Université de Rennes

Processus aléatoires discrets

Conditionnement
Martingales
Chaînes de Markov

M1 Mathématiques fondamentales

Jean-Christophe Breton
Université de Rennes

Septembre–Décembre 2022

version du 28 novembre 2023

Table des matières

Introduction

Les suites de variables aléatoires indépendantes sont étudiées dans les cours de probabilités de niveau L3, comme par exemple [Bre-proba], avec comme résultats phares la loi des grands nombres (LGN) et le théorème central limite (TCL). Dans ces notes, on étudie des suites de variables qui ont une forme de dépendance : les martingales et les chaînes de Markov.

Pour cela, la notion de conditionnement est d'abord étudiée dans la partie I avec une approche élémentaire (Chapitre 1) et une approche plus générale fondée sur la notion d'espérance conditionnelle (Chapitre 2).

Dans la partie II, on introduit les martingales dans le Chapitre 3 et on en étudie le comportement asymptotique dans le Chapitre 4.

Les chaînes de Markov sont l'objet de la partie III. Dans le chapitre 5, on définit les chaînes de Markov, et on présente la propriété de Markov. La classification des états d'une chaîne de Markov est détaillée en Chapitre 6 et le régime invariant est étudié en Chapitre 7.

Ces notes ont de nombreuses sources d'inspiration, parmi lesquelles des notes de cours de Jürgen Angst et d'autres de Mihai Gradinaru. Des références à la fois pour les martingales et les chaînes de Markov sont : [BL, Bei, BEL, BC, FF, Kal, Ouv]. Des références pour les martingales sont : [JP, Wil]. Des références pour les chaînes de Markov sont : [Gra, HPS, Nor, Pri].

Rappels

Dans ce chapitre, on rappelle quelques résultats de théorie de la mesure (Section 0.1) et de probabilités (Section 0.2).

Les résultats sont cités ici sans preuve. On renvoie à tout cours de niveau Licence de Mathématiques pour une présentation plus détaillée, par exemple [BP] ou [Bre-Leb] pour la théorie de la mesure et [Ouv] ou [Bre-proba] pour les probabilités.

Dans toute la suite, $(X, \mathcal{A}, \mu)$ désigne un espace mesuré et $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité.

0.1 Rappels de théorie de la mesure

Classe monotone

Dans cette section, on rappelle l'argument standard de classe monotone.

Définition 0.1 (Classe monotone ou λ-système) *Une famille $\mathcal{M}$ de parties de X est appelée classe monotone si*

i) $X \in \mathcal{M}$;

ii) *$\mathcal{M}$ est stable par différence propre : lorsque $A, B \in \mathcal{M}$ et $B \subset A$, alors $A \setminus B \in \mathcal{M}$;*

iii) *$\mathcal{M}$ est stable par réunion dénombrable croissante ($A_j \in \mathcal{M}$, $j \geq 1$, $A_j \subset A_{j+1} \Rightarrow \bigcup_{j\geq 1} A_j \in \mathcal{M}$).*

La classe monotone engendrée par une partie $\mathcal{E}$ est la plus petite classe monotone $\mathcal{M}(\mathcal{E})$ contenant $\mathcal{E}$.

Théorème 0.2 (des classes monotones) *Soit $\mathcal{E}$ une famille de parties de X stable par intersection finie (ie. $\mathcal{E}$ est un π-système). Alors $\mathcal{M}(\mathcal{E}) = \sigma(\mathcal{E})$.*

En pratique, on utilise le résultat sous la forme suivante :

Corollaire 0.3 (Classes monotones) *Soit $\mathcal{M}$ une classe monotone contenant la famille de parties $\mathcal{E}$, stable par intersection finie (ie. $\mathcal{E}$ est un π-système). Alors $\sigma(\mathcal{E}) \subset \mathcal{M}$.*

Démonstration : Par le Th. 0.2, on a $\sigma(\mathcal{E}) = \mathcal{M}(\mathcal{E})$. Mais comme $\mathcal{M}$ est une classe monotone contenant $\mathcal{E}$ on a aussi $\mathcal{M}(\mathcal{E}) \subset \mathcal{M}$ par définition de $\mathcal{M}(\mathcal{E})$. Finalement,

$\sigma(\mathcal{E}) \subset \mathcal{M}$. $\square$

Une application fréquente du théorème des classes monotones est pour constuire des mesures par extension comme suit :

Théorème 0.4 (Dynkin) *Soit deux mesures finies μ_1 et μ_2 sur $(X, \mathcal{A})$ de même poids ($\mu_1(X) = \mu_2(X) < +\infty$), qui coïncident sur $\mathcal{C} \subset \mathcal{A}$, sous-famille stable par intersections finies (π-système) et qui engendre $\mathcal{A}$. Alors μ_1 et μ_2 sont égales sur $\mathcal{A}$.*

Une version analogue du Théorème 0.4 existe pour les mesures σ-finies. Cette version assure par exemple l'unicité de la mesure de Lebesgue en utilisant le π-système $\mathcal{C}$ donné par l'ensemble des intervalles de $\mathbb{R}$ et en observant que $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$.

Théorèmes de Fubini

On considère deux espaces mesurables $(X, \mathcal{A})$ et $(Y, \mathcal{B})$ et des mesures σ-finies μ sur $(X, \mathcal{A})$ et ν sur $(Y, \mathcal{B})$. On rappelle que $\mu \otimes \nu$ (mesure produit) désigne l'unique mesure sur $X \times Y$ (espace produit) muni de $\mathcal{A} \otimes \mathcal{B} = \sigma(A \times B : A \in \mathcal{A}, B \in \mathcal{B})$ (tribu produit) qui étend la définition suivante :

$$(\mu \otimes \nu)(A \times B) = \mu(A)\nu(B), \quad A \in \mathcal{A}, B \in \mathcal{B}.$$

Comme $\mathcal{M} = \{A \times B : A \in \mathcal{A}, B \in \mathcal{B}\}$ est stable par intersection finie (π-système), le théorème de Dynkin (Th. 0.4), version σ-finie, assure l'unicité de la mesure $\mu \otimes \nu$ sur la tribu produit $\mathcal{A} \otimes \mathcal{B}$ (pour l'existence, il y a plus de travail).

Théorème 0.5 (Fubini-Tonelli et Fubini) *Si f est $(\mathcal{A} \otimes \mathcal{B})$-mesurable et positive (Fubini-Tonelli) ou $(\mu \otimes \nu)$-intégrable (Fubini) alors*

$$\int_{X \times Y} f(x,y)\,(\mu \otimes \nu)(dx, dy) = \int_X \Big(\int_Y f(x,y)\,\nu(dy) \Big)\,\mu(dx) = \int_Y \Big(\int_X f(x,y)\,\mu(dx) \Big)\,\nu(dy). \tag{F}$$

Théorème de Radon-Nikodym

On considère maintenant deux mesures μ, ν sur le même espace mesurable $(X, \mathcal{A})$. On rappelle que ν est absolument continue par rapport μ ($\nu \ll \mu$) lorsque $\mu(A) = 0$ entraîne $\nu(A) = 0$.

Théorème 0.6 (Radon-Nikodym) *Si $\nu \ll \mu$ alors il existe une fonction mesurable $f = \frac{d\nu}{d\mu}$ appelée dérivée de Radon-Nikodym telle que*

$$\nu(A) = \int_A f\,d\mu, \quad A \in \mathcal{A}. \tag{RN1}$$

De plus, si g est une fonction mesurable positive ou dans $L^1(X, \mathcal{A}, \nu)$

$$\int g\,d\nu = \int g f\,d\mu. \tag{RN2}$$

0.2 Rappels probabilistes

On rappelle que pour une variable aléatoire $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, la tribu engendrée par X est $\sigma(X) = \sigma\big(X^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\big)$. On rappelle aussi le résultat suivant fort utile :

Théorème 0.7 (Doob-Dynkin) *Une variable aléatoire Y est $\sigma(X)$-mesurable si et seulement s'il existe $h : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable telle que $Y = h(X)$.*

Démonstration : $\boxed{\Leftarrow}$ Le sens indirect est immédiat par composition d'applications mesurables : si $Y = h(X)$ alors

$$Y^{-1}(B) = X^{-1}(h^{-1}(B)) \in \sigma(X)$$

puisque $h^{-1}(B) \in \mathcal{B}(\mathbb{R})$ par mesurabilité de h.

$\boxed{\Rightarrow}$ Pour le sens direct, si $Y = \mathbf{1}_A$ est $\sigma(X)$-mesurable alors $A \in \sigma(X)$ est de la forme $A = X^{-1}(B)$ avec $B \in \mathcal{B}(\mathbb{R})$ et $Y = \mathbf{1}_{X^{-1}(B)} = \mathbf{1}_B(X)$ est de la forme requise avec $h = \mathbf{1}_B$ est mesurable. Si $Y = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$ est simple positive alors $A_i \in \sigma(X)$ et d'après le cas précédent $A_i = X^{-1}(B_i)$ et on a $Y = \sum_{i=1}^n \alpha_i \mathbf{1}_{B_i}(X)$ de la forme requise avec $h = \sum_{i=1}^n \alpha_i \mathbf{1}_{B_i}$. Si Y est $\sigma(X)$-mesurable positive alors $Y = \sup_{n\geq 1} Y_n$ avec $(Y_n)_{n\geq 1}$ suite croissante de variables aléatoires simples positives. D'après le cas précédent, $Y_n = h_n(X)$ avec h_n mesurable et alors $Y = h(X)$ avec $h = \sup_{n\geq 1} h_n$ mesurable, en tant que sup des fonctions h_n mesurables. Enfin, si Y est $\sigma(X)$-mesurable de signe quelconque alors on applique le cas précédent à $Y^+ = \max(Y, 0)$ et à $Y^- = \max(-Y, 0)$ variables aléatoires qui s'écrivent alors $Y^+ = h_1(X)$ et $Y^- = h_2(X)$. On pose alors $h(x) = h_1(x) - h_2(x)$ si $x \in \mathcal{S}(X)$ (support de X) et $h(x) = 0$ sinon. La fonction h est mesurable car h_1, h_2 le sont et $\mathcal{S}(X) \in \mathcal{B}(\mathbb{R})$. Noter que comme on n'a pas simultanément $Y^+ > 0$ et $Y^- > 0$ alors on a $h_1(x) = h_2(x) = +\infty$ pour aucun $x \in \mathcal{S}(X)$ et $h(x)$ est bien définie pour tout $x \in \mathbb{R}$. On a alors

$$Y = Y^+ - Y^- = h_1(X) - h_2(X) = h(X),$$

avec h fonction mesurable. □

Indépendances

Définition 0.8 (Indépendances)

- *Deux évènements A et B sont indépendants si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. On note $A \perp\!\!\!\perp B$.*
- *Deux tribus $\mathcal{A}$ et $\mathcal{B}$ sont indépendantes si pour tout $A \in \mathcal{A}$ et $B \in \mathcal{B}$ on a $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. On note alors $\mathcal{A} \perp\!\!\!\perp \mathcal{B}$.*
- *Deux variables aléatoires X, Y sont indépendantes si les tribus qu'elles engendrent le sont : $\sigma(X) \perp\!\!\!\perp \sigma(Y)$.*
- *On dit que des variables aléatoires X_i, $i \in I$, sont mutuellement indépendantes si pour tout $k \geq 1$ et $i_1, \dots, i_k$ distincts dans I, $B_{i_1}, \dots, B_{i_k} \in \mathcal{B}(\mathbb{R})$:*

$$\mathbb{P}\big(X_{i_1} \in B_{i_1}, \dots, X_{i_k} \in B_{i_k}\big) = \mathbb{P}(X_{i_1} \in B_{i_1}) \dots \mathbb{P}(X_{i_k} \in B_{i_k}).$$

— *On dit que des variables aléatoires X_i, $i \in I$, sont deux à deux indépendantes lorsque pour tout couple d'indice i, j distincts dans I, on a $X_i \perp\!\!\!\perp X_j$.*

L'indépendance mutuelle implique l'indépendance deux à deux mais la réciproque est fausse comme le montre l'exemple suivant : on considère une urne avec 4 boules, une bleue, une blanche, une rouge et une tricolore et on fait des tirages successifs avec remise. On note alors

— A : on tire une boule avec du bleue;
— B : on tire une boule avec du blanche;
— C : on tire une boule avec du rouge.

On observe aisément que $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = 1/2$, $\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = 1/4$ et $\mathbb{P}(A \cap B \cap C) = 1/4$ si bien que les évènements A, B, C sont deux à deux indépendants mais pas mutuellement indépendants. De même pour les variables aléatoires $\mathbf{1}_A$, $\mathbf{1}_B$, $\mathbf{1}_C$.

Variables et vecteurs gaussiens

Définition 0.9 (Variable aléatoire gaussienne (normale)) *Une variable aléatoire réelle X suit la loi normale standard $\mathcal{N}(0,1)$ si elle admet pour densité*

$$x \mapsto \frac{1}{\sqrt{2\pi}} \exp(-x^2/2).$$

De façon générale, si $m \in \mathbb{R}$ et $\sigma^2 > 0$, une variable aléatoire réelle X suit la loi normale $\mathcal{N}(m, \sigma^2)$ si elle admet pour densité

$$x \mapsto \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

Si $\sigma^2 = 0$, la loi est dégénérée et la variable aléatoire X est constante égale à m. Sa loi est un Dirac en m : $\mathbb{P}_X = \delta_m$.

On rappelle que $\int_{-\infty}^{+\infty} e^{-x^2/2} dx = \sqrt{2\pi}$ justifie la normalisation de la loi $\mathcal{N}(0,1)$. Par ailleurs, rappelons qu'une variable aléatoire $X \sim \mathcal{N}(m, \sigma^2)$ peut se voir comme la translatée et dilatée d'une variable aléatoire X_0 de loi normale standard $\mathcal{N}(0,1)$ par

$$X = m + \sigma X_0.$$

Autrement dit si $X \sim \mathcal{N}(m, \sigma^2)$, $\sigma^2 > 0$, on définit la variable centrée réduite $\widetilde{X} = (X - m)/\sigma$, de loi $\mathcal{N}(0,1)$. Rappelons également qu'une variable aléatoire X de loi $\mathcal{N}(m, \sigma^2)$ a pour

— espérance : $\mathbb{E}[X] = m$;
— variance : $\mathrm{Var}(X) = \sigma^2$;
— fonction caractéristique : $\varphi_X(t) = \mathbb{E}[e^{itX}] = \exp\big(imt - \sigma^2 t^2/2\big)$.

Proposition 0.10 *Soit* $X_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ *et* $X_2 \sim \mathcal{N}(m_2, \sigma_2^2)$ *indépendantes. Alors* $X_1 + X_2 \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.

Dans la suite, pour simplifier la présentation, on note sous la forme de transposée de vecteurs lignes les vecteurs colonnes : $X = (X_1, \dots, X_d)^t$. On considère le produit scalaire euclidien : pour $x = (x_1, \dots, x_d)^t, y = (y_1, \dots, y_d)^t \in \mathbb{R}^d$, on a $\langle x, y\rangle = x^t y = \sum_{i=1}^d x_i y_i$.

On décrit maintenant la version multidimensionnelle des variables normales.

Définition 0.11 (Vecteur gaussien) *Un vecteur aléatoire* $X = (X_1, \dots, X_d)^t$ *est gaussien si et seulement si toutes les combinaisons linéaires de ses coordonnées* $\langle a, X\rangle = a_1 X_1 + \dots + a_d X_d$ *suit une loi gaussienne dans* $\mathbb{R}$ *(pour tout* $a = (a_1, \dots, a_d)^t \in \mathbb{R}^d$*).*

Pour un vecteur gaussien $X = (X_1, \dots, X_d)^t$, tous les moments sont définis et on appelle

- espérance de X le vecteur $\mathbb{E}[X] = \big(\mathbb{E}[X_1], \dots, \mathbb{E}[X_d]\big)^t$;
- matrice de covariance de X la matrice carrée symétrique, positive

$$K = \big(\operatorname{Cov}(X_i, X_j)\big)_{1 \le i,j \le d}.$$

On observe facilement que la loi de X est caractérisée par m et K et on note $X \sim \mathcal{N}(m, K)$ sa loi. Si $\mathbb{E}[X] = 0$, le vecteur X est dit centré.

Proposition 0.12 *Soit* $X \sim \mathcal{N}_d(m, K)$ *un vecteur gaussien de dimension* d *et* $A \in \mathcal{M}_{p,d}(\mathbb{R})$ *alors* $AX \sim \mathcal{N}_p(Am, AKA^t)$.

Proposition 0.13 (Vecteurs gaussiens et indépendance)

(1) Soit (X, Y) *un couple gaussien. Alors* X *et* Y *sont indépendantes si et seulement si* $\operatorname{Cov}(X, Y) = 0$.

(2) Soit $(X_1, \dots, X_{d_1}, Y_1, \dots, Y_{d_2})^t$ *un vecteur gaussien de dimension* $d_1 + d_2$*. Les deux vecteurs aléatoires gaussiens* $X = (X_1, \dots, X_{d_1})^t$ *et* $Y = (Y_1, \dots, Y_{d_2})^t$ *sont indépendants si et seulement si les covariances* $\operatorname{Cov}(X_i, Y_j)$, $1 \le i \le d_1, 1 \le j \le d_2$, *sont toutes nulles.*

Première partie

Conditionnement

Chapitre 1

Conditionnement discret

La notion de conditionnement est essentielle dans la suite du cours pour définir les martingales (Chapitre 3) et les chaînes de Markov (Chapitre 5). On introduit cette notion dans ce chapitre dans un cadre élémentaire discret. L'approche plus générale sera l'objet du Chapitre 2. On considère un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

1.1 Probabilité conditionnelle discrète

Probabilité sachant un évènement

On commence par le cas très simple du conditionnement par un évènement [1] non négligeable :

Définition 1.1 (Probabilité conditionnelle) *Soit B un évènement de probabilité non nulle $\mathbb{P}(B) \neq 0$. Pour tout évènement A, on définit la probabilité conditionnelle de A sachant B par*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \tag{1.1}$$

L'intérêt de cette notion vient du fait que souvent, compte tenu des informations disponibles dans un modéle probabiliste, il peut être plus facile d'attribuer une valeur à la probabilité conditionnelle $\mathbb{P}(A|B)$ que de calculer $\mathbb{P}(A \cap B)$ ou $\mathbb{P}(A)$. Si $A \perp\!\!\!\perp B$, évidemment, on a $\mathbb{P}(A|B) = \mathbb{P}(A)$, ie. le conditionnement par B est sans effet.

En fait, la probabilité conditionnelle est une probabilité :

Proposition 1.2 *Soit $B \in \mathcal{F}$ avec $\mathbb{P}(B) > 0$. La fonction d'ensemble $\mathbb{P}(*|B) : A \in \mathcal{F} \mapsto \mathbb{P}(A|B)$ est une nouvelle probabilité sur $(\Omega, \mathcal{F})$.*

Démonstration : Il est clair que $\mathbb{P}(A|B) \geq 0$, $\mathbb{P}(\emptyset|B) = 0$. La σ-additivité découle de celle de $\mathbb{P}$: soit $(A_i)_{i\geq 1}$ une suite d'évènements deux à deux disjoints, on a

$$\mathbb{P}\Big(\bigcup_{i=1}^{+\infty} A_i \Big| B\Big) = \frac{\mathbb{P}\Big(\Big(\bigcup_{i=1}^{+\infty} A_i\Big) \cap B\Big)}{\mathbb{P}(B)} = \frac{\mathbb{P}\Big(\bigcup_{i=1}^{+\infty} (A_i \cap B)\Big)}{\mathbb{P}(B)}$$

1. sic

$$= \sum_{i=1}^{+\infty} \frac{\mathbb{P}\big(A_i \cap B\big)}{\mathbb{P}(B)} = \sum_{i=1}^{+\infty} \mathbb{P}(A_i|B).$$

Il en résulte que $\mathbb{P}(*|B)$ est bien une mesure. Il s'agit d'une probabilité puisque $\mathbb{P}(\Omega|B) = \mathbb{P}(\Omega \cap B)/\mathbb{P}(B) = \mathbb{P}(B)/\mathbb{P}(B) = 1$. □

Propriétés des probabilités sachant des évènements

La Prop.1.2 assure que l'on dispose pour les probabilités conditionnelles de toutes les propriétés habituelles d'une probabilité. En plus, on a les propriétés spécifiques suivantes :

Proposition 1.3 (Règle des conditionnements successifs) *Soit n évènements $A_1, \dots, A_n$ tels que $\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_{n-1}) \neq 0$. Alors*

$$\begin{aligned}
&\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) \\
&\quad = \mathbb{P}(A_1)\,\mathbb{P}(A_2|A_1)\,\mathbb{P}(A_3|A_1 \cap A_2) \times \dots \times \mathbb{P}(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}). \qquad (1.2)
\end{aligned}$$

Démonstration : Il suffit d'utiliser la définition (1.1) pour chaque probabilité conditionnelle et de simplifier. □

Enchaîner des conditionnements est équivalent à conditionner par l'intersection :

Proposition 1.4 (Conditionnement en cascade) *Étant donné des évènements A, B, C avec $\mathbb{P}(B \cap C) > 0$, en notant $\mathbb{P}_C = \mathbb{P}(\cdot|C)$, on a $\mathbb{P}_C(A|B) = \mathbb{P}(A|B \cap C)$.*

Démonstration :

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

□

Dans la suite, on utilise $I \subset \mathbb{N}$ pour désigner un ensemble dénombrable. Celui-ci peut être fini $I = \{1, \dots, n\}$ ou infini $I = \mathbb{N}$.

Définition 1.5 (Système complet) *On appelle* **système complet** *d'évènements toute suite dénombrable $(B_i)_{i \in I}$ d'évènements deux à deux disjoints et dont la somme des probabilités vaut 1 :*

$$\sum_{i \in I} \mathbb{P}(B_i) = 1.$$

Le système est dit fini si I est fini, infini si I est infini.

Proposition 1.6 (Formule des probabilités totales) *Étant donné* $(B_i)_{i\in I}$ *un système complet dénombrable de* Ω *avec* $\mathbb{P}(B_i) > 0$ *pour tout* $i \in I$*, pour tout* $A \in \mathcal{F}$ *on a*

$$\mathbb{P}(A) = \sum_{i\in I} \mathbb{P}(A|B_i)\,\mathbb{P}(B_i). \tag{1.3}$$

Démonstration : Notons $\Omega_0 = \bigsqcup_{i\in I} B_i$. Comme les B_i, $i \in I$, forment un système complet, on a $\mathbb{P}(\Omega_0) = \sum_{i\in I} \mathbb{P}(B_i) = 1$. Dès lors comme les $(A\cap B_i)$, $i \in I$, sont disjoints

$$\mathbb{P}(A) = \mathbb{P}(A\cap\Omega_0) = \mathbb{P}\Big(\bigcup_{i\in I}(A\cap B_i)\Big) = \sum_{i\in I}\mathbb{P}(A\cap B_i) = \sum_{i\in I}\mathbb{P}(A|B_i)\,\mathbb{P}(B_i).$$

□

Lorsque l'on sait calculer les probabilités conditionnelles $\mathbb{P}(A|B_i)$ pour tout un système de partition $(B_i)_{i\in I}$, on peut chercher les probabilités conditionnelles avec les conditionnements inverses $\mathbb{P}(B_i|A)$. Elles sont données par :

Proposition 1.7 (Formule de Bayes[2]) *Étant donné* $(B_i)_{i\in I}$ *un système complet de* Ω *avec* $\mathbb{P}(B_i) > 0$ *pour tout* $i \in I$*, pour tout évènement* A *de probabilité non nulle, on a :*

$$\forall j \in I, \quad \mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j)\,\mathbb{P}(B_j)}{\sum_{i\in I}\mathbb{P}(A|B_i)\,\mathbb{P}(B_i)}. \tag{1.4}$$

Démonstration : Pour tout $j \in I$, on a :

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(B_j\cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|B_j)\,\mathbb{P}(B_j)}{\sum_{i\in I}\mathbb{P}(A|B_i)\ \mathbb{P}(B_i)}$$

en utilisant la formule des probabilités totales (1.3) au dénominateur. □

Cette formule toute simple est à l'origine de tout un pan des statistiques qui consiste à inverser des conditionnements en manipulant des probabilités dites *a priori* ou *a posteriori*, il s'agit des *statistiques bayésiennes*.

Probabilité sachant une variable aléatoire discrète

En prenant $B = \{Y = y\}$ où Y est une variable aléatoire discrète et y un de ses atomes, (1.1) donne un sens à

$$\mathbb{P}(A|Y = y) = \frac{\mathbb{P}(A, Y = y)}{\mathbb{P}(Y = y)}$$

(souvent A prend même la forme $A = \{X = x\}$ lorsque X est une variable aléatoire discrète). On peut aussi définir la probabilité conditionnelle « sachant Y » plutôt que « sachant $Y = y$ » on ne conditionne alors plus par un évènement du type $\{Y = y\}$ mais par une variable aléatoire, ainsi $\mathbb{P}(A|Y)$ définit une nouvelle variable aléatoire !

Définition 1.8 (Probabilité conditionnelle discrète) *Étant donné une variable aléatoire discrète Y de support $\mathcal{S}(Y) = \{y_j : j \in J\}$, on appelle probabilité conditionnelle sachant Y la fonction d'ensemble*

$$\mathbb{P}(*|Y) : \left\{ \begin{array}{ccl} \mathcal{F} & \to & [0,1]^\Omega \\ A & \mapsto & \sum_{j\in J} \mathbb{P}(A|Y=y_j)\mathbf{1}_{\{Y=y_j\}}. \end{array} \right.$$

Ainsi $\mathbb{P}(A|Y) = \mathbb{P}(A|Y=y)$ sur l'évènement $\{Y=y\}$.

Dans le cas général, par exemple lorsque Y est une variable aléatoire à densité, la définition de $\mathbb{P}(*|Y)$ est plus compliquée car les conditionnements par $\{Y=y\}$ sont singuliers (évènements négligeables) et la définition (1.1) ne s'applique pas.

1.2 Espérance conditionnelle discrète

Étant donné un évènement B non négligeable, on définit l'espérance conditionnelle sachant B comme l'espérance par rapport à la probabilité $\mathbb{P}(*|B)$:

Définition 1.9 (Espérance conditionnelle élémentaire) *L'espérance conditionnelle sachant B d'une variable aléatoire X positive ou $X \in L^1$ est définie par*

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

On a

Proposition 1.10 *Soit X une variable aléatoire intégrable et $B \in \mathcal{F}$ non-négligeable. Alors on a*

$$\mathbb{E}[X|B] = \frac{\mathbb{E}[X\mathbf{1}_B]}{\mathbb{P}(B)}. \tag{1.5}$$

Démonstration : Pour $X = \mathbf{1}_A$, il s'agit de la définition de $\mathbb{P}(A|B)$ en (1.1). Le résultat s'étend alors par linéarité à $X = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$ étagée positive ($\alpha_i \geq 0$) puis par convergence monotone à $X \geq 0$. On traite le cas de X variable aléatoire de signe quelconque (intégrable) en écrivant $X = X^+ - X^-$ et en appliquant le cas positif précédent à X^+ et à X^-, en notant $X^+ = \max(X,0)$ et $X^- = \max(-X,0)$. La différence $\mathbb{E}[X^+\mathbf{1}_B] - \mathbb{E}[X^-\mathbf{1}_B]$ a bien un sens car X est intégrable. □

Si X est une variable aléatoire réelle discrète de support $\mathcal{S}(X) = \{x_i : i \in I\}$, alors $X = \sum_{i\in I} x_i \mathbf{1}_{\{X=x_i\}}$ et avec (1.5) on a

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

Dans le cas où Y est une autre variable aléatoire discrète de support $\mathcal{S}(Y) = \{y_j : j \in J\}$ (avec donc $J \subset \mathbb{N}$), on définit de cette façon

$$\mathbb{E}[X|Y = y_j] = \sum_{i\in I} x_i\, \mathbb{P}(X = x_i|Y = y_j), \tag{1.6}$$

et comme on l'a fait en Déf. 1.8 pour une probabilité conditionnelle, on peut généraliser l'espérance conditionnelle « sachant $Y = y_j$ » à « sachant Y » par :

Définition 1.11 (Espérance conditionnelle discrète) *Soit X une variable aléatoire intégrable et Y une variable aléatoire discrète. L'espérance conditionnelle de X sachant Y est définie par*

$$\mathbb{E}[X|Y] = \sum_{j\in J} \mathbb{E}[X|Y = y_j]\, \mathbf{1}_{\{Y=y_j\}}, \tag{1.7}$$

ie. $\mathbb{E}[X|Y] = \mathbb{E}[X|Y = y_j]$ *sur l'évènement* $\{Y = y_j\}$.

Il faut bien comprendre que l'espérance conditionnelle $\mathbb{E}[X|Y = y_j]$ en (1.6) est un réel alors que l'espérance conditionnelle $\mathbb{E}[X|Y]$ en (1.7) est une variable aléatoire.

En combinant (1.6) et (1.7), on a aussi

$$\mathbb{E}[X|Y] = \sum_{(i,j)\in I\times J} x_i\, \mathbb{P}(X = x_i|Y = y_j)\mathbf{1}_{\{Y=y_j\}}. \tag{1.8}$$

Due à la Définition 1.9 qui assure $\mathbb{E}[\mathbf{1}_A|B] = \mathbb{P}(A|B)$ pour un évènement B non-négligeable, il est facile de vérifier, lorsque Y est discrète, que l'espérance conditionnelle $\mathbb{E}[*|Y]$ en Définition 1.11 et la probabilité conditionnelle $\mathbb{P}(*|Y)$ en Définition 1.8 sont naturellement liées par

$$\mathbb{E}[\mathbf{1}_A|Y] = \mathbb{P}(A|Y).$$

Exemple 1.12

— Lorsque Y est une variable constante (presque sûrement) alors $\mathbb{E}[X|Y] = \mathbb{E}[X]$ ps. En effet en notant y l'unique atome de Y, comme $\{Y = y\}$ est un évènement presque sûr, $\mathbb{P}(X = x_i|Y = y) = \mathbb{P}(X = x_i)$ et $\mathbf{1}_{\{Y=y\}} = 1$ ps si bien que (1.8) se réduit à

$$\begin{aligned}\mathbb{E}[X|Y] &= \sum_{i\in I} x_i\, \mathbb{P}(X = x_i|Y = y)\mathbf{1}_{\{Y=y\}}\\ &= \sum_{i\in I} x_i\, \mathbb{P}(X = x_i) = \mathbb{E}[X] \text{ ps.}\end{aligned}$$

Le même résultat reste vrai lorsque $X \perp\!\!\!\perp Y$.

— Si X est $\sigma(Y)$-mesurable, alors $\mathbb{E}[X|Y] = X$. En effet, d'après le Théorème de Doob-Dynkin (Th. 0.7), on a $X = h(Y)$ pour une fonction h mesurable et donc

X est discrète avec pour valeurs $h(y_j)$, $j \in J$ (mais possiblement avec des répétitions). On a la partition $I = \bigsqcup_{i\in I} J_i$ où $J_i = \{j \in J : h(y_j) = x_i\}$, $i \in I$, et $\{X = x_i\} = \bigsqcup_{j\in J_i}\{Y = y_j\}$. Par (1.8), on a :

$$\begin{aligned}
\mathbb{E}[X|Y] &= \sum_{(i,j)\in I\times J} x_i\,\mathbb{P}(X = x_i|Y = y_j)\mathbf{1}_{\{Y=y_j\}} \\
&= \sum_{(i,j)\in I\times J} x_i\,\mathbb{P}\Big(\bigcup_{j'\in J_i}\{Y = y_{j'}\}|Y = y_j\Big)\mathbf{1}_{\{Y=y_j\}} \\
&= \sum_{(i,j)\in I\times J} x_i \sum_{j'\in J_i} \underbrace{\mathbb{P}(Y = y_{j'}|Y = y_j)}_{=\delta_{j,j'}}\mathbf{1}_{\{Y=y_j\}} \\
&= \sum_{(i,j)\in I\times J_i} x_i\,\mathbf{1}_{\{Y=y_j\}} = \sum_{i\in I} x_i \underbrace{\sum_{j\in J_i}\mathbf{1}_{\{Y=y_j\}}}_{=\mathbf{1}_{\{X=x_i\}}} \\
&= \sum_{i\in I} X\mathbf{1}_{\{X=x_i\}} = X.
\end{aligned}$$

Avec les définitions données, on vérifie sans difficulté que $\mathbb{E}\big[\mathbb{E}[X|Y]\big] = \mathbb{E}[X]$. En effet, par linéarité de l'espérance et par (1.5), on a

$$\mathbb{E}\big[\mathbb{E}[X|Y]\big] = \sum_{j\in J}\mathbb{E}[X|Y = y_j]\mathbb{P}(Y = y_j) = \sum_{j\in J}\mathbb{E}[X\mathbf{1}_{\{Y=y_j\}}]) = \mathbb{E}[X]$$

puisque $\sum_{j\in J}\mathbf{1}_{\{Y=y_j\}} = 1$ ps, $\{y_j : j \in J\}$ étant le support de Y. Plus généralement, on a la propriété de conditionnements en cascade :

Proposition 1.13 (Conditionnements en cascade) *Soit X, Y, Z des variables aléatoires discrètes. On a*

$$\mathbb{E}[X|Y] = \mathbb{E}\big[\mathbb{E}[X|Y, Z]\,|Y\big]. \tag{1.9}$$

Démonstration : On note $\mathcal{S}(X) = \{x_i : i \in I\}$, $\mathcal{S}(Y) = \{y_j : j \in J\}$, $\mathcal{S}(Z) = \{z_k : k \in K\}$ les supports discrets de X, Y, Z. Comme (Y, Z) est un vecteur discret, l'expression (1.8) s'écrit

$$\mathbb{E}[X|Y, Z] = \sum_{(i,j,k)\in I\times J\times K} x_i\,\mathbb{P}(X = x_i|Y = y_j, Z = z_k)\mathbf{1}_{\{Y=y_j,Z=z_k\}}.$$

La variable aléatoire $U = \mathbb{E}[X|Y, Z]$ prend les valeurs $u_{j,k} = \sum_{i\in I} x_i\,\mathbb{P}(X = x_i|Y = y_j, Z = z_k)$, $(j, k) \in J \times K$. Comme il y a possiblement des répétitions, on réindexe en notant $\mathcal{U} = \{u_\ell : \ell \in L\} = \{u_{j,k} : (j, k) \in J \times K\}$ avec $L \subset \mathbb{N}$. Pour $\ell \in L$, on note $A_\ell = \{(j, k) \in J \times K : u_{j,k} = u_\ell\}$. On a $J \times K = \bigsqcup_{\ell\in L} A_\ell$ car $u_\ell \neq u_{\ell'}$ pour $\ell \neq \ell'$

entraîne $A_\ell \cap A_{\ell'} = \emptyset$ et comme on a $u_{j,k} \in \mathcal{U}$ pour tout (j,k), l'union fait bien $J \times K$. L'expression (1.8) donne encore

$$\begin{aligned} \mathbb{E}\big[\mathbb{E}[X|Y,Z]\,|Y\big] &= \mathbb{E}[U|Y] \\ &= \sum_{\ell\in L, j\in J} u_\ell\, \mathbb{P}(U = u_\ell | Y = y_j) \mathbf{1}_{\{Y=y_j\}}. \end{aligned} \tag{1.10}$$

Sachant $Y = y_j$, nécessairement U ne peut prendre comme valeurs que $u \in \mathcal{U}_j = \{u_{j,k} : k \in K\}$ et cela exige $\ell \in L_j = \{\ell \in L : u_\ell \in \mathcal{U}_j\}$. Comme $\bigsqcup_{j\in J} L_j = L$, (1.10) s'écrit.

Dans ce cas pour $\ell \in L_j$, avoir $U = u_\ell$ sachant $Y = y_j$ est équivalent à avoir $Z = z_k$ pour $k \in K(\ell, j) := \{k \in K : u_{j,k} = u_\ell\}$:

$$\{U = u_\ell\} \cap \{Y = y_j\} = \Big(\bigsqcup_{k\in K(\ell,j)} \{Z = z_k\}\Big) \cap \{Y = y_j\}.$$

On note que $K = \bigsqcup_{\ell\in L, j\in J} K(\ell, j)$. Il s'ensuit pour (1.10)

$$\begin{aligned}
&\mathbb{E}\big[\mathbb{E}[X|Y,Z]\,|Y\big] = \sum_{\ell\in L, j\in J} u_\ell\, \mathbb{P}\Big(\bigsqcup_{k\in K(\ell,j)} \{Z = z_k\} | Y = y_j\Big) \mathbf{1}_{\{Y=y_j\}} \\
&= \sum_{j\in J}\sum_{\ell\in L_j} \underbrace{\Big(\sum_{i\in I} x_i\, \mathbb{P}(X = x_i | Y = y_j, Z = z_k)\Big)}_{=u_{j,k}=u_\ell \text{ lorsque } k\in K(\ell,j), \ell\in L_j, j\in J} \sum_{k\in K(\ell,j)} \mathbb{P}(Z = z_k | Y = y_j) \mathbf{1}_{\{Y=y_j\}} \\
&= \sum_{j\in J}\sum_{\ell\in L_j}\sum_{k\in K(\ell,j)} \Big(\sum_{i\in I} x_i\, \mathbb{P}(X = x_i | Y = y_j, Z = z_k)\Big) \mathbb{P}(Z = z_k | Y = y_j) \mathbf{1}_{\{Y=y_j\}} \\
&= \sum_{(j,k)\in J\times K} \Big(\sum_{i\in I} x_i\, \mathbb{P}(X = x_i | Y = y_j, Z = z_k)\Big) \mathbb{P}(Z = z_k | Y = y_j) \mathbf{1}_{\{Y=y_j\}} \\
&\qquad (\text{car } K = \textstyle\bigsqcup_{\ell\in L, j\in J} K(\ell, j)) \\
&= \sum_{(i,j)\in I\times J} x_i \left(\sum_{k\in K} \mathbb{P}(X = x_i | Y = y_j, Z = z_k) \mathbb{P}(Z = z_k | Y = y_j)\right) \mathbf{1}_{\{Y=y_j\}}
\end{aligned}$$

(par la formule des probabilités totales (1.3) avec les conditionnements successifs, cf. Prop. 1.4)

$$\begin{aligned}
&= \sum_{(i,j)\in I\times J} x_i\, \mathbb{P}(X = x_i | Y = y_j) \mathbf{1}_{\{Y=y_j\}} \\
&= \mathbb{E}[X|Y] \quad (\text{avec (1.8) pour obtenir (1.9)}).
\end{aligned}$$

□

Avec cette approche discrète des espérances conditionnelles, on observe les propriétés qui serviront à définir l'espérance conditionnelle dans le cas général (Chap. 2).

Proposition 1.14 *On a*

(1) $\mathbb{E}[X|Y]$ *est* $\sigma(Y)$*-mesurable ;*

(2) $\mathbb{E}[\mathbf{1}_A X] = \mathbb{E}\big[\mathbf{1}_A \mathbb{E}[X|Y]\big]$ $\forall A \in \sigma(Y)$.

(3) Lorsque les espérances sont bien définies, pour toute variable aléatoire Z *qui est* $\sigma(Y)$*-mesurable, on a* $\mathbb{E}[ZX] = \mathbb{E}\big[Z\mathbb{E}[X|Y]\big]$.

Démonstration : Le premier point 1) découle directement de l'expression (1.7) puisque $\mathbf{1}_{\{Y=y_j\}} = \mathbf{1}_{\{y_j\}}(Y)$ est $\sigma(Y)$-mesurable. Pour le deuxième point 2), comme $A \in \sigma(Y)$ s'écrit $A = \{Y \in B\}$ pour B mesurable, et comme Y est discrète, il suffit de considérer le cas $A = \{Y = y\}$ pour un atome y de Y. Dans cas

$$\mathbf{1}_A \mathbb{E}[X|Y] = \mathbf{1}_{\{Y=y\}} \Big(\sum_{j\in J} \mathbb{E}[X|Y = y_j] \mathbf{1}_{\{Y=y_j\}} \Big) = \mathbb{E}[X|Y = y] \mathbf{1}_{\{Y=y\}}$$

et

$$\mathbb{E}\big[\mathbf{1}_A \mathbb{E}[X|Y]\big] = \mathbb{E}\big[\mathbb{E}[X|Y = y]\mathbf{1}_{\{Y=y\}}\big] = \mathbb{E}[X|Y = y]\mathbb{P}(Y = y) = \mathbb{E}[X\mathbf{1}_{\{Y=y\}}] = \mathbb{E}[\mathbf{1}_A X].$$

Le point 3) se prouve de la même façon : d'après le Théorème de Doob-Dynkin (Th. 0.7, $Z = h(Y)$ où h est une fonction mesurable. On a alors

$$Z\mathbb{E}[X|Y] = h(Y)\mathbb{E}[X|Y] = \sum_{j\in J} h(y_j)\mathbb{E}[X|Y = y_j]\mathbf{1}_{\{Y=y_j\}}$$

et

$$\begin{aligned} \mathbb{E}\big[Z\mathbb{E}[X|Y]\big] &= \sum_{j\in J} h(y_j)\mathbb{E}[X|Y = y_j]\mathbb{P}(Y = y_j) = \sum_{j\in J} h(y_j)\mathbb{E}[X\mathbf{1}_{\{Y=y_j\}}] \\ &= \mathbb{E}\Big[X \underbrace{\Big(\sum_{j\in J} h(y_j)\mathbf{1}_{\{Y=y_j\}}\Big)}_{=h(Y)=Z}\Big] = \mathbb{E}[XZ]. \end{aligned}$$

□

1.3 Lois conditionnelles discrètes

Soit (X, Y) un couple de variables aléatoires discrètes sur $(\Omega, \mathcal{F}, \mathbb{P})$. On note $(E, \mathcal{E})$ l'espace des valeurs de X.

Définition 1.15 (Loi conditionnelle) *Pour* y *tel que* $\mathbb{P}(Y = y) \neq 0$ *(*y *atome de* Y*), on appelle loi conditionnelle de* X *sachant* $Y = y$*, l'application définie par*

$$\mathbb{P}(X \in A|Y = y) = \frac{\mathbb{P}(X \in A, Y = y)}{\mathbb{P}(Y = y)} = \sum_{x\in A} \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}, \quad A \in \mathcal{E}. \tag{1.11}$$

Plus généralement, on appelle loi conditionnelle de X *sachant* Y *l'application définie par*

$$\mathbb{P}_X(A|Y) = \mathbb{P}(X \in A|Y) = \sum_{y\in\mathcal{S}(Y)} \mathbb{P}(X \in A|Y = y)\mathbf{1}_{\{Y=y\}}, \quad A \in \mathcal{E}. \tag{1.12}$$

Si $y \notin S(Y)$, $\mathbb{P}(X = x|Y = y)$ n'est pas définie en (1.11) ; on pourra éventuellement décider de lui donner une valeur arbitraire (par exemple zéro).

D'après les définitions des espérances conditionnelles en (1.7) et des lois conditionnelles en (1.12) sachant Y, l'espérance conditionnelle sachant Y coïncide avec l'espérance par rapport à la loi conditionnelle sachant Y :

$$\mathbb{E}[h(X)|Y] = \int h(x)\,\mathbb{P}_X(dx|Y). \tag{1.13}$$

Proposition 1.16 *Si $X \perp\!\!\!\perp Y$ alors la loi conditionnelle de X sachant Y est la même que celle de X :*

$$\forall y \in \mathcal{S}(Y)\ :\ \mathbb{P}_X(*|Y = y) = \mathbb{P}_X \qquad et \qquad \mathbb{P}_X(*|Y) = \mathbb{P}_X.$$

Autrement dit : le conditionnement par une variable aléatoire indépendante est sans effet sur la loi d'une variable aléatoire.

Démonstration : C'est une conséquence directe de $\mathbb{P}(A|B) = \mathbb{P}(A)$ lorsque $A \perp\!\!\!\perp B$, en effet pour tout $A \in \mathcal{E}$, on a $\{X \in A\} \perp\!\!\!\perp \{Y = y\}$ et donc

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

et il suit de (1.12) que $\mathbb{P}_X(A|Y) = \mathbb{P}_X(A)$. □

Lois conditionnelles à densité

Soit (X, Y) un couple de densité f sur $\mathbb{R}^2$. On rappelle que X et Y ont alors pour densités respectives

$$f_X(x) = \int_{\mathbb{R}} f(x, y)\,dy, \qquad \text{et} \qquad f_Y(y) = \int_{\mathbb{R}} f(x, y)\,dx.$$

Dans cette situation, on a un analogue de (1.11) pour les densités avec la notion de densité conditionnelle :

Définition 1.17 (Densité conditionnelle) *Soit (X, Y) un couple de variables aléatoires réelles de densité $f : \mathbb{R}^2 \to \mathbb{R}$ et $y \in \mathcal{S}(Y)$. On définit la densité conditionnelle de X sachant $Y = y$ par*

$$f_{X|Y=y}(x) = \frac{f(x, y)}{f_Y(y)}. \tag{1.14}$$

La densité conditionnelle $f_{X|Y=y}$ définit la loi conditionnelle $\mathcal{L}(X|Y = y)$ de X sachant $Y = y$ (on le verra en Prop. 2.35). Il s'agit d'une fonction de la seule variable x ; la variable y y apparaît seulement comme un paramètre. Comme pour la Prop. 1.16, on a :

Proposition 1.18 *Si les variables aléatoires X et Y sont indépendantes de densité f_X et f_Y alors les densités conditionnelles sont les densités marginales :*

$$f_{X|Y=y}(x) = f_X(x) \quad \forall y \in \mathcal{S}(Y).$$

À nouveau, le conditionnement est sans effet car les variables aléatoires sont indépendantes.

Démonstration : Lorsque $X \perp\!\!\!\perp Y$, $f(x,y) = f_X(x)f_Y(y)$ et l'affirmation suit immédiatement de la forme de la densité conditionnelle (1.14). $\square$

Chapitre 2

Espérance conditionnelle

Dans ce chapitre, on définit la notion d'espérance conditionnelle sachant une sous-tribu. Les conditionnements par une variable aléatoire ou par un évènement du Chapitre 1 seront alors vus comme des cas particuliers du conditionnement par une sous-tribu.

À la notion de conditionnement sont associées celles de probabilités conditionnelles, de lois conditionnelles et d'espérances conditionnelles. On introduit dans ce chapitre ces objets et on explique leurs liens.

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité.

2.1 Introduction et définition

Étant donné une sous-tribu $\mathcal{G}$ de $\mathcal{F}$, on définit la notion d'espérance conditionnelle sachant $\mathcal{G}$ d'une variable aléatoire X.

Définition 2.1 (Espérance conditionnelle) *On appelle espérance conditionnelle de X sachant $\mathcal{G}$, notée $\mathbb{E}[X|\mathcal{G}]$, la variable aléatoire Y presque sûrement unique vérifiant*

(i) Y est $\mathcal{G}$-mesurable,

(ii) pour tout $A \in \mathcal{G}$,

$$\mathbb{E}\big[X\mathbf{1}_A\big] = \mathbb{E}\big[Y\mathbf{1}_A\big]. \tag{2.1}$$

Remarque 2.2

1. Attention, l'espérance conditionnelle $\mathbb{E}[X|\mathcal{G}]$ est définie presque sûrement seulement.
2. L'espérance conditionnelle de X est définie dès que les espérances dans (2.1) sont bien définies, typiquement pour X positive ou X intégrable.
3. Intuitivement, on interprète une tribu comme une quantité d'information. Ainsi quand on dispose de l'information de $\mathcal{G}$ (ie. pour tout $A \in \mathcal{G}$, on sait si A est réalisé ou pas), l'espérance conditionnelle $\mathbb{E}[X|\mathcal{G}]$ représente la « meilleure » estimation de X compte tenu de l'information disponible sachant $\mathcal{G}$.

Cette Définition 2.1 nécessite une justification :

Proposition 2.3 (Existence et unicité) *Soit X une variable aléatoire positive ou intégrable.*

1. *Il existe une variable aléatoire Y vérifiant (i)-(ii) dans la Définition 2.1.*
2. *Si Y, Y' sont deux variables aléatoires vérifiant (i)-(ii) dans la Définition 2.1 alors $Y = Y'$ ps.*

Démonstration : (1) L'existence de l'espérance conditionnelle est assurée par le théorème de Radon-Nikodym (Th. 0.6).

On suppose d'abord $X \geq 0$ et on définit une mesure sur $\mathcal{G}$ par $\mathbb{Q}(A) = \mathbb{E}[X\mathbf{1}_A]$, $A \in \mathcal{G}$. Il est immédiat que pour $A \in \mathcal{G}$: $\mathbb{P}(A) = 0 \Rightarrow \mathbb{Q}(A) = 0$. On a donc $\mathbb{Q} \ll \mathbb{P}_\mathcal{G}$ (restriction de $\mathbb{P}$ à $\mathcal{G}$) et le théorème de Radon-Nikodym (Th. 0.6) sur $(\Omega, \mathcal{G}, \mathbb{P}_\mathcal{G})$ assure l'existence d'une variable aléatoire $\mathcal{G}$-mesurable (dérivée de Radon-Nikodym)

$$Y(\omega) := \frac{d\mathbb{Q}}{d\mathbb{P}_\mathcal{G}}(\omega)$$

telle que $\mathbb{Q}(A) = \int_A Y d\mathbb{P}_\mathcal{G}$, c'est à dire (2.1).

Si X est intégrable, on écrit $X = X^+ - X^-$ et on applique le cas précédent aux variables aléatoires X^+ et X^- positives. On note alors $Y_1 = \mathbb{E}[X^+|\mathcal{G}]$, $Y_2 = \mathbb{E}[X^-|\mathcal{G}]$ et on pose $Y = Y_1 - Y_2$. La variable aléatoire Y est $\mathcal{G}$-mesurable car différence de telles fonctions et pour tout $A \in \mathcal{G}$:

$$\begin{aligned} \mathbb{E}[\mathbf{1}_A X] &= \mathbb{E}[X^+\mathbf{1}_A] - \mathbb{E}[X^-\mathbf{1}_A] \qquad (2.2)\\ &= \mathbb{E}[Y_1\mathbf{1}_A] - \mathbb{E}[Y_2\mathbf{1}_A] \\ &= \mathbb{E}[(Y_1 - Y_2)\mathbf{1}_A] = \mathbb{E}[Y\mathbf{1}_A], \end{aligned}$$

ce qui justifie que Y vérifie (i)–(ii) dans la Définition 2.1 et $Y = \mathbb{E}[X|\mathcal{G}]$ ps. Noter que l'intégrabilité de X assure celle de X^+ et de X^- et donc la finitude de $\mathbb{E}[X^+\mathbf{1}_A]$ et $\mathbb{E}[X^-\mathbf{1}_A]$ justifiant que la différence dans (2.2) a bien un sens.

(2) Soit Y, Y' deux variables aléatoires vérifiant la Définition 2.1. Pour tout $A \in \mathcal{G}$, on a $\mathbb{E}[Y\mathbf{1}_A] = \mathbb{E}[Y'\mathbf{1}_A]$. En particulier pour $\varepsilon > 0$, $A_\varepsilon = \{Y - Y' \geq \varepsilon\} \in \mathcal{G}$ et

$$0 = \mathbb{E}[(Y - Y')\mathbf{1}_{A_\varepsilon}] \geq \mathbb{E}[\varepsilon \mathbf{1}_{A_\varepsilon}] = \varepsilon \mathbb{P}(A_\varepsilon).$$

Cela exige $\mathbb{P}(A_\varepsilon) = 0$ et

$$\mathbb{P}(Y > Y') = \mathbb{P}\Big(\bigcup_{\varepsilon \in \mathbb{Q}_+} \{(Y - Y') \geq \varepsilon\}\Big) = \sum_{\varepsilon \in \mathbb{Q}_+} \mathbb{P}(A_\varepsilon) = 0.$$

Ainsi $Y \leq Y'$ ps et en échangeant les rôles de Y, Y', on a aussi $Y = Y'$ ps. □

Proposition 2.4 *La condition (2.1) ((ii) dans la Définition 2.1) est équivalente à avoir pour toute variable aléatoire $\mathcal{G}$-mesurable Z telle que les espérances aient un sens :*

$$\mathbb{E}[XZ] = \mathbb{E}[YZ]. \tag{2.3}$$

Démonstration : On a immédiatement (2.3) implique (2.1) en prenant $Z = \mathbf{1}_A$, $A \in \mathcal{G}$. Réciproquement, si (2.1) est vraie alors (2.3) l'est aussi successivement pour $Z = \mathbf{1}_A$, $A \in \mathcal{G}$, puis pour Z variable aléatoire simple et finalement pour Z positive par convergence monotone si tout est positif ou par convergence dominée dans le cas intégrable. $\square$

Notations

— Lorsque $\mathcal{G} = \sigma(Y)$, on note $\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)]$. Lorsque Y est une variable discrète, la Prop. 1.14 assure que la définition $\mathbb{E}[X|Y]$ du Chapitre 1 coïncide avec ce qui est définie en Déf. 2.1.

— On note $\mathbb{P}(A|\mathcal{G}) = \mathbb{E}[\mathbf{1}_A|\mathcal{G}]$.

2.2 Exemples d'espérance conditionnelle

Exemples simples mais fondamentaux

1. ($\mathcal{G}$-mesurabilité) Si X est $\mathcal{G}$-mesurable alors X vérifie directement (i)-(ii) dans la Définition 2.1 ! On a alors $\mathbb{E}[X|\mathcal{G}] = X$ ps.
   Autrement dit si $\mathcal{G}$ est connue, on connaît tous les $X^{-1}(A)$, $A \in \mathcal{G}$. Ainsi toutes les occurrences de X sont connues, ce qui signifie que X est connue, et sa meilleure approximation est elle même !
2. En particulier, si $X = c$ (constante), $\mathbb{E}[X|\mathcal{G}] = X$.
3. (Indépendance) Si $X \perp\!\!\!\perp \mathcal{G}$ alors $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$ ps. En effet avec $A \in \mathcal{F}$, on a
   $$\mathbb{E}[\mathbf{1}_A X] = \mathbb{E}[\mathbf{1}_A]\mathbb{E}[X] = \mathbb{E}\big[\mathbf{1}_A \mathbb{E}[X]\big],$$
   comme en plus une constante est bien $\mathcal{G}$-mesurable, on a bien $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$ ps.
   Dans le cas indépendant la connaissance de $\mathcal{G}$ ne donne aucune information sur X et la meilleure estimation de X (sachant $\mathcal{G}$) est alors sa moyenne $\mathbb{E}[X]$.
4. Dans le cas $\mathcal{G} = \{\emptyset, \Omega\}$ (tribu grossière), on vérifie facilement que $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$ puisque les variables aléatoires $\{\emptyset, \Omega\}$-mesurables sont des constantes et (2.1) n'est à vérifier que pour $A = \emptyset$ et $A = \Omega$ pour lesquels c'est immédiat.

Conditionnement et système complet

On considère un système complet dénombrable $(\Omega_i)_{i\in I}$ (Déf. 1.5 : avec $\Omega_i \cap \Omega_j = \emptyset$ pour $i \neq j$ et $\sum_{i\in I} \mathbb{P}(\Omega_i) = 1$, $I \subset \mathbb{N}$) avec $\mathbb{P}(\Omega_i) > 0$ pour chaque $i \in I$ et on note $\Omega_0 = \bigsqcup_{i\in I} \Omega_i$ de sorte que $\Omega_* = \Omega_0^c$ est négligeable. On considère alors la sous-tribu $\mathcal{G} = \sigma(\Omega_i : i \in I)$.

Lemme 2.5 *La tribu $\mathcal{G} = \sigma(\Omega_i : i \in I)$ engendrée par un système complet dénombrable $(\Omega_i)_{i\in I}$ est $\{\bigcup_{j\in J} \Omega_j : J \subset I \cup \{*\}\}$.*

Démonstration : Comme il est clair que $\mathcal{H} := \{\bigcup_{j\in J} \Omega_j : J \subset I \cup \{*\}\} \subset \mathcal{G}$, il suffit de montrer que $\mathcal{H}$ est une tribu, ce qui est le cas puisque

— $\emptyset \in \mathcal{H}$ car pour $J = \emptyset$ on a $\bigcup_{j\in J} \Omega_j = \emptyset$;

— $\mathcal{H}$ est stable par complémentarité puisque pour $\bigcup_{j\in J} \Omega_j \in \mathcal{H}$ alors

$$\Big(\bigcup_{j\in J} \Omega_j\Big)^c = \Big(\bigcup_{i\in I\cup\{*\}} \Omega_i\Big) \setminus \Big(\bigcup_{j\in J} \Omega_j\Big) = \Big(\bigcup_{k\in I\cup\{*\}\setminus J} \Omega_k\Big) \in \mathcal{H};$$

— $\mathcal{H}$ est stable par union dénombrable puisque si $\bigcup_{j\in J_k} \Omega_j \in \mathcal{H}$ pour des $J_k \subset I\cup\{*\}$ alors

$$\bigcup_{k\in\mathbb{N}} \Big(\bigcup_{j\in J_k} \Omega_j\Big) = \bigcup_{j\in\bigcup_{k\in\mathbb{N}} J_k} \Omega_j \quad \text{avec } J = \textstyle\bigcup_{k\in\mathbb{N}} J_k \subset I \cup \{*\}.$$

□

Un cas particulier de système complet est la partition $\Omega = B \sqcup B^c$ avec $B \in \mathcal{F}$ tel que $\mathbb{P}(B) \in]0, 1[$; dans ce cas $\sigma(B) = \{\emptyset, B, B^c, \Omega\}$.

Proposition 2.6 *Soit X une variable aléatoire positive ou intégrable. Pour la tribu $\mathcal{G} = \sigma(\Omega_i : i \in I)$ associée à un système complet dénombrable $(\Omega_i)_{i\in I}$ avec $\mathbb{P}(\Omega_i) > 0$, $i \in I$, on a*

$$\mathbb{E}[X\,|\mathcal{G}] = \sum_{i\in I} \frac{\mathbb{E}[X\mathbf{1}_{\Omega_i}]}{\mathbb{P}(\Omega_i)}\mathbf{1}_{\Omega_i} \qquad ps. \tag{2.4}$$

c'est à dire $\mathbb{E}[X|\mathcal{G}] = \frac{\mathbb{E}[X\mathbf{1}_{\Omega_i}]}{\mathbb{P}(\Omega_i)}$ sur Ω_i.

Démonstration : On montre (2.4) dans le cas d'une partition simple $\Omega = B \sqcup B^c$ avec $\mathbb{P}(B) \in]0, 1[$ pour laquelle (2.4) se réduit à deux termes :

$$\mathbb{E}\big[X\,|\sigma(B)\big] = \frac{\mathbb{E}[X\mathbf{1}_B]}{\mathbb{P}(B)}\mathbf{1}_B + \frac{\mathbb{E}[X\mathbf{1}_{B^c}]}{\mathbb{P}(B^c)}\mathbf{1}_{B^c}. \tag{2.5}$$

Comme $Z := \frac{\mathbb{E}[X\mathbf{1}_B]}{\mathbb{P}(B)}\mathbf{1}_B + \frac{\mathbb{E}[X\mathbf{1}_{B^c}]}{\mathbb{P}(B^c)}\mathbf{1}_{B^c}$ est $\mathcal{G}$-mesurable, pour voir (2.5), il reste à montrer que pour tout $C \in \mathcal{G} = \sigma(B) = \{\emptyset, B, B^c, \Omega\}$ on a :

$$\mathbb{E}[X\mathbf{1}_C] = \mathbb{E}[Z\mathbf{1}_C]. \tag{2.6}$$

— Pour $C = B$: $Z\mathbf{1}_C = \frac{\mathbb{E}[X\mathbf{1}_B]}{\mathbb{P}(B)}\mathbf{1}_B$ et (2.6) s'écrit

$$\mathbb{E}[X\mathbf{1}_B] = \mathbb{E}\Big[\frac{\mathbb{E}[X\mathbf{1}_B]}{\mathbb{P}(B)}\mathbf{1}_B\Big],$$

ce qui est vraie.

— Pour $C = B^c$: $Z\mathbf{1}_C = \frac{\mathbb{E}[X\mathbf{1}_{B^c}]}{\mathbb{P}(B^c)}\mathbf{1}_{B^c}$ et (2.6) s'écrit

$$\mathbb{E}[X\mathbf{1}_{B^c}] = \mathbb{E}\Big[\frac{\mathbb{E}[X\mathbf{1}_{B^c}]}{\mathbb{P}(B^c)}\mathbf{1}_{B^c}\Big],$$

ce qui est encore vraie.

— Pour $C = \emptyset$: (2.6) se réduit à $0 = 0$, vraie !

— Pour $C = \Omega$: le membre de gauche de (2.6) s'écrit $\mathbb{E}[X\mathbf{1}_C] = \mathbb{E}[X]$ et celui de droite

$$\begin{aligned}
\mathbb{E}[Z\mathbf{1}_C] &= \mathbb{E}[Z] \\
&= \mathbb{E}\Big[\frac{\mathbb{E}[X\mathbf{1}_B]}{\mathbb{P}(B)}\mathbf{1}_B + \frac{\mathbb{E}[X\mathbf{1}_{B^c}]}{\mathbb{P}(B^c)}\mathbf{1}_{B^c}\Big] \\
&= \frac{\mathbb{E}[X\mathbf{1}_B]}{\mathbb{P}(B)}\mathbb{P}(B) + \frac{\mathbb{E}[X\mathbf{1}_{B^c}]}{\mathbb{P}(B^c)}\mathbb{P}(B^c) \\
&= \mathbb{E}[X\mathbf{1}_B] + \mathbb{E}[X\mathbf{1}_{B^c}] = \mathbb{E}[X],
\end{aligned}$$

ce qui prouve (2.5).

Pour un système complet général, Z prend la forme

$$Z = \sum_{i\in I}\frac{\mathbb{E}[X\mathbf{1}_{\Omega_i}]}{\mathbb{P}(\Omega_i)}\mathbf{1}_{\Omega_i}$$

qui est bien $\mathcal{G}$-mesurable et il reste à vérifier (2.6) pour $C \in \mathcal{G} = \sigma(\Omega_i : i \in I)$ donc d'après le Lemme 2.5 de la forme $C = \bigcup_{j\in J}\Omega_j$ pour $J \subset I \cup \{*\}$ (cf. notation $*$ dans le Lemme 2.5). Pour cela :

$$\begin{aligned}
\mathbb{E}[Z\mathbf{1}_C] &= \mathbb{E}\Big[\Big(\sum_{i\in I}\frac{\mathbb{E}[X\mathbf{1}_{\Omega_i}]}{\mathbb{P}(\Omega_i)}\mathbf{1}_{\Omega_i}\Big)\mathbf{1}_C\Big] = \sum_{j\in J}\mathbb{E}\Big[\Big(\sum_{i\in I}\frac{\mathbb{E}[X\mathbf{1}_{\Omega_i}]}{\mathbb{P}(\Omega_i)}\mathbf{1}_{\Omega_i}\Big)\mathbf{1}_{\Omega_j}\Big] \\
&= \sum_{j\in J}\sum_{i\in I}\frac{\mathbb{E}[X\mathbf{1}_{\Omega_i}]}{\mathbb{P}(\Omega_i)}\mathbb{E}[\mathbf{1}_{\Omega_j}\mathbf{1}_{\Omega_i}] = \sum_{i\in I}\sum_{j\in J}\frac{\mathbb{E}[X\mathbf{1}_{\Omega_i}]}{\mathbb{P}(\Omega_i)}\mathbb{P}(\Omega_i)\delta_{i,j} \\
&= \sum_{j\in J}\mathbb{E}[X\mathbf{1}_{\Omega_j}] = \mathbb{E}[X\mathbf{1}_C].
\end{aligned}$$

Cela assure que Z satisfait bien la Définition 2.1 de $\mathbb{E}[X|\mathcal{G}]$ et prouve la Prop. 2.6. □

Remarque 2.7 On retrouve l'approche élémentaire du Chapitre 1.

— On fait le lien avec (1.1) en notant que $\mathbb{P}(A|B)$ est donnée par la valeur de $\mathbb{E}[\mathbf{1}_A|\sigma(B)]$ sur B.

— Si Y est une variable aléatoire discrète de support $\mathcal{S}(Y) = \{y_j : j \in J\}$, en prenant le système complet des $\Omega_j = \{Y = y_j\}$, $j \in J$, on retrouve la Définition 1.11 en la combinant avec (1.5).

2.3 Propriétés de l'espérance conditionnelle

Proposition 2.8 (Linéarité) *Pour $X, Y \in L^1(\mathcal{F})$, et $a, b \in \mathbb{R}$, on a*

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

L'égalité est valable dès que le terme $a\mathbb{E}[X \,|\mathcal{G}] + b\mathbb{E}[Y \,|\mathcal{G}]$ a bien un sens.

Attention, le presque sûr ci-dessus dépend de a et b si bien qu'il est incorrect d'affirmer que presque sûrement $\mathbb{E}[\cdot \,|\mathcal{G}]$ est linéaire.

Démonstration : Il est clair que $a\mathbb{E}[X \,|\mathcal{G}] + b\mathbb{E}[Y \,|\mathcal{G}]$ est $\mathcal{G}$-mesurable. Puis pour $A \in \mathcal{G}$, par linéarité de l'espérance, on a

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

d'où (ii) dans la Définition 2.1, ce qui assure $\mathbb{E}[aX + bY \,|\mathcal{G}] = a\mathbb{E}[X \,|\mathcal{G}] + b\mathbb{E}[Y \,|\mathcal{G}]$ ps. □

Proposition 2.9 (Monotonie) *Si $X \leq Y$ alors $\mathbb{E}[X|\mathcal{G}] \leq \mathbb{E}[Y|\mathcal{G}]$.*

Démonstration : On commence par voir que si $Y \geq 0$ alors $\mathbb{E}[Y|\mathcal{G}] \geq 0$.
Pour cela, en prenant $A_\varepsilon = \big\{\mathbb{E}[Y|\mathcal{G}] \leq -\varepsilon\big\} \in \mathcal{G}$ pour $\varepsilon > 0$, dans la propriété (2.1) de l'espérance conditionnelle, on a

$$0 \leq \mathbb{E}\big[Y\mathbf{1}_{A_\varepsilon}\big] = \mathbb{E}\big[\mathbb{E}[Y|\mathcal{G}]\mathbf{1}_{A_\varepsilon}\big] \leq \mathbb{E}[(-\varepsilon)\mathbf{1}_{A_\varepsilon}] = -\varepsilon\mathbb{P}(A_\varepsilon) \leq 0.$$

On a donc $\mathbb{P}(A_\varepsilon) = 0$ et $\mathbb{P}\big(\bigcup_{\varepsilon \in \mathbb{Q}_+^*} A_\varepsilon\big) = 0$. Mais comme

$$\bigcup_{\varepsilon \in \mathbb{Q}_+^*} A_\varepsilon = \big\{\mathbb{E}[Y|\mathcal{G}] < 0\big\},$$

il vient $\mathbb{E}[Y|\mathcal{G}] \geq 0$ ps.

De façon générale, on suppose que les espérances conditionnelles $\mathbb{E}[X|\mathcal{G}]$ et $\mathbb{E}[Y|\mathcal{G}]$ existent (variables aléatoires X, Y positives ou intégrables). Si $\mathbb{E}[Y|\mathcal{G}] = +\infty$ ou $\mathbb{E}[X|\mathcal{G}] = -\infty$, alors la conclusion est immédiate. Sinon, alors $-\infty < \mathbb{E}[X|\mathcal{G}]$ et $\mathbb{E}[Y|\mathcal{G}] < +\infty$, et on déduit de $Y = (Y - X) + X$ par linéarité (sans forme indéterminée dans le cas considéré) que

$$\mathbb{E}[Y|\mathcal{G}] = \mathbb{E}[Y - X|\mathcal{G}] + \mathbb{E}[X|\mathcal{G}].$$

En appliquant la première partie à $Y - X \geq 0$, il vient :

$$\mathbb{E}[Y|\mathcal{G}] - \mathbb{E}[X|\mathcal{G}] = \mathbb{E}[Y - X|\mathcal{G}] \geq 0 \quad \text{ps.}$$

□

Corollaire 2.10 *On a* $|\mathbb{E}[X|\mathcal{F}]| \leq \mathbb{E}[|X|\ |\mathcal{F}]$ *ps.*

Démonstration : Comme $X \leq |X|$, on a $\mathbb{E}[X|\mathcal{F}] \leq \mathbb{E}[|X|\ |\mathcal{F}]$ ps. De même, comme $-X \leq |X|$, on a aussi $-\mathbb{E}[X|\mathcal{F}] = \mathbb{E}[-X|\mathcal{F}] \leq \mathbb{E}[|X|\ |\mathcal{F}]$ ps. Et finalement, $|\mathbb{E}[X|\mathcal{F}]| \leq \mathbb{E}[|X|\ |\mathcal{F}]$ ps. □

Proposition 2.11 (Espérance et espérance conditionnelle) *(1) On a*

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

(2) Si $X \in L^1(\Omega, \mathcal{F})$ *alors* $\mathbb{E}[X|\mathcal{G}] \in L^1(\Omega, \mathcal{G})$ *et l'espérance conditionnelle est une contraction de* L^1 *:* $\mathbb{E}\big[|\mathbb{E}[X|\mathcal{G}]|\big] \leq \mathbb{E}[|X|]$.

Démonstration : 1) Par la Définition 2.1, on a $\mathbb{E}\big[\mathbb{E}[X|\mathcal{G}]\mathbf{1}_A\big] = \mathbb{E}[X\mathbf{1}_A]$ pour tout $A \in \mathcal{G}$. En particulier, en prenant $A = \Omega \in \mathcal{G}$, il vient $\mathbb{E}\big[\mathbb{E}[X|\mathcal{G}]\big] = \mathbb{E}[X]$!

2) On note $Y = \mathbb{E}[X|\mathcal{G}]$ et $A = \{Y > 0\} \in \mathcal{G}$. On a

$$\mathbb{E}[\mathbf{1}_A Y] = \mathbb{E}[\mathbf{1}_A X] \leq \mathbb{E}[\mathbf{1}_A |X|]$$

et

$$\mathbb{E}[\mathbf{1}_{A^c}(-Y)] = -\mathbb{E}[\mathbf{1}_{A^c} X] = \mathbb{E}[\mathbf{1}_{A^c}(-X)] \leq \mathbb{E}[\mathbf{1}_{A^c}|X|]$$

ce qui assure

$$\mathbb{E}\big[|Y|\big] = \mathbb{E}\big[\mathbf{1}_A Y\big] + \mathbb{E}\big[\mathbf{1}_{A^c}(-Y)\big] \leq \mathbb{E}\big[\mathbf{1}_A|X|\big] + \mathbb{E}\big[\mathbf{1}_{A^c}|X|\big] \leq \mathbb{E}\big[|X|\big].$$

□

Théorème 2.12 (Cascade) *Soit* $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F}$ *des sous-tribus ordonnées (par inclusion). Alors, on a :*

(1) $\mathbb{E}\big[\mathbb{E}[X\,|\mathcal{G}_1]\,|\mathcal{G}_2\big] = \mathbb{E}[X\,|\mathcal{G}_1]$ *ps ;*

(2) $\mathbb{E}\big[\mathbb{E}[X\,|\mathcal{G}_2]\,|\mathcal{G}_1\big] = \mathbb{E}[X\,|\mathcal{G}_1]$ *ps.*

Démonstration : Comme la variable aléatoire $\mathbb{E}[X|\mathcal{G}_1]$ est $\mathcal{G}_1$-mesurable, elle est *a fortiori* $\mathcal{G}_2$-mesurable et (1) suit de l'exemple simple 1 en Section 2.2.
Pour (2), en prenant $A \in \mathcal{G}_1 \subset \mathcal{G}_2$, on a

$$\mathbb{E}\big[\mathbf{1}_A\mathbb{E}[X\,|\mathcal{G}_1]\big] = \mathbb{E}\big[\mathbf{1}_A X\big] = \mathbb{E}\big[\mathbf{1}_A\mathbb{E}[X|\mathcal{G}_2]\big] \quad \text{ps}$$

car $A \in \mathcal{G}_1$ (à gauche) et $A \in \mathcal{G}_2$ (à droite). D'où on déduit $\mathbb{E}\big[\mathbb{E}[X\,|\mathcal{G}_2]\,|\mathcal{G}_1\big] = \mathbb{E}[X\,|\mathcal{G}_1]$ ps. □

Remarque 2.13 Attention, en général on n'a pas

$$\mathbb{E}\big[\mathbb{E}[X\,|\mathcal{G}_2]\,|\mathcal{G}_1\big] = \mathbb{E}\big[\mathbb{E}[X\,|\mathcal{G}_1]\,|\mathcal{G}_2\big] \quad \text{ps.} \tag{2.7}$$

On considère l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \lambda)$ et on prend $\mathcal{G}_1 = \sigma([0,1/2]) = \{\emptyset, [0,1/2],]1/2,1], [0,1]\}$, $\mathcal{G}_2 = \sigma([0,1/3]) = \{\emptyset, [0,1/3],]1/3,1], [0,1]\}$. Pour $X = \mathbf{1}_{[1/4,3/4]}$, on a presque sûrement :

$$\begin{aligned}
\mathbb{E}[X|\mathcal{G}_1] &= \frac{\mathbb{E}[X\mathbf{1}_{[0,1/2]}]}{\mathbb{P}([0,1/2])}\mathbf{1}_{[0,1/2]} + \frac{\mathbb{E}[X\mathbf{1}_{]1/2,1]}]}{\mathbb{P}(]1/2,1])}\mathbf{1}_{]1/2,1]}\\
&= \frac{\lambda([1/4,1/2])}{\lambda([0,1/2])}\mathbf{1}_{[0,1/2]} + \frac{\lambda(]1/2,3/4])}{\lambda(]1/2,1])}\mathbf{1}_{]1/2,1]}\\
&= \frac{1}{2}\mathbf{1}_{[0,1/2]} + \frac{1}{2}\mathbf{1}_{]1/2,1]} = \frac{1}{2},
\end{aligned}$$

et on a donc $\mathbb{E}\big[\mathbb{E}[X|\mathcal{G}_1]|\mathcal{G}_2\big] = \frac{1}{2}$ ps. Puis presque sûrement :

$$\begin{aligned}
\mathbb{E}[X|\mathcal{G}_2] &= \frac{\mathbb{E}[X\mathbf{1}_{[0,1/3]}]}{\mathbb{P}([0,1/3])}\mathbf{1}_{[0,1/3]} + \frac{\mathbb{E}[X\mathbf{1}_{]1/3,1]}]}{\mathbb{P}(]1/3,1])}\mathbf{1}_{]1/3,1]}\\
&= \frac{\lambda([1/4,1/3])}{\lambda([0,1/3])}\mathbf{1}_{[0,1/3]} + \frac{\lambda(]1/3,3/4])}{\lambda(]1/3,1])}\mathbf{1}_{]1/3,1]}\\
&= \frac{1}{4}\mathbf{1}_{[0,1/3]} + \frac{5}{8}\mathbf{1}_{]1/3,1]},
\end{aligned}$$

et on a donc ps :

$$\begin{aligned}
&\mathbb{E}\big[\mathbb{E}[X|\mathcal{G}_2]|\mathcal{G}_1\big] = \mathbb{E}\Big[\frac{1}{4}\mathbf{1}_{[0,1/3]} + \frac{5}{8}\mathbf{1}_{]1/3,1]}\Big|\mathcal{G}_1\Big]\\
&= \frac{\mathbb{E}\Big[\Big(\frac{1}{4}\mathbf{1}_{[0,1/3]} + \frac{5}{8}\mathbf{1}_{]1/3,1]}\Big)\mathbf{1}_{[0,1/2]}\Big]}{\mathbb{P}([0,1/2])}\mathbf{1}_{[0,1/2]} + \frac{\mathbb{E}\Big[\Big(\frac{1}{4}\mathbf{1}_{[0,1/3]} + \frac{5}{8}\mathbf{1}_{]1/3,1]}\Big)\mathbf{1}_{]1/2,1]}\Big]}{\mathbb{P}(]1/2,1])}\mathbf{1}_{]1/2,1]}\\
&= 2\Big(\frac{1}{4}\times\frac{1}{3} + \frac{5}{8}\times\frac{1}{6}\Big)\mathbf{1}_{[0,1/2]} + 2\Big(\frac{1}{4}\times 0 + \frac{5}{8}\times\frac{1}{2}\Big)\mathbf{1}_{[1/2,1]}\\
&= \frac{3}{8}\mathbf{1}_{[0,1/2]} + \frac{5}{8}\mathbf{1}_{[1/2,1]},
\end{aligned}$$

ce qui fournit bien un contre-exemple à (2.7).

