Université de Rennes 1

CMMA

# CHAÎNES DE MARKOV
# &
# MARTINGALES

Auteur
Jean-Christrophe Breton

Notes de cours
Victor Lecerf

2021–2022

# Table des matières

# Chapitre 1

# Conditionnement

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

## 1.1 Probabilité sachant un évènement

**Définition 1.1.1.** *Soit $B \in \mathcal{F}$ tel que $\mathbb{P}(B) \neq 0$. On définit la* ***probablité de $A$ sachant $B$*** *comme la quantité*

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

**Remarques.** Dans la théorie des probabilités, il est préférable de s'intéresser à des probabilités conditionnelles que d'essayer de déterminer $\mathbb{P}(A)$ ou $\mathbb{P}(A|B)$. Remarquons de plus que si $A$ et $B \in \mathcal{F}$ sont deux évènements indépendants, alors $\mathbb{P}(A|B) = \mathbb{P}(A)$.

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

Bien évidemment, une probabilité conditionnelle porte bien son nom. Cela justifie la proposition suivante.

**Proposition 1.1.1.** *Soit $B$ un évènement de probabilité non nulle. Alors, l'application $\mathbb{P}(\,\cdot\,|B) : \mathcal{F} \longrightarrow [0,1]$ est une probabilité sur $\mathcal{F}$.*

*Démonstration.* Il est clair que la fonction $\mathbb{P}(\,\cdot\,|B)$ est positive. La $\sigma$-additivité de $\mathbb{P}$ se transmet. Soit $(A_n)_{n\in\mathbb{N}}$ une famille d'évènements deux à deux disjoints. On a

$$\mathbb{P}\left(\bigsqcup_{n\in\mathbb{N}} A_n \middle| B\right) = \frac{\mathbb{P}\left(\left(\bigsqcup_{n\in\mathbb{N}} A_n\right) \cap B\right)}{\mathbb{P}(B)} = \mathbb{P}\left(\bigsqcup_{n\in\mathbb{N}} (A_n \cap B)\right) = \sum_{n=0}^{\infty} \frac{\mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_{n=0}^{\infty} \mathbb{P}(A_n|B).$$

Enfin, $\mathbb{P}(\Omega|B) = \frac{\mathbb{P}(\Omega\cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$.

$\square$

**Proposition 1.1.2.** *Soient $A_1, \dots, A_n$ $n$ évènements tels que $\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_{n-1}) \neq 0$ ? Alors,*

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}).$$

*Démonstration.* Récurrence avec la définition de probabilité conditionnelle.

□

À partir de maintenant, $J \subset \mathbb{N}$ désignera un ensemble qui sera soit $\mathbb{N}$, soit de la forme $[\![1, n]\!]$.

**Définition 1.1.2.** *On appelle système complet d'évènement*[a] *toute suite dénombrable $(B_j)_{j \in J} \in \mathcal{F}^J$ d'évènements deux à deux disjoints telle que $\sum_{j \in J} \mathbb{P}(B_j) = 1$. Le système est dit fini (*resp. *infini) si $J$ est fini (*resp. *infini).*

*a.* Généralement appelé *quasi-complet* puisqu'ici la famille n'est pas une partition de $\Omega$.

**Proposition 1.1.3.** *Soit $(B_j)_{j \in J}$ un système complet d'évènements tel que $\mathbb{P}(B_j) > 0$ pour tout $j \in J$. Alors, pour tout $A \subset \bigcup_{j \in J} B_j$, on a*

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

*Démonstration.* Posons $\Omega_0 = \bigcup_{j \in J} B_j$. On a $\mathbb{P}(A) = \mathbb{P}(A \cap \Omega_0)$. On termine avec la $\sigma$-additivité de $\mathbb{P}$.

□

**Exemple.** Reprenons le même exemple, en supposant de plus qu'on ait autant de chance de piocher dans une urne que dans une autre. On a $\mathbb{P}(B) = 2/3 \times 1/2 + 1/4 \times 1/2 = 11/24$.

**Proposition 1.1.4** (formule de Bayes)**.** *Soit $(B_j)_{j \in J}$ un système complet tel que $\mathbb{P}(B_j) > 0$ pour tout $i \in I$. Alors, pour tout $A \in \mathcal{F}$ non négligeable, et $j \in J$, on a*

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\sum_{k \in J} \mathbb{P}(A|B_k)\mathbb{P}(B_k)}.$$

## 1.2 Espérance conditionnelle

**Définition 1.2.1.** *Soit $B \in \mathcal{F}$ un évènement de probabilité non nulle. On définit* ***l'espérance conditionnelle sachant*** *$B$ comme l'espérance de la probabilité $\mathbb{P}(\cdot|B)$, notée $\mathbb{E}[\cdot|B]$. Ainsi, si $X$ est une variable aléatoire positive ou intégrable, on a*

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

**Proposition 1.2.1.** *Soient $X$ une variable aléatoire intégrable, et $B$ un évènement non négligeable. Alors,*

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

*Démonstration.* Lorsque $X$ est une indicatrice, le résultat est immédiat (par définition). On peut ensuite montrer par linéarité le résultat lorsque $X$ est une fonction étagée positive. Par convergence monotone, on peut montrer le résultat lorsque $X$ est positive. Enfin, on traite le cas de $X$ en considérant les parties positive et négative de $X$.

□

Si $X$ est une variable aléatoire réelle discrète de support $\mathcal{S}(X) = \{x_j \,|\, j \in J\}$, alors $X = \sum_{j\in J} x_j \mathbb{1}_{X=x_j}$. On a donc

$$\mathbb{E}[X|B] = \sum_{j\in J} x_j \mathbb{P}(X = x_j|B).$$

Si $Y$ est une variable aléatoire discrète de support $\mathcal{S}(Y) = \{y_k \,|\, k \in K\}$, on définit

$$\mathbb{E}[X|Y = y_k] = \sum_{j\in J} x_j \mathbb{P}(X = x_j|Y = y_k).$$

On généralise l'espérance conditionnelle "sachant $Y = y_k$" à "sachant $Y$" de la manière suivante.

**Définition 1.2.2** (espérance conditionnelle discrète)**.** *Soit $X$ une variable aléatoire intégrable et $Y$ une variable aléatoire discrète de support $\mathcal{S}(Y) = \{y_k \,|\, k \in K\}$. L'espérance conditionnelle de $X$ sachant $Y$ est définie par*

$$\mathbb{E}[X|Y] = \sum_{k\in K} \mathbb{E}[X|Y = y_k]\mathbb{1}_{\{Y=y_k\}}.$$

*En d'autres termes, $\mathbb{E}[X|Y] = \mathbb{E}[X|Y = y_k]$ sur l'évènement $\{Y = y_k\}$.*

**Remarques.**

— $\mathbb{E}[X|Y]$ est une *variable aléatoire.*

— Avec les définitions données, on observe que $\mathbb{E}\big[\mathbb{E}[X|Y]\big] = \mathbb{E}[X]$. En effet, par linéarité de l'espérance et la formule des probabilités totales données précédemment,

$$\mathbb{E}\big[\mathbb{E}[X|Y]\big] = \sum_{k\in K} \mathbb{E}[X|Y = y_k]\mathbb{P}(Y = y_k) = \sum_{k\in K} \frac{\mathbb{E}[X\mathbb{1}_{\{Y=y_k\}}]}{\mathbb{P}(Y = y_k)}\mathbb{P}(Y = y_k).$$

Ainsi,

$$\mathbb{E}\big[\mathbb{E}[X|Y]\big] = \mathbb{E}\left[X\left(\sum_{k\in K} \mathbb{1}_{\{Y=y_k\}}\right)\right] = \mathbb{E}[X].$$

En effet, $\sum_{k\in K} \mathbb{1}_{\{Y=y_k\}} = 1$ presque sûrement par définition des $y_k$ comme le support de $Y$.

### 1.2.1 Probabilité sachant une variable aléatoire discrète

De la même manière que dans la définition 1.2.2, on généralise la probabilité conditionnelle “sachant $Y = y$” par “sachant $Y$”. On va donc définir une nouvelle variable aléatoire du type “$\mathbb{P}(A|Y)$”. Soit encore une fois une variable aléatoire $Y$ discrète de support $\mathcal{S}(Y) = \{y_j | j \in J\}$.

**Définition 1.2.3.** *On appelle probabilité conditionnelle sachant $X$ la fonction*

$$\mathbb{P}(\,\cdot\,|Y) : \left| \begin{array}{lcl} \mathcal{F} & \longrightarrow & [0,1]^\Omega \\ A & \longmapsto & \displaystyle\sum_{j \in J} \mathbb{P}(A|Y = y_j) \mathbb{1}_{\{Y = y_j\}}. \end{array} \right.$$

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

Puisque la définition de l'espérance conditionnelle assure $\mathbb{E}[\mathbb{1}_A|B] = \mathbb{P}(A|B)$ pour un évènement $B$ non négligeable, on peut vérifier lorsque $Y$ est discrète que l'espérance conditionnelle $\mathbb{E}[\,\cdot\,|Y]$ en définition de l'espérance conditionnelle discrète, et la probabilité conditionnelle $\mathbb{P}(\,\cdot\,|Y)$ définie ci-dessus sont naturellement liées par $\mathbb{E}[\mathbb{1}_A|Y] = \mathbb{P}(A|Y)$. Dans le cas général, par exemple lorsque $Y$ est une variable à densité, la définition de $\mathbb{P}(\,\cdot\,|Y)$ est plus complexe car les conditionnements par $\{Y = y\}$ sont singuliers (évènements négligeables).

### 1.2.2 Lois conditionnelles discrètes

Soit $(X, Y)$ un couple de variables aléatoires discrètes sur $(\Omega, \mathcal{F}, \mathbb{P})$.

**Définition 1.2.4** (loi conditionnelle)**.** *Soit $y$ tel que $\mathbb{P}(Y = y) \neq 0$. On appelle loi conditionnelle de de $X$ sachant $Y = y$ l'application définie sur le support de $X$ par*

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

*pour tout $x \in \mathcal{S}(X)$.*

Si $\mathbb{P}(Y = y) = 0$, on peut éventuellement lui donner une valeur arbitraire (généralement zéro).

**Proposition 1.2.2.** *Si $X$ et $Y$ sont indépendantes, alors la loi conditionnelle de $X$ sachant $Y = y$ ($y \in \mathcal{S}(Y)$) est la même que celle de $X$, i.e que*

$$\mathbb{P}_X(\,\cdot\,|Y = y) = \mathbb{P}_X.$$

*Autrement dit, le conditionnement par une variable aléatoire indépendante est sans effet.*

### 1.2.3 Lois conditionnelles à densité

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

$$x \longmapsto f_X(x) = \int_{\mathbb{R}} f(x, y)\,\mathrm{d}y \quad \text{et} \quad y \longmapsto f_Y(y) = \int_{\mathbb{R}} f(x, y)\,\mathrm{d}x.$$

Dans cette situation, on a un analogue de la formule de Bayes pour les densités avec la notion de densité conditionnelle.

**Définition 1.2.5.** *Soit $(X, Y)$ un couple de variables aléatoires réelles de densité $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$. On définit la densité conditionnelle de $X$ sachant $Y = y$ par*

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

*pour tout $x \in \mathbb{R}$.*

**Remarque.** La densité conditionnelle $f_{X|Y=y}$ définit la loi conditionnelle $\mathscr{L}(X|Y = y)$ de $X$ sachant $Y = y$. Il s'agit d'une fonction de la seule variable $x$ (la variable $y$ y apparaît seulement comme un paramètre). Comme pour la proposition 1.2.2, on a la proposition suivante.

**Proposition 1.2.3.** *Si les variables aléatoires $X$ et $Y$ sont indépendantes de densité $f_X$ et $f_Y$, alors les densités conditionnelles sont les densités marginales sont telles que pour tout $x \in \mathbb{R}$,*

$$f_{X|Y=y}(x) = f_X(x).$$

*Encore une fois, le conditionnement est sans effet quand les variables sont indépendantes.*

*Démonstration.* Lorsque $X$ et $Y$ sont indépendantes, $f(x,y) = f_X(x) = f_Y(y)$ pour tout $x, y \in \mathbb{R}$.

$\square$

# Chapitre 2

# Espérance conditionnelle

Soit $\mathcal{G}$ une sous-tribu de $\mathcal{F}$.

## 2.1 Introduction

**Définition 2.1.1.** *Soit $\mathcal{G}$ une sous-tribu de $\mathcal{F}$ et $X$ une variable aléatoire positive ou intégrable. L'espérance conditionnelle $\mathbb{E}[X|\mathcal{G}]$ est la presque sûrement unique variable aléatoire $Y$ telle que*

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

*(ii) Pour tout $A \in \mathcal{G}$, $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A]$.*

**Remarque.** Cette variable aléatoire est seulement définie presque sûrement. De plus, $X$ doit être intégrable positive ou intégrable pour donner un sens à $\mathbb{E}[X|\mathcal{G}]$.

*Démonstration.* — On suppose que $Y$ et $Y'$ vérifient les points $(i)$ et $(ii)$. Le second point assure que $\mathbb{E}[Y\mathbb{1}_A] = \mathbb{E}[Y'\mathbb{1}_A]$ pour tout $A \in \mathcal{G}$. Soit $\varepsilon > 0$, on pose $A_\varepsilon = \{Y - Y' \geq \varepsilon\} \in \mathcal{G}$. On a

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

Ainsi, $\mathbb{P}(A_\varepsilon) = 0$. On en déduit que

$$0 \leq \mathbb{P}\underbrace{\left(\bigcup_{\varepsilon\in\mathbb{Q}^+} A_\varepsilon\right)}_{=\{Y-Y'>0\}} \leq \sum_{\varepsilon\in\mathbb{Q}^+} \mathbb{P}(A_\varepsilon) = 0.$$

On en déduit que $\mathbb{P}(Y - Y' > 0) = 0$. Par symétrie des rôles, on détermine l'inégalité stricte inverse, et alors $\mathbb{P}(Y - Y' \neq 0) = 0$.

$\square$

**Proposition 2.1.1.** *Dans la définition précédente, le point $(ii)$ est équivalent à demander : (ii') Pour toute variable aléatoire $Z$ $\mathcal{G}$-mesurable bornée, on a*

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

*Démonstration.* — $(ii') \implies (ii)$. Prendre $Z = \mathbb{1}_A$ où $A \in \mathcal{G}$. En effet, $Z$ est alors $\mathcal{G}$-mesurable bornée.

— $(ii) \implies (ii')$. Par les arguments standards de la théorie de la mesure. En effet, $(ii)$ assure $(ii')$ pour une dans le cas où $Z$ est une indicatrice.

□

**Notations.**

— Lorsque $\mathcal{G} = \sigma(Y)$ (*i.e* que $\mathcal{G}$ est la tribu engendrée par la variable aléatoire $Y$), on note $\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)]$.

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

## 2.2 Exemples et premières propriétés

Des exemples simples mais fondamentaux sont les suivants.

(i) Si $X$ est une variable aléatoire $\mathcal{G}$-mesurable, alors $\mathbb{E}[X|\mathcal{G}] = X$. En effet, $X$ est $\mathcal{G}$-mesurable, et $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[X\mathbb{1}_A]$.

(ii) Si $X$ est une variable aléatoire constante à $X = c$, alors $\mathbb{E}[X|\mathcal{G}] = c$.

(iii) Si $X$ est indépendant de $\mathcal{G}$, alors $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$. En effet, $Y = \mathbb{E}[X]$ est constante et $\mathcal{G}$-mesurable. De plus, $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[X]\mathbb{E}[1_A] = \mathbb{E}[\mathbb{E}[X]\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A]$.

(iv) On prend la tribu grossière $\mathcal{G} = \{\varnothing, \Omega\}$. Alors, $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$.

(v) Si $\mathcal{G} = \mathcal{F}$, on a $\mathbb{E}[X|\mathcal{G}] = X$.

(vi) Soit $\Omega = \bigsqcup_{j\in J} \Omega_i$ avec $\mathbb{P}(\Omega_j) > 0$ pour tout $j \in J$. On considère la tribu $\mathcal{G} = \sigma\left(\{\Omega_j \mid j \in J\}\right)$. Alors, on a pour toute variable aléatoire $X$ positive ou intégrable,

$$\mathbb{E}[X|\mathcal{G}] \overset{p.s}{=} \sum_{j\in J} \frac{\mathbb{E}[X\mathbb{1}_{\Omega_j}]}{\mathbb{P}(\Omega_j)} \mathbb{1}_{\Omega_j},$$

*i.e* que la variable aléatoire $\mathbb{E}[X|\mathcal{G}]$ vaut $\frac{\mathbb{E}[X\mathbb{1}_{\Omega_j}]}{\mathbb{P}(\Omega_j)}$ sur $\Omega_j$. Le même énoncé reste vrai avec un système complet.

*Démonstration.* Montrons cette formule dans le cas simple où $\Omega = B \sqcup B^c$. On montre uniquement le résultat pour $X = \mathbb{1}_A$. On pose

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

$Y$ est bien $\mathcal{G}$-mesurable. De plus, pour tout $C \in \mathcal{G} = \{\varnothing, B, B^c, \Omega\}$.

— Si $C = \varnothing$, on a $\mathbb{1}_C = 0$, et on se trouve à montrer que $0 = 0$.

— Si $C = B$, on a $\mathbb{E}[\mathbb{1}_A\mathbb{1}_C] = \mathbb{P}(A \cap B)$. De plus,

$$\mathbb{E}[Y\mathbb{1}_C] = \mathbb{E}\left[\frac{\mathbb{E}[\mathbb{1}_A\mathbb{1}_B]}{\mathbb{P}(B)}]\mathbb{1}_B^2 + 0\right].$$

— Si $C = B^c$, *idem.*

— Si $C = \Omega$, $\mathbb{1}_C = 1$ et donc $\mathbb{E}[\mathbb{1}_1\mathbb{1}_C] = \mathbb{P}(A)$. Ainsi,

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

Ainsi, $\mathbb{E}[Y\mathbb{1}_C] = \mathbb{E}[\mathbb{1}_A(\mathbb{1}_B + \mathbb{1}_{B^c})] = \mathbb{P}(A)$ car $\mathbb{1}_B + \mathbb{1}_{B^c} = 1$. Pour le cas général, on utilisera la méthode habituelle (fonctions étagées, linéarité, convergence monotone, ...).

□

**Proposition 2.2.1.** *Si $X \in L^1(\mathcal{F})$, alors $\mathbb{E}[X|\mathcal{G}] \in L^1(\mathcal{G})$. De, plus, on a l'inégalité*

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

*Démonstration.* Soit $Y = \mathbb{E}[X|\mathcal{G}]$, et $A = \{Y > 0\}$. On a

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

De plus,

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

En combinant les inégalités, on obtient que

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

Or, $Y\mathbb{1}_A = Y^+$, et $Y\mathbb{1}_{A^c} = Y^+$. Ainsi,

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

En rassemblant, tous nos calculs, on trouve l'inégalité cherchée.

□

**Proposition 2.2.2.** *Soient $X$ et $Y$ deux variables aléatoires indépendantes, et soit $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ une fonction mesurable telle que $\mathbb{E}[|f(X,Y)|] < \infty$. On pose $g(y) = \mathbb{E}[f(X,y)]$ (fonction mesurable). Alors,*

$$\mathbb{E}[f(X,Y)|Y] = g(Y).$$

*Démonstration.* $g(Y)$ est bien $\sigma(Y)$-mesurable. Montrons maintenant que pour tout $A \in \sigma(Y)$,

$$\mathbb{E}[\mathbb{1}_A f(X,Y)] = \mathbb{E}[\mathbb{1}_A g(Y)].$$

Soit $A \in \sigma(Y)$, que l'on écrit sous la forme $A = Y^{-1}(B)$ où $B \in \mathcal{B}(\mathbb{R})$ (tribu borélienne sur $\mathbb{R}$). Alors,

$$\begin{aligned}
\mathbb{E}[\mathbb{1}_A f(X,Y)] &= \mathbb{E}[\mathbb{1}_B(Y) f(X,Y)] \\
&= \iint_{\mathbb{R}^2} \mathbb{1}_B(y) f(x,y) \mathbb{P}(\mathrm{d}x, \mathrm{d}y) \\
&= \int_{\mathbb{R}} \mathbb{1}_B(y) \left( \int_{\mathbb{R}} f(x,y) \mathrm{d}\mathbb{P}_X(x) \right) \mathrm{d}\mathbb{P}_Y(y) \\
&= \int_B g(y) \mathrm{d}\mathbb{P}_Y(y) \\
&= \mathbb{E}[\mathbb{1}_B(Y) g(Y)] = \mathbb{E}[\mathbb{1}_A g(Y)].
\end{aligned}$$

□

**Proposition 2.2.3** (linéarité). *Pour tout $X, Y \in L^1(\mathcal{F})$, et $a, b \in \mathbb{R}$,*

$$\mathbb{E}[aX + bY|\mathcal{G}] \overset{p.s}{=} a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Y|\mathcal{G}].$$

*On remarquera que l'égalité presque sûre a lieu sur un support qui dépend de $a$ et de $b$.*

*Démonstration.* D'abord, $a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Y|\mathcal{G}]$ est bien $\mathcal{G}$-mesurable. Pour tout $A \in \mathcal{G}$,

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

□

**Proposition 2.2.4** (monotonie). *Si $X$ et $Y$ sont deux variables aléatoires telles que $X \leq Y$, alors $\mathbb{E}[X|\mathcal{G}] \leq \mathbb{E}[Y|\mathcal{G}]$ presque sûrement.*

*Démonstration.* On montre d'abord que $Y \geq 0$ implique $\mathbb{E}[Y|\mathcal{G}] \geq 0$ presque sûrement. Soit $A_\varepsilon = \{\mathbb{E}[Y|\mathcal{G}] \leq -\varepsilon\} \in \mathcal{G}$ où $\varepsilon > 0$. Alors,

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

Ainsi, $\mathcal{P}(A_\varepsilon) = 0$, et donc

$$\mathbb{P}\left(\bigcup_{\varepsilon \in \mathbb{Q}^+} A_\varepsilon\right) = 0$$

($\bigcup_{\varepsilon \in \mathbb{Q}^+} A_\varepsilon = \{\mathbb{E}[Y|\mathcal{G}] < 0\}$). On en conclut que $\mathbb{P}(\mathbb{E}[Y|\mathcal{G}] \geq 0) = 1$. De manière générale, en supposant que les espérances conditionnelles existent, si $\mathbb{E}[Y|\mathcal{G}] = \infty$ ou $\mathbb{E}[X|\mathcal{G}] = -\infty$, la conclusion est immédiate. Sinon, $\mathbb{E}[X|\mathcal{G}] > -\infty$ et $\mathbb{E}[Y|\mathcal{G}] < \infty$. Par linéarité en écrivant $Y = (Y - X) + X$, on a

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

Or, nous avons déjà montré que $\mathbb{E}[(Y - X)|\mathcal{G}] \geq 0$.

□

**Proposition 2.2.5** (inégalité de TCHEBYCHEV). *Pour toute variable aléatoire $X$, et $\varepsilon > 0$,*

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

*Démonstration.* On a $\varepsilon^2\mathbb{1}_{\{|X|\geq\varepsilon\}} \leq |X|^2$, à intégrer avec $\mathbb{E}[\,\cdot\,|\mathcal{G}]$.

□

**Proposition 2.2.6** (convergence monotone conditionnelle (Beppo-Levi))**.** *Soit* $(X_n)_{n\in\mathbb{N}}$ *une suite de variables aléatoires* $L^1$ *positives croissantes telle que les* $X_n$ *aient pour limite supérieure la variable aléatoire* $X$*, telle que* $X$ *soit* $L^1$*. Alors, les* $\mathbb{E}[X_n|\mathcal{G}]$ *ont pour limite supérieure la variable aléatoire* $\mathbb{E}[X|\mathcal{G}]$*.*

*Démonstration.* On pose $Y_n = X - X_n \geq 0$. Montrons que $(Z_n)_n := (\mathbb{E}[Y_n|\mathcal{G}])_n$ a pour limite inférieure 0. Par monotonie de $\mathbb{E}[\,\cdot\,|\mathcal{G}]$, on sait que $(Z_n)_n$ décroît et est positive. Ainsi, $(Z_n)_n$ converge vers une certaine variable aléatoire $Z_\infty$ positive et $\mathcal{G}$-mesurable. Pour tout $A \in \mathcal{G}$, puisque $Y_n$ est dominée par $X$ (qui est $L^1$), on a par convergence dominée de l'espérance que

$$\mathbb{E}[\mathbb{1}_A Z_n] = \mathbb{E}[\mathbb{1}_A Y_n] \xrightarrow[n\to\infty]{} 0.$$

Or, $(Z_n)_n$ décroît donc pour tout $n \in \mathbb{N}^*$, donc $0 \leq Z_n \leq Z_1$. On en déduit que

$$\mathbb{1}_A Z_n \xrightarrow[n\to\infty]{p.s} \mathbb{1}_A Z_\infty.$$

Ainsi, par convergence dominée, on obtient que

$$\mathbb{E}[\mathbb{1}_A Z_n] \xrightarrow[n\to\infty]{} \mathbb{E}[\mathbb{1}_A Z_\infty].$$

Or, par unicité de la limite, nécessairement, $\mathbb{E}[\mathbb{1}_A Z_\infty] = 0$. On pose $A_\varepsilon = \{Z_\infty \geq \varepsilon\} \in \mathcal{G}$. On a alors

$$0 \leq \varepsilon\mathbb{P}(A_\varepsilon) = \mathbb{E}[\varepsilon\mathbb{1}_{A_\varepsilon}] \leq \underbrace{\mathbb{E}[Z_\infty \mathbb{1}_{A_\varepsilon}]}_{=0}.$$

Ainsi, $\mathbb{P}(A_\varepsilon) = 0$. Ainsi,

$$\mathbb{P}\left(\bigcup_{\varepsilon\in\mathbb{Q}^+} A_\varepsilon\right) = 0,$$

*i.e* que $Z_\infty \leq 0$ presque sûrement, donc $Z_\infty = 0$ presque sûrement.

□

**Lemme 2.2.1** (Fatou conditionnel)**.** *Soit* $(X_n)_{n\geq 1}$ *une suite de variables aléatoires positives telle que* $\liminf_{n\to\infty} X_n$ *soit* $L^1$*. Alors,*

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n \middle| \mathcal{G}\right] \leq \liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}].$$

*Démonstration.* (i) On suppose d'abord $(X_n)_{n\geq 1}$ bornée. On a, en remarquant que $(\inf_{k\geq n} X_k)_{n\geq 1}$ est une suite croissante bornée, que

$$\begin{aligned}
\mathbb{E}\left[\liminf_{n\to\infty} X_n \middle| \mathcal{G}\right] &= \mathbb{E}\left[\liminf_{n\to\infty}\left(\inf_{k\geq n} X_k\right)\middle| \mathcal{G}\right] \\
&= \liminf_{n\to\infty} \mathbb{E}\left[\inf_{k\geq n} X_k \middle| \mathcal{G}\right] \quad \text{par convergence monotone} \\
&\leq \liminf_{n\to\infty} \mathbb{E}\left[X_n|\mathcal{G}\right],
\end{aligned}$$

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

(ii) Dans le cas général, on applique le cas borné à $(\inf(X_n, p))_{n\geq 1}$ pour $p \in \mathbb{N}$. On a ainsi

$$\mathbb{E}\left[\liminf_{n\to\infty} \inf(X_n, p)\Big|\mathcal{G}\right] \leq \liminf_{n\to\infty} \underbrace{\mathbb{E}\left[\inf(X_n, p)|\mathcal{G}\right]}_{\leq \mathbb{E}[X_n|\mathcal{G}]}.$$

La suite $(\inf(X_n, p))_{p\geq 1}$ est croissante, donc il en est de même pour $(\liminf_{n\to\infty} \inf(X_n, p))_{p\geq 1}$. Alors, $(\mathbb{E}\left[\liminf_{n\to\infty} \inf(X_n, p)|\mathcal{G}\right])_{p\geq 1}$ est croissante aussi. Ainsi, en faisant tendre $p$ vers l'infini dans l'inégalité précédente, on obtient le lemme.

□

**Théorème 2.2.1** (convergence dominée conditionnelle)**.** *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires avec $|X_n| \leq Z \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ avec $X_n \xrightarrow[n\to\infty]{p.s} X$. Alors,*

$$\mathbb{E}[X_n|\mathcal{G}] \xrightarrow[n\to\infty]{p.s} \mathbb{E}[X|\mathcal{G}].$$

*Démonstration.* L'idée de la démonstration est la même que le théorème de convergence dominée classique. Par le lemme de Fatou conditionnelle,

$$\begin{aligned}
\mathbb{E}[2Z|\mathcal{G}] &\leq \mathbb{E}\left[\liminf_{n\to\infty}(2Z - |X_n - X|)|\mathcal{G}\right] \\
&\leq \liminf_{n\to\infty} \mathbb{E}[2Z - |X_n - X||\mathcal{G}] \\
&\leq \liminf_{n\to\infty} \mathbb{E}[2Z|\mathcal{G}] + \liminf_{n\to\infty} \left(-\mathbb{E}[|X_n - X||\mathcal{G}]\right) \\
&\leq \liminf_{n\to\infty} \mathbb{E}[2Z|\mathcal{G}] - \limsup_{n\to\infty} \mathbb{E}[|X_n - X||\mathcal{G}].
\end{aligned}$$

Ainsi, $\limsup_{n\to\infty} \mathbb{E}[|X_n - X||\mathcal{G}] = 0$ et donc $\lim_{n\to\infty} \mathbb{E}[|X_n - X||\mathcal{G}] = 0$. Le résultat est immédiat car pour toute variable aléatoire $Y$ positive ou aléatoire, on a $|\mathbb{E}[Y|\mathcal{G}]| \leq \mathbb{E}[|Y||\mathcal{G}]$. En effet,

$$\begin{cases}
\mathbb{E}[Y|\mathcal{G}] \leq \mathbb{E}[|Y||\ \mathcal{G}], \\
-\mathbb{E}[Y|\mathcal{G}] = \mathbb{E}[-Y|\mathcal{G}] \leq \mathbb{E}[|Y||\mathcal{G}].
\end{cases}$$

□

**Proposition 2.2.7.** *On a $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$.*

*Démonstration.* Prendre $A = \Omega$, alors $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbb{1}_\Omega] = \mathbb{E}[X\mathbb{1}_\Omega]$.

□

**Proposition 2.2.8** (Cauchy-Schwarz conditionnel)**.** *On a*

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

*Démonstration.* On remarque que pour tout $t \in \mathbb{R}$, $(X + tY)^2 \geq 0$. Alors, $\mathbb{E}[(X + tY)^2|\mathcal{G}] \geq 0$ presque sûrement. En développant,

$$\mathbb{E}[Y^2|\mathcal{G}]t^2 + 2\mathbb{E}[XY|\mathcal{G}]t + \mathbb{E}[X^2|\mathcal{G}] \geq 0 \text{ p.s pour tout } t \in \mathbb{R}.$$

Alors,
$$\mathbb{E}[Y^2|\mathcal{G}]t^2 + 2\mathbb{E}[XY|\mathcal{G}]t + \mathbb{E}[X^2|\mathcal{G}] \geq 0 \text{ pour tout } t \in \mathbb{Q} \text{ p.s.}$$
On a donc *presque sûrement* un polynôme du second degré en $t$, positif pour tout $t \in \mathbb{Q}$, donc pour tout $t \in \mathbb{R}$ par continuité. Son discriminant est donc négatif, *i.e* que

$$4\mathbb{E}[XY|\mathcal{G}] \leq 4\mathbb{E}[X^2|\mathcal{G}]\ \mathbb{E}[Y^2|\mathcal{G}].$$

□

**Proposition 2.2.9** (Jensen conditionnel)**.** *Soit $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, et $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ convexe telle que $\varphi(X) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Alors,*

$$\varphi\left(\mathbb{E}[X|\mathcal{G}]\right) \leq \mathbb{E}[\varphi(X)|\mathcal{G}] \ \textit{p.s.}$$

*Démonstration.* L'inégalité est une égalité immédiate lorsque $\varphi(x) = ax+b$. Dans le cas générale, $\varphi$ étant convexe, elle est en dessous de ses tangeantes : pour $x, y \in \mathbb{R}$, et $d_y \in [\varphi'_g(y), \varphi'_d(y)]$,

$$\varphi(x) \geq \varphi(y) + d_y(x-y).$$
On prend par exemple $d_y = \frac{1}{2}(\varphi'_g(y) + \varphi'_d(y))$. On utilise cette inégalité avec $y = \mathbb{E}[X|G]$ et $x = X$. On a alors

$$\mathbb{E}[\varphi(X)|\mathcal{G}] \geq \mathbb{E}[\varphi(\mathbb{E}[X|\mathcal{G}]) + d_{\mathbb{E}[X|\mathcal{G}]}(X - \mathbb{E}[X|\mathcal{G}])|\mathcal{G}].$$
En remarquant que $d_{\mathbb{E}[X|\mathcal{G}]}$ est $\mathcal{G}$-mesurable [1], on voit que

$$\mathbb{E}[\varphi(\mathbb{E}[X|\mathcal{G}]) + d_{\mathbb{E}[X|\mathcal{G}]}(X - \mathbb{E}[X|\mathcal{G}])|\mathcal{G}] = \varphi(\mathbb{E}[X|\mathcal{G}]) + d_{\mathbb{E}[X|\mathcal{G}]}\mathbb{E}[(X - \mathbb{E}[X|\mathcal{G}])|\mathcal{G}].$$

Or, $E[(X - \mathbb{E}[X|\mathcal{G}])|\mathcal{G}] = 0$.

□

**Proposition 2.2.10.** *Soit $p \geq 1$.*

*(i) On a presque sûrement,*

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

*(ii) De plus,*

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

*Démonstration.* (i) Inégalité de Jensen conditionnelle avec $\varphi(x) = |x|^p$.

(ii) S'obtient en prenant l'espérance dans $(i)$ avec

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

□

1. On va voir cette propriété en 2.2.11.

**Proposition 2.2.11.** *Soit $X$ une variable aléatoire $\mathcal{G}$-mesurable, $Y$ un variable aléatoire $L^1$ telle que $XY$ est $L^1$. Alors,*

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

*Démonstration.* $X\mathbb{E}[Y|\mathcal{G}]$ est un variable aléatoire $\mathcal{G}$-mesurable. On montre en suite que pour tout $A \in \mathcal{G}$,

$$\mathbb{E}[(X\mathbb{E}[X|\mathcal{G}])\mathbb{1}_A] = \mathbb{E}[XY\mathbb{1}_A].$$

D'abord, si $X = \mathbb{1}_B$ pour $B \in \mathcal{G}$, alors

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

Par linéarité, on a donc le résultat lorsque $X$ est étagée et $\mathcal{G}$-mesurable. Pour continuer, on suppose que $Y \geq 0$. Si $X$ est $G$-mesurable positive, on écrit $X = \lim_{n\to\infty} X_n$ où $(X_n)_n$ est une suite de variables aléatoires étagées croissante. Alors,

$$\mathbb{E}[X_n\mathbb{E}[Y|\mathcal{G}]\mathbb{1}_A] = \mathbb{E}[(X_nY)\mathbb{1}_A].$$

Puisque $(\mathbb{1}_AX_nY)_n$ tend de manière croissante vers $\mathbb{1}_AXY$, alors $(\mathbb{1}_AX_n\mathbb{E}[Y|\mathcal{G}])_n$ tend aussi de manière croissante vers $\mathbb{1}_AX\mathbb{E}[Y|\mathcal{G}]$. Par convergence monotone, on en déduit que le résultat dans cas. Si $X$ est de signe quelconque, on écrit $X = X^+ - X^-$, et on applique le cas précédent sur $X^+$ et $X^-$. On a donc montré le résultat pour toute variable aléatoire $X$ $\mathcal{G}$-mesurable, et $Y$ positive. Finalement, si $Y$ est de signe quelconque, on écrit $Y = Y^+ - Y^-$ et le cas précédent s'applique à $Y^+$ et $Y^-$.

