Polycopié pour la partie “Probabilités” pour le cours MOPSI
(MOdéliser Programmer SImuler)

**Aurélien ALFONSI**

**Avertissement :** Ce polycopié est en cours d'écriture, et il contient probablement encore un bon nombre d'erreurs de tout ordre. Il faut donc le lire avec un certain esprit critique ! Toute remarque de lecteur est la bienvenue à l'adresse suivante poly.proba.mopsi@gmail.com

# Table des matières

# Chapitre 1

# Rappels de probabilités

## 1.1 Rudiments de théorie de la mesure

### 1.1.1 Espaces mesurables

L'objet de cette partie est de rappeler quelques notions de base de la théorie de la mesure, et de l'intégration.

**Définition 1.1.1.** *Soit $E$ un ensemble et $\mathcal{E} \subset \mathcal{P}(E)$ un ensemble de parties de $E$. On dit que $\mathcal{E}$ est une tribu ou $\sigma$-algèbre de $E$ si*

*i)* $\emptyset, E \in \mathcal{E}$,

*ii)* $A \in \mathcal{E} \iff E \backslash A \in \mathcal{E}$,

*iii) si* $(A_i)_{i\in\mathbb{N}} \in \mathcal{E}^{\mathbb{N}}$ *est une famille dénombrable d'éléments de* $\mathcal{E}$, $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{E}$.

En passant au complémentaire (propriété *ii*)), on obtient une définition équivalente en remplaçant *iii*) par :
*iii'*) si $(A_i)_{i\in\mathbb{N}} \in \mathcal{E}^{\mathbb{N}}$ est une famille dénombrable d'éléments de $\mathcal{E}$, $\bigcap_{i\in\mathbb{N}} A_i \in \mathcal{E}$.

Lorsqu'un ensemble $E$ est muni d'une tribu $\mathcal{E}$, on dit que $(E, \mathcal{E})$ est un *espace mesurable.* Si $\mathcal{E}$ et $\mathcal{E}'$ sont deux tribus de $E$ telles que $\mathcal{E}' \subset \mathcal{E}$, on dit que $\mathcal{E}'$ est une *sous-tribu* de $\mathcal{E}$. Voici deux exemples triviaux de tribus de $E$ : $\mathcal{P}(E)$ et $\{\emptyset, E\}$. Elles sont respectivement maximale et minimale au sens suivant : pour toute tribu $\mathcal{E}$ de $E$, $\mathcal{E}$ est une sous-tribu de $\mathcal{P}(E)$ (resp. $\{\emptyset, E\}$ est une sous-tribu de $\mathcal{E}$).

La proposition suivante est une conséquence immédiate de la définition d'une tribu.

**Proposition 1.1.2.** *Soient $E$ et $\Theta$ deux ensembles. Si $(\mathcal{E}_\theta)_{\theta\in\Theta}$ est une famille de tribus de $E$, $\bigcap_{\theta\in\Theta} \mathcal{E}_\theta$ est une tribu de $E$.*

**Corollaire 1.1.3.** *Pour $\mathcal{A} \subset \mathcal{P}(E)$, on définit*

$$\sigma(\mathcal{A}) = \bigcap_{\mathcal{E} \text{ tribu t.q.} \mathcal{A} \subset \mathcal{E}} \mathcal{E}.$$

*C'est la plus petite tribu de $E$ contenant $\mathcal{A}$.*

Lorsqu'on a deux espaces mesurables $(E_1, \mathcal{E}_1)$ et $(E_2, \mathcal{E}_2)$, on définit la *tribu produit* $\mathcal{E}_1 \otimes \mathcal{E}_2$ comme étant la plus petite tribu engendrée par $\mathcal{E}_1 \times \mathcal{E}_2$. On appelle $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ espace mesurable produit.

**Définition 1.1.4.** *Soient $E$ un ensemble, $(F, \mathcal{F})$ un espace mesurable et $g : E \to F$ une application. Alors,*

$$\sigma(g) = \sigma(\{g^{-1}(A),\ A \in \mathcal{F}\})$$

*est une tribu sur $E$ appelée tribu engendrée par $g$. (Rappelons ici que $g^{-1}(A) = \{x \in F, g(x) \in A\}$ est la préimage par $g$ de l'ensemble $A$.)*

Plus généralement, si l'on a une famille de fonctions $g_\theta : E \to F_\theta$ à valeurs dans des espaces mesurables $(F_\theta, \mathcal{F}_\theta)$, on désigne par $\sigma(g_\theta, \theta \in \Theta)$ la tribu engendrée par les fonctions $g_\theta$, c'est à dire : $\sigma(\{g_\theta^{-1}(A), \theta \in \Theta,\ A \in \mathcal{F}_\theta\})$.

**Définition 1.1.5.** *Soient $(E, \mathcal{E})$ et $(F, \mathcal{F})$ deux espaces mesurables. Une application $g : E \to F$ est dite mesurable si*

$$\forall A \in \mathcal{F},\ g^{-1}(A) \in \mathcal{E}.$$

Avec cette définition, il est aisé de voir que la plus petite tribu sur $E$ qui rende $g$ mesurable est $\sigma(g)$. On obtient aussi facilement que la composition de deux fonctions mesurables est mesurable. Par ailleurs, si $g_1 : E_1 \to F_1$ et $g_2 : E_2 \to F_2$ sont mesurables, alors $(g_1, g_2) : E_1 \times E_2 \to F_1 \times F_2$ est mesurable par rapport aux tribus produit.

Nous allons désormais introduire la notion de tribu borélienne. Rappelons que pour un ensemble $E$, $\mathcal{O} \subset \mathcal{P}(E)$ est une famille d'ouverts de $E$ si $\mathcal{O}$ contient les éléments $\emptyset$ et $E$, et si toute union (éventuellement non dénombrable) d'éléments de $\mathcal{O}$ est encore un élément de $\mathcal{O}$. L'espace $(E, \mathcal{O})$ est alors appelé un *espace topologique*. Très souvent, on ne travaille pas sur des espaces topologiques généraux mais sur des espaces normés ou munis d'une distance. La famille d'ouverts est alors simplement celle associée à la norme ou la distance de l'espace considéré.

**Définition 1.1.6.** *Soit $(E, \mathcal{O})$ un espace topologique. On appelle tribu borélienne la tribu engendrée par les ouverts de $E$ :*

$$\mathcal{B}(E) = \sigma(\mathcal{O}).$$

*Lorsque $E = \mathbb{R}^d$, $\mathcal{B}(\mathbb{R}^d)$ désigne la tribu engendrée par les ouverts induits par une norme de $\mathbb{R}^d$.*

**Exercice 1.1.7.** *Montrer que $\mathcal{B}(\mathbb{R}^d) = \sigma(\{\prod_{i=1}^d ]a_i, b_i[, a_i < b_i \in \mathbb{Q}\})$.*

Lorsqu'une fonction $f : E \to F$ à valeurs dans un espace topologique est mesurable par rapport à la tribu borélienne $\mathcal{B}(F)$, on dit qu'il s'agit d'une fonction *borélienne*. Nous rappelons enfin deux résultats important concernant les fonctions boréliennes.

**Proposition 1.1.8.** *Soient $(E, \mathcal{B}(E))$ et $(F, \mathcal{B}(F))$ deux espaces topologiques munis de leur tribu borélienne. Si $g : E \to F$ est continue, alors $g$ est borélienne.*

**Proposition 1.1.9.** *Soit $(E, \mathcal{B}(E))$ un espace mesurable et $(f_n)_n \in \mathbb{N}$ une suite de fonctions boréliennes de $E$, à valeurs dans $\mathbb{R}$. Alors, les fonctions $\sup_{n\in\mathbb{N}} f_n$, $\inf_{n\in\mathbb{N}} f_n$, $\limsup_{n\to+\infty} f_n$ et $\liminf_{n\to+\infty} f_n$ sont boréliennes. En particulier, si pour tout $x \in E$ la suite $(f_n(x))$ converge, $\lim_{n\to+\infty} f_n$ est une fonction borélienne.*

Enfin, rappelons que l'ensemble des fonctions mesurables à valeurs dans $\mathbb{R}^d$ (ou même dans un $\mathbb{R}$-espace vectoriel) forme un $\mathbb{R}$-espace vectoriel. En effet, pour $\lambda_1, \lambda_2 \in \mathbb{R}$, $(x, y) \mapsto \lambda_1 x + \lambda_2 y$ est continue et donc mesurable. Si $f_1 : E_1 \to \mathbb{R}^d$ et $f_2 : E_2 \to \mathbb{R}^d$ sont mesurables, alors $(f_1, f_2) : E_1 \times E_2 \to \mathbb{R}^d \times \mathbb{R}^d$ est mesurable en mettant la tribu produit sur $E_1 \times E_2$, et par conséquent $\lambda_1 f_1 + \lambda_2 f_2$ est mesurable comme composition de deux fonctions mesurables.

### 1.1.2 Mesure et intégration

**Définition 1.1.10.** *Soit $(E, \mathcal{E})$ un espace mesurable. Une mesure positive sur $(E, \mathcal{E})$ est une application $m : \mathcal{E} \to \bar{\mathbb{R}}_+$ telle que*

*i)* $m(\emptyset) = 0$,

*ii) Propriété de $\sigma$-additivité : si $(A_i)_{i\in\mathbb{N}} \in \mathcal{E}^{\mathbb{N}}$ est une famille dénombrable d'éléments disjoints de $\mathcal{E}$ (i.e. $A_i \cap A_j = \emptyset$ pour $i \neq j$), $m(\bigcup_{i\in\mathbb{N}} A_i) = \sum_{i\in\mathbb{N}} m(A_i)$.*

*On dit alors que $(E, \mathcal{E}, m)$ est un espace mesuré.*

On en déduit alors facilement les propriétés suivantes :

1. Si $A, B \in \mathcal{E}$, $m(A \cap B) + m(A \cup B) = m(A) + m(B)$, et en particulier $m(A \cup B) \leq m(A) + m(B)$.
2. Si $A, B \in \mathcal{E}$ et $B \subset A$, $m(A) = m(B) + m(A \backslash B)$, et en particulier $m(B) \leq m(A)$.
3. Si $(A_n) \in \mathcal{E}^{\mathbb{N}}$ est une suite croissante (i.e. $A_n \subset A_{n+1}$), $m(\bigcup_{n\in\mathbb{N}} A_n) = \lim_{n\to+\infty} m(A_n)$.

Soit $(E, \mathcal{E}, m)$ un espace mesuré. On appelle *ensemble négligeable* un élément de $\mathcal{N} = \{A \subset E, \exists B \in \mathcal{E} \text{ t.q. } A \subset B \text{ et } m(B) = 0\}$. C'est un ensemble qui est contenu dans un ensemble de mesure nulle mais qui n'est pas nécessairement mesurable. Pour des raisons techniques, il est commode d'avoir la mesurabilité de ces ensembles. On peut alors étendre la mesure $m$ sur la filtration engendrée par $\mathcal{E} \cup \mathcal{N}$ de telle sorte que $\bar{m}|_{\mathcal{E}} \equiv m$ et $\bar{m}|_{\mathcal{N}} \equiv 0$. Plutôt que d'invoquer systématiquement cette extension, on préfère en pratique supposer directement que la tribu $\mathcal{E}$ contient d'emblée tous les ensembles négligeables. Ainsi, on dira que $\mathcal{E}$ satisfait la *condition usuelle* lorsque $\mathcal{N} \subset \mathcal{E}$. Par défaut, nous supposerons cette condition toujours satisfaite. Une propriété sur $E$ sera dite vraie *$m$-presque partout* si elle est vérifiée sur un ensemble $A$ mesurable t.q. $m(E \backslash A) = 0$.

Nous admettrons le résultat suivant qui définit la mesure produit.

**Proposition 1.1.11.** *Soient* $(E_1, \mathcal{E}_1, m_1)$ *et* $(E_2, \mathcal{E}_2, m_2)$ *deux espaces mesurés. Alors, il existe une unique mesure* $m$ *définie sur* $\mathcal{E}_1 \otimes \mathcal{E}_2$ *telle que*

$$\forall A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2, m(A_1 \times A_2) = m_1(A_1)m_2(A_2).$$

*Cette mesure, notée* $m_1 \otimes m_2$*, est appelée mesure produit.*

**Définition 1.1.12.** *Soient* $(E, \mathcal{E}, m)$ *un espace mesuré et* $(F, \mathcal{F})$ *un espace mesurable. Si* $g : E \to F$ *est une application mesurable, alors l'application* $m \circ g^{-1} : \mathcal{F} \to \bar{\mathbb{R}}_+$ *définie par*

$$\forall A \in \mathcal{F}, \ m \circ g^{-1}(A) = m(g^{-1}(A))$$

*est une mesure sur* $\mathcal{F}$ *appelée mesure image de* $m$ *par l'application* $g$*.*

Désormais, nous allons brièvement rappeler le cheminement qui permet de construire l'intégrale de Lebesgue, puis rappeler les théorèmes fondamentaux que sont les théorèmes de convergence dominée et monotone. On se donne $(E, \mathcal{E}, m)$ un espace mesuré. La première étape consiste à construire l'intégrale pour les fonctions mesurables réelles à valeurs positives.

**Définition 1.1.13.** *Une fonction* $g : E \to \mathbb{R}$ *est dite étagée s'il existe une famille finie* $(A_i)_{1 \le i \le n} \in \mathcal{E}^n$ *d'ensembles mesurables deux à deux disjoints et* $(\alpha_i)_{1 \le i \le n} \in \mathbb{R}^n$ *telle que*

$$\forall x \in E, g(x) = \sum_{1 \le i \le n} \alpha_i \mathbf{1}_{A_i}(x).$$

*Si en outre* $g$ *est positive on définit l'intégrale de* $g$ *par rapport à la mesure* $m$ *par :*

$$\int_E g(x)m(dx) = \sum_{1 \le i \le n} m(A_i)g(\alpha_i)$$

*avec la convention* $0 \times +\infty = 0$*.*

Il est alors facile de voir que si $g_1 : E \to \mathbb{R}$ et $g_2 : E \to \mathbb{R}$ sont deux fonctions étagées positives telles que $g_1 \le g_2$, alors $\int_E g_1(x)m(dx) \le \int_E g_2(x)m(dx)$.

**Définition 1.1.14.** *Soit* $g : E \to \bar{\mathbb{R}}_+$ *une fonction mesurable positive. On définit la suite de fonctions étagées :* $x \in \mathbb{E}$, $\Phi_n(g)(x) = \sum_{0 \le i \le n2^n - 1} \frac{i}{2^n} \mathbf{1}_{g^{-1}((\frac{i}{2^n}, \frac{i+1}{2^n}])}(x) + n\mathbf{1}_{g^{-1}((n,+\infty])}(x)$*. On a pour* $n \in \mathbb{N}$ *et* $x \in E$, $\Phi_n(g)(x) \le \Phi_{n+1}(g)(x) \le g(x)$*. Par conséquent,* $(\int_E \Phi_n(g)(x)m(dx))_{n \in \mathbb{N}}$ *est une suite croissante qui admet une limite dans* $\bar{\mathbb{R}}_+$*, et on définit :*

$$\int_E gm(dx) = \lim_{n \to +\infty} \int_E \Phi_n(g)(x)m(dx).$$

On peut vérifier que pour les fonctions étagées, cette définition coincide avec la précédente. Il est facile de voir également que pour deux fonctions mesurables positives telles que $m(g_1 > g_2) = 0$ (on dit que $g_1 \leq g_2$ $m$-presque partout), on a $\int_E g_1 m(dx) \leq \int_E g_2 m(dx)$ puisque pour tout $n$, $\Phi_n(g_1)(x) \leq \Phi_n(g_2)(x)$.

**Théorème 1.1.15** (Théorème de convergence monotone (Beppo Levi))**.** *Soit $(g_p)_{p\in\mathbb{N}}$ une suite de fonctions positives mesurables de $E$ dans $\bar{\mathbb{R}}_+$ telle que $g_p(x) \leq g_{p+1}(x)$ $m$-presque partout, i.e. $(m(\{x \in E, g_p(x) > g_{p+1}(x)\}) = 0)$. Alors $g(x) = \lim_{p\to+\infty} g_p(x)$ est définie $m$-presque partout et on a :*

$$\lim_{p\to+\infty} \int_E g_p(x)m(dx) = \int_E g(x)m(dx).$$

*Preuve.* On a $m(\{x \in E, g_p(x) > g_{p+1}(x)\}) = 0$ et par conséquent $m(\bigcup_{p\in\mathbb{N}}\{x \in E, g_p(x) > g_{p+1}(x)\}) = 0$ par $\sigma$-additivité. Sur $E\setminus\bigcup_{p\in\mathbb{N}}\{x \in E, g_p(x) > g_{p+1}(x)\}$, $(g_p(x))_{p\in\mathbb{N}}$ est une suite croissante qui admet par conséquent une limite dans $\bar{\mathbb{R}}_+$. On obtient également que $(\int_E g_p(x)m(dx))_{p\in\mathbb{N}}$ est une suite croissante qui admet une limite dans $\bar{\mathbb{R}}_+$.

Maintenant, observons que $\forall x \in E, \Phi_n(g)(x) < g(x)$ ou $\Phi_n(g)(x) = g(x) = 0$. En posant $E_p = \{x \in E, g_p(x) \geq \Phi_n(g)(x)\}$, il vient que $\bigcup_{p\in\mathbb{N}} E_p = E\setminus N$ où $N$ est un ensemble négligeable (i.e. $m(N) = 0$). On a clairement

$$\int_E \Phi_n(g)(x)\mathbf{1}_{\{x\in E,\Phi_n(g)(x)\leq g_p(x)\}}m(dx) \leq \int_E g_p(x)m(dx) \leq \lim_{p\to+\infty} \int_E g_p(x)m(dx).$$

D'autre part, le terme de gauche est égal à :
$\sum_{0\leq i\leq n2^n-1} \frac{i}{2^n} m(g^{-1}((\frac{i}{2^n}, \frac{i+1}{2^n}])\cap E_p)+nm(g^{-1}((n, +\infty]\cap E_p))$, et il tend vers $\int_E \Phi_n(g)(x)m(dx)$ par $\sigma$-additivité, puisque $E_p$ est une suite croissante d'ensembles. On obtient ainsi que $\int_E \Phi_n(g)(x)m(dx) \leq \lim\limits_{p\to+\infty} \int_E g_p(x)m(dx)$ puis $\int_E g(x)m(dx) \leq \lim\limits_{p\to+\infty} \int_E g_p(x)m(dx)$. L'autre inégalité $\lim\limits_{p\to+\infty} \int_E g_p(x)m(dx) \leq \int_E g(x)m(dx)$ étant claire, cela conclut la preuve. $\square$

Une conséquence de ce résultat est la propriété de (semi-)linéarité : si $\lambda_1, \lambda_2 \geq 0$ et $g_1, g_2$ sont deux fonctions positives mesurables, $\int_E \lambda_1 g_1(x) + \lambda_2 g_2(x)m(dx) = \lambda_1 \int_E g_1(x)m(dx) + \lambda_2 \int_E g_2(x)m(dx)$. En effet, la linéarité est facile à vérifier pour les fonctions étagées, et $\lambda_1\Phi_n(g_1)(x) + \lambda_2\Phi_n(g_2)(x)$ est une suite croissante convergeant vers $\lambda_1 g_1(x) + \lambda_2 g_2(x)$.

Du Théorème de convergence monotone, on déduit facilement le Lemme de Fatou.

**Lemme 1.1.16** (Lemme de Fatou)**.** *Soit $(g_p)_{p\in\mathbb{N}}$ une suite de fonctions mesurables positives.*

$$\int_E \liminf_{p\to+\infty} g_p(x)m(dx) \leq \liminf_{p\to+\infty} \int_E g_p(x)m(dx)$$

On dira qu'une fonction mesurable réelle positive est *intégrable* lorsque $\int_E g(x)m(dx) < \infty$. Pour une fonction mesurable réelle, on utilise la décomposition $g = g^+ - g^-$ où $g^+ =$

$\max(g,0)$ et $g^- = \max(-g,0)$. On dira que $g$ est *intégrable* lorsque $g^+$ et $g^-$ le sont, et dans ce cas on définit l'intégrale de $g$ par :

$$\int_E g(x)m(dx) = \int_E g^+(x)m(dx) - \int_E g^-(x)m(dx).$$

Construite ainsi, l'intégrale est linéaire. Enfin, on déduit du Lemme de Fatou le Théorème de convergence dominée. Comme une preuve analogue est faite plus loin pour l'espérance conditionnelle (Section 2.2.2) nous ne la faisons pas ici.

**Théorème 1.1.17** (Théorème de convergence dominée (Lebesgue)). *Soit $(g_p)_{p\in\mathbb{N}}$ une suite de fonctions mesurables telle qu'il existe une fonction positive intégrable $h$ pour laquelle $|g_p| \le h$ $m$-presque partout (i.e. $m(\{|g_p| > h\}) = 0$). On suppose que $m$-presque partout, $g_p(x) \underset{p\to+\infty}{\to} g(x)$. Alors,*

$$\int_E g_p m(dx) \underset{p\to+\infty}{\to} \int_E g m(dx) \ \textit{et} \ \int_E |g_p - g| m(dx) \underset{p\to+\infty}{\to} 0.$$

### 1.1.3 Espace de probabilités et terminologie probabiliste

Un *espace de probabilité* $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace mesuré de masse totale $\mathbb{P}(\Omega) = 1$. Nous supposerons par défaut la condition usuelle satisfaite, à savoir que $\{A \subset \Omega, \exists B \in \mathcal{F} \ t.q. \ A \subset B \ et \ \mathbb{P}(B) = 0\} \subset \mathcal{F}$. Un ensemble $A \in \mathcal{F}$ est appelé un *événement*. Un événement $A$ est dit $\mathbb{P}$-*presque sûr* ($\mathbb{P}$-p.s.) si $\mathbb{P}(A) = 1$. Sauf s'il y a un risque de confusion, on omet en général la référence à $\mathbb{P}$.

Une application $X : \Omega \to E$ à valeurs dans un espace mesurable est une *variable aléatoire* (v.a.) à valeurs dans $E$ si c'est une fonction mesurable. La mesure image de $X$ est appelée la *loi* de $X$. Lorsque $E = \mathbb{R}$, on parle de variable aléatoire réelle (v.a.r.).

Pour une variable aléatoire réelle ou à valeurs dans $\mathbb{R}^d$, on définit l'espérance comme l'intégrale par rapport à la mesure $\mathbb{P}$. Lorsqu'elle est définie (i.e. si $X$ est une v.a.r. positive ou intégrable), on note ainsi :

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

L'espérance hérite donc de toutes les propriétés de l'intégrale. Les voici récapitulées.

1. **Linéarité** : si $X, Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, $\lambda, \mu \in \mathbb{R}, \mathbb{E}[\lambda X + \mu Y] = \lambda\mathbb{E}[X] + \mu\mathbb{E}[Y]$.
2. **Positivité** : si $X \ge 0$ p.s., $\mathbb{E}[X] \ge 0$. (et $\mathbb{E}[X] = 0 \implies X = 0$ p.s.)
3. **Théorème de convergence monotone (Beppo Levi).** Soit $(X_n, n \in \mathbb{N})$ une suite de v.a.r. positives telles que $X_n \le X_{n+1}$ p.s. Alors, $X_n$ converge p.s. et on a

$$\mathbb{E}[\lim_{n\to+\infty} X_n] = \lim_{n\to+\infty} \mathbb{E}[X_n].$$

4. **Lemme de Fatou.** Soit $(X_n, n \in \mathbb{N})$ une suite de v.a.r. positives. Alors,

$$\mathbb{E}[\liminf_{n\to+\infty} X_n] \leq \liminf_{n\to+\infty} \mathbb{E}[X_n].$$

5. **Théorème de convergence dominée (Lebesgue).** Soit $(X_n, n \in \mathbb{N})$ une suite de v.a.r. qui converge p.s. vers $X$ (i.e. $\mathbb{P}(\lim_{n\to+\infty} X_n = X) = 1$). On suppose qu'il existe une v.a.r. intégrable $V$ pour laquelle $\forall n \in \mathbb{N}, |X_n| \leq V$ p.s. Alors,

$$\mathbb{E}[X_n] \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[X] \text{ et } \mathbb{E}[|X_n - X|] \underset{n\to+\infty}{\longrightarrow} 0 \text{ (on dit que } X_n \text{ tend vers } X \text{ dans } L^1).$$

Enfin, en probabilités, on ne rappelle jamais (ou très rarement) la dépendance d'une variable aléatoire en fonction de $\omega$, et on fait rarement référence à l'ensemble $\Omega$ lui-même car on ne le "voit" qu'à travers les réalisations des variables aléatoires. On note ainsi $X$ la fonction $X(\omega)$, $\{X = x\}$ l'ensemble $\{\omega, X(\omega) = x\}$, etc...

| Analyse | Probabilités |
|---|---|
| intégrale | espérance |
| ensemble mesurable $A$ | événement $A$ |
| fonction mesurable $g$ | variable aléatoire $X$ |
| mesure image $m \circ g^{-1}$ | loi de $X$ $(\mathbb{P} \circ X^{-1})$ |
| $m$-presque partout | $\mathbb{P}$-presque sûrement |

TAB. 1.1 – Correspondance des terminologies en analyse et en probabilités

## 1.2 Loi d'une variable aléatoire, indépendance

Dans cette section, nous rappelons plusieurs caractérisations possibles de la loi d'une variable aléatoire.

**Théorème 1.2.1.** *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $X, Y : \Omega \to E$ deux variables aléatoires à valeurs dans un espace mesurable $(E, \mathcal{E})$. Alors, $X$ et $Y$ ont même loi si et seulement si :*

1. *$\forall A \in \mathcal{E}, \mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$.*
2. *Pour toute fonction $f : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ bornée et mesurable,*

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

*Lorsque $E = \mathbb{R}^d$, les conditions suivantes sont également équivalentes.*

3. $\forall a_1 < a'_1, a_2 < a'_2, \ldots, a_d < a'_d, \mathbb{P}(X \in \prod_{i=1}^d ]a_i, a'_i[) = \mathbb{P}(Y \in \prod_{i=1}^d ]a_i, a'_i[)$.
4. *Pour toutes fonctions* $f_1, \ldots, f_d : \mathbb{R} \to \mathbb{R}$ *bornées et* $\mathcal{C}^\infty$,

$$\mathbb{E}\left[\prod_{i=1}^d f_i(X_i)\right] = \mathbb{E}\left[\prod_{i=1}^d f_i(Y_i)\right].$$

5. $\forall u \in \mathbb{R}^d, \mathbb{E}[\exp(iu \centerdot X)] = \mathbb{E}[\exp(iu \centerdot Y)]$.

Chacune de ces caractérisations ont biensûr leur intérêt. Typiquement, lorsque l'on cherche à prouver que deux lois sont égales, on aura tendance à préférer si possible la caractérisation 3 (resp. 4 et 5) à la caractérisation 1 (resp. 2) car elle est a priori plus simple à montrer. En revanche, si l'on sait déjà que $X$ et $Y$ ont même loi, la caractérisation 1 (resp. 2) apporte une plus riche implication que la caractérisation 3 (resp. 4 et 5). Rappelons ici que si $X$ est une v.a. à valeurs dans $\mathbb{R}^d$, la fonction $u \in \mathbb{R}^d \mapsto \mathbb{E}[\exp(iu \centerdot X)]$ intervenant dans la caractérisation 5 est appelée la *fonction caractéristique* de $X$.

Rappelons un autre résultat qui est également important.

**Proposition 1.2.2.** *Soit* $f : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *une fonction mesurable. Si* $X, Y : \Omega \to E$ *sont deux v.a. qui ont même loi, alors* $f(X)$ *est intégrable ssi* $f(Y)$ *est intégrable, et dans ce cas* $\mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$.

Nous allons maintenant rappeler la définition d'indépendance de plusieurs ou d'une suite de v.a.

**Définition 1.2.3.** *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité et* $(E_i, \mathcal{E}_i)_{i \geq 1}$ *une suite d'espaces mesurables. On considère une suite de variables aléatoires* $X_i : \Omega \to E_i$.

– *On dit que les variables* $X_1, \ldots, X_n$ *sont indépendantes si pour toutes fonctions* $f_1, \ldots, f_n$ *bornées mesurables* ($f_i : E_i \to \mathbb{R}$)

$$\mathbb{E}\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n \mathbb{E}\left[f_i(X_i)\right],$$

*ou de manière équivalente si*

$$\forall A_1 \in \mathcal{E}_1, \ldots, A_n \in \mathcal{E}_n, \mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{i=1}^n \mathbb{P}(X_i \in A_i)$$

– *On dit que la suite de v.a.* $(X_i)_{i \geq 1}$ *est une suite de v.a. indépendantes si, quelque soit* $n \geq 2$, $X_1, \ldots, X_n$ *sont indépendantes.*

On a le résultat suivant qui est souvent utile.

**Proposition 1.2.4.** *On suppose que les v.a. $X_1, \ldots, X_n$ sont indépendantes à valeurs respectivement dans $E_1, \ldots, E_n$ et on considère $f_1, \ldots, f_n$ des fonctions mesurables ($f_i : E_i \to \mathbb{R}$). Alors,*

$$\prod_{i=1}^{n} f_i(X_i) \text{ est intégrable ssi } \forall i, f_i(X_i) \text{ est intégrable,}$$

*et dans ce cas $\mathbb{E}[\prod_{i=1}^{n} f_i(X_i)] = \prod_{i=1}^{n} \mathbb{E}[f_i(X_i)]$.*

**Exercice 1.2.5.** *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $X_1, X_2$ deux variables aléatoires. Soit $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ une copie de $(\Omega, \mathcal{F}, \mathbb{P})$ et $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}}) = (\Omega \times \tilde{\Omega}, \mathcal{F} \otimes \tilde{\mathcal{F}}, \mathbb{P} \otimes \tilde{\mathbb{P}})$. Montrer que $X_1$ et $X_2$ sont indépendantes si et seulement si $(X_1, X_2)$ a même loi que $(X_1, \tilde{X}_2)$ sous $\hat{\mathbb{P}}$. En déduire que pour une fonction $f$ réelle mesurable bornée, $\mathbb{E}[f(X_1, X_2)] = \mathbb{E}[\psi(X_2)]$ où $\psi(x) = \mathbb{E}[f(X_1, x)]$.*

Enfin nous terminons cette section en définissant l'indépendance d'une v.a. avec une tribu.

**Definition 1.2.6.** *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $\mathcal{F}'$ une sous-tribu de $\mathcal{F}$. Une variable aléatoire $X$ est dite indépendante de la tribu $\mathcal{F}'$ si elle est indépendante de toute variable aléatoire $\mathcal{F}'$-mesurable.*

## 1.3 Variables aléatoires réelles et à valeurs dans $\mathbb{R}^d$

### 1.3.1 Variables aléatoires discrètes et à densité

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $X : \Omega \to \mathbb{R}$ une variable aléatoire réelle. La loi de $X$ est donnée par la mesure image $\mathbb{P} \circ X^{-1}$ ou, autrement dit par :

$$\{\mathbb{P}(X \in (a, b)),\ a < b \text{ réels.}\}$$

On pose pour $x \in \mathbb{R}$, $F_X(x) = \mathbb{P}(X \leq x)$. Cette fonction est appelée *fonction de répartition* de $X$. Il s'agit d'une fonction croissante, qui satisfait grâce à la $\sigma$-additivité de la mesure $\mathbb{P}$ : $\lim_{x \to -\infty} F_X(x) = 0$, $\lim_{x \to +\infty} F_X(x) = 1$ et elle est continue à droite et limitée à gauche. On a pour $a < b$ et $n$ assez grand $\mathbb{P}(X \in (a, b - 1/n]) = F_X(b - 1/n) - F_X(a)$, et par conséquent $\mathbb{P}(X \in (a, b)) = \lim_{x \to b-} F_X(x) - F_X(a)$. La fonction de répartition caractérise donc la loi de $X$.

Parmi les variables aléatoires réelles, on distingue deux familles particulières. Les variables aléatoires *discrètes* sont celles qui prennent un nombre dénombrable de valeurs, c'est à dire qu'il existe une suite d'éléments distincts de $\mathbb{R}$, $(x_n)_{n \in \mathbb{N}}$, telle que $\mathbb{P}(X \in \{x_n, n \in \mathbb{N}\}) = 1$. Dans ce cas on peut écrire la loi comme une combinaison de mesures de Dirac $\mathbb{P}(X \in dx) = \sum_{n\in\mathbb{N}} \mathbb{P}(X = x_n)\delta_{x_n}(dx)$, et la fonction de répartition est constante par morceaux. Les variables aléatoires à *densité* sont celles dont la fonction de répartition s'écrit

$$F_X(x) = \int_{-\infty}^{x} p(z)dz$$

pour une fonction positive mesurable $p(x)$. Dans ce cas, $p$ est uniquement déterminée à un ensemble de mesure de Lebesgue nulle près, et on l'appelle *la densité de* $X$. Lorsqu'elle existe, cette densité caractérise entièrement la loi de $X$. Une v.a. à densité ne charge aucun point, i.e. $\mathbb{P}(X = x) = 0$ pour tout $x \in \mathbb{R}$. Il est important de noter que si les v.a. discrètes et les v.a. à densités sont deux familles disjointes, elle ne constituent pas une partition des v.a. et il existe des variables aléatoires ni discrètes, ni à densité. Par exemple, $F_X(x) = \mathbf{1}_{x\geq 0}\frac{2-e^{-x}}{2}$ est la fonction de répartition d'une telle v.a. Pour les v.a. à valeurs dans $\mathbb{R}^d$ on utilise également cette appellation : une v.a. discrète est une v.a. qui y prend un nombre dénombrable de valeurs tandis qu'une v.a. à densité satisfait $\mathbb{P}(X \in A) = \int_A p(x)dx$ pour tout ouvert $A$ de $\mathbb{R}^d$.

### 1.3.2 Espaces $L^p$ et $L^\infty$

Pour $p \in \mathbb{N}^*$, on définit $L^p(\Omega, \mathcal{F}, \mathbb{P}) = \{X : \Omega \to \mathbb{R}^d,\ t.q\ \mathbb{E}(|X|^p) < \infty\}$. Muni de la norme $\|X\|_p = \mathbb{E}(|X|^p)^{1/p}$, on peut montrer que $L^p(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de Banach. Dans le cas $p = 2$, c'est un espace de Hilbert muni du produit scalaire $X \cdot Y = \mathbb{E}[XY]$. De même, on définit $L^\infty(\Omega, \mathcal{F}, \mathbb{P}) = \{X : \Omega \to \mathbb{R}^d,\ t.q\ \exists M > 0, \mathbb{P}(|X| \leq M) = 1\}$. Muni de la norme $\|X\|_\infty = \inf\{M > 0, \mathbb{P}(|X| \leq M) = 1\}$, on peut montrer que $L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de Banach. Remarquons ici que bien que ces espaces dépendent de la dimension $d$, nous n'avons pas rappelé cette dépendance dans la notation $L^p(\Omega, \mathcal{F}, \mathbb{P})$. C'est parce que la plupart du temps, la dimension est claire dans le contexte étudié.

On se donne maintenant une v.a. $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ réelle. On définit sa variance et son écart-type par :

$$\mathbf{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] \geq 0 \text{ et } \sigma(X) = \sqrt{\mathbf{Var}(X)}.$$

Par linéarité de l'espérance, on a aussi l'égalité parfois plus commode :

$$\mathbf{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Si $Y$ est une seconde v.a. réelle de carré intégrable, on définit respectivement la covariance et la corrélation entre $X$ et $Y$ par

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

et $\rho(X, Y) = Cov(X, Y)/(\sigma(X)\sigma(Y))$ si $\sigma(X)\sigma(Y) > 0$ et $\rho(X, Y) = 0$ sinon. Ces quantités sont nulles lorsque $X$ est indépendante de $Y$, mais la réciproque est fausse.

**Exercice 1.3.1.** *Soit $X$ une v.a. réelle à densité $p(x)$. On suppose que $p(x) = p(-x)$ et que $\int_{\mathbb{R}} x^4 p(x)dx < \infty$. Montrer que $cov(X, X^2) = 0$ mais que les v.a. $X$ et $X^2$ ne sont pas indépendantes.*

On considère désormais une v.a. $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $\mathbb{R}^d$, et on note $X = (X_1, \ldots, X_d)'$. Pour $\alpha \in \mathbb{R}^d$, on a $\mathbf{Var}(\sum_{i=1}^d \alpha_i X_i) = \sum_{i=1}^d \sum_{j=1}^d \alpha_i \alpha_j (\mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\mathbb{E}[X_j])$. On définit alors la *matrice de covariance de $X$* par

$$i, j \in \{1, \ldots, d\},\ (\Gamma_X)_{i,j} = Cov(X_i, X_j).$$

Il s'agit d'une matrice symétrique qui est positive puisque $\alpha' \Gamma_X \alpha = \mathbf{Var}(\sum_{i=1}^d \alpha_i X_i) \geq 0$.

### 1.3.3 Variables aléatoires réelles usuelles

Voici deux tables récapitulatives des principales v.a. réelles.

| Nom | Param. | Loi | $\mathbb{E}[X]$ | $\mathbf{Var}(X)$ | fct. car. $\mathbb{E}[e^{iuX}]$ |
|---|---|---|---|---|---|
| Bernoulli $\mathcal{B}(p)$ | $p \in (0,1)$ | $\mathbb{P}(X=1) = p = 1 - \mathbb{P}(X=0)$ | $p$ | $p(1-p)$ | $1 + p(e^{iu} - 1)$ |
| Binômiale $\mathcal{B}(n,p)$ | $n \in \mathbb{N}^*, p \in (0,1)$ | $0 \leq k \leq n, \mathbb{P}(X=k) = C_n^k p^k (1-p)^{n-k}$ | $np$ | $np(1-p)$ | $(1 + p(e^{iu} - 1))^n$ |
| Poisson $\mathcal{P}(\lambda)$ | $\lambda > 0$ | $k \in \mathbb{N}, \mathbb{P}(X=k) = e^{-\lambda} \frac{\lambda^k}{k!}$ | $\lambda$ | $\lambda$ | $\exp(\lambda(e^{iu} - 1))$ |
| Géométrique $\mathcal{G}eo(p)$ | $p \in (0,1)$ | $k \in \mathbb{N}^*, \mathbb{P}(X=k) = p(1-p)^{k-1}$ | $1/p$ | $(1-p)/p^2$ | $\frac{pe^{iu}}{1-(1-p)e^{iu}}$ |

TAB. 1.2 – Variables aléatoires discrètes

| Nom | Param. | Loi (densité) | $\mathbb{E}[X]$ | $\mathbf{Var}(X)$ | fct. car. $\mathbb{E}[e^{iuX}]$ |
|---|---|---|---|---|---|
| Gaussienne $\mathcal{N}(m,\sigma^2)$ | $m \in \mathbb{R}, \sigma > 0$ | $\frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-m)^2}{2\sigma^2})$ | $m$ | $\sigma^2$ | $\exp(ium - \frac{\sigma^2}{2}u^2)$ |
| Uniforme $\mathcal{U}(a,b)$ | $a < b$ | $\frac{1}{b-a}\mathbf{1}_{[a,b]}(x)$ | $(a+b)/2$ | $(b-a)^2/12$ | $\frac{\sin((b-a)u/2)}{(b-a)u/2}e^{iu\frac{a+b}{2}}$ |
| Exponentielle $\mathcal{E}(\lambda)$ | $\lambda > 0$ | $\mathbf{1}_{\{x>0\}}\lambda e^{-\lambda x}$ | $1/\lambda$ | $1/\lambda^2$ | $\exp(\lambda(e^{iu}-1))$ |
| Cauchy $\mathcal{C}(a)$ | $a > 0$ | $\frac{a}{\pi}\frac{1}{x^2+a^2}$ | Non défini | Non défini | $e^{-a\|u\|}$ |
| Gamma $\Gamma(a,\theta)$ | $a, \theta > 0$ | $\frac{\theta^a}{\Gamma(a)}x^{a-1}e^{-\theta x}\mathbf{1}_{\{x>0\}}$ | $a/\theta$ | $a/\theta^2$ | |
| Beta $\beta(a,b)$ | $a, b > 0$ | $\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}\mathbf{1}_{\{0<x<1\}}$ | $\frac{a}{a+b}$ | $\frac{ab}{(a+b)^2(a+b+1)}$ | |
| Chi 2 $\chi^2(n)$ | $n \in \mathbb{N}^*$ | $\frac{(1/2)^{n/2}}{\Gamma(n/2)}x^{n/2-1}e^{-x/2}\mathbf{1}_{\{x>0\}}$ | $n$ | $2n$ | |

TAB. 1.3 – Variables aléatoires à densité

## 1.4 Les différents modes de convergence

Dans cette partie nous rappelons les nombreux modes de convergences utilisés en probabilités. On considère une suite $(X_n)_{n\in\mathbb{N}}$ de v.a. à valeurs dans $\mathbb{R}^d$. Lorsque $n \to +\infty$, on dit que la suite $(X_n)$ converge vers la v.a. $X$ :

- *presque sûrement (p.s)* si $\mathbb{P}(X_n \underset{n\to+\infty}{\longrightarrow} X) = 1$,
- *en probabilité* si $\forall \varepsilon > 0, \mathbb{P}(|X_n - X| \geq \varepsilon) \underset{n\to+\infty}{\longrightarrow} 0$.
- *dans* $L^p$ (en pratique, $p = 1, 2$ le plus souvent) lorsque $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ et $X_n \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ pour $n$ assez grand, et si l'on a : $\mathbb{E}(|X_n - X|^p) \underset{n\to+\infty}{\longrightarrow} 0$.
- *en loi* si pour toute fonction $f : \mathbb{R}^d \to \mathbb{R}$ continue bornée, $\mathbb{E}[f(X_n)] \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[f(X)]$.

Parmi ces différentes convergences, certaines sont plus fortes que d'autres. Nous avons les implications suivantes :

$$\boxed{\begin{array}{ccccccc} \text{cv } L^2 & \Longrightarrow & \text{cv } L^1 & \Longrightarrow & \text{cv en } Proba. & \Longleftarrow & \text{cv } p.s. \\ & & & & \Downarrow & & \\ & & & & \text{cv en loi} & & \end{array}}$$

TAB. 1.4 – Hierarchie des modes de convergence.

En outre, lorsque $X_n$ converge p.s vers $X$ et qu'il existe une v.a. intégrable telle que $|X_n|^p \leq V$, le théorème de convergence dominée nous assure alors que $X_n$ converge vers $X$ dans $L^p$. Enfin, lorsque $X_n$ converge en probabilité vers $X$, on peut montrer qu'il existe une sous-suite de $(X_n)_{n\in\mathbb{N}}$ qui converge presque sûrement vers $X$. C'est l'objet de l'exercice suivant.

**Exercice 1.4.1.** *On suppose que $X_n$ converge en probabilité vers $X$.*

1. *Montrer qu'il existe une sous-suite $\phi_1(n)$ telle que $\mathbb{P}(|X_{\phi_1(n)} - X| \geq 1) \leq 1/n^2$. En déduire que $\mathbb{E}[\sum_{n\in\mathbb{N}} \mathbf{1}_{\{|X_{\phi_1(n)} - X| \geq 1\}}] < +\infty$ puis que, presque sûrement, il existe un entier $N$ t.q. $|X_{\phi_1(n)} - X| < 1$ pour $n \geq N$ (i.e. $\mathbb{P}(\exists N, \forall n \geq N, |X_{\phi_1(n)} - X| < 1) = 1$).*

2. *Montrer qu'il existe une sous-suite $\phi_2(n)$ telle que $\mathbb{P}(|X_{\phi_1\circ\phi_2(n)} - X| \geq 1/2) \leq 1/n^2$. Puis, par récurrence, montrer qu'on peut construire des sous-suites $\phi_3(n), \ldots$ telles que pour tout $k$, $\mathbb{P}(|X_{\phi_1\circ\cdots\circ\phi_k(n)} - X| \geq 1/k) \leq 1/n^2$. En déduire que $\mathbb{P}(\exists N, \forall n \geq N, |X_{\phi_1\circ\cdots\circ\phi_k(n)} - X| < 1/k) = 1$.*
3. *On pose $\varphi(n) = \phi_1 \circ \cdots \circ \phi_n(n)$ (procédé diagonal). Montrer que $X_{\varphi(n)}$ converge presque sûrement vers $X$.*

*Donner un exemple de suite $(X_n)_{n\in\mathbb{N}}$ qui converge en loi vers une v.a. $X$ mais dont aucune sous-suite ne converge p.s. vers $X$.*

**Exercice 1.4.2.** *L'objet de cet exercice est de donner un exemple de suite de v.a. qui converge en probabilité mais pas presque sûrement. On se donne une suite $(X_n)_{n\geq 1}$ de v.a. indépendantes telle que $X_n \overset{loi}{=} \mathcal{B}(1/n)$, i.e $\mathbb{P}(X_n = 1) = 1/n = 1 - \mathbb{P}(X_n = 0)$.*

1. *Montrer que $X_n$ converge en probabilité vers $0$.*
2. *Soit $(x_n)_{n\geq 1}$ une suite à valeurs dans $\{0, 1\}$. Montrer que $\lim_{n\to+\infty} x_n = 0$ ssi $\exists N \in \mathbb{N}^*, \forall n \geq N, x_n = 0$.*
3. *En déduire que $\{\lim_{n\to+\infty} X_n = 0\} = \cup_{N\in\mathbb{N}^*}\cap_{n\geq N}\{X_n = 0\}$, puis que $\mathbb{P}(\lim_{n\to+\infty} X_n = 0) = \lim_{N\to+\infty} \mathbb{P}(\cap_{n\geq N}\{X_n = 0\})$.*
4. *Montrer que $\ln(\mathbb{P}(\cap_{N\leq n\leq N+k}\{X_n = 0\})) = \sum_{i=0}^{k} \ln(1 - \frac{1}{N+i})$. En déduire que $\mathbb{P}(\cap_{n\geq N}\{X_n = 0\}) = 0$ puis que $X_n$ ne converge pas p.s. vers $0$.*
5. *En utilisant la même méthode, montrer qu'en revanche la sous-suite $(X_{n^2})_{n\geq 1}$ converge p.s. vers $0$, i.e. $\mathbb{P}(\lim_{n\to+\infty} X_{n^2} = 0) = 1$.*

L'utilisation des fonctions caractéristiques donne une méthode parfois plus simple pour établir la convergence en loi d'une suite de variables aléatoires.

**Théorème 1.4.3.** *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. à valeurs dans $\mathbb{R}^d$. Alors, $X_n$ converge en loi vers la v.a. $X$ si et seulement si*

$$\forall u \in \mathbb{R}^d, \mathbb{E}[e^{iuX_n}] \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[e^{iuX}].$$

Enfin, pour une fonction $f$ bornée, on définit son ensemble de discontinuité par $Disc(f) = \{x, \liminf_x f < \limsup_x f\}$. Si $X_n$ converge en loi vers $X$ et si $\mathbb{P}(X \in Disc(f)) = 0$, alors on a $\mathbb{E}[f(X_n)] \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[f(X)]$.

Les preuves de beaucoup des implications énoncées dans la Table 1.4 reposent sur des inégalités fréquemment utiles en probabilité. Nous les rappelons ici pour des v.a. $X$ et $Y$ réelles.

– **Inégalité de Cauchy-Schwarz :** Si $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P}), \mathbb{E}[|XY|] \leq \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}$. En particulier ($Y \equiv 1$), $\mathbb{E}[|X|] \leq \sqrt{\mathbb{E}(X^2)}$.

- **Inégalité de Markov :** Si $X \in L^1(\Omega, \mathcal{F}, \mathbb{P}), x > 0, \mathbb{P}(|X| \geq x) \leq \frac{\mathbb{E}[|X|]}{x}$. Cela vient immédiatement de la majoration $\mathbf{1}_{|X|\geq x} \leq |X|/x$.
- **Inégalité de Bienaymé-Tchebychev :** Si $X \in L^2(\Omega, \mathcal{F}, \mathbb{P}), x > 0, \mathbb{P}(|X| \geq x) \leq \frac{\mathbb{E}[X^2]}{x^2}$. Découle de $\mathbf{1}_{|X|\geq x} \leq X^2/x^2$.

On termine cette section par un résultat très commode pour la convergence p.s. qui est souvent utilisé.

**Proposition 1.4.4.** *Une suite $(X_n)_{n\in\mathbb{N}}$ de v.a. à valeurs dans $\mathbb{R}^d$ converge p.s.*

$$\textit{(i.e. } \mathbb{P}(\forall \varepsilon > 0, \exists N \in \mathbb{N}^*, \forall n \geq N, |X_n - X| \leq \varepsilon) = 1\textit{)}$$

*si, et seulement si*

$$\forall \varepsilon > 0, \mathbb{P}(\exists N \in \mathbb{N}^*, \forall n \geq N, |X_n - X| \leq \varepsilon) = 1.$$

*Preuve.* On note $A_\varepsilon = \{\exists N \in \mathbb{N}^*, \forall n \geq N, |X_n - X| \leq \varepsilon\}$ et $A = \cap_{\varepsilon>0} A_\varepsilon$. Puisque pour tout $\varepsilon > 0$, $A \subset A_\varepsilon$, le sens direct est immédiat. Prouvons la réciproque, c'est à dire que si $\mathbb{P}(A_\varepsilon) = 1$ pour tout $\varepsilon > 0$, alors $\mathbb{P}(A) = 1$. Il est facile de voir que $A_\varepsilon \subset A_{\varepsilon'}$ pour $0 < \varepsilon < \varepsilon'$, et par conséquent $A = \cap_{n\in\mathbb{N}^*} A_{1/n}$. Par $\sigma$-additivité, $\mathbb{P}(A) = \lim_{n\to+\infty} \mathbb{P}(A_{1/n}) = 1$. □

## 1.5 Deux théorèmes fondamentaux

Dans cette section, nous rappelons les deux théorèmes fondamentaux que sont la loi forte des grands nombres (LFGN) et le théorème de la limite centrale ou théorème central limite (TCL).

**Théorème 1.5.1** (LFGN). *Soit $(X_n)_{n\geq 1}$ une suite de v.a. réelles i.i.d intégrables. On definit $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$ la moyenne empirique des $n$ premières réalisations. Alors, $\bar{X}_n$ converge p.s. vers $\mathbb{E}(X_1)$ lorsque $n \to +\infty$.*

**Théorème 1.5.2** (TCL). *Soit $(X_n)_{n\geq 1}$ une suite de v.a. réelles i.i.d de carré intégrables. On supose $\sigma = \sqrt{\mathbb{E}(X_1^2) - \mathbb{E}(X_1)^2} > 0$. On a alors :*

$$\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mathbb{E}(X_1)) \underset{n\to+\infty}{\overset{loi}{\to}} \mathcal{N}(0,1)$$

La loi forte des grands nombre donne la limite de la moyenne empirique tandis que le théorème de la limite centrale décrit la dispersion asymptotique de la moyenne empirique autour de cette limite.

Une application importante de ces théorèmes est la méthode de Monte-Carlo et la construction d'intervalles de confiance. Plus précisément, supposons que l'on observe (ou simule) la réalisation de variables aléatoires réelles $X_1, \dots, X_n$ i.i.d. de carré intégrable, et que l'on souhaite estimer l'espérance $\mathbb{E}(X_1)$ commune à ces v.a. Grâce au TCL, nous savons que pour $n$ "grand", $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mathbb{E}(X_1)) \overset{loi}{\approx} \mathcal{N}(0,1)$. Pour $\alpha \in (0,1)$, on considère le quantile de la loi gaussienne standard t.q. $\mathbb{P}(|N| > q_\alpha) = \alpha$. On dit alors qu'avec un *niveau de confiance* $1-\alpha$, $\frac{\sqrt{n}}{\sigma}|\bar{X}_n - \mathbb{E}(X_1)| \leq q_\alpha$ ou de manière équivalente :

$$\mathbb{E}(X_1) \in [\bar{X}_n - q_\alpha \frac{\sigma}{\sqrt{n}}, \bar{X}_n + q_\alpha \frac{\sigma}{\sqrt{n}}].$$

Cet intervalle est appelé *intervalle de confiance de niveau* $1-\alpha$. Typiquement, on prend les valeurs $\alpha = 1\%, 5\%, 10\%$ pour lesquelles on a respectivement $q_{1\%} \approx 2.58$, $q_{5\%} \approx 1,96$ et $q_{10\%} \approx 1,65$. Par exemple, l'intervalle de confiance à 95% est $[\bar{X}_n - 1.96\frac{\sigma}{\sqrt{n}}, \bar{X}_n + 1.96\frac{\sigma}{\sqrt{n}}]$. Cela signifie qu'avec 95% de chances, l'espérance que l'on souhaite estimer s'y trouve dedans. Néanmoins, déterminer cet intervalle nécessite de connaître la variance de la v.a. observée, ce qui souvent n'est pas le cas. On doit alors l'estimer. La loi forte des grands nombres assure alors que $\hat{V}_n := \frac{1}{n-1}\sum_{k=1}^n (X_k - \bar{X}_n)^2 = \frac{1}{n-1}\sum_{k=1}^n X_k^2 - \frac{n}{n-1}\bar{X}_n^2 \underset{n\to+\infty}{\overset{p.s.}{\to}} \sigma^2$, et on considère l'intervalle de confiance de niveau $1-\alpha$ :

$$\mathbb{E}(X_1) \in [\bar{X}_n - q_\alpha \frac{\sqrt{\hat{V}_n}}{\sqrt{n}}, \bar{X}_n + q_\alpha \frac{\sqrt{\hat{V}_n}}{\sqrt{n}}]$$

qui ne dépend que des réalisations $X_1, \ldots, X_n$. Lorsque ces réalisations sont simulées pour calculer l'espérance $E[X_1]$, on parle de méthode de Monte-Carlo.

