

# Processus Discrets

Année 2020 - 2021

François Simenhaus  
Bureau B 640  
[simenhaus@ceremade.dauphine.fr](mailto:simenhaus@ceremade.dauphine.fr)

# Table des matières

<b>1</b>	<b>Espérance conditionnelle</b>	<b>4</b>
1.1	Le cas $L^2$ . . . . .	4
1.2	Le cas $L^1$ . . . . .	7
1.3	Propriétés . . . . .	10
1.4	Cas particuliers . . . . .	12
<b>2</b>	<b>Martingales en temps discret</b>	<b>16</b>
2.1	Définitions : filtrations et martingales . . . . .	16
2.2	Théorèmes d'arrêt . . . . .	19
2.2.1	Temps d'arrêt . . . . .	20
2.2.2	Processus arrêté et transformation de processus . . . . .	21
2.2.3	Théorèmes d'arrêt . . . . .	24
2.3	Théorèmes de convergence . . . . .	26
2.3.1	Convergence presque sûre . . . . .	27
2.3.2	Convergence $L^2$ . . . . .	30
<b>3</b>	<b>Chaînes de Markov</b>	<b>32</b>
3.1	Définitions et équations de Chapman-Kolmogorov . . . . .	32
3.1.1	Définitions . . . . .	32
3.1.2	Récurrences aléatoires . . . . .	35
3.1.3	Équations de Chapman Kolmogorov . . . . .	36
3.2	Propriété de Markov forte . . . . .	42
3.3	Récurrence et transience . . . . .	44
3.4	Mesures stationnaires et probabilités invariantes . . . . .	50
3.4.1	Mesures stationnaires . . . . .	51
3.4.2	Probabilités invariantes . . . . .	55
3.5	Théorème ergodique et convergence à l'équilibre . . . . .	58
3.5.1	Excursions . . . . .	58
3.5.2	Théorème ergodique . . . . .	60
3.5.3	Convergence à l'équilibre des chaînes apériodiques . . . . .	62

# Introduction

Un amphi d'1h30 par semaine et une séance de TD de la même durée. La note est déterminée par

$$\text{Note} = \max(\text{Examen} ; 0,4 \text{ Partiel} + 0,6 \text{ Examen}). \quad (1)$$

Ces notes de cours sont largement inspirées de celles de M. Gubinelli et J. Lehec.

---

**Dans tout le cours  $(\Omega, \mathcal{F}, P)$  désigne un espace de probabilité. Tous les objets aléatoires considérés sont construits sur cet espace. On notera  $E$  (resp.  $P$ ) l'espérance (resp. la probabilité) associée. Si  $X$  est une variable aléatoire et  $A \in \mathcal{F}$  un événement, on utilisera parfois la notation**

$$E(X, A) := E(X1_A). \quad (2)$$

**Un processus discret est une suite de variables aléatoires définies sur un même espace de probabilités  $(\Omega, \mathcal{F}, P)$ .**

---

Le terme discret indique que l'ensemble qui indexe le processus est discret...plus simplement : le temps est discret. Un processus sert à décrire une évolution aléatoire dans le temps (ou l'espace). Les exemples d'applications sont nombreux dans des disciplines aussi diverses que l'écologie, l'économie, la biologie, la physique statistique ou encore la finance.

Ce cours comporte deux grand massifs : **la théorie des martingales** et celle des **processus de Markov**.

Les processus de Markov ont été introduits à la fin du 19 ème par le mathématicien russe Andreï Markov, avant donc la formalisation moderne des probabilités par Kolmogorov au début du 20 ème. Un processus est markovien s'il vérifie la propriété suivante : étant donné un temps  $n$ , la loi du

futur conditionné au passé est identique à la loi du futur conditionné au seul présent.

La théorie des martingales est plus récente. Les mathématiciens Levy, Borel ou encore Doob ont largement participé à son développement. Nous étudierons principalement les théorèmes d'arrêt et le comportement asymptotique de ces processus.

Le cours commence par un chapitre sur l'espérance conditionnelle. Il est indispensable de bien maîtriser cette notion pour suivre ce cours. Il est nécessaire de réviser également le cours de mesure et intégration ainsi que le lien entre cette théorie et celle des probabilités.

# Chapitre 1

## Espérance conditionnelle

On considère dans tout ce chapitre une variable aléatoire **intégrable**,

$$X : (\Omega, \mathcal{F}, P) \mapsto (\mathbb{R}, \mathcal{B}(R)), \quad (1.1)$$

i.e.  $E(|X|) < \infty$ . On considère également une sous-tribu  $\mathcal{B}$  de  $\mathcal{F}$ .

Il faut s'habituer à l'idée, présente dans tout ce cours, qu'une tribu ou une sous-tribu représente une **information**. Plus une tribu contient d'évènements plus elle contient d'information.

Le problème que l'on cherche à résoudre dans ce chapitre est de construire une variable aléatoire appelée **espérance conditionnelle de  $X$  sachant  $\mathcal{B}$**  qui actualise notre connaissance de  $X$  sachant l'information partielle  $\mathcal{B}$  dont on dispose.

### 1.1 Le cas $L^2$

On rappelle que  $L^2(\Omega, \mathcal{F}, P)$  désigne l'ensemble des variables aléatoires de carré intégrable :

$$X \in L^2(\Omega, \mathcal{F}, P) \Leftrightarrow E(X^2) < +\infty. \quad (1.2)$$

On munit cet espace du produit scalaire

$$\begin{aligned} \langle \cdot, \cdot \rangle : L^2 \times L^2 &\rightarrow \mathbb{R} \\ (X, Y) &\mapsto E(XY). \end{aligned} \quad (1.3)$$

Dans toute cette section on suppose que  $X \in L^2$ . On cherche à approximer  $X$  par une variable de  $L^2$  qui soit  $\mathcal{B}$ -mesurable. L'ensemble des variables aléatoires de  $L^2$  qui sont  $\mathcal{B}$ -mesurables est un sous-ensemble de  $L^2$ , c'est même un sous-espace vectoriel, on le notera  $L^2(\mathcal{B})$ . On cherche donc la variable

aléatoire de  $L^2(\mathcal{B})$  la plus proche pour la norme  $\|\cdot\|_2$  de  $X$  (qui n'est en général pas dans  $L^2(\mathcal{B})$  car seulement  $\mathcal{F}$ -mesurable mais pas  $\mathcal{B}$ -mesurable). **La solution de ce problème est de projeter orthogonalement  $X$  sur  $L^2(\mathcal{B})$ .**

L'espace  $L^2$  est un espace de Hilbert et nous disposons donc des outils pour réaliser cette projection. Commençons par un rappel de ces outils (dont je vous invite à réviser les preuves).

**Théorème 1** (Théorème de projection dans les espaces de Hilbert). *Soit  $(E, \langle \cdot, \cdot \rangle)$  un espace de Hilbert et  $C$  un convexe fermé non vide de  $E$ . On note  $\|\cdot\|$  la norme associée au produit scalaire. Soit  $x \in E$ . Il existe un unique  $y \in C$  tel que*

$$\|x - y\| = \inf\{\|x - z\|, z \in C\}. \quad (1.4)$$

*De plus,  $y$  est l'unique élément de  $C$  vérifiant :*

$$\forall z \in C, \quad \langle y - x, y - z \rangle \leq 0. \quad (1.5)$$

Pour tout  $x$  dans  $E$ , le théorème précédent définit un unique vecteur de  $C$  que l'on nomme projeté orthogonal de  $x$  sur  $C$  et que l'on note  $p_C^\perp(x)$ . Nous allons en fait utiliser ce théorème dans le cas particulier suivant :

**Corollaire 1.** *Dans le cas où  $C$  est un sous-espace vectoriel fermé, le théorème s'applique car un espace vectoriel est un convexe non vide. De plus  $p_C^\perp(x)$  est l'unique élément tel que*

$$\forall z \in C, \quad \langle x, z \rangle = \langle p_C^\perp(x), z \rangle. \quad (1.6)$$

Ce théorème nous assure donc que l'on peut projeter dans un espace de Hilbert. Il reste à s'assurer que  $L^2$  est bien un espace de Hilbert.

**Théorème 2.** *L'espace  $L^2(\Omega, \mathcal{F}, P)$  muni du produit scalaire  $\langle \cdot, \cdot \rangle$  est un espace de Hilbert.*

Le seul point délicat est de montrer que cet espace est complet. Il s'agit en fait d'un cas particulier du théorème de Riesz Fischer :

**Théorème 3.** *Pour tout  $1 \leq p \leq +\infty$ , l'espace  $(L^p, \|\cdot\|_p)$  est complet (on dit que c'est un espace de Banach), où pour tout  $X \in L^p$ ,*

$$\|X\|_p = E(|X|^p)^{1/p}. \quad (1.7)$$

Nous avons désormais tous les ingrédients pour définir l'espérance conditionnelle dans le cas  $L^2$  avec la proposition suivante.

**Proposition 1** (Définition de l'espérance conditionnelle dans le cas  $L^2$ ). Soit  $X \in L^2(\Omega, \mathcal{F}, P)$  et  $\mathcal{B} \subset \mathcal{F}$  une sous-tribu. Il existe une unique p.s. variable aléatoire  $Y$  satisfaisant les deux points suivants :

1.  $Y$  est  $\mathcal{B}$ -mesurable
2. Pour toute variable aléatoire  $Z$ ,  $\mathcal{B}$ -mesurable bornée  $E(XZ) = E(YZ)$

On note  $Y := E(X|\mathcal{B})$ .

Le second point peut sembler mystérieux et rendre cette définition abstraite. Il faut le comprendre comme une égalité de projection similaire à celle du Corollaire 1. En effet,  $E(XZ) = E(YZ)$  se réécrit  $\langle X - Y, Z \rangle = 0$ . Le fait que  $X - Y$  soit orthogonal à toutes les variables de  $L^2(\mathcal{B})$  caractérise bien  $Y$  comme le projeté orthogonal de  $X$  sur  $L^2(\mathcal{B})$ .

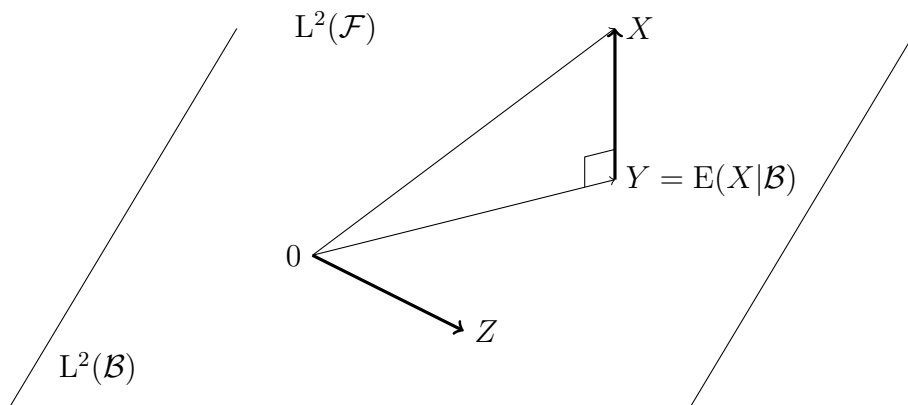


FIGURE 1.1: La variable  $X - Y$  est orthogonale à toutes les variables  $Z$  de  $L^2(\mathcal{B})$ . Evidemment le dessin est faux : les espaces que nous considérons ne sont pas de dimension finie.

*Preuve de la Proposition 1.* Pour l'**existence** on utilise le théorème de projection ou plutôt son corollaire. Vérifions les hypothèses :

1.  $L^2(\mathcal{F})$  est un espace de Hilbert (c'est le théorème de Riesz-Fischer).
2.  $L^2(\mathcal{B})$  est un espace vectoriel fermé. En effet, si une suite  $(T_n)_{n \geq 0}$  de variables dans  $L^2(\mathcal{B})$  converge dans  $L^2$  vers  $T$  alors on peut extraire une sous-suite de  $(T_n)_{n \geq 0}$  qui converge p.s. vers  $T$ . Or la convergence p.s. preserve la mesurabilité donc  $T$  est aussi  $\mathcal{B}$ -mesurable. On peut aussi simplement rappeler que  $L^2(\mathcal{B})$  est complet.

Le Corollaire 1 permet de définir

$$Y = p_{L^2(\mathcal{B})}^\perp(X), \quad (1.8)$$

vérifiant

1.  $Y \in L^2(\mathcal{B})$ ,
2. pour tout  $Z \in L^2(\mathcal{B})$  (donc en particulier pour tout  $Z$   $\mathcal{B}$ -mesurable borné),  $E(XZ) = E(YZ)$ .

L'unicité n'est pas assurée gratuitement par le Corollaire 1 puisque  $Y$  vérifie le second point pour toutes les variables bornées et non toutes les variables de  $L^2(\mathcal{B})$ . Une piste serait donc de montrer que  $Y$  satisfait en fait également 2 pour toute variable de  $L^2(\mathcal{B})$ . Nous allons plutôt utiliser le lemme suivant qui nous servira également plus tard.

**Lemme 1** (Monotonie dans le cas  $L^2$ ). *Soient  $X_1$  et  $X_2$  deux variables de  $L^2$  telles que  $X_1 \leq X_2$  presque sûrement. Soit  $Y_1$  vérifiant les deux points de la Proposition 1 pour  $X_1$  et  $Y_2$  vérifiant ces deux mêmes points pour  $X_2$ . Alors*

$$Y_1 \leq Y_2 \quad \text{p.s.} \quad (1.9)$$

*Preuve du Lemme 1.* On considère l'événement  $A := \{Y_1 - Y_2 \geq 0\}$ . La variable  $1_A$  est clairement  $\mathcal{B}$ -mesurable bornée donc d'après le second point

$$E((Y_1 - Y_2)1_A) = E((X_1 - X_2)1_A) \leq 0, \quad (1.10)$$

car  $(X_1 - X_2)1_A \leq 0$  p.s. Par ailleurs  $(Y_1 - Y_2)1_A \geq 0$  par la définition même de  $A$ . La variable  $(Y_1 - Y_2)1_A$  est donc positive p.s. d'espérance négative, on en déduit qu'elle est nulle p.s. Cela prouve  $Y_1 \leq Y_2$  p.s.  $\square$

En prenant  $X_1 = X_2 = X$ , on obtient l'unicité dans la proposition 1.  $\square$

## 1.2 Le cas $L^1$

On perd le produit scalaire, donc la structure hilbertienne...et donc les outils de projection orthogonale. On va néanmoins étendre la définition précédente.

**Proposition 2** (Définition de l'espérance conditionnelle dans le cas  $L^1$ ). *Soit  $X \in L^1(\Omega, \mathcal{F}, P)$  (**attention** : l'unique différence avec la Proposition 1 est ici !) et  $\mathcal{B} \subset \mathcal{F}$  une sous-tribu. Il existe une unique p.s. variable aléatoire  $Y$  satisfaisant les deux points suivants :*

1.  $Y$  est  $\mathcal{B}$ -mesurable
2. Pour toute variable aléatoire  $Z$ ,  $\mathcal{B}$ -mesurable bornée  $E(XZ) = E(YZ)$

On note  $Y := E(X|\mathcal{B})$ .



*Preuve.* L'unicité se prouve comme dans le cas  $L^2$ . Pour l'existence on commence par le cas  $X \geq 0$  p.s. On définit pour tout  $n \geq 0$

$$X_n = X \wedge n. \quad (1.11)$$

Pour tout  $n \geq 0$ , la variable  $X_n$  est dans  $L^2$  et on peut donc définir son espérance conditionnelle  $Y_n := E(X_n | \mathcal{B})$ . De plus d'après le Lemme 1, comme la suite  $(X_n)_{n \geq 0}$  est croissante p.s., la suite  $(Y_n)_{n \geq 0}$  est également croissante p.s. On note

$$Y := \lim_{n \rightarrow \infty} \uparrow Y_n \in [0, +\infty]. \quad (1.12)$$

La variable  $Y$  satisfait bien les deux points de la proposition :

1.  $Y$  est  $\mathcal{B}$ -mesurable comme limite p.s. de variable  $\mathcal{B}$ -mesurable
2. Soit  $Z$   $\mathcal{B}$ -mesurable borné. On suppose dans un premier temps que  $Z \geq 0$  p.s. Par définition de l'espérance conditionnelle dans le cas  $L^2$ , pour tout  $n \geq 0$ ,

$$E(X_n Z) = E(Y_n Z). \quad (1.13)$$

Les suites  $(X_n Z)_{n \geq 0}$  et  $(Y_n Z)_{n \geq 0}$  sont positives et croissantes (vers  $XZ$  et  $YZ$ ). D'après le théorème de convergence monotone, on en déduit

$$E(XZ) = E(YZ). \quad (1.14)$$

Revenons au cas où  $Z$  n'est pas supposé positif. On a  $Z = Z^+ - Z^-$ , avec  $Z^+$  et  $Z^-$  positifs donc

$$E(XZ) = E(XZ^+) - E(XZ^-) = E(YZ^+) - E(YZ^-) = E(YZ). \quad (1.15)$$

**Remarque 1.** *On note qu'à ce stade de la preuve nous n'avons pas encore utilisé que  $X \in L^1$ ...on peut donc tout à fait définir l'espérance conditionnelle pour une variable aléatoire positive p.s. Cette espérance conditionnelle peut être infinie sur un ensemble de probabilité non nulle.*

On termine avec le cas général sans supposer que  $X$  est positive. On définit alors

$$E(X | \mathcal{B}) = E(X^+ | \mathcal{B}) - E(X^- | \mathcal{B}). \quad (1.16)$$

Cette définition a bien du sens, c'est-à-dire que nous ne sommes pas en train d'écrire  $\infty - \infty$ . En effet,  $X^+ \geq 0$  est dans  $L^1$  donc

$$E(E(X^+ | \mathcal{B}) 1_\Omega) = E(X^+ 1_\Omega) < +\infty. \quad (1.17)$$

On en déduit que  $E(X^+|\mathcal{B}) < +\infty$  p.s. La définition a donc du sens et, comme la variable proposée est  $\mathcal{B}$ -mesurable, il ne reste plus qu'à vérifier le second point. Pour tout  $Z$   $\mathcal{B}$ -mesurable borné, par linéarité,

$$E((E(X^+|\mathcal{B}) - E(X^-|\mathcal{B}))Z) = E((X^+ - X^-)Z) = E(XZ). \quad (1.18)$$

□

**Remarque 2.** *Dans la proposition-définition de l'espérance conditionnelle, on peut remplacer le point 2. par le point suivant :*

$$2'. \text{ Pour tout événement } A \in \mathcal{B}, E(X1_A) = E(Y1_A)$$

*Preuve.* Il est clair que le point 2. implique le point 2'. Pour la réciproque, on commence par supposer  $X \geq 0$  et  $Z \geq 0$  p.s. On utilise alors l'approximation de variables positives par des fonctions étagées puis le théorème de convergence monotone. On conclut dans le cas général par linéarité en décomposant  $Z$  et  $X$  comme somme de leur parties positives et négatives. □

**Remarque 3.** *On peut aussi prouver que si  $Z$  est  $\mathcal{B}$ -mesurable et tel que  $XZ$  est dans  $L^1$  alors le point 2. reste vrai :*

$$E(E(X|\mathcal{B})Z) = E(XZ). \quad (1.19)$$

Ces deux dernières remarques fonctionnent un peu en sens inverse ! On a envie de mettre la condition minimale dans l'égalité de projection afin de prouver facilement qu'une variable est l'espérance conditionnelle (on a ainsi « peu » de fonctions à tester)...c'est le rôle de la Remarque 2. Par contre, une fois qu'on a prouvé qu'une variable est l'espérance conditionnelle on voudrait pouvoir écrire l'égalité de projection pour un maximum de variables...et c'est le rôle de la Remarque 3.

Avec un léger abus de notation on utilise souvent la définition suivante :

**Définition 1.** *Soit  $Y$  une variable aléatoire. On définit*

$$E(X|Y) = E(X|\sigma(Y)). \quad (1.20)$$

On rappelle que  $\sigma(Y)$  désigne la plus petite tribu qui rende  $Y$  mesurable, ou encore,  $\sigma(Y) := \{Y \in A, A \in \mathcal{B}(\mathbb{R})\}$ . On notera que l'espérance conditionnelle d'une variable par rapport à  $Y$  ne dépend que de  $\sigma(Y)$  et non « réellement » de  $Y$ .