□

**Proposition 2.2.12** (cascasde)**.** *Soit $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F}$ des sous-tribus ordonnées. On a les résultats suivants.*

*(i)* $\mathbb{E}[\mathbb{E}[X|\mathcal{G}_1]|\mathcal{G}_2] = \mathbb{E}[X|\mathcal{G}_1]$ *presque sûrement ;*

*(ii)* $\mathbb{E}[\mathbb{E}[X|\mathcal{G}_2]|\mathcal{G}_1] = \mathbb{E}[X|\mathcal{G}_1]$ *presque sûrement.*

*Démonstration.* (i) $\mathbb{E}[X|\mathcal{G}_1]$ est $\mathcal{G}_2$-mesurable puisque $\mathcal{G}_1$-mesurable.

(ii) $\mathbb{E}[X|\mathcal{G}_1]$ est $\mathcal{G}_1$-mesurable. Pour tout $A \in \mathcal{G}_1(\subset \mathcal{G}_2)$, on a $\mathbb{E}[\mathbb{E}[X|\mathcal{G}_2]\mathbb{1}_A] = \mathbb{E}[X\mathbb{1}_A]$. Alors,

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}_1]\mathbb{1}_A] = \mathbb{E}[X\mathbb{1}_A].$$

□

**Remarque.** En général, il est impossible d'inverser les conditionnements, *i.e* écrire quelque chose du genre "$\mathbb{E}[\mathbb{E}[X|\mathcal{G}_1]|\mathcal{G}_2] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}_2]|\mathcal{G}_1]$".

**Contre-exemple.** Prendre $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \lambda_1)$, $\mathcal{G}_1 = \sigma([0,1/2])$ et $\mathcal{G}_2 = \sigma([0,1/3])$, et $X = \mathbb{1}_{[1/4,1/3]}$.

**Proposition 2.2.13.** *On a les propriétés suivantes.*

*(i) Soit $X$ une variable aléatoire $L^1$ positive indépendante de $\mathcal{G}$. Alors, $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$ presque sûrement.*

*(ii) Soit $\mathcal{G}$ et $\mathcal{H}$ des sous-tribus de $\mathcal{F}$, et $X$ une variable aléatoire $L^1$ telle que $\mathcal{H}$ soit indépendante de $\sigma(X, \mathcal{G}) = \sigma(\sigma(X) \cup \mathcal{G})$. Alors,*

$$\mathbb{E}[X|\sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X|\mathcal{G}].$$

*Démonstration.* (i) $\mathbb{E}[X]$ en tant que fonction est constante, et donc $\mathcal{G}$-mesurable. Si $A \in \mathcal{G}$, on a

$$\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[X]\mathbb{E}[\mathbb{1}_A] = \mathbb{E}[\mathbb{E}[X]\mathbb{1}_A]$$

par indépendance.

(ii) On note $Y = \mathbb{E}[X|\mathcal{G}]$. On pose

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

$\mathcal{M}$ est une classe monotone.

— $\Omega \in \mathcal{M}$.

— $\mathcal{M}$ est stable par différence : soient $A$ et $B \in \mathcal{M}$ avec $B \subset A$. On a

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

— Soit $(A_n)_n$ une suite croissante d'évènements dans $\mathcal{M}$. On a

$$\mathbb{1}_{\bigcup_{n\in\mathbb{N}} A_n} = \lim_{n\to\infty} \mathbb{1}_{A_n}.$$

Alors,

$$\mathbb{E}\left[Y\mathbb{1}_{\bigcup_{n\in\mathbb{N}} A_n}\right] = \mathbb{E}\left[\lim_{n\to\infty} Y\mathbb{1}_{A_n}\right] = \lim_{n\to\infty} \underbrace{\mathbb{E}\left[Y\mathbb{1}_{A_n}\right]}_{\mathbb{E}[X\mathbb{1}_{A_n}]} = \mathbb{E}\left[X\mathbb{1}_{\bigcup_{n\in\mathbb{N}} A_n}\right].$$

On considère $\mathcal{P} = \{B \cap C \,|\, (B, C) \in \mathcal{G} \times \mathcal{H}\}$. C'est un $\pi$-système (car stable par intersection), et $\mathcal{P} \subset \mathcal{M}$. Pour terminer, on observe que $\sigma(\mathcal{P}) = \sigma(\mathcal{G}, \mathcal{H})$ car alors $\sigma(\mathcal{G}, \mathcal{H}) \subset \mathcal{M}$, *i.e* que pour tout $A \in \sigma(\mathcal{G}, \mathcal{H})$,

$$\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A].$$

Pour montrer cela, on montre $\mathcal{P} \subset \sigma(\mathcal{G}, \mathcal{H})$, puis que $\mathcal{G} \subset \mathcal{P}$ et $\mathcal{H} \subset \mathcal{P}$.

□

**Remarque.** Sans l'indépendance entre $\mathcal{H}$ et $\sigma(X, \mathcal{G})$, la conclusion est fausse. Soit $\varepsilon_1$ et $\varepsilon_2$ indépendantes identiquement distribuées suivant une loi de Rademacher de paramètre 1/2. Soit $X = \varepsilon_1\varepsilon_2$, $\mathcal{H} = \sigma(\varepsilon_1)$ et $\mathcal{G} = \sigma(\varepsilon_2)$. $X$ est de même loi que $\varepsilon_1$ et $\varepsilon_2$. De plus, $\mathcal{H}$ est indépendant de $\mathcal{G}$ et de $\sigma(X)$.

## 2.3 Espérance conditionnelle dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$

Dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$, l'espérance s'interprète comme un produit scalaire :

$$\mathbb{E}[XY] = \langle X, Y \rangle .$$

**Théorème 2.3.1.** *Soit $X \in L^2(\mathcal{F})$. Alors, $\mathbb{E}[X|\mathcal{G}]$ est la projection orthogonale de $X$ sur $L^2(\mathcal{G})$.*

*Démonstration.* Soit $Y = \mathbb{E}[X|\mathcal{G}]$. On montre que $Y$ est la projection orthogonale de $X$ sur $L^2(\mathcal{G})$. On a

$$\begin{aligned} \|X - Z\|^2 &= \mathbb{E}[(X-Z)^2] \\ &= \mathbb{E}[(X - Y + Y - Z)^2] \\ &= \mathbb{E}[(X-Y)^2] + 2\underbrace{\mathbb{E}[(X-Y)(Y-Z)]}_{0} + \mathbb{E}[(Y-Z)^2]. \end{aligned}$$

En effet, $\mathbb{E}[(X-Y)(Y-Z)] = \mathbb{E}[\mathbb{E}[(X-Y)(Y-Z)|\mathcal{G}]] = \mathbb{E}[(Y-Z)\mathbb{E}[X-Y|\mathcal{G}]] = 0$ car $Y = \mathbb{E}[X|\mathcal{G}]$. On en déduit que l'infimum $\|X-Z\|^2$ est atteint en $Z = Y$.

□

**Proposition 2.3.1** (Cas gaussien). *On a les résultats suivants.*

*(i) Soit $(X, Y)$ un vecteur gaussien, centré, non dégénéré (i.e de matrice de covariance non inversible). Alors,*

$$\mathbb{E}[X|Y] = \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(Y)} Y.$$

*(ii) Dans le cas non centré,*

$$\mathbb{E}[X|Y] = \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(Y)} (Y - \mathbb{E}[Y]) + \mathbb{E}[X].$$

*(iii) Soit $Z = (Z_1, \ldots, Z_d)$ un vecteur gaussien centré de matrice de covariance $\Sigma$, i.e que $Z \sim \mathcal{N}(0, \Sigma)$. Soient $a = (a_1, \ldots, a_d)^T$ et $b = (b_1, \ldots, b_d)^T \in \mathbb{R}^d$ $X = \sum_{i=1}^d a_i Z_i$ et $Y = \sum_{i=1}^d b_i Z_i$. Alors,*

$$\mathbb{E}[X|Y] = \frac{a^T \Sigma b}{b^T \Sigma b} Y \ p.s.$$

**Remarque.** Nous avions vu dans le cas $L^2$ que l'espérance conditionnelle s'interprète comme une projection orthogonale. Encore mieux, l'espérance $\mathbb{E}[X|Y]$ conditionnelle dans le cas gaussien centré non dégénéré s'interprète comme la projection orthogonale de $X$ sur la droite portée par $Y$. Dans le cas centré, un moyen simple de déterminer (sans apprendre de formule) $\mathbb{E}[X_1|X_2, \ldots, X_n]$ est de voir que $\mathbb{E}[X_1|X_2, \ldots, X_n] \in \mathrm{Vect}(X_2, \ldots, X_n)$.

*Démonstration.* (i) Soit $c = \frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]}$. Le vecteur $(X - cY, Y)$ est gaussien car

$$\alpha(X - cY) + \beta Y = \alpha X + (\beta - c\alpha) Y$$

suit une loi normale puisque $(X, Y)$ est gaussien. On a

$$\operatorname{cov}(X - cY, Y) = \operatorname{cov}(X, Y) - c \operatorname{cov}(Y, Y) = 0 \text{ car } c \text{ a été bien choisi.}$$

Bilan : $(X - cY, Y)$ est gaussien et $\operatorname{cov}(X - cY, Y) = 0$, donc $X - cY$ et $Y$ sont orthogonaux. Montrons maintenant que $cY = P_{L^2(\sigma(Y))}(X)$. Soit $Z \in L^2(\sigma(Y))$. Selon le lemme de Doob-Dynken, il existe une fonction mesurable $h$ telle que $Z = h(Y)$. Ainsi,

$$\begin{aligned}
\langle X - cY, Z \rangle &= \mathbb{E}[(X - cY)Z] \\
&= \mathbb{E}[(X - cY)h(Y)] \\
&= \mathbb{E}[\mathbb{E}[X - cY h(Y)|Y]] \\
&= \mathbb{E}[\mathbb{E}[X - cY|Y]h(Y)] \\
&= \mathbb{E}[\mathbb{E}[X - cY]h(Y)] \text{ par indépendance de } X - cY \text{ et } Y. \\
&= 0.
\end{aligned}$$

Ainsi, $X - cY$ est orthogonal à $L^2(\sigma(Y))$, et $cY \in L^2(\sigma(Y))$. On en déduit que $cY = P_{L^2(\sigma(Y))}(X)$.

(ii) $(X - \mathbb{E}[X], Y - \mathbb{E}[X])$ est un vecteur gaussien centré, le premier point nous affirme donc que

$$\mathbb{E}[X|Y] - \mathbb{E}[X] = \mathbb{E}[X - \mathbb{E}[X]|Y] = \frac{\operatorname{cov}(X, Y)}{\operatorname{var}(Y)}(Y - \mathbb{E}[Y]).$$

(iii) Soit $X = \langle a, Z \rangle$ et $Y = \langle b, Z \rangle$ deux combinaisons linéaires des variables coordonnées de $Z$. Pour $\alpha, \beta \in \mathbb{R}$, on a

$$\alpha X + \beta Y = \sum_{i=1}^{d} (\alpha a_i + \beta b_i) Z_i.$$

D'après le premier point,

$$\mathbb{E}[X|Y] = \frac{\operatorname{cov}(X, Y)}{\operatorname{var}(Y)} Y \text{ p.s.}$$

Or, $\operatorname{var}(Y) = \operatorname{var}(\langle b, Z \rangle) = b^T \sigma b$. Ainsi,

$$\operatorname{cov}(X, Y) = \sum_{i,j=1}^{d} a_i b_j \underbrace{\operatorname{cov}(Z_i, Z_j)}_{\Sigma_{i,j}} = a^T \Sigma b.$$

□

## 2.4 Lois conditionnelles

### 2.4.1 Définitions et exemples

**Définition 2.4.1.** *Soit $(S,\mathcal{A})$ et $(T,\mathcal{B})$ deux espaces mesurables. On appelle* ***noyau de probabilité*** *(ou de transition) toute application*

$$\nu : \mathcal{A} \times T \longrightarrow [0,1]$$

*telle que*

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

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

**Exemples.**

— Soit $\mu$ une mesure $\sigma$-finie sur $(S,\mathcal{A})$ et $h : S \times T \longrightarrow \mathbb{R}_+$ une fonction mesurable telle que pour tout $y \in T$,

$$\int_S h(x,y)\,\mathrm{d}\mu(x) = 1.$$

Alors, l'application $\nu$ définie pour tout $(A,y) \in \mathcal{A} \times T$ par

$$\nu(A,y) = \int_A h(x,y)\,\mathrm{d}\mu(x)$$

définit un noyau de probabilité.

— Soit $(X,Y)$ un couple sur $\mathbb{R}^2$ de densité $(x,y) \longmapsto f(x,y)$. On reprend le même exemple avec $(S,\mathcal{A}) = (T,\mathcal{B}) = (\mathbb{R},\mathcal{B}(\mathbb{R}))$ et $\mu = \lambda$ la mesure de Lebesgue, et

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

Cela permet de définir le noyau pour tout $(A,y) \in \mathcal{A} \times T$ par

$$\nu(A,y) = \begin{cases} \displaystyle\int_A \frac{f(x,y)}{f_Y(y)}\,\mathrm{d}x & \text{lorsque } f_Y(y) \neq 0, \\ \delta_{s_0}(A) & \text{sinon, où } s_0 \in S = \mathbb{R} \text{ est quelconque.} \end{cases}$$

— Si $(X,Y)$ est un couple discret, on définit un noyau de probabilité en posant

$$\nu(A,y) = \begin{cases} \mathbb{P}(X \in A | Y = y) & \text{lorsque } \mathbb{P}(Y = y) \neq 0, \\ \delta_{s_0}(A) & \text{sinon.} \end{cases}$$

**Proposition 2.4.1.** *On a les propriétés suivantes.*

*(i) Soit $h$ une fonction mesurable positive ou bornée sur $(S, \mathcal{A})$. Alors, l'application définie pour tout $y \in T$ par*

$$\varphi(y) = \int_S h(x)\nu(\mathrm{d}x, y)$$

*est une fonction mesurable sur $(T, \mathcal{B})$.*

*(ii) Si $\eta$ est une probabilité sur $(T, \mathcal{B})$, alors l'application définie pour tout $A \in \mathcal{A}$ par*

$$\mu(A) = \int_T \nu(A, y)\,\mathrm{d}\eta(y)$$

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

*(i)* Si $h = \mathbb{1}_A$, alors $\varphi(y) = \nu(A, y)$. C'est une fonction mesurable par rapport à $y$ d'après la définition 2.4.1. Supposons maintenant $h$ étagée sous la forme $h = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}$. Alors par linéarité,

$$\varphi(y) = \sum_{i=1}^n \alpha_i \nu(A_i, y).$$

On suppose cette fois $h$ mesurable positive. Alors $h$ est limite croissante d'une suite de fonctions étages $(h_n)_n$ pour lesquelles

$$\varphi_n(y) = \int_S h_n(x)\nu(\mathrm{d}x, y)$$

est mesurable. Alors, par convergence, $\varphi = \lim_{n\to\infty} \varphi_n$ est mesurable.

*(ii)* On a $\mu(\varnothing) = \int_T \nu(\varnothing, y)\,\mathrm{d}\eta(y) = 0$. Soit maintenant $(A_i)_i \in \mathcal{A}^I$ une famille dénombrable de mesurables deux à deux disjoints. Alors,

$$\begin{aligned}
\mu\left(\bigsqcup_{i\in I} A_i\right) &= \int_T \nu\left(\bigsqcup_{i\in I} A_i, y\right)\mathrm{d}\eta(y) \\
&= \int_T \sum_{i\in I} \nu(A_i, y)\,\mathrm{d}\eta(y) \\
&= \sum_{i\in I} \int_T \nu(A_i, y)\,\mathrm{d}\eta(y) \\
&= \sum_{i\in I} \mu(A_i).
\end{aligned}$$

Enfin,

$$\mu(S) = \int_T \underbrace{\nu(S, y)}_{=1}\,\mathrm{d}\eta(y) = \int_T \mathrm{d}\eta(y) = 1$$

car $\eta$ est une probabilité.

**Définition 2.4.2.** *Soient $X$ et $Y$ 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*

$$\nu : \mathcal{A} \times T \longrightarrow [0,1]$$

*tel que pour toute fonction $h : S \longrightarrow \mathbb{R}$ mesurable et positive, on a*

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

### 2.4.2 Existence des lois conditionnelles.

L'existence des lois conditionnelles découle d'un théorème difficile et admis.

**Théorème 2.4.1** (Jirina)**.** *Soient $(S, \mathcal{A})$ et $(T, \mathcal{B})$ espaces mesurables. On suppose que $S$ est un espace polonais (i.e métrique, complet, et séparable), et que $\mathcal{A} = \mathcal{B}(S)$. Alors, il existe une loi conditionnelle de $X$ sachant $Y$ comme dans la définition précédente.*

### 2.4.3 Unicité des lois conditionnelles.

Si $\nu$ et $\nu'$ sont deux noyaux vérifiant la définition, alors pour tout $h = \mathbb{1}_A$,

$$\nu(A, y) = \mathbb{E}[\mathbb{1}_A(X)|Y] = \nu'(A, Y)\, p.s.$$

On en déduit que pour tout $A \in \mathcal{A}$, presque sûrement, $\nu(A, y) = \nu'(A, y)$ pour $\mathbb{P}_X$-presque chaque $y \in T$. Dans le cas où les mesures de probabilités sur $(S, \mathcal{A})$ sont caractérisées par leurs valeurs sur une famille dénombrables. On en déduit que

$$\nu(\,\cdot\,, y) = \nu'(\,\cdot\,, y)$$

pour $\mathbb{P}_Y$-presque chaque $y \in T$. C'est en ce sens qu'il y a unicité des lois conditionnelles.

### 2.4.4 Cadres usuels

On donne maintenant les différents cadres et possibilités dans lesquels on rencontre ces lois conditionnelles.

**Proposition 2.4.2** (loi conditionnelle à densité)**.** *Soit $(X, Y)$ un couple aléatoire de densité $f$. La loi conditionnelle de $X$ sachant $Y$ est donnée par le noyau de probabilité définie pour tout $(A, y) \in \mathcal{A} \times T$ par*

$$\nu(A, y) = \begin{cases} \displaystyle\int_A \frac{f(x,y)}{f_Y(y)}\, dx & \text{lorsque } f_Y(y) \neq 0, \\ \delta_{s_0}(A) & \text{sinon, où } s_0 \in S = \mathbb{R} \text{ est quelconque.} \end{cases}$$

**Remarque.** On a donc $f_{X|Y=y}(x) = \frac{f(x,y)}{f_Y(y)}$ lorsque $f_Y(y) \neq 0$.

*Démonstration.* Soit $h$ et $g$ des fonctions mesurables positives. On a

$$\begin{aligned}
\mathbb{E}\Bigg[\underbrace{\left(\int_S h(x)\nu(\mathrm{d}x,Y)\right)}_{\varphi(Y)} g(Y)\Bigg] &= \int_{\mathcal{S}(Y)} \varphi(y)g(y)\,\mathrm{d}\mathbb{P}_Y(y) \\
&= \int_{\mathcal{S}(Y)} \varphi(y)g(y)f_Y(y)\,\mathrm{d}(y) \\
&= \int_{\mathcal{S}(Y)} \left(\int_{\mathcal{S}(X)} h(x)\nu(\mathrm{d}x,y)\right) g(y)f_Y(y)\,\mathrm{d}y \\
&= \int_{\mathcal{S}(X)} h(x)f(x,y)g(y)\,\mathrm{d}x\mathrm{d}y.
\end{aligned}$$

On a donc établi que pour tout $g$ mesurable positive,

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

Dans le cas où $g = \mathbb{1}_B$, on a

$$\mathbb{E}[\varphi(Y)\mathbb{1}_B(X)] = \mathbb{E}[h(X)\mathbb{1}_B(Y)].$$

Autrement, dit pour tout $A = Y^{-1}(B) \in \sigma(Y)$, $\mathbb{E}[\varphi(Y)\mathbb{1}_A] = \mathbb{E}[h(X)]\mathbb{1}_A]$, d'où $\mathbb{E}[h(X)|\ Y] = \varphi(Y)$.

□

**Remarque.** La valeur $\nu$ en $(A, y)$ lorsque $f_Y(y) = 0$ n'a aucune importance, ce qui est pourquoi on peut lui donner une quelconque valeur.

**Proposition 2.4.3** (loi conditionnelle discrète)**.** *Soit $(X, Y)$ un couple discret. Alors, la loi conditionnelle de $X$ sachant $Y$ est donnée par le noyau de probabilité définie pour tout $(A, y) \in \mathcal{A} \times T$ par*

$$\nu(A,y) = \begin{cases} \mathbb{P}(X \in A|Y = y) & \textit{lorsque } \mathbb{P}(Y = y) \neq 0, \\ \delta_{s_0}(A) & \textit{sinon, où } s_0 \in S = \mathbb{R} \textit{ est quelconque.} \end{cases}$$

*Démonstration.* Il s'agit de montrer que

$$\mathbb{E}[h(X)|Y] = \int_{\mathcal{S}(X)} \nu(\mathrm{d}x, Y) = \varphi(Y).$$

D'abord, $\varphi(Y)$ est bien $\sigma(Y)$-mesurable. Ensuite, pour tout $A = Y^{-1}(B) \in \mathcal{S}(Y)$ (avec $B \in \mathcal{B}(\mathbb{R})$), $\mathbb{E}[h(X)\mathbb{1}_A] = \mathbb{E}[\varphi(Y)\mathbb{1}_A]$. On en déduit que $\mathbb{E}[h(X)\mathbb{1}_B(Y)] = \mathbb{E}[\varphi(Y)\mathbb{1}_B(Y)]$. En effet,

$$\begin{aligned}
\mathbb{E}[\varphi(Y)\mathbb{1}_B(Y)] &= \sum_{y\in\mathcal{S}(Y)} \varphi(y)\mathbb{1}_B(y)\mathbb{P}(Y = y) \\
&= \sum_{y\in\mathcal{S}(Y)} \left(\int_{\mathcal{S}(X)} h(x)\nu(\mathrm{d}x,y)\right) \mathbb{1}_B(y)\mathbb{P}(Y = y) \\
&= \sum_{y\in\mathcal{S}(Y)} \left(\sum_{x\in\mathcal{S}(X)} h(x)\mathbb{P}(X = x|Y = y)\right) \mathbb{1}_B(y)\mathbb{P}(Y = y) \\
&= \sum_{(x,y)\in\mathcal{S}(X)\times\mathcal{S}(Y)} h(x)\mathbb{1}_B(y)\mathbb{P}(X = x, Y = y) \\
&= \mathbb{E}[h(X)\mathbb{1}_B(Y)] = \mathbb{E}[h(X)\mathbb{1}_A].
\end{aligned}$$

□

### 2.4.5 Retour au conditionnement sachant "$Y = y$"

**Définition 2.4.3.** *Lorsque la loi conditionnelle de $X$ sachant $Y$ existe, on pose $\mathbb{P}_Y$-presque chaque $y \in T$,*

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

*Lorsque le noyau existe, on a*

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

*On appelle loi conditionnelle de $X$ sachant $Y = y$ pour $\mathbb{P}_Y$-presque chaque $y \in T$ la loi*

$$\mathscr{L}(X | Y = y) = \nu(\,\cdot\,, y).$$

**Proposition 2.4.4.** *$X$ et $Y$ sont indépendants si et seulement si $\mathscr{L}(X|Y = y)$ existe pour $\mathbb{P}_Y$-presque chaque $y \in T$ et ne dépend pas de $y$.*

*Démonstration.* $\boxed{\Longrightarrow}$ Supposons $X$ et $Y$ indépendants. Alors,

$$\mathbb{E}[h(X)|Y] = \mathbb{E}[h(X)] = \int_{\mathcal{S}(X)} h(x) \, \mathrm{d}\mathbb{P}_X(x).$$

On a donc un noyau $\nu(\,\cdot\,, y) = \mathbb{P}_X$.

$\boxed{\Longleftarrow}$ Réciproquement , si $\mathscr{L}(X|Y = y) = \nu(\,\cdot\,, y)$ ne dépend pas de $y$. On note donc $\nu(\,\cdot\,) = \nu(\,\cdot\,, y)$. Pour tout $A, B \in \mathcal{B}(\mathbb{R})$, on a

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

$\square$

**Proposition 2.4.5** (désintégration des lois)**.** *Soient $X$ et $Y$ des variables aléatoires à valeurs respectivement dans $(S, \mathcal{A})$ et $(T, \mathcal{B})$, admettant un noyau $\nu$ qui définit la loi conditionnelle de $X$ sachant $Y$. Alors, pour tout $(A, B) \in \mathcal{A} \times \mathcal{B}$,*

$$\mathbb{P}(X \in A, Y \in B) = \int_B \mathbb{P}(X \in A | Y = y) \, \mathrm{d}\mathbb{P}_Y(y).$$

*Démonstration.* On a

$$\begin{aligned}
\mathbb{P}(X \in A, Y \in B) &= \mathbb{E}[\mathbb{1}_A(X)\mathbb{1}_B(Y)] \\
&= \mathbb{E}[\nu(A,Y)\mathbb{1}_B(Y)] \\
&= \int_B \nu(A,y)\,\mathrm{d}\mathbb{P}_Y(y) \quad \text{selon le lemme de transfert.}
\end{aligned}$$

□

Selon les arguments standards de la théorie de la mesure, on peut généraliser le résultat.

**Proposition 2.4.6** (Fubini conditionnel)**.** *Soient $X$ et $Y$ des variables aléatoires à valeurs respectivement dans $(S,\mathcal{A})$ et $(T,\mathcal{B})$, admettant un noyau $\nu$ qui définit la loi conditionnelle de $X$ sachant $Y$.*

*(i) Pour tout fonction mesurable $f : (S \times T, \mathcal{A} \otimes \mathcal{B}) \longrightarrow \mathbb{R}$ **positive**, alors l'application*

$$y \longmapsto \int_S f(x,y)\mathbb{P}_X(\mathrm{d}x|Y=y)$$

*est mesurable, et*

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

*(ii) Pour tout fonction $f : (S\times T, \mathcal{A}\otimes\mathcal{B}) \longrightarrow \mathbb{R}$ $\mathbb{P}_{(X,Y)}$-**intégrable**, alors pour $\mathbb{P}_Y$-presque chaque $y \in T$, l'application $f(\,\cdot\,,y)$ est $\mathbb{P}_X(\mathrm{d}x|Y=y)$-intégrable. De plus,*

$$y \longmapsto \int_S f(x,y)\mathbb{P}_X(\mathrm{d}x|Y=y)$$

*est $\mathbb{P}_Y$-mesurable, et l'on a*

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

**Remarque.** En notant, $\varphi_f(y) = \int_S f(x,y)\mathbb{P}_X(\mathrm{d}x|Y=y)$, alors pour toute fonction $f$ mesurable telle que $f(X,Y) \in L^1(\mathcal{A}\otimes\mathcal{B})$, alors

$$\mathbb{E}[f(X,Y)|Y] = \varphi_f(Y).$$

Pour $B \in \mathcal{B}$, en appliquant le théorème de Fubini conditionnel à $\mathbb{1}_B(y)f(x,y)$, on a

$$\begin{aligned}
\mathbb{E}[\mathbb{1}_B(Y)f(X,Y)] &= \int_T\int_S \mathbb{1}_B(y)f(x,y)\mathbb{P}_X(\mathrm{d}x|Y=y)\mathbb{P}_Y(\mathrm{d}y) \\
&= \int_T \mathbb{1}_B(y)\varphi_f(y)\mathbb{P}_Y(\mathrm{d}y) \\
&= \mathbb{E}[\mathbb{1}_B(Y)\varphi_f(Y)].
\end{aligned}$$

Lorsque pour une fonction $h$ mesurable telle que $h(f(X,Y)) \in L^1(\mathcal{A}\otimes\mathcal{B})$, par transfert, on a

$$\mathbb{E}[h(f(X,Y))|Y] = \varphi_{h\circ f}(Y) = \int_S h(f(x,y))\nu(\mathrm{d}x,y) = \int ??????$$

Par conséquent,

$$\mathscr{L}(f(X,Y)|Y) = \nu(\,\cdot\,,Y) \circ f(\,\cdot\,,Y)^{-1}.$$

En faisant de même avec $h(f(X,y))$, on obtient que

$$\mathscr{L}(f(X,y)|Y) = \nu(\,\cdot\,,Y) \circ f(\,\cdot\,,y)^{-1}.$$

On obtient alors le résultat suivant.

**Proposition 2.4.7.** *On a*

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

*et*

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

# Chapitre 3

# Martingales et filtrations

On considère dans ce chapitre un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

## 3.1 Filtration et mesurabilité

**Définition 3.1.1.** *Soit $(\mathcal{F}_n)_{n\in\mathbb{N}}$ une suite de sous-tribus de $\mathcal{F}$. On dit que $(\mathcal{F}_n)_n$ est une* ***filtration*** *lorsque qu'elle est croissante pour l'inclusion.*

**Définition 3.1.2.** *Soit $(\mathcal{F}_n)_n$ une filtration. On dit qu'une suite de variables aléatoires $(X_n)_{n\in\mathbb{N}}$ est* ***adaptée*** *à la filtration $(\mathcal{F}_n)_n$ si pour tout $n \in \mathbb{N}$, $X_n$ est $\mathcal{F}_n$-mesurable.*

**Remarque.** Très souvent, on considéra les filtrations comme des suites indexées par $\mathbb{N}^*$ (car en général, $\mathcal{F}_0$ est la tribu grossière $\{\varnothing, \Omega\}$).

**Exemples.**

— Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires. On pose $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ (*i.e* la plus petite tribu rendant mesurable $X_1, \ldots, X_n$). On pose $\mathcal{F}_0 = \{\varnothing, \Omega\}$. Alors $(\mathcal{F}_n)$ est clairement une filtration, $(X_n)_n$ est adaptée à $(\mathcal{F}_n)_n$. La suite $(\mathcal{F}_n)_{n\geq 0}$ est appelée *filtration canonique* de la suite $(X_n)_{n\geq 1}$.

— *Filtration diadique.* On considère ici $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \lambda)$. On pose

$$\mathcal{F}_n = \sigma\Bigg(\underbrace{\left[\frac{i-1}{2^n}, \frac{i}{2^n}\right[ \;:\; 1 \leq i \leq 2^n}_{\mathcal{D}_n}\Bigg).$$

On a $\mathcal{D}_n \subset \sigma(\mathcal{D}_{n+1})$, donc $\sigma(\mathcal{D}_n) \subset \sigma(\mathcal{D}_{n+1})$. En effet, on a

$$\left[\frac{i-1}{2^{n+1}}, \frac{i}{2^{n+1}}\right[ = \underbrace{\underbrace{\left[\frac{2i-2}{2^{n+1}}, \frac{2i-1}{2^{n+1}}\right[}_{\in\mathcal{D}_n} \cup \underbrace{\left[\frac{2i-1}{2^{n+1}}, \frac{2i}{2^{n+1}}\right[}_{\in\mathcal{D}_{n+1}}}_{\in\sigma(\mathcal{D}_{n+1})}.$$

**Définition 3.1.3.** *Une suite de variable aléatoire $(H_n)_{n\geq 1}$ est dite* ***prévisible*** *pour la filtration $(\mathcal{F}_n)_{n\geq 0}$ si pour tout $n \geq 1$, $H_n$ est $\mathcal{F}_{n-1}$-mesurable.*

## 3.2 Temps d'arrêt

Dans cette section, on fixe une filtration $(\mathcal{F}_n)_{n\geq 0}$.

### 3.2.1 Définitions et exemples

**Définition 3.2.1.** *Une variable aléatoire $T$ à valeurs dans $\mathbb{N}\cup\{\infty\}$ est un $(\mathcal{F}_n)_{n\geq 0}$-temps d'arrêt si pour tout $n \geq 0$,*

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

**Exemples.**

— Soit $T \equiv n_0$. Alors, pour tout $n \geq 0$,

$$\{T \leq n\} = \begin{cases} \varnothing \in \mathcal{F}_n & \text{si } n < n_0, \\ \Omega \in \mathcal{F}_n & \text{si } n \geq n_0. \end{cases}$$

— *Temps d'attente.* Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires réelles, et $(\mathcal{F}_n)_{n\geq 1}$ la filtration canonique associée. Soit $A \in \mathcal{B}(\mathbb{R})$, on pose

$$T(\omega) = \min\{i \in \mathbb{N}^* \mid X_i(\omega) \in A\}.$$

Alors, on a

$$\{T \leq n\} = \bigcup_{i=1}^{n} \underbrace{\{X_i \in A\}}_{\in \mathcal{F}_i} \in \mathcal{F}_n.$$

— Dans le même contexte que l'exemple précédent, la variable aléatoire définie par $\widetilde{T}$

$$\widetilde{T}(\omega) = \max\{i \in \mathbb{N}^* \mid X_i(\omega) \in A\}$$

n'est pas un temps d'arrêt car

$$\{\widetilde{T} = n\} = \{\omega \in \Omega \mid X_n(\omega) \in A, \forall i > n, X_i(\omega) \notin A\} = \{X_n \in A\} \cap \bigcap_{i>n} \{X_i \notin A\}.$$

Il est clair que $\{X_n \in A\}$ est élément de $\mathcal{F}_n$, mais nous ne pouvons rien dire sur $\bigcap_{i>n}\{X_i \notin A\}$.

**Remarques.** Soit $T$ un temps d'arrêt.

— $\{T = n\}$ est élément de $\mathcal{F}_n$. En effet,

$$\{T = n\} = \underbrace{\{T \leq n\}}_{\in \mathcal{F}_n} \setminus \underbrace{\{T \leq n-1\}}_{\in \mathcal{F}_{n-1} \subset \mathcal{F}_n}.$$

— $\{T \geq n\}$ est élément de $\mathcal{F}_{n-1}$.

$$\{T \geq n\} = \{T \leq n-1\}^c \in \mathcal{F}_{n_1}.$$

— Soit $H_n = \mathbb{1}_{T\geq n}$. Alors, pour tout $n \in \mathbb{N}^*$, $H_n$ est $\mathcal{F}_{n-1}$-mesurable, donc la suite $(H_n)_{n\geq 1}$ est prévisible.

### 3.2.2 Propriétés élémentaires des temps d'arrêts

**Proposition 3.2.1.** *$T$ est un temps d'arrêt si et seulement si $\{T = n\} \in \mathcal{F}_n$ pour tout $n \geq 0$.*

*Démonstration.* On a déjà montré le sens direct. Pour le sens réciproque, on remarque simplement que

$$\{T \leq n\} = \bigcup_{k=0}^{n} \underbrace{\{T = k\}}_{\in \mathcal{F}_k \subset \mathcal{F}_n} \in \mathcal{F}_n.$$

□

**Proposition 3.2.2.** *Soient $T$ et $S$ des $(\mathcal{F}_n)_n$-temps d'arrêt. Alors, $\min(T, S)$, $\max(T, S)$ et $T + S$ sont des $(\mathcal{F}_n)_n$-temps d'arrêt.*

*Démonstration.* On fait les observations suivantes.

$$\{\min(T, S) \leq n\} = \underbrace{\{T \leq n\}}_{\in \mathcal{F}_n} \cup \underbrace{\{S \leq n\}}_{\in \mathcal{F}_n} \in \mathcal{F}_n,$$
$$\{\max(T, S) \leq n\} = \underbrace{\{T \leq n\}}_{\in \mathcal{F}_n} \cap \underbrace{\{S \leq n\}}_{\in \mathcal{F}_n} \in \mathcal{F}_n,$$

et

$$\{T + S = n\} = \bigcup_{i=0}^{n} \underbrace{\{T = i\}}_{\in \mathcal{F}_i \subset \mathcal{F}_n} \cup \underbrace{\{S = n - i\}}_{\in \mathcal{F}_{n-i} \subset \mathcal{F}_n} \in \mathcal{F}_n.$$

□

**Proposition 3.2.3.** *Soit $T$ un $(\mathcal{F}_n)_n$-temps d'arrêt, et $k \in \mathbb{N}$. Alors, $T \wedge k = \min(T, k)$ est un $(\mathcal{F}_n)_n$-temps d'arrêt.*

*Démonstration.* On a déjà montré qu'une variable aléatoire constante (à un entier) est un temps d'arrêt. On conclue avec la proposition précédente.