Versions conditionnelles de résultats classiques

Proposition 2.14 (Inégalité de Tchebychev) *Pour $\varepsilon > 0$, on a*

$$\mathbb{P}\big(|X| \geq \varepsilon\,|\mathcal{G}\big) \leq \frac{\mathbb{E}\big[X^2\,|\mathcal{G}\big]}{\varepsilon^2} \quad \textit{ps.}$$

Démonstration : Il s'agit d'appliquer la Prop. 2.9 à l'inégalité $\varepsilon^2\mathbf{1}_{\{|X|\geq\varepsilon\}} \leq X^2$. □

Proposition 2.15 (Convergence monotone conditionnelle) *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires positives telles que $X_{n-1}\leq X_n \nearrow X$. Alors*

$$\mathbb{E}\big[X_n|\mathcal{G}\big]\nearrow \mathbb{E}\big[X|\mathcal{G}\big],\quad n\to+\infty\quad ps.$$

Démonstration : Comme $0\leq X_n\nearrow X$, par la monotonie (Prop. 2.9), la suite $\big(\mathbb{E}[X_n\,|\mathcal{G}]\big)_{n\geq 1}$ est croissante et admet donc une limite Z positive, $\mathcal{G}$-mesurable car limite de variables aléatoires $\mathcal{G}$-mesurables. De plus, pour $A\in\mathcal{G}$, on a

$$\begin{aligned}
\mathbb{E}[Z\mathbf{1}_A] &= \lim_{n\to+\infty}\mathbb{E}\big[\mathbb{E}[X_n\,|\mathcal{G}]\mathbf{1}_A\big] \quad \text{(par convergence monotone classique)}\\
&= \lim_{n\to+\infty}\mathbb{E}[X_n\mathbf{1}_A] \quad \text{(par la Définition 2.1)}\\
&= \mathbb{E}[X\mathbf{1}_A] \quad \text{(par convergence monotone classique).}
\end{aligned}$$

Finalement, on a $Z=\mathbb{E}[X\,|\mathcal{G}]$ ps. □

On déduit alors successivement comme dans le cas non-conditionnel le lemme de Fatou conditionnel puis le théorème de convergence dominée conditionnel (cf. [Bre-Leb]) :

Proposition 2.16 (Lemme de Fatou conditionnel) *Si X_n, $n\geq 1$, sont des variables aléatoires positives alors*

$$\mathbb{E}\big[\liminf_{n\to+\infty}X_n\,|\mathcal{G}\big]\leq \liminf_{n\to+\infty}\mathbb{E}\big[X_n\,|\mathcal{G}\big]\quad ps. \tag{2.8}$$

Démonstration : On applique la Prop. 2.15 (convergence monotone conditionnelle) à la suite $\big(\inf_{k\geq n}X_k\big)_{n\geq 0}$ positive croissante vers $\liminf_{n\to+\infty}X_n$. On a alors

$$\begin{aligned}
\mathbb{E}\big[\liminf_{n\to+\infty}X_n\,|\mathcal{G}\big] &= \mathbb{E}\big[\lim_{n\to+\infty}\inf_{k\geq n}X_k\,|\mathcal{G}\big]=\lim_{n\to+\infty}\mathbb{E}\big[\inf_{k\geq n}X_k\,|\mathcal{G}\big] \quad \text{(Prop. 2.15)}\\
&= \liminf_{n\to+\infty}\mathbb{E}\big[\inf_{k\geq n}X_k\,|\mathcal{G}\big]\leq \liminf_{n\to+\infty}\mathbb{E}\big[X_n\,|\mathcal{G}\big] \quad \text{(monotonie, Prop. 2.9)}
\end{aligned}$$

puisque $\inf_{k\geq n}X_k\leq X_n$. □

Proposition 2.17 (Convergence dominée conditionnelle) *Soit X_n, $n\geq 1$, des variables aléatoires avec $|X_n|\leq Z\in L^1(\mathcal{F})$. On suppose que $X_n\to X$ ps. Alors*

$$\lim_{n\to+\infty}\mathbb{E}[X_n|\mathcal{G}]=\mathbb{E}[X|\mathcal{G}]\quad ps\ et\ dans\ L^1. \tag{2.9}$$

Démonstration : Même preuve que le théorème de convergence dominée standard en utilisant la convergence monotone conditionnelle (Prop. 2.15) et le lemme de Fatou conditionnel (Prop. 2.16) : on pose $Y_n=2Z-|X_n-X|$; on a $\lim_{n\to+\infty}Y_n=\liminf_{n\to+\infty}Y_n=2Z$. Le lemme de Fatou conditionnel (Prop. 2.16) appliqué aux variables aléatoires Y_n (intégrables) assure

$$2\mathbb{E}[Z\,|\mathcal{G}]=\mathbb{E}\Big[\liminf_{n\to+\infty}Y_n\,\Big|\mathcal{G}\Big]\leq \liminf_{n\to+\infty}\mathbb{E}[Y_n\,|\mathcal{G}]=2\mathbb{E}[Z\,|\mathcal{G}]-\limsup_{n\to+\infty}\mathbb{E}\big[|X_n-X|\,\big|\mathcal{G}\big].$$

En simplifiant par $\mathbb{E}[Z\,|\mathcal{G}] < +\infty$ ps, on a

$$\limsup_{n\to+\infty} \mathbb{E}\big[|X_n - X|\,\big|\mathcal{G}\big] \leq 0.$$

Comme par ailleurs

$$0 \leq \liminf_{n\to+\infty} \mathbb{E}\big[|X_n - X|\,\big|\mathcal{G}\big] \leq \limsup_{n\to+\infty} \mathbb{E}\big[|X_n - X|\,\big|\mathcal{G}\big],$$

on a donc $\lim_{n\to+\infty} \mathbb{E}\big[|X_n - X|\,\big|\mathcal{G}\big] = 0$ ps. et on conclut à (2.9) avec le Corollaire 2.10 :

$$\big|\mathbb{E}[X_n|\mathcal{G}] - \mathbb{E}[X|\mathcal{G}]\big| = \big|\mathbb{E}\big[(X_n - X)\,|\mathcal{G}\big]\big| \leq \mathbb{E}\big[|X_n - X|\,|\mathcal{G}\big].$$

La convergence L^1 est directe puisque par convergence dominée usuelle $X_n \xrightarrow{L^1} X$ et donc

$$\mathbb{E}\big[\big|\mathbb{E}[X_n\,|\mathcal{G}] - \mathbb{E}[X\,|\mathcal{G}]\big|\big] \leq \mathbb{E}\big[\mathbb{E}[|X_n - X|\,|\mathcal{G}]\big] = \mathbb{E}[|X_n - X|] \to 0, \quad n \to +\infty.$$

□

Proposition 2.18 (Inégalité de Cauchy-Schwarz conditionnelle) *On a*

$$\mathbb{E}[XY\,|\mathcal{G}]^2 \leq \mathbb{E}\big[X^2\,|\mathcal{G}\big]\,\mathbb{E}\big[Y^2\,|\mathcal{G}\big] \quad ps.$$

Démonstration : La preuve est la même que celle de l'inégalité de Cauchy-Schwarz usuelle. Comme $(X+\theta Y)^2 \geq 0$, la linearité (Prop. 2.8) et la monotonie (Prop. 2.9) assurent que

$$\mathbb{E}\big[Y^2\,|\mathcal{G}\big]\theta^2 + 2\mathbb{E}\big[XY\,|\mathcal{G}\big]\theta + \mathbb{E}\big[X^2\,|\mathcal{G}\big]\mathbb{E}\big[Y^2\,|\mathcal{G}\big] = \mathbb{E}\big[(X+\theta Y)^2\,|\mathcal{G}\big] \geq 0 \quad \text{ps.}$$

Le polynôme en θ est positif pour tout $\theta \in \mathbb{Q}$, presque sûrement (attention à l'échange entre le ps et le $\forall \theta \in \mathbb{Q}$ est licite car on manipule une collection dénombrable de θ). Cela exige que son discriminant soit négatif ps, ce qui prouve l'inégalité. □

Proposition 2.19 (Inégalité de Jensen conditionnelle) *Soit $X \in L^1(\mathcal{F})$ et $\varphi : \mathbb{R} \to \mathbb{R}$ une fonction convexe telle que $\mathbb{E}[|\varphi(X)|] < +\infty$. Alors*

$$\varphi\big(\mathbb{E}[X\,|\mathcal{G}]\big) \leq \mathbb{E}[\varphi(X)\,|\mathcal{G}] \quad ps. \tag{2.10}$$

Démonstration : Si $\varphi(x) = ax + b$ est linéaire, (2.10) est immédiate par la linéarité de l'espérance conditionnelle (Prop. 2.8). De façon générale, la convexité de φ assure qu'en tout point de son graphe la courbe de φ est au dessus de sa tangente, ie. pour tout $y \in \mathbb{R}$ et tout $d_y \in [\varphi'_g(y), \varphi'_d(y)]$ on a :

$$\forall x \in \mathbb{R}, \quad \varphi(x) \geq \varphi(y) + d_y(x-y). \tag{2.11}$$

Ci-dessous, on prend par exemple $d_y = (\varphi'_g(y) + \varphi'_d(y))/2$. En appliquant cette inégalité (2.11) avec $y = \mathbb{E}[X|\mathcal{G}]$ et $x = X(\omega)$ on a

$$\varphi(X) \geq \varphi\big(\mathbb{E}[X|\mathcal{G}]\big) + d_{\mathbb{E}[X|\mathcal{G}]}(X - \mathbb{E}[X|\mathcal{G}]),$$

puis en prenant l'espérance $\mathbb{E}[\cdot|\mathcal{G}]$, comme $d_{\mathbb{E}[X|\mathcal{G}]}$ est $\mathcal{G}$-mesurable, on a

$$\varphi(X) \geq \varphi\big(\mathbb{E}[X|\mathcal{G}]\big) + d_{\mathbb{E}[X|\mathcal{G}]}\underbrace{\mathbb{E}\big[X - \mathbb{E}[X|\mathcal{G}]\,|\mathcal{G}\big]}_{=0} = \varphi\big(\mathbb{E}[X|\mathcal{G}]\big).$$

□

Proposition 2.20 (Intégrabilité et contraction L^p) *Soit $p \geq 1$.*

(1) On a

$$\big|\mathbb{E}\big[X\,|\mathcal{G}\big]\big|^p \leq \mathbb{E}\big[|X|^p\,|\mathcal{G}\big] \quad ps.$$

(2) L'espérance conditionnelle est une contraction sur L^p :

$$\mathbb{E}\big[\big|\mathbb{E}\big[X\,|\mathcal{G}\big]\big|^p\big] \leq \mathbb{E}\big[|X|^p\big] \quad ps.$$

En particulier, $\mathbb{E}\big[\big|\mathbb{E}\big[X\,|\mathcal{G}\big]\big|\big] \leq \mathbb{E}\big[|X|\big]$ ps.

Démonstration : 1) C'est une conséquence de l'inégalité de Jensen (Prop. 2.19) avec la fonction convexe $\varphi(x) = |x|^p$ pour $p \geq 1$.

2) Avec la monotonie de l'espérance, **1)** ci-dessus donne

$$\mathbb{E}\big[\big|\mathbb{E}\big[X\,|\mathcal{G}\big]\big|^p\big] \leq \mathbb{E}\big[\mathbb{E}\big[|X|^p\,|\mathcal{G}\big]\big] = \mathbb{E}\big[|X|^p\big]$$

où l'égalité vient de la Prop. 2.11. □

Propriétés supplémentaires de l'espérance conditionnelle

Théorème 2.21 *Soit X, Y des variables aléatoires avec X $\mathcal{G}$-mesurable. Lorsque les espérances conditionnelles sont bien définies, c'est à dire*

(1) $X, Y \geq 0$,

(2) $Y \in L^1(\mathcal{F})$ et $XY \in L^1(\mathcal{F})$,

on a

$$\mathbb{E}\big[XY|\mathcal{G}\big] = X\mathbb{E}\big[Y|\mathcal{G}\big] \quad ps. \tag{2.12}$$

Démonstration : D'abord, $X\mathbb{E}\big[Y|\mathcal{G}\big]$ est $\mathcal{G}$-mesurable car produit de fonctions $\mathcal{G}$-mesurables. Ensuite, en supposant pour commencer $X = \mathbf{1}_B$ avec $B \in \mathcal{G}$, on a pour $A \in \mathcal{G}$

$$\mathbb{E}\big[X\mathbb{E}[Y|\mathcal{G}]\mathbf{1}_A\big] = \mathbb{E}\big[\mathbb{E}[Y|\mathcal{G}]\mathbf{1}_{A\cap B}\big] = \mathbb{E}[Y\mathbf{1}_{A\cap B}] = \mathbb{E}[(XY)\mathbf{1}_A]$$

ce qui justifie (2.12) lorsque $X = \mathbf{1}_B$, puis par linéarité pour X variable aléatoire simple.

Dans le cas 1), pour X variable aléatoire $\mathcal{G}$-mesurable positive, il existe $(X_n)_{n\geq 1}$ suite de variables aléatoires $\mathcal{G}$-mesurables positives telle que $X_n \nearrow X$, $n \to +\infty$. Comme $Y \geq 0$, on a aussi $X_nY \nearrow XY$ ($Y \geq 0$), et le théorème de convergence mononotone conditionnel (Prop. 2.15) assure alors

$$\mathbb{E}[XY|\mathcal{G}] = \lim_{n\to+\infty} \mathbb{E}[X_nY|\mathcal{G}] = \lim_{n\to+\infty} X_n\mathbb{E}[Y|\mathcal{G}] = X\mathbb{E}[Y|\mathcal{G}].$$

Dans le cas 2), on peut de nouveau trouver des suites de variables aléatoires $\mathcal{G}$-mesurables simples positives croissantes $(X'_n)_{n\geq 1}$ et $(X''_n)_{n\geq 1}$ qui convergent vers X^+ et X^- et on pose $X_n = X'_n - X''_n$. Comme X_n est simple, le cas précédent assure

$$\mathbb{E}[X_nY|\mathcal{G}] = X_n\mathbb{E}[Y|\mathcal{G}].$$

On conclut en passant à la limite dans cette égalité puisque $\lim_{n\to+\infty} X_n = \lim_{n\to+\infty} X'_n + \lim_{n\to+\infty} X''_n = X' + X'' = X$ ps et comme $|X_n| \leq X'_n + X''_n \leq X^+ + X^- = |X|$, on a $|X_nY| \leq |XY| \in L^1$ et par convergence dominée conditionnelle (Prop. 2.17) :

$$\lim_{n\to+\infty} \mathbb{E}[X_nY\,|\mathcal{G}] = \mathbb{E}[XY\,|\mathcal{G}].$$

□

Théorème 2.22 (Conditionnement et indépendance)

1. *Si* $X \perp\!\!\!\perp \mathcal{G}$ *et* $X \in L^1(\mathcal{F})$ *alors* $\mathbb{E}[X\,|\mathcal{G}] = \mathbb{E}[X]$.
2. *Soit* $X \perp\!\!\!\perp Y$ *et* $f : \mathbb{R} \to \mathbb{R}$ *mesurable telles que* $\mathbb{E}[|f(X,Y)|] < +\infty$. *On pose* $g(y) = \mathbb{E}[f(X,y)]$. *Alors*
$$\mathbb{E}\big[f(X,Y)|\sigma(Y)\big] = g(Y).$$
3. *Soit* $\mathcal{G}, \mathcal{H}$ *des sous-tribus de* $\mathcal{F}$ *et* X *une variable aléatoire intégrable telles que* $\mathcal{H} \perp\!\!\!\perp \sigma(X,\mathcal{G}) := \sigma\big(\sigma(X)\cup\mathcal{G}\big)$ *alors*
$$\mathbb{E}[X\,|\sigma(\mathcal{G},\mathcal{H})] = \mathbb{E}[X\,|\mathcal{G}]. \tag{2.13}$$

Démonstration : 1) Soit $A \in \mathcal{G}$. En utilisant la définition de l'espérance conditionnelle puis l'indépendance $A \perp\!\!\!\perp X$, on a

$$\mathbb{E}\big[\mathbb{E}[X|\mathcal{G}]\mathbf{1}_A\big] = \mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[X]\mathbb{E}[\mathbf{1}_A] = \mathbb{E}\big[\mathbb{E}[X]\mathbf{1}_A\big].$$

Comme $\mathbb{E}[X]$ est bien $\mathcal{G}$-mesurable car constante, $\mathbb{E}[X]$ vérifie la Définition 2.1 de $\mathbb{E}[X|\mathcal{G}]$.

2) D'abord, on note que g est bien une fonction mesurable par le théorème de Fubini (Th. 0.5). Ainsi $g(Y)$ est $\sigma(Y)$-mesurable, d'où (i) dans la Définition 2.1. Ensuite pour (ii) dans la Définition 2.1, on considère $A \in \sigma(Y)$. Cet ensemble s'écrit $A = Y^{-1}(C)$ pour un certain $C \in \mathcal{B}(\mathbb{R})$ et on a alors

$$\mathbb{E}[f(X,Y)\mathbf{1}_A] \quad = \quad \mathbb{E}[f(X,Y)\mathbf{1}_C(Y)]$$

$$
\begin{aligned}
&= \int\int \mathbf{1}_C(y) f(x,y)\, \mathbb{P}_{(X,Y)}(dx,dy) \quad \text{(transfert)}\\
&= \int\int \mathbf{1}_C(y) f(x,y)\, \mathbb{P}_X(dx)\mathbb{P}_Y(dy) \quad (X \perp\!\!\!\perp Y)\\
&= \int \mathbf{1}_C(y) \left(\int f(x,y)\, \mathbb{P}_X(dx)\right) \mathbb{P}_Y(dy) \quad \text{(Fubini)}\\
&= \int \mathbf{1}_C(y) g(y)\, \mathbb{P}_Y(dy) \quad \text{(définition de } g)\\
&= \mathbb{E}[g(Y)\mathbf{1}_C(Y)] \quad \text{(transfert)}\\
&= \mathbb{E}[g(Y)\mathbf{1}_A].
\end{aligned}
$$

3) On utilise un argument de classe monotone (Th. 0.2). On note $Y = \mathbb{E}[X|\mathcal{G}]$ et on pose

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

L'ensemble $\mathcal{M}$ est une classe monotone car
- — $\Omega \in \mathcal{M}$ puisque $\mathbb{E}[X] = \mathbb{E}[Y]$;
- — Si $A, B \in \mathcal{M}$ avec $B \subset A$ alors $A \setminus B \in \mathcal{M}$ car $\mathbf{1}_{A\setminus B} = \mathbf{1}_A - \mathbf{1}_B$ et

$$
\begin{aligned}
\mathbb{E}\big[Y\mathbf{1}_{A\setminus B}\big] &= \mathbb{E}\big[Y(\mathbf{1}_A - \mathbf{1}_B)\big] = \mathbb{E}\big[Y\mathbf{1}_A\big] - \mathbb{E}\big[Y\mathbf{1}_B\big]\\
&= \mathbb{E}\big[X\mathbf{1}_A\big] - \mathbb{E}\big[X\mathbf{1}_B\big] = \mathbb{E}\big[X(\mathbf{1}_A - \mathbf{1}_B)\big]\\
&= \mathbb{E}\big[X\mathbf{1}_{A\setminus B}\big];
\end{aligned}
$$

- — Si $A_j \in \mathcal{M}$ avec $A_j \subset A_{j+1}$, $j \geq 1$, alors $\bigcup_{j\geq 1} A_j \in \mathcal{M}$ car

$$
\begin{aligned}
\mathbb{E}\big[Y\mathbf{1}_{\bigcup_{j\geq1} A_j}\big] &= \mathbb{E}\big[Y \lim_{n\to+\infty} \mathbf{1}_{A_n}\big] = \lim_{n\to+\infty} \mathbb{E}\big[Y\mathbf{1}_{A_n}\big] \quad \text{(convergence dominée)}\\
&= \lim_{n\to+\infty} \mathbb{E}\big[X\mathbf{1}_{A_n}\big] = \mathbb{E}\big[X \lim_{n\to+\infty} \mathbf{1}_{A_n}\big] \quad \text{(convergence dominée)}\\
&= \mathbb{E}\big[X\mathbf{1}_{\bigcup_{j\geq1} A_j}\big].
\end{aligned}
$$

Par Définition 2.1 de Y, on a $\mathcal{G} \subset \mathcal{M}$. On a aussi $\mathcal{H} \subset \mathcal{M}$ car pour $A \in \mathcal{H}$ par indépendance $\mathcal{H} \perp\!\!\!\perp \mathcal{G} \ni Y$ on a $\mathbb{E}[Y\mathbf{1}_A] = \mathbb{E}[Y]\mathbb{P}(A)$ et par indépendance $\mathcal{H} \perp\!\!\!\perp \sigma(X)$ on a aussi $\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[X]\mathbb{P}(A)$, d'où l'égalité puisque $\mathbb{E}[Y] = \mathbb{E}[X]$ et $A \in \mathcal{M}$.

On note maintenant que $\mathcal{P} = \{B \cap C : B \in \mathcal{G}, C \in \mathcal{H}\}$ est un π-système (stable par intersection finie) : si $B_1 \cap C_1, B_2 \cap C_2 \in \mathcal{P}$ alors

$$(B_1 \cap C_1) \cap (B_2 \cap C_2) = (B_1 \cap B_2) \cap (C_1 \cap C_2) \in \mathcal{P}$$

car $B_1 \cap B_2 \in \mathcal{G}$ et $C_1 \cap C_2 \in \mathcal{H}$.

On a $\mathcal{P} \subset \mathcal{M}$, en effet pour $A = B \cap C \in \mathcal{P}$ avec $B \in \mathcal{G}$ et $C \in \mathcal{H}$ par indépendance $\mathcal{H} \perp\!\!\!\perp \sigma(X, \mathcal{G})$, il vient :

$$\mathbb{E}[Y\mathbf{1}_A] = \mathbb{E}[Y\mathbf{1}_B\mathbf{1}_C] = \mathbb{E}[Y\mathbf{1}_B]\,\mathbb{E}[\mathbf{1}_C] \quad (C \in \mathcal{H} \perp\!\!\!\perp \mathcal{G} \ni Y, B)$$

$$\begin{aligned}
&= \mathbb{E}[X\mathbf{1}_B]\,\mathbb{E}[\mathbf{1}_C] \quad (\text{définition de } Y = \mathbb{E}[X\,|\mathcal{G}])\\
&= \mathbb{E}[X\mathbf{1}_B\mathbf{1}_C] \quad (C \in \mathcal{H} \perp\!\!\!\perp \sigma(X,\mathcal{G}) \ni X, B)\\
&= \mathbb{E}[X\mathbf{1}_A].
\end{aligned}$$

Par le théorème de classe monotone (Th. 0.2), on a $\sigma(\mathcal{P}) \subset \mathcal{M}$.

On conclut en justifiant $\sigma(\mathcal{P}) = \sigma(\mathcal{G},\mathcal{H})$. En effet, on a $\mathcal{P} \subset \sigma(\mathcal{G},\mathcal{H})$ car $P \in \mathcal{P}$ s'écrit $P = B \cap C \in \sigma(\mathcal{G},\mathcal{H})$ puisque $B \in \mathcal{G} \subset \sigma(\mathcal{G},\mathcal{H})$, $C \in \mathcal{H} \subset \sigma(\mathcal{G},\mathcal{H})$. On a donc $\sigma(\mathcal{P}) \subset \sigma(\mathcal{G},\mathcal{H})$. Puis comme $\mathcal{G} \subset \mathcal{P}$ et $\mathcal{H} \subset \mathcal{P}$, on a $(\mathcal{G} \cup \mathcal{H}) \subset \mathcal{P}$ et

$$\sigma(\mathcal{G},\mathcal{H}) = \sigma\big(\mathcal{G} \cup \mathcal{H}\big) \subset \sigma(\mathcal{P})$$

et finalement $\sigma(\mathcal{P}) = \sigma(\mathcal{G},\mathcal{H})$.

La conclusion $\sigma(\mathcal{G},\mathcal{H}) \subset \mathcal{M}$ signifie alors que $\forall A \in \sigma(\mathcal{G},\mathcal{H})$ on a $\mathbb{E}[Y\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A]$ et donc $Y = \mathbb{E}[X\,|\sigma(\mathcal{G},\mathcal{H})]$, ce qui est la conclusion (2.13) cherchée. □

Remarque 2.23 (Contre-exemple) Attention, dans le 3) du Th. 2.22 il est insuffisant de supposer seulement $\mathcal{H} \perp\!\!\!\perp \mathcal{G}$ et $\mathcal{H} \perp\!\!\!\perp \sigma(X)$: en effet, avec $\varepsilon_i = \pm 1$, $i = 1, 2$, indépendantes, de loi $\mathbb{P}(\varepsilon_i = 1) = \mathbb{P}(\varepsilon_i = -1) = 1/2$, prendre $X = \varepsilon_1\varepsilon_2$ et $\mathcal{H} = \sigma(\varepsilon_1)$, $\mathcal{G} = \sigma(\varepsilon_2)$: on a X de loi de Rademacher $\mathcal{R}(1/2)$, ie. $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$ et

- $\mathcal{H} \perp\!\!\!\perp \mathcal{G}$,
- $\mathcal{H} \perp\!\!\!\perp \sigma(X)$ car par exemple

$$\begin{aligned}
\mathbb{P}(\varepsilon_1 = 1, \varepsilon_1\varepsilon_2 = 1) &= \mathbb{P}(\varepsilon_1 = 1, \varepsilon_2 = 1) = \mathbb{P}(\varepsilon_1 = 1)\mathbb{P}(\varepsilon_2 = 1) = \frac{1}{4}\\
&= \mathbb{P}(\varepsilon_1 = 1)\mathbb{P}(\varepsilon_1\varepsilon_2 = 1),
\end{aligned}$$

- mais on n'a pas $\mathcal{H} \perp\!\!\!\perp \sigma(X,\mathcal{G})$ puisque $\varepsilon_1 = X/\varepsilon_2$ est $\sigma(X,\mathcal{G})$-mesurable.

Dans ce cas, (2.13) ne tient effectivement pas puisque

$$\begin{aligned}
\mathbb{E}[X\,|\mathcal{G}] &= \mathbb{E}[\varepsilon_1\varepsilon_2\,|\varepsilon_2] = \varepsilon_2\mathbb{E}[\varepsilon_1] = 0\\
\mathbb{E}[X\,|\sigma(\mathcal{G},\mathcal{H})] &= \mathbb{E}[\varepsilon_1\varepsilon_2\,|\sigma(\varepsilon_1,\varepsilon_2)] = \varepsilon_1\varepsilon_2.
\end{aligned}$$

2.4 Espérance conditionnelle dans le cas L^2

Interprétation géométrique

Dans le cadre L^2, l'espérance conditionnelle s'interprète comme une projection L^2 pour le produit scalaire $\langle X, Y\rangle = \mathbb{E}[XY]$.

Théorème 2.24 (Espérance conditionnelle et projection L^2) *Soit $X \in L^2(\mathcal{F})$ et $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. L'espérance conditionnelle $\mathbb{E}[X|\mathcal{G}]$ est la projection de X sur $L^2(\mathcal{G})$:*

$$\mathbb{E}[X|\mathcal{G}] = P_{L^2(\mathcal{G})}(X). \tag{2.14}$$

Démonstration : Notons $Y = \mathbb{E}[X|\mathcal{G}]$. D'après la Prop. 2.20, comme $X \in L^2(\mathcal{F})$, on a aussi $Y \in L^2(\mathcal{G})$. Pour $Z \in L^2(\mathcal{G})$, il vient

$$\begin{aligned}
\mathbb{E}\big[(X-Z)^2\big] &= \mathbb{E}\big[(X-Y+Y-Z)^2\big] \\
&= \mathbb{E}\big[(X-Y)^2\big] + 2\mathbb{E}[(X-Y)(Y-Z)] + \mathbb{E}\big[(Y-Z)^2\big] \\
&= \mathbb{E}\big[(X-Y)^2\big] + \mathbb{E}\big[(Y-Z)^2\big],
\end{aligned}$$

car $Y-Z$ étant $\mathcal{G}$-mesurable

$$\mathbb{E}[(X-Y)(Y-Z)] = \mathbb{E}\big[\mathbb{E}[(X-Y)(Y-Z)\,|\mathcal{G}]\big] = \mathbb{E}\big[\overbrace{\mathbb{E}[X-Y|\mathcal{G}]}^{=0}(Y-Z)\big] = 0.$$

Ainsi

$$d\big(X, L^2(\mathcal{G})\big) = \inf_{Z\in L^2(\mathcal{G})} \mathbb{E}\big[(X-Z)^2\big]^{1/2} = \mathbb{E}\big[(X-Y)^2\big]$$

est atteint en $Z = Y \in L^2(\mathcal{G})$. Cela justifie que Y est la projection $P_{L^2(\mathcal{G})}(X)$ de $X \in L^2(\mathcal{F})$ sur $L^2(\mathcal{G})$. □

Variance conditionnelle

On définit de la même façon d'autres quantités conditionnelles telles que la variance conditionnelle $\mathrm{Var}(X|\mathcal{G})$.

Définition 2.25 (Variance conditionnelle) *La variance conditionnelle de $X \in L^2(\mathcal{F})$ sachant $\mathcal{G}$ est définie par*

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

On verra qu'il s'agit de la variance par rapport à la loi conditionnelle $\mathbb{P}(\cdot|\mathcal{G})$. Comme dans le cas usuel, on a l'identité de König [1] :

Proposition 2.26 (König) *Pour X variable aléatoire L^2 et une sous-tribu $\mathcal{G}$, on a :*

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

Démonstration : En effet,

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

□

Par le théorème de Pythagore, on a la décomposition de la variance sous la forme :

1. Johann Samuel König (Allemand, 1712–1757)

Théorème 2.27 (Décomposition de la variance) *Soit $X \in L^2(\mathcal{F})$ et $\mathcal{G}$ une sous-tribu de $\mathcal{F}$. Alors, on a*

$$\mathrm{Var}(X) = \mathbb{E}\big[\,\mathrm{Var}(X\,|\mathcal{G})\big] + \mathrm{Var}\,\big(\mathbb{E}[X\,|\mathcal{G}]\big). \tag{2.15}$$

Démonstration : En utilisant la Prop. 2.26, on a :

$$\begin{aligned}
\mathbb{E}\big[\,\mathrm{Var}(X\,|\mathcal{G})\big] &= \mathbb{E}\big[\mathbb{E}[X^2|\mathcal{G}] - \mathbb{E}[X|\mathcal{G}]^2\big] = \mathbb{E}[X^2] - \mathbb{E}\big[\mathbb{E}[X|\mathcal{G}]^2\big] \\
\mathrm{Var}\,\big(\mathbb{E}[X\,|\mathcal{G}]\big) &= \mathbb{E}\big[\mathbb{E}[X\,|\mathcal{G}]^2\big] - \mathbb{E}\big[\mathbb{E}[X\,|\mathcal{G}]\big]^2 = \mathbb{E}\big[\mathbb{E}[X\,|\mathcal{G}]^2\big] - \mathbb{E}[X]^2
\end{aligned}$$

dont on déduit (2.15) par addition. □

2.5 Conditionnement gaussien

Dans le cas de conditionnement gaussien, on a mieux que (2.14) : il suffit de projeter sur l'espace vectoriel engendré par les variables aléatoires (gaussiennes) qui conditionnent :

Proposition 2.28 (Espérance conditionnelle gaussienne)

1. *Soit (X, Y) un couple gaussien centré. Alors*

$$\mathbb{E}[X\,|Y] = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)} Y \qquad ps.$$

2. *Dans le cas non centré, on a*

$$\mathbb{E}[X\,|Y] = m_X - \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)} m_Y + \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)} Y.$$

3. *Soit $(Z_1, \dots, Z_d)$ un vecteur gaussien centré de covariance Σ et $(a_1, \dots, a_d)^t, (b_1, \dots, b_d)^t \in \mathbb{R}^d$. On considère $X = a_1Z_1 + \dots + a_dZ_d$ et $Y = b_1X_1 + \dots + b_dZ_d$. Alors*

$$\mathbb{E}[X\,|Y] = \frac{a^t\Sigma b}{b^t\Sigma b} Y.$$

4. *Soit $Y = (Y_1, \dots, Y_d)$ un vecteur gaussien centré de covariance Σ inversible et X variable aléatoire réelle tels que (X, Y) soit un vecteur gaussien centré. On note $d = \big(\mathrm{Cov}(X, Y_1), \dots, \mathrm{Cov}(X, Y_d)\big)^t$. Alors*

$$\mathbb{E}[X\,|Y] = \langle \Sigma^{-1}d, Y\rangle.$$

Démonstration : 1) Notons $c = \mathbb{E}[XY]/\mathbb{E}[Y^2]$. Comme (X, Y) est un vecteur gaussien centré, $(Y, X - cY)$ l'est aussi et par définition de c :

$$\mathrm{Cov}(X - cY, Y) = \mathrm{Cov}(X, Y) - c\,\mathrm{Var}(Y) = 0,$$

on a donc $Y \perp\!\!\!\perp (X - cY)$. Soit $Z \in L^2(\sigma(Y))$, d'après le théorème de Doob-Dynkin (Th. 0.4), $Z = h(Y)$ avec $h(Y) \in L^2(\sigma(Y))$. On a

$$\begin{aligned}\mathbb{E}\big[(X - cY)Z\big] &= \mathbb{E}\big[\mathbb{E}[(X - cY)h(Y)\,|Y]\big]\\ &= \mathbb{E}\big[h(Y)\mathbb{E}[(X - cY)\,|Y]\big]\\ &= \mathbb{E}\big[h(Y)\underbrace{\mathbb{E}[X - cY]}_{=0}\big] = 0\end{aligned}$$

d'après le Th. 2.22 puisque $X - cY \perp\!\!\!\perp Y$. On a donc $X - cY \perp Z$ pour tout $Z \in L^2(\sigma(Y))$, ie. $X - cY \perp L^2(\sigma(Y))$. Comme $cY \in L^2(\sigma(Y))$, on a $P_{L^2(\sigma(Y))}(X) = cY$ et d'après l'interprétation projection (2.14) de l'espérance conditionnelle dans le cadre L^2, on a bien

$$\mathbb{E}[X\,|Y] = P_{L^2(\sigma(Y))}(X) = cY.$$

2) Le cas non centré se déduit de 1) appliqué à $X - m_X$ et $Y - m_Y$.

3) En notant $\langle a, b\rangle = \sum_{i=1}^d a_i b_i$ le produit scalaire euclidien, on a $X = \langle a, Z\rangle$ et $Y = \langle b, Z\rangle$. Le vecteur (X, Y) est gaussien car image linéaire de Z. En notant Σ la matrice de covariance de Z, on a

$$\begin{aligned}\mathrm{Var}(X) &= a^t\Sigma a, \quad \mathrm{Var}(Y) = b^t\Sigma b,\\ \mathrm{Cov}(X, Y) &= \sum_{i,j=1}^d a_i b_j \,\mathrm{Cov}(Z_i, Z_j) = a^t\Sigma b.\end{aligned}$$

En appliquant le 1) au vecteur gaussien (X, Y), on a

$$\mathbb{E}[X\,|Y] = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(Y)}\,Y = \frac{a^t\Sigma b}{b^t\Sigma b}\,Y.$$

4) On note $c = \Sigma^{-1}d$. Le vecteur $\big(X - c^tY, Y\big)$ est gaussien car toutes combinaisons de ses marginales en est une de celles de (X, Y) donc de loi normale. Pour chaque $1 \le i \le d$, on a

$$\begin{aligned}\mathbb{E}\big[(X - c^tY)Y_i\big] &= \mathbb{E}\Big[\Big(X - \sum_{j=1}^d c_jY_j\Big)Y_i\Big] = \mathbb{E}[XY_i] - \sum_{j=1}^d c_j\mathbb{E}[Y_iY_j]\\ &= \mathrm{Cov}(X, Y_i) - (\Sigma c)_i = (d - \Sigma c)_i = 0,\end{aligned}$$

par choix de $c = \Sigma^{-1}d$. Il suit que $X - c^tY \perp\!\!\!\perp Y$. En prenant $Z = h(Y) \in L^2(\sigma(Y))$, on a

$$\begin{aligned}\mathbb{E}\big[(X - c^tY)Z\big] &= \mathbb{E}\big[\mathbb{E}[(X - c^tY)h(Y)\,|Y]\big]\\ &= \mathbb{E}\big[h(Y)\mathbb{E}[(X - c^tY)\,|Y]\big]\\ &= \mathbb{E}\big[h(Y)\underbrace{\mathbb{E}[X - c^tY]}_{=0}\big] = 0\end{aligned}$$

d'après le Th. 2.22 puisque $X - c^tY \perp\!\!\!\perp Y$. On a donc $X - c^tY \perp Z$ pour tout $Z \in L^2(\sigma(Y))$, ie. $X - c^tY \perp L^2(\sigma(Y))$. Comme $c^tY \in L^2(\sigma(Y))$, on a $P_{L^2(\sigma(Y))}(X) = c^tY$ et d'après l'interprétation projection (2.14) de l'espérance conditionnelle dans le cadre L^2, on a bien

$$\mathbb{E}[X\,|Y] = P_{L^2(\sigma(Y))}(X) = c^tY.$$

□

2.6 Lois conditionnelles

Dans cette section, on définit les lois conditionnelles. On admet cependant le résultat difficile d'existence de ces lois (dans un espace polonais) dû à Jiřina (Th. 2.34, cf. [Jir]).

Définition 2.29 (Noyau de probabilité) *Soit $(S, \mathcal{A})$ et $(T, \mathcal{B})$ deux espaces mesurables. On appelle noyau de probabilité (ou de transition) de T dans S toute application $\nu : \mathcal{A} \times T \to [0, 1]$ telle que*

(i) pour tout $y \in T$, $\nu(*, y)$ est une probabilité sur $(S, \mathcal{A})$;

(ii) pour tout $A \in \mathcal{A}$, $y \mapsto \nu(A, y)$ est $\mathcal{B}$-mesurable.

Exemple 2.30 1) Si μ est une mesure σ-finie sur $(S, \mathcal{A})$ et $h : S \times T \to \mathbb{R}_+$ est une fonction mesurable avec $\int_S h(x, y)\,\mu(dx) = 1$ pour tout $y \in T$ alors

$$\nu(A, y) = \int_A h(x, y)\,\mu(dx)$$

définit un noyau de probabilité. La σ-additivité dans (i) de la Définition 2.29 vient du théorème de convergence monotone. Le point (ii) est assuré par le théorème de Fubini-Tonelli (Th. 0.5).

2) **(Couple à densité)** Soit (X, Y) un couple aléatoire sur $\mathbb{R}^2$ de densité f. On peut appliquer le cas **1)** précédent avec $(S, \mathcal{A}) = (T, \mathcal{B}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $\mu = \lambda$ (mesure de Lebesgue) et

$$h(x, y) = \frac{f(x, y)}{f_Y(y)}$$

lorsque $f_Y(y) > 0$. On définit alors un noyau de probabilité par

$$\nu(A, y) = \int_A \frac{f(x, y)}{f_Y(y)}\,dx \tag{2.16}$$

lorsque y est dans le support de Y et $\nu(*, y) = \delta_{s_0}$ sinon pour $s_0 \in S$ quelconque.

3) **(Couple discret)** On considère (X, Y) un couple aléatoire discret. On définit un noyau de probabilité par

$$\nu(A, y) = \begin{cases} \mathbb{P}(X \in A|Y = y) = \frac{\mathbb{P}(X\in A, Y=y)}{\mathbb{P}(Y=y)} & \text{si } \mathbb{P}(Y = y) > 0 \\ \delta_{s_0}(A) & \text{sinon,} \end{cases} \tag{2.17}$$

où $s_0 \in S$ est quelconque.

Proposition 2.31 *Soit ν un noyau de probabilité.*
(1) Si h est mesurable positive (ou bornée) sur $(S,\mathcal{A})$ alors

$$\varphi(y)=\int h(x)\,\nu(dx,y),\quad y\in T, \tag{2.18}$$

est mesurable sur $(T,\mathcal{B})$.
(2) Si η est une mesure de probabilité sur $(T,\mathcal{B})$ alors

$$\mu(A)=\int \nu(A,y)\,\eta(dy),\quad A\in\mathcal{A},$$

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

Démonstration : (1) Si $h=\mathbf{1}_A$ alors $\varphi(y)=\nu(A,y)$ et la mesurabilité découle de (ii) dans la Définition 2.29. Par les arguments standards de théorie de la mesure, on étend le résultat pour h simple puis pour h mesurable positive (par approximation et convergence monotone).

(2) On a $\mu(A)\geq 0$ car $\nu(A,y)\geq 0$ et $\mu(S)=1$ car $\nu(S,y)=1$ et η est une probabilité. Enfin, μ est σ-additive en utilisant celle de $\nu(*,y)$ et le théorème de convergence monotone : pour des $A_i\in\mathcal{A}$, $i\geq 1$, deux à deux disjoints, on a

$$\begin{aligned}\mu\Big(\bigcup_{i=1}^{+\infty}A_i\Big) &= \int \nu\Big(\bigcup_{i=1}^{+\infty}A_i,y\Big)\,\eta(dy)=\int\sum_{i=1}^{+\infty}\nu(A_i,y)\,\eta(dy)\\ &= \sum_{i=1}^{+\infty}\int \nu(A_i,y)\eta(dy)=\sum_{i=1}^{+\infty}\mu(A_i).\end{aligned}$$

□

Définition 2.32 (Loi conditionnelle) *Soit X, Y des variables aléatoires à valeurs respectivement dans $(S,\mathcal{A})$ et $(T,\mathcal{B})$. On appelle loi conditionnelle de X sachant Y tout noyau de probabilité $\nu:\mathcal{A}\times T\to[0,1]$ telle que pour toute fonction h mesurable positive sur $(S,\mathcal{A})$ on a*

$$\mathbb{E}\big[h(X)\,|Y\big]=\int h(x)\ \nu(dx,Y). \tag{2.19}$$

Remarque 2.33 En utilisant la fonction φ donnée en (2.18), mesurable d'après la Prop. 2.31, on a $\mathbb{E}[h(X)\,|Y]=\varphi(Y)$. La fonction φ est donc la fonction mesurable du théorème de Doob-Dynkin (Th. 0.7) appliqué à la variable aléatoire $\mathbb{E}[h(X)\,|Y]$ qui est $\sigma(Y)$-mesurable.

Par la Définition 2.32, si ν est la loi conditionnelle de X sachant Y alors en prenant $h=\mathbf{1}_A$ dans (2.19), on a

$$\mathbb{P}\big(X\in A\,|Y\big)=\nu(A,Y)\quad \text{ps.} \tag{2.20}$$

Par (i) dans la Définition 2.29, $\mathbb{P}(X\in\cdot|Y)$ ainsi définie est bien une probabilité.

Existence et unicité de la loi conditionnelle

Existence. L'existence des lois conditionnelles est un résultat difficile. Elle est donnée par le résultat suivant dû à Jiřina (et admis). Il s'applique en particulier pour des variables aléatoires à valeurs dans $\mathbb{R}$ ou $\mathbb{R}^p$.

Théorème 2.34 (Jiřina) *Soit $(S, \mathcal{A})$ et $(T, \mathcal{B})$ deux espaces mesurables. On suppose que S est un espace polonais (métrique, complet, séparable) avec $\mathcal{A} = \mathcal{B}(S)$ (tribu borélienne). Alors il existe une loi conditionnelle de X sachant Y comme dans la Définition 2.32.*

Démonstration : Admis. □

Unicité. Si ν et ν' sont deux lois conditionnelles de X sachant Y telles que dans la Définition 2.32 alors pour tout $A \in \mathcal{A}$ (cf. (2.20)) :

$$\nu(A, Y) = \mathbb{P}\big(Y \in A\,|Y\big) = \nu'(A, Y) \quad \text{ps}$$

c'est à dire $\nu(A, y) = \nu'(A, y)$ pour $\mathbb{P}_Y$-presque chaque $y \in T$ (attention au *presque sûr* qui dépend de $A \in \mathcal{A}$). Dans le cas où les mesures de probabilités sur $(S, \mathcal{A})$ sont caractérisées par leurs valeurs sur une famille dénombrable d'évènements alors on a

$$\nu(*, y) = \nu'(*, y) \qquad \text{pour } \mathbb{P}_Y\text{-presque chaque } y \in T.$$

C'est le cas pour $(S, \mathcal{A}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$; plus généralement, c'est encore le cas lorsque S un espace polonais (métrique, complet, séparable) avec $\mathcal{A} = \mathcal{B}(S)$ (tribu borélienne). En ce sens, il y a **unicité de la loi conditionnelle de X sachant Y**.

Cadres usuels

On prolonge l'Exemple 2.30 avec les cas usuels discret et à densité. D'après le Théorème de Jiřina (Th. 2.34), dans ces cas il y a existence et unicité de la loi conditionnelle. On montre qu'alors les noyaux de probabilité sont donnés par (2.16) et (2.17).

Proposition 2.35 (Loi conditionnelle à densité) *Soit (X, Y) un couple aléatoire de densité f. Alors la loi conditionnelle de X sachant Y est donnée par le noyau de densité* (2.16).

Démonstration : On montre (2.19) avec $\nu(dx, y) = \frac{f(x,y)}{f_Y(y)}dx$ lorsque $y \in \mathcal{S}(Y)$ (probabilité quelconque sinon) en établissant que pour $A \in \sigma(Y)$, on a

$$\mathbb{E}\big[h(X)\mathbf{1}_A\big] = \mathbb{E}\Big[\Big(\int h(x)\ \nu(dx, Y)\Big)\mathbf{1}_A\Big].$$

Comme $A \in \sigma(Y)$ s'écrit $A = Y^{-1}(B)$ et $\mathbf{1}_A = \mathbf{1}_B(Y)$, on a :

$$\mathbb{E}\Big[\Big(\int h(x)\ \nu(dx, Y)\Big)\mathbf{1}_A\Big] \;=\; \mathbb{E}\Big[\Big(\int h(x)\ \nu(dx, Y)\Big)\mathbf{1}_B(Y)\Big]$$

$$\begin{aligned}
&= \int \Big(\int h(x)\ \nu(dx,y)dx\Big)\mathbf{1}_B(y)f_Y(y)dy \quad \text{(formule de transfert)}\\
&= \int_{\mathcal{S}(Y)} \Big(\int h(x)\ \frac{f(x,y)}{f_Y(y)}dx\Big)\mathbf{1}_B(y)f_Y(y)dy \quad \text{(définition de } \nu)\\
&= \int h(x)\mathbf{1}_B(y)f(x,y)\ dxdy \quad \text{(théorème de Fubini)}\\
&= \mathbb{E}\big[h(X)\mathbf{1}_B(Y)\big] = \mathbb{E}\big[h(X)\mathbf{1}_A\big].
\end{aligned}$$

Cela justifie que $\varphi(Y) = \int h(x)\ \nu(dx,Y)$ vérifie le ii) de la Définition 2.1 de $\mathbb{E}[h(X)|Y]$. Comme $\varphi(Y)$ est $\sigma(Y)$-mesurable, le i) est immédiat et on a bien $\varphi(Y) = \mathbb{E}[h(X)|Y]$, prouvant que $\nu(*,Y)$ est la loi conditionnelle de X sachant Y (Définition 2.32). □

Proposition 2.36 (Loi conditionnelle discrète) *Soit (X,Y) un couple aléatoire discret. Alors la loi conditionnelle de X sachant Y est donnée par le noyau discret (2.17).*

Démonstration : On montre (2.19) avec $\nu(x,y) = \frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(Y=y)}$ lorsque $\mathbb{P}(Y=y)>0$ (probabilité quelconque sinon) en établissant que pour $A\in\sigma(Y)$

$$\mathbb{E}\big[h(X)\mathbf{1}_A\big] = \mathbb{E}\Big[\Big(\int h(x)\ \nu(dx,Y)\Big)\mathbf{1}_A\Big].$$

Comme $A\in\sigma(Y)$ s'écrit $A = Y^{-1}(B)$ et $\mathbf{1}_A = \mathbf{1}_B(Y)$, on a :

$$\begin{aligned}
\mathbb{E}\Big[\Big(\int h(x)\ \nu(dx,Y)\Big)\mathbf{1}_A\Big] &= \mathbb{E}\Big[\Big(\int h(x)\ \nu(dx,Y)\Big)\mathbf{1}_B(Y)\Big]\\
&= \sum_{y\in\mathcal{S}(Y)} \Big(\int h(x)\ \nu(dx,y)\Big)\mathbf{1}_B(y)\mathbb{P}(Y=y)\\
&= \sum_{y\in\mathcal{S}(Y)} \Big(\sum_{x\in\mathcal{S}(X)} h(x)\frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(Y=y)}\Big)\mathbf{1}_B(y)\mathbb{P}(Y=y)\\
&= \sum_{(x,y)\in\mathcal{S}(X)\times\mathcal{S}(Y)} h(x)\mathbf{1}_B(y)\mathbb{P}(X=x,Y=y)\\
&= \mathbb{E}\big[h(X)\mathbf{1}_B(Y)\big] = \mathbb{E}\big[h(X)\mathbf{1}_A\big].
\end{aligned}$$

Commme précédemment, cela justifie successivement que $\varphi(Y) = \mathbb{E}[h(X)|Y]$, et que $\nu(*,Y)$ est la loi conditionnelle de X sachant Y (Définition 2.32). □

Conditionnement par $Y = y$

On revient au conditionnement par un évènement comme dans le Chapitre 1 et on donne un sens général à des probabilités conditionnelles du type $\mathbb{P}(X\in A|Y=y)$, même lorsque $\mathbb{P}(Y=y)=0$.

Définition 2.37 (Probabilité conditionnelle sachant $Y = y$) *Lorsque la loi conditionnelle ν de X sachant Y existe, on pose pour $\mathbb{P}_Y$-presque chaque y :*

$$\mathbb{P}(X \in A|Y = y) = \nu(A, y). \tag{2.21}$$

Lorsque Y est une variable aléatoire discrète, on a commencé par voir au Chapitre 1 avec la Définition 1.11 que

$$\begin{aligned} \mathbb{P}(X \in A|Y) = \mathbb{E}[\mathbf{1}_A(X)|Y] &= \sum_{y \in Y(\Omega)} \mathbb{E}[\mathbf{1}_A(X)|Y = y]\mathbf{1}_{\{Y=y\}} \\ &= \sum_{y \in Y(\Omega)} \mathbb{P}(X \in A|Y = y)\mathbf{1}_{\{Y=y\}}, \end{aligned}$$

si bien que $\mathbb{P}(X \in A|Y) = \mathbb{P}(X \in A|Y = y)$ sur $\{Y = y\}$, ce qui se retrouve avec le cadre plus général donné par (2.20) et (2.21) :

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

Dans le cas où $\mathbb{P}(Y = y) > 0$, la définition de $\mathbb{P}(*|Y = y)$ coïncide donc avec la définition élémentaire du Chapitre 1, cf. (2.17).

Lorsque $\mathbb{P}(Y = y) = 0$, le conditionnement par $\{Y = y\}$ n'est pas bien défini dans le Chapitre 1 (cf. par exemple $\mathbb{P}(X \in A\,|Y = y)$ en (1.11)) et on parle de **conditionnement singulier**.

Avec la Définition 2.37, on définit aussi :

Définition 2.38 (Loi conditionnelle sachant $Y = y$) *On définit la loi conditionnelle de X sachant $Y = y$ pour $\mathbb{P}_Y$-presque chaque y par :*

$$\mathcal{L}\big(X|Y = y\big) = \nu(*, y).$$

On peut alors définir les espérances conditionnelles sachant $Y = y$ comme l'espérance par rapport à la loi conditionnelle sachant $Y = y$, ie.

$$\mathbb{E}\big[h(X)|Y = y\big] = \int h(x)\,\mathbb{P}_X(dx\,|Y = y).$$

De plus, compte tenu de (2.20) et (2.21), on observe que si

$$\mathbb{E}\big[h(X)\,|Y = y\big] = \varphi(y)$$

alors

$$\mathbb{E}\big[h(X)\,|Y\big] = \varphi(Y).$$

Cette observation justifie que pour calculer $\mathbb{E}[h(X)|Y]$, on peut faire le calcul comme si Y était figé en y avec la loi conditionnelle de X sachant $Y = y$, une fois le résultat $\varphi(y)$ obtenu, on a le résultat final $\varphi(Y)$ en reprenant Y à la place de y.

Proposition 2.39 *Soit X, Y deux variables aléatoires. On a $X \perp\!\!\!\perp Y$ si et seulement si la loi conditionnelle $\mathcal{L}(X|Y=y)$ de X sachant $Y=y$ existe pour $\mathbb{P}_Y$-presque chaque y et ne dépend pas de y.*

Démonstration : On suppose d'abord $X \perp\!\!\!\perp Y$. On a

$$\mathbb{E}\big[h(X)\,|Y\big] = \mathbb{E}[h(X)] = \int h(x)\,\mathbb{P}_X(dx)$$

ce qui justifie $\mathcal{L}\big(X|Y\big) = \nu(\cdot, Y) = \mathbb{P}_X$ et, pour tout y, $\mathcal{L}\big(X|Y=y\big) = \nu(\cdot, y) = \mathbb{P}_X$, ce qui prouve le sens direct.

On suppose ensuite que la loi $\mathcal{L}(X|Y=y)$ de X sachant $Y=y$ existe pour $\mathbb{P}_Y$-presque chaque y et ne dépend pas de y, ie. $\mathcal{L}(X|Y=y) = \nu(*, y) = \nu(*)$, on a alors pour tout $A, B \in \mathcal{B}(\mathbb{R})$:

$$\begin{aligned}
\mathbb{P}(X \in A, Y \in B) &= \mathbb{E}[\mathbf{1}_A(X)\mathbf{1}_B(Y)] = \mathbb{E}\big[\mathbb{E}[\mathbf{1}_A(X)\mathbf{1}_B(Y)\,|Y]\big] = \mathbb{E}\big[\mathbb{E}[\mathbf{1}_A(X)\,|Y]\mathbf{1}_B(Y)\big] \\
&= \int \nu(A, y)\mathbf{1}_B(y)\,\mathbb{P}_Y(dy) = \int \nu(A)\mathbf{1}_B(y)\,\mathbb{P}_Y(dy) \\
&= \nu(A)\int \mathbf{1}_B(y)\,\mathbb{P}_Y(dy) = \nu(A)\,\mathbb{P}(Y \in B),
\end{aligned}$$

ce qui justifie $X \perp\!\!\!\perp Y$. □

Désintégration et Fubini conditionnel

Proposition 2.40 (Désintégration d'une loi) *Soit X, Y des variables aléatoires à valeurs respectivement dans $(S, \mathcal{A})$ et $(T, \mathcal{B})$ telle que la loi conditionnelle $\mathbb{P}(X \in *|Y=\cdot)$ est bien définie comme en Définition 2.32 avec un noyau de transition $\nu(*, \cdot)$ comme en Définition 2.29. Alors, pour tout $A \in \mathcal{A}$ et $B \in \mathcal{B}$ on a*

$$\mathbb{P}(X \in A, Y \in B) = \int_B \nu(A, y)\,\mathbb{P}_Y(dy) = \int_B \mathbb{P}(X \in A|Y=y)\,\mathbb{P}_Y(dy). \tag{2.22}$$

Démonstration : Pour $A \in \mathcal{A}$ et $B \in \mathcal{B}$, on a

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

en utilisant la notation (2.21) pour $\mathbb{P}(X \in A|Y=y)$. □

Plus généralement, avec les arguments standards de théorie de la mesure, on montre :

Théorème 2.41 (Fubini conditionnel) *Soit (X, Y) un couple aléatoire à valeurs dans $(S \times T, \mathcal{A} \otimes \mathcal{B})$ telle que la loi conditionnelle $\mathbb{P}(X \in * | Y = \cdot)$ est bien définie comme en Définition 2.32. Alors*

(1) Pour $f : (S \times T, \mathcal{A} \otimes \mathcal{B}) \to \mathbb{R}_+$ mesurable (positive), $y \mapsto \int_S f(x, y)\, \mathbb{P}_X(dx\, |Y = y)$ est mesurable et

$$\int_{S\times T} f(x, y)\, \mathbb{P}_{(X,Y)}(dx, dy) = \int_T \left(\int_S f(x, y)\, \mathbb{P}_X(dx\, |Y = y) \right) \mathbb{P}_Y(dy). \tag{2.23}$$

(2) Pour $f : (S \times T, \mathcal{A} \otimes \mathcal{B}) \to \mathbb{R}$ mesurable de signe quelconque, $\mathbb{P}_{(X,Y)}$-intégrable, alors pour $\mathbb{P}_Y$-presque chaque $y \in T$, l'application $f(\cdot, y)$ est $\mathbb{P}_X(*\, |Y = y)$-intégrable et l'application $y \mapsto \int_S f(x, y)\, \mathbb{P}_X(dx\, |Y = y)$ est $\mathbb{P}_Y$-intégrable avec l'égalité (2.23) encore vraie.

En notant

$$\varphi_f(y) = \int_S f(x, y)\, \mathbb{P}_X(dx\, |Y = y),$$

(2.23) assure que pour tout $B = Y^{-1}(C) \in \sigma(Y)$

$$\begin{aligned} \mathbb{E}[\mathbf{1}_B f(X, Y)] &= \mathbb{E}[\mathbf{1}_C(Y) f(X, Y)] = \int_T \mathbf{1}_C(y) \varphi_f(y)\, \mathbb{P}_Y(dy) \\ &= \mathbb{E}[\mathbf{1}_C(Y) \varphi_f(Y)] = \mathbb{E}[\mathbf{1}_B \varphi_f(Y)]. \end{aligned}$$

Comme $\varphi_f(Y)$ est $\sigma(Y)$-mesurable, on a

$$\mathbb{E}[f(X, Y)\ |Y] = \varphi_f(Y) = \left(\int_S f(x, y)\, \mathbb{P}_X(dx\, |Y = y) \right)_{y=Y} = \int f(x, Y)\, \nu(dx, Y).$$

Par conséquent lorsque $h(f(X, Y)) \in L^1$, on a par la formule de transfert

$$\begin{aligned} \mathbb{E}[h(f(X, Y))\, |Y] &= \int h(f(x, Y))\, \nu(dx, Y) \\ &= \int h(u) \big(\nu(*, Y) \circ f(*, Y)^{-1}\big)(dx) \end{aligned}$$

donc, d'après la Définition 2.32, la loi conditionnelle de $f(X, Y)$ sachant Y est

$$\mathcal{L}\big(f(X, Y)\, |Y\big) = \nu(*, Y) \circ f(*, Y)^{-1}. \tag{2.24}$$

De la même façon, par la formule de transfert

$$\begin{aligned} \mathbb{E}[h(f(X, y))\, |Y] &= \int h(f(x, y))\, \nu(dx, Y) \\ &= \int h(u) \big(\nu(*, Y) \circ f(*, y)^{-1}\big)(dx) \end{aligned}$$

et, d'après la Définition 2.32, la loi conditionnelle de $f(X, y)$ sachant Y est

$$\mathcal{L}\big(f(X, y)\, |Y\big) = \nu(*, Y) \circ f(*, y)^{-1}. \tag{2.25}$$

En comparant (2.24) et (2.25), compte tenu de la notation (2.21), on a prouvé :

Proposition 2.42 (Transfert conditionnel) *Sous les mêmes conditions que dans le Th. 2.41, on a :*

$$\mathbb{P}\big(f(X,Y) \in * \,|Y=y\big) = \mathbb{P}\big(f(X,y) \in * \,|Y=y)$$

ou

$$\mathcal{L}\big(f(X,Y)\,|Y=y\big) = \mathcal{L}\big(f(X,y)\,|Y=y\big).$$

Deuxième partie

Martingales

Chapitre 3

Martingales et filtrations

Dans ce chapitre, on introduit la notion de martingale. On commence par les notions de filtration en Section 3.1 et de temps d'arrêt en Section 3.2 avant de définir les martingales en Section 3.3 et d'en donner les premières propriétés en Section 3.4. On termine ce chapitre avec la notion de martingale arrêtée en Section 3.5.

Dans la suite, on considère un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

3.1 Filtration et mesurabilité

Filtration

Définition 3.1 (Filtration) *Soit $(\mathcal{F}_n)_{n\geq 0}$ une suite de sous-tribus de $\mathcal{F}$. On dit que $(\mathcal{F}_n)_{n\geq 0}$ est une filtration lorsque pour tout $n \geq 0$, on a $\mathcal{F}_n \subset \mathcal{F}_{n+1}$.*
Un espace de probabilité muni d'une filtration $\big(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P}\big)$ s'appelle un espace de probabilité filtré.

Définition 3.2 (Adapté) *On dit qu'une suite $(X_n)_{n\geq 0}$ est adaptée par rapport à une filtration $(\mathcal{F}_n)_{n\geq 0}$ si pour tout $n \geq 0$, X_n est $\mathcal{F}_n$-mesurable.*

Exemple 3.3 (Filtration canonique) Si $(X_n)_{n\geq 1}$ est une suite de variables aléatoires, on appelle filtration canonique ou naturelle la filtration $(\mathcal{F}_n)_{n\geq 1}$ des tribus engendrées par ces variables aléatoires :

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

Il s'agit de la plus petite tribu rendant chaque X_i mesurable pour $1 \leq i \leq n$. Si besoin, on complète la filtration par $\mathcal{F}_0 = \{\emptyset, \Omega\}$ (en général associé à un choix de X_0 constante). On parle aussi de la filtration engendrée par la suite $(X_n)_{n\geq 1}$. Par construction, la suite $(X_n)_{n\geq 1}$ est adaptée par rapport à sa filtration naturelle.

Exemple 3.4 (Filtration dyadique) Soit $(\Omega, \mathcal{F}, \mathbb{P}) = \big(]0,1[, \mathcal{B}(]0,1[), \lambda\big)$ où λ est la mesure de Lebesgue sur $]0,1[$ (probabilité uniforme). On pose

$$\mathcal{F}_n = \sigma\Big(\Big[\frac{i-1}{2^n}, \frac{i}{2^n}\Big[\ : i = 1, \dots, 2^n\Big), \quad n \geq 0,$$

et $(\mathcal{F}_n)_{n\geq 0}$ s'appelle alors la filtration dyadique de $[0,1]$. On a bien une filtration car comme pour tout $n \geq 1$ et $1 \leq i \leq 2^n$:

$$\Big[\frac{i-1}{2^n}, \frac{i}{2^n}\Big[= \Big[\frac{2i-2}{2^{n+1}}, \frac{2i-1}{2^{n+1}}\Big[\cup \Big[\frac{2i-1}{2^{n+1}}, \frac{2i}{2^{n+1}}\Big[,$$

on a

$$\mathcal{D}_n := \Big\{\Big[\frac{i-1}{2^n}, \frac{i}{2^n}\Big[\ : i = 1, \dots, 2^n\Big\} \subset \sigma(\mathcal{D}_{n+1}),$$

et donc $\mathcal{F}_n = \sigma(\mathcal{D}_n) \subset \mathcal{F}_{n+1} = \sigma(\mathcal{D}_{n+1})$.

Prévisibilité

Définition 3.5 (Prévisibilité) *Une suite de variables aléatoires $(H_n)_{n\geq 1}$ est dite prévisible pour une filtration $(\mathcal{F}_n)_{n\geq 0}$ si, pour tout $n \geq 1$, H_n est $\mathcal{F}_{n-1}$-mesurable.*

Remarque 3.6 (Interprétation sous forme d'information) En interprétant une sous-tribu $\mathcal{G} \subset \mathcal{F}$ comme une quantité d'information, il faut comprendre la $\mathcal{G}$-mesurabilité d'une variable aléatoire X comme la connaissance de cette variable aléatoire : X $\mathcal{G}$-mesurable est connue dès lors que la sous-tribu $\mathcal{G}$ l'est.

Ainsi on peut interpréter une filtration comme une quantité d'information qui évolue au cours du temps : $\mathbb{N}$ est le temps et $\mathcal{F}_n$ est l'information disponible à la date n.

Une suite $(X_n)_{n\geq 1}$ est alors adaptée si X_n est connue à la date n. Une suite $(H_n)_{n\geq 1}$ est prévisible si H_n peut être prédite avec l'information $\mathcal{F}_{n-1}$ disponible à la date $n-1$.

Des exemples typiques de suites prévisibles sont donnés avec la notion de temps d'arrêt qui suit, cf. (3.1).

3.2 Temps d'arrêt

La notion de temps d'arrêt est associée à une filtration $(\mathcal{F}_n)_{n\geq 0}$ qu'on commence par se fixer.

Définition 3.7 (Temps d'arrêt) *Une variable aléatoire T à valeurs dans $\mathbb{N} \cup \{+\infty\}$ est un $(\mathcal{F}_n)$-temps d'arrêt si pour tout $n \geq 0$ on a $\{T \leq n\} \in \mathcal{F}_n$.*

Remarque 3.8 À chaque date $n \geq 0$, on sait si la date aléatoire T est échue ou pas.

Exemple 3.9 (Temps d'arrêt)

1. Si T est constant égale à n_0 alors T est un temps d'arrêt.
2. Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires et $(\mathcal{F}_n)_{n\geq 1}$ sa filtration naturelle.
   — (Temps d'atteinte) $T = \min\big(i \geq 0 : X_i \in A\big)$ est un $(\mathcal{F}_n)$-temps d'arrêt pour $A \in \mathcal{B}(\mathbb{R})$. En effet
   $$\big\{T \leq n\big\} = \bigcup_{k=0}^{n} \big\{X_k \in A\big\} \in \mathcal{F}_n$$
   puisque $\{X_k \in A\} \in \mathcal{F}_k \subset \mathcal{F}_n$, $0 \leq k \leq n$.
   — Attention, $T = \max\big(i \geq 0 : X_i \in A\big)$ n'est pas un temps d'arrêt par rapport à la filtration naturelle. Par exemple,
   $$\big\{T = n\big\} = \big\{X_n \in A, X_{n+1} \notin A, X_{n+2} \notin A, \dots\big\} \notin \sigma\big(X_1, \dots, X_n\big) = \mathcal{F}_n.$$

Remarque 3.10 1. $\{T = n\} = \{T \leq n\} \setminus \{T \leq n-1\} \in \mathcal{F}_n$;
2. $\{T \geq n\} = \{T \leq n-1\}^c \in \mathcal{F}_{n-1}$;
3. Étant donné un $(\mathcal{F}_n)$-temps d'arrêt T, on définit une suite prévisible par
$$H_n = \mathbf{1}_{\{T\geq n\}}, \quad n \geq 0. \tag{3.1}$$

Proposition 3.11 (Propriétés des temps d'arrêt)

(1) T est un temps d'arrêt si et seulement si pour tout $n \geq 0$ on a $\{T = n\} \in \mathcal{F}_n$.

(2) Si T et S sont des $(\mathcal{F}_n)$-temps d'arrêt. Alors $T \wedge S$, $T \vee S$, $T + S$ en sont aussi.

(3) Si T est un $(\mathcal{F}_n)$-temps d'arrêt alors pour tout $k \geq 0$, $T \wedge k$ en est un aussi.

(4) Si $(T_p)_{p\geq 1}$ est une suite monotone de $(\mathcal{F}_n)$-temps d'arrêt alors $T = \lim_{p\to+\infty} T_p$ est aussi un temps d'arrêt.

(5) Soit $(T_p)_{p\geq 1}$ est une suite de $(\mathcal{F}_n)$-temps d'arrêt alors
$$\inf_{p\geq 1} T_p,\ \sup_{p\geq 1} T_p,\ \liminf_{p\to+\infty} T_p,\ \limsup_{p\to+\infty} T_p$$
sont des $(\mathcal{F}_n)$-temps d'arrêt.

Démonstration : **1)** Si T est un temps d'arrêt vérifiant la Définition 3.7 alors $\{T = n\} = \{T \leq n\} \setminus \{T \leq n-1\} \in \mathcal{F}_n$ puisque $\{T \leq n\} \in \mathcal{F}_n$ et $\{T \leq n-1\} \in \mathcal{F}_{n-1} \subset \mathcal{F}_n$. La réciproque vient de ce que $\{T \leq n\} = \bigcup_{k\leq n} \{T = k\} \in \mathcal{F}_n$ lorsque $\{T = k\} \in \mathcal{F}_k \subset \mathcal{F}_n$ pour $0 \leq k \leq n$.

2) En effet pour $n \in \mathbb{N}$, on écrit
$$\begin{aligned}
\big\{T \wedge S \leq n\big\} &= \big\{T \leq n\big\} \cup \big\{S \leq n\big\} \in \mathcal{F}_n;\\
\big\{T \vee S \leq n\big\} &= \big\{T \leq n\big\} \cap \big\{S \leq n\big\} \in \mathcal{F}_n;\\
\big\{T + S = n\big\} &= \bigcup_{k=0}^{n} \{T = k\} \cap \{S = n-k\} \in \mathcal{F}_n
\end{aligned}$$

car $\{T \leq n\} \in \mathcal{F}_n$, $\{S \leq n\} \in \mathcal{F}_n$, $\{T = k\} \in \mathcal{F}_k \subset \mathcal{F}_n$ et $\{S = n-k\} \in \mathcal{F}_{n-k} \subset \mathcal{F}_n$ pour $0 \leq k \leq n$ et on utilise la caractérisation 1).

2) Cela découle de 3) avec le temps d'arrêt $S = k$ (Exemple 3.9). Ou alors pour $n \geq k$, on $\{T \wedge k \leq n\} = \Omega \in \mathcal{F}_n$ et pour $n < k$, on a

$$\begin{aligned}
\{T \wedge k \leq n\} &= \big(\{T \wedge k \leq n\} \cap \{T \leq k\}\big) \cup \big(\{T \wedge k \leq n\} \cap \{T > k\}\big) \\
&= \big(\{T \leq n\} \cap \{T \leq k\}\big) \cup \big(\{k \leq n\} \cap \{T > k\}\big) \\
&= \{T \leq k \wedge n\} \cup \big(\{T \leq k\}^c \cap \emptyset\big) = \{T \leq k \wedge n\} \in \mathcal{F}_{n \wedge k} \subset \mathcal{F}_n.
\end{aligned}$$

4) Pour tout $n \in \mathbb{N}$, on a dans les cas croissant et décroissant respectivement

$$\begin{aligned}
\{T \leq n\} &= \Big\{\lim_{p \to +\infty} T_p \leq n\Big\} = \bigcap_{p \geq 1} \underbrace{\{T_p \leq n\}}_{\in \mathcal{F}_n} \in \mathcal{F}_n \\
\{T \leq n\} &= \Big\{\lim_{p \to +\infty} T_p \leq n\Big\} = \bigcup_{p \geq 1} \underbrace{\{T_p \leq n\}}_{\in \mathcal{F}_n} \in \mathcal{F}_n
\end{aligned}$$

puisque $\{T_p \leq n\} \in \mathcal{F}_p$ (tribu) en utilisant dans la deuxième partie que les T_p sont à valeurs entières.

