 *Etant donnée une fonction h positive vérifiant*

$$\begin{cases} h(x) \geq ph(x) + c(x), & x \in D; \\ h(x) \geq \phi(x), & x \in \partial D. \end{cases},$$

on a nécessairement

$$h \geq u_D + u_{\partial D}.$$

Preuve :

On introduit les versions tronquées de u_D et $u_{\partial D}$ suivantes. Pour tout $K \geq 0$,

$$u_D^K(x) := \sum_{y \in D} c(y) \left(\sum_{n=0}^{K-1} \mathbb{P}_x(X_n = y, T > n) \right),$$

et

$$u_{\partial D}^K(x) := \sum_{y \in \partial D} \phi(y) \left(\sum_{n=0}^{K-1} \mathbb{P}_x(X_n = y, T = n) \right).$$

En reprenant les calculs de fonction de Green effectués précédemment, on vérifie facilement que $u_D^{K+1}(x) = (pu_D^K)(x) + c(x)$ pour tout $x \in D$, tandis que $u_{\partial D}^{K+1}(x) = 0$ sur ∂D , tandis que $u_{\partial D}^{K+1}(x) = (pu_{\partial D}^K)(x)$ pour tout $x \in D$ et $u_{\partial D}^K(x) = \phi(x)$ sur ∂D . En posant $u^K(x) := u_D^K(x) + u_{\partial D}^K(x)$, on en déduit que $u^{K+1}(x) = (pu^K)(x) + c(x)$ pour tout $x \in D$, tandis que $u^K(x) = \phi(x)$ sur ∂D . Par ailleurs, on voit facilement à partir de la définition que, pour tout $x \in S$,

$$\lim_{K \rightarrow +\infty} u_D^K(x) = u_D(x) \text{ et } \lim_{K \rightarrow +\infty} u_{\partial D}^K(x) = u_{\partial D}(x),$$

d'où le fait que

$$\lim_{K \rightarrow +\infty} u^K(x) = u_D(x) + u_{\partial D}(x).$$

Supposons donnée une fonction h vérifiant les hypothèses du théorème. On a $h \geq u^0 = 0$. Ensuite, par récurrence, on voit que si l'on a $h \geq u^K$, on doit avoir, pour $x \in D$, $h(x) \geq ph(x) + c(x) \geq pu^K(x) + c(x) = u^{K+1}(x)$. Par ailleurs, si $x \in \partial D$, on a $h(x) \geq \phi(x) = u_{\partial D}^K(x)$. On en déduit que $h \geq u^K$ pour tout K , d'où $h \geq u_D + u_{\partial D}$ en passant à la limite. \square

Nous avons vu que la positivité ne suffisait pas en général pour obtenir l'unicité. Il est donc nécessaire d'introduire des hypothèses supplémentaires. Un résultat général d'unicité est le théorème suivant.

Théorème 8 *Si l'on suppose que $\mathbb{P}_x(T < +\infty) = 1$ pour tout $x \in D$, et qu'il existe une solution bornée de l'équation de Poisson, celle-ci est nécessairement égale à $u_D + u_{\partial D}$.*

Preuve :

On se donne une solution h bornée, et on utilise le théorème 6 pour déduire le fait que

$$M_n := h(X_{T \wedge n}) - h(X_0) + \sum_{k=0}^{n-1} c(X_{T \wedge k}) \mathbf{1}(T > k)$$

est une martingale. On a donc $\mathbb{E}_x(M_n) = \mathbb{E}_x(M_0) = 0$, d'où le fait que, pour tout $x \in S$,

$$h(x) = \mathbb{E}_x(h(X_{T \wedge K})) + \mathbb{E}_x \left(\sum_{k=0}^{(T-1) \wedge K} c(X_k) \right).$$

Lorsque K tend vers l'infini, le premier terme du membre de droite de l'équation ci-dessus converge vers $\mathbb{E}_x(h(X_T))$, par convergence dominée, du fait que h est supposée bornée et T fini presque sûrement. Comme $h \equiv \phi$ sur ∂D , $\mathbb{E}_x(h(X_T)) =$

$\mathbb{E}_x(\phi(X_T)) = u_{\partial D}(x)$. Le deuxième terme du membre de droite de l'équation ci-dessus converge, quant à lui, vers $\mathbb{E}_x\left(\sum_{k=0}^{T-1} c(X_k)\right) = u_D(x)$, par convergence monotone, en utilisant l'hypothèse que $c \geq 0$. On en déduit finalement que $h(x) = u_D(x) + u_{\partial D}(x)$. \square

Exercice 97 *Re-prouver le théorème 7 en utilisant la méthode de preuve du théorème ci-dessus.*

Corollaire 11 *SI l'ensemble S est fini et si le noyau est irréductible, l'équation de Poisson possède toujours une unique solution donnée par $u_D + u_{\partial D}$.*

On note que le théorème ne garantit pas en général l'existence de solutions bornées : sous les hypothèses du théorème, il existe une solution bornée si et seulement si $u_D + u_{\partial D}$ est bornée. Du fait que la valeur sur la frontière est imposée par l'équation, il ne peut exister de solution bornée que si la condition au bord ϕ est elle-même bornée. Dans ce dernier cas, on voit, à partir de la définition, que $u_{\partial D}$ est automatiquement bornée, ce qui ramène la question de l'existence de solutions bornées à l'étude du caractère borné ou non de u_D . En particulier, si l'on considère le problème de Dirichlet, $u_D \equiv 0$, et l'on a donc le résultat suivant.

Corollaire 12 *Lorsque ϕ est bornée, le problème de Dirichlet possède une et une seule solution bornée, donnée par $u_{\partial D}$.*

En revanche, lorsque c n'est pas identiquement nulle, il est possible qu'il n'existe aucune solution bornée de l'équation (rappelons qu'il suffit de vérifier que u_D n'est pas bornée pour que cela soit le cas). C'est bien entendu le cas si c n'est pas bornée, car l'existence d'une solution bornée de l'équation entraîne automatiquement, par différence, que c doit être bornée, mais il est possible qu'il n'existe pas de solution bornée même si c l'est.

Exercice 98 *Donner un exemple pour lequel u_D n'est pas bornée bien que c le soit.*

Par ailleurs, on note que, si l'on supprime l'hypothèse que la probabilité d'absorption par la frontière est égale à 1, le résultat d'unicité du théorème tombe en défaut. En effet, en prenant $\phi \equiv 1$ et $c \equiv 0$, on voit que $u_{\partial D}(x) := \mathbb{P}_x(T < +\infty)$, et que $u_{\partial D}$ n'est pas constante car $u_{\partial D}$ vaut 1 sur ∂D et l'on suppose qu'il existe un $x \in D$ tel que $u_{\partial D}(x) < 1$. Comme la fonction constante égale à 1 est aussi une solution bornée, on n'a pas unicité.

L'exercice suivant fournit une preuve alternative à l'unicité dans le cas où l'ensemble S est fini, en s'appuyant sur une version discrète du principe du maximum pour les fonctions harmoniques (*stricto sensu*, il s'agit plutôt un principe du minimum pour les fonctions sur-harmoniques).

Exercice 99 Supposons donné un ensemble fini S , une partition de S en $D \cup \partial D$, et un noyau p sur S tel que $\mathbb{P}_x(T < +\infty) = 1$ pour tout $x \in D$. On suppose que $h : S \rightarrow \mathbb{R}$ vérifie $h(x) \geq ph(x)$ pour tout $x \in D$, et $h(x) \geq 0$ sur ∂D . Nous cherchons à montrer que le minimum de h sur S est nécessairement atteint sur ∂D . Pour cela, on considère x_* tel que $h(x_*) = \min h$, et l'on suppose que $x_* \in D$. On introduit ensuite $D_0 := \{x_*\}$, et on définit par récurrence, pour tout $i \geq 1$, $S_i := \{y \in S; \exists x \in D_{i-1}, p(x, y) > 0\}$.

- 1) Montrer que, pour tout i , h est constante sur S_i égale à $\min h$.
- 2) Montrer qu'il existe nécessairement $i \geq 0$ tel que $S_i \cap \partial D \neq \emptyset$.
- 3) Conclure que le minimum de h est atteint sur ∂D .
- 4) En déduire qu'il ne peut exister qu'une solution à l'équation de Poisson.

Exercice 100 On s'intéresse à l'équation de Poisson modifiée suivante :

$$u(x) = \alpha pu(x) + c,$$

où $0 < \alpha < 1$ est un nombre réel fixé, et $c : S \rightarrow \mathbb{R}$ est une fonction donnée. Montrer que, si l'on suppose c bornée, l'unique solution de l'équation est fournie par la formule

$$u(x) := \mathbb{E}_x \left(\sum_{n=0}^{+\infty} \alpha^n c(X_n) \right).$$

Exercice 101 Supposons donné un ensemble fini S , une partition de S en $D \cup \partial D$, et un noyau p sur S tel que $\mathbb{P}_x(T < +\infty) = 1$ pour tout $x \in D$. On suppose que $h : S \rightarrow \mathbb{R}$ vérifie $h(x) \geq ph(x)$ pour tout $x \in D$, et $h(x) \geq 0$ sur ∂D . Nous cherchons à montrer que le minimum de h sur S est nécessairement atteint sur ∂D . Pour cela, on considère x_* tel que $h(x_*) = \min h$, et l'on suppose que $x_* \in D$. On introduit ensuite $D_0 := \{x_*\}$, et on définit par récurrence, pour tout $i \geq 1$, $S_i := \{y \in S; \exists x \in D_{i-1}, p(x, y) > 0\}$.

- 1) Montrer que, pour tout i , h est constante sur S_i égale à $\min h$.
- 2) Montrer qu'il existe nécessairement $i \geq 0$ tel que $S_i \cap \partial D \neq \emptyset$.
- 3) Conclure que le minimum de h est atteint sur ∂D .
- 4) En déduire qu'il ne peut exister qu'une solution à l'équation de Poisson.

3.1.4 Quelques exemples classiques

L'objectif de ces notes n'étant pas d'approfondir significativement l'étude de la théorie du potentiel pour les chaînes de Markov, nous nous contentons, dans cette partie, de donner quelques exemples classiques de quantités possédant ayant à la fois une interprétation probabiliste simple et un sens dans le cadre de la théorie du potentiel. Le point le plus important pour nous est la possibilité d'étudier ces quantités au moyen de systèmes d'équations linéaires.

Exercice 102 Etant donné deux sous-ensembles A et B de S disjoints et non-vides, appelons T_A le temps d'atteinte de A et T_B le temps d'atteinte de B . Pour tout $x \in S$, posons $u(x) := \mathbb{P}_x(T_A < T_B)$. On a donc $u \equiv 1$ sur A et $u \equiv 0$ sur B . En posant $D := S \setminus (A \cup B)$, montrer que u est harmonique sur D , i.e. $(pf)(x) = f(x)$ pour tout $x \in D$. Que peut-on dire du caractère (sur-, sous-)harmonique de u sur S tout entier ?

Exercice 103 Etant donné un sous-ensemble A de S , appelons T_A le temps d'atteinte de A . Pour tout $x \in S$, posons $v(x) := \mathbb{E}_x(T_A)$. On a donc $v \equiv 0$ sur A . En posant $D := S \setminus A$, montrer que v vérifie l'équation $v(x) = pv(x) + 1$ pour tout $x \in D$.

Exercice 104 Etant donné un ensemble fini S , proposer une méthode permettant de calculer les quantités étudiées dans les trois exercices précédents.

Exercice 105 (La ruine du joueur) On considère la marche simple sur \mathbb{Z} , associée au noyau définie par $p(x, x+1) = \alpha$ et $p(x, x-1) = 1 - \alpha$ pour tout $x \in \mathbb{Z}$, avec $0 < \alpha < 1$.

Etant donné $a, b \in \mathbb{Z}$ tels que $a < b$, on définit T comme le temps d'atteinte de l'ensemble $\{a, b\}$, soit $T := \inf\{n \geq 0; X_n \in \{a, b\}\}$.

1) Montrer (élémentairement) que, partant de $a \leq x \leq b$, $\mathbb{P}_x(T < +\infty) = 1$ et même $\mathbb{E}_x(T) < +\infty$.

2) En résolvant l'équation de Poisson associée, donner, pour tout $a \leq x \leq b$ la valeur de $\mathbb{E}_x(T)$, et de $\mathbb{P}_x(X_T = a)$ et $\mathbb{P}_x(X_T = b)$.

3) En faisant tendre a ou b vers $\pm\infty$, retrouver le critère de récurrence/transience de la marche. Dans le cas transient, comment s'exprime la probabilité de revenir en un point ?

4) Retrouver ces résultats en faisant directement appel aux martingales suivantes :

$$[X_n - n(2\alpha - 1)]_{n \geq 0}, \left[\left(\frac{1 - \alpha}{\alpha} \right)^{X_n} \right]_{n \geq 0} \text{ pour } \alpha \neq 1/2, [X_n^2 - n]_{n \geq 0} \text{ pour } \alpha = 1/2.$$

Quel rapport ces martingales entretiennent-elles avec la résolution de l'équation de Poisson de la question 2) ?

Exercice 106 On reprend la marche simple de l'exercice précédent, et l'on note $T(0)$ le premier temps d'atteinte de 0. Pour $\alpha \leq 1/2$, et $x \geq 1$, donner la valeur de $\mathbb{E}_x(\exp(-\lambda T))$ pour tout $\lambda > 0$.

Exercice 107 (Processus de naissance et de mort en temps discret) On considère un noyau de transition défini sur \mathbb{N} , tel que $p(x, y)$ est nul à moins que $y \in \{x - 1, x, x + 1\}$. On introduit les notations $\alpha_x := p(x, x - 1)$ (avec la convention $\alpha_0 := 0$), $\beta_x := p(x, x + 1)$ et $\gamma_x := p(x, x)$.

1) A quelle condition sur les α, β, γ le noyau est-il irréductible sur \mathbb{N} ?

On introduit la fonction g définie sur \mathbb{N} par

$$g(n) := \sum_{m=0}^{n-1} \prod_{j=1}^m \frac{\alpha_j}{\beta_j}.$$

Avec les conventions usuelles $\sum_{\emptyset} = 0$ et $\prod_{\emptyset} = 1$, on constate que $g(0) = 0$ et $g(1) = 1$.

2) Pour tous $a, b \in \mathbb{N}$ tels que $a < b$, et tout $a \leq x \leq b$, montrer que

$$\mathbb{P}_x(T(a) < T(b)) = \frac{g(b) - g(x)}{g(b) - g(a)},$$

et

$$\mathbb{P}_x(T(b) < T(a)) = \frac{g(x) - g(a)}{g(b) - g(a)}.$$

3) En déduire que, sous l'hypothèse que le noyau est irréductible, il y a récurrence si et seulement si

$$\sum_{m=0}^{+\infty} \prod_{j=1}^m \frac{\alpha_j}{\beta_j} = +\infty.$$

Est-il surprenant que les γ_j n'interviennent pas dans ce critère ?

4) Dans le cas transient, comment s'exprime $\mathbb{P}_0(T(0) < +\infty)$?

5) Discuter de la récurrence/transience lorsque $\gamma_j \equiv 0$ et que β_j est de la forme $\beta_j = 1/2 + \epsilon_j$ avec $\epsilon_j \sim Aj^{-u}$ pour $A, u > 0$.

3.2 Mesures et lois invariantes

La théorie du potentiel décrite précédemment a pour objet central l'étude des fonctions (sur-)harmoniques. Dans cette partie, nous nous intéressons à la notion correspondante pour les mesures, à savoir la notion de mesure (sous-)invariante. Comme le verrons dans la suite, les mesures invariantes, et en particulier les mesures de probabilité invariantes, jouent en fait un rôle fondamental dans l'étude du comportement asymptotique de la dynamique associée à une chaîne de Markov.

Dans ce qui suit, toutes les mesures sont supposées positives sauf mention contraire (nous employons donc simplement le terme de "mesure", pour désigner une mesure positive).

Etant donnée un ensemble fini ou dénombrable S , un noyau de transition $p(\cdot, \cdot)$, une mesure positive μ sur S , on dit que μ est **invariante par rapport à p** , ou encore **stationnaire vis-à-vis de p** , lorsque $\mu p = \mu$. On parle de mesure **sous-invariante** lorsque $\mu p \leq \mu$.

Dans le cas d'une **mesure de probabilité invariante**, on parle plus volontiers de **loi invariante**.

Deux mesures invariantes triviales sont la mesure constamment égale à 0 et la mesure constamment égale à $+\infty$. Toute autre mesure invariante sera dite **non-triviale**. Par ailleurs, nous appellerons **propre** une mesure positive sur S qui vérifie $0 < \mu(x) < +\infty$ pour tout $x \in S$. On a en fait le résultat suivant.

Proposition 28 *Si $p(\cdot, \cdot)$ est irréductible, toute mesure sous-invariante non triviale est propre.*

Preuve :

Supposons qu'il existe x tel que $\mu(x) = 0$. Par sous-invariance, on a, pour tout $n \geq 0$, $\sum_{y \in S} \mu(y) p^n(y, x) \leq \mu(x)$. Par positivité de μ , on a donc $\mu(y) = 0$ dès que $p^n(y, x) = 0$. Par irréductibilité, pour tout $y \in S$ existe un n tel que ce soit le cas. On raisonne de même en supposant que $\mu(x) = +\infty$ pour un x . \square

Avant de donner davantage de détails, mentionnons dès maintenant quelques conséquences, immédiates, mais importantes, de l'existence d'une mesure de probabilité invariante.

Loi invariante sur l'espace des trajectoires

Dans le cas où ν est une mesure de probabilité invariante, on voit que, par définition, $\mathbb{P}_{\nu, p}(X_1 = x) = \nu(x)$ pour tout $x \in S$, et, plus généralement, que $\mathbb{P}_{\nu, p}(X_m = x) = \nu(x)$ pour tout $m \geq 1$. En d'autres termes, si $(X_n)_{n \geq 0}$ est une chaîne de Markov de loi initiale ν et de noyau de transition p , pour tout $m \geq 1$, on a que $\text{loi}(X_m) = \nu$.

En se plaçant sur l'espace des trajectoires, on a en fait la proposition suivante.

Proposition 29 *Si ν est une mesure de probabilité invariante par rapport à p , la mesure $\mathbb{P}_{\nu, p}$ est invariante sur $(S^{\mathbb{N}}, \mathcal{H}^{\otimes \mathbb{N}})$ sous l'action des décalages θ_k , $k \geq 1$. En d'autres termes, $\mathbb{P}_{\nu, p}(\theta_k^{-1}(A)) = \mathbb{P}_{\nu, p}(A)$ pour tout $A \in \mathcal{H}^{\otimes \mathbb{N}}$, ou, en termes plus probabilistes : si $(X_n)_{n \geq 0}$ est une chaîne de Markov de loi initiale ν et de noyau de transition p , $P((X_0, X_1, \dots) \in A) = P((X_k, X_{k+1}, \dots) \in A)$ pour tout $k \geq 1$.*

Exercice 108 *Prouver la proposition ci-dessus.*

Ce résultat permet de voir une chaîne de Markov initialisée selon une loi de probabilité invariante comme un **système dynamique mesuré** sur l'espace des trajectoires, et autorise l'emploi des outils de la **théorie ergodique** pour l'étude d'une telle chaîne. Les résultats de théorie ergodiques qui nous seront nécessaires sont rappelés dans un appendice.

Loi invariante et opérateur linéaire associé au noyau

Commençons par un résultat concernant l'action de p (à droite) sur les fonctions.

Proposition 30 *Si ν est une loi de probabilité invariante associée au noyau p , pour tout nombre réel $s \geq 1$, et tout $f \in L^s(\nu)$, pf est bien défini et fini, et $pf \in L^s(\nu)$. De plus, l'action sur les fonctions*

$$p : L^s(\nu) \rightarrow L^s(\nu)$$

définit un opérateur linéaire continu et de norme 1.

Exercice 109 *Prouver la proposition. (Indication : inégalité de Jensen.)*

Le corollaire suivant laisse augurer du rôle de ν dans le comportement à long terme de la chaîne :

Corollaire 13 *Sous les hypothèses de la proposition ci-dessus, si $f \in L^s(\nu)$ avec $s \geq 1$,*

$$\|pf - \nu(f)\|_s \leq \|f - \nu(f)\|_s.$$

Concernant l'action sur les mesures, notons tout d'abord que l'invariance de ν signifie que ν est un vecteur propre associé à la valeur propre 1, par exemple pour l'action sur $\ell^1(S)$. (Noter que la fonction constante égale à 1 est, quant à elle, un vecteur propre associé à l'action duale sur les fonctions de $\ell^\infty(S)$.) Enonçons maintenant une contrepartie à la proposition 30 pour l'action (à gauche) de p sur les mesures. En fait, on fait agir p à gauche sur la densité des mesures par rapport à ν .

Proposition 31 *Si μ/ν (vue comme une fonction sur S) est dans $L^s(\nu)$ pour $s \geq 1$, alors (μp) est bien défini, et $(\mu p)/\nu \in L^s(\nu)$. L'action correspondante de p de $L^s(\nu)$ dans lui-même définit un opérateur linéaire continu et de norme 1.*

On a également le corollaire :

Corollaire 14 *Sous les hypothèses de la proposition ci-dessus, si $\mu/\nu \in L^s(\nu)$ avec $s \geq 1$,*

$$\|(\mu p)/\nu - 1\|_s \leq \|\mu/\nu - 1\|_s.$$

3.2.1 Renversé dans le temps d'un noyau par rapport à une mesure invariante

Etant donné un noyau de transition irréductible p , et une mesure invariante non-triviale ν de p , le noyau renversé dans le temps de p par rapport à ν est défini, pour tous $x, y \in S$, par :

$$\hat{p}(x, y) = p(y, x) \frac{\nu(y)}{\nu(x)}. \quad (3.11)$$

Comme ν est propre, cette définition est licite.

Proposition 32 *Tel que défini ci-dessus, \hat{p} est un noyau de transition irréductible vérifiant la propriété suivante : pour toute suite x_0, \dots, x_n d'éléments de S*

$$\nu(x_0)p(x_0, x_1) \times p(x_{n-1}, x_n) = \nu(x_n)\hat{p}(x_n, x_{n-1}) \times \hat{p}(x_1, x_0).$$

Exercice 110 *Prouver la proposition ci-dessus.*

Proposition 33 *Avec les notations ci-dessus, ν est invariante pour \hat{p} , et le renversé dans le temps de \hat{p} par rapport à ν n'est autre que p .*

Exercice 111 *Prouver la proposition ci-dessus.*

Exercice 112 *On considère un noyau de transition irréductible p possédant une loi de probabilité invariante ν .*

1) *Prouver que, pour tout $n \geq 0$, la loi de (X_n, \dots, X_0) sous $\mathbb{P}_{\nu, p}$ est identique à la loi de (X_0, \dots, X_n) sous $\mathbb{P}_{\nu, \hat{p}}$.*

2) *Montrer que, étant donnée une chaîne de Markov $(X_n)_{n \geq 0}$ de loi initiale ν et de noyau de transition p , il est possible, quitte à enrichir l'espace de probabilité sous-jacent, de définir une suite de variables aléatoires $(X_n)_{n \leq 0}$ telle que $(X_n)_{n \in \mathbb{Z}}$ soit une chaîne de Markov de noyau de transition p (i.e. telle que toute sous-suite finie $(X_n)_{a \leq n \leq b}$ soit une telle chaîne de Markov).*

Proposition 34 *La récurrence de p est équivalente à celle de \hat{p} .*

Preuve :

Il suffit par exemple de noter que, pour tous $x, y \in S$ et $n \geq 0$, $\nu(x)p^n(x, y) = \nu(y)\hat{p}^n(y, x)$, ou de considérer les probabilités de boucles partant de x et se terminant en x sans retoucher x entre temps. \square

L'un des intérêts de \hat{p} est de fournir une dualité entre les actions sur les mesures et les fonctions, comme l'explique la proposition suivante.

Proposition 35 *Pour toute mesure positive μ sur S , on a l'identité*

$$\frac{(\mu p)}{\nu} = \hat{p} \left(\frac{\mu}{\nu} \right).$$

Preuve :

Par définition, on a, pour tout $x \in S$,

$$\sum_{y \in S} \hat{p}(x, y) \frac{\mu(y)}{\nu(y)} = \sum_{y \in S} p(y, x) \frac{\mu(y)}{\nu(x)} = \frac{1}{\nu(x)} \sum_{y \in S} p(y, x) \mu(y).$$

\square

Exercice 113 *Prouver la proposition 31 ainsi que son corollaire à partir de la proposition 30 et de la dualité ci-dessus.*

Comme conséquence, on obtient par exemple la dualité suivante entre fonctions (sur-)harmoniques et mesures (sous-)invariantes.

Proposition 36 *Etant donné une fonction positive g telle que $0 < g(x) < +\infty$ pour tout $x \in S$, il y a équivalence entre les deux propriétés suivantes :*

- g est harmonique (resp. sur-harmonique) pour \hat{p}
- la mesure définie sur S par $\mu(x) := g(x)\nu(x)$ est invariante (resp. sous-invariante) pour p .

Preuve :

Conséquence directe de la proposition 35. □

3.2.2 Mesures invariantes et récurrence/transience

Comme nous allons le voir, la question de l'existence et de l'unicité des mesures (sous-)invariantes est liée à la question du caractère récurrent/transient de la chaîne considérée.

Nous allons d'abord voir que, dans le cas d'une chaîne récurrente, les fonctions de Green introduites dans la section précédente permettent de fabriquer des mesures invariantes.

Définition 6 *Etant donné $a \in S$, considérons le découpage de S en $D := S \setminus \{a\}$ et $\partial D := \{a\}$. On définit une mesure positive sur S par $\mu_a^{cycle} := G_{\partial D \rightarrow}(a, \cdot)$.*

Redonnons l'expression de $G_{\partial D \rightarrow}$ dans le cas particulier ci-dessus :

$$\mu_a^{cycle}(x) = \mathbb{E}_a \left[\sum_{i=0}^{T_1(a)-1} \mathbf{1}(X_i = x) \right] = \sum_{i=0}^{+\infty} \mathbb{P}_a(X_i = x, T_1(a) > i), \quad (3.12)$$

et notons que

$$\mu_a^{cycle}(a) = 1.$$

Proposition 37 *Supposons $p(\cdot, \cdot)$ est irréductible et récurrent. Alors la mesure positive définie sur S par $\mu_a^{cycle} := G_{\partial D \rightarrow}(a, \cdot)$ est non-triviale et invariante pour $p(\cdot, \cdot)$.*

Preuve :

En vertu de la section précédente, nous savons déjà que $\mu_a^{cycle} p(y) = \mu_a^{cycle}(y)$ pour tout $y \in S \setminus \{a\}$. Par ailleurs, l'identité (3.10) montre que la validité de l'identité pour $y = a$ équivaut à la validité de l'identité pour la fonction duale, soit

$p[G_{\rightarrow\partial D}(\cdot, a)](a) = G_{\rightarrow\partial D}(\cdot, a)$. Rappelons l'expression de $G_{\rightarrow\partial D}(x, y)$ pour $x \notin \partial D$ et $y \in \partial D$:

$$\mathbb{P}_x(X_T = y, T < +\infty),$$

où, dans notre contexte, T est le premier temps d'atteinte de a . Du fait de la récurrence de la chaîne, on a nécessairement que $X_T = a$ avec probabilité 1, d'où le fait que $G_{\rightarrow\partial D}(\cdot, a)$ est la fonction constante égale à 1. La fonction $G_{\rightarrow\partial D}(\cdot, a)$ est donc évidemment harmonique, ce qui entraîne le résultat voulu. \square

On note que la preuve nous permet de comprendre pourquoi on choisit un ensemble ∂D réduit à un seul point : c'est le moyen de garantir le fait que la fonction $G_{\rightarrow\partial D}(\cdot, a)$ est harmonique sur S tout entier et pas seulement sur D .

Proposition 38 *Si p est irréductible et transient, l'égalité caractérisant l'invariance de μ_a^{cycle} est satisfaite partout, sauf en a . En a , on a $\mu_a^{cycle}(a) = 1$ tandis que $\mu_a^{cycle}p(a) = \mathbb{P}_a(T_1 < +\infty) < 1$. La mesure μ_a^{cycle} n'est donc pas invariante, mais sous-invariante.*

Proposition 39 *La masse totale de μ_a^{cycle} n'est autre que $\mathbb{E}_a(T_1(a))$. Par conséquent, μ_a^{cycle} est de masse totale finie si et seulement si a est récurrent positif.*

La proposition ci-dessus donne une manière simple de vérifier que, si S est fini, toute chaîne irréductible est positivement récurrente.

Après la question de l'existence vient naturellement celle de l'unicité des mesures invariantes. Nous commençons par traiter le problème pour les fonctions (sur-)harmoniques, pour lesquelles les techniques de martingale fournissent une preuve élégante.

Proposition 40 *Si p est irréductible et récurrent, les seules fonctions sur-harmoniques positives (à valeurs finies) pour p sont les fonctions constantes.*

Preuve :

Soit f une fonction sur-harmonique positive pour p . En considérant l'espace canonique des trajectoires, on en déduit que, pour tout $b \in S$, la suite $[f(X_n)]_{n \geq 0}$ est une sur-martingale positive par rapport à la filtration $(\mathcal{F}_n)_{n \geq 0}$ et la probabilité \mathbb{P}_b . Le théorème de convergence de Doob entraîne donc que $f(X_n)$ converge presque sûrement dans \mathbb{R} lorsque n tend vers l'infini. En appelant L la limite (aléatoire) de cette martingale, et en notant le fait que tout élément de S est, par irréductibilité et récurrence de p , visité une infinité de fois, on obtient que $L = f(x)$ pour tout x , et donc que f est une fonction constante. \square

On note que ce résultat n'est pas vrai en général si l'on s'autorise à considérer

des fonctions de signe quelconque. Par exemple, toutes les fonctions affines sont harmoniques pour la marche simple symétrique sur \mathbb{Z} , qui est pourtant récurrente.

Nous pouvons à présent énoncer la contrepartie du résultat ci-dessus pour les mesures.

Proposition 41 *Si $p(\cdot, \cdot)$ est irréductible et récurrent, μ_a^{cycle} est, à une constante multiplicative près, l'unique mesure sous-invariante non-triviale, a étant choisi arbitrairement dans S . (Rappelons que, dans ce cas, μ_a^{cycle} n'est pas seulement sous-invariante mais exactement invariante.)*

Preuve :

Considérons le noyau \hat{p} renversé dans le temps de p par rapport à μ_a^{cycle} , et soit une mesure invariante non-triviale (et donc propre) μ . La fonction g définie sur S par $g(x) = \mu(x)/\mu_a^{cycle}(x)$ est, d'après les propositions précédentes, sur-harmonique pour \hat{p} . Comme p est récurrent, \hat{p} l'est aussi et, par conséquent, g doit être constante. Par conséquent, μ et μ_a^{cycle} sont proportionnelles. Pour une preuve élémentaire, voir par exemple [18]). \square

En particulier, on voit que, pour une chaîne récurrente, toutes les mesures de la forme μ_a^{cycle} , a décrivant S , sont en fait proportionnelles les unes aux autres.

Exercice 114 *(Saines lectures) Lire les preuves élémentaires (sans martingales) présentées dans [18, 7] de l'unicité d'une mesure invariante non-triviale pour un noyau récurrent.*

Pour faire une petite synthèse de ce qui précède :

Proposition 42 *Pour une chaîne irréductible, la récurrence entraîne l'existence et l'unicité (à une constante multiplicative près) d'une mesure invariante non-triviale. Le fait que cette mesure soit de masse finie est équivalent au fait que la chaîne soit positivement récurrente.*

On retrouve ainsi que, pour une chaîne irréductible, la récurrence positive d'un point entraîne celle de tous les autres points : la récurrence positive (i.e. le fait que $\mathbb{E}_x(T_1(x)) < +\infty$) d'un point entraîne la récurrence de la chaîne. Par conséquent, toutes les μ_a^{cycle} sont proportionnelles à μ_x^{cycle} (et propres), donc $\mathbb{E}_a(T_1(a)) < +\infty$ pour tout a .

On retrouve également le fait que toute chaîne irréductible sur un espace d'états fini est positivement récurrente. Dans ce cas, la recherche d'une mesure de probabilité invariante se ramène à la résolution d'un système linéaire de taille finie, que l'on peut, si ses dimensions restent raisonnables, résoudre à l'aide d'un ordinateur (Google utilise par exemple la mesure invariante de la chaîne de Markov définie à l'exercice 19

pour mesurer la popularité d'une page ouaibe, celle-ci étant régulièrement recalculée en résolvant un système linéaire dont la taille était, en 2002, de l'ordre de $10^9 \times 10^9$.

On a également le résultat suivant (qui est également une conséquence des résultats asymptotiques basés sur le théorème de renouvellement, que nous verrons plus loin) :

Proposition 43 *Si p est irréductible, l'existence d'une loi de probabilité invariante entraîne le fait que p est positivement récurrente.*

Preuve :

Soit ν ladite mesure de probabilité. L'invariance s'écrit $\sum_{x \in S} \nu(x) p^n(x, y) = \nu(y)$. En sommant, on obtient que $\sum_{x \in S} \nu(x) \sum_{n \geq 1} p^n(x, y) = +\infty$.

Rappelons à toutes fins utiles que $\sum_{n \geq 1} p^n(x, y) = \mathbb{E}_x(N(y))$. Or, pour $x \neq y$, $\mathbb{E}_x(N(y)) \leq \mathbb{E}_y(N(y)) + 1$ (par exemple par couplage, ou alors, formule exacte de la proposition 8). Ainsi, $\sum_{x \in S} \nu(x) (\mathbb{E}_x(N(y)) + 1) = +\infty$. Comme ν est une mesure de probabilité, on en déduit que $\mathbb{E}_x(N(y)) = +\infty$. Par conséquent, p est récurrent. Par unicité des mesures invariantes non-triviales, ν coïncide forcément avec μ_a^{cycle} normalisée à 1. Par conséquent, μ_a^{cycle} est de masse finie, donc $\mathbb{E}_a(T_1(a))$ est finie, et la chaîne est positivement récurrente. □

Corollaire 15 *Si p est un noyau irréductible possédant une loi de probabilité invariante ν ,*

$$\nu(x) = \frac{\mu_a^{cycle}(x)}{\mathbb{E}_a(T_1(a))}.$$

En faisant $a = x$, on obtient que

$$\nu(x) = \frac{1}{\mathbb{E}_x(T_1(x))}.$$

Nous retrouverons cette formule de manière différente dans la preuve par renouvellement des théorèmes limites.

On retient, entre autres choses, de ce qui vient d'être dit, que l'existence d'une loi de probabilité invariante équivaut (pour une chaîne irréductible) à la récurrence positive, la loi de probabilité invariante étant alors unique. Une question naturelle est alors de savoir si l'existence et l'unicité d'une mesure invariante non-triviale caractérisent la récurrence, et la réponse est négative, comme le montre l'exemple suivant.

Exercice 115 *(Marche simple sur \mathbb{N} avec réflexion en 0) On considère le noyau de transition sur \mathbb{N} suivant : pour $i \geq 1$, $p(i, i - 1) = p$, $p(i, i + 1) = 1 - p$, et*

$p(0, 1) = 1$. On voit facilement en écrivant les équations qu'il existe, à une constante multiplicative près, une unique mesure invariante non-triviale, donnée par $\nu(i) = \left(\frac{1-p}{p}\right)^{i-1}$ pour $i \neq 0$ et $\nu(0) = p$. Pour $p > 1/2$, la chaîne est transiente (par comparaison avec la marche simple non réfléchie, qui est transiente par la loi des grands nombres, et a donc une probabilité positive de ne jamais revenir en 0 en partant par la droite). Pour $p < 1/2$, la mesure obtenue est de masse finie, et on en déduit la récurrence positive. Le cas $p = 1/2$ est récurrent (comparaison avec la marche aléatoire non-réfléchie, ou calcul direct).

On peut en fait caractériser la récurrence en considérant les mesures sous-invariantes plutôt qu'invariantes. On obtient alors le résultat suivant.