## 1.3 Propriétés

**Proposition 3.** *L'opérateur espérance conditionnelle construit dans la partie précédente vérifie les points suivants :*

1. **Variable  $\mathcal{B}$ -mesurable.** *Si  $X \in L^1$  est  $\mathcal{B}$ -mesurable alors*

$$E(X|\mathcal{B}) = X. \quad (1.21)$$

2. **Espérance.** *Pour tout  $X \in L^1$ ,  $E(E(X|\mathcal{B})) = E(X)$ .*

3. **Linéarité.** *Pour tout  $\lambda \in \mathbb{R}$  et  $X, Y \in L^1$*

$$E(X + \lambda Y|\mathcal{B}) = E(X|\mathcal{B}) + \lambda E(Y|\mathcal{B}). \quad (1.22)$$

4. **Positivité.** *Pour tout  $X \in L^1$  tel que  $X \geq 0$  p.s.,  $E(X|\mathcal{B}) \geq 0$  p.s.*

5. **« Sortir ce qui est mesurable ».** *Soit  $X \in L^1$  et  $Y$  une variable aléatoire  $\mathcal{B}$ -mesurable. On suppose de plus que  $XY \in L^1$ . Alors*

$$E(XY|\mathcal{B}) = YE(X|\mathcal{B}). \quad (1.23)$$

6. **Inégalité de Jensen conditionnelle.** *Pour tout  $X \in L^1$  et toute fonction  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  borelienne convexe telle que  $\phi(X) \in L^1$ ,*

$$\phi(E(X|\mathcal{B})) \leq E(\phi(X)|\mathcal{B}). \quad (1.24)$$

7. **Contractante.** *Soit  $p \geq 1$  et  $X \in L^p$ . Alors  $E(X|\mathcal{B}) \in L^p$  et*

$$\|E(X|\mathcal{B})\|_p \leq \|X\|_p. \quad (1.25)$$

8. **Convergence monotone conditionnelle.** *Soit  $(X_n)_{n \geq 1}$  une suite croissante p.s. de variables aléatoires positives p.s. Alors*

$$\lim \uparrow E(X_n|\mathcal{B}) = E(\lim \uparrow X_n|\mathcal{B}) \quad \text{p.s.} \quad (1.26)$$

9. **Projections successives.** *Soit  $\mathcal{B}_1 \subset \mathcal{B}_2 \subset \mathcal{F}$  deux sous-tribus et  $X \in L^1$ . Alors*

$$E(E(X|\mathcal{B}_2)|\mathcal{B}_1) = E(E(X|\mathcal{B}_1)|\mathcal{B}_2) = E(X|\mathcal{B}_1). \quad (1.27)$$

*Éléments de preuve pour la Proposition 3.*

1. La variable  $X$  vérifie bien les deux points de la définition puisqu'elle est  $\mathcal{B}$ -mesurable et pour tout  $Z$   $\mathcal{B}$ -mesurable borné  $E(XZ) = E(XZ)$  (!).

2. Comme  $\Omega \in \mathcal{B}$ , d'après la définition

$$E(E(X|\mathcal{B})) = E(E(X|\mathcal{B})1_\Omega) = E(X1_\Omega) = E(X). \quad (1.28)$$

3. C'est immédiat en utilisant la linéarité de l'espérance.
4. On a déjà vu cette proposition dans  $L^2$ . La variable  $E(X|\mathcal{B})1_{\{E(X|\mathcal{B}) \leq 0\}}$  est négative p.s. et d'espérance positive (utiliser la définition de l'espérance conditionnelle). On en déduit qu'elle est nulle p.s. et donc que  $E(X|\mathcal{B}) \geq 0$  p.s. On conclut par symétrie.
5. Vérifions que  $YE(X|\mathcal{B})$  est bien l'espérance conditionnelle de  $XY$  sachant  $\mathcal{B}$ ,
- (a)  $YE(X|\mathcal{B})$  est bien  $\mathcal{B}$ -mesurable.
  - (b) Soit  $Z$   $\mathcal{B}$ -mesurable et bornée. Comme  $ZY$  est  $\mathcal{B}$ -mesurable et  $XZY \in L^1$ , en utilisant la Remarque 3,

$$E(ZYE(X|\mathcal{B})) = E(ZYX). \quad (1.29)$$

6. La preuve est similaire à celle de l'inégalité de Jensen. Dans le cas où  $\phi$  est  $\mathcal{C}^1$ , on a la preuve simple suivante. Comme  $\phi$  est au dessus de toutes ses tangentes,

$$\phi(E(X|\mathcal{B})) + \phi'(E(X|\mathcal{B}))(X - E(X|\mathcal{B})) \leq \phi(X) \quad p.s. \quad (1.30)$$

On passe à l'espérance conditionnelle de chaque côté. À gauche, le premier terme est inchangé car  $\mathcal{B}$ -mesurable. Pour le second, comme  $\phi'(E(X|\mathcal{B}))$  est  $\mathcal{B}$ -mesurable

$$E(\phi'(E(X|\mathcal{B}))(X - E(X|\mathcal{B}))|\mathcal{B}) = \phi'(E(X|\mathcal{B}))E((X - E(X|\mathcal{B}))|\mathcal{B}) = 0, \quad (1.31)$$

en utilisant les points 1 et 5 de cette proposition. On utilise enfin la positivité de l'espérance conditionnelle pour obtenir

$$\phi(E(X|\mathcal{B})) \leq E(\phi(X)|\mathcal{B}). \quad (1.32)$$

7. C'est une application de l'inégalité de Jensen conditionnelle avec la fonction convexe  $\phi : x \rightarrow |x|^p$ ,

$$\|E(X|\mathcal{B})\|_p = E(\phi(E(X|\mathcal{B})))^{1/p} \leq E(E(\phi(X)|\mathcal{B}))^{1/p} \leq \|X\|_p. \quad (1.33)$$

8. On note  $X = \lim \uparrow X_n$  au sens de la convergence p.s. Comme  $(Y_n)_{n \geq 0} = (E(X_n|\mathcal{B}))_{n \geq 0}$  est une suite croissante, on peut également définir sa limite  $Y$ . Pour montrer que  $Y = E(X|\mathcal{B})$  il nous faut vérifier deux choses

- (a)  $Y$  est bien  $\mathcal{B}$ -mesurable comme limite de variables  $\mathcal{B}$ -mesurables
- (b) Soit  $Z$   $\mathcal{B}$ -mesurable et bornée. Par le théorème de convergence monotone

$$E(XZ^+) = \lim E(X_n Z^+) = \lim E(Y_n Z^+) = E(Y Z^+), \quad (1.34)$$

et on a la même égalité avec  $Z^-$ . En décomposant  $Z = Z^+ - Z^-$ , on obtient  $E(XZ) = E(YZ)$ .

9. Dans le cas  $L^2$  où l'espérance conditionnelle coïncide avec la projection orthogonale, cette égalité est classique. Elle indique que si  $G \subset H$  sont deux sous espaces de  $L^2$  alors projeter sur  $H$  puis sur  $G$  est identique à projeter directement sur  $H$  (voir Figure 1.2). La preuve dans le cas général est l'objet d'un exercice dans la première feuille de TD.

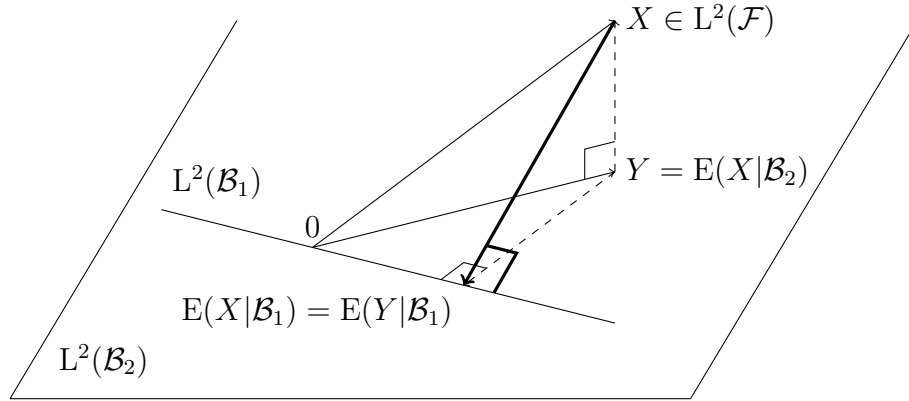


FIGURE 1.2: En gras la projection de  $X$  directement sur le petit espace  $L^2(\mathcal{B}_1)$  qui nous amène en  $E(X|\mathcal{B}_1)$ . En pointillés, les deux projections successives, d'abord de  $X$  sur  $L^2(\mathcal{B}_2)$  (on arrive en  $E(X|\mathcal{B}_2)$ ) puis de  $E(X|\mathcal{B}_2)$  sur  $L^2(\mathcal{B}_1)$  pour arriver en  $E(E(X|\mathcal{B}_2)|\mathcal{B}_1)$ . Les deux opérations nous amènent au même vecteur. (Rappelons que le dessin est faux...)

□

## 1.4 Cas particuliers

Dans cette section on étudie quelques cas importants pour lesquels on peut expliciter l'espérance conditionnelle.

1. Si  $X$  est indépendant de  $\mathcal{B}$  alors  $E(X|\mathcal{B}) = E(X)$ .

*Preuve.* Toujours deux points à vérifier :

- (a)  $E(X)$  est bien  $\mathcal{B}$ -mesurable (les variables constantes sont mesurables par rapport à n'importe quelle tribu puisque leur tribu réciproque ne contient que  $\emptyset$  et  $\Omega$ ),
- (b) pour tout  $Z$   $\mathcal{B}$ -mesurable borné, comme  $X$  et  $Z$  sont indépendants

$$E(XZ) = E(X)E(Z) = E(E(X)Z). \quad (1.35)$$

Ainsi si  $\mathcal{B} := \{\emptyset, \Omega\}$  est la tribu grossière, alors  $X$  est indépendant de  $\mathcal{B}$  et la meilleure approximation  $\mathcal{B}$ -mesurable que l'on puisse faire de  $X$  est  $E(X)$ . On retrouve là l'idée que la tribu grossière contient très peu d'informations.  $\square$

2.

**Proposition 4.** Soient  $X$  et  $Y$  deux variables aléatoires indépendantes et  $f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  borelienne tel que  $f(X, Y) \in L^1$ . Alors

$$E(f(X, Y)|Y) = g(Y), \quad (1.36)$$

*Preuve.* Commençons par rappeler un résultat important et très utile de théorie de la mesure :

**Lemme 2.** La variable aléatoire  $Z$  est  $\sigma(Y)$ -mesurable **si et seulement si** il existe une fonction  $h$  borelienne telle que  $Z = h(Y)$ .

On passe à la preuve avec...toujours deux points à vérifier :

- (a)  $g(Y)$  est bien  $\sigma(Y)$ -mesurable car fonction de  $Y$  (sens facile du lemme),
- (b) Soit  $Z$   $\sigma(Y)$ -mesurable borné. D'après le lemme (la réciproque), il existe  $h$  mesurable, et qu'on peut supposer borné, tel que  $Z = h(Y)$ . D'après le théorème de Fubini (version  $L^1$ ) :

$$E(f(X, Y)h(Y)) = \int f(x, y)h(y)dP_X \otimes dP_Y(x, y) \quad (1.37)$$

$$= \int h(y) \left( \int f(x, y)dP_X(x) \right) dP_Y(y) \quad (1.38)$$

$$= \int h(y)E(f(X, y))dP_Y(y) \quad (1.39)$$

$$= E(h(Y)g(Y)). \quad (1.40)$$

$\square$

3. **Le cas discret.** Lorsque la tribu  $\mathcal{B}$  est engendrée par une partition, ce qui est le cas quand  $\mathcal{B}$  est la tribu engendrée par une variable aléatoire discrète, la proposition suivante donne une forme explicite de l'espérance conditionnelle.

**Proposition 5.** Soit  $(A_i)_{i \in I}$  une partition dénombrable de  $\Omega$  et  $\mathcal{B} = \sigma(A_i, i \in I)$ . Alors

$$\mathbb{E}(X|\mathcal{B}) = \sum_{i \in I \text{ t.q. } P(A_i) > 0} \frac{\mathbb{E}(X1_{A_i})}{P(A_i)} 1_{A_i} \quad (1.41)$$

$$= \sum_{i \in I \text{ t.q. } P(A_i) > 0} \mathbb{E}(X|A_i) 1_{A_i}, \quad (1.42)$$

où  $\mathbb{E}(X|A_i)$  est une notation pour  $\frac{\mathbb{E}(X1_{A_i})}{P(A_i)}$ .

La preuve de cette proposition fait l'objet d'un exercice dans la première feuille de TD.

En particulier si  $Y$  est une variable aléatoire discrète à valeurs dans l'ensemble dénombrable  $(\alpha_i)_{i \in I}$  (supposés distincts) alors

$$\mathbb{E}(X|Y) = \sum_{i \in I \text{ t.q. } P(Y=i) > 0} \mathbb{E}(X|Y = \alpha_i) 1_{Y=\alpha_i}. \quad (1.43)$$

On notera que pour  $i \in I$ ,  $\mathbb{E}(X|Y = \alpha_i)$  peut se comprendre comme l'espérance de  $X$  par rapport à la probabilité

$$P(\cdot|Y = \alpha_i) := \frac{P(\cdot \cap \{Y = \alpha_i\})}{P(Y = \alpha_i)}. \quad (1.44)$$

Il est tentant dans ce cas particulier de définir la notion de probabilité conditionnelle mais on va s'en passer dans ce cours.

4. **Le cas à densité.** On suppose que le couple  $(X, Y)$  à valeurs dans  $\mathbb{R}^n \times \mathbb{R}^m$  admet une densité  $f_{X,Y}$  par rapport à la mesure de Lebesgue. On note  $f_X$  (défini pour tout  $x \in \mathbb{R}^n$  par  $f_X(x) = \int_{\mathbb{R}^m} f_{X,Y}(x, y) dy$ ) et  $f_Y$  les densités marginales. On définit la densité conditionnelle  $f_{X|Y=y}$  pour tout  $y$  tel que  $f_Y(y) > 0$  par

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \quad \text{pour tout } x \in \mathbb{R}^n. \quad (1.45)$$

On peut maintenant formuler le résultat suivant :

**Proposition 6.** Pour toute fonction réelle  $h$  mesurable tel que  $h(X) \in L^1$ ,

$$\mathbb{E}(h(X)|Y) = H(Y), \quad \text{où pour tout } y \quad (1.46)$$

$$H(y) = \int_{\mathbb{R}} h(x) f_{X|Y=y}(x) dx. \quad (1.47)$$

De nouveau, dans ce cas particulier simple, on pourrait définir facilement la notion de loi conditionnelle.

*Preuve.* De nouveau deux points à prouver

- (a)  $H(Y)$  est bien  $\sigma(Y)$ -mesurable car fonction de  $Y$ .
- (b) Soit  $Z$ ,  $\sigma(Y)$ -mesurable et bornée. Il existe  $\phi$  borelienne bornée telle que  $Z = \phi(Y)$ . On vérifie alors, en utilisant le théorème de Fubini  $L^1$ , que

$$\mathbb{E}(H(Y)\phi(Y)) = \int H(y)\phi(y)f_Y(y)dy \quad (1.48)$$

$$= \int \left( \int h(x) \frac{f_{X,Y}(x,y)}{f_Y(y)} dx \right) \phi(y) f_Y(y) dy \quad (1.49)$$

$$= \int h(x)\phi(y) f_{X,Y}(x,y) dx dy \quad (1.50)$$

$$= \mathbb{E}(h(X)\phi(Y)). \quad (1.51)$$

□



# Chapitre 2

## Martingales en temps discret

### 2.1 Définitions : filtrations et martingales

Dans tout le cours il faut s'habituer à penser une tribu comme une information (même si ça n'est pas facile!). Voici quelques exemples de traductions informelles :

1. dire que  $Y$  est  $\sigma(X)$ -mesurable c'est dire que, si on connaît  $X$ , on connaît également  $Y$ ,
2. dire que  $X$  est  $\mathcal{B}$ -mesurable c'est dire qu'il y a assez d'information dans la tribu  $\mathcal{B}$  pour connaître (« mesurer »)  $X$ ,
3. plus une tribu est grosse (i.e. contient d'événements) plus elle contient d'information. Ainsi si  $X$  est  $\mathcal{B}$ -mesurable et  $\mathcal{B}$  est inclus dans une tribu plus grosse  $\mathcal{G}$  alors  $X$  est également  $\mathcal{G}$ -mesurable.
4.  $\sigma(X)$  est l'information minimale dont on a besoin pour connaître  $X$  donc  $X$  est mesurable par rapport à une tribu  $\mathcal{B}$  si et seulement si  $\sigma(X) \subset \mathcal{B}$ .
5. ...

Il faut s'efforcer de poursuivre ce type de raisonnements dans toute la suite du chapitre.

**Définition 2.** Une **filtration** est une suite croissante pour l'inclusion  $(\mathcal{F})_{n \geq 0}$  de sous-tribus de  $\mathcal{F}$  :

$$\text{pour tout } n \geq 0, \quad \mathcal{F}_n \subset \mathcal{F}_{n+1}. \quad (2.1)$$

On peut comprendre une filtration comme une accumulation d'informations (on n'oublie rien!).

**Exemple 1** (Filtration naturelle d'un processus). Soit  $(X_n)_{n \geq 0}$  un processus. On définit pour tout  $n \geq 0$ ,

$$\mathcal{F}_n = \sigma(X_0, \dots, X_n), \quad (2.2)$$

la plus petite tribu rendant les variables  $X_i$ ,  $i = 1, \dots, n$  mesurables. La suite de tribus  $(\mathcal{F}_n)_{n \geq 0}$  est une filtration appelée **filtration naturelle du processus** ou **filtration engendrée par le processus**. Intuitivement,  $\mathcal{F}_n$  représente l'information minimale pour connaître la trajectoire du processus jusqu'au temps  $n$ .

**Définition 3.** Soit  $(\mathcal{F}_n)_{n \geq 0}$  une filtration et  $(X_n)_{n \geq 0}$  un processus. On dit que  $(X_n)_{n \geq 0}$  est

1. **adapté** si pour tout  $n \geq 0$ ,  $X_n$  est  $\mathcal{F}_n$ -mesurable,
2. **prévisible** si pour tout  $n \geq 0$ ,  $X_{n+1}$  est  $\mathcal{F}_n$ -mesurable et  $X_0$  est  $\mathcal{F}_0$ -mesurable.

Intuitivement, si un processus  $(X_n)_{n \geq 0}$  est adapté cela signifie que pour tout temps  $n$ , il y a assez d'information dans la tribu  $\mathcal{F}_n$  pour connaître la trajectoire du processus jusqu'au temps  $n$ . S'il est prévisible cela signifie qu'il y a assez d'information au temps  $n$  pour connaître la trajectoire jusqu'au temps (futur)  $n + 1$ .

**Exemple 2.**

Le processus  $(X_n)_{n \geq 0}$  est adapté à sa filtration naturelle. C'est même la plus petite (pour l'inclusion) filtration qui rende  $(X_n)_{n \geq 0}$  adapté.

Les processus prévisibles pourraient servir à modéliser des délits d'initié, c'est-à-dire le fait de participer à un marché en ayant connaissance d'informations qui permettent d'anticiper l'évolution des prix.

**Définition 4.** Soit  $(\mathcal{F}_n)_{n \geq 0}$  une filtration et  $(X_n)_{n \geq 0}$  un processus. On dit que  $(X_n)_{n \geq 0}$  est une **martingale** (ou  $(\mathcal{F}_n)$ -martingale) si on veut préciser la filtration) si

1. pour tout  $n \geq 0$ ,  $X_n$  est  $\mathcal{F}_n$ -mesurable (le processus est adapté),
2. pour tout  $n \geq 0$ ,  $X_n \in L^1$  (le processus est intégrable),
3. pour tout  $n \geq 0$ ,  $E(X_{n+1} | \mathcal{F}_n) = X_n$ .