□

**Proposition 3.2.4.** *Soit $(T_p)_{p\geq 1}$ une suite monotone de $(\mathcal{F}_n)_n$-temps d'arrêt. Alors, $T = \lim_{p\to\infty} T_p$ est un $(\mathcal{F}_n)_n$-temps d'arrêt.*

*Démonstration.* On suppose d'abord que $(T_p)_{p\geq 1}$ est croissante. Alors,

$$\{T \leq n\} = \bigcap_{p\geq 1} \underbrace{\{T_p \leq n\}}_{\in \mathcal{F}_n} \in F_n.$$

Si $(T_p)_{p\geq 1}$ est décroissante, alors

$$\{T \leq n\} = \left\{ \left( \lim_{p\to\infty} T_p \right) \leq n \right\} = \bigcup_{p\geq 1} \underbrace{\{T_p \leq n\}}_{\in \mathcal{F}_n} \in F_n.$$

□

**Proposition 3.2.5.** *Soit $(T_p)_{p\geq 1}$ une suite quelconque de $(\mathcal{F}_n)_n$-temps d'arrêt. Alors,* $\sup_{p\geq 1} T_p$, $\inf_{p\geq 1} T_p$, $\limsup_{p\to\infty} T_p$, *et* $\liminf_{p\to\infty} T_p$, *sont des $(\mathcal{F}_n)_n$-temps d'arrêt.*

*Démonstration.* On fait les observations suivantes.

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

□

**Définition 3.2.2.** *À un temps d'arrêt $T$, on associe la tribue*

$$\mathcal{F}_T = \left\{ A \in \mathcal{F} \,\middle|\, \forall n \geq 0,\, A \cap \{T \leq n\} \in \mathcal{F}_n \right\}.$$

**Remarques.**

— $\mathcal{F}_T$ désigne l'information disponible jusqu'à la date aléatoire $T$.

— $\mathcal{F}_T$ est bien une tribu.

- $\Omega \in \mathcal{F}_T$ car $\Omega \cap \{T \leq n\} = \{T \leq n\}$ pour tout $n \geq 0$.
- Soit $(A_i)_{i\in I}$ une famille dénombrables d'éléments de $\mathcal{F}_T$. Alors, pour tout $n \geq 0$,

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

- Soit $A \in \mathcal{F}_T$. Alors, pour tout $n \geq 0$,

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

### 3.2.3 Propriétés des tribus $\mathcal{F}_T$

**Proposition 3.2.6.** *Si $T \equiv k$, alors $\mathcal{F}_T = \mathcal{F}_k$.*

*Démonstration.* Soit $A \subset \Omega$. Alors,

$$A \in \mathcal{F}_T \iff \forall n \geq 0,\, A \cap \{T \leq n\} \in \mathcal{F}_n \iff \forall n\ \geq k,\, A \in \mathcal{F}_n \iff A \in \mathcal{F}_k.$$

— Si $n < k$, $\{T \leq n\} = \varnothing$, et donc $A \cap \{T \leq n\} = \varnothing \in \mathcal{F}_n$.

— Si $n \geq k$, alors $\{T \leq n\} = \Omega$ , et donc $A \cap \{T \leq n\} = A$.

□

**Proposition 3.2.7.** *Soient $T$ et $S$ des temps d'arrêt avec $T \leq S$. Alors, $\mathcal{F}_T \subset \mathcal{F}_S$.*

*Démonstration.* Soit $A \in \mathcal{F}_T$. Pour tout $n \geq 0$, on a

$$A \cap \{S \leq n\} = \underbrace{A \cap \{T \leq n\}}_{\substack{\in \mathcal{F}_n \\ \text{car } A \in \mathcal{F}_T}} \cap \underbrace{\{S \leq n\}}_{\substack{\in \mathcal{F}_n \\ \text{car } S \text{ t.a}}},$$

d'où $A \in \mathcal{F}_S$.

□

**Proposition 3.2.8.** *Soit $T$ un temps d'arrêt. Alors, $T$ est $\mathcal{F}_T$-mesurable.*

*Démonstration.* On montre que pour tout $p \geq 1$,

$$\{T \leq p\} \in \mathcal{F}_T.$$

On a pour tout $n \geq 0$, $\{T \leq p\} \cap \{T \leq n\} \in \mathcal{F}_n$ car

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

□

**Proposition 3.2.9.** *Soient $T$ et $S$ des temps d'arrêt. Alors,*

$$\mathcal{F}_{T \wedge S} = \mathcal{F}_T \cap \mathcal{F}_S.$$

**Exercice.** Soient $T$ et $S$ des temps d'arrêt. Alors, les évènements $\{T \leq S\}$, $\{S \leq T\}$, et $\{S = T\}$ sont éléments de $\mathcal{F}_{T \wedge S}$.

## 3.3 Martingales, sous-martingales et sur-martingales

**Définition 3.3.1.** *Étant donné une filtration $(\mathcal{F}_n)_{n \geq 0}$, on appelle $(\mathcal{F}_n)_{n \geq 0}$-martingale toute suite de variables aléatoires $(X_n)_{n \geq 0}$ telle que*

*(i) pour tout $n \in \mathbb{N}$, $\mathbb{E}[|X_n|] < \infty$ ;*

*(ii) $(X_n)_{n \geq 0}$ est adaptée à $(\mathcal{F}_n)_{n \geq 0}$ ;*

*(iii) (propriété de martingale) pour tout $n \geq 0$, $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$.*

**Définition 3.3.2.** *On définit de la même façon une sous-martingale et une sur-martingale en remplaçant (iii) par*

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

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

**Exemples.**

— *(Martingale fermée).* Soit $X \in L^1(\mathcal{F})$ et $(\mathcal{F}_n)_{n\geq 0}$ une filtration. Alors la suite de variables aléatoires définie pour tout $n \in \mathbb{N}$ par

$$X_n = \mathbb{E}[X|\mathcal{F}_n]$$

est une martingale, appelée *martingale de* Doob.

— *(Marche aléatoire).* Soit $(X_n)_{n\geq 0}$ des variables aléatoires intégrables indépendantes centrées. Soit pour tout $n \in \mathbb{N}$,

$$S_n = \sum_{k=0}^{n} X_k \quad \text{et} \quad \mathcal{F}_n = \sigma(X_0, \ldots, X_n).$$

Alors pour tout $n \geq 0$, $S_n$ est intégrable comme somme finie de variables intégrables, et est clairement $\mathcal{F}_n$-mesurable. De plus,

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

Or, $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mathbb{E}[X_{n+1}] = 0$ par indépendance de $X_{n+1}$ et $\mathcal{F}_n$ (par définition), et par le fait que les variables aléatoires ont été choisies centrées.

— *(Modèle auto-regressif).* Soit $(\varepsilon_n)_{n\geq 0}$ une suite de variables aléatoires i.i.d intégrables centrées, et soit $a \in \mathbb{R}^*$ et $x \in \mathbb{R}$. On définit par récurrence $X_{n+1} = aX_n + \varepsilon_{n+1}$ et $X_0 = x \in \mathbb{R}$. Soit alors $Y_n = X_n/a^n$ pour tout $n \geq 0$. La suite $(Y_n)_n$ définie une martingale pour la filtration canonique de la suite $(\varepsilon_n)_n$. En effet, de par une récurrence immédiate, les $X_n$ sont toutes intégrales, et que la suite est $(\mathcal{F}_n)_n$-adaptée. Enfin, on a

$$\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = \frac{1}{a^{n+1}}\mathbb{E}[aX_n + \varepsilon_{n+1}|\mathcal{F}_n] = \frac{1}{a^{n+1}}(aX_n + \mathbb{E}[\varepsilon_{n+1}|\mathcal{F}_n]).$$

Or, $\mathbb{E}[\varepsilon_{n+1}|\mathcal{F}_n] = \mathbb{E}[\varepsilon_{n+1}] = 0$. D'où $\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = Y_n$.

— *(*Galton-Watson*).* Soit $(X_{i,j})_{i,j\geq 1}$ une famille de variables aléatoires i.i.d à valeurs entières de loi $\mu$, intégrables et de moyenne $m$. On pose $Z_0 = 1$, et par récurrence

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

avec la convention que $Z_{n+1} = 0$ lorsque $Z_n = 0$. On appelle $\mu$ la *loi de reproduction.* On montre que $(Z_n/m^n)_{n\geq 0}$ est une martingale par rapport à $(\mathcal{F}_n)_{n\geq 0}$ où pour tout $n \in \mathbb{N}^*$,

$$\mathcal{F}_n = \sigma(X_{i,j} \,:\, i \leq n,\, j \geq 1) \quad \text{et} \quad \mathcal{F}_0 = \{\varnothing, \Omega\}.$$

Par récurrence, la suite est adaptée. En effet, $Z_0 = 1$ est $\mathcal{F}_0$-mesurable. En supposant que $Z_n$ est $\mathcal{F}_n$-mesurable, alors pour tout $A \subset \mathbb{N}$,

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

Ainsi, $Z_{n+1}$ est $\mathcal{F}_{n+1}$-mesurable, d'où $\frac{Z_{n+1}}{m^{n+1}}$ est $\mathcal{F}_{n+1}$-mesurable. Montrons maintenant la propriété de martingale (les espérances sont bien définies car les variables sont positives, nous montrons l'intégrabilité plus tard). Pour tout $n \in \mathbb{N}$,

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

(La seconde égalité a été démontré en TD). Or, $\mathbb{E}[X_{n+1,j}|\mathcal{F}_n] = m$ car $X_{n+1,j}$ est indépendant de $\mathcal{F}_n$. Ainsi,

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

On en déduit que

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

Par récurrence,

$$\mathbb{E}\left[\frac{Z_n}{m^n}\right] = \mathbb{E}\left[\frac{Z_0}{m^0}\right] = x$$

donc $\frac{Z_n}{m^n}$ est intégrable.

## 3.4 Propriétés des martingales

**Proposition 3.4.1.** *Soit $(X_n)_{n\geq 0}$ une $(\mathcal{F}_n)_{n\geq 0}$ martingale. Alors, $(X_n)_n$ est une $(\mathcal{G}_n)_n$-martingale, où pour tout $n \in \mathbb{N}$,*

$$\mathcal{G}_n = \sigma(X_0, \dots, X_n).$$

*Démonstration.* Par définition, les $X_n$ sont intégrables. De plus, $(\mathcal{G}_n)_n$ étant la filtration canonique, il est évident que $X_n$ est $\mathcal{G}_n$-mesurable pour tout $n \geq 0$. Enfin, par définition de $(X_n)_n$, pour tout $n \in \mathbb{N}$, $X_0, \dots, X_n$ sont $\mathcal{F}_n$-mesurables. Ainsi,

$$\mathcal{G}_n = \sigma(X_0, \dots, X_n) \subset \mathcal{F}_n$$

car $\mathcal{G}_n$ est la plus petite tribu (pour l'inclusion) rendant $X_0, \dots, X_n$ mesurables. Ainsi,

$$\mathbb{E}[X_{n+1}|\mathcal{G}_n] = \mathbb{E}[\underbrace{\mathbb{E}[X_{n+1}|\mathcal{F}_n]}_{X_n}|\mathcal{G}_n] = \mathbb{E}[X_n|\mathcal{G}_n] = X_n.$$

□

**Proposition 3.4.2.** *La propriété de martingale est équivalente à la propriété "Pour tout $n > m$, $\mathbb{E}[X_n|\mathcal{F}_m] = X_m$".*

*Démonstration.* Le sens réciproque est clair. Montrons le sens direct par récurrence sur $n > m$. Pour $n = m + 1$, c'est la propriété de martingale. Soit $n > m$, et supposons la propriété vraie au rang $n := m + k$. Alors,

$$\mathbb{E}[X_{m+k+1}|\mathcal{F}_m] = \mathbb{E}[\mathbb{E}[X_{m+k+1}|\mathcal{F}_{m+k}]|\mathcal{F}_m] = \mathbb{E}[X_{m+k}|\mathcal{F}_m] = X_m.$$

□

**Remarque.** On peut démontrer des résultats similaires pour les sous-martingales et les sur-martingales.

**Proposition 3.4.3.** *Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires.*

- *Si $(X_n)_{n\geq 0}$ est une martingale. Alors, la suite $(\mathbb{E}[X_n])_{n\geq 0}$ est constante.*
- *Si $(X_n)_{n\geq 0}$ est une sous-martingale. Alors, la suite $(\mathbb{E}[X_n])_{n\geq 0}$ est croissante.*
- *Si $(X_n)_{n\geq 0}$ est une sur-martingale. Alors, la suite $(\mathbb{E}[X_n])_{n\geq 0}$ est décroissante.*

*Démonstration.* Soit $n \in \mathbb{N}$. On a $\mathbb{E}[X_{n+1}] = \mathbb{E}[\mathbb{E}[X_{n+1}|\mathcal{F}_n]] = \mathbb{E}[X_n]$.

□

**Proposition 3.4.4.** *On a les propriétés suivantes.*

*(i) Soit $(X_n)_n$ une martingale, et $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction convexe telle que pour tout $n \in \mathbb{N}$, $Y_n := \varphi(X_n)$ est intégrable. Alors, la suite de variables aléatoires $(Y_n)_n$ forme une sous-martingale.*

*(ii) Soit $(X_n)_n$ une sous-martingale, et $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ une fonction convexe **croissante** telle que pour tout $n \in \mathbb{N}$, $Y_n := \varphi(X_n)$ est intégrable. Alors, la suite de variables aléatoires $(Y_n)_n$ forme aussi une sous-martingale.*

*Démonstration.* Par hypothèse, $Y_n$ est intégrable et est $\mathcal{F}_n$-mesurable car $X_n$ l'est et $\varphi$ est mesurable. Pour obtenir les propriétés de martingales et de sous-martingales, on utilise l'inégalité de Jensen.

□

**Conséquence 1.** Soit $(X_n)_n$ une martingale telles que les $X_n$ sont toutes de classe $L^p$ avec $p$ un nombre pair non nul. Alors, $(|X_n|^p)_{n\geq 0}$ est une martingale. En effet, la fonction $x \longmapsto |x|^p$ est convexe.

**Conséquence 2.** Si $a \in \mathbb{R}$ et $(X_n)_n$ est une sous-martingale, alors $((X_n - a)^+)_n$ est une sous-martingale. Pour le montrer, on observe que la fonction $x \longmapsto (x-a)^+$ est convexe.

**Conséquence 3.** Si $a \in \mathbb{R}$ et $(X_n)_n$ est une sur-martingale, alors $(X_n \wedge a)_n$ est une sur-martingale. Cette fois-ci, on utilise le second point de la proposition précédente à la sous-martingale $(-X_n)_n$ et $\varphi : x \longmapsto \max(x, -a)$.

**Proposition 3.4.5.** *Soit $(X_n)_n$ une sous-martingale (resp. une sur-martingale), et $(H_n)_n$ une suite prévisible bornée positive. Alors la suite $H \cdot X$ définie par $(H \cdot X)_0 = 0$ et*

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

*pour tout tout $n \geq 1$, est une sous-martingale (resp. sur-martingale). De plus, si $H$ est de signe quelconque et $(X_n)_n$ est de martingale, alors $H \cdot X$ est de martingale.*

*Démonstration.* (dans le cas où $(X_n)_n$ est une sous-martingale). Pour tout $n \in \mathbb{N}$, $(H \cdot X)_n$ est intégrale car pour tout $k \in [\![1, n]\!]$, $H_k(X_k - X_{k-1})$ est intégrale car $H_k$ est bornée. De plus, $(H \cdot X)_n$ est $\mathcal{F}_n$-mesurable de par sa définition. Enfin,

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

Or, $(X_n)_n$ étant une sous-martingale, on sait que $\mathbb{E}[X_{n+1} - X_n|\mathcal{F}_n]$ est une quantité positive. Puisque $H_{n+1}$ est positive, on obtient l'inégalité recherchée.

Notons que la démonstration est la même dans le cas d'une sur-martingale en remarquant que la quantité $\mathbb{E}[X_{n+1} - X_n|\mathcal{F}_n]$ est négative. Dans le cas martingale et $h$ de signe quelconque, $\mathbb{E}[X_{n+1} - X_n|\mathcal{F}_n] = 0$.

□

**Remarque.** On peut voir cette formule comme une intégration par parties (comme la transformée d'Abel est une forme d'intégration par parties dans le cas discret).

## 3.5 Martingale arrêtée

**Proposition 3.5.1.** *Soit $(X_n)_{n\geq 0}$ une suite $(\mathcal{F}_n)_n$-adaptée, et $T$ un $(\mathcal{F}_n)_{n\geq 0}$-temps d'arrêt. La variable aléatoire $\mathbb{1}_{\{T<\infty\}}X_T$ est $\mathcal{F}_T$-mesurable.*

**Remarque.** Si $T < \infty$ presque sûrement, on note simplement $X_T$. On travaillera d'ailleurs quasi-exclusivement avec de tels temps d'arrêt.

*Démonstration.* Soit $B \in \mathcal{B}(\mathbb{R})$. Pour tout $n \geq 0$,

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

On en déduit que $\{\mathbb{1}_{\{T<\infty\}}X_T \in B\} \in \mathcal{F}_T$. Ainsi,

$$\{\mathbb{1}_{\{T<\infty\}}X_T \in B\} \cap \{T \leq n\} = \bigcup_{p=0}^{n} \underbrace{\{\mathbb{1}_{\{T<\infty\}}X_T \in B\} \cap \{T = p\}}_{\in \mathcal{F}_p \subset \mathcal{F}_n}.$$

Ainsi, $\mathbb{1}_{\{T<\infty\}}X_T$ est $\mathcal{F}_T$-mesurable.

□

**Notation.** Étant donnés un $(\mathcal{F}_n)_n$ temps d'arrêt $T$ et une suite $(X_n)_n$, on pose

$$X^T = (X_{T\wedge n})_{n\geq 0}.$$

**Proposition 3.5.2.** *Soit $T$ un $(\mathcal{F}_n)_n$-temps d'arrêt.*

*(i) Si $(X_n)_n$ est adaptée, alors la variable aléatoire $X^T = (X_{T\wedge n})_n$ l'est aussi.*

*(ii) Si $(H_n)_n$ est prévisible, alors la variable aléatoire $H^T = (H_{T\wedge n})_n$ l'est aussi.*

*Démonstration.* $(i)$ Soit $\in B \in \mathcal{B}(\mathbb{R})$. On a

$$\begin{aligned}\left\{X_n^T \in B\right\} &= \bigcup_{p\in\mathbb{N}} \left\{X_n^T \in B, T = p\right\} \cup \left\{X_n^T \in B, T = \infty\right\} \\ &= \bigcup_{p\leq n} \left\{X_n^T \in B, T = p\right\} \cup \bigcup_{p>n} \left\{X_n^T \in B, T \geq n+1\right\}.\end{aligned}$$

En cherchant à déterminer à quelle tribu appartient chacun de ces éléments, on montre ainsi que $X^T$ est adaptée.

(*ii*) Soit $\in B \in \mathcal{B}(\mathbb{R})$, $n \geq 0$. On a

$$\left\{H_{n+1}^T \in B\right\} = \bigcup_{p=0}^{n} \left\{H_{n+1}^T \in B, T = p\right\} \cup \left\{H_{n+1}^T \in B, T \geq n\right\}.$$

Comme au point précédent, cette égalité permet de vérifier que $H^T$ est prévisible.

□

**Définition 3.5.1.** *Si $(X_n)_{n\geq 0}$ est une martingale, et $T$ un temps d'arrêt, on appelle* ***martingale arrêtée*** *la martingale*

$$X^T = (X_{T\wedge n})_{n\geq 0}.$$

*De même, on introduit des sur et sous-martingales arrêtées.*

En effet, si $X$ est une martingale, $X^T$ en est une aussi.

**Remarque.** Soit $H_n = \mathbb{1}_{\{T\geq n\}}$ $(n \geq 0)$. Cela définie une suite prévisible. De plus,

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

Ainsi, on a $(H \cdot X)_n + X_0$. La nature de $X$ (suite de variables aléatoires) se transmet à $H \cdot X$, et donc à $X^T$.

**Proposition 3.5.3** (théorème de temps d'arrêt faible)**.** *Soit $(X_n)_n$ un sous-martingale, et $T$ un temps d'arrêt borné par un entier $k$, i.e que $0 \leq T \leq k$. Alors $X_T$ est intégrale, et*

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

*Si $(X_n)_n$ est une sur-martingale,*

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

*Enfin, si $(X_n)_n$ est une martingale,*

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

*Démonstration.* (*i*) On suppose que $X$ est une sous-martingale. On a

$$X_T = \sum_{i=0}^{k} X_T \mathbb{1}_{\{T=i\}} = \sum_{i=0}^{k} X_i \mathbb{1}_{\{T=i\}}.$$

Ainsi,

$$\begin{aligned} |X_T| &\leq \sum_{i=0}^{k} |X_i| \mathbb{1}_{\{T=i\}} \\ &\leq \sum_{i=0}^{k} |X_i| \in L^1(\Omega, \mathcal{F}, \mathbb{P}). \end{aligned}$$

On en déduit que $X_T$ est intégrale. On considère la sous-martingale $X^T$. On a

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

En effet, $X_0 = X_{0\wedge T}$ et $X_{T\wedge k} = X_T$. Si $i \in [\![0, k]\!]$, alors $\{T = i\} \in \mathcal{F}_i$. On a

$$\sum_{i=0}^{k} \mathbb{E}\left[X_T \mathbb{1}_{\{T=i\}}\right] \leq \sum_{i=0}^{k} \mathbb{E}\left[X_k \mathbb{1}_{\{T=i\}}\right].$$

En effet, les $i$ sont inférieurs à $k$ et $X$ est une sous-martingale (on a en fait utilisé la propriété de sous-martingale). On en déduit que

$$\mathbb{E}\left[X_T \sum_{i=0}^{k} \mathbb{1}_{\{T=i\}}\right] \leq \mathbb{E}\left[X_k \sum_{i=0}^{k} \mathbb{1}_{\{T=i\}}\right].$$

Or, $\sum_{i=0}^{k} \mathbb{1}_{\{T=i\}} = 1$, donc

(*ii*) Si $X$ est une sur-martingale, on applique (*i*) à $-X$, et on rétablit les inégalités en multipliant par $-1$.

(*iii*) Si $X$ est un martingale, on a les inégalités des points (*i*) et (*ii*), qui ensemble deviennent un égalité.

□

**Contre-exemple.** La propriété est fausse sans l'hypothèse que $T$ est bornée. Par exemple, si les $X_i$ sont indépendantes identiquement distribuées selon une loi de Rademacher de paramètre $1/2$, si $S_0 = 0$,

$$S_n = \sum_{i=1}^{n} X_i,$$

et $T = \inf\{n \in \mathbb{N} \mid S_n = -1\}$. Alors, $S_0 = 0$, $S_T = -1$, donc $\mathbb{E}[S_0] = 0$ et $\mathbb{E}[S_T] = -1$. En effet,

$$\{T \geq n\} \subset \{X_1 = \cdots = X_n = 1\}.$$

Ainsi,

$$\mathbb{P}(T \geq n) \geq \mathbb{P}(X_1 = \cdots = X_n = 1) = \left(\frac{1}{2}\right)^n.$$

**Théorème 3.5.1** (d'arrêt faible)**.** *Soit $(X_n)_{n\geq 0}$ une sous-martingale et $T$ un temps d'arrêt. Supposons que $(X_n)_n$ et $T$ vérifient* ***l'une*** *des trois conditions suivantes.*

*(i) $T$ est bornée.*

*(ii) $X$ est bornée, i.e qu'il existe un $K > 0$ tel que pour tout $n \geq 0$, $|X_n| \leq K$, et $T$ est finie presque sûrement.*

*(iii) $\mathbb{E}[T] < \infty$, et il existe $K > 0$ tel que pour tout $n \in \mathbb{N}$, $|X_{n+1} - X_n| \leq K$ presque sûrement.*

*Alors $X_T$ est intégrale, et $\mathbb{E}[X_T] \geq \mathbb{E}[X_0]$. De plus, si $X$ est une martingale, alors sous (i), (ii), ou (iii), $X_T$ est intégrable, et $\mathbb{E}[X_T] = \mathbb{E}[X_0]$. Enfin, si $X$ est une sur-martingale, alors sous (i), (ii), (iii), ou sous la condition*

*(iv) $X_n$ est positive pour tout $n \in \mathbb{N}$ et $T$ est finie presque sûrement.*

*alors $X_T$ est intégrable et $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$.*

*Démonstration.* Soit $X$ une sous-martingale. On détaille selon l'hypothèse choisie.

($i$) Correspond au au résultat précédent.

($ii$) On considère $T \wedge n$ qui est un temps d'arrêt borné par $n$ auquel on applique ($i$). Alors,

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_{T\wedge n}]$$

et on passe à la limite lorsque $n$ tend vers l'infini. En effet, $X_{T\wedge n} \xrightarrow[n\to\infty]{} X_T$ presque sûrement car $T$ est finie presque sûrement. Ainsi,

$$|X_T| \leq \sum_{i=0}^{\infty} |X_T| \mathbb{1}_{\{T=i\}} \leq \sum_{i=0}^{\infty} |X_i| \mathbb{1}_{\{T=i\}} \leq K \left( \sum_{i=0}^{\infty} \mathbb{1}_{\{T=i\}} \right) = K.$$

Ainsi, $X_T$ est intégrale. De la même façon, $|X_{T\wedge n}| \leq K$. Par convergence dominée, on en déduit que

$$\mathbb{E}[X_{T\wedge n}] \xrightarrow[n\to\infty]{} \mathbb{E}[X_T].$$

En passant à la limite dans l'inégalité $\mathbb{E}[X_0] \leq \mathbb{E}[X_{T\wedge n}]$, on en déduit que $\mathbb{E}[X_T] \geq \mathbb{E}[X_0]$.

($iii$) On applique encore une fois le point ($i$) à $T \wedge n$, qui est un temps d'arrêt borné. Alors,

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

Or, $X_{T\wedge n} \xrightarrow[n\to\infty]{} X_T$ car $T$ intégrable donc finie presque sûrement. On a

$$X_{T\wedge n} - X_0 = \sum_{k=1}^{T\wedge n} (X_k - X_{k-1})$$

et donc

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

Bilan, $X_{T\wedge n} - X_0 \xrightarrow[n\to\infty]{} X_T - X_0$ presque sûrement, et $X_{T\wedge n} - X_0$ est dominée indépendant de $n$. Par convergence dominée,

$$\mathbb{E}[X_{T\wedge n} - X_0] \xrightarrow[n\to\infty]{} \mathbb{E}[X_T - X_0].$$

Or, $\mathbb{E}[X_{T\wedge n} - X_0]$ est une quantité positive car $\mathbb{E}[X_{T\wedge n}] \geq E[X_0]$. On en déduit que $\mathbb{E}[X_T - X_0] \geq 0$. De plus,

$$\mathbb{E}[|X_T - X_0|] = \lim_{n\to\infty} \mathbb{E}[|X_{T\wedge n} - X_0|]$$

mais $\mathbb{E}[|X_{T\wedge n} - X_0|] \leq \mathbb{E}[KT] < \infty$. Ainsi, $X_T - X_0$ est intégrable donc $X_T = X_0 + (X_T - X_0)$ est intégrable, et $\mathbb{E}[X_T] \geq \mathbb{E}[X_0]$ car $\mathbb{E}[X_T - X_0] \geq 0$.

On suppose maintenant que $X$ est une sur-martingale, et que la condition ($iv$) est vrai. Puisque $X$ est une sur-martingale, la suite des espérances des $X_n$ est décroissante, mais puisque $X_n \geq 0$ pour tout $n \geq 0$, on sait que $\mathbb{E}[X_T] \leq \mathbb{E}[X_0]$. On applique donc ($i$) à $T \wedge n$, et

$$\mathbb{E}[X_{T\wedge n}] \leq \mathbb{E}[X_0]$$

donc $X_{T\wedge n} \xrightarrow[n\to\infty]{} X_T$ car $T$ est finie. Ainsi,

$$\begin{aligned}\mathbb{E}[X_T] &= \mathbb{E}\left[\lim_{n\to\infty} X_{T\wedge n}\right] \\ &\leq \lim_{n\to\infty} \mathbb{E}[X_{T\wedge n}] \quad \text{(lemme de Fatou)} \\ &\leq \mathbb{E}[X_0].\end{aligned}$$

De plus, $X_T \geq 0$, et $\mathbb{E}[X_0] < \infty$, donc $X_T$ est intégrable.

□

## 3.6 Décomposition de Doob

**Théorème 3.6.1** (décomposition de Doob)**.** *Soit $(X_n)_n$ une sous-martingale. Alors, il existe $(M_n)_n$ une martingale, et $(A_n)_n$ une suite prévisible croissante avec $A_0 = 0$,* ***presque sûrement uniques*** *telles que pour tout $n \in \mathbb{N}$,*

$$X_n = M_n + A_n.$$

*Démonstration.* Si la décomposition a lieu, on doit avoir

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

Ainsi, $A_n - A_{n-1} = \mathbb{E}[X_n|\mathcal{F}_{n-1}] - X_{n-1}$. On en déduit que pour tout $n \geq 1$,

$$A_n = \sum_{k=0}^{n-1} \left(\mathbb{E}[X_{k+1}|\mathcal{F}_k] - X_k\right).$$

Or, $M_n = X_n - A_n$. Ainsi les suites $(M_n)_n$ et $(A_n)_n$ sont uniquement déterminées par $(X_n)_n$. Il suffit donc de s'assurer ces suites ainsi définies sont telles que $(A_n)_n$ est prévisible et $(M_n)_n$ est un martingale. Soit $n \in \mathbb{N}^*$, alors

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

car $X$ est une sous-martingale. De plus, on observe grâce à l'égalité ci-dessus que$A_n - A_{n-1}$ est $\mathcal{F}_{n-1}$-mesurable. De plus,

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

On en déduit que $(M_n)_n$ est une martingale.

□

**Remarque.** Par convexité de $x \longmapsto x^2$, si $X$ est un martingale alors $X^2 = (X_n^2)_n$ est une sous-martingale qui se décompose selon Doob sous la forme

$$X_n^2 = M_n + A_n$$

avec $(M_n)_n$ une martingale et $(A_n)_n$ une suite prévisible croissante.

**Définition 3.6.1.** *On appelle* ***compensateur*** *de la martingale $X$ la suite prévisible croissante de la décomposition de* Doob *de la sous-martingale $X^2$. On le note $\langle X, X\rangle = (\langle X, X\rangle_n)_{n\geq 0}$.*

**Remarque.** D'après le théorème de Doob, pour tout $n \geq 0$,

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

En effet, le calcul montre que $\mathbb{E}[(X_k - X_{k-1})^2 | \mathcal{F}_{k-1}] = \mathbb{E}[X_k^2 | \mathcal{F}_{k-1}] - X_{k-1}^2$.

**Exemple.** Supposons que les $X_i$ sont indépendantes identiquement distribuées, centrée, $L^2$ et de variance $\sigma^2$, et $\mathcal{F}_k = \sigma(X_1, \ldots, X_k)$. Soit alors

$$S_n = \sum_{i=1}^{n} X_i.$$

La suite des $S_n$ est une martingale, et

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

**Proposition 3.6.1.** *Soit $X$ une martingale $L^2$ (i.e que les $X_n$ sont $L^2$). Alors, la suite $X$ est bornée dans $L^2$ si et seulement si $\mathbb{E}[\langle X, X \rangle_\infty] < \infty$.*

**Proposition 3.6.2.** *Pour une martingale $X$ avec $X_n \in L^2(\mathcal{F})$, et $T$ un temps d'arrêt, on a*

$$\langle X, X \rangle^T = \left\langle X^T, X^T \right\rangle \ \text{p.s.}$$

*Démonstration.* La décomposition de Doob de la sous-martingale $X^2$

$$X^2 = X + \langle X, X \rangle .$$

En arrêtant,

$$(X^2)^T = (M + \langle X, X \rangle)^T.$$

Or,

$$(X^T)^2 = M^T + \langle X, X \rangle^T .$$

$(X^T)^2$ est une sous-martingale, $M^T$ est une martingale en tant que martingale arrêtée, et $\langle X, X \rangle^T$ est une suite prévisible croissante. Par unicité de la décomposition de Doob, on en déduit le résultat.

□

# Chapitre 4

# Convergences de martingales

On fixe $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ un espace de probabilité *filtré*.

## 4.1 Inégalités de martingales

### 4.1.1 Inégalité maximale de Doob

Soit $(X_n)_n$ une sous-martingale. On note

$$\overline{X}_n = \max_{0\leq k\leq n} X_k.$$

**Proposition 4.1.1** (Inégalité maximale de Doob)**.** *Pour toute sous-martingale* $(X_n)_{n\geq 0}$, $n \in \mathbb{N}$, *et* $x > 0$,

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

*Démonstration.* La seconde inégalité est immédiate. Pour la première, on pose $A = \{\overline{X}_n \geq x\}$, et $S = \inf\{k \in \mathbb{N} \mid X_k \geq x\}$. Remarquons que $S$ est un temps d'arrêt car c'est un temps d'attente. Soit $T = S \wedge n$ (temps d'arrêt borné). Selon la première version du théorème d'arrêt,

$$\mathbb{E}[X_T] \leq \mathbb{E}[X_n],$$

ce qui signifie que

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

Or, sur $A^c$ on a $\overline{X}_n < x$, donc $S < n$ et $T = n$. Ainsi, $\mathbb{E}[X_T \mathbb{1}_{A^c}] = \mathbb{E}[X_n \mathbb{1}_{A^c}]$. On a donc

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

Sur $A$, on $\overline{X}_n \geq x$, $T = S$, et $X_T = X_S \geq x$. On a donc

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

□

**Remarque.** Lorsque $(X_n)_n$ est une martingale, $(|X_n|)_n$ est une sous-martingale. De ce fait, on déduit le corollaire suivant.

**Corollaire 4.1.1.** *Soit $(X_n)_n$ une martingale. Alors, pour $n \in \mathbb{N}$ et $x > 0$,*

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

**Proposition 4.1.2** (inégalité maximale de KOLMOGOROV)**.** *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires centrées de variance finie. On note $S_n = \sum_{k=1}^n X_k$, et on suppose que $(\mathcal{F}_n)_n$ est la filtration associée à $(X_n)_{n\geq 1}$. Alors, pour tout $n \in \mathbb{N}^*$ et $x > 0$,*

$$\mathbb{P}\left(\max_{k \leq n} |S_k| \geq x\right) \leq \frac{\mathrm{var}[S_n]}{x^2}.$$

*Démonstration.* $(S_n)_n$ est une martingale, donc $(S_n^2)_N$ est une sous-martingale pour laquelle l'inégalité maximale de DOOB conclut que

$$\mathbb{E}[(S_n^2)^+] = \mathbb{E}[S_n^2] = \mathrm{var}(S_n).$$

□

**Remarque.** L'inégalité de BIENAYMÉ-TCHEBYCHEV affirme que

$$\mathbb{P}(|S_k| \geq x) \leq \frac{\mathrm{var}(S_k)}{x^2} \leq \mathrm{var}(S_n)$$

dès lors que $k \leq n$ et $x > 0$. Ainsi,

$$\max_{k \leq n} \mathbb{P}(|S_k| > x) \leq \frac{\mathrm{var}(S_n)}{x^2}$$

et donc que

$$\max_{k \leq n} \mathbb{P}(|S_k| \geq x) \leq \mathbb{P}\left(\bigcup_{k=1}^n \{|S_k| \geq x\}\right) = \mathbb{P}\left(\max_{k \leq n} |S_k| \geq x\right).$$

### 4.1.2 Inégalité maximale $L^p$ de DOOB

**Proposition 4.1.3.** *Soit $(X_n)_n$ une sous-martingale avec $X_0 \geq 0$ et soit $p > 1$. Alors,*

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

**Remarque.** Puisque $X_0 \geq 0$, par définition de $\overline{X}$, on a $\overline{X}_n \geq 0$ donc $\overline{X}_n$ est une variable aléatoire positive pour tout $n \in \mathbb{N}$, et $\mathbb{E}[\overline{X}_n^p]$ est bien définie.

