

CHAÎNES DE MARKOV & MARTINGALES

(CMMA)

Jean-Christophe BRETON

M1 maths fonda Université de Rennes 1



CHAPITRE 1 – CONDITIONNEMENT _____	1	4.3 Uniforme intégrabilité	26
1.1 Probabilité sachant un événement	1	4.4 Convergence L^1 et martingale fermée	28
1.2 Espérance conditionnelle	1	4.5 Martingales de carré intégrable	29
1.3 Lois conditionnelles	2	4.6 Théorème d'arrêt	31
CHAPITRE 2 – ESPÉRANCE CONDITIONNELLE _____	4	CHAPITRE 5 – DYNAMIQUE MARKOVIENNE _____	33
2.1 Introduction	4	5.1 Probabilité de transition	34
2.2 Exemples et premières propriétés	4	5.2 Exemples	35
2.3 Espérance conditionnelle dans le cas L^2	7	5.3 Premières propriétés	35
2.4 Cadre gaussien	8	5.4 Chaîne de MARKOV canonique	38
2.5 Probabilités conditionnelles	9	5.5 Propriétés de MARKOV	40
2.6 Lois conditionnelles	9	CHAPITRE 6 – RÉCURRENCE ET TRANSIENGE _____	43
CHAPITRE 3 – MARTINGALES ET FILTRATIONS _____	14	6.1 États récurrents et transitoires	44
3.1 Filtration et mesurabilité	14	6.2 Ensembles clos et irréductibilité	48
3.2 Temps d'arrêt	14	6.3 Classes de récurrence	50
3.3 Martingales, sous-martingales et sur-martingales	16	6.4 Absorption dans les classes de récurrence	52
3.4 Propriétés des martingales	17	CHAPITRE 7 – INVARIANCE ET ÉQUILIBRE _____	53
3.5 Martingale arrêtée	18	7.1 Mesures invariantes	53
3.6 Décomposition de DOOB	20	7.2 Invariance et récurrence	56
CHAPITRE 4 – CONVERGENCES DE MARTINGALE _____	22	7.3 Périodicité et forte réductibilité	60
4.1 Inégalités de martingale	22	7.4 Équilibre d'une chaîne de MARKOV	62
4.2 Convergence presque sûre de martingales	25	7.5 Théorème ergodique	63

Chapitre 1

CONDITIONNEMENT

1.1 Probabilité sachant un événement	1	1.3 Lois conditionnelles	2
1.2 Espérance conditionnelle	1	1.3.1 Cadre discret	3
		1.3.2 Cas à densité	3

Dans tout le cours, le triplet $(\Omega, \mathcal{F}, \mathbb{P})$ désignera un espace probabilisé.

1.1 PROBABILITÉ SACHANT UN ÉVÉNEMENT

DÉFINITION 1.1. Soit B un événement de probabilité non nulle. Pour tout événement A , on définit la probabilité de A sachant B comme la quantité

$$\mathbb{P}(A | B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

▷ **EXEMPLE.** On considère deux urnes notées U_1 et U_2 dont l'une possède 2 boules blanches (notées B) et 1 boule noire (notée N) et l'autre possède 1 boule blanche et 3 boules noires. Alors $\mathbb{P}(B | U_1) = 2/3$ et $\mathbb{P}(B | U_2) = 1/4$.

◇ **REMARQUE.** Si A et B sont deux événements indépendants, alors $\mathbb{P}(A | B) = \mathbb{P}(A)$.

PROPOSITION 1.2. Soit B un événement de probabilité non nulle. Alors la fonction d'ensemble

$$\mathbb{P}(\cdot | B) : \begin{cases} \mathcal{F} \longrightarrow [0, 1], \\ A \longmapsto \mathbb{P}(A | B) \end{cases}$$

est une probabilité.

Plus spécifiquement, il existe des propriétés concernant le conditionnement comme la règle des conditionnements successifs : si A_1, \dots, A_n sont des événements dont l'intersection est de probabilité non nulle, alors

$$\mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \mathbb{P}(A_2 | A_1) \mathbb{P}(A_3 | A_1 \cap A_2) \dots \mathbb{P}(A_n | A_1 \cap \dots \cap A_{n-1}).$$

Dans la suite, on notera $I \subset \mathbb{N}$ l'ensemble d'indexation et, en général, on aura $I = [1, n]$ avec $n \geq 1$ ou $I = \mathbb{N}$.

DÉFINITION 1.3. On appelle *système complet d'événement* toute suite dénombrable $(B_i)_{i \in I} \in \mathcal{F}^I$ d'événements deux à deux disjoints telle que $\sum_{i \in I} \mathbb{P}(B_i) = 1$.

PROPOSITION 1.4. Soit $(B_i)_{i \in I}$ un système complet tel que $\mathbb{P}(B_i) > 0$ pour tout $i \in I$. Alors pour tout événement A , on a

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A | B_i) \mathbb{P}(B_i).$$

Preuve Soit $\Omega_0 := \bigcup_{i \in I} B_i$. Comme $\mathbb{P}(\Omega_0) = 1$, on a $\mathbb{P}(A) = \mathbb{P}(A \cap \Omega_0)$ et il reste plus qu'à utiliser la σ -additivité de la probabilité \mathbb{P} . \square

▷ **EXEMPLE.** En reprenant l'exemple précédent et en supposant que l'on ait autant de chance de piocher dans une des deux urnes, on a $\mathbb{P}(B) = 2/3 \times 1/2 + 1/4 \times 1/2 = 11/24$.

PROPOSITION 1.5 (formule de BAYES). Soit $(B_i)_{i \in I}$ un système complet tel que $\mathbb{P}(B_i) > 0$ pour tout $i \in I$. Alors pour tout événement A non négligeable et tout $i \in I$, on a

$$\mathbb{P}(B_i | A) = \frac{\mathbb{P}(A | B_i) \mathbb{P}(B_i)}{\sum_{j \in I} \mathbb{P}(A | B_j) \mathbb{P}(B_j)}.$$

1.2 ESPÉRANCE CONDITIONNELLE

DÉFINITION 1.6. Soit B un événement de probabilité non nulle. On définit l'*espérance conditionnelle* sachant B l'espérance par rapport à la probabilité $\mathbb{P}(\cdot | B)$, notée $E[\cdot | B]$, c'est-à-dire si X est une variable aléatoire positive

ou intégrable, alors

$$\mathbb{E}[X | B] = \int_{\Omega} X(\omega) \mathbb{P}(d\omega | B).$$

PROPOSITION 1.7. Soient X une variable aléatoire intégrable et B un événement non négligeable. Alors

$$\mathbb{E}[X | B] = \frac{\mathbb{E}[X \mathbb{1}_B]}{\mathbb{P}(B)}.$$

Preuve Cette proposition est vraie si X est une indicatrice d'un événement. Par linéarité, elle reste vraie si X est étagée. Si X est une variable aléatoire positive, il suffit de l'approcher par une suite croissante de variables aléatoires étagées et d'utiliser le théorème de convergence monotone. Enfin, si X est intégrable, on utilise ses parties positive et négative. \square

▷ EXEMPLE. Soit X une variable aléatoire discrète dont on note le support $\{x_i | i \in I\}$. Alors pour tout événement non négligeable B , on a

$$\mathbb{E}[X | B] = \sum_{i \in I} x_i \mathbb{P}(X = x_i | B).$$

Soit Y une autre variable aléatoire discrète dont on note le support $\{y_j | j \in J\}$. Pour tout $j \in J$, on définit de cette façon

$$\mathbb{E}[X | Y = y_j] := \sum_{i \in I} x_i \mathbb{P}(X = x_i | Y = y_j).$$

De manière plus général, on peut définir l'espérance de X sachant Y .

DÉFINITION 1.8. Soient X un variable aléatoire intégrable et Y une variable aléatoire discrète dont on note $\{y_j | j \in J\}$ le support. On définit l'espérance conditionnelle de X sachant Y comme la variable aléatoire

$$\mathbb{E}[X | Y] := \sum_{j \in J} \mathbb{E}[X | Y = y_j] \mathbb{1}_{\{Y=y_j\}}.$$

◇ REMARQUE. Avec la définition précédente, on a $\mathbb{E}[\mathbb{E}[X | Y]] = \mathbb{E}[X]$. Il suffit d'utiliser la linéarité de l'espérance et la proposition précédente, on obtient

$$\begin{aligned} \mathbb{E}[\mathbb{E}[X | Y]] &= \mathbb{E}\left[\sum_{j \in J} \mathbb{E}[X | Y = y_j] \mathbb{1}_{\{Y=y_j\}}\right] \\ &= \sum_{j \in J} \mathbb{E}[X | Y = y_j] \mathbb{P}(\mathbb{1}_{\{Y=y_j\}}) \\ &= \sum_{j \in J} \frac{\mathbb{E}[X \mathbb{1}_{\{Y=y_j\}}]}{\mathbb{P}(Y = y_j)} \mathbb{P}(Y = y_j) \\ &= \mathbb{E}\left[X \sum_{j \in J} \mathbb{1}_{\{Y=y_j\}}\right] = \mathbb{E}(X). \end{aligned}$$

Précédemment, on a défini les quantités $\mathbb{P}(\cdot | B)$ pour des événements non négligeables B . On peut aussi définir la quantité $\mathbb{P}(\cdot | Y)$ pour une variable aléatoire discrète Y : on obtient une probabilité aléatoire.

NOTATION. On note $L^0(\Omega, \mathcal{F}, \mathbb{P})$ l'ensemble des variables aléatoires.

DÉFINITION 1.9. Soit Y une variable aléatoire discrète dont on note $\{y_j | j \in J\}$ le support. On définit la *probabilité conditionnelle* sachant Y comme la fonction d'ensemble

$$\mathbb{P}(\cdot | Y) : \begin{cases} \mathcal{F} \longrightarrow L^0(\Omega, \mathcal{F}, \mathbb{P}), \\ A \longmapsto \sum_{j \in J} \mathbb{P}(A | Y = y_j) \mathbb{1}_{\{Y=y_j\}}. \end{cases}$$

On a alors la relation $\mathbb{P}(A | Y) = \mathbb{E}[\mathbb{1}_A | Y]$ pour tout événement A .

1.3 LOIS CONDITIONNELLES

NOTATION. Pour toute variable aléatoire X , on note $\text{Sup } X$ son support.

1.3.1 Cadre discret

DÉFINITION 1.10. Soit (X, Y) un couple de variables aléatoires discrètes. Pour tout atome $y \in \text{Sup } Y$ (i. e. $\mathbb{P}(Y = y) \neq 0$), on appelle *loi conditionnelle* de X sachant $Y = y$ l'application

$$\mathbb{P}(X = \cdot | Y = y) : \begin{cases} \text{Sup } X \longrightarrow [0, 1], \\ x \longmapsto \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}. \end{cases}$$

Si $y \notin \text{Sup } Y$, on peut arbitrairement poser $\mathbb{P}(X = x | Y = y) := 0$ pour tout $x \in \text{Sup } X$.

◇ REMARQUE. Si X et Y sont indépendantes, alors la loi conditionnelle de X sachant Y est la loi de X .

1.3.2 Cas à densité

DÉFINITION 1.11. Soit (X, Y) un couple de variables aléatoires à densité f . Alors la variable aléatoire Y est de densité

$$y \longmapsto f_Y(y) := \int_{\mathbf{R}} f(x, y) dx$$

et de même pour la variable aléatoire X . Pour tout atome $y \in \text{Sup } y$, on appelle *loi conditionnelle* de X sachant $Y = y$ la loi de densité

$$x \longmapsto f_{X|Y=y}(x) := \frac{f(x, y)}{f_Y(y)}.$$

◇ REMARQUE. Si X et Y sont indépendantes, alors $f(x, y) = f_X(x)f_Y(y)$, donc $f_{X|Y=y}(x) = f_X(x)$ pour tous $x, y \in \mathbf{R}$

Chapitre 2

ESPÉRANCE CONDITIONNELLE

2.1 Introduction	4	2.4 Cadre gaussien	8
2.2 Exemples et premières propriétés	4	2.5 Probabilités conditionnelles	9
2.3 Espérance conditionnelle dans le cas L^2	7	2.6 Lois conditionnelles	9

2.1 INTRODUCTION

On considère \mathcal{G} une sous-tribu de \mathcal{F} . On commence par définir l'espérance conditionnelle sachant \mathcal{G} .

DÉFINITION-PROPOSITION 2.1. Soit X une variable aléatoire positive ou intégrable. On définit l'espérance conditionnelle de X sachant \mathcal{G} , notée $\mathbb{E}[X | \mathcal{G}]$, toute variable presque sûrement unique Y telle que

- (i) la variable Y est \mathcal{G} -mesurable;
- (ii) pour tout $A \in \mathcal{G}$, on a $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A]$.

◊ **REMARQUE.** Attention, l'espérance conditionnelle est définie presque sûrement. Il faut que la variable aléatoire X soit positive ou intégrable pour que la quantité $\mathbb{E}[X | \mathcal{G}]$ soit bien définie.

Preuve de l'existence et de l'unicité • **Existence.** On suppose d'abord $X \geq 0$. Pour $A \in \mathcal{G}$, on pose $\mathbf{Q}(A) := \mathbb{E}[X\mathbb{1}_A]$. L'application $\mathbf{Q} : \mathcal{G} \rightarrow [0, 1]$ est une probabilité vérifiant $\mathbf{Q} \ll \mathbb{P}$. Le théorème de RADON-NYKODYN assure alors qu'il existe une telle variable aléatoire Y .

On suppose désormais que X est intégrable. D'après le cas précédente, il existe deux variables aléatoires \mathcal{G} -mesurable Y^\pm telles que $\mathbb{E}[X^\pm \mathbb{1}_A] = \mathbb{E}[Y^\pm \mathbb{1}_A]$. Par linéarité, la variable aléatoire $Y := Y^+ - Y^-$ convient.

• **Unicité.** Soient Y et Y' deux telles variables aléatoires. Soit $\varepsilon > 0$. On pose $A_\varepsilon := \{Y - Y' > \varepsilon\} \in \mathcal{G}$. Avec le point (ii), on a alors $\mathbb{E}[\mathbb{1}_{A_\varepsilon}(Y - Y')] = 0$. Or $\mathbb{E}[\mathbb{1}_{A_\varepsilon}(Y - Y')] \geq \mathbb{E}[\mathbb{1}_{A_\varepsilon}\varepsilon] = \varepsilon\mathbb{P}(A_\varepsilon)$ ce qui implique $\mathbb{P}(A_\varepsilon) = 0$. Comme

$$\{Y - Y' > 0\} = \bigcup_{\varepsilon \in \mathbf{Q}_+} A_\varepsilon,$$

on en déduit $\mathbb{P}(Y - Y' > 0) = 0$. En se servant du cas symétrique, on en conclut $\mathbb{P}(Y - Y' \neq 0) = 0$. □

2.2 EXEMPLES ET PREMIÈRES PROPRIÉTÉS

- ▷ **EXEMPLES.** – Si X est un variable aléatoire \mathcal{G} -mesurable, alors $\mathbb{E}[X | \mathcal{G}] = X$.
- Pour toute constante c , on a $\mathbb{E}[c | \mathcal{G}] = c$.
- Si X est indépendante de la tribu \mathcal{G} , c'est-à-dire $\mathbb{P}(X \in A, G) = \mathbb{P}(X \in A)\mathbb{P}(G)$ pour tout $A \in \mathcal{B}(\mathbf{R})$ et tout $G \in \mathcal{G}$, alors $\mathbb{E}[X | \mathcal{G}] = \mathbb{E}[X]$. En effet, la variable aléatoire $Y := \mathbb{E}[X]$ vérifie bien les points (i) et (ii).

PROPOSITION 2.2. La condition (ii) est équivalente à avoir :

- (*) pour toute variable aléatoire \mathcal{G} -mesurable Z , on a $\mathbb{E}[XZ] = \mathbb{E}[YZ]$.

Preuve Pour passer de (*) à (ii), il suffit de prendre $Z = \mathbb{1}_A$ avec $A \in \mathcal{G}$. Réciproquement, de façon standard, on montre (*) pour des indicatrices, puis des variables étagées, puis des variable \mathcal{G} -mesurable. □

NOTATIONS. Pour deux variables aléatoires X et Y , on notera $\mathbb{E}[X | Y] := \mathbb{E}[X | \sigma(Y)]$ où la classe de parties

$$\sigma(Y) := \{Y^{-1}(B) | B \in \mathcal{B}(\mathbf{R})\}$$

est la tribu engendrée par Y . Pour un événement A , on notera $\mathbb{P}(A | \mathcal{G}) := \mathbb{E}[\mathbb{1}_A | \mathcal{G}]$.

- ▷ **EXEMPLE (conditionnement et partition).** On suppose avoir une partition dénombrable $\{\Omega_i\}_{i \in I}$ de Ω en ensembles non négligeables. Pour la tribu $\mathcal{G} := \sigma(\Omega_i | i \in I)$, on a

$$\mathbb{E}[X | \mathcal{G}] = \sum_{i \in I} \frac{\mathbb{E}[X\mathbb{1}_{\Omega_i}]}{\mathbb{P}(\Omega_i)} \mathbb{1}_{\Omega_i}.$$

En effet, montrons la relation dans le cas d'une partition $\{B, B^c\}$ avec $0 < \mathbb{P}(B) < 1$. Notons

$$Y := \frac{\mathbb{E}[X\mathbb{1}_B]}{\mathbb{P}(B)} \mathbb{1}_B + \frac{\mathbb{E}[X\mathbb{1}_{B^c}]}{\mathbb{P}(B^c)} \mathbb{1}_{B^c}$$

et montrons que $Y = \mathbb{E}[X | \mathcal{G}]$. La variable Y est bien \mathcal{G} -mesurable et la condition (ii) est vérifiée : on le montre pour $A \in \{\emptyset, B, B^c, \Omega\}$.

PROPOSITION 2.3. Si X est un variable aléatoire \mathcal{F} -intégrable, alors la variable $\mathbb{E}[X | \mathcal{G}]$ est G -intégrable.

Preuve On note $Y := \mathbb{E}[X | \mathcal{G}]$ et $A := \{Y > 0\} \in \mathcal{G}$. La condition (ii) donne

$$\begin{aligned} \mathbb{E}[Y \mathbb{1}_A] &= \mathbb{E}[X \mathbb{1}_A] \leq \mathbb{E}[|X| \mathbb{1}_A] \\ \text{et } \mathbb{E}[-Y \mathbb{1}_{A^c}] &= -\mathbb{E}[X \mathbb{1}_{A^c}] = \mathbb{E}[-X \mathbb{1}_{A^c}] \leq \mathbb{E}[|X| \mathbb{1}_{A^c}]. \end{aligned}$$

Comme $|Y| = Y \mathbb{1}_A - Y \mathbb{1}_{A^c}$, on en déduit $\mathbb{E}[|Y|] \leq \mathbb{E}[X \mathbb{1}_A] + \mathbb{E}[|X| \mathbb{1}_{A^c}] = \mathbb{E}[|X|]$ ce qui conclut. \square

PROPOSITION 2.4. Soient X et Y deux variables aléatoires indépendantes et f une fonction mesurable telles que $\mathbb{E}[|f(X, Y)|] < +\infty$. On pose $g := \mathbb{E}[f(X, \cdot)]$. Alors $\mathbb{E}[f(X, Y) | Y] = g(Y)$.

Preuve La fonction g est mesurable par le théorème de FUBINI, donc la variable $g(Y)$ est bien $\sigma(Y)$ -mesurable. Puis, on doit vérifier le point (ii). Soit $A \in \sigma(Y)$. On note $A = Y^{-1}(C)$ avec $C \in \mathcal{B}(\mathbf{R})$. Alors les théorèmes de transfert et de FUBINI donnent

$$\begin{aligned} \mathbb{E}[f(X, Y) \mathbb{1}_A] &= \mathbb{E}[\mathbb{1}_C(Y) f(X, Y)] = \iint_{\mathbf{R}^2} \mathbb{1}_C(y) f(x, y) \mathbb{P}_{(x, y)}(dx, dy) \\ &= \int_{\mathbf{R}} \mathbb{1}_C(y) \left(\int_{\mathbf{R}} f(x, y) \mathbb{P}_X(dx) \right) \mathbb{P}_Y(dy) \\ &= \mathbb{E}[\mathbb{1}_C(Y) g(Y)] = \mathbb{E}[g(Y) \mathbb{1}_A] \end{aligned}$$

ce qui montre le point (ii). \square

PROPOSITION 2.5 (linéarité). Soient X et Y deux variables aléatoires \mathcal{F} -intégrable et $a, b \in \mathbf{R}$. Alors

$$\mathbb{E}[aX + bY | \mathcal{G}] = a\mathbb{E}[X | \mathcal{G}] + b\mathbb{E}[Y | \mathcal{G}] \quad \text{presque sûrement.}$$

Preuve La variable aléatoire $a\mathbb{E}[X | \mathcal{G}] + b\mathbb{E}[Y | \mathcal{G}]$ est clairement \mathcal{G} -mesurable. Soit $A \in \mathcal{G}$. La linéarité de l'espérance donne

$$\begin{aligned} \mathbb{E}[(a\mathbb{E}[X | \mathcal{G}] + b\mathbb{E}[Y | \mathcal{G}]) \mathbb{1}_A] &= a\mathbb{E}[\mathbb{E}[X | \mathcal{G}] \mathbb{1}_A] + b\mathbb{E}[\mathbb{E}[Y | \mathcal{G}] \mathbb{1}_A] \\ &= a\mathbb{E}[X \mathbb{1}_A] + b\mathbb{E}[Y \mathbb{1}_A] = \mathbb{E}[(aX + bY) \mathbb{1}_A] \end{aligned}$$

ce qui assure le point (ii). \square

PROPOSITION 2.6 (monotonie). Soient X et Y deux variables aléatoires telles que $X \leq Y$. Alors $\mathbb{E}[X | \mathcal{G}] \leq \mathbb{E}[Y | \mathcal{G}]$.

Preuve Soit $A \in \mathcal{G}$. Comme $Y - X \geq 0$, on a

$$\mathbb{E}[\mathbb{1}_A(\mathbb{E}[X | \mathcal{G}] - \mathbb{E}[Y | \mathcal{G}])] = \mathbb{E}[\mathbb{1}_A \mathbb{E}[X - Y | \mathcal{G}]] = \mathbb{E}[\mathbb{1}_A(X - Y)] \leq 0.$$

Soit $\varepsilon > 0$. En prenant $A_\varepsilon := \{\mathbb{E}[X | \mathcal{G}] - \mathbb{E}[Y | \mathcal{G}] > \varepsilon\} \in \mathcal{G}$, on obtient $\varepsilon \mathbb{P}(A_\varepsilon) \leq 0$, donc $\mathbb{P}(A_\varepsilon) = 0$. Dès lors, on obtient $\mathbb{P}(\bigcup_{\varepsilon \in \mathbf{Q}_+^*} A_\varepsilon) = 0$ où $\bigcup_{\varepsilon \in \mathbf{Q}_+^*} A_\varepsilon = \{\mathbb{E}[X | \mathcal{G}] - \mathbb{E}[Y | \mathcal{G}] > 0\}$. \square

Les théorèmes classiques concernant l'espérance se généralise à l'espérance conditionnelle comme les inégalité de TCHEBYCHEV et de JENSEN et les théorèmes de convergences monotone et dominées.

COROLLAIRE 2.7 (inégalité de TCHEBYCHEV). Pour $\varepsilon > 0$, on a

$$\mathbb{P}(|X| > \varepsilon | \mathcal{G}) \leq \frac{\mathbb{E}[X^2 | \mathcal{G}]}{\varepsilon^2} \quad \text{presque sûrement.}$$

Preuve Il suffit d'appliquer la monotonie à l'inégalité $\varepsilon^2 \mathbb{1}_{\{|X| > \varepsilon\}} \leq X^2$. \square

THÉORÈME 2.8 (de convergence monotone conditionnel). Soit $(X_n)_{n \geq 1}$ une suite croissante de variables aléatoires \mathcal{F} -intégrables positives tendant presque sûrement vers une variable $X \in L^1(\mathcal{F})$. Alors $\mathbb{E}[X_n | \mathcal{G}] \rightarrow \mathbb{E}[X | \mathcal{G}]$ presque sûrement.

Preuve Pour $n \geq 1$, on pose $Y_n := X - X_n \geq 0$. Il suffit de montrer que la suite $(Z_n := \mathbb{E}[Y_n | \mathcal{G}])_{n \geq 1}$ décroît vers 0 presque sûrement. Comme la suite $(Y_n)_{n \geq 1}$ décroît, par monotonie, la suite $(Z_n)_{n \geq 1}$ décroît vers une variable aléatoire $Z_\infty \geq 0$. Montrons que $Z_\infty = 0$ presque sûrement. Soit $A \in \mathcal{G}$. Pour tout $n \geq 1$, on a $\mathbb{E}[Z_n \mathbb{1}_A] = \mathbb{E}[Y_n \mathbb{1}_A]$ et

$0 \leq Y_n \leq X \in L^1(\mathcal{F})$. Le théorème de convergence dominée pour l'espérance donne alors $\mathbb{E}[Z_n \mathbb{1}_A] \rightarrow 0$ et, de même, on a $\mathbb{E}[Z_n \mathbb{1}_A] \rightarrow \mathbb{E}[Z_\infty \mathbb{1}_A]$. Par unicité de la limite, on a $\mathbb{E}[Z_\infty \mathbb{1}_A] = 0$. Soit $\varepsilon > 0$. En prenant $A_\varepsilon := \{Z_\infty \geq \varepsilon\} \in \mathcal{G}$, on a $0 = \mathbb{E}[Z_\infty \mathbb{1}_A] \geq \varepsilon \mathbb{P}(A_\varepsilon)$. Comme précédemment, on obtient $\mathbb{P}(Z_\infty \neq 0) = 0$ ce qui termine la preuve. \square

LEMME 2.9 (FATOU conditionnel). Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires positives. Alors

$$\mathbb{E}[\liminf_{n \rightarrow +\infty} X_n \mid \mathcal{G}] \leq \liminf_{n \rightarrow +\infty} \mathbb{E}[X_n \mid \mathcal{G}] \quad \text{presque sûrement.}$$

THÉORÈME 2.10 (de convergence dominée conditionnel). Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires dominées par une variable $Z \in L^1(\mathcal{F})$ et tendant presque sûrement vers une variable $X \in L^1(\mathcal{F})$. Alors $\mathbb{E}[X_n \mid \mathcal{G}] \rightarrow \mathbb{E}[X \mid \mathcal{G}]$.

PROPOSITION 2.11 (inégalité de JENSEN conditionnelle). Soient $X \in L^1(\mathcal{F})$ et $\varphi: \mathbf{R} \rightarrow \mathbf{R}$ une fonction convexe telles que $\mathbb{E}[|\varphi(X)|] < +\infty$. Alors $\varphi(\mathbb{E}[X \mid \mathcal{G}]) \leq \mathbb{E}[\varphi(X) \mid \mathcal{G}]$.

Preuve C'est immédiat si la fonction φ est affine. De façon générale, on pose

$$S := \{(a, b) \in \mathbf{Q}^2 \mid \forall x \in \mathbf{R}, ax + b < \varphi(x)\}.$$

Pour tout $x \in \mathbf{R}$, on a $\varphi(x) = \sup_{(a,b) \in S} (ax + b)$. Comme $\varphi(X) \geq aX + b$ pour $(a, b) \in S$, on a

$$\mathbb{E}[\varphi(X) \mid \mathcal{G}] \geq \mathbb{E}[aX + b \mid \mathcal{G}] = a\mathbb{E}[X \mid \mathcal{G}] + b \quad \text{presque sûrement.}$$

Comme S est dénombrable, on a

$$\mathbb{E}[\varphi(X) \mid \mathcal{G}] \geq \sup_{(a,b) \in S} (a\mathbb{E}[X \mid \mathcal{G}] + b) = \varphi(\mathbb{E}[X \mid \mathcal{G}]) \quad \text{presque sûrement.} \quad \square$$

PROPOSITION 2.12. Soit $X \in L^1(\mathcal{F})$. Alors $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[X]$.

Preuve Il suffit de prendre $A = \Omega$ et d'appliquer le point (ii). \square

PROPOSITION 2.13 (inégalité de CAUCHY-SCHWARZ conditionnelle). Soient $X, Y \in L^1(\mathcal{F})$. Alors

$$\mathbb{E}[XY \mid \mathcal{G}]^2 \leq \mathbb{E}[X^2 \mid \mathcal{G}]\mathbb{E}[Y^2 \mid \mathcal{G}] \quad \text{presque sûrement.}$$

Preuve Soit $\theta \in \mathbf{R}$. Presque sûrement, on a $(X + \theta Y)^2 \geq 0$, donc $\mathbb{E}[(X + \theta Y)^2] \geq 0$, donc

$$\mathbb{E}[Y^2 \mid \mathcal{G}]\theta^2 + 2\mathbb{E}[XY \mid \mathcal{G}]\theta + \mathbb{E}[X^2 \mid \mathcal{G}] \geq 0.$$

Presque sûrement, on a donc

$$\forall \theta \in \mathbf{Q}, \quad \mathbb{E}[Y^2 \mid \mathcal{G}]\theta^2 + 2\mathbb{E}[XY \mid \mathcal{G}]\theta + \mathbb{E}[X^2 \mid \mathcal{G}] \geq 0.$$

Le discriminant de ce polynôme doit être négatif presque sûrement ce qui donne l'inégalité cherchée. \square

PROPOSITION 2.14. Soient $X \in L^1(\mathcal{F})$ et $p \geq 1$. Alors $|\mathbb{E}[X \mid \mathcal{G}]|^p \leq \mathbb{E}[|X|^p \mid \mathcal{G}]$.

Preuve Il suffit d'appliquer l'inégalité de JENSEN à la fonction convexe $x \mapsto |x|^p$. \square

PROPOSITION 2.15 (contraction). Soient $X \in L^1(\mathcal{F})$ et $p \geq 1$. Alors $\mathbb{E}[|\mathbb{E}[X \mid \mathcal{G}]|^p] \leq \mathbb{E}[|X|^p]$. Autrement dit, l'espérance conditionnelle est une contraction sur L^p .

THÉORÈME 2.16 (cascade). Soient $X \in L^1(\mathcal{F})$ et \mathcal{G}_1 et \mathcal{G}_2 deux tribus telles que $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F}$. Alors

1. $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}_1] \mid \mathcal{G}_2] = \mathbb{E}[X \mid \mathcal{G}_1]$ presque sûrement;
2. $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}_2] \mid \mathcal{G}_1] = \mathbb{E}[X \mid \mathcal{G}_1]$ presque sûrement.

Preuve 1. La variable $\mathbb{E}[X \mid \mathcal{G}_1]$ est \mathcal{G}_1 -mesurable, donc elle est \mathcal{G}_2 -mesurable, donc son espérance sachant \mathcal{G}_2 est elle-même.

2. Pour tout $A \in \mathcal{G}_1$, on a $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}_1] \mathbb{1}_A] = \mathbb{E}[X \mathbb{1}_A] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}_2] \mathbb{1}_A]$. On en déduit le point 2. \square

◊ REMARQUE. Attention, en général, on n'a pas $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}_1] \mid \mathcal{G}_2] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}_2] \mid \mathcal{G}_1]$. Pour avoir un contre-exemple, on se place dans l'espace probabilisé $([0, 1], \mathcal{B}([0, 1]), \lambda)$. On pose $\mathcal{G}_1 := \sigma([0, 1/2])$ et $\mathcal{G}_2 := \sigma([0, 1/3])$ et on considère la variable $X := \mathbb{1}_{[1/4, 3/4]}$.

THÉORÈME 2.17. Soient $Y \in L^1(\mathcal{F})$ et X une variable aléatoire telle que $XY \in L^1(\mathcal{F})$. Si X est \mathcal{G} -mesurable, alors $\mathbb{E}[XY | \mathcal{G}] = X\mathbb{E}[Y | \mathcal{G}]$.

Preuve D'abord, la variable $X\mathbb{E}[Y | \mathcal{G}]$ est bien \mathcal{G} -mesurable. De plus, montrons le point (ii). Dans le cas $X = \mathbb{1}_B$ avec $B \in \mathcal{G}$, pour tout $A \in \mathcal{G}$, on a

$$\mathbb{E}[\mathbb{1}_A X \mathbb{E}[Y | \mathcal{G}]] = \mathbb{E}[\mathbb{1}_A \mathbb{1}_B \mathbb{E}[Y | \mathcal{G}]] = \mathbb{E}[\mathbb{1}_{A \cap B} Y] = \mathbb{E}[\mathbb{1}_A XY].$$

Par linéarité, le point (ii) est vraie lorsque X est étagée. Maintenant, on suppose que Y est positive.

– Si X est positive, alors elle est limite simple d'une suite croissante $(X_n)_{n \geq 0}$ de variables aléatoires étagées. Comme $XY \in L^1(\mathcal{F})$, le théorème de convergence monotone conditionnelle donne $X_n \mathbb{E}[X_n Y | \mathcal{G}] \rightarrow \mathbb{E}[XY | \mathcal{G}]$. De plus, on a $X_n \mathbb{E}[Y | \mathcal{G}] \rightarrow X \mathbb{E}[Y | \mathcal{G}]$. Par unicité de la limite, on obtient $\mathbb{E}[XY | \mathcal{G}] = \mathbb{E}[Y | \mathcal{G}]$.

– Si X est de signe quelconque, on utilise les parties positive et négative de X , la linéarité de l'espérance conditionnelle et le cas précédent.

Désormais, on suppose que Y est de signe quelconque. On procède comme dans le point précédent. Cela montre le point (**). \square

THÉORÈME 2.18 (conditionnement et indépendance). Soit $X \in L^1(\mathcal{F})$.

1. Si X est indépendante de la tribu \mathcal{G} , alors $\mathbb{E}[X | \mathcal{G}] = \mathbb{E}[X]$.

2. Soient \mathcal{G} et \mathcal{H} deux sous-tribus de \mathcal{F} et X une variable aléatoire telle que \mathcal{H} soit indépendante de $\sigma(X, \mathcal{G})$. Alors $\mathbb{E}[X | \sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X | \mathcal{G}]$.

◇ **REMARQUE.** Pour le point 2, si la condition $\mathcal{H} \perp \sigma(X, \mathcal{G})$ n'est pas satisfaite, alors la conclusion n'est pas valable. En effet, prenons ε_1 et ε_2 deux variables aléatoires de RADEMACHER indépendantes. On pose $X := \varepsilon_1 \varepsilon_2$, puis $\mathcal{H} := \sigma(\varepsilon_1)$ et $\mathcal{G} := \sigma(\varepsilon_2)$. Alors $\mathcal{H} \perp \mathcal{G}$ et on montre $\mathcal{H} \perp \sigma(X)$. Pourtant, on n'a pas $\mathcal{H} \perp \sigma(X, \mathcal{G})$ car la variable $\varepsilon_1 = X/\varepsilon_2$ doit être \mathcal{G} -mesurable et donc

$$\mathbb{E}[X | \mathcal{G}] = \mathbb{E}[\varepsilon_1 \varepsilon_2 | \mathcal{G}] = \varepsilon_2 \mathbb{E}[\varepsilon_1 | \mathcal{G}] = \varepsilon_2 \mathbb{E}[\varepsilon_1] = 0 \quad \text{et} \quad \mathbb{E}[X | \sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[\varepsilon_1 \varepsilon_2 | \sigma(\varepsilon_1 \varepsilon_2)] = \varepsilon_1 \varepsilon_2 \neq 0.$$

Preuve Montrons le second point. On utilise un argument de classe monotone. On note $Y := \mathbb{E}[X | \mathcal{G}]$ et

$$\mathcal{M} := \{A \in \mathcal{F} \mid \mathbb{E}[Y \mathbb{1}_A] = \mathbb{E}[X \mathbb{1}_A]\}.$$

Par définition de y , on a $\mathcal{G} \subset \mathcal{M}$. De plus, on a $\mathcal{H} \subset \mathcal{M}$ car, pour tout événement $A \in \mathcal{G}$, on a $\mathbb{E}[Y \mathbb{1}_A] = \mathbb{E}[Y] \mathbb{P}(A)$ car $Y \perp A$ et de même pour X . Montrons que la classe \mathcal{M} est monotone.

– Comme $\mathbb{E}[Y] = \mathbb{E}[X]$, on a $\Omega \in \mathcal{M}$.

– Pour tous événements $A, B \in \mathcal{M}$ tels que $B \subset A$, on a $A \setminus B \in \mathcal{M}$. En effet, on a

$$\mathbb{E}[Y \mathbb{1}_{A \setminus B}] = \mathbb{E}[Y(\mathbb{1}_A - \mathbb{1}_B)] = \mathbb{E}[Y \mathbb{1}_A] - \mathbb{E}[Y \mathbb{1}_B] = \mathbb{E}[X \mathbb{1}_A] - \mathbb{E}[X \mathbb{1}_B] = \mathbb{E}[X \mathbb{1}_{A \setminus B}].$$

– Pour toute suite croissante $(A_i)_{i \geq 1}$, on a $\bigcup_{i \geq 1} A_i \in \mathcal{M}$. En effet, le théorème de convergence dominée assure

$$\mathbb{E}[Y \mathbb{1}_{\bigcup_{i \geq 1} A_i}] = \mathbb{E}[Y \lim_{i \rightarrow +\infty} \mathbb{1}_{A_i}] = \lim_{i \rightarrow +\infty} \mathbb{E}[Y \mathbb{1}_{A_i}] = \lim_{i \rightarrow +\infty} \mathbb{E}[X \mathbb{1}_{A_i}] = \mathbb{E}[X \mathbb{1}_{\bigcup_{i \geq 1} A_i}].$$

On note maintenant

$$\mathcal{P} := \{B \cap C \mid B \in \mathcal{G}, C \in \mathcal{H}\}.$$

Il s'agit d'un π -système inclus dans \mathcal{M} . En effet, pour tous $B \in \mathcal{G}$ et $C \in \mathcal{H}$, on a

$$\mathbb{E}[Y \mathbb{1}_{B \cap C}] = \mathbb{E}[Y \mathbb{1}_B \mathbb{1}_C] = \mathbb{E}[Y \mathbb{1}_B] \mathbb{E}[\mathbb{1}_C] = \mathbb{E}[X \mathbb{1}_B] \mathbb{E}[\mathbb{1}_C] = \mathbb{E}[X \mathbb{1}_{B \cap C}]$$

car $Y \mathbb{1}_B$ est \mathcal{G} -mesurable et $\mathcal{G} \perp \mathcal{H}$, puis $X \mathbb{1}_B$ est $\sigma(X, \mathcal{G})$ -mesurable et $\sigma(X, \mathcal{G}) \perp \mathcal{G}$. Le lemme des classes monotones assure $\sigma(\mathcal{P}) \subset \mathcal{M}$. On conclut en observant que $\sigma(\mathcal{P}) \supset \sigma(\mathcal{G}, \mathcal{H})$. Ainsi, on obtient que $\sigma(\mathcal{G}, \mathcal{H}) \subset \mathcal{M}$ ce qui montre $Y = \mathbb{E}[X | \sigma(\mathcal{G}, \mathcal{H})]$. Montrons alors que $\sigma(\mathcal{P}) \subset \mathcal{M}$. \square

◇ **REMARQUE.** On peut même montrer que $\sigma(\mathcal{P}) = \sigma(\mathcal{G}, \mathcal{H})$. En effet, pour tous $B \in \mathcal{G}$ et $C \in \mathcal{H}$, ces deux éléments appartiennent à $\mathcal{G} \cup \mathcal{H}$ et donc à $\sigma(\mathcal{G} \cup \mathcal{H})$, donc $B \cap C \in \sigma(\mathcal{G} \cup \mathcal{H})$.

2.3 ESPÉRANCE CONDITIONNELLE DANS LE CAS L^2

Dans le cas L^2 , on interprète l'espérance conditionnelle comme une projection orthogonale par rapport au produit scalaire défini par la relation $\langle X, Y \rangle := \mathbb{E}[XY]$. On note $\| \cdot \|$ la norme associée à ce produit scalaire.

THÉORÈME 2.19. Soient $X \in L^2(\mathcal{F})$ et \mathcal{G} une sous-tribu de \mathcal{F} . Alors l'espérance conditionnelle de X sachant \mathcal{G} est la projection orthogonale de X sur $L^2(\mathcal{G})$.

Preuve On note $Y := \mathbb{E}[X | \mathcal{G}]$. Comme $X \in L^2(\mathcal{F})$, la proposition 2.15 assure $Y \in L^2(\mathcal{G})$. Pour toute $Z \in L^2(\mathcal{G})$, on a

$$\begin{aligned} \|X - Z\|^2 &= \mathbb{E}[(X - Z)^2] = \mathbb{E}[(X - Y + Y - Z)^2] \\ &= \mathbb{E}[(X - Y)^2] + 2\mathbb{E}[(X - Y)(Y - Z)] + \mathbb{E}[(Y - Z)^2] \\ &= \mathbb{E}[(X - Y)^2] + 2\mathbb{E}[\mathbb{E}[(X - Y)(Y - Z) | \mathcal{G}]] + \mathbb{E}[(Y - Z)^2] \\ &= \mathbb{E}[(X - Y)^2] + 2\mathbb{E}[(Y - Z)\mathbb{E}[X - Y | \mathcal{G}]] + \mathbb{E}[(Y - Z)^2] \\ &= \mathbb{E}[(X - Y)^2] + \mathbb{E}[(Y - Z)^2] = \|X - Y\|^2 + \|Y - Z\|^2. \end{aligned}$$

On en déduit

$$d(X, L^2(\mathcal{G})) = \inf_{Z \in L^2(\mathcal{G})} \|X - Z\| = \|X - Y\|$$

ce qui montre le théorème. □

Variance conditionnelle

DÉFINITION 2.20. Soit \mathcal{G} une sous-tribu de \mathcal{F} . La *variance conditionnelle* d'une variable aléatoire $X \in L^2(\mathcal{F})$ sachant \mathcal{G} est la quantité

$$\text{Var}(X | \mathcal{G}) := \mathbb{E}[(X - \mathbb{E}[X | \mathcal{G}])^2 | \mathcal{G}].$$

PROPOSITION 2.21 (KÖNIG). Soient $X \in L^2(\mathcal{F})$. Alors $\text{Var}(X | \mathcal{G}) = \mathbb{E}[X^2 | \mathcal{G}] - \mathbb{E}[X | \mathcal{G}]^2$.