5) découle des propriétés précédentes en écrivant

$$\begin{aligned}
\inf_{p \geq 1} T_p &= \lim_{n \to +\infty} \min_{1 \leq p \leq n} T_p, & \liminf_{p \geq 1} T_p &= \sup_{n \geq 1} \inf_{k \geq n} T_k, \\
\sup_{p \geq 1} T_p &= \lim_{n \to +\infty} \max_{1 \leq p \leq n} T_p, & \limsup_{p \geq 1} T_p &= \inf_{n \geq 1} \sup_{k \geq n} T_k,
\end{aligned}$$

ou directement à partir de

$$\begin{aligned}
\Big\{\inf_{p \geq 1} T_p \leq n\Big\} &= \bigcup_{p \geq 1} \{T_p \leq n\}, & \Big\{\sup_{p \geq 1} T_p \leq n\Big\} &= \bigcap_{p \geq 1} \{T_p \leq n\}, \\
\Big\{\liminf_{p \to +\infty} T_p \leq n\Big\} &= \bigcup_{m=0}^{+\infty} \bigcap_{p \geq m} \{T_p \leq n\}, & \Big\{\limsup_{p \to +\infty} T_p \leq n\Big\} &= \bigcap_{m=0}^{+\infty} \bigcup_{p \geq m} \{T_p \leq n\}.
\end{aligned}$$

□

Définition 3.12 (Tribu d'un temps d'arrêt) *À un temps d'arrêt T, on associe la tribu*

$$\mathcal{F}_T = \big\{A \in \mathcal{F} : \forall n \in \mathbb{N}, A \cap \{T \leq n\} \in \mathcal{F}_n\big\}. \tag{3.2}$$

D'abord, on s'assure que la Définition 3.12 a bien un sens :

Proposition 3.13 *Lorsque T est un temps d'arrêt, $\mathcal{F}_T$ en (3.2) est bien une tribu.*

Démonstration : On a bien $\Omega \in \mathcal{F}_T$ puisque $\Omega \cap \{T \leq n\} = \{T \leq n\} \in \mathcal{F}_n$ pour tout $n \geq 0$. Puis si A_i, $i \in I \subset \mathbb{N}$, sont dans $\mathcal{F}_T$ alors

$$\Big(\bigcup_{i\in I} A_i\Big) \cap \{T \leq n\} = \bigcup_{i \in I} \underbrace{\big(A_i \cap \{T \leq n\}\big)}_{\in \mathcal{F}_n} \in \mathcal{F}_n$$

puisque $A_i \cap \{T \leq n\} \in \mathcal{F}_n$ (car $A_i \in \mathcal{F}_T$).

Enfin, si $A \in \mathcal{F}_T$ alors pour tout $n \geq 0$, on a

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

puisque $A \cap \{T \cap n\} \in \mathcal{F}_n$ et $\{T \leq n\} \in \mathcal{F}_n$. Cela assure $A^c \in \mathcal{F}_T$. Finalement, $\mathcal{F}_T$ est bien une tribu. □

Remarque 3.14 — On a $\sigma(T) \subset \mathcal{F}_T$ (voir ci-dessous, Prop. 3.15) donc T est en particulier $\mathcal{F}_T$-mesurable mais attention en général l'égalité est fausse.
- Il faut comprendre $\mathcal{F}_T$ de la manière suivante : quand T a lieu avant n, on sait à la date n si A est réalisé ou pas.
- Si $(\mathcal{F}_n)_{n\geq 0}$ est la filtration naturelle de $(X_n)_{n\geq 0}$ et T est un $(\mathcal{F}_n)$-temps d'arrêt alors l'information contenue dans $\mathcal{F}_T$ comprend, d'une part la valeur de T et d'autre part aussi, les valeurs de $X_1, \dots, X_T$.
- De façon générale, en suivant la Remarque 3.6, on peut interpréter la tribu $\mathcal{F}_T$ comme l'information disponible à la date aléatoire T.

Proposition 3.15 (Propriétés des tribus $\mathcal{F}_T$)

(1) Pour un temps d'arrêt constant $T = n_0$, alors on a bien $\mathcal{F}_T = \mathcal{F}_{n_0}$.

(2) À l'instar de 1) dans la Prop. 3.11, $A \in \mathcal{F}_T$ si et seulement si $A \cap \{T = n\} \in \mathcal{F}_n$.

(3) Si $T \leq S$ sont deux temps d'arrêt alors $\mathcal{F}_T \subset \mathcal{F}_S$.

(4) Un temps d'arrêt T est $\mathcal{F}_T$-mesurable.

(5) Pour T, S des temps d'arrêt, on a $\mathcal{F}_{T\wedge S} = \mathcal{F}_T \cap \mathcal{F}_S$. De plus $\{T \leq S\}, \{S \leq T\}, \{T = S\} \in \mathcal{F}_{T\wedge S}$.

(6) Pour $A \in \mathcal{F}$ et T un temps d'arrêt, posons $T_A(\omega) = T(\omega)$ si $\omega \in A$, $T_A(\omega) = +\infty$ sinon. Alors $A \in \mathcal{F}_T$ si et seulement si T_A est un temps d'arrêt.

Démonstration : 1) On a $A \in \mathcal{F}_T$ si et seulement si pour tout $n \geq 0$, $A \cap \{T \leq n\} \in \mathcal{F}_n$. Lorsque $T = n_0$,
- pour $n < n_0$, $\{T \leq n\} = \emptyset$ et $A \cap \{T \leq n\} = \emptyset \in \mathcal{F}_n$ (toujours vrai).
- pour $n \geq n_0$, $\{T \leq n\} = \Omega$ et $A \cap \{T \leq n\} = A$ est dans $\mathcal{F}_n$ pour tout $n \geq n_0$ si et seulement si $A \in \mathcal{F}_{n_0}$.

Finalement $A \in \mathcal{F}_T$ si et seulement si $A \in \mathcal{F}_{n_0}$, soit $\mathcal{F}_T = \mathcal{F}_{n_0}$.

2) découle des égalités :

$$A \cap \{T = n\} = \big(A \cap A \cap \{T \leq n\}\big) \setminus \big(A \cap \{T \leq n-1\}\big)$$
$$A \cap \{T \leq n\} = \bigcup_{k=0}^{n} A \cap \{T = k\}.$$

3) Soit $A \in \mathcal{F}_T$ et $n \geq 0$. Comme $T \leq S$, on a

$$A \cap \{S \leq n\} = \overbrace{\big(A \cap \{T \leq n\}\big)}^{\in \mathcal{F}_n} \cap \overbrace{\{S \leq n\}}^{\in \mathcal{F}_n} \in \mathcal{F}_n.$$

On a donc bien $A \in \mathcal{F}_S$.

4) Comme les ensembles $[0, t]$ engendrent la tribu $\mathcal{B}(\mathbb{R}_+)$, on montre que T est $\mathcal{F}_T$-mesurable en prouvant que $\{T \leq p\} \in \mathcal{F}_T$ pour tout $p \geq 0$. Pour cela, soit $n \geq 0$, on a

$$\{T \leq p\} \cap \{T \leq n\} = \big\{T \leq n \wedge p\big\} \in \mathcal{F}_{n \wedge p} \subset \mathcal{F}_n,$$

ce qui justifie que $\{T \leq p\} \in \mathcal{F}_T$.

5) D'après le 3) avec $T \wedge S \leq T$ et $T \wedge S \leq S$, on a $\mathcal{F}_{T \wedge S} \subset \mathcal{F}_T \cap \mathcal{F}_S$. Puis si $A \in \mathcal{F}_T \cap \mathcal{F}_S$ alors pour tout $n \geq 0$, on a $A \cap \{T \leq n\} \in \mathcal{F}_n$ et $A \cap \{S \leq n\} \in \mathcal{F}_n$ donc

$$A \cap \{T \wedge S \leq n\} = \big(A \cap \{T \leq n\}\big) \cup \big(A \cap \{S \leq n\}\big) \in \mathcal{F}_n$$

ce qui assure $A \in \mathcal{F}_{T \wedge S}$.

Compte tenu de la première partie, pour montrer $\{T \leq S\} \in \mathcal{F}_{T \wedge S} = \mathcal{F}_T \cap \mathcal{F}_S$, il suffit de montrer que $\{T \leq S\} \in \mathcal{F}_T$ et $\{T \leq S\} \in \mathcal{F}_S$.

On a

$$\{T \leq S\} \cap \{T \leq n\} = \{T \wedge n \leq S \wedge n\} \cap \{T \leq n\}$$

avec $\{T \wedge n \leq S \wedge n\} \in \mathcal{F}_n$ car $T \wedge n$ et $S \wedge n$ sont $\mathcal{F}_n$-mesurables et $\{T \leq n\} \in \mathcal{F}_n$. On montre que $\{T \leq S\} \cap \{S \leq n\} \in \mathcal{F}_n$ en montrant que $\{T \leq S\} \cap \{S = n\} \in \mathcal{F}_n$ pour chaque $n \geq 0$ (écrire $\{S \leq n\} = \bigcup_{k=0}^n \{S = k\}$) : comme $\{T \leq n\}, \{S = n\} \in \mathcal{F}_n$, on a bien

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

On a donc $\{T \leq S\} \in \mathcal{F}_T \cap \mathcal{F}_S = \mathcal{F}_{T \wedge S}$.

De la même façon, on a $\{S \leq T\} \in \mathcal{F}_{T \wedge S}$ et donc il vient

$$\big\{T = S\big\} = \{T \leq S\} \cap \{S \leq T\} \in \mathcal{F}_{T \wedge S}.$$

6) On a $A \in \mathcal{F}_T$ si et seulement si pour tout $n \geq 0$: $A \cap \{T \leq n\} \in \mathcal{F}_n$. Mais $A \cap \{T \leq n\} = \{T_A \leq n\}$ ce qui permet de conclure. □

Proposition 3.16 *Soit $(X_n)_{n\geq 1}$ une suite $(\mathcal{F}_n)$-adaptée et T un temps d'arrêt. Alors la variable aléatoire*

$$\mathbf{1}_{\{T<+\infty\}}\, X_T = \begin{cases} X_n & \text{si } T = n \\ 0 & \text{si } T = +\infty \end{cases}$$

est $\mathcal{F}_T$-mesurable. Lorsque $T < +\infty$ ps, il n'y a pas d'ambiguïté de notation et on écrit simplement X_T.

Démonstration : Pour $B \in \mathcal{B}(\mathbb{R})$ et $n \in \mathbb{N}$, on a

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

c'est à dire $\big\{\mathbf{1}_{\{T<+\infty\}} X_T \in B\big\} \in \mathcal{F}_T$ pour tout $B \in \mathcal{B}(\mathbb{R})$, ie. $\mathbf{1}_{\{T<+\infty\}} X_T$ est $\mathcal{F}_T$-mesurable. □

3.3 Martingales, sous-martingales et sur-martingales

Définitions

Définition 3.17 (Martingale) *Une suite de variables aléatoires $(X_n)_{n\geq 0}$ est une martingale par rapport à une filtration $(\mathcal{F}_n)_{n\geq 0}$ si*

(i) $\mathbb{E}\big[|X_n|\big] < +\infty$ *pour tout $n \geq 0$;*

(ii) la suite $(X_n)_{n\geq 0}$ est $(\mathcal{F}_n)$-adaptée ;

(iii) pour tout $n \geq 0$:

$$\mathbb{E}\big[X_{n+1}\,|\mathcal{F}_n\big] = X_n. \tag{3.3}$$

Définition 3.18 (Sur- et sous-martingales) *On parle de sous-martingales ou de sur-martingales quand (iii) dans la Définition 3.17 est remplacé respectivement par*

sous-martingale : $\mathbb{E}\big[X_{n+1}\,|\mathcal{F}_n\big] \geq X_n$ *pour tout $n \geq 0$;*

sur-martingale : $\mathbb{E}\big[X_{n+1}\,|\mathcal{F}_n\big] \leq X_n$ *pour tout $n \geq 0$.*

Dans la suite, lorsque il n'y a pas d'ambiguïté, on pourra omettre d'indiquer la filtration $(\mathcal{F}_n)_{n\geq 0}$ et on parlera simplement de martingales plutôt que de $(\mathcal{F}_n)$-martingale. Idem pour les sous ou sur-martingales.

Exemples de martingales, sous-martingales, sur-martingales

Exemple 3.19 (Martingale fermée) Soit $X \in L^1(\mathcal{F})$. On définit une martingale par $X_n = \mathbb{E}\big[X\,|\mathcal{F}_n\big]$, $n \geq 0$. La propriété de martingale suit facilement du Th. 2.12 (conditionnement par cascade). Une telle martingale sera dite fermée, cf. Définition 4.24. On parle aussi de martingale de **Doob**.

Exemple 3.20 (Marche aléatoire) Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires intégrables indépendantes centrées alors

$$S_n = X_1 + \dots + X_n,\ n \geq 1, \quad \text{et} \quad S_0 = 0,$$

est une (sur/sous)-martingale par rapport à la filtration naturelle $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$, $n \geq 1$ (avec $\mathcal{F}_0 = \{\emptyset, \Omega\}$), selon le signe ou la nullité de $\mathbb{E}[X_1]$.

En effet, S_n est clairement $\mathcal{F}_n$-mesurable et intégrable car les X_i, $1 \leq i \leq n$, le sont. Comme $X_{n+1} \perp\!\!\!\perp \mathcal{F}_n = \sigma(X_1, \dots, X_n)$, il vient par le Th. 2.22 :

$$\mathbb{E}\big[S_{n+1}\,|\mathcal{F}_n\big] \;=\; \mathbb{E}\big[S_n + X_{n+1}\,|\mathcal{F}_n\big] = S_n + \mathbb{E}\big[X_{n+1}\,|\mathcal{F}_n\big] = S_n + \mathbb{E}[X_{n+1}] = S_n,$$

dans le cas centré (on adapte facilement aux cas $\mathbb{E}[X_1] > 0$ et $\mathbb{E}[X_1] < 0$).

Dans cet exemple, on a

$$X_n = S_n - S_{n-1} = S_n - \mathbb{E}\big[S_n\,|\mathcal{F}_{n-1}\big].$$

Ainsi la suite $(X_n)_{n\geq 1}$ n'est pas un'e martingale mais une **différence de martingale**. De façon générale, une suite de variables aléatoires indépendantes est une différence de martingale. Le comportement asymptotique de martingales généralise ainsi celui de sommes de variables aléatoires indépendantes, cf. Section 4.2.

Exemple 3.21 (Modèle auto-régressif) Soit $(\varepsilon_n)_{n\geq 1}$ une suite de variables aléatoires *iid* intégrables centrées et $a \in \mathbb{R}^*$. On pose

$$X_{n+1} = aX_n + \varepsilon_{n+1},\ n \geq 0, \quad \text{et} \quad X_0 = x. \tag{3.4}$$

Alors $Y_n = X_n/a^n$, $n \geq 0$, forme une martingale par rapport à la filtration naturelle $\mathcal{F}_n = \sigma(\varepsilon_1, \dots, \varepsilon_n)$, $n \geq 1$, $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

En effet, par récurrence, les Y_n sont intégrables et $\mathcal{F}_n$-mesurables, $n \geq 1$. Puis

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

car X_n est $\mathcal{F}_n$-mesurable et $\varepsilon_{n+1} \perp\!\!\!\perp \mathcal{F}_n$ (Th. 2.22).

Exemple 3.22 (Galton-Watson) Soit $(X_{i,j})_{i,j\geq 1}$ une famille de variables aléatoires entières *iid* de loi μ (sur $\mathbb{N}$) admettant pour moyenne m. On pose $Z_0 = 1$ et pour $n \geq 1$

$$Z_{n+1} = \sum_{j=1}^{Z_n} X_{n+1,j}. \tag{3.5}$$

Alors $(Z_n/m^n)_{n\geq 0}$ est une martingale par rapport à la filtration donnée par $\mathcal{F}_n = \sigma(X_{i,j} : i \leq n, j \geq 1)$: d'abord, on observe par récurrence que Z_n est $\mathcal{F}_n$-mesurable. En effet, si Z_n l'est alors pour tout $A \in \mathbb{N}$, on a

$$\{Z_{n+1} \in A\} = \bigcup_{p=0}^{+\infty} \{Z_{n+1} \in A, Z_n = p\} = \bigcup_{p=0}^{+\infty} \left(\left\{ \sum_{j=1}^{p} X_{n+1,j} \in A \right\} \cap \{Z_n = p\} \right) \in \mathcal{F}_{n+1}$$

puisque $\sum_{j=1}^{p} X_{n+1,j}$ est $\mathcal{F}_{n+1}$-mesurable et Z_n aussi par hypothèse de récurrence. Puis la propriété de martingale est bien satisfaite :

$$\mathbb{E}[Z_{n+1}|\mathcal{F}_n] = \mathbb{E}\left[\sum_{j=1}^{Z_n} X_{n+1,j} \,|\mathcal{F}_n \right] = \sum_{j=1}^{Z_n} \mathbb{E}\big[X_{n+1,j} \,|\mathcal{F}_n\big] = Z_n \mathbb{E}\big[X_{n+1,j}\big] = Z_n m$$

car Z_n est $\mathcal{F}_n$-mesurable et $X_{n+1,j} \perp\!\!\!\perp \mathcal{F}_n$ (Th. 2.22). Il vient

$$\mathbb{E}\left[\frac{Z_{n+1}}{m^{n+1}} \,|\mathcal{F}_n \right] = \frac{Z_n}{m^n}.$$

À noter que les espérances conditionnelles sont bien définies puisque les variables aléatoires sont positives. *A posteriori*, on observe par récurrence que Z_n est bien intégrable puisque

$$\mathbb{E}[|Z_n|] = \mathbb{E}[Z_n] = \mathbb{E}\big[\mathbb{E}[Z_{n+1}|\mathcal{F}_n]\big] = m\,\mathbb{E}[Z_n] < +\infty.$$

Cette martingale modélise l'évolution d'une population avec loi de reproduction μ. Dans cette interprétation, $\mathbb{N}$ représente les numéros des générations successives, $X_{n+1,j}$ est le nombre d'enfants de l'individu j de la génération n pour former la génération $n+1$ et Z_n désigne la taille de la population à la génération n.

Exemple 3.23 (Wright-Fisher) Soit $E = \{0, 1 \dots, N\}$ et $X_0 \in \{1, 2 \dots, N-1\}$. Par récurrence, on définit les lois conditionnelles

$$\mathcal{L}\big(X_{n+1} \,|\mathcal{F}_n\big) = \mathcal{B}\big(N, X_n/N\big). \tag{3.6}$$

Cela définit une martingale par rapport à la filtration naturelle puisque

$$\mathbb{E}\big[X_{n+1} \,|\mathcal{F}_n\big] = N\left(\frac{X_n}{N}\right) = X_n.$$

Exemple 3.24 (Cascade aléatoire) Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires intégrables indépendantes telle que $\mathbb{E}[X_n] = 1$ pour tout $n \geq 1$. On note $(\mathcal{F}_n)_{n\geq 1}$ la filtration naturelle de $(X_n)_{n\geq 1}$ et on pose $Y_0 = 1$, et

$$Y_n = \prod_{i=1}^{n} X_i.$$

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

En effet, il est clair que Y_n est $\mathcal{F}_n$-mesurable et intégrable puisque par indépendance des X_i :

$$\mathbb{E}\big[|Y_n|\big] = \mathbb{E}\Big[\prod_{i=1}^n |X_i|\Big] = \prod_{i=1}^n \mathbb{E}\big[|X_i|\big] < +\infty.$$

Puis comme Y_n est $\mathcal{F}_n$-mesurable et $X_{n+1} \perp\!\!\!\perp \mathcal{F}_n$, avec le Th. 2.22 on a :

$$\mathbb{E}\big[Y_{n+1}\,|\mathcal{F}_n\big] = \mathbb{E}\big[Y_n X_{n+1}\,|\mathcal{F}_n\big] = Y_n \mathbb{E}\big[X_{n+1}\,|\mathcal{F}_n\big] = Y_n \mathbb{E}[X_{n+1}] = Y_n.$$

Lorsque les variables aléatoires X_i sont toutes positives, on a des résultats analogues pour des sous-martingales quand $\mathbb{E}[X_i] \geq 1$ pour tout $i \geq 1$ ou des sur-martingales quand $\mathbb{E}[X_i] \leq 1$ pour tout $i \geq 1$.

3.4 Propriétés des martingales

De façon générale, les énoncés pour les martingales s'adaptent pour des sous-martingales ou des sur-martingales.

Proposition 3.25 *Si $(X_n)_{n\geq 0}$ est une $(\mathcal{F}_n)$-martingale alors $(X_n)_{n\geq 0}$ est une $(\mathcal{G}_n)$-martingale pour $\mathcal{G}_n = \sigma(X_1, \dots, X_n)$. Le même énoncé reste vrai pour des sous-martingales ou sur-martingales.*

Démonstration : Puisque $(X_n)_{n\geq 0}$ est une $(\mathcal{F}_n)$-martingale, chaque X_n est intégrable. Puis par définition de $(\mathcal{G}_n)_{n\geq 0}$, $(X_n)_{n\geq 0}$ est $(\mathcal{G}_n)$-adaptée. Comme $X_1, \dots, X_n$ sont $\mathcal{F}_n$-mesurables, il est immédiat que $\mathcal{G}_n \subset \mathcal{F}_n$. Par le Th. 2.12 (conditionnement en cascade), on a alors

$$\mathbb{E}\big[X_{n+1}\,|\mathcal{G}_n\big] = \mathbb{E}\big[\overbrace{\mathbb{E}[X_{n+1}\,|\mathcal{F}_n]}^{=X_n}\,|\mathcal{G}_n\big] = \mathbb{E}\big[X_n|\mathcal{G}_n\big] = X_n$$

puisque $(X_n)_{n\geq 0}$ est une $(\mathcal{F}_n)$-martingale et X_n est $\mathcal{G}_n$-mesurable. □

Proposition 3.26 *Dans la définition d'une martingale (Définition 3.17), (3.3) est équivalente à : pour tout $n > m$, on a $\mathbb{E}\big[X_n\,|\mathcal{F}_m\big] = X_m$. Résultats analogues pour des sous ou sur-martingales.*

Démonstration : Par définition (Définition 3.18), le résultat est vrai pour $n = m + 1$. Si on suppose qu'il est vrai pour $n = m+k-1$, $k \geq 2$, alors par le Th. 2.12 (conditionnement en cascade) et la monotonie de l'espérance conditionnelle (Prop. 2.15), on a

$$\mathbb{E}\big[X_{n+k}\,|\mathcal{F}_m\big] = \mathbb{E}\big[\mathbb{E}[X_{n+k}\,|\mathcal{F}_{m+k-1}]\,|\mathcal{F}_m\big] = \mathbb{E}\big[X_{n+k-1}\,|\mathcal{F}_m\big]$$

et le résultat découle maintenant de l'hypothèse de récurrence. □

Proposition 3.27 *Soit $(X_n)_{n\geq 0}$ une martingale (resp. sous-martingale, sur-martingale). Alors $\mathbb{E}[X_{n+1}] = \mathbb{E}[X_n](= \mathbb{E}[X_0])$ (resp. $\mathbb{E}[X_{n+1}] \geq \mathbb{E}[X_n]$, $\mathbb{E}[X_{n+1}] \leq \mathbb{E}[X_n]$).*

Remarque 3.28 En quelque sorte, il faut retenir que les sous-martingales sont des analogues aléatoires des suites numériques croissantes ($\mathbb{E}[X_n] \leq \mathbb{E}[X_{n+1}]$ pour tout $n \geq 0$).

Démonstration : En prenant l'espérance dans la propriété de martingale (3.3), on a

$$\mathbb{E}[X_{n+1}] = \mathbb{E}\big[\mathbb{E}[X_{n+1} \,|\mathcal{F}_n]\big] = \mathbb{E}[X_n].$$

On adapte facilement l'argument pour les sous-martingales et pour les sur-martingales. □

Proposition 3.29 (Martingale et Jensen)

(1) Si $(X_n)_{n\geq 0}$ est une $(\mathcal{F}_n)$-martingale et $\varphi : \mathbb{R} \to \mathbb{R}$ est une fonction convexe telle que $\mathbb{E}\big[|\varphi(X_n)|\big] < +\infty$ pour tout $n \geq 0$ alors $Y_n = \varphi(X_n)$, $n \geq 0$, est une $(\mathcal{F}_n)$-sous-martingale.

(2) Si $(X_n)_{n\geq 0}$ est une $(\mathcal{F}_n)$-sous-martingale et $\varphi : \mathbb{R} \to \mathbb{R}$ est une fonction convexe et croissante telle que $\mathbb{E}\big[|\varphi(X_n)|\big] < +\infty$ pour tout $n \geq 1$ alors $Y_n = \varphi(X_n)$, $n \geq 0$, est une $(\mathcal{F}_n)$-sous-martingale.

Démonstration : (1) D'abord Y_n est clairement $\mathcal{F}_n$-mesurable puisque X_n l'est et φ convexe est mesurable. Puis, par l'inégalité de Jensen conditionnelle (Prop. 2.19), on a :

$$\mathbb{E}\big[Y_{n+1} \,|\mathcal{F}_n\big] = \mathbb{E}\big[\varphi(X_{n+1}) \,|\mathcal{F}_n\big] \geq \varphi\Big(\mathbb{E}\big[X_{n+1} \,|\mathcal{F}_n\big]\Big) = \varphi(X_n) = Y_n.$$

(2) La preuve du 1) s'adapte facilement en utilisant en plus la croissance de φ :

$$\mathbb{E}\big[Y_{n+1} \,|\mathcal{F}_n\big] = \mathbb{E}\big[\varphi(X_{n+1}) \,|\mathcal{F}_n\big] \geq \varphi\Big(\mathbb{E}[X_{n+1} \,|\mathcal{F}_n]\Big) \geq \varphi(X_n) = Y_n.$$

□

Corollaire 3.30 *Soit $(X_n)_{n\geq 0}$ une $(\mathcal{F}_n)$-martingale avec $\mathbb{E}\big[|X_n|^p\big] < +\infty$ pour tout $n \geq 0$. Alors $\big(|X_n|^p\big)_{n\geq 1}$ est une $(\mathcal{F}_n)$-sous-martingale.*

Démonstration : Application directe de la Prop. 3.29 avec la fonction convexe $\varphi(x) = |x|^p$. □

Corollaire 3.31 *Soit $a \in \mathbb{R}$.*

(1) Soit $(X_n)_{n\geq 1}$ une $(\mathcal{F}_n)$-sous-martingale. Alors $\big((X_n - a)^+\big)_{n\geq 0}$ est une $(\mathcal{F}_n)$-sous-martingale.

(2) Soit $(X_n)_{n\geq 1}$ une $(\mathcal{F}_n)$-sur-martingale. Alors $\big(\min(X_n, a)\big)_{n\geq 1}$ est une $(\mathcal{F}_n)$-sur-martingale.

Démonstration : 1) On applique la Prop. 3.29 avec la fonction convexe croissante $\varphi(x) = (x-a)^+$.

2) On applique 1) à la sous-martingale $(-X_n)_{n\geq 1}$ et la fonction convexe croissante $\varphi(x) = \max(x, -a)$ pour avoir que $\varphi(-X_n)$ est une sous-martingale. Il s'ensuit que $-\varphi(-X_n) = -\max(-X_n, -a) = \min(X_n, a)$, $n \geq 0$, forme une martingale. □

On rappelle que la notion de prévisibilité est donnée en Définition 3.5.

Proposition 3.32 *Soit $(X_n)_{n\geq 0}$ une $(\mathcal{F}_n)$-sous-martingale. Si $(H_n)_{n\geq 1}$ est une suite prévisible positive avec chaque H_n bornée alors $(H \cdot X)$ définie par $(H \cdot X)_0 = 0$ et*

$$(H \cdot X)_n = \sum_{k=1}^{n} H_k(X_k - X_{k-1}), \quad n \geq 1,$$

forme une $(\mathcal{F}_n)$-sous-martingale. La même affirmation est vraie pour une sur-martingale ou pour une martingale sans la restriction de positivité $H_n \geq 0$ dans le cas d'une martingale.

Démonstration : On observe sans difficulté que $(H \cdot X)_n$ est $\mathcal{F}_n$-mesurable pour tout $n \geq 1$: comme

$$(H \cdot X)_n = (H \cdot X)_{n-1} + H_n(X_n - X_{n-1}), \tag{3.7}$$

un argument par récurrence ramène à voir que $H_n(X_n - X_{n-1})$ est $\mathcal{F}_n$-mesurable ce qui est bien le cas puisque H_n, X_n, X_{n-1} le sont. Puis $(H \cdot X) \in L^1$ car chaque $H_k(X_k - X_{k-1}) \in L^1$ puisque $X \in L^1$ et H est bornée.

Ensuite, en utilisant (3.7) et la prévisibilité de H_n, on a :

$$\begin{aligned} \mathbb{E}\big[(H \cdot X)_{n+1} \,|\mathcal{F}_n\big] &= (H \cdot X)_n + \mathbb{E}\big[H_{n+1}(X_{n+1} - X_n) \,|\mathcal{F}_n\big] \\ &= (H \cdot X)_n + H_{n+1}\,\mathbb{E}\big[X_{n+1} - X_n \,|\mathcal{F}_n\big]. \end{aligned}$$

On conclut en observant que $\mathbb{E}\big[X_{n+1} - X_n \,|\mathcal{F}_n\big] \geq 0$ si $(X_n)_{n\geq 1}$ est une sous-martingale avec $H_n \geq 0$. On conclut de même avec $\mathbb{E}\big[X_{n+1} - X_n \,|\mathcal{F}_n\big] \leq 0$ si $(X_n)_{n\geq 1}$ est une sur-martingale et avec $\mathbb{E}\big[X_{n+1} - X_n \,|\mathcal{F}_n\big] = 0$ si c'est une martingale (sans condition sur le signe de H). □

Remarque 3.33 (Interprétation en termes financiers) On considère un actif risqué prenant la valeur X_n à la date n. Une suite prévisible $(H_n)_{n\geq 1}$ s'interprète dans ce contexte comme une stratégie d'investissement : il s'agit de la quantité H_n d'actif risqué acheté à la date n. La valeur du portefeuille à la date n est alors

$$(H \cdot X)_n = \sum_{i=1}^{n} H_i(X_i - X_{i-1}).$$

En effet $(H \cdot X)_n$ est la valeur $(H \cdot X)_{n-1}$ à la date $n-1$ plus la valeur du nouvel actif $H_n X_n$ moins le coût de l'achat $H_n X_{n-1}$.

On interprète également $(H \cdot X)_n$ comme une intégrale stochastique (discrète) de $(H_n)_{n\geq 1}$ contre la suite $(X_n)_{n\geq 1}$.

La prévisibilité de H s'interprète alors de la façon suivante : chaque jour, les ordres d'achat sont passés le matin et les prix re-actualisés au cours de la journée. Ainsi, le jour n, la quantité H_n d'actif risqué est achetée à la valeur X_{n-1} du $(n-1)$-ième jour. La décision d'acheter est donc prise avec l'information dont on dispose à la date $n-1$, ie. les X_i, $i \leq n-1$ (il n'y a pas de délit d'initié). Cela justifie que la variable aléatoire H_n doit être $\mathcal{F}_{n-1}$-mesurable.

3.5 Martingale arrêtée

Étant donné un $(\mathcal{F}_n)$-temps d'arrêt T et une suite $X = (X_n)_{n\geq 0}$, la suite $X^T = (X_{T\wedge n})_{n\geq 0}$ s'appelle la suite arrêtée.

Proposition 3.34 *Soit T un $(\mathcal{F}_n)$-temps d'arrêt.*

(1) Soit $(X_n)_{n\geq 0}$ une suite $(\mathcal{F}_n)$-adaptée. Alors $X^T = (X_{T\wedge n})_{n\geq 0}$ est encore une suite $(\mathcal{F}_n)$-adaptée.

(2) Soit $(H_n)_{n\geq 0}$ une suite $(\mathcal{F}_n)$-prévisible. Alors $H^T = (H_{T\wedge n})_{n\geq 0}$ est encore une suite prévisible.

Démonstration : 1) Pour $B \in \mathcal{B}(\mathbb{R})$ on a

$$\begin{aligned}\big\{X_n^T \in B\big\} &= \big\{X_{T\wedge n} \in B\big\}\\ &= \left(\bigcup_{p=0}^n \big\{X_p \in B, T = p\big\}\right) \cup \big\{X_n \in B, T \geq n+1\big\} \in \mathcal{F}_n\end{aligned}$$

puisque, pour $0 \leq p \leq n$, $\{X_p \in B\} \in \mathcal{F}_p \subset \mathcal{F}_n$, $\{T = p\} \in \mathcal{F}_p \subset \mathcal{F}_n$, $\{X_n \in B\} \in \mathcal{F}_n$, $\{T \geq n+1\} = \{T \leq n\}^c \in \mathcal{F}_n$.

2) Pour $B \in \mathcal{B}(\mathbb{R})$ on a

$$\begin{aligned}\big\{H_{n+1}^T \in B\big\} &= \big\{H_{T\wedge(n+1)} \in B\big\}\\ &= \left(\bigcup_{p=0}^n \big\{H_p \in B, T = p\big\}\right) \cup \big\{H_{n+1} \in B, T \geq n+1\big\} \in \mathcal{F}_n\end{aligned}$$

puisque, pour $0 \leq p \leq n$, $\{H_p \in B\} \in \mathcal{F}_{p-1} \subset \mathcal{F}_n$, $\{T = p\} \in \mathcal{F}_p \subset \mathcal{F}_n$, $\{H_{n+1} \in B\} \in \mathcal{F}_n$, $\{T \geq n+1\} = \{T \leq n\}^c \in \mathcal{F}_n$. □

Définition 3.35 (Martingale arrêtée) *Si $(X_n)_{n\geq 0}$ est une $(\mathcal{F}_n)$-martingale et T est un $(\mathcal{F}_n)$-temps d'arrêt. On appelle martingale arrêtée la suite $\big(X_n^T\big)_{n\geq 0}$ avec $X_n^T = X_{T\wedge n}$. On introduit des notions analogues pour les sous-martingales ou sur-martingales.*

En fait, on montre qu'une (sur/sous)-martingale arrêtée est une (sur/sous)-martingale.

Proposition 3.36 (Martingale arrêtée) *Si T est un $(\mathcal{F}_n)$-temps d'arrêt et $(X_n)_{n\geq 0}$ une $(\mathcal{F}_n)$-martingale, sur ou sous-martingale. Alors $X^T = \big(X_{T\wedge n}\big)_{n\geq 0}$ est une $(\mathcal{F}_n)$-martingale, sur ou sous-martingale. La (sous/sur)-martingale arrêtée est donc une (sous/sur)-martingale !*

Démonstration : On a vu en (3.1) que la suite $(H_n)_{n\geq 0}$ donnée par $H_n = \mathbf{1}_{\{T\geq n\}}$ est $(\mathcal{F}_n)$-prévisible. Dès lors, d'après la Prop. 3.32, on a :

$$(H\cdot X)_n = \sum_{k=1}^{n} \mathbf{1}_{\{T\geq k\}}(X_k - X_{k-1}) = \sum_{k=1}^{n\wedge T}(X_k - X_{k-1}) = X_{T\wedge n} - X_0, \quad n\geq 0,$$

est une $(\mathcal{F}_n)$-martingale, sur ou sous-martingale selon ce qu'est X, ce qui établit le résultat puisque la somme de martingale, sur ou sous-martingales est de même nature. □

Le théorème d'arrêt consiste à généraliser la propriété de martingale (ou de sur/sous martingale) à des dates $m \leq n$ données par des temps d'arrêt $S \leq T$, cf. Prop. 3.26. On commence par une version faible de cette propriété sur la constance (ou croissance/décroissance) des suites d'espérance, cf. Prop. 3.27. D'abord, on donne une première forme du théorème d'arrêt pour des temps d'arrêt **bornés** :

Théorème 3.37 *Soit $(X_n)_{n\geq 0}$ une sous-martingale et T un temps d'arrêt tel que $T \leq k$ ps pour un $k\in\mathbb{N}$ donné (ie. T est borné). Alors $X_T \in L^1$ et*

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_T] \leq \mathbb{E}[X_k]. \tag{3.8}$$

De plus,
- *il y a égalité dans (3.8) si $(X_n)_{n\geq 0}$ est une martingale ;*
- *Pour une sur-martingale, (3.8) est valable avec des bornes inversées.*

Exemple 3.38 (Contre-exemple au Th. 3.37) Soit $(S_n)_{n\geq 0}$ la marche aléatoire simple : $S_n = \sum_{i=1}^n X_i$ avec X_i *iid* de loi de Rademacher $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}$ et $S_0 = 0$. Il s'agit d'une martingale pour la filtration des $\mathcal{F}_n = \sigma(X_1,\dots,X_n)$, $n\geq 1$.

On note $T = \inf\big(n\geq 0 : S_n = -1\big)$. Il s'agit d'un temps d'arrêt par 2) dans Exemple 3.9. Alors $\mathbb{E}[S_0] = 0 > -1 = \mathbb{E}[S_T]$. On note que T n'est pas borné puisque

$$\{T\geq n\} \supset \{X_1 = 1, X_2 = 1, \dots, X_n = 1\}$$

et donc

$$\mathbb{P}(T\geq n) \geq \mathbb{P}(X_1 = 1, X_2 = 1, \dots, X_n = 1) = \frac{1}{2^n}.$$

Ainsi la première inégalité dans le Th. 3.37 n'est pas automatique si T n'est pas bornée.

Démonstration : On considère d'abord $X = (X_n)_{n\geq 0}$ une sous-martingale. Comme $0 \leq T \leq k$, on a $X_T = \sum_{i=0}^k X_i \mathbf{1}_{\{T=i\}}$ et il vient d'abord $X_T \in L^1$ puisque $|X_T| \leq \sum_{i=0}^k |X_i|$.

Par la Prop. 3.36, $(X_{T\wedge n})_{n\geq 0}$ est une sous-martingale. Ainsi comme $0 \leq T \leq k$ ps, en utilisant la croissance des espérances pour la sous-martingale arrêtée X^T (Prop. 3.27), on a :

$$\mathbb{E}[X_0] = \mathbb{E}\big[X_{T\wedge 0}\big] \leq \mathbb{E}\big[X_{T\wedge k}\big] = \mathbb{E}[X_T]$$

ce qui prouve la première inégalité de (3.8). Pour prouver la deuxième inégalité de (3.8), la propriété de sous-martingale donne pour tout $0 \leq i \leq k$: $X_i \leq \mathbb{E}[X_k|\mathcal{F}_i]$ ps et comme $\{T = i\} \in \mathcal{F}_i$:

$$\mathbb{E}\big[X_i \mathbf{1}_{\{T=i\}}\big] \leq \mathbb{E}\big[\mathbb{E}[X_k|\mathcal{F}_i]\mathbf{1}_{\{T=i\}}\big] = \mathbb{E}\big[X_k \mathbf{1}_{\{T=i\}}\big],$$

et donc

$$\mathbb{E}[X_T] = \mathbb{E}\Big[\sum_{i=0}^k X_T \mathbf{1}_{\{T=i\}}\Big] = \sum_{i=0}^k \mathbb{E}\big[X_i \mathbf{1}_{\{T=i\}}\big] \leq \sum_{i=0}^k \mathbb{E}\big[X_k \mathbf{1}_{\{T=i\}}\big] = \mathbb{E}[X_k].$$

Si $(X_n)_{n\geq 0}$ est une sur-martingale on applique le résultat (3.8) à la sous-martingale $(-X_n)_{n\geq 0}$ pour avoir $\mathbb{E}[X_0] \geq \mathbb{E}[X_T] \geq \mathbb{E}[X_k]$.

Si $(X_n)_{n\geq 0}$ est une martingale, on a (3.8) pour $(X_n)_{n\geq 0}$ et pour $(-X_n)_{n\geq 0}$, ce qui donne l'égalité. □

Théorème 3.39 *Soit $(X_n)_{n\geq 0}$ une sous-martingale et T un temps d'arrêt. Sous chacune des conditions suivantes, on a $X_T \in L^1$ et*

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_T]. \tag{3.9}$$

(1) T est borné (ie. il existe $C > 0$ tel que $T \leq C$ ps) ;

(2) la suite X est bornée (il existe $K > 0$ tel que $|X_n| \leq K$ $\forall n$ ps) et T est fini ps ;

(3) $\mathbb{E}[T] < +\infty$ et il existe $K > 0$ tel que $|X_{n+1} - X_n| \leq K$ ps pour tout $n \geq 0$.

De plus, si $(X_n)_{n\geq 0}$ une martingale, sous (1), (2) ou (3), on a l'égalité dans (3.9) :

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

Enfin si $(X_n)_{n\geq 0}$ est une sur-martingale, on a $\mathbb{E}[X_0] \geq \mathbb{E}[X_T]$ sous **(1)**, **(2)**, **(3)** *ou encore sous*

(4) $X_n \geq 0$ et T est fini ps.

Démonstration : On suppose d'abord que $(X_n)_{n\geq 0}$ est une sous-martingale.

1) découle du Th. 3.37 appliqué à la sous-martingale $(X_n)_{n\geq 0}$. On peut noter qu'on n'utilise que la partie facile du Th. 3.37 qui se réduit à : la sous-martingale arrêtee est une sous-martingale (Prop. 3.36) donc d'espérances croissantes.

2) Comme $T \wedge n$ est un temps d'arrêt borné, la partie **1)** s'applique avec $T \wedge n$ et donne

$$\mathbb{E}\big[X_{T\wedge n}\big] \geq \mathbb{E}[X_0]. \tag{3.10}$$

Quand $n \to +\infty$, on a $X_{T\wedge n} \to X_T$ ps (car $T < +\infty$ ps) et sous **2)**

$$\begin{aligned} X_{n\wedge T} &= \sum_{i=0}^{+\infty} X_{n\wedge T}\mathbf{1}_{\{T=i\}} = \sum_{i=0}^{+\infty} X_{n\wedge i}\mathbf{1}_{\{T=i\}} \\ \text{et} \quad |X_{n\wedge T}| &\leq \sum_{i=1}^{+\infty} |X_{n\wedge i}|\mathbf{1}_{\{T=i\}} \leq \sum_{i=1}^{+\infty} K\mathbf{1}_{\{T=i\}} \leq K. \end{aligned}$$

Comme $X_{T\wedge n} \xrightarrow{n\to+\infty} X_T$, on a aussi $|X_T| \leq K$. On a donc $X_T \in L^1$. Puis le théorème de convergence dominée s'applique pour donner (3.9) à partir de (3.10) :

$$\mathbb{E}[X_T] = \lim_{n\to+\infty} \mathbb{E}\big[X_{T\wedge n}\big] \geq \mathbb{E}[X_0].$$

3) Par **1)**, on a toujours (3.10) pour $T \wedge n$ avec $\lim_{n\to+\infty} X_{T\wedge n} = X_T$ ps ($T < +\infty$ ps). Sous **3)**, on peut écrire

$$\begin{aligned} X_T &= X_0 + \sum_{k=1}^{T\wedge n} (X_k - X_{k-1}) \\ \text{et} \quad |X_T| &\leq |X_0| + \sum_{k=1}^{T\wedge n} |X_k - X_{k-1}| \leq |X_0| + KT \in L^1 \end{aligned}$$

car $\mathbb{E}[T] < +\infty$. Le théorème de convergence dominée s'applique alors

$$\mathbb{E}[X_T] = \lim_{n\to+\infty} \mathbb{E}\big[X_{T\wedge n}\big] \geq \mathbb{E}[X_0],$$

ce qui assure (3.9). De plus, on a même

$$\mathbb{E}\big[|X_T|\big] = \lim_{n\to+\infty} \mathbb{E}\big[|X_{T\wedge n}|\big] \leq \mathbb{E}[|X_0|] + K\mathbb{E}[T] < +\infty$$

soit $X_T \in L^1$.

Enfin, si $(X_n)_{n\geq 0}$ est une martingale alors par le Th. 3.37, il y a égalité dans (3.10) et les passages à la limite précédents dans **2)**, **3)** les préservent. Puis comme en particulier, X est une sous-martingale, on a toujours $X_T \in L^1$ par le cas sous-martingale.

Puis, si $(X_n)_{n\geq 0}$ est une sur-martingale, on peut appliquer **1)**, **2)**, **3)** à la sous-martingale $(-X_n)_{n\geq 0}$ pour obtenir $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$ (noter que les hypothèses de **1)**, **2)**, **3)** sont insensibles aux changements de signe). Enfin, sous **4)**, partant de $\mathbb{E}[X_{T\wedge n}] \leq \mathbb{E}[X_0]$ pour la sur-martingale arrêtée $(X_n^T)_{n\geq 0}$ (avec $T\wedge n$ dû à **1)**, comme on a encore $\lim_{n\to+\infty} X_{T\wedge n} = X_T$ ($T < +\infty$ ps), le lemme de Fatou donne

$$\begin{aligned} \mathbb{E}[X_T] &= \mathbb{E}\Big[\lim_{n\to+\infty} X_{T\wedge n}\Big] = \mathbb{E}\Big[\liminf_{n\to+\infty} X_{T\wedge n}\Big] \\ &\leq \liminf_{n\to+\infty} \mathbb{E}\big[X_{T\wedge n}\big] \leq \mathbb{E}[X_0] \end{aligned}$$

en particulier comme $X_T \geq 0$, on a aussi $X_T \in L^1$ puisque $\mathbb{E}[X_0] < +\infty$. □

3.6 Décomposition de Doob

Théorème 3.40 (Décomposition de Doob) *Toute $(\mathcal{F}_n)$-sous-martingale $(X_n)_{n\geq 0}$ se décompose de façon (presque sûrement) unique sous la forme*

$$X_n = M_n + A_n \tag{3.11}$$

où $(M_n)_{n\geq 0}$ est une $(\mathcal{F}_n)$-martingale et $(A_n)_{n\geq 0}$ est une suite croissante $(\mathcal{F}_n)$-prévisible avec $A_0 = 0$ et donnée par

$$A_n = \sum_{k=1}^{n} \big(\mathbb{E}\big[X_k\,|\mathcal{F}_{k-1}\big] - X_{k-1}\big). \tag{3.12}$$

Démonstration : Existence. Pour avoir la décomposition (3.11), on doit nécessairement avoir

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

car $(M_n)_{n\geq 1}$ est une martingale et A_n est $\mathcal{F}_{n-1}$-mesurable. On pose donc

$$\begin{aligned}
A_n - A_{n-1} &:= \mathbb{E}\big[X_n\,|\mathcal{F}_{n-1}\big] - X_{n-1} \\
M_n &:= X_n - A_n,
\end{aligned} \tag{3.13}$$

ce qui définit les deux suites $(M_n)_{n\geq 0}$ et $(A_n)_{n\geq 0}$ en prenant en plus $A_0 = 0$ et $M_0 = X_0$. Il s'agit de vérifier que pour ce choix, $(M_n)_{n\geq 0}$ est bien une martingale et $(A_n)_{n\geq 0}$ est croissante, prévisible, la décomposition (3.11) étant satisfaite par construction.

Comme $(X_n)_{n\geq 1}$ est une sous-martingale, (3.13) assure

$$A_n - A_{n-1} = \mathbb{E}\big[X_n\,|\mathcal{F}_{n-1}\big] - X_{n-1} \geq 0, \tag{3.14}$$

et $A_n = (A_n - A_{n-1}) + A_{n-1}$ est bien $\mathcal{F}_{n-1}$-mesurable pour tout $n \geq 0$ par récurrence puisque $A_n - A_{n-1}$ l'est par (3.14). On note que $A_n \in L^1$ pour tout $n \geq 1$ puisque

$$\mathbb{E}\big[|A_n - A_{n-1}|\big] = \mathbb{E}\left[\big|\mathbb{E}\big[X_n\,|\mathcal{F}_{n-1}\big] - X_{n-1}\big|\right] \leq \mathbb{E}[|X_n|] + \mathbb{E}[|X_{n-1}|] < +\infty.$$

Pour montrer que $(M_n)_{n\geq 0}$ est une martingale, on observe d'abord que $M_n = X_n - A_n$ est $\mathcal{F}_n$-mesurable puis $M_n \in L^1$ car $X_n \in L^1$ ($(X_n)_{n\geq 0}$ sous-martingale) et $A_n \in L^1$. Pour la propriété de martingale (3.3), on utilise (3.13) et la $\mathcal{F}_{n-1}$-mesurabilité de A_n :

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

L'**unicité** est assurée par le raisonnement par condition nécessaire en début de démonstration. On peut aussi la vérifier directement en supposant qu'on a deux décompositions

$$X_n = M_n + A_n = M'_n + A'_n$$

avec $(M_n)_{n\geq 0}$, $(A_n)_{n\geq 0}$ et $(M'_n)_{n\geq 0}$, $(A'_n)_{n\geq 0}$ vérifiant les hypothèses du Th. 3.40. Alors $U_n = A_n - A'_n = M_n - M'_n$ définit une suite constante puisque

$$\begin{aligned} U_{n-1} &= M_{n-1} - M'_{n-1} = \mathbb{E}\big[M_n - M'_n\,|\mathcal{F}_{n-1}\big] \quad ((M_n)_{n\geq 0}, (M'_n)_{n\geq 0} \text{ martingales}) \\ &= \mathbb{E}\big[A_n - A'_n\,|\mathcal{F}_{n-1}\big] = A_n - A'_n = U_n \quad (A_n, A'_n \text{ sont } \mathcal{F}_{n-1}\text{-mesurables}). \end{aligned}$$

Comme $U_0 = 0$, on a $U_n = 0$ pour tout $n \geq 0$ et il vient $A_n = A'_n$ et $M_n = M'_n$ pour tout $n \geq 0$ presque sûrement. □

Remarque 3.41 La décomposition (3.11) est vraie pour toute suite $(X_n)_{n\geq 1}$ adaptée et L^1 avec $(M_n)_{n\geq 0}$ martingale et $(A_n)_{n\geq 0}$ prévisible partant de $A_0 = 0$. En fait $(X_n)_{n\geq 0}$ est une sous-martingale si et seulement si $(A_n)_{n\geq 0}$ est croissante, cf. (3.14).

Soit $(X_n)_{n\geq 0}$ une martingale de carré intégrable et nulle en 0. Comme $(X_n^2)_{n\geq 0}$ est une sous-martingale (Prop. 3.29), le Th. 3.40 donne la décomposition de Doob (3.11) suivante :

$$X_n^2 = M_n + A_n, \quad n \geq 0, \tag{3.15}$$

où $M = (M_n)_{n\geq 0}$ est une martingale et $A = (A_n)_{n\geq 0}$ est une suite prévisible croissante.

Définition 3.42 (Compensateur) *On note $\langle X, X\rangle$ le processus A croissant prévisible dans la décomposition de Doob (3.15) de X^2 et on l'appelle le compensateur de la martingale $X \in L^2$.*

Dans ce cas, l'expression (3.12) se réécrit :

$$\begin{aligned} \langle X, X\rangle_n &= \sum_{k=1}^n \big(\mathbb{E}\big[X_k^2\,|\mathcal{F}_{k-1}\big] - X_{k-1}^2\big) \\ &= \sum_{k=1}^n \mathbb{E}\big[(X_k - X_{k-1})^2\,|\mathcal{F}_{k-1}\big], \end{aligned} \tag{3.16}$$

où la formulation (3.16) vient du lemme suivant :

Lemme 3.43 (Formule de la variance conditionnelle) *Soit $(X_n)_{n\geq 1}$ une martingale telle que $\mathbb{E}[X_n^2] < +\infty$ pour tout $n \geq 1$. Alors pour $n \geq k$, on a*

$$\mathbb{E}\big[(X_n - X_k)^2\,|\mathcal{F}_k\big] = \mathbb{E}[X_n^2\,|\mathcal{F}_k] - X_k^2.$$

Démonstration : Pour $n \geq k$, on a :

$$\begin{aligned}\mathbb{E}\big[(X_n - X_k)^2\,|\mathcal{F}_k\big] &= \mathbb{E}\big[X_n^2 - 2X_nX_k + X_k^2\,|\mathcal{F}_k\big]\\ &= \mathbb{E}\big[X_n^2\,|\mathcal{F}_k\big] - 2X_k\mathbb{E}\big[X_n\,|\mathcal{F}_k\big] + X_k^2\\ &= \mathbb{E}\big[X_n^2\,|\mathcal{F}_k\big] - X_k^2.\end{aligned}$$

□

Dans l'expression (3.16), $\langle X, X\rangle_n$ apparaît comme la variance jusqu'à la date n et $\langle X, X\rangle_\infty$ (qui par croissance existe toujours quitte à avoir $+\infty$) est la variance totale de toute la suite $(X_n)_{n\geq 0}$.

Le comportement L^2 d'une martingale $(X_n)_{n\geq 0}$ de carré intégrable peut se lire sur son compensateur :

Proposition 3.44 (Martingale bornée dans L^2 et compensateur) *Soit $(X_n)_{n\geq 0}$ une martingale carré intégrable. Alors elle est bornée dans L^2 si et seulement si son compensateur vérifie $\mathbb{E}[\langle X, X\rangle_\infty] < +\infty$.*

Démonstration : Comme $\big(X_n^2 - \langle X, X\rangle_n)\big)_{n\geq 0}$ est une martingale, la propriété de martingale donne $\mathbb{E}\big[X_n^2 - \langle X, X\rangle_n] = \mathbb{E}[X_0^2 - \langle X, X\rangle] = \mathbb{E}[X_0^2$, c'est à dire

$$\mathbb{E}\big[\langle X, X\rangle_n\big] = \mathbb{E}[X_n^2] - \mathbb{E}[X_0^2],$$

pour tout $n \geq 0$. Comme $\langle X, X\rangle$ est une suite croissante, le théorème de convergence monotone donne alors

$$\mathbb{E}\big[\langle X, X\rangle_\infty\big] = \lim_{n\to+\infty} \mathbb{E}\big[\langle X, X\rangle_n\big] = \sup_{n\geq 0} \mathbb{E}\big[\langle X, X\rangle_n\big] = \Big(\sup_{n\geq 0} \mathbb{E}[X_n^2]\Big) - \mathbb{E}[X_0^2],$$

prouvant l'équivalence. □

Proposition 3.45 *Soit $(X_n)_{n\geq 0}$ une martingale de carré intégrable, nulle en 0 et T un temps d'arrêt. Alors*

$$\langle X^T, X^T\rangle = \langle X, X\rangle^T \qquad ps,$$

ie. le crochet de la martingale arrêtée est le crochet arrêté de la martingale.

Démonstration : Par la Prop. 3.36, $(X^T)^2 = M^T + \langle X, X\rangle^T$ est une sous-martingale, M^T est une martingale, et, par la Prop. 3.34, $\langle X, X\rangle^T$ est une suite croissante et prévisible. L'unicité (presque sûre) de la décomposition de Doob (3.15) de X^T exige $\langle X^T, X^T\rangle = \langle X, X\rangle^T$ ps. □

Exemple 3.46 (Somme de variables aléatoires *iid*) Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires *iid* centrées, de carrés intégrables avec $\mathrm{Var}(X_1) = \sigma^2$. Alors $S_n = \sum_{k=1}^n X_k$, $n \geq 1$, (avec $S_0 = 0$) est une martingale L^2 de compensateur $\langle S, S\rangle_n = n\sigma^2$:

En effet, d'après (3.16), en utilisant $X_k \perp\!\!\!\perp \mathcal{F}_{k-1} = \sigma(X_1, X_2, \dots, X_{k-1})$, on a

$$\langle S, S\rangle_n = \sum_{k=1}^n \mathbb{E}\big[(S_k - S_{k-1})^2\,|\mathcal{F}_{k-1}\big] = \sum_{k=1}^n \mathbb{E}\big[X_k^2\,|\mathcal{F}_{k-1}\big] = \sum_{k=1}^n \mathbb{E}[X_k^2] = n\sigma^2.$$

Chapitre 4

Convergences de martingales

Dans ce chapitre, on étudie les limites de martingales. On commence par les outils clef que sont les inégalités pour martingales en Section 4.1. On donne ensuite des résultats de convergence presque sûre en Section 4.2 puis en norme L^1 en Section 4.4 et en norme L^p en Section 4.5. On donne ensuite en Section 4.7 un résultat fondamental (théorème d'arrêt) qui généralise la propriété de martingale aux dates données par des temps d'arrêt.

Dans la suite, on considère un espace de probabilité filtré $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$. Par défaut, les (sur/sous)-martingales et temps d'arrêt sont par rapport à cette filtration $(\mathcal{F}_n)_{n\geq 0}$.

4.1 Inégalités de martingales

Les inégalités de martingales sont dues à Doob [1]. Essentiellement, elles donnent une bornes pour le sup d'une martingale sur $[0, n]$ par sa valeur en n, cf. (4.1)–(4.5), (4.11), (4.15). Étant donné une suite $(X_n)_{n\geq 0}$, on note $\overline{X}_n = \max_{0\leq k\leq n} X_k$.

4.1.1 Inégalité maximale de Doob

Théorème 4.1 (Inégalité maximale de Doob)

(1) Soit $(X_n)_{n\geq 0}$ une sous-martingale et $x > 0$. Alors

$$\mathbb{P}\big(\overline{X}_n \geq x\big) \leq \frac{\mathbb{E}\big[X_n \mathbf{1}_{\{\overline{X}_n \geq x\}}\big]}{x} \leq \frac{\mathbb{E}[X_n^+]}{x} \leq \frac{\mathbb{E}[|X_n|]}{x}. \tag{4.1}$$

(2) Soit $(X_n)_{n\geq 0}$ une sur-martingale, et $x > 0$. Alors

$$\mathbb{P}\big(\overline{X}_n \geq x\big) \leq \frac{\mathbb{E}[|X_0|] + \mathbb{E}[|X_n|]}{x}. \tag{4.2}$$

1. Joseph Leo Doob (1910–2004) probabiliste américain.

(3) Soit $(X_n)_{n\geq 0}$ une sur-martingale **positive**, *et $x>0$. Alors*

$$\mathbb{P}\big(\overline{X}_n \geq x\big) \leq \frac{\mathbb{E}[X_0]}{x}. \tag{4.3}$$

(4) Soit $(X_n)_{n\geq 0}$ une martingale, sous-martingale ou sur-martingale, et $x>0$, on a

$$\mathbb{P}\big(\max_{0\leq k\leq n} |X_k| \geq x\big) \leq \frac{\mathbb{E}[|X_0|]+2\mathbb{E}[|X_n|]}{x}. \tag{4.4}$$

(5) Pour une martingale $(X_n)_{n\geq 0}$, on peut améliorer (4.4) en :

$$\mathbb{P}\Big(\max_{0\leq k\leq n} |X_k| \geq x\Big) \leq \frac{\mathbb{E}\big[|X_n|\big]}{x}, \quad x>0. \tag{4.5}$$

Démonstration : 1) Les deux inégalités de droite dans (4.1) sont immédiates. Pour celle de gauche, on note $A=\big\{\overline{X}_n \geq x\big\}$ et on pose $S=\inf\big(k\geq 0 : X_k\geq x\big)$ et $T=S\wedge n$. Comme $T\leq n$, le Th. 3.37 assure $\mathbb{E}[X_T]\leq \mathbb{E}[X_n]$, ou encore

$$\mathbb{E}[X_T\mathbf{1}_A]+\mathbb{E}[X_T\mathbf{1}_{A^c}]=\mathbb{E}[X_T]\leq \mathbb{E}[X_n]=\mathbb{E}[X_n\mathbf{1}_A]+\mathbb{E}[X_n\mathbf{1}_{A^c}]. \tag{4.6}$$

Sur l'évènement $A^c=\big\{\overline{X}_n<x\big\}$, on a $S>n$ et donc $T=n$ et $X_T=X_n$. Il vient $\mathbb{E}[X_n\mathbf{1}_{A^c}]=\mathbb{E}[X_T\mathbf{1}_{A^c}]$ et (4.6) se réécrit $\mathbb{E}[X_T\mathbf{1}_A]\leq \mathbb{E}[X_n\mathbf{1}_A]$. De plus, sur l'évènement A, on a $S\leq n$ donc $T=S$ et par définition de S : $X_T=X_S\geq x$. Finalement, il vient

$$x\mathbb{P}(A)=\mathbb{E}[x\mathbf{1}_A]\leq \mathbb{E}[X_T\mathbf{1}_A]\leq \mathbb{E}[X_n\mathbf{1}_A],$$

ce qui prouve (4.1) puisque $x>0$.

2) et **3)** On adapte la preuve de 1) ci-dessus au cas d'une sur-martingale X. Pour $T\leq n$, le Th. 3.37 assure $\mathbb{E}[X_T]\leq \mathbb{E}[X_0]$, soit

$$\mathbb{E}[X_T\mathbf{1}_A]+\mathbb{E}[X_T\mathbf{1}_{A^c}]=\mathbb{E}[X_T]\leq \mathbb{E}[X_0]. \tag{4.7}$$

Comme précédemment, on a $\mathbb{E}[X_T\mathbf{1}_{A^c}]=\mathbb{E}[X_n\mathbf{1}_{A^c}]$ et $\mathbb{E}[X_T\mathbf{1}_A]\geq x\mathbb{P}(A)$ et (4.7) donne

$$x\mathbb{P}(A)\leq \mathbb{E}[X_0]-\mathbb{E}[X_n\mathbf{1}_{A^c}]\leq \begin{cases} \mathbb{E}[|X_0|]+\mathbb{E}[|X_n|] & \text{en général, d'où (4.2),}\\ \mathbb{E}[X_0] & \text{si } X_n\geq 0 \text{, d'où (4.3).}\end{cases}$$

4) Comme $\{\max_{k\leq n}|X_k|\geq x\}\subset\{\max_{k\leq n}X_k\geq x\}\cup\{\max_{k\leq n}(-X_k)\geq x\}$, on a

$$\mathbb{P}\big(\max_{0\leq k\leq n}|X_k|\geq x\big)\leq \mathbb{P}\big(\max_{0\leq k\leq n}X_k\geq x\big)+\mathbb{P}\big(\max_{0\leq k\leq n}(-X_k)\geq x\big) \tag{4.8}$$

et $(X_n)_{n\geq 0}$, $(-X_n)_{n\geq 0}$ sont des sous-martingales et sur-martingales (ou l'inverse) et on majore chaque terme de (4.8) par (4.1) et (4.2) pour avoir la conclusion (4.4).

5) Lorsque $(X_n)_{n\geq 0}$ est une martingale, on peut appliquer (4.1) à la sous-martingale positive $(|X_n|)_{n\geq 0}$ et avoir (4.5). □

Corollaire 4.2 (Inégalité maximale de Kolmogorov) *Soit $(X_n)_{n\geq 1}$ des variables aléatoires indépendantes centrées et de variances finies. On pose $S_n = X_1 + \cdots + X_n$. Alors pour $x > 0$, on a :*

$$\mathbb{P}\Big(\max_{1\leq k\leq n} |S_k| \geq x\Big) \leq \frac{\mathrm{Var}(S_n)}{x^2}. \tag{4.9}$$

Démonstration : Dans l'Exemple 3.20, on a vu que que $(S_n)_{n\geq 1}$ est une martingale pour la filtration canonique engendrée par la suite $(X_n)_{n\geq 1}$. Par le Corollaire 3.30, $Y_n = S_n^2$, $n \geq 1$, définit une sous-martingale à laquelle on applique l'inégalité maximale de Doob (4.1) avec $u = x^2$ (Th. 4.1). On obtient alors l'inégalité maximale de Kolmogorov (4.9) puisque $\mathbb{E}[S_n^2] = \mathrm{Var}(S_n)$. □

Remarque 4.3 (Comparaison avec Tchebychev) Dans le contexte du Corollaire 4.2, l'inégalité de Tchebychev donne pour tout $1 \leq k \leq n$:

$$\mathbb{P}(|S_k| \geq x) \leq \frac{\mathrm{Var}(S_k)}{x^2} \leq \frac{\mathrm{Var}(S_n)}{x^2}$$

car $\mathrm{Var}(S_k) = \sum_{i=1}^k \mathbb{E}[X_i^2] \leq \sum_{i=1}^n \mathbb{E}[X_i^2] = \mathrm{Var}(S_n)$. On a donc

$$\max_{1\leq k\leq n} \mathbb{P}(|S_k| \geq x) \leq \frac{\mathrm{Var}(S_n)}{x^2}. \tag{4.10}$$

Comme

$$\max_{1\leq k\leq n} \mathbb{P}(|S_k| \geq x) \leq \mathbb{P}\Big(\bigcup_{k=1}^n \{|S_k| \geq x\}\Big) = \mathbb{P}\Big(\max_{1\leq k\leq n} |S_k| \geq x\Big),$$

l'inégalité (4.9) est meilleure que (4.10).

4.1.2 Inégalité de moments de Doob

Théorème 4.4 (Inégalité de moments pour sous-martingale) *Soit $(X_n)_{n\geq 0}$ une sous-martingale avec $X_0 \geq 0$. Alors pour $p > 1$, on a*

$$\mathbb{E}\big[\overline{X}_n^p\big] \leq \Big(\frac{p}{p-1}\Big)^p \mathbb{E}\big[(X_n^+)^p\big]. \tag{4.11}$$

Démonstration : Comme $X_0 \geq 0$, on note que $\overline{X}_n \geq 0$ pour chaque $n \geq 0$. Si $\mathbb{E}\big[\overline{X}_n^p\big] = 0$, alors (4.11) est immédiate, on suppose donc $\mathbb{E}\big[\overline{X}_n^p\big] > 0$ et par convergence monotone, pour M assez grand, on a $\mathbb{E}\big[(\overline{X}_n \wedge M)^p\big] > 0$. On a donc $0 < \mathbb{E}\big[(\overline{X}_n \wedge M)^p\big] \leq M$. On utilise l'inégalité maximale de Doob (4.1) (Th. 4.1) avec des variables aléatoires tronquées au niveau $M > 0$:

$$\mathbb{E}\big[(\overline{X}_n \wedge M)^p\big] = \mathbb{E}\left[\int_0^{\overline{X}_n\wedge M} px^{p-1}dx\right] = \mathbb{E}\left[\int_0^M \mathbf{1}_{\{x\leq \overline{X}_n\}} px^{p-1}\,dx\right]$$

$$\begin{aligned}
&= \int_0^M px^{p-1}\mathbb{P}\big(\overline{X}_n \geq x\big)\,dx \quad \text{(Fubini-Tonelli)}\\
&\leq \int_0^M px^{p-1}\Big(x^{-1}\mathbb{E}\big[X_n^+\mathbf{1}_{\{\overline{X}_n\geq x\}}\big]\Big)\,dx \quad \text{(Th. 4.1-1)}\\
&= p\mathbb{E}\left[X_n^+\int_0^{\overline{X}_n\wedge M} x^{p-2}\,dx\right] \quad \text{(Fubini-Tonelli)}\\
&= \frac{p}{p-1}\mathbb{E}\big[X_n^+\big(\overline{X}_n\wedge M\big)^{p-1}\big]\\
&\leq \Big(\frac{p}{p-1}\Big)\mathbb{E}\big[\big(\overline{X}_n\wedge M\big)^p\big]^{(p-1)/p}\mathbb{E}\big[(X_n^+)^p\big]^{1/p} \quad \text{(Hölder)}.
\end{aligned}$$

En simplifiant la borne précédente, il vient

$$\mathbb{E}\big[(\overline{X}_n\wedge M)^p\big] \leq \Big(\frac{p}{p-1}\Big)^p\mathbb{E}\big[(X_n^+)^p\big].$$

Noter que pour simplifier la borne comme ci-dessus, il est nécessaire que $\mathbb{E}\big[(\overline{X}_n\wedge M)^p\big]$ soit fini et non nul, d'où l'importance de tronquer par M, assez grand. Finalement, on obtient (4.11) en faisant $M\to+\infty$ avec le théorème de convergence monotone. □

Remarque 4.5 Attention, l'inégalité maximale L^p (Th. 4.4) est fausse pour $p=1$ même avec une autre constante.

Pour des martingales, on spécialise le Th. 4.4 comme suit :

Corollaire 4.6 (Inégalité de moments pour martingale)

(1) Pour une martingale $(X_n)_{n\geq 0}$, on a :

$$\mathbb{E}\Big[\big(\max_{1\leq k\leq n}|X_k|\big)^p\Big] \leq \Big(\frac{p}{p-1}\Big)^p\mathbb{E}\big[|X_n|^p\big]. \tag{4.12}$$

(2) Pour une martingale $(X_n)_{n\geq 0}$ nulle en 0 (donc centrée) de carré intégrable, on a :

$$\mathbb{E}\Big[\big(\max_{1\leq k\leq n}|X_k|\big)^2\Big] \leq 4\mathbb{E}\big[\langle X,X\rangle_n^2\big] \tag{4.13}$$

$$\mathbb{E}\Big[\big(\sup_{n\geq 1}|X_n|\big)^2\Big] \leq 4\mathbb{E}\big[\langle X,X\rangle_\infty\big]. \tag{4.14}$$

Démonstration : On considère une martingale $(X_n)_{n\geq 0}$.

1) s'obtient directement en appliquant le Th. 4.4 à la sous-martingale $Y_n=|X_n|$, $n\geq 0$.

2) est une spécialisation de **1)** lorsque $p=2$. Dans ce cas, on utilise que $X^2-\langle X,X\rangle$ est une martingale centrée pour avoir $\mathbb{E}\big[X_n^2-\langle X,X\rangle_n\big]=0$, soit $\mathbb{E}\big[X_n^2\big]=\mathbb{E}[\langle X,X\rangle_n]$, ce qui assure (4.13) en l'injectant dans (4.12). Pour (4.14), on utilise la convergence monotone

$$\mathbb{E}\Big[\big(\sup_{n\geq 1}|X_n|\big)^2\Big] = \lim_{n\to+\infty}\mathbb{E}\Big[\big(\max_{1\leq k\leq n}|X_k|\big)^2\Big] \leq \lim_{n\to+\infty}4\mathbb{E}\big[\langle X,X\rangle_n^2\big] = 4\mathbb{E}\big[\langle X,X\rangle_\infty\big].$$

□

4.1.3 Nombre de montées

Une sous-martingale $(X_n)_{n\geq 1}$ a une tendance à croître comme l'indique la croissance de l'espérance $\mathbb{E}[X_n] \leq \mathbb{E}[X_{n+1}]$, $n \geq 0$. Cette croissance peut être contrôlée par l'inégalité sur le nombre de montées à travers un intervalle $[a, b]$ où $a < b$, cf. Th. 4.8.

Pour présenter cette inégalité, on pose $N_0 = 0$ et on définit les variables aléatoires N_k, $k \geq 1$, à valeurs dans $\mathbb{N} \cup \{+\infty\}$:

$$N_1 = \min\big(n \geq 1 : X_n \leq a\big), \quad N_2 = \min\big(n \geq N_1 : X_n \geq b\big),$$

et par récurrence pour $k \geq 2$:

$$N_{2k-1} = \min\big(n \geq N_{2k-2} : X_n \leq a\big), \quad N_{2k} = \min\big(n \geq N_{2k-1} : X_n \geq b\big),$$

avec la convention $\inf \emptyset = +\infty$. On a

$$N_1 < N_2 < \cdots < N_{2k-2} < N_{2k-1} < N_{2k} < \cdots$$

et $N_k \geq k$ ps.

Lemme 4.7 *Les variables aléatoires N_k, $k \geq 1$, sont des temps d'arrêt.*

Démonstration : D'abord

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

puisque $\{X_k \leq a\} \in \mathcal{F}_k \subset \mathcal{F}_n$ pour $0 \leq k \leq n$. Puis

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

puisque $\{N_1 = j\} \in \mathcal{F}_j \subset \mathcal{F}_n$ pour $1 \leq j \leq n-1$ et $\{X_k < b\} \in \mathcal{F}_k \subset \mathcal{F}_n$ pour $j+1 \leq k \leq n-1$ et $\{X_n \geq b\} \in \mathcal{F}_n$. Ensuite par récurrence sur k :

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

puisque $\{N_{2k-2} = j\} \in \mathcal{F}_j \subset \mathcal{F}_n$ pour $1 \leq j \leq n-1$ (hypothèse de récurrence) et $\{X_k > a\} \in \mathcal{F}_k \subset \mathcal{F}_n$ pour $j+1 \leq k \leq n-1$ et $\{X_n \leq a\} \in \mathcal{F}_n$. Et enfin, de même :

$$\big\{N_{2k} = n\big\} = \bigcup_{j=1}^{n-1} \{N_{2k-1} = j\} \cap \{N_{2k} = n\}$$

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

puisque $\{N_{2k-1} = j\} \in \mathcal{F}_j \subset \mathcal{F}_n$ pour $1 \leq j \leq n-1$ (hypothèse de récurrence) et $\{X_k < b\} \in \mathcal{F}_k \subset \mathcal{F}_n$ pour $j+1 \leq k \leq n-1$ et $\{X_n \geq b\} \in \mathcal{F}_n$. □

Comme

$$X_{N_{2k-1}} \leq a \quad \text{et} \quad X_{N_{2k}} \geq b,$$

entre les dates N_{2k-1} et N_{2k}, $(X_n)_{n\geq 1}$ monte d'au dessous de a à au dessus de b (exactement une fois). Notons $U_n([a,b])$ le nombre de telles montées le long de l'intervalle $[a,b]$ de la suite $(X_n)_{n\geq 1}$ jusqu'à la date n, c'est à dire

$$U_n([a,b]) = \sup\big(k \in \mathbb{N} : N_{2k} \leq n\big).$$

Comme $U_n([a,b])$ est le nombre de montées de $X_1, \dots, X_n$ le long de $[a,b]$, on a immédiatement $U_n([a,b]) \leq [n/2]$. En observant que $N_{2k} \leq n < N_{2k+2}$ signifie qu'il y a eu exactement k montées réalisées jusqu'à la date n, on peut écrire

$$U_n([a,b]) = \sum_{k=1}^{[n/2]} k\, \mathbf{1}_{\{N_{2k} \leq n < N_{2k+2}\}}.$$

Ainsi chaque $U_n([a,b])$ est positive, bornée par $[n/2]$ et donc intégrable. La suite $(U_n([a,b]))_{n\geq 1}$ est croissante. Le nombre de montées d'une sous-martingale est contrôlé (en moyenne) par l'inégalité suivante due à Doob :

Théorème 4.8 (Nombre de montées) *Soit $(X_n)_{n\geq 0}$ une sous-martingale. Alors pour tout $a < b$, on a :*

$$\mathbb{E}\big[U_n([a,b])\big] \leq \frac{\mathbb{E}[(X_n - a)^+]}{b-a}. \tag{4.15}$$

Remarque : Le Th. 4.8 s'applique pour une suite finie $(X_k)_{k=0,\dots,n}$ qui forme une martingale : $\mathbb{E}[X_{k+1}|\mathcal{F}_k] = X_k$ pour tout $0 \leq k < n$.