*Démonstration.* Soit $M > 0$. On a

$$\begin{aligned}
\mathbb{E}[(\overline{X}_n \wedge M)^p] &= \mathbb{E}\left[\int_0^{\overline{X}_n \wedge M} px^{p-1}\,\mathrm{d}x\right] \\
&= \mathbb{E}\left[\int_0^{M} \mathbb{1}_{\{\overline{X}_n \geq x\}} px^{p-1}\,\mathrm{d}x\right] \\
&= \int_0^{M} \mathbb{E}[\mathbb{1}_{\{\overline{X}_n \geq x\}}] px^{p-1}\,\mathrm{d}x.
\end{aligned}$$

Or, par l'inégalité maximale de Doob,

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

Ainsi,

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

Par l'inégalité de Hölder avec l'indice $\alpha = \frac{p}{p-1}$ (on note $\beta$ son conjugué), on a

$$\mathbb{E}[(\overline{X}_n \wedge M)^p] \leq \frac{p}{p-1} \mathbb{E}[(X_n^+)^\beta]^{\frac{1}{\beta}} \mathbb{E}[(M \wedge \overline{X}_n)^{(p-1)^\alpha}]^{\frac{1}{\alpha}}.$$

Or, le conjugué de $\alpha$ est $p$. On a donc

$$\mathbb{E}[\overline{X}_n \wedge M)^p] \leq \frac{p}{p-1} \mathbb{E}[(X_n^+)^p]^{\frac{1}{p}} \mathbb{E}[(M \wedge \overline{X}_n)^p]^{1-\frac{1}{p}}.$$

En simplifiant, on a.

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

On en conclut en faisant tendre $M$ vers l'infini par convergence monotone puisque

$$\lim_{n \to \infty} \mathbb{E}[(M \wedge \overline{X}_n)^p] = \mathbb{E}\left[\left(\lim_{M \to \infty} M \wedge \overline{X}_n\right)^p\right] = \mathbb{E}[\overline{X}_n^p].$$

□

**Corollaire 4.1.2.** *On a les propriétés suivantes.*

*(i) Pour une martingale $(X_n)_{n\geq 0}$ $L^p$ et pour tout $n \in \mathbb{N}$,*

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

*(ii) Pour une martingale $(X_n)_{n\geq 0}$ avec $X_0 = 0$, si l'on décompose selon* Doob *par $X_n^2 = M_n + \langle X, X\rangle_n$, alors pour tout $n \in \mathbb{N}$,*

$$\mathbb{E}\left[\max_{k\leq n} |X_k|^2\right] \leq 4\mathbb{E}\left[\langle X, X\rangle_n\right],$$

*et*

$$\mathbb{E}\left[\sup_{n\geq 0} |X_k|^2\right] \leq 4\mathbb{E}\left[\langle X, X\rangle_\infty\right].$$

On a $\mathbb{E}[X_n^2] = \mathbb{E}[\langle X, X\rangle_n]$. De plus,

*Démonstration.*

$$\begin{aligned}
\mathbb{E}\left[\sup_{n\geq 0} |X_k|^2\right] &= \mathbb{E}\left[\lim_{n\to\infty} \max_{k\leq n} |X_k|^2\right] \\
&= \lim_{n\to\infty} \mathbb{E}\left[\max_{k\leq n} |X_k|^2\right] \\
&\leq \lim_{n\to\infty} 4\mathbb{E}[\langle X, X\rangle_n] \\
&= 4\mathbb{E}\left[\lim_{n\to\infty} \langle X, X\rangle_n\right].
\end{aligned}$$

(On remarquera qu'on a utilisé le théorème de convergence monotone pour faire sortir puis rentrer la limite en $n$). □

### 4.1.3 Inégalité sur le nombres de montées

Soit $(X_n)_{n\geq 0}$ une sous-martingale, et $a < b$ des réels. On pose $N_0 = 0$,

$$N_1 = \min\{n \geq 1 \,|\, X_n \leq a\} \quad \text{et} \quad N_2 = \min\{n \geq N_1 \,|\, X_n \geq b\}.$$

Par récurrence, on pose

$$N_{2k-1} = \min\{n \geq N_{2k-2} \,|\, X_n \leq a\} \quad \text{et} \quad N_{2k} = \min\{n \geq N_{2k-1} \,|\, X_n \geq b\}.$$

**Lemme 4.1.1.** *On a les propriétés suivantes sur la suite $(N_k)_{k\geq 0}$.*

*(i) La suite $(N_k)_{k\geq 0}$ est strictement croissante.*

*(ii) Pour tout $k \in \mathbb{N}$, $N_k \geq k$.*

*(iii) Pour tout $k \in \mathbb{N}$, la variable aléatoire $N_k$ est un temps d'arrêt.*

*Démonstration.* Les deux premiers points sont immédiats. Montrons $(iii)$. On a

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

Ainsi, $N_1$ est un temps d'arrêt. On a

$$\{N_2 = n\} = \bigcup_{j=1}^{n-1} (\{N_1 = j\} \cap \ \{N_2 = n\}) = \bigcup_{j=1}^{n-1} \left( \{N_1 = j\} \cap \left( \bigcup_{k=j+1}^{n-1} \{X_k < b\} \cap \{X_n \geq b\} \right) \right).$$

De cet égalité, on en déduit que $N_2$ est un temps d'arrêt. Supposons maintenant que $N_p$ est un temps d'arrêt pour tout $p \leq 2k$. Montrons que $N_{2k+1}$ et $N_{2k+2}$ en sont aussi. On a

$$\{N_{2k+1} = n\} = \bigcup_{j=1}^{n-1} \{N_{2k} = j\} \cap \{N_{2k+1} = n\} = \bigcup_{j=1}^{n-1} \underbrace{\{N_{2k} = j\}}_{\in \mathcal{F}_j \subset \mathcal{F}_{n-1}} \cap \bigcap_{l=j+1}^{n-1} \underbrace{\{X_l > a\}}_{\in \mathcal{F}_l \subset \mathcal{F}_n} \cap \underbrace{\{X_n \leq a\}}_{\in \mathcal{F}_n}.$$

On en déduit que $\{N_{2k+1} = n\} \in \mathcal{F}_n$ et donc que $N_{2k+1}$ est un temps d'arrêt. ? ? ?

□

Comme $X_{N_{2k-1}} \leq a$ et $X_{N_{2k}} \geq b$, entre les dates $N_{2k-1}$ et $N_{2k}$, la suite $(X_n)_{n \geq 0}$ monte exactement une fois d'au dessous de $a$ vers au dessus de $b$. En notant $U_n([a, b])$ le nombre de montées d'au dessous de $a$ à au dessus de $b$ pour les $X_k$ avec $1 \leq k \leq n$, on a

$$U_n([a, b]) = \max\{k \in \mathbb{N} \,|\, N_{2k} \leq n\}.$$

**Théorème 4.1.1.** *Pour une sous-martingale $X$, on a pour tout $a < b$ des réels,*

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

**Remarque.** On observe que $U_n([a, b]) \leq \lfloor \frac{n}{2} \rfloor$, et que l'inégalité

$$N_{2k} \leq n < n_{2k+2}$$

signifie qu'il y a exactement $k$ montées. Cela signifie que

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

On pose pour tout $j \geq 1$,

$$Y_j = \begin{cases} 1 & \text{s'il existe } k \in \mathbb{N} \text{ tel que } N_{2k} < j \leq N_{2k+1}, \\ 0 & \text{sinon, } i.e \text{ qu'il existe } k \in \mathbb{N} \text{ tel que } N_{2k+1} < j \leq N_{2k+2}. \end{cases}$$

Ainsi, $Y_j$ vaut 1 *si et seulement si* $X_j$ est dans une phase de descente, et $Y_j$ vaut 0 *si et seulement si* $X_j$ est dans une phase de montée. On a donc

$$Y_1 = \mathbb{1}_{\{N_1 > 1\}} = 1 \quad \text{et} \quad Y_2 = \mathbb{1}_{\{N_1 \geq 2\}} = \mathbb{1}_{\{X_1 > a\}}.$$

De manière générale, on pour tout $j \geq 2$,

$$Y_j = \mathbb{1}_{\{Y_{j-1}=0, X_{j-1} \geq b\} \cup \{Y_{j-1}=1, X_{j-1} > a\}}.$$

L'évènement $\{Y_{j-1} = 0, X_{j-1} \geq b\}$ s'interprète comme "*à la date $j-1$, on monte et la montée s'achève*", et l'évènement $\{Y_{j-1} = 1, X_{j-1} > a\}$ s'interprète comme "*à la date $j - 1$, on descend et la descente n'est pas achevée*".

**Lemme 4.1.2.** *Soit $X_1, \dots, X_n$ des variables aléatoires, et $(Y_j)_{j\geq 1}$. Alors,*

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

*Démonstration.* Tout $k \geq 1$, $Y_k$ est $\mathcal{F}_{k-1}$-mesurable. On a donc

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

□

car tous les termes dans l'espérance de la dernière ligne sont positifs. Le lemme précédent nous assure donc que

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

donc

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

*Démonstration (lemme 4.1.2).* On distingue deux cas.

($i$) Si $U_n([a,b]) = 0$, $i.e$ qu'il n'y a pas de montée à la date $n$ (et donc $N_2 > n$), alors on distingue alors trois cas.

— Si $N_1 = 1$, alors, $Y_k = 0$ pour tout $k \in [\![2, n]\!]$. Alors,

$$\sum_{k=2}^{n} Y_k(X_k - X_{k-1}) = 0 \leq 0 + (X_n - a)^+.$$

— Si $1 < N_1 \leq n$, $Y_k = 1$ pour $k \in [\![1, N_1]\!]$ et $Y_k = 0$ si $k \in [\![N_1 + 1, n]\!]$. Alors,

$$\sum_{k=2}^{n} Y_k(X_k - X_{k-1}) = \sum_{k=2}^{N_1} Y_k(X_k - X_{k-1}) = X_{N_1} - X_1.$$

Or, $X_{N_1} \leq a$, et $X_1 > a$ car $N_1 > 1$. On a donc

$$\sum_{k=2}^{n} Y_k(X_k - X_{k-1}) \leq \underbrace{U_n([a,b])}_{0} + (X_n - a)^+.$$

— Si $N_1 > n$, alors $Y_k = 1$ pour tout $k \in [\![1, n]\!]$ donc

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

car $X_1 > A$ (puisque $N_1 > 1$).

$(ii)$ Si $U_n([a,b]) > 0$, $i.e$ qu'il y a au moins un montée, on distingue encore deux cas. Si $N_1 > 1$, alors $Y_k = 1$ pour tout $k \in [\![1, N_1]\!]$, $Y_k = 0$ pour tout $k \in [\![N_1+1, N_2]\!]$, et par définition de $N_1$, on a $X_{N_1} \leq a$ et $X_1 > a$ puisque $N_1 > 1$. Alors,

$$\sum_{k=2}^{n} Y_k(X_k - X_{k-1}) = \underbrace{\sum_{k=2}^{N_1}(X_k - X_{k-1})}_{X_{N_1}-X_1<0} + \sum_{k=N_2+1}^{n} Y_k(X_k - X_{k-1}).$$

On en déduit que

$$\sum_{k=2}^{n} Y_k(X_k - X_{k-1}) \leq \sum_{k=N_2+1}^{n} Y_k(X_k - X_{k-1}).$$

De même si $N_1 = 1$, on a toujours cette inégalité précise. On distingue[1] quelques cas.

— Si la date $n$ n'est pas dans une phase de montée, alors

$$\begin{aligned}\sum_{k=N_2+1}^{n} Y_k(X_k - X_{k-1}) &= \underbrace{(X_{N_3} - X_{N_2})}_{\leq a-b} + 0 \\ &+ \underbrace{(X_{N_5} - X_{N_4})}_{\leq a-b} + 0 \\ &+ \cdots \\ &+ \underbrace{(X_{N_{2t+1}} - X_{N_{2t}})}_{\leq a-b} + \underbrace{\sum_{k=N_{2k+1}+1}^{n} Y_k(X_k - X_{k-1})}_{0}.\end{aligned}$$

où $t = \max\{k \in \mathbb{N} \,|\, N_{2k} \leq n\} = U_n([a,b])$. AInsi,

$$\sum_{k=N_2+1}^{n} Y_k(X_k - X_{k-1}) \leq (a-b)U_n([a,b]).$$

— Si la date $n$ est dans une phase de descente, alors

$$\begin{aligned}\sum_{k=N_2+1}^{n} Y_k(X_k - X_{k-1}) &= \underbrace{(X_{N_3} - X_{N_2})}_{\leq a-b} + 0 \\ &+ \underbrace{(X_{N_5} - X_{N_4})}_{\leq a-b} + 0 \\ &+ \cdots \\ &+ \underbrace{(X_{N_{2t'+1}} - X_{N_{2t'}})}_{\leq a-b} + \underbrace{\sum_{k=N_{2k+1}+1}^{n} Y_k(X_k - X_{k-1})}_{X_n - X_{N_{2t'}}}.\end{aligned}$$

De plus, $X_n - X_{N_{2t'}} = a - X_{N_{2t'}} + X_n - a \leq a - b + (X_n - a)^+$. On a donc

1. Encore...

$$\sum_{k=N_2+1}^{n} Y_k(X_k - X_{k-1}) \leq t'(a-b) + (X_n - a)^+ \leq (a-b)U_n([a,b]) + (X_n - a)^+.$$

Grâce à l'inégalité montrée un peu plus haut dans la démonstration, on achève la démonstration.

□

## 4.2 Convergence presque sûre de martingale

**Théorème 4.2.1** (condition nécessaire presque sûre de sous-martingale)**.** *Soit* $(X_n)_n$ *une sous-martingale telle que*

$$\sup_{n\in\mathbb{N}} \mathbb{E}[X_n^+] < \infty.$$

*Alors,* $(X_n)_n$ *converge presque sûrement vers une variable aléatoire* $X$ *intégrable.*

*Démonstration.* Soient $a < b$ des réels. Puisque $U_n([a,b])$ est croissante en $n$, on note sa limite

$$U_\infty([a,b]) = \lim_{n\to\infty} U_n([a,b]) = \sup_{n\in\mathbb{N}} U_n([a,b])$$

(on peut remplacer la limite par un supremum puisque la suite est croissante). Par le théorème de convergence monotone et selon l'inégalité sur le nombre de montée, on a

$$\mathbb{E}[U_\infty[(a,b)]] = \sup_{n\in\mathbb{N}} \underbrace{\mathbb{E}[U_n([a,b])]}_{\leq \frac{1}{b-a}\mathbb{E}[(X_n-a)^+]} \leq \frac{1}{b-a} \sup_{n\geq 0} \mathbb{E}[(X_n - a)^+].$$

De plus, $(X_n - a)^+ \leq X_n^+ + |a|$. On a donc

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

On en déduit que $\mathbb{E}[U_\infty([a,b])] < \infty$ donc que $U_\infty([a,b]) < \infty$ presque sûrement. Pour tout $a < b$ des réels, on a

$$\left\{\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n\right\} \subset \{U_\infty([a,b]) = \infty\}.$$

En effet, dans le premier évènement, $\{\liminf_{n\to\infty} X_n < a\}$ signifie que les $X_n$ passent une infinité de fois sous $a$, et $\{b < \limsup_{n\to\infty} X_n\}$ signifie que les $X_n$ passent une infinité de fois au dessus de $b$. On en déduit que l'évènement

$$\left\{\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n\right\}$$

est négligeable. Ainsi,

$$\bigcup_{\substack{a,b\in\mathbb{Q}\\ a<b}} \left\{\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n\right\} = \left\{\liminf_{n\to\infty} X_n < \limsup_{n\to\infty} X_n\right\}$$

est aussi un évènement négligeable. On en déduit que

$$\liminf_{n\to\infty} X_n = \limsup_{n\to\infty} X_n$$

presque sûrement et donc que $(X_n)_n$ converge presque sûrement. Notons $X$ sa limite. On a

$$\begin{aligned}
\mathbb{E}[X_n^+] &= \mathbb{E}\left[\lim_{n\to\infty} X_n^+\right] \\
&\leq \liminf_{n\to\infty} \mathbb{E}[X_n^+] \\
&\leq \sup_{n\to\infty} \mathbb{E}[X_n^+] < \infty.
\end{aligned}$$

Or, $\mathbb{E}[X_n^-] = \mathbb{E}[X_n^+] - \mathbb{E}[X_n]$, et $\mathbb{E}[X_n] \geq \mathbb{E}[X_1]$ car $(X_n)_n$ est une sous-martingale. On a donc $\mathbb{E}[X_n^-] \leq \mathbb{E}[X_n^+] - \mathbb{E}[X_1]$. Ainsi,

$$\begin{aligned}
\mathbb{E}[X_n^-] &= \mathbb{E}\left[\lim_{n\to\infty} X_n^-\right] \\
&\leq \liminf_{n\to\infty} \mathbb{E}[X_n^-] \\
&\leq \liminf_{n\to\infty} \mathbb{E}[X_n^+] - \mathbb{E}[X_1] \\
&\leq \sup_{n\in\mathbb{N}} \mathbb{E}[X_n^+] - \mathbb{E}[X_1] < \infty.
\end{aligned}$$

On a en déduit que $X$ est intégrable car $\mathbb{E}[|X|] = \mathbb{E}[X^+] + \mathbb{E}[X^-]$.

□

**Corollaire 4.2.1.** *Soit $(X_n)_n$ une martingale (ou une sous-martingale, ou une sur-martingale) telle que*

$$\sup_{n\geq 0} \mathbb{E}[|X_n|] < \infty$$

*alors $(X_n)_n$ converge presque sûrement vers une variable aléatoire $X$ intégrable.*

*Démonstration.* Pour tout $n \geq 0$, $X_n^+ \leq |X_n|$. L'hypothèse assure celle du théorème précédent lorsque $X$ est une sous-martingale. Le corollaire est donc vrai pour une martingale (qui est aussi une sous-martingale). Pour une sur-martingale, on considère $(-X_n)_n$ qui est une sous-martingale vérifiant les hypothèses du théorème précédent, et alors $(-X_n)_n$ converge presque sûrement vers une variable aléatoire intégrable.

□

**Corollaire 4.2.2.** *Soit $(X_n)_n$ une* ***sur-martingale*** *positive, i.e que $X_n \geq 0$ pour tout $n \in \mathbb{N}$, alors $(X_n)_n$ converge presque sûrement vers une variable aléatoire $X$ telle que*

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

*Démonstration.* On applique le résultat de convergence à la sous-martingale $(Y_n = -X_n)_n$ en remarquant que $Y_n^+ = 0$ pour tout $n \in \mathbb{N}$. Il est donc évident que $\sup_{n\geq 0} \mathbb{E}[Y_n^+] < \infty$. Ainsi $(Y_n)_n$ et donc $(X_n)_n$ convergent presque sûrement. Soit

$$X = \lim_{n\to\infty} X_n.$$

Alors,

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}\left[\liminf_{n\to\infty} X_n\right] \\ &\leq \liminf_{n\to\infty} \mathbb{E}[X_n] \\ &\leq \mathbb{E}[X_0]\end{aligned}$$

(la dernière inégalité étant justifiée par le fait que $(X_n)_n$ est une sur-martingale).

□

**Attention.** Les martingales que l'on manipule converge presque sûrement vers des variables aléatoires, mais en aucun cas nous n'avons affirmé que la convergence pouvait être $L^1$. Par exemple, si $(X_n)_n$ est i.i.d de loi $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$. Alors, si $S_0 = 1$ et $S_{n+1} = S_n + X_{n+1}$ pour tout $n \geq 0$, et

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

La variable $T$ est un temps d'arrêt. En posant $Y_n = S_n^T = S_{n\wedge T}$ pour tout $n \geq 0$, alors $Y$ reste une martingale car c'est une martingale arrêtée positive. Ainsi, $Y$ converge presque sûrement vers une certaine variable $Y_\infty$. On a $Y_\infty = 0$, car $Y_n$ est à valeurs dans $\mathbb{N}$ et si $Y_n = k$, alors $Y_{n+1}$ vaut soit $k+1$, soit $k-1$. Pour converger, il faut que $Y$ atteigne 0. Sur $\{T = \infty\}$, il n'y a pas convergence. En effet,

$$|Y_{n+1} - Y_n| = |S_{n+1} - S_n| = 1,$$

donc la suite $(Y_n)_n$ n'est pas de Cauchy, donc pas convergente en $L^1$. Ainsi, $T < \infty$ presque sûrement, et il n'y a pas de convergence $L^1$ car si elle avait lieu, on aurait

$$1 = \mathbb{E}[Y_0] = \mathbb{E}[Y_n] \xrightarrow[n\to\infty]{} \mathbb{E}[Y_\infty] = 0.$$

## 4.3 Convergence $L^1$ de martingale

### 4.3.1 Uniforme intégrabilité

**Définition 4.3.1.** *Une suite de variables aléatoires intégrables $(X_i)_{i\in I}$ est dite **uniformément intégrable** si*

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

**Exemples.**

— Si $\operatorname{card} I = 1$ ou $\operatorname{card} I < \infty$, c'est toujours le cas.

— S'il existe $Z \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ une variable aléatoire telle que pour tout $i \in I$, $|X_i| \leq Z$, alors $(X_i)_{i\in I}$ est uniformément intégrable.

— S'il existe $\delta > 0$ tel que

$$\sup_{i\in I} \mathbb{E}[|X_i|^{1+\delta}] < \infty,$$

alors $(X_i)_{i\in I}$ est uniformément intégrable.

**Proposition 4.3.1** (critère d'uniforme intégrabilité)**.** *Une famille $(X_i)_{i\in I}$ est uniformément intégrable si et seulement si*

*(i) pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que pour tout $A \in \mathcal{F}$,*

$$(\mathbb{P}(A) < \eta) \implies (\forall i \in I,\, \mathbb{E}[|X_i|\mathbb{1}_A] < \varepsilon),$$

*(ii) et*

$$\sup_{i\in I} \mathbb{E}[|X_i|] < \infty.$$

**Théorème 4.3.1** (Vitali)**.** *Soit $(X_n)_n$ une suite variables aléatoires intégrables. Alors, les assertions suivantes sont équivalentes.*

*(i) $(X_n)_n$ converge en $L^1$.*

*(ii) $(X_n)_n$ converge en probabilité et $(X_n)_n$ est uniformément intégrable.*

**Exemple.** Soit $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, alors la famille

$$(\mathbb{E}[X|\mathcal{G}])_{\substack{\mathcal{G}\subset\mathcal{F} \\ \text{sous-tribu}}}$$

forme une famille uniformément intégrable.

### 4.3.2 Martingale fermée

**Définition 4.3.2.** *Une $(\mathcal{F}_n)_n$-martingale $(X_n)_n$ est dite fermée s'il existe une variable aléatoire $L^1(\Omega, \mathcal{F}, \mathbb{P})$ telle que pour tout $n \in \mathbb{N}$,*

$$X_n = \mathbb{E}[X|\mathcal{F}_n].$$

*En d'autres termes, $(X_n)_n$ est une martingale de* Doob.

**Théorème 4.3.2.** *Soit $(X_n)_n$ une sous-martingale, (ou martingale, ou sur-martingale). Sont équivalents.*

*(i) $(X_n)_n$ est uniformément intégrable.*

*(ii) $(X_n)_n$ converge presque sûrement et en norme $L^1$.*

*En particulier si $(X_n)_n$ est une martingale, alors ces conditions sont équivalentes à*

*(iii) $(X_n)_n$ est fermée.*

*Démonstration.* — *(i)* $\iff$ *(ii)*. ? ? ? ? ?

— *(iii)* $\implies$ *(i)*. La famille $(\mathbb{E}[X|\mathcal{F}_n] = X_n)_{n\geq 1}$ est uniformément intégrable quand $X$ est intégrable.

— *(ii)* $\implies$ *(iii)*. On suppose que $(X_n)_n$ est converge presque sûrement et dans $L^1$. Soit $X$ sa limite, alors pour tout $k \geq n$, puisque $\mathbb{E}[X_k|\mathcal{F}_n]$ (propriété de martingale), on sait que pour tout $A \in \mathcal{F}_n$, $\mathbb{E}[X_k\mathbb{1}_A] = \mathbb{E}[X_n\mathbb{1}_A]$. Or,

$$|\mathbb{E}[X_k \mathbb{1}_A] - \mathbb{E}[X \mathbb{1}_A]| \leq \mathbb{E}[|X_k - X|] \xrightarrow[k\to\infty]{} 0,$$

donc

$$\mathbb{E}[X_n \mathbb{1}_A] = \mathbb{E}[X_k \mathbb{1}_A] \xrightarrow[n\to\infty]{} \mathbb{E}[X \mathbb{1}_A].$$

Ainsi, $X_n = \mathbb{E}[X|\mathcal{F}_n]$. La martingale est donc bien fermée.

□

**Notation.** On note

$$\bigvee_{n\in\mathbb{N}} \mathcal{F}_n = \sigma\left(\bigcup_{n\in\mathbb{N}} \mathcal{F}_n\right).$$

On notera généralement cette tribu $\mathcal{F}_\infty$. Dans la suite, lorsque l'on affirmera que $\mathcal{F}_n$ croît en l'infini vers $\mathcal{F}_\infty$ si $(\mathcal{F}_n)_n$ est une filtration avec $\mathcal{F}_\infty = \bigvee_{n\in\mathbb{N}} \mathcal{F}_n$.

**Théorème 4.3.3.** *Soit $(\mathcal{F}_n)_n$ une filtration croissant vers $\mathcal{F}_\infty$ et $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Alors,*

$$\mathbb{E}[X|\mathcal{F}_n] \xrightarrow[n\to\infty]{p.s, L^1} \mathbb{E}[X|\mathcal{F}_\infty].$$

*Démonstration.* On pose $X_n = \mathbb{E}[X|\mathcal{F}_n]$. Alors, la suite $(X_n)_n$ est une martingale fermée donc est uniformément intégrable et converge presque sûrement et en $L^1$. Notons $Z$ sa limite. Alors, $Z$ est une variable aléatoire $\mathcal{F}_\infty$ mesurable car $Z$ est limite de $(X_n)_n$ et $X_n$ est $\mathcal{F}_\infty$-mesurable pour tout $n \geq 0$. Soit $A \in \mathcal{F}_n$, on a.

$$\mathbb{E}[X \mathbb{1}_A] = \mathbb{E}[X_n \mathbb{1}_A] \xrightarrow[n\to\infty]{} \mathbb{E}[Z \mathbb{1}_A]$$

puisque $X_n \xrightarrow[n\to\infty]{L^1} Z$. Ainsi, pour tout $A \in \bigcup_{n\geq 0} \mathcal{F}_n$, $\mathbb{E}[X \mathbb{1}_A] = \mathbb{E}[Z \mathbb{1}_A]$. Soit

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

Nous avons monté que $\bigcup_{n\in\mathbb{N}} \mathcal{F}_n \subset \mathcal{M}$, et $\bigcup_{n\in\mathbb{N}} \mathcal{F}_n$ est un $\pi$-système (il est stable par intersection car $(\mathcal{F}_n)_n$ est une filtration). De plus, $\mathcal{M}$ est une classe monotone. En effet,

— $\Omega \in \mathcal{M}$ car $\mathbb{E}[X] = \mathbb{E}[Z]$ ;

— $\mathcal{M}$ est stable par différence symétrique : pour tout $A, B \in \mathcal{M}$ avec $A \subset B$, on a

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

— enfin, si $(A_n)_n \in \mathcal{M}^{\mathbb{N}}$ est une suite croissante, alors

$$\begin{aligned}
\mathbb{E}\left[X\mathbb{1}_{\bigcup_{n\in\mathbb{N}} A_n}\right] &= \mathbb{E}\left[X\left(\lim_{n\to\infty}\mathbb{1}_{A_n}\right)\right]\\
&= \lim_{n\to\infty}\mathbb{E}[X\mathbb{1}_{A_n}]\\
&= \lim_{n\to\infty}\mathbb{E}[Z\mathbb{1}_{A_n}]\\
&= \mathbb{E}\left[Z\mathbb{1}_{\bigcup_{n\in\mathbb{N}} A_n}\right],
\end{aligned}$$

donc $\bigcup_{n\in\mathbb{N}} A_n$ est élément de $\mathcal{M}$.

Par le théorème de classe monotone, on a

$$\mathcal{F}_\infty = \bigvee_{n\in\mathbb{N}} \mathcal{F}_n = \sigma\left(\bigcup_{n\in\mathbb{N}} \mathcal{F}_n\right) = \mathcal{M},$$

donc pour tout $A \in \mathcal{F}_\infty$, $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[Z\mathbb{1}_A]$, ce qui signifie exactement que $Z = \mathbb{E}[X|\mathcal{F}_\infty]$.

□

### 4.3.3 Applications des convergences presque sûres et $L^1$

**Proposition 4.3.2** (Loi du 0-1 de Lévy)**.** *Soit $(\mathcal{F}_n)_n$ une filtration croissant vers $\mathcal{F}_\infty$. Alors, pour tout $A \in \mathcal{F}_\infty$,*

$$\mathbb{E}[\mathbb{1}_A|\mathcal{F}_n] \xrightarrow[n\to\infty]{p.s, L^1} \mathbb{1}_A.$$

*Démonstration.* C'est une conséquence de l'énoncé précédent avec $X = \mathbb{1}_A$ où $A \in \mathcal{F}_\infty$.

□

On a aussi la loi du 0-1 de Kolmogorov qui concerne la tribu asymptotique d'une suite de variables aléatoires indépendantes $(X_n)_n$. On appelle *tribus du futur* les tribus

$$\mathcal{F}^{(n)} = \sigma(X_n, X_{n+1}, \ldots).$$

De là, on définit la *tribu asymptotique* par

$$\mathcal{F}^{(\infty)} = \bigcap_{n\geq 0} \mathcal{F}^{(n)}.$$

**Proposition 4.3.3** (loi du 0-1 de Kolmogorov)**.** *Pour des variables aléatoires $(X_n)_n$ indépendantes, la tribu asymptotique est triviale. Ainsi, pour tout $A \in \mathcal{F}^{(\infty)}$, $\mathbb{P}(A) \in \{0,1\}$.*

*Démonstration.* Soit pour tout $n \geq 0$, $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$, $\mathcal{F}^{(n)} = \sigma(X_n, X_{n+1}, \ldots, )$ et

$$F_\infty = \bigvee_{n\in\mathbb{N}} \mathcal{F}_n \quad \text{et} \quad \mathcal{F}^{(\infty)} = \bigcap_{n\geq 0} \mathcal{F}^{(n)}.$$

Pour tout $n \geq 1$ et $A \in \mathcal{F}^{(\infty)}$. Alors, $(X_0, \ldots, X_n)$ est indépendant de $\mathbb{1}_A$ car $\mathcal{F}^{(\infty)} \subset \mathcal{F}^{(n+1)}$. Ainsi,

$$\mathbb{E}[\mathbb{1}_A] = \mathbb{E}[\mathbb{1}_A|\mathcal{F}_n] \xrightarrow[n\to\infty]{} \mathbb{E}[\mathbb{1}_A|\mathcal{F}_\infty].$$

On en déduit que

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

$\mathcal{F}^{(\infty)} \subset \mathcal{F}_\infty$. En effet,

$$\mathcal{F}^{(n)} = \sigma(X_n, X_{n+1}, \dots,) = \sigma\left(\bigcup_{p\geq n} \mathcal{F}_n^p\right)$$

où $\mathcal{F}_n^p = \sigma(X_n, \dots, X_p)$. On a $\mathcal{F}_n^p \subset \mathcal{F}_p \subset \mathcal{F}_\infty$ donc $\bigcup_{p\geq n} \subset \mathcal{F}_\infty$, ce qui implique que $\mathcal{F}^{(n)} \subset \mathcal{F}_\infty$. Ainsi, $\mathcal{F}^{(\infty)} \subset \mathcal{F}_\infty$. Bilan : on a $\mathbb{P}(A) \in \mathbb{1}_A(\Omega) \subset \{0,1\}$.

□

**Théorème 4.3.4** (convergence dominée pour l'espérance conditionnelle)**.** *Soit $(\mathcal{F}_n)_n$ une filtration croissant vers $\mathcal{F}_\infty$. Soit $(X_n)_n$ une suite de variables aléatoires convergeant presque sûrement vers une variable aléatoire $X$, et telle qu'il existe $Z$ une variable aléatoire intégrable aléatoire telle que $|X_n| \leq Z$ pour tout $n \in \mathbb{N}$. Alors,*

$$\mathbb{E}[X_n|\mathcal{F}_n] \xrightarrow[n\to\infty]{p.s} \mathbb{E}[X|\mathcal{F}_\infty].$$

*Démonstration.* Soit $N \in \mathbb{N}$ et $W_N = \sup_{n,m\geq N} |X_n - X_m|$. Alors, $W_N \leq 2Z$ et $2Z$ est intégrable. De plus, par le critère de Cauchy,

$$W_N \xrightarrow[N\to\infty]{p.s} 0.$$

Pour tout $n, m \geq N$, on a $|X_n - X_m| \leq W_N$. Ainsi, lorsque $m$ tend vers l'infini, on a $|X_n - X| \leq W_N$. Puisque $L^1$ est intégrable, on en déduit que

$$\mathbb{E}[|X_n - X||\mathcal{F}_n] \leq \mathbb{E}[W_N|\mathcal{F}_n] \xrightarrow[n\to\infty]{p.s} \mathbb{E}[W_N|\ \mathcal{F}_\infty].$$

On a donc

$$\limsup_{n\to\infty} \mathbb{E}[|X_n - X||\mathcal{F}_n] \leq \mathbb{E}[W_N|\mathcal{F}_\infty].$$

Or, $(W_N)_N$ tend vers 0 presque sûrement de manière décroissante, donc $(\mathbb{E}[W_N|\mathcal{F}_\infty])_{N\geq 1}$ décroit vers une limite $U \geq 0$. Puisque $W_N$ est intégrable, et $(W_N)_N$ tend vers 0, on a $(\mathbb{E}[W_N])_N$ tend de manière décroissante vers 0. De plus, pour tout $N \in \mathbb{N}$, $\mathbb{E}[W_N|\mathcal{F}_\infty] \leq 2\mathbb{E}[Z|\mathcal{F}_\infty] \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. Par convergence dominée,

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

Ainsi, $\mathbb{E}[U] = 0$ donc $U = 0$ par positivité. En faisant tendre $N$ vers l'infini dans l'inégalité

$$\limsup_{n\to\infty} \mathbb{E}[|X_n - X||\mathcal{F}_n] \leq \mathbb{E}[W_N|\mathcal{F}_\infty],$$

on a

$$0 \leq \liminf_{n\to\infty} \mathbb{E}[|X_n - X||\mathcal{F}_n] \leq \mathbb{E}[W_N|\mathcal{F}_\infty] \leq \limsup_{n\to\infty} \mathbb{E}[|X_n - X||\mathcal{F}_n] \leq \mathbb{E}[W_N|\mathcal{F}_\infty] = 0.$$

On en déduit que $\mathbb{E}[|X_n - X||\mathcal{F}_n] \xrightarrow[n\to\infty]{p.s} 0$, donc

$$|\mathbb{E}[X_n|\mathcal{F}_n] - \underbrace{E[X|\mathcal{F}_n]}_{\xrightarrow[n\to\infty]{}\mathbb{E}[X|\mathcal{F}_\infty]}| \leq \mathbb{E}[|X_n - X||\mathcal{F}_n] \xrightarrow[n\to\infty]{} 0.$$

Ainsi,

$$\mathbb{E}[X_n|\mathcal{F}_n] \xrightarrow[n\to\infty]{p.s} \mathbb{E}[X|\mathcal{F}_\infty].$$

□

**Théorème 4.3.5.** *Soit $(X_n)_n$ une sur-martingale (ou sous-martingale, ou martingale) telle qu'il existe $p > 1$ tel que*

$$\sup_{n\geq 1} \mathbb{E}[|X_n|^p] < \infty.$$

*Alors,*

$$X_n \xrightarrow[n\to\infty]{p.s, L^1} X.$$

*Dans le cas où $(X_n)_n$ est une martingale, alors la convergence est aussi $L^p$ vers $X$ avec $\mathbb{E}[X] = \mathbb{E}[X_n]$ pour tout $n \geq 0$.*

*Démonstration.* On a $\sup_{n\geq 1} \mathbb{E}[|X_n|^p] < \infty$ donc $(X_n)_n$ est uniformément intégrable puisque $p > 1$, on a donc la convergence $L^1$. De plus, l'inégalité de Hölder donne que $\sup_{n\geq 1} \mathbb{E}[|X_n|] < \infty$ donc la convergence est presque sûre. Dans le cas où $(X_n)_n$ est une martingale, l'inégalité maximales des moment de Doob pour $p > 1$ donne

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

Or, $(\sup_{k\leq n} |X_k|^p)_n$ tend de manière croissante vers $\sup_{k\geq 0} |X_k|^p$ lorsque $n$ tend vers l'infini. Ainsi, on

$$\mathbb{E}\left[\sup_{k\leq 0} |X_k|^p\right] \leq \left(\frac{p}{p-1}\right)^p \sup_{n\geq 1} \mathbb{E}[|X_n|^p] < \infty.$$

Ainsi, $Z$ est intégrable. Pour tout $k \in \mathbb{N}$, on a donc

$$|X_k - X|^p \leq 2^p Z^p.$$

Or $2^p Z^p$ est intégrable, et $(X_k)_k$ converge presque sûrement donc $(|X_k - X|)_k$ converge presque sûrement vers 0. Par convergence dominée,

$$\mathbb{E}[|X_k - X|^p] \xrightarrow[k\to\infty]{} 0$$

donc $(X_k)_k$ converge vers $X$ en $L^p$. Puisque $(X_n)_n$ converge en $L^1$, on a $\mathbb{E}[X] = \mathbb{E}[X_n]$ car $(X_n)_n$ est une martingale.