Pour conclure cette section, disons que la loi forte des grands nombres et le théorème de la limite centrale s'étendent au aux variables à valeurs dans $\mathbb{R}^d$. Pour la LFGN, cette extension est triviale puisque l'on applique coordonnées par coordonnées la LFGN unidimensionnelle. Par contre pour le TCL, nous devons au préalable définir les vecteurs gaussiens qui interviendront dans la loi limite de $\sqrt{n}(\bar{X}_n - \mathbb{E}(X_1))$.

## 1.6 Vecteurs gaussiens

**Définition 1.6.1.** *Un vecteur aléatoire $X : \Omega \to \mathbb{R}^d$ est un vecteur gaussien si toute combinaison linéaire de ses coordonnées est une v.a. réelle gaussienne.*

Si $X$ est un vecteur Gaussien et que $u \in \mathbb{R}^d$, $u \cdot X$ est donc une v.a. gaussienne de moyenne $u \cdot \mathbb{E}[X]$ et de variance $u \cdot \Gamma_X u$ où $\Gamma_X$ désigne la matrice de covariance de $X$. Par conséquent, on a $\mathbb{E}[e^{iu \cdot X}] = \exp(iu \cdot \mathbb{E}[X] - \frac{1}{2}u \cdot \Gamma_X u)$. Réciproquement, supposons que $X$ est un vecteur aléatoire dont la fonction caractéristique est $\exp(iu \cdot m - \frac{1}{2}u \cdot \Gamma u)$ où $\Gamma$ est une matrice symétrique positive et $m \in \mathbb{R}^d$. Alors, pour $u \in \mathbb{R}$ et pour tout $v \in \mathbb{R}$, $\mathbb{E}[e^{ivu \cdot X}] = \exp(ivu \cdot m - \frac{v^2}{2}u \cdot \Gamma u)$, et on en déduit que $u \cdot X$ est une v.a. gaussienne d'espérance $u \cdot m$ et de variance $u \cdot \Gamma u$. Par conséquent, $X$ est un vecteur gaussien.

**Proposition 1.6.2.** *Un vecteur aléatoire $X : \Omega \to \mathbb{R}^d$ est gaussien ssi il existe une matrice symétrique positive $\Gamma$ et $m \in \mathbb{R}^d$ t.q. $\mathbb{E}[e^{iu\cdot X}] = \exp(iu\cdot m - \frac{1}{2}u\cdot\Gamma u)$. Dans ce cas $m = \mathbb{E}[X]$ et $\Gamma$ est la matrice de covariance de $X$, et on note $\mathcal{N}_d(m,\Gamma)$ cette loi.*

Donnons maintenant un exemple de vecteurs gaussiens : on se donne $d$ v.a. gaussiennes indépendantes $X_1,\ldots,X_d$ et on pose $X = (X_1,\ldots,X_d)'$. Alors, pour $u \in \mathbb{R}^d$ on a : $\mathbb{E}[e^{iu\cdot X}] = \mathbb{E}[e^{i\sum_{k=1}^d u_k X_k}] = \prod_{k=1}^d \mathbb{E}[e^{iu_k X_k}] = \exp\left(i\sum_{k=1}^d u_k\mathbb{E}[X_k] - \frac{1}{2}\sum_{k=1}^d u_k^2 \mathbf{Var}(X_k)\right)$. Par conséquent, $X$ est un vecteur gaussien de moyenne $(\mathbb{E}[X_1],\ldots,\mathbb{E}[X_d])$ et de matrice de covariance $diag(\mathbf{Var}(X_1),\ldots,\mathbf{Var}(X_d))$.

Réciproquement, si $X$ est un vecteur gaussien d'espérance $m$ et de matrice de covariance diagonale $diag(\sigma_1^2,\ldots,\sigma_d^2)$, $\mathbb{E}[e^{iu\cdot X}] = \prod_{k=1}^d \exp\left(iu_k\mathbb{E}[X_k] - \frac{1}{2}u_k^2\sigma_k^2\right) = \prod_{k=1}^d \mathbb{E}[e^{iu_k X_k}]$ et ses coordonnées sont donc indépendantes. Nous venons de prouver le résultat suivant.

**Proposition 1.6.3.** *Les coordonnées d'un vecteur gaussien sont indépendantes si et seulement si sa matrice de covariance est diagonale.*

Une erreur classique consiste à penser qu'un vecteur est gaussien si ses coordonnées sont gaussiennes. Ceci est faux, voici un contre-exemple. On se donne $X_1 \sim \mathcal{N}(0,1)$ et on pose $X_2 = -X_1\mathbf{1}_{\{|X_1|\le 1\}} + X_1\mathbf{1}_{\{|X_1|>1\}}$. Il est facile de vérifier que $X_2$ suit une loi gaussienne standard, mais la combinaison linéaire $X_1 + X_2$ n'est pas gaussienne puisque $\mathbb{P}(X_1 + X_2 = 0) = \mathbb{P}(|X_1| \le 1) > 0$.

Le résultat suivant concernant les v.a. gaussiennes est important.

**Proposition 1.6.4.** *Si $(X_n)_{n\in\mathbb{N}}$ est une suite de v.a. gaussiennes réelles qui converge en loi vers $X$, alors $X$ est une v.a. gaussienne. Par conséquent, si $(X_n)_{n\in\mathbb{N}}$ est une suite de vecteurs gaussiens qui convergent en loi vers $X$, alors $X$ est un vecteur gaussien.*

Nous concluons ce rappel sur les vecteurs gaussiens en énonçant le TCL multidimensionnel.

**Théorème 1.6.5.** *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. i.i.d. à valeurs dans $\mathbb{R}^d$ et de carré intégrable. On note $\Gamma$ la matrice de covariance de $X_1$ et $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_i$ la moyenne empirique des $n$ premières réalisations. Alors,*

$$\sqrt{n}(\bar{X}_n - \mathbb{E}(X_1)) \underset{n\to+\infty}{\overset{loi}{\longrightarrow}} \mathcal{N}_d(0,\Gamma).$$

# Chapitre 2

# Espérance conditionnelle, loi conditionnelle

## 2.1 Un exemple introductif

Considérons un jeu de dés où l'on effectue deux lancés successifs qui rapporte le produit des deux résultats obtenus. On note $X_1$ et $X_2$ les deux résultats que l'on suppose indépendants, et de loi uniforme sur $\{1, 2, \ldots, 6\}$. La somme rapportée par le jeu est donc $S = X_1X_2$. On suppose qu'après le premier lancé, on a accès au résultat $X_1$. Une question naturelle est alors de savoir comment est distribuée la somme $S$ une fois que l'on connaît le premier lancé, et quel est l'espérance de gain. Dans ce cas, la réponse est simple : $S$ suit une loi uniforme sur $\{X_1, 2X_1, \ldots, 6X_1\}$ et l'espérance de gain est $\frac{1+2+\cdots+6}{6}X_1 = 3,5X_1$. On dira alors que la *loi conditionnelle de $S$ sachant* $X_1$ est la loi uniforme sur $\{X_1, 2X_1, \ldots, 6X_1\}$ tandis que l'*espérance conditionnelle de $S$ sachant* $X_1$ est $3,5X_1$.

Continuons maintenant cet exemple pour montrer où ces objets peuvent intervenir. On souhaite organiser le jeu suivant : pour jouer, le joueur doit payer le prix $p$ et après avoir vu le résultat du premier lancé, il peut soit récupérer la moitié de ce qu'il a payé, soit poursuivre le jeu et gagner la somme $S$. La question naturelle est la suivante : quel est le prix minimal à faire payer pour être sûr de ne pas perdre de l'argent en moyenne ? Pour cela, on suppose que les joueurs sont rationnels, et cherchent à maximiser leur espérance de gain. Par conséquent, ils choisissent de continuer le jeu si $3,5X_1 \geq p/2$, et reprennent la somme $p/2$ sinon. L'espérance de la somme touchée par le joueur rationnel est donc :

$$\mathbb{E}[\mathbf{1}_{\{3,5X_1 \geq p/2\}}X_1X_2 + \mathbf{1}_{\{3,5X_1 < p/2\}}p/2].$$

$\mathbb{E}[\mathbf{1}_{\{3,5X_1 \geq p/2\}}X_1X_2] = \frac{1}{6^2}\sum_{k_1,k_2 \in \{1,\ldots,6\}} \mathbf{1}_{\{3,5k_1 \geq p/2\}}k_1k_2 = \frac{1}{6}\sum_{k_1 \in \{1,\ldots,6\}} \mathbf{1}_{\{3,5k_1 \geq p/2\}}3,5k_1 = \mathbb{E}[\mathbf{1}_{\{3,5X_1 \geq p/2\}}3,5X_1]$. Comme on a $\mathbf{1}_{\{3,5X_1 \geq p/2\}}3,5X_1 + \mathbf{1}_{\{3,5X_1 < p/2\}}p/2 = \max(3,5X_1, p/2)$, l'espérance de la somme reçue par le joueur s'écrit $\mathbb{E}[\max(3,5X_1, p/2)]$. Du point de vue

de l'organisateur, l'espérance de gain est donc $G(p) = p - \mathbb{E}[\max(3,5X_1, p/2)]$, et pour que ce jeu lui soit rentable sur le long terme, il doit fixer $p$ pour que $G(p) \geq 0$ . Il est aisé de voir qu'il s'agit d'une fonction croissante de $p$.
Pour $0 \leq p \leq 7$, $G(p) = p - 3.5^2 \leq 0$.
Pour $7 \leq p \leq 14$, $G(p) = \frac{11}{12}p - \frac{70}{6}$ et $G$ s'annule pour $p = \frac{140}{11} \approx 12,73 \in [7, 14]$.
Par conséquent, le prix minimum à faire payer est $\frac{140}{11}$ pour ne pas perdre en moyenne.

**Remarque 2.1.1.** *Un joueur "non rationnel" prend la décision de continuer le jeu si $X_1 \in A$ et d'arrêter si $X_1 \notin A$ pour un certain ensemble $A \subset \{1, \ldots, 6\}$. Son gain est alors* $\mathbb{E}[X_1X_2\mathbf{1}_{\{X_1\in A\}} + (p/2)\mathbf{1}_{\{X_1\notin A\}}] = \frac{1}{6^2}\sum_{k_1,k_2\in\{1,\ldots,6\}}\left[k_1k_2\mathbf{1}_{\{k_1\in A\}} + k_1\mathbf{1}_{\{k_1\notin A\}}\right] = \frac{1}{6}\sum_{k_1\in\{1,\ldots,6\}}\left[3,5k_1\mathbf{1}_{\{k_1\in A\}} + k_1\mathbf{1}_{\{k_1\notin A\}}\right] = \mathbb{E}[3,5X_1\mathbf{1}_{\{X_1\in A\}} + (p/2)\mathbf{1}_{\{X_1\notin A\}}]$. *Puisqu'on a* $3,5X_1\mathbf{1}_{\{X_1\in A\}} + (p/2)\mathbf{1}_{\{X_1\notin A\}} \leq \max(3,5X_1, p/2)$, *il gagne moins que le joueur "rationnel", et génère ainsi plus de profit pour l'organisateur du jeu !*

## 2.2 Espérance conditionnelle

Avant de construire rigoureusement l'espérance conditionnelle, il est bon de donner une interprétation intuitive de cet objet. Il s'agit d'une espérance (et donc d'une intégrale) qui jouit par conséquent des mêmes propriétés que l'espérance usuelle. Rappelons qu'heuristiquement, l'espérance correspond à effectuer une moyenne sur tous les événements possibles $\mathcal{F}$ de $\Omega$. L'espérance conditionnelle, elle, consiste intuitivement à effectuer une moyenne sur tous les événements qui sont encore possibles (en général le futur), sachant que l'on a déjà observé certains événements (en général le passé). Il s'agit donc d'une moyenne qui tient compte des événements réalisés connus. Ces événements connus seront modélisés par une sous-tribu $\mathcal{F}'$ de $\mathcal{F}$, si bien que l'on dit souvent que la sous-tribu $\mathcal{F}'$ représente l'information disponible (dans l'exemple introductif, il s'agit du résultat du premier lancé).

### 2.2.1 Construction de l'espérance conditionnelle

L'objet de cette section est d'établir l'existence et l'unicité de l'espérance conditionnelle. Dans une première lecture, on pourra admettre directement les deux résultats principaux (Théorèmes 2.2.4 et 2.2.5).

On considère $X$ une variable aléatoire réelle intégrable (resp. positive) sur $(\Omega, \mathcal{F}, \mathbb{P})$ et $\mathcal{F}'$ une sous-tribu de $\mathcal{F}$. Nous allons montrer qu'il existe alors une unique variable aléatoire réelle $Y$, $\mathcal{F}'$-mesurable et intégrable (resp. positive), telle que

$$\forall A \in \mathcal{F}', \mathbb{E}[\mathbf{1}_A X] = \mathbb{E}[\mathbf{1}_A Y]. \tag{2.1}$$

Dans la suite du cours, on notera $E[X|\mathcal{F}']$ cette variable aléatoire *appelée espérance conditionnelle de $X$ sachant $\mathcal{F}'$.*

Nous commençons par énoncer un résultat d'unicité.

**Lemme 2.2.1.** *Si $Y_1$ et $Y_2$ sont deux variables aléatoires réelles $\mathcal{F}'$-mesurables, intégrables ou positives, telles que $\forall A \in \mathcal{F}', \mathbb{E}[\mathbf{1}_A Y_1] = \mathbb{E}[\mathbf{1}_A Y_2]$, alors $Y_1 = Y_2$ p.s.*

*Preuve.* Nous prouvons ce résultat par contraposée. Si $\mathbb{P}(Y_1 \neq Y_2) > 0$, cela signifie que $\mathbb{P}(Y_1 > Y_2) > 0$ ou $\mathbb{P}(Y_2 > Y_1) > 0$. Considérons par exemple le premier cas. Comme $\{Y_1 > Y_2\} = \bigcup_{m\in\mathbb{N}}\{Y_1 > Y_2, Y_2 \leq m\}$, il existe un entier $m \in \mathbb{N}$ tel que $\mathbb{P}(Y_1 > Y_2, Y_2 \leq m) > 0$. Les variables aléatoires $Y_1$ et $Y_2$ sont toutes deux $\mathcal{F}'$-mesurables et donc $\{Y_1 > Y_2, Y_2 \leq m\} \in \mathcal{F}'$. On en déduit alors $\mathbb{E}[\mathbf{1}_{\{Y_1>Y_2,Y_2\leq m\}}(Y_1 - Y_2)] > 0$. D'autre part, on a $\mathbb{E}[\mathbf{1}_{\{Y_1>Y_2,Y_2\leq m\}}Y_1] = \mathbb{E}[\mathbf{1}_{\{Y_1>Y_2,Y_2\leq m\}}Y_2]$ (cette quantité est finie que $Y_2$ soit positive ou intégrable). Ceci implique $\mathbb{E}[\mathbf{1}_{Y_1>Y_2,Y_2\leq m}(Y_1 - Y_2)] = 0$ ce qui est contradictoire. □

La proposition suivante donne un résultat d'existence de l'espérance conditionnelle lorsque la v.a. $X$ est de carré intégrable. Outre ce résultat, elle donne une autre interprétation intuitive de l'espérance conditionnelle dans ce cas : c'est la projection orthogonale sur le sous-espace des v.a.r. $\mathcal{F}'$-mesurables de carré intégrable.

**Proposition 2.2.2.** *Soient $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ une variable aléatoire réelle sur $(\Omega, \mathcal{F}, \mathbb{P})$ de carré intégrable et $\mathcal{F}'$ une sous-tribu de $\mathcal{F}$. Il existe une unique variable aléatoire réelle $Y$, $\mathcal{F}'$-mesurable et de carré intégrable, telle que*

$$\forall A \in \mathcal{F}', \mathbb{E}[\mathbf{1}_A X] = \mathbb{E}[\mathbf{1}_A Y].$$

*Preuve.* Il s'agit ici de prouver l'existence, l'unicité étant garantie par le Lemme 2.2.1 puisque $L^2(\Omega, \mathcal{F}, \mathbb{P}) \subset L^1(\Omega, \mathcal{F}, \mathbb{P})$. L'espace $L^2(\Omega, \mathcal{F}, \mathbb{P})$ muni du produit scalaire $X \centerdot Y = \mathbb{E}[XY]$ est un espace de Hilbert, et $L^2(\Omega, \mathcal{F}', \mathbb{P})$ est un sous-espace vectoriel fermé de $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Par conséquent, la projection orthogonale $\pi$ sur $L^2(\Omega, \mathcal{F}', \mathbb{P})$ est bien définie. Comme $X - \pi(X) \perp L^2(\Omega, \mathcal{F}', \mathbb{P})$, on a

$$\forall Z \in L^2(\Omega, \mathcal{F}', \mathbb{P}),\ \mathbb{E}[Z(X - \pi(X))] = 0.$$

En particulier, pour $A \in \mathcal{F}'$ on a $\mathbf{1}_A \in L^2(\Omega, \mathcal{F}', \mathbb{P})$ et donc $\mathbb{E}[\mathbf{1}_A X] = \mathbb{E}[\mathbf{1}_A \pi(X)]$. □

L'espérance conditionnelle est en réalité définie pour des v.a.r. plus générales, à savoir les v.a. positives et les v.a. intégrables. Comme pour la construction de l'espérance (ou de l'intégrale), nous allons commencer par construire l'espérance conditionnelle pour les v.a.r. positives, et ensuite étendre la construction aux v.a. intégrables. Nous avons besoin auparavant du résultat suivant.

**Proposition 2.2.3.** *On suppose que $X \geq 0$ p.s. est $\mathcal{F}$-mesurable et que $Y$ est une v.a.r. $\mathcal{F}'$-mesurable telle que $\forall A \in \mathcal{F}', \mathbb{E}[\mathbf{1}_A X] = \mathbb{E}[\mathbf{1}_A Y]$. Alors , $Y \geq 0$ p.s.*

*Preuve.* On considère $A = \{Y < 0\}$ et suppose par l'absurde que $\mathbb{P}(A) > 0$. Alors, puisque $\mathbf{1}_A X$ est une v.a. positive, $0 \leq \mathbb{E}[\mathbf{1}_A X] = \mathbb{E}[\mathbf{1}_A Y] < 0$. □

Une conséquence immédiate de ce résultat est que si deux variables aléatoires réelles intégrables $X_1$, $X_2$ sont telles que $X_1 \leq X_2$ p.s., alors $\mathbb{E}[X_1|\mathcal{F}'] \leq \mathbb{E}[X_2|\mathcal{F}']$ p.s. lorsque ces espérances conditionnelles existent (i.e. lorsque $X_1$ et $X_2$ sont intégrables ou positives).

**Théorème 2.2.4.** *Soient $X$ une variable aléatoire réelle positive sur $(\Omega, \mathcal{F}, \mathbb{P})$ et $\mathcal{F}'$ une sous-tribu de $\mathcal{F}$. Il existe une unique variable aléatoire réelle $Y$, $\mathcal{F}'$-mesurable et positive, telle que*

$$\forall A \in \mathcal{F}', \mathbb{E}[\mathbf{1}_A X] = \mathbb{E}[\mathbf{1}_A Y].$$

*En outre, $Y$ est intégrable si $X$ est intégrable.*

*Preuve.* L'unicité étant assurée par le Lemme 2.2.1, nous prouvons l'existence de $Y$. Pour tout $m \in \mathbb{N}$, $X \wedge m$ est une v.a. bornée et donc de carré intégrable. Elle admet donc une espérance conditionnelle, et on a pour tout $A \in \mathcal{F}'$,

$$\mathbb{E}[\mathbf{1}_A X \wedge m] = \mathbb{E}[\mathbf{1}_A \mathbb{E}[X \wedge m|\mathcal{F}']].$$

Grâce à la proposition précédente, $\mathbb{E}[X \wedge m|\mathcal{F}']$ est une suite croissante de v.a., et on définit $Y = \lim_{m\to+\infty} \mathbb{E}[X \wedge m|\mathcal{F}']$ qui est $\mathcal{F}'$-mesurable comme limite de fonctions $\mathcal{F}'$-mesurables. Par le théorème de convergence monotone, le terme de gauche converge vers $\mathbb{E}[\mathbf{1}_A X]$ tandis que le terme de droite converge vers $\mathbb{E}[\mathbf{1}_A Y]$, et on a donc $\mathbb{E}[\mathbf{1}_A X] = \mathbb{E}[\mathbf{1}_A Y]$ ce qui est le résultat voulu. En outre, si $X$ est intégrable, on a puisque $\Omega \in \mathcal{F}'$, $\mathbb{E}[X] = \mathbb{E}[Y] < +\infty$ et $Y$ est ainsi intégrable. □

Nous sommes désormais en position de prouver l'existence de l'espérance conditionnelle pour les v.a. intégrables.

**Théorème 2.2.5.** *Soient $X$ une variable aléatoire réelle intégrable sur $(\Omega, \mathcal{F}, \mathbb{P})$ et $\mathcal{F}'$ une sous-tribu de $\mathcal{F}$. Il existe une unique variable aléatoire réelle $Y$, $\mathcal{F}'$-mesurable et intégrable, telle que*

$$\forall A \in \mathcal{F}', \mathbb{E}[\mathbf{1}_A X] = \mathbb{E}[\mathbf{1}_A Y].$$

*Preuve.* L'unicité est toujours assurée par le Lemme 2.2.1. Pour l'existence, on utilise la décomposition $X = X^+ - X^-$. Comme $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, $X^+$ et $X^-$ sont également intégrables. Elles sont aussi positives et on définit :

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

Grâce à la proposition précédente, $Y$ est intégrable et satisfait $\forall A \in \mathcal{F}', \mathbb{E}[\mathbf{1}_A X] = \mathbb{E}[\mathbf{1}_A Y]$. □

Dans le cas particulier où $\mathcal{F}' = \{\emptyset, \Omega\}$ est la tribu grossière, et si $X$ est intégrable ou positive, il est facile de voir que $\mathbb{E}[X|\mathcal{F}'] = \mathbb{E}[X]$. Très souvent, l'information connue (la filtration) est donnée par l'observation de variables aléatoires. La définition suivante précise ce point.

**Définition 2.2.6.** *Soit $Y$ une variable aléatoire (resp. $(Y_\theta, \theta \in \Theta)$ une famille de v.a.). Si $X$ est une v.a.r. positive ou intégrable, on définit l'espérance conditionnelle de $X$ sachant $Y$ (resp. $(Y_\theta, \theta \in \Theta)$) comme l'espérance conditionnelle de $X$ sachant la tribu engendrée par $Y$ (resp. $(Y_\theta, \theta \in \Theta)$). On note alors*

$$\mathbb{E}[X|Y] \quad = \quad \mathbb{E}[X|\sigma(Y)] \quad (\textit{resp. } \mathbb{E}[X|Y_\theta, \theta \in \Theta] = \mathbb{E}[X|\sigma(Y_\theta, \theta \in \Theta)]).$$

**Remarque 2.2.7.** *Lorsque la tribu $\mathcal{F} = \sigma(\mathcal{A})$, on peut montrer que*

$$\forall A \in \mathcal{F}',\ \mathbb{E}[\mathbf{1}_A X] = \mathbb{E}[\mathbf{1}_A Z] \iff \forall A \in \mathcal{A},\ \mathbb{E}[\mathbf{1}_A X] = \mathbb{E}[\mathbf{1}_A Z].$$

*En particulier, si $X$ est une v.a.r. positive ou intégrable et $Y$ est une v.a. réelle,*

$$Z = \mathbb{E}[X|Y] \iff \forall a < b,\ \mathbb{E}[X\mathbf{1}_{\{Y\in(a,b)\}}] = \mathbb{E}[Z\mathbf{1}_{\{Y\in(a,b)\}}].$$

**Exercice 2.2.8** (lien avec les probabilités conditionnelles)**.** *On se place sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, et on considère $A, B \in \mathcal{F}$. On rappelle que la probabilité de $A$ sachant $B$ vaut $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$ lorsque $\mathbb{P}(B) \neq 0$ et $\mathbb{P}(A|B) = \mathbb{P}(A)$ si $\mathbb{P}(B) = 0$. Montrer que, presque sûrement, $\mathbb{E}[\mathbf{1}_A|\mathbf{1}_B] = \mathbb{P}(A|B)\mathbf{1}_B + \mathbb{P}(A|B^c)\mathbf{1}_{B^c}$ où $B^c = \Omega \backslash B$.*

De manière générale, lorsque $X$ est une v.a.r. positive ou intégrable et $Y$ est une v.a. discrète à valeurs dans un espace $F$, on a :

$$\mathbb{E}[X|Y] = \sum_{y\in F, \mathbb{P}(Y=y)>0} \frac{\mathbb{E}[X\mathbf{1}_{\{Y=y\}}]}{\mathbb{P}(Y=y)} \mathbf{1}_{\{Y=y\}}. \tag{2.2}$$

La preuve de ce résultat est laissée en exercice. On note $\mathbb{E}[X|Y=y] = \frac{\mathbb{E}[X\mathbf{1}_{\{Y=y\}}]}{\mathbb{P}(Y=y)}$ l'espérance de $X$ sachant que $Y = y$. Attention, cette quantité est, contrairement à l'espérance conditionnelle, déterministe : elle dépend de $y$ mais pas de la réalisation de $Y$.

**Exercice 2.2.9.** *Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. i.i.d. intégrables. On pose $S_n = \sum_{i=1}^n X_i$.*

1. *On prend $n \geq 2$. Montrer que $\mathbb{E}[X_1|S_n] = \mathbb{E}[X_2|S_n]$.*
2. *En déduire que $\mathbb{E}[X_1|S_n] = S_n/n$ pour $n \geq 1$, puis donner le comportement asymptotique de $\mathbb{E}[X_1|S_n]$. Ce résultat vous surprend-il ?*
3. *Donner un exemple de v.a. indépendantes $X$, $Y$ et de tribu $\mathcal{F}'$ telles que $\mathbb{E}[X|\mathcal{F}']$ et $\mathbb{E}[Y|\mathcal{F}']$ ne soient pas indépendantes.*

### 2.2.2 Propriétés de l'espérance conditionnelle

Ce qu'il faut retenir du paragraphe précédent sur la construction de l'espérance conditionnelle sont les deux choses suivantes. Comme l'espérance, l'espérance conditionnelle d'une variable aléatoire est définie dès lors qu'elle est positive ou intégrable. En outre, si cette variable est de carré intégrable, l'espérance conditionnelle s'interprète comme la projection orthogonale sur le sous espace vectoriel des v.a.r. de carré intégrable mesurables par rapport à la filtration à laquelle on effectue le conditionnement. D'un point de vue pratique et calculatoire, ce seront ensuite les propriétés énoncées dans ce paragraphe qui seront utiles.

On se place sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathcal{F}'$ désigne une sous-tribu de $\mathcal{F}$ et $X$ est une v.a. réelle positive ou intégrable.

1. $X$ est $\mathcal{F}'$-mesurable si et seulement si $\mathbb{E}[X|\mathcal{F}'] = X$ p.s.
2. Si $X$ est indépendante de $\mathcal{F}'$, $\mathbb{E}[X|\mathcal{F}'] = \mathbb{E}[X]$.
3. **Positivité :** si $X \geq 0$, $\mathbb{E}[X|\mathcal{F}'] \geq 0$. Si $X_1, X_2 \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ sont telles que $X_1 \geq X_2$ p.s., $\mathbb{E}[X_1|\mathcal{F}'] \geq \mathbb{E}[X_2|\mathcal{F}']$ p.s. Ces propriétés sont également vraies en remplaçant les inégalités larges par des inégalités strictes.
4. **Linéarité :** Si $\lambda_1, \lambda_2 \in \mathbb{R}$, $X_1, X_2 \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbb{E}[\lambda_1 X + \lambda_2 X_2|\mathcal{F}'] = \lambda_1 \mathbb{E}[X_1|\mathcal{F}'] + \lambda_2 \mathbb{E}[X_2|\mathcal{F}']$.
5. **Double conditionnement :** Si $\mathcal{F}''$ est une sous-tribu de $\mathcal{F}'$, $\mathbb{E}[X|\mathcal{F}''] = \mathbb{E}[\mathbb{E}(X|\mathcal{F}')|\mathcal{F}'']$. En particulier on a pour $\mathcal{F}'' = \{\emptyset, \Omega\}$, $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}(X|\mathcal{F}')]$.
6. **Inégalité de Jensen :** si $\phi : \mathbb{R} \to \mathbb{R}$ est une fonction convexe, et si $X, \phi(X) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$,
$$\mathbb{E}[\phi(X)|\mathcal{F}'] \geq \phi(\mathbb{E}[X|\mathcal{F}']) \; p.s.$$
Deux cas particuliers importants :
si $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbb{E}[|X||\mathcal{F}'] \geq |\mathbb{E}[X|\mathcal{F}']|$,
si $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, $\mathbb{E}[X^2|\mathcal{F}'] \geq \mathbb{E}[X|\mathcal{F}']^2$.
7. Si $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ et $Z$ est une v.a.r. bornée $\mathcal{F}'$-mesurable, $\mathbb{E}[ZX|\mathcal{F}'] = Z\mathbb{E}[X|\mathcal{F}']$.
8. **Théorème de convergence monotone :** soit $(X_n, n \in \mathbb{N})$ une suite de v.a.r. positives telles que $\forall n,\ X_n \leq X_{n+1}$ p.s. Soit $X = \lim_{n\to+\infty} X_n$. Alors
$$\mathbb{E}[X_n|\mathcal{F}'] \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[X|\mathcal{F}'] \; p.s.$$
9. **Lemme de Fatou :** si $(X_n, n \in \mathbb{N})$ est une suite de v.a.r. positives, on a
$$\mathbb{E}[\liminf_{n\to+\infty} X_n|\mathcal{F}'] \leq \liminf_{n\to+\infty} \mathbb{E}[X_n|\mathcal{F}'] \; p.s.$$

10. **Théorème de convergence dominée :** soit $(X_n, n \in \mathbb{N})$ une suite de v.a.r. uniformément bornée par une v.a.r. intégrable (i.e. $\exists V \in L^1(\Omega, \mathcal{F}, \mathbb{P}), \forall n \in \mathbb{N}, |X_n| \leq V$ *p.s.*) et qui converge p.s. vers $X$. Alors,

$$\mathbb{E}[X_n|\mathcal{F}'] \underset{n\to+\infty}{\to} \mathbb{E}[X|\mathcal{F}'] \ p.s. \text{ et dans } L^1(\Omega, \mathcal{F}, \mathbb{P}).$$

Insistons une dernière fois sur le fait que l'espérance conditionnelle jouit des mêmes propriétés que l'espérance (linéarité, positivité, inégalité de Jensen, théorèmes de convergence dominée et monotone et lemme de Fatou). Ce sont donc formellement les mêmes objets : l'espérance conditionnelle est bien une espérance !

*Preuve.* La preuve des points 1 à 5 ne pose pas de difficulté et est laissée en exercice. Celle du point 6 fait l'objet de l'exercice 2.2.10.

7 : Tout d'abord, l'égalité est évidente si $Z$ est l'indicatrice d'un événement. Comme $X$ est intégrable, on voit alors que l'égalité reste vraie si $Z$ est une combinaison linéaire finie d'indicatrices. Enfin pour une v.a. bornée générale $|Z| \leq K$, on peut se ramener à $Z \geq 0$ en considérant $Z + K$, puis on utilise l'approximation par des fonctions étagées comme pour la construction de l'intégrale (Section 1.1.2), et on obtient le résultat en passant à la limite.

8 : Grâce à la propriété de positivité, $0 \leq \mathbb{E}[X_n|\mathcal{F}'] \leq \mathbb{E}[X_{n+1}|\mathcal{F}']$ p.s. et ainsi la limite $\lim_{n\to+\infty} \mathbb{E}[X_n|\mathcal{F}']$ est bien définie p.s. Soit $A \in \mathcal{F}'$. On a $\mathbb{E}[\mathbf{1}_A X_n] = \mathbb{E}[\mathbf{1}_A \mathbb{E}[X_n|\mathcal{F}']]$, et par le théorème de convergence monotone pour l'espérance, $\mathbb{E}[\mathbf{1}_A X] = \mathbb{E}[\mathbf{1}_A \lim_{n\to+\infty} \mathbb{E}[X_n|\mathcal{F}']]$. D'après le lemme 2.2.1, on en déduit $\mathbb{E}[X|\mathcal{F}'] = \lim_{n\to+\infty} \mathbb{E}[X_n|\mathcal{F}']$ p.s.

9 : découle imédiatement de 8.

10 : Il s'agit de la preuve usuelle qui permet de déduire du Lemme de Fatou le théorème de convergence dominée. On a $|X| \leq V$ p.s. et $|X_n - X| \leq 2V$ p.s. Par conséquent, $\mathbb{E}[\liminf_{n\to+\infty}(2V - |X - X_n|)|\mathcal{F}'] \leq \liminf_{n\to+\infty} \mathbb{E}[(2V - |X - X_n|)|\mathcal{F}']$. On en déduit que $\mathbb{E}[\limsup_{n\to+\infty}|X - X_n||\mathcal{F}'] \geq \limsup_{n\to+\infty} \mathbb{E}[|X - X_n||\mathcal{F}']$ p.s. Comme $\limsup_{n\to+\infty}|X - X_n| = 0$, on obtient $\limsup_{n\to+\infty} \mathbb{E}[|X - X_n||\mathcal{F}'] = 0$ p.s. et donc $\mathbb{E}[|X - X_n||\mathcal{F}'] \underset{n\to+\infty}{\to} 0$. En appliquant l'inégalité de Jensen, on en déduit $|\mathbb{E}[X|\mathcal{F}'] - \mathbb{E}[X_n|\mathcal{F}']| \underset{n\to+\infty}{\to} 0$ p.s. Enfin, puisque $|\mathbb{E}[X|\mathcal{F}'] - \mathbb{E}[X_n|\mathcal{F}']| \leq 2\mathbb{E}[V|\mathcal{F}']$ et $\mathbb{E}[V|\mathcal{F}']$ est intégrable, on conclut par le théorème de convergence dominée usuel que $\mathbb{E}[|\mathbb{E}[X|\mathcal{F}'] - \mathbb{E}[X_n|\mathcal{F}']|] \underset{n\to+\infty}{\to} 0$, c'est à dire que $\mathbb{E}[X_n|\mathcal{F}']$ converge vers $\mathbb{E}[X|\mathcal{F}']$ dans $L^1(\Omega, \mathcal{F}, \mathbb{P})$. □

**Exercice 2.2.10.** *Soit $\phi : \mathbb{R} \to \mathbb{R}$ une fonction convexe. Montrer qu'il existe une suite $(a_n, b_n)$ de couples de réels telle que $\forall x \in \mathbb{R}, \phi(x) = \sup_{n\in\mathbb{N}} a_n x + b_n$. En déduire l'inégalité de Jensen pour l'espérance conditionnelle (point 6).*

**Exercice 2.2.11.** *On considère un vecteur gaussien* $(X_1, X_2)$ *centré de matrice de covariance* $\Gamma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$ *avec* $\sigma_1, \sigma_2 > 0$ *et* $\rho \in [-1, 1]$. *Montrer que* $X_1 - \rho\frac{\sigma_1}{\sigma_2}X_2$ *et* $X_2$ *sont indépendantes. En déduire* $\mathbb{E}[X_1|X_2]$ *puis* $\mathbf{Var}(X_1|X_2) := \mathbb{E}[X_1^2|X_2] - \mathbb{E}[X_1|X_2]^2$.

**Exercice 2.2.12.** *La propriété 2 assure que si* $X \perp\!\!\!\perp \mathcal{F}'$, *alors* $\mathbb{E}[X|\mathcal{F}'] = \mathbb{E}[X]$. *Le but de l'exercice est de montrer que la réciproque est fausse. On considère une v.a.r.* $Z$ *de mesure de probabilité sur* $\mathbb{R}$ : $\frac{1}{2}\delta_{-1}(dx) + \frac{1}{2}e^{-x}\mathbf{1}_{\{x>0\}}dx$. *On pose* $X = |Z|$ *et* $Y = \mathbf{1}_{Z>0}$. *Montrer que* $\mathbb{E}[X|Y] = \mathbb{E}[X]$ *mais que* $X$ *et* $Y$ *ne sont pas indépendants.*

Une conséquence de la propriété 7 est la suivante.

**Corollaire 2.2.13.** *Soient* $X$ *une v.a.r. intégrable et* $\Xi$ *une v.a.r. intégrable* $\mathcal{F}'$*-mesurable. Alors,*

$$\Xi = \mathbb{E}[X|\mathcal{F}'] \iff \forall Z \text{ v.a.r. bornée } \mathcal{F}'\text{-mesurable}, \mathbb{E}[XZ] = \mathbb{E}[\Xi Z].$$

Lorsque l'on conditionne par rapport à une variable aléatoire (resp. une famille de v.a.), on a également la formulation suivante, souvent plus commode.

**Corollaire 2.2.14.** *Soit* $Y$ *une variable aléatoire (resp.* $(Y_\theta, \theta \in \Theta)$ *une famille de v.a.). Soient* $X$ *une v.a.r. intégrable et* $\Xi$ *une v.a.r. intégrable* $\sigma(Y)$*-mesurable (resp.* $\sigma(Y_\theta, \theta \in \Theta)$*-mesurable). Alors,*

$$\begin{aligned} \Xi &= \mathbb{E}[X|Y] \iff \forall\varphi \text{ fonction mesurable réelle bornée}, \mathbb{E}[X\varphi(Y)] = \mathbb{E}[\Xi\varphi(Y)] \\ (\text{resp. } \Xi &= \mathbb{E}[X|Y_\theta, \theta \in \Theta] \iff \forall\varphi \text{ mes. bornée } \mathbb{E}[X\varphi(Y_\theta, \theta \in \Theta)] = \mathbb{E}[\Xi\varphi(Y_\theta, \theta \in \Theta)]). \end{aligned}$$

Toujours concernant la propriété 7, il est important de noter qu'elle est vraie aussi pour des v.a. non bornées dès lors que $ZX$ est intégrable. En effet, on a alors pour tout $n \in \mathbb{N}$,

$$\mathbb{E}[Z\mathbf{1}_{\{|Z|\leq n\}}X|\mathcal{F}'] = Z\mathbf{1}_{\{|Z|\leq n\}}\mathbb{E}[X|\mathcal{F}'].$$

Le terme de droite converge p.s. vers $Z\mathbb{E}[X|\mathcal{F}']$ tandis que celui de gauche converge p.s. vers $\mathbb{E}[ZX|\mathcal{F}']$ grâce au théorème de convergence dominée pour l'espérance conditionnelle. On déduit ainsi la proposition suivante.

**Proposition 2.2.15.** *Si* $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ *est une v.a.r.* $\mathcal{F}$*-mesurable et si* $Z$ *est une v.a.r.* $\mathcal{F}'$*-mesurable telle que* $ZX \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, *on a :*

$$\mathbb{E}[ZX|\mathcal{F}'] = Z\mathbb{E}[X|\mathcal{F}'].$$

En toute généralité, le calcul d'une espérance conditionnelle n'est pas aisé et on doit souvent avoir recours à des méthodes numériques pour la calculer. Néanmoins il existe quelques cas particuliers où l'on est capable de calculer explicitement l'espérance conditionnelle. La proposition suivante donne un cas en général très utile pour mener à bien des calculs.

**Proposition 2.2.16.** *Soient $X_1$ et $X_2$ deux variables aléatoires à valeurs respectivement dans les espaces mesurables $E_1$ et $E_2$ que l'on suppose indépendantes. Soit $f : E_1 \times E_2 \to \mathbb{R}$ une fonction borélienne telle que $f(X_1, X_2)$ est positive ou intégrable. Alors,*

$$\mathbb{E}[f(X_1, X_2)|X_2] = \psi(X_2)$$

*où $\psi(x) = \mathbb{E}[f(X_1, x)]$. Plus généralement, si $\mathcal{F}'$ est une tribu telle que $X_2$ est $\mathcal{F}'$-mesurable et $X_1$ est indépendante de $\mathcal{F}'$, on a également $\mathbb{E}[f(X_1, X_2)|\mathcal{F}'] = \psi(X_2)$.*

*Preuve.* Soit $A$ un événement de $\mathcal{F}'$. On a d'une part $\mathbb{E}[\mathbf{1}_A \psi(X_2)] = \int_\Omega \mathbf{1}_A(\omega)\psi(X_2(\omega))\mathbb{P}(d\omega)$ et d'autre part $\psi(x) = \int_\Omega f(X_1(\omega), x)\mathbb{P}(d\omega)$. Ainsi, en notant $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ une copie de $(\Omega, \mathcal{F}, \mathbb{P})$,

$$\begin{aligned} \mathbb{E}[\mathbf{1}_A \psi(X_2)] &= \int_\Omega \int_{\tilde{\Omega}} \mathbf{1}_A(\omega) f(X_1(\tilde{\omega}), X_2(\omega))\mathbb{P}(d\omega)\mathbb{P}(d\tilde{\omega}) \\ &= \int_\Omega \mathbf{1}_A(\omega) f(X_1(\omega), X_2(\omega))\mathbb{P}(d\omega) \text{ (par indépendance)} \\ &= \mathbb{E}[\mathbf{1}_A f(X_1, X_2)] \end{aligned}$$

On conclut que $\mathbb{E}[f(X_1, X_2)|X_2] = \psi(X_2)$ en utilisant la définition de l'espérance conditionnelle. □

**Remarque 2.2.17.** *On peut montrer de manière générale que toute variable aléatoire $Z$ mesurable par rapport à $X_2$ s'écrit sous la forme $g(X_2)$ pour une fonction $g$ mesurable. En particulier si $Y$ est une v.a. positive ou intégrable, il existe une fonction mesurable $g$ t.q. $\mathbb{E}[Y|X_2] = g(X_2)$. En toute généralité, il est impossible d'expliciter une telle fonction. La proposition précédente est donc un cas particulier où on a une écriture simple de $g$.*

Avant de conclure cette section, nous allons définir l'espérance conditionnelle pour une variable aléatoire à valeurs dans $\mathbb{R}^d$.

**Définition 2.2.18.** *Soit $X : \Omega \to \mathbb{R}^d$ une v.a. intégrable et $\mathcal{F}'$ une sous-tribu de $\mathcal{F}$. On définit $\mathbb{E}(X|\mathcal{F}') : \Omega \to \mathbb{R}^d$ comme l'espérance conditionnelle de chaque coordonnée, i.e. $\mathbb{E}(X|\mathcal{F}')_i = \mathbb{E}(X_i|\mathcal{F}')$ pour $i = 1, \dots, d$.*

**Exercice 2.2.19.** *On se place sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et on considère $\mathcal{F}'$ une sous-tribu de $\mathcal{F}$. Soit $N$ une variable aléatoire $\mathcal{F}'$-mesurable à valeurs entières et finie p.s. On se donne une suite de v.a. réelles $(X_i)_{i\in\mathbb{N}}$ et on pose $S = \sum_{i=0}^N X_i$. Montrer que si $S$ et les $X_i$ sont intégrables, ou si les $X_i$ sont des v.a. positives, $\mathbb{E}[S|\mathcal{F}']$ est bien définie et vaut :*

$$\mathbb{E}[S|\mathcal{F}'] = \sum_{i=0}^N \mathbb{E}[X_i|\mathcal{F}'].$$

*(Indication : revenir à la définition de l'espérance conditionnelle et décomposer selon les valeurs prises par $N$ : $N = \sum_{n\in\mathbb{N}} n\mathbf{1}_{\{N=n\}}$.)*

## 2.3 Loi conditionnelle

L'espérance conditionnelle, nous l'avons vu, est une moyenne qui tient compte des événements connus. De manière plus générale, il est intéressant de décrire comment est distribuée une v.a. une fois que l'on a eu accès à une certaine information. C'est pourquoi on introduit la notion de loi conditionnelle. Nous nous plaçons toujours sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

**Définition 2.3.1.** *Soit $X$ une variable aléatoire à valeurs dans un espace mesurable $E$ et $\mathcal{F}'$ une sous-tribu de $\mathcal{F}$. On appelle loi conditionnelle de $X$ sachant $\mathcal{F}'$ la loi caractérisée par :*

$$\{\mathbb{E}[\varphi(X)|\mathcal{F}'], \varphi : E \to \mathbb{R} \text{ borélienne bornée}\}.$$

*De même, la loi conditionnelle de $X$ sachant une ou plusieurs v.a. est la loi conditionnelle sachant la tribu qu'elles engendrent.*

**Exercice 2.3.2.** *Montrer que $X$ est indépendant de $\mathcal{F}'$ si et seulement si la loi de $X$ est la même que la loi de $X$ conditionnellement à $\mathcal{F}'$, c'est à dire si $\forall \varphi : E \to \mathbb{R}$ borélienne bornée, $\mathbb{E}[\varphi(X)|\mathcal{F}'] = \mathbb{E}[\varphi(X)|]$.*

Donnons un exemple simple. Soit $U$ une v.a. uniforme sur $[0, 1]$ et $G$ une v.a. gaussienne standard indépendante de $U$. On veut connaître la loi de $X = GU$ sachant $U$. On considère $\varphi : \mathbb{R} \to \mathbb{R}$ borélienne bornée. On a d'après la proposition 2.2.16, $\mathbb{E}[\varphi(X)|U] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \varphi(Uy) \exp(-y^2/2)dy = \frac{1}{\sqrt{2\pi U^2}} \int_{\mathbb{R}} \varphi(x) \exp(-x^2/(2U^2))dx$. Par conséquent, la loi conditionnelle de $X$ sachant $U$ est une loi normale centrée de variance $U^2$.

**Exercice 2.3.3.** *Montrer que la loi conditionnelle de $X$ sachant $G$ est une loi uniforme sur $[0, G]$ si $G \geq 0$ et sur $[G, 0]$ si $G < 0$.*

De manière plus générale, on considère un couple de v.a.r $(X, Y)$ qui admet une loi de densité $p(x, y)$. Rappelons que dans ce cas, $Y$ est une v.a.r. distribuée selon la densité $q(y) = \int_{\mathbb{R}} p(x, y)dx$. On souhaite calculer la loi de $X$ sachant $Y$. Soient $\varphi, \psi : \mathbb{R} \to \mathbb{R}$ deux fonctions boréliennes bornées. Nous commençons par calculer l'espérance conditionnelle $\mathbb{E}[\varphi(X)|Y]$. On a

$$\begin{aligned}
\mathbb{E}[\varphi(X)\psi(Y)] &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \varphi(x)\psi(y)p(x, y)dxdy \\
&= \int_{-\infty}^{+\infty} \psi(y) \left( \int_{-\infty}^{+\infty} \varphi(x) \frac{p(x, y)}{q(y)} dx \right) q(y)dy
\end{aligned}$$

En utilisant le Corollaire 2.2.14, il vient $\mathbb{E}[\varphi(X)|Y] = \int_{-\infty}^{+\infty} \varphi(x) \frac{p(x,Y)}{q(Y)} dx$ et la loi de $X$ sachant $Y$ est la loi distribuée selon la densité $\frac{p(x,Y)}{q(Y)}$.

**Exercice 2.3.4.** *Soient $\xi_1$ et $\xi_2$ deux variables aléatoires exponentielles indépendantes de paramètre* $1$. *Calculer la loi de $\xi_1$ sachant $S = \xi_1 + \xi_2$.*

**Exercice 2.3.5.** *On considère un vecteur gaussien $(X_1, X_2)$ centré de matrice de covariance $\Gamma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$ avec $\sigma_1, \sigma_2 > 0$ et $\rho \in [-1, 1]$. En utilisant la fonction caractéristique de $(X_1, X_2)$ montrer que la loi de $X_1$ sachant $X_2$ est une gaussienne de moyenne $\rho\frac{\sigma_1}{\sigma_2}X_2$ et de variance $\sigma_1^2(1-\rho^2)$. Retrouver ainsi les résultats de l'Exercice 2.2.11.*

**Exercice 2.3.6.** *On considère une v.a.r. $X$ et $Y$ une v.a. discrète à valeurs dans $F$. La loi de $(X, Y)$, donnée en toute généralité par une mesure $\mu(dx, dy)$, s'écrit également $\sum_{y \in F} \mu_y(dx)$ puisque $Y$ est une v.a. discrète. A l'aide de* (2.2)*, montrer que la loi de $X$ sachant $Y$ est la loi de mesure $\mu_Y(dx)$.*

# Chapitre 3

# Martingales en temps discret

## 3.1 Premières définitions

**Définition 3.1.1.** *On appelle filtration (discrète) de* $(\Omega, \mathcal{F}, \mathbb{P})$ *une suite croissante* $(\mathcal{F}_n)_{n\in\mathbb{N}}$ *de sous tribus de* $\mathcal{F}$ *:* $\forall n \in \mathbb{N},\ \mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}$.

*Une suite de variables aléatoires* $(X_n, n \in \mathbb{N})$ *est adaptée à la filtration* $(\mathcal{F}_n)_{n\in\mathbb{N}}$ *si pour tout* $n \in \mathbb{N}$*, la variable aléatoire* $X_n$ *est* $\mathcal{F}_n$*-mesurable. On dit alors que* $(X_n, n \in \mathbb{N})$ *est une suite* $(\mathcal{F}_n)$*-adaptée, ou encore que* $(X_n, n \in \mathbb{N})$ *est un processus* $(\mathcal{F}_n)$*-adapté.*

Lorsque $(\mathcal{F}_n)_{n\in\mathbb{N}}$ et $(\mathcal{F}'_n)_{n\in\mathbb{N}}$ sont deux filtrations de $(\Omega, \mathcal{F}, \mathbb{P})$, on dit que $(\mathcal{F}'_n)_{n\in\mathbb{N}}$ est une *sous-filtration* de $(\mathcal{F}_n)_{n\in\mathbb{N}}$ si $\forall n \in \mathbb{N}, \mathcal{F}'_n \subset \mathcal{F}_n$. Donnons un exemple important de filtration. Si $(X_n, n \in \mathbb{N})$ est une suite de variables aléatoires, la filtration $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$ est appelée *la filtration engendrée par* $(X_n, n \in \mathbb{N})$. C'est la plus petite filtration de $(\Omega, \mathcal{F}, \mathbb{P})$ qui rende la suite $(X_n, n \in \mathbb{N})$ adaptée : elle est sous-filtration de toute filtration rendant $(X_n, n \in \mathbb{N})$ adaptée.

En toute généralité, un processus $(X_n, n \in \mathbb{N})$ est donc une v.a. $\begin{array}{l} \Omega \to \mathbb{R}^{\mathbb{N}} \\ \omega \mapsto (X_n(\omega))_{n\in\mathbb{N}} \end{array}$. On munit $\mathbb{R}^{\mathbb{N}}$ de la tribu engendrée par $\cup_{n\in\mathbb{N}}\{A_0 \times \dots \times A_n \times \mathbb{R}^{\mathbb{N}},\ A_i$ boréliens de $\mathbb{R}\}$. Par conséquent, la loi d'un processus est déterminée en toute généralité par

$$\{\mathbb{E}[f(X_n, n \in \mathbb{N})], f : \mathbb{R}^{\mathbb{N}} \to \mathbb{R} \text{ bornée mesurable}\}.$$

Un résultat théorique important, le théorème de Kolmogorov assure que l'on peut se restreindre à des fonctions fini-dimensionnelles, c'est à dire que la loi est caractérisée par

$$\{\mathbb{E}[f(X_0, \dots, X_n)], n \in \mathbb{N}, f : \mathbb{R}^{n+1} \to \mathbb{R} \text{ bornée mesurable}\}.$$

D'un point de vue de la modélisation, un processus $(X_n, n \in \mathbb{N})$ décrit l'évolution d'un phénomène aléatoire. En général, cette évolution est en fonction du temps, si bien que l'indice $n$ est souvent appelé temps. La filtration engendrée par $(X_n, n \in \mathbb{N})$, $\sigma(X_0, \ldots, X_n)$, représente l'information donnée par les observations de ce processus jusqu'à l'instant $n$, tandis que la filtration $\mathcal{F}_n$ décrit toute l'information accessible à l'instant $n$.

**Définition 3.1.2.** *Un processus* $(X_n, n \in \mathbb{N})$ *à valeurs réelles ou dans* $\mathbb{R}^d$ *est dit :*
- *intégrable si* $\forall n \in \mathbb{N}, X_n \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.
- *de carré intégrable si* $\forall n \in \mathbb{N}, X_n \in L^2(\Omega, \mathcal{F}, \mathbb{P})$.
- *à valeurs bornées si* $\forall n \in \mathbb{N}, X_n \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$.

**Définition 3.1.3.** *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité muni d'une filtration* $(\mathcal{F}_n)_{n\in\mathbb{N}}$. *La suite de v.a. réelles* $(M_n, n \in \mathbb{N})$ *est une* $(\mathcal{F}_n)$*-martingale si* $\forall n \in \mathbb{N},\ M_n \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ *et* $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n$.

On en déduit immédiatement que pour une martingale, $\mathbb{E}[M_{n+p}|\mathcal{F}_n] = M_n$ pour $p, n \in \mathbb{N}$, et en particulier l'espérance est constante : $\mathbb{E}[M_{n+p}] = \mathbb{E}[M_n]$. Donnons deux exemples élémentaires de martingales. Si $(U_n)_{n\in\mathbb{N}^*}$ (resp. $(V_n)_{n\in\mathbb{N}^*}$) est une suite de v.a. réelles indépendantes intégrables (resp. bornées) telle que $E(U_n) = 0$ (resp. $E(V_n) = 1$), alors $S_n = \sum_{i=1}^n U_i$ avec $S_0 = 0$ (resp. $\Pi_n = \prod_{i=1}^n V_i$ avec $\Pi_0 = 1$) est une martingale par rapport à la filtration $\mathcal{F}_n = \sigma(U_1, \ldots, U_n)$ (resp. $\mathcal{F}_n = \sigma(V_1, \ldots, V_n)$), $\mathcal{F}_0$ désignant la filtration grossière. En effet, on vérifie facilement que le processus $(U_n, n \in \mathbb{N})$ (resp. $(V_n, n \in \mathbb{N})$) est intégrable, et on a $E(S_{n+1}|\mathcal{F}_n) = E(S_n + U_{n+1}|\mathcal{F}_n) = S_n + E(U_{n+1}|\mathcal{F}_n) = S_n$ (resp. $E(\Pi_{n+1}|\mathcal{F}_n) = E(\Pi_n V_{n+1}|\mathcal{F}_n) = \Pi_n E(V_{n+1}|\mathcal{F}_n) = \Pi_n$).