Comme les N_k, $k \geq 1$, sont des temps d'arrêt, on a

$$\big\{N_{2k+1} < j \leq N_{2k+2}\big\} = \big\{N_{2k+1} \leq j-1\big\} \cap \big\{N_{2k+2} \leq j-1\big\}^c \in \mathcal{F}_{j-1},$$

de sorte que

$$Y_j = \begin{cases} 1 & \text{si pour un } k \in \mathbb{N}, \text{ on a } N_{2k} < j \leq N_{2k+1} \\ 0 & \text{sinon, ie. pour un } k, \text{ on a } N_{2k+1} < j \leq N_{2k+2}, \end{cases}$$

$j \geq 1$, définit une suite prévisible.

- On a $Y_1 = 1$ car $0 = N_0 < 1 \leq N_1$ correspond à $k = 0$;
- On a $Y_2 = \mathbf{1}_{\{N_1 \geq 2\}} = \mathbf{1}_{\{X_1 > a\}}$ car $N_2 \geq 2$ exige d'avoir encore $k = 0$ dans la définition de Y_2 ;

— Pour $j \geq 2$, on a $Y_j = 1$ si et seulement si X_j est dans une descente (de b vers a) de la suite $(X_n)_{n\geq 0}$. Il y a 2 possibilités pour que X_j soit dans une phase de descente :
— soit X_{j-1} est dans une descente qui n'est pas finie ($Y_{j-1} = 1, X_{j-1} > a$) ;
— soit X_{j-1} est dans une montée qui termine ($Y_{j-1} = 0, X_{j-1} \geq b$).
On a donc aussi

$$Y_j = \mathbf{1}_{\{Y_{j-1}=0, X_{j-1}\geq b\}\cup\{Y_{j-1}=1, X_{j-1}>a\}}$$

et on retrouve (par récurrence) que Y_j est $\sigma(X_1, \dots, X_{j-1})$-mesurable, et donc la suite $(Y_j)_{j\geq 1}$ est prévisible.

On a alors :

Lemme 4.9 *Pour toute suite finie $X_1, X_2, \dots, X_n$, on a*

$$\sum_{k=2}^{n} Y_k(X_k - X_{k-1}) \leq (a-b)U_n([a,b]) + (X_n - a)^+. \tag{4.16}$$

Démonstration :[Lemme 4.9]
Cas 1 : $U_n([a,b]) = 0$, ie. il n'y a aucune montée jusqu'à la date n, donc $N_2 > n$.

(a) Si $N_1 = 1$ alors pour $k = 2, \dots, n$: $N_1 = 1 < k \leq n < N_2$ et $Y_k = 0$ et (4.16) est immédiate car son membre de gauche se réduit à 0.

(b) Si $1 < N_1 \leq n$ alors pour $2 \leq k \leq N_1$: $N_0 = 0 < 1 \leq k \leq N_1$ et $Y_k = 1$, et pour $N_1 < k \leq n$: $N_1 < k \leq n < N_2$ et $Y_k = 0$. On a donc :

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

puisque $X_{N_1} \leq a < X_1$ ($N_1 > 1$) et (4.16) est vraie.

(c) Si $N_1 > n$ alors pour $2 \leq k \leq n$: $N_0 = 0 < 2 \leq k \leq n < N_1$ et $Y_k = 1$. On a donc :

$$\sum_{k=2}^{n} Y_k(X_k - X_{k-1}) = X_n - X_1 \leq X_n - a \leq (X_n - a)^+.$$

Cas 2 : $U_n([a,b]) > 0$, ie. il y a au moins 1 montée avant la date n donc en particulier $N_1 < N_2 \leq n$.
Pour $N_0 = 0 < 1 \leq k \leq N_1$, on a $Y_k = 1$, et pour $N_1 < k \leq N_2$, on a $Y_k = 0$. Ainsi

$$\begin{aligned} \sum_{k=2}^{n} Y_k(X_k - X_{k-1}) &= \overbrace{(X_{N_1} - X_1)}^{\leq 0} + \sum_{k=N_2+1}^{n} Y_k(X_k - X_{k-1}) \\ &\leq \sum_{k=N_2+1}^{n} Y_k(X_k - X_{k-1}), \end{aligned} \tag{4.17}$$

car soit $N_1 = 1$ et $X_{N_1} - X_1 = 0$, soit $N_1 > 1$ et $X_{N_1} \leq a < X_1$.

Il y a maintenant deux sous-cas complémentaires à considérer dans (4.17) selon que

(a) soit la date n correspond à une phase de montée : pour un $\ell \in \mathbb{N}$, on a $N_{2\ell+1} < n \leq N_{2\ell+2}$ et $U_n([a,b]) = \ell$;

(b) soit la date n correspond à une phase de descente : pour un $\ell' \in \mathbb{N}$, on a $N_{2\ell'} < n \leq N_{2\ell'+1}$ et $U_n([a,b]) = \ell'$.

Dans le sous-cas (a) (phase de montée), on a

$$\begin{aligned}
\sum_{k=N_2+1}^{n} Y_k(X_k - X_{k-1}) &= (X_{N_3} - X_{N_2}) + \cdots + (X_{N_{2\ell+1}} - X_{N_{2\ell}}) + \overbrace{0}^{Y_k=0 \text{ pour les termes résiduels } k\in]N_{2\ell+1},n]} \\
&\leq (a-b)\ell = (a-b)U_n([a,b])
\end{aligned}$$

car $X_{N_{2s+1}} - X_{N_{2s}} \leq a - b$ pour chaque $s = 1, \dots, \ell$ et (4.16) suit dans ce sous-cas.

Dans le sous-cas (b) (phase de descente), on a

$$\begin{aligned}
&\sum_{k=N_2+1}^{n} Y_k(X_k - X_{k-1}) \\
&= (X_{N_3} - X_{N_2}) + \cdots + (X_{N_{2\ell'-1}} - X_{N_{2\ell'-2}}) + \overbrace{(X_n - X_{N_{2\ell'}})}^{Y_k=1 \text{ pour les termes résiduels } k\in]N_{2\ell'},n]} \\
&= \underbrace{(X_{N_3} - X_{N_2}) + \cdots + (X_{N_{2\ell'-1}} - X_{N_{2\ell'-2}})}_{\leq(\ell'-1)(a-b)} + \underbrace{(a - X_{N_{2\ell'}})}_{\leq a-b} + (X_n - a) \\
&\leq (a-b)\ell' + (X_n - a) \\
&\leq (a-b)U_n([a,b]) + (X_n - a)^+
\end{aligned}$$

car de nouveau $X_{N_{2s+1}} - X_{N_{2s}} \leq a - b$ pour chaque $s = 1, \dots, \ell' - 1$, $U_n(a,b) = \ell'$ et $a < b$. La conclusion (4.16) suit encore dans ce sous-cas, ce qui prouve le Lemme 4.9. □

Démonstration :[Th. 4.8] Comme $(X_n)_{n\geq1}$ est une sous-martingale et comme on a vu que Y_k est $\mathcal{F}_{k-1}$-mesurable positive, on a

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

par la propriété de sous-martingale et parce que $Y_k \geq 0$. Avec le Lemme 4.9, on a donc

$$0 \leq \mathbb{E}\left[\sum_{k=2}^{n} Y_k(X_k - X_{k-1})\right] \leq (a-b)\mathbb{E}\big[U_n([a,b])\big] + \mathbb{E}\big[(X_n - a)^+\big],$$

d'où il suit $(b-a)\mathbb{E}\big[U_n([a,b])\big] \leq \mathbb{E}\big[(X_n - a)^+\big]$ et donc (4.15) puisque $a < b$. □

4.2 Convergence presque sûre de martingales

Théorème 4.10 (Convergence ps de sous-martingale) *Soit $(X_n)_{n\geq 0}$ une sous-martingale telle que $\sup_{n\geq 0}\mathbb{E}[X_n^+]<+\infty$. Alors quand $n\to+\infty$, $(X_n)_{n\geq 0}$ converge ps vers une limite $X\in L^1$.*

Remarque 4.11 Comme on l'a vu en Remarque 3.28, les sous-martingales sont des analogues aléatoires de suites croissantes et le résultat précédent généralise à ces objets le résultat bien connu pour les suites majorées qui convergent !

Lemme 4.12 *Une suite $(x_n)_{n\geq 1}$ converge dans $\overline{\mathbb{R}}$ si et seulement si pour tout rationnels $a<b$, le nombre de montées de $(x_n)_{n\geq 1}$ le long de $[a,b]$ vérifie $U_\infty([a,b],x)<+\infty$.*

Démonstration : On a $(x_n)_{n\geq 1}$ diverge si et seulement si $\liminf_{n\to+\infty} x_n < \limsup_{n\to+\infty} x_n$ c'est à dire s'il existe $a<b$ rationnels tels que

$$\liminf_{n\to+\infty} x_n < a < b < \limsup_{n\to+\infty} x_n.$$

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

Démonstration : Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires. D'après le Lemme 4.12, $(X_n)_{n\geq 0}$ converge vers une limite (finie ou pas) si et seulement si $U_\infty([a,b])<+\infty$. Pour voir cela, on utilise le Th. 4.8 sur le nombre de montées.

Soit $a<b$. Comme le nombre de montées $U_n([a,b])$ le long de $[a,b]$ est une suite croissante, on note $U_\infty([a,b]) := \lim_{n\to+\infty} U_n([a,b])$ pour le nombre total de montées le long de $[a,b]$. Avec le théorème de convergence monotone, l'inégalité sur le nombre de montées (Th. 4.8) donne

$$\begin{aligned}
\mathbb{E}\big[U_\infty([a,b])\big] &= \lim_{n\to+\infty}\mathbb{E}\big[U_n([a,b])\big] = \sup_{n\geq 0}\mathbb{E}\big[U_n([a,b])\big]\\
&\leq \frac{|a|+\sup_{n\geq 0}\mathbb{E}[X_n^+]}{(b-a)} < +\infty
\end{aligned}$$

en utilisant $(X_n-a)^+\leq X_n^+ + |a|$ et l'hypothèse d'intégrabilité. On a donc $\mathbb{E}[U_\infty([a,b])] < +\infty$ et $U_\infty([a,b])<+\infty$ ps. Comme pour tout $a<b$ on a

$$\Big\{\liminf_{n\to+\infty} X_n < a < b < \limsup_{n\to+\infty} X_n\Big\} \subset \{U_\infty([a,b])=+\infty\},$$

il suit que l'évènement

$$\bigcup_{a,b\in\mathbb{Q}}\Big\{\liminf_{n\to+\infty} X_n < a < b < \limsup_{n\to+\infty} X_n\Big\}$$

est de probabilité nulle et donc $X := \lim_{n\to+\infty} X_n$ existe presque sûrement par le Lemme 4.12. Ensuite, par le lemme de Fatou

$$\mathbb{E}[X^+] \;=\; \mathbb{E}\Big[\lim_{n\to+\infty} X_n^+\Big] = \mathbb{E}\Big[\liminf_{n\to+\infty} X_n^+\Big]$$

$$\leq \liminf_{n\to+\infty} \mathbb{E}[X_n^+] \leq \sup_{n\geq 0} \mathbb{E}[X_n^+] < +\infty,$$

ce qui assure $X < +\infty$ ps. Pour s'assurer aussi de $X > -\infty$ ps, on utilise $X_n = X_n^+ - X_n^-$ et la propriété de sous-martingale pour $(X_n)_{n\geq 0}$:

$$\mathbb{E}[X_n^-] = \mathbb{E}[X_n^+] - \mathbb{E}[X_n] \leq \mathbb{E}[X_n^+] - \mathbb{E}[X_0].$$

Il suit alors encore par le lemme de Fatou

$$\begin{aligned}\mathbb{E}[X^-] &= \mathbb{E}\Big[\lim_{n\to+\infty} X_n^-\Big] = \mathbb{E}\Big[\liminf_{n\to+\infty} X_n^-\Big] \\ &\leq \liminf_{n\to+\infty} \mathbb{E}[X_n^-] < \sup_{n\geq 0} \mathbb{E}[X_n^+] - \mathbb{E}[X_0] < +\infty.\end{aligned}$$

On en déduit $X^- < +\infty$ et finalement $|X| = X^+ + X^- < +\infty$ ps et $\mathbb{E}[|X|] = \mathbb{E}[X^+] + \mathbb{E}[X^-] < +\infty$. □

Remarque 4.13 Dans la preuve, on a établi que si le nombre de montées de $(X_n)_{n\geq 1}$ sur $]a, b[$ est fini pour tout $a, b \in \mathbb{Q}$ alors la limite de X_n existe ps.

Corollaire 4.14 (Convergence ps de martingale bornée dans L^1) *Soit $(X_n)_{n\geq 0}$ une martingale ou sous/sur-martingale bornée dans L^1 ($\sup_{n\geq 0} \mathbb{E}[|X_n|] < +\infty$). Alors $(X_n)_{n\geq 0}$ converge presque sûrement.*

Démonstration : Il suffit de considérer le cas de $(X_n)_{n\geq 0}$ sous-martingale, les autres s'en déduisent facilement. Comme $\mathbb{E}[X_n^+] \leq \mathbb{E}[|X_n|]$, on a $\sup_{n\geq 0} \mathbb{E}[X_n^+] \leq \sup_{n\geq 0} \mathbb{E}[|X_n|]$ et la condition du Th. 4.10 est satisfaite lorsque la sous-martingale est bornée dans L^1, justifiant la convergence presque sûre. □

Le corollaire suivant généralise « une suite positive décroissante converge ! »

Corollaire 4.15 (Convergence ps de sur-martingale positive)
Soit $(X_n)_{n\geq 1}$ une sur-martingale positive ($X_n \geq 0$ pour tout $n \geq 0$). Alors, quand $n \to +\infty$, $X_n \xrightarrow{ps} X$ avec $\mathbb{E}[X] \leq \mathbb{E}[X_0]$.

Démonstration : On définit une sous-martingale négative en prenant $Y_n = -X_n$, $n \geq 0$. On a $\sup_{n\geq 1} \mathbb{E}[Y_n^+] = 0 < +\infty$ et $(Y_n)_{\geq 0}$ converge ps par le Th. 4.10. Comme par la propriété de sur-martingale $(\mathbb{E}[X_n])_{n\geq 0}$ décroît on a $\mathbb{E}[X_n] \leq \mathbb{E}[X_0]$. Le lemme de Fatou donne alors

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

□

La convergence presque sûre du Corollaire 4.15 peut ne pas être L^1 comme le montre l'exemple qui suit :

Exemple 4.16 (Martingale qui converge ps mais pas L^1)
On reprend l'Exemple 3.20 de la marche aléatoire symétrique avec une suite de variables aléatoires $(X_i)_{i\geq 1}$ *iid* de loi donnée par $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}$. On considère $S_n = S_{n-1} + X_n$ avec $S_0 = 1$. On pose $T = \inf\big(n \geq 0 : S_n = 0\big)$ le temps d'atteinte de 0 de la marche et on considère $Y_n = S_{T\wedge n}$. Comme on observe que T est un temps d'arrêt, il vient que $Y = (Y_n)_{n\geq 0}$ est une martingale (Prop. 3.36) positive (par définition de l'arrêt en T, temps d'atteinte de 0).

Par le Corollaire 4.15, on a $Y_n \xrightarrow{ps} Y_\infty$.

On doit avoir $Y_\infty = 0$: en effet, $Y_n \in \mathbb{N}$ et la convergence vers $k \in \mathbb{N}$ est impossible puisque si $Y_n = k > 0$ alors $Y_{n+1} = k\pm 1$. (Une suite entière convergeant ne peut converger vers un entier qu'en devenant stationnaire, ce qui n'est possible que si la marche a atteint 0 où elle est arrêtée !). Comme sur $\{T = +\infty\}$, on a $|S_{T\wedge(n+1)} - S_{T\wedge n}| = |S_{n+1} - S_n| = 1$, cela oblige d'avoir $\mathbb{P}(T = +\infty) = 0$ et donc $T < +\infty$ ps.

Comme par la propriété de martingale, $\mathbb{E}[Y_n] = \mathbb{E}[Y_0] = 1$, on ne peut pas avoir la convergence L^1 de Y_n vers $Y_\infty = 0$.

4.3 Uniforme intégrabilité

La notion d'uniforme intégrabilité qu'on introduit dans cette section sera utile pour étudier la convergence dans L^1 de martingales en Section 4.4. On pourra aussi consulter [Bre-proba].

Définition 4.17 (Uniforme intégrabilité) *Une suite de variables aléatoires intégrables $(X_n)_{n\geq 0}$ est dite uniformément intégrable (UI) si*

$$\lim_{c\to+\infty} \sup_{n\geq 0} \mathbb{E}\big[|X_n|\mathbf{1}_{\{|X_n|>c\}}\big] = 0.$$

Remarque 4.18
- La même définition s'applique à une famille non-dénombrable de variables aléatoires.
- Une famille de variables aléatoires avec un seul élément (intégrable) est uniformément intégrable (par convergence dominée !).
- Une suite de variables aléatoires dominées par une variable aléatoire Z intégrable est uniformément intégrable. En effet par croissance de $x \in \mathbb{R}_+ \mapsto x\mathbf{1}_{\{x>c\}}$, $|X_n|\mathbf{1}_{\{|X_n|>c\}} \leq Z\mathbf{1}_{\{Z>c\}}$, d'où :

$$\lim_{c\to+\infty} \sup_{n\geq 0} \mathbb{E}\big[|X_n|\mathbf{1}_{\{|X_n|>c\}}\big] \leq \lim_{c\to+\infty} \mathbb{E}[Z\mathbf{1}_{\{Z>c\}}] = 0,$$

  où la dernière limite s'obtient par convergence dominée avec $Z \in L^1$.
- Une suite finie de variables aléatoires intégrables est uniformément intégrable. En effet, une telle suite $(X_k)_{k=1,\dots,n}$ est dominée par $\sum_{k=1}^n |X_k|$ intégrable.
- Si $(X_n)_{n\geq 0}$ et $(Y_n)_{n\geq 0}$ sont deux suites de variables aléatoires avec $|X_n| \leq |Y_n|$ pour tout $n \geq 0$ et $(Y_n)_{n\geq 0}$ uniformément intégrable alors $(X_n)_{n\geq 0}$ l'est aussi.

Proposition 4.19 *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires dans L^1 telle que pour $\delta>0$ la suite est bornée dans $L^{1+\delta}$. Alors la suite $(X_n)_{n\geq 0}$ est uniformément intégrable.*

Démonstration : On a :

$$\begin{aligned}\sup_{n\geq 1}\mathbb{E}\big[|X_n|\mathbf{1}_{\{|X_n|>c\}}\big] &= \sup_{n\geq 1}\mathbb{E}\big[|X_n|\times 1\times \mathbf{1}_{\{|X_n|/c>1\}}\big]\\ &\leq \sup_{n\geq 1}\mathbb{E}\Big[|X_n|\times \frac{|X_n|^\delta}{c^\delta}\times \mathbf{1}_{\{|X_n|/c>1\}}\Big]\\ &\leq \frac{1}{c^\delta}\sup_{n\geq 1}\mathbb{E}\big[|X_n|^{1+\delta}\big]=O\big(c^{-\delta}\big)\to 0,\quad c\to+\infty.\end{aligned}$$

□

Rappelons la propriété suivante d'une variable aléatoire intégrable :

Lemme 4.20 *Soit X une variable aléatoire intégrable. Alors pour tout $\varepsilon>0$, il existe $\eta>0$ tel que si $A\in\mathcal{F}$ avec $\mathbb{P}(A)<\eta$ alors $\mathbb{E}[|X|\mathbf{1}_A]<\varepsilon$.*

Démonstration : Par le théorème de convergence dominée, pour c assez grand, on a $\mathbb{E}[|X|\mathbf{1}_{\{|X|>c\}}]\leq \varepsilon/2$; puis pour $\eta<\varepsilon/(2c)$, on a

$$\mathbb{E}\big[|X|\mathbf{1}_A\big]=\mathbb{E}\big[|X|\mathbf{1}_{A\cap\{|X|>c\}}\big]+\mathbb{E}\big[|X|\mathbf{1}_{A\cap\{|X|\leq c\}}\big]\leq\frac{\varepsilon}{2}+c\,\mathbb{P}(A)\leq\frac{\varepsilon}{2}+c\,\eta<\varepsilon.$$

□

Pour des variables aléatoires uniformément intégrables, on a la généralisation suivante de ce rappel :

Proposition 4.21 (Critère d'uniforme intégrabilité) *Une suite de variables aléatoires réelles $(X_n)_{n\geq 0}$ est uniformément intégrable si et seulement si*

(i) $\forall\varepsilon>0$, $\exists\eta>0$ *tel que pour* $A\in\mathcal{F}$ *avec* $\mathbb{P}(A)<\eta$ *on a* $\mathbb{E}[|X_n|\mathbf{1}_A]<\varepsilon$ *pour tout* $n\in\mathbb{N}$.

(ii) $\sup_{n\geq 0}\mathbb{E}[|X_n|]<+\infty$ *(ie. la famille est bornée dans L^1).*

Démonstration : On suppose d'abord que $(X_n)_{n\geq 0}$ est uniformément intégrable : pour tout $\varepsilon>0$, $\exists c>0$ tel que $\sup_{n\geq 0}\mathbb{E}\big[|X_n|\mathbf{1}_{\{|X_n|>c\}}\big]<\varepsilon/2$. Alors pour $A\in\mathcal{F}$ et $n\in\mathbb{N}$, on a

$$\mathbb{E}\big[|X_n|\mathbf{1}_A\big]=\mathbb{E}\big[|X_n|\mathbf{1}_{A\cap\{|X_n|>c\}}\big]+\mathbb{E}\big[|X_n|\mathbf{1}_{A\cap\{|X_n|\leq c\}}\big]\leq\frac{\varepsilon}{2}+c\mathbb{P}(A).$$

On obtient alors (i) avec $\eta=\varepsilon/(2c)$ et (ii) avec $A=\Omega$.

Réciproquement, on fixe $\varepsilon>0$ et on considère $\eta>0$ donné par (i) et $M=\sup_{n\geq 0}\mathbb{E}[|X_n|]<+\infty$ par (ii). D'après l'inégalité de Markov, pour tout $c\geq M/\eta$, on a

$$\mathbb{P}(|X_n|>c)\leq\frac{\mathbb{E}[|X_n|]}{c}\leq\frac{M}{c}\leq\eta.$$

En appliquant le (i) pour chaque $n \geq 0$ avec $A = \{|X_n| > c\}$, on a $\mathbb{E}\big[|X_n|\mathbf{1}_{\{|X_n|>c\}}\big] \leq \varepsilon$.
□

Proposition 4.22 *Soit $X \in L^1(\mathcal{F})$. Alors la famille (a priori non dénombrable) de variables aléatoires $\big\{\mathbb{E}[X|\mathcal{G}] : \mathcal{G} \subset \mathcal{F}$ sous-tribu de $\mathcal{F}\big\}$ est uniformément intégrable.*

Démonstration : Pour cela, notons $Z_{\mathcal{G}} = \mathbb{E}[|X|\,|\mathcal{G}]$. Comme $\{Z_{\mathcal{G}} > c\}$ est $\mathcal{G}$-mesurable, par définition de l'espérance conditionnelle $Z_{\mathcal{G}} = \mathbb{E}[|X|\,|\mathcal{G}]$, on a :

$$\mathbb{E}[Z_{\mathcal{G}}\mathbf{1}_{\{Z_{\mathcal{G}}>c\}}] = \mathbb{E}[|X|\mathbf{1}_{\{Z_{\mathcal{G}}>c\}}]. \tag{4.18}$$

Mais par l'inégalité de Markov

$$\mathbb{P}(Z_{\mathcal{G}} > c) \leq \frac{\mathbb{E}[Z_{\mathcal{G}}]}{c} = \frac{\mathbb{E}\big[\mathbb{E}[|X|\,|\mathcal{G}]\big]}{c} = \frac{\mathbb{E}[|X|]}{c}.$$

Puis comme X est intégrable, pour tout $\varepsilon > 0$, le Lemme 4.20 donne l'existence de $\delta > 0$ tel que si $\mathbb{P}(A) < \delta$ alors $\mathbb{E}[|X|\mathbf{1}_A] < \varepsilon$. Avec $c > \mathbb{E}[|X|]/\delta$, on a $\mathbb{P}(Z_{\mathcal{G}} > c) \leq \delta$ et donc $\mathbb{E}[|X|\mathbf{1}_{\{Z_{\mathcal{G}}>c\}}] < \varepsilon$. Finalement avec l'égalité (4.18), on a $\mathbb{E}[Z_{\mathcal{G}}\mathbf{1}_{\{Z_{\mathcal{G}}>c\}}] < \varepsilon$, c'est à dire

$$\lim_{c\to+\infty} \sup_{\mathcal{G} \text{ sous-tribu de } \mathcal{F}} \mathbb{E}\big[Z_{\mathcal{G}}\mathbf{1}_{\{Z_{\mathcal{G}}>c\}}\big] = 0,$$

ce qui prouve la Proposition 4.22. □

Théorème 4.23 (Vitali) *Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires intégrables. Il y a équivalence entre*

(1) $(X_n)_{n\geq 0}$ converge dans L^1 ;

(2) $(X_n)_{n\geq 0}$ est uniformément intégrable et $(X_n)_{n\geq 0}$ converge en probabilité.

Démonstration : (1)⇒(2). D'abord, la convergence L^1 entraîne la convergence en probabilité (par l'inégalité de Markov). Elle entraîne aussi que la suite $(X_n)_{n\geq 0}$ est bornée dans L^1 (point (ii) de Prop. 4.21). Ensuite pour tout $\varepsilon > 0$, comme $X_n \xrightarrow{L^1} X$, il existe n_0 tel que pour $n \geq n_0$, on a $\mathbb{E}[|X_n - X|] < \varepsilon/2$ et donc

$$\begin{aligned} \mathbb{E}[|X_n|\mathbf{1}_A] &\leq \mathbb{E}[|X_n - X|\mathbf{1}_A] + \mathbb{E}[|X|\mathbf{1}_A] \leq \mathbb{E}[|X_n - X|] + \mathbb{E}[|X|\mathbf{1}_A] \\ &\leq \varepsilon/2 + \mathbb{E}[|X|\mathbf{1}_A]. \end{aligned}$$

On a donc

$$\sup_{n\geq n_0} \mathbb{E}[|X_n|\mathbf{1}_A] \leq \varepsilon/2 + \mathbb{E}[|X|\mathbf{1}_A].$$

Avec $A = \Omega$, on déduit $\sup_{n\geq n_0} \mathbb{E}[|X_n|] < +\infty$. Puis comme X est intégrable, par le Lemme 4.20, il existe $\delta > 0$ tel que $\mathbb{P}(A) < \delta$ implique $\mathbb{E}[|X|\mathbf{1}_A] \leq \varepsilon/2$. On a donc

$\sup_{n\geq n_0}\mathbb{E}[|X_n|\mathbf{1}_A]\leq\varepsilon$ pour un tel A. Comme la suite finie $(X_n)_{n<n_0}$ est uniformément intégrable, il existe aussi $\delta'>0$ tel que $\mathbb{P}(A)<\delta'$ implique $\mathbb{E}[|X_k|\mathbf{1}_A]\leq\varepsilon$ pour $k<n_0$. Finalement lorsque $\mathbb{P}(A)\leq\min(\delta,\delta')$, on a $\sup_{n\geq 0}\mathbb{E}[|X_n|\mathbf{1}_A]\leq\varepsilon$. La Prop. 4.21 assure alors que $(X_n)_{n\geq 0}$ est uniformément intégrable.

(2)⇒(1). Comme $(X_n)_{n\geq 0}$ converge en probabilité vers X, presque sûrement, on peut extraire une sous-suite $(n_k)_{k\geq 1}$ telle que $X_{n_k}\xrightarrow{ps}X$, $k\to+\infty$. Le lemme de Fatou, avec l'uniforme intégrabilité, garantit $X\in L^1$:

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

d'après l'uniforme intégrabilité (critère de la Prop. 4.21). Puis pour tout $\varepsilon>0$, on a

$$\begin{aligned}\mathbb{E}[|X_n-X|] &\leq \mathbb{E}\big[|X_n-X|\mathbf{1}_{\{|X_n-X|\leq\varepsilon/3\}}\big]+\mathbb{E}\big[|X_n-X|\mathbf{1}_{\{|X_n-X|>\varepsilon/3\}}\big]\\ &\leq \mathbb{E}\big[|X_n-X|\mathbf{1}_{\{|X_n-X|\leq\varepsilon/3\}}\big]+\mathbb{E}\big[|X_n|\mathbf{1}_{\{|X_n-X|>\varepsilon/3\}}\big]\\ &\qquad +\mathbb{E}\big[|X|\mathbf{1}_{\{|X_n-X|>\varepsilon/3\}}\big]\\ &\leq \varepsilon/3+\mathbb{E}\big[|X_n|\mathbf{1}_{\{|X_n-X|>\varepsilon/3\}}\big]+\mathbb{E}\big[|X|\mathbf{1}_{\{|X_n-X|>\varepsilon/3\}}\big]. \end{aligned}\tag{4.19}$$

Comme $\{X,X_1,\dots,X_n,\dots\}$ est uniformément intégrable, par le critère de la Prop. 4.21 il existe $\eta>0$ tel que pour $\mathbb{P}(A)\leq\eta$:

$$\mathbb{E}[|X_n|\mathbf{1}_A]<\varepsilon/3,\quad \mathbb{E}[|X|\mathbf{1}_A]<\varepsilon/3.$$

Puis d'après la convergence en probabilité $X_n\xrightarrow{\mathbb{P}}X$ pour n assez grand $\mathbb{P}(|X_n-X|>\varepsilon/3)\leq\eta$ si bien que

$$\mathbb{E}\big[|X_n|\mathbf{1}_{\{|X_n-X|>\varepsilon/3\}}\big]<\varepsilon/3,\quad \mathbb{E}\big[|X|\mathbf{1}_{\{|X_n-X|>\varepsilon/3\}}\big]<\varepsilon/3.$$

Finalement pour n assez grand, (4.19) assure $\mathbb{E}[|X_n-X|]<\varepsilon$, ce qui prouve le 1). □

4.4 Convergence L^1 et martingales fermées

La convergence L^1 de martingale est liée à la fermeture de martingale qu'on définit :

Définition 4.24 (Martingale fermée) *Une $(\mathcal{F}_n)$-martingale $(X_n)_{n\geq 0}$ est dite fermée par une variable aléatoire $X\in L^1(\mathcal{F})$ si $X_n=\mathbb{E}[X\,|\mathcal{F}_n]$ pour tout $n\geq 0$.*

Théorème 4.25 (Sous-martingales UI) *Soit $(X_n)_{n\geq 0}$ une sous-martingale. Alors les assertions suivantes sont équivalentes :*

(1) $(X_n)_{n\geq 0}$ est uniformément intégrable ;

(2) $(X_n)_{n\geq 0}$ converge ps et dans L^1 ;

L'énoncé s'applique aussi aux sur-martingales et aux martingales.

Démonstration : (1)⇒(2). L'uniforme intégrabilité implique $\sup_{n\geq 0} \mathbb{E}[|X_n|] < +\infty$. Le théorème de convergence presque sûre des (sous-)martingales s'applique (Corollaire 4.14) et donne la convergence presque sûre. Cette convergence implique la convergence en probabilité. Dès lors avec le théorème de Vitali (Th. 4.23), l'uniforme intégrabilité implique la convergence L^1.

(2)⇒(1). Par le théorème de Vitali (Th. 4.23), la convergence L^1 implique l'uniforme intégrabilité. □

Pour des martingales, on peut compléter le Th. 4.25 avec la représentation des martingales avec le (3) ci-dessous :

Théorème 4.26 (Martingales UI et fermées) *Pour une martingale $(X_n)_{n\geq 0}$, les assertions suivantes sont équivalentes :*

(1) $(X_n)_{n\geq 0}$ est uniformément intégrable ;

(2) $(X_n)_{n\geq 0}$ converge presque sûrement et dans L^1 ;

(3) $(X_n)_{n\geq 0}$ est une martingale fermée, ie. il existe une variable aléatoire $X \in L^1(\mathcal{F})$ telle que $X_n = \mathbb{E}[X|\mathcal{F}_n]$ $\forall n \geq 0$.

Dans ce cas, on a $\mathbb{E}[X_n] = \mathbb{E}[X]$ pour tout $n \geq 0$.

Démonstration : Il reste à montrer que **(3)** est équivalente aux deux premiers points. On a **(3)⇒(1)** par la Prop. 4.22. Puis en supposant **(2)**, c'est à dire notamment $X_n \xrightarrow{L^1} X \in L^1$, comme $(X_n)_{n\geq 0}$ est une martingale, pour $k \geq n$ et $A \in \mathcal{F}_n$ on a :

$$\mathbb{E}[X_k \mathbf{1}_A] = \mathbb{E}[X_n \mathbf{1}_A].$$

Mais $\lim_{k\to+\infty} \mathbb{E}[X_k \mathbf{1}_A] = \mathbb{E}[X \mathbf{1}_A]$ car $\big|\mathbb{E}[X_k \mathbf{1}_A] - \mathbb{E}[X \mathbf{1}_A]\big| \leq \mathbb{E}[|X_k - X|]$ et donc pour tout $A \in \mathcal{F}_n$:

$$\mathbb{E}[X_n \mathbf{1}_A] = \mathbb{E}[X \mathbf{1}_A]$$

c'est à dire $X_n = \mathbb{E}[X\,|\mathcal{F}_n]$ par définition de l'espérance conditionnelle, on a donc **(2)⇒(1)**.

Lorsque ces conditions sont remplies, il y a convergence L^1 donc convergence des espérances $\mathbb{E}[X_n]$, $n \geq 0$, vers $\mathbb{E}[X]$. Comme pour une martingale les espérances sont constantes, la conclusion s'ensuit. □

Applications des convergences ps et L^1 de martingales

On rappelle que pour une famille de tribus $(\mathcal{G}_i)_{i\in I}$, on note $\bigvee_{i\in I} \mathcal{G}_i := \sigma\big(\bigcup_{i\in I} \mathcal{G}_i\big)$. Dans la suite, on écrit $\mathcal{F}_n \nearrow \mathcal{F}_\infty$ pour signifier qu'on considère une filtration $(\mathcal{F}_n)_{n\geq 0}$ avec $\mathcal{F}_\infty := \bigvee_{n\in\mathbb{N}} \mathcal{F}_n$, la tribu limite la plus grande quand $n \to +\infty$.

Théorème 4.27 (Continuité croissante du conditionnement) *Soit $\mathcal{F}_n \nearrow \mathcal{F}_\infty$. Alors pour $X \in L^1(\mathcal{F})$, on a*

$$\mathbb{E}[X|\mathcal{F}_n] \xrightarrow{ps,L^1} \mathbb{E}[X|\mathcal{F}_\infty], \quad n \to +\infty.$$

Démonstration : En notant $X_n = \mathbb{E}[X|\mathcal{F}_n]$ alors $(X_n)_{n\geq 0}$ est une martingale fermée donc convergente ps et dans L^1 vers une variable aléatoire limite $Z \in L^1$ par le Th. 4.26. Comme X_n est $\mathcal{F}_n$- donc $\mathcal{F}_\infty$-mesurable, $Z = \lim_{n\to+\infty} X_n$ est aussi $\mathcal{F}_\infty$-mesurable.

Pour $A \in \mathcal{F}_n$, on a $\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[X_n\mathbf{1}_A]$ (propriété de martingale) et $\mathbb{E}[X_n\mathbf{1}_A] \to \mathbb{E}[Z\mathbf{1}_A]$, $n \to +\infty$, (convergence L^1). Ainsi, pour tout $A \in \mathcal{F}_n$ et donc tout $A \in \bigcup_{n\geq 0} \mathcal{F}_n$, on a :

$$\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[Z\mathbf{1}_A]. \tag{4.20}$$

Comme $(\mathcal{F}_n)_{n\geq 1}$ est une filtration, $\bigcup_{n\geq 1} \mathcal{F}_n$ est stable par intersection (finie). Puis

$$\mathcal{M} = \big\{A \in \mathcal{F} \,:\, \mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[Z\mathbf{1}_A]\big\}$$

est une classe monotone car

— $\Omega \in \mathcal{M}$ puisque $\mathbb{E}[X] = \mathbb{E}[Z]$;

— si $A, B \in \mathcal{M}$ avec $A \subset B$ alors $B \setminus A \in \mathcal{M}$: comme $\mathbf{1}_{B\setminus A} = \mathbf{1}_B - \mathbf{1}_A$,

$$\begin{aligned} \mathbb{E}[X\mathbf{1}_{B\setminus A}] &= \mathbb{E}[X(\mathbf{1}_B - \mathbf{1}_A)] = \mathbb{E}[X\mathbf{1}_B] - \mathbb{E}[X\mathbf{1}_A] \\ &= \mathbb{E}[Z\mathbf{1}_B] - \mathbb{E}[Z\mathbf{1}_A] = \mathbb{E}[Z(\mathbf{1}_B - \mathbf{1}_A)] = \mathbb{E}[Z\mathbf{1}_{B\setminus A}]; \end{aligned}$$

— si $A_n \in \mathcal{M}$ avec $A_n \subset A_{n+1}$ alors $\bigcup_{n\geq 0} A_n \in \mathcal{M}$:comme $\mathbf{1}_{\bigcup_{n\geq 0} A_n} = \lim_{n\to+\infty} \mathbf{1}_{A_n}$,

$$\begin{aligned} \mathbb{E}\big[X\mathbf{1}_{\bigcup_{n\geq 0} A_n}\big] &= \mathbb{E}\big[X \lim_{n\to+\infty} \mathbf{1}_{A_n}\big] = \lim_{n\to+\infty} \mathbb{E}[X\mathbf{1}_{A_n}] \\ &= = \lim_{n\to+\infty} \mathbb{E}[Z\mathbf{1}_{A_n}] = \mathbb{E}\big[Z \lim_{n\to+\infty} \mathbf{1}_{A_n}\big] = \mathbb{E}\big[Z\mathbf{1}_{\bigcup_{n\geq 0} A_n}\big], \end{aligned}$$

en utilisant (2 fois) la convergene dominée avec $|X\mathbf{1}_{A_n}| \leq |X| \in L^1$ et $|Z\mathbf{1}_{A_n}| \leq |Z| \in L^1$.

le théorème des classes monotones (Th. 0.2), assure alors que

$$\mathcal{F}_\infty = \bigvee_{n\geq 0} \mathcal{F}_n = \sigma\Big(\bigcup_{n\geq 0} \mathcal{F}_n\Big) \subset \mathcal{M},$$

et on a donc (4.20) pour tout $A \in \mathcal{F}_\infty$.

Mais Z est $\mathcal{F}_\infty$-mesurable car limite des X_n qui sont $\mathcal{F}_n$ donc $\mathcal{F}_\infty$-mesurables. Finalement (4.20) assure $Z = \mathbb{E}[X|\mathcal{F}_\infty]$ et on a donc $X_n \xrightarrow{ps,L^1} \mathbb{E}[X|\mathcal{F}_\infty]$ quand $n \to +\infty$. □

Le résultat suivant généralise la Prop. 2.17 et le Th. 4.27 :

Théorème 4.28 (Convergence dominée pour l'espérance conditionnelle) *On suppose $X_n \xrightarrow{ps} X$ et $|X_n| \leq Z \in L^1$ pour tout $n \geq 1$. Alors si $\mathcal{F}_n \nearrow \mathcal{F}_\infty$, on a*

$$\mathbb{E}\big[X_n\,|\mathcal{F}_n\big] \xrightarrow{ps,L^1} \mathbb{E}\big[X\,|\mathcal{F}_\infty\big], \quad n \to +\infty.$$

Démonstration : On note $W_N = \sup\big(|X_n - X_m| : n, m \geq N\big)$. Comme on observe que $W_N \leq 2Z$, on a $\mathbb{E}[W_N] < +\infty$ pour tout $N \geq 1$. Pour $n \geq N$, on a $|X_n - X| \leq W_N$ et avec le Th. 4.27, il vient presque sûrement

$$\limsup_{n\to+\infty} \mathbb{E}\big[|X_n - X|\,|\mathcal{F}_n\big] \leq \lim_{n\to+\infty} \mathbb{E}[W_N\,|\mathcal{F}_n] = \mathbb{E}[W_N\,|\mathcal{F}_\infty]. \tag{4.21}$$

Comme $W_N \searrow 0$ ps avec $W_N \leq 2Z \in L^1$, la Prop. 2.17 assure $\lim_{N\to+\infty} \mathbb{E}[W_N\,|\mathcal{F}_\infty] = 0$ ps, soit avec (4.21) : $\lim_{n\to+\infty} \mathbb{E}\big[|X_n - X|\,|\mathcal{F}_n\big] = 0$ ps.
Par l'inégalité de Jensen conditionnelle, on a :

$$\big|\mathbb{E}[X_n\,|\mathcal{F}_n] - \mathbb{E}[X\,|\mathcal{F}_n]\big| \leq \mathbb{E}\big[|X_n - X|\,\big|\mathcal{F}_n\big] \xrightarrow{ps} 0, \quad n \to +\infty.$$

Par ailleurs toujours avec le Th. 4.27, on a $\mathbb{E}[X\,|\mathcal{F}_n] \xrightarrow{ps} \mathbb{E}[X\,|\mathcal{F}_\infty]$, $n \to +\infty$. Finalement,

$$\begin{aligned}\big|\mathbb{E}[X_n\,|\mathcal{F}_n] - \mathbb{E}[X\,|\mathcal{F}_\infty]\big| &= \big|\mathbb{E}[X_n\,|\mathcal{F}_n] - \mathbb{E}[X\,|\mathcal{F}_n] + \mathbb{E}[X\,|\mathcal{F}_n] - \mathbb{E}[X\,|\mathcal{F}_\infty]\big| \\ &\leq \big|\mathbb{E}[X_n\,|\mathcal{F}_n] - \mathbb{E}[X\,|\mathcal{F}_n]\big| + \big|\mathbb{E}[X\,|\mathcal{F}_n] - \mathbb{E}[X\,|\mathcal{F}_\infty]\big| \xrightarrow{ps} 0, \quad n \to +\infty.\end{aligned}$$

Pour la convergence L^1, on a

— $\mathbb{E}[X|\mathcal{F}_n] \xrightarrow{L^1} \mathbb{E}[X|\mathcal{F}_\infty]$ par la Prop. 4.27 et

— comme par convergence dominée usuelle $X_n \xrightarrow{L^1} X$:

$$\mathbb{E}\big[\big|\mathbb{E}[X_n\,|\mathcal{F}_n] - \mathbb{E}[X\,|\mathcal{F}_n]\big|\big] \leq \mathbb{E}\big[\mathbb{E}[|X_n - X|\,|\mathcal{F}_n]\big] = \mathbb{E}[|X_n - X|] \to 0, n \to +\infty,$$

on a aussi $\mathbb{E}[X_n\,|\mathcal{F}_n] - \mathbb{E}[X\,|\mathcal{F}_n] \xrightarrow{L^1} 0$.

Il suit donc $\mathbb{E}[X_n\,|\mathcal{F}_n] \xrightarrow{L^1} \mathbb{E}[X\,|\mathcal{F}_\infty]$, $n \to +\infty$. □

Comme conséquence du Th. 4.27, on a :

Théorème 4.29 (Loi du $0/1$ de Lévy) *On suppose $\mathcal{F}_n \nearrow \mathcal{F}_\infty$. Alors pour $A \in \mathcal{F}_\infty$, on a $\mathbb{E}[\mathbf{1}_A|\mathcal{F}_n] \xrightarrow{ps,L^1} \mathbf{1}_A$ quand $n \to +\infty$.*

Démonstration : Il suffit d'appliquer le Th. 4.27 avec $X = \mathbf{1}_A$, pour $A \in \mathcal{F}_\infty$ puisque dans ce cas $\mathbb{E}[\mathbf{1}_A|\mathcal{F}_\infty] = \mathbf{1}_A$. □

La loi du $0/1$ de Kolmogorov concerne la tribu asymptotique d'une suite de variables aléatoires indépendantes. On commence par rappeler pour une suite de variables aléatoires $(X_n)_{n\geq 0}$:

— les tribus du futur : $\mathcal{F}^{(n)} = \sigma\big(X_n, X_{n+1}, \dots\big)$, $n \geq 0$;

— la tribu asymptotique $\mathcal{F}^{(\infty)} = \bigcap_{n\geq 0} \mathcal{F}^{(n)}$.

Théorème 4.30 (Loi du $0/1$ de Kolmogorov) *Lorsque les variables aléatoires $(X_n)_{n\geq 0}$ sont indépendantes, la tribu asymptotique $\mathcal{F}^{(\infty)}$ se réduit, aux négligeables près, à $\{\emptyset, \Omega\}$, ie. $\forall A \in \mathcal{F}^{(\infty)}$, $\mathbb{P}(A) \in \{0, 1\}$.*

Démonstration : On note $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$ et $\mathcal{F}_\infty = \bigvee_{n\geq 0} \mathcal{F}_n$. Soit $A \in \mathcal{F}^{(\infty)}$. Pour tout $n \geq 1$, $(X_0, \dots, X_n)$ et $\mathbf{1}_A$ sont indépendantes car $A \in \mathcal{F}^{(n+1)}$. On a donc $\mathbb{E}[\mathbf{1}_A] = \mathbb{E}\big[\mathbf{1}_A \,|\mathcal{F}_n\big]$ et avec le Th. 4.27 :

$$\mathbb{E}[\mathbf{1}_A] = \lim_{n\to+\infty} \mathbb{E}\big[\mathbf{1}_A \,|\mathcal{F}_n\big] = \mathbb{E}\big[\mathbf{1}_A \,|\mathcal{F}_\infty\big] = \mathbf{1}_A \text{ ps}$$

en utilisant $A \in \mathcal{F}^{(\infty)} \subset \mathcal{F}_\infty$. En effet, on peut écrire $\mathcal{F}^{(n)} = \sigma\big(\bigcup_{p\geq n} \mathcal{F}_n^p\big)$ où $\mathcal{F}_n^p = \sigma(X_i : n \leq i \leq p)$. Comme $\mathcal{F}_n^p \subset \mathcal{F}_p \subset \mathcal{F}_\infty$, on a $\mathcal{F}^{(n)} \subset \mathcal{F}_\infty$ pour tout $n \geq 0$ et donc $\mathcal{F}^{(\infty)} \subset \mathcal{F}_\infty$.

Finalement, on a $\mathbf{1}_A = \text{Cte}$ ps, avec nécessairement $\text{Cte} \in \{0, 1\}$, c'est à dire $\mathbb{P}(A) \in \{0, 1\}$. □

4.5 Convergence L^p de martingales pour $p > 1$

Théorème 4.31 (Convergence L^p de martingale) *Soit $(X_n)_{n\geq 0}$ une sous ou sur-martingale telle que $\sup_{n\geq 0} \mathbb{E}\big[|X_n|^p\big] < +\infty$ pour $p > 1$. Alors $X_n \xrightarrow{ps, L^1} X$ quand $n \to +\infty$. Si de plus, $(X_n)_{n\geq 0}$ est une martingale alors il y a convergence vers X ps et dans L^p, avec $\mathbb{E}[X] = \mathbb{E}[X_n]$ pour tout $n \geq 0$.*

Démonstration : Soit $(X_n)_{n\geq 0}$ une sur- ou sous-martingale. Comme $\sup_{n\geq 0} \mathbb{E}\big[|X_n|^p\big] < +\infty$, $(X_n)_{n\geq 0}$ est uniformément intégrable par la Prop. 4.19 et donc elle converge ps et L^1 par le Th. 4.25.

Si de plus, $(X_n)_{n\geq 0}$ est une martingale, par l'inégalité de moments (4.12) du Corollaire 4.6, on a

$$\mathbb{E}\Big[\sup_{0\leq k\leq n} |X_k|^p\Big] \leq \Big(\frac{p}{p-1}\Big)^p \mathbb{E}\big[|X_n|^p\big].$$

En faisant $n \to +\infty$ avec le théorème de convergence monotone, on obtient

$$\mathbb{E}\Big[\sup_{k\geq 0} |X_k|^p\Big] \leq \Big(\frac{p}{p-1}\Big)^p \sup_{n\geq 0} \mathbb{E}\big[|X_n|^p\big] < +\infty, \tag{4.22}$$

c'est à dire $Z := \sup_{n\geq 0} |X_n| \in L^p$.

Maintenant, comme $|X_n - X|^p \leq (2Z)^p \in L^1$ et toujours $X_n \xrightarrow{ps} X$, le théorème de convergence dominée implique alors

$$\lim_{n\to+\infty} \mathbb{E}\big[|X_n - X|^p\big] = 0.$$

L'égalité des espérances est due au Théorème 4.26. □

4.6 Martingales carré-intégrables

On rappelle qu'une martingale $(X_n)_{n\geq 0}$ carré-intégrable admet un compensateur $(\langle X, X\rangle_n)_{n\geq 0}$ (Déf. 3.42) et qu'elle est bornée dans L^2 si et seulement si $\mathbb{E}\big[\langle X, X\rangle_\infty\big] < +\infty$ (Prop. 3.44).

Convergence ps de martingale L^2 et compensateur

Théorème 4.32 (Convergence ps de martingale L^2 et compensateur) *Soit $(X_n)_{n\geq 0}$ une martingale de carré intégrable et de compensateur $\langle X, X\rangle$.*

(1) Pour presque tout $\omega \in \Omega$ tel que $\langle X, X\rangle_\infty(\omega) < +\infty$, $\lim_{n\to+\infty} X_n(\omega)$ existe.

(2) Si $(X_n)_{n\geq 0}$ a des accroissements bornés, ie. il existe $K < +\infty$ tel que $|X_n - X_{n-1}| \leq K$ ps, alors pour presque tout ω tel que $\lim_{n\to+\infty} X_n(\omega)$ existe dans $\mathbb{R}$, on a $\langle X, X\rangle_\infty(\omega) < +\infty$.

Démonstration : Sans perte de généralité, on suppose pour simplifier que $X_0 = 0$.
1) On note $A_n = \langle X, X\rangle_n$. Comme $(A_n)_{n\geq 0}$ est prévisible, $S_k = \inf\big(p \geq 0 : A_{p+1} > k\big)$ est un temps d'arrêt : pour tout $n \geq 0$,

$$\{S_k \leq n\} = \bigcup_{p\leq n} \underbrace{\{A_{p+1} > k\}}_{\in \mathcal{F}_p} \in \mathcal{F}_n.$$

Par la Prop. 3.34, la suite $A^{S_k} = \big(A_{S_k\wedge n}\big)_{n\geq 0}$ est prévisible et, par la Prop. 3.45, $A^{S_k} = \langle X^{S_k}, X^{S_k}\rangle$. Par la propriété de martingale, il vient :

$$\mathbb{E}\big[(X^{S_k})_n^2 - A_n^{S_k}\big] = \mathbb{E}\big[(X^{S_k})_0^2 - A_0^{S_k}\big] = 0,$$

soit par définition de l'arrêt S_k :

$$\mathbb{E}\big[(X^{S_k})_n^2\big] = \mathbb{E}\big[A_n^{S_k}\big] \leq k$$

pour tout $n \geq 0$ et donc X^{S_k} est bornée dans L^2. Par l'inégalité de Cauchy-Schwarz, X^{S_k} est aussi bornée dans L^1 et le Corollaire 4.14 assure la convergence presque sûre.

Comme $\{A_\infty < +\infty\} = \bigcup_{k\geq 0}\{S_k = +\infty\}$, si $\omega \in \{A_\infty < +\infty\}$ alors il existe $k(\omega)$ tel que $S_{k(\omega)}(\omega) = +\infty$ et $X_n^{S_{k(\omega)}(\omega)}(\omega) = X_n(\omega)$. Pour presque chaque tels $\omega \in \{A_\infty < +\infty\}$, $(X_n(\omega))_{n\geq 0}$ converge.

2) On raisonne par l'absurde et on suppose que

$$\mathbb{P}\Big(A_\infty = +\infty \text{ et } \sup_{n\geq 0} |X_n| < +\infty\Big) > 0. \tag{4.23}$$

En notant $T_c = \inf(n \geq 0 : |X_n| > c)$, on a $\{T_c = +\infty\} = \{\sup_{n\geq 0} |X_n| \leq c\}$ et par convergence monotone

$$\mathbb{P}\big(A_\infty = +\infty \text{ et } T_c = +\infty\big)$$

$$= \mathbb{P}\big(A_\infty = +\infty, \sup_{n\geq 0}|X_n| \leq c\big) \nearrow \mathbb{P}\big(A_\infty = +\infty, \sup_{n\geq 0}|X_n| < +\infty\big).$$

Pour $c > 0$ assez grand, on déduit donc de (4.23) que

$$\mathbb{P}\big(A_\infty = +\infty \text{ et } T_c = +\infty\big) > 0. \tag{4.24}$$

D'après le théorème d'arrêt borné (Th. 3.37) avec le temps d'arrêt borné $T_c \wedge n$ et la martingale $X^2 - A$, on a

$$\mathbb{E}[X^2_{T_c\wedge n}] - \mathbb{E}[A_{T_c\wedge n}] = 0.$$

Mais $X^2_{T_c\wedge n}$ est bornée par $(c+K)^2$: en effet, à la date $(T_c\wedge n)-1 < T_c$, on a $X_{(T_c\wedge n)-1} \leq c$ et comme un accroissement (de la date $(T_c \wedge n) - 1$ à la date $(T_c \wedge n)$) est borné par K, on a

$$\mathbb{E}[A_{T_c\wedge n}] = \mathbb{E}[X^2_{T_c\wedge n}] \leq (c+K)^2.$$

En faisant $n \to +\infty$ par le théorème de convergence monotone, on a

$$\mathbb{E}[A_{T_c}] \leq (c+K)^2,$$

ce qui est en contradiction avec (4.24) et donc (4.23) car en notant $B = \{A_\infty = +\infty \text{ et } T_c = +\infty\}$, on a

$$\mathbb{E}[A_{T_c}] \geq \mathbb{E}[A_{T_c}\mathbf{1}_B] = \mathbb{E}[A_\infty \mathbf{1}_B] = +\infty$$

puisque $\mathbb{P}(B) > 0$ et $A_\infty = +\infty$ sur B.

Finalement, on doit avoir $A_\infty < +\infty$ dès que $\sup_{n\geq 0}|X_n| < +\infty$, en particulier dès que $(X_n)_{n\geq 1}$ converge dans $\mathbb{R}$. □

Martingale bornée dans L^2

Dans le cas de martingale L^2, le Théorème 4.31 s'écrit

Corollaire 4.33 (Convergence de martingale bornée dans L^2) *Soit $(X_n)_{n\geq 0}$ une martingale bornée dans L^2 (ou de façon équivalente, de compensateur $\langle X, X\rangle$ intégrable en $+\infty$). Alors $X_n \xrightarrow{ps} X$ ps et dans L^2. De plus, $\mathbb{E}[X_0] = \mathbb{E}[X_n] = \mathbb{E}[X]$.*

Le résultat suivant est une LGN presque sûre pour les martingales. Il complète le Th. 4.32.

Théorème 4.34 (LGN pour martingale) *Soit $(X_n)_{n\geq 0}$ une $(\mathcal{F}_n)$-martingale de carré intégrable avec $X_0 = 0$. Alors sur $\{\langle X, X\rangle_\infty = +\infty\}$, on a*

$$\frac{X_n}{\langle X, X\rangle_n} \xrightarrow{ps} 0, \quad n \to +\infty. \tag{4.25}$$

Remarque 4.35 (Cas *iid*) Lorsque $(X_n)_{n\geq 1}$ est une suite de variables aléatoires *iid* centrées de carrés intégrables, on retrouve la LGN forte L^2 : $S_n = \sum_{k=1}^n X_k$ est une martingale L^2 de compensateur $\langle S, S\rangle_n = n\sigma^2$ (où $\sigma^2 = \mathbb{E}[X_1^2]$, voir Exemple 3.46) et (4.25) se réécrit

$$\frac{S_n}{n} \xrightarrow{ps} 0, \quad n \to +\infty.$$

Démonstration : On considère la suite prévisible $H = \big((1+\langle X, X\rangle_n)^{-1}\big)_{n\geq 0}$, et $W = H\cdot X$ donnée par $W_0 = 0$ et

$$W_n = \sum_{k=1}^n \frac{X_k - X_{k-1}}{1+\langle X, X\rangle_k}, \quad n \geq 1.$$

Comme H est bornée par 1, la Prop. 3.32 assure que $W = (W_n)_{n\geq 0}$ définit une martingale. En utilisant l'expression (3.16) du compensateur, celui de W est donné par

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

On en déduit

$$\langle W, W\rangle_n = \sum_{k=1}^n \big(\langle W, W\rangle_k - \langle W, W\rangle_{k-1}\big) \leq 1 - \frac{1}{1+\langle X, X\rangle_n} \leq 1 \quad \text{ps.}$$

En passant à la limite, on a $\langle W, W\rangle_\infty \leq 1$ ps et par le Th. 4.32 la martingale W converge ps. Par le lemme de Kronecker (Lemme 4.37) avec $a_k = 1 + \langle X, X\rangle_k$, il suit

$$\frac{X_n}{1+\langle X, X\rangle_n} \xrightarrow{ps} 0, \quad n \to +\infty,$$

ou de façon équivalente

$$\frac{X_n}{\langle X, X\rangle_n} \xrightarrow{ps} 0, \quad n \to +\infty.$$

□

On prouve ci-dessous les lemmes de Césaro (Lemme 4.36) et de Kronecker (Lemme 4.37) pour la convergence de séries numériques.

Lemme 4.36 (Césaro) *Soit $(u_n)_{n\geq 1}$ une suite d'un espace vectoriel normé E qui converge vers ℓ et $(\alpha_n)_{n\geq 1}$ une suite positive, de sommes partielles $a_n = \sum_{k=1}^n \alpha_n$ qui tendent vers $+\infty$. Alors on a*

$$\lim_{n\to+\infty} \frac{1}{a_n}\sum_{k=1}^n \alpha_k u_k = \ell.$$

Démonstration : Soit $\varepsilon > 0$ et $n_0 \geq 1$ tel que, pour $n \geq n_0$, on a $|u_n - \ell| < \varepsilon$. Comme $\sum_{k=1}^n \alpha_k = a_n$, on a

$$\begin{aligned}
\left\|\frac{1}{a_n}\Big(\sum_{k=1}^n \alpha_k u_k\Big) - \ell\right\| &= \left\|\frac{1}{a_n}\sum_{k=1}^n \alpha_k (u_k - \ell)\right\| \\
&\leq \frac{1}{a_n}\sum_{k=1}^{n_0} \alpha_k \|u_k - \ell\| + \frac{1}{a_n}\sum_{k=n_0+1}^{n} \alpha_k \|u_k - \ell\| \\
&\leq \frac{1}{a_n}\sum_{k=1}^{n_0} \alpha_k \|u_k - \ell\| + \frac{a_n - a_{n_0}}{a_n}\varepsilon.
\end{aligned}$$

Comme $\frac{a_n - a_{n_0}}{a_n} \leq 1$ et $\lim_{n\to+\infty} \frac{1}{a_n}\sum_{k=1}^{n_0} \alpha_n \|u_k - \ell\| = 0$ ($a_n \to +\infty$), on a

$$\limsup_{n\to+\infty} \left\|\frac{1}{a_n}\sum_{k=1}^n \alpha_n u_k - \ell\right\| \leq \varepsilon,$$

ce qui permet de conclure puisque $\varepsilon > 0$ est arbitraire :

$$0 \leq \liminf_{n\to+\infty} \left\|\frac{1}{a_n}\sum_{k=1}^n \alpha_n u_k - \ell\right\| \leq \limsup_{n\to+\infty} \left\|\frac{1}{a_n}\sum_{k=1}^n \alpha_n u_k - \ell\right\| = 0,$$

soit

$$\lim_{n\to+\infty} \left\|\frac{1}{a_n}\sum_{k=1}^n \alpha_n u_k - \ell\right\| = 0.$$

□

Lemme 4.37 (Kronecker) *Soit $(x_n)_{n\geq 1}$ une suite dans un espace vectoriel normé et $(a_n)_{n\geq 1}$ une suite croissante strictement positive convergeant vers $+\infty$. Si la série $\sum_{n=1}^{+\infty} \frac{x_n}{a_n}$ converge alors*

$$\lim_{n\to+\infty} \frac{1}{a_n}\sum_{k=1}^n x_k = 0.$$

Démonstration : On note $S_k = \sum_{n=1}^k \frac{x_n}{a_n}$. En écrivant $x_k/a_k = S_k - S_{k-1}$, on a par une transformation d'Abel :

$$\sum_{k=1}^n x_k = \sum_{k=1}^n a_k(S_k - S_{k-1}) = \sum_{k=1}^{n-1} (a_k - a_{k+1})S_k + a_n S_n,$$

avec $S_0 = 0$ par convention et donc

$$\frac{1}{a_n}\sum_{k=1}^{n} x_k = S_n - \sum_{k=1}^{n-1} \frac{a_{k+1}-a_k}{a_n} S_k. \tag{4.26}$$

Comme S_k converge vers S, le Lemme 4.36 (Césaro) avec $\alpha_k = a_k - a_{k-1} \geq 0$ de somme partielle $a_n \to +\infty$ assure

$$\sum_{k=1}^{n-1} \frac{a_{k+1}-a_k}{a_n} S_k \to S, \quad n \to +\infty,$$

ce qui conclut le Lemme 4.37 (Kronecker) quand on passe à la limite dans (4.26). □

4.7 Théorème d'arrêt

Dans cette section, on cherche à généraliser la propriété de martingale (3.3) à des temps d'arrêt. On rappelle que si T, S sont des temps d'arrêt avec $S \leq T$, alors pour une sous-martingale $(X_n)_{n\geq 0}$ on n'a pas toujours $\mathbb{E}[X_S] \leq \mathbb{E}[X_T]$.

On rappelle le contre-exemple (Exemple 3.38) de la marche aléatoire simple $(S_n)_{n\geq 0}$ partant de $S_0 = 0$ avec $T = \inf\big(n \geq 1 : S_n = -1\big)$ (temps d'atteinte de -1) : on a $\mathbb{E}[S_T] = -1$ et $\mathbb{E}[S_0] = 0$ alors que $T \geq S = 0$.

Au contraire si $S \leq T$ avec T borné (ie. $\mathbb{P}(T \leq k) = 1$ pour un $k \in \mathbb{R}_+$) alors pour une sous-martingale $(X_n)_{n\geq 1}$, on a $\mathbb{E}[X_S] \leq \mathbb{E}[X_T]$. En effet, $X_n^T = X_{T\wedge n}$, $n \geq 0$, est une sous-martingale arrêtée donc une sous-martingale (Prop. 3.36). Le Th. 3.37 appliqué à X^T avec le temps d'arrêt S borné donne

$$\mathbb{E}[X_S] = \mathbb{E}[X_{T\wedge S}] = \mathbb{E}\big[X_S^T\big] \leq \mathbb{E}\big[X_k^T\big] = \mathbb{E}[X_{T\wedge k}] = \mathbb{E}[X_T].$$

En fait, on a un résultat plus général :

Proposition 4.38 *Soit $(X_n)_{n\geq 0}$ une sous-martingale uniformément intégrable et T un temps d'arrêt. Alors :*

(1) $X^T = (X_{T\wedge n})_{n\geq 0}$ est une sous-martingale uniformément intégrable

(2) $X_{T\wedge n} \xrightarrow{ps, L^1} X_T$, $n \to +\infty$

(3) $\mathbb{E}[X_0] \leq \mathbb{E}[X_T] \leq \mathbb{E}[X_\infty]$ où $X_\infty = \lim_{n\to+\infty} X_n$.

Démonstration : D'abord, comme $(X_n)_{n\geq 0}$ est une sous-martingale uniformément intégrable, le Th. 4.25 assure que $(X_n)_{n\geq 0}$ converge ps (et dans L^1) vers X_∞, X_T est bien défini même pour $T = +\infty$.

Ensuite, on sait déjà que $X^T = (X_{T\wedge n})_{n\geq 0}$ est une martingale par la Prop. 3.36. On montre d'abord qu'elle converge ps vers $X_T \in L^1$ en appliquant le Th. 4.10 en vérifiant

$$\sup_{n\geq 0} \mathbb{E}\big[X_{T\wedge n}^+\big] < +\infty \tag{4.27}$$

En effet, comme $\varphi(x) = x^+$ est croissante et convexe, $(X_n^+)_{n\geq 0}$ est aussi une sous-martingale (Prop. 3.29). En appliquant le Th. 3.37 à la sous-martingale $(X_n^+)_{n\geq 0}$ et au temps d'arrêt $T \wedge n$ borné, on a

$$\mathbb{E}\big[X_{T\wedge n}^+\big] \leq \mathbb{E}\big[X_n^+\big] \leq \mathbb{E}\big[|X_n|\big].$$

Comme $(X_n)_{n\geq 0}$ est uniformément intégrable, on a

$$\sup_{n\geq 0} \mathbb{E}\big[X_{T\wedge n}^+\big] \leq \sup_{n\geq 0} \mathbb{E}\big[|X_n|\big] < +\infty,$$

ce qui assure (4.27) et donc la convergence ps de $X_{T\wedge n}$ vers $X_T \in L^1$.

Ensuite, on établit que X^T est uniformément intégrable : pour tout $c > 0$, on a :

$$\begin{aligned} &\mathbb{E}\big[|X_{T\wedge n}|\mathbf{1}_{\{|X_{T\wedge n}|\geq c\}}\big] \\ &\quad = \mathbb{E}\big[|X_T|\mathbf{1}_{\{|X_T\geq c\}}\mathbf{1}_{\{T\leq n\}}\big] + \mathbb{E}\big[|X_n|\mathbf{1}_{\{|X_n|\geq c\}}\mathbf{1}_{\{T>n\}}\big] \\ &\quad \leq \mathbb{E}\big[|X_T|\mathbf{1}_{\{|X_T|\geq c\}}\big] + \mathbb{E}\big[|X_n|\mathbf{1}_{\{|X_n|\geq c\}}\big]. \end{aligned} \tag{4.28}$$

Comme $\big\{X_n : n \in \mathbb{N}\big\} \cup \{X_T\}$ est uniformément intégrable, la famille $\big\{X_{T\wedge n} : n \in \mathbb{N}\big\}$ l'est aussi, ce qui prouve 1).

Finalement, le Th. 4.25 s'applique à la sous-martingale X^T et prouve la convergence L^1 vers X_T, ce qui achève de prouver 2).

Le Th. 3.37 pour les temps d'arrêt bornés donne

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_{T\wedge n}] \leq \mathbb{E}[X_n].$$

Par 1), $\lim_{n\to+\infty} \mathbb{E}[X_{T\wedge n}] = \mathbb{E}[X_T]$ et par le le Th. 4.25 $\lim_{n\to+\infty} \mathbb{E}[X_n] = \mathbb{E}[X_\infty]$. On obtient donc 3) en faisant $n \to +\infty$. □

De (4.28) dans la preuve précédente, on déduit en particulier :

Corollaire 4.39 *Soit $(X_n)_{n\geq 0}$ une sous-martingale. Si $\mathbb{E}[|X_T|] < +\infty$ et $(X_n\mathbf{1}_{\{T>n\}})_{n\geq 0}$ est uniformément intégrable. Alors $X^T = (X_{T\wedge n})_{n\geq 0}$ est uniformément intégrable.*

On arrive à la forme générale du théorème d'arrêt :

Théorème 4.40 (Arrêt de Doob) *Soit $S \leq T$ des temps d'arrêt et $(X_n)_{n\geq 0}$ une sous-martingale uniformément intégrable. Alors $\mathbb{E}[X_S] \leq \mathbb{E}[X_T]$ et*

$$X_S \leq \mathbb{E}[X_T\,|\mathcal{F}_S]. \tag{4.29}$$

On a des énoncés analogues pour les martingales et sur-martingales.

Remarque 4.41 Si $T \leq k$ ps alors

$$\mathbb{E}\big[|X_T|\big] = \sum_{n=1}^{k} \mathbb{E}\big[|X_n| \mathbf{1}_{\{T=n\}}\big] \leq k\mathbb{E}[|X_k|] < +\infty$$

et $X_n \mathbf{1}_{\{T>n\}}$ est uniformément intégrable puisque la famille est finie (pour $n > k$, $X_n \mathbf{1}_{\{T>n\}} = 0$). D'après le Corollaire 4.39, $X^T = (X_{T\wedge n})_{n\geq 0}$ est uniformément intégrable et le théorème d'arrêt (Th. 4.40) s'applique à $X^T = (X_{T\wedge n})_{n\geq 0}$ quand T est borné : on retrouve bien l'exemple introductif de la section dans ce théorème général.

Démonstration :[Th. 4.40] On pose $Y_n = X_{T\wedge n}$, $n \geq 0$. D'après la Prop. 4.38, $(Y_n)_{n\geq 0}$ est une sous-martingale uniformément intégrable et on peut lui appliquer le 3) de la Prop. 4.38 avec le temps d'arrêt S pour avoir $\mathbb{E}[Y_S] \leq \mathbb{E}[Y_\infty]$, soit comme $S \leq T$, $Y_S = X_S$ et $Y_\infty = X_T$:

$$\mathbb{E}[X_S] \leq \mathbb{E}[X_T],$$

ce qui prouve la première partie du Th. 4.40.

On observe que X_S est $\mathcal{F}_S$-mesurable puisque pour tout $B \in \mathcal{B}(\mathbb{R})$, on a $X_S^{-1}(B) = \{X_S \in B\} \in \mathcal{F}_S$: $\{X_S \in B\} \cap \{S = n\} = \{X_n \in B\} \cap \{S = n\} \in \mathcal{F}_n$.

Pour la deuxième partie (4.29), on considère $A \in \mathcal{F}_S$ et on pose

$$U = S\mathbf{1}_A + T\mathbf{1}_{A^c}.$$

On observe que U est un temps d'arrêt car

$$\begin{aligned}\{U \leq n\} &= \big(\{U \leq n\} \cap A\big) \cup \big(\{U \leq n\} \cap A^c\big)\\ &= \big(\{S \leq n\} \cap A\big) \cup \big(\{T \leq n\} \cap A^c\big).\end{aligned}$$

Mais $\{S \leq n\} \in \mathcal{F}_n$ et donc, comme $A \in \mathcal{F}_S$, $\{S \leq n\} \cap A \in \mathcal{F}_n$. Puis $\{T \leq n\} \in \mathcal{F}_n$ et donc, comme $A \in \mathcal{F}_S \subset \mathcal{F}_T$, $\{T \leq n\} \cap A^c \in \mathcal{F}_n$. On a donc bien $\{U \leq n\} \in \mathcal{F}_n$, justifiant que U est temps d'arrêt.