□

## 4.4 Convergence $L^2$ de martingales

On rappelle qu'une martingale $(X_n)_n$ de carré intégrable admet un compensateur $(\langle X, X\rangle_n)_n$ (c'est une suite prévisible croissante nulle pour $n = 0$) tel que pour tout $n \in \mathbb{N}$,

$$X_n^2 = M_n + \langle X, X\rangle_n$$

où $(M_n)_n$ est une martingale. On a vu aussi que $(X_n)_n$ est bornée dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$ *si et seulement si* $\mathbb{E}[\langle X, X\rangle_\infty] < \infty$.

**Théorème 4.4.1.** *Soit $(X_n)_n$ une martingale $L^2$. On a les propriétés suivantes.*

*(i) Pour tout $\omega \in \Omega$ tel que $\langle X, X\rangle_\infty (\omega) < \infty$, alors*

$$\lim_{n\to\infty} X_n(\omega)$$

*existe.*

*(ii) Si $(X_n)_n$ a des accroissements bornés, i.e qu'il existe $K > 0$ tel que pour tout $n \geq 0$,*

$$|X_{n+1} - X_n| \leq K,$$

*alors pour presque chaque $\omega \in \Omega$ 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 $X_0 = 0$.

($i$) On note $(A_n)_n = \langle X, X\rangle$ (suite prévisible), et soit $S_k = \inf\{p \in \mathbb{N} \mid A_{p+1} > k\}$ pour $k \in \mathbb{N}$. Alors, $(S_k)_k$ est un temps d'arrêt. En effet, pour tout $k, n \in \mathbb{N}$,

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

De plus, la suite $A^{S_k} = (A_{S_k\wedge n})_n$ est aussi une suite prévisible. En effet, c'est le compensateur de $X^{S_k}$ car

$$\left\langle X^{S_k}, X^{S_k}\right\rangle = \langle X, X\rangle^{S_k} = A^{S_k}.$$

Selon la décomposition de Doob, $(X^{S_k})^2 - A^{S_k}$ est une martingale. Pour tout $n \in \mathbb{N}$, on a

$$\mathbb{E}[(X_n^{S_k})^2 - A_n^{S_k}] = \mathbb{E}[(X_0^{S_k})^2 - A_0^{S_k}] = 0,$$

donc

$$\mathbb{E}[(X_n^{S_k})^2] = \mathbb{E}[A_{S_k\wedge n}] \leq k$$

car $A_{S_k\wedge n} \leq A_{S_k} \leq k$ par définition de $S_k$. Ainsi, $(X_n^{S_k})_n$ est bornée $L^2$ donc ans $L^1$, d'où la convergence presque sûre de $X^{S_k}$. Puisque

$$\{A_\infty < \infty\} = \bigcup_{k\geq 0} \{S_k = \infty\},$$

pour tout $\omega \in \Omega$ tel que $\langle X, X\rangle_\infty (\omega) = A_\infty(\omega) < \infty$, il existe $k(\omega)$ tel que $S_{k(\omega)}(\omega) = \infty$. Pour un tel $\omega$, on a $S_{k(\omega)}(\omega) = \infty$ et donc $X^{S_{k(\omega)}} = X$ converge.

(*ii*) On montre que la convergence de $(X_n(\omega))_n$ dans $\mathbb{R}$ implique que $\langle X, X\rangle_\infty(\omega) < \infty$. Par l'absurde, on suppose que que

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

Soit

$$T_c = \inf\{n \in \mathbb{N} \,|\, |X_n| \geq c\}.$$

Alors,

$$\{T_c = \infty\} = \left\{\sup_{n\in\mathbb{N}}|X_n| \leq c\right\}.$$

On a

$$\mathbb{P}(A_\infty = \infty, T_c = \infty) = \mathbb{P}\left(A_\infty = \infty, \sup_{n\geq 0}|X_n| < c\right) > 0.$$

Or,

$$\lim_{c\to\infty}\mathbb{P}\left(A_\infty = \infty, \sup_{n\geq 0}|X_n| < c\right) = \mathbb{P}\left(A_\infty = \infty, \sup_{n\geq 0}|X_n| < c\right).$$

Ainsi, pour $c$ suffisamment grand, on a

$$\mathbb{P}\left(A_\infty = \infty,\, T_c = \infty\right) > 0.$$

Par le théorème d'arrêt avec $T_c \wedge n \leq n$ appliqué à $X^2 - A$, on a

$$\mathbb{E}[X^2_{T_c\wedge n} - A_{T_c\wedge n}] = \mathbb{E}[X_0^2 - A_0] = 0.$$

Or, on a $X_{T_c\wedge n} = X_{(T_c\wedge n)-1} + X_{T_c\wedge n} - X_{(T_c\wedge n)-1}$, $|X_{(T_c\wedge n)-1}| \leq c$ et $|X_{T_c\wedge n} - X_{(T_c\wedge n)-1}| \leq K$. Ainsi, $\mathbb{E}[A_{T_c\wedge n}] = \mathbb{E}[X^2_{T_c\wedge n}] \leq (K+c)^2$. Par convergence monotone, on trouve que $\mathbb{E}[A_{T_c}] \leq (K+c)^2$. Mais $\mathbb{E}[A_{T_c}] = \mathbb{E}[A_\infty \mathbb{1}_B] = \infty$.

□

**Conséquence.** Si $X$ est une martingale $L^2$ et $\langle X, X\rangle_\infty < \infty$ alors $(X_n)_n$ converge presque sûrement.

**Proposition 4.4.1.** *Soit $(X_n)_n$ une martingale $L^2$ avec $X_0 = 0$. Sur l'évènement, $\{\langle X, X\rangle_\infty = \infty\}$, on a*

$$\frac{X_n}{\langle X, X\rangle_n} \xrightarrow[n\to\infty]{p.s} 0.$$

**Remarque.** Soient $(Y_n)_n$ des variables i.i.d de carrés intégrables, et soit $S_n = \sum_{i=1}^n Y_i$. Alors,

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{p.s} \mathbb{E}[Y_1]$$

(c'est la loi des grands nombres *forte* $L^2$). En effet, en notant $Z_i = Y_i - \mathbb{E}[Y_i]$ et $S'_n = \sum_{i=1}^n Z_i$, alors $(S_n)_n$ est une martingale ($S'_0 = 0$) avec $\langle S', S'\rangle_n = n\,\mathrm{var}(Z)$. D'après la loi des grands nombres pour les martingales, on a

$$\langle S', S' \rangle_n \xrightarrow[n\to\infty]{} \infty \text{ p.s.}$$

Ainsi,

$$\frac{S'_n}{\langle S', S' \rangle_n} = \frac{S_n - n\mathbb{E}[Y_1]}{n\,\mathrm{var}(Y_1)}$$

donc

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{p.s} \mathbb{E}[Y_1].$$

*Démonstration (de la loi des grands nombres martingale).* Soit $H_n = (1 + \langle X, X \rangle_n)^{-1}$ pour tout $n \in \mathbb{N}$. La suite $(H_n)_n$ est prévisible. On s'intéresse à $W = H \cdot X$ avec $W_0 = 0$. On a pour tout $n \geq 1$

$$W_n = (H \cdot X)_n = \sum_{k=1}^{n} \frac{X_k - X_{k-1}}{1 + \langle X, X \rangle_k}.$$

$W$ est une martingale $L^2$ de compensateur tel que

$$\langle W, W \rangle_n - \langle W, W \rangle_{n-1} = \mathbb{E}[(W_n - W_{n-1})^2 | \mathcal{F}_{n-1}]$$

(par définition du compensateur). On a donc

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

En effet, $\frac{1}{(1+\langle X,X \rangle_n)^2}$ est $\mathcal{F}_{n-1}$-mesurable. On en déduit que

$$\langle W, W \rangle_n - \langle W, W \rangle_{n-1} = \frac{1}{(1 + \langle X, X \rangle_{n-1})^2} - \frac{1}{(1 + \langle X, X \rangle_n)^2}.$$

Ainsi, en sommant,

$$\langle W, W \rangle_n = \sum_{k=1}^{n} (\langle W, W \rangle_k - \langle W, W \rangle_{k-1}) \leq \frac{1}{1 + \langle X, X \rangle_0} - \frac{1}{1 + \langle X, X \rangle_n} \leq 1$$

(la dernière égalité est due au fait que $\langle X, X \rangle_n \geq 0$). On a donc $\langle W, W \rangle_\infty \leq 1$ donc la martingale $W$ converge presque sûrement. Ainsi, la suite

$$\left( \sum_{k=1}^{n} \frac{X_k - X_{k-1}}{1 + \langle X, X \rangle_k} \right)$$

converge presque sûrement. La conclusion va suivre du lemme suivant.

□

**Lemme 4.4.1** (Kronecker). *Soit $E$ un espace vectoriel normé, $(x_n)_{n\geq 1}$ une suite d'éléments de $E$, et $(a_n)_{n\geq 1}$ une suite de réels positifs croissante vers l'infini. On suppose que la suite*

$$\left(\sum_{k=1}^{n} \frac{x_k}{a_k}\right)_{n\geq 1}$$

*converge. Alors,*

$$\frac{1}{a_n}\sum_{k=1}^{n} x_k \xrightarrow[n\to\infty]{} 0.$$

*Démonstration.* Soit $S_k = \sum_{i=1}^{k} \frac{x_i}{a_i}$. On a

$$\sum_{k=1}^{n} x_k = \sum_{k=1}^{n} a_k(S_k - S_{k-1})$$

car $\frac{x_k}{a_k} = S_k - S_{k-1}$. On a donc

$$\sum_{k=1}^{n} x_k = \sum_{k=1}^{n-1} (a_k - a_{k+1})S_k + a_n S_n.$$

On a donc

$$\frac{1}{a_n}\sum_{k=1}^{n} a_k = \frac{1}{a_n}\sum_{k=1}^{n-1} (a_k - a_{k-1})S_k + S_n.$$

On conclut par Cesàro.

□

*Démonstration (fin de la démonstration précédente).* On applique le lemme à $x_k = X_k - X_{k-1}$ et $a_k = 1 + \langle X, X\rangle_k$. On en déduit que

$$\frac{1}{1+\langle X, X\rangle_n}\sum_{k=1}^{n} (X_k - X_{k_1}) = \langle X_n, 1 + \langle X, X\rangle_n\rangle \xrightarrow[n\to\infty]{p.s} 0.$$

Alors, $\langle X, X\rangle_n + 1 \sim_{n\to\infty} \langle X, X\rangle_n$.

□

**Théorème 4.4.2** (théorème central limite pour les martingales). *Soit $X$ une martingale $L^2$. On suppose qu'il existe $(a_n)_n$ une suite de réels positifs croissante vers l'infini telle que*

*(i) Il existe $l \in \mathbb{R}^*$ tel que*

$$\frac{\langle X, X\rangle_n}{a_n} \xrightarrow[n\to\infty]{\mathbb{P}} l\ ;$$

*(ii) (hypothèse de* Lindeberg*) pour tout $\varepsilon > 0$,*

$$\frac{1}{a_n}\sum_{k=1}^{n} \mathbb{E}[|\Delta X_k|^2 \mathbb{1}_{\{|\Delta X_k| > \varepsilon\sqrt{a_n}\}} | \mathcal{F}_{k-1}] \xrightarrow[n\to\infty]{\mathbb{P}} 0$$

*(où $\Delta X_k = X_k - X_{k-1}$).*

*Alors, on a*

$$\frac{X_n}{\langle X, X\rangle_n} \xrightarrow[n\to\infty]{\mathscr{L}} \mathcal{N}(0,1).$$

*Démonstration.* Résultat admis.

□

**Remarque.** Ce théorème central limite couvre le théorème central limite habituel pour les variables i.i.d centrées de carrés intégrables. En effet, si $(Y_n)_n$ est une famille de telles variables, alors $(S_n)_n = (\sum_{k=1}^n Y_i)_n$ (avec $S_0 = 0$) est une martingale $L^2$, et $\langle S, S\rangle_n = n\,\mathrm{var}(X_1)$. L'hypothèse $(i)$ est vérifiée avec $a_n = n$. De plus, on a

$$\mathbb{E}[Y_k^2 \mathbb{1}_{\{|Y_k| > \varepsilon\sqrt{n}\}} | \mathcal{F}_{k-1}] = \mathbb{E}[Y_k^2 \mathbb{1}_{\{|Y_k| > \varepsilon\sqrt{n}\}}]$$

et $Y_k^2 \mathbb{1}_{\{|Y_k| > \varepsilon\sqrt{n}\}}$ tend vers 0 lorsque $n$ tend vers l'infini par convergence dominée. Cela assure donc le point $(ii)$ et donc

$$\frac{S_n}{\sqrt{n\,\mathrm{var}(X_1)}} \xrightarrow[n\to\infty]{\mathscr{L}} \mathbb{N}(0,1).$$

## 4.5 Théorèmes d'arrêt

Il s'agît de voir que pour une sous-martingale, lorsqu'on a $T \leq S$ deux temps d'arrêts, alors $\mathbb{E}[X_T] \leq \mathbb{E}[X_S]$. Bien qu'elle semble évidente, elle n'est pas toujours vraie. En effet, si $(X_k)_k$ est une martingale de variables aléatoires à valeurs $\{-1, 1\}$, et si $M_n = \sum_{k=1}^n X_k$. Soit $T$ le temps d'arrêt constant à 0 et $S = \inf\{n \in \mathbb{N} \,|\, M_n = -1\}$. Alors, $S$ est un temps d'arrêt et $(M_n)_n$ est une martingale. Pourtant, $\mathbb{E}[M_S] = -1$ et $\mathbb{E}[M_T] = 0$.

**Proposition 4.5.1.** *Soit $(X_n)_n$ une sous-martingale uniformément intégrable et $T$ un temps d'arrêt. Alors, $X^T = (X_{T\wedge n})_n$ est une sous-martingale uniformément intégrable.*

*Démonstration.* Soit $\varphi : \mathbb{R} \to \mathbb{R}_+$, $x \mapsto x^+$ (fonction convexe croissante). Ainsi, $X_n^+ = \varphi(X_n)$ définit une nouvelle sous-martingale et $(X_{T\wedge n}^+)_n$ est encore une sous-martingale telle que

$$\mathbb{E}[X_{T\wedge n}^+] \leq \mathbb{E}[X_n^+] \leq \mathbb{E}[|X_n|].$$

Par le théorème d'arrêt avec le temps d'arrêt avec le temps d'arrêt borné $T \wedge n \leq n$, on a

$$\sup_{n\geq 0} \mathbb{E}[X_{T\wedge n}^+] \leq \sup_{n\geq 0} \mathbb{E}[|X_n|] < \infty$$

(car $(X_n)_n$ est uniformément intégrable). AInsi, $(X_{T\wedge n})_n$ converge presque sûrement vers $X_T$ et $X_T$ est $L^1$. Soit $c > 0$,

$$\mathbb{E}[|X_{T\wedge n}|\mathbb{1}_{\{X_{T\wedge n}>c\}}] = \underbrace{\mathbb{E}[|X_T|\mathbb{1}_{\{X_T>c\}}\mathbb{1}_{\{T\leq n\}}]}_{\leq \mathbb{E}[|X_T|\mathbb{1}_{\{X_T>c\}}]} + \underbrace{\mathbb{E}[|X_n|\mathbb{1}_{\{X_n>c\}}\mathbb{1}_{\{T>n\}}]}_{\leq \mathbb{E}[|X_n|\mathbb{1}_{\{X_n>c\}}]}.$$

Or, $\mathbb{E}[|X_T|\mathbb{1}_{\{X_T>c\}}]$ tend vers 0 lorsque $c$ tend vers l'infini par convergence dominée puisque $X_T$ est intégrable, et $\mathbb{E}[|X_n|\mathbb{1}_{\{X_n>c\}}]$ tend aussi vers 0 lorsque $c$ tend vers l'infini car $X$ est uniformément intégrable. Ainsi,

$$\sup_{n\geq 1} \mathbb{E}[|X_{T\wedge n}|\mathbb{1}_{\{X_{T\wedge n}>c\}}] \leq \sup_{n\geq 1} (\mathbb{E}[|X_T|\mathbb{1}_{\{X_T>c\}}] + \mathbb{E}[|X_n|\mathbb{1}_{\{X_n>c\}}]) \xrightarrow[c\to\infty]{} 0.$$

Ainsi, $(X_{T\wedge n})_n$ est uniformément intégrable.

□

**Proposition 4.5.2.** *Soit $(X_n)_n$ une sous-martingale uniformément intégrable. Alors, pour tout temps d'arrêt $T$,*

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

*Si $(X_n)_n$ est une sur-martingale uniformément intégrable, alors*

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

*Ce sont des égalités lorsque $X$ est une martingale uniformément intégrable.*

*Démonstration.* On montre le résultat dans le cas d'une sous-martingale. Soit $(X_n)_n$ une sous-martingale uniformément intégrable convergente presque sûrement et $L^1$ vers $X_\infty$ une variable aléatoire intégrable. Soit $n \in \mathbb{N}$. Alors, $0 \leq T \wedge n \leq n$ est un temps d'arrêt borné. On a par le théorème d'arrêt dans le cas borné,

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_{T\wedge n}] \leq \mathbb{E}[X_n].$$

Or, $\mathbb{E}[X_n] \xrightarrow[n\to\infty]{} \mathbb{E}[X_\infty]$ car $(X_n)_n$ converge en $L^1$. On a vu précédemment que $(X_{T\wedge n})_n$ est une sous-martingale uniformément intégrable donc elle converge presque sûrement et $L^1$ vers $X_T$. Ainsi, $\mathbb{E}[X_{T\wedge n}] \xrightarrow[n\to\infty]{} \mathbb{E}[X_T]$. On conclut en faisant tendre $n$ vers l'infini dans l'inégalité

$$\mathbb{E}[X_0] \leq \mathbb{E}[X_{T\wedge n}] \leq \mathbb{E}[X_n].$$

□

**Théorème 4.5.1** (théorème d'arrêt de Doob)**.** *Soit $(X_n)_n$ une sous-martingale uniformément intégrable, et $S$ et $T$ des temps d'arrêts tels que $S \leq T$. Alors, $\mathbb{E}[X_S] \leq \mathbb{E}[X_T]$ et $X_S \leq \mathbb{E}[X_T|\mathcal{F}_S]$ presque sûrement.*

**Remarques.**

— On a vu que la propriété de sous-martingale s'écrit aussi "$X_k \leq \mathbb{E}[X_n|\mathcal{F}_k]$" pour tout $k \leq n$. Le résultat $X_S \leq \mathbb{E}[X_T|\mathcal{F}_S]$ presque sûrement n'est donc pas surprenant.

— Pour une sur-martingale $(X_n)_n$ uniformément intégrable, on a dès lors que $S \leq T$ que

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

— Pour une martingale $(X_n)_n$ uniformément intégrable, on a dès lors que $S \leq T$ que

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

*Démonstration.* On montre le résultat pour une sous-martingale uniformément intégrable. Soit $Y_n = X_{T\wedge n}$. Alors, $(Y_n)_n$ est une sous-martingale uniformément intégrable à laquelle on applique le résultat précédent avec le temps d'arrêt $S$. Ainsi,

$$\mathbb{E}[Y_0] \leq \mathbb{E}[Y_S] \leq \mathbb{E}[Y_\infty].$$

On a $Y_S = X_{S\wedge T} = X_S$ et $Y_\infty = X_{\infty\wedge T} = X_T$. Alors, $\mathbb{E}[X_S] \leq \mathbb{E}[X_T]$. Montrons maintenant la seconde inégalité. Soit $A \in \mathcal{F}_S$ et $U = S\mathbb{1}_A + T\mathbb{1}_{A^c}$. $U$ est un temps d'arrêt car pour tout $n \geq 0$,

$$\{U \leq n\} = \underbrace{(\{S \leq n\} \cap \underbrace{A}_{\in\mathcal{F}_S})}_{\in\mathcal{F}_n} \cup \underbrace{(\{T \leq n\} \cap \underbrace{A^c}_{\in\mathcal{F}_S\subset\mathcal{F}_T})}_{\in\mathcal{F}_n} \in \mathcal{F}_n.$$

Puisque $U \leq T$, la première inégalité (déjà montrée) assure que $\mathbb{E}[X_U] \leq \mathbb{E}[X_T]$. On a donc que

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

En simplifiant, il vient que $\mathbb{E}[X_S\mathbb{1}_A] \leq \mathbb{E}[X_T\mathbb{1}_A] = \mathbb{E}[Z\mathbb{1}_A]$ où $Z = \mathbb{E}[X_T|\mathcal{F}_S]$. En choisissant $A = \{Z - X_S < 0\} \in \mathcal{F}_S$ (car $Z$ et $X_S$ sont $\mathcal{F}_S$-mesurables). Puisque $\mathbb{E}[X_S\mathbb{1}_A] \leq \mathbb{E}[X_T\mathbb{1}_A]$, on en déduit que

$$\mathbb{E}[(Z - X_S)\mathbb{1}_A] \geq 0$$

mais par le choix de $A$, on a $Z - X_S \leq 0$ donc $\mathbb{E}[(Z - X_S)\mathbb{1}_A] = 0$. Ainsi, $(Z - X_S)\mathbb{1}_A = 0$ presque sûrement car cette variable aléatoire est de signe constant et d'espérance nulle. Or sur $A$, $(Z - X_S)\mathbb{1}_A < 0$. Ainsi, $A$ est négligeable donc on en déduit que $X_S \leq Z$ presque sûrement. $\square$

# Chapitre 5

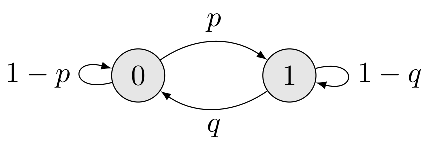
# Dynamique markovienne

On considère un système qui peut être dans un nombre au plus dénombrable d'états. L'ensemble de ces états sera noté $E$, appelé espace d'états. $E$ sera possiblement de la forme $[\![1, n]\!]$, $E = \mathbb{N}$, ou $E \subset \mathbb{N}$. On suppose le système observé à des dates $n \in \mathbb{N}$, et l'état du système à la date $n$ est décrit par $X_n$. Pour étudier le système aléatoire $(X_n)_{n \geq 0}$, il est nécessaire des faire des hypothèses sur les variables $X_n$. La propriété la plus simple est l'indépendance des variables aléatoires $X_n$, mais une telle hypothèse s'avère trop restrictive pour être intéressante. En fait, de nombreux systèmes évoluent au cours du temps de la façon suivante : le système à un date $n+1$ ne dépend que des valeurs précédentes $X_k$ pour $k \leq n$, qu'à travers $n$. Une telle façon d'évoluer dans le temps est appelée *propriété de* MARKOV et un tel système est dit *markovien.*

**Exemple.** Prenons l'exemple simple d'un système markovien (ou chaîne de MARKOV) prenant deux valeurs (généralement 0 et 1). On considère une machine qui soit fonctionne (état 1), soit est en panne. Chaque jour $(n)$, l'état de la machine est décrit par $X_n \in \{0, 1\}$. On suppose que

— si la machine est panne le jour $n$, elle est réparée et fonctionne le jour $n+1$ avec probabilité $p$ ;

— si la machine fonctionne le jour $n$, il y a une probabilité $q$ qu'elle tombe en panne et ne fonctionne pas le jour $n+1$.




**Objectif.** Décrire l'évolution de la machine, *i.e* déterminer pour tout $n \in \mathbb{N}$,

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

On suppose donné $\mu_0 = (\mu_0(0), \mu_0(1))$ qui décrit l'état initial de la machine. Remarquons qu'on a

$$\mathbb{P}(X_{n+1} = 0) = \mathbb{P}(X_{n+1} = 0 | X_n = 0)\mathbb{P}(X_n = 0) + \mathbb{P}(X_{n+1} = 0 | X_n = 1)\mathbb{P}(X_n = 1).$$

On en déduit que

$$\mathbb{P}(X_{n+1} = 0) = (1-p)\mathbb{P}(X_n = 0) + q\mathbb{P}(X_n = 1) = (1-p-q)\mathbb{P}(X_n = 0) + q.$$

On a donc une récurrence arithmético-géométrique puisque $p + q \neq 0$,

$$\mathbb{P}(X_n = 0) = (1-p-q)^n = \frac{q}{p+q} + (1-p-q)^n \left(\mathbb{P}(X_0 = 0) - \frac{q}{p+q}\right)$$

et

$$\mathbb{P}(X_n = 1) = (1-p-q)^n = \frac{p}{p+q} + (1-p-q)^n \left(\mathbb{P}(X_0 = 1) - \frac{p}{p+q}\right).$$

— Si $|1-(p+q)| < 1$, alors $\mathbb{P}(X_n = 0) \xrightarrow[n\to\infty]{} \frac{q}{p+q}$ et $\mathbb{P}(X_n = 1) \xrightarrow[n\to\infty]{} \frac{p}{p+q}$. On en déduit que $\mu_n$ converge en loi $\left(\frac{q}{p+q}, \frac{p}{p+q}\right)$ (on parle de loi asymptotique, ou d'équilibre).

— Si $p = q = 1$, il est facile de voir que $\mathbb{P}(X_{2n=0}) = \mu_0(0)$ et $\mathbb{P}(X_{2n+1} = 0) = \mu_1(0) = \mu_0(1)$. Alors, $\mu_{2n} = \mu_0$, et $\mu_{2n+1} = (\mu_0(1), \mu_0(0)) = 1 - \mu_0$. La loi de la machine alterne et il n'y a pas de convergence en loi du système.

— Si $p = q = 1$, alors $\mathbb{P}(X_n = 0) = \mu_0(0)$ et $\mathbb{P}(X_n = 1) = \mu_0(1)$. Ainsi, $\mu_n = \mu_0$. Le système n'évolue pas au cours du temps.

**Remarque.** Si $\mu_0(0) = \frac{q}{p+q}$ et $\mu_0(1) = \frac{p}{p+q}$, alors selon nos calculs dans le cas général, on a

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

Ainsi, $\mu_n = \mu_0$ pour tout $n \geq 0$. On dit que la loi asymptotique $\left(\frac{q}{p+q}, \frac{p}{p+q}\right)$ est une loi *invariante* (ou stationnaire).

**Approche matricielle.** Soit $\mu_n = (\mathbb{P}(X_n = 0), \mathbb{P}(X_n = 1))$ pour tout $n \in \mathbb{N}$. Soit la matrice

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}.$$

Le vecteur colonne $(1,1)^T$ est vecteur propre de $P$ associé à la valeur propre 1. Ainsi, on trouve que $\mathrm{Sp}(P) = \{1, 1-p-q\}$. Soit alors

$$D = \begin{pmatrix} 1-p & 0 \\ 0 & 1-p-q \end{pmatrix}.$$

1 est valeur propre dont un vecteur propre est $(1,1)^T$ et $1-p-q$ admet comme vecteur propre $(p,-q)^T$. Ainsi, un matrice de passage pour diagonaliser $P$ est

$$A = \begin{pmatrix} 1 & p \\ 1 & -q \end{pmatrix},$$

et

$$A^{-1} = \frac{1}{p+q} \begin{pmatrix} q & p \\ 1 & -1 \end{pmatrix}.$$

On en déduit que pour tout $n \in \mathbb{N}$,

$$P^n = A \begin{pmatrix} 1 & 0 \\ 0 & (1-p-q)^n \end{pmatrix} A^{-1} = \frac{1}{p+q} \begin{pmatrix} q & p \\ q & p \end{pmatrix} + \frac{(1-p-q)^n}{p+q} \begin{pmatrix} p & -q \\ -q & p \end{pmatrix}.$$

## 5.1 Probabilités de transition

On rappelle la définition de noyau de probabilité (ou noyau de transition).

**Définition 5.1.1.** *Soient $(E, \mathcal{E})$ et $(F, \mathcal{F})$ deux espaces mesurables. On appelle noyau de probabilité de $E$ dans $F$ toute application*

$$\nu : E \times \mathcal{F} \longrightarrow [0, 1]$$

*telle que*

*(i) pour tout $x \in E$, $\nu(x, \cdot\,)$ est un probabilité sur $(F, \mathcal{F})$ ;*

*(ii) pour tout $A \in \mathcal{F}$, $\nu(\,\cdot\,, A)$ est une fonction mesurable sur $(E, \mathcal{E})$.*

Dans la suite on prendre $E = F$ un ensemble dénombrable.

**Définition 5.1.2.** *On appelle matrice stochastique sur $E$ toute famille $(P(x, y))_{x,y\in E}$ des réels positives tels que*

*(i) pour tout $x, y \in E$, $0 \leq P(x, y) \leq 1$ ;*

*(ii) pour tout $x \in E$,*

$$\sum_{y\in E} P(x, y) = 1.$$

Bien évidemment, une matrice stochastique peut être vue comme une... matrice (infinie).

**Remarque.** Dans notre cas, la donnée d'une matrice stochastique est équivalente à la donnée d'une noyau de probabilité. Étant donné une matrice stochastique $P$, la formule

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

définit un noyau de probabilité $\nu$. Réciproquement, étant donné un noyau $\nu$, la formule

$$P(x, y) = \nu(x, \{y\})$$

définit une matrice stochastique. Dans les deux cas, on a toujours $P(x, \cdot\,) = \nu(x, \cdot\,)$ pour tout $x \in E$.

**Notations.**

— Si $f : E \longrightarrow \mathbb{R}$ est une fonction, on note $Pf$ la fonction

$$Pf : \left| \begin{array}{ccl} E & \longrightarrow & \mathbb{R} \\ x & \longmapsto & \displaystyle\sum_{y\in E} P(x, y) f(y). \end{array} \right.$$

On interprète les fonctions $f$ comme des vecteurs colonnes. Alors, on peut écrire

$$(Pf) = (P)(f).$$

On peut aussi écrire pour tout $x \in E$, $Pf(x) = \mathbb{E}_{P(x,\,\cdot\,)}[f]$.

— Soit $\mu$ une mesure sur $E$. On interprète $\mu$ comme un vecteur ligne et on définit $\mu P$ par

$$\mu P(y) = \sum_{x \in E} \mu(x) P(x,y)$$

pour tout $y \in E$. On a alors

$$(\mu P) = (\mu)(P).$$

— Étant donnés deux matrices stochastiques $P$ et $Q$, on définit la matrice $PQ$ pour tout $x, y \in E$ par

$$(PQ)(x,y) = \sum_{z \in E} P(x,z) Q(z,y).$$

$PQ$ est une matrice stochastique. Dans la suite, on notera $(P_n)_n$ la suite définie par $P_0(x,y) = \mathbb{1}_{x=y}$, $P_1 = P$, et $P_n = P^n$ pour tout $n \in \mathbb{N}^*$.

**Définition 5.1.3.** *On appelle chaîne de* MARKOV *sur un espace d'états $E$ dénombrable toute suite de variables aléatoires $(X_n)_n$ à valeurs dans $E$ telle que*

$$\mathscr{L}(X_{n+1}|X_0, \dots, X_n) = \mathscr{L}(X_{n+1}|X_n),$$

*i.e que pour tout $x_0, \dots, x_n \in E$ tel que*

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \neq 0,$$

*alors pour tout $y \in E$,*

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

*De plus, la chaîne de* MARKOV *est dite* ***homogène*** *s'il existe une matrice $P$ telle que*

$$\mathscr{L}(X_{n+1}|X_0, \dots, X_n) = P(X_n, \cdot),$$

*ou de manière équivalente si pour tout $x_0, \dots, x_n \in E$ tel que*

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \neq 0,$$

*alors pour tout $y \in E$,*

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

## 5.2 Exemples de chaînes MARKOV

**Exemple 1.** Soit $(X_n)_n$ une suite de variables aléatoires de loi $\mu$ à valeurs dans $E$. Alors, $(X_n)_n$ est une chaîne de MARKOV avec $P(x,y) = \mu(y)$. En effet,

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

et

$$\mathbb{P}(X_{n+1} = y | X_n = x_n) = \mu(y).$$

**Exemple 2.** *Marche aléatoire sur $\mathbb{Z}^d$.* Soit $(X_n)_n$ une suite de variables aléatoires i.i.d de loi $\mu$ sur $\mathbb{Z}^d$. On considère $(S_n = \sum_{i=0}^n X_i)_n$ appelée marche aléatoire de position initiale $X_0$ de pas de loi $\mu$. Ici, $(S_n)_n$ est une chaîne de Markov de matrice stochastique $P(x,y) = \mu(y-x)$. On a en fait

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

De même,

$$\mathbb{P}(S_{n+1} = y | S_n = x_n) = \mu(y - x_n).$$

**Cas particulier ($d=1$).** On a $E = \mathbb{Z}$. Soit $p = \mu(1)$, $q = \mu(-1)$, et $\mu(0) = r$. On a $p+q+r = 1$, et la marche $(S_n)_n$ est une chaîne de Markov sur $\mathbb{Z}$ avec

$$P(x,y) = \begin{cases} p & \text{si } y = x+1\,; \\ q & \text{si } y = x-1\,; \\ r & \text{si } y = x+1\,; \\ 0 & \text{si } y = x. \end{cases}$$

???

## 5.3 Probabilités trajectorielles

### 5.3.1 Propriétés

On considère une chaîne de Markov $(X_n)_n$ sur un espace d'état $E$ *dénombrable* et de matrice stochastique $P$ (ou matrice de transition, ou noyau de transition, ou encore noyau markovien).