On introduit également la notion de surmartingale et de sous-martingale.

**Définition 3.1.4.** *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité muni d'une filtration* $(\mathcal{F}_n)_{n\in\mathbb{N}}$. *La suite de v.a. réelles* $(M_n, n \in \mathbb{N})$ *est une surmartingale (resp sous-martingale) si* $\forall n \in \mathbb{N},\ M_n \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ *et* $\mathbb{E}[M_{n+1}|\mathcal{F}_n] \leq M_n$ *(resp.* $\mathbb{E}[M_{n+1}|\mathcal{F}_n] \geq M_n$*).*

Pour se souvenir du sens de l'inégalité, il est utile de remarquer que l'inégalité n'est pas dans le sens dans lequel on pourrait s'attendre. Lorsque $(M_n, n \in \mathbb{N})$ est une surmartingale (resp. sous-martingale), on a également $\mathbb{E}[M_{n+p}|\mathcal{F}_n] \leq M_n$ (resp. $\mathbb{E}[M_{n+p}|\mathcal{F}_n] \geq M_n$) pour $p \in \mathbb{N}$.

**Exercice 3.1.5.**
1. *Montrer que si* $(M_n, n \in \mathbb{N})$ *est une martingale et* $\phi$ *une fonction convexe (resp. concave) telle que* $\forall n \in \mathbb{N}, \phi(M_n) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, *alors* $X_n = \phi(M_n)$ *est une sous-martingale (resp. surmartingale).*
2. *On suppose désormais seulement que* $(M_n, n \in \mathbb{N})$ *est une sous-martingale (resp. surmartingale). Montrer que* $X_n = \phi(M_n)$ *est une sous-martingale (resp. surmartingale) si* $\phi$ *est convexe (resp. concave) croissante.*

Enfin, on définit une martingale à valeurs dans $\mathbb{R}^d$ de manière tout à fait analogue aux martingales réelles.

**Définition 3.1.6.** *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité muni d'une filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$. La suite $(M_n, n \in \mathbb{N})$ de v.a. à valeurs dans $\mathbb{R}^d$ est une martingale si $\forall n \in \mathbb{N}$, $M_n \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ et $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n$.*

# 3.2 Constructions usuelles de martingales

Avant de présenter trois importantes constructions de martingales, nous introduisons la notion de suite prévisible.

**Définition 3.2.1.** *Une suite de variables aléatoires $(X_n, n \in \mathbb{N})$ est un processus prévisible par rapport à la filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ si pour tout $n \in \mathbb{N}^*$, la variable aléatoire $X_n$ est $\mathcal{F}_{n-1}$-mesurable et si $\exists x_0, \mathbb{P}(X_0 = x_0) = 1$.*

## 3.2.1 Martingales de carré intégrable, martingales exponentielles

**Définition 3.2.2.** *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité muni d'une filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$. Une martingale $(M_n, n \in \mathbb{N})$ (réelle ou à valeurs dans $\mathbb{R}^d$) est dite de carré intégrable si $\forall n \in \mathbb{N}$, $M_n \in L^2(\Omega, \mathcal{F}, \mathbb{P})$.*

**Proposition 3.2.3.** *Soit $(M_n, n \in \mathbb{N})$ une martingale réelle de carré intégrable sur $(\Omega, \mathcal{F}, \mathbb{P})$ muni de la filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$. Il existe une unique suite prévisible croissante $(C_n)_{n\in\mathbb{N}}$ de v.a.r. telle que*

$$M_n^2 - C_n \text{ est une } (\mathcal{F}_n)\text{-martingale}$$

*et $C_0 = 0$. On note $[M]_n$ ou $[M, M]_n$ ce processus appelé crochet de la martingale $(M_n, n \in \mathbb{N})$. Il est défini par :*

$$[M]_n = [M, M]_n = \sum_{k=0}^{n-1} \mathbb{E}[(M_{k+1} - M_k)^2|\mathcal{F}_k].$$

*Preuve.* Nous avons $M_{n+1}^2 = M_n^2 + 2M_n(M_{n+1} - M_n) + (M_{n+1} - M_n)^2$. Par linéarité et puisque $M_n$ est $\mathcal{F}_n$-mesurable,

$$\begin{aligned} \mathbb{E}[M_{n+1}^2|\mathcal{F}_n] &= M_n^2 + 2M_n\mathbb{E}[M_{n+1} - M_n|\mathcal{F}_n] + \mathbb{E}[(M_{n+1} - M_n)^2|\mathcal{F}_n] \\ &= M_n^2 + \mathbb{E}[(M_{n+1} - M_n)^2|\mathcal{F}_n]. \end{aligned} \tag{3.1}$$

Par ailleurs, si $(C_n)_{n\in\mathbb{N}}$ est une suite prévisible t.q. $M_n^2 - C_n$ est une $(\mathcal{F}_n)$-martingale, on a $\mathbb{E}[M_{n+1}^2 - C_{n+1}|\mathcal{F}_n] = M_n^2 - C_n$ et $\mathbb{E}[M_{n+1}^2 - C_{n+1}|\mathcal{F}_n] = \mathbb{E}[M_{n+1}^2|\mathcal{F}_n] - C_{n+1}$. En combinant avec (3.1), on obtient $C_{n+1} - C_n = \mathbb{E}[(M_{n+1} - M_n)^2|\mathcal{F}_n]$ ce qui donne à la fois l'existence et l'unicité du crochet, ainsi que sa croissance grâce à la propriété de positivité de l'espérance conditionnelle. □

L'exercice suivant est une généralisation de cette construction.

**Exercice 3.2.4.** *Soit* $(X_n, n \in \mathbb{N})$ *une sous-martingale réelle.*

1. *Montrer que* $C_n = \sum_{k=0}^{n-1} \mathbb{E}[X_{k+1}|\mathcal{F}_k] - X_k$ *définit l'unique suite croissante prévisible telle que* $(X_n - C_n, n \in \mathbb{N})$ *est une* $(\mathcal{F}_n)$*-martingale et* $C_0 = 0$*. Vérifier que ce processus est intégrable.*
2. *En utilisant l'exercice 3.1.5, retrouver le résultat de la Proposition 3.2.3.*

On considère désormais une martingale réelle $(M_n, n \in \mathbb{N})$ sur $(\Omega, \mathcal{F}, \mathbb{P})$ telle que $\forall n \in \mathbb{N}, \mathbb{E}[e^{M_n}] < \infty$. On cherche de manière analogue une suite $(C_n, n \in \mathbb{N})$ $(\mathcal{F}_n)$-prévisible telle que $e^{M_n - C_n}$ soit une martingale. Alors, on a d'une part $\mathbb{E}[e^{M_{n+1}}|\mathcal{F}_n] = e^{M_n}\mathbb{E}[e^{M_{n+1}-M_n}|\mathcal{F}_n]$ et d'autre part $\mathbb{E}[e^{M_{n+1}-C_{n+1}}|\mathcal{F}_n] = e^{M_n - C_n}$. Par conséquent, on a nécessairement $C_{n+1} - C_n = \ln(\mathbb{E}[e^{M_{n+1}-M_n}|\mathcal{F}_n])$ qui est bien défini p.s. puisque par la propriété de positivité de l'espérance conditionnelle, $\mathbb{E}[e^{M_{n+1}-M_n}|\mathcal{F}_n] > 0$. En outre, en appliquant l'inégalité de Jensen, il vient que $\mathbb{E}[e^{M_{n+1}-M_n}|\mathcal{F}_n] \geq e^{\mathbb{E}[M_{n+1}-M_n|\mathcal{F}_n]} = 1$ et par conséquent $C_{n+1} \geq C_n$. On en déduit le résultat suivant :

**Proposition 3.2.5.** *Soit* $(M_n, n \in \mathbb{N})$ *une martingale réelle sur* $(\Omega, \mathcal{F}, \mathbb{P})$ *muni de la filtration* $(\mathcal{F}_n)_{n\in\mathbb{N}}$ *telle que* $\forall n \in \mathbb{N}, \mathbb{E}[e^{M_n}] < \infty$*. Il existe une unique suite prévisible croissante* $(C_n)_{n\in\mathbb{N}}$ *de v.a.r. telle que*

$$e^{M_n - C_n} \text{ est une } (\mathcal{F}_n)\text{-martingale}$$

*et* $C_0 = 0$*, c'est :* $C_n = \sum_{k=0}^{n-1} \ln(\mathbb{E}[e^{M_{k+1}-M_k}|\mathcal{F}_k])$*.*

La construction du crochet est “arithmétique” alors que la construction de la martingale exponentielle $e^{M_n - C_n}$ est “géométrique”. L'exercice suivant généralise cette construction.

**Exercice 3.2.6.** *Soit* $(X_n, n \in \mathbb{N})$ *une sous-martingale réelle t.q.* $\forall n, X_n > 0$*.*

1. *Montrer que* $C_n = \prod_{k=0}^{n-1} \frac{\mathbb{E}[X_{k+1}|\mathcal{F}_k]}{X_k}$ *définit l'unique suite croissante prévisible telle que* $(X_n/C_n, n \in \mathbb{N})$ *est une* $(\mathcal{F}_n)$*-martingale et* $C_0 = 1$*.*
2. *En utilisant l'exercice 3.1.5, retrouver le résultat de la Proposition 3.2.5.*

### 3.2.2 Intégrale stochastique discrète

On se place comme précédemment dans un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$ muni d'une filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$. On se donne un processus réel $(\mathcal{F}_n)$-adapté $(H_n, n \in \mathbb{N})$ et une $(\mathcal{F}_n)$-martingale $(M_n, n \in \mathbb{N})$.

On définit alors le processus $((H \bullet M)_n, n \in \mathbb{N})$ par :

$$(H \bullet M)_n = \sum_{k=0}^{n-1} H_k (M_{k+1} - M_k). \tag{3.2}$$

On l'appelle *intégrale stochastique discrète du processus* $(H_n, n \in \mathbb{N})$ *par rapport à la martingale* $(M_n, n \in \mathbb{N})$. Il s'agit clairement d'un processus $(\mathcal{F}_n)$-adapté.

**Proposition 3.2.7.** *Si* $(H_n, n \in \mathbb{N})$ *est un processus réel à valeurs bornées (i.e.* $\forall n, H_n \in L^\infty(\Omega, \mathcal{F}, \mathbb{P})$*), le processus* $((H \bullet M)_n, n \in \mathbb{N})$ *est une* $(\mathcal{F}_n)$*-martingale.*

*Preuve.* De l'inégalité $|(H \bullet M)_n| \leq \sum_{k=0}^{n-1} |H_k||M_{k+1} - M_k|$, il vient que $(H \bullet M)_n$ est intégrable puisque $|H_k|$ est bornée p.s. et $|M_{k+1} - M_k|$ est intégrable. Ensuite pour $n \geq 1$, $\mathbb{E}[(H \bullet M)_n | \mathcal{F}_{n-1}] = (H \bullet M)_{n-1} + H_{n-1}\mathbb{E}[M_n - M_{n-1}|\mathcal{F}_{n-1}] = (H \bullet M)_{n-1}$ et $((H \bullet M)_n, n \in \mathbb{N})$ est bien une $(\mathcal{F}_n)$-martingale. □

La condition imposée sur le processus intégré $(H_n, n \in \mathbb{N})$ est assez restrictive. On peut définir l'intégrale stochastique pour une famille plus large de processus, à condition de renforcer l'hypothèse sur la martingale $(M_n, n \in \mathbb{N})$. Nous allons désormais supposer qu'elle est de carré intégrable, et que le processus $(H_n, n \in \mathbb{N})$ satisfait

$$\forall n \in \mathbb{N},\ \mathbb{E}\left[\sum_{k=0}^{n-1} H_k^2([M]_{k+1} - [M]_k)\right] < \infty. \tag{3.3}$$

**Exercice 3.2.8.** *Vérifier que si* $(H_n, n \in \mathbb{N})$ *est à valeurs bornées, il satisfait cette condition.*

Sous ces hypothèses, $(H \bullet M)_n$ est de carré intégrable. En effet, par convexité on a $(H \bullet M)_n^2 \leq n \sum_{k=0}^{n-1} H_k^2 (M_{k+1} - M_k)^2$, et par conséquent

$$\begin{aligned}
\mathbb{E}[(H \bullet M)_n^2] &\leq n \sum_{k=0}^{n-1} \mathbb{E}[H_k^2 (M_{k+1} - M_k)^2] \\
&= n \sum_{k=0}^{n-1} \mathbb{E}[H_k^2 \mathbb{E}[(M_{k+1} - M_k)^2 | \mathcal{F}_k]] \text{ par double conditionnement} \\
&= n \sum_{k=0}^{n-1} \mathbb{E}[H_k^2([M]_{k+1} - [M]_k)] < \infty.
\end{aligned}$$

Le processus $((H \bullet M)_n, n \in \mathbb{N})$ est en particulier intégrable et c'est une martingale puisque $\mathbb{E}[(H \bullet M)_n | \mathcal{F}_{n-1}] = (H \bullet M)_{n-1} + H_{n-1}\mathbb{E}[M_n - M_{n-1}|\mathcal{F}_{n-1}] = (H \bullet M)_{n-1}$. Cette martingale étant de carré intégrable, on peut également calculer son crochet. On a $\mathbb{E}[(H \bullet M)_n^2 | \mathcal{F}_{n-1}] = \mathbb{E}[(H \bullet M)_{n-1}^2 + 2(H \bullet M)_{n-1} H_{n-1}(M_n - M_{n-1}) + H_{n-1}^2 (M_n - M_{n-1})^2 | \mathcal{F}_{n-1}] = (H \bullet M)_{n-1}^2 + H_{n-1}^2([M]_n - [M]_{n-1})$. Par conséquent, on obtient

$$\begin{aligned}
&[H \bullet M]_n - [H \bullet M]_{n-1} = H_{n-1}^2([M]_n - [M]_{n-1}) \text{ et} \qquad (3.4)\\
&\mathbb{E}[(H \bullet M)_n^2] = \mathbb{E}[[H \bullet M]_n] = \mathbb{E}\left[\sum_{k=0}^{n-1} H_k^2([M]_{k+1} - [M]_k)\right].
\end{aligned}$$

La proprosition suivante résume ces résultats.

**Proposition 3.2.9.** *Si* $(M_n, n \in \mathbb{N})$ *est une martingale de carré intégrable et si* $(H_n, n \in \mathbb{N})$ *est un processus réel satisfaisant la condition* (3.3) *le processus* $((H \bullet M)_n, n \in \mathbb{N})$ *est une* $(\mathcal{F}_n)$*-martingale de carré intégrable. Son crochet satisfait alors* (3.4)*.*

### 3.2.3 Un exemple important : la marche aléatoire sur $\mathbb{Z}$

On considère une suite i.i.d de variables aléatoires $(X_n, n \geq 1)$ de loi $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$ et on note $\mathcal{F}_0$ la tribu triviale et pour $n \geq 1$, $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. On pose $\Sigma_0 = 0$ et $\Sigma_n = \sum_{k=1}^n X_k$. Le processus $(\Sigma_n, n \in \mathbb{N})$ est appelé *marche aléatoire (standard)* sur $\mathbb{Z}$. Il est à valeurs bornées puisque $|\Sigma_n| \leq n$, et c'est une $(\mathcal{F}_n)$-martingale.

En particulier c'est une martingale de carré intégrable et $[\Sigma]_n - [\Sigma]_{n-1} = \mathbb{E}[X_n^2 | \mathcal{F}_{n-1}] = 1$ puisque $X_n^2 = 1$, et donc $[\Sigma]_n = n$. Ainsi,

$$(\Sigma_n^2 - n, n \in \mathbb{N}) \text{ est une } (\mathcal{F}_n)\text{-martingale}$$

qui est également à valeurs bornées. Calculons également pour $\lambda \in \mathbb{R}$, la martingale a exponentielle associée à $\lambda S_n$. Comme $X_{n+1}$ est indépendante de $\mathcal{F}_n$, $\mathbb{E}[e^{\lambda(\Sigma_{n+1} - \Sigma_n)} | \mathcal{F}_n] = \frac{e^\lambda + e^{-\lambda}}{2} = \cosh(\lambda)$, et donc

$$(\exp\left[\lambda \Sigma_n - n \ln(\cosh(\lambda))\right], n \in \mathbb{N}) \text{ est une } (\mathcal{F}_n)\text{-martingale}$$

à valeurs bornées.

Enfin, dans le cas de la marche aléatoire, si un processus $(H_n, n \in \mathbb{N})$ est tel que $\mathbb{E}[H_n^2] < +\infty$ pour tout $n$, l'intégrale stochastique discrète $(H \bullet \Sigma)_n = \sum_{i=1}^n H_{i-1} X_i$ est une martingale de carré intégrable, et son crochet vaut :

$$[H \bullet \Sigma]_n = \sum_{i=1}^n H_{i-1}^2.$$

## 3.3 Théorème d'arrêt

Dans cette section, on se place toujours sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ muni d'une filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$.

**Définition 3.3.1.** *Un temps aléatoire est une v.a.* $T : \Omega \to \bar{\mathbb{N}}$ *à valeurs dans* $\bar{\mathbb{N}} = \mathbb{N} \cup \{+\infty\}$*. Un temps aléatoire* $\tau$ *est un* $(\mathcal{F}_n)$*-temps d'arrêt si*

$$\forall n \in \mathbb{N}, \{\tau \leq n\} \in \mathcal{F}_n,$$

*ou de manière équivalente si* $\forall n \in \mathbb{N}, \{\tau > n\} \in \mathcal{F}_n$*.*

Pour un temps d'arrêt $\tau$, on a, pour $k \leq n$, $\{\tau \leq k\} \in \mathcal{F}_k \subset \mathcal{F}_n$. Par conséquent, les événements $\{\tau = k\}$ pour $k \leq n$ et $\{\tau > n\}$ sont $\mathcal{F}_n$-mesurables. Un temps d'arrêt est donc un temps aléatoire dont on connaît, à l'instant $n$, s'il est passé et, s'il est passé, l'instant auquel il s'est produit.

Donnons désormais quelques exemples et contre-exemples de temps d'arrêts. On se donne une suite i.i.d. $(X_n, n \in \mathbb{N})$ de variable aléatoire de Bernoulli de paramètre $1/2$ et on note $(\mathcal{F}_n, n \in \mathbb{N})$ la filtration engendrée par cette suite. On considère également $T$ une variable aléatoire indépendante de loi de Poisson de paramètre 1. On pose $S_n = \sum_{k=0}^{n} X_k$. Il s'agit d'un processus croissant $(\mathcal{F}_n)$-adapté qui croît au plus d'une unité : $0 \leq S_n - S_{n-1} \leq 1$. On pose pour $l \in \mathbb{N}$, $\tau_l = \inf\{k \geq 0, S_k \geq l\} = \inf\{k \geq 0, S_k = l\}$ et $\sigma_l = \sup\{k \geq 0, S_k \leq l\} = \sup\{k \geq 0, S_k = l\}$ qui sont respectivement le premier temps d'atteinte et le dernier temps d'atteinte du niveau $l$. Le temps aléatoire $\tau_l$ est un temps d'arrêt :

$$\{\tau_l > n\} = \{S_n < l\} \in \mathcal{F}_n.$$

En revanche $\sigma_l$ n'est pas un temps d'arrêt. En effet, $\{\sigma_l \leq n\} = \{S_{n+1} \geq l + 1\}$ n'est pas $\mathcal{F}_n$-mesurable puisqu'il fait intervenir la v.a. $X_{n+1}$. Prouvons cela rigoureusement. On a $\{S_{n+1} \geq l + 1\} = \{S_n \geq l + 1\} \sqcup \{S_n = l, X_{n+1} = 1\}$. Si cet ensemble est $(\mathcal{F}_n)$-mesurable, $\{S_n = l, X_{n+1} = 1\} = \{S_{n+1} \geq l + 1\} \cap \{S_n < l + 1\}$ l'est également. Ainsi, on obtient la contradiction suivante $\mathbf{1}_{\{S_n=l,X_{n+1}=1\}} = \mathbb{E}(\mathbf{1}_{\{S_n=l,X_{n+1}=1\}}|\mathcal{F}_n) = \frac{1}{2}\mathbf{1}_{\{S_n=l\}}$ en utilisant l'indépendance entre $X_{n+1}$ et $\mathcal{F}_n$.

Il est facile de voir qu'un temps aléatoire déterministe (i.e. un entier positif) est un temps d'arrêt. Enfin, le temps aléatoire $T$ n'est pas un temps d'arrêt puisque l'événement $\{T \leq n\}$ est indépendant de $\mathcal{F}_n$.

La proposition suivante est une conséquence immédiate de la définition d'un temps d'arrêt, et sa preuve et laissée en exercice.

**Proposition 3.3.2.** *Soit $\tau^1, \tau^2$ (resp. $\tau^i, i \in \mathbb{N}$) deux $(\mathcal{F}_n)$-temps d'arrêt (resp. une famille dénombrable de temps d'arrêt). Alors $\tau^1 \wedge \tau^2$ et $\tau^1 \vee \tau^2$ (resp. $\inf\limits_{n\in\mathbb{N}} \tau^i$ et $\sup\limits_{n\in\mathbb{N}} \tau^i$) sont également deux $(\mathcal{F}_n)$-temps d'arrêt.*

On considère $(M_n, n \in \mathbb{N})$ une $(\mathcal{F}_n)$-martingale. On sait que pour tout $n \in \mathbb{N}$, $\mathbb{E}[M_n] = \mathbb{E}[M_0]$, et une question naturelle est de savoir si l'on peut étendre cette propriété pour un temps aléatoire. Par exemple, pour un temps aléatoire $T$ indépendant de la martingale $(M_n, n \in \mathbb{N})$ et fini presque sûrement (i.e. $\mathbb{P}(T < +\infty) = 1$, on a par double conditionnement $\mathbb{E}[M_T] = \mathbb{E}[\mathbb{E}[M_T|T]] = \mathbb{E}[M_0]$ puisque $\mathbb{E}[M_T|T] = \sum_{n\in\mathbb{N},\mathbb{P}(T=n)>0} \frac{\mathbb{E}[M_n \mathbf{1}_{\{T=n\}}]}{\mathbb{P}(T=n)} \mathbf{1}_{\{T=n\}} = \mathbb{E}[M_0]$ p.s. Ce résultat est néanmoins d'un intérêt limité puisque l'indépendance qu'il requiert est une hypothèse forte. Nous allons montrer que l'on a un résultat analogue pour un temps d'arrêt borné. On commence par introduire la notion de martingale arrêtée.

**Proposition 3.3.3.** *Soit $(M_n, n \in \mathbb{N})$ une $(\mathcal{F}_n)$-martingale et $\tau$ un $(\mathcal{F}_n)$-temps d'arrêt. Alors le processus $(M_{\tau\wedge n}, n \in \mathbb{N})$ est une $(\mathcal{F}_n)$-martingale, appelée souvent martingale $M_n$ arrêtée en $\tau$.*

*Preuve.* On a $M_{\tau\wedge n} - M_{\tau\wedge(n-1)} = \mathbf{1}_{\{\tau>n-1\}}(M_n - M_{n-1})$. On reconnait une intégrale stochastique discrète : en posant $H_n = \mathbf{1}_{\{\tau>n\}}$, $M_{\tau\wedge n} = (H \bullet M)_n + M_0$ ce qui permet de conclure en utilisant la proposition 3.2.7. □

**Théorème 3.3.4** (Théorème d'arrêt)**.** *Soit $(M_n, n \in \mathbb{N})$ une $(\mathcal{F}_n)$-martingale et $\tau$ un $(\mathcal{F}_n)$-temps d'arrêt borné (i.e $\exists \nu \in \mathbb{N}, \mathbb{P}(\tau \leq \nu) = 1$). Alors on a :*

$$\mathbb{E}[M_\tau] = \mathbb{E}[M_0].$$

*Preuve.* D'après la proposition précédente, $(M_{\tau\wedge n}, n \in \mathbb{N})$ est une $(\mathcal{F}_n)$-martingale, et par conséquent $\mathbb{E}[M_{\tau\wedge\nu}] = \mathbb{E}[M_{\tau\wedge 0}] = \mathbb{E}[M_0]$. Mais par ailleurs, puisque $\tau$ est borné par $\nu$, $M_{\tau\wedge\nu} = M_\tau$ ce qui donne le résultat. □

**Exercice 3.3.5.** *On considère la marche aléatoire sur $\mathbb{Z}$ introduite à la section 3.2.3. Pour $a \in \mathbb{Z}$, on note*

$$\tau_a = \inf\{n \in \mathbb{N}^*, \Sigma_n = a\}.$$

1. *Montrer que $\tau_a$ est un temps d'arrêt pour la filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$.*
2. *Soit $n \in \mathbb{N}$. Montrer que $\mathbb{E}[\exp(\lambda\Sigma_{\tau_a\wedge n} - (\tau_a \wedge n)\ln(\cosh(\lambda)))] = 1$ pour $\lambda \in \mathbb{R}$.*
3. *On suppose $a \in \mathbb{N}^*$ et $\lambda > 0$. Montrer que $\mathbb{E}[\exp(\lambda a - \tau_a \ln(\cosh(\lambda)))\mathbf{1}_{\{\tau_a<+\infty\}}] = 1$.*
4. *En déduire que $\mathbb{P}(\tau_a < +\infty) = 1$, puis que $\sup_{n\in\mathbb{N}} \Sigma_n = +\infty$ p.s. Le temps d'arrêt $\tau_a$ peut-il être borné ?*
5. *Montrer que pour $y > 0$,*

   $$\mathbb{E}[e^{-y\tau_a}] = \exp(-a\operatorname{acosh}(e^y)).$$

   *On a ainsi déterminé la loi de $\tau_a$ car la transformée de Laplace, comme la fonction caractéristique, caractérise la loi. Montrer alors que $\mathbb{E}[\tau_a] = +\infty$.*
6. *Par contraposée du Théorème de convergence dominée, montrer que $\mathbb{E}[\sup_{0\leq n\leq\tau_a} |\Sigma_n|] = +\infty$ puis que $\mathbb{E}[\inf_{0\leq n\leq\tau_a} \Sigma_n] = -\infty$.*
7. *On considère le jeu de pile ou face suivant : on mise la somme $1$, et on gagne $1$ si l'on obtient pile et on perd sa mise si l'on voit la face de la pièce. On joue successivement à ce jeu. Montrer que le gain de ce jeu est une marche aléatoire sur $\mathbb{Z}$. Donner une stratégie où on est sûr de gagner la somme $a \in \mathbb{N}^*$, mais où on doit attendre en moyenne un temps infini pour gagner, et où on doit avoir une trésorerie illimitée.*
8. *On cherche désormais à connaître la loi de $\tau_a$ pour $a \leq 0$. Montrer que $\tau_a$ et $\tau_{-a}$ ont même loi. Pour $a = 0$, en déduire que $\mathbb{E}[e^{-y\tau_0}] = e^{-y}\exp(-\operatorname{acosh}(e^y))$. En déduire que $\mathbb{P}(\tau_0 < +\infty) = 1$ et $\mathbb{E}[\tau_0] = +\infty$.*

## 3.4 Inégalités maximales

Comme précédement, on se place sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ muni d'une filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$.

**Lemme 3.4.1.** *Si $(X_n, n \in \mathbb{N})$ une sous-martingale positive,*

$$\forall x \in \mathbb{R}_+^*, \mathbb{P}(\max_{0\leq k\leq n} X_k > x) \leq \frac{1}{x}\mathbb{E}[\mathbf{1}_{\{\max\limits_{0\leq k\leq n} X_k > x\}} X_n] \leq \frac{1}{x}\mathbb{E}[X_n].$$

*Preuve.* Pour $n = 0$, il s'agit de l'inégalité de Markov ($\mathbf{1}_{\{X_0 > x\}} \leq \frac{X_0}{x}$). Nous allons montrer ce résultat par récurrence et supposer le résultat vrai jusqu'au rang $n$. On a

$$\{\max_{0\leq k\leq n+1} X_k > x\} = \{\max_{0\leq k\leq n} X_k > x\} \sqcup \{\max_{0\leq k\leq n} X_k \leq x, X_{n+1} > x\}$$

et par conséquent, en utilisant l'hypothèse de récurrence puis $X_n \leq \mathbb{E}[X_{n+1}|\mathcal{F}_n]$, il vient

$$\begin{aligned}
\mathbb{P}(\max_{0\leq k\leq n+1} X_k > x) &= \mathbb{P}(\max_{0\leq k\leq n} X_k > x) + \mathbb{P}(\max_{0\leq k\leq n} X_k \leq x, X_{n+1} > x) \\
&\leq \frac{1}{x}\mathbb{E}[\mathbf{1}_{\{\max\limits_{0\leq k\leq n} X_k > x\}} X_n] + \mathbb{E}[\mathbf{1}_{\{\max\limits_{0\leq k\leq n} X_k \leq x, X_{n+1} > x\}} \frac{X_{n+1}}{x}] \\
&\leq \frac{1}{x}\left(\mathbb{E}[\mathbf{1}_{\{\max\limits_{0\leq k\leq n} X_k > x\}} X_{n+1}] + \mathbb{E}[\mathbf{1}_{\{\max\limits_{0\leq k\leq n} X_k \leq x, X_{n+1} > x\}} X_{n+1}]\right) \\
&= \frac{1}{x}\mathbb{E}[\mathbf{1}_{\{\max\limits_{0\leq k\leq n+1} X_k > x\}} X_{n+1}]. \quad \square
\end{aligned}$$

**Théorème 3.4.2** (Inégalité de Doob)**.** *Soit $(M_n, n \in \mathbb{N})$ une martingale réelle de carré intégrable. On a :*

$$\mathbb{E}[\max_{0\leq k\leq n} M_k^2] \leq 4\mathbb{E}[M_n^2].$$

*En particulier, $\mathbb{E}[\sup\limits_{n\in\mathbb{N}} M_n^2] \leq 4\sup\limits_{n\in\mathbb{N}} \mathbb{E}[M_n^2]$, ces termes pouvant être éventuellement infinis.*

*Preuve.* D'après l'exercice 3.1.5, $(|M_n|, n \in \mathbb{N})$ est une sous-martingale positive et on peut donc utiliser le lemme précédent. Puisque $\max\limits_{0\leq k\leq n} M_k^2 = (\max\limits_{0\leq k\leq n} |M_k|)^2 = 2\int_0^{+\infty} x\mathbf{1}_{\{x < \max\limits_{0\leq k\leq n} |M_k|\}} dx$, on obtient :

$$\begin{aligned}
\mathbb{E}[\max_{0\leq k\leq n} M_k^2] &= 2\int_0^{+\infty} x\mathbb{P}(\max_{0\leq k\leq n} |M_k| > x) dx \text{ (Théorème de Fubini)} \\
&\leq 2\int_0^{+\infty} \mathbb{E}[\mathbf{1}_{\{\max\limits_{0\leq k\leq n} |M_k| > x\}} |M_n|] dx \text{ (d'après le lemme précédent)} \\
&= 2\mathbb{E}\left[\int_0^{+\infty} \mathbf{1}_{\{\max\limits_{0\leq k\leq n} |M_k| > x\}} |M_n| dx\right] \text{ (Théorème de Fubini)} \\
&= 2\mathbb{E}[\max_{0\leq k\leq n} |M_k||M_n|] \leq 2\sqrt{\mathbb{E}[\max_{0\leq k\leq n} M_k^2]}\sqrt{\mathbb{E}[M_n^2]}. \text{ (par Cauchy-Schwarz)}
\end{aligned}$$

Il vient alors $\mathbb{E}[\max_{0\leq k\leq n} M_k^2] \leq 4\mathbb{E}[M_n^2]$. De cette inégalité, on déduit $\mathbb{E}[\max_{0\leq k\leq n} M_k^2] \leq 4\sup_{n\in\mathbb{N}}\mathbb{E}[M_n^2]$, puis par convergence monotone $\mathbb{E}[\sup_{n\in\mathbb{N}} M_n^2] \leq 4\sup_{n\in\mathbb{N}}\mathbb{E}[M_n^2]$. □

**Remarque 3.4.3.** *En examinant la preuve, on voit que ce résultat est valable également si $M_n$ est une sous martingale positive de carré intégrable. D'autre part, on peut montrer de façon similaire que pour $p \geq 1$,*

$$\mathbb{E}[\max_{0\leq k\leq n} M_k^p] \leq \left(\frac{p}{1-p}\right)^p \mathbb{E}[M_n^p]$$

*dès lors que $M_n$ est une martingale (ou sous-martingale) positive telle que $\forall n, \mathbb{E}[|M_n|^p] < \infty$.*

## 3.5 Théorèmes de convergence des Martingales

Cette section est dévouée à des résultats importants de convergence des martingales. Comme précédemment, on se place sur un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$ muni d'une filtration $(\mathcal{F}_n, n \in \mathbb{N})$.

**Théorème 3.5.1.** *Soit $(M_n, n \in \mathbb{N})$ une $(\mathcal{F}_n)$-martingale réelle de carré intégrable. Si on a $\sup_{n\in\mathbb{N}}\mathbb{E}[M_n^2] < +\infty$, alors $M_n$ converge p.s. et dans $L^2$ vers une variable notée $M_\infty$, i.e.*

$$M_n \underset{n\to+\infty}{\to} M_\infty \text{ p.s. et } \mathbb{E}[(M_n - M_\infty)^2] \underset{n\to+\infty}{\to} 0.$$

*Enfin, on a $\forall n \in \mathbb{N}, \mathbb{E}[M_\infty|\mathcal{F}_n] = M_n$.*

*Preuve.* Nous commençons par prouver la convergence p.s. Rappelons tout d'abord qu'une suite $(x_n)_{n\in\mathbb{N}}$ de réels converge si et seulement si c'est une suite de Cauchy, i.e. $\sup_{p,q\in\mathbb{N}} |x_{n+p} - x_{n+q}| \underset{n\to+\infty}{\to} 0$. Nous allons montrer que, presque sûrement $X_n = \sup_{p,q\in\mathbb{N}} |M_{n+p} - M_{n+q}| \underset{n\to+\infty}{\to} 0$ ce qui assurera que p.s., $(M_n)_{n\in\mathbb{N}}$ est une suite de Cauchy et donc que $M_n$ converge p.s. La suite $(X_n)$ est par construction positive et décroissante, elle converge donc p.s. vers une v.a que l'on note $X_\infty$. Par ailleurs, on a grâce à l'inégalité triangulaire

$$X_n \leq 2\sup_{p\in\mathbb{N}} |M_{n+p} - M_n|.$$

Il est aisé de voir que $(M_{n+p} - M_n)_{p\in\mathbb{N}}$ est une martingale par rapport à la filtration $(\mathcal{F}_{n+p})_{p\in\mathbb{N}}$. Grâce à l'inégalité de Doob, il vient

$$\mathbb{E}[X_n^2] \leq 4\mathbb{E}[\sup_{p\in\mathbb{N}}(M_{n+p} - M_n)^2] \leq 16\sup_{p\in\mathbb{N}}\mathbb{E}[(M_{n+p} - M_n)^2].$$

Par ailleurs, $\mathbb{E}[(M_{n+p}-M_n)^2] = E[M_{n+p}^2]-2\mathbb{E}[\mathbb{E}[M_nM_{n+p}|\mathcal{F}_n]]+E[M_n^2] = E[M_{n+p}^2]-E[M_n^2]$. Par conséquent $(E[M_n^2])_{n\in\mathbb{N}}$ est une suite croissante, et $\sup_{k\in\mathbb{N}}\mathbb{E}[M_k^2] = \lim_{k\to+\infty}\mathbb{E}[M_k^2]$. On en déduit :

$$\mathbb{E}[X_n^2] \leq 16(\lim_{k\to+\infty}\mathbb{E}[M_k^2] - \mathbb{E}[M_n^2])$$

Mais par convergence dominée ($0 \leq X_n \leq X_0$), $\mathbb{E}[X_n^2] \underset{n\to+\infty}{\to} \mathbb{E}[X_\infty^2]$ alors que le terme de droite tend clairement vers 0. On en déduit $\mathbb{E}[X_\infty^2] = 0$, puis $X_\infty = 0$ p.s. ce qui donne la convergence p.s.

Montrons maintenant la convergence $L^2$. On a clairement $|M_\infty| \leq \sup_{n\in\mathbb{N}}|M_n|$ et donc $M_\infty$ est de carré intégrable puisque l'inégalité de Doob donne

$$\mathbb{E}[M_\infty^2] \leq \mathbb{E}[\sup_{n\in\mathbb{N}}M_n^2] \leq 4\sup_{n\in\mathbb{N}}\mathbb{E}[M_n^2]$$

La suite $(M_\infty - M_n)^2$ converge p.s. vers 0 et est dominée par $4\sup_{n\in\mathbb{N}}M_n^2$ et le théorème de Lebesgue assure $\mathbb{E}[(M_\infty - M_n)^2] \underset{n\to+\infty}{\to} 0$.

Il reste a prouver que $\mathbb{E}[M_\infty|\mathcal{F}_n] = M_n$. Soit $A \in \mathcal{F}_n$. On a pour $p \geq 0$, $\mathbb{E}[M_{n+p}\mathbf{1}_A] = \mathbb{E}[M_n\mathbf{1}_A]$, puis en faisant tendre $p \to +\infty$, on obtient par le théorème de convergence dominée $\mathbb{E}[M_\infty\mathbf{1}_A] = \mathbb{E}[M_n\mathbf{1}_A]$. □

Voici deux exercices qui mettent directement en application ce résultat.

**Exercice 3.5.2.** *On considère un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ muni d'une filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ et $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. On pose $M_n = \mathbb{E}[X|\mathcal{F}_n]$. Montrer que $M_n$ converge p.s. et dans $L^2$.*

**Exercice 3.5.3.** *On considère une suite de v.a.r. $(X_n, n \in \mathbb{N}^*)$ i.i.d. et on note pour $\alpha > 0$ et $n \in \mathbb{N}^*$, $S_n^\alpha = \sum_{k=1}^n \frac{1}{k^\alpha}X_k$ et $S_0^\alpha = 0$.*

1. *On suppose de $X_1$ est bornée et $\alpha > 1$. Montrer que $S_n^\alpha$ converge presque sûrement vers une v.a. $S_\infty^\alpha$ qui est finie, p.s.*
2. *On suppose que $X_1$ est de carré intégrable t.q. $\mathbb{E}[X_1] = 0$ et $\alpha > 1/2$. Montrer que $S_n^\alpha$ converge presque sûrement vers une v.a. $S_\infty^\alpha$ qui est de carré intégrable. Calculer $\mathbb{E}[(S_\infty^\alpha)^2]$.*

Le but de l'exercice suivant est de donner un résultat de convergence pour les martingales (et surmartingales) positives.

**Exercice 3.5.4.**
1. *Soit $(Y_n, n \in \mathbb{N})$ une surmartingale positive par rapport à une filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$. Montrer que $X_n = e^{-Y_n}$ est une sous-martingale telle que $\forall n, 0 \leq X_n \leq 1$. (on pourra utiliser l'exercice 3.1.5).*
2. *D'après l'exercice 3.2.4, nous savons qu'il existe un unique processus $(C_n, n \in \mathbb{N})$ prévisible, intégrable et croissant tel que :*

– $C_0 = 0$,
– $M_n = X_n - C_n$ *est une* $(\mathcal{F}_n)$*-martingale.*

*On pose pour* $p \in \mathbb{N}^*$, $\tau_p = \inf\{n \in \mathbb{N}^*, C_{n+1} \geq p\}$ $(\inf \emptyset = +\infty)$*. Montrer* $\tau_p$ *est un temps d'arrêt par rapport à la filtration* $(\mathcal{F}_n)_{n\in\mathbb{N}}$.

3. *Montrer que* $C_n$ *converge p.s. vers une v.a.* $C_\infty$ *lorsque* $n \to +\infty$*. Montrer que* $\mathbb{E}(C_n) \leq 2$*, puis que* $\mathbb{E}(C_\infty) \leq 2$*. En déduire que* $\mathbb{P}(C_\infty < +\infty) = 1$*, puis que* $\Omega = \cup_{p\in\mathbb{N}^*}\{\tau_p = +\infty\}$.
4. *On considère* $M_{\tau_p \wedge n}$ *la martingale* $M_n$ *arrêtée en* $\tau_p$*. Montrer que* $M_{\tau_p\wedge n}$ *converge p.s. vers une v.a.* $\xi^p$ *de carré intégrable telle que* $\mathbb{E}[\xi^p|\mathcal{F}_n] = M_{\tau_p\wedge n}$*. En déduire que* $\lim_n M_n$ *existe sur* $\{\tau_p = +\infty\}$*, puis sur* $\Omega$*. On note* $M_\infty$ *cette limite.*
5. *Montrer enfin que la sous-martingale* $(X_n, n \in \mathbb{N})$ *converge p.s vers* $X_\infty = M_\infty - C_\infty$*. En déduire que la surmartingale que* $Y_n$ *converge p.s. vers* $Y_\infty = -\ln(X_\infty)$*, et montrer que l'on a :*
$$\forall n \in \mathbb{N}, \mathbb{E}[Y_\infty|\mathcal{F}_n] \leq Y_n.$$
6. *Application : on se donne une suite de v.a.* $(Z_n)_{n\geq 1}$ *i.i.d. de loi* $\mathbb{P}(Z_1 = 3/2) = \mathbb{P}(Z_1 = 1/2) = 1/2$*. On définit* $Y_0 = 1$ *et* $Y_n = \prod_{i=1}^n Z_i$ *pour* $n \geq 1$*. Vérifier que* $Y_n$ *est une martingale par rapport à la filtration* $\mathcal{F}_n = \sigma(Z_1, \ldots, Z_n)$*. Montrer que* $Y_n$ *converge p.s. vers* $Y_\infty$*. En étudiant* $\log(Y_n)$*, retrouver ce résultat et montrer que* $Y_\infty = 0$*. La martingale* $(Y_n, n \in \mathbb{N})$ *vérifie-t-elle* $\sup_{n\in\mathbb{N}} \mathbb{E}[Y_n^2] < +\infty$ *?*

# Chapitre 4

# Chaînes de Markov à temps discret

## 4.1 Premières définitions

Nous nous plaçons comme précédemment sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

**Définition 4.1.1.** *Soit $E$ un espace d'état au plus dénombrable. On dit que $(P(x,y), x \in E, y \in E)$ est une matrice de transition d'une chaîne de Markov si pour tout $x \in E$, $(P(x,y), y \in E)$ est une loi de probabilité sur $E$ (i.e. $\forall y, P(x,y) \geq 0$ et $\sum_{y \in E} P(x,y) = 1$.)*

Donnons trois exemples de matrices de transition :

- $E = \{0,1\}$, on se donne $p, q \in [0,1]$ et on définit $P(0,1) = p = 1 - P(0,0)$ et $P(1,0) = q = 1 - P(1,1)$.
- $E = \mathbb{Z}$, on pose pour $x \in \mathbb{Z}$, $P(x, x+1) = P(x, x-1) = 1/2$ et $P(x,y) = 0$ si $y \notin \{x-1, x+1\}$. C'est la matrice de transition de la marche aléatoire standard sur $\mathbb{Z}$.
- Pour un ensemble $E$ quelconque, $I(x,y) = \mathbf{1}_{\{x=y\}}$ est une matrice de transition appelée identité.

Si $P$ et $Q$ sont deux matrices de transition sur $E$, il est facile de voir que la matrice produit $PQ$ définie par

$$\forall x, z \in E, PQ(x,z) = \sum_{y \in E} P(x,y)Q(y,z)$$

est également une matrice de transition. Pour $l \in \mathbb{N}$, on note $P^l$ la matrice de transition correspondant au produit $\underbrace{P \ldots P}_{l \text{ fois}}$ avec la convention $P^0(x,y) = I(x,y)$. D'autre part, si $(\mu(x), x \in E)$ est une loi de probabilité sur $E$ (i.e. $\forall x, \mu(x) \geq 0$ et $\sum_{x \in E} \mu(x) = 1$) et si $P$ est une matrice de transition sur $E$, on définit $(\mu P(x), x \in E)$ par :

$$\forall x \in E, \mu P(x) = \sum_{z \in E} \mu(z) P(z,x).$$

Il s'agit d'une loi de probabilité sur $E$ puisque pour tout $x \in E$, $\mu P(x) \geq 0$ et

$$\sum_{x \in E} \mu P(x) = \sum_{z \in E} \mu(z) \left( \sum_{x \in E} P(z,x) \right) = \sum_{z \in \mathbb{E}} \mu(z) = 1.$$

Enfin, pour une fonction $f : E \to \mathbb{R}$ et une probabilité $(\mu(x), x \in E)$ sur $E$, on définit

$$\mu f = \sum_{x \in E} \mu(x) f(x) \in \mathbb{R}$$

lorsque cette somme a un sens, i.e. lorsque $f$ est positive ou $\sum_{x \in E} \mu(x)|f(x)| < \infty$. Pour une matrice de transition $P$ sur $E$, $(P(x,y), y \in E)$ est une loi de probabilité sur $E$ et on définit $Pf : E \to \mathbb{R}$ la fonction $Pf(x) = \sum_{y \in E} P(x,y)f(y)$ lorsque cette somme a un sens.

**Définition 4.1.2.** *Soit $E$ un espace d'état au plus dénombrable. On dit que le processus $(X_n, n \in \mathbb{N})$ est une chaîne de Markov sur $E$ si pour tout $x_0, \ldots, x_{n+1} \in E$,*

$$\mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n, \ldots, X_0 = x_0) = \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n). \tag{4.1}$$

*Cet égalité est appelée propriété de Markov.*

Autrement dit, pour une chaîne de Markov, la loi du prochain état visité ne dépend de sa trajectoire passée qu'à travers l'état dans lequel il se trouve.

**Remarque 4.1.3.** *Plus généralement, lorsque l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ est muni d'une filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$, on dit qu'un processus $(X_n, n \in \mathbb{N})$ à valeurs dans $E$ est une chaîne de Markov sur $E$ par rapport à la filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$ si c'est un processus $(\mathcal{F}_n)$-adapté tel que $\forall y \in E, \mathbb{P}(X_{n+1} = y | \mathcal{F}_n) = \mathbb{P}(X_{n+1} = y | X_n)$. Il est facile de voir qu'une $(\mathcal{F}_n)$-chaîne de Markov est une chaîne de Markov au sens donné ci-dessus. En revanche, si une chaîne de Markov est toujours une chaîne de Markov par rapport à la filtration engendrée par $(X_n, n \in \mathbb{N})$, elle ne l'est pas nécessairement pour une filtration plus grande.*

Lorsque $(X_n, n \in \mathbb{N})$ est une chaîne de Markov sur $E$, on note pour $x, y \in E$, $P_n(x,y) = \mathbb{P}(X_{n+1} = y | X_n = x)$. C'est une matrice de transition, appelée matrice de transition à l'instant $n$. Lorsque cette matrice de transition ne dépend pas de $n$, on dit que $(X_n, n \in \mathbb{N})$ est une chaîne de Markov *homogène en temps*. Dans ce cours on ne considèrera que des chaînes de Markov homogènes en temps.

**Définition 4.1.4.** *On dit qu'une chaîne de Markov $(X_n, n \in \mathbb{N})$ est homogène de matrice de transition $P$ si :*

$$\forall n \in \mathbb{N}, \forall x, y \in E,\ \mathbb{P}(X_{n+1} = y | X_n = x) = P(x,y).$$

Pour une chaîne de Markov de matrice de transition $P$, on dira qu'une probabilité (ou plus généralement une mesure) $\mu$ sur $E$ est *invariante* si $\mu P = \mu$, i.e.

$$\forall y \in E,\ \sum_{x \in E} \mu(x) P(x,y) = \mu(y).$$

Une telle probabilité existe toujours lorsque $E$ est fini (cf exercice 4.1.5). Si $X_n$ suit une loi de probabilité invariante $\mu$, alors $\mathbb{P}(X_{n+1} = y) = \sum_{x \in E} \mathbb{P}(X_{n+1} = y, X_n = x) = \sum_{x \in E} \mathbb{P}(X_{n+1} = y | X_n = x)\mathbb{P}(X_n = x) = \sum_{x \in E} \mu(x)P(x,y) = \mu(y)$ et $X_{n+1}$ suit également la loi $\mu$. Par conséquent, si $X_0$ suit une loi invariante $\mu$, pour tout $n \in \mathbb{N}$, $X_n$ est distribuée selon la même loi invariante.

**Exercice 4.1.5.** *On suppose que $E$ est un espace avec un nombre fini d'états et on considère $(X_n, n \in \mathbb{N})$ est une chaîne de Markov sur $E$ de matrice de transition $P$.*

1. *Vérifier que l'ensemble des probabilités sur $E$ forme un ensemble convexe compact.*
2. *Soit $\mu$ une probabilité sur $E$. On pose $\mu_n = \frac{1}{n}\sum_{k=0}^{n-1} \mu P^k$. Vérifier que $\mu_n$ est une probabilité sur $E$ et montrer que $\mu_n P - \mu_n \underset{n \to +\infty}{\longrightarrow} 0$ dans $\mathbb{R}^E$.*
3. *En déduire alors qu'il existe une probabilité invariante pour la matrice de transition $P$.*

La proposition suivante décrit la loi d'un processus de Markov.

**Proposition 4.1.6.** *Un processus $(X_n, n \in \mathbb{N})$ est une chaîne de Markov sur $E$ de matrice de transition $P$ et de loi initiale $\mu$ (i.e $\forall x \in E, \mu(x) = \mathbb{P}(X_0 = x)$) si, et seulement si, pour tout $x_0, \dots, x_n \in E$,*

$$\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = \mu(x_0)P(x_0, x_1) \dots P(x_{n-1}, x_n). \tag{4.2}$$

*Preuve.* Commençons par le sens direct et supposons que $(X_n, n \in \mathbb{N})$ est une chaîne de Markov sur $E$. Nous allons montrer (4.2) par récurrence sur $n$. Pour $n = 0$, il n'y a rien à montrer et nous supposons le résultat vrai au rang $n-1$. On a en utilisant la propriété de Markov

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

et on conclut en utilisant la propriété de récurrence.

Réciproquement, si $(X_n, n \in \mathbb{N})$ satisfait (4.2), on a pour $n \in \mathbb{N}$ et $x_0, \dots, x_{n+1} \in E$,

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

D'autre part on a,

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

Ainsi, $\mathbb{P}(X_{n+1}=x_{n+1}|X_n=x_n,\dots,X_0=x_0)=\mathbb{P}(X_{n+1}=x_{n+1}|X_n=x_n)=P(x_n,x_{n+1})$.
□

De cette proposition, nous déduisons trois corollaires intéressants.

**Corollaire 4.1.7.** *Si $P$ est une matrice de transition sur $E$ et $(X_n, n\in\mathbb{N})$ un processus tel que $\mathbb{P}(X_{n+1}=x_{n+1}|X_n=x_n,\dots,X_0=x_0)=P(x_n,x_{n+1})$ pour $x_0,\dots,x_{n+1}\in E$, alors $(X_n,n\in\mathbb{N})$ est une chaîne de Markov sur $E$ de matrice de transition $P$.*

*Preuve.* Il s'agit de montrer que $\mathbb{P}(X_{n+1}=x_{n+1}|X_n=x_n,\dots,X_0=x_0)=\mathbb{P}(X_{n+1}=x_{n+1}|X_n=x_n)$. On observe que dans le sens direct de la preuve précédente, on n'utilise que la propriété $\mathbb{P}(X_{n+1}=x_{n+1}|X_n=x_n,\dots,X_0=x_0)=P(x_n,x_{n+1})$. Par conséquent, on obtient ici de façon similaire que $\mathbb{P}(X_0=x_0,\dots,X_n=x_n)=\mathbb{P}(X_0=x_0)P(x_0,x_1)\dots P(x_{n-1},x_n)$. Cela prouve, en utilisant le sens réciproque de la proposition précédente, que $(X_n,n\in\mathbb{N})$ est une chaîne de Markov sur $E$ de matrice de transition $P$. □

**Corollaire 4.1.8.** *Soit $q\in\mathbb{N}$ et $(X_n,n\in\mathbb{N})$ une chaîne de Markov sur $E$ de matrice de transition $P$ et de loi initiale $\mu$. Alors, le processus $(X_{n+q},n\in\mathbb{N})$ est une chaîne de Markov sur $E$ de matrice de transition $P$ et de loi initiale $\mu P^q$. En outre, conditionnellement à $X_q$, $(X_{n+q},n\in\mathbb{N})$ et $(X_0,\dots,X_q)$ sont indépendants.*

On résume en général ce dernier résultat en disant que "le passé et le futur d'une chaîne de Markov sont indépendants conditionnellement au présent". On peut montrer facilement qu'un processus $(X_n,n\in\mathbb{N})$ à valeurs dans $E$ qui satisfait cette propriété vérifie la propriété de Markov de la définition 4.1.2. Ces deux propriétés sont en fait équivalentes, et on les appelle toutes deux "propriété de Markov". Nous avons prouvé ici le sens réciproque seulement dans le cas homogène en temps, mais la preuve s'étend sans problème au cas non-homogène.