Preuve On a

$$\begin{aligned} \text{Var}(X | \mathcal{G}) &= \mathbb{E}[X^2 - 2X\mathbb{E}[X | \mathcal{G}] + \mathbb{E}[X | \mathcal{G}]^2 | \mathcal{G}] \\ &= \mathbb{E}[X^2 | \mathcal{G}] - 2\mathbb{E}[X\mathbb{E}[X | \mathcal{G}] | \mathcal{G}] + \mathbb{E}[\mathbb{E}[X | \mathcal{G}]^2 | \mathcal{G}] \\ &= \mathbb{E}[X^2 | \mathcal{G}] - 2\mathbb{E}[X | \mathcal{G}]\mathbb{E}[X | \mathcal{G}] + \mathbb{E}[X | \mathcal{G}]^2 \\ &= \mathbb{E}[X^2 | \mathcal{G}] - \mathbb{E}[X | \mathcal{G}]. \end{aligned}$$

□

PROPOSITION 2.22 (décomposition de la variance). Soit $X \in L^2(\mathcal{F})$. Alors

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X | \mathcal{G})] + \text{Var}(\mathbb{E}[X | \mathcal{G}]).$$

Preuve On a

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \mathbb{E}[(X - \mathbb{E}[X | \mathcal{G}] + \mathbb{E}[X | \mathcal{G}] - \mathbb{E}[X])^2] \\ &= \mathbb{E}[(X - \mathbb{E}[X | \mathcal{G}])^2] + 2\mathbb{E}[(X - \mathbb{E}[X | \mathcal{G}])(\mathbb{E}[X | \mathcal{G}] - \mathbb{E}[X])] + \mathbb{E}[(\mathbb{E}[X | \mathcal{G}] - \mathbb{E}[X])^2] \\ &= \mathbb{E}[\mathbb{E}[(X - \mathbb{E}[X | \mathcal{G}])^2 | \mathcal{G}]] + 2\mathbb{E}[\mathbb{E}[(X - \mathbb{E}[X | \mathcal{G}])(\mathbb{E}[X | \mathcal{G}] - \mathbb{E}[X]) | \mathcal{G}]] + \mathbb{E}[(\mathbb{E}[X | \mathcal{G}] - \mathbb{E}[X])^2] \\ &= \mathbb{E}[\text{Var}(X | \mathcal{G})] + 2\mathbb{E}[(\mathbb{E}[X | \mathcal{G}] - \mathbb{E}[X])\mathbb{E}[X - \mathbb{E}[X | \mathcal{G}] | \mathcal{G}]] + \mathbb{E}[(\mathbb{E}[X | \mathcal{G}] - \mathbb{E}[X])^2] \\ &= \mathbb{E}[\text{Var}(X | \mathcal{G})] + \mathbb{E}[(\mathbb{E}[X | \mathcal{G}] - \mathbb{E}[X])^2] \quad (\text{car } \mathbb{E}[X - \mathbb{E}[X | \mathcal{G}] | \mathcal{G}] = 0) \\ &= \mathbb{E}[\text{Var}(X | \mathcal{G})] + \mathbb{E}[(\mathbb{E}[X | \mathcal{G}] - \mathbb{E}[\mathbb{E}[X | \mathcal{G}])^2] \\ &= \mathbb{E}[\text{Var}(X | \mathcal{G})] + \text{Var}(\mathbb{E}[X | \mathcal{G}]). \end{aligned}$$

□

2.4 CADRE GAUSSIEN

Dans le cas gaussien, il suffit de projeter sur des droites pour obtenir $\mathbb{E}[X | Y]$.

PROPOSITION 2.23. 1. Soit (X, Y) un couple gaussien centré tel que $\text{Var}(Y) \neq 0$. Alors

$$\mathbb{E}[X | Y] = \frac{\text{Cov}(X, Y)}{\text{Var}(Y)} Y \quad \text{presque sûrement.}$$

2. Dans le cas non centré, on a

$$\mathbb{E}[X | Y] = \mathbb{E}[X] - \frac{\text{Cov}(X, Y)}{\text{Var}(Y)} \mathbb{E}[Y] + \frac{\text{Cov}(X, Y)}{\text{Var}(Y)} Y \quad \text{presque sûrement.}$$

3. Soient (Z_1, \dots, Z_d) un vecteur gaussien de covariance Σ et $a := (a_1, \dots, a_d), b := (b_1, \dots, b_d) \in \mathbf{R}^d$. On pose

$$X := \sum_{i=1}^d a_i Z_i \quad \text{et} \quad Y := \sum_{i=1}^d b_i Z_i.$$

Alors

$$\mathbb{E}[X | Y] = \frac{{}^t a \Sigma b}{{}^t b \Sigma b} Y \quad \text{presque sûrement.}$$

Preuve 1. On pose $c := \text{Cov}(X, Y) / \text{Var}(Y)$. Comme le couple (X, Y) est gaussien, le couple $(Y, X - cY)$ l'est aussi et il vérifie $\text{Cov}(Y, X - cY) = 0$. On en déduit $X \perp\!\!\!\perp X - cY$. Soit $A \in \sigma(Y)$ qu'on écrit $A = Y^{-1}(B)$ avec $B \in \mathcal{B}(\mathbf{R})$. On a

$$\begin{aligned} 0 &= \mathbb{E}[(X - cY)\mathbb{1}_A] = \mathbb{E}[\mathbb{E}[(X - cY)\mathbb{1}_A | Y]] \\ &= \mathbb{E}[\mathbb{1}_A \mathbb{E}[X - cY | Y]] \\ &= \mathbb{E}[\mathbb{1}_A \mathbb{E}[X - cY]] = 0. \end{aligned}$$

Cela montre que $cY = \mathbb{E}[X | Y]$. □

2.5 PROBABILITÉS CONDITIONNELLES

L'objectif est de définir une probabilité sachant une sous-tribu \mathcal{G} de \mathcal{F} . Souvent, cette sous-tribu sera celle engendrée par une variable aléatoire ce qui permettra de parler de probabilité sachant une variable aléatoire. Naturellement, on définit cela comme suivant.

DÉFINITION 2.24. On définit la probabilité sachant \mathcal{G} par la relation

$$\mathbb{P}(A | \mathcal{G}) := \mathbb{E}[\mathbb{1}_A | \mathcal{G}] \quad \text{presque sûrement,} \quad A \in \mathcal{F}.$$

Dans cette définition, notons que le « presque sûrement » dépend de l'événement $A \in \mathcal{F}$. De la même façon qu'au chapitre 1, l'application $\mathbb{P}(\cdot | B)$ avec $\mathbb{P}(B) > 0$ est une probabilité, on souhaite que l'application $\mathbb{P}(\cdot | \mathcal{G})$ en soit aussi une presque sûrement. Cependant, on n'a pas le fait que la variable $\mathbb{P}(\cdot | \mathcal{G})$ est presque sûrement σ -additive. En fait, un choix adéquat de chaque $\mathbb{P}(A | \mathcal{G})$ dans la classe d'équivalence de la relation « être égale presque sûrement » permet d'assurer qu'on peut considérer que l'application $\mathbb{P}(\cdot | \mathcal{G})$ est presque sûrement une probabilité : c'est la notion de *probabilité conditionnelle régulière*.

DÉFINITION 2.25. Une fonction $\mathbb{P}(\star | \mathcal{G})(\cdot)$ définie sur $\Omega \times \mathcal{F}$ est appelée *probabilité conditionnelle régulière* lorsqu'on a les deux conditions suivantes :

- (i) pour tout $A \in \mathcal{F}$, on a $\mathbb{P}(A | \mathcal{G})(\omega) = \mathbb{E}[\mathbb{1}_A | \mathcal{G}](\omega)$ pour \mathbb{P} -presque tout $\omega \in \Omega$;
- (ii) pour \mathbb{P} -presque tout $\omega \in \Omega$, la fonction $\mathbb{P}(\star | \mathcal{G})(\omega)$ est une mesure de probabilité sur (Ω, \mathcal{F}) .

À défaut de toujours exister, les probabilités conditionnelles régulières existent dans de nombreux cadres considérés en pratique.

THÉORÈME 2.26 (JIRINA). Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité polonais, *i. e.* métrique, complet et séparable, où la tribu \mathcal{F} est la tribu borélienne. Alors les probabilités conditionnelles régulières existent pour toute sous-tribu \mathcal{G} de \mathcal{F} et elles sont presque sûrement uniques, c'est-à-dire que, pour toutes probabilité conditionnelle régulières ν et ν' , on a

$$\nu(\cdot, \omega) = \nu'(\cdot, \omega) \quad \text{pour } \mathbb{P}\text{-presque tout } \omega \in \Omega.$$

2.6 LOIS CONDITIONNELLES

DÉFINITION 2.27. Soient (S, \mathcal{A}) et (T, \mathcal{B}) deux espaces mesurables. On appelle *noyau de probabilité* de S par T toute application $\nu: \mathcal{A} \times T \rightarrow [0, 1]$ telle que

- (i) pour tout $y \in T$, l'application $\nu(\star, y)$ est une probabilité sur (S, \mathcal{A}) ;

(ii) pour tout $A \in \mathcal{A}$, l'application $\sigma(A, \cdot)$ est \mathcal{B} -mesurable.

▷ EXEMPLES. – Soient μ une mesure σ -finie sur (S, \mathcal{A}) et $h: S \times T \rightarrow \mathbf{R}_+$ une fonction $\mathcal{A} \otimes \mathcal{B}$ -mesurable vérifiant

$$\int_S h(x, y) \mu(dx) = 1, \quad \forall y \in T.$$

Alors l'application

$$v: \begin{cases} \mathcal{A} \times T \rightarrow [0, 1], \\ (A, y) \mapsto \int_A h(x, y) \mu(dx) \end{cases}$$

est un noyau de probabilité de S par T . En effet, la σ -additivité pour montrer le point (i) est assuré par le théorème de convergence dominée. Le point (ii) est assuré par le théorème de FUBINI-TONELLI.

– • *Cas à densité.* Soit (X, Y) un couple de variables aléatoires sur \mathbf{R}^2 de densité $f: \mathbf{R}^2 \rightarrow \mathbf{R}$. On considère la fonction $h: \mathbf{R}^2 \rightarrow \mathbf{R}$ définie par la relation

$$h(x, y) = \begin{cases} f(x, y) / f_Y(y) & \text{si } f_Y(y) > 0, \\ 0 & \text{sinon,} \end{cases} \quad x, y \in \mathbf{R}.$$

Alors l'application

$$v: \begin{cases} \mathcal{B}(\mathbf{R}) \times \mathbf{R} \rightarrow [0, 1], \\ (A, y) \mapsto \int_A h(x, y) dx. \end{cases}$$

est un noyau de probabilité de \mathbf{R} par lui-même muni de sa tribu borélienne.

– • *Cas discret.* Soit (X, Y) un couple de variables aléatoires discrètes. On définit un noyau de probabilité par

$$v(A, y) = \begin{cases} \mathbb{P}(X \in A | Y = y) & \text{si } \mathbb{P}(Y = y) > 0, \\ 0 & \text{sinon.} \end{cases}$$

PROPOSITION 2.28. Soit v un noyau de probabilité de S par T .

1. Soit h une fonction mesurable positive sur (S, \mathcal{A}) . Alors la fonction

$$\varphi: \begin{cases} T \rightarrow \mathbf{R}, \\ y \mapsto \int_S h(x) v(dx, y) \end{cases}$$

est mesurable sur (T, \mathcal{B}) .

2. Soit η une probabilité sur (T, \mathcal{B}) . Alors l'application

$$\mu: \begin{cases} \mathcal{A} \rightarrow \mathbf{R}, \\ A \mapsto \int_T v(A, y) \eta(dy) \end{cases}$$

est une probabilité sur (S, \mathcal{A}) .

Preuve 1. C'est vraie si h est une indicatrice, puis ça l'est également pour toute fonction mesurable h par les arguments usuels de la théorie de la mesure.

2. L'application μ envoie clairement \emptyset sur 0 et elle est positive. Montrons qu'elle est σ -additive. Soit $(A_i)_{i \in \mathbf{N}}$ une suite d'éléments de \mathcal{A} deux à deux disjoints. Alors

$$\begin{aligned} \mu\left(\bigsqcup_{i \in \mathbf{N}} A_i\right) &= \int_T v\left(\bigsqcup_{i \in \mathbf{N}} A_i, y\right) \eta(dy) \\ &= \int_T \sum_{i \in \mathbf{N}} v(A_i, y) \eta(dy) && \text{(car } v(\star, y) \text{ est un probabilité)} \\ &= \sum_{i \in \mathbf{N}} \int_T v(A_i, y) \eta(dy) && \text{(par le théorème de convergence dominée)} \\ &= \sum_{i \in \mathbf{N}} \mu(A_i). \end{aligned}$$

D'où la σ -additivité. Enfin, on a

$$\mu(S) = \int_T v(S, y) \eta(dy) = \int_T \eta(dy) = \eta(T) = 1. \quad \square$$

DÉFINITION 2.29. Soient X et Y deux variables aléatoires respectivement à valeurs dans (S, \mathcal{A}) et (T, \mathcal{B}) . On appelle *loi conditionnelle* de X sachant Y tout noyau de probabilité $\nu: \mathcal{A} \times T \rightarrow [0, 1]$ tel que, pour toute fonction mesurable positive h sur S , on a

$$\mathbb{E}[h(X) | Y] = \int_S h(x) \nu(dx, Y).$$

◇ **REMARQUE.** Soit ν une loi conditionnelle de X sachant Y . Pour tout $A \in \mathcal{A}$, en prenant $h = \mathbb{1}_A$, on obtient

$$\mathbb{P}(X \in A | Y) = \mathbb{E}[\mathbb{1}_A(Y) | Y] = \nu(A, Y) \quad \text{presque sûrement.}$$

Alors l'application $A \mapsto \nu(A, Y)$ est bien une probabilité presque sûrement. En fait, par le théorème de JIRINA, la loi conditionnelle de X sachant Y existe si et seulement si les probabilités conditionnelles régulières de X sachant Y existe. De plus, ce même théorème donne l'unicité presque sûrement.

PROPOSITION 2.30. Soit (X, Y) un couple de variables aléatoires sur \mathbf{R}^2 de densité f . Alors la loi conditionnelle de X sachant Y existe et est donnée par le noyau

$$\nu(dx, y) = f(x, y) dx \quad \text{avec} \quad \tilde{f}(x, y) = \begin{cases} f(x, y) / f_Y(y) & \text{si } f_Y(y) > 0, \\ 0 & \text{sinon.} \end{cases}$$

Preuve Il s'agit de prouver que le noyau $\nu(dx, y) = \tilde{f}(x, y) dx$ vérifie bien la définition précédente. Pour cela, on établit que, pour toute fonction mesurable positive h et toute fonction g telles que $h(X)g(Y) \in L^1(\mathcal{F})$, on a

$$\mathbb{E}[h(X)g(Y)] = \mathbb{E}\left[\int_{\mathbf{R}} h(x) \nu(dx, Y) g(Y)\right]$$

ce qui montrera la proposition. En effet, soient h et g deux telles fonctions. Le théorème de transfert donne

$$\begin{aligned} \mathbb{E}\left[\int_{\mathbf{R}} h(x) \nu(dx, Y) g(Y)\right] &= \int_{\mathbf{R}} \left(\int_{\mathbf{R}} h(x) \nu(dx, y)\right) g(y) f_Y(y) dy \\ &= \int_{\mathbf{R}} \left(\int_{\mathbf{R}} h(x) \frac{f(x, y)}{f_Y(y)} dy\right) f_Y(y) dy \\ &= \int_{\mathbf{R}} \int_{\mathbf{R}} h(x) g(y) f(x, y) dx dy = \mathbb{E}[h(X)g(Y)]. \quad \square \end{aligned}$$

PROPOSITION 2.31. Soit (X, Y) un couple de variables aléatoires discrètes. Alors la loi conditionnelle de X sachant Y existe et est donnée par le noyau

$$\nu(A, y) = \begin{cases} \mathbb{P}(X \in A | Y = y) & \text{si } \mathbb{P}(Y = y) > 0, \\ 0 & \text{sinon.} \end{cases}$$

Preuve Procédons comme dans la preuve précédente. Pour $x, y \in \mathbf{R}$, on pose

$$\nu(x, y) := \begin{cases} \mathbb{P}(X = x | Y = y) & \text{si } \mathbb{P}(Y = y) > 0, \\ 0 & \text{sinon.} \end{cases}$$

Soient h une fonction mesurable positive et g une fonction telle que $h(X)g(Y) \in L^1(\mathcal{F})$. Alors

$$\begin{aligned} \mathbb{E}\left[\int_{\mathbf{R}} h(x) \nu(dx, Y) g(Y)\right] &= \sum_{y \in Y(\Omega)} \int_{\mathbf{R}} h(x) \nu(dx, y) g(y) \mathbb{P}(Y = y) \\ &= \sum_{y \in Y(\Omega)} \sum_{x \in X(\Omega)} h(x) \mathbb{P}(X = x | Y = y) \mathbb{P}(Y = y) g(y) \\ &= \sum_{y \in Y(\Omega)} \sum_{x \in X(\Omega)} h(x) g(y) \mathbb{P}(X = x, Y = y) = \mathbb{E}[h(X)g(Y)]. \quad \square \end{aligned}$$

Retour au conditionnement du type sachant $\{Y = y\}$

On suppose que la loi conditionnelle de X sachant Y existe. On la note $\nu: \mathcal{A} \times T \rightarrow [0, 1]$.

DÉFINITION 2.32. Pour \mathbb{P}_Y -presque tout $y \in \mathbf{R}$ et pour tout $A \in \mathcal{A}$, on pose

$$\mathbb{P}(X \in A | Y = y) := \nu(A, y).$$

◇ **REMARQUES.** – Si Y est discrète, alors cette définition coïncide bien avec celle du chapitre 1.

– Avec l’approche de ce chapitre, on donne un sens à la notation $\mathbb{P}(X \in A \mid Y = y)$ dès que la loi existe et même lorsque $\mathbb{P}(Y = y) = 0$. Dans ce dernier cas, on parle de probabilité conditionnelle *singulière*.

On suppose que la loi conditionnelle ν de X sachant Y existe. Alors la loi de X sachant Y est $\nu(\star, Y)$. Pour \mathbb{P}_Y -presque tout $y \in T$, on peut ensuite définir les espérances conditionnelles sachant $Y = y$ comme l’espérance par rapport à la loi conditionnelle sachant $Y = y$, *i. e.*

$$\mathbb{E}[h(X) \mid Y = y] = \int_S h(x)\nu(dx, y) =: \varphi(y)$$

pour toute fonction mesurable positive h . On observe alors que $\mathbb{E}[h(X) \mid Y] = \varphi(Y)$. Cela justifie que, pour calculer $\mathbb{E}[h(X) \mid Y]$, on peut calculer en figeant Y en y et, une fois le résultat $\varphi(y)$ trouvé, et on obtient le résultat final avec $\varphi(Y)$.

PROPOSITION 2.33. Alors $X \perp\!\!\!\perp Y$ si et seulement si, pour \mathbb{P}_Y -presque tout $y \in S$, la loi de X sachant $Y = y$ ne dépend pas de y .

Preuve \Rightarrow On suppose $X \perp\!\!\!\perp Y$. Alors pour toute fonction mesurable positive h , on a

$$\mathbb{E}[h(X) \mid Y] = \mathbb{E}[h(X)] = \int_S h(x)\mathbb{P}_X(dx)$$

ce qui montre $\nu(\star, Y) = \mathbb{P}_X$ et conclut.

\Leftarrow On suppose que, pour \mathbb{P}_Y -presque tout $y \in S$, la loi de X sachant $Y = y$ ne dépend pas de y . Soient $A \in \mathcal{A}$ et $B \in \mathcal{B}$. Alors en notant ν la loi commune de X sachant $Y = y$ pour \mathbb{P}_Y -presque tout $y \in S$, on a

$$\begin{aligned} \mathbb{P}(X \in A, Y \in B) &= \mathbb{E}[\mathbb{1}_A(X)\mathbb{1}_B(Y)] \\ &= \mathbb{E}[\mathbb{1}_B(Y)\mathbb{E}[\mathbb{1}_A(X) \mid Y]] \\ &= \mathbb{E}\left[\mathbb{1}_B(Y) \int_A \nu(dx, Y)\right] \\ &= \mathbb{E}[\mathbb{1}_B(Y)\nu(A)] = \mathbb{P}(Y \in B)\nu(A) = \mathbb{P}(Y \in B)\mathbb{P}(X \in A). \end{aligned}$$

D’où $X \perp\!\!\!\perp Y$. □

PROPOSITION 2.34 (désintégration d’une loi). Soient $A \in \mathcal{A}$ et $B \in \mathcal{B}$. Alors

$$\mathbb{P}(X \in A, Y \in B) = \int_B \mathbb{P}(X \in A \mid Y = y)\mathbb{P}_Y(dy).$$

On en déduit

$$\mathbb{P}_X = \int_T \mathbb{P}(X \in \star \mid Y = y)\mathbb{P}_Y(dy).$$

Preuve Le théorème de transfert assure

$$\begin{aligned} \mathbb{P}(X \in A, Y \in B) &= \mathbb{E}[\mathbb{1}_A(X)\mathbb{1}_B(Y)] \\ &= \mathbb{E}[\mathbb{1}_B(Y)\mathbb{E}[\mathbb{1}_A(X) \mid Y]] \\ &= \mathbb{E}[\mathbb{1}_B(Y)\nu(A, Y)] \\ &= \mathbb{E}[\mathbb{1}_B(Y)\mathbb{P}(X \in A \mid Y)] \\ &= \int_B \mathbb{P}(X \in A \mid Y = y)\mathbb{P}_Y(dy). \end{aligned} \quad \square$$

THÉORÈME 2.35 (FUBINI conditionnel). 1. Soit $f: S \times T \rightarrow \mathbf{R}_+$ une fonction $\mathcal{A} \otimes \mathcal{B}$ -mesurable. Alors la fonction

$$\varphi: y \in T \rightarrow \int_S f(x, y)\mathbb{P}_X(dx \mid Y = y)$$

est mesurable et

$$\int_{S \times T} f(x, y)\mathbb{P}_{(X, Y)}(dx, dy) = \int_T \left(\int_S f(x, y)\mathbb{P}_X(dx \mid Y = y) \right) \mathbb{P}_Y(dy). \quad (*)$$

2. Soit $f: S \times T \rightarrow \mathbf{R}$ une fonction $\mathbb{P}_{(X, Y)}$ -intégrable. Alors pour \mathbb{P}_Y -presque tout $y \in T$, la fonction $f(\cdot, y)$ est $\mathbb{P}_X(dx \mid Y = y)$ -intégrable et la fonction φ est \mathbb{P}_Y -intégrable avec la relation (*) valable.

Preuve Le théorème résulte de la proposition précédente par les arguments standards de la théorie de la mesure. □

Soit $B \in \sigma(Y)$ qu'on note $B = Y^{-1}(C)$. En appliquant la relation (*) à la fonction $(x, y) \mapsto \mathbb{1}_C(y)f(x, y)$, on obtient $\mathbb{E}[\mathbb{1}_B f(X, Y)] = \mathbb{E}[\mathbb{1}_C(Y)f(X, Y)] = \mathbb{E}[\mathbb{1}_C(Y)\varphi(Y)] = \mathbb{E}[\mathbb{1}_B\varphi(Y)]$ et on en déduit

$$\mathbb{E}[f(X, Y) | Y] = \varphi(Y) = \int_T f(x, Y)\nu(dx, Y).$$

Pour toute fonction mesurable positive h telle que $h(f(X, Y)) \in L^1(\mathcal{F})$, la formule de transfert donne alors

$$\mathbb{E}[h(f(X, Y)) | Y] = \int_S h(f(x, Y))\nu(dx, Y) = \int_S h(u)\nu(\star, Y) \circ f(\star, Y)^{-1}(du).$$

où la mesure $\nu(\star, Y) \circ f(\star, Y)^{-1}$ est la mesure image de $\nu(\star, Y)$ par $f(\star, Y)$. Par conséquent, la loi de $f(X, Y)$ par rapport à Y est $\nu(\star, Y) \circ f(\star, Y)^{-1}$. De même, pour \mathbb{P}_Y -presque tout $y \in S$, on a

$$\mathbb{E}[h(f(X, y)) | Y] = \int_S h(u)\nu(\star, Y) \circ f(\star, y)^{-1}(du)$$

et la loi de $f(X, Y)$ par rapport à Y est $\nu(\star, Y) \circ f(\star, y)^{-1}$. Compte tenu des définitions, la loi de X sachant Y et la loi de X sachant $Y = y$ sont respectivement $\nu(\star, Y)$ et $\nu(\star, y)$. D'où

PROPOSITION 2.36. Pour \mathbb{P}_Y -presque tout $y \in S$, la loi de $f(X, Y)$ sachant $Y = y$ est la même que la loi de $f(X, y)$ sachant $Y = y$.

Chapitre 3

MARTINGALES ET FILTRATIONS

3.1 Filtration et mesurabilité	14	3.4 Propriétés des martingales	17
3.2 Temps d'arrêt	14	3.5 Martingale arrêtée	18
3.3 Martingales, sous-martingales et sur-martingales	16	3.6 Décomposition de DOOB	20

3.1 FILTRATION ET MESURABILITÉ

DÉFINITION 3.1. Une *filtration* est une suite croissante $(\mathcal{F}_n)_{n \geq 0}$ de sous-tribus de \mathcal{F} . Étant donnée une filtration, le quadruplet $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, \mathbb{P})$ est appelé un espace probabilisé *filtré*. Une suite $(X_n)_{n \geq 0}$ de variables aléatoires est dite $(\mathcal{F}_n)_{n \geq 0}$ -*adaptée* si, pour tout $n \geq 0$, la variable aléatoire X_n est \mathcal{F}_n -mesurable.

▷ EXEMPLES. – Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires. On appelle *filtration canonique* de la suite $(X_n)_{n \geq 1}$ la filtration $(\mathcal{F}_n)_{n \geq 0}$ définie par $\mathcal{F}_0 := \{\emptyset, \Omega\}$ et

$$\mathcal{F}_n := \sigma(X_1, \dots, X_n) = \sigma\left(\bigcup_{i=1}^n \sigma(X_i)\right), \quad n \geq 1.$$

Il est immédiat que la suite $(X_n)_{n \geq 1}$ est adaptée par rapport à sa filtration canonique.

– On se place dans l'espace probabilisé $([0, 1], \mathcal{B}([0, 1]), \lambda)$. On appelle *filtration dyadique* la filtration $(\mathcal{F}_n)_{n \geq 0}$ définie par

$$\mathcal{F}_n := \sigma\left(\left[\frac{i-1}{2^n}, \frac{i}{2^n}\right] \mid i \in \llbracket 1, 2^n \rrbracket\right), \quad n \geq 0.$$

DÉFINITION 3.2. Une suite $(H_n)_{n \geq 1}$ de variables aléatoires est dite *prévisible* par rapport à une filtration $(\mathcal{F}_n)_{n \geq 0}$ si, pour tout $n \geq 1$, la variable aléatoire H_n est \mathcal{F}_{n-1} -mesurable.

3.2 TEMPS D'ARRÊT

Dans la suite, on fixe une filtration $(\mathcal{F}_n)_{n \geq 0}$ et on définit la notion de temps d'arrêt par rapport à cette filtration fixée.

DÉFINITION 3.3. Un $(\mathcal{F}_n)_{n \geq 0}$ -*temps d'arrêt* est une variable aléatoire T à valeurs dans $\mathbf{N} \cup \{+\infty\}$ telle que, pour tout $n \geq 0$, on a $\{T \leq n\} \in \mathcal{F}_n$.

▷ EXEMPLES. – Toute constante $n_0 \in \mathbf{N} \cup \{+\infty\}$ est un temps d'arrêt.
– Soient $(X_n)_{n \geq 0}$ une suite de variables aléatoires et sa filtration naturelle. Pour $A \in \mathcal{B}(\mathbf{R})$, la variable aléatoire

$$T_A := \min\{n \geq 1 \mid X_n \in A\}$$

est un temps d'arrêt.

◇ REMARQUES. Soit T un temps d'arrêt. Alors
– pour tout $n \geq 0$, l'ensemble $\{T = n\} = \{T \leq n\} \setminus \{T \leq n-1\}$ est un événement de \mathcal{F}_n ;
– pour tout $n \geq 1$, l'ensemble $\{T \geq n\} = \{T \leq n-1\}^c$ est un événement de \mathcal{F}_{n-1} . Ainsi, la suite $(\mathbb{1}_{\{T \geq n\}})_{n \geq 1}$ est prévisible.

PROPOSITION 3.4. 1. Soient T et S deux temps d'arrêts. Alors les variables $\max(S, T)$, $\min(S, T)$ et $S + T$ sont des temps d'arrêts.

2. Soit T un temps d'arrêt. Alors pour tout $k \geq 0$, la variable $\min(T, k)$ est un temps d'arrêt borné.

3. Soit $(T_n)_{n \geq 1}$ une suite monotone de temps d'arrêt. Alors la variable $T := \lim_{n \rightarrow +\infty} T_n$ est un temps d'arrêt.

4. Soit $(T_n)_{n \geq 1}$ une suite temps d'arrêt. Alors les variables $\inf_{n \geq 1} T_n$, $\sup_{n \geq 1} T_n$, $\liminf_{n \rightarrow +\infty} T_n$ et $\limsup_{n \rightarrow +\infty} T_n$ sont des temps d'arrêt.

DÉFINITION 3.5. À un temps d'arrêt T donné, on associe la tribu

$$\mathcal{F}_T := \{A \in \mathcal{F} \mid \forall n \geq 0, A \cap \{T \leq n\} \in \mathcal{F}_n\}.$$

Preuve Justifions qu'il s'agit d'une tribu. Il est clair que $\Omega \in \mathcal{F}_T$ car, pour $n \geq 0$, on a $\Omega \cap \{T \leq n\} = \{T \leq n\} \in \mathcal{F}_n$. Soit $(A_i)_{i \geq 1}$ une suite de \mathcal{F}_T . Pour tout $n \geq 0$, on a

$$\left(\bigcup_{i \geq 1} A_i \right) \cap \{T \leq n\} = \bigcup_{i \geq 1} (A_i \cap \{T \leq n\}) \in \mathcal{F}_n$$

ce qui montre $\bigcup_{i \geq 1} A_i \in \mathcal{F}_T$. Soit $A \in \mathcal{F}_T$. Pour tout $n \geq 0$, on a

$$A^c \cap \{T \leq n\} = \{T \leq n\} \setminus (A \cap \{T \leq n\}) \in \mathcal{F}_n$$

ce qui montre $A^c \in \mathcal{F}_T$. Il s'agit donc d'une tribu. \square

Cette tribu s'interprète comme l'information disponible à la date T .

PROPOSITION 3.6. Soient T et S deux temps d'arrêt.

1. Si T est constante égale à $n_0 \in \mathbf{N}$, alors $\mathcal{F}_T = \mathcal{F}_{n_0}$.
2. Si $T \leq S$, alors $\mathcal{F}_T \subset \mathcal{F}_S$.
3. Le temps d'arrêt T est \mathcal{F}_T -mesurable.
4. On a $\mathcal{F}_{\min(T,S)} = \mathcal{F}_T \cap \mathcal{F}_S$. De plus, les ensembles $\{T \leq S\}$ et $\{T = S\}$ sont des événements de $\mathcal{F}_{\min(T,S)}$.
5. Soit $A \in \mathcal{F}$. On pose

$$T_A := \begin{cases} T & \text{sur } A, \\ +\infty & \text{sinon.} \end{cases}$$

Alors $A \in \mathcal{F}_T$ si et seulement si T_A est un temps d'arrêt.

Preuve 1. Comme la suite $(\mathcal{F}_n)_{n \geq 0}$ est croissante, on a les équivalences suivantes :

$$\begin{aligned} A \in \mathcal{F}_T &\iff \forall n \geq 0, \mathcal{F}_n \ni A \cap \{T \leq n\} = \begin{cases} A \cap \emptyset = \emptyset & \text{si } n < n_0, \\ A \cap \Omega = A & \text{sinon} \end{cases} \\ &\iff \forall n \geq n_0, A \in \mathcal{F}_n \\ &\iff A \in \mathcal{F}_{n_0}. \end{aligned}$$

2. On suppose $T \leq S$. Soit $A \in \mathcal{F}_T$. Alors pour tout $n \geq 0$, on a $A \cap \{T \leq n\} \in \mathcal{F}_n$, donc

$$A \cap \{S \leq n\} = A \cap \{T \leq n\} \cap \{S \leq n\} \in \mathcal{F}_n.$$

D'où $A \in \mathcal{F}_S$.

3. Soit $p \geq 0$. Pour tout $n \geq 0$, on a

$$\{T \leq p\} \cap \{T \leq n\} = \{T \leq \min(p, n)\} \in \mathcal{F}_{\min(p,n)} \subset \mathcal{F}_n.$$

Cela montre que $\{T \leq p\} \in \mathcal{F}_T$, i. e. la variable T est \mathcal{F}_T -mesurable.

4. Comme $\min(T, S) \leq T$ et $\min(T, S) \leq S$, on a $\mathcal{F}_{\min(T,S)} \subset \mathcal{F}_T$ et $\mathcal{F}_{\min(T,S)} \subset \mathcal{F}_S$, donc $\mathcal{F}_{\min(T,S)} \subset \mathcal{F}_T \cap \mathcal{F}_S$. Réciproquement, soit $A \in \mathcal{F}_T \cap \mathcal{F}_S$. Alors pour tout $n \geq 0$, on a

$$\begin{aligned} A \cap \{\min(T, S) \leq n\} &= A \cap (\{T \leq n\} \cup \{S \leq n\}) \\ &= (A \cap \{T \leq n\}) \cup (A \cap \{S \leq n\}) \in \mathcal{F}_n. \end{aligned}$$

Cela montre $A \in \mathcal{F}_{\min(T,S)}$. Finalement, on a $\mathcal{F}_T \cap \mathcal{F}_S \subset \mathcal{F}_{\min(T,S)}$ ce qui conclut.

Montrons que $\{T \leq S\} \in \mathcal{F}_T$. Pour cela, il suffit de montrer que

$$\forall n \geq 0, \{T \leq S\} \cap \{T \leq n\} \in \mathcal{F}_n.$$

Soit $n \geq 0$. Alors

$$\begin{aligned} \{T \leq S\} \cap \{T \leq n\} &= \{\min(T, n) \leq S\} \cap \{T \leq n\} \\ &= (\{\min(T, n) \leq S\} \cap \{S \leq n\} \cap \{T \leq n\}) \cup (\{\min(T, n) \leq S\} \cap \{S > n\} \cap \{T \leq n\}) \\ &= (\{\min(T, n) \leq \min(S, n)\} \cap \{S \leq n\} \cap \{T \leq n\}) \cup (\{S \leq n\} \cap \{T \leq n\}) \in \mathcal{F}_n \end{aligned}$$

car, comme $\min(T, n)$ est $\mathcal{F}_{\min(T,n)}$ -mesurable, elle est aussi \mathcal{F}_n -mesurable et de même pour $\min(S, n)$. Cela conclut. De même, montrons que $\{T \leq S\} \in \mathcal{F}_S$. On a

$$\{T \leq S\} \cap \{S \leq n\} = \{\min(T, n) \leq \min(S, n)\} \cap \{S \leq n\} \in \mathcal{F}_n.$$

Cela conclut également.

5. Le point 5 est clair. \square

PROPOSITION 3.7. Soient $(X_n)_{n \geq 0}$ une suite de variables aléatoires adaptée et T un temps d'arrêt. Alors la variable $\mathbb{1}_{\{T < +\infty\}} X_T$ est \mathcal{F}_T -mesurable.

Preuve Soit $B \in \mathcal{B}(\mathbf{R})$. Montrons que $\{\mathbb{1}_{\{T < +\infty\}} X_T \in B\} \in \mathcal{F}_T$. Pour tout $n \geq 0$, on a

$$\{\mathbb{1}_{\{T < +\infty\}} X_T \in B\} \cap \{T = n\} = \{X_n \in B\} \cap \{T = n\} \in \mathcal{F}_n$$

ce qui termine la preuve. \square

3.3 MARTINGALES, SOUS-MARTINGALES ET SUR-MARTINGALES

DÉFINITION 3.8. Une suite $(X_n)_{n \geq 0}$ des variables aléatoires est une $(\mathcal{F}_n)_{n \geq 0}$ -martingale lorsque qu'on a les trois conditions suivantes :

- (i) pour tout $n \geq 1$, on a $\mathbb{E}[|X_n|] < +\infty$;
- (ii) la suite $(X_n)_{n \geq 0}$ est $(\mathcal{F}_n)_{n \geq 0}$ -adaptée;
- (iii) pour tout $n \geq 0$, on a $\mathbb{E}[X_{n+1} | \mathcal{F}_n] = X_n$.

On définit une sous-martingale (respectivement une sur-martingale) lorsque le point (iii) est remplacé par

- (iii') pour tout $n \geq 0$, on a $\mathbb{E}[X_{n+1} | \mathcal{F}_n] \geq X_n$ (respectivement $\mathbb{E}[X_{n+1} | \mathcal{F}_n] \leq X_n$).

▷ EXEMPLES. – • *Martingale fermée.* Soient X une variable aléatoire intégrable et $(\mathcal{F}_n)_{n \geq 0}$ une filtration. Alors la suite $(X_n)_{n \geq 0}$ définie par $X_n := \mathbb{E}[X | \mathcal{F}_n]$ pour tout $n \geq 0$ est une martingale. Le point (iii) est assuré par la propriété du conditionnement en cascade.

– • *Marche aléatoire (ou de DOOB).* Soit $(X_n)_{n \geq 1}$ une suite indépendante de variables aléatoires intégrables centrées. Pour $n \geq 1$, on pose $S_n := X_1 + \dots + X_n$. Alors la suite $(S_n)_{n \geq 1}$ est une martingale par rapport à la filtration canonique^{§1}. Les points (i) et (ii) sont assez clairs. Justifions le point (iii). Pour tout $n \geq 0$, on a

$$\begin{aligned} \mathbb{E}[S_{n+1} | \mathcal{F}_n] &= \mathbb{E}[S_n + X_{n+1} | \mathcal{F}_n] \\ &= \mathbb{E}[S_n | \mathcal{F}_n] + \mathbb{E}[X_{n+1} | \mathcal{F}_n] = S_n + \mathbb{E}[X_{n+1}] = S_n \end{aligned}$$

puisque $X_{n+1} \perp \mathcal{F}_n = \sigma(X_1, \dots, X_n)$.

– • *Modèle auto-régressif.* Soit $(\varepsilon_n)_{n \geq 1}$ une suite indépendante de variables aléatoires identiques distribuées, intégrable et centrée. Soient $x, a \in \mathbf{R}$. La suite $(Y_n)_{n \geq 0} := (X_n / a^n)_{n \geq 0}$ définie par

$$X_0 := x \quad \text{et} \quad X_{n+1} = aX_n + \varepsilon_{n+1}, \quad n \geq 0$$

est une martingale par rapport à la tribu engendrée par la suite $(\varepsilon_n)_{n \geq 1}$. En effet, les points (i) et (ii) se montrent par récurrence puisque. Pour le point (iii), pour tout $n \geq 0$, on a

$$\begin{aligned} \mathbb{E}[Y_{n+1} | \mathcal{F}_n] &= \mathbb{E}\left[\frac{X_{n+1}}{a^{n+1}} \mid \mathcal{F}_n\right] \\ &= \frac{1}{a^{n+1}} \mathbb{E}[aX_n + \varepsilon_{n+1} \mid \mathcal{F}_n] \\ &= \frac{1}{a^{n+1}} (aX_n + \mathbb{E}[\varepsilon_{n+1} | \mathcal{F}_n]) = \frac{X_n}{a^n} = Y_n \end{aligned}$$

puisque $\mathbb{E}[\varepsilon_{n+1} | \mathcal{F}_n] = \mathbb{E}[\varepsilon_{n+1}] = 0$.

– • *Modèle de GALTON-WATSON.* Soit $(X_{i,j})_{i,j \geq 1}$ une famille indépendante de variables aléatoires L^1 identiquement distribuées dont on note μ la loi, appelée loi de reproduction. Soit $(Z_n)_{n \geq 0}$ la suite définie par

$$Z_0 := 1 \quad \text{et} \quad Z_{n+1} = \sum_{j=1}^{Z_n} X_{n+1,j}, \quad n \geq 0.$$

On note $m := \mathbb{E}[X_{1,1}]$. Alors la suite $(Z_n / m^n)_{n \geq 1}$ est une martingale par rapport à la filtration $(\mathcal{F}_n)_{n \geq 1}$ définie par

$$\mathcal{F}_n := \sigma(X_{i,j} \mid i \leq n, j \geq 1), \quad n \geq 1.$$

Vérifions le point (ii). Soit $n \geq 0$. Raisonnons par récurrence. Si Z_n est \mathcal{F}_n -mesurable, alors

$$\begin{aligned} \forall A \in \mathcal{B}(\mathbf{R}), \quad \{Z_{n+1} \in A\} &= \bigcup_{p \in \mathbf{N}} \{Z_{n+1} \in A, Z_n = p\} \\ &= \bigcup_{p \in \mathbf{N}} \left\{ \sum_{i=0}^p X_{n+1,j} \in A \right\} \cap \{Z_n = p\} \in \mathcal{F}_{n+1}. \end{aligned}$$

§1. Remarquons qu'on ne précise par quelle filtration canonique puisque $\sigma(S_1, \dots, S_n) = \sigma(X_1, \dots, X_n)$ pour tout $n \geq 1$.

Enfin, vérifions le point (iii). Pour tout $n \geq 0$, on a

$$\begin{aligned} \mathbb{E}\left[\frac{Z_{n+1}}{m^{n+1}} \mid \mathcal{F}_n\right] &= \frac{1}{m^{n+1}} \mathbb{E}\left[\sum_{j=1}^{Z_n} X_{n+1,j} \mid \mathcal{F}_n\right] \\ &= \frac{1}{m^{n+1}} \mathbb{E}\left[\sum_{k=0}^{+\infty} \sum_{j=1}^{Z_n} X_{n+1,j} \mathbb{1}_{\{Z_n=k\}} \mid \mathcal{F}_n\right] \\ &= \frac{1}{m^{n+1}} \sum_{k=0}^{+\infty} \sum_{j=1}^k \mathbb{E}[X_{n+1,j} \mid \mathcal{F}_n] \mathbb{1}_{\{Z_n=k\}} \\ &= \frac{1}{m^{n+1}} \sum_{k=0}^{+\infty} \sum_{j=1}^{Z_n} \mathbb{E}[X_{n+1,j} \mid \mathcal{F}_n] \mathbb{1}_{\{Z_n=k\}} \\ &= \frac{1}{m^{n+1}} \sum_{j=1}^{Z_n} \mathbb{E}[X_{n+1,j} \mid \mathcal{F}_n] \\ &= \frac{1}{m^{n+1}} \sum_{j=1}^{Z_n} \mathbb{E}[X_{n+1,j}] = \frac{mZ_n}{m^{n+1}} = \frac{Z_n}{m^n}. \end{aligned}$$

Le calcul est bien licite car toutes les variables aléatoires considérées sont positives. Une récurrence immédiate montre alors que les variables Z_n/m^n sont intégrables.