Comme $U \leq T$, par la première partie de la preuve, il vient

$$\mathbb{E}[X_U] \leq \mathbb{E}[X_T]. \tag{4.30}$$

Comme $U = S$ sur A et $U = T$ sur A^c, on a

$$\begin{aligned}\mathbb{E}[X_U] &= \mathbb{E}\big[X_U\mathbf{1}_A\big] + \mathbb{E}\big[X_U\mathbf{1}_{A^c}\big]\\ &= \mathbb{E}\big[X_S\mathbf{1}_A\big] + \mathbb{E}\big[X_T\mathbf{1}_{A^c}\big] \qquad (4.31)\\ &\leq \mathbb{E}[X_T] = \mathbb{E}\big[X_T\mathbf{1}_A\big] + \mathbb{E}\big[X_T\mathbf{1}_{A^c}\big] \qquad (4.32)\end{aligned}$$

où (4.32) vient de (4.30). En simplifiant (4.31)≤(4.32), il vient pour tout $A \in \mathcal{F}_S$

$$\mathbb{E}\big[X_S\mathbf{1}_A\big] \leq \mathbb{E}\big[X_T\mathbf{1}_A\big] = \mathbb{E}\big[Z\mathbf{1}_A\big],$$

en notant $Z = \mathbb{E}[X_T|\mathcal{F}_S]$. On a donc $\mathbb{E}[(Z - X_S)\mathbf{1}_A] \geq 0$ pour tout $A \in \mathcal{F}_S$. Comme $A = \{Z - X_S < 0\} \in \mathcal{F}_S$ (Z, X_S sont $\mathcal{F}_S$-mesurables) on a $(Z - X_S)\mathbf{1}_A \leq 0$ et donc $\mathbb{E}\big[(Z - X_S)\mathbf{1}_A\big] = 0$ ce qui exige $(Z - X_S)\mathbf{1}_A = 0$ ps et comme $Z - X_S < 0$ sur A, il vient $\mathbf{1}_A = 0$ ps, c'est à dire $\mathbb{P}(A) = 0$ et donc $X_S \leq Z$ ps, achevant la preuve du théorème d'arrêt (Th. 4.40). □

On a un résultat analogue sans l'hypothèse d'uniforme intégrabilité pour les sur-martingales qui sont positives :

Proposition 4.42 *Soit $(X_n)_{n\geq 1}$ une sur-martingale* **positive** *et T un temps d'arrêt. Alors $\mathbb{E}[X_T] \leq \mathbb{E}[X_0] + \mathbb{E}[X_\infty]$ où la limite $X_\infty = \lim_{n\to+\infty} X_n$ existe par le Corollaire 4.15.*

Démonstration : Par le Corollaire 4.15, la sur-martingale $(X_n)_{n\geq 0}$ converge ps vers X_∞. Comme $X^T = (X_{T\wedge n})_{n\geq 0}$ est aussi une sur-martingale (Prop. 3.36), son espérance décroît :

$$\mathbb{E}[X_{T\wedge n}] \leq \mathbb{E}[X_{T\wedge 0}] = \mathbb{E}[X_0]. \tag{4.33}$$

Comme la sur-martingale est positive, par convergence monotone, on a

$$\begin{aligned} \mathbb{E}\big[X_T\mathbf{1}_{\{T<+\infty\}}\big] &= \lim_{n\to+\infty} \mathbb{E}\big[X_T\mathbf{1}_{\{T\leq n\}}\big] = \lim_{n\to+\infty} \mathbb{E}\big[X_{T\wedge n}\mathbf{1}_{\{T\leq n\}}\big] \\ &\leq \lim_{n\to+\infty} \mathbb{E}\big[X_{T\wedge n}\big] \leq \mathbb{E}[X_0] \end{aligned} \tag{4.34}$$

en utilisant (4.33). On a aussi

$$\mathbb{E}\big[X_T\mathbf{1}_{\{T=+\infty\}}\big] = \mathbb{E}\big[X_\infty\mathbf{1}_{\{T=+\infty\}}\big] \leq \mathbb{E}[X_\infty]. \tag{4.35}$$

Ainsi, en combinant (4.34) et (4.35), on a

$$\begin{aligned} \mathbb{E}[X_T] &= \mathbb{E}\big[X_T\mathbf{1}_{\{T<+\infty\}}\big] + \mathbb{E}\big[X_T\mathbf{1}_{\{T=+\infty\}}\big] \\ &\leq \mathbb{E}[X_0] + \mathbb{E}[X_\infty]. \end{aligned}$$

□

Troisième partie

Chaînes de Markov

Chapitre 5

Dynamique markovienne

Introduction

On considère un système qui peut être dans un nombre fini ou infini dénombrable d'états. L'ensemble des états, noté E, est appelé **espace d'états** et on supposera dans ce cours que E est dénombrable (souvent, E sera $\mathbb{N}$ ou une partie de $\mathbb{N}$). On suppose le système observé en des temps discrets $n = 0, 1, 2, \dots$ et l'état du système à la date n est noté X_n.

Comme on s'intéresse aux systèmes non déterministes, on considère des suites de variables aléatoires $(X_n)_{n\geq 0}$. Pour étudier de tels systèmes aléatoires $(X_n)_{n\geq 0}$, on suppose que le système –ou son évolution– satisfait certaines propriétés.

La propriété la plus simple est de supposer que les variables aléatoires X_n, $n \geq 0$, sont (mutuellement) indépendantes, c'est à dire de supposer que les états du système sont tous indépendants. En pratique, une telle hypothèse est trop restrictive pour modéliser nombre de phénomènes intéressants.

En fait, de nombreux systèmes ont la propriété –plus générale– suivante : l'état présent du système étant connu, les états passés n'ont pas d'influence sur les états futurs. Autrement dit le système n'évolue pas indépendamment dans le temps mais évolue sans mémoire (seul le présent, et non le passé, influe sur le futur). Cette propriété est dite **propriété de Markov** et les systèmes qui la vérifient sont des **chaînes de Markov** (Définition 5.5).

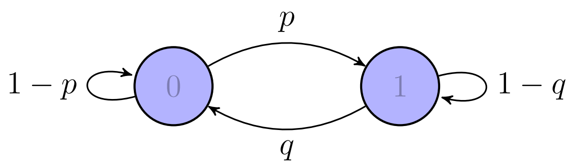
Pour formaliser ce type de propriété, on commence par un exemple très simple de système markovien ne pouvant prendre que deux valeurs.

Exemple 5.1 (Chaînes de Markov à deux états) Considérons une machine qui au début de chaque jour est soit en état de fonctionnement (état 1) soit en panne (état 0). On note alors $X_n = 1$ si la machine fonctionne au début du n-ème jour, $X_n = 0$ sinon. On suppose que si la machine est en panne le n-ème jour, la probabilité qu'elle soit réparée et fonctionne au début du $(n+1)$-ème jour est $p \in]0, 1[$. On suppose aussi que si la machine fonctionne le n-ème jour, la probabilité qu'un problème survienne et qu'elle soit en panne au début du $(n+1)$-ème jour est $q \in]0, 1[$. (Attention, avec ces notations, il n'y a pas de raison que $p+q = 1$ comme c'est souvent le cas.) Enfin, on décrit par $\mu_0 = (\mu_0(0), \mu_0(1))$

l'état initial de la machine, ie. $\mu_0(0)$ est la probabilité que la machine soit en panne au début du 0-ème jour et $\mu_0(1)$ la probabilité qu'elle fonctionne (ou encore : $X_0 \sim \mu_0$). Le modèle se réécrit

$$\mathbb{P}(X_{n+1} = 1 | X_n = 0) = p, \qquad \mathbb{P}(X_{n+1} = 0 | X_n = 1) = q$$

avec $\mathbb{P}(X_0 = 0) = \mu_0(0)$ et il se représente par le graphe de transitions suivant :



Comme les probabilités conditionnelles sont des probabilités, on en déduit immédiatement

$$\mathbb{P}(X_{n+1} = 0 | X_n = 0) = 1 - p, \qquad \mathbb{P}(X_{n+1} = 1 | X_n = 1) = 1 - q$$

et la probabilité d'être en fonction à la date 0 est $\mu_0(1) = 1 - \mu_0(0)$.

On calcule $\mathbb{P}(X_n = 0)$ et $\mathbb{P}(X_n = 1)$ par la formule des probabilités totales (1.3) :

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

Lorsque $p + q \neq 0$, comme $\mathbb{P}(X_0 = 0) = \mu_0(0)$, on en déduit par récurrence :

$$\begin{aligned}
\mathbb{P}(X_n = 0) &= (1 - p - q)^n \mu_0(0) + q \sum_{j=0}^{n-1} (1 - p - q)^j \\
&= \frac{q}{p + q} + (1 - p - q)^n \Big(\mu_0(0) - \frac{q}{p + q} \Big). \qquad (5.1)
\end{aligned}$$

Comme $\mathbb{P}(X_n = 1) = 1 - \mathbb{P}(X_n = 0)$, on a aussi

$$\mathbb{P}(X_n = 1) = \frac{p}{p + q} + (1 - p - q)^n \Big(\mu_0(1) - \frac{p}{p + q} \Big). \qquad (5.2)$$

Si $|1 - p - q| < 1$, en passant à la limite $n \to +\infty$, on a

$$\lim_{n \to +\infty} \mathbb{P}(X_n = 0) = \frac{q}{p + q}, \qquad \lim_{n \to +\infty} \mathbb{P}(X_n = 1) = \frac{p}{p + q}.$$

Ces probabilités s'interprètent comme une sorte de régime asymptotique :

$$X_n \overset{n \to +\infty}{\Longrightarrow} \mu_\infty := \Big(\frac{q}{p + q}, \frac{p}{p + q} \Big).$$

Si $p = q = 1$, alors il est facile de voir que le système est périodique de période 2 avec

$$\mathbb{P}(X_{2n} = 0) = \mu_0(0), \quad \mathbb{P}(X_{2n} = 0) = \mu_0(1);$$
$$\mathbb{P}(X_{2n+1} = 0) = \mu_0(1), \quad \mathbb{P}(X_{2n+1} = 0) = \mu_0(0)$$

ie. $X_{2n} \sim \mu_0$ et $X_{2n+1} \sim 1 - \mu_0$. Dans ce cas, $(X_n)_{n\geq 0}$ ne converge pas en loi.

Enfin, si $p = q = 0$ et on a facilement $\mathbb{P}(X_n = 0) = \mu_0(0)$ et $\mathbb{P}(X_n = 1) = \mu_0(1)$ et le système n'évolue donc pas, la loi reste donnée en tout temps par μ_0 et donc à la limite aussi : $X_n \sim \mu_0$ pour tout $n \geq 0$.

Observons qu'on peut aussi obtenir les probabilités limites $p/(p+q)$ et $q/(p+q)$ autrement : si on choisit $\mu_0(0)$ et $\mu_0(1)$ de façon que $\mathbb{P}(X_n = 0)$ et $\mathbb{P}(X_n = 1)$ ne dépendent pas de n alors on constate dans les expressions de $\mathbb{P}(X_n = 0)$ en (5.1) et $\mathbb{P}(X_n = 1)$ en (5.2) que nécessairement

$$\mu_0(0) = \frac{q}{p+q}, \quad \mu_0(1) = \frac{p}{p+q} \tag{5.3}$$

et dans ce cas si la chaîne $(X_n)_{n\geq 0}$ démarre avec la distribution donnée par (5.3), on a pour tout $n \geq 0$:

$$\mathbb{P}(X_n = 0) = \frac{q}{p+q}, \quad \mathbb{P}(X_n = 1) = \frac{p}{p+q}.$$

La distribution (5.3) s'interprète comme une distribution stationnaire (ou invariante) et on constate donc que les distributions limite et stationnaire coïncident.

Approche matricielle

Le modèle se représente aussi matriciellement en notant $\mu_n = \big(\mathbb{P}(X_n = 0), \mathbb{P}(X_n = 1)\big)$ et

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}.$$

La formule des probabilités totales (1.3) s'écrit

$$\mu_{n+1} = \mu_n P, \quad \text{et par récurrence} \quad \mu_n = \mu_0 P^n.$$

Pour calculer P^n, on observe que le spectre de P est $\mathrm{Sp}(P) = \{1, 1-p-q\}$ avec les espaces propres $\mathbb{R}^2 = \mathrm{Vect}\big((1,1)^t\big) \otimes \mathrm{Vect}\big((p,-q)^t\big)$. On en déduit la matrice de passage A et son inverse

$$A = \begin{pmatrix} 1 & p \\ 1 & -q \end{pmatrix} \qquad \text{et} \qquad A^{-1} = \frac{1}{p+q}\begin{pmatrix} q & p \\ 1 & -1 \end{pmatrix}.$$

On a donc

$$P = A \begin{pmatrix} 1 & 0 \\ 0 & 1-p-q \end{pmatrix} A^{-1}$$

et

$$P^n = A \begin{pmatrix} 1 & 0 \\ 0 & (1-p-q)^n \end{pmatrix} A^{-1}$$

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

Enfin puisque $\mu_n = \mu_0 P^n$, il vient

$$\mu_n = \left(\frac{q}{p+q} + (1-p-q)^n\Big(\mu_0(0) - \frac{q}{p+q}\Big), \frac{p}{p+q} + (1-p-q)^n\Big(\mu_0(1) - \frac{p}{p+q}\Big)\right)$$

ce qui retrouve (5.1) et (5.2).

5.1 Probabilités de transition

On rappelle la notion de noyau de probabilité aussi appelé noyau de transition ou noyau markovien (cf. Définition 2.29) :

Définition 5.2 (Noyau de transition) *Étant donné deux espaces mesurables $(E, \mathcal{E})$ et $(F, \mathcal{F})$, on appelle noyau de transition (ou noyau de probabilité ou noyau markovien) de E dans F toute application $\nu : E \times \mathcal{F} \to [0,1]$ qui vérifie*

(i) $\forall x \in E$, $\nu(x, \cdot)$ est une probabilité sur $(F, \mathcal{F})$;

(ii) $\forall A \in \mathcal{F}$, $\nu(\cdot, A)$ est $\mathcal{E}$-mesurable.

Dans la suite, on considérera $E = F$ et on supposera l'ensemble E au plus dénombrable avec $\mathcal{E} = \mathcal{F} = \mathcal{P}(E)$ si bien que le point (ii) sera automatique.

Définition 5.3 (Matrice stochastique) *On appelle matrice stochastique sur E (éventuellement infinie) toute famille $(P(x,y))_{x,y\in E}$ de réels tels que*

(i) $\forall (x,y) \in E^2$, $0 \le P(x,y) \le 1$

(ii) $\forall x \in E$, $\sum_{y\in E} P(x,y) = 1$.

Remarque 5.4

— Dire qu'une matrice P (à coefficients positifs) est stochastique, c'est dire que $(1, \dots, 1)^t$ est vecteur propre de P associé à la valeur propre 1.

— Si l'espace E est fini de cardinal d, on peut supposer sans perte de généralité que $E = \{1, \dots, d\}$ et alors $(P(x,y))_{x,y\in E} = (P_{x,y})_{x,y\in E}$ est une « vraie » matrice de taille $d \times d$, à coefficients positifs dont les sommes des lignes valent toutes 1.

— Dans le cas où l'ensemble E est dénombrable les notions de noyau de transition et de matrice stochastique sont équivalentes. Ainsi :

(i) Si P est une matrice stochastique, $\nu(x, A) = \sum_{y\in A} P(x,y)$ définit un noyau de transition ;

(ii) Si ν est un noyau de transition, $P(x,y) = \nu\big(x, \{y\}\big)$ définit une matrice stochastique.

En particulier, noter que pour chaque $x \in E$: $P(x, \bullet)$ est une loi de probabilité.

Quelques notations :

— Si $f : E \to \mathbb{R}_+$, on note Pf la fonction de E dans $\mathbb{R}_+$ donnée par

$$Pf(x) = \sum_{y\in E} P(x, y)f(y), \quad x \in E.$$

En interprétant les fonctions (donc ici f et Pf) comme des vecteurs colonnes, le vecteur colonne Pf s'obtient par le produit matriciel (à droite) et $Pf(x) = \mathbb{E}_{P(x,\bullet)}[f]$.

— En interprétant une mesure μ sur E comme un vecteur (ligne), on définit la mesure μP par

$$(\mu P)(y) = \sum_{x\in E} \mu(x)P(x, y), \quad y \in E.$$

Le vecteur ligne μP s'obtient par produit matriciel (à gauche).

— On définit le produit PQ de deux matrices stochastiques P, Q par

$$(PQ)(x, y) = \sum_{z\in E} P(x, z)Q(z, y). \tag{5.4}$$

On voit facilement que PQ est une matrice stochastique puisque pour tout $x \in E$:

$$\begin{aligned}\sum_{y\in E}(PQ)(x, y) &= \sum_{y\in E}\sum_{z\in E} P(x, z)Q(z, y) = \sum_{z\in E}\sum_{y\in E} P(x, z)Q(z, y)\\ &= \Big(\sum_{z\in E} P(x, z)\underbrace{\Big(\sum_{y\in E} Q(z, y)\Big)}_{=1}\Big) = 1.\end{aligned}$$

— Dans la suite, on considère P^n la puissance n-ème de P dans le sens du produit matriciel (5.4) : pour $n = 0$ $P^0(x, y) = \mathbf{1}_{\{x=y\}}$, pour $n = 1$ $P^1 = P$ et pour $n \geq 2$:

$$P_{n+1}(x, y) = \sum_{z\in E} P_n(x, z)P(z, y) \tag{5.5}$$

$$= \sum_{y_1\in E}\cdots\sum_{y_n\in E} P(x, y_1)P(y_1, y_2)\dots P(y_{n-1}, y_n)P(y_n, y) \tag{5.6}$$

où (5.6) vient de (5.5) par une récurrence immédiate.

Définition 5.5 (Chaîne de Markov) *On appelle chaîne de Markov sur un espace d'états E dénombrable toute suite de variables aléatoires $(X_n)_{n\in\mathbb{N}}$ à valeurs dans E telle que les lois conditionnelles vérifient pour tout $n \geq 0$*

$$\mathcal{L}(X_{n+1}|X_0, X_1, \dots, X_n) = \mathcal{L}(X_{n+1}|X_n) \tag{5.7}$$

c'est à dire pour tout $x_0, x_1, \dots, x_n \in E$ tel que $\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) > 0$ et pour $y \in E$

$$\mathbb{P}\big(X_{n+1} = y\,|X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\big) = \mathbb{P}\big(X_{n+1} = y\,|X_n = x_n\big). \tag{5.8}$$

On dit que la chaîne de Markov est **homogène** *s'il existe une matrice stochastique P telle que $\mathcal{L}(X_{n+1}|X_n) = P(X_n, \bullet)$ ou $\mathbb{P}\big(X_{n+1} = y\,|X_n = x_n\big) = P(x_n, y)$, dans ce cas, (5.7) et (5.8) se réécrivent*

$$\begin{aligned} \mathcal{L}(X_{n+1}|X_0, X_1, \dots, X_n) &= P(X_n, \bullet) \\ \mathbb{P}\big(X_{n+1} = y\,|X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\big) &= P(x_n, y). \end{aligned}$$

Remarque 5.6 — Autrement dit, si la chaîne est en x_n à la date n, peu importe de savoir où elle était avant pour connaître sa probabilité d'aller en y à la date suivante $n+1$, dans tous les cas cette probabilité est $P(x, y)$. Ainsi, $P(x, y)$ est appelée aussi probabilité de transition (en une étape) de la chaîne de Markov $(X_n)_{n\geq 0}$ de x vers y.

— On dit encore que le futur ne dépend du passé que par le présent où passé, présent, futur proviennent des interprétations suivantes :

$$\underbrace{X_0, \dots, X_{n-1}, \overbrace{X_n}^{\text{présent}}}_{\text{passé}}, \underbrace{X_{n+1}, X_{n+2}, \dots}_{\text{futur}}$$

— (Homogénéité) Dans ce cours, la matrice de transition P ne dépend pas de n et on parle de chaîne de Markov **homogène**. On pourrait envisager une chaîne dont les transitions sont données par

$$\mathbb{P}\big(X_{n+1} = y\,|X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\big) = P^{(n)}(x_n, y)$$

c'est à dire avec un noyau de transition $P^{(n)}$ qui dépend de n ; on parlerait alors de chaîne de Markov **inhomogène**.

5.2 Exemples de chaîne de Markov

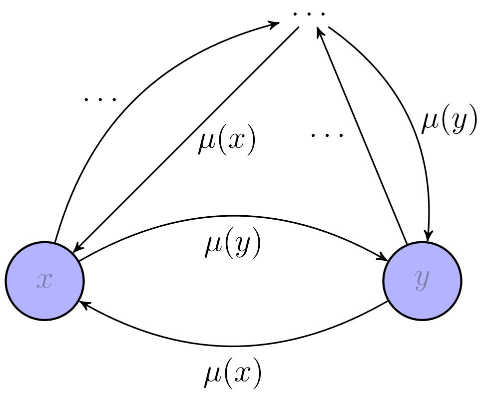
Exemple 5.7 (Suite de variables aléatoires indépendantes et identiquement distribuées) Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires indépendantes et identiquement distribuées de loi f à valeurs dans E alors $(X_n)_{n\geq 0}$ est une chaîne de Markov de matrice de transition $P(x, y) = f(y)$ pour tout $x, y \in E$.

Il s'agit d'une chaîne de Markov de matrice stochastique $P(x, y) = f(y)$ car

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

puisque $\{X_{n+1} = y\} \perp\!\!\!\perp \{X_0 = x_0, \dots, X_n = x\}$ et de même

$$\mathbb{P}(X_{n+1} = y\,|X_n = x) = \mathbb{P}(X_{n+1} = y) = f(y).$$



Exemple 5.8 (Marche aléatoire sur $\mathbb{Z}^d$) Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires indépendantes et identiquement distribuées à valeurs entières de distribution f (ie. $\mathbb{P}(X_1 = x) = f(x)$). On considère une variable aléatoire X_0 de loi μ_0 indépendante de la suite $(X_n)_{n\geq 0}$ et on note $S_n = X_0 + X_1 + \cdots + X_n$. La suite $(S_n)_{n\geq 0}$ est appelée une **marche aléatoire** : $S_0 = X_0$ est la position initiale d'un marcheur et X_n est le pas effectué à la date n qui l'amène à sa position S_n à la date n. C'est une chaîne de Markov d'espace d'états $E = \mathbb{N}$ et de noyau de transition $P(x, y) = f(y - x)$ car

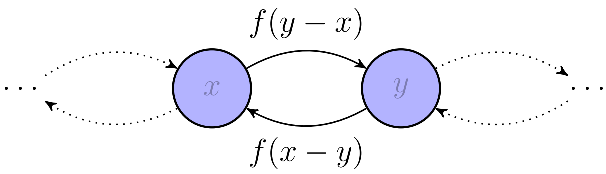
$$\begin{aligned}\mathbb{P}(S_{n+1} = y|S_0 = x_0, S_1 = x_1, \dots, S_n = x) &= \mathbb{P}(X_{n+1} = y - x|S_0 = x_0, S_1 = x_1, \dots, S_n = x)\\ &= \mathbb{P}(X_{n+1} = y - x) = f(y - x).\end{aligned}$$

De même

$$\mathbb{P}(S_{n+1} = y|S_n = x) = \mathbb{P}(X_{n+1} = y - x|S_n = x) = \mathbb{P}(X_{n+1} = y - x) = f(y - x).$$

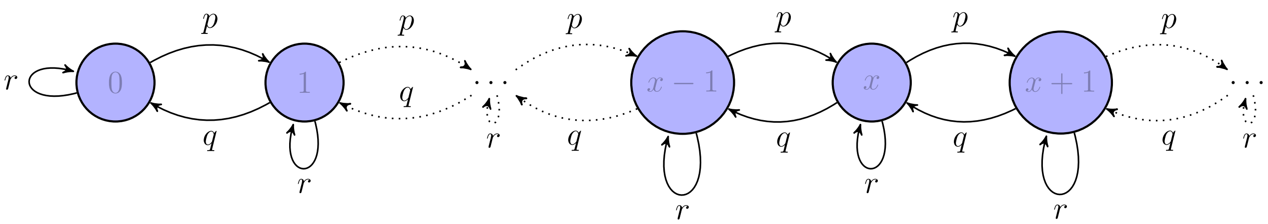
On a aussi la probabilité d'une trajectoire jusqu'à la date n :

$$\begin{aligned}\mathbb{P}(S_0 = x_0, \dots, S_n = x_n) &= \mathbb{P}(X_0 = x_0, X_1 = x_1 - x_0, \dots, X_n = x_n - x_{n-1})\\ &= \mathbb{P}(X_0 = x_0)\mathbb{P}(X_1 = x_1 - x_0) \dots \mathbb{P}(X_n = x_n - x_{n-1})\\ &= \mu_0(x_0) f(x_1 - x_0) \dots f(x_n - x_{n-1}).\end{aligned}$$



On peut considérer le cas spécial d'une **marche simple** sur $\mathbb{Z}$ avec $f(1) = p, f(-1) = q$ et $f(0) = r$ avec $p + q + r = 1$ (le marcheur fait un pas sur la droite $(+1)$ ou sur la gauche (-1) ou reste sur place (0) avec probabilités respectives p, q, r). Dans ce cas, les transitions sont gouvernées par

$$P(x, y) = \begin{cases} p & \text{si } y = x + 1\\ q & \text{si } y = x - 1\\ r & \text{si } y = x\\ 0 & \text{sinon.}\end{cases}$$



Exemple 5.9 (Marche aléatoire sur un graphe) On se donne un graphe au plus dénombrable $(E, \mathcal{A})$ où E désigne l'ensemble des sommets et $\mathcal{A}$ celui des arêtes. On note $\mathcal{A}_x$ l'ensemble des arêtes issues de $x \in E$. On suppose que pour tout $x \in E$, $\mathcal{A}_x$ est fini et non vide. On pose alors

$$P(x,y) = \begin{cases} 1/\mathrm{card}(\mathcal{A}_x) & \text{si } (x,y) \in \mathcal{A} \\ 0 & \text{sinon.} \end{cases}$$

Une chaîne de Markov de transition P est appelée marche aléatoire simple sur le graphe $(E, \mathcal{A})$.

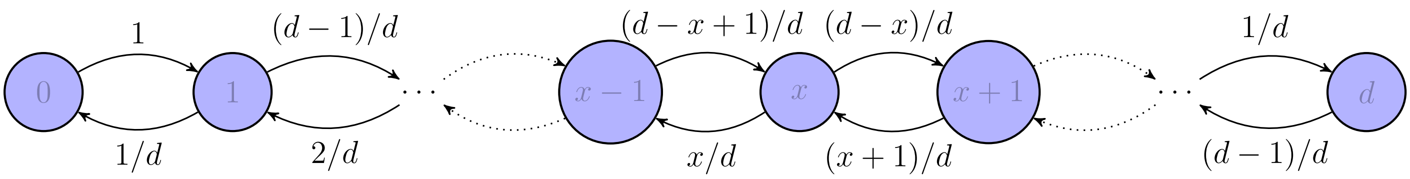
Exemple 5.10 (Ehrenfest) Il s'agit d'un modéle élémentaire d'échange de molécules de gaz entre deux corps isolés introduit par le (couple de) physiciens Ehrenfest [1].

Considérons deux boîtes A et B et d boules numérotées de 1 à d. On suppose qu'à l'origine certaines boules sont dans A, les autres dans B. À chaque étape (et indépendamment des autres étapes), on choisit au hasard une boule parmi $1, 2, \dots, d$ et elle est retirée de sa boîte pour être placée dans l'autre. On note X_n le nombre de boules présentes dans la boîte A après n étapes.

Il s'agit d'une chaîne de Markov à espace d'états $E = \{0, \dots, d\}$. Si on suppose que $X_n = x$, alors avec une probabilité x/d on tire une boule de la boîte A pour la déplacer en B si bien que $X_{n+1} = x - 1$. Avec probabilité $(d-x)/d$, on a $X_{n+1} = x + 1$. On en déduit les probabilités de transition

$$P(x,y) = \mathbb{P}(X_{n+1} = y | X_n = x) = \begin{cases} x/d & \text{si } y = x-1 \\ (d-x)/d & \text{si } y = x+1 \\ 0, & \text{sinon.} \end{cases}$$

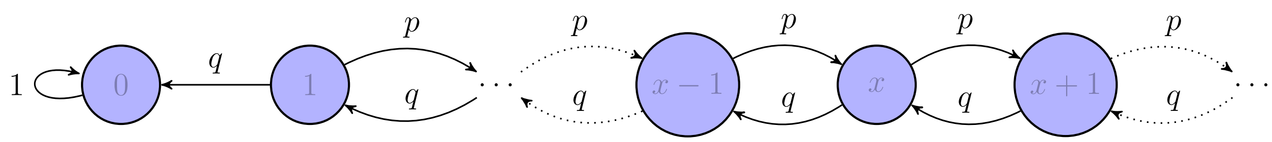
Noter qu'en une étape la chaîne d'Ehrenfest ne peut passer de l'état $x \notin \{0, d\}$ qu'à l'état $x - 1$ ou $x + 1$ tandis que 0 mène à 1, d à $d - 1$.



1. Paul Ehrenfest (autrichien, 1880–1933) et Tatiana Ehrenfest-Afanaseva (russo-néerlandaise, 1876–1964).

Exemple 5.11 (Ruine du joueur) On considère un joueur qui commence une partie avec un capital en euro (€) et fait une série de paris de 1 €. On suppose qu'il a une probabilité p de gagner chaque pari, $q = 1 - p$ de le perdre et que si son capital atteint 0 alors il est ruiné et doit arrêter. On note X_n son capital après le n-ème pari. C'est une chaîne de Markov avec 0 comme état absorbant, d'espace d'états $E = \mathbb{N}$ et de fonction de transition donnée par $P(0, 0) = 1$ (et $P(0, y) = 0$ pour $y > 0$) et si $x > 0$

$$P(x, y) = \mathbb{P}(X_{n+1} = y | X_n = x) = \begin{cases} q & \text{si } y = x - 1 \\ p & \text{si } y = x + 1 \\ 0 & \text{sinon.} \end{cases}$$



On parle de la chaîne de la **ruine du joueur** sur $E = \mathbb{N}$. On pourrait rajouter un deuxième état absorbant en d en demandant au joueur d'arrêter si son capital atteint d.

On peut aussi supposer que deux parieurs jouent l'un contre l'autre par des paris de 1 € avec un capital total fixe de d € dont la répartition entre les deux joueurs évolue en fonction des résultats des paris.

Définition 5.12 (État absorbant) *On appelle état absorbant d'une chaîne de Markov de noyau de transition P tout état $a \in E$ tel que $P(a, a) = 1$, ie. si la chaîne arrive en a, elle y reste !*

5.3 Probabilités trajectorielles

On considère dans cette section une chaîne de Markov $(X_n)_{n \geq 0}$ d'espace d'états E dénombrable et de noyau de transition P.

Proposition 5.13 *Une suite de variables aléatoires $(X_n)_{n \geq 0}$ à valeurs dans E est une chaîne de Markov de matrice de transition P si et seulement si pour tout $n \geq 0$ et pour tout $x_0, \dots, x_n \in E$*

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_0 = x_0) P(x_0, x_1) \dots P(x_{n+1}, x_n). \tag{5.9}$$

Remarque 5.14 Les lois jointes d'une chaîne de Markov (homogène) sont donc entièrement déterminées si on donne sa distribution initiale μ_0 (point de départ) et son noyau de transition P (évolution au cours du temps). Voir aussi la Déf. 5.34.

Démonstration :[Prop. 5.13] Si $(X_n)_{n \geq 0}$ est une chaîne de Markov de noyau de transition P alors l'identité (5.9) s'obtient par une récurrence immédiate : en effet, la récurrence est automatiquement initialisée pour $n = 0$; puis si (5.9) est vraie pour le rang n alors d'abord lorsque $\mathbb{P}\big(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\big) \neq 0$ on a :

$\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n, X_{n+1} = x_{n+1})$

$$\begin{aligned}
&= \mathbb{P}\big(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\big)\mathbb{P}\big(X_{n+1} = x_{n+1}\,|X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\big)\\
&= \mathbb{P}\big(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\big)P(x_n, x_{n+1})\\
&= \mathbb{P}(X_0 = x_0)P(x_0, x_1)\dots P(x_{n+1}, x_n)P(x_n, x_{n+1}) \quad \text{(hyp. récurrence (5.9) pour } n).
\end{aligned}$$

Puis lorsque $\mathbb{P}\big(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\big) = 0$ alors d'une part

$$\mathbb{P}(X_0 = x_0)P(x_0, x_1)\dots P(x_{n+1}, x_n) = 0$$

donc $\mathbb{P}(X_0 = x_0)P(x_0, x_1)\dots P(x_{n+1}, x_n)P(x_n, x_{n+1}) = 0$ et d'autre part

$$\mathbb{P}\big(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n, X_{n+1} = x_{n+1}\big) = 0,$$

si bien que (5.9) reste vraie, ce qui achève d'établir complètement (5.9) par récurrence.

Réciproquement, si (5.9) est vraie pour tout $n \geq 0$ alors pour $x_0, x_1, \dots, x_n \in E$ tels que $\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \neq 0$, on a

$$\begin{aligned}
&\mathbb{P}\big(X_{n+1} = x_{n+1}\,|X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\big)\\
&\quad= \frac{\mathbb{P}\big(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n, X_{n+1} = x_{n+1}\big)}{\mathbb{P}\big(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n\big)}\\
&\quad= \frac{\mathbb{P}(X_0 = x_0)P(x_0, x_1)\dots P(x_{n-1}, x_n)P(x_n, x_{n+1})}{\mathbb{P}(X_0 = x_0)P(x_0, x_1)\dots P(x_{n-1}, x_n)}\\
&\quad= P(x_n, x_{n+1})
\end{aligned}$$

et donc la Définition 5.5 d'une chaîne de Markov est bien satisfaite. □

Proposition 5.15 *(1) Soit $(X_n)_{n\geq 0}$ une chaîne de Markov sur E de noyau de transition P. Alors, pour tout $n \geq 0$ et $f : E \to \mathbb{R}$, on a*

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

(2) Plus généralement, pour tout $i_1, \dots, i_k \in \{0, \dots, n-1\}$, on a

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

Démonstration : (1) Comme l'espérance conditionnelle est l'espérance par rapport à la conditionnelle (cf. (1.13)), d'après la Définition 5.5, on a

$$\mathbb{E}\big[f(X_{n+1})|X_0, \dots, X_n\big] = \mathbb{E}\big[f(X_{n+1})|X_n\big] = \sum_{y\in E} P(X_n, y)f(y) = Pf(X_n).$$

(2) Ensuite, si $i_1, \dots, i_k \in \{0, \dots, n-1\}$, par conditionnement en cascade (Prop. 1.13 ou Th. 2.12) avec $\mathcal{G}_1 := \sigma(X_{i_1}, \dots, X_{i_k}, X_n) \subset \mathcal{G}_2 := \sigma(X_0, \dots, X_n)$, on a :

$$\begin{aligned}
\mathbb{E}\big[f(X_{n+1})\,|X_{i_1}, \dots, X_{i_k}, X_n\big] &= \mathbb{E}\Big[\mathbb{E}[f(X_{n+1})\,|X_0, \dots, X_n]\,\Big|X_{i_1}, \dots, X_{i_k}, X_n\Big]\\
&= \mathbb{E}\big[Pf(X_n)\,|X_{i_1}, \dots, X_{i_k}, X_n\big]\\
&= Pf(X_n),
\end{aligned}$$

puisque $Pf(X_n)$ est $\sigma(X_n)$ donc $\sigma(X_{i_1}, \dots, X_{i_k}, X_n)$-mesurable. □

Transition en n étapes

Le noyau de transition en n étapes donne, pour tout $x, y \in E$, la probabilité d'aller de x en y en n étapes ie. $P_n(x,y) = \mathbb{P}(X_n = y | X_0 = x)$. Il est donné par

$$P_n = P^n \tag{5.10}$$

où pour rappel P^n est la puissance n-ème de P, dans le sens du produit matriciel (5.4), cf. (5.6)).

En effet $P_0(x,y) = \delta_x(y)$, $P_1(x,y) = P(x,y)$ et pour $n \geq 2$, avec la partition

$$\{X_n = y\} = \bigsqcup_{x_1 \in E, \dots, x_{n-1} \in E} \big\{X_1 = x_1, X_2 = x_2, \dots, X_{n-1} = x_{n-1}, X_n = y\big\},$$

par additivité de $\mathbb{P}(\cdot | X_0 = x)$, on a :

$$\begin{aligned} P_n(x,y) &= \sum_{x_1 \in E, \dots, x_{n-1} \in E} \mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_{n-1} = x_{n-1}, X_n = y | X_0 = x) \\ &= \sum_{x_1 \in E, \dots, x_{n-1} \in E} \frac{\mathbb{P}(X_0 = x, X_1 = x_1, X_2 = x_2, \dots, X_{n-1} = x_{n-1}, X_n = y)}{\mathbb{P}(X_0 = x)} \\ &= \sum_{x_1 \in E, \dots, x_{n-1} \in E} P(x, x_1) P(x_1, x_2) \dots P(x_{n-1}, y) = P^n(x,y), \end{aligned} \tag{5.11}$$

en utilisant (5.9). Ainsi, on a :

Proposition 5.16 (Chapman-Kolmogorov) *Le noyau de transition en n étapes vérifie une propriété de semi-groupe (dite relation de Chapman-Kolmogorov) : $P_{n+m} = P_n P_m$ (dans le sens du produit (5.4)), ie.*

$$P_{n+m}(x,y) = \sum_{z \in E} P_n(x,z) P_m(z,y). \tag{5.12}$$

Démonstration : C'est immédiat par (5.10) puisque $P_{n+m} = P^{n+m} = P^n P^m = P_n P_m$; cela se retrouve aussi directement par le calcul à partir de l'expression (5.11), on a :

$$\begin{aligned} P_{n+m}(x,y) &= \sum_{y_1 \in E} \cdots \sum_{y_n \in E} \sum_{y_{n+m-1} \in E} P(x, y_1) P(y_1, y_2) \dots P(y_{n-1}, y_n) P(y_n, y_{n+1}) \\ &\qquad\qquad P(y_{n-2}, y_{n-1}) \dots P(y_{n+m-1}, y) \\ &= \sum_{y_n \in E} \left(\sum_{y_1 \in E} \cdots \sum_{y_{n-1} \in E} P(x, y_1) P(y_1, y_2) \dots P(y_{n-1}, y_n) \right) \\ &\qquad \left(\sum_{y_{n+1} \in E} \cdots \sum_{y_{n+m-1} \in E} P(y_n, y_{n+1}) P(y_{n-2}, y_{n-1}) \dots P(y_{n+m-1}, y) \right) \end{aligned}$$

$$= \sum_{y_n \in E} P_n(x, y_n) P_m(y_n, y).$$

□

Remarque 5.17 (Semi-groupe) — La formule de Chapman-Kolmogorov (5.12) montre que P_n est la puissance n-ème de P en terme de produit matriciel : $P_n = P^n$. On utilise indifféremment l'une ou l'autre notation dans la suite.

— Dans le cas E espace d'états fini, il s'agit de vraies matrices et de vrais produits matriciels. Dans le cas E dénombrable, il s'agit d'une généralisation naturelle aux matrices infinies.

Si on note $\mu_0 = (\mathbb{P}(X_0 = x))_{x \in E}$ la distribution initiale de la chaîne et $\mu_n = (\mathbb{P}(X_n = x))_{x \in E}$ celle de l'état X_n à la date n, alors pour tout $y \in E$:

$$\begin{aligned} \mathbb{P}(X_n = y) &= \sum_{x \in E} \mathbb{P}(X_n = y | X_0 = x) \mu_0(x) \\ &= \sum_{x \in E} \mu_0(x) P^n(x, y), \end{aligned}$$

c'est à dire $\mu_n = \mu_0 P^n$. On a aussi $\mu_n = \mu_{n-1} P$ puisque pour tout $y \in E$:

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

On détermine donc la distribution de X_n à partir de la distribution initiale μ_0 et du noyau de transition en n étapes P^n. On peut aussi calculer μ_n à partir de la loi à la date précédente μ_{n-1} avec transition en une étape :

Notations. Dans la suite, on utilise la notation $\mathbb{P}_\nu$ pour indiquer qu'on suppose que la loi initiale de la chaîne est ν ie. $X_0 \sim \mu_0 = \nu$. On note aussi $\mathbb{P}_x = \mathbb{P}_{\delta_x}$ lorsque la chaîne part de $X_0 = x$, autrement dit avec la distribution initiale $\mu_0 = \delta_x$, cf. après la Déf. 5.34.

Expressions explicites

La proposition suivante donne des expressions explicites pour les calculs de lois jointes conditionnelles :

Proposition 5.18 (Lois jointes d'une chaîne de Markov) *Pour une chaîne de Markov $(X_n)_{n \geq 0}$ d'espace d'états E (dénombrable) et de noyau de transition P, en supposant les probabilités conditionnelles bien définies, on a*

(1) Pour $x_0, \dots, x_{n-1}, x_n$ et $y_1, \dots, y_m$ dans E on a :

$$\begin{aligned} &\mathbb{P}(X_{n+1} = y_1, \dots, X_{n+m} = y_m \,|X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x_n) \\ &= P(x_n, y_1) P(y_1, y_2) \dots P(y_{m-1}, y_m); \end{aligned} \tag{5.13}$$

En particulier, on a

$$\mathbb{P}\big(X_{n+m} = y_m \,|X_0 = x_0, \dots, X_n = x_n\big) = \mathbb{P}\big(X_{n+m} = y_m \,|X_n = x_n\big) = P_m(x_n, y_m); \tag{5.14}$$

(2) Pour $A_0, \dots, A_{n-1} \subset E$, on a :

$$\begin{aligned} &\mathbb{P}(X_{n+1} = y_1, \dots, X_{n+m} = y_m \,|X_0 \in A_0, \dots, X_{n-1} \in A_{n-1}, X_n = x_n) \\ &= \quad P(x_n, y_1)P(y_1, y_2) \dots P(y_{m-1}, y_m); \end{aligned} \tag{5.15}$$

(3) Pour $A_0, \dots, A_{n-1} \subset E$ et $B_1, \dots, B_m \subset E$, on a :

$$\begin{aligned} &\mathbb{P}(X_{n+1} \in B_1, \dots, X_{n+m} \in B_m \,|X_0 \in A_0, \dots, X_{n-1} \in A_{n-1}, X_n = x_n) \\ &= \sum_{y_1 \in B_1} \cdots \sum_{y_m \in B_m} P(x_n, y_1)P(y_1, y_2) \dots P(y_{m-1}, y_m). \end{aligned} \tag{5.16}$$

La preuve va utiliser le résultat suivant :

Lemme 5.19 (Probabilités conditionnelles) *Soit $A \subset E$ et B_i des évènements disjoints non-négligeables. Lorsque $\mathbb{P}(A|B_i)$ ne dépend pas de $i \in I$, alors*

$$\mathbb{P}\Big(A \,\Big|\, \bigsqcup_{i\in I} B_i\Big) = \mathbb{P}(A|B_i) \quad \forall i \in I. \tag{5.17}$$

Démonstration :[Lemme 5.19] En notant $\mathbb{P}_C = \mathbb{P}(\cdot|C)$, observer que $\mathbb{P}_C(A|B) = \mathbb{P}(A|B\cap C)$. En effet

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

Lorsque $\mathbb{P}(A|B_i) = \alpha$ pour tout $i \in I$, la formule des probabilités totales (1.3) avec $\mathbb{P}_Big(\cdot\Big| \bigsqcup_{i\in I} B_i\Big) = \mathbb{P}_{\bigsqcup_{i\in I} B_i}$, on a

$$\begin{aligned} \mathbb{P}\Big(A \,\Big|\, \bigsqcup_{i\in I} B_i\Big) &= \sum_{i\in I} \mathbb{P}_{\bigsqcup_{i\in I} B_i}(A|B_i)\mathbb{P}_{\bigsqcup_{i\in I} B_i}(B_i) \\ &= \sum_{i\in I} \mathbb{P}(A|B_i)\mathbb{P}\Big(B_i \,\Big|\bigsqcup_{i\in I} B_i\Big) \\ &= \alpha \underbrace{\sum_{i\in I} \mathbb{P}\Big(B_i \,|\bigsqcup_{i\in I} B_i\Big)}_{=1} = \alpha. \end{aligned}$$

□

Démonstration :[Prop. 5.18] (1) Les lois jointes conditionnelles sont données par :

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

$$\begin{aligned}
&= \frac{\mathbb{P}(X_0 = x_0, \dots, X_n = x_n, X_{n+1} = x_{n+1}, \dots, X_{n+m} = x_{n+m})}{\mathbb{P}(X_0 = x_0, \dots, X_n = x_n)} \\
&= \frac{\mu_0(x_0)P(x_0, x_1) \cdots P(x_{n+m-1}, x_{n+m})}{\mu_0(x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n)} \\
&= P(x_n, x_{n+1}) \cdots P(x_{n+m-1}, x_{n+m})
\end{aligned}$$

qu'on peut réécrire sous la forme (5.13). Le cas particulier (5.14) s'obtient alors en faisant la somme sur $y_1, \dots, y_{p-1} \in E$ et avec la définition (5.11) de P_m et la définition (5.8) d'une chaîne de Markov.

(2) On écrit la partition

$$\big\{X_0 \in A_0, \dots, X_{n-1} \in A_{n-1}, X_n = x_n\big\} = \bigsqcup_{\substack{x_i \in A_i \\ 0 \le i \le n-1}} \big\{X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x_n\big\}.$$

Comme d'après (5.13), on a

$$\begin{aligned}
&\mathbb{P}(X_{n+1} = y_1, \dots, X_{n+m} = y_m | X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x_n) \\
&\quad = P(x_n, y_1) \dots P(y_{m-1}, y_m)
\end{aligned}$$

pour tout $x_i \in A_i$, $0 \le i \le n-1$, (5.17) dans le Lemme 5.19 s'applique et assure (5.15).

(3) Cela vient de (5.15) avec la σ-additivité de $\mathbb{P}$:

$$\begin{aligned}
&\mathbb{P}(X_{n+1} \in B_1, \dots, X_{n+m} \in B_m | X_0 \in A_0, \dots, X_{n-1} \in A_{n-1}, X_n = x_n) \\
&\quad = \sum_{y_1 \in B_1} \cdots \sum_{y_m \in B_m} \mathbb{P}(X_{n+1} = y_1, \dots, X_{n+m} = y_m | X_0 \in A_0, \dots, X_{n-1} \in A_{n-1}, X_n = x_n) \\
&\quad = \sum_{y_1 \in B_1} \cdots \sum_{y_m \in B_m} P(x_n, y_1)P(y_1, y_2) \dots P(y_{m-1}, y_m).
\end{aligned}$$

□

Approche récursive

Pour montrer qu'une suite de variables aléatoires à valeurs dans E est une chaîne de Markov, la proposition suivante est souvent plus pratique que de revenir à la Définition 5.5.

Proposition 5.20 (Suite récursive et Markov) *Soit X_0 une variable aléatoire à valeurs dans E de loi ν. Soit $(U_n)_{n\ge 0}$ une suite de variables aléatoires indépendantes et identiquement distribuées de loi μ à valeurs dans F, indépendantes de X_0. Pour une fonction $f : E \times F \to E$ mesurable, on définit par récurrence*

$$X_{n+1} = f\big(X_n, U_{n+1}\big), \quad n \ge 0. \tag{5.18}$$

Alors $(X_n)_{n\geq 0}$ est une chaîne de Markov (homogène) de transition

$$P(x,y)=\mathbb{P}\big(f(x,U)=y\big) \tag{5.19}$$

où $U\sim\mu$.

Démonstration : D'abord, on observe que pour chaque $n\geq 0$: X_n est $\sigma(X_0,U_1,\dots,U_n)$-mesurable. En effet, c'est clair pour $n=0$ et si c'est vrai pour X_n alors $X_{n+1}=f(X_n,U_{n+1})$ est mesurable par rapport à $\sigma(X_n,U_{n+1})\subset\sigma(X_0,U_1,\dots,U_n,U_{n+1})$. On en déduit $\sigma(X_0,\dots,X_n)\subset\sigma(X_0,U_1,\dots,U_n)$.

On vérifie la Définition 5.5 pour $(X_n)_{n\geq 0}$ définie en (5.18). Pour cela, soit $x_1,\dots,x_n,x_{n+1}\in E$ avec $\mathbb{P}(X_0=x_0,\dots,X_n=x_n)\neq 0$. On a

$$\begin{aligned}
&\mathbb{P}\big(X_{n+1}=x_{n+1}\,|X_0=x_0,\dots,X_n=x_n\big)\\
&\quad= \mathbb{P}\big(f(X_n,U_{n+1})=x_{n+1}\,|X_0=x_0,\dots,X_n=x_n\big)\\
&\quad= \mathbb{P}\big(f(x_n,U_{n+1})=x_{n+1}\,|X_0=x_0,\dots,X_n=x_n\big)\\
&\quad= \mathbb{P}\big(f(x_n,U_{n+1})=x_{n+1}\big)\quad(\text{car } U_{n+1}\perp\!\!\!\perp\sigma(X_0,U_1,\dots,U_n)\supset\sigma(X_0,\dots,X_n))\\
&\quad= P(x_n,x_{n+1}),
\end{aligned}$$

ce qui vérifie la Définition 5.5 d'une chaîne de Markov de matrice stochastique P. □

Exemple 5.21 (Suites récursives) On donne quelques exemples de chaînes de Markov données sous forme de suites récursives.

Marche aléatoire sur $\mathbb{Z}^d$. On reprend l'Exemple 5.8. Soit $X_0,X_1,\dots,X_n,\dots$ des variables aléatoires indépendantes et identiquement distribuées dans $\mathbb{Z}^d$ de loi μ. Alors, pour $n\geq 1$, $S_n=\sum_{i=0}^n X_i$ définit une chaîne de Markov de la forme $S_{n+1}=f(S_n,X_n)$ avec $f(x,y)=x+y$ et $(X_n)_{n\geq 1}$ indépendantes et identiquement distribuées. On retrouve le noyau de transition avec (5.19) :

$$P(x,y)=\mathbb{P}(x+X_1=y)=\mathbb{P}(X_1=y-x)=\mu(y-x).$$

En fait, la forme récursive (5.18) de la Prop. 5.20 est la forme typique d'une chaîne de Markov comme le justifie le résultat suivant :

Proposition 5.22 (Markov et suite récursive) *Une chaîne de Markov homogène à valeurs réelles peut être vue (en loi) comme une suite récurrente définie comme dans (5.18).*

La preuve de cette proposition repose sur le lemme suivant sur lequel se fonde la méthode dite d'inversion (voir [Bre-proba]) :

Lemme 5.23 (Méthode d'inversion) *Soit μ une loi de probabilité de fonction de répartition F. On pose $F^{-1}(u)=\inf(x\in\mathbb{R}:F(x)>u)$ pour $u\in]0,1[$. Alors pour $U\sim\mathcal{U}(]0,1[)$, on a $F^{-1}(U)\sim\mu$.*

Démonstration :[Proposition 5.22] Soit $(X_n)_{n\geq 0}$ une chaîne de Markov homogène de transition P. Il s'agit de trouver f et U_1 telles que $X_1 = f(x, U_1)$ si $X_0 = x$. La loi de X_1 est $P(x, \cdot)$. Soit U_1 une variable aléatoire de loi uniforme sur $]0, 1[$ indépendante de X_0 et $f(x, \cdot)$ l'inverse généralisé de la fonction de répartition de X_1 sachant $X_0 = x$ donnée par
$$f(x, u) = \inf\big(y \in \mathbb{R} : P(x,]-\infty, y]) > u\big), \ u \in]0, 1[.$$
Alors $f(x, U_1)$ a la même loi que $\mathcal{L}(X_1|X_0 = x)$ (par la méthode d'inversion du Lemme 5.23). Considérons $(U_i)_{i\geq 1}$ des variables aléatoires indépendantes et identiquement distribuées de loi uniforme sur $]0, 1[$ et indépendantes de X_0. On définit la chaîne $(\widetilde{X}_n)_{n\geq 0}$ par la récurrence (5.18) : $\widetilde{X}_{n+1} = f\big(\widetilde{X}_n, U_{n+1}\big)$ avec f comme ci-dessus et avec $\widetilde{X}_0 \sim \mu_0$. Soit $\widetilde{P}$ sa matrice stochastique. Par la Prop. 5.20, on a
$$\widetilde{P}(x, y) = \mathbb{P}\big(f(x, U_1) = y\big) = P(x, y).$$

□

Exemple 5.24 (Urne de Ehrenfest) On reprend l'Exemple 5.10 pour lequel on a $X_{n+1} = X_n + Y_{n+1}$ où
$$\mathbb{P}(Y_{n+1} = x|X_n) = \begin{cases} \frac{X_n}{d} & \text{si } x = -1 \\ \frac{d-X_n}{d} & \text{si } x = +1 \\ 0 & \text{sinon} \end{cases} \iff \mathcal{L}(Y_{n+1}|X_n) = \frac{X_n}{d}\delta_{-1} + \frac{d - X_n}{d}\delta_1.$$
En considérant $(U_n)_{n\geq 1}$ une suite *iid* de loi $\mathcal{U}([0, 1])$, on a
$$(2\mathbf{1}_{[0,(d-x)/d]}(U_n) - 1) \sim \frac{x}{d}\delta_{-1} + \frac{d - x}{d}\delta_1,$$
si bien qu'avec $f(x, u) = x + (2\mathbf{1}_{[0,(d-x)/d]}(u) - 1)$, on a $X_{n+1} \sim f(X_n, U_{n+1})$.

Temps d'atteinte

Une notion utile dans les calculs de loi pour les chaînes de Markov est celle de temps d'atteinte :

Définition 5.25 (Temps d'atteinte) *Soit $A \subset E$. Le temps d'atteinte de A est $T_A = \min(n \geq 0 : X_n \in A)$ avec par convention $\min \emptyset = +\infty$.*

Le temps d'atteinte T_A est la première date où la chaîne atteint A. En particulier, pour un état y, on définit $T_y = \min(n \geq 0 : X_n = y)$ le temps d'atteinte de y et $\widetilde{T}_y = \min(n > 0 : X_n = y)$ le temps d'atteinte de y après le départ. Sous $\mathbb{P}_x$ pour $x \neq y$ (ie. lorsque la chaîne part de $x \neq y$), on a $\widetilde{T}_y = T_y$. Sous $\mathbb{P}_y$, $T_y = 0$ et $\widetilde{T}_y$ désigne le temps de premier retour pour la chaîne qui part de y.

La proposition suivante donne une équation utile reliant les temps d'atteinte aux probabilités de transition.

Proposition 5.26 *Pour tout $x, y \in E$ avec $x \neq y$, et $n \geq 1$, on a*

$$P^n(x,y) = \sum_{k=1}^{n} \mathbb{P}_x(T_y = k) P^{n-k}(y,y). \tag{5.20}$$

Démonstration : Avec la partition $\{X_n = y\} = \bigsqcup_{k=1}^{n} \{T_y = m, X_n = y\}$ de $\{X_n = y\}$ (ie. $\{T_y = k, X_n = y\}$, $1 \leq k \leq n$, sont disjoints et de réunion $\{X_n = y\}$), on a :

$$\begin{aligned}
P^n(x,y) &= \mathbb{P}_x(X_n = y) = \sum_{k=1}^{n} \mathbb{P}_x(T_y = k, X_n = y) \\
&= \sum_{k=1}^{n} \mathbb{P}_x(T_y = k)\, \mathbb{P}(X_n = y | X_0 = x, T_y = k) \\
&= \sum_{k=1}^{n} \mathbb{P}_x(T_y = k)\, \mathbb{P}(X_n = y | X_0 = x, X_1 \neq y, \dots X_{k-1} \neq y, X_k = y) \\
&= \sum_{k=1}^{n} \mathbb{P}_x(T_y = k)\, \mathbb{P}(X_n = y | X_k = y) \qquad (5.21) \\
&= \sum_{k=1}^{n} \mathbb{P}_x(T_y = k)\, P^{n-k}(y,y),
\end{aligned}$$

en utilisant (5.15) en (5.21). □

En particulier pour un état absorbant a (Définition 5.12), on a la relation suivante :

Proposition 5.27 *Si a est un état absorbant, alors $P^n(x,a) = \mathbb{P}_x(T_a \leq n)$.*

Démonstration : Comme a est absorbant, on a $P^{n-m}(a,a) = 1$ pour tout $1 \leq m \leq n$ et (5.20) devient

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

□

Noter encore la relation $\mathbb{P}_x(T_y = 1) = \mathbb{P}_x(X_1 = y) = P(x,y)$ et

$$\mathbb{P}_x(T_y = 2) = \sum_{z \neq y} \mathbb{P}_x(X_1 = z, X_2 = y) = \sum_{z \neq y} P(x,z) P(z,y).$$

Plus généralement pour $n \geq 1$, on trouve $\mathbb{P}_x(T_y = n)$ par récurrence à partir de

$$\mathbb{P}_x(T_y = n+1) = \sum_{z \neq y} P(x,z) P_z(T_y = n)$$

puisque pour aller de x à y en exactement $n+1$ étapes, il faut aller de x à n'importe quel $z \neq y$ en 1 étape puis de ce z à y en exactement n étapes.

5.4 Chaîne de Markov canonique

On commence par expliquer que la donnée d'une loi uniforme $\mathcal{U}[0,1]$ est équivalente à la donnée d'une suite de lois de Bernoulli $b(1/2)$ indépendantes. Rappelons que tout $x \in [0,1]$ s'écrit en base 2 sous la forme

$$x = \sum_{n=1}^{+\infty} \frac{\varepsilon_n(x)}{2^n} \quad \text{avec } \varepsilon_n(x) \in \{0,1\},\ n \geq 1. \tag{5.22}$$

Lemme 5.28 (Poisson/Bernoulli) *Soit X une variable aléatoire à valeurs dans $[0,1]$ avec la décomposition (5.22). Alors X est de loi uniforme $\mathcal{U}[0,1]$ si et seulement si les variables aléatoires $\varepsilon_n := \varepsilon_n(X)$, $n \geq 1$, sont indépendantes et identiquement distribuées de loi de Bernoulli $b(1/2)$.*

Démonstration : D'abord, on note que X est mesurable si et seulement si les ε_n, $n \geq 1$, le sont. C'est clair dans le sens réciproque, X étant limite des sommes partielles qui sont alors mesurables ; dans le sens direct, on procède par récurrence en écrivant

$$\varepsilon_p = \left[2^p X - \sum_{k=1}^{p-1} 2^{p-k}\varepsilon_k\right]$$

où $[x]$ désigne la partie entière de x. On suppose que les ε_k sont indépendantes et identiquement distribuées de loi de Bernoulli $b(1/2)$ et on calcule la fonction caractéristique de X :

$$\begin{aligned}
\varphi(t) &= \mathbb{E}\left[\exp\left(i\sum_{k=1}^{+\infty}\frac{\varepsilon_k}{2^k}t\right)\right] = \mathbb{E}\left[\lim_{n\to+\infty}\exp\left(i\sum_{k=1}^{n}\frac{\varepsilon_k}{2^k}t\right)\right] = \lim_{n\to+\infty}\mathbb{E}\left[\exp\left(i\sum_{k=1}^{n}\frac{\varepsilon_k}{2^k}t\right)\right] \\
&\quad \text{(convergence dominée)} \\
&= \lim_{n\to+\infty}\prod_{k=1}^{n}\varphi_{\varepsilon_1}\left(\frac{t}{2^k}\right) = \lim_{n\to+\infty}\prod_{k=1}^{n}\frac{1+e^{it/2^k}}{2} = \lim_{n\to+\infty}\prod_{k=1}^{n}\left(e^{it/2^{k+1}}\cos\left(\frac{t}{2^{k+1}}\right)\right) \\
&= \exp\left(\sum_{k=1}^{+\infty}\frac{it}{2^{k+1}}\right)\prod_{k=1}^{+\infty}\cos\left(\frac{t}{2^{k+1}}\right) = e^{it/2}\prod_{k=1}^{+\infty}\cos\left(\frac{t}{2^{k+1}}\right).
\end{aligned}$$

Mais de $\sin t = 2\cos(t/2)\sin(t/2)$, on déduit $\sin(t/2) = 2^n\prod_{k=1}^n \cos\left(t/2^{k+1}\right)\times\sin\left(t/2^{n+1}\right)$ et donc

$$\prod_{k=1}^{+\infty}\cos\left(\frac{t}{2^{k+1}}\right) = \lim_{n\to+\infty}\frac{\sin(t/2)}{2^n\sin\left(t/2^{n+1}\right)} = \frac{2\sin(t/2)}{t} = \frac{e^{it/2}-e^{-it/2}}{it}.$$

On a alors

$$\varphi_X(t) = e^{it/2}\frac{e^{it/2}-e^{-it/2}}{it} = \frac{e^{it}-1}{it},$$

c'est à dire $X \sim \mathcal{U}[0,1]$.

Réciproquement, pour tout $n \geq 1$ et $a_i \in \{0,1\}$, $1 \leq i \leq n$, on a

$$\begin{aligned} \mathbb{P}\big(\varepsilon_1 = a_1, \dots, \varepsilon_n = a_n\big) &= \mathbb{P}\Big(\sum_{i=1}^n \frac{a_i}{2^i} \leq X < \sum_{i=1}^n \frac{a_i}{2^i} + \sum_{i>n} \frac{1}{2^i}\Big) \\ &= \mathbb{P}\Big(\sum_{i=1}^n \frac{a_i}{2^i} \leq X < \sum_{i=1}^n \frac{a_i}{2^i} + \frac{1}{2^n}\Big) = \frac{1}{2^n}, \end{aligned}$$

ce qui permet de voir par récurrence que $\mathbb{P}(\varepsilon_i = a_i) = 1/2$ et $\mathbb{P}\big(\varepsilon_1 = a_1, \dots, \varepsilon_n = a_n\big) = \mathbb{P}\big(\varepsilon_1 = a_1\big) \dots \mathbb{P}\big(\varepsilon_n = a_n\big)$, soit ε_i, $i \geq 1$, sont indépendantes et identiquement distribuées de loi $b(1/2)$. □

Lemme 5.29 (Suite de variables uniformes *iid***)** *L'espace de probabilité* $([0,1[, \mathcal{B}([0,1[), \lambda)$, *où* λ *est la mesure de Lebesgue sur* $[0,1[$, *supporte une suite* $(U_n)_{n\geq 0}$ *de variables aléatoires uniformes indépendantes et identiquement distribuées.*

Démonstration : Par le Lemme 5.28, $\omega \in [0,1[$ s'écrit en base 2 sous la forme (5.22) avec $\varepsilon_n := \varepsilon_n(\omega) \in \{0,1\}$, $n \geq 1$, indépendantes et de loi $b(1/2)$.

On considère une injection φ de $\mathbb{N} \times \mathbb{N}$ dans $\mathbb{N}$ et on pose $\eta_{i,j} = \varepsilon_{\varphi(i,j)}$. Les variables aléatoires $\eta_{i,j}$ restent indépendantes et identiquement distribuées de loi $b(1/2)$. On pose alors $U_i = \sum_{j=1}^{+\infty} \eta_{i,j} 2^{-j}$ et on observe par le théorème des coalitions ([Bre-proba, Th. 5.1.1]) que les variables aléatoires $U_0, U_1, \dots$ sont (mutuellement) indépendantes, de loi uniforme sur $[0,1[$ (Lemme 5.28). □

Proposition 5.30 (Construction d'une chaîne de Markov) *Soit* E *un espace au plus dénombrable et* $P = (P(x,y))_{x,y\in E}$ *une matrice stochastique. On peut trouver un espace de probabilité* $\big(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}}\big)$ *sur lequel il existe pour tout* $x \in E$ *une suite* $\big(\widetilde{X}_n^x\big)_{n\geq 0}$ *qui est une chaîne de Markov de transition* P *et qui est issue de* $\widetilde{X}_0^x = x$ *(ie.* $\mu_0 = \delta_x$*).*

Démonstration : On considère l'espace de probabilité $\big(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}}\big) = ([0,1[, \mathcal{B}([0,1[), \lambda)$ et la suite de variables aléatoires $(U_i)_{i\geq 1}$ indépendantes et de loi $\mathcal{U}[0,1]$ construites dans le Lemme 5.29. Soit $(y_n)_{n\geq 1}$ une énumération des éléments de E (supposé dénombrable). On pose $\widetilde{X}_0^x = x$ puis

$$\begin{array}{rcl} \widetilde{X}_1^x = y_k & \text{si} & \displaystyle\sum_{1\leq j<k} P(x, y_j) < U_1 \leq \sum_{1\leq j\leq k} P(x, y_j) \\ \vdots & & \vdots \\ \widetilde{X}_{n+1}^x = y_k & \text{si} & \displaystyle\sum_{1\leq j<k} P(\widetilde{X}_n^x, y_j) < U_{n+1} \leq \sum_{1\leq j\leq k} P(\widetilde{X}_n^x, y_j). \end{array}$$

Par construction, on a $\widetilde{\mathbb{P}}\big(\widetilde{X}_n^x = y \,\big|\, \widetilde{X}_{n-1}^x = z\big) = P(z,y)$ pour chaque $n \geq 1$.

En effet, pour $n = 1$: comme $\widetilde{X}_0^x = x$ est sûr

$$\begin{aligned}
\widetilde{\mathbb{P}}\big(\widetilde{X}_1^x = y_k \,|\, \widetilde{X}_0^x = x\big) &= \widetilde{\mathbb{P}}\big(\widetilde{X}_1^x = y_k\big) \\
&= \mathbb{P}\left(\sum_{1\leq j<k} P(x,y_j) < U_1 \leq \sum_{1\leq j\leq k} P(x,y_j)\right) \\
&= \sum_{1\leq j\leq k} P(x,y_j) - \sum_{1\leq j<k} P(x,y_j) = P(x,y_k).
\end{aligned}$$

Puis comme les variables aléatoires U_i, $i \geq 0$, sont indépendantes :

$$\begin{aligned}
&\widetilde{\mathbb{P}}\Big(\widetilde{X}_{n+1}^x = y_k \,\big|\, \widetilde{X}_0^x = x, \widetilde{X}_1^x = x_1, \dots, \widetilde{X}_n^x = x_n\Big) \\
&= \widetilde{\mathbb{P}}\Big(\sum_{1\leq j<k} P(\widetilde{X}_n^x, y_j) < U_{n+1} \leq \sum_{1\leq j\leq k} P(\widetilde{X}_n^x, y_j) \,\big|\, \widetilde{X}_0^x = x, \widetilde{X}_1^x = x_1, \dots, \widetilde{X}_n^x = x_n\Big) \\
&= \widetilde{\mathbb{P}}\Big(\sum_{1\leq j<k} P(x_n, y_j) < U_{n+1} \leq \sum_{1\leq j\leq k} P(x_n, y_j)\Big) \qquad (5.23) \\
&= \sum_{1\leq j\leq k} P(x_n,y_j) - \sum_{1\leq j<k} P(x_n,y_j) = P(x_n,y_k)
\end{aligned}$$

en utilisant $\big\{\widetilde{X}_0^x = x_0, \widetilde{X}_1^x = x_1, \dots, \widetilde{X}_n^x = x_n\big\} \in \sigma(U_1,\dots,U_n) \perp\!\!\!\perp U_{n+1}$ pour se débarrasser du conditionnement en (5.23). Ainsi par la Définition 5.5, $\big(\widetilde{X}_n^x\big)_{n\geq 0}$ est bien une chaîne de Markov issue de x et de matrice stochastique P. □

Dans la Prop. 5.30, le choix de l'espace de probabilité $([0,1[, \mathcal{B}([0,1[), \lambda)$ fait dans sa preuve est un peu arbitraire. On considère un espace vraiment canonique en prenant :

— $\Omega = E^{\mathbb{N}}$,

— $\mathcal{F}$ est la tribu cylindrique $\sigma(\mathrm{Cyl})$ engendrée par la famille Cyl des cylindres

$$C = \big\{\omega \in E^{\mathbb{N}} : \omega_{i_1} = x_{i_1}, \dots, \omega_{i_n} = x_{i_n}\big\} \qquad (5.24)$$

où $n \in \mathbb{N}$, $0 \leq i_0 < \dots < i_n$ et $x_{i_1}, \dots, x_{i_n} \in E$.

Sur cet espace mesurable $(\Omega, \mathcal{F})$, $\omega \in \Omega$ est une suite $\omega = (\omega_n)_{n\in\mathbb{N}}$ de E et on considère les applications coordonnées : $X_n(\omega) = \omega_n$, $n \geq 0$.

Lemme 5.31 *La tribu cylindrique $\sigma(Cyl)$ est la plus petite tribu rendant mesurables les applications coordonnées X_n, $n \geq 0$.*

Démonstration : On note $\mathcal{G}$ la plus petite tribu rendant mesurables les applications coordonnées X_n, $n \geq 0$, et on montre la double inclusion.

— Soit $x \in E$, alors $X_n^{-1}(\{x\}) = \{\omega \in E^{\mathbb{N}} : \omega_n = x\} \in \mathrm{Cyl} \subset \sigma(\mathrm{Cyl})$, ce qui justifie la mesurabilité de chaque X_n pour $\sigma(\mathrm{Cyl})$ et donc $\mathcal{G} \subset \sigma(\mathrm{Cyl})$.

— Soit $C \in \mathrm{Cyl}$ comme en (5.24). Comme $C = \bigcap_{p=1}^n X_{i_p}^{-1}(\{x_{i_p}\})$, on a $C \in \mathcal{G}$ et donc $\mathrm{Cyl} \subset \mathcal{G}$ et $\sigma(\mathrm{Cyl}) \subset \mathcal{G}$.

$\square$

La suite s'applique avec tout espace de probabilité $\big(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}}\big)$ vérifiant la Prop. 5.30. D'après la preuve de cette proposition, $([0,1[, \mathcal{B}([0,1[), \lambda)$ convient mais tout autre espace vérifiant la proposition ferait l'affaire. On rappelle que, ci-dessous, $(\Omega, \mathcal{F})$ désigne $(E^{\mathbb{N}}, \sigma(\mathrm{Cyl}))$.

Lemme 5.32 *Soit* $\psi : \big(\widetilde{\Omega}, \widetilde{\mathcal{F}}\big) \to (\Omega, \mathcal{F})$. *Alors* ψ *est mesurable si et seulement si* $X_n \circ \psi$ *est mesurable pour tout* $n \geq 0$.

Démonstration : $\boxed{\Rightarrow}$ Le sens direct est immédiat puisqu'il s'agit alors de composition d'applications mesurables, d'après le choix de $\mathcal{F} = \sigma(\mathrm{Cyl})$.
$\boxed{\Leftarrow}$ Pour le sens réciproque, la famille $\mathcal{G} = \big\{A \in \mathcal{F} : \psi^{-1}(A) \in \widetilde{\mathcal{F}}\big\}$ est une tribu qui contient tous les $X_n^{-1}(\{x\})$, $x \in E$, puisque $X_n^{-1}(\{x\}) = \{\omega \in E^{\mathbb{N}} : \omega_n = x\} \in \mathrm{Cyl} \subset \mathcal{F}$ et par hypothèse

$$\psi^{-1}\big(X_n^{-1}(\{x\})\big) = (X_n \circ \psi)^{-1}(\{x\}) \in \widetilde{\mathcal{F}}.$$

La tribu $\mathcal{G}$ rend donc mesurables les applications coordonnées X_n, $n \geq 0$. Par le Lemme 5.31, $\mathcal{F}$ étant la plus petite tribu rendant mesurables ces applications coordonnées X_n, $n \geq 0$, on a $\mathcal{F} \subset \mathcal{G}$ et finalement $\mathcal{G} = \mathcal{F}$, ce qui signifie que ψ est bien $(\widetilde{\mathcal{F}}, \mathcal{F})$-mesurable. $\square$

Théorème 5.33 (Chaîne canonique) *Soit* E *un espace d'états au plus dénombrable et* $P = (P(x,y))_{x,y \in E}$ *une matrice stochastique sur* E. *Pour toute loi de probabilité* ν *sur* E, *il existe une unique probabilité* $\mathbb{P}_\nu$ *sur* $(\Omega, \mathcal{F}) = (E^{\mathbb{N}}, \sigma(Cyl))$ *telle que sous* $\mathbb{P}_\nu$ *la suite des applications coordonnées* $(X_n)_{n \geq 0}$ *est une chaîne de Markov de matrice stochastique* P *et de loi initiale* ν.

Démonstration : Existence lorsque $\nu = \delta_x$. On commence par traiter le cas de $\nu = \delta_x$, pour $x \in E$, et on cherche une probabilité $\mathbb{P}_x$ telle que, sous $\mathbb{P}_x$, les applications coordonnées $(X_n)_{n \geq 0}$ forment une chaîne de Markov de matrice stochastique P partant de x.

D'après la Prop. 5.30, il existe un espace de probabilité $\big(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}}\big)$ et $(X_n^x)_{n \geq 0}$ une chaîne de Markov de matrice stochastique P avec $X_0^x = x$. On considère alors l'application

$$\psi_x : \begin{cases} (\widetilde{\Omega}, \widetilde{F}) & \longrightarrow & (\Omega, \mathcal{F}) \\ \quad \widetilde{\omega} & \longmapsto & \big(X_n^x(\widetilde{\omega})\big)_{n \geq 0}. \end{cases}$$

Par la première construction de la Prop. 5.30, pour chaque $n \geq 0$, $X_n \circ \psi_x = X_n^x$ est une variable aléatoire. Le Lemme 5.32 assure alors que ψ_x est une application mesurable. On définit alors

$$\mathbb{P}_x = \widetilde{\mathbb{P}} \circ \psi_x^{-1} \tag{5.25}$$

comme la mesure image de $\widetilde{\mathbb{P}}$ par ψ_x. Par définition de la mesure image, avec $C_0 = \{\omega \in \Omega : w_0 = x\}$, on a

$$\mathbb{P}_x(X_0 = x) = \mathbb{P}_x(C_0) = \widetilde{\mathbb{P}}\big(\psi_x^{-1}(C_0)\big) = \widetilde{\mathbb{P}}\big((X_n^x)_{n\geq 0} \in C_0\big) = \widetilde{\mathbb{P}}\big(X_0^x = x\big) = 1.$$

Puis, pour tout $n \geq 1$, $x_0, x_1, \dots, x_n \in E$, en notant $C_n = \{\omega \in \Omega : \omega_0 = x_0, \omega_1 = x_1, \dots, \omega_n = x_n\}$ le cylindre associé, on a

$$\begin{aligned}
\mathbb{P}_x\big(X_0 = x_0, \dots, X_n = x_n\big) &= \mathbb{P}_x(C_n) = \widetilde{\mathbb{P}}\big(\psi_x^{-1}(C_n)\big) = \widetilde{\mathbb{P}}\big((X_n^x)_{n\geq 0} \in C_n\big) \\
&= \widetilde{\mathbb{P}}\big(X_0^x = x_0, X_1^x = x_1, \cdots, X_n^x = x_n\big) \\
&= \widetilde{\mathbb{P}}\big(X_0^x = x_0\big) P(x_0, x_1) \dots P(x_{n-1}, x_n) \qquad (5.26) \\
&= \delta_{x,x_0} P(x_0, x_1) P(x_1, x_2) \dots P(x_{n-1}, x_n) \qquad (5.27)
\end{aligned}$$

en utilisant la Prop. 5.13 pour la chaîne de Markov $(X_n^x)_{n\geq 0}$ en (5.26). D'après cette même Prop. 5.13, (5.27) assure que sous $\mathbb{P}_x$, $(X_n)_{n\geq 0}$ est une chaîne de Markov de matrice stochastique P, et par construction, elle part de x.