*Preuve.* Puisque $\mathbb{P}(X_q=x_q)=\sum_{x_0,\dots x_{q-1}\in E}\mathbb{P}(X_0=x_0,\dots,X_q=x_q)$ on a d'après la proposition précédente,

$$\mathbb{P}(X_q=x_q)=\sum_{x_0,\dots,x_{q-1}\in E}\mu(x_0)P(x_0,x_1)\dots P(x_{q-1},x_q)=(\mu P^q)(x_q).$$

Donc $X_q$ est de loi $\mu P^q$. Par ailleurs, on a pour $x_q, \dots, x_{q+n} \in E$,
$\mathbb{P}(X_q = x_q, \dots, X_{q+n} = x_{q+n})$
$= \sum_{x_0,\dots,x_{q-1}\in E} \mu(x_0)P(x_0,x_1)\dots P(x_{q-1},x_q)P(x_q,x_{q+1})\dots P(x_{q+n-1},x_{q+n})$
$= (\mu P^q)(x_q)P(x_q,x_{q+1})\dots P(x_{q+n-1},x_{q+n})$ ce qui prouve la première partie du résultat.
Pour l'indépendance, on a pour $x_0, \dots, x_{q+n} \in E$,
$\mathbb{P}(X_0 = x_0, \dots X_q = x_q, \dots, X_{q+n} = x_{q+n} | X_q = x_q)$
$= \frac{1}{\mathbb{P}(X_q=x_q)}\mu(x_0)P(x_0,x_1)\dots P(x_{q-1},x_q)\frac{\mathbb{P}(X_q=x_q)}{\mathbb{P}(X_q=x_q)}P(x_q,x_{q+1})\dots P(x_{q+n-1},x_{q+n})$
$= \mathbb{P}(X_0 = x_0, \dots X_q = x_q | X_q = x_q)\mathbb{P}(X_q = x_q, \dots, X_{q+n} = x_{q+n} | X_q = x_q)$. □

**Corollaire 4.1.9.** *Soit $(X_n, n \in \mathbb{N})$ une chaîne de Markov sur $E$ de matrice de transition $P$. Soit $q \in \mathbb{N}$ et $A \in \sigma(X_0, \dots, X_q)$. On a*

$$\mathbb{P}(A, X_q = x_q, \dots, X_{n+q} = x_{n+q}) = \mathbb{P}(A, X_q = x_q)P(x_q,x_{q+1})\dots P(x_{q+n-1},x_{q+n}). \quad (4.3)$$

*Preuve.* C'est clair si $A = \{X_0 = x_0, \dots, X_{q-1} = x_{q-1}\}$ pour $x_0, \dots, x_{q-1} \in E$, puis par $\sigma$-additivité, on obtient ce résultat pour tout $A \in \sigma(X_0, \dots, X_q)$. □

**Définition 4.1.10.** *Soit $x \in E$. On note $\mathbb{P}_x$ la loi de probabilité sachant $\{X_0 = x\}$ ou, ce qui est équivalent, la loi de probabilité de la chaîne de Markov issue de $x$ (i.e. de loi initiale $\mu(x_0) = \mathbf{1}_{\{x_0=x\}}$). Sous $\mathbb{P}_x$, on a pour $x_0, \dots, x_n \in E$ :*

$$\mathbb{P}_x(X_0 = x_0, \dots, X_n = x_n) = \mathbf{1}_{\{x_0=x\}}P(x,x_1)\dots P(x_{n-1},x_n).$$

Nous allons désormais donner un moyen générique de construire des chaînes de Markov. On se donne un espace d'état $E$ au plus dénombrable, $F$ un ensemble quelconque et $\psi : E \times F \to E$ une application. On se donne également une suite de v.a. $(U_n, n \in \mathbb{N})$ i.i.d. à valeurs dans $F$ et $x_0 \in E$. On définit alors pour $n \in \mathbb{N}$,

$$\begin{cases} X_0 = x_0 \\ X_{n+1} = \psi(X_n, U_n) \end{cases}$$

C'est une chaîne de Markov de matrice de transition $P(x,y) = \mathbb{P}(\psi(x,U_0) = y)$. En effet,

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

puisque $U_n$ est indépendante de $X_0, \dots, X_n$.

Avant de conclure cette section, nous voulons donner un exemple de processus discret qui ne satisfait pas la propriété de Markov. Nous proposons ici une marche aléatoire sur $\mathbb{Z}$ qui tient compte du dernier mouvement effectué, c'est à dire qui a de l'inertie. On se donne

$\alpha \in (-1,1)\backslash\{0\}$ et une suite de v.a. $(B_n)_{n\in\mathbb{N}}$ i.i.d de loi de Bernoulli de paramètre $\frac{1+\alpha}{2}$. On pose $X_0 = 0$ et $X_1$ une v.a. indépendante telle que $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = -1) = 1/2$, et

$$\forall n \geq 1, \; X_{n+1} = X_n + (2B_n - 1)(X_n - X_{n-1}).$$

Il est facile de voir que $\forall n \in \mathbb{N}, |X_{n+1} - X_n| = 1$, et pour $x \in \mathbb{Z}$,

$$\mathbb{P}(X_{n+1} = x+1|X_n = x, X_{n-1} = x-1) = \frac{1+\alpha}{2}, \mathbb{P}(X_{n+1} = x-1|X_n = x, X_{n-1} = x-1) = \frac{1-\alpha}{2},$$

$$\mathbb{P}(X_{n+1} = x+1|X_n = x, X_{n-1} = x+1) = \frac{1-\alpha}{2}, \mathbb{P}(X_{n+1} = x-1|X_n = x, X_{n-1} = x+1) = \frac{1+\alpha}{2}.$$

S'il s'agissait d'une chaîne de Markov, on aurait $\mathbb{P}(X_{n+1} = x+1|X_n = x, X_{n-1} = x-1) = \mathbb{P}(X_{n+1} = x+1|X_n = x) = \mathbb{P}(X_{n+1} = x+1|X_n = x, X_{n-1} = x+1)$ ce qui est impossible car $\alpha \neq 0$.

**Exercice 4.1.11.** *Montrer que* $(X_{n+1} - X_n, n \in \mathbb{N})$ *est une chaîne de Markov sur* $\{-1, 1\}$ *dont on donnera la matrice de transition. Montrer que pour tout* $n$*,* $\mathbb{P}(X_{n+1} - X_n = 1) = \mathbb{P}(X_{n+1} - X_n = -1) = 1/2$.

**Exercice 4.1.12.** *Montrer que si* $\alpha = 0$*,* $(X_n, n \in \mathbb{N})$ *est une chaîne de Markov qui est la marche aléatoire standard sur* $\mathbb{Z}$ *issue de* $0$.

## 4.2 Modéliser avec une chaîne de Markov

### 4.2.1 Risque de crédit, et notations

Nous commençons par donner ici une application en finance des chaînes de Markov. Régulièrement au cours de l'année, des agences de notations émettent des notes (ratings) évaluant la qualité de crédit d'entreprises, d'états ou de collectivités locales. Ces notes reflètent la capacité de chaque entité à rembourser sa dette et sont établies grâce à des analyses financières effectuées sur leur activité. Dans le monde, les principales agences de notations s'appellent Fitch, Moody's et Standard & Poors. Chacune de ces agences a son propre système de notation, même si leur principe de notation est similaire. Nous considèrerons dans ce paragraphe le système de notation simplifié suivant :

- La note $A$ est réservée aux entités en très bonne santé financière et présentant très peu de risque de faillite.
- La note $B$ est attribuée à des entreprises ou collectivités en bonne situation financière, mais qui pourraient être affectées par une évolution défavorable du marché.
- La note $C$ est donnée à des entreprises étant sujettes à des difficultés financières.
- La note $D$ signifie que l'entité n'est plus capable de rembourser sa dette.

Pour simplifier également, on suppose que les notes sont publiées chaque année. On note $X_n$ la note d'une entreprise au cours de la $n+1$-ème année. On modélise $(X_n, n \in \mathbb{N})$ comme une chaîne de Markov sur $\{A, B, C, D\}$, et on suppose que la matrice de transition vaut

$$P = \begin{bmatrix} 0.9 & 0.09 & 0.009 & 0.001 \\ 0.14 & 0.7 & 0.14 & 0.02 \\ 0.01 & 0.4 & 0.4 & 0.19 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Par exemple, la probabilité de passer d'une note $A$ à la note $C$ vaut 0.009. Cette matrice est en général estimée à partir de données historiques.

On souhaite évaluer le prix d'une assurance qui paye la somme nominale 1 lorsque l'entreprise fait faillite au cours des cinq premières années. Ce type de contrat permet de se couvrir si on a prêté de l'argent à une entité susceptible de faire défaut. Ces contrats sont couramment échangés sur les marchés financiers avec une structure de paiement plus complexe et sont appelés Credit Default Swaps. La somme que paye ce contrat d'assurance est la variable aléatoire $\mathbf{1}_{\{X_4=D\}}$, et sa valeur moyenne est donc $\mathbb{E}(\mathbf{1}_{\{X_4=D\}}) = \mathbb{P}(X_4 = D) = \mu P^4 e_D$ où $\mu$ est la loi initiale et $e_D = [0\ 0\ 0\ 1]'$. Un calcul effectué sous Scilab donne

$$P^4 e_D = [0.0254844\ 0.1384316\ 0.3603744\ 1]'.$$

Le prix de l'assurance vaut donc 0.0254844 si l'entreprise était notée $A$ en 0, 0.1384316 si elle était notée $B$ et 0.3603744 si elle était notée $C$.

Nous voulons maintenant étudier le comportement de la chaîne en temps long ($n \to +\infty$). Nous avons vu que la loi de $X_n$ est donnée par $\mu P^n$. Pour calculer la puissance d'une matrice, une méthode efficace est de la diagonaliser biensûr si cela est possible. Dans le cas présent, nous faisons cela sous Scilab et obtenons $P = QDQ^{-1}$ avec

$$D = \begin{bmatrix} 0.9765546 & 0 & 0 & 0 \\ 0 & 0.7566726 & 0 & 0 \\ 0 & 0 & 0.2667728 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \ Q = \begin{bmatrix} -0.7155985 & -0.4242392 & 0.0315169 & 0.5 \\ -0.5680420 & 0.6085537 & -0.3165533 & 0.5 \\ -0.4065058 & 0.6705844 & 0.9480510 & 0.5 \\ 0 & 0 & 0 & 0.5 \end{bmatrix}.$$

Par conséquent, $P^n = QD^nQ^{-1} \underset{n\to+\infty}{\longrightarrow} Qdiag(0,0,0,1)Q^{-1} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$. Ainsi, pour toute probabilité initiale $\mu$, on obtient que $\mu P^n \underset{n\to+\infty}{\longrightarrow} [0\ 0\ 0\ 1]$, c'est à dire que $X_n$ converge en loi vers une variable aléatoire de loi $\mathbf{1}_{\{x=D\}}$. Autrement dit dans ce modèle, toute entité finit tôt au tard par faire faillite.

### 4.2.2 Modélisation de l'évolution d'une population

Dans ce paragraphe nous allons étudier un modèle proposé par Galton et Watson pour modéliser le nombre d'une population d'individu. C'est en en 1873 que Francis Galton s'est penché sur ce problème, car il avait remarqué l'extinction de noms de famille de la noblesse anglaise, et il souhaitait modéliser et comprendre ce phénomène. Un an plus tard, il proposa avec le concours d'Henry William Watson le modèle que nous allons présenter ici, et qui porte le nom de modèle de Galton-Watson dans la littérature.

Dans ce modèle, l'évolution de la population est modélisée par un processus $(X_n, n \in \mathbb{N})$ à valeurs dans $\mathbb{N}$, où $X_n$ représente le nombre d'individu à la $n$-ième génération. On suppose que chaque individu $i$ de la génération $n$ donne naissance à $\zeta_{i,n}$ enfants qui appartiennent à la génération $n+1$, et meurt. Par conséquent, le nombre d'individus à la génération $n+1$ est donné par :

$$X_{n+1} = \sum_{i=1}^{X_n} \zeta_{i,n}.$$

On fait de plus l'hypothèse que $(\zeta_{i,n}; i, n \in \mathbb{N})$ est une famille de v.a. i.i.d. à valeurs dans $\mathbb{N}$. On note $p_k = \mathbb{P}(\zeta_{1,1} = k)$ et $G$ la fonction génératrice de $\zeta_{1,1}$ :

$$x \in [0,1],\ G(x) = \mathbb{E}(x^{\zeta_{1,1}}) = \sum_{k \in \mathbb{N}} p_k x^k.$$

On suppose que $\zeta_{1,1}$ est de carré intégrable, si bien que $G$ est $\mathcal{C}^\infty$ sur $[0,1)$ et $\mathcal{C}^2$ sur $[0,1]$. On note $m = \mathbb{E}[\zeta_{1,1}]$ et $\sigma^2 = \mathbf{Var}[\zeta_{1,1}]$. Enfin, on prend $X_0 = 1$ comme population initiale.

Le processus $(X_n, n \in \mathbb{N})$ tel qu'il est défini est une chaîne de Markov sur $\mathbb{N}$. En effet, pour $x_0, \ldots, x_{n+1} \in \mathbb{N}$, on a

$$\begin{aligned}
\mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n, \ldots, X_0 = x_0) &= \frac{\mathbb{P}(\sum_{i=1}^{X_n} \zeta_{i,n} = x_{n+1}, X_n = x_n, \ldots, X_0 = x_0)}{\mathbb{P}(X_n = x_n, \ldots, X_0 = x_0)} \\
&= \frac{\mathbb{P}(\sum_{i=1}^{x_n} \zeta_{i,n} = x_{n+1}, X_n = x_n, \ldots, X_0 = x_0)}{\mathbb{P}(X_n = x_n, \ldots, X_0 = x_0)} \\
&= \frac{\mathbb{P}(\sum_{i=1}^{x_n} \zeta_{i,n} = x_{n+1})\mathbb{P}(X_n = x_n, \ldots, X_0 = x_0)}{\mathbb{P}(X_n = x_n, \ldots, X_0 = x_0)} \\
&= \mathbb{P}(\sum_{i=1}^{x_n} \zeta_{i,n} = x_{n+1}).
\end{aligned}$$

car $X_n$ est une fonction de $(\zeta_{i,p}, i \in \mathbb{N}, p \leq n-1)$ et est donc indépendant de $(\zeta_{i,n}, i \in \mathbb{N})$. On pose ainsi $P(x,y) = \mathbb{P}(\sum_{i=1}^{x} \zeta_{i,1} = y)$. C'est bien une matrice de transition (qui ne dépend pas de $n$), et on a $\mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n, \ldots, X_0 = x_0) = P(x_n, x_{n+1})$. Ainsi, $(X_n, n \in \mathbb{N})$ est bien une chaîne de Markov de matrice de transition $P$.

On s'intéresse à la probabilité d'extinction de la population. En posant $\tau = \inf\{n \in \mathbb{N}, X_n = 0\}$, elle s'exprime $\mathbb{P}(\tau < \infty)$. Il est facile de voir que si $X_n = 0$, alors $X_{n+k} = 0$ pour $k \in \mathbb{N}$. Ainsi on obtient :

$$\mathbb{P}(\tau \leq n) = \mathbb{P}(X_n = 0), \text{ puis } \mathbb{P}(X_n = 0) \underset{n\to+\infty}{\to} \mathbb{P}(\tau < \infty). \tag{4.4}$$

Enfin, il est clair que la population s'éteint immédiatement si $p_0 = 1$ et ne s'éteint jamais si $p_0 = 0$. On élimine ces cas triviaux et on suppose

$$0 < p_0 < 1.$$

Pour $n \in \mathbb{N}$, on écrit l'événement $\{X_{n+1} = 0\}$ comme union disjointe d'événements :

$$\{X_{n+1} = 0\} = \{X_n = 0\} \sqcup \left(\sqcup_{k\in\mathbb{N}^*}\{X_n = k, \forall 1 \leq i \leq k, \zeta_{i,n} = 0\}\right).$$

Ainsi, $\mathbb{P}(X_{n+1} = 0) = \mathbb{P}(X_n = 0) + \sum_{k\in\mathbb{N}^*} \mathbb{P}(X_n = k, \forall 1 \leq i \leq k, \zeta_{i,n} = 0)$. Comme $X_n$ est indépendant de $(\zeta_{i,n}, i \in \mathbb{N})$, on en déduit que $\mathbb{P}(X_n = k, \forall 1 \leq i \leq k, \zeta_{i,n} = 0) = \mathbb{P}(X_n = k)\mathbb{P}(\forall 1 \leq i \leq k, \zeta_{i,n} = 0) = \mathbb{P}(X_n = k)p_0^k$. On introduit $G_n(x) = \mathbb{E}[x^{X_n}] = \sum_{k\in\mathbb{N}} \mathbb{P}(X_n = k)x^k$, la fonction génératrice de $X_n$. On voit que

$$\mathbb{P}(X_{n+1} = 0) = G_n(p_0). \tag{4.5}$$

**Lemme 4.2.1.** *La fonction génératrice de $X_n$ est donnée par $G_n(x) = \underbrace{G \circ \cdots \circ G(x)}_{n \text{ fois}}$.*

- *Si $m \leq 1$, $\forall x \in (0,1]$, $G_n(x) \underset{n\to+\infty}{\to} 1$.*
- *Si $m > 1$, $G$ admet un unique point fixe $\eta \in (0,1)$ et $\forall x \in (0,1)$, $G_n(x) \underset{n\to+\infty}{\to} \eta$.*

Grâce à ce lemme, (4.4) et (4.5), il vient que la probabilité d'extinction dans le modèle de Galton-Watson vaut $\mathbb{P}(\tau < \infty) = 1$ si $m \leq 1$ et $\mathbb{P}(\tau < \infty) = \eta \in (0,1)$ si $m > 1$.

*Preuve.* On a par double conditionnement $G_{n+1}(x) = \mathbb{E}[x^{X_{n+1}}] = \mathbb{E}[\mathbb{E}[x^{\sum_{i=1}^{X_n} \zeta_{i,n}}|X_n]]$. Puis en utilisant que $X_n \perp\!\!\!\perp (\zeta_{i,n}, i \in \mathbb{N})$ et la proposition 2.2.16, il vient que $\mathbb{E}[x^{\sum_{i=1}^{X_n} \zeta_{i,n}}|X_n] = G(x)^{X_n}$, puis $G_{n+1}(x) = G_n(G(x))$. Puisque $G_0(x) = x$, on conclut par récurrence que $G_n(x) = \underbrace{G \circ \cdots \circ G(x)}_{n \text{ fois}}$.

Tout d'abord, remarquons que $G$ est convexe, strictement croissante, puis que $G(1) = 1$ et $G'(1) = m$. Elle est même strictement convexe dès qu'il existe $k \geq 2$ tel que $p_k > 0$, ce qui est nécessairement le cas si $m > 1$.

Si $m \leq 1$, $G$ étant au dessus de sa tangente en 1, il vient que $G(x) > x$ pour $x \in (0,1)$. C'est également le cas si $m = 1$. En effet supposons par l'absurde que $G(\eta) = \eta \in (0,1)$, alors puisque $G$ est convexe et au dessus de sa tangente en 1, il vient que $G(x) = x$ pour $x \in [\eta, 1]$. Par unicité du développement en série entière, il vient $p_1 = 1$ puis $p_0 = 0$ ce qui contredit l'hypothèse faite. Par conséquent pour $x \in (0,1)$, $x < G(x) < 1$. La suite $(G_n(x))_n$ est donc strictement croissante bornée par 1, elle admet une limite $l \in (0,1]$. La suite $(G_{n+1}(x))_n$ possède la même limite et par continuité de $G$ sur $[0,1]$, $G(l) = l$. Et nécessairement $l = 1$.

Si $m > 1$, on effectue une étude de la fonction $G(x) - x$. Elle est nulle en 1 et vaut $p_0 > 0$ en 0. Sa dérivée $G'(x) - 1$ est strictement croissante puisque $G$ est strictement convexe, vaut $p_1 - 1 < 0$ en 0 et $m - 1 > 0$ en 1. Elle est donc négative puis positive sur $[0,1]$. Ainsi, $G(x) - x$ est strictement décroissante puis strictement croissante sur $[0,1]$ et s'annule donc en un unique point $\eta \in (0,1)$. Si $x \in (0,\eta)$ (resp. $x \in (\eta,1)$), on a également $x < G(x) < \eta$ (resp. $\eta < G(x) < x$). La suite $(G_n(x))_n$ est strictement croissante (resp. strictement décroissante) et un raisonnement analogue à celui du cas $m \leq 1$ assure que $G_n(x) \underset{n\to+\infty}{\to} \eta$. □

L'exercice suivant permet de décrire plus précisément le comportement de l'évolution de la population lorsque $m > 1$.

**Exercice 4.2.2.** *On suppose $m > 1$, et on pose $M_n = m^{-n} X_n$.*

1. *Montrer que $(M_n, n \in \mathbb{N})$ est une martingale par rapport à la filtration engendrée par $(X_n, n \in \mathbb{N})$. En déduire la valeur de $\mathbb{E}[X_n]$.*
2. *Calculer $\mathbb{E}[X_{n+1}^2 | X_n]$. En déduire que $\mathbf{Var}(M_{n+1}) = \mathbf{Var}(M_n) + \sigma^2/m^{2+n}$ puis que $\mathbf{Var}(M_n) = \frac{\sigma^2}{m^{n+1}} \frac{m^n - 1}{m-1}$.*
3. *Montrer que la suite $(\mathbb{E}[M_n^2])_{n\in\mathbb{N}}$ est bornée. En déduire que $M_n$ converge p.s. et dans $L^2$ vers une v.a. positive $M_\infty$.*
4. *On note pour $x \in (0,1]$, $\phi(x) = \mathbb{E}[x^{M_\infty}]$. Vérifier que $\phi$ est bien définie, continue et croissante sur $(0,1]$. Montrer que $\mathbb{E}[x^{X_{n+1}}] = G(\mathbb{E}[x^{X_n}])$, puis en déduire que $\phi(x) = G(\phi(x^{1/m}))$.*
5. *Vérifier que $\lim_{x\to 0} \phi(x) = \mathbb{P}(M_\infty = 0)$. Calculer $\mathbb{E}[M_\infty]$, puis obtenir grâce à la question précédente que $\mathbb{P}(M_\infty = 0) = \eta$. En déduire que les événements $\{M_\infty = 0\}$ et $\{\tau < \infty\}$ sont identiques à un ensemble négligeable près. Que dire de l'évolution de la population si il n'y a pas extinction ?*

## 4.3 Propriété de Markov forte et excursions d'une chaîne de Markov

On considère un espace d'états au plus dénombrable $E$, $(X_n, n \in \mathbb{N})$ une chaîne de Markov de matrice de transition $P$ et $(\mathcal{F}_n)_{n\in\mathbb{N}}$ la filtration engendrée par cette chaîne de Markov.

On note $E^\infty = \cup_{n\in\mathbb{N}} E^{n+1}$ l'ensemble des suites finies sur $E$. C'est un espace dénombrable puisque c'est une union dénombrable d'espaces dénombrables. Pour $y \in E^\infty$, il existe un unique entier $l \in \mathbb{N}$ t.q. $y \in E^{l+1}$ que l'on note $l(y)$ et qui est appelée longueur de la suite $y$. La loi d'une variable aléatoire $Y : \Omega \to E^\infty$ est caractérisée par $(\mathbb{P}(Y = y), y \in E^\infty)$ ou, ce qui est équivalent, par $(\mathbb{P}[Y = (y_0, \dots, y_n)], y \in E^{\mathbb{N}}, n \in \mathbb{N})$. Comme $\mathbb{P}[Y = (y_0, \dots, y_n)] = \mathbb{P}[Y = (y_0, \dots, y_n), l(Y) = n]$, cette loi est également déterminée par

$$(\mathbb{P}[Y = (y_0, \dots, y_n), l(Y) = n], y \in E^{\mathbb{N}}, n \in \mathbb{N}),$$

ce qui est une écriture parfois plus commode que la précédente. Enfin, deux variables aléatoires à valeurs dans $E^\infty$ sont indépendantes si $\forall y^1, y^2 \in E^\infty, \mathbb{P}[Y_1 = y^1, Y_2 = y_2] = \mathbb{P}[Y_1 = y^1]\mathbb{P}[Y_2 = y^2]$, ou de manière équivalente si $\forall y^1, y^2 \in E^{\mathbb{N}}, \forall n_1, n_2 \in \mathbb{N}$,
$\mathbb{P}[Y_1 = (y^1_0, \dots, y^1_{n_1}), l(Y_1) = n_1, Y_2 = (y^2_0, \dots, y^2_{n_2}), l(Y_2) = n_2]$
$= \mathbb{P}[Y_1 = (y^1_0, \dots, y^1_{n_1}), l(Y_1) = n_1]\mathbb{P}[Y_2 = (y^2_0, \dots, y^2_{n_2}), l(Y_2) = n_2]$.

Ce type de variable aléatoire intervient lorsque l'on considère un processus jusqu'à un temps aléatoire. Nous allons en rencontrer beaucoup dans cette section.

**Théorème 4.3.1** (Propriété de Markov forte)**.** *Soit $\tau$ un temps d'arrêt par rapport à $(\mathcal{F}_n)_{n\in\mathbb{N}}$ tel que $\mathbb{P}(\tau < \infty) > 0$. Alors, conditionnellement à $\{\tau < \infty\}$, $(X_{\tau+n}, n \in \mathbb{N})$ est une chaîne de Markov de matrice de transition $P$ et de loi initiale celle de $X_\tau$ sachant $\{\tau < \infty\}$. En outre, conditionnellement à $\{\tau < \infty\}$ et $X_\tau$, la chaîne de Markov $(X_{\tau+n}, n \in \mathbb{N})$ est indépendante de $(X_0, \dots, X_\tau)$.*

*Preuve.* On se donne $x_0, \dots, x_n \in E$. On calcule : $\mathbb{P}(X_\tau = x_0, \dots, X_{\tau+n} = x_n | \tau < \infty)$
$= \frac{1}{\mathbb{P}(\tau<\infty)} \mathbb{P}(X_\tau = x_0, \dots, X_{\tau+n} = x_n, \tau < \infty)$
$= \frac{1}{\mathbb{P}(\tau<\infty)} \sum_{l\in\mathbb{N}} \mathbb{P}(X_\tau = x_0, \dots, X_{\tau+n} = x_n, \tau = l)$
$= \frac{1}{\mathbb{P}(\tau<\infty)} \sum_{l\in\mathbb{N}} \mathbb{P}(X_l = x_0, \dots, X_{l+n} = x_n, \tau = l)$
$= \frac{1}{\mathbb{P}(\tau<\infty)} \sum_{l\in\mathbb{N}} \mathbb{P}(X_l = x_0, \tau = l) P(x_0, x_1) \dots P(x_{n-1}, x_n)$ d'après le corollaire 4.1.9
$= \mathbb{P}(X_\tau = x_0 | \tau < \infty) P(x_0, x_1) \dots P(x_{n-1}, x_n)$,
puisque $\sum_{l\in\mathbb{N}} \mathbb{P}(X_l = x_0, \tau = l) = \mathbb{P}(X_\tau = x_0, \tau < \infty)$. Cela prouve la première partie du théorème. Remarquons qu'un calcul tout à fait analogue au précédent donne :

$$\mathbb{P}(X_\tau = x_0, \dots, X_{\tau+n} = x_n | \tau < \infty, X_\tau = x') = \mathbf{1}_{\{x_0 = x'\}} P(x_0, x_1) \dots P(x_{n-1}, x_n). \quad (4.6)$$

Il reste à établir l'indépendance. Pour cela, nous devons calculer la loi de $(X_0, \dots, X_{\tau+n})$ conditionnellement à $\{\tau < \infty\}$ et $X_\tau$. Soit $(x_n)_{n\in\mathbb{N}} \in E^{\mathbb{N}}$ et $l \in \mathbb{N}$.
$\mathbb{P}(X_0 = x_0, \dots, X_{l+n} = x_{l+n}, \tau = l | X_\tau = x', \tau < \infty)$
$= \frac{\mathbf{1}_{\{x_l = x'\}}}{\mathbb{P}(X_\tau = x', \tau<\infty)} \mathbb{P}(X_0 = x_0, \dots, X_{l+n} = x_{l+n}, \tau = l)$
$= \frac{\mathbf{1}_{\{x_l = x'\}}}{\mathbb{P}(X_\tau = x', \tau<\infty)} \mathbb{P}(X_0 = x_0, \dots, X_l = x_l, \tau = l) P(x', x_{l+1}) \dots P(x_{l+n-1}, x_{l+n})$ (corollaire 4.1.9)
$= \mathbb{P}(X_0 = x_0, \dots, X_l = x_l, \tau = l | X_\tau = x', \tau < \infty) \; \underbrace{\mathbf{1}_{\{x_l = x'\}} P(x', x_{l+1}) \dots P(x_{l+n-1}, x_{l+n})}_{= \mathbb{P}(X_\tau = x_l, \dots, X_{\tau+n} = x_{\tau+n} | X_\tau = x', \tau<\infty) \text{ d'après } (4.6)}$

Cela prouve l'indépendance voulue. □

Le corollaire suivant correspond à un cas particulier très important.

**Corollaire 4.3.2.** *Si $\tau$ est un $(\mathcal{F}_n)$-temps d'arrêt fini p.s tel que $X_\tau = x$ p.s., alors $(X_{\tau+n}, n \in \mathbb{N})$ est une chaîne de Markov indépendante de $(X_0, \ldots, X_\tau)$. Elle a pour matrice de transition $P$ et est issue de $x$ : elle suit donc la loi de probabilité $\mathbb{P}_x$.*

Nous sommes désormais en mesure de définir les excursions d'une chaîne de Markov. On considère toujours la même chaîne $(X_n, n \in \mathbb{N})$ et on définit pour $x \in E$ :

$$\tau_x^{(0)} = \inf\{n \in \mathbb{N}, X_n = x\} \text{ et } p \in \mathbb{N}, \tau_x^{(p)} = \inf\{n \geq \tau_x^{(p-1)} + 1, X_n = x\}.$$

Ce sont des temps d'arrêt par rapport à $(\mathcal{F}_n)_{n\in\mathbb{N}}$ et $\tau_x^{(p)}$ représente le $p+1$-ième temps d'atteinte de la valeur $x$ par la chaîne de Markov. Ces temps peuvent être finis ou infinis, et la suite $(\tau_x^{(p)})_{p\in\mathbb{N}}$ est strictement croissante jusqu'à ce qu'elle atteigne $+\infty$, si elle l'atteint. On définit également

$$\tau_x = \inf\{n \in \mathbb{N}^*, X_n = x\} = \mathbf{1}_{\{X_0=x\}}\tau_x^{(1)} + \mathbf{1}_{\{X_0\neq x\}}\tau_x^{(0)}$$

qui est le premier temps d'atteinte de $x$ après l'instant initial.

**Définition 4.3.3.** *On définit pour $n \geq 1$,*

$$E^{n,x} = \{(x_0, \ldots, x_n) \in E^{n+1} \text{ t.q. } x_0 = x_n = x \text{ et } x_i \neq x \text{ pour } 1 \leq i \leq n-1\},$$

*que l'on appelle ensemble des excursions autour de $x$ de longueur $n$ et*

$$\text{Exc}(x) = \cup_{n\geq 1} E^{n,x} \subset E^\infty$$

*l'ensemble des excursions autour de $x$.*

Si l'on se place sur $\{\tau_x^{(p)} < \infty\}$, on peut définir :
- $D^x = (X_0, \ldots, X_{\tau_x^{(0)}})$ le début de la chaîne de Markov. ($D^x = (x)$ si $X_0 = x$)
- Pour $l \in \{1, \ldots, p\}$, $\xi^{x,l} = (X_{\tau_x^{(l-1)}}, X_{\tau_x^{(l-1)}+1}, \ldots, X_{\tau_x^{(l)}})$ la $l$-ième excursion de la chaîne de Markov autour de $x$. Ce sont des variables aléatoires à valeurs dans $\text{Exc}(x)$.
- $R^x = (X_{\tau_x^{(p)}+n}, n \in \mathbb{N})$, le reste de la chaîne.

Nous avons le théorème général suivant.

**Théorème 4.3.4.** *Avec les notations précédentes, conditionnellement à $\{\tau_x^{(p)} < \infty\}$ on a :*
- *le début $D^x$, les excursions $\xi^{x,1}, \ldots, \xi^{x,p}$ et le reste $R^x$ sont indépendants,*
- *$R^x$ est une chaîne de Markov de loi $\mathbb{P}_x$,*
- *les excursions sont identiquement distribuées et suivent la loi de l'excursion $(X_0, \ldots, X_{\tau_x})$ sous $\mathbb{P}_x$ sachant $\{\tau_x < \infty\}$. On note $\xi^x$ cette loi.*

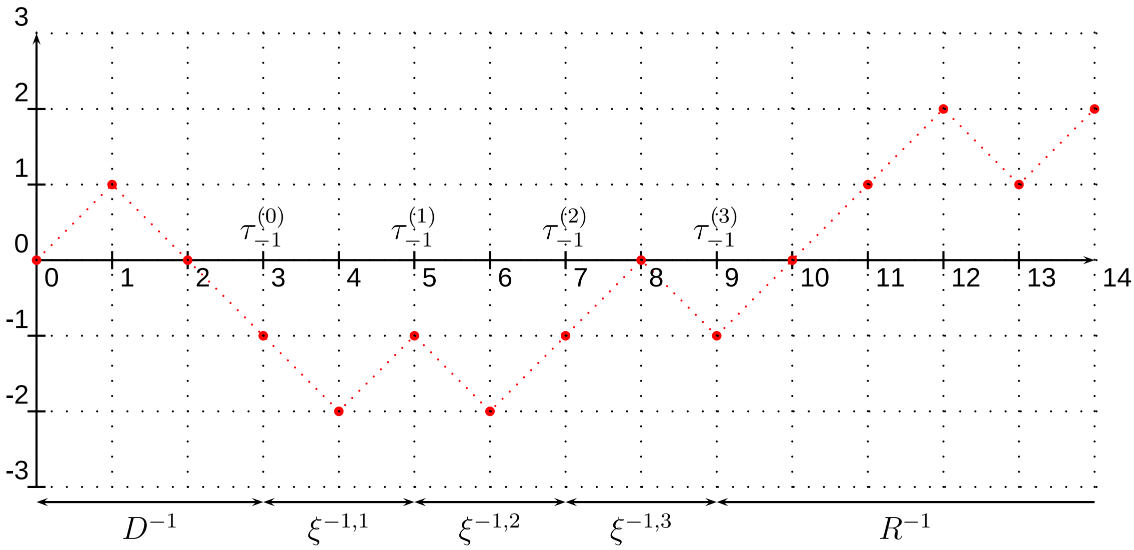


FIG. 4.1 – Illustration du découpage ci-dessus pour une marche aléatoire sur $\mathbb{Z}$ jusqu'à la troisième excursion autour de $-1$.

*Preuve.* Tout d'abord, nous notons pour $l \in \mathbb{N}$, $\tilde{\tau}_x^{(l)}$ le $l$-ème temps d'atteinte de $x$ par le processus $(X_{\tau_x^{(0)}+n}, n \in \mathbb{N})$ : $\tilde{\tau}_x^{(l)} = \inf\{n \geq \tilde{\tau}_x^{(l-1)} + 1, X_{\tau^{(0)}+n} = x\}$ et $\tilde{\tau}_x^{(0)} = 0$ sur $\{\tau_x^{(0)} < \infty\}$. On a clairement par construction :

$$\tau_x^{(l)} = \tilde{\tau}_x^{(l)} + \tau_x^{(0)}.$$

Pour $x \in E^{\mathbb{N}}, n \in \mathbb{N}, i_0 < i_1 < \cdots < i_p \in \mathbb{N}$, on calcule
$A = \mathbb{P}(D^x = (x_0, \dots, x_{i_0}), \xi^{x,1} = (x_{i_0}, \dots, x_{i_1}), \dots, \xi^{x,p} = (x_{i_{p-1}}, \dots, x_{i_p}), R_0^x = x_{i_p}, \dots, R_n^x = x_{i_p+n} | \tau_x^{(p)} < \infty) = \mathbb{P}(X_0 = x_0, \dots, X_{i_p+n} = x_{i_p+n}, \tau_x^{(0)} = i_0, \dots, \tau_x^{(p)} = i_p | \tau_x^{(p)} < \infty)$
$= \frac{\mathbb{P}(X_0=x_0,\dots,X_{i_p+n}=x_{i_p+n},\tau_x^{(0)}=i_0,\dots,\tau_x^{(p)}=i_p)}{\mathbb{P}(\tau_x^{(p)}<\infty)}$. Puisque $\tau_x^{(p)} = \tilde{\tau}_x^{(p)} + \tau_x^{(0)}$, $\{\tau_x^{(p)} < \infty\} = \{\tilde{\tau}_x^{(p)} < \infty, \tau_x^{(0)} < \infty\}$. En divisant numérateur et dénominateur par $\mathbb{P}(\tau_x^{(0)} < \infty)$, on obtient :
$A = \frac{\mathbb{P}(X_0=x_0,\dots,X_{i_p+n},\tau_x^{(0)}=i_0,\dots,\tau_x^{(p)}=i_p|\tau_x^{(0)}<\infty)}{\mathbb{P}(\tilde{\tau}_x^{(p)}<\infty|\tau_x^{(0)}<\infty)}$

$$= \frac{\mathbb{P}(\overbrace{X_0 = x_0, \dots, X_{i_0} = x_{i_0}, \tau_x^{(0)} = i_0}^{\text{dépend de } (X_0,\dots,X_{\tau_x^{(0)}})}, \overbrace{X_{i_0} = x_{i_0}, \dots, X_{i_p+n} = x_{i_p+n}, \tilde{\tau}_x^{(1)} = i_1 - i_0, \dots, \tilde{\tau}_x^{(p)} = i_p - i_0}^{\text{dépend de } (X_{\tau_x^{(0)}+n}, n\in\mathbb{N})} | \tau_x^{(0)}<\infty)}{\mathbb{P}(\tilde{\tau}_x^{(p)}<\infty|\tau_x^{(0)}<\infty)}$$

En appliquant la propriété de Markov forte au numérateur et au dénominateur, il vient :
$A = \mathbb{P}(X_0 = x_0, \dots, X_{i_0} = x_{i_0}, \tau_x^{(0)} = i_0 | \tau_x^{(0)} < \infty) \times \frac{\mathbb{P}_x(X_0=x_{i_0},\dots,X_{i_p+n-i_0}=x_{i_p+n},\tau_x^{(1)}=i_1-i_0,\dots,\tau_x^{(p)}=i_p-i_0)}{\mathbb{P}_x(\tau_x^{(p)}<\infty)}$
$= \mathbb{P}(X_0 = x_0, \dots, X_{i_0} = x_{i_0}, \tau_x^{(0)} = i_0 | \tau_x^{(0)} < \infty) \times \mathbb{P}_x(X_0 = x_{i_0}, \dots, X_{i_p+n-i_0} = x_{i_p+n}, \tau_x^{(1)} = i_1 - i_0, \dots, \tau_x^{(p)} = i_p - i_0 | \tau_x^{(p)} < \infty)$. En reprenant exactement le même raisonnement, on obtient que $\mathbb{P}_x(X_0 = x_{i_0}, \dots, X_{i_p+n-i_0} = x_{i_p+n}, \tau_x^{(1)} = i_1 - i_0, \dots, \tau_x^{(p)} = i_p - i_0 | \tau_x^{(p)} < \infty)$
$= \mathbb{P}_x(X_0 = x_{i_0}, \dots, X_{i_1-i_0} = x_{i_1}, \tau_x^{(1)} = i_1 - i_0 | \tau_x^{(1)} < \infty)$
$\times \mathbb{P}_x(X_0 = x_{i_1}, \dots, X_{i_p+n-i_1} = x_{i_p+n}, \tau_x^{(1)} = i_2 - i_1, \dots, \tau_x^{(p-1)} = i_p - i_1 | \tau_x^{(p-1)} < \infty)$ en appliquant la propriété de Markov forte pour le temps d'arrêt $\tau_x^{(1)}$. Puis en itérant ce même raisonnement encore $p-1$ fois, il vient :

$A = \mathbb{P}(X_0 = x_0, \ldots, X_{i_0} = x_{i_0}, \tau_x^{(0)} = i_0 | \tau_x^{(0)} < \infty)$
$\times \mathbb{P}_x(X_0 = x_{i_0}, \ldots, X_{i_1 - i_0} = x_{i_1}, \tau_x^{(1)} = i_1 - i_0 | \tau_x^{(1)} < \infty)$
$\times \cdots \times \mathbb{P}_x(X_0 = x_{i_{p-1}}, \ldots, X_{i_p - i_{p-1}} = x_{i_p}, \tau_x^{(1)} = i_p - i_{p-1} | \tau_x^{(1)} < \infty)$
$\times \mathbb{P}_x(X_0 = x_{i_p}, \ldots, X_{i_p+n} = x_{i_p+n})$ ce qui est le résultat voulu. □

Si les variables $D^x$, $\xi^{x,1}, \ldots, \xi^{x,p}$ sont indépendantes, leurs longueurs sont en particulier indépendantes ce qui donne le corollaire suivant.

**Corollaire 4.3.5.** *Conditionnellement à $\{\tau_x^{(p)} < \infty\}$, les v.a. $\tau_x^{(0)}, \tau_x^{(1)} - \tau_x^{(0)}, \ldots, \tau_x^{(p)} - \tau_x^{(p-1)}$ sont indépendantes, et les v.a. $\tau_x^{(1)} - \tau_x^{(0)}, \ldots, \tau_x^{(p)} - \tau_x^{(p-1)}$ sont identiquement distribuées.*

**Remarque 4.3.6.** *Il est important de remarquer que la loi d'une excursion $\xi^x$ ne dépend pas de la loi initiale de la chaîne de Markov. Pour $e = (x, x_1, \ldots, x_{n-1}, x) \in \textsc{Exc}(x)$, on a*

$$\begin{aligned}
\mathbb{P}(\xi^x = e) &= \mathbb{P}_x(X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, \tau_x = n | \tau_x < \infty) \\
&= \frac{\mathbb{P}_x(X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = x)}{\sum_{q \in \mathbb{N}^*} \mathbb{P}(\tau_x = q)} \\
&= \frac{P(x, x_1) P(x_1, x_2) \ldots P(x_{n-1}, x)}{\sum_{q \in \mathbb{N}^*} \sum_{z_1, \ldots, z_{q-1} \neq x} P(x, z_1) P(z_1, z_2) \ldots P(z_{q-1}, x)},
\end{aligned}$$

*et cette loi ne dépend que de la matrice de transition $P$.*

## 4.4 Récurrence, transience et irréductibilité

Comme précédemment, on considère une chaîne de Markov $(X_n, n \in \mathbb{N})$ de matrice de transition $P$. On note pour $x \in E$,

$$N_x = \sum_{n \in \mathbb{N}} \mathbf{1}_{\{X_n = x\}},$$

le nombre (éventuellement infini) de fois que la chaîne visite l'état $x$. Remarquons que l'on a également $N_x = \sup\{p \in \mathbb{N}, \tau_x^{(p)} < \infty\} + 1$ (avec pour convention $\sup \emptyset = -1$).

**Définition 4.4.1.** *Un état $x \in E$ est dit récurrent si $\mathbb{P}_x(\tau_x < \infty) = 1$, il est dit transient lorsque $\mathbb{P}_x(\tau_x < \infty) < 1$.*

On se place sous $\mathbb{P}_x$ si bien que $\tau_x = \tau_x^{(1)}$ et le début $D^x = (x)$. On suppose que $x$ est récurrent et nous allons réétablir dans ce cas particulier les résultats du Théorème 4.3.4. La preuve est en effet plus lisible dans ce cas, car il n'y a plus de conditionnement. Grâce au Corollaire 4.3.2, $(X_{\tau_x+n}, n \in \mathbb{N})$ est une chaîne de Markov issue de $x$, de matrice de transition $P$, indépendante de l'excursion $\xi^{x,1} = (X_0, \ldots, X_{\tau_x})$. Par conséquent $\tau_x^{(2)} - \tau_x^{(1)}$

qui est le premier temps d'atteinte après 0 de $x$ par $(X_{\tau_x^{(1)}+n}, n \in \mathbb{N})$ a même loi que $\tau^{(1)}$ et est fini p.s. En appliquant la propriété de Markov forte à cette nouvelle chaîne de Markov, il vient que la chaine $(X_{\tau_x^{(2)}+n}, n \in \mathbb{N})$ est indépendante de $\xi^{x,2}$, issue de $x$ et de matrice de transition $P$. En itérant $p$ fois, on obtient que :

$$\begin{array}{lcl}
\xi^{x,1} & \perp\!\!\!\perp & \xi^{x,2}, \ldots, \xi^{x,p}, (X_{\tau(p)+n}, n \in \mathbb{N}) \\
\xi^{x,2} & \perp\!\!\!\perp & \xi^{x,3}, \ldots, \xi^{x,p}, (X_{\tau(p)+n}, n \in \mathbb{N}) \\
 & \ldots & \\
\xi^{x,p} & \perp\!\!\!\perp & (X_{\tau(p)+n}, n \in \mathbb{N})
\end{array}$$

Par conséquent, $\xi^{x,1}, \xi^{x,2}, \ldots, \xi^{x,p}, (X_{\tau(p)+n}, n \in \mathbb{N})$ sont indépendants et la suite des excursions $(\xi^{x,n})_{n\in\mathbb{N}^*}$ est une suite de v.a. indépendantes et distribuées selon la même loi. En effet, pour $z \in \textsc{Exc}(x)$, $\mathbb{P}_x(\xi^{x,i} = z) = \mathbb{P}_x(X_0 = x, X_1 = z_1, \ldots, X_{n-1} = z_{n-1}, X_n = x) = P(x, z_1)P(z_1, z_2)\ldots P(z_{n-1}, x)$. En particulier, leurs longueurs

$$\tau_x^{(1)}, \ldots, \tau_x^{(p)} - \tau_x^{(p-1)}, \ldots \text{ sont i.i.d.}$$

Ainsi, puisque $\mathbb{P}_x(\tau_x < \infty) = 1$ et $\tau_x = \tau_x^{(1)}$ sous $\mathbb{P}_x$, elles sont toutes finies p.s. Par conséquent pour tout $p \in \mathbb{N}$, $\tau_x^{(p)}$ est fini p.s. et $N_x = \sup\{p \in \mathbb{N}, \tau_x^{(p)} < \infty\} + 1 = +\infty$.

Lorsque $x$ est transient et $k \in \mathbb{N}^*$, on obtient en utilisant la propriété de Markov forte :

$$\begin{aligned}
\mathbb{P}_x(N_x = k) &= \mathbb{P}_x(\tau_x < \infty, \sum_{n\in\mathbb{N}} \mathbf{1}_{\{X_{\tau_x+n}=x\}} = k-1) \\
&= \mathbb{P}_x(\tau_x < \infty)\mathbb{P}_x(\sum_{n\in\mathbb{N}} \mathbf{1}_{\{X_{\tau_x+n}=x\}} = k-1 | \tau_x < \infty) \\
&= \mathbb{P}_x(\tau_x < \infty)\mathbb{P}_x(N_x = k-1).
\end{aligned}$$

On en déduit le résultat suivant.

**Proposition 4.4.2.** *Si $x \in E$ est un point récurrent, $\mathbb{P}_x(N_x = +\infty) = 1$. Sous $\mathbb{P}_x$, la suite des excursions $(\xi^{x,i})_{i\in\mathbb{N}^*}$ autour de $x$ est bien définie et est i.i.d.*

*Sinon, $x$ est transient, et $N_x$ suit une loi géométrique de paramètre $1 - \mathbb{P}_x(\tau_x < \infty)$ :*

$$\forall k \in \mathbb{N}^*, \ \mathbb{P}_x(N_x = k) = \mathbb{P}_x(\tau_x < \infty)^{k-1}(1 - \mathbb{P}_x(\tau_x < \infty))$$

*sous $\mathbb{P}_x$. En particulier, $\mathbb{P}_x(N_x < +\infty) = 1$.*

Donnons un exemple en considérant la chaîne à deux états $\{0, 1\}$ t.q. $P(0, 1) = p = 1 - P(0, 0)$ et $P(1, 0) = q = 1 - P(1, 1)$. On suppose $p, q > 0$ et on considère les excursions $(\xi^{0,i})_{i\in\mathbb{N}^*}$ autour de 0. Alors,

$$\mathbb{P}_0(\xi^{0,i} = (0, \underbrace{1, \ldots, 1}_{n \text{ fois}}, 0)) = \begin{cases} p(1-q)^{n-1}q \text{ si } n > 0 \\ 1 - p \text{ si } n = 0. \end{cases}$$

On dit que l'état $x$ conduit à l'état $y$ (et on note $x \dashrightarrow y$) si l'on a une probabilité strictement positive d'aller en un temps fini à l'état $y$ en partant de $x$, i.e. si $\mathbb{P}_x(\tau_y < \infty) > 0$.

**Lemme 4.4.3.** *On a :* $x \dashrightarrow y \iff \exists n \in \mathbb{N}^*,\ P^n(x,y) > 0$.

*Preuve.* Supposons $x \dashrightarrow y$. Comme $\{\tau_y < \infty\} = \cup_{n\in\mathbb{N}^*}\{X_n = y\}$, $\mathbb{P}_x(\tau_y < \infty) \leq \sum_{n\in\mathbb{N}} P^n(x,y)$ et il existe $n \in \mathbb{N}^*$ tel que $P^n(x,y) > 0$. Réciproquement, comme $\{X_n = y\} \subset \{\tau_y \leq n\}$, $P^n(x,y) > 0$ implique nécessairement que $\mathbb{P}_x(\tau_y < \infty) > 0$. □

Nous avons le résultat suivant.

**Proposition 4.4.4.** *Soient $x, y \in E$. Si $x$ est un état récurrent et si $x \dashrightarrow y$, alors $y$ est un état récurrent et $y \dashrightarrow x$. En outre, $\mathbb{P}_x(N_y = +\infty) = \mathbb{P}_x(\tau_y < +\infty) = 1$ et la probabilité qu'une excursion autour de $x$ passe par $y$ est strictement positive : $\mathbb{P}(y \in \xi^x) > 0$.*

*Preuve.* On note $\{y \in \xi^{x,n}\}$ l'événement : "la $n$-ième excursion passe par $y$". Comme $x$ est récurrent, on a sous $\mathbb{P}_x$ : $\{\tau_y < \infty\} = \cup_{n\in\mathbb{N}}\{y \in \xi^{x,n}\}$. Ainsi, $0 < \mathbb{P}_x(\tau_y < \infty) \leq \sum_{n\in\mathbb{N}} \mathbb{P}_x(y \in \xi^{x,n}) = \sum_{n\in\mathbb{N}} \mathbb{P}(y \in \xi^x)$ et donc $\mathbb{P}(y \in \xi^x) > 0$. On remarque maintenant que pour tout $n \in \mathbb{N}$,

$$N_y \geq \sum_{i=1}^{n} \mathbf{1}_{\{y\in\xi^{x,i}\}}.$$

Les v.a. $\mathbf{1}_{\{y\in\xi^{x,i}\}}$ sont indépendantes et suivent une loi de Bernoulli de paramètre $\mathbb{P}(y \in \xi^x)$. Par la loi forte des grands nombres, il vient que $\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{y\in\xi^{x,i}\}} \underset{n\to+\infty}{\longrightarrow} \mathbb{P}(y \in \xi^x)$ $\mathbb{P}_x$-p.s., et en particulier $\sum_{i=1}^{n} \mathbf{1}_{\{y\in\xi^{x,i}\}} \underset{n\to+\infty}{\longrightarrow} +\infty$ $\mathbb{P}_x$-p.s. On en déduit que $\mathbb{P}_x(N_y = +\infty) = 1$. En particulier $\mathbb{P}_x(\tau_y < +\infty) = 1$, et en utilisant la propriété de Markov forte en $\tau_y$, il vient d'une part $\mathbb{P}_x(N_y = +\infty) = \mathbb{P}_x(\tau_y < +\infty)\mathbb{P}_y(N_y = +\infty)$ puis $\mathbb{P}_y(N_y = +\infty) = 1$ ($y$ est ainsi récurrent) et d'autre part $\mathbb{P}_x(N_x = +\infty) = \mathbb{P}_x(\tau_y < +\infty)\mathbb{P}_y(N_x = +\infty)$ puis $\mathbb{P}_y(N_x = +\infty) = 1$ (et donc $y \dashrightarrow x$). □

On dira que deux états $x, y \in E$ communiquent entre eux si $x \dashrightarrow y$ et $y \dashrightarrow x$, et on note dans ce cas $x \leftarrow\dashrightarrow y$. On dira qu'une chaîne de Markov est *irréductible* si

$$\forall x, y \in E,\ x \leftarrow\dashrightarrow y. \tag{4.7}$$

Dans ce cas et grâce à la proposition 4.4.4, les états sont soit tous récurrents, soit tous transients. Dans le premier cas, on dit que la chaîne est *irréductible récurrente.* Alors dans ce cas, toujours grâce à la proposition 4.4.4, on a $\mathbb{P}(N_x = +\infty) = \mathbb{P}(\tau_x < \infty) = 1$ **quelque soit la loi initiale de** $X_0$ car $\mathbb{P}(N_x = +\infty) = \sum_{y\in E} \mathbb{P}(X_0 = y)\mathbb{P}_y(N_x = +\infty) = 1$.