– • *Cascade aléatoire.* Soit $(X_n)_{n \geq 0}$ une suite indépendante de variable aléatoires d'espérance 1. La suite $(Y_n)_{n \geq 1}$ définie par

$$Y_0 := 1 \quad \text{et} \quad Y_n = X_1 \cdots X_n, \quad n \geq 0$$

est une martingale par rapport à la filtration canonique de la suite $(X_n)_{n \geq 0}$.

3.4 PROPRIÉTÉS DES MARTINGALES

PROPOSITION 3.9. Soit $(X_n)_{n \geq 0}$ une $(\mathcal{F}_n)_{n \geq 0}$ -martingale. Alors c'est aussi une $(\mathcal{G}_n)_{n \geq 0}$ -martingale où la filtration $(\mathcal{G}_n)_{n \geq 0}$ est la filtration canonique.

Preuve Les points (i) et (ii) sont automatiques. Montrons le point (i). Pour tout $n \geq 0$, comme la tribu \mathcal{G}_n est la plus petite tribu rendant les variables X_1, \dots, X_n mesurables, on a $\mathcal{G}_n \subset \mathcal{F}_n$ et donc

$$\mathbb{E}[X_{n+1} \mid \mathcal{G}_n] = \mathbb{E}[\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] \mid \mathcal{G}_n] = \mathbb{E}[X_n \mid \mathcal{G}_n] = X_n. \quad \square$$

PROPRIÉTÉ 3.10. Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires. Alors le point (iii) de la définition précédente est équivalent à l'assertion

$$\forall n \leq m, \quad \mathbb{E}[X_m \mid \mathcal{F}_n] = X_n. \quad (*)$$

Idem pour des sur- et sous-martingales.

Preuve Pour l'implication (iii) \implies (*), on effectue une récurrence sur l'entier m . L'implication, quant à elle, est trivial. \square

PROPOSITION 3.11. L'ensemble des $(\mathcal{F}_n)_{n \geq 0}$ -martingale est un espace vectoriel. Les ensembles des sous-martingales et des sur-martingales sont stable par combinaison linéaire à coefficients positifs.

PROPOSITION 3.12. Soit $(X_n)_{n \geq 0}$ une martingale (respectivement une sous-martingale et une sur-martingale). Alors la suite des espérances $(\mathbb{E}[X_n])_{n \geq 0}$ est constante (respectivement croissante et décroissante).

PROPOSITION 3.13. 1. Soient $(X_n)_{n \geq 0}$ une martingale et $\varphi: \mathbf{R} \rightarrow \mathbf{R}$ une fonction convexe telle que $\varphi(X_n) \in L^1(\mathcal{F})$ pour tout $n \geq 0$. Alors la suite $(\varphi(X_n))_{n \geq 0}$ est une sous-martingale.

2. Soient $(X_n)_{n \geq 0}$ une sous-martingale et $\varphi: \mathbf{R} \rightarrow \mathbf{R}$ une fonction convexe croissante telle que $\varphi(X_n) \in L^1(\mathcal{F})$ pour tout $n \geq 0$. Alors la suite $(\varphi(X_n))_{n \geq 0}$ est une sous-martingale.

Preuve Dans les deux cas, la suite $(\varphi(X_n))_{n \geq 0}$ est une suite adaptée de variables aléatoires intégrables. Plaçons-nous dans le cas 1. Pour tout $n \geq 0$, l'inégalité de JENSEN conditionnelle assure

$$\mathbb{E}[\varphi(X_{n+1}) \mid \mathcal{F}_n] \geq \varphi(\mathbb{E}[X_{n+1} \mid \mathcal{F}_n]) = \varphi(X_n).$$

Cela montre que la suite $(\varphi(X_n))_{n \geq 0}$ est une sous-martingale. \square

CONSÉQUENCES.

- Soit $(X_n)_{n \geq 0}$ une martingale telle que $X_n \in L^p(\mathcal{F})$ pour tout $n \geq 0$. Alors la suite $(|X_n|^p)_{n \geq 0}$ est une sous-martingale. En effet, la fonction $x \mapsto |x|^p$ est convexe.
- Soient $(X_n)_{n \geq 0}$ une sous-martingale et $a \in \mathbf{R}$. Alors la suite $((X_n - a)^+)_{n \geq 0}$ est une sous-martingale.
- Soient $(X_n)_{n \geq 0}$ une sur-martingale et $a \in \mathbf{R}$. Alors la suite $(\min(X_n, a))_{n \geq 0}$ est une sous-martingale. En effet, on applique la propriété précédente à la sous-martingale $(-X_n)_{n \geq 0}$ avec la fonction convexe $x \mapsto \max(x, -a)$.

PROPOSITION 3.14. Soit $X := (X_n)_{n \geq 0}$ une sous-martingale et $H := (H_n)_{n \geq 0}$ une suite prévisible des variables aléatoires positives et bornée. Alors la suite $H \cdot X$ définie par

$$(H \cdot X)_0 := 0 \quad \text{et} \quad (H \cdot X)_n := \sum_{m=1}^n H_m (X_m - X_{m-1})$$

est une sous-martingale. La même affirmation est vraie pour des sur-martingales et encore vraie pour des martingales sans condition de positivité sur H .

Preuve Pour tout $n \geq 0$, comme H_m est bornée et $X_m - X_{m-1} \in L^1(\mathcal{F})$ pour tout $m \in \llbracket 1, n \rrbracket$, on a $(H \cdot X)_n \in L^1(\mathcal{F})$. Le point (ii) se montre facilement par récurrence. Montrons le point (iii). Pour tout $n \geq 0$, on a

$$\begin{aligned} \mathbb{E}[(H \cdot X)_n \mid \mathcal{F}_{n-1}] &= \mathbb{E}[(H \cdot X)_{n-1} + H_n(X_n - X_{n-1}) \mid \mathcal{F}_{n-1}] \\ &= (H \cdot X)_{n-1} + H_n \mathbb{E}[X_n - X_{n-1} \mid \mathcal{F}_{n-1}] \\ &= (H \cdot X)_{n-1} + H_n(\mathbb{E}[X_n \mid \mathcal{F}_{n-1}] - X_{n-1}) \geq (H \cdot X)_{n-1} \end{aligned}$$

car, successivement, la variable H_n est \mathcal{F}_{n-1} -mesurable et la suite $(X_k)_{k \geq 0}$ est une sous-martingale. Ainsi la suite $H \cdot X$ est un sous-martingale. On procède de même dans les autres cas. \square

3.5 MARTINGALE ARRÊTÉE

NOTATION. Étant donné un temps d'arrêt T et une suite $X := (X_n)_{n \geq 0}$, on appelle *suite arrêtée* de $(X_n)_{n \geq 0}$ en T la suite $X^T := (X_n^T)_{n \geq 0}$ où, pour tout $n \geq 0$, on pose $X_n^T := X_{\min(T, n)}$.

PROPOSITION 3.15. Soit T un temps d'arrêt.

1. Soit $X := (X_n)_{n \geq 0}$ une suite adaptée de variables aléatoires. Alors la suite X^T est aussi adaptée.
2. Soit $H := (H_n)_{n \geq 1}$ une suite prévisible de variables aléatoires. Alors la suite H^T est aussi prévisible.

Preuve 1. Soit $n \geq 0$. Il suffit de montrer que la variable X_n^T est \mathcal{F}_n -mesurable. Soit $B \in \mathcal{B}(\mathbf{R})$. On a

$$\begin{aligned} \{X_n^T \in B\} &= \{X_{\min(T, n)} \in B\} = \bigcup_{p \geq 0} \{X_{\min(T, n)} \in B, T = p\} \\ &= \bigcup_{p \geq 0} \{X_{\min(p, n)} \in B, T = p\} \\ &= \bigcup_{p=0}^n \{X_p \in B, T = p\} \cup \bigcup_{p > n} \{X_n \in B, T = p\} \\ &= \bigcup_{p=0}^n \underbrace{\{X_p \in B, T = p\}}_{\in \mathcal{F}_p \subset \mathcal{F}_n} \cup \bigcup_{p > n} \underbrace{\{X_n \in B, T = p\}}_{\in \mathcal{F}_n} \in \mathcal{F}_n. \end{aligned}$$

2. De la même façon, soient $n \geq 1$ et $B \in \mathcal{B}(\mathbf{R})$. On a

$$\begin{aligned} \{H_n^T \in B\} &= \bigcup_{p \geq 0} \{H_{\min(p, n)} \in B, T = p\} \\ &= \bigcup_{p=0}^{n-1} \underbrace{\{H_p \in B, T = p\}}_{\in \mathcal{F}_p \subset \mathcal{F}_{n-1}} \cup \bigcup_{p \geq n} \underbrace{\{H_n \in B, T = p\}}_{\in \mathcal{F}_{n-1}} \in \mathcal{F}_{n-1}. \end{aligned} \quad \square$$

DÉFINITION 3.16. Soient X une martingale et T un temps d'arrêt. On appelle *martingale arrêtée* la suite X^T . Idem pour les sur- et sous-martingales.

PROPOSITION 3.17. Une martingale arrêtée est une martingale. Idem pour les sur- et sous-martingales.

Preuve Soient $X := (X_n)_{n \geq 0}$ une martingale et T un temps d'arrêt. Pour $n \geq 1$, on pose $H_n = \mathbb{1}_{\{T \geq n\}}$. Alors la suite $H := (H_n)_{n \geq 1}$ est prévisible et bornée. D'après la proposition 3.14, la suite $H \cdot X$ est une martingale. Pour tout $n \geq 1$, on a

$$(H \cdot X)_n = \sum_{m=1}^n \mathbb{1}_{\{T \geq m\}} (X_m - X_{m-1}) = \sum_{m=1}^{\min(n, T)} (X_m - X_{m-1}) = X_{\min(n, T)} - X_0.$$

Cela montre que la suite $X^T = H \cdot X + X_0$ est de même nature que la suite $H \cdot X$ et donc que la suite X □

THÉORÈME 3.18. Soient $X := (X_n)_{n \geq 0}$ une sous-martingale, $k \geq 0$ et T un temps d'arrêt tel que $T \leq k$. Alors

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_T] \leq \mathbb{E}[X_k].$$

Pour une martingale, il y a égalité et, pour une sur-martingale, les inégalités sont renversées.

Preuve Il suffit de montrer le cas sous-martingale, les autres s'en déduisant facilement. On a vu que la suite X^T est une sous-martingale, donc la suite de ses espérances est croissante. En particulier, on a $\mathbb{E}[X_0^T] \leq \mathbb{E}[X_k^T]$. Or on remarque que $X_0 = X_{\min(T, 0)}$ et $X_T = X_{\min(T, k)}$. On en déduit l'inégalité

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_T].$$

Montrons la second inégalité. Pour $n \geq 1$, on pose $K_n := \mathbb{1}_{\{T \leq n-1\}}$. Alors la suite $(K_n)_{n \geq 1}$ est prévisible, donc la suite $K \cdot X$ est une sous-martingale. Or pour tout $n \geq 0$, on a

$$\begin{aligned} (K \cdot X)_n &= \sum_{m=1}^n \mathbb{1}_{\{T \leq m-1\}} (X_m - X_{m-1}) \\ &= \sum_{m=T+1}^{\max(n, T)} (X_m - X_{m-1}) = X_{\max(n, T)} - X_T = X_n - X_{\min(T, n)}. \end{aligned}$$

Mais la suite des espérances de $K \cdot X$ est croissante, donc

$$\mathbb{E}[X_k - X_{\min(T, k)}] = \mathbb{E}[(K \cdot X)_k] \geq \mathbb{E}[(K \cdot X)_0] = 0.$$

On en déduit

$$\mathbb{E}[X_k] \geq \mathbb{E}[X_{\min(T, k)}] = \mathbb{E}[X_T]. \quad \square$$

CONTRE-EXEMPLE. Le théorème est faux si le temps d'arrêt T n'est pas déterministiquement borné. En effet, soit $(X_i)_{i \geq 1}$ une suite indépendante de variables aléatoires suivant une loi de RADEMACHER. Pour tout $n \geq 1$, on pose $S_0 := 0$ et $S_n := X_1 + \dots + X_n$. Alors la suite $(S_n)_{n \geq 0}$ est une martingale par rapport à la filtration canonique. On considère la variable

$$T := \inf\{n \geq 0 \mid S_n = -1\}$$

qui est un temps d'arrêt. Par définition, on a $S_T = -1$, donc $\mathbb{E}[S_T] = -1$ et, pourtant, on a $\mathbb{E}[S_0] = 0$.

THÉORÈME 3.19. Soient $X := (X_n)_{n \geq 0}$ une sous-martingale et T un temps d'arrêt. Sous chacune des conditions qui suivent, on a

$$X_T \in L^1(\mathcal{F}) \quad \text{et} \quad \mathbb{E}[X_0] \leq \mathbb{E}[X_T].$$

- (i) le temps d'arrêt T est déterministiquement borné, *i. e.* il existe $c \in \mathbf{R}_+$ tel que $T \leq c$ presque sûrement;
- (ii) la suite X est bornée presque sûrement et le temps d'arrêt T est presque sûrement fini;
- (iii) on a $\mathbb{E}[T] < +\infty$ et il existe $k \in \mathbf{R}_+$ tel que $|X_{n+1} - X_n| \leq k$ presque sûrement pour tout $n \geq 0$.

De plus, si X est une martingale, alors on a égalité sous les conditions (i), (ii) ou (iii). Si X est une sur-martingale, l'inégalité est renversée sous la condition (i), (ii) ou (iii) ou sous la condition

- (iv) le temps d'arrêt T est presque sûrement fini et les variables X_n sont positives.

Preuve (i) Cela découle du théorème précédent et on a

$$X_T = \sum_{k=1}^{\lfloor c \rfloor} X_k \mathbb{1}_{\{T \geq k\}}, \quad \text{donc} \quad |X_T| \leq \sum_{k=1}^{\lfloor c \rfloor} |X_k| \in L^1(\mathcal{F}).$$

(ii) Soit $n \geq 0$. On considère le temps d'arrêt $\min(T, n)$ la temps d'arrêt déterministiquement borné. D'après le cas (i), on a $\mathbb{E}[X_{\min(T, n)}] \geq \mathbb{E}[X_0]$. Comme $T < +\infty$ presque sûrement, la suite $(X_{\min(T, n)})_{n \geq 0}$ converge presque sûrement vers X_T . On veut ensuite passer à la limite. Pour tout $n \geq 0$, on observe que

$$X_{\min(T, n)} = \sum_{i=1}^{n-1} X_i \mathbb{1}_{\{T \geq i\}} + X_n \mathbb{1}_{\{T \geq n\}}, \quad \text{donc} \quad |X_{\min(T, n)}| = \sum_{i=1}^{n-1} |X_i| \mathbb{1}_{\{T \geq i\}} + |X_n| \mathbb{1}_{\{T \geq n\}} \leq k. \quad (*)$$

Le théorème de convergence dominée assure alors

$$\lim_{n \rightarrow +\infty} \mathbb{E}[X_{\min(T,n)}] = \mathbb{E}[\lim_{n \rightarrow +\infty} X_{\min(T,n)}] = \mathbb{E}[X_T].$$

On en déduit $\mathbb{E}[X_0] \leq \mathbb{E}[X_T]$. De plus, par la relation (*), on a donc $|X_T| \leq k$ ce qui montre $X_T \in L^1(\mathcal{F})$.

(iii) De même, comme $\mathbb{E}[|T|] < +\infty$, le temps d'arrêt T est presque sûrement fini, donc $X_{\min(T,n)} \xrightarrow{-ps} X_T$. D'après le cas (i), on a

$$\forall n \geq 0, \quad \mathbb{E}[X_{\min(T,n)}] \geq \mathbb{E}[X_0].$$

De plus, l'inégalité triangulaire assure $|X_T - X_0| \leq k \min(T,n) \leq kT \in L^1(\mathcal{F})$. Le théorème de convergence dominée assure alors

$$\mathbb{E}[X_T - X_0] = \lim_{n \rightarrow +\infty} \mathbb{E}[X_{\min(T,n)} - X_0] \geq 0$$

ce qui donne l'inégalité. De plus, comme $|X_T - X_0| \leq kT$, on a $X_T \in L^1(\mathcal{F})$.

• *Pour des martingales.* Lorsque X est une martingale, comme pour une sous-martingale, l'énoncé précédent s'applique et donne $\mathbb{E}[X_0] = \mathbb{E}[X_{\min(T,n)}]$ pour tout temps d'arrêt T . On montre alors, dans le cas (i), (ii) ou (ii), les égalités sont conservées par passage à la limite.

• *Pour des sur-martingales.* Lorsque X est une sur-martingale, alors $-X$ est une sous-martingale qui vérifie le même point (i), (ii) ou (iii) que lorsque X le vérifie. Il vient alors $\mathbb{E}[-X_0] \leq \mathbb{E}[-X_T]$, c'est-à-dire $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$.

Sous l'hypothèse (iv), on a $\mathbb{E}[X_{\min(T,n)}] \leq \mathbb{E}[X_0]$. Or $T < +\infty$ presque sûrement, donc $X_{\min(T,n)} \xrightarrow{-ps} X_T$ et le lemme de FATOU donne

$$\mathbb{E}[X_T] = \mathbb{E}[\liminf_{n \rightarrow +\infty} X_{\min(T,n)}] \leq \liminf_{n \rightarrow +\infty} \mathbb{E}[X_{\min(T,n)}] \leq \mathbb{E}[X_0]. \quad \square$$

3.6 DÉCOMPOSITION DE DOOB

THÉORÈME 3.20 (décomposition de DOOB). Toute sous-martingale $(X_n)_{n \geq 0}$ se décompose sous la forme (presque sûrement unique)

$$X_n = M_n + A_n, \quad \forall n \geq 0$$

où la suite $(M_n)_{n \geq 0}$ est une martingale vérifiant $M_0 = X_0$ et la suite $(A_n)_{n \geq 0}$ est prévisible et croissante vérifiant

$$A_n = \sum_{k=1}^n (\mathbb{E}[X_k | \mathcal{F}_{k-1}] - X_{k-1}), \quad \forall n \geq 0.$$

Preuve • *Existence.* On suppose avoir une telle décomposition. Alors pour tout $n \geq 1$, on a

$$\mathbb{E}[X_n | \mathcal{F}_{n-1}] = \mathbb{E}[M_n | \mathcal{F}_{n-1}] + A_n = M_{n-1} + A_n = X_{n-1} - A_{n-1} + A_n,$$

donc

$$A_n - A_{n-1} = \mathbb{E}[X_n | \mathcal{F}_{n-1}] - X_{n-1}. \quad (*)$$

On prend donc la suite $(A_n)_{n \geq 0}$ vérifiant (*) avec $A_0 = 0$. Pour $n \geq 0$, on pose $M_n := X_n - A_n$. Il reste ainsi à voir que la suite $(M_n)_{n \geq 0}$ est une martingale et que la suite $(A_n)_{n \geq 0}$ est prévisible. Comme $(X_n)_{n \geq 0}$ est une sous-martingale, on a $A_n - A_{n-1} \geq 0$ pour tout $n \geq 0$ ce qui montre la croissance de la suite $(A_n)_{n \geq 0}$. De plus, elle est prévisible. En effet, on peut montrer cela par récurrence : pour tout $n \geq 1$, si A_n est \mathcal{F}_{n-1} -mesurable, alors

$$A_{n+1} - A_n = \mathbb{E}[X_{n+1} | \mathcal{F}_n] - X_n + A_n$$

est \mathcal{F}_n -mesurable. Enfin, montrons que la suite $(M_n)_{n \geq 0}$ est une martingale. En effet, d'abord, pour tout $n \geq 0$, la variable $M_n = X_n - A_n$ est \mathcal{F}_n -mesurable. De plus, pour tout $n \geq 0$, comme

$$\forall k \in \llbracket 1, n \rrbracket, \quad |A_n - A_{n-1}| \leq |\mathbb{E}[X_n | \mathcal{F}_{n-1}]| + |X_{n-1}| \in L^1(\mathcal{F}),$$

donc $A_n = \sum_{k=1}^n (A_n - A_{n-1}) \in L^1(\mathcal{F})$, donc $M_n = X_n - A_n \in L^1(\mathcal{F})$. Enfin, pour tout $n \geq 1$, on a

$$\begin{aligned} \mathbb{E}[M_n | \mathcal{F}_{n-1}] &= \mathbb{E}[X_n - A_n | \mathcal{F}_{n-1}] \\ &= \mathbb{E}[X_n | \mathcal{F}_{n-1}] - A_n \\ &= (A_n - A_{n-1} + X_{n-1}) - A_n \\ &= X_{n-1} - A_{n-1} = M_{n-1}. \end{aligned}$$

Donc la suite $(M_n)_{n \geq 0}$ est une martingale.

• *Unicité.* L'unicité vient de la relation (*) dès lors que l'on a posé $A_0 = 0$. □

DÉFINITION 3.21. Soit $X := (X_n)_{n \geq 0}$ une martingale telle que $X_n \in L^2(\mathcal{F})$ pour tout $n \geq 0$. Alors la suite $(X_n^2)_{n \geq 0}$ est une sous-martingale. Alors il existe une martingale $(M_n)_{n \geq 0}$ et une suite croissante prévisible $(A_n)_{n \geq 0}$, presque sûrement uniques, telles que

$$X_n^2 = M_n + A_n, \quad n \geq 0.$$

La suite $\langle X, X \rangle := (A_n)_{n \geq 0}$ est le *compensateur* de la martingale X . Pour tout $n \geq 0$, on a

$$\begin{aligned} \langle X, X \rangle_n &= \sum_{k=1}^n (\mathbb{E}[X_k^2 | \mathcal{F}_{k-1}] - X_{k-1}^2) \\ &= \sum_{k=1}^n \mathbb{E}[(X_k - X_{k-1})^2 | \mathcal{F}_{k-1}]. \end{aligned} \quad (*)$$

PROPOSITION 3.22. Soit $X := (X_n)_{n \geq 0}$ une martingale telle que $X_0 = 0$ et $X_n \in L^2(\mathcal{F})$ pour tout $n \geq 0$. Alors la suite X est bornée dans $L^2(\mathcal{F})$ si et seulement si

$$\mathbb{E}[\langle X, X \rangle_\infty] := \mathbb{E}[\lim_{n \rightarrow +\infty} \langle X, X \rangle_n] < +\infty.$$

Preuve On sait qu'il existe une martingale $(M_n)_{n \geq 0}$ telle que $X_n = M_n + \langle X, X \rangle_n$ pour tout $n \geq 0$. Pour tout $n \geq 0$, on a $\mathbb{E}[M_n] = \mathbb{E}[M_0] = 0$. En passant à l'espérance et à la borne supérieure, on a

$$\sup_{n \geq 0} \mathbb{E}[X_n^2] = \sup_{n \geq 0} \mathbb{E}[\langle X, X \rangle_n].$$

Or le théorème de convergence monotone donne

$$\sup_{n \geq 0} \mathbb{E}[\langle X, X \rangle_n] = \lim_{n \rightarrow +\infty} \mathbb{E}[\langle X, X \rangle_n] = \mathbb{E}[\lim_{n \rightarrow +\infty} \langle X, X \rangle_n] = \mathbb{E}[\langle X, X \rangle_\infty].$$

Ceci donne alors l'équivalence. □

▷ EXEMPLE. Soient $\sigma \geq 0$ et $(X_n)_{n \geq 0}$ une suite indépendante de variables aléatoires identiquement distribuées appartenant à $L^2(\mathcal{F})$ et de variance σ^2 . Alors la suite $(S_n)_{n \geq 0}$ des sommes partielles est une martingale par rapport à la filtration canonique. En se servant de l'égalité (*), on obtient

$$\langle S, S \rangle_n = n\sigma^2, \quad n \geq 0.$$

PROPOSITION 3.23. Soit $X := (X_n)_{n \geq 0}$ une martingale telle que $X_n \in L^2(\mathcal{F})$ pour tout $n \geq 0$. Soit T un temps d'arrêt. Alors $\langle X, X \rangle^T = \langle X^T, X^T \rangle$ presque sûrement.

Preuve On écrit $X^2 = M + \langle X, X \rangle$ la décomposition de DOOB de X^2 . On peut alors écrire $(X^2)^T = M^T + \langle X, X \rangle^T$ où la suite M^T est une martingale et la suite $\langle X, X \rangle^T$ est une suite croissante prévisible. Comme $(X^2)^T = (X^T)^2$ est une sous-martingale, l'unicité de la décomposition de DOOB entraîne $\langle X, X \rangle^T = \langle X^T, X^T \rangle$ presque sûrement. □

Chapitre 4

CONVERGENCES DE MARTINGALE

4.1 Inégalités de martingale	22	4.3 Uniforme intégrabilité	26
4.1.1 Inégalité maximale de DOOB	22	4.4 Convergence L^1 et martingale fermée	28
4.1.2 Inégalité maximale dans L^p	23	4.5 Martingales de carré intégrable	29
4.1.3 Nombre de montée	23	4.6 Théorème d'arrêt	31
4.2 Convergence presque sûre de martingales	25		

4.1 INÉGALITÉS DE MARTINGALE

4.1.1 Inégalité maximale de DOOB

PROPOSITION 4.1 (inégalité maximale de DOOB pour une sous-martingale). Soit $X := (X_n)_{n \geq 0}$ une sous-martingale. Pour $n \geq 0$, on pose $\bar{X}_n := \max(X_0, \dots, X_n)$. Alors pour tous $n \geq 0$ et $x > 0$, on a

$$\mathbb{P}(\bar{X}_n \geq x) \leq \frac{\mathbb{E}[X_n \mathbb{1}_{\{\bar{X}_n \geq x\}}]}{x} \leq \frac{\mathbb{E}[X_n^+]}{x}.$$

Preuve La seconde inégalité est immédiate. Soient $n \geq 0$ et $x > 0$. On pose $A := \{\bar{X}_n \geq x\}$ et

$$S := \inf\{k \geq 0 \mid X_k \geq x\} \quad \text{et} \quad T := \min(T, n).$$

Alors les variables aléatoires T et S sont des temps d'arrêt et $T \leq n$. Par les résultats d'arrêt du chapitre précédente pour un temps d'arrêt T déterministiquement borné, on a

$$\mathbb{E}[X_T \mathbb{1}_A] + \mathbb{E}[X_T \mathbb{1}_{A^c}] = \mathbb{E}[X_T] \leq \mathbb{E}[X_n] = \mathbb{E}[X_n \mathbb{1}_A] + \mathbb{E}[X_n \mathbb{1}_{A^c}].$$

Sur A^c , on a $S > n$, donc $T = n$. D'où $\mathbb{E}[X_T \mathbb{1}_{A^c}] = \mathbb{E}[X_n \mathbb{1}_{A^c}]$. Il vient alors

$$\mathbb{E}[X_T \mathbb{1}_A] \leq \mathbb{E}[X_n \mathbb{1}_A].$$

Sur A , on a $S \leq n$, donc $T = S$ et $X_T = X_S \geq x$. On obtient alors

$$x \mathbb{P}(A) \leq \mathbb{E}[X_n \mathbb{1}_A].$$

Ceci montre l'inégalité. □

COROLLAIRE 4.2 (inégalité maximale de DOOB pour une martingale). Soit $X := (X_n)_{n \geq 0}$ une sous-martingale. Alors pour tous $n \geq 0$ et $x > 0$, on a

$$\mathbb{P}(\max_{k \leq n} |X_k| \geq x) \leq \frac{\mathbb{E}[|X_n|]}{x}.$$

Preuve Comme $(X_n)_{n \geq 0}$ est une martingale, la suite $(|X_n|)_{n \geq 0}$ est une martingale pour laquelle l'inégalité précédente donne celle recherchée puisque $|X_n|^+ \leq |X_n|$. □

COROLLAIRE 4.3 (inégalité maximale de KOLMOGOROV). Soit $(X_n)_{n \geq 0}$ une suite indépendante de variables aléatoires centrées de variances telle que $\sigma_n^2 := \mathbb{E}[X_n^2] < +\infty$ pour tout $n \geq 0$. Pour $n \geq 0$, on pose $S_n := X_0 + \dots + X_n$. Alors

$$\mathbb{P}(\max_{k \leq n} |S_k| \geq x) \leq \frac{\text{Var}(S_n)}{x^2}.$$

◇ REMARQUE. L'inégalité de TCHEBYCHEV permet d'écrire

$$\max_{k \leq n} \mathbb{P}(|S_k| \geq x) \leq \frac{\text{Var}(S_n)}{x^2}.$$

Ce corollaire permet de renforcer cette inégalité en faisant rentrer le maximum à l'intérieur de la probabilité.

Preuve La suite $(S_n)_{n \geq 0}$ est bien une martingale. D'après l'inégalité maximale de DOOB appliquée à la sous-martingale $(S_n^2)_{n \geq 0}$, on obtient

$$\mathbb{P}(\max_{k \leq n} S_k^2 \geq x^2) \leq \frac{\mathbb{E}[(S_n^2)^+]}{x^2}.$$

Or $\mathbb{E}[(S_n^2)^+] = \text{Var}(S_n)$ et $\{\max_{k \leq n} S_k^2 \geq x^2\} = \{\max_{k \leq n} |S_k| \geq x\}$. On en déduit le corollaire. □

4.1.2 Inégalité maximale dans L^p

PROPOSITION 4.4 (inégalité maximale dans L^p pour une sous-martingale). Soient $(X_n)_{n \geq 0}$ une sous-martingale et $p > 1$. Alors pour tout $n \geq 0$, on a

$$\mathbb{E}[\bar{X}_n^p] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[(X_n^+)^p].$$

Preuve On utilise l'inégalité maximale de DOOB à la suite tronquée à un niveau $M > 0$. Soient $n \geq 0$ et $p > 1$. Avec le théorème de FUBINI et l'inégalité de HÖLDER avec $q := p/(p-1)$, on a

$$\begin{aligned} \mathbb{E}[\min(\bar{X}_n^p, M^p)] &= \mathbb{E}[(\min(\bar{X}_n, M))^p] = \mathbb{E}\left[\int_0^{\min(\bar{X}_n, M)} px^{p-1} dx\right] \\ &= \mathbb{E}\left[\int_0^M px^{p-1} \mathbb{1}_{\{x \leq \bar{X}_n\}} dx\right] \\ &= \int_0^M px^{p-1} \mathbb{P}(\bar{X}_n \geq x) dx \\ &\leq p \int_0^M x^{p-1} x \mathbb{E}[X_n^+ \mathbb{1}_{\{\bar{X}_n \geq x\}}] dx \\ &= p \mathbb{E}\left[\int_0^M x^{p-2} X_n^+ \mathbb{1}_{\{\bar{X}_n \geq x\}} dx\right] \\ &= p \mathbb{E}\left[\left(\int_0^{\min(\bar{X}_n, M)} x^{p-2} dx\right) X_n^+\right] \\ &= \frac{p}{p-1} \mathbb{E}[\min(M, \bar{X}_n)^{p-1} X_n^+] \\ &\leq \frac{p}{p-1} \mathbb{E}[\min(M, \bar{X}_n)^p]^{(p-1)/p} \mathbb{E}[(X_n^+)^p]^{1/p}. \end{aligned}$$

On obtient alors

$$\mathbb{E}[\min(\bar{X}_n, M)^p] \leq \frac{p}{p-1} \mathbb{E}[\min(M, \bar{X}_n)^p]^{(p-1)/p} \mathbb{E}[(X_n^+)^p]^{1/p},$$

donc

$$\mathbb{E}[\min(\bar{X}_n, M)^p] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[(X_n^+)^p].$$

On achève la démonstration en faisant $M \rightarrow +\infty$ avec le théorème de convergence monotone. \square

COROLLAIRE 4.5 (inégalité maximale dans L^p pour une martingale). Soient $(X_n)_{n \geq 0}$ une martingale et $p > 1$. Alors pour tout $n \geq 0$, on a

$$\mathbb{E}[\max_{k \leq n} |X_k|^p] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}[|X_n|^p].$$

Preuve Il suffit d'appliquer l'inégalité précédente à la sous-martingale $(|X_n|)_{n \geq 0}$. \square

4.1.3 Nombre de montée

Soient $(X_n)_{n \geq 0}$ une sous-martingale et $a, b \in \mathbf{R}$ tels que $a < b$. On considère la suite $(N_k)_{k \geq 0}$ de variables aléatoires telle que

$$\begin{cases} N_0 = 0, \\ N_1 = \min\{n \geq 1 \mid X_n \leq a\}, \\ N_2 = \min\{n \geq N_1 \mid X_n \geq b\}, \\ N_{2k-1} = \min\{n \geq N_{2k-2} \mid X_n \leq a\}, \\ N_{2k} = \min\{n \geq N_{2k-1} \mid X_n \geq b\}, \quad k \geq 2 \end{cases}$$

LEMME 4.6. La suite $(N_k)_{k \geq 0}$ est strictement croissante. Par conséquent, on a $N_k \geq k$ pour tout $k \geq 0$. De plus, les variables aléatoires N_k sont toutes des temps d'arrêt.

Preuve Montrons que ce sont des temps d'arrêt. Soit $n \geq 0$. On a

$$\{N_1 \leq n\} = \bigcup_{k=1}^n \{X_k \leq a\} \in \mathcal{F}_n.$$

De plus, on a

$$\begin{aligned} \{N_2 = n\} &= \bigcup_{j=1}^{n-1} \{N_1 = j\} \cap \{N_2 = n\} \\ &= \bigcup_{j=1}^{n-1} \left(\{N_1 = j\} \cap \bigcap_{k=j+1}^{n-1} (\{X_k < b\} \cap \{X_n \geq b\}) \right) \in \mathcal{F}_n. \end{aligned}$$

On procède ensuite par récurrence sur l'entier $k \geq 0$. Soit $k \geq 2$. On suppose que N_{2k-2} est un temps d'arrêt. Alors pour tout $n \geq 0$, on a

$$\begin{aligned} \{N_{2k-1} = n\} &= \bigcup_{j=1}^{n-1} \{N_{2k-2} = j\} \cap \{N_{2k-1} = n\} \\ &= \bigcup_{j=1}^{n-1} \left(\{N_{2k-2} = j\} \cap \bigcap_{k=j+1}^{n-1} (\{X_k > a\} \cap \{X_n \geq b\}) \right) \in \mathcal{F}_n \end{aligned}$$

et, de même, on a $\{N_{2k} = n\} \in \mathcal{F}_n$. \square

Pour tout $k \geq 2$, comme $X_{N_{2k-1}} \leq a$ et $X_{N_{2k}} \geq b$, entre les dates N_{2k-1} et N_{2k} , la sous-martingale montre de en-dessous de a à au-dessus de b . Pour $n \geq 0$, on note $U_n(a, b)$ le nombre de montées de en-dessous de a à au-dessus de b jusqu'à la date n . Alors la suite $(U_n(a, b))_{n \geq 0}$ est croissante et $U_n(a, b) \leq \lfloor n/2 \rfloor$ pour tout $n \geq 0$. De plus, pour tous $n, k \geq 0$, écrire $N_{2k} \leq n \leq N_{2k+2}$ signifie $U_n(a, b) = k$, d'où

$$U_n(a, b) = \sum_{k=1}^{\lfloor n/2 \rfloor} k \mathbb{1}_{\{N_{2k} \leq n \leq N_{2k+2}\}}.$$

THÉORÈME 4.7 (nombre de montée). Soit $n \geq 0$. Alors

$$\mathbb{E}[U_n(a, b)] \leq \frac{\mathbb{E}[(X_n - a)^+]}{b - a}.$$

◇ REMARQUE. Le théorème est vrai pour des sous-martingales finies.

Preuve Pour $j \geq 1$, on pose

$$Y_j := \begin{cases} 1 & \text{s'il existe un entier } k \geq 0 \text{ tel que } N_{2k} < j \leq N_{2k+1}, \\ 0 & \text{sinon, i. e. s'il existe un entier } k \geq 0 \text{ tel que } N_{2k+1} < j \leq N_{2k+2}. \end{cases}$$

On observe que $Y_2 = \mathbb{1}_{\{N_1 \geq 2\}} = \mathbb{1}_{\{X_1 > a\}}$ et $Y_j = \mathbb{1}_{\{Y_{j-1}=0, X_{j-1} \geq b\} \cup \{Y_{j-1}=1, X_{j-1} > a\}}$ pour $j \geq 2$. En effet, on a $Y_j = 1$ lorsque X_j est dans une phase de descente et cela se produit

- soit lorsque X_{j-1} est dans une descente ($Y_{j-1} = 1$) qui n'est pas achevée ($X_{j-1} > a$),
- soit lorsque X_{j-1} est dans une montée ($Y_{j-1} = 0$) qui est achevée ($X_{j-1} \geq b$).

Des expressions des variables aléatoires Y_j , on remarque que la suite $(Y_j)_{j \geq 1}$ est prévisible.

LEMME 4.8. On a

$$\sum_{k=2}^n Y_k (X_k - X_{k-1}) \leq (a - b) U_n(a, b) + (X_n - a)^+.$$

Le preuve du théorème découle de ce lemme puisque, pour tout $k \geq 2$, on a

$$\begin{aligned} \mathbb{E}[Y_k (X_k - X_{k-1})] &= \mathbb{E}[\mathbb{E}[Y_k (X_k - X_{k-1}) \mid \mathcal{F}_{k-1}]] \\ &= \mathbb{E}[Y_k \mathbb{E}[X_k - X_{k-1} \mid \mathcal{F}_{k-1}]] \geq 0 \end{aligned}$$

car la suite $(X_k)_{k \geq 0}$ est une sous-martingale. On obtient alors

$$(a - b) \mathbb{E}[U_n(a, b)] + \mathbb{E}[(X_n - a)^+] \geq 0$$

ce qui donne le théorème. \square

Preuve du lemme • *Premier cas.* On suppose $U_n(a, b) = 0$, i. e. il n'y a pas de montées de a vers b jusqu'à la date n . Si $N_1 = 1$, alors $Y_k = 0$ pour tout $k \in \llbracket 2, n \rrbracket$ car on reste dans une phase de montée inachevée et l'inégalité est trivialement vraie. Si $1 < N_1 \leq n$, alors $Y_k = 1$ pour tout $k \leq N_1$ et $Y_k = 0$ pour $k \in \llbracket N_1 + 1, n \rrbracket$ et l'inégalité se réduit à

$$\sum_{k=2}^{N_1} (X_k - X_{k-1}) = X_{N_1} - X_1 < 0 \leq 0 + (X_n - a)^+$$

car $X_{N_1} \leq a$ et $X_1 > a$.

• *Second cas.* On suppose $U_n(a, b) > 0$, i. e. il y a au moins une phase de montée avant la date n . Alors $N_1 < N_2 \leq n$. Si $N_1 > 1$, alors $Y_j = 1$ pour tout $j \in \llbracket 1, N_1 \rrbracket$ et, comme $X_{N_1} \leq a < X_1$, on a

$$\begin{aligned} \sum_{j=2}^n Y_j(X_j - X_{j-1}) &= \sum_{j=1}^{N_1} \underbrace{(X_j - X_{j+1})}_{\leq 0} + \sum_{j=N_1+1}^n Y_j(X_j - X_{j-1}) \\ &\leq \sum_{j=N_1+1}^n Y_j(X_j - X_{j-1}). \end{aligned} \quad (*)$$

Ceci est même vraie pour $N_1 = 1$. Distinguons deux cas.

- Si la date n correspond à une phase de montée, alors $N_{2t} \leq N_{2t+1} < n \leq N_{2t+2}$ avec $t = U_n(a, b)$.
- Si la date n correspond à une phase de descente, alors $N_{2t'} < n \leq N_{2t'+1} \leq N_{2t'+2}$ avec $t' = U_n(a, b)$.

Dans le premier cas, on a

$$\begin{aligned} \sum_{j=N_2+1}^n Y_j(X_j - X_{j-1}) &= \sum_{j=N_2+1}^{N_3} Y_j(X_j - X_{j-1}) + \sum_{j=N_3+1}^{N_4} Y_j(X_j - X_{j-1}) + \dots \\ &= (X_{N_3} - X_{N_2}) + (X_{N_5} - X_{N_4}) + \dots + (X_{N_{2t+1}} - X_{N_{2t}}) + \underbrace{\sum_{j=N_{2t+1}+1}^n Y_j(X_j - X_{j-1})}_0 \\ &\leq t(a - b) = (a - b)U_n(a, b) \end{aligned}$$

et, avec l'inégalité (*), on obtient l'inégalité recherchée. Dans le second cas, on a

$$\begin{aligned} \sum_{j=N_2+1}^n Y_j(X_j - X_{j-1}) &= (X_{N_3} - X_{N_2}) + 0 + (X_{N_5} - X_{N_4}) + \dots + (X_{N_{2t'-1}} - X_{N_{2t'}}) + \underbrace{\sum_{j=N_{2t'}+1}^n Y_j(X_j - X_{j-1})}_0 \\ &\leq (a - b)(t' - 1) + X_n - X_{N_{2t'}} \\ &\leq (a - b)(t' - 1) + X_n - a + a - X_{N_{2t'}} \\ &\leq (a - b)t' + X_n - a \end{aligned}$$

ce qui conclut également. □

4.2 CONVERGENCE PRESQUE SÛRE DE MARTINGALES

THÉORÈME 4.9. Soit $(X_n)_{n \geq 0}$ une sous-martingale telle que $\sup_{n \geq 0} \mathbb{E}[X_n^+] < +\infty$. Alors cette suite $(X_n)_{n \geq 0}$ converge presque sûrement vers une variable aléatoire intégrable.