Existence dans le cas général. Étant donné une loi ν sur E, on considère

$$\mathbb{P}_\nu = \sum_{x\in E} \nu(x)\, \mathbb{P}_x. \qquad (5.28)$$

Comme $\sum_{x\in E} \nu(x) = 1$, $\mathbb{P}_\nu$ définit bien une probabilité sur $(\Omega, \mathcal{F}) = (E^{\mathbb{N}}, \sigma(\mathrm{Cyl}))$. De plus d'après (5.27), on a

$$\begin{aligned}
\mathbb{P}_\nu\big(X_0 = x_0, \dots, X_n = x_n\big) &= \sum_{x\in E} \nu(x) \mathbb{P}_x\big(X_0 = x_0, \dots, X_n = x_n\big) \\
&= \sum_{x\in E} \nu(x) \delta_{x,x_0} P(x_0, x_1) P(x_1, x_2) \dots P(x_{n-1}, x_n) \\
&= \nu(x_0)\, P(x_0, x_1) P(x_1, x_2) \dots P(x_{n-1}, x_n), \qquad (5.29)
\end{aligned}$$

ce qui caractérise une chaîne de Markov de loi initiale ν et de matrice stochastique P par la Prop. 5.13.

Unicité. Si une autre probabilité $\mathbb{P}'_\nu$ satisfait l'énoncé alors, (5.29) est vérifiée pour les deux probabilités, $\mathbb{P}_\nu$ et $\mathbb{P}'_\nu$. Cela signifie que $\mathbb{P}_\nu$ et $\mathbb{P}'_\nu$ coïncident sur les cylindres. Comme l'intersection de deux cylindres est encore un cylindre, Cyl est stable par intersection et forme donc un π-système. Par le théorème des classes monotones (Th. 0.2), on a $\mathbb{P}_\nu = \mathbb{P}'_\nu$ sur $\mathcal{F} = \sigma(\mathrm{Cyl})$ (tribu cylindrique engendrée par les cylindres). □

Du Th. 5.33, il résulte la définition suivante :

Définition 5.34 (Loi d'une chaîne de Markov) *La loi d'une chaîne de Markov homogène sur E de matrice de stochastique P et de loi initiale ν est l'unique probabilité $\mathbb{P}_\nu$ sur $(E^{\mathbb{N}}, \sigma(Cyl))$ du Th. 5.33.*

De plus, d'après la Prop. 5.13, la loi $\mathbb{P}_\nu$ est caractérisée par :

$$\mathbb{P}_\nu\big(\{\omega \in E^{\mathbb{N}} : \omega_0 = x_0, \dots, \omega_n = x_n\}\big) = \nu(x_0)P(x_0, x_1)\dots P(x_{n-1}, x_n),$$

pour tout $n \geq 1$ et $x_0, \dots, x_n \in E$. Dans la suite, on note $\mathbb{E}_\nu$ l'espérance $\mathbb{P}_\nu$ et lorsque $\nu = \delta_x$, on écrit $\mathbb{E}_x = \mathbb{E}_{\delta_x}$. De (5.28), on déduit

$$\mathbb{E}_\nu = \sum_{x \in E} \nu(x)\, \mathbb{E}_x. \tag{5.30}$$

Remarque 5.35 Si $(X'_n)_{n\geq 0}$ est une chaîne de Markov de loi initiale ν, de matrice stochastique P alors pour tout $B \in \mathcal{F} = \sigma(\mathrm{Cyl})$: $\mathbb{P}\big((X'_n)_{n\geq 0} \in B\big) = \mathbb{P}_\nu(B)$. Les résultats en loi obtenus pour la chaîne canonique se transposent donc à toute chaîne de Markov de même matrice de stochastique P et de même loi initiale ν.

5.5 Propriétés de Markov

Sur l'espace canonique $(\Omega, \mathcal{F}) = (E^{\mathbb{N}}, \sigma(\mathrm{Cyl}))$, on considère les opérateurs de décalage ou translation (ou *shift*) : si $k \in \mathbb{N}$,

$$\Theta_k\big((\omega_n)_{n\geq 0}\big) = (\omega_{k+n})_{n\geq 0}.$$

On a $\Theta_k = \Theta_1^{\circ k}$. Comme, pour tout $n \geq 0$, $X_n \circ \Theta_k = X_{n+k}$ est mesurable, le Lemme 5.32 assure que les Θ_k sont des applications mesurables de $(\Omega, \mathcal{F}) = (E^{\mathbb{N}}, \sigma(\mathrm{Cyl}))$) dans lui même. On note $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$, $n \geq 0$, la filtration naturelle associée à la suite $(X_n)_{n\geq 0}$, $\mathbb{E}_x$ l'espérance sous la probabilité $\mathbb{P}_x$ du Th. 5.33, ie. $\mathbb{E}_x[\mathbf{1}_A] = \mathbb{P}_x(A)$ pour $A \in \mathcal{F}$.

Théorème 5.36 (Markov faible) *Soit $G : \Omega \to \mathbb{R}$ une fonction mesurable positive ou bornée. Alors pour tout $x \in E$, on a :*

$$\mathbb{E}_x\big[G \circ \Theta_n \,|\mathcal{F}_n\big] = \mathbb{E}_{X_n}[G]. \tag{5.31}$$

De manière équivalente, pour toute fonction $\mathcal{F}_n$-mesurable $F : \Omega \to \mathbb{R}$ positive ou bornée, on a :

$$\mathbb{E}_x\big[F \times (G \circ \Theta_n)\big] = \mathbb{E}_x\big[F\, \mathbb{E}_{X_n}[G]\big]. \tag{5.32}$$

Les identités (5.31), (5.32) se généralisent au cas où $\mathbb{E}_x$ est remplacée par $\mathbb{E}_\nu$, l'espérance sous $\mathbb{P}_\nu$ pour toute loi initiale ν sur E.

Démonstration : On prouve la formulation (5.31) de la propriété de Markov faible. La formulation (5.32) en découle par la caractérisation de l'espérance conditionnelle de la

Prop. 2.4. Pour prouver (5.31), on commence par observer que $\mathbb{E}_{X_n}[G]$ est $\sigma(X_n)$ donc $\mathcal{F}_n$-mesurable en tant que composée de X_n et de $x \in E \mapsto \mathbb{E}_x[G]$. Ensuite, on établit

$$\mathbb{E}\big[\mathbf{1}_A\, G \circ \Theta_n\big] = \mathbb{E}\big[\mathbf{1}_A\, \mathbb{E}_{X_n}[G]\big], \quad \forall A \in \mathcal{F}_n. \tag{5.33}$$

Étape 1. On montre d'abord (5.33) pour $G = \mathbf{1}_B$ avec

$$B = \{X_0 = y_0, \dots, X_p = y_p\} \in \mathrm{Cyl}, \tag{5.34}$$

pour $p \in \mathbb{N}$, $y_0, \dots, y_p \in E$. Pour $y \in E$, on a

$$\begin{aligned} \mathbb{E}_y[G] &= \mathbb{E}_y\big[\mathbf{1}_{\{X_0=y_0,\dots,X_p=y_p\}}\big] \\ &= \mathbb{P}_y(X_0 = y_0, \dots, X_p = y_p) \\ &= \mathbf{1}_{\{y_0=y\}} P(y_0, y_1) \dots P(y_{p-1}, y_p). \end{aligned} \tag{5.35}$$

Lorsque $A \in \mathcal{F}_n$ est de de forme cylindrique

$$A = \{X_0 = x_0, \dots, X_n = x_n\} \tag{5.36}$$

pour $x_0, \dots, x_n \in E$, comme $G \circ \Theta_n = \mathbf{1}_{\{X_n=y_0,\dots,X_{n+p}=y_p\}}$, on a

$$\begin{aligned} \mathbb{E}_x\big[\mathbf{1}_A \times (G \circ \Theta_n)\big] &= \mathbb{E}_x\big[\mathbf{1}_{\{X_0=x_0,\dots,X_n=x_n\}}\ \mathbf{1}_{\{X_n=y_0,\dots,X_{n+p}=y_p\}}\big] \\ &= \mathbb{P}_x\big(X_0 = x_0, \dots, X_n = x_n, X_n = y_0, \dots, X_{n+p} = y_p\big) \\ &= \mathbf{1}_{\{x_0=x\}} P(x_0, x_1) \dots P(x_{n-1}, x_n) \mathbf{1}_{\{x_n=y_0\}} P(y_0, y_1) \dots P(y_{p-1}, y_p). \end{aligned}$$

Puis, en utilisant (5.35) on a aussi

$$\begin{aligned} \mathbb{E}_x\big[\mathbf{1}_A\ \mathbb{E}_{X_n}[G]\big] &= \mathbb{E}_x\big[\mathbf{1}_A \mathbf{1}_{\{y_0=X_n\}} P(y_0, y_1) \dots P(y_{p-1}, y_p)\big] \\ &= \mathbb{E}_x\big[\mathbf{1}_{\{X_0=x_0,\dots,X_n=x_n\}} \mathbf{1}_{\{y_0=X_n\}} P(y_0, y_1) \dots P(y_{p-1}, y_p)\big] \\ &= \mathbf{1}_{\{x=x_0\}} P(x_0, x_1) \dots P(x_{n-1}, x_n) \mathbf{1}_{\{x_n=y_0\}} P(y_0, y_1) \dots P(y_{p-1}, y_p), \end{aligned}$$

ce qui prouve bien (5.33) pour $G = \mathbf{1}_B$ avec (5.34) et $A \in \mathcal{F}_n$ donné par (5.36).

Comme la famille des cylindres Cyl est un π-système, par un argument de classe monotone (Th. 0.2), on étend (5.32) de A comme en (5.36) à $A \in \mathcal{F}_n$. En effet,

$$\mathcal{M}_1 = \Big\{A \in \mathcal{F} : \mathbb{E}_x[\mathbf{1}_A(\mathbf{1}_B \circ \Theta_n)] = \mathbb{E}_x\big[\mathbf{1}_A \mathbb{E}_{X_n}[\mathbf{1}_B]\big]\Big\}$$

est une classe monotone (linéarité de $\mathbb{E}$ et convergence monotone). Comme (5.33) est vraie pour $A, B \in \mathrm{Cyl}$ en (5.34) alors $\mathrm{Cyl} \cap \mathcal{F}_n \subset \mathcal{M}_1$. Puis comme $\mathrm{Cyl} \cap \mathcal{F}_n$ est stable par intersection, le théorème de classes monotones (Th. 0.2) assure $\mathcal{F}_n = \sigma(\mathrm{Cyl} \cap \mathcal{F}_n) \subset \mathcal{M}_1$. On a alors (5.33) pour tout $A \in \mathcal{F}_n$ et cela prouve (5.31) pour $G = \mathbf{1}_B$, $B \in \mathrm{Cyl}$.

Étape 2. On montre que (5.31) reste vraie pour $G = \mathbf{1}_B$ avec $B \in \mathcal{F}$. On pose

$$\mathcal{M}_2 = \Big\{B \in \mathcal{F} : \mathbb{E}_x[\mathbf{1}_A(\mathbf{1}_B \circ \Theta_n)] = \mathbb{E}_x\big[\mathbf{1}_A \mathbb{E}_{X_n}[\mathbf{1}_B]\big]\ \forall A \in \mathcal{F}_n\Big\}.$$

Il s'agit de nouveau d'une classe monotone, et, qui contient Cyl, par l'Étape 1. Comme Cyl est stable par intersection, le théorème de classes monotones (Th. 0.2) assure encore $\mathcal{F} = \sigma(\mathrm{Cyl}) \subset \mathcal{M}_2$ et on a alors (5.33) pour tout $A \in \mathcal{F}_n$ et $G = \mathbf{1}_B$, $B \in \mathcal{F}$, ce qui prouve (5.31) pour $G = \mathbf{1}_B$, $B \in \mathcal{F}$.

Étape 3. Enfin, par les arguments usuels de théorie de la mesure (linéarité pour passer aux fonctions simples, convergence monotone pour passer aux fonctions mesurables positives, parties positive et négative pour traiter le cas de fonctions de signes quelconques), on étend encore (5.31) aux fonctions $\mathcal{F}$-mesurables G pour lesquelles les espérances sont bien définies.

Finallement, lorsque (5.31) est vraie pour $\mathbb{E}_x$, on la déduit immédiatement pour $\mathbb{E}_\nu$ par sommation à partir de (5.30) :

$$\mathbb{E}_\nu\big[G \circ \Theta_n \,|\mathcal{F}_n\big] = \sum_{x\in E} \nu(x)\mathbb{E}_x\big[G \circ \Theta_n \,|\mathcal{F}_n\big] = \sum_{x\in E} \nu(x)\mathbb{E}_{X_n}[G] = \mathbb{E}_{X_n}[G].$$

□

La propriété de Markov reste vraie si on conditionne avec un temps d'arrêt T (Définition 3.7) :

Théorème 5.37 (Markov fort) *Soit T un temps d'arrêt de la filtration naturelle $(\mathcal{F}_n)_{n\geq 0}$ de la chaîne de Markov $(X_n)_{n\geq 0}$. Alors pour toute fonction mesurable $G : \Omega \to \mathbb{R}$ positive ou bornée, on a :*

$$\mathbb{E}_x\big[\mathbf{1}_{\{T<+\infty\}} G \circ \Theta_T \,|\mathcal{F}_T\big] = \mathbf{1}_{\{T<+\infty\}}\mathbb{E}_{X_T}[G]. \tag{5.37}$$

De manière équivalente, pour toute fonction $\mathcal{F}_T$-mesurable $F : \Omega \to \mathbb{R}$ positive ou bornée, on a :

$$\mathbb{E}_x\big[\mathbf{1}_{\{T<+\infty\}}\, F \times (G \circ \Theta_T)\big] = \mathbb{E}_x\big[\mathbf{1}_{\{T<+\infty\}}\, F\, \mathbb{E}_{X_T}[G]\big]. \tag{5.38}$$

De nouveau, (5.37) et (5.38) restent vrais si on y remplace $\mathbb{E}_x$ par $\mathbb{E}_\nu$, pour toute loi ν sur E.

Démonstration : D'abord, on observe que $\mathbf{1}_{\{T<+\infty\}}\mathbb{E}_{X_T}[G]$ est $\mathcal{F}_T$-mesurable. En effet pour tout borélien B

$$\big\{\mathbf{1}_{\{T<+\infty\}}\mathbb{E}_{X_T}[G] \in B\big\} \cap \{T \leq n\} = \bigcup_{k\leq n} \Big(\{\mathbb{E}_{X_k}[G] \in B\} \cap \{T = k\}\Big) \in \mathcal{F}_n$$

car pour $k \leq n$, $\{T = k\} \in \mathcal{F}_k \subset \mathcal{F}_n$, $\{\mathbb{E}_{X_k}[G] \in B\} \in \mathcal{F}_k \subset \mathcal{F}_n$ ($\mathcal{F}_k$-mesurabilité de $\mathbb{E}_{X_k}[G]$. On a donc $\{\mathbf{1}_{\{T<+\infty\}}\mathbb{E}_{X_T}[G] \in B\} \in \mathcal{F}_T$ et $\mathbf{1}_{\{T<+\infty\}}\mathbb{E}_{X_T}[G]$ est bien $\mathcal{F}_T$-mesurable.

Ensuite pour $A \in \mathcal{F}_T$, on a

$$A \cap \{T = n\} = \big(A \cap \{T \leq n\}\big) \setminus \big(A \cap \{T \leq n-1\}\big) \in \mathcal{F}_n$$

puisque $A \cap \{T \leq n\} \in \mathcal{F}_n$ et $A \cap \{T \leq n-1\} \in \mathcal{F}_{n-1} \subset \mathcal{F}_n$. On a alors

$$\begin{aligned}
\mathbb{E}_x\big[\mathbf{1}_A \mathbf{1}_{\{T<+\infty\}} G \circ \Theta_T\big] &= \sum_{n=0}^{+\infty} \mathbb{E}_x\big[\mathbf{1}_{A\cap\{T=n\}} G \circ \Theta_T\big] = \sum_{n=0}^{+\infty} \mathbb{E}_x\big[\mathbf{1}_{A\cap\{T=n\}} G \circ \Theta_n\big] \\
&= \sum_{n=0}^{+\infty} \mathbb{E}_x\big[\mathbf{1}_{A\cap\{T=n\}} \mathbb{E}_{X_n}[G]\big] = \sum_{n=0}^{+\infty} \mathbb{E}_x\big[\mathbf{1}_{A\cap\{T=n\}} \mathbb{E}_{X_T}[G]\big] \quad (5.39) \\
&= \mathbb{E}_x\big[\mathbf{1}_A \mathbf{1}_{\{T<+\infty\}} \mathbb{E}_{X_n}[G]\big]
\end{aligned}$$

en appliquant la propriété de Markov faible (5.31) dans (5.39).

Comme pour le Th. 5.36, on montre que (5.37), (5.38) restent vraies pour $\mathbb{E}_\nu$ à partir de (5.30). □

La situation la plus intéressante du Th. 5.37 est lorsqu'on sait que T est fini p.s. :

Corollaire 5.38 *Soit 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 $\mathbb{P}_x$, Θ_T est indépendante de $\mathcal{F}_T$ et a pour loi $\mathbb{P}_y$, ce qu'on peut écrire :*

$$\mathcal{F}_T \perp\!\!\!\perp_{\mathbb{P}_x} \Theta_T \sim \mathbb{P}_y.$$

Ce corollaire s'applique typiquement avec $T = T_y = \inf(n \geq 0 : X_n = y)$, le temps d'atteinte de $y \in E$ (récurrent, cf. Déf. 6.3).

Démonstration : Soit $A \in \mathcal{F}_T$ et $B \subset E$, alors

$$\begin{aligned}
\mathbb{P}_x(A, \Theta_T \in B) &= \mathbb{E}_x\big[\mathbf{1}_A\, \mathbf{1}_B \circ \Theta_T\big] = \mathbb{E}_x\big[\mathbf{1}_A\, \mathbb{E}_{X_T}[\mathbf{1}_B]\big] \\
&= \mathbb{E}_x\big[\mathbf{1}_A\, \mathbb{E}_y[\mathbf{1}_B]\big] = \mathbb{E}_x[\mathbf{1}_A]\, \mathbb{E}_y[\mathbf{1}_B] = \mathbb{P}_x(A)\, \mathbb{P}_y(B). \quad (5.40)
\end{aligned}$$

Avec $A = \Omega$, (5.40) donne $\mathbb{P}_x(\Theta_T \in B) = \mathbb{P}_y(B)$, c'est à dire $\Theta_T \overset{\mathbb{P}_x}{\sim} \mathbb{P}_y$. Et en ré-injectant cette égalité dans (5.40), on a pour tout $A \in \mathcal{F}_T$ et $B \subset E$

$$\mathbb{P}_x(A, \Theta_T \in B) = \mathbb{P}_x(A)\mathbb{P}_x(\Theta_T \in B)$$

soit $\mathcal{F}_T \perp\!\!\!\perp_{\mathbb{P}_x} \Theta_T$. □

Remarque 5.39 (Propriétés de Markov sous $\mathbb{P}_\nu$) Les propriétés de Markov faibles (5.31), (5.32) du Théorème 5.36 et fortes (5.37), (5.38) du Théorème 5.37 restent vraies si on remplace $\mathbb{E}_x$ par $\mathbb{E}_\nu$ pour toute loi initiale ν. En effet, on rappelle que $\mathbb{P}_\nu = \sum_{x\in E} \nu(x)\mathbb{P}_x$ et $\mathbb{E}_\nu = \sum_{x\in E} \nu(x)\mathbb{E}_x$, cf. 5.28). Ainsi en sommant convenablement par exemple l'égalité (5.38) on obtient

$$\mathbb{E}_\nu\big[F \times (G \circ \Theta_T)\big] = \mathbb{E}_\nu\big[F\ \mathbb{E}_{X_T}[G]\big]. \quad (5.41)$$

De même pour (5.31), (5.32) et (5.37) et pour le Corollaire 5.38 qui restent vrais pour $\mathbb{P}_\nu$ à la place de $\mathbb{P}_x$.

Reformulation de la propriété de Markov

En notant $\mathcal{L}_\nu\big((X_n)_{n\geq 0}\big)$ la loi de la chaîne de Markov $(X_n)_{n\geq 0}$ avec $X_0 \sim \nu$, la propriété de Markov (5.38) s'écrit

$$\mathcal{L}_\nu\big((X_n)_{n\geq T}|\mathcal{F}_T\big) = \mathcal{L}_{X_T}\big((X_n)_{n\geq 0}\big), \tag{5.42}$$

ou pour $B \in \mathcal{F} = \sigma(\mathrm{Cyl})$:

$$\mathbb{P}_\nu\big((X_n)_{n\geq T} \in B|\mathcal{F}_T\big) = \mathbb{P}_{X_T}\big((X_n)_{n\geq 0} \in B\big).$$

Lorsque T est un temps d'arrêt, il s'agit de Markov fort ; lorsque $T = p$ est déterministe, il s'agit de Markov faible.

Avec des indicatrices, les propriétés de Markov s'écrivent encore

Corollaire 5.40 *Pour tout $A \in \mathcal{F}$, $(x_n)_{n\geq 0} \in E^{\mathbb{N}}$, $y \in E$, et T temps d'arrêt (ps fini) :*

$$\begin{aligned}
\mathbb{P}_\nu\big(\Theta_p X \in A\,|X_0 = x_0, \dots, X_p = x_p\big) &= \mathbb{P}_{x_n}(X \in A) \quad \textit{(Markov faible)}\\
\mathbb{P}_\nu\big(\Theta_T X \in A\,|X_0 = x_0, \dots, X_T = y\big) &= \mathbb{P}_y(X \in A) \quad \textit{(Markov fort).}
\end{aligned}$$

Démonstration : On prouve la formulation Markov fort, celle-ci contient la formulation Markov faible quand on prend le temps d'arrêt constant $T = p$.

En appliquant (5.38) avec la fonction $\mathcal{F}_T$-mesurable $F = \mathbf{1}_{\{X_0=x_0,\dots,X_T=y\}}$ et la fonction mesurable $G = \mathbf{1}_{\{X\in A\}}$, on a

$$\begin{aligned}
\mathbb{E}_\nu\big[\mathbf{1}_{\{X_0=x_0,\dots,X_T=y\}}\mathbf{1}_{\{X\circ\theta_T\in A\}}\big] &= \mathbb{E}_\nu\big[\mathbf{1}_{\{X_0=x_0,\dots,X_T=y\}}\ \mathbb{E}_{X_T}[\mathbf{1}_{\{X\in A\}}]\big]\\
&= \mathbb{E}_\nu\big[\mathbf{1}_{\{X_0=x_0,\dots,X_T=y\}}\ \mathbb{E}_y[\mathbf{1}_{\{X\in A\}}]\big]\\
&= \mathbb{E}_\nu\big[\mathbf{1}_{\{X_0=x_0,\dots,X_T=y\}}\big]\mathbb{E}_y\big[\mathbf{1}_{\{X\in A\}}\big]\\
&= \mathbb{P}_\nu(X_0 = x_0, \dots, X_T = y)\mathbb{P}_y(X \in A).
\end{aligned}$$

On a donc

$$\begin{aligned}
\mathbb{P}_\nu\big(\Theta_T X \in A\,|X_0 = x_0, \dots, X_T = y\big) &= \frac{\mathbb{P}_\nu(\Theta_T X \in A, X_0 = x_0, \dots, X_T = y)}{\mathbb{P}_\nu(X_0 = x_0, \dots, X_T = y)}\\
&= \frac{\mathbb{E}_\nu[\mathbf{1}_{\{X_0=x_0,\dots,X_T=y\}}\mathbf{1}_{\{X\circ\theta_T\in A\}}]}{\mathbb{P}_\nu(X_0 = x_0, \dots, X_T = y)}\\
&= \mathbb{P}_y(X \in A).
\end{aligned}$$

□

La propriété de Markov justifie également que, pour une chaîne de Markov, passé et futur sont indépendants sachant le présent :

Corollaire 5.41 (Passé, présent, futur) *Soit $n \geq 1$ et $A \in \mathcal{F}_n$ et $B \in \sigma(X_k\ : k \geq n)$ alors*

$$\mathbb{P}(A \cap B\,|X_n) = \mathbb{P}(A\,|X_n)\,\mathbb{P}(B\,|X_n).$$

De la même façon, si T est un temps d'arrêt $\mathbb{P}_x$-ps fini. Alors pour $A \in \mathcal{F}_T$ et $B \in \{\Theta_T^{-1}(A) : A \in \mathcal{F}\}$, on a

$$\mathbb{P}(A \cap B\,|X_T) = \mathbb{P}(A\,|X_T)\,\mathbb{P}(B\,|X_T).$$

Remarque 5.42 Comme Θ_T est mesurable, $\{\Theta_T^{-1}(A) : A \in \mathcal{F}\}$ est une tribu qui contient les évènements réalisés après T. C'est la façon correcte d'écrire $\sigma(X_k : k \geq T)$ puisque les évènements typiques en sont $\{(X_1, \dots, X_n) \circ \Theta_T \in B\} = \{(X_{T+1}, \dots, X_{T+n}) \in B\}$ pour tout $B \in \sigma(\mathrm{Cyl})$.

Démonstration : Soit $A \in \mathcal{F}_n$-mesurable et $B \in \sigma(X_k : k \geq n)$. On peut écrire $B = Theta_n^{-1}(B')$. On a alors

$$\begin{aligned}
\mathbb{P}_x(A \cap B|X_n) &= \mathbb{P}_x(A \cap \Theta_n^{-1}(B')|X_n) = \mathbb{E}_x\big[\mathbf{1}_A \mathbf{1}_{B'} \circ \Theta_n|X_n\big] \\
&= \mathbb{E}_x\big[\mathbb{E}_x[\mathbf{1}_A \mathbf{1}_{B'} \circ \Theta_n|\mathcal{F}_n]\,|X_n\big] \\
&\quad \text{(conditionnement en cascade du Th. 2.12)} \\
&= \mathbb{E}_x\big[\mathbf{1}_A\, \mathbb{E}_x[\mathbf{1}_{B'} \circ \Theta_n|\mathcal{F}_n]\,|X_n\big] \\
&= \mathbb{E}_x\big[\mathbb{E}_x[\mathbf{1}_A\, \mathbb{E}_{X_n}[\mathbf{1}_{B'}]\,|\mathcal{F}_n]\,|X_n\big] \\
&\quad \text{(par la propriété de Markov (5.31))} \\
&= \mathbb{E}_x[\mathbf{1}_A|X_n]\, \mathbb{E}_{X_n}[\mathbf{1}_{B'}] \\
&= \mathbb{P}_x(A|X_n)\, \mathbb{P}_{X_n}(B'). \qquad (5.43)
\end{aligned}$$

Avec $A = \Omega$, (5.43) donne $\mathbb{P}(B|X_n) = \mathbb{P}_{X_n}(B')$, ce qu'en ré-injectant dans (5.43) donne

$$\mathbb{P}_x(A \cap B|X_n) = \mathbb{P}_x(A|X_n)\, \mathbb{P}(B|X_n)$$

pour tout $A \in \mathcal{F}_n$ et $B \in \sigma(X_k : k \geq n)$, ce qui prouve la première partie du Corollaire 5.41. La deuxième partie se prouve de la même façon avec la propriété de Markov forte (5.38) en (5.43). □

Chapitre 6

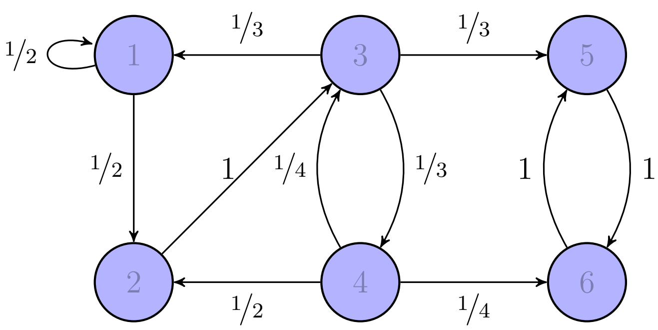
Récurrence et transience

Introduction et notations

Exemple 6.1 Sur l'espace $E = \{1, 2, 3, 4, 5, 6\}$, on considère une chaîne de Markov de matrice de transition

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 1/3 & 0 \\ 0 & 1/2 & 1/4 & 0 & 0 & 1/4 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Le graphe associé est alors

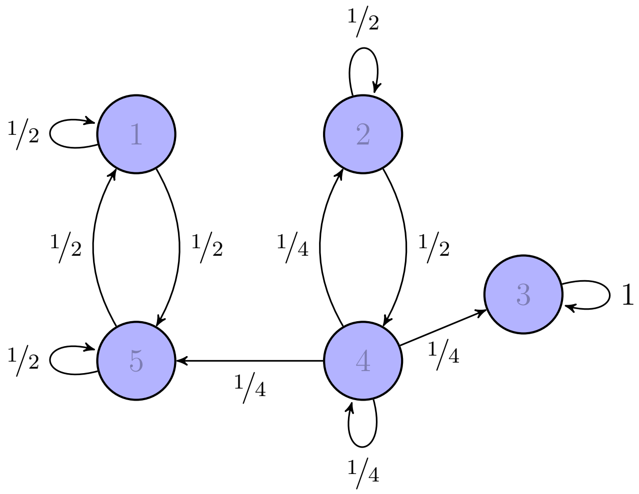


Les états $\{1, 2, 3, 4\}$ semblent visités un nombre fini de fois $\mathbb{P}_1$-ps. Au contraire, $\{5, 6\}$ sont visités une infinité de fois $\mathbb{P}_1$-ps.

Exemple 6.2 Sur l'espace $E = \{1, 2, 3, 4, 5\}$, on considère maintenant une chaîne de Markov de matrice de transition

$$P = \begin{pmatrix} 1/2 & 0 & 0 & 0 & 1/2 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1/4 & 1/4 & 1/4 & 1/4 \\ 1/2 & 0 & 0 & 0 & 1/2 \end{pmatrix}.$$

Le graphe associé est alors



Cette fois, $\mathbb{P}_2$-ps les états 2 et 4 semblent visités un nombre finis de fois, alors que $\{1,5\}$ et $\{3\}$ sont visités une infinité de fois mais 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

On considère $(X_n)_{n\geq 0}$ une chaîne de Markov d'espace d'états E et de matrice stochastique P. Si nécessaire, on travaille avec la chaîne canonique construite dans le Théorème 5.33. Dans la suite, on note $\mathbb{E}_x$ l'espérance par rapport à $\mathbb{P}_x$, c'est à dire on suppose que la chaîne part de x (ie. $\mu_0 = \delta_x$). Pour $y \in E$, avec la convention $\min \emptyset = +\infty$, on note

$$\begin{aligned} T_y &= \min\big(n \geq 0 : X_n = y\big) \quad \text{(temps d'atteinte de } y\text{)} \\ N(y) &= \sum_{k=0}^{+\infty} \mathbf{1}_{\{X_k = y\}} \quad \text{(nombre de visites en } y\text{)}. \end{aligned}$$

On note également

$$\begin{aligned} \widetilde{T}_y &= \min\big(n > 0 : X_n = y\big) \quad \text{(temps d'atteinte de } y\text{)} \\ \widetilde{N}(y) &= \sum_{k=1}^{+\infty} \mathbf{1}_{\{X_k = y\}} \quad \text{(nombre de visites de } y \text{ après le départ).} \end{aligned}$$

Les variables aléatoires T_y et $\widetilde{T}_y$ sont des temps d'arrêt pour la filtration canonique associée à la chaîne de Markov $(X_n)_{n\geq 0}$, cf. 2 dans l'Exemple 3.9.

On a les liens suivants selon le point de départ de la chaîne :

— sous $\mathbb{P}_x$, avec $x \neq y$: $\widetilde{T}_y = T_y$ est le temps d'atteinte de y et $\widetilde{N}(y) = N(y)$;

— sous $\mathbb{P}_y$: $\widetilde{T}_y > 0 = T(y)$ est le temps de retour de la chaîne en y et $\widetilde{N}(y) = N(y) - 1$;

On note également $\rho_{x,y} = \mathbb{P}_x\big(\widetilde{T}_y < +\infty\big)$ la probabilité que partant de $x \in E$ la chaîne puisse arriver en temps fini en $y \in E$. En particulier, $\rho_{x,x}$ est la probabilité que la chaîne partant de x finisse par y revenir.

Enfin, par récurrence, on définit les temps de retours successifs en $y \in E$ avec $T_y^{(0)} = 0$ (convention) et pour $k \geq 1$:

$$\begin{aligned} T_y^{(k)} &= \inf\big(n > T_y^{(k-1)} : X_n = y\big) && (6.1)\\ &= T_y^{(k-1)} + T_y^{(1)} \circ \Theta_{T_y^{(k-1)}}, && (6.2) \end{aligned}$$

où (6.2) vient de

$$\begin{aligned} T_y^{(k)} &= T_y^{(k)} + \inf\big(j > 0 : X_{j+T_y^{(k-1)}} = y\big) = T_y^{(k-1)} + \inf\big(j > 0 : X_j \circ \Theta_{T_y^{(k-1)}} = y\big)\\ &= T_y^{(k-1)} + \inf\big(j > 0 : X_j = y\big) \circ T_y^{(k-1)} = T_y^{(k-1)} + T_y^{(1)} \circ \Theta_{T_y^{(k-1)}}. \end{aligned}$$

— Observer que, par (6.1) ou par (6.2), lorsque $T_y^{(k-1)} = +\infty$ alors $T_y^{(k)} = +\infty$ aussi.

— Lorsqu' il est fini, l'intervalle de temps $[T_y^{(k-1)}, T_y^{(k)}]$ s'appelle une **excursion** de la chaîne entre deux visites en y.

— Les variables aléatoires $T_y^{(k)}$ sont des temps d'arrêt puisque pour tout $p \geq 0$:

$$\begin{aligned} \{T_y^{(k)} \leq p\} &= \{\text{la chaîne est passée } n \text{ fois en } y \text{ avant la date } p\}\\ &\in \sigma(X_1, \dots, X_p) = \mathcal{F}_p. \end{aligned}$$

6.1 États récurrents et transitoires

On distingue les états selon la propension qu'a la chaîne d'y revenir :

Définition 6.3 (Récurrence et transience) *Soit $(X_n)_{n\geq 0}$ une chaîne de Markov.*

— *Un état $x \in E$ est dit récurrent si $\rho_{x,x} = \mathbb{P}_x\big(\widetilde{T}_x < +\infty\big) = 1$.*

— *Un état $x \in E$ est dit transitoire (transient) si $\rho_{x,x} = \mathbb{P}_x\big(\widetilde{T}_x < +\infty\big) < 1$.*

En particulier, on appelle état absorbant tout état $x \in E$ tel que $P(x, x) = 1$ (Déf. **??**). Un tel état est récurrent puisque $\rho_{x,x} = \mathbb{P}_x\big(\widetilde{T}_x = 1\big) = P(x, x) = 1$.

Nombre de passages

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, avec la propriété de Markov forte, une infinité de fois. On formalise cette intuition dans la proposition suivante qui montre que le nombre de passages en un état x dépend fondamentalement de sa nature récurrente ou transitoire.

Proposition 6.4 (Nombre de passages en un état) *Pour tout état $x \in E$, on a l'alternative suivante :*

(1) Si x est récurrent alors $N(x) = +\infty$ $\mathbb{P}_x$-ps ($N(x) \overset{\mathbb{P}_x}{\sim} \delta_{+\infty}$).

(2) Si x est transitoire alors

$$N(x) \overset{\mathbb{P}_x}{\sim} \mathcal{G}(1-\rho_{x,x}). \tag{6.3}$$

En particulier, $N(x) < +\infty$ $\mathbb{P}_x$-ps et $\mathbb{E}_x[N(x)] = \frac{1}{1-\rho_{x,x}}$.

Démonstration : D'abord on a $\mathbb{P}_x(N(x) \geq 1) = 1$. Ensuite, on observe que sous $\mathbb{P}_x$, lorsque $\widetilde{T}_x < +\infty$, on a :

$$\begin{aligned} N(x) &= \sum_{k\geq 0} \mathbf{1}_{\{X_k=x\}} = 1 + \sum_{k\geq \widetilde{T}_x} \mathbf{1}_{\{X_k=x\}} = 1 + \sum_{j\geq 0} \mathbf{1}_{\{X_{j+\widetilde{T}_x}=x\}} \\ &= 1 + \sum_{j\geq 0} \mathbf{1}_{\{X_j\circ\Theta_{\widetilde{T}_x}=x\}} = 1 + \Big(\sum_{j\geq 0} \mathbf{1}_{\{X_j=x\}}\Big) \circ \Theta_{\widetilde{T}_x} = 1 + N(x)\circ\Theta_{\widetilde{T}_x}. \end{aligned} \tag{6.4}$$

Pour $k \geq 1$, on a donc $\mathbf{1}_{\{N(x)\geq k+1\}} = \mathbf{1}_{\{\widetilde{T}_x<+\infty\}} \mathbf{1}_{\{N(x)\geq k\}} \circ \Theta_{\widetilde{T}_x}$ $\mathbb{P}_x$-ps, et :

$$\begin{aligned} \mathbb{P}_x(N(x) \geq k+1) &= \mathbb{E}_x\big[\mathbf{1}_{\{N(x)\geq k+1\}}\big] = \mathbb{E}_x\big[\mathbf{1}_{\{\widetilde{T}_x<+\infty\}} \mathbf{1}_{\{N(x)\geq k\}} \circ \Theta_{\widetilde{T}_x}\big] \\ &= \mathbb{E}_x\big[\mathbf{1}_{\{\widetilde{T}_x<+\infty\}} \mathbb{E}_{X_{\widetilde{T}_x}}[\mathbf{1}_{\{N(x)\geq k\}}]\big] = \mathbb{P}_x\big(\widetilde{T}_x < +\infty\big)\mathbb{E}_x\big[\mathbf{1}_{\{N(x)\geq k\}}\big] \\ &\qquad \text{(propriété de Markov forte sous la forme du Corollaire 5.38)} \\ &= \rho_{x,x}\, \mathbb{P}_x(N(x) \geq k). \end{aligned}$$

Comme $\mathbb{P}_x(N(x) \geq 1) = 1$, on déduit d'une récurrence immédiate que

$$\mathbb{P}_x(N(x) \geq k) = \rho_{x,x}^{k-1}. \tag{6.5}$$

Dès lors,

— (1) lorsque x est récurrent alors $\rho_{x,x} = 1$ et en faisant $k \to +\infty$ par convergence monotone, on obtient

$$\mathbb{P}_x(N(x) = +\infty) = \lim_{n\to+\infty} \mathbb{P}_x(N(x) \geq k) = 1,$$

ie. $N(x) = +\infty$ $\mathbb{P}_x$-ps ;

— (2) lorsque x est transitoire alors $\rho_{x,x} < 1$ et (6.5) donne pour tout $k \geq 0$

$$\begin{aligned} \mathbb{P}_x(N(x) = k) &= \mathbb{P}_x(N(x) \geq k) - \mathbb{P}_x(N(x) \geq k+1) \\ &= \rho_{x,x}^{k-1} - \rho_{x,x}^k = \rho_{x,x}^{k-1}(1 - \rho_{x,x}) \end{aligned} \tag{6.6}$$

et il vient $N(x) \overset{\mathbb{P}_x}{\sim} \mathcal{G}(1 - \rho_{x,x})$. A fortiori, on a $\mathbb{E}_x[N(x)] = 1/(1 - \rho_{x,x})$ et $N(x) < +\infty$ $\mathbb{P}_x$-ps.

□

Lorsque la chaîne part d'un état x différent de l'état y où on considère les visites de la chaîne, la Prop. 6.4 prend la forme suivante :

Proposition 6.5 (Nombre de passages en y partant de $x \neq y$) *Soit x, y deux états avec $x \neq y$. Sous $\mathbb{P}_x$ (ie. lorsque la chaîne part de x),*

(1) Si y est récurrent ($\rho_{y,y} = 1$) 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 ($N(y) = +\infty$) :

$$N(y) \overset{\mathbb{P}_x}{\sim} (1 - \rho_{x,y})\delta_0 + \rho_{x,y}\delta_{+\infty}, \tag{6.7}$$

on a $\mathbb{P}_x(N(y) = +\infty) = \rho_{x,y}$.

(2) Si y est transitoire ($\rho_{y,y} < 1$) alors le nombre de passages en y est de loi

$$N(y) \overset{\mathbb{P}_x}{\sim} (1 - \rho_{x,y})\,\delta_0 + \rho_{x,y}\,\mathcal{G}(1 - \rho_{y,y}), \tag{6.8}$$

et on a $\mathbb{P}_x(N(y) < +\infty) = 1$.

Démonstration : On suppose que la chaîne part de $x \neq y$. On a $\{\widetilde{N}(y) \geq 1\} = \{\widetilde{T}_y < +\infty\}$ et

$$\mathbb{P}_x\big(\widetilde{N}(y) \geq 1\big) = \mathbb{P}_x\big(\widetilde{T}_y < +\infty\big) = \rho_{x,y}.$$

Étant donné $m_1, m_2 \geq 0$, la probabilité que la chaîne partant de x visite y la première fois à la date m_1 et n'y revienne qu'en date $m_1 + m_2$ est

$$\begin{aligned} &\mathbb{P}_x\big(T_y^{(1)} = m_1, T_y^{(2)} = m_1 + m_2\big) \\ &= \mathbb{E}_x\big[\mathbf{1}_{\{T_y^{(1)} = m_1\}}\,\mathbf{1}_{\{T_y^{(2)} = m_1 + m_2\}}\big] \\ &= \mathbb{E}_x\Big[\mathbf{1}_{\{T_y^{(1)} = m_1\}}\,\mathbf{1}_{\big\{T_y^{(1)} \circ \Theta_{T_y^{(1)}} = m_2\big\}}\Big] \quad \text{(car par (6.2) : } T_y^{(2)} = T_y^{(1)} + T_y^{(1)} \circ \Theta_{T_y^{(1)}}) \\ &= \mathbb{E}_x\Big[\mathbf{1}_{\{T_y^{(1)} = m_1\}}\mathbb{E}_{X_{T_y^{(1)}}}\big[\mathbf{1}_{\{T_y^{(1)} = m_2\}}\big]\Big] \quad \text{(par Markov fort avec le Corollaire 5.38)} \\ &= \mathbb{P}_x\big(\widetilde{T}_y = m_1\big)\,\mathbb{P}_y\big(\widetilde{T}_y = m_2\big) \quad \text{(car } X_{T_y^{(1)}} = y \text{ et donc } \mathbb{E}_{X_{T_y^{(1)}}} = \mathbb{E}_y). \end{aligned}$$

On a donc

$\mathbb{P}_x(N(y) \geq 2)$

$$
\begin{aligned}
&= \sum_{m_1,m_2=1}^{+\infty} \mathbb{P}_x\big(X \text{ visite } y \text{ la première fois à la date } m_1 \text{ et n'y revient qu'en date } m_1+m_2\big)\\
&= \sum_{m_1,m_2=1}^{+\infty} \mathbb{P}_x\big(T_y^{(1)}=m_1, T_y^{(2)}=m_1+m_2\big) = \sum_{m_1=1}^{+\infty}\sum_{m_2=1}^{+\infty} \mathbb{P}_x\big(\widetilde{T}_y=m_1\big)\,\mathbb{P}_y\big(\widetilde{T}_y=m_2\big)\\
&= \left(\sum_{m_1=1}^{+\infty} \mathbb{P}_x\big(\widetilde{T}_y=m_1\big)\right)\left(\sum_{m_2=1}^{+\infty} \mathbb{P}_y\big(\widetilde{T}_y=m_2\big)\right)\\
&= \mathbb{P}_x\big(\widetilde{T}_y<+\infty\big)\,\mathbb{P}_y\big(\widetilde{T}_y<+\infty\big) = \rho_{x,y}\,\rho_{y,y}.
\end{aligned}
$$

Plus généralement, un raisonnement analogue montre que pour $k \geq 1$:

$$\mathbb{P}_x(N(y) \geq k) = \rho_{x,y}\,\rho_{y,y}^{k-1}. \tag{6.9}$$

En effet, pour $m_1, \dots, m_k \geq 1$, on a

$$
\begin{aligned}
&\mathbb{P}_x\Big(\bigcap_{1\leq j\leq k} \{T_y^{(j)}=m_1+\dots+m_j\}\Big) = \mathbb{E}_x\left[\prod_{j=1}^{k} \mathbf{1}_{\{T_y^{(j)}=m_1+\dots+m_j\}}\right]\\
&= \mathbb{E}_x\left[\left(\prod_{j=1}^{k-1} \mathbf{1}_{\{T_y^{(j)}=m_1+\dots+m_j\}}\right)\mathbf{1}_{\{T_y^{(k)}=m_1+\dots+m_k\}}\right]\\
&= \mathbb{E}_x\left[\left(\prod_{j=1}^{k-1} \mathbf{1}_{\{T_y^{(j)}=m_1+\dots+m_j\}}\right)\mathbf{1}_{\{T_y^{(1)}=m_k\}}\circ\Theta_{T_y^{(k-1)}}\right]
\end{aligned}
$$

(car par (6.2) : $T_y^{(k)} = T_y^{(k-1)} + T_y^{(1)} \circ \Theta_{T_y^{(k-1)}}$)

$$
= \mathbb{E}_x\Big[\underbrace{\left(\prod_{j=1}^{k-1} \mathbf{1}_{\{T_y^{(j)}=m_1+\dots+m_j\}}\right)}_{\mathcal{F}_{T_y^{(k-1)}}\text{-mesurable}}\mathbb{E}_x\big[\mathbf{1}_{\{T_y^{(1)}=m_k\}}\circ\Theta_{T_y^{(k-1)}}\,\big|\mathcal{F}_{T_y^{(k-1)}}\big]\Big]
$$

(car, pour $j \leq k-1$, $T_y^{(j)}$ est $\mathcal{F}_{T_y^{(k-1)}}$-mesurable)

$$
= \mathbb{E}_x\left[\left(\prod_{j=1}^{k-1} \mathbf{1}_{\{T_y^{(j)}=m_1+\dots+m_j\}}\right)\mathbb{E}_{X_{T_y^{(k-1)}}}\big[\mathbf{1}_{\{T_y^{(1)}=m_k\}}\big]\right]
$$

(par Markov fort avec le Corollaire 5.38)

$$
\begin{aligned}
&= \mathbb{E}_x\left[\left(\prod_{j=1}^{k-1} \mathbf{1}_{\{T_y^{(j)}=m_1+\dots+m_j\}}\right)\right]\mathbb{E}_y\big[\mathbf{1}_{\{T_y^{(1)}=m_k\}}\big]\\
&= \mathbb{P}_x\Big(\bigcap_{1\leq j\leq k-1} \{T_y^{(j)}=m_1+\dots+m_j\}\Big)\,\mathbb{P}_y(T_y^{(1)}=m_k)\\
&= \left(\mathbb{P}_x(T_y^{(1)}=m_1)\prod_{j=2}^{k-1}\mathbb{P}_y(T_y^{(1)}=m_j)\right)\mathbb{P}_y(T_y^{(1)}=m_k)
\end{aligned}
$$

$$= \mathbb{P}_x(T_y^{(1)} = m_1) \prod_{j=2}^k \mathbb{P}_y(T_y^{(1)} = m_j)$$

par hypothèse de récurrence pour $\mathbb{P}_x\Big(\bigcap_{1\le j\le k-1}\{T_y^{(j)} = m_1 + \cdots + m_j\}\Big)$. On a alors

$$\begin{aligned}
&\mathbb{P}_x(N(y) \geq k)\\
&= \sum_{\substack{m_j\geq 1\\ 1\le j\le k}} \mathbb{P}_x\big(X \text{ visite } y \text{ la } j\text{-ème fois à la date } m_1+\cdots+m_j,\ \forall j \in [\![1,k]\!]\big)\\
&= \sum_{\substack{m_j\geq 1\\ 1\le j\le k}} \mathbb{P}_x\Big(\bigcap_{1\le j\le k}\{T_y^{(j)} = m_1+\cdots+m_j\}\Big)\\
&= \sum_{\substack{m_j\geq 1\\ 1\le j\le k}} \mathbb{P}_x(T_y^{(1)} = m_1) \prod_{j=2}^k \mathbb{P}_y(T_y^{(1)} = m_j)\\
&= \left(\sum_{m_1\geq 1} \mathbb{P}_x(T_y^{(1)} = m_1)\right) \prod_{j=2}^k \left(\sum_{m_j\geq 1} \mathbb{P}_y(T_y^{(1)} = m_j)\right)\\
&= \mathbb{P}_x(T_y^{(1)} < +\infty) \prod_{j=2}^k \mathbb{P}_y(T_y^{(1)} < +\infty) = \rho_{x,y}\, \rho_{y,y}^{k-1},
\end{aligned}$$

ce qui établit (6.9).
Puis comme $\mathbb{P}_x(N(y) = k) = \mathbb{P}_x(N(y) \geq k) - \mathbb{P}_x(N(y) \geq k-1)$, on a aussi

$$\mathbb{P}_x(N(y) = k) = \rho_{x,y}\rho_{y,y}^{k-1}(1-\rho_{y,y}), \quad k \geq 1, \tag{6.10}$$

et

$$\mathbb{P}_x(N(y) = 0) = 1 - \mathbb{P}_x(N(y) \geq 1) = 1 - \rho_{x,y}. \tag{6.11}$$

Dans le cas où y est récurrent ($\rho_{y,y} = 1$), on déduit de (6.9) par convergence monotone que

$$\mathbb{P}_x(N(y) = +\infty) = \lim_{k\to+\infty} \mathbb{P}_x(N(y) \geq k) = \rho_{x,y},$$

ce qui établit **1)**. Puis **2)** découle de (6.10). □

Remarque 6.6 — Attention, dans le cas $x = y$, les formules (6.10) et (6.11) sont remplacées par (6.6) dans la Proposition 6.4. 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 formules (6.10) et (6.11) sont intuitivement claires avec la description heuristique suivante : pour que partant de x la chaîne visite m fois y, elle commence à aller de x à y (facteur $\rho_{x,y}$) puis visite y $m-1$ fois (facteur $\rho_{y,y}$ pour chaque visite donc globalement $\rho_{y,y}^{m-1}$) et n'y retourne plus (facteur $1-\rho_{y,y}$).

— Les Propositions 6.4 et 6.5 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, 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, elle y repasse une infinité de fois. Si elle part d'ailleurs 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

À la chaîne de Markov $(X_n)_{n\geq 0}$ de matrice stochastique P, on associe :

Définition 6.7 (Potentiel/fonction de Green) *Soit $x, y \in E$, on note $G(x,y)$ le nombre moyen de passages en y de la chaîne partant de x :*

$$G(x,y) = \mathbb{E}_x\big[N(y)\big] = \sum_{k=0}^{+\infty} \mathbb{P}_x(X_k = y) = \sum_{k=0}^{+\infty} P_k(x,y).$$

(Les égalités ci-dessus viennent du théorème de convergence monotone et de l'expression $N(y) = \sum_{k\geq 0} \mathbf{1}_{\{X_k=y\}}$.)

Théorème 6.8 (Nature et potentiel)

(1) Si y est un état transitoire ($\rho_{y,y} < 1$), alors le potentiel (fonction de Green) $G(x,y)$ est fini pour tout état x et vaut

$$G(x,y) = \begin{cases} \frac{\rho_{x,y}}{1-\rho_{y,y}} & si\ x \neq y, \\ \frac{1}{1-\rho_{y,y}} & si\ x = y. \end{cases}$$

(2) Si y est un état récurrent alors $G(y,y) = +\infty$ et
  — si $\rho_{x,y} = 0$ alors $G(x,y) = 0$;
  — si $\rho_{x,y} > 0$, $G(x,y) = +\infty$.

Démonstration : La preuve vient de la Proposition 6.5 et des lois (6.7)–(6.8) de $N(y)$ sous $\mathbb{P}_x$ lorsque $x \neq y$ ou (6.3) lorsque $x = y$.

D'abord, si $x \neq y$, alors par la Prop. 6.5 :

(a) Soit y un état transitoire, $G(x,y) = \mathbb{E}_x[N(y)]$ est l'espérance de $(1-\rho_{x,y})\,\delta_0 + \rho_{x,y}\,\mathcal{G}(1-\rho_{y,y})$ donc vaut $\rho_{x,y}/(1-\rho_{y,y})$.

(b) Si y est récurrent, $G(x,y) = \mathbb{E}_x[N(y)]$ est l'espérance de $(1-\rho_{x,y})\delta_0 + \rho_{x,y}\delta_{+\infty}$ donc vaut 0 si $\rho_{x,y} = 0$ et $+\infty$ sinon.

Ensuite, si $x = y$, alors par la Prop. 6.4 :

(a) Soit y un état transitoire, $G(x,y) = \mathbb{E}_x[N(y)]$ est l'espérance de $\mathcal{G}(1-\rho_{y,y})$ donc vaut $1/(1-\rho_{y,y})$.

(b) Si y est récurrent, $G(x, y) = +\infty$ puisque $N(y) = +\infty$.

□

De l'alternative du Théorème 6.8, on déduit immédiatement un critère de récurrence à l'aide du potentiel :

Corollaire 6.9 (Récurrence et potentiel) *Un état x est récurrent si et seulement si $G(x, x) = +\infty$.*

Corollaire 6.10 *En convenant que $0 \times (+\infty) = 0$, pour $x \neq y$, on a :*

$$G(x, y) = \rho_{x,y}\, G(y, y). \tag{6.12}$$

La preuve du Corollaire 6.10 s'obtient des expressions explicites de $G(x, y)$ dans le Théorème 6.8. On peut aussi le prouver directement à partir de la propriété de Markov forte :

Démonstration : Soit $x \neq y$. Sous $\mathbb{P}_x$, lorsque $T_y = +\infty$ on a $N(y) = 0$ et lorsque $T_y < +\infty$

$$\begin{aligned}
N(y) &= \#(n \geq 0 \,:\, X_n = y) = \#(n \geq T_y \,:\, X_n = y)\\
&= \#\big(k \geq 0 \,:\, X_{k+T_y} = y\big) = \#\big(k \geq 0 \,:\, X_k \circ \Theta_{T_y} = y\big)\\
&= \#\big(k \geq 0 \,:\, X_k = y\big) \circ \Theta_{T_y} = N(y) \circ \Theta_{T_y}.
\end{aligned}$$

On a donc $N(y) = N(y) \circ \Theta_{T_y}$ $\mathbb{P}_x$-ps et par la propriété de Markov forte (5.38) (sous la forme du Corollaire 5.38), on a alors

$$\begin{aligned}
\mathbb{E}_x[N(y)] &= \mathbb{E}_x\big[\mathbf{1}_{\{T_y<+\infty\}}\, (N(y) \circ \Theta_{T_y})\big] = \mathbb{E}_x\big[\mathbf{1}_{\{T_y<+\infty\}}\, \mathbb{E}_x\big[(N(y) \circ \Theta_{T_y})\,\big|\mathcal{F}_{\Theta_{T_y}}\big]\big]\\
&= \mathbb{E}_x\big[\mathbf{1}_{\{T_y<+\infty\}}\, \mathbb{E}_{X_{T_y}}[N(y)]\big] = \mathbb{E}_x\big[\mathbf{1}_{\{T_y<+\infty\}}\, \mathbb{E}_y[N(y)]\big]\\
&= \mathbb{P}_x(T_y < +\infty)\, \mathbb{E}_y[N(y)]
\end{aligned}$$

ie. $G(x, y) = \rho_{x,y}\, G(y, y)$. □

On précise la notion de récurrence d'un état selon la durée moyenne d'un retour en cet état.

Définition 6.11 (Récurrence nulle et positive) *Un état x récurrent est dit récurrent positif si $m_x = \mathbb{E}_x\big[\widetilde{T}_x\big] < +\infty$. Il est dit récurrent nul si $m_x = \mathbb{E}_x\big[\widetilde{T}_x\big] = +\infty$.*

Ainsi

- si x est récurrent positif : $\widetilde{T}_x < +\infty$ $\mathbb{P}_x$-ps et $m_x = \mathbb{E}_x\big[\widetilde{T}_x\big] < +\infty$;
- si x est récurrent nul : $\widetilde{T}_x < +\infty$ $\mathbb{P}_x$-ps mais $m_x = \mathbb{E}_x\big[\widetilde{T}_x\big] = +\infty$;
- si x est transitoire : $\widetilde{T}_x = +\infty$ avec probabilité $\mathbb{P}_x$ positive et a fortiori $m_x = \mathbb{E}_x\big[\widetilde{T}_x\big] = +\infty$.

Définition 6.12 (Chaînes récurrente et transitoire) *Une chaîne est dite :*

— *transitoire si tous ses états sont transitoires ;*
— *récurrente si tous ses états sont récurrents ;*
— *récurrente positive si tous ses états sont récurrents positifs ;*
— *récurrente nulle si tous ses états sont récurrents nuls.*

Remarque 6.13 (1) Noter que si y est un état transitoire alors pour tout $x \in E$,

$$\lim_{n\to+\infty} P^n(x,y) = 0.$$

En effet cela découle de la convergence de la série $G(x,y) = \sum_{n=0}^{+\infty} P^n(x,y) < +\infty$.

(2) Si une chaîne $(X_n)_{n\geq 0}$ a un nombre fini d'états, alors nécessairement il y a au moins un état récurrent et la chaîne ne peut pas être transitoire. En effet, si tous les états étaient transitoires alors on aurait l'égalité absurde suivante :

$$0 = \sum_{y\in E} \lim_{n\to+\infty} P^n(x,y) = \lim_{n\to+\infty} \sum_{y\in E} P^n(x,y) = \lim_{n\to+\infty} \mathbb{P}_x(X_n \in E) = 1.$$

La première égalité vient de 1) ci-dessus, la deuxième du fait que la somme $\sum_{y\in E}$ est finie.

Excursions

On rappelle que les $T_y^{(k)}$, $k \geq 0$, désignent les dates de retours successifs de la chaîne en y et qu'ils sont définis en (6.1) et satisfont (6.2).

Proposition 6.14 (Indépendance des excursions) *Sachant $T_y^{(n)} < +\infty$ (qu'on suppose non négligeable), les variables aléatoires $\Delta_y^{(k)} = T_y^{(k)} - T_y^{(k-1)}$, $1 \leq k \leq n$, sont iid sous $\mathbb{P}_y$.*

Démonstration : Il s'agit de montrer pour des fonctions g_i, $1 \leq i \leq n$, mesurables bornées sur $\mathbb{R}_+$, on a :

$$\mathbb{E}_y\left[\prod_{i=1}^n g_i\big(\Delta_y^{(i)}\big)\,\Big|T_y^{(n)} < +\infty\right] = \prod_{i=1}^n \mathbb{E}_y\left[g_i\big(\Delta_y^{(1)}\big)\,\Big|T_y^{(n)} < +\infty\right]. \tag{6.13}$$

Comme $T_y^{(n)} = \sum_{i=1}^n \Delta_y^{(i)}$, on a

$$\{T_y^{(n)} < +\infty\} = \bigcap_{i=1}^n \{\Delta_y^{(i)} < +\infty\}, \tag{6.14}$$

et on commence par montrer que pour toutes fonctions g_i, $1 \leq i \leq n$, mesurables bornées sur $\mathbb{R}_+$:

$$\mathbb{E}_y\left[\prod_{i=1}^n \Big(g_i\big(\Delta_y^{(i)}\big)\mathbf{1}_{\{\Delta_y^{(i)}<+\infty\}}\Big)\right] = \prod_{i=1}^n \mathbb{E}_y\big[g_i\big(\Delta_y^{(1)}\big)\mathbf{1}_{\{\Delta_y^{(1)}<+\infty\}}\big]. \tag{6.15}$$

On procède par récurrence sur $n \geq 1$. Pour $n = 1$, l'égalité (6.15) est immédiate. On suppose alors (6.15) établie pour $n-1$ fixé et on la prouve pour n. Pour cela, on observe que

— les variables aléatoires $\Delta_y^{(1)}, \dots, \Delta_y^{(n-1)}$ sont $\mathcal{F}_{T_y^{(n-1)}}$-mesurables,
— $\Theta_{T_y^{(n-1)}}$ est indépendante de $\mathcal{F}_{T_y^{(n-1)}}$ et de loi $\mathbb{P}_y$ (propriété de Markov forte, Corollaire 5.38),
— $\Delta_y^{(n)} = \Delta_y^{(1)} \circ \Theta_{T_y^{(n-1)}}$; en effet par (6.2), on a $T_y^{(n)} = T_y^{(n-1)} + T_y^{(1)} \circ \Theta_{T_y^{(n-1)}}$ et on a donc
$$\Delta_y^{(n)} = T_y^{(n)} - T_y^{(n-1)} = T_y^{(1)} \circ \Theta_{T_y^{(n-1)}} = \Delta_y^{(1)} \circ \Theta_{T_y^{(n-1)}}.$$

En utilisant dans la 3-ème égalité la propriété de Markov fort (Corollaire 5.38), on a :

$$\begin{aligned}
&\mathbb{E}_y\left[\prod_{i=1}^{n}\left(g_i(\Delta_y^{(i)})\mathbf{1}_{\{\Delta_y^{(i)}<+\infty\}}\right)\right]\\
&= \mathbb{E}_y\left[\prod_{i=1}^{n-1}\left(g_i(\Delta_y^{(i)})\mathbf{1}_{\{\Delta_y^{(i)}<+\infty\}}\right) g_n\big(\Delta_y^{(1)}\circ\Theta_{T_y^{(n-1)}}\big)\mathbf{1}_{\left\{\Delta_y^{(1)}\circ\Theta_{T_y^{(n-1)}}<+\infty\right\}}\right]\\
&= \mathbb{E}_y\left[\prod_{i=1}^{n-1}\left(g_i(\Delta_y^{(i)})\mathbf{1}_{\{\Delta_y^{(i)}<+\infty\}}\right) \mathbb{E}_y\left[g_n\big(\Delta_y^{(1)}\circ\Theta_{T_y^{(n-1)}}\big)\mathbf{1}_{\left\{\Delta_y^{(1)}\circ\Theta_{T_y^{(n-1)}}<+\infty\right\}}\Big|\mathcal{F}_{T_y^{(n-1)}}\right]\right]\\
&= \mathbb{E}_y\left[\prod_{i=1}^{n-1}\left(g_i(\Delta_y^{(i)})\mathbf{1}_{\{\Delta_y^{(i)}<+\infty\}}\right) \mathbb{E}_{X_{T_y^{(n-1)}}}\left[g_n\big(\Delta_y^{(1)}\big)\mathbf{1}_{\left\{\Delta_y^{(1)}<+\infty\right\}}\right]\right]\\
&= \mathbb{E}_y\left[\prod_{i=1}^{n-1}\left(g_i(\Delta_y^{(i)})\mathbf{1}_{\{\Delta_y^{(i)}<+\infty\}}\right)\right] \mathbb{E}_y\big[g_n(\Delta_y^{(1)})\mathbf{1}_{\{\Delta_y^{(1)}<+\infty\}}\big]\\
&= \prod_{i=1}^{n-1}\mathbb{E}_y\left[g_i(\Delta_y^{(1)})\mathbf{1}_{\{\Delta_y^{(1)}<+\infty\}}\right] \mathbb{E}_y\big[g_n(\Delta_y^{(1)})\mathbf{1}_{\{\Delta_y^{(1)}<+\infty\}}\big]
\end{aligned}$$

en utilisant l'hypothèse de récurrence, ce qui prouve (6.15) par récurrence.

On déduit deux cas particuliers de (6.15) :
— Lorsque toutes les g_i sont égales à 1, (6.15) devient

$$\mathbb{P}_y\big(T_y^{(n)} < +\infty\big) = \mathbb{P}_y\big(\Delta_y^{(1)} < +\infty\big)^n. \tag{6.16}$$

— Lorsque les g_j sont égales à 1 pour tous les $j \neq i$, (6.15) devient

$$\begin{aligned}
\mathbb{E}_y\left[g_i\big(\Delta_y^{(i)}\big)\mathbf{1}_{\{T_y^{(n)}<+\infty\}}\right] &= \mathbb{E}_y\left[g_i\big(\Delta_y^{(i)}\big)\prod_{j=1}^{n}\mathbf{1}_{\{\Delta_y^{(j)}<+\infty\}}\right]\\
&= \mathbb{E}_y\left[\left(g_i\big(\Delta_y^{(i)}\big)\mathbf{1}_{\{\Delta_y^{(i)}<+\infty\}}\right)\prod_{j\neq i}\mathbf{1}_{\{\Delta_y^{(j)}<+\infty\}}\right]\\
&= \mathbb{E}_y\big[g_i\big(\Delta_y^{(1)}\big)\mathbf{1}_{\{\Delta_y^{(1)}<+\infty\}}\big] \times \prod_{j\neq i}\mathbb{E}_y\big[\mathbf{1}_{\{\Delta_y^{(1)}<+\infty\}}\big]\\
&= \mathbb{E}_y\big[g_i\big(\Delta_y^{(1)}\big)\mathbf{1}_{\{\Delta_y^{(1)}<+\infty\}}\big] \times \mathbb{P}_y(\Delta_y^{(1)} < +\infty)^{n-1}
\end{aligned} \tag{6.17}$$

Finalement pour montrer (6.13), on écrit :

$$\begin{aligned}
\mathbb{E}_y\left[\prod_{i=1}^n g_i\big(\Delta_y^{(i)}\big)\,\middle|T_y^{(n)}<+\infty\right] &= \frac{\mathbb{E}_y\left[\left(\prod_{i=1}^n g_i\big(\Delta_y^{(i)}\big)\right)\mathbf{1}_{\{T_y^{(n)}<+\infty\}}\right]}{\mathbb{P}_y(T_y^{(n)}<+\infty)}\\
&= \frac{\mathbb{E}_y\left[\prod_{i=1}^n\left(g_i\big(\Delta_y^{(i)}\big)\mathbf{1}_{\{\Delta_y^{(i)}<+\infty\}}\right)\right]}{\mathbb{P}_y(T_y^{(n)}<+\infty)} \qquad \text{(en utilisant (6.14))}\\
&= \frac{\prod_{i=1}^n\mathbb{E}_y\left[g_i\big(\Delta_y^{(1)}\big)\mathbf{1}_{\{\Delta_y^{(1)}<+\infty\}}\right]}{\mathbb{P}_y(T_y^{(n)}<+\infty)} \qquad \text{(en utilisant (6.15))}\\
&= \frac{\prod_{i=1}^n\left(\mathbb{E}_y\left[g_i\big(\Delta_y^{(1)}\big)\mathbf{1}_{\{\Delta_y^{(1)}<+\infty\}}\right]\right)}{\mathbb{P}_y(T_y^{(n)}<+\infty)}\underbrace{\frac{\prod_{i=1}^n\mathbb{P}_y(\Delta_y^{(1)}<+\infty)^{n-1}}{\mathbb{P}_y(T_y^{(n)}<+\infty)^{n-1}}}_{=1}\\
&= \frac{\prod_{i=1}^n\left(\mathbb{E}_y\left[g_i\big(\Delta_y^{(1)}\big)\mathbf{1}_{\{\Delta_y^{(1)}<+\infty\}}\right]\mathbb{P}_y(\Delta_y^{(1)}<+\infty)^{n-1}\right)}{\mathbb{P}_y(T_y^{(n)}<+\infty)^n} \qquad (6.18)
\end{aligned}$$

en notant grâce à (6.16) que $\prod_{i=1}^n\mathbb{P}_y(\Delta_y^{(1)}<+\infty)^{n-1}=\mathbb{P}_y(T_y^{(n)}<+\infty)^{n-1}$. En utilisant (6.17), on a alors

$$\begin{aligned}
\mathbb{E}_y\left[\prod_{i=1}^n g_i\big(\Delta_y^{(i)}\big)\,\middle|T_y^{(n)}<+\infty\right] &= \frac{\prod_{i=1}^n\mathbb{E}_y\left[g_i\big(\Delta_y^{(i)}\big)\mathbf{1}_{\{T_y^{(n)}<+\infty\}}\right]}{\mathbb{P}_y(T_y^{(n)}<+\infty)^n}\\
&= \prod_{i=1}^n\mathbb{E}_y\left[g_i\big(\Delta_y^{(i)}\big)\,\middle|T_y^{(n)}<+\infty\right].
\end{aligned}$$

Finalement, on a obtenu (6.13) ce qui prouve la Proposition 6.14. □

Temps passé en un état

On note $N_n(y)$ (resp. $\widetilde{N}_n(y)$) la variable aléatoire qui indique le temps passé en un état $y\in E$ jusqu'au temps n en comptant la date initiale (resp. en ne la comptant pas) :

$$N_n(y)=\sum_{k=0}^n\mathbf{1}_{\{X_k=y\}} \quad \text{et} \quad \widetilde{N}_n(y)=\sum_{k=1}^n\mathbf{1}_{\{X_k=y\}}, \qquad (6.19)$$

et $G_n(x,y)$ le temps moyen passé en y jusqu'au temps n lorsque la chaîne part de x :

$$G_n(x,y)=\mathbb{E}_x[N_n(y)]=\sum_{k=0}^n P^k(x,y).$$

Théorème 6.15 (Proportion du temps de visite) *Pour tout état $y \in E$ et toute loi initiale ν, on a*

$$\lim_{n\to+\infty} \frac{N_n(y)}{n} = \frac{\mathbf{1}_{\{\widetilde{T}_y<+\infty\}}}{m_y} \quad \mathbb{P}_\nu\text{-ps.} \tag{6.20}$$

De plus, pour tout $x \in E$, on a

$$\lim_{n\to+\infty} \frac{G_n(x,y)}{n} = \frac{\rho_{x,y}}{m_y}. \tag{6.21}$$

Remarque 6.16 Ces résultats se justifient heuristiquement : dès que la chaîne atteint un état y récurrent, elle revient en y en moyenne en m_y étapes. Donc si $\widetilde{T}_y < +\infty$, et n est grand, la proportion d'étapes parmi les n premières où la chaîne est en y est d'ordre $1/m_y$. Le résultat (6.21) vient de (6.20) en prenant l'espérance. Par ailleurs si y est transitoire, il y a un nombre fini de visite en y donc la proportion du temps passé en y est asymptotiquement nulle, ce qu'on retrouve avec $m_y = +\infty$ dans ce cas.

Démonstration : *a*) On commence par considérer le cas y récurrent et une chaîne de Markov qui démarre de ce y. Avec probabilité 1, la chaîne revient en y une infinité de fois (Proposition 6.4). Comme y est récurrent, $T_y^{(n)}$ est fini $\mathbb{P}_y$-ps pour tout $n \geq 0$ ($T_y^{(n)} = 0$), et la Proposition 6.14 donne que les variables aléatoires $\Delta_y^{(k)} = T_y^{(k)} - T_y^{(k-1)}$, $k \geq 1$, (durée entre la $(k-1)$-ème visite et la k-ème visite en y) sont *iid*. Comme $T_y^{(n)} = \Delta_y^{(1)} + \cdots + \Delta_y^{(n)}$, la loi des grands nombres (LGN) donne alors, $\mathbb{P}_y$-ps

$$\lim_{n\to+\infty} \frac{T_y^{(n)}}{n} = \lim_{n\to+\infty} \frac{\Delta_y^{(1)} + \cdots + \Delta_y^{(n)}}{n} = m_y. \tag{6.22}$$

En effet,

— si $m_y < +\infty$, alors $\Delta_y^{(1)} \in L^1$ et (6.22) s'obtient directement par la LGN ;

— si $m_y = +\infty$, on commence par appliquer la LGN aux variables aléatoires *iid* $\big(\Delta_y^{(i)} \wedge a\big) \in L^1$ où $a > 0$ quelconque est préalablement fixé :

$$\lim_{n\to+\infty} \frac{\big(\Delta_y^{(1)} \wedge a\big) + \cdots + \big(\Delta_y^{(n)} \wedge a\big)}{n} = \mathbb{E}_y\big[\Delta_y^{(1)} \wedge a\big].$$

Comme $\Delta_y^{(1)} \geq \Delta_y^{(1)} \wedge a$, on a

$$\begin{aligned}
&\liminf_{n\to+\infty} \frac{\Delta_y^{(1)} + \cdots + \Delta_y^{(n)}}{n} \\
&\quad \geq \liminf_{n\to+\infty} \frac{\big(\Delta_y^{(1)} \wedge a\big) + \cdots + \big(\Delta_y^{(n)} \wedge a\big)}{n} \\
&\quad = \mathbb{E}_y[\Delta_y^{(1)} \wedge a].
\end{aligned}$$

Mais comme par convergence monotone $\mathbb{E}_y[\Delta_y^{(1)} \wedge a] \nearrow \mathbb{E}_y[\Delta_y^{(1)}] = m_y = +\infty$ quand $a \to +\infty$, on a

$$\lim_{n\to+\infty} \frac{\Delta_y^{(1)} + \cdots + \Delta_y^{(n)}}{n} = +\infty,$$

ce qui correspond à (6.22) avec $m_y = +\infty$.