**Proposition 4.4.5.** *Soit $(X_n, n \in \mathbb{N})$ une chaîne de Markov de matrice de transition $P$ irréductible. Si elle possède une probabilité invariante $\mu$, alors la chaîne est récurrente.*

*Preuve.* On considère la chaîne de loi initiale $\mu$ et $y \in E$ tel que $\mu(y) > 0$. D'après la proposition 4.4.2, soit $y$ est récurrent et alors $E_y(N_y) = +\infty$, soit $y$ est transient et alors $E_y(N_y) < +\infty$ puisque $N_y$ suit une loi géométrique sous $\mathbb{P}_y$. On a :

$$\mathbb{E}[N_y] = \mathbb{E}\left[\sum_{n\in\mathbb{N}} \mathbf{1}_{\{X_n=y\}}\right] = \sum_{n\in\mathbb{N}} \mathbb{P}(X_n = y) = \sum_{n\in\mathbb{N}} \mu(y) = +\infty.$$

Par ailleurs, $\mathbb{E}[N_y] = \sum_{x\in E} \mathbb{P}(X_0 = x)\mathbb{E}[N_y|X_0 = x] = \sum_{x\in E} \mu(x)\mathbb{E}_x[N_y]$. Grâce à la propriété de Markov forte en $\tau_y^{(0)}$, $\mathbb{E}_x[N_y] = \mathbb{E}_x[\mathbb{E}_x[N_y|\mathbf{1}_{\{\tau_y^{(0)}<\infty\}}]] = \mathbb{E}_x[\mathbf{1}_{\{\tau_y^{(0)}<\infty\}}\mathbb{E}_y[N_y]] = \mathbb{P}_x(\tau_y^{(0)} < \infty)\mathbb{E}_y[N_y]$. Par conséquent,

$$+\infty = \mathbb{E}[N_y] = \left(\sum_{x\in E} \mu(x)\mathbb{P}_x(\tau_y^{(0)} < \infty)\right)\mathbb{E}_y[N_y]$$

Puisque la chaîne est irréductible, on a $0 < \mathbb{P}_x(\tau_y^{(0)} < \infty)$ et donc $0 < \sum_{x\in E} \mu(x)\mathbb{P}_x(\tau_y^{(0)} < \infty) \leq 1$. Par conséquent $\mathbb{E}_y[N_y] = +\infty$. L'état $y$ est donc récurrent et la chaîne est donc irréductible récurrente. □

Néanmoins, il existe des chaînes de Markov non irréductibles qui possèdent une probabilité invariante comme le montre l'exercice suivant.

**Exercice 4.4.6.** *Soit $E$ un espace d'état dénombrable et $y_0 \in E$. Vérifier que $P(x, y) = \mathbf{1}_{\{y=y_0\}}$ définit une matrice de transition. Montrer que cette chaîne n'est pas irréductible, mais admet $\mathbf{1}_{\{y=y_0\}}$ comme probabilité invariante.*

Grâce à l'exercice 4.1.5, nous déduisons de la proposition 4.4.5 le corollaire suivant.

**Corollaire 4.4.7.** *Une chaîne de Markov irréductible à valeurs dans un espace d'états fini est récurrente.*

L'exercice suivant propose une preuve plus directe de ce résultat.

**Exercice 4.4.8.** *On considère une chaîne de Markov irréductible sur un espace $E$ fini.*

1. *Montrer que quelque soit la loi initiale, $\sum_{y\in E} N_y = +\infty$ et en déduire que pour tout $x \in E$, il existe $y \in E$ t.q. $\mathbb{P}_x(N_y = +\infty) = 1$.*
2. *Montrer que $\mathbb{P}_x(N_y = +\infty) = \mathbb{P}_x(\tau_y^{(0)} < \infty)\mathbb{P}_y(N_y = +\infty)$ et en déduire le corollaire précédent.*

On conclut cette section avec un lemme utile concernant les mesures invariantes pour une chaîne de Markov récurrente. Une mesure invariante est une mesure $\nu$ sur $E$ telle que $\nu P = \nu$, c'est à dire $\forall x \in E, \nu(x) = \sum_{y\in E} \nu(y)P(y, x)$.

**Lemme 4.4.9.** *Si $\nu$ est une mesure invariante pour une chaîne de Markov irréductible telle qu'il existe $x_0 \in E$, $0 < \nu(x_0) < \infty$. Alors on a :*

$$\forall x \in E, 0 < \nu(x) < \infty.$$

*En particulier une probabilité invariante $\mu$ satisfait nécessairement $\forall x \in E, \mu(x) > 0$.*

*Preuve.* Il est facile d'établir par récurrence $\nu P^n = \nu$ pour $n \in \mathbb{N}$. Soit $x \in E$. La chaîne étant irréductible, $x_0 \dashrightarrow x$ et grâce au lemme 4.4.3, il existe $n \in \mathbb{N}$ tel que $P^n(x_0, x) > 0$. On en déduit $\nu(x) = \sum_{y \in E} \nu(y) P^n(y, x) \geq \nu(x_0) P^n(x_0, x) > 0$. Puisque on a aussi $x \dashrightarrow x_0$, le même raisonnement assure qu'il existe $m \in \mathbb{N}$ tel que $P^m(x, x_0) > 0$ et $\nu(x_0) \geq \nu(x) P^m(x, x_0)$ et donc $\nu(x) < \infty$. □

## 4.5 Théorème ergodique

Dans toute cette section, on considère une chaîne de Markov irréductible et récurrente de matrice de transition $P$. Nous avons la dichotomie suivante.

**Théorème 4.5.1.** *Soit $(X_n, n \in \mathbb{N})$ une chaîne de Markov de matrice de transition $P$ irréductible et récurrente. Alors, on est dans l'un des deux cas suivants :*

$$\forall x \in E,\ \mathbb{E}_x[\tau_x] < +\infty \text{ et la chaîne est dite récurrente positive,}$$
$$\forall x \in E,\ \mathbb{E}_x[\tau_x] = +\infty \text{ et la chaîne est dite récurrente nulle.}$$

Nous donnons ici une preuve probabiliste directe de ce résultat qui exploite et met en valeurs les résultats sur les excursions. Dans une première lecture, on pourra admettre ce résultat.

*Preuve.* On suppose que $\mathbb{E}_x[\tau_x] < \infty$ pour un certain $x \in E$, et il s'agit de montrer qu'alors, $\mathbb{E}_y[\tau_y] < \infty$ pour tout $y \in E$. La proposition 4.4.4 nous assure que $\mathbb{P}_x(\tau_y^{(0)} < \infty) = 1$. Puisque $\tau_y^{(1)} = \tau_y^{(0)} + \inf\{n > 0, X_{\tau_y^{(0)}+n} = y\}$, la propriété de Markov forte donne $\mathbb{E}_x[\tau_y^{(1)}] = \mathbb{E}_x[\tau_y^{(0)}] + \mathbb{E}_y[\tau_y]$. Nous allons montrer que $\mathbb{E}_x[\tau_y^{(0)}]$ puis $\mathbb{E}_x[\tau_y^{(1)}]$ sont finis ce qui prouvera le résultat. En fait, il suffirait même de montrer seulement que $\mathbb{E}_x[\tau_y^{(1)}] < \infty$, mais nous donnons d'abord la preuve de $\mathbb{E}_x[\tau_y^{(0)}] < \infty$ qui est plus facilement accessible.

On a $\tau_y^{(0)} = \sum_{p=1}^{\infty} \tau_y^{(0)} \mathbf{1}_{\{\tau_x^{(p-1)} < \tau_y^{(0)} < \tau_x^{(p)}\}} \leq \sum_{p=1}^{\infty} \tau_x^{(p)} \mathbf{1}_{\{\tau_x^{(p-1)} < \tau_y^{(0)} < \tau_x^{(p)}\}}$. Observons maintenant que $\mathbf{1}_{\{\tau_x^{(p-1)} < \tau_y^{(0)} < \tau_x^{(p)}\}} = \mathbf{1}_{\{y \notin \xi^{x,1}\}} \dots \mathbf{1}_{\{y \notin \xi^{x,p-1}\}} \mathbf{1}_{\{y \in \xi^{x,p}\}}$ et que, sous $\mathbb{P}_x$, $\tau_x^{(p)} = \sum_{i=1}^{p} \tau_x^{(i)} - \tau_x^{(i-1)} = \sum_{i=1}^{p} l(\xi^{x,i})$ où $l(\xi^{x,i})$ désigne la longueur de la $i$-ème excursion. On a donc, puisque les excursions sont i.i.d. de loi $\xi^x$,

$$\begin{aligned}\mathbb{E}_x[\tau_x^{(p)}\mathbf{1}_{\{\tau_x^{(p-1)}<\tau_y^{(0)}<\tau_x^{(p)}\}}] &= \sum_{i=1}^{p}\mathbb{E}_x[l(\xi^{x,i})\mathbf{1}_{\{y\notin\xi^{x,1}\}}\dots\mathbf{1}_{\{y\notin\xi^{x,p-1}\}}\mathbf{1}_{\{y\in\xi^{x,p}\}}] \\ &= (p-1)\mathbb{P}(y\notin\xi^x)^{p-2}\mathbb{P}(y\in\xi^x)\mathbb{E}[l(\xi^x)\mathbf{1}_{\{y\notin\xi^x\}}] \\ &\quad +\mathbb{P}(y\notin\xi^x)^{p-1}\mathbb{E}[l(\xi^x)\mathbf{1}_{\{y\in\xi^x\}}].\end{aligned}$$

Notons que $\mathbb{E}_x[\tau_x] = \mathbb{E}_x[\tau_x\mathbf{1}_{\{\tau_x<\tau_y\}}]+\mathbb{E}_x[\tau_x\mathbf{1}_{\{\tau_y<\tau_x\}}] = \mathbb{E}[l(\xi^x)\mathbf{1}_{\{y\in\xi^x\}}]+\mathbb{E}[l(\xi^x)\mathbf{1}_{\{y\notin\xi^x\}}]$, et par conséquent $\mathbb{E}[l(\xi^x)\mathbf{1}_{\{y\in\xi^x\}}]<\infty$ et $\mathbb{E}[l(\xi^x)\mathbf{1}_{\{y\notin\xi^x\}}]<\infty$. Ainsi, on obtient en sommant

$$\mathbb{E}_x[\tau_y^{(0)}] \leq \sum_{p=1}^{\infty}\mathbb{E}_x[\tau_x^{(p)}\mathbf{1}_{\{\tau_x^{(p-1)}<\tau_y^{(0)}<\tau_x^{(p)}\}}] < \infty.$$

La preuve pour établir $\mathbb{E}_x[\tau_y^{(1)}]<\infty$ suit exactement le même principe. On introduit pour une excursion $\xi^x$, $n_y(\xi^x) = \sum_{n=0}^{l(\xi^x)-1}\mathbf{1}_{\{\xi_n^x=y\}}$ le nombre de fois que l'excursion passe par $y$. On écrit :

$$\tau_y^{(1)} \leq \sum_{p=1}^{\infty}\tau_x^{(p)}\mathbf{1}_{\{\tau_x^{(p-1)}<\tau_y^{(1)}<\tau_x^{(p)}\}} = \sum_{p=1}^{\infty}\sum_{q=1}^{p}\sum_{i=1}^{p} l(\xi^{x,i})\mathbf{1}_{\{\tau_x^{(p-1)}<\tau_y^{(1)}<\tau_x^{(p)}\}}\mathbf{1}_{\{\tau_x^{(q-1)}<\tau_y^{(0)}<\tau_x^{(q)}\}}. \tag{4.8}$$

Remarquons que $\mathbf{1}_{\{\tau_x^{(p-1)}<\tau_y^{(1)}<\tau_x^{(p)}\}}\mathbf{1}_{\{\tau_x^{(q-1)}<\tau_y^{(0)}<\tau_x^{(q)}\}} = \mathbf{1}_{\{n_y(\xi^{x,q})=1\}}\mathbf{1}_{\{n_y(\xi^{x,p})\geq 1\}}\prod_{i<p,i\neq q}\mathbf{1}_{\{n_y(\xi^{x,i})=0\}}$ pour $q<p$ et $\mathbf{1}_{\{\tau_x^{(p-1)}<\tau_y^{(1)}<\tau_x^{(p)}\}}\mathbf{1}_{\{\tau_x^{(q-1)}<\tau_y^{(0)}<\tau_x^{(q)}\}} = \mathbf{1}_{\{n_y(\xi^{x,p})\geq 2\}}\prod_{i<p}\mathbf{1}_{\{n_y(\xi^{x,i})=0\}}$ pour $q=p$. Comme les excursions sont i.i.d, pour $q<p$, on a

$$\begin{aligned}&\mathbb{E}_x[l(\xi^{x,i})\mathbf{1}_{\{\tau_x^{(p-1)}<\tau_y^{(1)}<\tau_x^{(p)}\}}\mathbf{1}_{\{\tau_x^{(q-1)}<\tau_y^{(0)}<\tau_x^{(q)}\}}] \\ &= \begin{cases}\mathbb{P}(n_y(\xi^x)=1)\mathbb{P}(y\notin\xi^x)^{p-2}\mathbb{E}[l(\xi^x)\mathbf{1}_{\{n_y(\xi^x)\geq 1\}}] \text{ si } i=p \\ \mathbb{P}(n_y(\xi^x)\geq 1)\mathbb{P}(y\notin\xi^x)^{p-2}\mathbb{E}[l(\xi^x)\mathbf{1}_{\{n_y(\xi^x)=1\}}] \text{ si } i=q \\ \mathbb{P}(n_y(\xi^x)=1)\mathbb{P}(n_y(\xi^x)\geq 1)\mathbb{P}(y\notin\xi^x)^{p-3}\mathbb{E}[l(\xi^x)\mathbf{1}_{\{y\notin\xi^x\}}] \text{ sinon.}\end{cases}\end{aligned}$$

et pour $q=p$,

$$\mathbb{E}_x[l(\xi^{x,i})\mathbf{1}_{\{\tau_x^{(p-1)}<\tau_y^{(0)}<\tau_y^{(1)}<\tau_x^{(p)}\}}] = \begin{cases}\mathbb{P}(y\notin\xi^x)^{p-1}\mathbb{E}[l(\xi^x)\mathbf{1}_{\{n_y(\xi^x)\geq 2\}}] \text{ si } i=p \\ \mathbb{P}(n_y(\xi^x)\geq 2)\mathbb{P}(y\notin\xi^x)^{p-2}\mathbb{E}[l(\xi^x)\mathbf{1}_{\{y\notin\xi^x\}}] \text{ sinon.}\end{cases}$$

En sommant sur $i$, on obtient les majorations suivantes :

$$\begin{aligned}&\sum_{i=1}^{p}\mathbb{E}_x[l(\xi^{x,i})\mathbf{1}_{\{\tau_x^{(p-1)}<\tau_y^{(1)}<\tau_x^{(p)}\}}\mathbf{1}_{\{\tau_x^{(q-1)}<\tau_y^{(0)}<\tau_x^{(q)}\}}] \\ &\leq \begin{cases}2\mathbb{P}(y\in\xi^x)\mathbb{P}(y\notin\xi^x)^{p-2}\mathbb{E}[l(\xi^x)\mathbf{1}_{\{y\in\xi^x\}}]+(p-2)\mathbb{P}(y\in\xi^x)^2\mathbb{P}(y\notin\xi^x)^{p-3}\mathbb{E}[l(\xi^x)\mathbf{1}_{\{y\notin\xi^x\}}] \text{ si } q<p \\ (p-1)\mathbb{P}(y\notin\xi^x)^{p-1}\mathbb{E}[l(\xi^x)\mathbf{1}_{\{y\in\xi^x\}}]+\mathbb{P}(y\in\xi^x)\mathbb{P}(y\notin\xi^x)^{p-2}\mathbb{E}[l(\xi^x)\mathbf{1}_{\{y\notin\xi^x\}}] \text{ si } q=p\end{cases}\end{aligned}$$

En utilisant (4.8), et puisque $\mathbb{E}[l(\xi^x)\mathbf{1}_{\{y\in\xi^x\}}] < \infty$ et $\mathbb{E}[l(\xi^x)\mathbf{1}_{\{y\notin\xi^x\}}] < \infty$, on obtient que

$$\mathbb{E}_x[\tau_y^{(1)}] \le \sum_{p=1}^{\infty}\sum_{q=1}^{p}\sum_{i=1}^{p}\mathbb{E}_x[l(\xi^{x,i})\mathbf{1}_{\{\tau_x^{(p-1)}<\tau_y^{(1)}<\tau_x^{(p)}\}}\mathbf{1}_{\{\tau_x^{(q-1)}<\tau_y^{(0)}<\tau_x^{(q)}\}}] < \infty,$$

cette somme pouvant en effet être majorée par :
$\sum_{p=1}^{\infty}\Big\{(p-1)\Big[2\mathbb{P}(y\in\xi^x)\mathbb{P}(y\notin\xi^x)^{p-2}\mathbb{E}_x[l(\xi^x)\mathbf{1}_{\{y\in\xi^x\}}] + (p-2)\mathbb{P}(y\in\xi^x)^2\mathbb{P}(y\notin\xi^x)^{p-3}\mathbb{E}_x[l(\xi^x)\mathbf{1}_{\{y\notin\xi^x\}}]\Big]$
$+ (p-1)\mathbb{P}(y\notin\xi^x)^{p-1}\mathbb{E}_x[l(\xi^x)\mathbf{1}_{\{y\in\xi^x\}}] + \mathbb{P}(y\in\xi^x)\mathbb{P}(y\notin\xi^x)^{p-2}\mathbb{E}_x[l(\xi^x)\mathbf{1}_{\{y\notin\xi^x\}}]\Big\} < \infty.$ □

**Proposition 4.5.2.** *Soit $(X_n, n\in\mathbb{N})$ une chaîne de Markov de matrice de transition $P$ irréductible et récurrente. On pose pour $x\in E$ :*

$$\forall y\in E,\ \nu_x(y) = \mathbb{E}_x\left[\sum_{n=0}^{\tau_x-1}\mathbf{1}_{\{X_n=y\}}\right].$$

*La mesure $\nu_x$ est invariante (i.e. $\nu_x(y) = \sum_{z\in E}\nu_x(z)P(z,y)$) et telle que $\nu_x(x) = 1$. En outre, $\nu_x(E) = \mathbb{E}_x[\tau_x]$, et si $(X_n, n\in\mathbb{N})$ est récurrente positive, $(\frac{\nu_x(y)}{\mathbb{E}_x[\tau_x]})_{y\in E}$ est une probabilité invariante.*

Une autre écriture de $\nu$ est la suivante : $\nu_x(y) = \mathbb{E}_x\left[\sum_{n=0}^{l(\xi^{x,1})-1}\mathbf{1}_{\{\xi_n^{x,1}=y\}}\right] = \mathbb{E}\left[\sum_{n=0}^{l(\xi^x)-1}\mathbf{1}_{\{\xi_n^x=y\}}\right]$. C'est à dire que $\nu(y)$ est le nombre moyen de visites de l'état $y$ lors d'une excursion de la chaîne autour de $x$. Grâce au lemme 4.4.9, on sait que $0<\nu_x(y)<+\infty$ pour tout $y\in E$.

*Preuve.* Tout d'abord, remarquons que par le théorème de Fubini,

$$\nu_x(y) = \mathbb{E}_x\left[\sum_{n=0}^{\infty}\mathbf{1}_{\{n<\tau_x\}}\mathbf{1}_{\{X_n=y\}}\right] = \sum_{n=0}^{\infty}\mathbb{E}_x\left[\mathbf{1}_{\{X_1\neq x\}}\dots\mathbf{1}_{\{X_n\neq x\}}\mathbf{1}_{\{X_n=y\}}\right]. \tag{4.9}$$

On a utilisé ici que $\mathbf{1}_{\{\tau_x>n\}} = \mathbf{1}_{\{X_1\neq x\}}\dots\mathbf{1}_{\{X_n\neq x\}}$. D'autre part sous $\mathbb{P}_x$, puisque $\tau_x<\infty$ (la chaîne est récurrente), on a $X_0 = x = X_{\tau_x}$. Ainsi, $\nu_x(y) = \mathbb{E}_x\left[\sum_{n=1}^{\tau_x}\mathbf{1}_{\{X_n=y\}}\right] = \mathbb{E}_x\left[\sum_{n=1}^{\infty}\mathbf{1}_{\{n\le\tau_x\}}\mathbf{1}_{\{X_n=y\}}\right] = \mathbb{E}_x\left[\sum_{n=0}^{\infty}\mathbf{1}_{\{n+1\le\tau_x\}}\mathbf{1}_{\{X_{n+1}=y\}}\right] = \sum_{n=0}^{\infty}\mathbb{E}_x\left[\mathbf{1}_{\{n<\tau_x\}}\mathbf{1}_{\{X_{n+1}=y\}}\right] = \sum_{n=0}^{\infty}\mathbb{E}_x\left[\mathbf{1}_{\{X_1\neq x\}}\dots\mathbf{1}_{\{X_n\neq x\}}\mathbf{1}_{\{X_{n+1}=y\}}\right]$. En utilisant la proposition 4.1.6,

$$\begin{aligned}
\mathbb{E}_x\left[\mathbf{1}_{\{X_1\neq x\}}\dots\mathbf{1}_{\{X_n\neq x\}}\mathbf{1}_{\{X_{n+1}=y\}}\right] &= \sum_{x_1,\dots,x_n\neq x}P(x,x_1)P(x_1,x_2)\dots P(x_{n-1},x_n)P(x_n,y)\\
&= \sum_{x_n\neq x}\mathbb{E}_x\left[\mathbf{1}_{\{X_1\neq x\}}\dots\mathbf{1}_{\{X_{n-1}\neq x\}}\mathbf{1}_{\{X_n=x_n\}}\right]P(x_n,y)\\
&= \sum_{x_n\in E}\mathbb{E}_x\left[\mathbf{1}_{\{X_1\neq x\}}\dots\mathbf{1}_{\{X_n\neq x\}}\mathbf{1}_{\{X_n=x_n\}}\right]P(x_n,y).
\end{aligned}$$

Par conséquent, on a en utilisant (4.9) puis le théorème de Fubini, il vient

$$\nu_x(y) = \sum_{n=0}^{\infty}\sum_{z\in E}\mathbb{E}_x\left[\mathbf{1}_{\{X_1\neq x\}}\dots\mathbf{1}_{\{X_n\neq x\}}\mathbf{1}_{\{X_n=z\}}\right]P(z,y) = \sum_{z\in E}\nu_x(z)P(z,y).$$

Enfin, $\nu_x(E) = \sum_{y\in E} \nu_x(y) = \mathbb{E}_x\left[\sum_{n=0}^{\tau_x-1} 1\right] = \mathbb{E}_x[\tau_x]$ est fini lorsque la chaîne est récurrente positive, et dans ce cas $\frac{\nu_x(y)}{\nu_x(E)}$ est une probabilité sur $E$ qui est bien invariante. □

Nous somme désormais en mesure d'énoncer le résultat principal de cette section.

**Théorème 4.5.3.** *Soit $(X_n, n \in \mathbb{N})$ une chaîne de Markov de matrice de transition $P$ irréductible et récurrente. Soit $f : E \to \mathbb{R}$ telle que $\nu_x|f| < \infty$. Alors,*

$$\frac{\sum_{k=0}^n f(X_k)}{\sum_{k=0}^n \mathbf{1}_{\{X_k=x\}}} \underset{n\to+\infty}{\longrightarrow} \nu_x f \quad p.s.$$

*Preuve.* En utilisant la décomposition $f = f^+ - f^-$, il suffit de prouver ce résultat pour $f(x) \geq 0$. On note $N_x^n = \sum_{k=0}^n \mathbf{1}_{\{X_k=x\}}$, et on utilise le découpage de la chaîne introduit pour les excursions. On a :

$$\sum_{k=0}^n f(X_k) = \sum_{k=0}^{l(D^x)-1} f(D_k^x) + \sum_{i=1}^{N_x^n-1} \left( \sum_{k=0}^{l(\xi^{x,i})-1} f(\xi_k^{x,i}) \right) + \sum_{k=\tau_x^{(N_x^n)}}^{n} f(X_k).$$

Puisque $f \geq 0$, on a $\sum_{k=\tau_x^{(N_x^n)}}^n f(X_k) \leq \sum_{k=0}^{l(\xi^{x,N_x^n})-1} f(\xi_k^{x,N_x^n})$ et il vient l'encadrement :

$$\frac{1}{N_x^n} \sum_{i=1}^{N_x^n-1} \left( \sum_{k=0}^{l(\xi^{x,i})-1} f(\xi_k^{x,i}) \right) \leq \frac{\sum_{k=0}^n f(X_k)}{N_x^n} \leq \frac{1}{N_x^n} \sum_{i=1}^{N_x^n} \left( \sum_{k=0}^{l(\xi^{x,i})-1} f(\xi_k^{x,i}) \right) + \frac{\sum_{k=0}^{l(D^x)-1} f(D_k^x)}{N_x^n}.$$

Maintenant, on a puisque la chaîne est récurrente $N_x^n \underset{n\to+\infty}{\longrightarrow} N_x = +\infty$ p.s. et $l(D^x) < \infty$ p.s., ce qui implique $\sum_{k=0}^{l(D^x)-1} f(D_k^x) < \infty$ p.s. D'autre part, les excursions étant i.i.d., on a par la loi forte des grands nombres : $\frac{1}{q}\sum_{i=1}^q \left(\sum_{k=0}^{l(\xi^{x,i})-1} f(\xi_k^{x,i})\right) \underset{q\to+\infty}{\longrightarrow} \mathbb{E}[\sum_{k=0}^{l(\xi^x)-1} f(\xi_k^x)] = \mathbb{E}[\sum_{k=0}^{l(\xi^x)-1} \sum_{y\in E} f(y)\mathbf{1}_{\{\xi_k^x=y\}}] = \sum_{y\in E} f(y)\nu_x(y) = \nu_x f$, p.s. Par conséquent, on obtient grâce à l'encadrement : $\frac{\sum_{k=0}^n f(X_k)}{N_x^n} \underset{n\to+\infty}{\longrightarrow} \nu_x f$, p.s. □

On obtient immédiatement le corollaire suivant

**Corollaire 4.5.4.** *Soit $(X_n, n \in \mathbb{N})$ une chaîne de Markov de matrice de transition $P$ irréductible et récurrente. Soient $f, g : E \to \mathbb{R}$ telles que $\nu_x|f| < \infty$, $\nu_x|g| < \infty$ et $\nu_x g \neq 0$. Alors,*

$$\frac{\sum_{k=0}^n f(X_k)}{\sum_{k=0}^n g(X_k)} \underset{n\to+\infty}{\longrightarrow} \frac{\nu_x f}{\nu_x g} \quad p.s.$$

**Proposition 4.5.5.** *Soit $(X_n, n \in \mathbb{N})$ une chaîne de Markov de matrice de transition $P$ irréductible et récurrente.*

– *Si elle est récurrente positive, elle admet une unique probabilité invariante définie par*
$$\forall x \in E, \mu(x) = \frac{1}{\mathbb{E}_x(\tau_x)},$$
*et si $f : E \to \mathbb{R}$ est telle que $\mu|f| < \infty$,*
$$\frac{1}{n}\sum_{k=0}^{n} f(X_k) \underset{n\to+\infty}{\longrightarrow} \mu f \quad p.s.$$

– *Si elle est récurrente nulle, elle n'admet pas de probabilité invariante et si $f : E \to \mathbb{R}$ est t.q. $\nu_x|f| < \infty$,*
$$\frac{1}{n}\sum_{k=0}^{n} f(X_k) \underset{n\to+\infty}{\longrightarrow} 0 \quad p.s.$$

*Preuve.* Lorsque la chaîne est récurrente positive, $\nu_x(E) = \mathbb{E}_x(\tau_x) < \infty$. En prenant $g \equiv 1$, le corollaire 4.5.4 donne $\frac{1}{n}\sum_{k=0}^{n} f(X_k) \underset{n\to+\infty}{\longrightarrow} \frac{\nu_x f}{\mathbb{E}_x(\tau_x)}$, p.s. Si $\mu$ est une probabilité invariante et que $X_0$ a pour loi $\mu$, on a $\mathbb{E}[\mathbf{1}_{\{X_k=y\}}] = \mu(y)$ pour tout $k \geq 0$ et donc $\mathbb{E}[\frac{1}{n}\sum_{k=0}^{n} \mathbf{1}_{\{X_k=y\}}] = \frac{n+1}{n}\mu(y)$. Par ailleurs, $\frac{1}{n}\sum_{k=0}^{n} \mathbf{1}_{\{X_k=y\}}$ converge p.s. vers $\frac{\nu_x(y)}{\mathbb{E}_x(\tau_x)}$ et est dominé par $\frac{n+1}{n} \leq 2$. Ainsi, le théorème de convergence dominée assure que $\mathbb{E}[\frac{1}{n}\sum_{k=0}^{n} \mathbf{1}_{\{X_k=y\}}]$ tend vers $\frac{\nu_x(y)}{\mathbb{E}_x(\tau_x)}$, et par conséquent
$$\forall y \in E, \mu(y) = \frac{\nu_x(y)}{\mathbb{E}_x(\tau_x)}.$$
Il y a donc une unique probabilité invariante $\mu$. Puisque $\nu_x(x) = 1$, on a $\mu(x) = \frac{1}{\mathbb{E}_x(\tau_x)}$. Le choix de $x$ étant arbitraire, on a en fait $\forall x \in E, \mu(x) = \frac{1}{\mathbb{E}_x(\tau_x)}$.

Lorsque la chaîne est récurrente nulle, montrons tout d'abord que $\frac{1}{n}\sum_{k=0}^{n} f(X_k) \underset{n\to+\infty}{\longrightarrow} 0$ $p.s.$ lorsque $\nu_x|f| < \infty$. Quitte à utiliser la décomposition $f = f^+ - f^-$, on peut supposer $f$ positive. Soit $A \subset E$ un ensemble fini. On a clairement $\frac{1}{n+1}\sum_{k=0}^{n} f(X_k) \leq \frac{\sum_{k=0}^{n} f(X_k)}{\sum_{k=0}^{n} \mathbf{1}_A(X_k)}$, et par conséquent le corollaire 4.5.4 assure
$$\limsup_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n} f(X_k) \leq \frac{\nu_x f}{\nu_x(A)} \ p.s.,$$
et ce quelque soit $A$ fini. Puisque $E$ est dénombrable, il existe une suite croissante d'ensembles finis t.q. $\cup_q A_q = E$. Ainsi, $\nu_x(A_q) \underset{q\to+\infty}{\longrightarrow} \nu_x(E) = +\infty$ et on en déduit que $\limsup \frac{1}{n}\sum_{k=0}^{n} f(X_k) \leq 0$. Comme $f$ est positive, $\frac{1}{n}\sum_{k=0}^{n} f(X_k) \underset{n\to+\infty}{\longrightarrow} 0$ $p.s.$ Supposons désormais qu'il existe une probabilité invariante $\mu$ et considérons une chaîne de Markov issue de cette loi invariante. Le même raisonnement que précédemment donne alors que $\mathbb{E}\left[\frac{1}{n}\sum_{k=0}^{n} \mathbf{1}_{\{X_k=y\}}\right] = \frac{n+1}{n}\mu(y)$ tend à la fois vers $\mu(y)$ et 0 quel que soit $y \in E$, et par conséquent $\mu \equiv 0$ ce qui est contradictoire. $\square$

Une conséquence importante de cette proposition et de la proposition 4.4.5 est la suivante.

**Corollaire 4.5.6.** *Une chaîne de Markov irréductible admet une probabilité invariante si et seulement si elle est récurrente positive.*

En fait, de manière plus générale, lorsqu'une chaîne de Markov est récurrente, il existe une unique mesure invariante à une constante multiplicative près. C'est l'objet de l'exercice suivant.

**Exercice 4.5.7.** *On considère $(X_n, n \in \mathbb{N})$ une chaîne de Markov de matrice de transition $P$ irréductible et récurrente. On se donne une mesure invariante $\nu : E \to (0, +\infty)$ et $x_0 \in E$, nous allons montrer que $\forall y \in E, \nu(y) = \nu(x_0)\nu_{x_0}(y)$.*

1. *Montrer qu'il existe $\alpha : E \to (0, +\infty)$ t.q. $\nu_{x_0}(y) = \alpha(y)\nu(y)$.*
2. *On pose pour $x, y \in E$, $\tilde{P}(x,y) = \frac{\nu(y)P(y,x)}{\nu(x)}$. Vérifier que $\tilde{P}$ est bien définie sur $E \times E$ et est une matrice de transition. Montrer que pour $k \geq 1$, $\tilde{P}^k(x,y) = \frac{\nu(y)}{\nu(x)}P^k(y,x)$. En déduire qu'une chaîne de Markov $(\tilde{X}_n, n \in \mathbb{N})$ de matrice de transition $\tilde{P}$ est irréductible.*
3. *On pose $\tilde{N}_x = \sum_{n=0}^{\infty} \mathbf{1}_{\{\tilde{X}_n = x\}}$. Vérifier que $\mathbb{E}_{\tilde{X}_0 = x}[\tilde{N}_x] = \mathbb{E}_{X_0=x}[N_x]$, et en déduire que $(\tilde{X}_n, n \in \mathbb{N})$ est récurrente.*
4. *Vérifier que $\tilde{P}\alpha = \alpha$, puis montrer que sous $\mathbb{P}_{\tilde{X}_0 = x_0}$, $\alpha(\tilde{X}_n)$ est une martingale pour la filtration engendrée par $(\tilde{X}_n, n \in \mathbb{N})$.*
5. *Pour $x \in E$, on pose $\tilde{\tau}_x = \inf\{n \in \mathbb{N}^*, \tilde{X}_n = x\}$. Montrer que $\mathbb{E}_{\tilde{X}_0 = x_0}[\alpha(\tilde{X}_{\tilde{\tau}_x \wedge n})] = \alpha(x_0)$, puis en déduire que $\alpha(x) = \alpha(x_0)$. Conclure.*
6. *Prouver, à l'aide de ce résultat, le théorème 4.5.1 (observer d'abord que la mesure $\nu_x$ est proportionnelle à $\nu_{x_0}$).*

Au cours de cet exercice, nous avons vu que si $\nu$ est une mesure invariante pour la matrice de transition, $\tilde{P}(x,y) = \frac{\nu(y)P(y,x)}{\nu(x)}$ est également une matrice de transition. Si $\alpha$ est une mesure sur $E$, on a

$$\alpha P(x) = \sum_{y \in E} \alpha(y)P(y,x) = \sum_{y \in E} \alpha(y)\frac{\tilde{P}(x,y)\nu(x)}{\nu(y)} = \nu(x)\tilde{P}\frac{\alpha}{\nu}(x),$$

si bien que $\tilde{P}$ permet de transformer un produit mesure-matrice en un produit matrice-fonction. Pour cette raison, on appelle parfois $\tilde{P}$ la matrice duale de $P$.

Pour conclure cette section, nous faisons un tableau récapitulatif des résultats obtenus pour les chaînes de Markov irréductibles. **Lorsque $E$ est fini, une chaîne irréductible est récurrente positive (cf. exercice 4.1.5).**

| | Caractérisation | Proba. invariante | Théorème ergodique |
|---|---|---|---|
| réc. $> 0$ | $\forall x,\ \mathbb{E}_x[\tau_x] < \infty$ | oui, $\mu(x) = \frac{1}{\mathbb{E}_x[\tau_x]}$ | $\frac{1}{n}\sum_{k=0}^n f(X_k) \underset{n\to+\infty}{\longrightarrow} \mu f$ *p.s.* |
| réc. nulle | $\forall x,\ \mathbb{P}_x(\tau_x < \infty) = 1,\ \mathbb{E}_x[\tau_x] = \infty$ | non | $\frac{\sum_{k=0}^n f(X_k)}{\sum_{k=0}^n \mathbf{1}_{\{X_k = x\}}} \underset{n\to+\infty}{\longrightarrow} \nu_x f$ *p.s.* |
| transiente | $\forall x,\ \mathbb{P}_x(\tau_x < \infty) < 1$ | non | non |

TAB. 4.1 – Récapitulatif des propriétés des chaînes de Markov irréductibles.

## 4.6 Problèmes résolus autour de la théorie ergodique

Au cours de la section précédente, nous avons établi un résultat de convergence concernant des moyennes effectuées sur toute la trajectoire d'une chaîne de Markov $(X_n, n \in \mathbb{N})$. Néanmoins, dans certains cas, on voudrait avoir des résultats de convergence sur la loi de $X_n$ plutôt que sur une moyenne trajectorielle. Prenons l'exemple concret du battage de carte (on verra pourquoi on peut assimiler le battage de cartes à une chaîne de Markov). On souhaite en général que le jeu, après l'avoir battu $n$ fois, soit uniformément réparti. Ce qui nous intéresse est donc de connaître comment sont distribuées les cartes après $n$ battages, et le théorème ergodique établi à la section précédente ne nous permet pas de répondre à ce problème. Une autre question est de savoir à partir de combien de battages on peut considérer le jeu uniformément réparti.

Cette section est rédigée sous forme d'une succession de problèmes résolus. Son objectif est de prouver un résultat classique de convergence en loi des chaînes de Markov apériodiques, et de présenter des applications de ce résultat. Ces problèmes s'enchaînent de manière logique, et pour résoudre un problème, il est préférable d'avoir lu les problèmes précédents. Les résultats principaux des trois premiers problèmes sont récapitulés dans des propositions ou théorèmes.

### 4.6.1 Périodicité d'une chaîne de Markov

**Définition 4.6.1.** *Une chaîne de Markov de matrice de transition $P$ est dite périodique de période $d \in \mathbb{N}$ s'il existe une partition $C_1, \ldots, C_d$ (i.e. $\forall i, C_i \neq \emptyset$ et $\forall i, j, C_i \cap C_j = \emptyset$) telle que, en posant $C_0 = C_d$, on ait :*

$$\forall k \in \{1, \ldots, d\}, \forall x \in C_{k-1}, \sum_{y \in C_k} P(x, y) = 1. \tag{4.10}$$

*Une chaîne est dite apériodique lorsque sa plus grande période est* 1.

Autrement dit, lorsqu'une chaîne de Markov est périodique de période $d$, on peut découper l'espace d'états en sous-ensembles distincts $C_1, \ldots, C_d$ tels que la chaîne passe certainement d'un état de $C_k$ à un état de $C_{k+1}$ lorsque $k < d$, et d'un état de $C_d$ à un état de $C_1$. Le problème qui suit est une étude des chaînes de Markov irréductibles et périodiques ou apériodiques.

**Enoncé :** On considère une chaîne de Markov de matrice de transition $P$ irréductible.

1. **Caractérisation de la plus grande période.** Pour $x \in E$, on définit $\mathcal{T}(x) = \{n \in \mathbb{N}^*, P^n(x,x) > 0\}$. Comme la chaîne est irréductible, cet ensemble est non vide et on note $d_x$ le plus grand commun diviseur des entiers de cet ensemble.
   (a) Vérifier que si $m, n \in \mathcal{T}(x)$ et $q \in \mathbb{N}^*$, alors $m + n \in \mathcal{T}(x)$ et $qn \in \mathcal{T}(x)$.
   (b) Montrer que si $x, y \in E$, $d_y/d_x$ (rappelons que cela signifie $d_y$ divise $d_x$, i.e. $\exists \alpha \in \mathbb{N}^*, d_x = \alpha d_y$). En déduire que $d_x = d_y$, et par conséquent $\forall x \in E, d_x = \mathbf{d}$.
   (c) On suppose que la chaîne est périodique de période $d$, montrer que $d/\mathbf{d}$.
   (d) Soit $x_0 \in E$. On définit pour $l \in \{0, \dots, \mathbf{d} - 1\}$,
   $$C_l = \{y \in E, \exists n \in \mathbb{N}, P^{n\mathbf{d}+l}(x_0, y) > 0\},$$
   et $C_d = C_0$. Montrer que $C_1, \dots, C_d$ forme une partition de $E$, puis que la chaîne est périodique de période $\mathbf{d}$.
   (e) Déduire de ce qui précède que $\mathbf{d}$ est la plus grande période de la chaîne de Markov, et que elle est apériodique si et seulement si $\mathbf{d} = 1$.
2. **On suppose désormais la chaîne apériodique.** Soit $x \in E$.
   (a) Montrer qu'il existe $n_1, \dots, n_l \in \mathcal{T}(x)$ et $\alpha_1, \dots, \alpha_l \in \mathbb{Z}$ tels que $\sum_{i=1}^{l} \alpha_i n_i = 1$.
   (b) On pose $N = \sum_{i=1}^{l} |\alpha_i| n_i$. Montrer que $N \in \mathcal{T}(x)$ et $N + 1 \in \mathcal{T}(x)$.
   (c) Montrer que pour tout $n \geq N^2, n \in \mathcal{T}(x)$.
   (d) En déduire que : $\forall x, y \in E,\ \exists n_{xy} \in \mathbb{N}^*,\ \forall n \geq n_{xy}, P^n(x, y) > 0$.

Nous résumons les résultats principaux de cet exercice dans la proposition suivante.

**Proposition 4.6.2.** *Pour une chaîne de Markov de matrice de transition $P$ irréductible, le pgcd de $\mathcal{T}(x) = \{n \in \mathbb{N}^*, P^n(x,x) > 0\}$ ne dépend pas de $x$, et c'est la plus grande période de la chaîne. Lorsque la chaîne est apériodique, $\forall x, y \in E,\ \exists n_{xy},\ \forall n \geq n_{xy}, P^n(x, y) > 0$*

**Solution :**

1. (a) Il suffit de remarquer que $P^{m+n}(x,x) \geq P^m(x,x)P^n(x,x)$ pour avoir $m + n \in \mathcal{T}(x)$, et une récurrence immédiate donne $qn \in \mathcal{T}(x)$.
   (b) Soit $n \in \mathcal{T}(x)$. Puisque la chaîne est irréductible, on a par le lemme 4.4.3 qu'il existe $k, l \in \mathbb{N}^*$ tels que $P^k(y,x) > 0$ et $P^l(x,y) > 0$. Par conséquent, $P^{k+l}(y,y) \geq P^k(y,x)P^l(x,y) > 0$ et donc $k + l \in \mathcal{T}(y)$. De même, $P^{k+n+l}(y,y) \geq P^k(y,x)P^n(x,x)P^l(x,y) > 0$ et $k + l + n \in \mathcal{T}(y)$. Comme $d_y$ divise $k + l$ et $k + l + n$, on a $d_y/n$. L'entier $n$ étant arbitrairement choisi dans $\mathcal{T}(x)$, il vient $d_y/d_x$.
   (c) Si $d = 1$ il n'y a rien à vérifier. Sinon, il existe une partition $A_1, \dots, A_d$ telle que (en posant $A_0 = A_d$), $\forall k \in \{1, \dots, d\}, \forall x \in A_{k-1}, \forall y \notin A_k, P(x,y) = 0$. Soit $x \in A_0$. On note $\mathbf{i}(n)$ le reste de la division euclidienne de $n$ par $d$.
   $$P^n(x,x) = \sum_{x_1,\dots,x_{n-1} \in E} P(x,x_1)\dots P(x_{n-1},x) = \sum_{x_1 \in A_{\mathbf{i}(1)},\dots,x_{n-1} \in A_{\mathbf{i}(n-1)}} P(x,x_1)\dots P(x_{n-1},x)$$
   est nul si $\mathbf{i}(n-1) \neq d-1$, i.e. si $\mathbf{i}(n) \neq 0$. Par conséquent, $d/n$ si $n \in \mathcal{T}(x)$ et $d/d_x = \mathbf{d}$.

(d) Le fait que $\cup_l C_l = E$ provient directement du fait que la chaîne est irréductible. Montrons que l'on a bien une partition. On raisonne par l'absurde et on suppose qu'il existe $1 \le l < l'$ et $y \in C_l \cap C_{l'}$. Alors, par définition il existe $n, n' \in \mathbb{N}$, $P^{n\mathbf{d}+l}(x_0, y) > 0$ et $P^{n'\mathbf{d}+l'}(x_0, y) > 0$. Puisque la chaîne est irréductible, il existe $k \in \mathbb{N}^*$ tel que $P^k(y, x_0) > 0$. On a alors que $n\mathbf{d} + l + k \in \mathcal{T}(x_0)$ et $n'\mathbf{d} + l' + k \in \mathcal{T}(x_0)$, et donc $d_{x_0} / (n' - n)\mathbf{d} + (l' - l)$. Comme $d_{x_0} = \mathbf{d}$, il vient que $\mathbf{d}/l' - l$ ce qui est impossible puisque $0 < l' - l < \mathbf{d}$.

Enfin montrons que la chaîne est périodique : il s'agit de montrer que si $x \in C_l$ avec $l \in \{0, \ldots, \mathbf{d} - 1\}$ et si $P(x, y) > 0$, alors $y \in C_{l+1}$. Si $x \in C_l$, il existe $n$ t.q. $P^{n\mathbf{d}+l}(x_0, x) > 0$. Par conséquent, si $P(x, y) > 0$, $P^{n\mathbf{d}+l+1}(x_0, y) > 0$ et donc $y \in C_{l+1}$.

(e) La chaîne est périodique de période $\mathbf{d}$, et toute autre période divise $\mathbf{d}$. La période maximale de la chaîne est donc $\mathbf{d}$, et par définition la chaîne est apériodique ssi $\mathbf{d} = 1$.

2. (a) Puisque $d_x = 1$, il existe $n_1, \ldots, n_l \in \mathcal{T}(x)$ tels que leur pgcd est 1 et l'existence de $\alpha_1, \ldots, \alpha_l \in \mathbb{Z}$ est assurée par le Théorème de Bézout.

(b) En appliquant itérativement la question 1.(a), on obtient $N \in \mathcal{T}(x)$. De même, en écrivant $N+1 = \sum_{i=1}^{l}(|\alpha_i| + \alpha_i)n_i$ on obtient que $N + 1 \in \mathcal{T}(x)$ car $|\alpha_i| + \alpha_i \in \mathbb{N}$.

(c) Pour $0 \le k \le N - 1$, on écrit $N^2 + k = (N - k)N + k(N + 1)$ et grâce à la question précédente et la question 1.(a), il vient $N^2 + k \in \mathcal{T}(x)$. Ensuite, tout entier $n \ge N^2$ peut s'écrire $n = N^2 + k + qN$ avec $q \in \mathbb{N}$ et $0 \le k \le N - 1$. Par conséquent grâce à la question 1.(a), $n \in \mathcal{T}(x)$.

(d) Soit $y \in E$. La chaîne étant irréductible, il existe $k \in \mathbb{N}^*$ t.q. $P^k(x, y) > 0$. Pour tout $n \ge N^2 + k$, $P^n(x, y) \ge P^{n-k}(x, x)P^k(x, y) > 0$.

### 4.6.2 Couplage de chaînes de Markov

Le problème qui suit est une introduction au couplage, dont nous verrons une belle application au problème suivant pour prouver un théorème ergodique. La technique de couplage repose comme on le verra sur la propriété de Markov forte, et par conséquent peut s'appliquer à des processus Markoviens plus généraux que les chaînes de Markov.

**Enoncé :**

1. **Chaîne de Markov produit.** On considère deux espaces d'états dénombrables $E$ et $E'$. Soient $P$ et $P'$ deux matrices de transitions respéctivement sur $E$ et sur $E'$. On définit pour $(x, x'), (y, y') \in E \times E', Q((x, x'), (y, y')) = P(x, y)P'(x', y')$. Vérifier que $Q$ est une matrice de transition sur $E \times E'$. On appelle $Q$ matrice produit et une chaîne de Markov $(X_n, X'_n)$ associée une chaîne produit.
2. **Le couplage d'une chaîne de Markov.** Le couplage consiste à considérer le produit d'une chaîne avec elle-même, c'est à dire que l'on prend $E' = E$ et $P' = P$. Puis on introduit $\tau = \inf\{n \in \mathbb{N}, X_n = X'_n\}$.
   (a) Vérifier que $\tau$ est un temps d'arrêt par rapport à la filtration engendrée par $(X_n, X'_n)_{n \in \mathbb{N}}$.
   (b) Montrer que conditionnellement à $\{\tau < +\infty\}$, $(X_{\tau+n}, X'_{\tau+n})_{n\in\mathbb{N}}$ et $(X'_{\tau+n}, X_{\tau+n})_{n\in\mathbb{N}}$ sont des chaînes de Markov qui ont même loi.

On résume le résultat pricipal de cet exercice.

**Proposition 4.6.3.** *Soient $E$ un espace dénombrable, $P$ une matrice de transition sur $E$ et $(X_n, X'_n)$ une chaîne de Markov sur $E \times E$ construite par couplage, c'est à dire de matrice de transition $Q((x, x'), (y, y')) = P(x, y)P'(x', y')$. On pose $\tau = \inf\{n \in \mathbb{N}, X_n = X'_n\}$. Alors,* *conditionnellement à* $\{\tau < +\infty\}$, $(X_{\tau+n}, X'_{\tau+n})_{n\in\mathbb{N}}$ *et* $(X'_{\tau+n}, X_{\tau+n})_{n\in\mathbb{N}}$ *sont des chaînes de Markov qui ont même loi.*

**Solution :**

1. L'espace $E\times E'$ est dénombrable comme produit d'espaces dénombrables. Soit $(x,x') \in E\times E'$, on a $Q((x,x'),(y,y')) = P(x,y)P'(x',y') > 0$ pour $(y,y') \in E\times E'$ et $\sum_{y\in E, y'\in E'} P(x,y)P'(x',y') = \left(\sum_{y\in E} P(x,y)\right)\left(\sum_{y'\in E'} P'(x',y')\right) = 1$.
2. (a) $\{\tau \leq n\} = \cup_{i\leq n}\{X_i = X'_i\}$ est mesurable par rapport à la tribu $\sigma((X_i, X'_i), i \leq n)$.
   (b) Observons que les chaînes $(X_n, X'_n)_{n\in\mathbb{N}}$ et $(X'_n, X_n)_{n\in\mathbb{N}}$ ont toutes deux la même matrice de transition $Q$. D'après la propriété de Markov forte appliquée à $(X_n, X'_n)_{n\in\mathbb{N}}$, conditionnellement à $\{\tau < +\infty\}$, $(X_{\tau+n}, X'_{\tau+n})_{n\in\mathbb{N}}$ est une chaîne de Markov : sa matrice de transition est $Q$ et sa loi initiale est la loi de $(X_\tau, X'_\tau)$ sachant $\{\tau < +\infty\}$. De même en appliquant la propriété de Markov à la chaîne $(X'_n, X_n)_{n\in\mathbb{N}}$, on obtient que conditionnellement à $\{\tau < +\infty\}$, $(X'_{\tau+n}, X_{\tau+n})_{n\in\mathbb{N}}$ est une chaîne de Markov de matrice de transition est $Q$ et de loi initiale $(X'_\tau, X_\tau)$. Comme sur $\{\tau < +\infty\}$, $X_\tau = X'_\tau$, on a que conditionnellement à $\{\tau < +\infty\}$, $(X_\tau, X'_\tau)$ et $(X'_\tau, X_\tau)$ ont même loi (elle sont même égale p.s.). On en conclut que les chaînes $(X_{\tau+n}, X'_{\tau+n})_{n\in\mathbb{N}}$ et $(X'_{\tau+n}, X_{\tau+n})_{n\in\mathbb{N}}$ ont la même loi.

### 4.6.3 Un second théorème ergodique

**Enoncé :** On considère une chaîne de Markov $(X_n, n \in \mathbb{N})$ irréductible récurrente positive et apériodique. On note $\mu$ sa probabilité invariante (cf. Proposition 4.5.5).

1. Montrer qu'une chaîne $(X_n, X'_n)_{n\in\mathbb{N}}$ construite par couplage est irréductible récurrente positive. (On pourra utiliser la proposition 4.6.2.)
2. En déduire que $\tau = \inf\{n \in \mathbb{N}, X_n = X'_n\}$ est fini p.s., puis que les chaînes de Markov $(X_{\tau+n}, X'_{\tau+n})_{n\in\mathbb{N}}$ et $(X'_{\tau+n}, X_{\tau+n})_{n\in\mathbb{N}}$ ont même loi.
3. Soient $k \leq n$, $y \in E$. Montrer que $\mathbb{P}(X_n = y, \tau = k) = \mathbb{P}(X'_n = y, \tau = k)$, puis que $\mathbb{P}(X_n = y, \tau \leq n) = \mathbb{P}(X'_n = y, \tau \leq n)$. En déduire alors que
$$|\mathbb{P}(X_n = y) - \mathbb{P}(X'_n = y)| \leq \mathbb{P}(\tau > n).$$
4. En choisissant convenablement la loi de $X'_0$, en déduire que pour tout $y \in E$, $\mathbb{P}(X_n = y) \underset{n\to+\infty}{\longrightarrow} \mu(y)$.

Ce problème montre le résultat important suivant.

**Théorème 4.6.4.** *Soit* $(X_n, n \in \mathbb{N})$ *une chaîne de Markov irréductible, récurrente positive et apériodique. On note* $\mu$ *sa probabilité invariante. Alors,* $X_n$ *converge en loi vers une v.a. de loi* $\mu$.

**Solution :**

1. Tout d'abord, observons que $Q^n((x,x'),(y,y')) = P^n(x,y)P^n(x',y')$. Grâce à la proposition 4.6.2, on obtient que cette quantité est strictement positive pour $n$ assez grand, et la chaîne est donc irréductible. Pour prouver qu'elle est récurrente positive il suffit de prouver qu'elle a une probabilité invariante d'après le corollaire 4.5.6. Il facile de voir que $\mu(x)\mu(x')$ est une probabilité sur $E \times E$ invariante pour $Q$.
2. Soit $x \in E$. Puisque $(X_n, X'_n)_{n\in\mathbb{N}}$ est récurrente positive, $\tau^{(0)}_{(x,x)} = \inf\{n \in \mathbb{N}, X_n = X'_n = x\}$ est fini p.s., et comme $\tau \leq \tau^{(0)}_{(x,x)}$, $\tau$ est également fini p.s. Grâce à la question 2b du problème précédent, les chaînes de Markov $(X_{\tau+n}, X'_{\tau+n})$ et $(X'_{\tau+n}, X_{\tau+n})$ ont même loi.

3. Grâce à la question précédente, $\mathbb{P}(X_n = y, \tau = k) = \mathbb{P}(X_{\tau+n-k} = y, \tau = k) = \mathbb{P}(X'_{\tau+n-k} = y, \tau = k) = \mathbb{P}(X'_n, \tau = k)$, et en sommant sur $k$, $\mathbb{P}(X_n = y, \tau \leq n) = \mathbb{P}(X'_n = y, \tau \leq n)$. Ensuite, on écrit $\mathbb{P}(X_n = y) = \mathbb{P}(X_n = y, \tau \leq n) + \mathbb{P}(X_n = y, \tau > n) \leq \mathbb{P}(X'_n = y, \tau \leq n) + \mathbb{P}(\tau > n) \leq \mathbb{P}(X'_n = y) + \mathbb{P}(\tau > n)$. De la même manière, on obtient $\mathbb{P}(X'_n = y) \leq \mathbb{P}(X_n = y) + \mathbb{P}(\tau > n)$ et donc $|\mathbb{P}(X_n = y) - \mathbb{P}(X'_n = y)| \leq \mathbb{P}(\tau > n)$.
4. On choisit $\mu$ pour loi initiale de $X'_0$ si bien que pour tout $n$, $\mathbb{P}(X'_n = y) = \mu(y)$. Comme $\tau$ est fini p.s, $\mathbb{P}(\tau > n) \underset{n\to+\infty}{\to} 0$ et donc $|\mathbb{P}(X_n = y) - \mu(y)| \underset{n\to+\infty}{\to} 0$.