- ◇ **REMARQUES.** – L'hypothèse est satisfaite si $\sup_{n \geq 0} \mathbb{E}[|X_n|] < +\infty$. Pour une sous-martingale ou une sur-martingale, cette dernière hypothèse assure la convergence presque sûre.
- Une sur-martingale positive $(X_n)_{n \geq 0}$ converge presque sûrement vers une variable X telle que $\mathbb{E}[X] \leq \mathbb{E}[X_0]$. En effet, la suite $(Y_n := -X_n)_{n \geq 0}$ est une sous-martingale vérifiant $\mathbb{E}[Y_n^+] = 0$ pour tout $n \geq 0$. Elle converge donc presque sûrement vers une variable Y . Donc la sur-martingale $(X_n)_{n \geq 0}$ converge presque sûrement vers $X := -Y$. De plus, le lemme de FATOU donne

$$\mathbb{E}[X] = \mathbb{E}[\liminf_{n \rightarrow +\infty} X_n] \leq \liminf_{n \rightarrow +\infty} \mathbb{E}[X_n] \leq \mathbb{E}[X_0].$$

Preuve Une suite $x := (x_n)_{n \geq 0}$ converge dans $\bar{\mathbf{R}}$ si et seulement si, pour tous $a, b \in \mathbf{Q}$ tels que $a < b$, le nombre de montées de la suite de a vers b , noté $U_\infty([a, b], x)$, est fini. En effet, la suite diverge si ses limites inférieure et supérieure sont égales, i. e. il existe $a, b \in \mathbf{Q}$ tels que $a < b$ et

$$\liminf_{n \rightarrow +\infty} x_n < a < b < \limsup_{n \rightarrow +\infty} x_n,$$

et cela à lieu si et seulement si $U_\infty([a, b], x) = +\infty$.

Ainsi la suite $X := (X_n)_{n \geq 0}$ converge dans $\bar{\mathbf{R}}$ si et seulement si $U_\infty([a, b], X) < +\infty$ pour tous rationnels $a, b \in \mathbf{Q}$ tels que $a < b$. Soient $a, b \in \mathbf{Q}$ tels que $a < b$. Alors

$$U_\infty([a, b], X) = \lim_{n \rightarrow +\infty} U_n(a, b).$$

Le théorème de convergence monotone donne alors

$$\mathbb{E}[U_\infty([a, b], X)] = \lim_{n \rightarrow +\infty} \mathbb{E}[U_n(a, b)].$$

On en déduit

$$\mathbb{E}[U_\infty([a, b], X)] \leq \sup_{n \geq 0} \frac{\mathbb{E}[(X_n - a)^+]}{b - a} \leq \frac{\sup_{n \geq 0} \mathbb{E}[X_n^+] + |a|}{b - a} < +\infty.$$

Alors la variable $U_\infty([a, b], X)$ est presque sûrement finie et on a

$$\{\liminf_{n \rightarrow +\infty} X_n < a < b < \limsup_{n \rightarrow +\infty} X_n\} \subseteq \{U_\infty([a, b], X) = +\infty\}.$$

Ainsi l'ensemble

$$\{\liminf_{n \rightarrow +\infty} X_n < \limsup_{n \rightarrow +\infty} X_n\} = \bigcup_{a, b \in \mathbf{Q}} \{\liminf_{n \rightarrow +\infty} X_n < a < b < \limsup_{n \rightarrow +\infty} X_n\}$$

est négligeable, *i. e.* la suite $(X_n)_{n \geq 0}$ converge presque sûrement vers une variable aléatoire X .

Il reste à établir l'intégrabilité de X . Le lemme de FATOU donne

$$\mathbb{E}[X^+] = \mathbb{E}[\liminf_{n \rightarrow +\infty} X_n^+] \leq \liminf_{n \rightarrow +\infty} \mathbb{E}[X_n^+] \leq \sup_{n \geq 0} \mathbb{E}[X_n^+] < +\infty.$$

Par ailleurs, comme $(X_n)_{n \geq 0}$ est une sous-martingale, on a $\mathbb{E}[X_n] \geq \mathbb{E}[X_0]$, donc $\mathbb{E}[X_n^-] \leq \mathbb{E}[X_n^+] - \mathbb{E}[X_0]$. En appliquant encore le lemme de FATOU, on a

$$\mathbb{E}[X^-] = \mathbb{E}[\liminf_{n \rightarrow +\infty} X_n^-] \leq \liminf_{n \rightarrow +\infty} \mathbb{E}[X_n^-] \leq \liminf_{n \rightarrow +\infty} \mathbb{E}[X_n^+] - \mathbb{E}[X_0] < +\infty.$$

Finalement, on a $\mathbb{E}[|X|] = \mathbb{E}[X^+] + \mathbb{E}[X^-] < +\infty$, *i. e.* la limite X est intégrable. \square

- ◊ REMARQUE. La convergence n'est pas dans L^1 . Par exemple, prenons $(X_i)_{i \geq 0}$ une suite de variables aléatoires de RADEMACHER. Pour $n \geq 0$, on pose $S_n := X_1 + \dots + X_n + 1$. On considère le temps d'arrêt $T := \infty \{n \geq 0 \mid S_n = 0\}$. Alors la suite $(S_{\min(n, T)})_{n \geq 0}$ est une martingale qui ne converge dans L^1 .

COROLLAIRE 4.10. Soit $(X_n)_{n \geq 0}$ une martingale ou une sur/sous-martingale. Si elle est bornée dans L^1 , elle est converge presque sûrement.

4.3 UNIFORME INTÉGRABILITÉ

DÉFINITION 4.11. Une famille de variables aléatoires intégrables $(X_i)_{i \in I}$ est dite *uniformément intégrable* si

$$\lim_{c \rightarrow +\infty} \sup_{i \in I} \mathbb{E}[|X_i| \mathbb{1}_{\{|X_i| > c\}}] = 0.$$

- ◊ REMARQUES. – Pour toute variable aléatoire intégrable X , la famille (X) est uniformément intégrable par le théorème de convergence dominée.
– On suppose qu'il existe une variable aléatoire intégrable Z telle que $|X_i| \leq Z$ pour $i \in I$. Alors la famille $(X_i)_{i \in I}$ est uniformément intégrable car, comme la fonction $x \geq 0 \mapsto x \mathbb{1}_{\{x \geq c\}}$ est croissante pour tout $c \in \mathbf{R}$, donc

$$\sup_{i \in I} \mathbb{E}[|X_i| \mathbb{1}_{\{|X_i| > c\}}] \leq \mathbb{E}[Z \mathbb{1}_{\{Z > c\}}]$$

ce qui assure la limite.

– Si les variables X_i sont intégrables et l'ensemble I est fini, alors la famille est uniformément intégrable en appliquant le point précédent avec $Z := \sum_{i \in I} |X_i|$.

PROPOSITION 4.12. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires intégrables telle qu'il existe $\delta > 0$ tel que

$$\sup_{i \geq 0} \mathbb{E}[|X_i|^{1+\delta}] < +\infty.$$

Alors la suite $(X_n)_{n \geq 1}$ est uniformément intégrable.

Preuve Soit $c \in \mathbf{R}$. Pour tout $n \geq 1$, on a

$$\mathbb{E}[|X_n| \mathbb{1}_{\{|X_n| > c\}}] \leq \mathbb{E}\left[|X_n| \left(\frac{|X_n|}{c}\right)^\delta \mathbb{1}_{\{|X_n| > c\}}\right] \leq c^{-\delta} \mathbb{E}[|X_n|^{1+\delta}].$$

Ainsi on obtient

$$\sup_{n \geq 0} \mathbb{E}[|X_n| \mathbb{1}_{\{|X_n| > c\}}] \leq c^{-\delta} \sup_{n \geq 0} \mathbb{E}[|X_n|^{1+\delta}] = O(c^{-\delta}) \rightarrow 0.$$

D'où la proposition. \square

4.3. UNIFORME INTÉGRABILITÉ

LEMME 4.13. Soit X un variable aléatoire intégrable. Pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que, pour tout $A \in \mathcal{F}$ tel que $\mathbb{P}(A) < \delta$, on ait $\mathbb{E}[|X| \mathbb{1}_A] < \varepsilon$.

PROPOSITION 4.14. Une suite de variables aléatoires $(X_n)_{n \geq 0}$ est uniformément intégrable si et seulement si

- (i) pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que, pour tout $A \in \mathcal{F}$ tel que $\mathbb{P}(A) < \delta$, on ait $\sup_{n \geq 0} \mathbb{E}[|X_n| \mathbb{1}_A] < \varepsilon$;
- (ii) on a $\sup_{n \geq 0} \mathbb{E}[|X_n|] < +\infty$.

Preuve \Rightarrow On suppose que la suite $(X_n)_{n \geq 0}$ est uniformément intégrable. Soit $\varepsilon > 0$. Alors il existe $c_0 \in \mathbf{R}$ tel que

$$\forall c \geq c_0, \quad \sup_{n \geq 0} \mathbb{E}[|X_n| \mathbb{1}_{\{|X_n| > c\}}] < \varepsilon/2.$$

On pose $\delta := \varepsilon/2c_0$. Alors pour tout $A \in \mathcal{F}$ tel que $\mathbb{P}(A) < \delta$ et pour tout $n \geq 0$, on a

$$\begin{aligned} \mathbb{E}[|X_n| \mathbb{1}_A] &= \mathbb{E}[|X_n| \mathbb{1}_A \mathbb{1}_{\{|X_n| \leq c\}}] + \mathbb{E}[|X_n| \mathbb{1}_A \mathbb{1}_{\{|X_n| > c\}}] \\ &\leq c\mathbb{P}(A) + \varepsilon/2 \leq \varepsilon. \end{aligned}$$

D'où le point (i). En prenant $A = \Omega$, pour tout $n \geq 0$, on obtient $\mathbb{E}[|X_n|] \leq c_0 + \varepsilon/2$ ce qui assure le point (ii).

\Leftarrow On suppose les points (i) et (ii). Notons $M := \sup_{n \geq 0} \mathbb{E}[|X_n|] < +\infty$. Soit $\varepsilon > 0$. Alors il existe $\delta > 0$ tel que, pour tout $A \in \mathcal{F}$ tel que $\mathbb{P}(A) < \delta$, on ait $\sup_{n \geq 0} \mathbb{E}[|X_n| \mathbb{1}_A] < \varepsilon$. On pose $c_0 := M/\delta$. Alors pour tout $c \geq c_0$, on obtient $\mathbb{P}(|X_n| > c) \leq M/c < \delta$. On en déduit

$$\mathbb{E}[|X_n| \mathbb{1}_{\{|X_n| > c\}}] < \varepsilon, \quad \forall c \geq c_0, \forall n \geq 0.$$

Ceci assure l'uniforme intégrabilité. □

PROPOSITION 4.15. Soit X un variable aléatoire intégrable. Alors la famille $(\mathbb{E}[X | \mathcal{G}])_{\mathcal{G} \subset \mathcal{F}}$ est uniformément intégrable.

Preuve Pour toute sous-tribu \mathcal{G} de \mathcal{F} , la variable aléatoire $X_{\mathcal{G}} := \mathbb{E}[X | \mathcal{G}]$ vérifie

$$\mathbb{E}[|X_{\mathcal{G}}|] = \mathbb{E}[\mathbb{E}[|X| | \mathcal{G}]] = \mathbb{E}[|X|].$$

On obtient alors

$$\sup_{\mathcal{G} \subset \mathcal{F}} \mathbb{E}[|X_{\mathcal{G}}|] = \mathbb{E}[|X|] < +\infty$$

ce qui donne le point (i). Soit $\varepsilon > 0$. De plus, comme X est intégrable, il existe $\delta > 0$ tel que, pour tout $A \in \mathcal{F}$ tel que $\mathbb{P}(A) < \delta$, on ait $\mathbb{E}[|X| \mathbb{1}_A] < \varepsilon$. Alors pour tout $c > c_0 := \mathbb{E}[|X|]/\delta$ et pour toute sous-tribu \mathcal{G} de \mathcal{F} , on a

$$\mathbb{P}(X_{\mathcal{G}} > 0) \leq \frac{\mathbb{E}[X_{\mathcal{G}}]}{c} = \frac{\mathbb{E}[|X|]}{c} < \delta,$$

et, en prenant $A := \{X_{\mathcal{G}} > c\} \in \mathcal{G}$, on obtient

$$\mathbb{E}[X_{\mathcal{G}} \mathbb{1}_{\{X_{\mathcal{G}} > c\}}] = \mathbb{E}[|X| \mathbb{1}_{\{X_{\mathcal{G}} > c\}}] < \varepsilon.$$

Finalement, pour tout $c \geq c_0$, on a

$$\sup_{\mathcal{G} \subset \mathcal{F}} \mathbb{E}[X_{\mathcal{G}} \mathbb{1}_{\{X_{\mathcal{G}} > c\}}] < \varepsilon$$

ce qui donne le point (ii). Avec la proposition précédente, ceci assure l'uniforme intégrabilité de cette suite. □

THÉORÈME 4.16 (VITALI). Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires intégrables. Alors les propositions suivantes sont équivalentes :

- (i) la suite converge dans L^1 ;
- (ii) elle converge en probabilité et elle est uniformément intégrable.

Preuve \Rightarrow On suppose que la suite converge dans L^1 . Alors elle converge en probabilité par l'inégalité de MARKOV. Puis la convergence dans L^1 implique la convergence des moments, donc $\sup_{n \geq 0} \mathbb{E}[|X_n|] < +\infty$. De plus, comme X est intégrable, il existe $\delta > 0$ tel que, pour tout $A \in \mathcal{F}$ tel que $\mathbb{P}(A) < \delta$, on ait $\mathbb{E}[|X_n| \mathbb{1}_A] < \varepsilon/2$. Par ailleurs, comme la suite converge dans L^1 vers une variable aléatoire X , il existe $n_0 \geq 0$ tel que

$$\forall n \geq 0, \quad \mathbb{E}[|X_n - X|] < \varepsilon/2.$$

Ainsi l'inégalité triangulaire et un passage à la borne supérieure donne

$$\forall A \in \mathcal{F}, \quad \mathbb{P}(A) < \delta \quad \Rightarrow \quad \sup_{n \geq n_0} \mathbb{E}[|X_n| \mathbb{1}_A] \leq \varepsilon.$$

De plus, la famille finie de variables aléatoires $(X_i)_{i \leq n_0}$ est uniformément intégrable, donc il existe $\delta' > 0$ tel que

$$\forall A \in \mathcal{F}, \quad \mathbb{P}(A) < \delta' \quad \Rightarrow \quad \sup_{n \leq n_0} \mathbb{E}[|X_n| \mathbb{1}_A] \leq \varepsilon.$$

On obtient alors

$$\forall A \in \mathcal{F}, \quad \mathbb{P}(A) < \min(\delta, \delta') \implies \sup_{n \geq 0} \mathbb{E}[|X_n| \mathbb{1}_A] \leq \varepsilon.$$

Par le critère d'intégrabilité uniforme, la suite est uniformément intégrable.

\Leftarrow On suppose que la suite converge en probabilité vers une variable aléatoire X et qu'elle est uniformément intégrable. On peut alors en extraire une sous-suite $(X_{n_k})_{k \geq 0}$ qui converge presque sûrement vers X . Alors cette limite est intégrable car le lemme de FATOU assure

$$\mathbb{E}[|X|] = \mathbb{E}[\liminf_{k \rightarrow +\infty} |X_{n_k}|] \leq \liminf_{k \rightarrow +\infty} \mathbb{E}[|X_{n_k}|] \leq \sup_{n \geq 0} \mathbb{E}[|X_n|] < +\infty$$

puisque la suite est uniformément intégrable. Soit $\varepsilon > 0$. Pour tout $n \geq 0$, l'inégalité triangulaire donne

$$\mathbb{E}[|X_n - X|] \leq \mathbb{E}[|X_n - X| \mathbb{1}_{\{|X_n - X| \leq \varepsilon/3\}}] + \mathbb{E}[|X_n| \mathbb{1}_{\{|X_n - X| > \varepsilon/3\}}] + \mathbb{E}[|X| \mathbb{1}_{\{|X_n - X| > \varepsilon/3\}}]. \quad (*)$$

Comme la famille $(X) \cup (X_n)_{n \geq 0}$ est uniformément intégrable, il existe $\delta > 0$ tel que

$$\forall A \in \mathcal{F}, \quad \mathbb{P}(A) < \delta \implies \begin{cases} \mathbb{E}[|X| \mathbb{1}_A] < \varepsilon/3, \\ \mathbb{E}[|X_n| \mathbb{1}_A] < \varepsilon/3, \quad \forall n \geq 0. \end{cases} \quad (**)$$

Or $\mathbb{P}(|X_n - X| > \varepsilon/3) \rightarrow 0$, donc il existe $n_0 \geq 0$ tel que

$$\forall n \geq n_0, \quad \mathbb{P}(|X_n - X| > \varepsilon/3) < \delta.$$

En utilisant $A := \{|X_n - X| > \varepsilon/3\}$ dans la relation (**), on obtient alors

$$\begin{cases} \mathbb{E}[|X| \mathbb{1}_A] < \varepsilon/3, \\ \mathbb{E}[|X_n| \mathbb{1}_A] < \varepsilon/3, \quad \forall n \geq n_0. \end{cases}$$

En reprenant l'inégalité (*), pour tout $n \geq n_0$, on obtient $\mathbb{E}[|X_n - X|] < \varepsilon$ ce qui montre la convergence L^1 . \square

4.4 CONVERGENCE L^1 ET MARTINGALE FERMÉE

DÉFINITION 4.17. Une martingale $(X_n)_{n \geq 0}$ est dite *fermée* s'il existe une variable aléatoire intégrable X telle que

$$\forall n \geq 0, \quad X_n = \mathbb{E}[X | \mathcal{F}_n].$$

THÉORÈME 4.18 (convergence des sous-martingales uniformément intégrables). Soit $(X_n)_{n \geq 0}$ une sous-martingale. Alors les propositions suivantes sont équivalentes :

- (i) la martingale est uniformément intégrable;
- (ii) elle converge presque sûrement et dans L^1 .

Preuve Si elle est uniformément intégrable, alors elle est bornée dans L^1 , donc elle converge presque sûrement, donc elle converge en probabilité, donc elle converge dans L^1 d'après le théorème précédent puisqu'il y a intégrabilité uniforme. L'autre sens découle du théorème de VITALI. \square

THÉORÈME 4.19 (convergence des martingales uniformément intégrables). Soit $(X_n)_{n \geq 0}$ une martingale. Alors les propositions suivantes sont équivalentes :

- (i) la martingale est uniformément intégrable;
- (ii) elle converge presque sûrement et dans L^1 ;
- (iii) elle est fermée.

Preuve L'implication (iii) \Rightarrow (i) vient de la proposition 4.15 et l'implication (i) \Rightarrow (ii) du théorème précédent. On suppose le point (ii). Montrons le point (iii). Pour tous $k \geq n$ et $A \in \mathcal{F}_n$, on a

$$\mathbb{E}[X_k \mathbb{1}_A] = \mathbb{E}[X_n \mathbb{1}_A] \quad \text{et} \quad \mathbb{E}[X_k \mathbb{1}_A] \longrightarrow \mathbb{E}[X \mathbb{1}_A]$$

car la suite converge dans L^1 , donc $\mathbb{E}[X_n \mathbb{1}_A] = \mathbb{E}[X \mathbb{1}_A]$. Ceci étant vrai pour tout événement $A \in \mathcal{F}_n$, on obtient donc $X_n = \mathbb{E}[X | \mathcal{F}_n]$ pour tout $n \geq 0$. La martingale est donc fermée. \square

Quelques applications des résultats de convergence

PROPOSITION 4.20. Soit $(\mathcal{F}_n)_{n \geq 0}$ une filtration. On pose $\mathcal{F}_\infty := \sigma(\bigcup_{n \geq 0} \mathcal{F}_n)$. Soit X un variable aléatoire intégrable. Alors

$$\mathbb{E}[X | \mathcal{F}_n] \xrightarrow{-ps \& L^1} \mathbb{E}[X | \mathcal{F}_\infty].$$

Preuve La suite $(X_n)_{n \geq 0}$ avec $X_n := \mathbb{E}[X | \mathcal{F}_n]$ est une martingale fermée, donc elle converge presque sûrement et dans L^1 vers une variable aléatoire intégrable Z . Pour tous $n \geq 0$ et $A \in \mathcal{F}_n$, on a

$$\mathbb{E}[X \mathbb{1}_A] = \mathbb{E}[X_n \mathbb{1}_A] \longrightarrow \mathbb{E}[Z \mathbb{1}_A],$$

donc $\mathbb{E}[X \mathbb{1}_A] = \mathbb{E}[Z \mathbb{1}_A]$. Cette égalité est également vraie pour tout $A \in \bigcup_{n \geq 0} \mathcal{F}_n$. Posons

$$\mathcal{M} := \{A \in \mathcal{F} \mid \mathbb{E}[X \mathbb{1}_A] = \mathbb{E}[Z \mathbb{1}_A]\}.$$

Cette classe \mathcal{M} est monotone et contient $\bigcup_{n \geq 0} \mathcal{F}_n$ qui est un π -système car la suite $(\mathcal{F}_n)_{n \geq 0}$ est une filtration. On en déduit $\mathcal{F}_\infty \subset \mathcal{M}$. De plus, comme les variables X_n sont \mathcal{F}_∞ -mesurable, leur limite Z est \mathcal{F}_∞ -mesurable. D'où $Z = \mathbb{E}[X | \mathcal{F}_\infty]$ presque sûrement. \square

PROPOSITION 4.21 (loi du 0-1 de LÉVY). Soit $(\mathcal{F}_n)_{n \geq 0}$ une filtration. Pour tout $A \in \mathcal{F}_\infty$, on a

$$\mathbb{E}[\mathbb{1}_A | \mathcal{F}_n] \xrightarrow{-ps \& L^1} \mathbb{1}_A.$$

Preuve Le résultat est immédiat avec la proposition précédente en posant $X := \mathbb{1}_A$. \square

PROPOSITION 4.22 (loi du 0-1 de KOLMOGOROV). Soient $(X_n)_{n \geq 0}$ une suite indépendante de variables aléatoires et $(\mathcal{F}_n)_{n \geq 0}$ sa filtration naturelle associée. Pour $n \geq 1$, on pose $\mathcal{F}^{(n)} := \sigma(X_p \mid p \geq n)$. On pose $\mathcal{F}^{(\infty)} := \bigcap_{n \geq 1} \mathcal{F}^{(n)}$. Alors la tribu $\mathcal{F}^{(\infty)}$ est triviale, i. e. tout événement de $\mathcal{F}^{(\infty)}$ est négligeable ou presque sûr.

Preuve Soit $A \in \mathcal{F}^{(\infty)}$. Pour tout $n \geq 1$, les variables X_1, \dots, X_n sont indépendantes de l'indicatrice $\mathbb{1}_A$ puisqu'on a $A \in \mathcal{F}^{(\infty)} \subset \mathcal{F}^{(n+1)}$, donc

$$\mathbb{E}[\mathbb{1}_A | \mathcal{F}_n] = \mathbb{E}[\mathbb{1}_A] = \mathbb{P}(A).$$

Par ailleurs, la proposition précédente assure $\mathbb{E}[\mathbb{1}_A | \mathcal{F}_n] \xrightarrow{-ps} \mathbb{E}[\mathbb{1}_A | \mathcal{F}_\infty]$. Or $\mathcal{F}^{(\infty)} \subset \mathcal{F}_\infty$, on a $A \in \mathcal{F}_\infty$ ce qui implique $\mathbb{E}[\mathbb{1}_A | \mathcal{F}_\infty] = \mathbb{1}_A$. On obtient alors $\mathbb{P}(A) = \mathbb{1}_A$ presque sûrement, donc $\mathbb{P}(A) \in \{0, 1\}$.

Justifions juste le point $\mathcal{F}^{(\infty)} \subset \mathcal{F}_\infty$. Pour tout $n \geq 0$, on a $\mathcal{F}^{(n)} = \sigma(\bigcup_{p \geq 0} \mathcal{F}_n^p)$ avec $\mathcal{F}_n^p := \sigma(X_i \mid i \in \llbracket n, p \rrbracket)$. On remarque que, pour tous $n, p \geq 0$ tels que $n \leq p$, on a $\mathcal{F}_n^p \subset \mathcal{F}_n \subset \mathcal{F}_\infty$. On en déduit alors que $\mathcal{F}^{(n)} \subset \mathcal{F}_\infty$ pour tout $n \geq 0$, donc $\mathcal{F}^{(\infty)} \subset \mathcal{F}_\infty$. \square

THÉORÈME 4.23. Soient $(X_n)_{n \geq 1}$ une suite de variables aléatoires tendant presque sûrement vers une variable aléatoire X_∞ et Z une variable aléatoire intégrable dominant les variables aléatoires X_n . Alors

$$\mathbb{E}[X_n | \mathcal{F}_n] \xrightarrow{-ps} \mathbb{E}[X_\infty | \mathcal{F}_\infty].$$

Preuve Pour $N \geq 0$, on pose $W_N := \sup_{n, m \geq N} |X_n - X_m| \in L^1(\mathcal{F})$. Un passage à la limite donne $|X_n - X_\infty| \leq W_N$ pour tous $n, N \geq 1$ tels que $n \geq N$. Pour tout $N \geq 1$, la proposition 4.20 assure alors

$$\limsup_{n \rightarrow +\infty} \mathbb{E}[|X_n - X_\infty| | \mathcal{F}_n] \leq \lim_{n \rightarrow +\infty} \mathbb{E}[W_N | \mathcal{F}_n] = \mathbb{E}[W_N | \mathcal{F}_\infty]. \quad (*)$$

Comme la suite $(W_N)_{N \geq 1}$ décroît vers 0, la suite $(\mathbb{E}[W_N | \mathcal{F}_\infty])_{N \geq 1}$ converge presque sûrement vers une variable aléatoire $U \geq 0$. Comme $0 \leq W_N \leq 2Z$, le théorème de convergence dominée usuel assure $\mathbb{E}[W_N] \rightarrow 0$. De même, en appliquant le même théorème, comme $\mathbb{E}[W_N | \mathcal{F}_\infty] \xrightarrow{-ps} U$ et $0 \leq \mathbb{E}[W_N | \mathcal{F}_\infty] \leq 2\mathbb{E}[Z | \mathcal{F}_\infty]$ pour tout $N \geq 1$, on a

$$\mathbb{E}[W_N] = \mathbb{E}[\mathbb{E}[W_N | \mathcal{F}_\infty]] \rightarrow \mathbb{E}[U].$$

On en déduit $\mathbb{E}[U] = 0$ avec $U \geq 0$, donc $U = 0$ presque sûrement. De la relation (*), on obtient donc

$$\mathbb{E}[|X_n - X_\infty| | \mathcal{F}_n] \rightarrow 0.$$

Ainsi, avec l'inégalité triangulaire inversée, on obtient

$$|\mathbb{E}[X_n | \mathcal{F}_n] - \mathbb{E}[X_\infty | \mathcal{F}_n]| \rightarrow 0.$$

Avec la proposition 4.20, on a aussi $\mathbb{E}[X_\infty | \mathcal{F}_n] \xrightarrow{-ps} \mathbb{E}[X_\infty | \mathcal{F}_\infty]$. D'où $\mathbb{E}[X_n | \mathcal{F}_n] \xrightarrow{-ps} \mathbb{E}[X_\infty | \mathcal{F}_\infty]$. \square

4.5 MARTINGALES DE CARRÉ INTÉGRABLE

RAPPEL. Pour une martingale $X := (X_n)_{n \geq 0}$ dont les variables aléatoires sont de carré intégrable, la décomposition de DOOB assure que la suite $(X_n^2 - \langle X, X \rangle_n)_{n \geq 0}$ est une martingale.

THÉORÈME 4.24 (convergence presque sûre de martingale L^2). Soit $X := (X_n)_{n \geq 0}$ une martingale dont les variables aléatoires sont de carré intégrable. Alors

1. pour tout $\omega \in \Omega$ tel que $\langle X, X \rangle_\infty(\omega) < +\infty$, la suite $(X_n(\omega))_{n \geq 0}$ admet une limite;
2. si la martingale X a ses accroissements bornés (i. e. il existe $K > 0$ tel que $|X_n - X_{n-1}| \leq K$ pour tout $n \geq 1$), alors $\langle X, X \rangle_\infty(\omega) < +\infty$ pour tout $\omega \in \Omega$ tel que la suite $(X_n(\omega))_{n \geq 0}$ admette une limite.

Preuve 1. Sans perte de généralité, on suppose $X_0 = 0$. On note $A := \langle X, X \rangle$. Soit $k \geq 0$. On pose

$$S_k = \inf\{p \geq 0 \mid A_{p+1} \geq k\}$$

qui est un temps d'arrêt puisque, pour tout $n \geq 0$, on a

$$\{S_k \leq n\} = \bigcup_{p \leq n} \{A_{p+1} \geq k\} \in \mathcal{F}_n.$$

Considérons la martingale X^{S_k} . On a $A^{S_k} = \langle X^{S_k}, X^{S_k} \rangle$, donc la suite $(X^{S_k})^2 - A^{S_k}$ est une martingale nulle en 0, donc son espérance est nulle, i. e.

$$\forall n \geq 0, \quad \mathbb{E}[(X_n^{S_k})^2] = \mathbb{E}[A_n^{S_k}] \leq k.$$

Ceci assure

$$\sup_{n \geq 0} \mathbb{E}[(X_n^{S_k})^2] \leq k.$$

Le théorème de convergence presque sûrement des martingales donne alors la convergence presque sûrement de la martingale $(X_n^{S_k})_{n \geq 0}$. Sur $\omega \in \{A_\infty < +\infty\}$, il existe un entier $k(\omega) \geq 0$ tel que $S_{k(\omega)}(\omega) = +\infty$ et on obtient donc

$$X_n^{S_{k(\omega)}(\omega)}(\omega) = X_n(\omega),$$

c'est-à-dire que la suite $(X_n(\omega))_{n \geq 0}$ admet une limite.

2. Raisonnons par l'absurde et supposons

$$\mathbb{P}(A_\infty = +\infty, \sup_{n \geq 0} |X_n| < +\infty) > 0. \quad (*)$$

Soit $c \geq 0$. On considère le temps d'arrêt $T_c := \inf\{n \geq 0 \mid |X_n| \geq c\}$. Par l'hypothèse faite, pour $c \geq 0$ assez grand, on a $\mathbb{P}(A_\infty = +\infty, T_c = +\infty) > 0$. Soit $n \geq 0$. D'après le théorème d'arrêt borné appliqué au temps d'arrêt $\min(T_c, n)$ et à la martingale $X^2 - A$, on a

$$\mathbb{E}[X_{\min(T_c, n)}^2 - A_{\min(T_c, n)}] = \mathbb{E}[X_0^2 - A_0] = 0,$$

c'est-à-dire $\mathbb{E}[A_{\min(T_c, n)}] = \mathbb{E}[X_{\min(T_c, n)}^2]$. Comme juste avant T_c , la martingale X est plus petite que c puis il y a une accroissement d'au plus k , on a $X_{\min(T_c, n)} \leq c + k$, donc $\mathbb{E}[A_{\min(T_c, n)}] \leq (c + k)^2$. D'où $\mathbb{E}[A_{T_c}] \leq (c + k)^2$. Cela contredit l'hypothèse (*). Finalement, on doit avoir $A_\infty < +\infty$ dès que $\sup_{n \geq 0} |X_n| < +\infty$ et, en particulier, dès que la limite existe dans \mathbf{R} . \square

THÉORÈME 4.25 (loi des grands nombres pour une martingale). Soit $X := (X_n)_{n \geq 0}$ une martingale dont les variables aléatoires sont de carré intégrable avec $X_0 = 0$. Alors sur $\{\langle X, X \rangle_\infty = +\infty\}$, on a

$$\frac{X_n}{\langle X, X \rangle_n} \xrightarrow{-ps} 0.$$

- ◊ REMARQUE. Soit $(X_i)_{i \geq 0}$ une suite indépendante de variables aléatoires identiquement distribuées et de carré intégrable. Alors la suite $(X_0 + \dots + X_n)_{n \geq 0}$ est une martingale dont le compensateur est la suite $(n \text{Var}(X_1))_{n \geq 0}$. Le théorème précédent donne alors

$$\frac{X_1 + \dots + X_n}{n \text{Var}(X_1)} \xrightarrow{-ps} 0.$$

Preuve La suite $((1 + \langle X, X \rangle_n)^{-1})_{n \geq 1}$ est une suite prévisible bornée. La suite $W := (W_n)_{n \geq 0}$ définie par

$$W_0 := 0 \quad \text{et} \quad W_n := ((1 + \langle X, X \rangle_n)^{-1} \cdot X)_n, \quad n \geq 1$$

est une martingale. Pour tout $n \geq 1$, on a

$$\begin{aligned} \langle W, W \rangle_n - \langle W, W \rangle_{n-1} &= \mathbb{E}[(W_n - W_{n-1})^2 \mid \mathcal{F}_{n-1}] \\ &= \frac{1}{(1 + \langle X, X \rangle_n)^2} \mathbb{E}[(X_n - X_{n-1})^2 \mid \mathcal{F}_{n-1}] \\ &= \frac{1}{(1 + \langle X, X \rangle_n)^2} (\mathbb{E}[X_n^2 \mid \mathcal{F}_{n-1}] - \mathbb{E}[X_{n-1}^2 \mid \mathcal{F}_{n-1}]) \end{aligned}$$

4.6. THÉORÈME D'ARRÊT

$$\begin{aligned}
 &= \frac{1}{(1 + \langle X, X \rangle_n)^2} (\langle X, X \rangle_n - \langle X, X \rangle_{n-1}) \\
 &\leq \frac{1}{(1 + \langle X, X \rangle_n)(1 + \langle X, X \rangle_{n-1})} (\langle X, X \rangle_n - \langle X, X \rangle_{n-1}) \\
 &= \frac{1}{1 + \langle X, X \rangle_{n-1}} - \frac{1}{1 + \langle X, X \rangle_n},
 \end{aligned}$$

donc

$$\langle W, W \rangle_n = \sum_{k=1}^n (\langle W, W \rangle_k - \langle W, W \rangle_{k-1}) \leq 1.$$

D'après le théorème précédente, la suite $(W_n)_{n \geq 0}$ converge presque sûrement. Pour conclure, utilisons le lemme suivant.

LEMME 4.26 (KRONECKER). Soient $(x_n)_{n \geq 0}$ une suite d'un espace vectoriel normé et $(a_n)_{n \geq 0}$ une suite réelle croissante qui converge vers $+\infty$. Si la série $\sum_{n \geq 0} x_n / a_n$ converge, alors

$$\frac{1}{a_n} \sum_{k=1}^n x_k \longrightarrow 0.$$

On applique le lemme aux suites $(X_k - X_{k-1})_{k \geq 1}$ et $(1 + \langle X, X \rangle_k)_{k \geq 1}$ et on obtient

$$\frac{X_n}{1 + \langle X, X \rangle_n} \xrightarrow{\text{ps}} 0.$$

De là, on en déduit le théorème sur $\{\langle X, X \rangle_\infty = +\infty\}$. \square

Preuve du lemme Pour $n \geq 1$, posons $S_n = \sum_{k=1}^n x_k / a_k$ et $S_0 := 0$. On suppose qu'il existe $S \in E$ tel que $S_n \rightarrow S$. Alors pour tout $n \geq 2$, on a

$$\begin{aligned}
 \frac{1}{a_n} \sum_{k=1}^n x_k &= \frac{1}{a_n} \sum_{k=1}^n (S_k - S_{k-1}) a_k \\
 &= \frac{1}{a_n} \sum_{k=1}^{n-1} S_k (a_k - a_{k+1}) + S_n \longrightarrow 0
 \end{aligned}$$

par le lemme de CESÀRO. \square

THÉORÈME 4.27 (convergence presque sûre et dans L^2 de martingale). Soit $X := (X_n)_{n \geq 0}$ une martingale bornée dans L^2 , i. e. $\mathbb{E}[\langle X, X \rangle_\infty] < +\infty$. Alors la suite X converge presque sûrement et dans L^2 vers une certaine variable aléatoire X_∞ . De plus, on a $\mathbb{E}[X_\infty] = \mathbb{E}[X_n]$ pour tout $n \geq 0$.

Preuve Une martingale bornée dans L^2 est bornée dans L^1 , donc elle converge presque sûrement par les résultats de convergence presque sûrement. Montrons qu'elle converge dans L^2 . On pose $Z := \sup_{n \geq 0} |X_n|$. Admettons provisoirement que la variable aléatoire Z est dans L^2 . Alors le théorème de convergence dominée assure $\mathbb{E}[|X_n - X|^2] \rightarrow 0$ ce qui assure la convergence dans L^2 .

Effectivement, d'après l'inégalité maximale des moments, pour tout $n \geq 0$, on a

$$\mathbb{E}[\sup_{k \leq n} |X_k|^2] \leq \left(\frac{2}{2-1} \right)^2 \mathbb{E}[|X_n|^2].$$

Par convergence monotone, on obtient alors

$$\mathbb{E}[Z^2] = \lim_{n \rightarrow +\infty} \mathbb{E}[\sup_{k \leq n} |X_k|^2] \leq 2 \sup_{n \geq 0} \mathbb{E}[|X_n|^2] < +\infty. \quad \square$$

4.6 THÉORÈME D'ARRÊT

PROPOSITION 4.28. Soient $(X_n)_{n \geq 0}$ une sous-martingale uniformément intégrable et T un temps d'arrêt. Alors la martingale X^T est uniformément intégrable.

Preuve La fonction $x \mapsto x^+$ est croissante et convexe, donc la suite $(X_n^+)_{n \geq 0}$ est une sous-martingale. Alors pour tout $n \geq 0$, le théorème d'arrêt borné appliqué à cette dernière martingale et au temps d'arrêt $\min(T, n)$ donne

$$\mathbb{E}[X_{\min(T, n)}^+] \leq \mathbb{E}[X_n^+].$$

4.6. THÉORÈME D'ARRÊT

Comme $0 \leq X_n^+ \leq |X_n|$ pour tout $n \geq 0$, la suite $(X_n)_{n \geq 0}$ est uniformément intégrable, donc $\sup_{n \geq 0} \mathbb{E}[X_n^+] < +\infty$. D'où $\sup_{n \geq 0} \mathbb{E}[X_{\min(T,n)}^+] < +\infty$. Par le critère de convergence presque sûrement des martingales, ceci assure la convergence presque sûrement de la martingale X^T vers X_T et $\mathbb{E}[|X_T|] < +\infty$. Pour tous $n \geq 0$ et $c \in \mathbf{R}_+$, on a

$$\begin{aligned} \mathbb{E}[|X_{\min(T,n)}| \mathbb{1}_{\{|X_{\min(T,n)}| > c\}}] &= \mathbb{E}[|X_T| \mathbb{1}_{\{|X_T| > c\}} \mathbb{1}_{\{T < n\}}] + \mathbb{E}[|X_n| \mathbb{1}_{\{|X_n| > c\}} \mathbb{1}_{\{T \geq n\}}] \\ &\leq \mathbb{E}[|X_T| \mathbb{1}_{\{|X_T| > c\}}] + \mathbb{E}[|X_n| \mathbb{1}_{\{|X_n| > c\}}]. \end{aligned}$$

Comme la suite $(X_n)_{n \geq 0}$ est uniformément intégrable et la variable X_T est intégrable, la suite $(X) \cup (X_n)_{n \geq 0}$ est uniformément intégrable. Ceci implique

$$\lim_{c \rightarrow +\infty} \sup_{n \geq 0} \mathbb{E}[|X_{\min(T,n)}| \mathbb{1}_{\{|X_{\min(T,n)}| > c\}}] = 0$$

ce qui achève la preuve. □

THÉORÈME 4.29. Soient $(X_n)_{n \geq 0}$ une sous-martingale uniformément intégrable et T un temps d'arrêt. Alors

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_T] \leq \mathbb{E}[X_\infty] \quad \text{avec} \quad X_\infty := \lim_{n \rightarrow +\infty} X_n.$$

Preuve On observe que la limite X_∞ existe car la sous-martingale est uniformément intégrable. Pour tout $n \geq 0$, le théorème d'arrêt borné assure

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_{\min(T,n)}] \leq \mathbb{E}[X_n]. \quad (*)$$

Or la proposition précédente affirme que la martingale X^T est uniformément intégrable, donc elle converge presque sûrement et dans L^1 . On obtient $\mathbb{E}[X_{\min(T,n)}] \rightarrow \mathbb{E}[X_T]$ et la convergence presque sûrement et dans L^1 assure $\mathbb{E}[X_n] \rightarrow \mathbb{E}[X_\infty]$. On obtient la conclusion en laissant tendre n vers l'infini dans la relation (*). □

THÉORÈME 4.30 (d'arrêt de DOOB). Soient $(X_n)_{n \geq 0}$ une sous-martingale uniformément intégrable et S et T deux temps d'arrêt tels que $S \leq T$. Alors

$$\mathbb{E}[X_S] \leq \mathbb{E}[X_T] \quad \text{et} \quad X_S \leq \mathbb{E}[X_T | \mathcal{F}_S].$$

Idem pour des sur-martingales et des martingales.

Preuve Soit $Y := X^T$ la sous-martingale arrêtée en T . Elle est uniformément intégrable. Le résultat précédent affirme alors $\mathbb{E}[Y_S] \leq \mathbb{E}[Y_\infty]$. Mais comme $S \leq T$, on a $Y_S = X_S$ et $Y_\infty = X_T$. Ceci donne l'inégalité cherchée.