**Proposition 5.3.1.** *Soit $(X_n)_n$ une chaîne de* Markov. *Alors $(X_n)_n$ est de matrice stochastique $P$ si et seulement si pour tout $n \in \mathbb{N}$ et pour tout $x_0, \ldots, x_n \in E$, on a*

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) = \mathbb{P}(X_0 = x_0) P(x_0, x_1) P(x_1, x_2) \cdots P(x_{n-1}, x_n).$$

Cette propriété donne un lien entre propriété de Markov, et probabilité initiale et dynamique de transition.

*Démonstration.* $\boxed{\Longrightarrow}$ On procède par récurrence avec une initialisation automatique lorsque $n = 0$. Soit $n \in \mathbb{N}$ et supposons la formule vraie à ce rrang. Alors si $\mathbb{P}(X_0 = x_0, X_n = x_n) \neq 0$, alors si $p_n = \mathbb{P}(X_0 = x_0, \ldots, X_n = x_n)$

$$\mathbb{P}(X_0 = x_0, \ldots, X_n = x_n, X_{n+1} = x_{n+1}) = \underbrace{\mathbb{P}(X_{n+1} = x_{n+1} | X_0 = x_0, \ldots, X_n = x_n)}_{P(x_n, x_{n+1})} p_n.$$

On conclut en appliquant l'hypothèse de récurrence. Si maintenant $\mathbb{P}(X_0 = x_1, X_n = x_n) = 0$, alors

$$\mathbb{P}(X_0 = x_0, \ldots, X_n = x_n, X_{n+1} = x_{n+1}) = 0.$$

Par hypothèse de récurrence, on a

$$\mathbb{P}(X_0 = x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n) = 0$$

donc

$$\mathbb{P}(X_0 = x_0)P(x_0, x_1) \cdots P(x_n, x_{n+1}) = 0.$$

$\boxed{\Longleftarrow}$ Soit $x_0, \ldots, x_n \in E$ tel que $\mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) \neq 0$. On a

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

Ainsi, $(X_n)_n$ est une chaîne de Markov de matrice stochastique $P$.

□

**Proposition 5.3.2.** *Soit $X$ une chaîne de* Markov *de matrice $P$. On a les propriétés suivantes.*

*(i) Pour toute fonction mesurable $f : E \longrightarrow \mathbb{R}$, $n \geq 0$, et $x_0, \ldots, x_n \in E$ tels que $\mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) \neq 0$, alors*

$$\mathbb{E}[f(X_{n+1}) | X_0 = x_0, \ldots, X_n = x_n] = \mathbb{E}[f(X_{n+1}) | X_n = x_n] = Pf(x_n).$$

*(ii) Pour tout $i_1 < \ldots < i_k \in [\![0, n]\!]$, et $f : E \longrightarrow \mathbb{R}$ mesurable,*

$$\mathbb{E}[f(X_{n+1}) | X_{i_1}, \ldots, X_{i_k}, X_n] = (Pf)(X_n).$$

*Démonstration.* $(i)$ Pour montrer ce point, il suffit d'utiliser la définition.

$(ii)$ On a par propriété de conditionnement en cascade que

$$\mathbb{E}[f(X_{n+1}) | X_{i_1}, \ldots, X_{i_k}, X_n] = \mathbb{E}\big[\mathbb{E}[f(X_{n+1}) | X_0, \ldots, X_n] \big| \, X_{i_1}, \ldots, X_{i_k}, X_n\big].$$

Or, $\mathbb{E}[f(X_{n+1}) | X_0, \ldots, X_n] = Pf(X_n)$ d'après $(i)$, et

$$\mathbb{E}[Pf(X_n) | X_{i_1}, \ldots, X_{i_k}, X_n] = (Pf)(X_n).$$

□

**Remarque.** Le noyau de transition en $n$ étapes donne la probabilité de passer d'un état $x$ à un état $y$ en $n$ étapes :

$$P_n(x, y) = \mathbb{P}(X_n = y | X_0 = x_0) = \frac{\mathbb{P}(X_0 = x, X_n = y)}{\mathbb{P}(X_0 = x)}.$$

Or, on a la partition

$$\{X_0 = x, X_n = y\} = \bigsqcup_{\substack{x_k \in E \\ 1 \leq j \leq n-1}} \{X_0 = x, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = y\}.$$

On a donc que

$$P_n(x,y) = \sum_{\substack{x_k \in E \\ 1 \le j \le n-1}} \frac{\mathbb{P}(X_0 = x, X_1 = x_1, \dots, X_{n-1} = x_{n-1}, X_n = y)}{\mathbb{P}(X_0 = x)}.$$

Ainsi,

$$P_n(x,y) = \sum_{\substack{x_k \in E \\ 1 \le j \le n-1}} \frac{\mathbb{P}(X_0 = x)P(x,x_1)\cdots, P(x_{n-1},y)}{\mathbb{P}(X_0 = x)}.$$

On trouve donc que

$$P_n(x,y) = \sum_{\substack{x_k \in E \\ 1 \le j \le n-1}} P(x,x_1)P(x_1,x_2)\cdots P(x_{n-1},y) = P^n(x,y).$$

Autrement dit, $P_n = P^n$.

**Proposition 5.3.3.** *Le noyau de transition en n étapes d'une chaîne X de matrice stochastique P est donnée par la matrice $P^n$. On a une propriété de semi-groupe (équation de* Chapman-Kolmogorov*) avec*

$$P_{n+m}(x,y) = \sum_{z \in E} P_n(x,z)P_m(z,y),$$

*ce qui revient à dire que $P^{n+m} = P^n P^m$. En notant $\mu_n = \mathscr{L}(X_n)$, i.e*

$$\mu_n = (\mathbb{P}(X_n = x))_{x \in E},$$

*alors*

$$\mathbb{P}(X_n = y) = \sum_{x \in E} \mathbb{P}(X_0 = x)P_n(x,y).$$

*Ainsi, $\mu_n = \mu_0 P^n$. De plus,*

$$\mathbb{P}(X_n = y) \sum_{x \in E} \mathbb{P}(X_{n+1} = x)P(x,y).$$

*Ainsi, $\mu_n = \mu_{n-1}P$.*

**Notation.** On note $\mathbb{P}_0$ pour indiquer que la loi initiale de la chaîne est $\mu_0 = \nu$, et $\mathbb{P}_x = \mathbb{P}_{\delta_x}$ pour indiquer que la chaîne part de $x \in E$.

**Proposition 5.3.4** (boîte à outils)**.** *Soit $X$ une chaîne de* MARKOV *de matrice stochastique $P$. Lorsque les probabilités conditionnelles sont bien définies, on a les expressions suivantes.*

*(i) Pour $x_1, \dots, x_n, y_1, \dots, y_m \in E$, on a*

$$\begin{aligned}\mathbb{P}(X_{n+1} = y_1, X_{n+2} = y_2, \dots, X_{n+m} = y_m | X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \\ = P(x_n, y_1)P(y_1, y_2) \cdots P(y_{m-1}, y_m).\end{aligned}$$

*(ii) Soit $x_n, y_1, \dots, y_m \in E$ et $A_0, \dots A_{n-1} \subset E$. On a*

$$\begin{aligned}\mathbb{P}(X_{n+1} = y_1, \dots, X_{n+m} = y_m | X_0 \in A_0, \dots, X_{n-1} \in A_{n-1}, X_n = x_n) \\ = \mathbb{P}(X_{n+1} = y_1, \dots, X_{n+m} = y_m | X_n = x_n).\end{aligned}$$

*(iii) Pour tout $A_0, \dots, A_{n_1}, B_1, \dots, B_m \subset E$ et $x_n \in E$. Alors,*

$$\begin{aligned}\mathbb{P}(X_{n+1} \in B_1, \dots, X_{n+m} \in B_m | X_0 \in A_0, \dots, X_{n-1} \in A_{n-1}, X_n = x_n) \\ = \sum_{\substack{y_k \in B_k \\ 1 \leq k \leq m}} \mathbb{P}(X_{n+1} = y_1, \dots, X_{n+m} = y_m | X_n = x_n).\end{aligned}$$

**Remarque.** Le point $(i)$ montre que

$$\mathbb{P}(X_{n+m} = y_m | X_0 = x_0, \dots, X_n = x_n) = \mathbb{P}(X_{n+m} = y_m | X_n = x_n) = P^m(x_n, y_n).$$

En effet,

$$\begin{aligned}\mathbb{P}(X_{n+m} = y_m | X_0 = x_0, \dots, X_n = x_n) \\ = \sum_{\substack{y_k \in E \\ 1 \leq j \leq n-1}} \mathbb{P}(X_{n+1} = y_1, \dots, X_{n+m} = y_m | X_0 = x_0, \dots, X_n = x_n).\end{aligned}$$

*Démonstration.* C'est pénible, peut-être un jour aurai-je le temps d'écrire cette démonstration. $\square$

### 5.3.2 Approche récursive d'une chaîne de Markov

**Proposition 5.3.5.** *Soit $X_0$ une variable aléatoire dans $E$ loi $\nu$ et $(U_n)_{n\geq 0}$ une suite de variables aléatoires i.i.d de loi $\mu$ dans $F$ indépendante de $X_0$. Pour toute fonction $f : E \times F \longrightarrow E$ mesurable, alors la suite suite $(X_n)_n$ définie récursivement pour tout $n \in \mathbb{N}$ par*

$$X_{n+1} = f(X_n, U_{n+1})$$

*est une chaîne de* Markov *de matrice stochastique $P$ telle pour tout $x, y \in E$,*

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

*Démonstration.* Soit $x_0, \ldots, x_n \in E$ tel que $\mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) \neq 0$. Alors,

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

(le passage de la première à la seconde ligne étant dû au fait que les $(U_n)_n$ sont indépendantes de $X_0$). On a donc bien une chaîne de Markov de matrice $P$.

□

**Exemple.** *Marche aléatoire.* Soit $(X_i)_{i\in\mathbb{N}}$ une suite variables aléatoires i.i.d, et $S_n = \sum_{i=0}^{n} X_i$ pour tout $n \in \mathbb{N}$, et soit la fonction $f : \mathbb{R}^2 \to \mathbb{R}$, $(x, y) \mapsto x + y$. Alors,

$$S_{n+1} = f(S_n, X_{n+1})$$

pour tout $n \in \mathbb{N}$. Puisque $(X_n)_n$ est i.i.d, et indépendante de $S_0 = X_0$, on retrouve avec cette écriture que $(S_n)_n$ est une chaîne de Markov.

**Proposition 5.3.6.** *Une chaîne de* Markov *homogène à valeurs dans $E \subset \mathbb{R}$ peut être vue en* **loi** *comme une suite récursive du type $X_{n+1} = f(X_n, U_{n+1})$ comme dans la propriété précédente.*

**Rappel.** Soit $Z$ une variable aléatoire de fonction de répartition $F$. On définit l'inverse généralisé par

$$F^{-1}(y) = \inf\{u \in \mathbb{R} \,|\, F(u) \geq y\}$$

pour tout $y \in F(\mathbb{R})$. Par exemple si $U$ suit une loi uniforme sur $]0, 1[$, alors $F^{-1}(U) \sim Z$ (c'est la méthode d'inversion).

*Démonstration.* Soit $(X_n)_n$ une chaîne de Markov de matrice stochastique $P$. Il s'agît de trouver une fonction $f$ mesurable et une variable aléatoire $U$ telles que $X_1 \stackrel{\mathscr{L}}{=} f(x, U_1)$ lorsque $X_0 = x$. On définit $f(x, \cdot)$ comme l'inverse généralisé de la fonction de répartition $P(x, \cdot)$ (qui est la loi de $X$ sachant $X_0 = x$). Dans notre cas, si $U_1$ suit une loi uniforme sur $]0, 1[$, alors $f(x, U_1) \sim U_1$ sachant que $X_0 = x$. On définit la suite $(X_n)_n$ par récurrence avec

$$\widetilde{X}_{n+1} = f(\widetilde{X}_n, U_{n+1})$$

pour tout $n \in \mathbb{N}$, où $f(x, \cdot)$ est l'inverse généralisé de la fonction de répartition $P(x, \cdot)$, et $(U_n)_n$ est une suite i.i.id de loi uniforme sur $]0, 1[$. Alors, $\widetilde{X}$ a pour matrice stochastique

$$\widetilde{P}(x, y) = \mathbb{P}(f(x, U_1) = y) = \mathbb{P}(X_1 = y | X_0 = x) = P(x, y).$$

□

## 5.4 Chaîne canonique

Cette section aura pour but, étant donné une matrice stochastique $P = (P(x, y))_{x,y \in E}$, on va construire sur un certain espace probabilisé une variable aléatoire $(X_n^x)_n$ de matrice stochastique $P$ issue de $X_0^x = x$, pour tout $x \in E$. On considère

$$(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}}) = ([0, 1], \mathcal{B}([0, 1]), \lambda)$$

et une suite $(U_n)_n$ i.i.d de loi uniforme sur $[0, 1]$ sur $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$. Puisque $E$ est dénombrable, notons $E = \{y_n \,|\, n \in \mathbb{N}\}$, et on construit une chaîne de Markov $(X_n^x)_{n \geq 0}$ partant de $x \in E$ de matrice stochastique $P$ (donc $X_0^x = x$).

**Proposition 5.4.1.** *Soit $E$ un ensemble dénombrable et $P$ une matrice stochastique sur $E$. Il existe un espace $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ sur lequel pour tout $x \in E$, il existe une chaîne de* Markov *$(X_n^x)_n$ issue de $x$ et de matrice $P$.*

*Démonstration.* Soit $(y_j)_j$ une énumération de $E$. Soit $(U_n)_n$ uniforme i.i.d de loi uniforme sur $[0, 1]$, $X_0^x = x$ et pour tout $n \in \mathbb{N}$, $X_{n+1}^x = y_k$ avec $y_k$ tel que

$$\sum_{j \leq k} P(X_n^x, y_j) < U_{n+1} \leq \sum_{j \leq k} P(X_n^x, y_j).$$

On a

$$\begin{aligned}
&\mathbb{P}(X_{n+1}^x = y_k | X_0^x = x, \ldots, X_n^x = x_n) \\
&\qquad = \mathbb{P}\left(Y_{n+1} \in \left]\sum_{j \leq k} P(X_n^x, y_j), \sum_{j \leq k} P(X_n^x, y_j).\right[ \;\middle|\; X_0^x = x, \ldots, X_n^x = x\right) \\
&\qquad\qquad = \sum_{j \leq k} P(x_n, y_j) - \sum_{j < k} P(x_n, y_j) = P(x_n, y_k).
\end{aligned}$$

Puisque la probabilité obtenue ne dépend pas des $x_k$ pour $k < n$, on a établi la propriété de Markov. On observe d'ailleurs que $P$ est la matrice stochastique de $(X_n^x)_n$.

□

**Autre construction.** On propose une seconde construction sur l'espace dit *canonique de la chaîne* $\Omega = E^{\mathbb{N}}$. Sur $\Omega$, on considère la tribu $\mathcal{F} = \mathscr{C}$ la plus petite tribu rendant mesurables les applications coordonnées définies par $X_n(\omega) = \omega_n$ pour tout $n \in \mathbb{N}$ et $\omega \in \Omega$. En fait, $\mathscr{C}$ est la tribu engendrée par la famille Cyl des cylindres, *i.e* des ensembles $C$ de la forme

$$C = \{\omega \in E^{\mathbb{N}} \,|\, \forall i \in [\![0, n]\!], \omega_i = x_i\}$$

où $n \in \mathbb{N}$, et $(x_i)_{0 \leq i \leq n} \in E^{n+1}$. Montrons donc que $\mathscr{C} = \sigma(\mathrm{Cyl})$. Soit $C$ un cylindre (de forme donnée ci-dessus). On remarque que

$$C = \bigcap_{i=0}^{n} X_i^{-1}(\{x_i\}) \in \mathscr{C}$$

donc $\mathrm{Cyl} \subset \mathscr{C}$, donc $\sigma(\mathrm{Cyl}) \subset \mathscr{C}$. De plus, chaque application coordonnée $X_i$ est mesurable pour $\sigma(\mathrm{Cyl})$. En effet,

$$X_i^{-1}(\{x\}) = \{\omega \in E^{\mathbb{N}} \,|\, \omega_i = x\} \in \mathrm{Cyl} \subset \sigma(\mathrm{Cyl})$$

pour tout $x \in E$ et $i \in \mathbb{N}$. Or, $\mathscr{C}$ est la plus petite telle tribu par définition donc $\mathscr{C} = \sigma(\mathrm{Cyl})$. Cette tribu $\mathscr{C} = \sigma(\mathrm{Cyl})$ s'appelle la tribu *cylindrique*.

**Lemme 5.4.1.** *Soit* $(\Omega, \mathcal{F}) = (E^{\mathbb{N}}, \mathscr{C})$ *et* $(\widetilde{\Omega}, \widetilde{\mathcal{F}})$ *un espace mesurable. Une application* $\psi : (\widetilde{\Omega}, \widetilde{\mathcal{F}}) \longrightarrow (\Omega, \mathcal{F})$ *est mesurable si et seulement si* $X_n \circ \psi$ *est mesurable pour tout* $n \in \mathbb{N}$.

*Démonstration.* Le sens direct est immédiat par composition d'applications mesurables. Pour la réciproque, on suppose que $X_n \circ \psi$ est mesurable pour tout $n \in \mathbb{N}$. On pose

$$\mathcal{G} = \{A \in \mathcal{F} \,|\, \psi^{-1}(A) \in \widetilde{F}\}.$$

$\mathcal{G}$ est une tribu qui contient tous les cylindres. En effet, $X_n^{-1}(\{y\}) \in \mathcal{G}$ pour tout $y \in E$ et $n \in \mathbb{N}$ car $X_n^{-1}(\{y\}) \in \mathrm{Cyl} \subset \mathcal{F} = \mathscr{C}$, donc

$$\psi^{-1}(X_n^{-1}(\{y\})) = (X_n \circ \psi)^{-1}(\{y\}) \in \widetilde{F}$$

(par hypothèse). Un cylindre $C$ s'écrit

$$C = \bigcap_{i=0}^{n} \underbrace{X_i^{-1}(\{y_i\})}_{\in \mathcal{G}}$$

et est donc élément de $\mathcal{G}$. On en déduit que $\mathrm{Cyl} \subset \mathcal{G}$, donc que $\mathcal{F} = \sigma(\mathrm{Cyl}) \subset \mathcal{G} \subset \mathcal{F}$. On a ainsi montré que $\mathcal{G} = \mathcal{F}$ et donc que $\psi$ est mesurable.

□

**Théorème 5.4.1** (définition de la chaîne canonique)**.** *Soit $E$ un ensemble dénombrable et $P$ une matrice stochastique sur $E$. Pour tout $x \in E$, il existe une unique mesure de probabilité $\mathbb{P}_x$ sur $(\Omega, \mathcal{F}) = (E^{\mathbb{N}}, \mathscr{C})$ telle que la suite $(X_n)_n$ des applications coordonnées est une chaîne de* Markov *issue de $x$ de matrice stochastique $P$.*

*Démonstration.* — *Existence.* D'après la première construction, il existe un espace de probabilité $(\widetilde{\Omega}, \widetilde{F}, \widetilde{\mathbb{P}})$ sur lequel on a construit $(X_n^x)_n$ une chaîne de Markov issue de $x$ avec $P$ pour matrice stochastique. On considère l'application

$$\psi_x : \left| \begin{array}{ccc} (\widetilde{\Omega}, \widetilde{\mathcal{F}}) & \longrightarrow & (\Omega, \mathcal{F}) \\ \tilde{\omega} & \longmapsto & (X_n^x(\tilde{\omega}))_n. \end{array} \right.$$

D'après le lemme précédent, $\psi_x$ est mesurable puisque pour tout $n \in \mathbb{N}$, $X_n \circ \psi_x = X_n^x$ est une variable aléatoire donc mesurable. Sur $(\Omega, \mathcal{F})$ on considère alors $\mathbb{P}_x = \widetilde{\mathbb{P}} \circ \psi_x^{-1}$ la mesure image de $\widetilde{\mathbb{P}}$ par $\psi_x$. On a $\{X_0 = x\} = C_0 = \{\omega \in E^{\mathbb{N}} \,|\, \omega_0 = x\}$, donc

$$\mathbb{P}(X_0 = x) = \mathbb{P}_x(C_0) = \widetilde{\mathbb{P}}(\psi_x^{-1}(C_0)) = \widetilde{\mathbb{P}}((X_n^x)_n \in C_0) = \widetilde{\mathbb{P}}(X_0^x = x) = 1$$

donc $(X_n)_n$ est bien issue de $x$. Soit maintenant $n \in \mathbb{N}^*$, et $(x_1, \ldots, x_n) \in E^n$. On pose

$$C_n = \{\omega \in \Omega \,|\, \omega_0 = x, \omega_1 = x_1, \ldots, \omega_n = x_n\} \in \mathrm{Cyl}\,.$$

On a

$$\begin{aligned}
\mathbb{P}_x(X_0 = x, X_1 = x_1, \ldots, X_n = x_n) &= \mathbb{P}_x(C_n) \\
&= \widetilde{\mathbb{P}}(\psi_x^{-1}(C_n)) \\
&= \widetilde{\mathbb{P}}((X_n^x)_n \in C_n) \\
&= \widetilde{\mathbb{P}}(X_0^x = x, X_1^x = x_1, \ldots, X_n^x = x_n) \\
&= P(x, x_1)P(x_1, x_2) \cdots P(x_{n-1}, x_n),
\end{aligned}$$

ce qui montre que $(X_n)_n$ est bien une chaîne de MARKOV de matrice stochastique $P$.

— *Unicité.* Si $\mathbb{P}_x$ et $\mathbb{P}'_x$ sont deux mesures de probabilités vérifiant la conclusion du théorème. Alors, pour tout $C \in \mathrm{Cyl}$, $\mathbb{P}_x(C) = \mathbb{P}'_x(C)$. Or, Cyl est un $\pi$-système donc en vertu du théorème de classe monotone, $\mathbb{P}_x$ et $\mathbb{P}'_x$ coïncident sur $\sigma(\mathrm{Cyl}) = \mathcal{F}$ donc sont égales.

□

**Remarque.** Étant donné une probabilité $\nu$ sur $E$, on définit $\mathbb{P}_\nu = \sum_{x\in E} \nu(\{x\})\mathbb{P}_x$. C'est une probabilité sur $(\Omega, \mathcal{F}) = (E^{\mathbb{N}}, \mathscr{C})$. Sous la probabilité $\mathbb{P}_\nu$, on a $X_0 \sim \nu$. En effet,

$$\mathbb{P}_\nu(X_0 \in A) = \sum_{x\in E} \nu(\{x\}) \underbrace{\mathbb{P}_x(X_0 \in A)}_{\mathbb{1}_A(x)} = \sum_{x\in A} \nu(\{x\}) = \nu(A).$$

## 5.5 Propriétés de MARKOV

On travaille sur l'espace canonique $(\Omega, \mathcal{F}) = (E^{\mathbb{N}}, \mathscr{C})$, l'espace des suites d'éléments de $E$ muni de la tribu cylindrique, sur lequel on introduit les opérateurs de décalage (ou translation, ou shift) : si $k \in \mathbb{N}$, on définit pour tout $\omega \in \Omega$ par

$$\Theta_k(\omega) = (\omega_{n+k})_n.$$

On a donc $\Theta_k = \Theta_1^{\circ k}$. L'application $\Theta_k = (\Omega, \mathcal{F}) \longrightarrow (\Omega, \mathcal{F})$ est mesurable car $X_n \circ \Theta_k = X_{n+k}$ pour tout $n, k \in \mathbb{N}$. Dans la suite, on note $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$ pour tout $n \in \mathbb{N}$ pour tout $n \in \mathbb{N}$, et $\mathbb{E}_x$ l'espérance sous $\mathbb{P}_x$, et $\mathbb{E}_\nu$ celle sous $\mathbb{P}_\nu$.

**Théorème 5.5.1** (MARKOV faible)**.** *Soit $G : \Omega \longrightarrow \mathbb{R}$ une application mesurable. Alors, pour tout $x \in E$,*

$$\mathbb{E}_x[G \circ \Theta_n \,|\, \mathcal{F}_n] = \mathbb{E}_{X_n}[G].$$

*De manière équivalente, pour toute fonction $F : \Omega \longrightarrow \mathbb{R}$ $\mathcal{F}_n$-mesurable,*

$$\mathbb{E}_x[F \times (G \circ \Theta_n)] = \mathbb{E}_x[F\mathbb{E}_{X_n}[G]].$$

**Remarque.** $\mathbb{E}_{X_n}[G]$ est la composée de $X_n$ et de $x \mapsto \mathbb{E}_x[G]$. $\mathbb{E}_{X_n}[G]$ est donc une variable aléatoire $\sigma(X_n)$-mesurable (donc $\mathcal{F}_n$-mesurable).

*Démonstration.* On démontre la seconde formulation. Il suffit de considérer

$$F = \mathbb{1}_{\{X_0=x_0,\ldots,X_n=x_n\}} \quad \text{et} \quad F = \mathbb{1}_{\{X_0=y_0,\ldots,X_n=y_p\}}$$

avec $(x_1,\ldots,x_n) \in E^n$ et $(y_1,\ldots,y_p) \in E^p$ des éléments quelconques. Puisque Cyl est un $\pi$-système, par un argument de classe monotone, on peut généraliser à $F = \mathbb{1}_A$ et $G = \mathbb{1}_B$ pour tout $A \in \mathcal{F}_n$ et $B \in \mathcal{F}$. Puis, par les arguments usuels de la théorie de la mesure, on étend à des applications étagées, puis mesurables positives et mesurables de signe quelconque.

Soit maintenant $y \in E$. On a

$$\begin{aligned}\mathbb{E}_y[G] &= \mathbb{P}_y(X_0 = y_0,\ldots,X_p = y_p)\\ &= \delta_{y,y_0}P(y,y_1)P(y_1,y_2)\cdots P(y_{p-1},y_p)\end{aligned}$$

et

$$\begin{aligned}\mathbb{E}_x[F \times (G \circ \Theta_n)] &= \mathbb{E}_x(\mathbb{1}_{\{X_0=x_0,\ldots,X_n=x_n\}}, \mathbb{1}_{\{X_n=y_0,X_{n+1}=y_1,\ldots,X_{n+p}=y_p\}})\\ &= \mathbb{P}_x(X_0 = x_0,\ldots,X_n = x_n, X_n = y_0,\ldots,X_{n+p} = y_p)\\ &= \delta_{x,x_0}P(x_1,x_2)P(x_{n-1},x_n)\delta_{x_n,y_0}P(y_{p-1},y_p).\end{aligned}$$

On a

$$\begin{aligned}\mathbb{E}_x[F\mathbb{E}_{X_n}[G]] &= \mathbb{E}_x[\mathbb{1}_{\{X_0=x_0,\ldots,X_n=x_n\}}\delta_{X_n,y_0}P(y_0,y_1)\cdots P(y_{p-1},y_p)]\\ &= \underbrace{\mathbb{P}_x(X_0 = x_0,\ldots,X_n = x_n)}_{\delta_{x,x_0}P(x,x_1)\cdots P(x_{n-1},x_n)}\delta_{x_n,y_0}P(y_0,y_1)\cdots P(y_{p-1},y_p),\end{aligned}$$

d'où l'égalité cherchée.

□

**Théorème 5.5.2** (Markov fort)**.** *Soit $T$ un temps d'arrêt pour la filtration canonique $(\mathcal{F}_n)_n$ de $(X_n)_n$. Alors, pour tout fonction mesurable $G : \Omega \longrightarrow \mathbb{R}$,*

$$\mathbb{E}_x[\mathbb{1}_{\{T<\infty\}}G \circ \Theta_T|\mathcal{F}_T] = \mathbb{1}_{\{T<\infty\}}\mathbb{E}_{X_T}[G].$$

*De manière équivalente, pour toute fonction $F : \Omega \longrightarrow \mathbb{R}$ $\mathcal{F}_T$-mesurable et $G : \Omega \longrightarrow \mathbb{R}$ mesurable,*

$$\mathbb{E}_x[F(G \circ \Theta_T)\mathbb{1}_{\{T<\infty\}}] = \mathbb{E}_x[F\mathbb{E}_{X_T}[G]\mathbb{1}_{\{T<\infty\}}].$$

*Démonstration.* Soit $n \geq 0$, et $F : \Omega \longrightarrow \mathbb{R}$ $\mathcal{F}_T$-mesurable. Alors, $\mathbb{1}_{\{T=n\}}F$ est $\mathcal{F}_n$-mesurable. En effet,

$$\{\mathbb{1}_{\{T=n\}} \in A\} = \underbrace{(\{T = n\} \cap \underbrace{\{F \in A\}}_{\in\mathcal{F}_T})}_{\in\mathcal{F}_n} \cup (\{T \neq n\} \cap \{0 \in A\}).$$

Par Markov faible avec $\mathbb{1}_{\{T=n\}}F$ qui est $\mathcal{F}_n$-mesurable, on a

$$\begin{aligned}\mathbb{E}_x[\mathbb{1}_{\{T=n\}}FG\circ\Theta_T] &= \mathbb{E}_x[\mathbb{1}_{\{T=n\}}FG\circ\Theta_n]\\ &= \mathbb{E}_x[\mathbb{1}_{\{T=n\}}FE_{X_n}[G]]\\ &= \mathbb{E}_x[\mathbb{1}_{\{T=n\}}F\mathbb{E}_{X_T}[G]].\end{aligned}$$

En sommant sur $n \geq 0$, on a

$$\mathbb{E}_x[\mathbb{1}_{\{T<\infty\}}FG\circ\Theta_T] = \mathbb{E}_x[\mathbb{1}_{\{T<\infty\}}F\mathbb{E}_{X_T}[G]]$$

car $\sum_{n=0}^{\infty}\mathbb{1}_{\{T=n\}} = \mathbb{1}_{\{T<\infty\}}$.

□

**Corollaire 5.5.1.** *Soit $T$ un temps d'arrêt tel que $\mathbb{P}_x(T<\infty)=1$. On suppose qu'il existe $y\in\mathbb{N}$ tel que $\mathbb{P}(X_T=y)=1$. Alors, sous $\mathbb{P}_x$, la variable $\Theta_T$ est indépendante de $\mathcal{F}_T$ et sa loi est $\mathbb{P}_y$.*

*Démonstration.* Pour toutes fonctions bornées $\mathcal{F}_T$-mesurable $F$ et $G$ mesurable, la propriété de MARKOV forte donne

$$\mathbb{E}_x[F\times(G\circ\Theta_T)] = \mathbb{E}_x[F\mathbb{E}_{X_T}[G]] = \mathbb{E}_x[F\mathbb{E}_y[G]] = \mathbb{E}_x[F]\mathbb{E}_y[F].$$

Dans le cas particulier $F=1$, on trouve que

$$\mathbb{E}_x[G\circ\Theta_T] = \mathbb{E}_y[G]$$

qui une fois réinjecté dans nos calculs permet de dire que $\mathbb{E}_x[F\times(G\circ\Theta_T)] = \mathbb{E}_x[F]\mathbb{E}_x[G\circ\Theta_T]$. Ceci étant vrai pour toutes fonctions bornées $\mathcal{F}_T$-mesurable $F$ et $G$ mesurable, ceci montre que $\Theta_T$ et $\mathcal{F}_T$ sont indépendantes sous $\mathbb{P}_x$. Enfin, la dernière égalité signifie $\mathbb{P}_x\circ\Theta^{-1}=\mathbb{P}_y$, c'est-à-dire que $\Theta_T$ est de loi $\mathbb{P}_y$ sur $E^{\mathbb{N}}$.

□

**Reformulation de la propriété de MARKOV.** En notant $\mathscr{L}_\nu((X_n)_n)$ la loi de la chaîne de MARKOV $(X_n)_n$ avec $X_0\sim\nu$, la seconde propriété de MARKOV forte s'écrit

$$\mathscr{L}_\nu((X_n)_{n\geq T}|\mathcal{F}_T) = \mathscr{L}_{X_T}((X_n)_{n\geq 0}),$$

ou pour tout $B\in\mathscr{C}$,

$$\mathbb{P}_\nu((X_n)_{n\geq T}\in B|\mathcal{F}_T) = \mathbb{P}_{X_T}((X_n)_n\in B).$$

Lorsque $T$ est un temps d'arrêt, il s'agit de MARKOV fort. Si $T=p$ est déterministe, il s'agit de MARKOV faible. On peut encore reformuler les propriétés faible et forte de MARKOV, comme dans le corollaire suivant.

**Corollaire 5.5.2.** *Soient $A\in\mathscr{C}$, $(x_n)_n\in E^{\mathbb{N}}$, $y\in E$, et $T$ un temps d'arrêt presque sûrement fini. Alors,*

$$\mathbb{P}(\Theta_pX\in A|X_0=x_0,\ldots,X_p=x_p) = \mathbb{P}_{x_p}(X\in A)\quad\text{(MARKOV \textit{faible})},$$

*et*

$$\mathbb{P}(\Theta_TX\in A|X_0=x_0,\ldots,X_T=y) = \mathbb{P}_y(X\in A)\quad\text{(MARKOV \textit{fort})}.$$

La propriété de MARKOV justifie également que pour une chaîne de MARKOV, passé et futur sont indépendants sachant le présent.

**Corollaire 5.5.3.** *Soit $n \geq 1$, $A \in \mathcal{F}_n$, 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 $\mathbb{P}_x$-presque sûrement fini, alors pour tout $A \in \mathcal{F}_T$ et $B \in \{\Theta^{-1}(A) \,|\, A \in \mathcal{F}\}$, on a*

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

**Remarque.** Puisque $\Theta_T$ est mesurable, $\{\Theta^{-1}(A) \,|\, A \in \mathcal{F}\}$ est une tribu qui contient les évènements réalisés après $T$. C'est une manière adéquate pour remplacer $\sigma((X_k)_{k\geq T})$ puisque $\{(X_1, \ldots, X_n) \circ \Theta_T \in B\} = \{(X_{T+1}, \ldots, X_{T+n}) \in B\}$.

*Démonstration.* Soit $F$ une fonction $\mathcal{F}_n$-mesurable et $G$ une fonction $\sigma((X_k)_{k\ \geq n})$-mesurable. On peut écrire $G = \widetilde{G} \circ \Theta_n$ (selon le théorème de DINKYN). On a donc

$$\begin{aligned} \mathbb{E}_x[FG|X_n] &= \mathbb{E}_x[F(\widetilde{G} \circ \Theta_n)|X_n] \\ &= \mathbb{E}_x[\mathbb{E}_x[F(\widetilde{G} \circ \Theta_n)|\mathcal{F}_n]|X_n] \\ &= \mathbb{E}_x[F\mathbb{E}_x[\widetilde{G} \circ \Theta_n|\mathcal{F}_n]|X_n] \\ &= \mathbb{E}_x[F\mathbb{E}_{X_n}[\widetilde{G}]|X_n] \\ &= \mathbb{E}_x[F|X_n]\mathbb{E}_{X_n}[\widetilde{G}]. \end{aligned}$$

Le cas particulier $F = 1$ donne $\mathbb{E}_{X_n}[\widetilde{G}] = \mathbb{E}_x[G|X_n]$ ce qui une fois réinjecté dans nos calculs montre que

$$\mathbb{E}_x[FG|X_n] = \mathbb{E}_x[F|X_n]\mathbb{E}_x[G|X_n].$$

On a donc montré la première partie du corollaire. La seconde se fait de la même façon avec la propriété de MARKOV forte.