**Exercice 4.6.5.** *On considère un jeu de carte composé de $N \geq 2$ cartes que l'on souhaite battre afin qu'il soit réparti uniformément. On numérote arbitrairement ces cartes de 1 à $N$, et une configuration du jeu sera modélisée par une permutation $\sigma \in \mathfrak{S}_N$ : $\sigma(1)$ est la première carte du dessus, $\sigma(2)$ la seconde, etc. On propose le mode opératoire suivant pour battre les cartes : on prend la première carte que l'on remet dans le paquet de manière aléatoire et uniforme entre la deuxième et la dernière place. Pour $\sigma, \tilde{\sigma} \in \mathfrak{S}_N$, on note $P(\sigma, \tilde{\sigma})$ la probabilité d'obtenir après un battage la configuration $\tilde{\sigma}$ à partir de la configuration $\sigma$.*

1. *Donner la valeur de $P(\sigma, \tilde{\sigma})$ et vérifier qu'il s'agit d'une matrice de transition.*
2. *Montrer qu'une chaîne de Markov $(\sigma_n, n \in \mathbb{N})$ de matrice de transition $P$ est irréductible, et apériodique dès que $N \geq 3$ (utiliser la proposition 4.6.2).*
3. *Vérifier que $\mu \equiv \frac{1}{N!}$ est une probabilité invariante. En déduire que $\sigma_n$ converge en loi vers une v.a. uniforme sur $\mathfrak{S}_N$.*

**Exercice 4.6.6.** *Donner un exemple de chaîne de Markov irréductible récurrente telle que la chaîne construite par couplage n'est pas irréductible.*

### 4.6.4 L'algorithme d'Hastings-Metropolis et recuit simulé

On se place sur un espace dénombrable $E$, et on se donne une mesure de probabilité $\mu$ sur $E$. On souhaite simuler une v.a. de loi $\mu$ sur $E$. Commençons par remarquer que si l'on connaît $\mu$ parfaitement, il est facile de simuler une telle loi. En effet, $E$ étant dénombrable, on peut numéroter les états : $E = \{x_0, x_1, \ldots\}$ et on note $p_i = \mu(x_i)$. Alors, si on tire une loi uniforme $U$ sur $[0, 1]$, et que l'on choisit $x_i$ si $\sum_{j=0}^{i-1} p_j \leq U < \sum_{j=0}^{i} p_j$, on simule bien une v.a. de loi $\mu$ sur $E$.

L'algorithme d'Hastings-Metropolis permet de simuler $\mu$ lorsqu'on ne la connaît qu'à une constante multiplicative près. Cette hypothèse peut sembler surprenante d'un premier abord, mais est fréquente, notamment en physique statistique. En effet, souvent dans ce cas $E$ représente les différentes configurations possibles d'un système et l'on sait que la probabilité de chaque configuration est $\mu(x) = c \exp(-\beta H(x))$ où $H(x) \geq 0$ décrit l'énergie du système dans la configuration $x$. Le paramètre $\beta$ et la fonction $H(x)$ sont en général bien connus. La constante $c$ vaut alors bien évidemment $1/(\sum_{x\in E} \exp(-\beta H(x)))$, mais comme le nombre de configurations possibles est en général "très grand", ce calcul est trop coûteux. L'idée de l'algorithme d'Hastings-Metropolis est de construire une chaîne de Markov $(X_n, n \in \mathbb{N})$ apériodique irréductible de probabilité invariante $\mu$. Ainsi pour $n$ grand, grâce au théorème ergodique 4.6.4, $X_n$ suivra une loi proche de $\mu$. Une application importante de cet algorithme, on le verra, est de fournir un moyen d'approcher le minimum de la fonction $H$. Dans ce cas, on parle de recuit simulé.

**Enoncé :** Soit $\mu$ une probabilité sur $E$ telle que $\forall x \in E, \mu(x) > 0$. On se donne une matrice de transition $P$ irréductible et apériodique et telle que $P(x,y) > 0 \implies P(y,x) > 0$.

1. On considère une famille de réels $(a(x,y))_{x,y\in E}$ telle que $0 < a(x,y) \leq 1$. On pose $Q(x,y) = a(x,y)P(x,y)$ si $x \neq y$ et $Q(x,x) = 1 - \sum_{y\neq x} Q(x,y)$.
   (a) Vérifier que $Q$ est une matrice de transition sur $E$. Montrer qu'elle est irréductible et apériodique.
   (b) Montrer que si $\mu(x)Q(x,y) = \mu(y)Q(y,x)$, alors $\mu$ est une probabilité invariante de $Q$. Donner un choix de $(a(x,y))_{x,y\in E} \in (0,1]^{E\times E}$ qui permet de satisfaire cette condition.
   (c) On prend $a(x,y) = \min(1, \frac{\mu(y)P(y,x)}{\mu(x)P(x,y)})$ si $P(x,y) > 0$ et $a(x,y) = 1$ sinon. Ce choix ne nécessite en effet de ne connaître $\mu$ qu'à une constante multiplicative près. Montrer que $\mu$ est une probabilité invariante de $Q$ et qu'une chaîne de Markov $(X_n, n \in \mathbb{N})$ de matrice de transition $Q$ est récurrente positive. Donner le comportement asymptotique de $X_n$.
2. **Algorithme d'Hastings-Metropolis.** On souhaite désormais être capable de simuler une chaîne de Markov de matrice de transition $Q$. On se donne une famille de v.a. $(Y_{n,x}, n \in \mathbb{N}, x \in E)$ à valeurs dans $E$, indépendantes et de loi $\mathbb{P}(Y_{n,x} = y) = P(x,y)$, et une suite de v.a. indépendantes $(U_n, n \in \mathbb{N})$ de loi uniforme sur $[0,1]$.
   (a) Comment simuler $Y_{n,x}$ à partir d'une loi uniforme sur $[0,1]$ ?
   (b) Montrer que $X_{n+1} = Y_{n,X_n}\mathbf{1}_{\{U_n \leq a(X_n, Y_{n,X_n})\}} + X_n\mathbf{1}_{\{U_n > a(X_n, Y_{n,X_n})\}}$ définit une chaîne de Markov de matrice de transition $Q$.
3. **Recuit simulé.** On considère désormais une probabilité donnée par une fonction $H : E \to \mathbb{R}^+$ :
   $$\mu_\beta(x) = c_\beta \exp(-\beta H(x)).$$
   On définit $H_{\min} = \min\{H(x), x \in E\}$ et $\mathbf{H} = \{x \in E, H(x) = H_{\min}\}$.
   (a) Montrer que si $H(x) > H_{\min}$, $\mu_\beta(x) \underset{\beta\to+\infty}{\to} 0$.
   (b) On suppose $E$ fini, montrer que $\mathbf{H} \neq \emptyset$, et que si $x \in \mathbf{H}$, $\mu_\beta(x) \underset{\beta\to+\infty}{\to} 1/(\#\mathbf{H})$. Que se passe-t-il alors si l'on utilise l'algorithme d'Hastings-Metropolis pour un $\beta$ "grand" ?

**Solution :**

1. (a) Comme $a(x,y) > 0$, il est facile de voir que $P(x,y) > 0 \implies Q(x,y) > 0$, puis que $P^n(x,y) > 0 \implies Q^n(x,y) > 0$. Par conséquent, $Q$ est irréductible. En outre, comme $P$ est apériodique, grâce à la proposition 4.6.2 il existe $n_{xx} \in \mathbb{N}$ t.q. $\{n \geq n_{xx}\} \subset \{n \in \mathbb{N}, P^n(x,x) > 0\} \subset \{n \in \mathbb{N}, Q^n(x,x) > 0\}$. Par conséquent, le pgcd de $\{n \in \mathbb{N}, Q^n(x,x) > 0\}$ est 1 et $Q$ est apériodique.

   (b) Si $\mu(x)Q(x,y) = \mu(y)Q(y,x)$, $\sum_{y\in E}\mu(x)Q(x,y) = \mu(y)\sum_{y\in E}Q(y,x) = \mu(y)$. Cette condition s'écrit $\mu(x)a(x,y)P(x,y) = \mu(y)a(y,x)P(y,x)$. On pourrait prendre $a(x,y) = \mu(y)P(y,x) + \mathbf{1}_{\{P(x,y)=0\}}$ pour satisfaire cette condition. Mais on prend $a(x,y) = \min(1, \frac{\mu(y)P(y,x)}{\mu(x)P(x,y)})$ si $P(x,y) \neq 0$ et $a(x,y) = 1$ sinon, car ce choix ne nécessite en effet de ne connaître $\mu$ qu'à une constante multiplicative près.

   (c) D'après la question précédente, $\mu$ est une probabilité invariante de $Q$ et grâce au corollaire 4.5.6, $(X_n, n \in \mathbb{N})$ est récurrente positive. Comme elle est également apériodique, le théorème 4.6.4 assure que $X_n$ converge en loi vers une v.a. de loi $\mu$.

2. (a) $Y_{n,x}$ suit la loi discrète connue $P(x,y)$ sur $E$, et il a été rappelé au début de la section 4.6.4 comment simuler une telle v.a.

   (b) En utilisant l'indépendance des différentes variables, on a si $y \neq x$ $\mathbb{P}(X_{n+1} = y|X_n = x) = \mathbb{P}(Y_{n,x} = y, U_n \leq a(x, Y_{n,x})) = \mathbb{P}(Y_{n,x} = y, U_n \leq a(x,y)) = \mathbb{P}(Y_{n,x} = y)\mathbb{P}(U_n \leq a(x,y)) = P(x,y)a(x,y) = Q(x,y)$ et donc $\mathbb{P}(X_{n+1} = x|X_n = x) = 1 - \sum_{y\neq x}\mathbb{P}(X_{n+1} = y|X_n = x) = Q(x,x)$.

3. (a) Si $H(x) > H_{\min}$, il existe $x' \in E$ t.q. $H(x') < H(x)$. Or, la constante de normalisation satisfait $c_\beta = 1/(\sum_{x\in E}\exp(-\beta H(x))) \leq \exp(\beta H(x'))$ et on a donc $\mu_\beta(x) \leq \exp(-\beta[H(x) - H(x')]) \underset{\beta\to+\infty}{\to} 0$.

   (b) Comme $E$ fini, $H$ atteint son minimum. On a donc :

$$\mu_\beta(x) = \frac{1}{\#\mathbf{H} + \sum_{x\notin\mathbf{H}}\exp(-\beta[H(x) - H_{\min}])} \underset{\beta\to+\infty}{\to} \frac{1}{\#\mathbf{H}}.$$

   En prenant $\beta$ "grand", la mesure de probabilité $\mu_\beta$ se concentre sur les éléments de $E$ qui réalisent le minimum de $H$. Comme $X_n$ converge en loi vers une v.a. de loi $\mu_\beta$, pour $n$ grand, $H(X_n)$ sera proche du minimum $H_{\min}$. L'algorithme d'Hastings-Metropolis permet donc ici d'approcher un minimiseur de la fonction $H$. On parle alors de recuit simulé.

Une application intéressante du recuit simulé est donnée dans l'exercice suivant.

**Exercice 4.6.7.** *(Problème du voyageur de commerce.) Un voyageur de commerce doit passer par $N$ villes (numérotées de 1 à $N$), et souhaite connaître le plus court trajet pour effectuer son périple. On modélise un parcours par une permutation $\sigma \in \mathfrak{S}_N$ : le voyageur part de la ville $\sigma(1)$, va à $\sigma(2)$, et ainsi de suite jusqu'à $\sigma(N)$ et il retourne alors à la ville initiale. La distance à minimiser est donc*

$$\sigma \in \mathfrak{S}_N,\ H(\sigma) = d(\sigma(1),\sigma(2)) + \cdots + d(\sigma(N-1),\sigma(N)) + d(\sigma(N),\sigma(1)).$$

1. *Donner une estimation de 20!. En déduire que minimiser $H$ en examinant tous les parcours possibles demande un temps excessif dès que $N \geq 20$.*
2. *Proposer une matrice de transition $P$ sur $\mathfrak{S}_N$ qui soit irréductible et apériodique. (On pourra s'inspirer de l'exercice 4.6.5.)*
3. *Construire alors une chaîne de Markov $(\sigma_n, n \in \mathbb{N})$ à valeurs dans $\mathfrak{S}_N$ irréductible apériodique et de loi invariante $\mu_\beta(\sigma) = c_\beta \exp(-\beta H(\sigma))$. Si l'on choisit $\beta$ "grand", expliquer pourquoi $\sigma_n$ est proche d'un minimum de $H$ lorsque $n$ est grand.*

### 4.6.5 Vitesse de convergence vers la loi invariante

Dans cette section, nous nous posons la question de savoir à quelle vitesse une chaîne de Markov $(X_n, n \in N)$ irréductible, apériodique, récurrente positive converge-t-elle vers la probabilité invariante $\mu$. Autrement dit, à partir de quelle valeurs de $n$ peut on considérer la loi de $X_n$ "proche" de $\mu$ ? C'est une question qui est importante et qui intervient naturellement lorsque l'on souhaite mettre en oeuvre l'algorithme de Hastings-Metropolis. Pour donner une prémière réponse à ce problème, nous allons introduire la norme en variation d'une application $m : E \to \mathbb{R}$ par

$$\mathbf{V}(m) = \sum_{x \in E} |m(x)|, \tag{4.11}$$

et étudier la convergence de la loi de $X_n$, $\mathbb{P}(X_n = x)$, vers $\mu(x)$ pour cette norme.

**Enoncé :** Soit $E$ un espace d'états dénombrable.

1. **Etude de la norme en variation.**
   (a) Vérifier que $\mathcal{V}(E) = \{m : E \to \mathbb{R}, \mathbf{V}(m) < \infty\}$ est un $\mathbb{R}$-espace vectoriel contenant toutes les probabilités sur $E$, et sur lequel $\mathbf{V}$ est une norme. Montrer que l'espace $(\mathcal{V}(E), \mathbf{V})$ est complet.
   (b) Soit $(X_n)_n$ une suite de v.a. à valeurs dans $E$. On pose $p_n(x) = \mathbb{P}(X_n = x)$ pour $x \in E$. Montrer que si $\mu$ est une probabilité sur $E$ t.q. $\mathbf{V}(p_n - \mu) \underset{n\to\infty}{\to} 0$, alors $X_n$ converge en loi vers une v.a. de loi $\mu$ sur $E$.
2. **La condition de Doeblin.**[1] On considère désormais une chaîne de Markov $(X_n, n \in \mathbb{N})$ de matrice de transition $P$, qui satisfait la condition de Doeblin :

   il existe une probabilité $\pi$ sur $E,\ l \in \mathbb{N}^*,\ \alpha > 0,$ t.q. $\forall x, y \in E,\ P^l(x, y) \geq \alpha\pi(y)$.

   (a) Montrer que si $\mu$ et $\mu'$ sont des probabilités sur $E$,
   $$\mathbf{V}(\mu P^l - \mu' P^l) \leq (1 - \alpha)\mathbf{V}(\mu - \mu').$$
   (b) En déduire qu'il existe une unique probabilité invariante $\mu$ pour $P^l$, puis montrer que $\mu$ est également l'unique probabilité invariante pour $P$. Que dire de la chaîne si elle est irréductible ?
   (c) Montrer que $\mathbf{V}(\mu - p_n) \leq (1 - \alpha)^{\lfloor n/l \rfloor} \max_{j=0,\ldots,l-1} \mathbf{V}(\mu - p_j)$. En déduire que $X_n$ converge en loi vers une v.a. de loi $\mu$, et décrire qualitativement cette convergence.

---

[1]Une recherche "Wolfgang Doeblin" ou "Vincent Doeblin" sur votre moteur de recherche favori vous permettra de découvrir la vie brève, riche et mouvementée de ce mathématicien.

**Solution :**

1. **Etude de la norme en variation.**
   (a) Grâce à l'inégalité triangulaire, l'espace $\mathcal{V}(E)$ est un $\mathbb{R}$-espace vectoriel et $\mathbf{V}$ une norme sur $\mathcal{V}(E)$. Il contient toute probabilité $\mu$ sur $E$ puisque $\sum_{x\in E}|\mu(x)| = 1$. Maintenant, considérons une suite d'applications $m_n : E \to \mathbb{R}$ qui satisfait le critère de Cauchy. Soit $\varepsilon > 0$. Il existe $N > 0$ t.q. $\forall p, q \geq N, \mathbf{V}(m_p - m_q) \leq \varepsilon$. Soit $x \in E$. On a $|m_p(x) - m_q(x)| \leq \mathbf{V}(m_p - m_q)$, et par conséquent $(m_n(x))_n$ est une suite de Cauchy réelle. On note $m(x)$ sa limite. Grâce au lemme de Fatou par rapport à la mesure de comptage, $\mathbf{V}(m) \leq \liminf_{n\to+\infty} \mathbf{V}(m_n) < \infty$. Toujours par le lemme de Fatou, on a pour $p \geq N$,

$$\mathbf{V}(m_p - m) \leq \liminf_{q\to+\infty} \mathbf{V}(m_p - m_q) \leq \varepsilon.$$

   (b) Pour une fonction $f$ bornée ($|f| \leq K$), on a $|\sum_{x\in E} f(x)p_n(x) - \sum_{x\in E} f(x)\mu(x)| = |\sum_{x\in E} f(x)(p_n(x) - \mu(x))| \leq \sum_{x\in E} |f(x)||p_n(x) - \mu(x)| \leq K\mathbf{V}(p_n - \mu) \underset{n\to\infty}{\to} 0$. Ainsi, $\mathbb{E}[f(X_n)] \underset{n\to\infty}{\to} \sum_{x\in E} f(x)\mu(x)$.

2. (a) Soient $\mu$, $\mu'$ deux probabilités sur $E$. On a :
   $\mu P^l(y) - \mu' P^l(y) = \sum_{x\in E}[\mu(x) - \mu'(x)]P^l(x,y) = \sum_{x\in E}[\mu(x) - \mu'(x)][P^l(x,y) - \alpha\pi(y)]$ (remarquer que $\sum_{x\in E}[\mu(x) - \mu'(x)]\pi(y) = 0$). Par conséquent, $|\mu P^l(y) - \mu' P^l(y)| \leq \sum_{x\in E} |\mu(x) - \mu'(x)|[P^l(x,y) - \alpha\pi(y)]$, et en sommant sur $y \in E$, il vient $\mathbf{V}(\mu P^l - \mu' P^l) \leq (1-\alpha)\mathbf{V}(\mu - \mu')$.

   (b) Si $\mu$ et $\mu'$ sont invariantes pour $P^l$, on a $\mathbf{V}(\mu-\mu') = \mathbf{V}(\mu P^l - \mu' P^l) \leq (1-\alpha)\mathbf{V}(\mu-\mu')$, et donc $\mathbf{V}(\mu-\mu') = 0$ i.e. $\mu = \mu'$. Il reste à montrer l'existence d'une probabilité invariante pour $P^l$. Soit $\mu'$ une probabilité, on a $\mathbf{V}(\mu' P^{l(n_1+n_2)} - \mu' P^{ln_1}) \leq (1-\alpha)^{n_1}\mathbf{V}(\mu' P^{ln_2} - \mu') \leq 2(1-\alpha)^{n_1}$. Par conséquent, $(\mu' P^{ln})_n$ satisfait le critère de Cauchy et converge dans $\mathcal{V}(E)$. On note $\mu$ cette limite. Comme $\mu' P^{ln} \geq 0$ et $\mathbf{V}(\mu' P^{ln}) = 1$, il vient que $\mu$ est une probabilité sur $E$. Enfin, puisque $\mathbf{V}(\mu' P^{l(n+1)} - \mu' P^{ln}) \leq (1-\alpha)^n \mathbf{V}(\mu' P^l - \mu') \underset{n\to+\infty}{\to} 0$, on obtient $\mathbf{V}(\mu P^l - \mu) = 0$ et $\mu$ est invariante pour $P^l$.

   Il est clair qu'une probabilité invariante pour $P$ est invariante pour $P^l$, est $\mu$ est donc la seule probabilité invariante possible pour $P$. On a $\mathbf{V}(\mu' P^{ln+1} - \mu' P^{ln}) = \mathbf{V}((\mu' P)P^{ln} - \mu' P^{ln}) \leq (1-\alpha)^n \mathbf{V}(\mu' P - \mu') \underset{n\to+\infty}{\to} 0$, et donc $\mathbf{V}(\mu P - \mu) = 0$. Ainsi $\mu P = \mu$, et si la chaîne est irréductible, elle est récurrente positive.

   (c) Soit $n \in \mathbb{N}^*$. On écrit $n = \lfloor n/l \rfloor l + j$ avec $0 \leq j \leq l-1$. $\mathbf{V}(\mu - p_n) = \mathbf{V}(\mu P^{l\lfloor n/l\rfloor} - p_j P^{l\lfloor n/l\rfloor}) \leq (1-\alpha)^{\lfloor n/l\rfloor}\mathbf{V}(\mu - p_j)$. La convergence est donc géométrique de raison $(1-\alpha)^{1/l}$. Plus $\alpha$ est proche de 1 et plus $l$ est petit, plus la convergence est rapide.

**Exercice 4.6.8.** *1. Montrer qu'une chaîne de Markov qui satisfait la condition de Doeblin est apériodique.*

*2. On suppose un espace d'état $E$ fini. Montrer qu'une chaîne irréductible et apériodique satisfait la condition de Doeblin. (On pourra montrer que sa matrice de transition satisfait $\forall x, y \in E, P^l(x,y) > 0$ pour $l$ assez grand.)*

# Chapitre 5

# Le Mouvement Brownien

## 5.1 Le mouvement brownien

Le mouvement brownien porte le nom du botaniste Robert Brown qui, en 1827, s'intéressa au mouvement du pollen dans l'eau. La particularité des trajectoires observées est d'être très erratique si bien que la notion de vitesse n'a alors plus de sens. Au début du XXème siècle et de façon tout a fait indépendante, Louis Bachelier et Albert Einstein ont utilisé le mouvement brownien pour modéliser d'autres trajectoires. Le premier pour décrire le cours d'une action en bourse tandis que le second s'intéressait à la diffusion de particules. Mais ce n'est qu'en 1923 avec Norbert Wiener que le mouvement brownien a été défini de manière rigoureuse d'un point de vue mathématique.

**Définition 5.1.1.** *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace de probabilité et* $I = \mathbb{R}_+$ *ou* $[0, T]$. *On appelle mouvement brownien (standard) réel sur* $I$ *une famille de variables aléatoires réelles* $(W_t, t \in I)$ *telle que :*

1. $W_0 = 0$
2. *Pour tout* $s, t \in I$ *t.q.* $s \leq t$, $W_t - W_s$ *suit une loi gaussienne centrée de variance* $t - s$.
3. *Pour tout* $n \in \mathbb{N}^*$ *et tout* $t_1, \ldots, t_n \in I$ *tels que* $0 = t_0 < t_1 < t_2 < \cdots < t_n$, *les accroissements* $(W_{t_i} - W_{t_{i-1}}, 1 \leq i \leq n)$ *sont indépendants.*
4. *Presque sûrement, les trajectoires* $t \mapsto W_t$ *sont continues. C'est à dire que* $\mathbb{P}(\{\omega, t \mapsto W_t(\omega)\ \mathcal{C}^0\}) = 1$.

Dans ce cours, nous admettrons l'existence du mouvement brownien. Le mouvement brownien est en général noté $(W_t, t \geq 0)$ en référence à Wiener ou $(B_t, t \geq 0)$ en référence à Brown. D'après la définition, $W_t$ est une v.a. réelle de loi $\mathcal{N}(0, t)$, elle a donc pour densité :

$$\forall x \in \mathbb{R}, \frac{1}{\sqrt{2\pi t}} \exp(-\frac{x^2}{2t}).$$

Voici quelques calculs immédiats : $\mathbb{E}[W_t] = 0$, $\mathbb{E}[W_t^2] = t$, et

$$\forall t, s \in I, \mathbb{E}[W_t W_s] = s \wedge t. \tag{5.1}$$

Grâce aux propriétés 2 et 3 de la définition du mouvement brownien standard, on obtient que $(W_{t_i} - W_{t_{i-1}}, 1 \leq i \leq n)$ est un vecteur gaussien de matrice de covariance $diag(t_1, t_2 - t_1, \ldots, t_n - t_{n-1})$. Par conséquent, $(W_{t_1}, \ldots, W_{t_n})$ est un vecteur gaussien centré et de matrice de covariance $(t_i \wedge t_j)_{1 \leq i,j \leq n}$.

**Définition 5.1.2.** *Soit $I$ un intervalle de $\mathbb{R}$. On appelle processus sur $I$ à valeurs dans un espace mesurable $(E, \mathcal{E})$ une famille de v.a. $(X_t, t \in I)$ telle que pour $t \in I$, $X_t : \Omega \to E$. Lorsque $E = \mathbb{R}$ (resp. $E = \mathbb{R}^d$), on dit que le processus $(X_t, t \in I)$ est à valeurs réelles (resp. vectorielles). Dans ce cas, on dit que le processus est continu si ses trajectoires $t \mapsto X_t(\omega)$ sont p.s. continues.*

**Définition 5.1.3.** *Un processus réel $(X_t, t \in I)$ est un processus gaussien si pour tout $n \in \mathbb{N}^*$ et tout $t_1, \ldots, t_n \in I$, $(X_{t_1}, \ldots, X_{t_n})$ est un vecteur gaussien.*

Avec ces définitions, un mouvement brownien standard est un processus gaussien continu.

**Exercice 5.1.4.** *Soit $(W_t, t \geq 0)$ un mouvement brownien standard. Pour $T > 0$, on appelle pont brownien entre $0$ et $T$ le processus $p_t^T = W_t - \frac{t}{T} W_T$ défini pour $t \in [0, T]$. Montrer que $(W_t, t \geq T)$ est indépendant du processus $(p_t^T, t \in [0, T])$, c'est à dire que $(W_{u_1}, \ldots, W_{u_n})$ est indépendant de $(p_{t_1}^T, \ldots, p_{t_n}^T)$ pour tout $n \in \mathbb{N}^*$, $t_1, \ldots, t_n \in [0, T]$ et $u_1, \ldots, u_n \geq T$.*

Le mouvement brownien satisfait les propriétés suivantes.

**Proposition 5.1.5.** *Soit $(W_t, t \geq 0)$ un mouvement brownien standard réel sur $\mathbb{R}_+$. Nous avons les propriétés suivantes.*

1. *(Symétrie) $(-W_t, t \geq 0)$ est un mouvement brownien standard.*
2. *(Invariance par translation) Si $T \geq 0$, $(W_{t+T} - W_T, t \geq 0)$ est un mouvement brownien standard.*
3. *(Changement d'échelle) Si $\lambda > 0$, $(\frac{1}{\sqrt{\lambda}} W_{\lambda t}, t \geq 0)$ est un mouvement brownien standard.*
4. *(Retournement du temps) Si $T > 0$, $(W_T - W_{T-t}, t \in [0, T])$ est un mouvement brownien standard.*

La preuve de ces propriétés est une application directe de la définition du mouvement brownien standard et est laissée en exercice. Ces propriétés se visualisent sur les figures 5.1 et 5.2 : si on enlève les graduations en temps et en espaces, les deux trajectoires qualitativement se ressemblent et sans elles on serait incapable de mettre une échelle (propriété

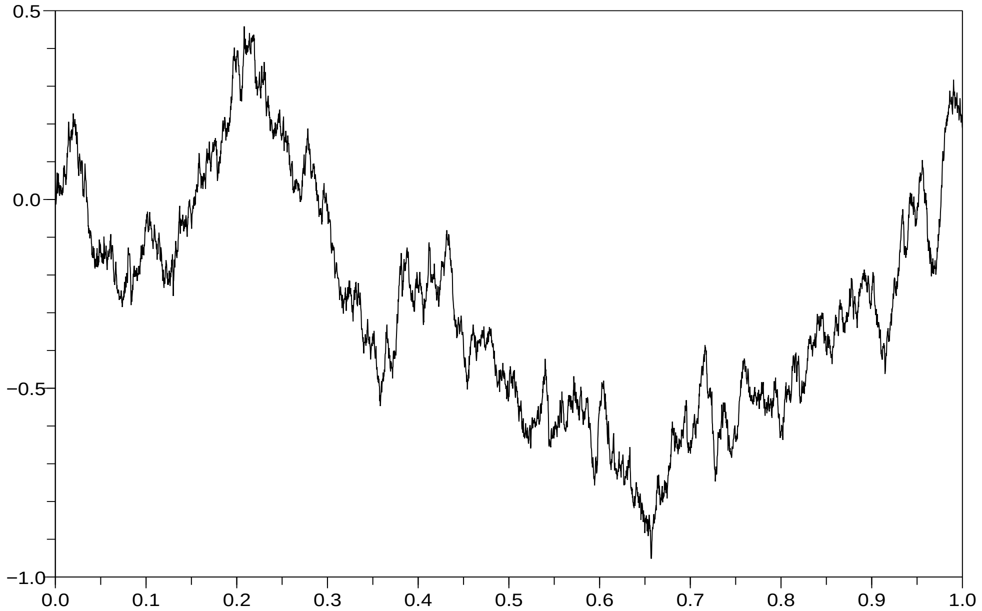


FIG. 5.1 – Un exemple de trajectoire brownienne (standard) sur $[0, 1]$.

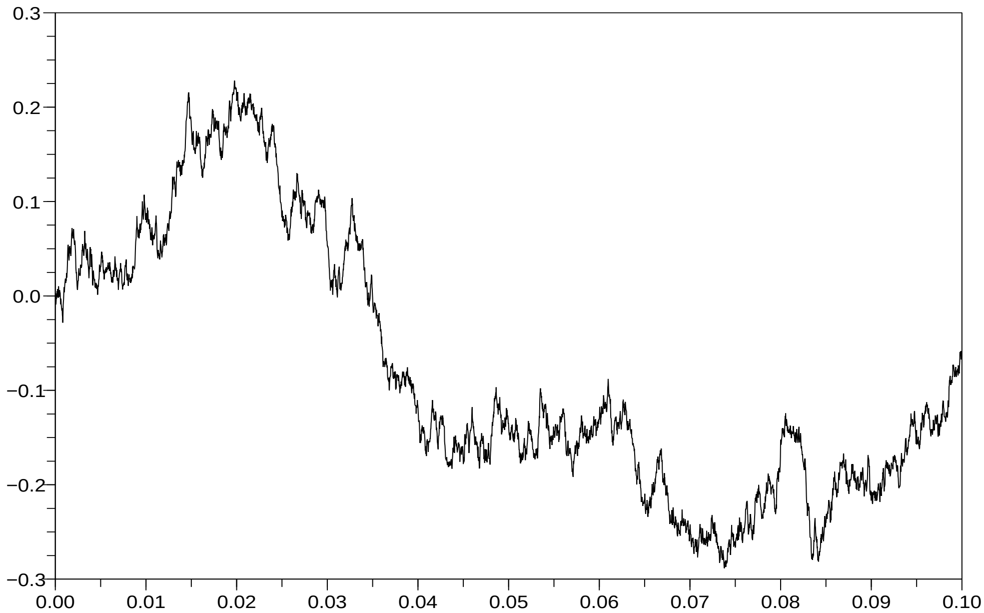


FIG. 5.2 – La même trajectoire brownienne sur $[0, 1/10]$.

3), de dire si les trajectoires sont parcourues ou pas dans le sens du temps (propriété 4). Enfin, même avec les graduations, on est bien incapable de dire si on a tracé $W_t$ ou $-W_t$ (propriété 1) ou même $W_{t+T} - W_T$ pour un certain $T > 0$ (propriété 2).

**Définition 5.1.6.** *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $I = \mathbb{R}_+$ ou $[0, T]$. On appelle mouvement brownien réel sur $I$ issu de $x \in \mathbb{R}$, de dérive $\mu \in \mathbb{R}$ et de coefficient de diffusion $\sigma > 0$ une famille de variables aléatoires réelles $(W_t, t \in I)$ telle que :*

1. $W_0 = x$
2. *Pour tout $s, t \in I$ t.q. $s \leq t$, $W_t - W_s$ suit une loi gaussienne de moyenne $\mu(t-s)$ et de variance $\sigma^2(t-s)$.*
3. *Pour tout $n \in \mathbb{N}^*$ et tout $t_1, \ldots, t_n \in I$ tels que $0 = t_0 < t_1 < t_2 < \cdots < t_n$, les accroissements $(W_{t_i} - W_{t_{i-1}}, 1 \leq i \leq n)$ sont indépendants.*
4. *Presque sûrement, les trajectoires $t \mapsto W_t$ sont continues.*

On peut montrer (nous l'admettrons ici) qu'un processus $(Z_t, t \geq 0)$ à valeurs réelles qui satisfait les propriétés suivantes :

1. $\mathbb{P}(Z_0 = x) = 1$,
2. (Stationnarité des accroissements) $\forall t, u \geq 0, Z_{t+u} - Z_u \overset{loi}{=} Z_t - Z_0$.
3. (Indépendance des accroissements) pour tout $n \in \mathbb{N}^*$ et tout $0 = t_0 < t_1 < t_2 < \cdots < t_n$, les accroissements $(Z_{t_i} - Z_{t_{i-1}}, 1 \leq i \leq n)$ sont indépendants,
4. les trajectoires de $(Z_t, t \geq 0)$ sont p.s continues,

est un mouvement brownien issu de $x$ pour une certaine dérive $\mu$ et un certain coefficient de diffusion $\sigma$. Un processus qui vérifie les propriétés 1 à 3 est appelé un *processus à accroissements indépendants et stationnaires (PAIS)*. C'est le fait d'imposer la continuité (propriété 4) qui permet d'obtenir que les accroissements suivent des loi gaussiennes. Il existe des PAIS non continus, et le processus de Poisson que nous verrons plus tard en est un exemple. Tous les PAIS à valeurs réelles sont en fait bien répertoriés et portent dans la littérature le nom de Processus de Lévy.

**Exercice 5.1.7.** *Soit $(x, r, \sigma) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+^*$.*

1. *Montrer que si $(W_t, t \geq 0)$ est un mouvement brownien standard, alors $(x + \mu t + \sigma W_t, t \geq 0)$ est un mouvement brownien issu de $x$, de dérive $\mu$ et de coefficient de diffusion $\sigma$.*
2. *Montrer que si $(W_t, t \geq 0)$ est un mouvement brownien issu de $x$, de dérive $\mu$ et de coefficient de diffusion $\sigma$, $(\frac{W_t - \mu t - x}{\sigma}, t \geq 0)$ est un mouvement brownien standard.*

Grâce à cet exercice, nous voyons qu'un mouvement brownien général peut s'exprimer à partir d'un mouvement brownien standard à travers une transformation affine. Pour cette raison, **nous ne considèrerons par la suite que des mouvements browniens réels standards et appellerons par défaut mouvement brownien réel un mouvement brownien standard**.

Nous donnons désormais la définition d'un mouvement brownien à valeurs dans $\mathbb{R}^d$ :

**Définition 5.1.8.** *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $I = \mathbb{R}_+$ ou $[0, T]$. On appelle mouvement brownien à valeurs dans $\mathbb{R}^d$ sur $I$ issu de $x \in \mathbb{R}^d$, de dérive $\mu \in \mathbb{R}^d$ et de matrice de diffusion $\Gamma \in \mathcal{S}_+(\mathbb{R}^d)$ une famille de variables aléatoires réelles $(W_t, t \in I)$ telle que :*

1. *$W_0 = x$*
2. *$\forall s, t \in I$ t.q. $s \leq t$, $W_t - W_s$ est un vecteur gaussien de loi $\mathcal{N}_d(\mu(t-s), \Gamma(t-s))$.*
3. *Pour tout $n \in \mathbb{N}^*$ et tout $t_1, \ldots, t_n \in I$ tels que $0 = t_0 < t_1 < t_2 < \cdots < t_n$, les accroissements $(W_{t_i} - W_{t_{i-1}}, 1 \leq i \leq n)$ sont indépendants.*
4. *Presque sûrement, les trajectoires $t \mapsto W_t$ sont continues.*

*Lorsque $x = 0$, $\mu = 0$ et $\Gamma = I_d$, on dit que le mouvement brownien est standard.*

**Exercice 5.1.9.** *Soit $(W_t, t \geq 0)$ un mouvement brownien standard dans $\mathbb{R}^d$. Soient $x \in \mathbb{R}^n$, $\mu \in \mathbb{R}^n$ et $M \in \mathcal{M}_{n,d}(\mathbb{R})$ une matrice réelle. Montrer que $(x + \mu t + MW_t, t \geq 0)$ est un mouvement brownien dans $\mathbb{R}^n$ issu de $x$, de dérive $\mu$ et de matrice de diffusion $MM^*$.*

Avant de conclure cette section, nous allons expliquer comment simuler une trajectoire brownienne sur une grille de discrétisation régulière $t_i^N = iT/N$, $i = 0, \ldots N$. Commençons par le cas du mouvement brownien réel standard. Dans ce cas, les v.a. $W_{t_{i+1}^N} - W_{t_i^N}$, $i = 0, \ldots, N-1$ sont indépendantes et de loi $\mathcal{N}(0, T/N)$. Par conséquent, si $G_1, \ldots, G_N$ sont $N$ gaussiennes centrées réduites,

$$(W_{t_{i+1}^N} - W_{t_i^N}, i = 0, \ldots, N-1) \overset{loi}{=} \left( \sqrt{\frac{T}{N}} G_1, \ldots, \sqrt{\frac{T}{N}} G_N \right)$$

et donc :

$$(W_{t_i^N}, i = 1, \ldots, N) \overset{loi}{=} \left( \sqrt{\frac{T}{N}} G_1, \sqrt{\frac{T}{N}} (G_1 + G_2), \ldots, \sqrt{\frac{T}{N}} \sum_{i=1}^{N} G_i \right).$$

Pour simuler une trajectoire brownienne, il est ainsi suffisant d'être capable de simuler une suite de v.a. gaussiennes centrées réduites indépendantes. Rappelons ici que la plupart des langages de programmation fournissent une fonction random qui génère une suite de nombres (pseudo)-aléatoires qui se comporte comme une suite de v.a. uniformes sur $[0, 1]$ indépendantes. Pour simuler à partir de cette fonction random des v.a. gaussiennes centrées réduites indépendantes, plusieurs méthodes existent et nous rappelons ici la méthode de Box-Muller : si $U_1$ et $U_2$ sont deux v.a. indépendantes uniformes sur $[0,1]$,

$$N_1 = \sqrt{-2\ln(U_1)}\cos(2\pi U_2) \text{ et } N_2 = \sqrt{-2\ln(U_1)}\sin(2\pi U_2)$$

sont deux v.a. gaussiennes centrées réduites indépendantes.

Faisons maintenant quelques remarques à propos de la méthode de simulation présentée ci-dessus. Tout d'abord, il est facile de voir que la méthode proposée s'étend aisément au cas où la grille de discrétisation $0 = t_0^N < t_1^N < \cdots < t_N^N = T$ n'est pas uniforme : il suffit de renormaliser $G_i$ par $\sqrt{t_i^N - t_{i-1}^N}$ au lieu de $\sqrt{T/N}$. Une autre remarque importante est que l'on peut raffiner la grille au fur et à mesure. En effet, pour $t \in (t_i^N, t_{i+1}^N)$, $W_t - W_{t_i^N} - \frac{t-t_i^N}{t_{i+1}^N - t_i^N}(W_{t_{i+1}^N} - W_{t_i^N})$ est indépendant de $(W_{t_j^N}, j = 1, \ldots, N)$. Cela peut être vu comme une conséquence du résultat sur le pont brownien établi à l'exercice 5.1.4, mais nous le redémontrons directement. Puisque $(W_t - W_{t_i^N} - \frac{t-t_i^N}{t_{i+1}^N - t_i^N}(W_{t_{i+1}^N} - W_{t_i^N}), W_{t_j^N}, j = 1, \ldots, N)$ est un vecteur gaussien, il s'agit de vérifier que $\mathbb{E}\left[\left(W_t - W_{t_i^N} - \frac{t-t_i^N}{t_{i+1}^N - t_i^N}(W_{t_{i+1}^N} - W_{t_i^N})\right) W_{t_j^N}\right] = 0$. Ceci découle immédiatement de 5.1. Par conséquent, $W_t - W_{t_i^N} - \frac{t-t_i^N}{t_{i+1}^N - t_i^N}(W_{t_{i+1}^N} - W_{t_i^N})$ est une v.a. gaussienne centrée de variance $\frac{(t_{i+1}^N - t)(t - t_i^N)}{t_{i+1}^N - t_i^N}$ indépendante de $(W_{t_j^N}, j = 1, \ldots, N)$. Par conséquent, une fois que $(W_{t_j^N}, j = 1, \ldots, N)$ sont simulées, on peut simuler $W_t$ en tirant une v.a. gaussienne $\tilde{G}$ centrée réduite indépendante des précédents tirages et en prenant :

$$W_t = W_{t_i^N} + \frac{t - t_i^N}{t_{i+1}^N - t_i^N}(W_{t_{i+1}^N} - W_{t_i^N}) + \sqrt{\frac{(t_{i+1}^N - t)(t - t_i^N)}{t_{i+1}^N - t_i^N}}\tilde{G}.$$

Dans le cas particulier $t = \frac{t_i^N + t_{i+1}^N}{2}$ le plus utilisé en pratique, $\frac{(t_{i+1}^N - t)(t - t_i^N)}{t_{i+1}^N - t_i^N} = \frac{1}{4}(t_{i+1}^N - t_i^N)$. En utilisant récursivement ce procédé, on peut ainsi obtenir sur un intervalle de temps fixé de plus en plus de points de la trajectoire brownienne. Il existe une preuve de l'existence du mouvement brownien constructiviste appelée construction de Paul Lévy qui consiste à définir récursivement le mouvement brownien aux points $(t_i^{2^n} = iT/2^n, i = 0, \ldots, 2^n)$, et de prouver que ces points forment à la limite $n \to +\infty$ une courbe continue qui satisfait effectivement les propriétés du mouvement brownien.

Maintenant que nous avons décrit précisément comment simuler un mouvement brownien réel standard, il est très simple de générer des trajectoires de mouvements browniens réels ou vectoriels généraux. Pour un mouvement brownien réel issu de $x$, de dérive $\mu$ et de coefficient de diffusion $\sigma$, il suffit de prendre $(x + \mu t_j^N + \sigma W_{t_j^N}, j = 1, \ldots, N)$ où $(W_{t_j^N}, j = 1, \ldots, N)$ est la trajectoire d'un mouvement brownien standard, grâce à l'exercice 5.1.7. Pour un mouvement brownien standard dans $\mathbb{R}^d$, il est suffisant de générer $d$ trajectoires browniennes (standard) réelles indépendantes. Ensuite, pour simuler un mouvement brownien général à valeurs dans $\mathbb{R}^d$ issu de $x$, de dérive $\mu$ et de matrice de diffusion $\Gamma$, il suffit grâce à l'exercice 5.1.9 de prendre $(x + \mu t_j^N + MW_{t_j^N}, j = 1, \ldots, N)$ où $(W_{t_j^N}, j = 1, \ldots, N)$ est un mouvement brownien standard dans $\mathbb{R}^d$, et $M \in \mathcal{M}_{d,d}(\mathbb{R})$ est une matrice telle que $MM^* = \Gamma$. Il existe en général plusieurs matrices $M$ qui satisfont cette équation. Une méthode souvent utilisée pour obtenir une telle matrice $M$ est la décomposition de Choleski. Elle est présentée dans l'exercice suivant.

**Exercice 5.1.10.** *Soit $S = (s_{ij})_{1 \leq i,j \leq d}$ une matrice symétrique définie positive sur $\mathbb{R}^d$. Nous allons montrer qu'il existe une unique matrice triangulaire inférieure $L = (l_{ij})_{1 \leq i,j \leq d}$ telle que $S = LL^*$.*

1. *Montrer que $s_{11} > 0$, et que l'on a nécessairement $l_{11} = \sqrt{s_{11}}$ et $l_{i1} = s_{i1}/\sqrt{s_{11}}$.*
2. *En écrivant que*
$$L = \begin{bmatrix} l_{11} & 0 \\ (l_{i1})_{2 \leq i \leq d} & I_{d-1} \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 0 & (l_{ij})_{2 \leq i,j \leq d} \end{bmatrix},$$
*donner une construction itérative de $L$, et vérifier que la matice $L$ est unique.*
3. *Donner une méthode de simulation pour un mouvement brownien en dimension 2 de dérive nulle et de matrice de diffusion $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ pour $\rho \in (-1, 1)$.*

*On veut étendre cette construction au cas où $S$ est positive, mais pas nécessairement définie. On suppose que $S$ est de rang $r < d$ et que ses $r$ premières colonnes sont indépendantes.*

5. *Montrer qu'il existe $A \in \mathcal{M}_{r,d-r}(\mathbb{R})$ telle que $(s_{ij})_{1 \leq i \leq d, r < j \leq d} = (s_{ij})_{1 \leq i \leq d, 1 \leq j \leq r} \times A$, puis en notant $S_r = (s_{ij})_{1 \leq i \leq r, 1 \leq j \leq r}$ que*
$$S = \begin{bmatrix} S_r & S_r A \\ A^* S_r & A^* S_r A \end{bmatrix}.$$
*Montrer que $S_r$ est de rang $r$ et en déduire qu'elle est définie positive. En déduire qu'il existe une matrice $L_r$ triangulaire inférieure telle que $L_r L_r^* = S_r$, puis que*
$$S = LL^* \text{ avec } L = \begin{bmatrix} L_r & 0 \\ A^* L_r & 0 \end{bmatrix}.$$

## 5.2 Les processus à temps continu

L'objectif de cette section est fournir un cadre théorique minimum à l'étude des processus. On se place sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et on considère un intervalle $I$. Nous allons commencer par préciser ce que l'on entend par la loi d'un processus.

Un processus $(X_t, t \in I)$ à valeurs dans l'espace mesurable $(E, \mathcal{E})$ est une famille de v.a. à valeurs dans $(E, \mathcal{E})$. Il peut également être vu comme une application partant de $\Omega$

à valeurs dans $E^I$,

$$\begin{aligned} X : \ & \Omega \to E^I \\ & \omega \mapsto (X_t(\omega))_{t\in I} \end{aligned}$$

et pour voir $X$ comme une variable aléatoire, il est nécessaire au préalable de munir $E^I$ d'une tribu qui rende $X$ mesurable. On définit ainsi $\mathcal{E}^{\otimes I}$ comme la tribu de $E^I$ engendrée par les ensembles

$$\{x \in E^I, x(t_1) \in A_1, \dots, x(t_n) \in A_n\}, n \in \mathbb{N}^*, A_1, \dots, A_n \in \mathcal{E}, t_1, \dots, t_n \in I.$$

Autrement dit, $\mathcal{E}^{\otimes I}$ est la plus petite tribu de $E^I$ qui rende les applications coordonnées

$$\begin{aligned} \xi_t : \ & (E^I, \mathcal{E}^{\otimes I}) \to (E, \mathcal{E}) \\ & (x(s))_{s\in I} \mapsto x(t) \end{aligned}$$

mesurables pour tout $t \in I$. On vérifie alors facilement que $X : (\Omega, \mathcal{F}) \to (E^I, \mathcal{E}^{\otimes I})$ est mesurable : en effet, pour $t_1, \dots, t_n \in I$ et $A_1, \dots, A_n \in \mathcal{E}$, $\cap_{i=1}^n \{X_{t_i} \in A_i\} \in \mathcal{F}$. Par conséquent, un processus peut être vu comme une variable aléatoire à valeurs dans $E^I$, et sa loi est caractérisée par

$$\mathbb{P}(X_{t_1} \in A_1, \dots, X_{t_n} \in A_n), n \in \mathbb{N}^*, A_1, \dots, A_n \in \mathcal{E}, t_1, \dots, t_n \in I.$$

On en déduit la proposition suivante.

**Proposition 5.2.1.** *Soient $(X_t, t \in I)$ et $(Y_t, t \in J)$ deux processus à valeurs dans $(E, \mathcal{E})$.*

- *Ils ont même loi si les intervalles $I$ et $J$ sont identiques et si pour tout $n \in \mathbb{N}^*$, $A_1, \dots, A_n \in \mathcal{E}$ et $t_1, \dots, t_n \in I$,*

$$\mathbb{P}(X_{t_1} \in A_1, \dots, X_{t_n} \in A_n) = \mathbb{P}(Y_{t_1} \in A_1, \dots, Y_{t_n} \in A_n).$$

- *Ils sont indépendants si pour tout $n \in \mathbb{N}^*$, $A_1, \dots, A_n, B_1, \dots, B_n \in \mathcal{E}$, $t_1, \dots, t_n \in I, s_1, \dots, s_n \in J$,*

$$\begin{aligned} & \mathbb{P}(X_{t_1} \in A_1, \dots, X_{t_n} \in A_n, Y_{s_1} \in B_1, \dots, Y_{s_n} \in B_n) \\ & = \mathbb{P}(X_{t_1} \in A_1, \dots, X_{t_n} \in A_n)\mathbb{P}(Y_{s_1} \in B_1, \dots, Y_{s_n} \in B_n). \end{aligned}$$

Comme toujours, nous avons l'écriture fonctionnelle équivalente qui est souvent plus commode. La preuve est laissée en exercice.

**Proposition 5.2.2.** *Soient $(X_t, t \in I)$ et $(Y_t, t \in J)$ deux processus à valeurs dans $(E, \mathcal{E})$.*

- *Ils ont même loi si $I = J$ et si pour tout $n \in \mathbb{N}^*$, $t_1, \dots, t_n \in I$ et $f : (E^n, \mathcal{E} \otimes \cdots \otimes \mathcal{E}) \to (E, \mathcal{E})$ bornée mesurable,*

$$\mathbb{E}[f(X_{t_1}, \dots, X_{t_n})] = \mathbb{E}[f(Y_{t_1}, \dots, Y_{t_n})].$$

– *Ils sont indépendants si pour tout $n \in \mathbb{N}^*$, $t_1, \ldots, t_n \in I$,$s_1, \ldots, s_n \in J$ et $f, g$ : $(E^n, \mathcal{E} \otimes \cdots \otimes \mathcal{E}) \to (E, \mathcal{E})$ bornées mesurables,*

$$\mathbb{E}[f(X_{t_1}, \ldots, X_{t_n})g(Y_{s_1}, \ldots, Y_{s_n})] = \mathbb{E}[f(X_{t_1}, \ldots, X_{t_n})]\mathbb{E}[g(Y_{s_1}, \ldots, Y_{s_n})].$$

La condition d'indépendance écrite ici pour deux processus, s'étend naturellement à plus de deux processus. En pratique nous ne considèrerons que des processus à valeurs dans $\mathbb{R}$ (resp. $\mathbb{R}^d$).

**Remarque 5.2.3.** *Parfois, pour prouver l'égalité en loi (resp. l'indépendance) de deux processus, il est commode d'avoir plus de régularité pour les fonctions $f$ et $g$. Lorsque $E = \mathbb{R}^d$, grâce qu Théorème 1.2.1 rappelé au premier chapitre, il est suffisant de prouver que pour $f_1, \ldots, f_n : \mathbb{R}^d \to \mathbb{R}$ (resp. $g_1, \ldots, g_n : \mathbb{R}^d \to \mathbb{R}$) $\mathcal{C}^\infty$ et bornées, on a :*

$$\mathbb{E}\left[\prod_{i=1}^n f_i(X_{t_i})\right] = \mathbb{E}\left[\prod_{i=1}^n f_i(Y_{t_i})\right]$$

$$(\textit{resp. } \mathbb{E}\left[\prod_{i=1}^n f_i(X_{t_i})g_i(Y_{s_i})\right] = \mathbb{E}\left[\prod_{i=1}^n f_i(X_{t_i})\right]\mathbb{E}\left[\prod_{i=1}^n g_i(Y_{s_i})\right]).$$

**Proposition 5.2.4.** *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité, $(W_t, t \geq 0)$ un mouvement brownien réel et $T > 0$. Alors, les processus $(W_t, t \in [0,T])$ et $(W_{T+t} - W_T, t \geq 0)$ sont deux mouvements browniens indépendants.*

*Preuve.* Il s'agit de vérifier l'indépendance, le fait que ce sont des mouvements browniens a déjà été vu (cf. proposition 5.1.5).