Par définition de $T_y^{(k)}$ et $\widetilde{N}_n(y)$, on a

$$T_y^{(\widetilde{N}_n(y))} \leq n < T_y^{(\widetilde{N}_n(y)+1)}$$

et donc pour n assez grand (pour assurer $\widetilde{N}_n(y) \geq 1$)

$$\frac{T_y^{(\widetilde{N}_n(y))}}{\widetilde{N}_n(y)} \leq \frac{n}{\widetilde{N}_n(y)} < \frac{T_y^{(\widetilde{N}_n(y)+1)}}{\widetilde{N}_n(y)}. \tag{6.23}$$

Mais comme $\widetilde{N}_n(y) \to +\infty$ $\mathbb{P}_y$-ps et $(\widetilde{N}_n(y)+1)/\widetilde{N}_n(y) \to 1$ quand $n \to +\infty$, la LGN (6.22) donne aussi :

$$\frac{T_y^{(\widetilde{N}_n(y))}}{\widetilde{N}_n(y)} \xrightarrow[n\to+\infty]{\mathbb{P}_y-\text{ps}} m_y \quad \text{et} \quad \frac{T_y^{(\widetilde{N}_n(y)+1)}}{\widetilde{N}_n(y)} \xrightarrow[n\to+\infty]{\mathbb{P}_y-\text{ps}} m_y. \tag{6.24}$$

Le théorème des gendarmes, (6.23) et (6.24) assurent alors que $\mathbb{P}_y$-ps :

$$\lim_{n\to+\infty} \frac{n}{\widetilde{N}_n(y)} = m_y,$$

ce qui prouve (6.20) dans ce cas (départ de la chaîne de y, récurrent) puisque $\widetilde{N}_n(y)/N_n(y) \to 1$ quand $n \to +\infty$.

b) On suppose maintenant que la chaîne part de $x \neq y$. Dans ce cas, la chaîne peut ne jamais rejoindre y. Cependant si elle rejoint y, l'argument précédent s'applique et se réécrit alors

$$\lim_{n\to+\infty} \frac{N_n(y)}{n} = \frac{\mathbf{1}_{\{\widetilde{T}_y<+\infty\}}}{m_y} \quad \mathbb{P}_x\text{-ps.}$$

On a donc (6.20) $\mathbb{P}_x$-presque sûrement pour tout $x \in E$ lorsque y est récurrent.

c) On considère y transitoire et la chaîne part d'un état x quelconque. Par la Prop. 6.4 et la Prop. 6.5, on a $\lim_{n\to+\infty} N_n(y) = N(y) < +\infty$ $\mathbb{P}_x$-presque sûrement pour tout $x \in E$, et donc

$$\lim_{n\to+\infty} \frac{N_n(y)}{n} = 0$$

ce qui correspond à (6.20) dans ce cas puisque $m_y = 0$. On a donc établi (6.20) $\mathbb{P}_x$-ps pour tout $x \in E$.

d) En notant que si $\mathbb{P}_x(A) = 1$ pour tout $x \in E$ alors par la définition de $\mathbb{P}_\nu$ en (5.28), on a :

$$\mathbb{P}_\nu(A) = \sum_{x\in E} \nu(x)\,\mathbb{P}_x(A) = \sum_{x\in E} \nu(x) = 1,$$

on a encore (6.20) $\mathbb{P}_\nu$-presque sûrement pour toute loi initiale ν.

e) Pour prouver (6.21), on observe que $0 \leq N_n(y)/n \leq 1$ puisque $0 \leq N_n(y) \leq n$. Dés lors, le théorème de convergence dominée permet d'obtenir (6.21) de (6.20) :

$$\lim_{n\to+\infty} \mathbb{E}_x\left[\frac{N_n(y)}{n}\right] = \mathbb{E}_x\left[\lim_{n\to+\infty} \frac{N_n(y)}{n}\right] = \mathbb{E}_x\left[\frac{\mathbf{1}_{\{\widetilde{T}_y<+\infty\}}}{m_y}\right] = \frac{\mathbb{P}_x(\widetilde{T}_y < +\infty)}{m_y} = \frac{\rho_{x,y}}{m_y}.$$

□

6.2 Ensembles clos et irréductibilité

Relations entre états

On note E_R l'ensemble des états récurrents et E_T l'ensemble des états transitoires. D'après la Proposition 6.4, on a

$$E = E_R \sqcup E_T. \tag{6.25}$$

Dans cette section, on précise cette partition de l'espace d'états E.

Définition 6.17 *Étant donné deux états $x, y \in E$, on dit que x peut mener à y et on note $x \rightsquigarrow y$ si $\rho_{x,y} = \mathbb{P}_x\big(\widetilde{T}_y < +\infty\big) > 0$.*

Proposition 6.18 *Pour des états x, y distincts, on a les équivalences :*

(1) $x \rightsquigarrow y$;

(2) $G(x, y) > 0$;

(3) $\exists n \geq 1$ tel que $P^n(x, y) > 0$, ie., avec une probabilité strictement positive, il existe un chemin de x à y en un nombre fini d'étape.

Démonstration : On suppose que la chaîne part de $x \neq y$.

1)$\Longleftrightarrow$2). Comme $\{N(y) \geq 1\} = \{\widetilde{T}_y < +\infty\}$, on a $\mathbb{P}_x(N(y) \geq 1) = \rho_{x,y}$. La condition $\rho_{x,y} > 0$ est alors équivalente à avoir $N(y) \geq 1$ avec probabilité $\mathbb{P}_x$ positive donc à $G(x, y) = \mathbb{E}_x[N(y)] > 0$.

2)$\Longleftrightarrow$3) suit immédiatement de $G(x, y) = \sum_{n\geq 0} P^n(x, y) = \sum_{n\geq 1} P^n(x, y)$ ($n \neq 0$ car $x \neq y$ et $P^0(x, y) = \delta_{x,y} = 0$). □

La relation $\rightsquigarrow$ est transitive :

Proposition 6.19 *Si $x \rightsquigarrow y$ et $y \rightsquigarrow z$ alors $x \rightsquigarrow z$.*

Démonstration : En effet, pour aller de x à z on peut en particulier aller de x à y et de y à z : plus précisement, on a

$$\begin{aligned}
\rho_{x,z} &= \mathbb{P}_x(T_z < +\infty) \\
&\geq \mathbb{P}_x\big(T_y < +\infty, T_z \circ \Theta_{T_y} < +\infty\big) \\
&= \mathbb{E}_x\big[\mathbf{1}_{\{T_y<+\infty\}}\,\mathbf{1}_{\{T_z\circ\Theta_{T_y}<+\infty\}}\big] = \mathbb{E}_x\big[\mathbf{1}_{\{T_y<+\infty\}}\,\mathbb{E}_x\big[\mathbf{1}_{\{T_z\circ\Theta_{T_y}<+\infty\}}|\mathcal{F}_{T_y}\big]\big] \\
&= \mathbb{E}_x\big[\mathbf{1}_{\{T_y<+\infty\}}\,\mathbb{E}_{X_{T_y}}\big[\mathbf{1}_{\{T_z<+\infty\}}\big]\big] = \mathbb{E}_x\big[\mathbf{1}_{\{T_y<+\infty\}}\big]\,\mathbb{E}_y\big[\mathbf{1}_{\{T_z<+\infty\}}\big] \\
&= \mathbb{P}_x(T_y < +\infty)\,\mathbb{P}_y(T_z < +\infty) = \rho_{x,y}\rho_{y,z} > 0,
\end{aligned}$$

due à la propriété de Markov forte (5.38) sous la forme du Corollaire 5.38.

Autre façon de faire, $x \rightsquigarrow y$ et $y \rightsquigarrow z$ impliquent $P^n(x, y) > 0$ et $P^m(y, z) > 0$ pour des entiers $n, m \geq 1$. En utilisant la relation de Chapman-Kolmogorov (5.12), on a $x \rightsquigarrow z$ car

$$P^{n+m}(x, z) = \sum_{w\in E} P^n(x, w)P^m(w, z) \geq P^n(x, y)P^m(y, z) > 0.$$

□

Théorème 6.20 *Soit $x \in E_R$ et $y \in E$ tel que $x \rightsquigarrow y$ (ie. $G(x,y) > 0$). Alors $y \in E_R$ et $\rho_{y,x} = \mathbb{P}_y(\widetilde{T}_x < +\infty) = 1$. En particulier, $y \rightsquigarrow x$ (ie. $G(y,x) > 0$) et on a même $\rho_{x,y} = 1$.*

Démonstration : Pour $y = x$, l'énoncé est immédiat ($\rho_{x,x} = 1$ car $x \in E_R$). On suppose donc $y \neq x$ et on dispose de la Proposition 6.18.

On commence par montrer que $\mathbb{P}_y(\widetilde{T}_x < +\infty) = 1$. Lorsque $\widetilde{T}_y < +\infty$ et $\widetilde{T}_x \circ \Theta_{\widetilde{T}_y} = +\infty$, on a nécessairement $N(x) \le \widetilde{T}_y$ (puisque après $\widetilde{T}_y$, il n'y a plus de visite en x), on a donc

$$\big\{\widetilde{T}_y < +\infty \text{ et } \widetilde{T}_x \circ \Theta_{\widetilde{T}_y} = +\infty\big\} \subset \{N(x) < +\infty\}.$$

Comme x est récurrent, on a

$$\begin{aligned}
0 &= \mathbb{P}_x(N(x) < +\infty) \ge \mathbb{P}_x\big(\widetilde{T}_y < +\infty \text{ et } \widetilde{T}_x \circ \Theta_{\widetilde{T}_y} = +\infty\big)\\
&= \mathbb{E}_x\big[\mathbf{1}_{\{\widetilde{T}_y<+\infty\}}\,(\mathbf{1}_{\{\widetilde{T}_x=+\infty\}} \circ \Theta_{\widetilde{T}_y})\big] = \mathbb{E}_x\big[\mathbf{1}_{\{\widetilde{T}_y<+\infty\}}\,\mathbb{E}_x\big[(\mathbf{1}_{\{\widetilde{T}_x=+\infty\}} \circ \Theta_{\widetilde{T}_y})|\mathcal{F}_{\widetilde{T}_y}\big]\big]\\
&= \mathbb{E}_x\big[\mathbf{1}_{\{\widetilde{T}_y<+\infty\}}\,\mathbb{E}_{X_{\widetilde{T}_y}}\big[\mathbf{1}_{\{\widetilde{T}_x=+\infty\}}\big]\big] = \mathbb{E}_x\big[\mathbf{1}_{\{\widetilde{T}_y<+\infty\}}\big]\mathbb{E}_y\big[\mathbf{1}_{\{\widetilde{T}_x=+\infty\}}\big]\\
&\quad \text{(propriété de Markov forte (5.38) sous la forme du Corollaire 5.38)}\\
&= \mathbb{P}_x\big(\widetilde{T}_y < +\infty\big)\,\mathbb{P}_y\big(\widetilde{T}_x = +\infty\big).
\end{aligned}$$

Comme $x \rightsquigarrow y$, on a $\rho_{x,y} = \mathbb{P}_x\big(\widetilde{T}_y < +\infty\big) > 0$, et cela exige $\mathbb{P}_y\big(\widetilde{T}_x = +\infty\big) = 0$ et donc $\rho_{y,x} = \mathbb{P}_y\big(\widetilde{T}_x < +\infty\big) = 1$, c'est à dire $y \rightsquigarrow x$.

On termine en montrant que $y \in E_R$. Comme par définition (Déf. 6.7) du potentiel G :

$$G(x,y) = \sum_{k\ge 0} P^k(x,y) > 0, \quad G(y,x) = \sum_{k\ge 0} P^k(y,x) > 0,$$

on peut trouver des entiers $n_1, n_2 \ge 1$ tels que

$$P^{n_1}(x,y) > 0, \quad P^{n_2}(y,x) > 0. \tag{6.26}$$

Pour tout entier $k \ge 0$, on a alors

$$P^{n_1+k+n_2}(y,y) \ge P^{n_2}(y,x)P^k(x,x)P^{n_1}(x,y)$$

et donc

$$G(y,y) \ge \sum_{k=0}^{+\infty} P^{n_1+k+n_2}(y,y) \ge P^{n_2}(y,x)\left(\sum_{k=0}^{+\infty} P^k(x,x)\right) P^{n_1}(x,y) = +\infty$$

puisque $\sum_{k=0}^{+\infty} P^k(x,x) = G(x,x) = +\infty$ et n_1, n_2 satisfont (6.26). On a donc $y \in E_R$.

Pour terminer, on obtient $\rho_{x,y}$ en échangeant les rôles de x et y puisqu'on sait que $y \in E_R$ et $y \rightsquigarrow x$. □

Remarque 6.21 **($x \in E_R \not\rightsquigarrow y \in E_T$)** Si $x \in E_R$ et $y \in E_T$ alors nécessairement par le Théorème 6.20 on a $G(x,y) = 0$: un état récurrent ne peut pas mener à un état transitoire !

Le résultat suivant précise le Théorème 6.20.

Théorème 6.22 *Soit x un état récurrent positif (resp. nul). Si $x \rightsquigarrow y$ alors y est récurrent positif (resp. nul).*

Démonstration : Soit $x \in E_R \rightsquigarrow y$. D'après le Théorème 6.20, on sait déjà que $y \in E_R$ et $y \rightsquigarrow x$.

D'abord, on suppose que x est récurrent positif. Il existe donc des entiers $n_1, n_2 \geq 1$ tels que

$$P^{n_1}(x,y) > 0, \quad P^{n_2}(y,x) > 0. \tag{6.27}$$

On a alors

$$P^{n_1+k+n_2}(y,y) \geq P^{n_2}(y,x)P^k(x,x)P^{n_1}(x,y),$$

puis en sommant sur $k = 0, 2, \dots, n$, et en divisant par n, on obtient

$$\begin{aligned}
G_{n_1+n+n_2}(y,y) - G_{n_1+n_2-1}(y,y) = \sum_{j=n_1+n_2}^{n_1+n_2+n} P^j(y,y) = \sum_{k=0}^{n} P^{n_1+k+n_2}(y,y) \\
\geq \sum_{k=0}^{n} P^{n_2}(y,x)P^k(x,x)P^{n_1}(x,y) = P^{n_2}(y,x)\left(\sum_{k=0}^{n} P^k(x,x)\right)P^{n_1}(x,y) \\
= P^{n_1}(y,x)P^{n_2}(x,y)G_n(x,x). \qquad (6.28)
\end{aligned}$$

Par le Théorème 6.15, quand $n \to +\infty$,

$$\begin{aligned}
\lim_{n\to+\infty} \frac{G_{n_1+n+n_2}(y,y)}{n} - \frac{G_{n_1+n_2}(y,y)}{n} &= \frac{1}{m_y} \\
\lim_{n\to+\infty} P^{n_2}(y,x)P^{n_1}(x,y)\frac{G_n(x,x)}{n} &= \frac{P^{n_2}(y,x)P^{n_1}(x,y)}{m_x},
\end{aligned}$$

et donc par (6.28)

$$\frac{1}{m_y} \geq \frac{P^{n_2}(y,x)P^{n_1}(x,y)}{m_x} > 0,$$

car $m_x < +\infty$ (x récurrent positif) et par choix de n_1, n_2 en (6.27), ce qui exige $m_y < +\infty$, c'est à dire y est récurrent positif.

Ensuite, dans le cas où x est récurrent nul, nécessairement y doit l'être aussi car si y était récurrent positif, comme $y \rightsquigarrow x$ (Th. 6.20), la première partie exigerait x récurrent positif, ce qui n'est pas le cas. □

Ensemble clos

Définition 6.23 (Ensemble clos) *Un ensemble d'états $C \subset E$ est dit clos si aucun état de C ne peut mener à l'extérieur de C, ie. $\rho_{x,y} = 0$, $\forall x \in C$, $y \notin C$ ou encore pour tout $n \geq 1$, $x \in C$, $y \notin C$, $P^n(x, y) = 0$.*

Exemple 6.24 Un état absorbant (Déf. 5.12) est un cas (très) particulier d'ensemble clos.

En fait, par récurrence on montre qu'il suffit de voir la propriété de la Définition 6.23 pour $n = 1$:

Proposition 6.25 *Si pour tout $x \in C$ et $y \notin C$ on a $P(x, y) = 0$, alors l'ensemble C est clos.*

Démonstration : On montre par récurrence que $P^n(x, y) = 0$ pour tout $x \in C$, $y \notin C$ et $n \geq 1$ dès que c'est vrai pour $n = 1$. Si c'est le cas pour P^{n-1} alors

$$P^n(x, y) = \sum_{z \in E} P(x, z)P^{n-1}(z, y) = \sum_{z \in C} P(x, z)P^{n-1}(z, y) = 0$$

par la relation de Chapman-Kolmogorov (5.12) pour la première égalité, l'hypothèse sur $x \in C$ dans la deuxième et par hypothèse de récurrence sur P^n pour la troisième. □

Irréductibilité

Définition 6.26 (Irréductibilité) *Un ensemble C clos est dit irréductible si pour tout $x, y \in C$ alors x peut mener à y (et y à x). Une chaîne est dite irréductible si l'espace d'états E entier l'est.*

Remarque 6.27 D'après la Proposition 6.18, la Définition 6.26 est équivalente à

- pour tout $x, y \in C$, on a $G(x, y) > 0$;
- pour tout $x, y \in C$, il existe $n = n(x, y) \geq 1$ tel que $P^n(x, y) > 0$;
- pour tout $x, y \in C$, il existe $n = n(x, y) \geq 1$ et $x_0 = x, x_1 \dots, x_n = y$ tels que $P(x_i, x_{i+1}) > 0$ pour tout $0 \leq i \leq n - 1$.

Proposition 6.28 *Dans un ensemble clos irréductible, tous les états sont de même nature : tous transitoires ou tous récurrents positifs ou tous récurrents nuls.*

Démonstration : S'il existe $x \in C$ récurrent positif (resp. récurrent nul) alors par le Théorème 6.22, tous les autres états de C sont récurrents positifs (resp. récurrents nuls). Sinon c'est que tous les états de C sont transitoires. □

En fait, on a :

Théorème 6.29 *Soit $(X_n)_{n\geq 0}$ une chaîne de Markov.*

(1) Soit C un ensemble clos irréductible d'états récurrents. Alors pour tout $x, y \in C$ on a $\rho_{x,y} = 1$, $\mathbb{P}_x(N(y) = +\infty) = 1$ et $G(x, y) = +\infty$.

(2) Soit C 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 nécéssairement récurrente positive.

Démonstration : 1) vient des Propositions 6.4 et 6.5 et du Théorème 6.8.

2) D'abord, un ensemble C fini et clos a au moins un état récurrent. En effet, de la même façon que dans la Remarque 6.13, comme la chaîne ne quitte pas C, si tous les états étaient transitoires alors on aurait en partant de $x \in C$:

$$0 = \sum_{y \in C} \lim_{n \to +\infty} P^n(x, y) = \lim_{n \to +\infty} \sum_{y \in C} P^n(x, y) = \lim_{n \to +\infty} \mathbb{P}_x(X_n \in C) = 1.$$

Cela exige donc que l'ensemble C contienne au moins un état récurrent.

Puis, il y a même nécessairement un état récurrent positif dans C : si la chaîne part de C, on a

$$n = \sum_{y \in C} N_n(y) \quad \text{et} \quad 1 = \sum_{y \in C} \frac{N_n(y)}{n}$$

et pour tout $x \in C$:

$$1 = \mathbb{E}_x \left[\sum_{y \in C} \frac{N_n(y)}{n} \right] = \sum_{y \in C} \mathbb{E}_x \left[\frac{N_n(y)}{n} \right] = \sum_{y \in C} \frac{G_n(x, y)}{n}.$$

Comme C est fini :

$$1 = \lim_{n \to +\infty} \sum_{y \in C} \frac{G_n(x, y)}{n} = \sum_{y \in C} \lim_{n \to +\infty} \frac{G_n(x, y)}{n} = \sum_{y \in C} \frac{\rho_{x,y}}{m_y}.$$

Il existe donc $y \in C$ avec $m_y < +\infty$, c'est à dire $y \in C$ est récurrent positif. Puis comme y mène à tout $x \in C$ (par la Déf. 6.23 de C clos), on a aussi x récurrent positif par le Théorème 6.22.

3) Appliquer **1)** avec $C = E$. □

Corollaire 6.30 (Irréductibilité, récurrence et transience) *Si $(X_n)_{n \geq 0}$ est une chaîne de Markov irréductible partant de $x \in E$, on a l'alternative :*

(1) ou bien la chaîne est récurrente : tous les états sont récurrents

$$\mathbb{P}_x\big(N(y) = +\infty \ \forall y \in E\big) = 1.$$

(2) ou bien la chaîne est transitoire : tous les éléments sont transitoires

$$\mathbb{P}_x\big(N(y) < +\infty \ \forall y \in E\big) = 1.$$

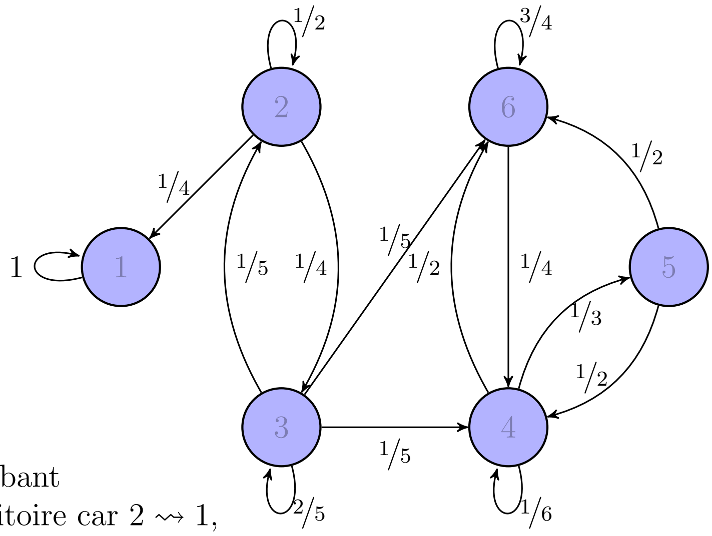
Une chaîne de Markov irréductible est donc soit transitoire soit récurrente positive soit récurrente nulle.

Démonstration : S'il existe un état x récurrent, le Théorème 6.20 montre que tous les états sont récurrents puisque par irréductibilité, x mène à tous les états y. De plus, puisque $G(x,y) > 0$ pour tout $x, y \in E$, il n'y a qu'une seule classe de récurrence. Le reste découle du Théorème 6.29. □

Définition 6.31 (Récurrence et irréductibilité) *Une chaîne de Markov irréductible dont tous les états sont récurrents est dite récurrente irréductible.*

Exemple 6.32 Classifier les états de la chaîne de Markov sur un espace d'états fini avec pour matrice stochastique

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & 0 & \frac{1}{5} \\ 0 & 0 & 0 & \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{3}{4} \end{pmatrix}.$$



On observe que
- 1 est absorbant
- 2 est transitoire car $2 \rightsquigarrow 1$,
- 3 est transitoire car $3 \rightsquigarrow 2 \rightsquigarrow 1$,
- $\{4,5,6\}$ forme une classe close irréductible qui est donc récurrente (positive).

On en déduit la classification :

$$\{1,2,3,4,5,6\} = \underbrace{\{1\} \cup \{4,5,6\}}_{\text{classes de récurrence}} \cup \underbrace{\{2,3\}}_{\text{états transitoires}}.$$

6.3 Classes de récurrence

Dans cette section, on précise les ensembles clos récurrents.

Définition 6.33 (Relation d'équivalence $\sim$) *On dit que deux états x et y communiquent et on note $x \sim y$ lorsque $x \rightsquigarrow y$ et $y \rightsquigarrow x$.*

Proposition 6.34 *Sur E_R la relation $\sim$ définit bien une relation d'équivalence. De plus on a :*

$$x \sim y \Leftrightarrow x \rightsquigarrow y \Leftrightarrow y \rightsquigarrow x \Leftrightarrow G(x,y) > 0 \Leftrightarrow G(y,x) > 0 \Leftrightarrow \rho_{x,y} > 0 \Leftrightarrow \rho_{y,x} > 0 \quad (6.29)$$

et dans ce cas $\rho_{x,y} = \rho_{y,x} = 1$.

Démonstration : On définit bien une une relation d'équivalence sur E_R puisque
- réflexivité ($x \sim x$) car $\rho_{x,x} = 1$ et donc $x \rightsquigarrow x$;
- symétrie par définition ;
- transitivité par la Proposition 6.19.

De plus d'après le Théorème 6.20 si $x \in E_R \rightsquigarrow y$ alors $y \in E_R$ et $y \rightsquigarrow x$, prouvant que $x \sim y$ est équivalent à $x \rightsquigarrow y$ pour des états récurrents. Le reste de (6.29) s'en déduit facilement, notamment avec la Proposition 6.18. □

On a la partition de l'ensemble des états récurrents E_R en classes d'équivalence de la relation d'équivalence $\sim$:

$$E_R = \bigsqcup_{i \in I} E_{R_i} = \Big(\bigsqcup_{i \in I_+} E_{R_i}\Big) \sqcup \Big(\bigsqcup_{i \in I_0} E_{R_i}\Big). \quad (6.30)$$

Les ensembles E_{R_i}, $i \in I$, sont appelés les **classes de récurrence** de la chaîne. Une classe de récurrence est close et irréductible et, d'après le Théorème 6.22, elle est soit récurrente positive (lorsque $i \in I^+$) soit récurrente nulle (lorsque $i \in I_0$). Les partitions (6.25) et (6.30) se combinent en la partition globale de l'espace d'états qu'on appelle **classification des états** de la chaîne de Markov

$$E = E_T \sqcup \Big(\bigsqcup_{i \in I_+} E_{R_i}\Big) \sqcup \Big(\bigsqcup_{i \in I_0} E_{R_i}\Big) \quad (6.31)$$

où E_T est l'ensemble (a priori non clos, non irréductible) des états transitoires, les classes E_{R_i} sont récurrentes positives pour $i \in I_+$ et récurrentes nulles pour $i \in I_0$. Et les classes de récurrence sont closes irréductibles.

Théorème 6.35 (Classes de récurrence) *Les classes de récurrence E_{R_i}, $i \in I$, de la partition (6.30) de l'ensemble des états récurrents E_R vérifient*

(1) Si $x \in E_{R_i}$ alors $\mathbb{P}_x$-ps
- *$N(y) = +\infty$ pour tout $y \in E_{R_i}$;*
- *$N(y) = 0$ pour tout $y \notin E_{R_i}$.*

(2) Si $x \in E_T$ *et* $T_{E_R} = \inf\big(n \geq 0 : X_n \in E_R\big)$ *alors*
— *ou bien* $T_{E_R} = +\infty$ *et* $\mathbb{P}_x$*-ps :* $N(y) < +\infty$ *pour tout* $y \in E$ *;*
— *ou bien* $T_{E_R} < +\infty$ *et* $\mathbb{P}_x$*-ps :* $\exists j \in I$ *(aléatoire) tel que pour tout* $n \geq T_{E_R}$ *on a* $X_n \in E_{R_j}$*.*

Démonstration : 1) Soit $x \in E_{R_i}$. On a $G(x,y) = 0$ pour tout $y \in E \setminus E_{R_i}$. En effet,
— si $y \in E_{R_j}$, $j \neq i$, la partition garantit que x et y ne communiquent pas et donc $G(x,y) = 0$ et $N(y) = 0$ $\mathbb{P}_x$-ps ;
— puis si $y \in E_T$, le Théorème 6.20 assure encore $G(x,y) = 0$. En particulier, $N(y) = 0$ $\mathbb{P}_x$-ps.

En revanche si $y \in E_{R_i}$, on a $\rho_{x,y} = \mathbb{P}_x(\widetilde{T}_y < +\infty) = 1$ d'après le Théorème 6.20 et par la Prop. 6.5, on a $\mathbb{P}_x(N(y) = +\infty) = 1$.

2) Soit maintenant $x \in E_T$.
— Si $T_{E_R} < +\infty$: la chaîne rentre dans $E_R = \bigsqcup_{j \in I} E_j$ donc dans une des classes E_{R_j} (pour un j aléatoire). D'après la propriété de Markov (5.38) (Théorème 5.37) et la première partie de l'énoncé, on a $X_n \in E_{R_j}$ pour tout $n \geq T_{E_{R_j}} = T_{E_R}$.
— Si $T_{E_R} = +\infty$, alors $N(y) = 0$ pour $y \in E_R$ (puisque $T_{E_R} = +\infty$) et $\mathbb{P}_x(N(y) < +\infty) = 1$ par la Prop. 6.5. □

Avec le Théorème 6.35, on peut préciser le Théorème 6.15 de la façon suivante :

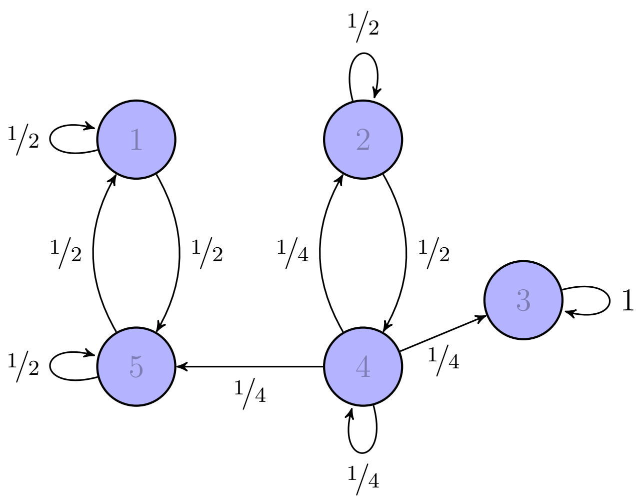
Corollaire 6.36 *Soit C un ensemble clos irréductible d'états récurrents. Alors*

$$\lim_{n \to +\infty} \frac{G_n(x,y)}{n} = \frac{1}{m_y}, \qquad x, y \in C, \tag{6.32}$$

et si $\mathbb{P}(X_0 \in C) = 1$ *alors avec probabilité* 1 *:*

$$\lim_{n \to +\infty} \frac{N_n(y)}{n} = \frac{1}{m_y}, \qquad y \in C. \tag{6.33}$$

Exemple 6.37 Retour sur l'Exemple 6.2 de l'introduction dont on rappelle le graphe de transition :



— Les classes de récurrence sont

$$E_{R_1} = \{1,5\}, \quad E_{R_2} = \{3\} \ \text{(ie. 3 est absorbant)}.$$

— Les états transitoires sont $\{2,4\} = E_T$.
En effet

$$P(1,5) > 0,\ P(5,1) > 0,\ P(1,1) > 0,\ P(5,5) > 0,$$
$$P(1,j) = 0,\ P(5,j) = 0 \quad \forall j \notin \{1,5\}$$

assurent $G(1,5) > 0$, $G(5,1) > 0$ et $G(1,j) = G(5,j) = 0$ $\forall j \neq 1,5$. Puis

$$P(3,3) > 0 \quad \text{et} \quad P(3,j) = 0 \quad \forall j \neq 3.$$

Par ailleurs $\mathbb{P}_1\big(\widetilde{T}_1 < +\infty\big) = 1$ car

$$\begin{aligned}
\mathbb{P}_1\big(\widetilde{T}_1 = +\infty\big) &= \mathbb{P}_1(X_1 = 5 \text{ et } X_n = 5\ \forall n \geq 2)\\
&= P(1,5)\mathbb{P}_5(X_n = 5\ \forall n \geq 1)\\
&\leq P(1,5) \lim_{N\to+\infty} \mathbb{P}_5\big(X_n = 5\ \forall n \in \{1,\dots,N\}\big)\\
&\leq P(1,5) \lim_{N\to+\infty} \Big(\frac{1}{2}\Big)^N = 0,
\end{aligned}$$

ie. $\mathbb{P}_1\big(\widetilde{T}_1 < +\infty\big) = 1$ et 1 est donc récurrent.

De la même façon $\mathbb{P}_5\big(\widetilde{T}_5 < +\infty\big) = 1$ ie. 5 est récurrent et $\mathbb{P}_3\big(\widetilde{T}_3 < +\infty\big) = 1 = P(3,3)$ donc 3 est récurrent.

L'état $\{2\}$ est transitoire car $\mathbb{P}_2\big(\widetilde{T}_2 < +\infty\big) < 1$, ie. $\mathbb{P}_2\big(\widetilde{T}_2 = +\infty\big) > 0$. En effet,

$$\mathbb{P}_2\big(\widetilde{T}_2 = +\infty\big) \geq \mathbb{P}_2(X_1 = 4 \text{ et } X_2 = 3) = P(2,4)P(4,3) = \frac{1}{3} \times \frac{1}{4} = \frac{1}{12}.$$

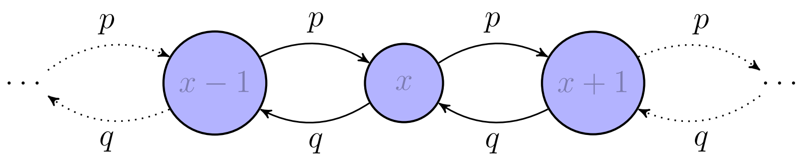
L'état $\{4\}$ est transitoire car

$$\mathbb{P}_4\big(\widetilde{T}_4 = +\infty\big) \geq \mathbb{P}_4(X_1 = 3) = \frac{1}{4} > 0.$$

On a alors la décomposition de l'espace d'états

$$E = \big\{1,2,3,4,5\big\} = \underbrace{\{2,4\}}_{E_T} \sqcup \underbrace{\underbrace{\{1,5\}}_{E_{R_1}} \sqcup \underbrace{\{3\}}_{E_{R_2}}}_{E_R}.$$

Exemple 6.38 (Marche aléatoire simple sur $\mathbb{Z}$) On considère $S_n = \sum_{i=1}^n X_i$ avec X_i *iid* de loi de Rademacher $(1-p)\delta_{-1} + p\delta_1$. La chaîne $(S_n)_{n\geq 0}$ a pour espace d'états $\mathbb{Z}$ et matrice stochastique $P(x,y) = (1-p)\mathbf{1}_{\{y=x-1\}} + p\mathbf{1}_{\{y=x+1\}}$.



On a immédiatement l'irréductibilité de la chaîne puisque pour tout $x \neq y$: $P^{|x-y|}(x,y) = p^{y-x} > 0$ (si $x < y$) ou $(1-p)^{x-y} > 0$ (si $y < x$) et $P^2(x,x) = 2p(1-p)$. La nature de la chaîne est donc déterminée par la nature d'un point quelconque. On détermine celle de l'état 0 (est-il récurrent ou transitoire ?).

Pour cela, en vertu du Corollaire 6.9, on se ramène à calculer $G(0,0)$:

$$G(0,0) = \sum_{n=0}^{+\infty} P_n(0,0) = \sum_{n=0}^{+\infty} P_{2n}(0,0)$$

car $P_{2n+1}(0,0) = 0$. De plus comme

$$P_{2n}(0,0) = \binom{2n}{n} p^n(1-p)^n = \frac{(2n)!}{(n!)^2} p^n(1-p)^n \sim \frac{(4p(1-p))^n}{\sqrt{\pi n}}$$

en utilisant la formule de Stirling $n! \sim (n/e)^n \sqrt{2\pi n}$. Ainsi,

— si $p \neq 1/2$, alors $(4p(1-p)) < 1$ et $G(0,0) = \sum_{n=0}^{+\infty} P_{2n}(0,0) < +\infty$: 0 est transitoire (et tous les états le sont !) ;

— si $p = 1/2$, alors $(4p(1-p)) = 1$ et $G(0,0) \sim \sum_{n=0}^{+\infty} (1/\sqrt{\pi n}) = +\infty$: 0 est récurrent (et tous les états le sont !).

Pour décider si la chaîne est récurrente positive ou nulle, voir le Chapitre 7 (Exemple 7.5 et Th. 7.25).

6.4 Absorption dans les classes de récurrence

Pour calculer les durées d'absorption dans les classes de récurrence lorsque la chaîne part de x, on introduit les quantités suivantes :

$$\begin{array}{rcl}
\text{(temps d'absorption) } S_i & = & \min\big(n \geq 0 : X_n \in E_{R_i}\big) \\
\text{(probabilité d'absorption) } \rho_i(x) & = & \mathbb{E}_x\big[\mathbf{1}_{\{S_i<+\infty\}}\big] = \mathbb{P}_x(S_i < +\infty) \\
\text{(temps moyen d'absorption tronqué) } \tau_i(x) & = & \mathbb{E}_x\big[S_i \mathbf{1}_{\{S_i<+\infty\}}\big] \\
\text{(temps moyen d'absorption) } t_i(x) & = & \mathbb{E}_x\big[S_i \,|S_i < +\infty\big].
\end{array}$$

Si $\rho_i(x) > 0$ alors $t_i(x)$ est bien défini et on a

$$t_i(x) = \mathbb{E}_x\big[S_i \,|S_i < +\infty\big] = \frac{\mathbb{E}_x\big[S_i \mathbf{1}_{\{S_i<+\infty\}}\big]}{\mathbb{P}_x(S_i < +\infty)} = \frac{\tau_i(x)}{\rho_i(x)}. \tag{6.34}$$

Immédiatement, on a :

— si $x \in E_{R_i}$, alors sous $\mathbb{P}_x$, $S_i = 0$ et donc $\rho_i(x) = 1$ et $\tau_i(x) = t_i(x) = 0$;
— si $x \in E_{R_j}$ pour $j \neq i$, alors sous $\mathbb{P}_x$, $S_i = +\infty$, et donc $\rho_i(x) = 0$ et $\tau_i(x) = 0$.

Le cas restant, intéressant à traiter, est le cas x transitoire. De plus, si $x \in E_T$ alors a fortiori $x \notin E_{R_i}$ et $S_i \geq 1$; dés lors $\rho_i(x) = 0$ implique $\tau_i(x) = 0$. Ainsi si $\rho_i(x) = 0$, on convient de prendre $t_i(x) = 0$.

Théorème 6.39 (Absorption) *Soit $x \in E_T$. Pour chaque $i \in I$, les probabilités d'absorption $\rho_i(x)$ et le temps moyen d'absorption $\tau_i(x)$, $t_i(x)$ sont solutions du système linéaire :*

$$\rho_i(x) = \sum_{y\in E} P(x,y)\rho_i(y) \tag{6.35}$$

$$\tau_i(x) = \rho_i(x) + \sum_{y\in E} P(x,y)\tau_i(y), \tag{6.36}$$

et lorsque $\rho_i(x) > 0$:

$$t_i(x) = 1 + \sum_{y\in E} P(x,y)\frac{\rho_i(y)}{\rho_i(x)}t_i(y). \tag{6.37}$$

Démonstration : Soit $x \in E_T$, on a $S_i \geq 1$ et on peut même écrire sous $\mathbb{P}_x$:

$$\begin{aligned} S_i &= \inf\big(n \geq 1 : X_n \in E_{R_i}\big) \\ &= 1 + \inf(k \geq 0 : X_{k+1} \in E_{R_i}\big) \\ &= 1 + S_i \circ \Theta_1. \end{aligned}$$

On utilise la propriété de Markov faible (5.32) en conditionnant par la première transition. Pour les probabilités d'absorption, on a :

$$\begin{aligned} \rho_i(x) &= \mathbb{P}_x\big(S_i < +\infty\big) = \mathbb{E}_x\big[\mathbf{1}_{\{S_i<+\infty\}}\big] \\ &= \mathbb{E}_x\big[\mathbf{1}_{\{1+S_i\circ\Theta_1<+\infty\}}\big] = \mathbb{E}_x\big[\mathbf{1}_{\{S_i\circ\Theta_1<+\infty\}}\big] \\ &= \mathbb{E}_x\big[\mathbb{E}_x\big[\mathbf{1}_{\{S_i\circ\Theta_1<+\infty\}}|\mathcal{F}_1\big]\big] = \mathbb{E}_x\big[\mathbb{E}_{X_1}\big[\mathbf{1}_{\{S_i<+\infty\}}\big]\big] \\ &\quad \text{(propriété de Markov faible (5.32) à la date } p=1) \\ &= \mathbb{E}_x\big[\rho_i(X_1)\big] = \sum_{y\in E} P(x,y)\rho_i(y), \end{aligned}$$

ce qui prouve le système (6.35). Puis pour les temps moyens d'absorption (tronqués), on a par un raisonnement analogue :

$$\begin{aligned} \tau_i(x) &= \mathbb{E}_x\big[S_i\mathbf{1}_{\{S_i<+\infty\}}\big] = \mathbb{E}_x\big[(1+S_i\circ\Theta_1)\mathbf{1}_{\{(1+S_i\circ\Theta_1)<+\infty\}}\big] \\ &= \mathbb{E}_x\big[\mathbb{E}_x\big[(1+S_i\circ\Theta_1)\mathbf{1}_{\{(1+S_i\circ\Theta_1)<+\infty\}}|\mathcal{F}_1\big]\big] \\ &= \mathbb{E}_x\big[\mathbb{E}_{X_1}\big[(1+S_i)\mathbf{1}_{\{S_i<+\infty\}}\big]\big] \\ &\quad \text{(propriété de Markov faible (5.32) à la date } p=1) \\ &= \mathbb{E}_x\Big[\mathbb{E}_{X_1}\big[S_i\mathbf{1}_{\{S_i<+\infty\}}\big] + \mathbb{E}_{X_1}\big[\mathbf{1}_{\{S_i<+\infty\}}\big]\Big] \end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}_x\big[\tau_i(X_1)\big] + \mathbb{E}_x\big[\rho_i(X_1)\big] \\
&= \sum_{y\in E} P(x,y)\tau_i(y) + \sum_{y\in E} P(x,y)\rho_i(y) \\
&= \sum_{y\in E} P(x,y)\tau_i(y) + \rho_i(x),
\end{aligned}
$$

en utilisant (6.35) déjà prouvée, ce qui prouve le système (6.36). Puis (6.37) découle de (6.36) avec (6.34) :

$$
t_i(x) = \frac{\tau_i(x)}{\rho_i(x)} = 1 + \sum_{y\in E} P(x,y)\frac{\tau_i(y)}{\rho_i(x)} = 1 + \sum_{y\in E} P(x,y)\frac{\rho_i(y)}{\rho_i(x)} t_i(y).
$$

□

Exemple 6.40 Retour sur l'Exemple 6.2 avec le calcul des **probabilités d'absorption**. On calcule $\rho_i(x)$ pour $i \in \big\{\{1,5\},\{3\}\big\}$ et $x \in \{2,4\}$.
Comme $\rho_1(1) = \rho_1(5) = 1$ et $\rho_1(3) = 0$, on a :

$$
\left\{\begin{array}{lcl} \rho_1(2) &=& \frac{1}{2}\rho_1(2) + \frac{1}{2}\rho_1(4) \\ \rho_1(4) &=& \frac{1}{4}\rho_1(2) + \frac{1}{4}\rho_1(3) + \frac{1}{4}\rho_1(4) + \frac{1}{4}\rho_1(5) \end{array}\right. \iff \left\{\begin{array}{lcl} \rho_1(2) &=& \rho_1(4) \\ 3\rho_1(4) &=& \rho_1(2) + 1 \end{array}\right.
$$

$$
\iff \boxed{\rho_1(2) = \rho_1(4) = \frac{1}{2}}
$$

Comme $\rho_2(1) = \rho_2(5) = 0$ et $\rho_1(3) = 1$, on a :

$$
\left\{\begin{array}{lcl} \rho_2(2) &=& \frac{1}{2}\rho_2(2) + \frac{1}{2}\rho_2(4) \\ \rho_2(4) &=& \frac{1}{4}\rho_2(2) + \frac{1}{4}\rho_2(3) + \frac{1}{4}\rho_2(4) + \frac{1}{4}\rho_2(5) \end{array}\right. \iff \left\{\begin{array}{lcl} \rho_2(2) &=& \rho_2(4) \\ 3\rho_2(4) &=& \rho_2(2) + 1 \end{array}\right.
$$

$$
\iff \boxed{\rho_2(2) = \rho_2(4) = \frac{1}{2}}
$$

Puis avec le calcul des **temps moyens d'absorption** : comme $\tau_1(1) = \tau_1(3) = \tau_1(5) = 0$, on a :

$$
\left\{\begin{array}{lcl} \tau_1(2) &=& \frac{1}{2}\tau_1(2) + \frac{1}{2}\tau_1(4) + \rho_1(2) \\ \tau_1(4) &=& \frac{1}{4}\tau_1(2) + \frac{1}{4}\tau_1(3) + \frac{1}{4}\tau_1(4) + \frac{1}{4}\tau_1(5) + \rho_1(4) \end{array}\right. \iff \left\{\begin{array}{lcl} \tau_1(2) &=& \tau_1(4) + 1 \\ 3\tau_1(4) &=& \tau_1(2) + 2 \end{array}\right.
$$

$$
\iff \boxed{\tau_1(2) = \frac{5}{2}, \tau_1(4) = \frac{3}{2}} \quad \text{et} \quad \boxed{\tau_1(2) = \frac{5}{2}\Big/\frac{1}{2} = 5, \tau_1(4) = \frac{3}{2}\Big/\frac{1}{2} = 3}
$$

Comme $\tau_2(1) = \tau_2(3) = \tau_2(5) = 0$, on a :

$$
\left\{\begin{array}{lcl} \tau_2(2) &=& \frac{1}{2}\tau_2(2) + \frac{1}{2}\tau_2(4) + \rho_2(2) \\ \tau_2(4) &=& \frac{1}{4}\tau_2(2) + \frac{1}{4}\tau_2(3) + \frac{1}{4}\tau_2(4) + \frac{1}{4}\tau_2(5) + \rho_2(4) \end{array}\right. \iff \left\{\begin{array}{lcl} \tau_2(2) &=& \tau_2(4) + 1 \\ 3\tau_2(4) &=& \tau_2(2) + 2 \end{array}\right.
$$

$$
\iff \boxed{\tau_2(2) = \frac{5}{2}, \tau_2(4) = \frac{3}{2}} \quad \text{et} \quad \boxed{\tau_2(2) = \frac{5}{2}\Big/\frac{1}{2} = 5, \tau_2(4) = \frac{3}{2}\Big/\frac{1}{2} = 3}
$$

Chapitre 7

Invariance et équilibre

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.

Génériquement, on considère dans ce chapitre une chaîne de Markov $(X_n)_{n\geq 0}$ sur un espace d'états au plus dénombrable E et avec une matrice stochastique $P = (P(x,y))_{x,y\in E}$.

7.1 Mesures invariantes

Invariance

Définition 7.1 (Mesure invariante) *Soit π une mesure (positive) sur E telle que $\pi(x) < +\infty$ pour tout $x \in E$ et $\pi \neq 0$. On dit que π est invariante (ou stationnaire) pour le noyau de transition P si π est solution de l'équation de Chapman-Kolmogorov :*

$$\pi = \pi\, P \quad \Longleftrightarrow \quad \forall y \in E,\ \text{on a } \pi(y) = \sum_{x\in E} \pi(x)P(x,y). \tag{7.1}$$

Par une récurrence immédiate de l'équation de Chapman-Kolmogorov (7.1), on a $\pi = \pi\, P^n$ pour tout $n \geq 0$. Dans le cas d'une probabilité, on a l'équivalence suivante pour l'invariance :

Proposition 7.2 *Soit $(X_n)_{n\geq 0}$ une chaîne de Markov. La loi de X_n est indépendante de n si et seulement si la distribution initiale μ_0 est une probabilité invariante π.*

Démonstration : Pour le sens direct, comme $\mu_n = \mu_0 P^n$, si la loi μ_n de X_n ne dépend pas de n alors en particulier $\mu_1 = \mu_0$ donne $\mu_0 = \mu_0 P$ et μ_0 vérifie l'équation de Chapman-Kolmogorov (7.1). Le sens réciproque est immédiat par (7.1) puisque si $\mu_0 = \pi$, on a

$$\mu_n = \pi P^n = (\pi P)P^{n-1} = \pi P^{n-1} = \cdots = \pi.$$

□

Remarque 7.3 (Probabilité et mesure invariantes) Attention, quand E est infini, il se peut qu'il existe une mesure invariante π mais pas de probabilité invariante, cf. l'exemple de la marche aléatoire simple qui suit. Dans ce cas, π est de poids $\pi(E) = +\infty$ et n'est pas normalisable en une probabilité.

Exemple 7.4 (Mesures invariantes de la chaîne à deux états) Pour la chaîne de Markov à deux états de l'Exemple 5.1 avec la matrice $P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$, on a vu que les mesures invariantes sont proportionnelles à (q, p) et il y a une seule probabilité invariante donnée par :

$$\pi = \left(\frac{q}{p+q}, \frac{p}{p+q}\right).$$

Exemple 7.5 (Marche aléatoire symétrique sur $\mathbb{Z}$) La mesure uniforme est l'unique mesure invariante pour la marche aléatoire simple sur $\mathbb{Z}$ mais il n'existe pas de probabilité invariante.

Supposons que π est invariante pour la marche aléatoire simple sur $\mathbb{Z}$. Alors pour tout $n \in \mathbb{Z}$, on a :

$$\pi(n) = \sum_{k\in\mathbb{Z}} P(k,n)\pi(k) = \frac{1}{2}\pi(n-1) + \frac{1}{2}\pi(n+1).$$

D'où

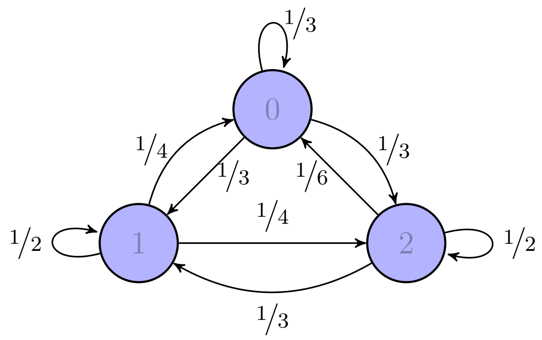
$$\pi(n+1) - \pi(n) = \pi(n) - \pi(n-1) = \pi(1) - \pi(0) := \alpha.$$

On déduit alors $\pi(n) = \pi(0) + n\alpha$ pour tout $n \in \mathbb{Z}$. Si $\alpha \neq 0$, c'est absurde pour n grand ou $-n$ grand car $\pi(n)$ devient négatif. Cela exige $\alpha = 0$ et $\pi(1) = \pi(0) = \pi(n)$ pour tout $n \in \mathbb{Z}$, ie. π est une mesure uniforme. Réciproquement, la mesure uniforme est bien solution de $\pi = \pi P$ donc invariante. À facteur multiplicatif près, il s'agit donc de l'unique mesure invariante de cette chaîne. De plus, il est impossible de normaliser la mesure uniforme sur $\mathbb{Z}$ en une mesure de probabilité : il n'existe donc pas de probabilité invariante pour la marche aléatoire simple symétrique.

Exemple 7.6 On considère une chaîne de Markov sur $E = \{0, 1, 2\}$ de matrice stochastique

$$\begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/4 & 1/2 & 1/4 \\ 1/6 & 1/3 & 1/2 \end{pmatrix}.$$

Le graphe de transitions est :



Une probabilité invariante π est solution du système $\pi = \pi P$:

$$\begin{cases} \frac{1}{3}\pi(0) + \frac{1}{4}\pi(1) + \frac{1}{6}\pi(2) & = & \pi(0) \\ \frac{1}{3}\pi(0) + \frac{1}{2}\pi(1) + \frac{1}{3}\pi(2) & = & \pi(1) \\ \frac{1}{3}\pi(0) + \frac{1}{4}\pi(1) + \frac{1}{2}\pi(2) & = & \pi(2) \\ \pi(0) + \pi(1) + \pi(2) & = & 1 \end{cases}$$

où la dernière équation est donnée par le fait que π est une probabilité. On en déduit facilement

$$\pi(0) = \frac{6}{25}, \quad \pi(1) = \frac{2}{5}, \quad \pi(2) = \frac{9}{25}.$$

On s'assure facilement que cette probabilité π est bien solution de (7.1), on a donc existence et unicité de la probabilité invariante.

De façon, générale, la recherche de mesure invariante π consiste à résoudre l'équation de Chapman-Kolmogorov (7.1) $\pi = \pi P$, c'est à dire un système linéaire de taille card(E) (possiblement infini). Même si E est fini, le système peut être de grande taille et difficile à résoudre !

Exemple 7.7 (Matrice bistochastique) Une matrice est dite **bistochastique** lorsque P^t est stochastique, ie. en plus de ses lignes, la somme de chacune de ses colonnes fait 1. Pour une telle matrice P, on observe que

$$(1, 1, \dots)P = (1, 1, \dots)$$

ie. $(1, 1, \dots)$ est vecteur propre à gauche de P pour la valeur propre 1. Comme ce vecteur correspond à la mesure uniforme sur E, cela signifie que la mesure uniforme est invariante pour une matrice bistochastique.

De plus, si E est fini de cardinal d alors on peut normaliser le vecteur $(1, 1, \dots)$ en $(1/d, \dots, 1/d)$ et dans ce cas, la probabilité uniforme sur E est invariante pour P bistochastique.

Proposition 7.8 *L'ensemble des mesures invariantes d'un noyau de transition est fermé et stable par combinaison linéaire à coefficients positifs. L'ensemble des probabilités invariantes est convexe, fermé et, si E est fini, il s'agit d'un compact.*

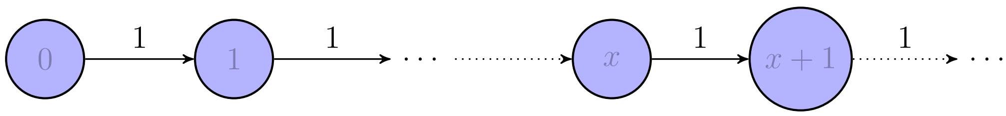
Démonstration : La linéarité et la fermeture suivent facilement de l'équation de Chapman-Kolmogorov (7.1) (linéaire et fermée). Si E est fini, l'ensemble des probabilités invariantes s'identifie à la partie de $\mathbb{R}_+^d$

$$\big\{(\mu_1, \dots, \mu_d) \in \mathbb{R}_+^d \ : \ \mu_1 + \cdots + \mu_d = 1\big\}. \tag{7.2}$$

Comme en dimension finie, la compacité est équivalente à être fermé et borné, l'ensemble (7.2) est clairement compact. Dès lors, la condition (7.1) définissant les mesures invariantes étant fermée, la compacité de l'ensemble des mesures invariantes suit. □

Il n'existe pas toujours de mesure invariante pour une chaîne de Markov comme on le voit dans les exemples suivants :

Exemple 7.9 (Chaîne de Markov sans mesure invariante) Soit $(X_n)_{n\geq 0}$ une chaîne de Markov sur $E = \mathbb{N}$ avec $P(i, i+1) = 1$ pour tout $i \geq 0$.



Si π est une mesure vérifiant l'équation de Chapman-Kolmogorov (7.1) $\pi = \pi P$, alors on doit avoir

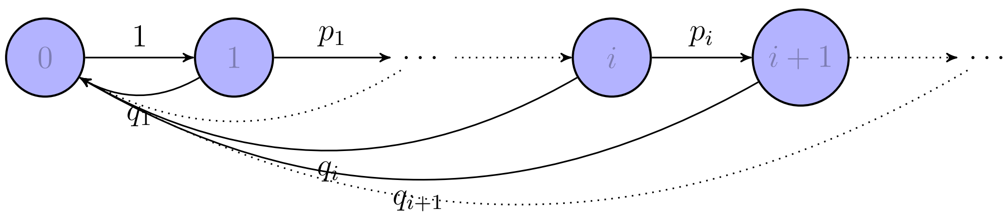
$$\begin{aligned}
\pi(i) &= \sum_{j\in\mathbb{N}} \pi(j)P(j,i) = \pi(i-1), \quad i \geq 1,\\
\text{et } \pi(0) &= \sum_{j\in\mathbb{N}} \pi(j)P(j,0) = 0 \quad \text{car } P(j,0) = 0 \text{ pour tout } j \in \mathbb{N}.
\end{aligned}$$

Il vient $\pi = 0$ et il n'existe pas de mesure invariante !

L'exemple suivant généralise l'Exemple 7.9 à une chaîne de Markov irréductible sans mesure invariante :

Exemple 7.10 (Chaînes de Markov irréductible sans mesure invariante) Soit $(p_i)_{i\geq 0}$ telle que $p_0 = 1$, $p_i > 0$ et $q_i := 1 - p_i > 0$ pour tout $i \geq 1$ avec $\sum_{i\geq 1} q_i < +\infty$ (par exemple $p_i = 1 - 1/(2i^2)$). On considère la chaîne de Markov $(X_n)_{n\geq 0}$ sur $E = \mathbb{N}$ avec le noyau de transition : $P(0,1) = p_0 = 1$ et

$$P(i, i+1) = p_i, \quad P(i, 0) = q_i, \ \forall i \geq 1.$$



Il s'agit d'une chaîne de Markov irréductible puisque pour tout $i, j \in \mathbb{N}$, le chemin $i \longrightarrow 0 \longrightarrow 1 \longrightarrow 2 \longrightarrow \dots \longrightarrow j-1 \longrightarrow j$ est possible lorsque $p_i \neq 1$ pour tout $i \in \mathbb{N}$.

Pour une mesure π invariante, l'équation de Chapman-Kolmogorov (7.1) se réduit à

$$\pi(0) = \sum_{i\geq 1} q_i \pi(i), \quad \pi(i) = p_{i-1}\pi(i-1), \quad \forall i \geq 1. \tag{7.3}$$

Une récurrence immédiate assure

$$\pi(i) = \pi(0) \prod_{j=0}^{i-1} p_j \tag{7.4}$$

ce qui, reinjectée dans (7.3), donne

$$\pi(0) = \pi(0) \sum_{i\geq 1} \Big(q_i \prod_{j=0}^{i-1} p_j \Big) = \pi(0) \sum_{i\geq 1} \Big(\prod_{j=0}^{i-1} p_j - \prod_{j=0}^{i} p_j \Big) = \pi(0) \Big(1 - \lim_{n\to+\infty} \prod_{j=0}^{n} p_j \Big).$$

Comme $\sum_{j\geq 1} q_j < +\infty$ implique $\prod_{j\geq 1} p_j > 0$, on doit avoir $\pi(0) = 0$ (sinon on aurait $\pi(0) < \pi(0)$) et donc par (7.4), il suit $\pi = 0$ et il n'y a pas de mesure invariante pour cette chaîne.

Réversibilité

Définition 7.11 (Réversibilité) *Une mesure π (positive) non nulle sur E telle que $\pi(x) < +\infty$ pour tout $x \in E$ est dite réversible pour le noyau P si pour tout $x, y \in E$:*

$$\pi(x)P(x, y) = \pi(y)P(y, x). \tag{7.5}$$

Par une récurrence simple, la définition (7.5) est équivalente à avoir pour toute suite $x_0, \dots, x_n \in E$:

$$\pi(x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n) = \pi(x_n)P(x_n, x_{n-1}) \cdots P(x_1, x_0). \tag{7.6}$$

On en déduit immédiatement un résultat qui éclaire la terminologie dans le cas d'une mesure de probabilité :

Proposition 7.12 (Réversibilité) *Une probabilité π est réversible pour un noyau 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*

$$\mathcal{L}_\pi(X_0, X_1, \dots, X_n) = \mathcal{L}_\pi(X_n, X_{n-1}, \dots, X_0).$$

Démonstration : Pour le sens direct, cela vient de (7.6) puisque pour tout $x_0, \dots, x_n \in E$, on a

$$\mathbb{P}_\pi\big(X_0 = x_0, X_1 = x_1, \dots X_n = x_n\big) = \pi(x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n)$$

$$\begin{aligned}\mathbb{P}_\pi\big(X_n = x_0, X_{n-1} = x_1, \dots X_0 = x_n\big) &= \mathbb{P}_\pi\big(X_0 = x_n, X_1 = x_{n-1}, \dots X_n = x_0\big)\\ &= \pi(x_n)P(x_n, x_{n-1})\cdots P(x_1, x_0).\end{aligned}$$

Pour le sens indirect, $\mathcal{L}(X_0, X_1) = \mathcal{L}(X_1, X_0)$ donne pour tout $x_0, x_1 \in E$:

$$\mathbb{P}_\pi\big(X_0 = x_0, X_1 = x_1\big) = \mathbb{P}_\pi\big(X_1 = x_0, X_0 = x_1\big) = \mathbb{P}_\pi\big(X_0 = x_1, X_1 = x_0\big),$$

soit $\pi(x_0)P(x_0, x_1) = \pi(x_1)P(x_1, x_0)$. □

Proposition 7.13 (Réversibilité et invariance) *Une mesure réversible pour un noyau markovien est invariante pour ce noyau.*

Démonstration : On vérifie immédiatement l'équation de Chapman-Kolmogorov (7.1) en utilisant la réversibilité (7.5) : pour une mesure π réversible, on a

$$(\pi P)(y) = \sum_{x\in E} \pi(x)P(x, y) = \sum_{x\in E} \pi(y)P(y, x) = \pi(y)\Big(\sum_{x\in E} P(y, x)\Big) = \pi(y),$$

ie. π est invariante. □

Exemple 7.14 (Mesures réversibles)

(1) Marche aléatoire sur $\mathbb{Z}$ avec $P(i, i+1) = p$, $P(i, i-1) = q = 1-p$ avec $p \in]0, 1[$ (cf. Exemple 6.38). La mesure définie par $\pi(i) = \big(\frac{p}{q}\big)^i$, $i \in \mathbb{Z}$, est réversible :

$$\pi(i)P(i, i+1) = \left(\frac{p}{q}\right)^i p = \frac{p^{i+1}}{q^i} = \frac{p^{i+1}}{q^{i+1}} q = \pi(i+1)P(i+1, i)$$

puis comme $P(i, j) = 0$ lorsque $|i-j| > 1$ alors $\pi(i)P(i, j) = 0 = \pi(j)P(j, i)$.

(2) Marche aléatoire sur un graphe de degré fini. La mesure définie par $\pi(x) = \mathrm{card}(A_x)$ est réversible : si $P(x, y) > 0$ et $P(y, x) > 0$

$$\pi(x)P(x, y) = \mathrm{card}(A_x)\frac{1}{\mathrm{card}(A_x)} = \mathrm{card}(A_y)\frac{1}{\mathrm{card}(A_y)} = \pi(y)P(y, x).$$

(3) Urne d'Ehrenfest sur $\{0, \dots, d\}$. Il s'agit de la chaîne de matrice stochastique

$$\begin{aligned}P(j, j+1) &= 1 - \frac{j}{d}, \quad 0 \le j \le d,\\ P(j, j-1) &= \frac{j}{d}, \quad 0 \le j \le d.\end{aligned}$$

Alors une mesure π est réversible si et seulement si

$$\pi(j)\Big(1 - \frac{j}{d}\Big) = \pi(j+1)\Big(\frac{j+1}{d}\Big) \quad \forall j \in \{0, \dots, d-1\}.$$

On vérifie sans peine que $\pi(j) = \binom{d}{j}$, $0 \le j \le d$, convient. Comme $\sum_{j=0}^d \binom{d}{j} = (1+1)^d = 2^d$, alors $\pi(j) = \binom{d}{j}/2^d$, $0 \le j \le d$, définit une probabilité invariante, c'est à dire $\pi = \mathcal{B}(d, 1/2)$.

Remarque 7.15 Pour trouver des mesures invariantes, il faut résoudre le système linéaire donné par l'équation de Chapman-Kolmogorov (7.1). En pratique, ce système peut être compliqué à résoudre (il est même infini si E l'est). Dans ce cas, il est intéressant de rechercher mieux en cherchant des mesures réversibles car l'équation (7.5) est en pratique plus simple à résoudre. On est alors assuré par la Proposition 7.13 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 7.16 (Convergence dominée) *Soit $(a(x))_{x\in E}\in\mathbb{R}^E$ une suite de réels positifs de somme finie et $(b_n(x))_{x\in E}$, $n\geq 1$, telle que, pour tout $x\in E$, $|b_n(x)|\leq 1$, $n\geq 1$ et $\lim_{n\to+\infty} b_n(x)=b(x)$. Alors on a*

$$\lim_{n\to+\infty}\sum_{x\in E} a(x)b_n(x)=\sum_{x\in E} a(x)b(x).$$

On commence par préciser le support d'une mesure ou probabilité invariante :

Proposition 7.17 (Support d'une mesure invariante) *Soit π une mesure invariante d'une chaîne de Markov.*

(1) Si $\pi(x)>0$ alors on a aussi $\pi(y)>0$ pour tout y tel que $x\rightsquigarrow y$.

(2) Si la chaîne est irréductible, le support de π est E.

(3) Si en plus π est une probabilité (invariante) alors $\pi(x)=0$ pour tout x transitoire ou récurrent nul, π ne charge que les états récurrents positifs et son support est une union de classes de récurrence positives (closes, irréductibles).

Démonstration : 1) Soit $x\in E$ tel que $\pi(x)>0$ et $x\rightsquigarrow y$. Il existe $n\geq 1$ tel que $P^n(x,y)>0$. L'équation de Chapman-Kolmogorov (7.1) $\pi=\pi P^n$ pour ce n donne alors

$$\pi(y)=\sum_{z\in E}\pi(z)P^n(z,y)\geq \pi(x)P^n(x,y)>0.$$

2) Comme $\pi\neq 0$, il existe $x\in E$ tel que $\pi(x)>0$ et comme par irréductibilité, x mène à tout y, on a aussi $\pi(y)>0$ pour tout $y\in E$.
3) On suppose que π est une probabilité invariante. Si x est transitoire ou récurrent nul alors $m_x=\mathbb{E}_x\big[\widetilde{T}_x\big]=+\infty$ et la Proposition 6.15 donne pour tout $z\in E$:

$$\lim_{n\to+\infty}\frac{G_n(z,x)}{n}=\frac{\rho_{z,x}}{m_x}=0. \tag{7.7}$$

Comme par invariance de π, on a pour $k \geq 1$: $\pi P^k = \pi$, ie.

$$\pi(x) = \sum_{z \in E} \pi(z) P^k(z, x),$$

en sommant sur $k = 0, \dots, n$ et en divisant par $n+1$, on obtient

$$\pi(x) = \sum_{z \in E} \pi(z) \frac{G_n(z, x)}{n+1}.$$

Comme π est une probabilité, le Lemme 7.16 s'applique et il permet de passer à la limite avec (7.7) pour obtenir lorsque x est transitoire ou récurrent nul :

$$\pi(x) = \lim_{n \to +\infty} \left(\sum_{z \in E} \pi(z) \frac{G_n(z, x)}{n+1} \right) = \sum_{z \in E} \pi(z) \left(\lim_{n \to +\infty} \frac{G_n(z, x)}{n+1} \right) = 0.$$

La probabilité π ne charge que des états récurrents positifs et comme d'après le **1)**, son support contient les classes de récurrence de ses éléments, le support est donc exactement une union de classes de récurrence positive. □

Remarque 7.18 — L'utilisation du Lemme 7.16 dans la preuve précédente exige que $\sum_{x \in E} \pi(x) < +\infty$ et, à normalisation près, que π soit une probabilité. Le 3) dans la Prop. 7.17 ne concerne donc que les probabilités invariantes (ou mesures invariantes finies mais pas les mesures invariantes de poids infinis !).
— Par conséquent, une chaîne qui n'a pas d'états récurrents positifs n'a pas de probabilité invariante (une probabilité ne peut pas être concentrée sur des points qu'elle ne charge pas !).

Proposition 7.19 (Invariance et transience/récurrence nulle) *Pour une chaîne de Markov transitoire ou récurrente nulle, une mesure invariante est de poids infini. En particulier, l'espace d'états doit être infini et il n'y a pas de probabilité invariante.*

Démonstration : Si π est une mesure invariante de poids $\pi(E) < +\infty$, alors $\widetilde{\pi} = \pi/\pi(E)$ serait une probabilité. D'après 3) dans la Prop. 7.17, $\widetilde{\pi}$ ne charge pas les états transitoires ou récurrents nuls, ce qui est absurde puisqu'il n'y a que de tels états. On doit donc avoir $\pi(E) = +\infty$. Le reste en découle facilement. □

Exemple 7.20 (Mesure invariante d'une chaîne transitoire) On reprend l'exemple de la marche aléatoire simple sur $\mathbb{Z}$ avec $P(i, i+1) = p$, $P(i, i-1) = q = 1-p$ avec $p \in]0, 1[$. D'après l'Exemple 6.38, la marche est transitoire si $p \neq 1/2$. D'après l'Exemple 7.14, la mesure donnée par $\pi(i) = \left(\frac{p}{q}\right)^i$, $i \in \mathbb{Z}$, est invariante (car réversible) :

On montre maintenant qu'on peut associer une mesure invariante à tout état x récurrent 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 de mesures invariantes. On rappelle que $\widetilde{T}_x = \inf(n > 0 : X_n = x)$ (date de premier retour en x sous $\mathbb{P}_x$) et on pose pour $y \in E$:

$$\begin{align} \nu_x(y) &= \mathbb{E}_x\left[\sum_{k=0}^{\widetilde{T}_x-1} \mathbf{1}_{\{X_k=y\}}\right] \tag{7.8} \\ &= \mathbb{E}_x\left[\sum_{k=1}^{\widetilde{T}_x} \mathbf{1}_{\{X_k=y\}}\right]. \tag{7.9} \end{align}$$

L'égalité entre (7.9) et (7.8) vient de ce que sous $\mathbb{P}_x$, on a $X_0 = X_{\widetilde{T}_x} = x$. À titre d'exemple, lorsque x est absorbant, on a $\nu_x = \delta_x$. Pour étudier la mesure ν_x, on utilise le lemme technique suivant :

Lemme 7.21 *Pour tout $p \geq 0$, et $x, y \in E$, on a :*

$$\mathbb{E}_x\left[\sum_{k=0}^{(p+1)\wedge(\widetilde{T}_x-1)} \mathbf{1}_{\{X_k=y\}}\right] = \sum_{z\in E} \mathbb{E}_x\left[\sum_{k=0}^{p\wedge(\widetilde{T}_x-1)} \mathbf{1}_{\{X_k=z\}}\right] P(z,y) + \mathbb{P}_x\big(\widetilde{T}_x > p+1\big)\,\delta_{x,y}. \tag{7.10}$$

Démonstration : D'abord, on observe que

$$\begin{align*} \sum_{k=0}^{(p+1)\wedge(\widetilde{T}_x-1)} \mathbf{1}_{\{X_k=y\}} - \sum_{k=1}^{(p+1)\wedge\widetilde{T}_x} \mathbf{1}_{\{X_k=y\}} &= \mathbf{1}_{\{X_0=y\}} - \mathbf{1}_{\{X_{\widetilde{T}_x}=y\}} \mathbf{1}_{\{\widetilde{T}_x\leq p+1\}} \\ &= \begin{cases} 0 & \text{si } y \neq x \\ 1 - \mathbf{1}_{\{\widetilde{T}_x\leq p+1\}} = \mathbf{1}_{\{\widetilde{T}_x> p+1\}} & \text{si } y = x. \end{cases} \end{align*}$$

On a alors

$$\begin{align} \mathbb{E}_x\left[\sum_{k=0}^{(p+1)\wedge(\widetilde{T}_x-1)} \mathbf{1}_{\{X_k=y\}}\right] &= \mathbb{E}_x\left[\sum_{k=1}^{(p+1)\wedge\widetilde{T}_x} \mathbf{1}_{\{X_k=y\}} + \mathbf{1}_{\{\widetilde{T}_x>p+1\}}\,\delta_{x,y}\right] \notag \\ &= \mathbb{E}_x\left[\sum_{k=0}^{p\wedge(\widetilde{T}_x-1)} \mathbf{1}_{\{X_{k+1}=y\}}\right] + \mathbb{P}_x(\widetilde{T}_x > p+1)\,\delta_{x,y}. \tag{7.11} \end{align}$$

Par la propriété de Markov faible (5.32) :

$$\mathbb{E}_x\left[\sum_{k=0}^{p\wedge(\widetilde{T}_x-1)} \mathbf{1}_{\{X_{k+1}=y\}}\right] = \sum_{z\in E}\sum_{k=0}^{p} \mathbb{E}_x\left[\mathbf{1}_{\{X_k=z\}} \mathbf{1}_{\{\widetilde{T}_x-1\geq k\}} \mathbf{1}_{\{X_{k+1}=y\}}\right]$$

$$= \sum_{z\in E}\sum_{k=0}^{p}\mathbb{E}_x\Big[\underbrace{\mathbf{1}_{\{X_k=z\}}\mathbf{1}_{\{\widetilde{T}_x-1\geq k\}}}_{\mathcal{F}_k\text{-mesurable}}\mathbb{E}_x[\mathbf{1}_{\{X_{k+1}=y\}}|\mathcal{F}_k]\Big]$$

$$= \sum_{z\in E}\sum_{k=0}^{p}\mathbb{E}_x\Big[\mathbf{1}_{\{X_k=z\}}\mathbf{1}_{\{\widetilde{T}_x-1\geq k\}}\underbrace{\mathbb{E}_{X_k}[\mathbf{1}_{\{X_1=y\}}]}_{\mathbb{E}_z[\mathbf{1}_{\{X_1=y\}}]=P(z,y)}\Big]$$

$$= \sum_{z\in E}\sum_{k=0}^{p}\mathbb{E}_x\Big[\mathbf{1}_{\{X_k=z\}}\mathbf{1}_{\{\widetilde{T}_x-1\geq k\}}\Big]P(z,y)$$

$$= \sum_{z\in E}\mathbb{E}_x\left[\sum_{k=0}^{p\wedge(\widetilde{T}_x-1)}\mathbf{1}_{\{X_k=z\}}\right]P(z,y). \tag{7.12}$$

Les égalités (7.11) et (7.12) concluent à (7.10). □

Proposition 7.22 (Mesure invariante d'un état récurrent) *On considère une chaîne de Markov $(X_n)_{n\geq 0}$ et x un état récurrent de la chaîne (s'il y en a !).*

(1) La mesure ν_x est invariante et $\nu_x(x) = 1$.

(2) La mesure ν_x a pour support la classe de récurrence de x : $\nu_x(y) > 0$ si et seulement si y appartient à la même classe de récurrence que x.