□

# Chapitre 6

# Récurrence et transciences

On considère une chaîne de Markov $(X_n)_n$ de matrice de transition $P$ sur un espace d'état $E$. Sauf mention contraire, cette chaîne sera la chaîne canonique construite à la fin du chapitre précédent. On note pour tout $y \in E$,

$$T_y = \inf\{n \in \mathbb{N} \mid X_n = y\} \quad \text{et} \quad \widetilde{T}_y = \inf\{n \in \mathbb{N}^* \mid X_n = y\}.$$

Ce sont des temps d'arrêt, le premier étant un temps d'atteinte de $y$, et le second un temps d'atteinte de $y$ après le départ. Enfin, on note

$$\rho_{x,y} = \mathbb{P}_x(\widetilde{T}_y < \infty)$$

la probabilité que, partant de l'état $x$, la chaîne arrive en temps fini à l'état $y$. Ainsi, la quantité $\rho_{x,x}$ est la probabilité que la chaîne partant de $x$, finisse par y revenir en temps fini.

## 6.1 États récurrents et transitoires

**Définition 6.1.1.** *Un état $x \in E$ est dit*

- ***récurrent*** *si $\rho_{x,x} = 1$ ;*
- ***transitoire*** *si $\rho_{x,x} < 1$.*

**Remarque.** Le nombre de passages sur l'état $x \in E$ est donné par la variable aléatoire $N(x) = \sum_{k=0}^{\infty} \mathbb{1}_{\{X_k = x\}}$.

### 6.1.1 Nombre de passages

**Proposition 6.1.1.** *Pour tout $x \in E$, on a l'alternative suivante.*

- *Si $x$ est récurrent, alors $N(x) = \infty$ $\mathbb{P}_x$-presque sûrement.*
- *Si $x$ est transitoire, alors $N(x) < \infty$ $\mathbb{P}_x$-presque sûrement, et*

$$\mathbb{E}_x[N(x)] = \frac{1}{1 - \rho_{x,x}}.$$

*Démonstration.* Remarquons d'abord que $\mathbb{P}_x(N(x) \geq 1) = 1$. Soit $k \geq 1$. Selon la propriété de Markov forte, on a

$$\begin{aligned}\mathbb{P}_x(N(x) \geq k+1) = \mathbb{E}_x[\mathbb{1}_{\{N(x)\geq k+1\}}] &= \mathbb{E}_x[\mathbb{1}_{\{\widetilde{T}_x<\infty\}}\mathbb{1}_{\{N(x)\geq k\}} \circ \Theta_{\widetilde{T}_x}]\\ &= \mathbb{P}_x(\widetilde{T}_x) < \infty\mathbb{E}[\mathbb{1}_{\{N(x)\geq k\}}]\\ &= \rho_{x,x}\mathbb{P}_x(N(x) \geq k).\end{aligned}$$

Par récurrence, il est clair que pour tout $k \geq 1$,

$$\mathbb{P}_x(N(x) \geq k) = \rho_{x,x}^{k-1}.$$

Si $x$ est un état récurrence, alors $\mathbb{P}_x(N(x) \geq k)$ pour tout $k \geq 1$ et on conclut par convergence monotone. Si $x$ est un état transitoire, alors pour tout $k \geq 1$,

$$\mathbb{P}_x(N(x) = k) = \mathbb{P}_x(N(x) \geq k) - \mathbb{P}_x(N(x) \geq k+1) = \rho_{x,x}^{k-1}(1-\rho_{x,x}).$$

On a donc aussi le second point. □

Soit $y \in E$ et $\tilde{N}(y) = \sum_{k=1}^{\infty} \mathbb{1}_{\{}X_k = y\}$ le nombre de passage de la chaîne par l'état $y$ après le départ. Si elle part d'un état $x \in E\backslash\{y\}$, alors $\tilde{N}(y) = N(y)$ et $\tilde{T}_y = T_y$. Si elle part de l'état $x = y$ alors $\tilde{N}(y) = N(y) - 1$, et le temps d'arrêt $\tilde{T}_y$ est le temps de premier retour en $y$. Soit alors la suite de temps d'arrêt $(T_y^{(k)})_{k\geq 0}$ définie récursivement par $T_y^{(0)} = 0$, et

$$T_y^{(k+1)} = \min\{n > T_y^{(k)} \mid X_n = y\}$$

pour tout $k \in \mathbb{N}$.

**Proposition 6.1.2.** *Soient $y \in E$ et $n \in \mathbb{N}$. Sur l'évènement $\{T_y^{(n)} < \infty\}$, les variables aléatoires $\Delta_y^{(k)} = T_y^{(k)} - T_y^{(k-1)}$ pour tout $k \in [\![1, n]\!]$ sont indépendantes et identiquement distribuées sous $\mathbb{P}_y$.*

*Démonstration.* Puisque $\{T_y^{(n)} < \infty\} = \bigcap_{i=1}^n \{\Delta_y^{(i)} < \infty\}$, il suffit de montrer pour tout $k \in [\![1, n]\!]$ que

$$\mathbb{E}_y\left[\prod_{i=1}^k g(\Delta_y^{(i)})\mathbb{1}_{\{\Delta^{(i)}<\infty\}}\right] = \prod_{i=1}^k \mathbb{E}_y\left[g(\Delta_y^{(i)})\mathbb{1}_{\{\Delta^{(i)}<\infty\}}\right]$$

pour toutes fonctions $(g_i)_{1\leq i\leq k}$ mesurables et bornées sur $\mathbb{R}_+$. Pour cela, on travaille par récurrence sur $k \in [\![1, n]\!]$. Le cas $k = 1$ étant évident, on montre directement l'hérédité. Soit $k \in [\![2, n]\!]$, supposons la relation vraie au rang $k-1$ et montrons là au rang $k+1$. Observons avant tout les faits suivants :

— les variables aléatoires $\Delta_y^{(1)}, \dots, \Delta_y^{(k-1)}$ sont $\mathcal{F}_{T_y^{(k-1)}}$-mesurables ;

— la variable aléatoire $\Theta_{T_y^{(k-1)}}$ est indépendante de la tribu $\mathcal{F}_{T_y^{(k-1)}}$ et est de loi $\mathbb{P}_y$ (d'après la propriété forte de Markov) ;

— $\Delta_y^{(k)} = \Delta_y^{(1)} \circ \Theta_{T_y^{(k-1)}}$.

Soient donc $g_1, \dots, g_k$ des fonctions mesurables et bornées sur $\mathbb{R}_+$. La propriété de Markov forte donne

$$
\begin{aligned}
&\mathbb{E}_y\left[\prod_{i=1}^{k} g_i\left(\Delta_y^{(i)}\right)\mathbb{1}_{\{\Delta_y^{(i)}<\infty\}}\right] \\
&\qquad = \mathbb{E}_y\left[\prod_{i=1}^{k-1}\left(g_i\left(\Delta_y^{(i)}\right)\mathbb{1}_{\{\Delta_y^{(i)}<\infty\}}\right) g\left(\Delta_y^{(1)}\circ\Theta_{T_y^{(k-1)}}\right)\mathbb{1}_{\left\{\Delta^{(1)}\circ\Theta_{T_y^{(k-1)}}<\infty\right\}}\right] \\
&\qquad = \mathbb{E}_y\left[\prod_{i=1}^{k-1}\left(g_i\left(\Delta_y^{(i)}\right)\mathbb{1}_{\{\Delta_y^{(i)}<\infty\}}\right)\right] \\
&\qquad \times \mathbb{E}_y\left[g_k\left(\Delta_y^{(1)}\circ\Theta_{T_y^{(k-1)}}\right)\mathbb{1}_{\left\{\Delta^{(1)}\circ\Theta_{T_y^{(k-1)}}<\infty\right\}}\right] \\
&\qquad = \mathbb{E}_y\left[\prod_{i=1}^{k-1}\left(g_i\left(\Delta_y^{(i)}\right)\mathbb{1}_{\{\Delta_y^{(i)}<\infty\}}\right)\right]\mathbb{E}_y\left[g_k\left(\Delta_y^{(k)}\right)\mathbb{1}_{\left\{\Delta_y^{(k)}<\infty\right\}}\right] \\
&\qquad = \prod_{i=1}^{k-1}\mathbb{E}_y\left[g_i\left(\Delta_y^{(i)}\right)\mathbb{1}_{\left\{\Delta_y^{(i)}<\infty\right\}}\right]\mathbb{E}_y\left[g_k\left(\Delta_y^{(k)}\right)\mathbb{1}_{\left\{\Delta_y^{(k)}<\infty\right\}}\right].
\end{aligned}
$$

□

**Proposition 6.1.3** (nombres de passages en $y$ partant de $x \neq y$)**.** *Soient $x, y \in E$ deux états distincts. Sous la probabilité $\mathbb{P}_x$,*

*(i) si $y$ est transitoire,*

$$N(y) \overset{\mathbb{P}_x}{\sim} (1-\rho_{x,y})\delta_0 + \rho_{x,y}\mathcal{G}(1-\rho_{y,y})$$

*et $\mathbb{P}_x(N(y) < \infty) = 1$ ;*

*(ii) si $y$ est récurrent et la chaîne part de $x$, ou bien elle ne rejoint jamais $y$ (i.e $N(y) = 0$), ou bien elle le rejoint une fois, puis une infinité de fois (i.e $N(y) = \infty$). De plus,*

$$N(y) \overset{\mathbb{P}_x}{\sim} (1-\rho_{x,y})\delta_0 + \rho_{x,y}\delta_\infty,$$

*$\mathbb{P}_y(N(y) = \infty) = 1,$ et $\mathbb{P}_x(N(y) = \infty) = \rho_{x,y}$.*

**Remarque.** Les propositions 6.1.1 et 6.1.3 permettent de décrire et différencier les états récurrents et transitoires.

— Si $y$ est transitoire, alors peu importe l'état initial de la chaîne, le nombre de passage en $y$ est fini. Le nombre moyen de passage en $y$ est aussi fini.

— Si $y$ est récurrent, alors si la chaîne part de cet état, elle y repasse un infinité de fois. Si elle ne part pas de $y$, ou bien elle n'y va jamais, ou bien elle y passe une fois puis une infinité de fois.

### 6.1.2 Potentiel ou fonction de Green

**Définition 6.1.2** (potentiel ou fonction de Green)**.** *Pour tout $x, y \in E$, on note*

$$G(x,y) = \mathbb{E}_x[N(y)] = \sum_{k=0}^{\infty} \mathbb{P}_x(X_k = y) = \sum_{k=0}^{\infty} P^k(x,y).$$

*La fonction $G : E^2 \longrightarrow \mathbb{R}$ est appelé* ***potentiel*** *(ou fonction de* Green*)*

**Interprétation.** $G(x,y)$ est le nombre de moyen de passages en $y$ lorsque la chaîne part de $x$.

**Théorème 6.1.1** (récurrence, transcience et potentiel)**.** *On a les propriétés suivantes.*

*(i) Si $y$ est un état transitoire, alors le potentiel $G(x,y)$ est fini pour tout état $x$. Plus précisément,*

$$G(x,y) = \begin{cases} \dfrac{\rho_{x,y}}{1-\rho_{x,y}} & \text{si } x \neq y\,; \\ \dfrac{1}{1-\rho_{y,y}} & \text{sinon.} \end{cases}$$

*(ii) Si $y$ est un état récurrent, alors $G(y,y) = \infty$ et*

$$G(x,y) = \begin{cases} 0 & \text{si } \rho_{x,y} = 0\,; \\ \infty & \text{sinon.} \end{cases}$$

De ce théorème se déduit facilement un critère de récurrence.

**Corollaire 6.1.1** (récurrence et potentiel)**.** *Un état $x \in E$ est récurrent si et seulement si $G(x,y) = \infty$.*

**Corollaire 6.1.2.** *En convenant que $0 \times \infty = 0$, pour tout $x, y \in E$ distincts*

$$G(x,y) = \rho_{x,y} G(y,y).$$

**Définition 6.1.3.** *Une chaîne est dite transitoire (resp. récurrente) si tous ses états sont transitoires (resp. récurrents).*

**Définition 6.1.4.** *Un état $x \in E$ est dit*

- ***récurrent positif*** *si $m_x = \mathbb{E}_x[\widetilde{T}_x] < \infty$.*
- ***récurrent nul*** *s'il est récurrent mais $m_x = \infty$.*

**Interprétation foireuse.** Un élément est récurrent positif s'il arrive souvent (récurrent) et qu'on attend généralement pas trop longtemps pour revenir à l'état $x$. Au contraire un élément récurrent est récurrent nul s'il arrive souvent mais qu'il faut attendre le dégel entre deux passages en $x$.

**Exemple.** Un état $x \in E$ est dit absorbant lorsque $P(x,x) = 1$. Un tel état est récurrent positif puisque dans ce cas, $\widetilde{T}_x = 1$ $\mathbb{P}_x$-presque sûrement.

**Définition 6.1.5.** *Une chaîne de* Markov *est dite récurrente positive (resp. récurrente nulle, resp. transitoire) si tous ses états sont récurrents positifs (resp. récurrents nuls, resp. transitoires).*

**Exemple.** Lorsque $E$ est fini, alors la chaîne ne peut pas être transitoire car au moins un état est récurrent. En effet, on rappelle que si $y$ est un état transitoire, alors

$$G(x,y) = \sum_{n=0}^{\infty} P^n(x,y) < \infty$$

pour tout $x \in E$. Alors, $P^n(x,y) \xrightarrow[n\to\infty]{} 0$. Si la chaîne était transitoire, on aurait

$$0 = \sum_{y\in E} \lim_{n\to\infty} P^n(x,y) = \lim_{n\to\infty} \sum_{y\in E} P^n(x,y) = \mathbb{P}_x(X_n \in E) = 1.$$

Il existe donc un état récurrent.

### 6.1.3 Temps passé sur un état.

Soit $y \in E$, on note

$$N_n(y) = \sum_{k=0}^{n} \mathbb{1}_{\{X_k = y\}}$$

le nombre de passages en $y$ à la date $n$ et

$$G_n(x,y) = \mathbb{E}_x[N_n(y)] = \sum_{k=0}^{n} P^k(x,y).$$

**Théorème 6.1.2.** *Pour tout $y \in E$ et toute loi initiale $\nu$,*

$$\lim_{n\to\infty} \frac{N_n(y)}{n} = \frac{\mathbb{1}_{\{\widetilde{T}_y < \infty\}}}{m_y}$$

*$\mathbb{P}_\nu$-presque sûrement. De plus, pour tout $x \in E$,*

$$\lim_{n\to\infty} \frac{G_n(x,y)}{n} = \frac{\rho_{x,y}}{m_y}.$$

*Démonstration.* D'abord, on suppose $y$ récurrent et que la chaîne part de $y$ : $\mathbb{1}_{\{\widetilde{T}_y<\infty\}} = 1$ $\mathbb{P}_y$-presque sûrement. On note

$$T_y^{(k)} \min\{n \geq 1 \,|\, N_n(y) = k\}.$$

On a vu précédemment que les $\Delta_y^{(k)} = T_y^{(k)} - T_y^{(k-1)}$ sont des variables i.i.d [1]. Puisque $T_y^{(n)} = \sum_{k=1}^{n} \Delta_y^{(k)}$, on a par la loi des grands nombres

1. Elles le sont sachant $T_y^{(n)} < \infty$, ce qui est vrai selon nos hypothèses.

$$\lim_{n\to\infty} \frac{T_y^{(n)}}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \Delta_y^{(k)} = \mathbb{E}_y[\Delta_y^{(1)}] = \mathbb{E}_y[\widetilde{T}_y] = m_y$$

($\mathbb{P}_y$-presque sûrement). Si $m_y < \infty$, l'utilisation de loi des grands nombres est licite (car alors $\Delta_y^{(k)}$ est intégrable). Si $m_y = \infty$, on applique la loi des grands nombres à $\Delta_y^{(k)} \wedge a$ (variable aléatoire tronquée au niveau $a > 0$). On aurait dans ce cas

$$\frac{1}{n} = \sum_{k=1}^{n} (\Delta_y^{(k)} \wedge a) \xrightarrow[n\to\infty]{} \mathbb{E}_y[\Delta_y^{(n)} \wedge a].$$

On a donc

$$\liminf_{n\to\infty} \frac{\Delta_y^{(1)} + \cdots + \Delta_y^{(n)}}{n} \geq \liminf_{n\to\infty} \sum_{k=1}^{n} (\Delta_y^{(k)} \wedge a) = \mathbb{E}_y[\Delta_y^{(1)} \wedge a].$$

En faisant tendre $a$ vers $\infty$, on a

$$\mathbb{E}_y[\Delta_y^{(1)} \wedge a] \xrightarrow[a\to\infty]{} \mathbb{E}_y[\Delta_y^{(1)}] = m_y = \infty.$$

Ainsi,

$$\lim_{n\to\infty} \frac{\Delta_y^{(1)} + \cdots + \Delta_y^{(n)}}{n} = \infty.$$

Or, on a

$$T_y^{(N_n(y))} \leq n < T_y^{(N_n(y)+1)}$$

et $\frac{T_y^{(N_n(y))}}{N_n(y)} \xrightarrow[n\to\infty]{} m_y$ car $N_n(y) \xrightarrow[n\to\infty]{} \infty$. De plus,

$$\frac{T_y^{(N_n(y)+1)}}{N_n(y)} = \frac{T_y^{(N_n(y)+1)}}{N_n(y)+1} \frac{N_n(y)+1}{N_n(y)} \xrightarrow[n\to\infty]{} m_y \cdot 1.$$

On a donc par encadrement

$$\frac{n}{N_n(y)} \xrightarrow[n\to\infty]{} m_y.$$

Si la chaîne part de $x$, soit elle rejoint une première fois $y$ et on applique ce que l'on vient de démontrer, soit la chaîne n'atteint jamais $y$ et $\frac{N_n(y)}{n} = 0$ pour tout $n \in \mathbb{N}^*$. On a donc

$$\frac{N_n(y)}{n} \xrightarrow[n\to\infty]{} \frac{\mathbb{1}_{\{\widetilde{T}_y < \infty\}}}{m_y}$$

$\mathbb{P}_x$-presque sûrement. Pour une loi initiale $\nu$ quelconque, $\mathbb{P}_\nu = \sum_{x\in E} \nu(\{x\})\mathbb{P}_x$. Si $\mathbb{P}_x(A) = 1$ pour tout $x \in E$, alors

$$\mathbb{P}_\nu(A) = \sum_{x\in E} \nu(\{x\})\mathbb{P}_x(A) = 1.$$

Lorsque $y$ est transitoire ($m_y = \infty$),

$$N_n(y) \xrightarrow[n\to\infty]{} N(y) < \infty$$

$\mathbb{P}_x$-presque sûrement. On a donc

$$\frac{N_n(y)}{n} \xrightarrow[n\to\infty]{} 0 = \frac{\mathbb{1}_{\{\widetilde{T}_y<\infty\}}}{m_y}$$

$\mathbb{P}_x$-presque sûrement, et donc aussi $\mathbb{P}_\nu$-presque sûrement.

Pour la seconde partie, on remarque que $0 \leq \frac{N_n(y)}{n} \leq 1$ et on applique de la théorème de convergence dominée. Alors,

$$\frac{G_n(x,y)}{n} = \mathbb{E}_x\left[\frac{N_n(y)}{n}\right] \xrightarrow[n\to\infty]{} \mathbb{E}_x\left[\frac{\mathbb{1}_{\{\widetilde{T}_y<\infty\}}}{m_y}\right] = \frac{\rho_{x,y}}{m_y}.$$

□

## 6.2 Ensemble clos et irréductible

Notons $E_R$ l'ensemble des états récurrents de $E$, et $E_T$ l'ensemble des états transitoires de $E$. On a alors $E = E_R \sqcup E_T$.

**Définition 6.2.1.** *Étant donné $x, y \in E$, on dit que $x$ peut mener à $y$ ($x \rightsquigarrow y$) si $\rho_{x,y} = \mathbb{P}_x(\widetilde{T}_y < \infty) > 0$.*

**Proposition 6.2.1.** *Si $x$ et $y$ sont deux états distincts, on a les équivalences suivantes.*

*(i)* $x \rightsquigarrow y$ ;

*(ii)* $G(x,y) > 0$ ;

*(iii)* *il existe $n \geq 1$ tel que $P^n(x,y) >$, i.e qu'il existe un chemin de $x$ à $y$ en $n$ étapes de probabilité non nulle.*

*Démonstration.* On suppose que la chaîne part de $x \neq y$.

— $(i) \implies (ii)$. On a $\{N(y) \geq 1\} = \{\widetilde{T}_y < \infty\}$, donc

$$\mathbb{P}_x(N(y) \geq 1) = \mathbb{P}_x(\widetilde{T}_y < \infty) > 0.$$

Ainsi, $G(x,y) = \mathbb{E}_x[N(y)] > 0$.

— $(ii) \implies (iii)$. $G(x,y) = \sum_{k=0}^{\infty} P^k(x,y)$, il existe donc $n \geq 1$ tel que $P^n(x,y) > 0$.

— $(ii) \implies (i),(ii)$. S'il existe $n \geq 1$ tel que $P^n(x,y) > 0$, alors $G(x,y) > 0$ et donc $\mathbb{P}_x(N(y) \geq 1) > 0$ et $\rho_{x,y} > 0$.

□

**Proposition 6.2.2** (transitivité)**.** *Soient $x, y, z \in E$. Si $x \rightsquigarrow y$ et $y \rightsquigarrow z$, alors $x \rightsquigarrow z$.*

*Démonstration.* Il existe $n \geq 1$ tel que $P^n(x,y) > 0$ et $m \geq 1$ tel que $P^m(y,z) > 0$. Alors,

$$P^{n+m}(x,z) > P^n(x,y)P^m(y,z) > 0$$

et donc $x \rightsquigarrow z$.

□

**Théorème 6.2.1.** *Soit $x \in E_R$ et $y \in E$ tels que $x \rightsquigarrow y$. Alors, $y \in E_R$ et $\rho_{y,x} = 1$. En particulier, $y \rightsquigarrow x$ et $\rho_{x,y} = 1$.*

*Démonstration.* Il suffit de montrer que $y \in E_R$ et $\rho_{y,x} = 1$. La seconde partie vient automatiquement par symétrie des rôles avec le premier fait. Si $y = x$, l'énoncé est immédiat. Supposons donc $x \neq y$. On commence par voir que $\mathbb{P}_y(T_x < \infty) = 1 = \rho_{y,x}$. Puisque $x$ est récurrent,

$$\begin{aligned}
0 &= \mathbb{P}(N(x) < \infty) \\
&\geq \mathbb{P}_x(T_y < \infty, T_x \circ \Theta_{T_y} = \infty) \\
&= \mathbb{E}_x[\mathbb{1}_{\{T_y<\infty\}} \times \mathbb{1}_{\{T_x=\infty\}} \circ \Theta_{T_y}] \\
&= \mathbb{E}_x[\mathbb{1}_{\{T_y<\infty\}}]\mathbb{E}_y[\mathbb{1}_{\{T_x=\infty\}}] \quad (\text{Markov fort}) \\
&= \mathbb{P}_x(T_y < \infty)\mathbb{P}_y(T_x = \infty) \\
&= 0
\end{aligned}$$

car $\mathbb{P}_y(T_x = \infty)$ puisque $x \rightsquigarrow y$. On a donc $\mathbb{P}_y(T_x < \infty) = 1$, *i.e* $\rho_{y,x} = 1$. On montre enfin que $y \in E_R$ en établissant $G(y, y) = \infty$. Il existe $n_1, n_1 \geq 1$ des entiers tels que $P^{n_1}(x, y) > 0$ et $P^{n_2}(y, x) > 0$. On a pour tout $k \geq 0$,

$$P^{n_1+k+n_2}(y, y) \geq P^{n_2}(y, x)P^k(x, x)P^{n_1}(x, y).$$

On a donc

$$\sum_{k=0}^{\infty} P^{n_1+k+n_2}(y, y) \geq \sum_{k=0}^{\infty} P^{n_2}(y, x)P^k(x, x)P^{n_1}(x, y).$$

Ainsi,

$$G(y, y) - \sum_{j=0}^{n_1+n_2-1} P^j(y, y) \geq P^{n_2}(y, x)G(x, x)P^{n_1}(x, y).$$

Or, $G(x, x) = \infty$, $P^{n_2}(y, x) > 0$, et $P^{n_1}(x, y) > 0$. Ainsi, $G(y, y) = \infty$ donc $y \in E_R$.

□

**Théorème 6.2.2.** *Soit $x \in E$ un état récurrent positif et $y \in E$ tel que $x \rightsquigarrow y$. Alors, $y$ est récurrent positif.*

*Démonstration.* Comme précédemment, il existe $n_1, n_2 \geq 1$ deux entiers tels que $P^{n_1}(x, y) > 0$ et $P^{n_2}(y, x) > 0$. Pour tout $k \in \mathbb{N}$,

$$P^{n_1+k+n_2}(y, y) \geq P^{n_2}(y, x)P^k(x, x)P^{n_1}(x, y).$$

On a donc

$$\sum_{k=0}^{\infty} P^{n_1+k+n_2}(y, y) \geq \sum_{k=0}^{\infty} P^{n_2}(y, x)P^k(x, x)P^{n_1}(x, y).$$

Ainsi,

$$G_{n_1+n_2+n}(y, y) - G_{n_1+n_2-1}(y) \geq P^{n_2}(y, x)G_n(x, x)P^n(x, y).$$

En divisant par $n$,

$$\frac{G_{n_1+n_2+n}(y,y)}{n} - \frac{G_{n_1+n_2-1}(y)}{n} \geq P^{n_2}(y,x)\frac{G_n(x,x)}{n}P^n(x,y).$$

En passant à la limite lorsque $n$ tend vers l'infini,

$$\frac{1}{m_y} \geq \frac{P^{n_2}(y,x)P^{n_1}(x,y)}{m_x} > 0.$$

Ainsi, $m_y < \infty$, *i.e* que $y$ est récurrent positif.

□

**Définition 6.2.2.** *Un ensemble d'états $C \subset E$ est dit clos si aucun état de $C$ ne mène à l'extérieur de $C$. Autrement dit, pour tout $x \in C$ et $y \in E \backslash C$, $\rho_{x,y} = 0$.*

**Exemple.** Si $a \in E$ est un état absorbant, alors $\{a\}$ est clos.

**Remarque.** D'après la définition, $C$ est clos si pour tout $x \in C$ et $y \notin C$ et $n \geq 1$, $P^n(x,y) = 0$. En fait, la définition est équivalente à dire que pour tout $x \in C$ et $y \in E\backslash C$, $P(x,y) = 0$. En effet, si pour tout $(x,y) \in C \times (E\backslash C)$, $P(x,y) = 0$, alors par récurrence sur $n$, on montre que $P^(x,y) = 0$ pour tout $n \geq 1$. Si c'est vrai pour $n \in \mathbb{N}$, alors pour tout $x \in C$ et $y \in C^c$,

$$P^{n+1}(x,y) = \sum_{z\in E} P(x,z)P^n(z,x) = \sum_{z\in C} P(x,z)\underbrace{P^n(z,x)}_{=0} = 0.$$

**Définition 6.2.3.** *Un ensemble clos $C \subset E$ est dit irréductible si pour tout $x, y \in C$, $x$ peut mener à $y$. En particulier, une chaîne de* Markov *est dite irréductible si l'espace d'états $E$ est irréductible.*

**Théorème 6.2.3.** *On a les affirmations suivantes.*

*(i) Soit $C \subset E$ un ensemble clos irréductible constitué d'états récurrents (ou qui contient un état récurrent), alors pour tout $x, y \in C$, $\rho_{x,y} = 1$, $\mathbb{P}_x(N(y) = \infty) = 1$ et $G(x,y) = \infty$.*

*(ii) Soit $C \subset E$ un ensemble clos irréductible* ***fini****. Alors, tous les états de $C$ sont récurrents positifs.*

*(iii) En particulier, une chaîne irréductible sur un espace d'état fini est récurrente positive.*

*Démonstration.* On ne montre que le point $(ii)$. Si tous les états états $y$ de $C$ sont transitoires, alors pour tout $x \in E$,

$$P^n(x,y) \xrightarrow[n\to\infty]{} 0$$

car le potentiel est fini. Ainsi,

$$\begin{aligned}
0 &= \sum_{y\in C} \lim_{n\to\infty} P^n(x,y) \\
&= \lim_{n\to\infty} \sum_{y\in C} P^n(x,y) \\
&= \lim_{n\to\infty} \underbrace{\mathbb{P}_x(X_n \in C)}_{=1} \\
&= 1.
\end{aligned}$$

C'est impossible. Il existe donc un état récurrent, et ils le sont tous par irréductibilité. Remarquons maintenant que $n = \sum_{y\in C} N_n(y)$ lorsque la chaîne part de $x \in C$. On a donc

$$1 = \mathbb{E}_x\left[\sum_{y\in C} \frac{N_n(y)}{n}\right] = \sum_{y\in C} \frac{\mathbb{E}_x[N_n(y)]}{n}.$$

Or, $\frac{\mathbb{E}_x[N_n(y)]}{n} = \frac{G_n(x,y)}{n} \xrightarrow[n\to\infty]{} \frac{\rho_{x,y}}{m_y}$. Ainsi,

$$1 = \sum_{y\in C} \frac{\rho_{x,y}}{m_y} = \sum_{y\in C} \frac{1}{m_y}.$$

Il existe donc $y \in C$ tel que $m_y < \infty$, *i.e* que $y$ est positif.

□

**Corollaire 6.2.1.** *Soit $(X_n)_n$ une chaîne de* Markov *irréductible partant de $x \in E$. On a l'alternative suivante.*

— *Soit la chaîne est récurrente,*

$$\mathbb{P}_x(N(y) = \infty, \forall y \in E) = 1,$$

— *soit la chaîne est transitoire,*

$$\mathbb{P}_x(N(y) < \infty, \forall y \in E) = 1.$$

## 6.3 Classes de réccurence

**Proposition 6.3.1.** *Sur $E_R$, $\rightsquigarrow$ définit une relation d'équivalence.*

**Définition 6.3.1.** *On dit que deux états $x, y \in E$ communiquent (noté $x \sim y$) lorsque $x \rightsquigarrow y$ et $y \rightsquigarrow x$.*

*Démonstration.* Remarquons que sur $E_R$, la relation $x \rightsquigarrow y$ se réduit à $x \sim y$. On a bien $x \sim x$ car $G(x,x) = \infty$, on a donc bien $x \rightsquigarrow x$. Par définition, $\sim$ est réflexive. Enfin, la transitivité a déjà été démontrée.

□

**Remarque.** Cette relation induit une partition de $E_R$ en classes d'équivalence pour $\sim$ :

$$E_R = \bigsqcup_{i \in I} E_{R_i} = \left( \bigsqcup_{i \in I_+} E_{R_i} \right) \sqcup \left( \bigsqcup_{j \in I_0} E_{R_j} \right)$$

où $I_+$ désigne les indices des classes de récurrence positive, et $I_0$ les classes de récurrence nulle. On a donc

$$E = E_T \sqcup \left( \bigsqcup_{i \in I_+} E_{R_i} \right) \sqcup \left( \bigsqcup_{j \in I_0} E_{R_j} \right).$$

**Théorème 6.3.1.** *On a le comportement suivant sur une chaîne de* Markokv *partant de $x \in E$.*

*(i) Si $x \in E_{R_i}$ (où $i \in I$), alors $\mathbb{P}_x$-presque sûrement, $N(y) = \infty$ pour tout $y \in E_{R_i}$ et $N(y) = 0$ pour tout $y \in E_{R_i}^c$.*

*(ii) Si $x \in E_T$, en notant $T_{E_R} = \inf\{n \geq 0 \,|\, X_n \in E_R\}$, alors $\mathbb{P}_x$-presque sûrement,*

- *ou bien $T_{E_R} = \infty$ et $N(y) < \infty$ pour $\mathbb{P}_x$-presque tout $y \in E$ ;*
- *ou bien $T_{E_R} < \infty$, alors il existe $j \in I$ tel que pour tout $n \geq T_{E_R}$,*

$$X_n \in E_{R_j}.$$

*Démonstration.* $(i)$ Soit $x \in E_{R_i}$. Alors, $G(x, y) = 0$ pour tout $y \notin E_{R_i}$. En effet, si $y \in E_T$ alors $G(x, y) = 0$ et si $y \in E_R \backslash E_{R_i}$ alors $x$ et $y$ ne communiquent pas, et donc $G(x, y) = 0$. On en conclut que $N(y) = 0$ $\mathbb{P}_x$-presque sûrement. S $y \in E_{R_i}$, on a $\mathbb{P}_x(\widetilde{T}_y < \infty) = 1$. On a donc

$$\mathbb{P}_x(N(y) = \infty) = \mathbb{P}_x(\widetilde{T}_y < \infty)\mathbb{P}_y(N(y) = \infty) = 1$$

car $\mathbb{P}_x(\widetilde{T}_y < \infty) = 1$ puisque $x \rightsquigarrow y$ et $\mathbb{P}_y(N(y) = \infty) = 1$ puisque $y$ est récurrent.

$(ii)$ Si $x \in E_T$, alors $T_{E_R} = \infty$ ou $T_{E_R} < \infty$. Si $T_{E_R} = \infty$, alors la chaîne ne visite aucun $y \in E_R$, donc $N(y) = 0$ pour un certain $y \in E_R$. De plus, pour tout $y \in E_T$, on a $N(y) < \infty$.
Si $T_{E_R} < \infty$, il existe $j \in I$ tel que $T_{E_{R_j}} < \infty$. D'après la propriété de Markov et le premier point de l'énoncé, on a $N(y) = \infty$ pour tout $y \in E_{R_j}$ et la chaîne reste dans $E_{R_j}$ pour tout $n \geq T_{E_{R_j}}$.

□

**Exemple.** *Marche aléatoire simple sur $\mathbb{Z}$.* Une marche aléatoire sur $\mathbb{Z}$ est dite simple si la probabilité de passer de la marche $n$ à la marche $n+1$ est $p$, et de $n$ à $n-1$ est $1-p$. On a donc pour tout $x, y \in E = \mathbb{Z}$,

$$\begin{cases} p & \text{si } y = x+1\,; \\ 1-p & \text{si } y = x-1\,; \\ 0 & \text{sinon.} \end{cases}$$

Cette chaîne est irréductible puisque pour tout $x, y \in E$, $P^{|x-y|}(x, y) > 0$. Par irréductibilité de la chaîne, tous les états sont de même nature. On cherche la nature de l'état 0 en calculant $G(0, 0)$. On a

$$G(0,0) = \sum_{k=0}^{\infty} P^k(0,0) = \sum_{k=0}^{\infty} P^{2k}(0,0).$$

En effet, on ne peut revenir à un état qu'en un nombre pair d'étapes. De plus, $P^{2k}(0,0) = \binom{2k}{k} p^k(1-p)^k$ (un chemin de 0 à 0 se fait en choisissant $k$ pas à droite et $k$ pas à gauche, chemin uniquement déterminé par le nombre de pas à droite). On a donc

$$G(0,0) = \sum_{k=0}^{\infty} \frac{(2k)!}{(k!)^2}(p(1-p))^k.$$

Selon Stirling, on a

$$\frac{(2k)!}{(k!)^2}(p(1-p))^k \sim \frac{(4p(1-p))^k}{\sqrt{\pi k}}.$$

Si $p \neq 1/2$, alors $G(0,0) < \infty$ : la marche est transitoire. Si $p = 1/2$, alors $G(0,0) = \infty$ car $\sum_{k=1}^{\infty} \frac{1}{\sqrt{\pi k}} = \infty$, et la marche est récurrente [2]

## 6.4 Absorption dans les classes de récurrence

On pose $S_i = \inf\{n \geq 0 \,|\, X_n \in E_{R_i}\}$ la plus petite date d'absorption dans la classe de récurrence, et

$$\rho_i(x) = \mathbb{E}_x[\mathbb{1}_{\{S_i<\infty\}}] = \mathbb{P}_x(S_i < \infty)$$

la probabilité d'absorption. Enfin on pose $\tau_i(x) = \mathbb{E}_x[S_i \mathbb{1}_{\{S_i<\infty\}}]$ le temps moyen d'absorption. On a aussi $\mathbb{E}_x[S_i|S_i < \infty] = \frac{\tau_i(x)}{\rho_i(x)}$. On a facilement les faits suivants : pour tout $x \in E_{R_i}$, $S_i = 0$, $\rho_i(x) = 1$ et $\tau_i(x) = 1$ ; pour tout $x \in E_{R_j}$ $(j \neq i)$, $S_i = \infty$, $\rho_i(x) = 1$, et $\tau_i(x) = 0$.