On observe d'abord que la variable aléatoire X_S est \mathcal{F}_S -mesurable. En effet, pour $B \in \mathcal{B}(\mathbf{R})$, on a $X_S^{-1}(B) \in \mathcal{F}_S$ puisque, pour tout $n \geq 0$, on a $X_S^{-1}(B) \cap \{S = n\} = \{X_n \in B\} \cap \{S = n\} \in \mathcal{F}_n$. Pour $A \in \mathcal{F}_S$, on considère le temps d'arrêt $U := T \mathbb{1}_{A^c} + S \mathbb{1}_A$. Comme $U \leq T$, la première inégalité donne $\mathbb{E}[X_U] \leq \mathbb{E}[X_T]$, c'est-à-dire

$$\mathbb{E}[X_T \mathbb{1}_{A^c}] + \mathbb{E}[X_S \mathbb{1}_A] \leq \mathbb{E}[X_T \mathbb{1}_{A^c}] + \mathbb{E}[X_T \mathbb{1}_A],$$

donc

$$\mathbb{E}[X_S \mathbb{1}_A] \leq \mathbb{E}[X_T \mathbb{1}_A] = \mathbb{E}[\mathbb{E}[X_T | \mathcal{F}_S] \mathbb{1}_A]. \quad (*)$$

Considérons l'événement $A := \{Z - X_S < 0\} \in \mathcal{F}_S$ avec $Z := \mathbb{E}[X_T | \mathcal{F}_S]$. Alors l'inégalité (*) donne $\mathbb{E}[(Z - X_S) \mathbb{1}_A] \geq 0$ ce qui implique $(Z - X_S) \mathbb{1}_A = 0$ presque sûrement. Comme $(Z - X_S) \mathbb{1}_A < 0$, cela exige $\mathbb{P}(A) = 0$. Finalement, on a obtenu $Z \geq X_S$ presque sûrement. □

PROPOSITION 4.31. Soient $(X_n)_{n \geq 0}$ une sur-martingale positive et T un temps d'arrêt. Alors

$$\mathbb{E}[X_T] \leq \mathbb{E}[X_0] + \mathbb{E}[X_\infty].$$

Preuve On sait déjà que la sur-martingale converge presque sûrement vers une variable aléatoire X_∞ . De plus, la suite X^T est une sur-martingale pour laquelle $\mathbb{E}[X_n^T] \leq \mathbb{E}[X_0^T]$ pour tout $n \geq 0$. On a alors

$$\mathbb{E}[X_T \mathbb{1}_{\{T < +\infty\}}] = \lim_{n \rightarrow +\infty} \mathbb{E}[X_{\min(T,n)} \mathbb{1}_{\{T \leq n\}}] \leq \lim_{n \rightarrow +\infty} \mathbb{E}[X_{\min(T,n)}] \leq \mathbb{E}[X_0]$$

et

$$\mathbb{E}[X_T \mathbb{1}_{\{T = +\infty\}}] = \mathbb{E}[X_\infty \mathbb{1}_{\{T = +\infty\}}] \leq \mathbb{E}[X_\infty].$$

En sommant ces deux dernières lignes, on obtient l'inégalité. □

Chapitre 5

DYNAMIQUE MARKOVIENNE

5.1 Probabilité de transition	34	5.4 Chaîne de MARKOV canonique	38
5.2 Exemples	35	5.5 Propriétés de MARKOV	40
5.3 Premières propriétés	35		

INTRODUCTION

On considère un système qui peut prendre un nombre fini ou dénombrable de valeurs. L'ensemble de ces valeurs, appelées *états*, est appelé l'*espace d'états* et noté E . Dans la suite, on supposera que cet ensemble E est dénombrable (typiquement $E \subset \mathbf{Z}$). De plus, on suppose le système observé en des dates entières $n \in \mathbf{N}$ dont on notera X_n son état en cette date. Pour des systèmes non déterministes, les éléments X_n seront des variables aléatoires. Pour étudier le système, c'est-à-dire la suite $(X_n)_{n \in \mathbf{N}}$, il faut faire des hypothèses sur son évolution au cours du temps. Par exemple, on peut supposer que le système évolue des manières complètement indépendantes, *i. e.* les variables aléatoires X_n sont indépendantes. En pratique, une telle hypothèse est trop restrictive pour nombre de phénomènes intéressants. En fait, de nombreux systèmes évoluent avec la propriété suivante : l'état présent du système étant connu, les états passés n'ont aucune influence sur l'état futur du système. Un tel système est dit *sans mémoire* et sont appelés des *chaînes de MARKOV* : seul le présent, et non le passé, influe sur le futur.

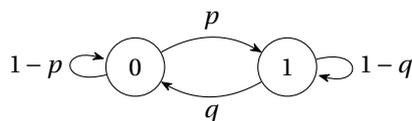
▷ EXEMPLE (*chaîne de MARKOV à deux états*). On considère $E := \{0, 1\}$ et une machine qui, au début de chaque journée, est en fonctionnement (1) ou en panne (0). Pour tout date $n \in \mathbf{N}$, on note $X_n \in E$ son état. Soit $p, q \in [0, 1]$. On suppose que, pour tout $n \in \mathbf{N}$,

- si la machine est en panne au n -ième jour, alors la probabilité qu'elle soit réparée au $(n + 1)$ -ième jour soit p ;
- si la machine fonctionne n -ième jour, alors la probabilité qu'elle soit en panne au $(n + 1)$ -ième jour soit q .

Mathématiquement, pour tout $n \in \mathbf{N}$, cela revient à supposer les relations

$$\mathbb{P}(X_{n+1} = 1 \mid X_n = 0) = p \quad \text{et} \quad \mathbb{P}(X_{n+1} = 0 \mid X_n = 1) = q.$$

On peut également représenté ces hypothèses sous la forme du schéma suivant.



Enfin, on déduit l'état initial X_0 de la machine à l'aide des deux réels $\mu_0(0) := \mathbb{P}(X_0 = 0)$ et $\mu_0(1) := \mathbb{P}(X_0 = 1)$. On souhaite chercher les lois des variables aléatoires X_n en procédant par récurrence. Pour tout $n \in \mathbf{N}$, la formule de probabilité totale donne

$$\begin{aligned} \mathbb{P}(X_{n+1} = 0) &= \mathbb{P}(X_{n+1} = 0 \mid X_n = 0)\mathbb{P}(X_n = 0) + \mathbb{P}(X_{n+1} = 0 \mid X_n = 1)\mathbb{P}(X_n = 1) \\ &= (1 - p)\mathbb{P}(X_n = 0) + q\mathbb{P}(X_n = 1) \\ &= q + (1 - p - q)\mathbb{P}(X_n = 0). \end{aligned}$$

Lorsque $p + q \neq 0$, pour tout $n \in \mathbf{N}$, on a facilement

$$\begin{aligned} \mathbb{P}(X_n = 0) &= \frac{q}{p + q} + (1 - p - q)^n \left(\mu_0(0) - \frac{q}{p + q} \right) \quad \text{et} \\ \mathbb{P}(X_n = 1) &= \frac{p}{p + q} + (1 - p - q)^n \left(\mu_0(0) - \frac{p}{p + q} \right) \end{aligned}$$

et, si $|1 - p - q| < 1$, alors

$$\mathbb{P}(X_n = 0) \longrightarrow \frac{q}{p + q} \quad \text{et} \quad \mathbb{P}(X_n = 1) \longrightarrow \frac{p}{p + q}$$

et la suite $(X_n)_{n \in \mathbf{N}}$ converge en loi vers la loi

$$\mu_\infty := \frac{q}{p + q} \delta_0 + \frac{p}{p + q} \delta_1.$$

Lorsque $p = q = 1$, le système est périodique de période 2 et, pour tout $n \in \mathbf{N}$, on a

$$\begin{aligned} \mathbb{P}(X_{2n} = 0) &= \mu_0(0) & \mathbb{P}(X_{2n+1} = 0) &= \mu_0(1), \\ \mathbb{P}(X_{2n} = 1) &= \mu_0(1) & \mathbb{P}(X_{2n+1} = 1) &= \mu_0(0). \end{aligned}$$

Lorsque $p = q = 0$, le système ne change pas et, pour tout $n \in \mathbf{N}$, on a

$$\mathbb{P}(X_n = 0) = \mu_0(0) \quad \text{et} \quad \mathbb{P}(X_n = 1) = \mu_0(1).$$

• *Approche matricielle.* Pour $n \in \mathbf{N}$, on pose $\mu_n := (\mathbb{P}(X_n = 0), \mathbb{P}(X_n = 1)) \in \mathbf{R}^2$. On pose également

$$P := \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \in \mathcal{M}_2(\mathbf{R}).$$

On observe alors que $\mu_{n+1} = \mu_n P$ pour tout $n \in \mathbf{N}$. De plus, les valeurs propres de P sont 1 et $1 - p - q$ associées à des vecteurs propres $(1, 1)$ et $(p, -q)$. On peut alors réduire P sous la forme

$$P = A \begin{pmatrix} 1 & 0 \\ 0 & 1-p-q \end{pmatrix} A^{-1} \quad \text{avec} \quad A := \begin{pmatrix} 1 & p \\ 1 & -q \end{pmatrix}.$$

Pour tout $n \in \mathbf{N}$, une simple récurrence donne $\mu_n = \mu_0 P^n$ avec

$$P^n = \frac{1}{p+q} \begin{pmatrix} q & p \\ q & p \end{pmatrix} + (1-p-q)^n \begin{pmatrix} p & -p \\ -q & 1 \end{pmatrix}.$$

On retrouve alors les expressions précédentes grâce à cette approche matricielle.

5.1 PROBABILITÉ DE TRANSITION

RAPPEL. Étant données deux espaces mesurables (E, \mathcal{E}) et (F, \mathcal{F}) , on appelle noyau de probabilité (ou de transition) de E dans F toute application $\nu: E \times \mathcal{F} \rightarrow [0, 1]$ telle que

- (i) pour tout $x \in E$, l'application $\nu(x, \cdot)$ est une probabilité sur (F, \mathcal{F}) ;
- (ii) pour tout $A \in \mathcal{F}$, l'application $\nu(\cdot, A)$ est \mathcal{E} -mesurable.

Dans la suite, on considérera $E = F$ où l'ensemble E est dénombrable. Dans ce cas, le point (ii) est toujours satisfait.

DÉFINITION 5.1. On appelle *matrice stochastique* sur E toute famille $P := (P(x, y))_{x, y \in E}$ telle que

- (i) pour tous $x, y \in E$, on a $0 \leq P(x, y) \leq 1$;
- (ii) pour tout $x \in E$, on a $\sum_{y \in E} P(x, y) = 1$.

- ◊ REMARQUES. – La suite $(1, 1, \dots)$ est un « vecteur propre » de P .
- Les matrices stochastiques et les noyaux de probabilités sont liés :
 - si P est une matrice stochastique, alors la relation $\nu(x, A) = \sum_{y \in A} P(x, y)$ définit un noyau de probabilité;
 - si ν est un noyau de probabilité, alors la relation $P(x, y) = \nu(x, \{y\})$ définit une matrice stochastique.

NOTATIONS. Dans la suite, pour tous $x \in E$ et $A \in \mathcal{F}$, on notera

$$P(x, A) := \sum_{y \in A} P(x, y)$$

de telle sorte que l'application $P(x, \cdot)$ soit une probabilité. De plus, pour toute fonction $f: E \rightarrow \mathbf{R}_+$, on pose

$$Pf: \begin{cases} E \rightarrow \mathbf{R}_+, \\ x \mapsto \sum_{y \in E} P(x, y) f(y) = \mathbb{E}_{P(x, \cdot)}[f] \end{cases}$$

Pour tout mesure μ sur P , on définit le mesure sur E

$$\mu^P: \begin{cases} E \rightarrow \mathbf{R}_+, \\ y \mapsto \sum_{x \in E} \mu(x) P(x, y). \end{cases}$$

où on fait les abus de notations $\mu(x)$ et $\mu^P(y)$ pour désigner les réels $\mu(\{x\})$ et $\mu^P(\{y\})$. Enfin, on définit la suite de matrices stochastiques $(P_n)_{n \geq 0}$ par les relations

$$\begin{aligned} P_0(x, y) &= \delta_x(y), \\ P_1(x, y) &= P(x, y), \\ P_{n+1}(x, y) &= \sum_{z \in E} P_n(x, z) P(z, y), \quad n \in \mathbf{N}, x, y \in E. \end{aligned}$$

DÉFINITION 5.2. On appelle *chaîne de MARKOV* sur un ensemble dénombrable E , appelée *espace d'états*, toute suite de variables aléatoires $(X_n)_{n \geq 0}$ à valeurs dans E telle que les lois conditionnelles vérifient

$$\mathcal{L}(X_{n+1} | X_0, \dots, X_n) = \mathcal{L}(X_{n+1} | X_n), \quad n \geq 0,$$

cela revient à avoir

$$\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) \neq 0 \implies \mathbb{P}(X_{n+1} = y | X_0 = x_0, \dots, X_n = x_n) = \mathbb{P}(X_{n+1} = y | X_n = x_n)$$

pour tous $n \geq 0$ et $x_0, \dots, x_n \in E$. De plus, elle est dite *homogène* s'il existe une matrice stochastique P telle que

$$\mathcal{L}(X_{n+1} | X_n) = P(X_n, \cdot), \quad n \geq 0.$$

- ◊ REMARQUE. Pour une chaîne de MARKOV pas forcément homogène, il existe une suite $(P^{(n)})_{n \geq 0}$ de matrices stochastiques telle que

$$\mathcal{L}(X_{n+1} | X_n) = P^{(n)}(X_n, \cdot), \quad n \geq 0.$$

5.2 EXEMPLES

– Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoires indépendantes. Alors cette suite est une chaîne de MARKOV puisque, pour tout $n \geq 0$, l'indépendance donne $\mathcal{L}(X_{n+1} | X_0, \dots, X_n) = \mathcal{L}(X_{n+1}) = \mathcal{L}(X_{n+1} | X_n)$. De plus, si la suite est identiquement distribuée, alors la chaîne est homogène et la relation $P(x, y) = \mathbb{P}(X_0 = y)$ définit la matrice stochastique associée.

– Soient $d \in \mathbf{N}^*$ et $(X_i)_{i \geq 0}$ une suite indépendante de variables aléatoires identiquement distribuées à valeurs dans \mathbf{Z}^d et de loi f . La marche aléatoire $(S_n)_{n \geq 0}$, la suite des sommes partielles, est une chaîne de MARKOV homogène de matrice stochastique donnée par $P(x, y) = f(y - x)$. En effet, pour tous $x, x_0, \dots, x_{n-1}, x_n \in \mathbf{Z}^d$, on a

$$\begin{aligned} \mathbb{P}(S_{n+1} = y | S_0 = x_0, \dots, S_n = x_n) &= \mathbb{P}(x_n + X_{n+1} = y | S_0 = x_0, \dots, S_n = x_n) \\ &= \mathbb{P}(X_{n+1} = y - x_n) = f(y - x_n). \end{aligned}$$

Un exemple en dimension une est la ruine du joueur.

5.3 PREMIÈRES PROPRIÉTÉS

On considère une chaîne de MARKOV homogène $(X_n)_{n \geq 0}$ à valeurs dans un ensemble dénombrable E . On note P son noyau de transition.

PROPOSITION 5.3. Une suite de variables aléatoires $(X_n)_{n \geq 0}$ est une chaîne de MARKOV de matrice P si et seulement si

$$\forall n \geq 0, \forall x_0, \dots, x_n \in E, \quad \mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = \mathbb{P}(X_0 = x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n).$$

En particulier, si $\mathbb{P}(X_0 = x_0) > 0$, alors $\mathbb{P}(X_n = x_n | X_0 = x_0) = P_n(x_0, x_n)$.

Preuve \implies On suppose que la suite $(X_n)_{n \geq 0}$ est une chaîne de MARKOV de matrice P . Montrons l'égalité par récurrence sur l'entier $n \geq 0$. Elle est clairement vraie pour $n = 0$. Soit $n \geq 0$. On suppose qu'elle est vraie au rang n . Soient $x_0, \dots, x_{n+1} \in E$. Si $\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = 0$, alors $\mathbb{P}(X_0 = x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n) = 0$, donc $\mathbb{P}(X_0 = x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n)P(x_n, x_{n+1}) = 0$, donc $\mathbb{P}(X_0 = x_0, \dots, X_{n+1} = x_{n+1}) = 0$ et l'égalité est vraie dans ce cas particulier. On suppose $\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) \neq 0$. Alors

$$\begin{aligned} \mathbb{P}(X_0 = x_0, \dots, X_{n+1} = x_{n+1}) &= \mathbb{P}(X_0 = x_0, \dots, X_n = x_n) \mathbb{P}(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) \\ &= \mathbb{P}(X_0 = x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n) \times P(x_n, x_{n+1}). \end{aligned}$$

Cela montre l'égalité au rang $n + 1$.

\Leftarrow Réciproquement, on suppose l'égalité. Soient $x_0, \dots, x_n \in E$ tels que $\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) \neq 0$. Alors

$$\begin{aligned} \mathbb{P}(X_{n+1} | X_0 = x_0, \dots, X_n = x_n) &= \frac{\mathbb{P}(X_0 = x_0, \dots, X_n = x_n, X_{n+1} = x_{n+1})}{\mathbb{P}(X_0 = x_0, \dots, X_n = x_n)} \\ &= \frac{\mathbb{P}(X_0 = x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n)P(x_n, x_{n+1})}{\mathbb{P}(X_0 = x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n)} = P(x_n, x_{n+1}). \end{aligned}$$

La suite $(X_n)_{n \geq 0}$ est donc une suite de MARKOV. □

COROLLAIRE 5.4. Toutes les lois jointes d'une chaîne de MARKOV sont connues dès qu'on est en possession d'une loi initiale $\mu_0 = \mathcal{L}(X_0)$ et une matrice stochastique P .

TRANSITION EN n ÉTAPES. Soient $n \geq 0$ et $x, y \in E$. La probabilité de transiter en n étapes de x à y est donnée par $P_n(x, y) = \mathbb{P}(X_n = y \mid X_0 = x)$. Autrement dit, c'est la quantité

$$P_n(x, y) = \sum_{y_1 \in E} \cdots \sum_{y_{n-1} \in E} P(x, y_1) \cdots P(y_{n-1}, y).$$

PROPOSITION 5.5 (CHAPMAN-KOLMOGOROV). Le noyau de transition en n étapes satisfait une propriété de semi-groupe : on a $P_{n+m} = P_n P_m$ pour tous $n, m \geq 0$. Ainsi

1. pour tout $n \geq 0$, on a $P_n = P^n$;
2. en notant $\mu_n = \mathcal{L}(X_n)$ pour tout $n \geq 0$, on a $\mu_n = \mu_{n-1} P$ pour tout $n \geq 1$.

Preuve Montrons le second point. Par la formule des probabilité totale puisque, pour tout $y \in E$, on a

$$\begin{aligned} \mu_n(y) &= \sum_{x \in E} \mathbb{P}(X_n = y \mid X_{n-1} = x) \mathbb{P}(X_{n-1} = x) \\ &= (\mu P_{n-1})(y). \end{aligned} \quad \square$$

NOTATION. Pour $x \in E$, on pose $\mathbb{P}_x := \mathbb{P}(\cdot \mid X_0 = x)$. De même, pour une loi ν , on note \mathbb{P}_ν pour signifier qu'on suppose $X_0 \sim \nu$.

PROPOSITION 5.6. Soient $(X_n)_{n \geq 0}$ une chaîne de MARKOV, $f: E \rightarrow \mathbf{R}$ une fonction mesurable et $n \geq 0$. Alors

$$\mathbb{E}[f(X_{n+1}) \mid X_0, \dots, X_n] = (Pf)(X_n).$$

De même, pour tous $0 \leq i_1 < \dots < i_p \leq n$, on a

$$\mathbb{E}[f(X_{n+1}) \mid X_{i_1}, \dots, X_{i_p}, X_n] = (Pf)(X_n).$$

Preuve La variable aléatoire $\mathbb{E}[f(X_{n+1}) \mid X_0, \dots, X_n]$ est l'espérance de f sous la loi $\mathcal{L}(X_{n+1} \mid X_n) = \mathbb{P}(X_n, \cdot)$, donc elle est égale à

$$\mathbb{E}[f(X_{n+1}) \mid X_n] = \sum_{y \in E} f(y) P(x_n, y) = (Pf)(X_n).$$

Maintenant, pour tous $0 \leq i_1 < \dots < i_p \leq n$, on a

$$\begin{aligned} \mathbb{E}[f(X_{n+1}) \mid X_{i_1}, \dots, X_{i_p}, X_n] &= \mathbb{E}[\mathbb{E}[f(X_{n+1}) \mid X_0, \dots, X_n] \mid X_{i_1}, \dots, X_{i_p}, X_n] \\ &= \mathbb{E}[(Pf)(X_n) \mid X_{i_1}, \dots, X_{i_p}, X_n] = (Pf)(X_n). \end{aligned} \quad \square$$

PROPOSITION 5.7 (boîte à outils). Soit $(X_n)_{n \geq 0}$ une chaîne de MARKOV de matrice de transition P . En supposant que les probabilités conditionnelles sont bien définies, on a les points suivants :

1. pour tous $x_0, \dots, x_n, y_1, \dots, y_m \in E$, on a

$$\mathbb{P}(X_{n+1} = y_1, \dots, X_{n+m} = y_m \mid X_0 = x_0, \dots, X_n = x_n) = P(x_n, y_1) \cdots P(y_{m-1}, y_m) ;$$

2. pour tous $A_0, \dots, A_{n-1} \subset E$ et $x_n, y_1, \dots, y_m \in E$, on a

$$\mathbb{P}(X_{n+1} = y_1, \dots, X_{n+m} = y_m \mid X_0 \in A_0, \dots, X_{n-1} \in A_{n-1}, X_n = x_n) = P(x_n, y_1) \cdots P(y_{n-1}, y_n)^{\S 1} ;$$

3. pour tous $A_0, \dots, A_{n-1}, B_1, \dots, B_m \subset E$ et $x_n \in E$, on a

$$\mathbb{P}(X_{n+1} \in B, \dots, X_{n+m} \in B_m \mid X_0 \in A_0, \dots, X_{n-1} \in A_{n-1}, X_n = x_n) = \sum_{y \in B_1} \cdots \sum_{y_m \in B_m} P(x_n, y_1) \cdots P(y_{m-1}, y_m).$$

Preuve 1. On a

$$\begin{aligned} \mathbb{P}(X_{n+1} = y_1, \dots, X_{n+m} = y_m \mid X_0 = x_0, \dots, X_n = x_n) &= \frac{\mathbb{P}(X_{n+1} = y_1, \dots, X_{n+m} = y_m, X_0 = x_0, \dots, X_n = x_n)}{\mathbb{P}(X_0 = x_0, \dots, X_n = x_n)} \\ &= \frac{\mathbb{P}(X_0 = x_0) P(x_0, x_1) \cdots P(x_{n-1}, x_n) P(x_n, y_1) \cdots P(y_{n-1}, y_n)}{\mathbb{P}(X_0 = x_0) P(x_0, x_1) \cdots P(x_{n-1}, x_n)} \\ &= P(x_n, y_1) \cdots P(y_{n-1}, y_n). \end{aligned}$$

\S 1. La dynamique markovienne est toujours à l'heure, j'oublie le passé.

2. Pour $C \in \mathcal{F}$, on note $\mathbb{P}_C := \mathbb{P}(\cdot | C)$ et alors $\mathbb{P}_C(A | B) = \mathbb{P}(A | B \cap C)$ pour tous $A, B \in \mathcal{F}$. En effet, on a

$$\mathbb{P}_C(A | B) = \frac{\mathbb{P}_C(A \cap B)}{\mathbb{P}_C(B)} = \frac{\mathbb{P}(A \cap B | C)}{\mathbb{P}(B | C)} = \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(C)} \frac{\mathbb{P}(C)}{\mathbb{P}(B \cap C)}.$$

Maintenant, si $B_1 \cap B_2 = \emptyset$ et $\mathbb{P}(B_1 \cup B_2) > 0$, alors

$$\begin{aligned} \mathbb{P}(A | B_1 \sqcup B_2) &= \mathbb{P}(A | B_1)\mathbb{P}(B_1 | B_1 \sqcup B_2) + \mathbb{P}(A | B_2)\mathbb{P}(B_2 | B_1 \sqcup B_2) \\ &= \frac{\mathbb{P}(A \cap B_1)}{\mathbb{P}(B_1 \sqcup B_2)} + \frac{\mathbb{P}(A \cap B_2)}{\mathbb{P}(B_1 \sqcup B_2)}. \end{aligned}$$

En plus, si $\mathbb{P}(A | B_1) = \mathbb{P}(A | B_2)$, alors $\mathbb{P}(A | B_1 \sqcup B_2) = \mathbb{P}(A | B_1)$.

Prouvons l'égalité. On remarque que

$$\{X_0 \in A_0, \dots, X_{n-1} \in A_{n-1}, X_n = x_n\} = \bigsqcup_{x_1 \in A_1} \dots \bigsqcup_{x_{n-1} \in A_{n-1}} \{X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x_n\}.$$

On applique ensuite la remarque et le point 1 avec cette partition.

3. Le troisième égalité découle du point 2 puisque

$$\begin{aligned} &\mathbb{P}(X_{n+1} \in B, \dots, X_{n+m} \in B_m | X_0 \in A_0, \dots, X_{n-1} \in A_{n-1}, X_n = x_n) \\ &= \sum_{y_1 \in B_1} \dots \sum_{y_m \in B_m} \mathbb{P}(X_{n+1} = y_1, \dots, X_{n+m} = y_m | X_0 \in A_0, \dots, X_{n-1} \in A_{n-1}, X_n = x_n) \\ &= \sum_{y_1 \in B_1} \dots \sum_{y_m \in B_m} P(x, y_1) \dots P(y_{m-1}, y_m). \end{aligned} \quad \square$$

Approche récursive

PROPOSITION 5.8. Soient X_0 un variable aléatoire à valeurs dans E suivant une loi ν et $(U_n)_{n \geq 0}$ une suite indépendante de variables aléatoires à valeurs dans F suivant un loi ν et indépendantes de X_0 . Soit $f: E \times F \rightarrow E$ une fonction mesurable. Alors la suite $(X_n)_{n \geq 0}$ vérifiant

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

est une chaîne de MARKOV de loi initiale ν et de noyau de transition donnée par

$$P(x, y) = \mathbb{P}(f(x, U_0) = y).$$

Preuve Soient $x_0, \dots, x_{n+1} \in E$ tels que $\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) \neq 0$. Alors

$$\begin{aligned} \mathbb{P}(X_{n+1} = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) &= \mathbb{P}(f(x_n, U_n) = x_{n+1} | X_0 = x_0, \dots, X_n = x_n) \\ &= \mathbb{P}(f(x_n, U_n) = x_{n+1}) \end{aligned}$$

car les variables $f(x_n, U_n)$ et X_0, \dots, X_n sont respectivement $\sigma(U_n)$ -mesurable et $\sigma(X_0, U_0, \dots, U_{n-1})$ -mesurable. Montrons ce dernier point par récurrence. En effet, la variable $X_n = f(X_{n-1}, U_{n-1})$ est $\sigma(X_{n-1}, U_{n-1})$ -mesurable. Donc pour tout $n \geq 1$, si X_{n-1} est $\sigma(X_0, U_0, \dots, U_{n-2})$ -mesurable, alors X_{n-1} est $\sigma(X_0, U_0, \dots, U_{n-1})$ -mesurable. Cela conclut. \square

PROPOSITION 5.9. Une chaîne de MARKOV homogène à valeurs réelles peut être vue en loi comme une suite récurrente du type (+) de la proposition précédentes.

Preuve Soit $(X_n)_{n \geq 0}$ une chaîne de MARKOV homogène de matrice P . Alors il s'agit de trouver une fonction mesurable $f: E \times \mathbf{R} \rightarrow E$ et une variable aléatoire U_1 tel que $X_1 = f(x, U_1)$ lorsque $X_0 = x$ pour $x \in E$. Soit $x \in E$. On sait que la loi de X_1 sachant $X_0 = x$ est $P(x, \cdot)$. On note $f(x, \cdot)$ l'inverse généralisée de la fonction de répartition de $P(x, \cdot)$, c'est-à-dire

$$f(x, y) := \inf\{u \in \mathbf{R} | P(x,]-\infty, u]) > y\}, \quad y \in \mathbf{R}.$$

Par la méthode d'inversion, on obtient $f(x, U_1) \sim \mathcal{L}(X_1 | X_0 = x)$ pour $U_1 \sim \mathcal{U}([0, 1])$. En considérant une suite indépendante de variable aléatoires $(U_n)_{n \geq 0}$ suivant chacune une loi uniforme sur $[0, 1]$ et indépendantes de X_0 , on définit la suite $(\tilde{X}_n)_{n \geq 0}$ telle que $\tilde{X}_{n+1} := f(\tilde{X}_n, U_{n+1})$ pour tout $n \geq 0$. Alors pour tous $n \geq 0$ et $x_n, x_{n+1} \in E$, la construction ainsi faite donne

$$\mathbb{P}(\tilde{X}_{n+1} = x_{n+1} | \tilde{X}_n = x_n) = \mathbb{P}(f(x_n, U_{n+1}) = x_{n+1}) = P(x_n, x_{n+1}).$$

Ainsi les chaînes $(X_n)_{n \geq 0}$ et $(\tilde{X}_n)_{n \geq 0}$ suivent la même dynamique. \square

▷ EXEMPLES. – On reprend l'exemple d'une marche aléatoire sur \mathbf{Z}^d . Soit $(X_n)_{n \geq 0}$ une suite indépendante de variables aléatoires identiquement distribuées. Alors la suite des sommes partielles $(S_n)_{n \geq 0}$ est une chaîne de MARKOV qui vérifie $S_{n+1} = f(S_n, X_{n+1})$ pour tout $n \geq 0$ avec $f(x, y) := x + y$.

– Soient $(X_{n,k})_{n \geq 1, k \geq 0}$ une suite indépendante de variables aléatoires identiquement distribuées. On considère la suite $(Z_n)_{n \geq 1}$ définie par $Z_{n+1} = \sum_{i=0}^{Z_n} X_{n,i}$ pour tout $n \geq 0$. Alors pour tout $n \geq 0$, on a

$$Z_{n+1} = f(Z_n, (X_{n,i})_{i \geq 0}) \quad \text{avec} \quad f(z, (x_i)_{i \geq 0}) := \sum_{i=0}^z x_i.$$

DÉFINITION 5.10. Soient $(X_n)_{n \geq 0}$ une chaîne de MARKOV et $A \subset E$. Le *temps d'atteinte* de A par la chaîne de MARKOV $(X_n)_{n \geq 0}$ est le temps d'arrêt

$$T_A := \inf(n \geq 0 \mid X_n \in A).$$

Pour tout $y \in E$, on notera $T_y := T_{\{y\}}$ et la variable $\tilde{T}_y := \inf(n > 0 \mid X_n = y)$ coïncide avec T_y lorsque $X_0 \neq y$ et diffère de T_y lorsque $X_0 = y$: c'est le temps de retour en y

PROPOSITION 5.11. Soient $(X_n)_{n \geq 0}$ une chaîne de MARKOV de matrice P , $n \geq 1$ et $x, y \in E$. Alors

$$P^n(x, y) = \sum_{m=1}^n \mathbb{P}_x(T_y = m) P^{n-m}(y, y).$$

De plus, pour tout élément absorbant $a \in E$, c'est-à-dire $P(a, a) = 1$, on a $P^n(x, a) = \mathbb{P}(T_a \leq n)$.

Preuve On peut écrire la partition

$$\{X_n = y\} = \bigsqcup_{m=1}^n \{T_y = m, X_n = y\}$$

et on obtient

$$\begin{aligned} P^n(x, y) &= \mathbb{P}_x(X_n = y) = \sum_{m=1}^n \mathbb{P}_x(T_y = m, X_n = y) \\ &= \sum_{m=1}^n \mathbb{P}_x(T_y = m) \mathbb{P}(X_n = y \mid X_0 \neq y, X_{m-1} \neq y, X_m = y) \\ &= \sum_{m=1}^n \mathbb{P}_x(T_y = m) P^{n-m}(y, y). \quad \square \end{aligned}$$

Maintenant soit $a \in E$ un élément absorbant. Alors pour tous $m, n \geq 1$ tels que $m \leq n$, on a $P^{n-m}(a, a) = 1$ et l'équation précédente devient

$$P^n(x, a) = \sum_{m=1}^n \mathbb{P}_x(T_a = m) P^{n-m}(a, a) = \sum_{m=1}^n \mathbb{P}_x(T_a = m) = \mathbb{P}_x(T_a \leq n).$$

5.4 CHAÎNE DE MARKOV CANONIQUE

PROPOSITION 5.12 (construction d'une chaîne de MARKOV). Soient E un ensemble au plus dénombrable et P une matrice stochastique. Alors il existe un espace probabilisé $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ sur lequel, pour tout $x \in E$, il existe une chaîne de MARKOV $(X_n^x)_{n \geq 0}$ de matrice de transition P et issue de $X_0^x = x$.

Preuve On considère l'espace probabilisé $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}) := ([0, 1], \mathcal{B}([0, 1]), \lambda)$. Commençons par montrer que cet espace supporte une suite de variables aléatoires uniformes indépendantes et identiquement distribuées. Pour cela, rappelons que tout réel $\omega \in [0, 1[$ s'écrit sous la forme

$$\omega = \sum_{n=0}^{+\infty} \varepsilon_n(\omega) 2^{-n-1}.$$

De plus, si ω suit une loi uniforme sur $[0, 1[$, alors les variables ε_n sont indépendantes et suivent une loi de BERNOULLI de paramètre $1/2$ (et c'est même équivalent).

Soit $\varphi: \mathbf{N} \times \mathbf{N} \rightarrow \mathbf{N}$ une bijection. Alors les variables aléatoires $\eta_{i,j} := \varepsilon_{\varphi(i,j)}$ sont toujours indépendantes. On observe alors que les variables aléatoires $U_i := \sum_{j=0}^{+\infty} \eta_{i,j} 2^{-(j+1)}$ sont indépendantes et suivent une loi uniforme sur $[0, 1[$. Soient $(y_n)_{n \geq 1}$ une énumération de E et $x \in E$. On considère alors la suite $(X_n^x)_{n \geq 0}$ définie par les relations $X_0^x = x$ et

$$X_{n+1}^x = y_k \quad \text{si} \quad \sum_{j=1}^{k-1} P(X_n^x, y_j) < U_{n+1} \leq \sum_{j=1}^k P(X_n^x, y_j), \quad n, k \geq 1.$$

Par construction, pour tous $n \geq 1$ et $y, z \in E$, on a $\tilde{\mathbb{P}}(X_n^x = y \mid X_{n-1}^x = z) = P(z, y)$. En effet, pour tout $k \geq 1$, on a

$$\begin{aligned} \tilde{\mathbb{P}}(X_1^x = y_k) &= \mathbb{P}\left(\sum_{j=1}^{k-1} P(x, y_j) < U_{n+1} \leq \sum_{j=1}^k P(x, y_j)\right) \\ &= \sum_{j=1}^k P(x, y_j) - \sum_{j=1}^{k-1} P(x, y_j) = P(x, y_k). \end{aligned}$$

Puis pour tous $n, k \geq 1$ et $x_1, \dots, x_n \in E$, on a

$$\begin{aligned} \tilde{\mathbb{P}}(X_{n+1}^x = y_k \mid X_0^x = x, X_1^x = x_1, \dots, X_n^x = x_n) &= \tilde{\mathbb{P}}\left(\sum_{j=1}^{k-1} P(X_n^x, y_j) < U_{n+1} \leq \sum_{j=1}^k P(X_n^x, y_j) \mid X_0^x = x, X_1^x = x_1, \dots, X_n^x = x_n\right) \\ &= \tilde{\mathbb{P}}\left(\sum_{j=1}^{k-1} P(X_n^x, y_j) < U_{n+1} \leq \sum_{j=1}^k P(X_n^x, y_j)\right) = P(x_n, y_k) \end{aligned}$$

en utilisant $\{X_0^x = x, X_1^x = x_1, \dots, X_n^x = x_n\} \in \sigma(U_1, \dots, U_n) \perp U_{n+1}$. Ceci montre que la suite $(X_n^x)_{n \geq 0}$ est bien une chaîne de MARKOV satisfaisant les conditions de la proposition. \square

- ◊ REMARQUE. Le choix de l'espace probabilisé est arbitraire. L'espace probabilisé canonique est $\tilde{\Omega} := E^{\mathbb{N}}$ associé à la plus petite tribu \mathcal{C} rendant les applications coordonnées X_n mesurables, i. e. c'est la tribu engendré par les cylindres

$$\{(\omega_k)_{k \geq 0} \in E^{\mathbb{N}} \mid \omega_0 = x_0, \dots, \omega_n = x_n\}, \quad n \geq 0, x_0, \dots, x_n \in E.$$

On note $\text{Cyl}(E, \mathbb{N})$ la famille des cylindres.

La suite s'applique sur un tel espace probabilisé $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$.

LEMME 5.13. Soit $\psi: (\tilde{\Omega}, \tilde{\mathcal{F}}) \rightarrow (\Omega, \mathcal{C})$ une fonction. Alors ψ est mesurable si et seulement si $X_n \circ \psi$ est mesurable pour tout $n \geq 0$

Preuve La classe de parties $\mathcal{G} := \{A \in \mathcal{C} \mid \psi^{-1}(A) \in \tilde{\mathcal{F}}\}$ est une tribu contenant toutes les préimages $X_n^{-1}(\{y\})$ avec $y \in E$ et $n \geq 1$ puisque $X_n^{-1}(\{y\}) \in \mathcal{C}$ et $\psi^{-1}(X_n^{-1}(\{y\})) = (X_n \circ \psi)^{-1}(\{y\}) \in \tilde{\mathcal{F}}$ par hypothèse. Ainsi la tribu \mathcal{G} rend mesurables les applications coordonnées X_n , donc $\mathcal{C} \subset \mathcal{G}$ et finalement $\mathcal{G} = \mathcal{C}$ ce qui assure que la fonction ψ est $(\tilde{\mathcal{F}}, \mathcal{C})$ -mesurable. \square

THÉORÈME 5.14 (chaîne canonique). Soient E un ensemble au plus dénombrable et P une matrice stochastique. Alors pour tout $x \in E$, il existe une unique probabilité \mathbb{P}_x sur $(E^{\mathbb{N}}, \mathcal{C})$ sous laquelle la suite $(X_n)_{n \geq 0}$ est une chaîne de MARKOV de matrice P avec $\mathbb{P}_x(X_0 = x) = 1$.

Preuve • *Existence.* Soit $x \in E$. D'après la proposition 5.12, il existe un espace probabilisé $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ et une chaîne de MARKOV $(X_n^x)_{n \geq 0}$ de matrice P telle que $X_0^x = x$. On considère alors la mesure image \mathbb{P}_x de $\tilde{\mathbb{P}}$ par l'application

$$\psi: \begin{cases} \tilde{\Omega} \longrightarrow E^{\mathbb{N}}, \\ \tilde{\omega} \longmapsto (X_n^x(\tilde{\omega}))_{n \geq 0}. \end{cases}$$

Par le lemme précédente, cette application ψ est bien mesurable car les fonctions $X_n \circ \psi = X_n^x$ sont des variables aléatoires. Par définition de la mesure image, on a $\mathbb{P}_x(X_0 = x) = \tilde{\mathbb{P}}(X_0^x = x) = 1$ et, pour tous $n \geq 0$ et $x_0, \dots, x_n \in E$, on a

$$\begin{aligned} \mathbb{P}_x(X_0 = x_0, \dots, X_n = x_n) &= \tilde{\mathbb{P}}(X_0^x = x_0, X_1^x = x_1, \dots, X_n^x = x_n) \\ &= \tilde{\mathbb{P}}(X_0^x = x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n) \\ &= \mathbb{P}(X_0 = x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n). \end{aligned} \quad (*)$$

D'après la proposition 5.3, on en déduit que, sous \mathbb{P}_x , la suite $(X_n)_{n \geq 0}$ est une chaîne de MARKOV de matrice P . Par construction, elle est issue de x .

• *Unicité.* Soit \mathbb{P}'_x une autre telle probabilité. L'égalité (*) assure que les probabilités \mathbb{P}_x et \mathbb{P}'_x coïncident sur les cylindres qui forment un π -système. D'après le lemme des classes monotones, on en déduit qu'elles coïncident sur \mathcal{C} . \square

- ◊ REMARQUES. – Pour tous $n \geq 0$ et $x, y \in E$, on a $\mathbb{P}_x(X_n = y) = P_n(x, y)$.
– Soit ν une probabilité sur E . Alors l'application $\mathbb{P}_\nu := \sum_{x \in E} \nu(x) \mathbb{P}_x$ est une probabilité sur $(E^{\mathbb{N}}, \mathcal{C})$. Sous celle-ci, la suite $(X_n)_{n \geq 0}$ est une chaîne de MARKOV de matrice P et de loi initiale ν . Pour $x \in E$, on note \mathbb{E}_x et \mathbb{E}_ν les espérances sous \mathbb{P}_x et \mathbb{P}_ν .

– Pour toute chaîne de MARKOV X de matrice P et de loi initiale ν , on a $\mathbb{P}(X \in B) = \mathbb{P}_\nu(B)$ pour tout $B \in \mathcal{F}$. Les résultats en loi obtenus pour la chaîne canonique se transposent donc à toute chaîne de même matrice P et de même loi initiale ν .

PROPOSITION 5.15 (loi d'une chaîne de MARKOV). La loi d'une chaîne de MARKOV homogène de matrice P et de loi initiale ν est l'unique probabilité \mathbb{P}_ν sur $(E^\mathbf{N}, \mathcal{C})$ telle que

$$\mathbb{P}_\nu(\{(x_k)_{k \in \mathbf{N}} \in E^\mathbf{N} \mid x_0 = y_0, \dots, x_n = y_n\}) = \nu(y_0)P(y_0, y_1) \cdots P(y_{n-1}, y_n), \quad n \in \mathbf{N}, y_0, \dots, y_n \in E.$$

Preuve Il suffit d'utiliser le théorème d'extension de KOLMOGOROV à partir de ces relations. \square

THÉORÈME 5.16 (d'extension de KOLMOGOROV). Soit $\mathcal{Q} := \{Q_t\}_{t \in T^p, p \in \mathbf{N}^*}$ une famille de lois fini-dimensionnelles vérifiant les conditions de comptabilités :

- pour tous $p \in \mathbf{N}^*$, $(t_1, \dots, t_p) \in T^p$, $\sigma \in S_p$ et $A_1, \dots, A_p \in \mathcal{B}(\mathbf{R})$, on a $Q_t(A_1 \times \cdots \times A_p) = Q_s(A_{\sigma(1)} \times \cdots \times A_{\sigma(p)})$;
- pour tous $p \in \mathbf{N}^*$, $t \in T^p$, $s \in T^{p-1}$ et $A \in \mathcal{B}(\mathbf{R}^{p-1})$, on a $Q_t(A \times \mathbf{R}) = Q_s(A)$.

Alors il existe une probabilité P sur $(\mathbf{R}^T, \sigma(\text{Cyl}(\mathbf{R}, T)))$ qui admet \mathbf{Q} pour famille de lois fini-dimensionnelles.

5.5 PROPRIÉTÉS DE MARKOV

En travaillant sur l'espace canonique $(E^\mathbf{N}, \mathbf{C})$, on peut introduire les opérateurs de décalage (ou de *shift*) : ce sont les applications

$$\Theta_k : \begin{cases} E^\mathbf{N} \longrightarrow E^\mathbf{N}, \\ (\omega_n)_{n \geq 0} \longmapsto (\omega_{k+n})_{n \geq 0}. \end{cases}$$

Le lemme 5.13 assure que ces applications Θ_k sont mesurables. De plus, on note $(\mathcal{F}_n)_{n \geq 0}$ la filtration associée à la suite $(X_n)_{n \geq 0}$.