Proposition 44 *Pour une chaîne irréductible, la récurrence équivaut à l'existence d'une unique mesure non-triviale sous-invariante.*

Preuve :

La preuve du fait que μ_a^{cycle} est l'unique mesure non-triviale sous-invariante dans le cas récurrent a déjà été faite plus haut. Inversement, supposons que p est transient, et considérons le renversé dans le temps de p par rapport à μ_a^{cycle} , défini par la formule 3.11. Comme μ_a^{cycle} n'est que sous-invariante, la formule donnant, \hat{p} ne définit plus un noyau de transition, mais un sous-noyau de transition. Plus précisément, pour $x \neq a$, on a bien l'égalité $\sum_{y \in S} \hat{p}(x, y) = 1$, mais $\sum_{y \in S} \hat{p}(a, y) = \mathbb{P}_{a,p}(T_1(a) < +\infty) < 1$. Pour remédier à ce problème, on adjoint un état $*$ à S , en définissant $\hat{p}(a, *) = 1 - \sum_{y \in S} \hat{p}(a, y)$, et l'on pose $\hat{p}(*, *) = 1$. On définit bien ainsi un noyau de transition sur $S \cup \{*\}$.

A présent, considérons $u \in S$ différent de a (S doit être infini pour qu'il puisse y avoir transience, donc on peut bien toujours trouver un tel u), et la fonction définie sur $S \cup \{*\}$ par $g_u(x) = \mathbb{P}_{x,\hat{p}}(T_1(u) < +\infty)$ pour $x \neq u$, et $g_u(u) = 1$. On vérifie que g_u est sur-harmonique par rapport à \hat{p} . Qui plus est, g_u n'est pas constante, car $g_u(u) = 1$ tandis que, du fait de la présence du point absorbant $*$, $\mathbb{P}_{a,\hat{p}}(T_1(u) < +\infty) < 1$. On vérifie alors que μ_a^{cycle} et la mesure définie par $\nu(x) = g_u(x)\mu_a^{cycle}(x)$ sont deux mesures non-triviales et sous-invariantes distinctes. \square

Les arguments utilisés pour prouver la proposition précédente permettent facilement de prouver la caractérisation "duale" suivante :

Proposition 45 *Pour une chaîne irréductible la récurrence équivaut au fait que toutes les fonctions sur-harmoniques positives sont constantes.*

Exercice 116 *Prouver la proposition ci-dessus.*

Pour une chaîne irréductible transiente il existe donc toujours au moins deux mesures sous-invariantes non-triviales. Pour ce qui est de l'existence et de l'unicité d'une mesure invariante non-triviale, toutes les situations sont possibles (pas de mesure invariante, une unique mesure invariante, plusieurs mesures invariantes), comme le montrent l'exercice 115, ainsi que les deux premiers exercices ci-dessous. Des conditions nécessaires et suffisantes concernant l'existence d'une mesure invariante pour les chaînes transientes peuvent être trouvées dans les articles [26, 52].

Exercice 117 (*L'échelle de l'ours*) On considère le noyau de transition sur \mathbb{N} suivant : pour $i \geq 0$, $p(i, i+1) = \alpha_i$, et $p(i, 0) = 1 - \alpha_i$, où $0 < \alpha_i < 1$ pour tout i . L'irréductibilité de la chaîne est immédiate. Partant de l'origine, la probabilité de ne jamais y revenir est égale au produit $\beta = \prod_{i=0}^{+\infty} \alpha_i$. Par conséquent, la chaîne est récurrente si $\beta = 0$, et transiente si $\beta > 0$. Étudions l'existence de mesures invariantes non-triviales pour p . L'invariance s'écrit, pour tout $i \geq 0$, $\nu(i+1) = \alpha_i \nu(i)$. Par conséquent, p possède au plus une mesure invariante non-triviale, qui doit vérifier $\nu(i) = \prod_{j=0}^{i-1} \alpha_j \nu(0)$. En 0, l'invariance s'écrit $\nu(0) = \sum_{i=0}^{+\infty} \nu(i)(1 - \alpha_i)$. Donc $\nu(0) = \nu(0) \sum_{i=0}^{+\infty} (1 - \alpha_i) \prod_{j=0}^{i-1} \alpha_j$ (avec la convention $\prod_{j=0}^{-1} \alpha_j = 1$), d'où, par télescopage, $\nu(0) = (1 - \beta)\nu(0)$. Par conséquent, si $\beta > 0$, il ne peut y avoir de mesure invariante non-triviale.

Exercice 118 (*Marche aléatoire simple avec (ou sans) biais sur \mathbb{Z}*) On considère le noyau de transition sur \mathbb{Z} suivant : pour tout i , $p(i, i-1) = p$, $p(i, i+1) = 1 - p$. La mesure constante est invariante par p . C'est également le cas de la mesure définie par $\nu(i) = \left(\frac{1-p}{p}\right)^i$. Pour $p \neq 1/2$, on a donc au moins deux mesures invariantes non-triviales, et c'est également une manière de déduire la transience. Au passage, on note que, sachant que le cas $p = 1/2$ est récurrent, le fait que l'on ait trouvé une mesure invariante de masse infinie entraîne automatiquement que la chaîne doit être récurrente nulle.

Exercice 119 Peut-il exister des mesures de probabilité invariantes pour une marche aléatoire sur le groupe \mathbb{Z}^d ? (Indication : utiliser la transformation de Fourier.) Qu'en conclure quant à la récurrence/transience ?

Exercice 120 (*Processus de naissance et de mort en temps discret II*) On considère un processus de naissance et de mort irréductible sur \mathbb{N} (voir l'exercice 107 pour les notations et les définitions).

1) Montrer que l'on définit une mesure invariante en posant

$$\mu(n) := \prod_{j=1}^n \frac{\beta_{j-1}}{\alpha_j}.$$

2) Montrer qu'il existe une probabilité invariante si et seulement si μ est de masse finie. Est-il surprenant que les γ_j n'interviennent pas dans ce critère ?

2) Discuter la récurrence nulle/positive dans le cas où $\gamma_j \equiv 0$ et α_j est de la forme $\alpha_j = 1/2 + \epsilon_j$ avec $\epsilon_j \sim Aj^{-u}$ pour $A, u > 0$.

3.2.3 Réversibilité

Définition 7 Etant donnée un noyau p sur un ensemble fini ou dénombrable S , on dit qu'une mesure positive ν sur S est réversible vis-à-vis de p si, pour tous $x, y \in S$,

$$\nu(x)p(x, y) = \nu(y)p(y, x).$$

On dit que p est réversible lorsqu'il existe une mesure positive non-triviale réversible.

Proposition 46 Une mesure réversible est nécessairement invariante.

En particulier (mais cela peut aussi se voir directement très facilement), une mesure réversible non-triviale pour un noyau irréductible est nécessairement propre.

Preuve :

On écrit que $(\nu p)(x) = \sum_{y \in S} \nu(y)p(y, x)$. D'un autre côté, $\nu(x) = \sum_{y \in S} \nu(x)p(x, y)$. La réversibilité signifie donc que l'on a égalité terme-à-terme entre ces deux sommes.

□

Remarque 16 Pour traduire l'égalité terme-à-terme figurant dans la preuve ci-dessus, on appelle parfois en anglais la condition définissant la réversibilité «detailed balance condition», par opposition à la condition définissant l'invariance, qui n'est que la «balance condition».

Remarque 17 La définition de la réversibilité est l'occasion de préciser quelque peu la notion vague d'«équilibre» associée à une mesure de probabilité invariante. Si une mesure invariante est réversible, on observera, en initialisant la chaîne avec cette loi, autant de transitions en moyenne de x vers y que de y vers x , pour tout couple x, y d'éléments de S . Lorsqu'une loi invariante n'est pas réversible, cette propriété n'est plus vérifiée, mais cela ne signifie pas qu'il n'y a pas «équilibre», puisque, en initialisant la chaîne avec cette loi, la probabilité d'observer un élément x donné est la même, quelque soit le nombre de pas considéré.

Exercice 121 Considérons la marche aléatoire sur $\mathbb{Z}/n\mathbb{Z}$ définie, pour tout $n \geq 2$ et $0 < p < 1$, par la loi suivante des incréments : $+1$ avec probabilité p , -1 avec probabilité $(1-p)$, $+0$ avec probabilité $1/2$. Montrer que le noyau est réversible pour $p = 1/2$, et non-réversible pour $p \neq 1/2$.

Remarque 18 *La recherche d'une mesure invariante pour une chaîne de Markov peut conduire à des systèmes linéaires difficiles à résoudre. En revanche, la recherche d'une mesure réversible, si elle existe, est nettement plus facile, puisque l'on peut la calculer de proche en proche ! En particulier, pour un noyau irréductible, il ne peut exister au plus qu'une mesure réversible non-triviale, à une constante multiplicative près.*

Exercice 122 *Donner un exemple d'un noyau pour lequel il existe deux mesures invariantes non-triviales distinctes dont l'une est réversible et l'autre ne l'est pas.*

Voici maintenant quelques caractérisations simples, mais importantes, de la réversibilité d'un noyau.

Proposition 47 *Un noyau irréductible est réversible si et seulement si, pour tous $n \geq 2$ et toute suite $(x_1, \dots, x_n) \in S^n$ vérifiant $x_n = x_1$, on a l'identité :*

$$p(x_1, x_2) \cdots p(x_{n-1}, x_n) = p(x_n, x_{n-1}) \cdots p(x_2, x_1).$$

Preuve :

Le fait que la réversibilité entraîne la condition est immédiat. Réciproquement, il suffit de fixer un point arbitrairement, disons a , et une valeur strictement positive arbitraire pour $\nu(a)$, puis d'étendre ν à tout x de la manière suivante : étant donné $x \in S$, il existe par irréductibilité un chemin de probabilité strictement positive x_1, \dots, x_n tel que $x_1 = a$ et $x_n = x$. On pose alors

$$\nu(x) := \nu(a) \frac{p(x_1, x_2) \cdots p(x_{n-1}, x_n)}{p(x_n, x_{n-1}) \cdots p(x_2, x_1)}.$$

On vérifie que ν est effectivement réversible. □

On note que, pour un noyau irréductible, la réversibilité entraîne que $p(y, x) > 0$ dès que $p(x, y) > 0$. Cette remarque fournit un argument simple permettant de montrer que certains noyaux ne sont pas réversibles.

Proposition 48 *Une mesure positive propre ν est réversible si et seulement si le renversé dans le temps \hat{p} de p par rapport à ν est égal à p .*

Preuve :

Immédiat. □

Corollaire 19 *Si ν est une loi réversible pour p , alors, par rapport à la probabilité \mathbb{P}_ν , les lois de (X_0, \dots, X_n) et de (X_n, \dots, X_0) sont identiques.*

Proposition 49 *Si ν est la loi invariante d'un noyau irréductible et positivement récurrent, l'adjoint de p vu comme un opérateur linéaire de $L^2(\nu)$ dans lui-même n'est autre que \hat{p} . Dans ce contexte, la réversibilité de ν signifie que p est un opérateur auto-adjoint.*

Exercice 123 *Prouver la proposition ci-dessus.*

Proposition 50 *Le noyau irréductible p est réversible si et seulement s'il est le noyau d'une marche aléatoire sur le graphe non-orienté $G = (V, E)$ avec $V = S$ et $E = S \times S$ quotienté par la relation d'équivalence identifiant (x, y) et (y, x) , associée à une pondération w , telle que définie dans l'exercice 11.*

Preuve :

Pour vérifier la réversibilité d'une marche aléatoire sur un graphe, il suffit par exemple de vérifier le critère de la proposition 47 portant sur les boucles. D'autre part, étant donné un noyau irréductible p possédant une mesure réversible ν , on vérifie qu'il suffit de poser $w(x, y) = \nu(x)p(x, y)$ pour définir une pondération correspondant à p . \square

Exercice 124 *Prouver que, dans le cas d'une marche aléatoire sur un graphe non-orienté associé à une pondération w , la mesure réversible non-triviale est fournie (à une constante multiplicative près) par la formule*

$$\nu(v) := \sum_{v'; \{v, v'\} \in E} w(\{v, v'\}).$$

Quelle forme prend cette expression dans le cas d'une marche aléatoire pour laquelle tous les poids sont égaux à 1 ?

Les chaînes de Markov réversibles forment une catégorie particulière de chaînes de Markov, pour l'étude desquelles des techniques nombreuses et variées, exploitant de manière importante la réversibilité, peuvent être utilisées. Nous ne ferons pas un exposé systématique des techniques propres au cas réversible, renvoyant pour cela à [1]. Pour ne mentionner que deux de ces techniques, citons l'analogie avec les réseaux électriques, présentée par exemple de manière très simple et très claire dans le splendide petit ouvrage [17], et l'utilisation du caractère autoadjoint de l'action de l'opérateur p dans L^2 .

3.3 Exercices supplémentaires

L'exercice suivant fournit une identité simple, mais très utile, reliant la loi du temps de retour en un point à la loi du temps d'atteinte de ce point en partant de la loi invariante, pour un noyau positivement récurrent.

Exercice 125 Soit p un noyau irréductible et positivement récurrent sur un ensemble fini ou dénombrable S , dont la loi invariante est notée ν . Le but de cet exercice est de prouver l'identité suivante, valable pour tout $x \in S$, et tout $n \geq 1$:

$$\mathbb{P}_\nu(T_1(x) = n) = \nu(x)\mathbb{P}_x(T_1(x) \geq n).$$

1) Vérifier l'identité pour $n = 1$. Quelle identité obtient-on en sommant sur toutes les valeurs de n ?

2) Montrer que, pour tout $n \geq 2$,

$$\mathbb{P}_\nu(T_1(x) = n) = \sum_{y \neq x} \mathbb{P}_\nu(X_1 = y)\mathbb{P}_y(T_1(x) = n - 1).$$

3) En déduire que

$$\mathbb{P}_\nu(T_1(x) = n) = \sum_{y \in S} \mathbb{P}_\nu(X_1 = y)\mathbb{P}_y(T_1(x) = n - 1) - \mathbb{P}_\nu(X_1 = x)\mathbb{P}_x(T_1(x) = n - 1),$$

puis que

$$\mathbb{P}_\nu(T_1(x) = n) = \mathbb{P}_\nu(T_1(x) = n - 1) - \nu(x)\mathbb{P}_x(T_1(x) = n - 1).$$

4) En déduire que, pour tout $n \geq 1$, et tout $k \geq 0$,

$$\mathbb{P}_\nu(T_1(x) = n) = \sum_{j=n}^{n+k} \mathbb{P}_x(T_1(x) = j) + \mathbb{P}_\nu(T_1(x) = n + k + 1).$$

5) Conclure.

Quelques conséquences de l'exercice précédent figurent dans l'exercice ci-dessous.

Exercice 126 On se place sous les mêmes hypothèses que dans l'exercice 125 ci-dessus. Utiliser le résultat de cet exercice pour traiter les questions suivantes.

1) Donner un exemple de noyau positivement récurrent pour lequel il existe un x vérifiant $\mathbb{E}_\nu(T_1(x)) = +\infty$.

2) Montrer que les conditions suivantes sont équivalentes :

- (i) Il existe $x \in S$ tel que $\mathbb{E}_x(T_1(x)^2) < +\infty$;
- (ii) Il existe $x \in S$ tel que $\mathbb{E}_\nu(T_1(x)) < +\infty$;
- (iii) Pour tout $x \in S$, $\mathbb{E}_\nu(T_1(x)) < +\infty$;
- (iv) Pour tout $x \in S$, $\mathbb{E}_x(T_1(x)^2) < +\infty$.

(Indication : pour (ii) \Rightarrow (iii), utiliser le Corollaire 8.

Exercice 127 (Cas d'un ensemble S fini) Lorsque p est un noyau irréductible sur un ensemble S fini, on cherche à prouver directement l'existence et l'unicité d'une mesure de probabilité invariante.

1) Soit $\hat{1}$ le vecteur de \mathbb{R}^S dont toutes les coordonnées sont égales à 1. Montrer que $p\hat{1} = \hat{1}$.

2) En déduire l'existence d'un vecteur non-nul $x = (x_s)_{s \in S} \in \mathbb{R}^S$ tel que $xp = x$.

3) Prouver que, pour un tel vecteur, le vecteur $|x| := (|x_s|)_{s \in S}$ vérifie également $|x|p = |x|$. (Montrer d'abord que $|x|p \leq |x|$ coordonnées par coordonnées, puis sommer).

4) Conclure quant à l'existence d'une probabilité invariante.

5) A présent, considérons $x = (x_s)_{s \in S} \in \mathbb{R}^S$ tel que $px = x$. En considérant s^* tel que $x_{s^*} = \max_{s \in S} x_s$, montrer que toutes les coordonnées de x sont constantes. (C'est une version du principe du maximum.) Comparer le plan de cette preuve avec celle donnée dans le cas général.

Exercice 128 On considère un noyau de transition sur \mathbb{N} vérifiant $p(0,1) = 1$, et, pour tout $i \geq 1$, $p(i, i-1) = p_i$ et $p(i, i+1) = 1 - p_i$. Etudier les mesures invariantes de ce noyau (existence, unicité, finitude).

Exercice 129 Etant donnée une marche aléatoire sur un graphe $G = (V, E)$ telle que définie dans l'exercice 11, montrer que la mesure définie par

$$\mu(x) := \sum_{e \ni x} w(e)$$

est invariante.

Exercice 130 Considérons une application f de S dans lui-même, et un noyau de transition p sur S , irréductible et positivement récurrent. Montrer que, si p est invariant sous l'action de f , c'est également le cas de la loi invariante de p .

Exercice 131 (Théorème arbre-matrice pour les chaîne de Markov, d'après [3]) Considérons un ensemble fini S et un noyau de transition irréductible p sur S . Il existe donc une unique loi invariante ν sur S . Considérons à présent le graphe orienté $G = (V, E)$ obtenu en posant $V = S$ et $E = \{(x, y); p(x, y) > 0\}$. Etant donné $x \in S$, un sous-graphe T de G sera appelé arbre couvrant de G enraciné en x s'il ne contient pas de cycles et si chaque élément de $V \setminus \{x\}$ est le sommet initial d'une et une seule arête de T . Nous noterons $\mathcal{T}(x)$ l'ensemble des arbres couvrants enracinés en x , et $\mathcal{T} := \bigcup_{x \in S} \mathcal{T}(x)$ l'ensemble de tous les arbres couvrants de G . A tout $T \in \mathcal{T}$, on associe un poids $p(T)$ défini de la manière suivante :

$$p(T) := \prod_{(x,y) \in E(T)} p(x, y),$$

où $E(T)$ désigne l'ensemble des arêtes de T . L'objectif de cet exercice est d'établir la remarquable formule suivante pour la loi invariante ν de p : pour tout $x \in S$,

$$\nu(x) = \frac{\sum_{T \in \mathcal{T}(x)} p(T)}{\sum_{T \in \mathcal{T}} p(T)}.$$

Plusieurs preuves de ce résultat existent. Celle qui suit fournit une interprétation probabiliste intéressante.

On considère (voir l'exercice 112) le prolongement d'une chaîne de Markov $(X_n)_{n \geq 0}$ de loi initiale ν et de noyau p en une chaîne de Markov $(X_n)_{n \in \mathbb{Z}}$. On définit ensuite pour tout $n \geq 0$ une variable aléatoire T_n à valeurs dans \mathcal{T} de la manière suivante. La racine de T_n est choisie égale à X_n . Pour $x \in S$, on définit $S_n(x) := \sup\{k \leq n-1; X_k = x\}$, et les arêtes présentes dans T_n sont exactement les arêtes de la forme $(X_{S_n(x)}, X_{S_n(x)+1})$, x décrivant S .

1) Montrer que T_n est bien défini, et mesurable par rapport à la tribu engendrée par les $(X_i)_{-\infty < i \leq n}$.

2) Montrer que l'on passe de T_n à T_{n+1} de la manière suivante : on ajoute l'arête (X_n, X_{n+1}) et l'on supprime l'unique arête issue de X_{n+1} . (Faire un dessin !) En appelant Y_n l'extrémité de cette arête supprimée, montrer que (X_n, X_{n+1}) est l'unique arête de T_{n+1} se terminant en X_{n+1} et que l'on peut atteindre à partir de Y_n en suivant les arêtes de T_{n+1} .

3) Quelle est la loi de X_{n+1} conditionnellement à T_0, \dots, T_n ?

4) Montrer que la suite $(T_n)_{n \geq 0}$ est une chaîne de Markov, irréductible sur \mathcal{T} . Nous noterons q son noyau de transition.

5) On définit le noyau de transition q' sur \mathcal{T} de la manière suivante : partant d'un arbre T enraciné en x , $q'(T, \cdot)$ est la loi de l'arbre obtenu en choisissant un sommet y selon la loi $p(x, \cdot)$, puis en ajoutant à T l'arête (x, y) et en supprimant l'unique arête se terminant en x et que l'on peut atteindre à partir de y en suivant les arêtes de T . Montrer que, pour tous $T, T' \in \mathcal{T}$,

$$p(T)q(T, T') = p(T')q'(T', T).$$

6) En déduire que la loi invariante de q est, à une constante multiplicative près, égale à p , et que q' est le noyau renversé dans le temps de q par rapport à cette loi.

7) Conclure.

Exercice 132 On considère un cavalier se déplaçant sur un échiquier ne comportant pas d'autres pièces, en choisissant uniformément au hasard à chaque pas l'une des positions autorisées par sa règle de déplacement (deux cases le long d'un axe, une le long de l'autre).

1) Montrer que la suite des positions occupées par le cavalier constitue une chaîne de Markov irréductible.

- 2) Montrer que cette chaîne est réversible. Quelle est sa loi invariante ?
 3) En déduire la moyenne du temps de retour du cavalier en son point de départ.

Exercice 133 Montrer que la chaîne de Markov associée à l'urne d'Ehrenfest (voir l'exercice 16) est réversible. Quelle est la loi invariante associée ?

Exercice 134 Montrer que la chaîne de Markov décrite dans l'exercice 18 est réversible. Quelle est la loi invariante associée ?

Exercice 135 (Marche aléatoire sur un arbre de Galton-Watson, d'après [35])

On considère une marche aléatoire sur un arbre de Galton-Watson enraciné. Rappelons qu'un arbre de Galton-Watson est construit de la manière suivante : étant donnée une loi de reproduction μ sur \mathbb{N} , on attribue d'abord à la racine un nombre aléatoire d'enfants de loi μ , qui forment le premier niveau de l'arbre. Ensuite, étant donné le n -ème niveau de l'arbre, on attribue indépendamment à chaque sommet de ce niveau un nombre aléatoire d'enfants de loi μ , pour former le $n + 1$ -ème niveau. Un tel arbre aléatoire étant construit, on considère ensuite la marche aléatoire sur le graphe correspondant, obtenue en attribuant à chaque arête un poids constant égal à 1. Il s'agit donc d'une marche aléatoire en milieu aléatoire, le milieu aléatoire étant fourni par la structure aléatoire de l'arbre sur lequel la marche évolue. Notons T l'arbre de Galton-Watson aléatoire considéré, et $(X_n)_{n \geq 0}$ la marche aléatoire sur T correspondante, initialisée en la racine. Si x désigne un sommet de T , nous noterons $T(x)$ l'arbre obtenu en déplaçant la racine de T en x . Avec cette notation, la notion d'environnement vu de la marche est donnée par la suite d'arbres aléatoires $(T(X_n))_{n \geq 0}$. Nous noterons \mathcal{T} l'ensemble des arbres localement finis (i.e. pour lesquels chaque sommet possède un nombre fini d'enfants enracinés), deux arbres étant considérés comme identiques s'ils ne diffèrent que par un ré-étiquetage des sommets préservant la racine, muni de la tribu engendrée par les sous-arbres finis issus de la racine.

- 1) Montrer que la suite de variables aléatoires $(T(X_n))_{n \geq 0}$ est une chaîne de Markov sur \mathcal{T} . (Attention : on ne travaille pas conditionnellement à la réalisation de T , mais en tenant compte de l'aléa engendrant T et de celui engendrant la marche.)
 2) On définit une probabilité Q sur \mathcal{T} de la manière suivante : on attribue à la racine un nombre aléatoire d'enfants de loi μ_+ définie par $\mu_+(n) := \mu(n + 1)$ pour tout $n \in \mathbb{N}$, puis on continue la construction comme pour l'arbre de Galton-Watson T , en utilisant la loi μ pour générer les nombres d'enfants de tous les sommets issus de la racine. Montrer¹ que Q est réversible pour $(T(X_n))_{n \geq 0}$.

¹Attention : il faut d'abord trouver la généralisation adéquate de la réversibilité dans le cas d'un espace mesurable général.

3) Supposons que l'espérance de μ est inférieure ou égale à 1. L'arbre T est donc fini avec probabilité 1. Quelle est la loi stationnaire de la marche aléatoire conditionnellement à T ? Comment ce résultat se compare-t-il au résultat de la question précédente?

Exercice 136 Si $(X_n)_{n \geq 0}$ est une chaîne irréductible et positivement récurrente, montrer que, pour tout $p \geq 1$, c'est également le cas de la chaîne $(Z_n)_{n \geq 0} := (X_n, \dots, X_{n+p})_{n \geq 0}$ (en définissant convenablement l'espace d'états). Comment s'exprime la loi invariante associée en fonction de celle de $(X_n)_{n \geq 0}$?

Exercice 137 Le but de cet exercice est de montrer que, si $(X_n)_{n \geq 0}$ est une chaîne de Markov irréductible positivement récurrente, la suite des lois des X_n est tendue, c'est-à-dire que, pour tout $\epsilon > 0$, il existe un sous-ensemble fini $A \subset S$ tel que, pour tout n , $P(X_n \in A^c) \leq \epsilon$. (Ici, ce sont les sous-ensembles finis de S qui jouent le rôle des parties compactes dans la définition de la tension.) Nous noterons p le noyau de transition de la chaîne, et ν la loi invariante.

1) En utilisant l'invariance de ν , montrer que, pour tout $\epsilon > 0$, on peut trouver un sous-ensemble fini $A \subset S$ tel que, pour tout $x \in S$, et tout $n \geq 0$, $\sum_x \nu(x) p^n(x, A^c) \leq \epsilon$.

2) Montrer que, pour tout $x \in S$, et tout $n \geq 0$, $P(X_n \in A^c) \leq P(T_1(x) \leq n) + \epsilon/\nu(x)$.

3) Conclure.

Preuve :

Comme ν est invariante, $\nu p^n = \nu$, et l'on voit que, pour tout ϵ , on dispose de A tel que $\sum_x \nu(x) p^n(x, A^c) \leq \epsilon$ pour tout n , d'où, en fixant x , $p^n(x, A^c) \leq \epsilon/\nu(x)$ pour tout n . Ainsi, en utilisant la propriété de Markov, on voit que $P(X_n \in A^c) \leq P(T_1(x) \leq n) + p^n(x, A^c)$. La conclusion en résulte. \square

Chapitre 4

Fonctionnelles additives : loi des grands nombres

Etant donnée une fonction f de S dans \mathbb{R} , et une chaîne de Markov $(X_n)_{n \geq 0}$, on parle de fonctionnelle additive de la chaîne pour désigner les sommes partielles de la forme $S_n(f) = f(X_0) + \dots + f(X_n)$. La connaissance du comportement asymptotique de quantités du type $S_n(f)$ lorsque n tend vers l'infini fournit de précieux renseignements sur les propriétés en temps longs des trajectoires de la chaîne. Pour prendre un exemple très simple, si $f = \mathbf{1}_A$ est la fonction indicatrice d'un sous-ensemble $A \subset S$, $S_n(f)$ n'est autre que le nombre de visites de l'ensemble A effectuées par la chaîne au cours des n premiers pas. Nous verrons que, dans le cas d'une chaîne positivement récurrente, le comportement asymptotique des fonctionnelles additives présente un certain nombre de points communs (mais également des différences !) avec celui des sommes de variables aléatoires i.i.d.

4.1 Résultat principal

Dans ce chapitre, nous nous intéressons au comportement du type loi des grands nombres ; autrement dit, à l'existence de limites de la forme

$$\lim_{n \rightarrow +\infty} \frac{S_n(f)}{n}.$$

Les théorèmes suivants montrent que, lorsque f est une fonction intégrable par rapport à la loi invariante de la chaîne, on obtient un comportement comparable à celui fourni par la loi des grands nombres usuelle pour les sommes de variables aléatoires i.i.d. intégrables.

Théorème 9 *Si p est un noyau irréductible et positivement récurrent de loi invariante ν , alors, pour tout $f \in L^1(\nu)$, et toute loi initiale μ , on a*

$$\lim_{n \rightarrow +\infty} \frac{S_n(f)}{n} = \nu(f), \mathbb{P}_\mu - p.s.$$

On notera que le résultat ci-dessus est valable quelle que soit la loi initiale μ . En revanche, le résultat suivant suppose que la loi initiale est précisément la loi invariante.

Théorème 10 *Si p est un noyau irréductible et positivement récurrent de loi invariante ν , alors, pour tout $f \in L^1(\nu)$, on a convergence de $\frac{S_n(f)}{n}$ vers $\nu(f)$ dans $L^1(\mathbb{P}_\nu)$, autrement dit,*

$$\lim_{n \rightarrow +\infty} \mathbb{E}_\nu \left(\left| \frac{S_n(f)}{n} - \nu(f) \right| \right) = 0.$$

La comparaison avec la loi des grands nombres habituelle pour les sommes de variables aléatoires i.i.d., dont les deux théorèmes ci-dessus constituent une généralisation, explique la nécessité de poser une hypothèse telle que l'intégrabilité de f par rapport à ν .

Une extension simple, mais utile du théorème 10 est fournie par le corollaire suivant.

Corollaire 20 *Si $\mu = g\nu$, g étant bornée supérieurement (par exemple si μ est à support fini), on a également convergence dans $L^1(\mathbb{P}_\mu)$, soit*

$$\lim_{n \rightarrow +\infty} \mathbb{E}_\mu \left(\left| \frac{S_n(f)}{n} - \nu(f) \right| \right) = 0.$$

Exercice 138 *Prouver que le corollaire ci-dessus est effectivement une conséquence du théorème 10.*

4.2 Preuves

Nous présenterons deux approches distinctes permettant de prouver le théorème 9. La première est basée sur la décomposition de renouvellement des trajectoires obtenue en considérant les portions de trajectoires séparant les retours successifs en un point donné. La seconde s'appuie sur la théorie ergodique.

Auparavant, expliquons comment le théorème 10 se déduit du théorème 9.

Preuve du théorème 10 à partir du théorème 9:

On note d'abord que $f(X_n) \in L^1(\mathbb{P}_\nu)$ pour tout n , compte-tenu du fait que la loi de X_n sous \mathbb{P}_ν est ν , et ce, quelle que soit la valeur de n . Ensuite, on procède par troncature. Pour $M \geq 0$, posons

$$f_M(x) := f(x)\mathbf{1}(|f(x)| \leq M).$$

On peut donc écrire

$$S_n(f) = S_n(f_M) + S_n(f - f_M), \quad \nu(f) = \nu(f_M) + \nu(f - f_M).$$

D'après le théorème 9, on a convergence presque sûre de $n^{-1}S_n(f_M)$ vers $\nu(f_M)$, presque sûrement par rapport à \mathbb{P}_ν . Comme f_M est une fonction bornée, la suite $(n^{-1}S_n(f_M))_{n \geq 0}$ est elle-même bornée, et, par conséquent, converge également vers $\nu(f_M)$ dans $L^1(\mathbb{P}_\nu)$. On note ensuite que

$$|S_n(f - f_M)| \leq S_n(|f - f_M|).$$

Compte-tenu de l'invariance de ν , on a $\mathbb{E}_\nu(|f - f_M|(X_n)) = \nu(|f - f_M|)$, et l'on déduit de ce qui précède que, pour tout n ,

$$|\mathbb{E}_\nu S_n(f - f_M)| \leq (n + 1)\nu(|f - f_M|).$$

Le théorème de convergence dominée entraînant le fait que $\lim_{M \rightarrow +\infty} \nu(|f - f_M|) = 0$, on en déduit que $n^{-1}S_n(f - f_M)$ et $\nu(f - f_M)$ tendent vers zéro dans $L^1(\mathbb{P}_\nu)$ lorsque n tend vers l'infini, ce qui achève la preuve. \square

4.2.1 Approche par renouvellement

L'idée de la preuve est d'utiliser une décomposition de la trajectoire selon les retours successifs en un point, en exploitant les propriétés d'indépendance des tronçons de trajectoire apparaissant dans cette décomposition.

Supposons donc donnés un noyau irréductible p , de loi invariante ν , et une fonction f définie sur S et à valeurs réelles, telle que $\nu(|f|) < +\infty$.

Fixons un point $a \in S$. Par récurrence de la chaîne, les $T_i(a)$ sont tous finis avec probabilité 1. Pour $i \geq 1$, posons

$$L_i(f) := \sum_{j=T_i(a)}^{T_{i+1}(a)-1} f(X_j).$$

Par ailleurs, définissons pour tout $n \geq 0$ la variable R_n par

$$R_n := \max\{i \geq 1; T_i(a) \leq n\},$$

(avec la convention $\max \emptyset = 0$). Pour tout $n \geq T_1(a)$, on peut alors écrire une **décomposition de renouvellement de $S_n(f)$** , sous la forme

$$S_n(f) = \sum_{j=0}^{T_1(a)-1} f(X_j) + \sum_{i=1}^{R_n-1} L_i(f) + \sum_{j=T_{R_n}(a)}^n f(X_j). \quad (4.1)$$

On sait que, quelle que soit la loi initiale, la suite $L_i(f)$, $i \geq 1$ est constituée de v.a. i.i.d. de même loi que celle de la variable $\sum_{j=0}^{T_1(a)-1} f(X_j)$ sous \mathbb{P}_a .

Lemme 21 *Du fait de l'intégrabilité de f par rapport à ν , on a nécessairement que $\mathbb{E}_a \left| \sum_{j=0}^{T_1(a)-1} f(X_j) \right| < +\infty$. L'identité suivante est alors vérifiée :*

$$\mathbb{E}_a \left(\sum_{j=0}^{T_1(a)-1} f(X_j) \right) = \mu_a^{cycle}(f).$$

Preuve :

Grâce à l'inégalité triangulaire, on a que

$$\mathbb{E}_a \left| \sum_{j=0}^{T_1(a)-1} f(X_j) \right| \leq \mathbb{E}_a \left(\sum_{j=0}^{T_1(a)-1} |f(X_j)| \right).$$

Ensuite, une interversion (légitime, du fait de la positivité), montre que

$$\mathbb{E}_a \left(\sum_{j=0}^{T_1(a)-1} |f(X_j)| \right) = \sum_{y \in S} |f(y)| \mathbb{E}_a \left[\sum_{j=0}^{T_1(a)-1} \mathbf{1}(X_j = y) \right].$$

Par conséquent,

$$\mathbb{E}_a \left(\sum_{j=0}^{T_1(a)-1} |f(X_j)| \right) = \mu_a^{cycle}(|f|).$$

En utilisant le fait que ν est proportionnelle à μ_a^{cycle} et que $f \in L^1(\nu)$, on obtient l'intégrabilité de $\left(\sum_{j=0}^{T_1(a)-1} f(X_j) \right)$ par rapport à \mathbb{P}_a . L'identité annoncée se déduit alors par le même calcul que ci-dessus. \square

Nous allons maintenant prouver le théorème 9 en étudiant séparément le comportement de chacun des trois termes qui constituent la décomposition de renouvellement (4.1).