On dit que  $(X_n)_{n \geq 0}$  est une **sous-martingale** si le dernier point est remplacé par

3. pour tout  $n \geq 0$ ,  $E(X_{n+1} | \mathcal{F}_n) \geq X_n$ .

On dit que  $(X_n)_{n \geq 0}$  est une **sur-martingale** si le dernier point est remplacé par

$$3. \text{ pour tout } n \geq 0, E(X_{n+1} | \mathcal{F}_n) \leq X_n.$$

On notera bien que la notion de martingale est relative à une filtration, qu'il faudra donc bien préciser lorsqu'il y a une ambiguïté. On note pour tout  $n \geq 1$ ,  $\Delta X_n := X_n - X_{n-1}$  le  $n$ -ième incrément du processus  $(X_n)_{n \geq 0}$ . On peut donc réécrire le point 3. dans la définition d'une martingale par :

$$3. \text{ pour tout } n \geq 1, E(\Delta X_n | \mathcal{F}_{n-1}) = 0.$$

**Remarque 4.** *On dit que le processus  $(X_n)_{n \geq 0}$  est intégrable si pour tout  $n \geq 0$ ,  $\|X_n\|_1 < \infty$ . C'est évidemment plus faible que de dire que le processus est borné dans  $L^1$  : il existe  $M \in \mathbb{R}$  tel que pour tout  $n \geq 0$ ,  $\|X_n\|_1 \leq M$ .*

**Remarque 5.** *Si  $(X_n)_{n \geq 0}$  est une martingale alors on prouve facilement à partir de la définition 4 que pour tout  $k \leq n$ ,*

$$E(X_n | \mathcal{F}_k) = X_k. \tag{2.3}$$

Voici 4 exemples de martingales qu'il faut connaître. Les trois premiers énoncés sont laissés en exercice. On considère une suite  $(Z_i)_{i \geq 1}$  de variables aléatoires i.i.d. et dans  $L^1$ . On note  $(\mathcal{F}_n)_{n \geq 0}$  la filtration engendrée par  $(Z_i)_{i \geq 1}$ .

1. **Martingale additive.** On suppose  $E(Z_1) = 0$ . La suite définie pour  $n \geq 1$  par,

$$X_n = \sum_{i=1}^n Z_i, \tag{2.4}$$

est une  $(\mathcal{F}_n)_{n \geq 0}$ -martingale.

2. **Martingale multiplicative.** On suppose  $E(Z_1) = 1$ . La suite définie pour  $n \geq 1$  par,

$$Y_n = \prod_{i=1}^n Z_i, \tag{2.5}$$

est une  $(\mathcal{F}_n)_{n \geq 0}$ -martingale.

3. **Martingale exponentielle.** Soit  $\alpha \in \mathbb{R}$ . On suppose  $E(e^{\alpha Z_1}) < +\infty$ . La suite définie pour  $n \geq 1$  par,

$$M_n = e^{\alpha X_n - n \ln E(e^{\alpha Z_1})}, \tag{2.6}$$

est une  $(\mathcal{F}_n)_{n \geq 0}$ -martingale.

4. **Martingale fermée.** Soit  $Z \in L^1$  et  $(\mathcal{B}_n)_{n \geq 0}$  une filtration. Alors le processus défini pour tout  $n \geq 1$  par

$$X_n = E(Z|\mathcal{B}_n) \quad (2.7)$$

est une  $(\mathcal{B}_n)_{n \geq 0}$ -martingale.

*Preuve du point 4.* Il faut vérifier les trois points de la définition :

- (a) le processus est clairement adapté par définition de l'espérance conditionnelle,
- (b) pour tout  $n \geq 1$ ,  $\|X_n\|_1 \leq \|Z\|_1 < +\infty$  car l'espérance conditionnelle est contractante. On en déduit que le processus est intégrable,
- (c) pour  $n \geq 1$ , en utilisant la propriété d'emboîtement de l'espérance conditionnelle,

$$E(X_{n+1}|\mathcal{B}_n) = E(E(Z|\mathcal{B}_{n+1})|\mathcal{B}_n) = E(Z|\mathcal{B}_n) = X_n. \quad (2.8)$$

□

**Proposition 7.** Soit  $(X_n)_{n \geq 0}$  une martingale (resp. une sur/sous-martingale). Alors  $(E(X_n))_{n \geq 0}$  est une suite constante (resp. décroissante / croissante).

*Preuve.* Il suffit de prendre l'espérance dans le troisième point de la définition d'une martingale. On obtient alors facilement que pour tout  $n \geq 0$ ,

$$E(X_{n+1}) = E(X_n) = E(X_0). \quad (2.9)$$

□

## 2.2 Théorèmes d'arrêt

Pour rendre la notion de martingale un peu plus intuitive on peut penser que ce processus décrit la fortune d'un joueur dans un **jeu équitable**. A chaque pas de temps  $(n, n+1)$  une manche du jeu a lieu et apporte au joueur le gain, positif ou négatif,  $\Delta X_{n+1}$ . Le processus  $(X_n)_{n \geq 0}$  apparait alors comme la somme des gains obtenus par le joueur. Le jeu est équitable au sens où pour tout  $n \geq 0$ ,  $E(\Delta X_{n+1}|\mathcal{F}_n) = 0$  c'est à dire qu'à chaque manche l'espérance du gain conditionnée par toute l'information passée est nulle.

Le but de cette partie est de comprendre l'espérance de la fortune obtenue dans un tel jeu lorsque le joueur adopte une stratégie admissible (il nous faudra définir ce que cela signifie). Par exemple si la stratégie consiste à s'arrêter

de jouer à un certain temps aléatoire  $T$  nous cherchons à étudier le gain associé qui est  $E(X_T)$ . Pour en rester à une formulation mathématique, on cherche des conditions telles que l'on puisse remplacer le temps déterministe  $n$  dans (2.9) par certains temps aléatoires.

### 2.2.1 Temps d'arrêt

Dans un premier temps nous ne considérons que des stratégies très simples qui consistent à jouer toutes les manches du jeu jusqu'à un certain temps  $T$  (qui peut être aléatoire) où on se retire définitivement du jeu. Ce type de stratégies sont modélisées par un temps aléatoire  $T$  à valeurs dans  $\mathbb{N}$ . Pour qu'on puisse considérer un tel temps comme une stratégie admissible il faut que pour tout temps  $n \in \mathbb{N}$  la décision de se retirer ou non du jeu soit prise uniquement avec l'information  $\mathcal{F}_n$  disponible à cet instant. Cela nous conduit à la définition suivante :

**Définition 5.** *La variable aléatoire  $T : (\Omega, \mathcal{F}, P) \rightarrow \mathbb{N} \cup \{+\infty\}$  est un  $(\mathcal{F}_n)_{n \geq 0}$ -temps d'arrêt si pour tout  $n \geq 0$ ,*

$$\{T = n\} \in \mathcal{F}_n. \quad (2.10)$$

On peut de façon équivalente remplacer la fin de la définition par

$$\{T \leq n\} \in \mathcal{F}_n. \quad (2.11)$$

On notera bien que, comme pour les martingales, il s'agit d'une notion qui est relative à une filtration qu'il faudra donc préciser lorsque il y a une ambiguïté. L'ensemble des temps d'arrêt est pour nous, dans un premier temps, l'ensemble des stratégies admissibles.

#### Exemple 3.

1. *Temps d'atteinte d'un borélien par un processus adapté. Soit  $(X_n)_{n \geq 0}$  un processus et  $(\mathcal{F}_n)_{n \geq 0}$  sa filtration naturelle. On considère un borélien  $B \in \mathcal{B}(\mathbb{R})$ . Le temps d'atteinte de  $B$  défini par*

$$T := \inf\{n \geq 0, X_n \in B\} \quad (2.12)$$

*est un  $(\mathcal{F}_n)_{n \geq 0}$ -temps d'arrêt.*

*Preuve.* Soit  $n \geq 0$ . On observe que

$$\{T \leq n\} = \{\exists k \leq n, X_k \in B\} = \cup_{k \leq n} \{X_k \in B\}. \quad (2.13)$$

Pour tout  $k \leq n$ ,  $\{X_k \in B\} \in \mathcal{F}_k$  car le processus est adapté. De plus  $\mathcal{F}_k \subset \mathcal{F}_n$  et une tribu est stable par union dénombrable donc  $\{T \leq n\} \in \mathcal{F}_n$ .  $\square$

2. Sous les mêmes hypothèses que le point précédent, le temps aléatoire défini par

$$T := \inf\{n \geq 0, X_{n+1} \in B\} \quad (2.14)$$

n'est pas en général un  $(\mathcal{F}_n)_{n \geq 0}$ -temps d'arrêt.

3. Sous les mêmes hypothèses que le point précédent, le temps aléatoire défini par

$$T := \inf\{n \geq 0, X_{n-3} \leq 0, X_{n-1} \geq 0\} \quad (2.15)$$

est un  $(\mathcal{F}_n)_{n \geq 0}$ -temps d'arrêt.

## 2.2.2 Processus arrêté et transformation de processus

**Définition 6.** Soit  $(X_n)_{n \geq 0}$  un processus et  $T$  un temps aléatoire (c'est-à-dire une variable aléatoire à valeurs dans  $\mathbb{N} \cup \{+\infty\}$ ). On définit le processus arrêté  $(X_{n \wedge T})_{n \geq 0}$  par :

$$\text{pour tout } \omega \in \Omega \text{ et tout } n \geq 0, \quad (X_{n \wedge T})(\omega) = X_{n \wedge T(\omega)}(\omega). \quad (2.16)$$

On a donc

$$X_{n \wedge T} = \begin{cases} X_n & \text{si } n \leq T, \\ X_T & \text{si } n \geq T. \end{cases} \quad (2.17)$$

De la même façon sur  $\{T < +\infty\}$ , on définit  $X_T(\omega) = X_{T(\omega)}(\omega)$ .

Le processus arrêté décrit la fortune du joueur qui a adopté la stratégie  $T$  : après le retrait du jeu au temps  $T$  le processus reste bloqué à  $X_T$ .

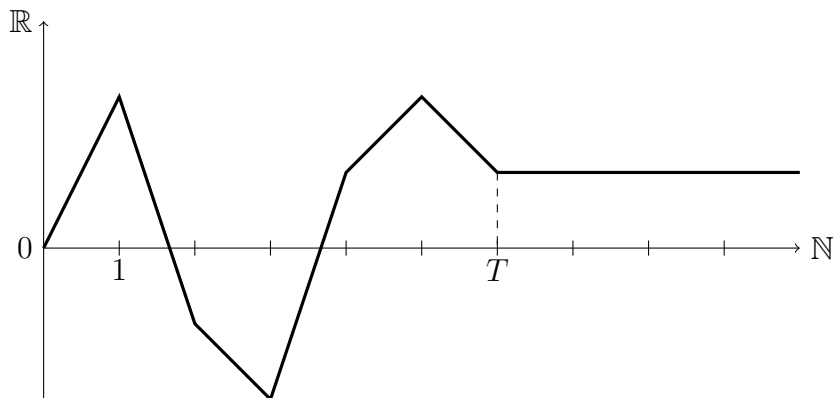


FIGURE 2.1: En gras, une réalisation de  $(X_{n \wedge T})_{n \geq 0}$ . On notera que le temps  $T$  à partir duquel le processus stationne est aléatoire et varie donc avec la réalisation.

**Proposition 8** (Un processus adapté arrêté par un temps d'arrêt est adapté). Soit  $(X_n)_{n \geq 0}$  un processus adapté à une filtration  $(\mathcal{F}_n)_{n \geq 0}$  et  $T$  un temps d'arrêt pour cette même filtration. Alors  $(X_{n \wedge T})_{n \geq 0}$  est également  $(\mathcal{F}_n)$ -adapté.

*Preuve.* Pour tout  $n \geq 0$

$$X_{n \wedge T} = \sum_{i=0}^n X_i 1_{T=i} + X_n 1_{T > n}. \quad (2.18)$$

Pour tout  $i \leq n$ ,  $X_i$  est  $\mathcal{F}_i$ -mesurable car le processus est adapté,  $\{T = i\} \in \mathcal{F}_i$  car  $T$  est un temps d'arrêt donc  $1_{T=i}$  est  $\mathcal{F}_i$ -mesurable. Comme pour tout  $i \leq n$ ,  $\mathcal{F}_i \subset \mathcal{F}_n$ , on obtient bien que la somme est  $\mathcal{F}_n$ -mesurable. Enfin  $\{T > n\} = \{T \leq n\}^c \in \mathcal{F}_n$  et on en déduit que  $X_n 1_{T > n}$  est  $\mathcal{F}_n$ -mesurable.  $\square$

**Proposition 9** (Un processus intégrable arrêté est intégrable). Soit  $(X_n)_{n \geq 0}$  un processus intégrable (i.e. pour tout  $n \geq 0$ ,  $X_n \in L^1$ ) et  $T$  un temps aléatoire (pas nécessairement un temps d'arrêt). Alors  $(X_{n \wedge T})_{n \geq 0}$  est également intégrable.

*Preuve.* Soit  $n \geq 0$ . D'après (2.18),

$$|X_{n \wedge T}| \leq \sum_{i=0}^n |X_i| + |X_n|, \quad (2.19)$$

donc  $E(|X_{n \wedge T}|) \leq \sum_{i=0}^n E(|X_i|) + E(|X_n|) < +\infty$ .  $\square$

**Proposition 10** (Une martingale arrêtée par un temps d'arrêt est une martingale). Soit  $(X_n)_{n \geq 0}$  une martingale relativement à une filtration  $(\mathcal{F}_n)_{n \geq 0}$  et  $T$  un temps d'arrêt pour cette même filtration. Alors  $(X_{n \wedge T})_{n \geq 0}$  est également une  $(\mathcal{F}_n)$ -martingale.

*Preuve.* Nous avons trois points à vérifier :

1. Le processus est adapté d'après la Proposition 8.
2. Le processus est intégrable d'après la Proposition 9.
3. Soit  $n \geq 0$ . On note que  $X_{(n+1) \wedge T} - X_{n \wedge T} = (X_{n+1} - X_n) 1_{\{n+1 \leq T\}}$ . Or, comme  $T$  est un temps d'arrêt,  $\{n+1 \leq T\} = \{T \leq n\}^c \in \mathcal{F}_n$  donc

$$E(X_{(n+1) \wedge T} - X_{n \wedge T} | \mathcal{F}_n) = 1_{\{n+1 \leq T\}} E((X_{n+1} - X_n) | \mathcal{F}_n) = 0, \quad (2.20)$$

où la dernière égalité est vraie car  $(X_n)_{n \geq 0}$  est une martingale.  $\square$

**Remarque 6.** On peut aisément adapter cette preuve pour étendre la proposition aux sur/sous-martingales.

Cette dernière proposition est par ailleurs un cas particulier d'un résultat plus général.

**Définition 7** (Transformation de processus). Soit  $(C_n)_{n \geq 1}$  et  $(X_n)_{n \geq 0}$  deux processus. On définit la transformation de  $X$  par  $C$ , notée  $((C \cdot X)_{n \geq 0})$  par

$$(C \cdot X)_0 = 0 \quad (2.21)$$

$$(C \cdot X)_n = \sum_{i=1}^n C_i \Delta X_i, \quad n \geq 1. \quad (2.22)$$

De façon équivalente on aurait pu définir le processus par ses incréments

$$\Delta(C \cdot X)_n = C_n \Delta X_n, \quad n \geq 1. \quad (2.23)$$

**Proposition 11.** Si  $(C_n)_{n \geq 1}$  est prévisible et dans  $L^\infty$  (i.e. pour tout  $n \geq 1$ ,  $\|C_n\|_\infty < +\infty$ ) et  $(X_n)_{n \geq 0}$  est une martingale relativement à une filtration  $(\mathcal{F}_n)_{n \geq 0}$  alors  $(C \cdot X)_{n \geq 0}$  est une martingale pour la même filtration. Si de plus  $C$  est positif, le résultat s'adapte aux sur/sous-martingales.

*Preuve.* La preuve est très proche de celle de la Proposition 10 ce qui n'est pas étonnant (voir la Remarque 7). On fixe  $n \geq 1$ .

1. Pour tout  $k \leq n$ ,  $C_k \Delta X_k$  est  $\mathcal{F}_k$ -mesurable donc  $\mathcal{F}_n$ -mesurable. On en déduit que  $(C \cdot X)_n$  est  $\mathcal{F}_n$ -mesurable comme somme finie de variables  $\mathcal{F}_n$ -mesurables.
2. Par ailleurs, pour tout  $k \leq n$ ,  $|C_k \Delta X_k| \leq \|C_k\|_\infty |\Delta X_k|$  p.s. donc  $C_k \Delta X_k$  est intégrable. Comme une somme finie de variables intégrables est encore intégrable, on en déduit que  $(C \cdot X)_n \in L^1$ .
3. Enfin puisque  $(C_n)_{n \geq 1}$  est prévisible,

$$\mathbb{E}(\Delta(C \cdot X)_n | \mathcal{F}_{n-1}) = \mathbb{E}(C_n \Delta X_n | \mathcal{F}_{n-1}) = C_n \mathbb{E}(\Delta X_n | \mathcal{F}_{n-1}) = 0. \quad (2.24)$$

Lorsque  $(C_n)_{n \geq 1}$  est positif, on peut facilement adapter le point 3 au sur/sous-martingale.  $\square$

**Remarque 7.** Cette notion permet d'étendre les stratégies admissibles en autorisant le joueur, à miser une quantité  $C_{n+1}$  à chaque manche  $(n, n+1)$ . Le gain associé au jeu  $(n, n+1)$  est alors donné par  $C_{n+1} \Delta X_{n+1}$  et on retrouve l'incrément du processus transformé...qui modélise donc bien la fortune associée à une telle stratégie. On notera qu'il est cohérent de demander que



le processus  $(C_n)_{n \geq 0}$  soit prévisible : il faut en effet choisir la mise  $C_{n+1}$  au début du tour  $(n, n+1)$  donc avec l'information  $\mathcal{F}_n$  disponible au temps  $n$ .

On retrouve la notion de processus arrêté en adoptant la stratégie qui consiste à miser 1 jusqu'au temps  $T$  (i.e. à la partie  $(T-1, T)$ ) puis à se retirer du jeu c'est-à-dire à miser 0 à partir du jeu  $(T, T+1)$ . Cette stratégie est décrite par le processus prévisible défini pour  $n \geq 0$  par

$$C_n = 1_{n \leq T}. \quad (2.25)$$

On vérifie alors que pour tout  $n \geq 0$ ,  $X_{n \wedge T} = (C \cdot X)_n$ . La Proposition 11 généralise donc bien la Proposition 10.

Si on revient à l'interprétation en terme de jeu équitable la Proposition 11 nous dit que toute stratégie de mise transforme un jeu équitable en...un autre jeu équitable.

### 2.2.3 Théorèmes d'arrêt

L'objectif de cette partie est de généraliser (2.9) à des temps d'arrêt. Sous quelles hypothèses a-t-on

$$E(X_T) = E(X_0)? \quad (2.26)$$

Est ce vrai pour toutes stratégies raisonnables (i.e. modélisées par les temps d'arrêt) ou peut-on espérer accroître sa fortune, en espérance, en adoptant une bonne stratégie dans un jeu pourtant équitable ?