THÉORÈME 5.17 (de MARKOV faible). Soient G une fonction mesurable sur Ω , $x \in E$ et $n \geq 0$. Alors

$$\mathbb{E}_x[G \circ \Theta_n \mid \mathcal{F}_n] = \mathbb{E}_{X_n}[G].$$

De manière équivalente, pour toute fonction \mathcal{F}_n -mesurable F , on a

$$\mathbb{E}_x[F \times (G \circ \Theta_n)] = \mathbb{E}_x[F \mathbb{E}_{X_n}[G]].$$

Ces deux identités sont encore valides lorsqu'on remplace \mathbb{E}_x par \mathbb{E}_μ .

- ◇ REMARQUE. Comme $\mathbb{E}_{X_n}[G]$ est la composée de X_n et $x \longmapsto \mathbb{E}_x[G]$, c'est une variable aléatoire $\sigma(X_n)$ -mesurable et donc \mathcal{F}_n -mesurable.

Preuve Il suffit de traiter le cas où les fonctions F et G sont des indicatrices de cylindres. Le cas général s'obtient ensuite par un argument de classe monotone.

Soient alors $x_0, \dots, x_n, y_0, \dots, y_p \in E$. On suppose $Q = \mathbb{1}_{\{X_0=x_0, \dots, X_n=x_n\}}$ et $F = \mathbb{1}_{\{X_0=y_0, \dots, X_p=y_p\}}$. Pour tout $y \in E$, on a alors

$$\mathbb{E}_y[G] = \delta_{y_0, y} P(y_0, y_1) \cdots P(y_{p-1}, y_p).$$

Comme $G \circ \Theta_n = \mathbb{1}_{\{X_n=x_0, \dots, X_{n+p}=x_n\}}$, on a

$$\begin{aligned} \mathbb{E}_x[F \times (G \circ \Theta_n)] &= \mathbb{P}_x(X_0 = x_0, \dots, X_n = x_n, X_n = y_0, \dots, X_{n+p} = y_p) \\ &= \delta_{x_0, x} P(x_0, x_n) \cdots P(x_{n-1}, x_n) \delta_{x_n, y_0} P(y_0, y_1) \cdots P(y_{p-1}, y_p). \end{aligned}$$

Par ailleurs, on a

$$\begin{aligned} \mathbb{E}_x[F \mathbb{E}_{X_n}[G]] &= \mathbb{E}_x[F \delta_{y_0, X_n} P(y_0, y_1) \cdots P(y_{p-1}, y_p)] \\ &= \mathbb{E}_x[\mathbb{1}_{\{X_0=y_0, \dots, X_p=y_p\}} \delta_{y_0, X_n} P(y_0, y_1) \cdots P(y_{p-1}, y_p)] \\ &= \delta_{x_0, x} P(x_0, x_n) \cdots P(x_{n-1}, x_n) \delta_{x_n, y_0} P(y_0, y_1) \cdots P(y_{p-1}, y_p) = \mathbb{E}_x[F \times (G \circ \Theta_n)]. \end{aligned} \quad \square$$

THÉORÈME 5.18 (de MARKOV fort). Soient T un temps d'arrêt, G une fonction mesurable sur Ω et $x \in E$. Alors

$$\mathbb{E}_x[\mathbb{1}_{\{T < +\infty\}} G \circ \Theta_T \mid \mathcal{F}_T] = \mathbb{1}_{\{T < +\infty\}} \mathbb{E}_{X_T}[G].$$

De manière équivalente, pour toute fonction \mathcal{F}_T -mesurable F , on a

$$\mathbb{E}_x[\mathbb{1}_{\{T < +\infty\}} F \times (G \circ \Theta_T)] = \mathbb{E}_x[\mathbb{1}_{\{T < +\infty\}} F \mathbb{E}_{X_T}[G]].$$

◊ REMARQUE. Avec des indicatrices, les propriétés de MARKOV faible et fort s'écrivent, pour tout $A \in \mathcal{C}$, tout temps d'arrêt T et tous $x_0, \dots, x_n, y \in E$, sous la forme

$$\begin{aligned}\mathbb{P}(\Theta^n X \in A \mid X_0 = x_0, \dots, X_n = x_n) &= \mathbb{P}_{x_n}(X \in A), \\ \mathbb{P}(\Theta^T X \in A \mid X_0 = x_0, \dots, X_T = y) &= \mathbb{P}_y(X \in A).\end{aligned}$$

Preuve Soit F une fonction \mathcal{F}_T -mesurable. Alors la fonction $\mathbb{1}_{\{T=n\}}F$ est \mathcal{F}_n -mesurable puisque, pour $A \in \mathcal{B}(\mathbf{R})$, comme $\{F \in A\} \cap \{T = n\} \in \mathcal{F}_n$, on a

$$\{\mathbb{1}_{\{T=n\}}F \in A\} = (\{F \in A\} \cap \{T = n\}) \cup (\{0 \in A\} \cap \{T = n\}) \in \mathcal{F}_n.$$

Dès lors, la propriété de MARKOV faible donne

$$\begin{aligned}\mathbb{E}_x[\mathbb{1}_{\{T=n\}}F \times (G \circ \Theta_T)] &= \mathbb{E}_x[\mathbb{1}_{\{T=n\}}F \times (G \circ \Theta_n)] \\ &= \mathbb{E}_x[\mathbb{1}_{\{T=n\}}F \mathbb{E}_{X_n}[G]] \\ &= \mathbb{E}_x[\mathbb{1}_{\{T=n\}}F \mathbb{E}_{X_T}[G]].\end{aligned}$$

En sommant sur les entiers $n \geq 0$, on obtient la conclusion. \square

COROLLAIRE 5.19. Soient $x \in E$ et T un temps d'arrêt tel que $\mathbb{P}_x(T < +\infty) = 1$. On suppose qu'il existe $y \in E$ tel que $\mathbb{P}_x(X_T = y) = 1$. Alors sous \mathbb{P}_x , la variable Θ_T est indépendante de \mathcal{F}_T et sa loi est \mathbb{P}_y .

Preuve Pour toutes fonctions \mathcal{F}_T -mesurables F et G , la propriété de MARKOV forte donne

$$\mathbb{E}_x[F \times (G \circ \Theta_T)] = \mathbb{E}_x[F \mathbb{E}_{X_T}[G]] = \mathbb{E}_x[F \mathbb{E}_y[G]] = \mathbb{E}_x[F] \mathbb{E}_y[G].$$

En particulier, pour $F = 1$, on obtient $\mathbb{E}_x[G \circ \Theta_T] = \mathbb{E}_y[G]$ (*) si bien que la relation précédente se réécrit

$$\mathbb{E}_x[F \times (G \circ \Theta_T)] = \mathbb{E}_x[F] \mathbb{E}_x[G \circ \Theta_T].$$

Ceci étant vrai pour toute fonction \mathcal{F}_T -mesurables F et G , on obtient l'indépendance de Θ_T avec \mathcal{F}_T . De plus, l'équation (*) implique $\mathbb{P}_x \circ \Theta_T^{-1} = \mathbb{P}_y$, c'est-à-dire que la variable Θ_T est de loi \mathbb{P}_y sur $E^{\mathbf{N}}$. \square

◊ REMARQUE. Les propriétés de MARKOV restent vraies si on remplace \mathbb{E}_x par \mathbb{E}_ν pour toute loi initiale ν . En effet, comme $\mathbb{E}_\nu = \sum_{x \in E} \nu(x) \mathbb{E}_x$, il suffit de sommer convenablement les diverses égalités.

REFORMULATION DE LA PROPRIÉTÉ DE MARKOV. Pour toute loi initiale ν , en notant $\mathcal{L}_\nu((X_n)_{n \geq 0})$ la loi de la chaîne $(X_n)_{n \geq 0}$ lorsque $X_0 \sim \nu$, la propriété de MARKOV forte s'écrit

$$\mathcal{L}_\nu((X_n)_{n \geq T} \mid \mathcal{F}_T) = \mathcal{L}_{X_T}((X_n)_{n \geq 0})$$

pour tout temps d'arrêt T . La propriété de MARKOV faible s'obtient lorsque le temps d'arrêt est déterministe.

COROLLAIRE 5.20. Soient $n \geq 0$, $A \in \mathcal{F}_n$ et $B \in \sigma(X_k \mid k \geq n)$. Alors

$$\mathbb{P}(A \cap B \mid \mathcal{F}_n) = \mathbb{P}(A \mid \mathcal{F}_n) \mathbb{P}(B \mid \mathcal{F}_n).$$

De plus, pour tout temps d'arrêt T qui est \mathbb{P}_x -presque sûrement fini, tous $A \in \mathcal{F}_T$ et $B \in \sigma(X_k \mid k \geq T)$, on a

$$\mathbb{P}(A \cap B \mid \mathcal{F}_T) = \mathbb{P}(A \mid \mathcal{F}_T) \mathbb{P}(B \mid \mathcal{F}_T).$$

Preuve Montrons uniquement la première partie, la seconde partie se montre comme dans la preuve du théorème 5.18. Soient F une fonction \mathcal{F}_n -mesurable et G une fonction $\sigma(X_k \mid k \geq n)$ -mesurable. Cette dernière peut s'écrire sous la forme $G = \tilde{G} \circ \Theta_n$. Soit $x \in E$. Par la proposition des conditionnements en cascade et le théorème de MARKOV faible, on a

$$\begin{aligned}\mathbb{E}_x[FG \mid X_n] &= \mathbb{E}_x[F(\tilde{G} \circ \Theta_n) \mid X_n] = \mathbb{E}_x[\mathbb{E}_x[F(\tilde{G} \circ \Theta_n) \mid \mathcal{F}_n] \mid X_n] \\ &= \mathbb{E}_x[F \mathbb{E}[\tilde{G} \circ \Theta_n \mid \mathcal{F}_n] \mid X_n] = \mathbb{E}_x[F \mathbb{E}_{X_n}[\tilde{G}] \mid X_n] \\ &= \mathbb{E}_x[F \mid X_n] \mathbb{E}_{X_n}[\tilde{G}].\end{aligned}$$

En particulier, pour $F = 1$, on obtient $\mathbb{E}_{X_n}[\tilde{G}] = \mathbb{E}_x[G \mid X_n]$. En réinjectant dans les calculs précédent, cela donne

$$\mathbb{E}_x[FG \mid X_n] = \mathbb{E}_x[F \mid X_n] \mathbb{E}_x[G \mid X_n]$$

ce qui montre la première égalité en passant à l'espérance. \square

Application au problème de DIRICHLET discret

Soit F un sous-ensemble de E , P un noyau de transition sur E et h une fonction bornée sur F . On note I la matrice identité $E^2 \rightarrow E$. On s'intéresse au problème de DIRICHLET discret

$$\begin{cases} (I - P)g = 0 & \text{sur } E \setminus F, \\ g = h & \text{sur } F. \end{cases}$$

Trouvons une solution g . Pour cela, on considère une chaîne de MARKOV $(X_n)_{n \geq 0}$ sur E de noyau de transition P . Considérons le temps d'arrêt

$$T_F = T_F(X_n)_{n \geq 0} := \inf\{n \geq 0 \mid X_n \in F\}.$$

Alors la fonction g définie par la relation

$$g(x) := \mathbb{E}_x[h(X_{T_F})\mathbb{1}_{\{T_F < +\infty\}}], \quad x \in E$$

est solution au problème.

Preuve Écrivons de quels termes de la suite $(X_n)_{n \geq 0}$ dépendent les variables aléatoires qu'on considère. Soit alors $x \in F$. Alors \mathbb{P}_x -presque sûrement, on a $T_F(X_0, \dots) = 0$ et $X_{T_F} = x$. Dans ce cas, on a

$$g(x) = \mathbb{E}_x[\mathbb{1}_{\{T_F < +\infty\}}] = \mathbb{E}_x[h(x)] = h(x).$$

Soit maintenant $x \in E \setminus F$. Alors \mathbb{P}_x -presque sûrement, on a $T_F(X_0, \dots) \geq 1$ et $T_F(X_0, \dots) = 1 + T_F(X_1, \dots)$. En utilisant la propriété de MARKOV forte en (*), on obtient

$$\begin{aligned} g(x) &= \mathbb{E}_x[h(X_{T_F})\mathbb{1}_{\{T_F < +\infty\}}] \\ &= \mathbb{E}_x[h(X_{1+T_F(X_1, \dots)})\mathbb{1}_{\{1+T_F(X_1, \dots) < +\infty\}}] \\ &= \mathbb{E}_x[h(X_{1+T_F(X_1, \dots)})\mathbb{1}_{\{T_F(X_1, \dots) < +\infty\}}] \\ &= \mathbb{E}_x[\mathbb{E}_x[h(X_{1+T_F(X_1, \dots)})\mathbb{1}_{\{T_F(X_1, \dots) < +\infty\}} \mid \mathcal{F}_1]] \\ &= \mathbb{E}_x[\mathbb{E}_{X_1}[h(X_{T_F(X_0, \dots)})\mathbb{1}_{\{T_F(X_0, \dots) < +\infty\}}]] \tag{*} \\ &= \mathbb{E}_x[g(X_1)] = \sum_{y \in E} g(y)\mathbb{P}_x(X_1 = y) = \sum_{y \in E} P(x, y)g(y) = Pg(x). \end{aligned}$$

Ceci montre que $(I - P)g(x) = 0$ pour tout $x \in E \setminus F$. □

Chapitre 6

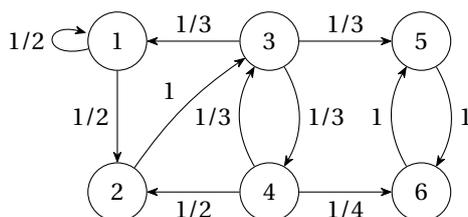
RÉCURRENCE ET TRANSIENNE

6.1 États récurrents et transitoires	44	6.3 Classes de récurrence	50
6.2 Ensembles clos et irréductibilité	48	6.4 Absorption dans les classes de récurrence	52

INTRODUCTION

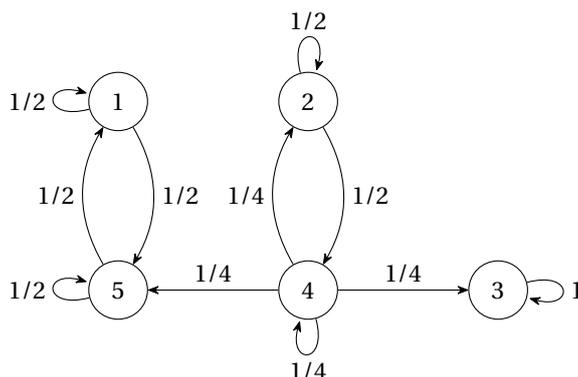
Exemples

Sur l'espace $E := [1, 6]$, on considère une chaîne de MARKOV donnée par le graphe de transition suivant.



Les états de 1 à 4 semblent être visités un nombre fini de fois \mathbb{P}_1 -presque sûrement. En revanche, les états 5 et 6 semblent être visités une infinité de fois \mathbb{P}_1 -presque sûrement.

Maintenant, considérons la chaîne suivant sur l'espace $E := [1, 5]$.



Cette fois-ci, les états 2 et 4 semblent être visités un nombre fini de fois \mathbb{P}_2 -presque sûrement, alors que les états 1/5 et 3 sont visités une infinité de fois \mathbb{P}_2 -presque sûrement et ceux-ci ne communiquent pas.

L'objet de cette section est de comprendre le comportement qualitatif d'une chaîne de MARKOV comme dans les exemples ci-dessus. Il s'agit d'un comportement qualitatif car les assertions précédentes ne semblent pas dépendre des probabilités de transition mais seulement de leur non-nullité.

On verra ensuite ce qu'on peut donner comme information quantitative sur la chaîne, par exemple la proportion de temps passé en un état.

Notations

Dans tout ce chapitre, on considérera une chaîne de MARKOV $(X_n)_{n \geq 0}$ sur un espace d'état E et de noyau de transition P . Sauf mention contraire, on travaillera avec la chaîne canonique construite dans le théorème 5.14. Dans la suite, pour $x \in E$, on note \mathbb{E}_x l'espérance sous \mathbb{P}_x , c'est-à-dire en supposant que la chaîne parte de x . Pour tout $y \in E$, on rappelle que les temps d'arrêt

$$T_y := \inf\{n \geq 0 \mid X_n = y\} \quad \text{et} \quad \tilde{T}_y := \inf\{n > 0 \mid X_n = y\}$$

sont respectivement les temps d'atteinte de y et d'atteinte de y après le départ. Lorsqu'on a prend deux éléments distincts $x, y \in E$, sous \mathbb{P}_x , le temps d'arrêt \tilde{T}_y est le temps d'atteinte T_y de y alors que, sous \mathbb{P}_y , il s'agit du temps de retour en y ; on notera

$$\rho_{x,y} := \mathbb{P}_x(\tilde{T}_y < +\infty)$$

la probabilité que, partant de l'état x , la chaîne puisse arriver en temps fini à l'état y . La quantité $\rho_{x,x}$ est alors la probabilité que la chaîne partant de x finisse par revenir.

6.1 ÉTATS RÉCURRENTS ET TRANSITOIRES

DEFINITION 6.1. Un état $x \in E$ est dit *récurrent* (respectivement *transitoire*) si $\rho_{x,x} = 1$ (respectivement $\rho_{x,x} < 1$).

REMARQUE. En particulier, un état absorbant est récurrent.

Nombre de passage

Pour un état transitoire x , une chaîne partant de x a une probabilité non nulle de ne jamais y revenir alors que, si l'état est récurrent, elle y reviendra une fois et donc, par récurrence et avec la propriété de Markov forte, une infinité de fois. On formalise cette intuition dans la proposition suivante. Elle éclaire sur la terminologie en montrant que le nombre de passage en un état x , noté

$$N(x) := \sum_{k=0}^{+\infty} \mathbb{1}_{\{X_k=x\}},$$

dépend fondamentalement de sa nature récurrente ou transitoire.

PROPOSITION 6.2 (nombre de passages en un état). Soit $x \in E$. Alors on a l'alternative suivante.

1. Si x est récurrent, alors $N(x) = +\infty$ \mathbb{P}_x -presque sûrement.
2. Si x est transitoire, alors $N(x)$ suit une loi géométrique de paramètre $1 - \rho_{x,x}$ sous \mathbb{P}_x . En particulier, la variable aléatoire $N(x)$ est \mathbb{P}_x -presque sûrement finie et

$$\mathbb{E}_x[N(x)] = \frac{1}{1 - \rho_{x,x}}.$$

Preuve D'abord, remarquons que $\mathbb{P}_x(N(x) \geq 1) = 1$. Puis pour tout $k \geq 1$, la propriété de MARKOV forte assure

$$\begin{aligned} \mathbb{P}_x(N(x) \geq k+1) &= \mathbb{E}_x[\mathbb{1}_{\{N(x) \geq k+1\}}] = \mathbb{E}_x[\mathbb{1}_{\{\tilde{T}_x < +\infty\}} \mathbb{1}_{\{N(x) \geq k\}} \circ \Theta_{\tilde{T}_x}] \\ &= \mathbb{P}_x(\tilde{T}_x < +\infty) \mathbb{E}[\mathbb{1}_{\{N(x) \geq k\}}] \\ &= \rho_{x,x} \mathbb{P}_x(N(x) \geq k). \end{aligned}$$

Par une récurrence immédiate, on obtient

$$\mathbb{P}_x(N(x) \geq k) = \rho_{x,x}^{k-1}, \quad k \geq 1.$$

Dès lors, on peut distinguer les cas.

1. On suppose que l'état x est récurrent. Alors $\rho_{x,x} = 1$ et, en laissant tendre k vers l'infini dans la relation précédente, le théorème de convergence monotone assure $\mathbb{P}_x(N(x) = +\infty) = 1$.
2. On suppose que l'état x est transitoire. Alors $\rho_{x,x} < 1$ et, pour tout $k \geq 1$, on a

$$\begin{aligned} \mathbb{P}_x(N(x) = k) &= \mathbb{P}_x(N(x) \geq k) - \mathbb{P}_x(N(x) \geq k+1) \\ &= \rho_{x,x}^{k-1} (1 - \rho_{x,x}) \end{aligned}$$

ce qui montre le point 2. □

Soit $y \in E$. Notons $\tilde{N}(y) := \sum_{k=1}^{+\infty} \mathbb{1}_{\{X_k=y\}}$ le nombre de passage de la chaîne dans l'état y après le départ. Si elle part d'un état $x \in E \setminus \{y\}$, alors $\tilde{N}(y) = N(y)$ et $\tilde{T}(y) = T(y)$. Mais si elle part de l'état y , alors $\tilde{N}(y) = N(y) - 1$ et le temps d'arrêt \tilde{T}_y est le temps de premier retour en y . Définissons la suite de temps d'arrêt $(T_y^{(k)})_{k \geq 0}$ par les relations

$$T_y^{(k)} = \min\{n > T_y^{(k-1)} \mid X_n = y\}$$

avec $T_y^{(0)} := 0$.

PROPOSITION 6.3. Soient $y \in E$ et $n \geq 0$. Sur $T_y^{(n)} < +\infty$, les variables aléatoires $\Delta_y^{(k)} := T_y^{(k)} - T_y^{(k-1)}$ pour $k \in \llbracket 1, n \rrbracket$ sont indépendantes et identiquement distribuées sous \mathbb{P}_y .

Preuve Comme $\{T_y^{(n)} < +\infty\} = \bigcap_{i=1}^n \{\Delta_y^{(i)} < +\infty\}$, il suffit de montrer que

$$\mathbb{E}_y \left[\prod_{i=1}^k g(\Delta_y^{(i)}) \mathbb{1}_{\{\Delta_y^{(i)} < +\infty\}} \right] = \prod_{i=1}^k \mathbb{E}_y [g(\Delta_y^{(i)}) \mathbb{1}_{\{\Delta_y^{(i)} < +\infty\}}]$$

pour des fonctions g_i mesurables et bornées sur \mathbf{R}_+ . Procédons par récurrence sur l'entier $k \geq 1$. Le cas $k = 1$ est trivial. Soit $k \geq 1$. On suppose que l'égalité au vrai au rang k . Observons d'abord que

- les variables $\Delta_y^{(1)}, \dots, \Delta_y^{(k-1)}$ sont $\mathcal{F}_{T_y^{(k-1)}}$ -mesurables;
- la variable $\Theta_{T_y^{(k-1)}}$ est indépendante de la tribu $\mathcal{F}_{T_y^{(k-1)}}$ et de loi \mathbb{P}_y d'après la propriété de MARKOV forte;
- on a $\Delta_y^{(k)} = \Delta_y^{(1)} \circ \Theta_{T_y^{(k-1)}}$.

Soient g_1, \dots, g_k des fonctions mesurables et bornées sur \mathbf{R}_+ . La propriété de MARKOV forte donne

$$\begin{aligned} \mathbb{E}_y \left[\prod_{i=1}^k g(\Delta_y^{(i)}) \mathbb{1}_{\{\Delta_y^{(i)} < +\infty\}} \right] &= \mathbb{E}_y \left[\prod_{i=1}^{k-1} \left(g(\Delta_y^{(i)}) \mathbb{1}_{\{\Delta_y^{(i)} < +\infty\}} \right) g_k(\Delta_y^{(1)} \circ \Theta_{T_y^{(k-1)}}) \mathbb{1}_{\{\Delta_y^{(1)} \circ \Theta_{T_y^{(k-1)}} < +\infty\}} \right] \\ &= \mathbb{E}_y \left[\prod_{i=1}^{k-1} g(\Delta_y^{(i)}) \mathbb{1}_{\{\Delta_y^{(i)} < +\infty\}} \right] \mathbb{E}_y [g_k(\Delta_y^{(1)} \circ \Theta_{T_y^{(k-1)}}) \mathbb{1}_{\{\Delta_y^{(1)} \circ \Theta_{T_y^{(k-1)}} < +\infty\}}] \\ &= \mathbb{E}_y \left[\prod_{i=1}^{k-1} g(\Delta_y^{(i)}) \mathbb{1}_{\{\Delta_y^{(i)} < +\infty\}} \right] \mathbb{E}_y [g_k(\Delta_y^{(k)}) \mathbb{1}_{\{\Delta_y^{(k)} < +\infty\}}] \\ &= \prod_{i=1}^{k-1} \mathbb{E}_y [g_k(\Delta_y^{(i)}) \mathbb{1}_{\{\Delta_y^{(i)} < +\infty\}}] \mathbb{E}_y [g_k(\Delta_y^{(k)}) \mathbb{1}_{\{\Delta_y^{(k)} < +\infty\}}] \end{aligned}$$

ce qui termine l'hérédité. Cela montre alors la proposition. \square

PROPOSITION 6.4 (nombre de passages en y partant de $x \neq y$). Soient $x, y \in E$ deux états distincts. Sous \mathbb{P}_x ,

1. si y est transitoire, alors

$$N(y) \stackrel{\mathbb{P}_x}{\sim} (1 - \rho_{x,y})\delta_0 + \rho_{x,y}\mathcal{G}(1 - \rho_{y,y}) \quad \text{et} \quad \mathbb{P}_x(N(y) < +\infty) = 1;$$

2. si y est récurrent, alors partant de x , soit la chaîne ne rejoint pas y ($N(y) = 0$), soit elle le rejoint une fois puis alors une infinité de fois ($N(y) = +\infty$). De plus, on a

$$N(y) \stackrel{\mathbb{P}_x}{\sim} (1 - \rho_{x,y})\delta_0 + \rho_{x,y}\delta_{+\infty}, \quad \mathbb{P}_y(N(y) = +\infty) = 1 \quad \text{et} \quad \mathbb{P}_x(N(y) = +\infty) = \rho_{x,y}.$$

Preuve Supposons que la chaîne part de x . On a $\mathbb{P}_x(\tilde{N}(y) \geq 1) = \mathbb{P}_x(\tilde{T}_y < +\infty) = \rho_{x,y}$. Soient $m, n \geq 0$. La probabilité que la chaîne partant de x visite y la première fois à la date m et n'y revienne qu'à la date $m + n$ est

$$\begin{aligned} \mathbb{P}_x(T_y^{(1)} = m, T_y^{(2)} = m + n) &= \mathbb{E}_x [\mathbb{1}_{\{T_y^{(1)} = m\}} \mathbb{1}_{\{T_y^{(2)} = m + n\}}] \\ &= \mathbb{E}_x [\mathbb{1}_{\{T_y^{(1)} = m\}} \mathbb{1}_{\{T_y^{(1)} \circ \Theta_{T_y^{(1)}} = m + n\}}] \\ &= \mathbb{E}_x [\mathbb{1}_{\{T_y^{(1)} = m\}}] \mathbb{E}_{X_{T_y^{(1)}}} [\mathbb{1}_{\{T_y^{(1)} = n\}}] \\ &= \mathbb{P}_x(\tilde{T}_y = m) \mathbb{P}_y(\tilde{T}_y = n) \end{aligned}$$

en utilisant la propriété de MARKOV forte et puisque $X_{T_y^{(1)}} = y$. On a donc

$$\begin{aligned} \mathbb{P}_x(N(y) \geq 2) &= \sum_{m=1}^{+\infty} \sum_{n=1}^{+\infty} \mathbb{P}_x(\text{la chaîne visite } y \text{ pour la 1}^{\text{re}} \text{ fois à la date } m \text{ et n'y revient qu'à la date } m + n) \\ &= \sum_{m=1}^{+\infty} \sum_{n=1}^{+\infty} \mathbb{P}_x(T_y^{(1)} = m, T_y^{(2)} = m + n) \\ &= \sum_{m=1}^{+\infty} \sum_{n=1}^{+\infty} \mathbb{P}_x(\tilde{T}_y = m) \mathbb{P}_y(\tilde{T}_y = n) = \mathbb{P}_x(\tilde{T}_y < +\infty) \mathbb{P}_y(\tilde{T}_y < +\infty) = \rho_{x,y} \rho_{y,y}. \end{aligned}$$

Plus généralement, on montre de la même façon que

$$\mathbb{P}_x(N(y) \geq m) = \rho_{x,y} \rho_{y,y}^{m-1}, \quad m \geq 1 \quad (*)$$

puis

$$\mathbb{P}_x(N(y) = m) = \rho_{x,y} \rho_{y,y}^{m-1} (1 - \rho_{y,y}), \quad m \geq 1 \quad (**)$$

et

$$\mathbb{P}_x(N(y) = 0) = 1 - \mathbb{P}_x(N(y) \geq 1) = 1 - \rho_{x,y}. \quad (***)$$

Dans le cas récurrence, on déduit de la relation (*) que $\mathbb{P}_x(N(y) = +\infty) = \rho_{x,y}$ par convergence monotone. \square

- ◇ REMARQUES. – Dans le cas $x = y$, les relations (**) et (***) sont remplacées par $\mathbb{P}_x(N(x) = k) = \rho_{x,x}^{k-1}(1 - \rho_{x,x})$. La différence vient du fait que, sous \mathbb{P}_x , on a $N(x) \geq 1$ puisque la chaîne part de x . Il y a donc un décalage dans le compte des passages en x dû au point de départ.
 - En fait, ces relations (**) et (***) sont intuitivement claires avec la description heuristique suivante : pour que, partant de x , la chaîne visite m fois y , elle commence par aller de x à y (facteur $\rho_{x,y}$) puis visite $y - 1$ fois (facteur $\rho_{y,y}^{m-1}$) et n’y retourne plus (facteur $1 - \rho_{y,y}$).
 - Les propositions 6.2 et 6.4 décrivent la différence fondamentale entre un état transitoire et un état récurrent.
 - Si l’état y est transitoire, alors quelque soit l’état initial de la chaîne, alors il y aura un nombre fini de passages en y et le nombre moyen de passages est fini aussi.
 - Si l’état y est récurrent alors quand la chaîne part de cet état, alors elle y repasse une infinité de fois. Si elle part d’ailleurs, alors soit elle n’y va jamais, soit elle y va une fois et alors elle y retourne nécessairement une infinité de fois.

Potentiel ou fonction de GREEN

DÉFINITION 6.5. Pour $x, y \in E$, on note

$$G(x, y) := \mathbb{E}_x[N(y)] = \sum_{k=0}^{+\infty} \mathbb{P}_x(X_k = y) = \sum_{k=0}^{+\infty} P_k(x, y)$$

le nombre moyen de passage en y partant de x .

- ◇ REMARQUE. Les égalités se montrent par le théorème de convergence monotone.

THÉORÈME 6.6 (récurrence, transience et potentiel). 1. Si un état y est transitoire, alors le potentiel $G(x, y)$ est fini pour tout état x et

$$G(x, y) = \begin{cases} \rho_{x,y}/(1 - \rho_{y,y}) & \text{si } x \neq y, \\ 1/(1 - \rho_{y,y}) & \text{sinon.} \end{cases}$$

2. Si un état y est récurrent, alors $G(y, y) = +\infty$ et

$$G(x, y) = \begin{cases} 0 & \text{si } \rho_{x,y} = 0, \\ +\infty & \text{sinon.} \end{cases}$$

Preuve On utilise les propositions précédentes 6.2 et 6.4 selon les cas. Soient $x, y \in E$. On suppose $x \neq y$.

1. Si y est transitoire, alors $G(x, y) = \mathbb{E}_x[N(y)]$ est l’espérance de la loi $(1 - \rho_{x,y})\delta_0 + \rho_{x,y}\mathcal{G}(1 - \rho_{y,y})$, égale à $\rho_{x,y}/(1 - \rho_{y,y})$.

2. Si y est récurrent, alors $G(x, y) = \mathbb{E}_x[N(y)]$ est l’espérance de la loi $(1 - \rho_{x,y})\delta_0 + \rho_{x,y}\delta_{+\infty}$, égale à

$$\begin{cases} 0 & \text{si } \rho_{x,y} = 0, \\ +\infty & \text{sinon.} \end{cases}$$

On procède de même si $x = y$. □

De l’alternative de ce théorème, on déduit immédiatement un critère de récurrence à l’aide du potentiel.

COROLLAIRE 6.7 (récurrence et potentiel). Un état x est récurrent si et seulement si $G(x, x) = +\infty$.

COROLLAIRE 6.8. En convenant que $0 \times (+\infty) = 0$, pour tous états $x, y \in E$ distincts, on a

$$G(x, y) = \rho_{x,y}G(y, y).$$

Preuve Sous \mathbb{P}_x , on a $N(y) = 0$ lorsque $T_y = +\infty$, donc

$$N(y) = \mathbb{1}_{\{T_y < +\infty\}}N(y) = \mathbb{1}_{\{T_y < +\infty\}}N(y) \circ \Theta_{T_y}.$$

Par la propriété de MARKOV, on obtient alors

$$\begin{aligned} G(x, y) &= \mathbb{E}_x[N(y)] = \mathbb{E}_x[\mathbb{1}_{\{T_y < +\infty\}}(N(y) \circ \Theta_{T_y})] \\ &= \mathbb{E}_x[\mathbb{1}_{\{T_y < +\infty\}}\mathbb{E}_y[N(y)]] = \mathbb{P}_x(T_y < +\infty)\mathbb{E}_y[N(y)] = \rho_{x,y}G(y, y). \end{aligned} \quad \square$$

- ◇ REMARQUE. Notons que, pour tout état transitoire $y \in E$ et tout état $x \in E$, on a $P^n(x, y) \rightarrow 0$. En effet, cela vient de la convergence de la série $\sum_{n \geq 0} P^n(x, y)$.

DÉFINITION 6.9. Une chaîne est dite transitoire (respectivement récurrente) si tous ses états sont transitoires (respectivement récurrents).

- ◊ **REMARQUE.** Si une chaîne a un nombre fini d'états, alors nécessairement il existe au moins un état récurrent et la chaîne ne peut être transitoire. En effet, si tous les états étaient transitoires, alors la remarque précédente indique

$$0 = \sum_{y \in E} \lim_{n \rightarrow +\infty} P^n(x, y) = \lim_{n \rightarrow +\infty} \sum_{y \in E} P^n(x, y) = \lim_{n \rightarrow +\infty} \mathbb{P}_x(X_n \in E) = 1.$$

pour tout $x \in E$ car la somme est finie ce qui est impossible.

DÉFINITION 6.10. Un état récurrente $x \in E$ est dit *positif* (respectivement *nul*) si $m_x := \mathbb{E}_x[\tilde{T}_x] < +\infty$ (respectivement $m_x = +\infty$).

Temps passé sur un état

Pour $n \geq 1$ et $x, y \in E$, on note respectivement

$$N_n(y) := \sum_{k=0}^n \mathbb{1}_{\{X_k=y\}} \quad \text{et} \quad G_n(x, y) := \mathbb{E}_x[N_n(y)] = \sum_{k=0}^n P^k(x, y)$$

la variable aléatoire qui indique le temps passé en l'état y jusqu'au temps n et le temps moyen passé en y jusqu'au temps n lorsqu'on part de x .

THÉORÈME 6.11 (proportion de visite). Soient $y \in E$ et ν une distribution initiale. Alors

$$\frac{N_n(y)}{n} \xrightarrow{-\mathbb{P}_\nu\text{-ps}} \frac{\mathbb{1}_{\{\tilde{T}_y < +\infty\}}}{m_y}.$$

De plus, pour tout $x \in E$, on a

$$\frac{G_n(x, y)}{n} \xrightarrow{-\mathbb{P}_x\text{-ps}} \frac{\rho_{x,y}}{m_y}.$$

- ◊ **REMARQUE.** Ces résultats se justifient par l'heuristique suivante : dès que la chaîne atteint un état récurrent y , elle y revient en moyenne en m_y étapes. Donc si $\tilde{T}_y < +\infty$, pour un grand entier $n \geq 0$, la proportion d'étapes parmi les n premières où la chaîne est en y est d'ordre $1/m_y$.

Preuve On commence par considérer le cas où l'état y est récurrent et où la chaîne démarre de ce point. Avec probabilité 1, la chaîne revient en y une infinité de fois. Pour $k \geq 1$, on pose

$$T_y^{(k)} := \min\{n \geq 1 \mid N_n(y) = k\}$$

la date de sa k -ième visite en y et on pose $T_y^{(0)} := 0$. Rappelons que les variables aléatoires $\Delta_y^{(k)} := T_y^{(k)} - T_y^{(k-1)}$ sont indépendantes et identiquement distribuées. Comme $T_y^{(k)} = \Delta_y^{(1)} + \dots + \Delta_y^{(k)}$, montrons que

$$\frac{T_y^{(n)}}{n} \xrightarrow{-\mathbb{P}_y\text{-ps}} m_y. \tag{*}$$

Si $m_y < +\infty$, alors $\Delta_y^{(1)} \in L^1$ et la relation (*) s'obtient avec la loi des grands nombre. On suppose $m_y = +\infty$. Pour tout $a > 0$, la loi des grands nombres donne

$$\frac{\min(\Delta_y^{(1)}, a) + \dots + \min(\Delta_y^{(n)}, a)}{n} \xrightarrow{-\mathbb{P}_y\text{-ps}} \mathbb{E}_y[\min(\Delta_y^{(1)}, a)].$$

Mais comme $\Delta_y^{(1)} \geq \min(\Delta_y^{(1)}, a)$, on a \mathbb{P}_y -presque sûrement

$$\liminf_{n \rightarrow +\infty} \frac{\Delta_y^{(1)} + \dots + \Delta_y^{(n)}}{n} \geq \mathbb{E}_y[\min(\Delta_y^{(1)}, a)].$$

Le théorème de converge monotone assure alors

$$\frac{\Delta_y^{(1)} + \dots + \Delta_y^{(n)}}{n} \xrightarrow{-\mathbb{P}_y\text{-ps}} +\infty$$

ce qui montre la limite (*) dans le cas $m_y = +\infty$.

Maintenant, par définition des variables aléatoires $T_y^{(k)}$ et $N_n(y)$, pour $n \geq 0$, on a $T_y^{(N_n(y))} \leq n < T_y^{(N_n(y)+1)}$. Pour $n \geq 0$ assez grand, afin d'assurer $N_n(y) \geq 1$, on a

$$\frac{T_y^{(N_n(y))}}{N_n(y)} \leq \frac{n}{N_n(y)} < \frac{T_y^{(N_n(y)+1)}}{N_n(y)}.$$

Comme \mathbb{P}_y -presque sûrement $N_n(y) \rightarrow +\infty$, la loi des grands nombres assure

$$\frac{n}{N_n(y)} \xrightarrow{-\mathbb{P}_y\text{-ps}} m_y$$

ce qui montre la première égalité du théorème.

Toujours dans le cas récurrent, supposons maintenant que la distribution initiale ν est quelconque. Dans ce cas, la chaîne ne peut jamais rejoindre y . Cependant, si elle rejoint y , l'argument précédent s'applique et se réécrit

$$\frac{N_n(y)}{n} \xrightarrow{-\mathbb{P}_\nu\text{-ps}} \frac{\mathbb{1}_{\{\tilde{T}_y < +\infty\}}}{m_y},$$

i. e. c'est la première égalité.

De plus, pour tout $n \geq 1$, comme $0 \leq N_n(y) \leq n$, on a $0 \leq N_n(y)/n \leq 1$ et le théorème de convergence dominée la deuxième égalité car, \mathbb{P}_ν -presque sûrement, on a

$$\lim_{n \rightarrow +\infty} \frac{G_n(x, y)}{n} = \lim_{n \rightarrow +\infty} \mathbb{E}_x \left[\frac{N_n(y)}{n} \right] = \mathbb{E}_x \left[\frac{\mathbb{1}_{\{\tilde{T}_y < +\infty\}}}{m_y} \right] = \frac{\mathbb{P}_x(\tilde{T}_y < +\infty)}{m_y} = \frac{\rho_{x,y}}{m_y}.$$

Achevons le preuve dans le cas où l'état y est transitoire et où la chaîne démarre de ce point. Soit $x \in E$. La proposition 6.4 assure que, \mathbb{P}_x -presque sûrement, on a $N_n(y) \rightarrow N(y) < +\infty$ et $G_n(x, y) \rightarrow G(x, y) < +\infty$. On en déduit

$$\frac{N_n(y)}{n} \xrightarrow{-\mathbb{P}_x\text{-ps}} 0 \quad \text{et} \quad \frac{G_n(x, y)}{n} \xrightarrow{-\mathbb{P}_y\text{-ps}} 0$$

puisque $m_y = +\infty$. Le résultat reste vrai \mathbb{P}_ν -presque sûrement. \square

6.2 ENSEMBLES CLOS ET IRRÉDUCTIBILITÉ

On note E_R et E_T l'ensemble des états récurrents et transitoires. Ces deux parties forment une partition de E , partition qu'on va préciser dans la suite.

DÉFINITION 6.12. On dit qu'un état $x \in E$ peut mener à un état $y \in E$ si $\rho_{x,y} > 0$. Dans ce cas, on note $x \rightsquigarrow y$;

PROPOSITION 6.13. Soient $x, y \in E$ deux états. Alors les propositions suivantes sont équivalentes :

- (i) $x \rightsquigarrow y$;
- (ii) $G(x, y) > 0$;
- (iii) il existe $n \geq 1$ tel que $P^n(x, y) > 0$.

Preuve On suppose que la chaîne part de l'état $x \neq y$. Montrons les équivalences cycliquement. Supposons (i). Comme $\{N(y) \geq 1\} = \{\tilde{T}_y < +\infty\}$, on a $\mathbb{P}_x(N(y) \geq 1) = \rho_{x,y} > 0$, donc $G(x, y) = \mathbb{E}_x[N(y)] > 0$. Supposons (ii). Comme $G(x, y) = \sum_{n=0}^{+\infty} P^n(x, y) > 0$, il existe bien un entier $n \geq 1$ tel que $P^n(x, y) > 0$. Supposons (iii). S'il existe un entier $n \geq 1$ tel que $P^n(x, y) > 0$. Alors $\rho_{x,y} = \mathbb{P}_x(N(y) \geq 1) > 0$. \square

PROPOSITION 6.14. La relation \rightsquigarrow est transitive.