Pour commencer, on a, en utilisant simplement le fait que $\mathbb{P}_\mu(T_1(a) < +\infty) = 1$, le fait que

$$\lim_{n \rightarrow +\infty} \frac{\sum_{j=0}^{T_1(a)-1} f(X_j)}{n} = 0, \mathbb{P}_\mu - p.s.$$

Le premier "terme de bord" de la décomposition peut donc être négligé. Ensuite, en utilisant le fait que la suite $(T_{i+1}(a) - T_i(a))_{i \geq 1}$ est i.i.d. et de même loi que $T_1(a)$ sous \mathbb{P}_a , et le fait que $\mathbb{E}_a(T_1(a)) < +\infty$ d'après la récurrence positive, on obtient, en appliquant la loi des grands nombres, que

$$\lim_{K \rightarrow +\infty} \frac{T_K(a)}{K} = \mathbb{E}_a(T_1(a)), \mathbb{P}_\mu - p.s. \quad (4.2)$$

D'autre part, en notant que l'inégalité $n \geq T_i(a)$ entraîne que $R_n \geq i$, on voit facilement que, lorsque n tend vers l'infini,

$$\lim_{n \rightarrow +\infty} R_n = +\infty, \mathbb{P}_\mu - p.s.$$

En partant de l'inégalité, valable par définition, $T_{R_n}(a) \leq n \leq T_{R_n+1}(a)$, on obtient que, pour tout $n \geq 1$,

$$\frac{T_{R_n}(a)}{R_n} \leq \frac{n}{R_n} \leq \frac{T_{R_n+1}(a)}{R_n}.$$

En utilisant (4.2) et le fait que R_n tend vers l'infini avec probabilité 1 sous \mathbb{P}_μ , on déduit que

$$\lim_{n \rightarrow +\infty} \frac{R_n}{n} = \frac{1}{\mathbb{E}_a(T_1(a))}, \mathbb{P}_\mu - p.s.$$

Enfin, le lemme 21 et la loi des grands nombres nous permettent d'établir que

$$\lim_{M \rightarrow +\infty} \frac{\sum_{i=1}^M L_i(f)}{M} = \mu_a^{cycle}(f), \mathbb{P}_\mu - p.s.$$

En combinant cette limite avec la précédente, on en déduit finalement que

$$\lim_{n \rightarrow +\infty} \frac{\sum_{i=1}^{R_n} L_i(f)}{n} = \frac{\mu_a^{cycle}(f)}{\mathbb{E}_a(T_1(a))} = \nu(f), \mathbb{P}_\mu - p.s.$$

Il nous reste à montrer que l'on peut négliger le second "terme de bord" de la décomposition. Pour cela, notons la majoration

$$\left| \sum_{j=T_{R_n}(a)}^n f(X_j) \right| \leq L_{R_n}(|f|).$$

A présent, on peut montrer comme ci-dessus que, avec probabilité 1 sous \mathbb{P}_μ ,

$$\lim_{n \rightarrow +\infty} \frac{\sum_{i=1}^{R_n-1} L_i(|f|)}{n} = \lim_{n \rightarrow +\infty} \frac{\sum_{i=1}^{R_n} L_i(|f|)}{n} = \nu(|f|), \mathbb{P}_\mu - p.s.$$

Par différence, on en déduit que

$$\lim_{n \rightarrow +\infty} \frac{|\sum_{j=T_{R_n}(a)}^n f(X_j)|}{n} = 0, \mathbb{P}_\mu - p.s.$$

La conclusion du théorème en résulte.

Exercice 139 *Cet exercice propose de reprouver des résultats sur les chaînes irréductibles positivement récurrentes à partir des arguments employés dans la preuve du théorème donnée ci-dessus.*

1) *En reprenant l'argumentation de la preuve ci-dessus, montrer (sans utiliser les preuves données au chapitre précédent) que, pour une chaîne irréductible positivement récurrente, toute loi invariante est nécessairement être proportionnelle à μ_a^{cycle} .*

2) *De même, reprendre l'argumentation de la preuve ci-dessus pour montrer que, pour une chaîne irréductible positivement récurrente, $\mu_a^{cycle} / \mathbb{E}_a(T_1(a))$ est une loi de probabilité invariante.*

On note en particulier que l'exercice ci-dessus fournit une autre justification à l'expression de la loi invariante en termes de μ_a^{cycle} que celle qui résulte de la preuve donnée dans le chapitre précédent.

4.2.2 Approche par la théorie ergodique

Pour se placer dans le cadre de la théorie ergodique, on considère l'espace canonique $(S^{\mathbb{N}}, \mathcal{H}^{\otimes \mathbb{N}}, \mathbb{P}_\nu)$. Par invariance de ν , la probabilité \mathbb{P}_ν est invariante sous l'action du décalage θ . Par ailleurs, pour $\omega = (x_0, x_1, \dots) \in S^{\mathbb{N}}$,

$$S_n(f)(\omega) = f \circ X_0(\omega) + f \circ X_0 \circ \theta(\omega) + \dots + f \circ X_0 \circ \theta^n(\omega).$$

Par conséquent, d'après le théorème ergodique de Birkhoff, lorsque n tend vers l'infini $n^{-1}S_n(f)$ converge presque sûrement et dans $L^1(\mathbb{P}_\nu)$ vers l'espérance de $f \circ X_0$ conditionnelle à la tribu des invariants de Ω sous l'action de θ . Nous allons montrer qu'un ensemble (mesurable) invariant est nécessairement de probabilité égale à 0 ou 1, i.e. que θ définit une action ergodique sur $(S^{\mathbb{N}}, \mathcal{H}^{\otimes \mathbb{N}}, \mathbb{P}_\nu)$, ce qui conclura la preuve lorsque la loi initiale est égale à ν . Si A est un tel ensemble invariant, $(x_0, x_1, \dots) \in A$ est équivalent au fait que $(x_1, x_2, \dots) \in A$. Par conséquent, pour tout k , $\mathbb{P}_\nu(A|\mathcal{F}_k) = \mathbb{P}_\nu((\theta^k)^{-1}(A)|\mathcal{F}_k) = \mathbb{P}_{X_k}(A)$. Si $\mathbb{P}_\nu(A) \neq 0$, il existe x tel que $\mathbb{P}_x(A) > 0$. Pour un text x , on a donc que $\mathbb{P}_\nu(A|\mathcal{F}_k) \geq \mathbb{P}_x(A)\mathbf{1}(X_k = x)$. D'après le théorème de Lévy (voir [18, 49]), $\mathbb{P}_\nu(A|\mathcal{F}_k) \rightarrow \mathbf{1}(A)$ \mathbb{P}_ν -presque sûrement lorsque k tend vers l'infini. Or, par récurrence de la chaîne, on voit que, avec probabilité 1 sous \mathbb{P}_ν , $\liminf_{k \rightarrow +\infty} (\mathbf{1}(X_k = x)) = 1$. Par conséquent, $\mathbf{1}(A) = 1$ \mathbb{P}_ν -p.s.

Comme ν est propre, on voit que la convergence p.s. sous ν entraîne la convergence p.s. sous μ pour toute loi initiale μ .

4.3 Exercices

La loi des grands nombres pour les fonctionnelles additives présentée dans ce chapitre implique une importante propriété de régularité des trajectoires, comme expliqué dans l'exercice suivant.

Exercice 140 *On considère un noyau de transition p sur un espace fini ou dénombrable S , que l'on suppose irréductible et positivement récurrent, ν désignant sa loi invariante, et μ une loi initiale quelconque sur S .*

1) *Montrer que, pour tout $A \subset S$,*

$$\lim_{n \rightarrow +\infty} \frac{\text{card} \{0 \leq i \leq n; X_i \in A\}}{n} = \nu(A) \quad \mathbb{P}_\mu - p.s.$$

2) *En déduire que, pour tout $\ell \geq 1$ et tout $x_{0:\ell} \in S^\ell$,*

$$\lim_{n \rightarrow +\infty} \frac{\text{card} \{0 \leq i \leq n; X_{i:i+\ell} = x_{0:\ell}\}}{n} = \nu(x_0) \prod_{i=0}^{\ell-1} p(x_i, x_{i+1}) \quad \mathbb{P}_\mu - p.s.$$

3) En déduire que, pour une chaîne réversible, la proportion asymptotique de transitions d'un état x vers un état y le long d'une trajectoire est égale à la proportion de transitions de y vers x .

Exercice 141 Quelle méthode de simulation la loi des grands nombres pour les fonctionnelles additives suggère-t-elle pour estimer la valeur de $\nu(f)$?

Exercice 142 (Tiré de [1].)

Cet exercice propose de prouver diverses identités intéressantes reliant loi invariante et temps d'atteintes. Dans toute la suite, on suppose donné un noyau de transition p défini sur un ensemble S fini ou dénombrable, irréductible et positivement récurrent.

1) Soit T un temps d'arrêt de $(\mathcal{F}_n)_{n \geq 0}$ (sur l'espace canonique des trajectoires). On suppose que, pour un $x \in S$, on a $\mathbb{E}_x(T) < +\infty$ et $X_T = x$ \mathbb{P}_x -p.s. Etant donné $y \in S$, on définit $N_T(y)$ comme le nombre de visites en y avant le temps T , ou, plus formellement,

$$N_T(y) := \text{card}\{0 \leq j \leq T - 1; X_j = y\}.$$

Montrer qu'alors, pour tout $y \in S$, on a

$$\mathbb{E}_x(N_T(y)) = \nu(y)\mathbb{E}_x(T).$$

2) Quelle relation obtient-on en posant $T := T_1(y)$?

3) Prouver la relation suivante : pour tous $x, y \in S$ tels que $x \neq y$,

$$\mathbb{P}_x(T_1(y) < T_1(x)) = \frac{1}{\nu(x)(\mathbb{E}_x(T_1(y)) + \mathbb{E}_y(T_1(x)))}.$$

(Indication : choisir un T approprié et appliquer le résultat du 1).)

4) Prouver la relation suivante : pour tous $x, y, z \in S$ tels que x, y, z sont deux-à-deux distincts,

$$\mathbb{P}_x(T_1(y) < T_1(z)) = \frac{\mathbb{E}_x(T_1(z)) + \mathbb{E}_z(T_1(y)) - \mathbb{E}_x(T_1(y))}{\mathbb{E}_y(T_1(z)) + \mathbb{E}_z(T_1(y))}.$$

(Indication : choisir un T approprié et appliquer le résultat du 1).)

Exercice 143 (Marche aléatoire dans un milieu périodique) On considère un noyau de transition p sur \mathbb{Z}^d vérifiant la propriété de périodicité suivante : il existe des entiers $\ell_1, \dots, \ell_d > 0$ tel que, pour tous $x, y \in \mathbb{Z}^d$, et tous $u_1, \dots, u_d \in \mathbb{Z}$,

$$p(x + (u_1\ell_1, \dots, u_d\ell_d), y + (u_1\ell_1, \dots, u_d\ell_d)) = p(x, y).$$

On suppose en outre que, pour tout $x \in \mathbb{Z}^d$, la loi de X_1 sous \mathbb{P}_x possède une espérance finie, ou, plus explicitement, que $\sum_{y \in \mathbb{Z}^d} |y|p(x, y) < +\infty$. Montrer qu'il existe $v \in \mathbb{R}^d$ tel que, pour tout $x \in \mathbb{Z}^d$,

$$\lim_{n \rightarrow +\infty} \frac{X_n}{n} = v, \mathbb{P}_x - \text{p.s.}$$

Exercice 144 Dans cet exercice, on considère le noyau markovien p sur \mathbb{N} défini de la manière suivante : $p(0,1) = 1$ et, pour tout $n \geq 1$, $p(n,i) = P(Z = i)$, où Z désigne une variable aléatoire de loi binomiale de paramètres $2n$ et $1/2$.

- 1) Montrer que p est irréductible et apériodique.
- 2) Posons $f(x) = x$ pour $x \in \mathbb{N}$. Calculer la fonction pf . En déduire que p est récurrent.
- 3) Prouver qu'il existe une constante $C \geq 0$ telle que, pour tout $n \geq 0$,

$$\mathbb{E}_0(X_n^2) \leq Cn^2.$$

(Indication : on peut raisonner par récurrence.)

- 4) Considérons une variable aléatoire $X \geq 0$ telle que $0 < E(X^2) < +\infty$. Prouver l'inégalité suivante :

$$P\left(X \geq \frac{1}{2}E(X)\right) \geq \frac{1}{4} \frac{E(X)^2}{E(X^2)}.$$

(Indication : que peut-on dire de $E(X \mathbf{1}_{[X \geq \frac{1}{2}E(X)]})$?)

- 5) Montrer que, si p est positivement récurrent, il doit exister une constante $c > 0$ telle que pour tout $n \geq 0$,

$$\mathbb{E}_0(X_n) \geq cn.$$

- 6) En faisant la synthèse des questions précédentes, montrer, en raisonnant par l'absurde, que p est récurrent nul.

Chapitre 5

Comportement asymptotique de la loi de X_n

Le résultat fondamental concernant le comportement asymptotique de la loi de X_n lorsque n tend vers l'infini est présenté dans les deux théorèmes suivants, le premier traitant du cas récurrent positif, tandis que le second traite les cas transients et récurrent nul.

Théorème 11 *Si p est un noyau irréductible et apériodique positivement récurrent, alors, en désignant par ν la loi invariante de la chaîne, on a, pour toute loi initiale μ , la convergence en loi suivante : pour tout $A \subset S$,*

$$\lim_{n \rightarrow +\infty} \mathbb{P}_\mu(X_n \in A) = \nu(A).$$

On constate donc que l'influence du point de départ X_0 sur la loi de X_n s'estompe à mesure que le nombre de pas effectués par la chaîne grandit : on dit que la chaîne «oublie» son point de départ.

Définition 8 *Une chaîne de Markov irréductible, apériodique, et récurrente positive, est appelée ergodique.*

La définition ci-dessus n'est pas très heureuse, car elle ne coïncide pas avec la définition de l'ergodicité dans le cadre que nous avons utilisé au chapitre précédent. Tant pis!

Théorème 12 *Si p est un noyau irréductible et apériodique, récurrent nul ou transient, alors, on a, pour toute loi initiale μ , et tout $x \in S$,*

$$\lim_{n \rightarrow +\infty} P(X_n = x) = 0.$$

On note que l'on ne peut avoir une formulation du résultat ci-dessus du type : pour tout $A \subset S$, $\lim_{n \rightarrow +\infty} P(X_n \in A) = 0$! Cette limite est cependant réalisée lorsque A est, par exemple, un sous-ensemble fini de S . En revanche, les deux formulations (avec $x \in S$ ou avec un sous-ensemble général $A \subset S$) sont équivalentes dans le cas où la limite est une loi de probabilité, comme expliqué dans l'exercice suivant.

Exercice 145 *On se donne une suite de lois μ_n sur un ensemble fini ou dénombrable S , et l'on suppose qu'il existe une mesure de probabilité ν sur S telle que, pour tout $x \in S$,*

$$\lim_{n \rightarrow +\infty} \mu_n(x) = \nu(x).$$

1) *Montrer que la suite $(\mu_n)_{n \geq 0}$ est tendue. (Indication : pour $\epsilon > 0$ donné, choisir A tel que $\nu(A) \geq 1 - \epsilon$.)*

2) *En déduire que, pour tout $B \subset S$,*

$$\lim_{n \rightarrow +\infty} \mu_n(B) = \nu(B).$$

Exercice 146 *Montrer, à l'aide des théorèmes 11 et 12, que, si $(X_n)_{n \geq 0}$ est une chaîne de Markov irréductible et apériodique sur un ensemble fini ou dénombrable S , la suite $(X_n)_{n \geq 0}$ est tendue si et seulement si elle converge en loi vers une mesure de probabilité.*

Exercice 147 *Montrer comment déduire des théorèmes 11 et 12 les faits suivants, pour une chaîne de Markov irréductible et apériodique :*

- *l'existence d'une probabilité invariante entraîne la récurrence positive ;*
- *il ne peut exister plus d'une probabilité invariante.*

Les deux théorèmes ci-dessus concernent le comportement asymptotique des lois de probabilité μp^n . Dans le cas ergodique tout au moins, on peut en déduire simplement le résultat suivant, en termes d'action du noyau sur les fonctions.

Théorème 13 *Si p est un noyau ergodique de loi invariante ν , on a le résultat suivant : pour tout $1 \leq q < +\infty$, et $f \in L^q(\nu)$,*

$$\lim_{n \rightarrow +\infty} p^n f = \nu(f),$$

la limite ayant lieu au sens de l'espace vectoriel normé $L^q(\nu)$.

Preuve :

Sans perte de généralité, nous prouverons le résultat pour toute $f \in L^q(\nu)$ telle que $\nu(f) = 0$. On se rappelle d'abord que, pour tout $x \in S$, $p^n f(x) = \mathbb{E}_x(f(X_n))$.

Ensuite, pour tout $\epsilon > 0$, il existe un ensemble fini A tel que $\sum_{x \notin A} |f(x)|^q \nu(x) \leq \epsilon$.
Ecrivons

$$\mathbb{E}_x(f(X_n)) = \mathbb{E}_x(f(X_n)\mathbf{1}(X_n \in A)) + \mathbb{E}_x(f(X_n)\mathbf{1}(X_n \notin A)).$$

Nous allons montrer séparément la convergence dans $L^q(\nu)$ des deux termes du membre de droite de l'égalité ci-dessus. L'inégalité de Jensen entraîne que

$$\sum_{x \in S} \nu(x) |\mathbb{E}_x(f(X_n)\mathbf{1}(X_n \notin A))|^q \leq \sum_{x \in S} \nu(x) \mathbb{E}_x(|f(X_n)|^q \mathbf{1}(X_n \notin A)).$$

Par invariance de ν , on a que

$$\sum_{x \in S} \nu(x) \mathbb{E}_x(|f(X_n)|^q \mathbf{1}(X_n \notin A)) = \sum_{x \notin A} |f(x)|^q \nu(x) \leq \epsilon.$$

A présent, pour tout $x \in S$, on a, en vertu du théorème 11, le fait que

$$\lim_{n \rightarrow +\infty} \mathbb{E}_x(f(X_n)\mathbf{1}(X_n \in A)) = \sum_{x \in A} f(x)\nu(x).$$

Qui plus est, A étant un ensemble fini, f est bornée sur A , et le théorème de convergence dominée permet de conclure que

$$\lim_{n \rightarrow +\infty} \sum_{x \in S} \nu(x) \left| \mathbb{E}_x(f(X_n)\mathbf{1}(X_n \in A)) - \sum_{x \in A} f(x)\nu(x) \right|^q = 0.$$

En utilisant le fait que $\sum_{x \in S} \nu(x)f(x) = 0$ (par hypothèse) et le fait que

$$\left| \sum_{x \notin A} f(x)\nu(x) \right| \leq \left(\sum_{x \notin A} |f(x)|^q \nu(x) \right)^{1/q} \leq \epsilon^{1/q},$$

on en déduit que

$$\left| \sum_{x \in A} f(x)\nu(x) \right| \leq \epsilon^{1/q},$$

et la conclusion en résulte facilement. \square

Exercice 148 *En utilisant la dualité fournie par le renversement dans le temps de p par rapport à ν , énoncer et prouver un résultat analogue à celui présenté ci-dessus pour l'action sur les mesures.*

5.1 Plan de cette partie

Les théorèmes ci-dessus sont de nature asymptotique. Une question naturelle en rapport avec ceux-ci est celle de la rapidité de la convergence dans les limites qui y figurent. Cette question peut elle-même être abordée de manière plus ou moins qualitative et plus ou moins asymptotique. Les différentes approches permettant de prouver ces théorèmes conduisent naturellement à différentes manières d'aborder le problème, et c'est là l'une de nos motivations pour présenter (ou au moins mentionner l'existence d') approches de natures variées. Nous présenterons donc successivement différentes approches permettant de prouver les théorèmes ci-dessus, en plus ou moins grande généralité. A savoir : le renouvellement, le couplage, la construction de temps stationnaires forts, la décomposition spectrale, les estimations sur la forme de Dirichlet, la décroissance de l'entropie relative. Il ne s'agit certainement pas d'une liste exhaustive, et bien d'autres approches sont possibles, telles que, par exemple, l'utilisation d'inégalités fonctionnelles autres que l'inégalité de Poincaré présentée ici (voir [48] ou [37] pour une présentation de diverses approches de ce type dans le cas où S est fini), ou l'approche par dualité d'ensembles en évolution (développée dans [38], voir également [15] et [37]).

Comme nous le verrons que ces approches se prêtent à des définitions variées de la manière de mesurer la proximité entre la loi de X_n et sa limite, et conduisent à des manières différentes d'estimer ces mesures.

5.2 Preuve par renouvellement

L'idée est de s'appuyer sur le découpage des trajectoires produit par les retours successifs en un point, les lois de probabilité des temps de retour jouant alors le rôle principal.

Remarquons d'abord que le théorème 12, dans le cas où la chaîne est transiente, est une conséquence directe de la proposition 26, et du théorème de convergence dominée.

On peut prouver aussi bien le théorème 12 dans le cas récurrent nul que le théorème 11 en faisant appel au théorème du renouvellement, que nous citons ici sous la forme adaptée à notre contexte.

Théorème 14 *Si ξ_1, ξ_2, \dots , est une suite de v.a. i.i.d. à valeurs entières strictement positives, et telles que le p.g.c.d. du support de la loi des ξ_i soit égal à 1. Alors*

$$\lim_{n \rightarrow +\infty} P(n \in \{0, \xi_1, \xi_1 + \xi_2, \dots\}) = \frac{1}{E(\xi_1)},$$

avec la convention $1/+\infty = 0$.

On prouve généralement le théorème du renouvellement au moyen de méthodes analytiques.

Exercice 149 Lire une preuve analytique du théorème de renouvellement, par exemple dans [20] (XIII,11), ou, pour une version plus générale (non limitée au cas discret), [21] (XI).

Supposons donc la chaîne récurrente. On note que pour toute loi initiale μ , et $n \geq 1$,

$$\mathbb{P}_\mu(X_n = x) = \mathbb{P}_\mu(n - T_1(x) \in \{0, T_2(x) - T_1(x), T_3(x) - T_1(x), \dots\}).$$

Posons

$$a(k, x) = \mathbb{P}_x(k \in \{0, T_1(x), T_2(x), T_3(x), \dots\}).$$

On note que, en utilisant le caractère i.i.d. de la suite $(T_{i+1}(x) - T_i(x))_{i \geq 1}$ par rapport à la probabilité \mathbb{P}_x , et en posant $\gamma_x :=$ loi de probabilité de $T_1(x)$ sous \mathbb{P}_x , i.e. $\gamma_x(n) = \mathbb{P}_x(T_1(x) = n)$, on peut réécrire, pour tout $k \geq 0$,

$$a(k, x) = \sum_{i=0}^{+\infty} \gamma_x^{\star i}(k), \quad (5.1)$$

où $\star i$ désigne le produit de convolution itéré i fois, et où l'on pose, pour $i = 0$, $\gamma^{\star 0} := \delta_0$.

Grâce à la propriété forte de Markov appliquée à $T_1(x)$, on obtient l'identité fondamentale suivante

$$\mathbb{P}_\mu(X_n = x) = \mathbb{E}_\mu(a(n - T_1(x), x)).$$

D'après le théorème du renouvellement appliqué à la suite définie par $\xi_i = T_i(x) - T_{i-1}(x)$, (avec la convention $T_0(x) = 0$) qui est i.i.d. sous \mathbb{P}_x , la condition sur le p.g.c.d. étant exactement la condition d'apériodicité de la chaîne, on a $\lim_{n \rightarrow +\infty} a(n - T_1(x), x) = 1/E_{\mathbb{P}_x}(T_1(x))$. Le théorème de convergence dominée permet de conclure que $\lim_{n \rightarrow +\infty} \mathbb{P}_\mu(X_n = x) = \nu(x)$. Ceci suffit à prouver le théorème 12 dans le cas récurrent nul. Pour prouver le théorème 11 dans le cas ergodique, il nous faut étendre la limite valable pour tout x à une limite valable pour tout sous-ensemble $A \subset S$, pas forcément fini. Cette extension est par exemple une conséquence de l'exercice 145 (mais c'est aussi une conséquence de la tension de la suite de lois $(\mathbb{P}_\mu(X_n = \cdot))_{n \geq 0}$, que l'on sait établir simplement en utilisant la récurrence positive, voir l'exercice 137).

On note que l'on retrouve grâce à cette preuve, la formule liant la mesure invariante et l'espérance du temps de retour en un point.

On note que, pour préciser la vitesse de convergence dans la preuve ci-dessus, on doit chercher à la fois à préciser la vitesse de convergence dans le théorème de

renouvellement, et disposer d'un contrôle sur la décroissance de la queue de la variable $T_1(x)$. Ce point de vue est illustré dans l'exercice suivant. Voir [36] pour plus de détails sur cette question.

Exercice 150 (D'après [36].) *On considère un noyau irréductible, apériodique, et positivement récurrent sur un ensemble fini ou dénombrable S , et l'on fixe $a \in S$.*

1) *Montrer que, pour tous $x, y \in S$, et $n \geq 0$, on a la représentation suivante de $\mathbb{P}_x(X_n = y)$:*

$$\mathbb{P}_x(X_n = y, T_1(a) \geq n) + \sum_{j=1}^n \left[\sum_{k=1}^j \mathbb{P}_x(X_k = a, T_1(a) \geq n) \mathbb{P}_a(X_{j-k} = a) \right] \mathbb{P}_a(X_{n-j} = y).$$

2) *On introduit les notations suivantes.*

$$\alpha_x(n) := \mathbb{P}_x(T_1(a) = n), \quad u(n) := \mathbb{P}_a(X_n = a), \quad t_y(n) := \mathbb{P}_a(X_n = y, T_1(a) \geq n).$$

Montrer que l'on a, pour tous $x, y \in S$, l'inégalité suivante :

$$|\mathbb{P}_x(X_n = y) - \nu(y)| \leq \mathbb{P}_x(X_n = y, T_1(a) \geq n) + |\alpha_x * u - \nu(a)| * t_y(n) + \nu(a) \sum_{j=n+1}^{+\infty} t_y(j).$$

3) *Montrer que $\lim_{n \rightarrow +\infty} |\mathbb{P}_x(X_n = y) - \nu(y)| = 0$ en analysant séparément chacun des termes de l'inégalité ci-dessus. De quelles informations a-t-on besoin pour contrôler la vitesse de convergence ?*

5.3 Preuve par couplage et distance en variation totale

Deux références sur le couplage (pas seulement dans le contexte des chaînes de Markov) sont les monographies consacrées au sujet [34, 51].

5.3.1 Distance en variation totale entre deux mesures de probabilité et couplage

Dans toute cette section, S désigne un ensemble fini ou dénombrable. Si μ et ν sont deux mesures de probabilité sur S , on définit

$$d_{VT}(\mu, \nu) := \sup_{A \subset S} |\mu(A) - \nu(A)|.$$

On dit que $d_{VT}(\mu, \nu)$ est la **distance en variation totale** entre μ et ν , et l'on note que $0 \leq d_{VT}(\mu, \nu) \leq 1$.

Proposition 51 *Sur l'ensemble des mesures de probabilité sur S , d_{VT} définit une distance.*

Exercice 151 *Prouver la proposition ci-dessus.*

Proposition 52 *Etant donnée deux mesures de probabilité μ et ν sur S , on a l'identité*

$$d_{VT}(\mu, \nu) = \frac{1}{2} \sum_{x \in S} |\mu(x) - \nu(x)|.$$

Preuve :

On introduit l'ensemble $B = \{x : \mu(x) \leq \nu(x)\}$, et l'on vérifie qu'il réalise le maximum dans le supremum définissant d_{VT} . \square

On voit donc que d_{VT} définit une distance sur les mesures de probabilités sur S , vues comme un sous-ensemble de l'espace normé $\ell^1(S)$. Une manière différente d'écrire cette distance, dans le cas particulier où ν est une mesure propre (ce qui sera toujours le cas lorsque ν est la loi invariante d'un noyau irréductible) est la suivante :

$$d_{VT}(\mu, \nu) = 1/2 \sum_{x \in S} |\mu(x)/\nu(x) - 1| \nu(x),$$

et $d_{VT}(\mu, \nu)$ apparaît donc comme la norme (au facteur $1/2$ près) dans $L^1(\nu)$ de la fonction $x \mapsto (\mu(x)/\nu(x) - 1)$.

Exercice 152 *Montrer que $d_{VT}(\mu, \nu) = 1$ si et seulement si μ et ν sont de supports disjoints, i.e. il existe $A \subset S$ tel que $\mu(A) = 1$ et $\nu(A^c) = 1$.*

Exercice 153 *Montrer que la convergence en loi sur S équivaut à la convergence en variation totale lorsque S est fini ou dénombrable, mais que l'on a seulement une implication dans le cadre des espaces de probabilité généraux (en prenant le sup sur tous les événements A de la tribu considérée pour définir d_{VT}).*

L'exercice ci-dessus montre donc que, qualitativement, convergence en loi et convergence en variation totale sont équivalentes (sur un espace fini ou dénombrable). Cependant, contrôler la distance en variation totale fournit une information quantitative que ne contient pas le simple fait de savoir que la convergence en loi a lieu.

L'un des principes de base de l'approche par couplage est le suivant : pour prouver que deux lois de probabilité sur S , disons μ_1 et μ_2 , sont proches, on fabrique sur un même espace de probabilité deux variables aléatoires, disons Z_1 et Z_2 , de lois respectives μ_1 et μ_2 , de telle manière que la probabilité que les deux variables aléatoires en question soient proches ait une valeur proche de 1. Si l'on choisit simplement comme notion de proximité entre éléments de S l'égalité entre éléments (ce qui est loisible sur un ensemble S dénombrable général sur lequel on ne suppose pas donnée une structure particulière), on obtient un contrôle sur la distance en variation totale entre μ_1 et μ_2 , comme le précise la proposition suivante.

Proposition 53 Si Z_1 et Z_2 sont deux variables aléatoires définies sur un même espace de probabilité (Ω, \mathcal{F}, P) , de lois respectives μ_1 et μ_2 , alors

$$d_{VT}(\mu_1, \mu_2) \leq P(Z_1 \neq Z_2). \quad (5.2)$$

Preuve :

On écrit que

$$|\mu_1(A) - \mu_2(A)| = |E\mathbf{1}(Z_1 \in A) - E\mathbf{1}(Z_2 \in A)| = |E(\mathbf{1}(Z_1 \in A) - \mathbf{1}(Z_2 \in A))|.$$

Par conséquent,

$$|\mu_1(A) - \mu_2(A)| \leq E(|\mathbf{1}(Z_1 \in A) - \mathbf{1}(Z_2 \in A)|).$$

A présent, on note que, si l'événement $Z_1 = Z_2$, $\mathbf{1}(Z_1 \in A) - \mathbf{1}(Z_2 \in A) = 0$. D'autre part, on a toujours que $|\mathbf{1}(Z_1 \in A) - \mathbf{1}(Z_2 \in A)| \leq 1$. Le résultat s'ensuit. \square

De manière intéressante, il est toujours possible de fabriquer un couplage réalisant l'égalité dans l'inégalité prouvée ci-dessus. On en déduit une interprétation de la distance en variation totale en termes de couplage.

Proposition 54 On a l'identité $d_{VT}(\mu, \nu) = \min P(X \neq Y)$, où le minimum porte sur l'ensemble des paires de v.a. à valeurs dans S , définies sur un même espace de probabilité (Ω, \mathcal{F}, P) et telles que la loi de X est μ et la loi de Y est ν .

Exercice 154 Prouver la proposition ci-dessus. (Indication : utiliser un découpage d'intervalles.)

La preuve est un exercice facile de découpage d'intervalles (faire un dessin!).

Exercice 155 Si μ_1 et μ_2 sont deux mesures de probabilité sur S , $1 \leq s \leq +\infty$ et si $f : S \rightarrow \mathbb{R}$ est une fonction telle que $f \in L^s(\mu_1) \cap L^s(\mu_2)$, montrer que l'on a

$$|E_{\mu_1}(f(X_0)) - E_{\mu_2}(f(X_0))| \leq [\|f\|_{L^s(\mu_1)} + \|f\|_{L^s(\mu_2)}] [d_{VT}(\mu, \nu)]^{1/r},$$

où r est l'exposant conjugué de s . (Indication : appliquer la représentation par couplage de la distance en variation totale, et utiliser les inégalités de Hölder et Minkowski.)

5.3.2 Approche par couplage pour la convergence des chaînes de Markov

Nous avons décrit dans la section précédente le lien existant entre distance en variation totale et couplage pour deux variables aléatoires données. Dans le contexte de la convergence des chaînes de Markov, qui constituent des suites de variables aléatoires, nous utiliserons la notion suivante.

Définition 9 Si $X^1 = (X_k^1)_k$ et $X^2 = (X_k^2)_k$ sont deux suites de v.a. définies sur un même espace de probabilité et à valeurs dans S , et si T est une v.a. à valeurs dans \mathbb{N} et définie sur ce même espace de probabilité, on dit que T est un temps de couplage de X^1 et X^2 si, pour tout n , on a $n \geq T \Rightarrow X_n^1 = X_n^2$.

Proposition 55 Sous les hypothèses de la définition ci-dessus, on a, pour tout n , l'inégalité :

$$d_{VT}(P(X_n^1 = \cdot), P(X_n^2 = \cdot)) \leq P(T > n).$$

Preuve :

Il suffit de vérifier que le couplage que l'on obtient entre X_n^1 et X_n^2 que l'on obtient est tel que $P(X_n^1 \neq X_n^2) \leq P(T > n)$. \square

Preuve par couplage des théorèmes 11 et 12

Nous allons maintenant donner une preuve par couplage du théorème 11. Pour prouver la convergence en loi d'une chaîne de Markov ergodique vers sa loi invariante, nous allons fabriquer un temps de couplage entre deux chaînes de Markov $(X_n^1)_{n \geq 0}$ et $(X_n^2)_{n \geq 0}$, la première de loi initiale μ quelconque, et la seconde de loi initiale ν , où ν est la loi invariante. Une fois un tel temps de couplage T construit, le théorème de convergence dominée entraîne alors que $\lim_{n \rightarrow +\infty} P(T > n) = 0$, et, du fait que X_n^3 a pour loi ν quelle que soit la valeur de n , on en déduit la conclusion du théorème 11 en appliquant la proposition 55.

Pour effectuer notre construction, supposons données sur le même espace de probabilité (Ω, \mathcal{F}, P) deux suites de variables aléatoires, indépendantes l'une de l'autre, $(X_n^1)_{n \geq 0}$ et $(X_n^3)_{n \geq 0}$, telles que $(X_n^1)_{n \geq 0}$ soit une chaîne de Markov de noyau de transition p et de loi initiale μ quelconque, et $(X_n^3)_{n \geq 0}$ une chaîne de Markov de noyau de transition p et de loi initiale ν (la loi invariante).

Exercice 156 Justifier le fait qu'il est toujours possible de trouver un espace de probabilité (Ω, \mathcal{F}, P) et deux suites $(X_n^1)_{n \geq 0}$ et $(X_n^3)_{n \geq 0}$ vérifiant les propriétés ci-dessus.

Définissons ensuite $T = \inf\{n : X_n^1 = X_n^3\}$, et définissons pour tout n X_n^2 par $X_n^2 = X_n^1$ pour $n \leq T$ et $X_n^2 = X_n^3$ pour $n > T$.

Lemme 22 La suite $(X_n^2)_{n \geq 0}$ est une chaîne de Markov de loi initiale μ et de noyau de transition p , et T est un temps de couplage de $(X_n^1)_{n \geq 0}$ et $(X_n^2)_{n \geq 0}$.

Exercice 157 Prouver le résultat ci-dessus.

A présent, remarquons que la suite $(X_n^1, X_n^3)_{n \geq 0}$ est une chaîne de Markov sur $S \times S$, de noyau de transition $p \otimes p$ défini par $p \otimes p((x_1, x_2), (y_1, y_2)) = p(x_1, y_1)p(x_2, y_2)$, et que T n'est autre que le temps de première atteinte de l'ensemble $D := \{(x, x); x \in S\}$ par cette chaîne.