On note tout d'abord que l'hypothèse  $T < +\infty$  p.s. est indispensable pour que la variable  $X_T$  soit bien définie. Mais ça n'est pas suffisant et l'égalité (2.26) peut être fautive sans hypothèses supplémentaires, comme le prouve l'exemple suivant :

**Exemple 4.** On définit la marche aléatoire simple  $(S_n)_{n \geq 0}$  sur  $\mathbb{Z}$  partant de  $x \in \mathbb{Z}$  de la façon suivante : on se donne une suite  $(Z_i)_{i \geq 1}$  de variables aléatoires i.i.d. telles que  $P(Z_1 = \pm 1) = 1/2$  puis on définit

$$\begin{aligned} S_0 &= x \quad \text{et pour } n \geq 1, \\ S_n &= x + \sum_{i=1}^n Z_i. \end{aligned} \quad (2.27)$$

Il s'agit bien d'une martingale pour sa filtration naturelle  $(\mathcal{F}_n)_{n \geq 0}$ . On considère dans cet exemple la marche partant de 0. Le temps d'atteinte  $T$  de 1 est un temps d'arrêt pour  $(\mathcal{F}_n)_{n \geq 0}$ . On montrera plus loin que  $T < +\infty$  p.s. (c'est une conséquence du caractère récurrent de la marche en dimension 1), on

peut donc définir  $S_T$  et il est clair que  $S_T = 1$  p.s. On a donc pour cet exemple

$$1 = E(S_T) \neq E(S_0) = 0. \quad (2.28)$$

**Théorème 4** (Théorème d'arrêt de Doob). *Soit  $(\mathcal{F}_n)_{n \geq 0}$  une filtration,  $(X_n)_{n \geq 0}$  une martingale et  $T$  un temps d'arrêt. On suppose de plus que l'on est dans l'un des trois cas suivants :*

1.  $T$  est borné,
2.  $E(T) < +\infty$  (i.e.  $T \in L^1$ ) et  $(\Delta X_n)_{n \geq 1}$  est borné dans  $L^\infty$ ,
3.  $T < +\infty$  p.s. et  $(X_{n \wedge T})_{n \geq 0}$  est borné dans  $L^\infty$ .

Alors  $X_T \in L^1$  et

$$E(X_T) = E(X_0). \quad (2.29)$$

*Preuve.* Notons tout d'abord que d'après la Proposition 10, pour tout  $n \geq 0$

$$E(X_{n \wedge T}) = E(X_{0 \wedge T}) = E(X_0). \quad (2.30)$$

1. Il existe un entier  $N$  tel que  $T \leq N$  p.s. D'après la proposition 9, en prenant  $n = N$ , comme  $N \wedge T = T$ , on obtient  $X_T = X_{n \wedge T} \in L^1$ . Par ailleurs, en prenant  $n = N$  dans (2.30), on obtient

$$E(X_T) = E(X_{N \wedge T}) = E(X_0). \quad (2.31)$$

3. On note  $M \in \mathbb{R}$  une borne du processus  $(X_{n \wedge T})_{n \geq 0}$  dans  $L^\infty$ . Comme  $T < +\infty$  p.s., la suite  $(X_{n \wedge T})_{n \geq 0}$  converge p.s. vers  $X_T$  (elle est même stationnaire p.s. mais attention le rang à partir duquel elle stationne dépend de la réalisation du processus). De plus les variables  $(X_{n \wedge T})_{n \geq 0}$  sont dominées par la variable aléatoire constante égale à  $M$  (qui est bien dans  $L^1$ ). On en déduit, d'après le théorème de convergence dominée que  $X_T$  est dans  $L^1$  et que la suite  $(E(X_{n \wedge T}))_{n \geq 0}$  converge vers  $E(X_T)$ . D'après (2.30) cette suite est par ailleurs constante égale à  $E(X_0)$  et on obtient le résultat.
2. La preuve du troisième point est identique à celle du second mais il faut adapter la domination. On note  $K \in \mathbb{R}$  une borne  $L^\infty$  du processus des incréments  $(\Delta X_n)_{n \geq 1}$ . Pour  $n \geq 0$ ,

$$|X_{n \wedge T}| \leq |X_0| + \sum_{k=1}^{n \wedge T} |\Delta X_k| \leq |X_0| + K(n \wedge T) \leq |X_0| + KT. \quad (2.32)$$

Comme  $T$  et  $X_0$  sont dans  $L^1$ ,  $|X_0| + KT$  est aussi dans  $L^1$  et on peut conclure comme pour le point précédent.

□

**Remarque 8.** 1. *Qu'en conclure en termes de jeu ? Que sous des hypothèses raisonnables il n'est pas possible d'augmenter sa fortune moyenne dans un jeu équitable en adoptant une stratégie admissible.*

2. *Il existe d'autres théorèmes d'arrêt. Cette liste de conditions suffisantes n'est pas exhaustive.*

3. *Revenons à l'Exemple 4. On peut en déduire que  $E(T) = +\infty$ . En effet si  $E(T) < +\infty$ , comme les incréments sont bornés, les hypothèses du second point du théorème d'arrêt seraient satisfaites et (2.29) serait vrai.*

*En terme de jeu on voit que la stratégie qui consiste à s'arrêter dans un pile ou face au premier instant où on a eu un succès de plus que d'échecs nécessite un temps d'espérance infini. On peut par ailleurs prouver (cf TD) que l'espérance de perte maximale associée à cette stratégie est également infinie...*

4. *Notons enfin que le théorème d'arrêt s'adapte aussi facilement pour les sous/sur-martingales.*

**Exemple 5** (calcul de la probabilité de sortie par un de deux bords d'un intervalle par la marche aléatoire simple sur  $\mathbb{Z}$ ). *On considère de nouveau la marche simple sur  $\mathbb{Z}$ ,  $(S_n)_{n \geq 0}$  partant de  $x \in \{0, \dots, N\}$  où  $N \geq 1$  est un entier. On rappelle que c'est une martingale pour sa filtration naturelle  $(\mathcal{F}_n)_{n \geq 0}$ . On cherche à calculer la probabilité que la marche sorte de l'intervalle par 0 plutôt que par  $N$ . On introduit le temps de sortie de l'intervalle*

$$T = \inf\{n \geq 0, S_n \in \{0, N\}\}. \quad (2.33)$$

*On peut montrer (cf TD) que  $T < +\infty$  p.s. et comme, de plus,  $|S_{n \wedge T}| \leq N$ , on peut utiliser le point 2 du théorème d'arrêt. On en déduit que  $E(S_0) = E(S_T)$ . Or  $E(S_0) = x$  et  $E(S_T) = 0 \times P(S_0 = 0) + NP(S_T = N) = NP(S_T = N)$ . On en déduit*

$$P(S_T = N) = \frac{x}{N} \quad \text{et} \quad P(S_T = 0) = \frac{N - x}{N}. \quad (2.34)$$

## 2.3 Théorèmes de convergence

Dans la dernière section de ce chapitre on s'intéresse au comportement asymptotique des martingales pour les modes de convergence p.s. ou  $L^p$  ( $p \geq 1$ ). Notons pour commencer qu'on ne peut pas espérer de résultats sans hypothèses additionnelles. Par exemple la marche aléatoire simple sur  $\mathbb{Z}$  ne converge ni p.s. ni dans  $L^1$  (simplement car elle n'est pas de Cauchy!).

### 2.3.1 Convergence presque sûre

Pour commencer cette partie nous énonçons dans le Lemme 3 une caractérisation de la convergence pour des suites réelles **déterministes**. Nous définissons tout d'abord le nombre de traversées ascendantes d'un intervalle :

**Définition 8.** Soit  $(x_i)_{i \geq 0}$  une suite à valeurs dans  $\mathbb{R}$  et  $a < b$  deux réels. On note pour  $n \geq 0$ ,  $U_n(a, b)$  le nombre de traversées ascendantes de l'intervalle  $[a, b]$  par la suite  $(x_i)_{i \geq 0}$  :  $U_n(a, b)$  est donc le plus grand entier  $k$  tel qu'il existe une suite strictement croissante d'indices  $i_1 < \dots < i_{2k}$  dans  $\{0, \dots, n\}$  vérifiant pour tout  $\ell \in \{1, \dots, k\}$ ,

$$x_{i_{2\ell-1}} < a \quad \text{et} \quad x_{i_{2\ell}} > b. \quad (2.35)$$

On note  $U(a, b) = \lim_{n \rightarrow +\infty} \uparrow U_n(a, b)$  le nombre total de traversées.

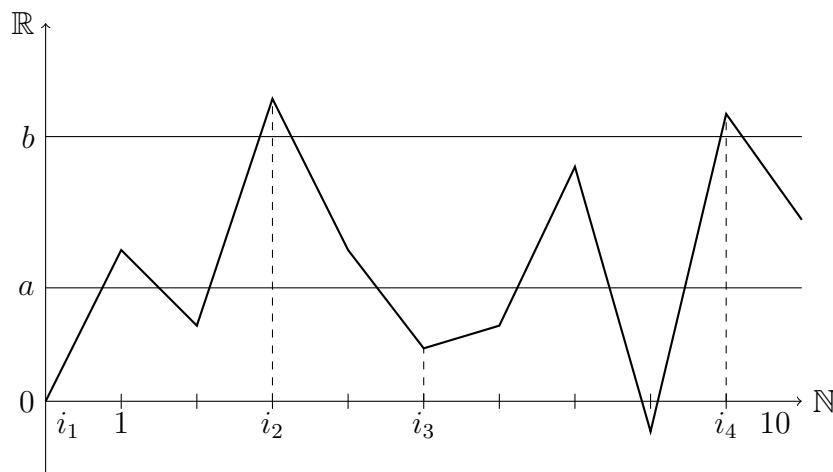


FIGURE 2.2: Sur cet exemple  $U_{10}(a, b) = 2$  car la suite traverse deux fois  $[a, b]$  avant le temps 10.

On commence donc par le résultat déterministe suivant

**Lemme 3.** Soit  $(x_i)_{i \geq 0}$  une suite telle que pour tout couple de rationnels  $p < q$ ,  $U(p, q) < +\infty$ . Alors  $(x_n)_{n \geq 0}$  admet une limite dans  $\mathbb{R} \cup \{\pm\infty\}$ .

*Preuve.* On note  $a = \liminf x_i$  et  $b = \limsup x_i$ . Supposons par l'absurde que  $a < b$ . Il existe donc  $p, q$  dans  $\mathbb{Q}^2$  tels que  $a < p < q < b$ . Par définition de la limite inférieure (et supérieure),  $|\{i \geq 0, x_i < p\}| = +\infty$  et  $|\{i \geq 0, x_i > q\}| = +\infty$ . On en déduit donc que  $U(p, q) = +\infty$  ce qui est absurde.

□

Pour obtenir une convergence p.s. nous allons utiliser ce lemme en établissant que, sous de bonnes hypothèses,

$$\mathbb{P} - \text{p.s.}, \quad \forall p, q \in \mathbb{Q} \text{ avec } p < q, \quad U(p, q) < +\infty. \quad (2.36)$$

L'étape la plus importante de cette partie consiste à établir un contrôle de l'espérance du nombre de traversées d'un intervalle par une martingale.

**Lemme 4.** *Soit  $(X_n)_{n \geq 0}$  une sur-martingale,  $a < b$  deux réels et  $n \in \mathbb{N}$ . Alors*

$$\mathbb{E}(U_n(a, b)) \leq \frac{\mathbb{E}(X_n - a)_-}{b - a}. \quad (2.37)$$

Pour comprendre ce lemme on peut garder en tête que les traversées ascendantes permettent de gagner dans le jeu en adoptant la stratégie suivante : je joue à partir du moment où la martingale passe sous  $a$  et jusqu'à ce qu'elle repasse au dessus de  $b$ . Comme le jeu est en moyenne défavorable ces traversées ne peuvent être nombreuses en espérance.

*Preuve.* Le processus suivant correspond à la stratégie que l'on vient de décrire :

$$\begin{aligned} C_1 &= 1_{\{X_0 < a\}} \quad \text{et, pour tout } n \geq 2, \\ C_n &= 1_{\{C_{n-1}=0, X_{n-1} < a\}} + 1_{\{C_{n-1}=1, X_{n-1} \leq b\}}. \end{aligned} \quad (2.38)$$

On peut vérifier que  $(C_n)_{n \geq 0}$  est prévisible pour  $(\mathcal{F}_n)_{n \geq 0}$ . Le processus de fortune associée à cette stratégie est donnée par la transformation suivante  $(W_n)_{n \geq 0} = ((C \cdot X)_n)_{n \geq 0}$ . On rappelle que  $W_0 = 0$  et pour  $n \geq 1$ ,  $\Delta W_n = C_n \Delta X_n$ . Comme  $(C_n)_{n \geq 0}$  est prévisible et positif,  $(W_n)_{n \geq 0}$  est une sur-martingale donc, d'après la Proposition 11, pour tout  $n \geq 0$ ,

$$\mathbb{E}(W_n) \leq \mathbb{E}(W_0) = 0. \quad (2.39)$$

Par ailleurs, par définition du nombre de traversées (voire Figure 2.3), pour tout  $n \geq 0$

$$W_n \geq (b - a)U_n(a, b) - (X_n - a)_-. \quad (2.40)$$

Finalement, en passant à l'espérance

$$\mathbb{E}(W_n) \geq (b - a)\mathbb{E}(U_n(a, b)) - \mathbb{E}((X_n - a)_-), \quad (2.41)$$

et, avec (2.39) on obtient le résultat.  $\square$

Nous pouvons maintenant énoncer le résultat de convergence presque sûre :

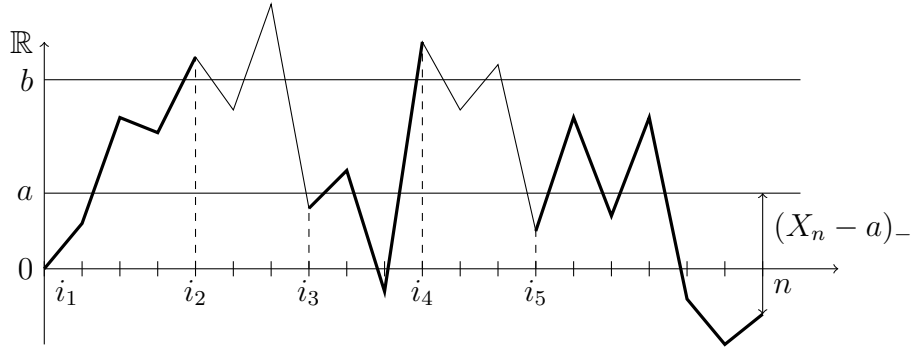


FIGURE 2.3: En gras les incréments de  $(X_n)_{n \geq 0}$  pris en compte pour construire  $(W_n)_{n \geq 0}$ . Chaque traversée apporte donc un gain supérieur à  $(b-a)$  sauf la dernière qui, au pire représente une perte inférieure à  $(X_n - a)_-$ .

**Théorème 5** (Convergence presque sûre de martingales). *Soit  $(X_n)_{n \geq 0}$  une sur-martingale **bornée dans  $L^1$**  (i.e.  $\sup_n E|X_n| < +\infty$ ). Alors **il existe  $X \in L^1$  telle que  $(X_n)_{n \geq 0}$  converge presque sûrement vers  $X$ .***

*Preuve.* Il existe  $C \in \mathbb{R}$  tel que pour tout  $n \geq 0$ ,  $E|X_n| < C$ . Soit  $a < b$  deux réels. D'après le Lemme 4, pour tout  $n \geq 0$

$$E(U_n(a, b)) \leq \frac{E(X_n - a)_-}{b - a} \leq \frac{E(|X_n| + |a|)}{b - a} \leq \frac{C + |a|}{b - a}, \quad (2.42)$$

et donc, d'après le théorème de convergence monotone,

$$E(U(a, b)) \leq \frac{C + |a|}{b - a} < +\infty. \quad (2.43)$$

On en déduit que  $U(a, b) < +\infty$  p.s. On a prouvé à ce stade

$$\forall p, q \in \mathbb{Q}, \quad \text{P - p.s.}, \quad U(p, q) < +\infty. \quad (2.44)$$

Comme  $\mathbb{Q}^2$  est dénombrable on peut échanger le quantificateur et le p.s. et on obtient

$$\text{P - p.s.}, \quad \forall p, q \in \mathbb{Q}, \quad U(p, q) < +\infty. \quad (2.45)$$

En utilisant le Lemme 3 on en déduit que  $(X_n)_{n \geq 0}$  converge p.s. vers  $X \in \mathbb{R} \cup \{\pm\infty\}$ . Il reste à montrer que  $X \in L^1$ . Comme  $|X| = \lim_{p.s.} |X_n| = \liminf_{p.s.} |X_n|$ , on peut utiliser le lemme de Fatou

$$E|X| \leq \liminf E|X_n| \leq C, \quad (2.46)$$

et on a bien  $X \in L^1$ . □

**Remarque 9.** 1. Le théorème assure le même résultat pour une martingale ou une sous-martingale.

2. La convergence n'a pas forcément lieu dans  $L^1$ . On peut par exemple définir la martingale multiplicative  $(X_n)_{n \geq 0}$  associée à la suite i.i.d.  $(Z_i)_{i \geq 0}$  de loi  $\frac{1}{2}\delta_0 + \frac{1}{2}\delta_2$ . On a alors convergence presque sûre de  $(X_n)_{n \geq 0}$  vers 0 mais pas convergence  $L^1$  puisque pour tout  $n \geq 0$ ,  $E(X_n) = 1$ .

Du théorème précédent on déduit le

**Corollaire 2.** Une sous-martingale majorée converge presque sûrement. Une sur-martingale minorée converge presque sûrement.

*Preuve.* Soit  $(X_n)_{n \geq 0}$  une surmartingale minorée par  $m \in \mathbb{R}$ . Pour tout  $n \geq 0$

$$E|X_n - m| = E(X_n - m) \leq E|X_0| + |m|, \quad (2.47)$$

et on en déduit

$$E|X_n| \leq E|X_n - m| + |m| \leq E|X_0| + 2|m|. \quad (2.48)$$

La surmartingale  $(X_n)_{n \geq 0}$  est donc bornée dans  $L^1$  et on en déduit la convergence p.s. vers une variable de  $L^1$ .  $\square$

### 2.3.2 Convergence $L^2$

**Théorème 6.** Soit  $(X_n)_{n \geq 0}$  une martingale bornée dans  $L^2$ . Alors **il existe**  $X \in L^2$  **tel que**  $(X_n)_{n \geq 0}$  **converge dans**  $L^2$  **et p.s. vers**  $X$ .

*Preuve.* La convergence p.s. est assurée par le théorème précédent puisque la martingale est bornée dans  $L^1$ . Pour la convergence  $L^2$  nous allons montrer que la suite est de Cauchy, ce qui est suffisant puisque  $L^2$  est complet. Il existe  $C \in \mathbb{R}$  tel que pour tout  $n \geq 0$ ,  $\|X_n\|_2^2 \leq C$ .

On commence en montrant que les incréments d'une martingale de carré intégrable sont orthogonaux dans  $L^2$ . Pour tout couple d'entiers  $1 \leq i < j$ ,

$$\langle \Delta X_i, \Delta X_j \rangle = E(\Delta X_i \Delta X_j) = E(E(\Delta X_i \Delta X_j | \mathcal{F}_i)) \quad (2.49)$$

$$= E(\Delta X_i E(\Delta X_j | \mathcal{F}_i)) \quad (\text{car } \Delta X_i \text{ est } \mathcal{F}_i\text{-mesurable}) \quad (2.50)$$

$$= 0 \quad (\text{car } (X_n)_{n \geq 0} \text{ est une martingale}). \quad (2.51)$$

Pour tous entiers  $n < p$ , on a donc par le théorème de Pythagore,

$$\|X_p - X_n\|_2^2 = \left\| \sum_{i=n+1}^p \Delta X_i \right\|_2^2 = \sum_{i=n+1}^p \|\Delta X_i\|_2^2, \quad (2.52)$$

et on déduit que  $\|X_p - X_n\|_2^2 \leq \sum_{i=n+1}^{+\infty} \|\Delta X_i\|_2^2$ . Or la série de terme général  $\|\Delta X_i\|_2^2$  est convergente car

$$\sum_{i=1}^n \|\Delta X_i\|_2^2 = \|X_n - X_0\|_2^2 \leq 2(\|X_n\|_2^2 + \|X_0\|_2^2) \leq 4C. \quad (2.53)$$

On en déduit que  $(X_n)_{n \geq 0}$  est de Cauchy dans  $L^2$  et donc converge dans  $L^2$ .  $\square$

**Remarque 10.** 1. Rappelons que la convergence  $L^2$  est plus forte que la convergence  $L^1$ .

2. Rappelons aussi que l'hypothèse « bornée dans  $L^1$  » est insuffisante pour obtenir la convergence  $L^1$ .

3. Pour étudier en général la convergence  $L^p$ ,  $p > 1$ , on a besoin de l'inégalité de Doob (cf TD).



# Chapitre 3

## Chaînes de Markov

Ajouter la déf d'une matrice sto. Ecrire le processus canonique explicitement.