Soient $n \in \mathbb{N}^*$, $t_1 < \cdots < t_n$, $s_1 < \ldots, s_n$ et $f, g : \mathbb{R}^n \to \mathbb{R}$ mesurables bornées. On pose

$$A = \mathbb{E}[f(W_{t_1}, \ldots, W_{t_n})g(W_{T+s_1} - W_T, \ldots, W_{T+s_n} - W_T)].$$

On considère l'application linéaire $\alpha : (x_1, \ldots, x_n) \mapsto (x_1, x_1 + x_2, \ldots, x_1 + \cdots + x_n)$. Il vient que : $A = \mathbb{E}[f \circ \alpha(W_{t_1}, W_{t_2} - W_{t_1}, \ldots, W_{t_n} - W_{t_{n-1}})g \circ \alpha(W_{T+s_1} - W_T, W_{T+s_2} - W_{T+s_1}, \ldots, W_{T+s_n} - W_{T+s_{n-1}})]$. En utilisant le fait que les accroissements $(W_{t_1}, W_{t_2} - W_{t_1}, \ldots, W_{t_n} - W_{t_{n-1}}, W_T - W_{t_n}, W_{T+s_1} - W_T, W_{T+s_2} - W_{T+s_1}, \ldots, W_{T+s_n} - W_{T+s_{n-1}})$ sont indépendants, il vient que
$A = \mathbb{E}[f \circ \alpha(W_{t_1}, W_{t_2} - W_{t_1}, \ldots, W_{t_n} - W_{t_{n-1}})]$
$\times\, \mathbb{E}[g \circ \alpha(W_{T+s_1} - W_T, W_{T+s_2} - W_{T+s_1}, \ldots, W_{T+s_n} - W_{T+s_{n-1}})]$
$= \mathbb{E}[f(W_{t_1}, \ldots, W_{t_n})]\mathbb{E}[g(W_{T+s_1} - W_T, \ldots, W_{T+s_n} - W_T)]$. □

**Exercice 5.2.5.** *Montrer que $(W_t, t \geq 0)$ est un mouvement brownien standard à valeurs dans $\mathbb{R}^d$ si et seulement si ses coordonnées $(W_t^i, t \geq 0)$, $i = 1, \ldots, d$ sont des mouvements browniens réels indépendants. (On pourra se contenter de faire le cas $d = 2$)*

Comme dans le cadre discret nous allons définir la notion de filtration. Dans la suite de cette section, nous nous plaçons sur un intervalle

$$\boxed{I = [0, T] \text{ ou } \mathbb{R}_+.}$$

On appelle *filtration* une famille $(\mathcal{F}_t)_{t \in I}$ croissante de sous-tribus de $\mathcal{F}$, c'est à dire vérifiant :

$$\forall s, t \in I, s < t \implies \mathcal{F}_s \subset \mathcal{F}_t.$$

Heuristiquement, la filtration $(\mathcal{F}_t)_{t \in I}$ représente l'information connue sur les aléas à l'instant $t$. Un processus $(X_t, t \in I)$ à valeurs dans un espace mesurable $(E, \mathcal{E})$ est dit $(\mathcal{F}_t)$-adapté si

$$\forall t \in I, X_t \text{ est } \mathcal{F}_t\text{-mesurable.}$$

Lorsque l'on a un processus $(X_t, t \in I)$, on peut construire la filtration suivante :

$$(\sigma(X_s, s \leq t))_{t \in I}.$$

On l'appelle *filtration engendrée par le processus* $(X_t, t \in I)$, et c'est la plus petite filtration qui rende le processus $(X_t, t \in I)$ adapté.

Pour des processus $(X_t, t \in I)$ à valeurs réelles (ou $\mathbb{R}^d$), il est naturel dans certains problèmes de vouloir considérer des quantités comme l'intégrale de ce processus. Mais pour que cela ait un sens, il est necessaire de savoir si $t \mapsto X_t(\omega)$ est mesurable. On dira que le processus $(X_t, t \in I)$ est *progressivement mesurable* si pour tout $t \in I$,

$$\begin{aligned} X :\ & ([0, t] \times \Omega, \mathcal{B}([0, t]) \otimes \mathcal{F}_t) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ & (s, \omega) \mapsto X_s(\omega) \end{aligned}$$

est mesurable. Dans ce cas, $\int_0^t X_s ds$ est bien définie sur $\{\int_0^t |X_s| ds < \infty\}$. Un processus progressivement mesurable est adapté, mais la réciproque n'est pas vraie. Néanmoins, on peut montrer qu'un processus adapté continu (ou même seulement continu à droite) est progressivement mesurable. Et comme nous ne considèrerons que des processus continus, nous n'aurons pas à nous soucier de ce détail technique.

**Définition 5.2.6.** *On se place sur un espace de probabilité* $(\Omega, \mathcal{F}, \mathbb{P})$ *muni d'une filtration* $(\mathcal{F}_t)_{t \in I}$. *Un temps aléatoire est une v.a* $\tau : \Omega \to I \cup \{+\infty\}$. *Un temps aléatoire* $\tau$ *est un* $(\mathcal{F}_t)$*-temps d'arrêt si*

$$\forall t \in I, \{\tau \leq t\} \in \mathcal{F}_t.$$

**Exercice 5.2.7.** *Montrer que si* $\tau$ *est un* $(\mathcal{F}_t)$*-temps d'arrêt, on a pour tout* $t > 0$, $\{\tau < t\} \in \mathcal{F}_t$ *et* $\{\tau = t\} \in \mathcal{F}_t$. *(On pourra écrire que* $\{\tau < t\} = \cup_{n \in \mathbb{N}^*} \{\tau \leq t - 1/n\}$.*)*

Il est immédiat de voir que si $\tau$ est déterministe (i.e. $\mathbb{P}(\tau = s) = 1$ pour un certain $s \in I$), $\tau$ est un temps d'arrêt. Donnons un exemple plus intéressant de temps d'arrêt. Soit $(W_t, t \geq 0)$ un mouvement brownien et $(\mathcal{F}_t)_{t\geq 0}$ la filtration qu'il engendre. On pose pour $a > 0$,

$$\tau_a = \inf\{t \geq 0, W_t \geq a\}, (\inf \emptyset = +\infty)$$

le premier temps où le mouvement brownien passe au dessus de $a$. Comme les trajectoires du mouvement brownien sont p.s. continues, on voit que $\tau_a = \inf\{t \geq 0, W_t = a\}$. Pour montrer que $\tau_a$ est un temps d'arrêt, il suffit de montrer que pour $t \geq 0$, $\{\tau \leq t\} \in \mathcal{F}_t$. On a $\{\tau \leq t\} = \cup_{s\in[0,t]}\{W_s \geq a\} = \cap_{n\in\mathbb{N}^*} \cup_{s\in[0,t]} \{W_s > a - 1/n\} = \cap_{n\in\mathbb{N}^*} \cup_{s\in[0,t]\cap\mathbf{Q}} \{W_s > a-1/n\}$. L'ensemble $\cup_{s\in[0,t]\cap\mathbf{Q}}\{W_s > a-1/n\}$ est $\mathcal{F}_t$-mesurable comme union dénombrable d'ensembles $\mathcal{F}_t$-mesurables, et $\{\tau \leq t\}$ est $\mathcal{F}_t$-mesurable comme intersection dénombrable d'ensembles $\mathcal{F}_t$-mesurables. Ici nous avons utiliser le fait que les trajectoires du mouvements brownien sont p.s. continues pour obtenir que $\cup_{s\in[0,t]}\{W_s \geq a\} = \cap_{n\in\mathbb{N}^*} \cup_{s\in[0,t]} \{W_s > a - 1/n\}$ et $\cup_{s\in[0,t]}\{W_s > a - 1/n\} = \cup_{s\in[0,t]\cap\mathbf{Q}}\{W_s > a - 1/n\}$, ces égalités étant vraies à un ensemble négligeable près.

En revanche $\sigma = \sup\{t \in [0, 1], W_t \leq 0\}$ $(\sup \emptyset = +\infty)$ n'est pas un temps d'arrêt. En effet, on a $\{\sigma \leq t\} = \{\forall u \in (t, 1], W_u > 0\}$, et heuristiquement, nous devons connaître le futur de $W$ entre les instants $t$ et 1 pour déterminer si $\sigma \leq t$. L'événement $\{\sigma \leq t\}$ ne peut donc être $\mathcal{F}_t$-mesurable car $\mathcal{F}_t$ ne contient que l'information sur le passé de la trajectoire. Une preuve rigoureuse de ce résultat est proposée à l'exercice 5.4.8.

**Proposition 5.2.8.** *Si $\tau_1$ et $\tau_2$ (resp. $(\tau_n)_{n\in\mathbb{N}^*}$) sont deux (resp. est une suite de) $(\mathcal{F}_t)$-temps d'arrêt, $\tau_1 \wedge \tau_2$ et $\tau_1 \vee \tau_2$ (resp. $\wedge_{n\in\mathbb{N}^*}\tau_n$ et $\vee_{n\in\mathbb{N}^*}\tau_n$) sont deux $(\mathcal{F}_t)$-temps d'arrêt.*

La preuve de ce résultat est identique à celle pour les temps d'arrêt dans les filtrations discrètes et est laissé en exercice.

**Définition 5.2.9.** *Soient $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité muni d'une filtration $(\mathcal{F}_t)_{t\in I}$ et $\tau$ un $(\mathcal{F}_t)$-temps d'arrêt. On définit :*

$$\mathcal{F}_\tau = \{A \subset \Omega \text{ t.q. } \forall t \in I, A \cap \{\tau \leq t\} \in \mathcal{F}_t\}.$$

*C'est une tribu qui représente l'information connue jusqu'en $\tau$.*

Le fait qu'il s'agit d'une tribu est laissé en exercice. On peut également vérifier que si $\tau = t$ p.s. (i.e. $\tau$ est un temps d'arrêt déterministe), $\mathcal{F}_\tau = \mathcal{F}_t$ si bien que les deux notations coïncident.

**Definition 5.2.10.** *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité muni d'une filtration $(\mathcal{F}_t)_{t\in I}$. On dit qu'un processus $(X_t, t \in I)$ à valeurs réelles (ou dans $\mathbb{R}^d$) est une $(\mathcal{F}_t)$-martingale si*

- $\forall t \in I,\ X_t$ *est* $\mathcal{F}_t$*-mesurable et intégrable.*
- $\forall s, t \in I, s \leq t \implies \mathbb{E}[X_t|\mathcal{F}_s] = X_s$.

Comme dans le cas discret, on appelle $(\mathcal{F}_t)$-surmartingale (resp. $(\mathcal{F}_t)$-sous-martingale) un processus $(X_t, t \in I)$ adapté tel que pour tout $t \in I$, $X_t$ est intégrable et pour tout $0 \leq s \leq t$, $\mathbb{E}[X_t|\mathcal{F}_s] \leq X_s$ (resp. $\mathbb{E}[X_t|\mathcal{F}_s] \geq X_s$).

Il est facile de voir qu'un mouvement brownien réel (ou dans $\mathbb{R}^d$) est une martingale par rapport à la filtration qu'il engendre. En effet soient $s, t \geq 0$ tels que $s \leq t$. Grâce à la proposition 5.2.4, $W_t - W_s$ est indépendant de $(W_u, u \in [0, s])$ et donc de la tribu $\mathcal{F}_s = \sigma(W_u, u \in [0, s])$. On en déduit :

$$\mathbb{E}[W_t|\mathcal{F}_s] = \mathbb{E}[W_t - W_s|\mathcal{F}_s] + W_s = \mathbb{E}[W_t - W_s] + W_s = W_s.$$

**Remarque 5.2.11.** *Lorsqu'on a un espace de probabilité* $(\Omega, \mathcal{F}, \mathbb{P})$ *muni d'une filtration* $(\mathcal{F}_t)_{t\in I}$*, on dit que* $(W_t, t \in I)$ *est un mouvement brownien par rapport à la filtration* $(\mathcal{F}_t)_{t\in I}$ *si l'on a les trois propriétés suivantes.*

- *Le processus* $(W_t, t \geq 0)$ *est un mouvement brownien.*
- *Le processus* $(W_t, t \geq 0)$ *est adapté par rapport à la filtration* $(\mathcal{F}_t)_{t\in I}$.
- *Pour tout* $s \geq 0$, $(W_{t+s} - W_s, t \geq 0)$ *est indépendant de* $\mathcal{F}_s$.

*Bien évidemment, un mouvement brownien est un mouvement brownien par rapport à la filtration qu'il engendre.*

**Exercice 5.2.12.** *(Inégalités maximales) On se place sur un espace de probabilité* $(\Omega, \mathcal{F}, \mathbb{P})$ *muni d'une filtration* $(\mathcal{F}_t)_{t\geq 0}$

1. *Soient* $(M_t, t \geq 0)$ *une* $(\mathcal{F}_t)$*-martingale continue,* $T > 0$ *et* $N \in \mathbb{N}^*$. *Montrer que* $(M_{\frac{kT}{N}}, k \in \mathbb{N})$ *est une martingale par rapport à la filtration discrète* $(\mathcal{F}_{\frac{kT}{N}}, k \in \mathbb{N})$ *définie par* $\mathcal{F}_k^N = \mathcal{F}_{\frac{kT}{N}}$. *En déduire grâce à l'inégalité de Doob que :*
$$\mathbb{E}[\max_{0\leq k\leq N} M^2_{\frac{kT}{N}}] \leq 4\mathbb{E}[M_T^2].$$
2. *En déduire que :*
$$\mathbb{E}[\max_{t\in[0,T]} M_t^2] \leq 4\mathbb{E}[M_T^2].$$
3. *De manière analogue, montrer l'inégalité maximale pour les sous-martingales continues positives i.e. si* $(X_t, t \geq 0)$ *est une* $(\mathcal{F}_t)$*-sous-martingale continue positive,*
$$\mathbb{P}(\max_{t\in[0,T]} X_t > x) \leq \frac{\mathbb{E}[X_T]}{x}.$$

## 5.3 Propriété de Markov du mouvement brownien

Dans cette section, nous allons étudier la propriété de Markov du mouvement brownien. Comme dans le cas discret, un processus $(X_t, t \geq 0)$ satisfait la propriété de Markov si pour tout $T > 0$, la loi du processus futur $(X_{T+t}, t \geq 0)$ sachant la trajectoire passée $(X_t, 0 \leq t \leq T)$ est la même que la loi du processus futur $(X_{T+t}, t \geq 0)$ sachant l'état présent $X_T$. En d'autres termes, la meilleure description que l'on a de l'évolution future du processus à partir du passé est aussi la meilleure description que l'on a connaissant seulement l'état présent. On dit ainsi parfois que l'information utile du passé pour décrire le futur se résume au présent. Voici une définition mathématique de cela.

**Définition 5.3.1.** *Un processus $(X_t, t \geq 0)$ à valeurs dans un espace mesurable $(E, \mathcal{E})$ est un processus de Markov si pour tout $T > 0$, $T \leq t_1 < \cdots < t_n$ et $f : (E^n, \mathcal{E}^{\otimes n}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable bornée,*

$$\mathbb{E}[f(X_{t_1}, \ldots, X_{t_n})|(X_t, 0 \leq t \leq T)] = \mathbb{E}[f(X_{t_1}, \ldots, X_{t_n})|X_T].$$

Pour un processus de Markov, le futur et le passé sont indépendants conditionnellement au présent. Soient $T > 0$, $T \leq t_1 < \cdots < t_n$ et $0 \leq u_1 < \cdots < u_n \leq T$. Il s'agit de montrer que pour $f, g : (E^n, \mathcal{E}^{\otimes n}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurables bornées, on a

$$\mathbb{E}[f(X_{t_1}, \ldots, X_{t_n})g(X_{u_1}, \ldots, X_{u_n})|X_T] = \mathbb{E}[f(X_{t_1}, \ldots, X_{t_n})|X_T]\mathbb{E}[g(X_{u_1}, \ldots, X_{u_n})|X_T]. \tag{5.2}$$

On utilise un double conditionnement :
$\mathbb{E}[f(X_{t_1}, \ldots, X_{t_n})g(X_{u_1}, \ldots, X_{u_n})|X_T]$
$= \mathbb{E}[\mathbb{E}[f(X_{t_1}, \ldots, X_{t_n}) \underbrace{g(X_{u_1}, \ldots, X_{u_n})}_{\sigma(X_t, 0 \leq t \leq T) \text{ mesurable}} |(X_t, 0 \leq t \leq T)]|X_T]$
$= \mathbb{E}[\mathbb{E}[f(X_{t_1}, \ldots, X_{t_n})|(X_t, 0 \leq t \leq T)]g(X_{u_1}, \ldots, X_{u_n})|X_T]$
$= \mathbb{E}[\mathbb{E}[f(X_{t_1}, \ldots, X_{t_n})|X_T]g(X_{u_1}, \ldots, X_{u_n})|X_T]$ grâce à la propriété de Markov
$= \mathbb{E}[f(X_{t_1}, \ldots, X_{t_n})|X_T]\mathbb{E}[g(X_{u_1}, \ldots, X_{u_n})|X_T]$ car $\mathbb{E}[f(X_{t_1}, \ldots, X_{t_n})|X_T] \in \sigma(X_T)$.

Réciproquement, si un processus $(X_t, t \geq 0)$ satisfait (5.2), on a :
$\mathbb{E}[\mathbb{E}[f(X_{t_1}, \ldots, X_{t_n})|X_T]g(X_{u_1}, \ldots, X_{u_n})]$
$= \mathbb{E}[\mathbb{E}[\mathbb{E}[f(X_{t_1}, \ldots, X_{t_n})|X_T]g(X_{u_1}, \ldots, X_{u_n})|X_T]]$
$= \mathbb{E}[\mathbb{E}[f(X_{t_1}, \ldots, X_{t_n})|X_T]\mathbb{E}[g(X_{u_1}, \ldots, X_{u_n})|X_T]]$
$= \mathbb{E}[\mathbb{E}[f(X_{t_1}, \ldots, X_{t_n})g(X_{u_1}, \ldots, X_{u_n})|X_T]]$ grace à (5.2)
$= \mathbb{E}[f(X_{t_1}, \ldots, X_{t_n})g(X_{u_1}, \ldots, X_{u_n})]$.
Par conséquent, $\mathbb{E}[f(X_{t_1}, \ldots, X_{t_n})|X_T] = \mathbb{E}[f(X_{t_1}, \ldots, X_{t_n})|(X_t, 0 \leq t \leq T)]$ et $(X_t, t \geq 0)$ est un processus de Markov. Nous venons ainsi de prouver la proposition suivante.

**Proposition 5.3.2.** *Un processus $(X_t, t \geq 0)$ est un processus de Markov si et seulement pour tout $T > 0$, les processus $(X_t, 0 \leq t \leq T)$ et $(X_{T+t}, t \geq 0)$ sont indépendants conditionnellement à $X_T$ (equation (5.2)).*

**Proposition 5.3.3.** *Le mouvement brownien satisfait la propriété de Markov.*

*Preuve.* D'après la Proposition 5.2.4, $(W_t, 0 \leq t \leq T)$ et $(W_{t+T} - W_T, t \geq 0)$ sont deux mouvements browniens indépendants. Soient $f : \mathbb{R}^n \to \mathbb{R}$ mesurable bornée et $T \leq t_1 < \cdots < t_n$. On a donc par la Proposition 2.2.16 $\mathbb{E}[f(W_{t_1}, \ldots, W_{t_n})|(W_t, 0 \leq t \leq T)] = \psi(W_T)$ où

$$\psi(x) = \mathbb{E}[f(W_{t_1-T} + x, \ldots, W_{t_n-T} + x)],$$

et par conséquent $\mathbb{E}[f(W_{t_1}, \ldots, W_{t_n})|(W_t, 0 \leq t \leq T)] = \mathbb{E}[f(W_{t_1}, \ldots, W_{t_n})|W_T]$. □

Nous allons désormais montrer la propriété de Markov forte. Comme dans le cas discret, on parle de propriété de Markov forte lorsque l'on peut écrire la propriété de Markov pour un temps d'arrêt $\tau$ au lieu d'un temps déterministe $T$. Puisque dans ce cours nous ne verrons cette propriété que pour le mouvement brownien, nous nous contentons de l'énoncer dans ce cadre sans donner la définition générale.

**Proposition 5.3.4.** *Soit $(W_t, t \geq 0)$ un mouvement brownien et $\tau$ un temps d'arrêt par rapport à la filtration $(\mathcal{F}_t)_{t\geq 0}$ engendrée par ce mouvement brownien tel que $\mathbb{P}(\tau < \infty) = 1$. Alors, $(W_{t+\tau} - W_\tau, t \geq 0)$ est un mouvement brownien indépendant de $\mathcal{F}_\tau$. En particulier sachant $\mathcal{F}_\tau$, $(W_{t+\tau}, t \geq 0)$ est un mouvement brownien issu de $W_\tau$.*

*Preuve.* Nous voulons montrer que $(W_{t+\tau} - W_\tau, t \geq 0)$ est un mouvement brownien indépendant de $\mathcal{F}_\tau$, c'est à dire que

$$\mathbb{E}[\mathbf{1}_A f(W_{\tau+s_1} - W_\tau, \ldots, W_{\tau+s_n} - W_\tau)] = \mathbb{P}(A)\mathbb{E}[f(W_{s_1}, \ldots, W_{s_n})] \tag{5.3}$$

pour toute fonction $f : \mathbb{R}^n \to \mathbb{R}$ bornée mesurable et tout ensemble $A \in \mathcal{F}_\tau$. Commençons par le cas où $\tau$ prend un nombre dénombrable de valeurs $T_k, k \in \mathbb{N}$. Soient $A \in \mathcal{F}_\tau$, $0 \leq s_1 < \cdots < s_n$ et $f : \mathbb{R}^n \to \mathbb{R}$ une fonction bornée mesurable.

$$\begin{aligned}
&\mathbb{E}[\mathbf{1}_A f(W_{\tau+s_1} - W_\tau, \ldots, W_{\tau+s_n} - W_\tau)] \\
&= \mathbb{E}[\textstyle\sum_{k\in\mathbb{N}} \mathbf{1}_{A\cap\{\tau=T_k\}} f(W_{T_k+s_1} - W_{T_k}, \ldots, W_{T_k+s_n} - W_{T_k})] \\
&= \textstyle\sum_{k\in\mathbb{N}} \mathbb{E}[\underbrace{\mathbf{1}_{A\cap\{\tau=T_k\}}}_{\in\mathcal{F}_{T_k}} f(W_{T_k+s_1} - W_{T_k}, \ldots, W_{T_k+s_n} - W_{T_k})] \\
&= \textstyle\sum_{k\in\mathbb{N}} \mathbb{E}[\mathbf{1}_{A\cap\{\tau=T_k\}}]\mathbb{E}[f(W_{T_k+s_1} - W_{T_k}, \ldots, W_{T_k+s_n} - W_{T_k})] \\
&= \left[\textstyle\sum_{k\in\mathbb{N}} \mathbb{P}(A\cap\{\tau = T_k\})\right] \mathbb{E}[f(W_{s_1}, \ldots, W_{s_n})] = \mathbb{P}(A)\mathbb{E}[f(W_{s_1}, \ldots, W_{s_n})]
\end{aligned}$$

Maintenant nous souhaitons prouver cela pour $\tau$ quelconque. Pour $N \in \mathbb{N}^*$, on pose $\tau^{(N)} = \frac{\lceil \tau N \rceil}{N}$ où $\lceil x \rceil$ désigne pour $x \in \mathbb{R}$ la partie entière supérieure de $x$. Il s'agit d'un temps d'arrêt par rapport à $(\mathcal{F}_t)_{t\geq 0}$. En effet, pour $k \in \mathbb{N}^*$

$$\{\tau^{(N)} = k/N\} = \{\frac{k-1}{N} < \tau \leq \frac{k}{N}\} \in \mathcal{F}_{\frac{k}{N}},$$

et $\{\tau^{(N)} \leq t\} = \cup_{k,k/N\leq t}\{\tau^{(N)} = k/N\} \in \mathcal{F}_t$. En outre, pour $A \in \mathcal{F}_\tau$, on a $A\cap\{\tau^{(N)} = k/N\} \in \mathcal{F}_{k/N}$. Par conséquent on a

$$\mathbb{E}[\mathbf{1}_A f(W_{\tau^{(N)}+s_1} - W_{\tau^{(N)}}, \ldots, W_{\tau^{(N)}+s_n} - W_{\tau^{(N)}})] = \mathbb{P}(A)\mathbb{E}[f(W_{s_1}, \ldots, W_{s_n})] \tag{5.4}$$

puisque $\tau^{(N)}$ prend un nombre dénombrable de valeurs. Grâce à la remarque 5.2.3, il est suffisant en réalité de prouver (5.3) pour $f(x_1, \ldots, x_n) = \prod_{i=1}^n f_i(x_i)$ où les fonctions $f_i$ sont $\mathcal{C}^\infty$ et bornées. Dans ce cas, puisque $\tau^{(N)} \underset{N\to+\infty}{\to} \tau$, $f(W_{\tau^{(N)}+s_1} - W_{\tau^{(N)}}, \ldots, W_{\tau^{(N)}+s_n} - W_{\tau^{(N)}}) \underset{N\to+\infty}{\to} f(W_{\tau+s_1} - W_\tau, \ldots, W_{\tau+s_n} - W_\tau)$ par continuité de $f$ et du mouvement brownien. La fonction $f$ étant bornée il existe $M_f > 0$ t.q. $\mathbf{1}_A|f(W_{\tau+s_1} - W_\tau, \ldots, W_{\tau+s_n} - W_\tau)| \leq M_f$. En appliquant le théorème de convergence dominée, il vient que : $\mathbb{E}[\mathbf{1}_A f(W_{\tau^{(N)}+s_1} - W_{\tau^{(N)}}, \ldots, W_{\tau^{(N)}+s_n} - W_{\tau^{(N)}})] \underset{N\to+\infty}{\to} \mathbb{E}[\mathbf{1}_A f(W_{\tau+s_1} - W_\tau, \ldots, W_{\tau+s_n} - W_\tau)]$ puis (5.3) grâce à (5.4).

Il nous reste à montrer que sachant $\mathcal{F}_\tau$, $(W_{t+\tau}, t \geq 0)$ est un mouvement brownien issu de $W_\tau$. Soient $f : \mathbb{R}^n \to \mathbb{R}$ mesurable bornée et $0 \leq t_1 < \cdots < t_n$. Puisque $(W_{t+\tau} - W_\tau, t \geq 0)$ est un mouvement brownien indépendant de $\mathcal{F}_\tau$, on a $\mathbb{E}[f(W_{\tau+t_1}, \ldots, W_{\tau+t_n})|\mathcal{F}_\tau] = \psi(W_\tau)$ où $\psi(x) = \mathbb{E}[f(W_{t_1} + x, \ldots, W_{t_n} + x)]$, ce qui prouve le résultat. □

**Exercice 5.3.5.** *(Principe de réflexion) On se donne $(W_t, t \geq 0)$ un mouvement brownien. Pour $t \geq 0$, on pose $S_t = \sup_{s\in[0,t]} W_s$ et pour $b \geq 0$, on pose $\tau_b = \inf\{t \geq 0, W_t \geq b\}$.*

1. *Montrer que $\{S_t \geq b\} = \{\tau_b \leq t\}$.*
2. *On admet que $\mathbb{P}(\tau_b < \infty) = 1$ (cf. (5.6)). Montrer que $(W_{\tau_b+t} - W_{\tau_b}, t \geq 0)$ et $(W_{\tau_b} - W_{\tau_b+t}, t \geq 0)$ sont deux mouvements browniens qui sont chacun indépendants de $\mathcal{F}_{\tau_b}$.*
3. *Pour $a \leq b$, montrer que $\mathbb{P}(S_t \geq b, W_t \leq a) = \mathbb{P}(\tau_b \leq t, W_t - W_{\tau_b} \leq a - b)$, puis que $\mathbb{P}(S_t \geq b, W_t \leq a) = P(W_t \geq 2b - a)$.*
4. *En déduire que $\mathbb{P}(S_t \geq b) = 2\mathbb{P}(W_t \geq b)$, puis que pour tout $t > 0$, $S_t$ et $|W_t|$ ont même loi. Les processus $(S_t, t \geq 0)$ et $(|W_t|, t \geq 0)$ ont-ils même loi ?*
5. *Calculer la loi de $\tau_b$ et retrouver l'expression donnée en (5.7).*
6. *Calculer la densité de la loi de $(S_t, W_t)$ et obtenir ensuite la loi de $S_t - W_t$ et de $2S_t - W_t$. En écrivant $S_t - W_t = \sup_{s\in[0,t]}(W_s - W_t)$, retrouver cette loi par un argument de retournement du temps.*

## 5.4 Martingales usuelles associées au mouvement brownien

Dans cette section nous allons construire des martingales à partir du mouvement brownien. Avant cela, nous allons énoncer et prouver le théorème d'arrêt pour les martingales continues, qui est tout à fait analogue à celui rencontré pour les martingales discrètes.

**Théorème 5.4.1.** *On se donne un espace de probabilité* $(\Omega, \mathcal{F}, \mathbb{P})$ *muni d'une filtration* $(\mathcal{F}_t)_{t\geq 0}$ *et* $(M_t, t \geq 0)$ *une* $(\mathcal{F}_t)$*-martingale à trajectoire p.s. continues. Soit* $\tau$ *un* $(\mathcal{F}_t)$*-temps d'arrêt borné :* $\exists K > 0, \mathbb{P}(\tau \leq K) = 1$*. Sous ces hypothèses, on a :*

$$\mathbb{E}[M_\tau] = \mathbb{E}[M_0].$$

*Preuve.* Soit $N \in \mathbb{N}^*$. Il est facile de vérifier que $(M_{\frac{k}{N}}, k \in \mathbb{N})$ est une martingale par rapport à la filtration discrète $(\mathcal{F}_{\frac{k}{N}})_{k\in\mathbb{N}}$. On pose $\tau^{(N)} = \lceil N\tau \rceil$ : il s'agit d'un temps d'arrêt dans cette filtration discrète. En effet,

$$\{\tau^{(N)} = k\} = \{k - 1 < N\tau \leq k\} = \{\frac{k-1}{N} < \tau \leq \frac{k}{N}\} \in \mathcal{F}_{\frac{k}{N}}.$$

En outre, $\tau^{(N)} \leq \lceil NK \rceil$ est borné, et d'après le Théorème d'arrêt des Martingale discrètes il vient que :

$$\mathbb{E}[M_{\frac{\tau^{(N)}}{N}}] = \mathbb{E}[M_0]. \tag{5.5}$$

A ce stade, puisque la martingale $M$ est continue et puisque $\frac{\tau^{(N)}}{N}$ converge vers $\tau$ lorsque $N \to +\infty$, $M_{\frac{\tau^{(N)}}{N}}$ converge presque sûrement vers $M_\tau$ (remarquer ici qu'en fait, on utilise seulement la continuité à droite de la martingale puisque $\frac{\tau^{(N)}}{N} \geq \tau$). On voudrait ainsi passer à la limite dans (5.5) pour obtenir le résultat. Malheureusement, ce passage à la limite est délicat à prouver. Il fait l'objet de l'exercice suivant. □

**Exercice 5.4.2.** *Cette exercice permet de justifier le passage à la limite dans* (5.5) *pour obtenir le théorème d'arrêt.*

1. *Montrer que si* $\sigma$ *est un* $(\mathcal{F}_t)$*-temps d'arrêt borné prenant un nombre dénombrable de valeurs, et si* $t$ *est tel que* $\mathbb{P}(\sigma \leq t) = 1$, $\mathbb{E}[M_t|\mathcal{F}_\sigma] = M_\sigma$.
2. *Montrer que* $\frac{\tau^{(N)}}{N}$ *est un* $(\mathcal{F}_t)$*-temps d'arrêt borné par* $K + 1$*. Déduire de la question précédente que* $\mathbb{E}[M_{K+1}|\mathcal{F}_{\frac{\tau^{(N)}}{N}}] = M_{\frac{\tau^{(N)}}{N}}$.
3. *Montrer que si* $X$ *est intégrable, on a :*

$$\forall\varepsilon > 0, \exists\delta > 0, \forall A \in \mathcal{F}, 0 < \mathbb{P}(A) \leq \delta \implies \mathbb{E}[|X|\mathbf{1}_A] \leq \varepsilon.$$

4. *Soit* $\varepsilon > 0$*. Puisque* $M_{K+1}$ *est intégrable, il existe* $\delta > 0$ *tel que*

$$\forall A \in \mathcal{F}, 0 < \mathbb{P}(A) \leq \delta \implies \mathbb{E}[|M_{K+1}|\mathbf{1}_A] \leq \varepsilon.$$

*Montrer à l'aide de la question 2 que pour* $C \geq \frac{\mathbb{E}[|M_{K+1}|]}{\delta}$*, on a*

$$\mathbb{E}[M_{\frac{\tau^{(N)}}{N}} \mathbf{1}_{\{|M_{\frac{\tau^{(N)}}{N}}|>C\}}] \leq \varepsilon.$$

*(On dit que la famille* $(M_{\frac{\tau^{(N)}}{N}}, N \in \mathbb{N}^*)$ *est uniformément intégrable.)*

5. *On rappelle que* $M_{\frac{\tau(N)}{N}}$ *converge presque sûrement vers* $M_\tau$. *Vérifier en utilisant le Lemme de Fatou que* $M_\tau$ *est intégrable. En déduire que l'on a* $\mathbb{E}[M_\tau \mathbf{1}_{\{|M_\tau|>C\}}] \leq \varepsilon$, *quitte éventuellement à prendre* $C$ *plus grand. Montrer que pour* $N$ *assez grand,* $\mathbb{E}[|M_{\frac{\tau(N)}{N}} - M_\tau|] \leq 3\varepsilon$ *puis que* $\mathbb{E}[|M_{\frac{\tau(N)}{N}} - M_\tau|] \underset{N\to+\infty}{\longrightarrow} 0$. *En déduire à partir de* (5.5) *que*
$$\mathbb{E}[M_\tau] = \mathbb{E}[M_0].$$
*(Indication : poser* $\phi_C(x) = x\mathbf{1}_{\{x\in[C,C]\}} + C\mathbf{1}_{\{x>C\}} - C\mathbf{1}_{\{x<-C\}}$ *et écrire* $|M_{\frac{\tau(N)}{N}} - M_\tau| \leq |M_{\frac{\tau(N)}{N}} - \phi_C(M_{\frac{\tau(N)}{N}})| + |\phi_C(M_{\frac{\tau(N)}{N}}) - \phi_C(M_\tau)| + |\phi_C(M_\tau) - M_\tau|$.*)*

Nous allons maintenant introduire les martingales browniennes usuelles.

**Proposition 5.4.3.** *Soit* $(W_t, t \geq 0)$ *un mouvement brownien réel et* $(\mathcal{F}_t, t \geq 0)$ *la filtration qu'il engendre. Alors,*

1. $(W_t^2 - t, t \geq 0)$ *est une* $(\mathcal{F}_t)$*-martingale,*
2. $(\exp(\lambda W_t - \frac{\lambda^2}{2}t), t \geq 0)$ *est une* $(\mathcal{F}_t)$*-martingale, quelque soit* $\lambda \in \mathbb{R}$ *(Martingales exponentielles).*

*Preuve.* Il est clair que $W_t^2 - t$ est intégrable pour tout $t \geq 0$, et pour $s \in [0,t]$, on a $W_t^2 = W_s^2 + 2W_s(W_t - W_s) + (W_t - W_s)^2$, si bien que l'on obtient :
$$\mathbb{E}[W_t^2|\mathcal{F}_s] = W_s^2 + (t-s).$$
On en déduit alors que $(W_t^2 - t, t \geq 0)$ est une $(\mathcal{F}_t)$-martingale.

De même, il est facile de voir que $\exp(\lambda W_t)$ est intégrable, et en écrivant $\exp(\lambda W_t) = \exp(\lambda W_s)\exp(\lambda(W_t - W_s))$, il vient que
$$\mathbb{E}[\exp(\lambda W_t)|\mathcal{F}_s] = \exp(\lambda W_s)\exp\left(\frac{\lambda^2}{2}(t-s)\right).$$
On obtient ainsi que $\mathbb{E}[\exp(\lambda W_t - \frac{\lambda^2}{2}t)|\mathcal{F}_s] = \exp(\lambda W_s - \frac{\lambda^2}{2}s)$. □

Une application importante des martingales exponentielles est qu'elles permettent de calculer les lois des temps d'atteinte du mouvement brownien. Pour $a > 0$, on pose
$$\tau_a = \inf\{t \geq 0, W_t \geq a\} = \inf\{t \geq 0, W_t = a\}.$$
Nous avons vu qu'il s'agit d'un temps d'arrêt par rapport à la filtration brownienne. Quelque soit $t > 0$, $\tau_a \wedge t$ est donc également un temps d'arrêt, évidemment borné par $t$. Le théorème d'arrêt appliqué aux Martingales exponentielles donne donc pour $\lambda \in \mathbb{R}$ :
$$\mathbb{E}[\exp(\lambda W_{\tau_a\wedge t} - \frac{\lambda^2}{2}\tau_a \wedge t)] = 1.$$

On se restreint alors à $\lambda > 0$. Puisque $W_{\tau_a \wedge t} \leq a$, il vient que $|\exp(\lambda W_{\tau_a \wedge t} - \frac{\lambda^2}{2}\tau_a \wedge t)| \leq \exp(\lambda a)$. En outre, $\exp(\lambda W_{\tau_a \wedge t} - \frac{\lambda^2}{2}\tau_a \wedge t) \underset{t\to+\infty}{\longrightarrow} \exp(\lambda a - \frac{\lambda^2}{2}\tau_a)\mathbf{1}_{\tau_a<\infty}$, et le théorème de convergence dominée assure donc que

$$\mathbb{E}\left[\exp\left(\lambda a - \frac{\lambda^2}{2}\tau_a\right)\mathbf{1}_{\tau_a<\infty}\right] = 1.$$

Puisque $|\exp\left(\lambda a - \frac{\lambda^2}{2}\tau_a\right)\mathbf{1}_{\tau_a<\infty}| \leq \exp(a)$ pour $\lambda \in (0,1]$ et $\exp\left(\lambda a - \frac{\lambda^2}{2}\tau_a\right)\mathbf{1}_{\tau_a<\infty} \underset{\lambda\to 0}{\longrightarrow} \mathbf{1}_{\tau_a<\infty}$, le théorème de convergence dominée assure que

$$\mathbb{P}(\tau_a < +\infty) = 1. \tag{5.6}$$

On pose alors $u = \lambda^2/2$. On obtient la loi de $\tau_a$ à travers sa transformée de Laplace :

$$\forall u > 0, \mathbb{E}[\exp(-u\tau_a)] = \exp(-\sqrt{2u}a).$$

Il s'agit d'une loi stable, et on peut montrer que sa densité est

$$\frac{a}{\sqrt{2\pi t^3}}\exp(-\frac{a^2}{2t})\mathbf{1}_{\{t>0\}}. \tag{5.7}$$

L'exercice qui suit prouve ce résultat.

**Exercice 5.4.4.** *Soit $\sigma_a$ une loi de densité $\frac{a}{\sqrt{2\pi t^3}}\exp(-\frac{a^2}{2t})\mathbf{1}_{\{t>0\}}$, nous voulons montrer que $\mathbb{E}[\exp(-u\sigma_a)] = \exp(-\sqrt{2u}a)$, pour $u > 0$ ce qui assurera que $\tau_a \overset{loi}{=} \sigma_a$.*

1. *En faisant un changement de variable convenable, montrer que $\mathbb{E}[\exp(-u\sigma_a)] = f(ua^2)$ où $f(x) = \mathbb{E}[\exp(-x/N^2)]$ avec $N \sim \mathcal{N}(0,1)$.*
2. *Montrer que $f'(x) = -\frac{1}{\sqrt{2x}}f(x)$, $f(0) = 1$ et en déduire le résultat. (Indication : faire le changement de variable $y \leftrightarrow \frac{\sqrt{2x}}{y}$.)*

**Exercice 5.4.5.** *Pour $a < 0$, on pose*

$$\tau_a = \inf\{t \geq 0, W_t \leq a\} = \inf\{t \geq 0, W_t = a\}.$$

1. *Montrer sans calcul que $\tau_a$ et $\tau_{-a}$ ont la même loi.*
2. *En reprenant la même démarche que pour $a > 0$, prouver directement que pour $a < 0$, $\mathbb{E}\left[\exp\left(-\frac{\lambda^2}{2}\tau_a\right)\right] = \exp(\lambda a)$ lorsque $\lambda < 0$.*

Pour conclure, nous résumons les résultats sur les temps d'atteinte dans la proposition suivante.

**Théorème 5.4.6.** *Soit $(W_t, t \geq 0)$ un mouvement brownien et pour $a \in \mathbb{R}$, on note $\tau_a = \inf\{t \geq 0, W_t = a\}$ le premier temps d'atteinte de $a$. Alors,*

$$\mathbb{P}(\tau_a < \infty) = 1 \textit{ et } \mathbb{E}[\exp(-u\tau_a)] = \exp(-\sqrt{2u}|a|).$$

**Exercice 5.4.7.** *Montrer en utilisant les transformées de Laplace que $\tau_a$ a même loi que $a^2\tau_1$. Retrouver ce résultat en utilisant la propriété de changement d'échelle du mouvement brownien à partir de l'expression $\tau_a = \inf\{t \geq 0, W_t \leq a\}$.*

**Exercice 5.4.8.** *On se donne un mouvement brownien et $\mathcal{F}_t$ la filtration qu'il engendre. L'objet de cet exercice est de montrer que $\sigma = \sup\{t \in [0,1], W_t \leq 0\}$ n'est pas un temps d'arrêt. On considère $t \in ]0,1[$.*

1. *Montrer que $\{\sigma < t\} = \{W_t > 0, \inf\{u \geq t, W_u \leq 0\} > 1\}$.*
2. *Montrer que $\tilde{W}_u = W_{t+u} - W_t$ est un mouvement brownien indépendant de $\mathcal{F}_t$, puis que $\{\sigma < t\} = \{W_t > 0, \inf\{u \geq 0, \tilde{W}_u \leq -W_t\} > 1 - t\}$.*
3. *On pose $\tau_a = \inf\{u \geq 0, \tilde{W}_u = a\}$ et $p(a) = \mathbb{P}(\tau_a > 1-t)$. Vérifier que pour tout $a \neq 0$, $0 < p(a) < 1$. Montrer que $\mathbb{E}[\mathbf{1}_{\{\sigma<t\}}|\mathcal{F}_t] = p(-W_t)\mathbf{1}_{\{W_t>0\}}$ et en déduire que $\{\sigma < t\} \notin \mathcal{F}_t$.*

**Exercice 5.4.9.** *Soit $(W_t, t \geq 0)$ un mouvement brownien et pour $a \geq 0$, $\tau_a = \inf\{t \geq 0, W_t = a\}$. Montrer que $(\tau_a, a \geq 0)$ est un processus à accroissements indépendants et stationnaires, c'est à dire que :*

– *pour $0 \leq a_1 < a_2 < \cdots < a_n$,*

$$\tau_{a_1}, \tau_{a_2} - \tau_{a_1}, \ldots, \tau_{a_n} - \tau_{a_{n-1}} \text{ sont indépendantes,}$$

– *(stationnarité) $\tau_{a+h} - \tau_a \stackrel{loi}{=} \tau_{b+h} - \tau_b$ pour $a, b, h \geq 0$. En d'autres termes, la loi de $\tau_{a+h} - \tau_a$ ne dépend pas de $a$.*

## 5.5 Etude des trajectoires du mouvement brownien

Dans cette section, nous allons étudier les trajectoires browniennes et établir notamment que ces trajectoires sont continues mais dérivables nulle part. Nous allons commencer par prouver que le mouvement brownien oscille, passe une infinité de fois par 0 lorsque $t \to +\infty$, et prend presque sûrement des valeurs arbitrairement grandes.

**Proposition 5.5.1.** *Soit $(W_t, t \geq 0)$ un mouvement brownien standard. Alors, on a p.s.*

$$\limsup_{t\to+\infty} W_t = +\infty \text{ et } \liminf_{t\to+\infty} W_t = -\infty.$$

*En particulier,* $\limsup_{t\to+\infty} |W_t| = +\infty$ *et* $\liminf_{t\to+\infty} |W_t| = 0$ *presque sûrement.*

*Preuve.* Rappelons que par définition, on a $\limsup_{t\to+\infty} W_t = \lim_{t\to+\infty} \sup_{u\geq t} W_u$ et de même $\liminf_{t\to+\infty} W_t = \lim_{t\to+\infty} \inf_{u\geq t} W_u$. Par symétrie du mouvement brownien, $(W_t, t \geq 0)$ a même loi que $(-W_t, t \geq 0)$. Par conséquent, $\liminf_{t\to+\infty} -W_t$ et $\liminf_{t\to+\infty} W_t$ ont même loi. Puisque $\liminf_{t\to+\infty} -W_t = -\limsup_{t\to+\infty} W_t$, on en déduit que $\limsup_{t\to+\infty} W_t$ a même loi que $-\liminf_{t\to+\infty} W_t$ et il est suffisant de prouver que $\limsup_{t\to+\infty} W_t = +\infty$ presque sûrement.

Nous allons en réalité prouver que $\sup_{u\geq t} W_u = +\infty$ p.s. ce qui prouvera le résultat. On écrit $\sup_{u\geq t} W_u = \sup_{u\geq 0}(W_{t+u} - W_t) + W_t$. D'après la proposition 5.2.4, $(W_{t+u} - W_t, u \geq 0)$ est un mouvement brownien indépendant de $W_t$. On pose pour $a \in \mathbb{R}$, $\tau_a = \inf\{u, W_{t+u} - W_t \geq a\}$ : $\tau_a$ est indépendant de $W_t$, et grâce au théorème 5.4.6, $\mathbb{P}(\tau_a < \infty) = 1$. Ainsi, $\mathbb{P}(\sup\limits_{u\geq 0}(W_{t+u} - W_t) \geq a) = 1$, et comme ceci est vrai pour tout $a > 0$, il vient que $\mathbb{P}(\sup\limits_{u\geq 0}(W_{t+u} - W_t) = +\infty) = 1$. Comme $W_t$ prend p.s. des valeurs finies, on en déduit que

$$\mathbb{P}(\sup_{u\geq t} W_u = +\infty) = 1.$$

Il est clair que $\limsup_{t\to+\infty} |W_t| = +\infty$ et il reste à montrer que $\liminf_{t\to+\infty} |W_t| = 0$ p.s. Puisque $\limsup_{t\to+\infty} W_t = +\infty$ et $\liminf_{t\to+\infty} W_t = -\infty$, on peut construire p.s. une suite $t_1 < \cdots < t_n < \dots$ d'instants tendant vers $+\infty$ tels que $W_{t_{2i}} \geq 1$ et $W_{t_{2i+1}} \leq -1$. Le mouvement brownien étant continu p.s., le théorème des valeurs intermédiaires assure que pour tout $i \in \mathbb{N}^*$, il existe $u_i \in (t_i, t_{i+1})$ tel que $W_{u_i} = 0$. On en déduit alors que $\liminf_{t\to+\infty} |W_t| = 0$ p.s. □

**Proposition 5.5.2.** *Si $(W_t, t \geq 0)$ est un mouvement brownien, alors pour $\alpha > 1/2$,*

$$\frac{1}{t^\alpha} \sup_{0\leq s\leq t} |W_s| \underset{t\to+\infty}{\to} 0, p.s.$$

*En particulier $\frac{W_t}{t} \underset{t\to+\infty}{\to} 0$, p.s. et $(tW_{1/t}, t \geq 0)$ est également un mouvement brownien.*

*preuve.* Nous allons commencer par prouver la convergence en probabilité vers 0 qui est plus facile à établir. Observons que $W_u^t = \frac{1}{t^\alpha} W_{ut^{2\alpha}}$ est un mouvement brownien grâce à la propriété de changement d'échelle. En outre, on a :

$$\frac{1}{t^\alpha} \sup_{0\leq s\leq t} |W_s| = \sup_{0\leq u\leq 1/t^{1-2\alpha}} |W_u^t|.$$

Par conséquent, pour tout $\varepsilon > 0$, $\mathbb{P}(\frac{1}{t^\alpha} \sup_{0\leq s\leq t} |W_s| > \varepsilon) = \mathbb{P}(\sup_{0\leq u\leq 1/t^{1-2\alpha}} |W_u| > \varepsilon)$. Or, puisque le mouvement brownien $(W_t, t \geq 0)$ est continu p.s. et tel que $W_0 = 0$, il vient que $\sup_{0\leq u\leq 1/t^{1-2\alpha}} |W_u|$ converge p.s. vers 0 lorsque $t \to +\infty$ (et donc en probabilité). Ainsi,

$$\mathbb{P}(\frac{1}{t^\alpha} \sup_{0\leq s\leq t} |W_s| > \varepsilon) \underset{t\to+\infty}{\to} 0.$$

La convergence p.s. est plus délicate à prouver : il s'agit de montrer que $\tilde{\Omega} = \{\forall \epsilon > 0, \exists T, \forall t \geq T, \frac{1}{t^\alpha} \sup_{0\leq s\leq t} |W_s| \leq \varepsilon\}$ est un événement de probabilité 1. On pose $t_n =$

$C^n$ pour $C > 1$. Il est en réalité suffisant de montrer que $\frac{1}{t_n^\alpha} \sup_{0\leq s\leq t_n} |W_s|$ converge p.s vers 0. En effet, si tel est le cas, p.s. pour tout $\varepsilon > 0$ il existe $N \in \mathbb{N}^*$ tel que $\frac{1}{t_n^\alpha} \sup_{0\leq s\leq t_n} |W_s| \leq \varepsilon/C^\alpha$ pour $n \geq N$. Ainsi, pour $t \geq C^N$, $t \in [t_n, t_{n+1})$ pour un certain $n \geq N$ et $\frac{1}{t^\alpha} \sup_{0\leq s\leq t} |W_s| \leq \frac{1}{C^{n\alpha}} \sup_{0\leq s\leq C^{n+1}} |W_s| \leq \varepsilon$.

On pose pour $\varepsilon > 0$, $S = \sum_{n=1}^\infty \mathbf{1}_{\{\frac{1}{t_n^\alpha} \sup_{0\leq s\leq t} |W_s| > \varepsilon\}}$. Nous allons montrer que $\mathbb{E}[S] < \infty$ ce qui assurera que $S$ est fini p.s. et donnera la convergence p.s. En prenant l'espérance et en utilisant la propriété d'échelle comme précédemment, il vient

$$\mathbb{E}[S] = \sum_{n=1}^\infty \mathbb{P}(\sup_{0\leq u\leq 1/t_n^{1-2\alpha}} |W_u| > \varepsilon).$$

On note $\tau_a = \inf\{t \geq 0, W_t = a\}$. On a

$$\{\sup_{0\leq u\leq 1/t_n^{1-2\alpha}} |W_u| > \varepsilon\} \subset \{\tau_\varepsilon \leq 1/t_n^{1-2\alpha}\} \cup \{\tau_{-\varepsilon} \leq 1/t_n^{1-2\alpha}\}.$$

Par conséquent, $\mathbb{P}(\sup_{0\leq u\leq 1/t_n^{1-2\alpha}} |W_u| > \varepsilon) \leq \mathbb{P}(\tau_\varepsilon \leq 1/t_n^{1-2\alpha}) + \mathbb{P}(\tau_{-\varepsilon} \leq 1/t_n^{1-2\alpha}) = 2\mathbb{P}(\tau_\varepsilon \leq 1/t_n^{1-2\alpha}) = 2\int_0^{t_n^{1-2\alpha}} \frac{\varepsilon}{2\pi t^3} e^{-\frac{\varepsilon^2}{2t}} dt$ car $\tau_\varepsilon$ et $\tau_{-\varepsilon}$ ont même loi. On pose $\varphi(t) = \frac{1}{t^3} e^{-\frac{\varepsilon^2}{2t}}$. On a $\varphi'(t) = \frac{1}{2t^5} e^{-\frac{\varepsilon^2}{2t}} (\varepsilon^2 - 6t)$, et $\varphi(t)$ atteint son maximum en $t = \frac{\varepsilon^2}{6}$ qui vaut : $\left(\frac{6}{\varepsilon^2}\right)^3 e^{-3}$. Par conséquent, $\mathbb{E}[S] \leq \sum_{n=1}^\infty \frac{\varepsilon}{\pi} \left(\frac{6}{\varepsilon^2}\right)^3 e^{-3} t_n^{1-2\alpha} < \infty$.

Puisque $\frac{1}{t} \sup_{0\leq s\leq t} |W_s| \underset{t\to+\infty}{\longrightarrow} 0, p.s.$ , la convergence p.s. de $W_t/t$ vers 0 est alors triviale, et assure que $tW_{1/t}$ tend p.s. vers 0 lorsque $t \to 0$. Le processus $(tW_{1/t}, t \geq 0)$ satisfait également les autres propriétés (2, 3 et 4 de la définition 5.1.1) du mouvement brownien, ce qui est laissé en exercice. □

Nous sommes désormais en mesure de prouver que le mouvement brownien est dérivable nulle part.

**Proposition 5.5.3.** *Soit $(W_t, t \geq 0)$ un mouvement brownien et $t \geq 0$. On a*

$$\limsup_{h\to 0} \left|\frac{W_{t+h} - W_t}{h}\right| = +\infty \ et \ \liminf_{h\to 0} \left|\frac{W_{t+h} - W_t}{h}\right| = 0 \ p.s.$$

*En particulier, le mouvement brownien n'est p.s. pas dérivable en $t$.*

*Preuve.* Le processus $\tilde{W}_u = W_{t+u} - W_t$ est un mouvement brownien, et grâce à la proposition précédente $\hat{W}_u = u\tilde{W}_{1/u}$ également. Or, $\left|\frac{W_{t+h}-W_t}{h}\right| = \left|\frac{\tilde{W}_h}{h}\right| = |\hat{W}_{1/h}|$. On conclut alors grâce à la proposition 5.5.1. □

## 5.6 Le processus de Poisson (exercice corrigé)

Cette section a pour objet de présenter le processus de Poisson, qui intervient dans la modélisation d'événements ponctuels, par exemple pour modéliser des files d'attente.

On se donne une suite de v.a. i.i.d. $\xi_1, \dots, \xi_n, \dots$, distribuées selon une loi exponentielle de paramètre $\lambda > 0$. Pour $k \in \mathbb{N}^*$, on pose $T_k = \sum_{j=1}^k \xi_j$, et on définit pour $t \geq 0$,

$$N_t = \sum_{k=1}^{\infty} \mathbf{1}_{\{T_k \leq t\}}.$$

Le processus $(N_t, t \geq 0)$ est appelé processus de Poisson d'intensité $\lambda$. On note $(\mathcal{F}_t, t \geq 0)$ la filtration qu'il engendre.