Lemme 23 *L'irréductibilité et l'apériodicité de p entraînent l'irréductibilité de $p \otimes p$.*

Preuve :

Etant donnés (x_1, x_2) et (y_1, y_2) dans $S \times S$, l'irréductibilité de p entraîne l'existence de $a_1 \geq 1$ et $a_2 \geq 1$ tels que $p^{a_1}(x_1, y_1) > 0$ et $p^{a_2}(x_2, y_2) > 0$. Notre but est de trouver $m_1, m_2 \geq 0$ tel que $p^{m_1}(y_1, y_1) > 0$, $p^{m_2}(y_2, y_2) > 0$, et $a_1 + m_1 = a_2 + m_2$. Par apériodicité de p , il existe $d_1^1, \dots, d_{k_1}^1 \geq 1$, et $u_1^1, \dots, u_{k_1}^1 \in \mathbb{Z}$ tels que $p^{d_j^1}(y_1, y_1) > 0$ pour tout $1 \leq j \leq k_1$, et $u_1^1 d_1^1 + \dots + u_{k_1}^1 d_{k_1}^1 = 1$. De même, il existe $d_1^2, \dots, d_{k_2}^2 \geq 1$, et $u_1^2, \dots, u_{k_2}^2 \in \mathbb{Z}$ tels que $p^{d_j^2}(y_2, y_2) > 0$ pour tout $1 \leq j \leq k_2$, et $u_1^2 d_1^2 + \dots + u_{k_2}^2 d_{k_2}^2 = 1$. On vérifie que, en choisissant $m_1 := a_2(u_1^1 d_1^1 + \dots + u_{k_1}^1 d_{k_1}^1) + K d_1^1 \dots d_{k_1}^1 d_1^2 \dots d_{k_2}^2$ et $m_2 := a_1(u_1^2 d_1^2 + \dots + u_{k_2}^2 d_{k_2}^2) + K d_1^1 \dots d_{k_1}^1 d_1^2 \dots d_{k_2}^2$, pour un entier K suffisamment grand (de manière à compenser les valeurs éventuellement négatives de u_i^j en faisant apparaître un facteur positif devant chaque d_i^j), on parvient au résultat recherché. \square

Exercice 158 *Donner un exemple où p est irréductible et où $p \otimes p$ ne l'est pas.*

A présent, notons que $\nu \otimes \nu$ est une probabilité invariante pour le noyau $p \otimes p$. Au vu du lemme 23 ci-dessus, $p \otimes p$ est irréductible, et l'existence d'une loi invariante entraîne automatiquement que $p \otimes p$ doit être positivement récurrent. En particulier, T est fini avec une probabilité égale à 1, ce qui prouve que T est un temps de couplage, et conclut la preuve du théorème 11.

Nous allons à présent donner une preuve par couplage du théorème 12 dans le cas où p est récurrent nul. A cette fin, on reprend l'étude du noyau $p \otimes p$ défini ci-dessus, et dont le lemme 23 prouve l'irréductibilité. Dans la situation précédente, la récurrence positive de p entraînait celle de $p \otimes p$, mais, dans le cas présent, $p \otimes p$ peut se révéler récurrent ou transient.

Exercice 159 *Donner un exemple où p est récurrent nul et $p \otimes p$ est transient, et un exemple où p est récurrent nul et $p \otimes p$ est récurrent. (Indication : essayer des marches aléatoires simples sur \mathbb{Z}^d .) Est-il possible que $p \otimes p$ soit récurrent positif dans ce cas ?*

Supposons d'abord que $p \otimes p$ est transient. On en déduit que, pour tout $x \in S$,

$$\lim_{n \rightarrow +\infty} \mathbb{P}_{\mu \otimes \mu, p \otimes p}(X_n = (x, x)) = 0.$$

Mais, par définition, $\mathbb{P}_{\mu \otimes \mu, p \otimes p}(X_n = (x, x)) = (\mathbb{P}_{\mu, p}(X_n = x))^2$, et l'on a donc que, pour tout $x \in S$,

$$\lim_{n \rightarrow +\infty} \mathbb{P}_{\mu, p}(X_n = x) = 0,$$

ce qui est la conclusion du théorème 12.

Supposons maintenant que $p \otimes p$ est récurrent. En reprenant la définition de T avec des chaînes $(X_n^1)_{n \geq 0}$ et $(X_n^3)_{n \geq 0}$ issues respectivement de deux lois initiales μ et ν quelconques, on en déduit de la récurrence de $p \otimes p$ le fait que T est fini $\mathbb{P}_{\mu \otimes \nu, p \otimes p}$ -presque sûrement, d'où, en appliquant la proposition 55, le fait que, pour toutes lois initiales μ et ν ,

$$\lim_{n \rightarrow +\infty} d_{VT}(\mathbb{P}_\mu(X_n = \cdot), \mathbb{P}_\nu(X_n = \cdot)) = 0. \quad (5.3)$$

A présent, fixons arbitrairement $a \in S$, et considérons la mesure μ_a^{cycle} . Comme p est supposé récurrent nul, cette mesure est invariante pour p (et donc propre), et de masse infinie. Etant donné $x \in S$, donnons-nous un sous-ensemble fini $A \subset S$ contenant x . On a donc $0 < \mu_a^{cycle}(A) < +\infty$, et il est possible de définir une mesure de probabilité sur S en posant

$$\begin{cases} \mu_A(x) = \frac{\mu_a^{cycle}(x)}{\mu_a^{cycle}(A)}, & x \in A, \\ \mu_A(x) = 0, & x \notin A. \end{cases}$$

D'après cette définition, on vérifie facilement que, pour tout $x \in S$,

$$\mu_A(x) \leq \frac{\mu_a^{cycle}(x)}{\mu_a^{cycle}(A)}.$$

En composant avec p^n , on en déduit que, pour tout $x \in S$,

$$\mu_{Ap^n}(x) \leq \frac{\mu_a^{cycle} p^n(x)}{\mu_a^{cycle}(A)} = \frac{\mu_a^{cycle}(x)}{\mu_a^{cycle}(A)}.$$

Pour un x fixé, du fait que μ_a^{cycle} est de masse infinie, on peut choisir A tel que $\mu_a^{cycle}(A)$ soit arbitrairement grand, disons, tel que $\mu_A(x) = \frac{\mu_a^{cycle}(x)}{\mu_a^{cycle}(A)} \leq \epsilon$. En appliquant (5.3) pour contrôler $|\mathbb{P}_\mu(X_n = x) - \mathbb{P}_{\mu_A}(X_n = x)|$, on en déduit que

$$\limsup_{n \rightarrow +\infty} \mathbb{P}_\mu(X_n = x) \leq \epsilon,$$

ce qui conclut la preuve.

Remarquons que, dans la section précédente, nous n'avons pas fourni de preuve du théorème du renouvellement. Ayant maintenant établi par couplage les théorèmes 11 et 12, nous pouvons les utiliser afin de donner une preuve du théorème du renouvellement sous la forme que nous avons énoncée dans la section précédente.

Exercice 160 Soit ζ une variable aléatoire ζ à valeurs dans \mathbb{N}^* telle que le pgcd des éléments du support de S soit égal à 1.

1) Montrer que, pour toute variable aléatoire, il est possible de construire une chaîne de Markov apériodique et récurrente sur \mathbb{N} telle que la loi de $T_1(0)$ sous la probabilité \mathbb{P}_0 est la loi de ζ .

2) En utilisant les théorèmes 11 et 12, en déduire le théorème du renouvellement énoncé dans la section précédente.

Vitesse de convergence et couplage

De manière générale, la proposition 55 transforme la question de la vitesse de convergence en variation totale en celle de la vitesse de décroissance de la queue du temps de couplage T . Nous présenterons ultérieurement des développements utilisant cette idée. Pour le moment, contentons-nous de remarquer que le temps de couplage T que nous avons défini dans la preuve possède une structure extrêmement simple : les deux trajectoires évoluent de manière indépendante jusqu'à ce qu'elles se touchent. On obtient ainsi un temps de couplage valable dans un cadre très général, mais qui n'est pas forcément très efficace au sens où, pour une chaîne donnée, il est vraisemblable que d'autres couplages, exploitant la structure particulière de la chaîne considérée, conduisent à des temps de couplages plus brefs, donnant donc lieu à des estimations plus précises de la vitesse de convergence. Nous verrons au moins un exemple de tel couplage dans le contexte des méthodes de Monte-Carlo par chaînes de Markov. Pour un exemple «high-tech», faisant intervenir entre autres des couplages anticipatifs (où la variable aléatoire T n'est donc pas, contrairement à notre exemple, un temps d'arrêt vis-à-vis de $(X_n^1, X_n^2)_{n \geq 0}$), voir par exemple [30]. Il faut également noter que les techniques de couplage jouent un rôle crucial dans certaines méthodes de simulation (simulation par couplage à partir du passé, à la Propp-Wilson) que nous décrirons dans une autre partie.

Exercice 161 Dans le cas où S est fini, montrer que T possède une queue sous-géométrique, et que la convergence a donc lieu à vitesse au moins exponentielle. Montrer que l'on peut même établir une borne indépendante de la loi initiale. Montrer que, génériquement, la vitesse de la convergence est effectivement exponentielle.

Exercice 162 Montrer sur un exemple que, lorsque S est infini, il n'existe pas toujours de borne uniforme vis-à-vis de la loi initiale pour la convergence en variation totale.

Nous avons vu que, pour tout temps de couplage entre deux versions de la chaîne dont l'une est initialisée selon la loi invariante ν , l'autre l'étant selon une loi quelconque μ , on a l'inégalité $d_{VT}(\mathbb{P}_\mu(X_n = \cdot), \nu) \leq P(T > n)$. Dans le cas d'une chaîne

de Markov ergodique sur un ensemble S fini, on peut en fait prouver qu'il existe toujours un temps de couplage tel qu'il y ait égalité dans l'inégalité ci-dessus. Un tel temps de couplage est donc minimal, en ce sens que tout autre temps de couplage T' doit vérifier que $P(T > n) \leq P'(T' > n)$, autrement dit, tout temps de couplage est minoré stochastiquement par T . Pour une preuve de ce résultat, dont l'intérêt est essentiellement théorique, voir [34].

Notons qu'en partant simplement de l'existence d'un couplage, tel que celui défini dans la preuve du théorème, il est possible d'obtenir des informations qualitatives sur la vitesse d'oubli de la condition initiale en variation totale. En considérant le couplage à partir de $(x, y) \in S \times S$, on constate que

$$d_{VT}(\delta_x p^n, \delta_y p^n) \leq P(T > n)$$

et que

$$\sum_{n=0}^{+\infty} P(T > n) < +\infty$$

car $p \otimes p$ est positivement récurrent. Nous pouvons reformuler ceci sous la forme d'une proposition.

Proposition 56 *Etant donnée une chaîne de Markov ergodique, on a, pour tous $x, y \in S$,*

$$\sum_{n=0}^{+\infty} d_{VT}(\delta_x p^n, \delta_y p^n) < +\infty.$$

Nous verrons dans le chapitre suivant que cette propriété n'est pas toujours vérifiée si l'on considère $\sum_{n=0}^{+\infty} d_{VT}(\delta_x p^n, \nu)$. On peut en tout cas noter que le fait que T soit d'espérance finie si l'on part non pas d'une loi concentrée en un point (en l'occurrence (x, y)), mais d'une loi dont le support est potentiellement infini telle que $\delta_x \otimes \nu$, n'est pas a priori une conséquence de la récurrence positive de $p \otimes p$.

Nous aurons l'occasion par la suite d'utiliser les notations :

$$d_1(n) := \sup_{x \in S} d_{VT}(p^n(x, \cdot), \nu),$$

et

$$\bar{d}_1(n) := \sup_{x, y \in S} d_{VT}(p^t(x, \cdot), p^t(y, \cdot)) = \delta(p^n).$$

5.4 Temps stationnaires forts et distance en séparation

Dans toute cette partie, nous nous placerons sur un espace d'états S de cardinal fini, et nous considérons un noyau ergodique p de loi invariante ν . La présentation donnée ici s'inspire fortement de celle de Brémaud [7] Chap. 6 Sec. 5. Voir également [1]

Définition 10 *Un temps stationnaire fort pour une chaîne de Markov ergodique $(X_n)_{n \geq 0}$ de loi invariante ν est une variable aléatoire T telle que :*

- (i) *T est un temps d'arrêt par rapport à une famille de tribus (\mathcal{G}_n) de la forme $\mathcal{G}_n := \sigma(X_0, \dots, X_n, Z)$, où Z est une variable aléatoire indépendante de $(X_k)_{k \geq 0}$ (on autorise donc les événements de la forme $T = n$ à dépendre de X_0, \dots, X_n et d'une randomisation indépendante de toute la chaîne) ;*
- (ii) *$P(T < +\infty) = 1$;*
- (iii) *$P(X_T = x | T = n) = \nu(x)$.*

En d'autres termes, la propriété (iii) signifie que X_T est distribué selon la loi ν et est de plus indépendant de la valeur de T . D'un point de vue algorithmique, un tel temps fournit une réponse à la question : combien de pas d'une chaîne de Markov ergodique doit-on simuler avant d'atteindre la loi invariante avec une précision satisfaisante. En effet, en simulant un nombre (aléatoire !) de pas égal à T , on obtient une variable aléatoire distribuée exactement selon cette loi stationnaire. Cette approche est à la base de l'algorithme de simulation parfaite de Fill que nous décrirons plus bas. Pour l'instant, expliquons comment l'existence d'un temps stationnaire fort permet de prouver la convergence vers la loi invariante.

Nous avons vu précédemment le lien étroit entre couplage et distance en variation totale. Dans le contexte des temps stationnaires forts, la notion appropriée de distance entre lois s'avère être celle de distance en séparation :

Définition 11 *La distance en séparation de deux mesures de probabilité μ et ν sur S est définie par $\text{sép.}(\mu, \nu) = \max_{x \in S} 1 - \frac{\mu(x)}{\nu(x)}$ (avec la convention $\mu(x)/\nu(x) = 1$ si $\mu(x) = \nu(x) = 0$).*

Attention : cette définition n'est pas symétrique en μ et ν , et donc la terminologie est trompeuse car elle ne correspond pas à une véritable distance. On vérifie que $0 \leq \text{sép.}(\mu, \nu) \leq 1$, et que l'on a l'inégalité suivante :

Proposition 57 *Pour deux mesures de probabilité μ et ν sur S , $d_{VT}(\mu, \nu) \leq \text{sép.}(\mu, \nu)$*

Preuve :

Immédiat en introduisant l'ensemble $B = \{x; \mu(x) \leq \nu(x)\}$. □

Proposition 58 *Si T est un temps stationnaire fort, alors pour tout $n \geq 0$, $\text{sép.}(\mu p^t, \nu) \leq P(T > n)$.*

Preuve :

On écrit que $P(X_n = x) = P(X_n = x, T \leq n) + P(X_n = x, T > n)$. En partant de la définition d'un temps stationnaire fort, on vérifie facilement que $(X_{T+i})_{i \geq 0}$

est une chaîne de Markov de noyau de transition p et de loi ν , indépendante de T . On en déduit que $P(X_n = x, T \leq n) = \nu(x)(1 - P(T > n))$, d'où le fait que $P(X_n = x)/\nu(x) \geq 1 - P(T > n)$, ce qui conclut la preuve. \square

Comme pour les temps de couplage vis-à-vis de la distance en variation totale, il existe également un temps stationnaire fort minimal, au sens où celui réalise l'égalité dans l'inégalité ci-dessus, et donc au sens où il minore stochastiquement tous les autres temps stationnaires forts.

Proposition 59 *Pour toute loi initiale μ sur S , il existe une chaîne de Markov de noyau de transition p et de loi initiale μ possédant un temps stationnaire fort T tel que l'on ait, pour tout $n \geq 0$,*

$$\text{sép.}(\mu p^n, \nu) = P(T > n).$$

Preuve :

Voir [7]. \square

La construction de ce temps stationnaire fort permet en particulier de retrouver la convergence en loi de la chaîne vers sa loi stationnaire.

Nous avons vu que la distance en séparation domine de manière générale la distance en variation totale. Dans le contexte de la comparaison des temps (de couplage et stationnaires forts), on a le résultat suivant.

Proposition 60 *Tout temps stationnaire fort est un temps de couplage (quitte à enrichir l'espace de probabilité sous-jacent).*

Preuve :

Partons d'un temps stationnaire fort T , et notons $(X_n^1)_{n \geq 0}$ la chaîne de Markov issue de la loi initiale μ correspondante. Ensuite, partant de X_T^1 , fabriquons indépendamment de ce qui précède une trajectoire $X_T^1 =: Y_0, \dots, Y_T$ selon le noyau renversé dans le temps de p par rapport à ν . On vérifie alors que la suite de variables aléatoires $(X_n^2)_{n \geq 0}$ définie pour $0 \leq n \leq T$ par $X_n^2 := Y_{n-T}$, et pour $n > T$ par $X_n^2 := X_n^1$ est une chaîne de Markov de noyau de transition p issue de la loi stationnaire ν , et que T est un temps de couplage entre $(X_n^1)_{n \geq 0}$ et $(X_n^2)_{n \geq 0}$. \square

Citons à présent quelques propriétés générales de la distance en séparation dans le contexte des chaînes de Markov.

Définition 12 *Pour une chaîne irréductible de loi invariante ν , on définit, pour $t \geq 0$, $d_{\text{sép.}}(t) = \max_{x \in S} \text{sép.}(p^n(x, \cdot), \nu)$.*

Proposition 61 *Pour toute loi initiale μ , on a $\text{sép.}(\mu p^n, \nu) \leq d_{\text{sép.}}(t)$.*

Preuve :

Convexité. □

Avec ces notations, on a donc, puisque $d_{VT} \leq \text{sép.}$:

Proposition 62 *Pour une chaîne irréductible de loi invariante ν , pour tout $t \geq 0$,*

$$d_1(t) \leq d_{\text{sép.}}(t).$$

Dans le cas d'une chaîne réversible, on a la borne en sens inverse suivante :

Proposition 63 *Pour une chaîne irréductible de loi invariante ν réversible, pour tout $t \geq 0$,*

$$d_{\text{sép.}}(2t) \leq 1 - (1 - \bar{d}_1(t))^2.$$

Preuve :

$$\frac{p^{2t}(x, y)}{\nu(y)} = \sum_{z \in S} \frac{p^t(x, z)p^t(z, y)}{\nu(y)} = \sum_{z \in S} \nu(z) \frac{p^t(x, z)p^t(y, z)}{\nu(z)^2},$$

la dernière égalité utilisant la réversibilité. Par Cauchy-Schwarz, on en déduit que

$$\frac{p^{2t}(x, y)}{\nu(y)} \geq \left(\sum_{z \in S} \nu(z) \frac{(p^t(x, z))^{1/2}(p^t(y, z))^{1/2}}{\nu(z)} \right)^2,$$

d'où

$$\frac{p^{2t}(x, y)}{\nu(y)} \geq \left(\sum_{z \in S} \min_z(p^t(x, z), (p^t(y, z))) \right)^2.$$

On vérifie ensuite facilement (pensez à l'ensemble B) que $\sum_{z \in S} \min(p^t(x, z), (p^t(y, z))) = 1 - d_{VT}(p^t(x, \cdot), p^t(y, \cdot))$. Le résultat s'ensuit. □

Proposition 64 *La fonction $t \mapsto d_{\text{sép.}}(t)$ est sous-multiplicative (et donc en particulier décroissante).*

Preuve :

On écrit que $p^t(x, y) = (1 - d_{\text{sép.}}(t))\nu(y) + d_{\text{sép.}}(t)q_t(x, y)$, et l'on vérifie que q_t est un noyau de transition pour lequel ν est invariante. On obtient alors en calculant que $p^{t+s}(x, y) = (1 - d_{\text{sép.}}(t)d_{\text{sép.}}(s))\nu(y) + d_{\text{sép.}}(t)d_{\text{sép.}}(s) \sum_{z \in S} q_s(x, z)q_t(z, y)$, d'où le fait que $p^{t+s}(x, y) \geq (1 - d_{\text{sép.}}(t)d_{\text{sép.}}(s))\nu(y)$, ce qui conclut la preuve. □

Ainsi, on peut par exemple utiliser $\tau_{\text{sép.}} = \inf\{t; d_{\text{sép.}}(t) \leq 1/2\}$ pour définir un temps de relaxation pour la distance en séparation.

Pour intéressante qu'elle soit en théorie, la construction de temps stationnaires forts analysables n'est possible que dans des cas très particuliers, possédant de fortes propriétés de symétrie. Voici deux exemples simples et croquignoles (cités par [7]).

5.4.1 Deux exemples de temps stationnaires forts

Le battage de cartes «top to random» (voir [2]).

Etant donné un paquet de N cartes empilées, on prend la carte du dessus et on l'insérant au hasard en-dessous de l'une des cartes du paquet, en choisissant uniformément la position d'insertion parmi les N positions possibles. On obtient ainsi une chaîne de Markov irréductible et apériodique sur l'ensemble des permutations possibles des N cartes ayant pour loi invariante la loi uniforme. Appelons f la carte initialement située en bas du paquet, posons $T_0 = 0$ et, pour $i = 1, \dots, N - 1$, appelons T_i le premier instant où une carte est insérée en-dessous de f pour la i -ème fois. Clairement, les v.a. $T_i - T_{i-1}$ sont indépendantes, et la loi de $T_i - T_{i-1}$ est une géométrique de paramètre i/N . De plus, on vérifie que, conditionnellement à la valeur de T_i , l'ordre des i cartes situées sous f à l'instant T_i est distribué uniformément sur l'ensemble des permutations possibles de ces i cartes (pour le voir, il suffit de vérifier que, pour tout k , l'insertion d'une carte supplémentaire à une position uniformément choisie parmi les $k + 1$ possibles, dans un jeu de k cartes dont l'ordre est une permutation uniforme, donne une permutation uniforme de ces $k + 1$ cartes). On en déduit que $1 + T_{N-1}$ est un temps stationnaire fort pour la chaîne considérée.

La marche aléatoire sur $\mathbb{Z}/2^a\mathbb{Z}$ (voir [15]).

On considère la marche aléatoire sur $\mathbb{Z}/2^a\mathbb{Z}$ dont les pas sont $+1, 0$ ou -1 avec une probabilité de $1/3$ pour chacun. Clairement, il s'agit d'une chaîne de Markov irréductible et apériodique dont la loi stationnaire est la loi uniforme. En définissant $T_{-1} = 0$ et, pour $i = 1, \dots, a - 1$, T_i comme le premier instant postérieur à T_{i-1} où l'on se trouve à distance 2^{a-i-1} de $X_{T_{i-1}}$, on vérifie par récurrence (en utilisant la symétrie du problème) que, conditionnellement à la valeur de T_i , la v.a. X_{T_i} est uniformément distribuée sur son support, qui, pour $i = a - 1$, est l'ensemble des points à distance impaire de X_0 . En définissant T_a comme le premier instant postérieur à T_{a-1} où l'on fait un pas égal à 0 ou $+1$, on voit que T_a est un temps stationnaire fort pour la chaîne.

Nous discuterons dans un autre chapitre les conclusions quantitatives qui peuvent être tirées des temps stationnaires forts introduits ci-dessus.

5.5 Approche spectrale

En se rappelant que l'objet que l'on étudie, à savoir la loi de X_n lorsque n est grand, peut se réécrire : $P(X_n = \cdot) = \mu p^n$, où μ désigne la loi initiale, et p le noyau de transition, on conçoit naturellement que les propriétés spectrales de

l'opérateur linéaire p (sur des espaces à préciser) puissent jouer un rôle important dans cette étude. Dans le cas où S est fini, on peut par exemple appliquer à p la théorie classique de Perron-Frobenius des matrices positives, pour établir les propriétés de base du spectre, comme nous le verrons ci-dessous, et en déduire le comportement asymptotique de μp^n .

Dans le cas où S est infini, l'étude des propriétés spectrales de p (en faisant agir p sur des espaces convenables) est également intéressante, mais fait appel à une sophistication technique bien plus grande, et nécessite des hypothèses supplémentaires sur le noyau considéré. Dans le cas réversible, on note que l'action de p sur $L^2(\nu)$ est auto-adjointe, et il est donc possible d'utiliser la théorie générale des opérateurs continus auto-adjoints sur les espaces de Hilbert, cette théorie (faisant appel à la notion de mesure spectrale) étant présentée par exemple dans [11, 47]. Dans ce qui suit, nous nous contenterons de présenter la théorie dans le cas où S est un ensemble fini.

5.5.1 Approche spectrale dans le cas où S est fini

Dans le cas où S est fini, on peut se contenter de considérer que p agit à droite et à gauche sur $\mathbb{C}^{|S|}$, vu soit comme l'espace des fonctions de S dans \mathbb{C} , soit comme l'espace des mesures complexes sur S , $[p(x, y)]_{(x, y) \in S^2}$ étant alors la matrice correspondant à l'action sur les fonctions (à droite), exprimée dans la base canonique, la transposée de cette matrice correspondant à l'action sur les mesures.

On a alors le théorème suivant :

Théorème 15 *Soit p un noyau de transition sur un ensemble S fini. Alors :*

- toutes les valeurs propres (dans \mathbb{C}) de p sont de module inférieur ou égal à 1 ;
- 1 est valeur propre de p ;
- si p est irréductible, la multiplicité (géométrique et algébrique) de 1 est égale à 1, le sous-espace propre associé (à droite) étant celui des fonctions constantes, et à gauche celui des mesures complexes invariantes ;
- si p est irréductible et apériodique, 1 est l'unique valeur propre de module 1.

Preuve :

On note qu'en ce qui concerne le spectre (valeurs propres, multiplicités algébriques et géométriques), une matrice et sa transposée ont les mêmes caractéristiques. On peut donc indifféremment étudier l'action à droite ou à gauche de p . On vérifie d'abord que $|fp| \leq |f|$, où, pour $f : S \rightarrow \mathbb{C}$, $|f| = \sum_{x \in S} |f(x)|$. On en déduit que toutes les valeurs propres complexes de p sont de module ≤ 1 . Le fait que 1 soit valeur propre se voit en vérifiant que la fonction constante égale à 1 est fixée par l'action de p . Le fait que la multiplicité géométrique de 1 soit égale à 1 si p est

irréductible se vérifie en considérant $f : S \rightarrow \mathbb{C}$ telle que $pf = f$, ce que l'on ramène, en considérant parties réelles et imaginaires sur lesquelles p agit séparément, à une fonction f à valeurs réelles, puis en introduisant s^* tel que $f(s^*) = \max_{s \in S} f(s)$ et en vérifiant que nécessairement, $f(s) = f(s^*)$ pour tout s tel que $p(s^*, s) > 0$, grâce à l'invariance de f sous l'action de p , puis en itérant. En appelant ν la probabilité invariante, on note que l'ensemble des fonctions F_0 des fonctions f vérifiant $\sum_{s \in S} f(s)\nu(s) = 0$ est stable par p , et constitue un supplémentaire de l'ensemble des fonctions constantes. On voit alors que 1 ne peut être valeur propre de p restreint à F_0 , ce qui prouve que 1 est également de multiplicité algébrique égale à 1. (On peut aussi considérer les mesures complexes de masse nulle en lieu et place de F_0 , et considérer l'action à gauche.)

Considérons maintenant un nombre complexe λ de module 1, et supposons l'existence d'une mesure complexe μ non-nulle telle que $\mu p = \lambda \mu$, p étant supposé irréductible et apériodique.

Nous allons prouver que l'égalité suivante a lieu pour tout $x \in S$ et $n \geq 1$:

$$\left| \sum_{y \in S} \mu(y) p^n(y, x) \right| = \sum_{y \in S} |\mu(y) p^n(y, x)|. \quad (5.4)$$

Notons d'abord que, pour x fixé, $\left| \sum_{y \in S} \mu(y) p^n(y, x) \right| \leq \sum_{y \in S} |\mu(y)| p^n(y, x)$. Or, par hypothèse, $\mu p = \lambda \mu$, d'où le fait que $\sum_{y \in S} \mu(y) p^n(y, x) = \lambda^n \mu(x)$. On en déduit que $\left| \sum_{y \in S} \mu(y) p^n(y, x) \right| = |\mu(x)|$, et, en sommant, que $\sum_{x \in S} \left| \sum_{y \in S} \mu(y) p^n(y, x) \right| = \sum_{x \in S} |\mu(x)|$. Par ailleurs, $\sum_{x \in S} \sum_{y \in S} |\mu(y)| p^n(y, x) = \sum_{x \in S} |\mu(x)|$. On en déduit l'égalité (5.4). Celle-ci entraîne en particulier que $\nu(x) := |\mu(x)|$ définit une mesure positive non-triviale invariante pour p , qui est donc propre. Par conséquent, $\mu(x) \neq 0$ pour tout $x \in S$.

A présent, rappelons qu'une égalité de la forme $|x_1 + \dots + x_k| = |x_1| + \dots + |x_k|$, où $x_1, \dots, x_k \in \mathbb{C}$, équivaut au fait que tous les x_i non-nuls possèdent le même argument (Pour $k = 2$, c'est le cas d'égalité dans l'inégalité de Cauchy-Schwarz, et le cas général s'en déduit par récurrence). Considérons donc $x \in S$, et $n \geq 1$ tel que $p^n(x, x) > 0$. On voit que les arguments de $\sum_{y \in S} \mu(y) p^n(y, x) = \lambda^n \mu(x)$ et de $p^n(x, x) \mu(x)$ doivent être égaux, ce qui entraîne que $\lambda^n = 1$. En utilisant le fait que la chaîne est apériodique, on en déduit que nécessairement $\lambda = 1$, ce qui conclut la preuve. \square

Remarque 24 *Au passage, on a montré que, lorsque la chaîne est périodique de période d , ses valeurs propres de module 1 sont nécessairement des racines d -èmes de l'unité.*

Le théorème précédent entraîne à présent automatiquement le théorème 11 : il suffit par exemple d'effectuer une réduction de Jordan, pour constater que, partant

d'une loi initiale quelconque μ , $P(X_n = \cdot)$ converge, lorsque n tend vers l'infini, vers la projection de μ sur le sous-espace propre associé à la valeur propre 1 en ayant pris comme supplémentaire les mesures signées de masse nulle. On retrouve le fait qualitatif que la convergence vers la loi stationnaire a lieu à vitesse exponentielle, et l'on constate que la connaissance du spectre de p , ou tout au moins de la distance de $\text{Spec}(p) \setminus \{1\}$ au cercle unité de \mathbb{C} , conditionne la vitesse de convergence.

Nous avons donné une preuve du théorème précédent utilisant le fait que la matrice associée à p est stochastique. Si l'on considère seulement des matrices positives, le théorème de Perron-Frobenius, dont il existe différentes versions, montre que certaines des propriétés ci-dessus demeurent valables, à condition d'être convenablement adaptées. Nous citons ci-dessous, sans donner de preuve, une version du théorème de Perron-Frobenius que nous aurons l'occasion d'utiliser. Rappelons qu'une matrice carrée $A = (a_{xy})$ indexée par $S \times S$ à coefficients dans \mathbb{R} est dite positive si ses coefficients sont tous positifs ou nuls, et irréductible si, pour tous $x, y \in S$, il existe $n \geq 0$ tel que $a_{xy}^{(n)} > 0$, où $A^n =: (a_{xy}^{(n)})$.

Théorème 16 *Soit A une matrice positive irréductible. Il existe une valeur propre $\lambda \in \mathbb{R}$ telle que*

- $\lambda > 0$;
- toute autre valeur propre de A possède un module $\leq \lambda$;
- il existe un vecteur propre de A associé à λ dont toutes les coordonnées sont strictement positives;
- la multiplicité géométrique de λ est égale à 1.

Exercice 163 *Prouver le théorème ci-dessus (ou aller en lire une preuve).*

5.6 Théorie L^2

5.6.1 Généralités

Dans toute cette partie p désigne un noyau ergodique de loi invariante ν . On rappelle que l'action de p sur les fonctions envoie $L^2(\nu)$ dans lui-même, et que l'adjoint de p pour le produit scalaire de $L^2(\nu)$ n'est autre que le noyau renversé dans le temps \hat{p} . Dans ce contexte, la distance entre deux fonctions est mesurée par la norme de $L^2(\nu)$, i.e. $\|f - g\|_2 := (\sum_{x \in S} (f(x) - g(x))^2)^{1/2}$, et la convergence vers la loi stationnaire ν pourra par exemple être mesurée sur les fonctions de $L^2(\nu)$, par $\|p^n f - \nu(f)\|_2$, qui n'est autre que l'écart-type de $f(X_n)$ sous \mathbb{P}_ν .

On peut également définir une distance L^2 entre μp^n et ν par la formule $\sum_{x \in S} |\mu(x)/\nu(x) - 1|^2 \nu(x)$ (avec la convention $0/0 = 1$), mais, dans le cas où S est infini, il se peut que cette distance prenne une valeur infinie. Notez que le fait que μ/ν (vue comme une

fonction sur S) se trouve dans $L^2(\nu)$ signifie que μ (vue comme une fonction sur S) se trouve dans $L^2(1/\nu)$.

On peut alors poser $d_2(t) := \sup_{x \in S} \left(\sum_{y \in S} |p^t(x, y)/\nu(y) - 1|^2 \nu(y) \right)^{1/2}$, et l'on vérifie que, par convexité, pour toute loi initiale μ , $\left(\sum_{x \in S} |\mu p^t(x)/\nu(x) - 1|^2 \nu(x) \right)^{1/2} \leq d_2(t)$.

Par ailleurs, on voit que l'on a l'inégalité

$$2d_{VT}(\mu, \nu) = \sum_{x \in S} \nu(x) |\mu(x)/\nu(x) - 1| \leq \left(\sum_{x \in S} |\mu(x)/\nu(x) - 1|^2 \nu(x) \right)^{1/2}.$$

D'autre part, on vérifie que \hat{p} est, comme p , ergodique de loi invariante ν , et que $(\mu p)/\nu = \hat{p}(\mu/\nu)$. Par conséquent, la convergence de μp^n vers ν peut être analysée dans notre contexte en étudiant la convergence dans $L^2(\nu)$ de l'action de \hat{p}^n sur les fonctions. Nous nous restreindrons donc à aborder le problème du point de vue de l'action sur les fonctions.

On se rappelle au passage que la suite $(\|p^n f - \nu(f)\|_2)_{n \geq 0}$ est décroissante du fait que la norme de l'action de p de L^2 dans lui-même est de norme ≤ 1 .

5.6.2 Formes de Dirichlet

La présentation de cette section est calquée sur celle de Jerrum [27], elle-même reprenant un argument dû à Mihail.

Etant donné un noyau p ergodique, et ν sa loi invariante, on définit la forme de Dirichlet associée à p comme la forme quadratique suivante sur $L^2(\nu)$:

$$\mathcal{E}_p(f, f) = \frac{1}{2} \sum_{x, y \in S} \nu(x) p(x, y) (f(y) - f(x))^2.$$

Il s'agit d'une forme quadratique sur $L^2(\nu)$ (on vérifie facilement qu'elle est bien définie, du fait que pf et f sont dans $L^2(\nu)$).

La forme de Dirichlet représente donc une mesure de la variation « locale » de f : partant d'un élément de S choisi selon la loi invariante ν , on mesure dans L^2 la variation de f lorsque l'on fait un pas de la chaîne.

On vérifie que $\mathcal{E}_p(f, f) = \mathcal{E}_{\hat{p}}(f, f)$.

Partant d'un noyau p , nous allons considérer le noyau $p_{par.}$ sur S (par. pour paresseux) défini par $p_{par.}(x, y) := (1/2)p(x, y)$ si $x \neq y$, et $p_{par.}(x, x) := 1/2 + 1/2p(x, x)$. On vérifie que $p_{par.}$ est également ergodique, et possède également ν pour loi invariante. En outre, $p_{par.}$ est réversible si et seulement si p l'est.