### 3.1 Définitions et équations de Chapman-Kolmogorov

#### 3.1.1 Définitions

On travaille toujours sur un espace de probabilité  $(\Omega, \mathcal{F}, P)$ . On considère aussi dans ce chapitre un ensemble  $E$  dénombrable (fini ou infini) que l'on appelle **espace d'états** dans lequel le processus prend ses valeurs.

**Définition 9.** Soit  $(X_n)_{n \geq 0}$  un processus à valeurs dans  $E$ . On dit qu'il est **markovien** si pour tout  $n \geq 0$  et tout  $(x_0, \dots, x_{n-1}, x, y) \in E^{n+2}$  tel que  $P(X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x) > 0$ ,

$$P(X_{n+1} = y | X_n = x, \dots, X_0 = x_0) = P(X_{n+1} = y | X_n = x) \quad (3.1)$$

Si de plus il existe  $Q : E \times E \rightarrow [0, 1]$  tel que

$$P(X_{n+1} = y | X_n = x) = Q(x, y) \quad (3.2)$$

on dit que la chaîne est **homogène en temps**.

L'application  $Q$  de la définition s'appelle la **matrice de transition** du processus (même si ça n'est pas vraiment une matrice lorsque  $E$  est dénombrable infini). Pour tout  $x, y \in E$ ,  $Q(x, y)$  est la probabilité lorsque la chaîne est en  $x$  (quelque soit le temps) de sauter en  $y$ .

On peut résumer informellement cette propriété en disant que la loi du futur d'un processus markovien sachant tout le passé est identique à la loi du

futur sachant uniquement la position présente. Il n'y a donc pas plus d'information dans l'ensemble de la trajectoire passée que dans la seule position présente. La probabilité conditionnelle d'un saut à un certain temps exprimée dans le terme de gauche de (3.1) est donc fonction uniquement de  $x, y$  et  $n$  (dans le terme de droite).

Si de plus cette loi ne dépend pas non plus du temps présent mais uniquement de la position présente (la probabilité conditionnelle est alors uniquement fonction de  $x$  et  $y$  dans le terme de droite de (3.2)), on dit que la chaîne est homogène. Ainsi, si le processus passe en  $x \in E$  à deux temps distincts, la probabilité qu'il saute en  $y$  au temps suivant est la même dans les deux cas.

---

**Dans la suite, on ne considérera que (même si j'oublie de le préciser !) des chaînes de Markov homogènes.**

---

**Exemple 6.** *Il est commode de représenter une chaîne de Markov par un graphe dont les sommets sont indexés par  $E$  et les arrêtes correspondent aux transitions non nulles de la matrice  $Q$ . Par exemple dans le cas où  $E = \{1, 2, 3, 4\}$  et*

$$Q = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1/3 & 0 & 2/3 \\ 3/4 & 0 & 0 & 1/4 \end{pmatrix}, \quad (3.3)$$

On obtient le graphe suivant :

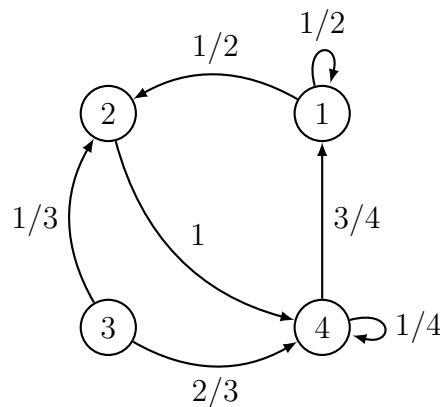


FIGURE 3.1: Le graphe correspondant à la matrice  $Q$  définie en (3.3).

La matrice  $Q$  est **stochastique**, ce qui signifie que la somme des éléments

de n'importe quelle ligne est 1. En effet pour tout  $x \in E$

$$\sum_{y \in E} Q(x, y) = \sum_{y \in E} P(X_1 = y | X_0 = x) = 1. \quad (3.4)$$

**Remarque 11.** *S'il existe  $Q : E \times E \rightarrow [0, 1]$  tel que pour tout  $n \geq 0$  et tout  $(x_0, \dots, x_{n-1}, x, y) \in E^{n+2}$  tel que  $P(X_0 = x_0, \dots, X_n = x) > 0$ ,*

$$P(X_{n+1} = y | X_n = x, \dots, X_0 = x_0) = Q(x, y), \quad (3.5)$$

*alors le processus  $(X_n)_{n \geq 0}$  est une chaîne de Markov homogène de matrice de transition  $Q$ . La différence avec la définition est ténue : il s'agit seulement de vous faire remarquer que (3.5) suffit pour que  $P(X_{n+1} = y | X_n = x) = Q(x, y)$ .*

**Exemple 7 (Marche aléatoire simple sur  $\mathbb{Z}^d$ ).** *Un exemple très important de processus markovien homogène est la marche aléatoire simple sur  $\mathbb{Z}^d$ ,  $d \geq 1$ . On dit que  $x$  et  $y$  dans  $\mathbb{Z}^d$  sont voisins (et on note  $x \sim y$ ) si  $\|x - y\|_1 = 1$ . Il s'agit donc de la notion intuitive de voisins sur le réseau. La marche simple est le processus markovien, partant de 0, et sautant à chaque pas de temps vers un des  $2d$  voisins du site où elle se trouve. Ce processus est caractérisé par la matrice de transition*

$$Q(x, y) = \frac{1}{2d}, \quad \text{pour tout } x \sim y \in \mathbb{Z}^d. \quad (3.6)$$

**Exemple 8.** *Un processus qui interagit non trivialement avec son passé à plus d'un pas est non markovien. Le processus à valeurs dans  $\{-1, +1\}$  défini par  $X_0 = X_1 = 1$  et pour  $n \geq 1$ ,*

$$X_{n+1} = X_n X_{n-1} \xi_{n+1}, \quad (3.7)$$

*où les variables  $(\xi_n)_{n \geq 2}$  sont i.i.d. de loi  $P(\xi = 1) = 1 - P(\xi = -1) = 2/3$  et indépendantes de  $X_0$  et  $X_1$ , n'est pas markovien. On note tout d'abord que pour tout  $n \geq 2$ ,  $X_n$  est  $\sigma(\xi_2, \dots, \xi_n)$  mesurable. On en déduit que pour  $n \geq 1$ ,*

$$\begin{aligned} P(X_{n+1} = 1 | X_n = 1, X_{n-1} = 1) &= P(\xi_{n+1} = 1 | X_n = 1, X_{n-1} = 1) \\ &= P(\xi_{n+1} = 1) \\ &= \frac{2}{3}, \end{aligned} \quad (3.8)$$

où on a utilisé que  $\xi_{n+1}$  est indépendant de  $(X_{n-1}, X_n)$ . Le même calcul montre que pour  $n \geq 2$ ,

$$\begin{aligned} P(X_{n+1} = 1 | X_n = 1, X_{n-1} = -1) &= P(\xi_{n+1} = -1 | X_n = 1, X_{n-1} = -1) \\ &= P(\xi_{n+1} = -1) \\ &= \frac{1}{3}, \end{aligned} \tag{3.9}$$

et on en déduit que  $(X_n)_{n \geq 0}$  n'est pas markovien.

**Exemple 9.** Un processus fonction d'un processus markovien n'est pas forcément markovien. Par exemple si  $(X_n)_{n \geq 0}$  est la marche aléatoire simple sur  $\mathbb{Z}$ , le processus  $(Y_n)_{n \geq 0} = (1_{X_n \geq 0})_{n \geq 0}$  n'est pas markovien. En effet

$$P(Y_3 = 1 | Y_2 = 1, Y_1 = 0) = P(X_3 = 1 | X_2 = 0, X_1 = -1) = \frac{1}{2}, \tag{3.10}$$

alors que

$$P(Y_3 = 1 | Y_2 = 1, Y_1 = 1) = P(X_3 \geq 0 | X_1 = 1) = \frac{3}{4}. \tag{3.11}$$

### 3.1.2 Récurrences aléatoires

Les récurrences aléatoires constituent une définition alternative et souvent commode des chaînes de Markov.

**Proposition 12.** Soit  $(\xi_n)_{n \geq 1}$  une suite de variables aléatoires i.i.d. à valeurs dans un ensemble  $I$  (cette suite s'appelle la suite des **innovations**),  $X_0$  une variable aléatoire à valeurs dans  $E$  indépendante de la suite des innovations, et enfin une fonction  $f : E \times I \rightarrow E$ . Alors la récurrence aléatoire définie par

$$X_{n+1} = f(X_n, \xi_{n+1}), \quad \text{pour tout } n \geq 0, \tag{3.12}$$

est une chaîne de Markov homogène de matrice de transition :

$$Q(x, y) = P(f(x, \xi) = y), \quad x, y \in E. \tag{3.13}$$

*Preuve.* Soit  $n \geq 0$  et  $(x_0, \dots, x_{n-1}, x, y) \in E^{n+2}$  tel que  $P(X_0 = x_0, \dots, X_n = x) > 0$ . On a

$$P(X_{n+1} = y | X_n = x, \dots, X_0 = x_0) = \frac{P(f(x, \xi_{n+1}) = y, X_n = x, \dots, X_0 = x_0)}{P(X_n = x, \dots, X_0 = x_0)}. \tag{3.14}$$

L'événement  $\{f(x, \xi_{n+1}) = y\} \in \sigma(\xi_{n+1})$  est indépendant de  $\{X_n = x, \dots, X_0 = x_0\} \in \sigma(X_0, \xi_1, \dots, \xi_n)$  donc,

$$P(X_{n+1} = y | X_n = x, \dots, X_0 = x_0) = P(f(x, \xi_{n+1}) = y), \quad (3.15)$$

et on conclut avec la Remarque 11.  $\square$

Nous allons maintenant établir la réciproque : pour toute chaîne de Markov homogène  $(X_n)_{n \geq 0}$  il existe une récurrence aléatoire qui a même loi que  $(X_n)_{n \geq 0}$ .

**Proposition 13.** *Soit  $(X_n)_{n \geq 0}$  une chaîne de Markov homogène sur  $E = \{y_i, i \geq 1\}$  de matrice de transition  $Q$ . On considère une suite  $(\xi_n)_{n \geq 0}$  de variables aléatoires i.i.d. uniformes sur  $[0, 1]$ . On définit de plus la fonction*

$$f : E \times I \rightarrow E \\ (x, u) \mapsto y_n \text{ si } u \in \left[ \sum_{i=1}^{n-1} Q(x, y_i), \sum_{i=1}^n Q(x, y_i) \right[$$

Alors la récurrence aléatoire définie par  $f, \xi$  et  $X_0$  est markovienne de matrice de transition  $Q$  (et loi initiale  $X_0$ ).

*Preuve.* Le caractère markovien est une conséquence de la proposition précédente. Il suffit donc de vérifier que la matrice de transition est la bonne. Or pour tout  $x \in E$  et  $n \geq 1$ ,

$$P(f(x, \xi) = y_n) = P(\xi \in \left[ \sum_{i=1}^{n-1} Q(x, y_i), \sum_{i=1}^n Q(x, y_i) \right]) = Q(x, y_n), \quad (3.16)$$

car  $Q$  est stochastique donc  $0 \leq \sum_{i=1}^{n-1} Q(x, y_i) \leq \sum_{i=1}^n Q(x, y_i) \leq 1$ .  $\square$

Cette proposition permet aussi de montrer l'existence de chaînes de Markov : pour toute fonction  $Q : E \times E \rightarrow [0, 1]$  stochastique et toute loi initiale  $\mu$ , il existe bien un processus markovien de matrice  $Q$  et de loi initiale  $\mu$  puisque la Proposition 13 permet de le construire explicitement. Nous verrons dans la section suivante que cela signifie que la récurrence aléatoire a même loi que  $(X_n)_{n \geq 0}$ .

### 3.1.3 Équations de Chapman Kolmogorov

Les **équations de Chapman Kolmogorov** décrivent la loi d'une chaîne de Markov à partir de sa matrice de transition et de sa condition initiale.

**Proposition 14** (Probabilité de suivre une trajectoire fixée). *Soit  $(X_n)_{n \geq 0}$  une chaîne de Markov homogène de matrice de transition  $Q$ . On note  $\mu$  la loi de  $X_0$ . Alors, tout entier  $n \geq 0$  et tous  $x_0, \dots, x_n$  dans l'espace d'états  $E$ ,*

$$P(X_n = x_n, \dots, X_0 = x_0) = \mu(x_0) \prod_{i=0}^{n-1} Q(x_i, x_{i+1}). \quad (3.17)$$

*Preuve.* Par récurrence. L'initialisation ne pose pas de problème et si l'hypothèse est satisfaite au rang  $n \in \mathbb{N}$  alors

$$P(X_{n+1} = x_{n+1}, \dots, X_0 = x_0) \quad (3.18)$$

$$= P(X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_0 = x_0) P(X_n = x_n, \dots, X_0 = x_0) \quad (3.19)$$

$$\stackrel{\text{Markov}}{=} Q(x_n, x_{n+1}) P(X_n = x_n, \dots, X_0 = x_0) \quad (3.20)$$

$$\stackrel{\text{HR}_n}{=} Q(x_n, x_{n+1}) \mu(x_0) \prod_{i=0}^{n-1} Q(x_i, x_{i+1}) \quad (3.21)$$

$$= \mu(x_0) \prod_{i=0}^n Q(x_i, x_{i+1}). \quad (3.22)$$

□

Cette proposition établit donc que la loi d'une chaîne de Markov homogène est entièrement déterminée par

1. la loi  $\mu$  de la condition initiale,
2. la matrice de transition  $Q$ .

Autrement dit, deux chaînes de Markov qui ont même matrice de transition et même loi initiale ont même loi.

**Dans la suite du cours, à chaque fois que l'on introduira un processus markovien  $(X_n)_{n \geq 0}$  de matrice  $Q$  sur notre espace de probabilité  $(\Omega, \mathcal{F}, P)$ , on supposera qu'il existe une collection de probabilités  $(P_x)_{x \in E}$  sur  $(\Omega, \mathcal{F})$  telle que pour tout  $x \in E$ , sous  $P_x$ ,  $(X_n)_{n \geq 0}$  est une chaîne de Markov de matrice  $Q$  et de condition initiale  $x$ , i.e.  $P_x(X_0 = x) = 1$ . On supposera plus généralement que, pour toute probabilité  $\mu$  sur  $E$ , il existe une probabilité  $P_\mu$  sur  $(\Omega, \mathcal{F})$  telle que, sous  $P_\mu$ ,  $(X_n)_{n \geq 0}$  est une chaîne de Markov de matrice  $Q$  et  $X_0$  a pour loi  $\mu$  i.e. pour tout  $x \in E$ ,  $P_\mu(X_0 = x) = \mu(x)$ .**

Dans le cas particulier où  $\mu = \delta_x$ , on a  $P_\mu = P_x$  et on préférera bien sûr

cette dernière notation. Enfin on utilisera la notation  $E_x$  pour l'espérance sous  $P_x$  et  $E_\mu$  pour celle sous  $P_\mu$ . Par ailleurs, notons que pour tout  $x, y$  dans  $E$ ,

$$Q(x, y) = P_x(X_1 = y | X_0 = x) = P_x(X_1 = y).$$

Enfin, les notations peuvent sembler redondantes puisque si  $X_0$  a pour loi  $\mu$  sous  $P$  alors  $(X_n)_{n \geq 0}$  a même loi sous  $P$  et sous  $P_\mu$ . Cela nous permet cependant d'alléger un peu les notations et d'utiliser  $P$  lorsqu'on n'a pas besoin de mentionner la condition initiale.

Celles et ceux qui souhaitent comprendre pourquoi on peut faire la supposition à l'origine de ce paragraphe peuvent consulter les paragraphes consacrés au processus canonique dans les polys de Joseph Lehec<sup>1</sup> ou de Zhan Shi<sup>2</sup>.

*Notations matricielles.* On considère une chaîne de Markov homogène de matrice  $Q$ . Par analogie avec le cas où  $E$  est de cardinal fini, on introduit les **notations matricielles** :

1. **Noyau (produit à droite).** On note  $\mathcal{F}(E, \mathbb{R})$  l'ensemble des fonctions bornées de  $E$  dans  $\mathbb{R}$ . On appelle noyau l'opérateur

$$Q : \begin{array}{ccc} \mathcal{F}(E, \mathbb{R}) & \rightarrow & \mathcal{F}(E, \mathbb{R}) \\ f & \mapsto & Qf \end{array} \quad (3.23)$$

où  $Qf$  est défini pour tout  $x \in E$  par

$$Qf(x) = \sum_{y \in E} Q(x, y) f(y). \quad (3.24)$$

On a aussi la formulation équivalente

$$Qf(x) = \sum_{y \in E} f(y) P_x(X_1 = y) \quad (3.25)$$

$$= E_x(f(X_1)). \quad (3.26)$$

Cet opérateur caractérise la loi de  $X_1$  sous  $P_x$  et donc la loi de la chaîne. Dans le cas où  $|E| < +\infty$  on retrouve bien sûr le produit matriciel entre la matrice  $Q$  et le vecteur colonne  $(f(x))_{x \in E}$ .

2. **Produit à gauche.** Si  $\mu$  désigne une mesure ou une probabilité sur  $E$ , on définit la mesure  $\mu Q$  sur  $E$  par

$$\mu Q(x) = \sum_{y \in E} \mu(y) Q(y, x) \quad (3.27)$$

$$= \sum_{y \in E} \mu(y) P_y(X_1 = x). \quad (3.28)$$

---

1. <https://www.ceremade.dauphine.fr/~lehec/processus/poly2017.pdf>

2. <https://www.lpsm.paris/pageperso/levy/4M011.Poly.pdf>

A nouveau dans le cas où  $|E| < +\infty$  on retrouve le produit matriciel classique entre le vecteur ligne  $\mu$  et la matrice carré  $Q$ . Comme  $Q$  est stochastique, la mesure  $\mu Q$  a même masse que la mesure  $\mu$  :

$$\begin{aligned}\mu Q(E) &= \sum_{x \in E} \sum_{y \in E} \mu(y) P_y(X_1 = x) \\ &= \sum_{y \in E} \mu(y) \sum_{x \in E} P_y(X_1 = x) \\ &= \sum_{y \in E} \mu(y) = \mu(E).\end{aligned}\tag{3.29}$$

Lorsque  $\mu$  est une probabilité, on note que pour tout  $x \in E$ ,  $\mu Q(x) = P_\mu(X_1 = x)$ . La probabilité  $\mu Q$  est donc la loi de  $X_1$  sous  $P_\mu$ .

Lorsque  $\mu$  est une mesure mais pas une probabilité, cette interprétation probabiliste ne fonctionne plus. On peut en revanche imaginer que chaque site  $x \in E$  reçoit une masse  $\mu(x)$  de sable. L'action de la chaîne répartit le tas présent en  $x$  sur les sites voisins en le découpant selon  $Q(x, \cdot)$ . La mesure  $\mu Q$  représente alors la masse présente en chaque site lorsqu'on a fini toutes les répartitions pour chaque site. De même,  $\mu Q^n$  représente la répartition du sable après  $n$  opérations de la chaîne.

3. **Produit.** On définit la matrice  $(Q^2(x, y))_{(x, y) \in E^2}$  par

$$Q^2(x, y) = \sum_{z \in E} Q(x, z) Q(z, y)\tag{3.30}$$

et par récurrence pour tout  $n \geq 2$

$$Q^{n+1}(x, y) = \sum_{z \in E} Q^n(x, z) Q(z, y).\tag{3.31}$$

Et on retrouve à nouveau le produit matriciel dans le cas où  $|E| < +\infty$ . On veillera bien à ne pas confondre  $Q^n(x, y)$ , l'entrée  $(x, y)$  de la matrice  $Q^n$ , avec  $(Q(x, y))^n$ , la puissance  $n$ -ième de l'entrée  $(x, y)$  de la matrice  $Q$ .