1. Montrer que $T_k \underset{k\to+\infty}{\longrightarrow} +\infty$, p.s. En déduire que $\forall t \geq 0, \mathbb{P}(N_t < +\infty) = 1$ puis que $\mathbb{P}(\forall t \geq 0, N_t < +\infty) = 1$.
2. Montrer que $N_t$ suit une loi de Poisson de paramètre $\lambda t$.
3. Calculer pour $t_1 < \cdots < t_n$ et $k_1, \dots, k_n \in \mathbb{N}$,
$$\mathbb{P}(N_{t_1} = k_1, N_{t_2} - N_{t_1} = k_2, \dots, N_{t_n} - N_{t_{n-1}} = k_n).$$
En déduire que $N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}}$ sont indépendantes et suivent une loi de Poisson dont on précisera le paramètre, puis que le processus de Poisson est un processus à accroissements stationnaires indépendants.
4. Montrer que $(N_t, t \in [0,T])$ et $(N_{T+t} - N_T, t \geq 0)$ sont indépendants et que $(N_{T+t} - N_T, t \geq 0)$ a même loi que $(N_t, t \geq 0)$.
5. Montrer que $(N_t - \lambda t, t \geq 0)$ est une $(\mathcal{F}_t)$-martingale.
6. Montrer que $\frac{\mathbb{P}(N_{T+t}-N_T=1)}{t} \underset{t\to 0}{\longrightarrow} \lambda$ et $\frac{\mathbb{P}(N_{T+t}-N_T=k)}{t} \underset{t\to 0}{\longrightarrow} 0$ pour $k \geq 2$. En déduire que pour toute fonction $f : \mathbb{R} \to \mathbb{R}$,
$$\frac{\mathbb{E}[f(N_{T+t}) - f(N_T)|\mathcal{F}_T]}{t} \underset{t\to 0}{\longrightarrow} \lambda(f(N_T+1) - f(N_T)).$$

**Solution :**

1. D'après la loi forte des grands nombres, $\frac{T_k}{k} \underset{k\to+\infty}{\longrightarrow} 1/\lambda$ p.s. et en particulier, $T_k \underset{k\to+\infty}{\longrightarrow} +\infty$ p.s. Par conséquent pour $t \geq 0$, l'événement $\{\omega, \exists N, \forall k \geq N, T_k \geq t\}$ est de probabilité 1, et donc $\mathbb{P}(N_t < \infty) = 1$. Enfin,
$$\Omega \backslash \{\forall t \geq 0, N_t < \infty\} = \{\exists t \geq 0, N_t = +\infty\} = \cup_{t>0}\{N_t = +\infty\} = \cup_{k\in\mathbb{N}}\{N_k = +\infty\},$$
la dernière égalité venant du fait que le processus de Poisson est un processus croissant. Par conséquent, il vient par $\sigma$-additivité
$$\mathbb{P}(\Omega \backslash \{\forall t \geq 0, N_t < \infty\}) = 0 \text{ puis } \mathbb{P}(\{\forall t \geq 0, N_t < \infty\}) = 1.$$

2. $\mathbb{P}(N_t = 0) = \mathbb{P}(\xi_1 > t) = \exp(-\lambda t)$.
$\mathbb{P}(N_t = k) = \mathbb{P}(T_k \leq t < T_{k+1}) = \int_0^\infty \dots \int_0^\infty \lambda^{k+1} \exp(-\lambda \sum_{i=1}^{k+1} x_i) \mathbf{1}_{\{\sum_{i=1}^{k} x_i \leq t < \sum_{i=1}^{k+1} x_i\}} dx_1 \dots dx_{k+1}$
On effectue le changement de variable
$$\begin{bmatrix} s_1 \\ \vdots \\ \vdots \\ s_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ 1 & \dots & \dots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ \vdots \\ x_{k+1} \end{bmatrix}$$
de Jacobien 1, et de domaine image $\{0 < s_1 < s_2 < \dots < s_{k+1}\}$. On obtient ainsi
$\mathbb{P}(N_t = k) = \int_{\{0<s_1<s_2<\dots<s_{k+1}\}} \lambda^{k+1} e^{-\lambda s_{k+1}} \mathbf{1}_{\{s_k \leq t \leq s_{k+1}\}} ds_1 \dots ds_{k+1} = \lambda^k e^{-\lambda t} \int_{\{0<s_1<s_2<\dots<s_k \leq t\}} ds_1 \dots ds_k = \frac{\lambda^k t^k}{k!} e^{-\lambda t}$. On a utilisé ici l'aire $\int_{\{0<s_1<s_2<\dots<s_k \leq t\}} ds_1 \dots ds_k = \frac{t^k}{k!}$.
3. On effectue le même calcul que précédemment :
$\mathbb{P}(N_{t_1} = k_1, N_{t_2} - N_{t_1} = k_2, \dots, N_{t_n} - N_{t_{n-1}} = k_n) = \mathbb{P}(T_{k_1} \leq t_1 < T_{k_1+1}, \dots, T_{k_1+\dots+k_n} \leq t_n < T_{k_1+\dots+k_n+1})$
$= \int_{\{0<s_1<\dots<s_{\sum_{i=1}^n k_i+1}\}} \lambda^{\sum_{i=1}^n k_i+1} e^{-\lambda s_{\sum_{i=1}^n k_i+1}} \mathbf{1}_{\{s_{k_1} \leq t_1 < s_{k_1+1}\}} \dots \mathbf{1}_{\{s_{\sum_{i=1}^n k_i} \leq t_n < s_{\sum_{i=1}^n k_i+1}\}} ds_1 \dots ds_{\sum_{i=1}^n k_i+1}$
$= \int_{\{0<s_1<s_2<\dots<s_{\sum_{i=1}^n k_i}<t_n\}} \lambda^{\sum_{i=1}^n k_i} e^{-\lambda t_n} \mathbf{1}_{\{s_{k_1} \leq t_1 < s_{k_1+1}\}} \dots \mathbf{1}_{\{s_{\sum_{i=1}^{n-1} k_i} \leq t_n < s_{\sum_{i=1}^{n-1} k_i+1}\}} ds_1 \dots ds_{\sum_{i=1}^n k_i}$
$= \lambda^{\sum_{i=1}^n k_i} e^{-\lambda t_n} \frac{t_1^{k_1}}{k_1!} \frac{(t_2-t_1)^{k_2}}{k_2!} \dots \frac{(t_n-t_{n-1})^{k_n}}{k_n!} = \frac{e^{-\lambda t_1} \lambda^{k_1} t_1^{k_1}}{k_1!} \frac{e^{-\lambda(t_2-t_1)} \lambda^{k_2} (t_2-t_1)^{k_2}}{k_2!} \dots \frac{e^{-\lambda(t_n-t_{n-1})} \lambda^{k_n} (t_n-t_{n-1})^{k_n}}{k_n!}$.
Les accroissements du processus de Poisson sont donc indépendants et suivent une loi de Poisson de paramètre $\lambda$ fois la taille de l'intervalle. En particulier, $N_{t+h} - N_t$ et $N_{s+h} - N_s$ ont même loi pour $s, t, h \geq 0$ : la loi des accroissements est donc stationnaire. Le processus de Poisson est donc à accroissements indépendants et stationnaires.
4. La preuve donnée pour le mouvement brownien (Proposition 5.2.4) s'adapte telle quelle.
5. La vérification de l'intégrabilité est immédiate. Grâce à la question précédente, on a pour $t > s$ $\mathbb{E}[N_t|\mathcal{F}_s] = N_s + \mathbb{E}[N_t - N_s|\mathcal{F}_s] = N_s + \lambda(t-s)$, ce qui donne la propriété de martingale.
6. Puisque $(N_{T+t} - N_T, t \geq 0)$ a même loi que $(N_t, t \geq 0)$, $\mathbb{P}(N_{T+t} - N_T = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$ et on obtient alors facilement les limites. En outre, par indépendance de $(N_{T+t} - N_T, t \geq 0)$ et $(N_t, t \in [0,T])$, on a que $\mathbb{E}[f(N_{T+t}) - f(N_T)|\mathcal{F}_T] = \psi(N_T)$ où
$$\psi(x) = \mathbb{E}[f(N_{T+t} - N_T + x) - f(x)] = \mathbb{E}[f(N_t + x) - f(x)] = e^{-\lambda t} \sum_{k=1}^\infty \frac{(\lambda t)^k}{k!} [f(x+k) - f(x)].$$
Le passage à la limite est alors immédiat.

# Chapitre 6

# Interprétation probabiliste d'EDP

Ce chapitre a pour objectif de donner une brève introduction au calcul stochastique afin de faire le lien entre probabilités et EDP. Il ne s'agit donc pas ici de présenter une construction rigoureuse et détaillée de l'intégrale stochastique, mais plutôt de donner une idée intuitive de cet objet. Surtout, nous voulons présenter les principaux résultats associés à cette intégrale, et notamment la formule d'Itô. Dans la première section, nous allons faire le lien entre le mouvement brownien et l'équation de la chaleur. Ensuite, afin de généraliser ce lien entre processus aléatoires et équations aux dérivées partielles, nous sommes amenés à donner quelques bases du calcul d'Itô. Cela nous permet de présenter les équations différentielles stochastiques, et leur lien naturel avec des EDP paraboliques (Feynman-Kac). Nous nous bornerons à étudier cela en dimension 1, bien que la plupart des résultats présentés s'étendent sans encombre à une dimension quelconque.

## 6.1 EDP de Feynman-Kac pour le mouvement brownien

Dans cette section, nous notons pour $t > 0$ et $x \in \mathbb{R}$, $p(t,x,y)$ la densité à l'instant $t$ du mouvement brownien issu de $x$, c'est à dire :

$$\forall y \in \mathbb{R},\ p(t,x,y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(y-x)^2}{2t}\right). \tag{6.1}$$

Nous avons la propriété remarquable suivante.

**Proposition 6.1.1.** *Pour $t > 0, x \in \mathbb{R}, y \in \mathbb{R}$, nous avons :*

$$\partial_t p(t,x,y) = \frac{1}{2}\partial_x^2 p(t,x,y).$$

*Preuve.* Il suffit de dériver et de vérifier ce résultat. Plus précisément, on a $\partial_t p(t,x,y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(y-x)^2}{2t}\right)\left[\frac{(y-x)^2}{2t^2} - \frac{1}{2t}\right]$. D'autre part, on a en dérivant deux fois par rapport à $x$, $\partial_x p(t,x,y) = \frac{y-x}{t}\frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(y-x)^2}{2t}\right)$ puis $\partial_x^2 p(t,x,y) = 2\partial_t p(t,x,y)$. □

Nous en déduisons le résultat suivant.

**Théorème 6.1.2.** *Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction réelle telle qu'il existe $\alpha \in (0,2)$ et $c, M > 0$ pour lesquels $\forall x \in \mathbb{R}$, $\frac{f(x)}{e^{c|x|^\alpha}} \leq M$. On pose $u(t,x) = \mathbb{E}[f(x+W_t)]$. Alors, $u$ est de classe $\mathcal{C}^\infty$ sur $\mathbb{R}^* \times \mathbb{R}$ et est solution de l'équation aux dérivées partielles (équation de la chaleur) :*

$$\begin{cases} \partial_t u(t,x) = \frac{1}{2}\partial_x^2 u(t,x) \\ u(0,x) = f(x). \end{cases}$$

*Pour $T > 0$, en posant $v(t,x) = \mathbb{E}[f(x+W_{T-t})]$, $v$ est $\mathcal{C}^\infty$ sur $[0,T) \times \mathbb{R}$, et résout l'EDP de Feynman-Kac :*

$$\begin{cases} \partial_t v(t,x) + \frac{1}{2}\partial_x^2 v(t,x) = 0 \\ v(T,x) = f(x). \end{cases}$$

*Preuve.* Pour $t > 0, x \in \mathbb{R}$, on a $u(t,x) = \int_{\mathbb{R}} f(y)p(t,x,y)dy$ et il s'agit de vérifier que $\partial_t u(t,x) = \int_{\mathbb{R}} f(y)\partial_t p(t,x,y)dy$ et $\partial_x^2 u(t,x) = \int_{\mathbb{R}} f(y)\partial_x^2 p(t,x,y)dy$ ce qui prouvera que $u$ satisfait l'EDP grâce à la proposition précédente.

Il est donc suffisant de justifier la dérivation sous le signe intégral à l'aide du théorème de convergence dominée. Pour la dérivation par rapport au temps, on a pour $s \in (t-\varepsilon, t+\varepsilon)$, $\left|\frac{p(s,x,y)-p(t,x,y)}{s-t}\right| \leq \sup_{s\in(t-\varepsilon,t+\varepsilon)} |\partial_t p(s,x,y)|$. En utilisant l'expression de $\partial_t p$, on obtient

$$\left|\frac{p(s,x,y)-p(t,x,y)}{s-t}\right| \leq \frac{1}{\sqrt{2\pi(t-\varepsilon)}} \exp\left(-\frac{(y-x)^2}{2(t+\varepsilon)}\right)\left[\frac{(y-x)^2}{2(t-\varepsilon)^2} + \frac{1}{2(t-\varepsilon)}\right]$$

et $|f(y)|\left|\frac{p(s,x,y)-p(t,x,y)}{s-t}\right|$ peut ainsi être dominée par une fonction intégrable en $y$. De façon tout à fait analogue, on justifie la dérivation par rapport à $x$. Enfin, le caractère $\mathcal{C}^\infty$ s'obtient de la même manière en justifiant par récurrence la dérivation sous le signe intégral. Enfin, puisque pour $t \in [0,T]$, $v(t,x) = u(T-t,x)$, on obtient immédiatement que $v$ résout l'EDP de Feynman-Kac. □

**Corollaire 6.1.3.** *Soient $\mu \in \mathbb{R}$ et $\sigma > 0$. Pour $f$ satisfaisant la même hypothèse de domination qu'au théorème précédent, on pose $u(t,x) = \mathbb{E}[f(x+\mu t+\sigma W_t)]$. Alors, $u$ est de classe $\mathcal{C}^\infty$ sur $\mathbb{R}^* \times \mathbb{R}$ et est solution de l'équation aux dérivées partielles :*

$$\begin{cases} \partial_t u(t,x) = \mu\partial_x u(t,x) + \frac{\sigma^2}{2}\partial_x^2 u(t,x) \\ u(0,x) = f(x). \end{cases}$$

*Pour $T > 0$, en posant $v(t,x) = \mathbb{E}[f(x + \mu(T-t) + \sigma(W_T - W_t))]$, $v$ est $\mathcal{C}^\infty$ sur $[0,T) \times \mathbb{R}$, et résout l'EDP de Feynman-Kac :*

$$\begin{cases} \partial_t v(t,x) + \mu \partial_x v(t,x) + \frac{\sigma^2}{2} \partial_x^2 v(t,x) = 0 \\ v(T,x) = f(x). \end{cases}$$

*Preuve.* Pour $t > 0, x \in \mathbb{R}$, on a $u(t,x) = \mathbb{E}[f(x + \mu t + W_{\sigma^2 t})]$ puisque $W_{\sigma^2 t}$ a même loi que $\sigma W_t$. En posant $\tilde{u}(t,x) = \mathbb{E}[f(x + W_t)]$, on voit que $u(t,x) = \tilde{u}(\sigma^2 t, x + \mu t)$. Grâce au théorème précédent, $\tilde{u}$ est de classe $\mathcal{C}^\infty$ et résout $\partial_t \tilde{u}(t,x) = \frac{1}{2}\partial_x^2 \tilde{u}(t,x)$ pour $t > 0$ et $x \in \mathbb{R}$. Ainsi, $\partial_t u(t,x) = \sigma^2 \partial_t \tilde{u}(\sigma^2 t, x + \mu t) + \mu \partial_x \tilde{u}(\sigma^2 t, x + \mu t) = \sigma^2 \partial_x^2 \tilde{u}(\sigma^2 t, x + \mu t) + \mu \partial_x \tilde{u}(\sigma^2 t, x + \mu t) = \sigma^2 \partial_x^2 u(t,x) + \mu \partial_x u(t,x)$ ce qui prouve le résultat pour $u$ et pour $v(t,x) = u(T-t,x)$. □

Nous déduisons de ces résultats l'EDP de Feynman-Kac pour le mouvement brownien géométrique. On pose pour $t < T$ et $x > 0$,

$$S_T^{t,x} = x \exp[(r - \frac{\sigma^2}{2})(T-t) + \sigma(W_T - W_t)]$$

et $v(t,x) = \mathbb{E}[e^{-r(T-t)} f(S_T^{t,x})]$. Il s'agit du modèle de Black-Scholes, et $S_T^{t,x}$ représente la valeur d'un actif risqué à la date $T$ sachant qu'il vaut $x$ à la date $t$. Ici $r$ désigne le taux d'intérêt et $\sigma$ la volatilité de l'actif. La fonction $f$ décrit le “payoff” d'une option d'échéance $T$, et $v$ représente le prix à la date $t$ de cette option sachant que le cours de l'actif vaut $x$. On pose $\tilde{f} = f \circ \exp$ et on a :

$$v(t,x) = e^{-r(T-t)} \mathbb{E}[\tilde{f}(\ln(x) + (r - \frac{\sigma^2}{2})(T-t) + \sigma W_{T-t})] = e^{-r(T-t)} \tilde{v}(t, \ln(x)),$$

où $\tilde{v}(t,\xi) = \mathbb{E}[\tilde{f}(\xi + (r - \frac{\sigma^2}{2})(T-t) + \sigma W_{T-t})]$. On suppose que $\tilde{f}$ satisfait l'hypothèse de domination du corollaire ci-dessus, si bien que pour $t \in [0,T), \xi \in \mathbb{R}$,

$$\partial_t \tilde{v}(t,\xi) + (r - \frac{\sigma^2}{2}) \partial_\xi \tilde{v}(t,\xi) + \frac{\sigma^2}{2} \partial_\xi^2 \tilde{v}(t,\xi) = 0.$$

On a $\tilde{v}(t,\xi) = e^{r(T-t)} v(t, e^\xi)$ et donc $\partial_t \tilde{v}(t,\xi) = e^{r(T-t)} \left(\partial_t v(t,e^\xi) - r v(t,e^\xi)\right)$, $\partial_\xi \tilde{v}(t,\xi) = e^{r(T-t)} e^\xi \partial_x v(t,e^\xi)$ et $\partial_\xi^2 \tilde{v}(t,\xi) = e^{r(T-t)}(e^\xi \partial_x v(t,e^\xi) + e^{2\xi} \partial_x^2 v(t,e^\xi))$. Par conséquent,

$$\forall t \in [0,T), \xi \in \mathbb{R},\ \partial_t v(t,e^\xi) - r v(t,e^\xi) + (r - \frac{\sigma^2}{2}) e^\xi \partial_x v(t,e^\xi) + \frac{\sigma^2}{2}(e^\xi \partial_x v(t,e^\xi) + e^{2\xi} \partial_x^2 v(t,e^\xi)) = 0$$

et donc avec $x = e^\xi$,

$$\begin{cases} \partial_t v(t,x) - r v(t,x) + r x \partial_x v(t,x) + \frac{\sigma^2}{2} x^2 \partial_x^2 v(t,x) = 0, t \in [0,T), x > 0 \\ v(T,x) = f(x). \end{cases} \tag{6.2}$$

## 6.2 Introduction au calcul d'Itô

Dans cette section nous allons présenter quelques rudiments du calcul d'Itô et de l'intégrale stochastique. Pour poser le problème, nous commençons par étudier la variation quadratique du mouvement brownien. Celle-ci est définie pour $t > 0$ par

$$\langle W \rangle_t = \lim_{n\to+\infty} \sum_{i=0}^{2^n-1} (W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t})^2.$$

Nous allons montrer que cette limite existe presque sûrement et vaut $t$. Nous commençons par montrer que la convergence a lieu dans $L^2$, et on regarde
$\mathbb{E}\left[\left(\sum_{i=0}^{2^n-1}(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t})^2 - t\right)^2\right]$
$= \mathbb{E}\left[\left(\sum_{i=0}^{2^n-1}\left\{(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t})^2 - (\frac{i+1}{2^n}t - \frac{i}{2^n}t)\right\}\right)^2\right]$
$= \sum_{i=0}^{2^n-1}\mathbb{E}\left[\left\{(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t})^2 - (\frac{i+1}{2^n}t - \frac{i}{2^n}t)\right\}^2\right]$
$= \sum_{i=0}^{2^n-1}\frac{t^2}{2^{2n}}\mathbb{E}[(G^2-1)^2]$ où $G \sim \mathcal{N}(0,1)$
$= \frac{t^2}{2^{n-1}}$ car $\mathbb{E}[G^4] = 3$ et par conséquent $\mathbb{E}[(G^2-1)^2] = 2$.

Maintenant, nous sommes en mesure de prouver la convergence presque sûre. Pour tout $\varepsilon > 0$, nous allons prouver qu'il existe p.s. un entier à partir duquel on a

$$\left|\sum_{i=0}^{2^n-1}(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t})^2 - t\right| \leq \varepsilon.$$

On pose

$$X_\varepsilon = \sum_{n\geq 1} \mathbf{1}_{\left\{\left|\sum_{i=0}^{2^n-1}(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t})^2 - t\right| > \varepsilon\right\}}.$$

Grâce à l'inégalité de Bienaymé-Tchebychev, on a

$$\mathbb{E}(X_\varepsilon) \leq \frac{1}{\varepsilon^2}\sum_{n\geq 1}\mathbb{E}\left[\left(\sum_{i=0}^{2^n-1}(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t})^2 - t\right)^2\right] < \infty,$$

et par conséquent $X_\varepsilon$ est finie p.s. ce qui prouve le résultat voulu. Nous résumons ce résultat dans la proposition suivante.

**Proposition 6.2.1.** *Soit $(W_t, t \geq 0)$ un mouvement brownien. Pour $t > 0$, la série $\sum_{i=0}^{2^n-1}(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t})^2$ converge presque sûrement vers $t$. On appelle cette limite la variation quadratique jusqu'à l'instant $t$ du mouvement brownien et on note*

$$\langle W \rangle_t = t.$$

**Exercice 6.2.2.** *Soit $(H_t, t \geq 0)$ un processus continu adapté par rapport à la filtration engendrée par $(W_t, t \geq 0)$ et tel que : $\forall t > 0, \sup_{s\in[0,t]} \mathbb{E}[H_s^2] < \infty$. Montrer que $\frac{t}{2^n}\sum_{i=0}^{2^n-1} H^2_{\frac{i}{2^n}t}$ converge p.s. vers $\int_0^t H_s^2 ds$. En adaptant la preuve de la proposition précédente, montrer que $\sum_{i=0}^{2^n-1} H^2_{\frac{i}{2^n}t}(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t})^2 - \frac{t}{2^n}\sum_{i=0}^{2^n-1} H^2_{\frac{i}{2^n}t}$ converge p.s vers 0. En déduire que*

$$\sum_{i=0}^{2^n-1} H^2_{\frac{i}{2^n}t}(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t})^2 \underset{n\to+\infty}{\longrightarrow} \int_0^t H_s^2 ds \text{ p.s.}$$

**Exercice 6.2.3.** *Soit $f : \mathbb{R}_+ \to \mathbb{R}$. On fixe $t > 0$.*

1. *On suppose que $f$ est lipschitzienne sur $[0,t]$. Montrer que*

$$\lim_{n\to+\infty} \sum_{i=0}^{2^n-1} (f(t\frac{i+1}{2^n}) - f(t\frac{i}{2^n}))^2 = 0.$$

   *En déduire que p.s, les trajectoires du mouvement brownien ne sont pas lipschitziennes (et en particulier $\mathcal{C}^1$).*

2. *On suppose que $f$ est monotone. On note pour $s \in (0,t)$, $f(s-) = \lim_{u\uparrow t} f(u)$ et $f(s+) = \lim_{u\downarrow t} f(u)$ Montrer que l'ensemble des discontinuité de $f$, $Disc_{[0,t]}(f) = \{s \in (0,t), f(s-) \neq f(s+)\}$ est dénombrable, puis que*

$$\lim_{n\to+\infty} \sum_{i=0}^{2^n-1} (f(t\frac{i+1}{2^n}) - f(t\frac{i}{2^n}))^2 = \sum_{s\in Disc_{[0,t]}(f)} (f(s+) - f(s-))^2.$$

Maintenant nous allons définir l'intégrale stochastique d'un processus par rapport à un mouvement brownien $(W_t, t \geq 0)$. On note $(\mathcal{F}_t, t \geq 0)$ la filtration engendrée par le mouvement brownien. Nous admettrons le résultat suivant.

**Théorème 6.2.4.** *On se donne un processus $(H_t, t \geq 0)$ qui est supposé $(\mathcal{F}_t)$-adapté et tel que*

$$\forall t > 0, \mathbb{E}\left[\int_0^t H_s^2 ds\right] < \infty.$$

*Alors, pour tout $t > 0$, les sommes $\sum_{i=0}^{2^n-1} H_{\frac{i}{2^n}t}(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t})$ convergent dans $L^2$, et on définit alors l'intégrale stochastique $\int_0^t H_s dW_s$ comme cette limite. L'intégrale stochastique $(\int_0^t H_s dW_s, s \geq 0)$ est une martingale continue et satisfait la propriété d'isométrie*

$$\mathbb{E}\left[\left(\int_0^t H_s dW_s\right)^2\right] = \mathbb{E}\left[\int_0^t H_s^2 ds\right].$$

La propriété d'isométrie peut "se voir" sur les sommes finies. En effet, on a par linéarité :

$$\mathbb{E}\left[\left(\sum_{i=0}^{2^n-1} H_{\frac{i}{2^n}t}(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t})\right)^2\right] = \sum_{i=0}^{2^n-1}\sum_{j=0}^{2^n-1} \mathbb{E}\left[H_{\frac{i}{2^n}t}H_{\frac{j}{2^n}t}(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t})(W_{\frac{j+1}{2^n}t} - W_{\frac{j}{2^n}t})\right].$$

Si $i \neq j$, par exemple si $i < j$, on a en conditionnant par rapport à $\mathcal{F}_{t_j}$, puisque le processus $(H_t, t \geq 0)$ est $(\mathcal{F}_t)$-adapté : $\mathbb{E}\left[H_{\frac{i}{2^n}t}H_{\frac{j}{2^n}t}(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t})(W_{\frac{j+1}{2^n}t} - W_{\frac{j}{2^n}t})|\mathcal{F}_{\frac{j}{2^n}t}\right] = H_{\frac{i}{2^n}t}H_{\frac{j}{2^n}t}(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t})\mathbb{E}\left[(W_{\frac{j+1}{2^n}t} - W_{\frac{j}{2^n}t})|\mathcal{F}_{\frac{j}{2^n}t}\right] = 0$. Et pour $i = j$, $\mathbb{E}\left[H^2_{\frac{i}{2^n}t}(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t})^2\right] = \mathbb{E}\left[H^2_{\frac{i}{2^n}t}\mathbb{E}\left[(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t})^2|\mathcal{F}_{\frac{i}{2^n}t}\right]\right] = \frac{t}{2^n}\mathbb{E}\left[H^2_{\frac{i}{2^n}t}\right]$. Par conséquent, il vient que

$$\mathbb{E}\left[\left(\sum_{i=0}^{2^n-1} H_{\frac{i}{2^n}t}(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t})\right)^2\right] = \frac{t}{2^n}\sum_{i=0}^{2^n-1}\mathbb{E}\left[H^2_{\frac{i}{2^n}t}\right],$$

et la propriété d'isométrie en découle grâce à la convergence $L^2$ de la somme vers l'intégrale stochastique.

De même, la propriété de martingale se comprend aisément à partir des sommes finies dans la mesure où l'on a pour $s < t$,

$$\mathbb{E}[\sum_{i=0}^{2^n-1} H_{\frac{i}{2^n}t}(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t})|\mathcal{F}_s] = \sum_{i=0}^{\lfloor 2^n\frac{s}{t}\rfloor-1} H_{\frac{i}{2^n}t}(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t}) + H_{\frac{\lfloor 2^n\frac{s}{t}\rfloor}{2^n}}(W_s - W_{\frac{\lfloor 2^n\frac{s}{t}\rfloor}{2^n}}).$$

Le terme de gauche converge vers $\mathbb{E}\left[\int_0^t H_u dW_u|\mathcal{F}_s\right]$ et la somme de droite converge vers $\int_0^s H_u dW_u$.

Nous allons désormais vérifier que l'intégrale stochastique est une martingale sur un cas particulier : celui où $H_t = W_t$. Nous avons $\mathbb{E}[\int_0^t W_s^2 ds] = \int_0^t s ds = t^2/2$, et l'hypothèse d'intégrabilité requise sur le processus $(H_t, t \geq 0)$ est satisfaite. Par ailleurs, nous avons : $W_t^2 - W_0^2 = \sum_{i=0}^{2^n-1}(W^2_{\frac{i+1}{2^n}t} - W^2_{\frac{i}{2^n}t}) = \sum_{i=0}^{2^n-1}(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t})(W_{\frac{i+1}{2^n}t} + W_{\frac{i}{2^n}t}) = 2\sum_{i=0}^{2^n-1} W_{\frac{i}{2^n}t}(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t}) + \sum_{i=0}^{2^n-1}(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t})^2$ Ainsi, puisque la variation quadratique du mouvement brownien converge dans $L^2$ (et p.s.) vers $t$,

$$\sum_{i=0}^{2^n-1} W_{\frac{i}{2^n}t}(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t}) \underset{n\to+\infty}{\to} \frac{1}{2}(W_t^2 - t) \text{ dans } L^2,$$

et donc $\int_0^t W_s dW_s = \frac{1}{2}(W_t^2 - t)$. Il s'agit bien d'une $(\mathcal{F}_t)$-martingale continue.

Maintenant, observons la relation que nous venons d'établir dans le cas du mouvement brownien :

$$t \geq 0,\ W_t^2 = 2\int_0^t W_s dW_s + t.$$

Si le calcul stochastique était un calcul intégral "usuel", nous aurions du avoir simplement $W_t^2 = 2\int_0^t W_s dW_s$. Ses règles de calculs sont donc différentes. Précisons cela. Pour une fonction $f : \mathbb{R} \to \mathbb{R}$ de classe $\mathcal{C}^2$, on a $f(W_t) = f(W_0) + \sum_{i=0}^{2^n-1} f(W_{\frac{i+1}{2^n}t}) - f(W_{\frac{i}{2^n}t}) = f(W_0) + \sum_{i=0}^{2^n-1} f'(W_{\frac{i}{2^n}t})(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t}) + \frac{1}{2}\sum_{i=0}^{2^n-1} f''(W_{\frac{i}{2^n}t})(W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t})^2 + \sum_{i=0}^{2^n-1} o((W_{\frac{i+1}{2^n}t} - W_{\frac{i}{2^n}t})^2)$. En passant à la limite $n \to +\infty$, la première somme converge vers $\int_0^t f'(W_s)dW_s$, la seconde vers $\frac{1}{2}\int_0^t f''(W_s)ds$ (analogue à l'exercice 6.2.2), et on peut montrer que la dernière converge vers 0. Ainsi, nous avons la formule d'Itô pour le mouvement brownien :

$$f(W_t) = f(W_0) + \int_0^t f'(W_s)dW_s + \frac{1}{2}\int_0^t f''(W_s)ds.$$

Nous allons maintenant généraliser cette formule pour des processus appelés processus d'Itô. Un processus $(X_t, t \geq 0)$ est appelé processus d'Itô s'il existe deux processus adaptés $(G_t, t \geq 0)$ et $(H_t, t \geq 0)$ tels que

$$\forall t \geq 0, X_t = X_0 + \int_0^t G_s ds + \int_0^t H_s dW_s. \tag{6.3}$$

et $\forall t > 0$, $\int_0^t |G_s|ds < \infty$ et $\int_0^t H_s^2 ds < \infty$ p.s. Nous admettrons que sous cette hypothèse plus faible, l'intégrale stochastique $\int_0^t H_s dW_s$ reste bien définie. De même, nous admettrons que la variation quadratique d'un processus d'Itô, définie comme la limite de la somme des accroissements au carré existe, et vaut

$$\langle X\rangle_t = \int_0^t H_t^2 d\langle W\rangle_t = \int_0^t H_t^2 dt.$$

Intuivement, cela découle du fait que :

$$[X_{t+\varepsilon} - X_t]^2 \approx [H_t(W_{t+\varepsilon} - W_t) + G_t\varepsilon]^2 \approx H_t^2(W_{t+\varepsilon} - W_t)^2 \approx H_t^2(\langle W\rangle_{t+\varepsilon} - \langle W\rangle_t),$$

au premier ordre en $\varepsilon$.

Souvent, on utilise la notation plus concise $dX_t = G_t dt + H_t dW_t$ pour désigner le processus d'Itô (6.3). Cette notation est souvent plus légère pour mener à bien les calculs, néanmoins elle n'a aucune signification propre et ne doit être vue que comme une notation condensée de (6.3).

Nous avons le résultat suivant.

**Théorème 6.2.5.** *Soient $X_t = X_0 + \int_0^t G_s ds + \int_0^t H_s dW_s$ un processus d'Itô et $f : \mathbb{R} \to \mathbb{R}$ une fonction $\mathcal{C}^2$. Alors $f(X_t)$ est un processus d'Itô et satisfait :*

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)G_s ds + \int_0^t f'(X_s)H_s dW_s + \frac{1}{2}\int_0^t f''(X_s)H_s^2 ds,$$

*ou de manière condensée* $df(X_t) = f'(X_t)dX_t + \frac{1}{2}f''(X_t)d\langle X\rangle_t$. *Lorsque* $f$ *dépend également du temps et est continûment dérivable en temps, on a*

$$\begin{aligned} f(t,X_t) &= f(0,X_0) + \int_0^t \partial_t f(s,X_s)ds + \int_0^t \partial_x f(s,X_s)G_s ds \\ &\quad + \int_0^t \partial_x f(s,X_s)H_s dW_s + \frac{1}{2}\int_0^t \partial_x^2 f(s,X_s)H_s^2 ds, \end{aligned}$$

*ou de façon concise* $df(t,X_t) = \partial_t f(t,X_t)dt + \partial_x f(t,X_t)dX_t + \frac{1}{2}\partial_x^2 f(t,X_t)d\langle X\rangle_t$.

## 6.3 Equations différentielles stochastiques et EDP

Nous sommes désormais en mesure de définir les équations différentielles stochastiques. Nous nous plaçons sur une espace de probabilité $(\Omega,\mathcal{F},\mathbb{P})$ muni d'une filtration $(\mathcal{F}_t, t\geq 0)$ et de $(W_t, t\geq 0)$, un $(\mathcal{F}_t)$-mouvement brownien. Nous admettons le résultat suivant.

**Théorème 6.3.1.** *Soit* $X_0$ *une condition initiale* $\mathcal{F}_0$*-mesurable de carré intégrable. Soient* $b : \mathbb{R}_+\times\mathbb{R}\to\mathbb{R}$ *et* $\sigma : \mathbb{R}_+\times\mathbb{R}\to\mathbb{R}_+$ *deux fonctions continues telles que il existe* $K>0$, $|b(t,x)-b(t,y)|+|\sigma(t,x)-\sigma(t,y)|\leq K|x-y|$ *et* $|b(t,0)|+|\sigma(t,0)|\leq K$ *pour tout* $t\geq 0, x,y\in\mathbb{R}$. *Alors, il existe un unique processus* $(X_t, t\geq 0)$ $(\mathcal{F}_t)$*-adapté tel que :*

$$X_t = X_0 + \int_0^t b(s,X_s)ds + \int_0^t \sigma(s,X_s)dW_s, \tag{6.4}$$

*c'est à dire que tout autre processus* $(\tilde{X}_t, t\geq 0)$ *satisfaisant cette équation différentielle stochastique est tel que* $\mathbb{P}(\forall t\geq 0, X_t = \tilde{X}_t) = 1$ *(on dit que* $X$ *et* $\tilde{X}$ *sont indistingables). En outre, les trajectoires de* $X$ *sont p.s. continues, et on a :* $\forall T>0$, $\mathbb{E}[\sup_{t\in[0,T]} X_s^2] < \infty$.

Intuitivement, si l'on se donne une grille de discrétisation sur $[0,T]$, $0 = t_0^n < t_1^n < \cdots < t_n^n = T$ telle que $\max_{i=0,\dots,n-1} t_{i+1}^n - t_i^n \underset{n\to+\infty}{\to} 0$ (cette condition est satisfaite pour la grille régulière $t_i^n = iT/n$), la solution $(X_t, t\in[0,T])$ de l'EDS (6.4) est la limite en un sens que nous ne précisons pas lorsque $n\to+\infty$ des processus continus $(\hat{X}_t^n, t\in[0,T])$ définis par $\hat{X}_0^n = X_0$ et

$$\hat{X}_t^n = \hat{X}_{t_i^n}^n + b(t_i^n, \hat{X}_{t_i^n}^n)(t-t_i^n) + \sigma(t_i^n, \hat{X}_{t_i^n}^n)(W_t - W_{t_i^n}) \text{ pour } t\in(t_i^n, t_{i+1}^n] \tag{6.5}$$

D'ailleurs, pour simuler en pratique une EDS, la méthode usuelle consiste à approcher la solution exacte par $\hat{X}_t^n$, et le schéma de discrétisation (6.5) s'appelle schéma d'Euler-Maruyama.

Donnons deux exemples où on est capable de résoudre explicitement une EDS. Le premier est lorsque $b(t,x) = \mu$ et $\sigma(t,x) = \sigma$ sont des fonctions constantes. Il est alors immédiat de voir que $x + \mu t + \sigma W_t$ est la seule solution de l'EDS de condition initiale $X_0 = x$. Un second exemple plus intéressant est le cas où $b(t, x) = \mu x$ et $\sigma(t, x) = \sigma x$. Grâce au théorème précédent, nous savons que pour $x > 0$, il existe une unique solution de

$$X_t = x + \int_0^t \mu X_s ds + \int_0^t \sigma X_s dW_s.$$

Alors, grâce à la formule d'Itô, $Y_t = \ln(X_t)$ satisfait

$$Y_t = \ln(x) + \int_0^t (\mu - \frac{\sigma^2}{2}) ds + \int_0^t \sigma dW_s,$$

et par conséquent $Y_t = \ln(x) + (\mu - \frac{\sigma^2}{2})t + \sigma W_t$. On en déduit alors que :

$$X_t = x \exp\left( (\mu - \frac{\sigma^2}{2})t + \sigma W_t \right).$$

Il s'agit du mouvement brownien géométrique. Pour $\mu = 0$, nous reconnaissons une martingale exponentielle du mouvement brownien et vérifions à nouveau sur ce cas particulier qu'une intégrale stochastique est une $(\mathcal{F}_t)$-martingale.

Revenons sur le théorème d'existence et d'unicité de la solution d'une EDS. Puisque la solution trajectorielle de $(X_t, t \geq 0)$ est unique, ce résultat nous dit intuitivement que $X_t$ est une fonction mesurable de la condition initiale $X_0$ et de la trajectoire brownienne $(W_s, s \in [0, t])$. On note $X_t = F_t(X_0, W_s, s \in [0, t])$. Ainsi, si $\tilde{X}_0$ est une condition initiale $\mathcal{F}_0$-mesurable de même loi que $X_0$ et $(\tilde{W}_t, t \geq 0)$ un autre mouvement brownien, le processus $\tilde{X}_t = F_t(\tilde{X}_0, \tilde{W}_s, s \in [0, t])$ est la solution de l'EDS de coefficient $b$ et $\sigma$ associée à la condition initiale $\tilde{X}_0$ et au mouvement brownien $(\tilde{W}_s, s \in [0, t])$. Il a même loi que le processus $(X_t, t \geq 0)$. En effet, si $0 \leq t_1 < \cdots < t_n$, et $f : \mathbb{R}^n \to \mathbb{R}$ est une fonction mesurable,
$\mathbb{E}[f(X_{t_1}, \ldots, X_{t_n})] = \mathbb{E}[f(F_{t_1}(X_0, W_s, s \in [0, t_1]), \ldots, F_{t_n}(X_0, W_s, s \in [0, t_n]))]$
$= \mathbb{E}[f(F_{t_1}(\tilde{X}_0, \tilde{W}_s, s \in [0, t_1]), \ldots, F_{t_n}(\tilde{X}_0, \tilde{W}_s, s \in [0, t_n]))] = \mathbb{E}[f(\tilde{X}_{t_1}, \ldots, \tilde{X}_{t_n})]$. En outre, puisque $\mathcal{F}_0$ est indépendant de $(W_t, t \geq 0)$, $\mathbb{E}[f(X_{t_1}, \ldots, X_{t_n}) | \mathcal{F}_0] = \psi(X_0)$ où $\psi(x) = \mathbb{E}[f(X^x_{t_1}, \ldots, X^x_{t_n})]$, $X^x_t$ désignant la solution de condition initiale $x \in \mathbb{R}$.

Nous venons de donner les arguments qui permettent de prouver que l'unicité trajectorielle obtenue au théorème précédent impliquent l'unicité faible des solutions, c'est à dire l'unicité en loi. Nous résumons ce résultat dans la proposition suivante.

**Proposition 6.3.2.** *Sous les hypothèses du théorème précédent, si $X_0$ et $\tilde{X}_0$ sont deux conditions initiales $\mathcal{F}_0$-mesurables de même loi et $(W_t, t \geq 0)$ et $(\tilde{W}_t, t \geq 0)$ deux $(\mathcal{F}_t)$-mouvements browniens, les processus $(X_t, t \geq 0)$ et $(\tilde{X}_t, t \geq 0)$ respectivement solutions des EDS*

$$X_t = X_0 + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dW_s \text{ et } \tilde{X}_t = \tilde{X}_0 + \int_0^t b(s, \tilde{X}_s) ds + \int_0^t \sigma(s, \tilde{X}_s) d\tilde{W}_s$$

*ont même loi. Leur loi est donc entièrement caractérisée par leur loi initiale et plus précisément, si $X_t^x$ désigne une solution de condition initiale $x \in \mathbb{R}$,*

$$\mathbb{E}[f(X_{t_1}, \ldots, X_{t_n})|\mathcal{F}_0] = \psi(X_0)$$

*où $\psi(x) = \mathbb{E}[f(X_{t_1}^x, \ldots, X_{t_n}^x)]$.*

Nous sommes désormais en mesure de prouver la propriété de Markov pour les solutions des EDS.

**Proposition 6.3.3.** *On se donne $b$ et $\sigma$ deux coefficients satisfaisant les conditions du théorème précédent et $(X_t, t \geq 0)$ la solution de l'EDS (6.4). Alors $(X_t, t \geq 0)$ satisfait la propriété de Markov et pour tout $T > 0$, $(X_{T+t}, t \geq 0)$ est la solution de l'EDS issue de $X_T$ et de coefficients $b(T+t, x)$ et $\sigma(T+t, x)$ associée au mouvement brownien $(W_{T+t} - W_T, t \geq 0)$.*

*Preuve.* Pour $T > 0$ et $t \geq 0$, on a en posant pour $s \geq 0$, $W_s^T = W_{T+s} - W_T$ :
$X_{T+t} = X_0 + \int_0^{T+t} b(s, X_s)ds + \int_0^{T+t} \sigma(s, X_s)dW_s$
$= X_T + \int_T^{T+t} b(s, X_s)ds + \int_T^{T+t} \sigma(s, X_s)dW_s$
$= X_T + \int_0^t b(T+s, X_{T+s})ds + \int_0^t \sigma(T+s, X_{T+s})dW_s^T$. On obtient ainsi que $(X_{T+t}, t \geq 0)$ est la solution de l'EDS issue de $X_T$ et de coefficients $b(T+t, x)$ et $\sigma(T+t, x)$ associée au mouvement brownien $(W_{T+t} - W_T, t \geq 0)$. Il est à remarquer que la filtration de référence pour cette EDS est $(\mathcal{F}_{T+t}, t \geq 0)$. Grâce à la proposition précédente, $\mathbb{E}[f(X_{T+t_1}, \ldots, X_{T+t_n})|\mathcal{F}_T] = \psi(X_T)$. Puisque les tribus engendrées par $X_T$ et $(X_t, t \in [0, T])$ sont incluses dans $\mathcal{F}_T$, il vient que

$$\mathbb{E}[f(X_{T+t_1}, \ldots, X_{T+t_n})|X_t, t \in [0, T]] = \mathbb{E}[f(X_{T+t_1}, \ldots, X_{T+t_n})|X_T],$$

ces espérances conditionnelles étant toutes deux égales à $\psi(X_T)$. Cela prouve que $(X_t, t \geq 0)$ satisfait la propriété de Markov. □

Nous allons maintenant établir la formule de Feynman-Kac. On se donne deux coefficients $b$ et $\sigma$ satisfaisant les conditions du théorème d'existence et d'unicité de la solution de l'EDS (6.4). On note $(X_s^{t,x}, s \geq t)$ la solution de

$$s \geq t, X_s^{t,x} = x + \int_t^s b(u, X_u^{t,x})du + \int_t^s \sigma(u, X_u^{t,x})dW_u.$$

On se donne une fonction $f$ régulière, un horizon $T > 0$, et on pose

$$t \in [0, T],\ v(t, x) = \mathbb{E}[f(X_T^{t,x})].$$

Grâce au résultat d'unicité des EDS, nous avons pour $t \leq u \leq T$,

$$X_T^{t,x} = X_T^{u, X_u^{t,x}} \ p.s,$$

et donc

$$v(t,x) = \mathbb{E}[f(X_T^{u,X_u^{t,x}})] = \mathbb{E}[\mathbb{E}[f(X_T^{u,X_u^{t,x}})|X_u^{t,x}]] = \mathbb{E}[v(u,X_u^{t,x})]. \tag{6.6}$$

Sous des conditions de régularités sur $f$ et sur $b$ et $\sigma$ que nous ne précisons pas, la fonction $v$ est suffisament dérivable pour lui appliquer la formule d'Itô. Il vient alors : $v(u,X_u^{t,x}) = v(t,x) + \int_t^u \partial_t v(s,X_s^{t,x})ds + \int_t^u \partial_x v(s,X_s^{t,x})\mu(s,X_s^{t,x})ds + \int_t^u \partial_x v(s,X_s^{t,x})\sigma(s,X_s^{t,x})dW_s + \frac{1}{2}\int_t^u \partial_x^2 v(s,X_s^{t,x})\sigma^2(s,X_s^{t,x})ds$. Par conséquent, pour $u = t+\delta t$ avec $\delta t \downarrow 0$, il vient : $v(t+\delta t, X_{t+\delta t}^{t,x}) \approx v(t,x) + \partial_t v(t,x)\delta t + \partial_x v(t,x)\mu(t,x)\delta t + \partial_x v(t,x)\sigma(t,x)(W_{t+\delta t} - W_t) + \frac{1}{2}\partial_x^2 v(t,x)\sigma^2(t,x)\delta t$, puis

$$\mathbb{E}[v(t+\delta t, X_{t+\delta t}^{t,x})] \approx v(t,x) + \delta t\left[\partial_t v(t,x) + \partial_x v(t,x)\mu(t,x) + \frac{1}{2}\partial_x^2 v(t,x)\sigma^2(t,x)\right].$$

En combinant avec (6.6), il vient que $v$ résout l'EDP suivante :

$$\begin{cases} \partial_t v(t,x) + \partial_x v(t,x)\mu(t,x) + \frac{1}{2}\partial_x^2 v(t,x)\sigma^2(t,x) = 0, \ t \in (0,T), x \in \mathbb{R} \\ v(T,x) = f(x) \end{cases} \tag{6.7}$$

Le corollaire 6.1.3 apparaît ainsi comme un cas particulier de ce résultat général.

**Exercice 6.3.4.** *Soit $r > 0$. Montrer que $\tilde{v}(t,x) = \mathbb{E}[e^{-r(T-t)}f(X_T^{t,x})]$ est solution de l'EDP :*

$$\begin{cases} \partial_t \tilde{v}(t,x) - r\tilde{v}(t,x) + \partial_x \tilde{v}(t,x)\mu(t,x) + \frac{1}{2}\partial_x^2 \tilde{v}(t,x)\sigma^2(t,x) = 0, \ t \in (0,T), x \in \mathbb{R} \\ \tilde{v}(T,x) = f(x) \end{cases}$$

*et retrouver l'équation* (6.2) *en spécifiant $\mu$ et $\sigma$.*

Nous allons maintenant établir un résultat d'unicité de l'EDP (6.7). Plus précisément, nous allons montrer que si $v(t,x)$ est une solution de (6.7) continûment dérivable en temps et deux fois continûment dérivable en espace, alors $v(t,x) = \mathbb{E}[f(X_T^{t,x})]$. Pour cela, il suffit d'appliquer la formule d'Itô à $v(s,X_s^{t,x})$ :

$$s \geq t, \ v(s,X_s^{t,x}) = v(t,x) + \int_t^s \partial_x v(u,X_u^{t,x})dW_u.$$

L'intégrale stochastique étant une martingale, on obtient que $\mathbb{E}[v(s,X_s^{t,x})] = v(t,x)$. En particulier pour $s = T$, $v(T,X_T^{t,x}) = f(X_T^{t,x})$ et donc $v(t,x) = \mathbb{E}[f(X_T^{t,x})]$, ce qui donne le résultat d'unicité.

Maintenant nous présentons succintement les équations de Kolmogorov en prenant le point de vue des distributions. Nous supposons pour cela que $X_T^{t,x}$ est une v.a. à densité $p(t,x,T,y)$ régulière, ce qui est le cas dès lors que $\forall t,x, \sigma(t,x) \geq \sigma_0 > 0$. Par conséquent, $v(t,x) = \int_{\mathbb{R}} f(y)p(t,x,T,y)dy$, et l'EDP de Feynman-Kac (6.7) peut s'écrire :

$$\int_{\mathbb{R}} f(y)[\partial_t p(t,x,T,y) + \partial_x p(t,x,T,y)\mu(t,x) + \frac{1}{2}\partial_x^2 p(t,x,T,y)\sigma^2(t,x)]dy = 0$$

et donc

$$\begin{cases} \partial_t p(t,x,T,y) + \partial_x p(t,x,T,y)\mu(t,x) + \frac{1}{2}\partial_x^2 p(t,x,T,y)\sigma^2(t,x) = 0, \ t < T, x, y \in \mathbb{R} \\ p(T,x,T,y) = \delta_x(dy) \text{ (mesure de Dirac)} \end{cases} \tag{6.8}$$

on parle d'équation de Kolmogorov "backward". Donnons les arguments pour établir l'équation "forward". Pour une fonction $f$ régulière à support compact, grâce à la formule d'Itô, $f(X^{t,x}_{T+\delta T}) = f(X^{t,x}_T) + \int_T^{T+\delta T} f'(X^{t,x}_u)\mu(u, X^{t,x}_u)du + \int_T^{T+\delta T} f'(X^{t,x}_u)\sigma(u, X^{t,x}_u)dW_u + \frac{1}{2}\int_T^{T+\delta T} f''(X^{t,x}_u)\sigma^2(u, X^{t,x}_u)du \approx f(X^{t,x}_T) + f'(X^{t,x}_T)\mu(T, X^{t,x}_T)\delta T + \sigma(T, X^{t,x}_T)(W_{T+\delta T} - W_T) + \frac{1}{2}\sigma^2(T, X^{t,x}_T)f''(X^{t,x}_T)\delta T$. Il vient donc

$$\frac{\mathbb{E}[f(X^{t,x}_{T+\delta T})] - \mathbb{E}[f(X^{t,x}_T)]}{\delta T} = \mathbb{E}[f'(X^{t,x}_T)\mu(T, X^{t,x}_T) + \frac{1}{2}\sigma^2(T, X^{t,x}_T)f''(X^{t,x}_T)]$$

et par conséquent lorsque $\delta T \downarrow 0$ :

$$\begin{aligned} \int_{\mathbb{R}} f(y)\partial_T p(t,x,T,y)dy &= \int_{\mathbb{R}} \left[f'(y)\mu(T,y) + \frac{1}{2}\sigma^2(T,y)f''(y)\right] p(t,x,T,y)dy \\ &= \int_{\mathbb{R}} f(y)\left[-\partial_y(\mu(T,y)p(t,x,T,y)) + \frac{1}{2}\partial_y^2(\sigma^2(T,y)p(t,x,T,y))\right] dy, \end{aligned}$$

par intégration par parties. On en déduit que $p(t,x,T,y)$ satisfait l'équation de Kolmogorov "forward" ou de Fokker-Planck :

$$\begin{cases} \partial_T p(t,x,T,y) + \partial_y(\mu(T,y)p(t,x,T,y)) - \frac{1}{2}\partial_y^2(\sigma^2(T,y)p(t,x,T,y)) = 0, \ t < T, x, y \in \mathbb{R} \\ p(t,x,t,y) = \delta_x(dy). \text{ (mesure de Dirac)} \end{cases} \tag{6.9}$$

Dans le cas du mouvement brownien ($\mu(t,x) = 0$ et $\sigma(t,x) = 1$), on a $p(t,x,T,y) = \frac{1}{\sqrt{2\pi(T-t)}} \exp\left(-\frac{(y-x)^2}{2(T-t)}\right)$ et les équations de Kolmogorov forward (6.9) et backward (6.8) ne sont alors que des réécritures de la proposition 6.1.1.

Pour conclure cette section, nous pouvons dire que nous avons établi un lien entre les équations différentielles stochastiques et certaines EDP. Nous n'avons présenté ni des preuves ni des hypothèses rigoureuses pour ces résultats car ceci dépasse le cadre du cours. Néanmoins, nous avons donné les principales idées qui permettent d'y arriver et nous avons notamment vu que la formule d'Itô joue un rôle clef pour établir ces EDP.

# Bibliographie


[1] Baldi, P., Mazliak, L., Priouret, P. (2001) Martingales et Chaînes de Markov, *Hermann.*

[2] Cottrell, M., Genon-Catalot, V., Duhamel, C. (1998) Exercices de Probabilités (licence, maîtrise, écoles d'ingénieurs), *Cassini.*

[3] Delmas, J.-F., Jourdain, B. (2006), Modèles aléatoires Applications aux sciences de l'ingénieur et du vivant, *Springer*

[4] Revuz, D. (1997) Mesure et intégration, *Hermann.*

[5] Williams, D. (1991) Probability with Martingales, *Cambridge Mathematical Textbooks.*