Exercice 164 *Partant d'une chaîne de Markov de noyau de transition p , comment fabriquer une chaîne de noyau de transition $p_{par.}$?*

Nous allons prouver le résultat suivant :

Proposition 65 *Pour tout $f \in L^2(\nu)$,*

$$\|p_{par}.f - \nu(f)\|_2^2 \leq \|f - \nu(f)\|_2^2 - (1/2)\mathcal{E}_p(f, f).$$

Preuve :

Sans perte de généralité, supposons $\nu(f) = 0$. On écrit que

$$p_{par}.f(x) = (1/2)f(x) + (1/2) \sum_{y \in S} p(x, y)f(y),$$

d'où

$$p_{par}.f(x) = (1/2) \sum_{y \in S} p(x, y)(f(x) + f(y)). \quad (5.5)$$

A présent, par définition,

$$\|p_{par}.f\|_2^2 = \sum_{x \in S} (p_{par}.f(x))^2 \nu(x).$$

En utilisant l'identité (5.5) établie plus haut, on en déduit que

$$\|p_{par}.f\|_2^2 = (1/4) \left(\sum_{x \in S} \left(\sum_{y \in S} p(x, y)^{1/2}(f(x) + f(y))p(x, y)^{1/2} \right)^2 \nu(x) \right).$$

Un petit coup d'inégalité de Cauchy-Schwarz, et nous obtenons que

$$\|p_{par}.f\|_2^2 = (1/4) \left(\sum_{x \in S} \left(\sum_{y \in S} p(x, y)(f(x) + f(y))^2 \right) \left(\sum_{z \in S} p(x, z) \right) \nu(x) \right),$$

et donc que

$$\|p_{par}.f\|_2^2 \leq (1/4) \left(\sum_{x \in S} \left(\sum_{y \in S} p(x, y)(f(x) + f(y))^2 \right) \nu(x) \right). \quad (5.6)$$

A présent, écrivons que

$$\|f\|_2^2 = (1/2) \sum_{x \in S} f(x)^2 \nu(x) + (1/2) \sum_{y \in S} f(y)^2 \nu(y),$$

soit encore, en utilisant l'invariance de ν par p pour le membre de droite,

$$\|f\|_2^2 = (1/2) \sum_{x, y \in S} f(x)^2 \nu(x) p(x, y) + (1/2) \sum_{x, y \in S} \nu(x) p(x, y) f(y)^2,$$

soit finalement

$$\|f\|_2^2 = (1/2) \sum_{x,y \in S} (f(x)^2 + f(y)^2) \nu(x) p(x,y). \quad (5.7)$$

En soustrayant (5.7) de (5.6), nous obtenons finalement que

$$\|f\|_2^2 - \|p_{par}.f\|_2^2 \geq (1/2) \mathcal{E}_p(f, f).$$

□

Ce résultat montre en particulier que la suite $(\|(p_{par.})^n f - \nu(f)\|_2^2)_{n \geq 1}$ est décroissante. De manière plus intéressante, notons que, si l'on dispose d'une inégalité du type $\mathcal{E}_p(f, f) \geq \gamma \|f - \nu(f)\|_2^2$, où $\gamma > 0$ on obtient des bornes exponentielles quantitatives sur la convergence vers la stationnarité. Une telle inégalité est appelée inégalité de Poincaré ; elle borne la variation L^2 globale d'une fonction, soit $\|f - \nu(f)\|_2^2$, en termes d'une somme de variations L^2 locales, soit $\mathcal{E}_p(f, f)$. Nous verrons plus loin comment obtenir des inégalités de Poincaré dans certaines situations.

Dans le cas où S est fini, une telle inégalité existe toujours, car on vérifie facilement que, si $\mathcal{E}_p(f, f) = 0$, f doit être une fonction constante, et donc on doit avoir $f = \nu(f)$. On conclut ensuite en utilisant l'homogénéité de $\mathcal{E}_p(f, f)$ et $\|f - \nu(f)\|_2^2$ pour se ramener à la sphère unité de $L^2(\nu)$, et un argument de compacité.

Nous verrons plus bas des techniques permettant de prouver ce type d'inégalité dans certains cas. Si cette inégalité vous paraît mystérieuse, vous pouvez noter que, en considérant une version en temps continu de la même chaîne, possédant exactement le même noyau de transition, et des taux de sauts constants égaux à 1, la forme de Dirichlet apparaît de manière plus naturelle : on a en effet l'identité

$$\frac{d}{dt} \|p_t f - \nu(f)\|_2^2 = -2\mathcal{E}_p(f, f),$$

qui suggère, par ailleurs, que l'inégalité de la proposition ci-dessus perd un facteur 4 par rapport à ce qui devrait être l'ordre de grandeur correct, un facteur 2 s'expliquant par le caractère paresseux de la chaîne considérée, un autre facteur 2 demeurant inexplicé.

5.6.3 Le cas réversible

Rappelons la propriété fondamentale des chaînes réversibles, en nous plaçant dans le cas ergodique : si ν désigne la loi invariante, l'action de p de $L^2(\nu)$ dans lui-même est autoadjointe. Cette propriété fait des chaînes réversibles un cadre intéressant pour l'utilisation des méthodes spectrales. Notons que, si S est un ensemble fini, cette propriété entraîne automatiquement que p est diagonalisable dans une base

orthonormée (pour le produit scalaire de $L^2(\nu)$) et que ses valeurs propres sont des nombres réels.

Cependant, hormis dans de rares cas, il est difficile, voire impossible de calculer précisément la diagonalisation d'un noyau réversible. Un exemple remarquable, dans le cas des marches aléatoires sur le groupe symétrique est fourni par [16] (voir également [14]), dans lequel l'analyse de Fourier sur le groupe symétrique (utilisant la théorie des représentations) est mise à contribution.

Proposition 66 *Pour une chaîne réversible, la forme de Dirichlet possède l'expression alternative suivante :*

$$\mathcal{E}_p(f, f) = \langle (I - p)f, f \rangle_{L^2(\nu)} .$$

Preuve :

Par définition, $\langle (I - p)f, f \rangle = \sum_{x, y \in S} \nu(x)p(x, y)f(x)(f(x) - f(y))$. Ceci se réécrit $\langle (I - p)f, f \rangle = \sum_{x, y \in S} \nu(y)p(y, x)f(y)(f(y) - f(x))$, et, par réversibilité, se réécrit encore $\langle (I - p)f, f \rangle = \sum_{x, y \in S} \nu(x)p(x, y)f(y)(f(y) - f(x))$. En écrivant $\langle (I - p)f, f \rangle$ comme la demi-somme de la première et de la dernière expression, on obtient la formule annoncée. \square

On note au passage qu'il est facile de vérifier que la valeur propre 1 est de multiplicité 1 dans ce contexte, en notant qu'une fonction f vérifiant $pf = f$ doit annuler la forme de Dirichlet d'après l'expression ci-dessus, ce qui entraîne très facilement le fait que f doit être constante, en utilisant la définition de $\mathcal{E}_p(f, f)$.

Dans le cas où S est fini, on a, en se rappelant que la valeur propre 1 est associée aux fonctions constantes et le fait que toutes les valeurs propres sont réelles et inférieures ou égales à 1 (et d'ailleurs > -1), la caractérisation classique suivante :

$$\inf\{1 - \lambda \in \text{Spec}(p); \lambda < 1\} = \inf \left\{ \frac{\mathcal{E}_p(f, f)}{\|f\|_2^2}; \langle f, 1 \rangle = 0, f \in L^2(\nu) \right\} .$$

Ou, en termes plus probabilistes, en notant que $\mathcal{E}_p(f, f)$ n'est pas modifiée par l'ajout à f d'une constante, et en notant que $\langle f, 1 \rangle = \sum_{x \in S} f(x)\nu(x)$,

$$\sup\{\lambda \in \text{Spec}(p); \lambda < 1\} = \inf \left\{ \frac{\mathcal{E}_p(f, f)}{\text{Var}_\nu(f)}; f \in L^2(\nu) \right\} .$$

(Cette caractérisation a encore un sens dans le cadre plus général de la décomposition spectrale en dimension infinie.)

On voit alors qu'une inégalité de Poincaré entraîne une borne sur la distance à 1 du spectre de p . On voit que le fait de considérer le noyau $p_{par.} = 1/2(I + p)$ permet de n'avoir que des valeurs propres entre 0 et 1 et que, par conséquent, l'inégalité de Poincaré entraîne une borne sur le «trou spectral» de $p_{par.}$, ce qui éclaire quelque peu les résultats de la partie précédente, dans le cas où p est réversible.

Exercice 165 *Fabriquez un exemple de chaîne ergodique réversible sur un ensemble S infini pour lequel $\inf \left\{ \frac{\mathcal{E}_p(f;f)}{\|f\|_2^2}; \langle f, 1 \rangle = 0, f \in L^2(\nu) \right\} = 0$. Pouvez-vous étudier la vitesse de convergence vers la loi stationnaire dans votre exemple ?*

5.7 Entropie relative

Il s'agit de l'approche sur laquelle nous en dirons le moins, notre but ici étant essentiellement de mentionner la possibilité d'utiliser d'autres distances entre mesures de probabilité que les distances L^p , et pour lesquelles d'autres outils que ceux que nous avons décrits peuvent être employés. En fait, il existe un vaste ensemble de techniques permettant d'étudier la convergence telle que mesurée par l'entropie relative, utilisant par exemple des inégalités fonctionnelles différentes de l'inégalité de Poincaré présentée ci-dessus (voir par exemple [48, 37] pour des exemples).

Définition 13 *Etant données deux mesures de probabilité μ et ν sur S , ν étant propre, on appelle entropie relative de μ par rapport à ν la quantité définie par*

$$H(\mu|\nu) := \sum_{x \in S} \mu(x) \log \frac{\mu(x)}{\nu(x)} = \sum_{x \in S} \frac{\mu(x)}{\nu(x)} \log \frac{\mu(x)}{\nu(x)} \nu(x)$$

(avec la convention suivante : $0 \log 0 = 0$).

Posons

$$\phi(u) := u \log u, \quad u \geq 0,$$

avec $\phi(0) := 0$. La définition de l'entropie relative se réécrit donc sous la forme

$$H(\mu|\nu) = \sum_{x \in S} \phi\left(\frac{\mu(x)}{\nu(x)}\right) \nu(x).$$

Comme ϕ est bornée inférieurement, on voit que $H(\mu|\nu)$ est toujours bien définie. Comme ϕ est convexe, l'inégalité de Jensen permet de voir que

$$H(\mu|\nu) \geq \phi\left(\sum_{x \in S} \frac{\mu(x)}{\nu(x)} \nu(x)\right) = \phi(1) = 0.$$

De plus, ϕ étant strictement convexe, l'égalité $H(\mu|\nu) = 0$ entraîne donc automatiquement que $\frac{\mu(y)}{\nu(y)} = \frac{\mu(z)}{\nu(z)}$ pour tous $y, z \in S$, d'où le fait que $\mu = \nu$.

On note que $H(\mu|\nu)$ ne définit pas une distance, car elle n'est pas symétrique vis-à-vis de μ et ν . L'entropie relative peut être comparée à la distance en variation totale grâce à l'inégalité importante suivante (dite de Csiszár-Kullback-Pinsker) :

Proposition 67

$$2d_{VT}(\mu, \nu) \leq (H(\mu|\nu))^{1/2}.$$

Preuve :

Posons $f(x) := 1 - \mu(x)/\nu(x)$. On a par définition $H(\mu|\nu) = \nu((1+f) \log(1+f))$. En notant que $\nu(f) = 0$, on a encore $H(\mu|\nu) = \nu((1+f) \log(1+f) - f)$. A présent, on utilise l'inégalité, valable pour tout $u \geq 0$, $(1+u) \log(1+u) - u \geq u^2/2(1+u/3)^{-1}$, d'où l'on tire que $H(\mu|\nu) \geq \nu(f^2/2(1+f/3)^{-1}) = \nu(f^2/2(1+f/3)^{-1})\nu(1+f/3)$. Par Cauchy-Schwarz, on en tire que $H(\mu|\nu) \geq (1/2)\nu(f)^2$. \square

On retrouve en particulier le fait que $H(\mu|\nu) \geq 0$, avec égalité si et seulement si $\mu = \nu$. Donnons-nous à présent un noyau ergodique p sur S . Par convexité, on a que, pour toute loi initiale μ ,

$$H(\mu p^n | \nu) \leq \sup_{x \in S} H(p^n(x \cdot) | \nu).$$

Une propriété fondamentale de l'entropie dans le contexte des chaînes de Markov est la suivante :

Proposition 68 *Etant donné un noyau ergodique p , soit ν sa loi invariante. Alors, pour toute loi de probabilité μ sur S ,*

$$H(\mu p | \nu) \leq H(\mu | \nu),$$

et, si $H(\mu p | \nu) = H(\mu | \nu) < +\infty$, $\mu = \nu$.

Preuve :

On écrit que

$$H(p\mu | \nu) = \sum_{x \in S} \nu(x) \phi \left(\frac{\sum_{y \in S} \mu(y) p(y, x)}{\nu(x)} \right).$$

On réécrit que

$$H(p\mu | \nu) = \sum_{x \in S} \nu(x) \phi \left(\sum_{y \in S} \frac{\mu(y)}{\nu(y)} \frac{\nu(y) p(y, x)}{\nu(x)} \right).$$

On voit réapparaître le noyau renversé dans le temps \hat{p} , qui permet d'obtenir que, par convexité de ϕ ,

$$H(p\mu | \nu) \leq \sum_{x \in S} \nu(x) \sum_{y \in S} \phi \left(\frac{\mu(y)}{\nu(y)} \right) \frac{\nu(y) p(y, x)}{\nu(x)} = H(\mu | \nu).$$

Par stricte convexité, on déduit que l'égalité dans l'inégalité ci-dessus entraîne que, pour tout x , $\frac{\mu(y)}{\nu(y)} = \frac{\mu(z)}{\nu(z)}$ pour tous $y, z \in S$ tels que $p(y, x) > 0$ et $p(z, x) > 0$. Le fait que $\mu = \nu$ s'en déduit par irréductibilité et apériodicité. \square

Dans le cas où S est fini, on peut redémontrer a minima le théorème 11 de la manière suivante. On note d'abord que l'ensemble des lois de probabilités sur S est

compact (pour la topologie héritée de celle de \mathbb{R}^S). En vertu de la proposition ci-dessus, la suite $(H(\mu p^n | \nu))_{n \geq 0}$ est décroissante, et minorée par zéro. Appelons h sa limite. Considérons maintenant une sous-suite convergente de $(\mu p^n)_{n \geq 0}$, et notons μ_1 sa limite. On vérifie facilement que $H(\mu_1 | \nu) = h$. De plus, on a également le fait que $(\mu p^{n+1})_{n \geq 0}$ converge vers $\mu_1 p$. Mais on doit également avoir que $H(\mu_1 p | \nu) = h$. Conclusion : $\mu_1 p = \mu_1$, et donc $\mu_1 = \nu$.

Exercice 166 *Peut-on utiliser un tel argument basé sur la stricte convexité avec les distances précédemment introduites ?*

Exercice 167 *Dans tout cet exercice, S désigne un ensemble fini, et p un noyau de transition sur S supposé irréductible et apériodique. La loi invariante de p est notée ν . Pour tout $x \in S$, on note $T(x) := \inf\{n \geq 0; X_n = x\}$ en se plaçant sur l'espace canonique des trajectoires, et avec la convention $\inf \emptyset = +\infty$.*

L'objectif principal de cet exercice est d'établir les deux identités suivantes, valables pour tous $x, y \in S$:

$$\nu(x) \mathbb{E}_\nu(T(x)) = Z(x, x), \quad (5.8)$$

$$\nu(y) \mathbb{E}_x(T(y)) = Z(y, y) - Z(x, y), \quad (5.9)$$

où

$$Z(x, y) := \sum_{n=0}^{+\infty} (p^n(x, y) - \nu(y)).$$

1) *Expliquer pourquoi la série définissant $Z(x, y)$ est absolument convergente pour tous $x, y \in S$. Montrer que, pour tout $x \in S$,*

$$\sum_{y \in S} Z(x, y) = 0. \quad (5.10)$$

2)

Considérons $x \in S$, un entier $n_0 \geq 0$ fixé, et définissons

$$S_0 := \inf\{n \geq n_0; X_n = x\},$$

avec la convention $\inf \emptyset = +\infty$. Montrer que S_0 est presque sûrement fini sous \mathbb{P}_x , et que $\mathbb{E}_x(S_0) < +\infty$. Montrer ensuite que

$$\mathbb{E}_x(\text{card } \{0 \leq j \leq S_0 - 1; X_j = x\}) = \nu(x) \mathbb{E}_x(S_0).$$

(Indication : utiliser par exemple une décomposition de renouvellement basée sur S_0 .)

3) *Déduire de la question précédente que*

$$\sum_{n=0}^{n_0-1} p^n(x, x) = \nu(x)(n_0 + \mathbb{E}_\mu(T(x))),$$

où $\mu(\cdot) = \mathbb{P}_x(X_{n_0} = \cdot)$.

4) En déduire l'identité (5.8).

5) Pour $x \neq y$, et un entier $n_0 \geq 0$ fixé, définissons

$$S'_0 := \inf\{n \geq T(x) + n_0; X_n = y\},$$

avec la convention $\inf \emptyset = +\infty$. Montrer que S'_0 est presque sûrement fini sous \mathbb{P}_y , et que $\mathbb{E}_y(S'_0) < +\infty$. Montrer ensuite que

$$\mathbb{E}_y(\text{card } \{0 \leq j \leq S'_0 - 1; X_j = y\}) = \nu(y)\mathbb{E}_y(S'_0).$$

(Indication : utiliser par exemple une décomposition de renouvellement basée sur S'_0 .)

6) En déduire que

$$\mathbb{E}_y(\text{card } \{0 \leq j \leq T(x) - 1; X_j = y\}) + \sum_{n=0}^{n_0-1} p^n(x, y)$$

est égal à

$$\nu(y) (\mathbb{E}_y(T(x)) + n_0 + \mathbb{E}_\mu(T(y))),$$

où $\mu(\cdot) = \mathbb{P}_x(X_{n_0} = \cdot)$.

7) En déduire l'identité

$$\mathbb{E}_y(\text{card } \{0 \leq j \leq T(x) - 1; X_j = y\}) = \nu(y)(\mathbb{E}_y(T(x)) + \mathbb{E}_x(T(y))). \quad (5.11)$$

8) Déduire (5.9) de ce qui précède.

9) Soit N la matrice indexée par $S \times S$ définie par $N(x, y) := \nu(y)$, soit I la matrice identité indexée par $S \times S$. En voyant p comme la matrice indexée par $S \times S$ définie par $p(x, y)$, montrer que $I - (p - N)$ est inversible et que

$$Z + N = (I - (p - N))^{-1}.$$

En déduire comment calculer numériquement les valeurs $Z(x, y)$ à partir de la connaissance du noyau p .

10) Peut-on généraliser ce qui précède au cas d'une chaîne ergodique sur un ensemble S dénombrable ? Quelles sont les obstructions éventuelles ?

Chapitre 6

Une première approche quantitative de l'ergodicité pour la distance en variation totale

Dans ce chapitre, nous discutons d'une première famille d'approches pour quantifier plus précisément l'ergodicité d'une chaîne de Markov pour la distance en variation totale. Cette approche est "quantitative" au sens où elle s'intéresse à la vitesse de convergence d'une chaîne vers sa loi stationnaire, mais elle ne donne de cette vitesse qu'une caractérisation plutôt qualitative, étant davantage destinée à délimiter des grandes classes de comportement asymptotique, plutôt qu'à fournir des bornes non-asymptotiques explicites sur l'écart à la stationnarité. Des exemples de telles bornes non-asymptotiques seront donnés dans un chapitre ultérieur.

Dans tout ce chapitre, p désigne un noyau ergodique sur un ensemble fini ou dénombrable S , ν étant la loi invariante.

6.1 Ergodicité de degré 2

Une première manière de quantifier la vitesse de convergence de la loi d'une chaîne vers sa loi stationnaire en utilisant la distance en variation totale, consiste à considérer la série

$$\sum_{n \geq 0} d_{VT}(\delta_x p^n, \nu p^n),$$

et à étudier quand la somme de celle-ci est finie.

Le théorème suivant fournit plusieurs caractérisations de cette propriété, et, lorsque l'une des conditions équivalentes ci-dessus est vérifiée, on dit que p est **ergodique de degré 2**.

Théorème 17 *Il y a équivalence entre les conditions suivantes :*

- (i) *Il existe $x \in S$ tel que $\mathbb{E}_x(T_1(x)^2) < +\infty$;*
- (ii) *Il existe $x \in S$ tel que $\mathbb{E}_\nu(T_1(x)) < +\infty$;*
- (iii) *Il existe $x \in S$ tel que $\mathbb{E}_\nu(T_1(x)) < +\infty$;*
- (iv) *Pour tout $x \in S$ tel que $\mathbb{E}_x(T_1(x)^2) < +\infty$;*
- (v) *Pour un $x \in S$,*

$$\sum_{n \geq 0} d_{VT}(\delta_x p^n, \nu) < +\infty;$$

- (vi) *Pour tout $x \in S$,*

$$\sum_{n \geq 0} d_{VT}(\delta_x p^n, \nu) < +\infty;$$

- (vii) *On a*

$$\sum_{x \in S} \nu(x) \left(\sum_{n=0}^{+\infty} d_{VT}(\delta_x p^n, \nu) \right) < +\infty.$$

Une conséquence immédiate de (vii) est qu'il existe une suite c_n de nombres positifs satisfaisant $\sum_{n=0}^{+\infty} c_n < +\infty$ et telle que, pour tout $x \in S$, et tout $n \geq 0$,

$$d_{VT}(\delta_x p^n, \nu) \leq \frac{c_n}{\nu(x)}.$$

On dispose donc en quelque sorte d'une borne uniforme sur la vitesse de convergence de $d_{VT}(\delta_x p^n, \nu)$ vers zéro lorsque n tend vers l'infini, modulée par la valeur de $\nu(x)$.

L'équivalence de (i)-(ii)-(iii)-(iv) fait l'objet de l'exercice 126, et résulte assez facilement de l'identité établie par l'exercice 125.

Preuve de (v) \Rightarrow (ii):

Introduisons, comme dans l'exercice 167, le temps d'arrêt

$$S_0 := \inf\{n \geq n_0; X_n = x\},$$

avec la convention $\inf \emptyset = +\infty$, n_0 étant un entier fixé. On montre comme dans l'exercice (qui est formulé dans le cas où l'espace est fini, mais l'adaptation au cas positivement récurrent sur un espace dénombrable ne pose pas de problème), que S_0 est presque sûrement fini sous \mathbb{P}_x , que $\mathbb{E}_x(S_0) < +\infty$, et que

$$\sum_{n=0}^{n_0-1} (p^n(x, x) - \nu(x)) = \nu(x) \mathbb{E}_\mu(T(x)).$$

où $\mu(\cdot) = \mathbb{P}_x(X_{n_0} = \cdot)$. Réécrivons

$$\mathbb{E}_\mu(T(x)) = \sum_{y \in S} \mathbb{P}_x(X_{n_0} = y) \mathbb{E}_y(T(x)).$$

Par ergodicité de p , on a, pour tout $y \in S$, la convergence $\lim_{n_0 \rightarrow +\infty} \mathbb{P}_x(X_{n_0} = y) = \nu(x)$. Le lemme de Fatou entraîne donc que

$$\liminf_{n_0 \rightarrow +\infty} \sum_{y \in S} \mathbb{P}_x(X_{n_0} = y) \mathbb{E}_y(T(x)) \geq \sum_{y \in S} \nu(y) \mathbb{E}_y(T(x)) = \mathbb{E}_\nu(T(x)),$$

d'où l'on déduit que

$$\liminf_{n_0 \rightarrow +\infty} \sum_{n=0}^{n_0-1} (p^n(x, x) - \nu(x)) \geq \mathbb{E}_\nu(T(x)).$$

Cette inégalité prouve automatiquement que (v) \Rightarrow (ii). □

On note que la preuve ci-dessus montre en fait que l'existence d'un x tel que $\sum_{n=0}^{n_0-1} |p^n(x, x) - \nu(x)| < +\infty$, qui est moins forte que (v), suffit à entraîner (ii), et se trouve donc en fait équivalente aux autres propriétés mentionnées dans le théorème. La preuve ci-dessus suffit également à montrer, (voir la première question de l'exercice 126), qu'il existe bel et bien des exemples de chaînes ergodiques pour lesquels (v) n'est pas vérifiée.

Nous allons maintenant prouver que l'une quelconque des propriétés équivalentes (i)-(ii)-(iii)-(iv) entraîne (vii), en utilisant l'approche par couplage pour contrôler la vitesse de convergence en variation totale.

Nous commençons par introduire et étudier un objet qui nous sera utile dans l'analyse du temps de couplage correspondant. Considérons une suite S_1, S_2, \dots de variables aléatoires i.i.d. à valeurs dans $\{1, 2, \dots\}$. Pour tout $n \geq 1$, notons

$$Z_n := S_1 + \dots + S_n,$$

posons également $Z_0 := 0$, et, étant donné un entier $t \geq 0$, soit

$$H_t := \inf\{Z_n - t; Z_n \geq t, n \geq 0\}.$$

Il est clair que H_t est une variable aléatoire presque sûrement finie. On définit le noyau de transition q sur \mathbb{N} par

$$q(t, s) := P(H_t = s).$$

Pour expliquer l'utilité de q , considérons deux suites de variables aléatoires i.i.d. indépendantes et de même loi que S_1 , soit S'_1, S'_2, \dots et S''_1, S''_2, \dots , et définissons récursivement une suite de variables aléatoires $(W_n)_{n \geq 0}$ par $W_0 := 0$, $W_1 := t$, puis, récursivement, pour tout $n \geq 0$,

$$W_{2n+2} := (\inf\{S'_1, S'_1 + S'_2, \dots\} \cap [W_{2n+1}, +\infty[),$$

et tout $n \geq 1$,

$$W_{2n+1} := (\inf\{t + S_1'', t + S_1'' + S_2'', \dots\} \cap [W_{2n}, +\infty[).$$

A présent, posons, pour tout $n \geq 1$, $L_n := W_n - W_{n-1}$. On a alors la proposition suivante :

Proposition 69 *La suite $(L_n)_{n \geq 1}$ est une chaîne de Markov de noyau de transition q .*

Exercice 168 *Prouver la proposition ci-dessus.*

Proposition 70 *Sous l'hypothèse que $E(S_1) < +\infty$ et que $\text{pgcd}(\{i \in \mathbb{N}; P(S_1 = i) > 0\}) = 1$, il existe $\beta_1, \beta_2 > 0$ tel que, pour tout $n \geq 0$,*

$$\sup_{t \geq 0} \mathbb{P}_{t,q}(T_1(0) \geq n) \leq \beta_1 \exp(-\beta_2 n).$$

Preuve :

En notant que $q(t, 0) = P(s \in (\inf\{S_1, S_1 + S_2, \dots\}))$, et en appliquant le théorème du renouvellement, nous en déduisons que

$$\lim_{t \rightarrow +\infty} q(t, 0) = 1/E(S_1) > 0. \quad (6.1)$$

Par ailleurs, pour tout $t \in \mathbb{N}$ donné, grâce à notre hypothèse sur le support de la loi de S_1 , nous pouvons prouver qu'il existe d_1, \dots, d_a et e_1, \dots, e_b tous dans le support de S_1 et tels que $t + d_1 + \dots + d_a = e_1 + \dots + e_b$. On en déduit, au vu de la proposition 69, qu'il existe $m \geq 1$ tel que

$$q^m(t, 0) > 0.$$

En combinant cette dernière propriété avec (6.1), on en déduit la conclusion. \square

Proposition 71 *Sous l'hypothèse que $E(S_1^2) < +\infty$, on a*

$$\sup_{t \geq 1} E(H_t) < +\infty.$$

Preuve :

En décomposant selon la valeur du plus grand Z_i strictement inférieur à t , on voit facilement que

$$P(H_t \geq u) = \sum_{0 \leq a \leq t-1} \sum_{\ell=0}^{+\infty} P(Z_\ell = a, S_{\ell+1} \geq t - a + u).$$

En notant $\alpha(x) := P(S_1 \geq x)$, le caractère i.i.d. des (S_i) entraîne que

$$P(Z_\ell = a, S_{\ell+1} \geq t - a + u) = P(Z_\ell = a)\alpha(t - a + u).$$

Comme on a de plus que

$$\sum_{\ell=0}^{+\infty} P(Z_\ell = a) = E\left(\sum_{\ell=0}^{+\infty} \mathbf{1}(Z_\ell = a)\right) \leq 1,$$

du fait que la suite (Z_i) est strictement croissante, on a l'inégalité

$$P(H_t \geq u) \leq \sum_{0 \leq a \leq t-1} \alpha(t - a + u) = \sum_{b=1}^t \alpha(u + b).$$

En sommant, on en déduit que, pour tout $t \geq 1$,

$$E(H_t) + 1 \leq \sum_{u=0}^{+\infty} \sum_{b=1}^{+\infty} \alpha(u + b) = E(S_1(S_1 + 1)/2).$$

□

Définissons

$$K := \inf\{k \geq 2; L_k = 0\},$$

avec la convention $\inf \emptyset = +\infty$. Au vu de la proposition 70, on déduit facilement que, sous les hypothèses de la proposition, K est presque sûrement fini, et même qu'il existe une constante c ne dépendant pas de t telle que

$$E(K) \leq c < +\infty.$$

Proposition 72 *Sous l'hypothèse que $E(S_1) < +\infty$ et que $\text{pgcd}(\{i \in \mathbb{N}; P(S_1 = i) > 0\}) = 1$, il existe une constante C ne dépendant pas de t telle que*

$$E(W_K - W_1) \leq C.$$

Preuve :

On écrit $W_K - W_1$ sous la forme

$$W_K - W_1 = \sum_{k=1}^{K-1} (W_{k+1} - W_k),$$

d'où

$$E(W_K - W_1) = \sum_{k=1}^{+\infty} E((W_{k+1} - W_k)\mathbf{1}(K > k)).$$

A présent, observons que, d'après sa définition, l'événement $K > k$ est mesurable par rapport à $\sigma(W_0, \dots, W_k)$, tandis que, conditionnellement à $\sigma(W_0, \dots, W_k)$, la loi de $W_{k+1} - W_k$ n'est autre que $q(L_{n-1}, \cdot)$. On en déduit que

$$E(W_K - W_1) \leq \left(\sup_{s \geq 1} E(H_s) \right) \sum_{k=1}^{+\infty} P(K > k) \leq \left(\sup_{s \geq 1} E(H_s) \right) \times c.$$

La proposition 71 montre ensuite que $\sup_{s \geq 1} E(H_s) < +\infty$. \square

Preuve de (i)-(ii)-(iii)-(iv) \Rightarrow (vii) :

Considérons le couplage employé dans 5.3, dont la loi du temps de couplage est celle du premier temps d'atteinte T de la diagonale par la chaîne $(X_n^1, X_n^3)_{n \geq 0}$ lorsque la loi initiale est la loi produit $\nu \otimes \nu$. Considérons $x \in S$, et définissons

$$B_0^1 := \inf\{n \geq 0; X_n^1 = x\}, \quad B_0^2 := \inf\{n \geq 0; X_n^3 = x\}.$$

Par récurrence, définissons également, pour tout $i \geq 1$,

$$B_i^1 := \inf\{n \geq B_{i-1}^1 + 1; X_n^1 = x\}, \quad B_i^2 := \inf\{n \geq B_{i-1}^2 + 1; X_n^3 = x\}.$$

A présent, si $B_0^1 \leq B_0^2$, on pose, pour tout $i \geq 1$,

$$S'_i := B_i^1 - B_{i-1}^1, \quad S''_i := B_i^2 - B_{i-1}^2.$$

Inversement, si $B_0^1 > B_0^2$, on pose, pour tout $i \geq 1$,

$$S'_i := B_i^2 - B_{i-1}^2, \quad S''_i := B_i^1 - B_{i-1}^1.$$

La propriété de Markov entraîne bien que, conditionnellement à B_0^1, B_0^2 , les suites (S'_i) et (S''_i) sont bien indépendantes, et i.i.d. ayant pour loi commune celle de $T_1(x)$ sous $\mathbb{P}_{x,p}$. Posons maintenant $t := \max(B_0^2 - B_0^1, B_0^1 - B_0^2)$, et notons que, avec les définitions précédentes, $X_{W_K}^1 = X_{W_K}^3 = x$ et donc que $T \leq \max(B_0^1, B_0^2) + (W_K - W_1)$, d'où

$$E(T) \leq E(\max(B_0^1, B_0^2)) + E(W_K - W_1).$$

L'hypothèse (ii) montre que $E(B_0^1) < +\infty$ et $E(B_0^2) < +\infty$, tandis l'hypothèse (i) permet d'appliquer la proposition 72 qui montre que $E(W_K) \leq C$. \square

On note que les implications restantes formulées dans le théorème 17 sont soit triviales, soit résultent de celles déjà prouvées ci-dessus.

Nous renvoyons par exemple à [40, 9, 36] pour plus d'informations ainsi que des références sur la notion d'ergodicité de degré 2.

6.2 Ergodicité géométrique

La notion d'ergodicité géométrique est l'une des formalisations possible de la convergence «à vitesse exponentielle» d'une chaîne ergodique vers sa loi invariante. Par rapport à la notion précédente, on ne demande plus seulement la convergence de la série de terme général $d_{VT}(\delta_x p^n, \nu p^n)$, mais la convergence exponentiellement rapide de ce terme vers zéro, soit le fait qu'il existe $r < 1$ tel que, lorsque $n \rightarrow +\infty$,

$$d_{VT}(\delta_x p^n, \nu p^n) = O(r^n)$$

Le théorème suivant fournit plusieurs caractérisations de cette propriété, et, lorsque l'une des conditions équivalentes ci-dessus est vérifiée, on dit que p est **géométriquement ergodique**.

Théorème 18 *Etant donné un noyau de transition ergodique p sur un ensemble S fini ou dénombrable, de loi invariante ν , il y a équivalence entre les propriétés suivantes :*

- (i) *Il existe $x \in S$ tel que $T_1(x)$ possède une queue sous-géométrique sous \mathbb{P}_x ;*
- (ii) *Pour tout $x \in S$, la queue de $T_1(x)$ est sous-géométrique sous \mathbb{P}_x ;*
- (iii) *Il existe $x \in S$, $0 < r < 1$ et $A \geq 0$ tel que, pour tout $n \geq 0$,*

$$d_{VT}(p^n(x, \cdot), \nu) \leq Ar^n;$$

- (iv) *Pour tout $x \in S$, il existe $0 < r < 1$ et $A \geq 0$ tel que, pour tout $n \geq 0$,*

$$d_{VT}(p^n(x, \cdot), \nu) \leq Ar^n;$$

- (v) *Il existe $A \geq 0$ et $0 < r < 1$ tels que pour tout $n \geq 0$,*

$$\sum_{x \in S} \nu(x) d_{VT}(p^n(x, \cdot), \nu) \leq Ar^n.$$

Une conséquence immédiate de (v) est qu'il existe $0 < r < 1$, $A \geq 0$ tels que, pour tout $x \in S$, et tout $n \geq 0$,

$$d_{VT}(\delta_x p^n, \nu) \leq \frac{Ar^n}{\nu(x)}.$$

On peut donc en quelque sorte choisir uniformément par rapport à x la vitesse de convergence exponentielle de $d_{VT}(\delta_x p^n, \nu)$ vers zéro lorsque n tend vers l'infini, modulée par la valeur de $\nu(x)$.

Exercice 169 *Montrer que l'ergodicité géométrique entraîne, pour tout $\ell \geq 1$, l'existence d'un $r < 1$ et une fonction $A_\ell : S \rightarrow \mathbb{R}_+$ telle que, pour tout $n \geq 0$,*

$$d_{VT}(p^n(x, \cdot), \nu) \leq A_\ell(x)r^n$$

et telle que l'on ait de plus

$$A_\ell \in L^\ell(\nu).$$

(Indication : noter que $d_{VT}(p^n(x, \cdot), \nu) \leq 1$ et que $u^\ell \leq u$ lorsque $0 \leq u \leq 1$.)

Avant de donner la preuve du théorème, procédons à quelques rappels sur les variables aléatoires à queue sous-géométrique.

Pour commencer, une variable aléatoire ξ à valeurs entières positives est dite posséder une **queue sous-géométrique** s'il existe deux nombres $a > 0$ et $b > 0$ tels que l'estimation suivante soit vérifiée :

$$\text{pour tout } x \geq 0, P(\xi \geq x) \leq a \exp(-bx).$$

Notons la caractérisation suivante :

Proposition 73 *La variable ξ possède une queue sous-géométrique si et seulement s'il existe $\theta > 1$ tel que $E(\theta^\xi) < +\infty$.*

Proposition 74 *La somme de deux variables aléatoires à queue sous-géométrique est à queue sous-géométrique.*

Exercice 170 *Prouver les deux propositions ci-dessus.*

Proposition 75 *Soit $(Y_i)_{i \geq 1}$ une suite de variables aléatoires définies sur un espace de probabilité (O, \mathcal{H}, Q) , et $(\mathcal{H}_i)_{i \geq 0}$ une famille croissante de sous-tribus de \mathcal{H} telles que $\mathcal{H}_0 = \{\emptyset, O\}$. Supposons que :*

- pour tout $i \geq 1$, Y_i est mesurable par rapport à \mathcal{H}_i ;
- il existe $0 < c_1, c_2 < +\infty$ tels que, pour tout $i \geq 1$ et $t \geq 0$, $Q(Y_i \geq t | \mathcal{H}_{i-1}) \leq c_1 \exp(-c_2 t)$.