**Proposition 15.** Soit  $(X_n)_{n \geq 0}$  une chaîne de Markov homogène de matrice  $Q$ . Pour tout  $n \geq 0$ , on note  $\mu_n$  la loi de  $X_n$ . Alors :

1. (**Loi des marginales**) Pour tout  $n \geq 0$ ,  $\mu_n = \mu_{n-1} Q = \mu_0 Q^n$ .
2. (**Transitions à  $n$  pas**) Pour tout  $x, y$  dans  $E$ ,  $P_x(X_n = y) = Q^n(x, y)$ .



3. (*caractérisation de la loi à n pas*) Pour toute fonction  $h : E \rightarrow \mathbb{R}$  bornée

$$\mathbb{E}_x(h(X_n)) = Q^n h(x). \quad (3.32)$$

*Preuve.* 1. Pour tout  $x \in E$ ,

$$\mu_n(x) = \mathbb{P}(X_n = x) = \sum_{y \in E} \mathbb{P}(X_n = x, X_{n-1} = y) \quad (3.33)$$

$$= \sum_{y \in E} Q(y, x) \mu_{n-1}(y) = \mu_{n-1} Q(x). \quad (3.34)$$

Et par récurrence on obtient  $\mu_n(x) = \mu_0 Q^n(x)$ .

2. On utilise le résultat précédent avec  $\mu_0 = \delta_x$  :

$$\mu_n(y) = \mathbb{P}_x(X_n = y) = \delta_x Q^n(y) = \sum_{z \in E} \delta_x(z) Q^n(z, y) = Q^n(x, y). \quad (3.35)$$

3. C'est une conséquence du point précédent :

$$\mathbb{E}_x(h(X_n)) = \sum_{y \in E} h(y) \mathbb{P}_x(X_n = y) \quad (3.36)$$

$$= \sum_{y \in E} h(y) Q^n(x, y) = Q^n h(x). \quad (3.37)$$

□

**Remarque 12.** Le produit matriciel  $Q^n(x, y)$  correspond bien à sommer les probabilités de toutes les trajectoires que le processus peut emprunter pour aller de  $x$  à  $y$  en  $n$  pas :

$$Q^n(x, y) = \mathbb{P}_x(X_n = y) \quad (3.38)$$

$$= \sum_{x_1, \dots, x_{n-1} \in E} \mathbb{P}_x(X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = y) \quad (3.39)$$

$$= \sum_{x_1, \dots, x_{n-1} \in E} Q(x, x_1) \cdots Q(x_{n-1}, y) \quad (3.40)$$

$$= Q^n(x, y). \quad (3.41)$$

En particulier le cas  $n = 2$  représente la somme sur tous les  $z \in E$  intermédiaires permettant d'aller de  $x$  à  $y$  en deux coups.

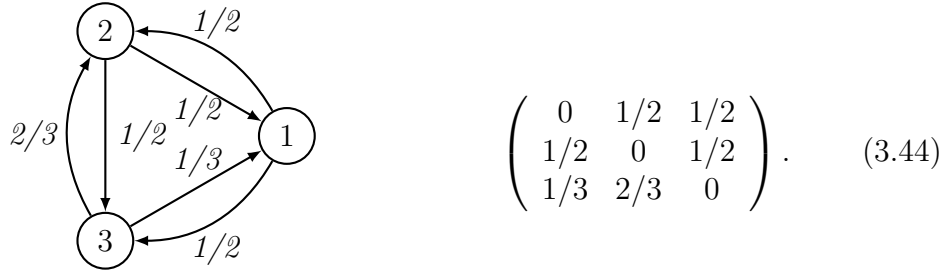
$$Q^2(x, y) = \sum_{z \in E} Q(x, z) Q(z, y). \quad (3.42)$$

C'est la même intuition pour comprendre que  $\mu Q$  est la loi de  $X_1$  quand  $X_0$  a pour loi  $\mu$  : la probabilité d'être en  $x$  au temps 1 est la somme sur tous les éléments  $y \in E$  de la probabilité d'être initialement en  $y$  puis de faire la transition  $y \rightarrow x$ . Avec les notations matricielles on a donc

$$P_\mu(X_1 = x) = \sum_{y \in E} P_\mu(X_1 = x | X_0 = y) P_\mu(X_0 = y) = \sum_{y \in E} Q(y, x) \mu(y) = \mu Q(x). \quad (3.43)$$

Et là aussi on peut généraliser avec  $n \geq 2$  en sommant sur tous les points de départ et toutes les trajectoires possibles.

**Exemple 10.** La matrice de transition peut donc être utile pour calculer les transitions à  $n$  pas. Considérons l'exemple suivant :



On vérifie bien que

$$\begin{aligned} P_1(X_2 = 1) &= P_1(X_2 = 1 | X_1 = 3) P_1(X_1 = 3) + P_1(X_2 = 1 | X_1 = 2) P_1(X_1 = 2) \\ &= \frac{1}{2} \frac{1}{3} + \frac{1}{2} \frac{1}{2} \\ &= Q^2(1, 1). \end{aligned} \quad (3.45)$$

Lorsque l'on considère plus de pas, l'utilisation de la matrice peut être commode. On obtient ainsi  $P_2(X_5 = 2) = Q^5(2, 2) = 1/3$ .

Notons enfin que la propriété de Markov est énoncée uniquement à un pas mais qu'elle peut aisément se généraliser à toute la trajectoire future :

**Proposition 16** (Propriété de Markov généralisée). Soit  $(X_n)_{n \geq 0}$  une chaîne de Markov,  $n, k$  deux entiers et  $x_0, \dots, x_n, y_1, \dots, y_k$  dans  $E$ . Alors

$$P(X_{n+1} = y_1, \dots, X_{n+k} = y_k | X_0 = x_0, \dots, X_n = x_n) = P_{x_n}(X_1 = y_1, \dots, X_k = y_k). \quad (3.46)$$

*Preuve.* On utilise les équations de Chapman Kolmogorov

$$P(X_{n+1} = y_1, \dots, X_{n+k} = y_k | X_0 = x_0, \dots, X_n = x_n) \quad (3.47)$$

$$= \frac{P(X_{n+1} = y_1, \dots, X_{n+k} = y_k, X_0 = x_0, \dots, X_n = x_n)}{P(X_0 = x_0, \dots, X_n = x_n)} \quad (3.48)$$

$$= \frac{P(X_0 = x_0) \prod_{i=0}^{n-1} Q(x_i, x_{i+1}) Q(x_n, y_1) \prod_{i=1}^{k-1} Q(y_i, y_{i+1})}{P(X_0 = x_0) \prod_{i=0}^{n-1} Q(x_i, x_{i+1})} \quad (3.49)$$

$$= Q(x_n, y_1) \prod_{i=1}^{k-1} Q(y_i, y_{i+1}) \quad (3.50)$$

$$= P_{x_n}(X_1 = y_1, \dots, X_k = y_k). \quad (3.51)$$

□

## 3.2 Propriété de Markov forte

Commençons par la propriété de Markov simple. Fixons un temps **deterministe**  $k \geq 0$  et  $x \in E$ . Alors, conditionnellement à  $\{X_k = x\}$ , passé et futur sont indépendants. De plus, le futur est également markovien avec même matrice de transition et condition initiale  $x$ .

**Proposition 17 (Markov simple).** *Soit*

1.  $(X_n)_{n \geq 0}$  une chaîne de Markov homogène de matrice de transition  $Q$ ,
2.  $k \geq 1$  un entier,
3.  $x \in E$  tel que  $P(X_k = x) > 0$ .

Alors le **processus décalé**  $(X_{k+n})_{n \geq 0}$ , sous la probabilité  $P(\cdot | X_k = x)$ , est :

1. indépendant de  $(X_0, \dots, X_k)$ ,
2. markovien de matrice de transition  $Q$  et condition initiale  $x$ .

*Preuve.* On fixe  $K \geq 1$  et  $x_0, \dots, x_{k-1}, y_1, \dots, y_K$  dans  $E$ . En utilisant la propriété de Markov (généralisée) au temps  $n$

$$P(X_k = x, X_{k+1} = y_1, \dots, X_{k+K} = y_K, X_0 = x_0, \dots, X_k = x | X_k = x) \quad (3.52)$$

$$= P_x(X_1 = y_1, \dots, X_K = y_K) \frac{P(X_0 = x_0, \dots, X_k = x)}{P(X_k = x)} \quad (3.53)$$

$$= P_x(X_1 = y_1, \dots, X_K = y_K) P(X_0 = x_0, \dots, X_k = x | X_k = x) \quad (3.54)$$

□

La propriété de Markov forte est similaire à l'énoncé précédent mais le temps  $k$  déterministe est remplacé par un un temps d'arrêt  $T$ .

**Proposition 18 (Markov fort).** *Soit*

1.  $(X_n)_{n \geq 0}$  une chaîne de Markov homogène de matrice de transition  $Q$ ,
2.  $T$  un temps d'arrêt pour la filtration engendrée par  $(X_n)_{n \geq 0}$ ,
3.  $x \in E$  tel que  $P(X_T = x, T < +\infty) > 0$ .

Alors le **processus décalé**  $(X_{T+n})_{n \geq 0}$ , sous la probabilité  $P(\cdot | X_T = x, T < +\infty)$ , est :

1. indépendant de  $(X_0, \dots, X_T)$ ,
2. markovien de matrice de transition  $Q$  et condition initiale  $x$ .

*Preuve.* On fixe  $k \geq 0$  et  $x_0, \dots, x_{k-1}, x \in E$  pour la trajectoire passée. On fixe également  $K \geq 0$  et  $y_1, \dots, y_K \in E$  pour la trajectoire future. On doit montrer que

$$\begin{aligned} & P((X_0, \dots, X_T) = (x_0, \dots, x_{k-1}, x), (X_T, \dots, X_{T+K}) = (x, y_1, \dots, y_K) | X_T = x, T < +\infty) \\ &= P((X_0, \dots, X_T) = (x_0, \dots, x_{k-1}, x) | X_T = x, T < +\infty) P_x((X_1, \dots, X_K) = (y_1, \dots, y_K)). \end{aligned} \quad (3.55)$$

Le terme de gauche se réécrit

$$\frac{P((X_0, \dots, X_k) = (x_0, \dots, x_{k-1}, x), T = k, (X_k, \dots, X_{k+K}) = (x, y_1, \dots, y_K))}{P(X_T = x, T < +\infty)}. \quad (3.56)$$

Comme  $T$  est un temps d'arrêt pour la filtration engendrée par  $(X_n)_{n \geq 0}$ ,  $\{T = k\}$  est dans  $\sigma(X_0, \dots, X_k)$  et il existe donc un ensemble de trajectoires  $A_k$  tel que

$$\{T = k\} = \{(X_0, \dots, X_k) \in A_k\}. \quad (3.57)$$

On déduit de cette observation que

$$\begin{aligned} & \{(X_0, \dots, X_k) = (x_0, \dots, x_{k-1}, x), T = k\} \\ &= \begin{cases} \emptyset & \text{si } (x_0, \dots, x_{k-1}, x) \notin A_k \\ \{(X_0, \dots, X_k) = (x_0, \dots, x_{k-1}, x)\} & \text{si } (x_0, \dots, x_{k-1}, x) \in A_k \end{cases} \end{aligned} \quad (3.58)$$

Si  $(x_0, \dots, x_k) \notin A_k$ , le terme de gauche dans (3.55) est donc nul et on montre de la même façon que  $P((X_0, \dots, X_T) = (x_0, \dots, x_{k-1}, x) | X_T = x, T < +\infty) = 0$ .

Si  $(x_0, \dots, x_k) \in A_k$ , le terme de gauche se réécrit

$$\begin{aligned}
& \frac{\mathbb{P}((X_0, \dots, X_k) = (x_0, \dots, x_{k-1}, x), (X_k, \dots, X_{k+K}) = (x, y_1, \dots, y_K))}{\mathbb{P}(X_T = x, T < +\infty)} \\
& \stackrel{\text{Markov en } k}{=} \frac{\mathbb{P}((X_0, \dots, X_k) = (x_0, \dots, x_{k-1}, x)) \mathbb{P}_x((X_1, \dots, X_K) = (y_1, \dots, y_K))}{\mathbb{P}(X_T = x, T < +\infty)} \\
& = \frac{\mathbb{P}((X_0, \dots, X_k) = (x_0, \dots, x_{k-1}, x), T = k) \mathbb{P}_x((X_1, \dots, X_K) = (y_1, \dots, y_K))}{\mathbb{P}(X_T = x, T < +\infty)} \\
& = \mathbb{P}((X_0, \dots, X_T) = (x_0, \dots, x_{k-1}, x) | T < +\infty, X_T = x) \mathbb{P}_x((X_1, \dots, X_K) = (y_1, \dots, y_K)). \tag{3.59}
\end{aligned}$$

□

**Remarque 13.** On notera bien que la démonstration utilise que  $T$  est un temps d'arrêt pour établir (3.55). On ne peut pas espérer que le résultat soit vrai pour un temps aléatoire  $T$  quelconque.

### 3.3 Récurrence et transience

On considère dans toute cette section une chaîne de Markov  $(X_n)_{n \geq 0}$  homogène sur un espace d'états  $E$ . On cherche à étudier le nombre de visites de  $(X_n)_{n \geq 0}$  en un point  $x \in E$  donné. On cherche en particulier à savoir si ce nombre de passages est fini ou non.

**Définition 10.**

1. On dit que  $x \in E$  communique avec  $y \in E$  s'il existe un entier  $n \geq 0$  tel que

$$\mathbb{P}_x(X_n = y) > 0 \quad \text{i.e.} \quad Q^n(x, y) > 0. \tag{3.60}$$

2. Si tous les éléments communiquent entre eux on dit que la chaîne est **irréductible**.

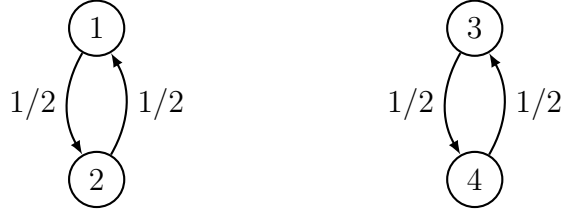
**Exemple 11.** 1. La marche aléatoire simple sur  $\mathbb{Z}^d$  est irréductible. En effet pour tout  $x, y$  dans  $\mathbb{Z}^d$ , on vérifie que

$$Q^{\|x-y\|_1}(x, y) = \left(\frac{1}{2d}\right)^{\|x-y\|_1} > 0. \tag{3.61}$$

2. La chaîne suivante n'est en revanche pas irréductible :

Pour  $z \in E$ , on appelle **temps d'atteinte** de  $z$  le temps aléatoire

$$T_z := \inf\{n \geq 1, X_n = z\}, \tag{3.62}$$



avec la convention (intuitive)  $\inf \emptyset = +\infty$ . De manière plus générale, pour  $k \geq 0$  on définit  $T_z^{(0)} = 0$ ,  $T_z^{(1)} = T_z$  et pour tout  $k \geq 2$ , on définit le  $k$ -ième **temps de passage** par

$$T_z^{(k)} = \inf\{n \geq T_z^{(k-1)} + 1, X_n = z\}. \quad (3.63)$$

On note que

$$1 \leq T_z \leq T_z^{(2)} \leq \dots \leq T_z^{(k)} \leq T_z^{(k+1)} \leq +\infty, \quad (3.64)$$

et que  $\{T_z^{(k)} = T_z^{(k+1)}\} = \{T_z^{(k)} = +\infty\}$ . Ces variables aléatoires peuvent être écrites comme des fonction du processus  $(X_n)_{n \geq 0}$ . On introduit explicitement ces fonctions qui nous seront utiles plus loin. On définit donc la fonction  $t_z$  qui, à une trajectoire, associe le premier temps où elle touche le point  $z$  :

$$\begin{aligned} t_z : E^{\mathbb{N}} &\rightarrow \mathbb{N} \cup \{+\infty\} \\ (x_n)_{n \geq 0} &\mapsto \inf\{n \geq 1, x_n = z\} \end{aligned} \quad (3.65)$$

avec la convention (cohérente)  $\inf \emptyset = +\infty$ .

De manière plus générale, on définit par récurrence les fonctions  $t_z^{(k)}$ , pour  $k \geq 0$  par  $t_z^{(0)} = 0$ ,  $t_z^{(1)} = t_z$  et pour tout  $k \geq 2$ ,

$$\begin{aligned} t_z^{(k)} : E^{\mathbb{N}} &\rightarrow \mathbb{N} \cup \{+\infty\} \\ (x_n)_{n \geq 0} &\mapsto \inf\{n \geq t_z^{(k-1)} + 1, x_n = z\}. \end{aligned} \quad (3.66)$$

On note que l'on a bien les égalités suivantes,

$$\begin{aligned} T_z^{(0)} = 0 \quad T_z^{(1)} = T_z = t_z((X_n)_{n \geq 0}) \quad \text{et pour } k \geq 2, \\ T_z^{(k)} = t_z^{(k)}((X_n)_{n \geq 0}). \end{aligned} \quad (3.67)$$

On introduit également **nombre de passages en  $z$**

$$N_z = \sum_{k=0}^{+\infty} 1_{\{X_k = z\}}, \quad (3.68)$$

et le **nombre de passages en  $z$  jusqu'au temps  $n$**

$$N_z(n) = \sum_{k=0}^n 1_{\{X_k=z\}}. \quad (3.69)$$

Là encore il sera commode d'expliciter ces variables comme fonctions de  $(X_n)_{n \geq 0}$ . Nous définissons donc

$$\begin{aligned} n_z : E^{\mathbb{N}} &\rightarrow \mathbb{N} \cup \{+\infty\} \\ (x_n)_{n \geq 0} &\mapsto \sum_{k=0}^{+\infty} 1_{\{x_k=z\}}, \end{aligned} \quad (3.70)$$

et pour tout  $n \geq 0$

$$\begin{aligned} n_z^{(n)} : E^{\mathbb{N}} &\rightarrow \mathbb{N} \cup \{+\infty\} \\ (x_n)_{n \geq 0} &\mapsto \sum_{k=0}^n 1_{\{x_k=z\}}. \end{aligned} \quad (3.71)$$

On vérifie que, pour tout  $n \geq 0$ ,

$$N_z = n_z((X_k)_{k \geq 0}) \quad \text{et} \quad N_z^{(n)} = n_z^{(n)}((X_k)_{k \geq 0}). \quad (3.72)$$

On note que, d'après le théorème de Fubini, pour tout  $x, y$  dans  $E$ ,

$$\mathbb{E}_x(N_y) = \mathbb{E}_x\left(\sum_{k=0}^{+\infty} 1_{\{X_k=y\}}\right) = \sum_{k=0}^{+\infty} \mathbb{P}_x(X_k = y) = \sum_{k=0}^{+\infty} Q^k(x, y). \quad (3.73)$$

Nous allons plus tard appliquer la propriété de Markov aux temps de passages et il faut s'assurer qu'il s'agit de temps d'arrêt :

**Proposition 19.** *Pour tout  $k \geq 0$  et tout  $x \in E$ ,  $T_x^{(k)}$  est un temps d'arrêt pour la filtration engendrée par  $(X_n)_{n \geq 0}$ .*

*Preuve.* Pour tous  $k, n$  entiers

$$\{T_x^{(k)} \leq n\} = \bigcup_{1 \leq i_1 < \dots < i_k \leq n} \{X_{i_1} = \dots = X_{i_k} = x\} \in \sigma(X_0, \dots, X_n). \quad (3.74)$$

□

On arrive maintenant à la définition importante de cette section :

**Définition 11.** *On dit que  $x$  est*

1. **récurrent** si  $P_x(T_x < +\infty) = 1$
2. **transient** si  $P_x(T_x < +\infty) < 1$ .