Démonstration : D'abord puisque sous $\mathbb{P}_x$, $X_0 = x$ et $X_k \neq x$ pour $1 \leq k \leq \widetilde{T}_x - 1$, on a $\sum_{k=0}^{\widetilde{T}_x-1}\mathbf{1}_{\{X_k=x\}} = 1$ et donc $\nu_x(x) = 1$ par la définition (7.8). Ensuite, si y n'est pas dans la classe de récurrence de x, alors x et y ne communiquent pas et $G(x,y) = 0$, d'où il vient :

$$0 = G(x,y) = \mathbb{E}_x\big[N_y\big] = \mathbb{E}_x\left[\sum_{k=0}^{+\infty}\mathbf{1}_{\{X_k=y\}}\right] \geq \mathbb{E}_x\left[\sum_{k=0}^{\widetilde{T}_x-1}\mathbf{1}_{\{X_k=y\}}\right] = \nu_x(y)$$

et donc nécessairement $\nu_x(y) = 0$. À ce stade, le support de ν_x est inclus dans la classe de récurrence de x. Lorsque l'invariance de ν_x sera établie, le 1) de la Prop. 7.17 donnera que le support de ν_x est exactement la classe de récurrence de x.

L'équation de Chapman-Kolmogorov (7.1) pour ν_x ($\nu_x = \nu_x P$) découle de (7.10) dans le Lemme 7.21 en passant à la imite $p \to +\infty$ par convergence monotone, en notant que comme x est récurrent, on a $\mathbb{P}_x(\widetilde{T}_x < +\infty) = 1$ et donc $\mathbb{P}_x(\widetilde{T}_x > p+1) \xrightarrow[p\to+\infty]{} 0$.

On achève de montrer que ν_x est une mesure invariante en montrant que $\nu_x(y) < +\infty$ pour tout $y \in E$:

— si y n'est pas dans la classe de x alors on a vu que $\nu_x(y) = 0$;

— si $y \sim x$ alors il existe $m \geq 1$ tel que $P^m(y,x) > 0$ et en itérant pour ce m la relation $\nu_x = \nu_x P$ obtenue précédemment, on a $\nu_x = \nu_x P^m$ et

$$\nu_x(x) = \sum_{z\in E}\nu_x(z)P^m(z,x) \geq \nu_x(y)P^m(y,x);$$

comme $\nu_x(x) = 1$ et $P^m(y,x) > 0$, cela exige $\nu_x(y) < +\infty$.
Finalement, $\nu_x(y) < +\infty$ pour tout $y \in E$ et ν_x est bien invariante (Définition 7.1). $\square$

Pour relier les mesures invariantes aux mesures ν_x, on commence par :

Proposition 7.23 *Soit $(X_n)_{n\geq 0}$ une chaîne de Markov admettant un état récurrent x. Si π est une mesure invariante alors pour tout $y \in E$, on a $\pi(y) \geq \pi(x)\nu_x(y)$ où ν_x est associée en (7.8) à l'état récurrent x. De plus si $y \rightsquigarrow x$, alors il y a égalité : $\pi(y) = \pi(x)\nu_x(y)$*

Démonstration : À l'aide du Lemme 7.21, on commence par montrer par récurrence que pour tout entier $p \geq 0$, et tout état $y \in E$, on a

$$\pi(y) \geq \pi(x)\,\mathbb{E}_x\left[\sum_{k=0}^{p\wedge(\widetilde{T}_x-1)} \mathbf{1}_{\{X_k=y\}}\right]. \tag{7.13}$$

D'abord, l'inégalité (7.13) est facilement vérifiée si $y = x$ ou $p = 0$:

— si $y = x$ alors $\sum_{k=0}^{p\wedge(\widetilde{T}_x-1)} \mathbf{1}_{\{X_k=x\}} = 1$ (terme d'indice 0) et l'inégalité (7.13) se réduit à $\pi(x) \geq \pi(x) \times 1$ (qui est vraie) ;

— si $p = 0$ alors $\sum_{k=0}^{p\wedge(\widetilde{T}_x-1)} \mathbf{1}_{\{X_k=y\}} = \mathbf{1}_{\{X_0=y\}}$ et l'espérance dans (7.13) vaut $\mathbb{P}_x(X_0 = y) = \delta_{x,y}$ (7.13) se réduit à $\pi(y) \geq \pi(x)\delta_{x,y}$ (qui est vraie).

Pour le cas général ($y \neq x$ et $p \geq 1$), on procède par récurrence sur p et on suppose que (7.13) est vraie pour un entier p fini et $y \in E$. Par le Lemme 7.21, on a

$$\begin{aligned}
\pi(x)\,\mathbb{E}_x\left[\sum_{k=0}^{(p+1)\wedge(\widetilde{T}_x-1)} \mathbf{1}_{\{X_k=y\}}\right] &= \sum_{z\in E} \pi(x)\mathbb{E}_x\left[\sum_{k=0}^{p\wedge(\widetilde{T}_x-1)} \mathbf{1}_{\{X_k=z\}}\right] P(z,y) \\
&\leq \sum_{z\in E} \pi(z)P(z,y) = \pi(y)
\end{aligned}$$

par hypothèse de récurrence (7.13) et invariance de π.
En faisant $p \to +\infty$ par convergence monotone dans (7.13), on obtient

$$\pi(y) \geq \pi(x)\,\mathbb{E}_x\left[\sum_{k=0}^{\widetilde{T}_x-1} \mathbf{1}_{\{X_k=y\}}\right] = \pi(x)\nu_x(y). \tag{7.14}$$

Pour la deuxième partie, en combinant alors l'invariance de π, (7.14), l'invariance de ν_x et $\nu_x(x) = 1$ (Proposition 7.22), on a pour tout $n \geq 1$:

$$\begin{aligned}
\pi(x) &= \sum_{z\in E} \pi(z)P^n(z,x) \\
&\geq \sum_{z\in E} \pi(x)\nu_x(z)P^n(z,x) \qquad (7.15)
\end{aligned}$$

$$= \quad \pi(x)\,\nu_x(x) = \pi(x). \tag{7.16}$$

Cela exige l'égalité ci-dessus dans (7.15) et donc dans (7.14) dès que $P^n(z,x) > 0$. Ainsi pour $y \rightsquigarrow x$, il existe un tel n avec $P^n(y,x) > 0$, ce qui assure $\pi(y) = \pi(x)\nu_x(y)$ dès que $P^n(y,x) > 0$. □

Théorème 7.24 (Mesure invariante et classe de récurrence) *On considère une chaîne de Markov $(X_n)_{n\geq 0}$. Sur une classe de récurrence E_{R_i} (close, irréductible), il y a unicité à facteur multiplicatif près de la mesure invariante, en particulier π invariante s'écrit $\pi = \pi(x)\nu_x$ pour tout x de la classe de récurrence. On a alors l'alternative :*

(1) Si ces mesures sont de poids finis, il y a une unique probabilité invariante sur la classe de récurrence et elle est récurrente positive ; la probabilité invariante est donnée par

$$\forall x \in E_{R_i}, \qquad \pi(x) = \frac{1}{\mathbb{E}_x\big[\widetilde{T}_x\big]}. \tag{7.17}$$

(2) Si ces mesures sont de poids infinis, il n'y a pas de probabilité invariante sur la classe de récurrence et elle est récurrente nulle.

Démonstration : Étant donné un état x quelconque de la classe de récurrence E_{R_i} considérée, pour tout y de la classe, on a $y \rightsquigarrow x$. D'après le Lemme 7.23, on a $\pi(y) = \pi(x)\nu_x(y)$. Comme ν_x est concentrée sur la classe de récurrence (Proposition 7.22) et π aussi (par hypothèse), on a bien $\pi = \pi(x)\nu_x$. Les mesures invariantes sur E_{R_i} sont donc proportionnelles ; en particulier, elles sont toutes de poids finis ou toutes de poids infinis.

1) S'il existe une mesure invariante finie sur la classe récurrente considérée, elles le sont toutes et on note π l'unique probabilité invariante et ν_x la mesure invariante associée à un état x (récurrent !) en (7.8)

$$\nu_x(y) = \mathbb{E}_x\left[\sum_{k=0}^{\widetilde{T}_x-1} \mathbf{1}_{\{X_k=y\}}\right].$$

Comme π et ν_x sont proportionnelles (début de la démonstration), π est une probabilité et $\nu_x(x) = 1$ (Proposition 7.22), on a

$$\pi = \pi(x)\nu_x = \frac{\nu_x}{\nu_x(E)}.$$

On a donc $\pi(x) \neq 0$ et

$$\pi(x) = \frac{1}{\nu_x(E)}.$$

De plus, par convergence monotone :

$$\nu_x(E) \quad = \quad \sum_{y\in E} \nu_x(y) = \sum_{y\in E} \mathbb{E}_x\left[\sum_{k=0}^{\widetilde{T}_x-1} \mathbf{1}_{\{X_k=y\}}\right]$$

$$= \mathbb{E}_x\left[\sum_{k=0}^{\widetilde{T}_x-1}\Big(\sum_{y\in E}\mathbf{1}_{\{X_k=y\}}\Big)\right] = \mathbb{E}_x\big[\widetilde{T}_x\big] \tag{7.18}$$

puisque $\sum_{y\in E}\mathbf{1}_{\{X_k=y\}} = 1$. On a donc $\mathbb{E}_x\big[\widetilde{T}_x\big] < +\infty$ et x est récurrent positif et donc la classe de récurrence est récurrente positive.

2) Dans le cas alternatif où toute les mesures invariantes sont de poids infinis, alors ν_x est de poids infini et donc $\mathbb{E}_x\big[\widetilde{T}_x\big] = +\infty$ par le même calcul (7.18) que juste précédemment : x est récurrent nul, et sa classe est donc une classe de récurrente nulle. □

Chaînes de Markov irréductibles

Dans le cas irréductible, il n'y a qu'une classe de récurrence et le Théorème 7.24 se spécialise en le théorème suivant qui dresse un bilan des résultats précédents pour l'invariance d'une chaîne Markov irréductible. Attention, on rappelle que d'après l'Exemple 7.10, il n'existe pas toujours de mesure invariante, même pour une chaîne irréductible.

Théorème 7.25 (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 (tous les états sont transitoires) : toute mesure invariante π est de poids infini et il n'y a pas de probabilité invariante ;

(2) La chaîne est récurrente nulle (tous les états sont récurrents nuls : pour tout $x \in E$, $\mathbb{E}_x\big[\widetilde{T}_x\big] = +\infty$) : les mesures invariantes sont toutes proportionnelles et de poids infinis ($\pi(E) = +\infty$) ; il n'existe alors pas de probabilité invariante ;

(3) La chaîne est récurrente positive (tous les états sont récurrents positifs : pour tout $x \in E$, $\mathbb{E}_x\big[\widetilde{T}_x\big] < +\infty$) : les mesures invariantes sont toutes proportionnelles et de poids finis ($\pi(E) < +\infty$), il existe une unique probabilité invariante et elle est donnée par

$$\forall x \in E, \quad \pi(x) = \frac{1}{\mathbb{E}_x\big[\widetilde{T}_x\big]}.$$

Remarque 7.26 (Probabilité invariante) — En particulier, une chaîne de Markov irréductible sur un espace d'états fini est récurrente positive et admet donc une unique probabilité invariante.

— On peut préciser 3) dans la Prop. 7.17 : Quand elle existe, l'unique probabilité invariante est donnée par $\pi(x) = 1/\mathbb{E}_x\big[\widetilde{T}_x\big]$, $x \in E$ (et $\pi(x) > 0$ si et seulement si x est récurrent positif). Elle est concentrée sur les états récurrents positifs.

On déduit du Théorème 7.25 la caractérisation suivante pour l'existence et unicité de la probabilité invariante

Corollaire 7.27 (Existence et unicité de probabilité invariante) *Pour une chaîne de Markov irréductible, les assertions suivantes sont équivalentes :*

(1) il existe une (unique) probabilité invariante π ;

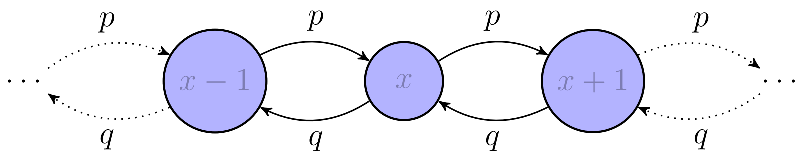
(2) la mesure π définie, pour $x \in E$ par $\pi(x) = 1/\mathbb{E}_x\big[\widetilde{T}_x\big]$ est la probabilité invariante ;

(3) il existe un état récurrent positif ;

(4) tous les états sont récurrents positifs.

Démonstration : (1)⇒(2)⇒(3) par le Théorème 7.25 et la Remarque 7.26 ; (3)⇒(4) par irréductibilité ; (4)⇒(1) par le Théorème 7.25. □

Exemple 7.28 (Marche aléatoire simple sur $\mathbb{Z}$) On considère la marche aléatoire sur $\mathbb{Z}$ aux plus proches voisins avec $P(x, x+1) = 1-P(x, x-1) = p$ et $P(x, y) = 0$ si $y \neq x \pm 1$.

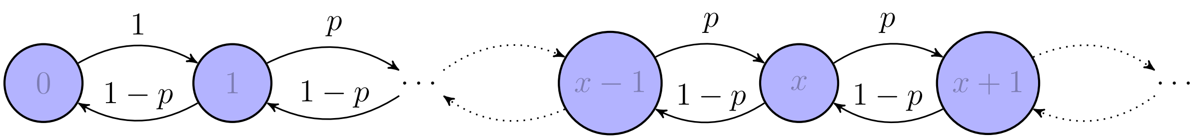


— D'après l'Exemple 6.38, la marche aléatoire non-symétrique ($p \neq 1/2$) est transitoire.

— D'après l'Exemple 7.5, π uniforme sur $\mathbb{Z}$ est une mesure invariante. Comme cette mesure est de poids infini, le Théorème 7.25 montre alors que cette chaîne est récurrente nulle.

Exemple 7.29 (Chaîne de naissance et de mort réfléchie) On considère la chaîne de Markov irréductible de transition

$$P(k, k+1) = p, \quad P(k, k-1) = 1-p, \quad P(0,1) = 1.$$



On vérifie que la mesure $\pi = \left(\left(\frac{p}{1-p}\right)^k\right)_{k\geq 0}$ est réversible donc invariante.

— Si $p < 1/2$ alors π est de poids fini et la chaîne est récurrente positive par le Théorème 7.25. La probabilité invariante correspondante, obtenue en normalisant π, est la loi géométrique $\mathcal{G}(p/(1-p))$ sur $\mathbb{N}$.

— Si $p \geq 1/2$ alors π est de poids infini et la chaîne est récurrente nulle ou transitoire. Pour trancher, il faut par exemple étudier en détails la probabilité de retour en 0 et on montre que

  — pour $p = 1/2$: la chaîne est récurrente nulle ;

  — pour $p > 1/2$: la chaîne est transitoire.

Chaînes de Markov non irréductibles

De façon générale, une chaîne de Markov admet une classification (6.31) 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 7.30 (Mesures invariantes pour les chaînes non irréductibles) *Soit $(X_n)_{n\geq 0}$ une chaîne de Markov non irréductible.*

(1) Sur chaque classe de récurrence, il existe une mesure invariante (unique à facteur multiplicatif près).

(2) Il existe une probabilité invariante sur cette classe si et seulement si la classe est récurrente positive et elle est alors donnée par $\pi(x) = 1/\mathbb{E}_x\big[\widetilde{T}_x\big]$ pour tout état x de la classe.

(3) L'ensemble des probabilités invariantes est donné par les combinaisons convexes des probabilités invariantes de chaque classe de récurrence positive.

Démonstration : (1) Toute mesure ν_x associée à un x de la classe par (7.8) est invariante sur la classe par la Proposition 7.22. Puis par le 1) dans le Théorème 7.24, toutes les mesures invariantes concentrées sur une classe de récurrence sont proportionnelles et s'écrivent $\pi = \pi(x)\nu_x$, pour tout état x de la classe.

(2) Cela découle de l'alternative du Théorème 7.24.

(3) Il est immédiat que toute combinaison convexe de probabilités invariantes est une probabilité invariante (l'équation de Chapman-Kolmogorov (7.1) et être une probabilité sont des notions stables par combinaison convexe).

Par le Théorème 7.24, il existe une probabilité invariante sur chaque classe de récurrence positive. Si π est une probabilité invariante arbitraire, il existe x, nécessairement récurrent positif par la Prop. 7.17 tel que $\pi(x) > 0$. On note E_{R_i} la classe de récurrence (positive) à laquelle x appartient et on montre que la restriction $\pi_i = \pi_{|E_{R_i}}$ de π à la classe E_{R_i} de x reste invariante. En effet, pour $y \in E_{R_i}$, on a

$$\sum_{z\in E} \pi_i(z)P(z,y) = \sum_{z\in E_{R_i}} \pi(z)P(z,y) = \sum_{z\in E} \pi(z)P(z,y) = \pi(y) = \pi_i(y)$$

où la première égalité vient de la restriction à E_{R_i}, la deuxième de ce qu'un état z qui peut mener à $y \in E_{R_i}$ est soit transitoire (et alors $\pi(z) = 0$ par la Prop. 7.17) soit dans la classe de récurrence E_{R_i}, les autres ne communiquent pas avec y, la troisième égalité vient de l'invariance de π. Puis pour $y \notin E_{R_i}$, on a $\pi_i(y) = 0$ et $P(z,y) = 0$ pour $z \in E_{R_i}$ si bien que

$$\sum_{z\in E} \pi_i(z)P(z,y) = \sum_{z\in E_{R_i}} \pi(z)P(z,y) = 0 = \pi_i(y),$$

assurant l'équation de Chapman-Kolmogorov $\pi_i = \pi_i P$ et l'invariance de π_i.

Comme par ailleurs pour tout $x_i \in E_{R_i}$, ν_{x_i} est une mesure invariante (finie) de support E_{R_i}, l'unicité à facteur multiplicatif près des mesures invariantes donnée par le Théorème 7.24 assure alors que $\pi_i = \pi(x_i)\nu_x = \pi(x_i)\nu_{x_i}(E)\widetilde{\nu}_{x_i}$ en notant $\widetilde{\nu}_{x_i} = \nu_{x_i}/\nu_{x_i}(E)$ la probabilité associée à ν_{x_i}. Finalement, comme π est concentrée sur les classes de récurrence positive (Prop. 7.17) qui sont disjointes, on a

$$\pi = \sum_{i \in I_+} \pi_i = \sum_{i \in I_+} \big(\pi(x_i)\nu_{x_i}(E)\big)\widetilde{\nu}_{x_i} \tag{7.19}$$

Comme on a des probabilités, il vient $\sum_{i \in I_+} \pi(x_i)\nu_{x_i}(E) = 1$ et (7.19) prouve que π est combinaison convexe des $\widetilde{\nu}_{x_i}$, $i \in I^+$. □

On déduit immédiatement du 3) dans le Théorème 7.30 que l'unicité d'une probabilité invariante est équivalente à l'unicité d'une classe de récurrence positive.

Corollaire 7.31 (Existence et unicité des probabilités invariantes) *Il existe une unique probabilité invariante si et seulement s'il existe une unique classe de récurrence positive. Dans ce cas, la probabilité invariante est donnée par $\pi(x) = 1/\mathbb{E}_x\big[\widetilde{T}_x\big]$, $x \in E$.*

7.3 Périodicité et forte irréductibilité

Périodicité

Définition 7.32 (Période) *On appelle période de l'état $x \in E$ d'une chaîne de Markov de matrice stochastique P l'entier*

$$d_x = PGCD\big(n \geq 1 \,:\, P^n(x,x) > 0\big)$$

avec la convention $d_x = 0$ si $P^n(x,x) = 0$ pour tout $n \geq 1$. Si $d_x = 1$, on dit que l'état x est **apériodique***.*

Exemple 7.33 (Période)

— Pour la marche aléatoire sur $\mathbb{Z}$, on a vu que $P^{2n+1}(0,0) = 0$ pour tout $n \geq 0$ et $P^{2n}(0,0) > 0$ pour tout $n \geq 1$ (Exemple 6.38). On a donc $d_0 = 2$.

— De même dans le modèle de l'urne d'Ehrenfest (Exemple 5.10, exercice).

Proposition 7.34 (Période commune des états communiquants) *Si $x \sim y$ alors $d_x = d_y$.*

Démonstration : Si $x \sim y$ alors il existe des entiers $n \geq 1$ et $m \geq 1$ tels que $P^n(x,y) > 0$ et $P^m(y,x) > 0$. Dès lors

$$P^{m+n+Nk}(x,x) \geq P^n(x,y)\big(P^k(y,y)\big)^N P^m(y,x).$$

Donc pour tout $k \geq 0$ tel que $P^k(y,y) > 0$ on a $P^{m+n+Nk}(x,x) > 0$, ie. d_x divise $n+m+Nk$ pour tout N ($d_x \mid n+m+Nk$). En particulier d_x divise k ($d_x \mid k$).

On en déduit que d_x divise d_y, et par symétrie d_y divise d_x, soit finalement $d_x = d_y$. $\square$

D'après la Proposition 7.34, la définition qui suit à un sens :

Définition 7.35 (Période) *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.*

Forte irréductibilité

Définition 7.36 (Forte irréductibilité) *Une chaîne de Markov $(X_n)_{n\geq 0}$ de matrice stochastique P est dite fortement irréductible s'il existe $k \geq 1$ tel que pour tout $x, y \in E$, on a $P^k(x, y) > 0$.*

D'après la Définition 6.26, un noyau P est irréductible si pour tout $x, y \in E$, il existe un chemin fini de x à y : $\exists n \geq 1$, $(x_i)_{0\leq i\leq n} \in E^n$ avec $x_0 = x$, $x_n = y$ et

$$P^n(x, y) \geq P(x, x_1)P(x_1, x_2) \dots P(x_{n-1}, y) > 0.$$

Il y a forte irréductibilité quand il existe une **longueur commune** de chemin reliant tout $x, y \in E$.

Proposition 7.37 *Une chaîne de Markov fortement irréductible est irréductible et apériodique.*

Démonstration : Il est immédiat que la forte irréductibilité implique l'irréductibilité. On considère $(x, y) \in E^2$ tel que $P(y, x) > 0$. Avec l'indice k de la Déf. 7.36 de la forte irréductibilité, on a $P^k(x, x) > 0$ et $P^k(x, y) > 0$. On a donc

$$P^{k+1}(x, x) \geq P^k(x, y)P(y, x) > 0.$$

On a donc $k, k+1 \in R(x) := \{n \in \mathbb{N} : P^n(x, x) > 0\}$, ce qui assure $d_x = \mathrm{PGCD}(R(x)) = 1$ et donc la chaîne est apériodique. $\square$

Proposition 7.38 *Soit $(X_n)_{n\geq 0}$ chaîne de Markov irréductible et apériodique. Alors pour tout $x \in E$, il existe $n(x) \geq 1$ tel que pour tout $n \geq n(x)$, on a $P^n(x, x) > 0$.*

Démonstration : Par irréductibilité, pour tout $x, y \in E$, il existe $n(x, y) \geq 1$ tel que $P^{n(x,y)}(x, y) > 0$. Comme la chaîne est apériodique, il existe $n_1, \dots, n_k \in R(x)$ des entiers de PGCD égale à 1. Par le théorème de Bézout, il existe $q_1, \dots, q_k \in \mathbb{Z}$ tels que $q_1n_1 + \dots + q_kn_k = 1$. On note

$$a(x) = \sum_{i\,:\,q_i>0} q_in_i, \quad b(x) = -\sum_{i\,:\,q_i<0} q_in_i \geq 0,$$

de sorte que $a(x) = b(x) + 1$. Comme $n_i \in R(x)$, on a

$$P^{a(x)}(x,x) \geq \prod_{i\,:\,q_i>0} P^{n_i}(x,x)^{q_i} > 0, \quad P^{b(x)}(x,x) \geq \prod_{i\,:\,q_i<0} P^{n_i}(x,x)^{-q_i} > 0.$$

On a donc $b(x), b(x)+1 \in R(x)$. Pour tout $n \in \mathbb{N}$, on écrit la division euclidienne de n par $b(x)$:

$$n = qb(x) + r = (q-r)b(x) + ra(x) \tag{7.20}$$

avec $0 \leq r \leq b(x) - 1$. On pose alors $n(x) = b(x)^2 - 1$. Pour $n \geq n(x)$, on doit avoir $q \geq r$ car $ra(x) \leq (b(x)-1)(b(x)+1) = b(x)^2 - 1 = n(x)$. De l'écriture (7.20) de n, i suit

$$P^n(x,x) \geq P^{b(x)}(x,x)^{q-r} P^{a(x)}(x,x)^r > 0,$$

puisque $a(x), b(x) \in R(x)$. □

En fait, quand E est fini, on la caractérisation suivante de la forte irréductibilité :

Proposition 7.39 *Si E est fini, la forte irréductibilité est équivalente à l'irréductibilité plus l'apériodicité.*

Démonstration : On suppose $(X_n)_{n\geq 0}$ irréductible, apériodique sur E fini et il s'agit de montrer que $(X_n)_{n\geq 0}$ est fortement irréductible. Pour cela, avec les notations de la Prop. 7.38, soit $k = \sup\big(n(x) + n(x,y) : x, y \in E\big)$. Observer que k est fini car l'espace d'état E est fini. Pour $x, y \in E$ quelconques, on écrit $k = n(x) + j + n(x,y)$ avec $j = j(x,y) \geq 0$ et d'après la Prop. 7.38, on a

$$P^k(x,y) \geq P^{n(x)+j}(x,x) P^{n(x,y)}(x,y) > 0.$$

□

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 E dénombrable, de matrice stochastique P.

Théorème 7.40 (Convergence vers l'équilibre) *On considère une chaîne de Markov $(X_n)_{n\geq 0}$ irréductible récurrente positive et apériodique. Alors si π désigne l'unique probabilité invariante (Théorème 7.24), pour tout $x \in E$:*

$$\lim_{n\to+\infty} \sum_{y\in E} \Big|\mathbb{P}_x(X_n = y) - \pi(y)\Big| = 0. \tag{7.21}$$

On a aussi pour toute loi initiale ν

$$\lim_{n\to+\infty} \sum_{y\in E} \Big|\mathbb{P}_\nu(X_n = y) - \pi(y)\Big| = 0. \tag{7.22}$$

Remarque 7.41 — Les convergences (7.21) et (7.22) impliquent en particulier

$$\mathbb{P}_x(X_n = y) \xrightarrow{n\to+\infty} \pi(y) \quad \text{et} \quad \mathbb{P}_\nu(X_n = y) \xrightarrow{n\to+\infty} \pi(y)$$

c'est à dire la convergence en loi de $(X_n)_{n\geq 0}$ vers π sous $\mathbb{P}_x$ ou sous $\mathbb{P}_\nu$.

— Noter que $\mathbb{P}_x(X_n = y) = P^n(x,y)$ on a donc $P^n(x,y) \xrightarrow{n\to+\infty} \pi(y)$. De là vient que sous les conditions du Théorème 7.40, on a la convergence de P^n vers P^∞ dont toutes les lignes valent π.

— Le Théorème 7.40 implique que la convergence est uniforme en $y \in E$.

— La convergence (7.21) énonce en fait la convergence en variation totale de la loi $\mathcal{L}(X_n|X_0 = x)$ vers la probabilité invariante π : $\mu_n \xrightarrow{var} \pi$ sous $\mathbb{P}_x$.

La démonstration du Th. 7.40 utilise un argument de couplage fondé sur le lemme suivant :

Lemme 7.42 (Couplage) *Soit P une matrice stochastique sur E. Alors :*

(i) On définit une matrice stochastique $\overline{P} = P \otimes P$ sur E^2 avec

$$\overline{P}\big((x_1,x_2),(y_1,y_2)\big) = P(x_1,y_1)P(x_2,y_2), \tag{7.23}$$

et on a

$$\overline{P}^n\big((x_1,x_2),(y_1,y_2)\big) = P^n(x_1,y_1)P^n(x_2,y_2). \tag{7.24}$$

(ii) Si P est irréductible apériodique alors $\overline{P}$ aussi.

(iii) Si π est une probabilité invariante pour P alors $\pi\otimes\pi = (\pi(x)\pi(y))_{(x,y)\in E^2}$ en est une de $\overline{P}$.

(iv) Si $\big(\overline{X}_n\big)_{n\geq 0} = \big((X_n^{(1)}, X_n^{(2)})\big)_{n\geq 0}$ est une chaîne de Markov sur E^2 de matrice stochastique $\overline{P}$ et de loi initiale ν sur E^2 alors $(X_n^{(i)})_{n\geq 0}$ est une chaîne de Markov sur E de matrice stochastique P et de loi initiale ν_i (où $\nu_1(A) = \nu(A\times E)$ et $\nu_2(B) = \nu(E\times B)$), ie.

$$\overline{\mathbb{P}}_\nu\big(X_n^{(i)} = x\big) = \mathbb{P}_{\nu_i}(X_n = x) \quad i = 1, 2, \tag{7.25}$$

en notant de façon générique $(X_n)_{n\geq 0}$ une chaîne de Markov de matrice stochastique P.

Démonstration : i) Pour tout $(x_1,x_2) \in E^2$, on a :

$$\begin{aligned}
\sum_{(y_1,y_2)\in E\times E} \overline{P}\big((x_1,x_2),(y_1,y_2)\big) &= \sum_{(y_1,y_2)\in E\times E} P(x_1,y_1)P(x_2,y_2) \\
&= \Big(\sum_{y_1\in E} P(x_1,y_1)\Big)\Big(\sum_{y_2\in E} P(x_2,y_2)\Big) = 1\times 1 = 1,
\end{aligned}$$

ce qui justifie que $\overline{P}$ est une matrice stochastique car en plus $\overline{P}\big((x_1,x_2),(y_1,y_2)\big) \geq 0$. Puis on prouve (7.24) par récurrence : l'initialisation est due à (7.23) ; ensuite, en supposant (7.24) vraie pour l'entier $n-1$, on le montre pour l'entier n :

$$\begin{aligned}
\overline{P}^n\big((x_1,x_2),(y_1,y_2)\big) &= \sum_{(z_1,z_2)\in E^2} \overline{P}^{n-1}\big((x_1,x_2),(z_1,z_2)\big)\overline{P}\big((z_1,z_2),(y_1,y_2)\big)\\
&= \sum_{(z_1,z_2)\in E^2} P^{n-1}(x_1,z_1)P^{n-1}(x_2,z_2)P(z_1,y_1)P(z_2,y_2)\\
&= \left(\sum_{z_1\in E} P^{n-1}(x_1,z_1)P(z_1,y_1)\right)\left(\sum_{z_2\in E} P^{n-1}(x_2,z_2)P(z_2,y_2)\right)\\
&= P^n(x_1,y_1)P^n(x_2,y_2).
\end{aligned}$$

ii) Pour tout $(x_1,x_2),(y_1,y_2) \in E^2$, par irréductibilité de P, il existe $m_1 \geq 1$ tel que $P^{m_1}(x_1,y_1) > 0$ et par irréductibilité et apériodicité (Proposition 7.38), il existe $n_1 \geq 1$ tel que $P^{n_1+k}(x_1,x_1) > 0$ pour tout $k \geq 0$. De la même façon, 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 $k \geq 0$. Alors pour tout $k \geq 0$, on a :

$$\begin{aligned}
P^{n_1+m_1+k}(x_1,y_1) &\geq P^{n_1+k}(x_1,x_1)P^{m_1}(x_1,y_1) > 0,\\
P^{n_2+m_2+k}(x_2,y_2) &\geq P^{n_2+k}(x_2,x_2)P^{m_2}(x_2,y_2) > 0.
\end{aligned}$$

Cela assure que pour tout $n \geq \max(n_1+m_1, n_2+m_2)$, on a

$$\begin{aligned}
\overline{P}^n\big((x_1,x_2),(y_1,y_2)\big) &= P^n(x_1,y_1)P^n(x_2,y_2)\\
&= P^{n_1+m_1+k_1}(x_1,y_1)P^{n_2+m_2+k_2}(x_2,y_2) > 0,
\end{aligned}$$

en écrivant $n = n_1+m_1+k_1 = n_2+m_2+k_2$ avec $k_1,k_2 \geq 0$ et la matrice stochastique $\overline{P}$ est bien irréductible. De plus, de cette façon, on montre que chaque $(x_1,x_2) \in E^2$ pour n assez grand $\overline{P}^n\big((x_1,x_2),(x_1,x_2)\big) > 0$ et $\overline{P}^{n+1}\big((x_1,x_2),(x_1,x_2)\big) > 0$ si bien que la période de la chaîne (Prop. 7.34) divise n et $n+1$ et vaut donc 1.

iii) La mesure produit $\pi\otimes\pi$ est invariante pour $\overline{P}$ puisque pour tout $(y_1,y_2) \in E^2$ on a :

$$\begin{aligned}
&\sum_{(x_1,x_2)\in E^2} (\pi\otimes\pi)(x_1,x_2)\overline{P}\big((x_1,x_2),(y_1,y_2)\big) = \sum_{x_1,x_2\in E} \pi(x_1)\pi(x_2)P(x_1,y_1)P(x_2,y_2)\\
&= \left(\sum_{x_1\in E} \pi(x_1)P(x_1,y_1)\right)\left(\sum_{x_2\in E} \pi(x_2)P(x_2,y_2)\right)\\
&= \pi(y_1)\pi(y_2) = (\pi\otimes\pi)(y_1,y_2),
\end{aligned}$$

soit l'équation de Chapman-Kolmogorov (7.1) pour $\overline{P}$: $(\pi\otimes\pi) = (\pi\otimes\pi)\overline{P}$, et $\pi\otimes\pi \neq 0$ charge finiment tous les points. De plus $\pi\otimes\pi$ est une probabilité puisque

$$\sum_{(x_1,x_2)\in E^2} (\pi\otimes\pi)(x_1,x_2) = \sum_{x_1,x_2\in E} \pi(x_1)\pi(x_2) = \Big(\sum_{x_1\in E}\pi(x_1)\Big)\Big(\sum_{x_2\in E}\pi(x_2)\Big) = 1\times 1 = 1.$$

iv) On montre que $\overline{\mathbb{P}}_\nu\big(X_n^{(1)}=x\big)=\mathbb{P}_{\nu_1}\big(X_n^{(1)}=x\big)$ (on procèderait de la même façon pour $i=2$). Pour cela, on a :

$$\begin{aligned}
\overline{\mathbb{P}}_\nu\big(X_n^{(1)}=x\big) &= \sum_{y\in E}\overline{\mathbb{P}}_\nu\big(X_n^{(1)}=x, X_n^{(2)}=y\big)\\
&= \sum_{y\in E}\sum_{(x_0,y_0)\in E^2}\nu(x_0,y_0)\overline{P}^n\big((x_0,y_0),(x,y)\big)\\
&= \sum_{y\in E}\sum_{(x_0,y_0)\in E^2}\nu(x_0,y_0)P^n(x_0,x)P^n(y_0,y)\\
&= \sum_{x_0\in E}P^n(x_0,x)\Big(\sum_{y_0\in E}\nu(x_0,y_0)\sum_{y\in E}P^n(y_0,y)\Big)\\
&= \sum_{x_0\in E}\nu_1(x_0)P^n(x_0,x)=\mathbb{P}_{\nu_1}\big(X_n^{(1)}=x\big)
\end{aligned}$$

car $\sum_{y\in E}P^n(y_0,y)=1$ et en notant $\nu_1(x_0)=\nu(\{x_0\}\times E)$. □

Démonstration :[Théorème 7.40] On commence par prouver (7.21) en utilisant le couplage du Lemme 7.42. On considère la chaîne $\overline{X}=\big(X_n^{(1)},X_n^{(2)}\big)_{n\ge 0}$ de matrice stochastique $\overline{P}$ en (7.23) qui est irréductible, apériodique, de probabilité invariante $\pi^{\otimes 2}$ (Lemme 7.42). D'après le Théorème 7.25, $(\overline{X}_n)_{n\ge 0}$ est donc récurrente positive (car irréductible et existence d'une probabilité invariante, cf. 3) dans le Th.7.25). On note $\Delta=\{(x,x)\in E^2 : x\in E\}$ la diagonale de E^2. Comme la chaîne $(\overline{X}_n)_{n\ge 0}$ est récurrente et irréductible, elle atteint presque sûrement en temps fini tout état de E^2 quelque soit la loi initiale. Par conséquent, $T_\Delta=\inf\big(n\ge 0 : (X_n^{(1)},X_n^{(2)})\in\Delta\big)$ est un temps d'arrêt et il est fini presque sûrement pour toute loi initiale de $(\overline{X}_n)_{n\ge 0}$, en particulier $\pi\otimes\delta_x$.

L'identité (7.25) du Lemme 7.42 assure $\overline{\mathbb{P}}_{\pi\otimes\delta_x}\big(X_n^{(2)}=y\big)=\mathbb{P}_x(X_n=y)$ et $\overline{\mathbb{P}}_{\pi\otimes\delta_x}\big(X_n^{(1)}=y\big)=\pi(y)$. On a donc

$$\begin{aligned}
\mathbb{P}_x(X_n=y)-\pi(y) &= \overline{\mathbb{P}}_{\pi\otimes\delta_x}\big(X_n^{(2)}=y\big)-\overline{\mathbb{P}}_{\pi\otimes\delta_x}\big(X_n^{(1)}=y\big)\\
&= \overline{\mathbb{E}}_{\pi\otimes\delta_x}\Big[\mathbf{1}_{\{X_n^{(2)}=y\}}-\mathbf{1}_{\{X_n^{(1)}=y\}}\Big].
\end{aligned}$$

En distinguant selon les valeurs de T_Δ, on a :

$$\begin{aligned}
\mathbb{P}_x(X_n=y)-\pi(y) &= \overline{\mathbb{E}}_{\pi\otimes\delta_x}\Big[\mathbf{1}_{\{X_n^{(2)}=y\}}-\mathbf{1}_{\{X_n^{(1)}=y\}}\Big]\\
&= \overline{\mathbb{E}}_{\pi\otimes\delta_x}\Big[\mathbf{1}_{\{T_\Delta>n\}}\Big(\mathbf{1}_{\{X_n^{(2)}=y\}}-\mathbf{1}_{\{X_n^{(1)}=y\}}\Big)\Big]\\
&\quad+\sum_{k=0}^n\overline{\mathbb{E}}_{\pi\otimes\delta_x}\Big[\mathbf{1}_{\{T_\Delta=k\}}\Big(\mathbf{1}_{\{X_n^{(2)}=y\}}-\mathbf{1}_{\{X_n^{(1)}=y\}}\Big)\Big]\\
&= \overline{\mathbb{E}}_{\pi\otimes\delta_x}\Big[\mathbf{1}_{\{T_\Delta>n\}}\Big(\mathbf{1}_{\{X_n^{(2)}=y\}}-\mathbf{1}_{\{X_n^{(1)}=y\}}\Big)\Big]
\end{aligned}$$

$$+\sum_{k=0}^{n}\sum_{z\in E}\overline{\mathbb{E}}_{\pi\otimes\delta_x}\Big[\mathbf{1}_{\{T_\Delta=k\}}\mathbf{1}_{\{X_k^{(1)}=X_k^{(2)}=z\}}\Big(\mathbf{1}_{\{X_n^{(2)}=y\}}-\mathbf{1}_{\{X_n^{(1)}=y\}}\Big)\Big]. \tag{7.26}$$

En utilisant la propriété de Markov (Théorème 5.36) au temps k avec loi initiale $\pi\otimes\delta_x$ (cf. Remarque 5.39), on a

$$\begin{aligned}
&\overline{\mathbb{E}}_{\pi\otimes\delta_x}\Big[\mathbf{1}_{\{T_\Delta=k\}}\mathbf{1}_{\{X_k^{(1)}=X_k^{(2)}=z\}}\mathbf{1}_{\{X_n^{(2)}=y\}}\Big]\\
&= \overline{\mathbb{E}}_{\pi\otimes\delta_x}\Big[\overline{\mathbb{E}}_{\pi\otimes\delta_x}\Big[\underbrace{\mathbf{1}_{\{T_\Delta=k\}}\mathbf{1}_{\{X_k^{(1)}=X_k^{(2)}=z\}}}_{\mathcal{F}_k\text{-mesurable}}\mathbf{1}_{\{X_n^{(2)}=y\}}\big|\mathcal{F}_k\Big]\Big]\\
&= \overline{\mathbb{E}}_{\pi\otimes\delta_x}\Big[\mathbf{1}_{\{T_\Delta=k\}}\mathbf{1}_{\{X_k^{(1)}=X_k^{(2)}=z\}}\overline{\mathbb{E}}_{(X_k^{(1)},X_k^{(2)})}\big[\mathbf{1}_{\{X_{n-k}^{(2)}=y\}}\big]\Big] \quad \text{(par Markov faible, Théorème 5.36)}\\
&= \overline{\mathbb{E}}_{\pi\otimes\delta_x}\Big[\mathbf{1}_{\{T_\Delta=k\}}\mathbf{1}_{\{X_k^{(1)}=X_k^{(2)}=z\}}\overline{\mathbb{E}}_{(z,z)}\big[\mathbf{1}_{\{X_{n-k}^{(2)}=y\}}\big]\Big]\\
&= \overline{\mathbb{E}}_{\pi\otimes\delta_x}\Big[\mathbf{1}_{\{T_\Delta=k\}}\mathbf{1}_{\{X_k^{(1)}=X_k^{(2)}=z\}}\Big]\overline{\mathbb{E}}_{(z,z)}\Big[\mathbf{1}_{\{X_{n-k}^{(2)}=y\}}\Big]\\
&= \overline{\mathbb{E}}_{\pi\otimes\delta_x}\Big[\mathbf{1}_{\{T_\Delta=k\}}\mathbf{1}_{\{X_k^{(1)}=X_k^{(2)}=z\}}\Big]P_{n-k}(z,y)\\
&= \overline{\mathbb{E}}_{\pi\otimes\delta_x}\Big[\mathbf{1}_{\{T_\Delta=k\}}\mathbf{1}_{\{X_k^{(1)}=X_k^{(2)}=z\}}\mathbf{1}_{\{X_n^{(1)}=y\}}\Big] \quad \text{(par symétrie entre } X^{(1)} \text{ et } X^{(2)})
\end{aligned}$$

ce qui assure pour tout $z\in E$

$$\overline{\mathbb{E}}_{\pi\otimes\delta_x}\Big[\mathbf{1}_{\{T_\Delta=k\}}\mathbf{1}_{\{X_k^{(1)}=X_k^{(2)}=z\}}\Big(\mathbf{1}_{\{X_n^{(2)}=y\}}-\mathbf{1}_{\{X_n^{(1)}=y\}}\Big)\Big]=0$$

et donc (7.26) se réduit à

$$\mathbb{P}_x(X_n=y)-\pi(y)=\overline{\mathbb{E}}_{\pi\otimes\delta_x}\Big[\mathbf{1}_{\{T_\Delta>n\}}\Big(\mathbf{1}_{\{X_n^{(2)}=y\}}-\mathbf{1}_{\{X_n^{(1)}=y\}}\Big)\Big].$$

Il vient alors

$$\begin{aligned}
\sum_{y\in E}\big|\mathbb{P}_x(X_n=y)-\pi(y)\big| &= \sum_{y\in E}\Big|\overline{\mathbb{E}}_{\pi\otimes\delta_x}\Big[\mathbf{1}_{\{T_\Delta>n\}}\Big(\mathbf{1}_{\{X_n^{(2)}=y\}}-\mathbf{1}_{\{X_n^{(1)}=y\}}\Big)\Big]\Big|\\
&\leq \sum_{y\in E}\overline{\mathbb{E}}_{\pi\otimes\delta_x}\Big[\mathbf{1}_{\{T_\Delta>n\}}\big|\mathbf{1}_{\{X_n^{(2)}=y\}}-\mathbf{1}_{\{X_n^{(1)}=y\}}\big|\Big]\\
&\leq \sum_{y\in E}\overline{\mathbb{E}}_{\pi\otimes\delta_x}\Big[\mathbf{1}_{\{T_\Delta>n\}}\Big(\mathbf{1}_{\{X_n^{(2)}=y\}}+\mathbf{1}_{\{X_n^{(1)}=y\}}\Big)\Big]\\
&= \overline{\mathbb{E}}_{\pi\otimes\delta_x}\Big[\mathbf{1}_{\{T_\Delta>n\}}\sum_{y\in E}\Big(\mathbf{1}_{\{X_n^{(2)}=y\}}+\mathbf{1}_{\{X_n^{(1)}=y\}}\Big)\Big]\\
&= 2\,\overline{\mathbb{E}}_{\pi\otimes\delta_x}\big[\mathbf{1}_{\{T_\Delta>n\}}\big]=2\,\overline{\mathbb{P}}_{\pi\otimes\delta_x}(T_\Delta>n)
\end{aligned} \tag{7.27}$$

puisque $\sum_{y\in E}\mathbf{1}_{\{X_n^{(1)}=y\}}=\sum_{y\in E}\mathbf{1}_{\{X_n^{(2)}=y\}}=1$.

Comme T_Δ est fini $\overline{\mathbb{P}}_{\pi\otimes\delta_x}$-ps, par convergence monotone on a

$$\lim_{n\to+\infty}\overline{\mathbb{P}}_{\pi\otimes\delta_x}(T_\Delta>n)=0,$$

et donc par (7.27)
$$\lim_{n\to+\infty}\sum_{y\in E}\big|\mathbb{P}_x(X_n=y)-\pi(y)\big|=0.$$

On termine avec la preuve de (7.22) : on rappelle la définition (5.28) de $\mathbb{P}_\nu$ ($\mathbb{P}_\nu = \sum_{x\in E}\nu(x)\mathbb{P}_x$) et on a
$$\begin{aligned}\big|\mathbb{P}_\nu(X_n=y)-\pi(y)\big| &= \Big|\sum_{x\in E}\nu(x)\mathbb{P}_\nu(X_n=y)-\Big(\sum_{x\in E}\nu(x)\Big)\pi(y)\Big|\\ &= \left|\sum_{x\in E}\big(\mathbb{P}_x(X_n=y)-\pi(y)\big)\,\nu(x)\right|\end{aligned}$$

et donc
$$\begin{aligned}\sum_{y\in E}|\mathbb{P}_\nu(X_n=y)-\pi(y)| &\leq \sum_{y\in E}\sum_{x\in E}\big|\mathbb{P}_x(X_n=y)-\pi(y)\big|\,\nu(x)\\ &= \sum_{x\in E}\Big(\sum_{y\in E}\big|\mathbb{P}_x(X_n=y)-\pi(y)\big|\Big)\,\nu(x).\end{aligned}\tag{7.28}$$

En appliquant le Lemme 7.16 avec $a=\nu$ de poids fini et $b_n(x)=\sum_{y\in E}|\mathbb{P}_x(X_n=y)-\pi(y)|$ vérifiant
$$0\leq b_n(x)\leq\sum_{y\in E}\mathbb{P}_x(X_n=y)+\sum_{y\in E}\pi(y)=2,$$
avec $\lim_{n\to+\infty}b_n(x)=0$ par (7.21), on obtient la conclusion (7.22) en passant à la limite dans (7.28). □

Remarque 7.43 1) L'apériodicité est essentielle, sans elle, le couplage échoue en général. Par exemple pour la chaîne à deux états sur $E=\{0,1\}$ (Exemple 5.1)
$$P=\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad P^{2n}=\begin{pmatrix}1&0\\0&1\end{pmatrix},\quad P^{2n+1}=\begin{pmatrix}0&1\\1&0\end{pmatrix},\quad n\geq 0,$$
et si $X_0^{(1)}=0$, $X_0^{(2)}=1$, on aura $X_n^{(1)}\neq X_n^{(2)}$ pour tout $n\geq 0$ (ie. pas de couplage possible).

Cas périodique

Dans le cas d'une chaîne de Markov périodique, on généralise le Théorème 7.40 comme suit :

Théorème 7.44 *Soit $(X_n)_{n\geq 0}$ une chaîne irréductible récurrente positive, périodique de période d et de probabilité invariante π. Alors pour toute paire d'états $x,y\in E$, il existe $r\in[\![0,d-1]\!]$ tel que $P^n(x,y)=0$ si $n\neq r$ mod d, sinon $n=md+r$ et on a*
$$\lim_{m\to+\infty}P^{md+r}(x,y)=d\pi(y).$$

Exemple 7.45 (Chaîne de de naissance et de mort) On considère une chaîne de naissance et de mort irréductible récurrente positive de période 2 . Si $y-x$ est pair alors $P^{2m+1}(x,y)=0$ pour tout $m\geq 0$ et

$$\lim_{m\to+\infty} P^{2m}(x,y)=2\pi(y).$$

Si $y-x$ est impair alors $P^{2m}(x,y)=0$ pour tout $m\geq 0$ et

$$\lim_{m\to+\infty} P^{2m+1}(x,y)=2\pi(y).$$

Démonstration : On commence par une extension du Théorème 7.40 pour une classe close irréductible récurrente positive apériodique. Soit E_i une telle classe et soit $\pi^{(i)}$ la probabilité invariante concentrée sur E_{R_i} (Théorème 7.30). En considérant la chaîne restreinte à E_{R_i}, on a d'après le Théorème 7.40 :

$$\lim_{n\to+\infty} P^n(x,y)=\pi^{(i)}(y)=\frac{1}{\mathbb{E}_y\big[\widetilde{T}_y\big]},\quad x,y\in E_{R_i}.$$

En particulier si y est un état récurrent positif de période 1 alors en choisissant pour E_{R_i} la classe close irréductible contenant y, on voit que :

$$\lim_{n\to+\infty} P^n(y,y)=\frac{1}{\mathbb{E}_y\big[\widetilde{T}_y\big]}. \tag{7.29}$$

On prouve maintenant le Théorème 7.44 pour le cas périodique. Soit donc $(X_n)_{n\geq 0}$ une chaîne irréductible récurrente positive et périodique de période $d>1$. En posant $(Y_m)_{m\geq 0}=(X_{md})_{m\geq 0}$, on définit une chaîne de Markov de matrice stochastique $Q=P^d$. Soit $y\in E$, alors

$$\begin{aligned}\mathrm{PGCD}\big(m\ :\ Q^m(y,y)>0\big) &= \mathrm{PGCD}\big(m\ :\ P^{md}(y,y)>0\big)\\ &= \frac{1}{d}\,\mathrm{PGCD}\big(n\ :\ P^n(y,y)>0\big)=1.\end{aligned}$$

Les états sont donc de période 1 pour la chaîne $(Y_m)_{m\geq 0}$, qui est donc apériodique.

On suppose que la chaîne $(X_n)_{n\geq 0}$, et donc aussi $(Y_m)_{m\geq 0}$, démarre de y. Comme la chaîne $(X_n)_{n\geq 0}$ revisite y la première fois à un multiple de d, la durée d'un retour moyen en y pour la chaîne $(Y_m)_{m\geq 0}$ est $d^{-1}\mathbb{E}_y\big[\widetilde{T}_y\big]$ où $\mathbb{E}_y\big[\widetilde{T}_y\big]$ est la durée d'un retour moyen en y de la chaîne $(X_n)_{n\geq 0}$. En particulier, y est récurrent positif pour toute chaîne de Markov de matrice stochastique Q. En appliquant le résultat préliminaire (7.29) à cette matrice stochastique, on a

$$\lim_{m\to+\infty} Q^m(y,y)=\frac{d}{m_y}=d\pi(y),$$

soit

$$\lim_{m\to+\infty} P^{md}(y,y)=d\pi(y),\quad y\in E. \tag{7.30}$$

Soit $x, y \in E$, par irréductibilité de P, il existe $n \geq 1$ tel que $P^n(x,y) > 0$. On pose alors $r_1 = \min(n \geq 0 \, : \, P^n(x,y) > 0)$. On a en particulier $P^{r_1}(x,y) > 0$.

On montre que $P^n(x,y) > 0$ seulement si $n - r_1$ est multiple de d : par irréductibilité, on choisit $n_1 > 0$ tel que $P^{n_1}(y,x) > 0$, alors

$$P^{r_1+n_1}(y,y) \geq P^{n_1}(y,x)P^{r_1}(x,y) > 0$$

et donc $r_1 + n_1$ est multiple de d. Réciproquement, si $P^n(x,y) > 0$ alors de la même façon,

$$P^{n+n_1}(x,x) \geq P^n(x,y)P^{n_1}(y,x) > 0$$

et $n + n_1$ doit être un multiple de d ; par conséquent $n - r_1 = (n + n_1) - (r_1 + n_1)$ aussi. Finalement, $n - r_1$ doit être multiple de d et $n = kd + r_1$ pour un certain $k \in \mathbb{N}$.

Il existe $m_1 \in \mathbb{N}$ tel que $r_1 = m_1 d + r$ avec $r \in [\![0, d-1]\!]$. D'après ce qui précède, on a $P^n(x,y) = 0$ si et seulement si $n \neq r \mod d$. On déduit maintenant que

$$P^{md+r}(x,y) = \sum_{k=0}^{m} \mathbb{P}_x\big(\widetilde{T}_y = kd + r\big) P^{(m-k)d}(y,y). \tag{7.31}$$

On pose

$$b_m(k) = \begin{cases} P^{(m-k)d}(y,y) & \text{si } 0 \leq k \leq m \\ 0 & \text{si } k > m \end{cases}$$

Alors par (7.30), pour chaque k fixé, on a $\lim_{m\to+\infty} b_m(k) = d\pi(y)$. En appliquant le Lemme 7.16 (convergence dominée) avec $E = \mathbb{N}$, $b(k) = d\pi(y)$ et $a(k) = \mathbb{P}_x\big(\widetilde{T}_y = kd+r\big)$ (sommable) pour passer à la limite en $m \to +\infty$ dans (7.31), on a

$$\begin{aligned} \lim_{m\to+\infty} P^{md+r}(x,y) &= d\pi(y) \sum_{k=0}^{+\infty} \mathbb{P}_x\big(\widetilde{T}_y = kd + r\big) \\ &= d\pi(y)\mathbb{P}_x\big(T_y < +\infty\big) \\ &= d\pi(y), \end{aligned}$$

ce qui achève la preuve du Théorème 7.44. Noter que $\sum_{k=0}^{+\infty} \mathbb{P}_x\big(\widetilde{T}_y = kd + r\big) = \mathbb{P}_x\big(T_y < +\infty\big)$ vient de $P^n(x,y) = 0$ si $n \neq r \mod d$.

□

7.5 Théorème ergodique

Le théorème ergodique relie moyenne temporelle et moyenne spatiale.

Théorème 7.46 (Ergodique) *Soit $(X_n)_{n\geq 0}$ une chaîne de Markov récurrente irréductible et soit π une mesure invariante. Soit $f, g \in L^1(\pi)$ avec $\int_E g\, d\pi \neq 0$. Alors pour toute loi initiale ν sur E, on a*

$$\lim_{n\to+\infty} \frac{\sum_{k=0}^n f(X_k)}{\sum_{k=0}^n g(X_k)} = \frac{\int_E f\, d\pi}{\int_E g\, d\pi} \quad \mathbb{P}_\nu\text{-ps}. \tag{7.32}$$

Remarque 7.47 En fait le résultat reste vrai si f est positive avec $\int_E f\,d\pi = +\infty$, il suffit de prendre des fonctions $f_k \nearrow f$ avec $\int_E f_k\,d\pi < +\infty$ et d'utiliser le théorème de convergence monotone : Comme $f \geq f_k$, on a :

$$\liminf_{n\to+\infty} \frac{\sum_{i=0}^n f(X_i)}{\sum_{i=0}^n g(X_i)} \geq \liminf_{n\to+\infty} \frac{\sum_{i=0}^n f_k(X_i)}{\sum_{i=0}^n g(X_i)} = \frac{\int_E f_k\,d\pi}{\int_E g\,d\pi} \quad \mathbb{P}_\nu\text{-ps.}$$

Puis comme $\int_E f_k\,d\pi \nearrow +\infty$ quand $k \to +\infty$, il vient

$$\lim_{n\to+\infty} \frac{\sum_{i=0}^n f(X_i)}{\sum_{i=0}^n g(X_i)} = +\infty \quad \mathbb{P}_\nu\text{-ps.}$$

Corollaire 7.48 (Ergodique) *Soit $(X_n)_{n\geq 0}$ une chaîne de Markov irréductible récurrente positive et π son unique probabilité invariante. Alors pour toute loi initiale ν sur E et $f \in L^1(\pi)$, on a :*

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^n f(X_k) = \int_E f\,d\pi \quad \mathbb{P}_\nu\text{-ps.} \tag{7.33}$$

Démonstration : Comme la chaîne $(X_n)_{n\geq 0}$ est irréductible récurrente positive, il existe une unique probabilité invariante π par le Théorème 7.25 et on a $\mathbf{1} \in L^1(\pi)$. Le Théorème 7.46 s'applique alors avec $g(x) = 1$ et assure (7.33). □

Remarque 7.49 Cette limite (7.33) est l'essence même de la notion d'ergodicité : la moyenne de f le long de la trajectoire de la chaîne, ie. sa moyenne temporelle $\frac{1}{n}\sum_{k=0}^n f(X_k)$, converge en temps long ($n \to +\infty$) vers la moyenne spatiale $\int_E f\,d\pi$ de f (par rapport à la probabilité invariante).

Puis en appliquant le Théorème 7.46 avec $f(y) = \mathbf{1}_{\{y=x\}}$ et $g(y) = 1$ (avec la Remarque 7.47 lorsque $g \notin L^1(\pi)$), on récupère immédiatement le Théorème 6.15 :

Corollaire 7.50 (Ergodique) *Soit $(X_n)_{n\geq 0}$ une chaîne de Markov récurrente irréductible. Alors pour toute loi initiale ν sur E :*

(1) Dans le cas récurrent positif :

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^n \mathbf{1}_{\{X_k=x\}} = \pi(x) \quad \mathbb{P}_\nu\text{-ps ;}$$

(2) Dans le cas récurrent nul :

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^n \mathbf{1}_{\{X_k=x\}} = 0 \quad \mathbb{P}_\nu\text{-ps.}$$

Démonstration :[Théorème 7.46] D'abord on observe qu'il suffit de voir la convergence (7.32) pour $\mathbb{P}_x$ pour tout $x \in E$ pour la récupérer pour toute loi ν par (5.28) : si $\mathbb{P}_x(A) = 1$ pour tout $x \in E$ alors $\mathbb{P}_\nu(A) = \sum_{x\in E} \nu(x)\mathbb{P}_x(A) = \sum_{x\in E} \nu(x) = 1$.

Pour la suite, on fixe $x \in E$. On commence par observer que $\pi(x) > 0$: en effet, par le Théorème 7.24, les mesures invariantes $\pi = \pi(x)\nu_x$ où ν_x est la mesure invariante associée à l'état x récurrent dans (7.8). Comme $\pi \neq 0$, cela exige $\pi(x) > 0$.

On utilise les dates de retour de la chaîne $(X_n)_{n\geq 0}$ en x définies par récurrence en (6.1) par $T_x^{(n)} = \inf\big(k > T_x^{(n-1)} : X_k = x\big)$ et satisfaisant (6.2)

$$T_x^{(n+1)} = T_x^{(n)} + T_x^{(1)} \circ \Theta_{T_x^{(n)}}. \tag{7.34}$$

Puisque x est récurrent, les temps $(T_x^{(n)})_{n\geq 0}$ sont $\mathbb{P}_x$-ps finis (Déf. 6.3 + Markov fort (5.37)). Pour $k \geq 1$, on pose alors

$$Z_k(f) = \sum_{i=T_x^{(k-1)}}^{T_x^{(k)}-1} f(X_i). \tag{7.35}$$

Lemme 7.51 *La variable aléatoire $Z_k(f)$ est $\mathcal{F}_{T_x^{(k)}}$-mesurable.*

Démonstration :[Lemme 7.51] On montre que pour tout $A \in \mathcal{B}(\mathbb{R})$, on a $\{Z_k(f) \in A\} \in \mathcal{F}_{T_x^{(k)}}$. Pour cela, il faut et il suffit de voir que pour tout $n \geq 1$ $\{Z_k(f) \in A\} \cap \{T_x^{(k)} = n\} \in \mathcal{F}_n$ (cf. 2) dans la Prop. 3.15).

$$\begin{aligned}\{Z_k(f) \in A\} \cap \{T_x^{(k)} = n\} &= \bigcup_{m=1}^{n-1} \Big(\{Z_k(f) \in A\} \cap \{T_x^{(k)} = n\} \cap \{T_x^{(k-1)} = m\}\Big)\\ &= \bigcup_{m=1}^{n-1} \left(\Big\{\sum_{j=m}^{n-1} f(X_j) \in A\Big\} \cap \{T_x^{(k)} = n\} \cap \{T_x^{(k-1)} = m\}\right) \in \mathcal{F}_n\end{aligned}$$

car $\Big\{\sum_{j=m}^{n-1} f(X_j) \in A\Big\} \in \mathcal{F}_{n-1} \subset \mathcal{F}_n$, $\{T_x^{(k-1)} = m\} \in \mathcal{F}_m \subset \mathcal{F}_n$, $\{T_x^{(k)} = n\} \in \mathcal{F}_n$. □

Lemme 7.52 *Les variables aléatoires $\big(Z_k(f)\big)_{k\geq 1}$ sont iid. En particulier avec $f = 1$, on obtient que les variables aléatoires $T_x^{(n)} - T_x^{(n-1)}$, $n \geq 1$, sont iid et on retrouve la Proposition 6.14.*

Démonstration :[Lemme 7.52] Pour tout $k \geq 1$, il suffit de voir pour des fonctions g_i mesurables bornées sur $\mathbb{R}_+$ $(1 \leq i \leq k)$:

$$\mathbb{E}_x\left[\prod_{i=1}^k g_i\big(Z_i(f)\big)\right] = \prod_{i=1}^k \mathbb{E}_x\big[g_i\big(Z_1(f)\big)\big]. \tag{7.36}$$

Pour cela, on raisonne par récurrence sur l'entier k.

Pour $k = 1$, l'égalité (7.36) est immédiate. On suppose alors (7.36) établie pour k fixé et on la prouve pour $k + 1$. Pour cela, on observe que

— les variables aléatoires $Z_1(f), Z_2(f), \dots, Z_k(f)$ sont $\mathcal{F}_{T_x^{(k)}}$-mesurables (Lemme 7.51),
— $\Theta_{T_x^{(k)}}(\omega)$ est indépendante de $\mathcal{F}_{T_x^{(k)}}$ et de loi $\mathbb{P}_x$ (propriété de Markov forte sous la forme du Corollaire 5.38),
— $Z_{k+1}(f) = Z_1(f) \circ \Theta_{T_x^{(k)}}$. En effet, avec (7.34) on a :

$$\begin{aligned} Z_{k+1}(f) &= \sum_{i=T_x^{(k)}}^{T_x^{(k+1)}-1} f(X_i) = \sum_{j=0}^{T_x^{(k+1)}-T_x^{(k)}-1} f\big(X_{j+T_x^{(k)}}\big) \\ &= \sum_{j=0}^{T_x^{(1)}\circ\Theta_{T_x^{(k)}}-1} f\big(X_j \circ \Theta_{T_x^{(k)}}\big) = \Big(\sum_{j=0}^{T_x^{(1)}-1} f(X_j)\Big) \circ \Theta_{T_x^{(k)}} \\ &= Z_1(f) \circ \Theta_{T_x^{(k)}}. \end{aligned}$$

En utilisant la propriété de Markov fort (Théorème 5.37), on a donc

$$\begin{aligned} \mathbb{E}_x\left[\prod_{i=1}^{k+1} g_i(Z_i(f))\right] &= \mathbb{E}_x\left[\Big(\prod_{i=1}^{k} g_i(Z_i(f))\Big)\, g_{k+1}\big(Z_1(f)\circ\Theta_{T_x^{(k)}}\big)\right] \\ &= \mathbb{E}_x\left[\Big(\prod_{i=1}^{k} g_i(Z_i(f))\Big)\, \mathbb{E}_x\Big[g_{k+1}\big(Z_1(f)\circ\Theta_{T_x^{(k)}}\big)\big|\mathcal{F}_{T_x^{(k)}}\Big]\right] \\ &= \mathbb{E}_x\left[\Big(\prod_{i=1}^{k} g_i(Z_i(f))\Big)\, \mathbb{E}_{X_{T_x^{(k)}}}\Big[g_{k+1}\big(Z_1(f)\big)\Big]\right] \\ &= \mathbb{E}_x\left[\prod_{i=1}^{k} g_i(Z_i(f))\right] \mathbb{E}_x\big[g_{k+1}(Z_1(f))\big] \\ &= \left(\prod_{i=1}^{k} \mathbb{E}_x\big[g_i(Z_i(f))\big]\right) \mathbb{E}_x\big[g_{k+1}(Z_1(f))\big] \quad \text{(par hyp. de récurrence),} \end{aligned}$$

ce qui prouve (7.36) pour $k+1$ et donc le Lemme 7.52 par récurrence. □

Suite de la preuve du théorème ergodique (Théorème 7.46). Afin d'appliquer la loi des grands nombres (LGN) aux variables aléatoires $(Z_k(f))_{k\geq 1}$ *iid* (Lemme 7.52), on montre qu'elles sont L^1 lorsque $f \in L^1(\pi)$: en effet

$$\mathbb{E}_x\big[|Z_1(f)|\big] \leq \mathbb{E}_x\left[\sum_{k=0}^{T_x^{(1)}-1} |f(X_k)|\right] = \mathbb{E}_x\left[\sum_{k=0}^{T_x^{(1)}-1} \sum_{y\in E} |f(y)| \mathbf{1}_{\{X_k=y\}}\right] \tag{7.37}$$

$$= \sum_{y\in E} |f(y)|\, \mathbb{E}_x\left[\sum_{k=0}^{T_x^{(1)}-1} \mathbf{1}_{\{X_k=y\}}\right] = \sum_{y\in E} |f(y)|\nu_x(y) = \frac{\int_E |f|\, d\pi}{\pi(x)} \tag{7.38}$$

puisque $\pi = \pi(x)\nu_x$. Le même calcul avec f à la place de $|f|$, et égalité dans ce ce cas dans (7.37), donne

$$\mathbb{E}_x\big[Z_1(f)\big] = \mathbb{E}_x\left[\sum_{k=0}^{T_x^{(1)}-1} f(X_k)\right] = \frac{\int_E f\,d\pi}{\pi(x)} < +\infty.$$

La LGN assure alors

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^n Z_k(f) = \frac{\int_E f\,d\pi}{\pi(x)} \quad \mathbb{P}_x\text{-ps.} \tag{7.39}$$

Maintenant, on note $\widetilde{N}_x(n) = \sum_{k=1}^n \mathbf{1}_{\{X_k=x\}}$ de sorte que

$$T_x^{(\widetilde{N}_x(n))} \leq n < T_x^{(\widetilde{N}_x(n)+1)}.$$

Lorsque f est une fonction **positive**, on a

$$\frac{\sum_{k=0}^{T_x^{(\widetilde{N}_x(n))}-1} f(X_k)}{\widetilde{N}_x(n)} \leq \frac{\sum_{k=0}^n f(X_k)}{\widetilde{N}_x(n)} \leq \frac{\sum_{k=0}^{T_x^{(\widetilde{N}_x(n)+1)}-1} f(X_k)}{\widetilde{N}_x(n)}.$$

En regroupant les paquets $Z_j(f)$ définis en (7.35), on a :

$$\sum_{k=0}^{T_x^{(\widetilde{N}_x(n))}-1} f(X_k) = \sum_{j=1}^{\widetilde{N}_x(n)} \Big(\sum_{i=T_x^{(j-1)}}^{T_x^{(j)}-1} f(X_i)\Big) = \sum_{j=1}^{\widetilde{N}_x(n)} Z_j(f),$$

et donc

$$\frac{\sum_{j=1}^{\widetilde{N}_x(n)} Z_j(f)}{\widetilde{N}_x(n)} \leq \frac{\sum_{k=0}^n f(X_k)}{\widetilde{N}_x(n)} \leq \frac{\sum_{j=1}^{\widetilde{N}_x(n)+1} Z_j(f)}{\widetilde{N}_x(n)}. \tag{7.40}$$

Comme x est récurrent, $\widetilde{N}_x(n) \to +\infty$ et (7.39) assurent que les termes de gauche et de droite de (7.40) convergent $\mathbb{P}_x$-ps vers $\int_E f\,d\pi/\pi(x)$ et donc par le théorème des gendarmes

$$\lim_{n\to+\infty} \frac{1}{\widetilde{N}_x(n)}\sum_{k=0}^n f(X_k) = \frac{\int_E f\,d\pi}{\pi(x)} \quad \mathbb{P}_x\text{-ps.} \tag{7.41}$$

Si $f \in L^1(\pi)$ est de signe quelconque, on applique (7.41) à $f^+ = \max(f,0)$ et à $f^- = \max(-f,0)$ et par différence on obtient (7.41) pour $f = f^+ - f^-$ (la différence dans (7.41) a bien un sens car $f \in L^1(\pi)$) :

$$\begin{aligned}
\frac{1}{\widetilde{N}_x(n)}\sum_{k=0}^n f(X_k) &= \frac{1}{\widetilde{N}_x(n)}\sum_{k=0}^n \big(f^+(X_k) - f^-(X_k)\big)\\
&= \frac{1}{\widetilde{N}_x(n)}\sum_{k=0}^n f^+(X_k) - \frac{1}{\widetilde{N}_x(n)}\sum_{k=0}^n f^-(X_k)
\end{aligned}$$

$$\xrightarrow[n\to+\infty]{} \frac{\int_E f^+\,d\pi}{\pi(x)} - \frac{\int_E f^-\,d\pi}{\pi(x)} = \frac{\int_E (f^+-f^-)\,d\pi}{\pi(x)} = \frac{\int_E f\,d\pi}{\pi(x)} \quad \mathbb{P}_x\text{-ps.}$$

De la même façon, on obtient pour la fonction g :

$$\lim_{n\to+\infty} \frac{1}{\widetilde{N}_x(n)} \sum_{k=0}^n g(X_k) = \frac{\int_E g\,d\pi}{\pi(x)} \neq 0 \quad \mathbb{P}_x\text{-ps.}$$

Le rapport des deux limites donne alors

$$\lim_{n\to+\infty} \frac{\sum_{k=0}^n f(X_k)}{\sum_{k=0}^n g(X_k)} = \frac{\int_E f\,d\pi}{\int_E g\,d\pi} \quad \mathbb{P}_x\text{-ps,}$$

prouvant le Théorème 7.46. □

Exemple 7.53 (MCMC) La méthode de Monte Carlo par chaîne de Markov (Monte Carlo Markov Chains) vise à estimer une somme $S := \sum_{x\in E} \nu(x)f(x)$ où ν est une probabilité, et $f \in L^1(\nu)$, a priori difficile à calculer, en trouvant une chaîne de Markov $(X_n)_{n\geq 0}$ (irréductible, récurrente positive) admettant ν comme probabilité invariante. On a alors

$$\frac{1}{n}\sum_{k=0}^n f(X_k) \xrightarrow[n\to+\infty]{ps} \mathbb{E}_\nu[f] = S$$

et pour n assez grand $\frac{1}{n}\sum_{k=0}^n f(X_k)$ est une bonne estimation de la somme S. Cf. [Rob].

Bibliographie

[BL] Philippe Barbe, Michel Ledoux. *Probabilité*. EDP sciences, 2007.

[Bei] Frank Beichelt. *Stochastic processes in Science, Engineering and Finance*. Chapman & Hall, 2006.

[BEL] Michel Benaïm, Nicole El Karoui. *Promenade Aléatoire*, Ed. École Polytechnique, 2007.

[BC] Bernard Bercu, Djalil Chafaï. *Modélisation stochastique et simulation*. Dunod Ed., 2007.

[Bre-Leb] Jean-Christophe Breton. *Intégrale de Lebesgue*. Notes de cours de L3 Mathématiques, Université de Rennes 1, 2014. http://perso.univ-rennes1.fr/jean-christophe.breton/Fichiers/Integrale_Lebesgue.pdf

[Bre-proba] Jean-Christophe Breton. *Probabilités*. Notes de cours de L3 Mathématiques, Université de Rennes 1, 2014. http://perso.univ-rennes1.fr/jean-christophe.breton/Fichiers/proba_base.pdf

[BP] Marc Briane, Gilles Pagès. *Théorie de l'intégration*, 5ème édition. Coll. Vuibert Supérieur, Ed. Vuibert, 2012.

[Dud] Richard M. Dudley. *Real analysis and Probability*. Cambridge studies in advanced mathematics, vol 74, 2002.

[FF] Dominique Foata, Aimé Fuch. *Processus stochastiques*. Dunod, 2004.

[Gra] Carl Graham. *Chaînes de Markov*. Dunod Ed., 2008.

[HPS] Paul G. Hoel, Sidney C. Port et Charles J. Stone. *Introduction to stochastic processes*. Waveland Ed., 1972.

[JP] Jean Jacod, Philipp Protter. *L'essentiel en théorie des probabilités*. Vuibert, 2003.

[Jir] Miloslav Jiřina. *Conditional probabilities on strictly separable σ-algebras*. (Russian, with English summary) Czechoslovak Math. J. no. 4, vol. 79 pp. 372–380, 1954.

[Kal] Olav Kallenberg. *Foundations of modern probability*. 2nd Edition, Springer Series in Statistics. Probability and its Applications, 2002.

[Nev] Jacques Neveu. *Martingales à temps discret*, Masson, 1972.

[Nor] James Norris. *Markov Chains*. Cambridge studies in advanced mathematics, 1997.

[Ouv] Jean-Yves Ouvrard. *Probabilités*. Tomes 1 et 2. Cassini, 2008.

[Pri] Nicolas Privault. *Understanding Markov Chains – Examples and Applications*. Second Edition, Springer Undergraduate Mathematics Series, Springer, 2018.

[Rob] Christian Robert. *Méthodes de Monte Carlo par chaîne de Markov*, Ed. Economica., 1996

[Rue] Alain Ruegg. *Processus Stochastique*. Presse Universitaire Romande, 1989.

[Wil] David Williams. *Probability with martingales*. Cambridge mathematical textbooks, 1991.

[Yca] Bernard Ycart. *Modèles et Algorithmes Markoviens*, Ed. Springer, 2002.