Alors, pour tout $t > c_1/c_2$, il existe $0 < c_3, c_4 < +\infty$ tels que, pour tout $n \geq 1$,

$$Q(Y_1 + \dots + Y_n \geq nt) \leq c_3 \exp(-c_4 n)$$

Preuve de la proposition 75:

Pour $0 < \lambda < c_2$, on a, en utilisant une intégration par parties, le fait que

$$E(\exp(\lambda Y_i) | \mathcal{H}_{i-1}) \leq 1 + \int_0^{+\infty} \lambda \exp(\lambda t) Q(Y_i \geq t | \mathcal{H}_{i-1}) dt \leq 1 + c_1 \lambda (c_2 - \lambda)^{-1}.$$

On en déduit que

$$E(\exp(\lambda(Y_1 + \dots + Y_n))) \leq (1 + c_1 \lambda (c_2 - \lambda)^{-1})^n.$$

Ensuite, d'après l'inégalité de Markov,

$$Q(Y_1 + \dots + Y_n \geq nt) \leq \exp(-n\lambda t) E(\exp(\lambda(Y_1 + \dots + Y_n))),$$

si bien que

$$Q(Y_1 + \dots + Y_n \geq nt) \leq \exp(-n(\lambda t + \log(1 + c_1\lambda(c_2 - \lambda)^{-1}))).$$

En choisissant λ assez petit, on obtient le résultat. □

Nous allons commencer par prouver que (i) \Rightarrow (vi).

La preuve consiste, comme dans la section précédente, à obtenir une borne sur le temps de couplage. Au lieu de la proposition 71, nous utiliserons la proposition suivante.

Proposition 76 *Si $T_1(x)$ possède une queue sous-géométrique sous \mathbb{P}_x , c'est également le cas pour $q(t, \cdot)$, uniformément par rapport à t , ou, plus précisément : il existe $f, h > 0$ tels que, pour tout $t \geq 0$, et tout $s \geq 0$,*

$$q(t, s) \leq f \exp(-hs).$$

Preuve de la proposition 76:

Supposons que, pour tout $x \geq 0$, $P(S_1 \geq x) \leq a \exp(-bx)$. Nous allons montrer, avec les notations de la section précédente, qu'il existe $f > 0$ tel que, pour tout $s \geq 0$,

$$P(H_t \geq k) \leq f \exp(-bk). \tag{6.2}$$

Reprenant l'argument employé pour la preuve de la proposition 71, on a que

$$P(H_t \geq u) \leq \sum_{0 \leq a \leq t-1} \alpha(t-a+u) = \sum_{w=1}^t \alpha(u+w).$$

D'après notre hypothèse,

$$\alpha(u+w) \leq a \exp(-b(u+w)),$$

et l'on a donc

$$P(H_t \geq u) \leq \exp(-bu) \sum_{w=1}^t \exp(-bw) \leq f \exp(-bu),$$

en posant $f = \sum_{w=1}^t \exp(-bw) = \exp(-b)(1 - \exp(-b))^{-1}$. □

De même, la proposition 72 est remplacée par la suivante.

Proposition 77 *Sous l'hypothèse que S_1 possède une queue sous-géométrique et que $\text{pgcd}(\{i \in \mathbb{N}; P(S_1 = i) > 0\}) = 1$, il existe deux constantes $d, g > 0$ ne dépendant pas de t telles que, pour tout $k \geq 0$,*

$$P(W_K - W_1 \geq k) \leq d \exp(-gk).$$

Preuve :

Supposons donc que, pour tout $x \geq 0$, $P(S_1 \geq x) \leq a \exp(-bx)$, et écrivons W_K sous la forme

$$W_K - W_1 := \sum_{k=0}^{K-1} (W_{k+1} - W_k).$$

Considérons à présent $\gamma > a/b$, et $m := \lfloor k/\gamma \rfloor$. D'après la proposition 70,

$$P(K \geq m) \leq \beta_1 \exp(-\beta_2 m).$$

D'autre part, du fait que la suite $(W_j)_j$ est croissante et que, par définition, $k \geq m\gamma$, on a

$$P(W_K - W_1 \geq k, K \leq m) \leq P(W_m - W_1 \geq k) \leq P(W_m - W_1 \geq m\gamma).$$

En appliquant la proposition 75, on obtient que

$$P(W_m - W_1 \geq m\gamma) \leq c_3 \exp(-c_4 m).$$

La proposition en résulte. □

Preuve de (i) \Rightarrow (v):

On reprend l'argument de la preuve donnée dans la section précédente, pour la partie du théorème 17 affirmant l'implication (i)-(ii)-(iii)-(iv) \Rightarrow (vii). On obtient ainsi que

$$T \leq \max(B_0^1, B_0^2) + W_K - W_1.$$

La proposition 77 entraîne que $W_K - W_1$ possède une queue sous-géométrique. D'autre part, l'identité établie dans l'exercice 125 montre que, si $T_1(x)$ possède une queue sous-géométrique par rapport à \mathbb{P}_x , c'est également le cas par rapport à \mathbb{P}_ν . On arrive à la conclusion annoncée. □

Le fait que (v) \Rightarrow (iv), (v) \Rightarrow (iii), (ii) \Rightarrow (i) est immédiat.

Preuve de (iii) \Rightarrow (i):

La preuve présentée ici reprend la preuve donnée dans [36] du théorème de Kendall sur le renouvellement (p. 358). Cette preuve est basée sur des considérations

d'analyse complexe¹, et il ne semble pas exister de preuve faisant appel à des considérations plus probabilistes, par exemple par couplage.

Supposons (iii) vérifiée. On en déduit facilement que la série entière

$$F(z) := \sum_{n=1}^{+\infty} (\mathbb{P}_x(X_n = x) - \mathbb{P}_x(X_{n-1} = x)) z^n,$$

possède un rayon de convergence $r_0 > 1$.

Introduisons également

$$U(z) := \sum_{n=0}^{+\infty} \mathbb{P}_x(X_n = x) z^n,$$

dont le rayon de convergence est nécessairement ≥ 1 , vu que les coefficients sont des probabilités. Dans le disque $\{|z| < 1\}$, on a l'égalité

$$F(z) = (1 - z)U(z) - 1. \tag{6.3}$$

Introduisons à présent la série

$$P(z) := \sum_{n=0}^{+\infty} \mathbb{P}_x(T_1(x) = n) z^n,$$

Grâce à l'identité (5.1), on vérifie que l'on a, dans le disque $\{|z| < 1\}$, l'identité

$$U(z) = (1 - P(z))^{-1}.$$

On en déduit alors de (6.3) que, sur $\{|z| < 1\}$, $(1 - z)(1 - P(z))^{-1} = 1 + F(z)$, d'où

$$P(z) = (z + F(z))(F(z) + 1)^{-1}. \tag{6.4}$$

Nous avons vu que F est analytique sur $\{|z| < r_0\}$. Par conséquent, pour tout $0 < \epsilon < r_0 - 1$, $1 + F(z)$ ne peut avoir qu'un nombre fini de zéros dans $\{|z| < r_0 - \epsilon\}$ (sauf à être identiquement nulle, mais nous allons voir que ce n'est pas le cas). Montrons qu'aucun de ces zéros ne peut appartenir à $\{|z| \leq 1\}$. Par l'absurde, soit z_0 un tel zéro. Notons que nécessairement $z_0 \neq 1$, car $1 + F(1) = \nu(x)$. Par conséquent, lorsque $z \rightarrow z_0$ en restant dans $\{|z| < 1\}$, $z + F(z) \rightarrow z_0 + F(z_0) = z_0 - 1 \neq 0$, et donc $|P(z)| \rightarrow +\infty$ d'après (6.4). Mais P est bornée sur $\{|z| \leq 1\}$, car ses coefficients sont des probabilités. Contradiction. Par conséquent, toujours d'après (6.4), il existe $r_1 > 1$ tel que P soit analytique sur $\{|z| < r_1\}$, ce qui prouve (i). \square

La preuve présentée ci-dessus montre en fait que, si l'on dispose d'un x pour lequel la propriété énoncée dans (iii) est vérifiée, la propriété énoncée dans (i) est vérifiée pour le même x . En utilisant le fait que (i) entraîne (v), et que (v) entraîne que la propriété énoncée dans (ii) est vérifiée pour tout x , on en déduit que (i) entraîne (ii).

On conclut ainsi la preuve des équivalences contenues dans le théorème 18.

¹Et non pas des considérations complexes d'analyse!

Exercice 171 Donner une preuve élémentaire que (i) entraîne (ii), en utilisant l'identité établie dans l'exercice 125 et la stratégie du Corollaire 8.

Nous renvoyons par exemple à [45, 36] pour plus de détails sur la notion d'ergodicité géométrique.

6.2.1 Ergodicité uniforme

La notion d'ergodicité uniforme correspond à la convergence uniforme de

$$d_{VT}(p^n(x, \cdot), \nu)$$

vers zéro par rapport à x , lorsque n tend vers l'infini. Avant d'énoncer un théorème de caractérisation de cette propriété, nous commençons par établir quelques résultats relatifs aux propriétés de contraction du noyau p vis-à-vis de la distance en variation totale. Ceci est naturel, puisque nous nous intéressons à la convergence de p vers un point fixe, à savoir la loi invariante.

Propriétés de contraction de p

Définition 14 On définit, pour tout noyau de transition p , le coefficient de Dobrushin de p par $\delta(p) = \sup_{x,y \in S} d_{VT}(p(x, \cdot), p(y, \cdot))$.

On vérifie immédiatement que $0 \leq \delta(p) \leq 1$.

Exercice 172 On note que $\delta(p)$ s'exprime de manière explicite en termes des probabilités de transition. Comment ce coefficient s'interprète-t-il en termes de couplage ?

Proposition 78 Si p et q sont deux noyaux de transition, on a $\delta(pq) \leq \delta(p)\delta(q)$.

Preuve :

Facile avec l'interprétation par couplage de la distance en variation totale (voir par exemple [1]). On peut également donner une preuve élémentaire (voir par exemple [7]).

□

Exercice 173 Donner les détails de la preuve de la proposition.

Proposition 79 Si μ_1 et μ_2 sont deux mesures de probabilité sur S , $d_{VT}(\mu p, \nu p) \leq \delta(p)d_{VT}(\mu, \nu)$.

Preuve :

Idem.

□

Exercice 174 Donner les détails de la preuve de la proposition.

Corollaire 25 Le coefficient de Dobrushin est le coefficient de contraction de p vu comme une application de l'ensemble des mesures de probabilité sur S dans lui même, muni de la distance en variation totale. Plus explicitement :

$$\delta(p) := \sup_{\mu_1 \neq \mu_2} \frac{d_{VT}(\mu_1 p, \mu_2 p)}{d_{VT}(\mu_1, \mu_2)}.$$

Exercice 175 Donner la preuve du corollaire ci-dessus.

Rappelons les notations :

$$d_1(n) := \sup_{x \in S} d_{VT}(p^n(x, \cdot), \nu),$$

et

$$\bar{d}_1(n) := \sup_{x, y \in S} d_{VT}(p^n(x, \cdot), p^n(y, \cdot)) = \delta(p^n).$$

On prouve facilement à partir de ce qui précède la proposition suivante.

Proposition 80 Avec les notations et hypothèses précédentes, on a :

1. pour tout n , $d_1(n) \leq \bar{d}_1(n) \leq 2d_1(n)$;
2. pour tous $n, m \geq 0$, $\bar{d}_1(n+m) \leq \bar{d}_1(n)\bar{d}_1(m)$;
3. pour tous $n, m \geq 0$, $d_1(n+m) \leq d_1(n)\bar{d}_1(m)$;
4. pour tous $n, m \geq 0$, $d_1(n+m) \leq 2d_1(n)d_1(m)$;
5. les suites $(\bar{d}_1(n))_{n \geq 0}$ et $(d_1(n))_{n \geq 0}$ sont décroissantes.

Exercice 176 Prouver la proposition ci-dessus.

Caractérisations de l'ergodicité uniforme

Théorème 19 Etant donné un noyau de transition ergodique p sur un ensemble S fini ou dénombrable, de loi invariante ν , il y a équivalence entre les propriétés suivantes :

- (i) Il existe $x \in S$ tel que $\sup_{y \in S} \mathbb{E}_y(T_1(x)) < +\infty$;
- (ii) Il existe $x \in S$ et $\theta > 1$ tels que $\sup_{y \in S} \mathbb{E}_y(\theta^{T_1(x)}) < +\infty$;
- (iii) On a

$$\lim_{n \rightarrow +\infty} \sup_{x \in S} d_{VT}(p^n(x, \cdot), \nu) = 0;$$

- (iv) Il existe $r < 1$ et $A \geq 0$ tel que, pour tout $n \geq 0$,

$$\sup_{x \in S} d_{VT}(p^n(x, \cdot), \nu) \leq Ar^n;$$

(v) On a

$$\lim_{n \rightarrow +\infty} \delta(p^n) = 0;$$

(vi) Il existe $r < 1$ et $A \geq 0$ tel que, pour tout $n \geq 0$,

$$\delta(p^n) \leq Ar^n;$$

(vii) Il existe $n \geq 1$ pour lequel $\sup_{x \in S} d_{VT}(p^n(x, \cdot), \nu) < 1/2$;

(viii) Il existe $n \geq 1$ pour lequel $\delta(p^n) < 1$;

(ix) Il existe $n \geq 1$ tel que p^n vérifie la condition de Doeblin, i.e. il existe une mesure de probabilité π sur S et $\epsilon > 0$ telle que $p^n(x) \geq \epsilon\pi(x)$ pour tout $x \in S$.

L'implication (i) \Rightarrow (v) se prouve en reprenant la stratégie de preuve par couplage employé dans la preuve de l'implication (i)-(ii)-(iii)-(iv) \Rightarrow (vii) du Théorème 17. On obtient ainsi une borne uniforme sur l'espérance du temps de couplage uniforme vis-à-vis du point de départ $(y_1, y_2) \in S^2$. La propriété de sous-multiplicativité énoncée dans la proposition 80 montre que (v) et (vi) sont équivalentes. En utilisant l'ensemble des propriétés énoncées dans la proposition, on déduit l'équivalence entre (iii), (iv), (v), (vi), (vii), et (viii). Inversement, (iii) permet de voir qu'il existe un x et un n tels que $\inf_{z \in S} \mathbb{P}_z(X_n = x) > 0$. On en déduit facilement le fait qu'il existe une borne sous-géométrique uniforme en le point de départ pour la queue de $T_1(x)$, ce qui n'est autre que (ii). On a par ailleurs évidemment que (ii) entraîne (i). On obtient ainsi l'équivalence entre les propriétés (i) à (viii).

Il reste donc à traiter le cas de (ix). Le fait que (ix) implique (viii) est une conséquence de l'interprétation que l'on peut donner de (ix) en termes de couplage. Sous (ix), on peut fabriquer une famille de variables $(Z_x)_{x \in S}$ et une variable Y à valeurs dans S , de telle sorte que, pour tout $x \in S$, Z_x suit la loi $p^n(x, \cdot)$, tandis que Y est de loi π , et telles que la probabilité que l'on ait $Z_x = Y$ pour tout $x \in S$ soit supérieure ou égale à ϵ . Inversement, (iii) implique facilement (ix) : il suffit de choisir $x \in S$, puis n tel que

$$\sup_{y \in S} d_{VT}(p^n(y, \cdot), \nu) \leq \nu(x)/2,$$

d'où l'on tire que $p^n(y, x) \geq \nu(x)/2$ pour tout $y \in S$.

Exercice 177 *Faire en détail les preuves esquissées ci-dessus.*

Exercice 178 *Prouver (sans faire appel au théorème ci-dessus) que, pour une chaîne de Markov ergodique sur un ensemble S fini, il existe n tel que $\delta(p^n) < 1$. Donnez un exemple où $\delta(p) = 1$ (et donc $n = 2$ au minimum). Donner également un exemple de chaîne ergodique avec S infini et pour laquelle $\delta(p^n) = 1$ pour tout t .*

On retrouve en particulier, grâce à cet exercice le résultat qualitatif déjà obtenu en étudiant la queue du temps de couplage T introduit dans la partie précédente : dans le cas d'un ensemble fini, la convergence en loi d'une chaîne ergodique vers sa loi stationnaire a toujours lieu à vitesse exponentielle.

Exercice 179 *On considère un noyau ergodique p .*

1) *Montrer l'existence d'un $r \in [0, +\infty]$ tel que*

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \log \bar{d}_1(n) = -r.$$

(Indication : utiliser le lemme classique sur les suites sous-additives.)

2) *Montrer que l'on a également*

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \log d_1(n) = -r.$$

3) *Donner un exemple pour lequel il existe $t \geq 0$ tel que $\bar{d}_1(t) = 0$ (et donc en particulier $r = +\infty$). (Indication : faire une tentative avec un espace à deux états.)*

Ainsi, on peut par exemple, dans le cas où le noyau est uniformément ergodique, (par exemple dans le cas ergodique avec S fini), utiliser $\tau_{VT} = \inf\{t; \bar{d}_1(t) \leq e^{-1}\}$ pour estimer la vitesse de convergence en variation totale, et définir un temps de relaxation en variation totale permettant des comparaisons entre chaînes différentes, et possédant une interprétation immédiate en termes de contrôle de la vitesse de convergence.

Chapitre 7

Fonctionnelles additives : théorème de la limite centrale

La situation est un peu plus complexe pour le théorème de la limite centrale que pour la loi des grands nombres, et nous présenterons trois approches, basées respectivement sur le renouvellement, les martingales et l'équation de Poisson, et la méthode de Nagaev en exploitant les propriétés spectrales. Nous renvoyons par exemple à [9, 36, 45] pour plus d'informations.

Nous reprenons dans cette section les notations définies dans la partie sur la loi des grands nombres. Donnons-nous donc $f : S \rightarrow \mathbb{R}$, un noyau de transition irréductible et positivement récurrent p de loi invariante ν , et une loi initiale μ quelconque. Nous nous restreindrons essentiellement à des fonctions f vérifiant

$$\nu(f) = 0, \nu(f^2) < +\infty, \quad (7.1)$$

et notre objectif sera donc de prouver des théorèmes limites affirmant que, sous \mathbb{P}_μ , on a la convergence en loi suivante lorsque n tend vers l'infini :

$$n^{-1/2} \sum_{i=0}^n f(X_i) \xrightarrow{\text{loi}} \mathcal{N}(0, v), \quad (7.2)$$

où $v \geq 0$ (avec la convention $\mathcal{N}(0, v) = \delta_0$).

Notons qu'une condition telle que (7.1) n'est pas nécessaire pour que la propriété ci-dessus ait lieu. Un exemple très simple (quoique dans un cas dégénéré) est le suivant : partant d'une suite de v.a. i.i.d. à valeurs réelles $(Y_n)_{n \geq 0}$, on fabrique la chaîne $(X_n)_{n \geq 0} = (Y_n, Y_{n+1})_{n \geq 0}$, dont on vérifie facilement qu'elle est ergodique. On pose ensuite $f(x, y) := y - x$, et l'on voit que l'on a toujours un théorème de la limite centrale (dégénéré, avec $v = 0$) pour f appliquée à la chaîne $(X_n)_{n \geq 0}$, car $n^{-1/2} S_n(f)$ tend vers zéro en probabilité. On peut pourtant facilement choisir la loi de (Y_0) de telle manière que f ne soit pas intégrable (et l'on n'a donc alors même

pas $f \in L^1(\nu)$!). Nous verrons également qu'une condition telle que (7.1) n'est pas non plus suffisante en général.

Commençons par un résultat préliminaire montrant que l'on peut, pour prouver un théorème de la limite centrale, choisir la loi initiale à notre guise.

Proposition 81 *Si une limite telle que (7.2) a lieu pour une loi initiale μ donnée, la même limite a lieu pour toute loi initiale.*

Preuve :

Pour le voir lorsqu'il y a apériodicité, il suffit par exemple de considérer un couplage entre deux versions de la chaîne tel que celui défini dans la preuve par couplage du Théorème 11, et de constater que la différence entre les deux sommes renormalisées que l'on considère tend vers zéro avec probabilité 1. Lorsqu'il y a périodicité, on décompose en classes et l'on se ramène à l'argument précédent en considérant un itéré apériodique du noyau de départ. \square

7.1 Approche par renouvellement

Théorème 20 *Supposons qu'il existe $a \in S$ tel que*

$$\mathbb{E}_a \left(\sum_{i=0}^{T_1(a)-1} f(X_i) \right) = 0, \quad \mathbb{E}_a \left(\sum_{i=0}^{T_1(a)-1} f(X_i) \right)^2 < +\infty. \quad (7.3)$$

Alors, lorsque n tend vers l'infini, on a, pour toute loi initiale μ , la convergence en loi

$$n^{-1/2} \sum_{i=0}^n f(X_i) \xrightarrow{\text{loi}} \mathcal{N}(0, v),$$

où

$$v = \frac{\mathbb{E}_a \left(\sum_{i=0}^{T_1(a)-1} f(X_i) \right)^2}{\mathbb{E}_a(T_1(a))}.$$

On notera que, sous réserve que f soit intégrable par rapport à ν , la première hypothèse du théorème signifie que $\nu(f) = 0$ (en se rappelant que $\nu = \frac{\mu_a^{\text{cycle}}}{\mathbb{E}_a(T_1(a))}$). De fait, si f est intégrable par rapport à ν , la loi des grands nombres pour les fonctionnelles additives montre que l'on doit avoir $\nu(f) = 0$ pour espérer obtenir un TCL.

Notons que la condition $\nu(f^2) < +\infty$ se traduit par

$$\mathbb{E}_a \left(\sum_{i=0}^{T_1(a)-1} f^2(X_i) \right) < +\infty,$$

ce qui n'est pas du tout la même chose que la deuxième hypothèse du théorème.

Remarque 26 *L'exemple donné précédemment montre qu'il est tout-à-fait possible que la variance asymptotique soit égale à 0 sans que f soit une fonction constante. Ceci étant, une variance nulle signifie que $\sum_{i=0}^{T_1(a)-1} f(X_i) = 0$ avec probabilité 1 sous \mathbb{P}_a , ce qui restreint tout de même la classe des fonctions f pouvant donner lieu à une variance nulle ! Par ailleurs, l'exemple très simple donné précédemment montre que la variance de la gaussienne obtenue à la limite dans le TCL n'est pas nécessairement égale à $\nu(f^2)$, comme une analogie naïve avec la loi des grands nombres pourrait le faire croire : la dépendance entre les termes successifs de la chaîne fait que la variance limite n'est pas en général identique à celle que l'on obtiendrait si les variables aléatoires $(f(X_i))_{i \geq 0}$ étaient i.i.d.*

Dans cette preuve, nous travaillerons sous la probabilité \mathbb{P}_a , ce qui est licite car nous avons vu que l'on pouvait choisir arbitrairement la loi initiale sous laquelle travailler pour prouver un théorème de la limite centrale dans notre contexte. Nous reprenons les notations déjà employée dans le chapitre sur la loi des grands nombres, à savoir

$$S_n(f) = \sum_{j=0}^n f(X_j),$$

et, pour tout $i \geq 0$,

$$L_i(f) = \sum_{j=T_i(a)}^{T_{i+1}(a)-1} f(X_j),$$

(avec la convention $T_0(a) := 0$), ainsi que

$$R_n = \max\{i \geq 0; T_i(a) \leq n\}.$$

On reprend la décomposition déjà utilisée pour prouver la loi des grands nombres, qui se simplifie quelque peu par le fait de travailler sous \mathbb{P}_a :

$$S_n(f) = \sum_{i=0}^{R_n-1} L_i(f) + \sum_{j=T_{R_n}(a)}^n f(X_j). \quad (7.4)$$

Nous allons d'abord montrer que le terme «de bord»

$$n^{-1/2} \sum_{j=T_{R_n}(a)}^n f(X_j)$$

tend vers zéro en probabilité lorsque n tend vers l'infini, et il nous suffira d'établir ensuite le TCL pour

$$n^{-1/2} \sum_{i=0}^{R_n-1} L_i(f).$$

En effet, rappelons que, si une suite $(Z_n)_{n \geq 0}$ de variables aléatoires réelles converge en loi vers une loi limite μ , il en va de même de toute suite de la forme

$$(Z_n + V_n)_{n \geq 0}$$

telle que V_n tende vers zéro en probabilité lorsque n tend vers l'infini. (Pour le voir, passer par exemple par la transformée de Fourier.) Nous utiliserons alors le fait suivant : si une suite de variables aléatoires $(Z_n)_{n \geq 0}$ est tendue, et si $(\epsilon_n)_{n \geq 0}$ est une suite (déterministe) tendant vers zéro, alors

$$(\epsilon_n Z_n)_{n \geq 0}$$

tend vers zéro en probabilité lorsque n tend vers l'infini. Commençons donc par prouver le résultat général suivant.

Lemme 27 *Pour toute f définie sur S , la suite de variables aléatoires*

$$\left(\sum_{j=T_{R_n}(a)}^n f(X_j) \right)_{n \geq 0}$$

est tendue.

On note que cette preuve ne fait aucune hypothèse sur f , et améliore donc celle donnée dans la preuve de la loi des grands nombres, qui utilisait le fait que $\nu(|f|) < +\infty$. En particulier, on voit que, dans la loi des grands nombres, les termes «de bord» peuvent toujours être négligés, que f soit intégrable sous ν ou non.

Preuve :

Partons de l'inégalité

$$\left| \sum_{j=T_{R_n}(a)}^n f(X_j) \right| \leq \sum_{j=T_{R_n}(a)}^n |f(X_j)|.$$

A présent,

$$\mathbb{P}_a \left(\sum_{j=T_{R_n}(a)}^n |f(X_j)| \geq t \right) = \sum_{k=0}^n \mathbb{P}_a \left(\sum_{j=k}^n |f(X_j)| \geq t, T_{R_n}(a) = k \right),$$

et donc cette probabilité est encore égale à

$$\sum_{k=0}^n \mathbb{P}_a \left(\sum_{j=k}^n |f(X_j)| \geq t, X_k = a, X_{k+1} \neq a, \dots, X_n \neq a \right).$$

En appliquant la propriété de Markov, on en déduit que

$$\mathbb{P}_a \left(\sum_{j=T_{R_n}(a)}^n |f(X_j)| \geq t \right) = \sum_{k=0}^n \mathbb{P}_a(X_k = a) \mathbb{P}_a \left(\sum_{j=0}^{n-k} |f(X_j)| \geq t, T_1(a) > n - k \right).$$

En utilisant le fait que l'événement

$$\left\{ \sum_{j=0}^{n-k} |f(X_j)| \geq t, T_1(a) > n - k \right\}$$

entraîne l'événement

$$\left\{ \sum_{j=0}^{T_1(a)} |f(X_j)| \geq t, T_1(a) > n - k \right\},$$

et en posant $m := n - k$, on en déduit que

$$\mathbb{P}_a \left(\sum_{j=T_{R_n}(a)}^n |f(X_j)| \geq t \right) \leq \sum_{m=0}^{+\infty} \mathbb{P}_a \left(\sum_{j=0}^{T_1(a)} |f(X_j)| \geq t, T_1(a) > m \right).$$

A présent, pour m fixé, on a $\lim_{t \rightarrow +\infty} \mathbb{P}_a \left(\sum_{j=0}^{T_1(a)} |f(X_j)| \geq t \right) = 0$. D'autre part, pour tout m , $\mathbb{P}_a \left(\sum_{j=0}^{T_1(a)} |f(X_j)| \geq t, T_1(a) > m \right) \leq \mathbb{P}_a(T_1(a) > m)$, et l'on a que $\sum_{m=0}^{+\infty} \mathbb{P}_a(T_1(a) > m) < +\infty$ car $\mathbb{E}_a(T_1(a)) < +\infty$ par hypothèse. Le théorème de convergence dominée (appliqué à la mesure de dénombrement sur \mathbb{N}) permet donc de conclure que

$$\lim_{t \rightarrow +\infty} \sum_{m=0}^{+\infty} \mathbb{P}_a \left(\sum_{j=0}^{T_1(a)} |f(X_j)| \geq t, T_1(a) > m \right) = 0,$$

et donc que

$$\lim_{t \rightarrow +\infty} \sup_{n \geq 0} P \left(\left| \sum_{j=T_{R_n}(a)}^n f(X_j) \right| \geq t \right) = 0.$$

Ceci entraîne la tension de la suite $\left(\sum_{j=T_{R_n}(a)}^n f(X_j) \right)_{n \geq 0}$. □

Preuve du théorème 20:

D'après le lemme ci-dessus,

$$n^{-1/2} \sum_{j=T_{R_n}(a)}^n f(X_j)$$

tend vers zéro en probabilité. Il nous reste donc à prouver le théorème de la limite centrale pour

$$n^{-1/2} \sum_{i=1}^{R_n} L_i(f).$$

Les hypothèses que nous avons formulées sur f se traduisent par

$$\mathbb{E}_a(L_1(f)) = 0, \quad \mathbb{E}_a((L_1(f))^2) < +\infty,$$

et l'on peut donc appliquer le théorème de la limite centrale pour les v.a.i.i.d. centrées et de carré intégrable à

$$k^{-1/2} \sum_{i=0}^k L_i(f)$$

lorsque k tend vers l'infini, mais le fait que R_n soit aléatoire et certainement pas indépendante des $L_i(f)$ ne nous permet pas de faire $k = R_n$ dans la limite correspondante et d'en déduire le résultat directement, comme cela était possible dans la preuve de la loi des grands nombres, dans laquelle on pouvait considérer le comportement asymptotique presque sûre des suites qui intervenaient. Nous utilisons donc l'argument un peu plus sophistiqué qui suit.

D'après la preuve par renouvellement donnée dans la partie sur la loi des grands nombres, on a la convergence suivante :

$$\lim_{n \rightarrow +\infty} n^{-1} R_n = \rho, \quad \mathbb{P}_a - p.s.,$$

où $\rho = (\mathbb{E}_a(T_1(a)))^{-1}$. Par conséquent, pour tout $0 < \alpha < 1$ et $\epsilon > 0$, nous avons que, pour tout n assez grand,

$$\mathbb{P}_a [(1 - \alpha)\rho \leq n^{-1} R_n \leq (1 + \alpha)\rho] \geq 1 - \epsilon.$$

En utilisant le fait que les $L_i(f)$ sont i.i.d., centrées, et de carré intégrable sous \mathbb{P}_a , on voit facilement que la suite $(G_k^1)_{k \geq 0}$ définie par

$$G_k^1 := \sum_{j=\lfloor \rho n \rfloor + 1}^{\lfloor \rho n \rfloor + k} L_j(f)$$

est une martingale, et que l'inégalité maximale pour les martingales (voir par exemple [49, 18], ou le memento sur les martingales donné dans ces notes) entraîne que, pour tout $t \geq 0$,

$$\mathbb{P}_a \left(\max_{k=0}^{\lceil \alpha \rho n \rceil + 1} |G_k^1| \geq tn^{1/2} \right) \leq t^{-2} n^{-1} \mathbb{E}_a \left(|G_{\lceil \alpha \rho n \rceil + 1}^1|^2 \right).$$

On voit également que, pour tout $k \geq 0$,

$$\mathbb{E} \left(|G_k^1|^2 \right) = k E(|L_1(f)|^2),$$

d'où finalement le fait que

$$\mathbb{P}_a \left(\max_{k=0}^{\lceil \alpha \rho n \rceil + 1} |G_k^1| \geq tn^{1/2} \right) \leq t^{-2} n^{-1} (\lceil \alpha \rho n \rceil + 1) E(|L_1(f)|^2).$$

En définissant $G_k^2 := \sum_{j=\lfloor \rho n \rfloor - k}^{\lfloor \rho n \rfloor} L_j(f)$, un argument similaire montre que

$$\mathbb{P}_a \left(\max_{k=0}^{\lceil \alpha \rho n \rceil + 1} |G_k^2| \geq tn^{1/2} \right) \leq t^{-2} n^{-1} (\lceil \alpha \rho n \rceil + 1) E(|L_1(f)|^2).$$

Posons alors

$$d_n := n^{-1/2} \left(\sum_{j=0}^{\lfloor \rho n \rfloor} L_j(f) - \sum_{j=0}^{R_n} L_j(f) \right),$$

et notons que, sur l'événement $\{(1 - \alpha)\rho \leq n^{-1}R_n \leq (1 + \alpha)\rho\}$,

$$|d_n| \leq \max_{k=0}^{\lceil \alpha \rho n \rceil + 1} |G_k^1| + \max_{k=0}^{\lceil \alpha \rho n \rceil + 1} |G_k^2|.$$

On obtient maintenant facilement le fait que d_n tend vers zéro en probabilité lorsque n tend vers l'infini. On conclut en appliquant le théorème de la limite centrale pour les suites i.i.d. centrées de carré intégrable à $\lfloor \rho n \rfloor^{-1/2} \sum_{i=0}^{\lfloor \rho n \rfloor} L_i(f)$. \square

De manière remarquable, il est possible de prouver une réciproque au théorème que nous venons de prouver.

Théorème 21 *S'il existe une loi initiale pour laquelle la suite*

$$(n^{-1/2} S_n(f))_{n \geq 1}$$

est tendue, alors nécessairement, pour tout $a \in S$,

$$\mathbb{E}_a \left(\sum_{i=0}^{T_1(a)-1} f(X_i) \right) = 0, \quad \mathbb{E}_a \left(\sum_{i=0}^{T_1(a)-1} f(X_i) \right)^2 < +\infty.$$

On note que le résultat est en fait plus fort qu'une simple réciproque, puisqu'il ne suppose que la tension pour en déduire les hypothèses du théorème 20. Par conséquent, au vu des deux théorèmes 20 et 21, la tension de la suite

$$(n^{-1/2} S_n(f))_{n \geq 1}$$

entraîne automatiquement la validité du théorème de la limite centrale. En particulier, on ne peut avoir de loi limite autre que gaussienne dans notre contexte. Qui plus est, le théorème de la limite centrale peut toujours être obtenu grâce à l'approche par renouvellement. On obtient également que la validité de l'hypothèse 7.3 pour un

$a \in S$ entraîne sa validité pour tout $a \in S$, ce qui n'est pas évident a priori. Une preuve directe de ce dernier fait peut être trouvée dans [10].

Le plan de la preuve du théorème 21 est le suivant : on prouve d'abord que la suite

$$\left(n^{-1/2} \sum_{i=0}^n L_i(f) \right)_{n \geq 1}$$

est tendue sous \mathbb{P}_a , puis on utilise l'argument général suivant :

Théorème 22 *Soit $(Z_n)_{n \geq 0}$ une suite de variables aléatoires i.i.d. Alors la suite*

$$\left(n^{-1/2} (Z_0 + \dots + Z_n) \right)_{n \geq 1}$$

est tendue si et seulement si $E(Z_0^2) < +\infty$ et $E(Z_0) = 0$.

Preuve du théorème 22:

Le fait que les hypothèses $E(Z_0^2) < +\infty$ et $E(Z_0) = 0$ entraînent la tension de la suite

$$\left(n^{-1/2} (Z_0 + \dots + Z_n) \right)_{n \geq 1}$$

est une conséquence immédiate du théorème de la limite centrale pour les variables aléatoires i.i.d. centrées de carré intégrable.