On note que le point 2. est équivalent à  $P_x(T_x = +\infty) > 0$ , c'est-à-dire au fait que la **probabilité d'échapper** soit strictement positive. Le théorème suivant propose une caractérisation de la transience et de la récurrence extrêmement importante et utile :

**Théorème 7.** *Pour tout  $x \in E$ ,*

1.  $x$  **est récurrent**  $\Leftrightarrow N_x = +\infty$   $P_x - p.s.$
2.  $x$  **est transient**  $\Leftrightarrow N_x \rightsquigarrow Geo(P_x(T_x = +\infty))$  sous  $P_x$

On a donc, d'après le Théorème 7, la caractérisation suivante de la récurrence :

$$\begin{aligned}
x \text{ est transient} &\Leftrightarrow E_x(N_x) < +\infty \\
&\Leftrightarrow N_x < +\infty \quad P_x - p.s. \\
&\Leftrightarrow P_x(N_x < +\infty) > 0 \\
&\Leftrightarrow P_x(N_x < +\infty) = 1.
\end{aligned} \tag{3.75}$$

Ces équivalences sont non triviales : elles sont la conséquence du théorème précédent et fausses pour une variable aléatoire quelconque.

*Preuve.* Soit  $k \geq 1$  un entier. On note que sur  $\{T_x^{(k-1)} < +\infty\}$ ,

$$T_x^{(k)} = T_x^{(k-1)} + t_x((X_{T_x^{(k-1)}+n})_{n \geq 0}), \tag{3.76}$$

et on en déduit que

$$\{N_x \geq k+1\} \stackrel{P_x - p.s.}{=} \{T_x^{(k)} < +\infty\} = \{T_x^{(k-1)} < +\infty, t_x((X_{T_x^{(k-1)}+n})_{n \geq 0}) < +\infty\}. \tag{3.77}$$

On peut appliquer la propriété de Markov forte au temps d'arrêt  $T_x^{(k-1)}$ , en notant que  $\{T_x^{(k-1)} < +\infty\} \subset \{X_{T_x^{(k-1)}} = x\}$ , et on obtient,

$$\begin{aligned}
P_x(N_x \geq k+1) &= P_x(t_x((X_{T_x^{(k-1)}+n})_{n \geq 0}) < +\infty | T_x^{(k-1)} < +\infty, X_{T_x^{(k-1)}} = x) \\
&= P_x(T_x^{(k-1)} < +\infty, X_{T_x^{(k-1)}} = x) \\
&\stackrel{\text{Markov Fort}}{=} P_x(t_x((X_n)_{n \geq 0}) < +\infty) P_x(T_x^{(k-1)} < +\infty) \\
&= P_x(T_x < +\infty) P_x(T_x^{(k-1)} < +\infty)
\end{aligned} \tag{3.78}$$

et en itérant

$$P_x(N_x \geq k+1) = P_x(T_x < +\infty)^k. \tag{3.79}$$

On en déduit :



1. **si  $x$  est récurrent** : pour tout  $k \geq 1$ ,  $P_x(N_x \geq k) = 1$  et donc

$$P_x(N_x = +\infty) = P_x\left(\bigcap_{k \rightarrow +\infty} \downarrow \{N_x \geq k\}\right) = \lim_{k \rightarrow +\infty} P_x(N_x \geq k) = 1. \quad (3.80)$$

2. **si  $x$  est transient** : on reconnaît la queue de distribution d'une géométrique à valeurs dans  $\{1, 2, \dots\}$  et de paramètre de succès  $1 - P_x(T_x < +\infty) = P_x(T_x = +\infty) > 0$ .

□

**Proposition 20.** 1. Lorsque la chaîne est **irréductible**, tous les états sont du même type. On parle alors de chaîne récurrente ou transiente.

2. Lorsque la chaîne est (irréductible) **récurrente**, pour tout  $x, y \in E$

$$P_x(N_y = +\infty) = 1 \text{ en particulier } P_x(T_y < +\infty) = 1. \quad (3.81)$$

3. Lorsque la chaîne est (irréductible) **transiente**, pour tout  $x, y \in E$ ,

$$P_x(N_y < +\infty) = 1. \quad (3.82)$$

*Preuve.* 1. Soit  $x$  un élément de  $E$  récurrent. On considère un autre élément  $y \in E$  et on veut montrer qu'il est également récurrent. En utilisant la caractérisation du Théorème 7, on doit donc montrer que  $E_y(N_y) = +\infty$ . Comme  $(X_n)_{n \geq 0}$  est irréductible, il existe  $m_1$  et  $m_2$  tels que

$$Q^{m_1}(x, y) > 0 \quad \text{et} \quad Q^{m_2}(y, x) > 0. \quad (3.83)$$

Pour  $n \geq m_1 + m_2$

$$Q^n(y, y) \geq Q^{m_1}(y, x)Q^{n-(m_1+m_2)}(x, x)Q^{m_2}(x, y). \quad (3.84)$$

Comme la série de terme général  $Q^{n-(m_1+m_2)}(x, x)$  diverge, c'est également le cas pour celle de terme général  $Q^n(y, y)$  également et donc  $E_y(N_y) = +\infty$ . On en déduit que  $y$  est récurrent.

2. Comme la chaîne est irréductible il existe  $n \geq 0$  tel que  $Q^n(y, x) > 0$ . En décomposant sur tous les chemins possibles, on en déduit qu'il existe un chemin  $x_1, \dots, x_n = x$  tel que  $P_y(X_1 = x_1, \dots, X_n = x) > 0$ . Quitte à démarrer à la dernière occurrence de  $y$  dans le chemin, on peut supposer que pour tout  $1 \leq i < n$ ,  $x_i \neq y$ . On remarque que

$$\{X_1 = x_1, \dots, X_n = x, t_y((X_{n+k})_{k \geq 0}) = +\infty\} \stackrel{P_y\text{-p.s.}}{\subset} \{T_y = +\infty\}. \quad (3.85)$$

En appliquant Markov faible au temps  $n$  nous obtenons donc

$$\begin{aligned} P_y(T_y = +\infty) &\geq P_y(X_1 = x_1, \dots, X_n = x, t_y((X_{n+k})_{k \geq 0}) = +\infty) \\ &\geq P_y(X_1 = x_1, \dots, X_n = x)P_x(T_y = +\infty). \end{aligned} \quad (3.86)$$

Comme  $P_y(X_1 = x_1, \dots, X_n = x) > 0$ , on en déduit  $P_x(T_y = +\infty) = 0$ . Par ailleurs,

$$\begin{aligned} P_x(N_y = +\infty) &= P_x(T_y < +\infty, n_y((X_{T_y+k})_{k \geq 0}) = +\infty) \\ &= P_x(T_y < +\infty)P_y(N_y = +\infty), \end{aligned} \quad (3.87)$$

en appliquant Markov fort au temps d'arrêt  $T_y$  et en notant que sur  $\{T_y < +\infty\}$ ,  $X_{T_y} = y$ . Comme  $P_x(T_y < +\infty) = P_y(N_y = +\infty) = 1$ , cela conclut la preuve.

3. Enfin lorsque la chaîne est transiente, en utilisant à nouveau (3.87), et  $P_y(N_y = +\infty) = 0$ , on obtient  $P_x(N_y = +\infty) = 0$ . □

On obtient un premier résultat pour les chaînes sur un espace d'états fini.

**Proposition 21.** *On considère une chaîne de Markov  $(X_n)_{n \geq 0}$  sur un espace d'états  $E$  **fini**. Alors  $(X_n)_{n \geq 0}$  admet au moins un état récurrent. Si  $(X_n)_{n \geq 0}$  est de plus irréductible elle est récurrente.*

*Preuve.* Comme  $E$  est fini, pour tout  $x \in E$ ,

$$P_x - p.s. \quad \exists y \in E, \quad N_y = +\infty, \quad (3.88)$$

on en déduit que  $P_x(\bigcup_{y \in E} \{N_y = +\infty\}) = 1$  et donc qu'il existe  $y \in E$  tel que  $P_x(N_y = +\infty) > 0$ . D'après (3.87),  $P_x(N_y = +\infty) = P_x(T_y < +\infty)P_y(N_y = +\infty)$ , on en déduit que  $P_y(N_y = +\infty) > 0$ . D'après le Théorème 7, cela prouve  $y$  est récurrent. Si la chaîne est irréductible la Proposition 20 nous assure que tous les états sont récurrents. □

**Exemple 12 (La marche aléatoire simple sur  $\mathbb{Z}^d$ ).** *Le célèbre théorème suivant, dû à **Pólya**, détermine la récurrence ou transience de la marche selon la dimension.*

**Théorème 8** (Récurrence-transience de la MAS sur  $\mathbb{Z}^d$ ). *La marche aléatoire simple sur  $\mathbb{Z}^d$  est récurrente si et seulement si  $d \leq 2$ .*

*Preuve.* i.  $\mathbf{d} = \mathbf{1}$ . Montrons que  $E_0(N_0) = +\infty$ . Rappelons pour commencer que  $E_0(N_0) = \sum_{n \geq 0} P_0(X_n = 0)$ . On note que si  $n$  est impair cette probabilité est nulle tandis que pour tout  $k \geq 0$

$$P_0(X_{2k} = 0) = \frac{\binom{2k}{k}}{2^{2k}} \sim \frac{1}{\sqrt{\pi k}}, \quad (3.89)$$

où l'équivalent est obtenu grâce à la formule de Stirling. La série de terme général  $P_0(X_n = 0)$  diverge donc  $E_0(N_0) = +\infty$  et la marche est récurrente.

ii.  $\mathbf{d} = \mathbf{2}$ . On utilise la même méthode : donner un équivalent pour  $P_0(X_{2k} = 0)$ . On peut calculer directement cette probabilité en dénombrant les chemins. On peut sinon prouver que les projections  $U_n$  et  $V_n$  de  $X_n$  dans la base  $((1/2, 1/2); (-1/2, 1/2))$  sont deux marches aléatoires indépendantes en dimension 1. On obtient alors

$$P_0(X_{2k} = 0) = P_0(U_{2k} = 0, V_{2k} = 0) \sim \frac{1}{\pi k}. \quad (3.90)$$

De nouveau il s'agit du terme général d'une série divergente et on conclut comme dans le cas précédent.

iii.  $\mathbf{d} \geq \mathbf{3}$ . En dénombrant les chemins, on peut montrer qu'il existe  $C > 0$  tel que

$$P_0(X_{2k} = 0) \sim \frac{C}{k^{3/2}}. \quad (3.91)$$

La série est cette fois convergente et cela prouve que la marche est transiente...ça ne fait qu'empirer avec la dimension. □

### 3.4 Mesures stationnaires et probabilités invariantes

De nouveau, dans toute cette partie, on considère une chaîne de Markov  $(X_n)_{n \geq 0}$ , sur un espace d'états dénombrable  $E$  de matrice de transition  $Q$ .

**Définition 12.** On dit d'une mesure (de probabilité ou non) sur  $E$  qu'elle est stationnaire ou invariante si

$$\mu Q = \mu. \quad (3.92)$$

Dans le cas où  $\mu$  est une mesure de probabilité, on rappelle que  $\mu Q$  désigne la loi de probabilité de  $X_1$  lorsque  $X_0$  est distribuée selon  $\mu$ . Dire que  $\mu$  est probabilité invariante est donc dire que si un marcheur est distribué initialement selon  $\mu$  alors il est encore distribué selon cette même mesure  $\mu$  pour tout temps. Si on considère la mesure  $\mu$  comme une répartition d'une masse de sable, alors  $\mu$  est stationnaire si la masse reste distribuée de la même façon après l'action redistributive de la chaîne.

Quelques remarques avant de démarrer notre étude :

1. Si  $\mu$  et  $\nu$  sont deux mesures stationnaires alors  $\mu + \nu$  est encore une mesure stationnaire. De même  $\lambda\mu$  est stationnaire pour tout  $\lambda \in [0, +\infty]$ .
2. Si  $\mu$  est une mesure stationnaire de masse finie, i.e.  $\mu(E) < +\infty$  alors  $\nu = \mu/\mu(E)$  est une probabilité invariante.
3. Si  $\mu$  est une mesure stationnaire alors par récurrence  $\mu Q^k = \mu$  pour tout  $k \geq 0$ . On en déduit que sous  $P_\mu$ ,  $X_k$  a pour loi  $\mu$  pour tout  $k \geq 0$ .

**L'objectif de cette partie est d'étudier l'existence et l'unicité de mesures stationnaires ou de probabilités invariantes.**

### 3.4.1 Mesures stationnaires

Nous commençons notre étude par un premier résultat d'**existence**.

**Proposition 22.** *Soit  $(X_n)_{n \geq 0}$  une chaîne de Markov et  $x \in E$  un état récurrent. Alors la mesure  $\mu_x$  définie par*

$$\mu_x(y) = E_x(N_y(T_x - 1)), \quad \text{pour tout } y \in E, \quad (3.93)$$

*est stationnaire. On dit parfois que c'est la mesure associée à l'excursion de  $x$  à  $x$ .*

On rappelle que pour  $n \geq 1$  le **nombre de passages en  $y$  jusqu'au temps  $n$**  est défini par

$$N_y(n) = \sum_{k=0}^n 1_{\{X_k=y\}}. \quad (3.94)$$

On note que

$$\mu_x(x) = E_x \left( \sum_{k=0}^{T_x-1} 1_{\{X_k=x\}} \right) = 1, \quad (3.95)$$

et

$$\mu_x(E) = \sum_{y \in E} \mu_x(y) = \sum_{y \in E} \mathbb{E}_x \left( \sum_{k=0}^{T_x-1} 1_{\{X_k=y\}} \right) = \mathbb{E}_x(T_x). \quad (3.96)$$

Il est important de noter que l'existence de la mesure stationnaire par la Proposition 22 n'assure pas l'existence d'une probabilité invariante puisqu'il n'est pas toujours vrai que  $\mathbb{E}_x(T_x) < +\infty$  et donc que l'on puisse renormaliser  $\mu_x$ .

*Preuve.* On veut prouver que pour tout  $y \in E$

$$\mu_x(y) = \sum_{z \in E} \mu_x(z) Q(z, y). \quad (3.97)$$

Or, pour tout  $z \in E$ ,

$$\mu_x(z) = \mathbb{E}_x \left( \sum_{k=0}^{T_x-1} 1_{X_k=z} \right) = \mathbb{E}_x \left( \sum_{k=0}^{+\infty} 1_{X_k=z} 1_{T_x > k} \right) = \sum_{k=0}^{+\infty} \mathbb{P}_x(X_k = z, T_x > k), \quad (3.98)$$

On en déduit

$$\sum_{z \in E} \mu_x(z) Q(z, y) = \sum_{z \in E} \sum_{k=0}^{+\infty} \mathbb{P}_x(X_k = z, T_x > k) Q(z, y). \quad (3.99)$$

Or, pour  $z \in E$  et  $k \geq 0$  tel que  $\mathbb{P}_x(T_x > k, X_k = z) > 0$ ,

$$\mathbb{P}_x(X_k = z, T_x > k, X_{k+1} = y) = \mathbb{P}_x(X_{k+1} = y | T_x > k, X_k = z) \mathbb{P}_x(T_x > k, X_k = z). \quad (3.100)$$

Comme  $\{T_x > k\} = \{X_1 \neq x, \dots, X_k \neq x\}$ , on peut appliquer Markov simple au temps  $k$  et on obtient

$$\mathbb{P}_x(X_{k+1} = y | T_x > k, X_k = z) = Q(z, y). \quad (3.101)$$

Finalement, on obtient

$$\mathbb{P}_x(X_k = z, T_x > k, X_{k+1} = y) = \mathbb{P}_x(X_k = z, T_x > k) Q(z, y), \quad (3.102)$$

et cette égalité est encore valable lorsque  $\mathbb{P}_x(T_x > k, X_k = z) = 0$ .

En assemblant tous les bouts, on obtient pour  $z \in E$  et  $k \geq 0$ ,

$$\begin{aligned}
\sum_{z \in E} \mu_x(z) Q(z, y) &= \sum_{z \in E} \sum_{k=0}^{+\infty} P_x(X_k = z, T_x > k, X_{k+1} = y) \\
&= \sum_{k=0}^{+\infty} P_x(T_x \geq k+1, X_{k+1} = y) \\
&= \sum_{k=1}^{+\infty} P_x(T_x \geq k, X_k = y).
\end{aligned} \tag{3.103}$$

On doit comparer cette dernière quantité à

$$\mu_x(y) = \sum_{k=0}^{+\infty} P_x(T_x > k, X_k = y). \tag{3.104}$$

Or

$$\begin{aligned}
\sum_{k=1}^{+\infty} P_x(T_x \geq k, X_k = y) &= \sum_{k=0}^{+\infty} (P_x(T_x > k, X_k = y) + P_x(T_x = k, X_k = y)) - P_x(X_0 = y) \\
&= \mu_x(y) + P_x(X_{T_x} = y) - P_x(X_0 = y),
\end{aligned} \tag{3.105}$$

puisque  $T_x < +\infty$ ,  $P_x$ -p.s. car  $x$  est récurrent. On note que :

- si  $y \neq x$ ,  $P_x(X_{T_x} = y) = P_x(X_0 = y) = 0$ ,
- si  $y = x$ ,  $P_x(X_{T_x} = y) = P_x(X_0 = y) = 1$ .

Dans tous les cas, on a bien,

$$\sum_{k=1}^{+\infty} P_x(T_x \geq k, X_k = y) = \mu_x(y), \tag{3.106}$$

et cela conclut la preuve. □

Nous poursuivons avec un résultat d'**unicité**. Avant d'énoncer le théorème principal de cette partie nous avons besoin de quelques résultats intermédiaires mais qui ont aussi leur intérêt propre.

**Proposition 23.** *Si  $(X_n)_{n \geq 0}$  est irréductible et  $\mu$  est une mesure stationnaire non triviale (c'est-à-dire que  $\mu$  n'est pas identiquement égale à 0 ou à  $+\infty$ ) alors pour tout  $x \in E$*

$$0 < \mu(x) < +\infty. \tag{3.107}$$

*Preuve.* Soit  $x$  et  $y$  deux éléments de  $E$ . Comme la chaîne est irréductible il existe  $k \geq 0$  tel que  $Q^k(x, y) > 0$ . On a donc

$$\mu(y) = \mu Q^k(y) \geq \mu(x) Q^k(x, y). \quad (3.108)$$

On en déduit facilement le résultat puisque si  $\mu(x) > 0$ , il est clair que  $\mu(y) > 0$ . De même si  $\mu(x) = +\infty$ , il est clair que  $\mu(y) = +\infty$ .  $\square$

**Corollaire 3.** *Soient  $\mu$  et  $\nu$  deux mesures stationnaires d'une chaîne irréductible telles que*

1.  $\mu \geq \nu$  (i.e. pour tout  $x \in E$ ,  $\mu(x) \geq \nu(x)$ ),
2. il existe  $x \in E$  tel que  $\mu(x) = \nu(x)$ .

Alors  $\mu = \nu$ .

*Preuve.* On vérifie facilement que  $\mu - \nu$  est une mesure stationnaire pour la chaîne  $(X_n)_{n \geq 0}$  irréductible. Comme de plus  $(\mu - \nu)(x) = 0$ , on en déduit que  $\mu - \nu$  est la mesure triviale identiquement nulle.  $\square$

**Théorème 9.** *Soit  $(X_n)_{n \geq 0}$  une chaîne de Markov irréductible et récurrente. Alors toutes les mesures stationnaires non triviales sont multiples les unes des autres.*

*Preuve.* On considère un élément  $x \in E$ . D'après la Proposition 22,  $\mu_x$  est une mesure stationnaire (non triviale puisque  $\mu_x(x) = 1$ ). Soit  $\tilde{\nu}$  une mesure stationnaire non triviale. D'après la Proposition 23,  $0 < \tilde{\nu}(x) < +\infty$ . On peut donc renormaliser cette mesure en posant  $\nu = \tilde{\nu}/\tilde{\nu}(x)$  de telle sorte que  $\nu(x) = 1$ . On doit montrer que  $\mu_x = \nu$ . On va utiliser le Corollaire 3 et nous devons donc montrer :

1.  $\nu \geq \mu_x$ ,
2.  $\mu_x(x) = \nu(x) = 1$ .