Preuve On peut utiliser la relation de CHAPMAN-KOLMOGOROV avec le point (iii) de la proposition précédente. Un autre manière est de remarquer que, par la propriété de MARKOV forte, on a

$$\begin{aligned} \rho_{x,y} &= \mathbb{P}_x(T_z < +\infty) \\ &\geq \mathbb{P}_x(T_y < +\infty, T_z \circ \Theta_{T_y} < +\infty) \\ &= \mathbb{E}_x[\mathbb{1}_{\{T_y < +\infty\}} \mathbb{1}_{\{T_z \circ \Theta_{T_y} < +\infty\}}] \\ &= \mathbb{E}_x[\mathbb{1}_{\{T_y < +\infty\}}] \mathbb{E}_y[\mathbb{1}_{\{T_z < +\infty\}}] = \rho_{x,y} \rho_{y,z}. \end{aligned} \quad \square$$

THÉORÈME 6.15 (un état récurrent mène à une état récurrent). Soient $x \in E_R$ et $y \in E$ tels que $x \rightsquigarrow y$. Alors $y \in E_R$ et $\rho_{y,x} = 1$. En particulier, on a $y \rightsquigarrow x$ et $\rho_{x,y} = 1$.

CONSÉQUENCE. Un état récurrent ne mène jamais à un état transitoire.

Preuve L'énoncé est immédiat si $x = y$. On suppose donc $x \neq y$. Commençons par montrer $\mathbb{P}_y(T_x < +\infty) = 1$. Comme x est récurrente, on a

$$\begin{aligned} 0 = \mathbb{P}_x(N(x) < +\infty) &\geq \mathbb{P}_x(T_y < +\infty, T_x \circ \Theta_{T_y} = +\infty) \\ &= \mathbb{E}_x[\mathbb{1}_{\{T_y < +\infty\}}(\mathbb{1}_{\{T_x = +\infty\}} \circ \Theta_{T_y})] \\ &= \mathbb{E}_x[\mathbb{1}_{\{T_y < +\infty\}}] \mathbb{E}_y[\mathbb{1}_{\{T_x = +\infty\}}] \\ &= \mathbb{P}_x(T_y < +\infty) \mathbb{P}_y(T_x = +\infty). \end{aligned}$$

Comme $x \rightsquigarrow y$, on a $\rho_{x,y} > 0$. Cela exige $\mathbb{P}_y(T_x = +\infty) = 1$, donc $\mathbb{P}_y(T_x < +\infty) = 1$. D'où $G(y, x) > 0$ et $y \rightsquigarrow x$.

Montrons maintenant que $y \in E_R$. Comme $G(x, y) > 0$ et $G(y, x) > 0$, on peut trouver des entiers $n_1, n_2 \geq 0$ tels que $P^{n_1}(x, y) > 0$ et $P^{n_2}(y, x) > 0$. Pour tout $k \geq 0$, on a alors

$$P^{n_1+k+n_2}(y, y) \geq P^{n_2}(y, x) P^k(x, x) P^{n_1}(x, y).$$

En sommant, on obtient alors

$$G(y, y) \geq \sum_{k=0}^{+\infty} P^{n_1+k+n_2}(y, y) = P^{n_1}(y, x) G(x, x) P^{n_1}(x, y) = +\infty.$$

D'où $y \in E_R$. □

THÉORÈME 6.16. Soient $x \in E$ un état récurrent positif et $y \in E$. Si $x \rightsquigarrow y$, alors l'état y est récurrent positive.

DÉFINITION 6.17. Un ensemble d'états $C \subset E$ est *clos* si aucun état de C ne peut mener à un état en dehors de C , *i. e.*

$$\forall x \in C, \forall y \notin C, \rho_{x,y} = 0.$$

◇ REMARQUE. Le singleton constitué d'un état absorbant est un ensemble clos.

PROPOSITION 6.18. Soit $C \subset E$ tel que $P(x, y) = 0$ pour tous $x \in C$ et $y \notin C$. Alors l'ensemble C est clos.

Preuve On montre que $P^n(x, y) = 0$ pour tous $x \in C$, $y \notin C$ et $n \geq 1$ en utilisant la relation de CHAPMANN-KOLMOGOROV. □

DÉFINITION 6.19. Un ensemble d'états clos $C \subset E$ est dit *irréductible* si, pour tous $x, y \in C$, l'état x peut mener à l'état y . Un chaîne est dite irréductible si son espace d'états l'est.

◇ REMARQUE. La définition est équivalente à avoir $G(x, y) > 0$ pour tous $x, y \in C$.

On a vu que, dans un ensemble clos irréductible, tous les états sont de même nature : tous transitoires, tous récurrents positifs ou tous récurrent nuls. En effet, on a le théorème suivant.

THÉORÈME 6.20. 1. Soit $C \subset E$ un ensemble clos irréductible d'états récurrent. Alors pour tous $x, y \in C$, on a

$$\rho_{x,y} = 1, \mathbb{P}_x(N(y) = +\infty) = 1 \text{ et } G(x, y) = +\infty.$$

2. Soit $C \subset E$ un ensemble fini d'états clos irréductible. Alors tous les états de C sont récurrents (positifs).

3. En particulier, une chaîne irréductible sur un espace d'états fini est nécessairement récurrente.

Preuve 1. Il suffit d'appliquer les proposition 6.2 et 6.4 et le théorème 6.6.

2. Un ensemble fini et clos contient au moins un état récurrent et, par irréductibilité, tous les états sont récurrents.

3. Il suffit d'appliquer le point 2 avec $C = E$. □

COROLLAIRE 6.21 (irréductibilité, récurrence et transience). On suppose que le chaîne $(X_n)_{n \geq 0}$ est irréductible et part d'un point $x \in E$. Alors on a l'alternative suivant :

- (i) ou bien la chaîne est récurrente, *i. e.* tous les états sont récurrents ;
- (ii) ou bien la chaîne est transitoires, *i. e.* tous les états sont transitoires ;

Une chaîne irréductible est donc soit transitoire, soit récurrente positive, soit récurrente nulle.

Preuve S'il existe un état récurrent $x \in E$, alors le théorème 6.15 montre que tous les états sont récurrents puisque, par irréductibilité, l'état x mène à tous les états. De plus, comme $G(x, y) > 0$ pour tous $x, y \in E$, il n'y a qu'une seule classe de récurrence. Le reste découle du théorème précédent. □

6.3 CLASSES DE RÉCURRENCE

PROPOSITION 6.22. Sur E_R , la relation \rightsquigarrow devient une relation d'équivalence.

DÉFINITION 6.23. Deux états $x, y \in E$ communiquent si $G(x, y) > 0$ et $G(y, x) > 0$. On note alors $x \sim y$.

Avec la proposition 6.13 et le théorème 6.15, pour deux états récurrents $x, y \in E$, on a

$$x \sim y \iff x \rightsquigarrow y \iff y \rightsquigarrow x \iff G(x, y) > 0 \iff G(y, x) > 0 \iff \rho_{x,y} > 0 \iff \rho_{y,x} > 0.$$

On peut alors partitionner l'ensemble des états récurrents E_R en classes d'équivalence. On note $\{E_{R,i}\}_{i \in I}$ cette partition. Les ensembles de cette partition sont appelées les *classes de récurrence de la chaîne*. On peut alors exhiber une partition globale de l'espace d'états qui est $\{E_{R,i}\}_{i \in I} \cup \{E_T\}$, appelée la *classification des états*. Remarquons qu'une classe de récurrence est close est irréductible et, de plus, elle est soit récurrente positive soit récurrente nulle.

THÉORÈME 6.24 (classes de récurrence). 1. Soient $i \in I$ et $x \in E_{R,i}$. Alors \mathbb{P}_x -presque sûrement, on a

– pour tout $y \in E_{R,i}$, on a $N(y) = +\infty$;

– pour tout $y \in E \setminus E_{R,i}$, on a $N(y) = 0$.

2. Soient $x \in E_T$ et $T_{E_R} := \inf\{n \geq 0 \mid X_n \in E_R\}$. Alors \mathbb{P}_x -presque sûrement, on a

– ou bien $T_{E_R} = +\infty$ et $N(y) < +\infty$ pour tout $y \in E$;

– ou bien $T_{E_R} < +\infty$ et il existe un entier aléatoire $j \in I$ tel que, pour tout $n \geq T_{E_R}$, on ait $X_n \in E_{R,j}$

Preuve 1. Pour tout $y \in E \setminus E_{R,i}$, on a $G(x, y) = 0$ puisque

– si $y \in E_{R,j}$ pour $j \neq i$, alors la partition garantit que les états x et y ne communiquent pas et donc $G(x, y) = 0$;

– si $y \in E_T$, le théorème 6.15 assure encore $G(x, y) = 0$ et, en particulier, on a $N(y) = 0$ \mathbb{P}_x -presque sûrement.

En revanche, pour tout $y \in E_{R,i}$, on a $\mathbb{P}_x(\tilde{T}_y < +\infty)$ d'après le théorème 6.15 et, par la propriété de MARKOV forte, comme $y \in E_R$, on obtient

$$\begin{aligned} \mathbb{P}_x(N(y) = +\infty) &= \mathbb{E}_x[\mathbb{1}_{\{\tilde{T}_y < +\infty\}}(\mathbb{1}_{\{N(y) = +\infty\}} \circ \Theta_{T_y})] \\ &= \mathbb{P}_x(\tilde{T}_y < +\infty)\mathbb{P}_y(N(y) = +\infty) = 1. \end{aligned} \quad (*)$$

ce qui assure la conclusion.

2. On suppose $T_{E_R} < +\infty$. Alors la chaîne rentre dans l'ensemble d'état E_R et donc dans des classes $E_{R,j}$. Dans ce cas, la propriété de MARKOV et le point 1 assure $X_n \in E_{R,j}$ pour tout $n \geq T_{E_R} = T_{E_R}$. On suppose $T_{E_R} = +\infty$. Alors pour tout $y \in E$, on a l'alternative :

– si $y \in E_R$, alors $N(y) = 0$ puisque $T_{E_R} = +\infty$;

– si $y \in E_T$, alors $\mathbb{P}_y(N(y) = +\infty) = 0$ et, comme dans l'égalité (*), on a

$$\mathbb{P}_x(N(y) = +\infty) = \mathbb{P}_x(\tilde{T}_y < +\infty)\mathbb{P}_y(N(y) = +\infty) = 0$$

ce qui conclut $\mathbb{P}_x(N(y) < +\infty) = 1$. □

On peut alors préciser le théorème 6.11.

COROLLAIRE 6.25. Soit C un ensemble clos et irréductible d'états récurrents. Alors

$$\frac{G_n(x, y)}{n} \longrightarrow \frac{1}{m_y}, \quad x, y \in C.$$

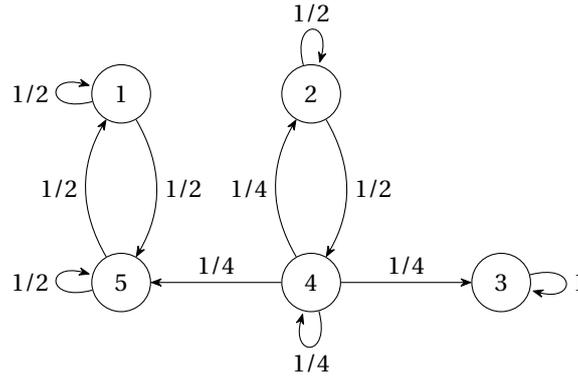
De plus, si $\mathbb{P}(X_0 \in C) = 1$, alors

$$\frac{N_n(y)}{n} \longrightarrow \frac{1}{m_y}, \quad y \in C.$$

avec probabilité 1.

DÉFINITION 6.26. Un état $x \in E$ est *absorbant* si le singleton $\{x\}$ est une classe de récurrence, i. e. si $P(x, x) = 1$.

▷ EXEMPLE. Retournons au second exemple de l'introduction dont on rappelle le graphe des transitions.



Ses classes de récurrences sont $\{1, 5\}$ et $\{3\}$. Ses états transitoires sont 2 et 4. En effet, les relations

$$P(1, 5), P(5, 1), P(1, 1), P(5, 5) > 0 \quad \text{et} \quad P(1, j) = P(5, j) = 0, \quad \forall j \in \{2, 3, 4\}$$

assure les relations

$$G(1, 5), G(5, 1) > 0, \quad G(1, j) = G(5, j) = 0, \quad \forall j \in \{2, 3, 4\}.$$

De plus, l'état 3 est bien absorbant puisque $P(3, 3) > 0$ et $P(3, j) = 0$ pour tout $j \in \{1, 2, 4, 5\}$. Par ailleurs, l'état 1 est récurrent puisque

$$\begin{aligned} \mathbb{P}_1(\tilde{T}_1 = +\infty) &= \mathbb{P}_1(X_1 = 5 \text{ et } X_n = 5, \forall n \geq 2) \\ &= P(1, 5)\mathbb{P}_5(X_n = 5, \forall n \geq 2) \\ &\leq P(1, 5) \lim_{N \rightarrow +\infty} \mathbb{P}_Y(X_n = 5, \forall n \in \{1, \dots, N\}) \\ &\leq P(1, 5) \lim_{N \rightarrow +\infty} (1/2)^N = 0 \end{aligned}$$

et donc $\mathbb{P}_1(\tilde{T}_1 < +\infty) = 1$. De la même façon, les états 5 et 3 sont récurrents. L'état 2 est transitoire car

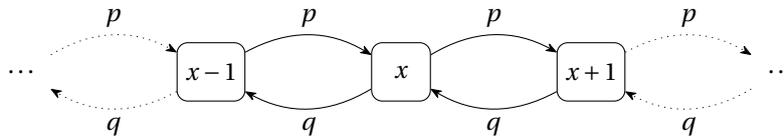
$$\mathbb{P}_2(T_Z = +\infty) \geq \mathbb{P}_2(X_1 = 4 \text{ et } X_2 = 3) = P(2, 4)P(4, 3) = 1/8$$

et donc $\mathbb{P}_2(T_2 < +\infty) < 1$. De même, l'état 4 est transitoire.

▷ **EXEMPLE (marche aléatoire sur \mathbf{Z}).** Soient $p \in]0, 1[$ et $(X_i)_{i \geq 1}$ une suite indépendante de variables aléatoires suivant la loi $(1-p)\delta_{-1} + p\delta_1$. On pose $q := 1-p$ et $S_n := X_1 + \dots + X_n$ pour $n \geq 0$. Alors la suite $(S_n)_{n \geq 0}$ est une chaîne sur l'espace d'état \mathbf{N} dont le noyau de transition P vérifie

$$P(x, y) = q\mathbb{1}_{\{y=x-1\}} + p\mathbb{1}_{\{y=x+1\}}.$$

On peut représenter cette chaîne par le graphe suivant.



Cette chaîne est irréductible puisque, pour tous $x, y \in \mathbf{N}$ tels que $x \neq y$, on a

$$P^{|x-y|}(x, y) = \begin{cases} p^{y-x} > 0 & \text{si } x < y, \\ (1-p)^{x-y} > 0 & \text{sinon} \end{cases}$$

et $P^2(x, x) = 2p(1-p) > 0$. La nature de la chaîne est donc déterminée par nature d'un point quelconque. Déterminons celle de l'état 0 : est-il récurrent ou transitoire? Pour cela, on calcule $G(0, 0)$. On remarque que, pour tout $n \geq 0$, on a $P_{2n+1}(0, 0) = 1$ et

$$P_{2n}(0, 0) = \binom{2n}{n} p^n q^n \sim \frac{(4pq)^n}{\sqrt{\pi n}}$$

en utilisant la formule de STIRLING. Ainsi deux cas se distinguent :

- si $p \neq 1/2$, alors $4pq < 1$ et $G(0, 0) < +\infty$, i. e. l'état 0 est transitoire;
- si $p = 1/2$, alors $4pq = 1$ et $G(0, 0) = +\infty$, i. e. l'état 0 est récurrent.

Évolution qualitative d'une chaîne de MARKOV

Le décomposition de l'espace d'états

$$E = \bigsqcup_{i \in I} E_{R,i} \sqcup E_T.$$

permet de comprendre l'évolution d'une chaîne de MARKOV. Deux cas de figure apparaissent.

- Si la chaîne démarre dans un des ensembles clos irréductibles récurrents $E_{R,i}$, alors elle y restera en visitant chacun de ses états infiniment et elle ne visite aucun autre état.
- Si la chaîne démarre dans E_T , alors soit elle y reste (ceci est possible seulement si l'ensemble E_T est infinie), soit elle rejoint une des classes $E_{R,i}$ et y reste en visitant alors chacun de ses états infiniment.

6.4 ABSORPTION DANS LES CLASSES DE RÉCURRENCE

Pour calculer les durées d'absorption dans les classes de récurrences, on introduit, pour $x \in E$ et $i \in I$, les quantités

$$\begin{aligned} S_i &:= \inf\{n \geq 0 \mid X_n \in E_{R,i}\}, \\ \rho_i(x) &:= \mathbb{E}_x[\mathbb{1}_{\{S_i < +\infty\}}] = \mathbb{P}_x(S_i < +\infty), \\ \tau_i(x) &:= \mathbb{E}_x[S_i \mathbb{1}_{\{S_i < +\infty\}}] \end{aligned}$$

correspondant respectivement au temps d'absorption, à la probabilité d'absorption et au temps moyen d'absorption. Immédiatement, on a

- pour tous $i \in I$ et $x \in E_{R,i}$, alors \mathbb{P}_x -presque sûrement $S_i = 0$ et donc $\rho_i(x) = 1$ et $\tau_i(x) = 0$;
- pour tous $i, j \in I$ tels que $i \neq j$ et tout $x \in E_{R,j}$, alors \mathbb{P}_x -presque sûrement $S_i = +\infty$ et donc $\rho_i(x) = \tau_i(x) = 0$.

Il reste à traiter le cas d'un état transitoire $x \in E_T$.

THÉORÈME 6.27 (absorption). Soient $x \in E_T$ et $i \in I$. Alors la probabilité d'absorption $\rho_i(x)$ et le temps moyen d'absorption $\tau_i(x)$ sont solutions du système linéaire

$$\begin{aligned} \rho_i(x) &= \sum_{y \in E} P(x, y) \rho_i(y), \\ \tau_i(x) &= \rho_i(x) + \sum_{y \in E} P(x, y) \tau_i(y). \end{aligned}$$

Preuve On a $S_i \geq 1$ et on peut écrire \mathbb{P}_x -presque sûrement $S_i = 1 + S_i \circ \Theta_1$. Avec la propriété de MARKOV faible à l'égalité (*) et le théorème de transfert, on obtient

$$\begin{aligned} \rho_i(x) &= \mathbb{E}_x[\mathbb{1}_{\{S_i < +\infty\}}] \\ &= \mathbb{E}_x[\mathbb{1}_{\{1 + S_i \circ \Theta_1 < +\infty\}}] \\ &= \mathbb{E}_x[\mathbb{1}_{\{S_i \circ \Theta_1 < +\infty\}}] \\ &= \mathbb{E}_x[\mathbb{E}_x[\mathbb{1}_{\{S_i \circ \Theta_1 < +\infty\}} \mid \mathcal{F}_1]] \\ &= \mathbb{E}_x[\mathbb{E}_{X_1}[\mathbb{1}_{\{S_i < +\infty\}}]] \\ &= \mathbb{E}_x[\rho_i(X_1)] = \sum_{y \in E} P(x, y) \rho_i(y). \end{aligned} \tag{*}$$

Par les mêmes arguments, on a

$$\begin{aligned} \tau_i(x) &= \mathbb{E}_x[(1 + S_i \circ \Theta_1) \mathbb{1}_{\{1 + S_i \circ \Theta_1 < +\infty\}}] \\ &= \mathbb{E}_x[\mathbb{E}_{X_1}[(1 + S_i) \mathbb{1}_{\{S_i < +\infty\}}]] \\ &= \mathbb{E}_x[\mathbb{E}_{X_1}[S_i \mathbb{1}_{\{S_i < +\infty\}}] + \mathbb{E}_{X_1}[\mathbb{1}_{\{S_i < +\infty\}}]] \\ &= \mathbb{E}[\tau_i(X_1)] + \mathbb{E}_x[\rho_i(X_1)] \\ &= \sum_{y \in E} P(x, y) \tau_i(y) + \sum_{y \in E} P(x, y) \rho_i(y) \\ &= \sum_{y \in E} P(x, y) \tau_i(y) + \rho_i(x). \end{aligned} \quad \square$$

Chapitre 7

INVARIANCE ET ÉQUILIBRE

7.1 Mesures invariantes	53	7.3 Périodicité et forte réductibilité	60
7.1.1 Invariance	53	7.3.1 Périodicité	60
7.1.2 Réversibilité	55	7.3.2 Forte irréductibilité	61
7.2 Invariance et récurrence	56	7.4 Équilibre d'une chaîne de MARKOV	62
		7.5 Théorème ergodique	63

Dans ce chapitre, on étudie les mesures qui sont invariantes pour une chaîne de MARKOV. On fait le lien entre ces mesures, les états récurrents (positifs) et leur temps de retour. On étudie le comportement en temps long des chaînes de MARKOV et la convergence vers un régime d'équilibre, en lien avec le théorème ergodique

Par défaut, on considère, dans ce chapitre, une chaîne de MARKOV $(X_n)_{n \geq 0}$ sur un espace d'états au plus dénombrable E et de noyau de transition P .

7.1 MESURES INVARIANTES

7.1.1 Invariance

DÉFINITION 7.1 (mesure invariante). Une mesure positive non nulle π sur E telle que $\pi(x) < +\infty$ pour $x \in E$ est dite *invariante* (ou *stationnaire*) pour le noyau de transition P si elle vérifie l'équation de CHAPMAN-KOLMOGOROV

$$\pi = \pi P,$$

c'est-à-dire

$$\forall y \in E, \quad \pi(y) = \sum_{x \in E} \pi(x)P(x, y).$$

◇ **REMARQUE.** Par une récurrence immédiate, une mesure invariante π vérifie également les équations $\pi = \pi P^n$ pour $n \geq 0$

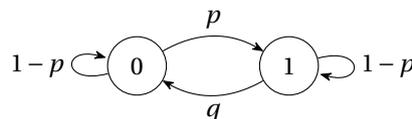
PROPOSITION 7.2. Alors la loi de X_n est indépendante de n si et seulement si la distribution initiale μ_0 est une probabilité invariante

Preuve Pour le sens direct, pour $n \geq 0$, comme $\mu_n = \mu_0 P^n$, si la loi μ_n de X_n ne dépend pas de n , alors $\mu_1 = \mu_0$ et donc $\mu_0 = \mu_0 P$ ce qui montre l'invariance de la mesure μ_0 . Réciproquement, si μ_0 est une probabilité invariante, alors $\mu_n = \mu_0$ pour tout $n \geq 0$ par une récurrence immédiate. □

◇ **REMARQUE.** Attention, lorsque l'espace d'état E est infini, il se peut qu'il existe une mesure invariante mais pas de probabilité invariante (comme c'est le cas dans l'exemple suivant). Dans ce cas, cette mesure invariante est de masse infini et n'est donc pas normalisable en une probabilité.

▷ **EXEMPLES.** – On considère la chaîne de MARKOV à deux états de matrice de transition

$$\begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}.$$



Alors elle n'a qu'une seule probabilité invariante qui est

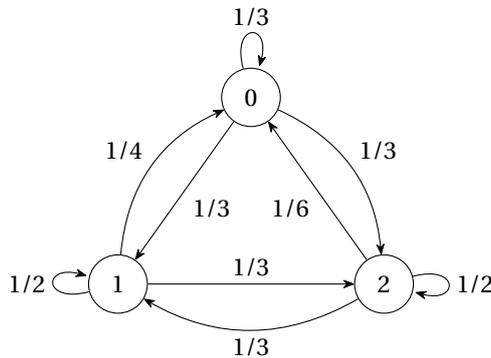
$$\pi := \left(\frac{q}{p+q}, \frac{p}{p+q} \right).$$

– Le mesure uniforme est l'unique mesure invariante pour la marche aléatoire simple symétrique sur \mathbf{Z} . En particulier, il n'existe pas de probabilité invariante puisque l'espace d'état \mathbf{Z} est infini.

Soit π une mesure invariante pour la marche aléatoire simple symétrique sur \mathbf{Z} . Alors pour tout $n \in \mathbf{Z}$, elle vérifie $\pi(n) = \frac{1}{2}\pi(n-1) + \frac{1}{2}\pi(n+1)$. Ainsi la suite $(\pi(n+1) - \pi(n))_{n \in \mathbf{Z}}$ est constante égale à un certain réel $\alpha \in \mathbf{R}$.

On en déduit alors $\pi(n) = \pi(0) + n\alpha$ pour tout $n \in \mathbf{Z}$. Or la mesure π est positive, cela implique $\alpha = 0$. Ainsi La mesure π est la mesure uniforme sur \mathbf{Z} et, réciproquement, c'est bien une mesure invariante.

– On considère la chaîne de MARKOV sur $E := \{0, 1, 2\}$ donnée par le graphe suivant.



Une probabilité invariante π pour cette chaîne est solution du système linéaire

$$\begin{cases} \frac{1}{3}\pi(0) + \frac{1}{2}\pi(1) + \frac{1}{6}\pi(2) = \pi(0), \\ \frac{1}{3}\pi(0) + \frac{1}{2}\pi(1) + \frac{1}{3}\pi(2) = \pi(1), \\ \frac{1}{3}\pi(0) + \frac{1}{4}\pi(1) + \frac{1}{2}\pi(2) = \pi(0), \\ \pi(0) + \pi(1) + \pi(2) = 1. \end{cases}$$

On en déduit facilement $\pi = (6/25, 2/5, 9/25)$ et on vérifie que la mesure π est bien une probabilité vérifiant le système précédent. Dans ce cas, on a existence et unicité de la probabilité invariante.

- ◊ REMARQUE. Pour une matrice bistochastique P , i. e. telles que les matrices P et tP soient stochastiques, on observe que le vecteur $(1, \dots, 1)$ est un vecteur propre à gauche de P pour la valeur propre 1. Ce vecteur correspond à la mesure uniforme sur E . Ainsi la mesure uniforme est invariante pour un noyau de transition bistochastique. De plus, si l'espace d'état E est fini de cardinal $d \geq 1$, alors on peut normaliser le vecteur $(1, \dots, 1)$ et on trouve que la probabilité uniforme $(1/d, \dots, 1/d)$ est invariante.

PROPOSITION 7.3. L'ensemble des mesures invariantes d'un noyau de transition est stable par combinaison linéaire à coefficients positifs. Si E est fini, alors l'ensemble des probabilités invariantes est un compact. De plus, l'ensemble des probabilités invariantes est convexe.

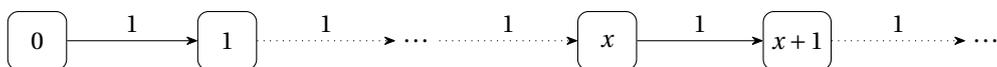
Preuve Cela suit immédiatement de la linéarité de l'équation de CHAPMAN-KOLMOGOROV. Si E est fini, alors l'ensemble des probabilités invariantes s'identifie à la partie

$$\{(\mu_1, \dots, \mu_d) \in \mathbf{R}_+^d \mid \mu_1 + \dots + \mu_d = 1\}$$

qui est fermée et bornée ce qui en fait une partie compacte. Dès lors, la compacité de l'ensemble des mesures invariantes suit. □

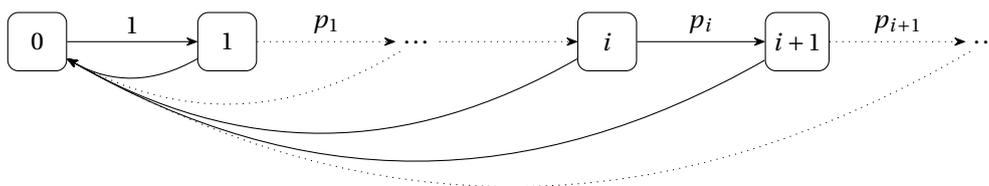
Il n'existe pas toujours de mesure invariante pour une chaîne de MARKOV comme on peut le voir dans les exemples suivants.

- ▷ EXEMPLES. – On considère le noyau de transition P sur \mathbf{N} donné par $P(i, i + 1) = 1$ pour tout $i \geq 0$. Alors il n'existe pas de mesure invariante. En effet, si une mesure invariante π existait, alors on aurait $\pi(i) = \pi(i - 1)$ pour tout $i \geq 1$ et $\pi(0) = 0$, donc $\pi = 0$ ce qui est impossible.



– Soit $(p_i)_{i \geq 1}$ une suite de $[0, 1]$ telle que $\sum_{i \geq 1} (1 - p_i) < 1$. On considère la chaîne de MARKOV $(X_n)_{n \geq 0}$ sur \mathbf{N} dont le noyau de transition est donnée par $P(0, 1) = 1$ et

$$P(i, i + 1) = p_i \quad \text{et} \quad P(i, 0) = 1 - p_i, \quad i \geq 1.$$



Il s'agit d'une chaîne de MARKOV irréductible puisque, pour tous $i, j \in \mathbf{N}$, le chemin $i \rightarrow 0 \rightarrow \dots \rightarrow j$ est possible lorsque $p_k \neq 1$ pour $k \in \mathbf{N}$. Raisonnons par l'absurde et supposons qu'il existe une mesure invariante π . Alors elle vérifie

$$\pi(0) = \sum_{i=1}^{+\infty} (1 - p_i)\pi(i) \quad \text{et} \quad \pi(i) = p_{i-1}\pi(i-1), \quad i \geq 1.$$

Une récurrence immédiate assure $\pi(i) = \pi(0)p_1 \cdots p_{i-1}$ pour tout $i \geq 1$ ce qui, en réinjectant dans l'expression de $\pi(0)$, donne

$$\pi(0) = \sum_{i=1}^{+\infty} \left((1 - p_i) \prod_{j=1}^{i-1} p_j \right).$$

Si tous les réels p_i valent 1, alors $\pi(0) = 0$ puis $\pi(i) = 0$ pour tout $i \geq 0$ ce qui est impossible. Donc il existe $i \geq 1$ tel que $p_i < 1$. Par conséquent, si $0 < \pi(0) < +\infty$, on obtient $\pi(0) \leq \pi(0) \sum_{i=1}^{+\infty} (1 - p_i) < \pi(0)$ ce qui est impossible. On doit donc avoir $\pi(0) = 0$ et donc $\pi(i) = 0$ pour tout $i \geq 0$. Donc le mesure π est nulle si bien qu'il n'y a pas de mesure invariante pour cette chaîne.

7.1.2 Réversibilité

DÉFINITION 7.4 (mesure réversible). Une mesure positive non nulle π sur E telle que $\pi(x) < +\infty$ pour $x \in E$ est dite *réversible* pour le noyau de transition P si elle vérifie

$$\pi(x)P(x, y) = \pi(y)P(y, x), \quad \forall x, y \in E.$$

De manière équivalente, cela revient à avoir

$$\pi(x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n) = \pi(x_n)P(x_n, x_{n-1}) \cdots P(x_1, x_0), \quad \forall n \geq 0, \forall x_0, \dots, x_n \in E. \quad (*)$$

PROPOSITION 7.5. Une loi π est réversible pour un noyau de transition P si et seulement si, pour toute chaîne de MARKOV $(X_n)_{n \geq 0}$ de noyau P et de loi initiale π et tout entier $n \geq 0$, on a

$$\mathcal{L}(X_0, X_1, \dots, X_n) = \mathcal{L}(X_n, X_{n-1}, \dots, X_0).$$

Preuve Pour le sens direct, cela vient de la définition équivalente (*) puisque, pour tous $x_0, \dots, x_n \in E$, on a

$$\begin{aligned} \mathbb{P}(X_0 = x_0, \dots, X_n = x_n) &= \pi(x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n), \\ \text{et } \mathbb{P}(X_n = x_0, \dots, X_0 = x_n) &= \mathbb{P}(X_0 = x_n, \dots, X_n = x_0) = \pi(x_n)P(x_n, x_{n-1}) \cdots P(x_1, x_0). \end{aligned}$$

Réciproquement, comme $\mathcal{L}(X_0, X_1) = \mathcal{L}(X_1, X_0)$, pour $x_0, x_1 \in E$, on a $\mathbb{P}(X_0 = x_0, X_1 = x_1) = \mathbb{P}(X_0 = x_1, X_1 = x_0)$, c'est-à-dire $\pi(x_0)P(x_0, x_1) = \pi(x_1)P(x_1, x_0)$. \square

PROPOSITION 7.6 (réversibilité et invariance). Une mesure réversible pour un noyau markovien est invariante pour ce noyau.

Preuve On vérifie immédiate l'équation de CHAPMAN-KOLMOGOROV en utilisant la réversibilité. \square

► **EXEMPLES.** – Soit $p \in]0, 1[$ On considère la marche aléatoire sur \mathbf{Z} dont le noyau est donnée par $P(i, i+1) = p$ et $P(i, i-1) = q := 1 - p$ pour tout $i \in \mathbf{Z}$. Alors la mesure définie par $\pi(i) = (p/q)^i$ pour $i \geq 0$ est réversible car, pour tout $i \in \mathbf{Z}$, on a

$$\pi(i)P(i, i+1) = \left(\frac{p}{q}\right)^i p = \frac{p^{i+1}}{q^{i+1}} q = \pi(i+1)P(i+1, i).$$

Comme $P(i, j) = 0$ dès que $|i - j| > 1$, on en déduit $\pi(i)P(i, j) = \pi(j)P(j, i)$ pour tous $i, j \geq 0$.

– On reprend le modèle de WRIGHT-FISCHER. Comme les mesures invariantes sont concentrées sur des états absorbants et les mesures δ_0 et δ_N sont réversibles, ces dernières sont les seules mesures réversibles pour cette chaîne.

- ◊ REMARQUE. Pour trouver des mesures invariantes, il faut résoudre le système linéaire donné par l'équation de CHAPMAN-KOLMOGOROV. En pratique, ce système peut être compliqué à résoudre (il est même infini si E l'est). Dans ce cas, il peut être pertinent de rechercher mieux en cherchant des mesures réversibles car l'équation associée est, en pratique, plus simple à résoudre. On est alors assuré par la Proposition précédente qu'une solution serait aussi invariante.

7.2 INVARIANCE ET RÉCURRENCE

Dans cette section, on discute de l'existence et de l'unicité de mesures ou probabilités invariantes pour des chaînes de MARKOV. On voit en particulier que les mesures invariantes sont liées aux états récurrents (positifs). Ci-dessous, les échanges de limites et de sommes sont justifiés par le résultat d'interversion suivant (dû au théorème de convergence dominée).

LEMME 7.7 (convergence dominée). Soit $(a(x))_{x \in E}$ une suite de réels positifs de somme finie et $(b_n(x))_{x \in E, n \geq 1}$ une famille de $[-1, 1]$ telle que, pour tout $x \in E$, la suite $(b_n(x))_{n \geq 1}$ converge vers un réel $b(x) \in [-1, 1]$. Alors

$$\lim_{n \rightarrow +\infty} \sum_{x \in E} a(x) b_n(x) = \sum_{x \in E} a(x) b(x).$$

Le premier résultat indique d'une probabilité invariante ne charge que les états récurrents positifs.

PROPOSITION 7.8 (support d'une probabilité invariante). Soit π une probabilité invariante. Alors $\pi(x) = 0$ pour tout état $x \in E$ transitoire ou récurrent nul.

Preuve Soit $x \in E$ un état transitoire ou récurrent nul. Alors $m_x = +\infty$ et le corollaire 6.25 donne

$$G_n(z, x) / n \rightarrow 0, \quad z \in E.$$

Mais comme π est invariante, pour tout $k \geq 1$, on a $\pi P^k = \pi$ ce qui se réécrit sous la forme

$$\pi(x) = \sum_{z \in E} \pi(z) P^k(z, x)$$

et, en sommant, on obtient

$$\pi(x) = \sum_{z \in E} \pi(z) \frac{G_n(z, x)}{n}.$$

Comme π est une probabilité, le lemme précédent s'applique et permet de passer à la limite dans cette dernière également pour obtenir $\pi(x) = 0$. \square

- ◊ REMARQUE. L'utilisation de lemme dans la preuve précédente exige d'avoir $\sum_{x \in E} \pi(x) < +\infty$ et, à une normalisation près, que la mesure π soit une probabilité. La proposition précédente ne concerne donc que les probabilités invariantes. Par conséquent, une chaîne qui n'a pas d'états récurrents positifs n'a pas de probabilité invariante.

PROPOSITION 7.9 (invariance et irréductibilité). Pour une chaîne de MARKOV irréductible, une mesure invariante charge tous les points.

Preuve Soient $(X_n)_{n \geq 0}$ une chaîne de MARKOV irréductible et π une mesure invariante pour son noyau de transition P . Comme $\pi \neq 0$, il existe $x \in E$ tel que $\pi(x) > 0$. Pour tout $y \in E$, comme la chaîne est irréductible, il existe $n \geq 1$ tel que $P^n(x, y) > 0$ et, avec l'équation de CHAPMAN-KOLMOGOROV, on a donc

$$\pi(y) = \sum_{z \in E} \pi(z) P^n(z, y) \geq \pi(x) P^n(x, y) > 0. \quad \square$$

PROPOSITION 7.10 (invariance et transience). Pour une chaîne de MARKOV irréductible transitoire, une mesure invariante est de masse infinie. En particulier, l'espace doit être infini et il n'existe pas de probabilité invariante.

Preuve Soient $(X_n)_{n \geq 0}$ une chaîne de MARKOV irréductible transitoire et π une mesure invariante pour son noyau de transition P . Comme la chaîne est transitoire, pour tous $x, y \in E$, on a $G(x, y) < +\infty$, donc $P^n(x, y) \rightarrow 0$. Si $\pi(E) < +\infty$, alors un passage à la limite dans l'équation $\pi = \pi P^n$ à l'aide du lemme 7.7 donne

$$\pi(y) = \lim_{n \rightarrow +\infty} \sum_{x \in E} \pi(x) P^n(x, y) = \sum_{x \in E} \pi(x) \lim_{n \rightarrow +\infty} P^n(x, y) = 0$$

ce qui est contradiction avec la proposition précédente. D'où $\pi(E) = +\infty$. Le reste en découle facilement. \square

On montre maintenant qu'on peut associer une mesure invariante à tout état récurrent x en calculant, pour chaque état y , le nombre moyen de visite de cet état dans l'excursion de la chaîne entre deux visites de x . Il s'agit d'une construction trajectorielle des mesures invariantes. À un état $x \in E$, on associe la mesure ν_x définie par

$$\nu_x(y) := \mathbb{E}_x \left[\sum_{k=0}^{\tilde{T}_x-1} \mathbb{1}_{\{X_k=y\}} \right], \quad y \in E.$$

Remarquons que, pour un état absorbant $x \in E$, on a $\nu_x = \delta_x$.

PROPOSITION 7.11 (mesure invariante d'un état récurrent). Soit $x \in E$ un état récurrent.

1. La mesure ν_x est invariante et vérifie $\nu_x(x) = 1$.
2. Elle a pour support la classe de récurrence de l'état x , i. e. pour tout $y \in E$, on a $\nu_x(y) > 0$ si et seulement si l'état y appartient à la classe de récurrence de l'état x .

Preuve Puisque sous \mathbb{P}_x , on a $X_0 = X_{\tilde{T}_x} = x$, donc la quantité $\nu_x(y)$ pour $y \in E$ se réécrit

$$\nu_x(y) = \mathbb{E}_x \left[\sum_{k=1}^{\tilde{T}_x} \mathbb{1}_{\{X_k=y\}} \right]. \quad (*)$$

Pour tout état $y \in E$ qui n'est pas dans la classe de récurrence de x , les états x et y ne communiquent pas et donc

$$G(x, y) = \mathbb{E}_x[\tilde{N}_y] = \mathbb{E}_x \left[\sum_{k=0}^{+\infty} \mathbb{1}_{\{X_k=y\}} \right] = 0$$

ce qui implique $\nu_x(y) = 0$. Ensuite, puisque sous \mathbb{P}_x , on a $X_0 = x$ et $X_k \neq x$ pour $k \in \llbracket 1, \tilde{T}_x - 1 \rrbracket$, on a $\nu_x(x) = 1$.

Montrons qu'elle est invariante. Pour tout $y \in E$, en utilisant la relation (*), le théorème de FUBINI-TONELLI puis la propriété de MARKOV faible, on obtient

$$\begin{aligned} \nu_x(y) &= \mathbb{E}_x \left[\sum_{k=1}^{\tilde{T}_x} \mathbb{1}_{\{X_k=y\}} \right] \\ &= \mathbb{E}_x \left[\sum_{z \in E} \sum_{k=1}^{\tilde{T}_x} \mathbb{1}_{\{X_k=y, X_{k-1}=z\}} \right] \\ &= \sum_{z \in E} \sum_{k=1}^{+\infty} \mathbb{E}_x[\mathbb{1}_{\{\tilde{T}_x \geq k\}} \mathbb{1}_{\{X_k=y\}} \mathbb{1}_{\{X_{k-1}=z\}}] \\ &= \sum_{z \in E} \sum_{k=1}^{+\infty} \mathbb{E}_x[\mathbb{1}_{\{\tilde{T}_x \geq k\}} \mathbb{1}_{\{X_{k-1}=z\}} \mathbb{E}_x[\mathbb{1}_{\{X_k=y\}} \mid \mathcal{F}_{k-1}]] \quad (\text{car } \mathbb{1}_{\{\tilde{T}_x \geq k\}} \mathbb{1}_{\{X_{k-1}=z\}} \text{ est } \mathcal{F}_{k-1}\text{-mesurable}) \\ &= \sum_{z \in E} \sum_{k=1}^{+\infty} \mathbb{E}_x[\mathbb{1}_{\{\tilde{T}_x \geq k\}} \mathbb{1}_{\{X_{k-1}=z\}} \mathbb{E}_{X_{k-1}}[\mathbb{1}_{\{X_k=y\}}]] \\ &= \sum_{z \in E} \sum_{k=1}^{+\infty} \mathbb{E}_x[\mathbb{1}_{\{\tilde{T}_x \geq k\}} \mathbb{1}_{\{X_{k-1}=z\}} P(z, y)] \\ &= \sum_{z \in E} P(z, y) \mathbb{E}_x \left[\sum_{k=1}^{\tilde{T}_x} \mathbb{1}_{\{X_{k-1}=z\}} \right] \\ &= \sum_{z \in E} P(z, y) \nu_x(z) = \nu_x P(y). \end{aligned}$$

Cela montre $\nu_x P = \nu_x$.