Pour la réciproque, considérons une suite de variables aléatoires $(Z'_n)_{n \geq 0}$ indépendante de $(Z_n)_{n \geq 0}$, et de même loi. Posons, pour tout $i \geq 0$,

$$W_i = Z_i - Z'_i,$$

et, pour un paramètre $K \geq 0$ fixé, posons, pour tout $i \geq 0$,

$$Y_i = W_i \mathbf{1}(|W_i| \leq K).$$

On note que la suite $(Y_i)_{i \geq 0}$ est une suite de variables aléatoires indépendantes bornées, et vérifiant donc $E(Y_i^2) < +\infty$. De plus, la définition de W_i entraîne que W_i et $-W_i$ ont la même loi, ce qui entraîne que l'on a nécessairement $E(Y_i) = 0$. On peut donc appliquer le théorème de la limite centrale pour les variables aléatoires i.i.d. centrées de carré intégrable pour déduire que, lorsque n tend vers l'infini, la loi de

$$n^{-1/2} (Y_0 + \dots + Y_n)$$

converge vers une loi gaussienne centrée et de variance égale à

$$E(W_0^2 \mathbf{1}(|W_0| \leq K)).$$

Par ailleurs, observons que, toujours en vertu de la définition symétrique de W_i , pour tout i , on a l'identité en loi suivante :

$$(Y_i, W_i \mathbf{1}(|W_i| > K)) \stackrel{\text{loi}}{=} (Y_i, -W_i \mathbf{1}(|W_i| > K)).$$

On en déduit que

$$\left(\sum_{i=0}^n Y_i, \sum_{i=0}^n W_i \mathbf{1}(|W_i| > K) \right) \stackrel{\text{loi}}{=} \left(\sum_{i=0}^n Y_i, - \sum_{i=0}^n W_i \mathbf{1}(|W_i| > K) \right).$$

Par conséquent, pour tout $t \in \mathbb{R}$, on a

$$P \left(\sum_{i=0}^n Y_i \geq t, \sum_{i=0}^n W_i \mathbf{1}(|W_i| > K) > 0 \right) = P \left(\sum_{i=0}^n Y_i \geq t, \sum_{i=0}^n W_i \mathbf{1}(|W_i| > K) < 0 \right).$$

On en déduit que

$$P \left(\sum_{i=0}^n Y_i \geq t, \sum_{i=0}^n W_i \mathbf{1}(|W_i| > K) \geq 0 \right) \geq (1/2) P \left(\sum_{i=0}^n Y_i \geq t \right).$$

En se rappelant que $Y_i + W_i \mathbf{1}(|W_i| > K) = W_i$, on en déduit finalement que

$$P \left(\sum_{i=0}^n W_i \geq t \right) \geq (1/2) P \left(\sum_{i=0}^n Y_i \geq t \right) \tag{7.5}$$

A présent, rappelons nous que, par hypothèse, les suites

$$\left(n^{-1/2}(Z_0 + \dots + Z_n) \right)_{n \geq 1}$$

et

$$\left(n^{-1/2}(Z'_0 + \dots + Z'_n) \right)_{n \geq 1}$$

sont tendues. Il en résulte facilement que la suite

$$\left(n^{-1/2}(W_0 + \dots + W_n) \right)_{n \geq 1}$$

est elle-même tendue. Il existe donc, par exemple, un $t \geq 0$ tel que

$$P \left(n^{-1/2}(W_0 + \dots + W_n) \geq t \right) \leq 1/8$$

pour tout $n \geq 1$. Supposons que $E(W_0^2) = +\infty$. On peut alors, par convergence monotone, trouver K tel que $E(W_0^2 \mathbf{1}(|W_0| \leq K))$ soit arbitrairement grand. En appliquant le théorème de la limite centrale à

$$n^{-1/2}(Y_0 + \dots + Y_n),$$

on en déduit que l'on peut faire en sorte que, pour n assez grand,

$$P \left(\sum_{i=0}^n Y_i \geq t \right) \geq 1/3$$

(il suffit de choisir K de telle sorte que la variance limite soit assez grande, et l'on peut s'approcher aussi près que l'on veut de $1/2$). Mais l'inégalité (7.5) entraîne alors que

$$P\left(\sum_{i=0}^n W_i \geq t\right) \geq (1/2) \times (1/3) > 1/8,$$

ce qui conduit à une contradiction. On en déduit que $E(W_0^2) < +\infty$. En conditionnant par un événement de la forme $a \leq Z'_0 \leq b$ de probabilité non-nulle, on en déduit que $E(Z_0^2) < +\infty$. A présent, on sait que $E(|Z_0|) < +\infty$, et l'on doit nécessairement avoir $E(Z_0) = 0$, sans quoi la loi forte des grands nombres contredit manifestement l'hypothèse sur la tension de la loi de

$$\left(n^{-1/2}(Z_0 + \dots + Z_n)\right)_{n \geq 0}.$$

□

Preuve du théorème 21:

Comme dans la preuve du théorème 20, nous fixons $a \in S$, et nous nous plaçons sous la probabilité \mathbb{P}_a . L'hypothèse du théorème nous permet de prouver la tension de $(n^{-1/2}(S_n(f)))_{n \geq 1}$ sous \mathbb{P}_a , en utilisant un argument semblable à celui qui permet de prouver la proposition 81. Nous allons en déduire la tension sous \mathbb{P}_a de la suite

$$\left(n^{-1/2}\left(\sum_{i=0}^n L_i(f)\right)\right)_{n \geq 1},$$

et le résultat s'ensuivra. Fixons donc

$$\lambda < (\mathbb{E}_a(T_1(a)))^{-1},$$

et expliquons comment établir la tension de la suite

$$\left(n^{-1/2} \sum_{i=0}^{\lceil n\lambda \rceil - 1} L_i(f)\right)_{n \geq 1} = \left(n^{-1/2} S_{T_{\lceil n\lambda \rceil}(a)-1}(f)\right)_{n \geq 1},$$

ce qui entraîne le résultat souhaité. En appliquant la borne de la réunion, on obtient que

$$\mathbb{P}_a\left(S_{T_{\lceil n\lambda \rceil}(a)-1}(f) \geq 2Mn^{1/2}\right)$$

est inférieure ou égale à

$$\mathbb{P}_a(R_n \leq \lceil \lambda n \rceil) + \mathbb{P}_a\left(S_{T_{\lceil n\lambda \rceil}(a)-1}(f) \geq 2Mn^{1/2}, R_n \geq \lceil \lambda n \rceil\right).$$

La loi des grands nombres déjà utilisée dans la preuve du théorème 20 entraîne que

$$\lim_{n \rightarrow +\infty} \mathbb{P}_a(R_n \leq \lceil \lambda n \rceil) = 0.$$

D'autre part,

$$\mathbb{P}_a \left(S_{T_{\lceil n\lambda \rceil}(a)-1}(f) \geq 2Mn^{1/2}, R_n \geq \lceil n\lambda \rceil \right) \leq \mathbb{P}_a \left(\max_{1 \leq k \leq R_n} |S_{T_k(a)-1}(f)| \geq 2Mn^{1/2} \right).$$

Définissons

$$\tau := \inf \left\{ T_k(a); k \geq 1, |S_{T_k(a)-1}(f)| \geq 2Mn^{1/2} \right\}$$

(avec la convention $\inf \emptyset = +\infty$). On constate que τ est un temps d'arrêt de la chaîne $(X_i)_{i \geq 0}$ tel que $X_\tau = a$ si $\tau < +\infty$, et que

$$\{\tau \leq n\} = \left\{ \max_{1 \leq k \leq R_n} |S_{T_k(a)-1}(f)| \geq 2Mn^{1/2} \right\}.$$

Par ailleurs, notons que l'on a l'inclusion

$$\{\tau \leq n\} \cap \left\{ \left| \sum_{i=\tau}^n f(X_i) \right| \leq Mn^{1/2} \right\} \subset \left\{ |S_n(f)| \geq Mn^{1/2} \right\}.$$

A présent, notons que, grâce à la propriété forte de Markov, on a, pour tout $1 \leq u \leq n$, sur l'événement $\{\tau = u\}$, l'égalité \mathbb{P}_a -p.s. suivante

$$\mathbb{P}_a \left(\left| \sum_{i=\tau}^n f(X_i) \right| \leq Mn^{1/2} \middle| \mathcal{F}_\tau \right) = \mathbb{P}_{X_u} \left(\left| \sum_{i=0}^{n-u} f(X_i) \right| \leq Mn^{1/2} \right).$$

En utilisant le fait que $X_\tau = a$ si $\tau < +\infty$, on en déduit que le membre de droite de l'identité précédente est supérieur ou égal à

$$c := \min_{0 \leq m \leq n-1} \mathbb{P}_a \left(|S_m(f)| \leq Mn^{1/2} \right).$$

On déduit de ce qui précède que

$$\mathbb{P}_a \left(\max_{1 \leq k \leq R_n} |S_{T_k(a)-1}(f)| \geq 2Mn^{1/2} \right) \times c \leq \mathbb{P}_a \left(|S_n(f)| \geq Mn^{1/2} \right).$$

La tension de la suite

$$\left(n^{-1/2}(S_n(f)) \right)_{n \geq 1}$$

nous montre que, pour tout $M \geq 0$ assez grand, on a, pour tout $n \geq 0$,

$$\mathbb{P}_a(|S_n(f)| \geq Mn^{1/2}) \leq 1/2.$$

Pour un tel M , on a que $c \geq 1/2$, et, par conséquent,

$$\mathbb{P}_a \left(\max_{1 \leq k \leq R_n} |S_{T_k(a)-1}(f)| \geq 2Mn^{1/2} \right) \leq 2\mathbb{P}_a \left(|S_n(f)| \geq Mn^{1/2} \right).$$

Nous déduisons ainsi le fait que la suite

$$\left(n^{-1/2} S_{T_{\lceil n\lambda \rceil}(a)-1}(f) \right)_{n \geq 1}$$

est tendue. \square

Donnons à présent quelques exemples de situations dans lesquelles il est possible de vérifier les hypothèses du TCL tel que nous l'avons énoncé précédemment.

Proposition 82 *Si p est ergodique de degré 2 et si f est bornée et satisfait $\nu(f) = 0$, alors les hypothèses du théorème 20 sont satisfaites.*

Preuve :

Il est immédiat que $\left| \sum_{j=0}^{T_1(a)-1} |f(X_j)| \right| \leq T_1(a) \sup |f|$. Ensuite, l'ergodicité de degré 2 entraîne le fait que $\mathbb{E}_a(T_1(a)^2) < +\infty$. On en déduit que

$$\mathbb{E}_a \left(\sum_{j=0}^{T_1(a)-1} |f(X_j)| \right)^2 < +\infty.$$

Il en résulte en particulier que

$$\mathbb{E}_a \left(\sum_{j=0}^{T_1(a)-1} f(X_j) \right)^2 < +\infty$$

et, en se rappelant que $\nu = \frac{\mu_a^{cycle}}{\mathbb{E}_a(T_1(a))}$, que

$$\mathbb{E}_a \left(\sum_{j=0}^{T_1(a)-1} f(X_j) \right) = \nu(f) = 0.$$

\square

Il se trouve que la proposition ci-dessus admet une réciproque (voir [9] pour une preuve) : si le TCL a lieu pour toutes les fonctions bornées, alors on doit avoir ergodicité de degré 2.

Proposition 83 *Si $f \in L^{2+\epsilon}(\nu)$ pour un $\epsilon > 0$ donné, satisfait $\nu(f) = 0$, et si p est géométriquement ergodique, alors les hypothèses du théorème 20 sont satisfaites.*

Preuve :

On a

$$\left| \sum_{j=0}^{T_1(a)-1} f(X_j) \right| \leq \sum_{j=0}^{T_1(a)-1} |f(X_j)|.$$

En développant le carré, et en échangeant sommation et espérance, on obtient que

$$\left(\sum_{j=0}^{T_1(a)-1} |f(X_j)| \right)^2 = \sum_{i,j \geq 0} \mathbb{E}_a(|f|(X_i)\mathbf{1}(T > i)|f|(X_j)\mathbf{1}(T > j)).$$

Posons

$$u_i := \mathbb{E}_a(|f|^{2+\epsilon}(X_i)\mathbf{1}(T > i))$$

En choisissant $s > 1$ tel que $1/(2 + \epsilon) + 1/(2 + \epsilon) + 1/s + 1/s = 1$, l'inégalité de Hölder entraîne que, pour tous $i, j \geq 0$,

$$\mathbb{E}_a(|f|(X_i)\mathbf{1}(T > i)|f|(X_j)\mathbf{1}(T > j)) \leq u_i^{1/(2+\epsilon)} u_j^{1/(2+\epsilon)} \mathbb{P}_a(T > i)^{1/s} \mathbb{P}_a(T > j)^{1/s}.$$

On en déduit que

$$\left(\sum_{j=0}^{T_1(a)-1} |f(X_j)| \right)^2 \leq \left(\sum_{i=0}^{+\infty} u_i^{1/(2+\epsilon)} \mathbb{P}_a(T > i)^{1/s} \right)^2.$$

Or, d'après l'hypothèse que $f \in L^{2+\epsilon}(\nu)$, et en se rappelant que $\nu = \frac{\mu_a^{cycle}}{\mathbb{E}_a(T_1(a))}$, on voit que

$$\sum_{i=0}^{+\infty} \mathbb{E}_a(|f|^{2+\epsilon}(X_i)\mathbf{1}(T > i)) = \sum_{i=0}^{+\infty} c_i < +\infty.$$

Par ailleurs, l'ergodicité géométrique entraîne que $\mathbb{P}_a(T > i)$ tend exponentiellement rapidement vers zéro lorsque $i \rightarrow +\infty$. Toujours d'après l'inégalité de Hölder,

$$\sum_{i=0}^{+\infty} c_i^{1/(2+\epsilon)} \mathbb{P}_a(T > i)^{1/s} \leq \left(\sum_{i=0}^{+\infty} c_i \right)^{1/(2+\epsilon)} \left(\sum_{i=0}^{+\infty} \mathbb{P}_a(T > i)^{u/s} \right)^{1/u},$$

où u est défini par $1/(2 + \epsilon) + 1/u = 1$. On conclut ainsi que

$$\mathbb{E}_a \left(\sum_{j=0}^{T_1(a)-1} |f(X_j)| \right)^2 < +\infty.$$

La preuve s'achève comme celle de la proposition précédente. \square

On peut construire des contre-exemples à cette proposition dans le cas où l'on suppose seulement le fait que $f \in L^2(\nu)$ (voir par exemple [24, 6]). En revanche on a toujours un TCL lorsque $f \in L^2(\nu)$ si l'on suppose en plus que p^k est uniformément ergodique, ou la réversibilité (voir [45]).

7.2 Approche par les martingales et l'équation de Poisson

Le résultat central de cette section est le suivant.

Théorème 23 *Etant donnée $f \in L^2(\nu)$, s'il existe une fonction $g \in L^2(\nu)$ solution de l'équation de Poisson*

$$g - pg = f,$$

alors, lorsque n tend vers l'infini, on a, pour toute loi initiale μ , la convergence en loi

$$n^{-1/2} \sum_{i=0}^n f(X_i) \xrightarrow{\text{loi}} \mathcal{N}(0, v),$$

où

$$v = \nu(g^2) - \nu((pg)^2) = \mathbb{E}_\nu [(g(X_1) - pg(X_0))^2] < +\infty.$$

La preuve du théorème s'appuie sur le théorème de la limite centrale pour les martingales, dont nous donnons ici la version qui nous sera utile (voir par exemple [18, 49]) pour une preuve et des énoncés plus généraux.

Théorème 24 *Soit $(M_n)_{n \geq 0}$ une martingale par rapport à une filtration $(\mathcal{F}_n)_{n \geq 0}$ vérifiant les conditions suivantes :*

- $M_0 = 0$;
- pour tout $n \geq 0$, $E(M_n^2) < +\infty$;
- on a la convergence en probabilité suivante :

$$\lim_{n \rightarrow +\infty} n^{-1} \sum_{k=1}^n E((M_k - M_{k-1})^2 | \mathcal{F}_{k-1}) = v \in \mathbb{R}_+;$$

- pour tout $\epsilon > 0$, on a la convergence en probabilité suivante :

$$\lim_{n \rightarrow +\infty} n^{-1} \sum_{k=1}^n E\left((M_k - M_{k-1})^2 \mathbf{1}(|M_k - M_{k-1}| \geq \epsilon n^{1/2}) | \mathcal{F}_{k-1}\right) = 0.$$

Alors, lorsque n tend vers l'infini, on a, pour toute loi initiale μ , la convergence en loi

$$n^{-1/2} M_n \xrightarrow{\text{loi}} \mathcal{N}(0, v).$$

Nous ne donnerons pas de preuve de ce résultat, mais au moins une petite idée (très grossière) de preuve, afin que le résultat n'apparaisse pas comme trop mystérieux.

Preuve (esquisse):

Pour $\lambda \in \mathbb{R}$, on introduit

$$C_n := \exp(i\lambda n^{-1/2} M_n),$$

$$G_n(\lambda) := \prod_{1 \leq k \leq n} E\left(\exp(i\lambda n^{-1/2}(M_k - M_{k-1})) | \mathcal{F}_{k-1}\right),$$

et

$$\phi(\lambda) = \exp(-\lambda^2 v^2 / 2).$$

Pour prouver le théorème, on cherche à prouver que, pour tout λ ,

$$\lim_{n \rightarrow +\infty} E(C_n(\lambda)) = \phi(\lambda).$$

Supposons que nous ayons prouvé que, au sens de la convergence en probabilité,

$$\lim_{n \rightarrow +\infty} G_n(\lambda) = \phi(\lambda).$$

Compte-tenu du fait que, grâce à la définition,

$$E(C_n(\lambda)/G_n(\lambda)) = 1,$$

on peut écrire que

$$|E(C_n(\lambda) - \phi(\lambda))| = |E(C_n(\lambda)/G_n(\lambda) (G_n(\lambda) - \phi(\lambda)))|.$$

En utilisant que $G_n(\lambda)$ tend en probabilité vers $\phi(\lambda) \neq 0$ lorsque n tend vers l'infini, on en déduit que

$$\lim_{n \rightarrow +\infty} E(C_n(\lambda)) = \phi(\lambda).$$

A présent, en effectuant un développement à l'ordre 2 dans l'exponentielle, on voit que, lorsque $n \rightarrow +\infty$,

$$E\left(\exp(i\lambda n^{-1/2}(M_k - M_{k-1})) | \mathcal{F}_{k-1}\right) = 1 - (\lambda^2/2)E((M_k - M_{k-1})^2 | \mathcal{F}_{k-1}) + o(\lambda^2),$$

où l'absence de terme d'ordre 1 en λ s'explique par le fait que, par la propriété de martingale, $E(M_k - M_{k-1} | \mathcal{F}_{k-1}) = 0$. En négligeant les termes d'erreurs, on en conclut que

$$G_n(\lambda) \approx \prod_{1 \leq k \leq n} (1 - (\lambda^2/2)E((M_k - M_{k-1})^2 | \mathcal{F}_{k-1})),$$

puis que

$$G_n(\lambda) \approx \prod_{1 \leq k \leq n} \exp\left(-(\lambda^2/2) \sum_{k=1}^n E((M_k - M_{k-1})^2 | \mathcal{F}_{k-1})\right),$$

d'où l'on déduit le résultat, à condition de pouvoir convenablement contrôler les termes d'erreur que nous avons négligés. \square

Preuve du théorème 23:

On note que la finitude de v sous l'hypothèse que $g \in L^2(\nu)$, ainsi que l'identité donnant une seconde expression de v , sont immédiates.

Partant de la relation $f = g - pg$, on peut écrire que

$$f(X_0) + \cdots + f(X_n) = \sum_{i=0}^n g(X_i) - pg(X_i),$$

d'où

$$f(X_0) + \cdots + f(X_n) = g(X_0) - pg(X_n) + M_n, \quad (7.6)$$

où l'on a posé

$$M_n := \sum_{i=0}^{n-1} g(X_{i+1}) - pg(X_i).$$

On constate que $(M_n)_n$ est une martingale par rapport à la filtration naturelle de la chaîne $(\mathcal{F}_n)_{n \geq 0}$. Avant d'appliquer le théorème de la limite centrale pour les martingales à M_n , expliquons comment contrôler les deux termes de bord apparaissant dans l'équation (7.6)

On a évidemment que $n^{-1/2}g(X_0)$ tend p.s. vers zéro lorsque n tend vers l'infini. Quant à $n^{-1/2}pg(X_n)$, on lui règle son compte en utilisant par exemple le fait que, comme par hypothèse $\nu((pg)^2) < +\infty$, on a la loi des grands nombres pour

$$n^{-1}((pg)^2(X_0) + \cdots + (pg)^2(X_n))$$

d'où le fait que, presque sûrement,

$$\lim_{n \rightarrow +\infty} n^{-1}(pg)^2(X_n) = 0.$$

(On peut également par exemple partir sous la loi ν et utiliser le fait que la loi de $pg(X_n)$ ne dépend alors pas de n .) Il reste à vérifier que l'on peut appliquer le TCL pour les martingales à M_n . On vérifie que

$$\mathbb{E}_\nu((g(X_{i+1}) - pg(X_i))^2 | \mathcal{F}_i) = p(g^2)(X_i) - (pg(X_i))^2.$$

En appliquant la loi des grands nombres à la fonctionnelle additive obtenue avec la fonction h définie par

$$h(x) := p(g^2)(x) - (pg(x))^2,$$

nos hypothèses assurant que $h \in L^1(\nu)$, on obtient que, presque sûrement,

$$\lim_{n \rightarrow +\infty} n^{-1} \sum_{i=0}^{n-1} \mathbb{E}_\nu((g(X_{i+1}) - pg(X_i))^2 | \mathcal{F}_i) = \nu(h) = v.$$

Pour conclure, il reste à vérifier que

$$n^{-1} \sum_{i=0}^{n-1} \mathbb{E}_\nu((g(X_{i+1}) - pg(X_i))^2 \mathbf{1}(|g(X_{i+1}) - pg(X_i)| \geq \epsilon n^{1/2}) | \mathcal{F}_i)$$

tend presque sûrement vers zéro, pour tout $\epsilon > 0$. On réapplique pour cela la loi des grands nombres à la fonctionnelle additive obtenue en prenant avec la fonction h_c définie par

$$h_c(x) := \mathbb{E}_x((g(X_1) - pg(X_0))^2 \mathbf{1}((g(X_1) - pg(X_0))^2 \geq c)),$$

et l'on obtient que, presque sûrement,

$$\lim_{n \rightarrow +\infty} n^{-1} \sum_{i=0}^{n-1} \mathbb{E}_\nu((g(X_{i+1}) - pg(X_i))^2 \mathbf{1}((g(X_{i+1}) - pg(X_i))^2 \geq c) | \mathcal{F}_i) = \nu(h_c).$$

Comme

$$\nu(h_c) = \mathbb{E}_\nu(((g(X_1) - pg(X_0))^2 \mathbf{1}((g(X_1) - pg(X_0))^2 \geq c))),$$

le théorème de convergence dominée entraîne facilement que

$$\lim_{c \rightarrow +\infty} \nu(h_c) = 0,$$

ce qui permet de conclure. □

Etant donnée $f \in L^2(\nu)$ telle que f puisse se mettre sous la forme $g - pg$, où $g \in L^2(\nu)$, on voit que f doit nécessairement vérifier $\nu(f) = 0$. Une bonne question est donc, étant donnée $f \in L^2(\nu)$ telle que $\nu(f) = 0$, y a-t-il existence (et, pourquoi pas, unicité) de solutions de l'équation de Poisson associée à f ? Notons que nous ne sommes pas dans le cadre de l'équation de Poisson avec frontière absorbante étudié dans un chapitre antérieur.

Concernant l'unicité, nous avons le résultat suivant :

Proposition 84 *Si g_1 et g_2 sont deux solutions de l'équation de Poisson dans $L^2(\nu)$, $g_1 - g_2$ est constante.*

Preuve :

On a alors $p(g_1 - g_2) = g_1 - g_2$. En itérant, on en déduit que $p^n(g_1 - g_2) = g_1 - g_2$. La loi des grands nombres appliquée à $g_1 - g_2$ entraîne alors que $g_1 - g_2 = \nu(g_1 - g_2)$. □

Concernant l'existence, formellement, il semble naturel de tenter d'inverser l'opérateur $Id - p$ en posant

$$g = \sum_{k=0}^{+\infty} p^k f.$$

Lorsque la série ci-dessus converge au sens de $L^2(\nu)$, on obtient effectivement une solution de l'équation étudiée.

On obtient ainsi la proposition ci-dessous.

Proposition 85 *Etant donnée $f \in L^2(\nu)$, si $\sum_{k=0}^{+\infty} p^k f$ converge dans $L^2(\nu)$, alors en notant g la somme de la série, on a $f = g - pg$.*

Exercice 180 *Vérifier la proposition ci-dessus.*

Dans le cas d'une chaîne ergodique, on a $\lim_{k \rightarrow +\infty} p^k f = 0$ dans $L^2(\nu)$, et il apparaît donc que la question de la convergence de la série ci-dessus est liée à la vitesse de convergence de la chaîne vers sa loi invariante.

Notons que, dans le cas général d'une fonction f de $L^2(\nu)$ qui ne vérifie pas $\nu(f) = 0$, on doit chercher à appliquer le résultat ci-dessus à la fonction $f - \nu(f)$. Dans ce cadre, on voit que la fonction g obtenue dans la proposition ci-dessus est (formellement) obtenue à partir de l'action sur f de la fonction de Green

$$G(x, y) = \sum_{n=0}^{+\infty} (p^n(x, y) - \nu(y)).$$

Celle-ci diffère de celles étudiées dans le chapitre d'introduction à la théorie du potentiel dans le cas sans frontière, par la soustraction de ν , et l'objet défini (formellement) ci-dessus est appelé **potentiel récurrent**. Il permet, à condition que la convergence de la chaîne vers sa loi stationnaire soit suffisamment rapide, d'obtenir un objet pouvant jouer le rôle d'une fonction de Green, alors même que la série $\sum_{n=0}^{+\infty} p^n(x, y)$ possède toujours une valeur infinie. Nous avons en fait déjà rencontré cet objet dans l'exercice 167.

Mentionnons que l'approche du théorème de la limite centrale basée sur l'équation de Poisson a donné lieu à l'important résultat de Kipnis et Varadhan (voir [29]), qui, dans le cas réversible, fournit un théorème de la limite centrale pour des chaînes de Markov ergodiques à valeurs dans des espaces d'états généraux. La preuve de ce résultat repose également sur l'utilisation de la théorie spectrale de l'opérateur auto-adjoint associé à l'action du noyau sur $L^2(\nu)$.

Notons que cette approche permet de traiter les deux exemples vus dans la partie précédente.

Proposition 86 *Si p est ergodique de degré 2 et si f est bornée et satisfait $\nu(f) = 0$, alors les hypothèses de la proposition 85, et donc du théorème 23, sont satisfaites.*

Exercice 181 *Prouver la proposition ci-dessus.*

Proposition 87 Si $f \in L^{2+\epsilon}(\nu)$ pour un $\epsilon > 0$ donné, satisfait $\nu(f) = 0$, et si p est géométriquement ergodique, alors les hypothèses de la proposition 85, et donc du théorème 23, sont satisfaites.

Preuve :

Grâce au résultat de l'exercice 155 appliqué avec $\mu_1 = p^k(x, \cdot)$ et $\mu_2 = \nu$. on obtient que, pour tout $x \in S$, $|p^k f(x)| \leq [(p^k |f|^2(x))^{1/2} + \|f\|_{L^2(\nu)}] [d_{VT}(\delta_x p^k, \nu)]^{1/2}$. On en déduit que

$$\sum_{x \in S} |p^k f(x)|^2 \nu(x) \leq \sum_{x \in S} 2((p^k |f|^2(x)) + \|f\|_{L^2(\nu)}^2) d_{VT}(\delta_x p^k, \nu) \nu(x).$$

D'après nos hypothèses, il est clair que $\sum_{k \geq 0} \sum_{x \in S} \|f\|_{L^2(\nu)}^2 d_{VT}(\delta_x p^k, \nu) \nu(x) < +\infty$. D'autre part, en appliquant l'inégalité de Hölder, on en déduit que

$$\sum_{x \in S} (p^k |f|^2(x)) d_{VT}(\delta_x p^k, \nu) \nu(x)$$

est inférieur ou égal au produit

$$\left[\sum_{x \in S} (p^k |f|^2(x))^{1+\epsilon/2} \nu(x) \right]^{2/(2+\epsilon)} \times \left[\sum_{x \in S} d_{VT}(\delta_x p^k, \nu)^{1+2/\epsilon} \nu(x) \right]^{1/(1+2/\epsilon)}.$$

Ensuite, en utilisant l'inégalité $(p^k |f|^2(x))^{1+\epsilon/2} \leq p^k |f|^{2+\epsilon}(x)$, on peut conclure que la série $\sum_n p^n f$ converge normalement dans $L^2(\nu)$. \square

7.3 Calculs asymptotiques de variance

Les deux approches présentées ci-dessus conduisent à deux expressions a priori différentes de la variance limite dans le TCL. Etant donnée $f \in L^2(\nu)$ telle que $\nu(f) = 0$, dans l'approche par renouvellement, la variance limite est donnée par

$$v = \frac{\mathbb{E}_a \left(\sum_{i=0}^{T_1(a)-1} f(X_i) \right)^2}{\mathbb{E}_a(T_1(a))},$$

tandis que, dans l'approche par l'équation de Poisson, on a

$$v = \nu(g^2) - \nu((pg)^2).$$

Dans ce dernier cas, on peut montrer le résultat suivant.

Proposition 88 *Etant donnée $f \in L^2(\nu)$, si $\sum_{k=0}^{+\infty} p^k f$ converge dans $L^2(\nu)$, alors en notant g la somme de la série, on a*

$$\nu(g^2) - \nu((pg)^2) = \mathbb{V}_\nu(f(X_0)) + 2 \sum_{k=1}^{+\infty} \mathbb{C}ov(f(X_0), f(X_k)).$$

(La convergence de la série ci-dessus fait partie de la conclusion du théorème.)

Exercice 182 *Prouver la proposition ci-dessus.*

A présent, calculons la variance de $S_n(f)$ sous \mathbb{P}_ν . Posons pour $k \geq 0$,

$$c(k) = \mathbb{E}_\nu(f(X_0)f(X_k)) = \mathbb{C}ov_\nu(f(X_0), f(X_k))$$

(rappelons-nous que $\nu(f) = 0$). Un calcul facile obtenu en développant le carré, utilisant la stationnarité de ν , montre que

$$\mathbb{V}_\nu \left[n^{-1/2} S_n(f) \right] = c(0) + 2 \sum_{k=1}^n (1 - k/n) c(k).$$

Si la série $\sum_{k=1}^{+\infty} c_k$ converge, on en déduit facilement que

$$\lim_{n \rightarrow +\infty} \mathbb{V}_\nu \left[n^{-1/2} S_n(f) \right] = c(0) + 2 \sum_{k=1}^{+\infty} c_k,$$

et l'on retrouve alors l'expression de la variance obtenue dans la proposition précédente. On note que celle-ci fait apparaître le terme que l'on aurait eu si les $(f(X_i))_{i \geq 0}$ étaient i.i.d., à savoir $c(0)$, auquel s'ajoutent des termes de covariance liés à la dépendance existant entre ces variables.

Proposition 89 *Si la suite $(c(0) + 2 \sum_{k=1}^n (1 - k/n) c(k))_{n \geq 0}$ converge lorsque $n \rightarrow +\infty$, et, en particulier, si la série $\sum_{k=1}^{+\infty} c_k$ converge, $n^{-1/2}(f(X_0) + \dots + f(X_n))$ converge en loi vers une gaussienne centrée.*

Preuve :

On observe qu'une telle hypothèse entraîne la tension de la suite

$$\left[n^{-1/2}(f(X_0) + \dots + f(X_n)) \right]_{n \geq 1},$$

et les théorèmes 20 et 21 entraînent donc la validité du TCL. \square

Un problème est que l'on ne peut pas garantir en général que, lorsque la proposition ci-dessus s'applique, la variance asymptotique dans le TCL (celle de la gaussienne) est effectivement égale à la limite de la variance des sommes partielles. Il est

pour cela nécessaire d'ajouter des hypothèses. Nous avons vu dans la partie précédente que, si $\sum_{k=1}^{+\infty} p^k f$ converge dans $L^2(\nu)$, alors c'est effectivement le cas. C'est donc ce qui se produit dans les deux exemples que nous avons donnés (ergodicité de degré 2 + fonction bornée, ou ergodicité géométrique + fonction de $L^{2+\epsilon}(\nu)$). Il est en fait possible (voir [9]) de prouver le résultat plus fort suivant :

Proposition 90 *Si la série $\sum p^k f$ converge dans $L^1(\nu)$, alors la série $c(0) + 2 \sum_{k=1}^{+\infty} c(k)$ converge, et la variance asymptotique dans le TCL est égale à sa limite.*

Nous retiendrons que, de manière générale, les hypothèses permettant d'obtenir le TCL portent sur la rapidité de la convergence de p^k vers l'équilibre, de manière assez analytique dans cette partie, de manière plus implicite dans l'approche par renouvellement.

Notons qu'il est également possible d'obtenir des principes d'invariance, c'est-à-dire la convergence en loi des trajectoires renormalisées vers un mouvement brownien.

Pour une discussion plus détaillée des liens entre l'existence d'un TCL et les différentes expressions possibles de la variance, nous renvoyons à [25].

Chapitre 8

Critères de dérive

Les critères de dérive – «drift criteria» en anglais, encore appelés critères de fonctions de Lyapounov – fournissent l'un des outils utilisés en pratique pour étudier le comportement en temps long des chaînes de Markov. Ils ont l'avantage de ne faire appel qu'à une analyse «à un pas» des transitions de la chaîne. Notons l'analogie étroite entre les théorèmes qui suivent et ceux qui ont cours dans le contexte des fonctions de Lyapounov associées aux équations différentielles. Nous ne présentons ici que les exemples les plus simples de ces critères, dont il existe de nombreux raffinements. Nous renvoyons à [36, 19] ou encore [7], pour plus de détails et des exemples d'utilisation de ces critères.

8.1 Un critère de non-réurrence positive.

Théorème 25 *Considérons un noyau de transition irréductible sur un ensemble fini ou dénombrable S , et supposons qu'il existe une fonction $V : S \rightarrow \mathbb{R}$ vérifiant les propriétés suivantes :*

- (i) *pour tout $x \in S$, $V(x) \geq 0$;*
- (ii) *V n'est pas constante sur S ;*
- (iii) *pour tout $x \in S$, $\mathbb{E}_x(V(X_1)) < +\infty$;*
- (iv) *pour tout $x \in S$, $\mathbb{E}_x(V(X_1) - V(X_0)) \geq 0$;*
- (v) *il existe $C \geq 0$ tel que, pour tout $x \in S$, $\mathbb{E}_x|V(X_1) - V(X_0)| \leq C$.*

Alors la chaîne est soit transiente, soit récurrente nulle.

Preuve :

Supposons la récurrence positive, et considérons $x, y \in S$. On a

$$V(X_{T_1(x)}) = V(X_0) + \sum_{i=0}^{T_1(x)-1} (V(X_{i+1}) - V(X_i)),$$

soit

$$V(X_{T_1(x)}) = V(X_0) + \sum_{i=0}^{+\infty} (V(X_{i+1}) - V(X_i)) \mathbf{1}(T_1(x) > i).$$

On voit alors facilement que

$$\sum_{i=0}^{+\infty} \mathbb{E}_y (|V(X_{i+1}) - V(X_i)| \mathbf{1}(T_1(x) > i)) \leq C \sum_{i=0}^{+\infty} \mathbb{P}_y(T_1(x) > i) < +\infty,$$

en utilisant (v) et la récurrence positive. Par Fubini, on a donc

$$\mathbb{E}_y \left(\sum_{i=0}^{+\infty} (V(X_{i+1}) - V(X_i)) \mathbf{1}(T_1(x) > i) \right) = \sum_{i=0}^{+\infty} \mathbb{E}_y ((V(X_{i+1}) - V(X_i)) \mathbf{1}(T_1(x) > i)),$$

et le fait que

$$\mathbb{E}_y(V(X_{i+1}) - V(X_i) | \mathcal{F}_i) \geq 0$$

d'après (iv) entraîne que

$$\mathbb{E}_y ((V(X_{i+1}) - V(X_i)) \mathbf{1}(T_1(x) > i)) \geq 0,$$

car $T_1(x) > i$ est un événement de \mathcal{F}_i . Or, avec probabilité 1 sous \mathbb{P}_y , on a $V(X_0) = V(y)$, et, en utilisant la récurrence supposée, $V(X_{T_1(x)}) = V(x)$. On déduit donc de cette positivité le fait que $V(x) \geq V(y)$ pour tous x et y , donc V doit être constante. Contradiction. \square

Remarque 28 *L'exemple défini sur $S = \mathbb{N}$ par $p(i, 0) = 1/2$ et $p(i, 2i + 1) = 1/2$, et $V(i) = i$ montre que les hypothèses (i) à (iv) ne suffisent pas à assurer la non-récurrence positive.*

8.2 Un critère de transience

Théorème 26 *Considérons un noyau de transition irréductible sur un ensemble fini ou dénombrable S , et supposons qu'il existe une fonction $V : S \rightarrow \mathbb{R}$ vérifiant les propriétés suivantes*

- (i) pour tout $x \in S$, $V(x) \geq 0$;
- (ii) V n'est pas constante sur S ;
- (iii) pour tout $x \in S$, $\mathbb{E}_x(V(X_1)) < +\infty$;
- (iv) pour tout $x \in S$, $\mathbb{E}_x(V(X_1) - V(X_0)) \leq 0$;

Alors la chaîne est transiente. La réciproque est vraie.