On se concentre en fait sur le premier point puisque le second est trivial. Soit  $y \neq x$  un élément de  $E$ .

$$\nu(y) = \nu Q(y) = \sum_{z_1 \in E} \nu(z_1) Q(z_1, y) = Q(x, y) + \sum_{z_1 \neq x} \nu(z_1) Q(z_1, y). \quad (3.109)$$

On note ensuite que pour tout  $z_1 \neq x$ ,

$$\nu(z_1) = Q(x, z_1) + \sum_{z_2 \neq x} \nu(z_2) Q(z_2, z_1), \quad (3.110)$$

et on obtient

$$\nu(y) = Q(x, y) + \sum_{z_1 \neq x} Q(x, z_1)Q(z_1, y) + \sum_{z_1 \neq x, z_2 \neq x} \nu(z_2)Q(z_2, z_1)Q(z_1, y). \quad (3.111)$$

Pour tout  $k \geq 1$ , on peut itérer cette procédure et on obtient

$$\begin{aligned} \nu(y) = & Q(x, y) + \sum_{z_1 \neq x} Q(x, z_1)Q(z_1, y) + \cdots + \sum_{z_1, \dots, z_{k-1} \neq x} Q(x, z_{k-1})Q(z_{k-1}, z_{k-2}) \cdots Q(z_1, y) \\ & + \sum_{z_1, \dots, z_k \neq x} \nu(z_k)Q(z_k, z_{k-1}) \cdots Q(z_1, y). \end{aligned} \quad (3.112)$$

On note que le dernier terme dans la somme est positif, donc, en utilisant Chapman Kolmogorov,

$$\begin{aligned} \nu(y) & \geq P_x(X_1 = y) + P_x(X_1 \neq x, X_2 = y) + \cdots + P_x(X_1 \neq x, \dots, X_{k-1} \neq x, X_k = y) \\ & = \sum_{i=1}^k P_x(X_i = y, i < T_x). \end{aligned} \quad (3.113)$$

Comme cette majoration vaut pour tout  $k \geq 1$ , on en déduit en faisant tendre  $k$  vers  $+\infty$  que

$$\begin{aligned} \nu(y) & \geq \sum_{k=1}^{+\infty} P_x(X_k = y, k < T_x) \\ & = E_x\left(\sum_{k=1}^{T_x-1} 1_{X_k=y}\right) \\ & = \mu_x(y). \end{aligned} \quad (3.114)$$

□

### 3.4.2 Probabilités invariantes

On s'intéresse dans cette partie à l'existence et l'unicité de probabilités invariantes, c'est-à-dire de mesures stationnaires de masse 1. On commence par le cas où **l'espace d'états est de cardinal fini**.

**Théorème 10.** *On suppose que  $|E| < +\infty$ . Alors*

1. *il existe une probabilité invariante,*
2. *si de plus la chaîne est irréductible, alors cette probabilité est unique.*



*Preuve.* Commençons par le premier point relatif à l'existence. On considère une mesure de probabilité  $\nu$  sur  $E$  (par exemple  $\nu = \delta_x$  où  $x$  est un élément de  $E$ ). Pour tout  $n \geq 1$ , on définit la probabilité  $\nu_n$  par la somme de Césaro suivante

$$\nu_n = \frac{1}{n} \sum_{k=1}^n \nu Q^k. \quad (3.115)$$

La suite  $(\nu_n)_{n \geq 1}$  est à valeurs dans le compact  $[0, 1]^{|E|}$ , on peut donc extraire une sous-suite convergente  $(\nu_{\phi(n)})_{n \geq 1}$ . On note  $\pi$  sa limite. En particulier, pour tout  $x \in E$ ,  $\nu_{\phi(n)}(x) \rightarrow \pi(x)$  et on en déduit, en passant à la limite dans la somme (finie), que  $\sum_{x \in E} \pi(x) = 1$ . La mesure  $\pi$  est donc une probabilité et il nous reste à prouver qu'elle est invariante. Notons pour cela que pour tout  $n \geq 1$ ,

$$\nu_n Q = \frac{1}{n} \sum_{k=1}^n \nu Q^{k+1} = \nu_n + \frac{1}{n} (\nu Q^{n+1} - \nu Q). \quad (3.116)$$

Donc pour tout  $x \in E$ ,  $|(\nu_n Q - \nu_n)(x)| \leq 2/n$  et  $|(\nu_n Q - \nu_n)(x)| \rightarrow 0$  quand  $n \rightarrow +\infty$ . Comme il en va de même pour l'extraction, on obtient, en passant à la limite,  $\pi Q(x) = \pi(x)$  pour tout  $x \in E$ .

L'unicité pour le point 2. est une conséquence du Théorème 9 puisqu'il existe une unique probabilité parmi les multiples d'une mesure.  $\square$

On s'intéresse maintenant au cas où **l'espace d'états est infini (toujours avec l'hypothèse qu'il est dénombrable)**. On a besoin d'affiner la notion de récurrence pour pouvoir énoncer le résultat principal de cette section.

**Définition 13.** Soit  $x$  un état récurrent. On dit que  $x$  est

1. **récurrent nul** si  $E_x(T_x) = +\infty$ ,
2. **récurrent positif** si  $E_x(T_x) < +\infty$ .

**Théorème 11.** On suppose que la chaîne est irréductible. Alors les trois propositions suivantes sont équivalentes :

1. il existe un état récurrent positif
2. tous les états sont récurrents positifs
3. la chaîne admet une probabilité invariante  $\pi$ .

Dans ce cas, la chaîne est dite **récurrente positive**. De plus  $\pi$  est unique et pour tout  $x \in E$ ,

$$\pi(x) = \frac{1}{E_x(T_x)}. \quad (3.117)$$

*Preuve.* La stratégie de la preuve est d'établir la suite d'implications 1.  $\Rightarrow$  3.  $\Rightarrow$  2.  $\Rightarrow$  1. On ne montre en fait que les deux premières puisque la dernière est triviale.

1.  $\Rightarrow$  3. Soit  $x \in E$  récurrent positif, i.e.  $E_x(T_x) < +\infty$ . La mesure  $\mu_x$  est stationnaire d'après la Proposition 22. De plus  $\mu_x(E) = E_x(T_x) < +\infty$ . On peut donc renormaliser  $\mu_x$  en posant

$$\pi = \frac{\mu_x}{E_x(T_x)}, \quad (3.118)$$

de manière à obtenir une probabilité invariante.

3.  $\Rightarrow$  2. On suppose qu'il existe une probabilité invariante  $\pi$  et on veut montrer que la chaîne est récurrente positive. Nous allons procéder en deux étapes : montrer qu'elle est récurrente puis montrer qu'elle est récurrente positive.

(a) **La chaîne est récurrente.** On considère  $x \in E$ . On note que

$$E_\pi(N_x) = E_\pi\left(\sum_{k=0}^{+\infty} 1_{X_k=x}\right) = \sum_{k=0}^{+\infty} E_\pi(1_{X_k=x}) = \sum_{k=0}^{+\infty} \pi Q^k(x) = \sum_{k=0}^{+\infty} \pi(x). \quad (3.119)$$

Or  $\pi$  est une mesure non triviale puisque c'est une probabilité et la chaîne est irréductible donc, d'après la Proposition 23,  $\pi(x) > 0$  et on déduit que  $E_\pi(N_x) = +\infty$ . On doit en déduire  $E_x(N_x) = +\infty$ . Commençons par rappeler que

$$E_\pi(N_x) = \sum_{y \in E} \pi(y) E_y(N_x). \quad (3.120)$$

De plus, pour  $y \neq x$ , si  $P_y(T_x < +\infty) = 0$  on a  $E_y(N_x) = 0$  et sinon,  $P_y(\cdot | T_x < +\infty)$ -p.s.,  $N_x = n_x((X_{T_x+k})_{k \geq 0})$ , (et on notera que c'est faux si  $y = x$ ) et on obtient donc,

$$\begin{aligned} E_y(N_x) &= E_y(n_x((X_{T_x+k})_{k \geq 0}) | T_x < +\infty) P_y(T_x < +\infty) + E_y(N_x, T_x = +\infty) \\ &= E_x(N_x) P_y(T_x < +\infty) \\ &\leq E_x(N_x) \end{aligned} \quad (3.121)$$

où on a utilisé Markov fort au temps d'arrêt  $T_x$  pour la dernière égalité. Finalement, de (3.120) on obtient,

$$E_\pi(N_x) \leq \sum_{y \in E} \pi(y) E_x(N_x) = E_x(N_x), \quad (3.122)$$

et donc  $E_x(N_x) = +\infty$ .

- (b) **La chaîne est récurrente positive.** Comme  $x$  est récurrent, la mesure  $\mu_x$  est stationnaire. Comme la chaîne est irréductible et récurrente, et qu'aucune des mesures  $\mu_x$  et  $\pi$  n'est triviale, ces deux mesures sont proportionnelles. De plus  $\mu_x(x) = 1$ , on en déduit que

$$\mu_x = \frac{\pi}{\pi(x)}. \quad (3.123)$$

Or,

$$E_x(T_x) = \mu_x(E) = \frac{\pi(E)}{\pi(x)} = \frac{1}{\pi(x)} < +\infty, \quad (3.124)$$

car  $\pi(x) > 0$ . On en déduit que  $x$  est récurrent positif.

On a terminé la preuve des équivalences. L'unicité est une conséquence du Théorème 9 puisque la chaîne est récurrente et irréductible. Enfin l'égalité

$$E_x(T_x) = \frac{1}{\pi(x)} \quad (3.125)$$

vient de (3.124). □

**Proposition 24.** *Toute chaîne irréductible sur un espace d'états fini est récurrente positive.*

*Preuve.* C'est une conséquence du théorème précédent et du Théorème 10. □

**Fin du cours 2020 - 2021.**

## 3.5 Théorème ergodique et convergence à l'équilibre

### 3.5.1 Excursions

On considère une chaîne  $(X_n)_{n \geq 0}$  sur un espace d'états  $E$  et  $z \in E$  **un état récurrent**. On découpe la trajectoire du processus en différents morceaux de trajectoires (qui correspondent aux boucles de  $z$  à  $z$ ) de telle sorte que ces morceaux soient i.i.d. sous  $P_z(\cdot)$ . Commençons par définir les fonctions qui vont nous permettre de réaliser ce découpage.

On définit

$$\mathcal{T} = \{x \in E^{\mathbb{N}}, x_0 = z, \sum_{i=0}^{+\infty} 1_{\{X_i=z\}} = +\infty\}, \quad (3.126)$$

l'espace des trajectoires passant un infinité de fois en  $z$  et

$$\mathcal{U} = \{x \in E^{\mathbb{N}}, x_0 = z, t_z(x) < +\infty, x_i = z \quad \forall i \geq t_z(x)\}, \quad (3.127)$$

l'espace des boucles de  $z$  à  $z$ .

Pour  $i \geq 1$ , la  $i$ -ème excursion (de  $z$  à  $z$ ) d'une trajectoire est la boucle donnée par la fonction

$$\begin{aligned} b^{(i)} : \mathcal{T} &\rightarrow \mathcal{U} \\ x = (x_n)_{n \geq 0} &\mapsto (x_{(t_z^{(i-1)}(x)+n) \wedge t_z^{(i)}(x)})_{n \geq 0} \end{aligned} \quad (3.128)$$

avec toujours la convention  $\inf \emptyset = +\infty$ .

On définit enfin la famille de processus  $(B^{(i)})_{i \geq 1}$  par

$$(B_n^{(i)})_{n \geq 0} = (b^{(i)}((X_n)_{n \geq 0}))_{n \geq 0} \quad (3.129)$$

Pour  $i \geq 1$ , le processus  $B^{(i)}$  correspond à la boucle faite par  $(X_n)_{n \geq 0}$  entre  $T_z^{(i-1)}$  et  $T_z^{(i)}$ . On note que ces processus sont bien définis puisque, comme  $z$  est récurrent, les trajectoires de  $(X_n)_{n \geq 0}$  sont  $P_z$ -p.s. dans  $\mathcal{T}$ .

**Proposition 25.** *Soit  $z \in E$  un point récurrent. Sous  $P_z$ , les processus  $(B^{(i)})_{i \geq 1}$  sont i.i.d.*

*Preuve.* Soit  $k \geq 1$  et  $b_1, \dots, b_{k+1}$  des boucles de  $\mathcal{U}$ . Comme  $z$  est récurrent  $P_z(\cdot) = P_z(\cdot | T_z^{(k)} < +\infty, X_{T_z^{(k)}} = z)$ . De plus

$$\{B^{(k+1)} = b_{k+1}\} = \{B^{(1)}((X_{T^{(k)}+n})_{n \geq 0}) = b_{k+1}\}. \quad (3.130)$$

On obtient donc, en appliquant la propriété de Markov au temps d'arrêt  $T_z^{(k)}$ ,

$$\begin{aligned} P_z(B^{(k+1)} = b_{k+1}, \dots, B^1 = b_1) &= P_z(b^{(1)}((X_{T^{(k)}+n})_{n \geq 0}) = b_{k+1}, B^{(k)} = b_k, \dots, B^{(1)} = b_1) \\ &= P_z(B^{(1)} = b_{k+1}) P_z(B^{(k)} = b_k, \dots, B^{(1)} = b_1) \\ &= \prod_{i=1}^{k+1} P_z(B^{(1)} = b_i), \end{aligned} \quad (3.131)$$

où la dernière égalité est obtenue en itérant et en utilisant que les processus  $(B^{(i)})_{i \geq 1}$  sont i.i.d.  $\square$

### 3.5.2 Théorème ergodique

**Théorème 12.** Soit  $(X_n)_{n \geq 0}$  une chaîne de Markov irréductible récurrente positive. On note  $\pi$  son unique probabilité invariante. Soit  $F : E \rightarrow \mathbb{R}$  tel que  $F \in L^1(\pi)$  (i.e.  $E_\pi(|F|) < +\infty$ ). Alors (quelque soit la distribution initiale)

$$\frac{1}{n} \sum_{i=0}^{n-1} F(X_i) \rightarrow E_\pi(F) \quad p.s. \quad (3.132)$$

**Remarque 14.** 1. On peut reformuler la conclusion du théorème de la façon suivante :

On note pour tout  $n \geq 1$ ,  $\nu_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ . On a alors

$$\nu_n \xrightarrow{(loi)} \pi \quad p.s. \quad (3.133)$$

Mais on notera que cette reformulation n'est pas évidente puisqu'elle nécessite l'échange du p.s. et de la quantification sur  $F$  qui porte sur un espace non dénombrable.

2. En prenant  $F \equiv 1_{\{y\}}$  dans le théorème ergodique on obtient

$$\frac{1}{n} \sum_{i=0}^{n-1} 1_{\{X_i=y\}} \rightarrow \pi(y) \quad p.s. \quad (3.134)$$

Ce résultat indique donc que le temps moyen passé en  $y$  converge vers  $\pi(y)$ .

*Preuve.* On fixe un point  $x \in E$ . L'unique probabilité invariante s'écrit donc  $\pi = \mu_x / E_x(T_x)$ . On définit par ailleurs la fonction  $\Phi$

$$\begin{aligned} \Phi : \mathcal{T} &\rightarrow \mathbb{R} \\ (X_n)_{n \geq 0} &\mapsto \sum_{n=0}^{T_x-1} F(X_n). \end{aligned} \quad (3.135)$$

D'après la Proposition 25, les variables  $(\Phi(B^{(i)}))_{i \geq 1}$  sont i.i.d. sous  $P_x$ . Pour pouvoir appliquer la loi des grands nombres, il nous faut vérifier que  $\Phi(B) \in L^1$ .

$$\begin{aligned} E_x |\Phi(B)| &\leq E_x \sum_{n=0}^{T_x-1} |F|(X_n) \\ &= E_x \sum_{n=0}^{T_x-1} \sum_{y \in E} |F(y)| 1_{\{X_n=y\}} \\ &= \sum_{y \in E} |F(y)| E_x(N_y(T_x - 1)). \end{aligned} \quad (3.136)$$

Comme  $E_x(N_y(T_x - 1)) = \mu_x(y) = \pi(y)E_x(T_x)$ , on obtient

$$E_x|\Phi(B)| \leq E_x(T_x) \sum_{y \in E} |F(y)|\pi(y) = E_x(T_x)E_\pi|F| < +\infty. \quad (3.137)$$

Nous pouvons donc utiliser la loi des grands nombres. Avec un calcul similaire, on prouve que  $E_x(\Phi(B)) = E_\pi F$  et on obtient donc

$$\frac{1}{k} \sum_{i=0}^{T_x^{(k)}-1} F(X_i) = \frac{1}{k} \sum_{i=0}^k \Phi(B^{(i)}) \xrightarrow[k \rightarrow +\infty]{} E_x(T_x)E_\pi(F) \quad P_x - p.s. \quad (3.138)$$

En prenant la fonction  $F = 1$  dans l'équation précédente on obtient

$$\frac{1}{k} T_x^{(k)} \xrightarrow[k \rightarrow +\infty]{} E_x(T_x) \quad P_x - p.s. \quad (3.139)$$

et donc finalement

$$\frac{1}{T_x^{(k)}} \sum_{i=0}^{T_x^{(k)}-1} F(X_i) \xrightarrow[k \rightarrow +\infty]{} E_\pi(F) \quad P_x - p.s. \quad (3.140)$$

On a donc établi notre résultat sur la sous-suite des temps aléatoires  $(T_x^{(k)})_{k \geq 0} \dots$  il faut maintenant l'étendre à toute la suite en contrôlant les temps intermédiaires.

Pour  $n \geq 0$ , on note  $k(n)$  l'unique entier tel que

$$T_x^{(k(n))} \leq n < T_x^{(k(n)+1)}. \quad (3.141)$$

Cela signifie donc que  $n$  est dans la  $(k(n) + 1)$ -ième excursions. On notera que  $k(n)$  est aléatoire et que, comme  $x$  est récurrent,  $k(n) \rightarrow +\infty$   $P_x - p.s.$  On suppose dans un premier temps que  $F \geq 0$ , on a donc

$$\frac{1}{T_x^{(k(n)+1)}} \sum_{i=0}^{T_x^{(k(n))}-1} F(X_i) \leq \frac{1}{n} \sum_{i=0}^{n-1} F(X_i) \leq \frac{1}{T_x^{(k(n))}} \sum_{i=0}^{T_x^{(k(n)+1)}-1} F(X_i). \quad (3.142)$$

Comme les variables  $(T_x^{(k+1)} - T_x^{(k)})_{k \geq 0}$  sont i.i.d. sous  $P_x$  et que  $E_x(T_x) < +\infty$ , on obtient par la loi des grands nombres que,  $\frac{T_x^{(k(n)+1)}}{T_x^{(k(n))}} \xrightarrow[n \rightarrow +\infty]{} 1$  p.s. En utilisant (3.140), le terme de gauche converge donc vers  $E_\pi(F)$ . Avec le même argument on obtient la convergence vers la même limite pour le terme de droite. Finalement, par encadrement,

$$\frac{1}{n} \sum_{i=0}^{n-1} F(X_i) \xrightarrow[n \rightarrow +\infty]{} E_\pi(F). \quad (3.143)$$

Comme la limite ne dépend pas de  $x$  choisit arbitrairement en début de démonstration, on en déduit que cette convergence vaut pour toute distribution initiale  $\mu$  :

$$\begin{aligned}
 P_\nu \left( \frac{1}{n} \sum_{i=0}^{n-1} F(X_i) \xrightarrow{n \rightarrow +\infty} E_\pi(F) \right) &= \sum_{x \in E} \nu(x) P_x \left( \frac{1}{n} \sum_{i=0}^{n-1} F(X_i) \xrightarrow{n \rightarrow +\infty} E_\pi(F) \right) \\
 &= \sum_{x \in E} \nu(x) \\
 &= 1.
 \end{aligned}
 \tag{3.144}$$

On conclut dans le cas général en décomposant  $F = F^+ - F^-$ .

□

### 3.5.3 Convergence à l'équilibre des chaînes apériodiques

À compléter.