Montrons désormais le second point. Soit $y \in E$. L'invariance assure

$$\nu_x(x) = \sum_{z \in E} \nu_x(z) P(z, x) = \sum_{z \in E} \nu_x(z) P^n(z, x).$$

On suppose que les états x et y appartiennent à la même classe de récurrence. Alors il existe un entier $n \geq 0$ tel que $P^n(x, y) > 0$ et, avec la relation précédente, on doit nécessairement avoir $\nu_x(y) < +\infty$. Et comme $x \sim y$, il existe aussi un entier $m \geq 1$ tel que $P^m(x, y) > 0$ et on obtient alors

$$\nu_x(y) = \sum_{z \in E} \nu_x(z) P^m(z, y) \geq \nu_x(x) P^m(x, y) = P^m(x, y) > 0.$$

On montre identiquement la réciproque ce qui conclut cette preuve. \square

PROPOSITION 7.12. On considère une chaîne de MARKOV $(X_n)_{n \geq 0}$ irréductible récurrente. Alors la mesure invariante est unique à une constante multiplicative près. Ainsi pour toute mesure invariante π , on a $\pi = \pi(x)v_x$ pour tout $x \in E$.

Preuve Soit π une mesure invariante. Montrons, par récurrence sur l'entier $p \geq 0$, que

$$\pi(y) \geq \pi(x) \mathbb{E}_x \left[\sum_{k=0}^{\min(p, \tilde{T}_x - 1)} \mathbb{1}_{\{X_k=y\}} \right], \quad x, y \in E. \quad (*)$$

D'abord, l'inégalité est simple lorsque $y = x$ ou $p = 0$:

- si $y = x$, alors le terme dans l'espérance vaut 1 et on a bien $\pi(x) \geq \pi(x) \times 1$;
- si $p = 0$, alors ce dernier vaut $\mathbb{1}_{\{X_0=y\}}$ et on a bien $\pi(y) \geq \pi(x) \delta_{x,y} = \pi(x) \mathbb{E}_x[\mathbb{1}_{\{X_0=y\}}]$.

Dans le cas général, soit $p \geq 1$ et supposons que l'état est vrai au rang p . Soient $x, y \in E$ deux états distincts. Alors l'invariance de π , l'hypothèse de récurrence, la propriété de MARKOV faible puis le théorème de FUBINI-TONELLI assurent successivement

$$\begin{aligned} \pi(y) &= \sum_{z \in E} \pi(z) P(z, y) \\ &\geq \pi(x) \sum_{z \in E} \mathbb{E}_x \left[\sum_{k=0}^{\min(p, \tilde{T}_x - 1)} \mathbb{1}_{\{X_k=y\}} \right] P(z, y) \\ &= \pi(x) \sum_{z \in E} \sum_{k=0}^p \mathbb{E}_x[\mathbb{1}_{\{X_k=z\}} \mathbb{1}_{\{\tilde{T}_x - 1 \geq k\}}] P(z, y) \\ &= \pi(x) \sum_{z \in E} \sum_{k=0}^p \mathbb{E}_x[\mathbb{1}_{\{X_k=z\}} \mathbb{1}_{\{\tilde{T}_x - 1 \geq k\}}] \mathbb{E}_{X_k}[\mathbb{1}_{\{X_1=y\}}] \\ &= \pi(x) \sum_{z \in E} \sum_{k=0}^p \mathbb{E}_x[\mathbb{1}_{\{X_k=z\}} \mathbb{1}_{\{\tilde{T}_x - 1 \geq k\}}] \mathbb{E}_x[\mathbb{1}_{\{X_{k+1}=y\}} | \mathcal{F}_k] \\ &= \pi(x) \sum_{z \in E} \sum_{k=0}^p \mathbb{E}_x[\mathbb{1}_{\{X_k=z\}} \mathbb{1}_{\{\tilde{T}_x - 1 \geq k\}}] \mathbb{1}_{\{X_{k+1}=y\}} \\ &= \pi(x) \mathbb{E}_x \left[\sum_{k=0}^{\min(p, \tilde{T}_x - 1)} \mathbb{1}_{\{X_{k+1}=y\}} \right] \\ &= \pi(x) \mathbb{E}_x \left[\sum_{k=0}^{\min(p, \tilde{T}_x - 1)} \mathbb{1}_{\{X_k=y\}} \right] = \pi(x) \mathbb{E}_x \left[\sum_{k=0}^{\min(p+1, \tilde{T}_x - 1)} \mathbb{1}_{\{X_k=y\}} \right]. \quad (\text{car } x \neq y) \end{aligned}$$

Ceci termine la récurrence. En laissant tendre p vers l'infini dans l'inégalité (*), le théorème de convergence monotone assure

$$\pi(y) \geq \pi(x) \mathbb{E}_x \left[\sum_{k=0}^{\tilde{T}_x - 1} \mathbb{1}_{\{X_k=y\}} \right] = \pi(x) v_x(y), \quad x, y \in E.$$

Soient $x \in E$ et $n \geq 1$. En combinant alors l'invariance de π , celle de v_x et le fait que $v_x(x) = 1$, on obtient

$$\begin{aligned} \pi(x) &= \sum_{z \in E} \pi(z) P^n(z, x) \\ &\geq \sum_{z \in E} \pi(x) v_x(z) P^n(z, x) \\ &= \pi(x) v_x(x) = \pi(x). \end{aligned}$$

Finalement, l'inégalité ci-dessus est une égalité et, dès que $P^n(z, x) > 0$, il en va de même pour l'inégalité

$$\pi(z) \geq \pi(x) v_x(z).$$

Par irréductibilité, il existe bien un tel entier $n \geq 1$ ce qui montre finalement $\pi(z) = \pi(x) v_x(z)$ pour tout $z \in E$. \square

Le théorème suivant dresse un bilan des résultats précédents pour l'invariance d'une chaîne de MARKOV. Attention, on rappelle que, d'après un exemple précédent, il n'existe pas toujours de mesure invariante pour une chaîne irréductible.

THÉORÈME 7.13 (invariance et irréductibilité). On considère une chaîne de MARKOV $(X_n)_{n \geq 0}$ irréductible. Alors il y a trois cas distincts :

1. La chaîne est transitoire : toute mesure invariante est de masse infinie et il n'existe pas de probabilité invariante.
2. La chaîne est récurrente nulle : toute mesure invariante est de masse infinie et il n'existe pas de probabilité

invariante.

3. La chaîne est récurrente positive : il existe une unique probabilité invariante π qui est donnée par

$$\pi(x) = 1/\mathbb{E}_x[\tilde{T}_x], \quad x \in E.$$

◇ REMARQUES. – En particulier, une chaîne de MARKOV irréductible sur un espace d'état fini est récurrente positive et admet une unique probabilité invariante.

– De plus, on peut préciser la proposition 7.8. En effet, quand elle existe, l'unique probabilité invariante est donnée par $\pi(x) = 1/\mathbb{E}_x[\tilde{T}_x]$ pour $x \in E$ et elle est concentrée sur les états récurrents positifs.

Preuve Le premier cas suit de la proposition 7.10. Ensuite pour une chaîne irréductible récurrente, la proposition 7.12 assure que les mesures toutes proportionnelles. Ainsi soit il en existe une finie et elle le sont toutes, soit elles sont toutes infinies. Cela justifie l'alternative entre les deuxième et troisième points. Il reste à voir qu'ils correspondent respectivement à la récurrence nulle et à la récurrence positive.

On suppose qu'il existe une mesure invariante finie. Notons π l'unique probabilité invariante. Soit $x \in E$ un état récurrent. Comme π et ν sont proportionnelles, on a nécessairement $\pi = \nu_x/\nu_x(E)$ et, comme $\nu_x(x) = 1$, il vient $\pi(x) = 1/\nu_x(E)$. De plus, le théorème de convergence monotone donne

$$\nu_x(E) = \sum_{y \in E} \nu_x(y) = \sum_{y \in E} \mathbb{E}_x \left[\sum_{k=0}^{\tilde{T}_x-1} \mathbb{1}_{\{X_k=y\}} \right] = \mathbb{E}_x \left[\sum_{k=0}^{\tilde{T}_x-1} \sum_{y \in E} \mathbb{1}_{\{X_k=y\}} \right] = \mathbb{E}_x[\tilde{T}_x].$$

D'où $\mathbb{E}_x[\tilde{T}_x] < +\infty$, i. e. l'état x est récurrent positif et donc la chaîne l'est aussi.

Dans le cas alternative où toutes les mesures invariantes sont infinies, la mesure ν_x est alors de masse infinie, donc $\mathbb{E}_x[\tilde{T}_x] = +\infty$ par le même calcul que précédemment, donc la chaîne est récurrente nulle. □

▷ EXEMPLE. Soit $p \in]0, 1[$. On considère la chaîne de MARKOV dont le noyau de transition est donné par

$$P(k, k+1) = p, \quad P(k, k-1) = q := 1-p, \quad P(0, 1) = 1, \quad k \geq 1.$$

On peut vérifier que la mesure π donnée par

$$\varphi(k) = (p/q)^{k-1}, \quad \varphi(0) = q, \quad k \geq 1$$

est réversible et donc invariante. Deux cas se présentent.

– Si $p < q$, alors la mesure π est finie et la chaîne est récurrente positive par le théorème précédent. La probabilité invariante correspondant, obtenu en normalisant la mesure π , est la loi géométrique sur \mathbf{N} de paramètre p/q .

– Si $p > q$, alors elle est infinie et la chaîne est récurrente nulle.

Cas des chaînes réductibles

De façon générale, une chaîne de MARKOV admet une classification non triviale de ses états. Dans cette partition, les classes récurrentes positives $E_{R^+, i}$ sont les seules classes irréductibles qui portent une probabilité invariante π_i . L'ensemble des probabilités invariantes est alors donné par les combinaisons convexes de ces probabilités invariantes π_i .

THÉORÈME 7.14 (invariance pour les chaînes non irréductibles). On considère une chaîne de MARKOV $(X_n)_{n \geq 0}$ non irréductible. Sur chaque classe de récurrence, il existe une mesure invariante (qui est unique à un facteur multiplicatif près). De plus, il existe une probabilité invariante si et seulement si la classe est récurrente positive et, dans ce cas, elle est donnée par $\pi(x) = 1/\mathbb{E}_x[\tilde{T}_x]$.

Preuve Soient R^* une classe de récurrence et $x \in R^*$. On considère une mesure invariante π qui chaque x . Par la proposition 7.8, la mesure ν_x est une telle mesure de support R^* . Par ailleurs, pour tout $y \in E$, on a déjà montré $\pi(y) \geq \pi(x)\nu_x(y)$. Cela exige que le support de π est R^* . En effet,

– si $\pi(y) = 0$, alors $\nu_x(y) = 0$ et $y \notin R^*$ car $\pi(x) > 0$ par la proposition 7.8;

– si $y \in R^*$, alors $\pi(y) \geq \pi(x)\nu_y(x) > 0$ et l'état y appartient au support de π .

Puis en reprenant l'argument de la fin de la preuve de la proposition 7.8, on doit avoir $\pi(y) = \pi(x)\nu_x(y)$ dès qu'il existe $n \geq 1$ tel que $P^n(y, x) > 0$. Comme un tel entier n existe pour tout $y \in R^*$, on a bien $\pi = \pi(x)\nu_x$ et il y a unicité de la mesure invariante sur R^* à un facteur multiplicatif près.

Le même argument que dans la preuve du théorème 7.13 prouve la dernière partie sur la probabilité invariante. □

Bilan pour la classification des états

On reprend la partie de l'espace d'état

$$E = E_T \sqcup \bigsqcup_{i \in I} E_{R,i}.$$

Avec ce qui précède, sur chaque classe de récurrence $E_{R,i}$, il existe une mesure invariante qui est donnée par

$$\nu_x^{(i)}(y) = \mathbb{E}_x \left[\sum_{k=0}^{\tilde{T}_x-1} \mathbb{1}_{\{X_k=y\}} \right], \quad x, y \in E_{R,i}.$$

Si la classe $E_{R,i}$ est récurrente positive, alors il existe une unique probabilité invariante concentrée sur $E_{R,i}$, c'est la mesure $\nu_x^{(i)}/\nu_x^{(i)}(E_{R,i})$. Si la classe $E_{R,i}$ est récurrente nulle, il n'existe pas de probabilité invariante sur $E_{R,i}$ (la mesure $\nu_x^{(i)}$ est de masse infinie et donc ne se normalise pas en une probabilité). De plus, toutes combinaisons linéaires à coefficients positifs des mesures $\nu_x^{(i)}$ fournit une mesure invariante sur E .

Pour une chaîne de MARKOV irréductible, tous les états sont de même nature (récurrents positifs, récurrents nuls ou transitoires).

Lorsque la chaîne de MARKOV est irréductible, elle admet au plus une probabilité invariante. De plus, si π est une telle probabilité, alors $\pi(x) > 0$ pour tout $x \in E$ par la proposition 7.9.

Lorsque la chaîne de MARKOV est irréductible récurrente, il existe une unique mesure invariante (à coefficient multiplicatif près). Si cette mesure est finie, alors la chaîne est positive et la probabilité invariante est donnée par $\pi(x) = 1/\mathbb{E}_x[\tilde{T}_x]$ et, dans ce cas, la chaîne est récurrente positive d'après le théorème 7.13

COROLLAIRE 7.15. On considère une chaîne de MARKOV irréductible. Alors les assertions suivantes sont équivalentes :

- (i) il existe une unique probabilité invariante π ;
- (ii) la mesure π définie par $\pi(x) = 1/\mathbb{E}_x[\tilde{T}_x]$ pour $x \in E$ est la probabilité invariante;
- (iii) il existe un état récurrent positif;
- (iv) tous les états sont récurrents positifs.

Preuve Ces équivalences découlent du théorème 7.13. □

7.3 PÉRIODICITÉ ET FORTE RÉDUCTIBILITÉ

7.3.1 Périodicité

DÉFINITION 7.16. La *période* d'un état $x \in E$ d'une chaîne de MARKOV de transition P est l'entier

$$d_x := \text{pgcd}(n \geq 1 \mid P^n(x, x) > 0)$$

avec la convention $d_x = 0$ si $P^n(x, x) = 0$ pour tout $n \geq 1$. On dit que l'état x est apériodique lorsque $d_x = 1$.

▷ **EXEMPLE.** Pour la marche aléatoire simple sur \mathbf{Z} , on a vu que $P_{2n+1}(0, 0) = 0$ si $n \geq 0$ et $P_{2n}(0, 0) > 0$ si $n \geq 1$. On en déduit que la période de l'état 0 est $d_0 = 2$.

PROPOSITION 7.17 (période commune des états communiquant). Deux états communiquant ont la même période.

Preuve Soient $x, y \in E$ deux états communiquant. Alors il existe $n, m \geq 1$ tels que $P^n(x, y) > 0$ et $P^m(y, x) > 0$. Soit $N, k \geq 0$. On obtient alors

$$P^{m+n+Nk}(x, x) \geq P^n(x, y)P^k(y, y)^N P^m(y, x).$$

On choisit $k \geq 0$ tel que $P^k(y, y) > 0$. Alors $P^{m+n+Nk}(x, x) > 0$, donc $d_x \mid m + n + Nk$ et ceci pour tout $N \geq 0$. En appliquant ce résultat pour N et $N + 1$, on a

$$d_x \mid (m + n + [N + 1]k) - (m + n + Nk) = k$$

D'où $d_x \mid d_y$. De même, on montre que $d_y \mid d_x$ ce qui prouve l'égalité $d_x = d_y$. □

COROLLAIRE 7.18. Si la chaîne de MARKOV est irréductible, tous les états ont même période, appelée période de la chaîne. Si cette période est 1, on dit que la chaîne est apériodique.

◇ REMARQUE. Soient P un noyau de transition et $p \in]0, 1[$. On considère le noyau

$$P_p := (1 - p)P + pI \quad \text{avec} \quad I := (\delta_{x,y})_{x,y \in E}.$$

Le graphe des transitions de ce noyau P_p s'obtient à partir de celui du noyau P en rajoutant une boucle en chaque état $x \in E$ de poids p s'il n'y en a pas ou une rajoutant le poids p à la boucle existante s'il y en a déjà une (toutes autres transitions sont multipliées par $1 - p$). Cela permet à la chaîne de noyau P_p de rester en tout état x avec une probabilité strictement positive et la loi $P_p(x, \cdot)$ est le mélange des lois $P(x, \cdot)$ et δ_x par la loi de BERNOULLI de paramètre p . Comme $P_p(x, x) \geq p > 0$ pour tout $x \in E$, on en déduit que la chaîne de noyau P est apériodique et aussi irréductible si le noyau P l'est. De plus, un simple calcul montre que les noyaux P et P_p ont les mêmes mesures invariantes. On parle ici de *perturbation* car $P_p(x, y) \rightarrow P(x, y)$ lorsque $p \rightarrow 0$

7.3.2 Forte irréductibilité

DÉFINITION 7.19. Une chaîne de MARKOV de noyau de transition P est dite *fortement irréductible* s'il existe un entier $k \geq 1$ tel que, pour tous $x, y \in E$, on ait $P^k(x, y) > 0$.

◇ REMARQUE. L'irréductibilité, c'est lorsque qu'il existe des chemins finis entre tous couples de points, mais ces chemins ne sont pas nécessairement de la même longueur. Pour la forte irréductibilité, on impose que les longueurs de ces chemins soient majorées.

PROPOSITION 7.20. Une chaîne de MARKOV de noyau P fortement irréductible est irréductible et apériodique.

Preuve Il est immédiat que la forte irréductibilité implique l'irréductibilité. Soient $x, y \in E$ tels que $P(y, x) > 0$. Soit $k \geq 1$ un entier comme dans la définition précédent. Alors $P^k(x, x) > 0$, donc $P^{k+1}(x, x) > 0$. Cela montre qu'on a $k, k + 1 \in R(x) := \{n \geq 0 \mid P^n(x, x) > 0\}$. Or les entiers k et $k + 1$ sont premiers entre eux ce qui assure l'égalité $d_x = \text{pgcd } R(x) = 1$. Ceci montre l'apériodicité. □

PROPOSITION 7.21. On considère une chaîne de MARKOV irréductible et apériodique de noyau P . Alors pour tout $x \in E$, il existe un entier $n(x) \geq 1$ tel que, pour tout $n \geq n(x)$, on ait $P^n(x, x) > 0$.

Preuve Par l'irréductibilité, pour tous $x, y \in E$, il existe un entier $n(x, y) \geq 1$ tel que $P^{n(x,y)}(x, y) > 0$. Soit $x \in E$. Par l'apériodicité, on peut considérer des entiers $n_1, \dots, n_k \in R(x)$ dont le PGCD vaut 1. Par le théorème de BÉZOUT, il existe $q_1, \dots, q_k \in \mathbb{Z}^*$ tels que $q_1 n_1 + \dots + q_k n_k = 1$. Notons

$$a(x) := \sum_{i \in Q_+} q_i n_i \quad \text{et} \quad b(x) := \sum_{i \in Q_-} q_i n_i$$

avec $Q_+ := \{i \in \llbracket 1, k \rrbracket \mid q_i > 0\}$ et $Q_- := \{i \in \llbracket 1, k \rrbracket \mid q_i < 0\}$ de sorte que $a(x) = b(x) + 1$. Comme les entiers n_i appartiennent à $R(x)$, on a

$$P^{a(x)}(x, x) \geq \prod_{i \in Q_+} P^{n_i}(x, x)^{q_i} > 0 \quad \text{et} \quad P^{b(x)}(x, x) \geq \prod_{i \in Q_-} P^{n_i}(x, x)^{-q_i} > 0.$$

D'où $b(x), b(x) + 1 \in R(x)$. Soit $n \in \mathbb{N}$. Écrivons la division euclidienne de n par $b(x)$

$$n = qb(x) + r = (q - r)b(x) + ra(x)$$

avec $0 \leq r \leq b(x) - 1$. Posons $n(x) := b(x)^2 - 1$ et supposons $n \geq n(x)$. Comme

$$ra(x) \leq (b(x) - 1)(b(x) + 1) = b(x)^2 - 1 = n(x),$$

on doit avoir $q \geq r$. Comme $a(x), b(x) \in R(x)$, il vient alors

$$P^n(x, x) \geq P^{b(x)}(x, x)^{q-r} P^{a(x)}(x, x)^r > 0. \quad \square$$

PROPOSITION 7.22. Si E est fini, alors la forte irréductibilité est équivalente à l'irréductibilité plus l'apériodicité.

Preuve Le sens direct vient de la proposition 7.20. Réciproquement, on considère une chaîne de MARKOV irréductible et apériodique de noyau P . Montrons qu'elle est fortement irréductible. Avec les notations de la preuve précédente et comme E est fini, on peut considérer l'entier

$$k := \sup\{n(x) + n(x, y) \mid x, y \in E\} < +\infty.$$

Pour $x, y \in E$, en notant $k = n(x) + j + n(x, y)$ avec $j \geq 1$, on a alors

$$P^k(x, y) \geq P^{n(x)+j}(x, x) P^{n(x,y)}(x, y) > 0$$

ce qui prouve la forte irréductibilité. □

7.4 ÉQUILIBRE D'UNE CHAÎNE DE MARKOV

Dans cette section, on considère une chaîne de MARKOV $(X_n)_{n \geq 0}$ sur un espace d'états dénombrable E de noyau de transition P . Dans le cas fini, c'est la forte irréductibilité qui a joué un rôle crucial pour la convergence en loi de la chaîne. Dans le cas dénombrable général, c'est l'apériodicité qui joue un rôle similaire.

THÉORÈME 7.23 (convergence vers l'équilibre). On suppose que la chaîne est irréductible, récurrente positive et apériodique. Soit π l'unique probabilité invariante (qui existe d'après la proposition 7.12). Soit $x \in E$. Alors

$$\sum_{y \in E} |\mathbb{P}_x(X_n = y) - \pi(y)| \longrightarrow 0.$$

Il en va de même pour une loi initiale quelconque ν .

- ◊ REMARQUES. – Ce théorème implique que la convergence est uniforme par rapport aux états $y \in E$.
 – La convergence donnée énonce en fait la convergence en variation totale de la loi $\mathcal{L}(X_n | X_0 = x)$ vers la probabilité invariante π .

Preuve Sur l'espace $E \times E$, on considère la chaîne $(X_n^{(1)}, X_n^{(2)})_{n \geq 0}$ de noyau $\bar{P} := P \otimes P$, i. e.

$$\bar{P}((x_1, x_2), (y_1, y_2)) = P(x_1, y_1)P(x_2, y_2), \quad x_1, x_2, y_1, y_2 \in E.$$

Montrons que cette chaîne est irréductible. Soient $x_1, x_2, y_1, y_2 \in E$. Comme P est irréductible, il existe $m_1 \geq 1$ tel que $P^{m_1}(x_1, y_1) > 0$. Comme P est irréductible et apériodique, la proposition 7.21 assure qu'il existe $n_1 \geq 0$ tel que $P^{n_1+k}(x_1, x_2) > 0$ pour tout $k \geq 0$. De même, il existe $m_2, n_2 \geq 1$ tels que $P^{m_2}(x_2, y_2) > 0$ et $P^{n_2+k}(x_2, x_2) > 0$ pour tout $k \geq 0$. Ainsi pour tout $k \geq 0$, on a alors

$$\begin{aligned} P^{n_1+k+m_1}(x_1, y_1) &\geq P^{n_1+k}(x_1, x_1)P^{m_1}(x_1, y_1) > 0, \\ P^{n_2+k+m_2}(x_2, y_2) &\geq P^{n_2+k}(x_2, x_2)P^{m_2}(x_2, y_2) > 0. \end{aligned}$$

Cela assure alors

$$\forall n \geq \max(n_1 + m_1, n_2 + m_2), \quad \bar{P}^n((x_1, x_2), (y_1, y_2)) > 0.$$

Cela montre l'irréductibilité du noyau \bar{P} .

Par ailleurs, la mesure produit $\pi \otimes \pi$ est invariante pour la chaîne $(X_n^{(1)}, X_n^{(2)})_{n \geq 0}$ puisque la mesure π l'est pour la chaîne $(X_n)_{n \geq 0}$. Finalement, cette dernière chaîne est irréductible et admet une probabilité invariante, donc le théorème 7.13 assure qu'elle est récurrente positive.

Soit $n \geq 0$. Remarquons d'abord que, pour toutes probabilités ν_1 et ν_2 et tout état $x \in E$, on a

$$\bar{\mathbb{P}}_{\nu_1 \otimes \nu_2}(X_n^{(1)} = x) = \mathbb{P}_{\nu_1}(X_n^{(1)} = x) \quad \text{et} \quad \bar{\mathbb{P}}_{\nu_1 \otimes \nu_2}(X_n^{(2)} = x) = \mathbb{P}_{\nu_2}(X_n^{(2)} = x).$$

Soit $\Delta := \{(x, y) \in E^2 \mid x = y\}$ la diagonale de E^2 . La chaîne étant récurrente, la variable aléatoire

$$T_\Delta := \inf\{n \geq 0 \mid (X_n^{(1)}, X_n^{(2)}) \in \Delta\}$$

est un temps d'arrêt presque sûrement fini. Soient $x, y \in E$. Avec la remarque précédente, on a

$$\begin{aligned} \mathbb{P}_x(X_n = y) - \pi(y) &= \bar{\mathbb{P}}_{\pi \otimes \delta_x}(X_n^{(2)} = y) - \bar{\mathbb{P}}_{\pi \otimes \delta_x}(X_n^{(1)} = y) \\ &= \bar{\mathbb{E}}_{\pi \otimes \delta_x}[\mathbb{1}_{\{X_n^{(2)}=y\}} - \mathbb{1}_{\{X_n^{(1)}=y\}}]. \end{aligned}$$

En distinguant selon les valeurs du temps d'arrêt T_Δ , on obtient

$$\begin{aligned} \mathbb{P}_x(X_n = y) - \pi(y) &= \bar{\mathbb{E}}_{\pi \otimes \delta_x}[\mathbb{1}_{\{T_\Delta > n\}}(\mathbb{1}_{\{X_n^{(2)}=y\}} - \mathbb{1}_{\{X_n^{(1)}=y\}})] + \sum_{k=0}^n \bar{\mathbb{E}}_{\pi \otimes \delta_x}[\mathbb{1}_{\{T_\Delta=k\}}(\mathbb{1}_{\{X_n^{(2)}=y\}} - \mathbb{1}_{\{X_n^{(1)}=y\}})] \\ &= \bar{\mathbb{E}}_{\pi \otimes \delta_x}[\mathbb{1}_{\{T_\Delta > n\}}(\mathbb{1}_{\{X_n^{(2)}=y\}} - \mathbb{1}_{\{X_n^{(1)}=y\}})] + \sum_{k=0}^n \sum_{z \in E} \bar{\mathbb{E}}_{\pi \otimes \delta_x}[\mathbb{1}_{\{T_\Delta=k\}} \mathbb{1}_{\{X_k^{(1)}=X_k^{(2)}=z\}}(\mathbb{1}_{\{X_n^{(2)}=y\}} - \mathbb{1}_{\{X_n^{(1)}=y\}})]. \end{aligned} \quad (*)$$

Pour tout $k \in [0, n]$ et tout $z \in E$, en utilisant la propriété de MARKOV au temps k pour la loi initiale $\pi \otimes \delta_x$, on a

$$\begin{aligned} \bar{\mathbb{E}}_{\pi \otimes \delta_x}[\mathbb{1}_{\{T_\Delta=k\}} \mathbb{1}_{\{X_k^{(1)}=X_k^{(2)}=z\}} \mathbb{1}_{\{X_n^{(2)}=y\}}] &= \bar{\mathbb{E}}_{\pi \otimes \delta_x}[\bar{\mathbb{E}}_{\pi \otimes \delta_x}[\mathbb{1}_{\{T_\Delta=k\}} \mathbb{1}_{\{X_k^{(1)}=X_k^{(2)}=z\}} \mathbb{1}_{\{X_n^{(2)}=y\}} \mid \mathcal{F}_k]] \\ &= \bar{\mathbb{E}}_{\pi \otimes \delta_x}[\mathbb{1}_{\{T_\Delta=k\}} \mathbb{1}_{\{X_k^{(1)}=X_k^{(2)}=z\}} \bar{\mathbb{E}}_{\pi \otimes \delta_x}[\mathbb{1}_{\{X_n^{(2)}=y\}} \mid \mathcal{F}_k]] \\ &= \bar{\mathbb{E}}_{\pi \otimes \delta_x}[\mathbb{1}_{\{T_\Delta=k\}} \mathbb{1}_{\{X_k^{(1)}=X_k^{(2)}=z\}} \bar{\mathbb{E}}_{(X_k^{(1)}, X_k^{(2)})}[\mathbb{1}_{\{X_{n-k}^{(2)}=y\}}]] \\ &= \bar{\mathbb{E}}_{\pi \otimes \delta_x}[\mathbb{1}_{\{T_\Delta=k\}} \mathbb{1}_{\{X_k^{(1)}=X_k^{(2)}=z\}} \bar{\mathbb{E}}(z, z)[\mathbb{1}_{\{X_{n-k}^{(2)}=y\}}]] \end{aligned}$$

$$\begin{aligned}
&= \bar{\mathbb{E}}_{\pi \otimes \delta_x} [\mathbb{1}_{\{T_\Delta = k\}} \mathbb{1}_{\{X_k^{(1)} = X_k^{(2)} = z\}}] \bar{\mathbb{E}}_{(z, z)} [\mathbb{1}_{\{X_{n-k}^{(2)} = y\}}] \\
&= \bar{\mathbb{E}}_{\pi \otimes \delta_x} [\mathbb{1}_{\{T_\Delta = k\}} \mathbb{1}_{\{X_k^{(1)} = X_k^{(2)} = z\}}] P_{n-k}(z, y)
\end{aligned}$$

et, par symétrie, on en déduit

$$\bar{\mathbb{E}}_{\pi \otimes \delta_x} [\mathbb{1}_{\{T_\Delta = k\}} \mathbb{1}_{\{X_k^{(1)} = X_k^{(2)} = z\}} (\mathbb{1}_{\{X_n^{(2)} = y\}} - \mathbb{1}_{\{X_n^{(1)} = y\}})] = 0.$$

La relation (*) se réduit alors à l'égalité

$$\mathbb{P}_x(X_n = y) - \pi(y) = \bar{\mathbb{E}}_{\pi \otimes \delta_x} [\mathbb{1}_{\{T_\Delta > n\}} (\mathbb{1}_{\{X_n^{(2)} = y\}} - \mathbb{1}_{\{X_n^{(1)} = y\}})].$$

Il vient alors

$$\begin{aligned}
\sum_{y \in E} |\mathbb{P}_x(X_n = y) - \pi(y)| &= \sum_{y \in E} |\bar{\mathbb{E}}_{\pi \otimes \delta_x} [\mathbb{1}_{\{T_\Delta > n\}} (\mathbb{1}_{\{X_n^{(2)} = y\}} - \mathbb{1}_{\{X_n^{(1)} = y\}})]| \\
&\leq \sum_{y \in E} \bar{\mathbb{E}}_{\pi \otimes \delta_x} [\mathbb{1}_{\{T_\Delta > n\}} |\mathbb{1}_{\{X_n^{(2)} = y\}} - \mathbb{1}_{\{X_n^{(1)} = y\}}|] \\
&\leq \sum_{y \in E} \bar{\mathbb{E}}_{\pi \otimes \delta_x} [\mathbb{1}_{\{T_\Delta > n\}} (\mathbb{1}_{\{X_n^{(2)} = y\}} + \mathbb{1}_{\{X_n^{(1)} = y\}})] \\
&= \bar{\mathbb{E}}_{\pi \otimes \delta_x} [\mathbb{1}_{\{T_\Delta > n\}} \sum_{y \in E} (\mathbb{1}_{\{X_n^{(2)} = y\}} + \mathbb{1}_{\{X_n^{(1)} = y\}})] \\
&= 2 \bar{\mathbb{P}}_{\pi \otimes \delta_x} [T_\Delta > n] = 2 \bar{\mathbb{P}}_{\pi \otimes \delta_x} (T_\Delta > n).
\end{aligned}$$

Comme le temps d'arrêt T_Δ est $\mathbb{P}_{\pi \otimes \delta_x}$ -presque sûrement fini, le théorème de convergence monotone assure

$$\mathbb{P}_{\pi \otimes \delta_x} (T_\Delta > n) \longrightarrow 0$$

ce qui assure la conclusion

$$\sum_{y \in E} |\mathbb{P}_x(X_n = y) - \pi(y)| \longrightarrow 0.$$

Terminons la preuve. Soit ν une mesure de probabilité. Rappelons que $\mathbb{P}_\nu = \sum_{x \in E} \nu(x) \mathbb{P}_x$. Alors

$$\begin{aligned}
\sum_{y \in E} |\mathbb{P}_\nu(X_n = y) - \pi(y)| &= \sum_{z \in E} \left| \sum_{x \in E} [\mathbb{P}_x(X_n = y) - \pi(y)] \nu(x) \right| \\
&\leq \sum_{x \in E} \left(\sum_{y \in E} |\mathbb{P}_x(X_n = y) - \pi(y)| \right) \nu(x). \tag{**}
\end{aligned}$$

Appliquons le lemme 7.7 avec $a(x) = \nu(x)$ et

$$b_n(x) = \sum_{y \in E} |\mathbb{P}_x(X_n = y) - \pi(y)|$$

vérifiant bien $0 \leq b_n(x) \leq 1$ et $b_n(x) \longrightarrow 0$ par le premier point du théorème pour tous $n \geq 0$ et $x \in E$. On obtient ensuite la conclusion en passant à la limite dans l'inégalité (**). \square

◇ REMARQUE. L'apériodicité est essentielle sinon le couplage échoue en général. Par exemple, pour la chaîne à deux états, on a

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad P^{2n} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{et} \quad P^{2n+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad n \geq 0.$$

En supposant $X_0^{(1)} = 0$ et $X_0^{(2)} = 1$, on aura $X_n^{(1)} \neq X_n^{(2)}$ pour tout $n \geq 0$.

7.5 THÉORÈME ERGODIQUE

THÉORÈME 7.24 (ergodique). Soit $(X_n)_{n \geq 0}$ une chaîne de MARKOV récurrente et irréductible. Soient π une mesure invariante et $f, g \in L^1(\pi)$ telle que $\int_E g \, d\pi \neq 0$. Alors pour tout $x \in E$, on a

$$\frac{\sum_{k=0}^n f(X_k)}{\sum_{k=0}^n g(X_k)} \xrightarrow{-\mathbb{P}_x\text{-ps}} \frac{\int_E f \, d\pi}{\int_E g \, d\pi}.$$

COROLLAIRE 7.25 (ergodique). Soit $(X_n)_{n \geq 0}$ une chaîne de MARKOV récurrente positive et irréductible. Soient π son unique mesure invariante et $f \in L^1(\pi)$. Alors pour tout $x \in E$, on a

$$\frac{1}{n} \sum_{k=0}^n f(X_k) \xrightarrow{-\mathbb{P}_x\text{-ps}} \int_E f \, d\pi.$$

Preuve Comme la chaîne est irréductible et récurrente positive, il existe une unique probabilité invariante π par le théorème 7.13. Il suffit d'appliquer le théorème avec $g := 1 \in L^1(\pi)$. \square

◊ REMARQUE. Cette dernière limite est l'essence même de la notion d'ergodicité : la moyenne de la fonction f le long de la trajectoire de la chaîne, *i. e.* sa moyenne temporelle, converge en temps long vers sa moyenne spatiale (par rapport à la probabilité invariante).

COROLLAIRE 7.26 (ergodique). Soient $(X_n)_{n \geq 0}$ une chaîne de MARKOV récurrente irréductible et $x \in E$. Alors

$$\frac{1}{n} \sum_{k=0}^n \mathbb{1}_{\{X_k=x\}} = \begin{cases} \pi(x) & \text{si l'état } x \text{ est récurrent positif,} \\ 0 & \text{sinon} \end{cases} \quad \mathbb{P}_x\text{-presque sûrement.}$$

Preuve Il suffit d'appliquer le théorème avec $f := \delta_x$ et $g := 1$. \square

Preuve du théorème On commence par observer que $\pi(x) > 0$. En effet, les mesures π et ν_x sont proportionnelles. Comme $\nu_x(x) = 1$, on doit avoir $\pi(x) > 0$. Définissons les dates de retour

$$\begin{aligned} T_x^{(0)} &:= 0, \\ T_x^{(1)} &:= \inf\{k \geq 1 \mid X_k = x\}, \\ T_x^{(n+1)} &:= \inf\{k \geq T_x^{(n)} \mid X_k = x\}, \quad n \geq 1. \end{aligned}$$

Puisque l'état x est récurrent, les temps d'arrêt $T_x^{(n)}$ sont \mathbb{P}_x -presque fini d'après la propriété de MARKOV fort. Pour $k \geq 1$, on pose alors

$$Z_k(f) := \sum_{i=T_x^{(k-1)}}^{T_x^{(k)}} f(X_i).$$

Admettons provisoirement le lemme suivant.

LEMME 7.27. Les variables aléatoires $Z_k(f)$ sont indépendantes et identiquement distribuées. En particulier, pour $f = 1$, on retrouve la proposition 6.3.

Afin d'appliquer la loi des grands nombres à la suite $(Z_k(f))_{k \geq 1}$, montrons que les variables aléatoires $Z_k(f)$ sont intégrables. En effet, comme $\pi = \pi(x)\nu_x$, on a

$$\begin{aligned} \mathbb{E}_x[|Z_1(f)|] &\leq \mathbb{E}_x \left[\sum_{i=T_x^{(0)}}^{T_x^{(1)}} |f(X_i)| \right] \\ &= \mathbb{E}_x \left[\sum_{i=T_x^{(0)}}^{T_x^{(1)}} \sum_{y \in E} |f(y)| \mathbb{1}_{\{X_i=y\}} \right] \\ &= \sum_{y \in E} |f(y)| \mathbb{E}_x \left[\sum_{i=T_x^{(0)}}^{T_x^{(1)}} \mathbb{1}_{\{X_i=y\}} \right] = \sum_{y \in E} |f(y)| \nu_x(y) = \frac{1}{\pi(x)} \int_E |f| d\pi. \end{aligned}$$

Cela montre l'intégrabilité des variables aléatoires $Z_k(f)$. Le même calcul donne

$$\mathbb{E}[Z_1(f)] = \frac{1}{\pi(x)} \int_E f d\pi < +\infty.$$

On peut donc appliquer la loi des grands nombres et on obtient

$$\frac{1}{n} \sum_{k=0}^n Z_k(f) \xrightarrow{\mathbb{P}_x\text{-ps}} \frac{1}{\pi(x)} \int_E f d\pi. \quad (*)$$

Soit $n \geq 1$. On pose $N_x(n) := \sum_{k=1}^n \mathbb{1}_{\{X_k=x\}}$ de sorte que $T_x^{(N_x(n))} \leq n < T_x^{(N_x(n)+1)}$. Lorsque la fonction f est positive, en sommant cette dernière relation, on a

$$\frac{1}{N_x(n)} \sum_{k=1}^{T_x^{(N_x(n))}-1} f(X_k) \leq \frac{1}{N_x(n)} \sum_{k=0}^n f(X_k) \leq \frac{1}{N_x(n)} \sum_{k=0}^{T_x^{(N_x(n)+1)}-1} f(X_k)$$

ce qui, en regroupant les paquets $Z_j(f)$, devient

$$\frac{1}{N_x(n)} \sum_{j=1}^{N_x(n)} Z_j(f) \leq \frac{1}{N_x(n)} \sum_{k=0}^n f(X_k) \leq \frac{1}{N_x(n)} \sum_{j=1}^{N_x(n)+1} Z_j(f).$$

Comme l'état x est récurrent, on a $N_x(n) \rightarrow +\infty$ et la convergence (*) assure

$$\frac{1}{N_x(n)} \sum_{k=0}^n f(X_k) \xrightarrow{-\mathbb{P}_x\text{-ps}} \frac{1}{\pi(x)} \int_E f d\pi. \quad (**)$$

Lorsque la fonction f est de signe quelconque, on applique la convergence (**) aux fonctions f^- et f^+ et on obtient la même conclusion. De même avec la fonction g . D'où

$$\frac{\sum_{k=0}^n f(X_k)}{\sum_{k=0}^n g(X_k)} \xrightarrow{-\mathbb{P}_x\text{-ps}} \frac{\int_E f d\pi}{\int_E g d\pi}. \quad \square$$

Preuve du lemme 7.27 Par récurrence, montrons que, pour tout entier $k \geq 1$ et toutes fonctions mesurables bornées g_1, \dots, g_k sur \mathbf{R}_+ , on a

$$\mathbb{E}_x \left[\prod_{i=1}^k g_i(Z_i(f)) \right] = \prod_{i=1}^k \mathbb{E}_x [g_i(Z_1(f))]. \quad (*)$$

L'égalité (*) est immédiate pour $k = 1$. Soit $k \geq 1$. On suppose que la relation (*) est vérifiée au rang k pour toutes fonctions mesurables bornées g_1, \dots, g_k sur \mathbf{R}_+ . Observons d'abord que

- les variables aléatoires $Z_1(f), \dots, Z_k(f)$ sont $\mathcal{F}_{T_x^{(k)}}$ -mesurables;
- la variable aléatoire $\Theta_{T_x^{(k)}}$ est indépendante de la tribu $\mathcal{F}_{T_x^{(k)}}$ et de loi \mathbb{P}_x par la propriété de MARKOV forte;
- on a $Z_{k+1}(f) = Z_1(f) \circ \Theta_{T_x^{(k)}}$.

Soient g_1, \dots, g_k des fonctions mesurables bornées sur \mathbf{R}_+ . En utilisant la propriété de MARKOV forte, on obtient

$$\begin{aligned} \mathbb{E}_x \left[\prod_{i=1}^{k+1} g_i(Z_i(f)) \right] &= \mathbb{E}_x \left[\left(\prod_{i=1}^k g_i(Z_i(f)) \right) g_{k+1}(Z_1(f) \circ \Theta_{T_x^{(k)}}) \right] \\ &= \mathbb{E}_x \left[\left(\prod_{i=1}^k g_i(Z_i(f)) \right) \mathbb{E}_x [g_{k+1}(Z_1(f) \circ \Theta_{T_x^{(k)}}) \mid \mathcal{F}_{T_x^{(k)}}] \right] \\ &= \mathbb{E}_x \left[\left(\prod_{i=1}^k g_i(Z_i(f)) \right) \mathbb{E}_{X_{T_x^{(k)}}} [g_{k+1}(Z_1(f))] \right] \\ &= \mathbb{E}_x \left[\left(\prod_{i=1}^k g_i(Z_i(f)) \right) \right] \mathbb{E}_x [g_{k+1}(Z_1(f))] \end{aligned}$$

et l'hypothèse de récurrence permet de conclure la relation (**) au rang $k + 1$. Cela montre le lemme. \square