Preuve :

On remarque que les hypothèses entraînent automatiquement que $\mathbb{E}_y(V(X_n)) < +\infty$ pour tous y et n . Ce résultat a en fait déjà été vu : V est une fonction sur-harmonique positive non-constante sous \mathbb{P}_x , ce qui est impossible si la chaîne est récurrente. Pour la réciproque, on suppose qu'il y a transience, et l'on fixe un point a et on prend la fonction définie par $V(x) := \mathbb{P}_x(T_1(a) < +\infty)$ si $x \neq a$, et $V(a) := 1$. \square

8.3 Un critère de récurrence

Théorème 27 *Considérons un noyau de transition irréductible sur un ensemble fini ou dénombrable S , et supposons qu'il existe une fonction $V : S \rightarrow \mathbb{R}$ vérifiant les propriétés suivantes*

- (i) pour tout $x \in S$, $V(x) \geq 0$;
- (ii) pour tout $x \in S$, $\mathbb{E}_x(V(X_1)) < +\infty$;
- (iii) il existe un sous-ensemble fini C de S , tel que, pour tout $x \notin C$,

$$\mathbb{E}_x(V(X_1) - V(X_0)) \leq 0;$$

- (iv) pour tout K , l'ensemble $\{x \in S; V(x) \leq K\}$ est fini.

Alors la chaîne est récurrente.

Preuve :

Considérons $y \in S \setminus C$. On remarque que l'hypothèse entraîne facilement que $\mathbb{E}_y(V(X_n)) < +\infty$ pour tout n . Posons $T_1(C) = \inf\{n \geq 0; X_n \in C\}$ (avec la convention habituelle $\inf \emptyset = +\infty$). On pose

$$Y_n = V(X_n)\mathbf{1}(T_1(C) > n).$$

En utilisant la positivité de V , on obtient que, pour tout $n \geq 0$,

$$\mathbb{E}_y(V(X_{n+1})\mathbf{1}(T_1(C) > n+1)|\mathcal{F}_n) \leq \mathbb{E}_y(V(X_{n+1})\mathbf{1}(T_1(C) > n)|\mathcal{F}_n).$$

Comme $T_1(C)$ est un temps d'arrêt, l'événement $\{T_1(C) > n\}$ est mesurable par rapport à \mathcal{F}_n , et l'on a donc

$$\mathbb{E}_y(V(X_{n+1})\mathbf{1}(T_1(C) > n)|\mathcal{F}_n) = \mathbb{E}_y(V(X_{n+1})|\mathcal{F}_n)\mathbf{1}(T_1(C) > n).$$

Comme, compte-tenu de l'hypothèse (iii), on a, sur l'événement $X_n \notin C$, l'inégalité

$$\mathbb{E}_y(V(X_{n+1})|\mathcal{F}_n) \leq V(X_n),$$

on en déduit finalement que

$$\mathbb{E}_y(Y_{n+1}|\mathcal{F}_n) \leq Y_n,$$

et donc le fait que (Y_n) est une surmartingale positive par rapport à (\mathcal{F}_n) , pour la probabilité \mathbb{P}_y . Par le théorème de convergence des surmartingales, Y_n tend donc avec probabilité 1 (sous \mathbb{P}_y) vers une limite (a priori aléatoire) finie lorsque n tend vers l'infini, et donc, sur l'événement $\{T_1(C) = +\infty\}$, $V(X_n)$ tend avec probabilité 1 vers une limite finie.

Par ailleurs, si nous supposons le noyau transient, l'hypothèse (iv) entraîne que, pour tout entier K l'ensemble, fini d'après nos hypothèses, $\{x \in S; V(x) \leq K\}$ n'est visité qu'un nombre fini de fois par la chaîne, et par conséquent on doit avoir $\lim_{n \rightarrow +\infty} V(X_n) = +\infty$ avec probabilité 1. En revenant à (Y_n) , on constate donc que l'événement $\{T_1(C) = +\infty\}$ doit donc avoir une probabilité nulle sous \mathbb{P}_y , y pouvant être choisi arbitrairement hors de l'ensemble C . Par irréductibilité de la chaîne, et en utilisant le fait que C est un ensemble fini, on en déduit facilement que la chaîne est récurrente (en cas de doute, regarder la preuve du théorème suivant), d'où une contradiction. On conclut donc finalement que p est transient. \square

Remarque 29 *Bien observer la différence avec le théorème précédent! Par ailleurs, on peut seulement supposer que V est bornée inférieurement, quitte à lui ajouter une constante.*

8.4 Un critère de récurrence positive

Théorème 28 *(Critère de Foster) Considérons un noyau de transition irréductible sur un ensemble fini ou dénombrable S , et supposons qu'il existe une fonction $V : S \rightarrow \mathbb{R}$ vérifiant les propriétés suivantes*

- (i) pour tout $x \in S$, $V(x) \geq 0$;
- (ii) pour tout $x \in S$, $\mathbb{E}_x(V(X_1)) < +\infty$;
- (iii) il existe $\epsilon > 0$ et un sous-ensemble fini C de S , tel que, pour tout $x \notin C$, $\mathbb{E}_x(V(X_1) - V(X_0)) \leq -\epsilon$.

Alors la chaîne est récurrente positive. La réciproque est vraie.

Preuve :

On remarque que l'hypothèse entraîne automatiquement que $\mathbb{E}_y(V(X_n)) < +\infty$ pour tous y et n . On reprend les notations utilisées dans la preuve du théorème précédent. D'après celui-ci, on sait déjà qu'il y a récurrence. On vérifie facilement que les hypothèses entraînent le fait que, pour $x \notin C$, $\mathbb{E}_x(Y_{n+1}) \leq \mathbb{E}_x(Y_n) - \epsilon \mathbb{P}_x(T_1(C) >$

n). On en déduit facilement en itérant et en utilisant la positivité de Y , que $\mathbb{E}_x(Y_0) \geq \epsilon \sum_{i=0}^n \mathbb{P}_x(T_1(C) > i)$, d'où le fait que

$$\mathbb{E}_x(T_1(C)) \leq V(x)/\epsilon.$$

Notons que, pour $x \in C$, on peut écrire que

$$\mathbb{E}_x(T_1(C)) = 1 + \sum_{y \notin C} p(x, y) \mathbb{E}_y(T_1(C)),$$

d'où

$$\mathbb{E}_x(T_1(C)) \leq 1 + \sum_{y \notin C} p(x, y) V(y)/\epsilon \leq 1 + \sum_{y \in S} p(x, y) V(y)/\epsilon,$$

la dernière expression étant finie grâce à l'hypothèse (ii). On obtient ainsi une borne explicite sur $\mathbb{E}_x(T_1(C))$ en termes de V et ϵ .

On conclut ensuite à la récurrence positive de la façon suivante. Partant de $y \in C$, définissons une suite de temps aléatoires par $\tau_0 := 0$ et, pour tout $i \geq 0$, $\tau_{i+1} := \inf\{n \geq \tau_i + 1; X_n \in C\}$. On vérifie facilement que les τ_i sont des temps d'arrêt, et que la suite $(X_{\tau_i})_{i \geq 0}$ est une chaîne de Markov sur l'ensemble C , irréductible car la chaîne de départ l'est, et donc positivement récurrente du fait que C est fini. Pour $x \in C$, appelons $H(y) := \inf\{i \geq 1; X_{\tau_i} = y\}$. Nous savons donc que $\mathbb{E}_y(H_y) < +\infty$. Pour tout $k \geq 0$, posons $S_k := \tau_{k+1} - \tau_k$. Notons que l'on a l'identité

$$T_1(y) = \tau_{H(y)} = \sum_{k=0}^{H(y)} S_k = \sum_{k=0}^{+\infty} S_k \mathbf{1}(H(y) > k).$$

A présent, notons que l'événement $H(y) > k$ est mesurable par rapport à la tribu \mathcal{F}_{τ_k} . Par conséquent, la propriété forte de Markov appliquée à l'instant τ_k entraîne que

$$\mathbb{E}_x(S_k \mathbf{1}(H(y) > k)) \leq M \mathbb{P}_x(H(y) > k),$$

où

$$M := \sup_{x \in C} \mathbb{E}_x(T_1(C)),$$

en utilisant le fait que $X_{\tau_k} \in C$. D'après ce qui précède, $\mathbb{E}_x(T_1(C)) < +\infty$ pour tout $x \in S$, et C est un ensemble fini, donc $M < +\infty$. On en déduit que $\mathbb{E}_x(T_1(y)) \leq M \sum_{k=0}^{+\infty} \mathbb{P}_x(H(y) > k) = M \mathbb{E}_y(H(y)) < +\infty$.

Pour la réciproque, on prend $a \in S$, et l'on pose $V(x) = \mathbb{E}_x(T_1(a))$ pour $x \neq a$, $V(a) = 0$, et $C = \{a\}$.

□

8.4.1 Un critère d'ergodicité géométrique

Théorème 29 *Considérons un noyau de transition irréductible sur un ensemble fini ou dénombrable S , et supposons qu'il existe une fonction $V : S \rightarrow \mathbb{R}$ vérifiant les propriétés suivantes*

- (i) pour tout $x \in S$, $V(x) \geq 1$;
- (ii) pour tout $x \in S$, $E_x(V(X_1)) < +\infty$;
- (iii) il existe $0 < \epsilon$ et un sous-ensemble fini C de S , tel que, pour tout $x \notin C$,

$$\mathbb{E}_x(V(X_1) - V(X_0)) \leq -\epsilon V(x).$$

Alors il existe $a, b > 0$ tels que, tout $n \geq 0$, $\mathbb{P}_x(T_1(x) \geq n) \leq a \exp(-bn)$.

Preuve :

Notons que la positivité de V impose que $\epsilon \leq 1$. On reprend les notations précédentes, en posant cette fois

$$Y_n = \rho_0^n V(X_n) \mathbf{1}(T_1(C) > n),$$

où l'on a choisi $\rho_0 > 1$ tel que $\rho_0(1 - \epsilon) \leq 1$. On vérifie que, pour $x \notin C$,

$$\mathbb{E}_x(Y_{n+1} | \mathcal{F}_n) \leq \rho_0(1 - \epsilon) Y_n,$$

et, par conséquent, pour tout $n \geq 0$, $\mathbb{E}_x(Y_n) \leq \mathbb{E}_x(Y_0) \leq V(x)$. En utilisant l'hypothèse (i), on en déduit que, pour tout $n \geq 0$,

$$\rho_0^n \mathbb{P}_x(T_1(C) > n) \leq V(x),$$

d'où, pour tout $\rho < \rho_0$, une inégalité de la forme

$$\mathbb{E}_x(\rho^{T_1}(C)) \leq cV(x),$$

où $c \geq 0$ dépend de ρ_0 et ρ , mais pas de x . A présent, pour tout $x \in C$, on a

$$\mathbb{E}_x(\rho^{T_1}(C)) = \rho \sum_{y \notin C} p(x, y) \mathbb{E}_y(\rho^{T_1}(C)) + \sum_{y \in C} p(x, y) \rho$$

d'où

$$\mathbb{E}_x(\rho^{T_1}(C)) \leq \rho + \sum_{y \notin C} p(x, y) \mathbb{E}_y(\rho^{T_1}(C)) \leq \rho + c \mathbb{E}_x(V(X_1)) < +\infty.$$

On reprend la décomposition utilisée dans la preuve du théorème précédent :

$$T_1(y) = \tau_{H(y)} = \sum_{k=0}^{H(y)} S_k.$$

En utilisant le fait que $(X_{\tau_i})_{i \geq 0}$ est une chaîne de Markov sur un ensemble fini, on déduit que, pour $y \in C$, $H(y)$ possède une queue sous-géométrique sous \mathbb{P}_y . Ensuite, pour tout k , la propriété forte de Markov entraîne que $\mathbb{E}_y(\rho^{S_k} | \mathcal{F}_{\tau_k}) = \mathbb{E}_{X_{\tau_k}}(\rho^{T_1(C)})$. Or, en vertu de ce qui précède, C étant un ensemble fini, $\sup_{x \in C} \mathbb{E}_x(\rho^{T_1(C)}) < +\infty$. Au vu de ce résultat, et des estimations précédentes, on constate que l'on peut appliquer la proposition 75 à la suite $(S_k)_{k \geq 0}$.

A présent, pour tout $\alpha > 0$, nous pouvons écrire que, par la borne de la réunion,

$$\mathbb{P}_y(T_1(y) \geq n) \leq \mathbb{P}_y(H(y) \geq \alpha n) + \mathbb{P}_y\left(\sum_{k=0}^{\lfloor \alpha n \rfloor} S_k \geq n\right).$$

Comme $H(y)$ possède une queue sous-géométrique, $\mathbb{P}_y(H(y) \geq \alpha n)$ est exponentiellement petit en n pour tout $\alpha > 0$, tandis que le lemme ci-dessus nous montre que $\mathbb{P}_y(\sum_{k=0}^{\lfloor \alpha n \rfloor} S_k \geq n)$ est exponentiellement petit en n pourvu que α soit suffisamment petit. On en déduit que, pour tout $y \in S$, il existe $a, b > 0$ tels que $\mathbb{P}_y(T_1(y) \geq n) \leq a \exp(-bn)$, ce qui prouve le résultat.

Pour prouver la réciproque, on considère un point a tel que $T_1(a)$ possède une queue sous-géométrique sous \mathbb{P}_a . On utilise le théorème 18 et le résultat de l'exercice 125, qui entraîne facilement l'existence de $r > 1$ tel que $E_x(r^{T_1(a)}) < +\infty$ pour tout $x \in S$. On pose alors $V(x) = E_x(r^{T_1(a)})$, pour $x \neq a$, $V(a) = 1$, et $C = \{a\}$. \square

Chapitre 9

Principe des méthodes de Monte-Carlo par chaînes de Markov

On désigne par le nom générique de méthodes de Monte-Carlo par chaînes de Markov (ou Markov Chain Monte Carlo, souvent abrégé en MCMC, en anglais les méthodes consistant à simuler une loi de probabilité ν à partir d'une chaîne de Markov ergodique de loi invariante ν . Le principe de base de la méthode est le suivant : en simulant n pas de la chaîne, avec n suffisamment grand, on s'attend à ce que X_n fournisse une variable aléatoire approximativement distribuée selon la loi ν . Alternativement, la loi des grands nombres entraîne que $n^{-1}S_n(f)$ fournit une approximation de $\nu(f)$.

9.1 Intérêt des méthodes MCMC

L'un des intérêts des méthodes MCMC est qu'elles peuvent être mises en œuvre dans des situations où une simulation directe de ν par les méthodes usuelles s'avère impossible ou prohibitive en temps de calcul. Par exemple lorsque ν n'est connue qu'à une constante multiplicative près, et que l'espace sur lequel ν est définie comporte un grand nombre d'éléments. Cette situation n'est pas rare, et elle est en fait caractéristique des mesures de Gibbs, qui apparaissent en physique statistique, et des probabilités conditionnelles à un événement complexe, qui apparaissent naturellement en statistique bayésienne. Dans une telle situation, il est en général aisé de construire et de simuler des chaînes de Markov ergodiques possédant ν pour loi invariante. Deux exemples canoniques (mais très loin d'être les seuls) sont l'algorithme de Metropolis et l'échantillonneur de Gibbs décrits ci-après.

Notons que la méthode classique de simulation directe par rejet, suppose égale-

ment que la loi à simuler n'est connue qu'à une constante multiplicative près. Cependant, les méthodes MCMC peuvent être mises en œuvre dans diverses situations où la probabilité de rejet obtenue par cette méthode directe est prohibitivement petite, permettant en quelque sorte une approche progressive de la loi à simuler n'entraînant que des probabilités de rejet raisonnables à chaque pas.

9.2 Qualité de l'approximation

Une question qui se pose naturellement dans le contexte des méthodes MCMC est celle de la précision de l'approximation de ν obtenue. Pour un n donné, quelle est la qualité des approximations obtenues ? Pour une précision donnée, comment doit-on choisir n ?

Divers types d'approches peuvent être envisagés pour aborder ces questions, parmi lesquels :

- une approche théorique, consistant à établir mathématiquement des bornes sur la qualité de l'approximation obtenue, c'est-à-dire sur la rapidité de la convergence de la chaîne vers l'équilibre. Cette approche peut être menée à bien dans plusieurs situations intéressantes, et nous en verrons quelques exemples dans la partie «Analyse quantitative de la convergence». Il faut cependant mentionner que, dans de nombreux exemples intéressants, il est très difficile, voire impossible, d'obtenir des bornes exploitables en pratique. Les problèmes mathématiques correspondants se révèlent trop difficiles !
- une approche exploitant des critères empiriques, testés au fur et à mesure de la simulation, et censés permettre de «diagnostiquer» le fait que la chaîne est proche de l'équilibre. Le plus souvent, ces critères sont basés sur un mélange de résultats mathématiques (par exemple des résultats asymptotiques) et d'hypothèses plus ou moins bien justifiées empiriquement, mais que l'on ne sait pas prouver dans le contexte envisagé. Les critères de ce type sont en général condamnés à être mis en défaut dans certaines situations «pathologiques» pour lesquelles ces hypothèses ne sont pas vérifiées, fournissant alors des résultats incorrects.
- l'approche dite de simulation «exacte», ou «parfaite», dans laquelle on cherche, en raffinant la méthode de simulation employée, à obtenir des variables aléatoires distribuées exactement selon la loi ν . Ces approches ne sont praticables que lorsque certaines hypothèses supplémentaires (par exemple, mais pas exclusivement, des propriétés de monotonie) sont vérifiées. Nous verrons deux exemples de cette approche : la méthode de Propp-Wilson et l'algorithme de Fill.

9.3 Deux exemples

Dans ce qui suit, nous donnons deux exemples classiques de chaînes de Markov se prêtant à une utilisation dans le contexte des méthodes MCMC. Ces deux méthodes peuvent être mises en oeuvre sous des hypothèses très générales, ce qui ne signifie pas qu'elles soient toujours performantes ou qu'elles constituent un standard absolu dans le domaine. On notera qu'elles ont pour particularité de conduire à des chaînes de Markov réversibles.

9.3.1 Algorithme de Metropolis

On suppose donnée une mesure de probabilité ν sur un ensemble S vérifiant $\nu(x) > 0$ pour tout $x \in S$, et un noyau de transition q irréductible et apériodique sur S et vérifiant $q(x, y) > 0$ si et seulement si $q(y, x) > 0$. Les transitions de l'algorithme de Metropolis peuvent être décrites de la manière suivante : partant de x , on génère Y selon la loi $q(x, \cdot)$. Si $\nu(Y)q(Y, x) \geq \nu(x)q(x, Y)$, on va en Y à coup sûr. Sinon, on va en Y avec une probabilité de $\nu(Y)q(Y, x)/\nu(x)q(x, Y)$, et l'on reste en x sinon. On vérifie que le noyau obtenu est irréductible et apériodique (les mouvements possibles sont les mêmes pour l'algorithme de Metropolis et pour q), et que ν est réversible pour ledit noyau.

Le choix du noyau q ainsi que de la condition initiale est laissé à l'utilisateur de la méthode ; il peut s'avérer critique pour l'efficacité de l'algorithme obtenu, et diverses préconisations, plus ou moins fondées sur des considérations théoriques ou empiriques, existent, quant à la manière d'effectuer celui-ci en fonction de ce que l'on sait de la loi ν .

9.3.2 Échantillonneur de Gibbs

On suppose ici que S est de la forme $S = A^V$, où A et V sont des ensembles finis, et que l'ensemble S_0 des $x \in S$ vérifiant $\nu(x) > 0$ est connexe pour la relation de voisinage définie sur S par les modifications d'une seule coordonnée. On appelle S_0 cet ensemble, et l'on construit une chaîne de Markov sur S_0 dont les transitions sont définies de la manière suivante.

Pour $x \in S_0$, $v \in V$ et $a \in A$, on note $x_{v,a}$ l'élément de S_0 défini par $x_{v,a}(w) = x(w)$ pour tout $w \neq v$, et $x_{v,a}(v) = a$.

Les transitions de l'échantillonneur de Gibbs sont définies ainsi : partant de $x \in S_0$, on choisit uniformément au hasard un élément v de V , et l'on définit y par $y(w) = x_{v,a}$, a étant choisi selon la probabilité $\nu(x_{v,a})(\sum_{b \in A} \nu(x_{v,b}))^{-1}$, c'est-à-dire la loi sous ν de la v -ème coordonnée d'un élément de S conditionné à être égal à x en toutes les coordonnées différentes de v . On vérifie que ce conditionnement est bien défini pour $x \in S_0$, et qu'un élément y qui ne serait pas dans S_0 a une probabilité

nulle d'être sélectionné par ce mécanisme de transition, qui laisse donc S_0 stable. On vérifie par ailleurs aisément le caractère irréductible et apériodique du noyau ainsi obtenu sur S_0 , ainsi que la réversibilité de ν par rapport à celui-ci.

Chapitre 10

Appendice : Mémento sur les martingales

Dans cet appendice, nous récapitulons brièvement les définitions et les propriétés relatives aux martingales que nous aurons l'occasion d'utiliser dans ce cours. Attention ce qui suit omet plusieurs résultats et notions importants, car nous nous contentons de mentionner ce dont nous avons explicitement besoin. Les ouvrages classiques sur la théorie des probabilités contiennent en général un chapitre (ou plus) consacré aux martingales, et la lecture d'un tel chapitre est vivement recommandée. Mentionnons par exemple [18, 49, 53].

Définition 15 Une suite de variables aléatoires $(M_n)_{n \geq 0}$ définie sur un espace de probabilité (Ω, \mathcal{F}, P) est une martingale par rapport à une filtration $(\mathcal{F}_n)_{n \geq 0}$ de \mathcal{F} (c'est-à-dire une famille de sous-tribus de \mathcal{F} telle que, pour tous $n \leq m$, $\mathcal{F}_n \subset \mathcal{F}_m$) si, pour tout $n \geq 0$, les conditions suivantes sont vérifiées :

1. $E|M_n| < +\infty$;
2. M_n est mesurable par rapport à \mathcal{F}_n ;
3. $E(M_{n+1}|\mathcal{F}_n) = M_n$ $P - p.s.$

Lorsque la filtration considérée est la filtration naturelle de $(M_n)_{n \geq 0}$, c'est-à-dire lorsque $\mathcal{F}_n = \sigma(M_0, \dots, M_n)$, on dit simplement que $(M_n)_{n \geq 0}$ est une martingale. On note qu'une martingale par rapport à une certaine filtration est toujours une martingale par rapport à sa filtration naturelle. Si l'on remplace la condition 3) ci-dessus par $E(M_{n+1}|\mathcal{F}_n) \geq M_n$ $P - p.s.$, on dit que l'on a affaire à une sous-martingale, et l'on parle de sur-martingale lorsque la condition est remplacée par $E(M_{n+1}|\mathcal{F}_n) \leq M_n$ $P - p.s.$. La définition dans le cas d'un indice continu est très similaire : il suffit de demander que 1) et 2) aient lieu pour tout indice, et de remplacer la condition 3) par le fait que $E(M_t|\mathcal{F}_s) = M_s$ $P - p.s.$ pour tout couple $s \leq t$.

Proposition 91 (*Sous-version du théorème de Doob*) Si T est un temps d'arrêt de la filtration $(\mathcal{F}_n)_{n \geq 0}$, et que $(M_n)_{n \geq 0}$ est une martingale pour cette filtration (resp. sous-martingale, resp. sur-martingale), alors la suite $(M_{T \wedge n})_{n \geq 0}$ est également une martingale (resp. sous-martingale, resp. sur-martingale) pour cette filtration. (Avec la notation usuelle : $a \wedge b = \min(a, b)$.)

Corollaire 30 Si T est un temps d'arrêt borné de la filtration $(\mathcal{F}_n)_{n \geq 0}$, et que $(M_n)_{n \geq 0}$ est une martingale pour cette filtration (resp. sous-martingale, resp. sur-martingale), alors $E(X_T) = E(X_0)$ (resp. \geq , resp. \leq).

Exercice 183 Faire la preuve de la proposition.

Théorème 30 (*Inégalité maximale de Doob*)

Si $(M_n)_{n \geq 0}$ est une martingale, on a, pour tout $p > 1$, et tout $n \geq 0$, l'inégalité

$$P \left[\max_{0 \leq k \leq n} |M_k| \geq \lambda \right] \leq \frac{E(|M_n|^p)}{\lambda^p}.$$

Théorème 31 (*Théorème de convergence de Doob*) Si $(M_n)_{n \geq 0}$ est une martingale, une sur-martingale ou une sous-martingale, telle que $\sup_{n \geq 0} E|M_n| < +\infty$, alors, avec probabilité 1, la limite $M_{+\infty} := \lim_{n \rightarrow +\infty} M_n$ existe dans \mathbb{R} , et $E|M_{+\infty}| < +\infty$.

Corollaire 31 La conclusion du théorème de convergence ci-dessus est vérifiée si $(M_n)_{n \geq 0}$ est une martingale positive, ou une sur-martingale positive.

Chapitre 11

Appendice : Mémento sur la théorie ergodique

Dans cet appendice, nous récapitulons brièvement les définitions et les propriétés relatives à la théorie ergodique que nous aurons l'occasion d'utiliser dans ce cours. Attention ce qui suit omet plusieurs résultats et notions importants, car nous nous contentons de mentionner ce dont nous avons explicitement besoin. Voir par exemple [18, 49] pour une introduction à la théorie, et par exemple [41] pour une présentation plus approfondie.

L'objet de base de la théorie est la donnée d'un espace probabilisé (Ω, \mathcal{F}, P) muni d'une application mesurable $T : S \rightarrow S$ qui préserve la probabilité P , c'est-à-dire que la mesure-image de P par T est égale à P .

Définition 16 *Un sous-ensemble $A \in \mathcal{F}$ est dit invariant si $T^{-1}(A) = A$. On note \mathcal{I} l'ensemble des sous-ensembles invariants de \mathcal{F} , et l'on vérifie qu'il constitue une sous-tribu de \mathcal{F} .*

Définition 17 *On dit que T est ergodique lorsque \mathcal{I} est triviale pour P , i.e. $P(A) \in \{0, 1\}$ pour tout $A \in \mathcal{I}$.*

Etant donnée une variable aléatoire X définie sur (Ω, \mathcal{F}, P) et à valeurs réelles, définissons la variable $S_n(X)$ par

$$S_n(X) := \sum_{i=0}^n X \circ T^i,$$

où T^i désigne la fonction obtenue en composant i fois par T . Le théorème fondamental que nous utiliserons est le suivant, connu sous le nom de théorème ergodique de Birkhoff.

Théorème 32 *Etant donnée une variable aléatoire X définie sur (Ω, \mathcal{F}, P) et à valeurs réelles, on a la convergence suivante :*

$$\lim_{n \rightarrow +\infty} \frac{S_n(X)}{n} = E(X | \mathcal{I}), \text{ } P\text{-p.s. et dans } L^1(P).$$

On voit en particulier que, si T est ergodique, la limite dans le théorème ci-dessus n'est autre que $E(X)$, et est donc égale à une constante (alors que $E(X | \mathcal{I})$ est génériquement une variable aléatoire).

Bibliographie

- [1] D. J. Aldous and J. A. Fill. *Reversible Markov Chains and Random Walks on Graphs*. Book in preparation. Drafts available at <http://www.stat.berkeley.edu/users/aldous/RWG/book.html>.
- [2] David Aldous and Persi Diaconis. Shuffling cards and stopping times. *Amer. Math. Monthly*, 93(5) :333–348, 1986.
- [3] V. Anantharam and P. Tsoucas. A proof of the Markov chain tree theorem. *Statist. Probab. Lett.*, 8(2) :189–192, 1989.
- [4] William J. Anderson. *Continuous-time Markov chains*. Springer Series in Statistics : Probability and its Applications. Springer-Verlag, New York, 1991. An applications-oriented approach.
- [5] Vlad Stefan Barbu and Nikolaos Limnios. *Semi-Markov chains and hidden semi-Markov models toward applications*, volume 191 of *Lecture Notes in Statistics*. Springer, New York, 2008. Their use in reliability and DNA analysis.
- [6] Richard C. Bradley, Jr. Information regularity and the central limit question. *Rocky Mountain J. Math.*, 13(1) :77–97, 1983.
- [7] Pierre Brémaud. *Markov chains*, volume 31 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 1999. Gibbs fields, Monte Carlo simulation, and queues.
- [8] Olivier Cappé, Eric Moulines, and Tobias Rydén. *Inference in hidden Markov models*. Springer Series in Statistics. Springer, New York, 2005. With Randal Douc’s contributions to Chapter 9 and Christian P. Robert’s to Chapters 6, 7 and 13, With Chapter 14 by Gersende Fort, Philippe Soulier and Moulines, and Chapter 15 by Stéphane Boucheron and Elisabeth Gassiat.
- [9] Xia Chen. Limit theorems for functionals of ergodic Markov chains with general state space. *Mem. Amer. Math. Soc.*, 139(664) :xiv+203, 1999.
- [10] Kai Lai Chung. *Markov chains with stationary transition probabilities*. Second edition. Die Grundlehren der mathematischen Wissenschaften, Band 104. Springer-Verlag New York, Inc., New York, 1967.

- [11] John B. Conway. *A course in functional analysis*, volume 96 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1990.
- [12] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 1998.
- [13] P. Diaconis and D. Freedman. de Finetti's theorem for Markov chains. *Ann. Probab.*, 8(1) :115–130, 1980.
- [14] Persi Diaconis. *Group representations in probability and statistics*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11. Institute of Mathematical Statistics, Hayward, CA, 1988.
- [15] Persi Diaconis and James Allen Fill. Strong stationary times via a new form of duality. *Ann. Probab.*, 18(4) :1483–1522, 1990.
- [16] Persi Diaconis and Mehrdad Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, 57(2) :159–179, 1981.
- [17] Peter G. Doyle and J. Laurie Snell. *Random walks and electric networks*, volume 22 of *Carus Mathematical Monographs*. Mathematical Association of America, Washington, DC, 1984.
- [18] Richard Durrett. *Probability : theory and examples*. Duxbury Press, Belmont, CA, second edition, 1996.
- [19] G. Fayolle, V. A. Malyshev, and M. V. Men'shikov. *Topics in the constructive theory of countable Markov chains*. Cambridge University Press, Cambridge, 1995.
- [20] William Feller. *An introduction to probability theory and its applications. Vol. I*. Third edition. John Wiley & Sons Inc., New York, 1968.
- [21] William Feller. *An introduction to probability theory and its applications. Vol. II*. Second edition. John Wiley & Sons Inc., New York, 1971.
- [22] James Allen Fill. An interruptible algorithm for perfect sampling via Markov chains. *Ann. Appl. Probab.*, 8(1) :131–162, 1998.
- [23] Xavier Guyon. *Random fields on a network*. Probability and its Applications (New York). Springer-Verlag, New York, 1995. Modeling, statistics, and applications, Translated from the 1992 French original by Carenne Ludeña.
- [24] Olle Häggström. On the central limit theorem for geometrically ergodic Markov chains. *Probab. Theory Related Fields*, 132(1) :74–82, 2005.
- [25] Olle Häggström and Jeffrey S. Rosenthal. On variance conditions for Markov chain CLTs. *Electron. Comm. Probab.*, 12 :454–464 (electronic), 2007.
- [26] T. E. Harris. Transient Markov chains with stationary measures. *Proc. Amer. Math. Soc.*, 8 :937–942, 1957.

- [27] Mark Jerrum. *Counting, sampling and integrating : algorithms and complexity*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2003.
- [28] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.
- [29] C. Kipnis and S. R. S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.*, 104(1) :1–19, 1986.
- [30] Yevgeniy Kovchegov. Mixing times via super-fast coupling. *math.PR/0609568*.
- [31] Thomas M. Liggett. *Stochastic interacting systems : contact, voter and exclusion processes*, volume 324 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999.
- [32] Thomas M. Liggett. *Interacting particle systems*. Classics in Mathematics. Springer-Verlag, Berlin, 2005. Reprint of the 1985 original.
- [33] Torgny Lindvall. On Strassen’s theorem on stochastic domination. *Electron. Comm. Probab.*, 4 :51–59 (electronic), 1999.
- [34] Torgny Lindvall. *Lectures on the coupling method*. Dover Publications Inc., Mineola, NY, 2002. Corrected reprint of the 1992 original.
- [35] Russell Lyons, Robin Pemantle, and Yuval Peres. Ergodic theory on Galton-Watson trees : speed of random walk and dimension of harmonic measure. *Ergodic Theory Dynam. Systems*, 15(3) :593–619, 1995.
- [36] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1993.
- [37] Ravi Montenegro and Prasad Tetali. Mathematical aspects of mixing times in markov chains. volume 1 :3 of *Foundations and Trends in Theoretical Computer Science*, pages 237–354. NOW Publishers, 2006.
- [38] B. Morris and Yuval Peres. Evolving sets, mixing and heat kernel bounds. *Probab. Theory Related Fields*, 133(2) :245–266, 2005.
- [39] J. R. Norris. *Markov chains*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998. Reprint of 1997 original.
- [40] Esa Nummelin. *General irreducible Markov chains and nonnegative operators*, volume 83 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1984.

- [41] Karl Petersen. *Ergodic theory*, volume 2 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1989. Corrected reprint of the 1983 original.
- [42] James Gary Propp and David Bruce 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.
- [43] D. Revuz. *Markov chains*, volume 11 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, second edition, 1984.
- [44] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.
- [45] Gareth O. Roberts and Jeffrey S. Rosenthal. Geometric ergodicity and hybrid Markov chains. *Electron. Comm. Probab.*, 2 :no. 2, 13–25 (electronic), 1997.
- [46] L. C. G. Rogers and David Williams. *Diffusions, Markov processes, and martingales. Vol. 1*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. Foundations, Reprint of the second (1994) edition.
- [47] Walter Rudin. *Functional analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill Inc., New York, second edition, 1991.
- [48] Laurent Saloff-Coste. Lectures on finite Markov chains. In *Lectures on probability theory and statistics (Saint-Flour, 1996)*, volume 1665 of *Lecture Notes in Math.*, pages 301–413. Springer, Berlin, 1997.
- [49] A. N. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1996. Translated from the first (1980) Russian edition by R. P. Boas.
- [50] Frank Spitzer. *Principles of random walks*. Springer-Verlag, New York, second edition, 1976. Graduate Texts in Mathematics, Vol. 34.
- [51] Hermann Thorisson. *Coupling, stationarity, and regeneration*. Probability and its Applications (New York). Springer-Verlag, New York, 2000.
- [52] William Veech. The necessity of Harris' condition for the existence of a stationary measure. *Proc. Amer. Math. Soc.*, 14 :856–860, 1963.
- [53] David Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.
- [54] D. B. Wilson. *Web Site for Perfectly Random Sampling with Markov Chains*. <http://dbwilson.com/exact/>.