**Théorème 6.4.1** (absorption)**.** *Soit $x \in E_T$. Les probabilités d'absorption $\rho_i(x)$ et les temps d'absorption $\tau_i(x)$ sont solutions des systèmes linéaires*

$$\rho_i(x) = \sum_{y\in E} P(x,y)\rho_i(y)$$

*et*

$$\tau_i(x) = \rho_i(x) + \sum_{y\in E} P(x,y)\tau_i(y).$$

*Démonstration.* Si $x \in E_T$, alors $S_i \geq 1$ $\mathbb{P}_x$-presque sûrement. De plus, on a $S_i = 1 + S_i \circ \Theta_1$. On a

2. Seulement lorsque $p = 1/2$, la chaîne est en équilibre et aucun côté (gauche ou droite) n'est préféré.

$$\begin{aligned}
\rho_i(x) &= \mathbb{P}_x(S_i < \infty) \\
&= \mathbb{E}_x[\mathbb{1}_{\{S_i<\infty\}}] \\
&= \mathbb{E}_x[\mathbb{1}_{\{1+S_i\circ\Theta_1<\infty\}}] \\
&= \mathbb{E}_x[\mathbb{E}_x[\mathbb{1}_{\{1+S_i\circ\Theta_1<\infty\}}|\mathcal{F}_i]] \\
&= \mathbb{E}_x[\underbrace{\mathbb{E}_{X_i}[\mathbb{1}_{\{S_i<\infty\}}]}_{p_i(X_1)}] \\
&= \sum_{y\in E} P(x,y)p_i(y).
\end{aligned}$$

De plus,

$$\begin{aligned}
\tau_i(x) &= \mathbb{E}_x[S_i\mathbb{1}_{\{S_i<\infty\}}] \\
&= \mathbb{E}_x[(1+S_i\circ\Theta_1)\mathbb{1}_{\{1+S_i\circ\Theta_1<\infty\}}] \\
&= \mathbb{E}_x[\mathbb{E}_x[(1+S_i\circ\Theta_1)\mathbb{1}_{\{1+S_i\circ\Theta_1<\infty\}}|\mathcal{F}_i]] \\
&= \mathbb{E}_x[\mathbb{E}_{X_i}[(1+S_i)\mathbb{1}_{\{S_i<\infty\}}]] \\
&= \mathbb{E}_x[\underbrace{\mathbb{E}_{X_i}[\mathbb{1}_{\{S_i<\infty\}}]}_{\rho_i(X_1)} + \underbrace{\mathbb{E}_{X_i}[S_i\mathbb{1}_{\{S_i<\infty\}}]}_{\tau_i(X_1)}] \\
&= \mathbb{E}_x[\rho_i(X_1)] + \mathbb{E}_x[\tau_i(X_1)] \\
&= \rho_i(x) + \sum_{y\in E} P(x,y)\tau_i(y).
\end{aligned}$$

□

# Chapitre 7

# Invariance et équilibre

## 7.1 Mesures invariantes

**Définition 7.1.1.** *Une mesure $\pi$ (positive) sur $E$ telle que $\pi(x) < \infty$ pour tout $x \in E$ est dite* ***invariante*** *pour une matrice stochastique $P$ si $\pi$ est solution de l'équation de* Chapman-Kolmogorov

$$\pi = \pi P$$

*(au sens du produit matriciel).*

Autrement dit, pour tout $y \in E$,

$$\pi(y) = \sum_{x \in E} \pi(x) P(x, y).$$

**Proposition 7.1.1.** *Soit $(X_n)_n$ une chaîne de* Markov*. La loi de $X_n$ ne dépend pas de $n$ si et seulement si la loi initiale $\mu_0$ est invariante.*

*Démonstration.* $\boxed{\Longrightarrow}$ On a $\mathscr{L}(X_1) = \mathscr{L}(X_0)$ donc $\mu_0 P = \mu_0$, *i.e* que $\mu_0$ est invariante.

$\boxed{\Longleftarrow}$ On a $\mathscr{L}(X_n) = \mu_n = \mu_0 P^n = \mu_0$ pour tout $n \in \mathbb{N}$. □

**Exemples.**

— Soit $E = \{0, 1\}$ et $P$ la matrice de transition

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}.$$

Alors, $(q, p)$ est une mesure invariante pour $P$, et $\left(\frac{q}{p+q}, \frac{p}{p+q}\right)$ est une probabilité invariante.

— *Marche aléatoire symétrique sur $\mathbb{Z}$.* On considère une marche aléatoire sur $\mathbb{Z}$ avec probabilité $1/2$ de passer de $x$ à $x+1$ et probabilité $1/2$ de passer de $x$ à $x-1$. Une mesure invariante $\pi$ vérifie

$$\pi(y) = \sum_{y \in E} \pi(x) P(x, y) = \frac{\pi(y-1) + \pi(y+1)}{2},$$

*i.e* $\pi(y+1) - \pi(y) = \pi(y) - \pi(y-1)$. On en déduit que la mesure est invariante *si et seulement si* elle est de la forme $\pi(y) = \alpha y + \beta$. Nécessairement $\alpha = 0$ (car si $\alpha > 0$ alors $\pi$

devient négative lorsque $y$ tend vers $-\infty$,si $\alpha < 0$ alors $\pi$ devient négative lorsque $y$ tend vers $\infty$). On en déduit qu'une telle mesure invariante $\pi$ est une mesure uniforme, mais ne peut être une probabilité. Cette chaîne de Markov n'a pas de probabilité invariante. Remarquons par ailleurs que pour une matrice bistochastique, les mesures uniformes sont invariantes (par définition d'une matrice stochastique).

— $E = \mathbb{N}$. On considère la matrice stochastique telle que $P(x, x+1) = 1$ pour tout $x \in E$. Soit $\pi$ une mesure invariante pour $P$ donc telle que $\pi(y) = \sum_{x \in E} \pi(x)P(x, y)$. Si $y > 0$, $\pi(y) = \pi(y-1)$ et $\pi(0) = 0$. Nécessairement, $\pi = 0$. Il n'y a donc pas de mesure invariante dans ce système.

— $E = \mathbb{N}$. On considère la dynamique markovienne de matrice stochastique définie par $P(n, n+1) = p_n$ et $P(n, 0) = 1 - p_n$ pour tout $n \in \mathbb{N}$ où $0 < p_n < 1$ et $\sum_{i=1}^{\infty}(1 - p_i)$. Cette dynamique est irréductible car il est clair que tout $n \in \mathbb{N}$ mène à tout $m \in \mathbb{N}$. Une mesure invariante $\pi$ de ce système est solution de $\pi(y) = \sum_{x \in E} \pi(x)P(x, y)$ pour tout $y \in E$. On a $\pi(0) = \sum_{i=1}^{\infty}(1 - p_i)\pi_i$ et $\pi(i) = p_{i-1}\pi(i-1)$ pour tout $i > 0$. On a donc $\pi(i) = \pi(0) \prod_{j=1}^{i-1} p_j$, et alors

$$\pi(0) = \pi(0) \sum_{i=1}^{\infty} \left( (1 - p_i) \prod_{j=1}^{i-1} p_j \right).$$

Or, si $0 < \pi(0) < \infty$, on a

$$1 = \sum_{i=1}^{\infty}(1 - p_i) \underbrace{\prod_{j=1}^{i-1} p_j}_{\leq 1} \leq \sum_{i=1}^{\infty}(1 - p_i) < 1.$$

On en déduit que pour cette chaîne de Markov (irréductible), il n'y a pas de mesure invariante.

**Définition 7.1.2.** *Une mesure positive $\pi$ non nulle sur $E$ avec $\pi(x) < \infty$ pour tout $x \in E$ est dite* ***réversible*** *pour $P$ si pour tout $x, y \in E$,*

$$\pi(x)P(x, y) = \pi(y)P(y, x).$$

**Remarque.** Par récurrence immédiate, pour tout $x_0, \ldots, x_n \in E$, on a

$$\pi(x_0)P(x_0, x_1) \cdots P(x_{n-1}, x_n) = \pi(x_n)P(x_n, x_{n-1}) \cdots P(x_1, x_0).$$

**Proposition 7.1.2.** *Une probabilité $\pi$ est réversible pour $P$ si et seulement si pour toute chaîne de* Markov *de noyau $P$ de loi initiale $\pi$, on a*

$$\mathscr{L}(X_0, X_1, \ldots, X_n) = \mathscr{L}(X_n, X_{n-1}, \ldots, X_0).$$

**Proposition 7.1.3.** *Une mesure $\pi$ réversible est invariante.*

*Démonstration.* On a pour tout $y \in E$,

$$(\pi P)(y) = \sum_{x \in E} \pi(x)P(x, y) = \sum_{x \in E} \pi(y)P(y, x) = \pi(y) \sum_{x \in E} P(y, x) = \pi(y).$$

□

**Exemples.**

— *Marche aléatoire sur* $\mathbb{Z}$. On considère la dynamique markovienne définie par

$$P(x,y) = \begin{cases} p & \text{si } y = x+1\,; \\ 1-p & \text{si } y = x-1\,; \\ 0 & \text{sinon.} \end{cases}$$

Alors, la mesure définie $\pi(i) = \left(\frac{p}{1-p}\right)^i$ est réversible.

— *Urne d'*EHRENFEST. On considère sur l'espace d'états $E = [\![0,d]\!]$ la dynamique markovienne de noyau

$$P(x,y) = \begin{cases} \frac{d-x}{d} & \text{si } y = x+1\,; \\ \frac{x}{d} & \text{si } y = x-1\,; \\ 0 & \text{sinon.} \end{cases}$$

Si $\pi$ est une mesure réversible pour $P$, alors

$$\pi(x)P(x,x+1) = \pi(x+1)P(x+1,x)$$

donc $\pi(x+1) = \frac{d-x}{x+1}\pi(x)$ pour tout $x \in [\![0,d]\!]$. On en déduit que $\pi(x) = \binom{d}{x}\pi(0)$. On peut normaliser $\pi$ et considérer la probabilité invariante définie par $\pi(x) = \binom{d}{x}2^{-d}$ donc $\pi \sim \mathcal{B}(d,1/2)$.

## 7.2 Invariance et récurrence

### 7.2.1 Théorème

**Lemme 7.2.1** (convergence dominée)**.** *Soit* $(a(x))_{x\in E}$ *une suite de réels positifs sommable et* $(b_n(x))_{x\in E}$ *une suite telle que* $b_n(x) \le 1$ *pour tout* $x \in E$ *et* $n \in \mathbb{N}$. *Et on suppose que*

$$b_n(x) \xrightarrow[n\to\infty]{} b(x)$$

*pour tout* $x \in E$. *Alors,*

$$\sum_{x\in E} a(x)b_n(x) \xrightarrow[n\to\infty]{} \sum_{x\in E} a(x)b(x).$$

*Démonstration.* Tout est dans le titre du lemme.

□

**Théorème 7.2.1** (support d'une mesure invariante)**.** *Soit* $\pi$ *une mesure invariante pour une chaîne de* MARKOV *de matrice stochastique* $P$.

— *Si* $x \in E$ *est tel que* $\pi(x) > 0$, *alors on a aussi* $\pi(y) > 0$ *pour tout* $y \in E$ *tel que* $x \rightsquigarrow y$.

— *Si la chaîne est irréductible, le support de* $\pi$ *et* $E$.

— *Si* $\pi$ *est une probabilité, alors* $\pi(x) = 0$ *dès que* $x$ *est transitoire ou récurrent nul. Une probabilité invariante ne charge que les états récurrents positifs. Le support d'une probabilité invariante est une union de classes de récurrence positives.*

*Démonstration.* Si $x \in E$, $\pi(x) > 0$, et $x \rightsquigarrow y$, alors il existe $n \geq 1$ tel que $P(x,y) > 0$. On a $\pi = \pi P^n$, donc

$$\pi(y) = \sum_{z \in E} \pi(z) P^n(z,y) \geq \pi(x) P^n(x,y) > 0.$$

Si la chaîne est irréductible, tous les états communiquent. La mesure étant non nul, au moins un état est charge la mesure $\pi$.

Pour le dernier point, supposons que $\pi$ est une probabilité et $x$ est transitoire ou récurrent nul. Alors, $m_x = \mathbb{E}_x[\widetilde{T}_x] = \infty$. On a vu que pour tout $z \in E$,

$$\frac{G_n(z,x)}{n} \xrightarrow[n\to\infty]{} \frac{\rho_{z,x}}{m_x} = 0$$

(car $m_x = \infty$). Par invariance de $\pi$, $\pi = \pi P^k$ pour tout $k \geq 1$. On a alors

$$\pi(x) = \sum_{z \in E} P^k(z,x).$$

En sommant sur $k \in [\![1,n]\!]$ et en divisant par $n$,

$$\pi(x) \sum_{z \in E} \pi(z) \frac{G_n(z,x)}{n} \xrightarrow[n\to\infty]{} 0$$

selon le lemme précédent. Ainsi, $\pi(x) = 0$ pour tout $x \in E_T \cup E_{R_0}$.

$\square$

**Exemple.** On considère une marche aléatoire sur $\mathbb{Z}$ avec $P(n,n+1) = p$ et $P(n,n-1) = 1-p$ où $p \in ]0,1[$.

- Si $p \neq 1/2$, la marche est transitoire et $\pi = \left(\left(\frac{p}{1-p}\right)^x\right)_{x\in\mathbb{Z}}$ est invariante. Ici, $\pi$ est de poids infini. Pour cette marche, il n'y a pas de probabilité invariante.
- Si $p = 1/2$, la marche est récurrente, et $\pi$ est uniforme et invariante de poids infini. Il n'y a donc pas de probabilité invariante.

### 7.2.2 Construction de mesures invariantes

Soit $x \in E$ un état récurrent. Alors, $\widetilde{T}_x$ est fini $\mathbb{P}_x$-presque sûrement. On pose

$$\nu_x(y) = \mathbb{E}_x\left[\sum_{k=0}^{\widetilde{T}_x - 1} \mathbb{1}_{\{X_k = y\}}\right] = \mathbb{E}_x\left[\sum_{k=1}^{\widetilde{T}_x} \mathbb{1}_{\{X_k = y\}}\right]$$

le nombre moyen de passage en $y$ avec l'arrivée en $x$.

**Proposition 7.2.1.** *Soit $x \in E_R$. Alors, $\nu_x$ est une mesure invariante, et avec $\nu_x(x) = 1$ de support la classe de récurrence de $x$ (i.e $\nu_x(y) > 0$ si et seulement si $x \rightsquigarrow y$).*

*Démonstration.* On a $\nu_x(x) = 1$ car $\sum_{k=0}^{\widetilde{T}_x - 1} \mathbb{1}_{\{X_k = x\}} = 1$ $\mathbb{P}_x$-presque sûrement. Si $y$ n'est pas dans la classe de $x$,

$$0 = G(x,y) = \mathbb{E}_x\left[\sum_{k=0}^{\infty} \mathbb{1}_{\{X_k = y\}}\right] \geq \mathbb{E}_x\left[\sum_{k=0}^{\widetilde{T}_x - 1} \mathbb{1}_{\{X_k = y\}}\right] = \nu_x(y).$$

Soit $y \in E$, on a

$$\begin{aligned}
\nu_x(y) &= \mathbb{E}_x\left[\sum_{k=1}^{\widetilde{T}_x} \mathbb{1}_{\{X_k=y\}}\right] \\
&= \mathbb{E}_x\left[\sum_{z\in E}\sum_{k=1}^{\infty} \mathbb{1}_{\{\widetilde{T}_x\geq k\}}\mathbb{1}_{\{X_k=y\}}\mathbb{1}_{\{X_{k-1}=z\}}\right] \\
&= \sum_{z\in E}\sum_{k=1}^{\infty}\mathbb{E}_x\left[\mathbb{E}_x\left[\mathbb{1}_{\{\widetilde{T}_x\geq k\}}\mathbb{1}_{\{X_k=y\}}\mathbb{1}_{\{X_{k-1}=z\}}|\mathcal{F}_{k-1}\right]\right]
\end{aligned}$$

Permettez moi d'interrompre cette démonstration pour observer quelques secondes le tableau que j'essaye de copier ici.

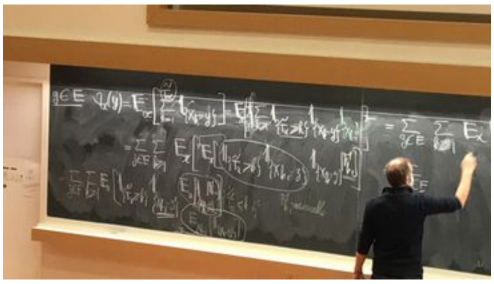


FIGURE 7.1 – Jean-Christophe BRETON essayant de communiquer. 2021, colorized.

□

**Lemme 7.2.2.** *Soit $x \in E$ un état récurrent et $\pi$ une mesure invariante. Pour tout $y \in E$,*

$$\pi(y) \geq \pi(x)\nu_x(y).$$

*De plus, si $x \sim y$ alors il y a égalité.*

*Démonstration.* Admis pour mon plus grand plaisir.

□

**Théorème 7.2.2.** *On considère une classe de récurrence $E_{R_i}$ (donc close irréductible). Il y a unicité à multiple près d'une mesure invariante concentrée sur cette classe (i.e dont le support est $E_{R_i}$). De plus,*

— *Si ces mesures sont finies, alors il y a unique probabilité invariante sur $E_{R_i}$. Cette probabilité est donnée pour tout $x \in E$ par*

$$\pi(x) = \frac{1}{\mathbb{E}_x[\widetilde{T}_x]}.$$

*Dans ce cas, la classe est récurrente positive.*

— *Si ces mesures sont infinies, il n'y a pas de probabilité invariante sur $E_{R_i}$ et la classe est récurrente nulle.*

*Démonstration.* Soit $x \in E_{R_i}$. Cet état est donc récurrent et on considère la mesure invariante $\nu_x$ qui lui est associé (de support $E_{R_i}$). D'après le lemme, on sait que pour tout $y \in E_{R_i}$, $x \rightsquigarrow y$

et $\pi(y) = \pi(x)\nu_x(y)$. Puisque $\nu_x$ et $\pi$ ont pour support $E_{R_i}$, on a $\pi = \pi(x)\nu_x$. On étudie deux cas.

— Si ces mesures sont de poids fini, il existe une unique probabilité invariante notée $\pi$ telle que $\pi = \pi(x)\nu_x$. On a donc $1 = \pi(E) = \pi(x)\nu_x(E)$ donc $\pi(x) > 0$ et $\pi(x) = \frac{1}{\nu_x(E)}$. Cela signifie que $x$ est récurrent positif. Ainsi, $E_{R_i}$ est une classe récurrente positive. On a

$$\begin{aligned}\nu_x(E) &= \sum_{y\in E} \nu_x(y) \\ &= \sum_{y\in E} \mathbb{E}_x\left[\sum_{k=0}^{\widetilde{T}_x-1} \mathbb{1}_{\{X_k=y\}}\right] \\ &= \mathbb{E}_x\left[\sum_{k=0}^{\widetilde{T}_x-1}\sum_{y\in E} \mathbb{1}_{\{X_k=y\}}\right] \\ &= \mathbb{E}_x[\widetilde{T}_x] = m_x.\end{aligned}$$

— Si ces mesures sont infinies, il n'y a pas de probabilité. De plus, pour tout $x \in E_{R_i}$,

$$\infty = \nu_x(E) = m_x.$$

Ainsi, $x$ est récurrent nul et la classe donc récurrente nulle.

□

### 7.2.3 Cas d'une chaîne irréductible.

Dans ce cas, tous les états sont de même nature et il y a lors trois situations.

— La chaîne est transitoire et toute mesure invariante (s'il y en a) est de poids infini. Il n'y a pas de probabilité invariante.

— La chaîne est récurrente nulle. Les mesures invariantes sont toutes proportionnelles et de poids infinis. Il n'y a pas de probabilité invariante.

— La chaîne est récurrente positive. Les mesures invariantes sont toutes proportionnelles de poids finis. Il y a une unique probabilité $\pi$ donnée pour tout $x \in E$ par

$$\pi(x) = \frac{1}{\mathbb{E}_x[\widetilde{T}_x]}.$$

**Corollaire 7.2.1.** *Pour une chaîne de* MARKOV *irréductible, on a équivalence entre les assertions suivantes.*

*(i) Il existe une unique probabilité invariante.*

*(ii) La mesure définie par $\pi(x) = \frac{1}{\mathbb{E}_x[\widetilde{T}_x]}$ pour tout $x \in E$ est une probabilité invariante.*

*(iii) Il existe un état récurrent positif.*

*(iv) Tous les états sont récurrents positifs.*

**Exemple.** On considère maintenant une marche aléatoire sur $\mathbb{N}$ avec $P(n, n+1) = p$ pour $n \geq 1$, $P(n, n-1) = q = 1 - p$ pour tout $n \geq 2$, et $P(0, 1) = 1$. La mesure définie par $\pi(0) = q$ et $\pi(n) = \left(\frac{p}{q}\right)^{n-1}$ pour tout $n \geq 1$, est réversible. Si $p < q$, $\pi$ est finie et la chaîne est récurrente positive. Si $p \geq q$, $\pi$ est de poids infini et la chaîne est récurrente nulle ou transitoire.

### 7.2.4 Cas d'une chaîne non irréductible.

On a toujours

$$E = E_T \sqcup \left( \bigsqcup_{i \in I_+} E_{R_i} \right) \sqcup \left( \bigsqcup_{j \in I_0} E_{R_j} \right)$$

et $E_R = \left( \bigsqcup_{i \in I_+} E_{R_i} \right) \sqcup \left( \bigsqcup_{j \in I_0} E_{R_j} \right)$.

**Proposition 7.2.2.** *Pour une chaîne non irréductible, on a les propriétés suivantes.*

- *Sur chaque classe de récurrence, il existe une mesure invariante (unique à un facteur multiplicatif près).*
- *Sur une classe de récurrence, il existe une unique probabilité invariante si et seulement si elle récurrente positive. Dans ce cas, elle est donnée pour tout $x$ dans cette classe par*

$$\pi(x) = \frac{1}{\mathbb{E}_x[\widetilde{T}_x]}.$$

- *L'ensemble des probabilités est donné par les combinaisons convexes des probabilités invariantes de chaque classe récurrente positive.*

*Démonstration.* Les deux premiers points découlent de résultats précédents. On montre que toute probabilité invariante $\pi$ est combinaison convexe des uniques probabilités invariantes de chaque classe de récurrence positive. Soit $E_{R_i}$ une classe de récurrence positive telle que $\pi(E_{R_i}) > 0$, donc il existe un état dans $E_{R_i}$ chargé par la mesure pour $\pi$. On en déduit que pour tout $x \in E_{R_i}$, $\pi(x) > 0$. De plus, $\pi_{|E_{R_i}}$ est toujours invariante car pour tout $y \in E_{R_i}$,

$$\pi(y) = \sum_{z \in E} \pi(z) P(z,y) = \sum_{z \in E_{R_i}} P(z,y).$$

En effet, $\pi(z) = 0$ si $z$ n'est pas récurrent positif, et seuls les $z \in E_{R_i}$ sont tels que $P(z,y) > 0$. $\pi_{|E_{R_i}}$ est proportionnelle à $\nu_x$ et $\pi_{E_{R_i}} = \pi(x)\nu_x = \pi(x)\nu_x(E_{R_i})\tilde{\nu}_{x,i}$, où $\tilde{\nu}_{x,i} = \nu_x/\nu_x(E_{R_i})$ est la mesure normalisée de sorte que $\tilde{\nu}_{x,i}$ soit une probabilité sur $E_{R_i}$. On a

$$\pi = \sum_{i \in I_+} \pi_{|E_{R_i}} = \sum_{i \in I_+} \pi(x)\nu_x(E_{R_i})\tilde{\nu}_{x,i}.$$

C'est bien une combinaison convexe de probabilités invariantes chacune sur une classe de récurrence positive

□

## 7.3 Asymptotique d'une chaîne de MARKOV

### 7.3.1 Périodicité

**Définition 7.3.1.** *On appelle période de l'état $x \in E$ pour une chaîne de MARKOV de transition $P$ l'entier*

$$d_x = \operatorname{pgcd}\{n \geq 1 \mid P^n(x,x) > 0\}$$

*avec pour contient $d_x = 0$ si $P^n(x,x) = 0$ pour tout $n \geq 1$. Si $d_x = 1$, on dit que $x$ est apériodique.*

**Remarque.** On considère sur $\mathbb{Z}$ la marche aléatoire de matrice stochastique $P$ telle que $P(n, n+1) = p \in ]0,1[$ et $P(n, n-1) = 1-p$. Pour tout $x \in \mathbb{Z}$, $d_x = 2$. Cela est dû à un fait plus général.

**Proposition 7.3.1.** *Si $x \rightsquigarrow y$, alors $d_x = d_y$.*

*Démonstration.* Si $x \rightsquigarrow y$, il existe $n, m \geq 1$ tels que $P^n(x,y) > 0$ et $P^m(y,x) > 0$. Alors, pour tout $N, k \in \mathbb{N}$,

$$P^{m+n-Nk}(x,x) \geq P^n(x,y)P^k(y,y)^N P^m(y,x).$$

Pour $k \geq 1$ tel que $P^k(y,y) > 0$, alors $P^{m+n+Nk}(x,x) > 0$. Ainsi, $d_x$ divise $m+n+Nk$ pour tout $N \geq 1$. Cela signifie donc que $d_x$ divise $k$ mais $P^k(y,y) > 0$. On a donc montré que $d_x$ divise $d_y$. On conclut par symétrie des rôles.

□

**Définition 7.3.2.** *Si une chaîne de MARKOV est irréductible, on appelle période la chaîne la période de tous ses états. Lorsque cette période est de 1, la chaîne est dite apériodique.*

**Remarque.** Soit $P$ une matrice stochastique et $p \in ]0,1[$. La matrice stochastique perturbée $P_p = (1-p)P + pI$ avec $I = (\delta_{x,y})_{x,y}$. Perturber de cette façon la matrice $P$ consiste à rajouter à chaque état une boucle. $P_p$ est une matrice de transition apériodique.

**Théorème 7.3.1** (convergence à l'équilibre)**.** *On considère une chaîne de MARKOV irréductible récurrente positive et apériodique. Alors, si $\pi$ désigne l'unique probabilité invariante. pour tout $x \in E$,*

$$\lim_{n\to\infty} \sum_{y\in E} |\mathbb{P}_x(X_n = y) - \pi(y)| = 0.$$

*On a aussi pour la loi $\nu$ initiale,*

$$\lim_{n\to\infty} \sum_{y\in E} |\mathbb{P}_\nu(X_n = y) - \pi(y)| = 0.$$

**Remarques.**

— On a $P^n(x,y) = \mathbb{P}_x(X_n = y) \xrightarrow[n\to\infty]{} \pi(y)$ pour tout $y \in E$ (et la convergence est uniforme en $y$) : $(\mathscr{L}(X_n|X_0 = x))_n$ converge en loi vers la loi $\pi$. On en déduit que $(P^n)_n$ tend vers une matrice dont les lignes sont toutes $\pi$.

— En fait, on a la convergence en variation totale de $(\mathscr{L}(X_n|X_0 = x))_n$ vers la mesure invariante $\pi$.

*Démonstration.* C'est long. Cependant, on peut montrer le second résultat. On a $\mathbb{P}_\nu = \sum_{x\in E} \nu(x)\mathbb{P}_x$. On a donc

$$\begin{aligned}\mathbb{P}_\nu(X_n = y) - \pi(y) &= \sum_{x\in E} \nu(x)\mathbb{P}(X_n = y) - \left(\sum_{x\in E} \nu(x)\pi(y)\right)\\ &= \sum_{x\in E} \nu(x)(\mathbb{P}_x(X_n = y) - \pi(y)).\end{aligned}$$

On a donc

$$\begin{aligned}\sum_{y\in E} |\mathbb{P}_\nu(X_n = y) - \pi(y)| &\leq \sum_{y\in E}\sum_{x\in E} |\mathbb{P}_x(X_n = y) - \pi(y)|\\ &= \sum_{x\in E}\left(\sum_{y\in E} |\mathbb{P}_x(X_n = y) - \pi(y)|\right)\nu(x).\end{aligned}$$

Soit alors $a(x) = \nu(x)$ et $b_n(x) = \sum_{y\in E} |\mathbb{P}_x(X_n = y) - \pi(y)|$. On a bien $\sum_{x\in E} |b_n(x)| \leq 2$ et alors

$$\sum_{x\in E} a(x)b_n(x) \xrightarrow[n\to\infty]{} \sum_{x\in E} a(x) \times 0 = 0.$$

□

**Théorème 7.3.2** (convergence d'une chaîne périodique)**.** *Soit $(X_n)_n$ une chaîne irréductible récurrente positive et périodique de période $d$. On note $\pi$ l'unique probabilité invariante. Alors, pour toute pair d'états $x \neq y$, il existe $r \in [\![0, d-1]\!]$ un entier tel que $P^n(x,y) = 0$ si $n \not\equiv r \mod d$. Si $(\varphi(n))_n$ désigne la suite croissante des entiers valant $r$ modulo $d$, alors*

$$P^{\varphi(n)}(x,y) \xrightarrow[n\to\infty]{} d\pi(y).$$

### 7.3.2 Théorème ergodique

**Théorème 7.3.3** (ergodique)**.** *Soit $(X_n)_n$ une chaîne de* Markov *irréductible récurrente, et soit $\pi$ une mesure invariante. Soient $f, g \in L^1(\pi)$ avec $\int_E g \,\mathrm{d}\pi = \sum_{x\in E} g(x)\pi(x) \neq 0$. Alors, pour tout $x \in E$,*

$$\lim_{n\to\infty} \frac{\sum_{k=1}^n f(X_k)}{\sum_{k=1}^n g(X_k)} = \frac{\int_E f \,\mathrm{d}\pi}{\int_E g \,\mathrm{d}\pi}$$

*$\mathbb{P}_x$-presque sûrement. Pour n'importe quelle loi initiale, on a le même résultat $\mathbb{P}_\nu$-presque sûrement.*

**Corollaire 7.3.1** (ergodique)**.** *Soit $(X_n)_n$ une chaîne de* MARKOV *irréductible récurrente positive, d'unique probabilité invariante $\pi$. Pour tout $f \in L^1(\pi)$, on a pour tout $x \in E$ (resp. pour toute loi initiale $\nu$),*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} f(X_k) = \int_E f \mathrm{d}\pi$$

*$\mathbb{P}_x$-presque sûrement (resp. $\mathbb{P}_\nu$-presque sûrement).*

*Démonstration.* Prendre $g$ constante à 1 dans le théorème ergodique. □

*Démonstration (du théorème).* La seconde partie découle de

$$\mathbb{P}_\nu(A) = \sum_{x \in E} \nu(x) \mathbb{P}_x(A).$$

Si pour tout $\in E$, $\mathbb{P}_x(A) = 1$, alors $\mathbb{P}_\nu(A) = 1$. Un évènement $\mathbb{P}_x$-sûre pour tout $x \in E$ est $\mathbb{P}_\nu$ sûre pour toute loi initiale $\nu$. Pour la première partie, on fixe $x \in E$. On observe que $\pi(x) > 0$ et on dispose de la mesure invariante $\nu_x$. Par unicité à un facteur près, il existe $\alpha > 0$ tel que $\pi = \alpha \nu_x$ (en fait, $\alpha = \pi(x)$ car $\nu_x(x) = 1$). On suppose que la chaîne part de $x$ et on note $T_x^{(k)}$ les dates de retours successifs en $x$ : $T_x^{(0)} = 0$ et $T_x^{(n+1)} = \inf\{k > T_x^{(n)} \mid X_k = x\}$. Puisque $x$ est récurrent, les $T_x^{(n)}$ sont finis $\mathbb{P}_x$-presque sûrement. On note

$$Z_k(f) = \sum_{i=T_x^{(k-1)}}^{T_x^{(k)}-1} f(X_i)$$

pour tout $k \geq 1$. Montrons avant tout que les $Z_k(f)$ sont i.i.d lorsque $k$ parcourt $\mathbb{N}^*$. Soit $(g_i)_{1 \leq i \leq k}$ une famille de fonctions mesurables bornées. On montre par récurrence sur $k$ que

$$\mathbb{E}_x \left[ \prod_{i=1}^{k} g_i(Z_i(f))) \right] = \prod_{i=1}^{k} \mathbb{E}_x \left[ g_i(Z_1(f)) \right].$$

Pour $k = 1$, c'est immédiat. On suppose le résultat vrai au rang $k$ et on l'établit au rang $k+1$. On observe que $Z_1(f), \ldots, Z_k(f)$ sont $\mathcal{F}_{T_x^{(k)}}$-mesurables, et $Z_{k+1}(f) = Z_1(f) \circ \Theta_{T_x^{(k)}}$. On a

$$\begin{aligned}
\mathbb{E}_x \left[ \prod_{i=1}^{k+1} g_i(Z_i(f)) \right] =& \mathbb{E}_x \left[ \left( \prod_{i=1}^{k} g_i(Z_i(f)) \right) \times g_{k+1}(Z_1(f) \circ \Theta_{T_x^{(k)}} w) \right] \\
&= \mathbb{E}_x \left[ \left( \prod_{i=1}^{k} g_i(Z_i(f)) \right) \mathbb{E}_x \left[ g_{k+1}(Z_1(f) \circ \Theta_{T_x^{(k)}}) \middle| \mathcal{F}_{T_x^{(k)}} \right] \right] \\
&= \mathbb{E}_x \left[ \prod_{i=1}^{k} g_i(Z_i(f)) \right] \times \mathbb{E}_{X_{T_x^{(k)}}} \left[ g_{k+1}(Z_1(f)) \right].
\end{aligned}$$

On conclut en remarquant que $X_{T_x^{(k)}} = x$. Pour appliquer la loi des grands nombres aux variables aléatoires $Z_k(f)$ $(k \geq 1)$, on montre que $Z_1(f) \in L^1(\mathbb{P}_x)$. On a

$$
\begin{aligned}
\mathbb{E}_x[|Z_1(f)|] &= \mathbb{E}_x\left[\left|\sum_{k=0}^{T_x^{(1)}-1} f(X_k)\right|\right] \\
&\leq \mathbb{E}_x\left[\sum_{k=0}^{T_x^{(1)}-1} |f(X_k)|\right] \\
&= \mathbb{E}_x\left[\sum_{k=0}^{T_x^{(1)}-1} \sum_{y\in E} |f(y)|\, \mathbb{1}_{\{X_k=y\}}\right] \\
&= \sum_{y\in E} \mathbb{E}_x\left[\sum_{k=0}^{T_x^{(1)}-1} \mathbb{1}_{\{X_k=y\}}\right] |f(y)| \\
&= \sum_{y\in E} |f(y)| \frac{\pi(y)}{\pi(x)} \\
&= \frac{\int_E |f|\, \mathrm{d}\pi}{\pi(x)} < \infty.
\end{aligned}
$$

Le même calcul sans valeurs absolues montre que $\mathbb{E}_x[Z_1(f)] = \frac{\int_E |f|\, \mathrm{d}\pi}{\pi(x)}$. En appliquant la loi des grands nombres,

$$
\frac{1}{n}\sum_{k=1}^{n} Z_k(f) \xrightarrow[n\to\infty]{\mathbb{P}_x-p.s} \frac{\int_E f\, \mathrm{d}\pi}{\pi(x)}.
$$

On s'intéresse à $\frac{1}{N}\sum_{k=1}^{N} f(X_k)$. J'arrête cette démonstration subitement puisque le tableau a commencé à ressembler à Guernica.

$\square$